ADVANCES IN ATMOSPHERIC SCIENCES, VOL. 29, NO. 2, 2012, 422–434

On the Application of a Genetic Algorithm to the Predictability Problems Involving "On–Off" Switches

ZHENG Qin*[1,2] (郑 琴), DAI Yi[1] (戴 毅), ZHANG Lu[1] (张 露),
SHA Jianxin[1] (沙建新), and LU Xiaoqing [1] (陆小庆)

[1]*Institute of Science, PLA University of Science and Technology, Nanjing* 211101

[2]*State Key Laboratory of Numerical Modeling for Atmospheric Sciences and Geophysical Fluid Dynamics, Institute of Atmospheric Physics, Chinese Academy of Sciences, Beijing* 100029

(Received 30 May 2011; revised 19 October 2011)

ABSTRACT

The lower bound of maximum predictable time can be formulated into a constrained nonlinear optimization problem, and the traditional solutions to this problem are the filtering method and the conditional nonlinear optimal perturbation (CNOP) method. Usually, the CNOP method is implemented with the help of a gradient descent algorithm based on the adjoint method, which is named the ADJ-CNOP. However, with the increasing improvement of actual prediction models, more and more physical processes are taken into consideration in models in the form of parameterization, thus giving rise to the on–off switch problem, which tremendously affects the effectiveness of the conventional gradient descent algorithm based on the adjoint method. In this study, we attempted to apply a genetic algorithm (GA) to the CNOP method, named GA-CNOP, to solve the predictability problems involving on–off switches. As the precision of the filtering method depends uniquely on the division of the constraint region, its results were taken as benchmarks, and a series of comparisons between the ADJ-CNOP and the GA-CNOP were performed for the modified Lorenz equation. Results show that the GA-CNOP can always determine the accurate lower bound of maximum predictable time, even in non-smooth cases, while the ADJ-CNOP, owing to the effect of on–off switches, often yields the incorrect lower bound of maximum predictable time. Therefore, in non-smooth cases, using GAs to solve predictability problems is more effective than using the conventional optimization algorithm based on gradients, as long as genetic operators in GAs are properly configured.

Key words: predictability, on–off switch, conditional nonlinear optimal perturbation (CNOP), genetic algorithm (GA)

Citation: Zheng, Q., Y. Dai, L. Zhang, J. X. Sha, and X. Q. Lu, 2012: On the application of a genetic algorithm to the predictability problems involving "on–off" switches. *Adv. Atmos. Sci.*, **29**(2), 422–434, doi: 10.1007/s00376-011-1054-z.

1. Introduction

According to Lorenz (1975), the issue of predictability of climate falls into two categories. One is bound up with the initial error that the model is supposedly perfect (or very close to perfect), while the other is concerned with the model error in which the initial field is supposedly perfect (Lorenz, 1975; Mu et al., 2010). The first category involves the primary problem of numerical weather and climate prediction. Mu et al. (2002), approached this problem based on actual demands by dividing it into three subproblems: (1) the problem of the lower bound of maximum predictable time; (2) the problem of the upper bound of maximum prediction error, and (3) the problem of the lower bound of maximum allowable initial error and parameter error. These three problems, when utilizing a numerical model to make the prediction, can be formulated into three constrained nonlinear optimization problems. Therefore, it is of theoretical and practical significance for checking how well an optimization algorithm can perform in solving a predictability prob-

*Corresponding author: ZHENG Qin, qinzheng@mail.iap.ac.cn

lem.

To solve the problem of the lower bound of maximum predictable time, the current methods most frequently used are the filtering method and the conditional nonlinear optimal perturbation (CNOP) method. The filtering method refers to the division of a constraint region of the initial perturbation superposed upon initial basic state, after which the maximum predictable time corresponding to every grid point is calculated, and then the minimum value of these maximum predictable times is considered as the lower bound of maximum predictable time. This method is simple and easy to implement. However, it applies only to the theoretical research, partly because its solution to the problem rests upon the step length of the division, and partly because it is negatively affected by high time consumption and low practicability. Based on a simple ENSO Model WF96, Duan and Luo (2010) recently put forward a new solution to the problem of the lower bound of maximum predictable time using the CNOP with a gradient descent algorithm based on the adjoint method, known as the ADJ-CNOP method. This method can first formulate the problem of the lower bound of maximum predictable time into a constrained nonlinear optimization perturbation problem, and then utilize a modified sequence quadratic programming (SQP; Powell, 1982; Mu and Zhang, 2006) optimization algorithm to solve the problem, in which the gradient information required by SQP is provided by integrating backward the relevant adjoint model. The numerical test results in Duan and Luo (2010) verify the feasibility and the effectiveness of this method to determine how much time the calculation takes and how precise the solution is. However, with the increasing improvement of actual prediction models, more and more physical processes are taken into consideration in models in the form of parameterization, thus giving rise to the discontinuousness or non-differentiability in prediction models. This phenomenon is commonly called the on–off switch problem, which tremendously affects the efficiency of the conventional adjoint method (Xu, 1996; Mu and Zheng, 2005; Zheng and Mu, 2006; Zheng and Dai, 2009). In the study of variation data assimilation with on–off switches, it is demonstrated, from the theoretical and numerical experimentation perspective, that the conventional discretization of on–off processes in the governing equation will generate zigzag oscillations in both the numerical solution of the forward model and the associated cost function (CF). Furthermore, the model errors generated by the conventional discretization of on–off processes could, on one hand, make the solution of the corresponding tangent linear model obtained using the conventional approach not a good first-order linear approximation of the nonlinear perturbation solution of the governing equation. On the other hand, it could cause the discrete CF gradient (even one-sided gradient for the CF) at some initial conditions not to exist. At this time, the gradient information supplied by the adjoint model is unable to provide a correct descent direction for the optimization algorithm, which consequently leads to the imperfectness of optimized results. Therefore, it is of vital importance to search for a new global optimization algorithm capable of handling with the constrained optimization problems in non-smooth situations to solve ultimately the problem of the lower bound of maximum predictable time.

Zheng et al. (2011), based on a simple single-grid-line model with discontinuous on–off switches, recently proposed a new genetic algorithm (GA) in which adaptive selection and mutation operators, a blend crossover operator, and the elitist strategy are integrally used. The numerical experiments performed using these elements showed that the new GA is effective to solve the problem of variation data assimilation involving discontinuous on–off switch processes. When a GA is applied to the solution of the constrained optimization problem, dealing with the constraint in a proper way can contribute to the improvement in the performance of the GA. The widely used approach to handling the constraint is the penalty method, the core of which lies in using a penalty term to formulate the constrained optimization problem into an unconstrained one. Fang and Zheng (2009) applied the penalty method to the simple single-grid-line model involving the on–off switch process, with the purpose of studying how effectively a GA can solve the problem of CNOPs. Despite the fact that the penalty method is simple and easy to perform, the optimization results depend largely upon the selection of the penalty parameters. To avoid difficulties of this kind, Deb (2000) devised a constraint handling method based on the tournament selection mechanism and niche strategy for GAs. By putting into use both the GA adopted by Zheng et al. (2011) and the constraint handling method put forward by Deb (2000), this study attempted to solve the problem of the lower bound of maximum predictable time in the model involving discontinuous on–off switch processes and to determine how effectively and feasibly the GA-CNOP works, for the sake of offering a new approach to solving the predictability problems in actual predication models.

This paper is structured as follows: In section 2 the problem of the lower bound of maximum predictable time, the notion of the CNOP, and the approach to utilizing a CNOP to solve the problem of the lower bound of maximum predictable time are briefly introduced. In addition, this section presents the description of the modified Lorenz equation. Section 3 mainly describes the GA-CNOP method. In section 4 the filtering method, the ADJ-CNOP method and GA-CNOP method, respectively, are applied to solve the problem of the lower bound of maximum predictable time in the modified Lorenz model. Then the precision of these three results are compared and analyzed. Finally, section 5 provides a discussion and summary of the study results.

2. Lower bound of maximum predictable time, CNOP, and nonlinear model

2.1 *Lower bound of maximum predictable time*

The lower bound of maximum predictable time is briefly described as follows [see Mu et al. (2002) for details]. Let $\boldsymbol{u}_{\mathrm{tr},0}$ and $\boldsymbol{u}_{\mathrm{tr},t}$ be the true value of the state at the initial time and t time respectively, μ_{tr} the parameter true value, and $\boldsymbol{M}_t$ the nonlinear propagator from the initial time to t time. Under the assumption of the perfect model, we have

$$\boldsymbol{u}_{\mathrm{tr},t} = \boldsymbol{M}_t(\boldsymbol{u}_{\mathrm{tr},0}, \mu_{\mathrm{tr}}) \,. \tag{1}$$

Suppose that the maximum allowable prediction error of a weather or climate event is predetermined to be less than or equal to $\varepsilon>0$ (the allowable prediction precision). Then the maximum predictable time T corresponding to the initial observation $\boldsymbol{u}_{\mathrm{obs},0}$ and the first guess of the model parameter μ_{g} are defined as follows:

$$T = \max\{t | \|(\boldsymbol{M}_\tau(\boldsymbol{u}_{\mathrm{obs},0}, \mu_{\mathrm{g}}) - \boldsymbol{u}_{\mathrm{tr},\tau}\| \leqslant \varepsilon \,, \ \ 0 \leqslant \tau \leqslant t\} \,, \tag{2}$$

where $||\cdot||$ is the norm measuring the prediction error, which is taken as the L^2 norm in this study. However, in realistic problems, it is impossible to obtain the true value of the state $\boldsymbol{u}_{\mathrm{tr}}$; instead, the information about the errors in the initial observations and the initial assumed values of the parameters can be known with the following levels of tolerance:

$$\|\boldsymbol{u}_{\mathrm{tr},0} - \boldsymbol{u}_{\mathrm{obs},0}\|_{\mathrm{A}} \leqslant \sigma_1 \,, \quad \|\mu_{\mathrm{tr}} - \mu_{\mathrm{g}}\|_{\mathrm{B}} \leqslant \sigma_2 \,, \tag{3}$$

where $\|\cdot\|_{\mathrm{A}}$ and $\|\cdot\|_{\mathrm{B}}$ are norms measuring the errors in the initial conditions and parameters of the model. Therefore, Mu et al. (2002) presented a lower bound estimation for the maximum predictable time T:

$$T_{\mathrm{l}} = \min_{\boldsymbol{u}_0 \in \mathrm{B}_{\sigma_1}, \mu \in \mathrm{B}_{\sigma_2}} \{T_{\boldsymbol{u}_0,\mu} | T_{\boldsymbol{u}_0,\mu} = \max t : \|\boldsymbol{M}_\tau(\boldsymbol{u}_0, \mu) - \boldsymbol{M}_\tau(\boldsymbol{u}_{\mathrm{obs},0}, \mu_{\mathrm{g}})\| \leqslant \varepsilon, 0 \leqslant \tau \leqslant t\} \,, \tag{4}$$

where B_{σ_1} and B_{σ_2} are constraint balls with centers at $\boldsymbol{u}_{\mathrm{obs},0}, \mu_{\mathrm{g}}$, and radii $\sigma_1 > 0, \sigma_2 > 0$, respectively. Because $\boldsymbol{u}_{\mathrm{tr},0} \in \mathrm{B}_{\sigma_1}$, $\mu_{\mathrm{tr}} \in \mathrm{B}_{\sigma_2}$, the maximum predictable time T corresponding to the initial observation $\boldsymbol{u}_{\mathrm{obs},0}$ and first guess of the model parameter μ_{g} satisfies $T_{\mathrm{l}} \leqslant T$.

2.2 *Conditional nonlinear optimal perturbation (CNOP)*

Nonlinearity effect must be taken into consideration when studying the predictability and stability of the atmospheric and oceanic motion. Thus, Mu et al. proposed the notion and theory of CNOP (Mu et al., 2003). A CNOP refers to the initial perturbation, among all the initial perturbations that satisfy certain constraint conditions, which has the largest nonlinear evolution at the end of the time period of concern and is an extension of the linear singular vector to the nonlinear case.

Assume the model simulating atmosphere or ocean motion as follows:

$$\begin{cases} \dfrac{\partial \boldsymbol{w}}{\partial t} + \boldsymbol{F}(\boldsymbol{w}) = 0, \\ \boldsymbol{w}|_{t=0} = \boldsymbol{w}_0, \end{cases} \quad \text{in} \quad \Omega \times [0, T] \,, \tag{5}$$

where $\boldsymbol{w}(\boldsymbol{x}, t) = (w_1(\boldsymbol{x}, t), w_2(\boldsymbol{x}, t), \cdots, w_m(\boldsymbol{x}, t))^{\mathrm{T}}$ is a m-dimension vector; the superscript T represents transpose; $\boldsymbol{F}$ is the nonlinear operator; $\boldsymbol{w}_0(\boldsymbol{x})$ is the initial state; Ω is a domain of $R^n; (\boldsymbol{x}, t) \in \Omega \times [0, T]$. $\boldsymbol{x} = (x_1, x_2, \cdots, x_n)$ and t are respectively the spatial and temporal variables; $t = 0$ is the initial time and $t = T$ with $T < +\infty$ is a future time. If $\boldsymbol{M}_t$ denotes the propagator of model (5) from 0 to $t (0 \leqslant t \leqslant T)$, then the solution of model (5) at time t can be given by

$$\boldsymbol{w}(\boldsymbol{x}, t) = \boldsymbol{M}_t(\boldsymbol{w}_0) \,. \tag{6}$$

When superposing an initial perturbation $\delta \boldsymbol{w}_0(\boldsymbol{x})$ upon $\boldsymbol{w}_0(\boldsymbol{x})$ and denoting $\delta \boldsymbol{w}(\boldsymbol{x}, t)$ the nonlinear evolution of the initial perturbation with time t, we have

$$\delta \boldsymbol{w}(\boldsymbol{x}, t) = \boldsymbol{M}_t(\boldsymbol{w}_0 + \delta \boldsymbol{w}_0) - \boldsymbol{M}_t(\boldsymbol{w}_0) \,. \tag{7}$$

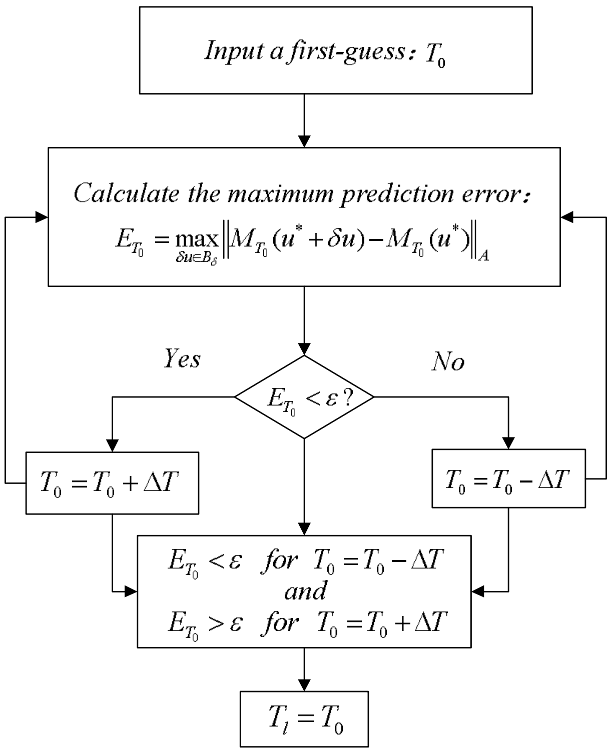

Fig. 1. Flow chart for solving the lower bound of maximum predictable time.

With a specified norm $\|\cdot\|$ measuring the nonlinear evolution of the initial perturbation $\delta \boldsymbol{w}(\boldsymbol{x}, t)$, the objective function is defined as follows:

$$J(\delta \boldsymbol{w}_0) = \|\boldsymbol{M}_T(\boldsymbol{w}_0 + \delta \boldsymbol{w}_0) - \boldsymbol{M}_{\mathrm{T}}(\boldsymbol{w}_0)\| \ . \quad (8)$$

The perturbation $\delta \boldsymbol{w}_0^*$ satisfying

$$J(\delta \boldsymbol{w}_0^*) = \max_{\|\delta \boldsymbol{w}_0\| \leqslant \sigma} J(\delta \boldsymbol{w}_0) \ , \quad (9)$$

is called a CNOP at prediction time T, where $\sigma>0$ is the radius of the ball constraining initial perturbations.

2.3 *Attaining the lower bound of maximum predictable time by means of CNOPs*

Recently, Duan and Luo (2010) devised a numerical scheme (Fig. 1) to solve the lower bound of maximum predictable time in their predictability study.

For a given first guess T_0 of T_1, a constrained nonlinear optimal algorithm, such as the spectral projected gradient method (version 2, SPG2; Birgin et al., 2000) or the SQP, was used to calculate the maximum prediction error at T_0 in the constraint region B_σ of the initial error, noted as E_{T_0}. If $E_{T_0}>\varepsilon$ (ε represents the allowable prediction precision), we tried a smaller $T_0 = T_0 - \Delta T (\Delta T>0$ is a specified constant), and calculated the maximum prediction error at the update time T_0. If $E_{T_0}<\varepsilon$, then we tried a larger $T_0 = T_0 + \Delta T$ and calculated the maximum prediction error at the update time T_0. The procedure lasted until T_0 satisfied both $E_{T_0+\Delta T}>\varepsilon$ and $E_{T_0-\Delta T}\leqslant\varepsilon$. The corresponding T_0 was considered to be the lower bound of maximum predictable time satisfying the allowable prediction precision ε under the constraint of the given initial error.

According to the definition of a CNOP, the initial error that causes the maximum prediction error E_{T_0} at an update time T_0 is just a CNOP in the constraint ball B_δ. Thus, the problem of solving the maximum prediction error can be reduced to the optimization problem of searching for the CNOP.

Duan and Luo (2010) used the modified SQP solver based on the adjoint method to capture the CNOP and ultimately obtained the lower bound of maximum predictable time, and this method is called the ADJ-CNOP.

2.4 *Modified Lorenz model*

The following modified Lorenz model was adopted by Xu and Gao (1999) to study the influence of on–off switches on the conventional adjoint minimization:

$$\begin{cases} \dfrac{dx}{dt} = -ax + ay \\ \dfrac{dy}{dt} = -xz + rx - y \\ \dfrac{dz}{dt} = xy - bz \end{cases} , \quad (10)$$

where $r = r_0 + H(c)r_1$ is the modified Rayleigh number that contains a jump controlled by the threshold condition $c = y - y_c>0$, and $H(\cdot)$ is the Heaviside unit-step function, which simulates on–off switches in the parameterized processes and is defined as following:

$$H(x) = \begin{cases} 1, \ x > 0 \\ 0, \ x \leqslant 0 \end{cases} , \quad (11)$$

where a and b are related to the Prandtl number and the aspect ratio geometry (in the original model), respectively. In this study, we utilize this model with parameters $a = 10, b = 8/3, r_0 = 28$, and $r_1 = -18$ to study the feasibility and effectiveness of the GA-CNOP method to determine the lower bound of maximum predictable time.

If we let the three equations in model (10) equal zero, we get the following three stationary points:

$$\begin{cases} o : (x, y, z) = (0, 0, 0) \\ c_1 : (x, y, z) = (-\sqrt{b(r-1)}, -\sqrt{b(r-1)}, r-1) \\ c_2 : (x, y, z) = (\sqrt{b(r-1)}, \sqrt{b(r-1)}, r-1) \ . \end{cases} \quad (12)$$

When the forward difference scheme with the traditional numerical treatment of on–off switches is used to discretize model (10), we get the following numerical model:

$$\begin{cases} x_{k+1} = x_k + \Delta t(-ax_k + ay_k) \\ y_{k+1} = y_k + \Delta t(-x_k z_k + rx_k - y_k) \\ x_{k+1} = z_k + \Delta t(-x_k y_k - bz_k) \end{cases} \quad (13)$$

where $k = 0, 1, \cdots, N$, and N is the number of integral steps with time step $\Delta t = 0.005$. In the original Lorenz equation, the behavior of the solution changes with the model parameters. When $1<r<24.74$, stationary points c_1 and c_2 are stable; when $r > 24.74$, the stationary points become unstable (Kalnay, 2003). Figures 2 and 3 demonstrate that stationary point c_1 changes its regime from unstable to stable when r decreases from 28 to 10.

3. The genetic strategy in the GA-CNOP method

A GA is a global optimization algorithm inspired by the biological heredity mechanism and the natural selection principle, in which the group search strategy and optimization procedure are independent of gradient information. This guarantees its superiority to the conventional optimization algorithms in universality and effectiveness when used to handle constrained optimization problems. When applied to an optimization problem, a standard GA performs the following operations: encoding, initializing the population, evaluating the fitness of populations, and evolving populations. During the evolution processes, three kinds of genetic operators (i.e., selection, crossover, and mutation) are at work. The different configurations of genetic operators in a GA exert a great influence on the performance of the GA. [For a detailed description of a standard GA manipulation, refer to Barth (1992) and Zheng et al. (2011).]

GAs exhibit three superior properties when they are used to solve the CNOP problem. First, GAs do not depend on any gradient information relevant to objective functions, thus they are still applicable when the non-differentiability or even discontinuity of the objective functions occur due to the physical parameterization in actual models. Second, capturing a CNOP belongs to a constrained nonlinear optimization problem in which GAs continue to work very well no matter how constrained the conditions are. And third, GAs have an inherent parallel computation characteristic in processing information. Thus, the computational efficiency of a GA will be effectively improved by devising a parallel execution strategy and constructing the corresponding parallel algorithm.

GAs deal with constraint conditions using many approaches: the space constraint method, the rejection method, the restoration method, the modified operator method, the penalty method, to name but a few. Each method has its own limitations, however. In other words, until now there has not been a universal method that can deal with all kinds of constraint conditions in an effective manner. Deb (2000) proposed a constraint-handling method based on the penalty function approach, in which the tournament selection mechanism and niche strategy are adopted in the selection operation of a GA. Its benefits are obvious, partly because it is unnecessary to adjust the penalty parameters by trail and error, and partly because it is easy to handle and implement. Deb's comparison criteria in the tournament selection operator are as follows (Deb, 2000):

(1) Any feasible solution is preferred to any infeasible solution.

(2) Among two feasible solutions, the one having better objective function value is preferred.

(3) Among two infeasible solutions, the one having smaller constraint violation is preferred.

To maintain diversity among feasible solutions, Deb (2000) used a simple niche strategy in the tournament selection operator, which can be described in the following way: Let $\bar{d}$ and n_f respectively be a specific critical distance and the number of feasible solutions to be checked. When comparing two feasible solutions $\boldsymbol{X}_i = (x_{i,1}, x_{i,2}, \cdots, x_{i,m})$ and $\boldsymbol{X}_j = (x_{j,1}, x_{j,2}, \cdots, x_{j,m})$, where m is the gene number of the individual, the normalized Euclidean distance

$$d_{ij} = \sqrt{\frac{1}{m}\sum_{l=1}^{m}\left(\frac{x_{i,l} - x_{j,l}}{\bar{x}_l - \underline{x}_l}\right)^2}$$

is measured between them, where $\bar{x}_l$ and $\underline{x}_l (1 \leqslant l \leqslant m)$ are specified constants and stand respectively for the upper and lower bounds of the lth variable, i.e., $\underline{x}_l \leqslant x_{i,l}, x_{j,l} \leqslant \bar{x}_l (1 \leqslant l \leqslant m)$. If $d_{ij} < \bar{d}$, the solutions $\boldsymbol{X}_i$ and $\boldsymbol{X}_j$ are compared with their objective function values. Otherwise, they are not compared, and another solution $\boldsymbol{X}_j$ is checked. If n_f feasible solutions are checked and none is found to qualify within the critical distance, the solution $\boldsymbol{X}_i$ is declared the winner.

This kind of selection mechanism does not need to take both the objective function and the extent of constraint violation into consideration, and it does not need to calculate the objective function value for infeasible solutions, which will be used in the GA-CNOP.

The fitness function in the GA-CNOP is specified

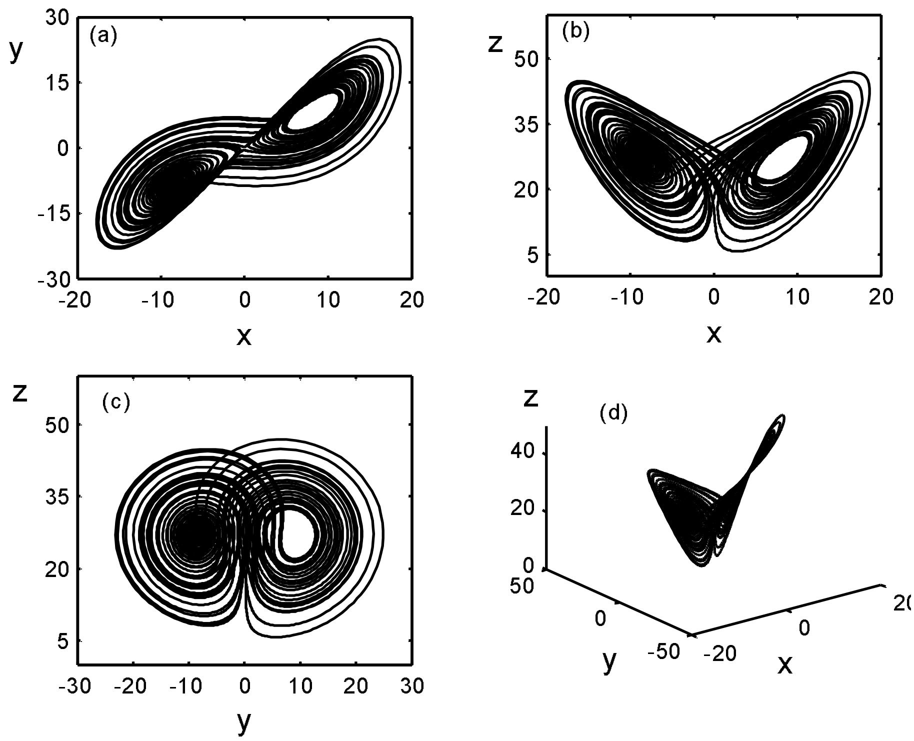

Fig. 2. Evolution of the model solution with time, where $r = 28$ and the initial condition is $(-8.48, -8.48, 27)$, which corresponds to c_1 superposed an initial perturbation.

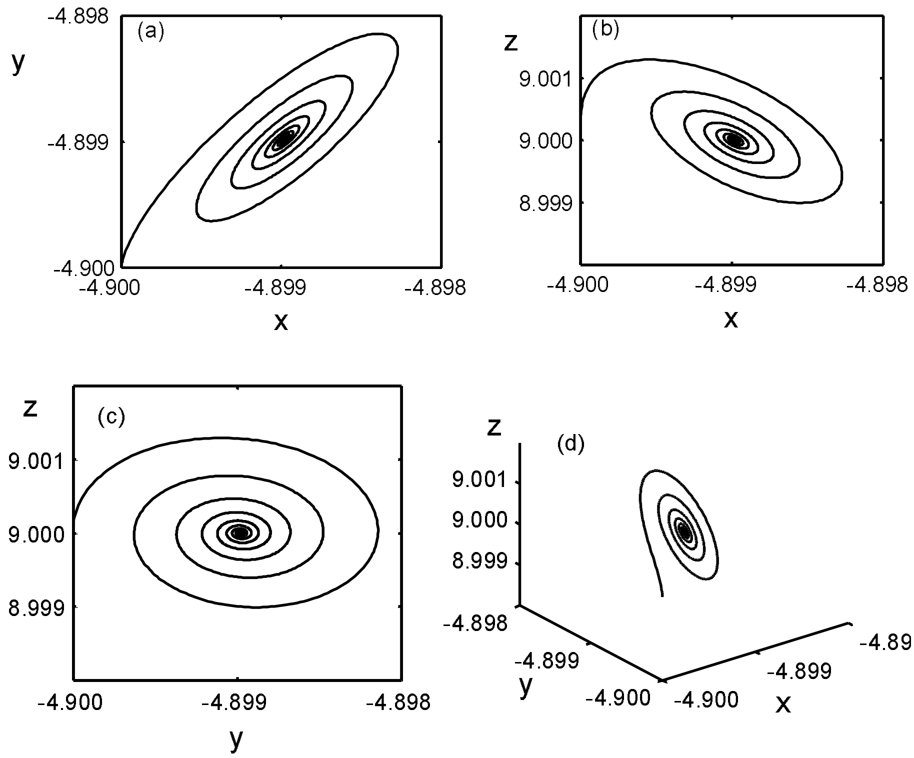

Fig. 3. Same as Fig. 2, except with $r = 10$ and the initial condition $(-4.49, -4.89, 9)$.

as

$$f = J(\delta \boldsymbol{X}(0)) = \|(\delta \boldsymbol{X}(T))\|_2^2 \ , \tag{14}$$

i.e., the fitness function is taken as the objective function, where $\delta \boldsymbol{X}(0) = (\delta x(0), \delta y(0), \delta z(0))$ is the initial perturbation, $\delta \boldsymbol{X}(T) = (\delta x(T), \delta y(T), \delta z(T)) = \boldsymbol{M}_T(\boldsymbol{X}_0 + \delta \boldsymbol{X}(0)) - \boldsymbol{M}_T(\boldsymbol{X}_0)$ denotes the nonlinear evolution of $\delta \boldsymbol{X}(0)$, $\boldsymbol{X}_0$ is a given initial basic state, and $\boldsymbol{M}_T$ is the nonlinear propagator of model (13), $\|\delta \boldsymbol{X}(T)\|_2^2 = \delta x^2(T) + \delta y^2(T) + \delta z^2(T)$.

The iteration procedure that uses the GA with the particular genetic operators to locate the CNOP of the model (13) can be illustrated as follows.

Step0. Encoding and population initialization

Real encoding is adopted in this study, i.e., all optimization variables (initial perturbations) are coded in decimal numbers, which is the direct description of optimization problems.

Set generation $k = 0$ and initialize population $\Gamma(0)$, which is a set of possible solutions (i.e., initial perturbation guesses),

$$\Gamma(0) = \left[\delta \boldsymbol{X}_{1(0)}, \delta \boldsymbol{X}_{2(0)}, \ldots, \delta \boldsymbol{X}_{n(0)}\right] \ , \tag{15}$$

$$\delta \boldsymbol{X}_{i(0)} = (\delta x_{i(0)}(0), \delta y_{i(0)}(0), \delta z_{i(0)}(0)) \ , \tag{16}$$

where $\delta \boldsymbol{X}_{i(0)}$ denotes the ith individual of the initial generation and n is the size of population. There are two ways to initialize population (i.e., absolute stochastic initialization and stochastic initialization combined with prior knowledge), and the first was used in our numerical experiments, i.e., $\delta x_{i(0)}(0), \delta y_{i(0)}(0)$, and $\delta z_{i(0)}(0)$ were randomly generated in their constraint range. Notably, however, the key of initialization is the diversity of population, that is to say, the initial population should cover the whole solution space.

Step 1. Evaluating the population fitness

Calculate the objective function values $J\left(\delta \boldsymbol{X}_{i(k)}\right)$, $i = 1, \cdots, n$ according to Eq. (14) by integrating the model (13) with the initial condition $\boldsymbol{X}_0 + \delta \boldsymbol{X}_{i(k)}$, and select the best individual $\delta \boldsymbol{X}^*_{(k)}$ of current population:

$$\delta \boldsymbol{X}^*_{(k)} = \arg \max_{1 \leqslant i \leqslant n} J\left(\delta \boldsymbol{X}_{i(k)}\right) \ , \tag{17}$$

Step 2. Determine whether $\delta \boldsymbol{X}^*_{(k)}$ satisfies the stop criterion of

$$\sum_{j=k-\nu}^{k-1} \left| J(\delta \boldsymbol{X}^*_{(j+1)}) - J(\delta \boldsymbol{X}^*_{(j)}) \right| < \gamma \quad \text{or} \quad k \geqslant k_{\mathrm{m}} \ , \tag{18}$$

where ν is a given positive integer, $\gamma>0$ is a real number, k_{m} is the specified maximum generation, and $k_{\mathrm{m}} = 500$ in our numerical experiments. If Eq. (18) is satisfied, then output $\delta \boldsymbol{X}^*_{(k)}$, which is the CNOP of the model (13) corresponding to the initial basic state $\boldsymbol{X}_0$, and stop the procedure. If not, go to step 3.

Step 3. Evolving populations

The population evolution is carried out by the following genetic operators.

(1) Selection operator

A selection operator determines whether an individual is selected for the next operation. Deb's tournament selection mechanism and niche strategy were used in selection operators. In detail, two individuals are picked at random from the current population and are compared based on Deb's comparison criteria and niche strategy. The better solution (or the winner) is chosen and kept in the intermediate population. This process is continued until all n population slots are filled.

(2) Crossover operator

A GA implements the crossover operation according to crossover probability. In this study, the following self-adaptive crossover probability proposed by Srinivas and Patnaik (1994) was adopted:

$$P_{\mathrm{c}} = \begin{cases} P_{\mathrm{c1}} - \dfrac{(P_{\mathrm{c1}} - P_{\mathrm{c2}})(f' - f_{\mathrm{avg}})}{f_{\max} - f_{\mathrm{avg}}}, & f' \geqslant f_{\mathrm{avg}} \\ P_{\mathrm{c1}}, & f' < f_{\mathrm{avg}} \end{cases} , \tag{19}$$

where $f_{\max}$ and f_{avg} are the maximum fitness value and average fitness value of the present generation, f' is the bigger fitness value of the two individuals to be crossed over, $P_{\mathrm{c1}} = 0.95$, $P_{\mathrm{c2}} = 0.8$ in this study.

The crossover operator imitates the gene recombination process of natural sexual breeding, and it is a main search operator in a GA because it exploits the available information in the previous population to influence the future search. As a result, many different types of crossover operators, such as two-point crossover, uniform crossover, arithmetical crossover, geometrical crossover, simulated binary crossover (SBX), BLX-α, and so on, have been proposed in different contexts. Considering the high quality of search ability which it has for multimodal functions, the BLX-α operator proposed by Eshelman and Schaffer (1992) was adopted in the GA-CNOP.

The BLX-α operates on two parent individuals. For the two selected parent individuals, $\delta \boldsymbol{X}_i^{(k)} = (\delta x_i^{(k)}(0),\ \delta y_i^{(k)}(0),\ \delta z_i^{(k)}(0))$ and $\delta \boldsymbol{X}_j^{(k)} = (\delta x_j^{(k)}(0), \delta y_j^{(k)}(0), \delta z_j^{(k)}(0))$, denoting respectively $u_{i,l}^{(k)}$ and $u_{j,l}^{(k)}$ their lth genes, where k is the generation number. To determine whether or not the crossover between $u_{i,l}^{(k)}$ and $u_{j,l}^{(k)}$ is to be performed, a uniform random real number R in $[0,1]$ is chosen. If $R<P_{\mathrm{c}}$ (the crossover probability corresponding to $\delta \boldsymbol{X}_i^{(k)}$ and $\delta \boldsymbol{X}_j^{(k)}$), the crossover between $u_{i,l}^{(k)}$ and $u_{j,l}^{(k)}$ is operated as follows.

Two uniform random numbers R_1 and R_2 are generated in interval $\left[u_{j,l}^{(k)} - \alpha\left(u_{i,l}^{(k)} - u_{j,l}^{(k)}\right),\right.$ $\left.u_{j,l}^{(k)} + \alpha\left(u_{i,l}^{(k)} - u_{j,l}^{(k)}\right)\right]$ for $u_{j,l}^{(k)} < u_{i,l}^{(k)}$ or in $\left[u_{i,l}^{(k)} - \alpha\left(u_{j,l}^{(k)} - u_{i,l}^{(k)}\right),\quad u_{i,l}^{(k)} + \alpha(u_{j,l}^{(k)} - u_{i,l}^{(k)})\right]$ for $u_{i,l}^{(k)} < u_{j,l}^{(k)}$, and they are respectively used as the lth genes $u_{i,l}^{(k+1)}$ and $u_{j,l}^{(k+1)}$ of the offspring individuals, i.e., $u_{i,l}^{(k+1)} = R_1$, $u_{j,l}^{(k+1)} = R_2$, where $\alpha > 0$ is a constant and $\alpha = 0.5$ normally.

If $R \geqslant P_{\rm c}$, $u_{i,l}^{(k)}$ and $u_{j,l}^{(k)}$ are directly copied to $u_{i,l}^{(k+1)}$ and $u_{j,l}^{(k+1)}$, respectively.

After three operations, offspring individuals $\delta \boldsymbol{X}_i^{(k+1)}$ and $\delta \boldsymbol{X}_j^{(k+1)}$ can be obtained.

(3) Mutation operator

A GA implements mutation operation according to mutation probability also. In this study, we adopted the following self-adaptive mutation probability (Srinivas and Patnaik, 1994):

$$P_{\rm m} = \begin{cases} P_{\rm m1} - \dfrac{(P_{\rm m1} - P_{\rm m2})(f - f_{\rm avg})}{f_{\max} - f_{\rm avg}}, & f \geqslant f_{\rm avg} \\ P_{\rm m1}, & f < f_{\rm avg} \end{cases}, \tag{20}$$

where $f_{\max}$ and $f_{\rm avg}$ are the same as in crossover probability, f is the fitness of the individual going to mutate, $P_{\rm m1} = 0.1$, $P_{\rm m2} = 0.01$ in this study.

The mutation operator simulates the mutation of some gene in a given chromosome during biological evolution, which introduces new information and maintains diversity in the population, thus preventing the search process from plunging into a local minimum. The non-uniform operator, which is considered one of the most suitable mutation operators for real encoding of GAs (Herrera et al., 1998) and is used in the GA-CNOP, is described as follows.

For a given parent individual $\delta \boldsymbol{X}_i^{(k)} = (\delta x_i^{(k)}(0), \delta y_i^{(k)}(0), \delta z_i^{(k)}(0))$, based on the circulation of gene bits to determine whether its each gene has mutated or not according to mutation probability, a uniform random real number R in $[0,1]$ is first chosen. If $R \geqslant P_{\rm m}$, then the lth gene $u_{i,l}^{(k)}$ does not mutate, and it is directly copied to the next generation; otherwise the operation (i.e., $R < P_{\rm m}$) creates a mutation according to one of following formulas:

$$u_{i,l}^{(k)} + \Delta\left(k, \bar{x}_l - u_{i,l}^{(k)}\right) \quad \text{or} \quad u_{i,l}^{(k)} - \Delta\left(k, u_{i,l}^{(k)} - \underline{x}_l\right), \tag{21}$$

where $\bar{x}_l$ and $\underline{x}_l$ are the upper bound and lower bound of $u_{i,l}^{(k)} (i = 1, 2, \cdots, n)$, which can be determined by the range of the corresponding physical quantity, also by the problem itself. $\Delta(k, y)$ is given by

$$\Delta(k, y) = y\upsilon\left(1 - k/\omega\right)^{\vartheta}, \tag{22}$$

where υ is a random number on the interval $[0,1], \omega$ is the maximum genetic generation number, and ϑ is the parameter to determine the non-uniform degree, which is 2 in this study.

4. Numerical experiments and their results analyses

To demonstrate the effectiveness of the GA-CNOP method for solving the problem of the lower bound of maximum predictable time in discontinuous cases, the numerical experiments using the filtering method, the ADJ-CNOP and the GA-CNOP, respectively, were conducted. Because the precision of the filtering method depends uniquely on the division of the constraint region, its results were taken as benchmarks, and the results yielded by the ADJ-CNOP and the GA-CNOP were compared with them.

Because this study focused on the impacts of initial errors on the maximum predictable time, the model parameter errors were neglected, and the lower bound estimation for the maximum predictable time T [defined in Eq. (4) for Lorenz model (13)] became

$$T_{\rm l} = \min_{\|\delta \boldsymbol{X}(0)\|_{\rm A} \leqslant \sigma} \left\{T_{\delta \boldsymbol{X}(0)} | T_{\delta \boldsymbol{X}(0)} = \max t : \|\boldsymbol{M}_\tau(X_0 + \delta \boldsymbol{X}(0)) - \boldsymbol{M}_\tau(X_0)\|_2 \leqslant \varepsilon, 0 \leqslant \tau \leqslant t\right\}. \tag{23}$$

When the norm measuring the initial perturbation is taken as the two norm, the constraint

$$\|\delta \boldsymbol{X}(0)\|_2 = \sqrt{\delta x^2(0) + \delta y^2(0) + \delta z^2(0)} \leqslant \sigma$$

is called a ball constraint. When the norm is taken as the infinite norm, the constraint $\|\delta \boldsymbol{X}(0)\|_\infty = \max\{|\delta x(0)|, |\delta y(0)|, |\delta z(0)|\} \leqslant \sigma$ is called a box constraint.

When the filtering method was used, cubic meshes of certain sizes were used to discretize the constraint domain of initial perturbations $\delta \boldsymbol{X}(0)$. In the context of a ball constraint, the constraint domain is a ball, and the circumscribed cube of this ball is considered. Each mesh point outside the ball is connected with the center of the ball (the origin), and then the intersection point of this line with the boundary of the ball replaces the mesh point outside the ball. From every initial perturbation $\delta \boldsymbol{X}(0)$ corresponding to a mesh point inside the constraint ball or an intersection point, a maximum predictable time $T_{\delta \boldsymbol{X}(0)}$ was obtained by integrating discretized Lorenz model (13) and using the following formula:

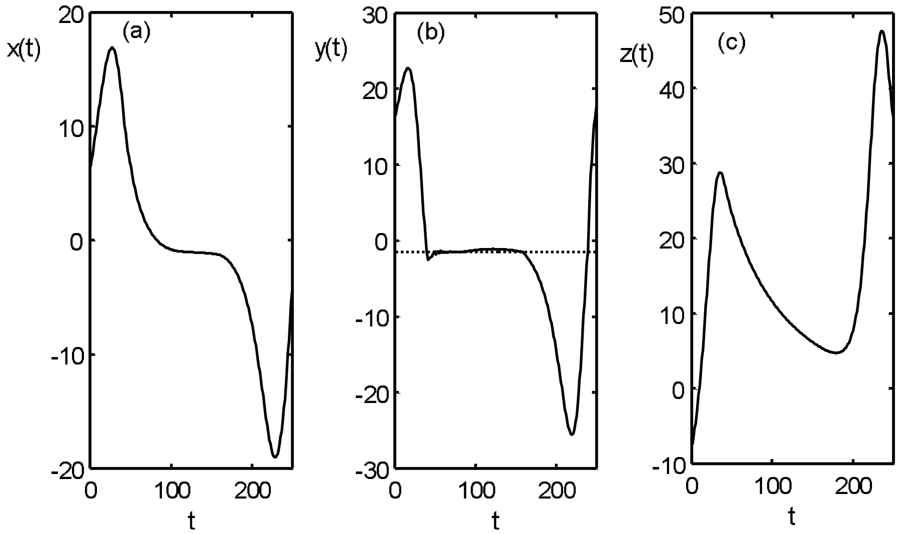

Fig. 4. Time evolutions of three components of the numerical solution of model (13), with initial basic state $\boldsymbol{A}(6.0,\ 16.0,\ -8.0)$, in which the horizontal dotted line denotes the threshold. From the second patch corresponding to y, it can be clearly seen that on–off switches are triggered from 40th time step.

$$\{T_{\delta \boldsymbol{X}(0)} = \max t : \|\boldsymbol{M}_\tau(X_0 + \delta\boldsymbol{X}(0)) - \boldsymbol{M}_\tau(X_0)\|_2 \leqslant \varepsilon, 0 \leqslant \tau \leqslant t\}\ .$$

For all of these maximum predictable times, the smallest one was considered the lower bound of maximum predictable time T_l. Obviously, the precision of the filtering method depends upon the cubic-mesh size (cube-mesh length), which was set as 0.001 in the numerical experiments.

The relevant parameters used in the numerical experiments were as follows: the allowable prediction precision ε was taken as 0.6, 1.0, 1.4, 1.8, 2.2 respectively; the radius σ of the constraint ball of initial perturbations was taken as 0.005, 0.01, 0.02, 0.04, 0.08, and 0.16 respectively; and the population size was 60.

When the initial basic state was stationary point $\boldsymbol{X}_0 = O$ and threshold was $y_c = -1.5$, to every value of ε and σ, with either the ball constraint or the box constraint, the ADJ-CNOP and GA-CNOP both gave the same lower bound of maximum predictable time, which is consistent with the result of the filtering method with sufficiently fine division. We took the CNOP, corresponding to the lower bound of maximum predictable time in the numerical experiment, as an initial perturbation, and superposed it upon the initial basic state and integrated the model (13). In this case, because the model solution did not trigger "switches", the ADJ-CNOP and GA-CNOP could solve both prediction problems effectively.

To study the influence of switches on the ADJ-CNOP and GA-CNOP, the initial basic state was taken at $\boldsymbol{X}_0 = \boldsymbol{A} = (6.0, 16.0, -8.0)$, and threshold was still $y_c = -1.5$. We plotted the time evolution of x, y, z components of the model solution (Fig. 4), from which we can see that the on–off switches were triggered from the 40th time step.

For the initial state $\boldsymbol{A} = (6.0, 16.0, -8.0)$ and different values of ε and σ, we tested the performance of the ADJ-CNOP and GA-CNOP in yielding the lower bound of maximum predictable time under the ball constraint, and we compared the results with those of the filtering method. The following three tables demonstrate the results of the three methods.

The numbers in bold italic in tables show distinct lower bounds of maximum predictable time compared with those of the filtering method. The results show that when the model contained discontinuous on–off switches, the results of the GA-CNOP were almost consistent with those of the filtering method, except for a one-step difference in the lower bound of maximum predictable time at some points, which stems from the one-step precision of the filtering method. However, in the results listed Table 3, the ADJ-CNOP only shares 50% of the same lower bounds of maximum predictable time with those of the filtering method, while the remaining 50% of results are all bigger, even reaching the biggest 31 division steps corresponding to $\sigma = 0.04$, $\varepsilon = 2.2$. To further test the correctness of the lower bound of maximum predictable time via the

Table 1. The lower bounder of maximal prediction time yielded by the filtering approach under the ball constraint.

ε	$\delta = 0.005$	$\delta = 0.01$	$\delta = 0.02$	$\delta = 0.04$	$\delta = 0.08$	$\delta = 0.16$
0.6	192	192	192	40	40	40
1.0	203	203	203	40	40	40
1.4	211	211	211	198	198	190
1.8	221	221	220	203	202	195
2.2	239	239	239	208	207	199

Table 2. Same as Table 1 except the lower bounder of maximum predictability time is yielded by the GA-CNOP, in which the numbers in bold italic show distinct lower bounds of maximum predictable time compared with those by the filtering method.

ε	$\delta = 0.005$	$\delta = 0.01$	$\delta = 0.02$	$\delta = 0.04$	$\delta = 0.08$	$\delta = 0.16$
0.6	192	192	***191***	40	40	40
1.0	203	203	***202***	40	40	40
1.4	211	211	211	***197***	***197***	190
1.8	220	220	***219***	203	202	195
2.2	239	239	239	208	***206***	199

Table 3. Same as Table 2 except the lower bounder of maximum predictability time is yielded by the ADJ-CNOP.

ε	$\delta = 0.005$	$\delta = 0.01$	$\delta = 0.02$	$\delta = 0.04$	$\delta = 0.08$	$\delta = 0.16$
0.6	192	***196***	***195***	40	40	40
1.0	203	***209***	***205***	40	40	40
1.4	211	***215***	***216***	***215***	***200***	***197***
1.8	221	***226***	***221***	***211***	***203***	***196***
2.2	239	239	239	***239***	***209***	***208***

GA-CNOP method, we used the CNOPs, obtained at every time $n\Delta t$, $n = 0, 1, \cdots, N$ of the three methods, as initial perturbations, and superposed them on the initial basic state $\boldsymbol{X}_0 = \boldsymbol{A}$ to integrate the model (13). Thus we obtained the prediction errors at every time increment and their time evolutions (Fig. 5). It is obvious that the prediction error reached the limit $\varepsilon = 1.4$ at $179\Delta t$ for the CNOP gained from the GA-CNOP, while for the CNOP gained from the ADJ-CNOP, the prediction error continued to search forward without reaching the limit and eventually output the incorrect lower bound of maximum predictable time, $T = 215\Delta t$.

These results were calculated under the ball constraint. To reveal the effectiveness of the GA-CNOP in attaining the lower bound of maximum predictable time in discontinuous cases and its independence of the form of a constraint condition, we further implemented numerical experiments with the box constraints using the filtering method, the ADJ-CNOP method, and the GA-CNOP method with the same parameters used for the ball constraint. Tables 4, 5, and 6 show the related results, respectively.

Test results listed in Tables 4, 5, and 6 reveal that, under the box constraint, the GA-CNOP method was also able to obtain the precise lower bound of maximum predictable time, while the ADJ-CNOP method yielded the biggest deviation of 155 time steps of the lower bound of maximum predictable time with respect to that of the filtering method, which is true for $\sigma = 0.04$, $\varepsilon = 1.0$. In addition, similar to the analysis shown in Fig. 5, Fig. 6 shows the time evolutions of prediction errors based on the filtering method, the ADJ-CNOP method, and the GA-CNOP method under the box constraint. It can be seen clearly that prediction error reached the limit $\varepsilon = 1.0$ at $40\Delta t$ time for the CNOP gained from the GA-CNOP, while for the CNOP gained from the ADJ-CNOP, the prediction error continued to search forward without reaching the limit and eventually output the incorrect lower bound of maximum predictable time, $T = 195\Delta t$.

5. Discussion and conclusion

Using the modified Lorenz equation as the prediction model and taking the results gained by using the filtering method as benchmarks, this study investigated how effectively the ADJ-CNOP method and the GA-CNOP method can solve the problem of the lower bound of maximum predictable time in non-smooth cases. Numerical experiment results show that, due to the effects of discontinuous on–off switches, the ADJ-

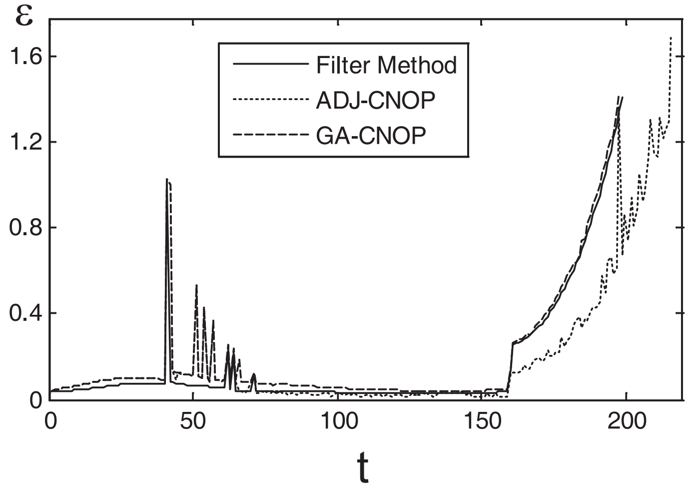

Fig. 5. Time evolutions of the prediction errors based on the CNOPs gained with the three methods under the ball constraint, in which $\sigma = 0.04$, $\varepsilon = 1.4, y_{\rm c} = -1.50$.

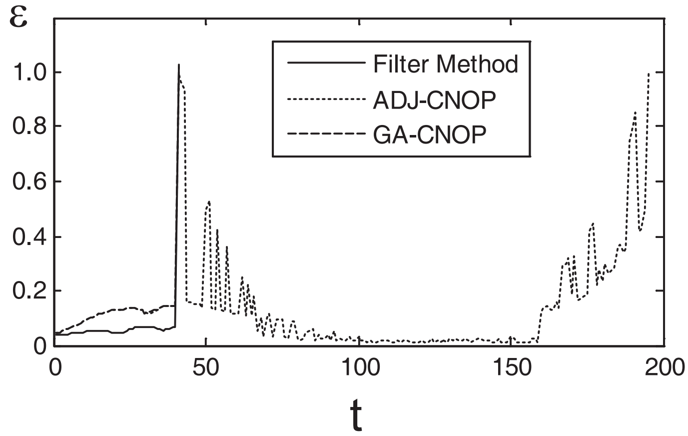

Fig. 6. Same as Fig. 5 except $\sigma = 0.04$, $\varepsilon = 1.0$, $y_{\rm c} = -1.50$ and under the box constraint.

CNOP method, which uses the gradient information provided by the conventional adjoint method, is often unable to work out the accurate lower bound of maximum predictable time. On the contrary, thanks to the proper configuration of genetic operators and the effective handling of constraint conditions, the GA-CNOP method reveals its superiority in solving the problem of the lower bound of maximum predictable time. This method is capable of providing the accurate solution to the problem whether the effect of on–off switches is present or absent. The GA-CNOP method, compared to the other two methods, is more effective and feasible in its capacity to solve the problem of the lower bound of maximum predictable time; it is worth disseminating among our fellow researchers and applying to predictability problems that involve on–off switch processes. However, this method still requires further testing in terms of whether it can solve the problem of the lower bound of maximum predictable time effectively in a real high-dimensional atmospheric model in spite of the fact that in the modified Lorenz equation, the GA-CNOP method can solve the problem. In addition, the computational time consumed by the GA-CNOP is much greater than that of the ADJ-CNOP; a GA starting from a set of candidate solutions must integrate the prediction model many more times compared with the conventional adjoint method, thus consuming much more time when using single central processing unit (CPU). Nevertheless, this deficiency can be overcome by executing the parallel computation for GAs, because many genetic operations of different individuals in one generation are independent. This can be done among different CPUs, taking full advantage of fast computational parallel technology. Because the dimension of the Lorenz equation used as the prediction model is just three, the parallel com-

Table 4. The lower bounder of maximal prediction time yielded by the filtering approach under the box constraint.

ε	$\delta = 0.005$	$\delta = 0.01$	$\delta = 0.02$	$\delta = 0.04$	$\delta = 0.08$	$\delta = 0.16$
0.6	198	197	40	40	40	40
1.0	209	209	40	40	40	40
1.4	215	214	214	206	204	194
1.8	239	239	220	209	208	201
2.2	239	239	239	212	211	208

Table 5. Same as Table 4 except the lower bounder of maximum predictability time is yielded by the GA-CNOP, in which the numbers in bold italic show distinct lower bounds of maximum predictable time compared with those by the filtering method.

ε	$\delta = 0.005$	$\delta = 0.01$	$\delta = 0.02$	$\delta = 0.04$	$\delta = 0.08$	$\delta = 0.16$
0.6	198	197	40	40	40	40
1.0	209	209	40	40	40	40
1.4	***214***	214	214	206	***203***	194
1.8	239	239	***222***	209	208	201
2.2	239	239	239	212	211	208

Table 6. Same as Table 5 except the lower bounder of maximum predictability time is yielded by the ADJ-CNOP.

ε	$\delta = 0.005$	$\delta = 0.01$	$\delta = 0.02$	$\delta = 0.04$	$\delta = 0.08$	$\delta = 0.16$
0.6	198	197	40	40	40	40
1.0	***211***	209	40	***195***	***41***	40
1.4	***216***	***219***	214	***207***	***207***	***195***
1.8	239	239	***239***	***215***	***211***	***207***
2.2	239	239	239	***214***	***213***	***212***

putation in the GA-CNOP was not performed in this study. Therefore, better GAs and the parallel of GAs in solving the problems of this kind need to be further explored.

Acknowledgements. This work was supported by the National Natural Science Foundation of China (Grant Nos. 40975063 and 40830955).

REFERENCES

Barth, N. H., 1992: Oceanographic experiment design II: Genetic algorithms. *J. Atmos Oceanic Technol.*, **9**, 434–443.

Birgin, E. G., J. M. Martınez, and M. Raydan, 2000: Nonmonotone spectral projected gradient methods on convex sets. *SIAM Journal on Optimization*, **10**, 1196–1211.

Deb, K., 2000: An efficient constraint handing method for genetic algorithms. *Computer Methods in Applied Mechanics and Engineering*, **186**, 311–338.

Duan, W. S., and H. Y. Luo, 2010: A new strategy for solving a class of constrained nonlinear optimization problems related to weather and climate predictability. *Adv. Atmos. Sci.*, **27**(4), 741–749, doi: 10.1007/s00376-009-9141-0.

Eshelman, L. J., and J. D. Schaffer, 1992: Real-coded genetic algorithms and interval schemata. *FOGA-2 Foundations of Genetic Algorithms-2*, Morgan Kaufmann Publishers, San Mateo, 187–202.

Fang, C. L., and Q. Zheng, 2009: The effectiveness of a genetic algorithm in capturing conditional nonlinear perturbation with parameterization "on-off" switches included by a model, *J. Trop. Meteor*, **13**, 13–19.

Herrera, F., M. Lozano, and J. L.Verdegay, 1998: Tackling real-coded genetic algorithms: operators and tools for the behaviour analysis. *Artificial Intelligence Review*, **12**, 265–319.

Kalnay, E., 2003: *Atmospheric Modeling, Data Assimilation and Predictability.* 1st ed., Cambridge University Press, England, 364pp.

Lorenz, E. N., 1975: Climate predictability. *The Physical Basis of Climate Modeling*, WMO GARP Publ. Ser. No. 16, WMO, 132–136,

Mu, M., and Q. Zheng, 2005: Zigzag oscillations in variational data assimilation with physical "on–off" processes. *Mon. Wea. Rev.*, **133**, 2711–2720.

Mu, M., and Z. Y. Zhang, 2006: Conditional nonlinear optimal perturbation of a two-dimensional quasigeostrophic model. *J. Atmos. Sci.*, **63**, 1587–1604.

Mu, M., W. S. Duan and J. C. Wang, 2002: The predictability problems in numerical weather and climate prediction. *Adv. Atmos. Sci.*, **19**, 191–204.

Mu, M., W. S. Duan and B. Wang, 2003: Conditional nonlinear optimal perturbation and its applications. *Nonlinear Processes in Geophysics*, **10**, 493–501

Mu, M., W. S. Duan, Q. Wang, and R. Zhang, 2010: An extension of conditional nonlinear optimal perturbation approach and its applications. *Nonlinear Processes in Geophysics*, **17**, 211–220.

Powell, M. J. D., 1982: VMCWD: A FORTRAN subroutine for constrained optimization. DAMTP Report 1982/NA4, University of Cambridge, England.

Srinivas, M., and L. M. Patnaik, 1994: Adaptive probabilities of crossover and mutation in genetic algorithms. *IEEE Transaction on Systems, Man and Cybernetics*, **24**, 656–667.

Xu, Q., 1996: Generalized adjoint for physical process with parameterized discontinuities, Part I: Basic issues and heuristic examples. *J. Atmos. Sci.*, **53**, 1123–1142.

Xu, Q., and J. D. Gao, 1999: Generalized adjoint for physical processes with parameterized discontinuities. Part VI: Minimization problems in multidimensional space. *J. Atmos. Sci.*, **56**, 994–1002.

Zheng, Q., and M. Mu, 2006: The effects of the model errors generated by discretization of "on-off" processes on VDA. *Nonlinear Processes in Geophysics*, **13**, 309–320.

Zheng, Q., and Y. Dai, 2009: On the feasibility of nonlinear perturbation equation method in variational data assimilation with "on-off" oscillations. *Chinese Progress in Natural Science*, **19**, 81–88.

Zheng, Q., J. X. Sha, and C. L. Fang, 2011: An effective genetic algorithm to VDA with discontinuous "on-off" switches. *Science in China* (D), doi: 10.1007/s11430-011-4300-4.