**Annales
Geophysicae**

Parameter estimation using the genetic algorithm and its impact on quantitative precipitation forecast

Y. H. Lee[1]**, S. K. Park**[2]**, and D.-E. Chang**[1]

[1]Meteorological Research Institute, Korea Meteorological Administration, Seoul, Republic of Korea
[2]Department of Environmental Science and Engineering, Ewha Womans University, Seoul, Republic of Korea

Abstract. In this study, optimal parameter estimations are performed for both physical and computational parameters in a mesoscale meteorological model, and their impacts on the quantitative precipitation forecasting (QPF) are assessed for a heavy rainfall case occurred at the Korean Peninsula in June 2005. Experiments are carried out using the PSU/NCAR MM5 model and the genetic algorithm (GA) for two parameters: the reduction rate of the convective available potential energy in the Kain-Fritsch (KF) scheme for cumulus parameterization, and the Asselin filter parameter for numerical stability. The fitness function is defined based on a QPF skill score. It turns out that each optimized parameter significantly improves the QPF skill. Such improvement is maximized when the two optimized parameters are used simultaneously. Our results indicate that optimizations of computational parameters as well as physical parameters and their adequate applications are essential in improving model performance.

1 Introduction

Numerical weather/climate prediction models contain numerous parameterizations for physical processes and numerical stability. Parameterizations are based on physical laws but typically contain parameters whose values are not known precisely. The values of the parameters directly or indirectly affect the performance of model, and thus uncertainties in parameter values may lead to sensitive results, especially with high resolution and sophisticated microphysics (e.g., Park and Droegemeier, 1999). Accordingly, optimal estimation of parameters is one of the essential factors in improving the accuracy of numerical forecasts.

Correspondence to: S. K. Park
(spark@ewha.ac.kr)

Recently, efforts have been made to obtain better estimation of parameters for numerical forecast models using various methods such as the variational technique using a full-physics adjoint model (Zhu and Navon, 1999), the Bayesian stochastic inversion (Jackson et al., 2004), the downhill simplex method (Severijns and Hazeleger, 2005), and the ensemble Kalman filter (EnKF) (Aksoy et al., 2006).

The genetic algorithm (GA) has also been applied to some parameter estimation problems. Compared to traditional optimization methods based on the gradient of a function, the GA is more appropriate when the function includes some complexities and/or discontinuities (Barth, 1992). Major advantages of the GA include that: 1) derivatives of a fit function with respect to model parameters are not required; and 2) nonlinearity between the model and its parameters can be handled (Holland, 1975; Goldberg, 1989; Charbonneau, 2002).

The parameter estimation problems have been explored to a wide scope including the land surface parameters (Jackson et al., 2004), the radiation and cloud parameters (Severijns and Hazeleger, 2005), the vertical eddy mixing coefficient (Aksoy et al., 2006), and even for the purpose of experiment design (Barth, 1992). However, application has seldom been made on the quantitative precipitation forecasting (QPF).

This study focuses on optimal parameter estimation to improve the QPF skill in a mesoscale meteorological model using the GA. Section 2 describes the model and experiments, and Sect. 3 explains the details of GA for parameter estimation. Results are discussed in Sect. 4, and conclusions are provided in Sect. 5.

2 Case, model and experiments

A heavy rainfall case in Korea is selected for experiments. It occurred in the west-central part of the Korean peninsula, associated with a summer monsoon front, with a local

maximum 6-h accumulated rainfall of 100 mm in Seoul from 12:00 UTC to 18:00 UTC 26 June 2005.

In this study, the 5th-generation PSU/NCAR Mesoscale Model (MM5) version 3.6.3 (Grell et al., 1994) is employed. The computational domain consists of 218×181 grids in the horizontal, with a resolution $\Delta x = \Delta y = 18$ km, and 35 layers in the vertical. The MM5 is integrated up to 12 h starting from 06:00 UTC 26 June 2005, with a timestep $\Delta t = 45$ s. Schemes for physical processes include: the MRF PBL, the Kain-Fritsch (KF) cumulus parameterization, the Dudhia radiation and RRTM, the Schultz microphysics, and the five-layer soil scheme (see Grell et al., 1994).

For the parameter estimation study, we focus on the closure assumption of the KF parameterization and the Asselin filter coefficient in MM5. The "closure" in the KF parameterization relates the intensity of convective activity to the resolved-scale properties in a model, and assumes that convection consumes at least 90% of the environmental convective available potential energy (CAPE) over an advective time period (Kain, 2003). However, Saito et al. (2006) found that this setting tended to overstabilize the model atmosphere, making rainfall decrease with time. In this study, an experiment will be carried out to obtain the optimal value of the reduction rate of CAPE (ε) in the KF scheme.

The temporal differencing in MM5 consists of leapfrog steps with an Asselin filter (Asselin, 1972). Splitting of the solution associated with the leapfrog scheme can be avoided by using this filter. It is applied to all variables α as

$$\hat{\alpha}^t = (1 - 2\nu)\alpha^t + \nu \left(\alpha^{t+1} + \hat{\alpha}^{t-1} \right) , \qquad (1)$$

where $\hat{\alpha}$ is the filtered variables, and $\nu \in [0, 1]$ is the Asselin filter parameter. The value of ν is set to 0.1 in MM5 for all variables (Grell et al., 1994). However, Bryan and Fritsch (2000) found that the Asselin filter parameter used in MM5 is a source of the unphysical thermodynamic structures. Another experiment in this study will focus on the optimal estimation of the Asselin coefficient.

3   Methodology of parameter estimation

This study aims at performing optimal estimations of two parameters in MM5 using the GA. The GA is a global optimization approach based on the Darwinian principles of natural selection. Developed from the concept of Holland (1975), it seeks the extrema of complex function efficiently – see Goldberg (1989) for a detailed description. Deb (2000) discussed an efficient constraint handling method for the GA.

A key concept in the GA is the chromosome. A chromosome contains a group of numbers that completely specifies a candidate during the optimization process. Typically, the GA uses crossover, mutation, and reproduction to provide structure to a random search. The GA also uses randomization heavily in choosing a chromosome that will propagate to future generations. In general, the average fitness of individuals increases with each generation, through the process of natural selection. In each successive generation, individuals with just good genes propagate their genetic code. The genetic code that determines the fitness of an individual is termed, logically enough, the chromosome of that individual. Given a chromosome, the GA should be able to ascertain its fitness.

For the parameter estimation experiments in this study, a GA package called the PIKAIA (Charbonneau, 2002) is employed. Each generation has 20 chromosomes. The crossover probability is set to 0.85, implying that 85% of the chromosomes in a generation are allowed to crossover in an average sense. The maximum and minimum mutation probability is set to 0.05 and 0.005, respectively.

Internally, the PIKAIA seeks to maximize a function $f(X)$ in a bounded $n-$dimensional space,

$$X \equiv (x_1, x_2, \ldots, x_n), \quad x_k \in [0.0, 1.0] \, \forall \, k . \qquad (2)$$

In our problem, there exist two adjustable parameters, i.e., $n=2$. Then we may associate the reduction assumption of the KF scheme ε with x_1 and the Asselin filter parameter ν with x_2. The ranges of parameters are $0 \leq \varepsilon \leq 0.95$ and $0.01 \leq \nu \leq 0.3$.

The function to be optimized (i.e., Fitness) is defined by using a QPF skill score, the equitable treat score (ETS) (Schaefer, 1990),

$$\text{Fitness} = \sum_i \text{ETS}_i, \quad i = 1, 2, \ldots, 100 , \qquad (3)$$

where i is the precipitation threshold in mm. Here, the ETS is defined as:

$$\text{ETS} = \frac{H - R}{F + O - H - R} , \qquad (4)$$

where H is the number of hits, F and O are the numbers of samples in which the precipitation amounts are greater than the specified threshold in forecast and observation, respectively, and R is the expected number of hits in a random forecast – $R = FO/N$, where N is total number of points being verified.

Each generation includes 20 individual MM5 runs as a function of ε and ν. Every individual run with the two parameters is encoded by chromosomes and returns the accumulated rainfall to determine the fitness; thus the fitness function is dynamically coupled to the MM5 model. In each successive generation, the two parameters make independent search for the optimal solution concurrently; hence there exists no feedback between the two parameters.

4   Results

In this study, two parameters in MM5 are optimized (i.e., ε and ν) to improve the QPF skill. The GA converged to solutions $\varepsilon=0.0111781 \approx 0.01$ and $\nu=0.2498580 \approx 0.25$, through global optimization in the fitness function space which has

multiple minima (not shown). The performance of average chromosomes improved exponentially, up to the second generation (i.e., 60 runs of MM5), as the GA discovers and populates the best regions in the search space. This implies that evolution for only a few generations is sufficient to obtain optimal estimations of parameters.

Figure 1 compares the ETSs computed for the 6-h accumulated rainfall at forecast period of 6–12 h from the following five experiments using: 1) the default parameter (CNTL; ε=0.9, ν=0.1); 2) no convective parameterization scheme (NC; ε=none, ν=0.1); 3) the revised KF parameter (KF; ε=0.01, ν=0.1); 4) the revised Asselin filter parameter (AF; ε=0.9, ν=0.25); and 5) the revised parameters for both the KF scheme and the Asselin filter (KF-AF; ε=0.01, ν=0.25). The ETSs of the default run dropped rapidly with increasing threshold values reaching lower than 0.1 at thresholds larger than 30 mm. In general, it is noticed that the GA-estimated parameters give positive effect on increasing the QPF skill, either independently or together.

In the original KF scheme, ε is set to 0.9; that is, the convection consumes the pre-existing CAPE by 90%. However, the GA-estimated value (i.e., ε=0.01) is quite different from the original. This implies that the convective rainfall in the selected case requires almost no consumption of the pre-existing CAPE. It is notified that, compared with convective systems in the North America, those in the East Asia include air columns that are thermodynamically more neutral and nearly saturated up to the mid-troposphere; thus resulting in a smaller amount of CAPE, especially prior to and during heavy rainfall (see Lee et al., 1998; Hong, 2004). Therefore, in applying the KF scheme to convective rainfalls in the East Asia, it might be essential to assume slow or almost no consumption of the pre-existing CAPE (e.g., Saito et al., 2006); however, it does not necessarily mean that the KF scheme is not applicable to the QPF study in this region.

Compared with the no-convective parameterization experiment (i.e., NC), the KF scheme revised with the GA-estimation (i.e., KF) shows much higher ETSs at thresholds larger than 40 mm (see Fig. 1). It suggests that the KF scheme is still useful but with an optimized value of ε in accordance with the environment that consumes the CAPE slowly for a heavy rainfall event in the East Asia.

The revised Asselin filter (i.e., AF; ν=0.25) also brings about improvement in the ETSs for thresholds of 15–50 mm. Generally, the Asselin filter with ν=0.25 removes $2\Delta t$ waves and reduces the amplitude of $4\Delta t$ waves by half, but with little effect on longer-period waves; that is, it acts as a low-pass filter in time. In contrast, the Asselin filter with default value (ν=0.1) serves as a high-pass filter so that some short-period waves, including gravity waves, are not filtered out. Although the Asselin filter is used for the purpose of numerical stability, the result indicates that its impacts on the QPF are considerable; thus it should be treated with care.

The experiment using both parameters estimated through the GA (i.e., KF-AF) produced the highest ETSs for almost
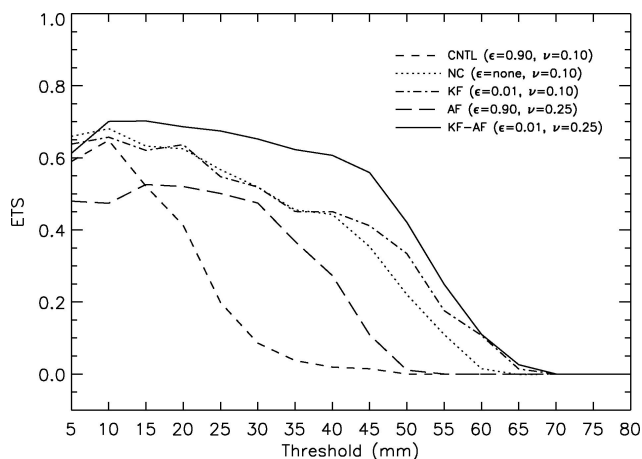

Fig. 1. The ETSs of precipitation forecasts in terms of various thresholds (in mm). Curves denote scores for the 6–12 h accumulated rainfall forecasts for experiments CNTL (control run with default values), NC (no convective parameterization), KF (using optimized parameter in the KF scheme), AF (using optimized parameter in the Asselin filter), and KF-AF (using optimized parameters from both the KF scheme and the Asselin filter).

all thresholds, exceeding 0.6 at thresholds lower than 45 mm. It is notable that the QPF skill increases prominently when the two revised parameters are used together in the model. This suggests that simultaneous use of all optimized parameters, both physical and computational, are essential in improving model performances.

Figure 2 represents a 6-h accumulated rainfall for the forecast time from 6 to 12 h for two experiments: 1) with the default values (i.e., ε=0.9, ν=0.1) and 2) with the GA-estimated values (i.e., ε=0.01, ν=0.25). During the 6-h period between 12:00 UTC and 18:00 UTC 26 June 2006, a heavy rainfall occurred in the west-central part of the Korean peninsula with a local maximum of 100 mm in Seoul (Fig. 2a). The default experiment failed to simulate the amount of rainfall – only 25 mm at the region where more than 90 mm is observed (Fig. 2b). Meanwhile, the experiment with the GA-estimated parameters simulated the localized heavy rainfall quite well with 70 mm peak rainfall (Fig. 2c).

The uniqueness problem in parameter estimation is ultimately related to the issue of parameter identifiability (Navon, 1997). Since the GA is basically a random search algorithm, the parameter indentifiability can be assessed by repeating the GA run, each composed of 200 MM5 runs (i.e., 10 generations × 20 chromosomes), until the best chromosomes start to repeat with some regularity.
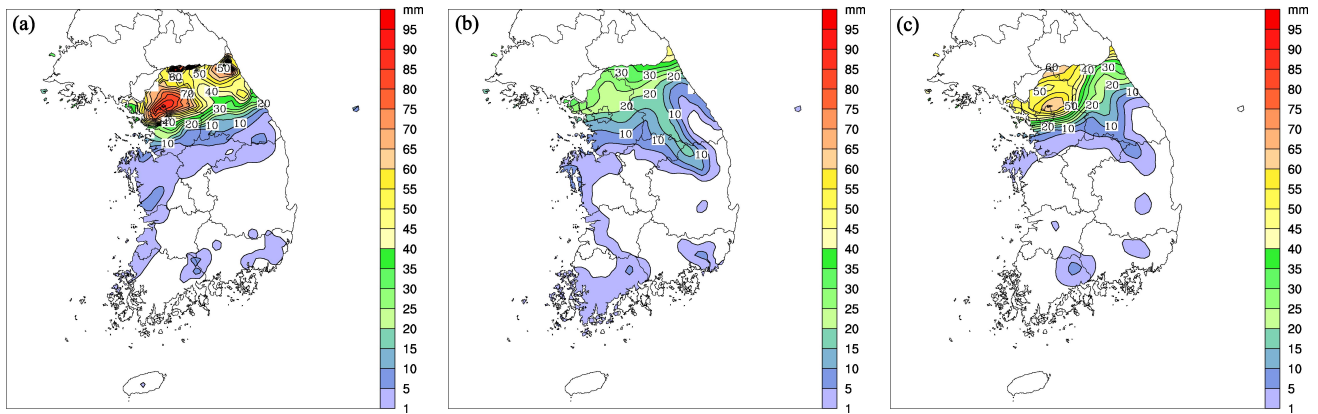
Fig. 2. A 6-h accumulated precipitation (in mm) of **(a)** observation from 592 rain gauge data, and experiments with **(b)** the default parameters and **(c)** the GA-estimated parameters, ending at 18:00 UTC 26 June 2005. The model output is interpolated to the rain gauge locations for verification.

5 Conclusions

In this study, optimal estimation of parameters in a mesoscale meteorological model (MM5) is performed in the purpose of improving the quantitative precipitation forecast (QPF) skills for a heavy rainfall case in the Korean peninsula, employing a global optimization technique called the genetic algorithm (GA). The GA is applied to find out optimal parameters directly using the QPF skill score as a fitness (cost) function.

The GA is robust to complexity and nonlinearity in the model and thus provides more flexible and direct way of solving in parameter estimation. Therefore, nonlinear relations between the fitness function and the model parameters are well treated in the GA. However, evolutions in the GA must accommodate physical constraints associated with development and growth so that all possible paths would not be searched in the genetic parameter space (see Deb, 2000).

The model parameters selected for optimal estimation are the reduction rate of the convective available potential energy (CAPE) in the Kain-Fritsch (KF) scheme, ε, for convective parameterization (i.e., physical parameter) and the Asselin filter coefficient, ν, for numerical stability (i.e., computational parameter). The optimized solutions are $(\varepsilon, \nu) = (0.01, 0.25)$. The GA discovered and populated the best regions in the search space only in a few generations.

Each optimized parameter exerted a favorable influence on the heavy rainfall forecast by improving the QPF skill. Further significant improvement was achieved when two optimized parameters were used simultaneously in the model. This implies that an interaction between optimized physical and computational parameters works favorably to bring about potentially best performance of a numerical model. Therefore, optimizations of computational parameters as well as physical parameters and adequate use of optimized parameters are essential in improving model performance.

References

Aksoy, A., Zhang, F., and Nielsen-Gammon, J. W.: Ensemble-based simulations state and parameter estimation with MM5, Geophys. Res. Lett., 33, L12801, doi:10.1029/2006GL026186, 2006.

Asselin, R.: Frequency filter for time integrations, Mon. Weather. Rev., 100, 487–490, 1972.

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

Bryan, G. H. and Fritsch, J. M.: Unphysical thermodynamic structures in explicitly simulated thunderstorms, presented at: 10th PSU/NCAR Mesoscale Model User's Workshop, NCAR, Boulder, CO, available from http://www.mmm.ucar.edu/mm5, 2000.

Charbonneau, P.: An introduction to genetic algorithms for numerical optimization, NCAR Tech. Note TN-450+IA, 2002.

Deb, K.: An efficient constraint handling method for genetic algorithms, Comput. Methods Appl. Mech. Eng., 186, 311–338, 2000.

Goldberg, D. E.: Genetic algorithms in search, optimization and machine learning, Addison-Wesley, MA, 1989.

Grell, G. A., Dudhia, J., and Stauffer, D. R.: A description of the fifth-generation Penn State/NCAR Mesoscale Model (MM5), NCAR Tech. Note TN-398+STR, 1994.

Jackson, C., Sen, M. K., and Stoffa, P. L.: An efficient stochastic Bayesian approach to optimal parameter and uncertainty estimation for climate model predictions, J. Climate, 17, 2828–2841, 2004.

Holland, J.: Adaptation in natural and artificial systems, University of Michigan Press, Ann Arbor, 1975.

Hong, S.-Y.: Comparison of heavy rainfall mechanisms in Korea and the Central US, J. Meteor. Soc. Japan, 5, 1469–1479, 2004.

Kain, J.: The Kain-Fritsch convective parameterization: An update, J. Appl. Meteoro., 43, 170–181, 2003.

Lee, D.-K., Kim, H.-R., and Hong, S.-Y.: Heavy rainfall over Korea during 1980–1990, Korean J. Atmos. Sci., 1, 32–50, 1998.

Navon, I. M.: Practical and theoretical aspects of adjoint parameter estimation and indetifiability in meteorology and oceanography, Dyn. Atmos. Ocean., 27, 55–79, 1997.

Park, S. K. and Droegemeier, K. K.: Sensitivity analysis of a moist 1D Eulerian cloud model using automatic differentiation, Mon. Weather Rev., 127, 2180–2196, 1999.

Saito, K., Fujita, T., Yamada, U., Ishida, J.-I., Kumagai, Y., Aranami, K., Ohmori, S., Nagasawa, R., and Kumaga, S.: The operational JMA nonhydrostatic mesoscale model, Mon. Weather Rev., 134, 1266–1298, 2006.

Schaefer, J. T.: The critical success index as indicator of warning skill, Weather Forecasting, 5, 570–575, 1990.

Severijns, C. A. and Hazeleger, W.: Optimizing parameters in an atmospheric general circulation model, J. Climate, 18, 3527–3535, 2005.

Zhu, Y. and Navon, I. M.: Impact of parameter estimation on the performance of the FSU global spectral model using its full-physics adjoint, Mon. Weather Rev., 127, 1497–1517, 1999.