Proceedings of ERAD (2002): 416–421 © Copernicus GmbH 2002

ERAD 2002

Using genetic algorithm for distributed parameters optimisation of conceptual rainfall-runoff model with radar data as a high-resolution input

J. Szturc, A. Jurczyk, and K. Osrodka

Institute of Meteorology and Water Management, Branch of Katowice, PL 40-045 Katowice, ul. Bratków 10, Poland

Abstract. The paper is a continuation of the earlier works concerning an introduction of radar data as a high-resolution input into hydrological models. The application of full radar rainfall information provokes to distribute all the rainfall-runoff model parameters onto individual areas corresponding to the radar pixels. If there are thousands of parameters to optimise, more efficient methods involving probabilistic rules of choice are required. The genetic algorithm (GA), used in this paper, may be one of possible optimisation procedures. Investigations of the properties of the genetic algorithm have been performed. The quality of the model using GA has not been better than using integrated parameters and the spatial distribution of parameters optimised in this way has not corresponded to any physical or geographical features of the catchment. For this reason further investigations have been directed towards a limitation of the number of parameters to optimise by either decreasing of both data and parameters spatial resolution or assuming a relationship between the model parameters and the distance to the outlet.

1 Introduction

Introduction of complete information about spatial distribution of precipitation field into a hydrological model should result in an improvement of forecasts quality. However in consequence all model parameters have to be also spatially distributed. Such an approach leads to thousands of parameters to optimise. That complicates procedures of parameters optimisation, therefore more efficient methods, for example the concept of artificial intelligence, may be required.

An example of such a model is presented in this paper. The model has been tested for the catchment of the Vistula river up to Skoczów level gauge (Szturc et al., 2002). It is rural mountainous catchment in the Beskid Ślaski Mountains covering an area of 296.7 km^2 with height varying from 300 to 1220 m above sea level. It is a source region of the Vistula River, the biggest one in Poland.

Rainfall data has been provided by C-band weather radar in Katowice with spatial resolution 0.5 km at range of 100 km. Hourly rainfall accumulations have been calculated on the basis of SRI method (Surface Rainfall Intensity), i.e. rainfall measured 0.5 km above the ground, with time step 15 minutes.

2 Conceptual rainfall-runoff model

Rainfall-runoff conceptual models are the most frequently based on a parametrical equation expressing a relationship between the outflow from a river basin, so called the direct runoff $Q(t)$, and the excess rainfall $R_e(t)$. The appropriate transfer function, so called the response function $h(t)$, is used to transfer excess rainfall into outflow:

$$Q(t) = \int_0^t h(t-\tau) R_e(\tau) d\tau. \tag{1}$$

The response function $h(t)$ used in this paper is well known in hydrology as the Nash model (Nash, 1958):

$$h(t) = \frac{1}{k\Gamma(n)} \left(\frac{t}{k}\right)^{n-1} e^{-\frac{t}{k}}, \tag{2}$$

where $\Gamma(n)$ is the gamma function, and n, k are the model parameters.

In practice, a continuous record of rainfall is not available. For this reason the transformation formula (Eq. 1) is limited to the following discrete form:

$$Q(t) = \sum_{\tau=0}^{t} h(t-\tau) R_e(\tau) \Delta t. \tag{3}$$

The time step of calculations Δt depends on the catchment size and in our case it is 1 hour.

Every catchment pixel is considered as independent of the others (see Fig. 1). Then the total runoff from the basin is

Correspondence to: J. Szturc (Jan.Szturc@imgw.pl)

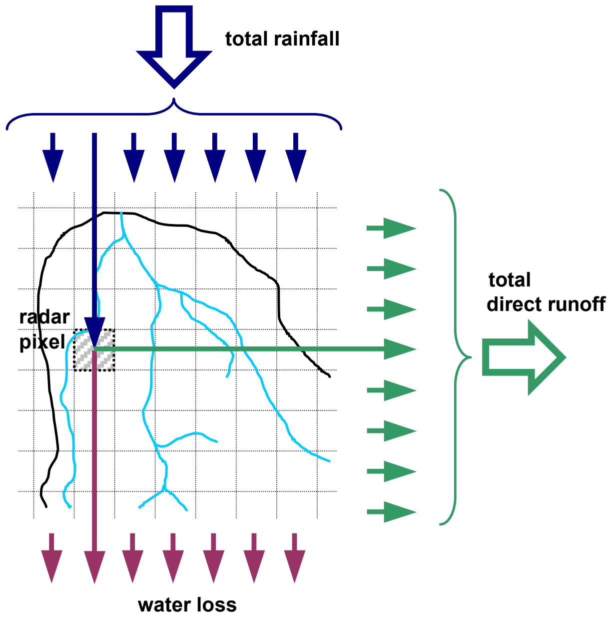

Fig. 1. Scheme of the spatially distributed model according to radar resolution.

calculated as the sum (all over the catchment area) of runoffs from individual pixels; the runoffs being in turn sums (all over the rainfall time) of runoffs at consecutive time steps:

$$Q(t) = \sum_{g=1}^{N_g} \sum_{\tau=0}^{t} h(g,\, t-\tau) R_e(g,\, \tau) \Delta t \qquad (4)$$

where g is a number of a pixel (1... N_g). Each parameter of the model constitutes an array of N_g elements, what makes the optimisation difficult.

The excess rainfall has been estimated basing on SCS formula (SCS, 1986) with parameters calculated from total direct runoff.

Quality of the model has been estimated by integrated square error ISE (Weglarczyk, 1998):

$$ISE = \frac{RMSE}{Q^m_{obs}} \cdot 100\% \qquad (5)$$

where $RMSE$ is the root mean square error, and Q^m_{obs} is the mean observed discharge.

The conceptual rainfall-runoff models require optimisation of their parameters. A frequent difficulty is a lack of possibility of physical interpretation of the parameters, what would help in the estimation of their values.

The application of the simple form of rainfall-runoff model to the following analysis results from the aim of the job, which was not an examination of hydrological features of the catchment, but investigation of possibilities of the model parameters optimisation.

3 Problem of parameters optimisation

The response function in any conceptual model has a number of parameters, for instance two parameters in Nash model: n and k, which do not have any physical meaning. Then a mathematical optimisation basing on historical data is necessary to estimate them.

When the number of parameters is small enough – up to three according to Liong et al. (2001) – they can be optimised using trial-and-error method. But the bigger the number of parameters is, the more efficient optimisation algorithms should be applied. The simplest method to optimise the big number of parameters is so called Uniform Random Sampling (URS). In this method the great number of parameters samples is randomly generated and the best one is taken as the optimum.

The rainfall-runoff model quality may be estimated by an objective function, for instance expressed by Eq. (5). This function has usually more than one minimum so there is more than one solution of the optimisation problem. The procedures that stop after having found a local minimum are not appropriate for application in this case. Duan et al. (1992) and Gan and Biftu (1996) described and verified other methods applying empirical material, e.g. exhaustive gridding (EG), shuffle complex evolution (SCE-UA), multiple start Simplex (MSX) and local Simplex methods. At the outside 15 parameters were optimised in these papers. Efficiency of the methods results from an introduction of randomness for the search of the global minimum.

The genetic algorithm (GA), used in this paper, may be one of more advanced optimisation procedures. Only few papers have considered hydrological applications of GA. To our knowledge the first attempt was undertaken by Wang (1991) who applied Xinanjiang model with 7 parameters. Next GAs were used in this field by Franchini (1996) and Franchini and Galeati (1997). Liong et al. (2001) applied the method of genetic programming, which is an extension of GA. The trials of using GA undertaken by authors of this paper are reported in (Jurczyk et al., 2001).

Usually a compromising approach, where the number of parameters to optimise has been limited, is chosen. Although a direct interpretation of the parameters does not exist, an empirical dependence of the parameters on the following features may be introduced: the catchment concentration time (Chander and Fattorelli, 1991), travel path to the outlet, which consists of hill slope and stream fractions (Garrote and Bras, 1995), or the distance to the closest watercourse (Corral et al., 2000; 2001). Therefore only a few or several parameters remain for optimisation.

4 Description of the genetic algorithm (GA)

Genetic algorithms (GAs) are the algorithms of optimisation and search basing on analogy to the mechanisms of natural selection and heredity. They differ from the traditional methods in the following features: they do not directly treat

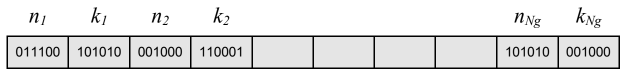

Fig. 2. Coding of a chromosome (total number of pixels Ng = 1,251).

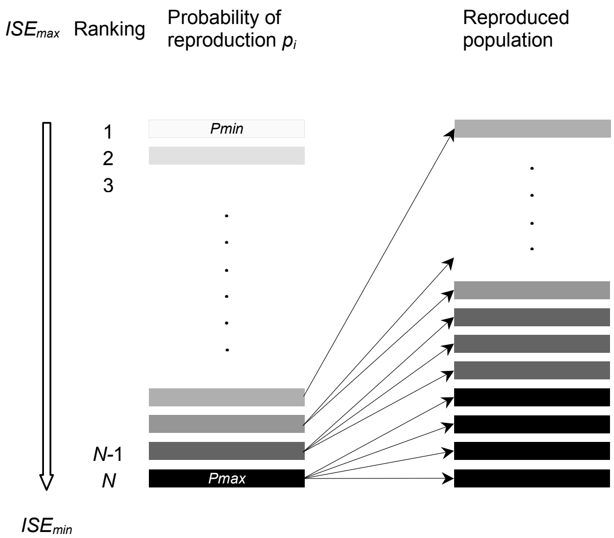

Fig. 3. Scheme of selection and reproduction.

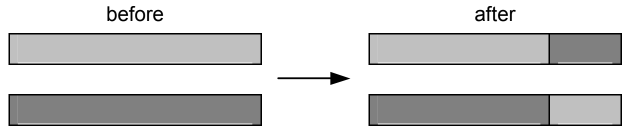

Fig. 4. Scheme of crossover.

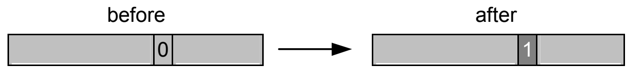

Fig. 5. Scheme of mutation.

the task parameters but their coded form, conduct the search starting not from a single point but from certain population of points, and do not use deterministic but probabilistic rules of choice. GAs are utilised in many problems containing the search, optimisation and teaching (Goldberg, 1989). Elementary GA is composed of four operations: (a) selection, (b) reproduction, (c) crossover, and (d) mutation.

The total number of parameters to be optimised is equal to the doubled number of pixels at the catchment area, as Nash model adopted here introduces two parameters. For our river basin there are 1,251 pixels × 2 parameters = 2,502 parameters. All those mutually independent parameters have been coded as one binary number in such a way that they have been written one after another, pixel after pixel, every parameter covering a predefined number of bits. If the number of bits is 8, the parameters get values from the interval 0.1 ÷ 25.6 with the step 0.1. In this way a chromosome of the length of 20,016 genes – or binary digits – has been defined (Fig. 2).

All chromosomes form a population, which size has been optimised (from 100 to 500). The initial population has been created by a random generation of the appropriate number of chromosomes.

1. Selection. The first step consists in a selection of chromosomes, performed randomly, but favouring better chromosomes, i.e. having higher values of the fitness function (Fig. 3).

   For the purpose of this paper the fitness function F has been defined as:

   $$F = 1 - \frac{ISE}{100\%} \tag{6}$$

   where ISE is the error of the model taking the parameters included in a given chromosome.

   There are many methods of population selection, e.g. roulette, tournament and ranking. The last method has seemed to be the most efficient in this case. The ranking method consists in ordering of all chromosomes according to their fitness function values: from the worst to the best one. Then a probability of duplication p_i is attributed to each chromosome according to the formula:

   $$p_i = \frac{i}{\sum\limits_{i=1}^{N} i} \tag{7}$$

   where i is number of the chromosome in the series, N is the number of chromosomes in the population.

2. Reproduction (Fig. 3). Thus, the better a chromosome is, the higher is its probability to be selected, so it is more frequently duplicated. In this way the population is reproduced.

3. Crossover (Fig. 4). The next step is the crossover, performed also in random way. The crossover is applied to randomly matched couples of chromosomes with the determined probability, with one or more also randomly selected crossover point.

4. Mutation (Fig. 5). The mutation – as well as the crossover – has to prevent the stopping of the optimisation procedure after attaining a local maximum of the fitness function. Both the mutation index and the number of mutations have to be optimised. If the mutation index is assumed to be e.g. 0.1, and the number of mutations 50, it means that fifty digits are randomly altered into opposite in one chromosome among 10.

In presented paper up to 100 reproductive cycles have been performed. Described below investigations of the genetic algorithm properties have been performed for a single freshet of 6–7 June 2000. It has been caused by intensive storms connected with cold front spreading from the west. The precipitation (30 mm on average over the whole catchment) on 6 June resulted in an increase of water level up to the range of high level.

5 Optimisation of fully distributed parameters

First the model errors have been calculated for integrated parameters, i.e. when they have had the same optimised values for all over the catchment. The best ISE has amounted to 5.50% with integrated parameters n = 1.0 and k = 17.1.

Before trying to use more sophisticated methods to optimise fully distributed parameters, the simplest optimising method URS has been used. Thus 500 samples of parameters have been randomly generated. The lowest ISE value using the best parameters has equalled about 6.6% (Jurczyk et al., 2001).

The next step was using GA method. Applying of this method requires estimation of relevant GA parameters. Therefore the following properties of GA having influence on the algorithm efficiency have been investigated.

1. The number of bits recording the parameters n and k (admissible parameter interval). The optimum values are 5 bits for parameter n and 7 bits for parameter k.

2. The initial population size. The probability of generation of chromosomes having high values of the fitness function increases in larger population. On the other hand the time of optimisation procedure gets longer because a model has to be started for each chromosome. A compromising value of 500 chromosomes has been assumed.

3. The initial population type. The initial population has been generated in different ways: either fully randomly or partially imposing extreme values. The introduction of chromosomes with partially imposed values into the initial population is aimed to obtain better differentiation of the population. The extreme chromosome values are to give to GA a stimulus for further search. The GA efficiency is significantly better when the initial population is created using the second method, however only in some tens of steps.

4. The method of reproduction. The ranking method has been used. When the population size is big (here N = 500 chromosomes), the probability of duplication is very low for all chromosomes. It leads to excessive averaging of population. In order to avoid it, the probability of duplication for better chromosomes has been amplified multiplying their probabilities of duplication by constant factor or squaring these probabilities. The amplification generally improves the GA efficiency at the several first steps of its running. Later on this improvement is impeded because values of the fitness functions in the population equalise.

5. The number of crossover points in one chromosome and crossover coefficient. The optimum values of the both parameters have been set to 1. The crossover coefficient value equal 1 means that each chromosome is crossovered.

6. The number of mutations and mutation coefficient. The optimum values of the number of mutation and mutation coefficient have been set to 500 and 0.1 respectively. Increase in number of mutations has not improved optimisation quality. It may be caused by the fact that too many mutations apart from knocking out population from averaged values additionally as a side effect lead to a breakage of some well-adjusted chromosomes.

The Nash model quality with fully distributed parameters optimised with GA (ISE = 5.49%) has not turned out to be better than with integrated parameters (ISE = 5.50%).

In spite of mathematically correct fitness, the spatial distribution of parameters optimised using GA seems to be completely chaotic what excludes any trials to connect their values with physical and geographical features of the catchment. Many different combinations of n and k values can give similarly good results of the model.

There is a variety of techniques which may be applied to smoothing of data set, for instance various forms of kriging. Results described for instance by Whelan et al. (2001) show that the form of spatial estimation chosen for mapping has a significant influence on the final estimated image. Kriging with a local semivariogram performs the best in the case of large data sets, and using block kriging gives a well smoothed map. The VESPER software (Minasny et al., 2002) has been used for our analysis. Smoothing of parameters values (optimised with GA) using kriging and block kriging have had no effect on improving their spatial distribution.

Therefore next works have been directed towards a limitation of the number of parameters to optimise. Two compromising approaches have been proposed: (1) using lower spatial resolution, and (2) introducing *a priori* a relationship between model parameters and some physical or geographical features of the catchment.

6 Decrease in number of parameters by decrease in spatial resolution

GA is a very efficient method however it does not cope with hundreds and more parameters to optimise. Thus the next undertaken step has been to test its efficiency with lower number of parameters. The decrease in number of rainfall-runoff model parameters has been achieved by reduction in their spatial resolution. The resolution of radar input data has been always the same as the resolution of model parameters.

The spatial resolution has been changed gradually from 0.5 to 8 km, so the number of parameters has varied from 2502 down to 10. The best result has been obtained with 1 km resolution (ISE = 4.61%), whereas the worst with 4 km resolution (ISE = 7.86%). It is difficult to interpret these results. It might be effect of specific freshet which the algorithm has been tested on.

7 Decrease in number of parameters by partially imposing of parameters values

Next approach has consisted in introducing a relationship connecting model parameters to some physical or geographical feature of catchment. A strict relationship between the values of the parameters n and k and the distance l from the pixel to the catchment outlet has been assumed as follows:

$$n_g = a_1 l_g + a_2 \tag{8}$$

$$k_g = a_3 l_g + a_4 \tag{9}$$

where g is a number of a pixel. The values of l_g have been normalised to unity, i.e. $l_g \in$ (0, 1).

It is a simplification with respect to the assumption accepted by Corral et al. (2000; 2001). In such a way the number of parameters has been limited to only a few. After applying assumption described by Eqs. (8) and (9) only four parameters a_1, a_2, a_3 and a_4 have remained to optimise. Then the size of the chromosome drastically has decreased to 28 bits with 7-bits record for each parameter (from a_1 to a_4). The other GA parameters have had also to be changed. Optimum GA parameters have been set to: the population size 250, number of crossover points in chromosome 1, crossover coefficient 0.25, number of mutations 8, and mutation coefficient 0.1.

At once the result has appeared to be better. The model error ISE has been 5.07%.

8 Summary

The application of full radar rainfall information provokes to distribute all the rainfall-runoff model parameters onto individual areas corresponding to the radar pixels. However such an approach leads to thousands of parameters to optimise. Using fully distributed parameters optimised with GA has resulted in no improvement of the Nash model quality (ISE = 5.49%, whereas with integrated parameters ISE = 5.50%). The spatial distribution of parameters optimised using GA seems to be completely chaotic what excludes any trials to connect their values with physical and geographical features of the catchment. It indicates that GA does not cope with the optimisation problem although fitness is mathematically correct. Many different combinations of n and k values can give similarly good results of the model. Smoothing of optimised parameters values using kriging and block kriging has had no effect on improving their spatial distribution.

Thus next two approaches consisting in radical diminishing of the number of parameters to optimise have been proposed: (1) using lower spatial resolution, and (2) introducing a relationship between model parameters and some physical or geographical features of the catchment. Decrease in number of parameters by decrease in spatial resolution has affected improving GA efficiency but only slightly. The best result has been obtained with 1 km resolution when ISE = 4.61%. Introducing relationship connecting model parameters with the distance to the catchment outlet has resulted in lower improvement of GA efficiency (the model error ISE = 5.07%).

Table 1. The results of optimisation with different ways of parameters distributing

Method of distribution of model parameters	Optimisation method	ISE [%]
Integrated parameters	Trial-and-error	5.50
Fully distributed parameters	URS	about 6.6$*$
Fully distributed parameters, highest resolution 0.5 km	GA	5.49
Fully distributed parameters, the resolution from 0.5 to 8 km	GA	4.61$**$
Distributed, partially imposed parameters	GA	5.07

$*$ depends on specific randomly generated samples.
$**$ with 1-km resolution.

References

Chander, S. and Fattorelli, S.: Adaptive grid-square-based geometrically distributed flood-forecasting model. In Hydrological applications of weather radar, eds. I.D. Cluckie and C.G. Collier, Ellis Horwood, Chichester, 424-439, 1991.

Corral, C., Sempere-Torres, D., Revilla, M. and Berenguer, M.: A semi-distributed hydrological model using rainfall estimates by radar. Application to Mediterranean Basins. Phys. Chem. Earth (B), 25, 1133-1136, 2000.

Corral, C., Sempere-Torres, D. and Berenguer, M.: A distributed rainfall runoff model to use in Mediterranean Basins with radar rainfall estimates. 30th International Conference on Radar Meteorology, 19-24 July 2001, Munich, Germany, 6-8, 2001.

Duan, Q., Sorooshian, S. and Gupta, V.: Effective and efficient global optimization for conceptual rainfall-runoff models. Water Resour. Res., 28, 1015-1031, 1992.

Franchini, M.: Use of a genetic algorithm combined with a local search method for the automatic calibration of conceptual rainfall-runoff models. Hydrol. Sci. J., 41, 21-39, 1996.

Franchini, M. and Galeati, G.: Comparing several genetic algorithm schemes for the calibration of conceptual rainfall-runoff models. Hydrol. Sci. J., 42, 357-380, 1997.

Gan, T.Y. and Biftu, G.F.: Automatic calibration of conceptual rainfall-runoff models: optimization algorithms, catchment conditions, and model structure. Water Resour. Res., 32, 3513-3524, 1996.

Garrote, L. and Bras, R.L.: A distributed model for real-time flood forecasting using digital elevation models. J. Hydrol., 167, 279-306, 1995.

Goldberg, D.E.: Genetic algorithm in search, optimization, and machine learning. Addison-Wesley, Reading, Mass., 1989.

Jurczyk, A., Osrodka, K. and Szturc, J.: Distributed conceptual rainfall-runoff model with radar data as a high resolution input – using genetic algorithm for parameters optimization, Fifth International Symposium on Hydrological Applications of Weather Radar "Radar Hydrology", Proceedings, November 19-22, 2001, Heian-Kaikan, Kyoto, Japan, 395-400, 2001.

Liong, S.Y., Van Nguyen, V.T., Gautam, T.R. and Wee, L.: Alternative well calibrated rainfall-runoff model: genetic programming scheme. Paper presented at World Water and Environmental Resources Congress 2001 (papers on CD-ROM), 20-24 May 2001, Orlando, Florida, USA, 2001.

Minasny, B., McBratney, A.B. and Whelan, B.M.: VESPER version 1.5. Australian Center for Precision Agriculture, The University of Sydney (http://www.usyd.edu.au/su/agric/acpa), 2002.

Nash, J.E.: The form of the instantaneous unit hydrograph. IASH AISH Publ., 42, 114-118, 1958.

Soil Conservation Service: Urban Hydrology for Small Watersheds. U.S. Dept. of Agriculture, Soil Conservation Service, Engineering Division, Technical Release 55, June, 1986.

Szturc, J., Barczyk, A., Osrodka, K., Jurczyk, A. and Moszkowicz, S.: Selected case studies for the Vistula River catchment in the frame of the MUSIC Project, Wiadomosci IMGW (in press), 2002.

Wang, Q.J.: The genetic algorithm and its application to calibrating conceptual rainfall-runoff models. Water Resour. Res., 27, 2467-2471, 1991.

Weglarczyk, S.: The independence and applicability of some statistical quality measures for hydrological models. J. Hydrol., 206, 98-103, 1998.

Whelan, B.M., McBratney, A.B. and Minasny, B.: VESPER spatial prediction software for Precision Agriculture. Proc. of the 3rd European Conference on Precision Agriculture, Montpellier, France, 139-144, 2001.