STATISTICAL APPROACH OF THE SCALE EFFECT IN HYDROLOGICAL MODELING

Prof. Radu DROBOT
Technical University of Civil Engineering Bucharest

Abstract.

The calibration, respectively the validation of the hydrological or hydrogeological parameters of the mathematical models is a difficult task, having not till now an adequate response. The parameters used in physically based mathematical models have a physical meaning and are measured in-situ. Still, using these measured parameters does not lead automatically to a correct response of the modeled process, a time-consuming process of parameters calibration and validation being necessary. Although there are different computer codes based on automatic search, the specialists usually prefer a "trial-and-error" procedure, because this way they can control better the optimization process, matching the calculated values with their own perception about the plausible parameters. In the paper, several considerations on the statistical characteristics of the parameters, function of the grid size, are presented. Because the error function has usually local optima, our purpose is to identify a family of near-optimal solutions, for which the errors obtained during the mathematical modeling process are rather close, placed in a relatively narrow interval. During the simulations, as a consequence, the results, either state values or output values, are obtained as an interval. If an unique value is needed, it can be achieved based on a linear combination of the partial results provided by each set of parameters from the quasi-optimal vectors family.

Key words: parameters, statistical distribution, calibration, sub-optimal solutions, linear convex combination.

1. Physical considerations

In the case of physically based hydrological models, the parameters have a physical meaning and can thus be measured in the field. However, a contradiction between the strictly local character of the parameters thus obtained and the mean value, corresponding to the representative elementary surface used in hydrological modeling, respectively the representative elementary volume in hydrogeological modeling appears.

That is why the discretization of the equations describing the physical process leads to the necessity of parameters calibration, even if a great number of measurements *in situ* are available; as a result, their measured values only have an indicative role, referring to the order of magnitude or to the range in which the effective parameters used in modeling can vary. The role of the measured parameters diminishes as the discretization of the hydrological or hydrogeological basin becomes rougher ,as in the models with quasi-distributed or global parameters (fig. 1).

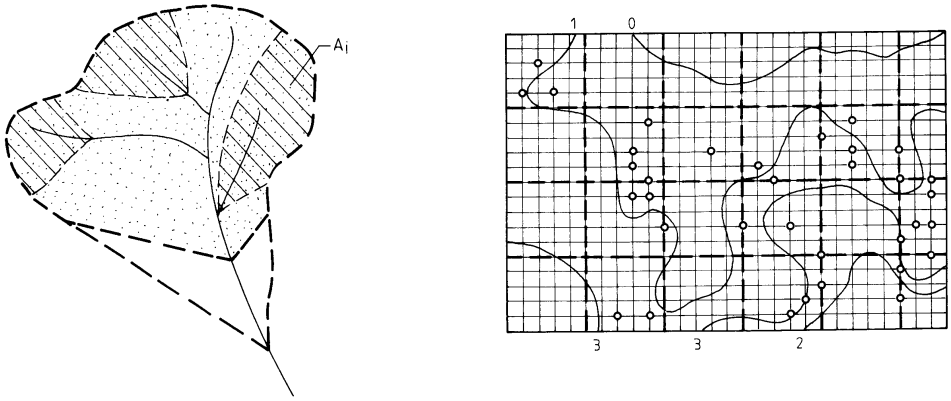

Fig. 1 – Discretization of the study domain
a) – river basin; b) hydrogeological basin

At the limit, in the case of hydrological models with global parameters, respectively hydrogeological models for homogenous aquifers, the values determined *in situ* for some parameters are of no practical use. Besides, in these cases, even the model's type is modified, getting a conceptual character, to the detriment of the physical meaning of the processes (P. Serban, 1996).

2. Statistical relationship: *discretization mesh size – parameters value*

The obvious need of using field data must be connected to the necessity of a quick calibration of the parameters. On the other hand, it is proved that hydrological or hydrogeological parameters have a statistical distribution. One of the first models taking this into account is the Stanford model (Serban and others, 1989), in which the repartition function of infiltration capacity linearly varies between a null value, for impervious areas, and a maximum value, for the most permeable areas; thus, in this model, the repartition density of local infiltration on the river basin surface is supposed to be uniform.

In most cases however, the wide range of genetic factors, as well as their additional or multiplying effect over the parameters' value usually lead to normal or lognormal distribution functions. We can mention here the lognormal distribution of hydraulic conductivities, noticed while processing data from measurements of some regional aquifers (Fl. Zamfirescu, 1996).

The most important element that influences the values of the calibration parameters within distributed models is the size of the discretization element, A_i, which may be thus considered as explanatory variable (fig. 2)

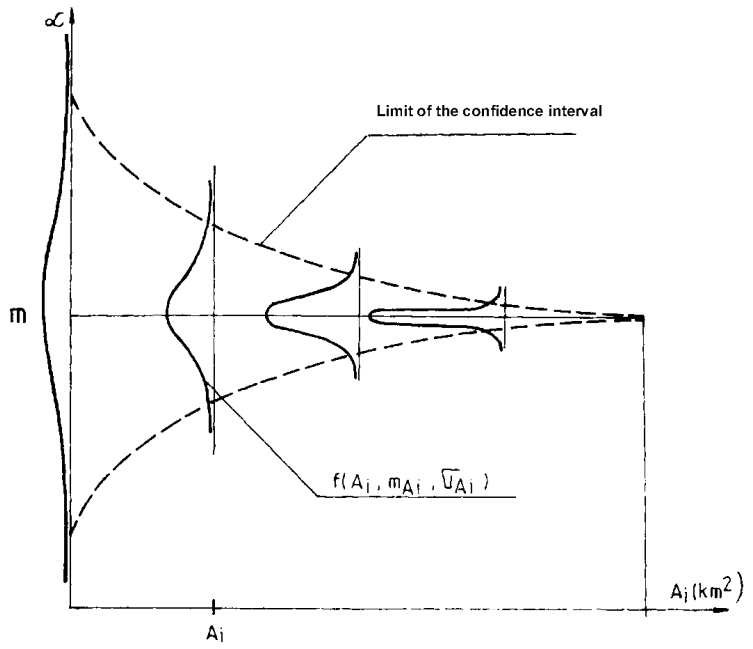

Fig. 2 One-dimensional distribution functions
for the discretization element's different values.

The parameters range reaches its greatest value in the case of some experimental determinations, which have a strictly local character (e.g. the infiltration rate into the soil) or quasi-local respectively (like the hydraulic conductivities or transmissivities for regional aquifers, obtained by pumping tests, lasting a limited period of time and which partially mobilize the aquifer's capacity of water transfer or water storage– V. Pietraru, 1977).

However, the range of parameters diminishes with the increase of surfaces or volumes of the discretization elements. Denoting by α any parameter of the models, its repartition density for different values of the discretization element's surface may be plotted as in figure 2.

For every discretization element we introduce by modeling an average value of local state variables, as well as for the parameters resulted from the calibration procedure. During the calibration process, the parameters values will be selected not only by purely mathematical reasons but also depending on other available information, referring to physiographical, geological or hydrogeological characteristics. The quasi-local parameters, attached to the discretization elements, will have a smaller range, corresponding to this averaging process. However, the statistical distribution type is supposed to remain the same, obviously with other statistical parameters.

The greater is the size of discretization elements, the more reduced is the variation of parameters. At the limit, when the discretization element is the whole modeled basin, the value of the searched parameter is unique. As a result of the process implied by a rougher discretization, the limit value of the analyzed parameter may be supposed to be the statistical average of punctual values obtained by measurements.

The previously presented facts lead to the idea of using a two-dimensional repartition of every parameter's repartition; the plot in fig. 2 suggests in fact such a distribution, the shown distribution of the densities being obtained by its intersection with various planes A_i = ct.

3. *Discretization - quality of modeling* relationship

There is however a direct relationship between the size of the discretization element and the quality of the simulation, which requires the use of a mesh as smooth as possible. Calculus precision or simulation quality is usually defined by one of the following two indicators (A. Musy, 1998):

- the root mean squared error between the calculated and the measured values , or the standard deviation, given by the relation:

$$\varepsilon = \sqrt{\frac{F}{n}} = \sqrt{\frac{\sum_{i=1}^{n} (y_i^m - y_i^c)^2}{n}}; \quad (1)$$

- the R^2 Nash-Suitcliff efficiency criterion:

$$R^2 = 1 - \frac{F}{F_0} \quad (2),$$

The notations have the following meaning:

y_i^m and y_i^c are the measured, respectively calculated values, of the system's output values (discharges) or of the state variables (e.g. levels of the groundwater);

F - the sum of the squares of the differences between the calculated values and the measured ones;

F_0 - the "initial" deviation of the observed values in relation to their average:

$$F_0 = \sum_{i=1}^{n} (y_i^m - \bar{y})^2 \quad (3)$$

The more refined is the discretization in the case of the models with distributed parameters, the better is the precision, meaning that the error function (1) is more reduced, respectively the R^2 Nash-Suitcliff efficiency criterion is closer to the unity (fig. 3); at the opposite side there are the models with global parameters, for which the average square deviation of the errors ε has a maximum value, respectively R^2 criterion is minimum.

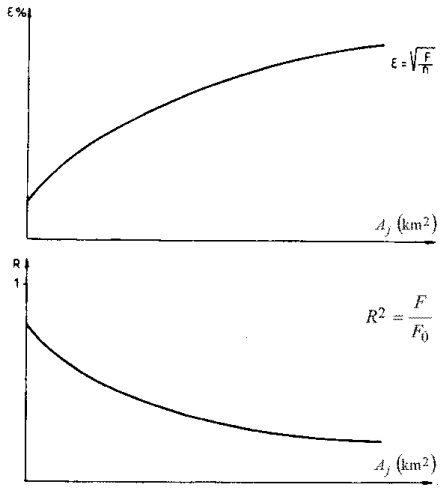

Fig.3. Modeling precision vs. the discretization element's surface

4. Statistical distribution of parameters

Further on, for a simplified approach, we consider that the repartition of punctual values (respectively parameters repartitions obtained after rougher discretizations) have a gaussian behavior with the same mean value of the parameter *m,* which is the value corresponding to the whole studied domain. Depending on the grid size, the limits of the confidence interval which contain (with a given probability) the calibration parameters values, might be established for the chosen distribution (see fig. 2).

For the probability density function of parameters, normal or lognormal distribution may be chosen:

$$f(\alpha, A;\ \ m\ \ , \sigma_A) = \frac{1}{\sigma_A\ \sqrt{2\pi}}\ \exp\left[-\frac{1}{2}\cdot\frac{\left(\alpha - m\ \right)^2}{\sigma_A^2}\right] \tag{4}$$

respectively

$$f(\alpha, A;\ \mu\ \ , \sigma'_A) = \frac{1}{\sigma'_A\ \ x\sqrt{2\pi}}\ \ \exp\left[-\frac{1}{2}\cdot\frac{\left(\ln\alpha - \mu\ \right)^2}{(\sigma'_A)^2}\right] \tag{5}$$

The parameters σ_A and σ'_A depend non-linearly on the surface of discretization elements, and they could be defined by the following expressions:

$$\sigma_A = a_0 + a_1\ A + a_2\ A^2, \tag{6}$$

respectively

$$\sigma'_A = a_0 + a_1\ \ln A + a_2\ \ln A^2 \tag{7}$$

In the relations (5) and (6), as well as in the example shown in figure 2, the parameters m or μ were taken as constant, independently of the grid size. A linear or even non-linear trend line might as well be admitted for the average value, depending on the surface; in this case, in the relations (5) and (6), instead of m or μ, the notations m_A, μ_A respectively will be used, while the definitions are similar to (6).

5. The random character of the system's response

Generating artificial values with a chosen distribution for every parameter having a physical meaning will lead to a random output of the system; we can expect that those parameter values leading to the minimum values of the residuals are also the most credible. Accepting as valid the sets of parameters which lead to relatively close optimal error functions (in a range of differences of 5-10%), a family of quasi-optimal type of solution vectors is obtained, each vector corresponding to a local optimum of the objective function defined by the relations (1) or (2). We notice that in the case of using of function (1) its minimum values are searched, while for the other optimization criterion the goal is to find the parameters configurations that maximize the function (2).

The choice of the quasi-optimal calibration vectors from the sets of parameters kept in the previous phase must be performed taking into account the physical constraints as well, which reflect the nature of the studied system.

Finally, in order to validate the quasi-optimal parameters' family with the sets selected after this last trial, simulations with input data not used in the calibration process will be performed. Corresponding to each set of parameters j ($j = \overline{1,m}$) from the selected family, a weight function reflecting the quality of the obtained results will be calculated (Shamseldin et al, 2000); as weight functions, the following expression can be used:

$$w_j = 1 - \frac{\varepsilon_j}{\sum_{j=1}^{m} \varepsilon_j}, \tag{8}$$

respectively

$$w_j = \frac{R_j^2}{\sum_{j=1}^{m} R_j^2} \tag{9}$$

In the simulation phase of any hydrological or hydrogeological model, the obtained results for each set j of parameters lead to an interval of the system's possible responses; thus, the response value could be statistically characterized (average value, value characterized by a certain exceeding probability, average square deviation, confidence interval with a certain probability etc.). If one wishes to obtain a unique value of the response, the partial results are to be weighted with the w_j

coefficients (the WAM - Weighted Average Method), the final result being a linear convex combination of them.

Surpassing the limits of classical optimization, based on the premise of a global optimum, is thus being tried; the so-called flexible optimization, by means of accepting some sub-optimal solutions, may be an attractive alternative.

Conclusions

The statistical approach for the scale effect means to introduce supplementary parameters for each parameter of the examined system which have a physical meaning: m or μ (may be m_A or μ_A) for the average, respectively a_0, a_1 and a_2 for the average square deviation. Obviously, for more than a few parameters, the calibration process becomes prohibitive from the computation time point of view.

The approach presented here infers artificial generating of the parameters' sets, their trial based on repeated simulations during the calibration and validation process or based on physical reasons, as well as the use of a WAM procedure of weighting the results obtained for the sets of parameters having been retained as quasi-optimal solutions.

Bibliography

A. Musy – Hydrologie Appliquée, 1998 - Editions HGA, Bucarest.

V. Pietraru, 1997 – Seepage calculation (Calculul infiltratiilor). Editura Ceres.Bucuresti

P. Serban, V. Al. Stanescu, P. Roman, 1989 – Dynamic Hydrologyu (Hidrologie dinamica). Editura Tehnica. Bucuresti

P. Serban, 1996 – Deterministic hydrologic models (Modele hidrologice deterministe). Editura Didactica si Pedagogica. Bucuresti

A. Shamseldin, A. Ahmet, K. O'Connor – A Seasonally Varying Weighted Averege Method for the Combination ofRiver Flow Forecasts. XX Conference of the Danubian Countries on Hydrological Forecasting and Hydrological Bases of Water Management. Bratislava, Slovakia, 2000.

Fl. Zamfirescu, 1997 – Basics of the groundwater dynamics (Elemente de baza in dinamica apelor subterane) Editura Didactica si Pedagogica, Bucuresti.