Reducing of model formulation errors as an effective remedy for improving simulation results

A. SPALVINS, J. SLANGENS, R. JANBICKIS & I. LACE
Riga Technical University, Environment Modelling Centre,
1, k.4 Meza Street, Riga, LV-1048, Latvia
e-mail: emc@egle.cs.rtu.lv

Abstract If a hydrogeological model (HM) contains serious errors than no calibration can improve results of such HM. Reliability of HM can be increased if these errors are subdued by a methodology and special software tools aimed at such a purpose. Often data interpolation programmes apply inapt algorithms. The Environment Modelling Centre (EMC) of the Riga Technical University has developed tools to fight most of these faults.

INTRODUCTION

Results of HM may be wrong if it includes serious errors, as after-effects of careless pre-processing. Specialists involved in modelling and software tools unproperly developed or applied cause these errors. To narrow the scope of problems to be discussed, only data pre-processing for semi-3D steady state HM will be reviewed; mistakes of data processing and post-processing will not be considered.

The *xyz*-grid of HM is built of ($h*h*h_z$)-sized blocks (h is block plane size; h_z is a variable block height). They constitute a rectangular s-tiered xy-layer system (s is the number of strata involved). Its four vertical sides compose the shell. The relief and the lower side of the model are its geometrical top and bottom.

In HM, the vector j of the piezometric head is the numerical solution of the corresponding boundary field problem, which is approximated, in nodes of the HM grid, by the following algebraic equation system:

$$A\, j = b,\ A = A_{xy} + A_z - G,\ b = b - G\, y,\ b = b_{in} + b_{bot} + b_{sh} + b_w,\ b_y = G\,(j - y) \quad (1)$$

where the matrices A_{xy}, A_z, G, represent, correspondingly, horizontal links (transmissivities) of aquifers, vertical connections (leakages) of aquitards, elements connecting nodes of the grid with the piezometric boundary conditions y; the vector b accounts for boundary flows; b_w is the flow passing through the elements of G; b_w is the water production rate in wells; b_{in}, b_{bot} and b_{sh} are the boundary flows specified on top, bottom and shell surfaces, accordingly. If the y-condition is fixed on some boundary areas (y-surfaces) then, on the remaining part (b-surface), the flow b should be specified.

The elements a_{xy}, a_z of the matrices A_{xy}, A_z, which represent the links in the grid, are computed by using digital maps of thicknesses m, m_0 and permeabilities k, k_0 of aquifers and aquitards, accordingly. The following formulae are applied:

$$a_{xy} = k\, m,\ \ a_z = h^2 k_0 / m_0,\ \ k, k_0\ ³\ 0,\ \ m, m_0 = H_{i-1} - H_i\ ³\ 0,\ \ i = 1,2,\ldots,s \quad (2)$$

where H_{i-1} and H_i are the elevation distributions of neighbouring bottom surfaces of the geological strata; H_0 represents the landscape relief with the hydrographical set included. The set of the H_i maps describes the full geometry of HM. It is built in the consecutive way: $H_0 \rightarrow H_1 \rightarrow \ldots \rightarrow H_s$, by keeping m, m_0 ³ 0. To simplify our discussion, it is conditionally assumed that the thicknesses m, m_0 are independent from the solution j.

To subdue errors of pre-processing, the following six principles have been formulated and applied by EMC.

1. It is easier to avoid possible mistakes than to fight and discover them in the running model where their after-effects are intermixed.

2. To pare down casual human errors, a modeller should not be needlessly involved in routine operations performed rightly by a properly designed software.

3. To cope with difficult problems to be settled, a complex model should be prepared by a team of experts in overlapping fields (initial data treatment and interpolation, hydrogeology, computer software, etc.).

4. To ensure automatic regeneration of interpolation results, all input maps should be fully computer created.

5. The choice of an interpolation method to be used depends on the type of the input map to be prepared. A wrongly applied method may cause considerable errors.

6. The unreliable boundary flows b_{in} , b_{bot} and b_{sh} should be substituted for the more veritable flow b_y.

We are going to discuss problems of preparing H_i, (k, k_0), y, and b maps, as inputs for building of the system (1). Pre-processing of data, in some modelling programmes, is estimated.

TREATMENT OF INITIAL DATA

The common mistakes associated with initial data are such, as:

- direct errors of data caused by technical faults of measurements or by a careless attitude of personnel;
- data from different distant sources may be wrongly referenced in space if their co-ordinates are not transformed to the topographical base applied in HM;
- mismatched physical dimensions of data are incautiously used.

No method can detect automatically all erroneous initial data. To recognise them, the EMC team uses the following indications:

- incorrect data stand out, as a visible discordance, against the nearest true ones;
- a wrong location of data with respect to known landmarks (roads, buildings, rivers, etc.) that are present on the base map of the HM area;
- in the vertical direction, data are incorrectly attached to the geological strata included in the model;
- if values of a seasonally dependent head are not observed in wells simultaneously then such data are worthless for interpolation purposes;
- crude errors of boundary conditions can be detected (Atruskievics *et al*., 1994) by a visual inspection of j-portraits related to separate objects (rivers, lakes, the flow b_w; etc.); this approach is based on the following superposition:

$$j = \sum_1^n j_i , \qquad A\, j_i = b_i , \qquad b = \sum_1^n b_i \qquad (3)$$

where j_i, b_i are the j-portrait and the boundary condition of the i-th object, correspondingly; for example, the depression cone is the j-portrait of the flow b_w.

These symptoms are outcomes of tests on data compatibility, within rather simple systems, where the true information dominates over the false one, and a modeller stands a good chance of detecting incorrect data.

The safety and amount of initial information increase considerably if data lines are applied. For example, the border of a geological layer is the isoline of its zero thickness: m, $m_0 = 0$. A line can also carry any variable function. For instant, the long profile of a river is obtained by interpolating along the river between the points of observed water levels. Evidently, this profile is much more informative then the initial point data not incorporated in the line.

EMC uses data lines mainly and has developed software (Atruskievics *et al*., 1995) for creating them in the most general grid-independent form. It allows to compute the interpolated function in any random point of the line, particularly, in the crosspoints where the line intersects rows and columns of the grid applied for the system (1). These values adequately account for the influence of the line and are used as the input of the Geological Data Interpolation (GDI) programme (Spalvins & Slangens, 1994). In GDI, lines can also be used as impervious borders, which dissect, if necessary, the HM grid area (Spalvins *et al*., 1995).

The figure of a data line is scanned from a map. Sometimes, shapes of arbitrary lines are chosen in the course of interpolating. Man-made maps of the y-distributions cannot be used as sources of isolines to be scanned. The y-conditions should be obtained as solutions of systems like (1) where the observed head values are applied only as calibration targets.

Tested and corrected initial information (points, lines) should be applied as the computer driven database.

INTERPOLATION OF DATA

Even under an idealistic assumption that the initial data contain no errors, the result of digital mapmaking may be wrong due to inapt interpolation methods used. This problem is only scantly mentioned in manuals on interpolation.

For H_i, (k, k_0), and b maps, interpolation surfaces represent geometric images. Their quality describes the residue of values (difference between the initial data and interpolated values).

It is often overlooked that the formal mass balance of an interpolation surface is always disturbed at locations where initial data are applied as the input. This feature does no harm for H_i, (k, k_0), and b maps, but it is a disaster for the y-maps, because their virtue depends on two residues (value and mass balance), and y-maps cannot be created by applying methods, which are good for other input maps.

In Table 1, distinctive interpolation features of input maps are summarised.

Table 1 Features of interpolated HM input maps.

Type of input map	Initial data as input	Mass balance required	Other conditions
elevations H_i	yes	no	$m, m_0 = H_{i-1}\text{-}H_i{}^{3}\ 0$
permeability (k, k_0)	yes	no	$k, k_0{}^{3}\ 0$
y-distribution	no	yes	limited b_y
b-distribution	yes	no	interpolated b_w

The features, regarding the role of data input and a mass balance, have been considered above. Other conditions, included in Table 1, require some explanations.

The condition m, m_0 ³ 0 is a serious burden, especially, when: surfaces H_i, H_{i-1} are uneven; within some subareas, the thicknesses m or m_0 are zero. These difficult tasks have been solved by the GDI programme. If m, m_0-maps are created directly (not involving H_i, H_{i-1}) then considerable errors of the HM vertical size emerge. For this reason, EMC never uses such a method when HM is applied for driving contaminant transport models, which require high accuracy of the HM geometry.

Permeabilities k, k_0 are always nonnegative. These maps are the main subjects for calibration procedures, because the HM geometry is usually kept unchanged. Double density grids (plane block size equals $0.5\ h$) should be used for interpolating calibrated k-maps, to ensure a controllable solitary value for each element a_{xy} of (2).

To prevent unrealistic immense flows b_y caused by a locally contradictory y-distribution, one should be careful in fixing piezometric boundary conditions at the sites where the formula (2) gives large values of a_z or a_{xy} (as elements of G). A typical case is a hydrogeological window ($m_0=0$). Often excessive flows arise on the HM shell. The shell is used by EMC, as an interpolator for producing a conformable y-distribution here.

Water production rates b_w should be properly interpolated to nodes of the grid (especially, for regional HM where $h=0.5\text{-}4.0$ km), but not roughly moved into the nearest node. EMC has developed the programme for this purpose (Lace *et. al*., 1998).

It follows from the above material that a lot of special software tools has been developed by EMC for data interpolation purposes. They are included in the modelling programme REMO (Spalvins *et al*., 1995).

None of available interpolation programmes are perfectly fit for making complex HM. It is shown in (Spalvins *et al*., 1998), why such wide-known interpolation methods as: Criging, Minimal Curvature and Inverse Distance sometimes provide wrong results. There are two basic reasons:

- interpolated values do not hold the maximum/minimum principle towards initial data applied;
- the interpolation surface does not carry a minimal amount of energy.

Because these methods cannot provide strictly controllable results, the GDI programme has been was developed. It creates an interpolation, as the numerical solution of an arbitrary boundary field problem, which is formulated on the HM grid for the parameter, to be interpolated (H_i, (k, k_0), b). The initial data of this parameter are applied as piezometric boundary conditions. The GDI code possesses the following advantages (Spalvins & Slangens, 1995):

- the shape of any interpolation surface can be exactly ruled, because it automatically satisfies the principles of maximum/minimum and minimal energy;
- processes in REMO and GDI are in a perfect synchronism, because system solvers and storage arrangements are identical here;
- not only pointwise initial data, but also all kinds of data lines can serve as the input of the GDI code;
- the GDI code can be applied in the sequential mode - during the current interpolation step, new information is applied, but the results of the previous step serve as the base; the number of steps are not limited.

The sequential GDI mode is very effective when complex surfaces should be created, or the severe condition: m, $m_0=0$ secured. This mode usually follows some natural geological and/or industrial scenario, which gradually alters the simple shape of a surface into the complex one. In this case, the set of initial data is much simpler than if one tries to obtain the result at once.

SUBSTITUTION OF BOUNDARY CONDITIONS FOR FLOWS

The boundary flows b_{in}, b_{bot}, and b_{sh} for the corresponding b-surfaces are the most unreliable parts of the system (1). They cannot be well defined, by reason of missing credible experimental data about their spatial distribution. Moreover, any b-surface distorts shapes of depression cones created by the flow b_w. Their areas of influence and values of the drawdown are always considerably overestimated under the inducement of the b-surfaces.

Fortunately, there exists no unavoidable obligation to simulate straightforwardly these uncertain boundary conditions. They can be substituted by far more accurate flows b_y, as follows:

- the infiltration b_{in} is simulated, as the leakage $b_{in} = G_{z0}(j_a\text{-}H_0)$ through the unsaturated aeration zone; the diagonal matrice G_{z0} contains vertical links between the relief H_0-map and the j_a-distribution, which exists just under the aeration zone; G_{z0} should be calibrated to obtain j_a as a target;

- the flow b_{bot} is given, as the leakage $b_{bot} = G_{zs}(j_{s-1}\text{-}y_s)$ through the s-th aquitard; the diagonal matrice G_{zs} contains vertical links between the j_{s-1} and y_s-distributions;

- to substitute the flow b_{sh}, the initial HM, in the xy-plane, must be encircled by the buffer zone; on the shell of in this way enlarged HM, the appropriate y-distribution should be applied; the zone must be wide enough, so that the depression cone is insignificantly affected by the introduced y-shell.

EMC has successfully used the above substitutions, especially, the first one (Spalvins *et al.*, 1995), where the H_0 map controls the infiltration flow. In all three above cases, the model itself automatically computes the unknown b_y-distributions. The gain in accuracy is due to much higher reliability of data, which determine these distributions. The flow b_y is self-regulative, because it depends also on components of j.

ESTIMATION OF PRE-PROCESSING IN SOME MODELLING SYSTEMS

In view on pre-processing, EMC has compared the modelling systems: REMO, Groundwater Vistas (GV), and Visual MODFLOW (VM) by building an intricate model described in (Riga Technical University, 1998). All input maps were prepared on REMO and imported into GV and VM. The summary of some estimates is presented by Table 2.

Table 2 The summary of data pre-processing estimates.

System code	Map import	Map regeneration	Zone principle	Approximation scheme	Interpolated b_w	Human involvement
REMO	yes	yes	no	semi-3D	yes	slight
GV[1)]	yes	no	yes	semi-3D	no	medium
VM[2)]	limited	no	yes	3D	no	strong

[1)]Environment Simulation, 1997; [2)] Waterloo Hydrogeologic, 1998.

Table 2 shows that GV and VM programmes are under notable influence of principles, which have been applied in the era of slow mainframe computers (zone principle used to save memory, evident expectations about simpleness of a model, needlessly strong human involvement). The harmful impact of the zone principle is obvious if the computer generated (k, k_0)-maps are considered. It is especially true for the VM code where the k and k_0-maps are not separated (3D scheme used). The typical ratio $k_{max}/k_{min} \cong 10^6$ is large. Such a wide sector of permeability values cannot be covered satisfactorily with only 5000 zones available in GV. For this reason, the REMO properly created (k, k_0)-maps were seriously distorted if imported into the VM programme. A limited ability to import beforehand prepared input maps (only MODFLOW codes accepted) is the other serious drawback of the GV system. There exists a lot of other minor critical remarks about data pre-processing in the GV and VM systems. However, such an analysis is a theme for another discussion.

CONCLUSIONS

Model formulation errors can be subdued if data pre-processing is focused on this purpose. The following main measures are of importance:

- testing and correcting of initial point and line data bases;
- using computer based interpolation methods for creating maps;
- avoiding use of uncertain boundary flows;
- excluding a needless involvement of a modeller in routine processes;
- special software tools should be developed for data interpolation purposes.

We hope that results reported in this paper will be of some interest for modellers involved in creating of complex hydrogeological models.

REFERENCES

Atruskievics, J., Janbickis, R., Krutofal, T., Lace, I., Levina, N., Slangens, J., Spalvins, A. & Viksne, Z. (1994) Computer based hydrogeological model "Large Riga". In: *Proceedings of Latvian - Danish Seminar on "Groundwater and Geothermal Energy"*, Riga - Copenhagen, vol. 2, 203 - 224, (*Boundary Field Problems and Computers; 35-th issue*).

Atruskievics, J., Viksne, Z., Spalvins A. & Lace, I. (1995) Lakes and rivers algorithms in the hydrogeological model "Large Riga". In: *Proceedings of International Seminar on "Environment Modelling"*, Riga – Copenhagen, vol.1, 193-200, (*Boundary Field problems and computers; 36-th issue*).

Environmental Simulations, Inc. (1997) Guide to Using Groundwater Vistas.

Lace, I., Spalvins, A. & Slangens, J. (1995) Algorithms of accounting groundwater discharge in the regional hydrogeological model and interpolation of simulation results at observation wells. In: *Proceedings of International Seminar on "Environment Modelling"*, Riga - Copenhagen, vol. 1, 201 - 216, (*Boundary Field Problems and Computers; 36-th issue*).

Riga Technical University (1998) Modelling of groundwater flow dynamics and contamination transport processes at area of Vilnius oil storage. *Contract No 6274/98 Report*, Riga.

Spalvins, A., Janbickis, R., Slangens, J., Gosk, E., Lace, I., Viksne, Z., Atruskievics, J., Levina, N. & Tolstovs, J. (1995) Development of regional hydrogeological model "Large Riga". In: *Proceedings of International Seminar on "Environment Modelling"*, Riga - Copenhagen, vol. 1, 145 - 158, (*Boundary Field Problems and Computers; 36-th issue*).

Spalvins, A. & Slangens, J. (1994) Numerical interpolation of geological environment data. In: *Proceedings of Latvian - Danish Seminar on "Groundwater and Geothermal Energy"*, Riga - Copenhagen, vol. 2, p. 181 - 196, (*Boundary Field Problems and Computers; 35-th issue*).

Spalvins, A. & Slangens, J. (1995) Updating of hydrogeological data interpolation programme. In: *Proceedings of International Seminar on "Environment Modelling"*, Riga - Copenhagen, vol. 1, 201 - 216, (*Boundary Field Problems and Computers; 36-th issue*).

Spalvins, A., Slangens, J. & Janbickis, R. (1998) Preprocessing of initial data for creating hydrogeological models. In: *Proceedings of the 9-th International Symposium on "System - Modelling - Control"*, Zakopane, Poland, 7 pages (on SMC'98 CD-ROM).

Waterloo Hydrogeologic (1998) Visual MODFLOW. User's Manual.