Pergamon

Wat. Sci. Tech. Vol. 36, No. 5, pp. 61–68, 1997.

Printed in Great Britain.
0273–1223/97 $17·00 + 0·00

PII: S0273-1223(97)00462-9

MULTIPLE DATA PARAMETER IDENTIFICATION FOR NONLINEAR CONCEPTUAL MODELS

Hermann Eberl, Amar Khelil and Peter Wilderer

Institute for Water Quality Control and Waste Management, Technical University of Munich, Am Coulombwall, 85747 Garching, Germany

ABSTRACT

A numerical method for the identification of parameters of nonlinear higher order differential equations is presented, which is based on the Levenberg-Marquardt algorithm. The estimation of the parameters can be performed by using several reference data sets simultaneously. This leads to a multicriteria optimization problem, which will be treated by using the Pareto optimality concept. In this paper, the emphasis is put on the presentation of the calibration method. As an example identification of the parameters of a nonlinear hydrological transport model for urban runoff is included, but the method can be applied to other problems as well.

KEYWORDS

Lumped parameter; nonlinear model; parameter identification; urban hydrology.

INTRODUCTION

In urban hydrology, lumped parameter conceptual models are widely used, because they need substantially less computing time than the more detailed hydrodynamic models. This is of major importance for implementing a real time control system. Another purpose is their usage as a supplementary tool in hydrodynamic simulation of simplified descriptions of sewer systems (Eberl *et al*, 1996). The disadvantage of this kind of models, however, is, that they are not capable of describing hydraulic problems in the sewer system like backwater effects or submerging. The basic idea behind them is to consider the whole sewer system as one single reservoir or as a combination of some reservoirs. Besides the equation describing the conservation of mass in the reservoir an additional equation must be established, that describes the volume of water inside the reservoir as a function of its outflow (Becker and Glos, 1969) or *vice versa*. In the past, mainly linear models were used but we should relax this restriction because of the nonlinearity of the flow processes in the sewer system. Since these models have only poor physical meanings – in fact, they can be considered as nothing but a structured black box – model parameters cannot be derived directly from observations but have to be estimated using measurements of the entire process or empirical and statistical relations. For linear models the solution of the governing differential equation could be undertaken in a closed form and parameter identification reduces to classical nonlinear regression. Here, however, we are concerned with nonlinear models and therefore with the problem of parameter identification for ordinary differential equations.

The method discussed in the sequel can be used for the estimation of the parameter vector $\alpha = (\alpha_1, .., \alpha_n)^T$ of any higher order lumped model

$$\frac{d^m y}{dt^m} = \mathcal{F}\left(t, y, \frac{dy}{dt}, .., \frac{d^{m-1}y}{dt^{m-1}}; u, \alpha\right) \tag{1}$$

where t is the independent variable (*time in our case*), $u = u(t)$ the model input and $y = y(t; \alpha)$ the model output, i.e. the outflow of the sewer system in our case. Equation (1) can be transformed into a system of m first order differential equations for the vector valued function $\psi = (\psi_1, .., \psi_m)^T$ with $\psi_m \equiv y$. It follows

$$\frac{d\psi_r}{dt} = f_r(t, \psi; u, \alpha) \quad , \qquad r = 1, .., m \tag{2}$$

For our purposes, equation (2) is the basic model equation. Combination of several reservoirs immediately leads to (2) as the formulation of the total model. Equation (1) will not be needed in this case. More than one model input could be taken into consideration, but surface runoff will be the only one in hydrological conceptual sewer system models.

PARAMETER IDENTIFICATION

Objective function for parameter identification

The identification of the model parameter vector $\alpha = (\alpha_1, .., \alpha_n)^T$ will be executed by mathematical optimization. For this purpose, an objective function must be established. Our goal is calibrating the model by the use of k reference data sets simultaneously. We can not expect, however, to achieve the same parameter for each of these data sets by separately solving the related scalar minimization problems. We, therefore, will formulate and solve a multicriteria optimization problem. As a criterion for the comparison of reference data and computed model output we will use the least square criterion. In runoff hydrology this choice is often rejected with the argument, that even a small time shift in the computed solution can result in a big least square error and hence other criteria are proposed (e.g. Khelil and Semke, 1991). But indeed, this is not a problem of the least square criterion: if a systematical time shift error is observed then the model can be reformulated by adding a time shift parameter $\alpha_{n+1} = \tau$. Especially, if our ultimate goal is to apply the calibrated model for forecasting within real time control, any time shift should be avoided.

The vector optimization problem can be formulated as

$$\min_{\alpha \in \mathcal{A}} \int_0^T \left(\hat{y}_i(t) - y_i(t; \alpha)\right)^2 dt, \qquad i = 1, .., k \tag{3}$$

where $\hat{y}_i(t)$ denotes the measured values for the system outflow belonging to the input data $u_i(t)$ and $y_i(t; \alpha) \equiv \psi_{m,i}(t; \alpha)$ is the model output of equations (1) or (2), respectively, with model parameter α. $\mathcal{A}$ is the set of admissible α. No optimization concept analogous to scalar optimization exists for this type of problem, but there are some other optimality concepts for multicriteria optimization, like Pareto Optimum, Nash Equilibrium or the Min-Max-Solution (Stadler, 1988). Which of them can be successfully employed depends on the nature of the problem. We will use the Pareto concept, because it is suitable for coalition problems, where all the scalar objectives should be minimized simultaneously, whereas the remaining concepts are more suitable for noncoalition problems, where the minimization of one scalar objective can be achieved at the cost of maximization of any of the others. A parameter $\tilde{\alpha}$ is called Pareto minimal if and only if for all $\alpha \in \mathcal{A}$ holds

$$\int_0^T \left(\hat{y}_i(t) - y_i(t;\alpha)\right)^2 dt \leq \int_0^T \left(\hat{y}_i(t) - y_i(t;\tilde{\alpha})\right)^2 dt \quad \forall\, i = 1,..,k \quad \Longrightarrow$$
$$\Longrightarrow \quad \int_0^T \left(\hat{y}_i(t) - y_i(t;\alpha)\right)^2 dt = \int_0^T \left(\hat{y}_i(t) - y_i(t;\tilde{\alpha})\right)^2 dt \quad \forall\, i = 1,..,k \tag{4}$$

It can not be expected to find a unique Pareto minimum. Moreover, an infinite number will exist. We are interested in only one of them, so an additional choice has to be made. This choice will be subjective in any case. It can be shown (e.g. Stadler, 1988) that minimization of any scalar function Z of the scalar partial objectives $Z_i = \int_0^T (\hat{y}(t)_i - y_i(t;\alpha))^2 dt$ results in a Pareto minimum if

$$\frac{\partial Z}{\partial Z_i} > 0, \qquad \forall\, i = 1,..,k \tag{5}$$

The most simple approach is to use a linear weighting scalarization function

$$Z(\alpha) = \sum_{i=1}^{k} w_i Z_i(\alpha) = \sum_{i=1}^{k} w_i \int_0^T \left(\hat{y}_i(t) - y_i(t;\alpha)\right)^2 dt, \qquad w_i > 0 \quad \forall\, i = 1,..,k \tag{6}$$

Compared to more complicated functions, a big advantage of this function Z – at least from a numerical point of view – is that the least square structure of the problem is conserved. One can imagine several choices for w_i, for example:

$$w_i = 1 \qquad \text{or} \qquad w_i \cdot \int_0^T \hat{y}_i dt = \sum_{j=1}^{k} \int_0^T \hat{y}_j dt \qquad \forall\, i = 1,..,k \tag{7.a,b}$$

In the first case, every data set has the same weight, in the second, they are weighted with regard to their total runoff.

Optimization scheme: the Levenberg-Marquardt algorithm

To solve the optimization problem (6) numerical methods are available which explicitly use the least square structure. Since the problem may be ill-posed (Groetsch, 1988) we should concentrate on methods with regularisation properties. A standard technique amongst them is the Levenberg-Marquardt algorithm which is described for example in Marquardt, 1963. It combines the major advantages of the steepest descent and the Gauss-Newton methods. On one hand, it initially does quick steps towards the minimum, on the other hand it provides quadratic convergence in the vicinity of the minimum. The interpolation is automatically done by a factor λ which decreases while converging towards the minimum and should be 0 in a sufficient small neighborhood of the minimum, which means that at the end of the iteration a strict Gauss-Newton method is used. In every Levenberg-Marquardt iteration step the gradient and the Hessian of the objective function must be computed once but in order to choose an optimal λ some more evaluations of the objective function itself may be necessary.

Depending on the construction of the conceptual model the admissible set $\mathcal{A}$ can be bounded by some constraints. So far, only linear constraints were obtained in our studies. For this purpose, the Levenberg-Marquardt algorithm, which is used in the interior of $\mathcal{A}$, is coupled with a method analogous to the projective gradient method, that is described in Ritter and Kredler (1992) in order to deal with these constraints at the boundaries of $\mathcal{A}$.

Computing the gradient and Hessian of the objective function

The determination of the gradient of the objective function Z with respect to the parameter α is of high importance for both, the Levenberg-Marquardt algorithm as well as the projective gradient method, because optimization actually is done by searching for the zeros of this gradient and, therefore, some emphasis should be placed on the accuracy of this computation. The gradient is given by

$$\frac{\partial Z}{\partial \alpha_j} = -2 \sum_{i=1}^{k} w_i \int_0^T (\hat{y}_i - y_i) \frac{\partial y_i}{\partial \alpha_j} dt, \qquad j = 1, .., n \tag{8}$$

and the Hessian, which is needed for the Levenberg-Marquardt algorithm, reads

$$\frac{\partial^2 Z}{\partial \alpha_j \partial \alpha_l} = -2 \sum_{i=1}^{k} w_i \int_0^T \Big((\hat{y}_i - y_i) \frac{\partial^2 y_i}{\partial \alpha_j \partial \alpha_l} - \frac{\partial y_i}{\partial \alpha_j} \frac{\partial y_i}{\partial \alpha_l} \Big) dt, \qquad j, l = 1, .., n \tag{9}$$

Because a closed form of the model output y of (1) is unknown and only a numerical approximation η can be achieved some considerations on the approximation of the $\frac{\partial y}{\partial \alpha_j}$, which appear in the gradient and in the Hessian, should be done.

The most simple approach of course is numerical differentiation of the numerical solution. This approach is well known and often used. Using a one-sided difference operator it is given by

$$\frac{\partial \eta}{\partial \alpha_j}(t, \alpha) \doteq \frac{\eta(t, \alpha + \Delta \alpha_j e_j) - \eta(t, \alpha)}{\Delta \alpha_j} \qquad j = 1, .., n, \tag{10}$$

where e_j is the unit vector in $j-$direction and $\Delta \alpha_j$ is the stepsize. But the problem of choosing the stepsize arises. Since α_j may vary from a small to a large value during the iteration – especially if no good initial value is *a priori* known – choosing $\Delta \alpha_j$ relative to the actual size of α_j (say 2%) may be convenient, but this will fail if α_j itself is approximately 0. Then an *a priori* fixed value $\Delta \alpha_j$ should be used in which case it might happen, that the step size is of the same order or even bigger than the value itself.

Another ansatz can be obtained from differentiating the model equation (2) with respect to α_j. In vector notation it yields

$$\frac{\partial^2 \psi}{\partial t \partial \alpha_j} = \frac{\partial f}{\partial \alpha_j} + \frac{\partial f}{\partial \psi} \frac{\partial \psi}{\partial \alpha_j}, \qquad j = 1, .., n, \tag{11}$$

where $f = (f_1, .., f_m)^T$. After substituting $\varphi_j \equiv \frac{\partial \psi}{\partial \alpha_j}$ the gradient of model output $y \equiv \psi_m$ can be computed by the linear differential equation systems

$$\frac{d \varphi_j}{dt} = \frac{\partial f}{\partial \alpha_j} + \frac{\partial f}{\partial \psi} \varphi_j, \qquad j = 1, ..n. \tag{12}$$

In this equation for every α_j the coefficient functions are approximately known *a priori*, since an approximation η for y can be computed. The disadvantage of the numerical differentiation in the previous approach is avoided here.

A third method for computing the gradient of the objective function is the differentiation of the numerical solver for the differential equation, which we will not discuss further in this paper.

In all three approaches, computing the gradient of Z for any parameter α requires the additional numerical solution of $k \cdot n$ differential equation systems: following the first one (Eq. (10)), the model

(2) must be computed to determine $\eta(t, \alpha + \Delta\alpha_j e_j)$, the second one (Eq. (11)) requires to solve the linear system (12) in order to get the φ_j.

Some remarks on the computing of the Hessian – which is needed in the Levenberg-Marquardt algorithm – should be added. Computing the Hessian plays a minor role, compared to computing the gradient, because using an approximation for the Hessian has no effect on the solution of the optimization problem. There are only effects on the convergence of the method and on the numerical effort. A whole class of methods for nonlinear algebraic equations, the so called quasi Newton methods, is based on this idea. If we take a look at the formula (9) for the Hessian, we realize that the second derivative term appears together with the factor $\hat{y}_i - y_i$. If our model describes the processes well we can assume, that at least in the vicinity of the optimal $\alpha \in \mathcal{A}$ this factor is very small. In addition we can assume, that the sign of this factor is not constant if we are concerned with real world measurements and therefore integrated along the time axis this error will compensate itself to some degree. Furthermore, if we neglect the effect of this error additional regularisation of the data will be achieved. Thus, we decided to neglect the second order derivation term in (9).

Solving the ordinary differential equation and computing of the objective function

The last module in our parameter identification scheme is a solver for ordinary differential equations to evaluate y and $\frac{\partial y}{\partial \alpha}$. Though methods with stepsize control and variable order like the Runge-Kutta-Fehlberg RKF7(8) method (Stoer and Bulirsch, 1990), are state of art, we will use a scheme with fixed step size and we do this for a good reason. Because we are interested in a least square solution and in most cases the reference data are only given by a discrete set, we should compute the solution and the gradient of the model for the same time set where reference data are available. If we are using automatic stepsize control techniques for this computation, interpolation of data will be necessary. This can be avoided by using a fixed time step scheme, for example the $4th$ order Runge-Kutta RK4 method (Stoer and Bulirsch, 1990). The approximation $\eta(t)$ of the solution of (2) obtained by a one-step method like RK4 can be interpreted as a piecewise linear function. In most cases the reference data function $\hat{y}(t)$ is obtained by piecewise *(linear)* interpolation of discrete data. Thus, the integrands of (3) are piecewise polynomials and the objective function can be calculated directly without additional discretization errors. The same holds for the computing of the gradient and the Hessian.

NONLINEAR CONCEPTUAL MODELS FOR SEWER SYSTEMS

We are concerned with nonlinear higher order conceptual models for sewer systems, which are constructed by combining m nonlinear single reservoir models. They are represented by

$$\frac{d\psi_i}{dt} = h_i(\psi_i; \hat{\alpha}_i) \left(\gamma_i u + \sum_{j=1}^{m-1} \beta_{ij} \psi_j - \psi_i \right), \quad i = 1, .., m \tag{13.a}$$

$$\sum_{i=1}^{m} \gamma_i = 1, \quad \sum_{i=1}^{m} \beta_{ij} = 1, \quad \gamma_i \geq 0, \quad \beta_{ij} \geq 0 \quad \forall\, j = 1, .., m-1\,, \quad i = 1, .., m \tag{13.b}$$

Parameters γ_i, $i = 1, .., m$, determine the distribution of the total input u on the m reservoirs and parameters β_{ij}, $i = 1, .., m-1$, $j = 1, .., m$ determine the portion of the outflow of the jth reservoir, which is an inflow for the ith reservoir, i.e. the β_{ij} define the combination of the single reservoirs. The outflow $\psi_m \equiv y$ is the total outflow of the system. The function h_i describes the ith reservoir and $\hat{\alpha}_i = (c_1, .., c_l)^T$ is the model parameter vector of the ith reservoir. Some types of reservoirs are

$$h = c \qquad \textit{linear reservoir, Becker and Glos (1969)} \tag{14.a}$$

$$h = c_1 \psi^{c_2} \qquad \textit{exponential reservoir, Becker and Glos (1969)} \tag{14.b}$$

$$h = \sqrt{c_1 + c_2 \psi},\ c_1 \geq 0 \qquad \textit{quadratic reservoir, Diskin et.al. (1984)} \tag{14.c}$$

The distribution parameters γ_i and β_{ij} can be either fixed *a priori* during the model building proces or *a posteriori* as part of the parameter identification process. Since we have not considered automatic optimization of model structures, we made *a priori* choices for the β_{ij}. For example, a cascade of m reservoirs with automatically distributed total inflow is described by $\beta_{ij} = 1$ for $j = i-1$, $i = 2,..,m$ and $\beta_{ij} = 0$ otherwise. The model parameter vector α, which has to be estimated, consists of the reservoir parameters $\hat{\alpha}_i$ and the *a posteriori* distribution parameters γ_i.

SOME NUMERICAL RESULTS

Constructing a model, one should be aware, that with increasing number of free parameters the probability to achieve a good fit for one single reference data set will increase, too. But the optimum parameter vector for one reference data set might differ from the best choice for another reference data set. Moreover, increasing the number of free parameters can increase the number of local or even global minima. For example, if we consider two parallel reservoirs of the same type, interchanging the parameters of the reservoirs will *de facto* yield the same model, but the minimization problem (6) has two different solutions. Therefore, the total number of parameters should be kept small.

In the following, the results for a small sewer system with a total drainage area of $0.16km^2$ and a total length of $2.5km$ will be given. The detailed description for hydrodynamic simulation has 62 elements. The model used was a cascade of identical exponential reservoirs with automatically distributed input. It was observed that best results are achieved with one or two reservoirs in this example. The latter case will be presented here. The model equation with free parameters $\alpha_1, \alpha_2, \alpha_3$ can be summarized as

$$\begin{aligned} \frac{d\psi_1}{dt} &= \alpha_1 \psi_1^{\alpha_2} \Big(\alpha_3 u(t) - \psi_1 \Big) \\ \frac{d\psi_2}{dt} &= \alpha_1 \psi_2^{\alpha_2} \Big((1-\alpha_3)\, u(t) + \psi_1(t) - \psi_2(t) \Big) \end{aligned} \qquad 0 \le \alpha_3 \le 1 \qquad (15)$$

It should be mentioned, that the classical cascade of linear reservoirs is included in this model for $\alpha_2 = 0$ and $\alpha_3 = 1$. Five storm events were used for calibration, another five ones for verification. Instead of real world measurement data, smooth and artificially disturbed (up to 15%) results from hydrodynamic simulation were used.

Table 1. *Parameters α_i and the least square error χ (relative to reference data) for single data calibration of five storm events (smooth data);*

storm	1	2	3	4	5
α_1	0.152	0.111	0.123	0.180	0.148
α_2	0.238	0.131	0.324	0.165	0.143
α_3	0.633	0.045	0.490	0.764	0.403
χ[%]	0.04	0.2	0.08	0.09	0.19

From Table 1 it can be seen, that no unique optimum parameter is achieved by single data calibration for several storm events. Identification results of multiple data calibration for different weighting factors (7.a), (7.b) are shown in Table 2, where Z_{rel} is the value of the objective function relative to reference data. Comparing Table 1 and Table 2 indicates, that none of the single data optimum parameters is also best choice for multiple data calibration. The optimum parameters are not the same for different weighting factors, but the differences are much smaller than for single data calibration. In both cases (7.a) and (7.b) the relative error is of about the same order, (see Table 2). Of course, it must be larger for perturbed data than for smooth data. The estimated parameters are nearly the

Table 2. *Results of multiple data calibration: parameters α_i and value of objective function Z_{rel} for weighting factors (7.a) and (7.b), smooth and disturbed data*

	(7.a) smooth	(7.b) smooth	(7.a) dist.	(7.b) dist.
α_1	0.1542	0.1575	0.1539	0.1575
α_2	0.1614	0.1949	0.1622	0.1952
α_3	0.4823	0.5924	0.4824	0.5926
Z_{rel}[%]	0.215	0.185	0.714	0.683

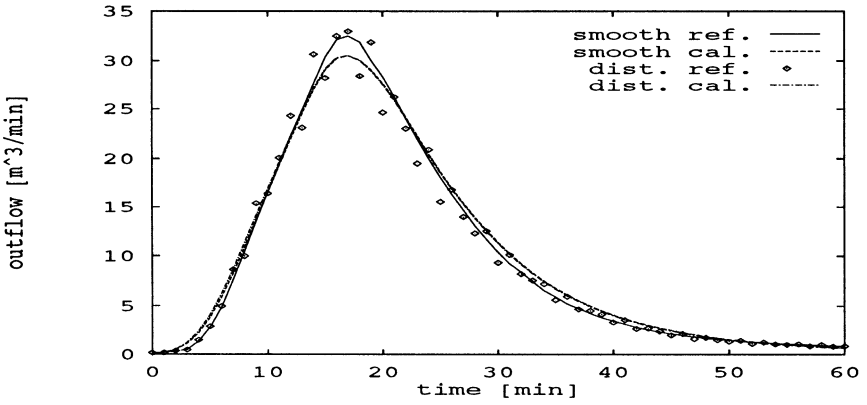

Figure 1. Calibrated model output and reference data for storm 1, smooth and disturbed data, weighting factors according to (7.a)

same for smooth and disturbed data in both cases (7.a) and (7.b). A graphical comparison of smooth and perturbed data calibration is given in Figure 1. It can be seen that the calculated outflow is identical in both cases.

To judge the results of the parameter identification and the model, however, we must take a look at the verification data sets (Table 3 and Figure 2). The least square errors in the second collection of these data sets No. 6–10, which are computed with the parameters estimated from data sets No. 1–5, are of the same order. This means that both, the model structure and the estimated parameters can be accepted.

Table 3. *Relative least square error χ for smooth calibration data sets 1–5 and smooth verification data sets 6–10; N is the number of data points*

storm	1	2	3	4	5	6	7	8	9	10
$\chi_{(7.a)}$[%]	0.258	0.227	0.155	0.222	0.207	0.154	0.193	0.163	0.093	0.374
$\chi_{(7.b)}$[%]	0.101	0.261	0.128	0.129	0.306	0.072	0.239	0.093	0.119	0.098
N	60	120	120	120	60	120	60	120	120	60

The values obtained for α_2 and α_3 indicate that in all cases the best choice calibrated model is distinct from the widely used classical cascade of linear reservoirs.

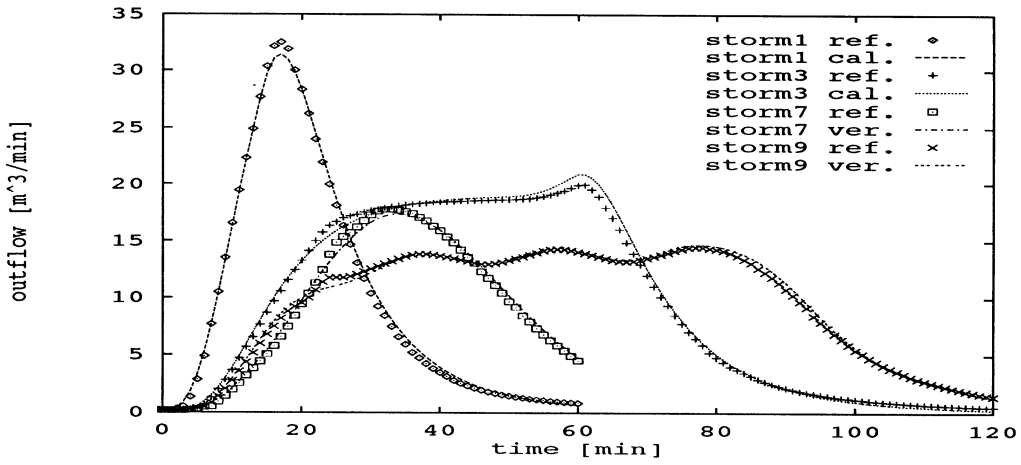

Figure 2. Results of model parameter identification: outflow for calibration data sets 1 and 3 (weighting factors according to (7.b)) and verification data sets 7 and 9 compared to reference data

CONCLUSION

A numerical method for multiple data parameter estimation for nonlinear ordinary differential equations of higher order was presented. It allows calibration of hydrological transport models for sewer systems which are much more comprehensive than the ones in current use. Moreover, the simple example given here shows that the method is able to deal with perturbations in reference data (e.g. oscillations) in a good way. From the general formulation (13) of hydrologic models one can hope to derive the best model structure automatically. But this has not been investigated.

REFERENCES

Becker, A, Glos, E (1969). *Grundlagen der Systemhydrologie.* Mitteilungen des Instituts für Wasserwirtschaft. VEB Verlag für Bauwesen Berlin, Heft 32

Diskin, M H, Boneh, A, Golan, A (1984). Identification of a Volterra series conceptual model based on a cascade of nonlinear reservoirs, *Journal of Hydrology*, **68**, 231-245

Eberl, H, Khelil, A, Wilderer, P (1996). Construction of simplified descriptions of a sewersystem to characterize its hydraulic behaviour, *Proc.7th Int.Conf. Urban Storm Drainage,* **3**, 1569–1574

Groetsch, C W (1993). *Inverse problems in the mathematical sciences.* Friedr. Vieweg & Sohn Verlagsgesellschaft mbH, Braunschweig/Wiesbaden

Khelil, A, Semke, M (1991). Quantifizierung der Unsicherheiten bei der Abflußberechnung in städtischen Entwässerungssystemen, *Report on the research project DFG SI 242/7-2 of the German Research Foundation, Vol. 2*

Marquardt, D W (1963). An algorithm for least-squares estimation of nonlinear parameters. *JSIAM*, **11**(2), 431–441

Ritter, K, Kredler, Ch (1992). *Nichtlineare Optimierung.* Schriftenreihe des Instituts für Angewandte Mathematik und Statistik, Technische Universität München, Nr.1, 2nd ed.

Stadler, W (1988). Fundamentals of multicriteria optimization. In: *Multicriteria optimization in engineering and in the sciences*, Stadler, W (Ed), Plenum Press, New York, 1–25

Stoer, J, Bulirsch, R (1990). *Numerische Mathematik II.* Springer-Verlag, 2nd ed.