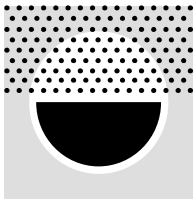
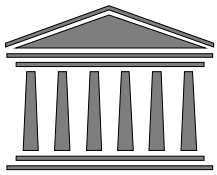




HANDOUT

MODELLING TERMINOLOGY AND METHODOLOGY FOR ACTIVATED SLUDGE MODELLING

Jacob Carstensen
Wolfgang Rauch

Dept. of Environmental Sc. & Eng., Technical University of Denmark

and

Peter A. Vanrolleghem

BIOMATH, University of Gent, Belgium

Kollekolle, Denmark, March 15 1998

Programme of the workshop:

10.00 - 10.15	Introduction
10.15 - 11.15	Model characterisation - constituents and attributes
11.15 - 12.15	Model building - methodology
12.15 - 13.45	*Lunch break*
13.45 - 14.45	Simulation tools
14.45 - 15.45	Model uncertainty
15.45 - 16.15	*Coffee break*
16.15 - 17.15	Model application
17.15 - 18.00	Discussion on model application

Table of contents

Annex

Carstensen, J., Vanrolleghem, P., Rauch, W. and Reichert, P. (1997). “Terminology and Methodology in Modelling for Water Quality Management - A Discussion Starter”, *Water Science & Technology, 36(5), pp. 157-168.*

1. Introduction

This workshop takes it's starting point in a 2000 years old Chinese mansion. This was during the Singapore biennial conference in 1996. For this venue many modellers from all around the world attended to present and discuss models as a means to improve water quality. The number of incidents with misunderstandings due to unclear modelling terminology and methodology was profound. This conference was by no means an exception in that respect. It is obvious that much confusion results from lack of common terminology. The present workshop and the paper in the annex for the WATERMATEX 97 conference is the result of the discussions that took place in Singapore.

The scientific society within activated sludge processes is comprised by a multitude of scientists with very different backgrounds, such as mathematicians, statisticians, microbiologists, environmental engineers, civil engineers, control engineers, ecologists, biologists, etc. These researchers have to communicate and collaborate across the linguistic, cultural and terminologic barriers in order to mitigate the pollution problems in urban areas. In fact, most pollution problems may only be solved by integrating several scientific disciplines. It is crucial to any kind of collaboration that a common reference of terminology is established and the communication gap is bridged.

There are two main source of scientific communication problems:

1. *misuse of existing terminology*
2. *use of unclear terminology*

The first main source is the most serious since it might potentially undermine the established terminology, while both sources lead to confusion. As a result, the message is often misunderstood by the recipient.

The terminology problem is closely related to the methodology problem - model building and application. It must be stressed that a universal model does not exist. Models have to be customised to the given application. Model building is a complicated task comprised of many steps and in each step there are many tools available. It is important to establish an overview of model types, model building, model application and the tools used in the processes.

The goal of this workshop is

- to clarify the use of modelling terms
- to establish an overview of model types
- to iterate the model building process
- to discuss the use of models (application and tools)

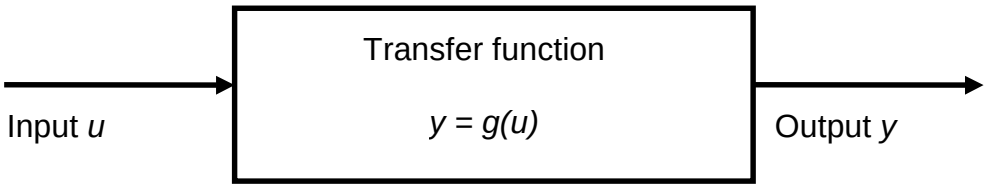
2. Model characterisation

The first fundamental question to be answered is: “what is a model?”. Philosophical speaking a model is an abstraction of reality. This means, that a model could be a scale-down physical model (e.g. pilot plant), an analogous representation (e.g. using electrical devices for simulation) or a mathematical model (e.g. a set of equations describing the real system), but it will never be a true representation of reality. In this workshop, we shall focus only on mathematical models and we shall refer to these as ‘models’.

2.1 Model structure

A model may be considered as a ‘machine’ transforming inputs (*u*) to outputs (*y*) by a set of relations (Casti, 1977; Chui and Chen, 1988). These relations can be formulated as algebraic equations, differential equations and/or partial differential equations.

If inputs *u* are transformed directly to outputs *y*, the model consists of transfer functions or alternatively, the structure of the model is *a transfer function*. This is illustrated below:

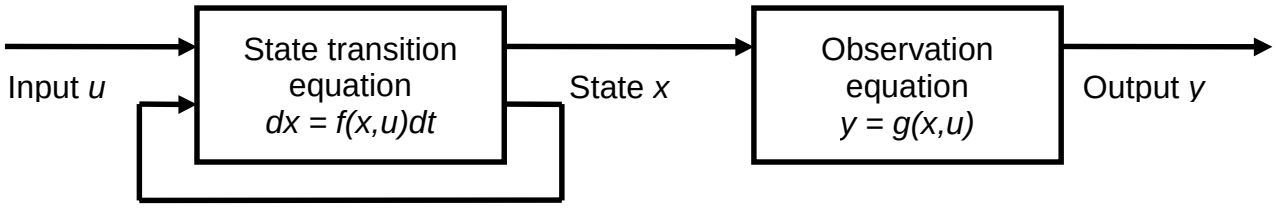


Consider the nitrification process

$$\frac{dS_{NH}}{dt} = -\frac{\mu_A}{Y_A} \bullet \frac{S_{NH}}{S_{NH} + K_{NH}} \bullet \frac{S_O}{S_O + K_{O,A}} \bullet X_{B,A}$$

This differential equation describes the transformation from one state to another in ammonia concentration as a function of the present ammonia and oxygen concentration. The autotrophic biomass is assumed to be constant.

The transfer function model structure may for some systems prove inadequate in representing the system, since the underlying mechanisms are not truly reflected in the model. For this reason, the *state-space model* approach was introduced in the early 1960s (Kalman, 1960).



The state-space model structure includes state variables (x) which act as mediators between the inputs and outputs. The state variables contain all information necessary to determine the state of the system. The state transition equations are in most cases formulated as differential equations, but difference equations are seen as well. Some of the state variables may not be observable (*unobserved components*).

Consider the nitrification process above. If we assume the autotrophic biomass to be a variable over time, X_{BA} and S_{NH} represent the internal state of the system, and S_{NH} is the output as well.

State transition equations:

$$\frac{dS_{NH}}{dt} = -\frac{\mu_A}{Y_A} \bullet \frac{S_{NH}}{S_{NH} + K_{NH}} \bullet \frac{S_O}{S_O + K_{O,A}} \bullet X_{B,A}$$

$$\frac{dX_{BA}}{dt} = \mu_A \bullet \frac{S_{NH}}{S_{NH} + K_{NH}} \bullet \frac{S_O}{S_O + K_{O,A}} \bullet X_{B,A} - b_{B,A} \bullet X_{B,A}$$

The observation equation becomes rather simple, since the measured output S_{NH} is equal to one of the state variables.

The structure of a model is given by the equations comprising the model (see later structural identifiability).

2.2 Model constituents

There are three basic components in a model:

1. variables
2. constants
3. parameters

Inputs, outputs and states are all *variables* in the model and as such also referred to as input variables, output variables and state variables. In the model above S_O is the input variable, S_{NH} and X_{BA} are state variables and S_{NH} is the output variable as well.

Model constituents which never change their value throughout all possible applications are termed as *constants*. In the example above there are no constants, but for example the gravity constant and molecular weights are constant within all applications of activated sludge processes.

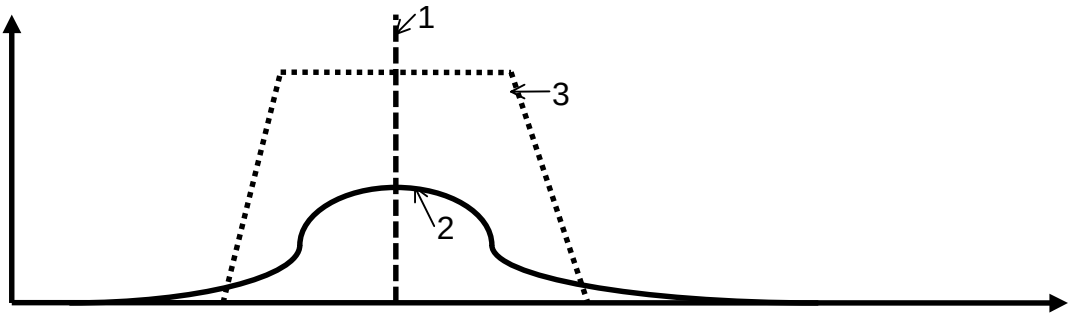
Model constituents which change their value according to the application are called parameters. The value of a parameter can be related to time, location or input variables which for the given application are constant (e.g. temperature and plant dependent parameters).

2.3 Data types

We shall consider the following three data types:

1. Deterministic data
2. Stochastic data
3. Fuzzy data

which are illustrated in the figure below:



Deterministic data is based on the assumption of exact knowledge and no uncertainty. A deterministic variable, parameter or constant is associated with one value as indicated on the figure above. It is common to assume model constituents to be deterministic data types, because the calculations involved become much simpler than using stochastic or fuzzy data types. In general, model constants can be assumed to be deterministic, while the asssumption of deterministic data types for model variables and parameters is more doubtfull.

Stochastic data is based on the theories of probability analysis. A stochastic variable or parameter is associated with a probability density function (pdf) as indicated on the figure above. The most commonly used pdf is the normal distribution which is characterised by a mean and a variance. In case the variance of a model constituent is neglectable relative to the mean and the application of the model, this model constituent can be assumed to be deterministic. Thus, the deterministic data type is a simplification of the stochastic data type where it is assumed, that the variation of the constituent is not significant relative to the application.

Fuzzy data is a relatively new data type which deviates significantly from the two previous. Similar to stochastic data, fuzzy data is based on the assumption of unexact knowledge (fuzziness), but unlike stochastic data fuzzy data does not depend on pdf's. Fuzzy data works on a membership function which should not be confused with a pdf. The membership function is indicated on the figure above. The fuzzy set theory is still not well developed and the application of fuzzy data types to activated sludge processes so far only relates to control issues. The fuzzy theory should to some extent resemble the human way of thinking (making fuzzy or unclear statements).

2.4 Model attributes

In the sections above, the building blocks of models have been described. However, there are an infinite number of ways these building blocks can be combined into a model. Therefore, it is necessary to add some adjectives or model attributes in order to characterise the model. Some model attributes have a clear and stringent definition which we shall refer to as *strong model attributes*, while some attributes are less stringent and thus termed *weak model attributes*.

2.4.1 Strong model attributes

Linear versus non-linear: These attributes relates to the structure of the model. Linear models are frequently used, because an analytical solution can be found. Non-linear models often require numerical methods to find the solution.

A model may be linear with respect to the variables:

$$Q = Q_{raw} + Q_{recycle} = (1 + r_{recycle}) \bullet Q_{raw}$$

where $r_{recycle}$ is the recycle rate, and as in

$$\frac{dX_{B,A}}{dt} = -b \bullet X_{B,A}$$

which is a linear differential equation for the decay of autotrophic biomass.

A model can be linear with respect to the parameters but non-linear with respect to the variables:

$$\frac{dS_S}{dt} = -\mu_H \bullet S_S \bullet X_{B,H}$$

where S_S and $X_{B,H}$ are the variables and μ_H is the parameter. In statistics the linearity in parameters is referred to as linear model.

Tong (1990) has also described a bi-linear model being a non-linear model which for fixed random coefficients is linear in the parameters and for fixed variables is linear in the parameters.

Dynamic versus static: In case the variables evolve over time the model is termed dynamic otherwise static. The flow balance above is a static relationship if Q_{raw} is constant over time. If the model is dynamic this is frequently formulated by indexing t to the variables or putting t in parentheses after the variable name.

$$Q(t) = Q_{raw}(t) + Q_{recycle}(t) = (1 + r_{recycle}) \bullet Q_{raw}(t)$$

Time-invariant versus time-variant: If the model parameters are constant in time the model is characterised as time-invariant, otherwise time-variant. In the nitrification process equation we may consider the autotrophic growth rate $\mu_A(t)$ to be a time varying parameter when the run length of the model covers a longer period of time.

Distributed parameter model: Similar to the time-dependency a model can have a space-dependency. This is frequently seen in clarifier models and models for the runoff from catchments. Such models are formulated as partial differential equations involving derivatives with respect to both time and space.

Discrete-time versus continuous-time: These attributes relates to whether the time-dependency is formulated by difference or differential equations. The model for substrate degradation above in discrete-time is as follows:

$$\Delta S_S = -\mu_H \bullet S_S \bullet X_{B,H} \bullet \Delta t$$

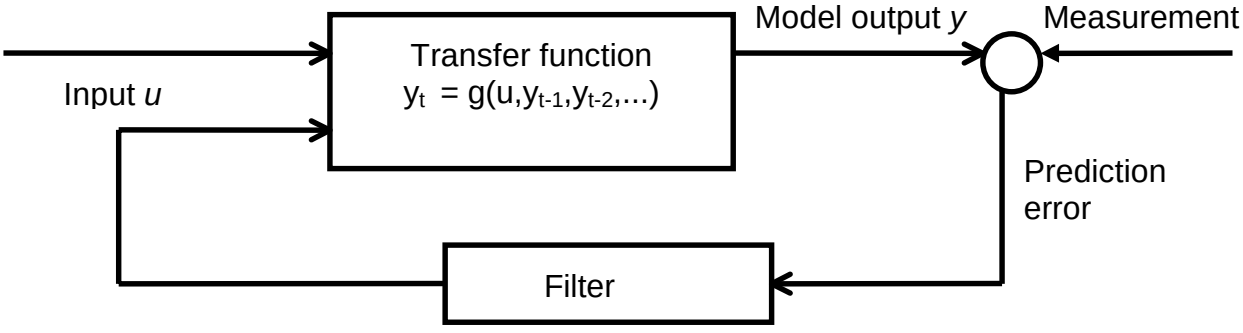
Most computer programs for activated sludge processes use a discretisation of the differential equations.

Discrete-space versus continuous-space: Similar to the above the partial differential equations can be formulated as difference or differential equations with respect to the spatial distribution. Most computer programs for activated sludge use a lattice discretisation for the partial differential equations.

Stochastic versus deterministic: If one constituent of a model is assumed to be of stochastic data type, the model is stochastic, and similarly if all the model constituents are assumed to be of deterministic data type, the model is deterministic. Model based on fuzzy data are not considered in the present context, since they are mainly used in controllers. Model constituents are frequently split up into a deterministic part (the mean value) and a random variation with zero mean:

$$X = \overline{X} + e \qquad \text{where e follows a random distribtion}$$

Adaptive models: This is a term frequently used in models for control. The term relates to a model which has a feedback from previous output measurements as input to the model. Since all measurements contain elements of randomness, adaptive models are in their nature also stochastic.



2.4.2 Weak model attributes

Simple versus complex: These attributes relates to the degree of simplicity or complexity in the model, which is a subjective measure of the number of equations, parameters and their relationship. The lack of clarity in using this attribute is obvious when considering the activated sludge model no. 1. This model is simple relative to model no. 2 but complex relative to reduced order models as in Jeppsson (1995). Thus, these terms only have a meaning when another model is given for reference (e.g. this model is more simple/complex than the reference model). For this reason the terms segregated and aggregated/lumped are commonly used when one model is derived from another.

Mechanistic versus phenomenological: These terms are used to describe the degree to which a model is based upon physical, chemical and biological laws. Mechasnistic, physical, white-box and transparent are attributes used to stress that a model is based upon laws of physics, chemistry and biology (deductive). Phenomenological, black-box, empirical and heuristic are terms used to denote that the model is based on empiricism (data-driven or inductive). Between the two categories are hybrid models which have a structure based on partly physical, chemical and biological laws and partly empiricism. These models are also referred to as grey-box models or semi-physical models.

Although the trend has been to use either deterministic, complex, mechanistic models or stochastic, simple, phenomenological models, these terms do have different meanings. These terms are also strongly connected to the concepts of identifiability as shall be introduced in the next section.

2.4.3 Examples of model attributes

Again, consider the Activated Sludge Model no. 1 (Henze *et al.*, 1986). This model should accordingly be termed a *non-linear, dynamic, time-invariant, continuous-time, deterministic, mechanistic model*. Whether the model is simple or complex depends on the model it is compared to (following the discussion above). Since the ASM1 and ASM2 are reference models known to the scientific community, these two models should be used for characterising other models.

In the next example we re-iterate the nitrification process once more. This model can be made stochastic by adding a stochastic error term (data type) to the differential equation:

$$\frac{dS_{NH}}{dt} = -\frac{\mu_A}{Y_A} \bullet \frac{S_{NH}}{S_{NH} + K_{NH}} \bullet \frac{S_O}{S_O + K_{O,A}} \bullet X_{B,A} + e_t$$

The error term in the equation above is known as a *structural error*. This error compensates for lack of explanatory power in the kinetic expression. Thus, the magnitude of this error term depends on how well the kinetic expression describe the ‘real world situation’.

The parameters of a model may also be considered to be of stochastic data type:

$$\frac{dS_{NH}}{dt} = -\frac{\mu_A + e}{Y_A} \bullet \frac{S_{NH}}{S_{NH} + K_{NH}} \bullet \frac{S_O}{S_O + K_{O,A}} \bullet X_{B,A}$$

In this case the autotrophic growth rate is associated with a mean value μ_A and an error term. The error term can be formulated as an unknown stochastic bias (unknown parameter or parametric uncertainty) in the equation or a stochastic process making the autotrophic yield coefficient a time-varying parameter (which in fact is a variable).

Finally, error terms can be added to input variables, which in the case above would be on the measurement of oxygen concentration. These different uncertainty sources in a model are investigated more thoroughly in section five in combination with tool for assessing magnitude and propagation.

An adaptive model can be obtained by including lagged values of e_t in the model as structural error:

$$\frac{\Delta S_{NH}}{\Delta t} = -\frac{\mu_A}{Y_A} \bullet \frac{S_{NH}}{S_{NH} + K_{NH}} \bullet \frac{S_O}{S_O + K_{O,A}} \bullet X_{B,A} + e_t + \theta e_{t-1}$$

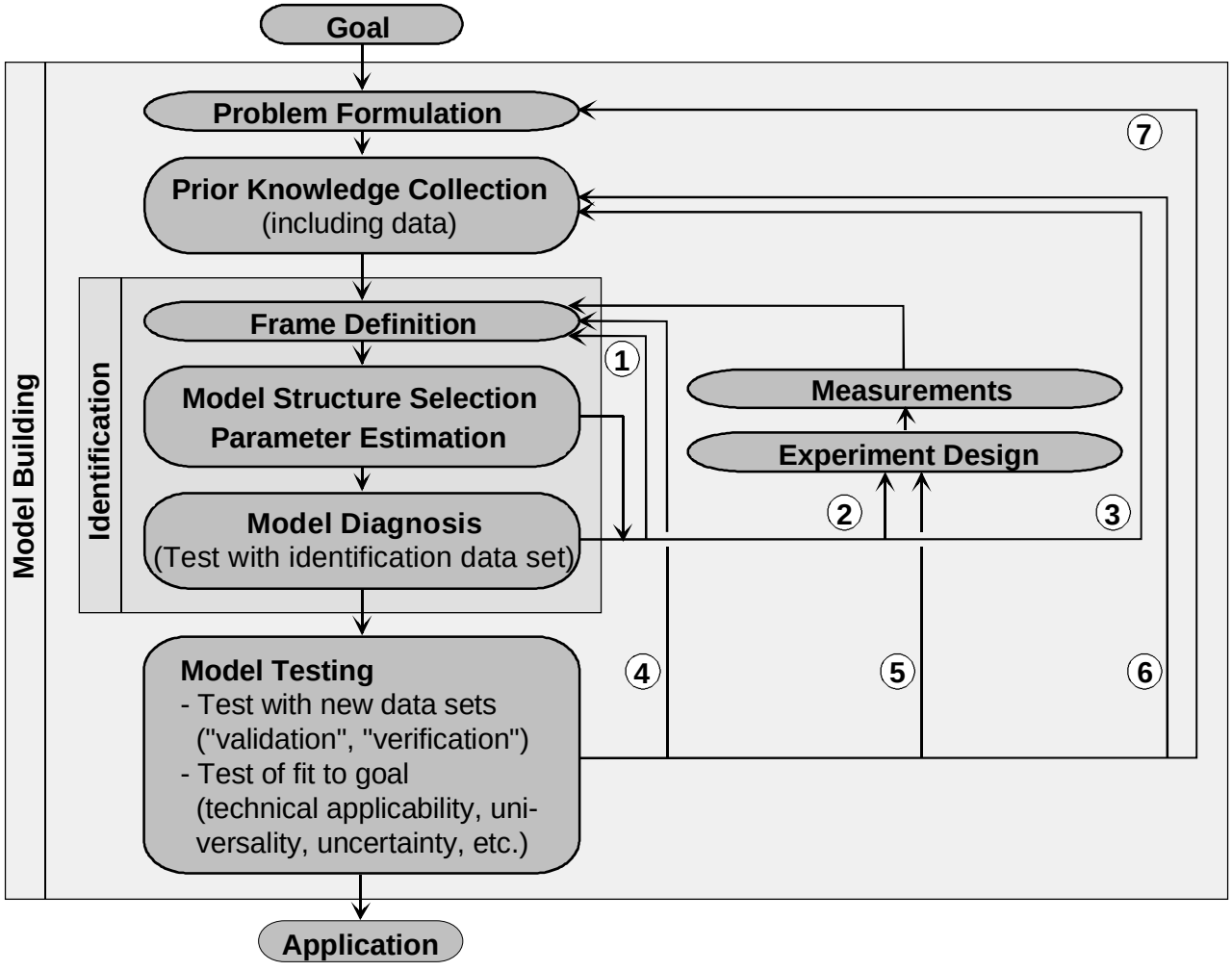
where the prediction error of the previous measurement is fed back as input to the model again.

3. Model building - methodology

Using the “story” of a model building exercise, a number of terms involved in this activity will be introduced within their appropriate context. At the same time a short review is given on the current state-of-the-art in modelling methodology. The diagram in the figure below summarises the aspects of model building which are described in detail below. Only when all the steps in the figure have been fulfilled successfully, the model can be applied for its intended purpose. These applications typically involve **simulation** that may be regarded as virtual experimentation with the virtual reality described by the model.

3.1 Problem formulation

An often forgotten task in a model building exercise is the clear formulation of the goal of the model that is to be constructed. While in most cases this task is rather intuitively performed by the modeller in case he is also the problem supplier, problem formulation or goal incorporation is much more difficult in case these persons are not the same. In this case, an important effort must be spent to answer such questions as accuracy of the results, degree of uncertainty in the provided answers, time scale of the solution, system boundaries, important variables, environmental conditions for which the model must give an answer, etc.



3.2 Prior knowledge collection.

The next task is to collect the available, relevant a priori knowledge from literature and experts or from a model building environment that supports re-use of model-encapsulated knowledge. At this (early) stage of the exercise, some experiments may be conducted or some data collected during previous experiments may be retrieved and stored in the experimental database.

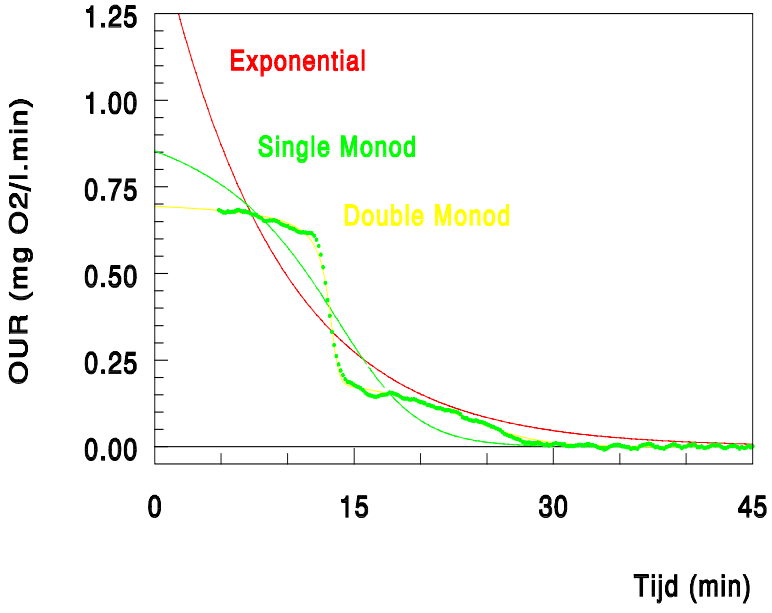
3.3 Frame definition.

As soon as these two tasks have been performed once, a first iteration of the model building can start. The **frame definition** phase aims to delineate the conditions under which a model will be used (e.g. temperature), to choose the class of models that seems fit for the task (time series, state-space, distributed parameter, stochastic...), to specify the variables that seem important to find a solution to the formulated problem (inputs, outputs, states), the range of time constants that need to be covered by the model, etc.

3.4 Model structure selection.

With this frame of work defined, candidate models may be constructed combining the collected a priori knowledge and the creativity of the model builder. In the figure above three candidate models have been fitted to the respirometric data set and the question, albeit rather easy to answer in this example is to decide which model is best for the purpose. The goal of **classical model structure selection** (**model structure characterisation**) is to select a unique model structure according to the principles of a quality of fit and parsimony (Spriet, 1985; Harvey, 1989). However, it is also possible to select a set of models that are attributed different weights reflecting their probability of appropriateness (Draper, 1995; Reichert and Omlin, 1996).

Most model structure selection criteria require the parameter values θ to be estimated (see below) on the basis of *Ndata* measurements and from that a measure of the residual errors (*SSR*).



However, structural selection criteria that only require basic data analysis also exist for particular applications (Vanrolleghem and Dochain, 1998). The advantage of these methods is that they don't require computationally intensive parameter estimations to be performed for the different candidate models.

Since the model selection criteria that require prior parameter estimation are more generally applicable, some details are given here on these methods. A more extensive overview is preseted in Vanrolleghem and Dochain (1998). First, the so-called information criteria should be mentioned, such as AIC (Akaike, 1974), FPE (Ljung, 1987) and BIC (Schwartz, 1978).

$$AIC = N \log\left(\frac{SSR}{Ndata}\right) + 2\dim(\theta)$$

$$FPE = \frac{SSR}{Ndata}\left(1 + \frac{2\dim(\theta)}{Ndata - \dim(\theta)}\right)$$

$$BIC = N \log\left(\frac{SSR}{Ndata}\right) + \dim(\theta)\log(Ndata)$$

Methods that go back to statistics can also be used for model selection, e.g. the well-known F-test in which the residuals *SSR* of two models, a complex one (M_j) and a simple one (M_i) are compared:

$$F_w = \left(\frac{SSR(M_j) - SSR(M_i)}{\dim(\theta_j) - \dim(\theta_i)}\right) \Big/ \frac{SSR(M_j)}{Ndata - \dim(\theta_j)}$$

This calculated value is to be compared with tabulated values for a significance levels α $F_\alpha(dim(\theta_j)\text{-}dim(\theta_i);Ndata\text{-}dim(\theta_j))$. Notice that quite some similarities exist between the F-test and the above information based criteria and also with the X^2 test (Söderström and Stoica, 1989)

Finally, very powerful techniques are available in which an analysis is made of the residuals between model predictions and measured data

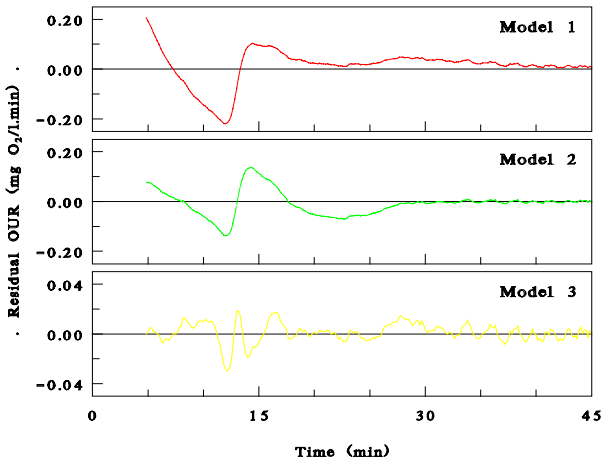
$$\varepsilon(t_k) = Y(t_k) - Y(\theta, t_k)$$

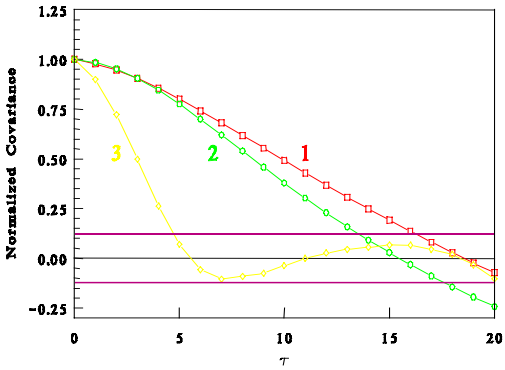
In the figure below, the residuals are plotted for the respirometric example introduced above. In many applications, residuals are supposed to be randomly distributed and independent of each other. This is obviously not the case for the first two models since consecutive residuals have nearly always the same sign. The so-called **runs test** evaluates the number of sign changes in the residual sequence and compares that to the expected number *N/2* (Söderström and Stoica, 1989).

Another method for evaluating the properties of the residuals sequence is the autocorrelation test. An autocorrelation with *time lag* τ quantifies the dependency of a variable at any time t_k and the variable at time (t_k-τ):

$$r_\varepsilon(\tau) = \sum_{k=1}^{Ndata-\tau} \varepsilon(t_k - \tau) \cdot \varepsilon(t_k)$$

In the figure below-right, the autocorrelations for the residuals of the figure below-left are plotted for time lags of 0 to 20. Obviously, a time lag of zero means that the correlation is perfect, i.e. $r_\varepsilon(0)$=1 and correlations over longer time lags decrease.





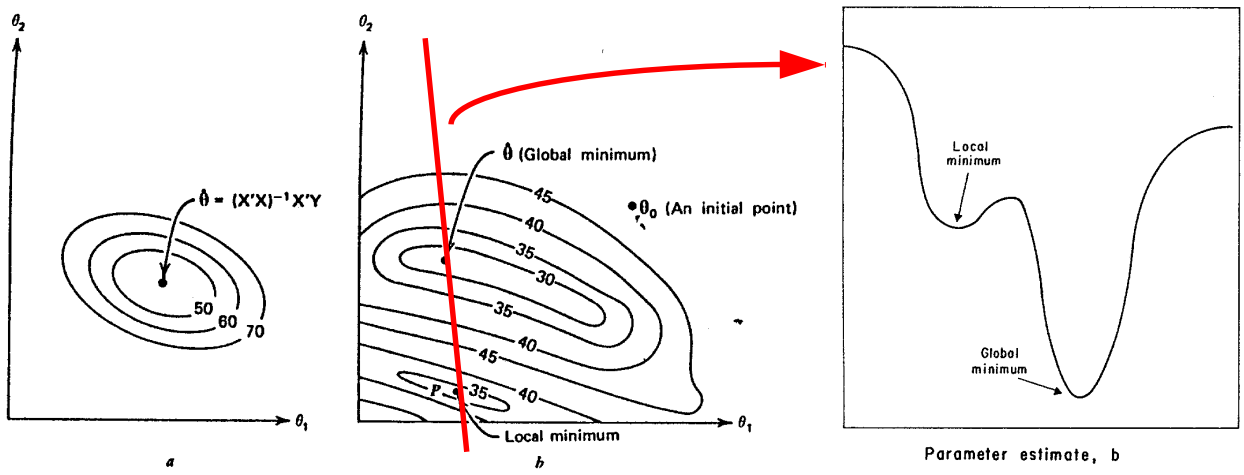
A suggested model selection criterion consists of comparing the value of the autocorrelation r_ε for each lag τ with the limit value $N(0,1)/\sqrt{(Ndata)}$ with $N(0,1)$ the standard normal distribution. For a significance level α, for instance, this means that only 5 percent of the autocorrelations may be larger than $1.96/\sqrt{(Ndata)}$. The lines for these critical levels are included in the autocorrelation plot given above. It is apparent that neither of the three models is acceptable from a residuals analysis point of view, but model 3 is obviously the better one.

3.5 Parameter estimation.

Parameter estimation is based on the maximisation or minimisation of a goodness-of-fit criterion such as Least Squares, Weighted Least Squares, Maximum Likelihood, etc. Below a quite general weighted least squares criterion is given in which multivariable (number of variables *Nvar*) datasets are available for *Nrespons* experiments.

$$J(\theta) = \sum_{k=1}^{Nrespons} w_k \sum_{j=1}^{N\,var_k} w_{jk} \sum_{i=1}^{Ndata_{jk}} w_{ijk} \left(Y_{ijk}(t_{ijk}) - Y_{ijk}(\theta, t_{ijk}) \right)^2$$

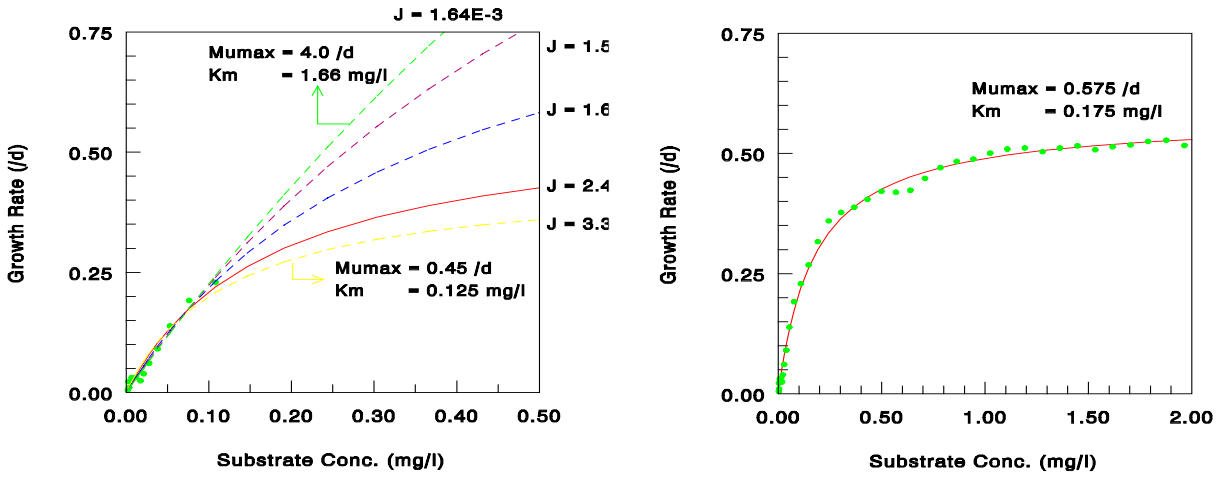
Optimisation of such criterion function aims to provide values for the parameters in the model and, in some cases, also values for the initial (and boundary) conditions of the state variables (in case a state space representation is adopted). Several powerful estimation (non-linear **regression**) algorithms are available. In case only one data point is used at a time to estimate parameter values, one uses the term **recursive estimation** and when the whole data set available is used at once to obtain the parameter values the term **batch estimation** is applied.



Estimation is a basic optimisation problem that is subject to a number of pitfalls, the most important being the presence of local minima in the criterion (or objective or cost) function. This is illustrated below. On the left the topography (contour lines of criterion function values) of the criterion functional is given for a linear optimisation problem. We see ellipses and can imagine that an optimisation algorithm will easily find the lowest value (imagine a ping pong ball rolling at once to the lowest point when starting from different initial locations (initial parameter guess)). In the middle of the figure the same topography is shown for a non-linear optimisation problem (as we are facing nearly always when dealing with wastewater treatment process models). On the right, a slice through the surface is made along the think line of the centre figure. One observes multiple minima and one can imagine that the optimiser (the algorithm that directs the "ping pong ball" to the minimum) may have problems in finding the global minimum, i.e. finding the best parameter estimates. The algorithm may get stuck in a local minimum, yielding a non-optimal parameter set.

The success of the parameter estimation under such difficult (non-linear) conditions is highly dependent on the experimental dataset available. To find out how the chances are for finding reliable parameter values one can (and should) perform an identifiability analysis.

The identifiability analysis performed prior to the parameter estimation itself can provide answers to the key question whether, given a set of measured variables, unique parameter values can be obtained. Two types of answers can be given depending on the applied method. In case **structural** (also termed **theoretical** or **a priori**) **identifiability** (Godfrey and DiStefano, 1987; Norton, 1986) is evaluated the answer is either yes or no, respectively meaning that the parameters can be given unique values or not at all (Dochain et al., 1995). However, it is not ensured that the data always contain sufficient information to provide unique estimates (e.g. in the model $Y=aX_1+bX_2+c(X_1+X_2)$ only the parameter combinations $a+c$ and $b+c$ are structurally identifiable and not the three parameters a, b and c. If in a model $Y=aX_1+bX_2$ measurements for Y are available for given X_1 and X_2, the model is structurally identifiable but it will be practically unidentifiable if the dataset contains only values of Y for $X_1=\alpha X_2$). Methods for structural identifiability analysis are mathematically sophisticated and are reviewed within an application context in bioprocess modelling in Vanrolleghem and Dochain (1998).

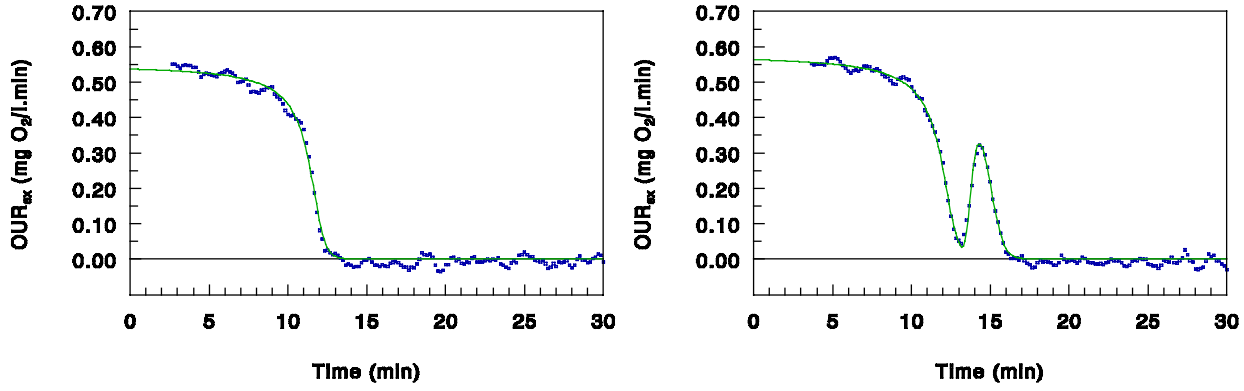


changing for very different values of parameter values, which implies that the optimisation algorithm will have difficulties in finding the best value. On the right hand side a much better data set for estimation of Monod parameters is illustrated.

Methods for the **practical** (also termed as **numerical** or **posteriori**) **identifiability** study are available and allow to evaluate the information content of the dataset intended for parameter estimation (Vanrolleghem and Dochain, 1998).

The basis of the methods for practical identifiability analysis is also underlying methods for **optimal experimental design** that can provide a solution to an encountered practical identifiability problem. This design procedure uses the model for which reliable parameter estimates are to be found to calculate experimental conditions such that sufficient information is contained in the data. An example taken from Vanrolleghem *et al.* (1995) illustrates that by simple extension of an experiment, considerable increases in estimation accuracy can be achieved. Below on the left a normal batch respirometric experiment is shown. The respiration rate profile is the result of the injection of substrate at the beginning of the experiment. On the right, a similar experiment is performed, but after 13 minutes, an additional pulse of substrate is injected in the sludge and one observes the respiration rate to increase again. Parameter estimation on both data sets revealed that the parameter variance decrease with a factor 2 that was predicted by the optimal experimental design procedure was indeed achieved.

One should note that a structural identifiability problem encountered cannot be solved without altering the candidate model or



frame definition (e.g. include other variables in the system description). **Model reduction** can lead to models that become less “data-hungry” and hence their identifiability properties may improve (Jeppsson and Olsson, 1993).

3.6 Model diagnosis.

Once the parameters are estimated it remains to investigate whether the identified model violates the assumptions made in the frame definition. For instance, statistical tests of systematic deviations between model results and measurements (residuals) and their distribution are frequently used (see above).

3.7 Model testing.

Fitness of a model can be evaluated by comparing its performance with data obtained under different conditions than the ones prevailing at the time of the data collection performed for model identification. This process of putting the model in jeopardy (Boyle and Berthouex, 1974) or, in other words, straining the model to its limits, may reveal model inadequacies that may be sufficient to conclude that the model is no longer “fit” for the purpose it was intended for. Hence, the whole model building process may have to start all over. Sometimes this may even lead to a reformulation of the problem as the modelling exercise has provided considerable insight in the system under study and its behaviour.

This process of putting the model into jeopardy by confronting it with new data is most often called **validation**, but serious arguments are put forward against this term. As a model only describes part of the reality (the one defined in the frame) in a simplified manner, it is obvious that a model never can describe reality completely. Therefore, there will always exist experimental conditions for which the model is not valid. Hence, validation of a model is utopian! A completely other approach is to term this process of jeopardising the model a model **falsification** step (Caswell, 1976; Reckhow and Chapra, 1983), which if answered negatively, provides more confidence in the selected model. However, the term falsification appears too negative and one has therefore looked for other terminology that is less pronounced (quantitative) as validation but still gives a qualitative insight in the level of confidence one has in a selected identified model. The terms put forth for this are **corroboration** and **confirmation** (Popper, 1980). Finally, the term **verification** is frequently interchanged with validation, but the use of the term validation is advocated here.

To quantify the validity of a model recurrence is again made quite often to the residual analysis methods mentioned above.

4. Simulation tools

Differential equations are fundamental to the mathematical description of the dynamic behaviour of systems and hence applied routinely in science and engineering. However, all but the most simple set of equations cannot be solved analytically but require the use of computers and numerical procedures for solving. Scientist and engineers are predominately concerned with the formulation of the problem, that is, with setting up the equations. The aspect of the numerical solution procedure is usually not given much consideration and therefore the use of rather simple numerical methods is preferred. However, the application of crude procedures such as the popular Euler method is, at best, inefficient. In the worst case the procedure is unstable, meaning that the numerical results deviate widely from the true solution with the error increasing exponentially in each time step.

The choice of numerical routine and its implementation has serious implication on the simulation result and the time of the simulation. This presentation deals with the different methods for solving these differential equations in a numerically stable manner by giving an overview on popular integration algorithms and their computer implementation.

4.1 Ordinary differential equations (ODE's)

4.1.1 Order and degree of differential equations

The main feature of ordinary differential equations is that they contain only one single derivative variable. In science and engineering this is usually time t. The general form of a system of first order differential equations is

$$\frac{dy_i}{dt} = f_i(t, y_1, \ldots, y_N, x_1, \ldots, x_m) \qquad i = 1,\ldots,N \qquad (1)$$

where	$y_1,\ldots y_N$	vector of dependent variables with N components
	$f_1,\ldots,f_N$	derivative vector
	t	independent variable (usually time)

Observe that the term order is related to the highest order of the derivatives in t. That is, a first order differential equation must contain only derivatives as dy_i/dt and no higher order derivatives as e.g. d^2y/dt^2. The following discussion is restricted to the fundamental problem of first order differential equations. Still, we can apply all methods discussed here also to higher order ODE's as these can generally be rewritten as 1st order ODE's (by introducing new variables).

One should also not misinterpret the term first order with the degree of the dependent variable y_i in the derivative functions. More precise, the degree of an equation is the greatest power to which the variable in the derivative function is raised. Thus e.g. the variable y_1 in the derivative function $f_1 = y_1^2$ is of 2nd degree but the ODE is still of first order.

Another important issue for differential equations is the concept of linearity. Generally an differential equation is said to be linear when the dependent variables and its derivatives appear (1) only to the zero or first degree in the equations and (2) no products of the variables and its derivatives occur. For example, $\frac{dy}{dt} = \frac{y}{k+y}t$ and $\frac{dy}{dt}y = t$ are both nonlinear equations whereas $\frac{dy}{dt} = ye^{-t}$ is linear.

4.1.2 Boundary conditions

For specifying the true nature of the problem one has also to define the boundary conditions of the problem. These are the predefined values of the vector y_i in equation 1 at specific (discrete) points. The solution procedure has to satisfy those conditions and therefore the nature of the boundary conditions is determining also the numerical method feasible. In essence there are only two categories:

A. Initial value problem:

Initial value problems are the most frequent ones in the mathematical description of physical systems. The problem is specified by a given set of values for y_i at some starting point $t = t_0$. The aim is to find y_i values at subsequent points t_f. The interval $[t_{0,...,} t_f]$ is also denoted as simulation timestep h and $y_{1(t0)},... y_{N(t0)}$ the initial condition vector.

B: Boundary value problem

Not surprisingly, in this kind of problem sets boundary conditions are specified for more than for one point t. Quite frequently we have a given set of values for y_i at the start and the end of the simulation period (2-point boundary problem). The solution of boundary value problems requires fairly complicated procedures (see e.g. Press et al, 1992).

4.1.3 General Classification of solution procedures

Since ODE's appear in nearly all fields of science and engineering, the topic of numerical solution has received much attention. The approaches developed are generally to be classified into the three categories (1) Single step (Runge Kutta) methods, that use information from a single point t_n for predicting the values of y_i at the next point t_{n+1} (2) Extrapolation (Richardson) methods and (3) Multistep (Predictor-Corrector) methods, in which information from several preceding timesteps is used to obtain the y_i values at the point t_{n+1} iteratively. Single step methods are most frequently applied as these methods are rather simple to understand and work adequately for a wide range of applications.

4.1.4 Accuracy and Stability

Accuracy and stability are the two fundamental characteristics of numerical procedures. Both issues are connected with the accumulation of the local truncation error of the procedure (i.e. the error that occurs during the computation of one time step) when proceeding the integration. The total error that has accumulated at the end of the simulation period is denoted as global truncation error. Accuracy is the problem of loss of precision in the numerical procedure as compared to the true solution. However, a

solution that is only inaccurate must not be fundamentally wrong. The prediction just deviates from the analytical solution of the mathematical formulation of the problem. If the loss of precision in the numerical method gets catastrophic the method is said to be unstable. A formal definition of the term stability is the constraint, that the global error (the difference between the numerical result and the analytical solution) must be bounded for all values of the independent variable:

$$\|y_n - Y_n\| = err_n \leq C \tag{2}$$

where: y_n is the numerical result at $t = t_n$
Y_n is the analytical solution at $t = t_n$
err is the global truncation error
C is the constant constraint for ensuring stability

Stringent stability analysis is generally limited to linear equations with constant coefficients. Nonlinear problem can hardly ever been analysed in such way that exact stability boundaries can be derived. One way to circumvent the problem is to linearize the equations and use the result as an approximation. However, frequently we simply try to bound the global error by adapting the stepsize h in order to maintain accuracy in the result - stability comes then as a subset of the approach (but is not guaranteed).

4.1.5 Finite differences - basic theory

The solution of differential equations by finite differences is generally one of the most important aspects in computational mathematics. The method is intuitive, relatively simple to implement and serves also as an adequate numerical procedure for a wide spectrum of differential equation problems. The following discussion is focusing on the aspect of ODE's where most of the traditional methods (Runge-Kutta, Gear's, Adams etc.) are actually subsets of the basic formulations.

Differential equations are formulated in a continuous domain, that is, in equation 1 the values of y_i are specified for any instance of the independent variable time. Finite difference methods use a numerical grid for transforming the mathematical expressions into a discrete domain, where the values of y_i are specified for only a finite number of points. These points are denoted as nodes and the numerical solution is then an approximation of the true solution at exactly those nodes. Finite difference approximations are accordingly discrete approximations of the derivative functions at specific points. The spacing of the independent variable is denoted as discretization.

Taylor series expansion is the fundamental underlying concept of finite difference methods. Consider the Taylor expansion for dy/dt = f(t,y) at point $t = t_0$:

$$y_{(t)} = y_{(t0)} + y^1{}_{(t0)}(t - t_0) + \frac{y^2{}_{(t0)}}{2!}(t - t_0)^2 + \frac{y^3{}_{(t0)}}{3!} + \ldots \tag{3}$$

with $h = t - t_0$ and using the more general subscript n for t_0 follows

$$y_{n+1} = y_n + hy^1{}_n + \frac{h^2}{2!}y^2{}_n + \frac{h3}{3!}y^3{}_n + \ldots \tag{4}$$

where $y^q{}_n = \frac{d^q y}{dt^q} = f^q(t_n, y_n)$ q = 1,2,.. is the order of the term written as $O(h^q)$

If the Taylor series expansions is truncated after the second term (first order term $0(h^1)$) we get Euler's method:

$$y_{n+1} = y_n + h y^1{}_n \tag{5}$$

The advantage of this theoretical sound way to derive a singe step procedure (here Euler's method) is, that we get also an estimate of the local truncation error involved. Eulers methode is of 1st order accuracy. The error results from leaving all higher order terms in the Taylor series expansion and is therefore said to be of 2nd order, i.e. $O(h^2)$.

4.1.6 Adaptive stepsize

From the theory of ODE's it is clear that accuracy and stability of numerical solution methods depend on an adequate choice of the stepsize. However, this choice of the stepsize is always a trade-off between the need for accuracy and stability on the one hand and our wish to minimise the computational effort on the other hand..

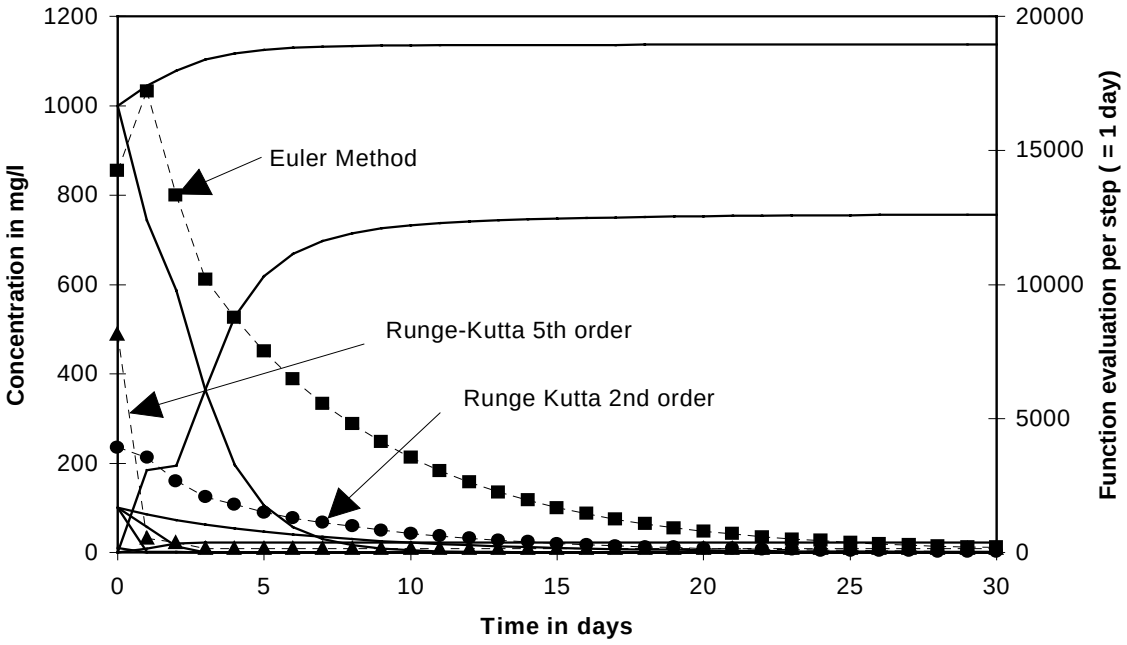


Figure 4.1: Computational efficiency of three algorithms expressed as number of function evaluations per timestep for a batch reactor problem.

Using a solution procedure with a constant stepsize requires that the stepsize is sufficiently small to guaranty accuracy in the worst case. But this may well be an overkill in a wide range of the simulation. The solution to this dilemma is an algorithm that is capable to adapt the stepsize to current requirement. The example in Figure 4.1 outlines the effect of such algorithms when proceeding through a simulation

4.2 Partial differential equations

4.2.1 General form of partial differential equations

Partial differential equations (PDE's) are distinguished from ODE's by the number of independent variables. While ODE's contain only one independent variable (which is usually time) PDE's have additional ones (usually in space). The concepts of order and degree of differential equations apply here as well. A general form of a system of partial differential equation in one-dimension is:

$$u_{i,t} = f_i(t, x, u_i, u_{i,x}, u_{i,xx}, \ldots u_{i,nx}) \qquad \text{with } i = 1 \ldots N \qquad (10)$$

where u_i = vector of dependent variables
x = independent variable in space dimension
t = independent variable in time
f_i = vector of derivative functions

Subscripts are used to indicate derivatives:

$$u_t = \frac{\partial u}{\partial t} \qquad u_x = \frac{\partial u}{\partial x} \qquad u_{xx} = \frac{\partial^2 u}{\partial t^2} \quad \text{etc.} \qquad (11)$$

4.2.2 Boundary conditions

For a specific solution of any differential equation, boundary conditions are required. In case of PDE's these are predefined values of the dependent variables u_i at specific time and space locations. Boundary conditions are of paramount importance for the dynamic behaviour of the system and the specification of those conditions is usually a critical issue in the numerical scheme. The two fundamental types of conditions are:

- Dirichlet boundary conditions: the vector u_i is specified in the x, t plane.
- Neumann boundary condition: the derivative functions of u_i is specified in the x, t plane.

A general rule for the number of required boundary conditions for a specific problem is the following: For each independent variable, there must be as many conditions as the highest order of the derivative for that variable.

The following example outlines several types of boundary conditions applied to the 1-d advection/diffusion equation. This equation is of paramount importance in water quality modelling as it specifies the dynamic transport of a non-reactive substance in an incompletely mixed, longitudinal system with uniform flow:

$$\frac{\partial u}{\partial t} + v\frac{\partial u}{\partial x} - D\frac{\partial^2 u}{\partial x^2} = 0 \qquad (12)$$

where u = concentration of the substance in ML^{-3}
v = uniform flow in LT^{-1}
D = longitudinal dispersion coefficient in L^2T^{-1}
x = space coordinate in L
t = time in T

According to above we need 3 boundary conditions, two in space (x) as there is a 2nd order derivative in x and one in time. A typical example of those three conditions is:

* initial conditions in time ($t = t_0$) for all x :	$u_{(t0, \bar{x})}$	Dirichlet condition
* upper boundary conditions ($x = x_0$) for all t:	$u_{(\bar{t}, x0)}$	Dirichlet condition
* lower boundary condition ($x = x_f$) for all t:	$\frac{\partial u}{\partial x}_{(\bar{t}, xf)}$	Neumann condition

The Neumann condition at the lower end of the system is either a specific vector of values or a constant, most frequently zero (i.e. gradient of the substance concentration = 0). This example of boundary conditions is represented in Figure 4.2. If the 2nd order dispersion term is dropped (assuming pure advective flow - plug flow reactor) then the PDE reduces to $\frac{\partial u}{\partial t} + v\frac{\partial u}{\partial x} = 0$ and the 3rd boundary condition (2nd condition for the space derivative) is no longer necessary.

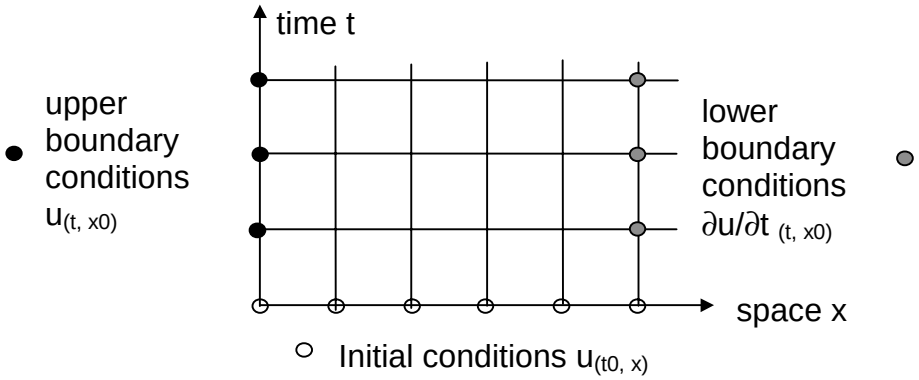


Figure 4.2: Boundary conditions for a PDE problem

4.2.3 Classification of PDE's

According to the nature of the differential equations it is possible to classify PDE's in a geometric way, i.e. as either of elliptic, parabolic or hyperbolic type, where the classification is relating to the characteristic curves of the solution. In a simplified manner the classification can be based on the following considerations:

* Elliptic PDE's:	only independent variables in space and none in time.
* Parabolic PDE's:	1st order derivative in time and 2nd order derivatives in space.
* Hyperbolic PDE's	1st order derivative in time and 1st order derivatives in space.

4.2.4 Solution methods

Numerical treatment of PDE's is a vast issue and the material written about it fills bookshelves. The method traditionally applied most often is finite differencing. The basic idea is to replace all derivatives (both in time and in space) with finite difference approximations. There are numerous ways to derive these approximations and consequently as many numerical schemes. These schemes are quite distinct with respect

to computational effort, accuracy and stability. To pick the best for a specific problem is an art more than dealing with a mathematical problem.

The most important method besides finite differencing is the finite element method. The numerical approach has been developed originally for problems in structural mechanics but the procedure has been applied since then to many other fields. However, finite elements are predominately suitable for solving boundary value problems, which appear seldom in water quality modelling.

A method that is applied quite frequently for solving initial value problems is the method of lines. The procedure is related to the finite difference method but replaces the spatial derivatives only. As the derivative with respect to time remains the result is a (huge) set of ODE's that is solved with one of mentioned ODE solvers.

There are two more popular methods for solving PDE's in the field of water quality modelling but both are slightly outdated and only applied rarely. The first one is the finite volume approach. The idea is to discretize the physical system into small segments and derive mass balance equations for each of those. The result is a set of algebraic equations similar to those derived by finite differencing. The second one is the method of characteristics. The underlying concept of the method of characteristics is to rearrange the set of differential equations in such way that we obtain a curve in the x, t plane where the system has some very specific features. Thus the original equations are simplified along those characteristic curves and it is possible to obtain simple numerical solutions.

4.3 Software tools

In many situations it is not advisable to spend effort to program these numerical methods by yourself. There are already quite a number of computer programs available for identification of systems and simulation of system behaviour. Rather than reinventing the wheel (and risk to spend endless time in debugging source code) such existing programs should be used whenever possible. Computers programs that are suitable for water quality management can be classified into the following three categories:

* Open model structure programs - general purpose simulation software

Programs that have an open model structure (general purpose simulation software) allow the user a large degree of freedom in model definition. That is variables, parameters, transfer functions and differential equations can be chosen more or less freely. The program provides the mathematical routines for solving the resulting set of ODEs (ordinary differential equations) and PDEs (partial differential equations). The main disadvantage of an open model structure is the significant amount of work left for the user and the required detailed knowledge on model formulation. Hence, this type of program is a typical research tool. Typical examples are MATLAB/Simulink, Mathematica and more directed at applications in wastewater engineering AQUASIM, GPS-X and WEST++.

* Closed model structure programs

Most commercial available programs for simulation of special systems have a closed structure. That is, the program concentrates on a given environmental system and implements a specific mathematical model of relevant processes. Apart from specifying

values for the given set of model parameters the user has no choice of freedom. Examples of wastewater treatment simulators are BIOWIN, DAISY, EFOR, SIMBA and STOAT.

* Software libraries

Software libraries are available for many task as e.g. solving differential equations, statistical analysis of results etc. The libraries can speed up the time for developing a program considerably as libraries are (at least in theory) debugged and numerically optimized. However, note that this still requires to write considerable amount of software in a programming language of your choice. The better numerical libraries are traditionally available in Fortran (e.g. NAG) but more advanced object orientated programming languages such as C++ provide more and better features for development and implementatioon of software libraries.

5. Model uncertainty

In Lei (1996) a comprehensive description of uncertainty analysis is given for urban rainfall-runoff modelling.

Uncertainty is an inherent property of modelling. It is not realistic to expect that a model performs perfectly. Modelling along with uncertainty analysis provides us detailed information about why the model is performing poorly. For example:

- The data quality is not good enough to be useful
- The model structure is wrong, thus the model should not be applied in the first place
- The information is not sufficient for calibrating all parameters

It follows from the points above that the uncertainty encapsulated in any model is a combination of (Beck, 1987):

- Uncertainty in the input variables
- Uncertainty in the model structure
- Uncertainty in the model parameters

and in the case of model simulation, there is uncertainty in the initial boundary conditions.

By analysing the model output error, we

1. acknowledge the fact that model output is bounded with an error that must be quantitatively assessed (i.e. uncertainty bounds) in order to be taken into account in decision making.
2. gain information and insight that guide us to avoid pitfalls in the process of system modelling, thus achieving successful model performance.
3. obtain a ranked list for resource allocation plan for reducing models uncertainty (i.e. improving the model performance).

5.1 Uncertainty propagation

This subsection deals with the propagation of errors through a model. As described above, a model is imprecise due to uncertainty in input variables, model structure and parameters. For simulation studies initial values of state variables in a state-space formulation are also contributing to the output uncertainty. Thus, this subsection is concerned with estimating the magnitude of the output uncertainty given the magnitude and distribution of the errors imposed to the model. The most common application of uncertainty propagation is to investigate the contribution of parameter uncertainty to the output. However, the techniques can be applied as well to address the propagation of uncertainty in input variables, initial conditions and structural error.

In assessing the uncertainty propagation to the output there are generally two approaches:

- linearisation techniques

- Monte Carlo simulation

The *linearisation* approach includes two types of uncertainty propagation: first-order analysis and statistical linearisation. The *Monte Carlo* approach involves ordinary Monte Carlo runs and a refinement to this called the Least Squares Linearisation.

5.1.1 Linearisation techniques

Consider the linear relationship for characterising the distribution of nitrogen in the raw wastewater (i.e. it is assumed that N is entering the treatment plant only in the form of ammonia or organic nitrogen):

$$N_{tot} = N_{NH4} + N_{org}$$

If N_{NH4} and N_{org} are stochastic variables with uncertainty $\sigma^2_{NH¤}$ and σ^2_{org}, the variance of the output N_{tot} would be according to the probability laws:

$$\sigma^2_{tot} = \sigma^2_{NH4} + \sigma^2_{org} + 2 \bullet \rho \bullet \sigma_{NH4} \bullet \sigma_{org}$$

where ρ is the crosscorrelation of the errors relation to N_{NH4} and N_{org}. Thus, for a linear model the propagation of uncertainty can be derived analytically.

Unfortunately, most models for activated sludge processes are non-linear. One way of addressing this uncertainty propagation in a non-linear model is to use a *first order Taylor expansion* of the model to linearise around a given point.

Consider the rate of nitrification r_{nit} which has been introduced previously. Investigating the propagation of uncertainty with respect to the four parameters around the expansion point ($Y_{A,0}$, $\mu_{A,0}$, $K_{NH,0}$ and $K_{O,0}$) yields the following expansion.

$$r_{nit}(Y_A,\ m_A,\ K_{NH},\ K_O) \approx r_{nit}(Y_{A,0},\ m_{A,0},\ K_{NH,0},\ K_{O,0}) + \frac{dr_{nit}}{dY_{A,0}}(Y_A - Y_{A,0})$$
$$+ \frac{dr_{nit}}{d\mu_{A,0}}(\mu_A - \mu_{A,0}) + \frac{dr_{nit}}{dK_{NH,0}}(K_{NH} - K_{NH,0}) + \frac{dr_{nit}}{dK_{O,0}}(K_O - K_{O,0})$$

The uncertainty of the output is obtained by applying the probability law for a linear model as shown above (the total variance of r_{nit} consists of 4 individual variances of the parameters and 6 crosscorrelation products).

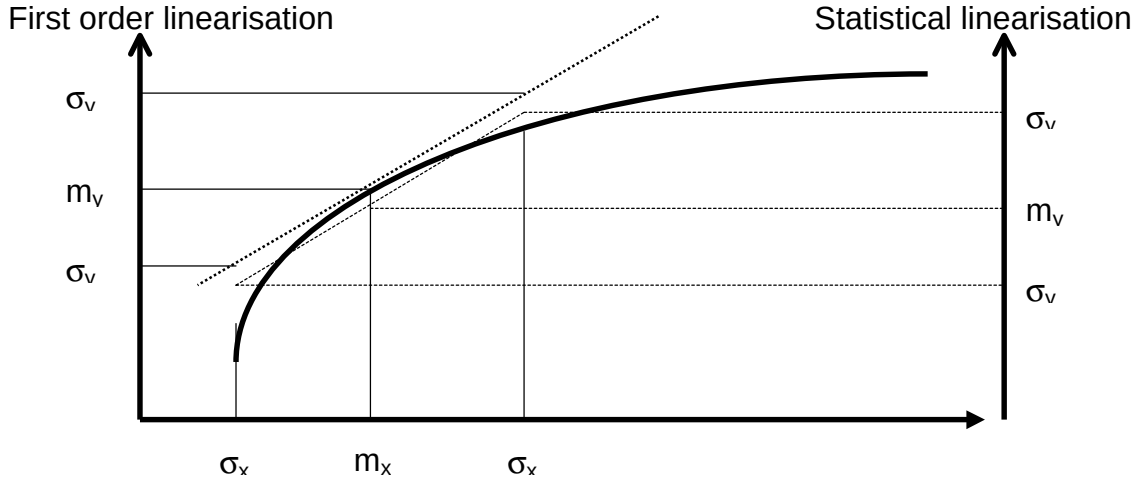
It is crucial to the Taylor expansion approach that the approximation is good and the remainder in the approximation is negligible. The approach is only valid for small deviations (errors) from the expansion point - investigating the propagation of larger errors through the equation will fail due to the non-linearity (i.e. the remainder of the approximation is no longer negligible).

The *statistical linearisation* is similar to the Taylor expansion with the exception that weights instead of derivatives are found based on a optimisation of the squared error. It is assumed that the nitrification rate can be expressed as

$$r_{nit}(Y_A,\ m_A,\ K_{NH},\ K_O) \approx w_1(Y_A - Y_{A,0}) + w_2(\mu_A - \mu_{A,0}) + w_3(K_{NH} - K_{NH,0}) + w_4(K_O - K_{O,0}) + b$$

where b is the mean value of the output in the investigated uncertainty area. The weights w_i and b are found by integration over the uncertainty surface of the investigated parameters. For most models this integration is solved numerically.

The difference between the first order linearisation and the statistical linearisation is illustrated on the figure below.



First order linearisation uses the derivative in examination point m_x, while statistical linearisation uses a regression line which minimises the squared errors. The mean value of the output as well as the uncertainty differs between the two methods. Statistical linearisation is giving the most correct results, but it is also requiring more computation time.

5.1.2 Monte Carlo simulation techniques

Uncertainty propagation can also be examined assuming the uncertainty components to be given by a joint distribution. The stochastic equations are interpreted as an infinite set of deterministic equations, when simulated yields the output distribution. After a sufficient large number of simulations, the empirical distribution function of the output can be determined. Thus, Monte Carlo simulation is a numerical integration technique for deriving the output distribution.

There are two major concerns of using Monte Carlo simulation:

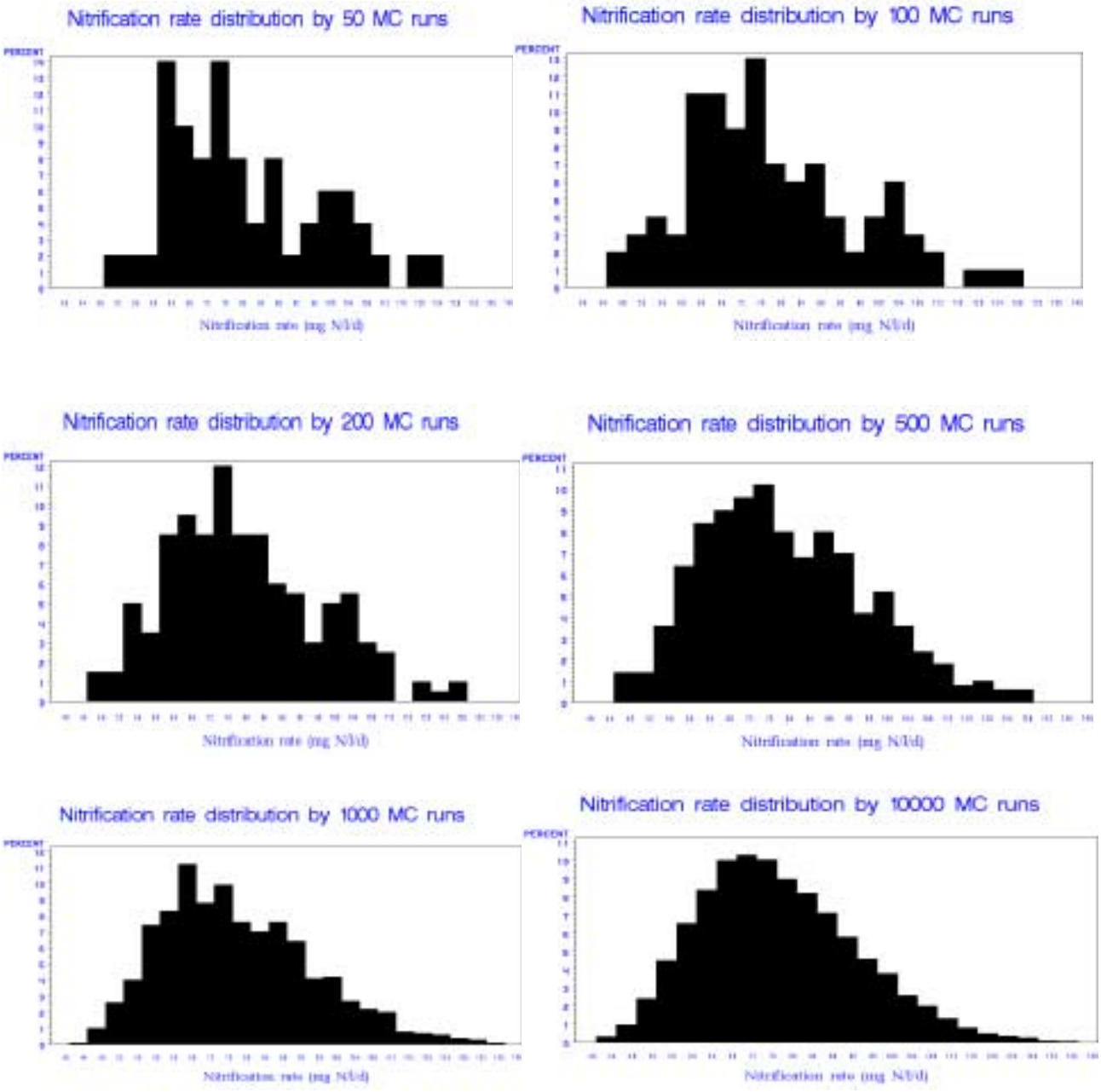
1. a joint distribution function for the uncertainty components is required.
2. a large number of simulation runs is required

Reconsider the uncertainty analysis of the nitrification rate with respect to the four parameters (Y_A, μ_A, K_{NH} and K_O). Assume that these parameters are uniformly and independently distributed in the following intervals:

$$Y_A \in [0.20;0.30] \qquad K_{NH} \in [0.50;1.00]$$

$\mu_A \in [0.40;0.60]$ $K_O \in [0.3;0.7]$

Given a constant biomass X_{BA}=100 g COD/m^3, ammonia concentration S_{NH}=1.0 g N/m^3 and oxygen concentration S_O=1.0 g O_2/m^3, the rate distribution is found by Monte Carlo simulation.



It is clear to see that it requires at least 500 Monte Carlo simulations to get an idea of the distribution, but the true shape only appear after approximately 10000 runs.

Monte Carlo Simulation has the flaw that the individual contributions of each independent variable cannot be determined, i.e. all uncertainty contributions are lumped into the output distribution.

Lei & Schilling (1996) have developed a method called least squares linearisation which is a combination of the statistical linearisation technique and the Monte Carlo approach. A number of Monte Carlo simulations are performed in order to obtain data by which the weights w_i and b in the statistical linearisation can be obtained by multivariate regression. This method has the advantage like statistical linearisation that the individual uncertainty contributions can be traced to the output and like the Monte Carlo approach it is computationally simple.

5.2 Uncertainty analysis

The previous section investigated the propagation of uncertainty through the model. It was implicitly assumed that the magnitude and distribution of the error components are well-known. This is often not the case. Uncertainty bounds for analysing the uncertainty propagation are frequently determined by experience or adequate bounds. Thus, as a result the output uncertainty highly depends on the apriori assumptions. The intervals where a parameter value is likely to be found are normally determined based on literature references of calibrating the model to measured data. On the other hand, structural errors are often neglected and as a consequence, the use of a wrong model structure will lead to increased parameter uncertainty. The uncertainty attributed to input data and initial boundary conditions are likewise neglected.

As described previously, for models that are practically identifiable, the structural and parametric uncertainties can be determined. However, this often requires a simpler model structure than desired or fixation of given parameters. Thus, in order to obtain practical identifiability

- the model should be fairly simple
- the data set for identification should be information rich (persistently excited)

5.2.1 Statistical estimation

Assume that there is a linear relationship between ammonia and total nitrogen in the raw wastewater. This can be formulated as a regression model:

$$S_{N\text{-}tot} = a \bullet S_{NH} + b + e_i \qquad i = 1.....n$$

A regression analysis would yield parameter estimates and uncertainties for the estimates as well as a residual uncertainty. For most models the residual variation is quickly looked over, but it can be regarded as a combination of uncertainty in input data (both for the dependent and independent variables) and structural uncertainty.

Moreover, dynamical models for describing the activated sludge processes formulated in a state-space form can be used to estimate the magnitude of several uncertainties components by application of a Kalman-filter (Kalman, 1960). Examples are given in Jeppsson (1996) and Carstensen (1994). In case the model is not linear, the problem can be addressed using the extended Kalman filter which uses a first-order Taylor approximation for the non-linearities (see e.g. Harvey, 1989).

If the model is practically identifiable, the covariance matrix of the parameters can be found either implicitly ($\sigma^2(X'\Sigma^{-1}X)^{-1}$ for generalised linear models) or explicitly from numerical optimisation (the inverse of the Hessian matrix in the optimisation point). Analyses of residuals and covariance matrix of the parameters may conclude that the model structure is wrong, and consequently a more adequate model has to be found. As an example, strong correlation in the parameter estimates (ill conditioned covariance matrix close to singularity) indicate a wrong model structure or bounds between model parameters.

5.2.2 The GLUE approach

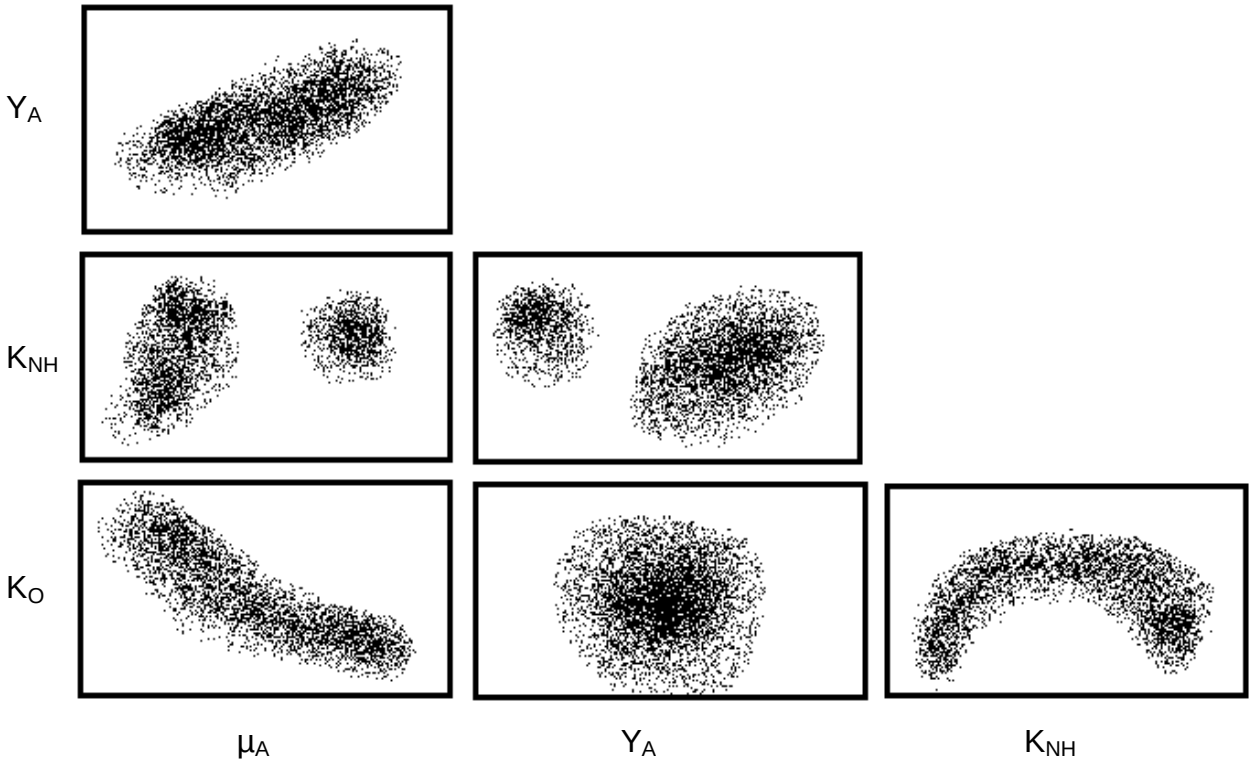
Beven & Binley (1992) have developed a technique called Generalised Likelihood Uncertainty Estimation (GLUE), which is a tool for calibration and uncertainty estimation. The basis of the GLUE approach is that any model with any parameter set combination that predicts output variables reasonably well must be considered equally likely. It is based upon making a large number of runs of a given model with different sets of parameter values, chosen randomly from specified distributions. In general, this technique can be applied to different model structures, thus enabling the assessment of structural errors. Each set of parameter for a given model structure is assigned a goodness-of-fit value of being the 'true' system simulator.

The GLUE approach is divided into the following steps:

1. Defining a goodness-of-fit function for output data. The choice of function can be crucial to the results of the procedure. Further, a criterion based on the goodness-of-fit function for accepting or rejecting a model must be determined.
2. Defining initial ranges or distributions of parameter values to be considered for a particular model structure.
3. the parameter space is sampled to obtain realisations or simulations of the model. It is most common to use Monte Carlo simulation with uniform parameter distributions.

Going through the steps above yields empirical joint distributions for model parameters. Again, this approach can be extrapolated to involve many different model structures.

The scatter plots below exemplifies the approach.


Y_A
K_{NH}
K_O
μ_A
Y_A
K_{NH}

6. Model application

In nearly every aspect of human life models are being used. Weather forecasting, car and aeroplane design, bridge construction, etc., are all examples in our daily life where we have learned that models are useful. The manifold applications of models can be categorised according to the modelling objective, and basically, there are only two:

1. Description of the past
2. Prediction of the future

Models applied in understanding aim at the increase of knowledge on system behavior. The objective is to develop a simple, however universal model of the system under consideration that gives an adequate *description* of reality as it was observed (Reichert, 1994). The use of models for the purpose of understanding is most frequent in research and education.

The *prediction* of either future or hypothetical system behavior is one of the basic tasks in practice. Models applied for prediction aim at providing an accurate and fast image of real systems behavior under different conditions than the ones prevailing during model building. The model use can either aim on forecasting future states of the system (simulation with new inputs) or on predicting system behavior under hypothetical scenarios (simulation with new parameter values). The latter mentioned application is most frequent in design.

For each system under study different appropriate models exist. The chosen appropriate model will depend on its purpose. The decision about the model type must be made before the model building begins. Hence, below an overview is given of different model application areas and some hints are presented on the requirements put forth for each area.

6.1 Criteria for model choice

Some of the criteria to consider while defining a certain application and deciding on the model to use are (Jeppsson, 1996):

1. the purpose of the model,
2. the system boundaries,
3. the time constraints and
4. the desired accuracy.

Evaluating these criteria before the (considerable) efforts of model building are made is a worthwhile undertaking as it may prevent that a considerable number of model adaptations (reducing or increasing the complexity of the model) is necessary.

6.1.1 Purpose of the model

Activated sludge models are typically used in research and development for analysis, design, control and decision support. No single model will be able to fulfil all types of applications. For example, simple models that may be suitable within model-based control algorithms , may be totally inadequate for simulating and predicting the entire process behaviour for safety analysis.

6.1.2 System boundaries

Properly choosing the system boundaries should ensure that all the important dynamics of the process are included in the model. A choice of the boundaries which includes too many insignificant details will lead to an unduly large model. This may cloud the understanding of how and why the system dynamics are occurring as well as lead to an unacceptable computational burden. Conversely, failing to include significant features of the real process could lead to inaccurate dynamic responses and a loss of confidence in the model. If it is unclear how to choose the boundaries, a criterion for boundary selection consists of evaluating whether the streams crossing the proposed boundary are easy to characterise. If this is the case, the system boundaries can be relied upon.

6.1.3 Time constraints

Activated sludge processes are characterised by a very wide range of dynamic activity with widely varying speeds of response to changes in process conditions. Characteristic time constants in the process may range over many orders of magnitude, e.g. dissolved oxygen concentrations change within tens of seconds, while the nitrifying population changes over periods of weeks.

Invariably, the modeller is only interested in a simulation over a defined period of time. To produce an appropriate model, he should therefore identify a ‘time-scale-of-interest’ and not model any latent dynamic effects outside this time-scale window. This window is described by a maximum and a minimum characteristic time constant. Selection of an appropriate time-scale window will also have the added advantage of possibly avoiding problems in the numerical solution of the model that are characteristic of so-called ‘stiff’ systems with a wide range of time constants.

When a time constant window of interest is chosen, structured procedures have been developed to adequately reduce a complex model to a model that only focuses on the dynamics with the considered time constants (Steffens & Lant, 1997). Basically, one freezes the slowly varying variables to the initial values, while the rapidly varying variables are considered to take new values instantaneously.

6.1.4 Desired accuracy

Depending on the model application accuracy levels will be set. Such accuracy level can be one number, but may also consist of a range of values that are to be considered for different conditions. For instance, in a certain operating range, a high accuracy is necessary, while under other conditions a lower accuracy is allowed. Obviously, the

accuracy sought for a model will affect the degree of simplification that is allowed while building the model. Again, it will be necessary to set the accuracy before model building is initiated and it should reflect the purpose of the model. During model validation, i.e. before the model is used for its intended purpose, the achieved accuracy should be evaluated.

6.2 Models for understanding

Models serve as a tool to develop and test hypotheses with the aim of gaining new knowledge about the system. Or as Stigter and Beck (1994) state in relation to their work on selecting models for increased understanding:

> *"a model is a more or less complex assembly of constituent hypotheses and the purpose of system identification is therefore that of reconciling this set of concepts with the given field data."*

Mathematical models are an excellent method of conceptualising knowledge about a process and to communicate it to other people. It can be stated that models are the ultimate and crisp summary of knowledge. Models are also useful to formulate hypotheses and for incorporating new ideas that can be refuted through validation exercises with new data that may be collected with this purpose in mind (see next section).

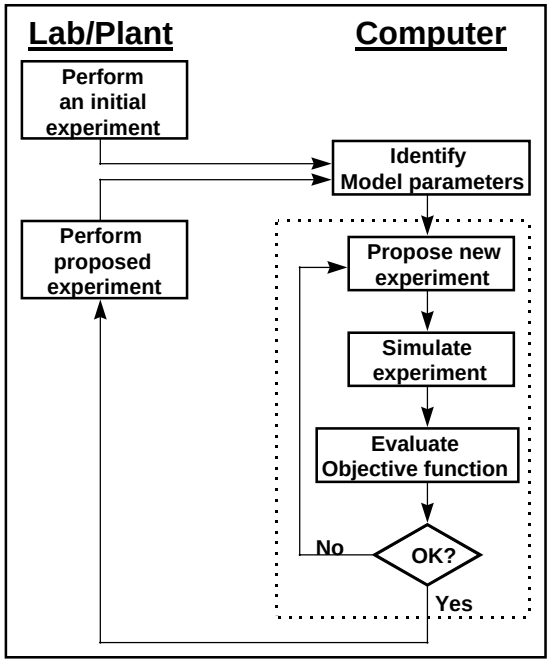
A very clear example of model use for increased understanding is the work performed on biological phosphate removal. In several iterations, a metabolic model could be constructed that allowed to describe the anaerobic and aerobic reactions of phosphate accumulating organisms (PAO) (Smolders *et al.*, 1995). In the mean time, the validity of this model was confirmed further by its applicability to denitrifying PAO's as only one parameter had to be changed in a predictable way to describe data collected under denitrifying conditions (Murnleitner *et al.*, 1997).

6.3 Model use in optimal experimental design

One of the main tasks in model building is the model identification. As explained above, identification of an activated sludge model (calibration) has to rely a lot on data because the *a priori* knowledge is rather restricted. Hence, due to the increased use of models, in the last few years considerable efforts have been made in finding better experimental techniques to

1. select among different model structures
2. obtain parameter values of the model
3. validate the identified model

Mostly, so far, the experiments were designed using intuition and empirical evaluation of the results obtained (Vanrolleghem *et al.*, 1998). However, techniques exist by which rational design of experiments can be performed with the aim of maximising the reliability of model selection, parameter estimation and model validation.



The basic idea of model-based optimal experimental design for parameter estimation is illustrated in the figure below (Vanrolleghem & Dochain, 1998). Given a first prototype model, e.g. only based on *a priori* knowledge and default parameter values, this model can be used to generate data (via simulation) for given experimental conditions. Such simulation predicts how data collected in a real experiment are expected to be. Supposing that these simulated data are the result of a real experiment one can apply a parameter estimation tool and for instance find out how reliable the parameter estimates would be if such data would be collected from a real experiment. The aim of the experimental design procedure is to find those experimental conditions that maximise, for instance, the reliability of the parameter estimates (or in other words, minimise its variance). In case of experimental design for parameter estimation use is often made of the Fisher Information Matrix that summarises the information content of an experiment and can easily be used to evaluated parameter estimate reliability. The search for the optimal experiment is performed via iterations with the model for an extended set of potentially applicable experimental conditions. The best experiment found can then be implemented in the lab or on the plant and the real data collected should allow an optimal estimation of the parameter values.

The aim of the experimental design must be specified in an objective function that can be calculated using the simulated data. This objective function can contain different factors that can be weighted in the overall function, e.g. costs of experiments, reliability of parameter estimates, length of experiments, required equipment, number of analyses needed, etc. In case model selection is the aim of the experiments (e.g. to choose between one or the other hypothesis on process behaviour), one will try to find those experimental conditions that maximise the difference between the responses of both models (Vanrolleghem & Van Daele, 1994). Even combined experimental design methods have been proposed in which an optimal compromise is sought for experiments that are good for both model selection and parameter estimation (Hill *et al.*, 1968). Generally speaking, however, the application of these model-based experimental design techniques is still in its infancy, but it certainly warrants further research and, foremost, application.

6.4 Models in static design of systems

This is probably the oldest application of models in wastewater treatment. The design rules (Vandevenne & Eckenfelder, 1980; ATV, 1990) that are typically used for setting up a treatment facility are simple but reliable models. They are summaries of experience and conceptual thinking and have proven their usefulness. Depending on the “school” to which people belong, different static models are being used and modifications are applied as knowledge on the processes increases (Wentzel *et al.*, 1990).

Limitations to this use of models is that due to their steady state approach considerable simplifications are made. However, this is accounted for by the safety margins that are included in the design models. On the other hand, the adaptation of these design models in case control systems or advanced operating modes are adopted, is difficult and can probably be seen as one of the factors that has and still is hampering their introduction.

6.5 Models in dynamic design (control engineering)

The development of control strategies is very much supported by models. In *model-based control design*, models allow to investigate the response of the controlled system to a wide range of inputs (disturbances, equipment failures, changing operating points) without endangering the actual plant. In this application the model is used to evaluate the performance of a new controller under such inputs.

A second use of models in control engineering consists of *model-based control*. In these applications the model is integrated in the control law and is therefore used in real-time as the process runs. Examples are such control engineering concepts as model-based predictive control, adaptive linearising control, state feedback control, etc. (Van Impe *et al.*, 1998). Models can also be present in a control loop via the software sensors that upgrade information from raw sensor data by combining it with the concepts summarised in models (Dochain & Perrier, 1998). Finally, models can also be very useful in error detection and diagnosis, allowing to increase the reliability of the data used in the control law (Bergh & Olsson, 1996; Hellinga *et al.,* 1998).

In model-based control design time constraints are not very severe and quite complex mechanistic models can be applied. Extrapolation power of the applied model should be reasonable because the evaluation of controller performance should be realistic even under conditions quite different from the normal operating point.

The models applied within model-based controllers, software sensors and error diagnosis systems, however, are severely affected by the computational demands they impose because these models have to be running in real-time with the process. Their extrapolation power, however, is less important as the control system should exactly ensure that the system remains around within certain bounds.

6.6 Models for decision support

A model used as a decision support tool is aiming at identifying constraints and objectives of a problem in the initial phase (purpose of understanding) and later on choosing the most suitable solution from a set of alternative techniques (by predicting the effect of those techniques). Quite clearly, interaction with the person taking the decisions is a necessity and attention is to be given that the model can provide outputs that are understandable by this user. (in control applications, this is not always a necessity !).

A typical decision support tool is a simulator with a dynamic activated sludge model available at a treatment plant where an operator can

1. evaluate different actions to deal with a certain problem,

2. obtain some background information on the causes of the problem he is facing by an expert system that is incorporated in the software tool and uses the simulation results for diagnosis (Kayser, 1990)
3. receive advice on potential actions that have been beneficial in similar situations in the past

Typically for such decision support tool, calculation speed is not critical, but should of course not be too excessive to allow for reasonable interaction with the user. Mechanistic models will typically be used to accommodate for the need for understandability of the outputs.

7. List of references

Akaike, H. (1974). A New Look at the Statistical Model Identification, *IEEE Transactions on Automatic Control,* 19(6), pp. 716-723.

ATV (1991). *ATV - A131E: Dimensioning of Single Stage Activated Sludge Plants upwards from 5000 Total Inhabitants and Population Equivalents,* Abwassertechnische Vereinigung e.V., Gesellschaft zur Förderung der Abwassertechnik e.V., St. Augustin, Germany. pp. 63.

Beck, M.B. (1987). Water Quality Modeling: A Review of the Analysis of Uncertainty, *Water Resources Research*, 23(8), pp. 1393-1442.

Bergh S.-G. and Olsson G. (1996) Knowledge Based Diagnosis of Solids-liquid Separation Problems, *Water Science & Technology,* 33(2), pp. 219-226.

Boyle, W.C. and Berthouex, P.M. (1974). Biological Wastewater Treatment Model Building: Fits and Misfits, *Biotechnology & Bioengineering*, 16, pp. 1139-1159.

Carstensen, J. (1994). *Identification of Wastewater Processes,* Ph.D. thesis, Technical University of Denmark, Denmark.

Casti J.L. (1977). *Dynamical Systems and their Application - Linear Theory,* Academic Press, New York.

Caswell, H. (1976). The Validation Problem, in: Patten, B.C., ed., *Systems Analysis and Simulation in Ecology*, Volume IV, Academic Press, New York, pp. 313-325.

Chui C.K. and Chen G. (1988). *Linear Systems and Optimal Control*. Springer-Verlag, Berlin.

Dochain D. and Perrier M. (1998). Monitoring and Adaptive Control of Bioprocesses, in: *Advanced Instrumentation, Data Interpretation and Control of Biotechnological Processes*. Eds. Van Impe J., Vanrolleghem P. & Iserentant D., Kluwer Academic Publishers, Dordrecht, The Netherlands. 347-400.

Dochain, D., Vanrolleghem, P.A. and Van Daele, M. (1995). Structural Identifiability of Biokinetic Models of Activated Sludge Respiration, *Water Research*, 29, pp. 2571-2579.

Draper, D. (1995). Assessment and Propagation of Model Uncertainty (with discussion). *Journal of Royal Statistical Society B,* 57(1), pp. 45-97.

Godfrey, K.R. and DiStefano, J.J. (1987). Identifiability of Model Parameters, Chapter 1 of Walters, E., ed., *Identifiability of Parametric Models*, Pergamon Press, Oxford, pp. 1-20.

Harvey, A. (1989). *Forecasting, Structural Time Series Models and the Kalman Filter*, Cambridge University Press, 554 pp.

Hellinga C., Romein B., Luyben K.Ch.A.M. and Heijnen J.J. (1998). Error Diagnosis and Data Reconciliation using Linear Conservation Relations, in: *Advanced Instrumentation, Data Interpretation and Control of Biotechnological Processes*. Eds. Van Impe J., Vanrolleghem P. & Iserentant D., Kluwer Academic Publishers, Dordrecht, The Netherlands, pp. 211-244.

Hill W.J., Hunter W.G. and Wichern D.W. (1968). A Joint Design Criterion for the Dual Problem of Model Discrimination and Parameter Estimation, *Technometrics*, 10, pp. 145-160.

Jeppsson, U. and Olsson, G. (1993). Reduced Order Models for On-line Parameter Identification of the Activated Sludge Process. *Water Science & Technology*, 28(11-12), pp. 173-183.

Jeppsson, U. (1996). *Modelling Aspects of Wastewater Treatment Processes*, Ph.D. thesis, Lund Institute of Technology, Sweden.

Kalman, R. (1960). A New Approach to Linear Filtering and Predition Problems, *Transactions of ASME, Series D, Journal of Basic Engineering,* 82, pp. 35-45.

Kayser R. (1990). Process Control and Expert Systems for Advanced Wastewater Treatment Plants, in: *Advances in Water Pollution Control*. Ed. Briggs R., Pergamon Press, London, pp. 203-210.

Lei, J.H. and Schilling, W. (1996). Preliminary Uncertainty Analysis - A Prerequisite for Assessing the Predictive Uncertainty of Hydrological Models, *Water Science & Technology,* 33(2), pp. 79-90.

Lei, J.H. (1996). *Uncertainty Analysis of Urban Rainfall-Runoff Modelling*, Ph.D. thesis, Norwegian University of Science and Technology, Norway.

Ljung, L. (1987). *System Identification - Theory for the User*, Prentice-Hall, Englewood Cliffs, New Jersey, 519 pp.

Murnleitner E., Kuba T., van Loosdrecht M.C.M. and Heijnen J.J. (1997). An Integrated Metabolic Model for the Aerobic and Denitrifying Biological Phosphorus Removal, *Biotechnology & Bioengineering*, 54, pp. 434-450.

Norton, J.P. (1986). *An Introduction to Identification*, Academic Press, London.

Popper, K.R. (1980). *The Logic of Scientific Discovery*, Hutchinson, London, 480 pp.

Press W.H., Flannery B.P., Teukolsky S.A. and Vetterling W.T. (1992): Numerical Recipies in C. 2nd Ed. Cambridge University Press.

Reckhow, K.H. and Chapra, S.C. (1983). Confirmation of Water Quality Models, *Ecological Modelling*, 20, pp. 113-133.

Reichert, P. and Omlin, M. (1996). On the Usefulness of Overparameterized Ecological Models, *Ecological Modelling* (in press).

Schwarz, G. (1978). Estimating the Dimension of a Model, *The Annals of Statistics*, 6(2), pp. 461-464.

Smolders G.J.F., van der Meij J., van Loosdrecht M.C.M. and Heijnen J.J. (1995). A Structured Metabolic Model for Anaerobic and Aerobic Stoichiometry and Kinetics of the Biological Phosphorus Removal Process. *Biotechnology & Bioengineering*, 47, pp. 277-287.

Söderström, T. and Stoica, P. (1989). *System Identification*, Prentice Hall, New York.

Spriet, J.A. (1985). Structure Characterization - An Overview, in: Barker, H.A. and Young, P.C., eds., *Identification and System Parameter Estimation 1985*, Proceedings of the 7th IFAC/IFORS Symposium, Volume 1, pp. 749-756.

Steffens M. and Lant P. (1997). Reducing the Complexity of Biological Wastewater Treatment Models, in: *Proceedings 7th IAWQ Workshop on ICA of Water and Wastewater Treatment and Transportation Systems*. Brighton, UK, July 6-9 1997, pp. 343-350.

Stigter J.D. and Beck M.B. (1994). A New Approach to the Identification of Model Structure, *Environmetrics*, 5, pp. 315-333.

Tong, H. (1990). *Non-linear Time Series - A Dynamical System Approach*, Clarendon Press, Oxford, 564 pp.

Van Impe J.F., Vanrolleghem P.A. and Iserentant D. (1998) *Advanced Instrumentation, Data Interpretation and Control of Biotechnological Processes*. Kluwer Academic Publishers, Dordrecht, The Netherlands. ISBN 0-7923-4860-5. pp. 464.

Vandevenne L. and Eckenfelder W.W.Jr. (1980). A Comparison of Models for Completely Mixed Activated Sludge Treatment Design and Operation, *Water Research*, 14, pp. 561-566.

Vanrolleghem P.A. and Dochain D. (1998). Bioprocess Model Identification, in: *Advanced Instrumentation, Data Interpretation and Control of Biotechnological Processes*. Eds. Van Impe J., Vanrolleghem P. & Iserentant D., Kluwer Academic Publishers, Dordrecht, The Netherlands, pp. 251-318

Vanrolleghem P.A. and Van Daele M. (1994). Optimal Experimental Design for Structure Characterization of Biodegradation Models: On-line Implementation in a Respirographic Biosensor. *Water Science & Technology,* 30(4), 243-253.

Vanrolleghem P.A., Spanjers H., Petersen B., Ginestet P. and Takacs I. (1998). Estimating (Combinations of) Activated Sludge Model No.1 Parameters and Components by Respirometry, in: *Proceedings 4th IAWQ Specialised Seminar on Modelling and Microbiology of Activated Sludge Processes*. Copenhagen, Denmark, March 16-18 1998, pp. 18.

Vanrolleghem, P.A., Van Daele, M., and Dochain, D. (1995). Practical Identifiability of a Biokinetic Model of Activated Sludge Respiration, *Water Research*, 29, pp. 2561-2570.

Wentzel M.C., Ekama G.A., Dold P.L. and Marais G.V.R. (1990). Biological Excess Phosphorus Removal - Steady State Process Design. *Water SA*, 16, pp. 29-48.

Annex

Carstensen, J., Vanrolleghem, P., Rauch, W. and Reichert, P. (1997). “Terminology and Methodology in Modelling for Water Quality Management - A Discussion Starter”, *Water Science & Technology, 36(5), pp. 157-168.*