EVOLUTIONARY OPTIMISERS IN CALIBRATION OF ACTIVATED SLUDGE MODELS

Jukka Keskitalo, Kauko Leiviskä

Control Engineering Laboratory
P.O Box 4300, FIN-90014, University of Oulu

jukka.keskitalo@oulu.fi (Jukka Keskitalo)

Abstract

In most applications of the Activated Sludge Models (ASM), calibration is based on more or less ad-hoc and trial and error approaches. There is a need for more systematic and automated approaches for ASM calibration. In this study, the applicability of real-coded genetic algorithms and differential evolution in calibration of a modified ASM No. 1 was evaluated. The evolutionary optimisers were used in parameter identification of a subset of the model parameters. The results were compared with a previously proposed calibration approach based on Monte Carlo simulations. All methods were capable of calibrating the model when given enough computation time. However, some of the evolutionary optimisation methods had a clear advantage in terms of computation time against the Monte Carlo method. The calibration was tested on an independent data set. The calibrated model provided accurate predictions on the testing data. However, it was found that the best parameter set for the calibration data did not result in the best performance for the testing data. Therefore the calibration and testing data sets have somewhat different optimal parameter values. As the optimal parameter values cannot be expected to remain constant for extended periods of time, the model will need recalibration in practical applications. This result emphasises the importance of methods which can automate the model calibration.

Keywords: pulp and paper wastewater, modified activated sludge model no. 1 (ASM1), parameter identifiability, model calibration, full-scale WWTP.

Presenting Author's biography

Jukka J. Keskitalo was born in Oulu, Finland in 1984. He received the M.Sc. degree in process engineering with honours from the University of Oulu, Finland in 2008. Currently he is a PhD student at the Control Engineering Laboratory, Department of Process and Environmental Engineering, University of Oulu and at the Graduate School in Chemical Engineering (GSCE). His current research interest is modelling and control of biological wastewater treatment. The research covers first principles and data driven approaches to biological wastewater treatment modelling, and combination of the two, the so called hybrid modelling.

1 Introduction

Modelling activated sludge systems has become an accepted practice in wastewater treatment plant (WWTP) design, teaching and research. Examples of model applications include education of process operators, evaluation of process alternatives in design phase and process optimisation [1]. Activated Sludge Models (ASM) published by the IWA Task Group on Mathematical Modelling for Design and Operation of Biological Wastewater Treatment [2] are by far the most widely used models for activated sludge systems. The ASMs have been implemented in a number of commercial WWTP simulators as well as general purpose simulation environments [1].

Despite the popularity of the ASMs, questions regarding their calibration remain to be answered. The fact that parameters of the ASMs are not universal, and that influent wastewater has to be characterised in terms of the model state variables, was acknowledged already in the original IWA Task Group reports [2]. The Task Group provided some guidelines for calibrating the models but no comprehensive calibration protocol was given. Experimental procedures for estimating model parameters and wastewater characterisation were developed by several researchers to fill the gaps in calibration methodology. Review on these methodologies can be found in [3]. Systematic calibration protocols were developed later to give a description of complete calibration procedure. Sin et al. have reviewed some of these protocols [4]. Despite all the effort and numerous publications over the years on calibrating the ASMs, none of the proposed calibration protocols has established the status as the standard calibration protocol. Review of literature on applications of the ASMs revealed that in practice calibration is often based on more or less ad-hoc approaches [5].

It is well-known that the parameters of large scale environmental models such as the ASMs are poorly identifiable [6]. For example, it was stated already in the IWA Task Group report on ASM1 that errors in the estimate of a certain parameter will be compensated when estimating values of other parameters [2]. The activated sludge models are practically unidentifiable for two reasons: the model structure does not allow identification of unique values for all parameters, and the available data is insufficient in quality and quantity [7]. Fortunately, only a subset of parameters has to be calibrated for a certain application. Determining suitable combinations of identifiable parameters has been mainly based on expert knowledge. As the combinations are not guaranteed to be uniquely identifiable, parameters have been adjusted manually one by one.

As alternatives to the experience based approach, systematic approaches have been proposed to address the problem of parameter identifiability in large scale non-linear models. These approaches systematically evaluate subsets of parameters to determine their identifiability given the available data and model structure. The most often cited and applied identifiability analysis methodologies were developed by Weijers and Vanrolleghem [8] and Brun et al. [6]. There have also been more recent developments in this field [9-10]. Methods described in [8], [6] and [9] have been successfully applied on the ASMs. The method of Weijers and Vanrolleghem [8] is based on the Fischer Information Matrix (FIM). Properties of the FIM are used to find identifiable subsets of parameters. The method of Brun et al. [6] is derived from linear regression diagnostics. Both methods make use of local sensitivity analysis.

Identifiability analysis methods partly automate the calibration process by systematically screening for identifiable parameter subsets instead of relying on expert judgment. Achieving a higher level of automation in the calibration procedure is desirable in order to avoid the problems associated with the commonly used trial and error approaches: the laboratory scale experiments to estimate parameters are time consuming and expensive and there is no established standard protocol for conducting the experiments. Moreover, variable operating conditions in some industrial treatment plants lead to the need of updating calibration. For example, by analysing process data from pulp mill WWTP it was found that dependencies between process variables differ significantly in different data clusters, which represent operating conditions extracted from the process data [11]. Frequent recalibration involving extensive measurement campaigns would make the model application too arduous. On the other hand, with feasible automatic calibration methods the recalibration would not become an issue.

Naturally, the next step in automating calibration procedure is to use numerical optimisation methods to calibrate the identifiable parameter subsets instead of manual parameter adjustment. So far, identifiability analysis has not been much used in combination with global optimisation methods. Monte Carlo based calibration approach, where parameter space was sampled with Latin hypercube sampling (LHS), was applied to calibrate a subset of ASM2d parameters in [7]. Sin et al. justify their use of Monte Carlo approach over other optimisation methods by citing a claim that global search methods such as genetic algorithms would be unusable due to their requirement of an enormous number of simulations [7].

Genetic algorithms (GA) and differential evolution (DE) have been previously used successfully in parameter identification of nonlinear bioprocess

models. In [12], GA was used to estimate parameter values of a model for E. Coli fed-batch cultivation process. Real-coded GA was used in [13] for parameter identification of chemostat bioprocess model. A modified DE was developed and applied to estimate parameters of a Monod-type batch fermentation model [14]. However, these successful applications using data from simulators or from laboratory scale experiments are no absolute proof that the methods are equally feasible in modelling full scale biological wastewater treatment processes.

This paper studies the applicability of genetic algorithms and differential evolution in calibration of a mechanistic model based on the ASM1 to simulate the biological treatment of pulp mill effluent by an activated sludge treatment plant. Identifiable subset of parameters is chosen by following the methodology of Brun et al. [6]. Performance of the evolutionary optimisation methods in calibration of the Activated Sludge Model is compared with the Monte Carlo based calibration approach proposed in [7].

2 Materials & methods

2.1 Process and data

The WWTP under study is fully aerobic activated sludge plant consisting of primary sedimentation, aeration and secondary sedimentation. The WWTP treats an average of 32 000 m^3 of wastewater per day from a pulp mill with an annual production of 340 000 tonne of fully bleached kraft pulp. Process data including online measurements and results from laboratory analyses was extracted from the mill databases. The data covers the period from 1 September 2009 to 31 January 2010.

Flow rates of influent wastewater, return sludge and wasted sludge, and total chemical oxygen demand (COD), total nitrogen (N) and total phosphorus (P) concentrations are included in the process data of the WWTP influent. The effluent process data consists of total COD, total N, total P, soluble N and soluble P concentrations. Flow rates in the data are daily average values. Total COD concentrations are analysed from 24 hour composite samples. N and P concentrations are analysed from seven day composite samples.

The process data has been split into calibration and testing data sets. First 77 days of the data were used as training data and the remaining 76 days as validation data. The calibration data set is used to identify the model parameters. The testing data set is used to check if the calibration results can be generalised.

2.2 Model structure

The applied model is modified from the original ASM1 to be more feasible in modelling the biological treatment of nutrient deficient wastewaters. The modifications are motivated by the fact that the original ASM1 was developed to describe simultaneous removal of excess organic nitrogen and carbon. On the other hand, nutrient content in pulp mill wastewaters is typically very low, and addition of nutrients is required to get biological treatment without nutrient limitation problems. The model was introduced in [15] and simplified in [16]. Description of the model given in [16] is used in this work.

When modelling biological wastewater treatment with the ASMs, reactors are modelled as continuous stirred tank reactors (CSTR). In this work the aeration basin was modelled as one CSTR having the complete volume of the aeration basin. Component balances for the liquid phase over each CSTR are written for every state variable of the model as

$$\frac{dC_i}{dt} = \frac{Q}{V}\left(C_{i,in} - C_i\right) + R \quad (1)$$

where C_i is the concentration of the ith component in mg l^{-1}, Q is the flow rate in m^3 d^{-1}, V is the reactor volume in m^3 and R_i is the reaction rate of the ith component in mg $(l\ d)^{-1}$. Reaction rates R_i can be formulated from the model matrix in Appendix 1. with

$$R_i = \sum_{j=1}^{n} y_{j,i} \cdot r_j \quad (2)$$

where $y_{j,i}$ is the stoichiometric coefficient of the ith component for the jth process, r_j is the rate equation for the jth process and n is the number of processes. Parameters are listed in Tab. 1 with their initial values. Initial parameters values come from [2] and [15]. As in [6], each parameter is assigned to an uncertainty class from 1 to 3. Class 1 represents accurately known parameters with 5% uncertainty, class 2 moderately known parameters with 20% uncertainty and class 3 poorly known parameters with 50% uncertainty.

Tab. 1. Stoichiometric and kinetic parameters.

Parameter	Unit	Initial value	Uncertainty class
Stoichiometric parameters			
Y_H	g COD (g COD)$^{-1}$	0.67	2
f_P	-	0.08	2
i_{XB}	g N (g COD)$^{-1}$	0.05	2
i_{XP}	g N (g COD)$^{-1}$	0.035	3
$i_{XBP,1}$	g P (g COD)$^{-1}$	0.005	3
$i_{XBP,2}$	g P (g COD)$^{-1}$	0.005	3
i_{XPP}	g P (g COD)$^{-1}$	0.004	3
Kinetic parameters			
$\hat{\mu}_H$	d^{-1}	12	3
b_H	d^{-1}	0.744	2
K_{OH}	mg O^2 l^{-1}	0.2	3
K_S	mg COD l^{-1}	20	3
K_{Pi}	mg P l^{-1}	0.2	3
K_{P1}	mg P l^{-1}	0.01	3
K_{P2}	mg P l^{-1}	0.5	3
K_{NH}	mg N l^{-1}	0.1	3
k_a	m^3 (g COD d)$^{-1}$	0.16	3
k_h	g COD (g COD d)$^{-1}$	9.0	3
K_X	g COD (g COD)$^{-1}$	0.09	3

The modified ASM1 describing biological part of the WWTP is coupled with a settler model to get predictions on quality of WWTP effluent that is discharged to sea. The settler model is necessary also for modelling the composition of return sludge. Secondary settler was modelled as described in the report on Benchmark Simulation Model no. 1 (BSM1) by the IWA Task Group on Benchmarking of Control Strategies for WWTPs [17]. The settler volume was divided into four layers instead of the original ten to reduce computation time. Primary settler was not included in the model, as the influent wastewater measurements used in calibration were made from the primary settler effluent.

2.3 Identifiability analysis

Analysis of practical identifiability of the applied model structure taking into account the available data is made by following the methodology described in [6]. In order to be identifiable, a parameter subset K with a number of k parameters has to be influential on the model outputs, i.e. the model outputs are sensitive to parameters in the subset K, and changes in certain parameters of the subset K may not cancel out the effect on model output by changes in other parameters.

In the identifiability methodology of Brun et al. [6], sensitivity of the model outputs to changes in parameters is calculated first. The method is a local sensitivity analysis, where small changes are made to parameters at a specific location in parameter space. Dimensionless scaled sensitivity functions are

$$s_{i,j}(t,\theta_0) = \frac{\Delta\theta_j}{sc_i}\frac{\partial y_i(t,\theta)}{\partial\theta_j} \quad (3)$$

where $\partial_{yi}/\partial\theta_j$ denotes the derivative of a model variable y_i with respect to the parameter θ_j evaluated at a specific point θ_{0j} in the parameter space, $\Delta\theta_j$ is an estimate of the reasonable range of θ_j and sc_i is a scale factor representing typical magnitudes of the output y_i. The a priori measures of reasonable ranges for parameters are calculated from the uncertainty classes in Tabs. 1 and 2. All parameters in this work were assigned to classes 2 and 3, as the applied model is rather experimental compared to the well established standard ASMs.

The sensitivity function in Eq. (3) was approximated with the finite difference method

$$\frac{\partial y_i}{\partial\theta_j} \approx \frac{y_i(t,\theta_j+\xi\theta_j)-y_i(t,\theta_j)}{\xi\theta_j} \quad (4)$$

where ξ is a perturbation factor. Sensitivity measures

$$\delta_j^{msqr} = \sqrt{\frac{1}{n}\sum_{i=1}^{n} s_j^2} \quad (5)$$

are calculated to quantify the influence of individual parameters on the model outputs [6]. Vector s_j is constructed by concatenating the sensitivity function vectors of each output variable i for the parameter j and n is the total number of measurements.

Collinearity index of parameter subset K is defined as

$$\gamma_K = \frac{1}{\sqrt{\tilde{\lambda}}} \quad (6)$$

where $\tilde{\lambda}$ is the smallest eigenvalue of matrix $\tilde{S}_K^T\tilde{S}_K$ [6]. $\tilde{S}_K$ is the normalised matrix of sensitivity functions, where columns are the vectors s_j belonging to the subset K. Purpose of the collinearity index is to reveal interdependencies of parameters in the subset K by quantifying the degree of collinearity between columns of matrix $\tilde{S}_K$. Determinant measure ρ_K is useful in identifiability analysis as it considers both parameter sensitivity and collinearity. Low collinearity index γ_K and high determinant measure ρ_K values are indicators of good identifiability. The determinant measure is defined as [6]

$$\rho_K = \det\left(S_K^T S_K\right)^{1/2k}. \quad (7)$$

Kinetic and stoichiometric parameters of the modified ASM1 were included in the identifiability analysis. Parameters of the settler model were not included in the identifiability analysis and calibration procedure, and they were therefore kept at their default values. Wastewater characterisation of total COD and nutrient concentrations to COD and nutrient components of the applied model is also a part of the model calibration procedure. Wastewater characterisation was included in the identifiability analysis and calibration procedure as influent wastewater characteristics of each WWTP are unique and using default values would not be reasonable.

Tab. 2. Influent wastewater characterisation used in identifiability analysis and calibration.

Parameter	Initial value	Uncertainty class
Biodegradable COD (α_{SS}+ α_{XS})	0.502	3
Ratio of S_S to biodegradable COD (α_{SS} (α_{SS}+ α_{XS})$^{-1}$)	0.468	3
Ratio of S_I to nonbiodegradable COD (α_{SI} (1-(α_{SS}+ α_{XS}))$^{-1}$)	0.843	3
Ratio of soluble N to total N	0.311	2
Ratio of soluble P to total P	0.473	2
Ratio of S_{NI} to soluble N	0.078	3
Ratio of S_{NH} to soluble N	0.027	2
N/COD in inert particulate matter (i_{XI})	0.015	3
P/COD in inert particulate matter (i_{XIP})	0.005	3

Wastewater characteristics in the model are given as ratios of model component concentrations to total COD, N and P concentrations. The sum of model component fractions for COD, N and P is one. The model component fractions cannot be used directly in identifiability analysis or calibration, as change in one fraction would have to be compensated in the other fractions to retain the sum of one. Therefore the

wastewater characterisation was defined in a new set of components which can be altered without the need to compensate the changes. The defined components are converted back into model components before running a simulation. The new set of components is given in Tab. 2. Initial values of wastewater characterisation come from [16].

2.4 Modelling objective

Objective of the model calibration is to fit the modelled concentrations of pollutants to the measured values. From the WWTP operator's point of view, the most important components in the effluent are the chemical oxygen demand, nitrogen and phosphorus. Discharges of these three are monitored by the authorities, and discharge limits are expected to become more stringent in the future.

An objective function that describes the problem well is necessary in order for the optimisation to succeed. The problem at hand has multiple objectives, as the residuals between the modelled and measured total COD, N and P have to be minimised simultaneously. The weighted sum of squared errors objective function is calculated as follows:

$$J(\theta)=\sum_{i=1}^{n}\left(\frac{Y_{i,COD}-\hat{Y}_{i,COD}(\theta)}{\sigma(Y_{COD})}\right)^2+\sum_{i=1}^{n}\left(\frac{Y_{i,N}-\hat{Y}_{i,N}(\theta)}{\sigma(Y_{N})}\right)^2 \quad (8)$$
$$+\sum_{i=1}^{n}\left(\frac{Y_{i,P}-\hat{Y}_{i,P}(\theta)}{\sigma(Y_{P})}\right)^2$$

where θ is the evaluated parameter subset, Y_i is the ith measurement, $\hat{Y}_i$ is the model output at the ith time instance and n is the number of measurements.

2.5 Optimisation methods

Genetic algorithms, differential evolution and Monte Carlo based calibration approach [7] are applied in calibration of the modified ASM1. Genetic algorithms are optimisation methods which operate on a population of potential solutions. The population is improved over several generations using genetic operators inspired by biological evolution. The potential solutions are encoded into chromosomes. Binary encoding is used most commonly, but other representations such as real valued encoding can be used. Information about the solution in the chromosomes must be decoded before fitness of the solution can be evaluated by an objective function. The chromosome without knowledge of the encoding provides no information about the optimisation problem. The optimisation process, however, operates on the encoded chromosomes using genetic operators. [18]

Individual solutions are chosen to reproduce the next generation with a probability related to their fitness value. Recombination operators are used to combine genetic information from the selected parents to produce offspring. Mutation operator is applied to the new population to guard against the loss of genetic material in the selection and crossover. The new population is then again evaluated with the objective function. The GA is usually terminated after a predefined number of generations has been achieved. [18]

Differential evolution is also based on a population of evolving solutions. The most novel aspect of DE is the differential mutation operator used in producing intermediary population. In differential mutation scaled difference of two randomly selected solutions is added to a third randomly selected solution. The intermediary population is recombined with the current population to produce a trial population. Potential solutions of the trial population compete with solutions of the current population for placement in the next generation. [19]

There are many variants of GA and DE, and the choice of a suitable variant for a given problem is far from obvious. The GA variant used in this study is adopted from [13]. The real-coded GA described in [13] allows encoding parameter values in chromosomes without conversion between real values and binary digits. Genetic operators are also applied as described in [13].

DE is inherently real-coded. Therefore it is directly applicable to the parameter optimisation problem at hand. Of the different DE variants DE/rand/1/bin is said to be more reliable while DE/best/1/bin converges faster on some problems. The terms "rand" and "best" after "DE" specify how the base vector is chosen in differential mutation. "rand" means that the base vector is chosen randomly from the current population and "best" means that the best vector from current population is chosen. "bin" means that uniform crossover is used in both variants. [19] Both DE/rand/1/bin and DE/best/1/bin are compared in this study.

The effect of constraint handling method on the performance of DE is evaluated by comparing three constraint handling methods: brick wall penalty, random reinitialisation of out-of-bounds parameters and bounce-back. Boundary constraints arise from the fact that parameters of the mechanistic model have limitations on their possible values. The brick wall penalty sets a very high objective function value for solutions with out-of-bounds parameters to make sure such solutions will not be selected to the next generation. Random reinitialisation replaces offending parameters with random values from the allowed range. The bounce-back method replaces offending parameters with values that lie between the mutation base vector and the violated bound [19].

The Monte Carlo approach for model calibration of Sin et al. [7] uses Latin hypercube sampling to draw a predefined number of samples from the parameter

space. The samples are evaluated with an objective function such as Eq. (8). The sample with the lowest objective function value is chosen as it describes the measured data best. In [7] the LHS is integrated into complete modelling procedure including the choice of parameters to be calibrated, definition of parameters space and testing the calibrated parameter set with data that was not used in the model calibration.

3 Results and discussion

3.1 Identifiability analysis

The identifiability analysis methodology of Brun et al. [6] presented in Chapter 2.3 was used to find an identifiable parameter subset. The 27 kinetic, stoichiometric and wastewater composition parameters listed in Tabs. 1 and 2 were included in the identifiability analysis.

The finite difference method (Eq. 4) used in calculating the sensitivity functions (Eq. 3) requires the choice of a proper perturbation factor ξ. The choice of ξ determines the quality of sensitivity function. Perturbation factor should approach zero, but in practice precision of calculations limits the choice [20]. Sensitivity analysis was run with ξ values of 0.05, 0.01, 0.005 and 0.001. ξ value of 0.01 was chosen, as smaller values produced inconsistent results.

The sensitivity function in Eq. 3 was calculated for all 27 parameters of Tabs. 1 and 2 using the calibration data set. Three model outputs, total COD, N and P, were considered in the sensitivity analysis. Sensitivity measure δ_j^{msqr} defined in Eq. 5 was calculated for all parameters using the sensitivity functions. The parameters were ranked according to their δ_j^{msqr} values, and collinearity index γ_K defined in Eq. 6 and determinant measure ρ_K defined in Eq. 7 were calculated for all parameter subsets K of sizes 2 to 22. The lowest ranking five parameters were removed as the least significant parameters have very little effect on model outputs, and to reduce computation time. Including all 27 parameters would have required γ_K and ρ_K to be calculated for 134217700 parameter subsets while including 22 parameters reduced the number of subsets to 4194281.

Results from calculation of γ_K and ρ_K for different subset sizes are shown in Tab. 3. Ranges of γ_K and ρ_K, ρ_K of subset with the smallest γ_K and proportion of subsets with $\gamma_K < 10$ are given for parameter subsets with sizes from 2 to 17. γ_K value of 10 was chosen as a threshold for subset identifiability in line with [6]. Subsets with $\gamma_K \geq 10$ are considered poorly identifiable. There were no subsets of 17 or more parameters with $\gamma_K < 10$.

Tab. 3. Identifiability analysis results.

Subset size	$\gamma_{K,min}$	$\gamma_{K,max}$	$\rho_{K,min}$	$\rho_{K,max}$	$\rho(\gamma_{min})$	$\gamma_K<10$ [%]
2	1.00	12.61	0.36	34.21	1.72	99.6
3	1.01	18.71	0.34	18.87	2.38	97.9
4	1.09	25.34	0.34	12.64	1.90	93.7
5	1.31	27.42	0.33	9.18	1.07	86.6
6	1.51	34.21	0.32	7.21	0.93	76.5
7	1.81	38.03	0.31	5.49	0.95	64.2
8	2.02	40.89	0.31	4.37	1.17	50.8
9	2.24	45.61	0.35	3.53	1.02	37.5
10	2.48	47.75	0.38	2.89	0.76	25.7
11	2.57	49.02	0.42	2.45	0.93	16.2
12	2.71	50.26	0.45	2.10	0.82	9.1
13	3.11	51.49	0.48	1.81	0.86	4.5
14	3.54	52.37	0.51	1.59	0.62	1.8
15	4.01	53.17	0.53	1.41	0.77	0.5
16	5.40	53.64	0.55	1.27	0.80	0.1
17	12.42	53.91	0.56	1.14	0.74	0.0

Parameter subsets with $\gamma_K < 10$ are found up to subset size of 16. However, there are no subsets with more than 10 parameters having $\gamma_K < 10$ in hundred best subsets ranked with ρ_K values. This indicates that the largest identifiable subsets contain mostly insignificant parameters. A subset of ten parameters was chosen to be identified to calibrate the model. The subset was chosen as it had both low γ_K and high ρ_K values. The chosen subset is given in Tab. 5. Parameters which were not included were given their initial values listed in Tabs. 1 and 2.

3.2 Model calibration

The modified ASM was calibrated by optimising the parameter subset which was chosen in the previous chapter. A genetic algorithm, two differential evolution variants with different boundary constraint handling methods and the Monte Carlo based calibration approach [7] were all applied and evaluated for their performance in calibration. Boundary constraints for the parameters were defined by the uncertainty classes in Tabs. 1 and 2.

Optimisation method parameters for both DE variants rand/1/bin and best/1/bin were population size = 50, scale factor F = 0.8 and crossover probability Cr = 0.9. The applied GA was adopted from [13] including the choice of GA parameters. Both GA and DE were run for 50 generations for a total of 2500 objective function (Eq. 8) evaluations. With the Monte Carlo based calibration approach, 2500 samples of parameter values were drawn using LHS. Initial values of the optimised model parameters are listed in Tabs. 1 and 2.

As the performance of evolutionary optimisers may vary from run to run, parameter optimisation was run ten times with each method. Results from the ten runs were averaged as the interest is on average performance of the methods instead of a single best

result. Progress of solution for the evolutionary optimisers is shown in Figs. 1 and 2. In the Figs. 1 and 2, objective function values of the full population averaged over ten runs is plotted as a function of generation. The initial randomly generated population is omitted from the Figs. 1 and 2. DE with brick wall penalty converges most slowly while the other methods have quite similar performance. The GA has variation until the end of the optimisation run due to its mutation operator. DE/best/1/bin with bounce-back boundary constraint handling method has the fastest convergence.

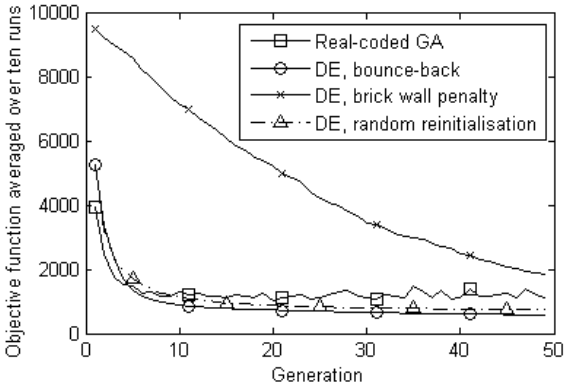

Fig. 1. Progress of solution for the GA and DE/rand/1/bin. Objective function values of the full population averaged over ten runs and plotted as a function of generation.

It can be seen from the Figs. 1 and 2 that with the best performing optimisation methods the result after 15 to 20 generations is already near the final result of full run of 50 generations. In practical applications it is not reasonable to run the optimisation further than necessary as computation times are considerable. On a fairly efficient desktop computer (AMD Athlon™ 64 X2 Dual Core 4200+, 2.00 GB RAM) 2500 objective function evaluations required approximately 18 hours of computation. The required computation time is directly proportional to the number of objective function evaluations, so stopping the optimisation after 15 generations (750 objective function evaluations) corresponds to computation time of approximately five and half hours.

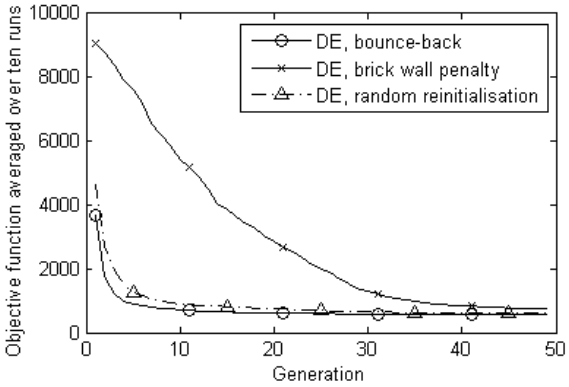

Fig. 2. Progress of solution for DE/best/1/bin. Objective function values of the full population averaged over ten runs and plotted as a function of generation.

Performance of the optimisation methods is compared at 750 objective function evaluations. For each method, best solution was taken from each of the ten runs. For these best solutions, mean and standard deviation of objective function values and sum of squared errors (SSE) of the three components in the objective function was calculated for both calibration and testing data to give the results in Tab. 4.

From Tab. 4 it can be seen that DE/best/1/bin and DE/rand/1/bin with bounce-back boundary constraint handling gave the best results for calibration data. The Monte Carlo method performed worse than

Tab. 4. Mean performance and standard deviations (in parentheses) of the best solutions averaged over the ten runs.

	Training data				Testing data			
	Obj. fun	COD, SSE	P, SSE	N, SSE	Obj. fun	COD, SSE	P, SSE	N, SSE
GA	721.7	$8.79*10^5$	3.5	28.5	322.9	$1.47*10^6$	14.1	270.7
	(39.2)	($4.40*10^4$)	(0.6)	(6.2)	(7.8)	($8.20*10^4$)	(0.9)	(16.0)
DE/rand/1/bin, bounce-back	636.3	$7.60*10^5$	2.69	30.8	325.5	$1.58*10^6$	13.0	287.1
	(29.9)	($6.25*10^4$)	(0.23)	(9.4)	(16.2)	($1.21*10^5$)	(0.9)	(49.1)
DE/rand/1/bin, brick wall penalty	852.6	$9.84*10^5$	3.6	45.5	328.2	$1.59*10^6$	14.3	255.1
	(65.5)	($1.97*10^5$)	(0.6)	(20.5)	(22.6)	($2.51*10^5$)	(0.5)	(57.9)
DE/rand/1/bin, random reinitialisation	707.6	$8.73*10^5$	3.0	32.8	328.0	$1.51*10^6$	14.1	279.2
	(52.7)	($1.03*10^5$)	(0.3)	(7.6)	(11.9)	($1.72*10^5$)	(1.0)	(30.1)
DE/best/1/bin, bounce-back	604.0	$7.05*10^5$	2.7	30.3	323.3	$1.47*10^6$	13.1	301.0
	(56.3)	($6.11*10^4$)	(0.3)	(4.9)	(9.8)	($1.04*10^5$)	(0.8)	(19.6)
DE/best/1/bin, brick wall penalty	808.2	$8.83*10^5$	3.4	49.4	319.9	$1.42*10^6$	14.7	253.8
	(76.9)	($1.23*10^5$)	(0.5)	(16.6)	(21.1)	($1.51*10^5$)	(0.8)	(62.4)
DE/best/1/bin, random reinitialisation	646.3	$7.64*10^5$	3.1	29.0	315.3	$1.48*10^6$	13.5	265.5
	(33.7)	($3.87*10^4$)	(0.5)	(4.1)	(11.4)	($1.21*10^5$)	(1.1)	(22.3)
LHS	767.0	$9.08*10^5$	3.5	35.8	328.4	$1.57*10^6$	14.1	266.5
	(45.2)	($1.50*10^5$)	(0.7)	(16.9)	(17.6)	($2.59*10^5$)	(1.0)	(35.1)

evolutionary optimisers except for DE with brick wall penalty. DE with brick wall penalty was the most inferior optimisation method. However, the best parameter set for the calibration data did not result in the best performance for the testing data. Therefore the calibration and testing data sets have different optimal parameter values.

Although the optimisation runs produced many very similar results with only slight variance in the optimised parameter values, simulations with one result are presented. Optimised parameter values of the result with the lowest objective function value for the calibration data are given in Tab. 5. The result was obtained with DE/best/1/bin with bounce-back boundary constraint handling. Comparison between the simulation results and measured WWTP effluent total COD, N and P concentrations are illustrated in Figs. 3-5. The dotted vertical line in the Figs. 3-5 represents the split between calibration and testing data. By visual comparison the measured and simulated values are in good agreement for both calibration and testing data.

Tab. 5. Optimised parameter values of the result with lowest objective function value for calibration data.

Parameter	Calibrated value
Y_H	0.804
i_{XB}	0.049
i_{XP}	0.025
$i_{XBP,1}$	0.0026
$i_{XBP,2}$	0.0059
K_S	14.01
K_X	0.123
Biodegradable COD	0.747
Ratio of S_I to nonbiodegradable COD	0.752
Ratio of S_{NI} to soluble N	0.043

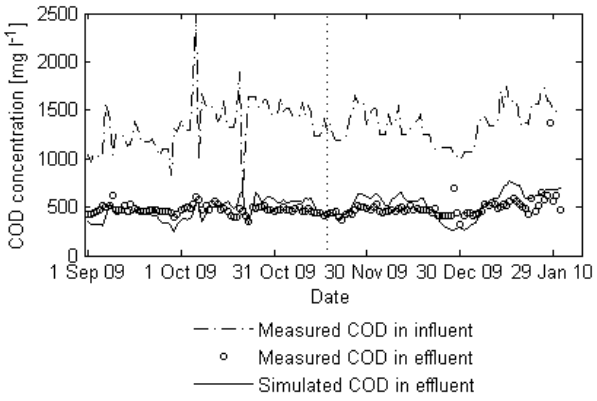

Fig. 3. The simulated COD concentration in the WWTP effluent and the measured COD concentration in the WWTP influent and effluent.

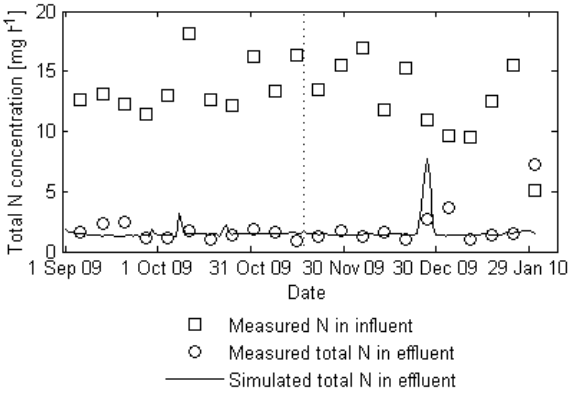

Fig. 4. The simulated N concentration in the WWTP effluent and the measured N concentration in the WWTP influent and effluent.

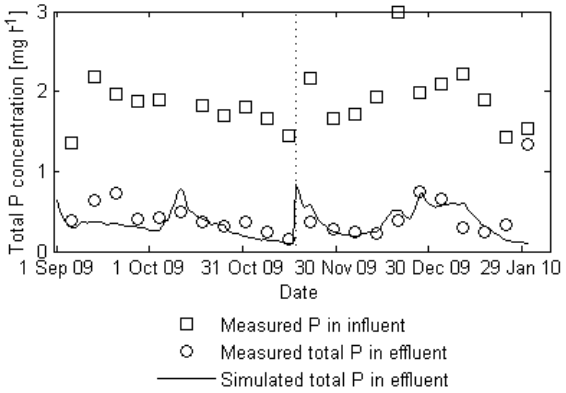

Fig. 5. The simulated P concentration in the WWTP effluent and the measured P concentration in the WWTP influent and effluent.

4 Conclusions

Given enough computation time, genetic algorithms, differential evolution and the Monte Carlo based calibration approach are all applicable for calibration of Activated Sludge Models. Some of the DE variants, DE/rand/1/bin and DE/best/1/bin with bounce-back boundary constraint handling, require much less computation time than less efficient methods to achieve sufficient convergence. This gives the evolutionary optimisers an edge in terms of computation time against the Monte Carlo methods. However, the Monte Carlo method is very simple to implement and the only required parameter for the optimisation method is the number of samples. Bad choice of evolutionary optimisation method and variant, as in this case DE with brick wall penalty, can actually lead to worse performance than the random search method.

It was found that the most optimal parameter set for the calibration data did not result in the best performance for testing data. Therefore calibration and testing data sets have somewhat different optimal parameter values. It is therefore necessary to recalibrate the model regularly in practical

applications. This result emphasises the importance of developing methods to automate the model calibration.

5 References

[1] K.V. Gernaye, M.C.M van Loosdrecht, M. Henze, M. Lind, and B. Jørgensen. Activated sludge wastewater treatment plant modelling and simulation: state of the art. *Environmental Modelling & Software*, 19:763-783, 2004.

[2] M. Henze, W. Gujer, T. Mino, and M.C.M. van Loosdrecht, eds. Activated Sludge Models ASM1, ASM2, ASM2d and ASM3. Scientific and Technical Report no 9. IWA Publishing 2000.

[3] B. Petersen, K. Gernaey, M. Henze, and P.A. Vanrolleghem. Calibration of activated sludge models: a critical review of experimental designs. In: S.N. Agathos and W. Reineke (Eds.), Biotechnology for the Environment: Wastewater Treatment and Modeling, Waste Gas Handling. Kluwer Academic Publishers. 2003.

[4] G. Sin, S.W.H. Van Hulle, D.J.W. De Pauw, A. Griensven, and P.A. Vanrolleghem. A critical comparison of systematic calibration protocols for activated sludge models: a SWOT analysis. *Water Research*, 39:2459-2474, 2005.

[5] J. Keskitalo and K. Leiviskä. Mechanistic modelling of pulp and paper mill wastewater treatment plants. Report A No. 41. University of Oulu, Control Engineering Laboratory. 2010. Available at: http://ntsat.oulu.fi/file.php?718

[6] R. Brun, M. Kühni, H. Siegrist, W. Gujer, and P. Reichert. Practical identifiability of ASM2d parameters – systematic selection and tuning of parameter subsets. *Water Research*, 36:4113-4127, 2002.

[7] G. Sin, D.J.W. De Pauw, S. Weijers, and P.A. Vanrolleghem. An efficient approach to automate the manual trial and error calibration of activated sludge models. *Biotechnology and Bioengineering*, 100:516-528, 2008.

[8] S.R. Weijers and P.A. Vanrolleghem. A Procedure for selecting best identifiable parameters in calibrating Activated Sludge Model no.1 to full-scale plant data. *Water Science and Technology*, 36:69-79, 1997.

[9] V.C. Machado, G. Tapia, D. Gabriel, J. Lafuente, and J.A. Baeza. Systematic identifiability study based on the Fisher Information Matrix for reducing the number of parameters calibration of an activated sludge model. *Environmental Modelling & Software*, 24:1274-1284, 2009.

[10] S. Hengl, C. Kreutz, J. Timmer, and T. Maiwald. Data-based identifiability analysis of non-linear dynamical models. *Bioinformatics*, 23:2612-2618, 2007.

[11] M. Heikkinen, T. Heikkinen, and Y. Hiltunen. Process states and their submodels using self-organizing maps in an activated sludge treatment plant. The 49th Scandinavian Conference on Simulation and Modeling (SIMS 2008), Oslo, Norway, October 7-8 2008.

[12] O. Roeva. Parameter estimation of a Monod-type model based on genetic algorithms and sensitivity analysis. In I. Lirkov, S. Margenov, and J. Waśniewski, editors, Large-Scale Scientific Computing. Springer Berlin / Heidelberg. 2008.

[13] A. Sorsa, R. Peltokangas, and K. Leiviskä. Real-coded genetic algorithms and nonlinear parameter identification. IEEE International Conference on Intelligent Systems (IS 2008), Varna, Bulgaria, September 6-8 2008.

[14] J.-P. Chiou and F.-S. Wang. Estimation of Monod model parameters by hybrid differential evolution. *Bioprocess and Biosystems Engineering*, 24:109-113, 2001.

[15] E. Lindblom, C. Rosen, P.A. Vanrolleghem, L.-E Olsson, and U. Jeppsson. Modelling a nutrient deficient wastewater treatment process. 4th IWA World Water Conference, Marrakech, Morocco, September 19-24 2004.

[16] J. Keskitalo, J. la C. Jansen, and K. Leiviskä. Calibration and validation of a modified ASM1 using long-term simulation of a full-scale pulp mill wastewater treatment plant. *Environmental Technology,* 31:555-566, 2010.

[17] J.B. Copp (ed.) The COST simulation benchmark: Description and simulator manual, Office for Official Publications of European Communities, Luxembourg. 2002.

[18] A. Chipperfield. Introduction to genetic algorithms. In: A.M.S. Zalzala and P.J. Fleming (Eds.), Genetic Algorithms in Engineering Systems, IEE London. 1997.

[19] K.V. Price, R.M. Storn, and J.A. Lampinen. Differential Evolution – A Practical Approach to Global Optimization. Springer Berlin Heidelberg. 2005.

[20] D.J.W. De Pauw and P.A. Vanrolleghem. Practical aspects of sensitivity function approximation for dynamic models. *Mathematical and Computer Modelling of Dynamical Systems*, 12:395-414, 2006.

Table 1. Stoichiometric coefficients of the modified ASM1 [15-16].

Component i → / Process j ↓	1. S_I	2. S_S	3. X_I	4. X_S	5. X_{BH}	6. X_P	7. S_O	8. S_{NH}	9. S_{ND}	10. X_{ND}	11. S_{NI}	12. X_{NB}	13. X_{NI}	14. X_{NP}	15. S_P	16. X_{PD}	17. X_{PP}	18. $X_{PB,1}$	19. $X_{PB,2}$
1. Aerobic growth of heterotrophs		$-Y_H^{-1}$			1		$1 - Y_H^{-1}$	$-i_{XB}$				i_{XB}			$-(i_{XPB,1} + i_{XPB,2} \cdot S_P \cdot (K_{Pi} + S_P)^{-1})$			$i_{XPB,1}$	$i_{XPB,2} \cdot S_P \cdot (K_{Pi} + S_P)^{-1}$
2. Decay of heterotrophs				$1 - f_P$	-1	f_P				$i_{XB} - f_P \cdot i_{XB}$		$-i_{XB}$		$f_P \cdot i_{XP}$		$(X_{PB,1} + X_{PB,2}) \cdot X_{BH}^{-1} - f_P \cdot i_{XPP}$	$f_P \cdot i_{XPP}$	$-X_{PB,1} \cdot X_{BH}^{-1}$	$-X_{PB,2} \cdot X_{BH}^{-1}$
3. Ammonification of soluble organic nitrogen								1	-1										
4. Hydrolysis of entrapped organics		1		-1															
5. Hydrolysis of entrapped organic nitrogen									1	-1									
6. Hydrolysis of entrapped organic phosphorus															1	-1			

Table 2. Process rates of the modified ASM1 [15-16].

Process j	Process rate r_j
1. Aerobic growth of heterotrophs	$\hat{\mu}_H \left(\frac{S_{NH}}{K_{NH} + S_{NH}} \right) \left(\frac{S_P}{K_{P1} + S_P} \alpha + \frac{S_P}{K_{P2} + S_P} (1-\alpha) \right) \left(\frac{S_S}{K_S + S_S} \right) \left(\frac{S_O}{K_{OH} + S_O} \right) X_{BH}$
2. Decay of heterotrophs	$b_H X_{BH}$
3. Ammonification of soluble organic nitrogen	$k_a S_{ND} X_{BH}$
4. Hydrolysis of entrapped organics	$k_h \frac{X_S X_{BH}^{-1}}{K_X + X_S X_{BH}^{-1}} \left(\frac{S_O}{K_{OH} + S_O} \right) X_{BH}$
5. Hydrolysis of entrapped organic nitrogen	$k_h \frac{X_S X_{BH}^{-1}}{K_X + X_S X_{BH}^{-1}} \left(\frac{S_O}{K_{OH} + S_O} \right) \left(\frac{X_{ND}}{X_S} \right) X_{BH}$
6. Hydrolysis of entrapped organic phosphorus	$k_h \frac{X_S X_{BH}^{-1}}{K_X + X_S X_{BH}^{-1}} \left(\frac{S_O}{K_{OH} + S_O} \right) \left(\frac{X_{PD}}{X_S} \right) X_{BH}$