Methods for anaerobic digestion model fitting—comparison between heuristic and automatic approach

Karol Postawa¹ · Jerzy Szczygieł¹ · Marek Kułczyński¹

Received: 14 May 2020 / Revised: 29 July 2020 / Accepted: 4 August 2020 / Published online: 29 August 2020
 © The Author(s) 2020

Abstract

The article demonstrates if automatic optimization can be better than manual adjustment. The subject of optimization was the temperature-phased anaerobic digestion (TPAD) model. A selection of 3 parameters per each reactor in the process chain was appointed—reaction rate for propionate conversion, acetate conversion, and hydrolysis. Overall, both methods provided very convergent results. However, the total summary error (TSE) for the automatic algorithm was always moderately lower than for manual—the difference varied between 16.16 and 57.05 percentage points. Although the manual method has significant advantages—adjustment was more homogenous and gave more uniform fitting. Finally, cross-validation was performed to unify the values between the experimental series. The result was a total number of 4 values for each optimized constant—for two temperature points in each of two methods. Due to inconclusive information about the accuracy, averaged values were calculated to use in further researches. The recommendation from this article is to connect the best aspect of both methods to achieve the most accurate results.

Keywords Biogas · TPAD · Model · Algorithm · Optimization

Abbreviations

δ_x	The total error for iteration X
θ	Mean value of constant, from both series and methods
f_x	The fractioning factor for a fraction x
$\text{fun}(k_{m;\text{pro}}, k_{m;\text{ac}}, k_{\text{dis}})$	The objective function in the optimization
k_{dis}	Hydrolysis rate
$k_{m;x}$	Uptake rate for substance X
$k_{x,T}$	Value of constant k_x for the temperature T
$Q_{\text{gas}}, Q_{\text{gas.exp}}$	Biogas flow for model or experiment
Q_{in}	Input biomass flow to the reactor
$S_x, S_{x,\text{exp}}$	The concentration of substance X , from model or experiment
t_{sr}	The additional retention time for solids in addition to the hydraulic retention time
V_{Rx}	The total volume of the reactor X
X_c	Total feeding value of biomass

x_{IT}	Value of measured constant for iteration IT and time n for the model (y_n) or experimental results (1_n)
θ	The temperature-dependence coefficient
MRE	Mean relative error
ODEs	Ordinary differential equations
R1, R2	Reactor 1 or reactor 2
T	Temperature
TSE	Total summary error

1 Introduction

The computing power of current systems is increasing, leading to a trend in which mathematical modeling becomes an actual competition for the in-scale experiment [1–3]. This includes also the bioenergy sector [4] and the designing of related bioprocesses [5]. However, the construction and optimization of the suitable model are not trivial [6] and need lots of human and computational resources—thus, many shortcuts in achieving the goal are used. One of them is utilizing automatic algorithms for optimizing of model constants, instead of adjusting them manually [7, 8]. This approach undeniably saves time, but can it be enough precise and trustworthy as the work of an experienced

✉ Karol Postawa
karol.postawa@pwr.edu.pl

¹ Department of Chemistry, Wrocław University of Science and Technology, Wybrzeże Wyspiańskiego 27, 50-370, Wrocław, Poland

scientist, which does not work in cold, machine way, but uses his knowledge to heuristically find the best result? This question lays in the basics of this work.

The manual technique applied in this work, to solve the nonlinear optimization problem (from the mathematical point of view), has all features of heuristic methods [9, 10]. They are used to make decisions basing on creative thinking, intuition, and logical combinations. In the results, heuristic methods let to find the way to follow, to achieve the intended target. On the one hand, manual optimization of constants needs a lot of work, as all changes in the model need to be by the researcher itself. It also limits the number of combinations possible to test, as the speed of human work is limited comparing with automatic algorithms. On the other hand, this approach gives the best control over the process, as at every step of optimization the researcher can take independent conclusions and change the direction of the further test, basing on obtained results [11]. This gives much higher flexibility and potentially protects from some kind of errors. However, it needs to be mentioned, that every change in the values of constants is based on the subjective sense of the scientist, which may lead to inaccuracy if his predictions will be incorrect or initially targeted to achieve desired results.

The automatic algorithm will be much faster than human work, as it can test hundreds of combinations within minutes [12]. It is also fully arbitral, and unidirectional as its decisions are made based on independent and universal mathematical definitions [13, 14], unaffected by any subjective aspects. However, these advantages are charged with a high price—in some cases, the algorithm can give results that make sense in mathematical meaning, but are incorrect from a biological or physical point of view. For example, the very high value of a selected constant can be satisfactory for optimization query but can exceed values in which it will be expected in a real experiment. This is a potential trap and shows how important is to construct correct, and studied statement for the function to minimize. Also if the reference data for curve fitting includes discontinuity or is not smooth, it can mislead the algorithm or deteriorate accuracy [15].

The fundamental difference between searching for a solution using exact and heuristic algorithms is that the first approach returns the real optimal solution (in the perfect case) while the second provides only approximate one, and only in particular instance exact [10]. Due to this, the exact methods are applied mostly in the cases that are well explored and recognized, while heuristics are more common in all places where there are not known the algorithms that would let to find the exact solution in enough time. The search is carried out along the shortest, the most likely way, avoiding the less promising patches. The effectiveness of heuristic steps cannot be fully proved

theoretically, it is possible only to experimentally show their relevance [16].

In this article, a detailed comparison between manual and automatic optimization of constants will be presented. As a reference, a mathematical model of the biogas plant, and the literature data from real in-scale installations will be utilized. To be more specific—the temperature-phased anaerobic digestion (TPAD) system will be considered [17, 18] (Fig. 1).

This concept uses the differences in environmental preferences between selected groups of microorganisms to split their populations (at least partially) between two steps of the process. The first reactor is kept in higher temperature (typically around 55 °C), to promote hydrolysis, which is particularly important for hardly decomposable feedstocks. However, these conditions are not proper for methanogenesis, which leads to the limited efficiency of biogas production, and potential issues in the stability of the process [19]. Thus in the second step, the temperature is kept in the mesophilic range (around 35 °C), to promote the further conversion of intermediate products from hydrolysis to the final fuel. By that means, all processes involved in the anaerobic digestion path—hydrolysis, acidogenesis, acetogenesis, and methanogenesis [20]—can occur in optimal conditions.

2 Materials and methods

The optimization in this article was performed for a biogas plant model of two-phase anaerobic digestion. This concept of installation introduces separating bioprocesses, which take place during fermentation, to two different tanks. To be more specified, in the case considered in our work, the phasing is performed by keeping different temperatures of the process between steps. The model was described in our previous works [21, 22] and consists of 29 ordinary differential equations (ODEs) per every reactor in the process chain.

Overall, the selection of constants to adjust vary between researchers [23–25]. However, some of them repeat in every study, namely, hydrolysis rate (k_{dis}), uptake rate for propionate ($k_{m;pro}$), and acetate degrading organisms ($k_{m;ac}$). The relation between temperature and value of first of the mentioned parameters is the clearest, as most of the researches agree that increasing the temperature—which takes place in the first reactor in the considered system—will raise the ratio of hydrolysis. For the remaining two parameters, it will be more complicated to declare correct correlations and detailed optimization is required.

Irrespective of the method of optimization—manual or automatic, all simulations were performed in pure MATLAB suite, without the support of Simulink. To keep

Fig. 1 The overall scheme of the TPAD concept

all calculations reproducible and equivalent, in all cases the same ODEs solver is utilized—namely ode15s [26]. This algorithm is dedicated to stiff systems, which fits perfectly in the requirements for biological simulation [27]. What's more, it provides relatively high accuracy, with very good performance.

Also, both optimization methods were split into two series—in the first, the full TPAD system was modeled. In the second, two independent steps were considered—one for mesophilic and one for the thermophilic reactor, both with shifting feeding rate. The starting from raw biomass in both reactors, instead of effluent from the previous reactor, let to prevent potential error accumulation in the modeling of the mesophilic step. In this way, we can be sure that the constants are optimized correctly, while still being tested in the complete process chain.

As a reference to simulation results, experimental data from the following studies [23, 25] were taken. In both series, the biogas was produced from sewage sludge—it is a common substrate, also in co-digestion [28]. In the case of hydrolysis constant, overall biogas production was chosen as a parameter to fit in the graph. For $k_{m;pro}$, the concentration of dissolved propionate was taken, and for $k_{m;ac}$ similarly—dissolved acetate concentration. The summary of the model configuration for both series of simulations is presented in Table 1.

2.1 A heuristic method for optimization.

The manual method of optimization was relatively simple, thus could not test a high number of parameter combinations [29]. In general, the sequence of steps was as follows: the constants were optimized starting from the $k_{m;pro}$, while the decision-making factor was the convergence between experimental points and model predictions plots. Parallel with the value of $k_{m;pro}$, also from 2 to 3 combinations of hydrolysis constants were considered, to check its eventual impact on the concentration of propionate.

If the best-fitting configuration were inside the tested range, extreme values from both ends were eliminated, and a new range with the higher resolution was generated, to

define more precisely the final optimization value. If the best value were on the edge of the range, its borders were slid to look for the solution in a new scope. The procedure was repeated as long as the fitting between experimental and modeled data was not assumed arbitrarily as satisfying. Then, analogical procedures for $k_{m;pro}$ and finally for k_{dis} were considered. A summary of the range of optimization and the number values per step is presented in Table 2.

For the first configuration, with two subsequential reactors usually between 110 and 140 combinations of constants for reactors were tested. While for the second system, where reactors were considered separately, and with shifting feeding rate, more iterations were required—20–35 per every reactor. Considering that on every iteration usually around 15 combinations are tested, the final number of trails can reach the level of 500. The flowchart of a single trail, the same for every constant and optimization series, is presented in Fig. 2.

2.2 Algorithm of automatic optimization.

The automatic optimization was performed for the same constants as previously described. However, basing on the knowledge from the earlier research, the range of optimization was slightly changed. Also, it was unified for both reactors in the process chain. These changes are necessary to speed up an optimization algorithm and its sum-up, as well as the initial condition for the algorithm, is presented in Table 3.

The automatic optimization can be performed with different algorithms. The Optimization Toolbox for MATLAB provides a wide selection of methods, from problem-based, through solver-based, ending with multi-objectives functions. As the model required the external solver for ODEs, the most logical and natural choice is the solver-based approach. In this category, 5 most common methods are available and in this study, the *fmincon* algorithm will be utilized [30, 31]. It can be used for nonlinear multivariable functions what is important in the considered case. What's more, it lets to declare upper and lower borders, which is very useful if a rough estimation of the range, in which the

Table 1 Model configuration for both simulation series

Parameter	Unit	Description	Series 1 (R1+R2)	Series 2, R1	Series 2, R2
X_c	kg COD/m ³	Feeding value	64.4	32	42
f_{pr_xc}	-	Protein fraction	0.247	0.18	0.18
f_{li_xc}	-	Lipids fraction	0.238	0.27	0.27
f_{ch_xc}	-	Fiber fraction	0.117	0.12	0.12
f_{XI_xc}	-	Inert fraction	0.398	0.43	0.43
T_1	°C	R1 temperature	55	55	—
T_2	°C	R2 temperature	35	—	35
V_{R1}	m ³	Volume of R1	0.19	0.016	—
V_{R2}	m ³	Volume of R2	0.94	—	0.025
Q_{in}	m ³ /day	Input feedstock flow	0.04	0.001786; 0.003232;	0.001212; 0.002438; 0.004002;
t_{sr}	day	Additional solids retention time	0	10	10

values should be searched, is known. Additionally instead of subsequential optimization, like for the manual method, the constants are changed independently in random order, which can lead to interesting results, impossible to achieve by organized, manual adjustment.

The algorithm needs at least 4 parameters—upper and lower borders, initial conditions, and a function to minimize. First, three aspects were discussed above, and the last one is strictly connected with the character of the process constant which will be optimized. As it was already mentioned, the sum of 3 process parameters, every bounded with the different measured value, is considered during optimization. All of them need to be included in the optimization query. The most simple and elementary function to minimize can look as follows:

$$\text{fun}(k_{m;pro}, k_{m;ac}, k_{dis}) = |Q_{gas} - Q_{gas_exp}| + |S_{pro} - S_{pro_exp}| + |S_{ac} - S_{ac_exp}| \quad (1)$$

where S_x means a concentration of dissolved substance x and Q_{gas} is a biogas output flow from the reactor. The indicator “exp” shows values from the experiment, while other parameters are obtained from the simulation. In general, this equation is correct and would lead to

some conclusions about the process, but it considers the concentration only at one point in time. Thus, it needs to be rewritten as follows:

$$\text{fun}(k_{m;pro}, k_{m;ac}, k_{dis}) = \sum_{i=1}^n |Q_{gas_n} - Q_{gas_exp_n}| + |S_{pro_n} - S_{pro_exp_n}| + |S_{ac_n} - S_{ac_exp_n}| \quad (2)$$

This form includes a sum of all differences between measured and modeled values, for all points of time where both are available. However, the order of the magnitude of selected parameters can vary significantly between them. Thus, it is justified to consider differences as a relative error. The function to minimize takes the form:

$$\text{fun}(k_{m;pro}, k_{m;ac}, k_{dis}) = \sum_{i=1}^n \frac{|Q_{gas_n} - Q_{gas_exp_n}|}{Q_{gas_exp_n}} + \frac{|S_{pro_n} - S_{pro_exp_n}|}{S_{pro_exp_n}} + \frac{|S_{ac_n} - S_{ac_exp_n}|}{S_{ac_exp_n}} \quad (3)$$

And this form of function for optimization query will be utilized in further study. It connects the best aspect of both described earlier, additionally preventing the directional errors from different orders of magnitude between the parts of the objective function. In the second series of optimization, additional correction is included—the relative error for propionate concentration is weighted by 0.5, to limit its excessive impact on biogas curve fitting.

However, the *fmincon* itself is an algorithm that looks for a local minimum, which means the results will strongly depend on the initial conditions and is unable to determine if the solution is global or local. In the matter of model optimization, it would be to search for a global minimum or at least its approximation. Thus, the method was not called directly, but by *GlobalSearch*, from MATLAB Global Optimization Toolbox. These extensions run the *fmincon*

Table 2 The optimization range for the manual method

Parameter	Unit	Min	Max	Tested values per step
Thermophilic reactor (55 °C)				
$k_{m;pro}$	day ⁻¹	0.9	60.9	5
$k_{m;ac}$	day ⁻¹	0.74	25.74	5
k_{dis}	day ⁻¹	0.1	0.9	2–5
Mesophilic reactor (35 °C)				
$k_{m;pro}$	day ⁻¹	0.5	13	5
$k_{m;ac}$	day ⁻¹	0.25	20.13	5
k_{dis}	day ⁻¹	0.00625	1.2	3–5

Fig. 2 The flowchart of a single trail of optimization

multiple times, to test many combinations of not only optimized variable but also initial conditions. In contrast to the similar, but simpler *MultiStart* approach, it not only generates a series of start points but also evaluate them with score function and reject those that would not improve the solution [30]. It can potentially save calculation time, which is very important for a rather complicated and demanding biological system. However, it needs to be noticed that the *GlobalSearch* algorithm uses a scatter-search mechanism [32], which is more complicated than in the case of the *MultiStart* [33] and cannot run parallel on multiple CPU cores, which can affect performance in some cases. As a result, the *GlobalSearch* returns a vector of solutions, which are a representation of points assume as the global minimum for the optimization problem.

Table 3 The optimization range for the automatic method

Parameter	Unit	Min	Max	Initial Conditions
$k_{m;pro}$	day ⁻¹	2.5	17.5	10
$k_{m;ac}$	day ⁻¹	10	25	15
k_{dis}	day ⁻¹	0.0005	1.5	0.25

3 Results and discussion

3.1 Results of manual and automatic optimization

As it was discussed, both techniques have benefits and disadvantages. In general, they should provide similar results, at least in the same order of magnitude. And indeed—the results show high convergence in the overall assessment, however, in detail, there are significant variations. Both methods led to an acceptable biogas production profile in the first series, which is presented in Fig. 3.

In both cases, the manual adjustment led to higher projected biogas production. However, considering relative error between experimental and modeled values, these changes seem to be justified. In the case of manual optimization of the first reactor (Fig. 3a), the error was equal to 36.60 % while using the model parameter suggested by the *fmincon* leads to a relative error on the level of 30.30%. This difference is statistically significant and can lead to preliminary conclusions that automatic optimization can be an interesting alternative to manual adjustments. The situation is even more interesting if the second reactor is considered (Fig. 3b). In this case, also the automatic

algorithm suggested constant values, which lead to lower biogas production profile, but this time the relative error is decreasing very clearly—from 114.78 to 83.10%. This could be surprising, basing on the fact how similar are both plots, and show the advantage of automatic simulation over the manual. It would be very easy to skip a better solution, basing only on subjective analysis.

In the matter of the other two constants and measured parameters connected with them, the relation between method and relative error was not so directly proportional. Despite the propionate concentration projection accuracy increased slightly in both reactors (0.31 and 2.9 percentage point respectively), for the acetate only in the first case, around 10 percentage point decrease in relative error was determined. For the second reactor, the relative error increased by about 8 percentage points. However, this fact does not exclude the usability of automatic methods, as the summary accuracy, considering all parameters, is better. To present it the most clearly, a total summary error (TSE) was calculated as:

$$\text{TSE} = \overline{\delta_{S_{\text{pro}}}} + \overline{\delta_{S_{\text{ac}}}} + \overline{\delta_{Q_{\text{biog}}}} \quad (4)$$

Fig. 3 Comparison between the manually and automatically optimized model in the first series

Table 4 Summary of the first series of optimization

Unit	Constant name	Initial value (day ⁻¹)	Manually adjusted value (day ⁻¹)	TSE (%)	Auto-adjusted value (day ⁻¹)	TSE (%)
R1	$k_{m;\text{pro}}$	13.0	9.775	146.04	10.160	129.88
	$k_{m;\text{ac}}$	8.0	15.188		17.933	
	k_{dis}	0.5	0.120		0.090	
R2	$k_{m;\text{pro}}$	13.0	6.48	241.98	5.498	215.89
	$k_{m;\text{ac}}$	8.0	13.625		15.323	
	k_{dis}	0.5	0.011		0.007	

where $\overline{\delta_{S_x}}$ means the mean value of relative error for all available points of time. The summary of obtained constants and related TSE for the first series is presented in Table 4.

The analogical analysis was performed for the second series of experimental data. As mentioned earlier, this time both reactors are considered separately, with independent feedstock input, to avoid potential error accumulation. What's more, the biomass flow is not constant and is increasing during the experiment, which is a benefit in

the matter of examining model response for additional changes. Again, both methods led to statistically correct values, while the values from an automatic algorithm result in lower biogas production for the thermophilic reactor. The production for the second, mesophilic, is nearly the same. The details are presented in Fig. 4.

For the first reactor (Fig. 4a), the overall mean relative error of biogas production was 30.35% for the manual method, and 25.10% for automatic. Despite weaker fitting in the second part of the simulation, there is a very good convergence in the first 30 days of simulation. On the other hand, the manual method is more balanced, shows similar fitting on both parts of the process.

For the second reactor (Fig. 4b), the biogas production was nearly the same, independently on the method utilized to achieve model parameters. The difference in relative error was just 0.52% in favor of the *fmincon*. In this situation, it is far too low, to determine if it is statistically relevant, so it had to be assumed that both methods did just as well.

Considering other optimized parameters— $k_{m,pro}$, and $k_{m,ac}$, the difference is the biggest in the aspect of propionate concentration, especially for the first reactor.

Fig. 4 Comparison between the manually and automatically optimized model in the second series

Table 5 Summary of the second series of optimization

Unit	Constant name	Initial value (day ⁻¹)	Manually adjusted value (day ⁻¹)	TSE (%)	Auto-adjusted value (day ⁻¹)	TSE (%)
R1	$k_{m,pro}$	13.0	15.150	402.66	16.910	346.36
	$k_{m,ac}$	8.0	15.383		14.550	
	k_{dis}	0.5	0.450		0.240	
R2	$k_{m,pro}$	13.0	13.781	162.18	15.563	105.13
	$k_{m,ac}$	8.0	11.087		17.358	
	k_{dis}	0.5	1.250		1.294	

The improvement in decreasing relative error between manual and automatic methods reached over 52 percentage points. For the second reactor, it was more subtle—as the error decreased from 75.46 to 56.21%. Ultimately in the matter of dissolved acetate concentration, the accuracy was slightly decreased in the first reactor (around 0.61 percentage point), and the accuracy was significantly improved for the second tank—over 37 percentage point

improvement. A summary of the proposed constant values for both steps and connected errors is presented in Table 5.

As previously, in all cases the TSE is decreasing, which suggests that the *fmincon* supported by *GlobalSearch*, can find more precise and accurate values, than a researcher using systematized but manual adjustment.

3.2 Cross-validation and final values

The consideration from Section 3.1 led to interesting results, however, they are still separated between steps and methods, giving a sum of four possible values for any constant in every reactor in the process chain. Additionally, they are correct only for these two fixed points of temperature, which is a significant limitation. Thus, there is a need to indicate full temperature dependencies for the model, so it can work in every temperature specified by the user, and to find a consensus value, between both mentioned methods.

The methodology of calculation was split into two parts. In the first, the unified values of constants for two fixed temperatures were calculated within both optimization methods. Then, in the second part, a consensus value for each constant was calculated, based on the results of both the *fmincon* and heuristic method. Finally, from this constant in fixed conditions, full temperature dependencies were calculated. The matrix of cross-validation (Fig 5) between series was the same for both methods. The simple algorithm based on taking always only one constant, and replacing it with its equivalent from another series. For example, after the first initial calculation with all default values, in the second iteration, the $k_{m,pro}$, was replaced with its value from the second series of optimization (where reactors were independent), then in 3rd series $k_{m,ac}$ was replaced, while $k_{m,pro}$ value was restored to default for this experiment, and so on. A total number of 14 different combinations were calculated.

The results of the calculations consistent with the above-mentioned matrix were required to determine the impact of selected changes on an eventual increase of relative error. For reference trials (iteration 1 and 8) the error is assumed as 0%, and the change for any other iterations is calculated in the respect of the corresponding test from these two. For example, the error for iterations 2–7 will be calculated as follows:

$$\overline{\delta x} = \frac{1}{n} \sum_{i=1}^n \frac{|x_{IT1,n} - x_{ITy,n}|}{x_{IT1,n}} * 100\% \quad (5)$$

where $x_{IT1,n}$ means the value of the corresponding parameter for measuring point n in the first iteration and $x_{ITy,n}$ in the tested configuration. For iterations 9 to 14, the scheme will be the same, only the reference will be changed to $x_{IT8,n}$. No differences between the automatic and heuristic method in this matter are present. The summary for both algorithms is presented in Table 6.

As it may be noticed, there is no simple pattern in the impact of the method, on the result of cross-validation. In the case of $k_{m,pro}$, in all optimized configurations, the manually adjusted value led to a lower error. However, for $k_{m,ac}$, the impact is more complicated—in thermophilic reactors, manually adjusted values were better, why for mesophilic this from automatic algorithm gave lower mean relative error. An opposite tendency can be denoted in the case of hydrolysis constant, wherein both thermophilic reactors, the automatic algorithm gives better value, while for mesophilic tanks, manually adjusted values led to overall lower mean relative error.

Basing on these dependencies it is impossible to arbitrary chose a better method. Both in the step of raw optimization, as well as in the cross-validation, manual and automatic method have their advantages. Therefore, both will be considered as an input to the final results. The next step was to calculate the values of constant, for every method

Fig. 5 Cross-validation matrix for determination of unified constants values within the selected method

Constant name	$k_{m,pro}$ (R1)	$k_{m,ac}$ (R1)	k_{dis} (R1)	$k_{m,pro}$ (R2)	$k_{m,ac}$ (R2)	k_{dis} (R2)
Series 1						
1						
2						
3						
4						
5						
6						
7						
Series 2						
8						
9						
10						
11						
12						
13						
14						
Series 1, Reactor 1 value			Series 1, Reactor 2 value			
Series 2, Reactor 1 value			Series 2, Reactor 2 value			

Table 6 The final results for cross-validation for both manual and automatic methods

Modified constant	Manual method MRE [%]	Automatic method MRE [%]	Trend	Modified constant	Manual method MRE [%]	Automatic method MRE [%]	Trend
Series 1							
Thermophilic reactor				Mesophilic reactor			
$k_{m;pro}$	69.18	70.05	↑	$k_{m;pro}$	51.76	65.66	↑
$k_{m;ac}$	2.18	41.03	↑	$k_{m;ac}$	23.79	11.87	↓
k_{dis}	139.15	129.45	↓	k_{dis}	479.09	820.61	↑
Series 2							
Thermophilic reactor				Mesophilic reactor			
$k_{m;pro}$	118.14	131.58	↑	$k_{m;pro}$	269.66	651.72	↑
$k_{m;ac}$	1.80	24.91	↑	$k_{m;ac}$	23.71	16.28	↓
k_{dis}	51.81	51.27	↓	k_{dis}	92.81	94.34	↑

separately, basing on the data from cross-validation. To do so, weighted arithmetic mean will be utilized, where the mean relative error will be treated as a weight. The equation below summarizes the idea:

$$\bar{k}_x = \frac{k_{x1,n} * \bar{\delta}_{1,n} + k_{x2,n} * \bar{\delta}_{2,n}}{\bar{\delta}_{1,n} + \bar{\delta}_{2,n}} \quad (6)$$

The $k_{x1,n}$ and $k_{x2,n}$ mean the value of constant from the first and second series of optimization, respectively. The $\bar{\delta}_{1,n}$ and $\bar{\delta}_{2,n}$ are mean relative errors connected with them. The results for both considered techniques of optimization are summarized in Table 7, additionally mean value for further calculation is determined.

However, the mean value of constant, calculated based on two series of experimental data, and two independent methods of optimization, can be treated like a kind of unification, this is still correct only for two fixed points of time. There is still a need to determine full temperature

dependencies, to let the model works in any condition programmed by the user. In general, the relation between temperature, and the value of discussed constants can be approximated to exponential [25]. Referring it to the initial value of constant, for one of the known temperature points (T_0), leads to the following conclusions:

$$k_{x,T} = k_{x,T0} * e^{\theta(T-T_0)} \quad (7)$$

Where $k_{x,Tn}$ describes the value of the selected constant in any temperature n , and $k_{x,T0}$ means its value in reference temperature (average value from both methods). The θ is the temperature-dependence coefficient that is individual for any of the considered constants. It can be calculated, by rewriting mentioned above equation, to include the value of constant for both temperatures considered in the optimization:

$$\theta = \frac{\ln \frac{k_{x,T1}}{k_{x,T0}}}{(T_1 - T_0)} \quad (8)$$

where T_0 refers to mesophilic conditions (35 °C) and T_1 to thermophilic (55 °C). According to the adopted nomenclature, all further calculations will be related to the calculated values for T_0 . The data are summarized in Table 8.

Table 7 Final constants from both methods of optimization

Unit	Constant name	Manually adjusted value (day ⁻¹)	Automatically adjusted value (day ⁻¹)	Mean Value (day ⁻¹)
R1	$k_{m,pro}$	11.760	12.505	12.133
	$k_{m,ac}$	15.295	15.825	15.560
	k_{dis}	0.360	0.198	0.279
R2	$k_{m,pro}$	7.964	6.419	7.192
	$k_{m,ac}$	12.354	16.181	14.267
	k_{dis}	1.049	1.162	1.105

Table 8 The summary of final optimization results

	Referential constant value for T_0	θ (°C ⁻¹)
$\bar{k}_{m;pro}$ [day ⁻¹]	7.192	0.02615
$\bar{k}_{m;ac}$ [day ⁻¹]	14.267	0.00434
$\bar{k}_{dis}$ [day ⁻¹]	1.105	− 0.06879

3.3 Comparison of methods

The determination if fully impartial, an automatic algorithm is better or worse than a manual, subjective adjustment by a scientist is not a trivial issue. As our research shows, there is no simple answer to a question, which method provides higher accuracy, better results, preferable fitting to experimental results. Certainly, it can be said that both approaches provide results, which are convergent in all the most important aspects. What's more, both give constant values in acceptable order of magnitude, leading to satisfactory fitting between experimental and simulated data.

However, also significant differences need to be noticed. For example, for most constant, the temperature tendency was the same, independently of the method. Nevertheless, for the constant that describes uptake rate for acetate degrading organisms, the manual method suggested increase with temperature, while automatic algorithm shows values leading to the conclusion that it is almost independent, or slightly decreasing.

Another interesting observation is related to the nature of both methods. The TSE is always lower for automatic algorithms, but it does not mean directly that this method is just better. This is the best visible for the second series of optimization, where the input flow of biomass is no constant (Fig. 4). The overall error for the first reactor in this series is 56.3 percentage points lower for automatic method, however, this is achieved mostly by very precise fitting in the first part of the simulation, while in days 30–50 the accuracy is significantly dropping. The manual method provides less accurate results in synthetic tests, but the fit is more regular throughout the entire measurement points. This can have significant consequences on the results of further, more complicated simulations.

On the other hand, the errors, connected with absolute nature of optimization function in automatic method, can be reduced by utilizing more complicated queries. In this research, a comparatively simple form of a sum of relative errors between experimental, and predicted values for all known points, was used. The more complex objective function, utilizing weight factors or other solutions forcing more equal fitting in a whole range of simulations could lead to significantly different results.

Irrespectively of presented above benefits and disadvantages, it should be kept in mind, that auto-optimization is a much faster process, as the computer can test hundreds of possible combinations within minutes. What's more, they can be shifted in a very complex pattern, which potentially may be more efficient than a simple manual, consequent order presented in this article. Ultimately, it can find solutions, which would be omitted by a researcher in manual adjustment.

In the presence of the above-mentioned facts, we suggested a cross-validation based method, which in the final

step connected both approaches. This will be a compromise between the absolute accuracy, and scientific sense, between impartiality and the experience. This denouement seems to be the one equitable in this draw between scientists and computers. The averaged value of constant, and the temperature-dependence coefficient, received by this methodology can be used in further simulations and researches.

4 Conclusions

Our recommendation is to combine the advantages of both approaches always when it is possible. The automatic method should be carried out first, as this approach is generally faster and does not require a detailed understanding of the process basics. If the desired accuracy and uniformity of result is not achieved, the manual method can be utilized independently, and eventually combined with previous calculations. As our results show, while the automatic algorithm indicates higher overall accuracy (in the meaning of the TSE), it can neglect some subtle correlations which can be recognized by a scientist. Nevertheless, in some cases where manual optimization is not applicable—for example, where there is a high number of parameters to optimize, the automatic method based on the *GlobalSearch* algorithm can provide good and acceptably accurate results, without external support of manual adjustment. Both of the examined methods are efficient for the optimization of bioprocess models.

Funding information This work was supported by The National Centre for Research and Development (Poland) under the program BIOSTRATEG (No. BIOSTRATEG2/298357/8/NCBR/2016) and co-financed by a statutory activity for the year 2019, from the Polish Ministry of Science and Higher Education for the faculty of Chemistry of Wrocław University of Science and Technology (grant no. 0049U/0073/19).

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

References

1. Bedoić R, Čuček L, Čosić B, Krajnc D, Smoljanić G, Kravanja Z, Ljubaš D, Pukšec T, Duić N (2019) Green biomass to biogas

- a study on anaerobic digestion of residue grass. *J Clean Prod* 213:700–709. <https://doi.org/10.1016/j.jclepro.2018.12.224>
2. Kalogirou SA (2001) Artificial neural networks in renewable energy systems applications: a review. *Renew Sust Energ Rev* 5(4):373–401. [https://doi.org/10.1016/S1364-0321\(01\)00006-5](https://doi.org/10.1016/S1364-0321(01)00006-5)
3. Ringkjøb HK, Haugan PM, Solbrenke IM (2018) A review of modelling tools for energy and electricity systems with large shares of variable renewables. *Renew Sust Energ Rev* 96:440–459. <https://doi.org/10.1016/j.rser.2018.08.002>
4. Razm S, Nickel S, Sahebi H (2019) A multi-objective mathematical model to redesign of global sustainable bioenergy supply network. *Comput Chem Eng* 128:1–20. <https://doi.org/10.1016/j.compchemeng.2019.05.032>
5. Ersahin ME (2018) Modeling the dynamic performance of full-scale anaerobic primary sludge digester using anaerobic digestion model no. 1 (adml). *Bioprocess Biosyst Eng* 41(10):1539–1545. <https://doi.org/10.1007/s00449-018-1981-5>
6. Marion G, Lawson G, Marion G (2008) An introduction to mathematical modelling. University of Bristol, Bristol
7. Sarkar D, Modak JM (2004) Optimization of fed-batch bioreactors using genetic algorithm: multiple control variables. *Comput Chem Eng* 28(5):789–798. <https://doi.org/10.1016/j.compchemeng.2004.02.018>
8. Tolson BA, Shoemaker CA (2007) Dynamically dimensioned search algorithm for computationally efficient watershed model calibration. *Water Resour Res* 43(1) <https://doi.org/10.1029/2005WR004723>. [Online; accessed 2019-08-06]
9. Gendreau M, Potvin JY (2005) Metaheuristics in combinatorial optimization. *Ann Oper Res* 140(1):189–213. <https://doi.org/10.1007/s10479-005-3971-7>
10. Martí R, Reinelt G (2011) The linear ordering problem: exact and heuristic methods in combinatorial optimization. Applied mathematical sciences. Springer, New York. oCLC: ocn668941895
11. Inooka H, Koitabashi T (1990) Experimental studies of manual optimization in control tasks. *IEEE Control Syst Mag* 10(5):20–23. <https://doi.org/10.1109/37.60418>
12. Vrugt JA, Gupta HV, Bastidas LA, Bouten W, Sorooshian S (2003) Effective and efficient algorithm for multiobjective optimization of hydrologic models. *Water Resour Res* 39(8) <https://doi.org/10.1029/2002WR001746>. [Online; accessed 2019-08-06]
13. Dragoi EN, Curteanu S, Galaction AI, Cascaval D (2013) Optimization methodology based on neural networks and self-adaptive differential evolution algorithm applied to an aerobic fermentation process. *Appl Soft Comput* 13(1):222–238. <https://doi.org/10.1016/j.asoc.2012.08.004>
14. Turgut MS, Turgut OE (2020) Global best-guided oppositional algorithm for solving multidimensional optimization problems. *Eng Comput* 36(1):43–73. <https://doi.org/10.1007/s00366-018-0684-5>
15. Rios LM, Sahinidis NV (2013) Derivative-free optimization: a review of algorithms and comparison of software implementations. *J Glob Optim* 56(3):1247–1293. <https://doi.org/10.1007/s10898-012-9951-y>
16. Oommen BJ, Rueda LG (2005) A formal analysis of why heuristic functions work. *Artif Intell* 164(1):1–22. <https://doi.org/10.1016/j.artint.2002.02.001>
17. Riau V, De la Rubia MA, Pérez M (2010) Temperature-phased anaerobic digestion (tpad) to obtain class a biosolids: a semi-continuous study. *Bioresour Technol* 101(8):2706–2712. <https://doi.org/10.1016/j.biortech.2009.11.101>
18. Sung S, Santha H (2003) Performance of temperature-phased anaerobic digestion (tpad) system treating dairy cattle wastes. *Water Res* 37(7):1628–1636
19. Mc W, Kw S, Zhang Y (2006) Influence of temperature fluctuation on thermophilic anaerobic digestion of municipal organic solid waste. *J Zhejiang Univ Sci B* 7(3):180–185. <https://doi.org/10.1631/jzus.2006.B0180>
20. Merlin Christy P, Gopinath LR, Divya D (2014) A review on anaerobic decomposition and enhancement of biogas production through enzymes and microorganisms. *Renew Sust Energ Rev* 34:167–173. <https://doi.org/10.1016/j.rser.2014.03.010>
21. Budzianowski WM, Postawa K (2017) Renewable energy from biogas with reduced carbon dioxide footprint: implications of applying different plant configurations and operating pressures. *Renew Sustain Energ Rev* 68 Part 2:852–868. <https://doi.org/10.1016/j.rser.2016.05.076>
22. Postawa K (2018) Novel solutions in modeling of anaerobic digestion process - two-phase ad models development and comparison. *Int J Chem React Eng* 16(8) <https://doi.org/10.1515/ijcre-2017-0139>. <https://www.degruyter.com/view/j/ijcre.2018.16.issue-8/ijcre-2017-0139/ijcre-2017-0139.xml>, [Online; accessed 2019-10-03]
23. Blumensaat F, Keller J (2005) Modelling of two-stage anaerobic digestion using the iwa anaerobic digestion model no. 1 (adml). *Water Res* 39(1):171–183. <https://doi.org/10.1016/j.watres.2004.07.024>
24. Lee MY, Suh CW, Ahn YT, Shin HS (2009) Variation of adm1 by using temperature-phased anaerobic digestion (tpad) operation. *Bioresour Technol* 100(11):2816–2822. <https://doi.org/10.1016/j.biortech.2008.12.025>
25. Siegrist H, Vogt D, Garcia-Heras JL, Gujer W (2002) Mathematical model for meso- and thermophilic anaerobic sewage sludge digestion. *Environ Sci Technol* 36(5):1113–1123. <https://doi.org/10.1021/es010139p>
26. Ashino R, Nagase M, Vaillancourt R (2000) Behind and beyond the matlab ode suite. *Comput Mathe Appl* 40(4):491–512. [https://doi.org/10.1016/S0898-1221\(00\)00175-9](https://doi.org/10.1016/S0898-1221(00)00175-9)
27. Postawa K, Szczygieł J, Kułczyński M (2020) A comprehensive comparison of ode solvers for biochemical problems. *Renew Energ* 156:624–633. <https://doi.org/10.1016/j.renene.2020.04.089>
28. Banafsha A, Vinay KT, Priyanka AAK, Kazmi AA (2020) Optimization of process parameters for enhanced biogas yield from anaerobic co-digestion of ofmsw and bio-solids. *Biomass Conversion and Biorefinery* <https://doi.org/10.1007/s13399-020-00919-3>
29. Postawa K, Szczygieł J, Kułczyński M (2020) Heuristic methods in optimization of selected parameters of two-phase anaerobic digestion (tpad) model. *Fuel* 281:118257. <https://doi.org/10.1016/j.fuel.2020.118257>
30. Agnarsson J, Sunde M, Ermilova I (2013) Parallel optimization in matlab project report. Tech. rep., Uppsala University. <https://doi.org/10.13140/rg.2.2.28603.87840>, [Online; accessed 2019-08-06]
31. Villaverde AF, Fröhlich F, Weindl D, Hasenauer J, Banga JR (2019) Benchmarking optimization methods for parameter estimation in large kinetic models. *Bioinformatics* 35(5):830–838. <https://doi.org/10.1093/bioinformatics/bty736>
32. Ugray Z, Lasdon L, Plummer J, Glover F, Kelly J, Martí R (2007) Scatter search and local nlp solvers: a multistart framework for global optimization. *Inform J Comput* 19(3):328–340. <https://doi.org/10.1287/ijoc.1060.0175>
33. Krityakierne T, Shoemaker CA (2017) Soms: surrogate multistart algorithm for use with nonlinear programming for global optimization. *Int Trans Oper Res* 24(5):1139–1172. <https://doi.org/10.1111/itor.12190>