An Augmented Lagrangian Pattern Search Method for Optimal WWTP Designs

[1]I. A. C. P. ESPÍRITO SANTO, [1]E. M. G. P. FERNANDES, [1]M. M. ARAÚJO, [2]E. C. FERREIRA
[1]Systems and Production Department, [2]Centre of Biological Engineering
Minho University
Campus de Gualtar, 4710-057 Braga
PORTUGAL

Abstract: - This paper describes a derivative-free method to solve a non-linear constrained optimization problem that arises from the mathematical formulation of an activated sludge system of a wastewater treatment plant (WWTP), in which the objective is to minimize the investment and operation costs. The method relies on an augmented Lagrangian function to penalize infeasible solutions and uses the original Hooke and Jeeves pattern search method for solving bound constrained subproblems. The experimental results are very promising and give economically attractive WWTP designs.

Key-Words: - Activated Sludge System, Secondary Settler, Pattern Search Method, Augmented Lagrangian.

1 Introduction

With the growing need to decrease the installation and operation costs of a wastewater treatment plant (WWTP), the search for a minimum cost design is becoming more and more challenging. To be able to achieve an optimal WWTP design, an optimization procedure that considers the mathematical modeling of an activated sludge system and the definition of a cost function is carried out.

Besides the densely populated and industrial regions, there are small and poor regions in the north of Portugal that produce high quality wines and have significant effluent variations in terms of amount of pollution and flow, during the vintage season. For these reasons, it is crucial to reduce, as much as possible, the costs associated with the design and operation of WWTPs, in such a way that the environmental law is accomplished.

A typical WWTP is usually defined by a primary treatment, a secondary treatment and in some cases a tertiary treatment. The cost of the primary treatment, usually a physical process that aims to eliminate the gross solids and grease, does not depend on the characteristics of the wastewater as significantly as the secondary treatment. So, the primary treatment is not included in the optimization procedure. However, its impact on the cost of the secondary treatment is analyzed. The secondary treatment is an activated sludge system, since this is the most common treatment process in WWTPs. This system consists of an aeration tank and a secondary settler. In the planning and design of a wastewater treatment plant the role played by the secondary settler is, most of the time, underestimated.

The most common models in literature to describe the aeration tank are the ASM kind models, in particular the ASM1 model [8]. To describe the secondary settler, the ATV [3] and the double exponential (DE) [13] models have been used in the past. The ATV model is usually used as a design procedure to new WWTPs. It is based on empirical equations obtained by experiments and does not contain any solid balances, although it contemplates peak wet weather flow (PWWF) events. The DE model is the most widely used in simulations and it produces results very close to reality. However, as it does not provide extra sedimentation area needed during PWWF events, the resulting design has to consider the use of security factors, many times inadequate.

The optimization procedures found in the literature involving secondary settlers are based on very simple models [14] or on the ATV model [1]. To the best of our knowledge, and until last year, there have been no optimization attempts using the DE model. However, simulation procedures have been carried out with the DE model aiming to find the best combination of the decision variables to achieve the minimum cost design, choosing one among some possible designs [11, 12].

Recently real optimization procedures using the DE model were carried out in order to obtain the best optimal WWTP design in the sense that a minimum cost is attained [6]. This work compares an innovative model that combines the ATV and DE models, with the two traditional models used separately. Numerical experiments show that the combined model provided the most equilibrated WWTP design. Further, the three resulting designs were introduced in the GPS-X simulator [15] and a stress condition, based on a PWWF value of about 5 times the normal flow, was imposed. Only the combined model was able to support this adverse condition maintaining the quality of the effluent under the values imposed by the portuguese law.

In the presence of non-smooth functions as encountered in the mathematical equations involved in the DE model, a derivative-free optimization technique is the most appropriate. Thus, a pattern search algorithm is proposed. This algorithm relies on an augmented Lagrangian function in order to obtain a solution that satisfies the equality and inequality constraints of the problem. The algorithm was coded in the C programming language and incorporates an interface to AMPL [7] to be able to read the problem coded in AMPL. The experimental results are very promising and give economically attractive WWTP designs.

This paper is organized as follows. Section 2 provides a brief discussion of the modeling process of the activated sludge system and Section 3 proposes the cost function. The augmented Lagrangian pattern search algorithm is discussed in Section 4 and Section 5 contains the results. Some conclusions are drawn in Section 6.

2 Modeling process of the activated sludge system

To model the aeration tank, where the biological reactions take place, the ASM1 model is used. The tank is considered a CSTR in steady state. As our purpose is to make cost predictions in a long term basis, it is reasonable to do so. The balances around this unit define some of the constraints of the mathematical model.

To model the secondary settler, which plays a crucial role in the wastewater treatment, a combination of the two traditional models ATV and DE is used. Previous work [4,6] on this matter shows that the model is prepared to overcome PWWF events without over dimensioning and overcomes the limitations and powers the advantages of the other two, when used separately. The combination proved to be robust in all situations.

The system behaviour, in terms of concentration and flows, may be predicted by balances. In order to achieve a consistent system, these balances must be done around the entire system and not only around each unitary process. They were done to the suspended matter, dissolved matter and flows.

In a real system, some state variables are, most of the time, not available for evaluation. Thus, another important group of constraints in the mathematical model is a set of linear equalities that define composite variables.

Some system variables definitions should be added to the model in order to define the system correctly. These definitions include the sludge retention time, the recycle rate, the hydraulic retention time, the recycle rate in a PWWF event, the recycle flow rate in a PWWF event and the maximum overflow rate.

All the variables in the model must be nonnegative, although more restricted bounds are imposed to some of them due to operational consistencies. These conditions define a set of simple bounds on the variables. For example, the dissolved oxygen has to be always greater or equal to 2 mg/L. For a more thorough analysis of all the equations the reader is referred to [4,5].

Finally, the quality of the effluent has to be imposed. The quality constraints are usually derived from law restrictions. The most used are related with limits in the *COD*, *N* and *TSS* at the effluent.

The obtained mathematical model has 64 parameters, 115 variables and 105 constraints, where 67 are nonlinear equalities, 37 are linear equalities and there is only 1 nonlinear inequality. 104 variables are bounded below and 11 are bounded below and above.

3 The cost function

The cost function is used to describe the installation and operation costs of a WWTP, in a way that reflects the behaviour of each unitary process. In the present study, only the aeration tank and the secondary settler are considered.

The basic structure of the cost function, based on the work done by Tyteca *et al.* [14], is $C = aZ^b$, where C represents the cost and Z the variable that most influences the design of the unitary process. The parameters a and b are estimated according to the costs associated with the unit under study and depend on the local conditions where the WWTP is being built. A least squares technique considering real data collected from a portuguese WWTP company builder was used. At the present, the collected data come from a set of WWTPs in design, therefore no operation data are available. However, from the experience of the company, it is known that the maintenance expenses for the civil construction are around 1% during the first 10 years and around 2% in the next 10. To the electromechanical components, the maintenance expenses are negligible, but all the material is usually replaced after 10 years. The energy cost is directly related with the air flow (G_S).

The total cost (TC) function is given by the sum of the operation and investment costs, in a present value basis, from the two considered units (the aeration tank and the secondary settler) and is given by

$$TC = 174.2 V_a^{1.07} + 12487\, G_S^{0.62} + 114.8 G_S + 955.5 A_s^{0.97} + 41.3 (A_s h)^{1.07} \qquad (1)$$

where V_a is the aeration tank volume, A_s and h are the secondary settler surface area and depth, respectively. Function (1) is the objective function of the mathematical problem.

4 An augmented Lagrangian pattern search algorithm

When the double exponential model is incorporated in the formulation of the optimization problem, some of the constraints are non-smooth functions. Thus, any optimization method that relies on derivative information may have convergence problems. To address this issue, a derivative-free algorithm based on an augmented Lagrangian paradigm is proposed.

The general form of a nonlinear constrained programming problem can be expressed as follows:

$$\underset{x \in \Psi \subset \Re^n}{\text{minimize}} \ f(x) \qquad (2)$$

where x is the vector that contains the decision variables of the problem, n is the number of variables, $\Psi = \{x : b(x) = 0, g(x) \le 0, l \le x \le u\}$ denotes the feasible region, $f : \Re^n \to \Re$ is the objective function, $b(x) = 0$ are the equality constraints and $g(x) \le 0$ are the inequality constraints. Some of the bounds l, u may not exist, meaning that $l_j = -\infty, u_j = +\infty$, for some $j \in \{1, \ldots, n\}$. Here we consider $b : \Re^n \to \Re^m$ and $g : \Re^n \to \Re^p$. To solve problem (2), a penalty multiplier technique is proposed.

4.1 The penalty multiplier technique

A penalty multiplier technique solves a sequence of very simple subproblems whose objective function penalizes all or some of the constraints violation. These functions are known as penalty functions. The following augmented Lagrangian penalty function

$$\Phi(x; \lambda, \delta, \mu) = f(x) + \sum_{i=1}^{m} \lambda_i b_i(x) + \frac{1}{2\mu} \sum_{i=1}^{m} b_i(x)^2 + \frac{\mu}{2} \sum_{i=1}^{p} \left(\max\left(0, \delta_i + \frac{g_i(x)}{\mu}\right)^2 - \delta_i^2 \right)$$

is used, where μ is a positive penalty parameter, $\lambda = (\lambda_1, \ldots, \lambda_m)^T$ and $\delta = (\delta_1, \ldots, \delta_p)^T$ are the Lagrange multipliers vectors associated with the equality and inequality constraints respectively [4]. The purpose of function $\Phi(x;.)$ is to penalize solutions that violate the equality and inequality constraints only. The simple bounds $l \le x \le u$ are left unchanged. Thus, the corresponding subproblems are

$$\underset{x \in \Omega \subset \Re^n}{\text{minimize}} \quad \Phi(x; \lambda^j, \delta^j, \mu^j) \qquad (3)$$

where the feasible region is now defined as $\Omega = \{x : l \le x \le u\}$. The solution of (3) for each set of fixed λ^j, δ^j and μ^j, gives an approximation to the solution of (2). Denote this approximation by x^{j+1}. Here, the index j is the iteration counter. As $j \to \infty$ and $\mu^j \to 0$, the solutions of subproblems (3) converge to the solution of (2). We refer to Bertsekas [2] for details. The Lagrange multipliers (λ^j, δ^j) are estimated in this iterative process according to the following updating formulae

$$\lambda_i^{j+1} = \lambda_i^j + \frac{b_i(x^{j+1})}{\mu^j}, i = 1, \ldots, m \qquad (4)$$

and

$$\delta_i^{j+1} = \max\left(0, \delta_i^j + \frac{g_i(x^{j+1})}{\mu^j}\right), i = 1, \ldots, p. \qquad (5)$$

This outer iterative process is halted when an approximation x^{j+1} that satisfies the first-order optimality conditions for problem (2) is found. These conditions define the convergence criteria of the algorithm and involve appropriate measures of feasibility, complementarity and optimality. The traditional augmented Lagrangian methods are locally convergent if the subproblems (3) are solved according to a certain tolerance, for sufficiently small values of the penalty parameter. To ensure global convergence, the penalty parameter must be driven to zero and the Lagrange multipliers estimates must have a reasonable behaviour.

In general, the choice of the penalty parameter μ^j has a significant impact on the performance of this outer iterative process. If μ is reduced too rapidly, the method may become too slow. Thus, a reduction of the parameter μ^j is carried out only if an appropriate feasibility and complementarity measure, here denoted for simplicity by $M_{f/c}$, computed at the new approximation, by solving (3), is not too small. This procedure aims to force a reduction in $M_{f/c}$ for the subproblem of the next iteration. On the other hand, if $M_{f/c}$ is small enough, according to a given tolerance (for instance, $M_{f/c} \le \eta^j$), the penalty parameter is maintained and the Lagrange multipliers λ^j are updated using (4). The updating of the multipliers δ^j (see (5)) is done in both cases and precedes them as the new estimates are required to evaluate $M_{f/c}$. The algorithm for solving problem (2) is the following.

Algorithm 1: (augmented Lagrangian multiplier method)
Initialize variables and algorithm parameters
While convergence criteria are not satisfied **do**
Inner iterative process: approximately solve subproblem (3) according to a tolerance ε^j
Update multipliers δ^j
  if $M_{f/c} \le \eta^j$ **then**
    Update multipliers λ^j
    Update_1 algorithm parameters
  else
    Reduce μ^j
    Update_2 algorithm parameters
  end if
$j \leftarrow j+1$
end while

Procedures Update_1 and Update_2 are slightly different. They are used to update, although in a different manner, the error tolerances ε^j and η^j required in the algorithm [4].

4.2 The Hooke and Jeeves pattern search

We now show how subproblem (3) is solved by a derivative-free method. We chose to implement a pattern search method [10]. In this method a series of exploratory moves about the current iterate, x^k, is conducted in order to find a new iterate, $x^{k+1} = x^k + \Delta^k s^k$, with a lower objective function value, where Δ^k represents the step length and s^k determines the direction of the step. The step $\Delta^k s^k$ is computed by the Hooke and Jeeves exploratory moves [9]. The index k is the iteration counter in this inner iterative process. The corresponding algorithm is as follows.

Algorithm 2: (pattern search method)
Initialize with $x^k \in \Omega$
while termination criterion is not satisfied **do**
Compute step $\Delta^k s^k$ using Hooke and Jeeves exploratory moves such that $x^k + \Delta^k s^k \in \Omega$
  if $\Phi(x^k;.) - \Phi(x^k + \Delta^k s^k;.) > 0$ **then**
  $x^{k+1} = x^k + \Delta^k s^k$
  else
  $x^{k+1} = x^k$
  end if
Update Δ^k and s^k
$k \leftarrow k+1$
end while

The exploratory moves to produce $\Delta^k s^k$ and the updating of Δ^k and s^k define a particular pattern search method and their choices are crucial to the success of the algorithm. An iteration with $\Phi(x^k;.) - \Phi(x^k + \Delta^k s^k;.) > 0$ is considered successful; otherwise, the iteration is unsuccessful. When an iteration is successful, the step length is not allowed to decrease, while in an unsuccessful iteration, Δ^k should be reduced.

The condition that halts this inner iterative process, here designated as termination criterion, is based on the size of the step length Δ^k. If condition $\Delta^k \le \varepsilon^j$ is true, the algorithm cannot make progress and should be terminated. The parameter ε^j is defined in the outer iterative process.

The Hooke and Jeeves exploratory moves differ from the traditional coordinate search, as they also include a 'pattern step' in an attempt to accelerate the process.

If the previous iteration was successful ($\Phi(x^{k-1};.) - \Phi(x^k;.) > 0$), the current iteration carries out a coordinate search about a speculative iterate $x^k + (x^k - x^{k-1})$, instead of x^k. This is the 'pattern step'. The idea is to investigate a possible progress along the promising direction $x^k - x^{k-1}$. For example, if $k > 0$ and $x^k \ne x^{k-1}$, this search takes the step $x^k - x^{k-1}$ from x^k. The function is evaluated at this new point that is temporarily accepted, even if $\Phi(x^k + (x^k - x^{k-1});.) \ge \Phi(x^k;.)$. The Hooke and Jeeves algorithm then carries a coordinate search about the temporary iterate $x^k + (x^k - x^{k-1})$. If this coordinate search is successful, the returned point is accepted as the new iterate x^{k+1}; otherwise, the 'pattern step' is rejected and the method is reduced to coordinate search about x^k.

We note that the Hooke and Jeeves exploratory moves must return a feasible iterate (recall (3)). If the new accepted iterate is infeasible, the point is reflected into the set Ω as follows:

$$x_i = \begin{cases} l_i + (l_i - x_i) & \text{se} \quad x_i < l_i \\ x_i & \text{se} \quad l_i \le x_i \le u_i \\ u_i - (x_i - u_i) & \text{se} \quad x_i > u_i \end{cases}$$

for all $i = 1,...,n$. If the iterate remains infeasible, then $x_i = (l_i + u_i)/2$.

5 Results and discussion

The following results were obtained using the data provided by a WWTP company builder, corresponding to four small towns in the interior north of Portugal – Alijó, Murça, Sabrosa and Sanfins do Douro. These are medium scale WWTPs, being the first slightly bigger than the others. The data collected from the four small towns are shown in Table 1.

Three experiences were carried out with the four WWTPs using the proposed method. First, no primary treatment is used. Then 50% and 70% primary treatment efficiencies are included.

Table 1 Data collected from the four small towns.

	Location of the WWTPs			
	Alijó	Murça	Sabrosa	Sanfins
population equivalent	6850	3850	2750	3100
influent flow (m^3/day)	1050	885	467.5	530
peak flow (m^3/h)	108	86.4	48.6	54
COD (Kg/m^3)	2000	1750	1250	1250
TSS (Kg/m^3)	750	660	610	610

Tables 2 - 5 show the optimal values for some of the decision variables, namely, the aeration tank volume, the air flow, the area and depth of the secondary settler, and *COD*, *TSS* and *N* at the effluent. The total costs (*TC*), in millions of euros, for the three cases, no primary treatment, 50% and 70% efficiencies in the primary treatment, are also reported.

We observe that the presence of the primary treatment causes a reduction in the total cost. As the primary treatment efficiency gets better, the cost of the secondary treatment decreases. The exception is the Sanfins do Douro WWTP, where the highest cost is obtained when the primary treatment efficiency is 70%. This is due to the extremely high value of the settling depth in the obtained design. We feel that this could be solved by imposing an upper bound on the settling depth.

Table 2 Results for the Alijó WWTP.

efficiency	0	50	70
V_a	2632	1155	1308
G_S	108	101	100
A_s	654	600	659
h	23.8	16.8	9.8
COD	41.1	22.8	43.5
TSS	22.8	19.9	16.4
N	6.4	8.2	6.4
TC	2.79	1.81	1.60

Table 3 Results for the Murça WWTP.

efficiency	0	50	70
V_a	1599	946	748
G_S	102	101	159
A_s	746	933	715
h	16.5	11.1	7.2
COD	12.8	40.0	63.7
TSS	10.4	20.1	10.6
N	3.1	6.8	6.8
TC	2.24	2.00	1.44

Table 4 Results for the Sabrosa WWTP.

efficiency	0	50	70
V_a	1344	1447	706
G_S	100	100	100
A_s	707	768	987
h	17.5	14.5	7.9
COD	88.7	63.2	63.7
TSS	20.1	20.4	12.6
N	9.1	8.6	8.1
TC	2.14	2.11	1.76

Table 5 Results for the Sanfins do Douro WWTP.

efficiency	0	50	70
V_a	1263	885	801
G_S	103	100	182
A_s	703	736	998
h	14.0	8.8	24.7
COD	51.7	28.7	87.8
TSS	24.0	13.8	16.5
N	10.5	9.5	9.4
TC	1.90	1.53	3.36

In all cases, the *COD*, *TSS* and *N* law limits (125, 35 and 15, respectively) were never achieved (see rows 6 - 8 in Tables 2 - 5), meaning that the obtained solutions are robust. We observe that the presence of the primary treatment reduces the aeration tank volume although the air flow is maintained at the same range of values.

6 Conclusions

This paper presents a new derivative-free optimization algorithm for solving a problem that is based on a WWTP design, in which the objective is to minimize the costs associated with the installation and operation of an activated sludge system.

Four real WWTPs were analyzed and the mathematical modeling of the activated sludge system was carried out using the ASM1 model for the aeration tank and a combination of the ATV and DE models for the settling tank. The cost function was obtained using real data provided by a portuguese WWTP company builder. This mathematical programming problem was coded in AMPL and solved using an augmented Lagrangian pattern search method. This method does not require any derivative information being the most appropriate when the mathematical problem involves non-smooth functions. The obtained optimal WWTP designs are economically attractive. The impact of the primary treatment on the total cost is also analyzed.

Future developments will focus on the optimization of a WWTP in operation in which only the operational parameters are considered as decision variables. The dynamic equations that appear in the mathematical formulation of the problem will be solved by a numerical integrator.

References:

[1] P. N. C. M. Afonso. Modelação Matemática de Reactores Biológicos no Tratamento Terciário de Efluentes. *PhD Thesis* (in portuguese), Universidade do Porto, Portugal, 2001.

[2] D. P. Bertsekas. *Constrained Optimization and Lagrange Multiplier Methods*, Athena Scientific, 1996

[3] G. A. Ekama, J. L. Barnard, F. W Günthert, P. Krebs, J. A. McCorquodale, D. S. Parker and E. J Wahlberg. Secondary Settling Tanks: Theory, Modeling, Design and Operation, *Technical Report 6. IAWQ - International Association on Water Quality*, 1978.

[4] I. A. C. P. Espírito Santo. Desenho Óptimo de Estações de Águas Residuais Através da Modelação de Funções Custo. *PhD Thesis* (in portuguese), Universidade do Minho, Portugal, 2007.

[5] I. A. C. P. Espírito Santo, E. M. G. P. Fernandes, M. M. Araújo and E. C. Ferreira. How Wastewater Processes can be Optimized Using LOQO. *Lecture Notes in Economics and Mathematical Systems,* A. Seeger (ed.), Springer-Verlag, No. 563, 2006, pp 435-455.

[6] I. A. C. P. Espírito Santo, E. M. G. P. Fernandes, M. M. Araújo and E. C. Ferreira. On the Secondary Settler Models Robustness by Simulation. *WSEAS Transactions on Information Science and Applications,* Issue 12, Vol. 3, 2006, pp. 2323-2330.

[7]R. Fourer, D. M. Gay and B. Kernighan. A modeling language for mathematical programming, *Management Science*, Vol. 36, No. 5, 1990, pp. 519-554.

[8] M. Henze, C. P. L. Grady Jr, G. V. R. Marais and T. Matsuo. Activated sludge model no 1, Technical Report 1, *IAWPRC Task Group on Mathematical Modelling for Design and Operation of Biological Wastewater Treatment*, 1986.

[9] R. Hooke and T. A. Jeeves. Direct Search Solution of Numerical and Statistical Problems, *Journal on Associated Computation*, 8, 1961, pp. 212-229.

[10] R. M. Lewis and V. Torkzon. Pattern Search Algorithms for Bound Constrained Minimization, *SIAM Journal on Optimization*, 9(4), 1999, pp. 1082-1099.

[11] R. Otterpohl, T. Rolfa and J. Londong. Optimizing Operation of Wastewater Treatment Plants by Offline and Online Computer Simulation, *Water Science and Technology*, Vol. 30, no. 2, 1994, pp. 165-174.

[12] A. Seco, J. Serralta and J. Ferrer. Biological Nutrient Removal Model no. 1 (BNRM1), *Water Science and Technology*, Vol. 50, no. 6, 2004, pp. 69-78.

[13] I. Takács, G. G. Patry and D. Nolasco. A Dynamic Model of the Clarification-Thickening Process, *Water Research*, Vol. 25, No. 10, 1991, pp. 1263-1271.

[14] D. Tyteca, Y. Smeers and E. J. Nyns. Mathematical Modeling and Economic Optimization of Wastewater Treatment Plants, *CRC Critical Reviews in Environmental Control*, Vol. 8, No. 1, 1977, pp. 1-89.

[15] http://www.hydromantis.com