

## ON OPTIMIZATION OF SUBSTRATE REMOVAL IN A BIOREACTOR WITH WALL ATTACHED AND SUSPENDED BACTERIA

ALMA MAŠIĆ

EAWAG, Swiss Federal Institute of Aquatic Science and Technology  
Überlandstrasse 133, P.O. Box 611, CH-8600 Dübendorf, Switzerland

HERMANN J. EBERL

Biophysics Interdepartmental Program and Department  
of Mathematics and Statistics  
University of Guelph, Guelph ON, N1G 2W1, Canada

(Communicated by Sergei S. Pilyugin)

**ABSTRACT.** We investigate the question of optimal substrate removal in a biofilm reactor with concurrent suspended growth, both with respect to the amount of substrate removed and with respect to treatment process duration. The water to be treated is fed externally from a buffer vessel to the treatment reactor. In the two-objective optimal control problem, the flow rate between the vessels is selected as the control variable. The treatment reactor is modelled by a system of three ordinary differential equations in which a two-point boundary value problem is embedded. The solution of the associated singular optimal control problem in the class of measurable functions is impractical to determine and infeasible to implement in real reactors. Instead, we solve the simpler problem to optimize reactor performance in the class of off-on functions, a choice that is motivated by the underlying biological process. These control functions start with an initial no-flow period and then switch to a constant flow rate until the buffer vessel is empty. We approximate the Pareto Front numerically and study the behaviour of the system and its dependence on reactor and initial data. Overall, the modest potential of control strategies to improve reactor performance is found to be primarily due to an initial transient period in which the bacteria have to adapt to the environmental conditions in the reactor, i.e. depends heavily on the initial state of the dynamic system. In applications, the initial state, however, is often unknown and therefore the efficiency of reactor optimization, compared to the uncontrolled system with constant flow rate, is limited.

**1. Introduction.** Biofilms are macrostructures of bacterial cells that are attached to an immersed surface and/or to each other in an aqueous environment. The first cells adhere to a surface, become sessile and start producing a gel-like matrix of extracellular polymeric substances [21]. Formation and growth of the macrostructure occurs through proliferation among the already sessile bacteria or through attachment of new bacteria from the surrounding aqueous phase. Eventually, bacteria

---

2010 *Mathematics Subject Classification.* Primary: 92D25; Secondary: 49N90.

*Key words and phrases.* Biofilm, suspended growth, substrate removal, treatment time, optimization.

may also leave the biofilm by means of various processes collectively known as detachment [19]. Bacteria exist in a suspended mode of growth in addition to biofilms [28].

Both biofilms and suspended bacteria are used in wastewater treatment for removal of unwanted compounds. In the activated sludge process bacteria are kept in suspension while substrates enter the reactor with the influent [34]. The treated water and bacterial sludge then leave the reactor with the effluent. Biofilm reactors provide a longer retention time for bacteria due to the surface attachment. These processes are particularly suitable for e.g. nitrogen and phosphorus removal, which are dominated by slow growing bacteria. Biofilm systems do not require sludge recirculation in order to maintain a high biomass density in the reactor, in contrast to the activated sludge process [29]. New technologies have combined biofilms and activated sludge in hybrid or integrated fixed-film activated sludge (IFAS) processes [17]. Such systems encourage growth of bacteria in both biofilm and suspended mode and are often implemented as an upgrade of an existing treatment plant. However, even pure biofilm reactors always contain a certain amount of suspended biomass which must be removed from the effluent before the treated water can be released to further treatment steps or into a receiving water body [15].

Mathematical models have long been used in wastewater treatment technologies to describe and analyze different processes [36]. Both activated sludge models and biofilm models have a well developed foundation and framework. Although suspended biomass is an integral part of biofilm reactor systems, it is generally neglected in biofilm models [26]. It is assumed that suspended biomass does not contribute to the biofilm system and need not be considered. We investigated the effects and role of suspended biomass in biofilm reactors in two previous papers. A one-dimensional single species single substrate model was presented in [22], where the substrate, biofilm and suspended bacteria exist in a continuous stirred tank reactor (CSTR). This model is a biofilm extension of the Freter model of microbial growth with wall attachment that has been studied in [3, 4, 10, 18, 32]. The model takes into account the exchange of biomass through attachment and detachment between the biofilm and the suspended biomass. The bacteria will either be completely washed out from the reactor or exist in both suspended and biofilm form. We studied the longterm dynamic behavior of the reactor, assuming an infinite constant flow with unlimited substrate. It was shown that suspended bacteria are relatively more efficient at substrate removal than the biofilm. The model was extended in a follow-up study in [23], where microbial complexity was introduced to represent the nitrification process for a multi-species multi-substrate case. Numerical simulations showed that suspended biomass need not be considered if the quantity of interest is the overall reactor performance with respect to ammonium removal, but it plays a significant role in the intermediate processes of nitrification.

In contrast to these earlier studies that focused on longterm reactor behavior with an infinite supply of substrates, we focus now on the question of optimizing the treatment of a finite amount of substrate in finite time. In the setup that we consider, the biofilm reactor is fed from a storage tank which is to be emptied. This leads to an optimal control problem with the reactor flow rate as control function. It should be chosen to minimize two objectives: (i) the amount of untreated substrate that is discharged from the reactor and (ii) the treatment time. These objectives are not co-operative and a suitable compromise must be established.

In the literature, several studies can be found in which optimal control and optimization are used to improve bioreactor and treatment efficiency. In [13], the substrate concentration in an activated sludge system with unlimited amount of incoming substrate and a given final time was minimized by choosing the optimal flow rate. A sequencing batch reactor (SBR), which has a cyclic filling and emptying pattern, was studied in [25]. The goal was to minimize the fill and reaction time in the reactor by varying the flow rate at which the reactor is filled, while satisfying the end condition at every cycle of achieving a minimum substrate concentration before the reactor is emptied. The unlimited substrate was supplied either in a constant manner or by a predefined function. A generic time and spatial optimal control was discussed in [31], where e.g. a maximization of the amount of biomass at a given final time in a well-mixed bioreactor by varying the flow was considered. Furthermore, in [11] the goal was to minimize the time it takes to decrease the substrate concentration in a polluted lake to a given value by manipulating the flow with which polluted water is pumped from the lake to a biological treatment reactor and clean water is pumped back into the lake. A chemostat model was used for the biological reactor.

In most of these studies it is assumed that substrate is unlimited and the goal is often to minimize process duration, i.e. a pure time-optimal problem. Our problem is not only a time-optimal problem, but also involves a second objective to increase removal efficiency in the reactor. The resulting optimal control problem is singular, leading to chattering control, and as such is difficult to treat with standard numerical methods for optimal control.

Moreover, the reactor type that we consider, inspired by modern wastewater treatment technologies, contains both sessile biofilms and suspended bacteria, while the studies discussed above largely focus on reactors with suspended biomass only. Few optimization studies can be found in the literature, in which the focus is on a biofilm *per se*, but not the overall performance of an entire reactor. These studies are set in the context of time-dependent periodic dosing strategies for the eradication of biofilms with antibiotics [7, 33].

In Section 2, we formulate the model as a dynamic system for biomass and substrate in the reactor, coupled with a two-point boundary value problem, the solution of which determines the substrate flux into the biofilm. We introduce the optimization problem for reactor performance with two objectives, for which we seek a compromise in the Edgeworth-Pareto sense. In Section 3, we remark on the problem in the traditional context of optimal control theory and on the difficulties to tackle the problem in this setting. In Section 4, based on biological considerations, we introduce a two-parametric class of control functions, and simplify the optimal control problem as a nonlinear vector optimization problem, for which we approximate the Pareto front numerically. Finally, in Section 5, we discuss briefly two alternative approaches, which, however, for computational reasons require a significant simplification of the underlying reactor model. Results and conclusion are summarized in Section 6.

## 2. Problem formulation.

**2.1. Mathematical model.** The single species single substrate model from [22] was constructed through a coupling between a CSTR mass balance and a traditional one-dimensional biofilm model. The bacteria are present in the reactor in suspended

and biofilm form. Microbial growth is limited by a single substrate, which is continuously supplied to the reactor. Bacterial cells can detach from the biofilm and become suspended, which in turn can (re-)attach to the biofilm. It is assumed that the biofilm covers the substratum in the reactor uniformly and that the substrate diffuses into the biofilm creating concentration gradients due to diffusion and reaction. The flow, with which substrate is supplied, also removes suspended biomass and unconsumed substrate with the effluent.

**Governing equations.** The model is formulated with respect to the three dependent variables substrate concentration in the aqueous phase  $S$  [g/m<sup>3</sup>], suspended biomass  $u$  [g] and biofilm thickness  $\lambda$  [m], and reads:

$$\dot{S} = \frac{1}{V} \left( Q(S_0 - S) - \frac{u\mu_u(S)}{\gamma} - AD_c j(\lambda, S) \right) \quad (1)$$

$$\dot{u} = u \left( \mu_u(S) - \frac{Q}{V} - k_u \right) + A\rho d(\lambda)\lambda - \alpha u \quad (2)$$

$$\dot{\lambda} = \frac{\gamma D_c}{\rho} j(\lambda, S) - \lambda k_\lambda + \frac{\alpha u}{A\rho} - d(\lambda)\lambda \quad (3)$$

where  $V$  [m<sup>3</sup>] is the constant treatment reactor volume,  $A$  [m<sup>2</sup>] the substratum area in the reactor,  $Q$  [m<sup>3</sup>/d] the influent and effluent flow rate,  $S_0$  [g/m<sup>3</sup>] the influent substrate concentration,  $\gamma$  [-] the biomass yield from the substrate,  $k_u, k_\lambda$  [1/d] the decay rates for suspended and biofilm biomass,  $D_c$  [m<sup>2</sup>/d] the diffusion coefficient of substrate in the biofilm,  $\rho$  [g/m<sup>3</sup>] the biofilm biomass density and  $\alpha$  [1/d] the attachment rate. The model is essentially a biofilm extension of the Freter model of microbial growth with wall attachment, which has been studied in [3, 4, 10, 18, 32]. The biofilm detachment function  $d(\lambda)$  that we use is defined as

$$d(\lambda) = E\lambda, \quad (4)$$

where  $E$  [1/md] is the erosion parameter. This is the most frequently used detachment rate in the biofilm modeling literature.

Using the same notation as in [22] the substrate net flux into the biofilm is defined through the function  $j(\lambda, S)$  [g/m<sup>4</sup>]

$$j(\lambda, S) := \begin{cases} \frac{\rho}{\gamma D_c} \int_0^\lambda \mu_\lambda(C(z)) dz, & \lambda > 0 \\ 0 & \lambda = 0, \end{cases} \quad (5)$$

where  $C(z)$  [g/m<sup>3</sup>] is the substrate concentration in the biofilm at depth  $0 < z \leq \lambda$ . The specific growth rates for suspended and biofilm biomass are  $\mu_u(S)$  [1/d] and  $\mu_\lambda(C(z))$  [1/d] respectively, defined as

$$\mu_u(S(t)) = \frac{\mu_u^{max} S(t)}{K_u + S(t)}, \quad \mu_\lambda(C(z)) = \frac{\mu_\lambda^{max} C(z)}{K_\lambda + C(z)} \quad (6)$$

where  $\mu_u^{max}, \mu_\lambda^{max}$  [1/d] are the maximum growth rates and  $K_u, K_\lambda$  [g/m<sup>3</sup>] the Monod half-saturation concentrations. The local substrate concentration  $C(z)$ ,  $0 \leq z \leq \lambda$ , is found as a solution to the quasi-steady state diffusion reaction equation posed as a boundary value problem

$$D_c C''(z) = \frac{\rho}{\gamma} \mu_\lambda(C(z)), \quad C'(0) = 0, C(\lambda) = S. \quad (7)$$

Here we made use of the standard time scale argument that diffusion and consumption in the biofilm is fast compared to microbial growth.

**Properties of the flux function.** We observe that the function  $j(\lambda, S)$  in (5) depends on  $S$  through the boundary condition in (7). As a quantity that is derived from the solution of a two-point boundary value problem,  $j(\lambda, S)$  is difficult to evaluate. However, many of its qualitative properties are known [2, 20, 22]: It is well-defined and twice differentiable for  $\lambda > 0, S > 0$ . We have  $j(\cdot, 0) = j(0, \cdot) = 0$ . Moreover,  $j(\lambda, S)$  is monotonically increasing in both its arguments and concave in  $S$ . Using standard comparison theorems, the flux function  $j(\lambda, S)$  can be estimated from below and above by

$$S\sqrt{\frac{\theta}{K_\lambda + S}} \tanh \sqrt{\frac{\theta\lambda^2}{K_\lambda + S}} \leq j(\lambda, S) \leq S\sqrt{\frac{\theta}{K_\lambda}} \tanh \sqrt{\frac{\theta\lambda^2}{K_\lambda}}$$

where  $\theta := \frac{\rho\mu_\lambda^{max}}{\gamma D_c}$ . The wastewater engineering literature contains several empirical formulae that approximate  $j(\lambda, S)$ . These have been developed based on experimental data but usually contain additional parameters that are reactor specific. Using two very different approaches [2, 20] one obtain the approximation

$$j(\lambda, S) = \mu_\lambda \frac{\lambda S}{K_\lambda + S} \quad (8)$$

that only involves parameters of the underlying two-point boundary value problem (7). Numerically it was verified in [2] that this approximation appears to work reasonably well for an extended range of realistic biofilm parameters in some realistic ranges for  $\lambda$  and  $S$ , however, primarily for thin biofilms. Below we will normally use the full flux calculation according to (5) in our numerical simulations and resort to the flux approximation (8) only where computational costs prevent us from using (5) in Sections 5.1 and 5.2.

With the above properties of  $j(\lambda, S)$ , (1)-(3) is formally an ordinary differential equation, albeit with a difficult to evaluate right hand side. In [22] its longterm behavior was analyzed for a constant flow rate  $Q$ , where also easy to evaluate conditions for the stability and instability of the washout equilibrium were derived.

**Solutions in the case of non-constant flow rate.** In the context of optimal control problems, the flow rate  $Q$  will be the control parameter, i.e. is no longer constant. The natural setting is to search for control functions  $Q$  in the class of measurable functions. Then, (1) - (3) becomes a non-autonomous system of differential equations. For non-negative  $S, u, \lambda$  and measurable  $Q$ , the right hand side of (1) - (3) is continuous in the dependent variable for fixed  $t$ , and it is measurable in  $t$  for fixed  $S, u, \lambda$ . Thus, solutions of (1)-(3) are understood in the generalized sense of Carathéodory [6]. Many of the standard results on ODEs carry over to Carathéodory differential equations accordingly [35]. The initial value problem of (1)-(3) with positive initial data possesses a unique solution that remains non-negative almost everywhere. Moreover, by forming a linear combination of (1), (2), (3), we obtain for non-negative initial data the upper estimate

$$V\gamma S(t) + u(t) + A\rho\lambda(t) \leq V\gamma S(0) + u(0) + A\rho\lambda(0) + \gamma S_0 \int_0^t Q(t)dt. \quad (9)$$

almost everywhere. Thus, biomass production is bounded by the amount of substrate supplied. Furthermore, if the initial data are positive, there is no interval  $(t', t'')$  such that either  $u(t) = 0, \lambda(t) > 0$  or  $u(t) > 0, \lambda(t) = 0$  for  $t' < t < t''$ . These results carry over directly from [22] to the setting of Carathéodory solutions, using

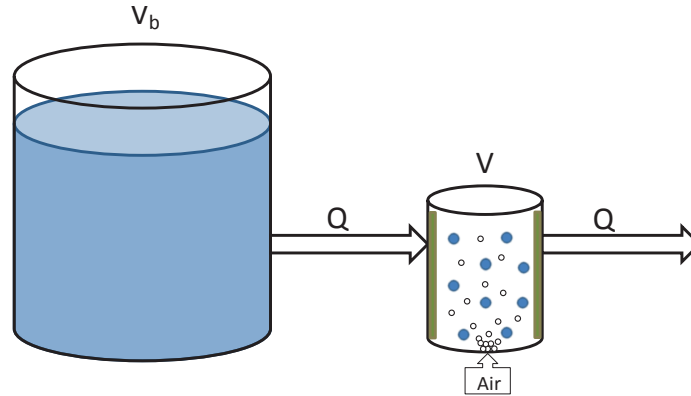


FIGURE 1. Wastewater from the buffer reactor with volume  $V_b$  is pumped to the treatment reactor of volume  $V$  at a flow rate  $Q$ , with treated water leaving the reactor at the same rate. The treatment reactor is aerated from the bottom and contains bacteria in the form of biofilms and suspended biomass, which ensure that the process is not oxygen limited but that substrate  $S$  is the only growth limiting factor.

the comparison theorems in [35]. On intervals  $(t', t'')$  on which  $Q(t)$  is continuous, the solutions of (1)-(3) are differentiable, i.e. traditional ODE solutions.

**2.2. Optimization problem. Reactor setup.** The analysis and investigation of the basic model (1)-(7) in [22] assumed a constant flow rate and essentially an infinite amount of substrate to be treated; the focus of the investigation was on longterm behavior, i.e the asymptotics of the model for  $t \rightarrow \infty$ . From an engineering perspective, it is useful to investigate the question of treating a finite amount of substrate in finite time. More specifically, one is interested in the problem, for example, to degrade a given amount of substrate as much as possible in as short a time as possible.

To this end we consider a treatment reactor, described by (1)-(7) that is fed from a buffer reactor that contains the volume  $V_b(t)$  [m<sup>3</sup>] of water with a bulk substrate concentration  $S_0$  [g/m<sup>3</sup>], see Figure 1. The emptying of the buffer reactor is described by the simple mass balance

$$\dot{V}_b = -Q, \quad (10)$$

where  $Q(t)$  is a non-negative function such that  $\int_0^T Q(t)dt = V_b(0)$  for some  $T > 0$  and  $Q(t) = 0$  for  $t > T$ . This restriction guarantees that the buffer reactor is emptied in finite treatment time  $T$  and that negative buffer volumes are avoided.

**Vector optimization problem.** The task at hand is now formulated as the vector optimization problem

$$\min_{Q \in \Omega} \begin{pmatrix} \int_0^T QS dt \\ T \end{pmatrix}, \quad (11)$$

where  $\Omega$  is the set of admissible flow rates  $Q : [0, T_{max}] \rightarrow \mathbb{R}_0^+$ . This set will be made more precise below.  $T_{max}$  is the maximum duration of the process we are prepared to tolerate.

The two objectives are not cooperative. In order to minimize the first objective, i.e. the amount of substrate in the effluent, one would expect that large process durations be advantageous, so the microbes have sufficient time to degrade almost all the substrate. This, however, obviously counters the objective to minimize process time.

The trade-off between both goals that we seek is a compromise in the sense of Edgeworth-Pareto optimality, with the positive cone as order cone [12, 16]. We denote the two objectives by

$$z_1(Q) := \int_0^T QS dt, \quad z_2(Q) := T,$$

and the vector of objectives by  $Z := (z_1, z_2)^T$ . An admissible flow rate  $Q^*(t)$  is said to be Edgeworth-Pareto optimal if a further improvement of one objective is only possible at the expense of increasing the other. We call then  $Z^* = (z_1(Q^*), z_2(Q^*))^T$  an EP-minimum. We must not expect that the EP-minimum is unique [12, 16]. Rather, we expect that there might be infinitely many such optima. The set of these optima can be graphically represented in the  $z_1$ - $z_2$  plane. They form the graph of a monotonically decreasing function, the so-called *Pareto Front*. Points below the graph are not attainable with an admissible flow rate, points above the graph are not optimal. To compute the Edgeworth-Pareto minima, the vector optimization problem is converted into a family of scalar optimization problems.

**Scalarization with linear functionals.** A standard approach to compute Edgeworth Pareto optima is to scalarize the vector objective function  $Z$  with a monotonically increasing functional  $\mathcal{F} : \mathbb{R}^2 \rightarrow \mathbb{R}$  and to solve the scalar problem

$$\min_{Q \in \Omega} \mathcal{F}(Z(Q)). \quad (12)$$

The solution of this scalar optimization problem is an Edgeworth-Pareto optimal solution of the vector optimization problem (11). Using linear functionals, the objective function of (12) becomes

$$J(Q) := \mathcal{F}(Z(Q)) = w_1 z_1 + w_2 z_2, \quad w_1 > 0, w_2 > 0. \quad (13)$$

Thus, we arrive at the following optimal control problem in Bolza form:

$$\min_{Q \in \Omega} w\beta \int_0^T QS dt + (1-w)T. \quad (14)$$

Here  $\beta := T_{max}/(S_0 V_b(0))$  is a constant that is introduced to rescale the problem such that both objectives have the same units. This allows to replace  $w_1$  and  $w_2$  in (13) by a single dimensionless weight  $w$  with  $0 < w < 1$ . The EP-optimum that is found this way as a solution of (14) depends on the choice of  $w$ . The Pareto Front is the union of these points. In the numerical realization, (14) is solved for a finite number of weights that is large enough to approximate the Pareto Front well.

**Modified Polak Algorithm.** An alternative approach to determining the Pareto Front numerically, which is particularly useful and easy to implement for our problem due to the special form of  $z_2$ , is the Modified Polak Algorithm [16].

The interval  $(0, T_{max}]$  is divided into a finite number of  $n$  subintervals by introducing  $0 < T_0 < T_1 < T_2 < \dots < T_n := T_{max}$ . For every  $T_i$  the problem

$$\min_{Q \in \Omega_i} \int_0^{T_i} Q S dt \quad (15)$$

is solved, where the sets

$$\Omega_i := \{Q \in \Omega : Q(t) = 0, t > T_i\} \subset \Omega$$

are the sets of admissible functions such that the reactor is emptied in finite time  $T_i$ . The Pareto Front can be approximated by interpolation after removing those segments of the curve that perturb monotonicity, if any.

In the next sections we will make use of both approaches to discuss the solutions of (11), depending on the situation.

**3. Remarks on the optimal control problem.** Admissible functions in optimal control theory are often defined in the class of measurable functions. Most standard theorems and results in optimal control theory are derived for these types of functions. Furthermore, realistically there exists an upper bound  $Q_{max}$  on the flow rate  $Q$  due to the pump capacity between the buffer reactor and the treatment reactor. We therefore proceed by reformulating our problem and studying the set of measurable, bounded flow rates.

The constraint on  $Q$ , which enforces emptying of the reactor, i.e.  $\int_0^T Q dt = V_b(0)$ , can be restated in the form of a differential equation for the buffer volume  $V_b$ . Thus, the system of equations (1)-(3) now reads

$$\dot{S} = \frac{1}{V} \left( Q(S_0 - S) - \frac{u\mu_u(S)}{\gamma} - AD_c j(\lambda, S) \right) \quad (16)$$

$$\dot{u} = u \left( \mu_u(S) - \frac{Q}{V} - k_u \right) + A\rho d(\lambda)\lambda - \alpha u \quad (17)$$

$$\dot{\lambda} = \frac{\gamma D_c}{\rho} j(\lambda, S) - \lambda k_\lambda + \frac{\alpha u}{A\rho} - d(\lambda)\lambda \quad (18)$$

$$\dot{V}_b = -Q \quad (19)$$

with initial conditions

$$S(0) = S_0, \quad u(0) = u_0, \quad \lambda(0) = \lambda_0, \quad V_b(0) = V_{b0} \quad (20)$$

and an end condition

$$V_b(T) = 0 \quad (21)$$

for the optimal control problem in Lagrange form

$$\min_{Q \in \Omega} \int_0^T (w\beta QS + (1-w)) dt, \quad \text{subject to } T_{min} \leq T \leq T_{max} \quad (22)$$

where  $T_{min} = V_b(0)/Q_{max}$  is the minimum possible time and  $\Omega$  is the set of admissible control functions. Let

$$\Omega = \{Q \text{ measurable} : 0 \leq Q \leq Q_{max} < \infty\}. \quad (23)$$

It can be shown by using standard existence results [9] that the optimal control problem (22) with the set of admissible functions  $\Omega$ , subject to (16)-(19) with initial conditions (20) and end condition (21) possesses an optimal solution  $Q^*$ . The



optimal solution may be characterized with Pontryagin’s minimum principle [9, 27]. The Hamiltonian is

$$\begin{aligned}
 H = & \lambda_1 \left( \frac{1}{V} \left( Q(S_0 - S) - \frac{u\mu_u(S)}{\gamma} - AD_c j(\lambda, S) \right) \right) \\
 & + \lambda_2 \left( u \left( \mu_u(S) - \frac{Q}{V} - k_u \right) + A\rho d(\lambda)\lambda - \alpha u \right) \\
 & + \lambda_3 \left( \frac{\gamma D_c}{\rho} j(\lambda, S) - \lambda k_\lambda + \frac{\alpha u}{A\rho} - d(\lambda)\lambda \right) \\
 & + \lambda_4(-Q) + w\beta QS + 1 - w
 \end{aligned} \tag{24}$$

where  $\lambda_1(t), \dots, \lambda_4(t)$  are the adjoint variables given by

$$\begin{aligned}
 \frac{d\lambda_1}{dt} = -\frac{\partial H}{\partial S} = & -w\beta Q + \frac{\lambda_1 Q}{V} + \left( \frac{\lambda_1}{V\gamma} - \lambda_2 \right) \mu'_u(S)u \\
 & + D_c \frac{\partial j(\lambda, S)}{\partial S} \left( \frac{\lambda_1 A}{V} - \frac{\lambda_3 \gamma}{\rho} \right)
 \end{aligned} \tag{25}$$

$$\frac{d\lambda_2}{dt} = -\frac{\partial H}{\partial u} = -\mu_u(S) \left( \frac{\lambda_1}{V\gamma} + \lambda_2 \right) + \lambda_2 \left( \frac{Q}{V} + k_u + \alpha \right) - \frac{\lambda_3 \alpha}{A\rho} \tag{26}$$

$$\begin{aligned}
 \frac{d\lambda_3}{dt} = -\frac{\partial H}{\partial \lambda} = & D_c \frac{\partial j(\lambda, S)}{\partial \lambda} \left( \frac{\lambda_1 A}{V} - \frac{\lambda_3 \gamma}{\rho} \right) \\
 & + (\lambda_3 - \lambda_2 A\rho) (d'(\lambda)\lambda + d(\lambda)) - \lambda_3 k_\lambda
 \end{aligned} \tag{27}$$

$$\frac{d\lambda_4}{dt} = -\frac{\partial H}{\partial V_b} = 0 \tag{28}$$

with the transversality conditions

$$\lambda_1(T) = \lambda_2(T) = \lambda_3(T) = 0 \text{ and } \lambda_4(T) = \text{free}. \tag{29}$$

Pontryagin’s minimum principle states that

$$H(S^*, u^*, \lambda^*, V_b^*, Q^*, \lambda_1^*, \lambda_2^*, \lambda_3^*, \lambda_4^*) \leq H(S^*, u^*, \lambda^*, V_b^*, Q, \lambda_1^*, \lambda_2^*, \lambda_3^*, \lambda_4^*) \tag{30}$$

for the optimal solution  $Q^*$ , where  $S^*, u^*, \lambda^*, V_b^*$  are the optimal state trajectories and  $\lambda_1^*, \lambda_2^*, \lambda_3^*, \lambda_4^*$  the optimal adjoint trajectories. Furthermore, since (22) is a free final time problem and the Hamiltonian does not explicitly depend on time, at the optimal solution the Hamiltonian must satisfy

$$H(S^*, u^*, \lambda^*, V_b^*, Q^*, \lambda_1^*, \lambda_2^*, \lambda_3^*, \lambda_4^*) = 0 \tag{31}$$

for all  $t \in [0, T^*]$ . The optimal control trajectory can be obtained by minimizing the Hamiltonian with respect to  $Q$ . To this end, the Hamiltonian can be rewritten as

$$H = \text{terms without } Q + Q \left( w\beta S + \frac{\lambda_1(S_0 - S)}{V} - \frac{\lambda_2 u}{V} - \lambda_4 \right). \tag{32}$$

Set  $h = w\beta S + \frac{\lambda_1(S_0 - S)}{V} - \frac{\lambda_2 u}{V} - \lambda_4$ , this is now the *switching function*.  $H$  is linear in  $Q$ , implying a bang-bang control as the optimal solution, switching between 0 and  $Q_{max}$ . The switching times can be determined with the signs of the switching function  $h$ .

Optimal control problems like (22) that are linear in the control variable, with bounded control sets, are known to lead to chattering [37]. Such controls are characterized by an infinite number of switches between the lower and upper bound in a

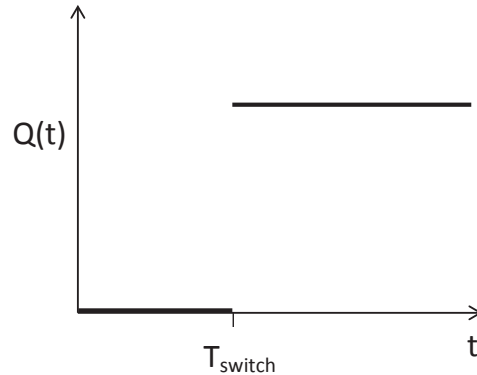


FIGURE 2. Flow rate  $Q(t)$  as an off-on function.

finite amount of time. This type of control regime is, however, not appropriate from a practical point of view, wherefore this control set will not be considered here.

The solutions to (22) in the class of measurable and bounded admissible functions (23) depend heavily on  $Q_{max}$ , the upper bound for the flow rate. Giving up this restriction, we permit  $Q$  to be unbounded, i.e. let  $\Omega$  be the set of measurable, non-negative flow rates  $Q$  defined on  $[0, T_{max}]$ . For the optimal control problem (14), i.e., a minimum problem with a linear integrand, and with this definition of  $\Omega$ , Fleming and Rishel [9] indicate that impulse control solutions often are found. These controls are characterized by impulses of the Dirac delta function type and cause instantaneous jumps between different values in the state variables. Such a behavior is, for practical purposes, not appropriate and will, therefore, not be pursued further.

#### 4. A finite-dimensional optimization problem.

**4.1. Off-on control functions.** The rate with which substrate can be degraded depends on both, the substrate concentration and the amount of bacterial biomass in the system. In a typical application, the bacterial populations, wall attached and suspended, are initially not fully developed, but must adjust to the environmental conditions in the reactor. Since the overall degradation rate scales with amount of biomass, in this initial period the reactor performance is limited by small population sizes. If the flow rate  $Q$  in this initial period is too high, much more substrate is supplied than can be utilized and more substrate might be washed out of the reactor than degraded. Therefore, a first and simple candidate class for admissible flow rates are off-on functions. The flow rate is initially turned off for a transient period during which the bacterial population can establish itself. Later it is switched on to a constant value until the buffer reactor is empty. With this definition the admissible functions are the functions

$$Q(t) = \begin{cases} 0, & \text{for } t < T_{\text{switch}} \\ \frac{V_b(0)}{T - T_{\text{switch}}}, & \text{for } T_{\text{switch}} < t \leq T \end{cases} . \quad (33)$$

An example of such a function is shown in Figure 2. Since for practical purposes the buffer reactor cannot be emptied instantaneously, it is appropriate to introduce the minimum required treatment time  $T_{min}$ . Every pair  $(T_{\text{switch}}, T)$  with

$T_{min} \leq T_{switch} < T \leq T_{max}$  defines uniquely one such function  $Q(t)$  and, conversely, every such function is defined by a unique pair  $(T_{switch}, T)$ . Therefore, the set of admissible functions is a two-parametric family of functions and finding the optimal  $Q^* \in \Omega$  is equivalent to finding the corresponding pair  $(T_{switch}^*, T^*)$ .

With this choice of admissible controls, the optimal control problem (11) reduces to a vector optimization problem in two unknown parameters. We use the Modified Polak Method to scalarize the problem: we solve for selected  $T_i, i = 0, \dots, n$  with  $T_{min} = T_0$  and  $T_{max} = T_n$ , and  $T_0 < T_1 < T_2 < \dots < T_n$  the problem

$$\min_{T_{min} \leq T_{switch} \leq T_i} \int_0^{T_i} Q(t)S(t)dt. \quad (34)$$

with  $Q$  according to (33). Thus, the problem reduces to finding the solutions of  $n+1$  scalar optimization problems with one argument for every given  $T_i$ . To compute the objective function requires to solve numerically (1)-(7) and to evaluate the integral (34).

**4.2. Computational setup.** The model parameters for the numerical simulations are found in Table 1, where, as in [22], the parameters were taken from Benchmark Problem 1 of the International Water Association's Taskgroup on Biofilm Modelling [36], which models an aerated reactor with carbon as the only substrate, and from [22]. As in [22], we use the same parameter values for both biofilm and suspended biomass, implying that suspended biomass is treated as unattached flocs of microbial communities.

All simulations, unless otherwise specified, are initiated with the following initial conditions:

$$S(0) = 30\text{g/m}^3, \quad u(0) = 0.05\text{g}, \quad \lambda(0) = 0.0001\text{m}. \quad (35)$$

We consider a small laboratory scale buffer reactor with volume  $V_b = 0.05\text{m}^3 = 50\text{l}$ . The minimum and maximum treatment periods were chosen as

$$T_{min} = T_0 = 0.1\text{days} = 2.4\text{hours} \quad \text{and} \quad T_{max} = T_n = 50\text{days}. \quad (36)$$

The treatment times  $T_i$  were selected on a non-equidistant grid with  $n = 62$  points. The grid points are chosen denser for small  $T$  than for large  $T$ . All simulations were executed in MATLAB R2012a using the built-in solvers `ode15s` (a variable order, multi-step solver based on the numerical differentiation formulas) for the solution of the system of ordinary differential equations (1)-(3), `bvp4c` (a finite difference code using the three-stage Lobatto IIIa formula, which is a collocation formula [24]) for the diffusion-reaction equation (7) and `fminbnd`, a derivative-free method based on golden section search and parabolic interpolation [24], for the minimization problem (34).

**4.3. Results and discussion.** The choice of using a constant flow rate  $Q$  requires no further adjustment after initiation until the buffer reactor is emptied. It will, therefore, serve as a baseline for comparison with other flow rates to measure efficiency of a control strategy.

**Solution of the vector optimization problem.** The Pareto Front for the optimal control problem (11), as determined by the Modified Polak Algorithm is shown in Figure 3a. This curve is monotonically decreasing from  $(\max z_1, \min z_2)$  to  $(\min z_1, \max z_2)$  with respect to  $z_2$ : decreasing  $z_2$  further will only be possible at the expense of increasing  $z_1$  and *vice versa*. Increasing the treatment time  $T$  allows a

TABLE 1. Model parameters.

Symbol	Name	Value	Unit	Reference
$A$	inner reactor area	0.0647	$\text{m}^2$	[22]
$\alpha$	attachment rate	1	$\text{d}^{-1}$	[22]
$D_c$	diffusion coefficient	$10^{-4}$	$\text{m}^2\text{d}^{-1}$	[36]
$E$	erosion parameter	1000	$\text{m}^{-1}\text{d}^{-1}$	[22]
$\gamma$	yield of biomass from substrate	0.63	—	[36]
$k_\lambda, k_u$	decay rates	0.4	$\text{d}^{-1}$	[36]
$K_\lambda, K_u$	Monod half-saturation concentrations	4	$\text{gm}^{-3}$	[36]
$\mu_\lambda^{max}, \mu_u^{max}$	maximum specific growth rates	6	$\text{d}^{-1}$	[36]
$\rho$	biomass density	10000	$\text{gm}^{-3}$	[36]
$V$	treatment reactor volume	0.001	$\text{m}^3$	[22]
$V_b$	buffer reactor volume	0.05	$\text{m}^3$	assumed

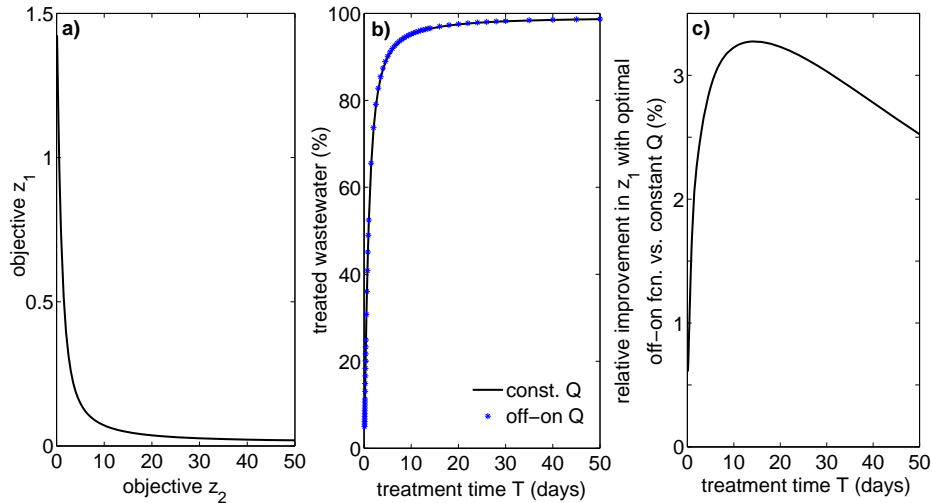


FIGURE 3. **a)** Edgeworth-Pareto minima of (11) with  $z_1 = \int_0^T Q S dt$  and  $z_2 = T$ . **b)** Percentage of treated wastewater in relation to initial amount of untreated wastewater as a function of treatment time  $T$  for constant flow rates  $Q = V_b/T$  (solid line) and optimal off-on functions (blue asterisks). **c)** Relative improvement (%) in objective  $z_1$  of off-on functions compared to the constant flow rate.

further reduction of substrate in the outflow. Therefore, we can view the amount of substrate that can be degraded as a function of treatment time  $T$ . A steep incline in the Pareto Front is visible for  $T < 5$ , after which the curve levels off. Short treatment times imply high flow rates and, therefore, high washout of substrate. Hence, reactor performance increases drastically with increased treatment time for small  $T$ . On the other hand, for large  $T$  ( $T > 10$  in our example) most substrate is removed, so that only a small further improvement in performance is noted for large  $T$ .

Each point on the Pareto curve is an EP minimum of (11) and represents a compromise between treatment time and reactor performance. In general it depends on user preferences which of these compromises to accept. One might argue that of all the EP minima a distinguished one is the point on the Pareto front that is closest in some suitable metric to the (unachievable) ideal target  $(T_{min}, 0)$ .

We found the solution of this *goal programming* problem [12, 16] by normalizing the  $z_1$ - and  $z_2$ -axis, computing the Euclidean distance

$$\sqrt{\left(\frac{z_1 - 0}{V_b(0)S_0}\right)^2 + \left(\frac{z_2 - T_{min}}{T_{max}}\right)^2}$$

and minimizing over  $T$ . The closest point on the Pareto Front was found at  $T = 5.09285$  days with  $z_1 = 0.14499$ .

We plot the optimal performance of the off-on-controlled reactors as a function of treatment time  $T$  in Figure 3b as blue asterisks, together with the performance of the uncontrolled reactor, i.e. the reactor with constant flow rate  $Q = V_b(0)/T$ . Both curves are very close to each other, with hardly any difference visible. When comparing the off-on functions with the constant flow rate  $Q$  we see a small improvement in reactor performance with off-on functions over constant  $Q$  in Figure 3c. The relative improvement ranges between approximately 0.5% and 3.5%. For short treatment times the improvement remains small and begins increasing when the treatment time is increased, reaching a maximum around  $T = 10$  days, after which it slowly decreases.

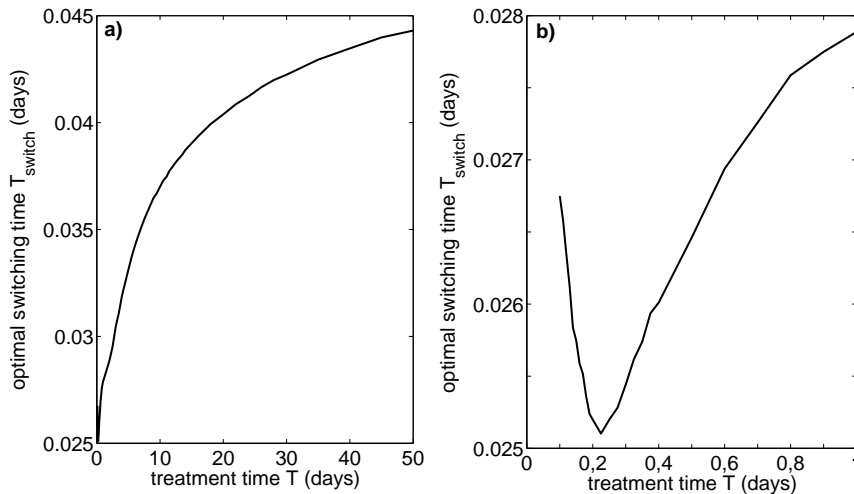


FIGURE 4. The optimal switching times  $T_{switch}$  determined by the modified Polak method for the optimal control problem (11) in a) with a close-up for  $0.1 \leq T \leq 1$  in b).

Figures 4a-b show the optimal switching times  $T_{switch}^*$  as a function of  $T$ . The function initially decreases as treatment time increases until a minimum is reached around 0.2 days. The function begins to increase again when the treatment time is further increased. An increased switching time implies a longer bacterial adjustment time during which the flow rate is zero. We find  $T_{switch}^* \ll T$ , with  $T_{switch}^*$  roughly

in the interval  $[0.025, 0.045]$  for  $0.1 \leq T \leq 50$ . A possible explanation for the initial drop in the switching time is that for very short treatment times  $T < 0.2$  the system is entirely dominated by washout.

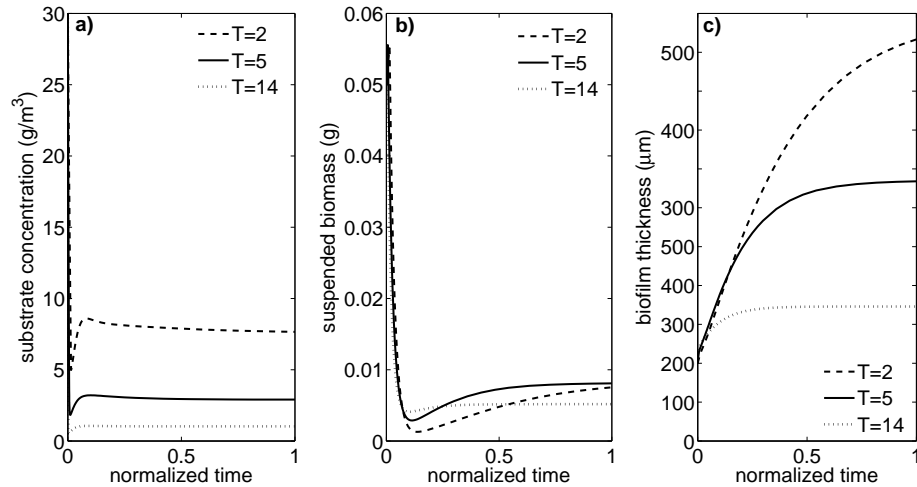


FIGURE 5. Time evolutions of **a)** substrate concentration  $S$  [ $\text{g/m}^3$ ], **b)** suspended biomass  $u$  [g] and **c)** biofilm thickness  $\lambda$  [ $\mu\text{m}$ ] for the treatment times  $T = 2$  (dashed),  $T = 5$  (solid) and  $T = 14$  days (dotted). The time-axis has been normalized to facilitate graph visibility.

In Figure 5 we plot the time evolutions of the three state variables  $S, u, \lambda$  for three selected treatment times  $T = 2, 5, 14$  days. The time-dependent functions are plotted on a normalized time axis. Initially, while the flow is turned off, the substrate in the treatment reactor is degraded by suspended and wall attached biomass, which increases. The flow is switched on rather early, which is clearly visible in Figure 5a. This brings fresh substrate into the reactor. Hence  $S$  increases. Simultaneously, the flow washes out some of the suspended biomass, whence  $u$  decreases. The biofilm thickness keeps increasing throughout. Eventually all three state variables level off. The treatment time  $T$  has an effect on the state that is attained by the reactor. The longer the treatment period, the smaller is the final substrate concentration. Smaller bulk substrate concentrations imply that the biofilm that can be sustained will be thinner, as observed in Figure 5a. While the effect of treatment time  $T$  on the suspended biomass is clearly visible, it is not as straightforward to summarize. For one, shorter process duration implies higher substrate concentrations and therefore faster growth of suspended biomass. On the other hand, shorter process duration implies higher flow rates and therefore higher washout of suspended biomass. The latter phenomenon is also reflected by the minimum to which  $u$  dips after the flow is turned on. For  $T > 5$ , the value that  $u$  levels off at does not change notably.

Complementing the above data, we plot in Figure 6 the quantities  $S(T_{switch}^*), u(T_{switch}^*), \lambda(T_{switch}^*)$  as functions of  $T$ , i.e. the state of the reactor at the optimal time to turn on the flow. Their respective initial values are plotted for comparison. The substrate concentration in Figure 6a shows that the final concentrations in the

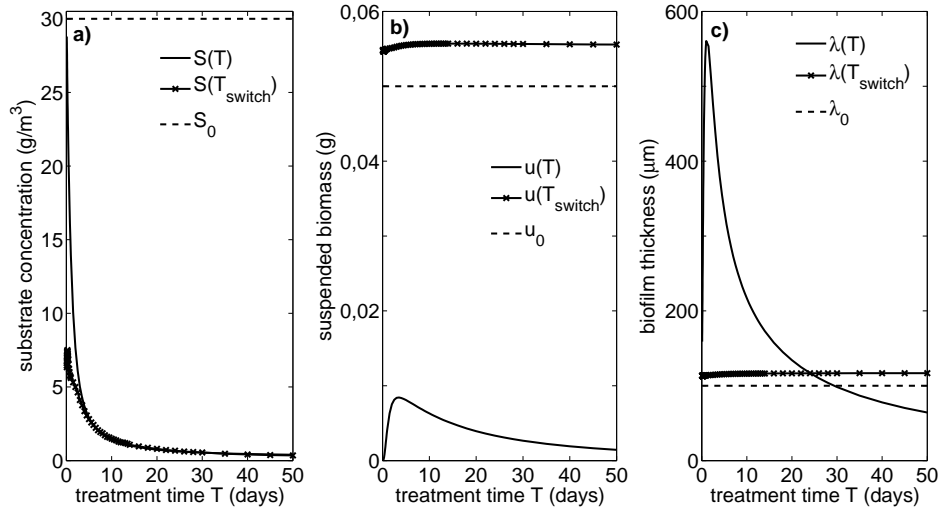


FIGURE 6. Values at  $T$  (solid) and  $T_{switch}$  (x-marks) with the initial values (dashed) for **a)** substrate concentration  $S$  [g/m<sup>3</sup>], **b)** suspended biomass  $u$  [g] and **c)** biofilm thickness  $\lambda$  [ $\mu\text{m}$ ] for different treatment times  $T$  [days].

reactor after the treatment are higher than the concentrations at the time of flow start,  $S(T) > S(T_{switch}^*)$ , for short treatment times  $T < 5$ . This was also observed in Figure 5a for  $T = 2$  days. However, for  $T > 5$  we have  $S(T) \approx S(T_{switch}^*)$ .

A completely different pattern is seen for the suspended biomass in Figure 6b, where  $u(T_{switch}^*) > u(T)$  for all  $T$ . By comparing this figure with Figure 5b we see that the considerable drop in biomass occurs after the start of the flow, i.e. right after  $T_{switch}$ . The final concentrations  $u(T)$  reach a peak around  $T = 5$  days after which they decrease.

For all  $T$ ,  $\lambda(T_{switch}^*)$  in Figure 6c, the biofilm thickness when the flow is turned on, is only slightly higher than the initial biofilm thickness. However, the final biofilm thickness  $\lambda(T)$  varies considerably for different  $T$ . For short treatment times  $T < 20$  we have  $\lambda(T) > \lambda(T_{switch}^*)$  with a maximum reached around  $T = 1$ . For  $T > 30$  we find thickness  $\lambda(T) < \lambda(0) < \lambda(T_{switch}^*)$ , due to substrate limitations in the biofilm.

In Figure 6 the maxima of suspended biomass and biofilm thickness in dependence on the length of the treatment period are most likely related to substrate depletion: For small  $T$  the biomass at the end of the treatment time increases because the more time the bacteria are given the more substrate they can remove and hence grow. However, for longer process durations, the system becomes eventually substrate depleted which explains the monotonic decrease in biomass for large  $T$  due to starvation and washout.

**Dependency of reactor performance on initial data.** The definition of off-on functions in 4.1 was motivated by an initial adaptation period for the bacteria before the flow was turned on, characterized by the switching time  $T_{switch}$ . Therefore, sensitivity of the reactor performance to initial data  $S(0), u(0), \lambda(0)$  is interesting. In order to investigate the effects on the solution by varying initial data we performed a study in which initial data for suspended biomass and biofilm

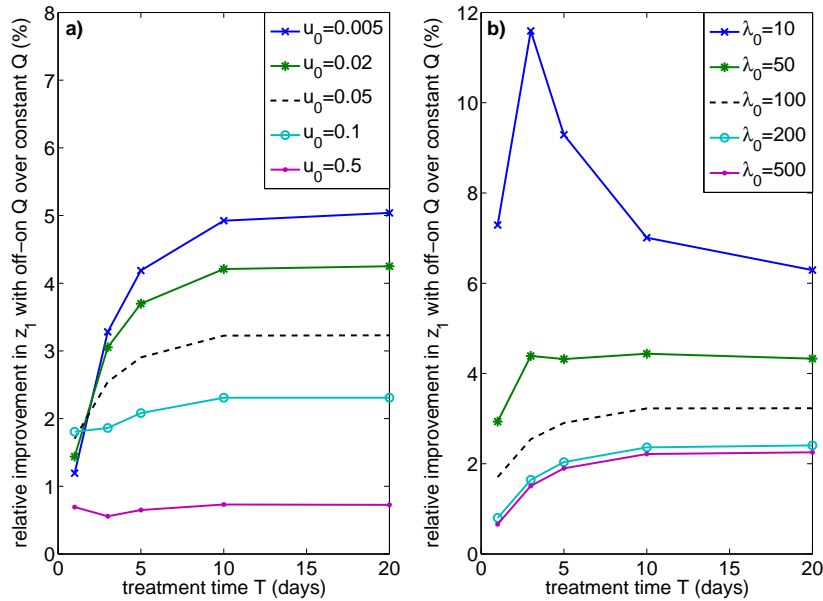


FIGURE 7. Relative improvement in objective  $z_1$  with optimal off-on flow rate  $Q$  compared with constant  $Q$  (%) for varying initial **a)** suspended biomass  $u_0$  and **b)** biofilm thickness  $\lambda_0$ .

thickness were varied. A Pareto Front was approximated for each variation (data not shown), and for comparison, the reactor performance was also calculated for the constant flow rate. Initial suspended biomass  $u_0$  was 0.005, 0.02, 0.1 and 0.5 g, compared with the standard 0.05 g; initial biofilm thickness  $\lambda_0$  was 10, 50, 200 and 500  $\mu\text{m}$ , compared with the standard 100  $\mu\text{m}$ . Figure 7 shows the relative improvement in objective  $z_1$  when using optimal off-on flow rate functions compared with constant  $Q$ . The dashed lines denote the standard values for  $u_0$  and  $\lambda_0$  that were used throughout the previous simulations. Only one initial value was varied in each instance. We see in Figure 7a that the largest improvement is obtained for the smallest amount of initial suspended biomass. The improvement increases with increasing treatment time and the overall improvement ranges between 0.5% and 5%. The less suspended biomass is initially present, the more pronounced is the effect of the flow control compared to the uncontrolled system for process durations  $T > 2$ . This statement does not hold for smaller process durations, where the washout rate after the flow is turned on might be substantial. Increasing process duration for  $T > 3$  leads to an increasing improvement by control, up to  $T \approx 10$  when the reactor performance levels off.

The results for varying biofilm biomass are plotted in Figure 7b. The overall improvement ranges approximately between 0.5% and 12%. For a very thin initial biofilm thickness  $\lambda(0) = 10\mu\text{m}$  the largest relative improvement can be found for all  $T$ . The improvement increases monotonically with respect to  $T$  for the three thickest initial biofilms. However, for the thinner biofilm with  $\lambda(0) = 10\mu\text{m}$  we see first an increase to a local maximum around  $T = 3$  days after which a decrease occurs.



In both simulation experiments run here, the relative gain by control with off-on functions is biggest if the initial amount of biomass is smallest. This confirms the reasoning that led us to choose this type of control functions: The initial off-period is required for the bacterial populations to attain a sufficiently large population size to be efficient in substrate degradation.

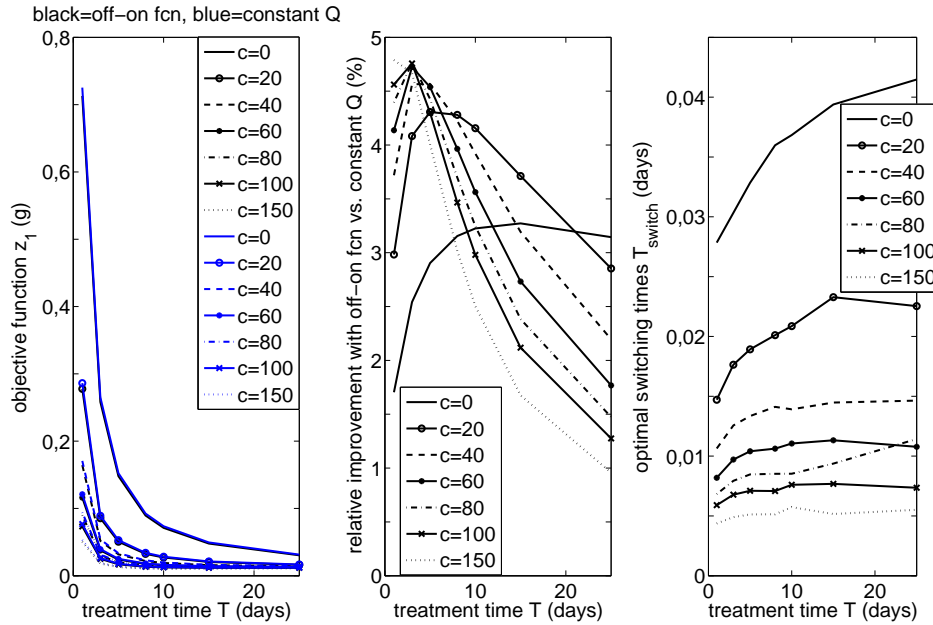


FIGURE 8. Effect of increased surface area on the controlled treatment process. For an increasing number of carrier chips  $c$  in the treatment vessel we show: **a)** Pareto Front of the off-on controlled system, along with the treatment performance of the uncontrolled reactor, **b)** Relative improvement of the controlled system compared to the uncontrolled system with, **c)** optimal switching time  $T_{switch}^*$ .

**Sensitivity with respect to biofilm growth area.** In the technical realization the wastewater treatment process can be accelerated by providing a larger substratum area  $A$  for biofilm to grow on. In Moving Bed Biofilm Reactors, this is achieved by adding suspended biofilm carrier chips into the completely mixed treatment vessel. To investigate the effect of control on such an improved reactor, we repeat the numerical control experiment for different numbers  $c$  of such carriers. The initial data were set to the default values again.

The results of this simulation experiment are plotted in Figure 8. The surface area was  $A_{min} = 0.0648\text{m}^2 \leq A \leq 1.08\text{m}^2 = A_{max}$ , corresponding to  $0 \leq c \leq 150$  carriers added. For  $T_{min} < T < T_{max} = 25$  we plot the Pareto Front along with the substrate removed by the uncontrolled ( $Q = const$ ) reactor in subplot a), the relative efficiency of the off-on controlled reactor compared to the uncontrolled reactor in subplot b), and the optimal switching time  $T_{switch}^*$  in subplot c). Qualitatively, the results for added carrier chips are the same as for the previously studied

case. As intended, adding biofilm growth area accelerates the treatment process, i.e. quantitative differences are observed. Increasing the biofilm growth area  $A$  makes treatment more efficient: the Pareto Fronts computed for larger values of  $c$  lie under the fronts for smaller values of  $c$ . For small treatment times  $T$ , adding carriers increases the efficiency of the controlled system compared to the uncontrolled one with  $Q = \text{const}$ . The improvement, however, remains modest, below 5%. For larger treatment times  $T$ , the improvement achieved by a carrier augmented controlled system compared to the uncontrolled system becomes smaller. The optimal switching time decreases with an increased number of carrier chips added.

**Discussion.** Using off-on functions to control the process, the largest improvement compared to an uncontrolled reactor was found for the mid-range treatment time of 10-15 days. For short treatment times the throughput in water is high and a substantial amount of substrate might be washed out. On the other hand, for long treatment times the bacteria are exposed to the substrate for a long period of time and can deplete to a great extent. A small improvement is obtained with the off-on functions by allowing the bacteria to establish themselves by consuming the substrate that is already in the reactor before the flow rate is turned on.

The dependency of the reactor performance on the initial suspended biomass and initial biofilm thickness was investigated. A significant difference could be observed between the smallest and largest amount of biomass initially present. Nevertheless, the relative improvement in reactor performance compared with the efficacy of the constant flow rate was smaller than 12%, with most cases ranging between 1% and 5%. Overall, the efficiency of controlling the process as implemented here is not great. Moreover, the initial state that has some effect on optimal reactor performance is very difficult to assess in applications, including controlled laboratory experiments. Therefore, for practical purposes it is doubtful that reactor performance can be reliably improved by controlling the flow rate, compared to the much simpler setup of emptying the buffer vessel at a constant flow rate.

**4.4. Sensitivity of reactor performance with respect to perturbations of the constant flow rate.** The previous results lead to the question how slight perturbations of the constant flow rate affect the reactor performance. To investigate this, we study for fixed  $T$  the effect of random perturbations of the constant flow rate  $Q$  on the amount of substrate in the effluent, i.e. objective  $z_1$ . For given  $T$  we calculate the constant flow rate  $Q = V_b/T$  and perform random perturbations of this value  $Q$  with up to 1%, 10% and 20%. To this end we divide the interval  $[0, T]$  into 10 subintervals in which the flow rate is constant. We always maintain  $\int_0^T Q(t)dt = V_b(0)$ . See Figure 9 for an example.

TABLE 2. Difference between maximum/minimum value of  $z_1$  for random perturbations of up to 1%, 10% and 20%, and the  $z_1$  obtained for a constant flow rate  $Q$ .

	1%	10%	20%
$ \max(z_1) - z_1(Q_{\text{const}}) /z_1(Q_{\text{const}})$ (%)	0.0513	0.6837	1.7189
$ \min(z_1) - z_1(Q_{\text{const}}) /z_1(Q_{\text{const}})$ (%)	0.0620	0.4688	0.7439

For each percentage class we randomly choose  $N = 1000$  such perturbations of the constant flow rate. In Figure 10 we plot the histograms of the reactor performance

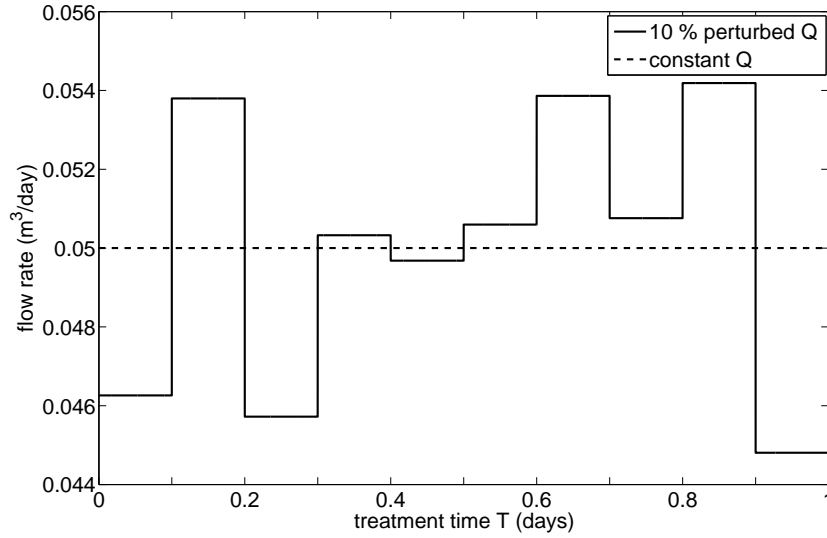


FIGURE 9. Random perturbation of the constant flow rate  $Q$  (dashed) by up to 10% (solid) as functions of the treatment time.

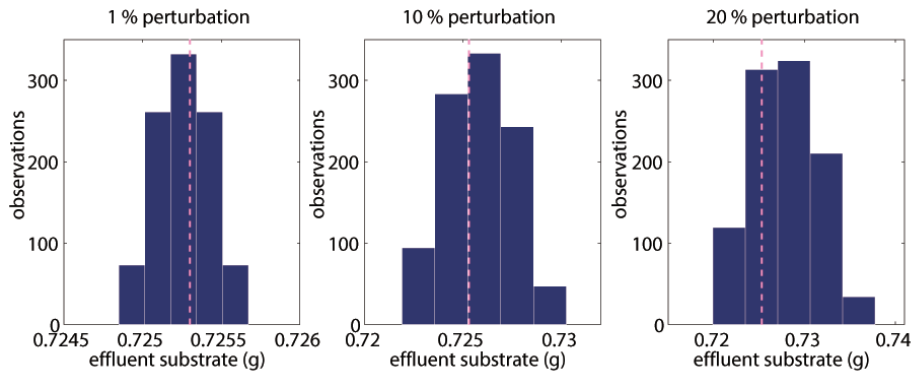


FIGURE 10. Values of objective function  $z_1$  for random perturbations of the constant flow rate  $Q$  with up to 1%, 10% and 20%, represented by histograms with five bins. The vertical dashed line marks the value of  $z_1$  obtained with constant  $Q$ . The value of  $z_1$  obtained with the optimal off-on function was 0.7129.

index. Table 2 shows the largest distances from  $z_1(Q_{const})$  for each of the percentage classes. Larger perturbations cause a wider range in  $z_1$ , i.e. the amount of untreated substrate in the effluent. However, this range is not symmetrically centered around the value obtained with constant  $Q$ . The higher the perturbation, the more the respective histograms appear skewed with respect to the constant flow rate. A larger perturbation is more likely to worsen the value of  $z_1$  than to improve it. This

effect can also be seen for the larger perturbations in Table 2, where the differences between the maximum values and constant  $Q$  are larger than the differences between the minimum values and constant  $Q$ . Nevertheless, the largest changes in  $z_1$  are 1-2 orders of magnitude smaller than the actual perturbations in  $Q$ , i.e. reactor performance is robust with respect to small deviations from the constant flow rate. In all cases, the optimal reactor performance obtained with off-on function control is lower than the results obtained in this experiment.

## 5. Other control functions.

**5.1.  $Q$  switching between 0 and  $Q_{max}$  in discrete time intervals.** Inspired by the chattering control for measurable, bounded  $Q$ , we now define  $\Omega$  as the set of non-negative piecewise constant functions, that can switch between 0 and  $Q_{max}$  a finite number of times. For a fixed final time  $T$  the interval  $[0, T]$  is divided into  $n$  equidistant subintervals  $I_i := [T_{i-1}, T_i)$  with  $T_0 < T_1 < \dots < T_n$ . In each subinterval, the function  $Q$  can either take the value 0 or  $Q_{max}$ . Moreover, it has to satisfy the reactor condition  $\int_0^T Q(t)dt = V_b(0)$ . Thus, the number  $m$  of intervals in which  $Q = Q_{max}$  is then  $m := nV_b(0)/(TQ_{max})$ , where the parameters need to be such that  $m < n$  is an integer. We refer to this class of functions as *zero-max* functions  $\Omega_Z$ .

**Numerical method.** An initial guess for  $Q$  in the class of zero-max functions is chosen randomly. We employ a greedy algorithm [8] to find the optimal solution with respect to the objective function  $\int_0^T Q S dt$ . The greedy algorithm explores the neighborhood of the current best guess, and makes a locally optimal choice to update the best guess. This process is iterated until no further improvements are possible. For this purpose, we define a neighborhood of a zero-max function as the set of all zero-max functions which differ from the current function in exactly one interval with  $Q = 0$  and one with  $Q = Q_{max}$ .

The Greedy Algorithm requires numerous solutions of the reactor model and is, therefore, computationally extremely expensive. The lion's share of computing time is here incurred by the computation of the substrate flux into the biofilm, requiring the solution of a two-point-boundary value problem. To ease this and make computations feasible, we use the flux approximation (8) instead of (5). The underlying model then, in fact, becomes the Freter model [10, 18], albeit with different detachment and re-attachment terms.

**Results.** In Figure 11a we plot a randomly chosen initial guess for  $Q$  along with the optimal solution found by the Greedy Algorithm after 104 local decisions in Figure 11b, for  $T = 10$ ,  $Q_{max} = 0.01$ ,  $n = 50$ ,  $m = 25$ . The optimal solution starts with a short off-period with  $Q = 0$ , switches to  $Q = Q_{max}$  until the reactor is emptied and then switches back to  $Q = 0$ . The Greedy algorithm by construction finds a local minimum. Therefore, we repeated the simulation five times with different randomly chosen initial data. In all cases the optimal solution found was the same.

Within the class of zero-max functions tested here, the optimal control function is the one that resembles the optimal off-on function found previously the most. Of course, the choice of switching time here is constrained by the discretization of the interval  $[0, T]$ . For comparison, we computed also the optimal, grid independent switching time from 0 to  $Q_{max}$  where it stays for a period of  $V_b(0)/Q_{max}$ . This value is found to be  $t = 0.1663$  days, within the first subinterval of the time grid. To assess the efficiency of this control we computed for comparison the optimal off-on function (33), but with the flux approximation (8) instead of (5). The optimal

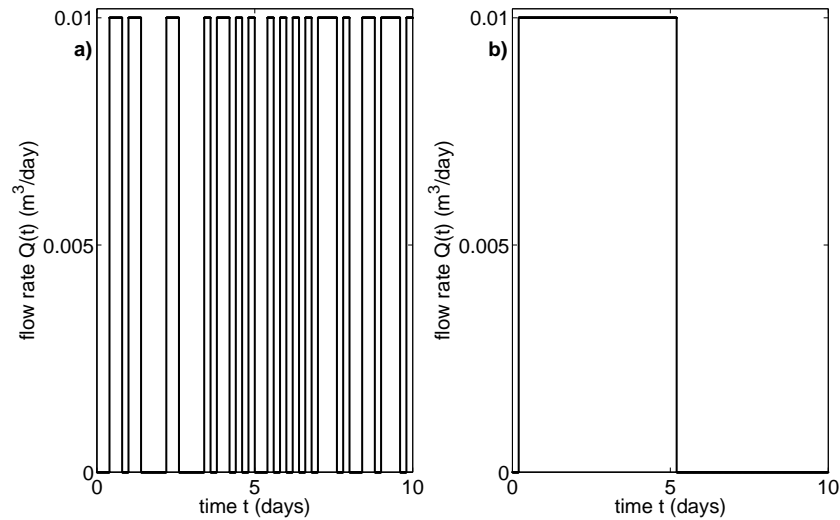


FIGURE 11. **a)** Initial guess of the piecewise constant flow rate  $Q(t)$  with randomly distributed switches between the lower and upper bound. **b)** Solution obtained with the greedy algorithm after 104 steps.

reactor performance for zero-max functions with  $m = 25$  was found to be about 1.5 times higher than the optimal reactor performance found for functions of type (33). This could be improved by refining the time grid, which however, would incur a substantial increase in computing time.

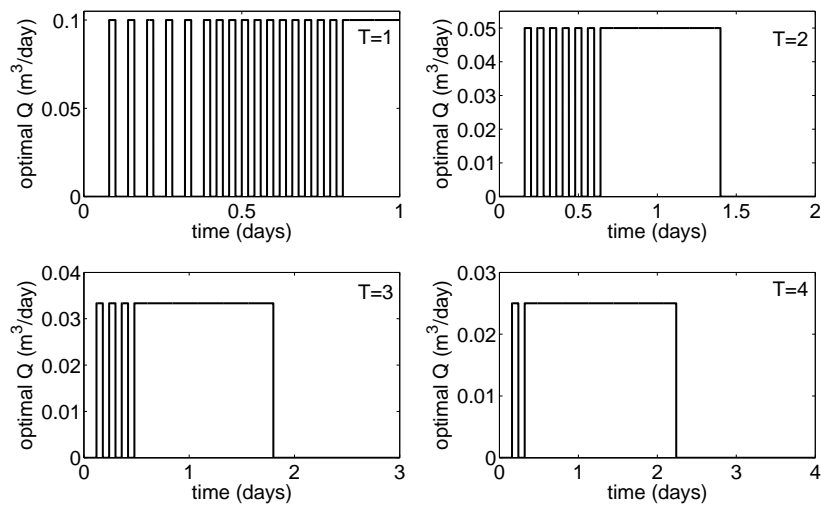


FIGURE 12. Optimal zero-max functions found with the greedy algorithm for different values  $T < 5$ .

We repeated these simulations for other values of  $T$ , keeping the number of subintervals of  $[0, T]$  the same. The result obtained for values  $T \geq 5$  are qualitatively the same as found for  $T = 10$  and documented above. In all cases the optimal control switches from 0 to  $Q_{max}$  after the first subinterval and stays there until the reactor is empty. The reactor performance is worse than for the optimal off-on functions.

For smaller  $T$ , on the other hand, the optimal zero-max control function is distinctively different, see Figure 12. It is characterized by switching several times between 0 and  $Q_{max}$  for small  $t \ll T$ . The number of such jumps in the control functions increases for smaller  $T$ , and is in part due to the smaller subintervals. Another important aspect is the increase of  $Q_{max}$  for smaller  $T$  due to  $m = 25$  being kept the same. In Figure 12  $Q_{max}$  varies between 0.1 and 0.025 for  $T = 1$  and  $T = 4$ , respectively. We performed additional simulations with a fixed  $Q_{max} = 0.05$  and a fixed number of max-intervals  $m = 25$  for  $T = 2$  and  $T = 4$ . In order to obtain 25 max-intervals the number of subintervals for  $T = 4$  was increased to  $n = 100$ . The optimal zero-max control function found by the greedy algorithm is the same for these two treatment times up to  $T = 2$ , see Figure 13. Thereafter, for  $T = 4$  the function remains at 0 until the end. Thus, it seems that the value of  $Q_{max}$  strongly influences the optimal zero-max solution rather than the length of the treatment time. The optimal zero-max solutions for small  $T$  are characterized by many rapid switches between 0 and the maximum feasible flow rate, mimicking the chattering control with infinitely many switching points that characterizes the solutions to the original infinite-dimensional control problem (22). In this finite-dimensional setup, the number of such switches possible is bounded by the number of degrees of freedom  $n$  and the minimum duration of a single on- or off-interval is bounded by  $T/n$ .

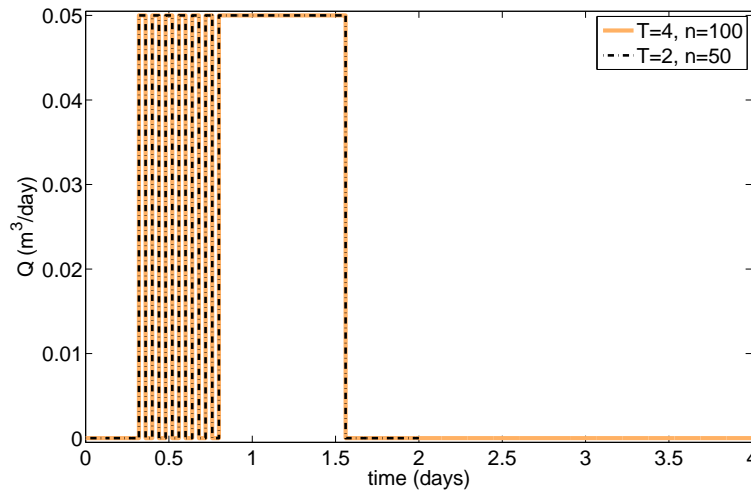


FIGURE 13. Optimal zero-max functions found with the greedy algorithm for  $Q_{max} = 0.05\text{m}^3/\text{day}$  and  $T = 2$  and 4 days.

The locally optimal solution found with the greedy algorithm can depend on the initial data. However, there is not much variation in reactor performance between such local optima. The Pareto Front obtained with zero-max functions with a

fixed number of subintervals, as described here, lies above the Pareto Front of the optimal off-on functions, and above the reactor performance achieved with constant flow rate, see Figure 14, implying a worse control of the reactor, even compared to the uncontrolled constant flow rate. Furthermore, the increased number of switches between 0 and  $Q_{max}$  in Figure 12 is unrealistic for practical applications.

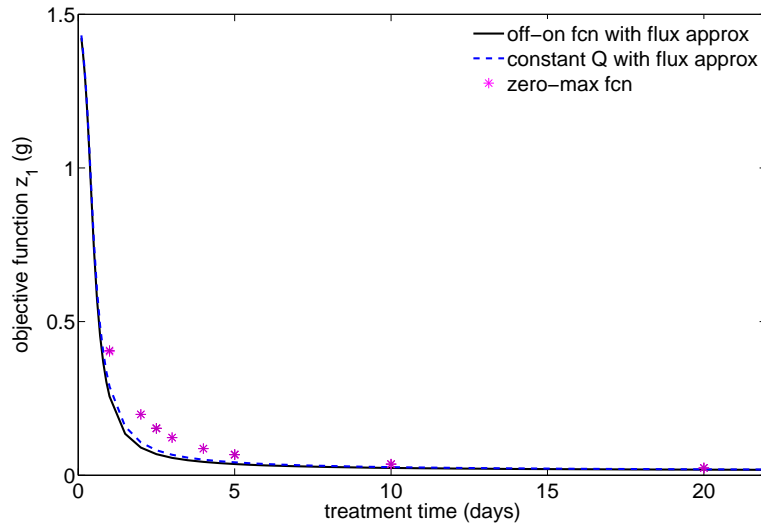


FIGURE 14. Pareto Fronts for zero-max functions and off-on functions with flux approximation (8). For comparison the data obtained with the constant flow rate for varying  $T$  are included.

**5.2. Optimal control software.** Optimal control problems with singular arcs, state constraints and discontinuous controls are known challenges for general purpose optimal control software [5]. Such software packages, in particular those based on the *direct method*, can be understood loosely as searching for the optimal control within a finite dimensional class of functions, usually described by the number of basis functions used to approximate the optimal control, e.g. the number of subintervals into which  $[0, T_{max}]$  is divided. For comparison purposes, we applied two such general purpose optimal control solvers to the problem at hand in Section 3: ACADO Toolkit (Version 1.0.2613beta, June 2011, [14]) and DIDO (Version 7.3.6, [30]). As in Section 5.1 also here the flux approximation was used. We were not able to obtain a robustly converging solution with the former software, probably because of the difficulties to mimic a chattering control. The latter, instead of computing the solution, is designed to determine candidates for solutions which are then to be subjected to further tests. In Figure 15 we plot two such candidates that we obtained for  $T = 10$  days, discretized with 50 and 100 nodes, with  $0 \leq Q \leq 0.05$   $\text{m}^3/\text{day}$ . There are oscillations in the beginning and in the end of the treatment period with a relatively smooth arc in between. The amount of oscillations is increased for the larger number of discretization nodes and the largest value attained by  $Q$  is  $\approx 0.016 \ll 0.05$   $\text{m}^3/\text{day}$ . Here, as for the off-on and zero-max functions, the flow rate is turned off for a few time segments in the initial phase of the treatment. As

in the case of zero-max functions, the oscillatory patterns probably might be interpreted as an attempt to mimic chattering in the finite dimensional class of control functions and the frequency of the oscillations increases with the number of degrees of freedom. Moreover, along the trajectory found we observe for the Hamiltonian  $H \approx 0$ , which in accordance with (31) might be taken as an indication that the computed solution is optimal for a neighboring regularized problem. To investigate this, however, is beyond the scope of this study. None of the numerical candidate functions that we were able to obtain for the problem at hand were competitive, with respect to reactor performance, with the much simpler optimal off-on function of Section 4.1.

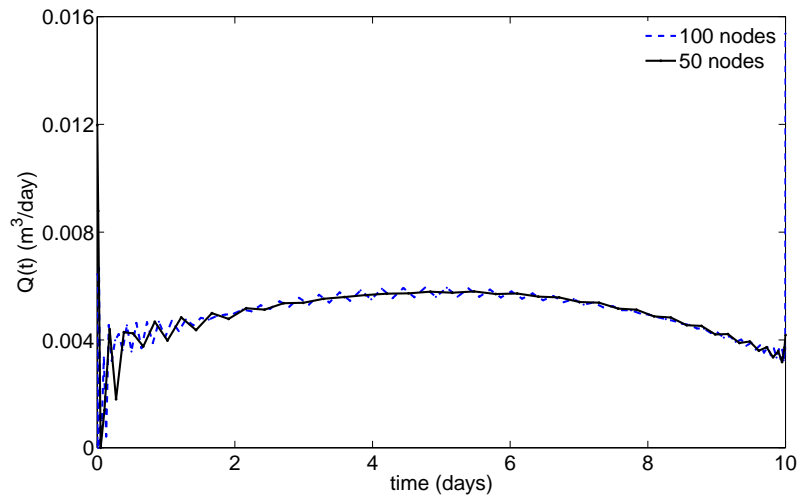


FIGURE 15. Optimal control  $Q$  found for the optimal control problem (14) with the flux approximation (8) by DIDO for 50 (solid black) and 100 (dashed blue) time discretization nodes of  $T = 10$  days, where  $0 \leq Q \leq 0.05 \text{ m}^3/\text{day}$ .

**5.3. Remarks on computational aspects.** One can argue that a finer discretization of the intervals  $[0, T]$  or  $[0, T_{max}]$  shall improve the performance of the approaches discussed in this section and that, eventually, the results might be competitive with the much simpler off-on function of section 4.1. Such a refinement, however, goes hand in hand with an increase in computing time required for the problem. For example, the number of neighbors to be computed in each iteration of the algorithm in Section 5.1 scales quadratically with the number of subintervals into which  $[0, T]$  is divided.

On a personal computer, the time to compute a single point of the Pareto Front with off-on functions, using the solution of the two-point boundary value problem for the substrate concentration (5) to compute the substrate flux into the biofilm, is in the order of minutes. Using instead the flux approximation (8) is in the order of seconds. On the other hand, computing the best zero-max function with the substrate flux approximation for a given  $T$  is in the order of hours; moreover, the solution depends on the initial guesses, i.e. the problem needs to be solved several



times. Similarly, the time it takes to solve for a single weight  $w$  the optimal control problem with optimal control software is in the order of days. These numbers will multiply drastically if the full two-point boundary value problem is solved to compute substrate fluxes. Taking these considerations into account, refining the time discretization in the methods of Sections 5.1, 5.2 to achieve results comparable to or better than the results achieved with the much simpler off-on function of Section 4.1 becomes prohibitive.

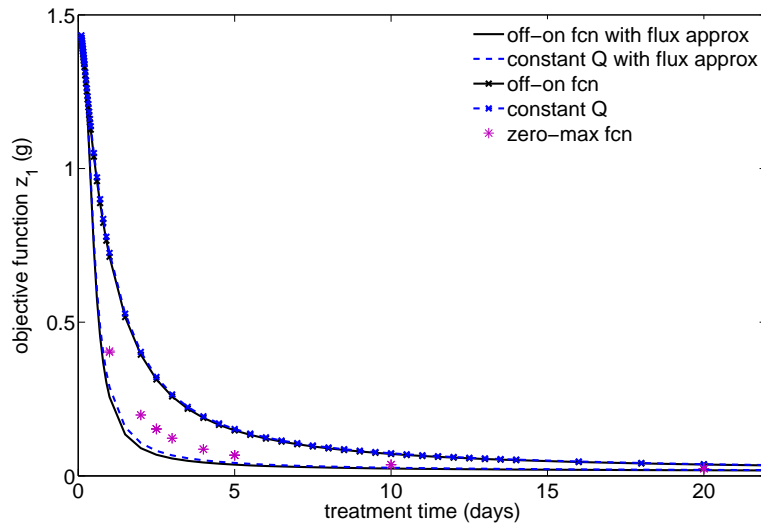


FIGURE 16. Pareto fronts for off-on functions computed with (8) and (5), substrate on effluent for constant flow rate  $Q = V_b(0)/T$  computed with (8) and (5), and the Pareto front for zero-max functions with (8).

Furthermore, the flux approximation (8) was used in Sections 5.1 and 5.2 to replace (5) because it is computationally much cheaper. It has been shown previously that this formula is able to give a good approximation for a wide range of realistic biofilm model parameters and a wide range of biofilm thicknesses and substrate concentrations. The flux approximation of the Monod kinetics resembles the flux of a zero-order kinetics biofilm model [2]. This is likely a good approximation in the very early stages of the control period, where the substrate concentration is still relatively large and the biofilm thickness small [20]. On the other hand, the goal of the optimization is to reduce the substrate concentration, i.e. to reach low substrate concentrations  $S(t)$ . At low substrate concentration, e.g. in the later stages of the control interval, Monod kinetics is known to be better approximated by first order kinetics [20]. To investigate the suitability of the approximation (8) for the problem at hand, we compare in Figure 16 the Pareto fronts for off-on functions computed with (8) and (5), substrate in effluent for constant flow rate  $Q = V_b(0)/T$  computed with (8) and (5), and the Pareto front for zero-max functions with (8). These results indicate, that substrate in effluent computed with the flux approximation overestimates substrate removal. This mandates that the computationally much more expensive (5) should be used instead of (8). In the context of the approaches

of Section 5.1, 5.2, however, this is prohibitive. Therefore, zero-max functions have no advantages compared to the off-on functions, which yield better solutions to the optimal control problem and are computationally more efficient.

**6. Conclusion.** In this paper, we formulated and studied an optimal control problem for a bioreactor with biofilm and suspended biomass with a limited amount of substrate and treatment time. The bioreactor model is a biofilm extension of the Freter model with wall attachment. For a one-dimensional single species single substrate model, the optimal control problem was formulated with two objectives: to minimize the treatment time and the amount of substrate in the effluent. The flow rate between the buffer reactor and the bioreactor was selected as the control for the problem. All results were compared against the constant flow rate, which characterizes the uncontrolled treatment. For computational reasons, of particular interest were the off-on functions, which allow the biomass in the bioreactor to establish itself while the flow is turned off before the flow is turned on and the buffer reactor is emptied. We reached the following conclusions:

- The Pareto Front for off-on functions shows that for small treatment times the optimal removal efficiency of the reactor changes drastically, whereas it levels off for larger  $T$ . For large  $T$  almost all the substrate can be removed, whereas for small  $T$  a considerable amount leaves the reactor untreated.
- The optimal switching times for the off-on functions, at which point the flow is turned on, are small relative to the treatment time. Furthermore, the improvement in reactor efficiency, compared with the constant flow rate, is modest.
- Other classes of admissible functions, such as measurable and bounded or measurable and unbounded control functions, led to solutions that are inapplicable/infeasible from a practical point of view.
- Calculations with an approximated expression for the substrate flux into the biofilm, used in order to reduce computational complexity, overestimated the substrate removal in the reactor.
- The potential to improve reactor performance is due to an initial transient period in which the bacterial populations adapt to the environmental conditions in the reactor. Therefore, control potential depends on the initial state of the system which is typically not known, even in controlled laboratory experiments.
- In summary, we come to the conclusion, that the performance increase that is gained by controlling the reactor flow rate is not a significant improvement compared to the uncontrolled, constant flow rate. Moreover, it was shown that the reactor performance is very robust with respect to small deviations from the constant flow rate.

These results are based on a comprehensive numerical study for which several hundreds of scalar optimization problems were solved, each of which required numerous simulations of the underlying mathematical model. In these simulations, a small laboratory reactor was mimicked and a generic set of reaction parameters was used. It remains to be investigated whether these results carry over to other setups as well, such as full scale treatment reactors.

**Acknowledgments.** This work was carried out while A.M. was Ph.D. student at the Centre for Mathematical Sciences, Lund University, Sweden. The authors

would like to thank Brian Ingalls (University of Waterloo, Canada) and Andrey Ghulchak (Lund University, Sweden) for their helpful comments and discussions, and the Centre for Mathematical Sciences at Lund for their support.

## REFERENCES

- [1] F. Abbas, R. Sudarsan and H. J. Eberl, [Longtime behavior of one-dimensional biofilm models with shear dependent detachment rates](#), *Math. Biosc. Eng.*, **9** (2012), 215–239.
- [2] F. Abbas and H. J. Eberl, [Analytical substrate flux approximation for the Monod boundary value problem](#), *Appl. Math. Comp.*, **218** (2011), 1484–1494.
- [3] M. M. Ballyk, D. A. Jones and H. L. Smith, [The biofilm model of Freter: A review](#), in *Structured Population Models in Biology and Epidemiology*, Lecture Notes in Math., 1936, Springer, Berlin, 2008, 265–302.
- [4] M. M. Ballyk, D. A. Jones and H. L. Smith, Microbial competition in reactors with wall attachment, *Microb. Ecol.*, **41** (2001), 210–221.
- [5] J. T. Betts, *Practical Methods for Optimal Control and Estimation using Nonlinear Programming*, SIAM series Adv. Design and Control, Philadelphia, 2010.
- [6] C. Carathéodory, *Vorlesungen über reelle Funktionen*, 3rd edition, Chelsea Publ, 1968.
- [7] N. G. Cogan, B. Szomolay and M. Dindo, [Effect of periodic disinfection on persisters in a one-dimensional biofilm model](#), *Bull. Math. Biol.*, **75** (2013), 94–123.
- [8] T. H. Cormen, C. E. Leiserson, R. L. Rivest and C. Stein, *Introduction to Algorithms*, 2nd edition, MIT Press, 2001.
- [9] W. H. Fleming and R. W. Rishel, *Deterministic and Stochastic Optimal Control*, Springer-Verlag New York Inc, 1975.
- [10] R. Freter, H. Brickner, J. Fekete, M. Vickerman and K. Carey, Survival and implantation of Escherichia coli in the intestinal tract, *Infect. Immun.*, **39** (1983), 686–703.
- [11] P. Gajardo, J. Harmand, H. Ramírez C. and A. Rapaport, [Minimal time bioremediation of natural water resources](#), *Automatica*, **47** (2011), 1764–1769.
- [12] A. Göpfert and R. Nehse, *Vektoroptimierung*, BSB Teubner Verlagsgesellschaft, Leipzig, 1990.
- [13] E. V. Grigorieva and E. N. Khailov, [Minimization of pollution concentration on a given time interval for the waste water cleaning plant](#), *J. Control Sci. Eng. Article*, **2010** (2010), 1–10.
- [14] B. Houska, H. J. Ferreau and M. Diehl, [ACADO toolkit – An open-source framework for automatic control and dynamic optimization](#), *Optim. Contr. Appl. Met.*, **32** (2010), 298–312.
- [15] I. Ivanovic and T. O. Leiknes, [Particle separation in Moving Bed Biofilm Reactor: Applications and opportunities](#), *Separ. Sci. Technol.*, **47** (2012), 647–653.
- [16] J. Jahn, *Vector Optimization: Theory, Applications and Extensions*, 2nd edition, Springer-Verlag, Berlin Heidelberg, 2011.
- [17] T. L. Johnson, J. P. McQuarrie and A. R. Shaw, [Integrated Fixed-film Activated Sludge \(IFAS\): The new choice for nitrogen removal upgrades in the United States](#), *Proceedings WEFTEC Session*, (2004), 296–318.
- [18] D. Jones, H. V. Kojouharov, D. Le and H. Smith, [The Freter model: A simple model of biofilm formation](#), *J. Math. Biol.*, **47**, (2003), 137–152.
- [19] J. B. Kaplan, [Biofilm dispersal: Mechanisms, clinical implications, and potential therapeutic uses](#), *J. Dent. Res.*, **89** (2010), 205–218.
- [20] I. Klapper, [Productivity and equilibrium in simple biofilm models](#), *Bull. Math. Biol.*, **74** (2012), 2917–2934.
- [21] Z. Lewandowski and H. Beyenal, *Fundamentals of Biofilm Research*, CRC Press, Boca Raton, 2007.
- [22] A. Mašić and H. Eberl, [Persistence in a single species CSTR model with suspended flocs and wall attached biofilms](#), *Bull. Math. Biol.*, **74** (2012), 1001–1026.
- [23] A. Mašić and H. Eberl, [A modeling and simulation study of the role of suspended microbial populations in nitrification in a biofilm reactor](#), *Bull. Math. Biol.*, **76** (2014), 27–58.
- [24] The Mathworks, MATLAB online documentation, <http://www.mathworks.com/help/matlab/>, accessed on January 18, 2013.
- [25] J. Moreno, Optimal time control of bioreactors for the wastewater treatment, *Optim. Control Appl. Meth.*, **20** (1999), 145–164.
- [26] E. Morgenroth, M. C. M. van Loosdrecht and O. Wanner, Biofilm models for the practitioner, *Water Sci. Technol.*, **41** (2000), 509–512.

- [27] L. S. Pontryagin, N. G. Boltyanski, R. V. Gamkrelidze and E. F. Mishchenko, *The Mathematical Theory of Optimal Processes*, Wiley, New York, 1962.
- [28] B. E. Rittmann and P. L. McCarty, *Environmental Biotechnology*, McGraw-Hill, New York, 2001.
- [29] M. E. Roberts and X.-M. Zhan, Moving-medium Biofilm Reactors, *Rev. Environ. Sci. Biotechnol.*, **2** (2003), 213–224.
- [30] I. M. Ross, *A beginner's guide to DIDO: A MATLAB application package for solving optimal control problems*, Elissar Global, Monterey, CA, 2007.
- [31] I. Y. Smets and J. F. Van Impe, [Optimal control of \(bio-\)chemical reactors: Generic properties of time and space dependent optimization](#), *Math. Comput. Simulat.*, **60** (2002), 475–486.
- [32] E. D. Stemmons and H. L. Smith, [Competition in a chemostat with wall attachment](#), *SIAM J. Appl. Math.*, **61** (2000), 567–595.
- [33] B. Szomolay, I. Klapper and M. Dindos, [Analysis of adaptive response to dosing protocols for biofilm control](#), *SIAM J. Appl. Math.*, **70** (2010), 3175–3202.
- [34] M. von Sperling, *Activated Sludge and Aerobic Biofilm Reactors*, IWA Publishing, London, 2007.
- [35] W. Walter, *Gewöhnliche Differentialgleichungen*, 7th ed., Springer, 2000.
- [36] O. Wanner, H. Eberl, E. Morgenroth, D. R. Noguera, C. Picioreanu, B. Rittmann and M. van Loosdrecht, *Mathematical Modeling of Biofilms*, Scientific and Technical Report No.18, IWA Publishing, 2006.
- [37] M. I. Zelikin and V. F. Borisov, *Theory of Chattering Control with Applications to Astronautics, Robotics, Economics, and Engineering*, Systems & Control: Foundations & Applications, Birkhäuser Boston, 1994.

Received February 10, 2013; Accepted March 12, 2014.

*E-mail address:* [alma.masic@eawag.ch](mailto:alma.masic@eawag.ch)

*E-mail address:* [heberl@uoguelph.ca](mailto:heberl@uoguelph.ca)