

MODELING AND SIMULATION FOR TOXICITY ASSESSMENT

CRISTINA ANTON*

Department of Mathematics and Statistics, Grant MacEwan University
Edmonton, Alberta, T5P2P7, Canada

JIAN DENG AND YAU SHU WONG AND YILE ZHANG

Department of Mathematical and statistical Sciences, University of Alberta
Edmonton, Alberta, T6G2G1, Canada

WEIPING ZHANG

Alberta Health, Edmonton, Alberta, T5J1S6, Canada

STEPHAN GABOS

Department of Laboratory Medicine and Pathology, University of Alberta
Edmonton, Alberta, T6G2B7, Canada

DOROTHY YU HUANG

Alberta Centre for Toxicology, University of Calgary
Calgary, Alberta, T2N4N1, Canada

CAN JIN

ACEA Biosciences Inc, San Diego, California, 92121, USA

(Communicated by Jianhong Wu)

ABSTRACT. The effect of various toxicants on growth/death and morphology of human cells is investigated using the xCELLigence Real-Time Cell Analysis High Throughput in vitro assay. The cell index is measured as a proxy for the number of cells, and for each test substance in each cell line, time-dependent concentration response curves (TCRCs) are generated. In this paper we propose a mathematical model to study the effect of toxicants with various initial concentrations on the cell index. This model is based on the logistic equation and linear kinetics. We consider a three dimensional system of differential equations with variables corresponding to the cell index, the intracellular concentration of toxicant, and the extracellular concentration of toxicant. To efficiently estimate the model's parameters, we design an Expectation Maximization algorithm. The model is validated by showing that it accurately represents the information provided by the TCRCs recorded after the experiments. Using stability analysis and numerical simulations, we determine the lowest concentration of toxin that can kill the cells. This information can be used to better design experimental studies for cytotoxicity profiling assessment.

2010 *Mathematics Subject Classification.* Primary: 93A30, 37N25; Secondary: 60G35.
Key words and phrases. Mathematical model, cytotoxicity, parameter estimation, persistence.

1. Introduction. Industrial chemicals might be environmental or health hazards, thus there is an increasing need for developing efficient methodologies for toxicity assessment. The traditional method of conducting *in vivo* assays is very efficient for understanding the mode of action of various toxicants, but it is expensive, time consuming, and there are moral issues regarding the use of animals. Cell-based *in vitro* assays [28] are attractive alternatives for the assessment of toxicity. We propose a mathematical model to study the effect of toxicants on human cells.

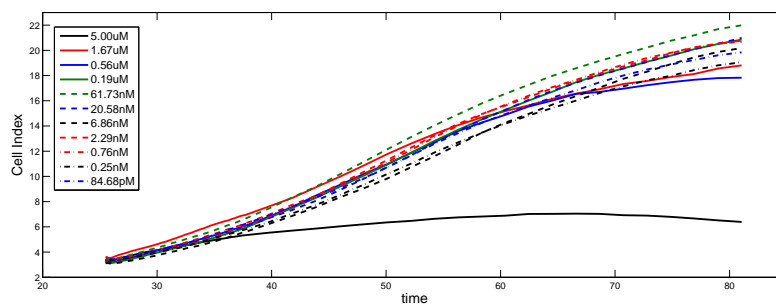
Our work is part of the cytotoxicity profiling project carried by Alberta Centre for Toxicology in which initially 63 chemicals [27] were investigated using the xCELLigence Real-Time Cell Analysis High Throughput (RTCA HT) Assay [19]. The xCELLigence RTCA HT system is developed by the ACEA Biosciences Inc. (San Diego, USA), utilizing 384 well electronic microtiter plates (E-Plates 384). Impedance is generated by cells attached to electrodes, impeding the electric current between electrodes. Therefore, the impedance signal is label free and non-invasive, enabling direct measurement of cellular status in real time. The impedance value is converted by a software to Cell Index, which closely reflects not only cell number, but also cell morphology and adhesion.

The cells were seeded into the wells of E-Plates 384 [19], and the test substances were added about 24 hours later to allow cells to attach to the bottom of the wells and adapt to growth. The 11 concentrations of the test substances were selected based on a literature study, and each of them was tested in at least quadruplicates [27]. Time-dependent concentration response curves (TCRCs) for each test substance in each cell line were generated [27].

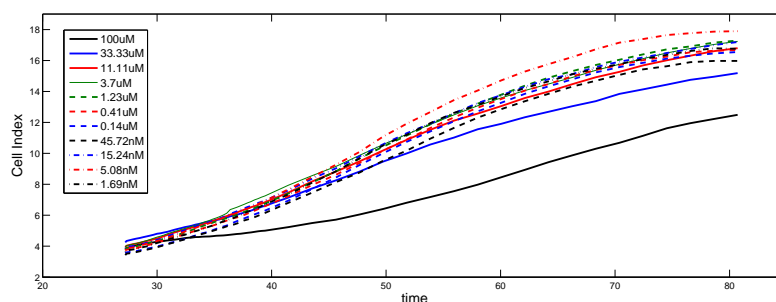
Series of physiological events, such as changes in cell population, cell morphology, and cellular functions, that characterize an adverse biological response are referred to as mode of action (MOA) [27]. According to the MOA, the tested substances were divided in 10 groups [27]. For example, the TCRCs for two chemicals from the same group, PF431396 and monastrol, are displayed in Fig. 1.

The main objectives of our work are to develop a mathematical model capable of reproducing the TCRCs and, based on this model, to investigate the effects of toxicants on the survival/death of the cells' population. We consider a three dimensional model for a cells' population subjected to an acute dose of toxicant. The state variables are the cell index, as a proxy for the number of cells, the concentration of toxicant in the cell, and the concentration of toxicant in the environment, and they are coupled by a linear dose-response function. This model is a special case of the class of models proposed in [4], and it is related to the models considered in [5, 10, 15]. However, since we consider an acute dose of toxicant instead of a chronic one, the analysis of the survival/death of the cells' population is different from the one done in the previously mentioned papers. We analyze the stability of this system of ordinary differential equations analytically using an approach similar with the one in [7], and we find sufficient conditions for the survival or death of the cells' population.

An important contribution of this paper is the efficient statistical methodology used to validate the proposed mathematical model. For parameter estimation and prediction we propose an approach based on the Expectation Maximization (EM) algorithm [3] and the Unscented Filter [12]. This method can also be used to predict the concentration of toxicant outside the cells. These predictions are important in any system for toxicity monitoring. For example, a similar cytotoxicity mathematical dynamic model [8] is applied for early detection and quantification of the



(a)



(b)

FIGURE 1. TCRCs for (a) PF431396 and (b) monastrol

presence of toxicants in the water supply system. The estimation of the model parameters and prediction of the concentrations of the toxic agent presented in water supply are done using the moving horizon estimator (MHE) and extended Kalman filter (EKF). Although the model considered in [8] is less complicated than the model presented in this paper, for parameter estimation the MHE and the EKF require extensive tuning, so the EM algorithm is a much more efficient alternative.

Since the parameters of the proposed model are estimated not measured directly, there is uncertainty regarding their precise values. We use Monte Carlo simulations and parameter sensitivity analysis for uncertainty quantification. Sensitivity analysis [13] refers to a broad group of methods that provide valuable insight regarding the dependence of the model output on its parameters. We apply a global sensitivity analysis (GSA) method [23] to test the robustness of the results of our model in the presence of uncertainty, and to increase the understanding of the relationships between input and output variables in the model.

In addition to proposing an efficient statistical method for parameter estimation and prediction of the concentration of toxicant outside the cells, an important application of the mathematical model is to generate numerically the TCRCs. Based on these numerical simulations and a mathematical analysis of the model, we can determine the threshold value for the initial concentration of a chemical after which the death of the cells' population happens (i.e. the chemical considered becomes toxic for the cells).

From the information included in the TCRCs, various parameters, such as LC_{50} [14], KC_{50} [29, 20], and AUC_{50} [19], were used to describe the time and concentration dependent cellular activity, and they were used in various statistical methods for clustering [27]. The success of clustering and classification methods, such as neural networks and support vectors machines, depends on providing TCRCs corresponding to a range of values of the initial concentration of the toxicant that include both cases when the concentration is low enough to not affect the cells' population, and cases when the concentration is high enough to determine the eventual death of the cells' population ([20, 27]). For example, in Fig. 1 (b) none of the concentrations used for monastrol was high enough to kill cells. Using the proposed mathematical model we can determine an appropriate range for the initial concentration of the toxicant. Our methodology based on stability analysis has better mathematical justification and gives more accurate results than the approach proposed in [20] and based on LC_{50} and linear regression.

The paper is organized as follows. We start by presenting the mathematical model in section 2 followed by the study of stability in section 3. Section 4 includes the methods used for parameter estimation and prediction. Next, in section 5, we validate the proposed model by comparing the experimental TCRCs with the predicted values for the cell index. The GSA and Monte Carlo Simulations are presented in section 6. The last section of the paper includes a summary and discussion of the main results and intended future work.

2. The mathematical model. The data set used in this study is from the cytotoxicity test on human hepatocellular carcinoma cell line HepG2 [19].

Without exposure to the test substances, a typical cell growth curve exhibits four phases: the lag, the log, the plateau and the decline phases. The lag phase is the phase when the cells do not divide and adapt to the culture conditions. In the log phase the cells actively proliferate and have an exponential increase in the cell density. In the plateau phase, the cellular proliferation reaches a steady-state level. Finally, in the decline phase, the cell death predominates and the number of viable cells declines due to the natural path of the cell cycle and depletion of the nutrient supplements.

We focus on the modeling of the log phase. After seeding the cells in E-Plates for 24 hours, eleven concentrations of each substance were applied. Negative controls were presented in each E-Plate 384, as the cells in these wells were treated with only assay buffer for substance dilution, but no testing substances. The cells were monitored totally for 89 hours. The TCRCs are truncated such that only the data from the log phase are kept.

To model the negative control curves, we start with the well-known logistic model describing cell growth in non-toxic environment [24]:

$$\frac{dn(t)}{dt} = \beta n(t) \left(1 - \frac{n(t)}{K}\right), \quad (1)$$

where $n(t)$ is the cell index at time t , which closely reflects the number of cells, β denotes the cell growth rate, and K is the capacity volume. The initial stage of growth is approximately exponential; then, as the population is closed to capacity K , the growth slows.

We consider the following model for cells growth/death in a toxic environment [4]. Let $C_0(t)$ and $CE(t)$ be the the intracellular and the extracellular concentrations of toxicant at time t , respectively. We suppose the death rate of cell is linearly

dependent on the intracellular concentration of toxicant with coefficient $\alpha > 0$, so we have

$$\frac{dn(t)}{dt} = \beta n(t) \left(1 - \frac{n(t)}{K}\right) - \alpha C_0(t)n(t). \tag{2}$$

We suppose that the only source of toxicant is the environment. Linear kinetics is known to be relevant for certain toxicants, but there is evidence that higher order formulations are also important [4]. In [6] the mechanisms of in vitro cytotoxicity are analyzed, and in the proposed models the toxicant’s uptake rate is considered to be a combination of a linear diffusive component and a saturable, carrier-mediated component (Michaelis-Menten like kinetics). The rapid on-line estimation of toxicant concentration in water supply proposed in [8] is also based on these models. However, after estimating the parameters, for some toxicants it appears that the Michaelis-Menten uptake is not significant [6], so we ignore carrier-mediated transportation and we use linear kinetics [4, 24]:

$$\frac{dC_0(t)}{dt} = \lambda_1^2 CE(t) - \eta_1^2 C_0(t), \tag{3}$$

where λ_1^2 represents the uptake rate of the toxicant from environment, and η_1^2 is the toxicant input rate to the environment.

As in [4] we assume that the cells absorb part of the toxicant, and we express the rate of change of the total amount of toxicant in the environment as the sum of the rates of input to and losses from the environment. Supposing that the inputs to the environment are proportional with $C_0(t)n(t)$, and the losses from the environment to the population are proportional with $CE(t)n(t)$, the concentration of toxicant outside the cells changes over time according the following equation:

$$\frac{dCE(t)}{dt} = \lambda_2^2 C_0(t)n(t) - \eta_2^2 CE(t)n(t) \tag{4}$$

where λ_2^2 is the toxicant uptake rate from cells, and η_2^2 represents the losses rate of toxicants absorbed by cells. Coupling the equations (2) – (4), we obtain the following toxicity model

$$\begin{aligned} \frac{dn(t)}{dt} &= \beta n(t) \left(1 - \frac{n(t)}{K}\right) - \alpha C_0(t)n(t) \\ \frac{dC_0(t)}{dt} &= \lambda_1^2 CE(t) - \eta_1^2 C_0(t) \\ \frac{dCE(t)}{dt} &= \lambda_2^2 C_0(t)n(t) - \eta_2^2 CE(t)n(t). \end{aligned} \tag{5}$$

TABLE 1. List of Variables and Parameters

Symbol	Definition
$n(t)$	cell index \approx cell population
$C_0(t)$	toxicant concentration inside the cell
$CE(t)$	toxicant concentration outside the cell
β	cell growth rate in the absence of toxicant
K	capacity volume
α	effect coefficient of toxicant on the cell’s growth
λ_1^2	the uptake rate of the toxicant from environment
λ_2^2	the toxicant uptake rate from cells
η_1^2	the toxicant input rate to the environment
η_2^2	the losses rate of toxicant absorbed by cells

The variables and parameters in this model are listed in Table 1.

Without loss of generality we suppose that the initial time is $t_0 = 0$, and we consider the initial cell index $0 < n(0) \leq K$. We assume that the toxicants do not exist in the cells before experiments, so the initial concentration of internal toxicant $C_0(0) = 0$, and the initial concentration of toxicants outside of the cells $CE(0) > 0$ equals to the concentration of toxicant used in the experiments.

As we have mentioned before, the model (5) is a particular case of the class of models introduce in [4]. More recently, similar models were studied in [5, 10, 15]. However, in these papers sufficient conditions for persistence or extinction of the population are expressed in terms of the toxicant input rate $u(t)$, so they do not apply to our model because we consider an acute dose of toxicant ($u = 0$), instead of a chronic one. In the next section we study the stability of the model (5) applying both techniques for studying the persistence as in [4, 5], and methods used to analyze the existence and stability of equilibria.

3. Stability analysis. We assume that all parameters of model (5) have positive values.

3.1. Existence and stability of equilibria. Equilibria are very important in the study of the asymptotic behavior of a dynamical system [7]. To determine the equilibria, we set the vector field in (5) equal to $\mathbf{0}$, and we solve the equilibria equations. To study the stability of the equilibria, we analyze the eigenvalues of the Jacobian matrix J of (5):

$$J = \begin{bmatrix} \beta - 2\beta n/K - \alpha C_0 & -\alpha n & 0 \\ 0 & -\eta_1^2 & \lambda_1^2 \\ \lambda_2^2 C_0 - \eta_2^2 CE & \lambda_2^2 n & -\eta_2^2 n \end{bmatrix}.$$

We have the following equilibria.

- **The extinction equilibria:** $\left[0, CE^* \frac{\lambda_1^2}{\eta_1^2}, CE^*\right]$ where $CE^* > 0$ is the concentration of toxicant outside the cells. For $\left[0, CE^* \frac{\lambda_1^2}{\eta_1^2}, CE^*\right]$ the eigenvalues are

$$\xi_1 = \beta - \alpha CE^* \frac{\eta_1^2}{\lambda_1^2}, \quad \xi_2 = -\eta_1^2, \quad \xi_3 = 0$$

To determine the stability and the qualitative behavior in a neighborhood of these non-hyperbolic equilibria, we study the behavior on the local center manifold [21] (see appendix A for details regarding the center manifold). The analytical center manifold is given by the line $\{(n, C_0, CE) : n = 0, \lambda_1^2 CE - \eta_1^2 C_0 = 0\}$. The extinction equilibria are unstable [7] if the first eigenvalue is positive, i.e. $0 < CE^* < \frac{\beta \eta_1^2}{\alpha \lambda_1^2}$.

- **The maximum capacity equilibrium:** $[K, 0, 0]$. For $[K, 0, 0]$ the eigenvalues are

$$\xi_1 = -\beta < 0$$

$$\xi_{2,3} = -\frac{1}{2} \left(\eta_1^2 + \eta_2^2 K \pm \sqrt{(\eta_1^2 + \eta_2^2 K)^2 + 4(\lambda_1^2 \lambda_2^2 - \eta_1^2 \eta_2^2) K} \right)$$

Notice that if $\lambda_1^2 \lambda_2^2 - \eta_1^2 \eta_2^2 < 0$, then all the eigenvalues are negative, so $[K, 0, 0]$ is locally asymptotically stable. If $\lambda_1^2 \lambda_2^2 - \eta_1^2 \eta_2^2 > 0$ then $\xi_1 < 0, \xi_2 > 0, \xi_3 < 0$, so $[K, 0, 0]$ is unstable. Finally, if $\lambda_1^2 \lambda_2^2 - \eta_1^2 \eta_2^2 = 0$ then $\xi_1 < 0, \xi_2 = 0, \xi_3 < 0$, so $[K, 0, 0]$ is a non-hyperbolic fixed point.

- **The interior equilibrium:** If $\lambda_1^2 \lambda_2^2 - \eta_1^2 \eta_2^2 = 0$, then we have the interior equilibria $\left[K \left(1 - \frac{\alpha \lambda_1^2 CE}{\beta \eta_1^2} \right), \frac{\lambda_1^2 CE}{\eta_1^2}, CE \right]$, where $CE > \frac{\beta \eta_1^2}{\alpha \lambda_1^2}$. For these equilibria the eigenvalues are

$$\xi_1 = 0, \quad \xi_2 = -\beta \left(1 - \frac{\alpha \lambda_1^2 CE}{\beta \eta_1^2} \right) < 0, \quad \xi_3 = -\eta_1^2 - \eta_2^2 K \left(1 - \frac{\alpha \lambda_1^2 CE}{\beta \eta_1^2} \right) < 0,$$

so these equilibria are non-hyperbolic fixed points.

3.2. Persistence and extinction. Following [5, 4] we define persistence and extinction as follows.

- The cell population is said to be uniformly persistent if there exist the constants $M_1 > 0, M_2 > 0$ such that $0 < M_1 \leq \liminf_{t \rightarrow \infty} n(t) \leq \limsup_{t \rightarrow \infty} n(t) \leq M_2 < +\infty$.
- The cell population is said to go to local extinction if $\lim_{t \rightarrow \infty} n(t) = 0$.

Intuitively, uniform persistence corresponds to the long term cells survival, and the cells death corresponds in the previous definition to local extinction.

Similarly with [5] we can prove the following lemma (see appendix B for a detailed proof).

Lemma 3.1. *If $\eta_1^2 \eta_2^2 - \lambda_1^2 \lambda_2^2 > 0$, then $0 < CE(t) \leq CE(0), 0 \leq C_0(t) \leq \frac{\lambda_1^2 CE(0)}{\eta_1^2}$, and $n(t) > 0$, for all $t \geq 0$.*

We have noticed that for the chemicals considered in this paper, the estimated values of the parameters $\eta_1, \eta_2, \lambda_1$, and λ_2 verify $\eta_1^2 \eta_2^2 - \lambda_1^2 \lambda_2^2 > 0$, so we focus on this case. Using lemma 3.1, we get the following result (see appendix B for the proof).

Theorem 3.2. *If $\eta_1^2 \eta_2^2 - \lambda_1^2 \lambda_2^2 > 0$, then $\lim_{t \rightarrow \infty} CE(t)$ exists and its value determines the asymptotic behavior of system (5) according to the following two cases:*

1. *If $\lim_{t \rightarrow \infty} CE(t) < \frac{\beta \eta_1^2}{\alpha \lambda_1^2}$ then the population is uniformly persistent and we have*

$$\lim_{t \rightarrow \infty} n(t) = K, \quad \lim_{t \rightarrow \infty} C_0(t) = \lim_{t \rightarrow \infty} CE(t) = 0.$$

2. *If $\lim_{t \rightarrow \infty} CE(t) > \frac{\beta \eta_1^2}{\alpha \lambda_1^2}$ then the population goes to local extinction and we have*

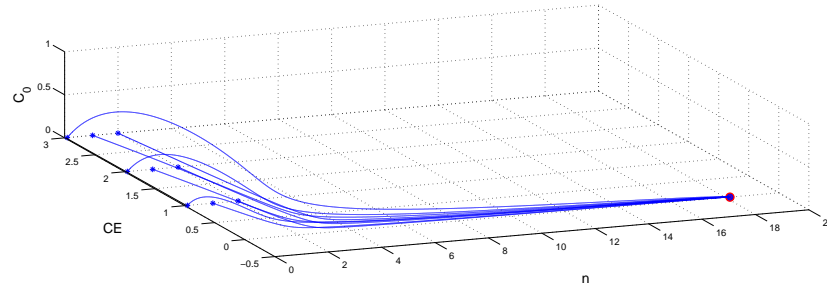
$$\lim_{t \rightarrow \infty} n(t) = 0, \quad \lim_{t \rightarrow \infty} C_0(t) = CE^* \frac{\lambda_1^2}{\eta_1^2}, \quad \lim_{t \rightarrow \infty} CE(t) = CE^* > \frac{\beta \eta_1^2}{\alpha \lambda_1^2},$$

Moreover, the population component is in $L^1[0, \infty)$, i.e. $|n|_1 = \int_0^\infty n(t) dt < \infty$.

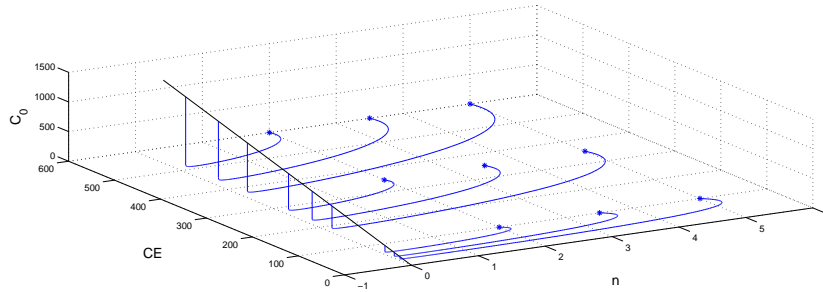
Notice that if $\eta_1^2 \eta_2^2 - \lambda_1^2 \lambda_2^2 > 0$ and $0 < CE(0) < \frac{\beta \eta_1^2}{\alpha \lambda_1^2}$ then, by Lemma 3.1, $\lim_{t \rightarrow \infty} CE(t) < \frac{\beta \eta_1^2}{\alpha \lambda_1^2}$, so we are in the first case and we have $\lim_{t \rightarrow \infty} n(t) = K > 0$. From the analysis of the stability of the equilibria we know that $[K, 0, 0]$ is a locally asymptotically stable fixed point, so there exists populations modeled by equations (5) that are uniformly persistent. These results are summarized in the following corollary.

Corollary 1. *If $\eta_1^2 \eta_2^2 - \lambda_1^2 \lambda_2^2 > 0$ and $0 < CE(0) < \frac{\beta \eta_1^2}{\alpha \lambda_1^2}$ then the population is uniformly persistent, and we have $\lim_{t \rightarrow \infty} n(t) = K$.*

This is illustrated numerically by the trajectories displayed in Fig. 2(a) for monastrol. These trajectories correspond to several sets of initial values $n(0) > 0$, $C_0(0) = 0$, $0 < CE(0) < \frac{\beta\eta_1^2}{\alpha\lambda_1^2} = 6.51$. We noticed that for this set of initial values the trajectories approach the equilibrium $[K, 0, 0] = [18.17, 0, 0]$.



(a)



(b)

FIGURE 2. Trajectories corresponding to monastrol and initial values $0 < n(0) < K$, $C_0(0) = 0$, and (a) $CE(0) < \frac{\beta\eta_1^2}{\alpha\lambda_1^2} = 6.51$. (b) $CE(0) > \frac{\beta\eta_1^2}{\alpha\lambda_1^2} = 6.51$

This means that if the initial concentration is sufficiently low, the presence of the chemical does not affect the convergence of the cell index towards the maximum capacity K . Moreover, we have a threshold, $\frac{\beta\eta_1^2}{\alpha\lambda_1^2}$, such that if the experiment is done using a chemical with initial concentration $CE(0) < \frac{\beta\eta_1^2}{\alpha\lambda_1^2}$, then the cell population is uniformly persistent.

If $\eta_1^2\eta_2^2 - \lambda_1^2\lambda_2^2 > 0$ and $CE(0) > \frac{\beta\eta_1^2}{\alpha\lambda_1^2}$, then we know that the population goes to local extinction if $|n|_1 < \infty$ (see Theorem 3.2). In Fig. 2 (b) this is illustrated numerically for monastrol by the trajectories that asymptotically approach the extinction equilibria $[0, CE^* \frac{\lambda_1^2}{\eta_1^2}, CE^*]$, with $CE^* > \frac{\beta\eta_1^2}{\alpha\lambda_1^2}$. We notice from Fig. 2 (b) that these extinction equilibria are on a line representing also the center manifold.

However, if $\eta_1^2\eta_2^2 - \lambda_1^2\lambda_2^2 > 0$, $CE(0) > \frac{\beta\eta_1^2}{\alpha\lambda_1^2}$ and $|n|_1 = \infty$, the population is uniformly persistent (see Theorem 3.2). Thus, the initial concentration of the

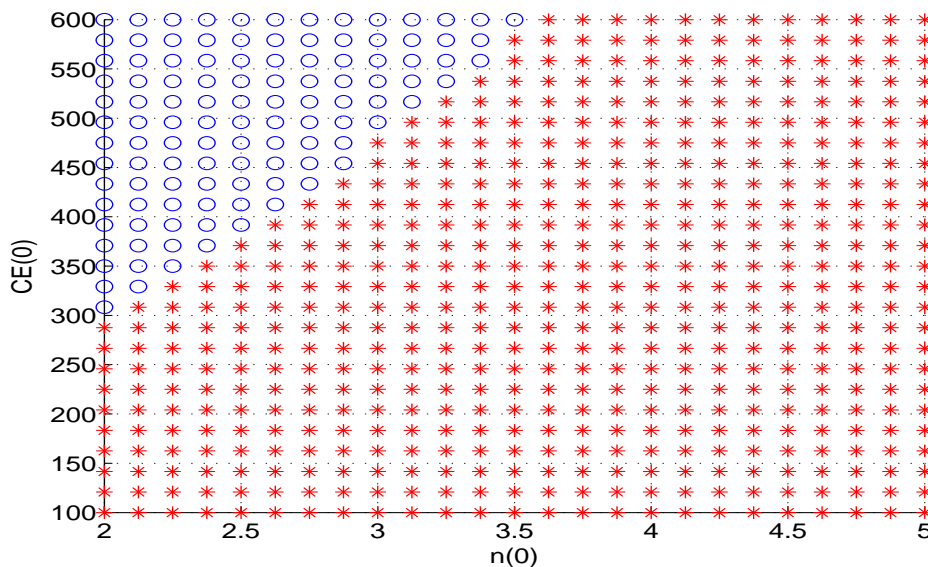


FIGURE 3. The separation between persistence and extinction according to the initial values $n(0)$ and $CE(0)$, red *: persistence; blue o: extinction.

chemical and the entire past history of the population are important to determine the persistence or extinction of the population. A similar conclusion was obtained in [4] for a reduced model (with $\lambda_2 = 0$), also corresponding to an acute dose of toxicant.

For monastrol, the separation between persistence and extinction according to the initial values $n(0) > 0$, $C_0(0) = 0$, $CE(0) > \frac{\beta\eta_1^2}{\alpha\lambda_1^2} = 6.51$ are presented in Fig. 3. Results of numerical simulations similar with the ones presented in Fig. 3 are important, because they can be applied to choose an appropriate range for the values of $CE(0)$ used in the experiments such that the TCRCs will illustrate the evolution from cells' population survival (corresponding to small values of $CE(0)$) to cells' population extinction (corresponding to large values of $CE(0)$).

In the next section we design an efficient method to estimate the parameters based on the TCRCs. With the proposed method, for a given chemical and a cell line, the parameters can be determined after only two experiments, namely the negative control and an experiment with a relatively high initial concentration of the testing substance. Thus, after two experiments we can determine the threshold $\frac{\beta\eta_1^2}{\alpha\lambda_1^2}$ and we can obtain numerically separation regions similar with one displayed in Fig. 3. This information can be used to save both experimental time and resources.

4. Parameter estimation. The non-linear system of ordinary differential equations (5) cannot be solved analytically, so to estimate the parameters we rewrite it in a state-space form. Notice that from the experimental data recorded in the TCRCs we get observations, possibly affected by measurement errors, only for the cell index $n(t)$. Thus, using the Euler integration scheme with time step h , the

associated discrete state-space system can be expressed as:

$$x_{k+1} = x_k + h \begin{bmatrix} \beta x_k[1] \left(1 - \frac{x_k[1]}{K}\right) - \alpha x_k[2] x_k[1] \\ \lambda_1^2 x_k[3] - \eta_1^2 x_k[2] \\ \lambda_2^2 x_k[2] x_k[1] - \eta_2^2 x_k[3] x_k[1] \end{bmatrix} + v_{k+1} \quad (6)$$

$$y_{k+1} = C x_{k+1} + w_{k+1} \quad (7)$$

Here $x_{k+1} = [n, C_0, CE]^t$ is the state of the system, y_{k+1} is the observation at time step $k + 1$, and $C = [1 \ 0 \ 0]$. We assume that the white noise vectors, v_k and w_k , are Gaussian and from uncorrelated white sequences such that $v_k \sim N(0, Q)$ and $w_k \sim N(0, R)$. The 3×3 diagonal covariance matrix Q , the variance $R > 0$, α , $\lambda_1, \lambda_2, \eta_1, \eta_2, \beta$ and K represent the unknown parameters, h is the time step and the initial values $x_1[1] = n(0)$ comes from the measured data at the initial time t_0 , $x_1[2] = C_0(0) = 0$ and $x_1[3] = CE(0)$ equals the concentration of toxicant used in the experiments.

The parameters β and K are also the parameters of the logistic model (1) for which we can easily get an analytic solution. Thus, using the experimental data corresponding to the negative control (no toxicant) we can estimate β and K using the nonlinear least square method based on the analytic solution of equation (1).

The system (5) cannot be solved analytically, so to estimate the remaining parameters $\Theta = \{Q, R, \alpha, \lambda_1, \lambda_2, \eta_1, \eta_2\}$ we use the state-space model (6)-(7) and the Expectation Maximization (EM) algorithm based on the the unscented filter (UF). The EM algorithm is a classical method for estimating the parameters of linear systems [22], and for a general nonlinear dynamics, the EM algorithm has been applied using the UF or the EKF smoothing [3, 25].

The EM algorithm is especially useful when it is easier to calculate the likelihood of the model using not only the observed data $Y_{obs} = \{y_k, k = 1, \dots, N\}$, but also the hidden data $Y_{hid} = \{x_k, k = 1, \dots, N\}$. Hence the EM is based upon a data augmentation scheme, such that the observed data are a mapping of the augmented data $Y_{obs} = \mathbf{m}(Y_{aug})$, where $Y_{aug} = \{Y_{obs}, Y_{hid}\}$.

The state equation (6) is non-linear, so an approximation is needed during the E-step. The likelihood and the conditional likelihood are approximated based on a linearization, and the UF is used for filtering and smoothing. The next two subsections briefly present the implementation of the UF and the EM algorithm for the state-space model (6)-(7).

4.1. The unscented filter. The UF was introduced by Julier and Uhlmann [12] as an alternative to the extended Kalman filter (EKF) for non-linear state-space models. This filter does not require the calculation of the Jacobians, and it is computationally at most as expensive as the EKF.

Both the EKF and the UF approximate the state distribution with a Gaussian one. However, instead of using the EKF linearization approach, the UF employs a deterministic sampling. The sample points completely capture the true mean and the true covariance. In contrast to the first-order accuracy of the EKF, the UF is capable to accurately capture the true posterior mean and the covariance up to the third order for a nonlinear system.

For the state-space model (6)-(7), let us denote $\bar{x}_i = E[x_i | y_1, \dots, y_i]$ and $\hat{x}_{i+1} = E[x_{i+1} | y_1, y_2, \dots, y_i]$ for the filtered and the predicted values, respectively, where

$E[\cdot|\cdot]$ represent the conditional expectation. The corresponding conditional covariances are $\bar{P}_i = E[(x_i - \bar{x}_i)(x_i - \bar{x}_i)^t | y_1, \dots, y_i]$ and $\hat{P}_{i+1} = E\left[(x_{i+1} - \hat{x}_{i+1})(x_{i+1} - \hat{x}_{i+1})^t | y_1, \dots, y_i\right]$. Based on the available observations $y_1 \dots y_N$, smoothed values $x_{i|N} = E[x_i | y_1 \dots y_N]$ and $P_{i|N} = E[(x_i - x_i^N)(x_i - x_i^N)^t | y_1 \dots y_N]$ can be calculated.

Let denote

$$f(x) = x + h \begin{bmatrix} \beta x[1](1 - \frac{x[1]}{K}) - \alpha x[2]x[1] \\ \lambda_1^2 x[3] - \eta_1^2 x[2] \\ \lambda_2^2 x[2]x[1] - \eta_2^2 x[3]x[1] \end{bmatrix}, x \in \mathbb{R}^3 \tag{8}$$

and consider the linearised system

$$x_{k+1} = f(\bar{x}_k) + \frac{\partial f}{\partial x}(\bar{x}_k)(x_k - \bar{x}_k) + v_{k+1}, \tag{9}$$

$$y_k = Cx_k + w_k, \tag{10}$$

Notice that for the linearized system we have

$$\bar{P}_{x_k x_{k+1}} = E[(x_k - \bar{x}_k)(x_{k+1} - \hat{x}_{k+1})^t | x_1, \dots, x_k] = \bar{P}_k \frac{\partial f^t}{\partial x}(\bar{x}_k), \tag{11}$$

and with

$$J_k = \bar{P}_{x_k x_{k+1}} \hat{P}_{k+1}^{-1}, \tag{12}$$

we have the following backward recursions

$$x_{k-1|N} = \bar{x}_{k-1} + J_{k-1}(x_{k|N} - \hat{x}_k), \tag{13}$$

$$P_{k-1|N} = \bar{P}_{k-1} + J_{k-1}(P_{k|N} - \hat{P}_k)J_{k-1}^t. \tag{14}$$

The UF filtering algorithm [11] can be used to compute \bar{x}_{k+1} , \bar{P}_{k+1} , \hat{x}_{k+1} , \hat{P}_{k+1} and $\bar{P}_{x_k x_{k+1}}$. The smoothed values $x_{k|N}$ and $P_{k|N}$, $k = 1, \dots, N$ can be computed using (12) and the backwards recursions given in (13)-(14), starting with $k = N$.

4.2. The EM algorithm. The algorithm starts with an initial guess θ_0 for the unknown parameters and iteratively compute the estimation θ^* . Each iteration consists of two steps: the expectation (E) and the maximization (M) step. Using the current estimation θ_n of the parameters, the E-step computes the conditional expectation of the augmented data log-likelihood $Q(\theta|\theta_n) = E[\log p(\theta|Y_{aug}) | Y_{obs}, \theta_n]$. The M-step performs a maximization with respect to the parameters θ :

$$\theta_{n+1} = \arg \max_{\theta} Q(\theta|\theta_n).$$

We briefly explain the implementation of the proposed EM-algorithm for estimating the parameters Θ of the state-space model (6)-(7). First, we augment the data with the hidden variables x_i , $i = 1, \dots, N$ and we calculate the complete log-likelihood:

$$\begin{aligned} \log(L) &= \log P(x_1, \dots, x_N, y_1, \dots, y_N) = \log P(y_N | y_{N-1}, \dots, y_1, x_N, \dots, x_1) \\ &+ \dots + \log P(y_1 | x_N, \dots, x_1) + \log P(x_N | x_{N-1}, \dots, x_1) + \dots + \log P(x_1). \end{aligned}$$

Using the linearized equations (9) and (10), we obtain:

$$\begin{aligned} \log(L) &= -5N \ln(2\pi) - \frac{1}{2} \sum_{i=1}^N \ln(|R|) - \frac{1}{2} \sum_{i=1}^N (y_i - Cx_i)^t R^{-1} (y_i - Cx_i) \\ &\quad - \frac{1}{2} \sum_{i=2}^N \ln(|Q|) - \frac{1}{2} \sum_{i=2}^N (x_i - F_{i-1}x_{i-1} - b_{i-1})^t Q^{-1} (x_i - F_{i-1}x_{i-1} - b_{i-1}) \\ &\quad - \frac{1}{2} \ln(|\Sigma|) - \frac{1}{2} (x_1 - \mu)^t \Sigma^{-1} (x_1 - \mu), \end{aligned}$$

where

$$F_i = \begin{bmatrix} 1 + h(\beta - 2\beta\bar{x}_i[1]/K - \alpha\bar{x}_i[2]) & -h\alpha\bar{x}_i[1] & 0 \\ 0 & 1 - \eta_1^2 h & \lambda_1^2 h \\ h\lambda_2^2 \bar{x}_i[2] - h\eta_2^2 \bar{x}_i[3] & h\lambda_2^2 \bar{x}_i[1] & 1 - h\eta_2^2 \bar{x}_i[1] \end{bmatrix}$$

and

$$b_i = \begin{bmatrix} h(\beta\bar{x}_i[1]^2/K + \alpha\bar{x}_i[2]\bar{x}_i[1]) \\ 0 \\ h(-\lambda_2^2 \bar{x}_i[2]\bar{x}_i[1] + \eta_2^2 \bar{x}_i[3]\bar{x}_i[1]) \end{bmatrix}.$$

The EM algorithm iteratively maximizes $\hat{E} = E[\log(L)|y_1, \dots, y_N]$. Using the previous formulas and $\text{tr}(AB) = \text{tr}(BA)$, where tr is the notation for the trace of a matrix, we get

$$\begin{aligned} \hat{E} &= -5N \ln(2\pi) - \frac{N}{2} \log(|R|) - \frac{1}{2} \sum_{n=1}^N \text{tr}[R^{-1}(y_n y_n^t - y_n x_{n|N}^t C^t - C x_{n|N} y_n^t \\ &\quad + C P_{n|N} C^t)] - \frac{1}{2} \sum_{n=2}^N \log(|Q|) - \frac{1}{2} \sum_{n=2}^N \text{tr}[Q^{-1}(P_{n|N} - P_{n,n-1|N} F_{n-1}^t - x_{n|N} b_{n-1}^t \\ &\quad - F_{n-1} P_{n,n-1|N}^t + F_{n-1} P_{n-1|N} F_{n-1}^t + F_{n-1} x_{n|N} b_{n-1}^t - b x_{n|N}^t + b_{n-1} x_{n|N}^t F_{n-1}^t \\ &\quad + b_{n-1} b_{n-1}^t)] - \frac{1}{2} \log(|\Sigma|) - \frac{1}{2} \text{tr}[\Sigma^{-1}(P_1 - \mu x_{1|N}^t - x_{1|N} \mu^t + \mu \mu^t)] \end{aligned}$$

The smoothed values $P_{n|N} = E[x_n x_n^t | y_1, \dots, y_N]$, $P_{n,n-1|N} = E[x_n x_{n-1}^t | y_1, \dots, y_N]$, and $x_{n|N} = E[x_n | y_1, \dots, y_N]$ can be calculated using the UF smoothing presented in the previous section. For the M-step, analytical update equations for the parameters Θ can be found taking derivatives with respect to the parameters Θ in the previous formula for \hat{E} . The implementation of the EM algorithm is summarized in Table 2

TABLE 2. The EM algorithm

Initialize the model parameters $\Theta = \{Q, R, \alpha, \lambda_1, \lambda_2, \eta_1, \eta_2\}$
Repeat until the log likelihood has converged
The E step
For k=1 to N
Run the UF filter to compute \bar{x}_{k+1} , \bar{P}_{k+1} , \hat{x}_{k+1} , \hat{P}_{k+1} and $\bar{P}_{x_k x_{k+1}}$
For k=N to 1
Calculate the smoothed values $x_{k N}$, and $P_{k N}$ using (13), (14)
The M step
Update the values of the parameters Θ to maximize \hat{E}

The main advantage of using the EM-algorithm is the guaranteed convergence [26, 18]. However, depending on the initial guess, the algorithm can only converge to a local maximum. Various strategies for choosing the initial guess are proposed in [1], and methods for accelerating the convergence are presented in [17].

5. Model validation. We divide the experimental data into a training set (containing N observations, where N equals around 70% of the data), and a test set. We use the observations in the training set to estimate the parameters. Once the parameters are estimated, we can predict the future values of x_i , $i = N + 1, N + 2, \dots$, and we validate the model by comparing these predictions with the experimental data in the test set.

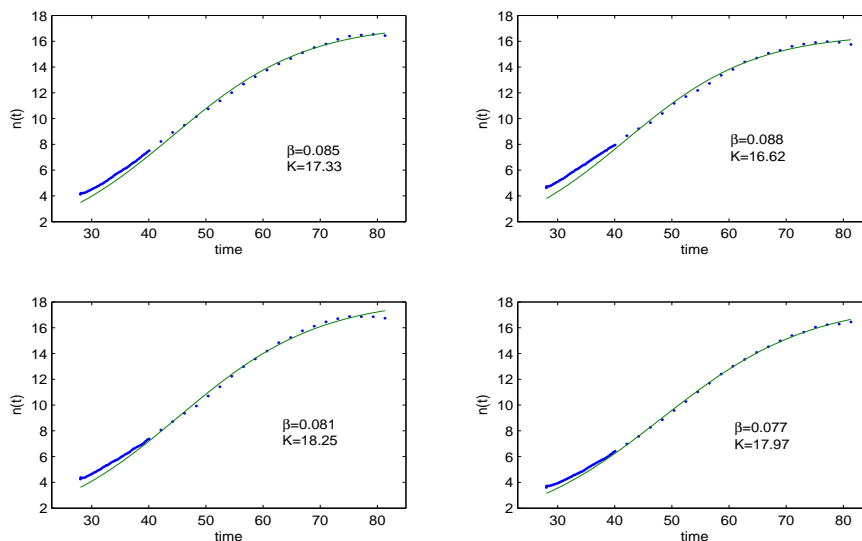


FIGURE 4. Negative control data fitted by logistic model, dot: experimental data, line: logistic model.

We start with the validation of the logistic model on the negative control cell growth data, and we estimate the cell growth rate β and capacity K (which we suppose are unchanged for the whole experiment). Since the logistic equation (1) can be solved analytically, the nonlinear least square method is an efficient and reliable method for parameter estimation and prediction.

Figure 4 shows the fitting of the negative control data of four replicates. The data predicted by the logistic model is in agreement with all the experimental data. Moreover the estimated parameters of all replicated experiments are consistent. Thus we conclude that the logistic equation (1) is a suitable model to represent the cell growth in non-responsive environment.

To test the responsive model (5), we present here the results obtained for PF431396, monastrol, ABT888, and HA1100 hydrochloride. These chemicals are selected from Cluster I: DNA/RNA-nucleic acid targets, and Cluster X: protein-motor targets. These two clusters are the top 2 most common clusters in the 63 compounds experiments, covering more than half of the cases.

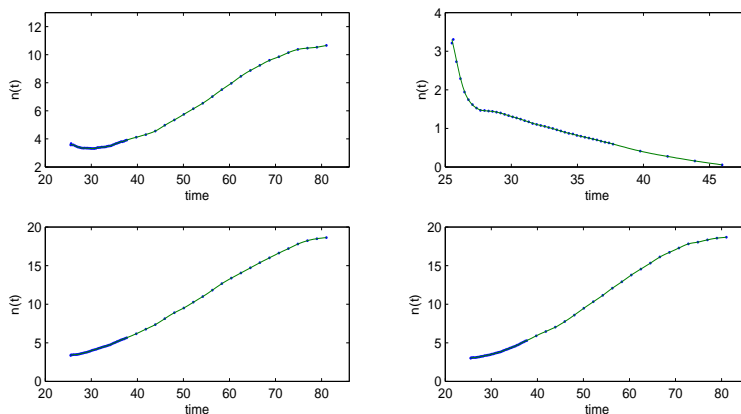


FIGURE 5. Smooth spline approximation, dot: experimental data, line: smooth spline

We perform preprocessing of the experiment data before testing the toxicity model. Since the cell index is not measured uniformly on time, we apply a smoothing spline to approximate the experimental data. Figure 5 presents a good agreement between the experimental data and the fitting spline.

After interpolation with a smoothing spline, we sample uniformly from the approximating spline, and we use the first 70% experimental data (i.e. the training set) to estimate the parameters, α , η_1 , η_2 , λ_1 and λ_2 with the EM algorithm. The estimated values of the parameters for the four chemicals mentioned before are given in Table 1. Using the estimated values of the parameters, we apply the UF to

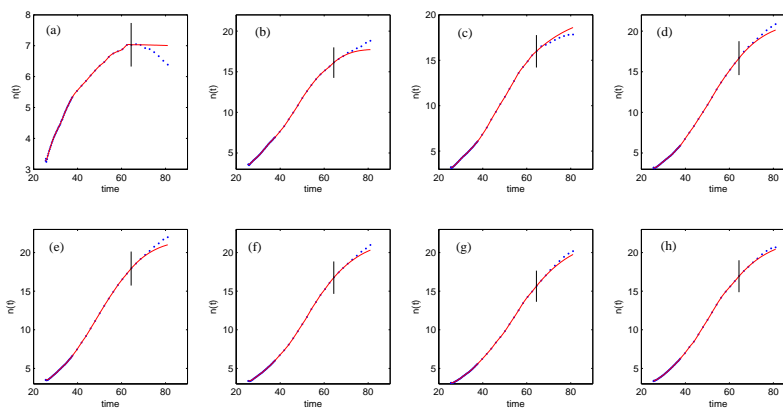


FIGURE 6. Estimation results for PF431396, dot: experimental data, line: filtered or predicted observations; (a) $CE(0)=5.00\mu\text{M}$, (b) $CE(0)=1.67\mu\text{M}$, (c) $CE(0)=0.56\mu\text{M}$, (d) $CE(0)=0.19\mu\text{M}$, (e) $CE(0)=61.73\text{nM}$, (f) $CE(0)=20.58\text{nM}$, (g) $CE(0)=6.86\text{nM}$, (h) $CE(0)=2.29\text{nM}$

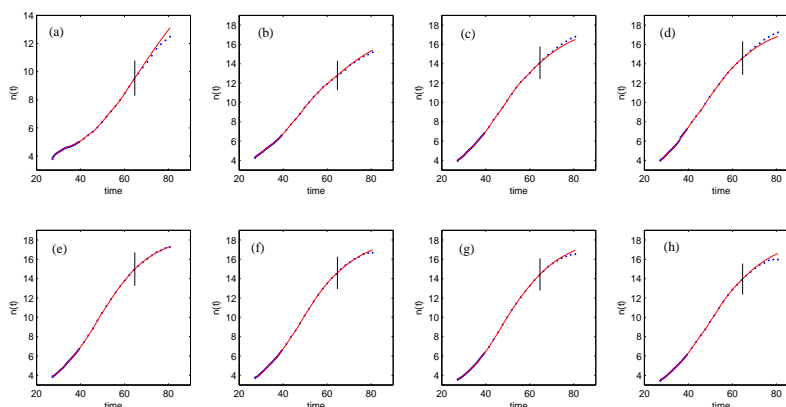


FIGURE 7. Estimation results for monastrol, dot: experimental data, line: filtered or predicted observations; (a) $CE(0)=100.00\mu\text{M}$, (b) $CE(0)=33.33\mu\text{M}$, (c) $CE(0)=11.11\mu\text{M}$, (d) $CE(0)=3.70\mu\text{M}$, (e) $CE(0)=1.23\mu\text{M}$, (f) $CE(0)=0.41\mu\text{M}$, (g) $CE(0)=0.14\mu\text{M}$, (h) $CE(0)=45.72\text{nM}$

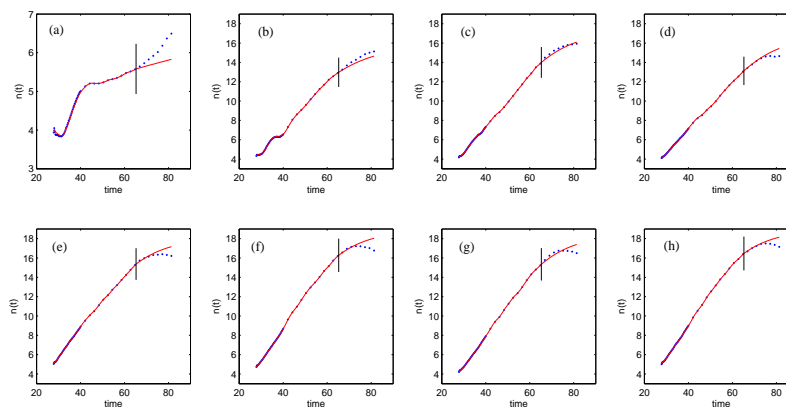


FIGURE 8. Estimation results for ABT888, dot: experimental data, line: filtered or predicted observations; (a) $CE(0)=308.00\mu\text{M}$, (b) $CE(0)=102.67\mu\text{M}$, (c) $CE(0)=34.22\mu\text{M}$, (d) $CE(0)=11.41\mu\text{M}$, (e) $CE(0)=3.80\mu\text{M}$, (f) $CE(0)=1.27\mu\text{M}$, (g) $CE(0)=0.42\mu\text{M}$, (h) $CE(0)=0.14\mu\text{M}$

predict the cell growth/death, and we compare the predicted values with the data in the test set (i.e. with the last 30% of the experimental data).

From Figs. 6-8, the predictions are in agreement with the experimental data. We have plotted only the TCRCs corresponding to the largest 8 out of 11 values of the initial concentration of the chemicals because the curves corresponding to the first 3 values are similar to the negative control. The observations from the initial

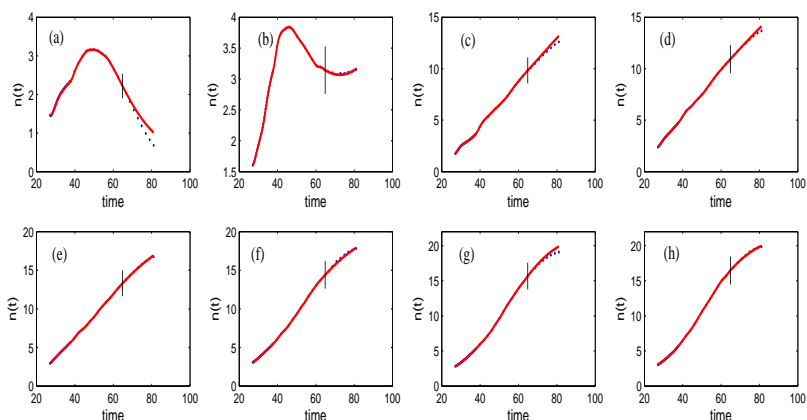


FIGURE 9. Estimation results for HA1100 hydrochloride, dot: experimental data, line: filtered or predicted observations; (a) $CE(0)=1.00\text{mM}$, (b) $CE(0)=0.33\text{mM}$, (c) $CE(0)=0.11\text{mM}$, (d) $CE(0)=37.04\text{uM}$, (e) $CE(0)=12.35\text{uM}$, (f) $CE(0)=4.12\text{uM}$, (g) $CE(0)=1.37\text{uM}$, (h) $CE(0)=0.46\text{uM}$

TABLE 3. Estimated Values of Parameters

Toxicant	Cluster	β	K	η_1	λ_1	λ_2	η_2	α
PF431396	X	0.077	21.912	0.273	0.058	0	0.008	0.238
monastrol	X	0.074	18.17	0.209	0.177	0.204	0.5	0.016
ABT888	I	0.083	17.543	0.079	0.177	0.205	0.5	0.005
HA1100 hydrochloride	I	0.077	21.913	0.143	0.0098	0.0786	0.147	0.351

time t_0 to the red vertical line represent the training set, and are used for parameter estimation. The test set is formed with the data after the vertical red line. In all the figures the experimental data is plotted using dotted lines, and the filtered or predicted observations are presented using plain lines. The concentration reduces by a third from top to bottom, and from left to right.

6. Parameter uncertainty and sensitivity analysis. In addition to the point estimates obtained with the EM algorithm, we analyze the effect of parameter uncertainty on predicting the cell index using Monte Carlo simulations and global sensitivity analysis (GSA). Sensitivity Analysis (SA) refers to a broad group of methods that ranks parameters by their effect on output variables. GSA is the most suitable group of sensitivity analysis methods for application in highly nonlinear dynamic models [13]. In GSA, variations of all input parameters simultaneously are considered. As a result, interactions among different inputs can be detected. There is a variety of GSA methods, and here we apply the Sobol's method [23].

6.1. Simulations with random parameters. To account for parameter uncertainty, we run simulations assuming that all parameters are random variables with triangle distributions on intervals centered around the estimated values of the parameters given in Table 1. Based on these simulations we determine the expected cell index and the extinction probability. We consider intervals corresponding to

perturbations of 10% , we run the model for 1000 realizations and we simulate for $t=60$ hours (similarly with the experimental settings). In Figs. 10(a), 11(a)

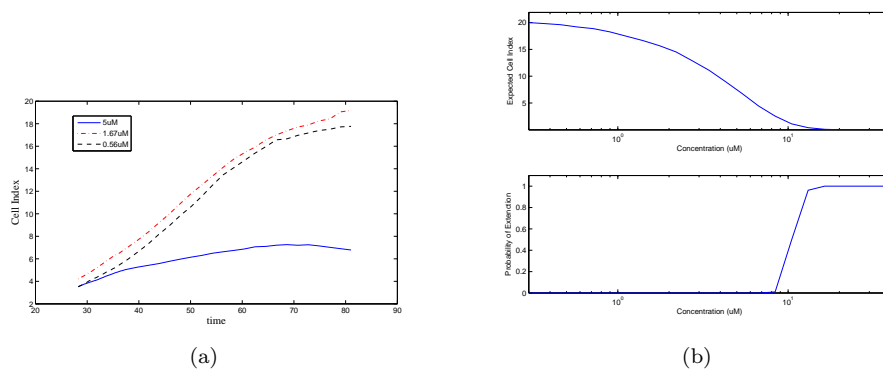


FIGURE 10. (a) Experimental TCRCs for PF431396 for $CE(0)=5\mu\text{M}$, $1.67\mu\text{M}$, $0.56\mu\text{M}$ (b) Expected cell index and probability of extinction for different concentrations for PF431396

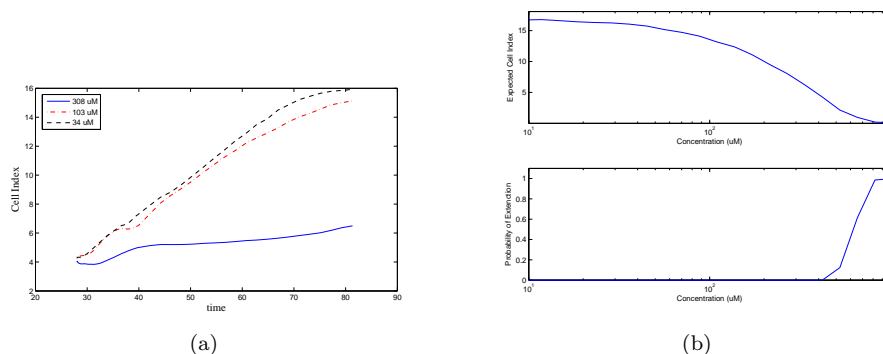


FIGURE 11. (a) Experimental TCRCs for ABT888 for $CE(0)=308\mu\text{M}$, $103\mu\text{M}$, $34\mu\text{M}$ (b) Expected cell index and probability of extinction for different concentrations for ABT888

we display some TCRCs recorded after experiments for PF431396 and ABT888. We compare these curves with the simulations results obtained using the proposed mathematical model.

In Figs. 10(b), 11(b) we plot the expected cell index and the probability of extinction against the concentration of PF431396 and ABT888. Notice that the values of the expected cell index are in good agreement with the experimental data represented in the TCRCs. The probability of extinction increases very steeply from 0 to 1 when the concentrations increase and pass a certain threshold. This illustrates the conclusions of the stability analysis presented in section 3.

From figures 6, 8, 9 we can notice that the TCRCs for PF431396, ABT888 and HA1100 hydrochloride cover both cases corresponding to persistence and to

extinction. However, from figures 7 we can notice that the experiments for monastrol are done with too small initial concentrations because all TCRCs show persistence. The results of this type of simulations can be used to determine an appropriate range for the initial concentration of chemicals $CE(0)$ used in the experiments such that both values smaller and larger than the threshold between extinction and persistence are included. They extend the stability analysis done in section 3 by considering also the effects of parameter uncertainty.

6.2. Global sensitivity analysis. Sobol's sensitivity measures [23] that utilize the analysis of variance (ANOVA) of the model output are among the most widely used GSA methods [13] [9]. In Sobol's method, the variance of the model output is decomposed into components that result from individual parameters as well as from parameter interactions.

Consider a mathematical model represented by the square integrable function $f(\mathbf{x})$, where $\mathbf{x} = (x_1, \dots, x_d)$ are the parameters of the model. Without loss of generality, we suppose x_1, \dots, x_d are independent, uniformly distributed random variables on $[0, 1]^d$, and $f(\mathbf{x})$ has the ANOVA type of decomposition

$$f(\mathbf{x}) = f_{\emptyset} + \sum_k f_{[k]}(x_k) + \sum_{k_1 < k_2} f_{[k_1, k_2]}(x_{k_1}, x_{k_2}) + \dots + f_{[1, \dots, d]}(x_1, \dots, x_d) \quad (15)$$

where f_{\emptyset} is a constant and $f_{[k_1, \dots, k_r]}(x_{k_1}, \dots, x_{k_r})$ is a function only depending on the variables x_{k_1}, \dots, x_{k_r} such that for $1 \leq i \leq r$,

$$\int_0^1 f_{[k_1, \dots, k_r]}(x_{k_1}, \dots, x_{k_r}) dx_{k_i} = 0.$$

Thus the summands in (15) are orthogonal to each other and we have

$$\int_{[0,1]^d} f(\mathbf{x}) d\mathbf{x} = f_{\emptyset}.$$

Let us define

$$D := \int_{[0,1]^d} f^2(\mathbf{x}) d\mathbf{x} - f_{\emptyset}^2$$

$$D_{k_1, \dots, k_r} := \int_{[0,1]^r} f_{[k_1, \dots, k_r]}^2(x_{k_1}, \dots, x_{k_r}) dx_{k_1} \dots dx_{k_r}.$$

We have $\text{Var}[f(\mathbf{x})] = D$, $\text{Var}[E[f(\mathbf{x})|x_i]] = D_i$, $\text{Var}[E[f(\mathbf{x})|x_i, x_j]] = D_i + D_j + D_{i,j}$, \dots , so D can be seen as the total variance and D_{k_1, \dots, k_r} as partial variances. The Sobol's indices are defined as

$$S_{k_1, \dots, k_r} = \frac{D_{k_1, \dots, k_r}}{D}.$$

Squaring (15), integrating over $[0, 1]^d$ and using the orthogonality of the summands [23], we get

$$\sum_{r=1}^d \sum_{k_1 < k_2 < \dots < k_r} D_{k_1, \dots, k_r} = D.$$

The GSA indices of a model can be evaluated using Monte Carlo numerical integration. To demonstrate the process, we present the estimation of first order indices D_i for $1 \leq i \leq d$, and the second order indices $D_{i,j}$ for $1 \leq i < j \leq d$.

Let $\xi_i = (\xi_1^i, \dots, \xi_d^i)$ for $i = 1, \dots, N$ be random samples of \mathbf{x} . Then

$$f_\emptyset \approx \frac{1}{N} \sum_{i=1}^N f(\xi_i), \quad D \approx \frac{1}{N} \sum_{i=1}^N f^2(\xi_i) - f_\emptyset^2.$$

Then we generate another group of samples $\hat{\xi}_i = (\hat{\xi}_1^i, \dots, \hat{\xi}_d^i)$ for $i = 1, \dots, N$, which are independent with ξ_i . The GSA indices are estimated as

$$D_j \approx \frac{1}{N} \sum_{i=1}^N f(\xi_1^i, \dots, \xi_d^i) f(\hat{\xi}_1^i, \dots, \hat{\xi}_{j-1}^i, \xi_j^i, \hat{\xi}_{j+1}^i, \dots, \hat{\xi}_d^i) - f_\emptyset^2$$

$$D_{j_1, j_2} \approx \frac{1}{N} \sum_{i=1}^N f(\xi_1^i, \dots, \xi_d^i) f(\hat{\xi}_1^i, \dots, \hat{\xi}_{j_1-1}^i, \xi_{j_1}^i \hat{\xi}_{j_1+1}^i, \dots, \hat{\xi}_{j_2-1}^i, \xi_{j_2}^i \hat{\xi}_{j_2+1}^i, \dots, \hat{\xi}_d^i) - f_\emptyset^2$$

We present the GSA studies for our proposed model. The estimated values of the parameters for different chemicals are given in Table 3. We consider that all parameters are random variables with uniform distribution on intervals corresponding to perturbations of 10%. To calculate the Sobol’s indices we use the Matlab toolbox named GSAT developed by Cannavó [2].

In Figs. (12), (13), we present the ranking for the GSA indices for PF431396 and ABT888 with various concentration, respectively. We can notice that when the chemicals’ concentration is small the capacity K is the most sensitive parameter of the model. This can be explained by the fact that for low concentrations of toxicant the cell population is persistent, and asymptotically the cell index n approaches the capacity K (see section 3). As the concentration increase to the transition stage from persistence to extinction, the cell population become more sensitive to the intake and outtake rates $\lambda_1, \eta_1, \lambda_2, \eta_2$.

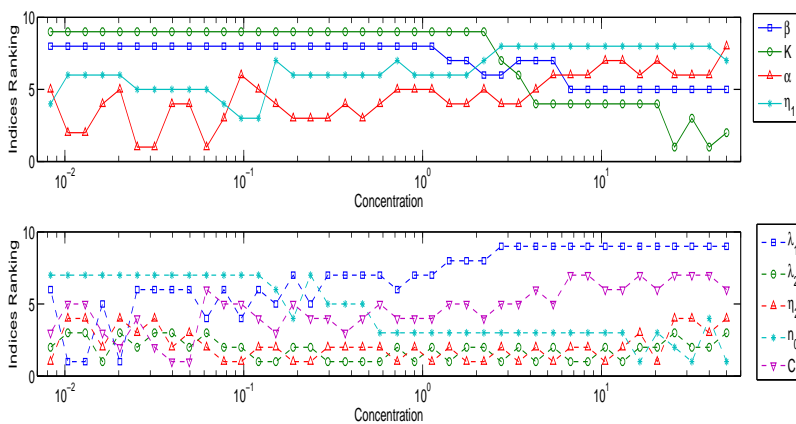


FIGURE 12. The first order GSA indices ranking for PF431396 (higher rank means more sensitive).

Figure (14) present the second order GSA indices for the same chemicals at the transition stage from persistence to extinction. The thickness of lines of any pair of parameters represents the rank of their co-sensitivity (the thicker the line, the greater the co-sensitivity). These pictures also show that the intakes and outtakes

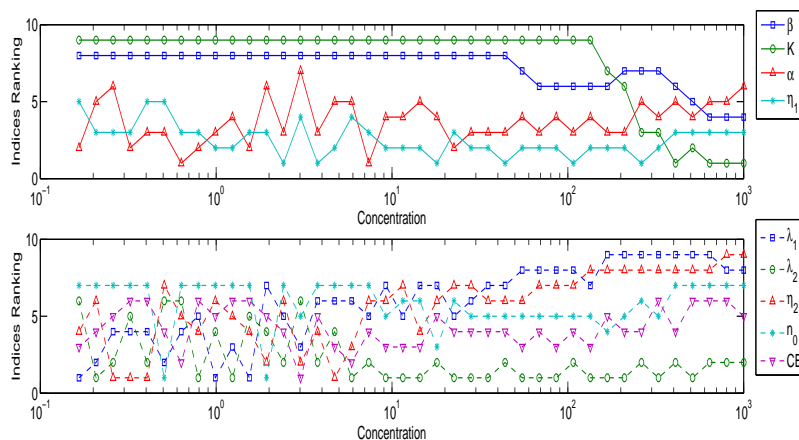


FIGURE 13. The first order GSA indices ranking for ABT888 (higher rank means more sensitive).

rates are the most sensitive parameters for the cells' population model. Similar results are obtained for monastrol and HA1100 hydrochloride.

7. Discussion and summary. The novelty of this paper is the application of a dual methodology, including an efficient statistical method for parameter estimation and prediction of the cell index and chemicals' concentration, followed by a numerical and analytical stability study for the dynamical system modeling the cells' growth/death.

We propose a mathematical model expressed by a three dimensional system of differential equations that can be expressed in a state-space form, and we design an EM algorithm based on the unscented filter for parameter estimation and prediction. Since it requires less tuning, the EM algorithm is a more efficient alternative for parameter estimation than the MHE or the EKF [8]. The results presented here for four chemicals with different MOAs illustrate the efficiency of the proposed algorithm and validate the mathematical model. Our approach is very efficient because the model parameters can be estimated based on TCRCs provided by only two experiments, namely the negative control and an experiment with a relatively large initial concentration of the chemical. Using the mathematical model, accurate TCRCs for other concentrations can be obtained numerically.

Based on the TCRCs, clustering methods or machine learning algorithms can be applied to classify the chemicals. However, some chemicals were tested with either too low concentrations (so no significant effect on the cells growth/death is noticed) or too high concentrations (so the only effect that can be noticed is the rapid cells' death). Consequently, the TCRCs generated after this kind of experiments do not have the features necessary for statistical classification. The main practical motivation of this study was to find a way to determine the lowest concentration of toxin that can kill the cells (i.e. the value of chemical's initial concentration that separate the persistent cells' populations from those that go to extinction).

Once the parameters are estimated, the stability analysis and numerical simulations can provide a range of values for the chemical's initial concentration that give TCRCs with useful features. The asymptotic analysis of the proposed model

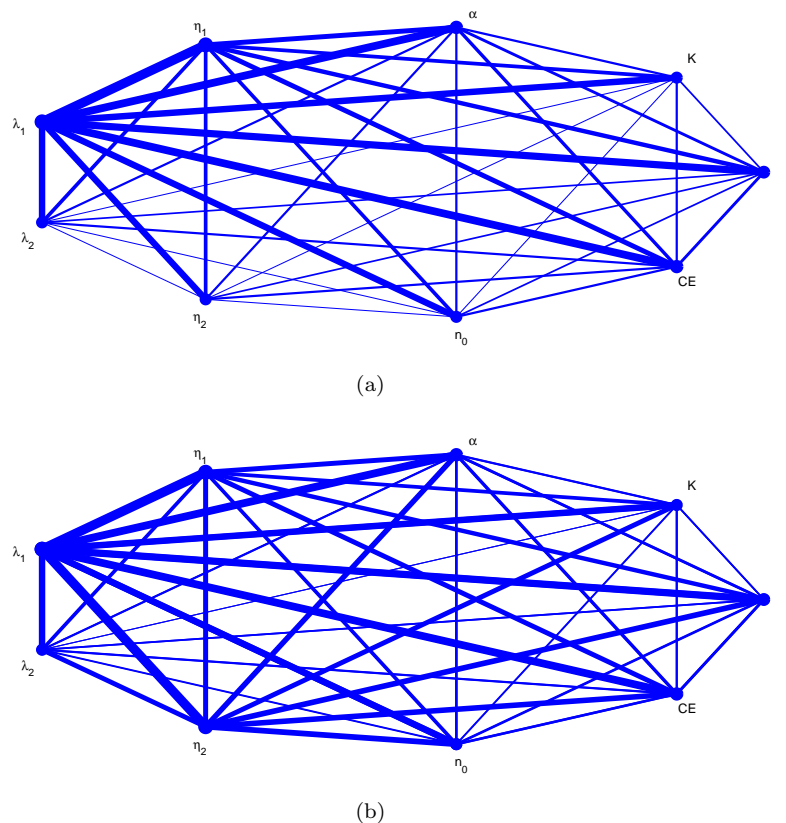


FIGURE 14. Network graph visualizing the second order GSA indices for (a) PF431396 with $CE(0)=10\mu\text{M}$ (b) ABT888 with $CE(0)=400\mu\text{M}$

shows that persistence occurs if the initial concentration of chemical is smaller than a certain threshold. For the persistent trajectories, the cell index tends toward the nominal carrying capacity of the population, and the internal and external chemical concentrations tend toward zero. Extinction depends on the initial concentration of toxicant and the entire past history of the population. The extinction behavior is more complex because when the cell index tends to zero, the internal and external toxicant concentrations can approach an infinite number of limiting values.

We illustrate this analytical and numerical analysis for monastrol, but the proposed model based on first order kinetics gives accurate predictions for the TCRCs corresponding to the 63 chemicals considered in this study. However, there might be limitations for different experimental settings or chemicals with different MOA.

We plan to further extend this study using uncertainty quantification methods. Although the proposed model was validated by experimental data, parameter uncertainty is inherent. To account for parameter uncertainty and calculate the probability of extinction, we have considered random perturbation of parameters' values, we have done Monte Carlo simulations, and we have calculated the Sobol's GSA indices to find out which parameters are important and sensitive to change.

Based on the results of the GSA we plan to introduce a model with random parameters, for example to consider a system of stochastic differential equations similar with the ones in [15, 16].

Acknowledgments. This work was supported by a Mathematics of Information Technology and Complex Systems (MITACS) Accelerate Clusters grant and by a grant from the Natural Sciences and Engineering Research Council of Canada.

Appendix A. Center manifold. Let $[0, CE^* \frac{\lambda_1^2}{\eta_1^2}, CE^*]$ be any extinction equilibrium with $CE^* \neq \frac{\beta\eta_1^2}{\alpha\lambda_1^2}$. To translate this fixed point to the origin, we consider two new variables $y_1 = C_0 - CE^* \frac{\lambda_1^2}{\eta_1^2}$ and $z_1 = CE - CE^*$. The corresponding model in terms of $[n, y_1, z_1]$ is given by the following system of equations

$$\begin{aligned} \frac{dn(t)}{dt} &= \left(\beta - \frac{\alpha\lambda_1^2 CE^*}{\eta_1^2} \right) n(t) - \frac{\beta n(t)^2}{K} - \alpha y_1(t)n(t) \\ \frac{dy_1(t)}{dt} &= \lambda_1^2 z_1(t) - \eta_1^2 y_1(t) \\ \frac{dz_1(t)}{dt} &= \frac{CE^* (-\eta_1^2 \eta_2^2 + \lambda_1^2 \lambda_2^2) n(t)}{\eta_1^2} + \lambda_2^2 y_1(t)n(t) - \eta_2^2 z_1(t)n(t). \end{aligned} \quad (16)$$

The diagonal form of this system is

$$\begin{aligned} \frac{dx}{dt} &= -\frac{(\alpha CE^* \lambda_1^2 - \beta \eta_1^2)}{\eta_1^2} x + F_1(x, y, z) \\ \frac{dy}{dt} &= -\eta_1^2 y + F_2(x, y, z) \\ \frac{dz}{dt} &= F_3(x, y, z) \end{aligned} \quad (17)$$

where the new variables are

$$\begin{aligned} x &= \frac{CE^* (\eta_1^2 \eta_2^2 - \lambda_1^2 \lambda_2^2)}{(\alpha CE^* \lambda_1^2 - \beta \eta_1^2)} n \\ y &= \frac{(\eta_1^2 \eta_2^2 - \lambda_1^2 \lambda_2^2) CE^* \lambda_1^2}{(\alpha CE^* \lambda_1^2 - \eta_1^2 - \beta \eta_1^2) \eta_1^2} n + y_1 - \frac{\lambda_1^2}{\eta_1^2} z_1 \\ z &= \frac{CE^* (\eta_1^2 \eta_2^2 - \lambda_1^2 \lambda_2^2)}{(\alpha CE^* \lambda_1^2 - \beta \eta_1^2)} n + z_1, \end{aligned}$$

and the non-linear terms are of the form $F_i(x, y, z) = c_1 x^2 + c_2 xy + c_3 xz$, for some coefficients c_i , $i = 1, \dots, 3$ that can be easily determined using MAPLE. The local center manifold at $[0, 0, 0]$ (see chapter 2.12 in [21]) is of the form

$$\begin{aligned} x &= h(z) = a_1 z^2 + a_2 z^3 + \dots \\ y &= g(z) = b_1 z^2 + b_2 z^3 + \dots \end{aligned} \quad (18)$$

To determine the coefficients a_i , b_i , $i = 1, \dots$, we substitute the previous expansions into the following equation (see Theorem 1, chapter 2.12 in [21])

$$F_3(h(z), g(z), z) D \begin{bmatrix} h(z) \\ g(z) \end{bmatrix} - \begin{bmatrix} -\frac{(\alpha CE^* \lambda_1^2 - \beta \eta_1^2)}{\eta_1^2} h(z) + F_1(h(z), g(z), z) \\ -\eta_1^2 g(z) + F_2(h(z), g(z), z) \end{bmatrix} = 0 \quad (19)$$

Simple, but long calculations give $a_i = b_i = 0, i = 1, \dots$, so the initial system is topologically conjugate (see Theorem 2, chapter 2.12 in [21]) to the system

$$\begin{aligned} \frac{dx}{dt} &= -\frac{(\alpha CE^* \lambda_1^2 - \beta \eta_1^2)}{\eta_1^2} x \\ \frac{dy}{dt} &= -\eta_1^2 y \\ \frac{dz}{dt} &= 0. \end{aligned} \tag{20}$$

and, in terms of the initial variables, the local center manifold is the line $\{(n, C_0, CE) : n = 0, \lambda_1^2 CE - \eta_1^2 C_0 = 0\}$.

Appendix B. The proofs. Before we give the proofs for the lemma and theorem included in section 3, we notice that the we have the following formulas for the solutions of system (5) (see also [4]). Since equations (3) and (4) are linear we have

$$C_0(t) = \lambda_1^2 e^{-\eta_1^2 t} \int_0^t CE(s) e^{\eta_1^2 s} ds \tag{21}$$

$$\begin{aligned} CE(t) &= CE(0) \exp\left(-\eta_2^2 \int_0^t n(s) ds\right) + \lambda_2^2 \exp\left(-\eta_2^2 \int_0^t n(s) ds\right) \\ &\quad \int_0^t C_0(s) n(s) \exp\left(\eta_2^2 \int_0^s n(l) dl\right) ds, \quad t \geq 0. \end{aligned} \tag{22}$$

Equation (2) is a Bernoulli equation, so we get

$$n(t) = \frac{n(0) \exp\left(\beta t - \alpha \int_0^t C_0(s) ds\right)}{1 + \frac{n(0)\beta}{K} \int_0^t \exp\left(\beta s - \alpha \int_0^s C_0(l) dl\right) ds}, \quad t \geq 0. \tag{23}$$

B.1. Proof of lemma 3.1. Assume that there exists $T_1 > 0$ such that $CE(T_1) > CE(0)$ or $CE(T_1) \leq 0$, and suppose that $[0, T]$ is the maximum interval for which $0 < CE(t) \leq CE(0)$. Then we should either have $CE(T) = CE(0)$ or $CE(T) = 0$.

From (21) we get for any $0 < t \leq T$

$$\begin{aligned} 0 \leq C_0(t) &= \lambda_1^2 e^{-\eta_1^2 t} \int_0^t CE(s) e^{\eta_1^2 s} ds \leq \lambda_1^2 e^{-\eta_1^2 t} CE(0) \int_0^t e^{\eta_1^2 s} ds \\ &= \frac{\lambda_1^2 CE(0)}{\eta_1^2} \left(1 - e^{-\eta_1^2 t}\right) \leq \frac{\lambda_1^2 CE(0)}{\eta_1^2} \end{aligned}$$

Thus, for any $0 < t \leq T$

$$0 \leq \int_0^t C_0(s) ds \leq \frac{\lambda_1^2 CE(0)t}{\eta_1^2} < \infty \tag{24}$$

and from equation (23) we have $0 < n(t) \leq n(0)e^{\beta t}$. Using this and equation (22) we get for any $0 < t \leq T$

$$CE(t) \geq CE(0) \exp\left(-\eta_2^2 \int_0^t n(s) ds\right) > 0, \tag{25}$$

so $CE(T) > 0$.

Moreover, if $CE(T) = CE(0)$, then from equation (4) we obtain

$$\begin{aligned} \left. \frac{dCE}{dt} \right|_{t=T} &= \lambda_2^2 C_0(T)n(T) - \eta_2^2 CE(T)n(T) \\ &\leq CE(0)n(T) \left(\frac{\lambda_1^2 \lambda_2^2}{\eta_1^2} - \eta_2^2 \right) < 0 \end{aligned} \quad (26)$$

So there exists $\delta > 0$ such that $CE(t) \leq CE(0)$ for all $t \in [T, T + \delta]$, and we have a contradiction with the definition of the interval $[0, T]$.

Thus, we have proved that $0 < CE(t) \leq CE(0)$ for any $t \geq 0$. Using this and (21) we also get $0 \leq C_0(t) \leq \frac{\lambda_1^2 CE(0)}{\eta_1^2}$, for any $t \geq 0$. Consequently inequalities (24) are true for any $t \geq 0$, and from (23) we get $n(t) > 0$, for all $t \geq 0$.

B.2. Proof of Theorem 3.2. Since from Lemma 3.1 we know that $n(t) > 0$, for any $t \geq 0$, we have $\int_0^\infty n(t)dt = \infty$ or $0 \leq \int_0^\infty n(t)dt < \infty$.

If $|n|_1 = \int_0^\infty n(t)dt < \infty$ then for any $t \geq 0$ we also have

$$\int_0^t C_0(s)n(s) \exp\left(\eta_2^2 \int_0^s n(l)dl\right) ds \leq \frac{\lambda_1^2 CE(0)}{\eta_1^2} \exp(\eta_2^2 |n|_1) |n|_1, \quad (27)$$

so $0 \leq \int_0^\infty C_0(s)n(s) \exp(\eta_2^2 \int_0^s n(l)dl) ds = M < \infty$. Thus from (22) we get

$$\lim_{t \rightarrow \infty} CE(t) = CE(0) \exp(-\eta_2^2 |n|_1) + \lambda_2^2 M \exp(-\eta_2^2 |n|_1) < \infty. \quad (28)$$

Consequently, there exists $T_1 > 0$ such that for any $t > T_1$ we have $CE(t) > CE(0) \exp(-\eta_2^2 |n|_1)/2$. This implies that $\int_0^\infty CE(s)e^{\eta_1^2 s} ds = \infty$ because for any $t > T_1$ we have

$$\int_0^t CE(s)e^{\eta_1^2 s} ds \geq \int_{T_1}^t CE(s)e^{\eta_1^2 s} ds \geq CE(0) \exp(-\eta_2^2 |n|_1)/2 \int_{T_1}^t e^{\eta_1^2 s} ds.$$

So we can apply L'Hospital rule in (21), and we get

$$\lim_{t \rightarrow \infty} C_0(t) = \frac{\lambda_1^2}{\eta_1^2} \lim_{t \rightarrow \infty} CE(t). \quad (29)$$

Thus, if $\int_0^\infty n(t)dt < \infty$, then $\lim_{t \rightarrow \infty} CE(t)$ and $\lim_{t \rightarrow \infty} C_0(t)$ exist and they are related by the previous equation.

Next we consider the case when $\int_0^\infty n(t)dt = \infty$. If $0 \leq \int_0^\infty C_0(s)n(s) \exp(\eta_2^2 \int_0^s n(l)dl) ds < \infty$, from (22) we get $\lim_{t \rightarrow \infty} CE(t) = 0$. On the other hand, if $\int_0^\infty C_0(s)n(s) \exp(\eta_2^2 \int_0^s n(l)dl) ds = \infty$, from L'Hospital rule in (22) we have

$$0 \leq \frac{\lambda_2^2}{\eta_2^2} \liminf_{t \rightarrow \infty} C_0(t) \leq \liminf_{t \rightarrow \infty} CE(t) \leq \limsup_{t \rightarrow \infty} CE(t) \leq \frac{\lambda_2^2}{\eta_2^2} \limsup_{t \rightarrow \infty} C_0(t) \quad (30)$$

Similarly, from (21) we either get that $\lim_{t \rightarrow \infty} C_0(t) = 0$ (if $\int_0^\infty CE(s)e^{\eta_1^2 s} ds < \infty$), or we have

$$0 \leq \frac{\lambda_1^2}{\eta_1^2} \liminf_{t \rightarrow \infty} CE(t) \leq \liminf_{t \rightarrow \infty} C_0(t) \leq \limsup_{t \rightarrow \infty} C_0(t) \leq \frac{\lambda_1^2}{\eta_1^2} \limsup_{t \rightarrow \infty} CE(t), \quad (31)$$

(if $\int_0^\infty CE(s)e^{\eta_1^2 s} ds = \infty$). All these possible cases give

$$\lim_{t \rightarrow \infty} C_0(t) = \lim_{t \rightarrow \infty} CE(t) = 0, \quad (32)$$

because $\eta_1^2 \eta_2^2 - \lambda_1^2 \lambda_2^2 > 0$. Thus, if $\int_0^\infty n(t)dt = \infty$, then $\lim_{t \rightarrow \infty} CE(t)$ and $\lim_{t \rightarrow \infty} C_0(t)$ exist and they are equal with zero.

In conclusion, we have shown that in both cases $\lim_{t \rightarrow \infty} CE(t)$ and $\lim_{t \rightarrow \infty} C_0(t)$ exist, and we have $\lim_{t \rightarrow \infty} C_0(t) = \frac{\lambda_1^2}{\eta_1^2} \lim_{t \rightarrow \infty} CE(t)$.

To prove the first part of the theorem we first show that if $\lim_{t \rightarrow \infty} CE(t) < \frac{\beta \eta_1^2}{\alpha \lambda_1^2}$ then $\int_0^\infty n(t)dt = \infty$. From the previous discussion we know that $\lim_{t \rightarrow \infty} CE(t) < \frac{\beta \eta_1^2}{\alpha \lambda_1^2}$ implies $\lim_{t \rightarrow \infty} C_0(t) < \frac{\beta}{\alpha}$, and consequently

$$\lim_{t \rightarrow \infty} \exp\left(\beta t - \alpha \int_0^t C_0(s)ds\right) = \infty, \quad \lim_{t \rightarrow \infty} \int_0^t \exp\left(\beta s - \alpha \int_0^s C_0(l)dl\right) ds = \infty$$

Hence, from L'Hospital rule in (23) we have

$$0 \leq \lim_{t \rightarrow \infty} n(t) = \frac{K(\beta - \alpha \lim_{t \rightarrow \infty} C_0(t))}{\beta} \tag{33}$$

If $\int_0^\infty n(t)dt < \infty$ then we should have $\lim_{t \rightarrow \infty} n(t) = 0$, and this implies $\lim_{t \rightarrow \infty} C_0(t) = \frac{\beta}{\alpha}$. But $\lim_{t \rightarrow \infty} C_0(t) < \frac{\beta}{\alpha}$, so $\int_0^\infty n(t)dt = \infty$. Hence $\lim_{t \rightarrow \infty} C_0(t) = \lim_{t \rightarrow \infty} CE(t) = 0$, and replacing in (33) we get $\lim_{t \rightarrow \infty} n(t) = K$. This completes the proof of the first part of the theorem.

Next, if $\lim_{t \rightarrow \infty} CE(t) = CE^* > \frac{\beta \eta_1^2}{\alpha \lambda_1^2} > 0$, then from the previous discussion we know that $|n|_1 = \int_0^\infty n(t)dt < \infty$. Moreover, $\lim_{t \rightarrow \infty} C_0(t) = \frac{\lambda_1^2}{\eta_1^2} CE^* > \frac{\beta}{\alpha} > 0$, so

$$\lim_{t \rightarrow \infty} \exp\left(\beta t - \alpha \int_0^t C_0(s)ds\right) = 0.$$

Thus $\lim_{t \rightarrow \infty} n(t) = 0$ because from (23) we have

$$0 < n(t) \leq n(0) \exp\left(\beta t - \alpha \int_0^t C_0(s)ds\right).$$

REFERENCES

[1] C. Biernacki, G. Celeux and G. Govaert, [Choosing starting values for the EM algorithm for getting the highest likelihood in multivariate gaussian mixture models](#), *Comput. Statist. Data Anal.*, **41** (2003), 561–575.

[2] F. Cannavó, Sensitivity analysis for volcanic source modeling quality assessment and model selection, *Computers & Geosciences.*, **44** (2012), 52–59.

[3] Z. Ghahramani and S. Roweis, Learning nonlinear dynamical systems using an EM algorithm, in *Advances in Neural Information Processing Systems* (eds. M. Kearns, S. Solla and C. D.A.), MIT Press, 1999, 599–605.

[4] T. Hallam, C. Clark and G. Jordan, Effects of toxicants on populations: A qualitative approach II. First order kinetics, *J. Math. Biology*, **18** (1983), 25–37.

[5] J. He and K. Wang, [The survival analysis for a population in a polluted environment](#), *Nonlinear Analysis: Real World Applications*, **10** (2009), 1555–1571.

[6] B. Huang and J. Xing, Dynamic modeling and prediction of cytotoxicity on microelectronic cell sensor array, *The Canadian Journal of Chemical Engineering*, **84** (2006), 393–405.

[7] Q. Huang, L. Parshotam, H. Wang, C. Bampfyld and M. Lewis, [A model for the impact of contaminants on fish population dynamics](#), *Journal of Theoretical Biology*, **334** (2013), 71–79.

[8] F. Ibrahim, B. Huang, J. Xing and S. Gabos, Early determination of toxicant concentration in water supply using MHE, *Water Research*, **44** (2010), 3252–3260.

[9] A. M. Jarrett, Y. Liu, N. Cogan and M. Y. Hussaini, [Global sensitivity analysis used to interpret biological experimental results](#), *Journal of Mathematical Biology*, **71** (2015), 151–170.

[10] J. Jiao, W. Long and L. Chen, [A single stage-structured population model with mature individuals in a polluted environment and pulse input of environmental toxin](#), *Nonlinear Analysis: Real World Applications*, **10** (2009), 3073–3081.

- [11] S. Julier, J. Uhlmann and H. Durrant-White, [A new method for the nonlinear transformation of means and covariances in filters and estimators](#), *IEEE Trans. Aut. Control*, **45** (2000), 477–482.
- [12] S. Julier, J. Uhlmann and H. Durrant-Whyte, [A new approach for filtering nonlinear systems](#), in *American Control Conference, Seattle, Washington*, 1995, 1628–1632.
- [13] A. Kiparissides, S. Kucherenko, A. Mantalaris and E. N. Pistikopoulos, [Global sensitivity analysis challenges in biological systems modeling](#), *Industrial & Engineering Chemistry Research*, **48** (2009), 7168–7180.
- [14] K. Kothawad, A. Pathan and M. Logad, Evaluation of in vitro anti-cancer activity of fruit lagenaria siceraria against MCF7, HOP62 and DU145 cell line, *Int. J. Pharm. & Technol.*, **4** (2012), 3909–4392.
- [15] M. Liu and K. Wang, Survival analysis of stochastic single-species population models in polluted environments, *Ecological Modeling*, **220** (2009), 1347–1357.
- [16] M. Liu, K. Wang and X. Liu, [Long term behaviors of stochastic single-species growth models in a polluted environment](#), *Applied Mathematical Modelling*, **35** (2011), 752–762.
- [17] X. Meng and D. Van Dyk, [The EM algorithm - an old folk-song to a fast new tune](#), *J.R. Statist. Soc.B.*, **59** (1997), 511–567.
- [18] R. Neal and G. Hinton, A view of the EM algorithm that justifies incremental, sparse, an other variants, in *Learning in Graphical Models* (ed. M. Jordan), **89** (1998), 355–368.
- [19] T. Pan, B. Huang, W. Zhang, S. Gabos, D. Huang and V. Devendran, Cytotoxicity assessment based on the AUC₅₀ using multi-concentration time-dependent cellular response curves, *Anal. Chim. Acta*, **764** (2013), 44–52.
- [20] T. Pan, S. Khare, F. Ackah, B. Huang, W. Zhang, S. Gabos, C. Jin and M. Stampff, In vitro cytotoxicity assessment based on KC₅₀ with real-time cell analyzer (RTCA) assay, *Comp. Biol. Chem.*, **47** (2013), 113–120.
- [21] L. Perko, *Differential Equations and Dynamical Systems*, Springer, New York, 2001.
- [22] R. Shumway and D. Stoffer, [An approach to time series smoothing and forecasting using the EM algorithm](#), *J. Time Ser. Anal.*, **3** (1982), 253–264.
- [23] I. M. Sobol, [Global sensitivity indices for nonlinear mathematical models and their Monte Carlo estimates](#), *Mathematics and Computers in Simulation*, **55** (2001), 271–280.
- [24] H. Thieme, *Mathematics in Population Biology*, Princeton Series in theoretical and Computational Biology., 2003.
- [25] E. A. Wan, R. Van der Merwe and A. T. Nelson, Dual estimation and the unscented transformation, in *Advances in Neural Information Processing Systems* (ed. M. I. J. et al.), MIT Press, 2000.
- [26] C. Wu, [On the convergence properties of the EM algorithm](#), *The Annals of Statistics*, **11** (1983), 95–103.
- [27] Z. Xi, S. Khare, A. Cheung, B. Huang, T. Pan, W. Zhang, F. Ibrahim, C. Jin and S. Gabos, [Mode of action classification of chemicals using multi-concentration time-dependent cellular response profiles](#), *Comp. Biol. Chem.*, **49** (2014), 23–35.
- [28] J. Xing, L. Zhu, S. Gabos and L. Xie, [Microelectronic cell sensor assay for detection of cytotoxicity and prediction of acute toxicity](#), *Toxicology in Vitro*, **20** (2006), 995–1004.
- [29] M. Zhang, D. Aguilera, C. Das, H. Vasquez, P. Zage, V. Gopalakrishnan and J. Wolff, Measuring cytotoxicity: A new perspective on LC₅₀, *Anticancer Res.*, **27** (2007), 35–38.

Received February 29, 2016; Accepted October 17, 2016.

E-mail address: popescuc@macewan.ca

E-mail address: deng2@ualberta.ca

E-mail address: yaushu.wong@ualberta.ca

E-mail address: yile2@ualberta.ca

E-mail address: weiping2@ualberta.ca

E-mail address: sgabos@ualberta.ca

E-mail address: yhuang@ucalgary.ca

E-mail address: cjin@aceabio.com