Parameter Estimation in a Cell Cycle Model for Frog Egg Extracts

Jason W. Zwolak*, **John J. Tyson****, and **Layne T. Watson***
Departments of Computer Science* and Biology**
Virginia Polytechnic Institute and State University
Blacksburg, Virginia 24061-0106
e-mail: jzwolak@vt.edu

Keywords: Computational biology, ordinary differential equations, parameter estimation

Abstract
Cell cycle models used in biology can be very complex, involving many parameters with initially unknown values. The values of the parameters vastly affect the accuracy of a model in representing real biological cells. Typically people search for the best parameters of a model using the computer only as a tool to run simulations. In this paper methods and results are described for a computer program that searches for parameters to a specific model using well tested algorithms. The code for this program uses ODRPACK for parameter estimation and LSODAR to solve the differential equations that comprise the model.

1. INTRODUCTION

Computational models of cell growth and division involve digital representation of a complex network of biochemical reactions within cells. These reactions can be described by a system of nonlinear ordinary differential equations, according to the principles of biochemical kinetics. Rate constants and binding constants enter as parameters in the differential equations, and must be estimated by fitting solutions of the equations to experimental data.

This work concerns some classical experiments on activation of MPF (M-phase promoting factor) in frog egg extracts. MPF is a dimer of cyclin and Cdc2 (a protein kinase that drives egg nuclei into mitosis). In the experimental preparation, a fixed amount of cyclin is added to an extract containing an excess of Cdc2 subunits. If the amount of cyclin added is below a threshold, MPF activity never appears. Above the threshold, MPF is activated but only after a characteristic time lag. The time lag decreases abruptly as total-cyclin-added increases above the threshold. The goal is to fit this data with a reasonable model of the underlying biochemistry, which keeps track of cyclin monomers, Cdc2 monomers, and the phosphorylation state of cyclin/Cdc2 dimers.

ODRPACK, based on the orthogonal distance between experimental data and the model, is used for the nonlinear regression to estimate the unknown rate constants (ODE parameters). The ability of this algorithm to arbitrarily weight data values, and to treat both the abscissa and ordinates as uncertain, is crucial, given the sparsity and uncertainty of available biological data. Constructing the model's predictions of experimental data requires simulating MPF activity as a function of time after addition of cyclin. These simulations yield the cyclin threshold for MPF activation, and the time lag (the time necessary for MPF activity to reach one-half of its asymptotic value, for supra-threshold amounts of cyclin added to the extracts).

The complete calculation is expensive, because the ODE's are stiff, and must be solved numerous times for the nonlinear regression. Also, because of local minima, the nonlinear regression must be done from many starting points to adequately explore the parameter space. There are potential sources for parallelism in the ODE solution itself, the estimation of partial derivatives of the ODE solution, and multiple starting points for regression. Numerical results are presented for a four-component, ten-parameter model.

The model described in this paper is a realistic model of the biochemical kinetics of MPF activity in frog egg extracts. However, the model ignores a number of other proteins that affect the cell cycle. To study more complete models of cell cycle control, more components must be added to the model, and other measurable phenomena incorporated in the cost function. As the modeling fidelity is increased, the mathematical and computational complexities of the problem grow rapidly. Efficient and accurate tools for parameter estimation will be needed to build computational models of the complex control networks operating within cells, which is one of the main goals of bioinformatics in the postgenomic era.

Section 2 outlines the biological model and provides the experimental data for said model. An overview of the code along with descriptions of the tools (ODRPACK and LSODAR) used by the code can be found in Section 3. Section 4 contains a more detailed pseudocode for the algorithm. The results of the parameter estimation are in Section 5. The conclusion and future work are in Section 6.

2. PROBLEM STATEMENT

The primary concern of this model is to keep track of the concentration of active MPF. This model uses data from an experiment done by J. Moore [9] that is based on an experiment done by Solomon et al. [14]. Wee1 and Cdc25 are the primary proteins affecting the activity of MPF. MPF also affects the activity of Wee1 and Cdc25. The system being modeled has two feedback loops (one involving Wee1 and one involving Cdc25). Goldbeter-Koshland functions, G, [5] are used to describe the feedback loops. The equations

$$\frac{dM}{dt} = k_d(C - M) - k_w M,$$
$$k_d = v'_d(1 - d) + v''_d d,$$
$$k_w = v'_w(1 - w) + v''_w w,$$
$$d = G(M, v'''_d, K_{md}, K_{mdr}),$$
$$w = G(v'''_w, M, K_{mw}, K_{mwr}),$$
$$G(v_1, v_2, K_1, K_2) = \frac{2v_1 K_2}{\beta + \sqrt{\beta^2 - 4v_1 K_2(v_2 - v_1)}},$$
$$\beta = v_2 - v_1 + v_1 K_2 + v_2 K_1.$$

are used to model MPF activity for this problem [8]. In these equations, C is the total concentration of cyclin (equivalent to the total concentration of MPF for this model), M is the concentration of active MPF, d is the scaled concentration of active Cdc25, and w is the scaled concentration of active Wee1. k_d and k_w represent the effect of the feedback loops on MPF from Cdc25 and Wee1, respectively. In the equations for k_d and k_w both active (d and w) and inactive ($(1 - d)$ and $(1 - w)$) forms of Cdc25 and Wee1 appear because the inactive forms may still have some activity. The concentrations of active Cdc25 and Wee1 are scaled so they represent the fraction of active concentration over total concentration; therefore the concentration of inactive Cdc25 and Wee1 can be written as $(1 - d)$ and $(1 - w)$.

The equations for active Cdc25 (d) and active Wee1 (w) come from differential equations using Michaelis-Menten kinetics. We assume that Cdc25 and Wee1 reach a steady state much faster than MPF. Therefore, the model can be approximated by setting the time rate of change of d and w to zero. The resulting equations can be solved for d and w [5]. The solutions for d and w have the form of the Goldbeter-Koshland function G.

There are ten parameters in the model—v'_d, v''_d, v'''_d, v'_w, v''_w, v'''_w, K_{md}, K_{mdr}, K_{mw}, and K_{mwr}. All the parameters can be varied to change the model behavior. Ultimately the model should be consistent with experimental data. There are six data points to be fit. Therefore not all the parameters can be estimated. The parameter estimates from Marlovits [8] are used as

the parameter values for some parameters and the initial guess for the remaining parameters. K_{md}, K_{mdr}, K_{mw}, and K_{mwr} are set to the estimates from Marlovits [8] and v'''_w is set equal to v'''_d. The number of free parameters is then five. The experimental data available are in Table 1.

Table 1. Experimental data for MPF activity in frog egg extracts [9].

Total Cyclin	Time Lag (min)
0.20	45
0.25	40
0.30	30
0.50	20
≤ 0.18	never activates
≥ 0.06	never inactivates

The activity of MPF versus time with MPF initially inactive looks like any one of the curves in Figure 1. Each curve corresponds to a different total concentration of cyclin (C). In most of the curves seen in Figure 1, MPF activates after a time lag. However, if C is below a threshold, MPF will never activate. This threshold is called the activation threshold. There is also an inactivation threshold. For values of C above the inactivation threshold MPF will never inactivate, if initially active. The first four data from Table 1 represent time lags for MPF activation. The fifth datum represents the activation threshold. The last datum represents the inactivation threshold. For the rest of this paper the inactivation threshold will be represented as $1/3$ the activation threshold instead of 0.06.

Neither the time lags nor the thresholds appear directly in the model's output. Therefore, they must be defined and calculated from the model. The time lag is defined as the time when active MPF is half way between its initial and final concentration. Time lag is not defined if MPF never changes from mostly active to mostly inactive or vice versa (where mostly is defined as more than 50 percent). The threshold for MPF activation is defined as the concentration of C below which MPF will remain inactive if initially inactive. For all concentrations of C above the threshold, MPF will activate after some time lag. The threshold for MPF inactivation is defined similarly. For all concentrations of C above the threshold MPF will remain active if initially active. For concentrations below the threshold MPF will inactivate after some time lag.
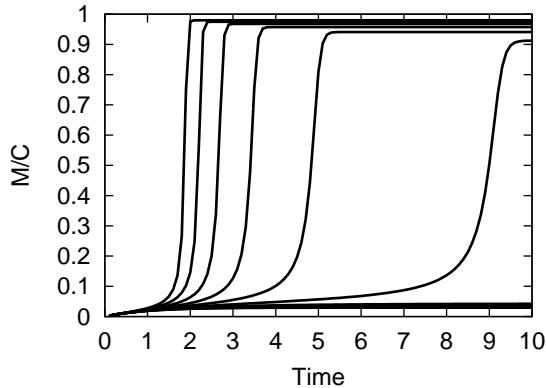
Figure 1. Percent total cyclin in active MPF M/C versus time t for multiple concentrations of total cyclin

3. METHODS

The first step is to use the ODE to calculate a model function f corresponding to the data in Table 1. The components of f differ depending on which data are being compared. In other words, for the first four data f evaluates the time lag as a function of total cyclin. For the fifth datum f evaluates the activation threshold; for the sixth datum f evaluates the ratio of the activation threshold to the inactivation threshold. The values of f are dependent on the parameters. For the time lags the values are also dependent on the total cyclin concentration. The input variables to f will be represented by x.

LSODAR is used to solve the ODE and to find the time lags and thresholds from the solution to the ODE (indirectly). $f(x)$ is used by ODRPACK to find the rate constants giving the $f(x)$ that best fits the experimental data in Table 1.

3.1. LSODAR

LSODAR is a variant of LSODE ([11], [6], [7]) that automatically switches between stiff and non-stiff methods and has a root finder. LSODAR starts with a non-stiff method and switches to a stiff method if necessary. LSODAR's root finder is used in this application to find the time lag for MPF activation.

For non-stiff problems LSODAR uses Adams-Moulton (AM) of orders 1 to 12. For stiff problems LSODAR uses backward differentiation formulas (BDF) of orders 1 to 5. With both methods LSODAR varies the step size and order. LSODAR switches from AM to BDF when AM is no longer stable for the problem or cannot meet the accuracy requirements efficiently [10].

The present problem uses LSODAR to solve for $M(t)$ (the concentration of active MPF with respect to time). The tolerances are set to 10^{-12} for both relative

and absolute error. A tolerance of 10^{-10} is used when calculating a root for a function of the form

$$M(t) - M_{root},$$

where M_{root} is the value of the function $M(t)$ for which a time, t, is desired.

LSODAR takes, as an argument, a user written function, GEX, that evaluates equations based on the variables involved in the ODE that LSODAR is solving. For this problem GEX evaluates $M - M_{root}$. GEX returns evaluations of its equations to LSODAR and LSODAR looks for roots for those equations. When a sign change is detected LSODAR has bracketed a root and begins an algorithm based on the ROOT function described below. After each iteration of ROOT, LSODAR must evaluate a point on the solution curve of the ODE as requested by ROOT. Each evaluation involves interpolating the ODE solution $M(t)$. This interpolation formula is defined as part of the AM [13] or BDF [4] method (depending on which is currently being used by LSODAR).

3.2. ODRPACK

ODRPACK is used to estimate the rate constants that fit time lag versus total cyclin to the experimental data in Table 1. ODRPACK finds an estimate for the rate constants by minimizing the weighted orthogonal distance between the experimental data and the calculated curve.

The present problem explicitly relates time lag to the total concentration of cyclin in the cell. Precisely,

$$y = f(x; \beta),$$

where y is time lag, x is total cyclin, and β is a vector of the rate constants. ODRPACK takes an equation of this form and experimental data for x and y to minimize

$$E = \min_{\beta,\delta,\epsilon} \left(\sum_{i=1}^{n} w_{\epsilon_i} \epsilon_i^2 + w_{\delta_i} \delta_i^2 \right),$$

where n is the number of experimental data points, ϵ_i is the error in the dependent variable y for point i, δ_i is the error in the explanatory variable x for point i, and w_{ϵ_i} and w_{δ_i} are the weights for ϵ_i and δ_i, respectively. E will be referred to as the weighted sum of squares. β, δ, and ϵ are subject to the constraints

$$y_i = f(x_i + \delta_i; \beta) - \epsilon_i,$$

where $i = 1, \ldots, n$ indexes the experimental data points.

ODRPACK actually minimizes a more general objective function

$$E = \min_{\beta,\delta,\epsilon} \left(\sum_{i=1}^{n} \epsilon_i^T w_{\epsilon_i} \epsilon_i + \delta_i^T w_{\delta_i} \delta_i \right),$$

where ϵ_i and δ_i are vectors for the errors in the dependent variable and errors in the explanatory variable, respectively. w_{ϵ_i} and w_{δ_i} are matrices of weights for ϵ_i and δ_i, respectively ([2], [1]). Note that x and y, from the previous description of ODRPACK, are vectors and the function f is a vector-valued function in the general case. The present problem can be thought of as using the scalar version of ODRPACK, since the present problem has w_{ϵ_i} and w_{δ_i} as matrices of one element and ϵ_i and δ_i as vectors of one element.

The function $f(x + \delta; \beta)$ is implemented in FORTRAN and used by ODRPACK. Constraints are put on β by setting a flag (when β is invalid) before returning from the user supplied function. This is used to prevent the rate constants from becoming negative, which does not make sense biologically.

ODRPACK uses a trust region Levenberg-Marquardt method with scaling to minimize the objective function [2]. In doing so ODRPACK needs to calculate the Jacobian matrices for β and δ. ODRPACK can calculate the Jacobian matrices by finite differences or by a user supplied routine. Finite differences were used here.

3.3. ROOT

ROOT is based on ZEROIN [12], which is in turn based on code by Dekker [3]. ROOT uses a combination of the secant and bisection methods where the secant method is used by default. ROOT has two working approximations of the root: A and B. The approximations always satisfy the constraint

$$g(A) * g(B) \le 0,$$

where $g(t) = M - M_{root}$ and t is time in this problem (note that M is dependent on t). Furthermore, A is the better approximation of the root of $g(t)$. A is replaced in each iteration by a better approximation and B remains the same or changes to the old A, whichever satisfies the above equation. ROOT switches to the bisection method under two circumstances: when the secant method is converging too slowly, or when a large error is introduced because of limitations in machine precision. Notice the bisection method will not suffer from large error because it computes

$$\frac{A + B}{2}$$

for each iteration.

The initial approximations, for A and B, come from LSODAR's evaluation of GEX before and after LSODAR noticed a sign change. ROOT then requests values for g at new times until the approximation for the root of g is within the requested relative and absolute error.

4. ALGORITHM

In this section the algorithm used is described in some detail using pseudocode. Many of the function arguments used with ODRPACK's subroutine DODRC and ODEPACK's subroutine LSODAR do not appear in the pseudocode. Most of these arguments were set to default values, and others are not relevant to understanding the methods used to solve the present problem.

The main program sets up the input for DODRC and is as follows:

begin

$s := 8$; s is the number of significant digits in the response variable of f. The significant digits here come from the reliability of the computations done by the computer, not by experiments.

$n := 6$; n is the number of experimental data points.

$x := (0.20, 0.25, 0.30, 0.50, 0, 0)$; the vector x contains the total cyclin components of the experimental data. The last two elements are 0 because they correspond to the thresholds which have no explanatory variables.

$y := (45, 40, 30, 20, 0.18, 3)$; the first 4 elements of the vector y contain the time lags corresponding to total cyclin concentrations from above. The fifth element is the threshold for MPF activation. The sixth element is the ratio of the activation and inactivation thresholds.

$w_\delta := (25, 16, 11.11, 4, 1, 1)$; w_δ contains the weights for the errors in x. The weights of the error in the explanatory variables for the threshold information do not matter. Values of 1 were chosen.

$w_\epsilon := (4.938 \cdot 10^{-4}, 6.25 \cdot 10^{-4}, 1.111 \cdot 10^{-3}, 2.4 \cdot 10^{-3}, 30.86, 0.1111)$; w_ϵ contains the weights for the errors in y. The weights are the squared reciprocals of the corresponding data values, which makes all the errors in the objective function relative instead of absolute. w_δ uses the same method of squared reciprocals for the weights.

$\beta := (0.017, 0.17, 0.05, 0.01, 1.0, 0.05, 0.1, 1.0, 1.0, 0.1)$; β contains the initial guess for the rate constants. After DODRC has been called β will contain ODRPACK's best estimate for β given the arguments to DODRC. The parameters in β are v'_d, v''_d, v'''_d, v'_w, v''_w, v'''_w, K_{md}, K_{mdr}, K_{mw}, and K_{mwr}, respectively.

$I_x := (1, 1, 1, 1, 0, 0)$; I_x specifies which x to vary when finding the orthogonal distance. A value of 1 tells ODRPACK to vary the corresponding x, a value of 0 tells ODRPACK not to vary the corresponding

x. This will ensure no error is introduced in the explanatory variables for the threshold information and no time will be wasted finding the orthogonal distance for the threshold information.

$I_\beta := (1, 1, 1, 1, 1, 0, 0, 0, 0, 0)$; I_β specifies which parameters ODRPACK is to vary. A 1 tells ODRPACK to vary the corresponding parameter in β, a 0 tells ODRPACK to fix that parameter. Note that parameter number 6 (v_w''') is not actually fixed but is set equal to v_d'''. ODRPACK does not need to vary v_w'''. The actual assignment of v_w''' to v_d''' is done in FEX (defined later).

DODRC(FCN, $n, s, x, y, w_\delta, w_\epsilon, \beta, I_x, I_\beta, \ldots$); the ODRPACK subroutine used is DODRC. FCN is defined below.

end

The function FCN takes the explanatory variables and parameters to the ODE and returns the dependent variables. ODRPACK does not give FCN the explanatory variables directly from the experimental data. Instead, ODRPACK gives FCN the explanatory variables plus some error δ. In most cases the dependent variables returned by FCN will not match the dependent variables from the experimental data. Errors in measurements in the experimental data contribute to this mismatch. ODRPACK handles this by labeling the output of FCN as $y + \epsilon$. Precisely, let $X = x + \delta$ and $Y = y + \epsilon$. FCN takes arguments β and X and returns Y. The code for FCN follows.

subroutine FCN
  for $i := 1$ **step** 1 **until** $n - 2$ **do**
    $Y(i) :=$TIMELAG$(0, \beta, X(i))$; Calculate the time lag for a given total cyclin concentration.
  enddo
  $Y(n-1) :=$THRESHOLD$(0, \beta)$; Calculate the threshold for MPF activation.
  $Y(n) := Y(n-1)/$THRESHOLD$(1, \beta)$; Divide the threshold for MPF activation by the threshold for MPF inactivation.
end subroutine FCN

The TIMELAG routine calculates the time MPF takes to activate or inactivate given a specific total cyclin concentration (the third parameter C). The first parameter a to TIMELAG specifies whether MPF is initially active or inactive. The second parameter to TIMELAG is β (the parameters to the model). β and C are given to the model before LSODAR is called (although not shown in the pseudocode).

subroutine TIMELAG
  $T := 0$; (the initial time)
  $R_{tol} := 10^{-12}$; (relative error tolerance)
  $A_{tol} := 10^{-12}$; (absolute error tolerance)
  $M_{init} := 0$; (initial MPF concentration)
  $T_{out} := 1440$; (solve for the MPF concentration at this time)
  $N_g := 0$; (no roots are desired from LSODAR)
  $M_{inf} :=$LSODAR(FEX, M_{init}, T, T_{out}, R_{tol}, A_{tol}, N_g, JEX, GEX, $\ldots$);
  if $M_{inf} < C/2$ **then**
    TIMELAG:= 1440; (pseudo-infinite-lag)
    return;
  endif
  $M_{root} := M_{inf}/2$; (find a root at $M_{inf}/2$)
  $N_g := 1$; (one root is desired from LSODAR)
  LSODAR(FEX, M_{init}, T, T_{out}, R_{tol}, A_{tol}, N_g, JEX, GEX, $\ldots$);
  TIMELAG:= T_{out}; (the root is returned in T_{out})
  return;
end subroutine TIMELAG

The number 1440, construed as a pseudo-infinite-lag, is used to put a limit on how long to search for MPF activation. Effectively, the (computed) curve in Figure 2 will be flat when it reaches 1440 minutes. The true physical curve continues to increase after 1440 minutes. This modification creates a curve that does not precisely match the actual curve, but this modification does not affect the computation. All the experimental data is well below 1440 minutes (1 day). ODRPACK looks for the point on the calculated curve that is closest to the experimental data when calculating the error. Since the initial guess is not closer to the horizontal line at 1440 minutes than to the real curve, the flat portion will not cause ODRPACK to make wrong estimates for the rate constants.

The THRESHOLD function calculates the threshold for MPF activation or inactivation for a given set of rate constants β. If the first parameter a to THRESHOLD is 1 then MPF is initially active. If the first parameter to THRESHOLD is 0 then MPF is initially inactive.

subroutine THRESHOLD
  $b := 0.01$; Initialize the lower bound for the threshold. The initial value of b may not be a lower bound. A more realistic lower bound will be found later, if b is not a lower bound.
  $c := 1$; Initialize the upper bound for the threshold. The initial value of c may not be an upper bound. A more realistic upper bound will be found later, if c is not an upper bound.

$e_{tol} := 10^{-10}$; The error tolerance for calculating the threshold.

Find the lower bound on the threshold.

while $((a = 0$ **and** $\text{TIMELAG}(a, \beta, b) < 1440)$ **or** $(a = 1$ **and** $\text{TIMELAG}(a, \beta, b) \geq 1440))$ **and** $(b > e_{tol})$
  do
    $b := b/2$;
  enddo
  if $b \leq e_{tol}$ **then**
    THRESHOLD$:= b$;
    return;
  endif

Find the upper bound on the threshold.

while $((a = 0$ **and** $\text{TIMELAG}(a, \beta, c) \geq 1440)$ **or** $(a = 1$ **and** $\text{TIMELAG}(a, \beta, c) < 1440))$ **and** $(c < 1/e_{tol})$
  do
    $c := c * 2$;
  enddo
  if $c \geq 1/e_{tol}$ **then**
    THRESHOLD$:= c$;
    return;
  endif

Begin bisecting the interval.

while $(c - b)/b > e_{tol}$ **do**
  $next = (c + b)/2$; Bisect the interval.
  if $(a = 0$ **and** $\text{TIMELAG}(a, \beta, next) \geq 1440)$ **or** $(a = 1$ **and** $\text{TIMELAG}(a, \beta, next) < 1440)$ **then**
    $b := next$;
  else
    $c := next$;
  endif
enddo
$THRESHOLD := b$;
end subroutine THRESHOLD

Subroutine FEX solves for the change in MPF concentration given MPF concentration, time, values for the parameters, and total cyclin concentration. Note that time does not appear directly in the ODE, but M is dependent on time. FEX is used by LSODAR when computing M numerically. FEX takes the MPF concentration M and returns the numeric derivative M_t of MPF concentration with respect to time at the point M. JEX computes the partial derivative P of the ODE with respect to the dependent variable M, and takes the same arguments as FEX. In this problem the partial derivatives are calculated numerically; therefore, JEX is empty. LSODAR returns a root for the function g evaluated in GEX. GEX takes the same arguments as FEX. Pseudocode for FEX, JEX, and GEX follow.

subroutine FEX
  $D := G(M, \beta_3, \beta_7, \beta_8)$; The function G here is defined in the problem statement. No pseudocode for G is presented in this paper.
  $W := G(\beta_3, M, \beta_9, \beta_{10})$; Note that β_3 is substituted for β_6 to satisfy the constraint $\beta_3 = \beta_6$. Earlier in the code β_6 was fixed so ODRPACK will only vary β_3.
  $K_D := \beta_1 * (1 - D) + \beta_2 * D$;
  $K_W := \beta_4 * (1 - W) + \beta_5 * W$;
  $M_t := K_D * (C - M) - K_W * M$;
end subroutine FEX

subroutine JEX
end subroutine JEX

subroutine GEX
  $g_1 := M - M_{root}$; M_{root} is set elsewhere to a desired value of the solution M to the ODE defined in FEX.
end subroutine GEX

5. RESULTS

The parameter estimates can be found in Table 2 (along with the initial guess). The experimental data along with the fitted curves are in Figures 2 and 3. The ratio of the inactivation threshold to the activation threshold is 2.9993. The total weighted sum of squares is $1.54 \cdot 10^{-2}$. A few thousand points were checked in the neighborhood of the parameter estimate in Table 2 to ensure that the sum of squares is really a minimum. All of the points yielded a sum of squares greater than $1.54 \cdot 10^{-2}$.

Table 2. Initial and final estimated parameter values for the ten parameter model.

Rate Constant	Initial	Final
v'_d	0.017	0.00123
v''_d	0.17	0.0475
v'''_d, v'''_w	0.05	0.0363
v'_w	0.01	0.000502
v''_w	1.0	0.305
K_{md}	0.1	0.1
K_{mdr}	1.0	1.0
K_{mw}	1.0	1.0
K_{mwr}	0.1	0.1

A sensitivity analysis was done against each of the parameters. Table 3 contains the partial derivative of the weighted sum of squares with respect to each parameter. These derivatives are all small with the exception of the derivative with respect to v'_d. The derivatives were calculated using a central difference formula with a step size of 10^{-4} relative to the parameter. The larger partial with respect to v'_d suggests that the local minimum may be different from the estimate in Table 2. The other partials are orders of magnitude smaller and seem reasonable given the discretization error. A plot of the weighted sum of squares versus v'_d was generated to explore the parameter space further. The plot shows a local minimum with respect to v'_d with a difference of 10^{-3} relative to the ODRPACK estimate. The sum of squares at this point differs by 10^{-5} relative to the sum of squares from the ODRPACK estimate. The error in the parameter estimate is well below the error in experimental data and therefore is acceptable. The last four parameters have small partials even though they were not estimated by ODRPACK suggesting the experimental data does not constrain those parameters.

Table 3. The partial derivatives of the weighted sum of squares with respect to each parameter at the estimated point in Table 2.

Rate Constant	$\partial E / \partial \beta_i$
v'_d	0.1393
v''_d	0.0051
v'''_d, v'''_w	-0.0089
v'_w	0.0080
v''_w	-0.0005
K_{md}	-0.0034
K_{mdr}	0.0013
K_{mw}	-0.0007
K_{mwr}	0.0022

Table 4 contains partial derivatives of the weighted orthogonal distance squared with respect to each parameter. For every combination of one data point with one parameter there is a partial derivative of the weighted orthogonal distance of the error with respect to the parameter. Table 4 only reports the partial derivatives for one data point for each parameter. The maximum partial for each parameter is reported. j represents the data point that yielded the largest partial. Indexing for j starts at 1. Refer to the pseudocode in Section 4 for the order of the data.

Table 4. The maximum partial derivatives of the weighted orthogonal distances with respect to each parameter at the estimated point in Table 2.

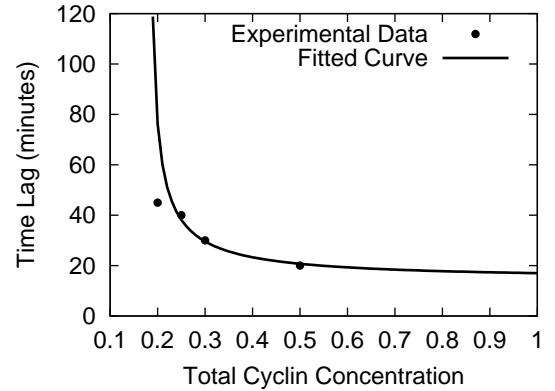| Rate Constant	j	$\max_j \frac{\partial (w_{\epsilon_j} * \epsilon_j^2 + w_{\delta_j} * \delta_j^2)}{\partial \beta_i}$
v'_d	1	32.84
v''_d	1	138.5
v'''_d, v'''_w	1	155.7
v'_w	4	0.520
v''_w	1	111.2
K_{md}	1	15.24
K_{mdr}	1	114.9
K_{mw}	1	18.55
K_{mwr}	1	3.100

Figure 2. Time lag for MPF activation versus total cyclin (as calculated using the parameters estimated for the ten parameter model).

6. CONCLUSION

The model uses Goldbeter-Koshland functions to represent the feedback loops between MPF, Wee1, and Cdc25. In this sense the model reasonably models the underlying biochemistry. However, the model is not complete. The model still must keep track of cyclin monomers, Cdc2 monomers, and more phosphorylation states of cyclin/Cdc2 dimers (MPF) to meet the goals set forth in this paper.

The parameter estimation was successful for the model presented in this paper. The biological data can have as much as 10 percent error. The fitted model has less error than the data. Any improvements of the model to describe the real biochemistry would first require more accurate data (or more data providing different constraints on the model).

The code for the parameter estimate presented in this paper took approximately 30 minutes to perform.
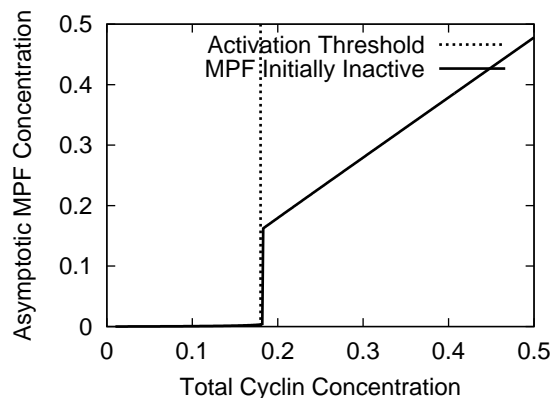
Figure 3. Asymptotic MPF concentration versus total cyclin concentration for the ten parameter model. The experimentally measured threshold is also present as a vertical line.

A previous paper was published using similar code and a simpler model [15]. The parameter estimate from the previous paper took approximately 2 minutes on the same computer. Another model is already under development which will demand another large increase in run time.

The model will be expanded in future work to meet the goals set in this paper. More data will be added to provide more constraints on the model. The parameter estimation will be parallelized to increase performance.

ACKNOWLEDGMENTS

REFERENCES

[1] Boggs, P.T.; R.H. Byrd; J.R. Donaldson; R.B. Schnabel. 1989. "Algorithm 676 — ODRPACK: Software for Weighted Orthogonal Distance Regression." *ACM Trans. Math. Software 15*, No. 4:348–364.

[2] Boggs, P.T.; R.H. Byrd; J.E. Rogers; R.B. Schnabel. 1992. *User's Reference Guide for ODRPACK Version 2.01: Software for Weighted Orthogonal Distance Regression*. Center for Computing and Applied Mathematics, U.S. Department of Commerce, Gaithersburg, MD.

[3] Dekker, T.J. 1969. "Finding a Zero by Means of Successive Linear Interpolation." In *Constructive Aspects of the Fundamental Theorem of Algebra* (Dejon, B., and Henrici, P., eds.). Wiley-Interscience, London.

[4] Gear, C.W. 1971. *Numerical Initial Value Problems in Ordinary Differential Equations*. Prentice-Hall, Englewood Cliffs, NJ.

[5] Goldbeter, A.; D. E. Koshland. 1981. "An Amplified Sensitivity Arising from Covalent Modification in Biological Systems." *Proc. Natl. Acad. Sci. USA 78*, No. 11:6840–6844.

[6] Hindmarsh, A.C. 1980. "LSODE and LSODI, Two New Initial Value Ordinary Differential Equation Solvers." *ACM SIGNUM Newsletter 15*, No. 4:10–11.

[7] Hindmarsh, A.C. 1983. "ODEPACK: A Systematized Collection of ODE Solvers." In *Scientific Computing* (R.S. Stepleman, et al., eds.). North Holland Publishing Co., New York, 55–64.

[8] Marlovits, G.; C.J. Tyson; B. Novak; J.J. Tyson. 1998. "Modeling M-phase control in Xenopus oocyte extracts: the surveillance mechanism for unreplicated DNA." *Biophys. Chem.*, 72, 169–184.

[9] Moore, J. 1997. Private Communication, Aug.

[10] Petzold, L. 1983. "Automatic Selection of Methods for Solving Stiff and Nonstiff Systems of Ordinary Differential Equations." *SIAM Journal on Scientific and Statistical Computing*, 4, 136–148.

[11] Radhakrishnan, K.; A.C. Hindmarsh. 1993. *Description and Use of LSODE, the Livermore Solver for Ordinary Differential Equations*. NASA Reference Publication 1327. Lawrence Livermore National Laboratory, Livermore, CA, Dec.

[12] Shampine, L.F.; R.C. Allen. 1973. *Numerical Computing: An Introduction*. W. B. Saunders Company, Philadelphia, PA.

[13] Shampine, L.F.; M.K. Gordon. 1975. *Computer Solution of Ordinary Differential Equations, The Initial Value Problem*. W. H. Freeman, San Francisco, CA.

[14] Solomon, M. J., et al. 1990. "Cyclin Activation of p34cdc2." *Cell*, 63, 1013–1024.

[15] Zwolak, J.W.; J.J. Tyson; L.T. Watson. 2001. "Estimating Rate Constants in Cell Cycle Models." In *Proc. High Performance Computing Symposium* (A. Tentner ed.). Soc. for Modeling and Simulation Internat., San Diego, CA, 53–57.