MATHEMATICAL BIOSCIENCES
AND ENGINEERING
Volume **00**, Number **0**, Xxxx **XXXX**

http://www.mbejournal.org/

pp. **1–??**

A MULTI-STAGE MODEL FOR QUANTITATIVE PCR

Emily Stone

Department of Mathematical Sciences, The University of Montana
Missoula, MT 59812

John Goldes

Department of Mathematical Sciences, The University of Montana
Missoula, MT 59812

Martha Garlick

Department of Mathematics and Statistics, Utah State University
Logan, UT 84322

Abstract. PCR (Polymerase Chain Reaction), a method which replicates a selected sequence of DNA, has revolutionized the study of genomic material, but mathematical study of the process has been limited to simple deterministic models or descriptions relying on stochastic processes. In this paper we develop a deterministic model for the reactions of quantitative PCR (Polymerase Chain Reaction) based on the law of mass action. Maps are created from DNA copy number in one cycle to the next, with ordinary differential equations describing the evolution of different molecular species during each cycle. The advantage of this type of model is the ability to vary the time spent in each stage of the reaction, which is critical to predicting optimal protocols. Qualitative analysis of the models are performed and parameters are estimated by fitting each model to data from Roche LightCycler (TM) runs.

1. **Introduction.** The Polymerase Chain Reaction (PCR), a technique for the enzymatic amplification of specific target segments of DNA, has revolutionized molecular biological approaches involving genomic material. This, in turn, has impacted research in human genetics, disease diagnosis, cancer detection, evolutionary and developmental biology, and pathogen detection, to name a few. The company Idaho Technology, Inc. has capitalized on the invention of fluorescent probe techniques to create fast, accurate devices for quantitative PCR. Quantitative PCR is a method where the amount of amplified DNA (or amplicon) is tracked throughout the reaction and the initial amount of sample DNA can then be estimated. Understanding the important parts of a complex reaction that is repeated tens of times, is critical in improving the design of these processes in the laboratory, and to date theoretical studies of quantitative PCR are limited. In this paper we present a suite of deterministic models for quantitative PCR, with parameters estimated from data provided from Roche LightCycler (TM) (a device developed by Idaho Technology) PCR runs. Determining the critical features of the model through construction of increasingly complex descriptions of the reaction, with the ability to vary the time spent in each stage of the reaction, is the overall goal of the project. These models

2000 *Mathematics Subject Classification.* 92C45, 92C50, 92D20.
Key words and phrases. polymerase chain reaction, mathematical model.

can then be used to develop optimum PCR protocols (heating and cooling regimes), the subject of ongoing research.

In PCR a reaction mixture containing a few copies of the target double-stranded DNA is first heated to separate the DNA into single strands. It is then rapidly cooled and held at a lower temperature briefly so that PCR primers (short single strands of DNA that have been designed for this purpose) anneal specifically to the template DNA. The enzyme Taq Polymerase recognizes these primer-template pairs and synthesizes a new strand of DNA, starting at the end of the annealed primer. In this way, a complementary strand is made from each strand of the original double-stranded DNA molecule. Under ideal reaction conditions the number of copies of this stretch of DNA in the sample is doubled in each heating-cooling cycle.

Instruments that perform real-time PCR usually detect the amplified DNA using fluorescent probes that are added to the PCR reagents before temperature cycling. These probes bind to the DNA and generally fluoresce more when bound than when free. When there is a sufficient quantity of DNA present in the sample (for example, after many temperature cycles), this change in fluorescence is detected using a fluorimeter. If the fluorescent signal of a sample rises above a background level, a sizable amount of DNA has been synthesized, indicating that the specific DNA was initially present.

Current methods for DNA quantification (for more information see the following references: Morrison et al. [5], Wittwer et al. [17, 18, 19], Weiss and Von Haeseler [14, 15], Sun [10], Sun et al. [11]) with PCR are based on comparing a set of successively diluted standards against unknown samples. The methods utilize the concentrations of the standards in a dilution series to determine the concentration of the unknown. The amount of DNA in successively diluted standards is typically decreased by factors of 2 or 10, and anywhere from three to ten standards are used. Figure 1 shows a set of six standards containing between one and 1,000,000 copies of initial DNA template. The fluorescence curve that crosses the threshold value at the smallest cycle number initially had 1,000,000 copies of DNA, the next curve to cross the threshold had 100,000 copies of DNA initially, and so on. Notice that the curve that does not cross the threshold is the control-no-template sample.

These amplification curves suggest that a more natural model for the PCR reaction would be logistic, which proceeds to saturation as a resource is depleted. Some device software fits the data to such a logistic map, and uses the result to estimate initial copy number of the template. For the purpose of this estimation both the exponential growth model and the logistic model are sufficient in many cases, and have the advantage of a limited number of free parameters, requiring a minimum amount parameter estimation. However, for the long range goal of developing a more complete model of the reaction that can lead to innovation in process design, we must look beyond these one dimensional approximations. We will also show that the data deviate from the logistic model in a consistent way for all the amplification curves, suggesting that the simplifications leading to it eliminate some critical features of the dynamics.

To our knowledge no deterministic model of the reactions of PCR that does not include assumptions about the kind of enzyme kinetics involved (i.e. Michaelis-Menten) are present in the literature. Stochastic models for estimating reaction efficiency and specificity can be found however, for instance, in Sun et al. [10, 11] a model for distributions of mutations and estimation of mutation rates during PCR is developed, using the theory of branching processes. Another such model

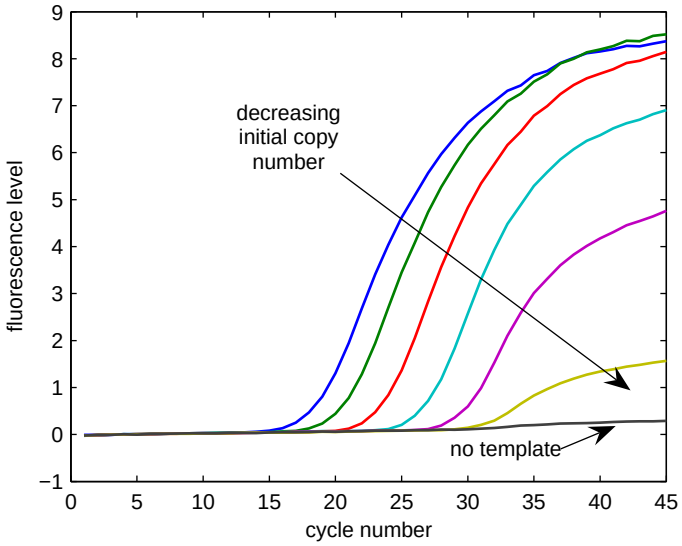

FIGURE 1. Fluorescence level vs. cycle number during PCR Roche Lightcycler run. Different lines are standard dilutions for quantification purposes, from 10^6 copies down to 10^1 and no template as a control.

is reported in Weiss and Von Haeseler [14, 15], where the accumulation of new molecules during PCR is treated as a randomly bifurcating tree to estimate overall error rates for the reaction. In Schnell and Mendoza [7], the reaction efficiency of quantitative competitive PCR (QC-PCR: a target and a competitor template are amplified simultaneously to provide an internal standard for identifying the initial target template amount) is computed using Michaelis-Menten type kinetics. Following this the authors produced a companion paper [8], which incorporates similar kinetic considerations of reduced enzyme efficiency into a simple continuous of model of PCR, resulting in accurate predictions of the synthesized products.

For more background, see tech report.

In the first part of our paper we analyze the utility of simple deterministic maps to capture the essential behavior of PCR. Maps allow for simple fitting and an appropriate number of parameters given the form of the data used in parameter estimation. We use data from a Roche LightCycler, provided by Idaho Technology to fit these parameters and estimate the initial amount of template in the reaction.

While predicting yield and estimating initial template amounts are important in the analysis of PCR, we focus in the second half of the paper on constructing a model that can be used in process optimization. By this we mean the ability to vary the PCR protocol (time spent in different stages of the reaction, specifically), so that a reaction can be tuned to give the best yield or the fastest results. To this end we develop continuous time models, ODEs, which we analyze and perform parameter estimation.

An outline of the paper is as follows.

Table I: List of Variables and Notation Used in PCR Models

Variable name	quantity
C	copy number
E	exponential efficiency of reaction
S, s	single stranded DNA (ssDNA), $s = [S]$
P, p	primer molecule, $p = [P]$
S', s'	primed ssDNA, $s' = [S']$
Q, q	Taq molecule, $q = [Q]$
C, c	enzyme complex, $c = [C]$
N, n	nucleotide sequence for the extension, $n = [N]$
D, d	double stranded DNA (dsDNA), $d = [D]$
k_{-1}, k_1	forward and backward reaction rates for annealing
k_{-2}, k_2	forward and backward reaction rates for complex formation
k_{-3}, k_3	forward and backward reaction rates for extension
ϵ	logistic differential equation growth parameter
γ	logistic map growth parameter
$\mathcal{K}$	carrying capacity of the logistic map and ODE
$\Gamma(d_i)$	growth parameter function for Taq model
e, α	parameters in $\Gamma(d_i)$
W_i	estimation of $\Gamma(d_i)$ from experimental data
Y_i	logarithmic regression variable
t_1	time in stage I of two stage model without Taq dynamics
t_2	time in stage II of two stage model without Taq dynamics
τ_I	re-scaled time in stage I of two stage model without Taq dynamics
τ_{II}	re-scaled time in stage II of two stage model without Taq dynamics
K, K_n, K_d, K_s	conserved quantities in the two stage models
KK	normalization parameter used for experimental data
β, γ	re-scaled reaction rates in the two stage model with Taq dynamics
s'^I, s'^{II}	s' in stage I and stage II respectively
$\bar{x}$	fixed point of the x variable
t_I, t_{II}	time in stage I and stage II in model with Taq dynamics

2. **Fitting Deterministic Maps to PCR Data.** In the introduction we mention a method for determining initial copy number by essentially bracketing the unknown runs with runs of known concentration. Then an exponential growth model can be used to determine the initial amount using only the initial part of the reaction when exponential growth dominates.

Another method fits data to a logistic map, as the PCR data do level-off in a sigmoidal fashion. This is physical since the reaction necessarily saturates, it will run out of either primers or nucleotides as more copies are synthesized. Primers are synthesized molecules and are much more expensive than nucleotides, so in most cases running out of primers will be what stops the reaction. Does the logistic model give a good fit to the PCR data? To test this we fit a dilution series run from a Roche LightCycler (TM) to a logistic map:

$$d_{i+1} = d_i + \gamma d_i(\mathcal{K} - d_i), \tag{1}$$

where d_i is the fluorescence level at cycle i.

The run was typical for these quantification experiments: it had 45 cycles, each consisting of a brief melt stage at 95 degrees C, a 10 second annealing stage at 55 degrees C, and a 30 second extension stage at 72 degrees C. The fluorescence acquisition occurred at the end of the annealing stage, and used a FRET (fluorescence resonance energy transfer) probe system. FRET probes are a pair of oligonucleotide labeled with fluorescent dyes. The pair are designed to hybridize to adjacent regions on the target DNA, and the marker dyes of each probe can only interact when they are in close proximity and bound to the target. The fluorophores are chosen so that the emission spectrum of one overlaps with the excitation spectrum of the other. The donor fluorophore is excited by a light source, transfers its energy to an acceptor fluorophore, which then emits light of a longer wavelength. This light is then detected during the fluorescence acquisition.

The parameter estimation was done in MATLAB using least squares to first compute K and γ, and a simplex search method based scheme (MATLAB's *fminsearch*) to find the initial fluorescence level. The objective function used in the nonlinear optimization was the two-norm of the difference between the model time series and the data. Percentage error was computed by dividing the final value of the objective function by the two-norm of the model time series.

Table II: Parameter Estimation for the Logistic Model (1)

	γ	$\mathcal{K}$	d_0	% error
run 1	0.0185	8.1473	2.9×10^{-2}	5.15
run 2	0.0189	8.2832	1.6×10^{-2}	5.8
run 3	0.0214	7.9001	4.2×10^{-3}	5.15
run 4	0.0296	6.678	5.5×10^{-4}	4.9
run 5	0.0457	4.5959	1.02×10^{-5}	5.3
run 6	0.1273	1.5445	4.12×10^{-5}	5.82

The results of the parameter estimation for the dilution series are present in figure 2 and Table II. We see that the model is more than adequate for predicting initial copy number, given the current practice of running standards simultaneously with samples to generate a map between initial copy number and fluorescence level. The drift in the growth constant for decreasing copy number indicates that some aspect of the dynamics is not captured by this map, and for the lowest copy number run (number 6) we see that the initial amount estimated is off by an order of magnitude. In this case competition from other reactions is thought to be the

culprit, but in all cases the map is a gross simplification to the actual dynamics. It clearly overestimates the growth for earlier cycles, and approaches saturation more quickly at later cycles. To verify this intuition quantitatively we preformed a version of logarithmic regression on the first five runs of the dilutions series (see figure 1). The basis for this regression is the separation of variables solution to the logistic differential equation:

$$\frac{dy}{dt} = \epsilon y(\mathcal{K} - y),$$

namely

$$\ln(\frac{y}{K-y}) = \epsilon \mathcal{K} t + \ln(\frac{y_0}{\mathcal{K}-y_0}).$$

If y follows logarithmic growth the variable $Y = \ln(\frac{y}{\mathcal{K}-y})$ will depend linearly on time t. Plots of the discretely sampled variable Y_i vs. cycle number i for standard data set are shown in figure 3, where it is obvious that they are not well-estimated by a linear function of i. There are several linear regions in these graphs, corresponding to a) low cycle number noise, b) a region where the initial exponential growth occurs, and c) a third region where saturation occurs. A linear fit of this data would have an intermediate slope causing an overestimate of data for low cycle number, an underestimate of the data for intermediate cycle number, and an overestimate of data for large cycle number. This is reflected in the quality of fit to the logistic data seen in figure 2. One explanation of this is a growth rate that depends on the number of molecules, e.g., the reaction becomes less efficient for larger numbers of molecules. This suggests a growth parameter $\Gamma(d)$ that varies with the amount of amplicon, d, so that

$$d_{i+1} = d_i + \Gamma(d_i) d_i(\mathcal{K} - d_i),$$

where $\Gamma(d)$ is a decreasing function of d. The shape of this function can be estimated by plotting the variable found by solving the above equation for Γ:

$$W_i = \frac{d_{i+1} - d_i}{d_i(\mathcal{K} - d_i)} = \Gamma(d_i).$$

That is, the graph of W_i vs. d_i will give an idea of $\Gamma(d_i)$. This is plotted for the first four traces in the dilution series in figure 4. The function Γ appears to be inversely proportional to d_i, so we fit a new map with $\Gamma(d_i) = \frac{e}{1+\alpha d_i}$, where e and α are positive constants, so

$$d_{i+1} = d_i + \frac{e d_i}{1 + \alpha d_i}(K - d_i). \tag{2}$$

The results are shown in figure 5, and in Table III. Along with the dilution series data we also fit two samples with originally unknown initial concentrations of template, each having five replicates. The error presented in Table III is the mean square error.

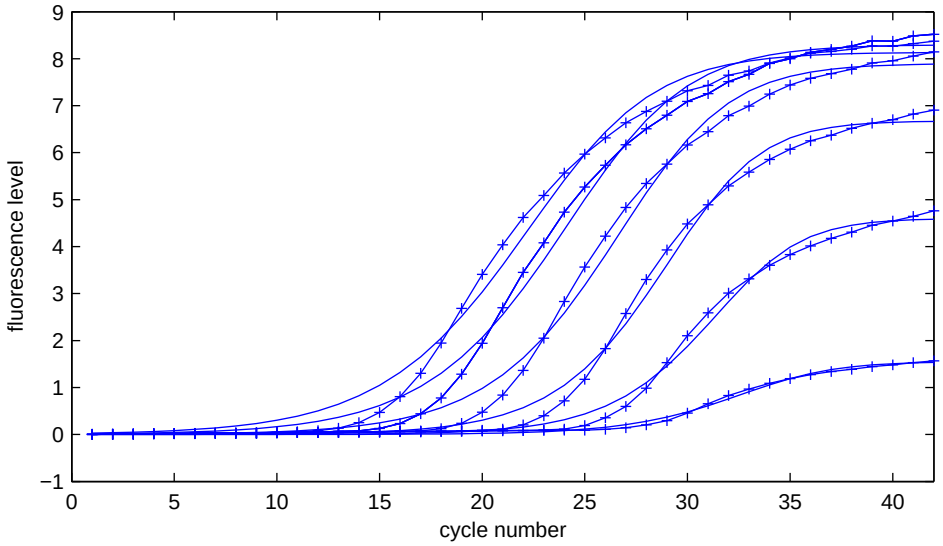

FIGURE 2. Fitting dilution series data with a logistic map. See text for parameter values. Solid line: model, (+++++): data.

Table III: Parameter Estimation of Taq Model (2).

	error	e	$\mathcal{K}$	α	d_0
run 1	0.1894	0.1481	8.5579	0.9148	4.91×10^{-7}
run 2	0.2055	0.1473	8.7559	0.8907	7.31×10^{-8}
run 3	0.2587	0.1227	8.3295	0.6096	1.20×10^{-7}
run 4	0.3639	0.1237	6.9411	0.4158	1.06×10^{-7}
run 5	0.3945	0.1592	4.7559	0.4111	7.53×10^{-8}
run 6	0.2394	0.3721	1.5740	0.6636	3.02×10^{-7}
sample 1a	0.2321	0.1413	9.3277	0.5918	3.98×10^{-8}
sample 1b	0.2377	0.1546	9.5827	0.7381	1.04×10^{-8}
sample 1c	0.2189	0.1231	9.7042	0.5957	1.57×10^{-7}
sample 1d	0.2582	0.1202	9.9448	0.6035	1.60×10^{-7}
sample 1e	0.2610	0.1196	10.0509	0.6391	1.62×10^{-7}
sample 2a	0.4427	0.1297	6.5511	0.3543	8.28×10^{-8}
sample 2b	0.4203	0.1222	6.8775	0.3390	9.03×10^{-8}
sample 2c	0.4328	0.1214	6.9499	0.3581	9.08×10^{-8}
sample 2d	0.4223	0.1173	7.1260	0.3541	1.07×10^{-8}
sample 2e	0.4270	0.1267	6.9205	0.3514	6.42×10^{-8}

The growth coefficient for this model (e), and the value for α are more consistent than for the regular logistic model, though the variation is more pronounced in smaller copy number runs, and suffers the same over-calculation of the initial fluorescence in run 6. (Also in run 6 the model coefficients are significantly different from the other runs). We also fitted the replicates of the unknown samples (sample 1 and sample 2) with reasonably consistent results, though it is clear there is a

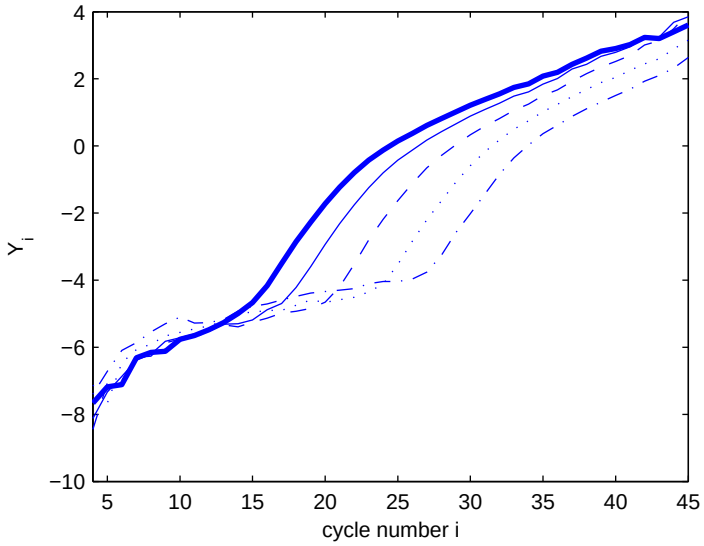

FIGURE 3. Logarithmic regression curves for five standard dilution series. The different line styles are used to indicate different dilutions, the largest initial concentration of template is denoted with the heaviest dark line

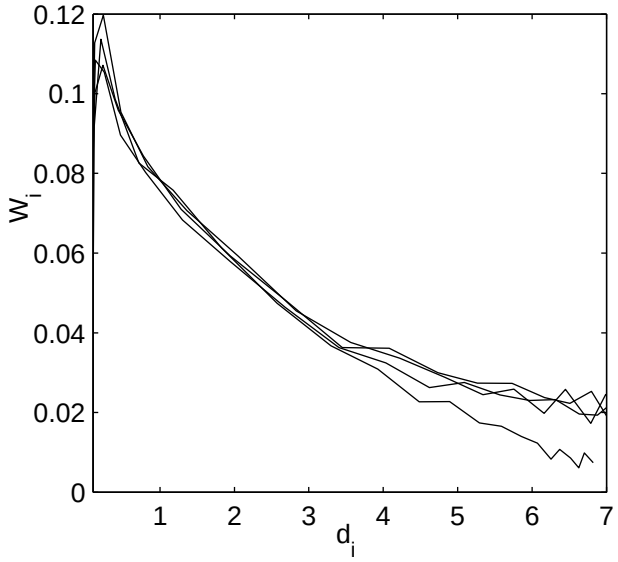

FIGURE 4. Plotting W_i vs. d_i to estimate $\Gamma(d_i)$; the first four dilution series runs.

trade-off between values of the growth constant e and the initial fluorescence, indicating hidden dependencies in the parameters that cannot be differentiated with this sort of data.

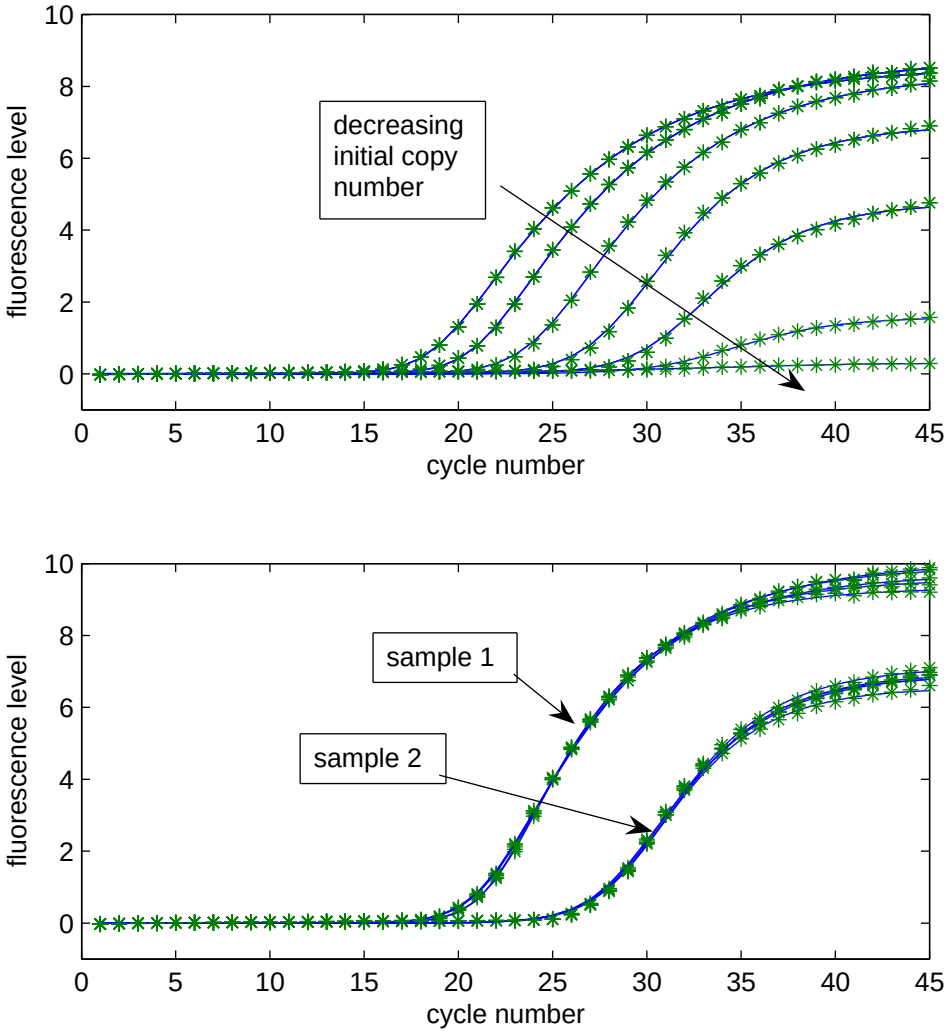

FIGURE 5. PCR data fitted with the Taq model (2), dilution series and unknown samples. See Table III for parameter and error information.

We should note here that this map is what is arrived at in the large Michaelis-Menten constant limit of the branching process model of PCR developed by Jagers and Klebaner [2]. If Γ is taken to be the efficiency discussed in Schnell and Mendoza [8], we see that it is of the same form as their derived expression. Our data thus supports both the models presented in [2] and [7].

This could be the end of the story, but an empirically determined rate function is not as satisfactory as a model that captures the behavior built-up directly from the reaction equations. Furthermore, we want a model where the time spent in each

phase of the reaction can be varied, allowing for incomplete priming and extension of the molecules in the case of shorter time. In the next section we construct such a model and parameterize it with the data.

3. **The Reactions of PCR.** PCR proceeds through repeated cycles of dissociation, annealing and extension by the enzyme Taq polymerase. During dissociation the sample is heated to approximately 90 degrees C where the template's DNA nucleotide base pairs unbind and the strand essentially unzips to form two half-strands (single stranded DNA). The sample is then cooled to a temperature where the primer reaction is optimal (about 60 degrees C), during which primer molecules, themselves sequences of single stranded DNA that have been designed to adhere to either end of the target sequence of the template, bind on. Then the sample is heated again to a temperature where Taq enzyme adds base pairs on the bracketed sequence to form a new double-stranded piece of DNA. The annealing/extension can done in one or two distinct steps, either with a continuous ramp-up to the Taq operating temperature (during which time the primers anneal) or with a lower temperature annealing stage followed by a higher temperature extension phase. We model the latter, but the model itself could easily be adapted for the one-step scenario.

These three phases, dissociation, annealing, and extension, are repeated typically 30-40 times to yield exponentially growing numbers of the target sequence, assuming the reaction runs as designed. Factors influencing the success of the reaction are competition from contaminants in the reaction mixture, primers that bind to themselves or other primer molecules (primer-dimers), or primers that can extend pieces of the template other than the target, to name a few. Naturally the reaction saturates, which is assumed to occur by complete depletion of primer molecules, since they are incorporated into the extended strands. The nucleotides in the mixture could also be used up, but typically they are present in great numbers to prevent this from occurring. Also, the initial amount of DNA to be amplified can not be either too large or too small. If it is too large the number of primers is not sufficient to completely prime the molecules, and if too small it can lose out to the competing amplification of undesired sequences.

The reaction equations for these phases can be written as follows.
Dissociation:

$$D \to 2S$$

Annealing:

$$S + P \underset{k_{-1}}{\overset{k_1}{\rightleftarrows}} S'$$

Extension:

$$S' + Q \underset{k_{-2}}{\overset{k_2}{\rightleftarrows}} C$$

$$C + N \underset{k_{-3}}{\overset{k_3}{\rightleftarrows}} D + Q$$

Here D is double-stranded DNA (or dsDNA), S is single-stranded DNA (ssDNA), P is primer, S', primed single-stranded DNA, Q, Taq polymerase, C, complex of primed single-stranded DNA and Taq, and N, nucleotides. The plus/minus k's represent the forward and backward reaction rate respectively. Ideally the reactions form a cascade, the product of one reaction continues into the next reaction and

the final double-stranded DNA cycles back to the dissociation phase. In reality the reactions occur simultaneously, with highest probability at their optimal temperature. For our purpose we will treat the phases as distinct and cascade the output of one phase to the input of the next. We also assume that the back reactions are negligible compared to the forward reactions in all but the creation of the enzyme complex, e.g. $k_{-1} = k_{-3} = 0$. This simplification emphasizes the importance of the dynamics of the Taq enzyme, and could be relaxed, leading to more complicated, but still tractable, models.

The most significant simplification in these models is the compression of a process that adds base pairs (BPs) one-at-a-time to a one that adds all the necessary BPs for the amplicon in one chunk. This assumption is equivalent to assume the process of adding BPs is very fast compared to annealing, and that it is irreversible. We made this simplification at the advice of Idaho Technology scientists [1]. A more complex model would add base pairs one at a time which would make possible an examination of the dependence of the reaction on template length and BP make-up. In this paper we will lay the ground work for such more complicated analysis (which should be done with stochastic processes for the extension).

4. **The Two Stage Model.** In our model of PCR we assume that the dissociation phase of the reaction (or "melt") is complete, and describe the phases of annealing primers and extending strands separately to create two distinct stages for each cycle. The law of mass action can then be invoked to create differential equations for the concentrations of the reactants in each stage. We use lower case letters to indicate the concentrations, e.g. $[S] = s$, $[D] = d$, etc.

For the annealing reaction we have:

$$\frac{ds}{dt} = -k_1 sp \tag{3}$$

$$\frac{dp}{dt} = -k_1 sp \tag{4}$$

$$\frac{ds'}{dt} = +k_1 sp. \tag{5}$$

And for the extension phase the equations are:

$$\frac{ds'}{dt} = -k_2 s'q + k_{-2} c \tag{6}$$

$$\frac{dq}{dt} = -k_2 s'q + k_{-2} c + k_3 cn \tag{7}$$

$$\frac{dc}{dt} = k_2 s'q - k_{-2} c - k_3 cn \tag{8}$$

$$\frac{dn}{dt} = -k_3 cn \tag{9}$$

$$\frac{dd}{dt} = k_3 cn. \tag{10}$$

This model makes the assumption of complete dissociation, so that the amount of ssDNA entering the annealing phase is equal to twice the amount of dsDNA from the previous extension phase, plus whatever ssDNA was leftover from the previous annealing phase. This eliminates the need for the equation that describes dissociation (or the "melt") phase of the reaction. We next assume that the annealing phase happens distinct from the extension phase and call this stage I. The equations

for stage I are (3)-(5). The extension phase we name stage II, and the equations are (6)-(10).

The initial amount of primed ssDNA in stage II is equal to the amount created in stage I, while the initial amount of nucleotide in stage II is whatever was left over from the previous cycle of stage II. Upon completion of stage II any unextended complex will break-up during dissociation, as will any primed ssDNA. Thus the initial amount of complex in stage II will be zero, and the amount of Taq enzyme will be the original amount from the beginning of the PCR reaction (Q). Stage I starts with no primed ssDNA, it is assumed to dissociate during the melt phase. The primer initial condition is the amount of unused primer from the previous cycle, plus the amount created during the dissociation of the complex and primed ssDNA during the melt phase. The initial amount of ssDNA will be that left from the previous annealing phase plus an amount equal to amount of complex left in stage II that dissociates, plus the ssDNA that results from the dissociation of the dsDNA created in the previous stage II, which is double the amount of dsDNA. In terms of equations these initial conditions can be stated:

Stage II

$$s'(0) = s'(t_{end},\ \text{stage I});\quad q(0) = Q;\quad c(0) = 0.0;\quad n(0) = n(t_{end}, \text{previous stage II});\quad d(0) = 0.0. \tag{11}$$

Stage I

$$s(0) = s(t_{end}, \text{prev. stage I}) + c(t_{end}, \text{prev. stage II}) + 2d(t_{end}, \text{prev. stage II}) + s'(t_{end}, \text{prev. stage II}); \tag{12}$$

$$p(0) = c(t_{end}, \text{previous stage II}) + s'(t_{end}, \text{previous stage II}) + p(t_{end}, \text{previous stage I});$$

$$s'(0) = 0.0.$$

A map for the reaction is created by integrating the ODEs in each stage and using the initial condition rules to link one stage to the other. However, insight can be gained by analyzing the dynamics of each stage separately and forming some special limiting cases for this map. See tech report for more details.

4.1. **Parameterizing the Full Two Stage Model with PCR Data.** We now investigate the parameterizations of the model with arbitrary time in stage I and stage II. The solution in the annealing phase (stage I) is the same as presented in equations (**??**),(**??**), and the linking initial conditions are (12). The stage II ODEs are given in equations (**??**)-(**??**) with re-scaled parameters and initial conditions:

$$s'(0) = \frac{s'(t_{end}, \text{stage I})}{N_0};\quad q(0) = \frac{Q}{N_0};\quad c(0) = 0.0;\quad n(0) = 1.0,\quad d(0) = 0.0.$$

Integrating the complete model occurs in phases, first computing the solution of stage I, eq. (**??**), then the value of the stage I variables at $t_{\text{end, stage I}}$ are used as initial conditions (eq. (11)) for the stage II integration. The following dissociation phase breaks up existing complex into ssDNA, primer and Taq, and dsDNA into twice as many ssDNA strands, and these are used as initial conditions for the next stage I (see eq.(12)). Examples of runs with varying amounts of initial template are shown in figure 6.

The parameter estimation was done using the same quantitative PCR data set in the previous sections, again using the Matlab function *fminsearch* to minimize the mean square error between the amplification data and the simulated time series. In performing the parameterizations we used the value determined for the initial s from fitting the two stage model without Taq dynamics. We fit the normalization

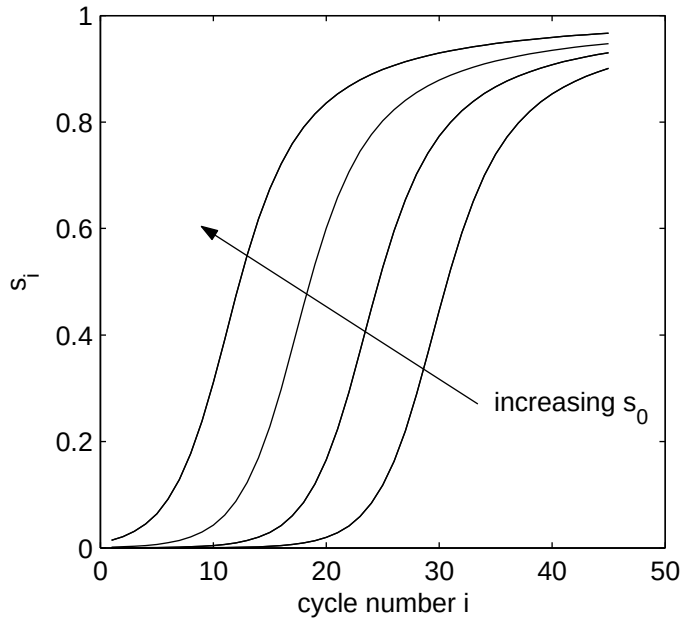

FIGURE 6. Integration of the full ODE model, with varying initial template amount, $s_0(0)$.

constant for the data, KK, (essentially a linear scaling between concentration of strands and the fluorescence level) the two reaction coefficients, β and γ, and the reaction times t_I and t_{II}. That leaves the initial amount of primer, $p_1(0)$, and of Taq, Q, relative to the initial amount of nucleotides. From the results of many parameterization runs we determined that the best fit was obtained when $p_1(0) = 1.0$, which is not what is indicated by Idaho Technology protocol, where a standard reaction set-up has 0.5 micromole of each primer and 0.8 millimole of dNTPs, the base pairs (BP) used in extension. Given an amplicon of 200 BPs this means about 4 micromole of completed segments, or 2 micromole of each complementary segment. The ratio of primers to nucleotides is about four-to-one, so we should set the initial primer amount to 0.25. This never achieved the same goodness-of-fit that the runs with higher initial amounts of primer did. On the other hand, the parameterization was relatively insensitive to the initial amount of Taq polymerase, so for simplicity we set it to 1.0. (The initial amount of Taq in a given reaction mixture is typically proprietary and difficult to obtain.) See Tables Va,b for the parameterization results.

In figure 7 a) we show a comparison of data to model with parameters found by the algorithm, for the dilution series. The logarithmic regression variable, $Y_n = log(\frac{y_n}{KK-y_n})$ was plotted vs. cycle number n for the model and the data in figure 7 b).

Table Va: Parameter Estimation: Two Stage Model with Taq Dynamics-Dilution Series

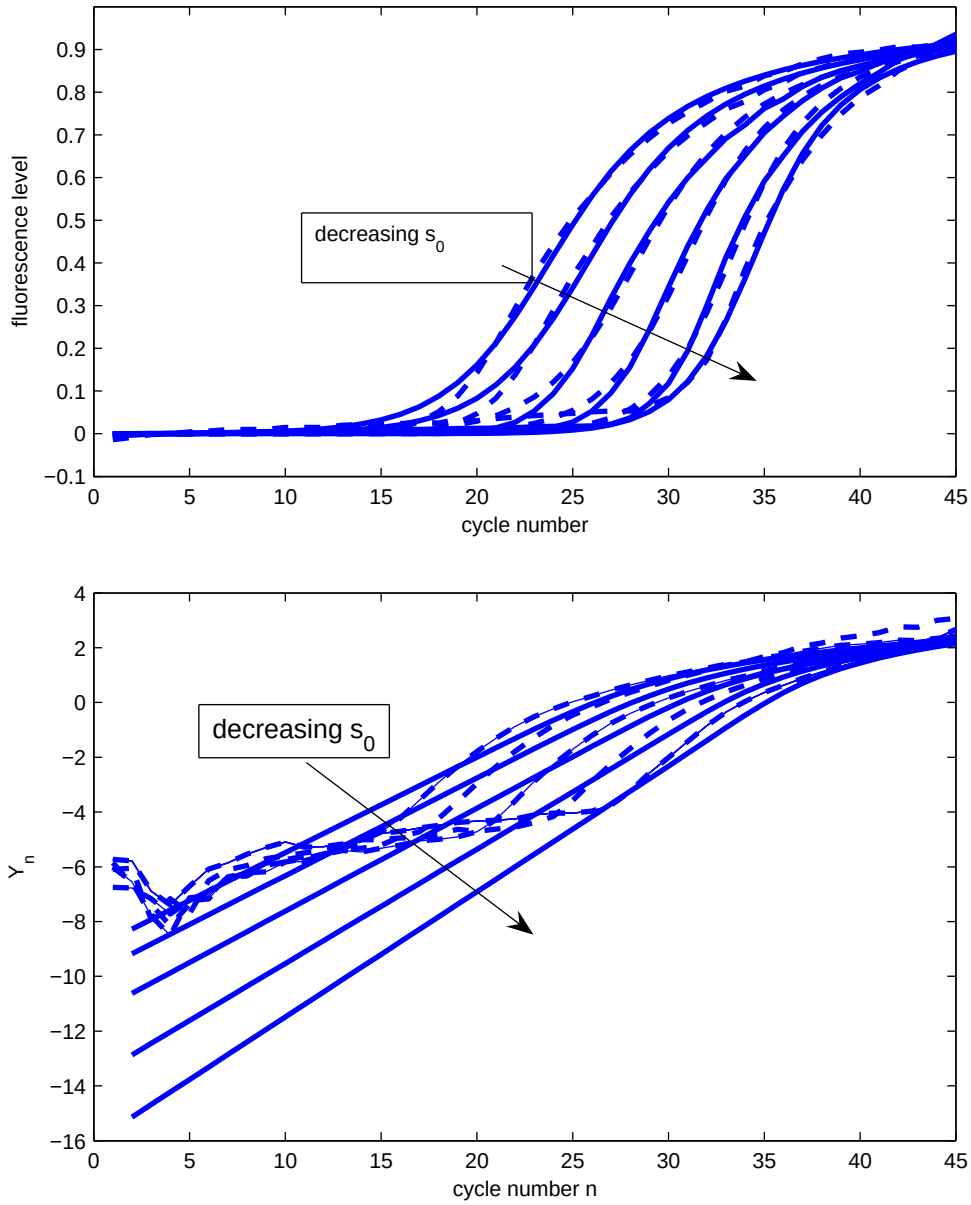

FIGURE 7. Comparison of two stage model with Taq dynamics to data. a) amplification curves, b) logarithmic regression curves. The data is represented with dashed lines in a) and with a dashed line overlaid by a thin line in b). See the text for parameter values.

	KK	$s_1(0)$	β	γ	t_I	t_{II}	% error
run 1	9.126	1.8e-04	4.0267	0.8650	2.2137	1.1521	2.57
run 2	9.3252	7.23e-05	5.4532	1.0404	2.2440	0.9242	2.62
run 3	8.9122	1.67e-05	4.8744	0.9028	2.2876	1.1484	2.66
run 4	7.5077	1.70e-06	4.6310	0.9777	2.5459	1.2457	2.43
run 5	5.0893	1.68e-07	4.3792	0.9423	2.0108	1.6888	3.05
run 6	1.6993	8.00e-08	5.9758	0.6248	1.9002	2.3773	5.52

We then performed the parameter estimation with data from the replicates with unknown initial concentrations, the results are presented in Table Vb.

Table Vb: Parameter Estimation: Two Stage Model with Taq Dynamics-Unknown Samples

	KK	$s_1(0)$	β	γ	t_I	t_{II}	% error
sample 1a	10.13	8.02e-05	0.8133	1.5316	2.5651	1.8070	3.42
sample 1b	10.49	7.87e-05	1.7851	0.7927	3.3509	1.6211	3.32
sample 1c	10.42	7.67e-05	3.1262	2.2279	4.3212	0.6089	2.86
sample 1d	10.73	7.57e-05	4.5382	1.3121	4.2464	0.7309	2.86
sample 1e	10.82	7.42e-05	5.4529	1.3825	3.1713	0.6814	2.90
sample 2a	7.18	1.51e-06	7.6983	1.2537	4.7358	0.7902	2.76
sample 2b	7.54	1.40e-06	4.6407	0.7659	4.2093	1.3862	2.60
sample 2c	7.60	1.40e-06	5.4337	0.9113	3.6034	1.1658	2.67
sample 2d	7.78	1.39e-06	6.7849	0.8755	3.4071	1.1475	2.64
sample 2e	7.50	1.50e-06	4.7158	0.7127	3.6256	1.4914	2.73

Comparing the fitted parameters for the different replicates in the unknown sample runs points out an obvious flaw with this parameterization. There is clearly more than one parameter set with an equally good fit to the data, indicating hidden dependencies in the parameters that simple re-scaling cannot uncover. Determining these dependencies through alternative re-scalings and singular perturbation analysis, and through other parameter estimation techniques is the subject of ongoing research.

Turning to the problem of realistic initial primer concentration, we found that for values much less than 1.0 the model did not capture the non-logistic behavior of the data. In figure 8 we illustrate this by graphing the logarithmic regression variable for differing values of initial primer, with all other parameters fixed. The concave down portion for the larger cycle numbers occurs only for larger initial primer concentration. In performing further parameter estimation trials with lower initial primer concentration, we found that we could not overcome this behavior by varying other reaction parameters.

We conclude that adding variable integration times for the model with Taq dynamics, while leading to a more flexible model and to improved goodness-of-fit, does not completely explain the non-logistic appearance of the data.

5. **Discussion and Conclusions.** In this paper we have analyzed a sequence of models for the reactions of PCR. Exponential growth, the first order approximation of simple doubling of the DNA strands, is replaced by a logistic model which captures the sigmoidal nature of the amplification curves. Both of these models are in common use in quantitative PCR devices. We postulated a variation on the logistic model where the efficiency decreases with copy number, and were able to fit the data with good results. This, however, is less satisfactory than a model built directly from the reactions that captures the decrease in efficiency as cycle number/copy number increases, and allows for variation of the time spent in each phase.

A parameterization of the full two stage model with Taq dynamics fit the data in the early and late stages of the reaction better than any of the simpler ODE models, though the best fit still corresponded to taking a larger than expected initial concentration of primer molecules. We note here (again) that the concentration of

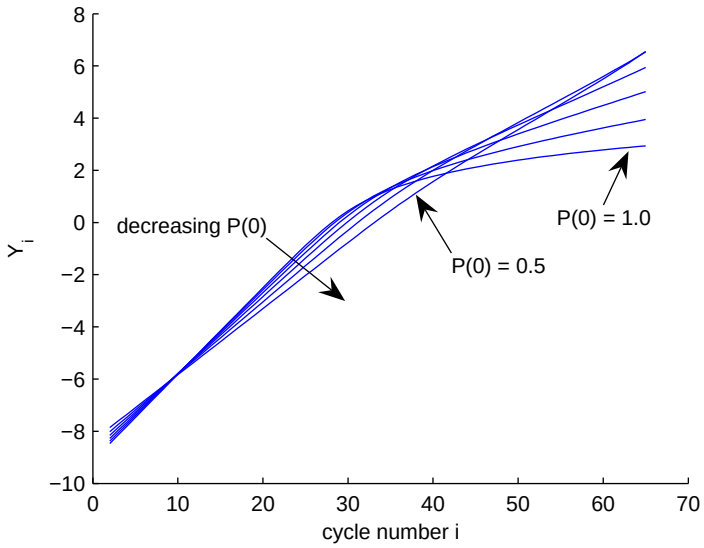

FIGURE 8. Integration of the two stage model with Taq Dynamics, with varying initial primer amount, $p_0(0)$.

Taq is typically proprietary in these devices, but since the model was relatively insensitive to the concentration of Taq, this is not the most important issue.

While we were not able to completely explain the variation from logistic behavior seen in the data with our suite of models, we were able to determine what portions of the model were important in capturing its non-logistic character. Competing reactions at higher cycle numbers most certainly will have an effect on the efficiency, especially with the lower initial copy number runs. Another possible explanation of the deviation from logistic at the end of the reaction is reduced efficiency of Taq after repeated thermal cycling, a fact that appears to be disputed among the scientists who perform PCR. We followed the information provided for us by Idaho Technology scientists, who believe that the 30 or 40 cycles in a typical PCR reaction would not degrade the enzyme significantly.

The main goal of the work, however, has been achieved, that is to build a model of PCR that allows the time in each reaction phase to be tuned. This model can now be used in studies of the efficiency of PCR protocols, which in turn can offer insight into the optimization of the reaction. We are currently engaged in work directed at using the two stage model to seek optimal reaction protocols, either minimizing the time to some level of amplification, or maximizing the yield for a fixed total run time.

Acknowledgements. We would like to thank Dr. David Eyre, Idaho Technology, for suggesting the problem and for providing us with Idaho Technology data and much consultation. This work was partially supported by a Utah State University Community/University Initiative grant, a USU Center for Integrated Biosystems seed grant, and one undergraduate (John Goldes) was supported by NSF-EPSCoR funds at the University of Montana-Missoula.

REFERENCES

[1] Eyre, D., private communication, 2002.
[2] Jagers, P. and F. Klebaner, Random variation and concentration effects in PCR. J. Theor. Bio. **224**, no. 3, (2003) 299–304.
[3] Lalam, N., Estimation of the reaction efficiency in polymerase chain reaction. J. Theor. Bio. (2006) *in press*.
[4] Moore, G.L. and C.D. Maranas, Modeling DNA mutation and recombination for directed evolution experiments. J. Theor. Bio. **205** (2000) 483-503.
[5] Morrison T., J. Weis and C. Wittwer, Quantification of low copy transcripts by continuous SYBR Green I monitoring during amplification. BioTechniques **24** (1998) 954–962.
[6] Pritchard, L., Corne, D. Kell, D., Rowland, J. and M. Winson, A general model of error-prone PCR. J. Theor. Bio. **234** (June 2005) 497–509.
[7] Schnell S., and C. Mendoza, Ensymological considerations for a theoretical description of the quantitative competitive polymerase chain reaction (QC-PCR). J. Theor. Biol. **184**(1997) 433–440.
[8] Schnell S., and C. Mendoza, Theoretical Description of the Polymerase Chain Reaction. J. Theor. Biol. **188** (1997) 313–318.
[9] Stolovitzky, G. and G. Cecchi, Efficiency of DNA replication in the polymerase chain reaction. Proc. Natl. Acad. Sci. USA. **93** (1996) 12947–12952.
[10] Sun, F., The polymerase chain reaction and branching processes. J. Comp. Bio. **2** no. 1 (1995) 63–86.
[11] Sun, F., D. Galas and M. Waterman, A mathematical analysis of in vitro molecular selection-amplification. J. Mol. Biol. **258** (1996) 650–660.
[12] Velikanov, M. and R. Kapral, Polymerase chain reaction: A markov process approach. J. theor. Biol. **201** (1999) 239–249.
[13] Viljoen, S., M. Griep, M. Nelson and H. Viljoen, A macroscopic kinetic model for DNA polymerase elongation and high-fidelity nucleotide selection. Comp. Bio. and Chem. **29** (2005) 101–110.
[14] Weiss, G and A. von Haeseler, Modeling the polymerase chain reaction. J. Comput. Biol. **2** (1995) 49–61.
[15] Weiss, G. and A. von Haeseler, A coalescent approach to the polymerase chain reaction. Nucleic Acids Res. **25** (1997) 3082–3087.
[16] Whitney, S., A. Sudhir, R. M. Nelson and H. Viljoen, Principles of rapid polymerase chain reactions: mathematical modeling and experimental verification. Comp. Bio. and Chem. **28** (2004) 195–209.
[17] Wittwer, C., G. Reed, and K. Ririe, Rapid cycle DNA amplification, in K Mullis, F. Ferre, and R Gibbs (Eds.), **The Polymerase Chain Reaction.** Springer-Verlag, Deerfield Beach, FL, (1994) 174–181.
[18] Wittwer C., M. Herrmann, A. Moss and R. Rasmussen, (Continuous fluorescence monitoring of rapid cycle DNA amplification. BioTechniques **22**(1997) 130–138.
[19] Wittwer C., K. Ririe and R. Rasmussen, Fluorescence monitoring of rapid cycle PCR for quantification. In: Ferre F, ed. **Gene Quantification.** Birkhauser, Boston MA (1998) 129–144.

E-mail address: stone@mso.umt.edu
E-mail address: jgoldes@mso.umt.edu
E-mail address: mgarlick@math.usu.edu