One Gene with Two Transcriptions: A Model of Two Populations

VIGRE Research Connections Program Report

Deborah Chun

## I. Background

In the most basic part of the brain, the leptin signaling pathway is a complex feedback loop which helps an organism determine when an intake of food is necessary or when the intake has been sufficient. In humans, this feedback loop can trigger sensations of hunger and satiation. Understanding this process can, in the future, help researchers understand and even chemically manipulate parts of this cycle, which aid our understanding of obesity. A paper by Qian Gao et. al. linked one particular gene transcription in this cycle to obesity, diabetes, infertility, and thermal dysregulation.

Around the world, researchers are studying this process in lab rats. Researchers measure gene and protein abundancies in rat populations over the course of two days, sampling every 4 hours. Using these data sets, a clearer picture of this process can be drawn. Andre Ptitsyn, a professor at Colorado State Univeristy, has been studying this process and began work with Jake Blanton, a Louisiana State University graduate student, on a mathematical model of this feedback loop in the summer of 2009.

## II. Introduction

This project is related to the work done June and July of 2009 by Jake Blanton. He has worked with Andre Ptitsyn to develop a model in MatLab of this feedback loop. At one stage of the leptin signaling pathway, phosphorylated STAT3 activates SOCS3 transcription in the cell nucleus. The SOC3 gene is translated by the cell's ribosome which builds a protein. Research data has revealed that two different transcriptions of the SOC3 gene are produced in this nucleus, a short transcription and a long transcription. When the global production of this gene appears constant, these two populations oscillate, behaving like two sinusoidal waves completely out of phase. We seek to model this behavior.

The main difference between these two gene populations is the decay rate. The long transcription is more often subject to predation by polymerase, an enzyme which breaks genes down into component nucleotides, which can then be reused to transcribe genes. The reason the same gene has two typical transcriptions, however, remains a mystery. While several labs around the world have produced data on these genes, every data set is very granular due to the technical difficulties of gathering this information.

## III. The Models

Because the causes of the variable populations are unknown, we develop multiple models. First, we will develop and refine a fixed life-span model. Then we will develop and refine a production and decay rate model. We will examine the models when the global transcription rate is a sinusoidal function, and we will examine the production and decay model when the global transcription rate is also a square function and test for behavior when a shock, an influx of one transcript of the gene SOC3, is introduced to the system.

**Model 1: Fixed life-span model.**

This model will use the following state variables and parameters.
The state variables:

$n_1(t)$ = amount of the short mRNA transcription of the gene (quantity)
$n_2(t)$ = amount of the long mRNA transcription of the gene (quantity)
$r(t)$ = global transcription rate of the gene (quantity per time)

The constant parameters:

$p$= probability that if a gene is transcribed, a short mRNA transcript results (%)
$L_1$ = average life span for the short transcription (time)
$L_2$ = average life span for the long transcription (time)

To find the amount of $n_i$ present, we will add the amount which was present to the amount produced and subtract the amount lost. We will consider how this happens in the time step, $\Delta t$, it takes to transcribe the SOC3 gene, although several copies of this gene can be simultaneously transcribed. In this model, the amount of $n_i$ produced is the global gene production multiplied by the probability that any gene transcribed will be type $n_i$. The amount of $n_i$ lost is the amount which was produced one lifespan ago. The formal equations follow.

$$n_1(t) = n_1(t-\Delta t) + r(t-\Delta t) * p * \Delta t - r(t-\Delta t-L_1) * p * \Delta t$$
$$n_2(t) = n_2(t-\Delta t) + r(t-\Delta t) * (1-p) * \Delta t - r(t-\Delta t-L_2) * (1-p) * \Delta t$$

These recursive models are more useful in the following form, found by taking the limit of these equations as the time step approaches 0.

$$\dot{n_1} = p * (r(t) - r(t-L_1))$$
$$\dot{n_2} = (1-p) * \big(r(t) - r(t-L_2)\big)$$

The solutions to these differential equations depend on the behavior of the global transcription rate. The simplest behavior rate can exhibit is simple on off switching. Experimental data suggest the behavior of this transcription rate oscillates over the course of a day, which we will represent as $2\pi$ radians. Figure 1 shows this simple switching.

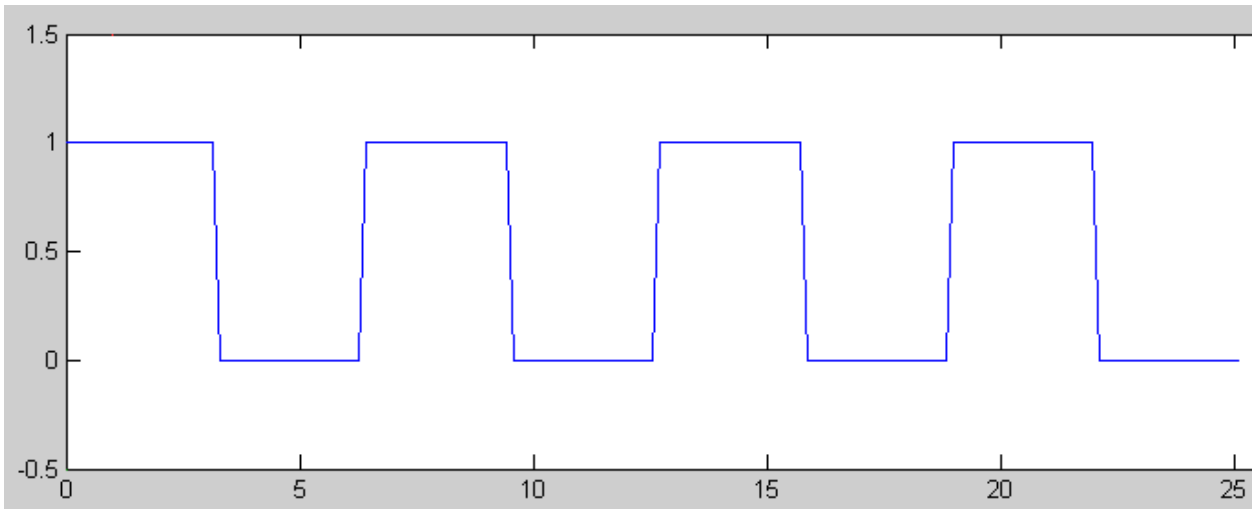


**Figure 1: On/Off switching, the square curve $r(t) = \{_{0 \ \ if \ \ 2\pi k+\pi \le t \le 2\pi(k+1)}^{1 \ \ if \ \ 2\pi k \le t \le 2\pi k+\pi}$.**

We will use this transcription rate in the abundance equations derived for this model. Choosing the following parameters, $n_1(0) = n_2(0) = 0; p = 30\%; L_1 = 1.5; L_2 = 1$, we get the behavior shown in Figure 2, with the blue line representing the abundance of the short SOC3 transcript, and the red line, the long SOC3 transcript. The black line represents the global production rate.

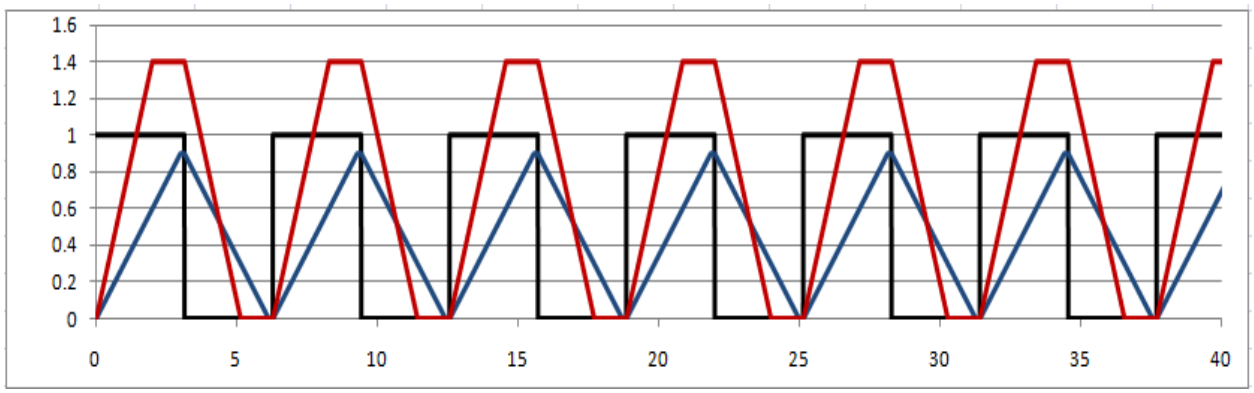


**Figure 2: Model: 1.0: Graph of $n_1$ (red) and $n_2$ (blue) populations when $r(t) = \{_{0 \ \ if \ \ 2\pi k+\pi \le t \le 2\pi(k+1)}^{1 \ \ if \ \ 2\pi k \le t \le 2\pi k+\pi}$ (black).**

This plot shows the piece-wise linear behavior of the populations when the global transcription rate is switching on and off. This rate may not just turn on and off over the course of a day; the rate may oscillate below full production.

We will assume that the global production rate is sinusoidal, letting $r(t) = 1 + \cos(t)$ and letting $2\pi$ radians represent the period of a day. See Figure 3.

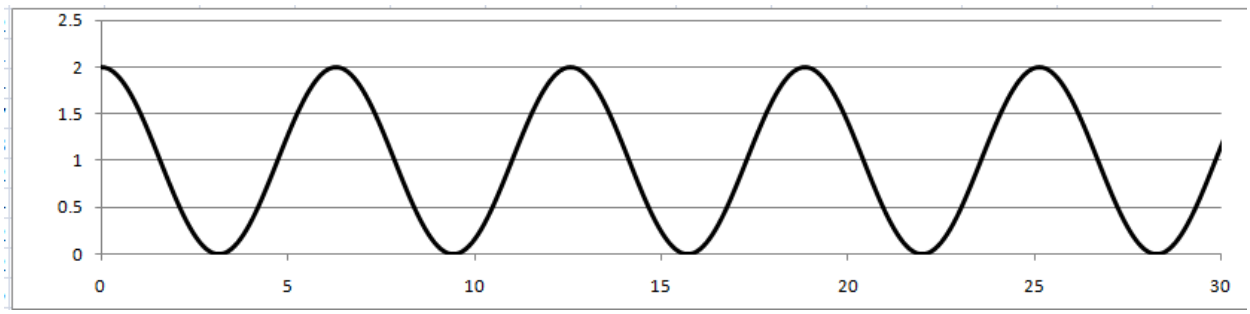


**Figure 3: Oscillating global transcription rate, the sinusoid curve $r(t) = 1 + \cos(t)$.**

Using this formula for the global transcription rate, we can find the abundances of the short and long SOC3 transcriptions in the closed form solutions below.

$$n_1(t) = p * ((\cos(t - L_1) - \cos(t)) - (\cos(L_1) - 1)) + n_1(0)$$
$$n_2(t) = (1 - p) * ((\cos(t - L_2) - \cos(t)) - (\cos(L_2) - 1)) + n_2(0)$$

Using these equations and inserting values for the parameters, we can model the behavior of the populations. The following plot was made by MatLab software written by Jake Blanton. Letting $r(t) = 1 + \cos(t)$, and choosing the following parameters: $p = 30\%; L_1 = 1.5; L_2 = 1$, we get the behavior shown in Figure 4, where the initial values have been chosen so the curves are non-negative. The blue line represents the population of the short SOC3 transcript, $n_1$, and the red line represents the long SOC3 transcript, $n_2$.

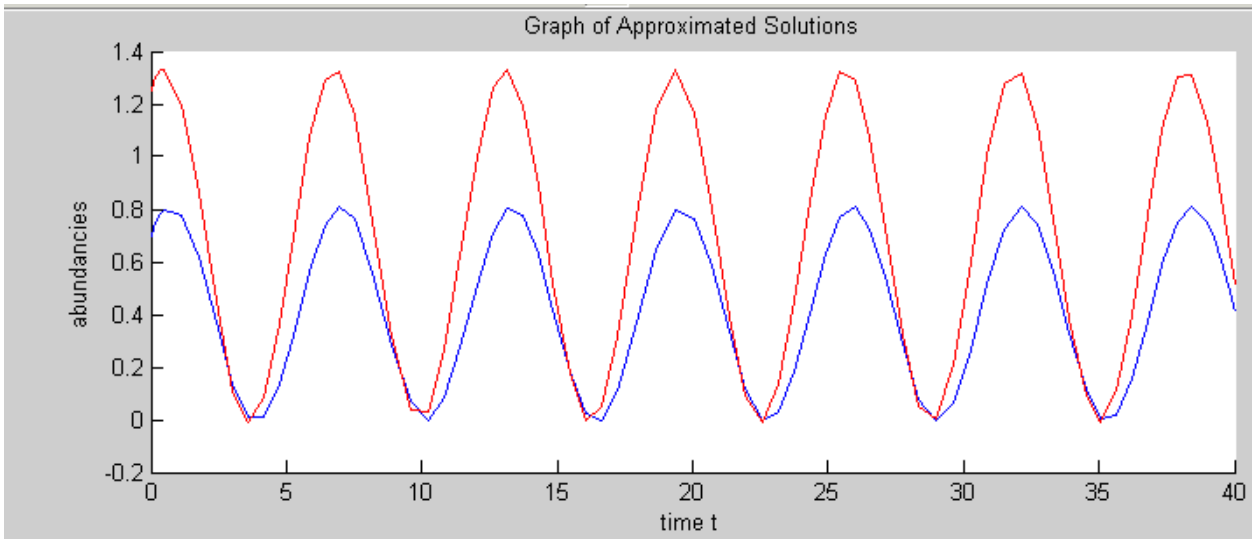


**Figure 4: Model: 1.0: Graph of $n_1$ (blue) and $n_2$ (red) populations when $r(t) = 1 + \cos(t)$.**

This plot shows sinusoidal behavior in the gene abundances, driven by the sinusoidal production rate. Notice the slight phase shift between the red and blue curves. This shift is very small, which does not reflect the experimental data. The experimental data indicate that the gene populations are completely out of phase, effectively adding to a constant function.

We can vary the parameter values to try to produce a plot closer to the experimental data. To accomplish this, we will maximize the phase shift between these curves. By taking a derivative, it is easy to see that the times corresponding to extrema of these graphs depend only on $L_1$ and $L_2$. If we fix the lifespan of the short SOC3 transcript at 6, just shy of one day, and allow the shorter lifespan of the long SOC3 transcript to vary, we can see how the phase shift changes. Figures 5 show how the phase shift can vary as this lifespan takes the values .5 (red), 1 (cyan), 2 (magenta), 3 (black), 4 (green), and 5 (yellow). The initial values for these curves have been set so the abundances are non-negative, and $p$=30%.

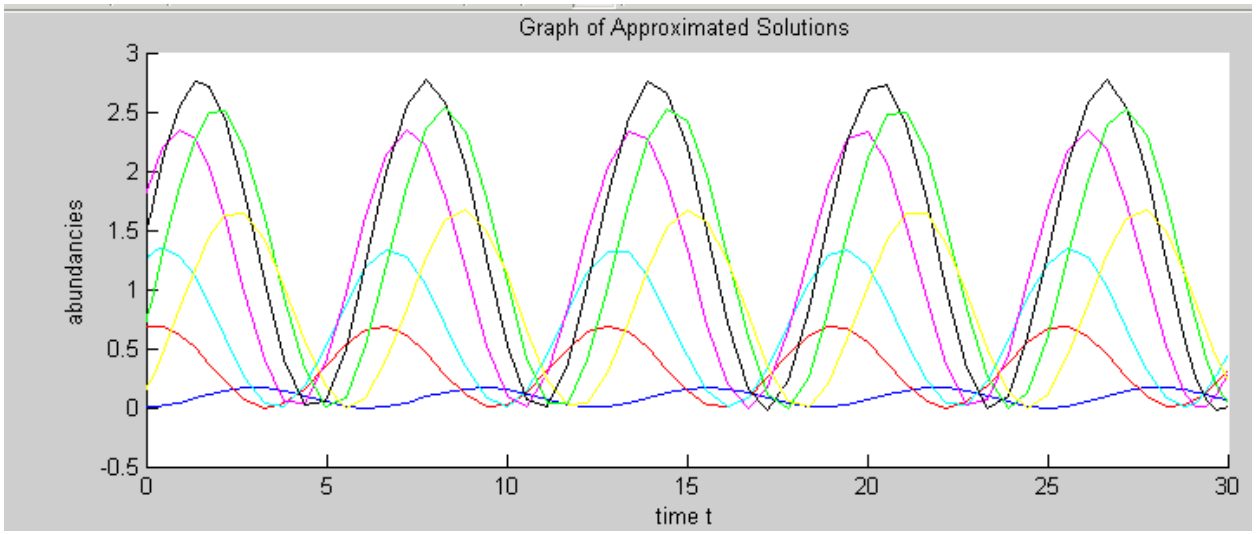


**Figure 5: Model: 1.0: Graph of $n_1$ (blue) and $n_2$ populations when $L_1 = 6$ and $L_2$ is set at .5 (red), 1 (cyan), 2 (magenta), 3 (black), 4 (green), and 5 (yellow).**

This plot shows that a shorter $L_2$ corresponds to a larger phase shift between the population curves for $n_1$ and $n_2$. The amplitude of the $n_2$ curve also decreases as this lifespan decreases. The curve which is closest to out of phase with the blue line, representing the population of the short SOC3 transcript, is the red line, reflecting the shortest $L_1$ value.

Because the phase shift between the two gene populations depends solely on the lifespans, we can calculate the relation. We will assume the curve $r(t) = 1 + \cos(t)$ has 0 phase shift. Because this sinusoid has a maximum at time 0, we will assume the first positive maximum of the population curve is the phase shift. To find this maximum, we will take the derivative and set it to zero.

$$n_1(t) = p * ((\cos(t - L_1) - \cos(t)) - (\cos(L_1) - 1)) + n_1(0)$$
$$\frac{d}{dt} n_1(t) = p * (-\sin(t - L_1) + \sin(t))$$

We can set this derivative equal to zero to find the extrema points.

$$0 = -\sin(t - L_1) + \sin(t)$$
$$0 = -\sin(t)\cos(L_1) + \cos(t)\sin(L_1) + \sin(t)$$

Notice that this equation depends only on the gene's life span, and this equation can be used to find the extrema values for either gene population. If we assume $0 \neq \cos(t)$, we can divide and simplify to find the following. Note that our arctan function provides an angle value $-\frac{\pi}{2} < t < \frac{\pi}{2}$.

$$t = \arctan\left(\frac{\sin(L_i)}{\cos(L_i) - 1}\right)$$

After comparing this equation with the phase shifts found in Figure 6, we can determine that these time values reflect not the maxima, but the minima. Because these values are 0.5π apart, the following equation gives us the relation between the lifespan and the phase shift (from $r(t) = 1 + \cos(t)$).

$$\theta = \arctan\left(\frac{\sin(L_i)}{\cos(L_i) - 1}\right) + \frac{\pi}{2}$$

This function is graphed in Figure 6.

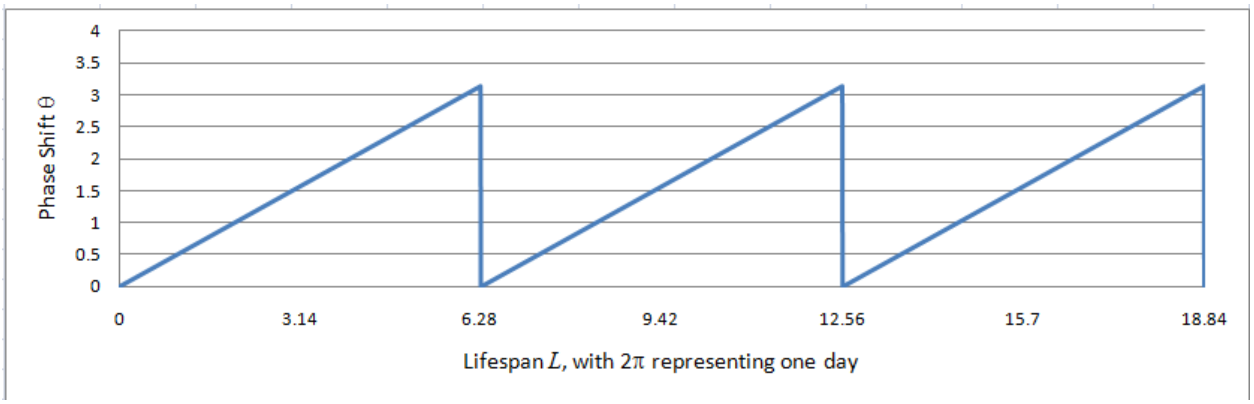


**Figure 6: Model: 1.0: The phase shift from $r(t) = 1 + \cos(t)$ to the sinusoid curve when a gene's lifespan is $L$.**

This plot shows the relation between the phase shift and the gene's lifespan is linear with slope 0.5 for $0 < L < 2\pi$ and the curve is periodic.

This linear behavior is a caused by the argument of the arctangent function being so similar to tangent. This curve shows that the population curve of one gene is never more than π radians from $\boldsymbol{r(t) = 1 + \cos(t)}$. One way to produce two curves which are almost perfectly out of phase would be to let one lifespan approach 2π from the left and the other from the right. The closed form solutions show that as a gene's lifespan approaches 2π, however, the population curve approaches a constant.

If we select lifespans close to 2π, we get the phase shift shown in Figure 7. This plot has been normalized so no negative abundances appear.

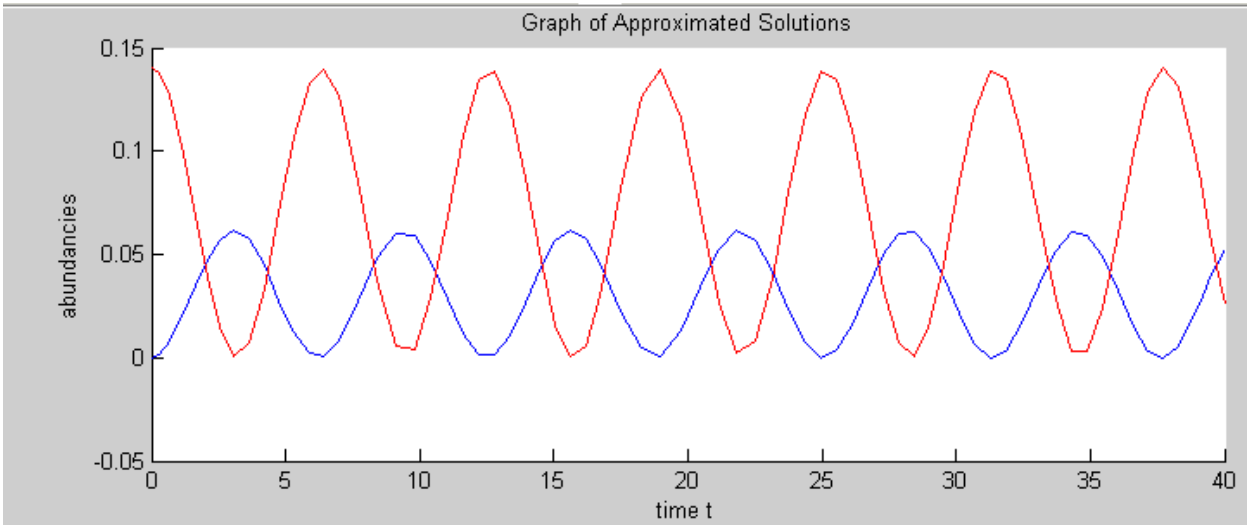


**Figure 7: Model: 1.0: Graph of $n_1$ (blue) and $n_2$ (red) populations when $r(t) = 1 + \cos(t)$ ; $L_1 = 6.18$; $L_2 = 6.28$.**

This plot shows nearly perfectly out of phase population curves. Notice the extremely small amplitudes of these curves. Figure 8 shows how these curve amplitudes change as lifespan changes when $p$=30%.

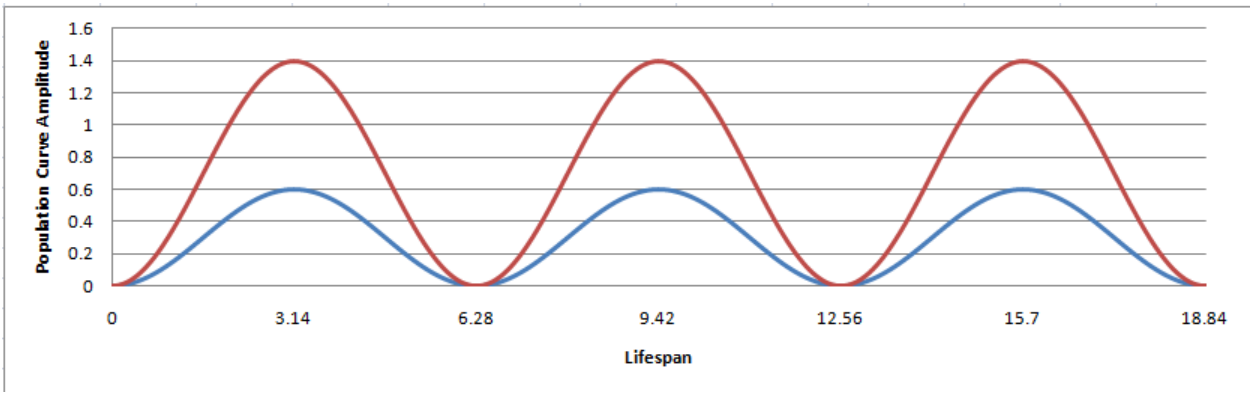


**Figure 8: Model: 1.0: The amplitude of $n_1$ (blue) and $n_2$ (red) populations depending on the gene's lifespan is $L_i$.**

This plot shows that as either gene's lifespan approaches $2\pi$, the amplitude of the population curve will approach 0. This trend can also be seen in Figure 5.

Now, we will refine this model by adding a predation parameter.

**Model 1.1: Fixed life-span model with a predation parameter.**

Model 1.1 adds another level of complexity to Model 1. We incorporate a predation constant, which affects only the long mRNA transcription. This model will use the following state variables and parameters.

The state variables:

$n_1(t)$ = amount of the short mRNA transcription of the gene (quantity)

$n_2(t)$ = amount of the long mRNA transcription of the gene (quantity)

$r(t)$ = global transcription rate of the gene (quantity per time)

The constant parameters:

$p$ = probability that if a gene is transcribed, a short mRNA transcript results (%)

$L_1$ = average life span for the short transcription (time)

$d$ = decay rate of the long transcription due to predation (% per time)

To find the amount of $n_i$ present, we will add the amount which was present to the amount produced and subtract the amount lost. We will consider how this happens in the time step, $\Delta t$, it takes to transcribe the SOC3 gene, although several copies of this gene can be simultaneously transcribed. In this model, the amount of $n_i$ produced is the global gene production multiplied by the probability that any gene transcribed will be type $n_i$. The amount of $n_1$ lost is the amount which was produced one lifespan ago. The amount of $n_2$ lost is the amount lost to predation, a fixed percentage of the amount which was present. The formal equations follow.

$$n_1(t) = n_1(t-\Delta t) + r(t-\Delta t) * p * \Delta t - r(t-\Delta t-L_1) * p * \Delta t$$
$$n_2(t) = n_2(t-\Delta t) + r(t-\Delta t) * (1-p) * \Delta t - d * \Delta t * n_2(t-\Delta t)$$

These basic recursive models are more useful in the following form, found by taking the limit of these equations as the time step approaches 0.

$$\dot{n_1} = p * (r(t) - r(t-L_1))$$
$$\dot{n_2} = (1-p) * r(t) - n_2(t) * d$$

We can assume the global transcription rate switches from on to off over the course of a day, represented by $2\pi$ radians, see Figure 1. Choosing the parameters $n_1(0) = n_2(0) = 0; p = 30\%; L_1 = 1; d = .4$, we get the behavior shown in Figure 9.

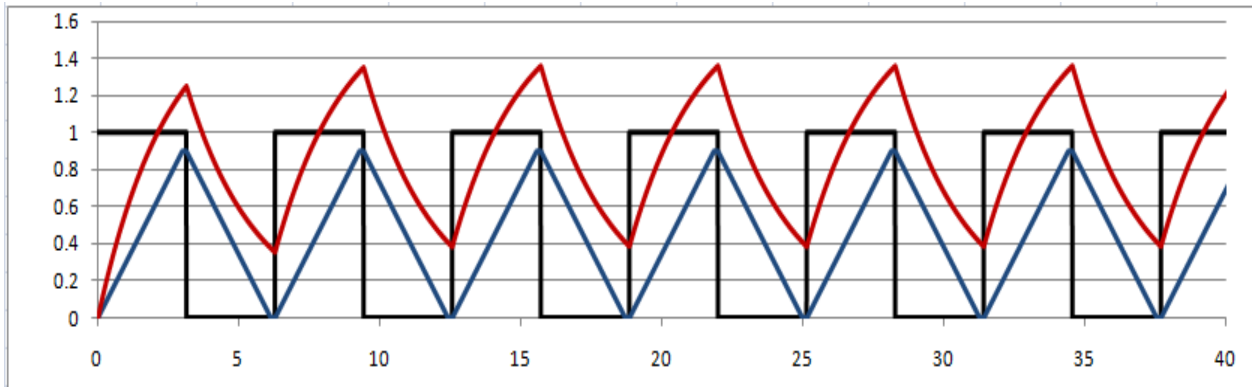


**Figure 9: Model 1.1: Graph of $n_1$ (blue) and $n_2$ (red) populations when $r(t) = \{_{0 \ \ if \ \ 2\pi k+\pi \le t \le 2\pi(k+1)}^{1 \ \ if \ \ 2\pi k \le t \le 2\pi k+\pi}$ (black).**

This plot shows the piece-wise linear behavior of short SOC3 transcript population and the similar behavior of the long SOC3 transcript population when the global transcription rate is switching on and off. This rate may not just turn on and off over the course of a day; the rate may oscillate below full production.

We will assume that the global production rate is sinusoidal, letting $r(t) = 1 + \cos(t)$ and letting $2\pi$ radians represent the period of a day. Using this equation, we can find the abundances of the short and long SOC3 transcriptions in the closed form solutions below.

$$n_1(t) = p * ((\cos(t - L_1) - \cos(t)) - (\cos(L_1) - 1)) + n_1(0)$$
$$n_2(t) = \frac{(1-p)}{(d_2)^2 + 1}(d_2 \cos(t) + \sin(t) - d_2) + n_2(0)$$

Using these equations and inserting values for the parameters, we can model the behavior of the populations. The following plot was made using MatLab software written by Jake Blanton. Letting $r(t) = 1 + \cos(t)$, and choosing the following parameters: $p = 30\%; L_1 = 1; d = .4$, we get the behavior shown in Figure 10, where the initial values have been chosen so the curves are non-negative. The blue line represents the population of $n_1$; the red line, $n_2$.

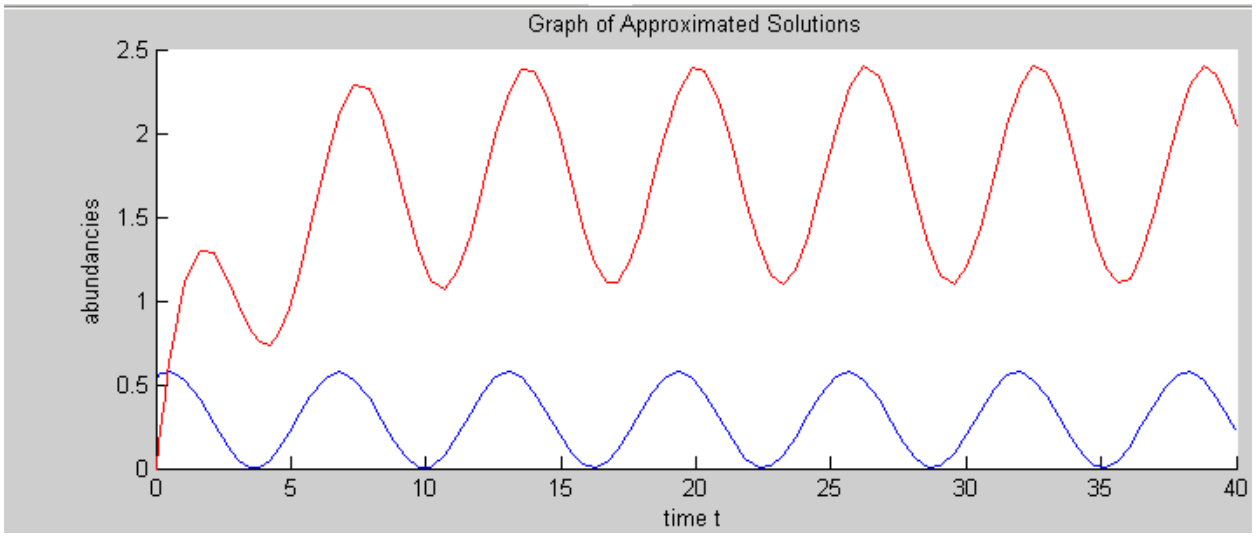


**Figure 10: Model 1.1: Graph of $n_1$ (blue) and $n_2$ (red) populations when $r(t) = 1 + \cos(t)$.**

This plot shows the sinusoidal behavior of the two gene populations. Again, notice the small phase shift, which does not reflect the larger phase shift apparent in the experimental data. Changing the parameters for this model within plausible ranges did not produce a large enough phase shift to cause these two curves to be completely out of phase.

There may be that more of the parameters vary with time. For example, it may be that the probability that a short transcription is produced may decline as the global SOC3 gene population increases. Since this version of the gene has a longer lifespan, this variation would help stabilize the global population. There is, however, no experimental data to support this hypothesis. Model 1 may be further refined in the future, but now we consider a different model.

**Model 2: The basic production rate and decay rate model.**

This model is a basic production rate and decay rate model. The relevant constants and variables follow. Separating the parameter modeled by lifespan into production and decay rates should better model the probabilistic behavior. We introduce this model assuming a fixed global production rate.

The state variables:

$n_1(t)$ = amount of the short mRNA transcript of the gene (quantity)

$n_2(t)$ = amount of the long mRNA transcript of the gene (quantity)

The constant parameters:

$r$ = global transcription rate of the gene (quantity per time)

$p$ = probability that if a gene is transcribed, a short mRNA transcript results (%)

$d_1$ = rate of decay for the short transcript (% per time)

$d_2$ = rate of decay for the long transcript (% per time)

To find the amount of $n_i$ present, we will add the amount which was present to the amount produced and subtract the amount lost. We will consider how this happens in the time step, $\Delta t$, it takes to transcribe the SOC3 gene, although several copies of this gene can be simultaneously transcribed. In this model, the amount of $n_i$ produced is the total amount of the gene production multiplied by the probability that any gene transcribed will be type $n_i$. The amount of $n_i$ lost is a fixed percentage of the amount which was present. The formal equations follow.

$$n_1(t) = n_1(t-\Delta t) + r * p * \Delta t - n_1(t-\Delta t) * d_1 * \Delta t$$
$$n_2(t) = n_2(t-\Delta t) + r * (1-p) * \Delta t - n_2(t-\Delta t) * d_2 * \Delta t$$

These basic recursive models are more useful in the following form, found by taking the limit of these equations as the time step approaches 0.

$$\dot{n_1} = r * p - n_1 * d_1$$
$$\dot{n_2} = r * (1-p) - n_2 * d_2$$

The closed-form solutions to these differential equations follow.

$$n_1(t) = \frac{1}{d_1} * (r * p + (n_1(0) * d_1 - r * p) * e^{-d_1 t})$$
$$n_2(t) = \frac{1}{d_2} * (r * (1-p) + (n_2(0) * d_2 - r * (1-p)) * e^{-d_2 t})$$

Using these equations and inserting values for the parameters, we can model the behavior of the populations. The following plots (see Figure 11) were made using MatLab software written by Jake Blanton. For the first plot, we set the following parameter values: $n_1(0) = n_2(0) = 0; r = 1; p = 30\%; d_1 = .2; d_2 = .5.$

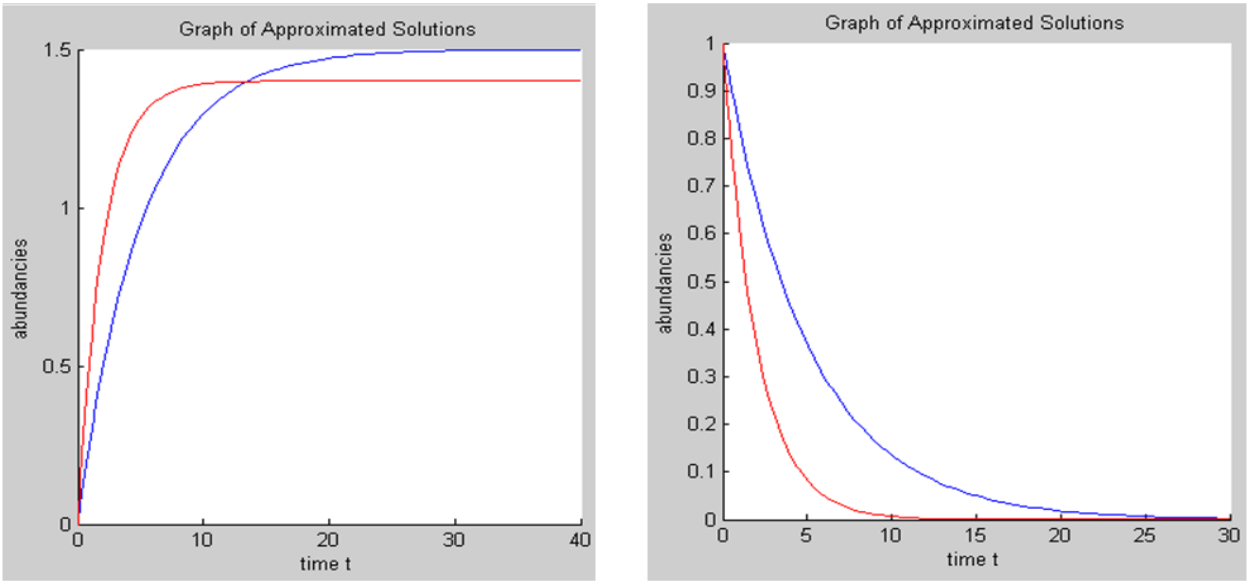


**Figure 11: Model 2.0: Graph of $n_1$ (blue) and $n_2$ (red) populations for constant input rate 1 (left) or 0 (right).**

These plots show the gene populations behaviors from initial population amounts to steady state abundances, with the blue line modeling the population of $n_1$ and the red line, $n_2$. In the second plot, on the right, we set the following parameter values: $n_1(0) = n_2(0) = 1; r = 0; p = 30\%; d_1 = .2; d_2 = .5$ to show the decay of the populations when the gene is not transcribed. Again, the blue line represents the population of $n_1$ and the red line, $n_2$. These plots show simple, exponential behavior. In order to better approximate the complexity of the actual process, we will need to allow some of these fixed parameters to become variables.

**Model 2.1: Introducing a variable production rate**.

Now, we can replace the fixed production rate with a variable production rate, $r(t)$. The state variables and constant parameters follow.

The state variables:

$n_1(t)$ = amount of the short mRNA transcript of the gene (quantity)

$n_2(t)$ = amount of the long mRNA transcript of the gene (quantity)

$r(t)$ = global transcription rate of the gene (quantity per time)

The constant parameters:

$p$ = probability that if a gene is transcribed, a short mRNA transcript results (%)

$d_1$ = rate of decay for the short transcript (% per time)

$d_2$ = rate of decay for the long transcript (% per time)

To find the amount of $n_i$ present, we will add the amount which was present to the amount produced and subtract the amount lost. We will consider how this happens in the time step, $\Delta t$, it takes to transcribe the SOC3 gene, although several copies of this gene can be simultaneously transcribed. In this model, the amount of $n_i$ produced is the total amount of the gene production multiplied by the

probability that any gene transcribed will be type $n_i$. The amount of $n_i$ lost is a fixed percentage of the amount which was present. The formal equations follow.

$$n_1(t) = n_1(t - \Delta t) + r(t - \Delta t) * p * \Delta t - n_1(t - \Delta t) * d_1 * \Delta t$$
$$n_2(t) = n_2(t - \Delta t) + r(t - \Delta t) * (1 - p) * \Delta t - n_2(t - \Delta t) * d_2 * \Delta t$$

These basic recursive models are more useful in the following form, found by taking the limit of these equations as the time step approaches 0.

$$\dot{n_1} = r(t) * p - n_1 * d_1$$
$$\dot{n_2} = r(t) * (1 - p) - n_2 * d_2$$

The solutions to these differential equations depend on the behavior of the global transcription rate. First, we will assume this rate switches on and off, and $r(t) = \{_{0 \ \ if \ \ 2\pi k + \pi \le t \le 2\pi(k+1)}^{1 \ \ if \ \ 2\pi k \le t \le 2\pi k + \pi}$ as shown in Figure 1. Using this global production rate, and the fixed parameters $n_1(0) = n_2(0) = 0; p = 30\%; d_1 = .2; d_2 = .5$, we get the behavior shown in Figure 12.

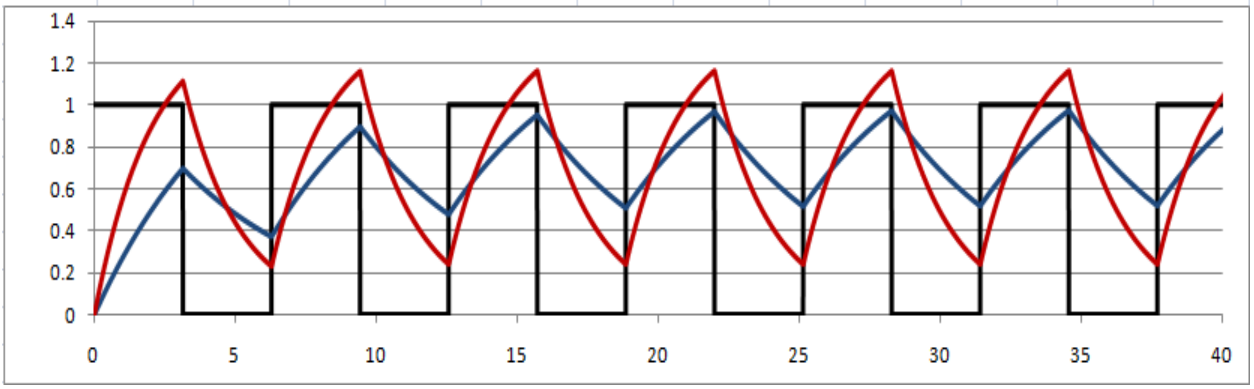


**Figure 12: Model 2.1: Graph of $n_1$ (blue) and $n_2$ (red) populations when $r(t) = \{_{0 \ \ if \ \ 2\pi k + \pi \le t \le 2\pi(k+1)}^{1 \ \ if \ \ 2\pi k \le t \le 2\pi k + \pi}$ (black).**

This plot shows the periodic behavior of the populations when the global transcription rate is switching on and off. This rate may not just turn on and off over the course of a day; the rate may oscillate below full production.

We will assume that the global production rate is sinusoidal, letting $r(t) = 1 + \cos(t)$ and letting $2\pi$ radians represent the period of a day. Using this equation, we can find the abundancies of the short and long SOC3 transcriptions in the closed form solutions below.

$$n_1(t) = n_1(0) + \frac{p}{(d_1)^2 + 1}(d_1 \cos(t) + \sin(t) - d_1)$$
$$n_2(t) = n_2(0) + \frac{(1 - p)}{(d_2)^2 + 1}(d_2 \cos(t) + \sin(t) - d_2)$$

Using the software developed by Jake Blanton, and setting the fixed parameters $n_1(0) = n_2(0) = 0; p = 30\%; d_1 = .2; d_2 = .5$, we find the behavior shown in Figure 13.

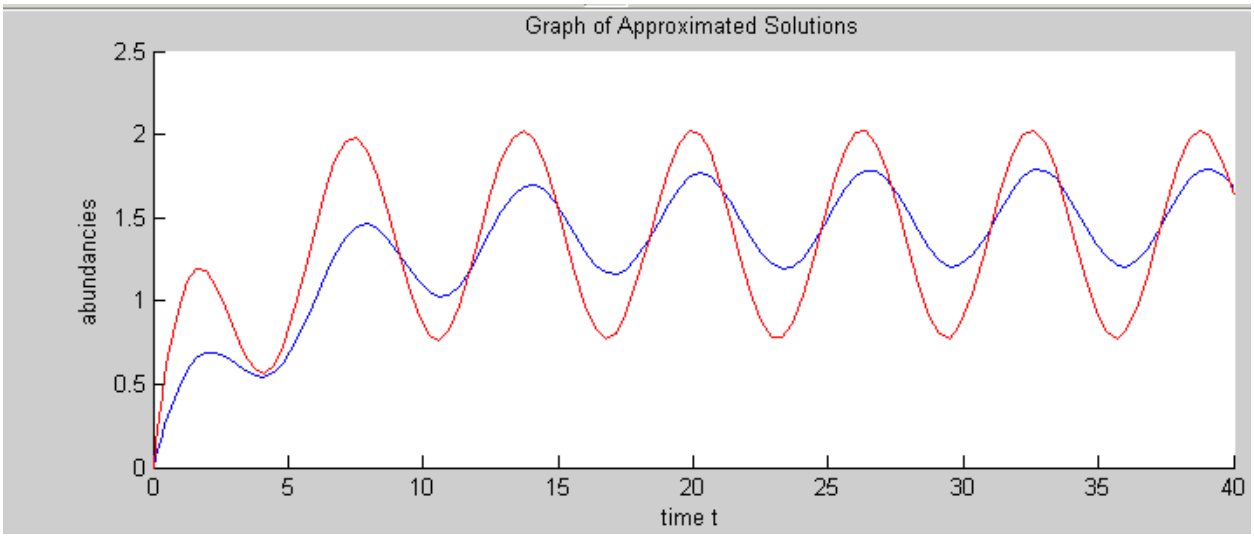


**Figure 13: Model 2.1: Graph of $n_1$ (blue) and $n_2$ (red) populations when $r(t) = 1 + \cos(t)$.**

This plot shows sinusoidal behavior in the gene abundancies, driven by the sinusoidal production rate. This model, like the previous models, shows a small phase shift between the red and blue curves. Again, this small shift does not reflect the experimental data, which indicate the curves should be completely out of phase.

We can vary the parameter values to try to produce a plot closer to the experimental data. To accomplish this, we will maximize the phase shift between these curves. By taking a derivative, it is easy to see that the times corresponding to extrema of these graphs depend only on $d_1$ and $d_2$. By varying these values between .1 and .9, the maximum phase shift that can be produced is just over .6. In a period of $2\pi$, this is still a very small phase shift.

This model is not flexible enough to produce the desired output. We can further refine the model by allowing another parameter to vary.

**Model 2.2: Introducing a variable** $p$.

Until this point, we have assumed the probability that a transcribed gene was a short transcript or a long transcript was fixed. It may be that, in order to maintain a constant global quantity of the gene, when the global amount of the gene is low, the hardier, longer-lived transcription is produced. We will assume the probability that a short transcript (the longer-lived version of the gene) is more likely to be produced if the global gene supply is low. The state variables and constant parameters follow. The state variables:

$n_1(t)$ = amount of the short mRNA transcript of the gene (quantity)
$n_2(t)$ = amount of the long mRNA transcript of the gene (quantity)
$r(t)$ = global transcription rate of the gene (quantity per time)
$p(t)$ = probability that if a gene is transcribed, a short mRNA transcript results (%)

The constant parameters:

$d_1$ = rate of decay for the short transcript (% per time)

$d_2$ = rate of decay for the long transcript (% per time)

To find the amount of $n_i$ present, we will add the amount which was present to the amount produced and subtract the amount lost. We will consider how this happens in the time step, $\Delta t$, it takes to transcribe the SOC3 gene, although several copies of this gene can be simultaneously transcribed. In this model, the amount of $n_i$ produced is the total amount of the gene production multiplied by the probability that any gene transcribed will be type $n_i$. The amount of $n_i$ lost is a fixed percentage of the amount which was present. The formal equations follow.

$$n_1(t) = n_1(t-\Delta t) + r(t-\Delta t) * p(t-\Delta t) * \Delta t - n_1(t-\Delta t) * d_1 * \Delta t$$
$$n_2(t) = n_2(t-\Delta t) + r(t-\Delta t) * \big(1 - p(t-\Delta t)\big) * \Delta t - n_2(t-\Delta t) * d_2 * \Delta t$$

These basic recursive models are more useful in the following form, found by taking the limit of these equations as the time step approaches 0.

$$\dot{n_1} = r(t) * p(t) - n_1 * d_1$$
$$\dot{n_2} = r(t) * (1 - p(t)) - n_2 * d_2$$

The solutions to these differential equations depend on the behavior of the global transcription rate and the behavior of the probability that when a gene is transcribed, a short transcription is produced.

We will assume the variable $p(t)$ tries to stabilize the global gene population at 2, and let $p(t) = 1 - .5 * (n_1 + n_2)$. First, we will assume the global production rate switches on and off, and $r(t) = \{_{0 \ \ if \ \ 2\pi k + \pi \le t \le 2\pi(k+1)}^{1 \ \ if \ \ 2\pi k \le t \le 2\pi k + \pi}$ as shown in Figure 1. Using the fixed parameters $n_1(0) = n_2(0) = 0; d_1 = .2; d_2 = .5$, we get the behavior shown in Figure 14.

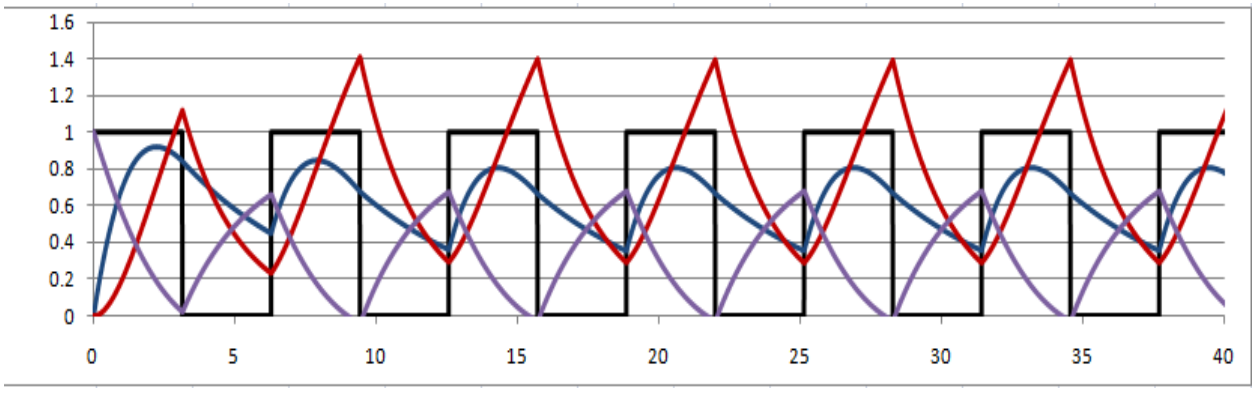


**Figure 14: Model 2.2: Graph of $n_1$ (blue) and $n_2$ (red) populations when $r(t) = \{_{0 \ \ if \ \ 2\pi k + \pi \le t \le 2\pi(k+1)}^{1 \ \ if \ \ 2\pi k \le t \le 2\pi k + \pi}$ (black) and $p(t) = 1 - .5 * (n_1 + n_2)$ (purple).**

This plot shows the periodic behavior of the populations when the global transcription rate is switching on and off and the probability that the short SOC3 transcript will be made changes based on the global population, as shown by the purple curve in Figure 10. The global production rate, however, may not just turn on and off over the course of a day; the rate may oscillate below full production.

We will assume that the global production rate is sinusoidal, letting $r(t) = 1 + \cos(t)$ and letting $2\pi$ radians represent the period of a day. We will let $p(t) = 1 - .3 * (n_1 + n_2)$. Using the software developed by Jake Blanton, and choosing the fixed parameters $n_1(0) = n_2(0) = 0; d_1 = .2; d_2 = .5$, we find the behavior shown in Figure 15.

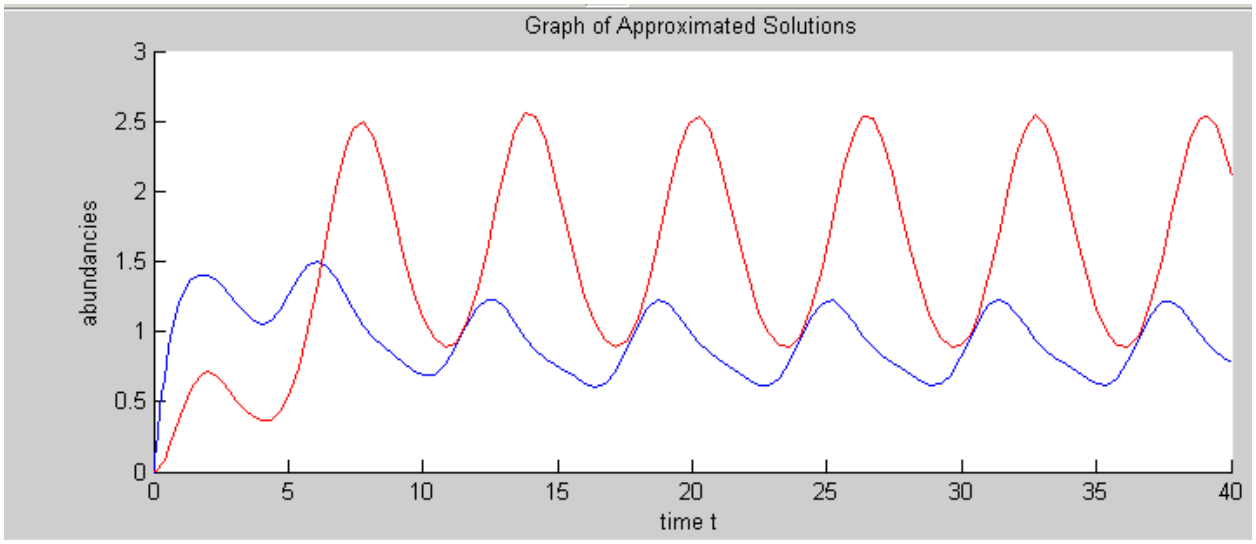


**Figure 15: Model 2.2: Graph of $n_1$ (blue) and $n_2$ (red) populations with $p(t) = 1 - .3 * (n_1 + n_2)$ and $r(t) = 1 + \cos(t)$.**

This plot shows periodic behavior in the gene abundances and *p(t)*. Unlike the behavior in previous models, the behavior of these curves is not purely sinusoidal. The red and blue curves are close enough to sinusoids, however, to see an apparent small phase shift. Again, this small shift does not reflect the experimental data, which indicates the curves should be completely out of phase.

## IV. An Alternate Theory

The models developed in this report all failed to produce population curves which are completely out of phase. This means that either there is another significant factor that none of these models has accounted for, or the original interpretation of the data is not correct. In an organism, SOC3 gene production is part of an appetite suppressor signal. It may make more sense to expect the global SOC3 gene production to be sinusoidal, and the populations of the different transcriptions to oscillate inside this sinusoid. Also, the extreme granularity of the data may have masked this oscillation.

Every model developed in this report, with the exception of Model 2.0 which lacked a variable global transcript rate, produces this type of oscillation. For example, we examine Model 2.1 with the following behavior of the populations of the short and long SOC3 transcripts.

$$n_1(t) = n_1(0) + \frac{p}{(d_1)^2 + 1}(d_1 \cos(t) + \sin(t) - d_1)$$
$$n_2(t) = n_2(0) + \frac{(1-p)}{(d_2)^2 + 1}(d_2 \cos(t) + \sin(t) - d_2)$$

We can normalize by the sum, or global population, $n_g(t) = n_1(t) + n_2(t)$. These curves are shown in Figure 16, with parameters: $n_1(0) = n_2(0) = 1; n_g(0) = 2; p = 30\%; d_1 = .2; d_2 = .5$. These parameters differ from those used to produce Figure 10, because now, we will require the initial values for the populations to be non-zero, since to normalize, we will divide by the global population, $n_g$ .

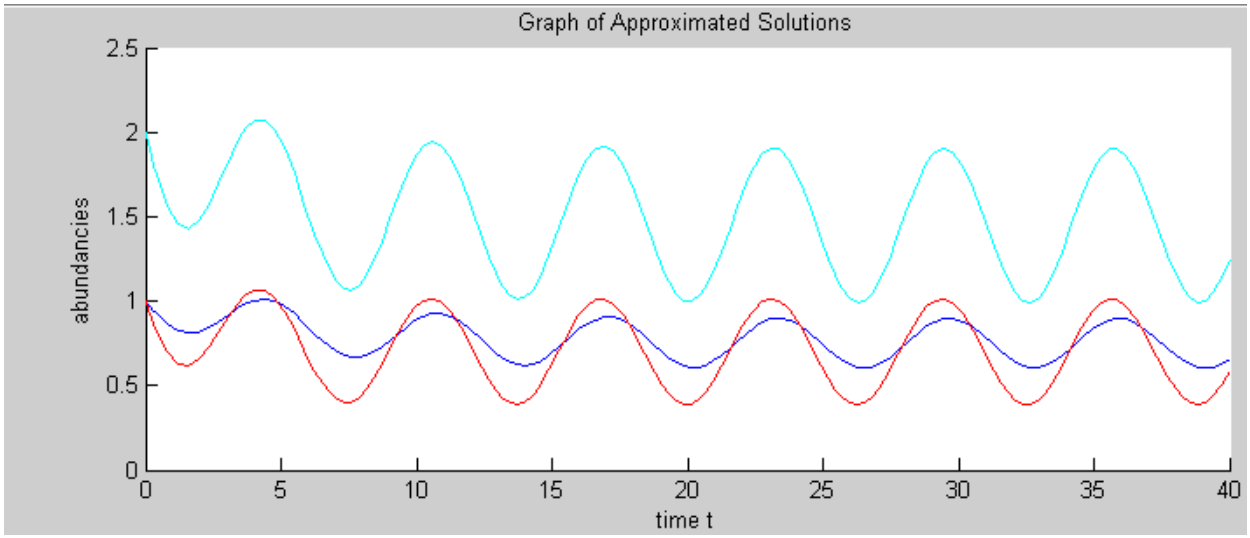


**Figure 16: Model 2.1: Graph of $n_1$ (blue), $n_2$ (red), and $n_g$ (cyan) populations.**

This plot shows the oscillation of the different transcripts inside the sinusoidal global population. It will be easier to see the oscillation inside the sinusoid after normalizing by the global gene abundancy, see Figure 16.

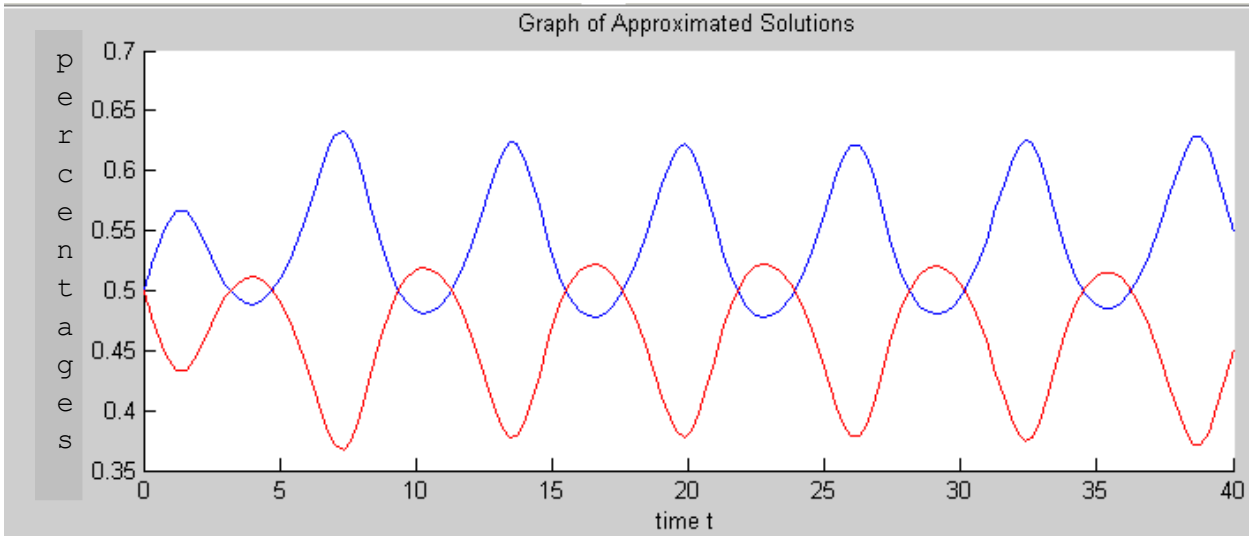


**Figure 16: Model 2.1: Graph of percentages of $n_1$ (blue) and $n_2$ (red) inside the global $n_g$ population.**

This plot shows the behavior of the percentages of the short and long transcripts. These curves add to a total of 100%, and behave like out of phase sinusoids.

## V. Shocking the Model

Now, we will consider the behavior of Model 2.1 when certain shocks are introduced. We will test the model's behavior when the system is infused with extra short or long SOC3 transcriptions or when these quantities are artificially reduced. Note that if we examine shocking Model 1.0, the effects would disappear after one lifespan. We will also test the model when the global production rate changes in certain ways.

The following plot (Figure 17) shows the population curves for this model using the parameters discussed in section III.

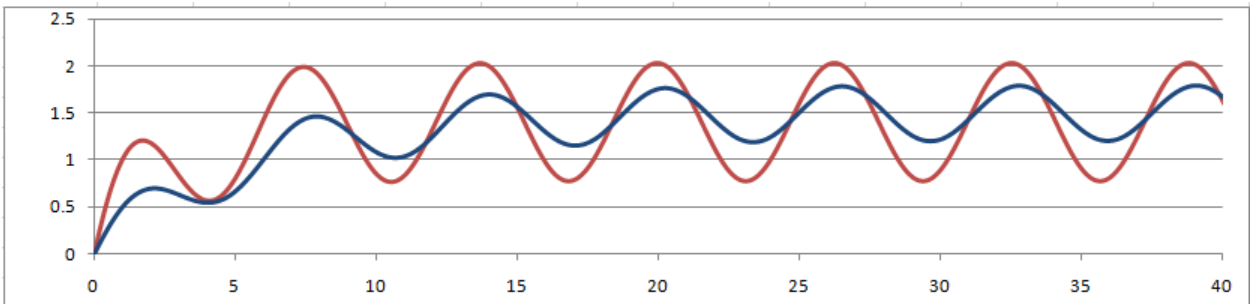


**Figure 17: Model 2.1: Graph of $n_1$ (blue) and $n_2$ (red) for $\boldsymbol{n_1(0) = n_2(0) = 0; p = 30\%; d_1 = .2; d_2 = .5; r(t) = 1 + \cos(t)}$.**

The populations appear to reach a stable oscillation by the third day, at time $t$=3*2π.

We will consider an influx of three units to the population of $n_1$, the short SOC3 transcript, to see how the populations might behave. See Figure 18.

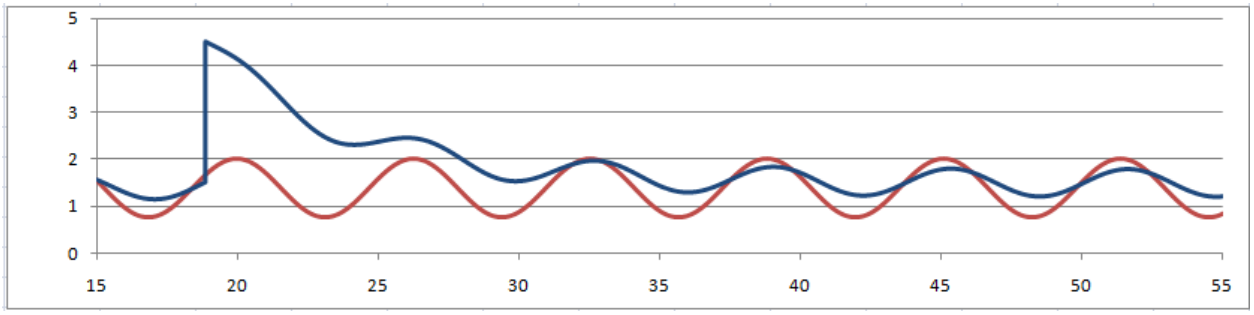


**Figure 18: Shocked Model 2.1: Graph of $n_1$ (blue) and $n_2$ (red) with an influx of $n_1$.**

This plot show that, while the immediate effect of the influx is dramatic, tripling the population of the short SOC3 transcript, the population recovers normal behavior relatively quickly, after less than three days (by $t$=6*2π). Isolating this effect by subtracting the population curve shown in blue in Figure 17, we can see the dampening very clearly. See Figure 19.

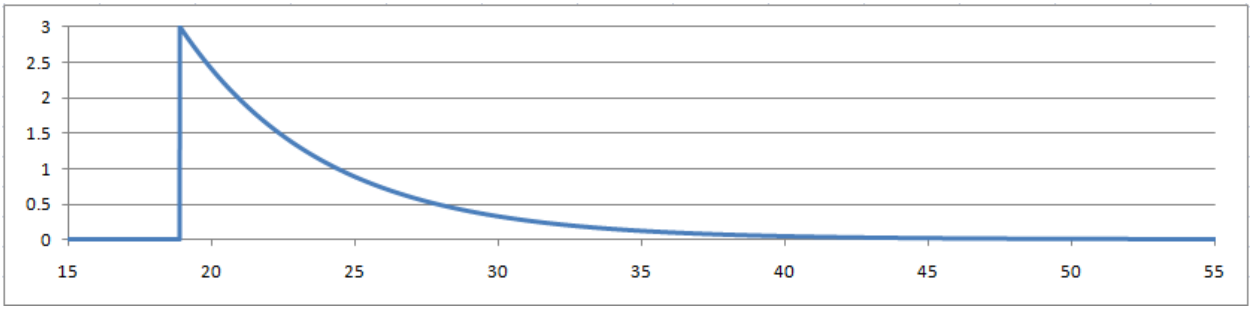


**Figure 19: Shocked Model 2.1: Graph of the increase in $n_1$ (blue) caused by the shock.**

This plot shows that the shock is dampened within approximately 3 days, represented by $\Delta t$=3*2$\pi$. This curve displays the exponential decrease caused by the decay rate, and, in fact, the height of this curve is decreases exactly as the exponential $e^{-0.2t}$. If this model accurately predicts the population behavior, the population will recover from the shock (if we assume within .03 of normal is recovered) based on the decay rate.

$$.03 = 3e^{-d*t}$$
$$t = \frac{\ln(.01)}{-d}$$

And for our decay rate, we expect full recovery in 3.66 days (for $\Delta$t=3.66*2$\pi$).

We will expect similar behavior if we shock the population of the long SOC3 transcript, with the exponential dampening behaving like $e^{-0.5t}$. See Figure 20.

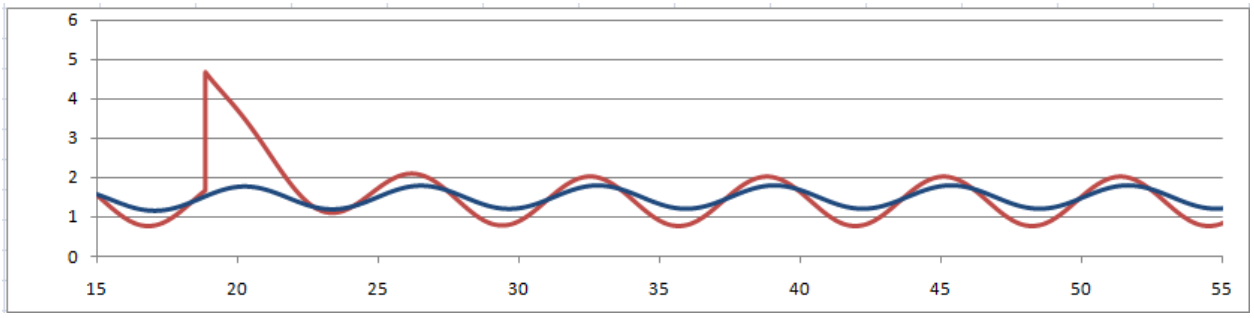


**Figure 20: Shocked Model 2.1: Graph of $n_1$ (blue) and $n_2$ (red) with an influx of $n_2$.**

This plot shows the influx of the long SOC3 transcript is dampened even more quickly than the influx of the short transcript was. This exponential damping behaves as predicted, like simple, exponential decay. The recovery time for this population curve is 1.47 days.

Because the populations are independent of each other in Model 2.1, shocking both at the same time produces the effects shown in Figures 19 and 20. See Figure 21.

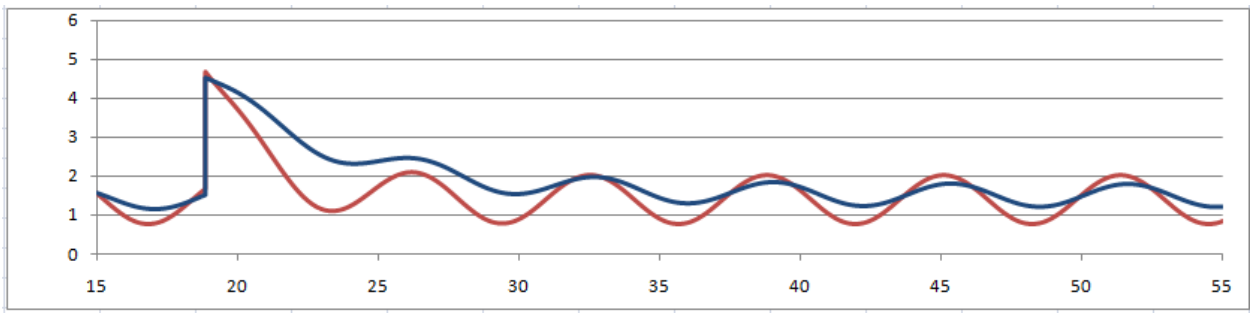


**Figure 21: Shocked Model 2.1: Graph of $n_1$ (blue) and $n_2$ (red) with an influx of $n_1$ and $n_2$.**

This plot shows the recovery of the population curves from influxes of genes.

This population system can also be shocked by a sudden decrease in the gene population. For example, the population of the short SOC3 transcript may suddenly drop to zero. See Figure 22.

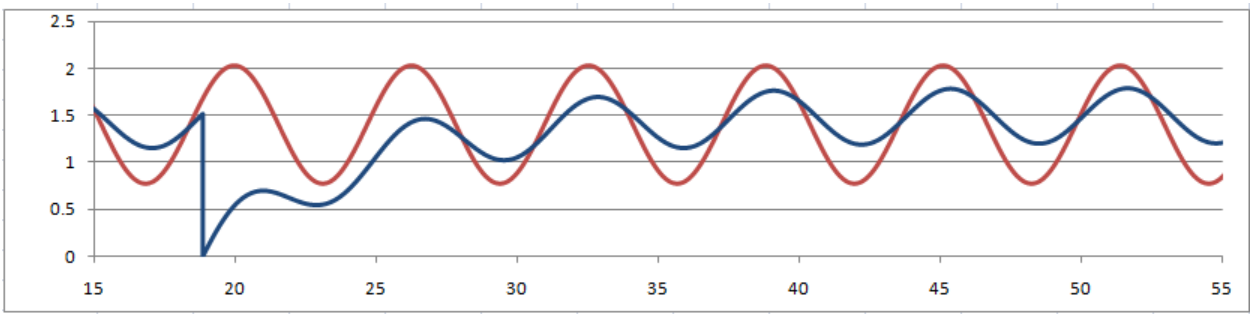


**Figure 22: Shocked Model 2.1: Graph of $n_1$ (blue) and $n_2$ (red) when $n_1$ is depleted.**

This plot shows that the population recovers from the depletion of the short SOC3 transcript. The difference between this curve and the normal curve is shown in Figure 23.

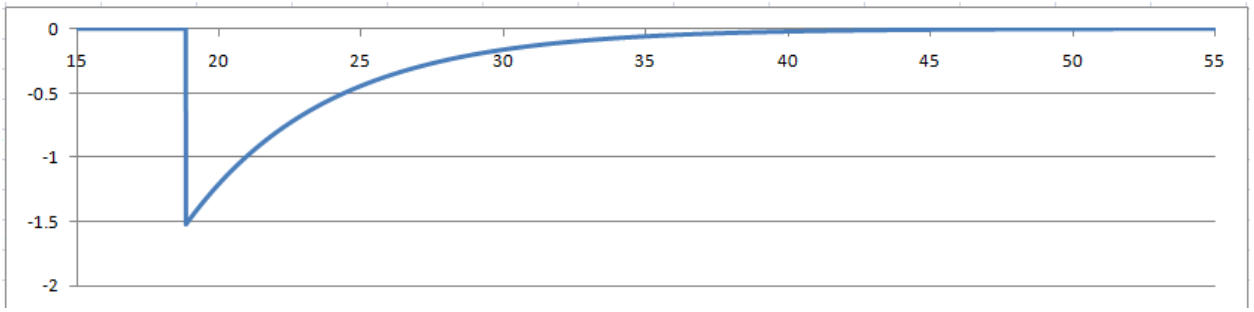


**Figure 23: Shocked Model 2.1: Graph of the decrease in $n_1$ (blue) caused by the shock.**

This plot shows that the shock is dampened within approximately 3 days, represented by $\Delta t=3*2\pi$. The population's deviation from normal decreases like the deviation from normal decreased with the influx of genes. This curve is just a scaled version of the negative of the curve in Figure 19. The deviation decreases exactly as the exponential $e^{-0.2t}$. If this model accurately predicts the population behavior, the population will recover from the shock (if we assume within .03 of normal is recovered) based on the decay rate.

$$-.03 = -1.52 * e^{-d*t}$$
$$t = \frac{\ln(.0197)}{-d}$$

And for our decay rate, we expect full recovery in 3.12 days (for $\Delta t = 3.12*2\pi$).

We will expect similar behavior if we shock the population of the long SOC3 transcript, with the exponential dampening behaving like $e^{-0.5t}$. See Figure 24.

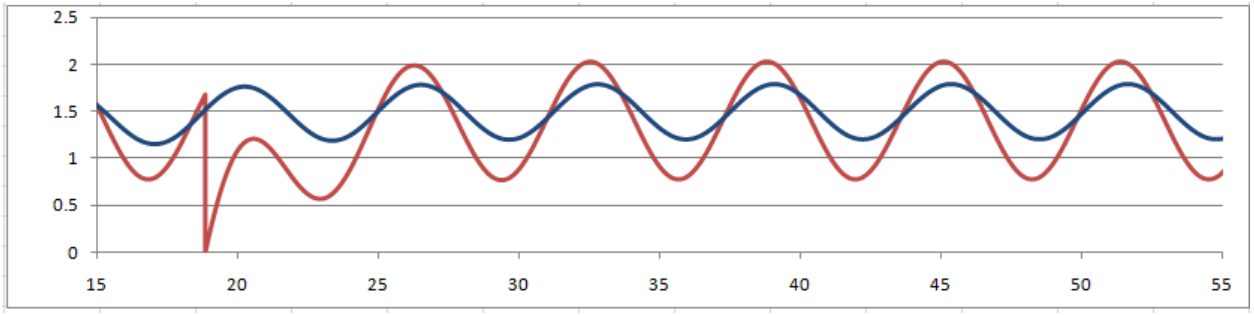


**Figure 24: Shocked Model 2.1: Graph of $n_1$ (blue) and $n_2$ (red) when $n_2$ is depleted.**

This plot shows the depletion of the long SOC3 transcript is corrected even more quickly than the depletion of the short transcript was. This exponential dampening behaves as predicted, like simple, exponential decay. The recovery time for this population curve is 1.25 days.

Another way the system can experience a shock is if the global transcription rate changes from the normal oscillating curve. The global production rate may persist at peak production for a full day, or production could stop for a day. The production oscillation could speed up or slow down.

We will consider what happens if the production oscillation changes after the gene populations have reached stable oscillation at $t=3*2\pi$. First we will let the oscillation speed to twice the normal rate for one day, from $t=3*2\pi$ to $t=4*2\pi$. See Figure 25.

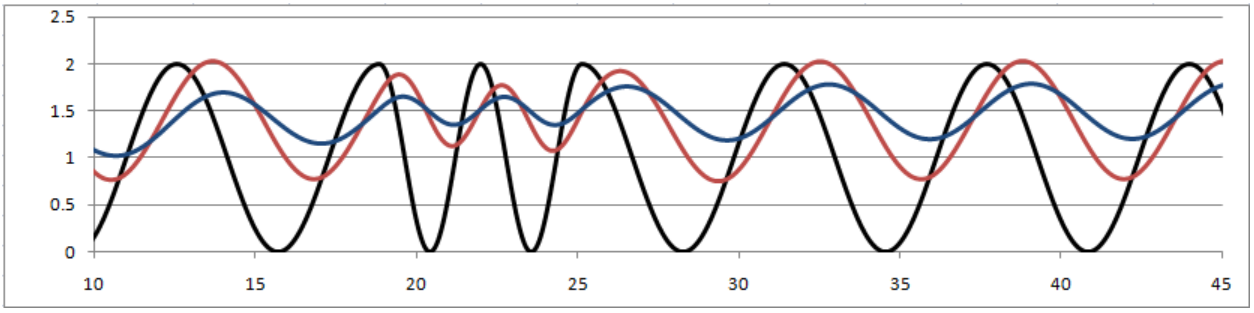


**Figure 25: Shocked Model 2.1: Graph of $n_1$ (blue) and $n_2$ (red) when the oscillation of r(t) (black) increases for a day.**

This plot shows that the faster oscillation of the global production rate causes the oscillation of the population curves to increase for that day, and the amplitude of oscillation decreases. After the global production rate returns to normal, the behavior of the population curves quickly returns to normal.

Now, we will consider how a slower oscillation of the global production rate will affect the gene populations. See Figure 26.

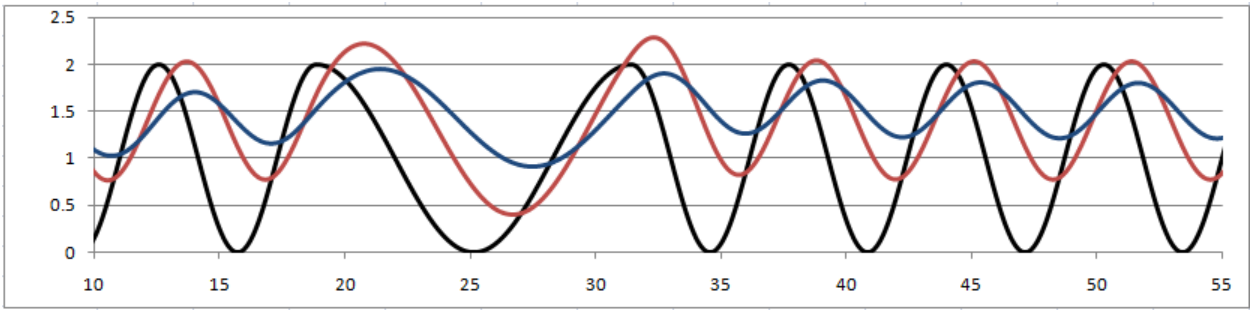


**Figure 26: Shocked Model 2.1: Graph of $n_1$ (blue) and $n_2$ (red) when the oscillation of r(t) (black) decreases for 2 days.**

This plot shows that the slower oscillation of the global production rate causes the oscillation of the population curves to decrease for that day, and the amplitude of oscillation increases. After the global production rate returns to normal, the behavior of the population curves quickly returns to normal.

Another anomaly the global production rate may display is that it may stop production for a day. See Figure 27.

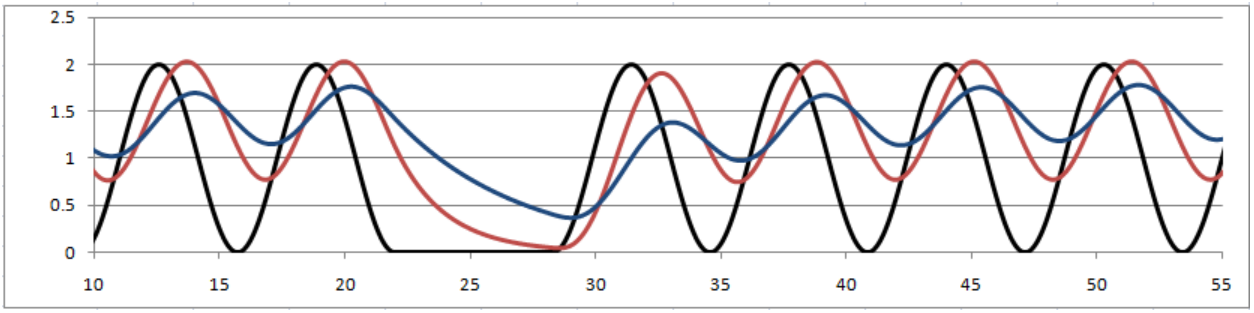


**Figure 27: Shocked Model 2.1: Graph of $n_1$ (blue) and $n_2$ (red) when r(t) (black) is zero for a day.**

This plot shows that when global production stops for a day, the gene populations decay. After the global production returns to normal, the population curves quickly resume their normal behavior.

Now, we will assume the SOC3 gene is produced at the maximum production rate for a day. See Figure 28.

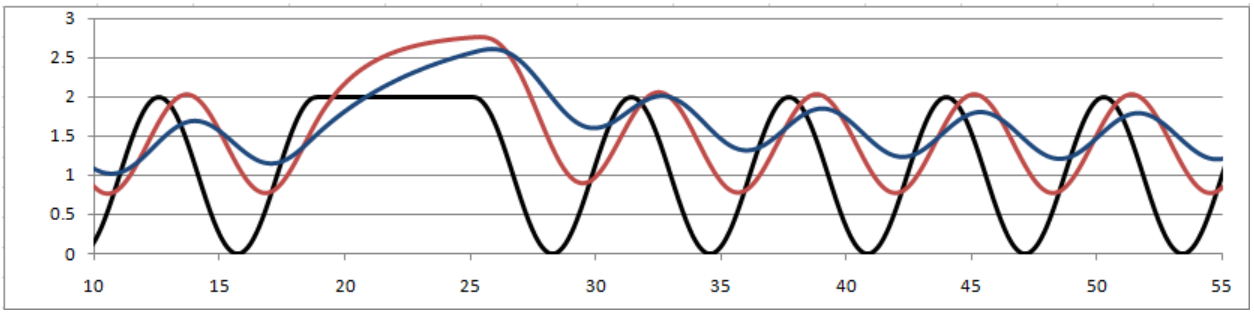


**Figure 28: Shocked Model 2.1: Graph of $n_1$ (blue) and $n_2$ (red) when r(t) (black) is two for one day.**

This plot shows that when the production rate is maximized for a day, the gene populations increase toward steady states. When the global production returns to normal, the population curves quickly resume their normal behavior.

## VI. Conclusions

These models were developed to try to create a coherent picture of the populations of two transcripts of the SOC3 gene. The data to which we would like to compare these models is highly granular, with only 13 data points over the course of two days. Because of this granularity, we cannot distinguish between most of the models presented. The behavior of the global transcription rate is periodic, but it may be a simple on-off switch or a smooth oscillation. We will need to check these models against less granular data to determine which picture fits best.

This research project started with an assumption that the global population of the SOC3 gene was constant, and none of the models developed were able to reflect this behavior while using reasonable parameters. Every model appeared to support a different hypothesis – the different transcription populations were oscillating inside a global population which also oscillated. We can conclude that the granularity of the data, which is taken every 4 hours over the course of two days, was masking an oscillatory behavior in the global population, and that Model 1.0, the fixed lifespan model, Model 2.1, the fixed decay rate model, and Model 1.1, which is a hybrid of both models, are good predictors of the population behavior.

We also shocked Model 2.1 in several different ways, but found that the normal gene population behavior persisted. The gene abundances in the model quickly returned to behavior the populations exhibited before the system shock.

## Acknowledgements

I would like to thank Jake Blanton, for allowing me to use his software and for bringing me up to speed on the project. I would like to thank Andre Ptitsyn for many helpful and insightful conversations and for hosting my trip to do this work at the Colorado State University in Fort Collins.

## Works Referenced

Friedman, Jeffrey M and Halaas, Jeffrey L, Leptin and the regulation of body weight in mammals, *Nature*. **395** (1998), 763-770.

Gao, Qian; Wolfgang, Michael J.; et al., Disruption of neural signal transducer and activator of transcription 3 causes obesity, diabetes, infertility, and thermal dysregulation, *Proceedings of the National Academy of Sciences of the United States of America*. **101.13** (2004), 4661-4666.

Gimble, J. M. and Ptitsyn, A. A. Analysis of circadian pattern reveals tissue-specific alternative transcription in leptin signaling pathway, *BMC Bioinforamtics*. **8**(Suppl 7) (2007): S15.