

MODELS OF LIFE AND THE LIFE OF MODELS

Stephen J. Merrill
Department of MSCS
Program in Computational Sciences
Marquette University
Milwaukee, WI

What is a “model”?

- Fashion model– a person who is employed to display, advertise and promote commercial products (notably fashion clothing)
- Civil Engineering – scale model to test design aspects of a structure.
- Biology – animal model which mimics important aspects of the human in disease or drug response.
- Mathematics – mathematical models employing gross simplifications but hopefully retaining important aspects. “First principles” builds the model.
- Simulation models – models nearing the complexity of the system being studied. Used as a way to describe the data in this system. Data validates the model.
- Statistical models – model descriptions created from regression (for instance). Data builds the model.
- Hybrid models – combinations of the above. e.g. a mouse with a human immune system or a simulation model with some mathematical models embedded.

And more ...

- Physical model – a physical representation of an object often for visualization.
- 3D modelling – a 3D polygonal representation of an object, usually displayed with a computer, e.g. FEM.
- Model building – a hobby centered around the construction of material replicas, usually scale models.
- Solid modelling – study of very accurate representations of the solid parts of an object, as in CAD.
- Conceptual Models – analogies used to aid understanding (my definition).
- Role models – an example for others
- Crash Test Dummy – see above
- Flight simulator – to provide repetitions and situations to increase experience.

And many more

Some Common Themes for Models

- A real system, or collection of real things that are of interest.

For instance, it may be “all bridges that could be built using a particular design”

- Specific use (purpose for which it was constructed). The nature of the specific use varies with the model.
- The “model” represents idealized and simplified version of the real system. The level of simplification is dictated by the type of model.
- The model allows uses not possible in the real system.

For instance, testing of a device under different failure modes.

- A test (sometimes heuristic, usually data driven) to determine if the model is sufficient at the current stage of development. If not, “improvements” are made and the test reapplied. This could be more than a validation.
- An understanding that the model is not the system – the nature of its differences and what difficulties those differences may cause.

The model is often confused with the real system in models that have been used for a long time.

Models of Life

- Many of the types of models are used to describe living systems.
- Each type of model has its strengths and weaknesses.

For instance, a simulation model is at nearly the complexity of the system it was designed to describe (which may have been too complicated to understand).

- Whether the model is a “good” one depends on the intended use. If one can address the questions that motivated the model – it is a good model. Note that it need not reproduce data perfectly as a criteria (unless that is a requirement).
- Limitations of data place real limits on the complexity of a model. Data gaps can force too many arbitrary decisions or parameter values to make the results believable for a complex model.
- The type of model that is appropriate for a given situation is often determined by the system under study and the nature of the questions – not the “bias of training.”

Other aspects

- Legacy models – created in the past, found useful, and continue to be used. Often associated with large projects.

Often associated with confusing the system with the model.
e.g. chemistry as described in CHEM001 where they are actually teaching a model of the real interactions.

- Single use (disposable) models – created to answer a specific question – it did its job, and now is no longer used.

An example would be competing models created to see which one gave best fit.

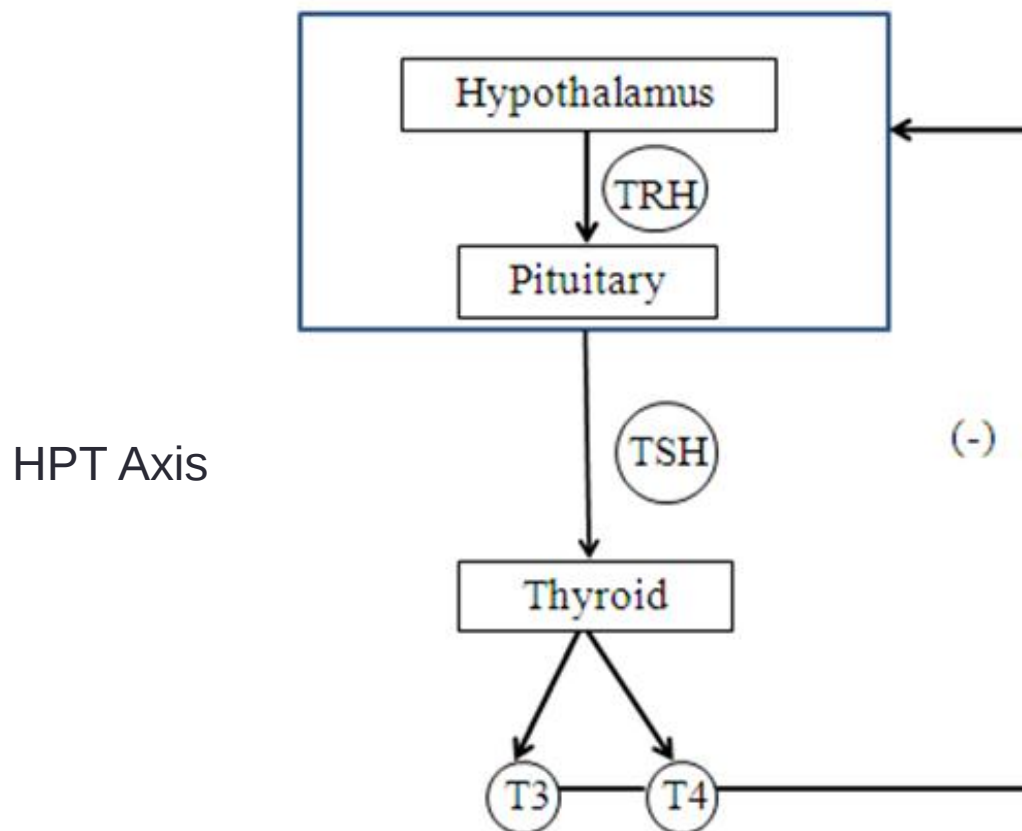
- Deterministic model – every simulation with the same parameters and initial conditions should result in the same result.
- Stochastic model – randomness, usually in addition to deterministic aspects, plays a role in the results.

A good example is autoregressive (AR) time series models.

Some Examples

- Clinical course in autoimmune thyroiditis (Hashimoto's)
Limited data, individual differences, interesting question
- Heart rate variability classification using Markov Chains
Method to identify patterns and changes in patterns (like stats model)
- Early stages in HIV infection
Stochastic model – identify key parameters
- Dynamics of engraftment in hematopoietic stem cell transplants
Data in the context of a model – discovery of dynamical surprise
- Spatial variation in cDNA microarray
simple description – complicated result

Hashimoto's – autoimmune Destruction of the thyroid



The Question

- Can we determine if the patient will eventually develop chronic hypothyroidism?

(If so, when do we start treatment to minimize effects of the disease)

The Data

- 119 patients with autoimmune antibody in Sicily. Each patient has 2-7 measurements of TSH and free T4 at irregular intervals over years.
- Although there are models of the HPT axis, none exist for this situation where the response of the thyroid is disrupted.

Possible Responses

- Impossible – need more data
- Impossible – individual differences reduce the usefulness of even the little data available
- Maybe a simple model can tell us something

$$\frac{dFT4}{dt} = \frac{k_3 T TSH}{k_d + TSH} - k_4 FT4$$

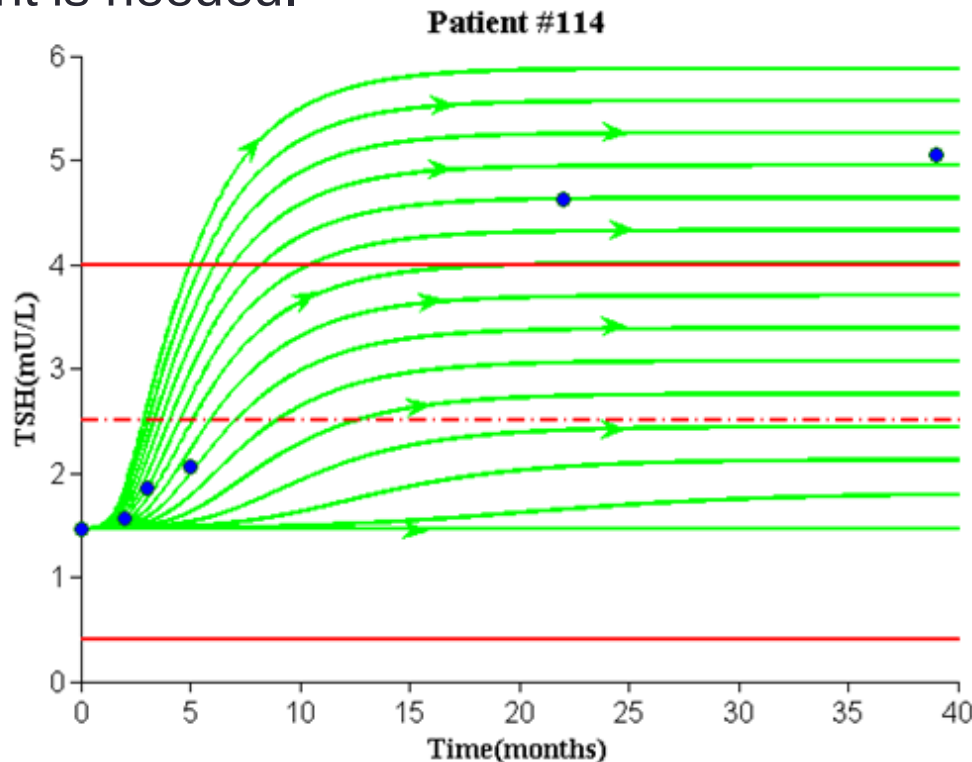
$$\frac{dTSH}{dt} = k_1 - \frac{k_1 FT4}{k_a + FT4} - k_2 TSH$$

$$\frac{dT}{dt} = k_5 \left(\frac{TSH}{T} - N \right) - k_6 Ab T$$

$$\frac{dAb}{dt} = k_7 Ab T - k_8 Ab$$

Results

1. Dynamics are simple – depending on only a few parameters
2. For each patient, an approach to the equilibrium can be determined
3. The position of the equilibrium will determine if (and when) treatment is needed.



HRV using Markov chains

- Heart Rate Variability (HRV) describes the beat-to-beat variation in the time interval between beats as seen on ECG (R-R interval).
- It is described by many different indices.
- The variability is due to several different control mechanisms in the systems
- Operation of the controls are affected by drugs (specifically here, anesthesia)

The Question

- Design a real-time monitor to detect a patient heading toward sudden cardiac arrest

The Data

- pediatric patients undergoing surgery
- Patient 29
2.5 years old
- Patient 55
7.5 years old
Early with halothane
Late with atropine

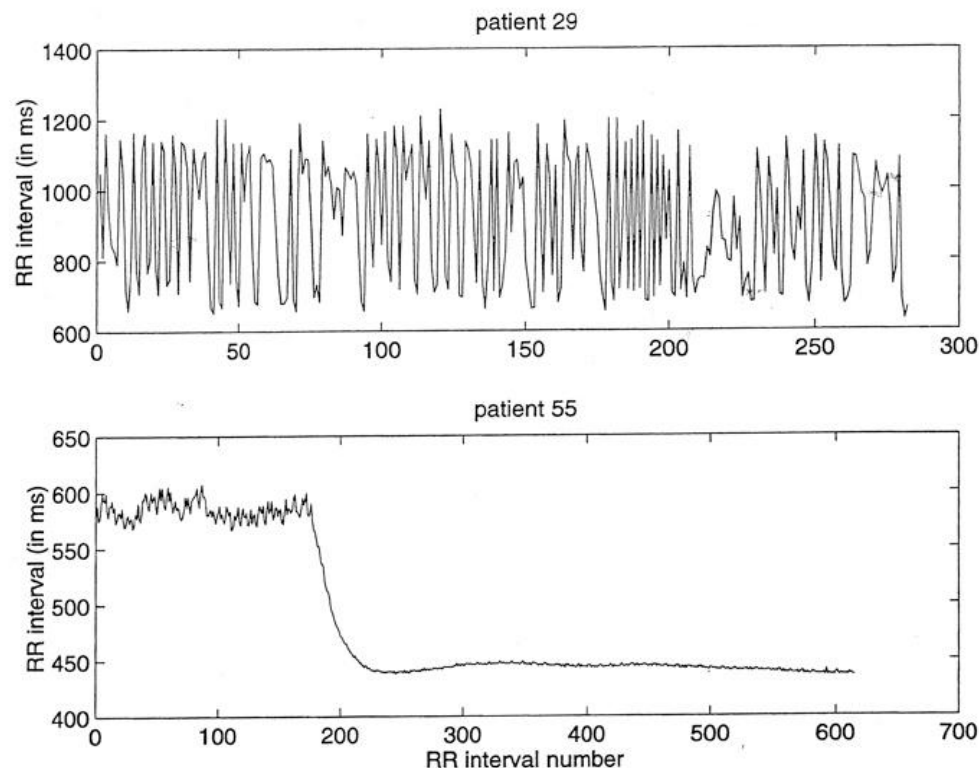


Figure 1 RR interval data from Patient 29 and Patient 55. In the analysis, the second set will be divided into an early and late phase.

Lag 1 maps

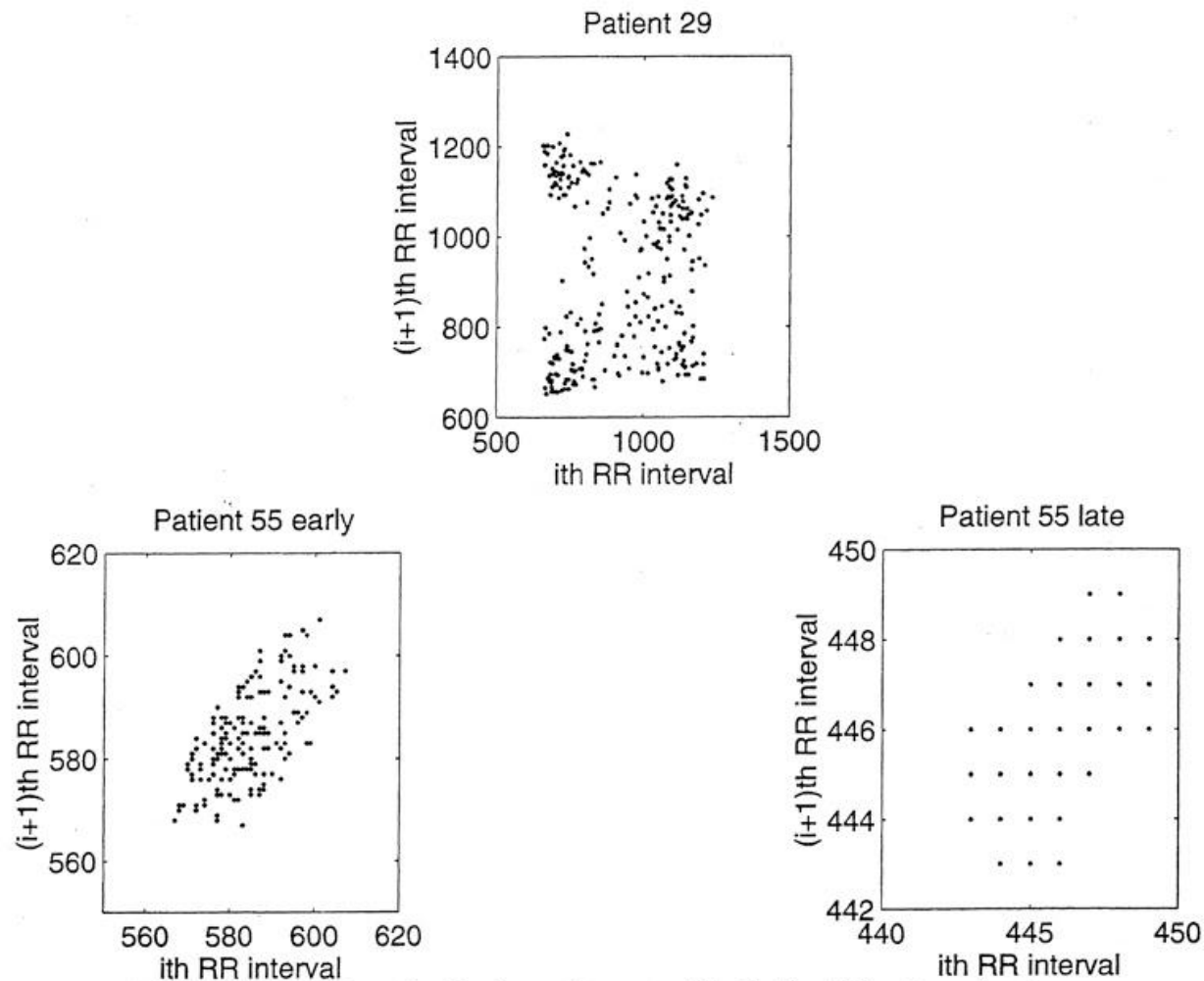


Figure 3 Lag 1 plots for the three data sets. The Patient 29 pattern has been described previously as a “complex pattern” (Woo et al. [1992]). Patient 55 early and late data plots would be classified as “torpedo patterns” by Woo et al. [1992].

Model of R-R interval data

- Create an empirical Markov chain. Data is in the form of sequence of numbers and lag 1 maps indicate first order structure.
- Need to define the bin size corresponding to the length of the data set (usual number of bins used was 10). Then estimating transition probabilities to get a transition matrix. Note that many possible transitions are not observed.
- Transient aspects of the chain are of interest (not asymptotic behavior). Characterization of the dynamics (or the resulting matrix) is desired.
- Basic idea is to use properties of the matrix (such as eigenvalues) to distinguish between cases.

Eigenvalue maps

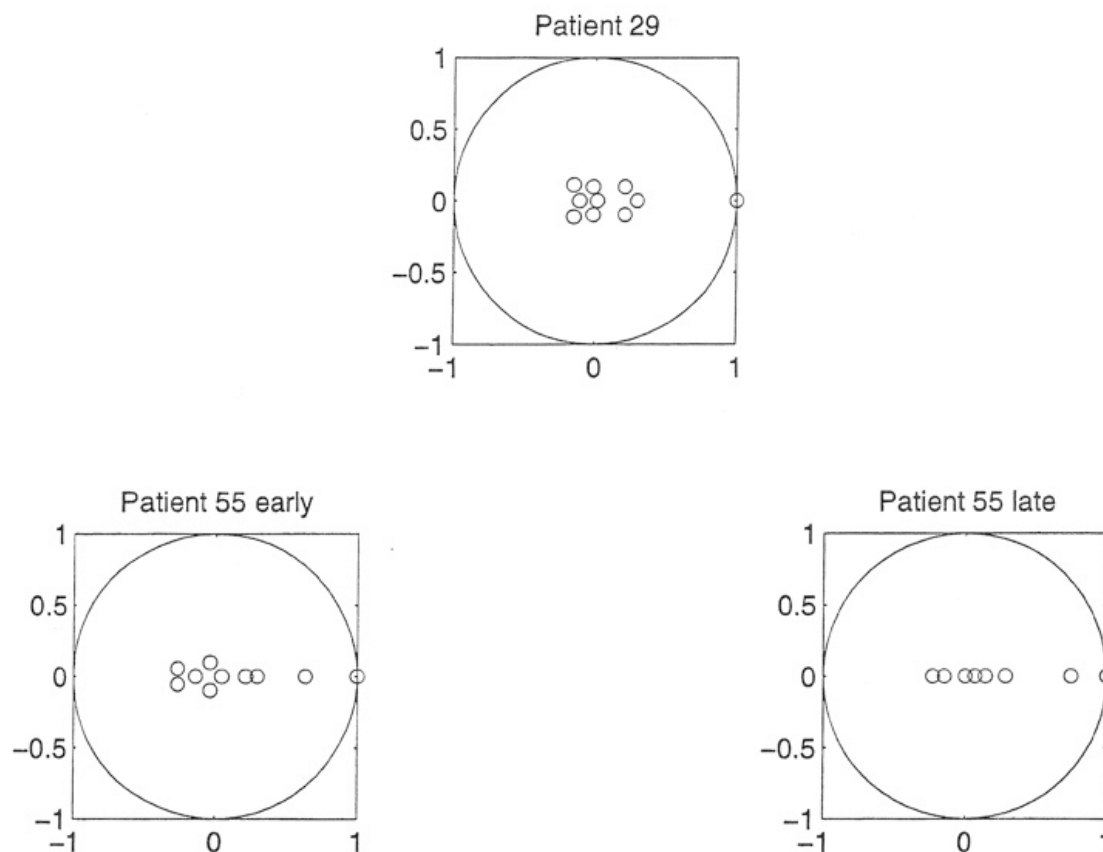


Figure 5 Eigenvalues for each of the three transition matrices are shown in relation to the unit circle. Besides the nearness of the “non-1” eigenvalues to the unit circle, notice the differences in the number of complex eigenvalues.

Early HIV infection

- Long term time-course of the infection depends on the “set point” -- related to the state of the infection at the time the immune response controls the initial acute infection.
- Interested in computing the incubation-time distribution (defined as the time from infection to a fixed clinical marker such as seroconversion -- the appearance of anti-HIV antibodies).

The Question

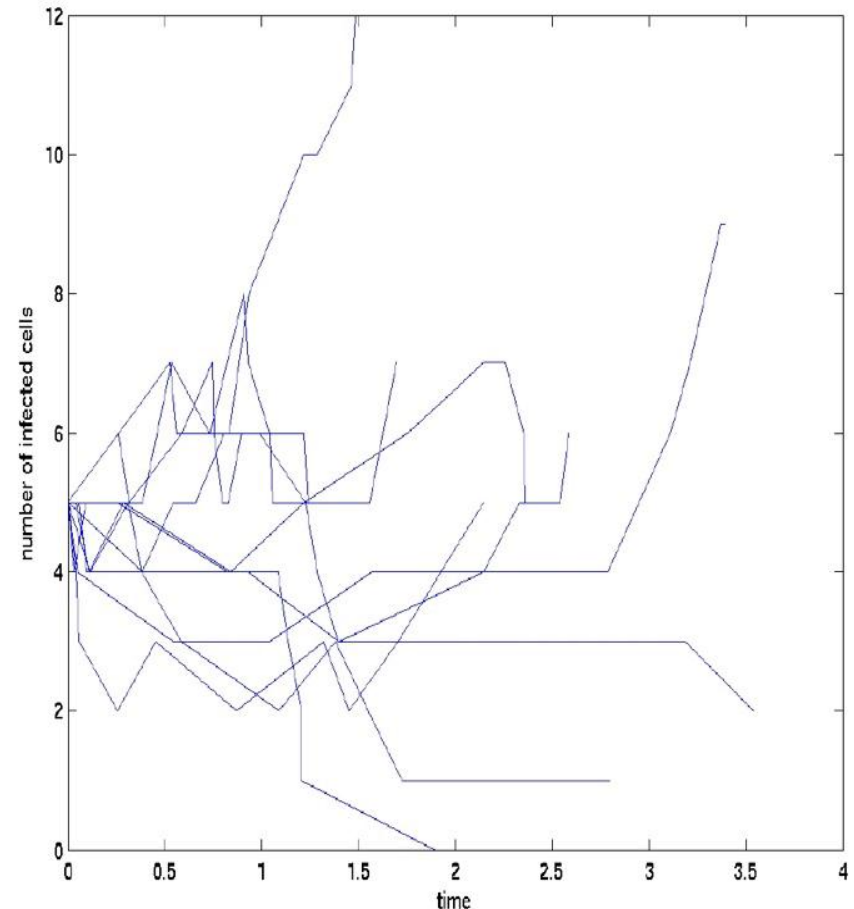
- What is the nature of this distribution (for instance its mean) and what are the critical parameters.

The Data

- Existing statistical descriptions of the distribution built from large numbers of patients (censoring is a problem here)

The Model

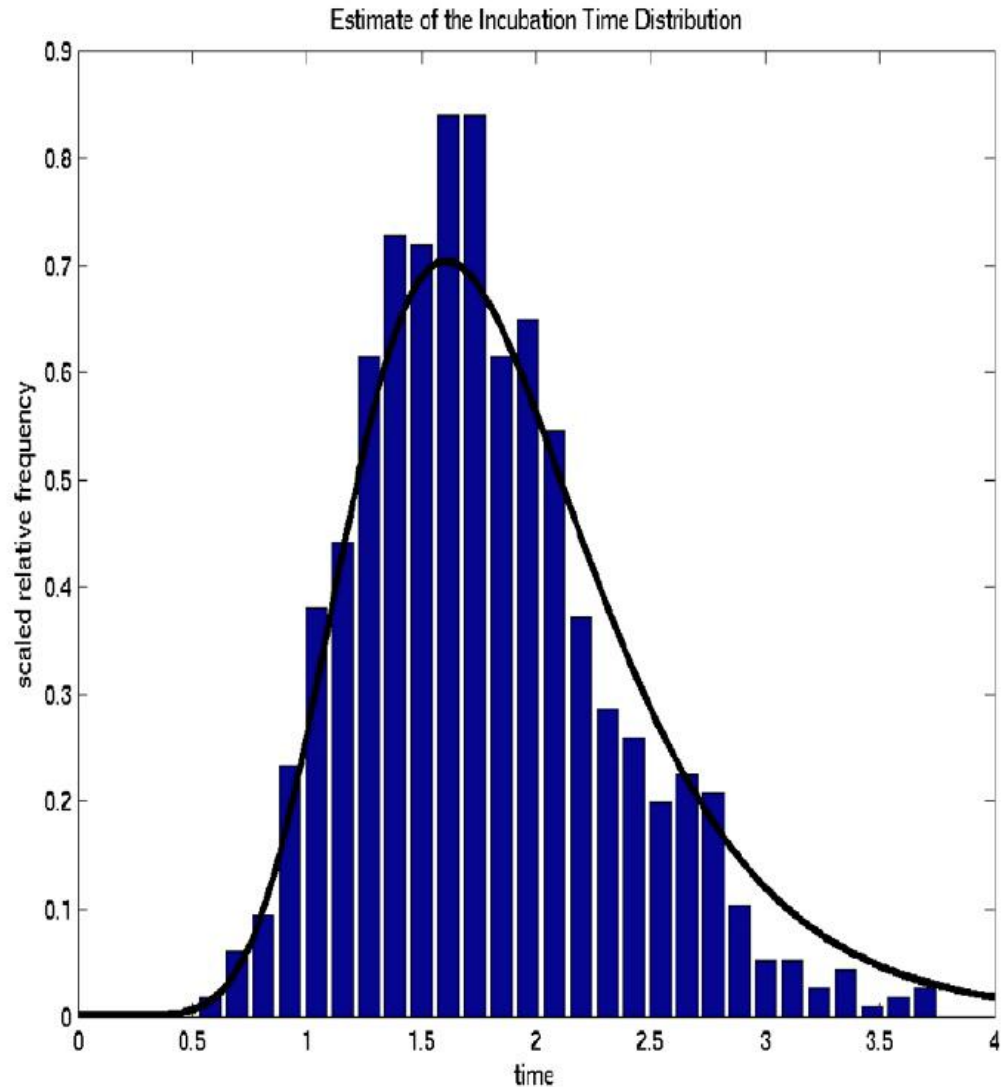
- Branching process with immigration. The basic quantity tracked is the number of infected T cells.
- Only data that exists is the distributional information
- Model designed to see what happens in the initial stages.



Sample Paths of branching model -- no immigration

Simulations

1000 simulations
of model,
recording the
time at which the
number of
infected cells
reach some fixed
Value (stopping time
distribution)



Dynamics of Engraftment

- Hematopoietic stem cells can be collected from blood (or bone marrow) for later infusion (transplantation) after high-dose chemotherapy.
- In autologous transplants, no rejection is present.
- Interested in monitoring engraftment (return to normal levels) of each cell type – primarily leukocytes (WBC in early counts), lymphocytes, platelets, and red cells.

The Question

- From daily blood counts, estimate “time to engraftment” and detect possible problems before they occur.

The Data

- Daily counts from 32 women following transplantation

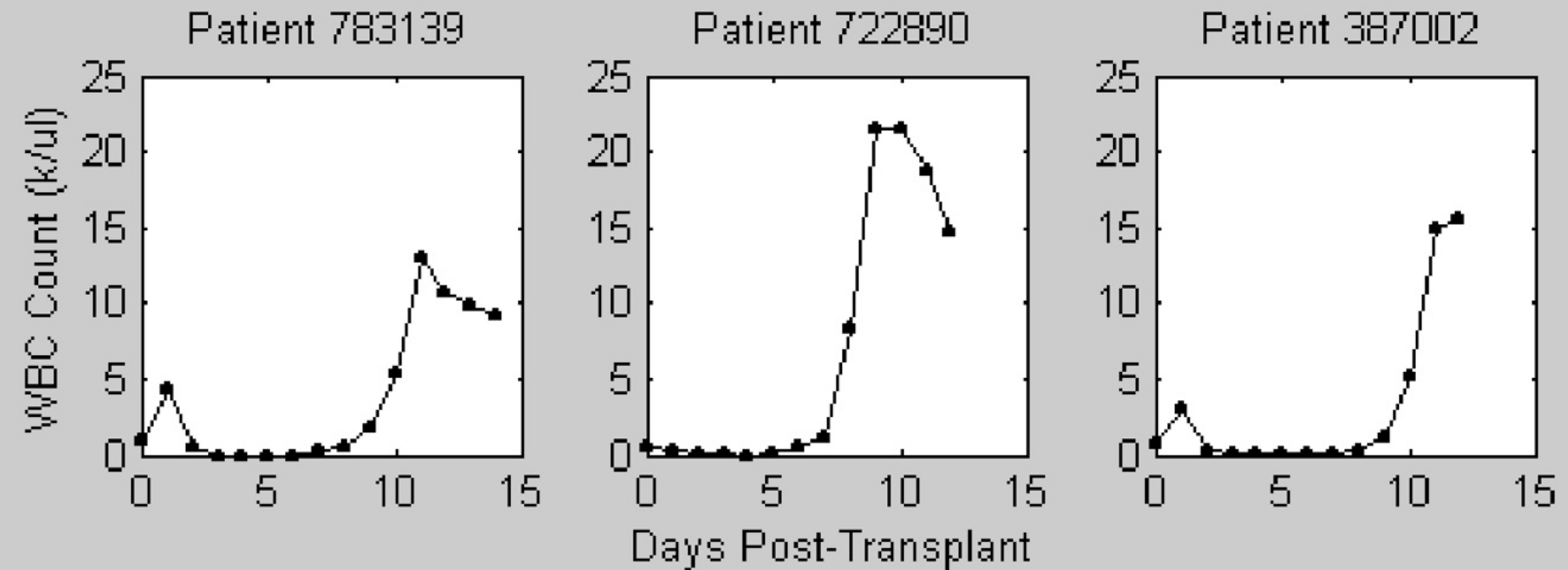
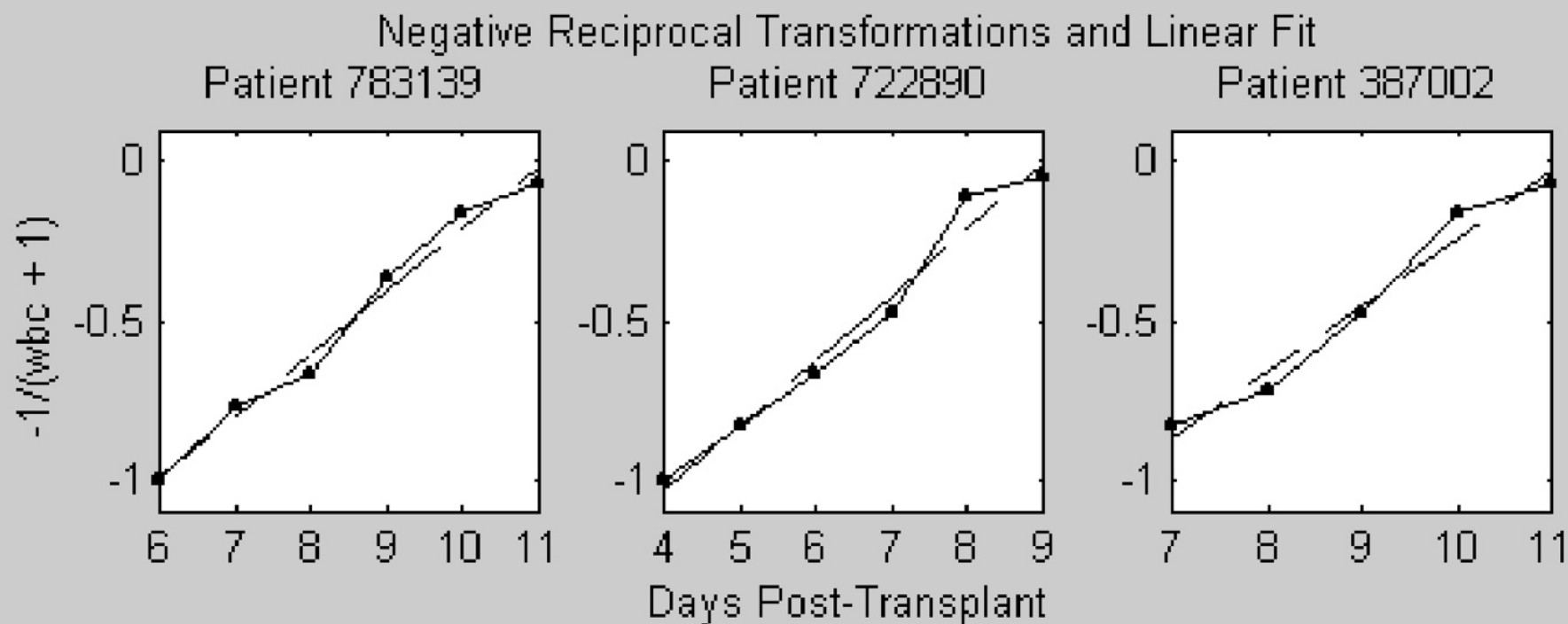


Figure 3.1. Typical WBC plots

The Model

- Reciprocal plot shows hyperbolic growth $r^2 = .94$

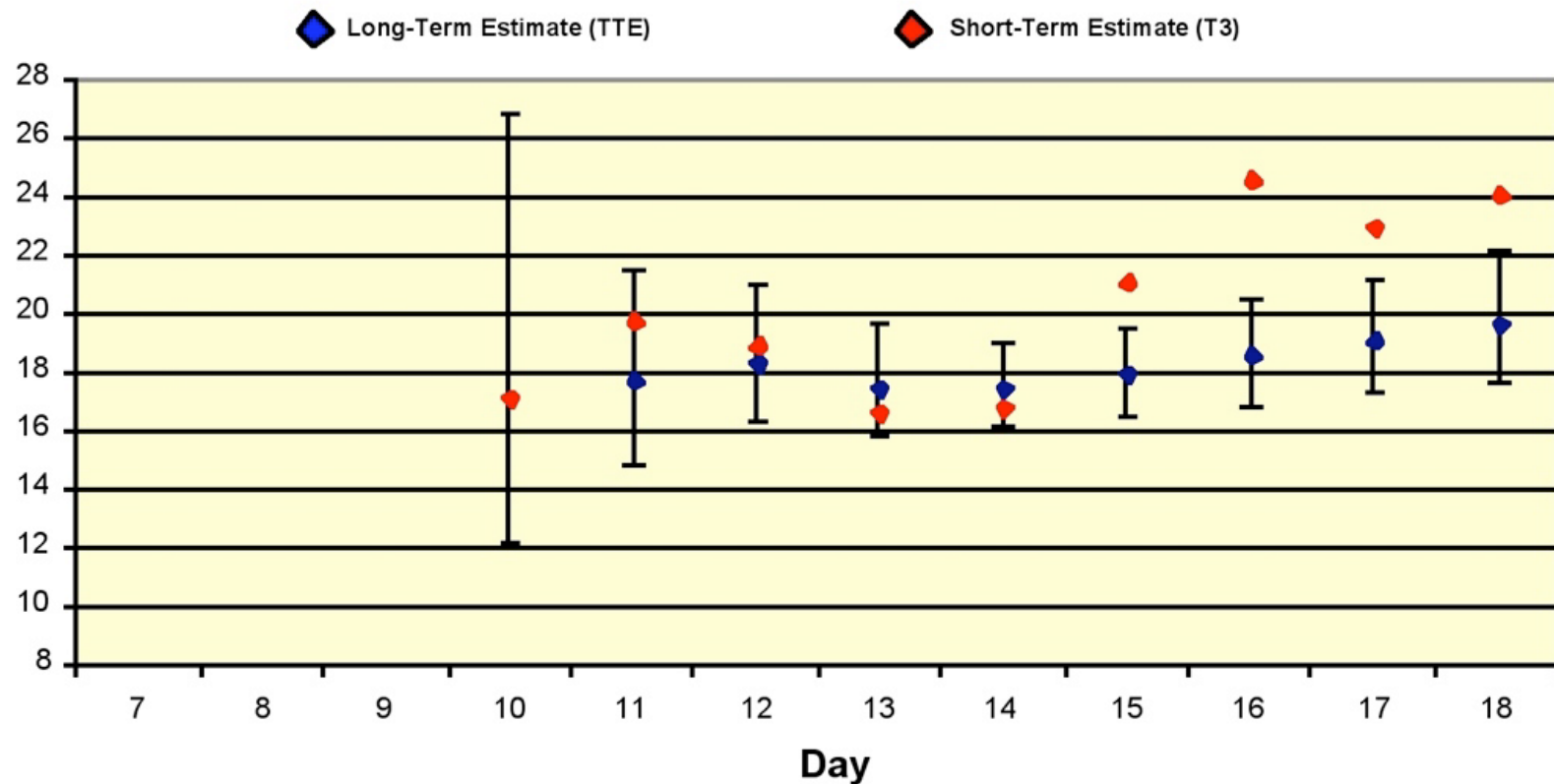


Results

- Estimating the position of the asymptote (and as it changes with each days data) allows the estimation of the time to engraftment -- and resulting release from the hospital.
- Changes in the estimates indicate problems, represented in a control chart.
- Similar results for lymphocytes. For platelets, polynomial growth was observed.

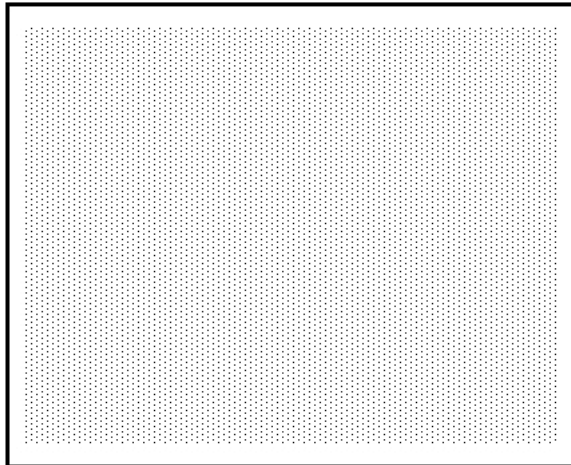
Final Result – a Control Chart

Patient 177 -- Engraftment Control Chart (80% C.I.)



Spatial variation in the microarray

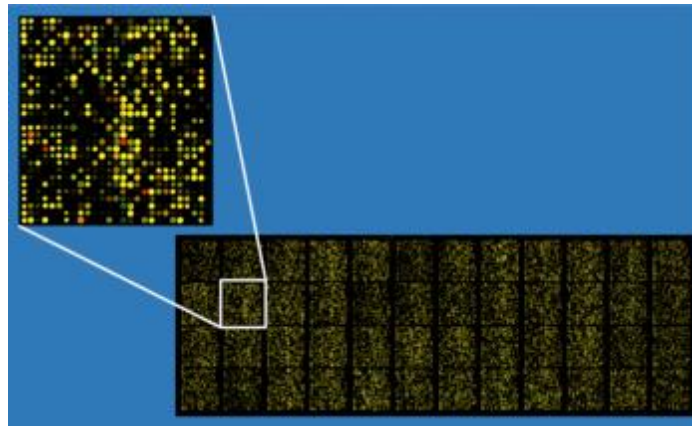
- cDNA microarray used to identify genes that are differentially under or over expressed in a sample (as seen through mRNA).



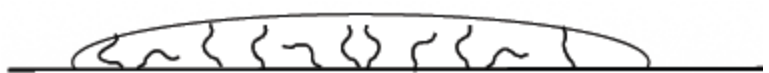
The Question

- Why are my edges bright? Or what “normalization” is the right one to use for this process.
- What is the source of the variability – how should replicates be done (and what statistics should be used)

The Data

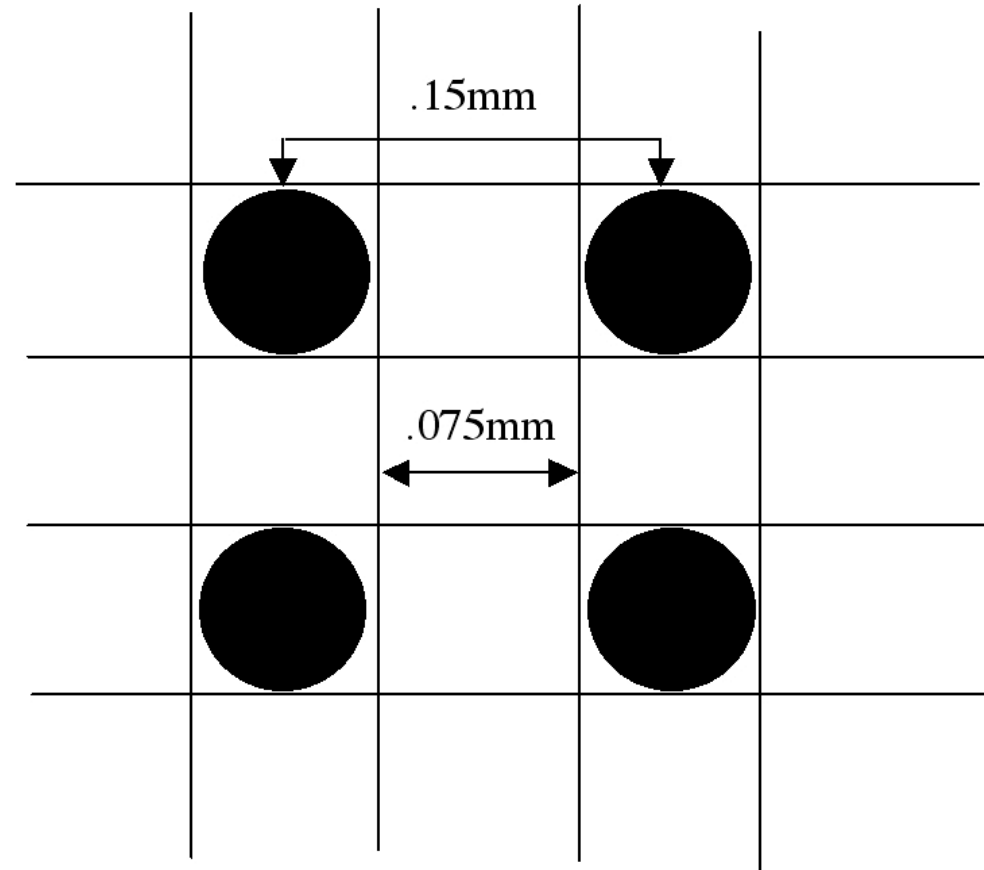


A bit of the process – one version

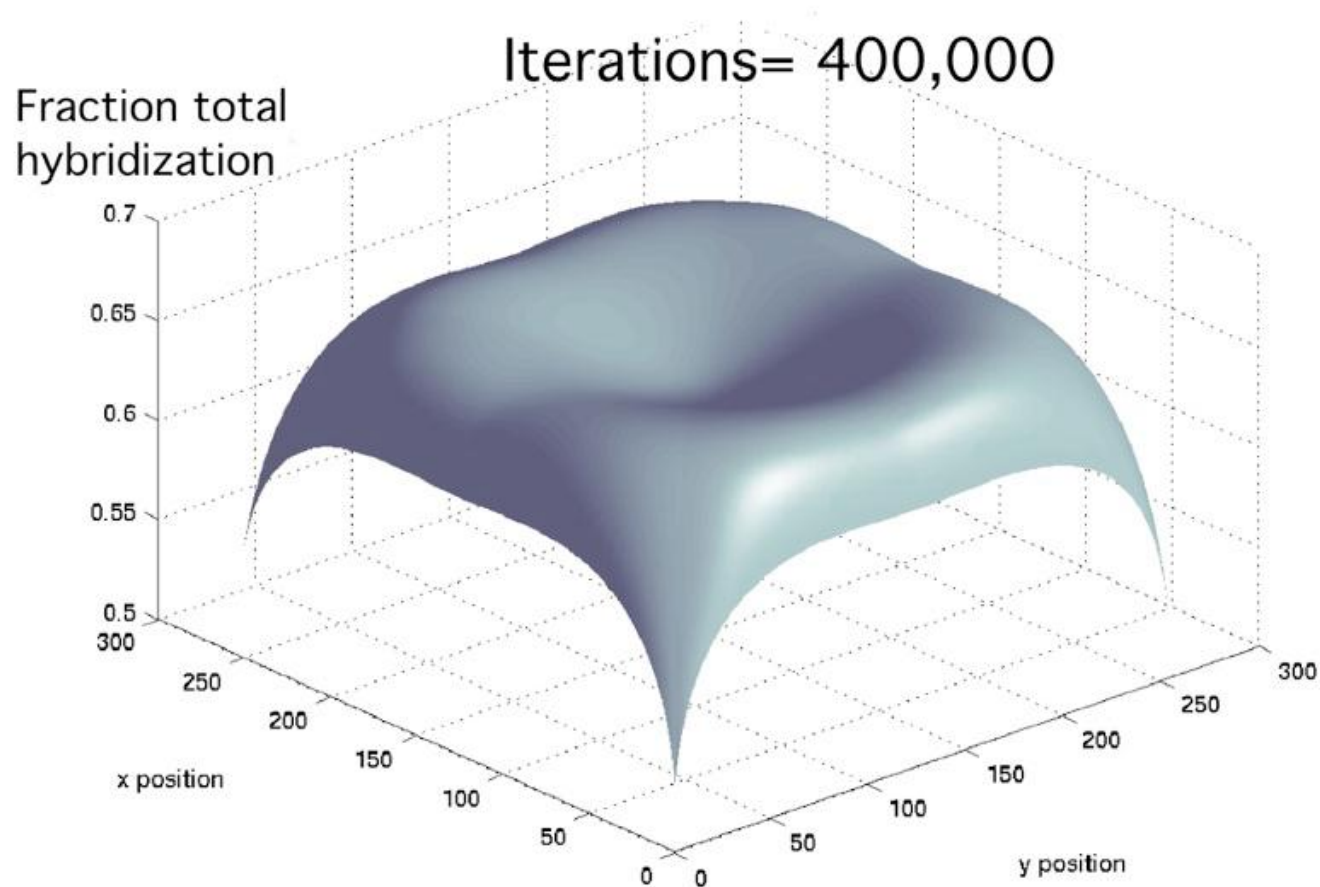
- the slide or chip is printed with a library of genes including those of special interest 
- collect mRNA under two different conditions. Using RT and two different fluorescent dyes, samples of labeled (“red” and “green”) DNA are produced.
- incubate the samples with the slide under a cover slip.
- scan the result to measure the amount of red and green fluorescence at each spot to measure the relative amount of mRNA present in the two samples.

The Model of microarray hybridization

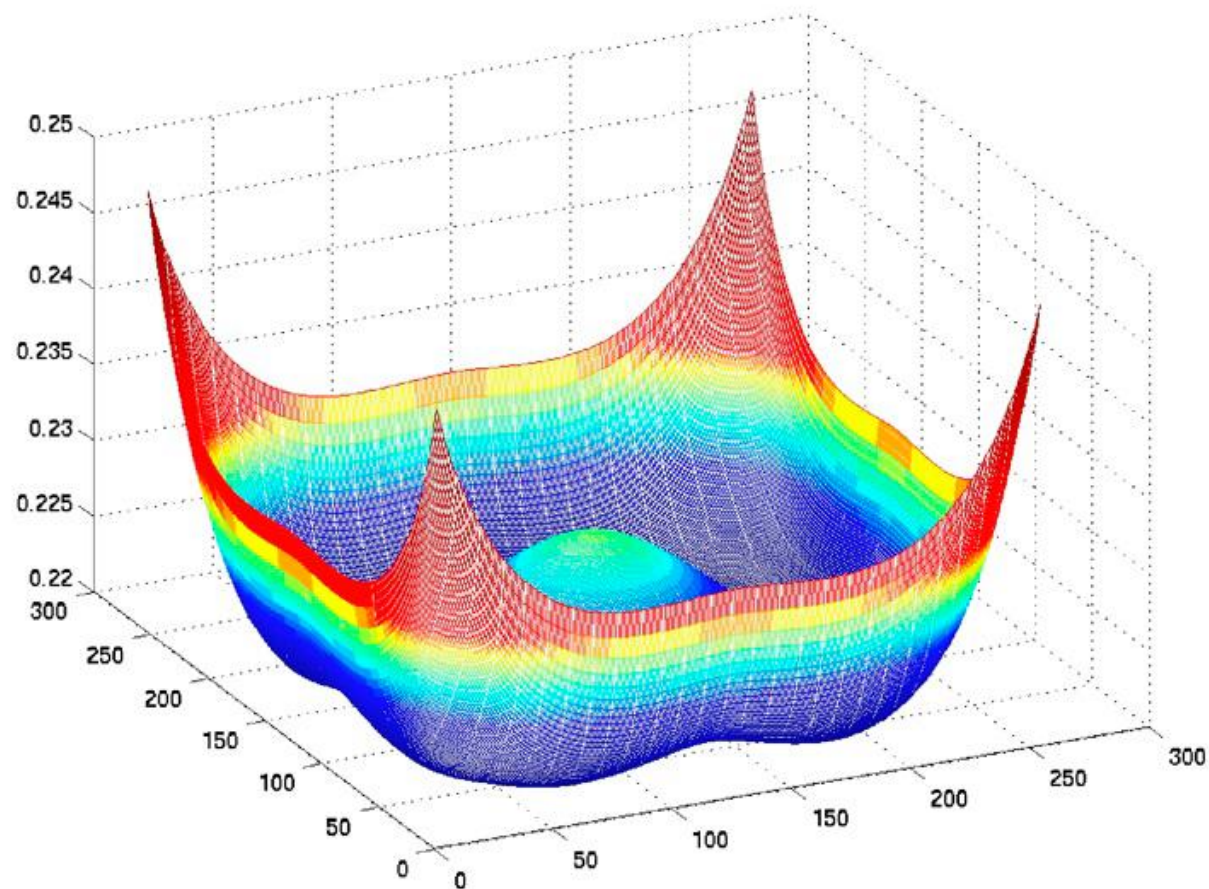
- Using the natural grid of positions on a slide, a Markov corresponding to each of the 16,000 dots is constructed. The goal being to compute the probability of absorption as a function of the transition number. Assume all dots are same.
- The transition probabilities are based on the “taxi-cab” metric on the grid.



12 hours of hybridization



Variance after 12 hours



Conclusions

- The idea of “model” is complicated.
- Important aspects:
 1. Definition of the Real System (scope of the model) – what does it cover and what does it not cover.
 2. Specific use – questions that the model is designed to address
 3. Tests – validation and refinement
 4. If the model is to be used over a period of time, software design principles should be used.
- Questions and nature of the system dictate the model form
- It is often a great surprise that (simple) models tell us anything about the real system
- The process of modeling (and solving problems in general) is a complicated one.

References

- S.J. Merrill & J.R. Cochran (1997) Markov chain methods in the analysis of heart rate variability, *Fields Institute Comm.* **11**:241-252.
- S.J. Merrill & B.M. Murphy (2002) Detection of autocatalytic dynamics in data modeled by a compartmental model, *Math. Biosci.* **180**:255-262.
- S.J. Merrill, S. Nelson, & C.A. Struble (2003) Spatial dependence of hybridization in the cDNA microarray, *Can. Appl. Math. Quart.* **11**: 321-337.
- S.J. Merrill (2005) The stochastic dance of early HIV infection, *J. Comp. Appl. Math.* **184**:242-257.
- S.J. Merrill (2007) To again feel the creative voice, *Int. J. of Sci. and Math. Ed.*, **5**:145-164.
- B. Pandiyan, S.J. Merrill, & S. Benvenga (2012) A patient-specific model of the negative feedback control of the hypothalamus-pituitary-thyroid (HPT) axis in autoimmune (Hashimoto's) thyroiditis, submitted.