Playing Newton

Ilya Nemenman

Departments of Physics and Biology
Emory University

JAMES S.
MCDONNELL
FOUNDATION

nemenmanlab.org

EMORY
UNIVERSITY

Thanks

- Bryan Daniels (Arizona State)
- Andre Levchenko (Yale)
- Will Ryu (Toronto)

Why?
(paraphrasing Richard Hamming)

1. What are the important problems in your field?

1. What are the important problems in your field?
2. What important problems are you working on?
3. Why are the answers to (1) and (2) different?

So:

What are the important problems in theoretical biophysics?

Of exactitude in science

...In that Empire, the craft of Cartography attained such Perfection that the Map of a Single province covered the space of an entire City, and the Map of the Empire itself an entire Province. In the course of Time, these Extensive maps were found somehow wanting, and so the College of Cartographers evolved a Map of the Empire that was of the same Scale as the Empire and that coincided with it point for point. Less attentive to the Study of Cartography, succeeding Generations came to judge a map of such Magnitude cumbersome, and, not without Irreverence, they abandoned it to the Rigours of sun and Rain. In the western Deserts, tattered Fragments of the Map are still to be found, Sheltering an occasional Beast or beggar; in the whole Nation, no other relic is left of the Discipline of Geography.

From Travels of Praiseworthy Men (1658) by J. A. Suarez Miranda (a fictional reference).
By Jorge Luis Borges and Adolfo Bioy Casares.
English translation quoted from J. L. Borges, A Universal History of Infamy,
Penguin Books, London, 1975.

At a recent meeting...

- A common opinion: "**The final** theory of biological systems will be a large multiscale computational model. We need more and more experimental data to specify details of these models."

openworm

A simulation platform to build digital in silico living systems -- starting with a c. elegans worm virtual organism simulation

- There's something wrong with this statement.
  - The "**final**" theory?
  - Do we need the theory of "**everything**" in any biological (or physical) system?

- The best material model of a cat is another, or preferably the same, cat.
  (*Philosophy of Science*, Wiener and Rosenblueth, 1945)

Physics analogy

- What is the final, complete theory of the chair you are sitting in?
  - How does it fall from the second floor?
  - How does the cloth seat age and tear?
  - How much weight would the chair hold before it breaks?
  - How does it conduct electricity?
  - How much food can I cook when I burn it?
  - …
- There's no such thing as "the full theory of the chair".
  - We build models tailored to answer **specific questions**.
  - The complete theory that answers **every** question would need to include quarks, superstrings…
  - Each modeling level needs its own **effective** degrees of freedom
    - "Don't model bulldozers with quarks." (Goldenfeld and Kadanoff, 1999)
- Models must loose details. Otherwise, just use the same cat…

So...

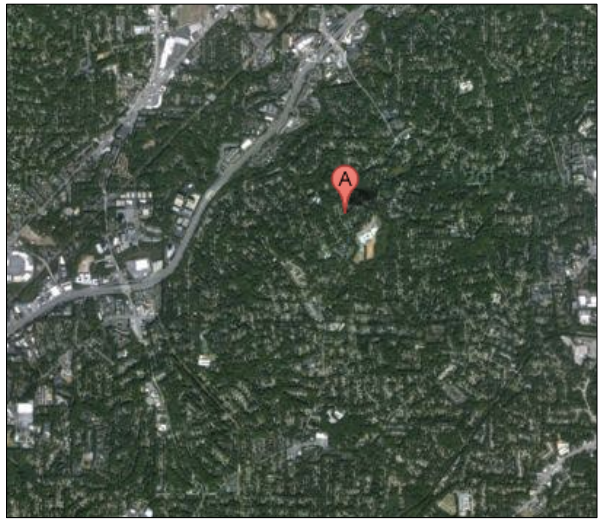
- **Are there *phenomenological, coarse-grained*, and yet functionally accurate representations of (some) biological dynamics, or are we forever doomed to every detail mattering?**
  - And, of course, these models would not answer *every* question, but specific questions on coarse scales.
    - E.g., not What is a position of this particular atom in the cell? But What is the whole system doing?
  - (Parenthetically): Without such coarse-grained descriptions, if everything is equally important, modern biology ceases to be a Western science.

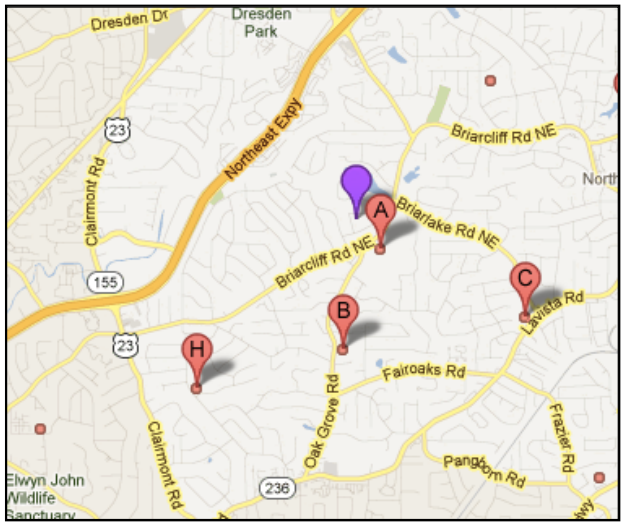
What do I mean by this?

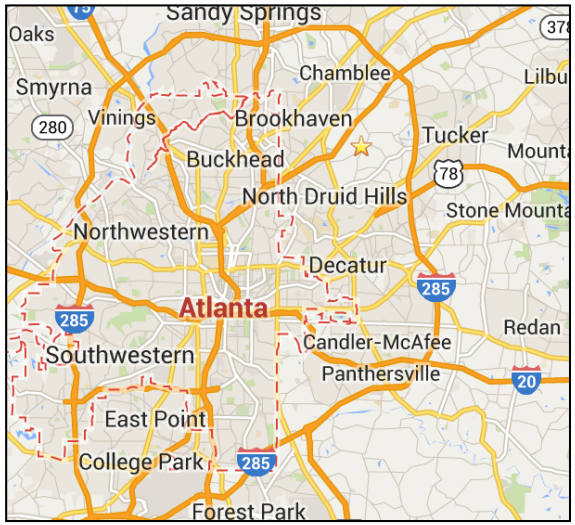
- Western tradition:
  - There are laws (of nature, of god, whatever). A rock is a rock everywhere. It falls the same in Pisa and in Atlanta.
  - There're causes and there are effects.
  - There is "useless information" (Oscar Wilde).
  - But this belief requires closing one's eyes to minor discrepancies
    - Two balls dropped from the Leaning Tower didn't actually land simultaneously.
    - "If we had the STM in the 1920s, there wouldn't be the Debye theory of solids." (H. Levine)
- Non-western tradition, e.g., buddhism
  - Pratityasamutpada: dependent origination: "*Pratitya samutpada* is sometimes called the teaching of cause and effect, but that can be misleading, because we usually think of cause and effect as separate entities, with cause always preceding effect, and one cause leading to one effect. According to the teaching of Interdependent Co-Arising, cause and effect co-arise (*samutpada*) and everything is a result of multiple causes and conditions... " — Thich Nhat Hanh

Where would new laws come from?

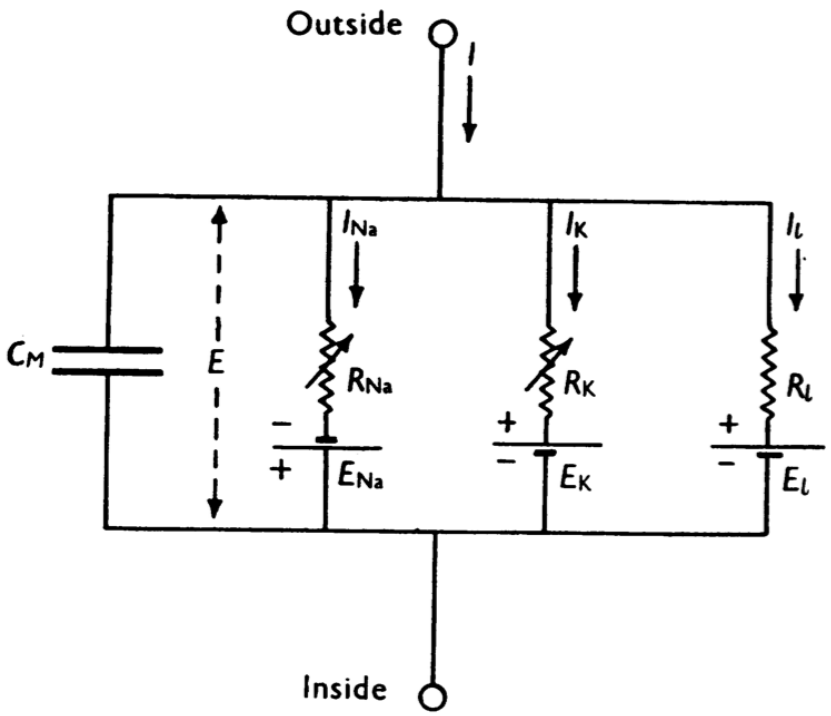
- **More is different!** (PW Anderson)
  - The **law of large numbers** produces universalities if the right questions are asked (e.g., about *large-scale* quantities).
- Important to use the right level of description.
  - Coarse-graining: Each modeling level needs its own **effective** degrees of freedom.
  - "Don't model bulldozers with quarks." (Goldenfeld and Kadanoff, 1999)
- This is already common in your **every-day life,** not just physics
  - Which level of description is **better for** driving to a local school?







A side note: This has worked before! Hodgkin-Huxley



$$I = C_M \frac{dV}{dt} + I_i, \quad (1)$$

$$I_{Na} = g_{Na}(V - V_{Na}), \quad (3)$$

$$I_K = g_K(V - V_K), \quad (4)$$

$$I_l = \bar{g}_l(V - V_l), \quad (5)$$

$$g_K = \bar{g}_K n^4, \quad (6)$$

$$\frac{dn}{dt} = \alpha_n(1 - n) - \beta_n n, \quad (7)$$

$$g_{Na} = m^3 h \bar{g}_{Na}, \quad (14)$$

$$\frac{dm}{dt} = \alpha_m(1 - m) - \beta_m m, \quad (15)$$

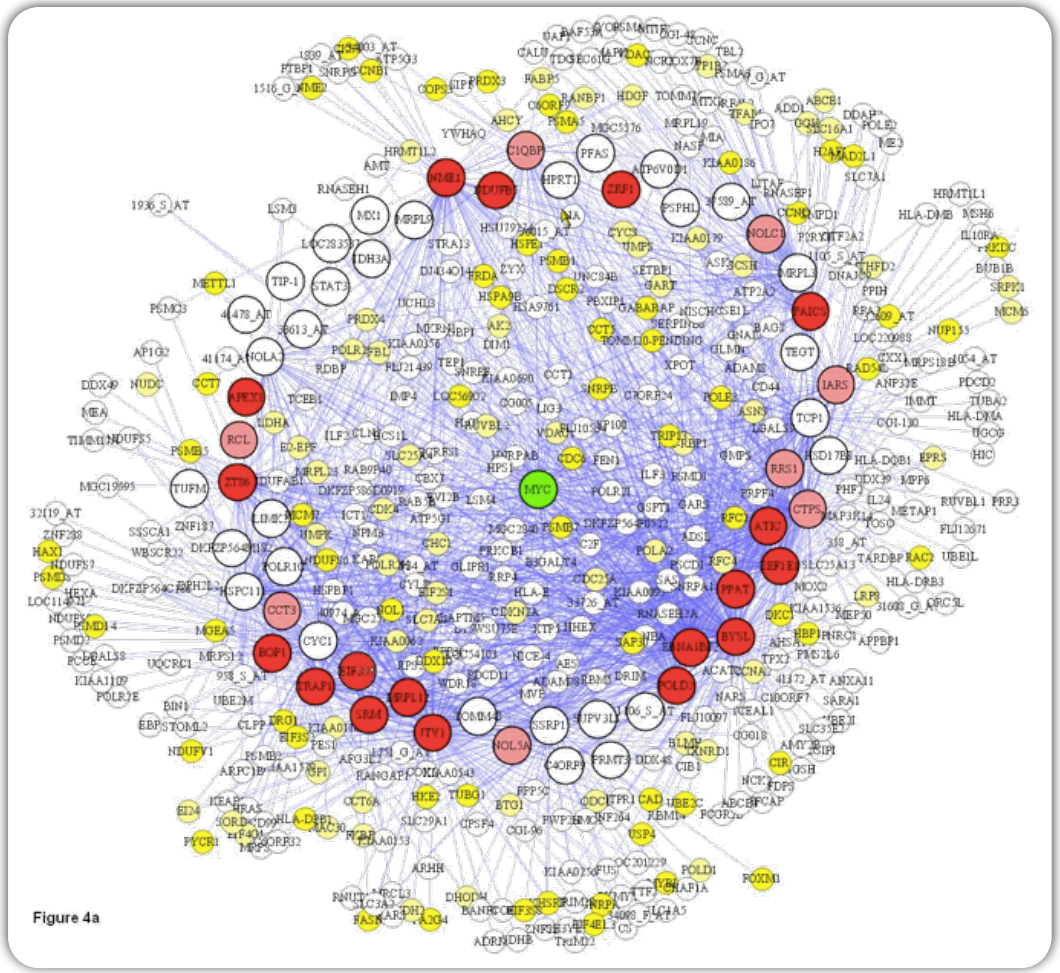
$$\frac{dh}{dt} = \alpha_h(1 - h) - \beta_h h, \quad (16)$$

- A good theory! (and only roughly correct)

Cellular networks: complex beasts
Large *N* is baked in

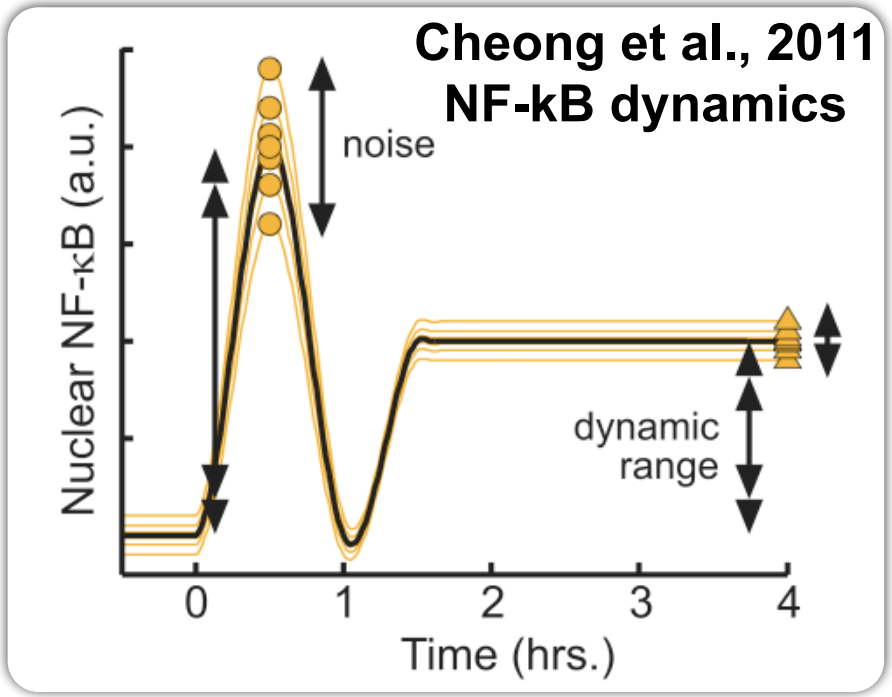
A culture's icons are a window onto its soul. Few would disagree that, in the culture of molecular biology that dominated much of the life sciences for the last third of the 20th century, the dominant icon was the double helix. In the present, post-modern, 'systems biology' era, however, it is, arguably, the hairball.

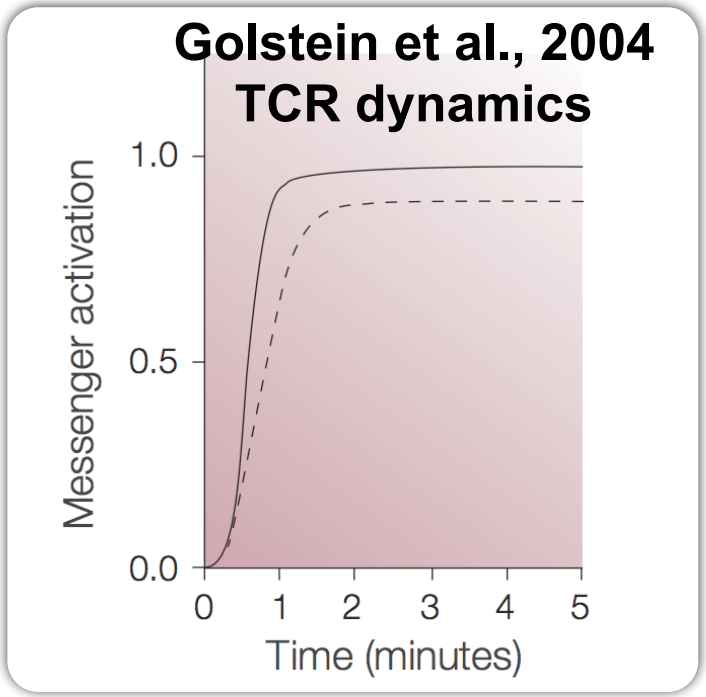
A.D. Lander. BMC Biology 2010, 8:40



Margolin et al., 2006

And yet, for typical inputs, their dynamics are rather simple





- A handful of parameters (time scales, amplitudes) describe responses of networks to most experimentally accessible perturbations.
- Relation of phenomenological to mechanistic parameters often unclear.
- **Do we need complex networks to describe simple dynamics?**

The field (and we) has worked on this for a while

- Bottom-up methods, reducing a known microscopic, mechanistic network.
  - with Sinitsyn et al., 2006-2010; with Munsky, Bel et al.,2009-2013; with Merchan, 2016.
  - Schwab and Mehta 2015
  - Large number of publications in chemical physics (Petzold et al.)
  - **Problem:** need to know and start with the microscopic model.
- Can we instead build **phenomenological models** top-down, from data directly, and without reconstructing a mechanistic network as an intermediate step?
  - Purpose: **predict responses** to exogenous signals.
  - Purpose: drive all of us, theorists, out of work?

First steps…

- Assume that dynamics of cellular networks is given by local **ordinary differential equations**.
  - Do not fit curves; **fit dynamics**.
- Neglect stochasticity, and spatial structure for now

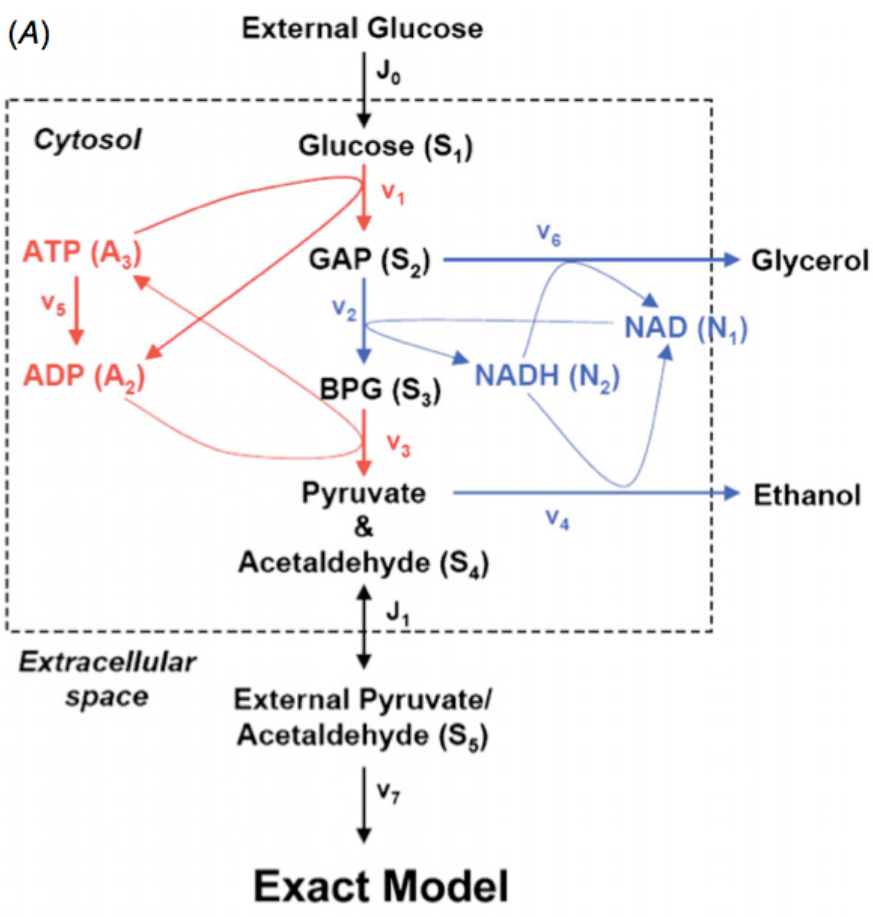
$$\begin{cases} \frac{dx_1}{dt} = f_1(x_1, x_2, \ldots, x_n) \\ \ldots \\ \frac{dx_n}{dt} = f_n(x_1, x_2, \ldots, x_n) \end{cases}$$

- Can we automatically fit these functions f_i from data?
  - How do we enumerate the set of all possible multivariate functions?
  - How do we search through this list?
  - How do we not overfit?

Prior (and posterior) art

- The full search approach for an exact model
  - Small systems dynamics — search for all possible models using S-systems formalism **(Voit et al, Theor Biol Med Model 2006).**
  - Searching for a control model from a (small) set of *a priori* allowed models **(Lillacci and Khammash, PLoS CB 2010).**
  - Searching for a stochastic model from a (small) set of *a priori* allowed models **(Munsky, et al., MSB 2009, Science 2013).**
  - **Eureqa:** exhaustive genetic algorithm search through all possible elementary function combinations, with selection of new experiments to optimize discriminability among models **(Lipson et al., Science 2009, Phys Biol 2011).**
  - **SINDy:** Compressed sensing approaches **(Brunton et al., 2016).**
  - **AutomatedStatistician.com:** compositional tree structures for stochastic data.
  - More at the 2017 MM meeting sessions on "Robot-scientist".
- Phenomenological search **(Crutchfield and McNamara, Compl Syst 1987).**
- Problems (limiting the analysis to only a few variables)
  - **data/computing demands** explode with the number of variables;
  - cannot handle **unobserved** variables.

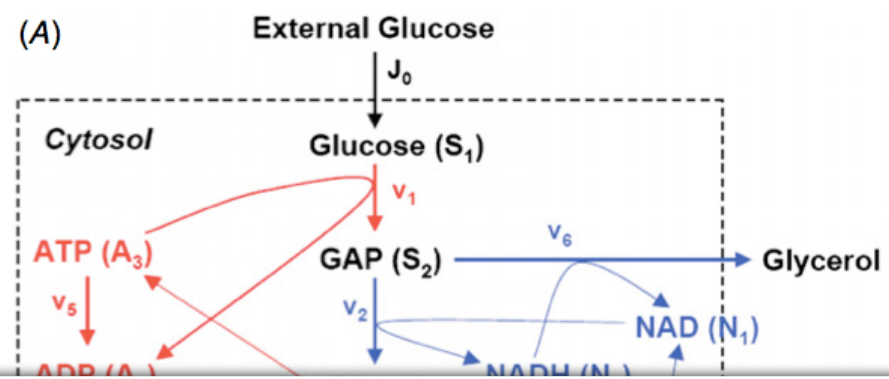
Test Model:
Yeast Glycolytic Oscillator



Ruoff et al., 2003

- 7 species, 28 parameters
- Complex rational dynamical laws

Eureqa performance: Yeast Glycolytic Oscillator



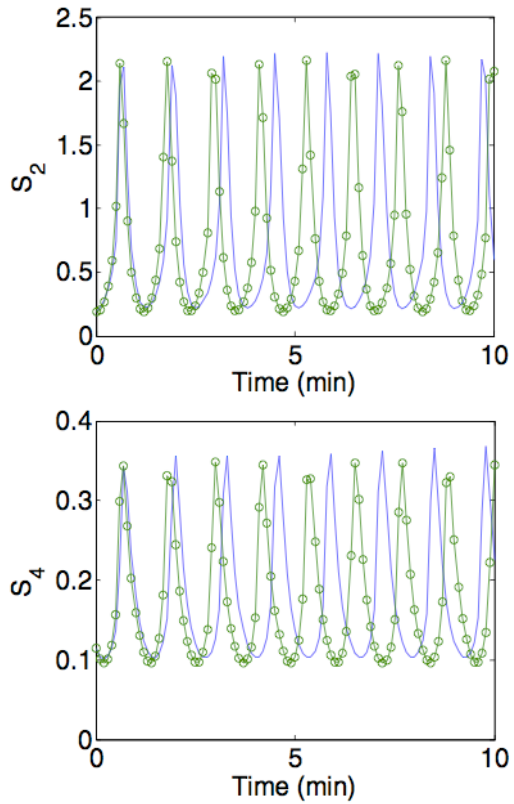
- 7 species, 28 variables
- Complex rational dynamical laws

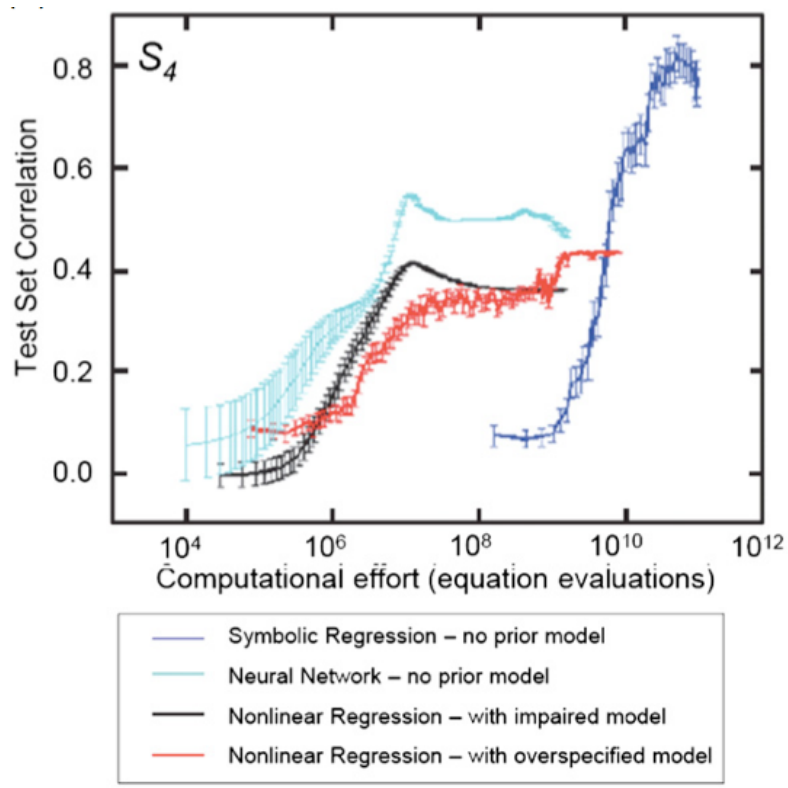
Amazing accuracy!

Original system	Automatically inferred system
$\frac{dS_1}{dt} = 2.5 - \frac{100*A_3S_1}{1+13.68*A_3^4}$	$\frac{dS_1}{dt} = 2.53 - \frac{98.79\cdot A_3S_1}{1+12.66\cdot A_3^4}$
$\frac{dS_2}{dt} = \frac{200*A_3S_1}{1+13.68*A_3^4} - 6*S_2 - 6*S_2N_2$	$\frac{dS_2}{dt} = \frac{200.23\cdot A_3S_1}{1+13.80\cdot A_3^4} - 6.87\cdot S_2 - 6.87\cdot N_2 + 0.95$
$\frac{dS_3}{dt} = 6*S_2 - 6*N_2S_2 - 64*S_3 + 16*A_3S_3$	$\frac{dS_3}{dt} = 6.00\cdot S_2 - 6.00\cdot N_2S_2 - 64.16\cdot S_3 + 16.08\cdot A_3S_3$
$\frac{dS_4}{dt} = 64*S_3 - 16*A_3S_3 - 13*S_4 - 100*N_2S_4 + 13*S_5$	$\frac{dS_4}{dt} = 64.04\cdot S_3 - 16.03\cdot A_3S_3 - 13.03\cdot S_4 - 100.11\cdot N_2S_4 + 13.21\cdot S_5$
$\frac{dN_2}{dt} = 6*S_2 - 18*N_2S_2 - 100*N_2S_4$	$\frac{dN_2}{dt} = -0.055 + 5.99\cdot S_2 - 17.94\cdot N_2S_2 - 98.82\cdot N_2S_4$
$\frac{dA_3}{dt} = -1.28*A_3 - \frac{200*A_3S_1}{1+13.68*A_3^4} + 128*S_3 + 32*A_3S_3$	$\frac{dA_3}{dt} = -1.12\cdot A_3 - \frac{192.24\cdot A_3S_1}{1+12.50\cdot A_3^4} + 124.92\cdot S_3 + 31.69\cdot A_3S_3$
$\frac{dS_5}{dt} = 1.3*S_4 - 3.1*S_5$	$\frac{dS_5}{dt} = 1.23\cdot S_4 - 2.91\cdot S_5$

Schmidt et al., Phys Biol 2011

But at the same time: Need derivatives, and…

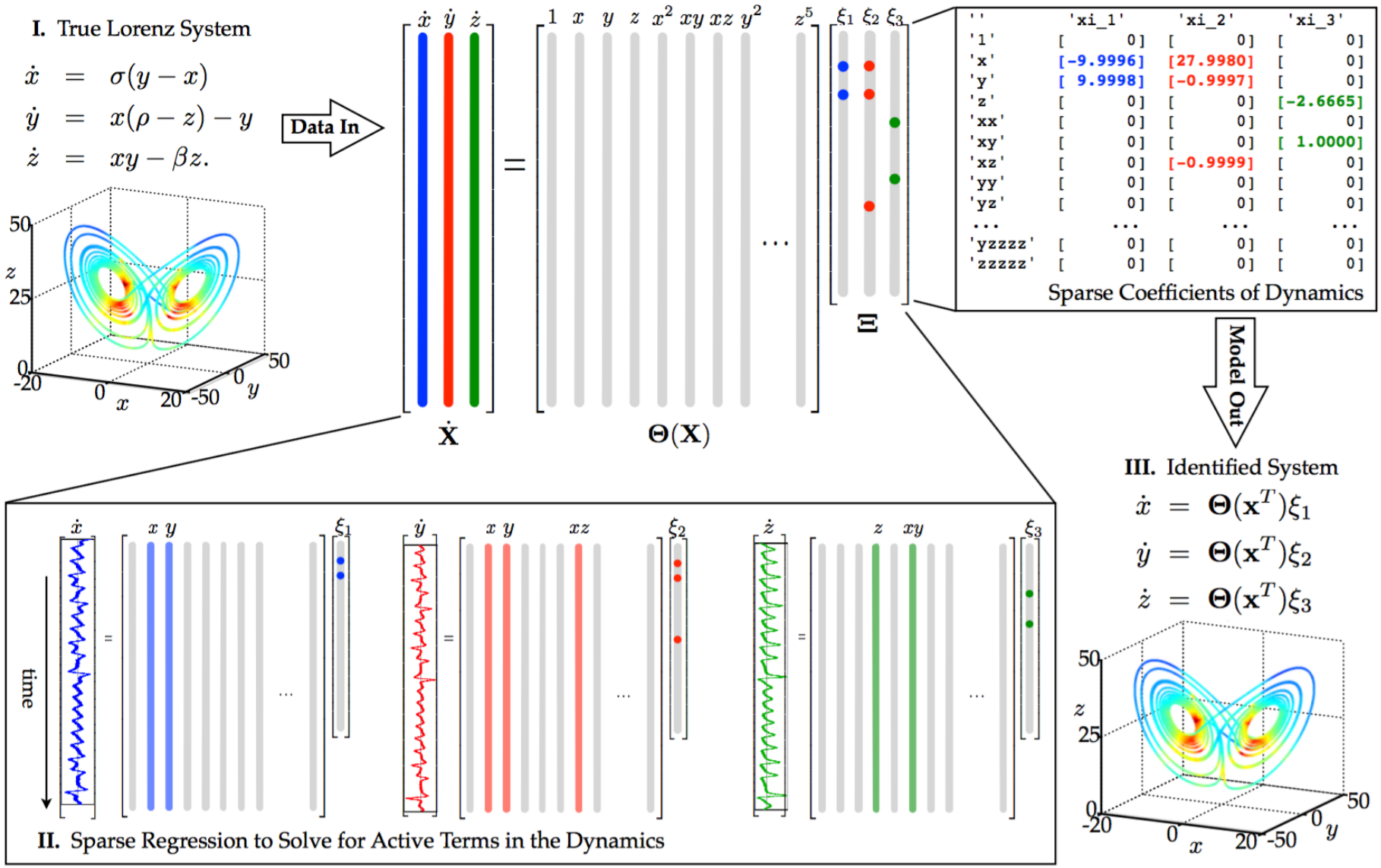




- Astronomical computation times -- **exhaustive search.**
  - **Overfitting** -- need astronomical sample sizes.
- Two exponential costs: **selecting** the best model family, **fitting** the best family with the model.

Schmidt et al., Phys Biol 2011

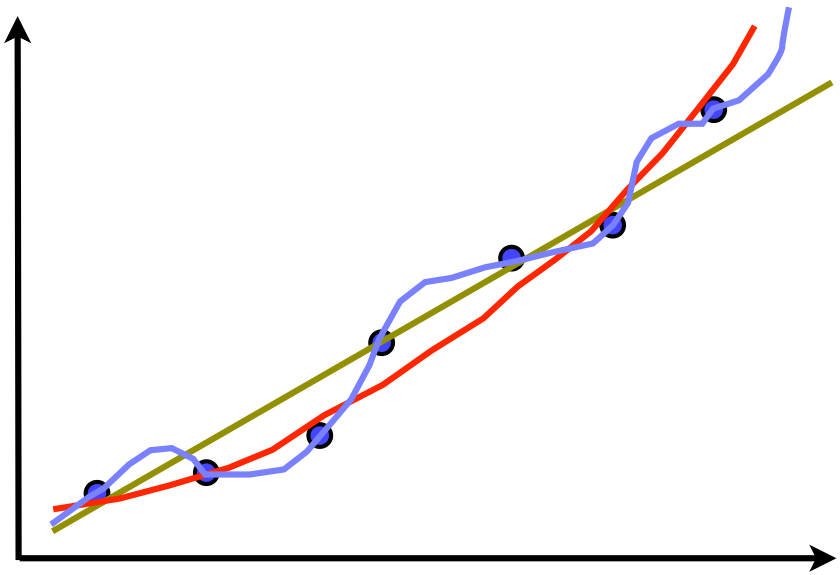
SINDy (Brunton et al., 2016)
(But needs derivatives and correct basis)



''	'xi_1'	'xi_2'	'xi_3'
'1'	[0]	[0]	[0]
'x'	[-9.9996]	[27.9980]	[0]
'y'	[9.9998]	[-0.9997]	[0]
'z'	[0]	[0]	[-2.6665]
'xx'	[0]	[0]	[0]
'xy'	[0]	[0]	[1.0000]
'xz'	[0]	[-0.9999]	[0]
'yy'	[0]	[0]	[0]
'yz'	[0]	[0]	[0]
...	...	...	...
'yzzzz'	[0]	[0]	[0]
'zzzzz'	[0]	[0]	[0]

Can we avoid the exhaustive search and the need for the correct basis?

- Do not need an exhaustive search or exact fits when fitting dimensional curves with progressively increasing complexity



$$y_K(x) = \sum_{k=1}^{K} A_k x^k + \text{noise}$$

  - Use **nested**, **complete** model families, e.g., Taylor series.
  - Use Bayesian model selection to limit the complexity of the search space (the value of maximum *K*).

Schwartz, Ann Stat 1978; MacKay, Neural Comp, 1992
Balasubramanian, Neural Comp1996; Nemenman, Neural Comp, 2005

Bayesian Model selection

- For large sample size *N,* averages done in the Laplace (saddle point) limit.

$$P(K|\{x_i\}) = \int d^K\vec{\alpha} P(\vec{\alpha}|\{x_i\}) = \int d^K\vec{\alpha} \frac{P(\{x_i\}|\vec{\alpha})\mathcal{P}(\alpha)}{P(\{x_i\})}$$
$$= \int d^k\vec{\alpha} \exp(-N\mathcal{L})$$

$$\log P(K|\{x_i\}) = \log P(\{x_i\}|\vec{\alpha}_{\mathrm{ML}}) - \tfrac{1}{2}\log\det N\mathcal{F} + O(N^0)$$

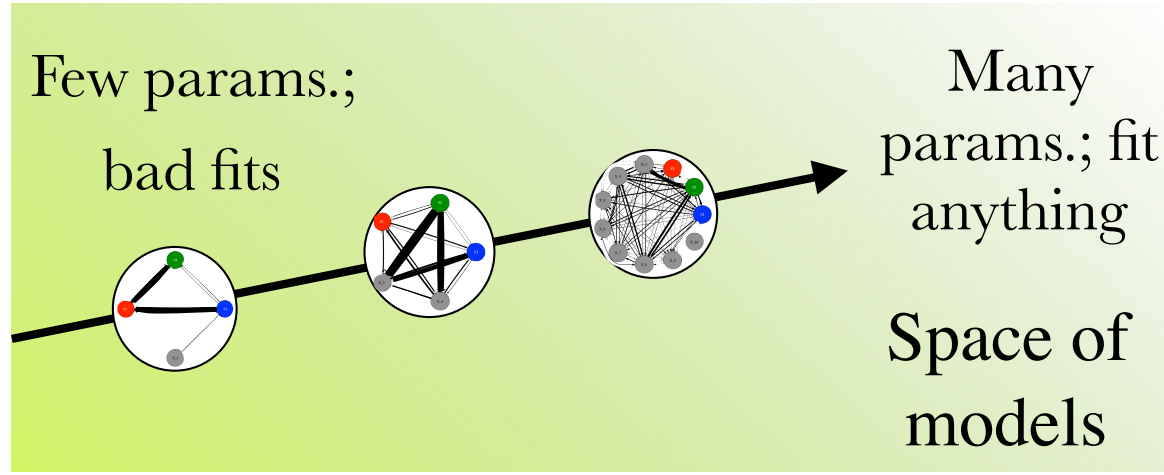
- Penalty for model complexity (the log term) “selects” the best model family.
- Not that simple in detail, but this description is roughly accurate.
- Beautiful consistency properties for **nested, complete** model families.

MacKay 1992, Balasubramanian 1996, Nemenman 2005

Why is fitting dynamics so hard?

$$\frac{d\vec{x}}{dt} = \begin{array}{l} A_{\{xx\}}\vec{x} \\ +A^{(2)}_{\{xx\}}\vec{x}\odot\vec{x} \\ +\dots \\ +A^{(K)}_{\{xx\}}\vec{x}\odot\cdots\odot\vec{x} \end{array}$$

More nonlinearities



More hidden variables

$$\frac{d\vec{x}}{dt} = A_{\{xx\}}\vec{x}$$

$$\begin{cases} \frac{d\vec{x}}{dt} = A_{\{xx\}}\vec{x} + B_{\{x\}1}\xi_1 + \cdots + B_{\{x\}K}\xi_K \\ \frac{d\xi_1}{dt} = A_{1\{x\}}\vec{x} + B_{11}\xi_1 + \cdots + B_{1K}\xi_K \\ \cdots \\ \frac{d\xi_K}{dt} = A_{K\{x\}}\vec{x} + B_{1K}\xi_1 + \cdots + B_{KK}\xi_K \end{cases}$$

- Hidden degrees of freedom and nonlinearities breaks nestedness -- no consistency.
- Choose any (reasonable) **complete** path through the model space
  - Good choice — good fits with few data; Bad choice — not worse than exhaustive search.

Two types of model families

- Both nested and complete.
- Account for nonlinearities **and** hidden variables.
- Biochemically reasonable.

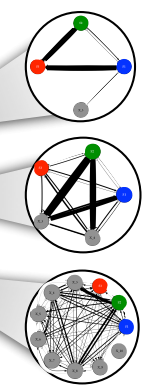
Degradation Interactions Input

Sigmoidal recurrent networks

Daniels and Beer, arXiv 2010

$$\frac{dx_i}{dt} = -x_i/\tau_i + \sum_{j=1}^{J} W_{ij}\ \xi(x_j + \theta_j) + \sum_{k=1}^{K} V_{ik} I_k$$

$$\text{with } \xi(y) = 1/(1+e^{-y})$$

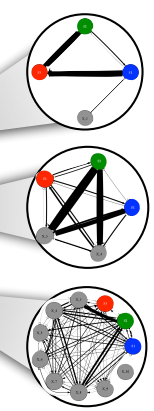


Interactions and input dependence

S-systems

Savageau et al., 1976-…

$$\frac{dx_i}{dt} = A_i \prod x_j^{\alpha_{ij}} \prod_k I_k^{a_{ik}} - B_i \prod x_j^{\beta_{ij}} \prod_k I_k^{b_{ik}}$$



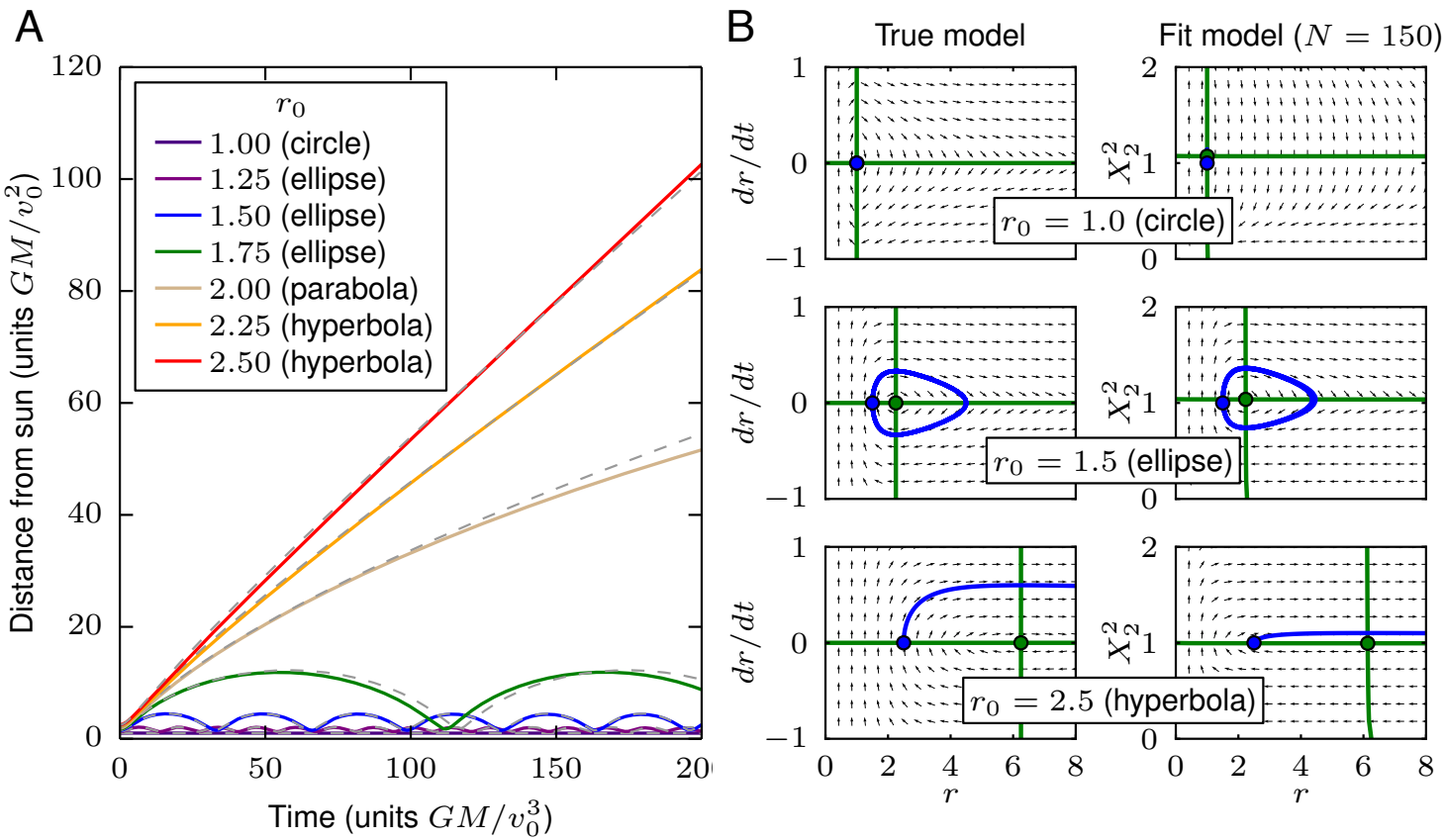
Daniels and Nemenman, Nature Comm 2015; PLoS ONE 2015

Algorithm

- Specify a particular hierarchy of model families.
- For given data:
  - Choose a model family within the hierarchy.
  - Fit for the best model within the family.
  - Calculate the posterior likelihood of the family using modified Bayesian criterion.
  - Choose more complex family and terminate when the modified likelihood starts to decrease.
- Algorithmic improvements to ensure that no complete re-fitting is done when move to the next family, or increase data set size.
- Two exponential complexities: **search of a model family**, and fitting a model within the family.
  - This only solves the **first**.
  - In practice works OK for both.

Finding laws that we already know: An automated Sir Isaac *(SirIsaac on GitHub)*

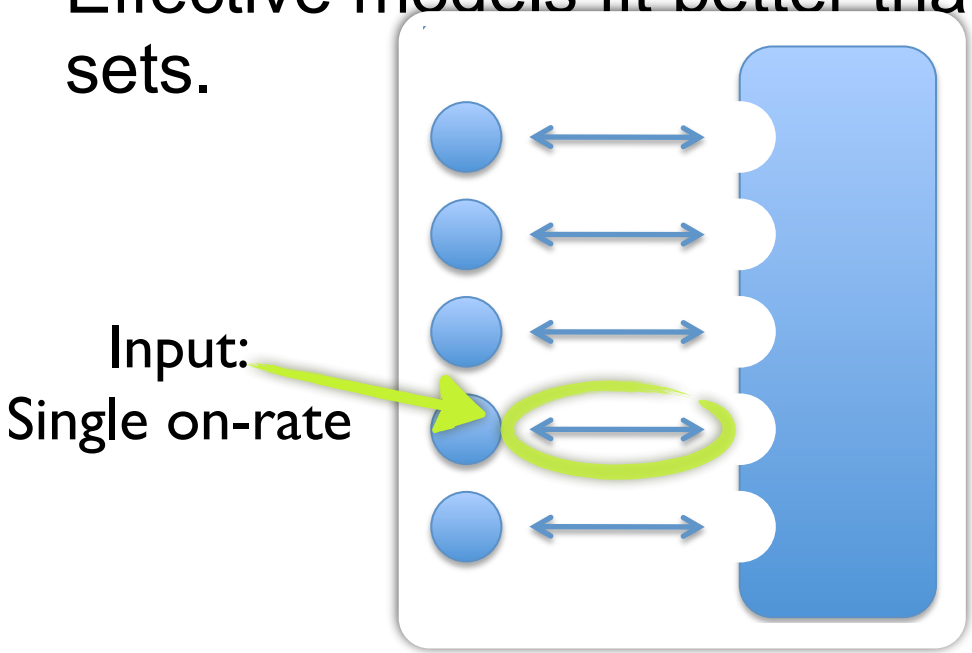
- Fin
  Nev
- Acc

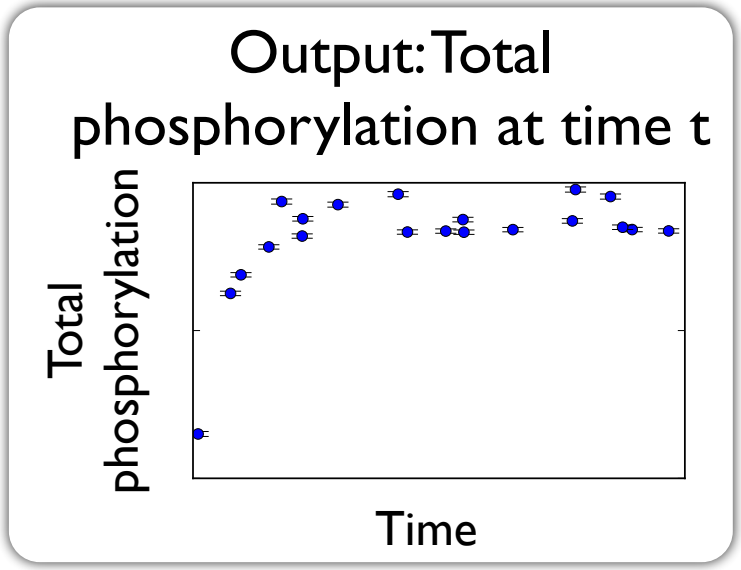


Daniels and Nemenman, Nature Comm 2015, PLoS ONE 2015

Simple dynamics from a complex network: Combinatorial multisite phosphorylation

- Effective models fit better than the true model for finite data sets.

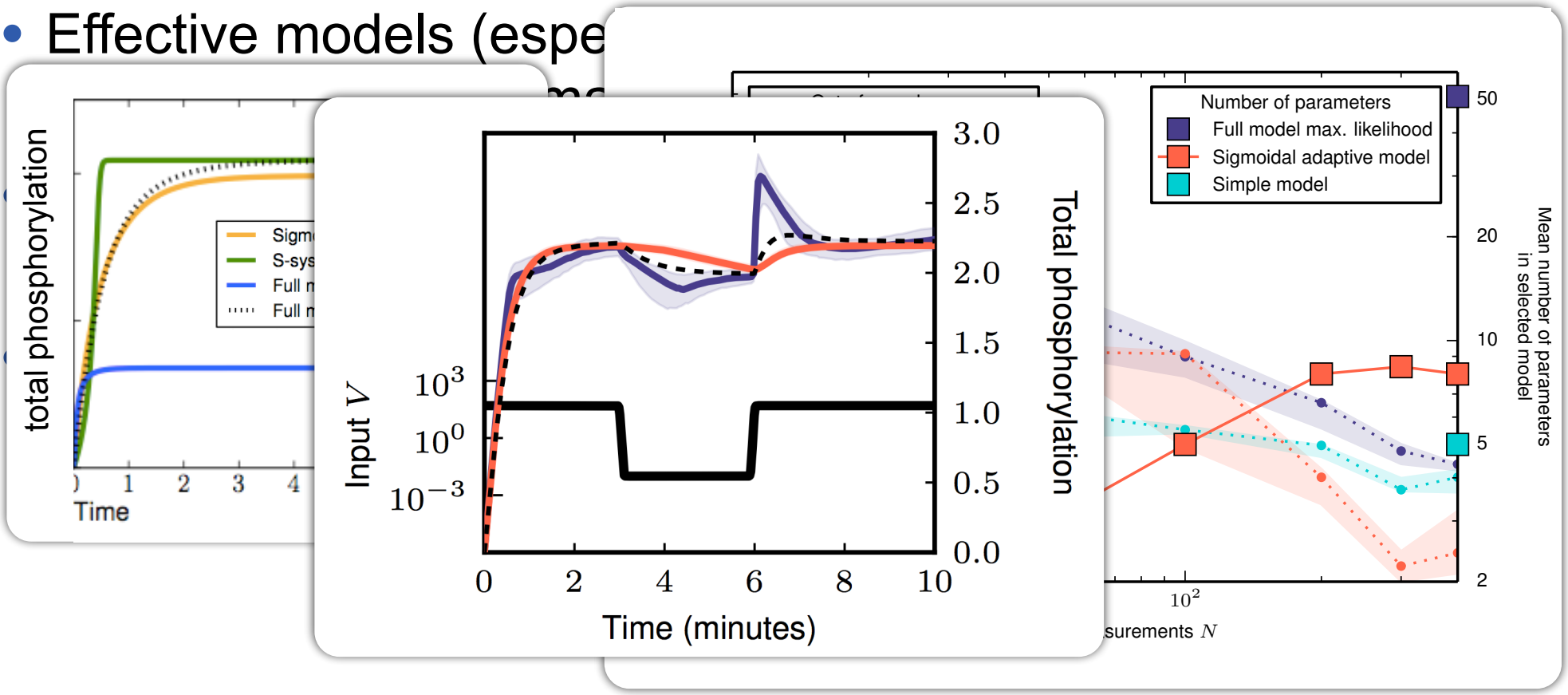




- Rates depend on occupancy of the nearby sites, about 50 parameters total.
- Caricature of some of the most combinatorially complex signaling models.
- Typically more parameters than data.

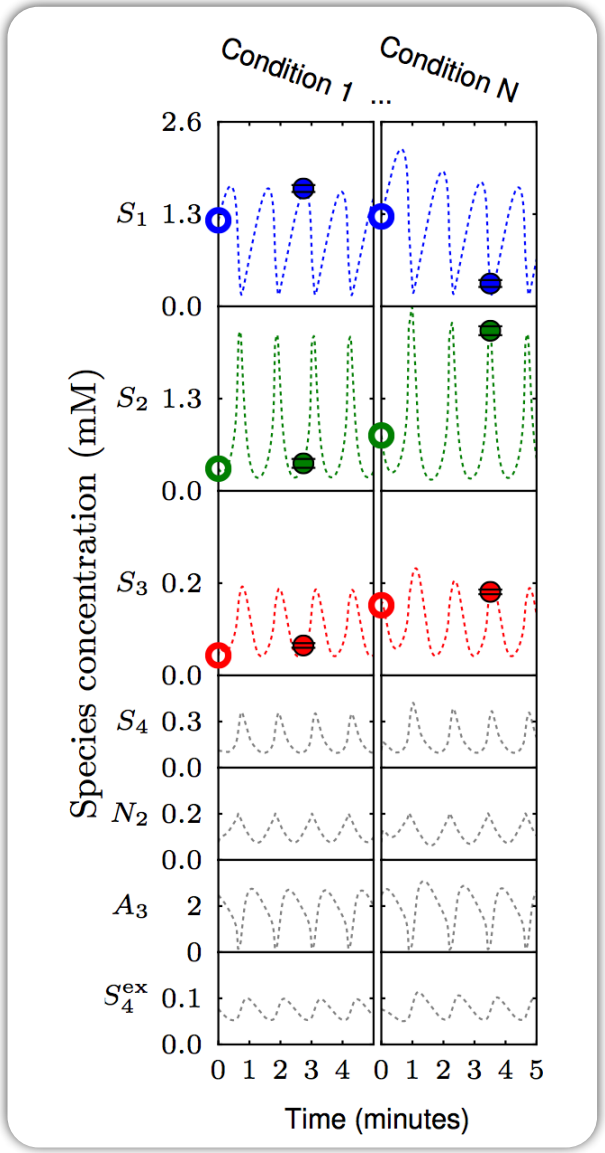
Effective, reduced model of multi-site phosphorylation

- Effective models (espe



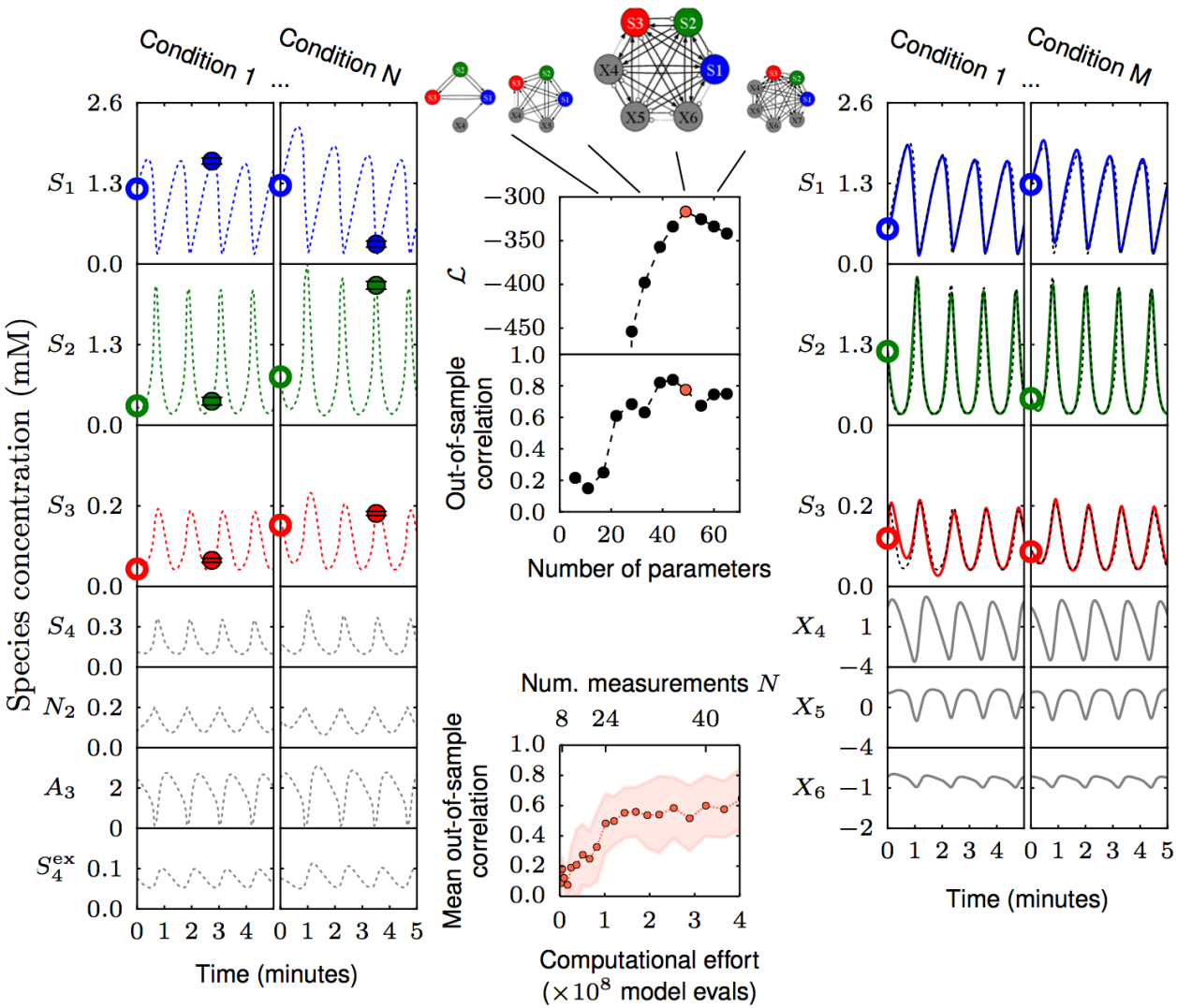
The yeast glycolytic oscillations: Complex dynamics needing complex structure

- Observe only 3/7 of variables; add 10%
- Data: *N* samples of structure
  - Initial condition of the 3 species;
  - Some random time later;
  - The value of these 3 species at that time.



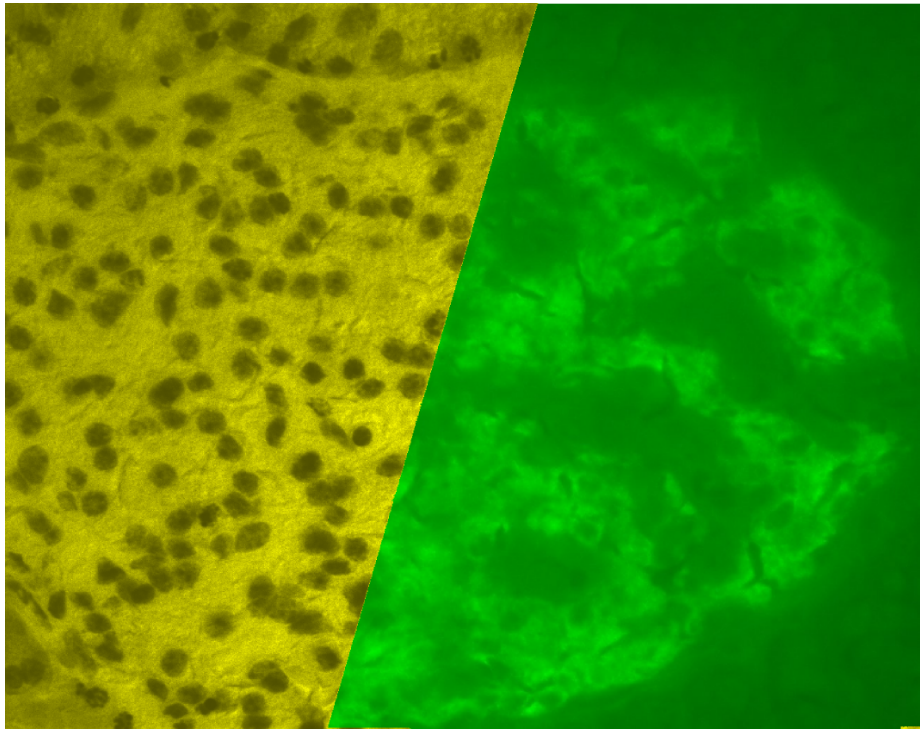
Results

Daniels and Nemenman, Nature Comm 2015; PLoS ONE 2015



- ~100x fewer evaluations for the same accuracy compared to full search.
- ~1000x fewer data points than full search.
- Better accuracy than curve fitting.
- Linear scaling with the amount of data and with the number of variables.

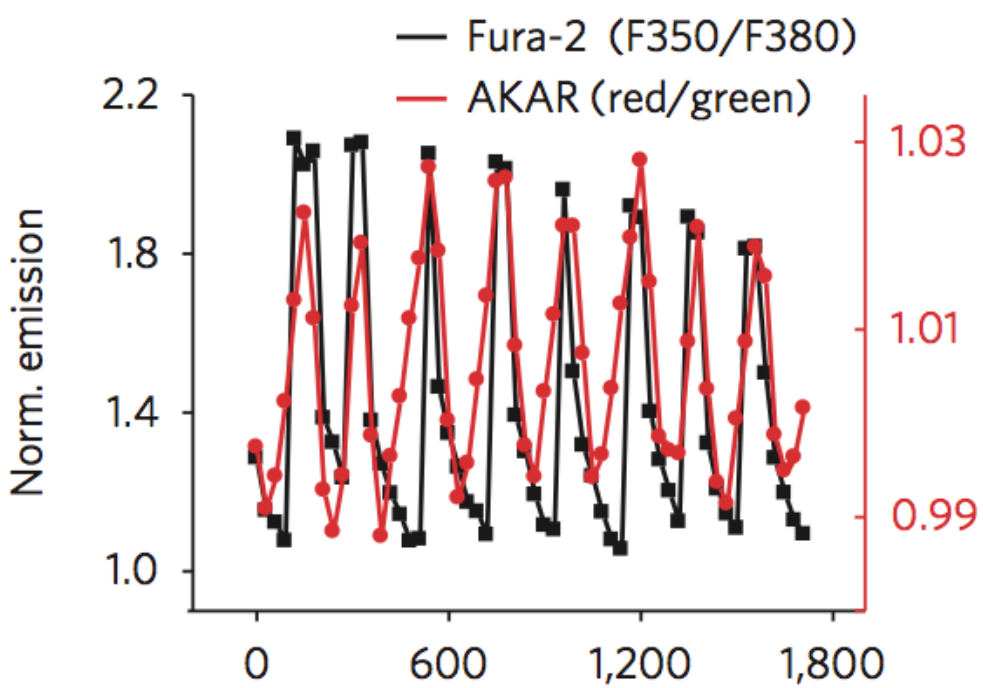
Learning new science: Ca++ oscillations in beta cells



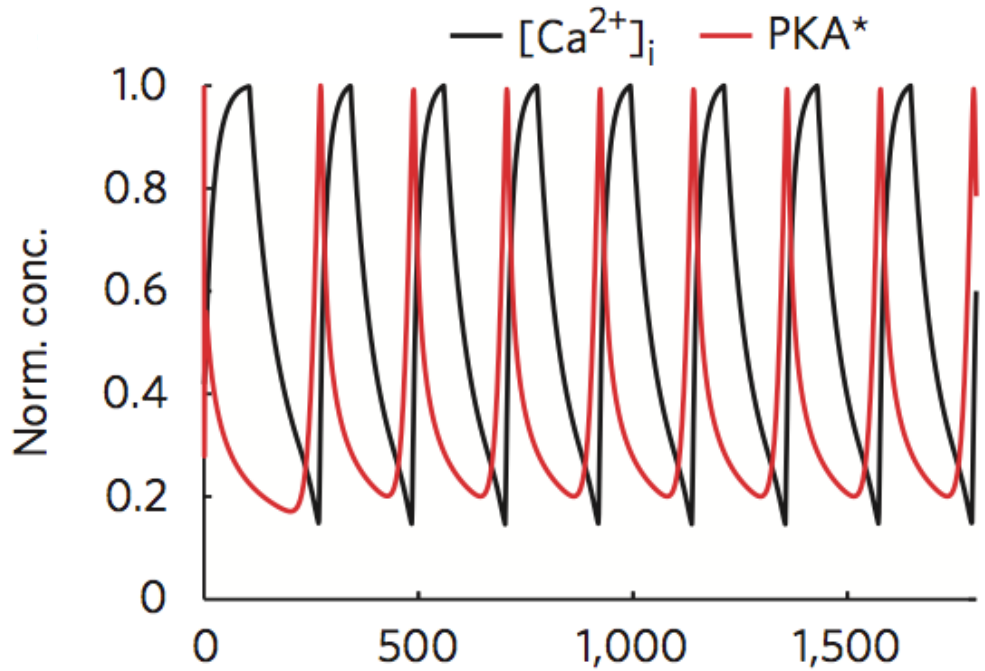
- Beta cells: cells in the pancreatic islets. Primary function: to store and release insulin.
- Insulin is released in pulses in synchrony with pulsatile activity of Ca++ spikes.
- Incorrect frequency / irregular pulsing of Ca++ / insulin measured in the portal vein, which connects pancreas to liver, prevents the liver from absorbing glucose from blood and is a precursor of type 2 diabetes.
- Protein Kinase A (PKA), cAMP, and Ca++ are a part of a synchronized oscillatory circuit.

Calcium-PKA oscillatory dynamics in beta cells

Experiment



Mechanistic model:
11 ODEs, 78+ parameters,
fits all cells with parameter adjustments

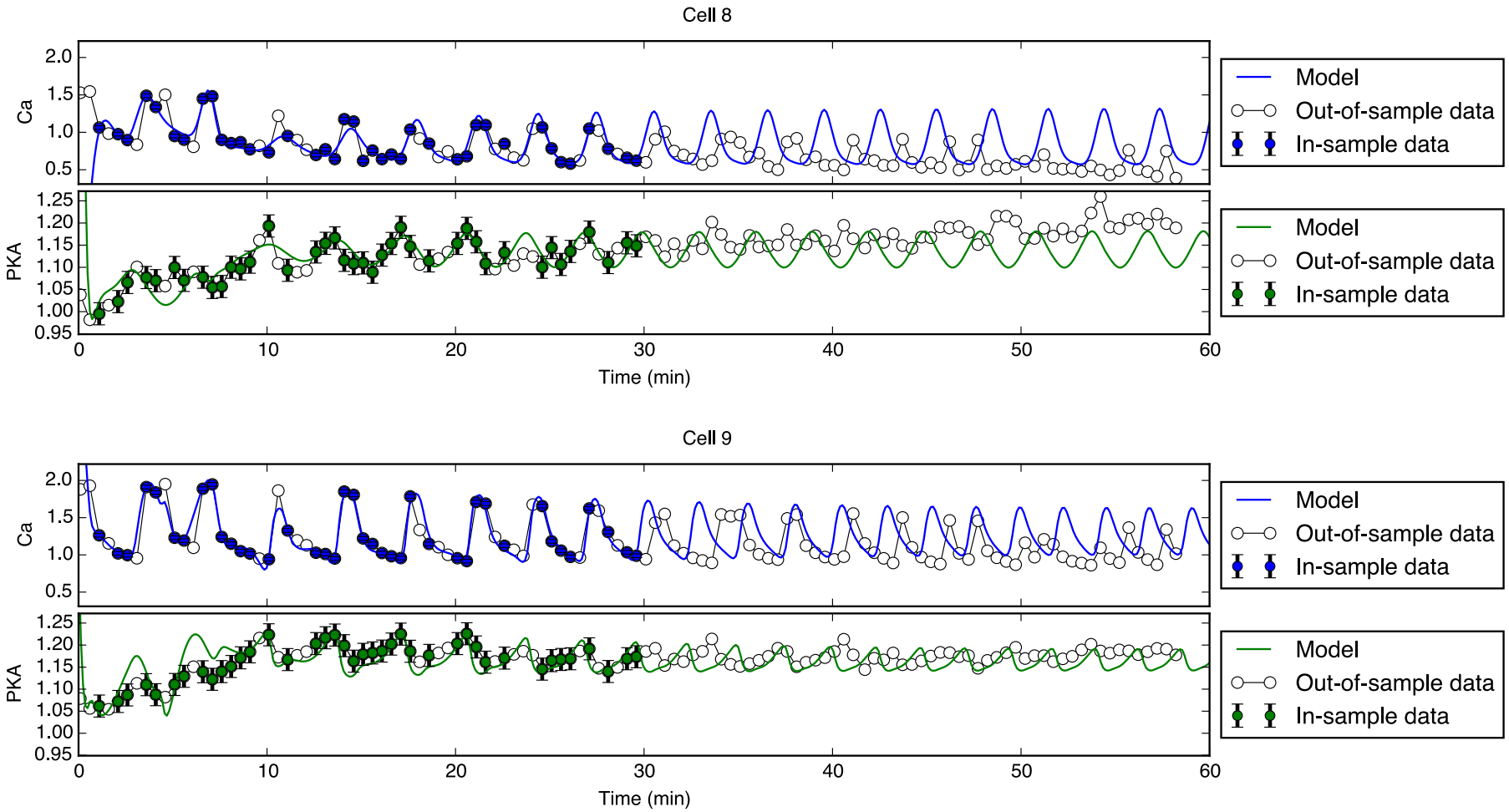


- Doesn't account for decreasing amplitude of oscillations.
- Different model for every cell.

Ni et al, NCB, 2010

Calcium-PKA oscillatory dynamics in beta cells: automatic inference

Daniels, IN, Levchenko, in prep.

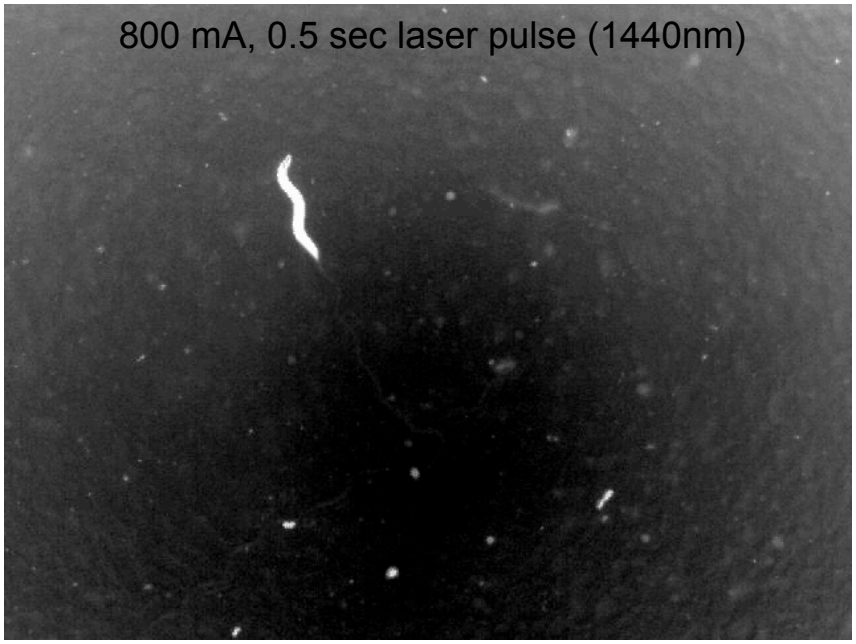


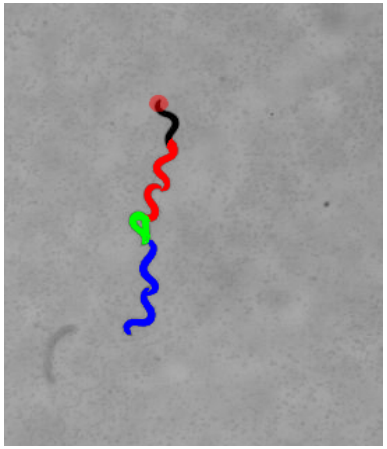
- 6 ODEs, <40 parameters.
- Fits every cell with <10 parameters individualized..

Modeling *C. elegans* temperature nociception escape response

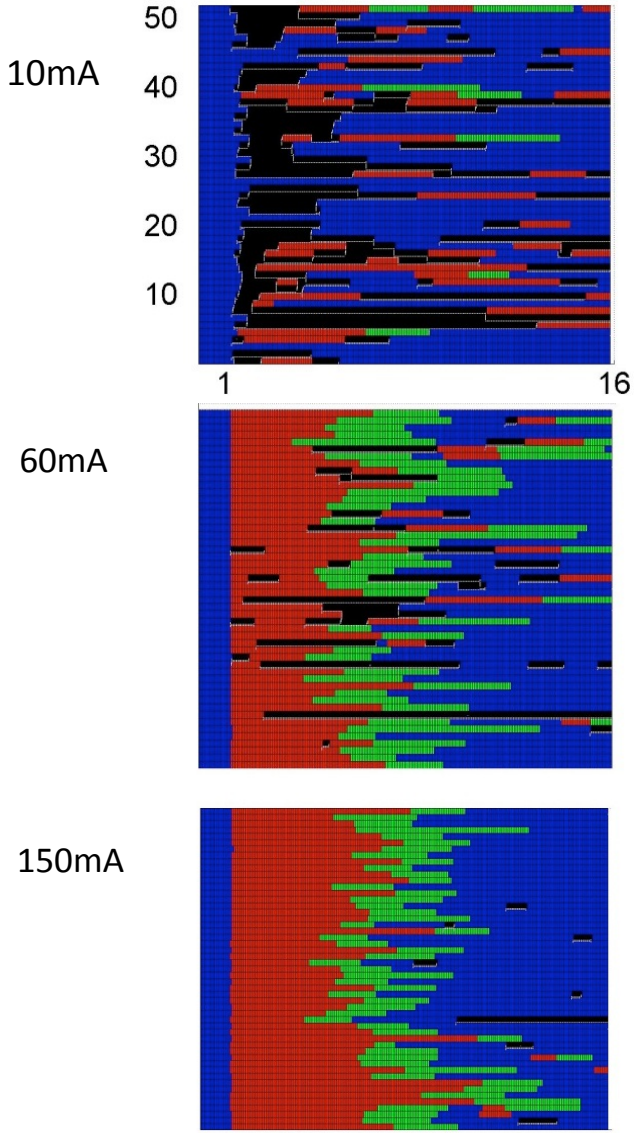
among behavioral
ral states?

Properties of the escape behavior

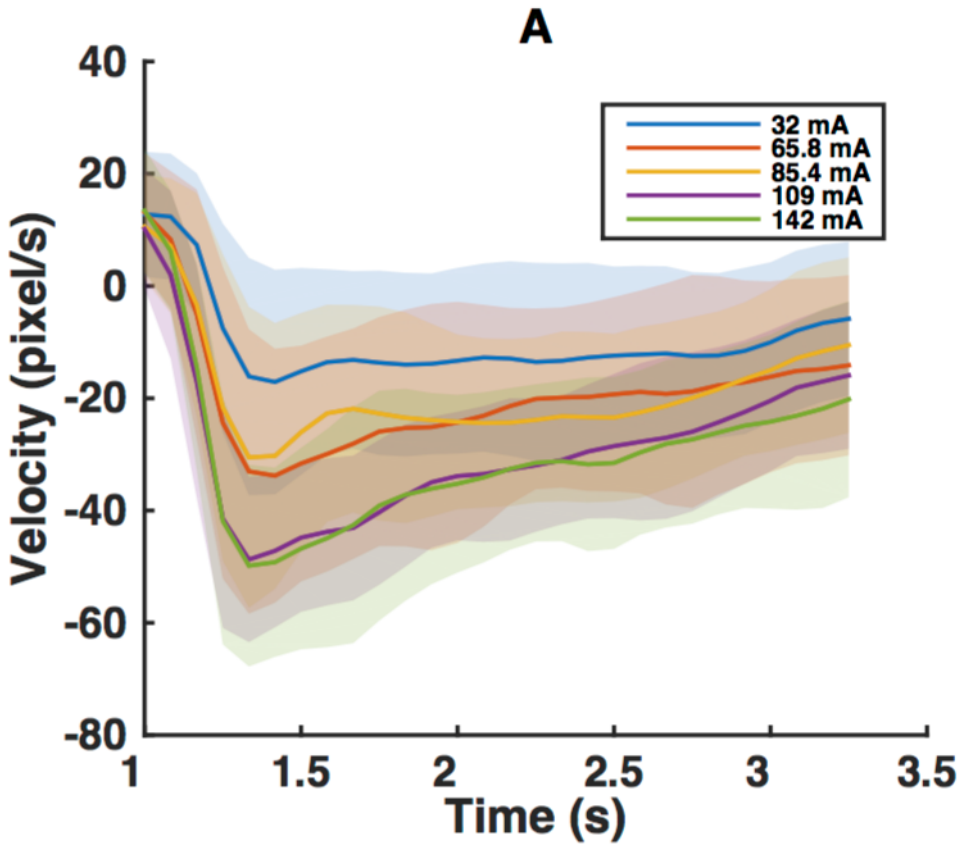


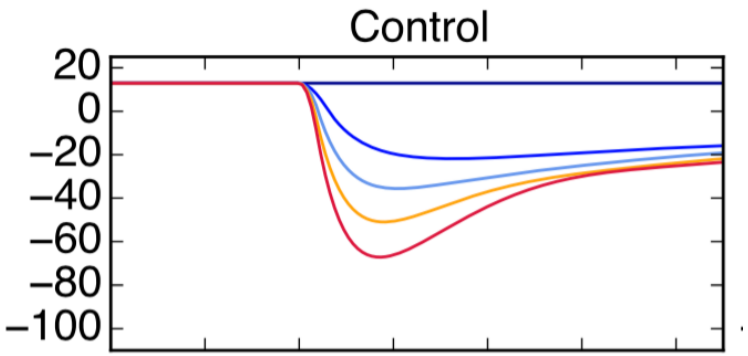


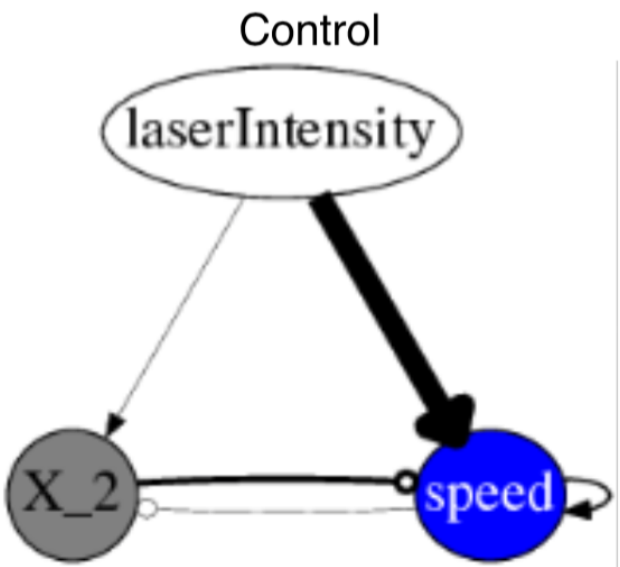
REVERSE
FORWARD
PAUSE



Worm nociception: Data and fits

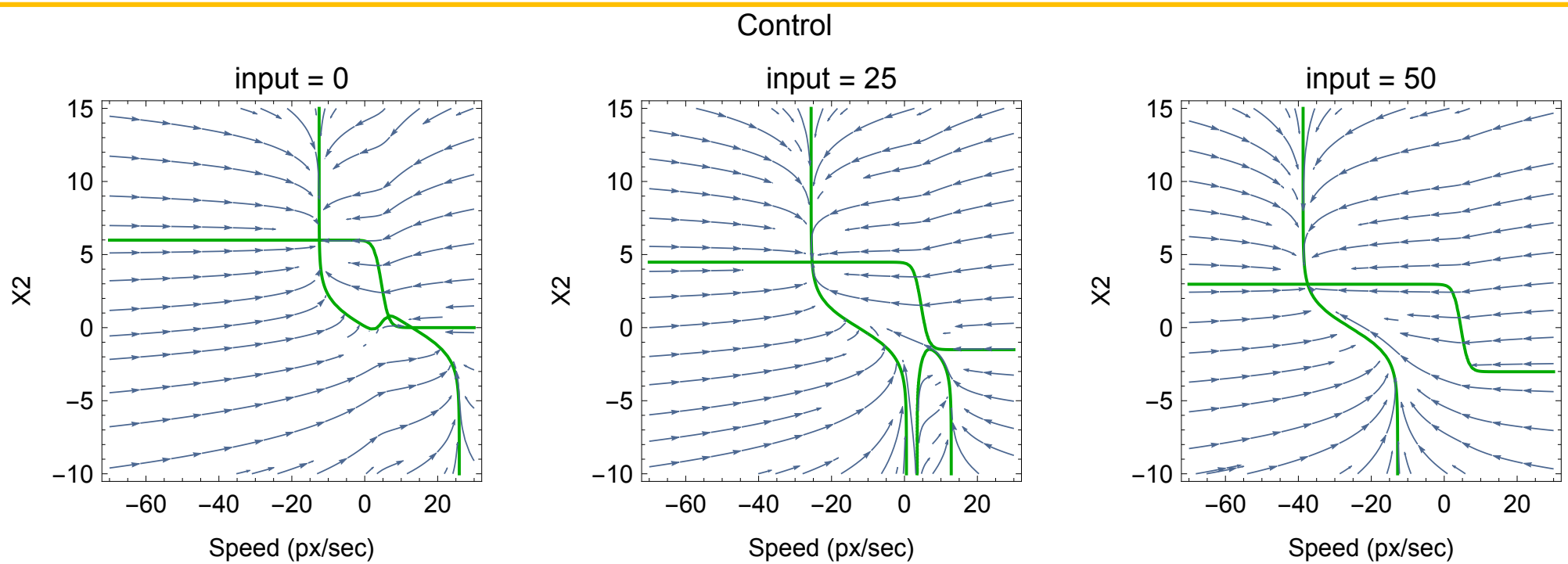






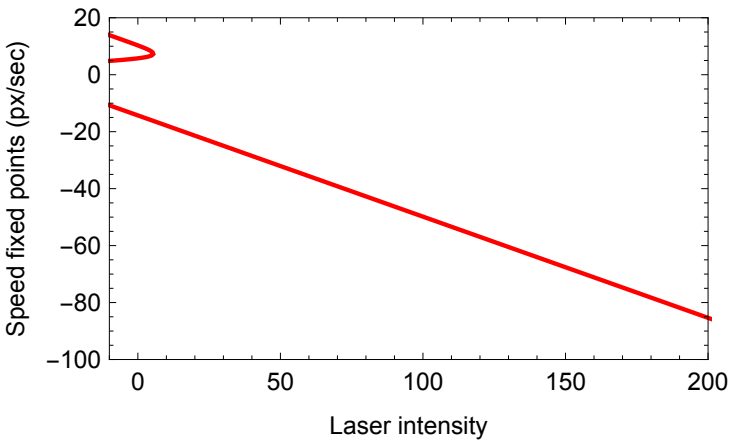
Leung, Mohhamadi, Ryu, IN, 2016

Worm nociception: phase space

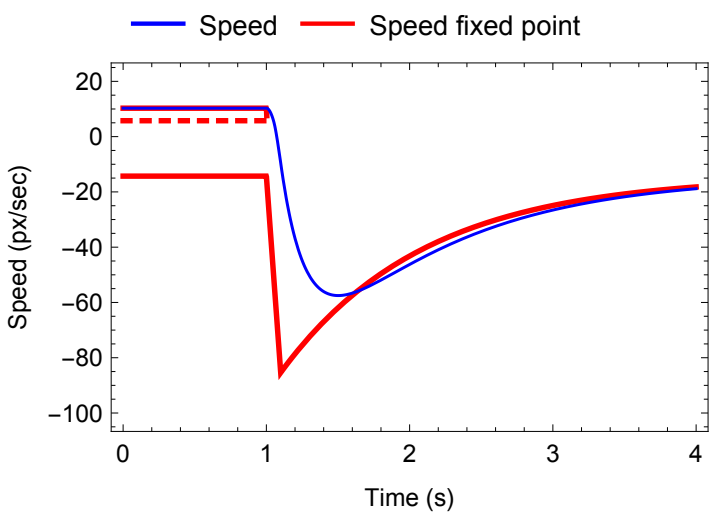


Daniels, Ryu, IN, in prep.

Worm nociception: phase space



- **Settle the debate:** Do stimuli bias transition probabilities among behavioral states, or do they create new behavioral states? In some sense, both.



Daniels, Ryu, IN, in prep.

Summary

- Focus on **refining phenomenological** dynamics.
- **Complete, nested** model families of dynamics allow to use Bayesian model selection to adapt model complexity to the available data.
- Such phenomenological models make accurate predictions in the undersampled regime, where true models overfit.
- Why do this?
  - **The duck test:** If it looks like a duck, swims like a duck, and quacks like a duck, then it probably is a duck **(Nemenman, Physics Today 2015)**.
  - **Find new phenomenological laws of nature**
    - Repeat Hookean approach in biology: build effective models of similar systems and look for patterns (e.g., chemotaxis in *C. elegans* and *E. coli*).