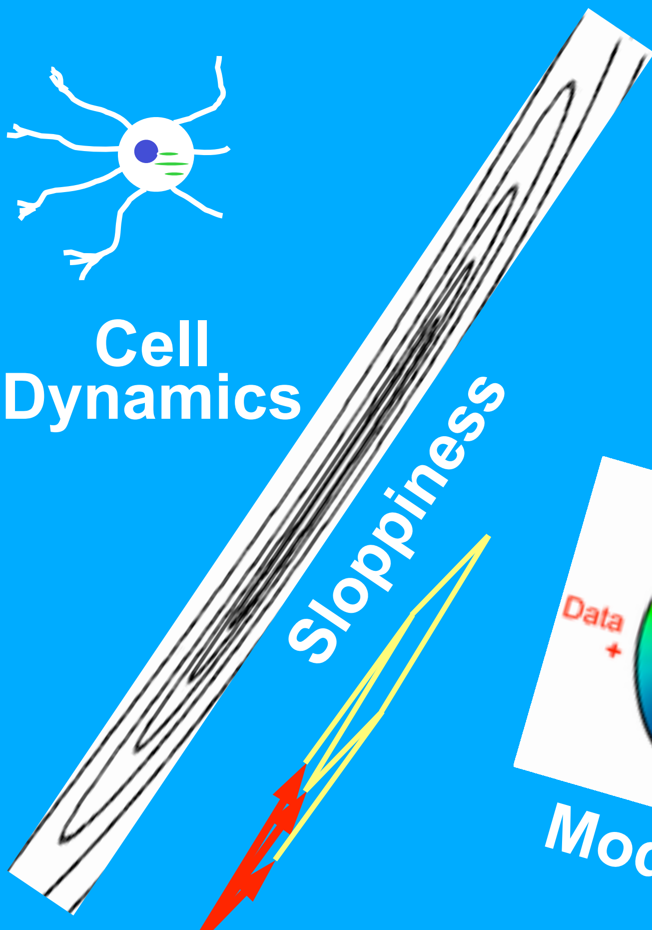
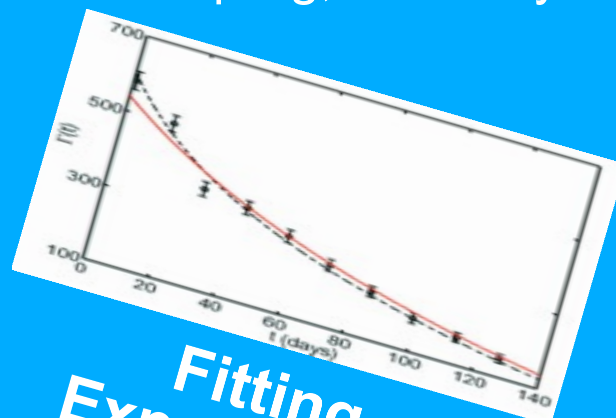
‘Sloppy Model’ Nonlinear Fits: Signal Transduction to Differential Geometry

JPS, Mark Transtrum, Ben Machta, Ricky Chachra, Lorien Hayden, Alex Alemi, Isabel Kloumann, Colin Clement, Kevin Brown, Ryan Gutenkunst, Josh Waterfall, Paul Ginsparg, Chris Myers, …

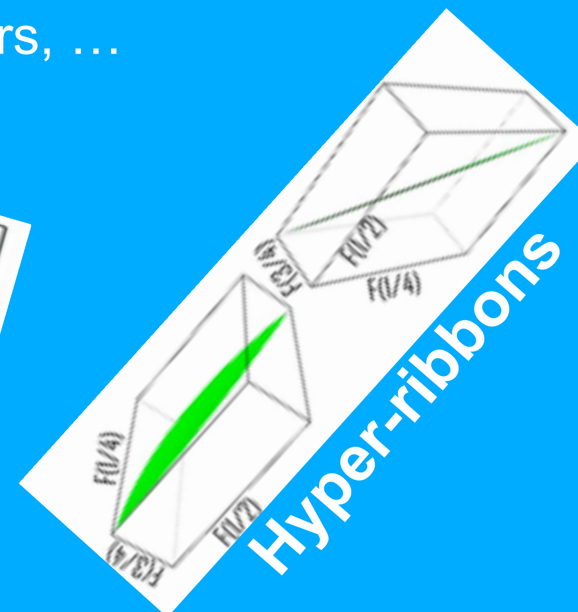


Cell Dynamics

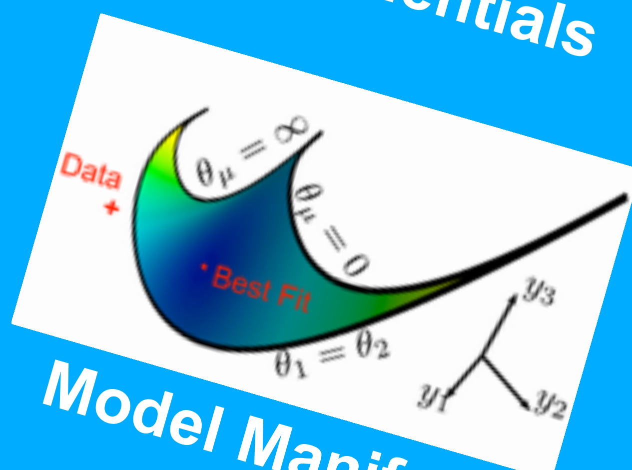
Sloppiness



Fitting Exponentials

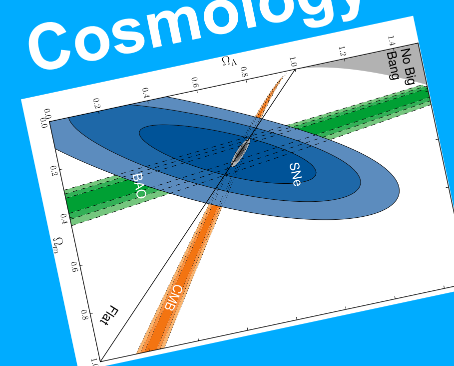


Hyper-ribbons



Model Manifold

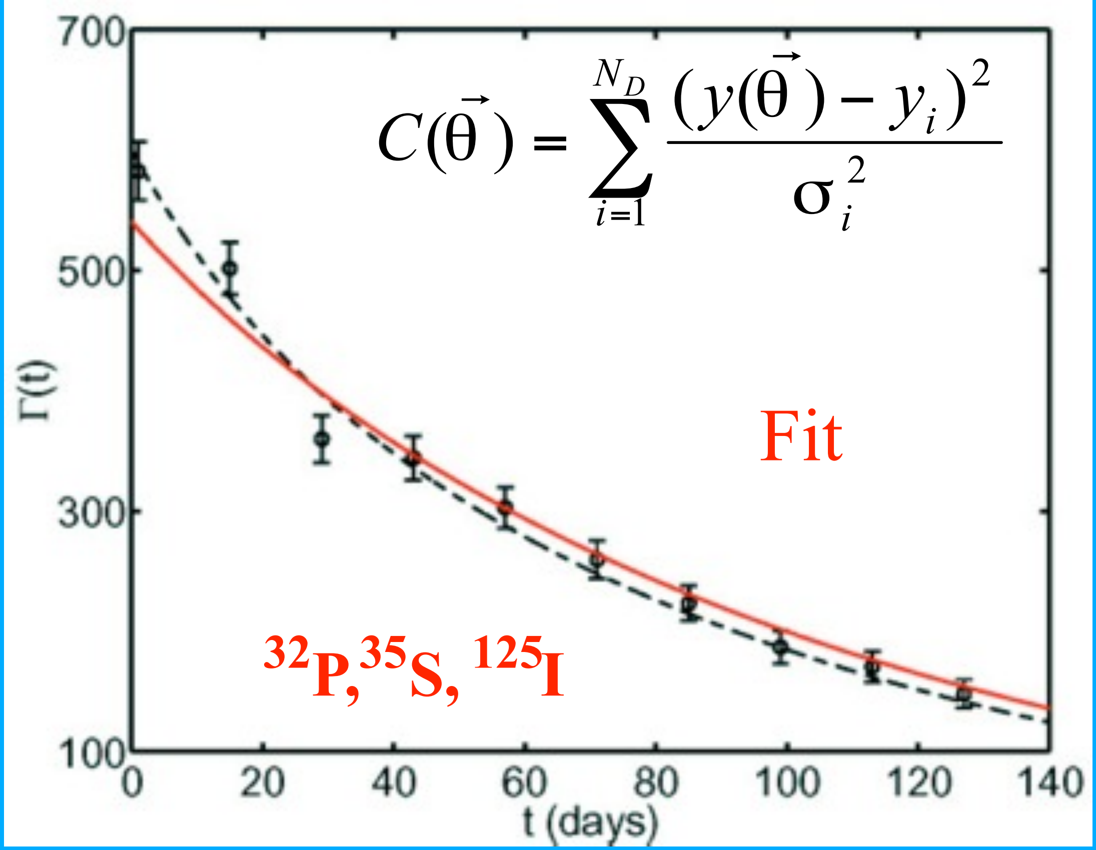
Cosmology



Fitting Decaying Exponentials

Classic ill-posed inverse problem

Given Geiger counter measurements from a radioactive pile, can we recover the identity of the elements and/or predict future radioactivity? Good fits with bad decay rates!



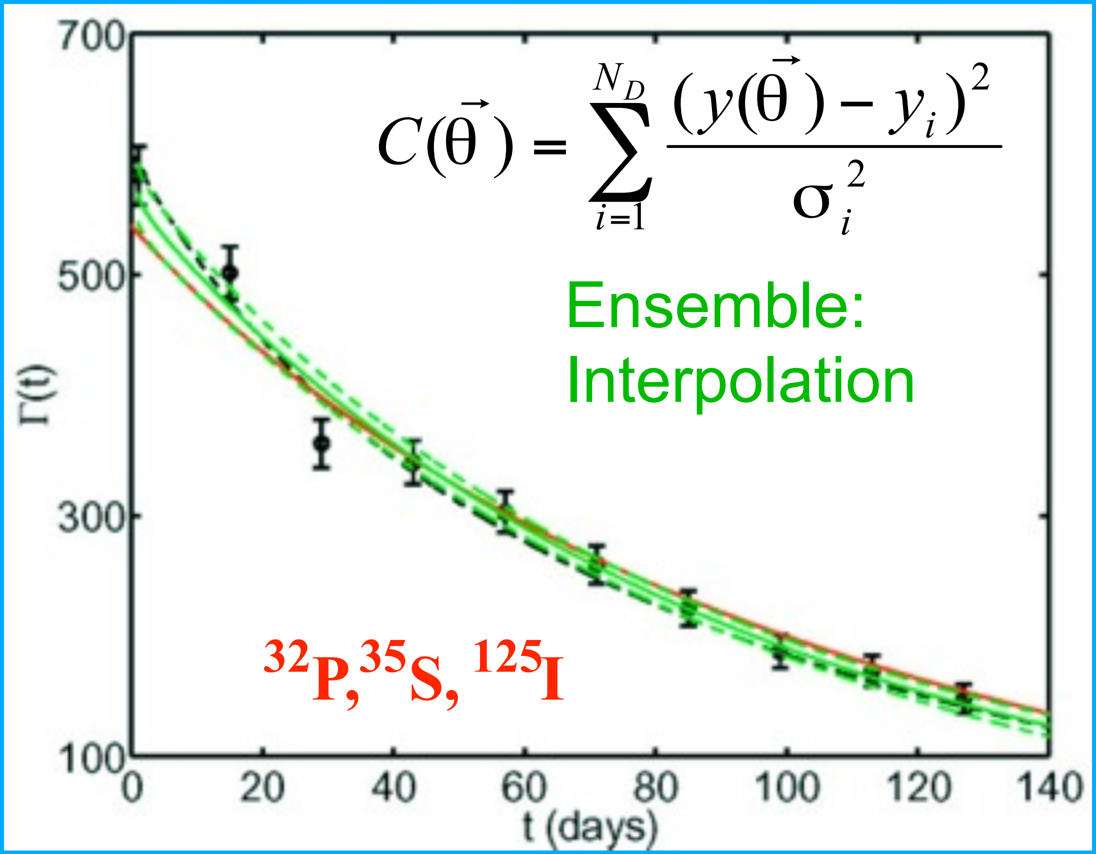
$$y(A,\gamma,t) = A_1 e^{-\gamma_1 t} + A_2 e^{-\gamma_2 t} + A_3 e^{-\gamma_3 t}$$

6 Parameter Fit

Fitting Decaying Exponentials

Classic ill-posed inverse problem

Given Geiger counter measurements from a radioactive pile, can we recover the identity of the elements and/or predict future radioactivity? Good fits with bad decay rates!



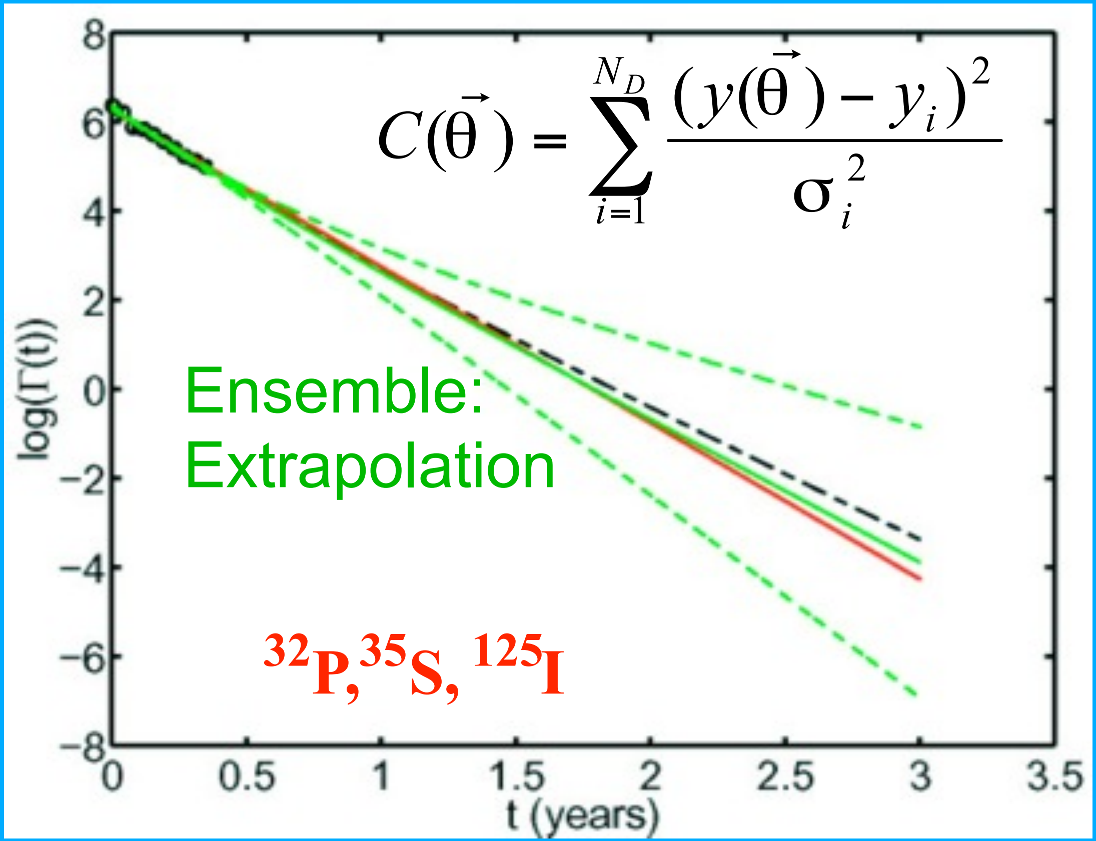
$$y(A,\gamma,t) = A_1 e^{-\gamma_1 t} + A_2 e^{-\gamma_2 t} + A_3 e^{-\gamma_3 t}$$

6 Parameter Fit

Fitting Decaying Exponentials

Classic ill-posed inverse problem

Given Geiger counter measurements from a radioactive pile, can we recover the identity of the elements and/or predict future radioactivity? Good fits with bad decay rates!



$$y(A,\gamma,t) = A_1 e^{-\gamma_1 t} + A_2 e^{-\gamma_2 t} + A_3 e^{-\gamma_3 t}$$

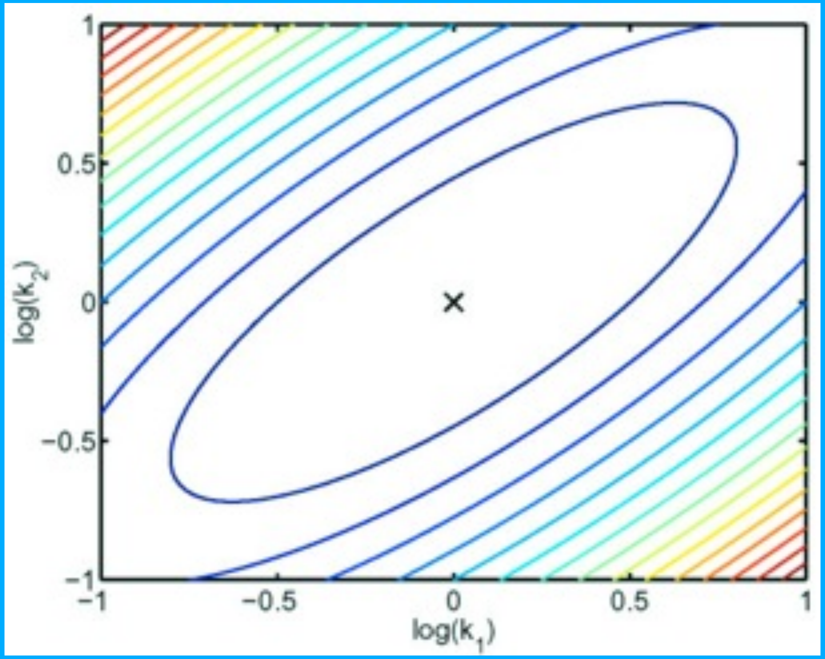
6 Parameter Fit

Ensemble of Models

Kevin Brown

We want to consider not just minimum cost fits, but all parameter sets consistent with the available data. New level of abstraction: *statistical mechanics in model space*.

Don't trust predictions that vary



$$H_{ij} = \partial^2 C / \partial\theta_i \partial\theta_j$$

Cost is least-squares fit

$$C(\vec{\theta}) = \frac{1}{2}\sum_{i=1}^{N_D} \frac{(y(\vec{\theta}) - y_i)^2}{\sigma_i^2}$$

Boltzmann weights exp(-C/T)

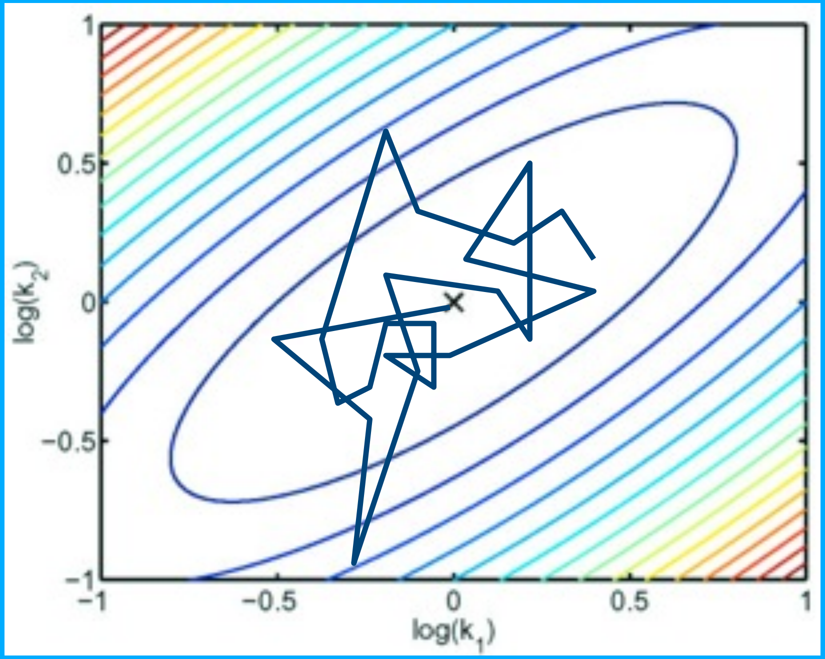
O is chemical concentration $y(t_i)$, or rate constant θ_n…

Ensemble of Models

Kevin Brown

We want to consider not just minimum cost fits, but all parameter sets consistent with the available data. New level of abstraction: *statistical mechanics in model space*.

Don't trust predictions that vary



$$H_{ij} = \partial^2 C / \partial\theta_i \partial\theta_j$$

Cost is least-squares fit

$$C(\vec{\theta}) = \frac{1}{2} \sum_{i=1}^{N_D} \frac{(y(\vec{\theta}) - y_i)^2}{\sigma_i^2}$$

Boltzmann weights exp(-C/T)

$$\langle O \rangle = \frac{1}{N_E} \sum_{i=1}^{N_E} O(\vec{\theta}_i)$$

$$\sigma_O^2 = \langle O^2(\vec{\theta}) \rangle - \langle O(\vec{\theta}) \rangle^2$$

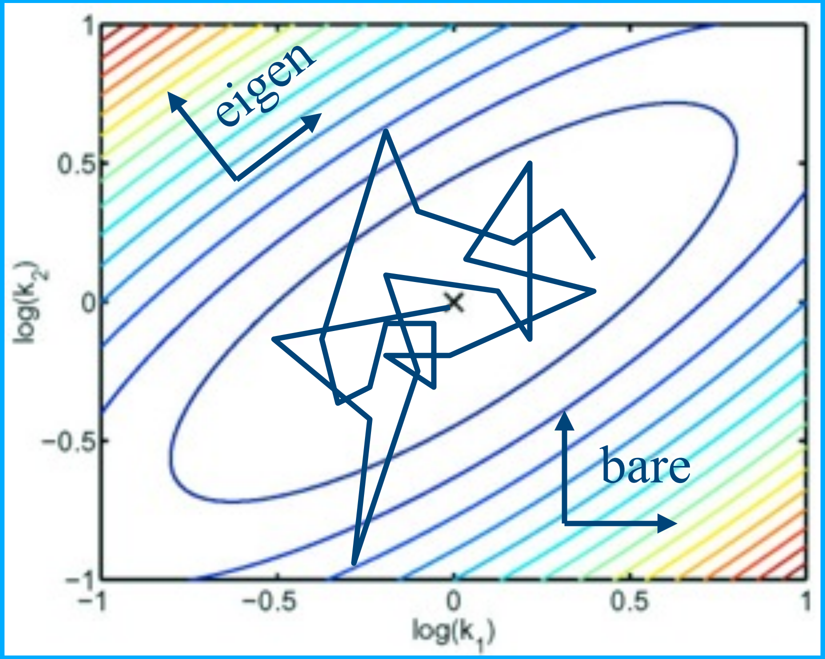
O is chemical concentration $y(t_i)$, or rate constant θ_n…

Ensemble of Models

Kevin Brown

We want to consider not just minimum cost fits, but all parameter sets consistent with the available data. New level of abstraction: *statistical mechanics in model space*.

Don't trust predictions that vary



$$H_{ij} = \partial^2 C / \partial\theta_i \partial\theta_j$$

Cost is least-squares fit

$$C(\vec{\theta}) = \frac{1}{2}\sum_{i=1}^{N_D} \frac{(y(\vec{\theta}) - y_i)^2}{\sigma_i^2}$$

Boltzmann weights exp(-C/T)

$$\langle O \rangle = \frac{1}{N_E}\sum_{i=1}^{N_E} O(\vec{\theta}_i)$$

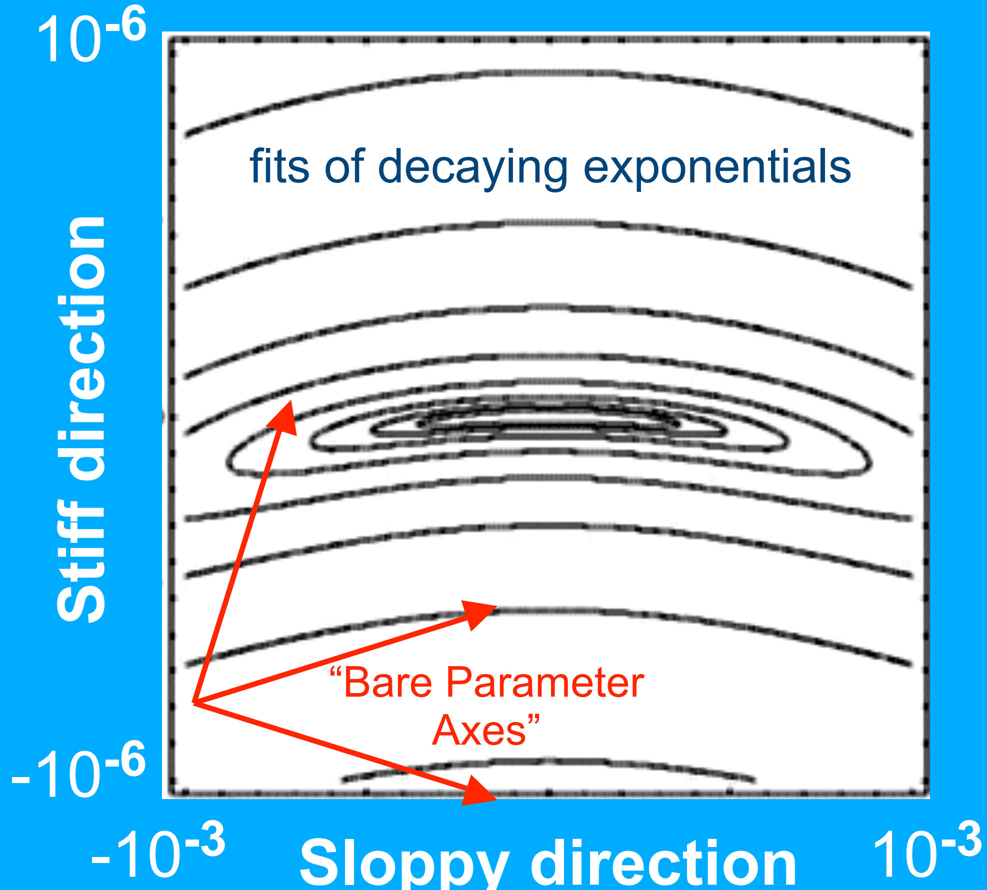
$$\sigma_O^2 = \langle O^2(\vec{\theta}) \rangle - \langle O(\vec{\theta}) \rangle^2$$

O is chemical concentration $y(t_i)$, or rate constant θ_n…

Parameter Indeterminacy and Sloppiness

Josh Waterfall, Ryan Gutenkunst, Chris Myers

Cost Contours

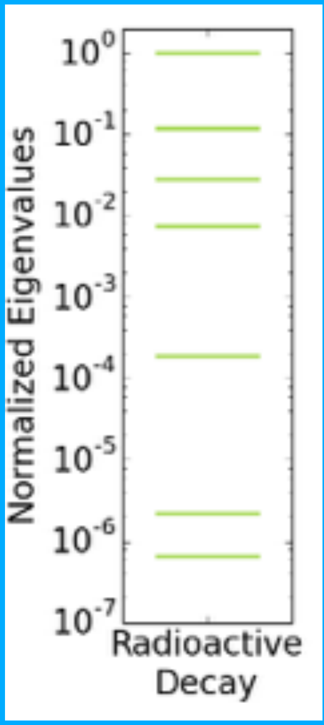


Horizontal scale shrunk by 1000 times

Aspect ratio = Human hair

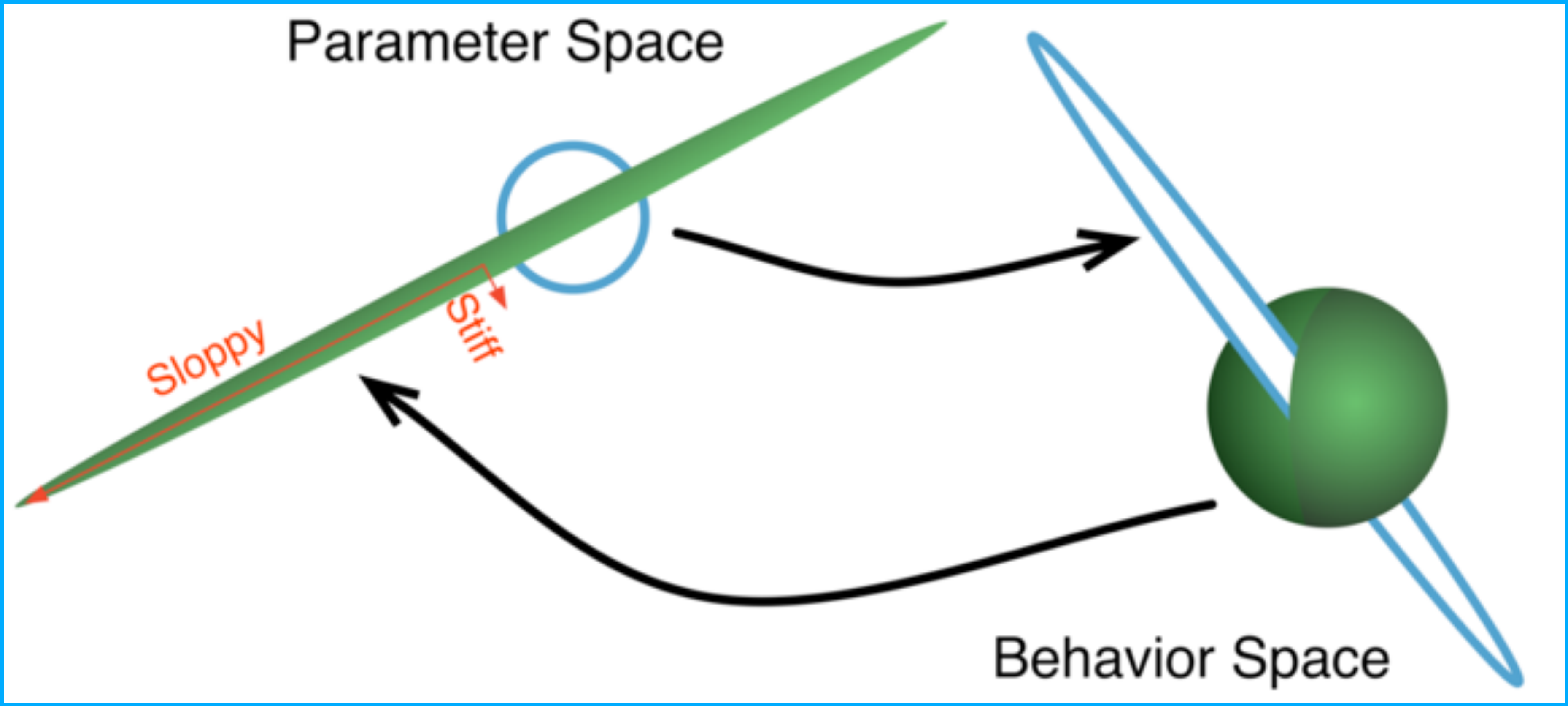
Many parameter sets give almost equally good fits

A few 'stiff' constrained directions allow model to remain predictive



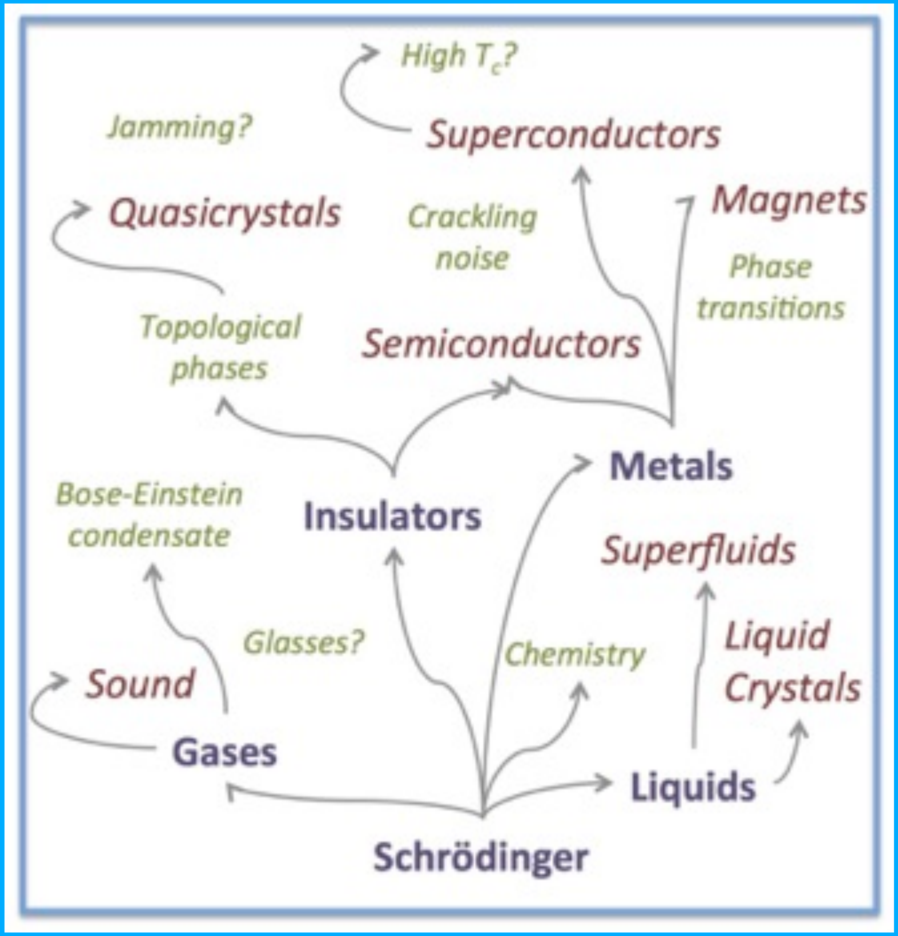
Few stiff, many sloppy directions

Models: Predictions about Data



Scientific model: Predictions about behavior depend on physical constants (parameters) in the model.
Sloppiness: the behavior only depends on a few stiff parameter combinations.

Emergence: More is Different



Emergence

Condensed Matter

Microscopic complexity

Simplicity emerges on long length and time scales, low energies

Emergent theory compresses microscopic details into a few governing parameters

Sloppiness and the Diffusion Equation

Ben Machta, Ricky Chachra, Mark Transtrum

What features of the microscopic hopping laws remain after several hops? Central limit theorem: only mean and variance.

Eigenvalues of parameter identifiability: Stiff = emergent, sloppy = microscopic

Diffusion equation

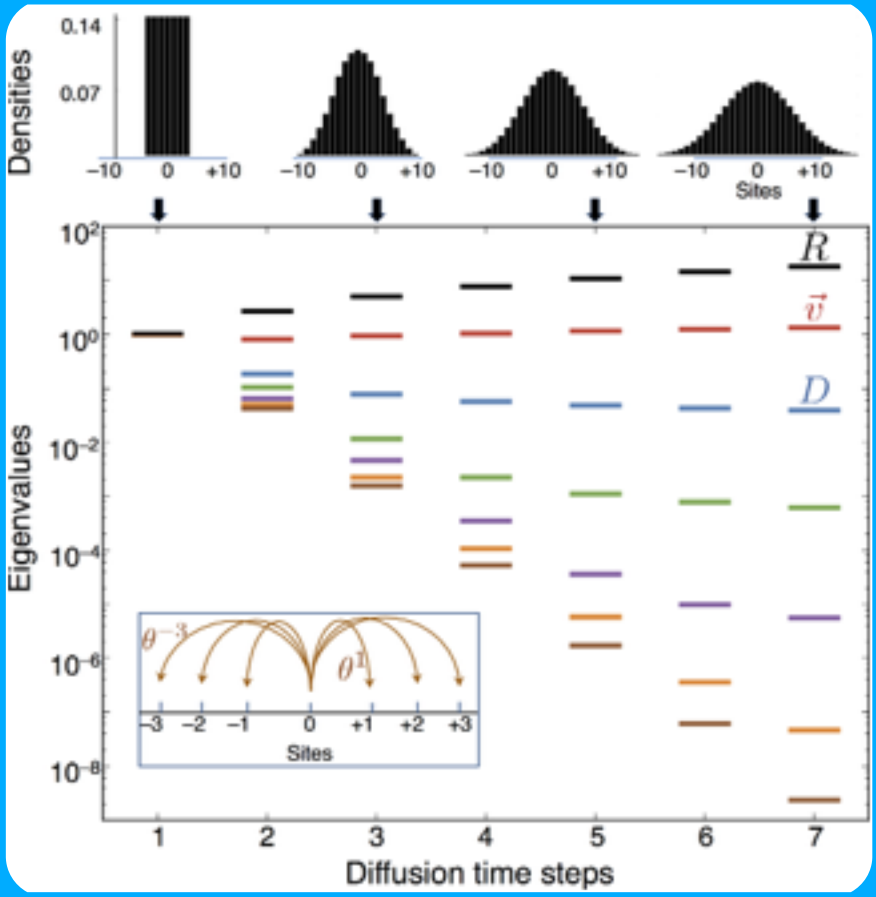
Microscopic long-range hopping model

Continuum limit

$\partial\rho/\partial t = R\rho - V\,\partial\rho/\partial x + D\,\partial^2\rho/\partial x^2$

One time step: all θ_h

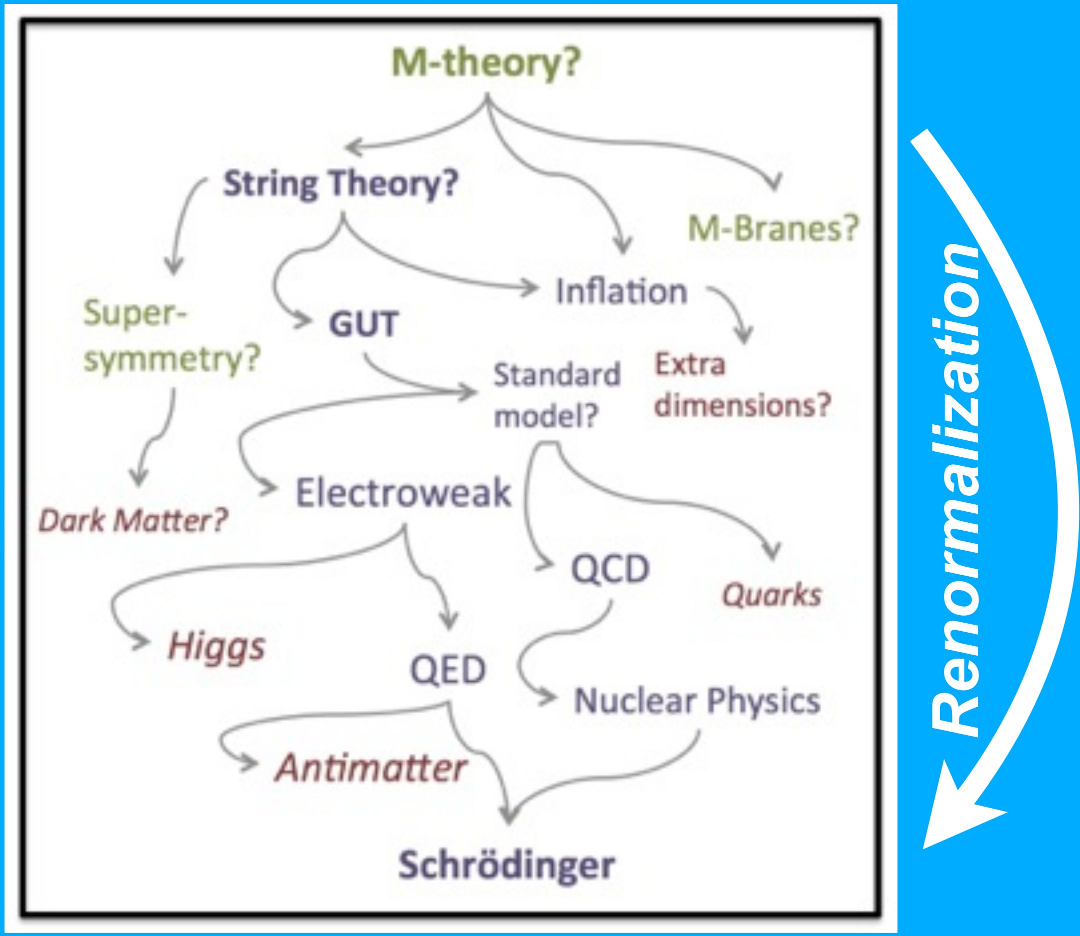
6 time steps: only R, V and D



Stiff

Sloppy

Renormalizability: Invisible underpinnings



Particle physics

Renormalizability: Low energy physics independent of cutoff theory

Underlying theory contributes only a few governing parameters

Can't see microscopic details at low energies: need big accelerators

Sloppiness and the Ising Model

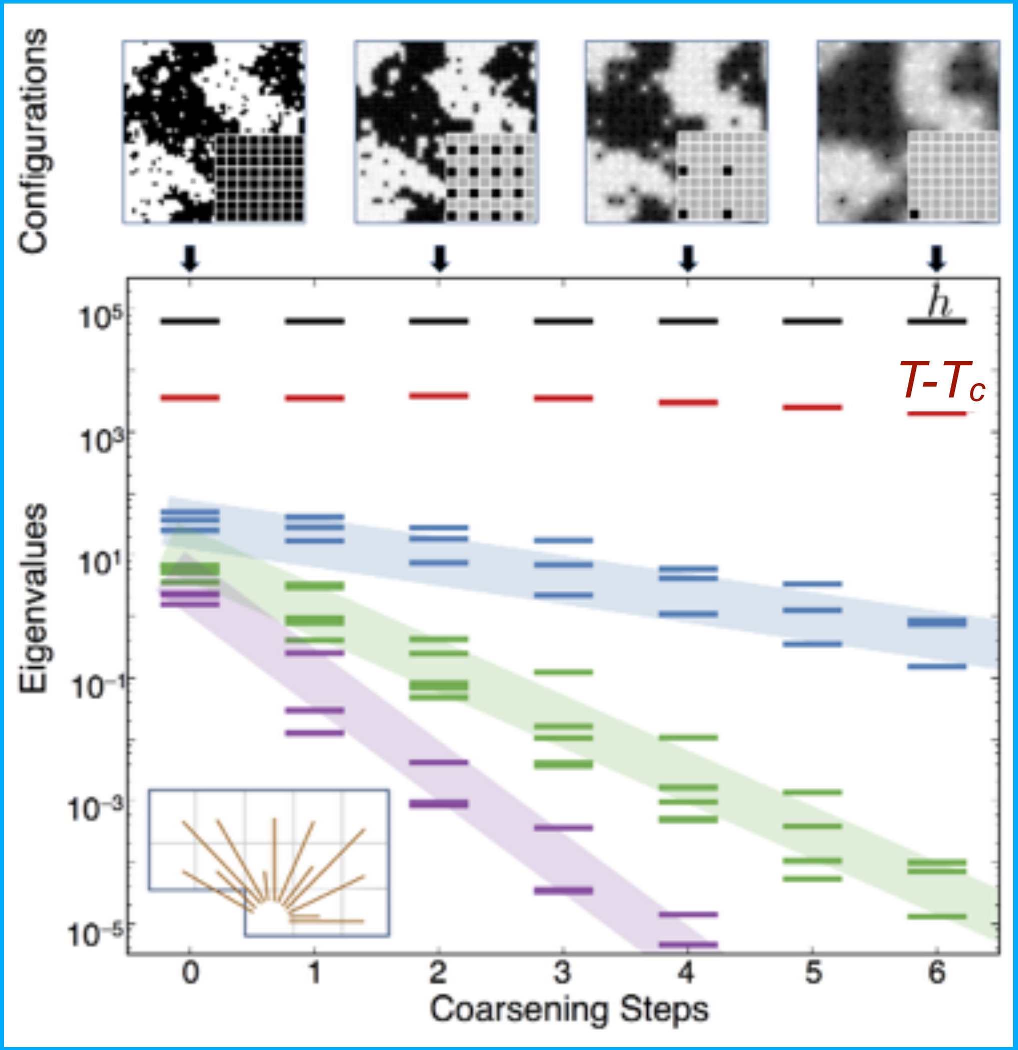
Ben Machta, Ricky Chachra, Mark Transtrum

What features of the microscopic interactions remain after coarse graining?

Renormalization group: only *h* and *T-T_c*

Eigenvalues of the Fisher Information matrix, Ising with long-range couplings. Only [left] eigenvector of relevant RG operators measurable

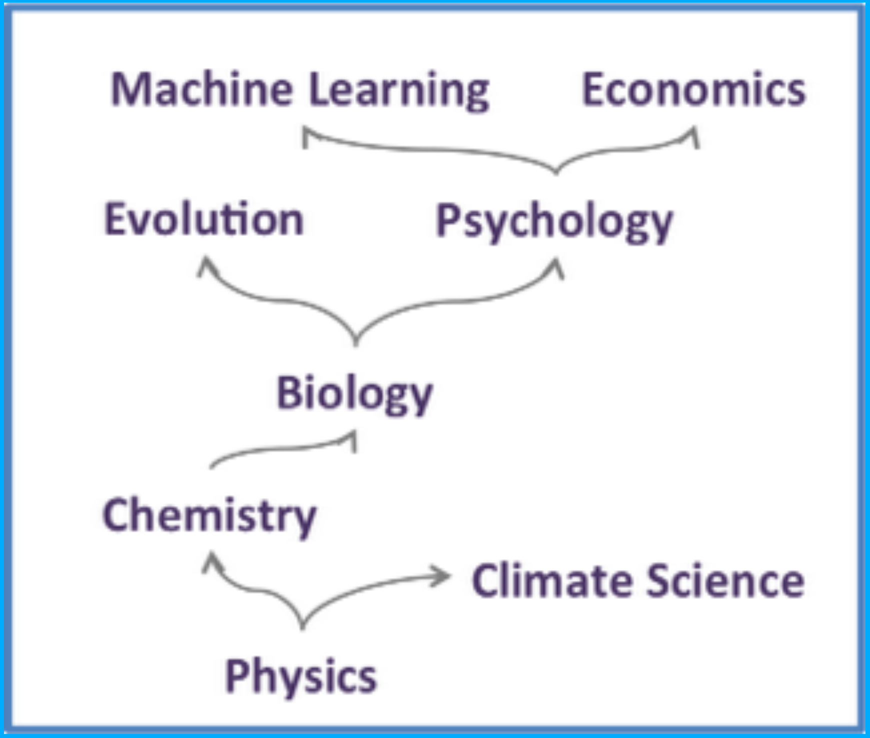
Sloppy after coarse graining in space



Stiff

Sloppy

Sloppiness and the rest of science

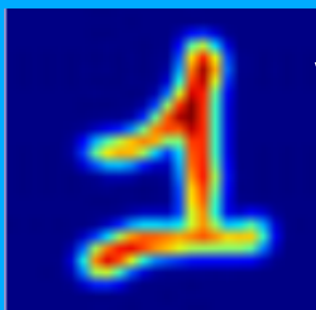


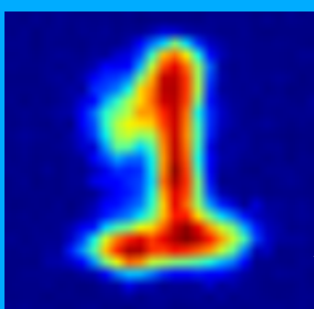
How is science possible, without small parameters like $1/L$, $T-T_c$?

Simple models succeed in describing complex behavior

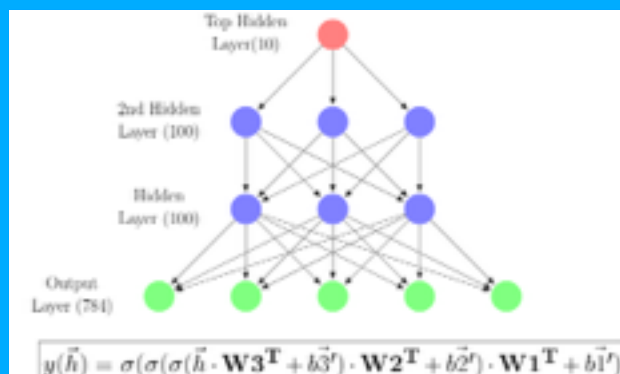
Neural Networks and the Model Manifold

Lorien Hayden, Alex Alemi

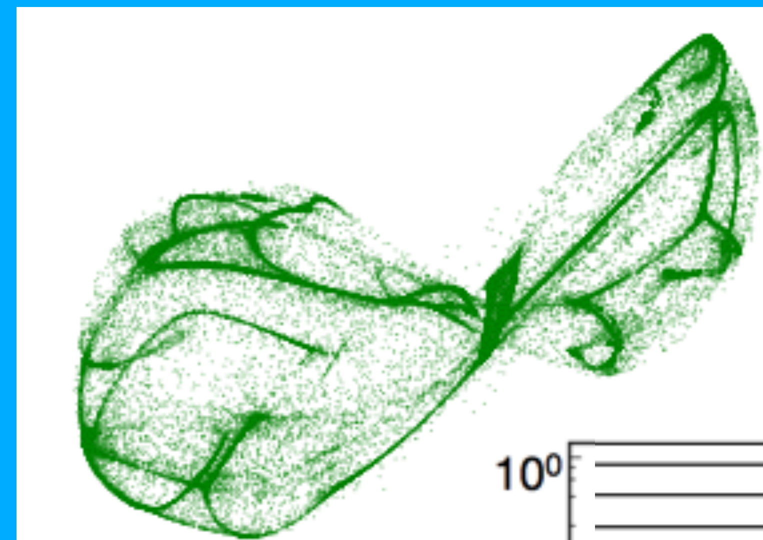


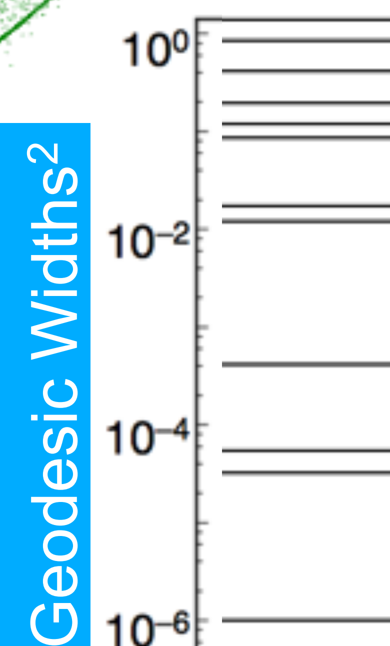


MNIST digits



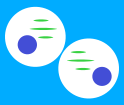
Stacked Denoising Autoencoder

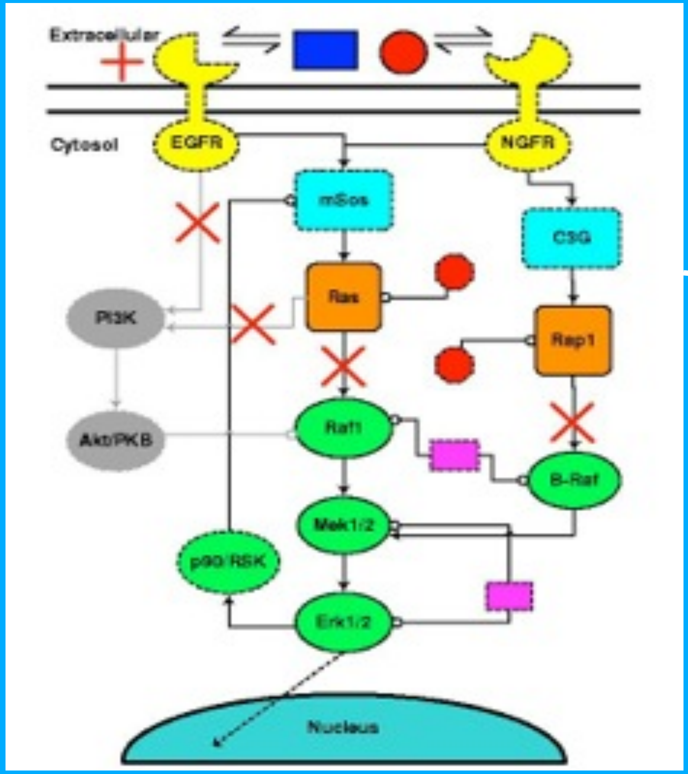


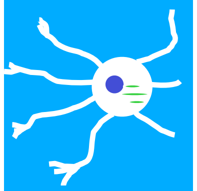


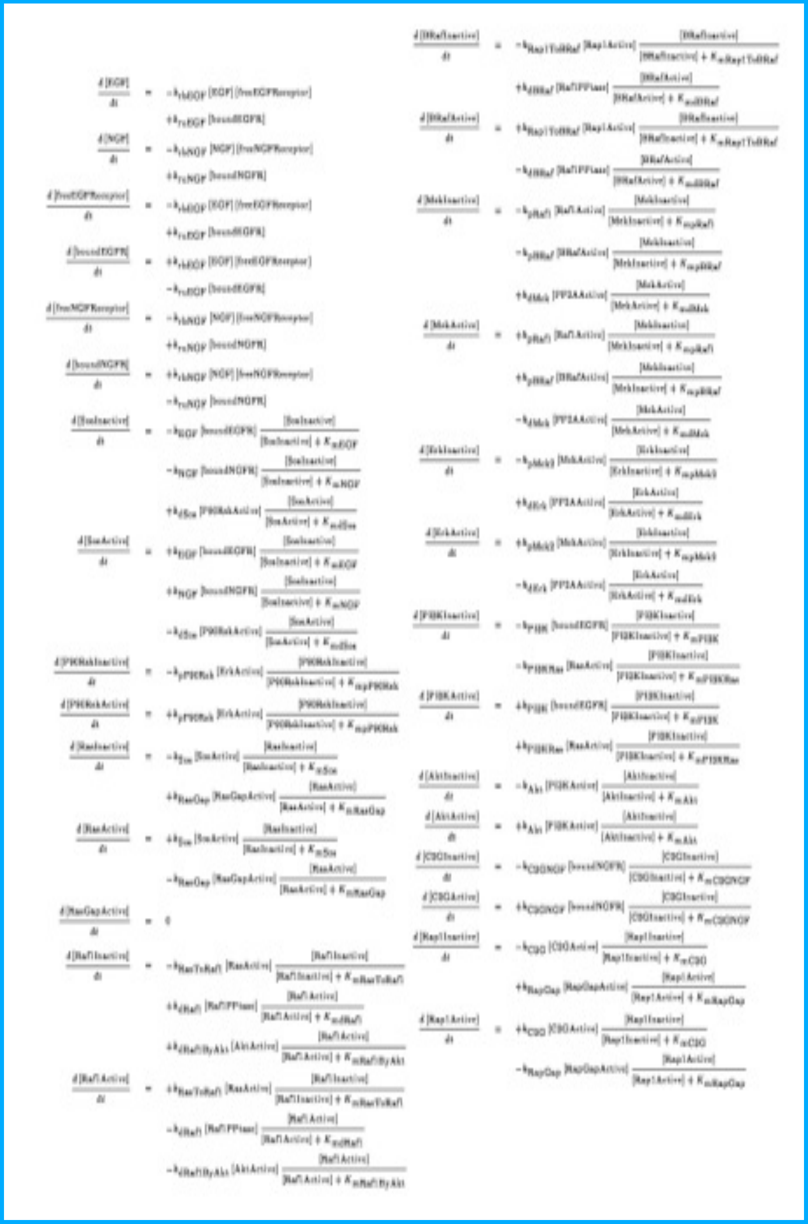
Systems Biology: Cell Protein Reactions

Kevin Brown, Rick Cerione





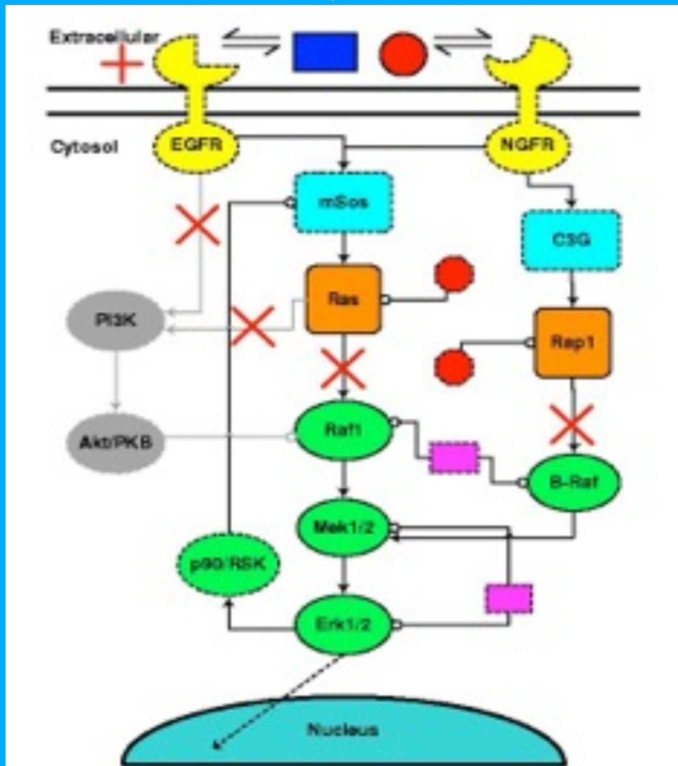


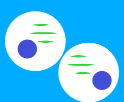


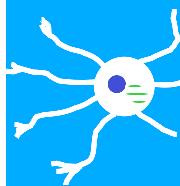
48 Parameter Fit!

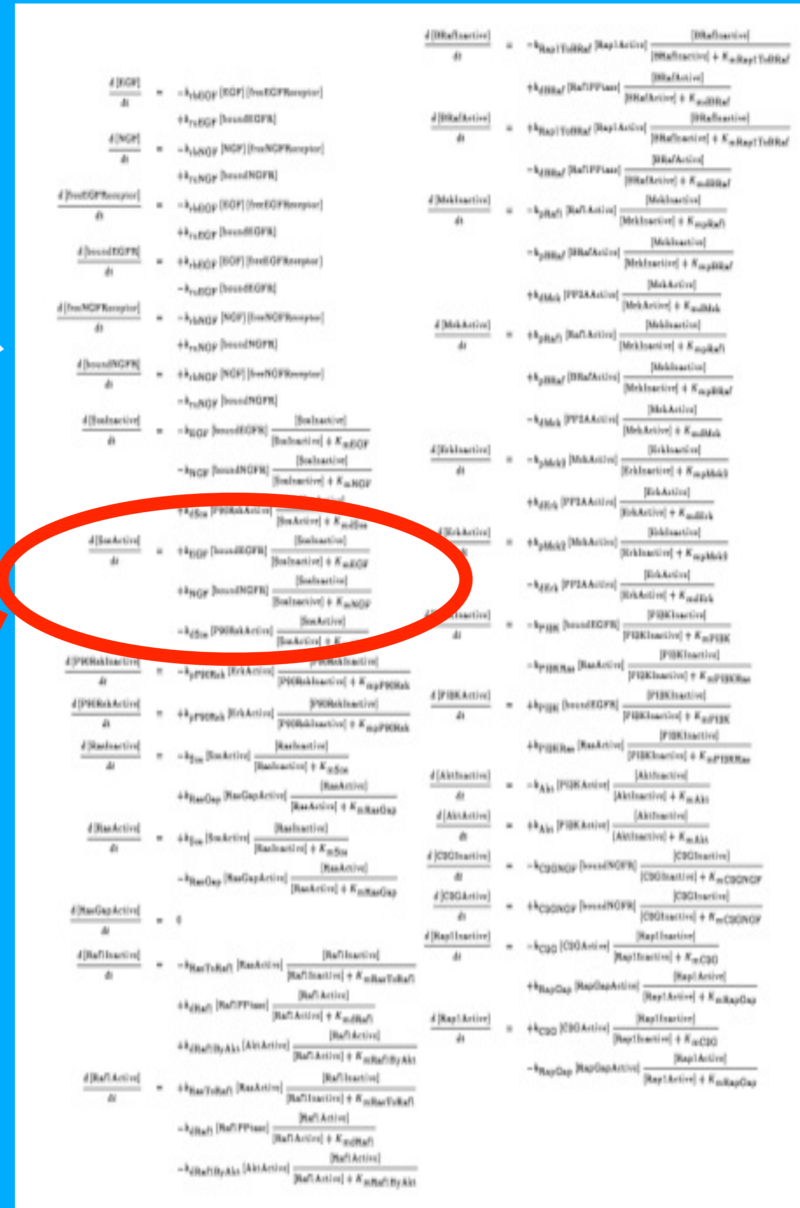
Systems Biology: Cell Protein Reactions

Kevin Brown, Rick Cerione





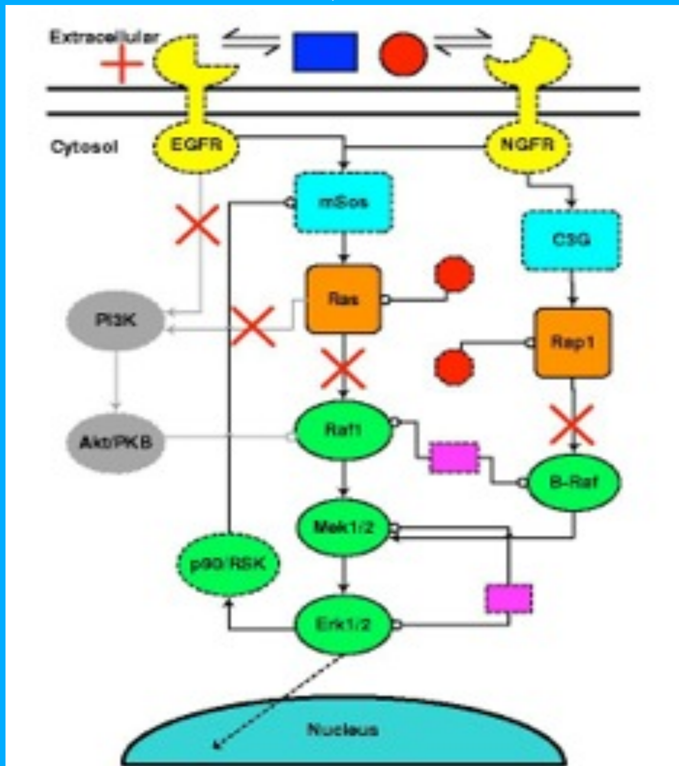


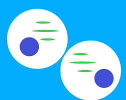


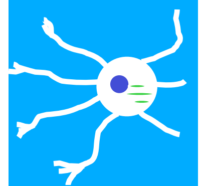
48 Parameter Fit!

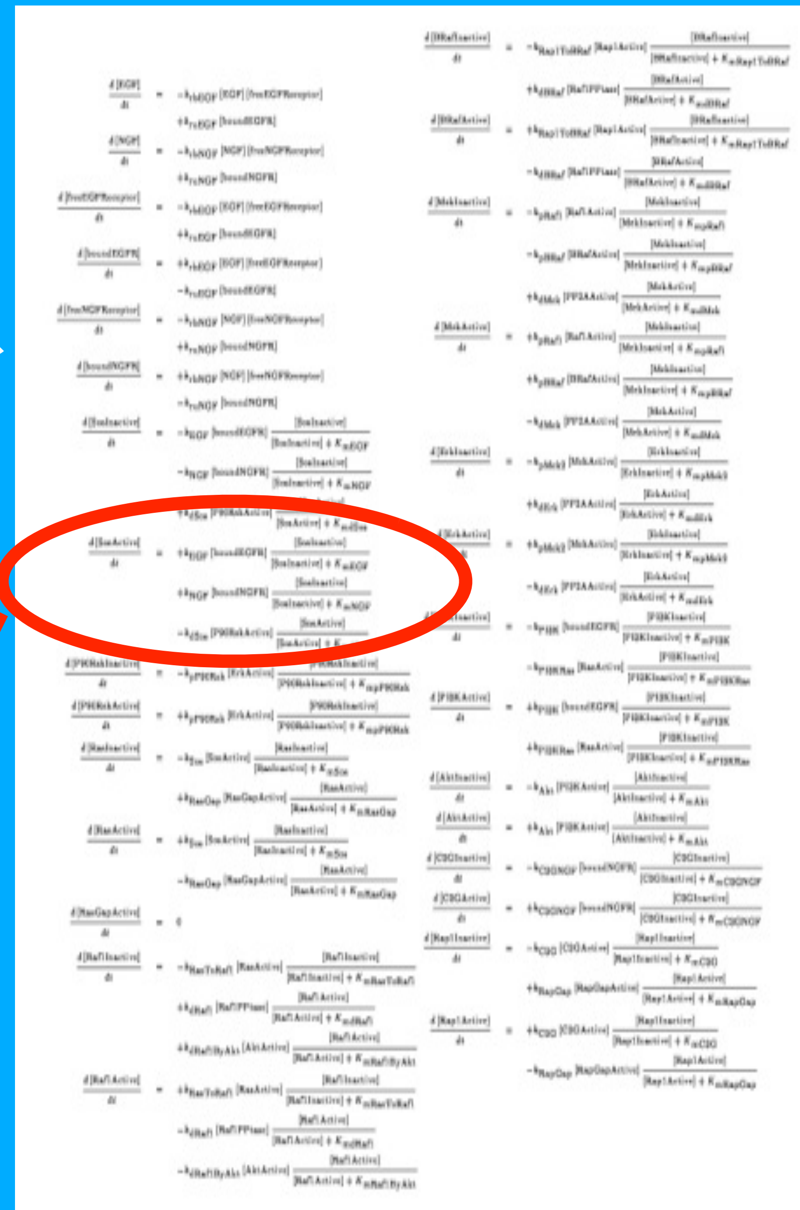
Systems Biology: Cell Protein Reactions

Kevin Brown, Rick Cerione







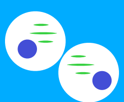


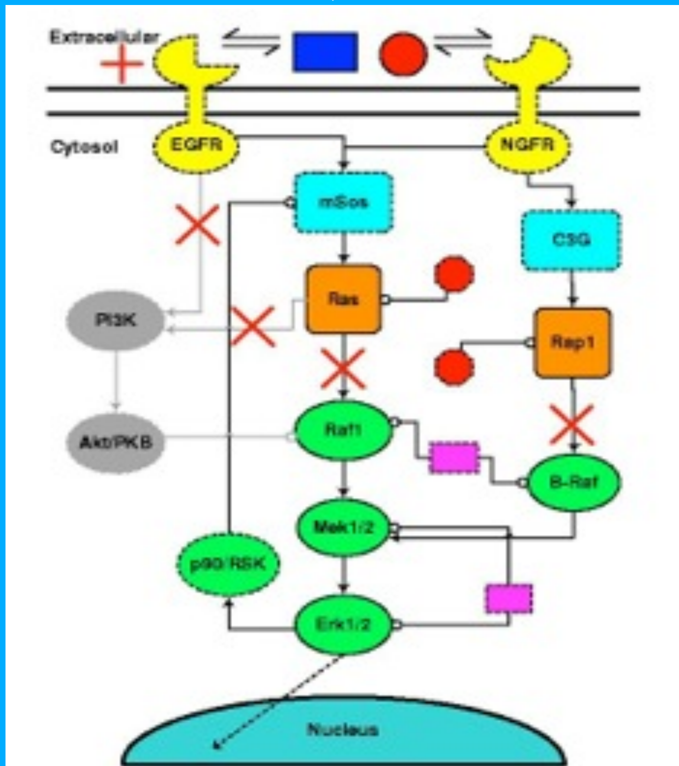
$$
\begin{aligned}
\frac{d\,[\mathrm{SosActive}]}{dt} = & +k_{\mathrm{EGF}}\,[\mathrm{boundEGFR}]\,\frac{[\mathrm{SosInactive}]}{[\mathrm{SosInactive}]+K_{\mathrm{mEGF}}} \\
& +k_{\mathrm{NGF}}\,[\mathrm{boundNGFR}]\,\frac{[\mathrm{SosInactive}]}{[\mathrm{SosInactive}]+K_{\mathrm{mNGF}}} \\
& -k_{\mathrm{dSos}}\,[\mathrm{P90RskActive}]\,\frac{[\mathrm{SosActive}]}{[\mathrm{SosActive}]+K_{\mathrm{mdSos}}}
\end{aligned}
$$

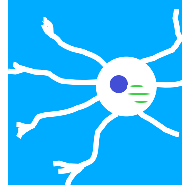
48 Parameter Fit!

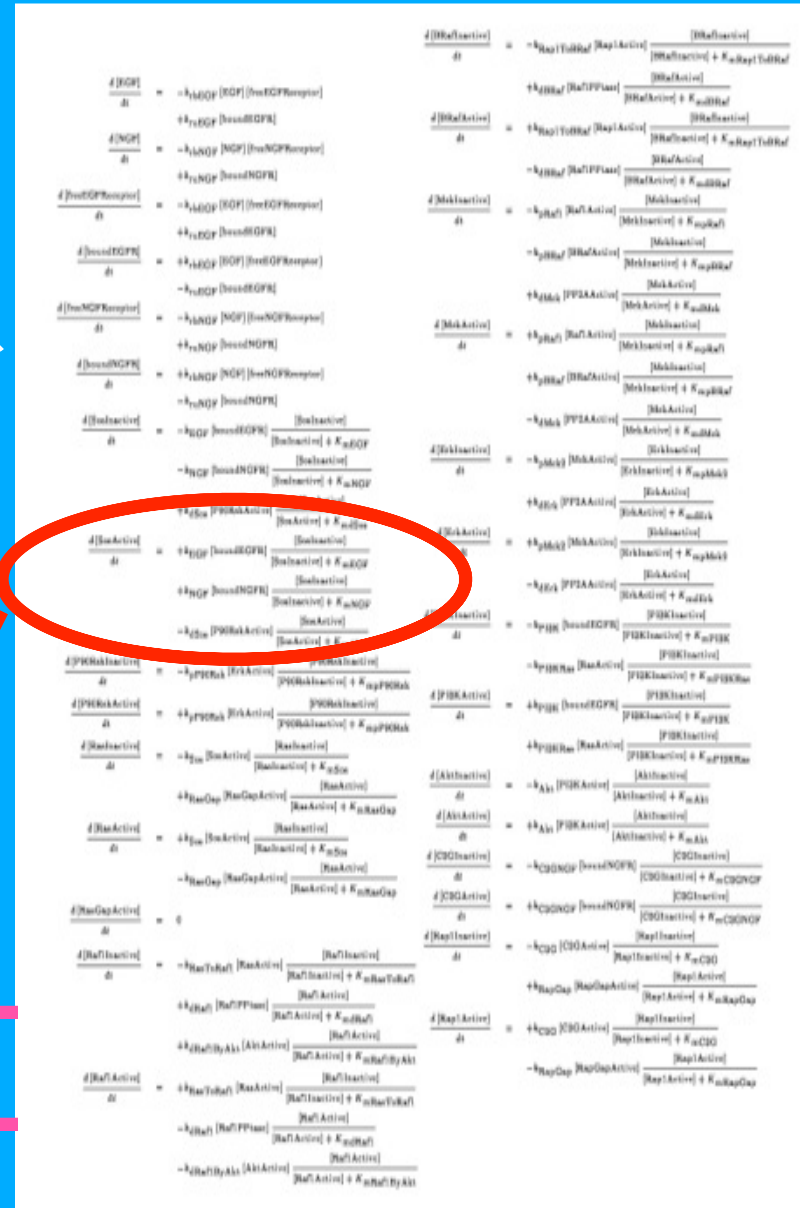
Systems Biology: Cell Protein Reactions

Kevin Brown, Rick Cerione





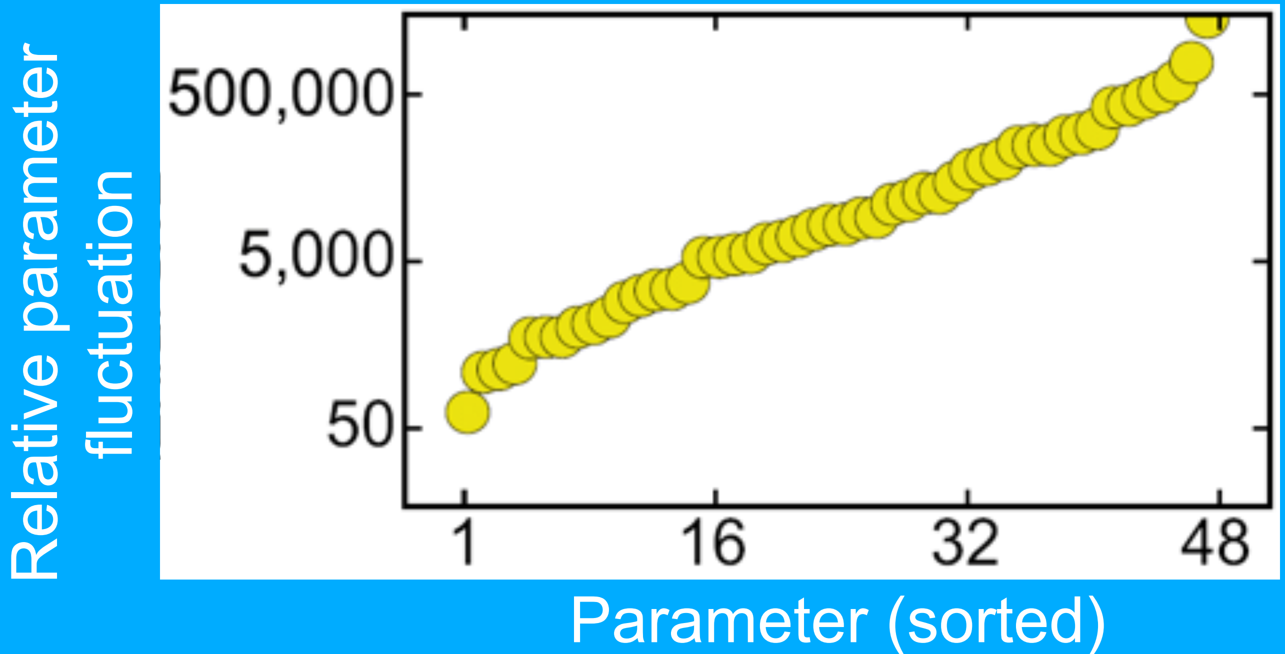




$$\frac{d\,[\mathrm{SosActive}]}{dt} = +k_{\mathrm{EGF}}\,[\mathrm{boundEGFR}]\,\frac{[\mathrm{SosInactive}]}{[\mathrm{SosInactive}] + K_{\mathrm{mEGF}}} + k_{\mathrm{NGF}}\,[\mathrm{boundNGFR}]\,\frac{[\mathrm{SosInactive}]}{[\mathrm{SosInactive}] + K_{\mathrm{mNGF}}} - k_{\mathrm{dSos}}\,[\mathrm{P90RskActive}]\,\frac{[\mathrm{SosActive}]}{[\mathrm{SosActive}] + K_{\mathrm{mdSos}}}$$

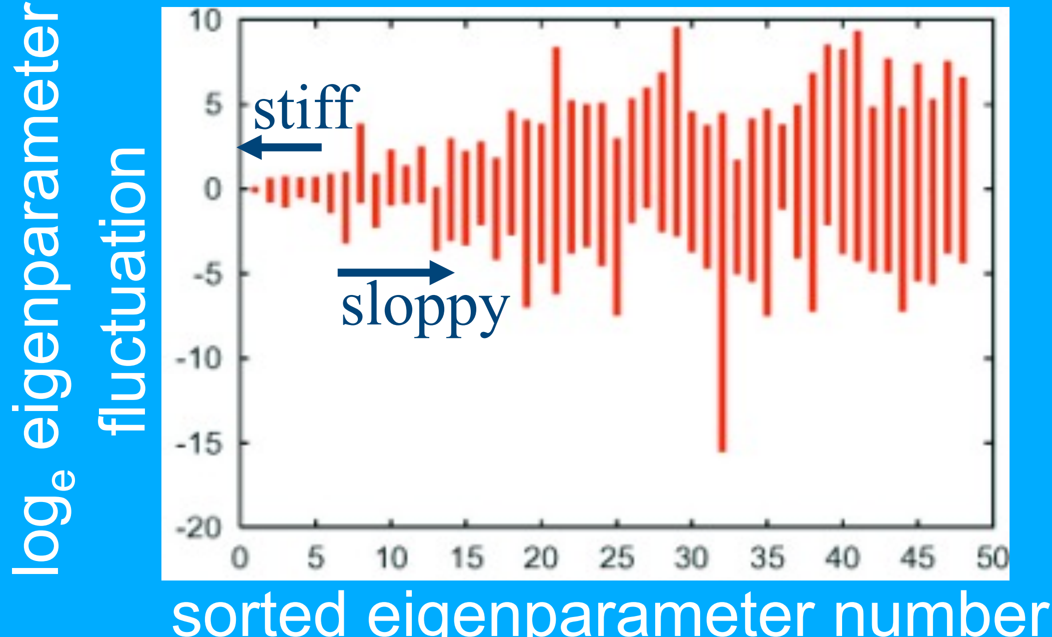
48 Parameter Fit!

Parameters Fluctuate over Enormous Range

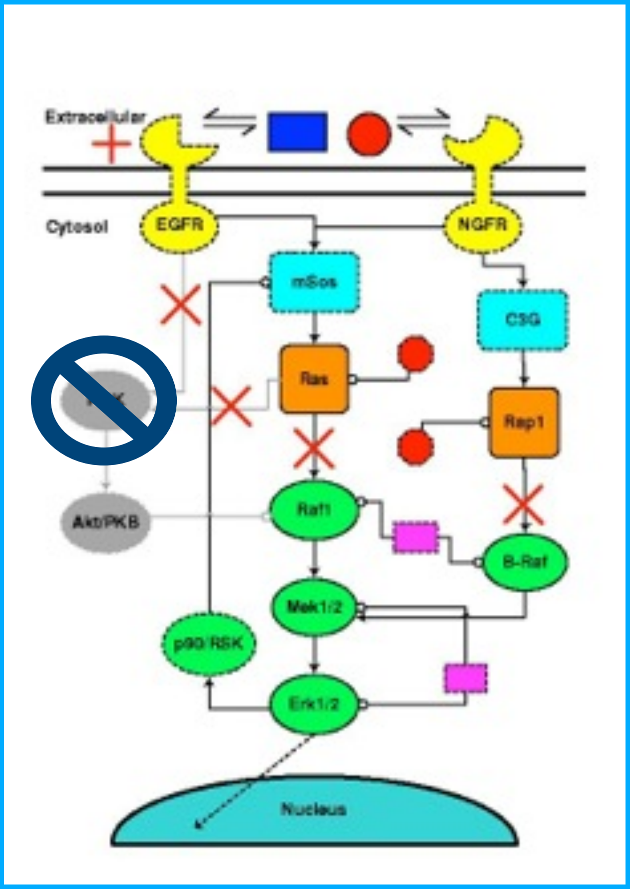


- All parameters vary by minimum factor of 50, some by a million
- Not robust: four or five "stiff" linear combinations of parameters; 44 sloppy

Are predictions possible?

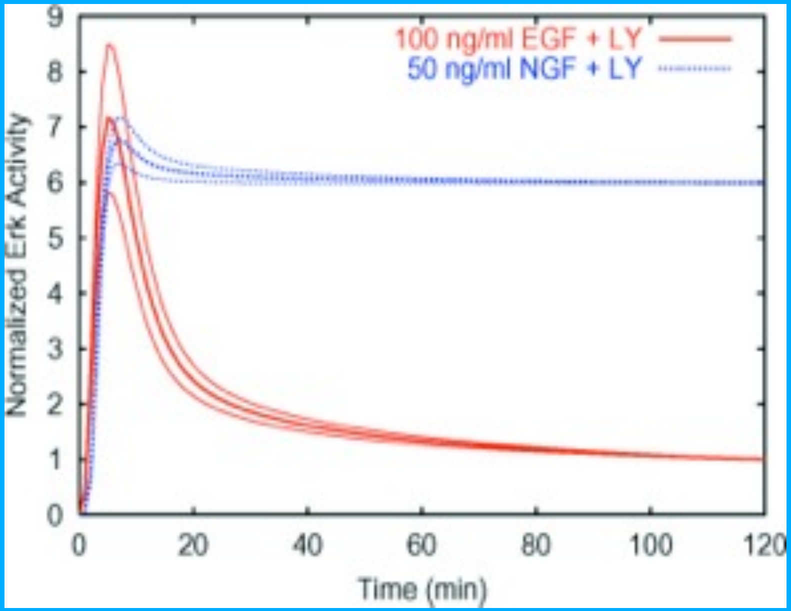


Predictions are Possible

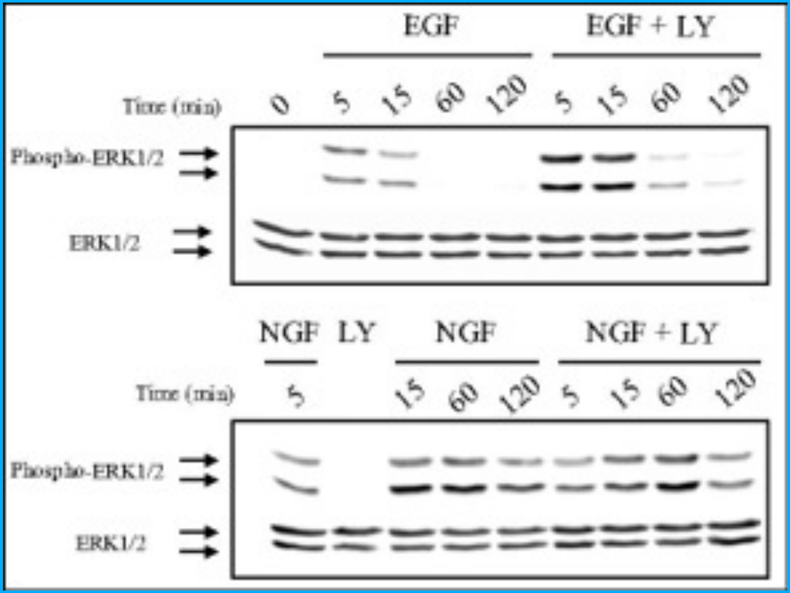


Model predicts that the left branch isn't important

Model Prediction



Brown's Experiment



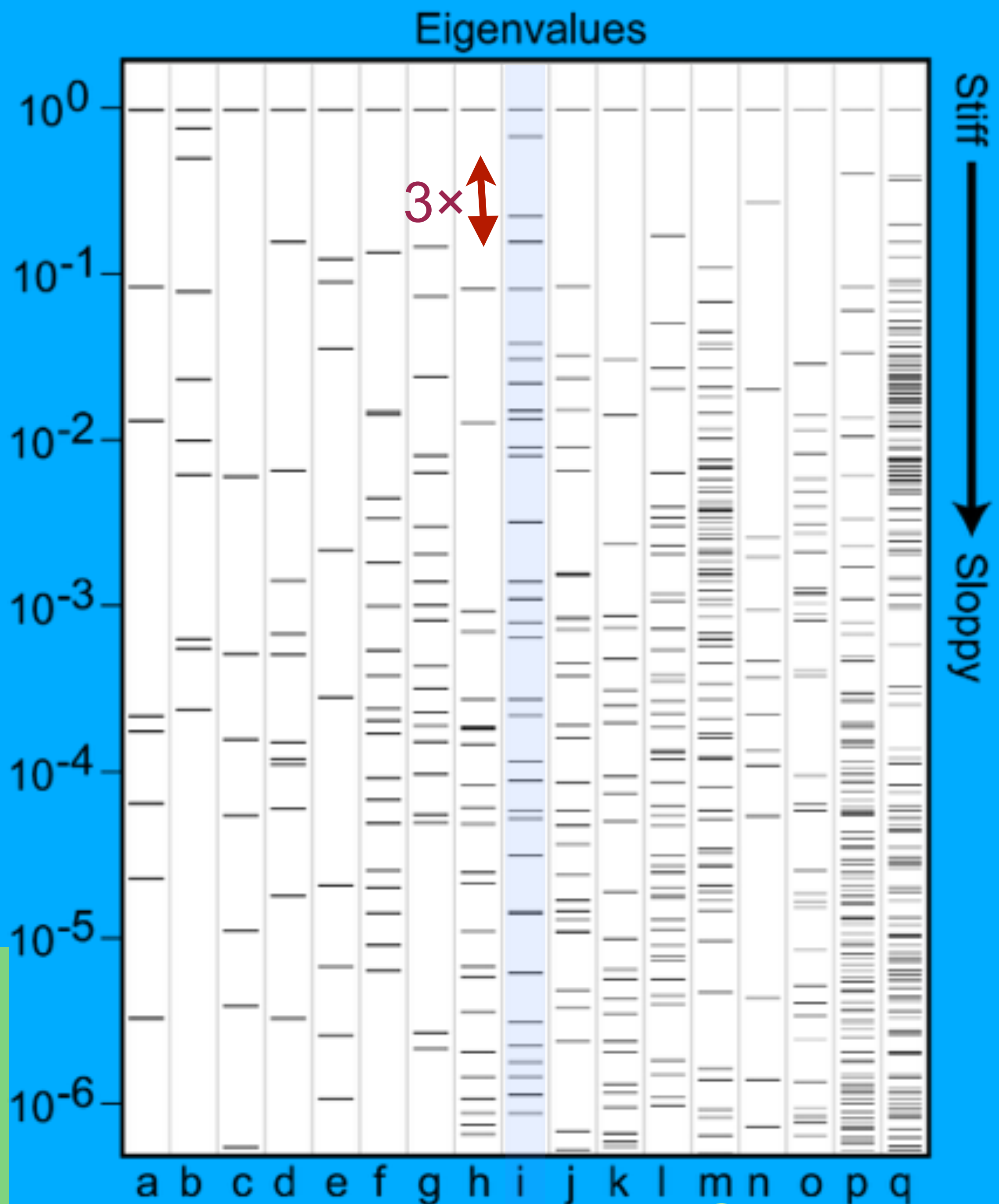
Parameters fluctuate orders of magnitude, but still predictive!

Systems Biology

Seventeen models

(a) eukaryotic cell cycle
(b) Xenopus egg cell cycle
(c) eukaryotic mitosis
(d) generic circadian rhythm
(e) nicotinic acetylcholine intra-receptor dynamics
(f) generic kinase cascade
(g) Xenopus Wnt signaling
(h) Drosophila circadian rhythm
(i) rat growth-factor signaling
(j) Drosophila segment polarity
(k) Drosophila circadian rhythm
(l) Arabidopsis circadian rhythm
(m) in silico regulatory network
(n) human purine metabolism
(o) Escherichia coli carbon metabolism
(p) budding yeast cell cycle
(q) rat growth-factor signaling

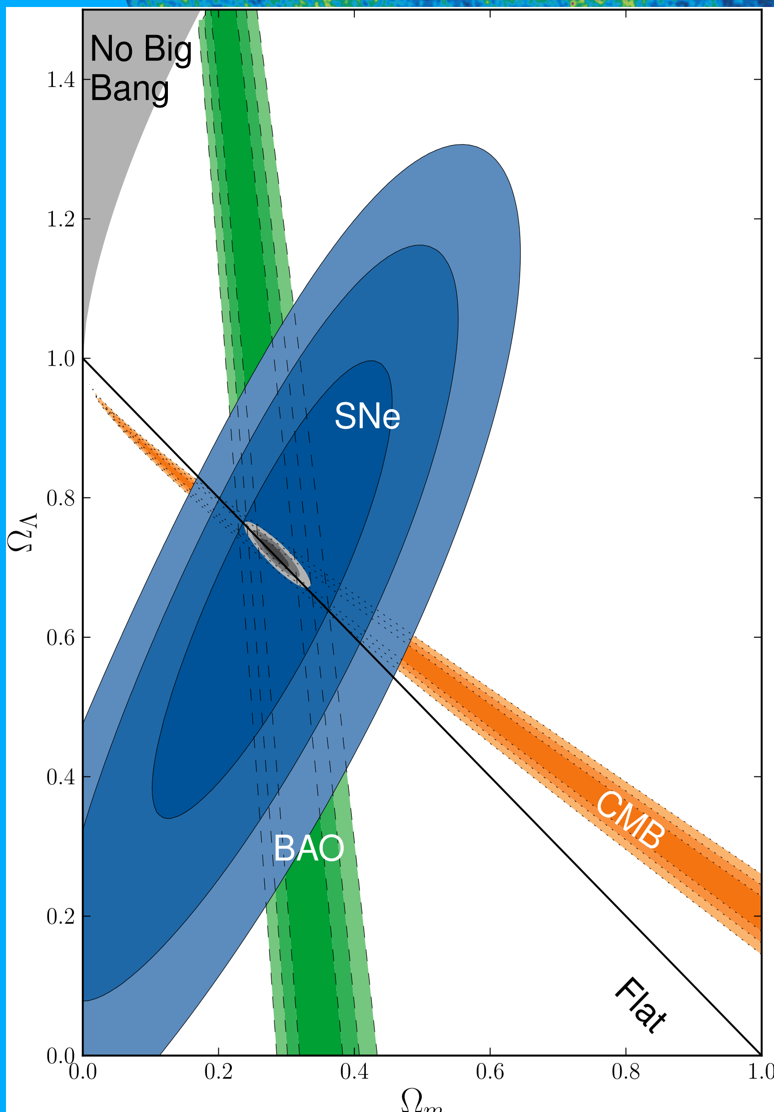
Enormous Ranges of Eigenvalues
(3^{48} is a big number)
Sloppy Range ~ $\sqrt{\lambda}$

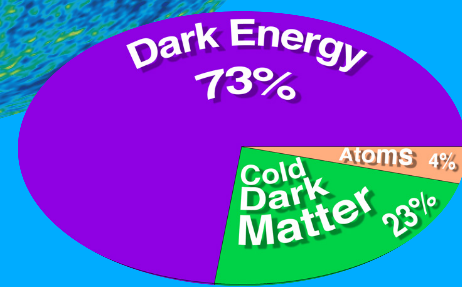


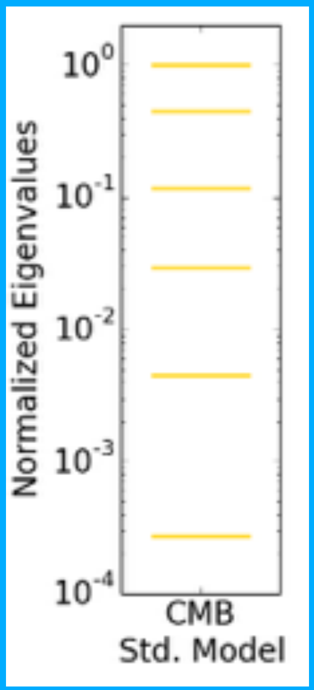
Ryan Gutenkunst

The Universe

ΛCDM fit for cosmic microwave background radiation







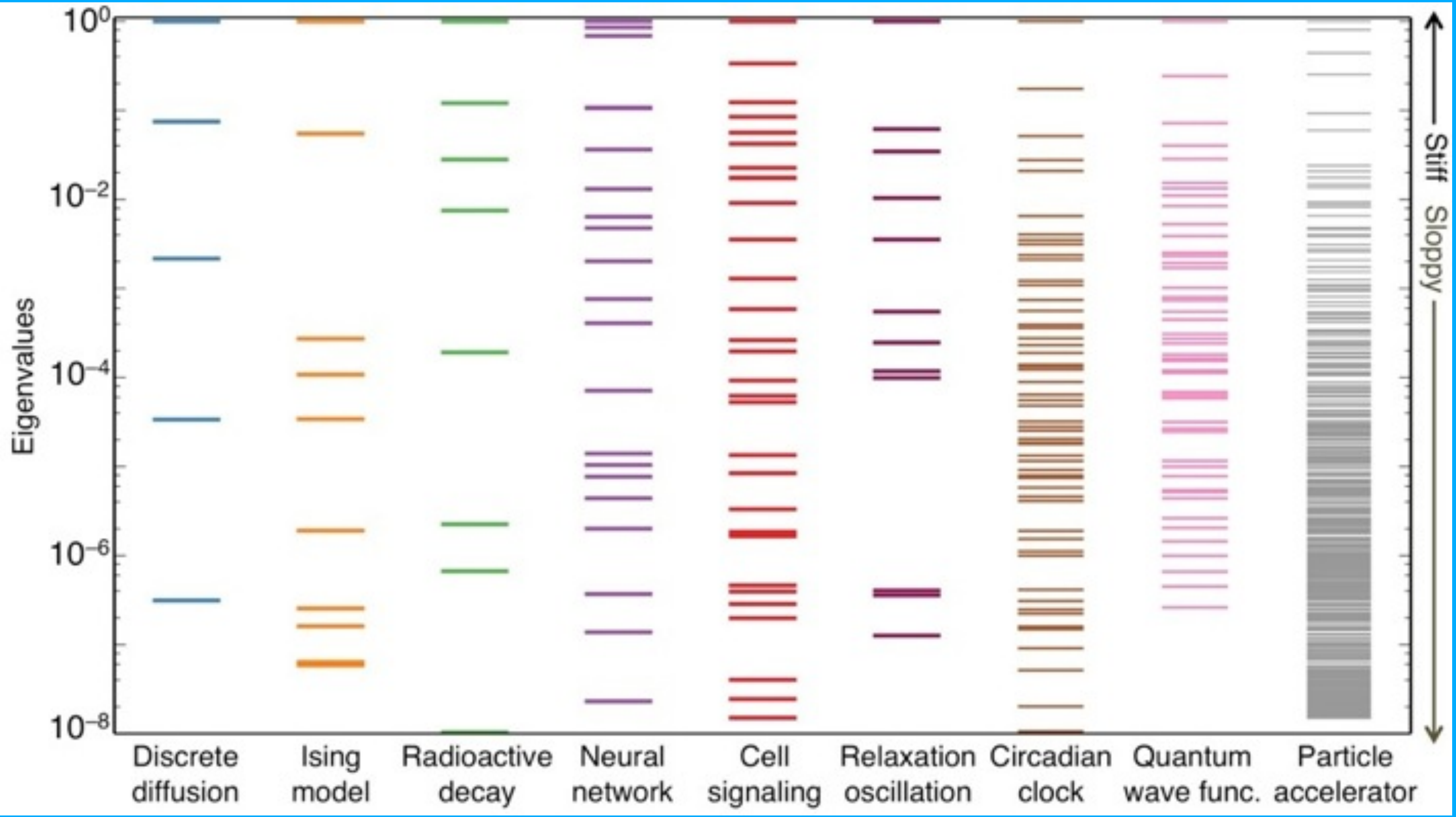
Universe is flat, mostly unknown dark stuff

- Six parameter ΛCDM model is sloppy fit to CMB; SNe and BAO determine
- More general models introduce worse degeneracies

Katherine Quinn, Michael Niemack, Francesco De Bernardis

Sloppy Universality Outside Bio

Waterfall, Gutenkunst, Chachra, Machta, Clement



Enormous range of eigenvalues; Roughly equal density in log; Observed in broad range of systems

The Model Manifold

Mark Transtrum, Ben Machta

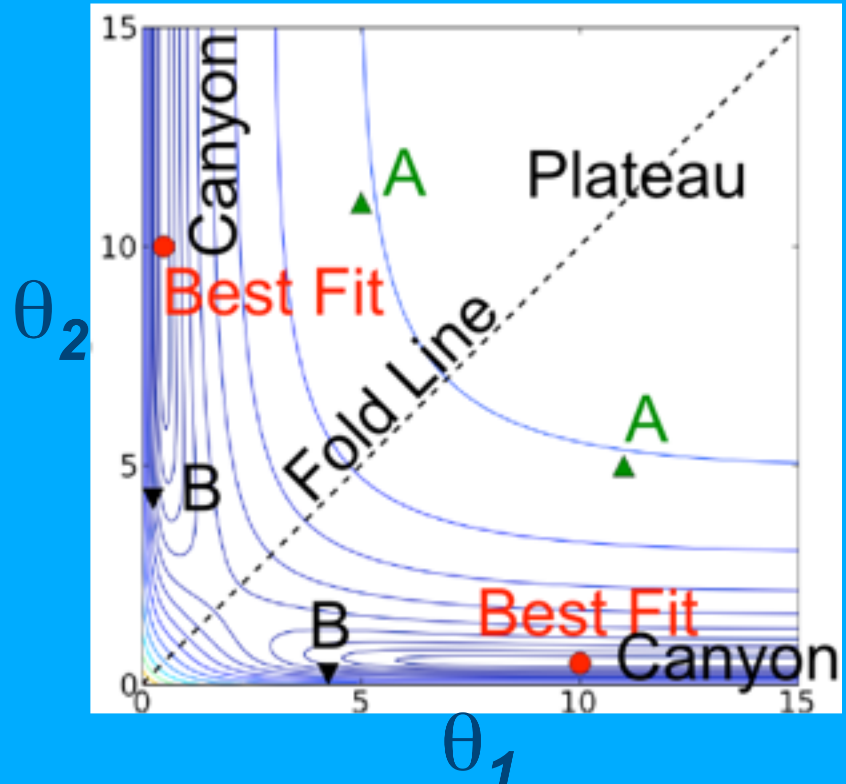
Two exponentials θ_α fit to three data points y_n,

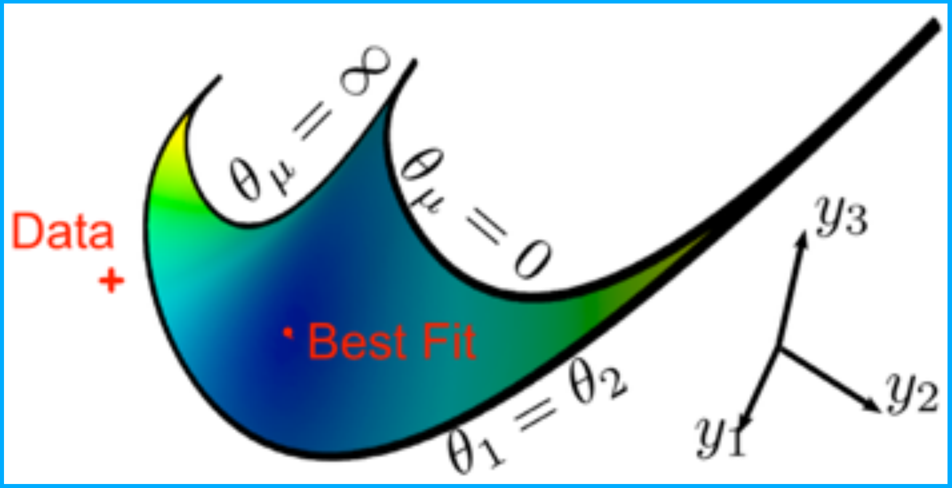
$$y_n = \exp(-\theta_1\, t_n) + \exp(-\theta_2\, t_n)$$

Parameter space

Stiff and sloppy directions

Canyons, Plateaus





Data space

Manifold of model predictions

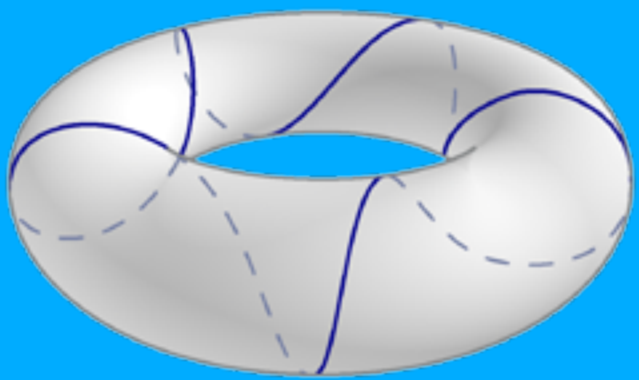
Parameters as coordinates

Model boundaries $\theta_n = \infty$, θ_m cause Plateaus

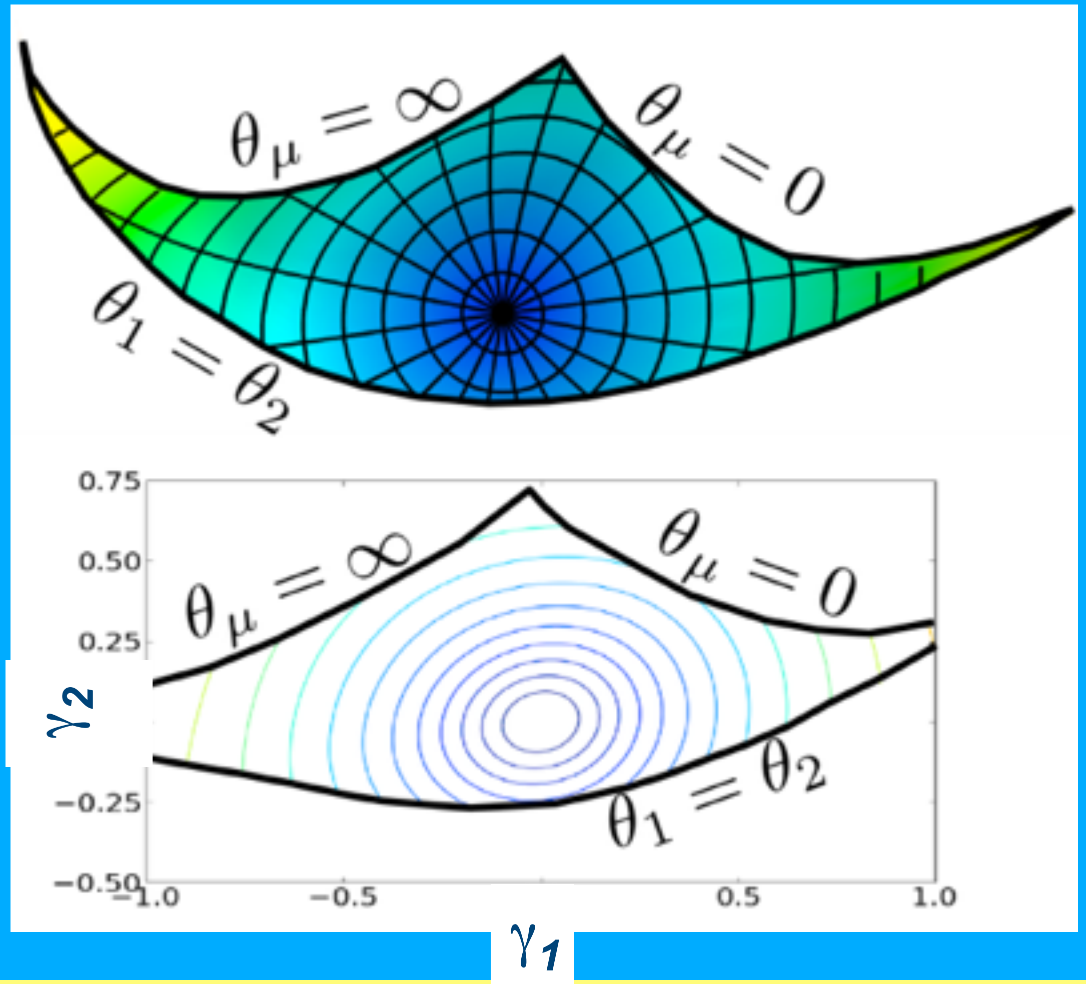
Metric $g_{\mu\nu}$ from distance to data

Geodesics

"Straight line" in curved space
Shortest path between points



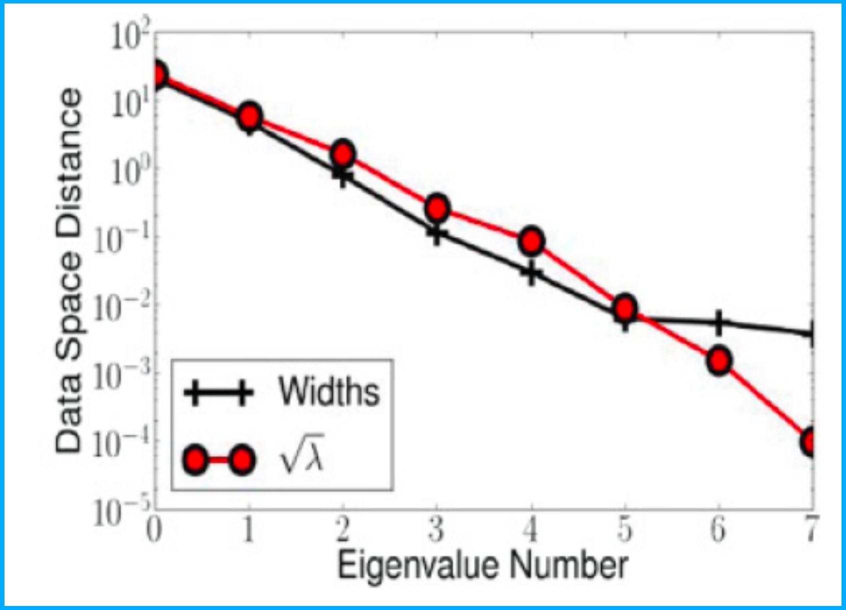
Easy to find cost minimum using polar *geodesic coordinates*
γ_1, γ_2



Cost contours in geodesic coordinates nearly concentric circles!
Use this for algorithms…

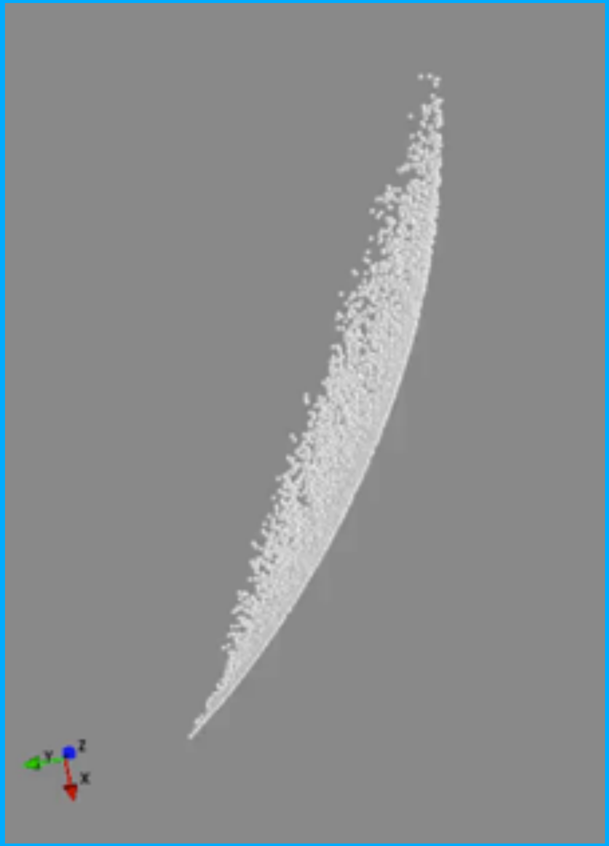
The Model Manifold is a *Hyper-Ribbon*

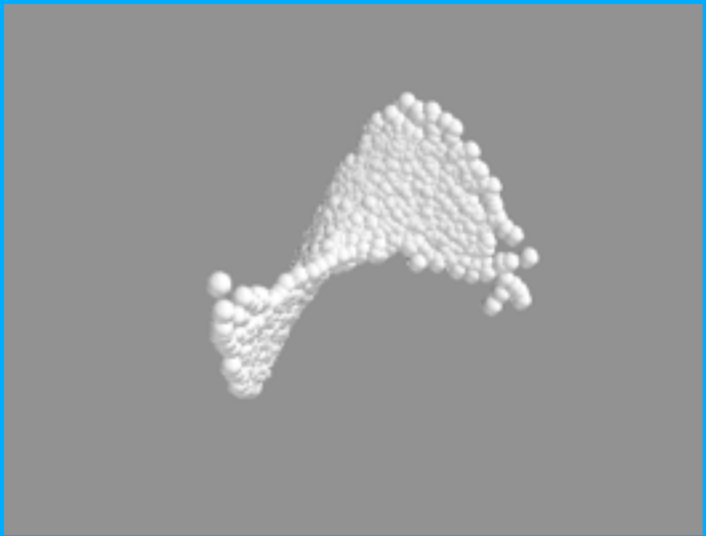
- Hyper-ribbon: object that is longer than wide, wider than thick, thicker than ...
- Thick directions traversed by stiff eigenparameters, thin as sloppy directions varied.



Widths along geodesics track eigenvalues almost perfectly!

Sum of many exponentials, fit to y(0), y(1) data predictions at y(1/4), y(1/2), y(3/4)

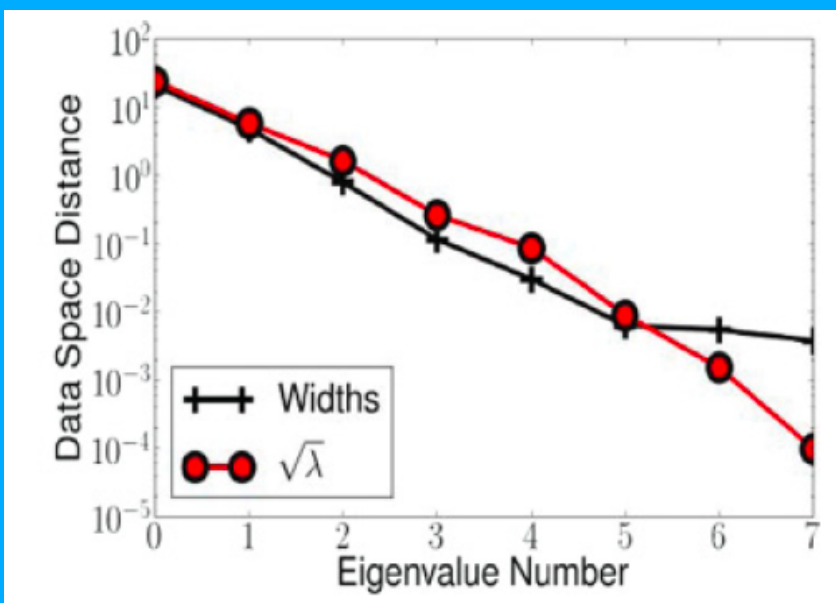




Diffusion equation after three time steps

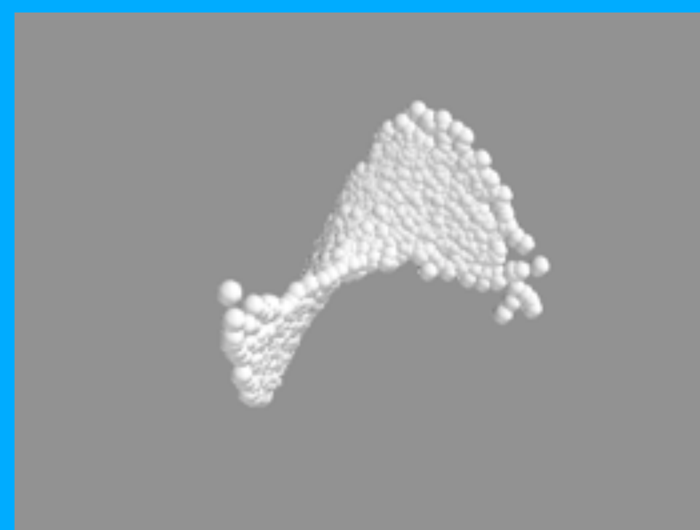
The Model Manifold is a *Hyper-Ribbon*

- Hyper-ribbon: object that is longer than wide, wider than thick, thicker than ...
- Thick directions traversed by stiff eigenparameters, thin as sloppy directions varied.



Widths along geodesics track eigenvalues almost perfectly!

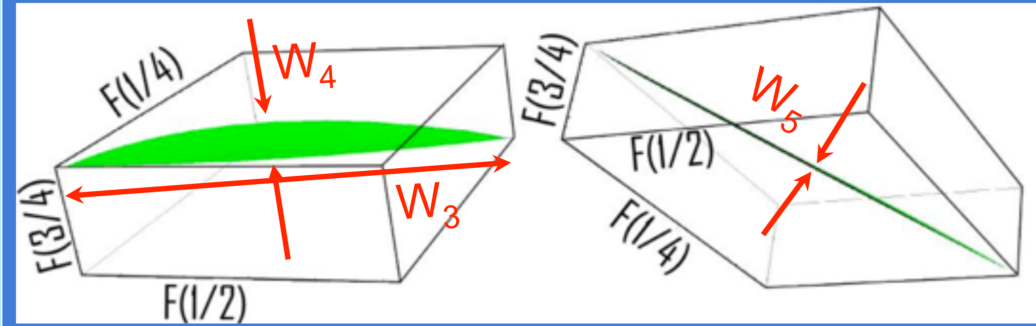
Sum of many exponentials, fit to y(0), y(1) data predictions at y(1/4), y(1/2), y(3/4)



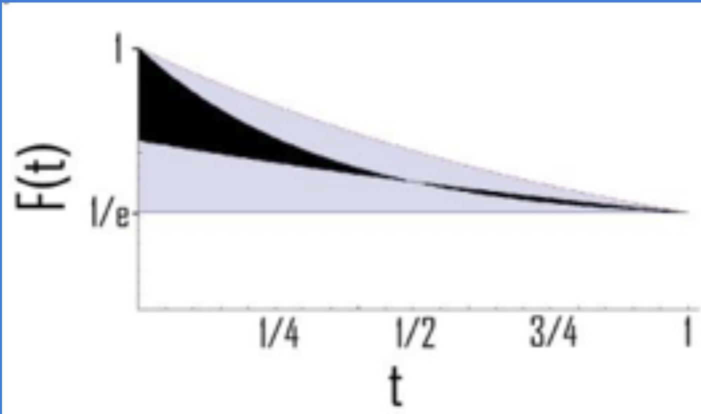
Diffusion equation after three time steps

Hierarchy of widths and curvatures

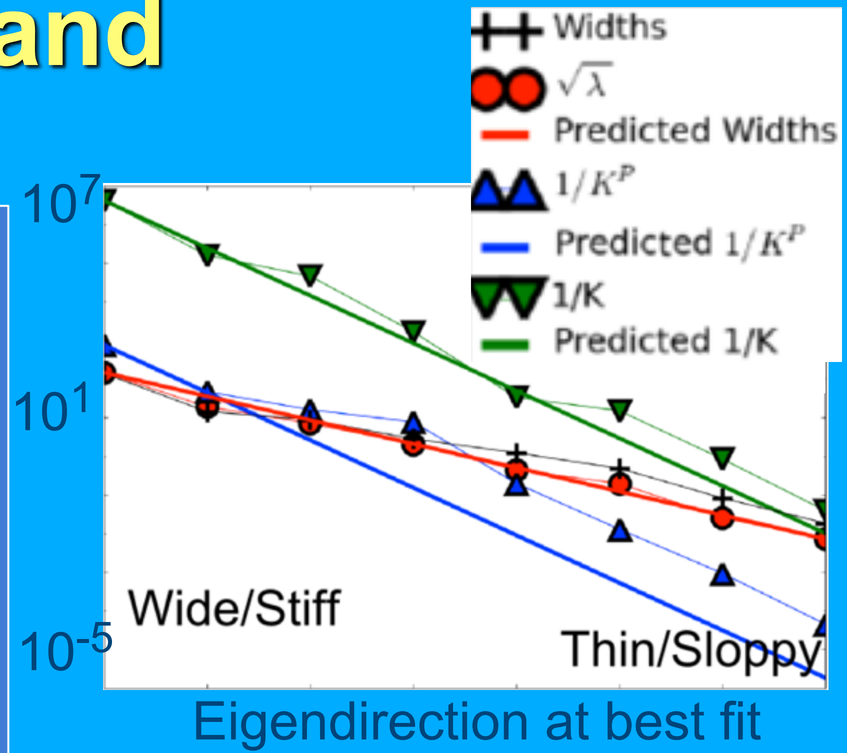
Hierarchy of widths



Cross sections: fixing f at 0, ½, 1



Theorem: interpolation good with many data points
Geometrical convergence



Multi-decade span of widths, curvatures, eigenvalues

Widths ~ √λ sloppy eigs

Parameter curvature
$K^P = 10^3 \times K$
>> extrinsic curvature

Why is it so thin and flat?

Model $f(t,\theta)$ analytic:

$f^{(n)}(t)/n! \le R^{-n}$

Polynomial fit $P_{m-1}(t)$

to $f(t_1),\ldots,f(t_m)$

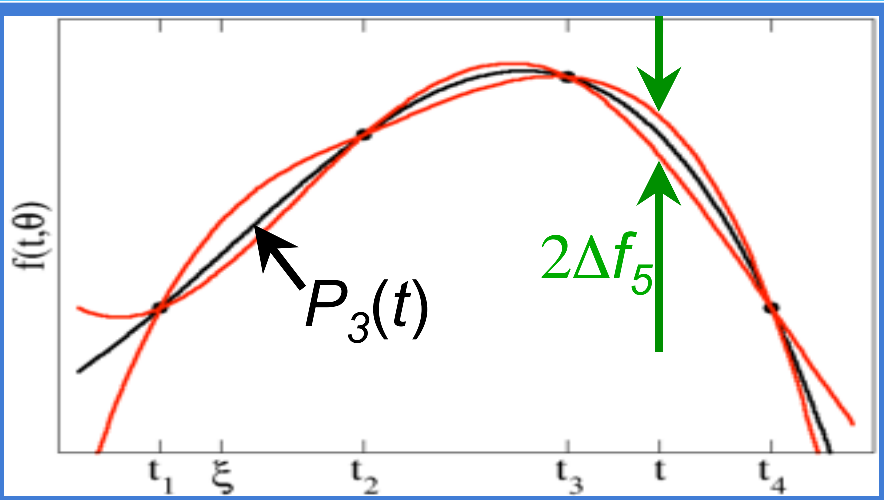
Interpolation convergence theorem

$\Delta f_{m+1} = f(t)-P_{m-1}(t)$

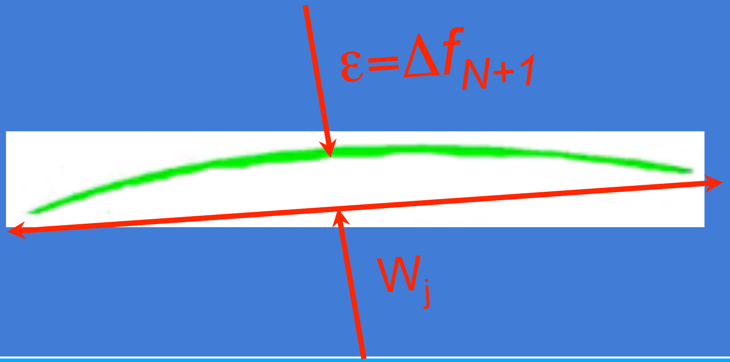
$< (t-t_1)\,(t-t_2)\ldots(t-t_m)\, f^{(m)}(\xi)/m!$

$\sim (\Delta t / R)^m$

More than one data per R

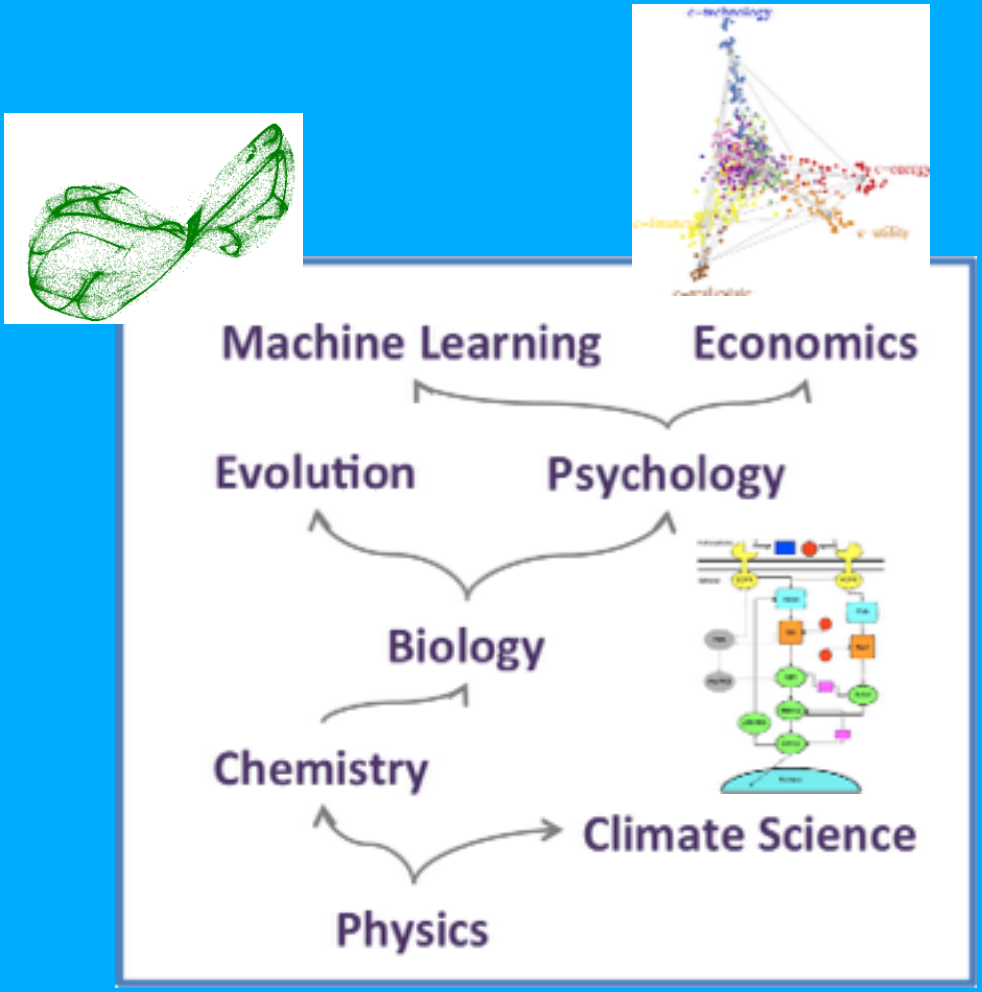


Hyper-ribbon: Cross section constraining m points has width $W_{m+1} \sim \Delta f_{m+1} \sim (\Delta t / R)^m$



Extrinsic flatness: N=M trivially flat, extra data deviates $\varepsilon \sim \Delta f_{N+1}$, so curvature $K \sim \varepsilon / W_j^2 \sim (\Delta t/R)^{N+1-j}/W_j$

Big Sloppiness Questions.



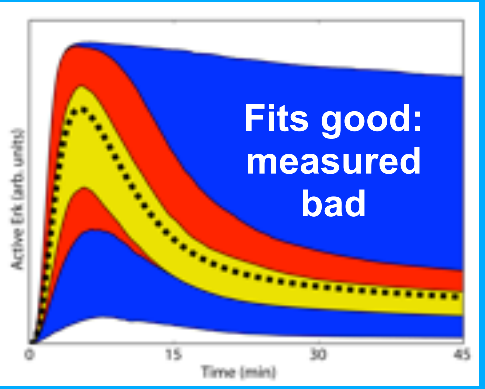
Science appears to rely on *parameter compression*: only a few stiff parameter combinations matter.

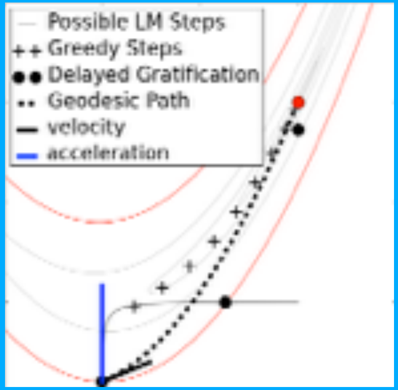
- How is our general explanation for the hierarchy of stiffness (interpolation theory) related to that in physics (small parameters)?
- Without sloppiness, science is hard. (If all the details matter, can't work toward the answer.) Is science selecting sloppy problems, or is everything sloppy?

Sloppy Applications

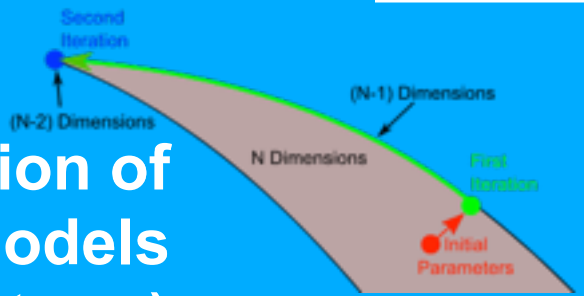
Several applications emerge

A. Fitting data vs. measuring parameters (Gutenkunst)

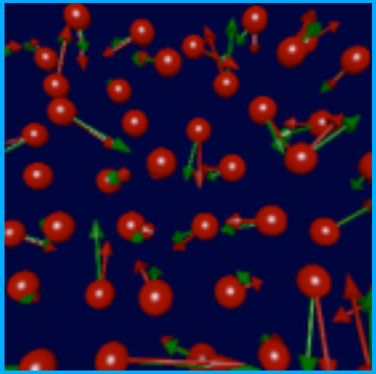


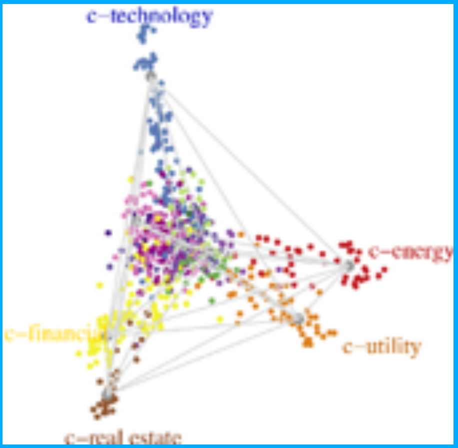


B. Finding best fits by geodesic acceleration (Transtrum)



C. Generation of reduced models (Transtrum)





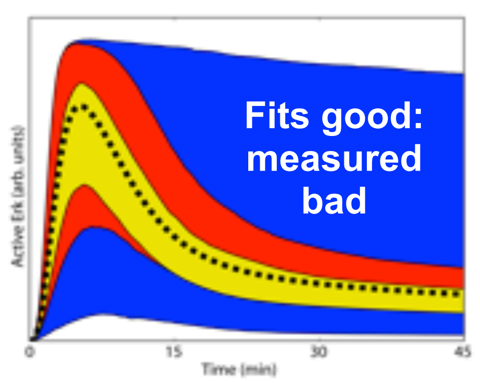
D. Unsupervised learning: Economic sectors from stock prices

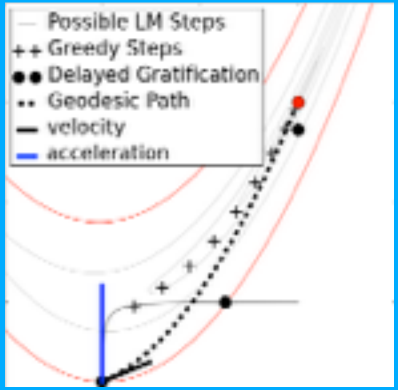
E. Estimating systematic errors: DFT and interatomic potentials (Jacobsen et al.)

Sloppy Applications

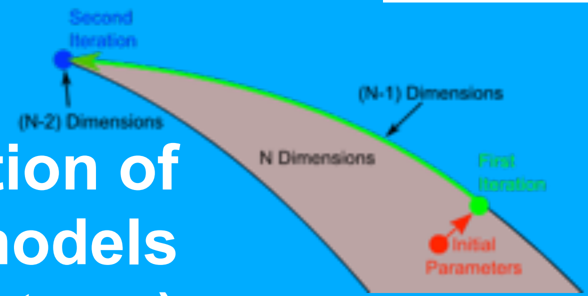
Several applications emerge

A. Fitting data vs. measuring parameters (Gutenkunst)

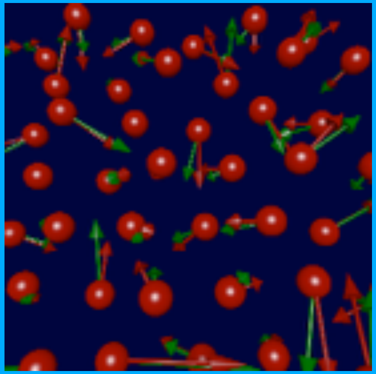




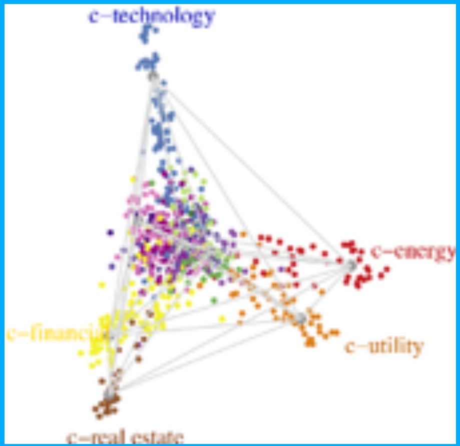
B. Finding best fits by geodesic acceleration (Transtrum)



C. Generation of reduced models (Transtrum)



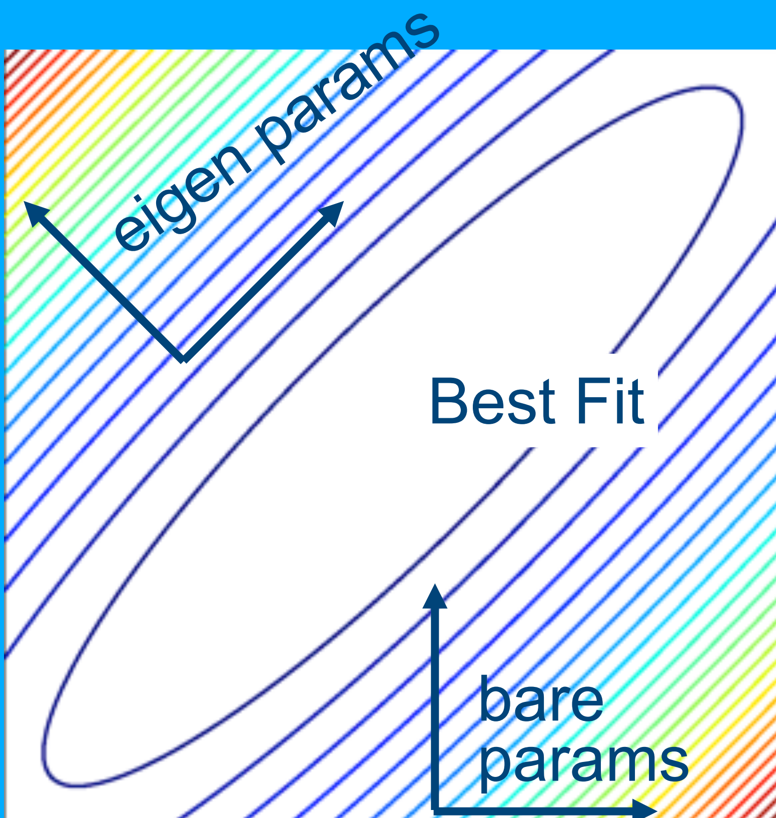
D. Unsupervised learning: Economic sectors from stock prices

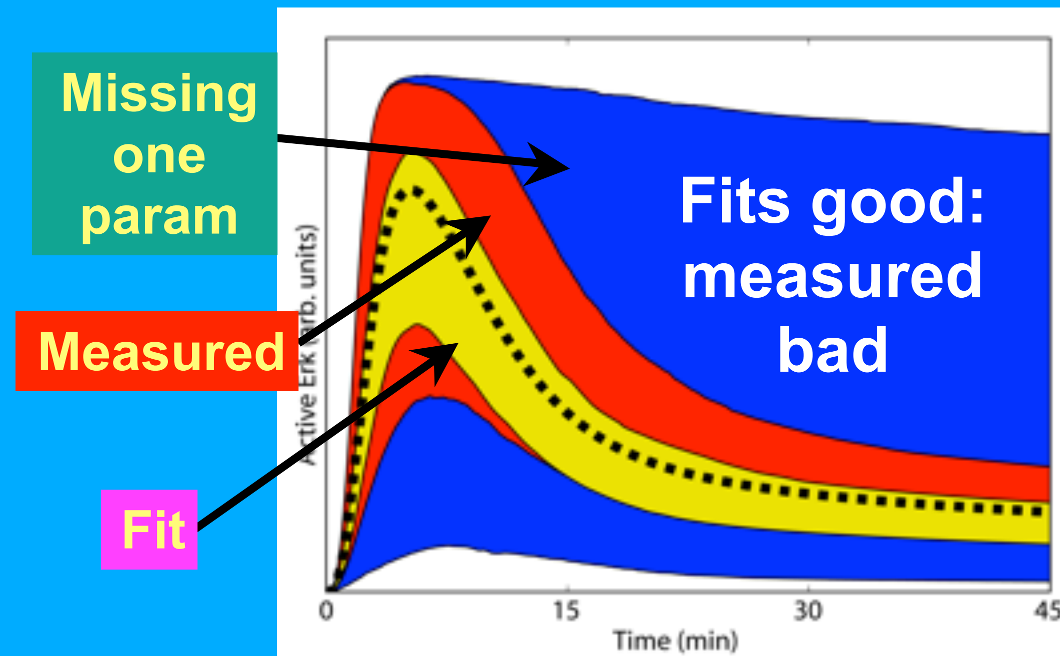


E. Estimating systematic errors: DFT and interatomic potentials (Jacobsen et al.)

A. Are rate constants useful?

Fits vs. measurements

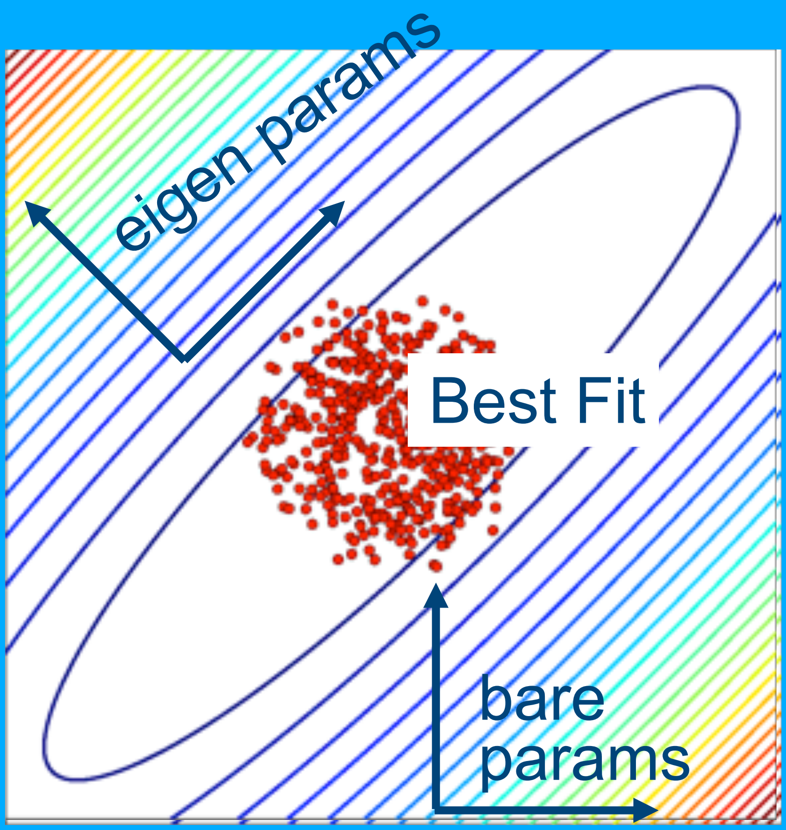


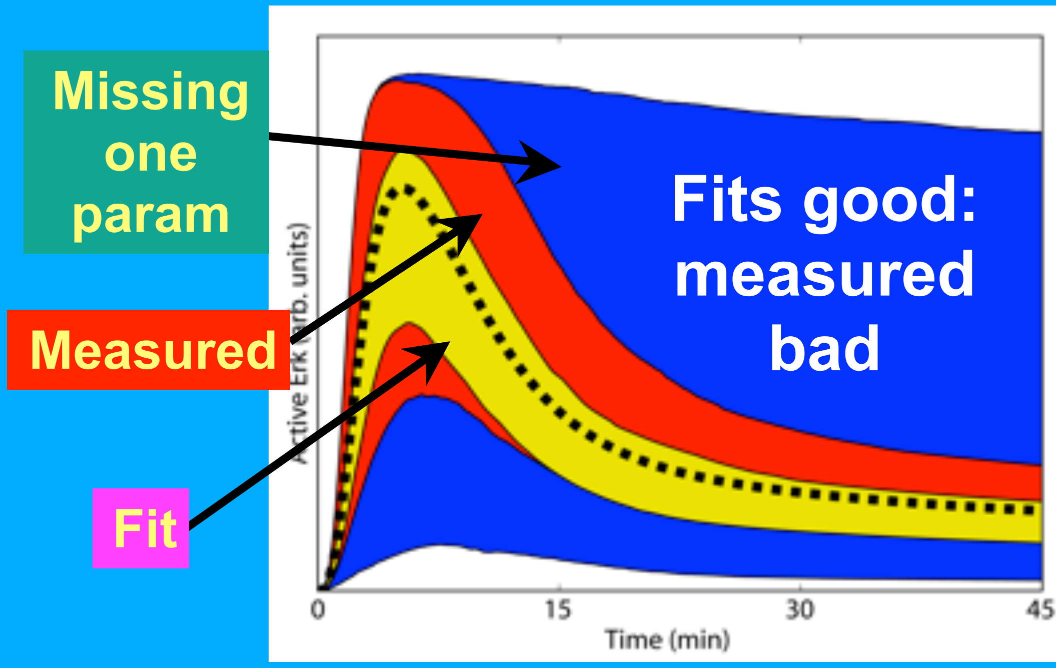


Monte Carlo (anharmonic)

- Easy to Fit (14 expts); Measuring huge job (48 params, 25%)
- One missing parameter measurement = No predictivity
- Sloppy Directions = Enormous Fluctuations in Parameters
- Sloppy Directions often do not impinge on predictivity

A. Are rate constants useful?
Fits vs. measurements



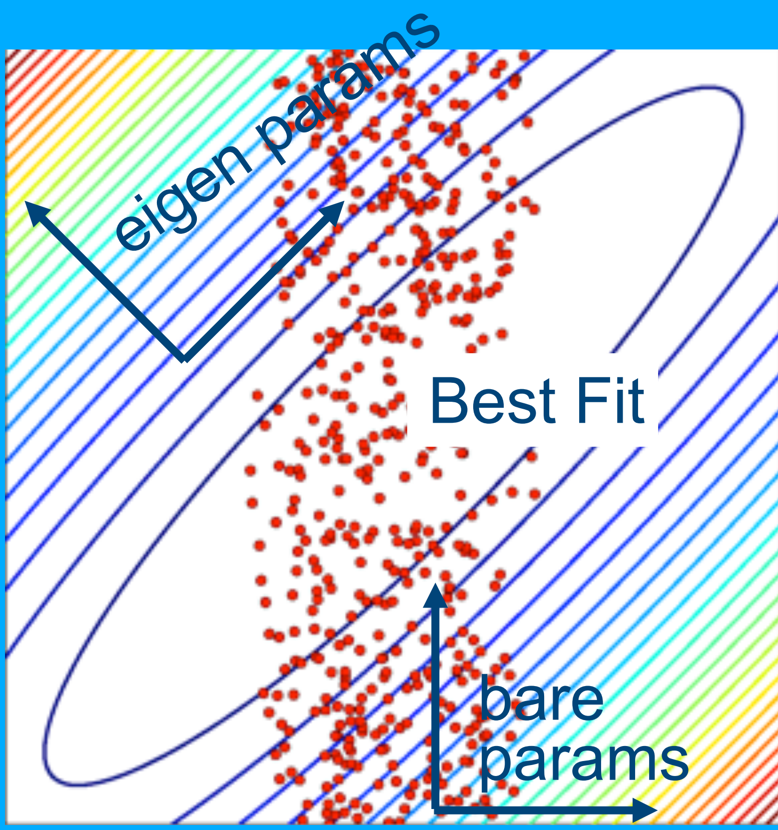


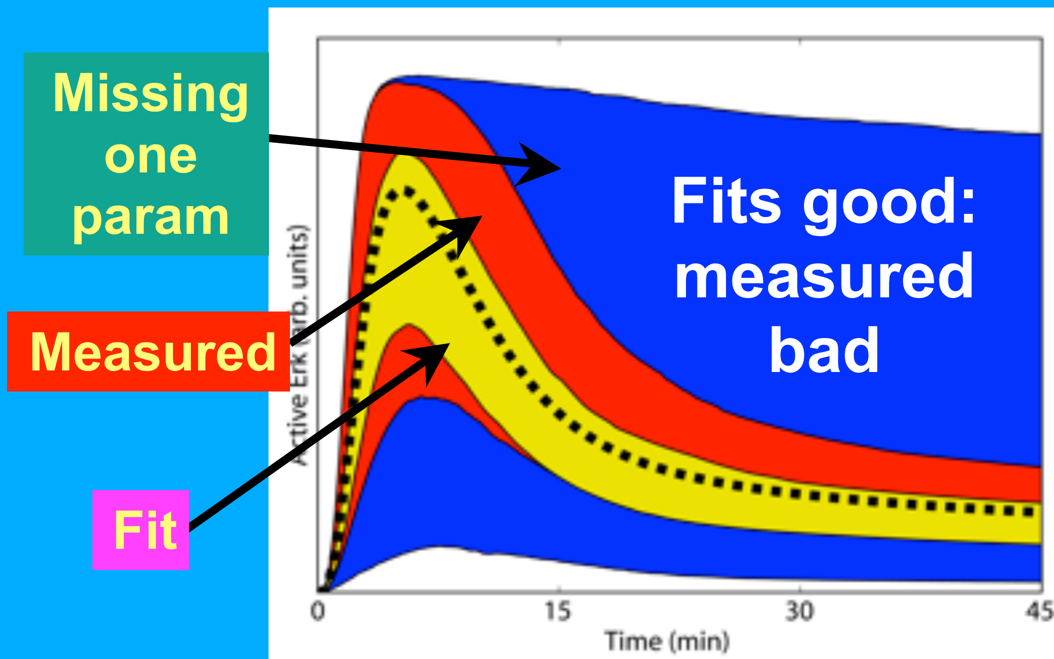
Monte Carlo (anharmonic)

- Easy to Fit (14 expts); Measuring huge job (48 params, 25%)
- One missing parameter measurement = No predictivity
- Sloppy Directions = Enormous Fluctuations in Parameters
- Sloppy Directions often do not impinge on predictivity

A. Are rate constants useful?

Fits vs. measurements



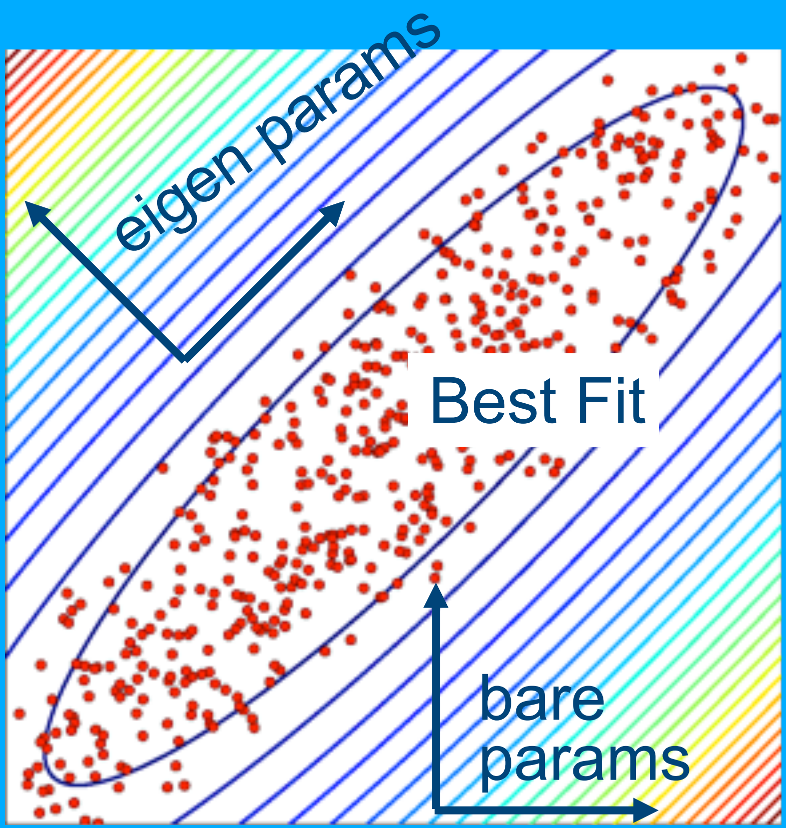


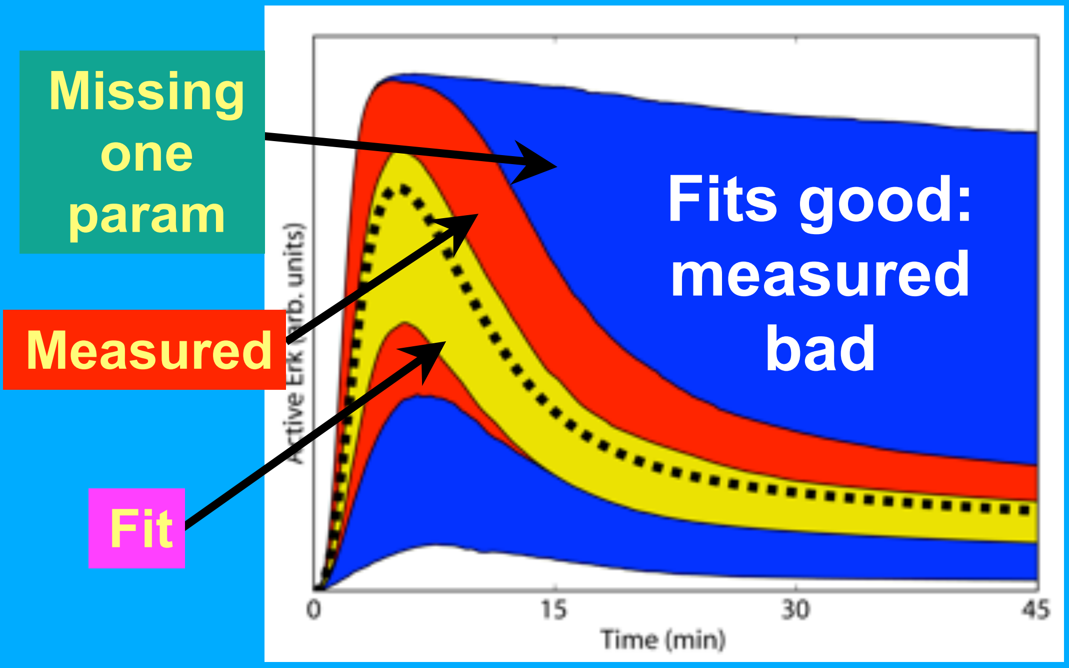
Monte Carlo (anharmonic)

- Easy to Fit (14 expts); Measuring huge job (48 params, 25%)
- One missing parameter measurement = No predictivity
- Sloppy Directions = Enormous Fluctuations in Parameters
- Sloppy Directions often do not impinge on predictivity

A. Are rate constants useful?

Fits vs. measurements

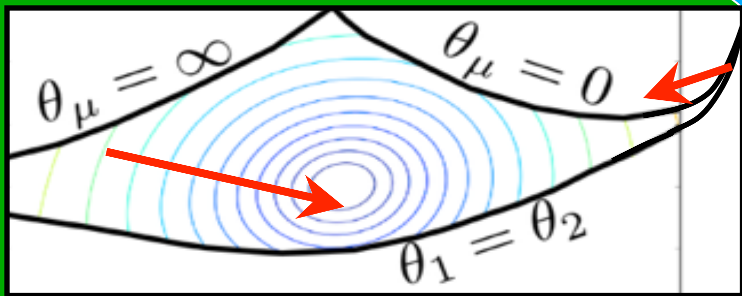




Monte Carlo (anharmonic)

- Easy to Fit (14 expts); Measuring huge job (48 params, 25%)
- One missing parameter measurement = No predictivity
- Sloppy Directions = Enormous Fluctuations in Parameters
- Sloppy Directions often do not impinge on predictivity

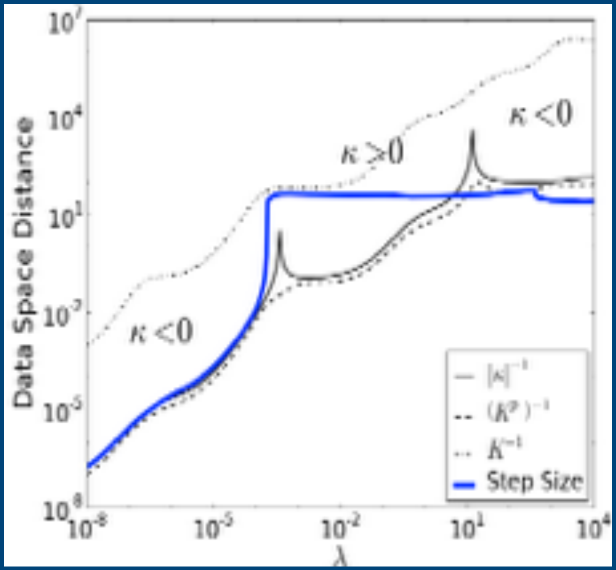
B. Finding best fits: Geodesic acceleration



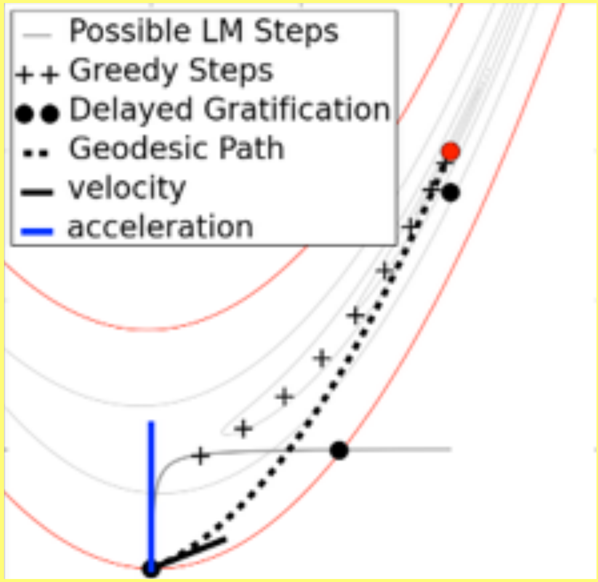
Geodesic Paths nearly circles
Follow local geodesic velocity?

$$\delta\theta^\mu = -g_{\mu\nu}\nabla_\nu C$$

➔ Gauss-Newton
➔ Hits manifold boundary



Model Graph
add weight λ of parameter metric yields Levenberg-Marquardt: Step size now limited by curvature



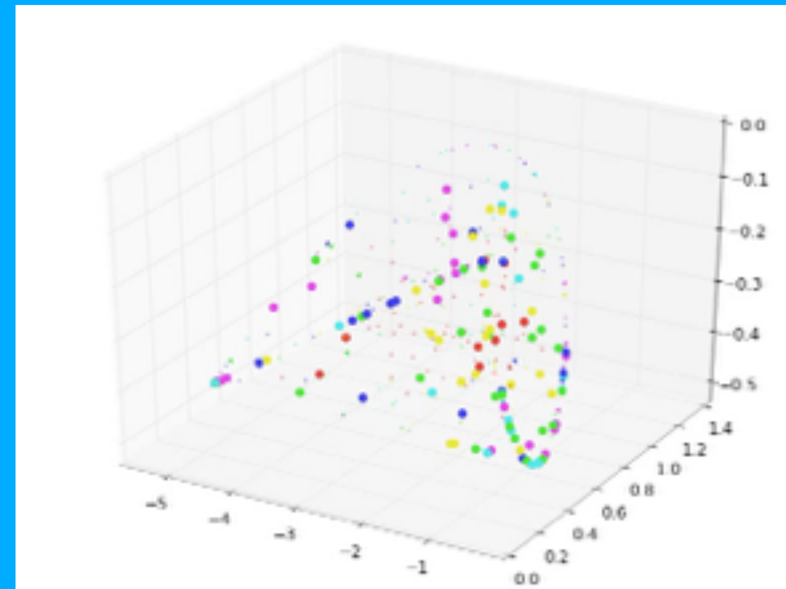
Algorithm	Success Rate	Mean njev	Mean nfev
Traditional LM + accel	65%	258	1494
Traditional LM	33%	2002	4003
Trust Region LM	12%	1517	1649
BFGS	8%	5363	5365

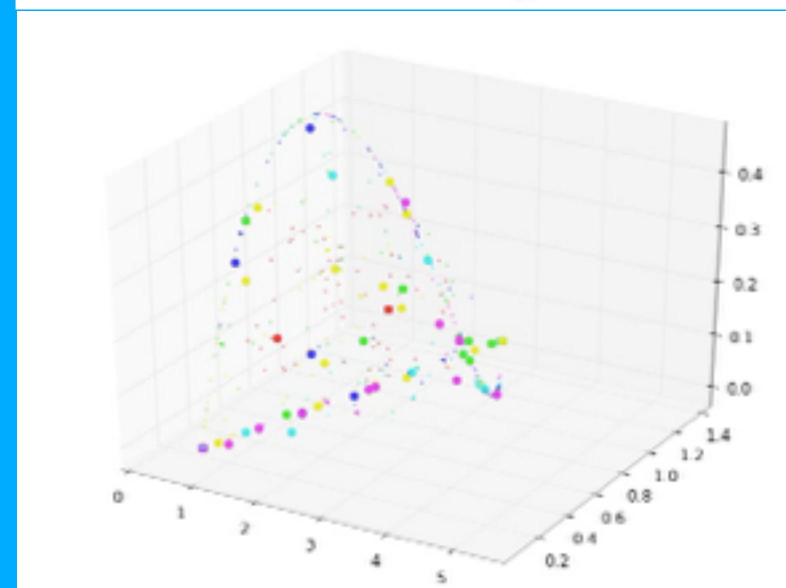
Follow parabola, ***geodesic acceleration***
Cheap to calculate; faster; more success

B. Finding best fits: Model manifold dynamics (Isabel Kloumann)

Dynamics on the model manifold: Searching for the best fit

- Jeffrey's prior plus noise
- Big noise concentrates on manifold edges
- Note scales: flat
- Top: Levenberg-Marquardt
- Bottom: Geodesic acceleration
- Large points: Initial conditions which fail to converge to best fit

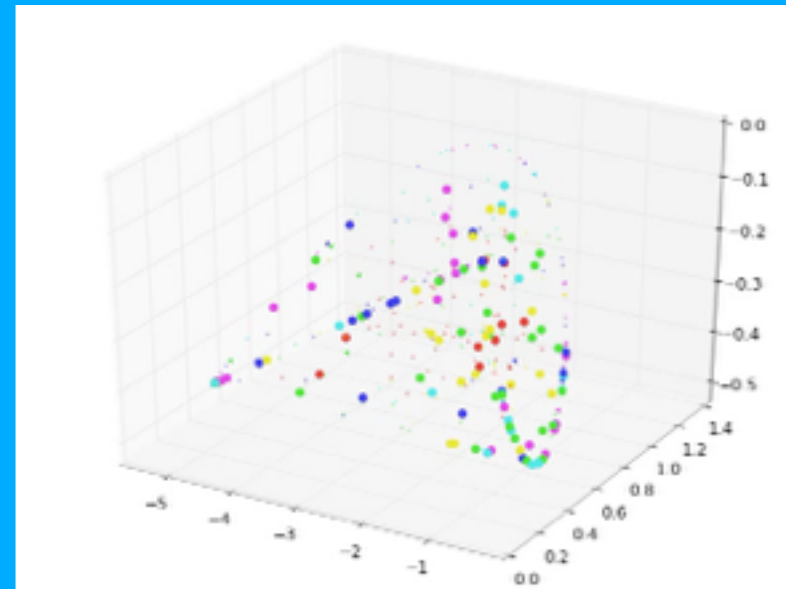


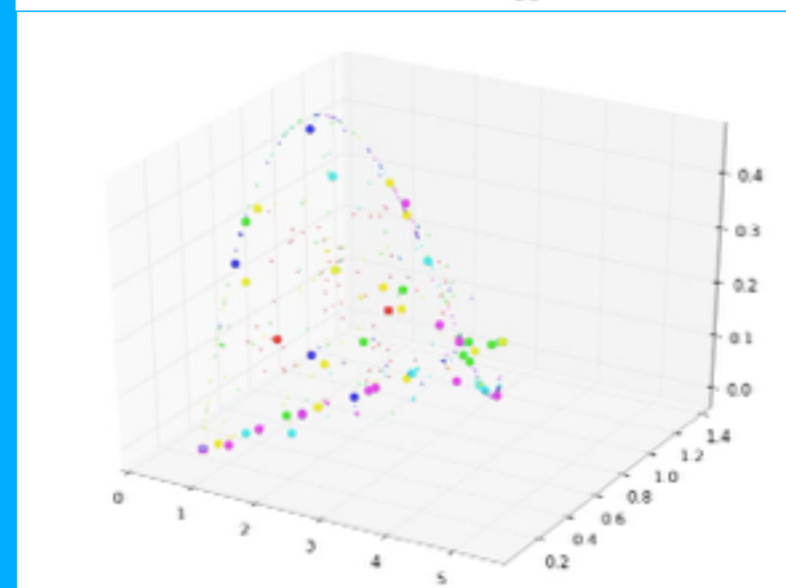


B. Finding best fits: Model manifold dynamics (Isabel Kloumann)

Dynamics on the model manifold: Searching for the best fit

- Jeffrey’s prior plus noise
- Big noise concentrates on manifold edges
- Note scales: flat
- Top: Levenberg-Marquardt
- Bottom: Geodesic acceleration
- Large points: Initial conditions which fail to converge to best fit

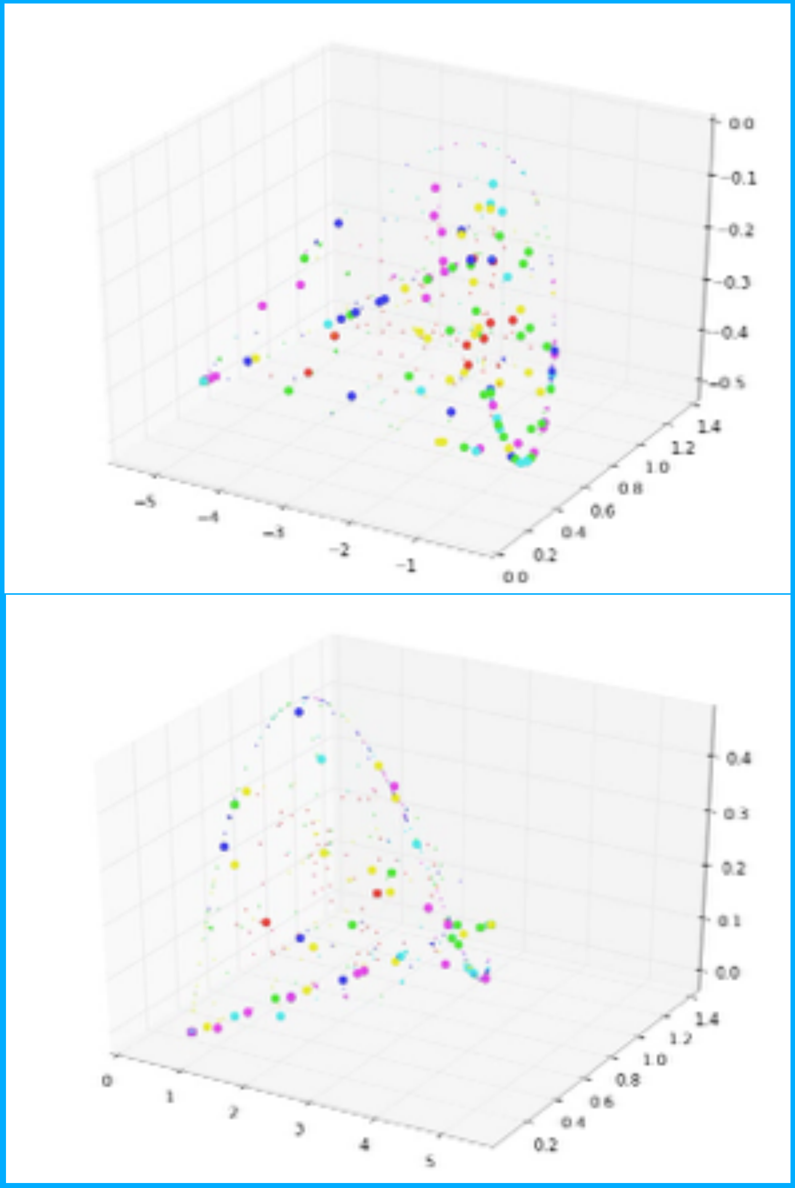




B. Finding best fits: Model manifold dynamics (Isabel Kloumann)

Dynamics on the model manifold: Searching for the best fit

- Jeffrey's prior plus noise
- Big noise concentrates on manifold edges
- Note scales: flat
- Top: Levenberg-Marquardt
- Bottom: Geodesic acceleration
- Large points: Initial conditions which fail to converge to best fit



C. Generation of Reduced Models

Mark Transtrum (not me)

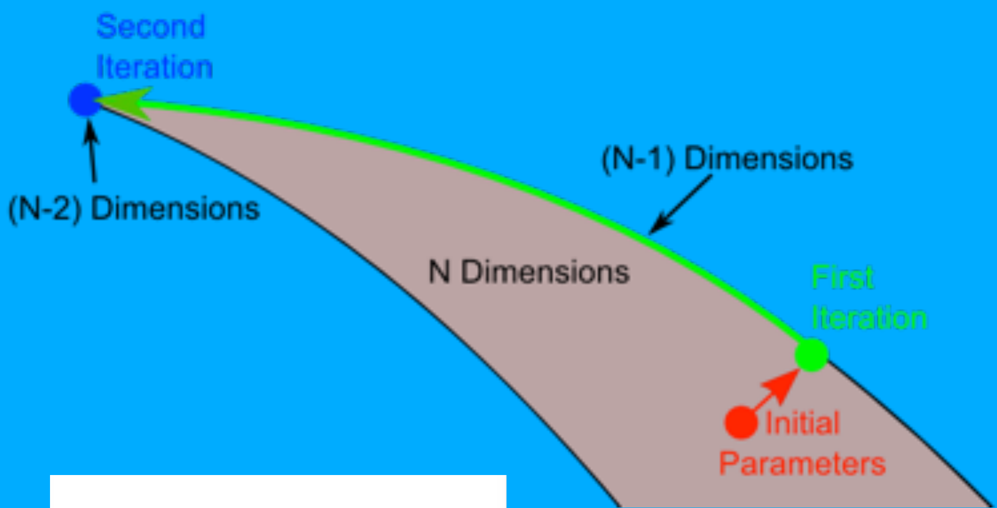
Can we coarse-grain sloppy models? If most parameter directions are useless, why not remove some?

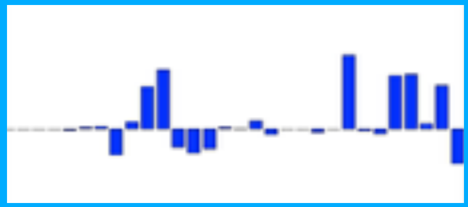
Transtrum has *systematic* method!

(1) Geodesic along sloppiest direction to nearby point on manifold boundary

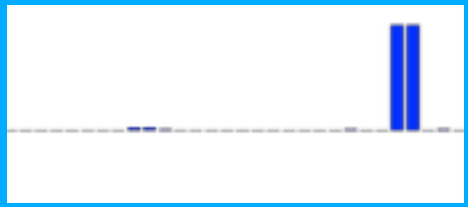
(2) Eigendirection simplifies at model boundary to chemically reasonable simplified model

Coarse-graining = boundaries of model manifold.

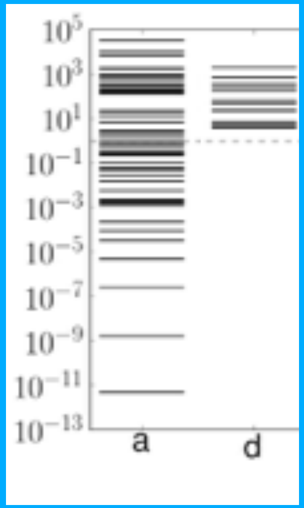




Sloppiest Eigendirection

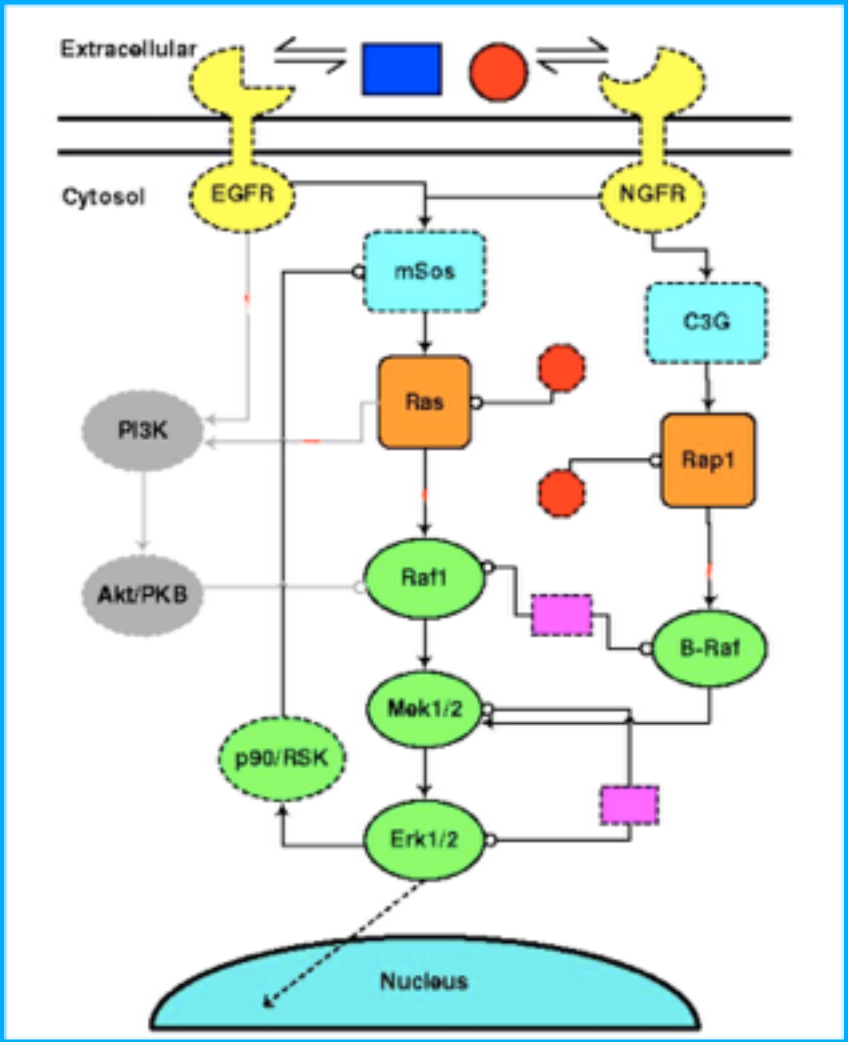


Simplified at Boundary
(Unsaturable reaction)

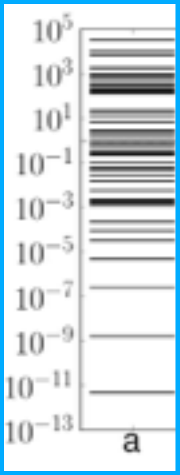


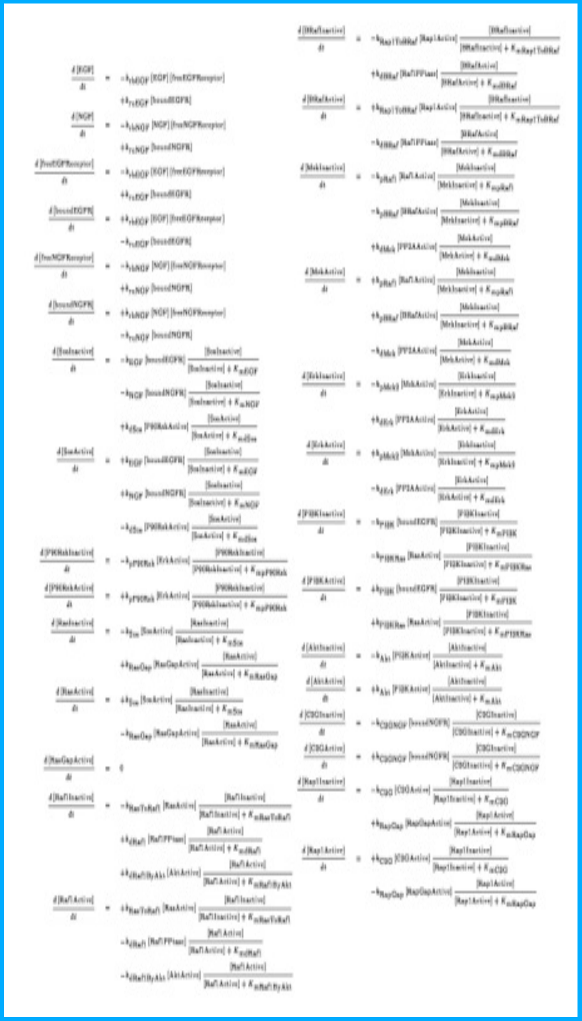
C. Generation of Reduced Models

Mark Transtrum (not me)



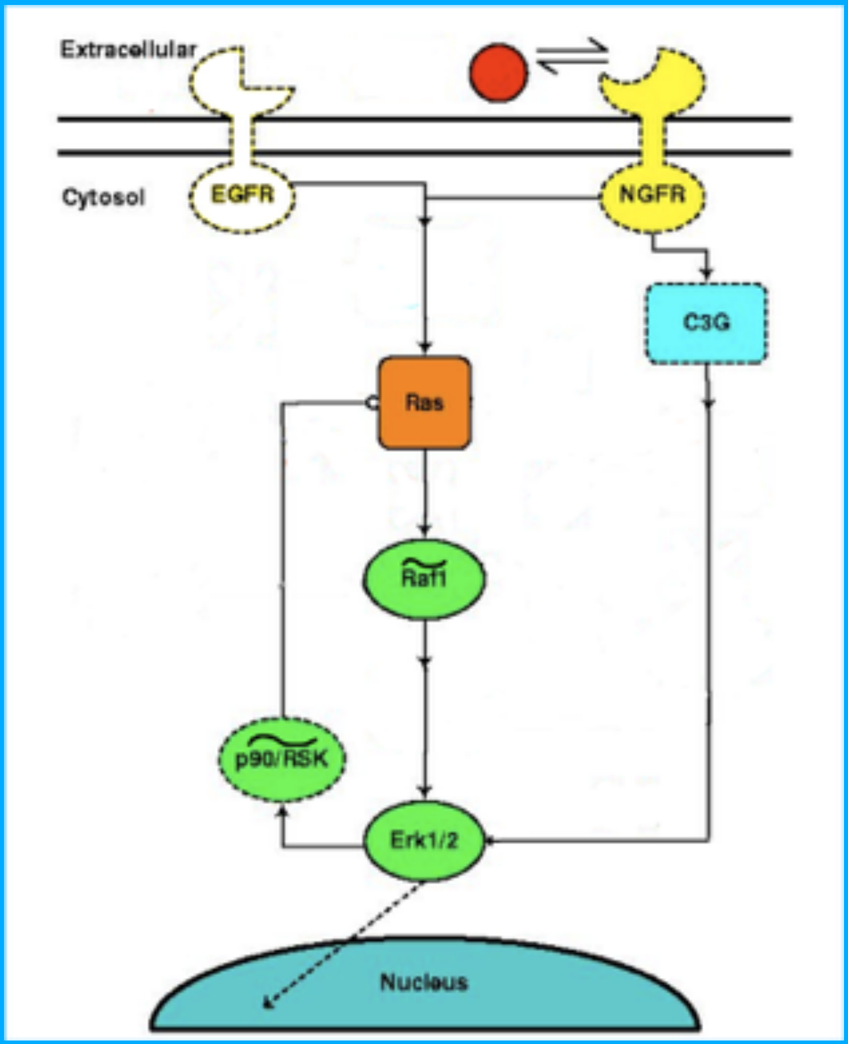
48 params
29 ODEs



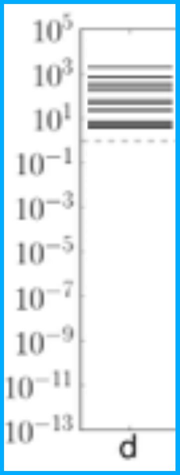


C. Generation of Reduced Models

Mark Transtrum (not me)



12 params
6 ODEs



$$[\mathrm{bEGFR}] = \begin{cases} 1 & \text{EGF Present} \\ 0 & \text{Otherwise} \end{cases}$$
$$\frac{d}{dt}[\mathrm{bNGFR}] = \theta_1[\mathrm{NGF}][\mathrm{fNGFR}]$$
$$\frac{d}{dt}[\mathrm{NGF}] = -\theta_1[\mathrm{NGF}][\mathrm{fNGFR}]$$
$$\frac{d}{dt}[\mathrm{RasA}] = -[\mathrm{RasA}][\widetilde{\mathrm{P90RskA}}] + \theta_2[\mathrm{bEGFR}] + \theta_3[\mathrm{bNGFR}]$$
$$\frac{d}{dt}[\widetilde{\mathrm{Raf1A}}] = \theta_4[\mathrm{RasA}] - \theta_5[\widetilde{\mathrm{Raf1A}}]/([\widetilde{\mathrm{Raf1A}}] + \theta_6)$$
$$\frac{d}{dt}[\mathrm{C3GA}] = \theta_7[\mathrm{bNGFR}][\mathrm{C3GI}]$$
$$[\mathrm{Rap1A}] = \theta_8[\mathrm{C3GA}]$$
$$[\mathrm{MekA}] = [\widetilde{\mathrm{Raf1A}}][\mathrm{MekI}] + \theta_9[\mathrm{Rap1A}]$$
$$\frac{d}{dt}[\mathrm{Erk}] = -\theta_{10}[\mathrm{ErkA}] + \theta_{11}[\mathrm{MekA}][\mathrm{ErkI}]$$
$$\frac{d}{dt}[\widetilde{\mathrm{P90RskA}}] = \theta_{12}[\mathrm{ErkA}]$$

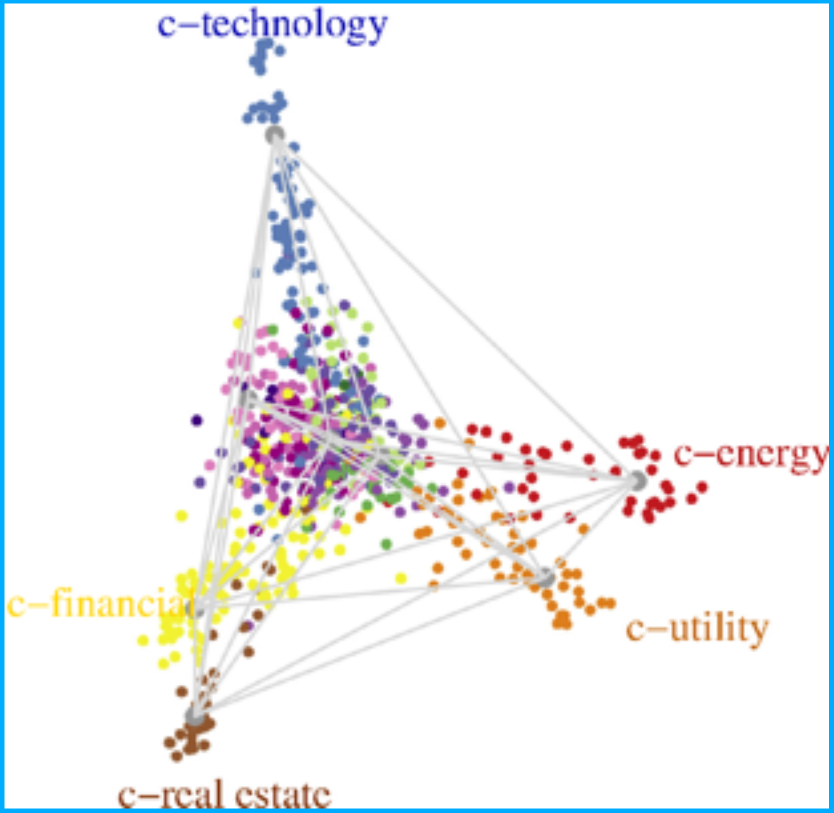
Reduced model fits all experimental data

$$\theta_9 = \frac{[\mathrm{BRafI}]\ \mathrm{kRap1toBRaf}\ \mathrm{KmdBRAF}\ \mathrm{kpBRaf}\ \mathrm{KmdMek}}{[\mathrm{PP2AA}]\ [\mathrm{Raf1PPtase}]\ \mathrm{kdBRaf}\ \mathrm{KmRap1toBRaf}\ \mathrm{kdMek}}$$

Effective ‘renormalized’ params

D. Machine Learning

Ricky Chachra, Alex Alemi, Paul Ginsparg



Stock Returns Decomposed into ‘Canonical’ Sectors (unsupervised learning)

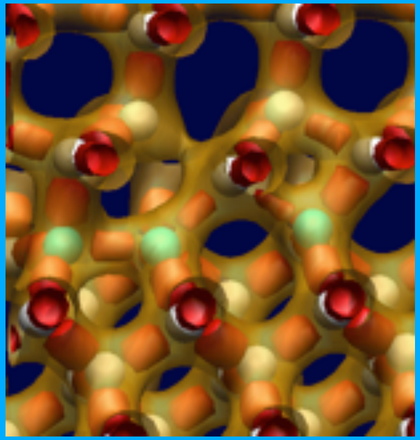
Low dimensional representations of high-dimensional data

Neural Networks categorize images

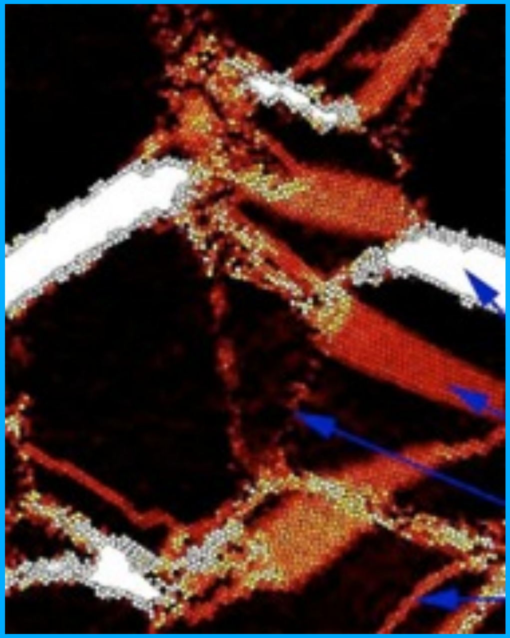
E. Bayesian Errors for Atoms

'Sloppy Model' Approach to Error Estimation of Interatomic Potentials

Søren Frederiksen, Karsten W. Jacobsen, Kevin Brown, JPS



Quantum Electronic Structure (Si)
90 atoms (Mo)
(Arias)



Atomistic potential
820,000 Mo atoms
(Jacobsen, Schiøtz)

Interatomic Potentials $V(\mathbf{r}_1, \mathbf{r}_2, \ldots)$

- Fast to compute
- Limit $m_e/M \to 0$ justified
- Guess functional form
  Pair potential $\sum V(\mathbf{r}_i - \mathbf{r}_j)$ poor
  Bond angle dependence
  Coordination dependence
- Fit to experiment (old)
- Fit to forces from electronic structure calculations (new)

17 Parameter Fit

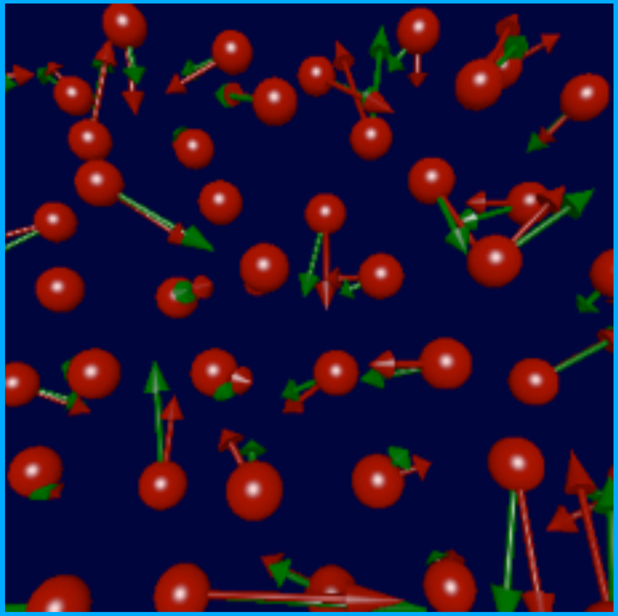
E. Interatomic Potential Error Bars

Ensemble of Acceptable Fits to Data

Not ***transferable***

Unknown errors

- 3% elastic constant
- 10% forces
- 100% fcc-bcc, dislocation core



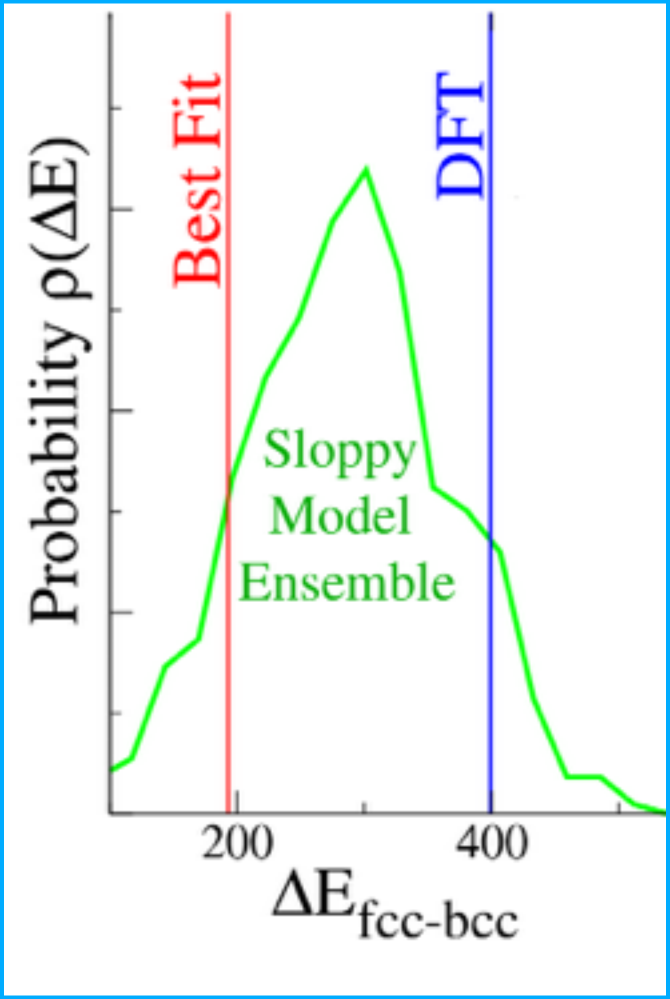
Green = DFT, Red = Fits

Best fit is ***sloppy***: ensemble of fits that aren't much worse than best fit.

Ensemble in Model Space!

T_0 set by equipartition energy = best cost

Error Bars from quality of best fit



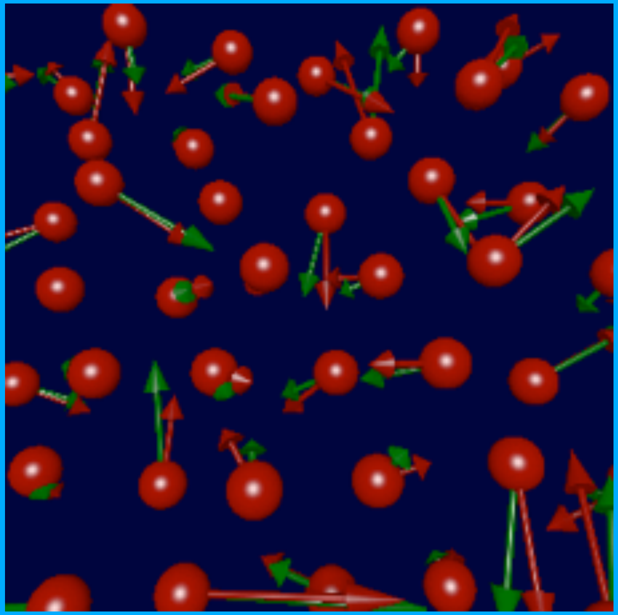
E. Interatomic Potential Error Bars

Ensemble of Acceptable Fits to Data

Not ***transferable***

Unknown errors

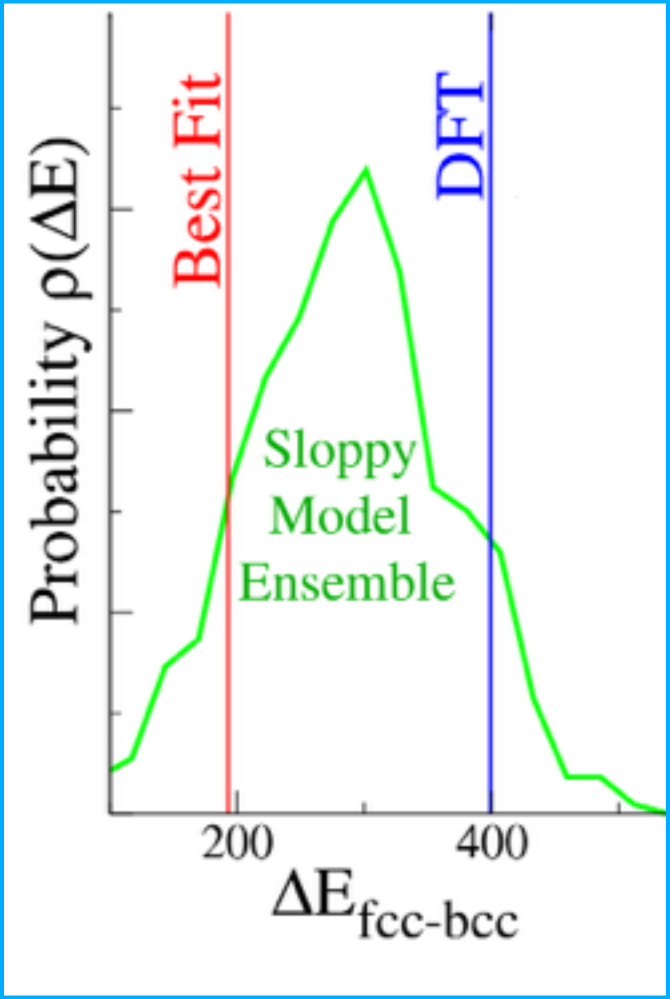
- 3% elastic constant
- 10% forces
- 100% fcc-bcc, dislocation core



Green = DFT, Red = Fits

Best fit is ***sloppy***: ensemble of fits that aren't much worse than best fit. **Ensemble in Model Space!** T_0 set by equipartition energy = best cost

Error Bars from quality of best fit



Sloppy Molybdenum: Does it Work?

Estimating *Systematic* Errors

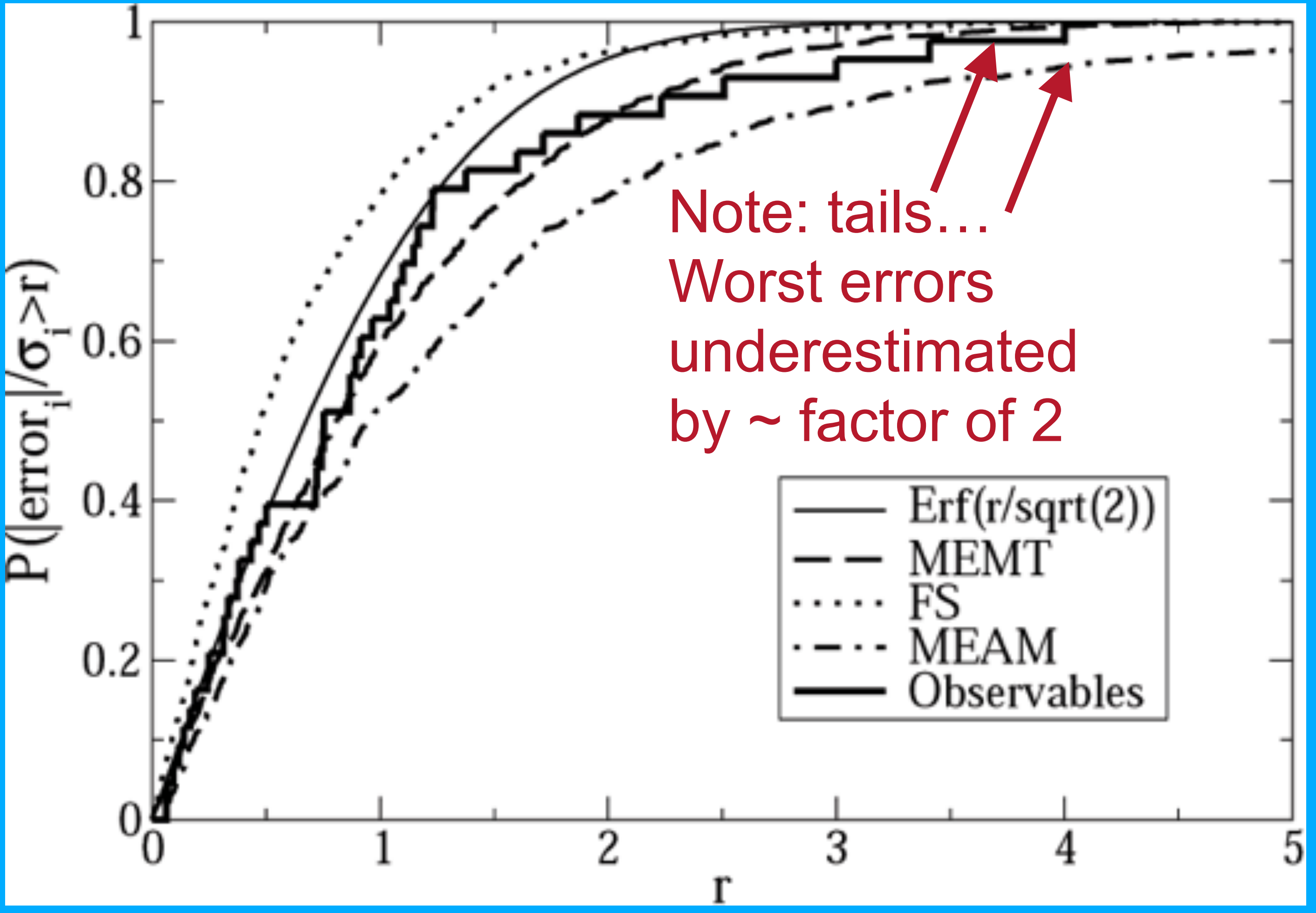
Bayesian error σ_i gives total error if ratio $r = \mathrm{error}_i/\sigma_i$ distributed as a Gaussian: cumulative distribution $P(r)=Erf(r/\sqrt{2})$

Three potentials

- Force errors
- Elastic moduli
- Surfaces
- Structural
- Dislocation core
- $7\% < \sigma_i < 200\%$

"Sloppy model" systematic error most of total

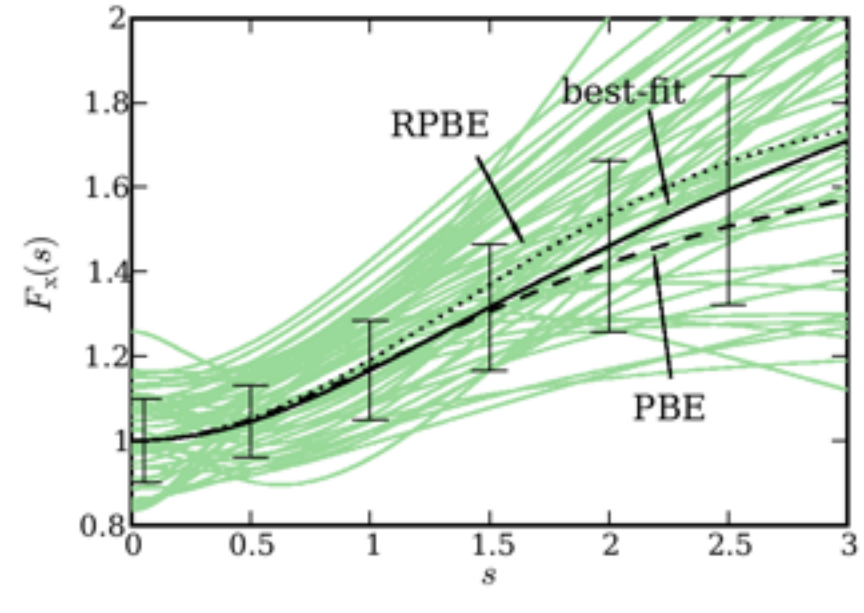
$\sim 2 \ll 200\%/7\%$



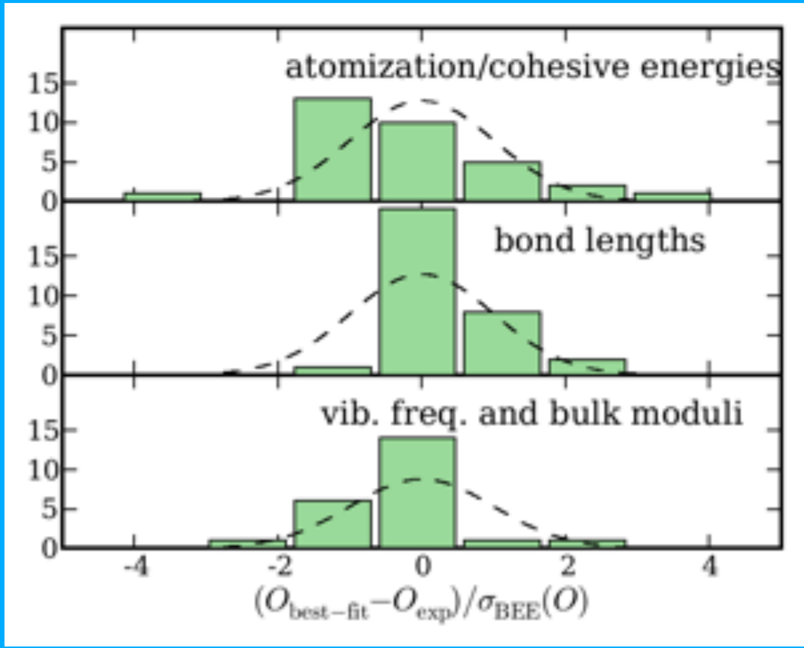
Systematic Error Estimates for DFT

GGA-DFT as Multiparameter Fit?

J. J. Mortensen, K. Kaasbjerg, S. L. Frederiksen,
J. K. Nørskov, JPS, K. W. Jacobsen,
(Anja Tuftelund, Vivien Petzold, Thomas Bligaard)



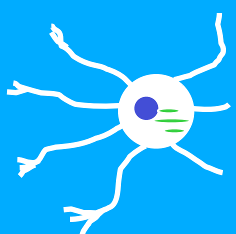
Enhancement factor $F_x(s)$
in the exchange energy E_x
Large fluctuations



Actual error / predicted error
Deviation from experiment well described by ensemble!

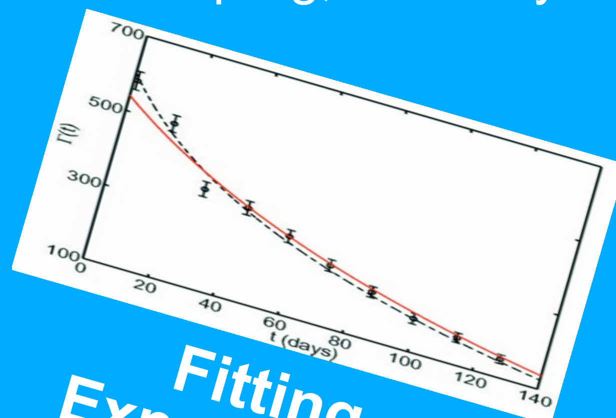
‘Sloppy Model’ Nonlinear Fits: Signal Transduction to Differential Geometry

JPS, Mark Transtrum, Ben Machta, Ricky Chachra, Lorien Hayden, Alex Alemi, Isabel Kloumann, Colin Clement, Kevin Brown, Ryan Gutenkunst, Josh Waterfall, Paul Ginsparg, Chris Myers, …

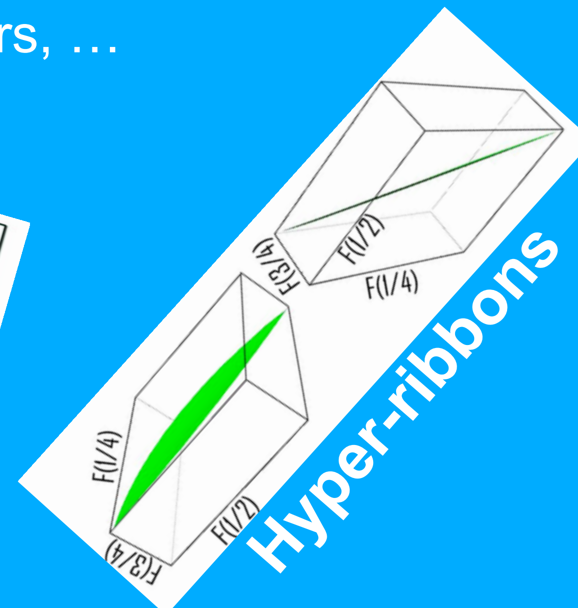


Cell Dynamics

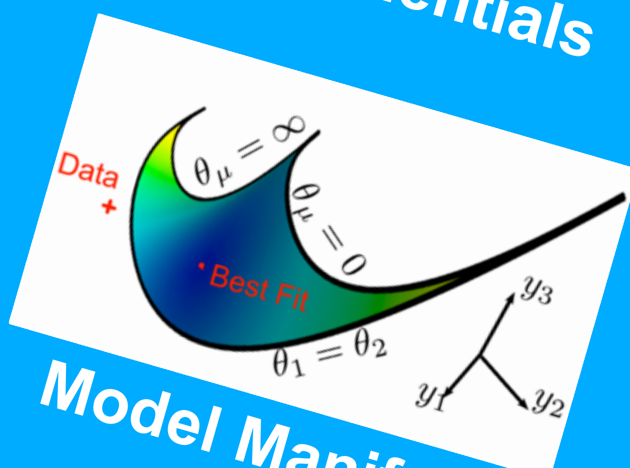
Sloppiness



Fitting Exponentials

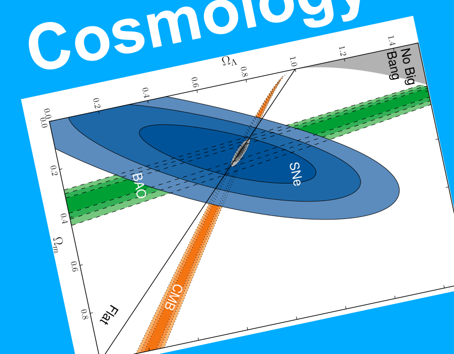


Hyper-ribbons



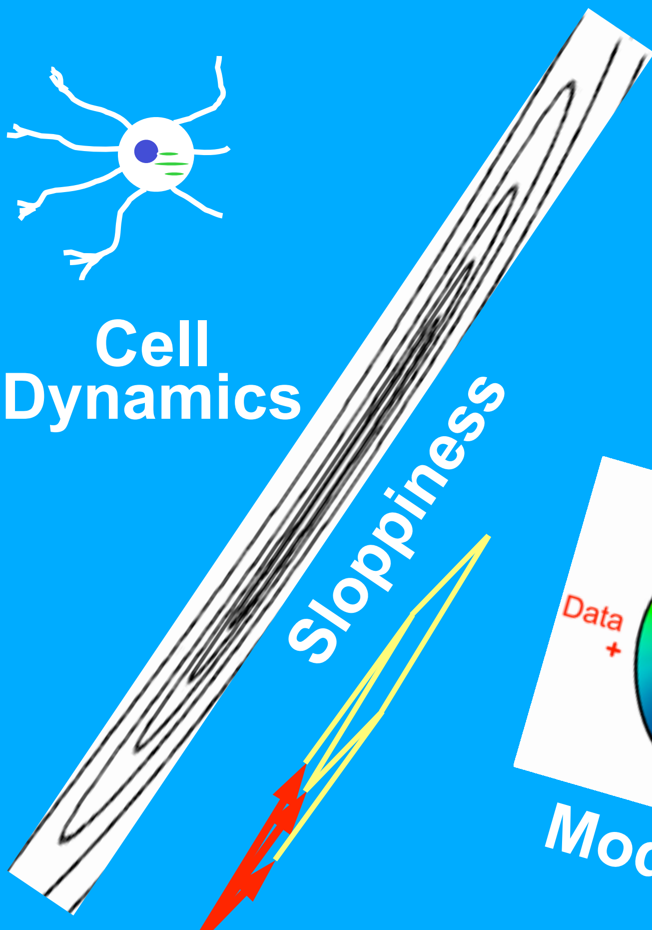
Model Manifold

Cosmology



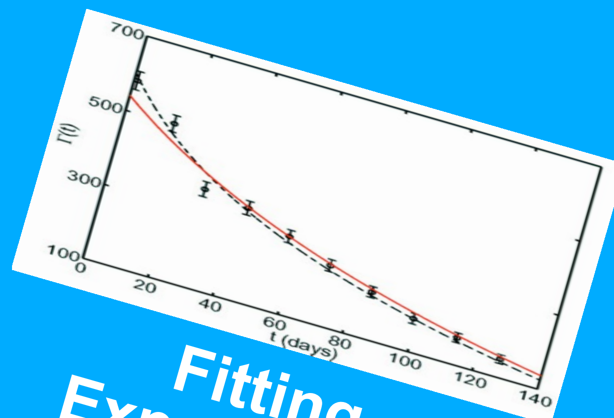
‘Sloppy Model’ Nonlinear Fits: Signal Transduction to Differential Geometry

JPS, Mark Transtrum, Ben Machta, Ricky Chachra, Isabel Kloumann, Kevin Brown, Ryan Gutenkunst, Josh Waterfall, Chris Myers, …

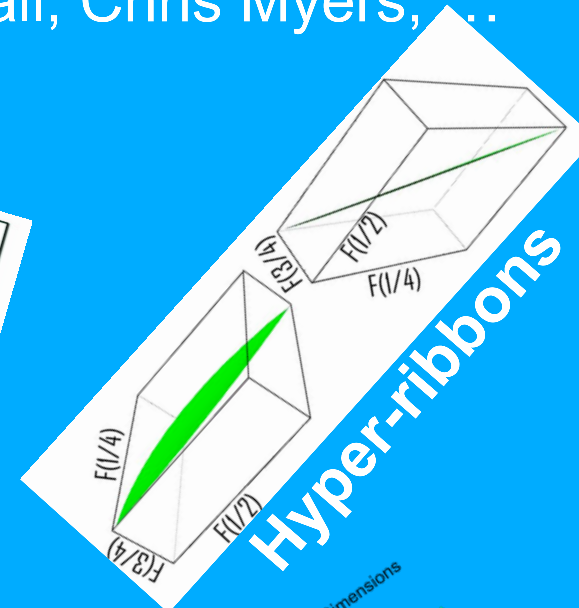


Cell Dynamics

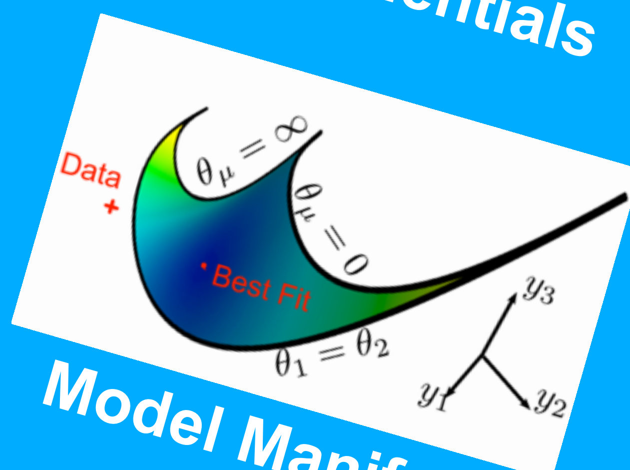
Sloppiness



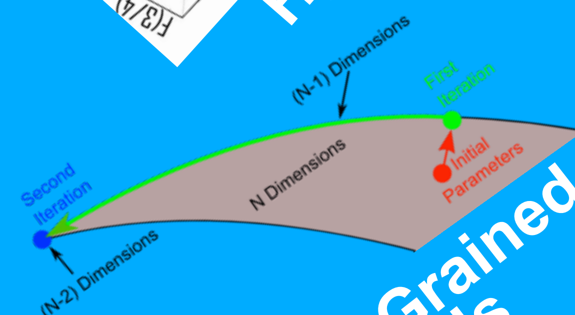
Fitting Exponentials



Hyper-ribbons



Model Manifold

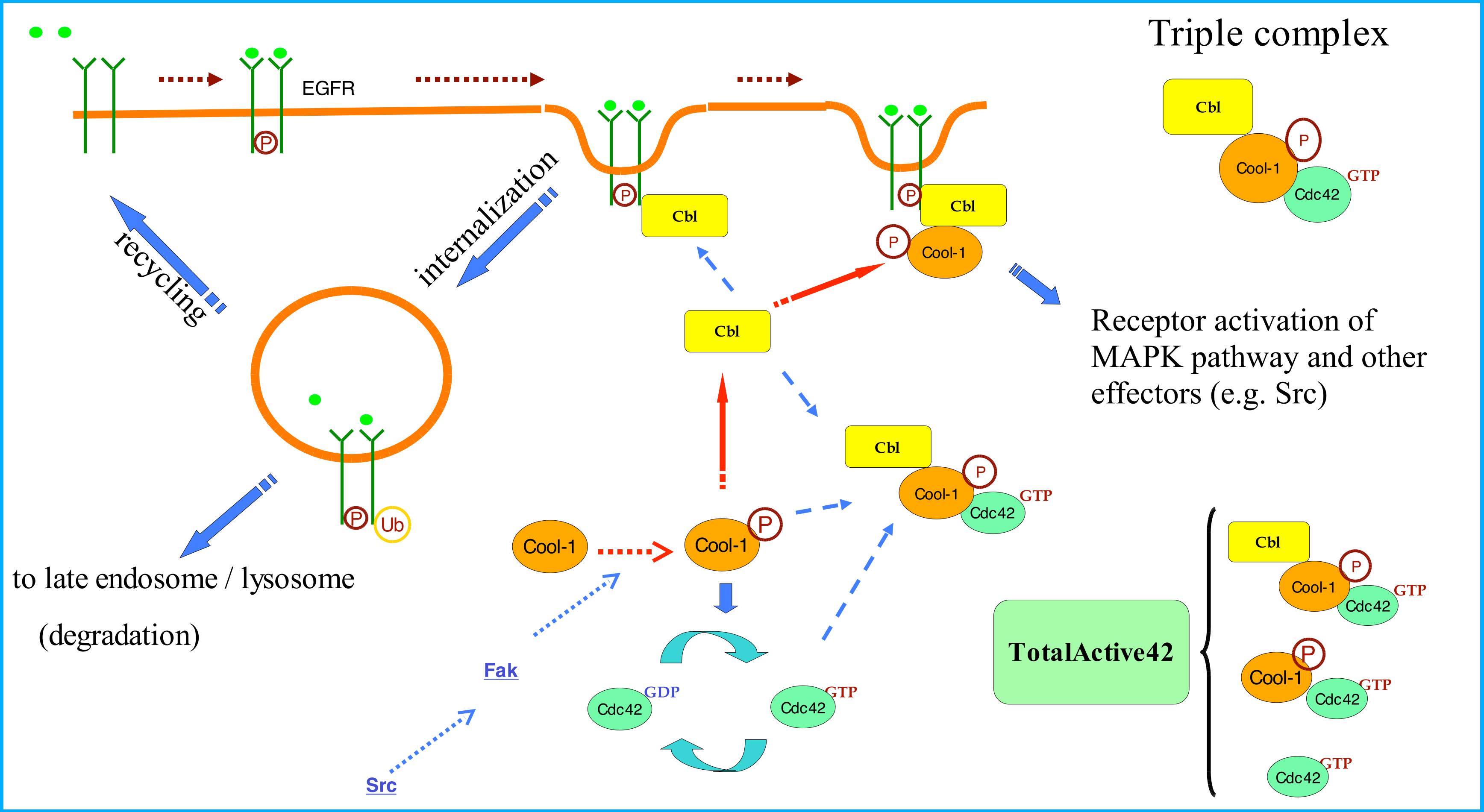


Coarse-Grained Models

C. EGFR Trafficking Model

Fergal Casey, Cerione lab

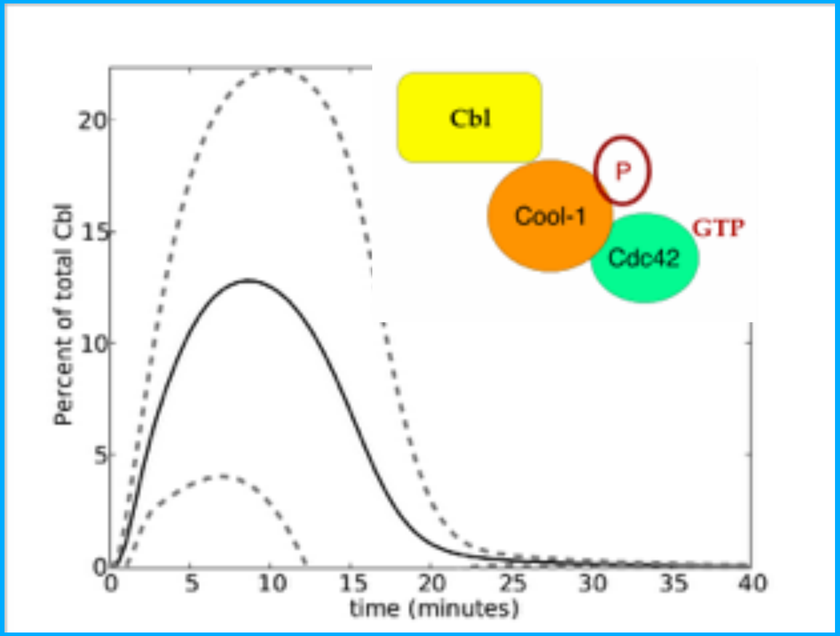
- Active research, Cerione lab: testing hypothesis, experimental design (Cool1≡β-PIX)
- 41 chemicals, 53 rate constants; only 11 of 41 species can be measured
- Does Cool-1 triple complex sequester Cbl, delay endocytosis in wild type NIH3T3 cells?

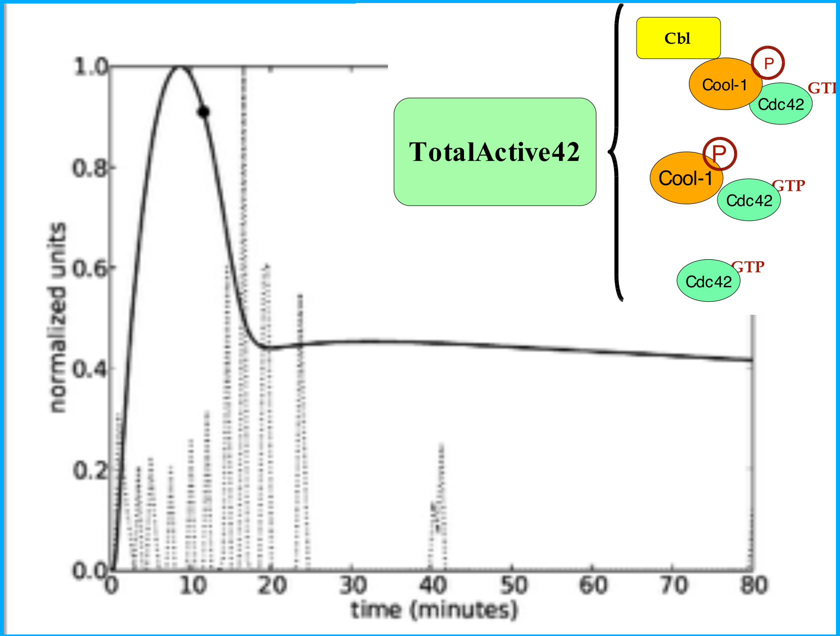


C. Trafficking: experimental design

Which experiment best reduces prediction uncertainty?

- Amount of triple complex was not well predicted
- V-optimal experimental design: single & multiple measurements
- Total active Cdc42 at 10 min.; Cerione independently concurs
- Experiment indicates significant sequestering in wild type
- Predictivity without decreasing parameter uncertainty

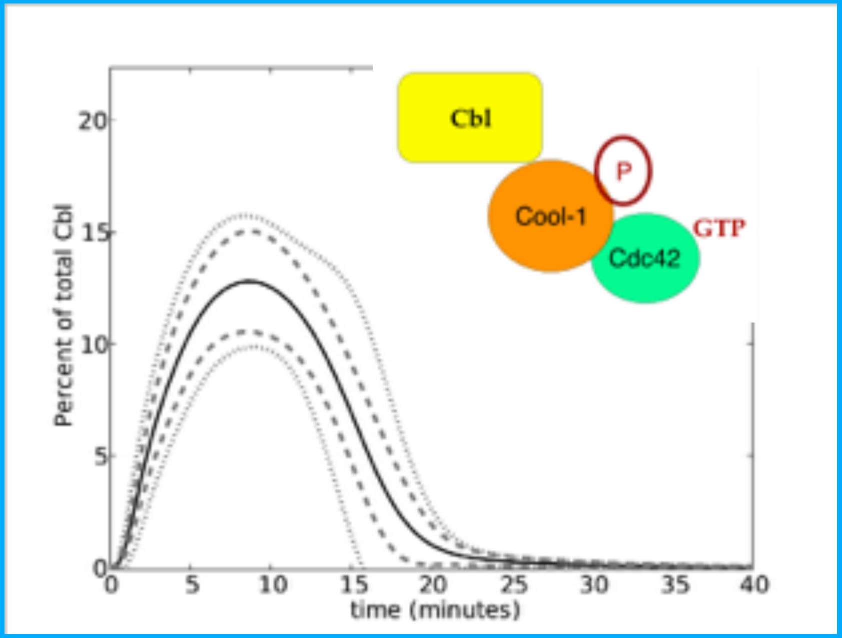


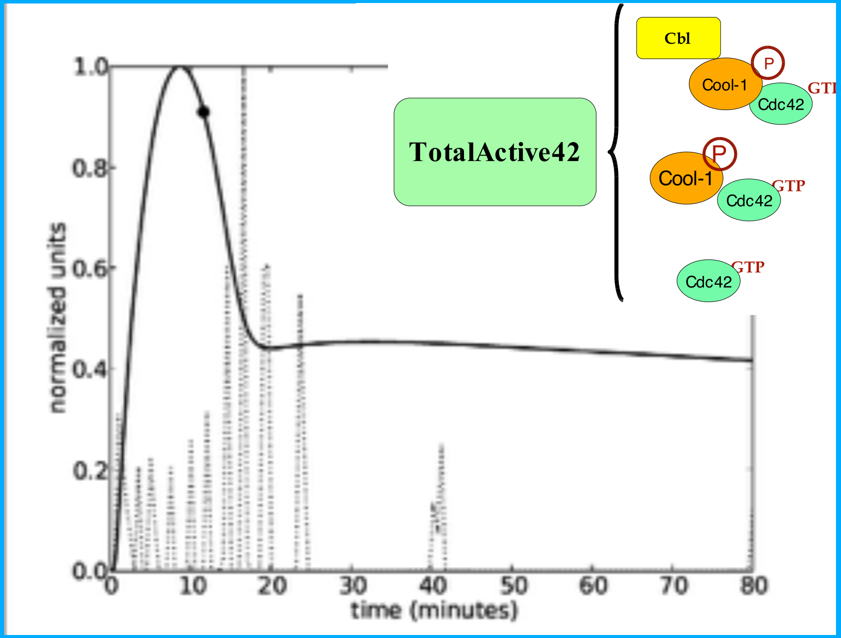


C. Trafficking: experimental design

Which experiment best reduces prediction uncertainty?

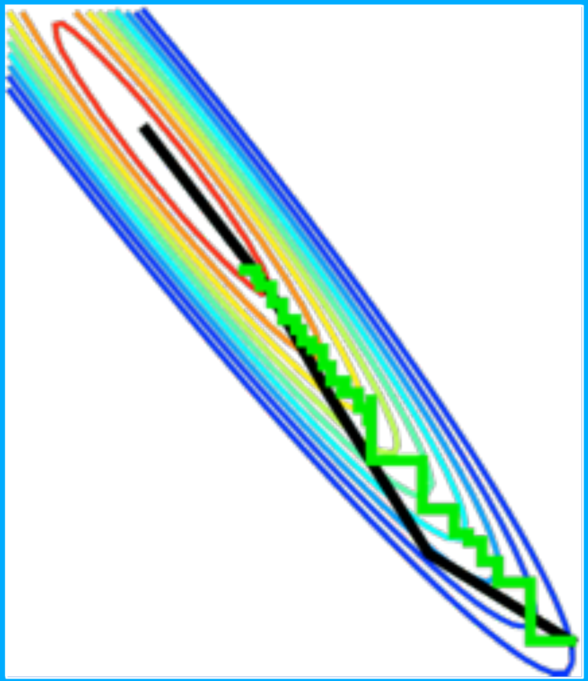
- Amount of triple complex was not well predicted
- V-optimal experimental design: single & multiple measurements
- Total active Cdc42 at 10 min.; Cerione independently concurs
- Experiment indicates significant sequestering in wild type
- Predictivity without decreasing parameter uncertainty

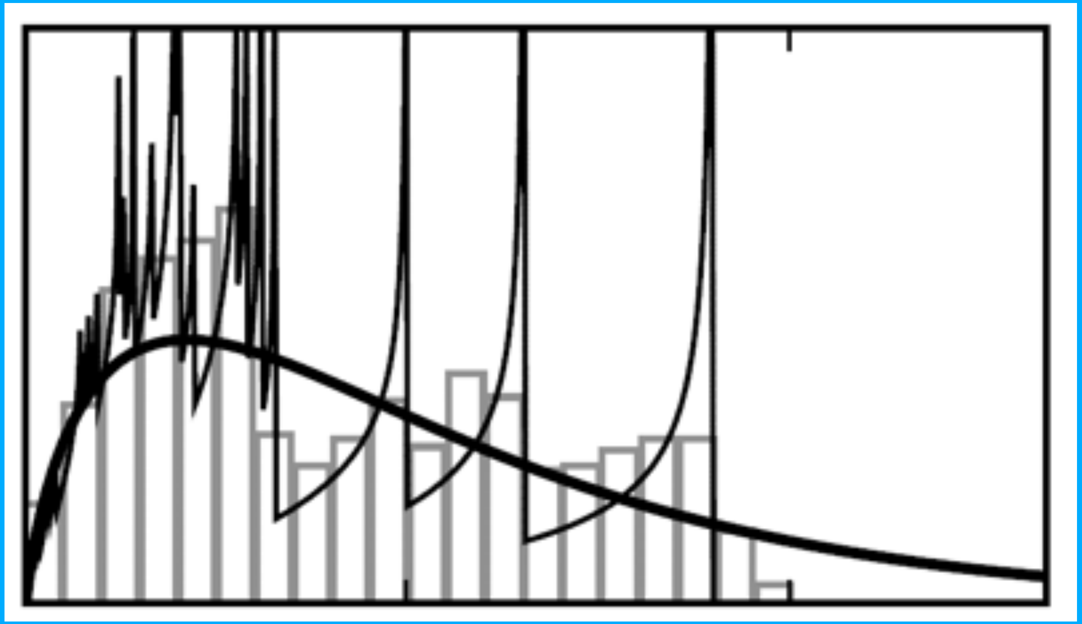




D. Evolution in Chemotype space

Implications of sloppiness?



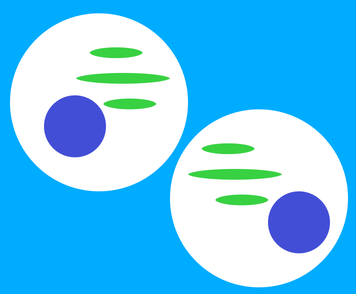


Fitness gain from first successful mutation

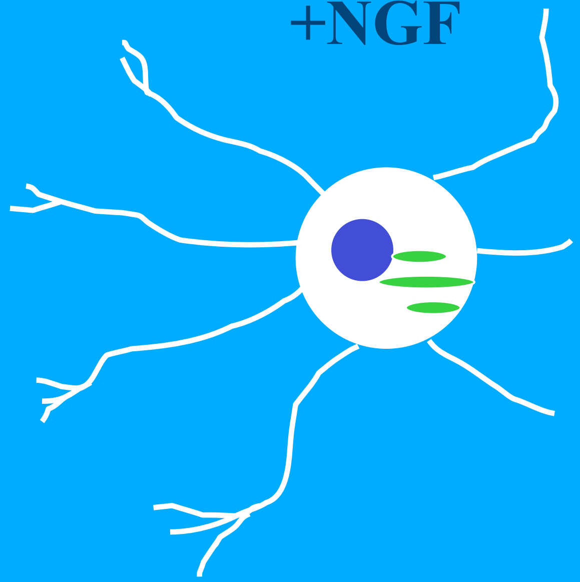
- Culture of identical bacteria, one mutation at a time
- Mutation changes one or two rate constants (no *pleiotropy*): orthogonal moves in rate constant (chemotype) space
- ***Cusps*** in first fitness gain (one for each rate constant, big gap)
- Multiple mutations get stuck on ridge in sloppy landscape

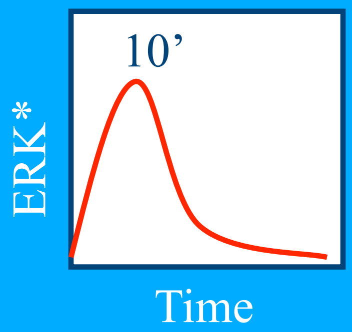
PC12 Differentiation

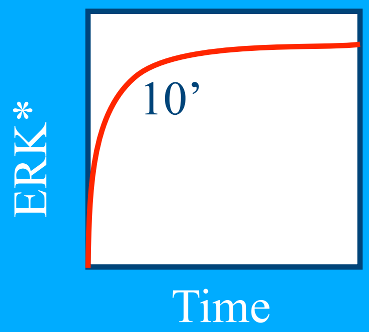
+EGF



+NGF

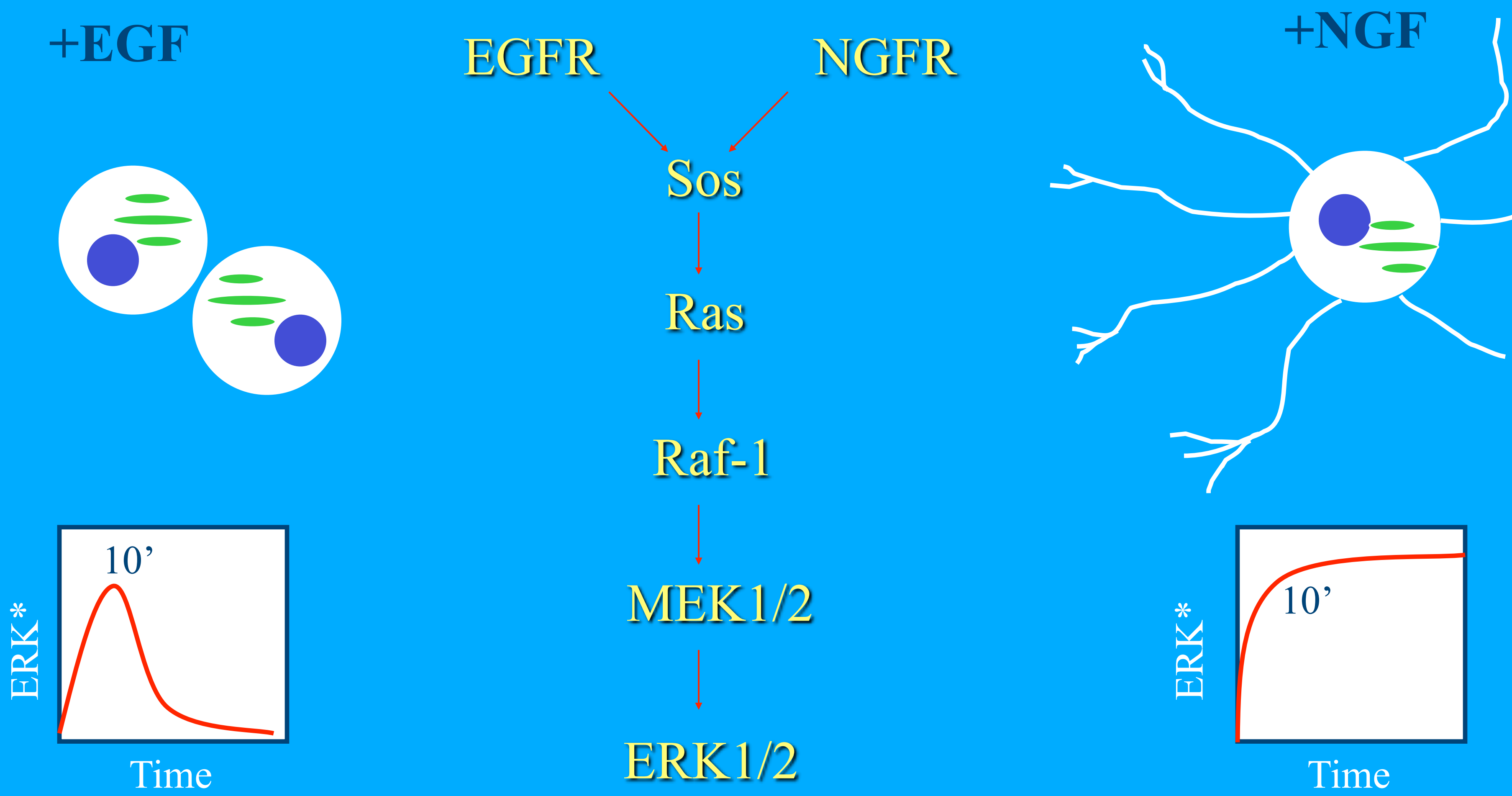






Biologists study which proteins talk to which. Modeling?

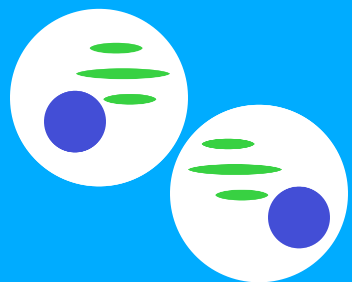
PC12 Differentiation

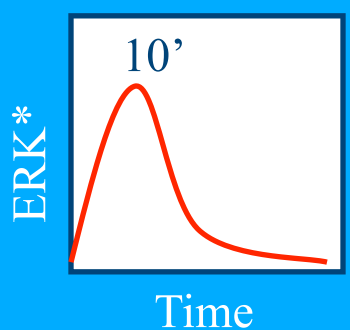


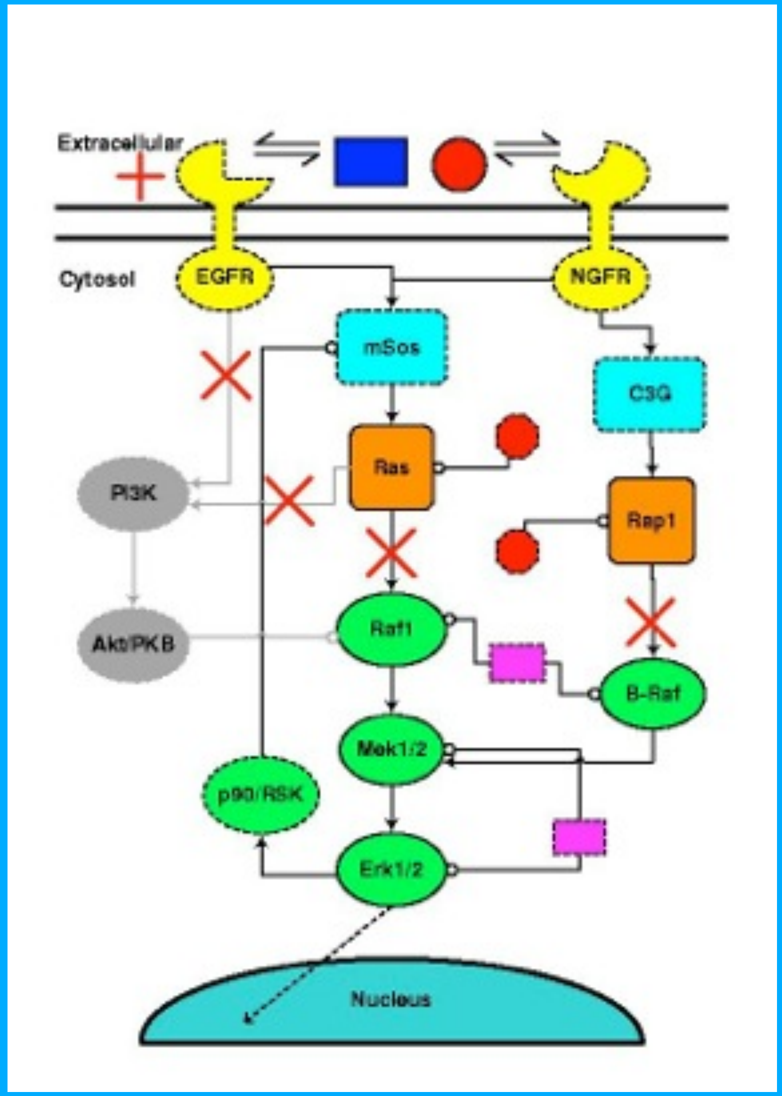
Biologists study which proteins talk to which. Modeling?

PC12 Differentiation

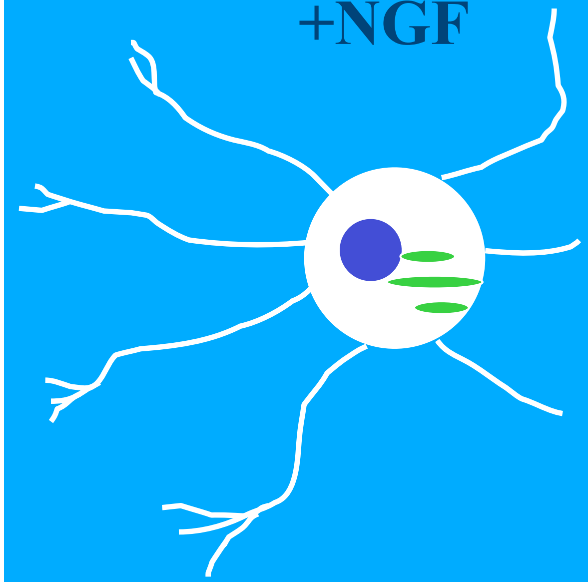
+EGF

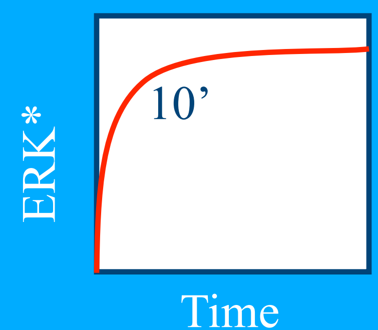






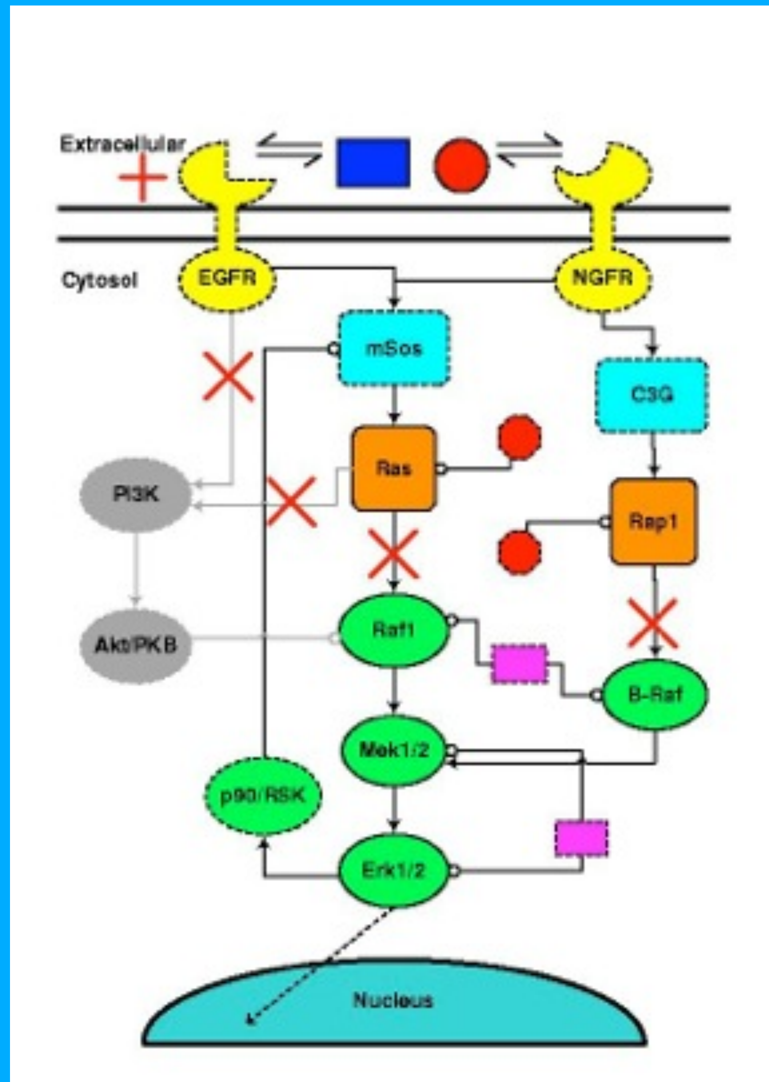
+NGF

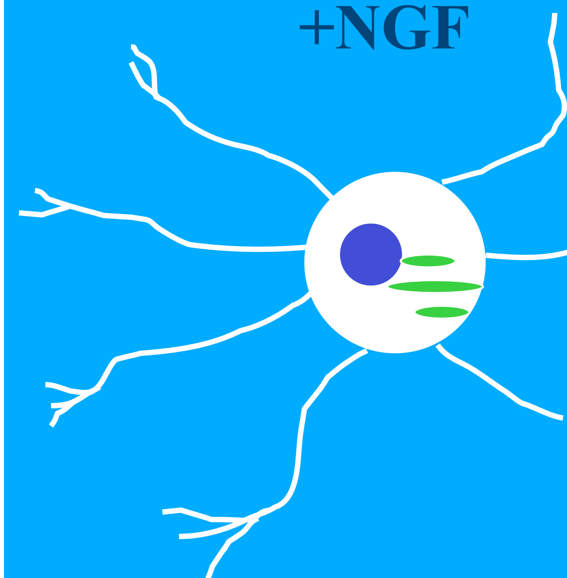


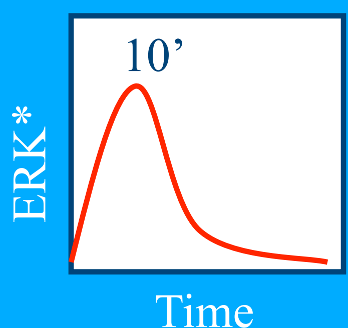


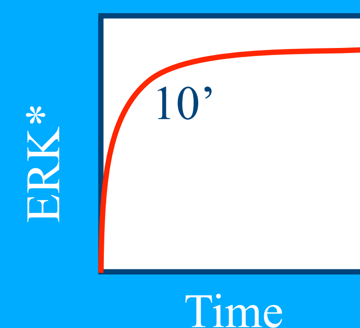
Biologists study which proteins talk to which. Modeling?

PC12 Differentiation



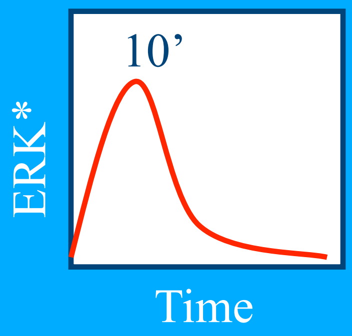


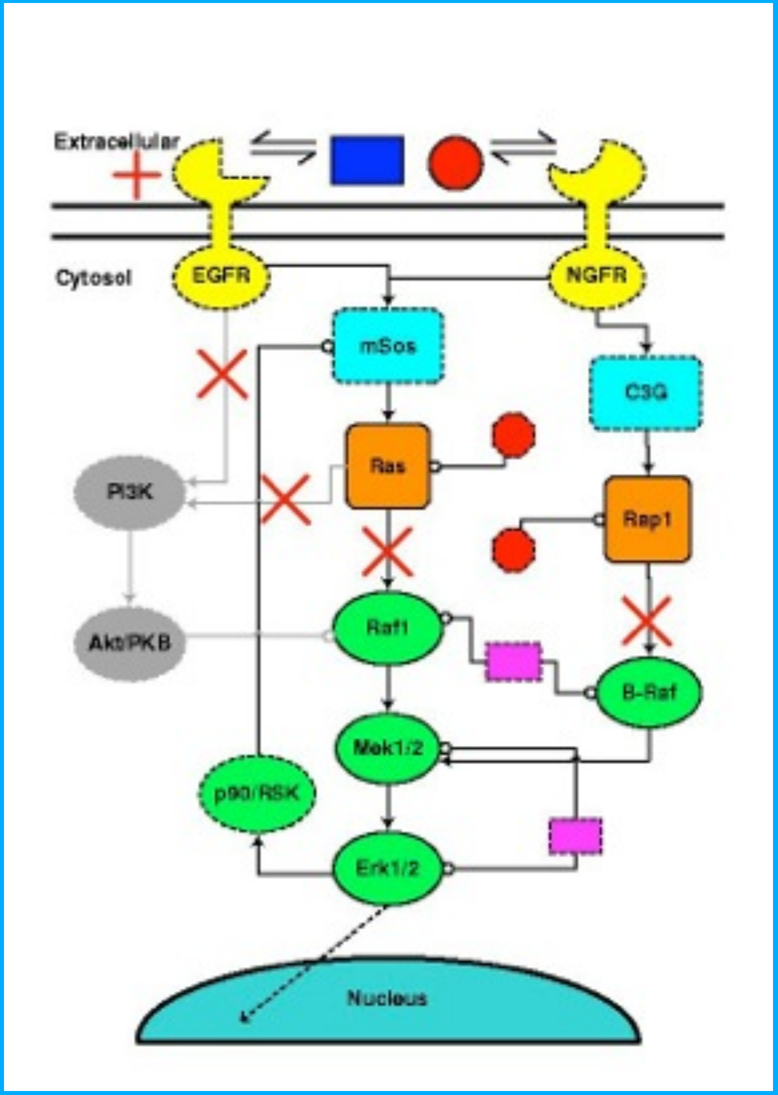


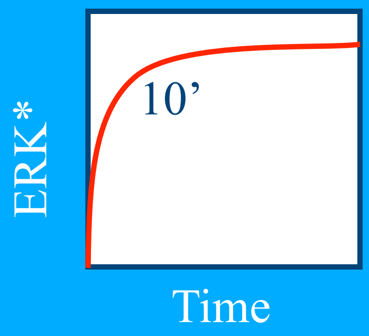


Biologists study which proteins talk to which. Modeling?

PC12 Differentiation



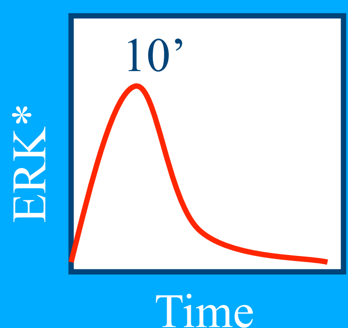


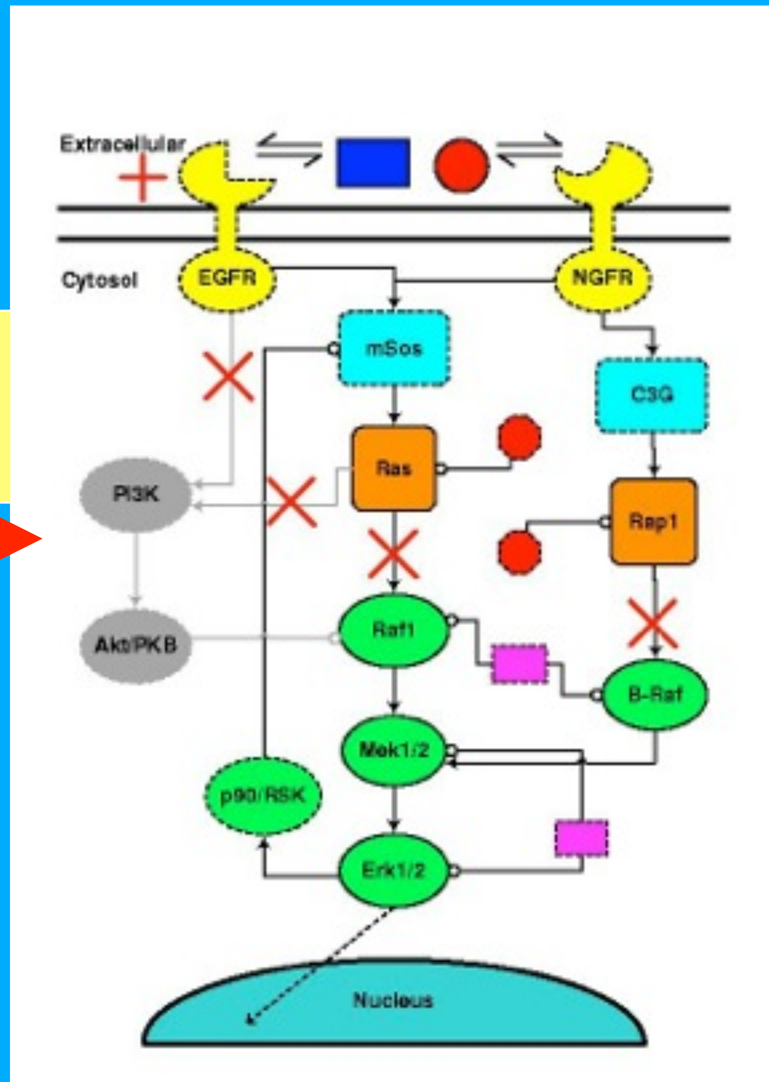


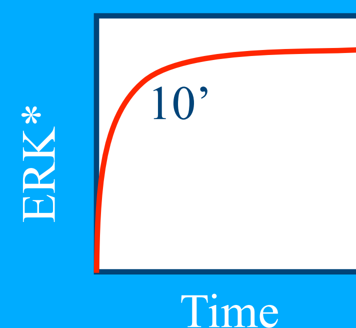
Biologists study which proteins talk to which. Modeling?

PC12 Differentiation

Tunes down signal (Raf-1)





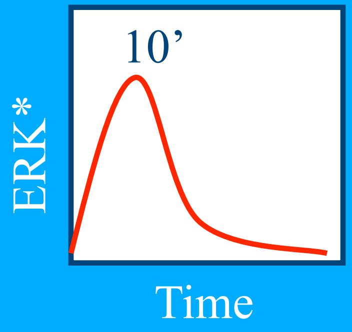


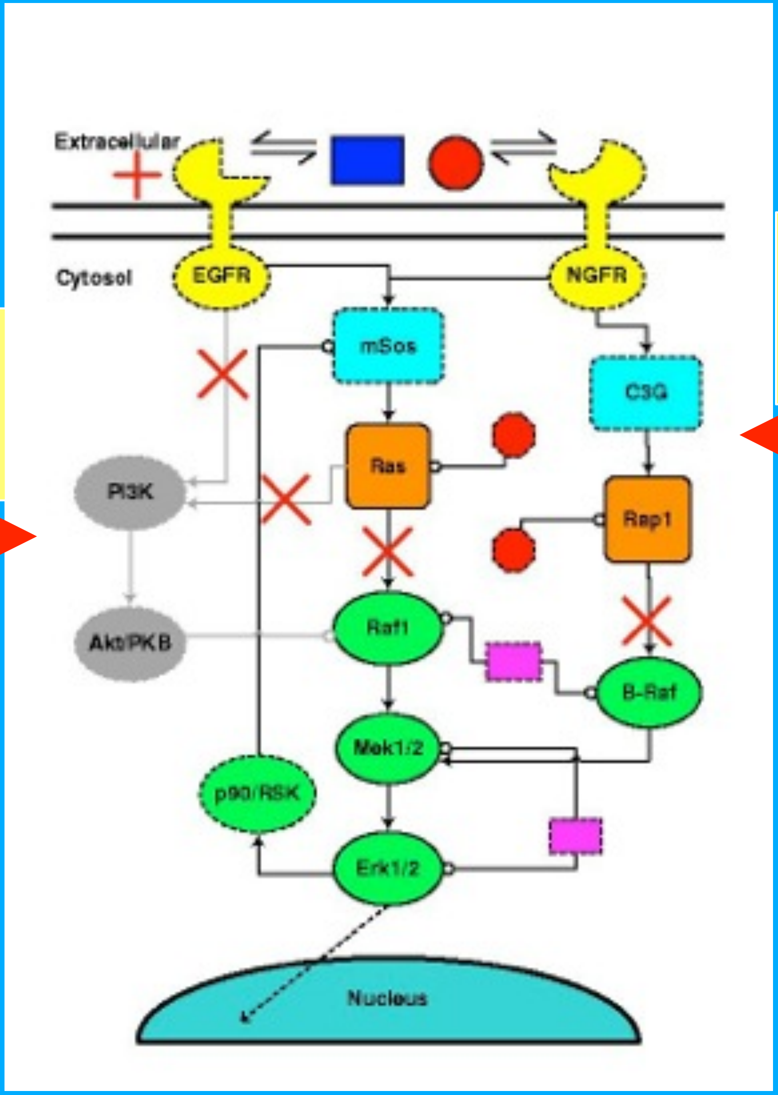
Biologists study which proteins talk to which. Modeling?

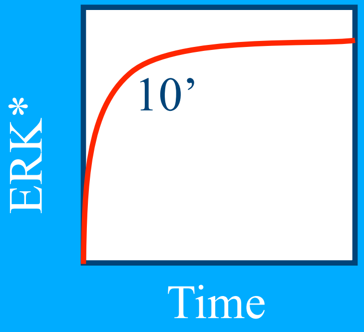
PC12 Differentiation

Tunes down signal (Raf-1)

Pumps up signal (Mek)



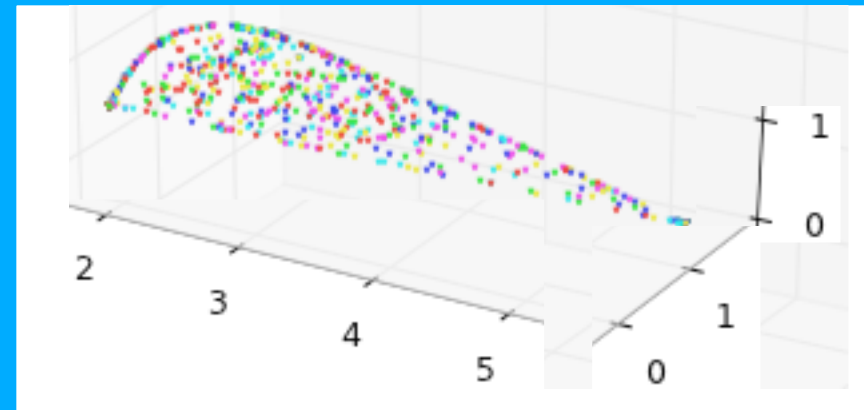


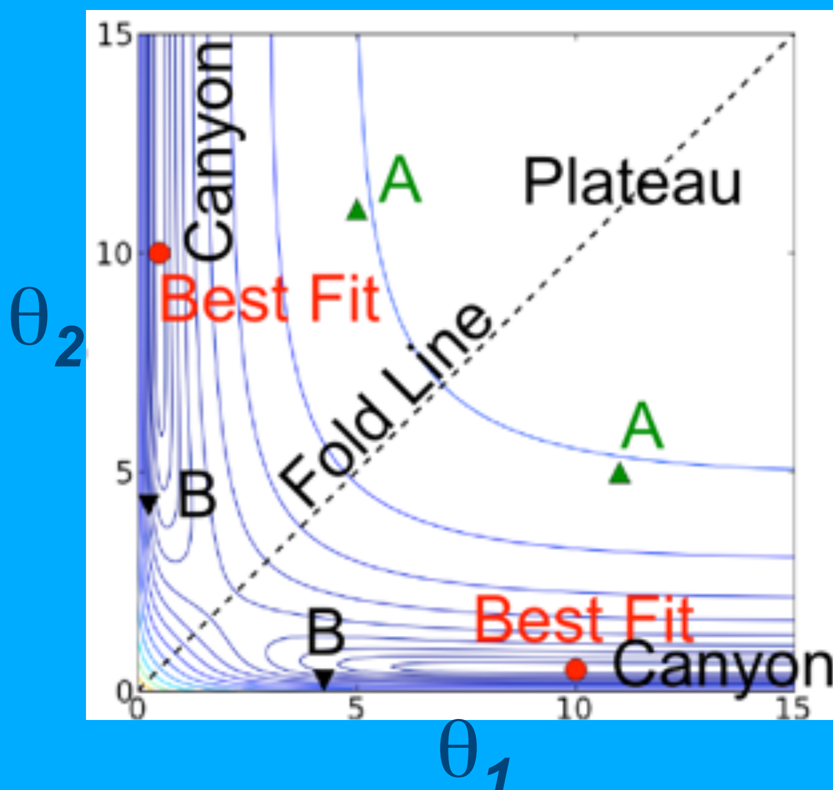


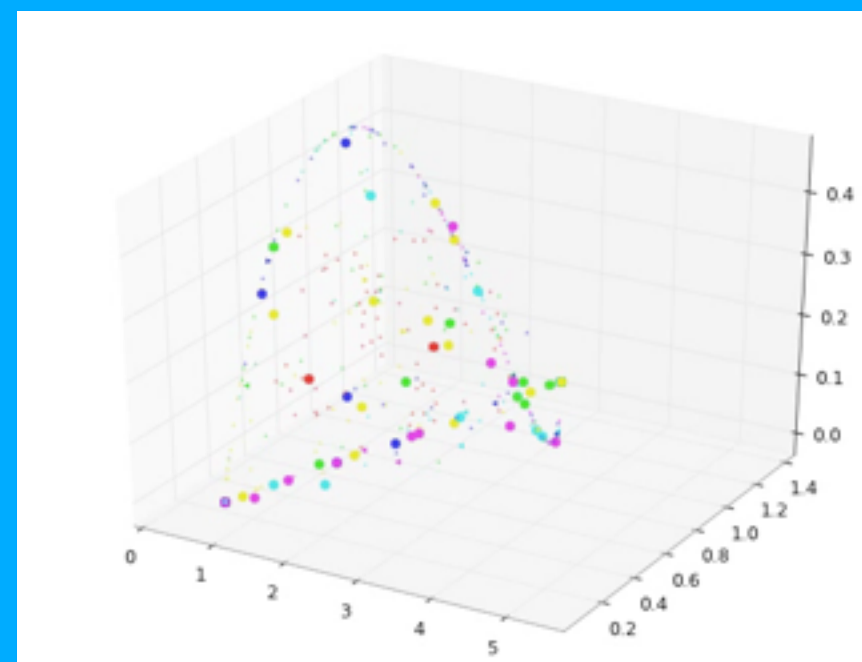
Biologists study which proteins talk to which. Modeling?

Edges of the model manifold

Fitting Exponentials
Top: Flat model manifold; articulated edges = plateau
Bottom: Stretch to uniform aspect ratio (Isabel Kloumann)

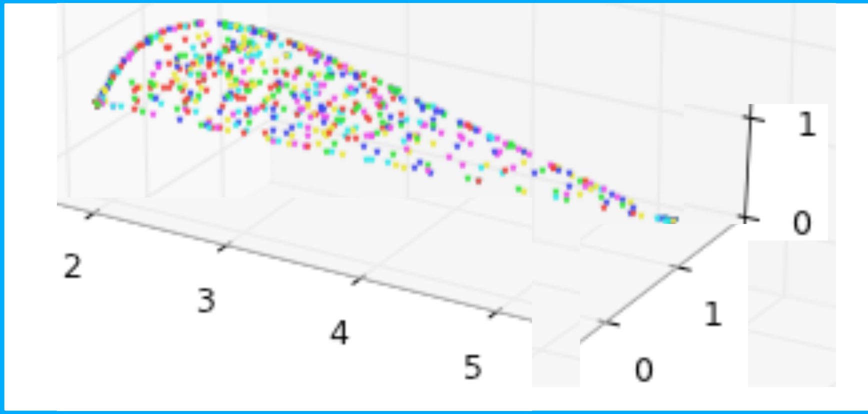


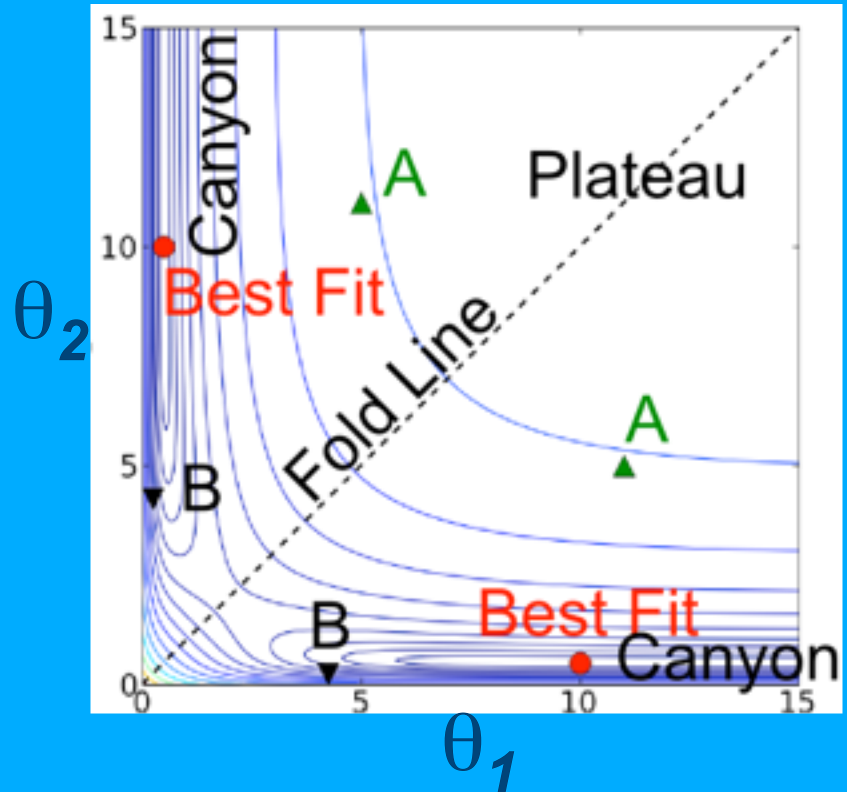


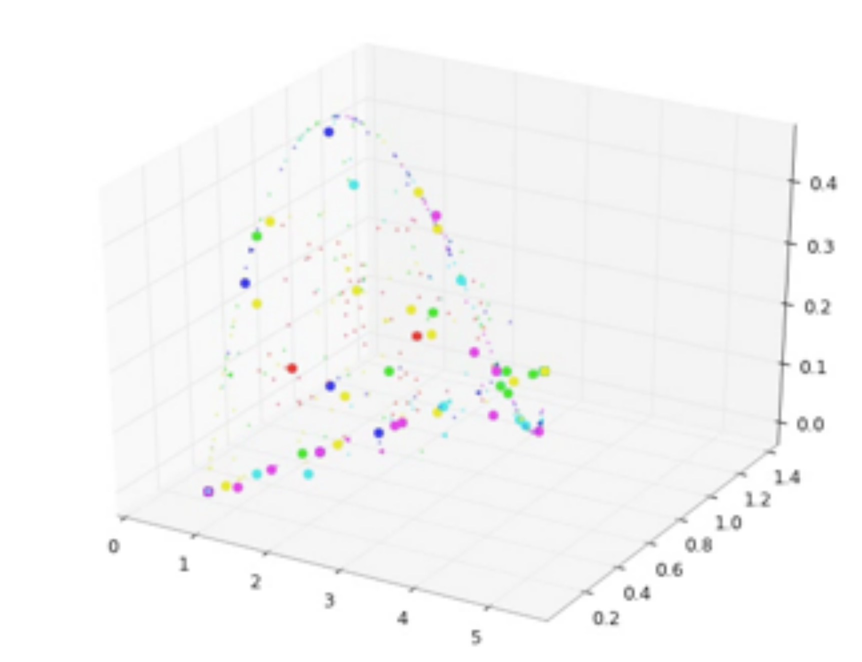


Edges of the model manifold

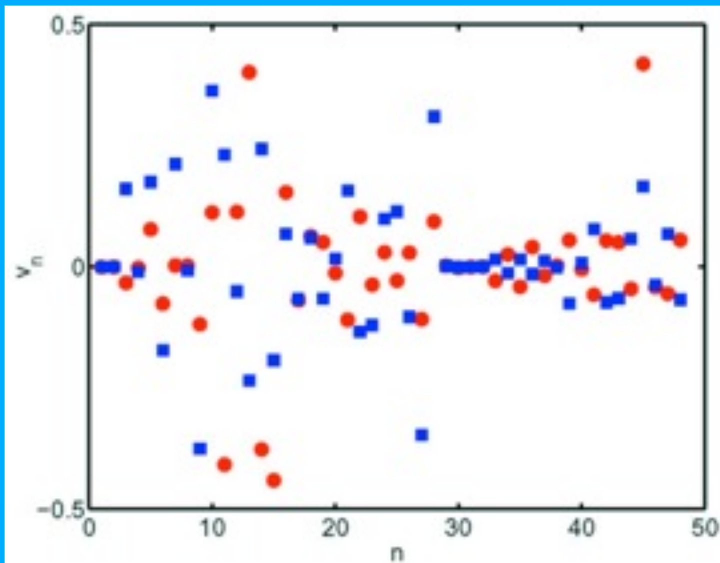
Fitting Exponentials
Top: Flat model manifold; articulated edges = plateau
Bottom: Stretch to uniform aspect ratio (Isabel Kloumann)



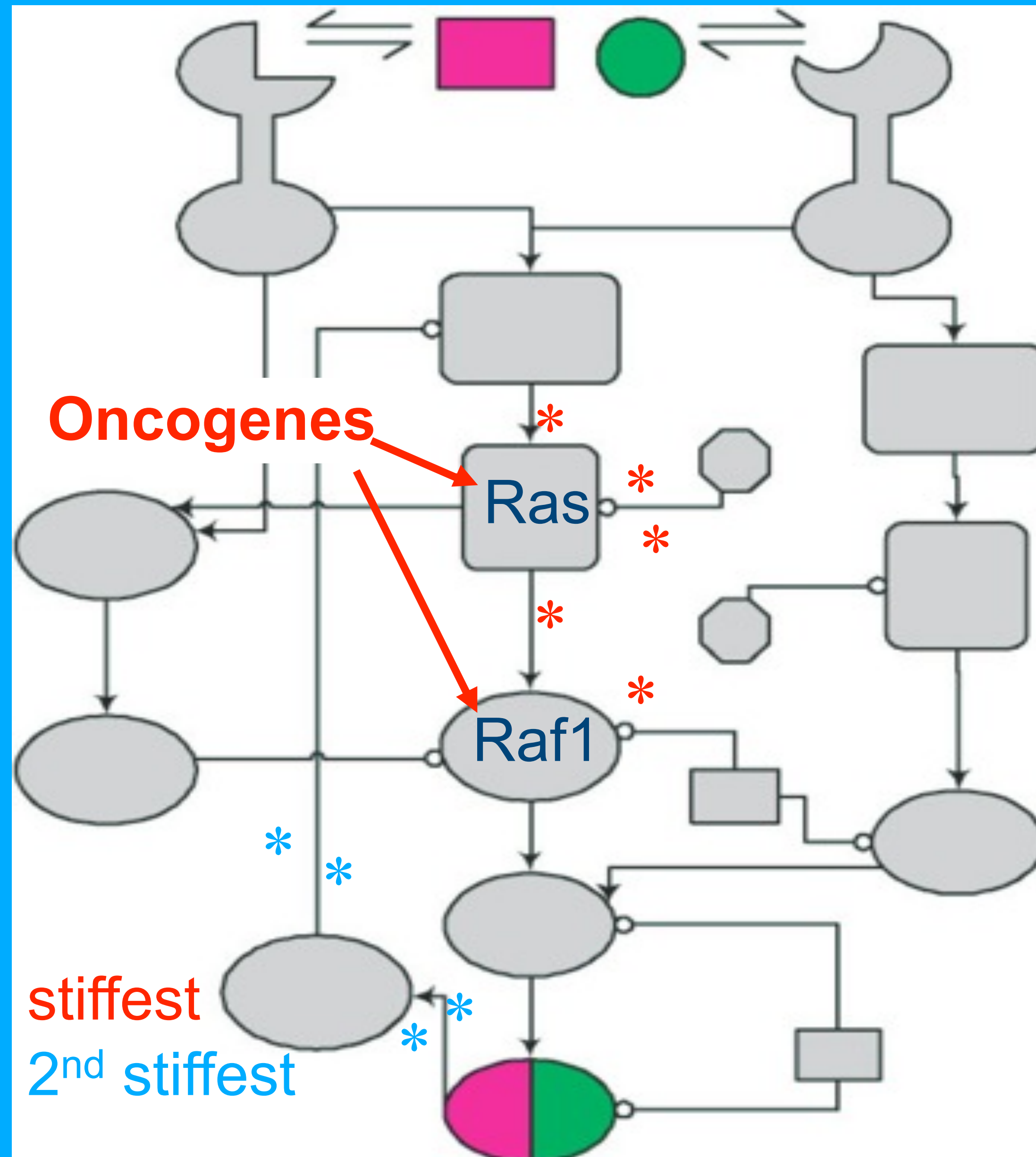




Which Rate Constants are in the Stiffest Eigenvector?

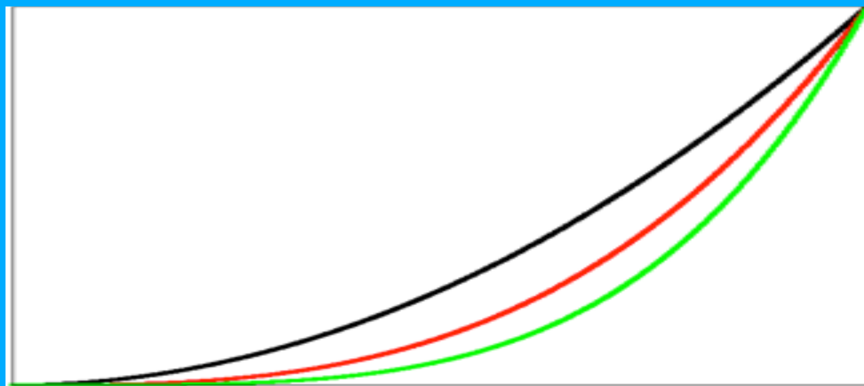


Eigenvector components along the bare parameters reveal which ones are most important for a given eigenvector.



Where is Sloppiness From?

Fitting Polynomials to Data



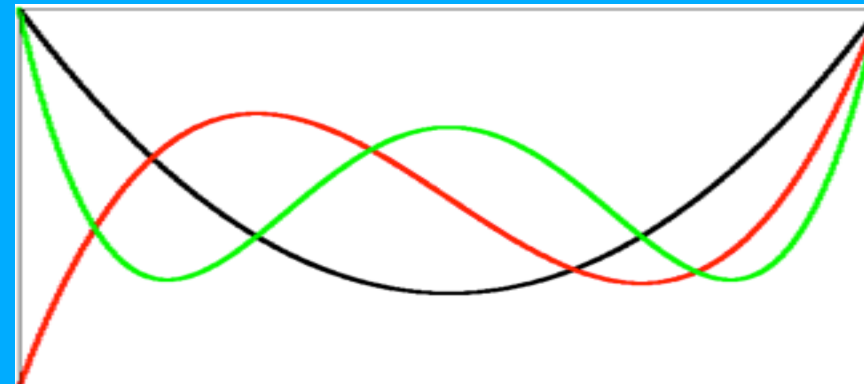
Fitting Monomials to Data

$y = \sum a_n x^n$

Functional Forms Same

Hessian $H_{ij} = 1/(i+j+1)$

Hilbert matrix: famous



Orthogonal Polynomials

$y = \sum b_n L_n(x)$

Functional Forms Distinct

Eigen Parameters

Hessian $H_{ij} = \delta_{ij}$

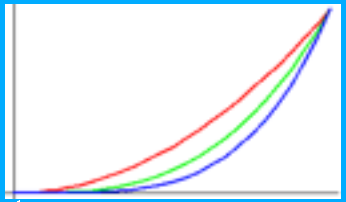
Sloppiness arises when bare parameters skew in eigenbasis

Small Determinant!

$|H| = \prod \lambda_n$

Proposed universal ensemble

Why are they sloppy?



Assumptions: (Not one experiment per parameter)

i. **Model predictions all depend on every parameter, *symmetrically*:** $y_i(\theta_1, \theta_2, \theta_3) = y_i(\theta_2, \theta_3, \theta_1)$

ii. **Parameters are nearly degenerate:** $\theta_\iota = \theta_0 + \varepsilon_i$

$$H = J^T J = V^T A^T A V$$

$$V = \begin{bmatrix} 1 & 1 & \cdots & 1 \\ \varepsilon_1 & \varepsilon_2 & \cdots & \varepsilon_N \\ \vdots & \vdots & \ddots & \vdots \\ \varepsilon_1^d & \varepsilon_2^d & \cdots & \varepsilon_N^d \end{bmatrix}$$

Vandermonde Matrix

$$\det(V) = \prod_{i<j} (\varepsilon_i - \varepsilon_j) \propto \varepsilon^{N(N-1)/2}$$

d=N-1

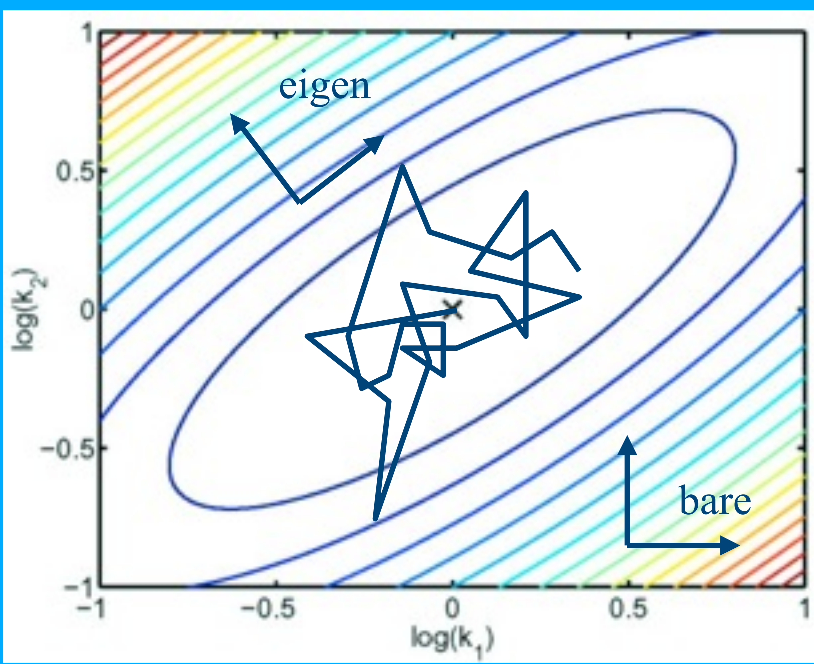
- Implies enormous range of eigenvalues
- Implies equal spacing of log eigenvalues
- Like universality for random matrices

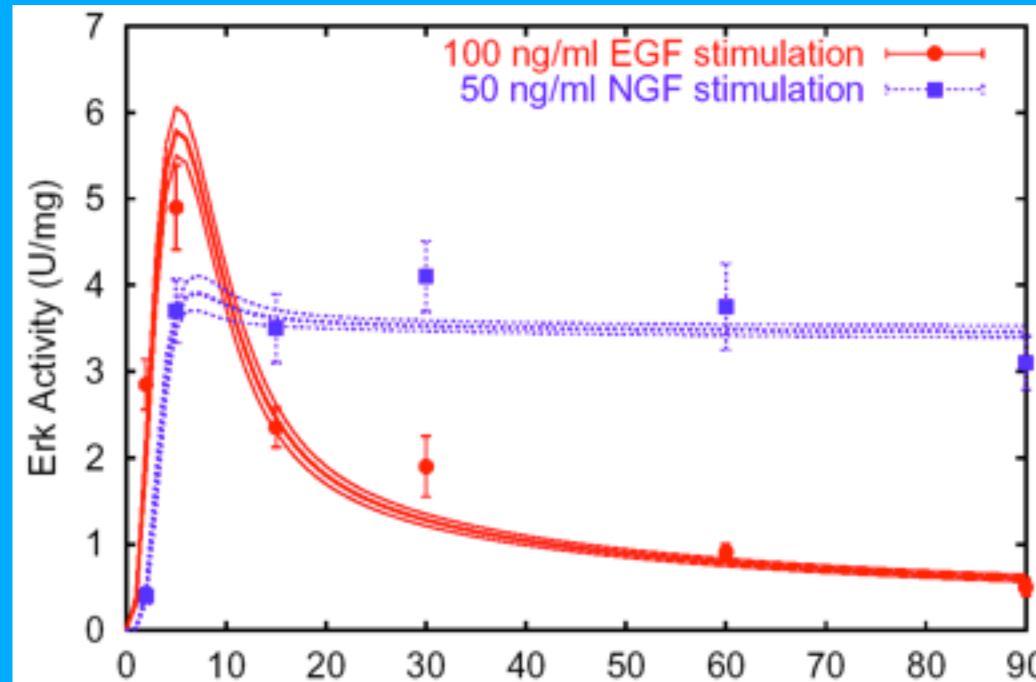
48 Parameter "Fit" to Data

Cost is Energy

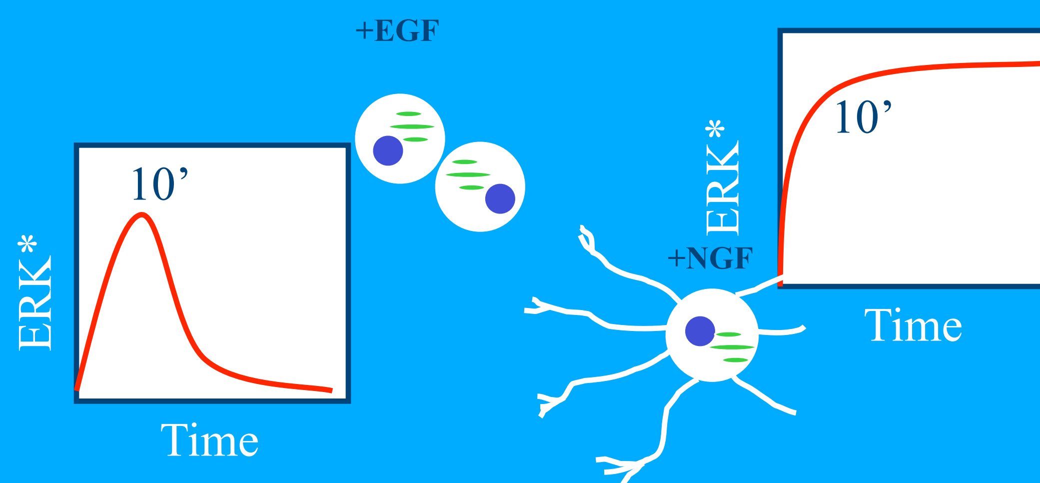
$$C(\vec{\theta}) = \frac{1}{2}\sum_{i=1}^{N_D} \frac{(y(\vec{\theta}) - y_i)^2}{\sigma_i^2}$$

Ensemble of Fits
Gives Error Bars





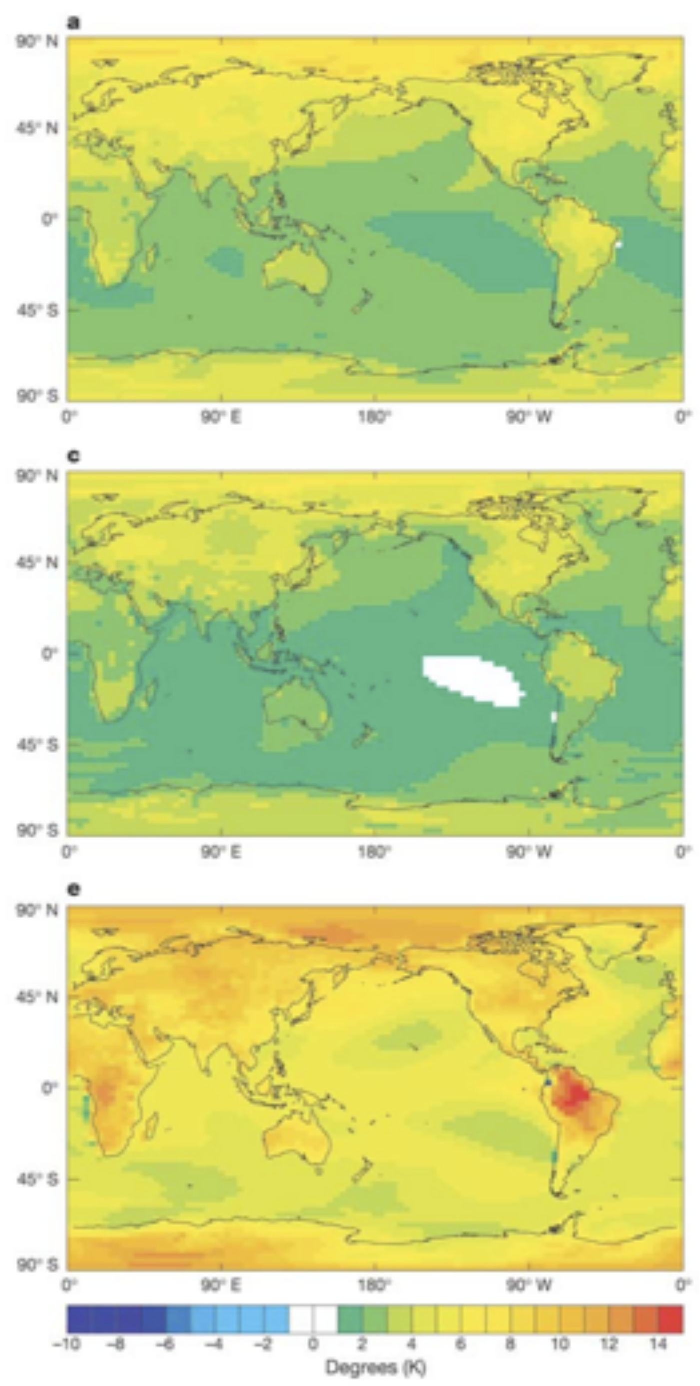
Error Bars from Data Uncertainty



Exploring Parameter Space

Rugged? More like Grand Canyon (Josh)

- Glasses: Rugged Landscape
  - Metastable Local Valleys
  - Transition State Passes
- Optimization Hell: Golf Course
- Sloppy Models
  - Minima: 5 stiff, N-5 sloppy
  - Search: Flat planes with cliffs



Climate Change

Climate models contain many unknown parameters, fit to data

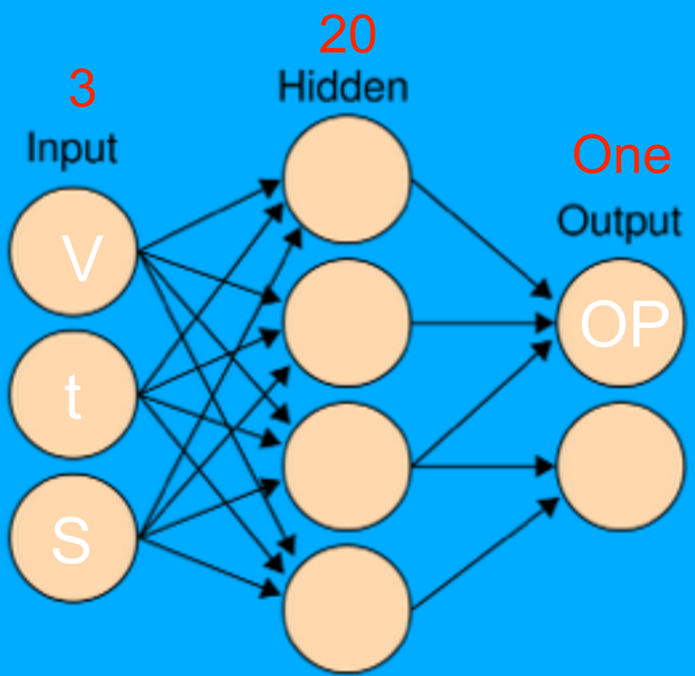
- General Circulation Model (air, oceans, clouds), exploring doubling of CO_2
- 21 total parameters
- Initial conditions and (only) 6 "cloud dynamics" parameters varied
- Heating typically 3.4K, ranged from < 2K to > 11K

Stainforth et al., *Uncertainty in predictions of the climate response to rising levels of greenhouse gases,* **Nature 433**, 403-406 (2005)

Yan-Jiun Chen

Neural Networks

Mark Transtrum



V	t	S	OP
0.20	5.0	75.	25.0000
0.40	5.0	93.	7.2537
0.40	15.0	79.	21.0225
0.66	10.0	91.	10.3957

…

- Neural net “trained” to predict Black-Scholes output option price OP, given inputs volatility V, time t, and strike S
- Each circular “neuron” has sigmoidal response signal s_j to input signals s_i:

$$s_j = \tanh(\Sigma_i \, w_{ij} \, s_i)$$

- Inputs and outputs scaled to [-1,1]
- 101 parameters w_{ij} fit to 1530 data points

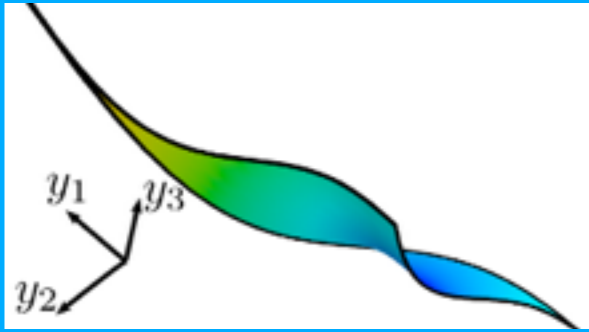
(http://www.scientific-consultants.com/nnbd.html)

Mark Transtrum

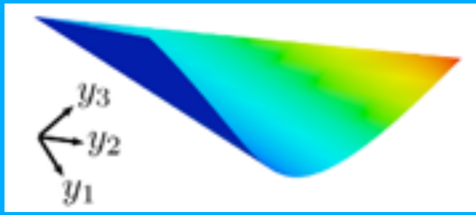
Curvatures

Intrinsic curvature $R^{\mu}{}_{\nu\alpha\beta}$

- determines geodesic shortest paths
- independent of embedding, parameters

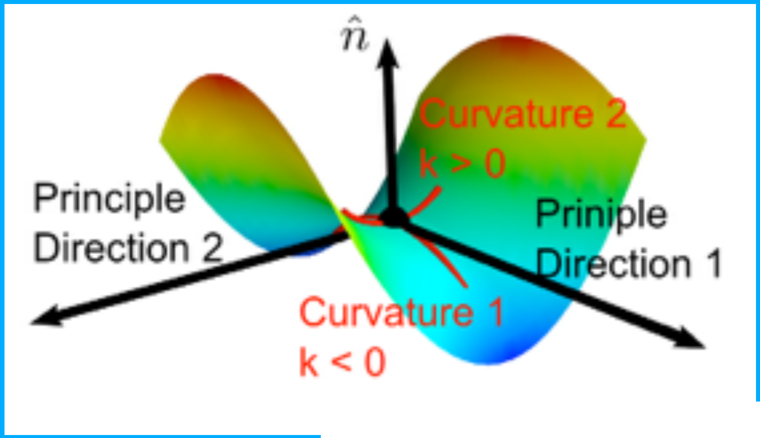


No intrinsic curvature



Extrinsic curvature

- also measures bending in embedding space (i.e., cylinder)
- independent of parameters
- Shape operator, geodesic curvature

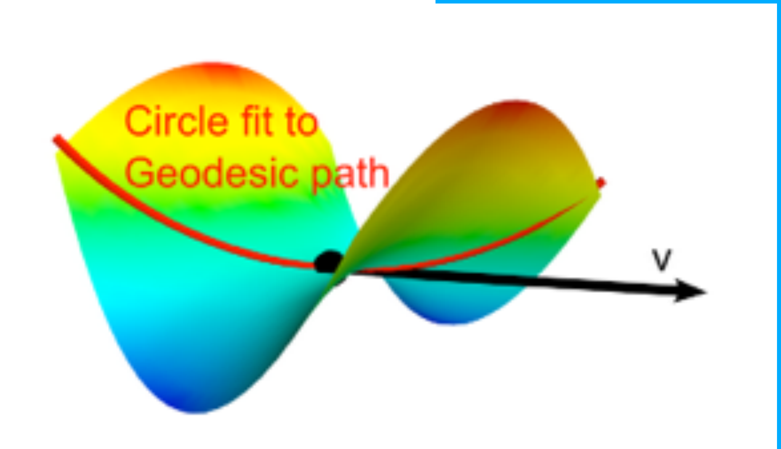


Shape Operator

Parameter effects "curvature"

- Usually much the largest
- Defined in analogy to extrinsic curvature (projecting out of surface, rather than into)

Geodesic Curvature



Why is it so thin?

Model $f(t,\theta)$ analytic:

$f^{(n)}(t)/n! \leq R^{-n}$

Polynomial fit $P_{m-1}(t)$

to $f(t_1),\ldots,f(t_m)$

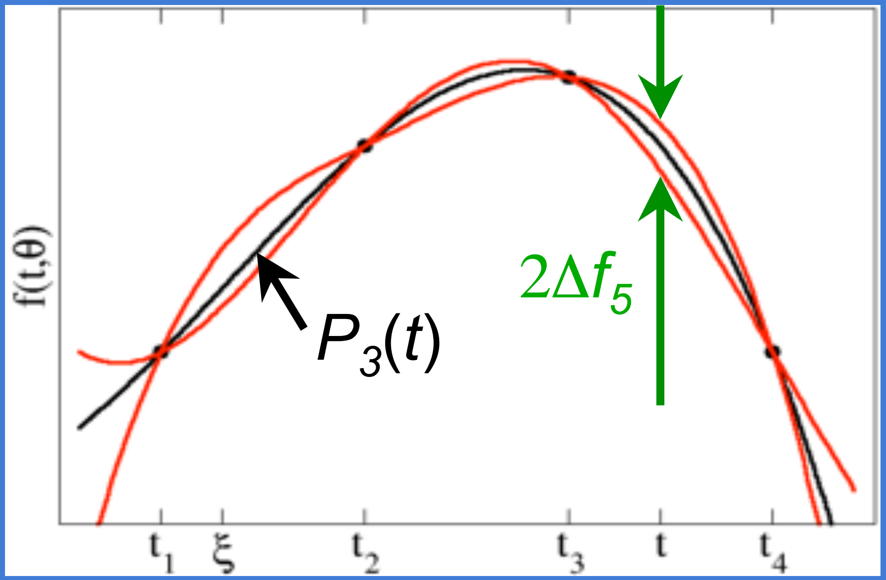
Interpolation convergence theorem

$\Delta f_{m+1} = f(t)-P_{m-1}(t)$

$< (t-t_1)\ldots(t-t_m)\, f^{(m)}(\xi)/m!$

$\sim (\Delta t / R)^m$

More than one data per R



Hyper-ribbon: Cross-section constraining m points has width $W_{m+1} \sim \Delta f_{m+1} \sim (\Delta t/R)^m$

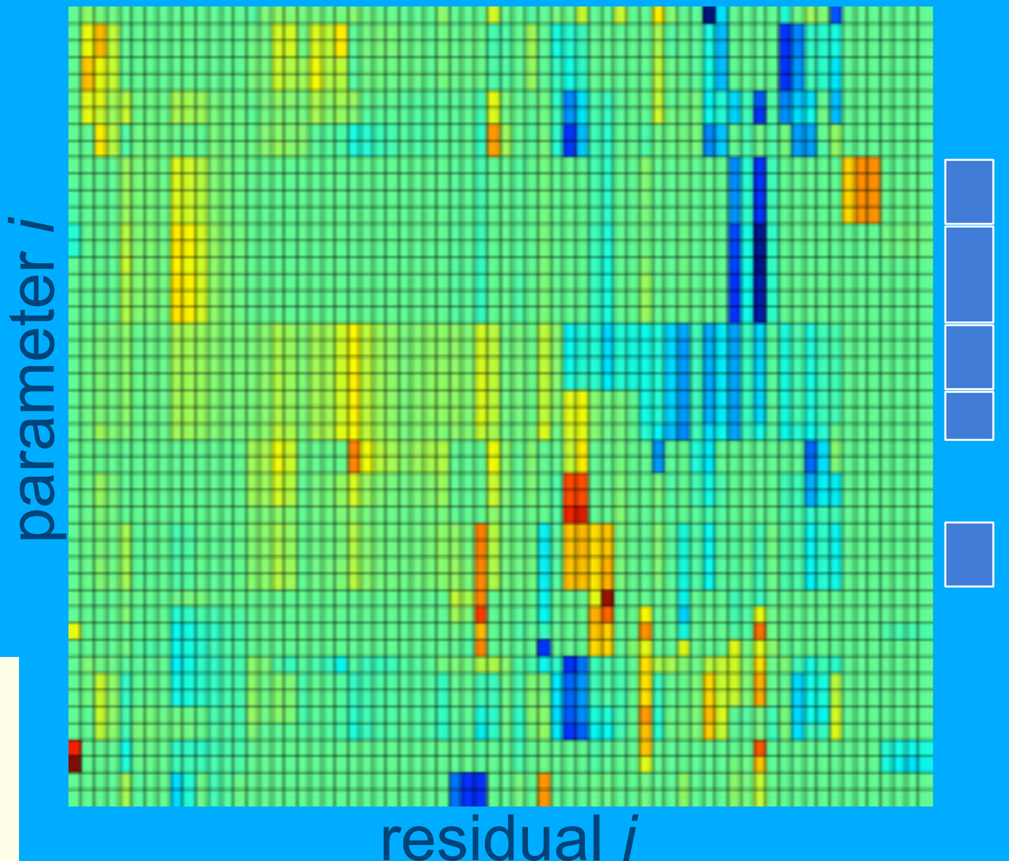
B. Finding sloppy subsystems

Model reduction?

- Sloppy model as multiple redundant parameters?
- Subsystem = subspace of parameters p_i with similar effects on model behavior
- Similar = same effects on residuals r_j
- Apply clustering algorithm to rows of $J_{ij}^T = \partial r_j / \partial p_i$

Continuum mechanics, renormalization group, Lyapunov exponents can also be viewed as sloppy model reduction

PC12 differentiation model



References

"***The sloppy model universality class and the Vandermonde matrix***", J. J. Waterfall, F. P. Casey, R. N. Gutenkunst, K. S. Brown, C. R. Myers, P. W. Brouwer, V. Elser, and James P. Sethna, http://arxiv.org/abs/cond-mat/0605387.

"***Sloppy systems biology: tight predictions with loose parameters***", Ryan N. Gutenkunst, Joshua J. Waterfall, Fergal P. Casey, Kevin S. Brown, Christopher R. Myers & James P. Sethna (submitted).

"***The Statistical Mechanics of Complex Signaling Networks: Nerve Growth Factor Signaling***", K. S. Brown, C. C. Hill, G. A. Calero, C. R. Myers, K. H. Lee, J. P. Sethna, and R. A. Cerione, Physical Biology 1, 184-195 (2004) .

"***Statistical Mechanics Approaches to Models with Many Poorly Known Parameters***", K. S. Brown and J. P. Sethna, Phys. Rev. E 68, 021904 (2003).

"***Bayesian Ensemble Approach to Error Estimation of Interatomic Potentials***", Søren L. Frederiksen, Karsten W. Jacobsen, Kevin S. Brown, and James P. Sethna, *Phys. Rev. Letters* **93**, 165501 (2004).

"***Bayesian Error Estimation in Density Functional Theory***", J. J. Mortensen, K. Kaasbjerg, S. L. Frederiksen, J. K. Norskov, James P. Sethna, K. W. Jacobsen, *Phys. Rev. Letters* **95**, 216401 (2005).

"***SloppyCell***" systems biology modeling software, Ryan N. Gutenkunst, Christopher R. Myers, Kevin S. Brown, Joshua J. Waterfall, Fergal P. Casey, James P. Sethna http://www.lassp.cornell.edu/sethna/GeneDynamics/, SourceForge repository at http://sloppycell.sourceforge.net/