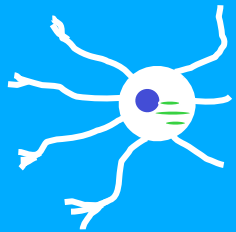
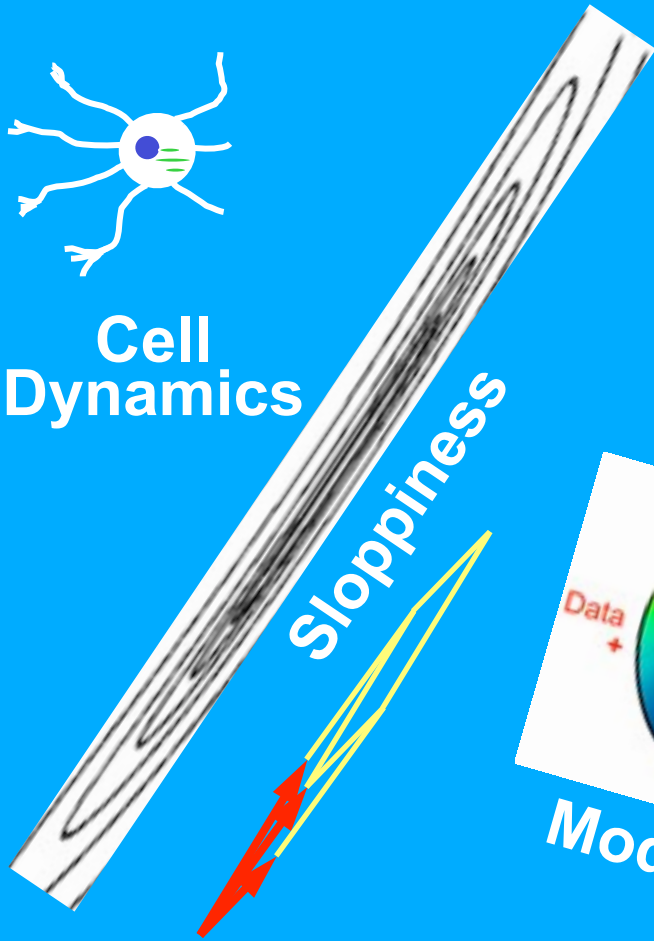


'Sloppy Model' Nonlinear Fits: Signal Transduction to Differential Geometry

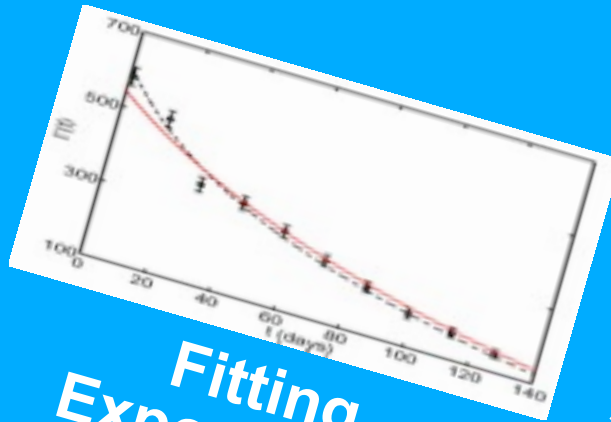
JPS, Mark Transtrum, Ben Machta, Ricky Chachra, Isabel Kloumann, Kevin Brown, Ryan Gutenkunst, Josh Waterfall, Chris Myers, ...



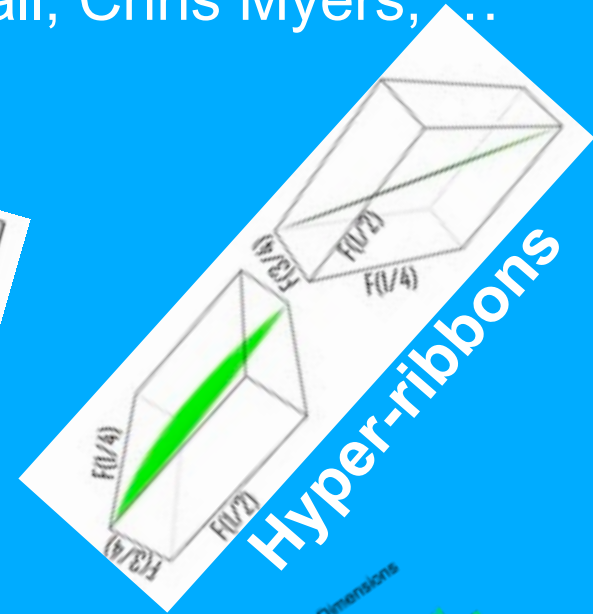
Cell Dynamics



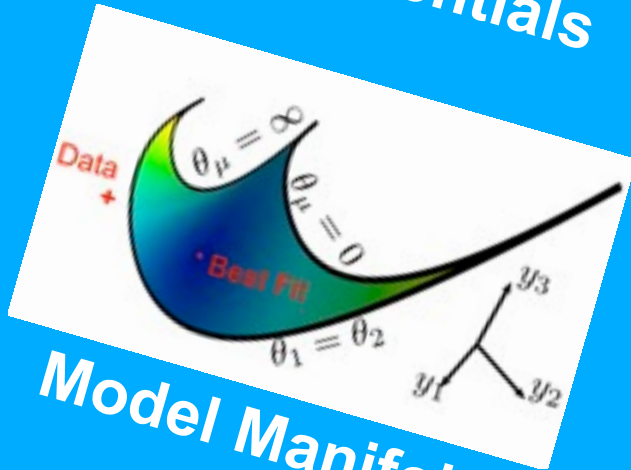
Sloppiness



Fitting Exponentials



Hyper-ribbons



Model Manifold

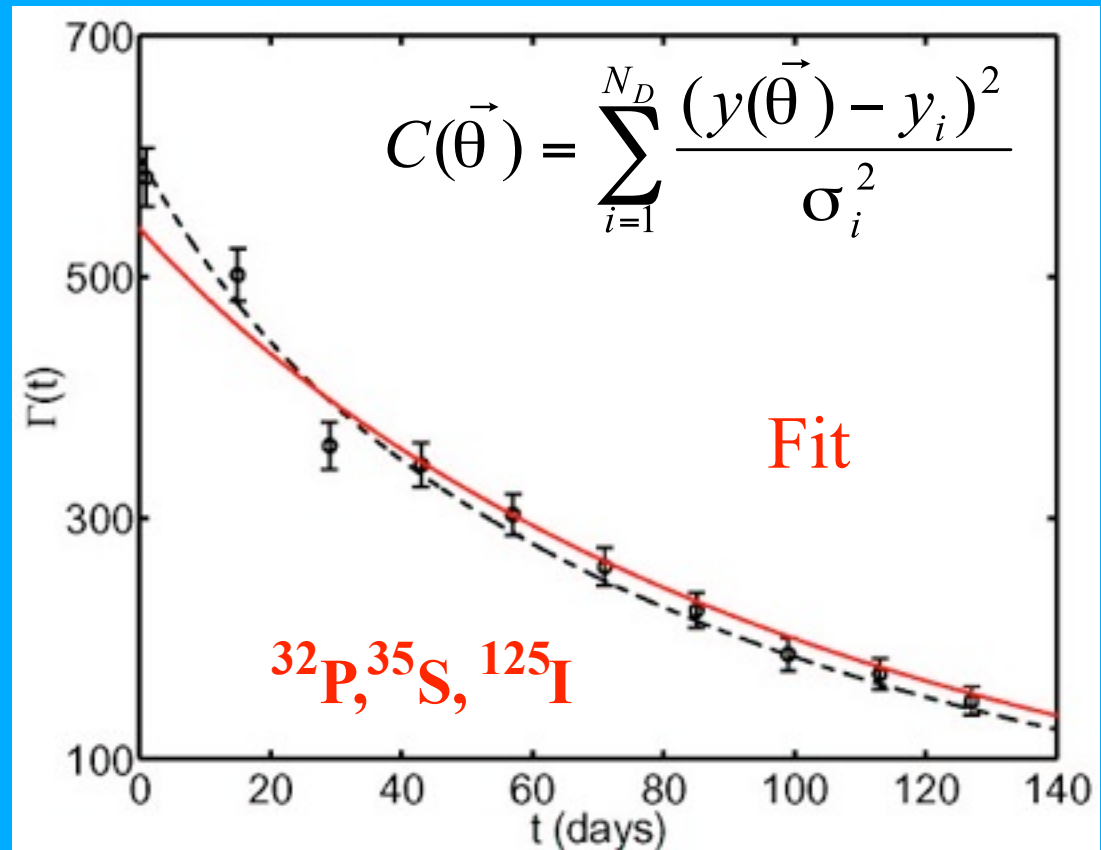


Coarse-Grained Models

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!



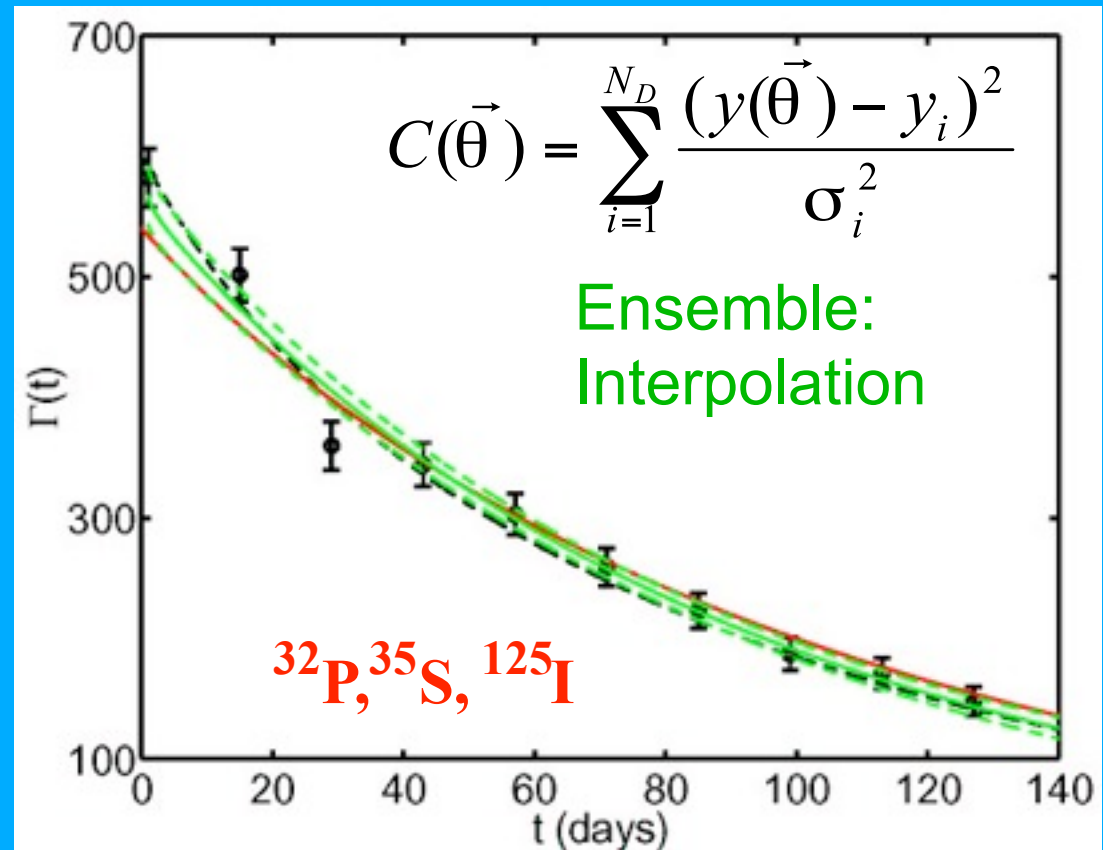
$$y(A, \gamma, t) = A_1 e^{-\gamma_1 t} + A_2 e^{-\gamma_2 t} + A_3 e^{-\gamma_3 t}$$

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!



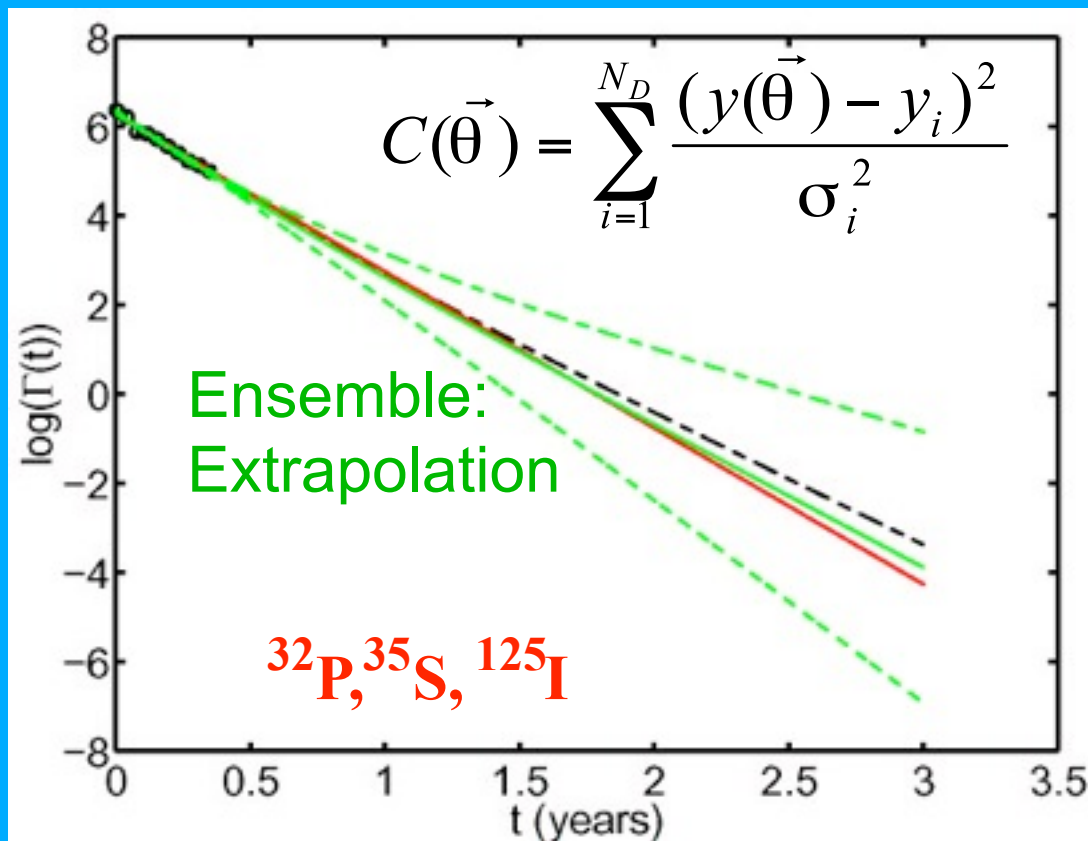
$$y(A, \gamma, t) = A_1 e^{-\gamma_1 t} + A_2 e^{-\gamma_2 t} + A_3 e^{-\gamma_3 t}$$

6 Parameter Fit

Fitting Decaying Exponentials

Classic ill-posed inverse problem

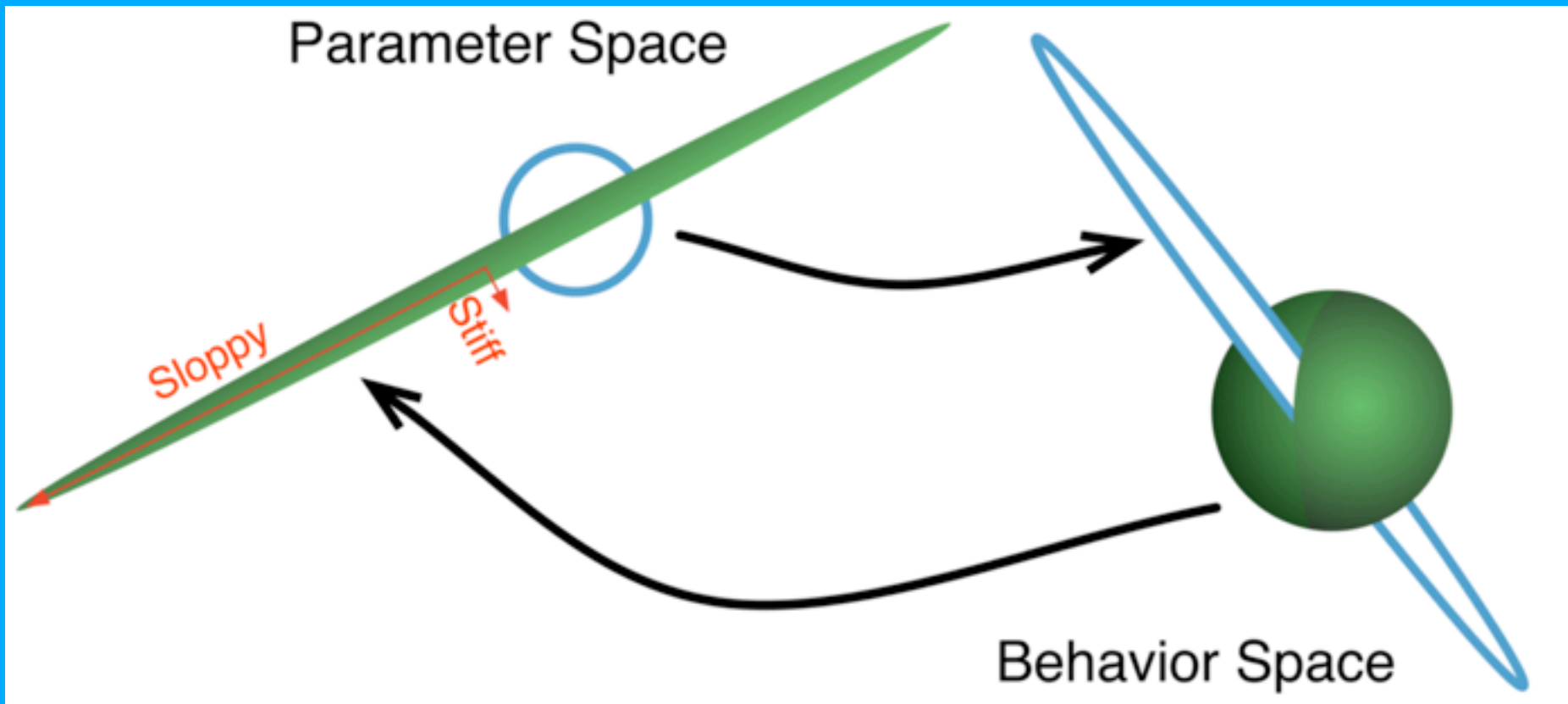
Given Geiger counter measurements from a radioactive pile, can we recover the identity of the elements and/or predict future radioactivity? Good fits with bad decay rates!



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

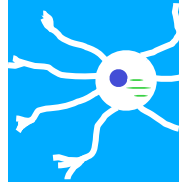
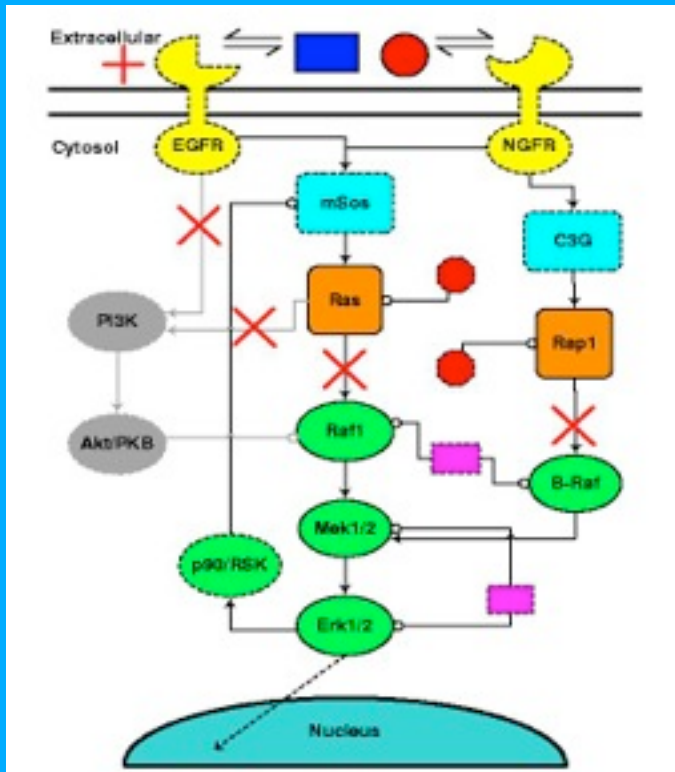
6 Parameter Fit

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.

Systems Biology: Cell Protein Reactions



$$\frac{d[EGF]}{dt} = -k_{deg} [EGF] + k_{prod} [EGF]$$

$$\frac{d[EGFR]}{dt} = -k_{deg} [EGFR] + k_{prod} [EGFR]$$

$$\frac{d[mSos]}{dt} = -k_{deg} [mSos] + k_{prod} [mSos]$$

$$\frac{d[Ras]}{dt} = -k_{deg} [Ras] + k_{prod} [Ras]$$

$$\frac{d[Raf1]}{dt} = -k_{deg} [Raf1] + k_{prod} [Raf1]$$

$$\frac{d[Mek1/2]}{dt} = -k_{deg} [Mek1/2] + k_{prod} [Mek1/2]$$

$$\frac{d[Erk1/2]}{dt} = -k_{deg} [Erk1/2] + k_{prod} [Erk1/2]$$

$$\frac{d[NGFR]}{dt} = -k_{deg} [NGFR] + k_{prod} [NGFR]$$

$$\frac{d[C3G]}{dt} = -k_{deg} [C3G] + k_{prod} [C3G]$$

$$\frac{d[Rap1]}{dt} = -k_{deg} [Rap1] + k_{prod} [Rap1]$$

$$\frac{d[B-Raf]}{dt} = -k_{deg} [B-Raf] + k_{prod} [B-Raf]$$

$$\frac{d[RSK]}{dt} = -k_{deg} [RSK] + k_{prod} [RSK]$$

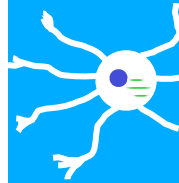
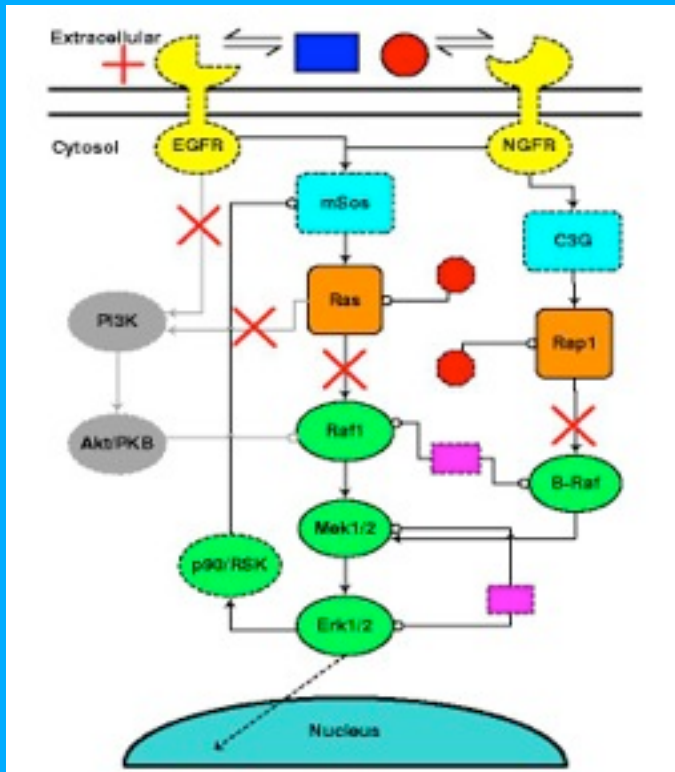
$$\frac{d[Akt/PKB]}{dt} = -k_{deg} [Akt/PKB] + k_{prod} [Akt/PKB]$$

$$\frac{d[p90/RSK]}{dt} = -k_{deg} [p90/RSK] + k_{prod} [p90/RSK]$$

$$\frac{d[Nucleus]}{dt} = -k_{deg} [Nucleus] + k_{prod} [Nucleus]$$

48 Parameter Fit!

Systems Biology: Cell Protein Reactions

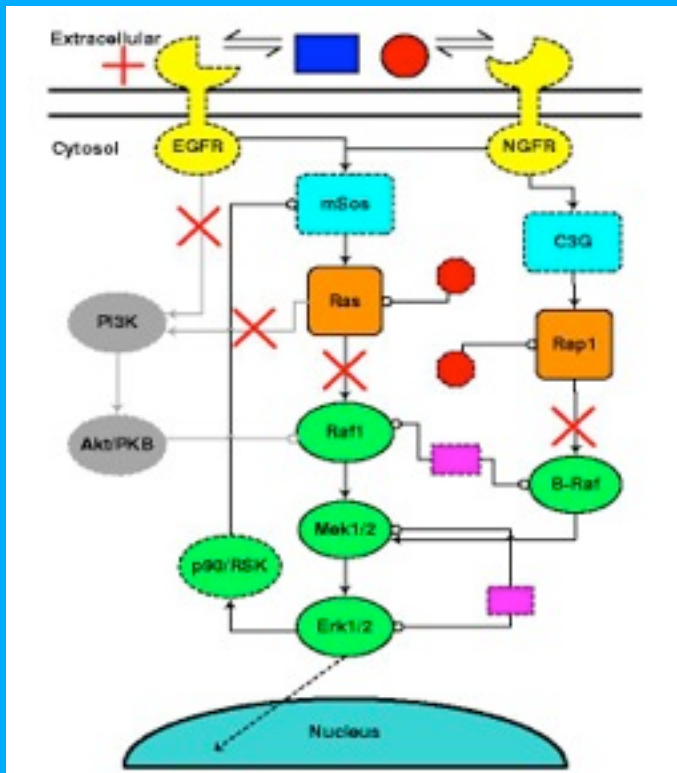


$\frac{d[EGF]}{dt} = -k_{EGF} [EGF] [mEGFR_{tot}] + k_{EGF} [mEGFR]$	$\frac{d[EGFR]}{dt} = -k_{EGFR} [EGFR] [mEGFR_{tot}] + k_{EGFR} [mEGFR]$	$\frac{d[EGFR_{tot}]}{dt} = -k_{EGFR_{tot}} [EGFR_{tot}] [mEGFR_{tot}] + k_{EGFR_{tot}} [mEGFR_{tot}]$	$\frac{d[EGFR_{tot}]}{dt} = -k_{EGFR_{tot}} [EGFR_{tot}] [mEGFR_{tot}] + k_{EGFR_{tot}} [mEGFR_{tot}]$
$\frac{d[EGFR]}{dt} = -k_{EGFR} [EGFR] [mEGFR_{tot}] + k_{EGFR} [mEGFR]$	$\frac{d[EGFR]}{dt} = -k_{EGFR} [EGFR] [mEGFR_{tot}] + k_{EGFR} [mEGFR]$	$\frac{d[EGFR_{tot}]}{dt} = -k_{EGFR_{tot}} [EGFR_{tot}] [mEGFR_{tot}] + k_{EGFR_{tot}} [mEGFR_{tot}]$	$\frac{d[EGFR_{tot}]}{dt} = -k_{EGFR_{tot}} [EGFR_{tot}] [mEGFR_{tot}] + k_{EGFR_{tot}} [mEGFR_{tot}]$
$\frac{d[mEGFR]}{dt} = -k_{mEGFR} [mEGFR] [EGFR] + k_{mEGFR} [EGFR] [mEGFR]$	$\frac{d[mEGFR]}{dt} = -k_{mEGFR} [mEGFR] [EGFR] + k_{mEGFR} [EGFR] [mEGFR]$	$\frac{d[EGFR_{tot}]}{dt} = -k_{EGFR_{tot}} [EGFR_{tot}] [mEGFR_{tot}] + k_{EGFR_{tot}} [mEGFR_{tot}]$	$\frac{d[EGFR_{tot}]}{dt} = -k_{EGFR_{tot}} [EGFR_{tot}] [mEGFR_{tot}] + k_{EGFR_{tot}} [mEGFR_{tot}]$
$\frac{d[EGFR_{tot}]}{dt} = -k_{EGFR_{tot}} [EGFR_{tot}] [mEGFR_{tot}] + k_{EGFR_{tot}} [mEGFR_{tot}]$	$\frac{d[EGFR_{tot}]}{dt} = -k_{EGFR_{tot}} [EGFR_{tot}] [mEGFR_{tot}] + k_{EGFR_{tot}} [mEGFR_{tot}]$	$\frac{d[EGFR_{tot}]}{dt} = -k_{EGFR_{tot}} [EGFR_{tot}] [mEGFR_{tot}] + k_{EGFR_{tot}} [mEGFR_{tot}]$	$\frac{d[EGFR_{tot}]}{dt} = -k_{EGFR_{tot}} [EGFR_{tot}] [mEGFR_{tot}] + k_{EGFR_{tot}} [mEGFR_{tot}]$
$\frac{d[Ras]}{dt} = -k_{Ras} [Ras] [Raf1] + k_{Ras} [Raf1] [Ras]$	$\frac{d[Ras]}{dt} = -k_{Ras} [Ras] [Raf1] + k_{Ras} [Raf1] [Ras]$	$\frac{d[Rap1]}{dt} = -k_{Rap1} [Rap1] [C3G] + k_{Rap1} [C3G] [Rap1]$	$\frac{d[Rap1]}{dt} = -k_{Rap1} [Rap1] [C3G] + k_{Rap1} [C3G] [Rap1]$
$\frac{d[Raf1]}{dt} = -k_{Raf1} [Raf1] [Ras] + k_{Raf1} [Ras] [Raf1]$	$\frac{d[Raf1]}{dt} = -k_{Raf1} [Raf1] [Ras] + k_{Raf1} [Ras] [Raf1]$	$\frac{d[B-Raf]}{dt} = -k_{B-Raf} [B-Raf] [Rap1] + k_{B-Raf} [Rap1] [B-Raf]$	$\frac{d[B-Raf]}{dt} = -k_{B-Raf} [B-Raf] [Rap1] + k_{B-Raf} [Rap1] [B-Raf]$
$\frac{d[Mek1/2]}{dt} = -k_{Mek1/2} [Mek1/2] [Raf1] + k_{Mek1/2} [Raf1] [Mek1/2]$	$\frac{d[Mek1/2]}{dt} = -k_{Mek1/2} [Mek1/2] [Raf1] + k_{Mek1/2} [Raf1] [Mek1/2]$	$\frac{d[Erk1/2]}{dt} = -k_{Erk1/2} [Erk1/2] [Mek1/2] + k_{Erk1/2} [Mek1/2] [Erk1/2]$	$\frac{d[Erk1/2]}{dt} = -k_{Erk1/2} [Erk1/2] [Mek1/2] + k_{Erk1/2} [Mek1/2] [Erk1/2]$

Note: The table above contains repeated text from the image. A red circle highlights a specific section of the table, and a red arrow points from this circle to the '48 Parameter Fit!' text below.

48 Parameter Fit!

Systems Biology: Cell Protein Reactions

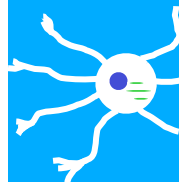
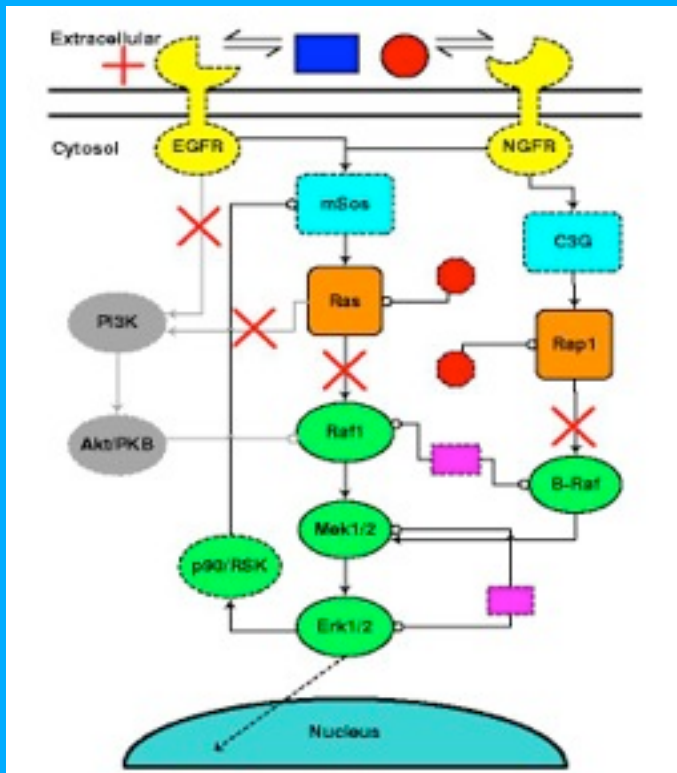


$\frac{d[EGF]}{dt} = -k_{EGF} [EGF] + k_{EGF} [boundEGFR]$	$\frac{d[boundEGFR]}{dt} = k_{EGF} [EGF] [EGFR] - k_{dEGF} [boundEGFR]$	$\frac{d[EGFR]}{dt} = k_{EGFR} [EGFR] - k_{dEGFR} [EGFR]$	$\frac{d[EGF]}{dt} = -k_{EGF} [EGF] + k_{EGF} [boundEGFR]$	$\frac{d[boundEGFR]}{dt} = k_{EGF} [EGF] [EGFR] - k_{dEGF} [boundEGFR]$	$\frac{d[EGFR]}{dt} = k_{EGFR} [EGFR] - k_{dEGFR} [EGFR]$
$\frac{d[NGF]}{dt} = -k_{NGF} [NGF] + k_{NGF} [boundNGFR]$	$\frac{d[boundNGFR]}{dt} = k_{NGF} [NGF] [NGFR] - k_{dNGF} [boundNGFR]$	$\frac{d[NGFR]}{dt} = k_{NGFR} [NGFR] - k_{dNGFR} [NGFR]$	$\frac{d[NGF]}{dt} = -k_{NGF} [NGF] + k_{NGF} [boundNGFR]$	$\frac{d[boundNGFR]}{dt} = k_{NGF} [NGF] [NGFR] - k_{dNGF} [boundNGFR]$	$\frac{d[NGFR]}{dt} = k_{NGFR} [NGFR] - k_{dNGFR} [NGFR]$
$\frac{d[SosInactive]}{dt} = -k_{EGF} [boundEGFR] \frac{[SosInactive]}{[SosInactive] + K_{mEGF}} - k_{NGF} [boundNGFR] \frac{[SosInactive]}{[SosInactive] + K_{mNGF}} + k_{dSos} [P90RskActive] \frac{[SosActive]}{[SosActive] + K_{mdSos}}$	$\frac{d[SosActive]}{dt} = k_{EGF} [boundEGFR] \frac{[SosInactive]}{[SosInactive] + K_{mEGF}} + k_{NGF} [boundNGFR] \frac{[SosInactive]}{[SosInactive] + K_{mNGF}} - k_{dSos} [P90RskActive] \frac{[SosActive]}{[SosActive] + K_{mdSos}}$	$\frac{d[RasInactive]}{dt} = -k_{Ras} [RasInactive] + k_{dRas} [RasActive]$	$\frac{d[RasActive]}{dt} = k_{Ras} [RasInactive] - k_{dRas} [RasActive]$	$\frac{d[boundRas]}{dt} = k_{Ras} [RasActive] [EGFR] - k_{dRas} [boundRas]$	$\frac{d[RasActive]}{dt} = k_{Ras} [RasInactive] - k_{dRas} [RasActive]$
$\frac{d[Rap1Inactive]}{dt} = -k_{Rap1} [Rap1Inactive] + k_{dRap1} [Rap1Active]$	$\frac{d[Rap1Active]}{dt} = k_{Rap1} [Rap1Inactive] - k_{dRap1} [Rap1Active]$	$\frac{d[boundRap1]}{dt} = k_{Rap1} [Rap1Active] [NGFR] - k_{dRap1} [boundRap1]$	$\frac{d[Rap1Active]}{dt} = k_{Rap1} [Rap1Inactive] - k_{dRap1} [Rap1Active]$	$\frac{d[boundRap1]}{dt} = k_{Rap1} [Rap1Active] [NGFR] - k_{dRap1} [boundRap1]$	$\frac{d[Rap1Active]}{dt} = k_{Rap1} [Rap1Inactive] - k_{dRap1} [Rap1Active]$
$\frac{d[Raf1Inactive]}{dt} = -k_{Raf1} [Raf1Inactive] + k_{dRaf1} [Raf1Active]$	$\frac{d[Raf1Active]}{dt} = k_{Raf1} [Raf1Inactive] - k_{dRaf1} [Raf1Active]$	$\frac{d[boundRaf1]}{dt} = k_{Raf1} [Raf1Active] [RasActive] - k_{dRaf1} [boundRaf1]$	$\frac{d[Raf1Active]}{dt} = k_{Raf1} [Raf1Inactive] - k_{dRaf1} [Raf1Active]$	$\frac{d[boundRaf1]}{dt} = k_{Raf1} [Raf1Active] [RasActive] - k_{dRaf1} [boundRaf1]$	$\frac{d[Raf1Active]}{dt} = k_{Raf1} [Raf1Inactive] - k_{dRaf1} [Raf1Active]$
$\frac{d[B-RafInactive]}{dt} = -k_{B-Raf} [B-RafInactive] + k_{dB-Raf} [B-RafActive]$	$\frac{d[B-RafActive]}{dt} = k_{B-Raf} [B-RafInactive] - k_{dB-Raf} [B-RafActive]$	$\frac{d[boundB-Raf]}{dt} = k_{B-Raf} [B-RafActive] [Rap1Active] - k_{dB-Raf} [boundB-Raf]$	$\frac{d[B-RafActive]}{dt} = k_{B-Raf} [B-RafInactive] - k_{dB-Raf} [B-RafActive]$	$\frac{d[boundB-Raf]}{dt} = k_{B-Raf} [B-RafActive] [Rap1Active] - k_{dB-Raf} [boundB-Raf]$	$\frac{d[B-RafActive]}{dt} = k_{B-Raf} [B-RafInactive] - k_{dB-Raf} [B-RafActive]$
$\frac{d[Mek1/2Inactive]}{dt} = -k_{Mek1/2} [Mek1/2Inactive] + k_{dMek1/2} [Mek1/2Active]$	$\frac{d[Mek1/2Active]}{dt} = k_{Mek1/2} [Mek1/2Inactive] - k_{dMek1/2} [Mek1/2Active]$	$\frac{d[boundMek1/2]}{dt} = k_{Mek1/2} [Mek1/2Active] [Raf1Active] + k_{Mek1/2} [Mek1/2Active] [B-RafActive] - k_{dMek1/2} [boundMek1/2]$	$\frac{d[Mek1/2Active]}{dt} = k_{Mek1/2} [Mek1/2Inactive] - k_{dMek1/2} [Mek1/2Active]$	$\frac{d[boundMek1/2]}{dt} = k_{Mek1/2} [Mek1/2Active] [Raf1Active] + k_{Mek1/2} [Mek1/2Active] [B-RafActive] - k_{dMek1/2} [boundMek1/2]$	$\frac{d[Mek1/2Active]}{dt} = k_{Mek1/2} [Mek1/2Inactive] - k_{dMek1/2} [Mek1/2Active]$
$\frac{d[Erk1/2Inactive]}{dt} = -k_{Erk1/2} [Erk1/2Inactive] + k_{dErk1/2} [Erk1/2Active]$	$\frac{d[Erk1/2Active]}{dt} = k_{Erk1/2} [Erk1/2Inactive] - k_{dErk1/2} [Erk1/2Active]$	$\frac{d[boundErk1/2]}{dt} = k_{Erk1/2} [Erk1/2Active] [Mek1/2Active] - k_{dErk1/2} [boundErk1/2]$	$\frac{d[Erk1/2Active]}{dt} = k_{Erk1/2} [Erk1/2Inactive] - k_{dErk1/2} [Erk1/2Active]$	$\frac{d[boundErk1/2]}{dt} = k_{Erk1/2} [Erk1/2Active] [Mek1/2Active] - k_{dErk1/2} [boundErk1/2]$	$\frac{d[Erk1/2Active]}{dt} = k_{Erk1/2} [Erk1/2Inactive] - k_{dErk1/2} [Erk1/2Active]$
$\frac{d[RSKInactive]}{dt} = -k_{RSK} [RSKInactive] + k_{dRSK} [RSKActive]$	$\frac{d[RSKActive]}{dt} = k_{RSK} [RSKInactive] - k_{dRSK} [RSKActive]$	$\frac{d[boundRSK]}{dt} = k_{RSK} [RSKActive] [Erk1/2Active] - k_{dRSK} [boundRSK]$	$\frac{d[RSKActive]}{dt} = k_{RSK} [RSKInactive] - k_{dRSK} [RSKActive]$	$\frac{d[boundRSK]}{dt} = k_{RSK} [RSKActive] [Erk1/2Active] - k_{dRSK} [boundRSK]$	$\frac{d[RSKActive]}{dt} = k_{RSK} [RSKInactive] - k_{dRSK} [RSKActive]$
$\frac{d[AktPKBInactive]}{dt} = -k_{AktPKB} [AktPKBInactive] + k_{dAktPKB} [AktPKBActive]$	$\frac{d[AktPKBActive]}{dt} = k_{AktPKB} [AktPKBInactive] - k_{dAktPKB} [AktPKBActive]$	$\frac{d[boundAktPKB]}{dt} = k_{AktPKB} [AktPKBActive] [RasActive] - k_{dAktPKB} [boundAktPKB]$	$\frac{d[AktPKBActive]}{dt} = k_{AktPKB} [AktPKBInactive] - k_{dAktPKB} [AktPKBActive]$	$\frac{d[boundAktPKB]}{dt} = k_{AktPKB} [AktPKBActive] [RasActive] - k_{dAktPKB} [boundAktPKB]$	$\frac{d[AktPKBActive]}{dt} = k_{AktPKB} [AktPKBInactive] - k_{dAktPKB} [AktPKBActive]$
$\frac{d[p90RSKInactive]}{dt} = -k_{p90RSK} [p90RSKInactive] + k_{dp90RSK} [p90RSKActive]$	$\frac{d[p90RSKActive]}{dt} = k_{p90RSK} [p90RSKInactive] - k_{dp90RSK} [p90RSKActive]$	$\frac{d[boundp90RSK]}{dt} = k_{p90RSK} [p90RSKActive] [RasActive] - k_{dp90RSK} [boundp90RSK]$	$\frac{d[p90RSKActive]}{dt} = k_{p90RSK} [p90RSKInactive] - k_{dp90RSK} [p90RSKActive]$	$\frac{d[boundp90RSK]}{dt} = k_{p90RSK} [p90RSKActive] [RasActive] - k_{dp90RSK} [boundp90RSK]$	$\frac{d[p90RSKActive]}{dt} = k_{p90RSK} [p90RSKInactive] - k_{dp90RSK} [p90RSKActive]$

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

48 Parameter Fit!

Systems Biology: Cell Protein Reactions



$\frac{d[EGF]}{dt} = -k_{EGF} [EGF] + k_{EGF} [boundEGFR]$	$\frac{d[boundEGFR]}{dt} = k_{EGF} [EGF] [EGFR] - k_{dEGF} [boundEGFR]$	$\frac{d[EGFR]}{dt} = -k_{EGFR} [EGFR] + k_{EGFR} [boundEGFR]$	$\frac{d[EGFR]}{dt} = -k_{EGFR} [EGFR] + k_{EGFR} [boundEGFR]$
$\frac{d[NGF]}{dt} = -k_{NGF} [NGF] + k_{NGF} [boundNGFR]$	$\frac{d[boundNGFR]}{dt} = k_{NGF} [NGF] [NGFR] - k_{dNGF} [boundNGFR]$	$\frac{d[NGFR]}{dt} = -k_{NGFR} [NGFR] + k_{NGFR} [boundNGFR]$	$\frac{d[NGFR]}{dt} = -k_{NGFR} [NGFR] + k_{NGFR} [boundNGFR]$
$\frac{d[mSos]}{dt} = -k_{mSos} [mSos] + k_{mSos} [boundEGFR]$	$\frac{d[boundEGFR]}{dt} = k_{mSos} [mSos] [boundEGFR] - k_{dEGFR} [boundEGFR]$	$\frac{d[EGFR]}{dt} = -k_{EGFR} [EGFR] + k_{EGFR} [boundEGFR]$	$\frac{d[EGFR]}{dt} = -k_{EGFR} [EGFR] + k_{EGFR} [boundEGFR]$
$\frac{d[Ras]}{dt} = -k_{Ras} [Ras] + k_{Ras} [boundEGFR]$	$\frac{d[boundEGFR]}{dt} = k_{Ras} [Ras] [boundEGFR] - k_{dEGFR} [boundEGFR]$	$\frac{d[EGFR]}{dt} = -k_{EGFR} [EGFR] + k_{EGFR} [boundEGFR]$	$\frac{d[EGFR]}{dt} = -k_{EGFR} [EGFR] + k_{EGFR} [boundEGFR]$
$\frac{d[Raf1]}{dt} = -k_{Raf1} [Raf1] + k_{Raf1} [boundEGFR]$	$\frac{d[boundEGFR]}{dt} = k_{Raf1} [Raf1] [boundEGFR] - k_{dEGFR} [boundEGFR]$	$\frac{d[EGFR]}{dt} = -k_{EGFR} [EGFR] + k_{EGFR} [boundEGFR]$	$\frac{d[EGFR]}{dt} = -k_{EGFR} [EGFR] + k_{EGFR} [boundEGFR]$
$\frac{d[Mek1/2]}{dt} = -k_{Mek1/2} [Mek1/2] + k_{Mek1/2} [boundEGFR]$	$\frac{d[boundEGFR]}{dt} = k_{Mek1/2} [Mek1/2] [boundEGFR] - k_{dEGFR} [boundEGFR]$	$\frac{d[EGFR]}{dt} = -k_{EGFR} [EGFR] + k_{EGFR} [boundEGFR]$	$\frac{d[EGFR]}{dt} = -k_{EGFR} [EGFR] + k_{EGFR} [boundEGFR]$
$\frac{d[Erk1/2]}{dt} = -k_{Erk1/2} [Erk1/2] + k_{Erk1/2} [boundEGFR]$	$\frac{d[boundEGFR]}{dt} = k_{Erk1/2} [Erk1/2] [boundEGFR] - k_{dEGFR} [boundEGFR]$	$\frac{d[EGFR]}{dt} = -k_{EGFR} [EGFR] + k_{EGFR} [boundEGFR]$	$\frac{d[EGFR]}{dt} = -k_{EGFR} [EGFR] + k_{EGFR} [boundEGFR]$
$\frac{d[C3G]}{dt} = -k_{C3G} [C3G] + k_{C3G} [boundNGFR]$	$\frac{d[boundNGFR]}{dt} = k_{C3G} [C3G] [boundNGFR] - k_{dNGFR} [boundNGFR]$	$\frac{d[NGFR]}{dt} = -k_{NGFR} [NGFR] + k_{NGFR} [boundNGFR]$	$\frac{d[NGFR]}{dt} = -k_{NGFR} [NGFR] + k_{NGFR} [boundNGFR]$
$\frac{d[Rep1]}{dt} = -k_{Rep1} [Rep1] + k_{Rep1} [boundNGFR]$	$\frac{d[boundNGFR]}{dt} = k_{Rep1} [Rep1] [boundNGFR] - k_{dNGFR} [boundNGFR]$	$\frac{d[NGFR]}{dt} = -k_{NGFR} [NGFR] + k_{NGFR} [boundNGFR]$	$\frac{d[NGFR]}{dt} = -k_{NGFR} [NGFR] + k_{NGFR} [boundNGFR]$
$\frac{d[B-Raf]}{dt} = -k_{B-Raf} [B-Raf] + k_{B-Raf} [boundNGFR]$	$\frac{d[boundNGFR]}{dt} = k_{B-Raf} [B-Raf] [boundNGFR] - k_{dNGFR} [boundNGFR]$	$\frac{d[NGFR]}{dt} = -k_{NGFR} [NGFR] + k_{NGFR} [boundNGFR]$	$\frac{d[NGFR]}{dt} = -k_{NGFR} [NGFR] + k_{NGFR} [boundNGFR]$
$\frac{d[RSK]}{dt} = -k_{RSK} [RSK] + k_{RSK} [boundEGFR]$	$\frac{d[boundEGFR]}{dt} = k_{RSK} [RSK] [boundEGFR] - k_{dEGFR} [boundEGFR]$	$\frac{d[EGFR]}{dt} = -k_{EGFR} [EGFR] + k_{EGFR} [boundEGFR]$	$\frac{d[EGFR]}{dt} = -k_{EGFR} [EGFR] + k_{EGFR} [boundEGFR]$
$\frac{d[Akt/PKB]}{dt} = -k_{Akt/PKB} [Akt/PKB] + k_{Akt/PKB} [boundEGFR]$	$\frac{d[boundEGFR]}{dt} = k_{Akt/PKB} [Akt/PKB] [boundEGFR] - k_{dEGFR} [boundEGFR]$	$\frac{d[EGFR]}{dt} = -k_{EGFR} [EGFR] + k_{EGFR} [boundEGFR]$	$\frac{d[EGFR]}{dt} = -k_{EGFR} [EGFR] + k_{EGFR} [boundEGFR]$
$\frac{d[p90/Rsk]}{dt} = -k_{p90/Rsk} [p90/Rsk] + k_{p90/Rsk} [boundEGFR]$	$\frac{d[boundEGFR]}{dt} = k_{p90/Rsk} [p90/Rsk] [boundEGFR] - k_{dEGFR} [boundEGFR]$	$\frac{d[EGFR]}{dt} = -k_{EGFR} [EGFR] + k_{EGFR} [boundEGFR]$	$\frac{d[EGFR]}{dt} = -k_{EGFR} [EGFR] + k_{EGFR} [boundEGFR]$

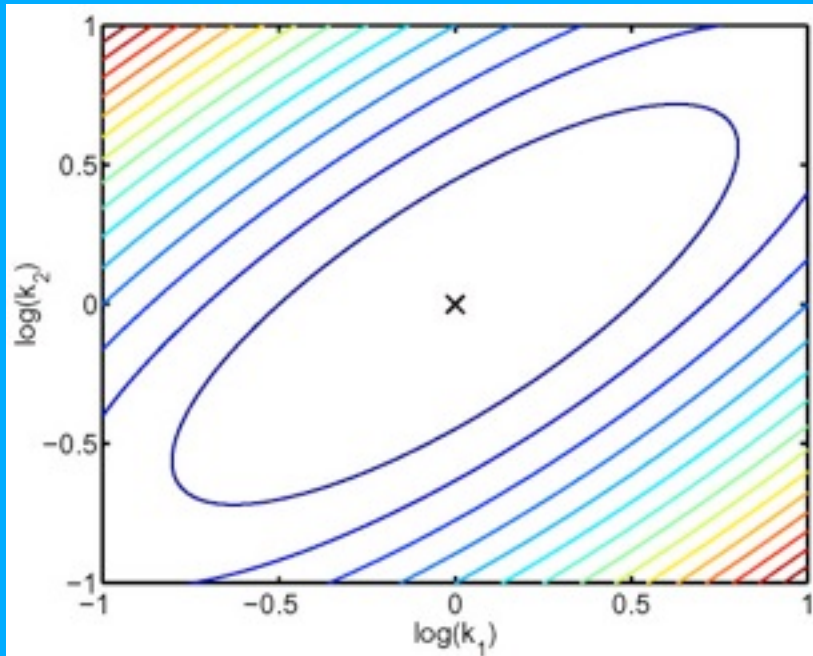
$$\frac{d[SosActive]}{dt} = +k_{EGF} [boundEGFR] \frac{[SosInactive]}{[SosInactive] + K_{mEGF}} + k_{NGF} [boundNGFR] \frac{[SosInactive]}{[SosInactive] + K_{mNGF}} - k_{dSos} [P90RskActive] \frac{[SosActive]}{[SosActive] + K_{mdSos}}$$

48 Parameter Fit!

Ensemble of Models

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*.

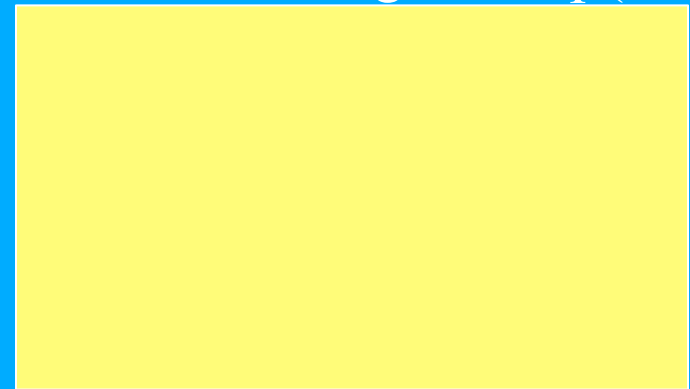
Don't trust predictions that vary



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 $\exp(-C/T)$



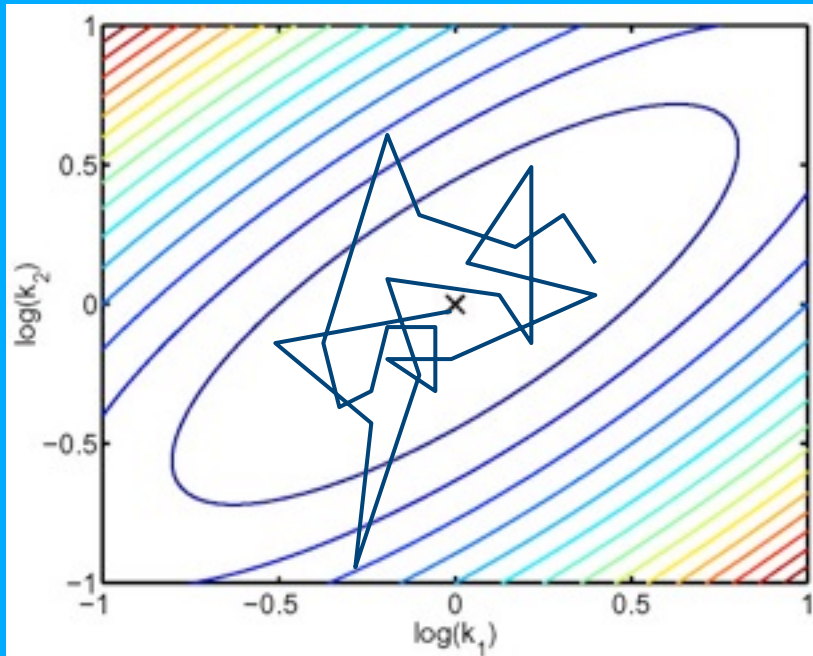
O is chemical concentration $y(t_i)$, or rate constant $\theta_n \dots$

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

Ensemble of Models

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



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 $\exp(-C/T)$

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

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

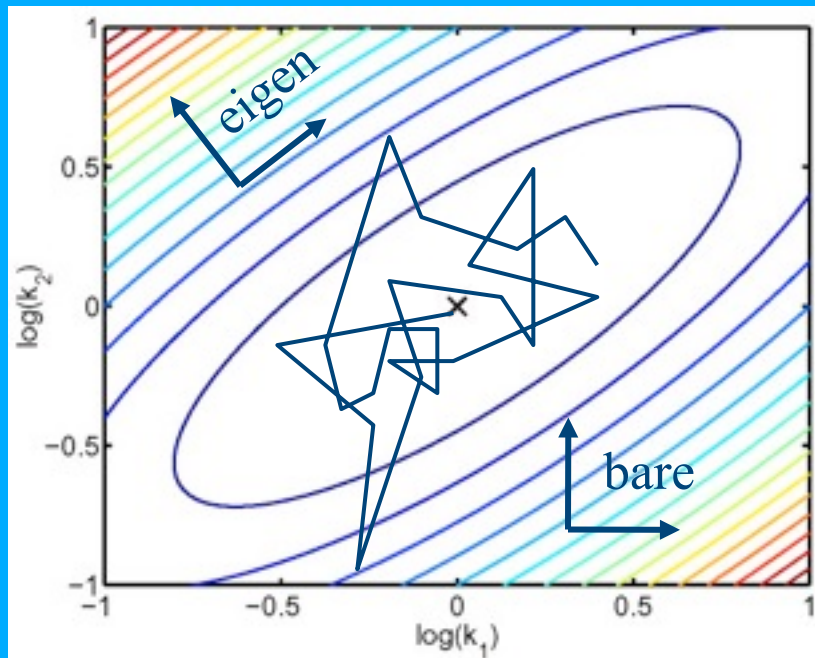
O is chemical concentration $y(t_i)$, or rate constant $\theta_n \dots$

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

Ensemble of Models

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



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 $\exp(-C/T)$

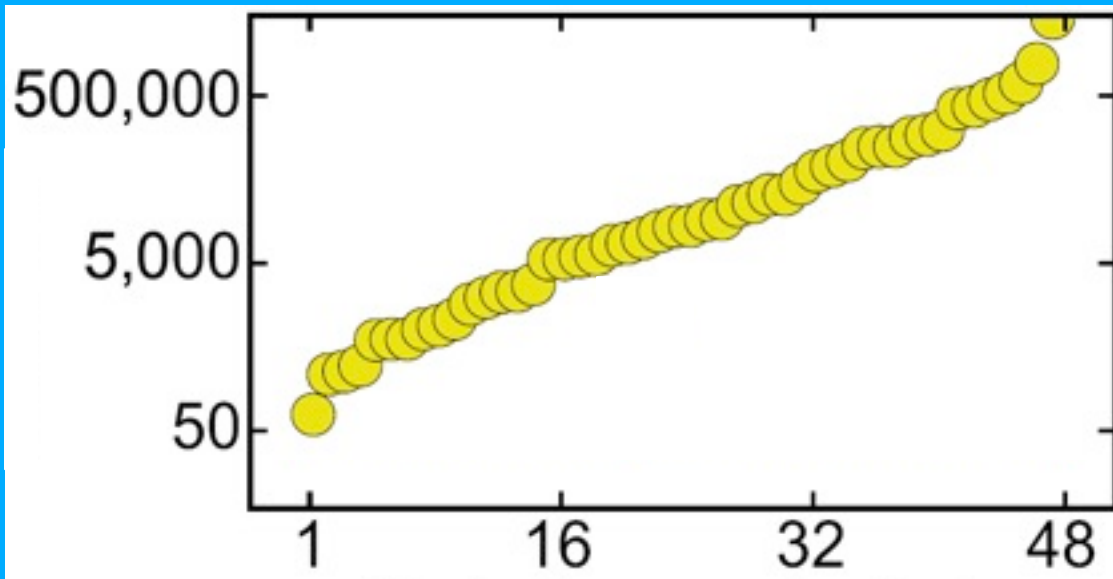
$$\langle O \rangle = \frac{1}{N_E} \sum_{i=1}^{N_E} O(\vec{\theta}_i)$$

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

O is chemical concentration $y(t_i)$, or rate constant $\theta_n \dots$

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

Relative parameter fluctuation



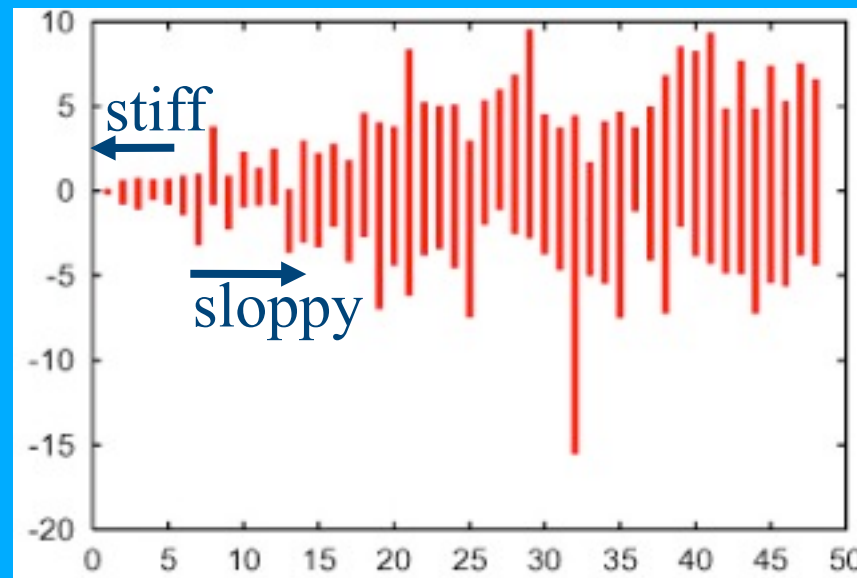
Parameter (sorted)

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

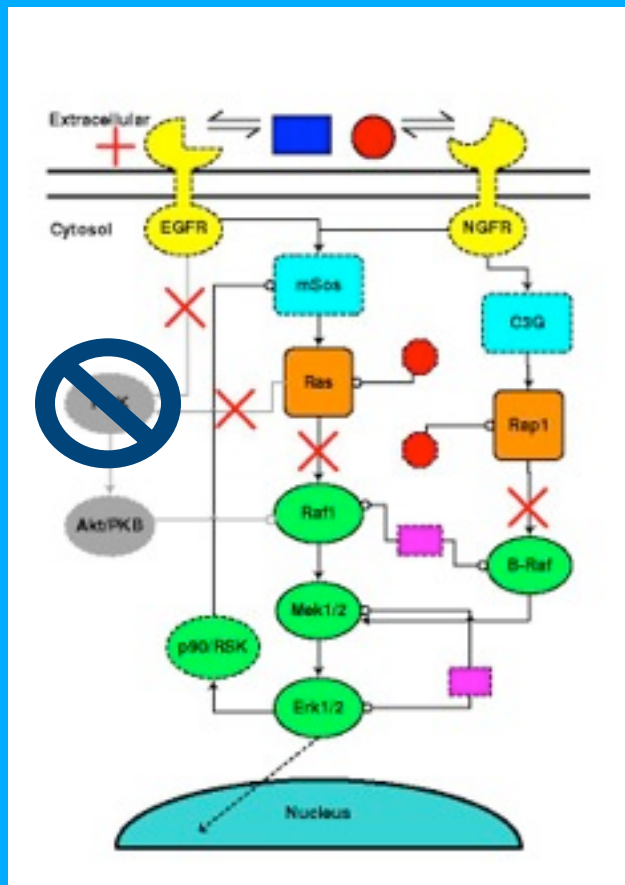
Are predictions possible?

\log_e eigenparameter fluctuation



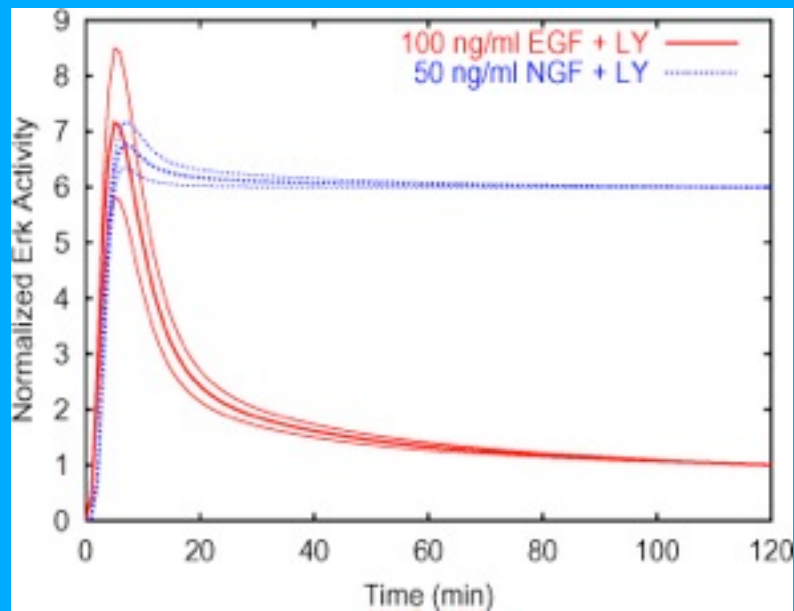
sorted eigenparameter number

Predictions are Possible

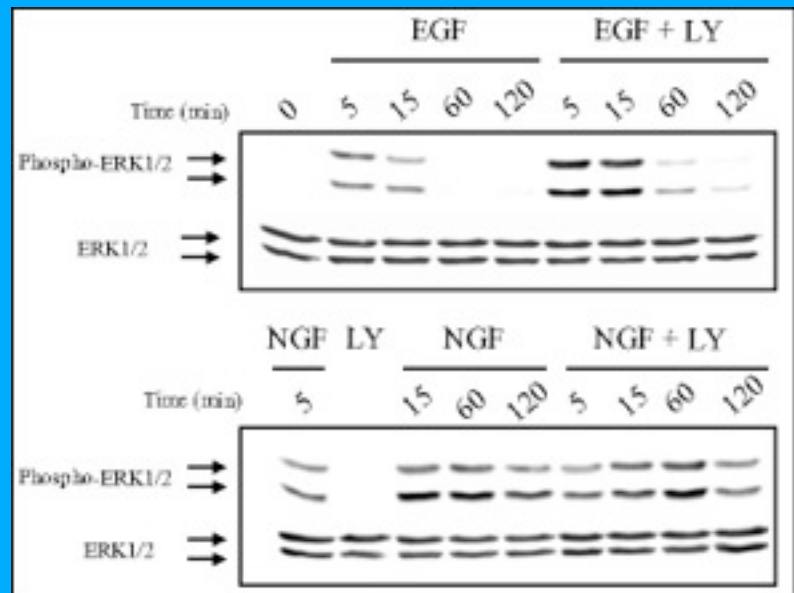


Model predicts that the left branch isn't important

Model Prediction



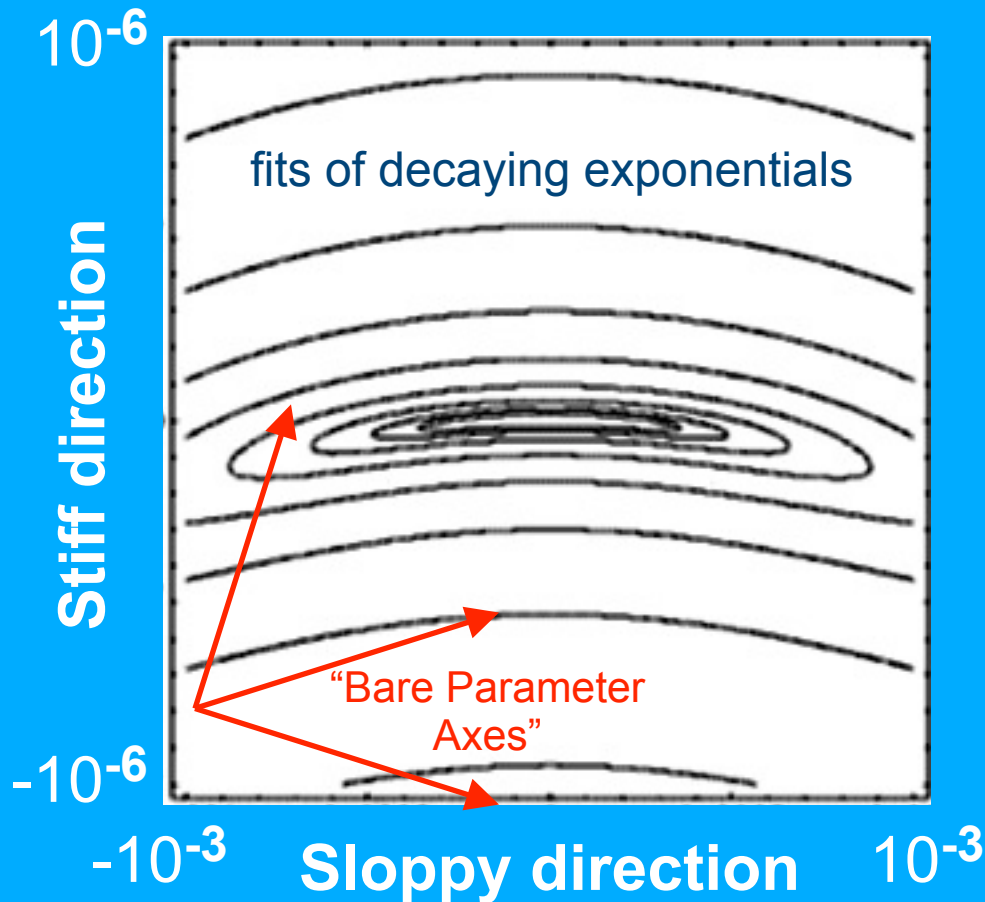
Brown's Experiment



Parameters fluctuate orders of magnitude, but still predictive!

Parameter Indeterminacy and Sloppiness

Cost Contours



~5 stiff, ~43 sloppy directions

Note: Horizontal scale
shrunk by 1000 times
Aspect ratio = Human hair

**48 parameter fits are
sloppy: Many parameter
sets give almost equally
good fits**

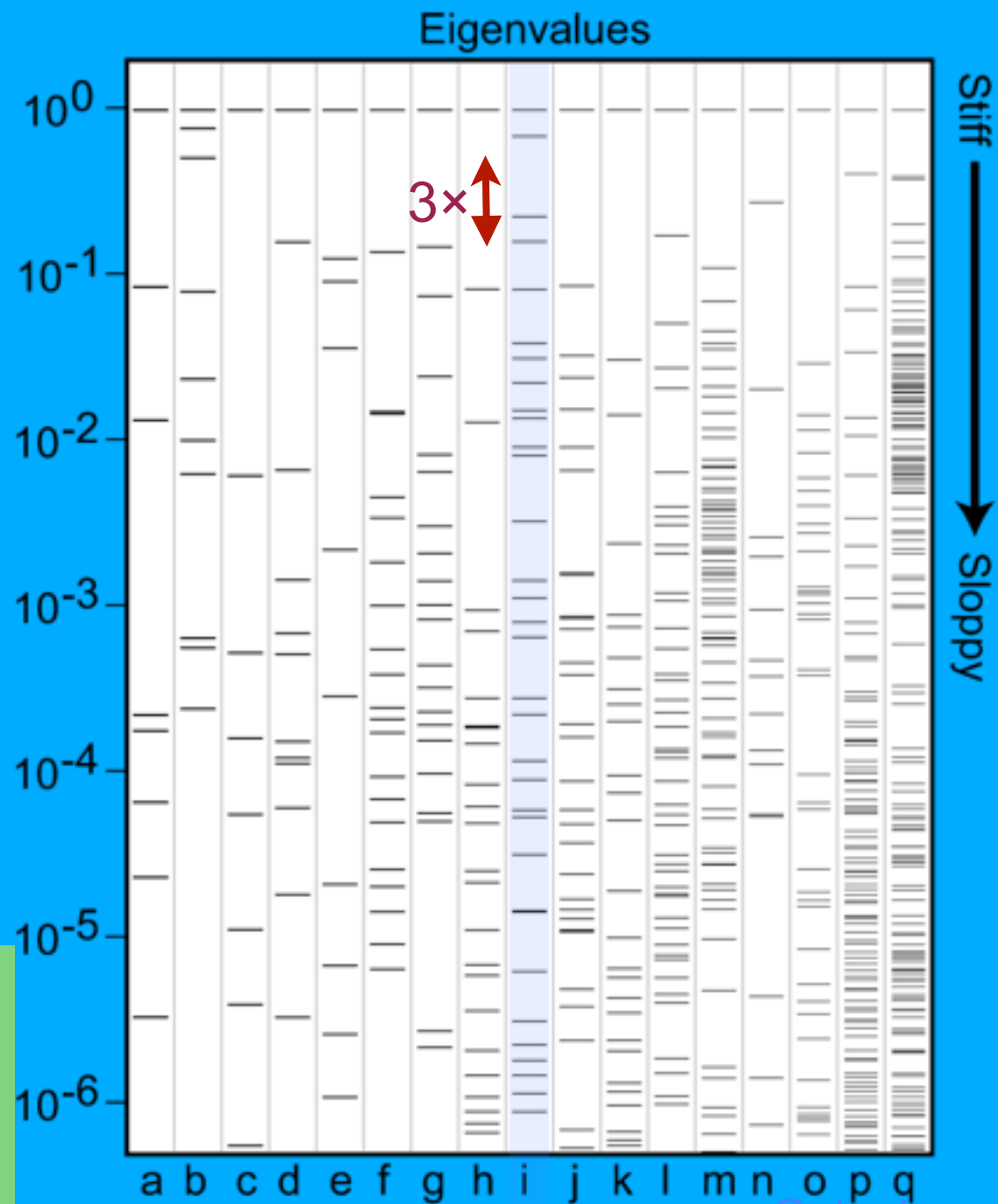
**A few 'stiff' constrained
directions allow model to
remain predictive**

Systems Biology

Seventeen models

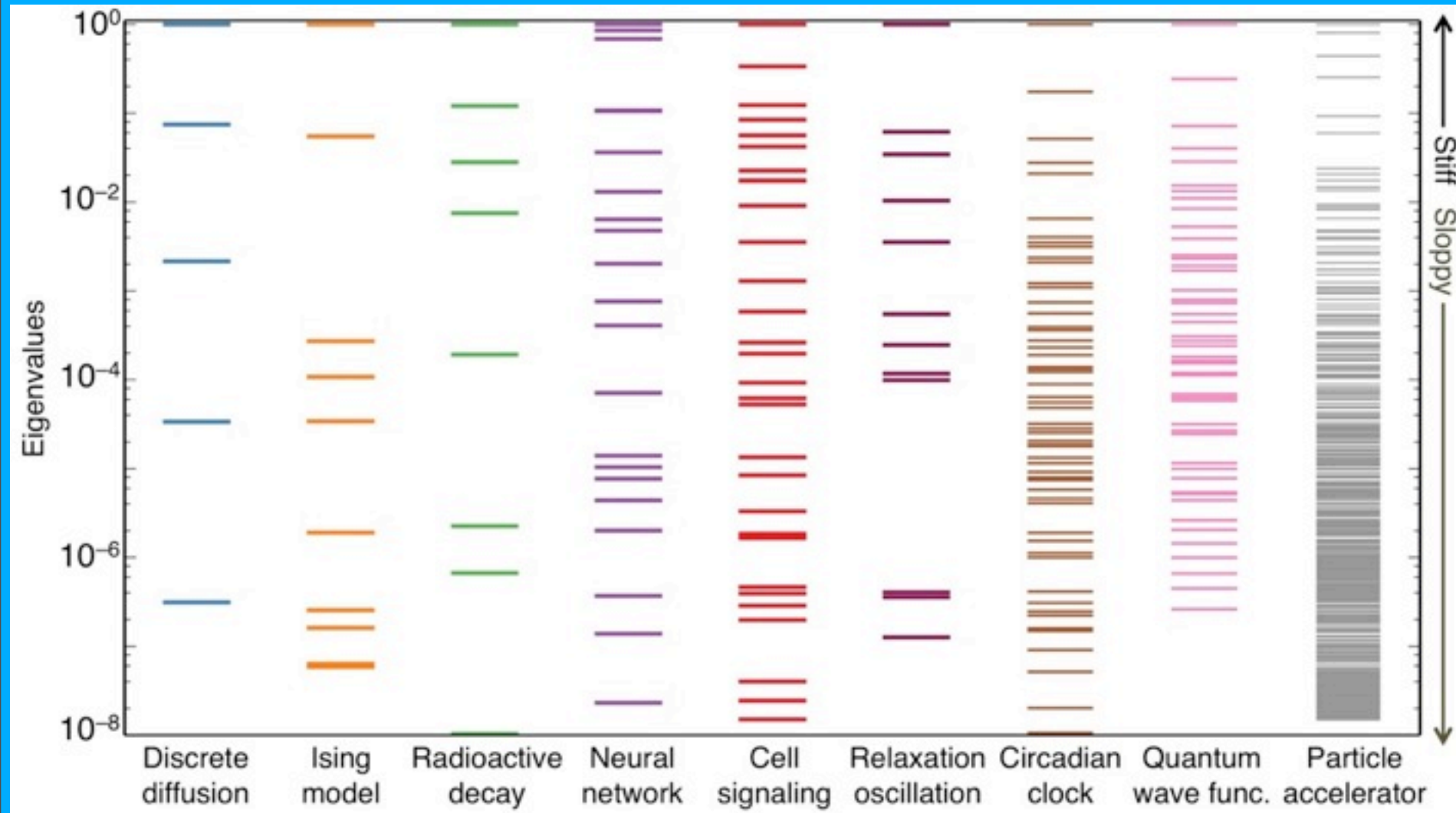
- (a) eukaryotic cell cycle
- (b) Xenopus egg cell cycle
- (c) eukaryotic mitosis
- (d) generic circadian rhythm
- (e) nicotinic acetylcholine intra-receptor dynamics
- (f) generic kinase cascade
- (g) Xenopus Wnt signaling
- (h) Drosophila circadian rhythm
- (i) rat growth-factor signaling
- (j) Drosophila 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 $\sim \sqrt{\lambda}$



Gutenkunst

Sloppy Universality Outside Bio



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

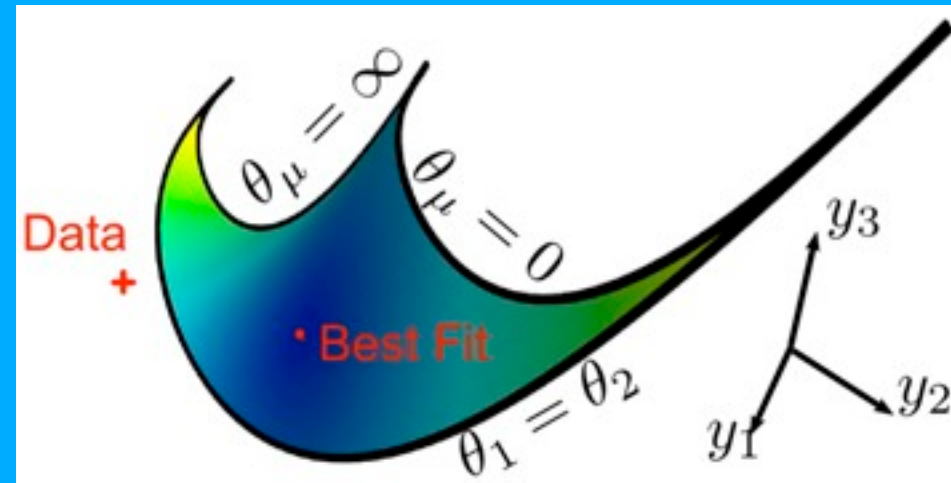
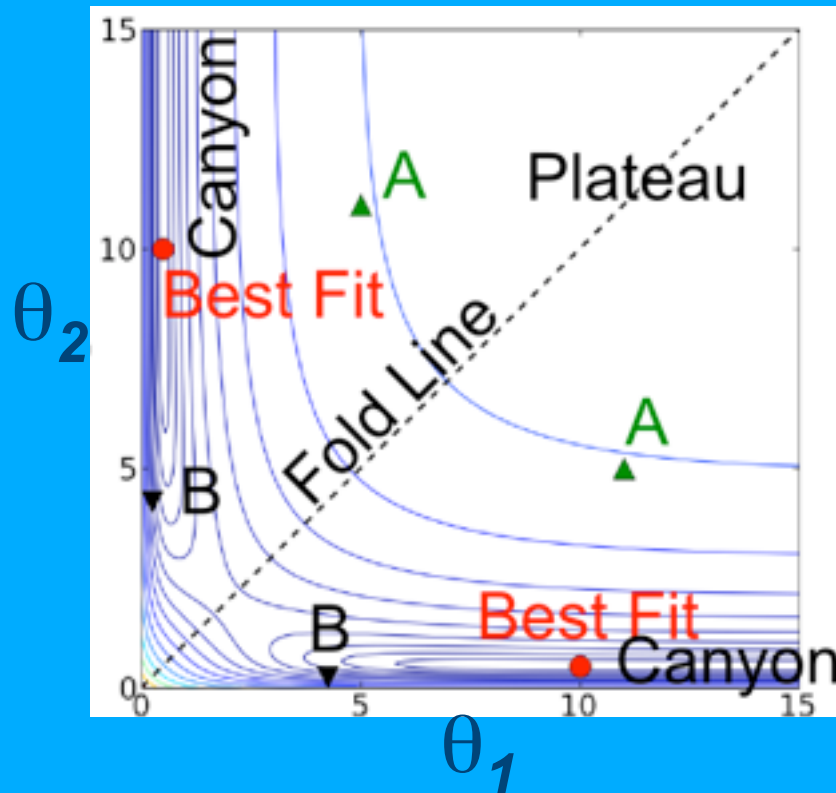
The Model Manifold

Two exponentials θ_1, θ_2
 fit to three data points y_1, y_2, y_3

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

Parameter space

Stiff and sloppy directions
 Canyons, Plateaus

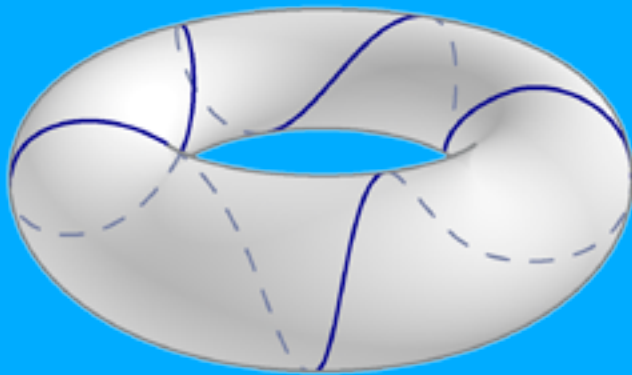


Data space

Manifold of model predictions
 Parameters as coordinates
 Model boundaries $\theta_n = \pm\infty, \theta_m$
 cause Plateaus
 Metric $g_{\mu\nu}$ from distance to data

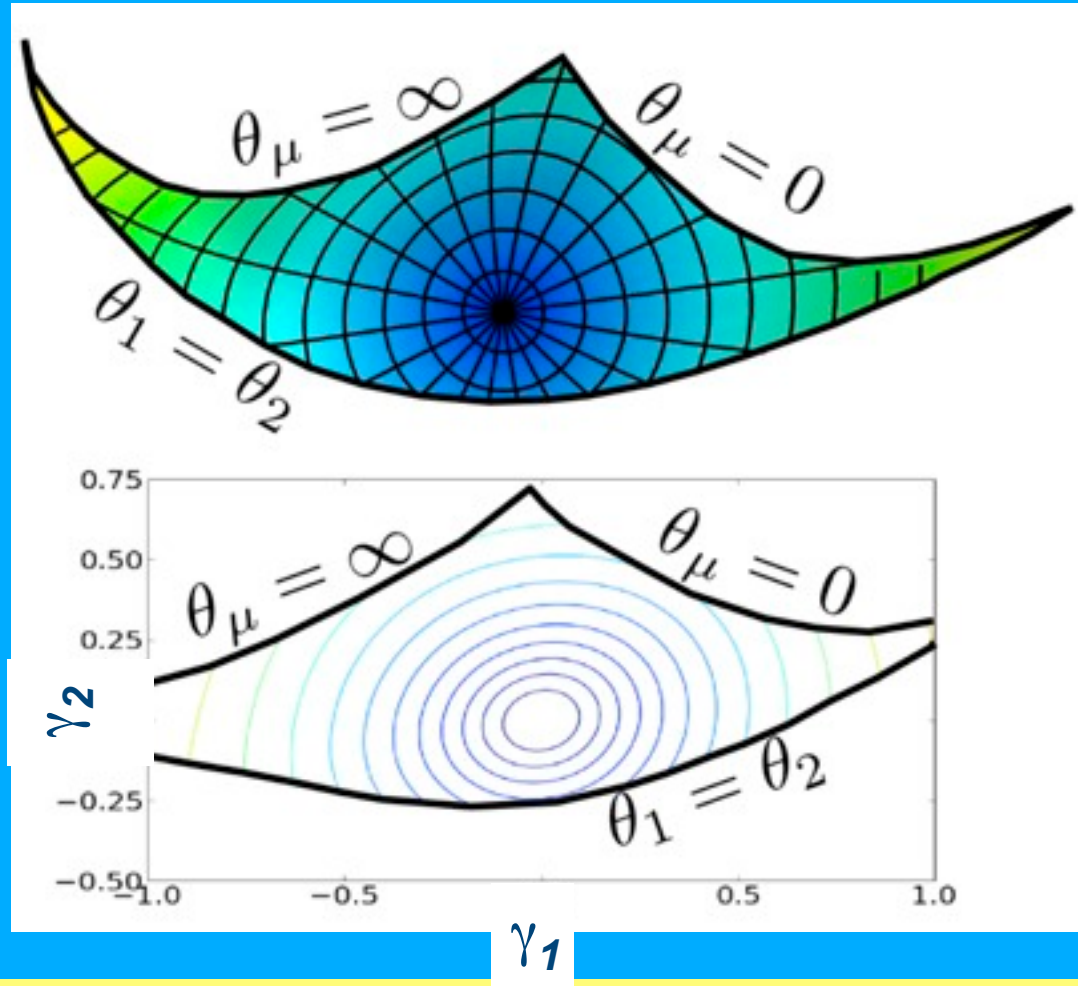
Geodesics

“Straight line” in curved space
Shortest path between points



Easy to find cost minimum using polar geodesic coordinates

γ_1, γ_2

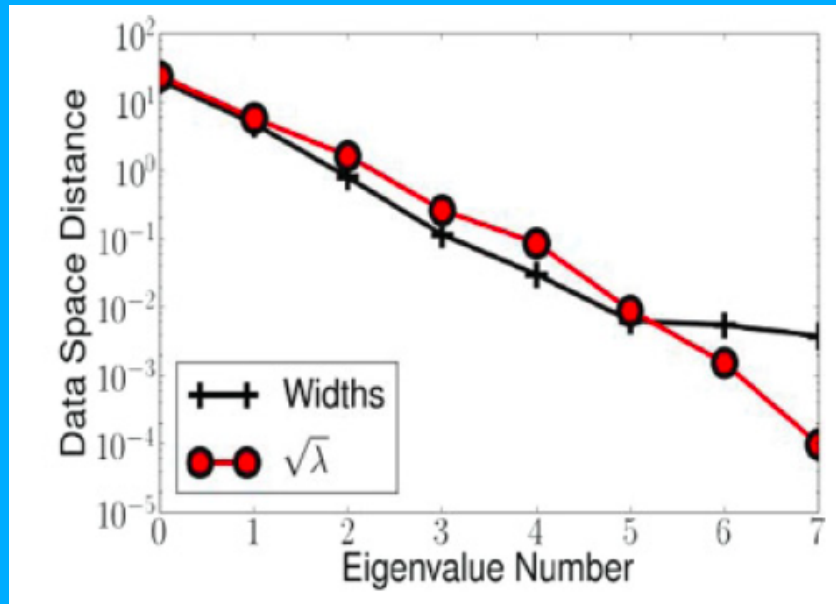
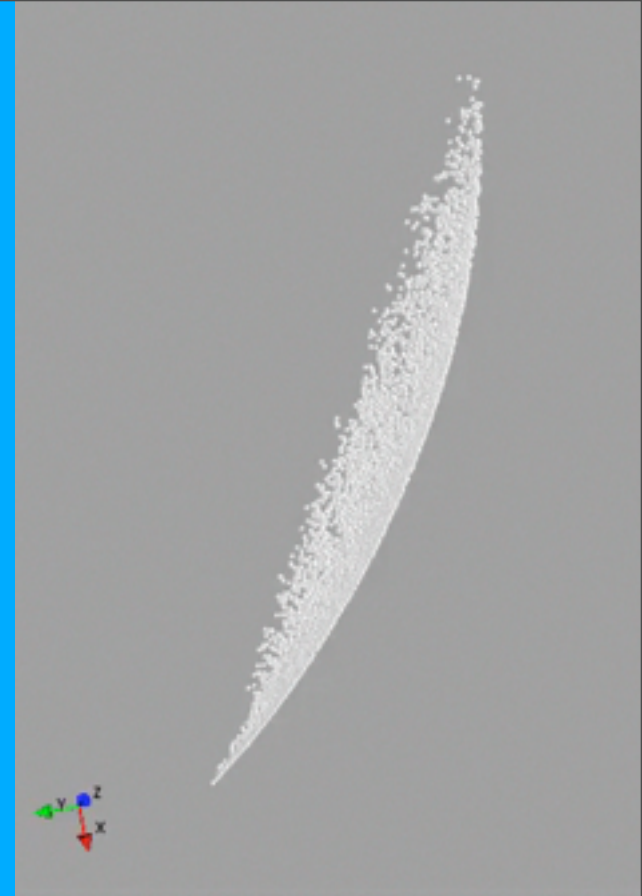


Cost contours in geodesic coordinates
nearly concentric circles!
Use this for algorithms...

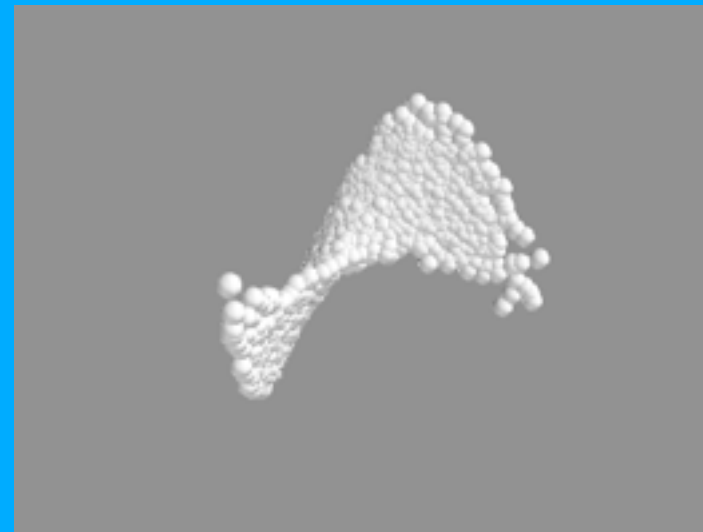
The Model Manifold is a *Hyper-Ribbon*

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

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



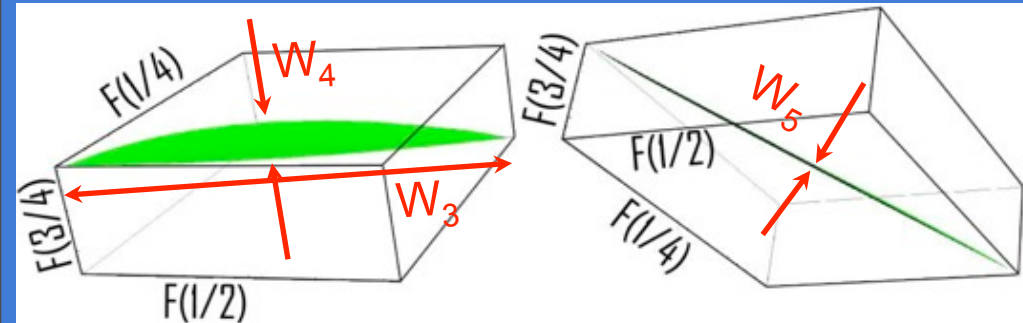
Widths along geodesics track eigenvalues almost perfectly!



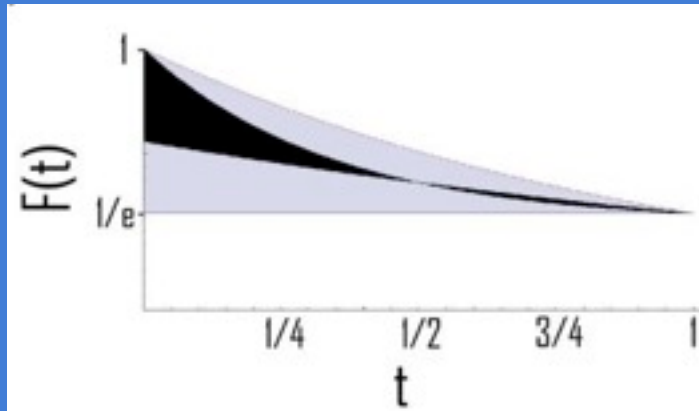
Diffusion equation after three time steps

Hierarchy of widths and curvatures

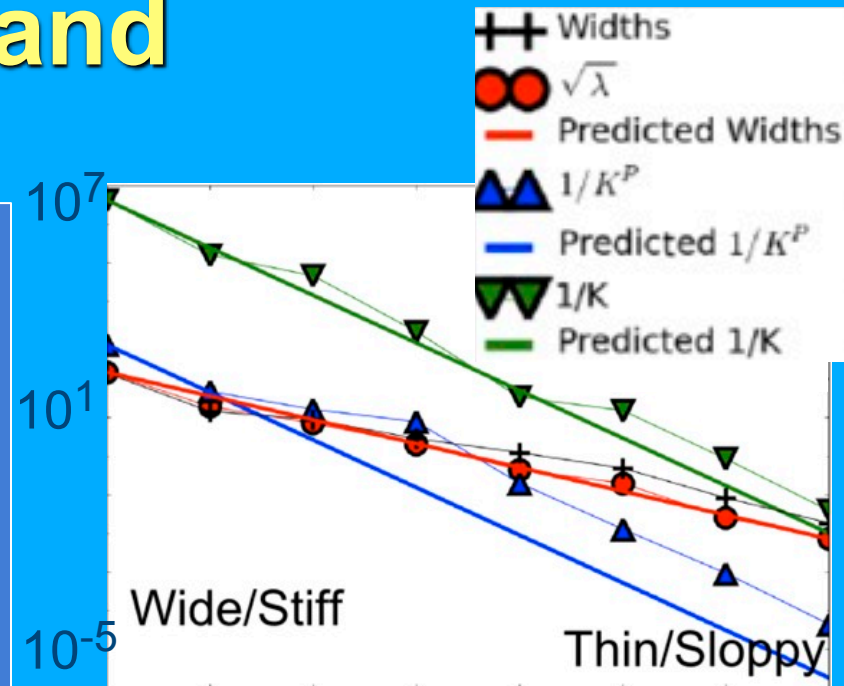
Hierarchy of widths



Cross sections: fixing f at 0, $\frac{1}{2}$, 1



Theorem: interpolation good with many data points
Geometrical convergence



Eigendirection at best fit

Multi-decade span of widths, curvatures, eigenvalues

Widths $\sim \sqrt{\lambda}$ sloppy eigs

Parameter curvature

$$K^P = 10^3 \times K$$

\gg extrinsic curvature

Why is it so thin and flat?

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

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

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

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

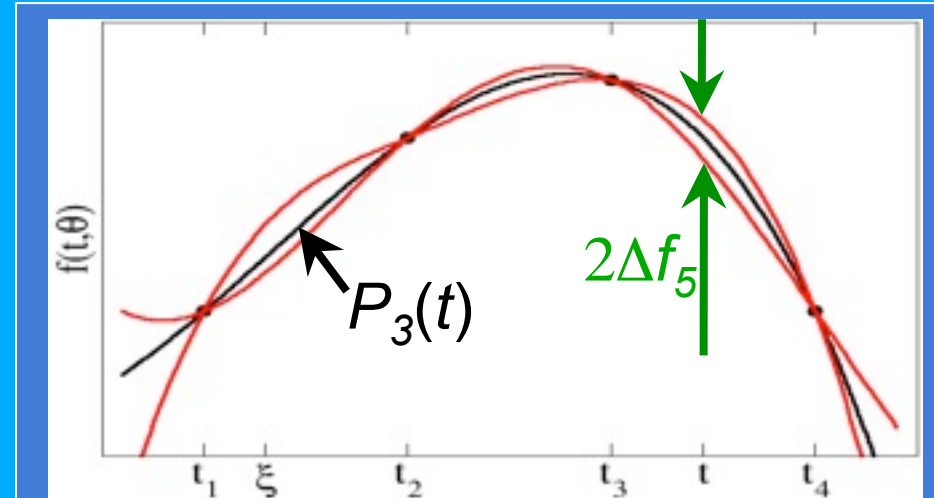
Interpolation convergence theorem

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

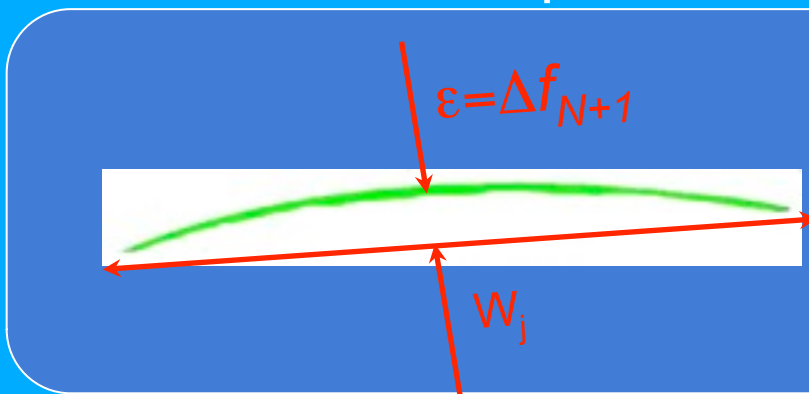
$$< (t-t_1)(t-t_2)\dots(t-t_m) f^{(m)}(\xi)/m!$$

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

More than one data per R



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

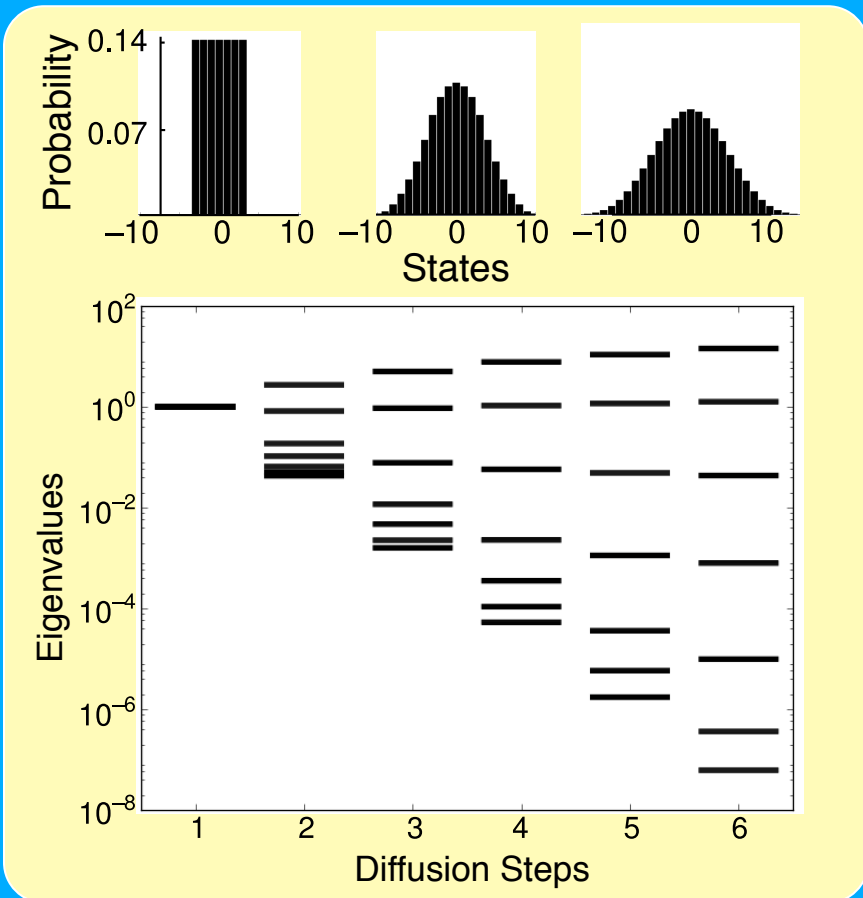
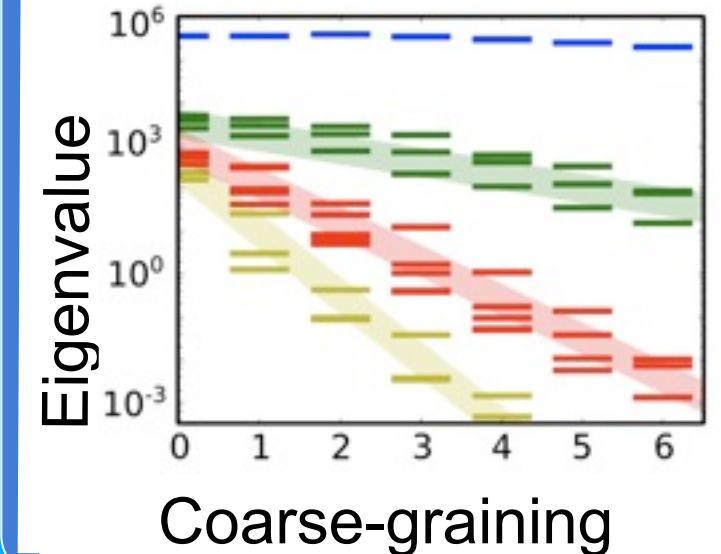
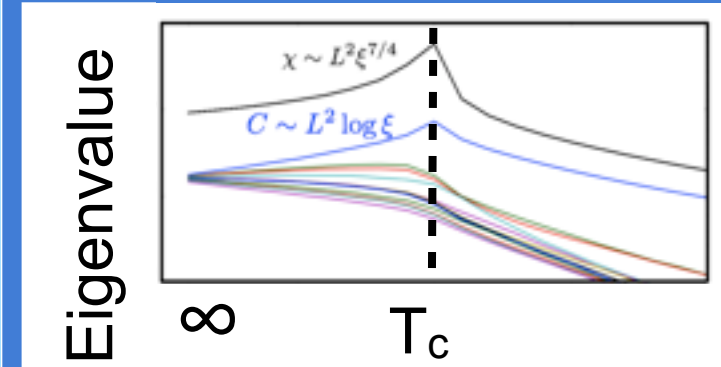
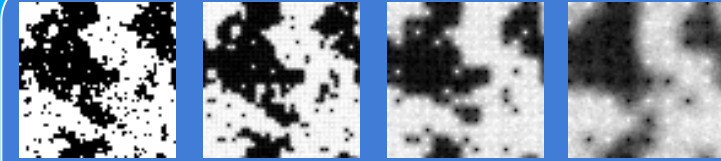


Extrinsic flatness: $N=M$ trivially flat, extra data deviates $\epsilon \sim \Delta f_{N+1}$, so curvature $K \sim \epsilon / W_j^2 \sim (\Delta t / R)^{N+1-j} / W_j$

Physics: Sloppiness and Emergence

Ben Machta, Ricky Chachra

Kullback-Liebler divergence metric
Sloppy after coarse graining (in
space for Ising, time for diffusion)



Generation of Reduced Models

Mark Transtrum (not me)

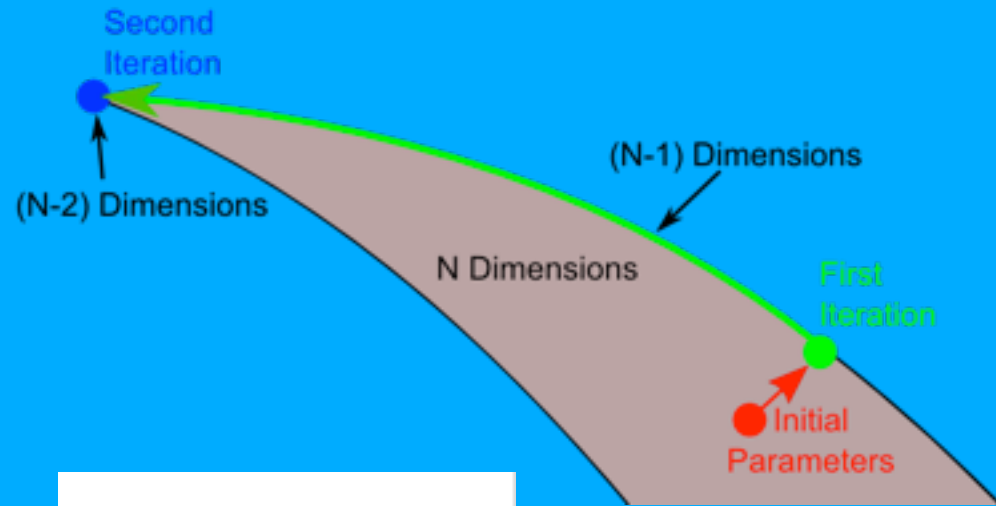
Can we coarse-grain sloppy models? If most parameter directions are useless, why not remove some?

Transtrum has *systematic* method!

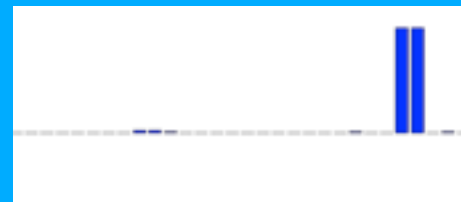
(1) Geodesic along sloppiest direction to nearby point on manifold boundary

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

Coarse-graining = boundaries of model manifold.



Sloppiest Eigendirection

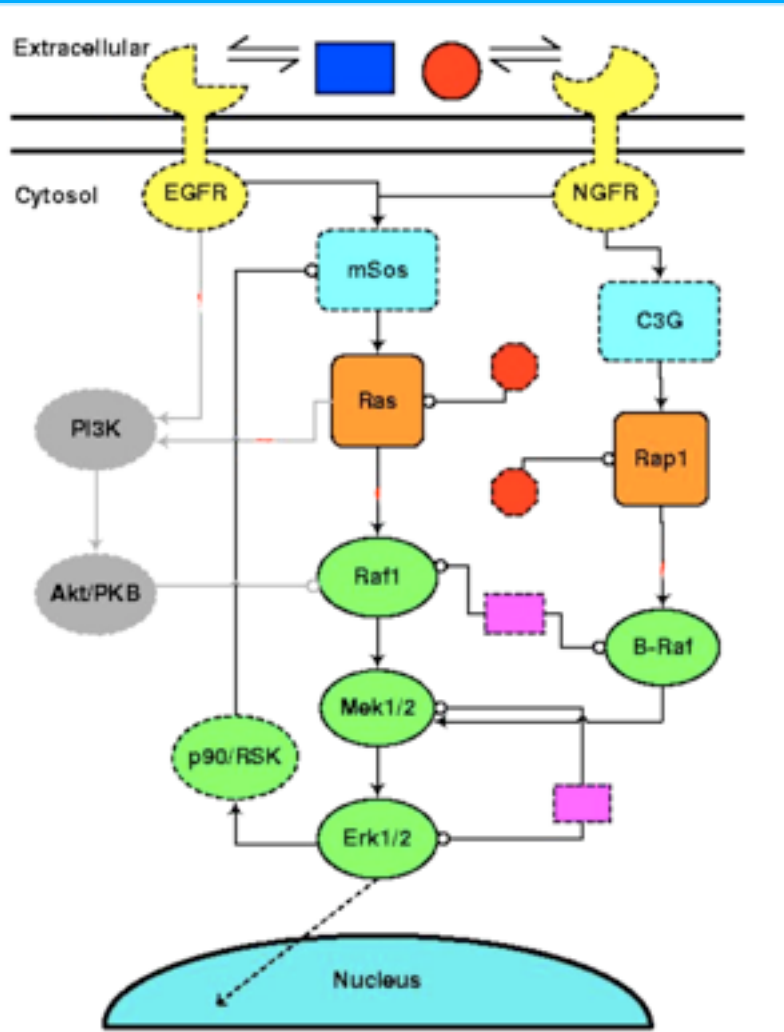


Simplified at Boundary (Unsaturation reaction)



Generation of Reduced Models

Mark Transtrum (not me)



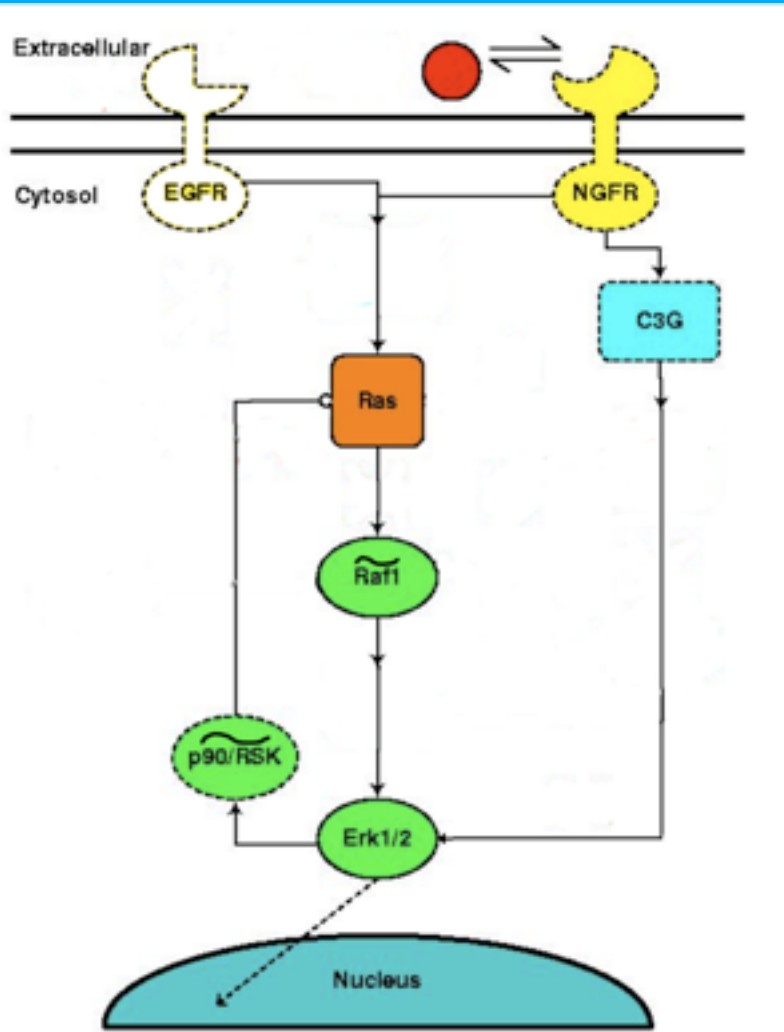
48 params
29 ODEs



ODE	Equation	ODE	Equation	ODE	Equation
1	$\frac{dE}{dt} = -k_{off} E + k_{on} E_{bound}$	17	$\frac{dRaf1}{dt} = k_{Raf1} - k_{Raf1} Raf1$	33	$\frac{dErk1/2}{dt} = k_{Erk1/2} - k_{Erk1/2} Erk1/2$
2	$\frac{dE_{bound}}{dt} = k_{on} E - k_{off} E_{bound}$	18	$\frac{dRaf1_{act}}{dt} = k_{Raf1_{act}} - k_{Raf1_{act}} Raf1_{act}$	34	$\frac{dErk1/2_{act}}{dt} = k_{Erk1/2_{act}} - k_{Erk1/2_{act}} Erk1/2_{act}$
3	$\frac{dEGFR}{dt} = k_{EGFR} - k_{EGFR} EGFR$	19	$\frac{dRaf1_{inact}}{dt} = k_{Raf1_{inact}} - k_{Raf1_{inact}} Raf1_{inact}$	35	$\frac{dErk1/2_{inact}}{dt} = k_{Erk1/2_{inact}} - k_{Erk1/2_{inact}} Erk1/2_{inact}$
4	$\frac{dEGFR_{act}}{dt} = k_{EGFR_{act}} - k_{EGFR_{act}} EGFR_{act}$	20	$\frac{dRaf1_{inact}}{dt} = k_{Raf1_{inact}} - k_{Raf1_{inact}} Raf1_{inact}$	36	$\frac{dErk1/2_{inact}}{dt} = k_{Erk1/2_{inact}} - k_{Erk1/2_{inact}} Erk1/2_{inact}$
5	$\frac{dEGFR_{bound}}{dt} = k_{EGFR_{bound}} - k_{EGFR_{bound}} EGFR_{bound}$	21	$\frac{dRaf1_{inact}}{dt} = k_{Raf1_{inact}} - k_{Raf1_{inact}} Raf1_{inact}$	37	$\frac{dErk1/2_{inact}}{dt} = k_{Erk1/2_{inact}} - k_{Erk1/2_{inact}} Erk1/2_{inact}$
6	$\frac{dEGFR_{bound}}{dt} = k_{EGFR_{bound}} - k_{EGFR_{bound}} EGFR_{bound}$	22	$\frac{dRaf1_{inact}}{dt} = k_{Raf1_{inact}} - k_{Raf1_{inact}} Raf1_{inact}$	38	$\frac{dErk1/2_{inact}}{dt} = k_{Erk1/2_{inact}} - k_{Erk1/2_{inact}} Erk1/2_{inact}$
7	$\frac{dEGFR_{bound}}{dt} = k_{EGFR_{bound}} - k_{EGFR_{bound}} EGFR_{bound}$	23	$\frac{dRaf1_{inact}}{dt} = k_{Raf1_{inact}} - k_{Raf1_{inact}} Raf1_{inact}$	39	$\frac{dErk1/2_{inact}}{dt} = k_{Erk1/2_{inact}} - k_{Erk1/2_{inact}} Erk1/2_{inact}$
8	$\frac{dEGFR_{bound}}{dt} = k_{EGFR_{bound}} - k_{EGFR_{bound}} EGFR_{bound}$	24	$\frac{dRaf1_{inact}}{dt} = k_{Raf1_{inact}} - k_{Raf1_{inact}} Raf1_{inact}$	40	$\frac{dErk1/2_{inact}}{dt} = k_{Erk1/2_{inact}} - k_{Erk1/2_{inact}} Erk1/2_{inact}$
9	$\frac{dEGFR_{bound}}{dt} = k_{EGFR_{bound}} - k_{EGFR_{bound}} EGFR_{bound}$	25	$\frac{dRaf1_{inact}}{dt} = k_{Raf1_{inact}} - k_{Raf1_{inact}} Raf1_{inact}$	41	$\frac{dErk1/2_{inact}}{dt} = k_{Erk1/2_{inact}} - k_{Erk1/2_{inact}} Erk1/2_{inact}$
10	$\frac{dEGFR_{bound}}{dt} = k_{EGFR_{bound}} - k_{EGFR_{bound}} EGFR_{bound}$	26	$\frac{dRaf1_{inact}}{dt} = k_{Raf1_{inact}} - k_{Raf1_{inact}} Raf1_{inact}$	42	$\frac{dErk1/2_{inact}}{dt} = k_{Erk1/2_{inact}} - k_{Erk1/2_{inact}} Erk1/2_{inact}$
11	$\frac{dEGFR_{bound}}{dt} = k_{EGFR_{bound}} - k_{EGFR_{bound}} EGFR_{bound}$	27	$\frac{dRaf1_{inact}}{dt} = k_{Raf1_{inact}} - k_{Raf1_{inact}} Raf1_{inact}$	43	$\frac{dErk1/2_{inact}}{dt} = k_{Erk1/2_{inact}} - k_{Erk1/2_{inact}} Erk1/2_{inact}$
12	$\frac{dEGFR_{bound}}{dt} = k_{EGFR_{bound}} - k_{EGFR_{bound}} EGFR_{bound}$	28	$\frac{dRaf1_{inact}}{dt} = k_{Raf1_{inact}} - k_{Raf1_{inact}} Raf1_{inact}$	44	$\frac{dErk1/2_{inact}}{dt} = k_{Erk1/2_{inact}} - k_{Erk1/2_{inact}} Erk1/2_{inact}$
13	$\frac{dEGFR_{bound}}{dt} = k_{EGFR_{bound}} - k_{EGFR_{bound}} EGFR_{bound}$	29	$\frac{dRaf1_{inact}}{dt} = k_{Raf1_{inact}} - k_{Raf1_{inact}} Raf1_{inact}$	45	$\frac{dErk1/2_{inact}}{dt} = k_{Erk1/2_{inact}} - k_{Erk1/2_{inact}} Erk1/2_{inact}$
14	$\frac{dEGFR_{bound}}{dt} = k_{EGFR_{bound}} - k_{EGFR_{bound}} EGFR_{bound}$	30	$\frac{dRaf1_{inact}}{dt} = k_{Raf1_{inact}} - k_{Raf1_{inact}} Raf1_{inact}$	46	$\frac{dErk1/2_{inact}}{dt} = k_{Erk1/2_{inact}} - k_{Erk1/2_{inact}} Erk1/2_{inact}$
15	$\frac{dEGFR_{bound}}{dt} = k_{EGFR_{bound}} - k_{EGFR_{bound}} EGFR_{bound}$	31	$\frac{dRaf1_{inact}}{dt} = k_{Raf1_{inact}} - k_{Raf1_{inact}} Raf1_{inact}$	47	$\frac{dErk1/2_{inact}}{dt} = k_{Erk1/2_{inact}} - k_{Erk1/2_{inact}} Erk1/2_{inact}$
16	$\frac{dEGFR_{bound}}{dt} = k_{EGFR_{bound}} - k_{EGFR_{bound}} EGFR_{bound}$	32	$\frac{dRaf1_{inact}}{dt} = k_{Raf1_{inact}} - k_{Raf1_{inact}} Raf1_{inact}$	48	$\frac{dErk1/2_{inact}}{dt} = k_{Erk1/2_{inact}} - k_{Erk1/2_{inact}} Erk1/2_{inact}$

Generation of Reduced Models

Mark Transtrum (not me)



12 params
6 ODEs



$$[bEGFR] = \begin{cases} 1 & \text{EGF Present} \\ 0 & \text{Otherwise} \end{cases}$$

$$\frac{d}{dt}[bNGFR] = \theta_1[NGF][iNGFR]$$

$$\frac{d}{dt}[NGF] = -\theta_1[NGF][iNGFR]$$

$$\frac{d}{dt}[RasA] = -[RasA][P90RskA] + \theta_2[bEGFR] + \theta_3[bNGFR]$$

$$\frac{d}{dt}[\widetilde{Raf1A}] = \theta_4[RasA] - \theta_5[\widetilde{Raf1A}]/([\widetilde{Raf1A}] + \theta_6)$$

$$\frac{d}{dt}[C3GA] = \theta_7[bNGFR][C3GI]$$

$$[Rap1A] = \theta_8[C3GA]$$

$$[MekA] = [\widetilde{Raf1A}][MekI] + \theta_9[Rap1A]$$

$$\frac{d}{dt}[Erk] = -\theta_{10}[ErkA] + \theta_{11}[MekA][ErkI]$$

$$\frac{d}{dt}[P90RskA] = \theta_{12}[ErkA]$$

Reduced model fits all
experimental data

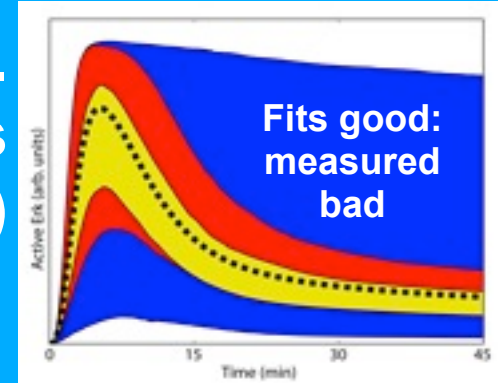
$$\theta_9 = \frac{[BRafI] kRap1toBRaf KmdBRAF kpBRaf KmdMek}{[PP2AA] [Raf1PPtase] kdBRaf KmRap1toBRaf kdMek}$$

Effective 'renormalized' params

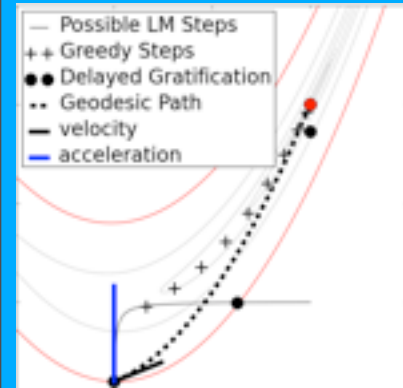
Sloppy Applications

Several applications emerge

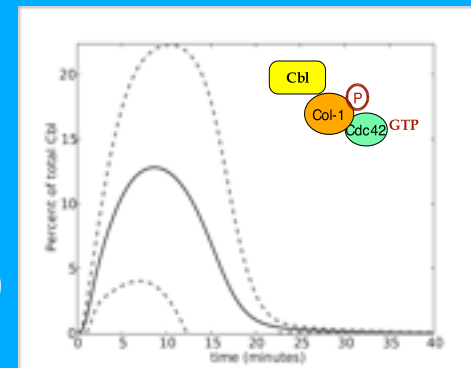
A. Fitting data vs. measuring parameters (Gutenkunst)



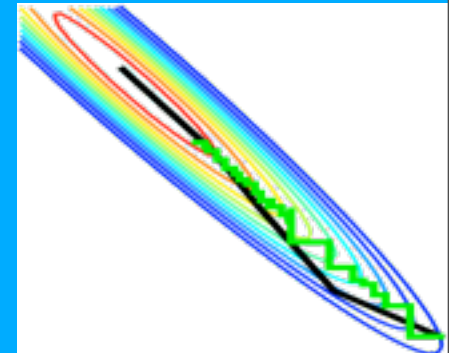
B. Finding best fits by geodesic acceleration (Transtrum)



C. Optimal experimental design (Casey)



D. Sloppy fitness and evolution (Gutenkunst)

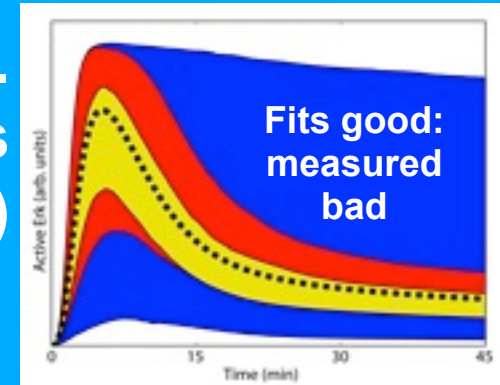


E. Estimating systematic errors: DFT and interatomic potentials (Jacobsen et al.)

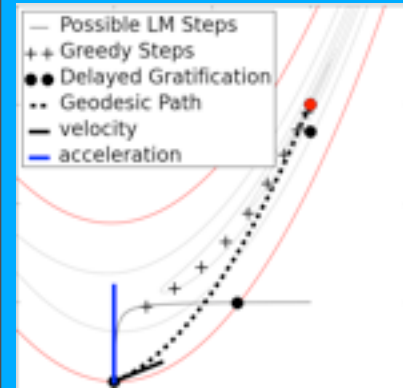
Sloppy Applications

Several applications emerge

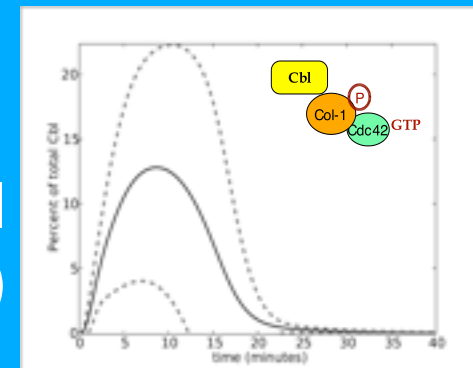
A. Fitting data vs. measuring parameters (Gutenkunst)



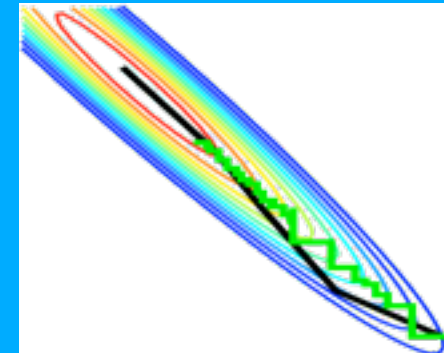
B. Finding best fits by geodesic acceleration (Transtrum)



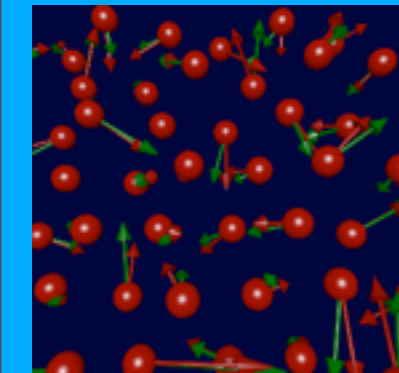
C. Optimal experimental design (Casey)



D. Sloppy fitness and evolution (Gutenkunst)

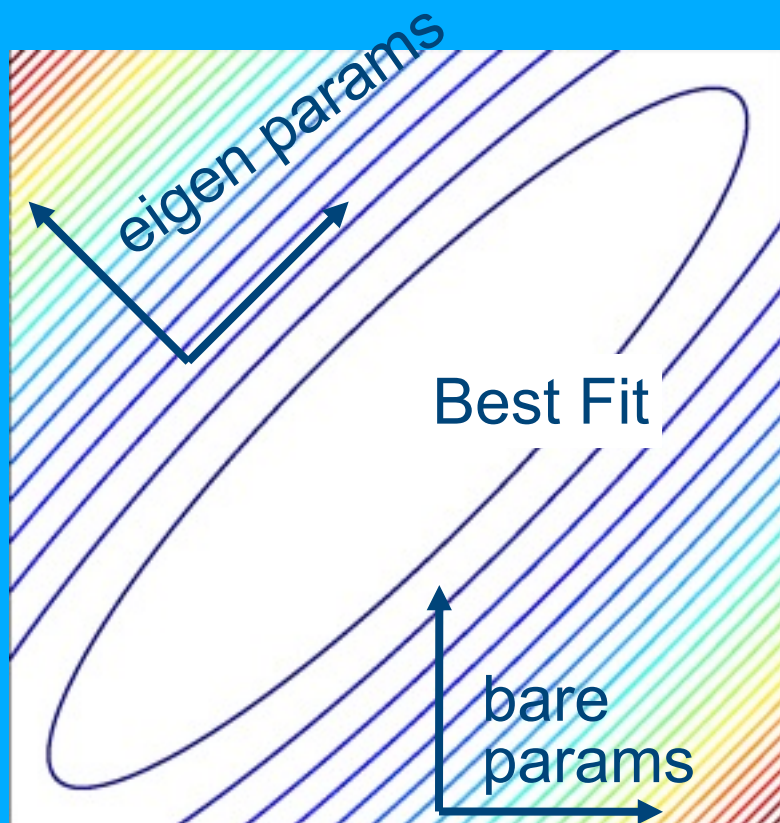


E. Estimating systematic errors: DFT and interatomic potentials (Jacobsen et al.)



A. Are rate constants useful?

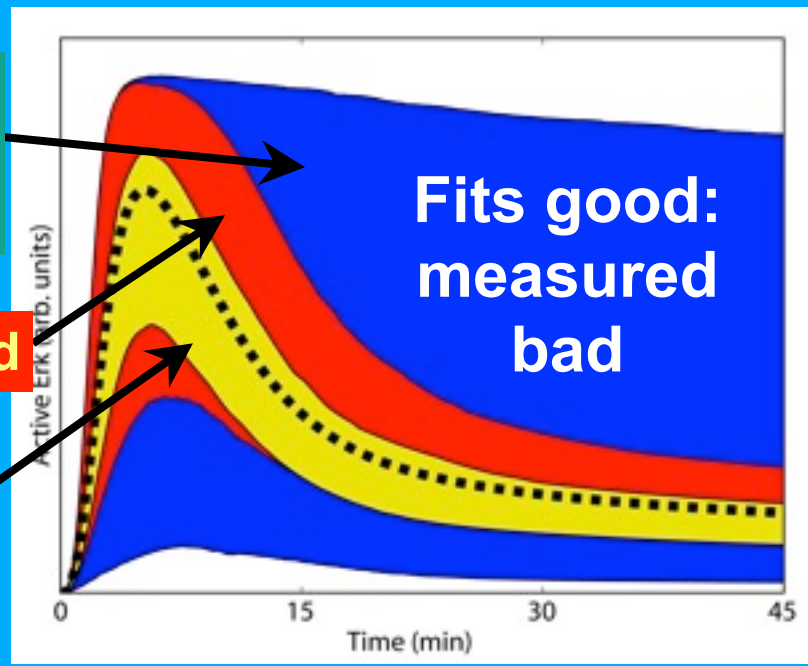
Fits vs. measurements



Missing one param

Measured

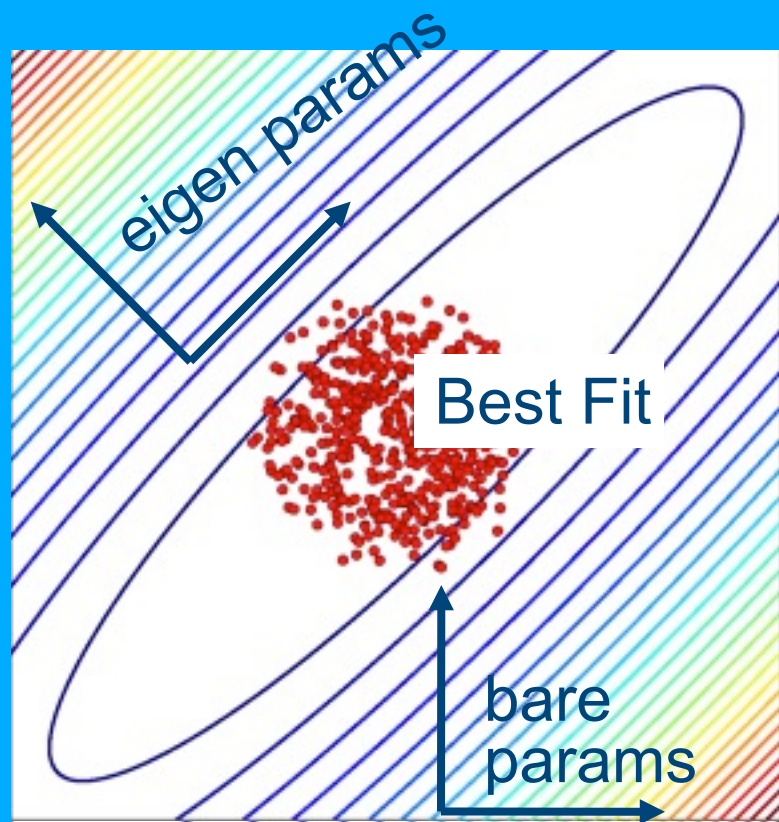
Fit



- 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?

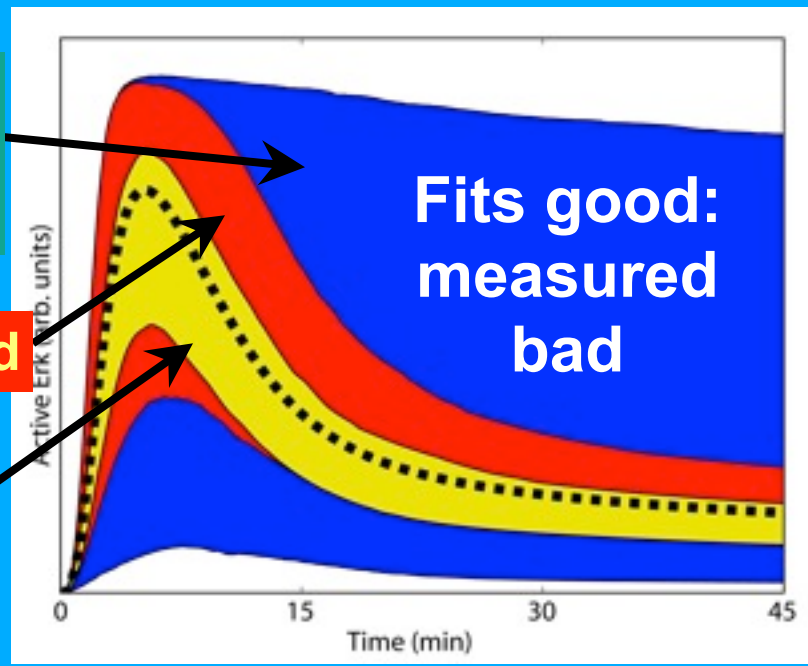
Fits vs. measurements



Missing one param

Measured

Fit

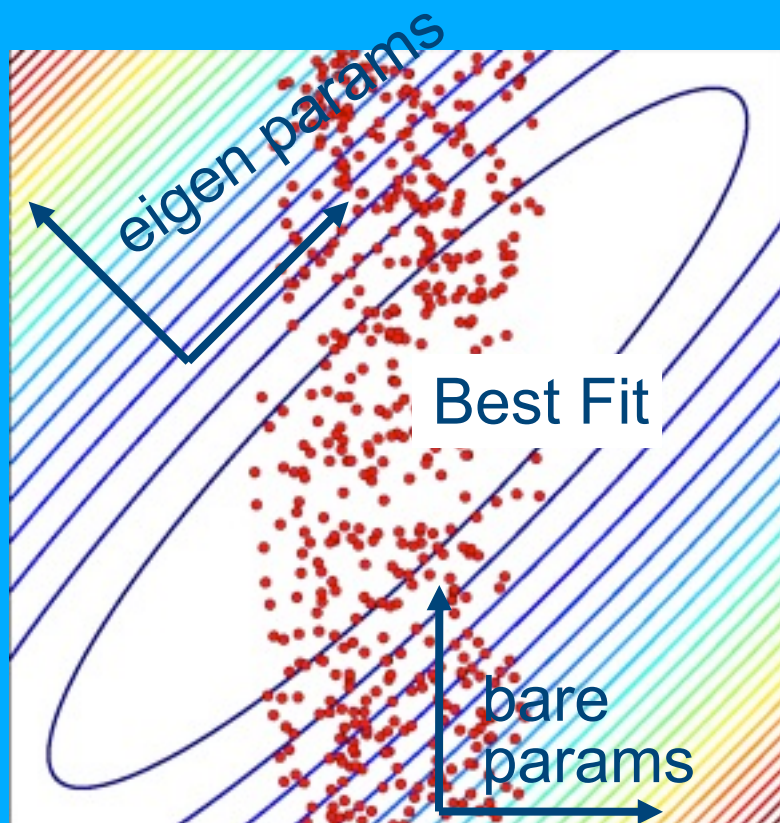


Monte Carlo (anharmonic)

- 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?

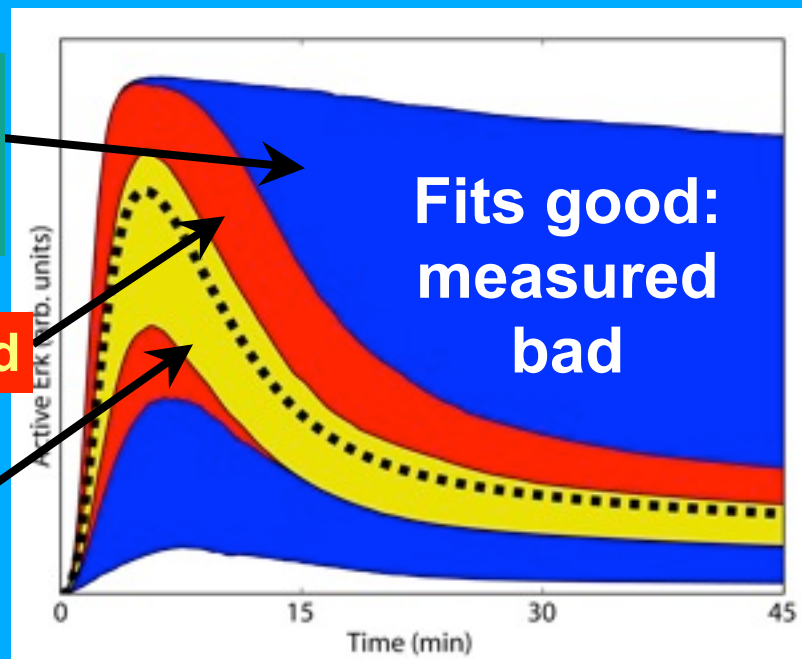
Fits vs. measurements



Missing one param

Measured

Fit

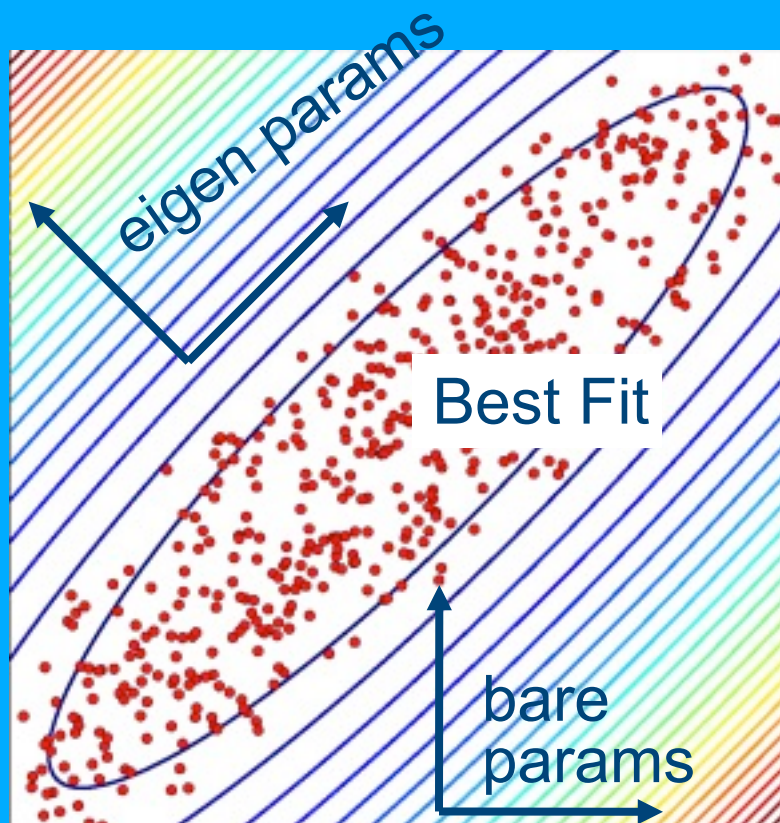


Monte Carlo (anharmonic)

- 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?

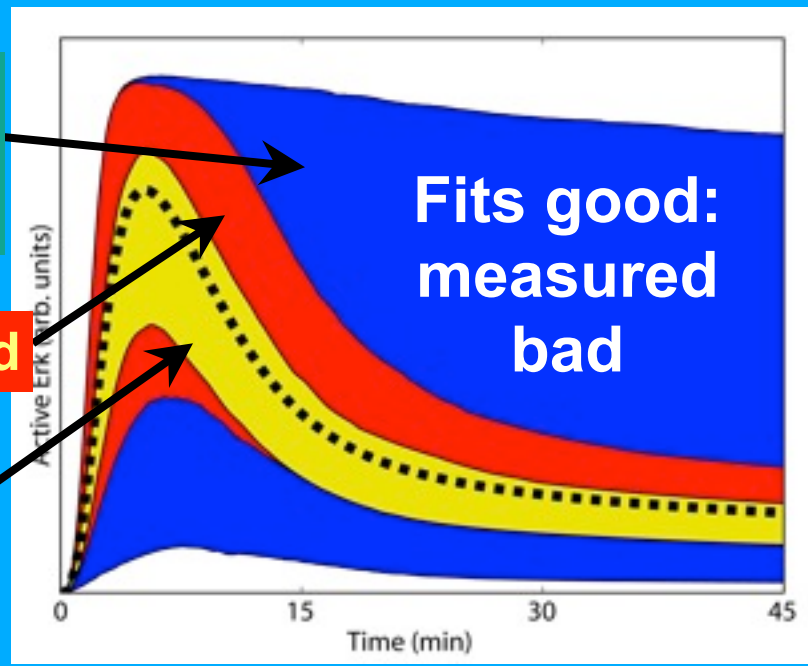
Fits vs. measurements



Missing one param

Measured

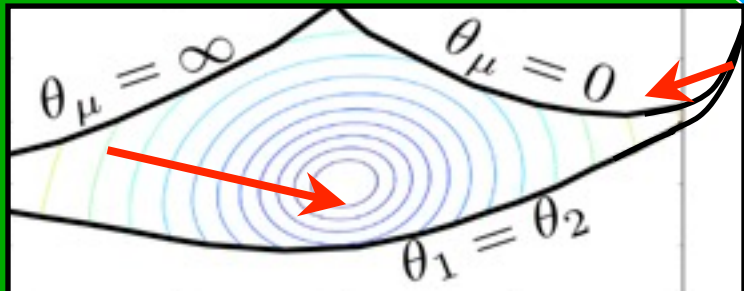
Fit



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

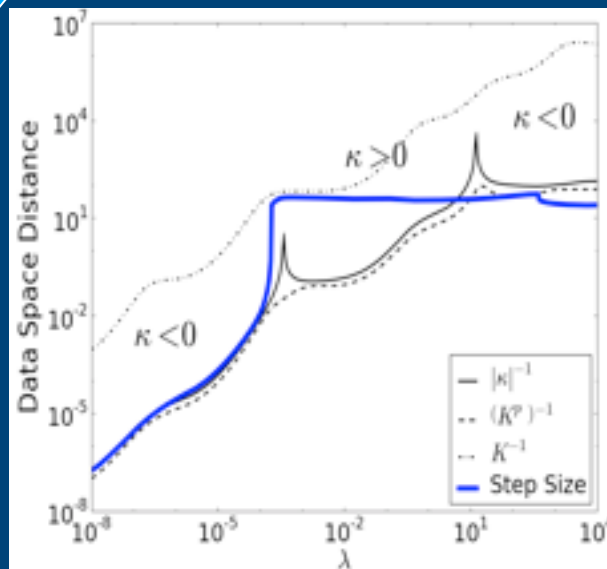


Geodesic Paths nearly circles
Follow local geodesic velocity?

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

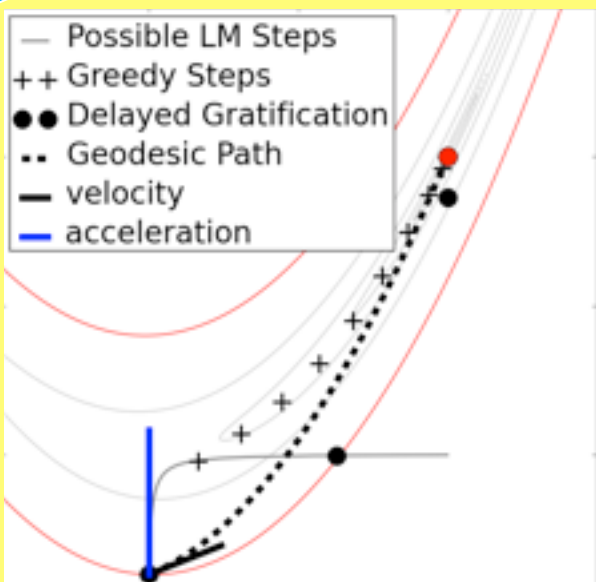
→ Gauss-Newton

→ 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)

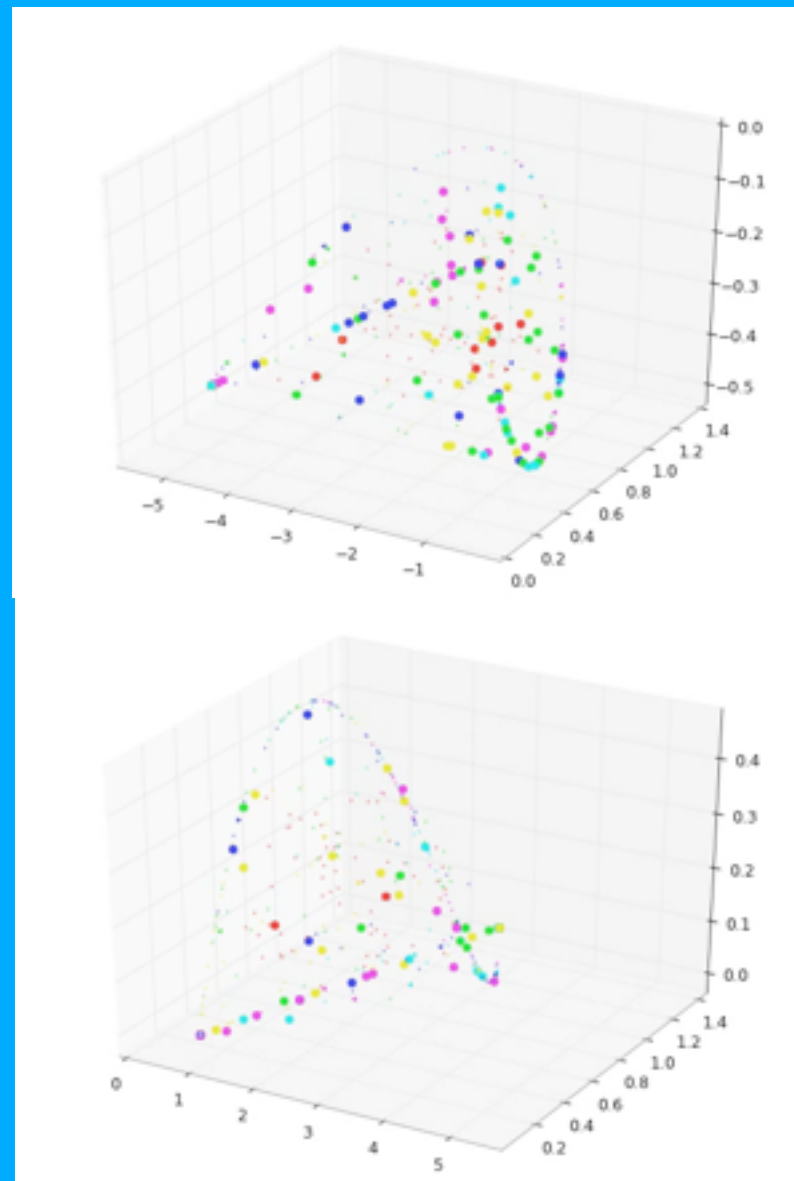
Dynamics on the model manifold: Searching for the best fit

- 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

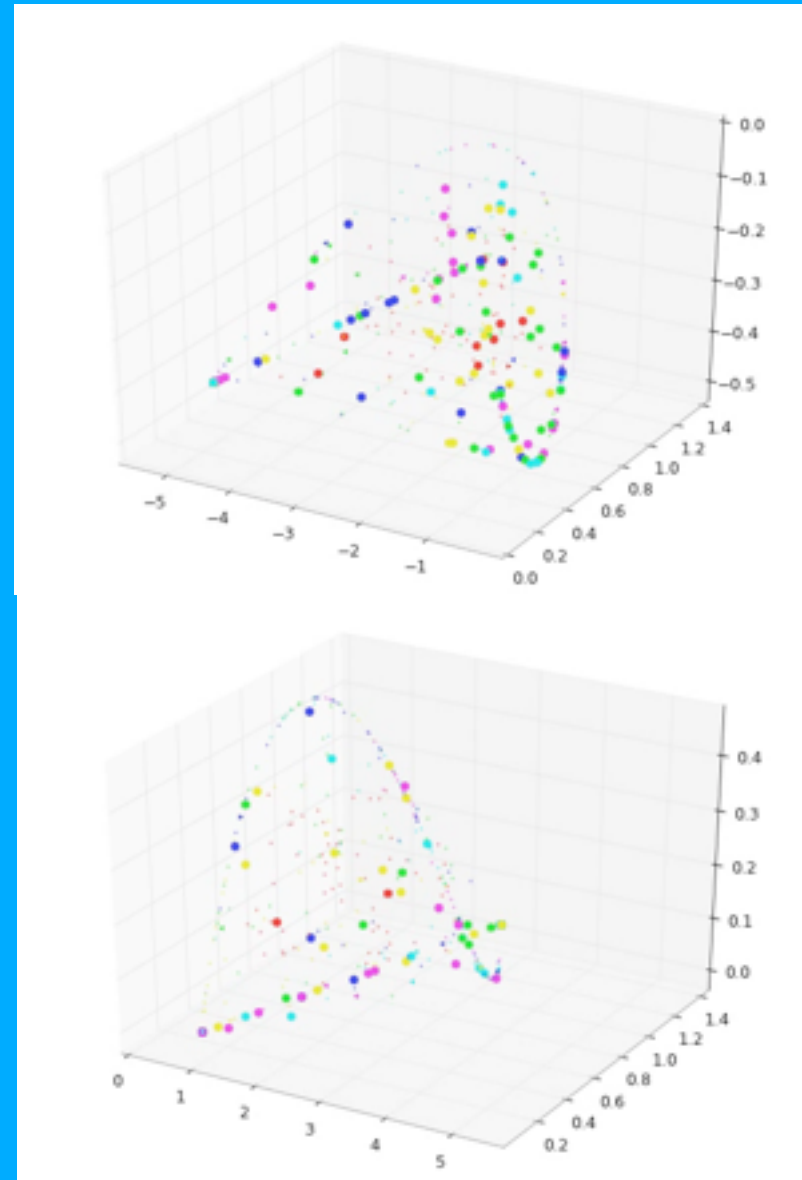
- Jeffrey's prior plus noise
- Big noise concentrates on manifold edges
- Note scales: flat
- Top: Levenberg-Marquardt
- Bottom: Geodesic acceleration
- Large points: Initial conditions which fail to converge to best fit



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

Dynamics on the model manifold: Searching for the best fit

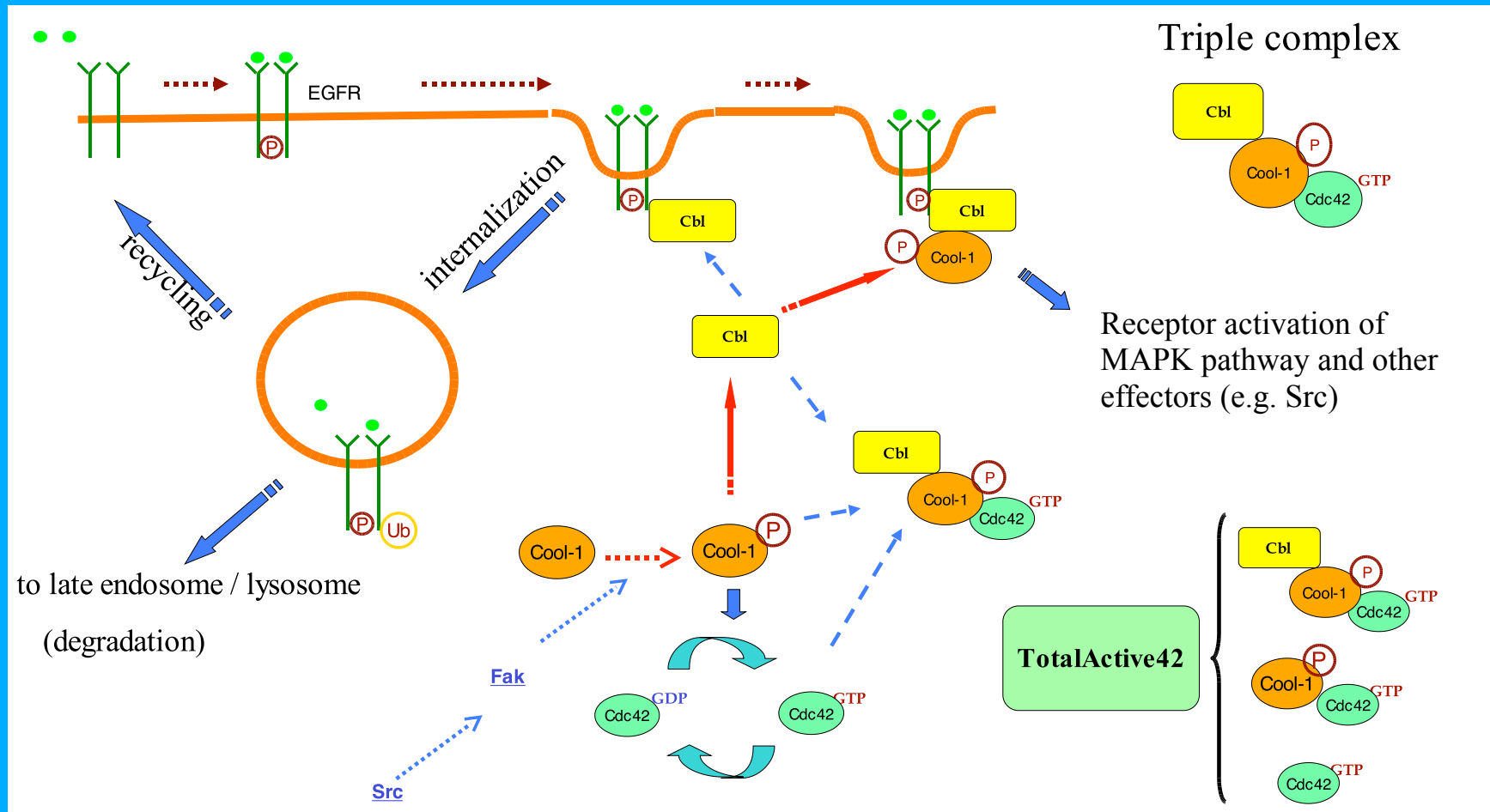
- Jeffrey's prior plus noise
- Big noise concentrates on manifold edges
- Note scales: flat
- Top: Levenberg-Marquardt
- Bottom: Geodesic acceleration
- Large points: Initial conditions which fail to converge to best fit



C. EGFR Trafficking Model

Fergal Casey, Cerione lab

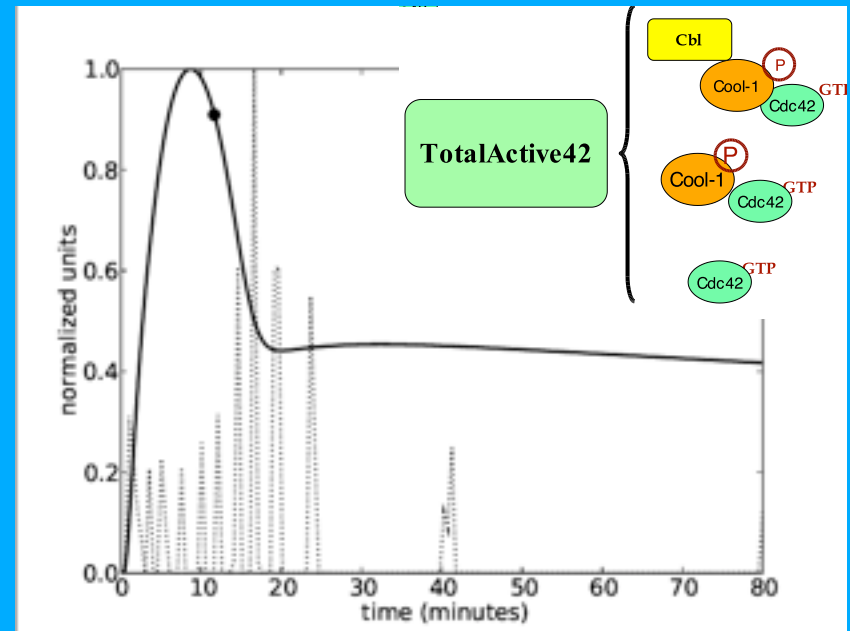
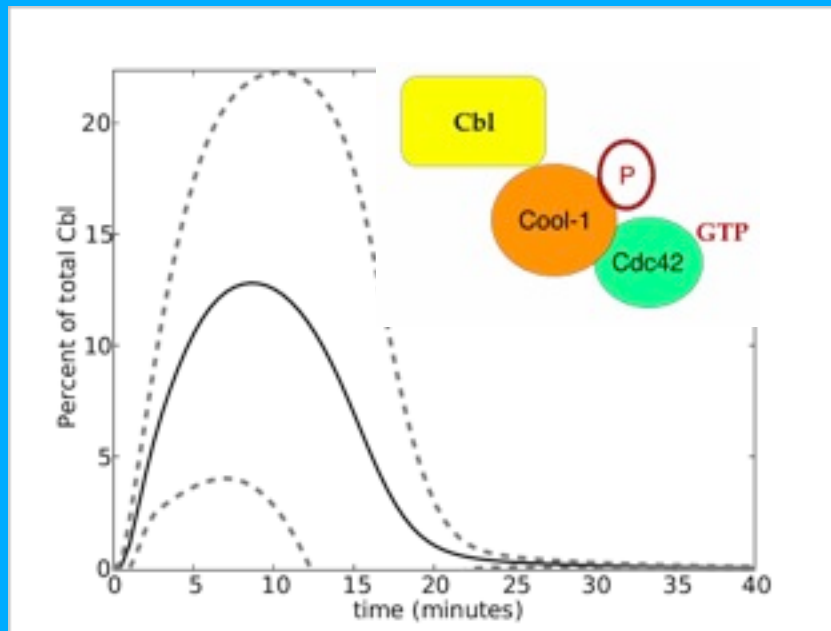
- Active research, Cerione lab: testing hypothesis, experimental design (Cool1 \equiv β -PIX)
- 41 chemicals, 53 rate constants; only 11 of 41 species can be measured
- Does Cool-1 triple complex sequester Cbl, delay endocytosis in wild type NIH3T3 cells?



C. Trafficking: experimental design

Which experiment best reduces prediction uncertainty?

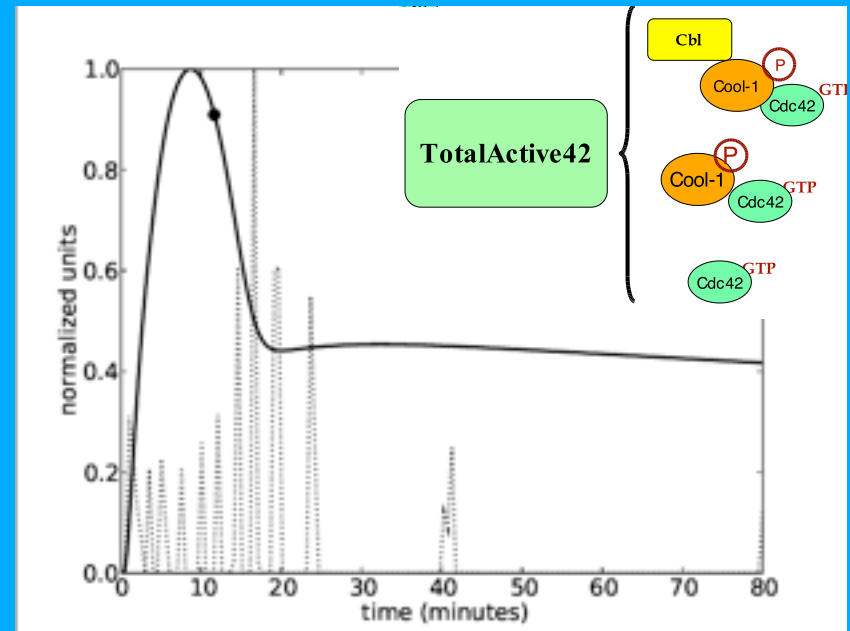
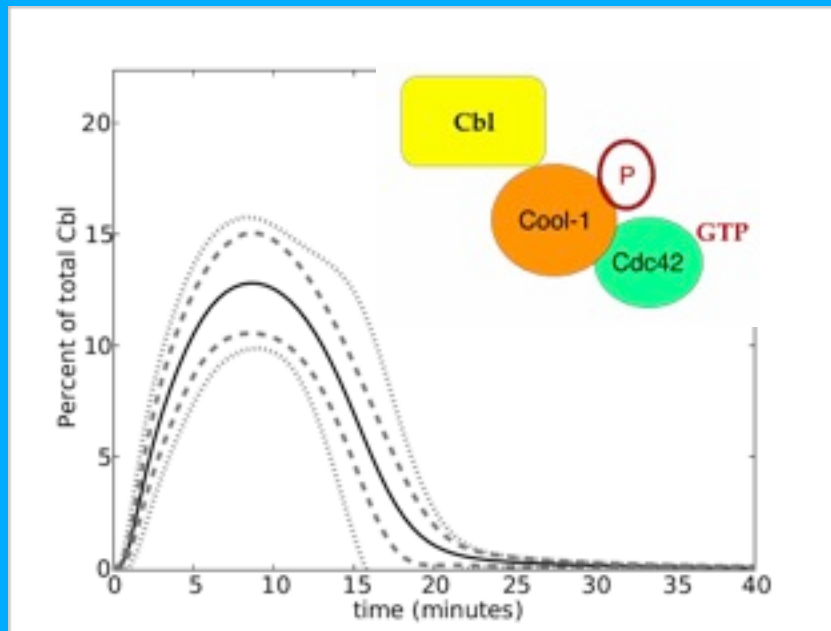
- Amount of triple complex was not well predicted
- V-optimal experimental design: single & multiple measurements
- Total active Cdc42 at 10 min.; Cerione independently concurs
- Experiment indicates significant sequestering in wild type
- Predictivity without decreasing parameter uncertainty



C. Trafficking: experimental design

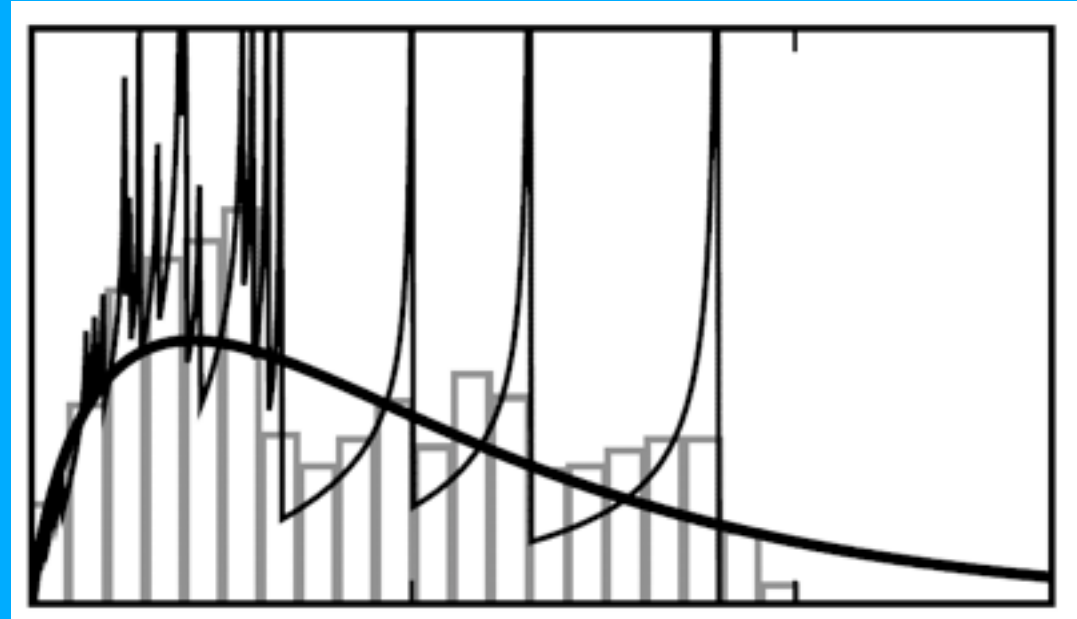
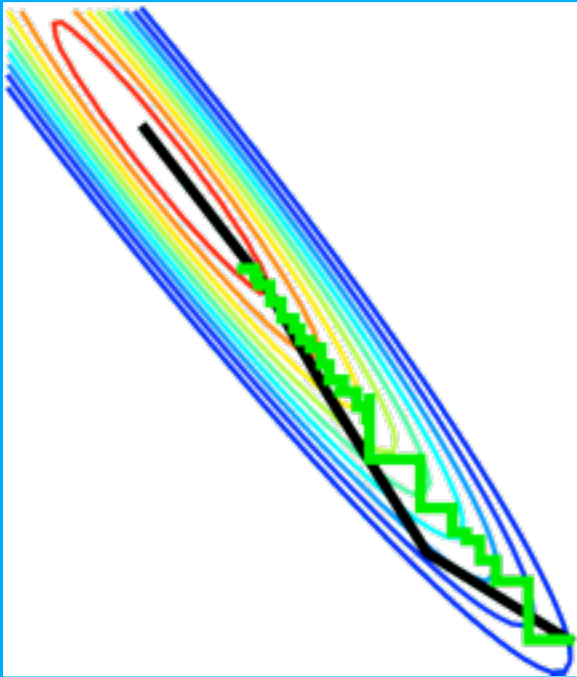
Which experiment best reduces prediction uncertainty?

- 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?



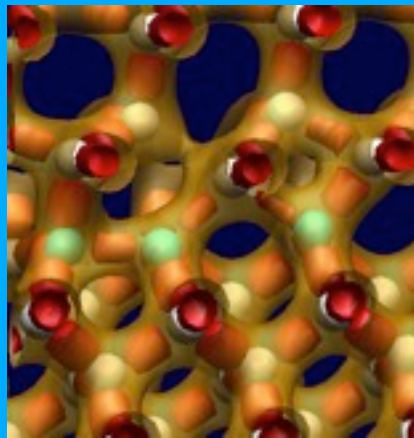
Fitness gain from first successful mutation

- 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

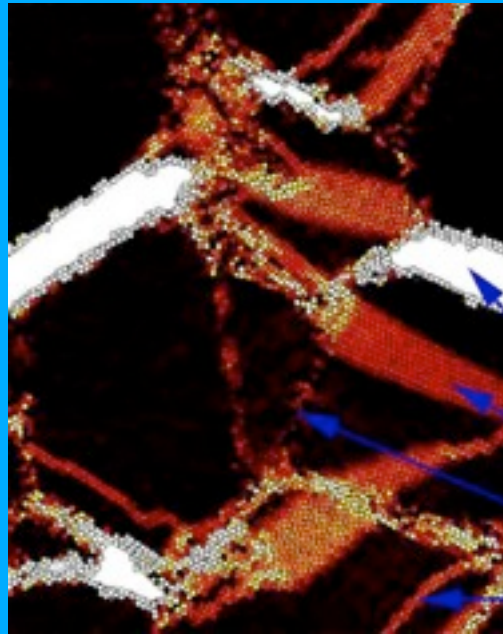
E. Bayesian Errors for Atoms

'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 $V(r_1, r_2, \dots)$

- Fast to compute
- Limit $m_e/M \rightarrow 0$ justified
- Guess functional form
 - Pair potential $\sum V(r_i - r_j)$ 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

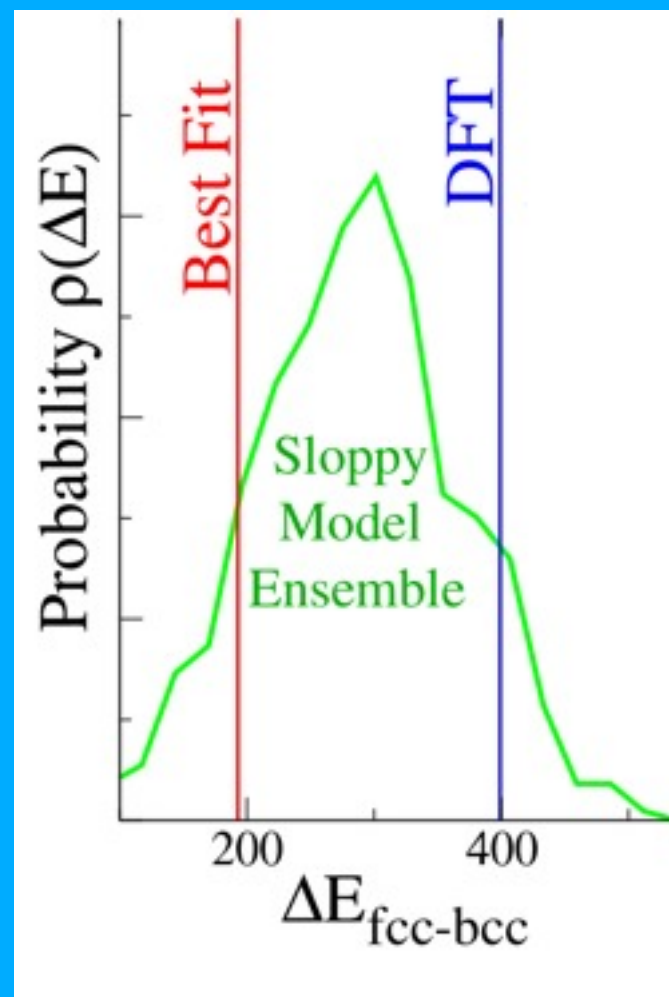
- 3% elastic constant
- 10% forces
- 100% fcc-bcc, dislocation core

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



Green = DFT, Red = Fits

E. Interatomic Potential Error Bars

Ensemble of Acceptable Fits to Data

Not *transferable*

Unknown errors

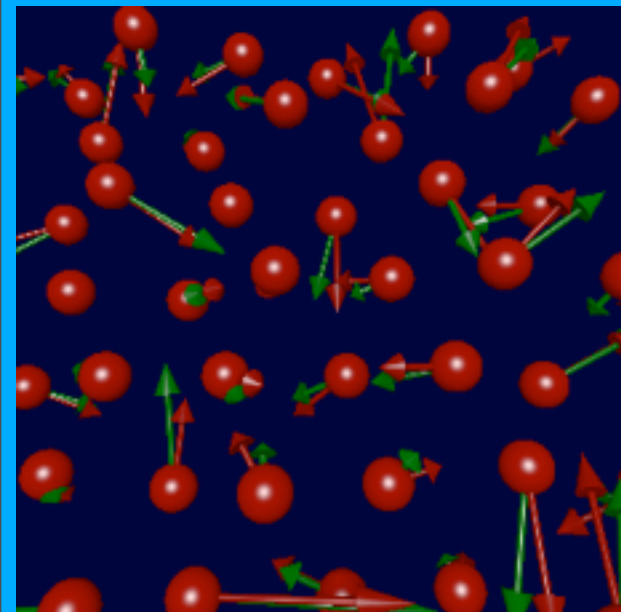
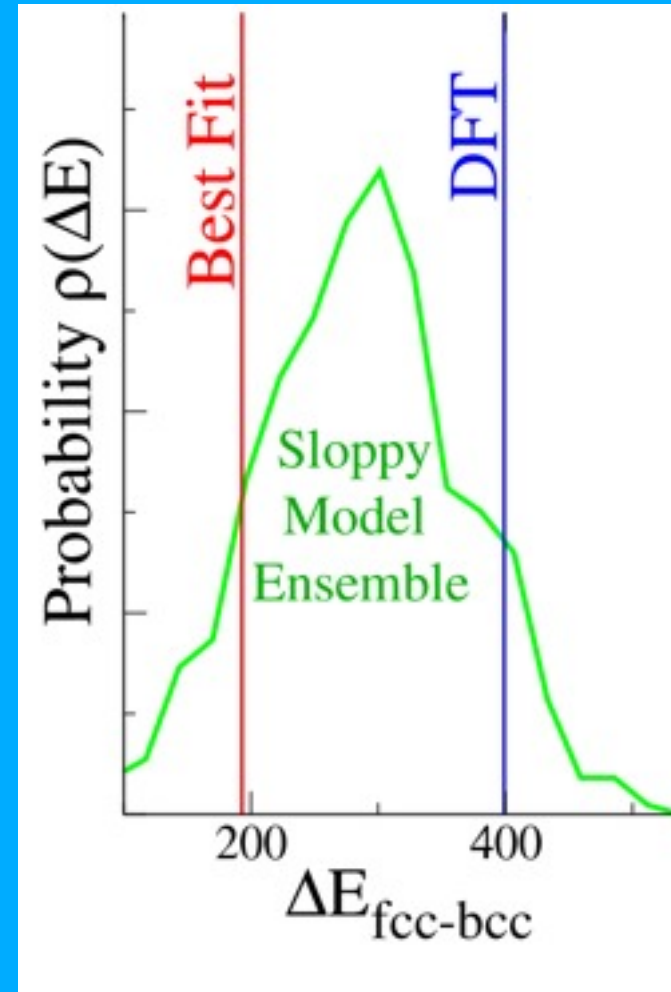
- 3% elastic constant
- 10% forces
- 100% fcc-bcc, dislocation core

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



Green = DFT, Red = Fits

Sloppy Molybdenum: Does it Work?

Estimating *Systematic* Errors

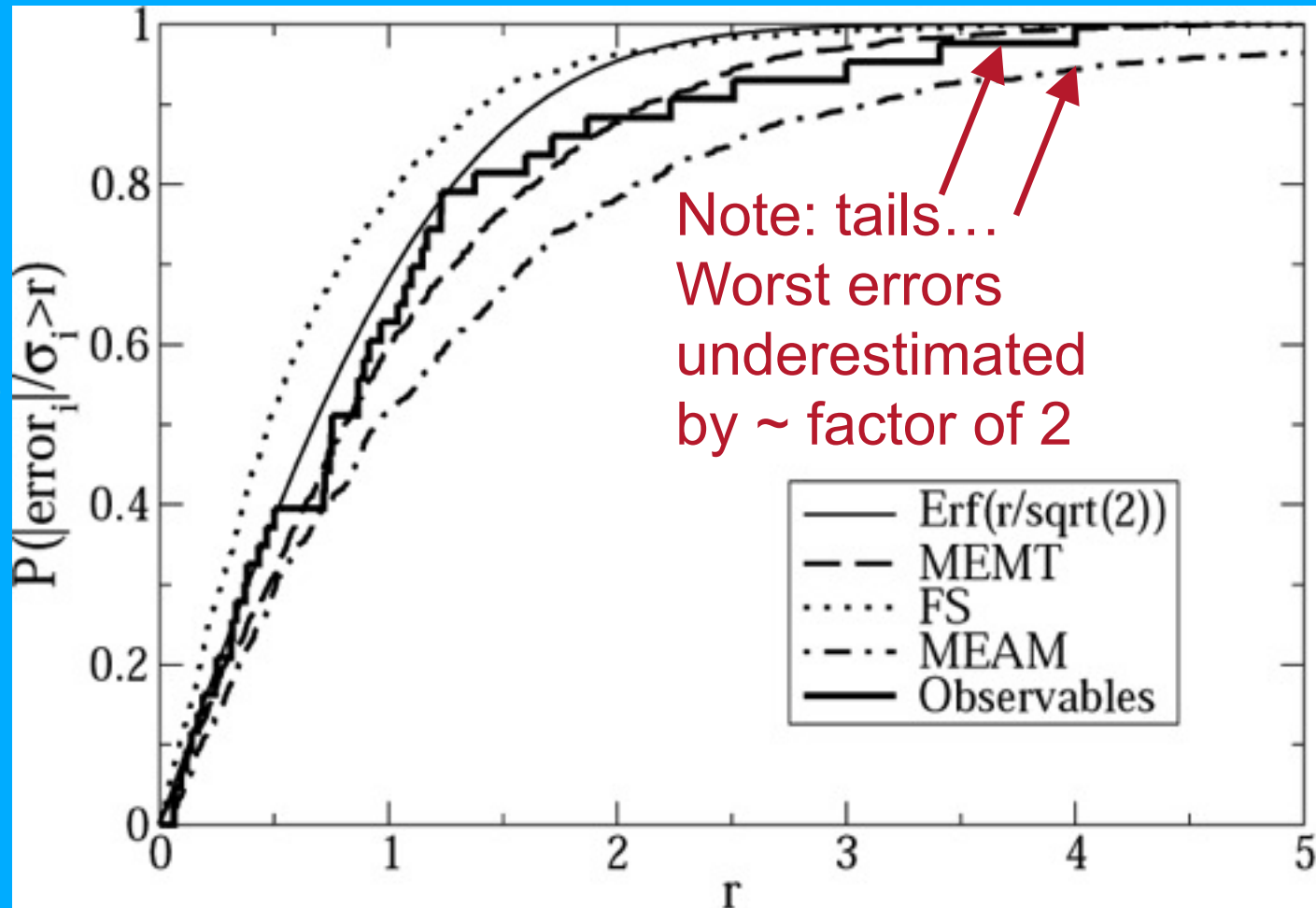
Bayesian error σ_i gives total error if ratio $r = \text{error}_i/\sigma_i$ distributed as a Gaussian: cumulative distribution $P(r) = \text{Erf}(r/\sqrt{2})$

Three potentials

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

“Sloppy model”
systematic
error most of
total

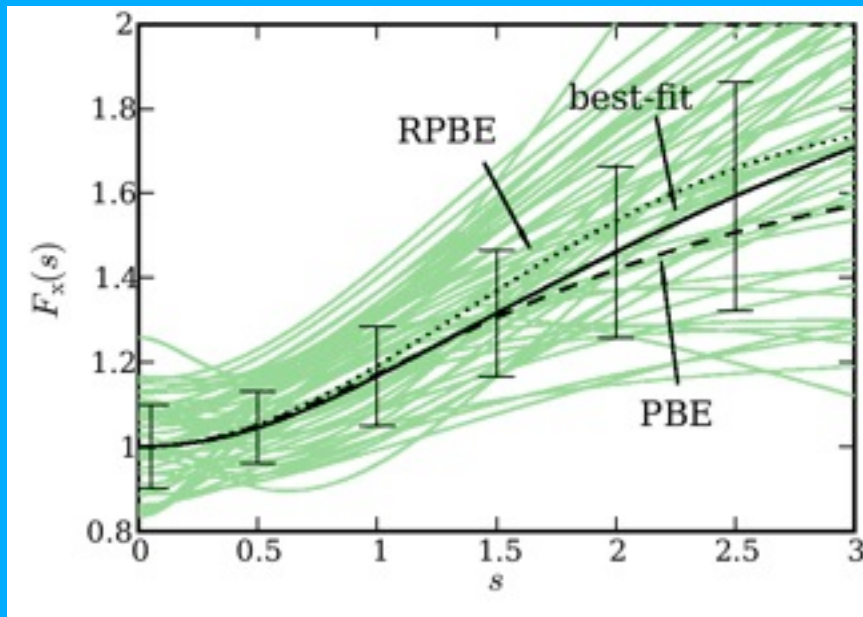
$\sim 2 \ll 200\%/7\%$



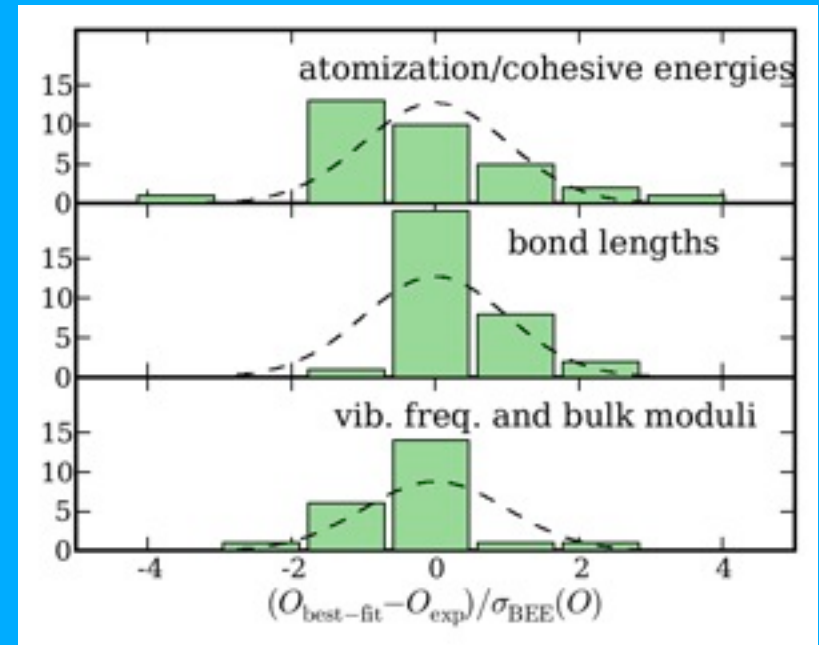
Systematic Error Estimates for DFT

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)



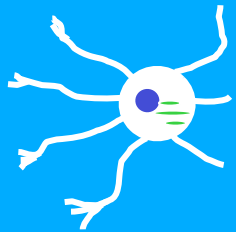
Enhancement factor $F_x(s)$
in the exchange energy E_x
Large fluctuations



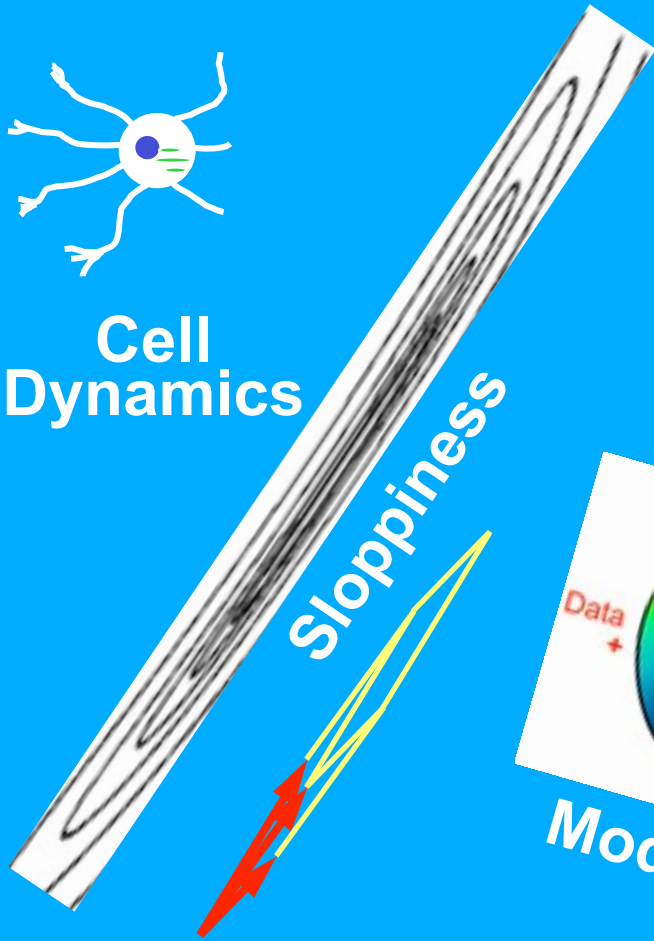
Actual error / predicted error
**Deviation from experiment
well described by ensemble!**

'Sloppy Model' Nonlinear Fits: Signal Transduction to Differential Geometry

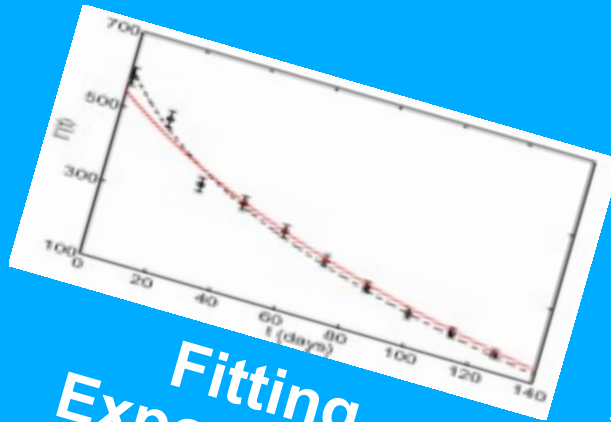
JPS, Mark Transtrum, Ben Machta, Ricky Chachra, Isabel Kloumann, Kevin Brown, Ryan Gutenkunst, Josh Waterfall, Chris Myers, ...



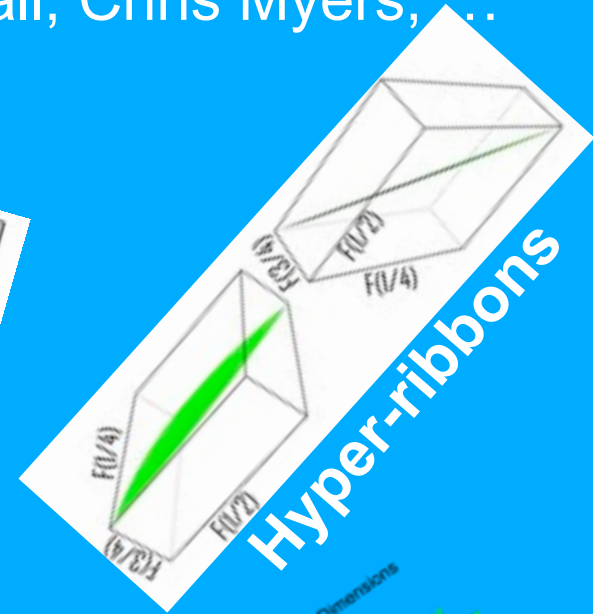
Cell Dynamics



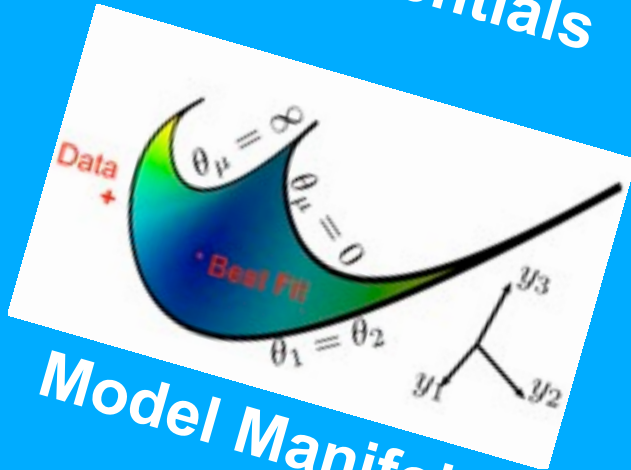
Sloppiness



Fitting Exponentials



Hyper-ribbons



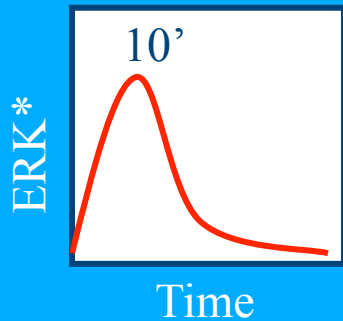
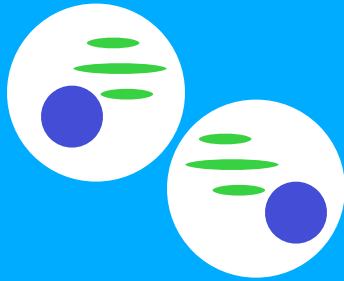
Model Manifold



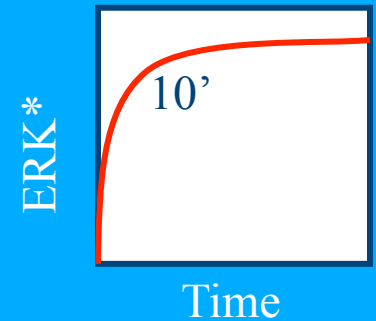
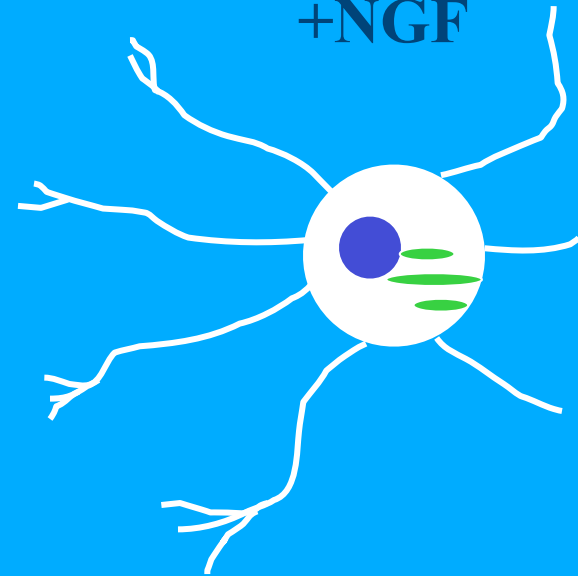
Coarse-Grained Models

PC12 Differentiation

+EGF



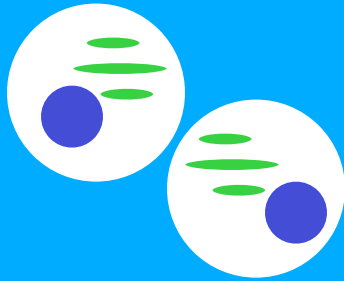
+NGF



Biologists study which proteins talk to which. Modeling?

PC12 Differentiation

+EGF



EGFR

NGFR

Sos

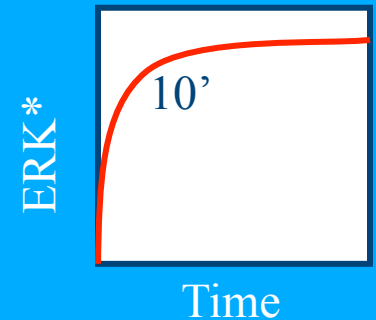
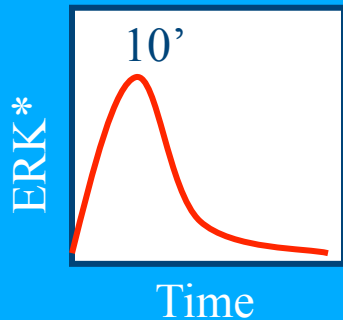
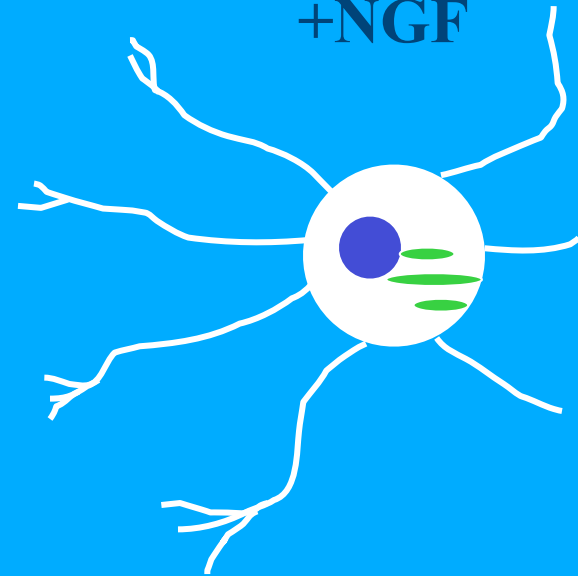
Ras

Raf-1

MEK1/2

ERK1/2

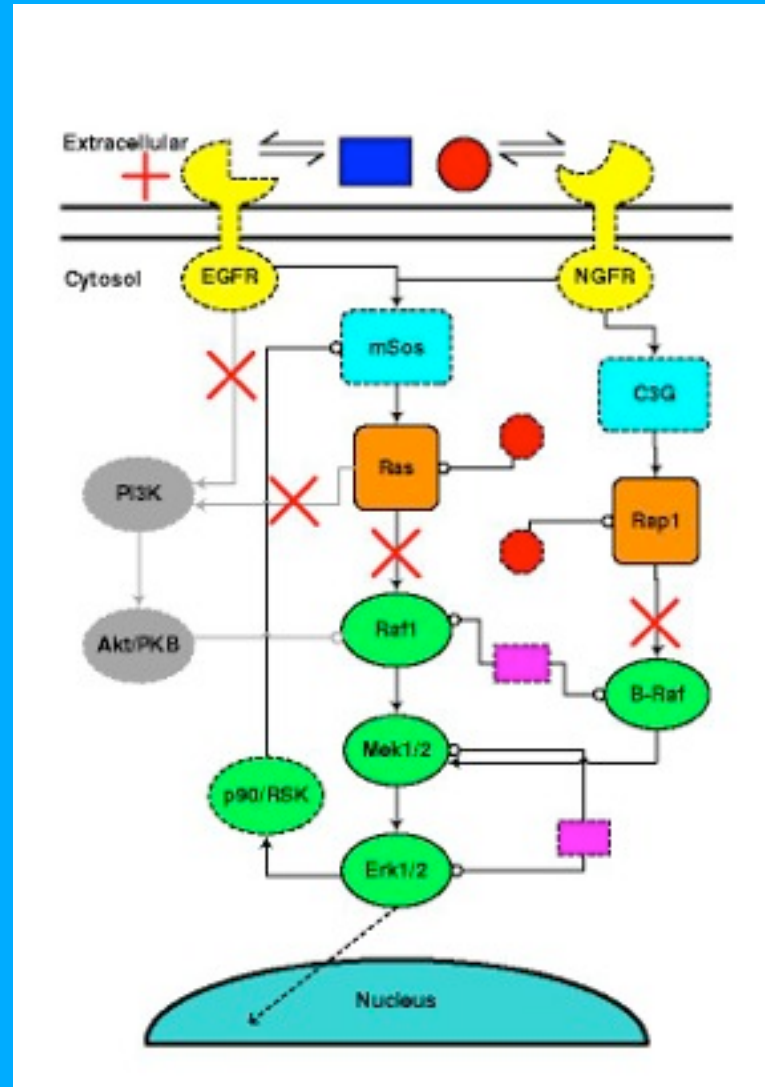
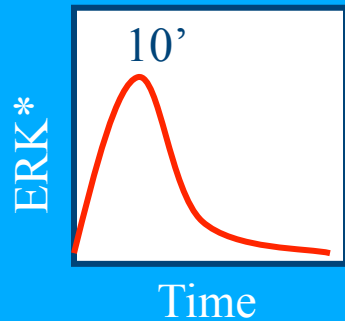
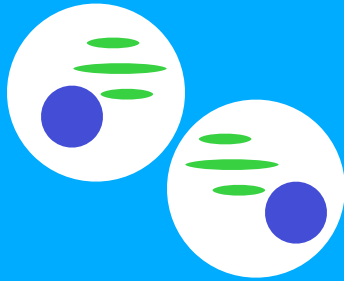
+NGF



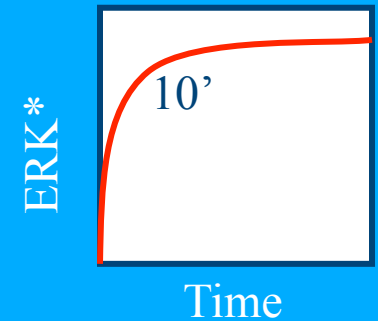
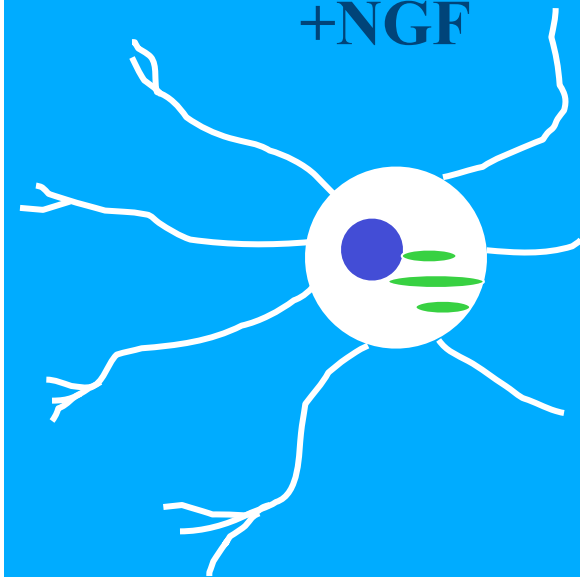
Biologists study which proteins talk to which. Modeling?

PC12 Differentiation

+EGF

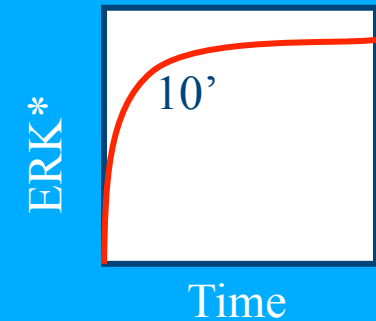
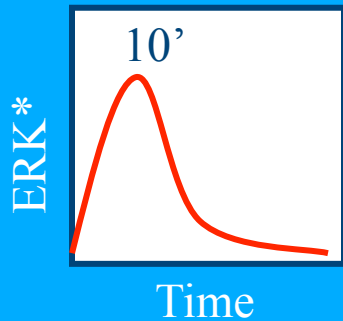
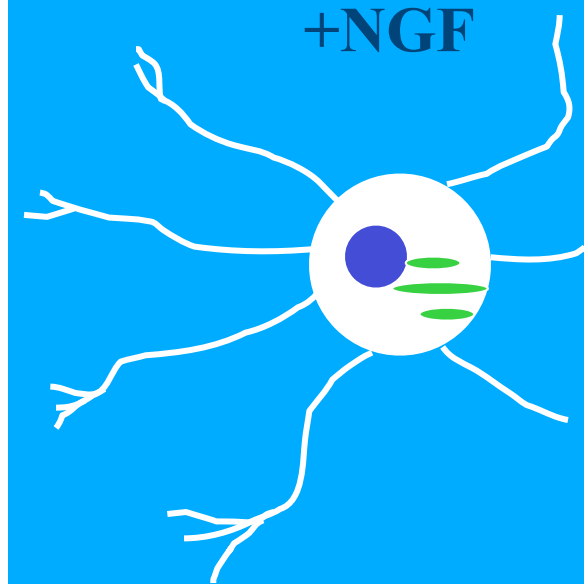
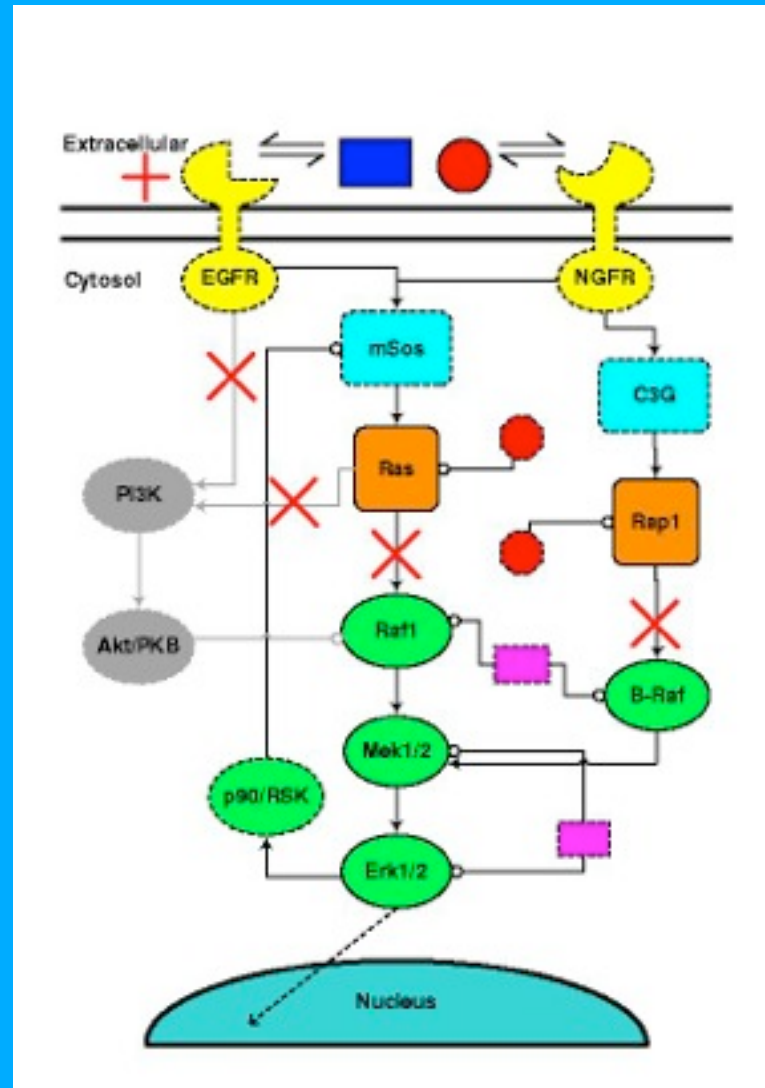


+NGF



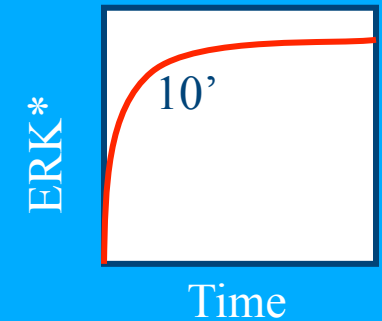
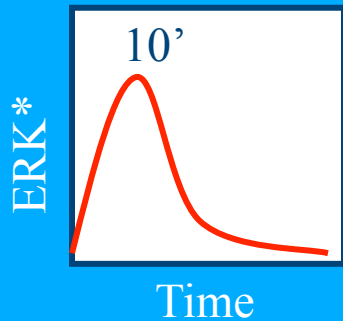
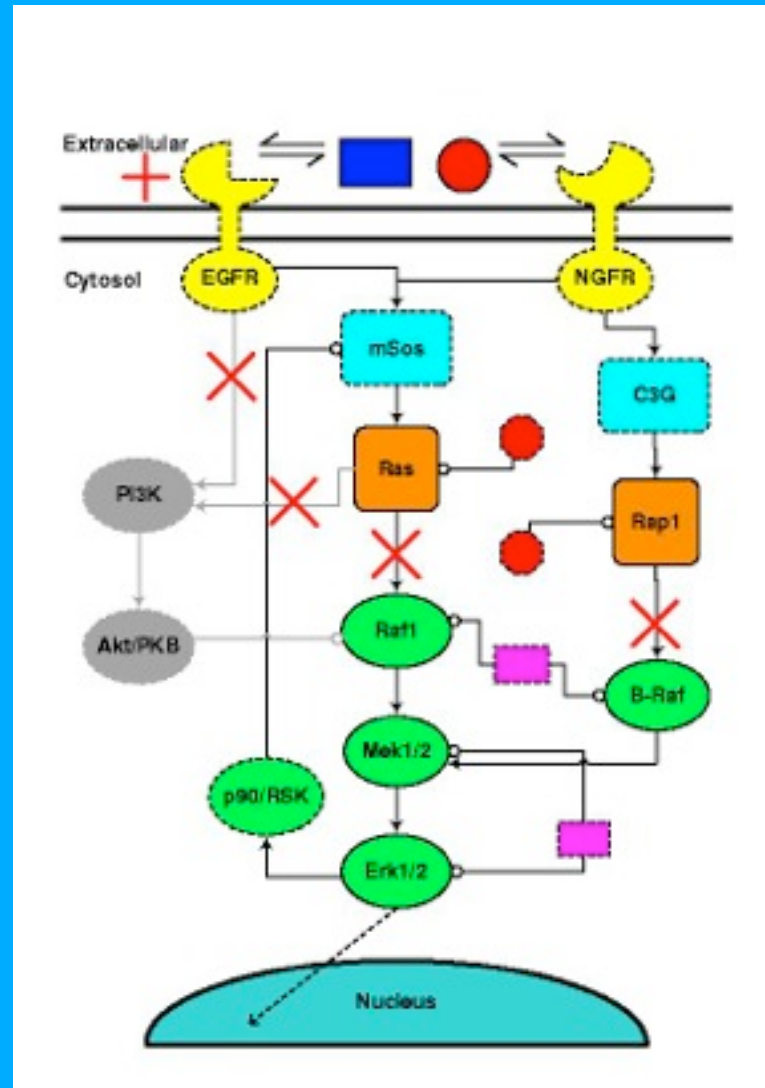
Biologists study which proteins talk to which. Modeling?

PC12 Differentiation



Biologists study which proteins talk to which. Modeling?

PC12 Differentiation



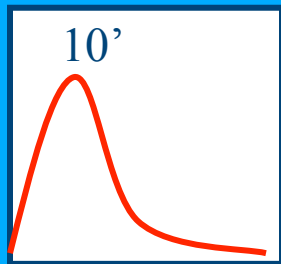
Biologists study which proteins talk to which. Modeling?

PC12 Differentiation

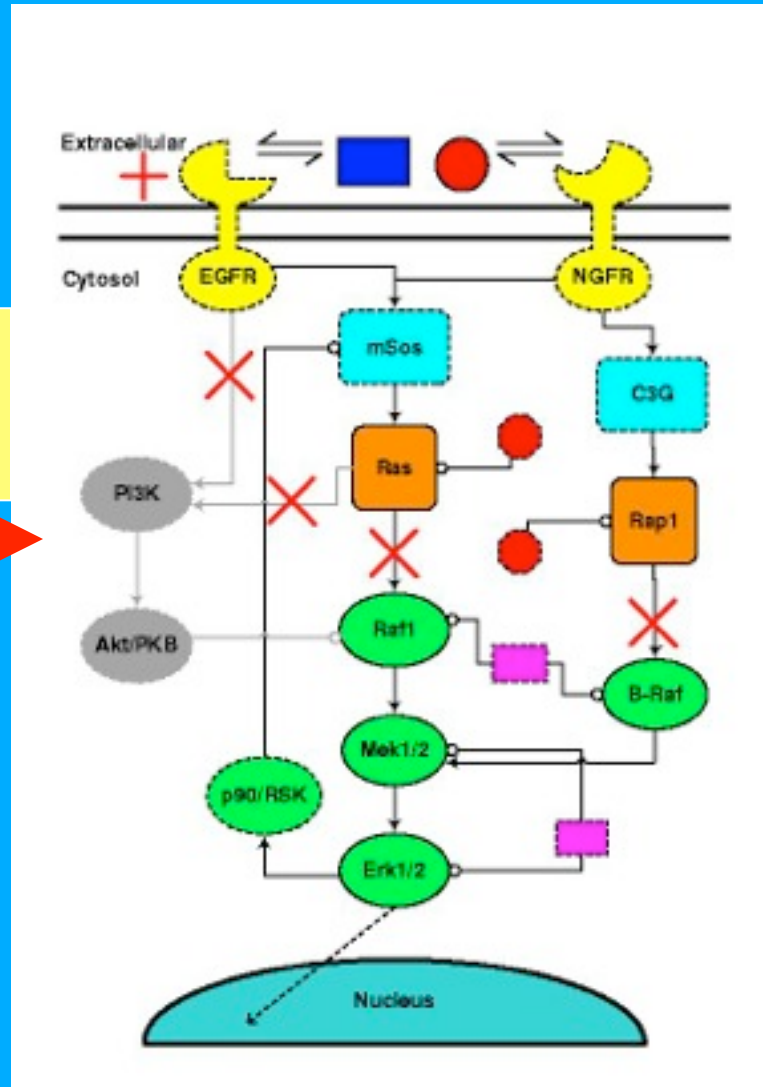
Tunes down signal (Raf-1)



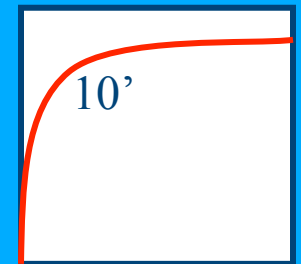
ERK*



Time



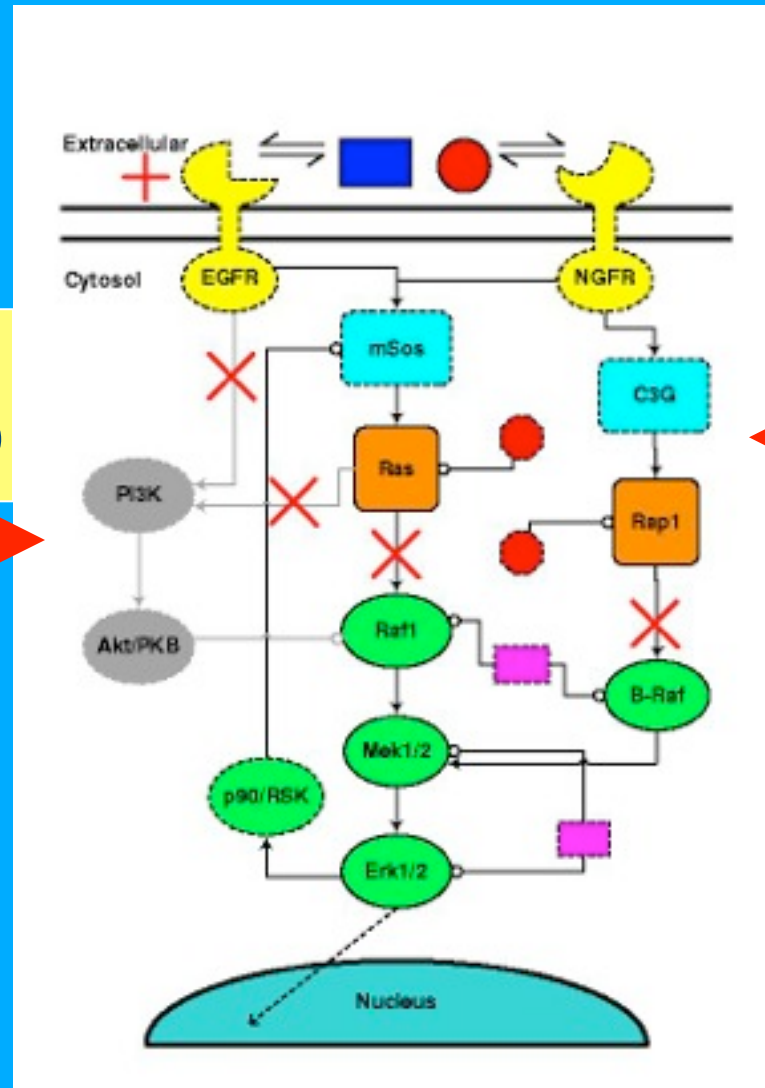
ERK*



Time

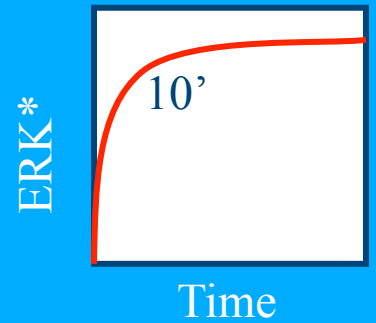
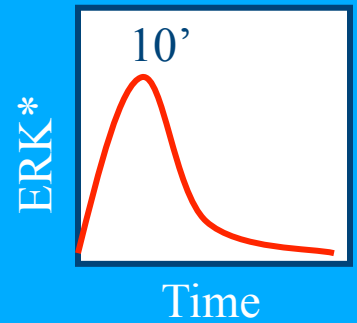
Biologists study which proteins talk to which. Modeling?

PC12 Differentiation



Pumps up signal (Mek)

Tunes down signal (Raf-1)

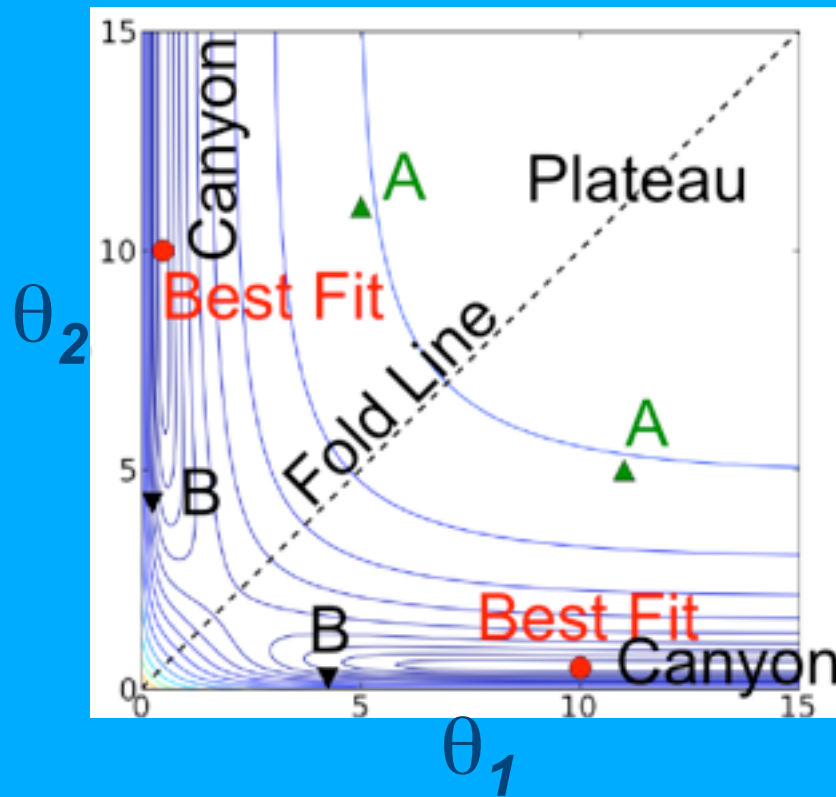
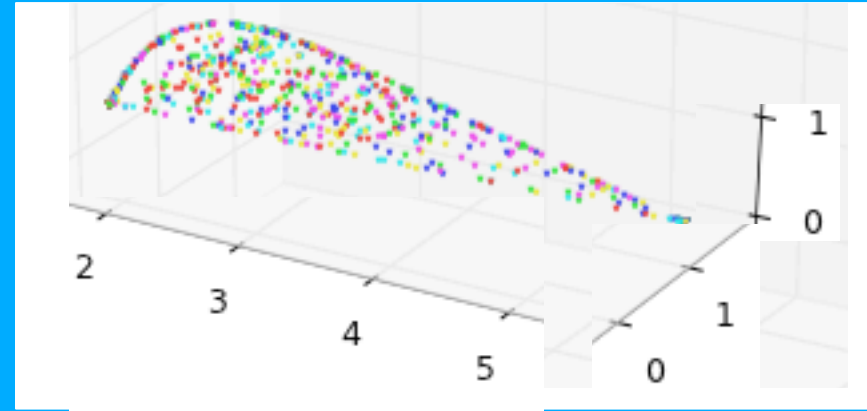


Biologists study which proteins talk to which. Modeling?

Edges of the model manifold

Fitting Exponentials

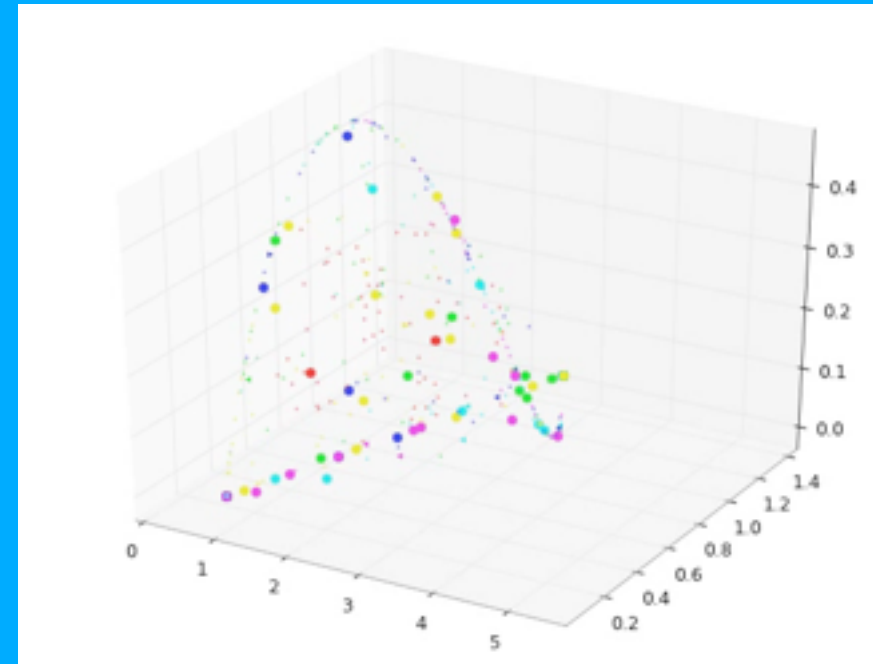
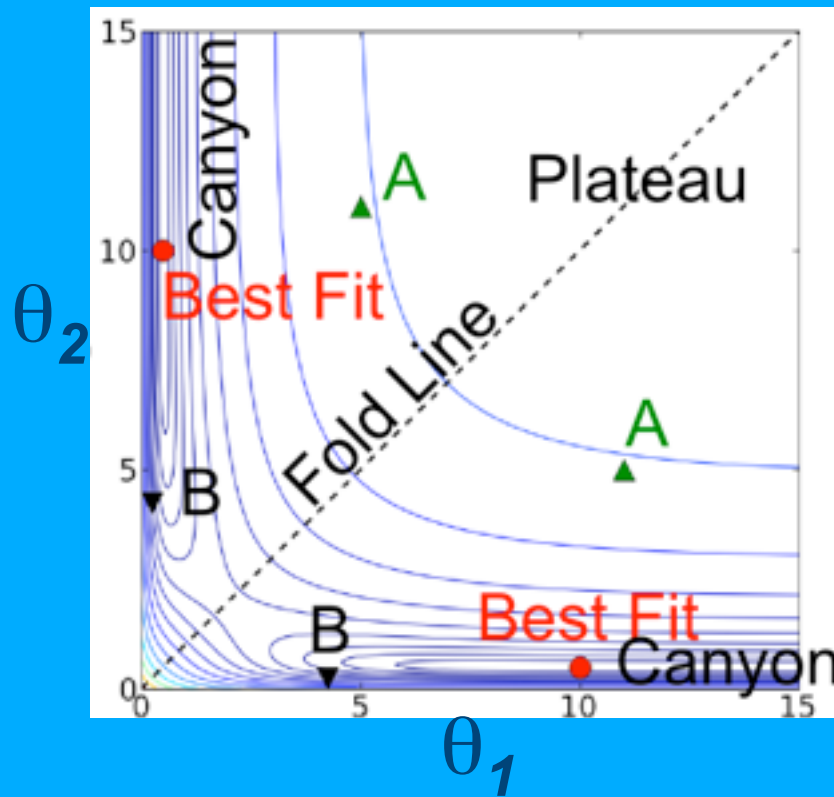
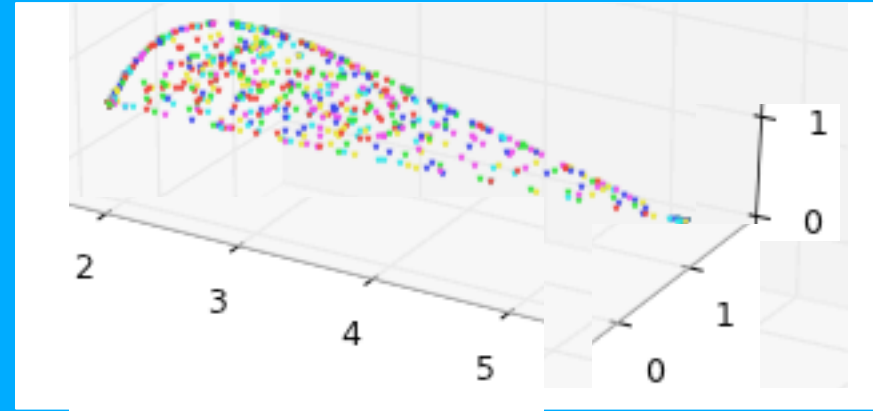
Top: Flat model manifold;
articulated edges = plateau
Bottom: Stretch to uniform
aspect ratio (Isabel Kloumann)



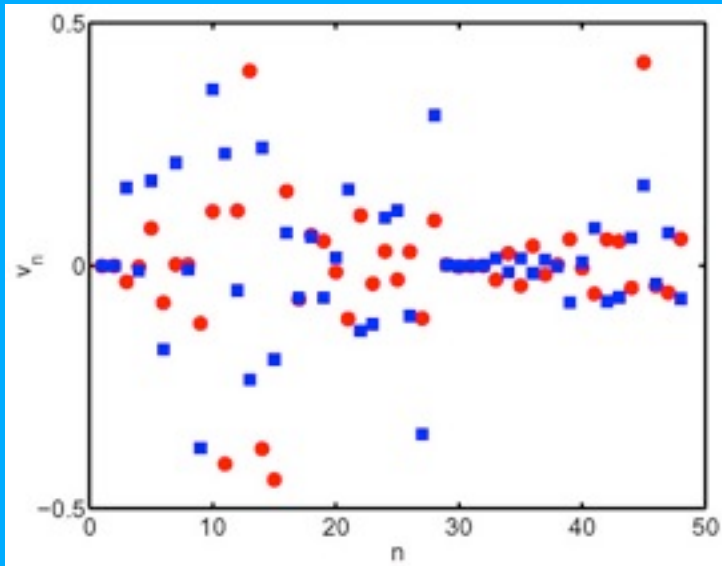
Edges of the model manifold

Fitting Exponentials

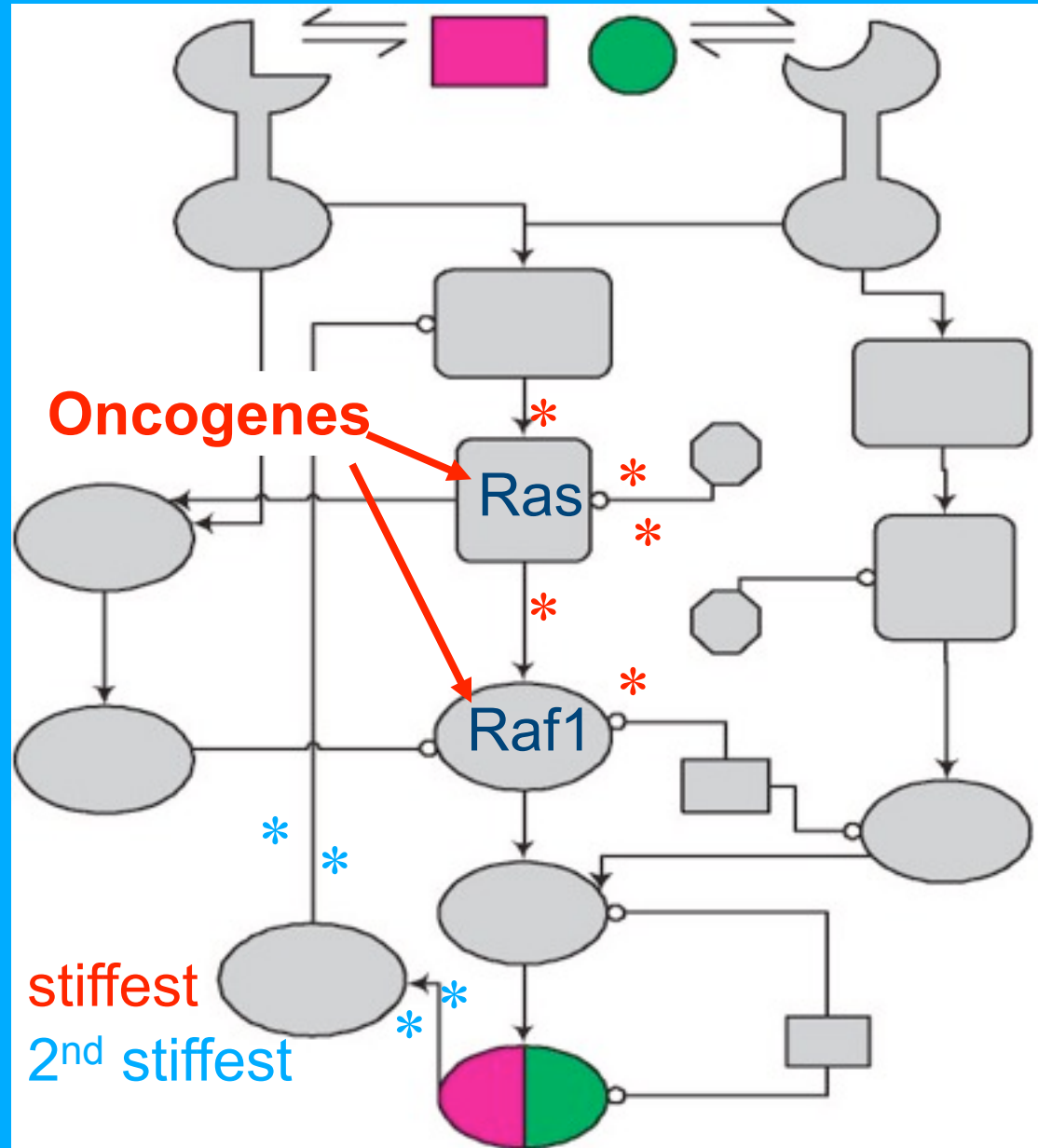
Top: Flat model manifold;
articulated edges = plateau
Bottom: Stretch to uniform
aspect ratio (Isabel Kloumann)



Which Rate Constants are in the Stiffest Eigenvector?

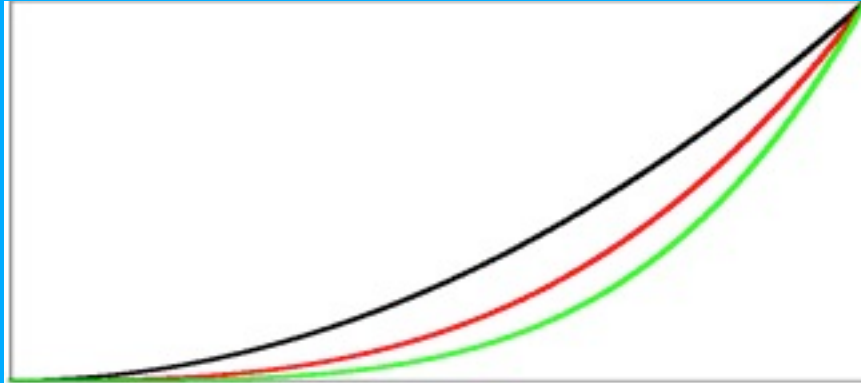


Eigenvector components along the bare parameters reveal which ones are most important for a given eigenvector.



Where is Sloppiness From?

Fitting Polynomials to Data



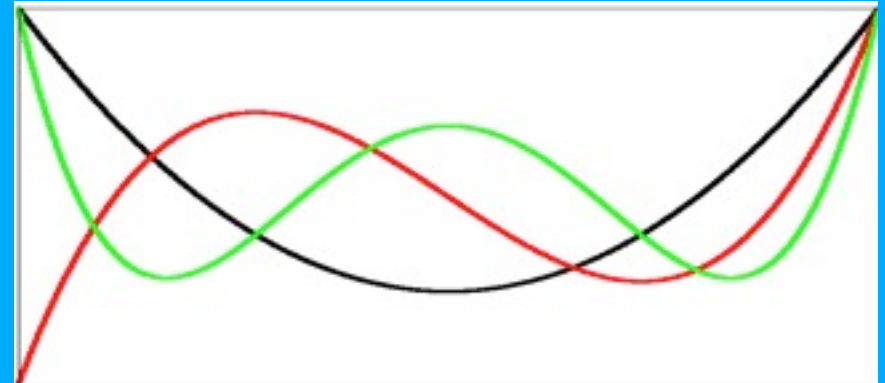
Fitting Monomials to Data

$$y = \sum a_n x^n$$

Functional Forms Same

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

Hilbert matrix: famous



Orthogonal Polynomials

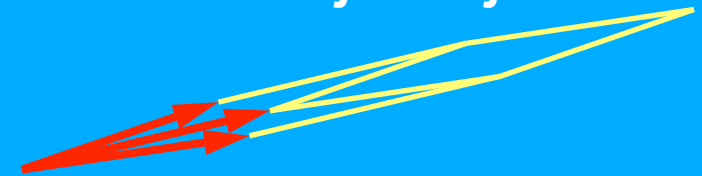
$$y = \sum b_n L_n(x)$$

Functional Forms Distinct

Eigen Parameters

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

Sloppiness arises when bare parameters skew in eigenbasis

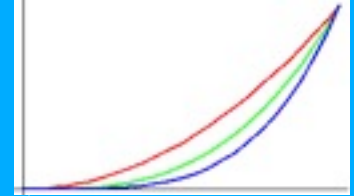


Small Determinant!

$$|H| = \prod \lambda_n$$

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: $\theta_i = \theta_0 + \varepsilon_i$

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

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

Vandermonde
Matrix

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

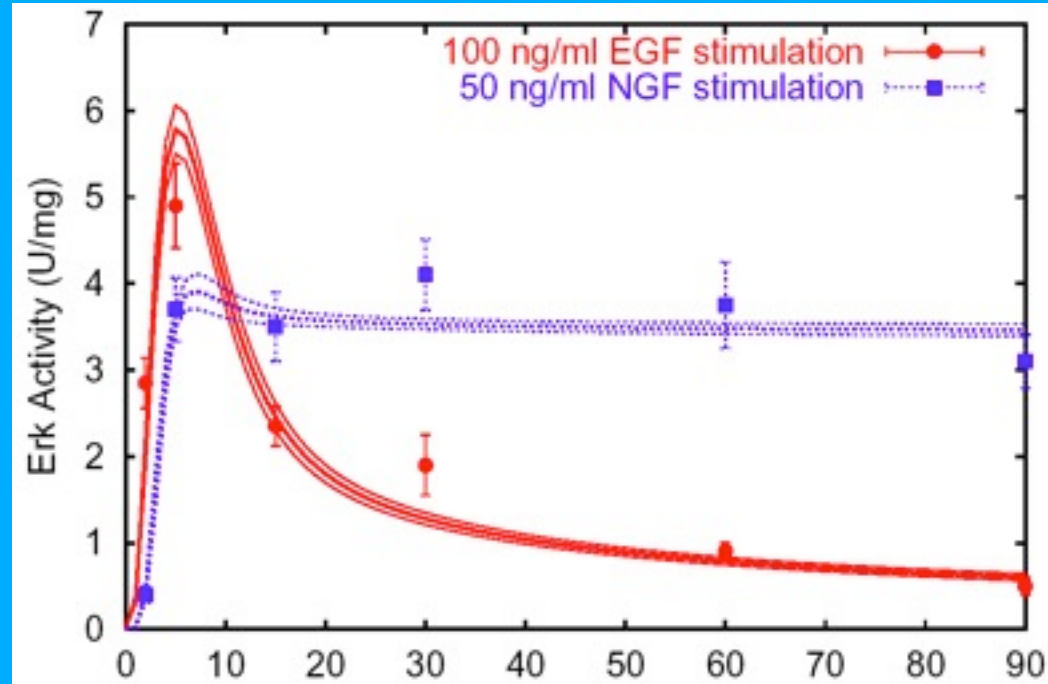
- Implies enormous range of eigenvalues
- Implies equal spacing of log eigenvalues
- Like universality for random matrices

48 Parameter "Fit" to Data

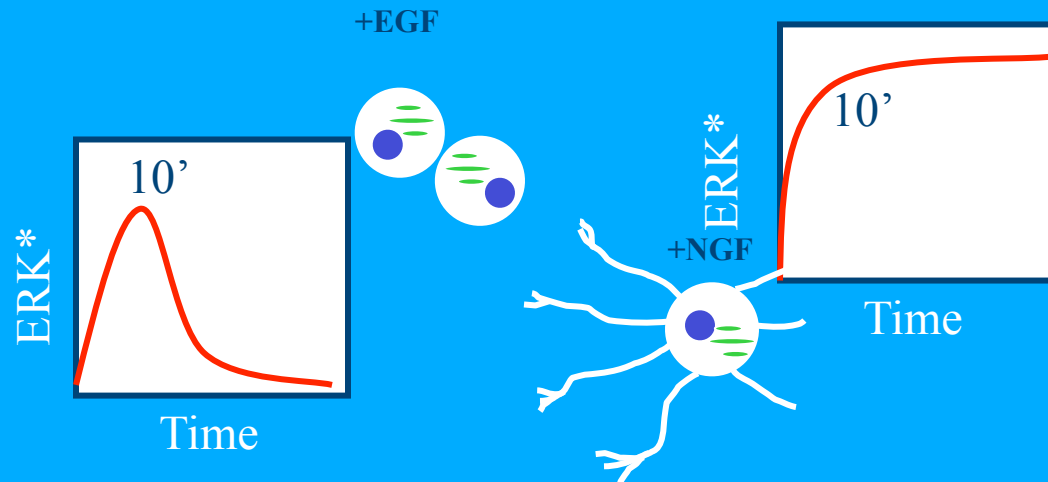
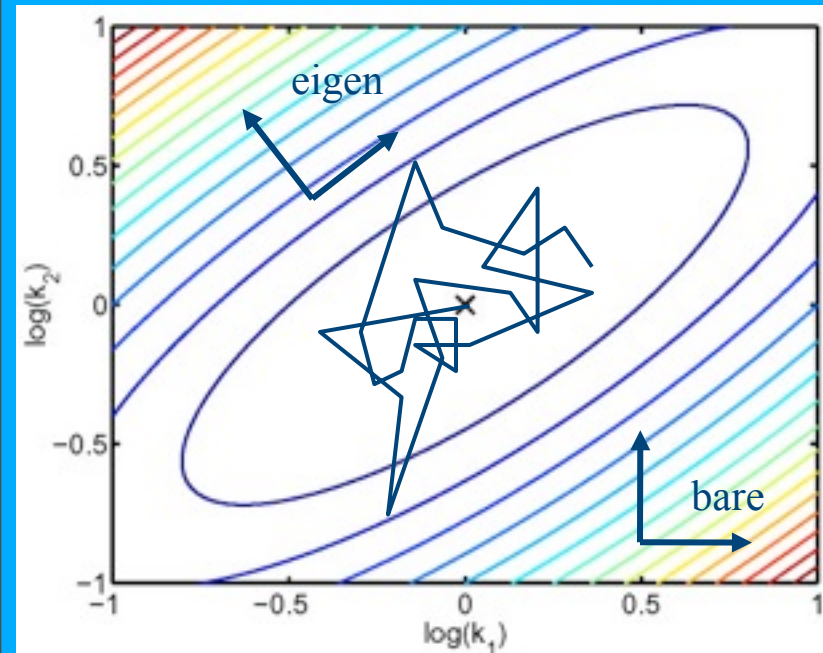
Cost is Energy

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

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

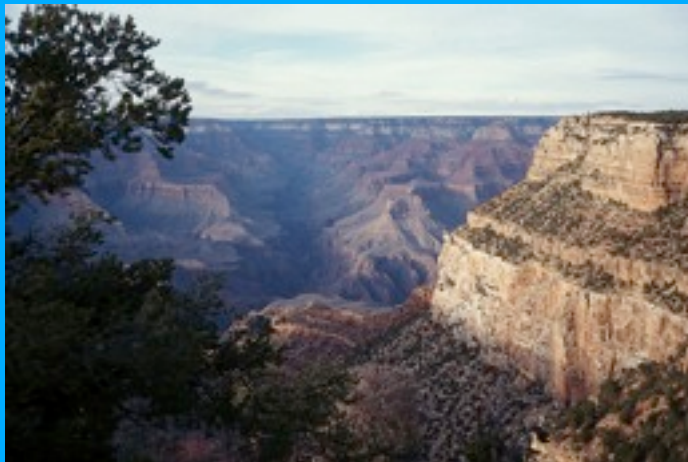
Transition State Passes

Optimization Hell: Golf Course

Sloppy Models

Minima: 5 stiff, N-5 sloppy

Search: Flat planes with cliffs



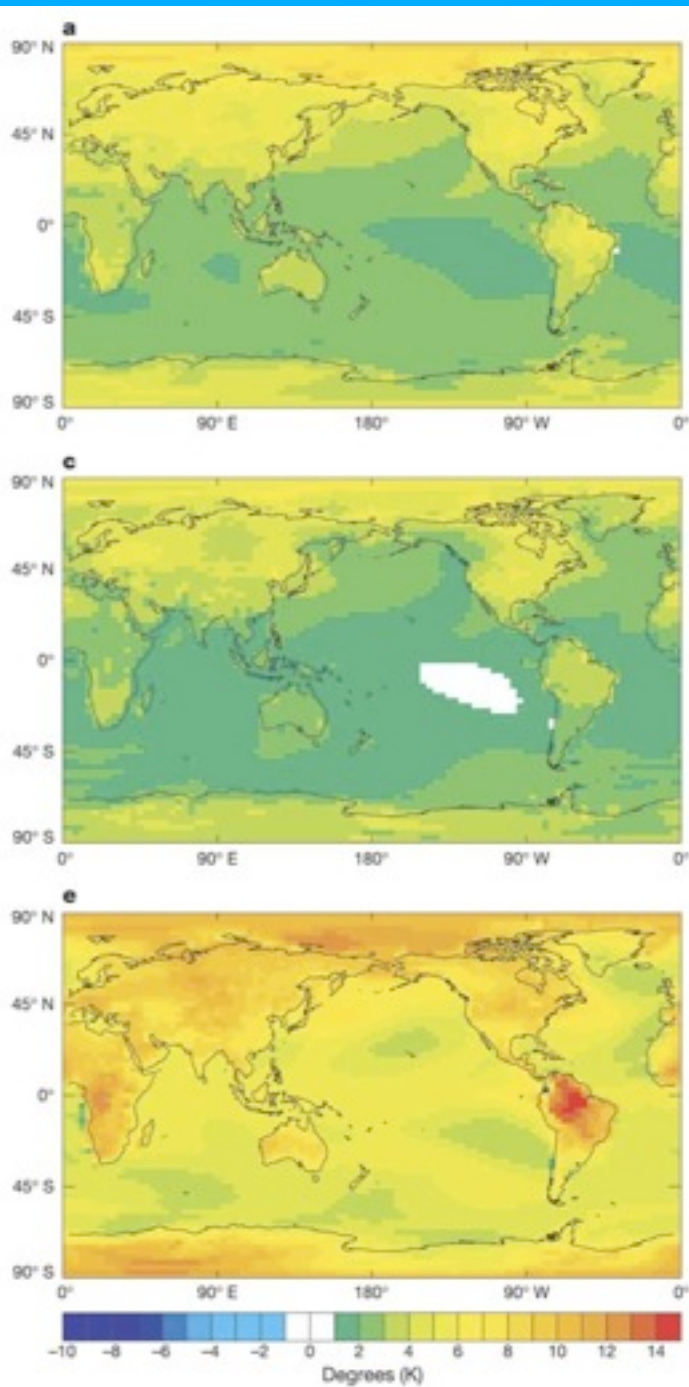
Climate Change

Climate models contain many unknown parameters, fit to data

- General Circulation Model (air, oceans, clouds), exploring doubling of CO₂
- 21 total parameters
- Initial conditions and (only) 6 “cloud dynamics” parameters varied
- Heating typically 3.4K, ranged from < 2K to > 11K

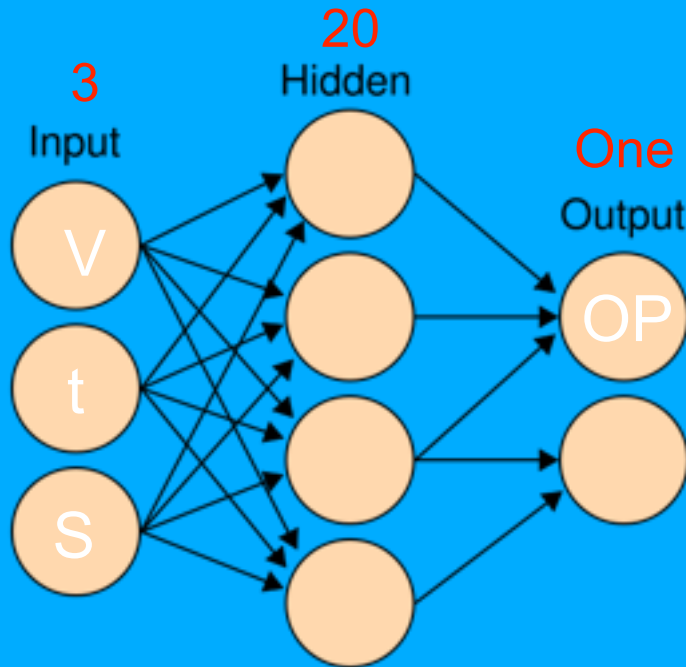
Stainforth et al., *Uncertainty in predictions of the climate response to rising levels of greenhouse gases*, **Nature** 433, 403-406 (2005)

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 :

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

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

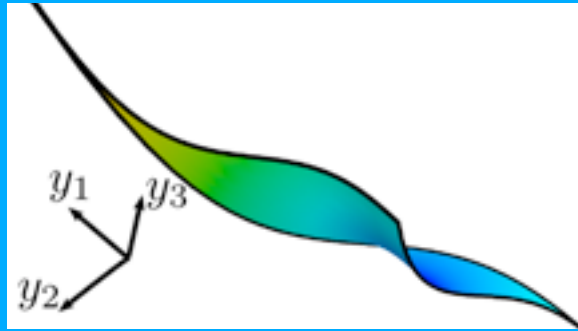
(<http://www.scientific-consultants.com/nnbd.html>)

Mark Transtrum

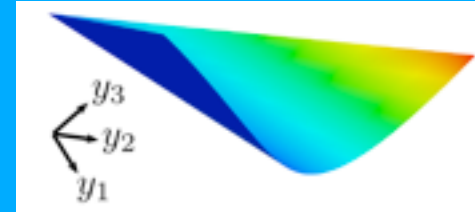
Curvatures

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

- determines geodesic shortest paths
- independent of embedding, parameters

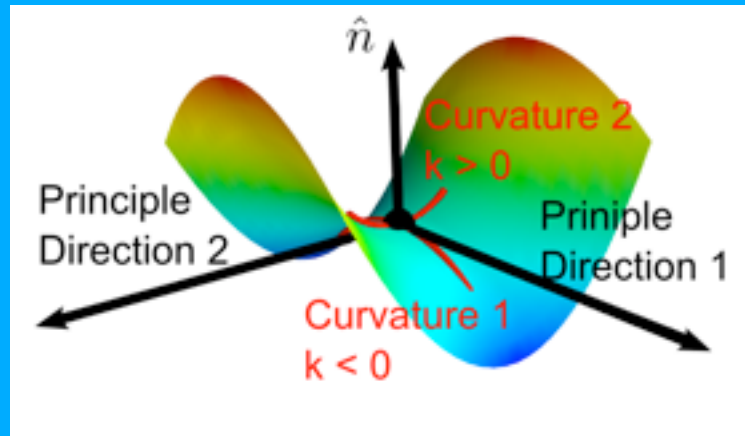


No intrinsic curvature



Extrinsic curvature

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

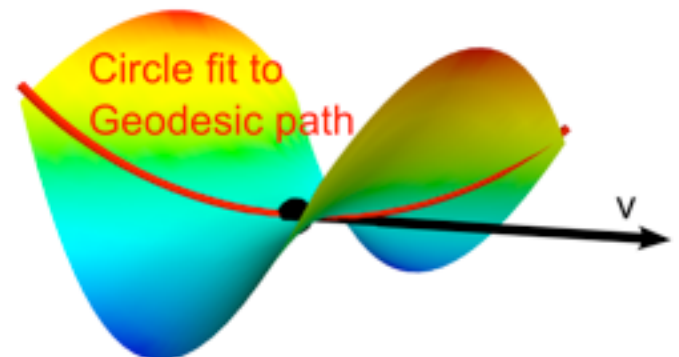


Shape Operator

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:

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

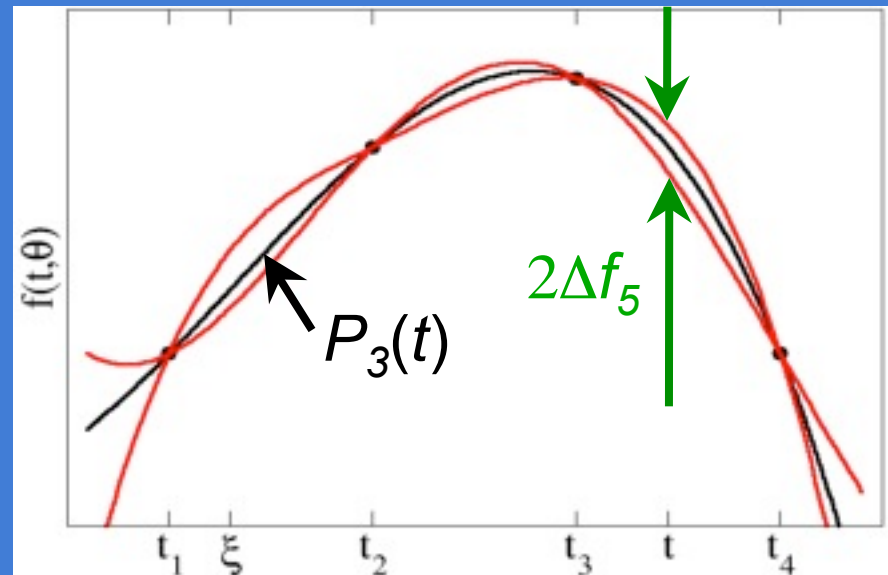
Polynomial fit $P_{m-1}(t)$

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

Interpolation convergence theorem

$$\begin{aligned} \Delta f_{m+1} &= f(t) - P_{m-1}(t) \\ &< (t-t_1) \dots (t-t_m) f^{(m)}(\xi)/m! \\ &\sim (\Delta t / R)^m \end{aligned}$$

More than one data per R



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

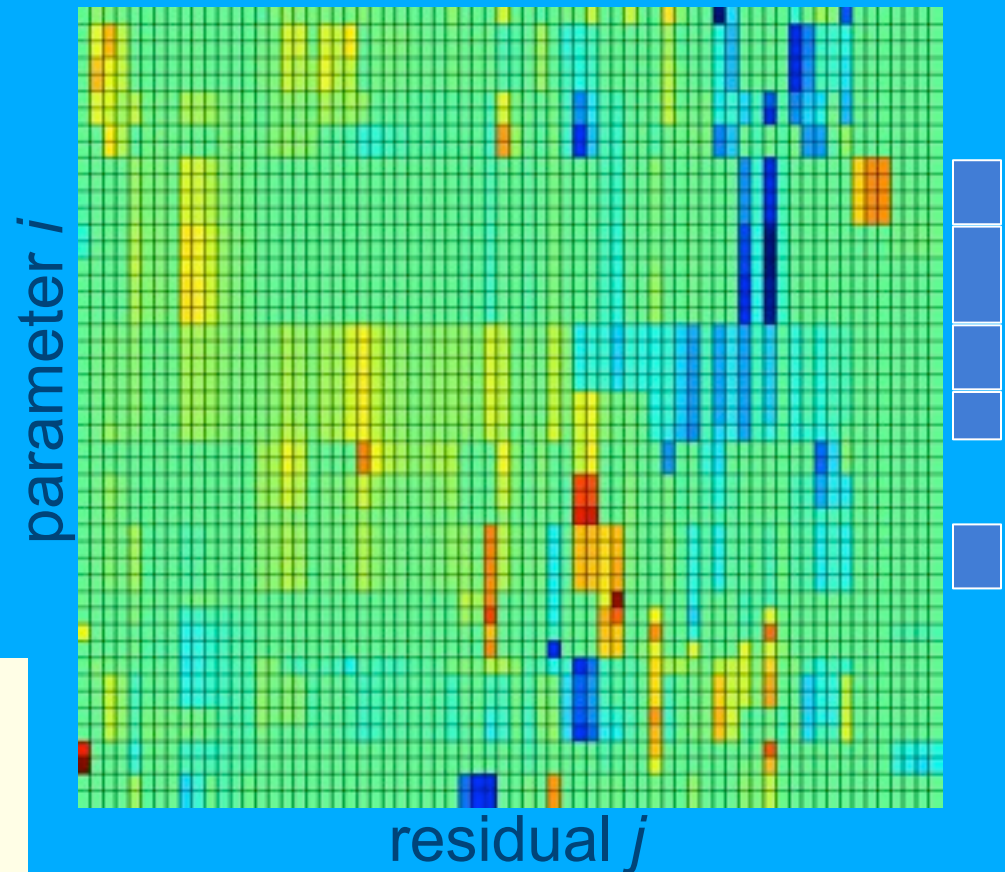
B. Finding sloppy subsystems

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. Nørskov, 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/>