Regression on Manifolds: Nonparametric system identification with applications in control and systems biology

Anil Aswani

Peter Bickel

Claire Tomlin

Joint work with Mark Biggin, Charless Fowlkes, Soile Keränen, and Jitendra Malik

Electrical Engineering and Computer Sciences, UC Berkeley

ACCESS Linnaeus Center, School of Electrical Engineering, KTH

April 2010

Experimental Platform: STARMAC

The **S**tanford **T**estbed of **A**utonomous **R**otorcraft for **M**ulti-**A**gent **C**ontrol

[Hoffmann, Waslander, Vitus, Huang, Gillula, Mercer, Bouffard, Li]

Case Study: Collision Avoidance

Pilots instructed to attempt to collide vehicles

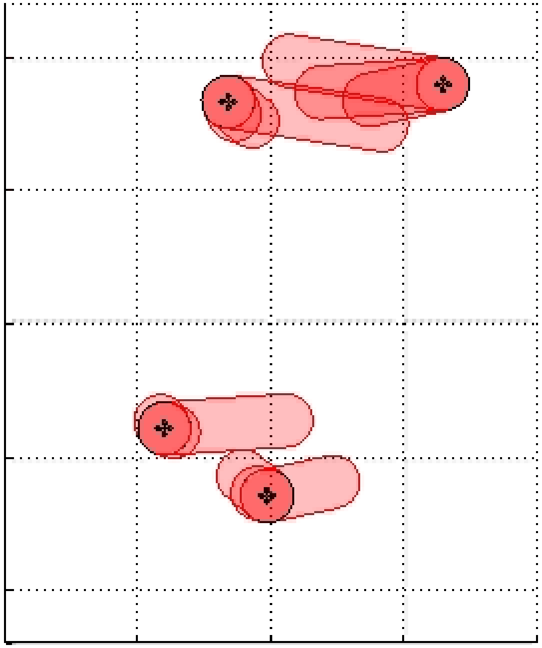
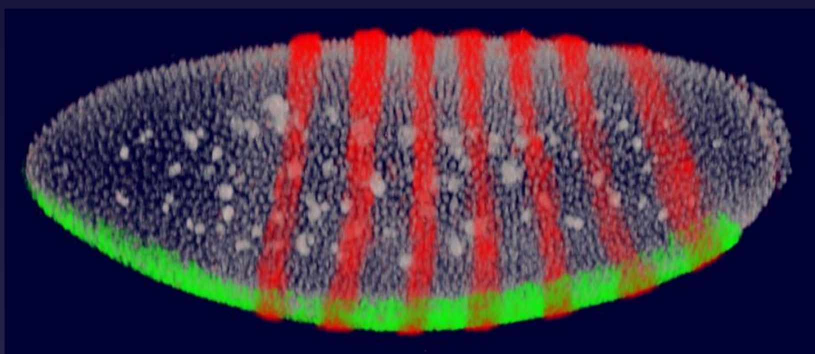
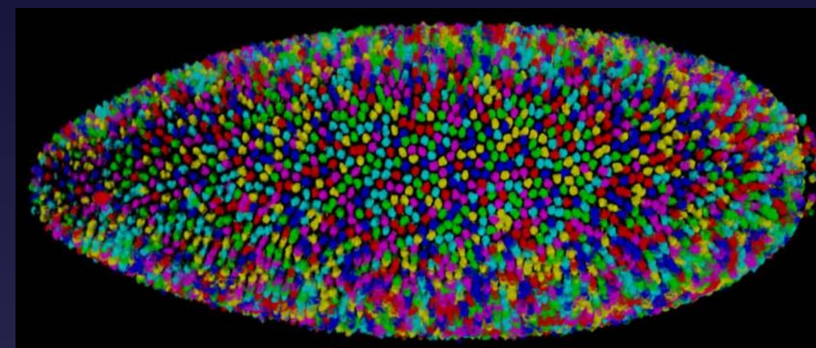


Image analysis can record 3D gene expression at cellular resolution

3D confocal images



3D segmentation mask



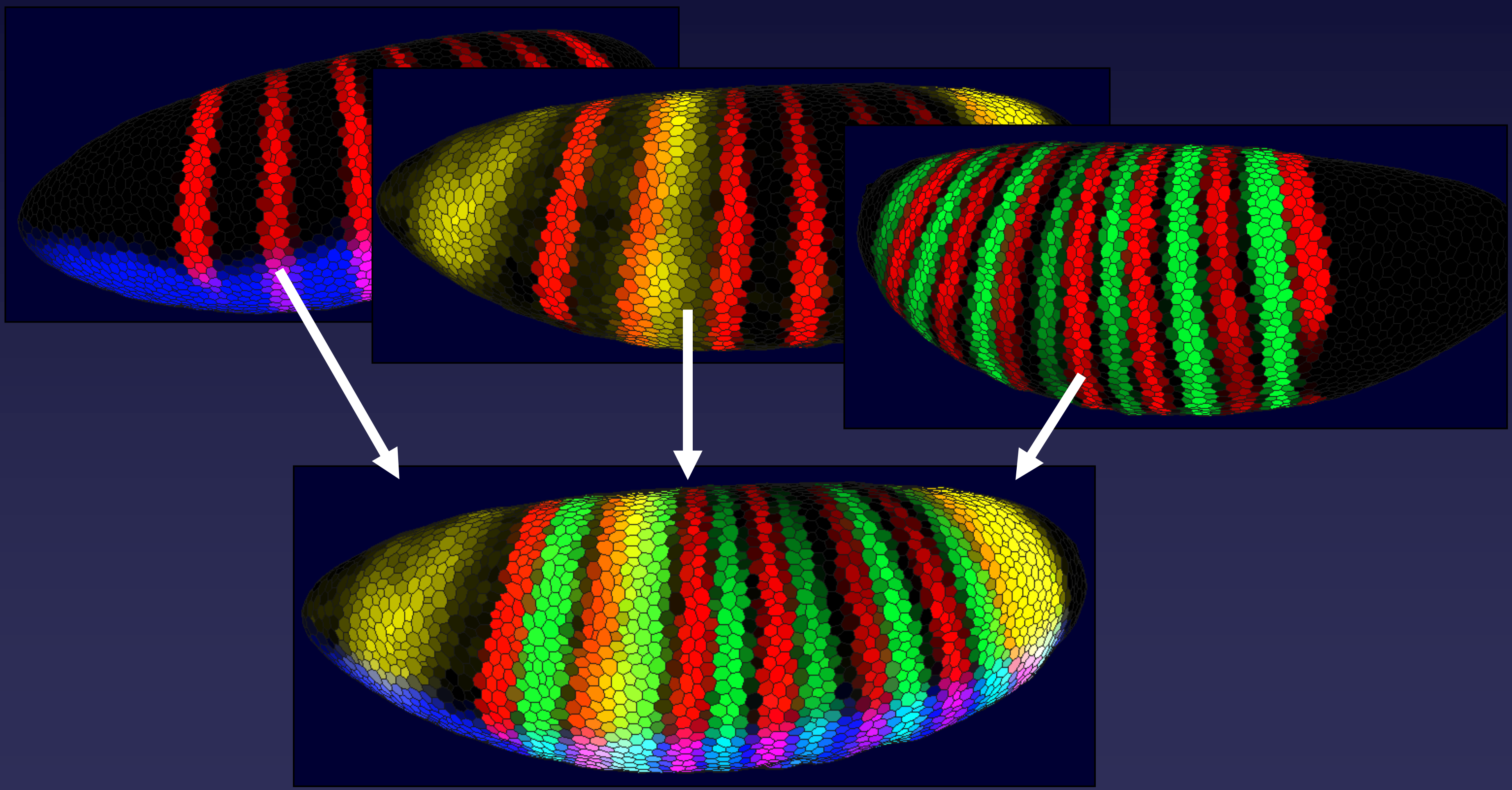
a “PointCloud” table

id	x	y	z	Nx	Ny	Nz	Vn	Vc	Sytox	Cy3_n	Cy3_a	Cy3_b	Cy3_g	Cou_n	Cou_a	Cou_b	Cou_g
1	102.36	142.14	112.00	-0.396	0.851	0.344	207.96	605.36	52.18	23.55	18.76	22.55	22.10	11.95	8.13	28.01	12.04
2	264.63	172.01	79.36	0.103	0.972	-0.208	281.73	599.90	82.12	31.67	34.97	15.95	31.93	21.06	12.56	41.40	19.12
3	225.91	174.99	88.65	-0.030	0.999	-0.015	185.79	418.35	85.32	35.63	31.27	14.77	34.00	19.59	20.53	38.80	21.35
4	318.42	48.34	138.91	0.095	-0.744	0.660	182.46	464.19	37.61	19.31	15.15	12.47	17.55	21.01	13.78	26.87	17.53
5	110.18	34.40	109.65	-0.186	-0.913	0.362	127.81	432.01	55.78	24.12	23.53	12.19	19.71	13.81	7.57	28.16	12.40
6	340.48	73.79	37.548	0.205	-0.299	-0.931	208.26	607.49	80.23	33.04	26.75	21.24	28.91	31.48	20.69	50.45	26.96

. . .

Luengo et al, 2006

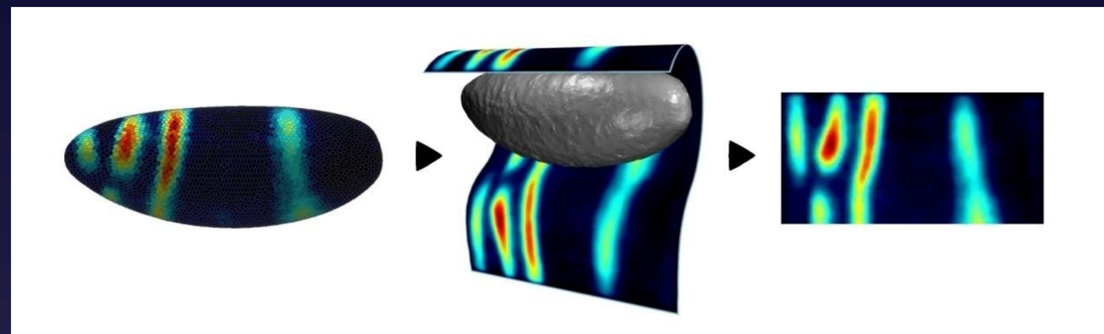
a 3D gene expression atlas

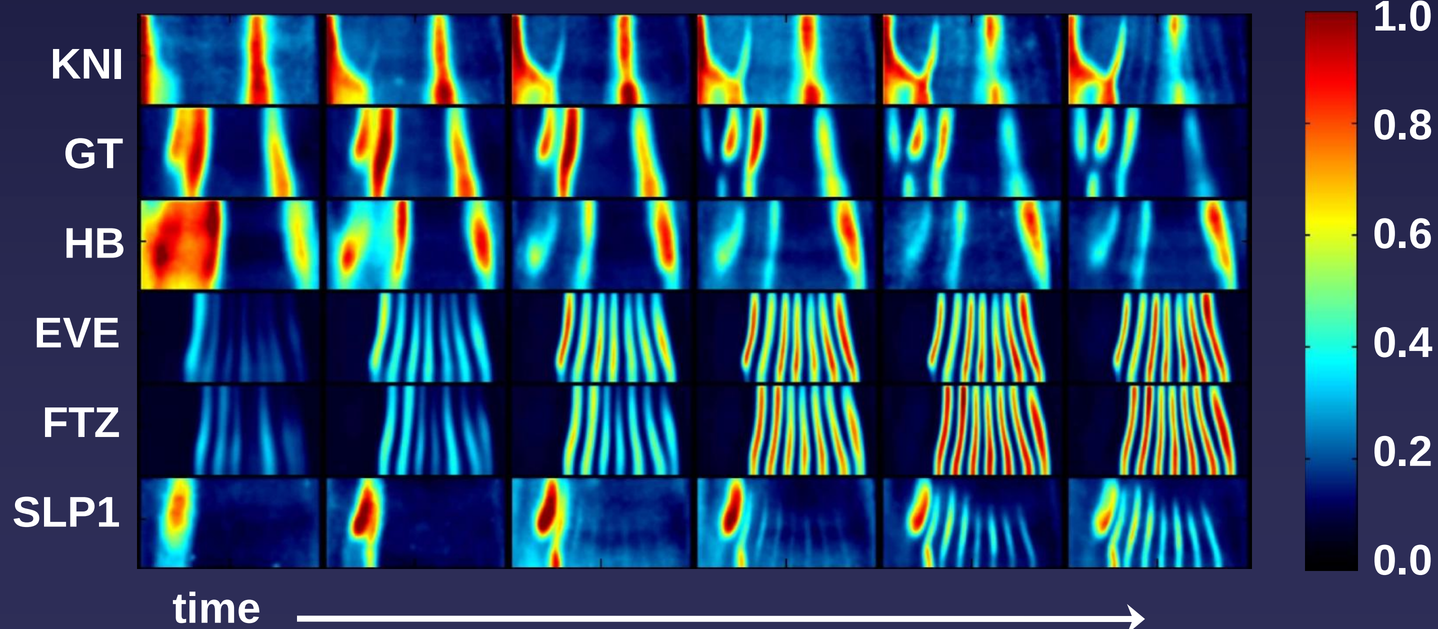


Fowlkes et al, 2008

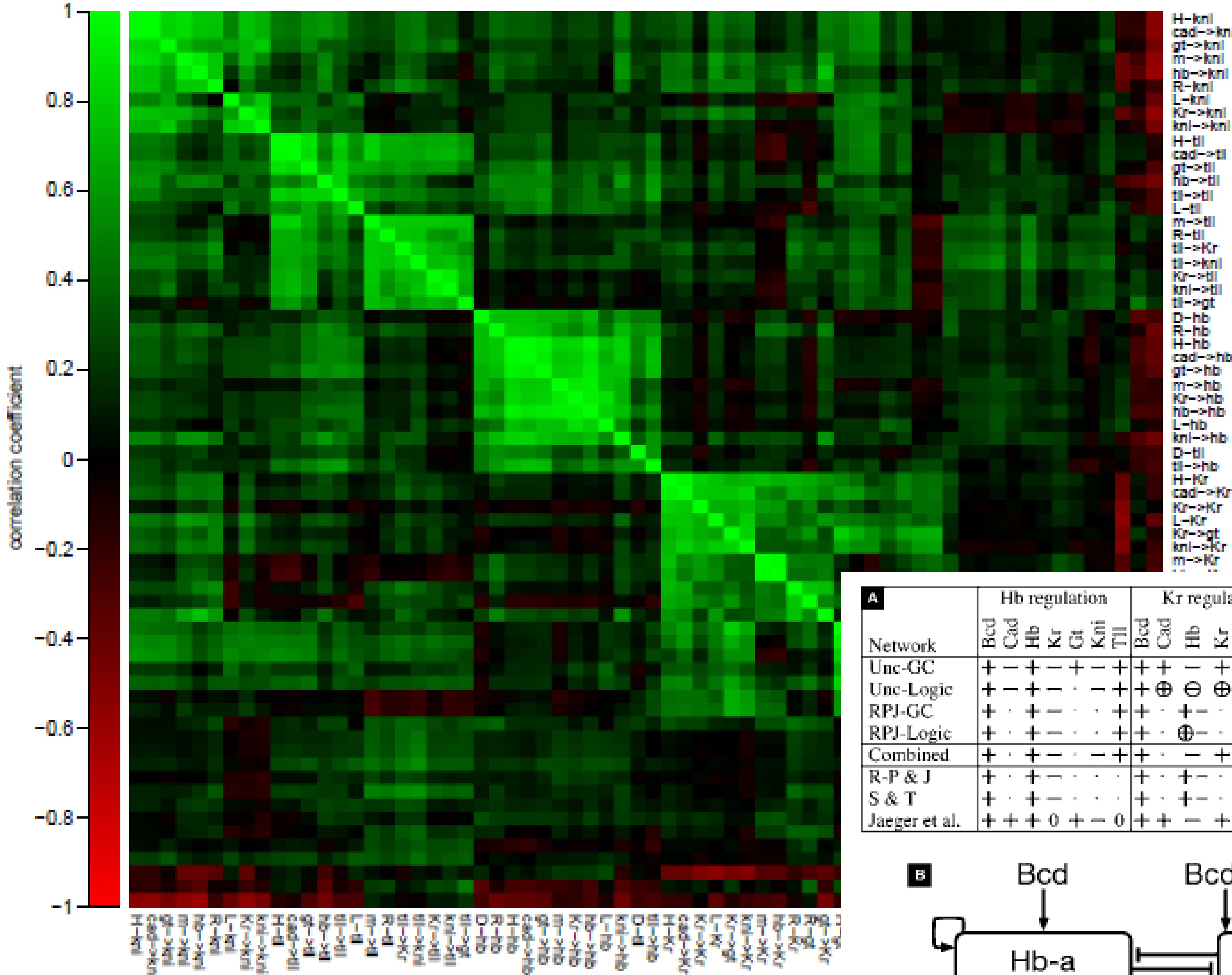
a 3D gene expression atlas

16 million cells
3,000 embryos
7 time points
protein 20 factors



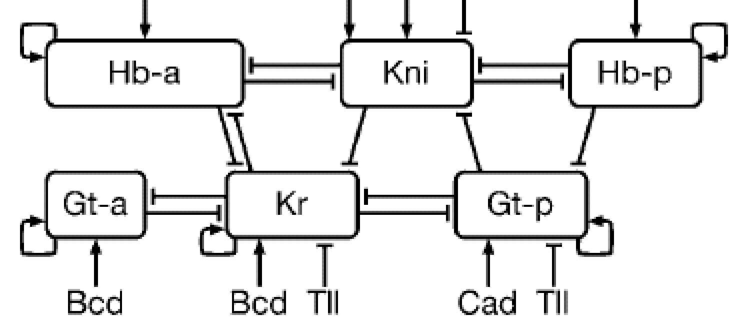


Fowlkes et al, 2008



A Network	Hb regulation							Kr regulation							Gt regulation							Kni regulation						
	Bcd	Cad	Hb	Kr	Gt	Kni	Tll	Bcd	Cad	Hb	Kr	Gt	Kni	Tll	Bcd	Cad	Hb	Kr	Gt	Kni	Tll	Bcd	Cad	Hb	Kr	Gt	Kni	Tll
Unc-GC	+	−	+	−	+	−	+	+	+	−	+	−	−	−	+	+	−	−	+	+	−	+	+	−	−	−	+	−
Unc-Logic	+	−	+	−	·	−	+	+	⊕	⊖	⊕	−	−	⊖	+	+	⊖	−	+	·	⊖	+	⊕	−	⊖	−	⊕	⊖
RPJ-GC	+	·	+	−	·	·	+	+	·	+−	·	−	−	−	+	+	·	−	·	⊖	−	⊕	+	−	⊕	−	·	−
RPJ-Logic	+	·	+	−	·	·	+	+	·	⊕−	·	−	−	−	+	+	·	−	·	⊖	−	+	⊕	−	⊕	⊖	·	⊖
Combined	+	·	+	−	·	−	+	+	·	−	+	−	−	−	+	+	−	−	+	·	−	+	+	−	·	−	·	−
R-P & J	+	·	+	−	·	·	·	+	·	+−	·	−	−	−	+	+	·	−	·	−	−	+	+	−	+	−	·	−
S & T	+	·	+	−	·	·	·	+	·	+−	·	−	−	·	+	+	−	−	·	·	·	+	+	−	·	−	·	·
Jaeger et al.	+	+	+	0	+	−	0	+	+	−	+	−	−	−	+	+	−	−	+	0	−	+	+	−	−	−	+	−

B

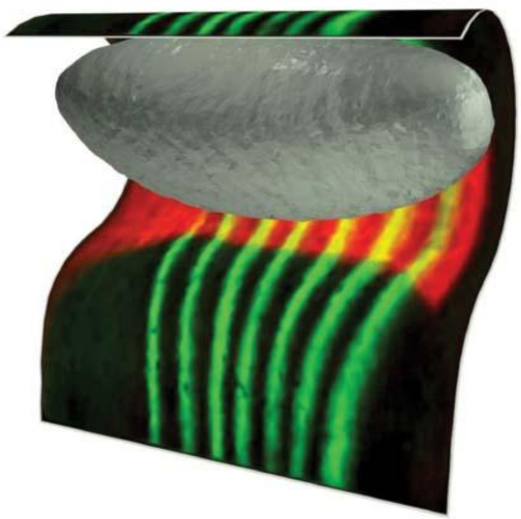


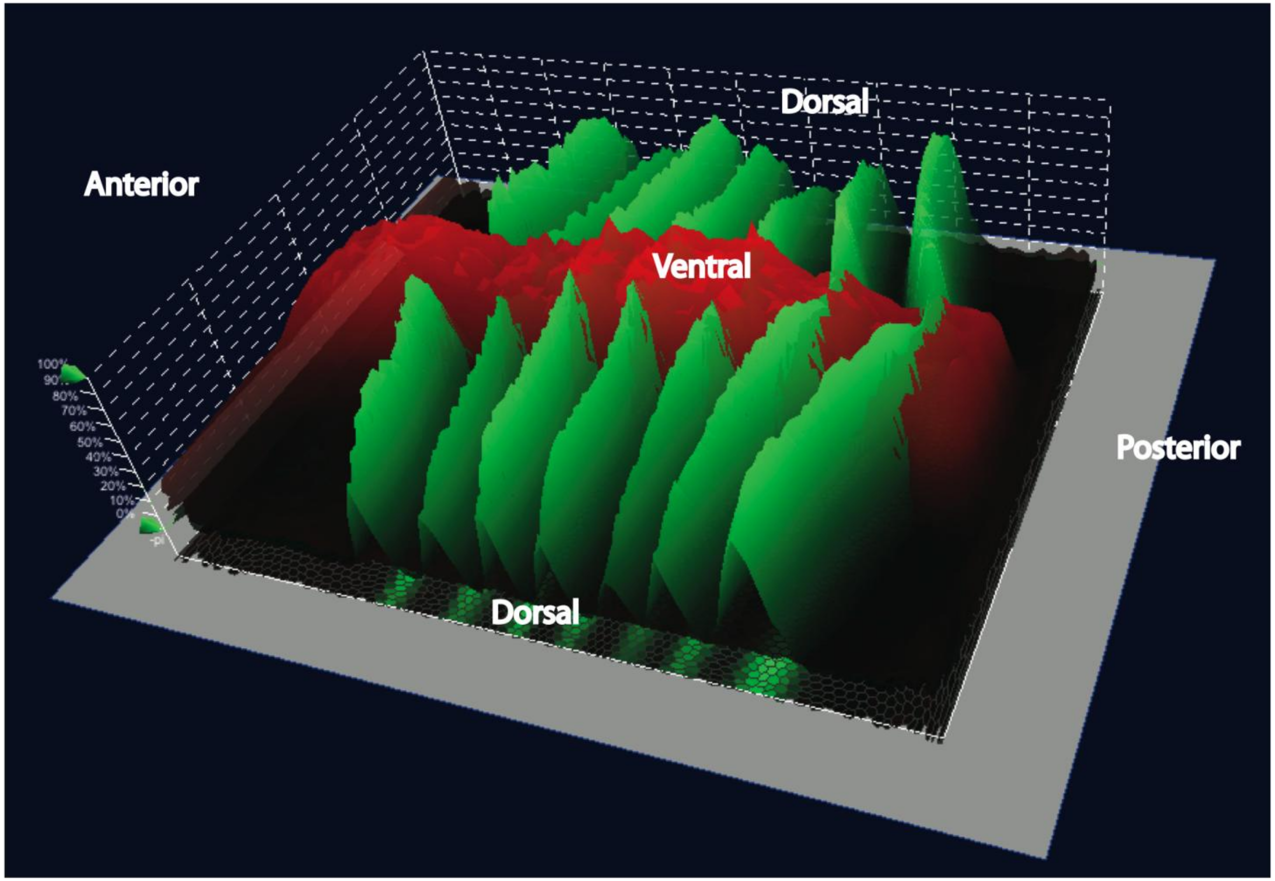
[Fomekong-Nanfack et al 2009, Perkins et al 2006]

Quantitative changes in expression are evident along both axes for almost all genes

visualizing expression along both axes

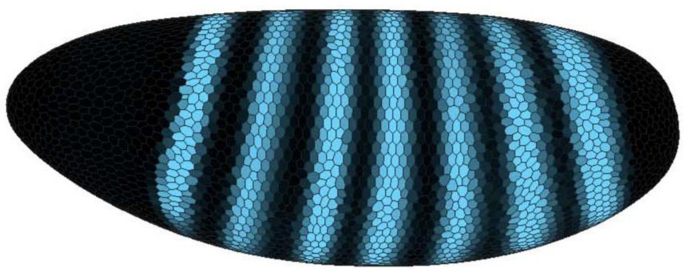
cylindrical projection



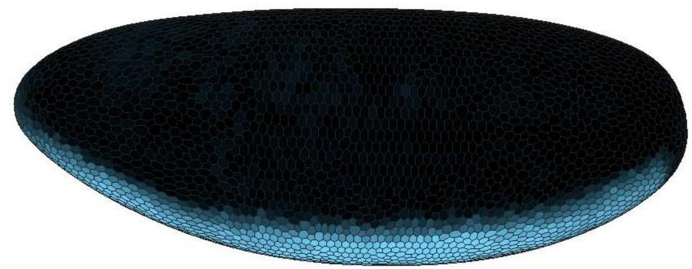


Quantitative changes in expression are evident along both axes for almost all genes

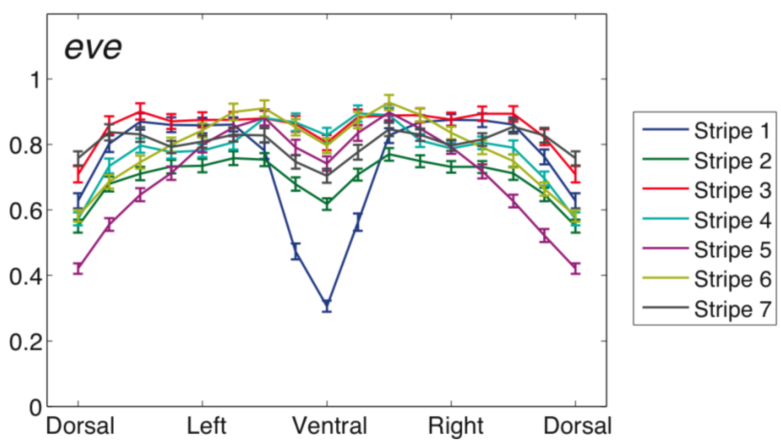
eve



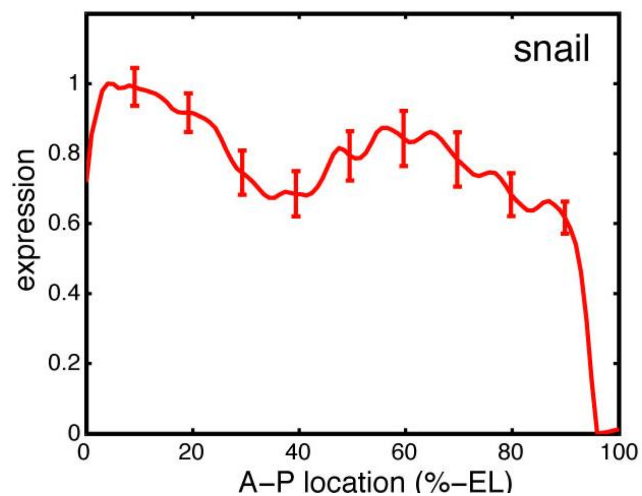
sna



eve expression along D/V axis

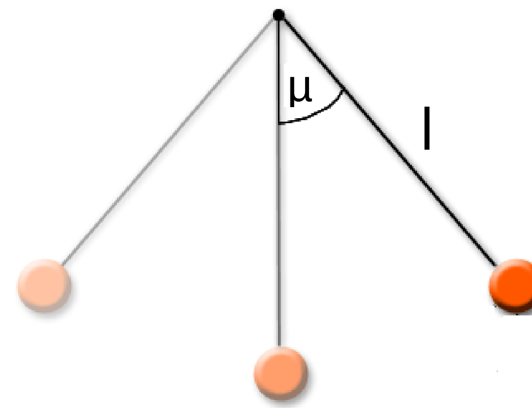


sna expression along A/P axis



Toy Example: Pendulum

- Dynamics: $\overset{\text{Ä}}{\mu} = ¡\ \frac{g}{l} \sin \mu$



$$x = \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} \theta \\ \dot{\theta} \end{bmatrix} \quad \rightarrow \quad \dot{x} = f(x)$$

$$f(x) = \begin{bmatrix} x_2 \\ -\frac{g}{l} \sin x_1 \end{bmatrix}$$

- Write this as: $\dot{x} = \beta X$

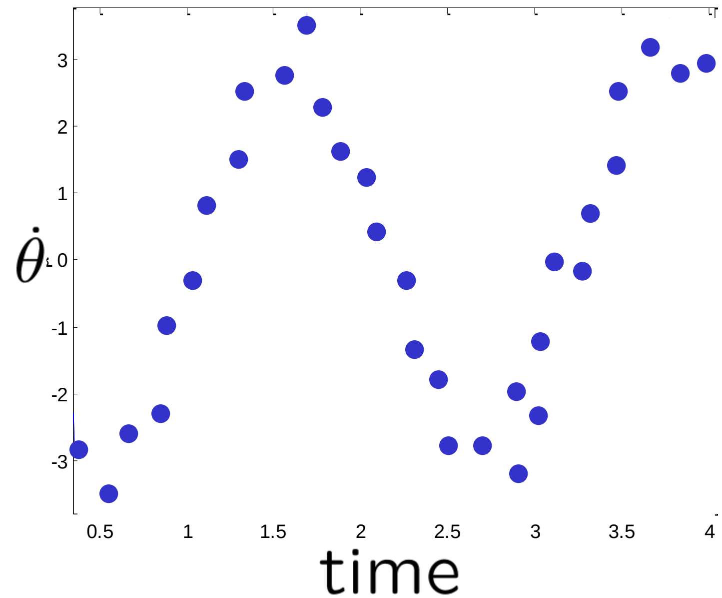
$$\beta = \begin{bmatrix} 0 & 1 \\ -\frac{g}{l} & 0 \end{bmatrix} \qquad X = \begin{bmatrix} \sin x_1 \\ x_2 \end{bmatrix}$$

[Jeremy Gillula]

Pendulum

Suppose:

- model is unknown
- noisy measurements are available of velocity



Identify a model $\dot{x} = \beta X$

where

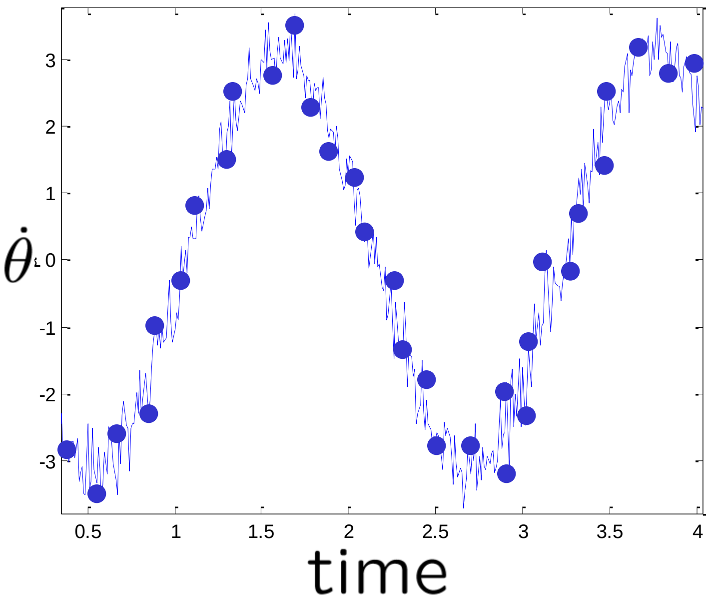
$$X = \left[\theta \quad \dot{\theta} \quad \sin\theta \quad \sin\dot{\theta} \quad \cos\theta \quad \cos\dot{\theta} \quad \theta^2 \quad \dot{\theta}^2 \quad \ldots\right]^T$$

and β is unknown

[Jeremy Gillula]

Pendulum

Learned Result:

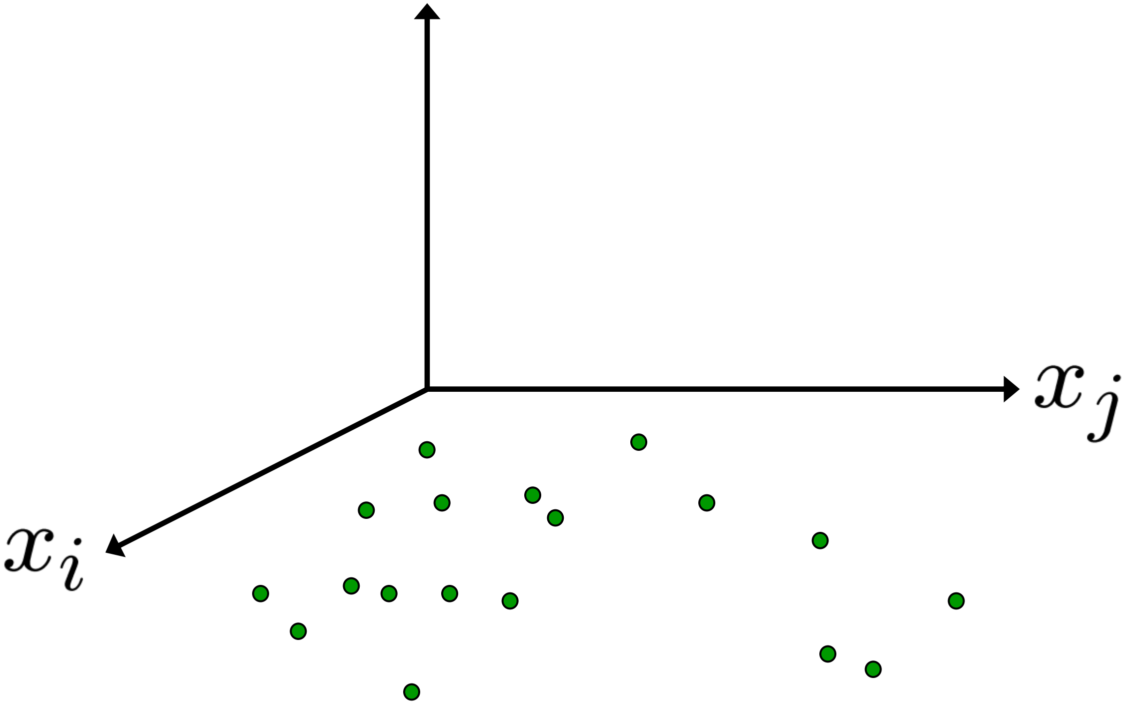



```
x1_dot = 0.23*x1 + 0.96*x2 + -0.27*sin(x1) +
  -0.02*cos(x1) + 0.15*sin(x2) + -0.01*cos(x2)
x2_dot = 3.66*x1 + 0.00*x2 + -13.92*sin(x1) +
  0.01*cos(x1) + -0.00*sin(x2) + 0.01*cos(x2)
```

→ **Important to prevent overfitting**

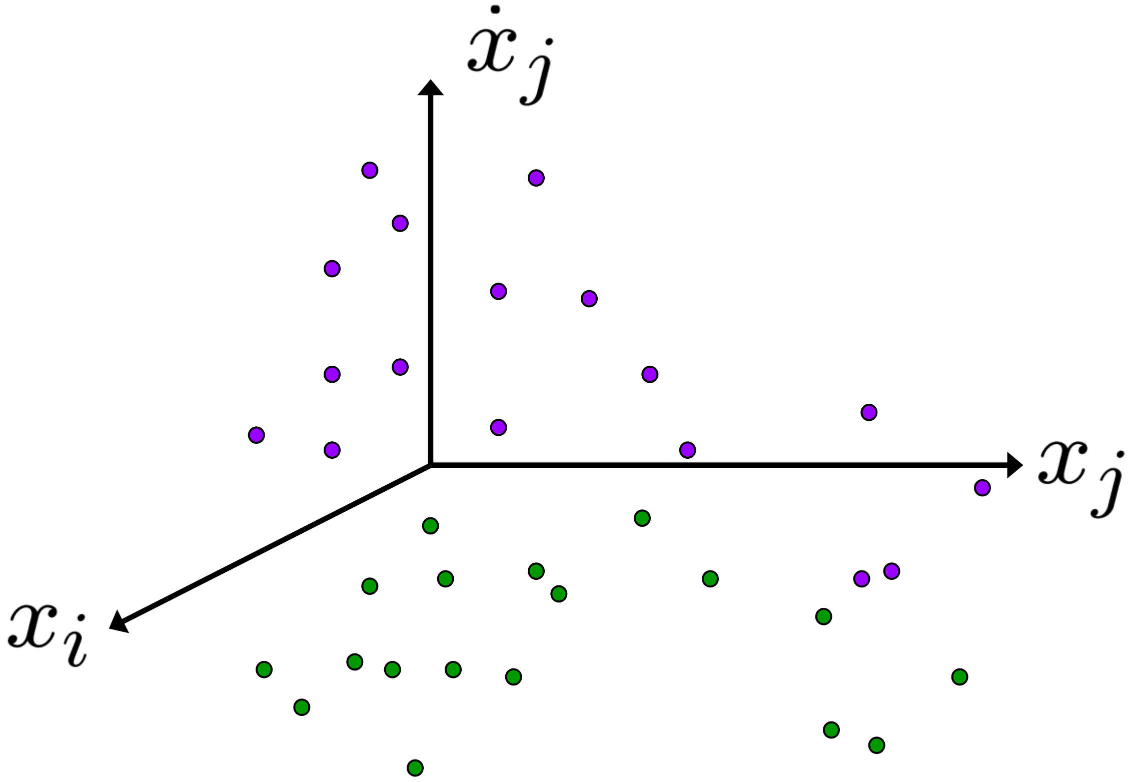
[Jeremy Gillula]

Online System Identification



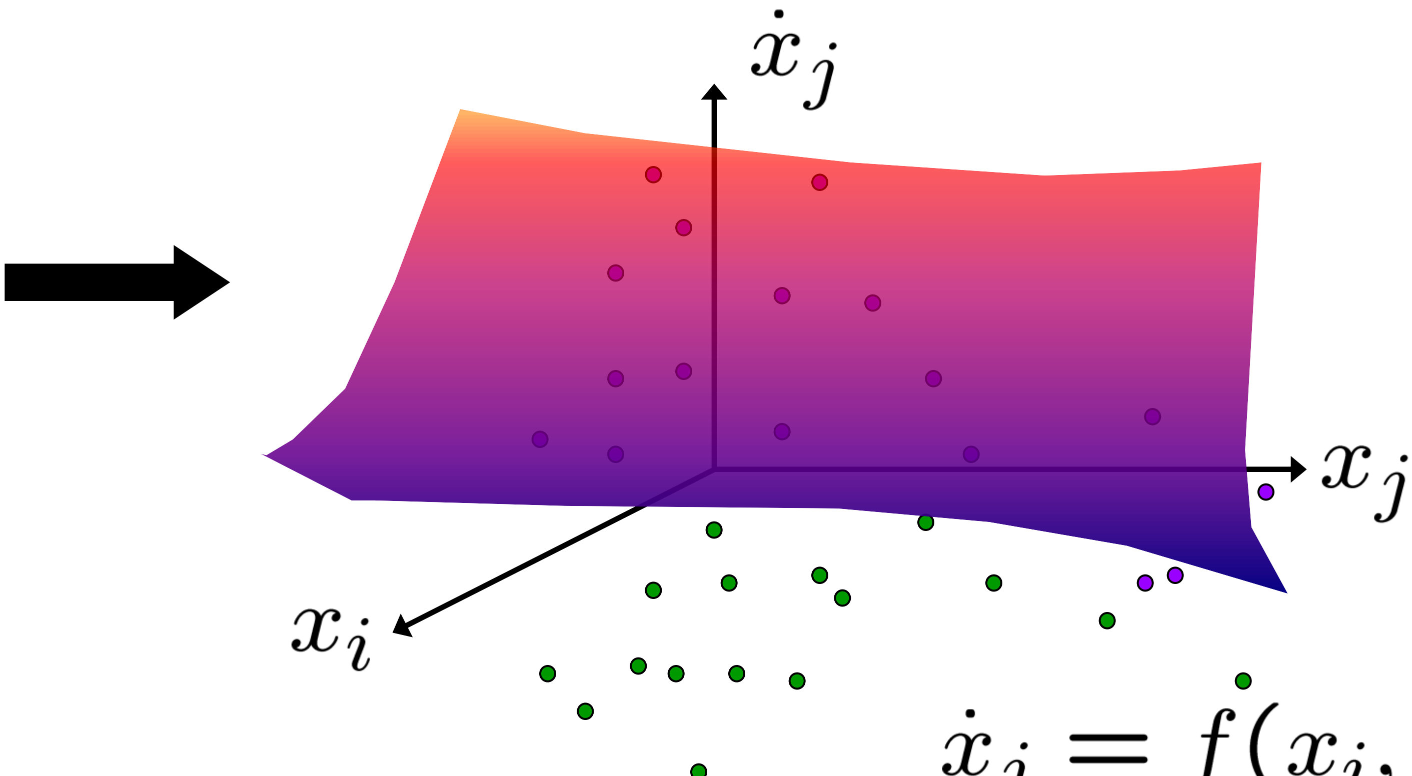
[Kloetzer and Belta; Ma, Vidal, and Sastry; Soatto; Vijayakumar; Atkeson; Ting; Hunt…]

Online System Identification



[Kloetzer and Belta; Ma, Vidal, and Sastry; Soatto; Vijayakumar; Atkeson; Ting; Hunt…]

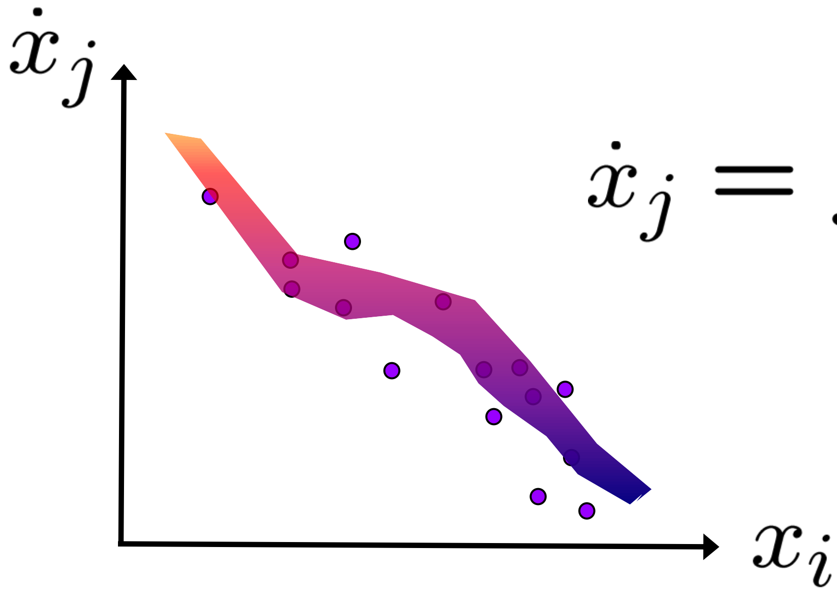
Online System Identification



$$\dot{x}_j = f(x_i, x_j)$$

[Kloetzer and Belta; Ma, Vidal, and Sastry; Soatto; Vijayakumar; Atkeson; Ting; Hunt…]

Online System Identification



$$\dot{x}_j = f(x_i, x_j)$$

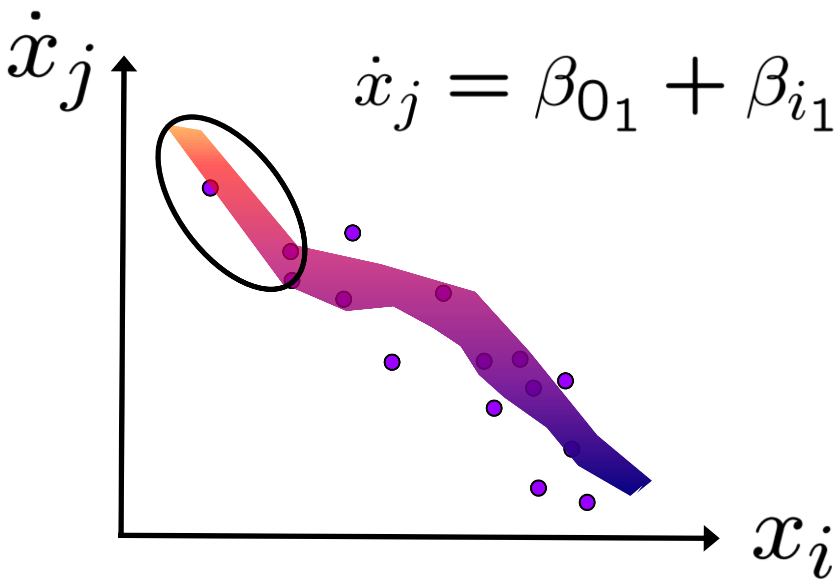
- Undersampling for high-dimensional systems
- Constrained dynamics
- Fast-slow dynamics

[Bickel and Li, 2007]

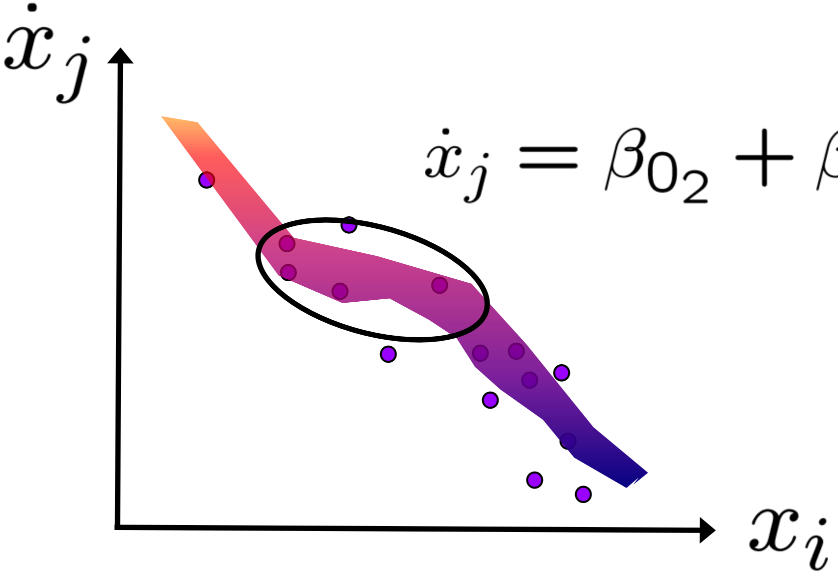
Online System Identification

$$\dot{x}_j = \beta_{0_1} + \beta_{i_1} x_i + \beta_{j_1} x_j$$

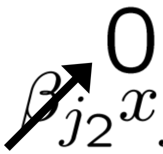
$\beta_{j_1} \rightarrow 0$



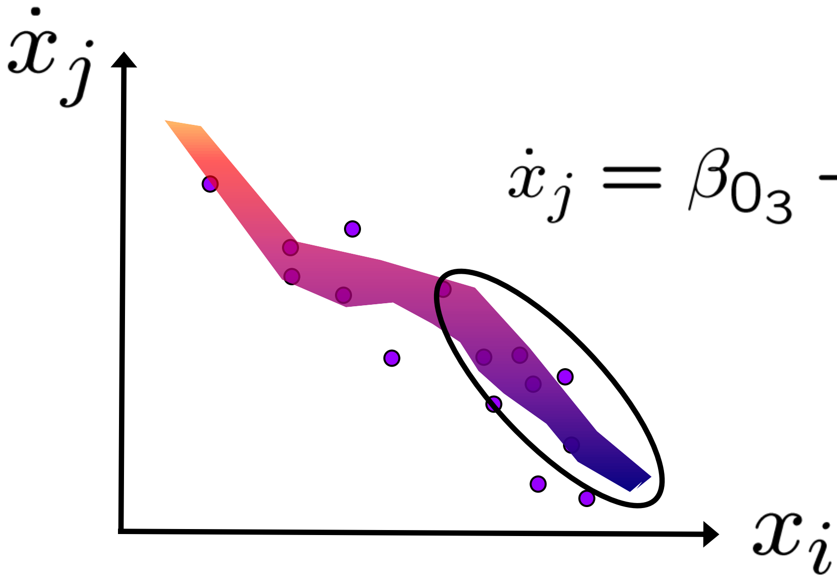
Online System Identification



$$\dot{x}_j = \beta_{0_2} + \beta_{i_2} x_i + \beta_{j_2} x_j$$



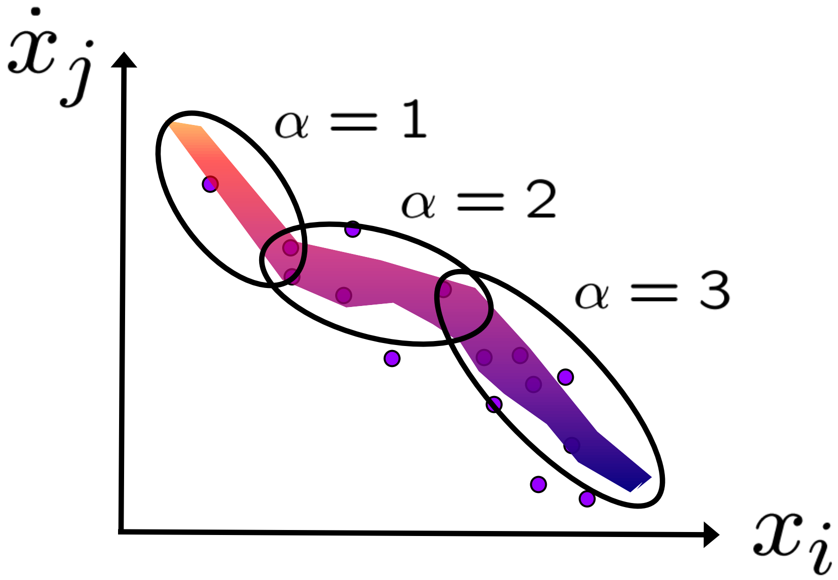
Online System Identification



0

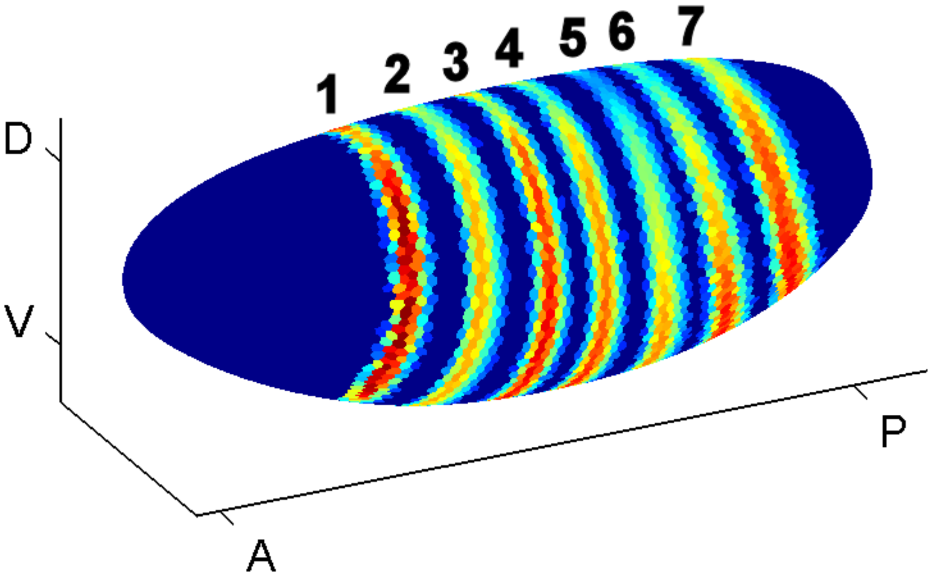
$$\dot{x}_j = \beta_{0_3} + \beta_{i_3} x_i + \beta_{j_3} x_j$$

Online System Identification

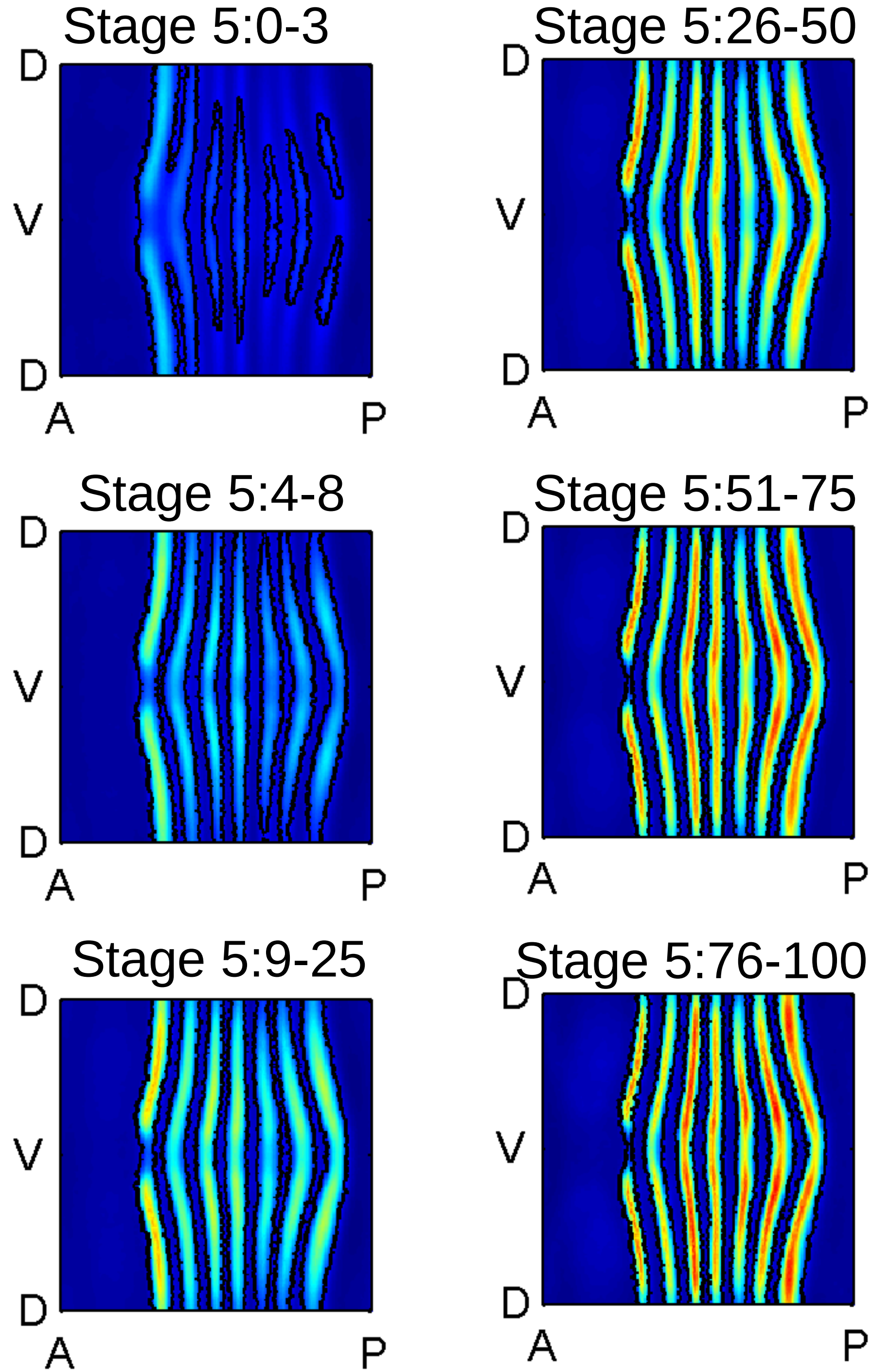


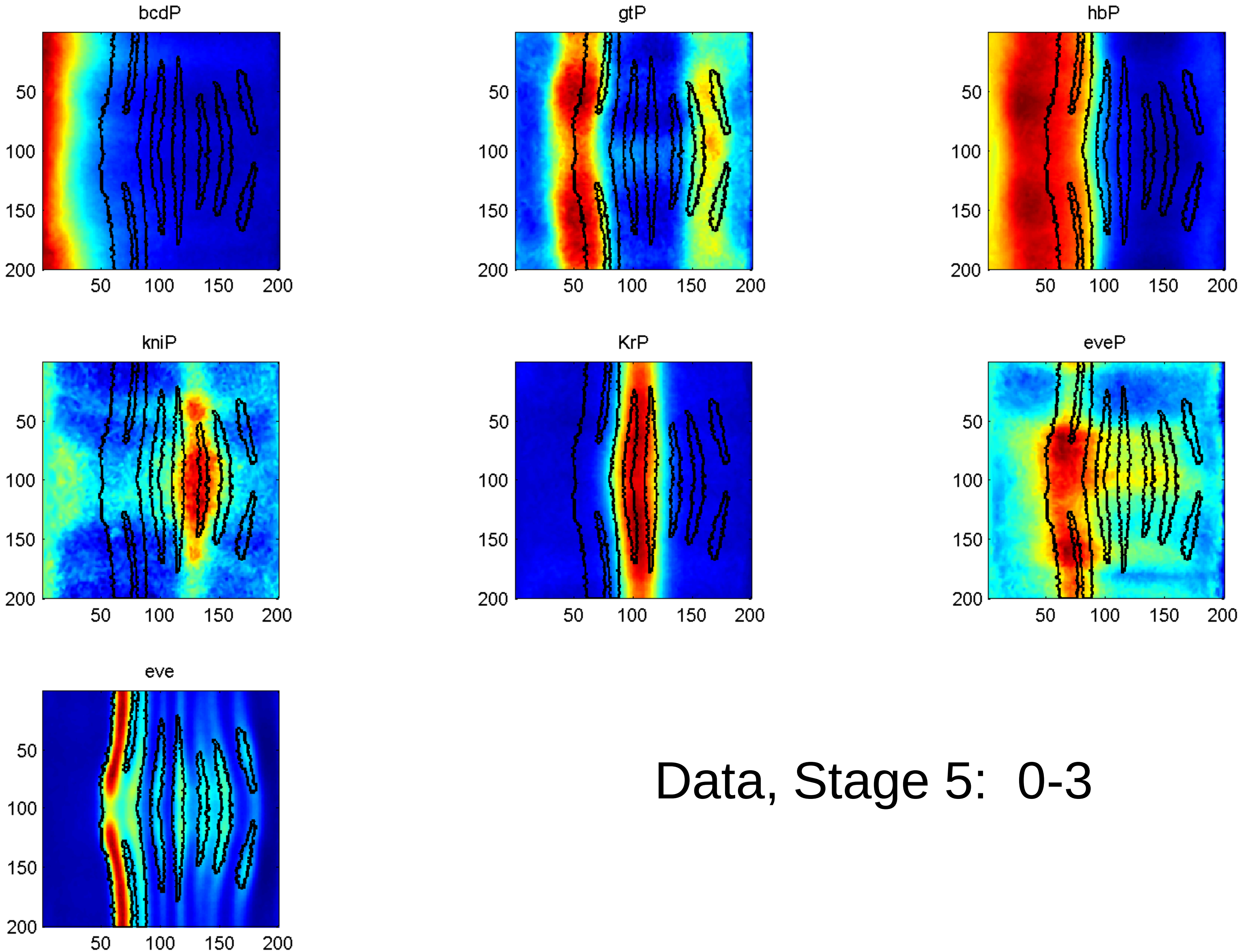
$$\dot{x} = A_\alpha x + b_\alpha$$

→ **Look for a geometric structure for sparsity**
Local linear models are easy to manipulate

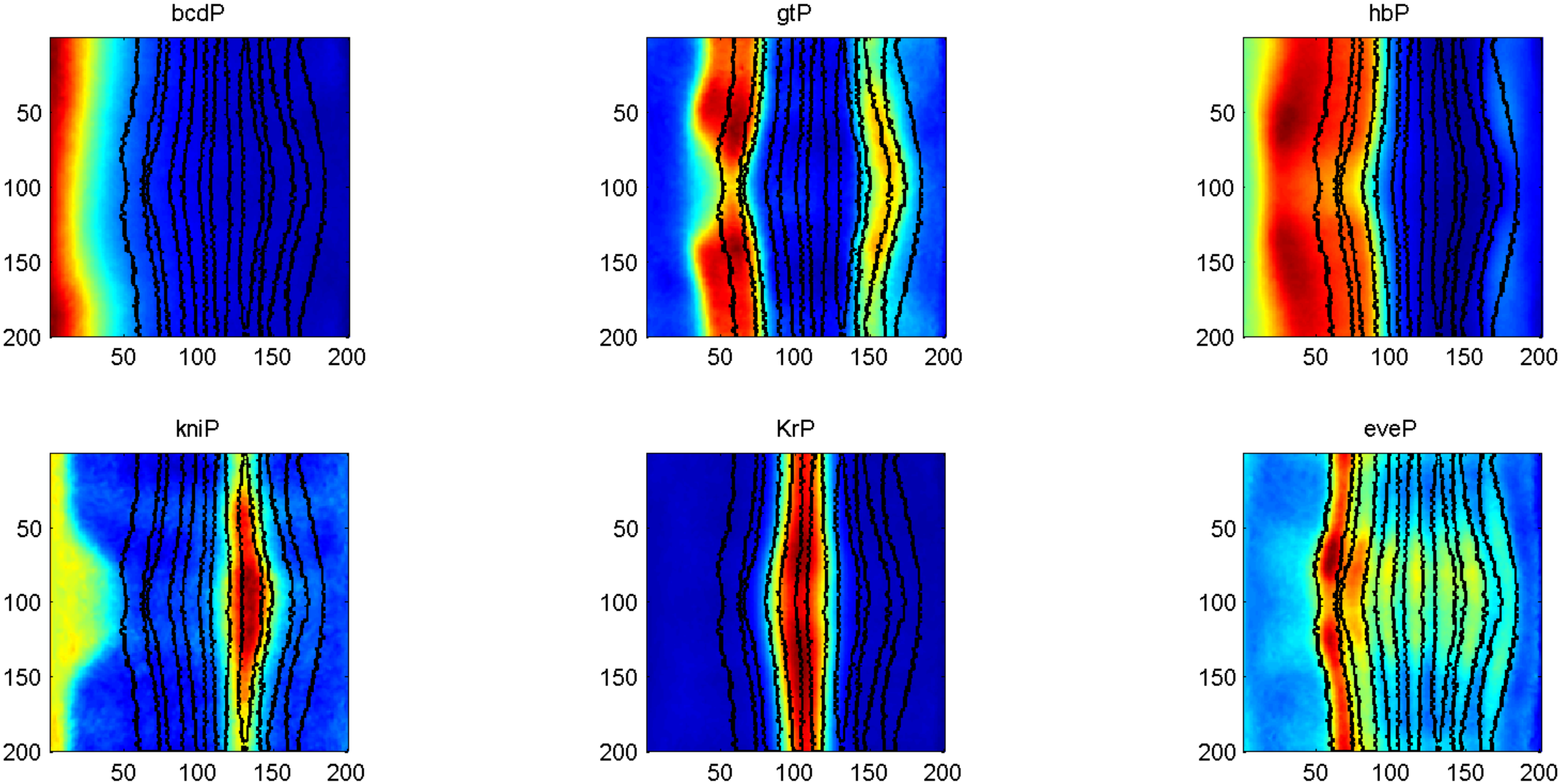


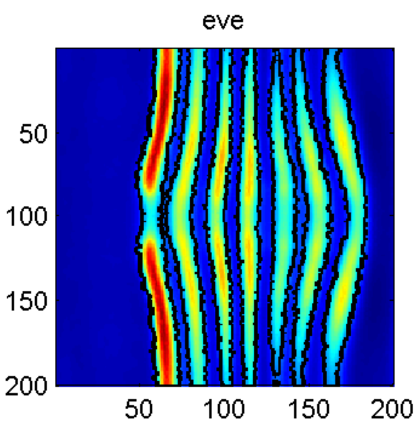
eve mRNA data shown in 3D and 2D projections



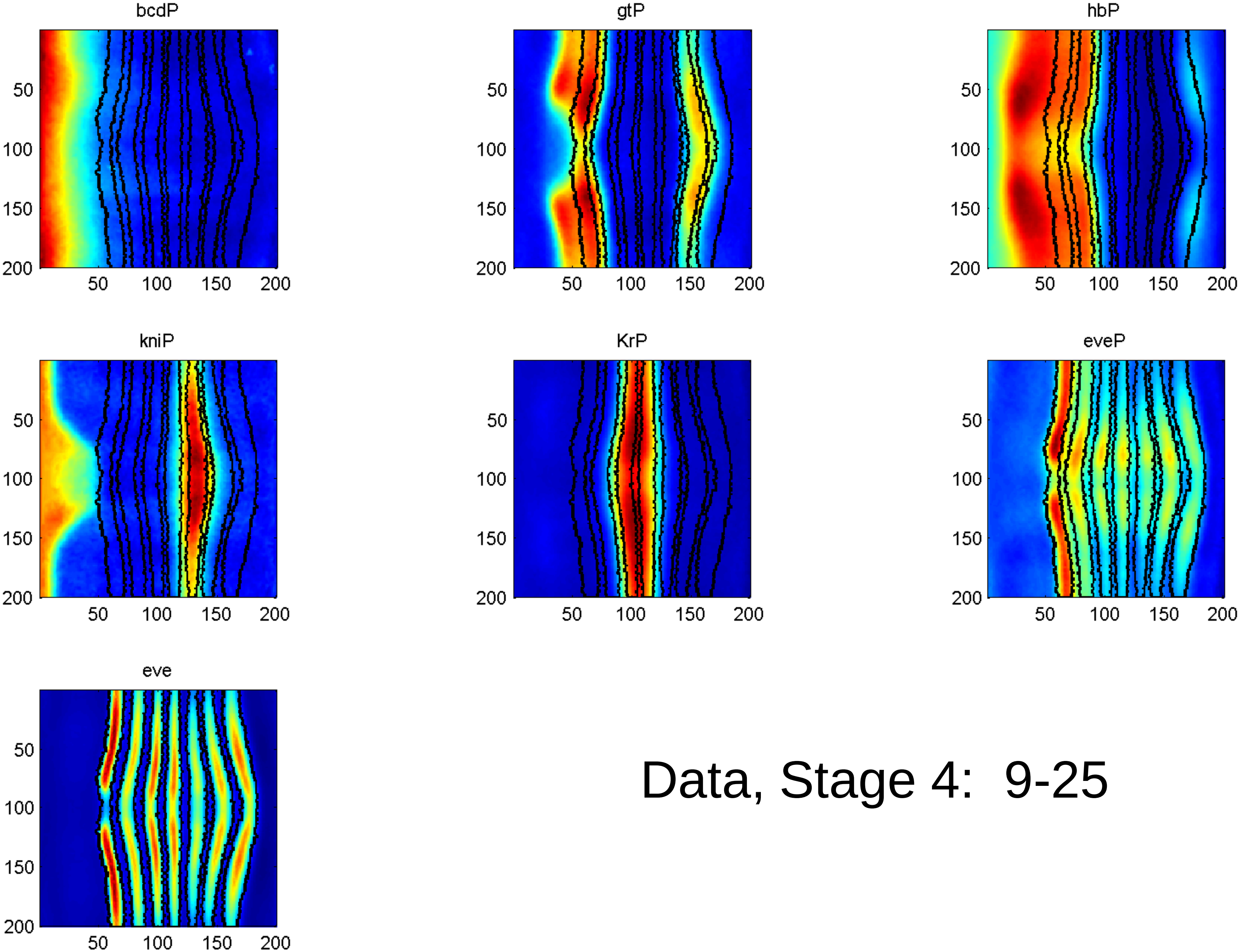


Data, Stage 5: 0-3





Data, Stage 5: 4-8



Data, Stage 4: 9-25

Local Linear Regression

Solve for (A_α, b_α) in $\dot{x} = A_\alpha x + b_\alpha$ for all α

Rewrite as: $Y = \beta X$

where

$$Y^T = [\dot{x}_1(t_1) \ldots \dot{x}_E(t_T)]$$

$$\beta = [A \quad b]$$

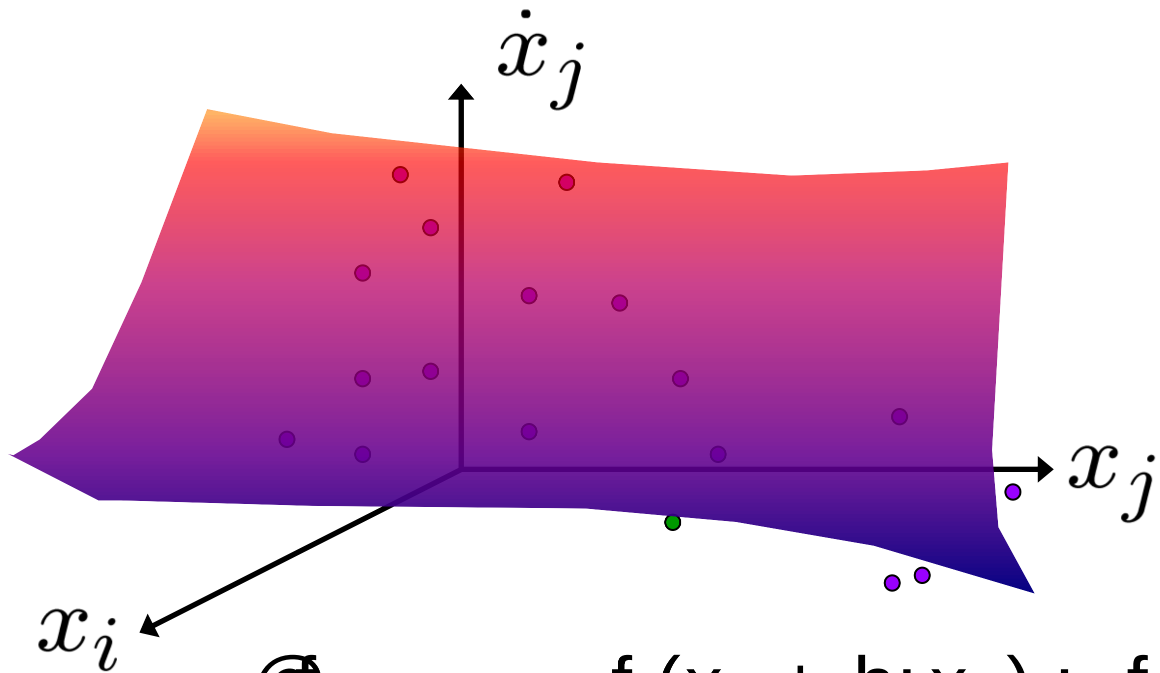
$$X^T = [x_1(t_1) \ldots x_E(t_T) \quad 1]$$

Existing Approaches

Estimator	**Considers geometry**	**Sparsity**	**High-dimensionality**
Moore-Penrose[1]	Yes		
Ridge[2]			
Principal Components Regression[3]	Yes		
Lasso[4,5]		Yes	Yes
Elastic Net[6]		Yes	Yes
Partial Least Squares[7]	Yes		

[1] (Knight and Fu, 2000); [2](Hoerl and Kannard, 1970); [3](Massy, 1965); [4](Tibshirani, 1996); [5](Zou, 2006); [6](Zou and Hastie, 2005); [7](Wold, 1975)

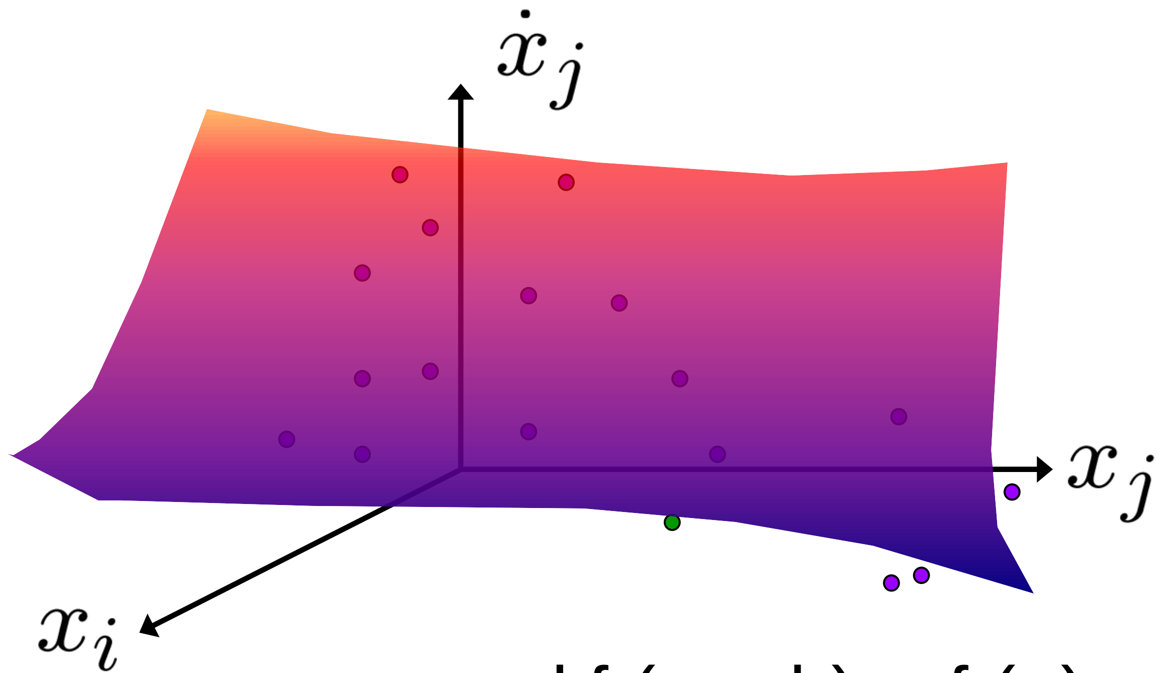
Online System Identification



$$\frac{\partial f}{\partial x_1} = \lim_{h \to 0} \frac{f(x_1 + h; x_2) - f(x_1; x_2)}{h}$$

- Difficulty in interpreting regression coefficients
- Gradient of function does not exist

Online System Identification

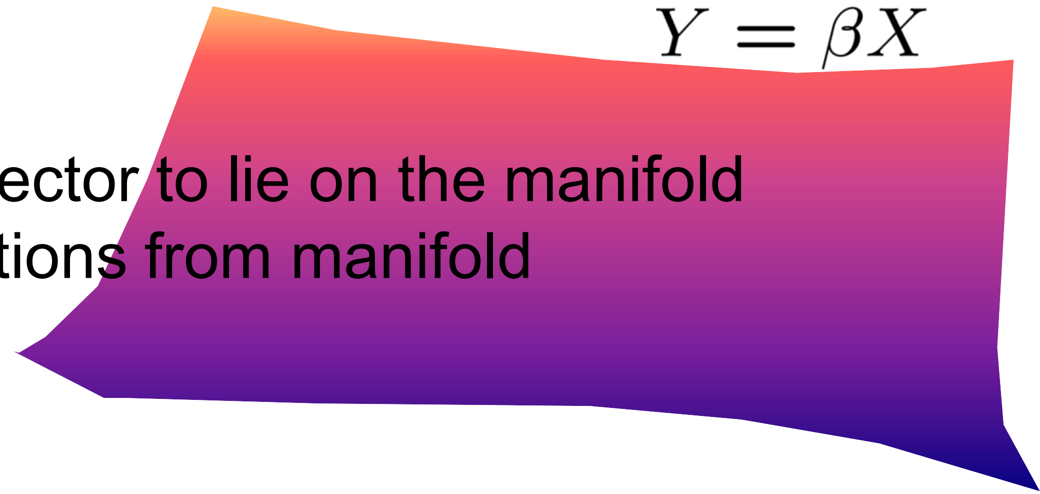


$$df = A : \lim_{\substack{\|h\| \to 0 \\ x+h \in M}} \frac{\|f(x+h) - f(x) - Ah\|}{\|h\|} = 0$$

Exterior derivative of function does exist

- Extension of gradients to manifolds
- Best local linear approximation of function on manifold

New Estimation Approach



- Locally learn manifold
- Constrain regression vector to lie on the manifold by penalizing for deviations from manifold

$$\hat{\beta} = \arg\min_{\beta}(||W(Y - \beta X)||_2^2 + \lambda||\Pi\beta||_2^2)$$

- Where Π is chosen to penalize β for lying off of the manifold

(Aswani , Bickel, Tomlin, 2010); (Bickel and Levina, 2008)

Correlation over space and time

eve mRNA expression

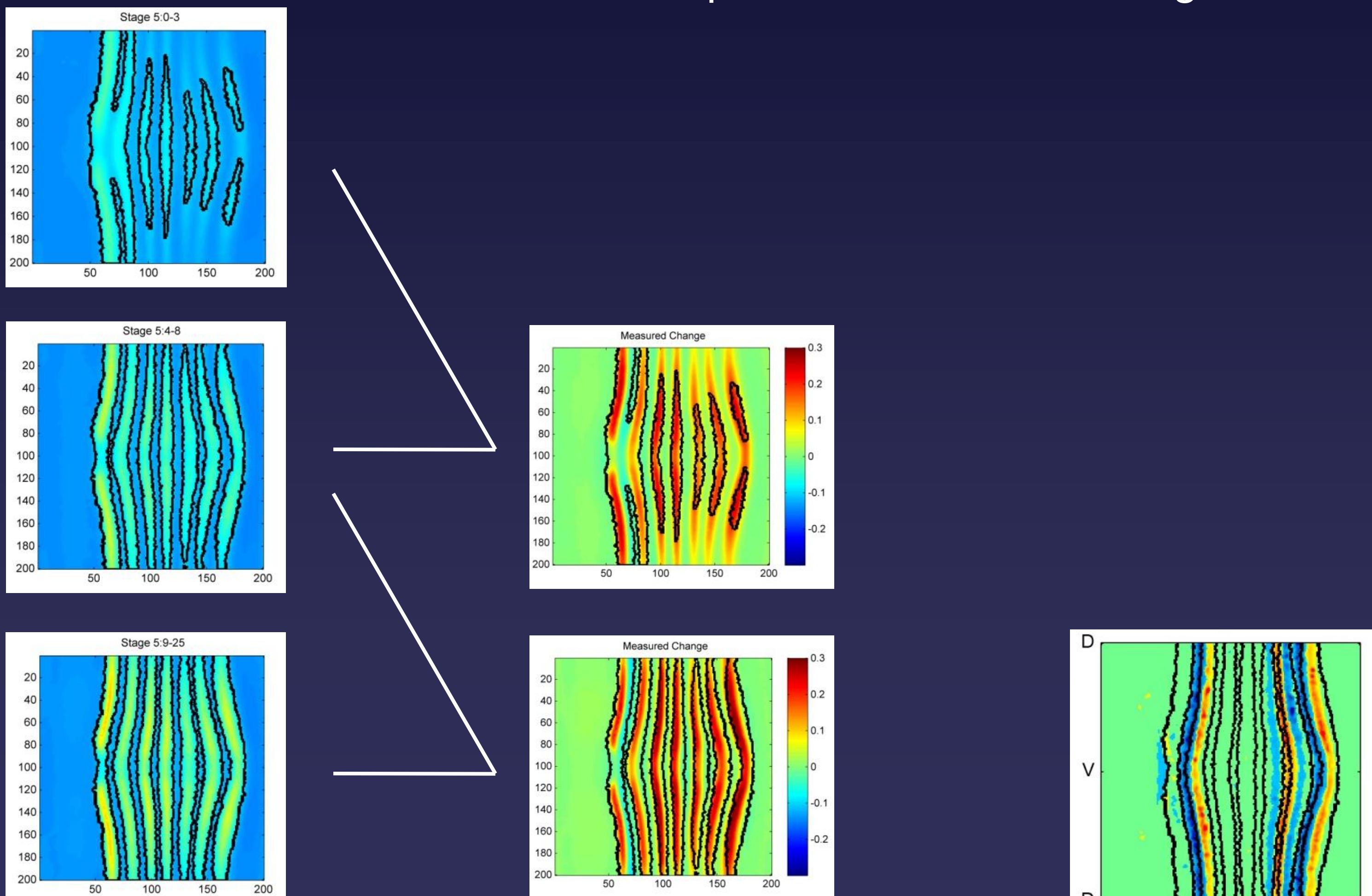
temporal change in mRNA expression

correlation of gt protein with change in eve mRNA

t=0

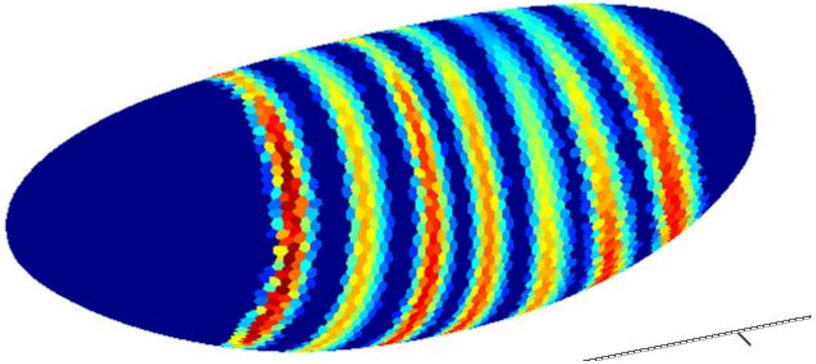
t=1

t=2



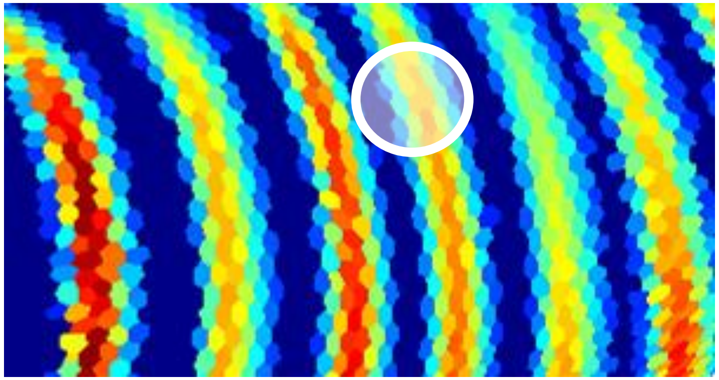
Drosophila Embryo, Stage 5

$$\frac{d[eve]}{dt} = f(bcdP, gtP, hbP, krP, kniP, eveP, eve)$$

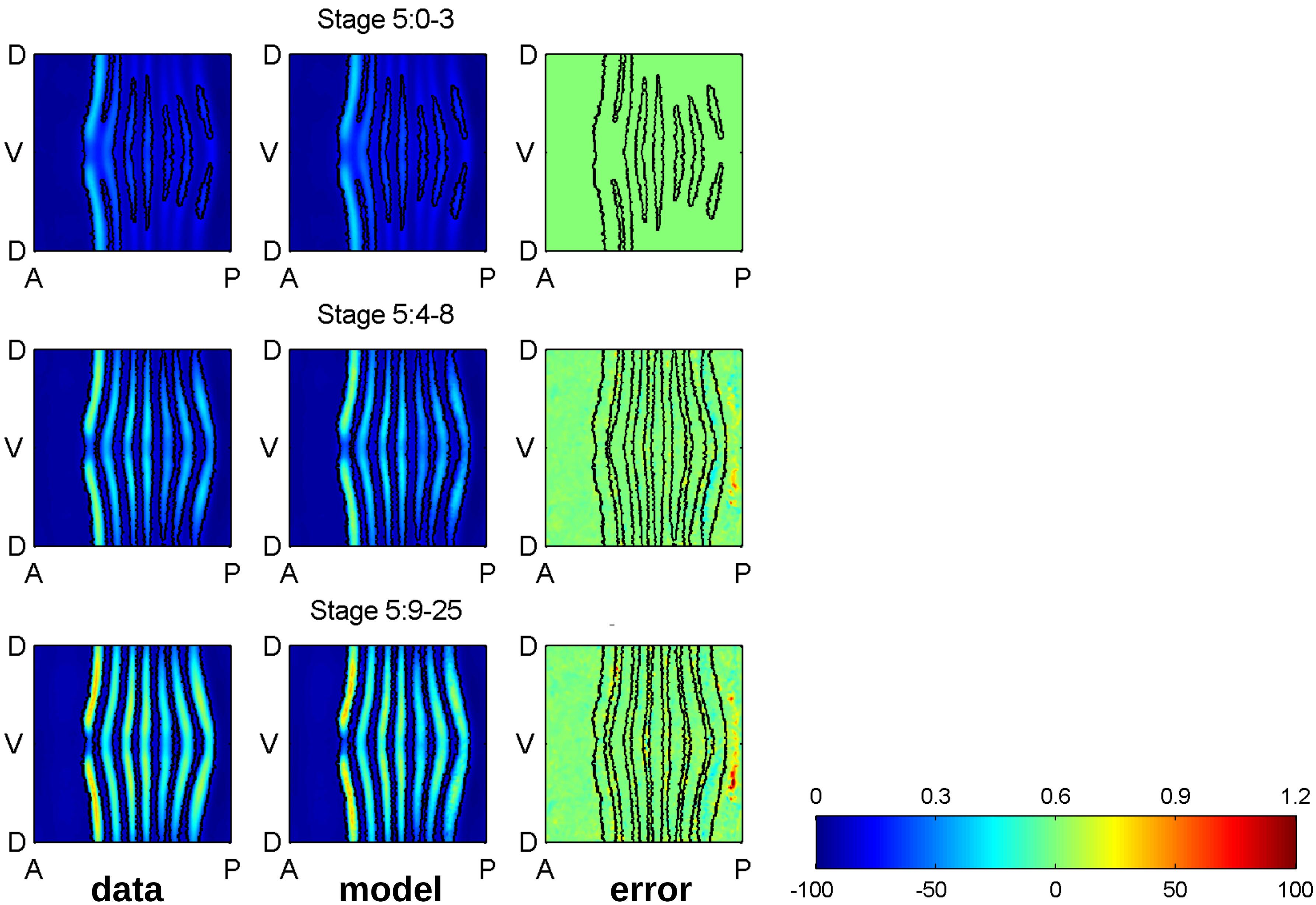


$$\frac{d[eve]}{dt} = \beta_0 + \beta_1[bcdP] + \beta_2[gtP] + \beta_3[hbP] + \beta_4[krP] + \ldots + \beta_5[kniP] + \beta_6[eveP] + \beta_7[eve]$$

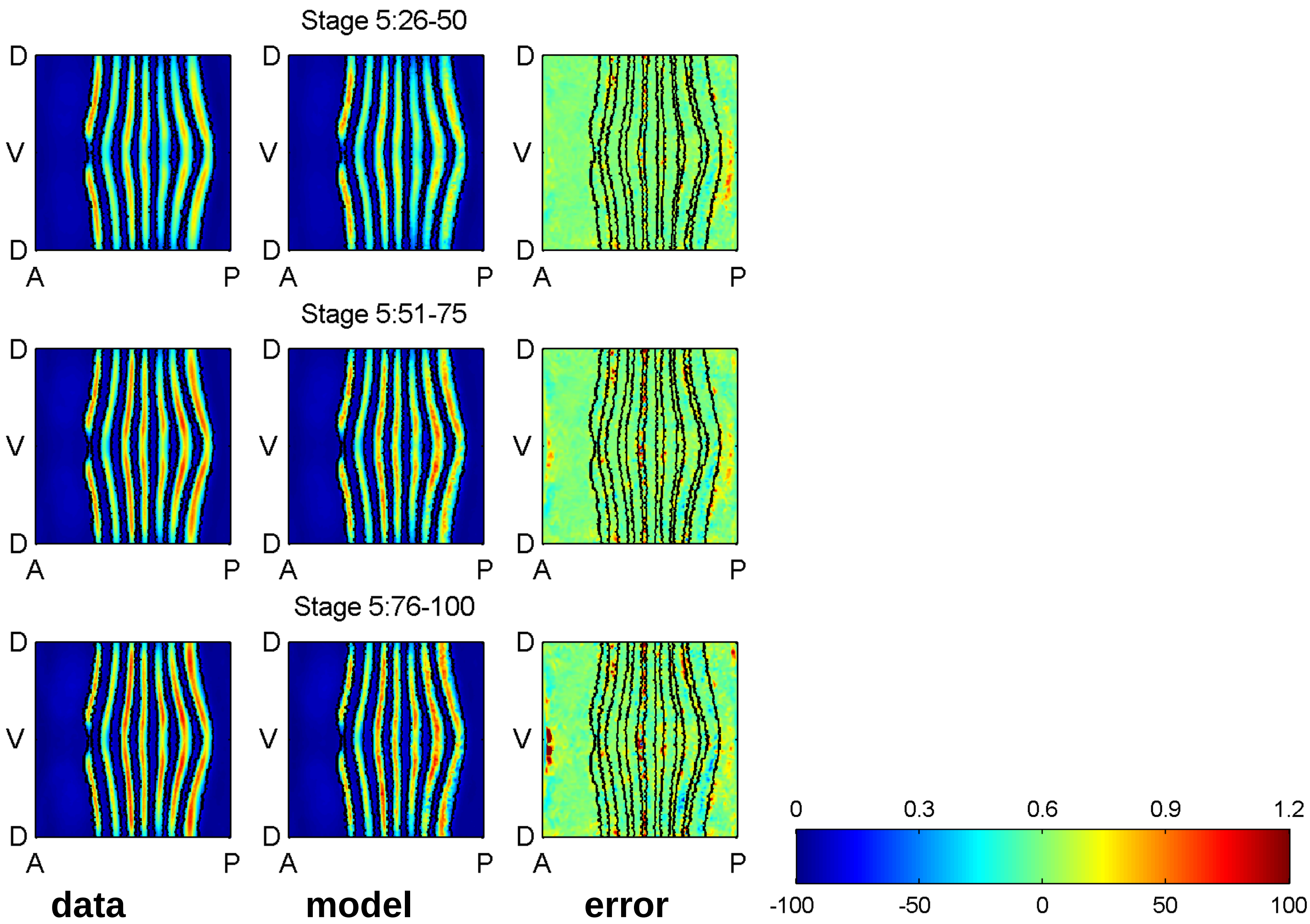
factor activity



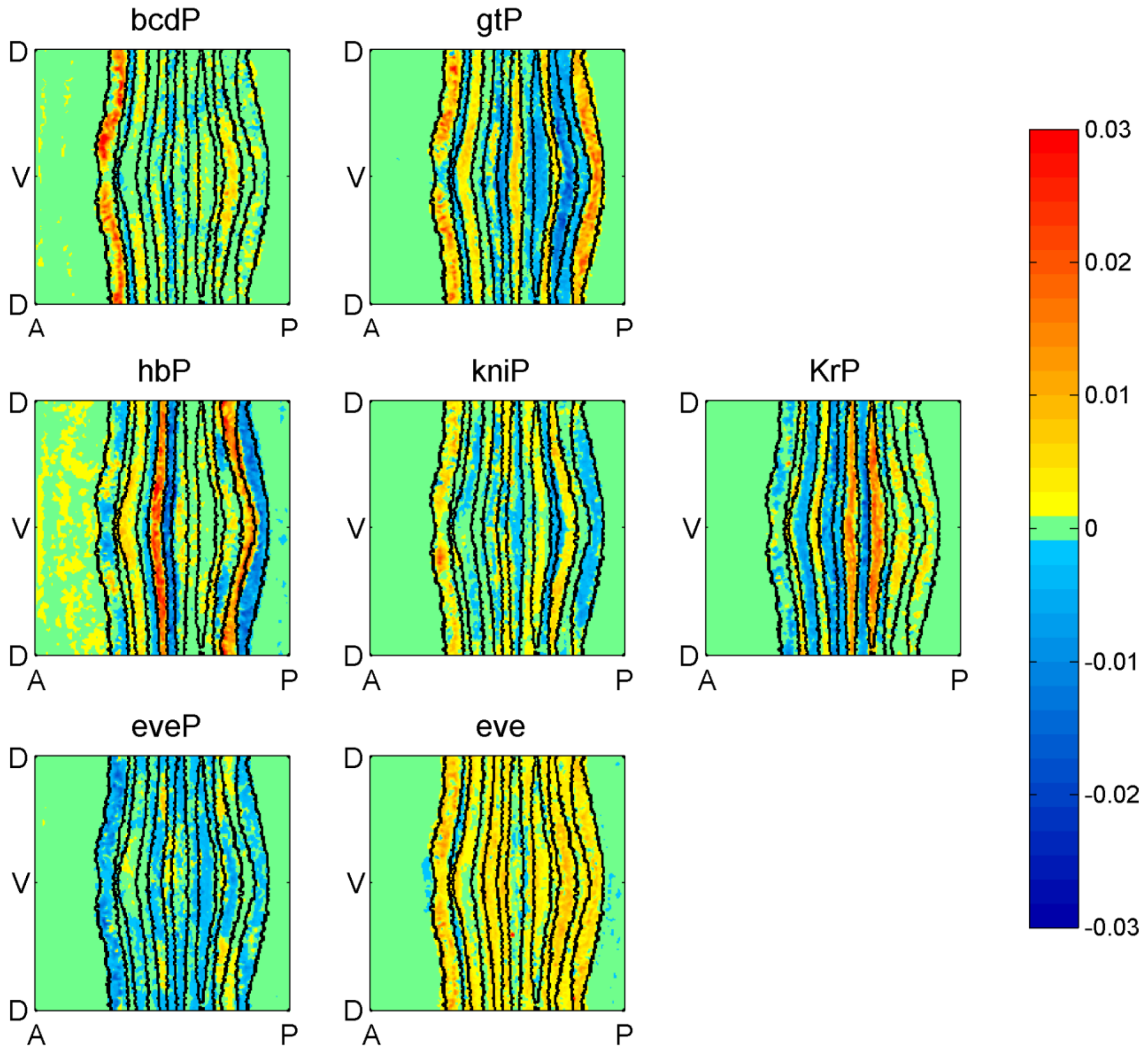
Results: eve, Stage 5: 0-25



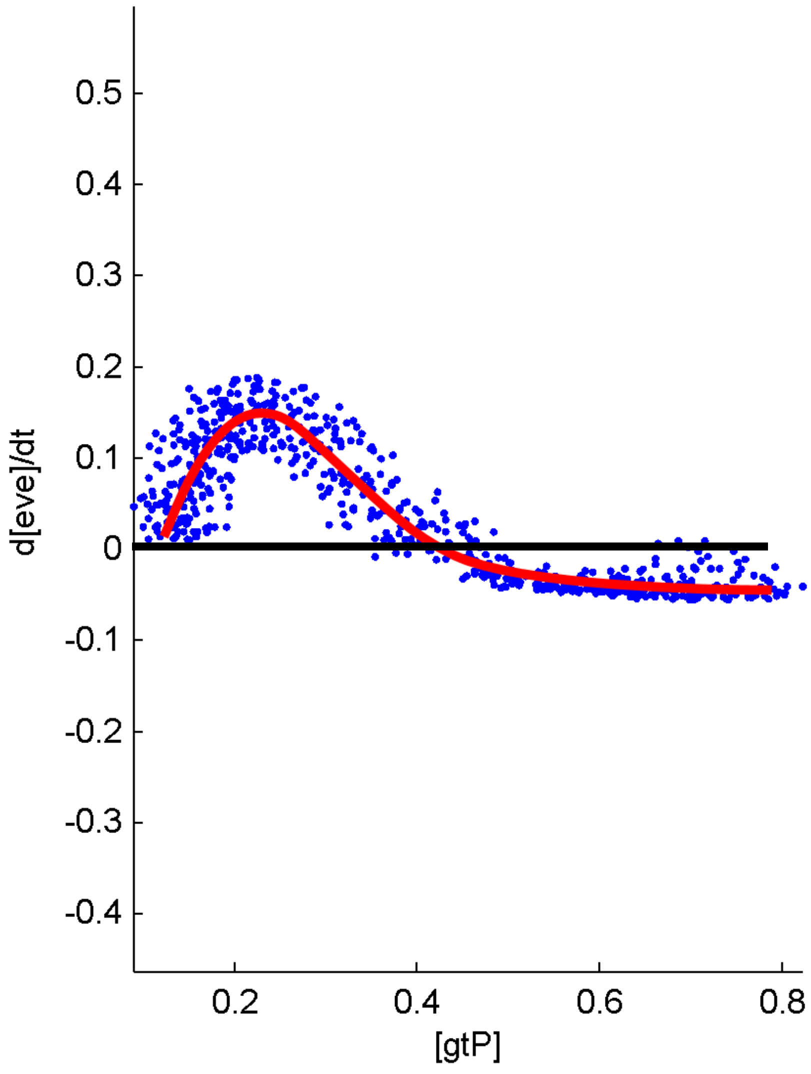
Results: eve, Stage 5: 26-100

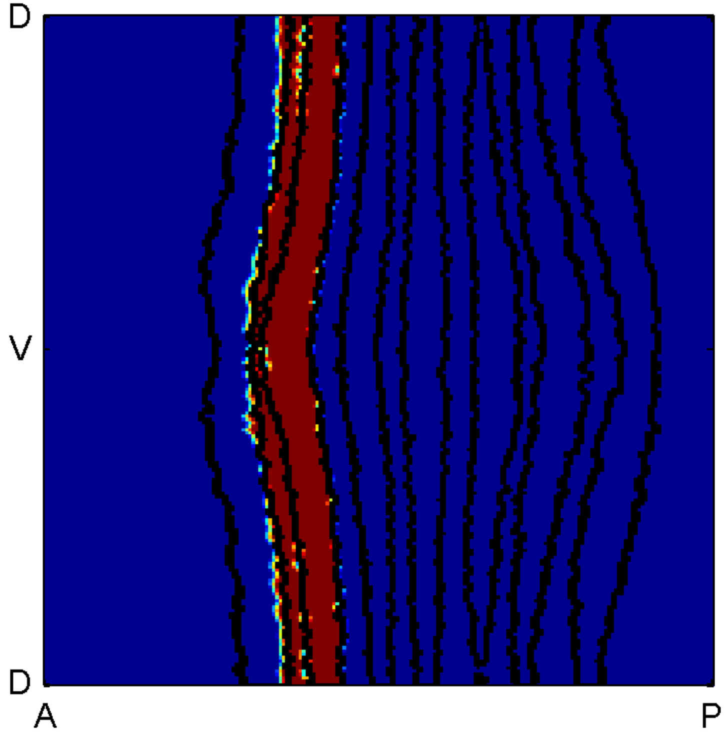


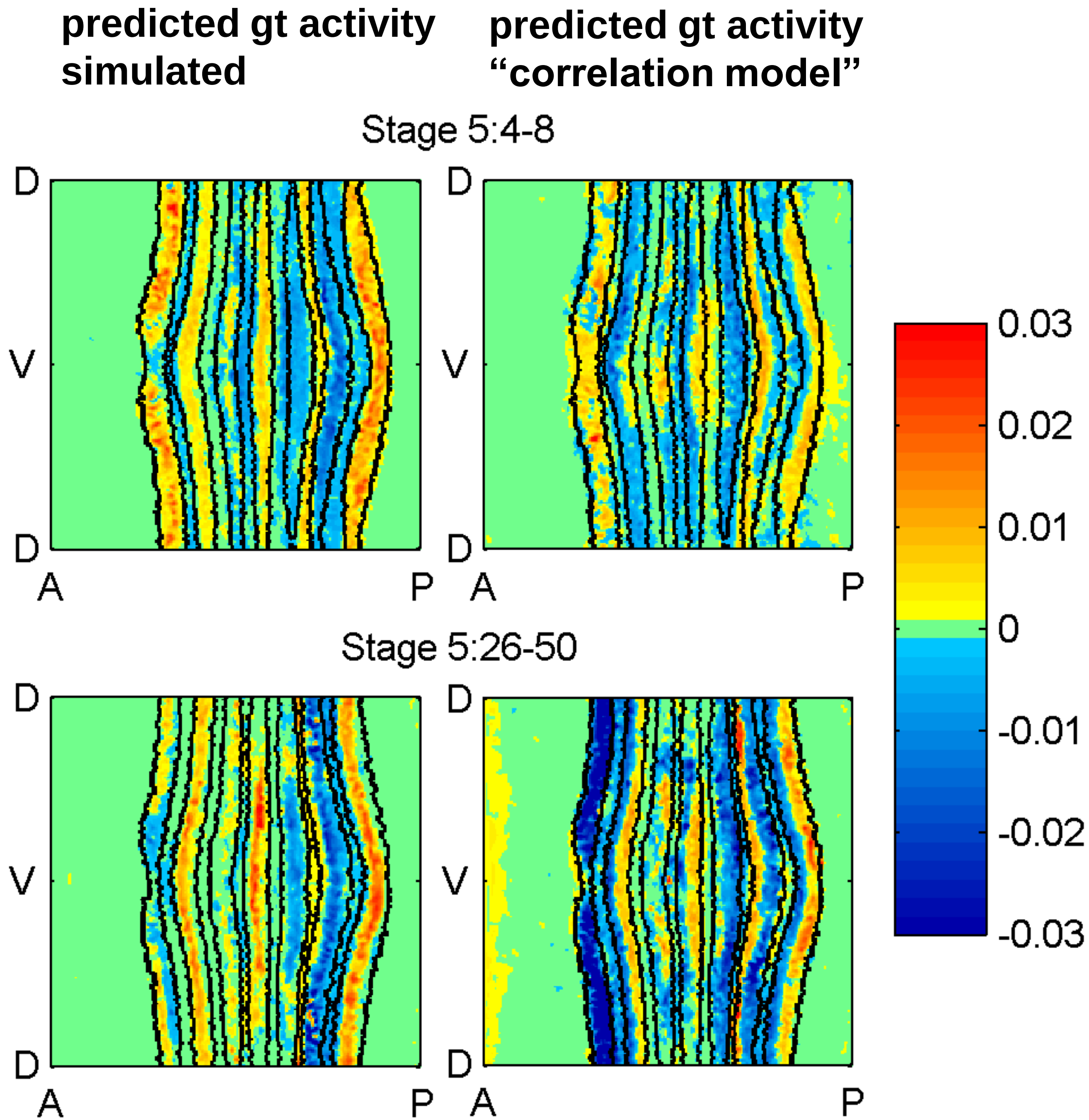
Factor activity, Stage 5: 4-8



Rate of eve production vs gtP, Stage 5: 4-8

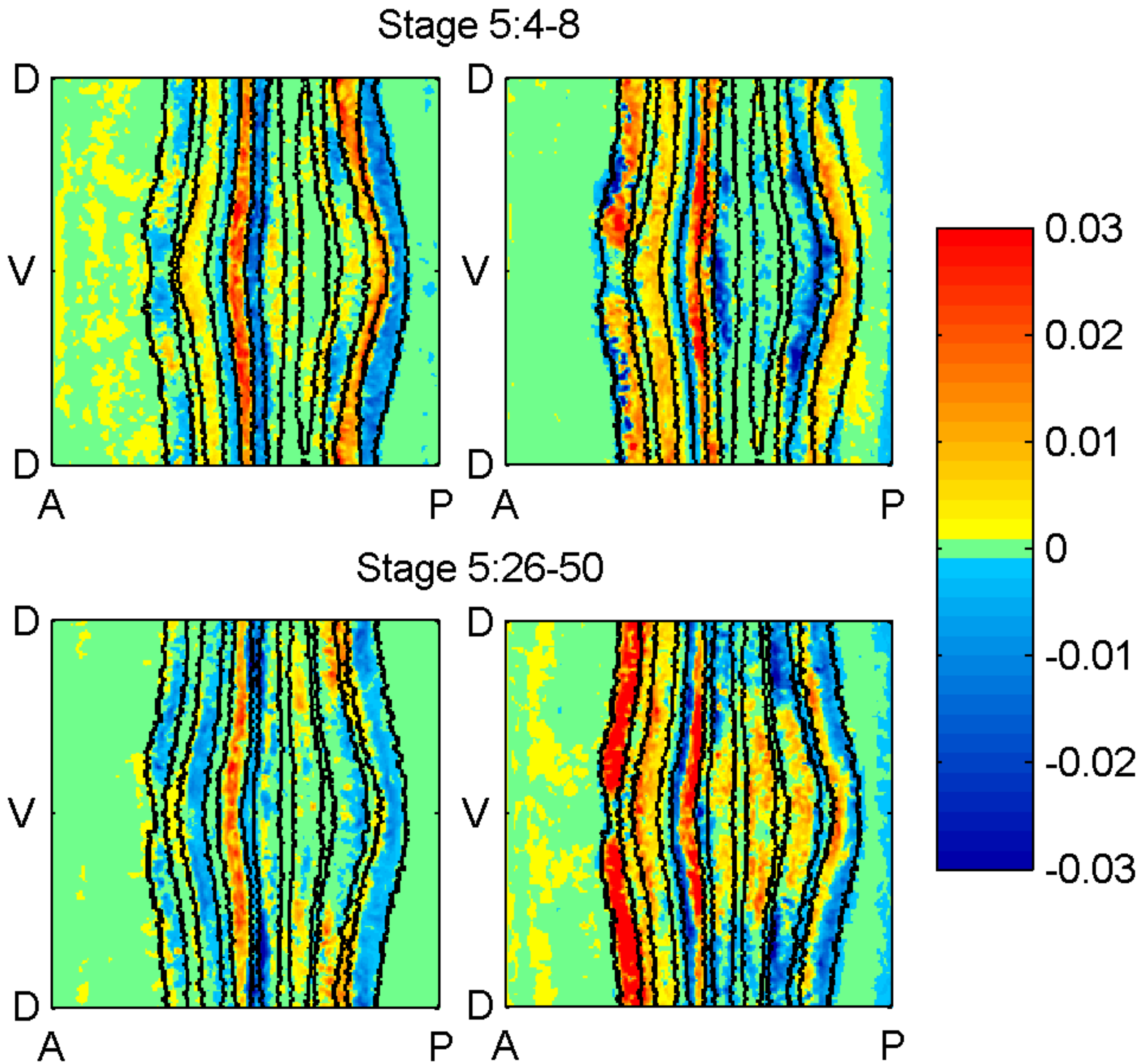


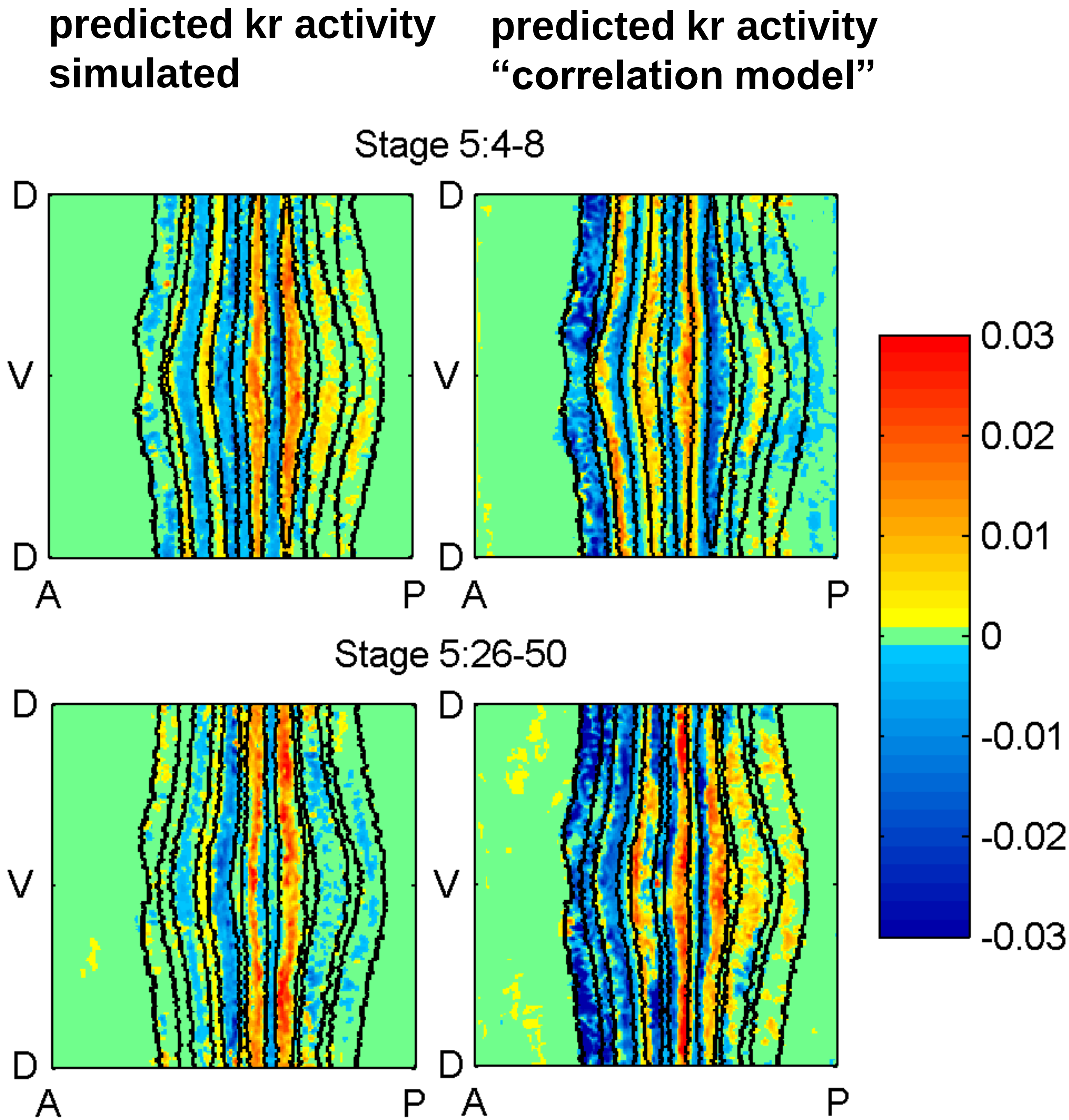



predicted gt activity
simulated
predicted gt activity
“correlation model”
Stage 5:4-8
Stage 5:26-50
D
V
D
A
P
0.03
0.02
0.01
0
-0.01
-0.02
-0.03

predicted hb activity simulated

predicted hb activity “correlation model”



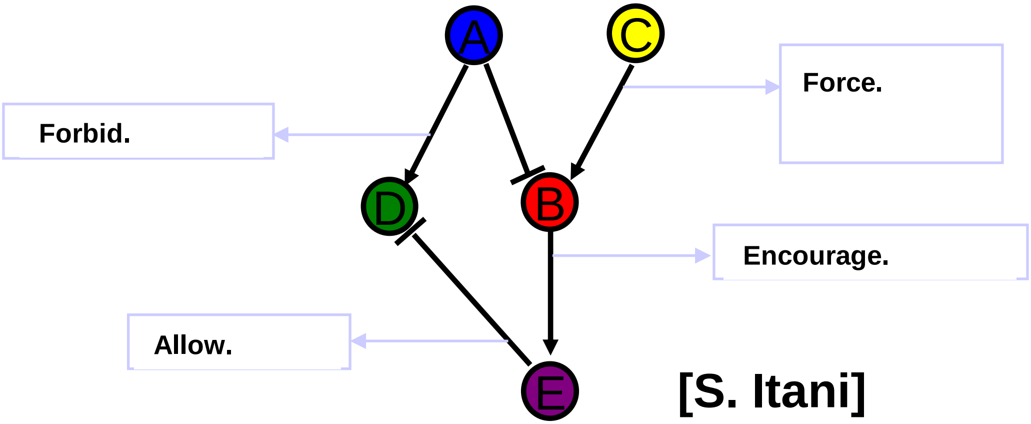

predicted kr activity
simulated
predicted kr activity
"correlation model"
Stage 5:4-8
Stage 5:26-50
D
V
D
A
P
0.03
0.02
0.01
0
-0.01
-0.02
-0.03

Potential insights

- factors appear to have concentration dependent effects
  - repressing at one concentration, activating at another
  - spurious correlations or real effects?
    - starting to analyze the other data sets (binding data)
  - if true, could add a new layer to the complexity of the signaling network
- model overlaps, but also gives some different results from the spatial correlation model
  - can distinguish between weakening of repression, and repression, for example

Summary …

- Method for local linear regression, designed for systems evolving on a manifold of lower dimension than overall space
  - Designed to prevent overfitting
  - Can be used as a tool to help identify network structure
- Another new project: network and parameter identification of HER2/3 network in cancer (with Joe Gray, Young-Hwan Chang, Steven Xie, and Soulaiman Itani)



[S. Itani]

Nonparametric Identification of Regulatory Interactions from Spatial and Temporal Gene Expression Data

Anil Aswani

Peter Bickel

Claire Tomlin

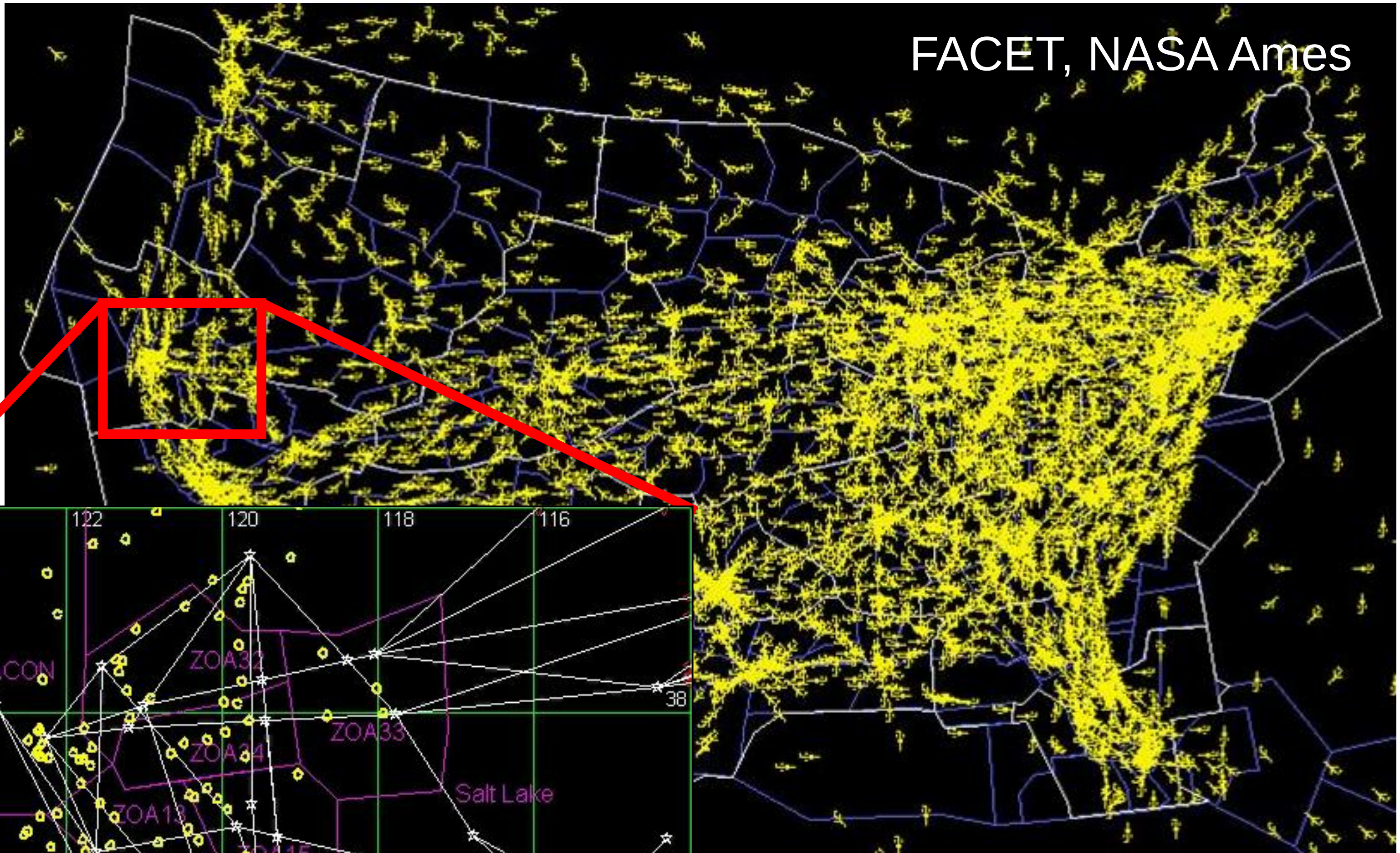
Joint work with Mark Biggin, Charless Fowlkes, Soile Keränen, and Jitendra Malik

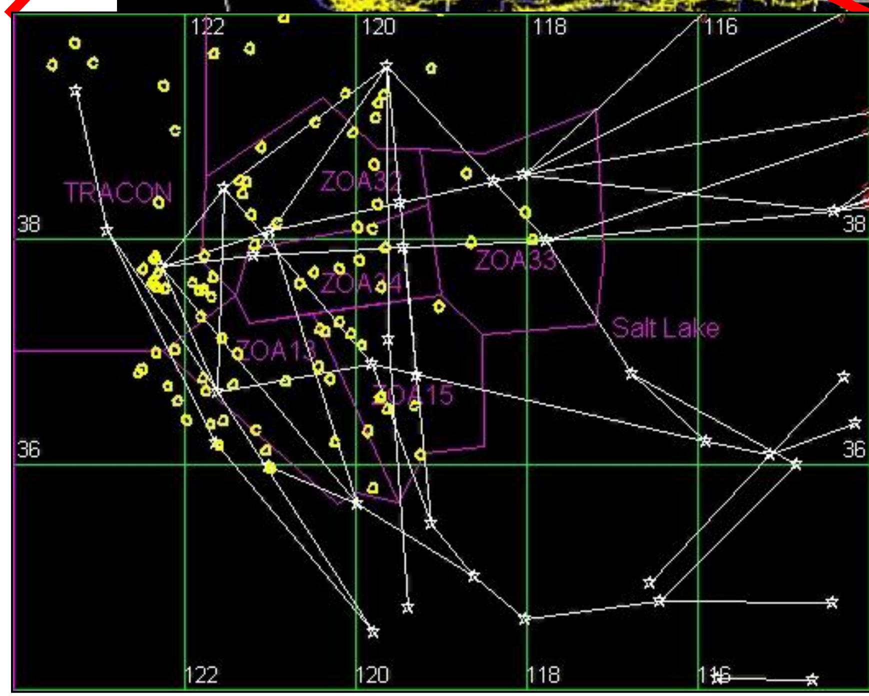
Electrical Engineering and Computer Sciences, UC Berkeley

ACCESS Linnaeus Center, School of Electrical Engineering, KTH

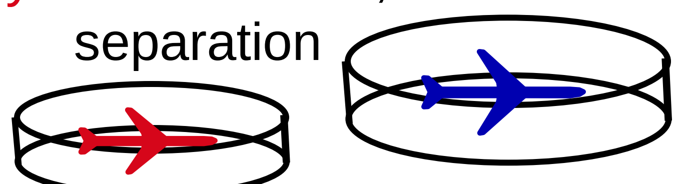
Thanks: NSF, NIH NCI

Air Traffic Control: Separation Assurance

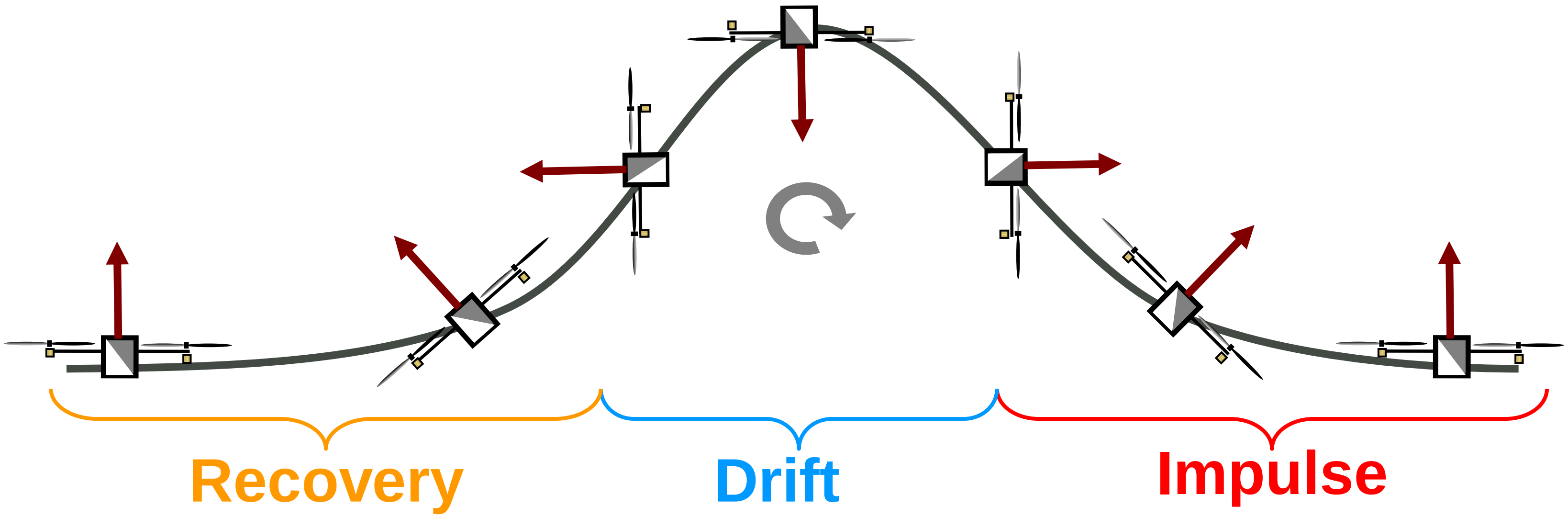




Safety: 5 mile lateral, 1000 ft vertical separation



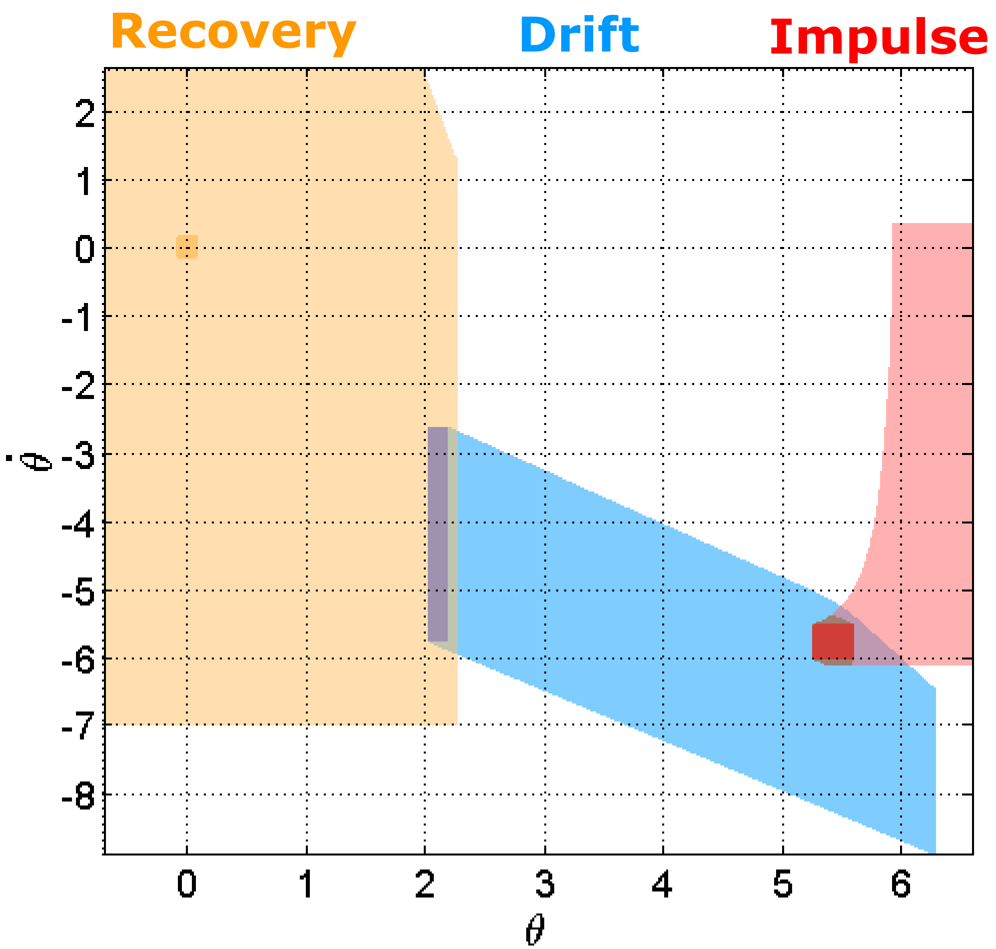
Case Study 2: Back-Flip



- Divide flip into three modes
- Hit desired target sets while avoiding unsafe sets

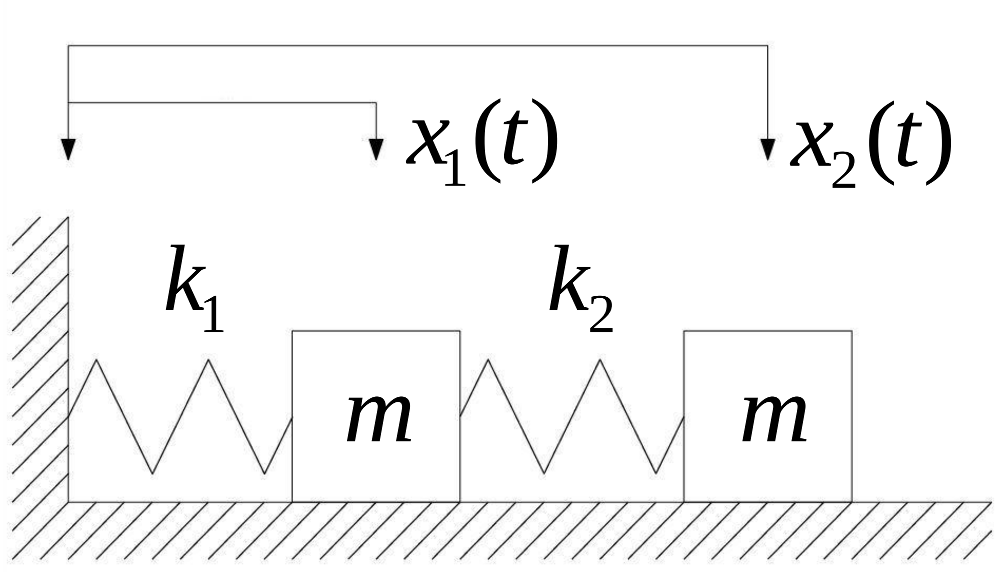
Back-flip: Method

- Identify target region in rotational state space for each mode
- Use reachable sets to calculate capture basin for each target
  - Dynamic game formulation accounts for worst-case disturbances
- Verify that target of each mode is contained by capture basin of next mode



STARMAC

Toy Example: Mass-Spring System



L = Length of uncompressed spring

$$\begin{bmatrix} \ddot{x}_1 \\ \ddot{x}_2 \end{bmatrix} = \begin{bmatrix} L(k_1 - k_2)/m & -(k_1 + k_2)/m & 0 \\ k_2 L / m & k_2 m & -k_2 m \end{bmatrix} \begin{bmatrix} 1 \\ x_1 \\ x_2 \end{bmatrix}$$

Mass-Spring System

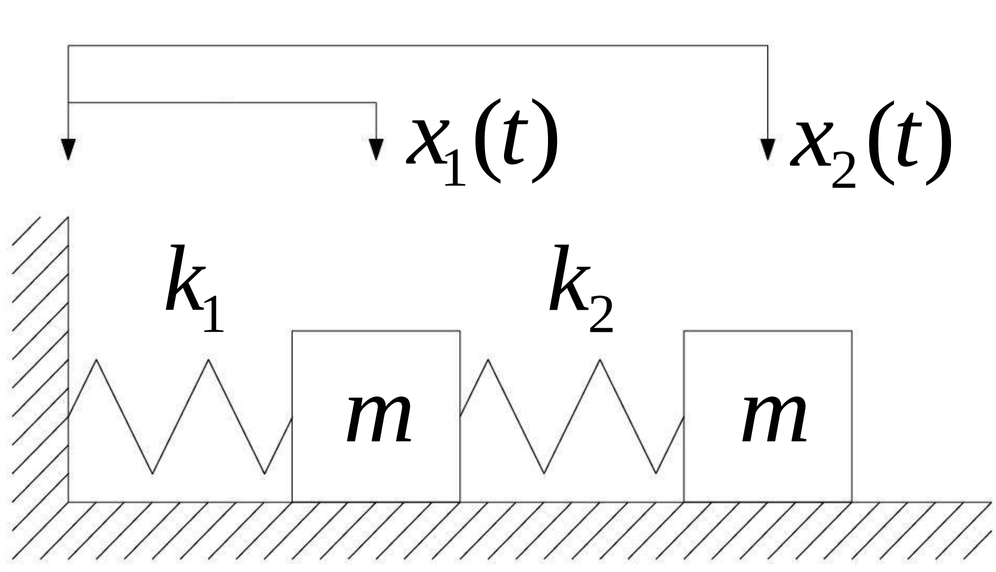
- X = matrix of low noise measurements of positions
- Y = vector of noisy measurements of acceleration

$$Y = \begin{bmatrix} \ddot{x}_1(t_1) \\ \vdots \\ \ddot{x}_1(t_N) \end{bmatrix} + \eta \qquad X = \begin{bmatrix} 1 & x_1(t_1) & x_2(t_1) \\ \vdots & \vdots & \vdots \\ 1 & x_1(t_N) & x_2(t_N) \end{bmatrix} + \begin{bmatrix} 0 & \upsilon_1 & \upsilon_2 \end{bmatrix}$$

- K = vector of estimated coefficients for the first ODE

$$\begin{aligned} K &= \arg\min_{\beta} \|Y - X\beta\|_2^{\,2} \\ &= \left(X^T X\right)^{-1} X^T Y \end{aligned}$$

Degenerate Mass-Spring System



L **= Length of uncompressed spring**

k_2 **is a very stiff spring**

$$\ddot{x}_1 = -\frac{k_1}{2m}\left(x_1 - L\right)$$

$$x_2 = x_1 + L$$

Degenerate Mass-Spring System

- X = matrix of low noise measurements of positions
- Y = vector of noisy measurements of acceleration
- K = vector of estimated coefficients for the first ODE

$$K = \arg\min_{\beta} \|Y - X\beta\|_2^2$$
$$= \left(X^T X\right)^{-1} X^T Y$$

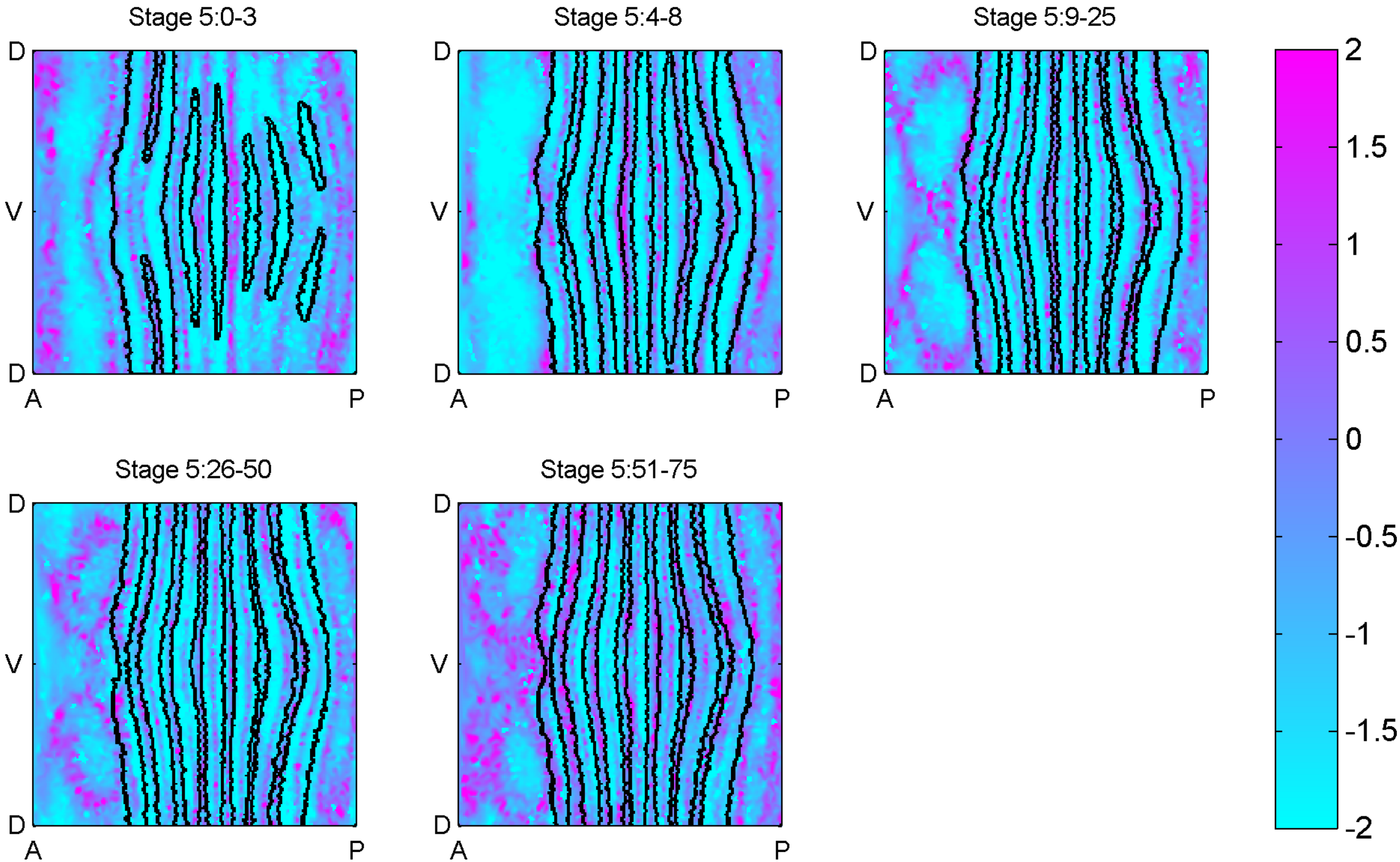
- PROBLEM: Covariance matrix is (nearly) singular
- CAUSE: States have geometric constraints

➡ **Important to design methods robust to this**

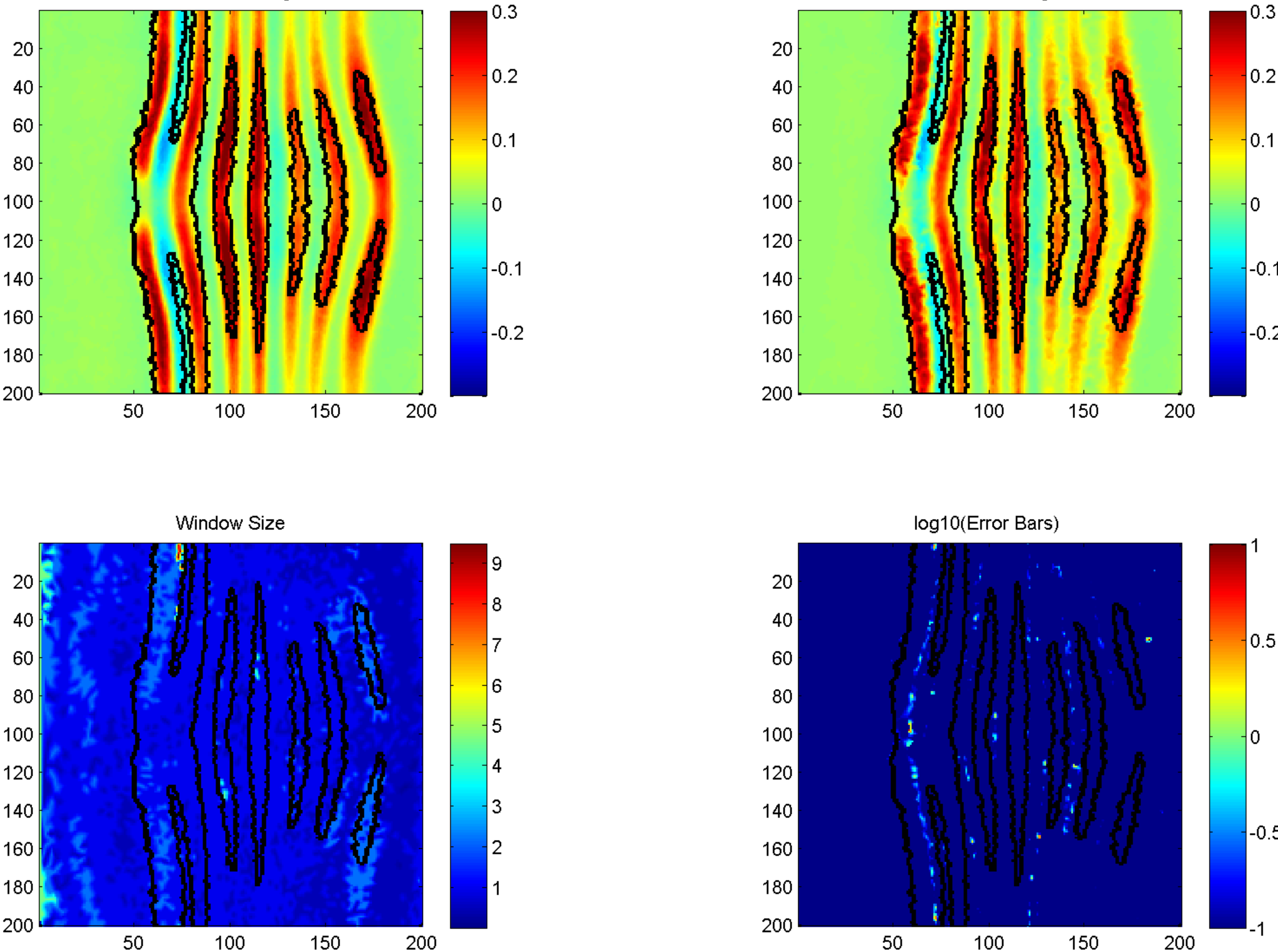
Mass-Spring Example

	Model (8.10), with $k_1 = 0.4, k_2 = 0.25, L = 1, m = 1$			
	$nMSE(\hat{B}, B_1)$		$nMSE(\hat{B}, B_2)$	
OLS/MP	0.096	(0.011)	1.422	(0.130)
RR	0.091	(0.009)	1.286	(0.115)
EN	0.091	(0.009)	1.286	(0.115)
PLS	0.096	(0.011)	1.422	(0.130)
PCR	0.096	(0.011)	1.422	(0.130)
EDE	0.091	(0.009)	1.286	(0.115)
ALEDE	0.091	(0.009)	1.286	(0.115)
EDEP	0.091	(0.009)	1.286	(0.115)
ALEDEP	0.091	(0.009)	1.286	(0.115)
	Model (8.11), with $k_1 = 0.4, k_2 = 10000, L = 1, m = 1$			
	$nMSE(\hat{B}, B_1)$		$nMSE(\hat{B}, B_2)$	
OLS/MP	1.000	(0.000)	0.231	(0.162)
RR	1.000	(0.000)	0.118	(0.058)
EN	1.000	(0.000)	0.135	(0.074)
PLS	1.000	(0.000)	0.160	(0.167)
PCR	1.000	(0.000)	0.162	(0.166)
EDE	1.000	(0.000)	0.112	(0.060)
ALEDE	1.000	(0.000)	0.129	(0.077)
EDEP	1.000	(0.000)	0.111	(0.060)
ALEDEP	1.000	(0.000)	0.128	(0.078)

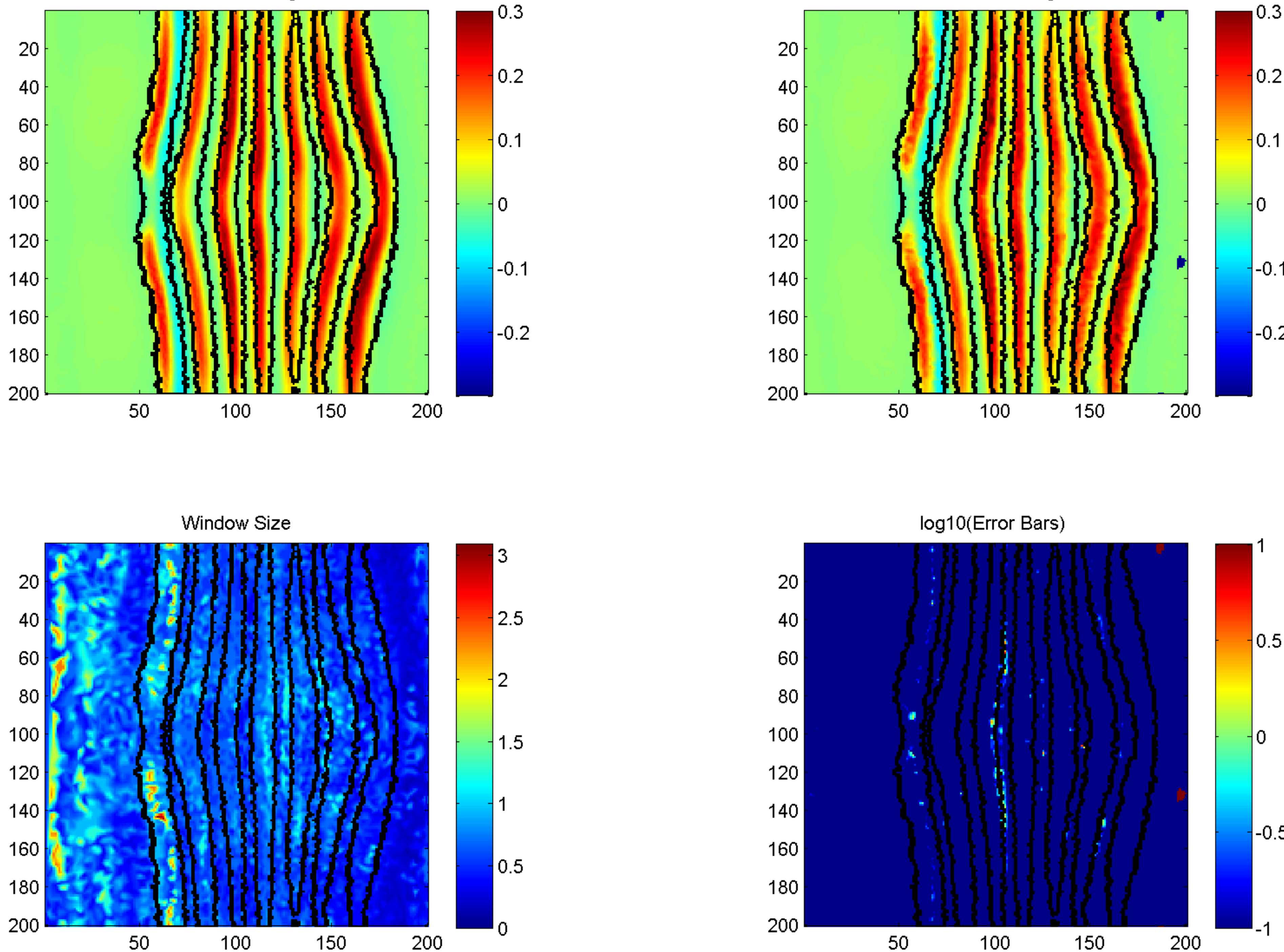
Error Bars



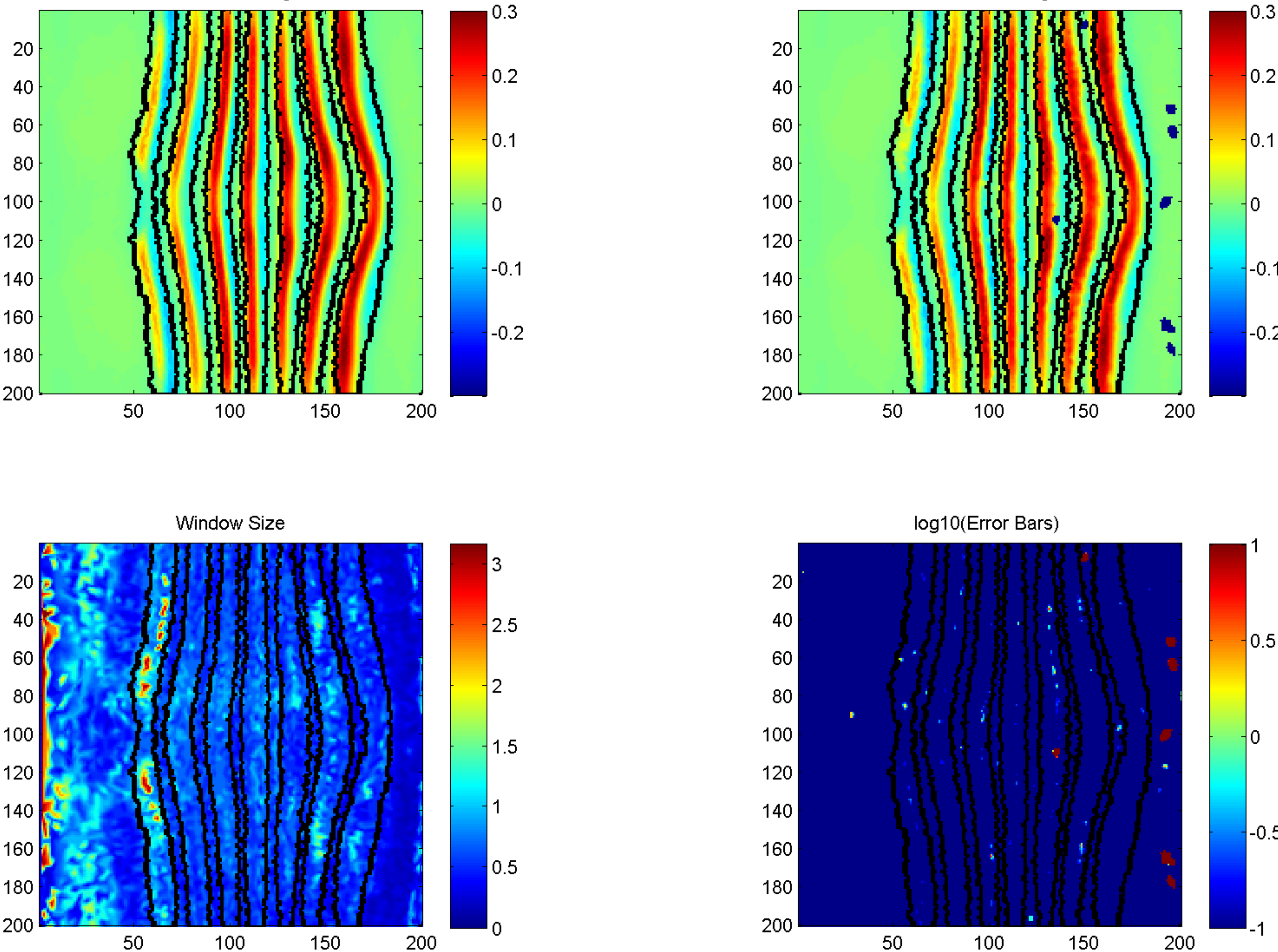
Stage 5:0-3
Stage 5:4-8
Stage 5:9-25
Stage 5:26-50
Stage 5:51-75
D
V
D
A
P
2
1.5
1
0.5
0
-0.5
-1
-1.5
-2



FIT: [0, 3]



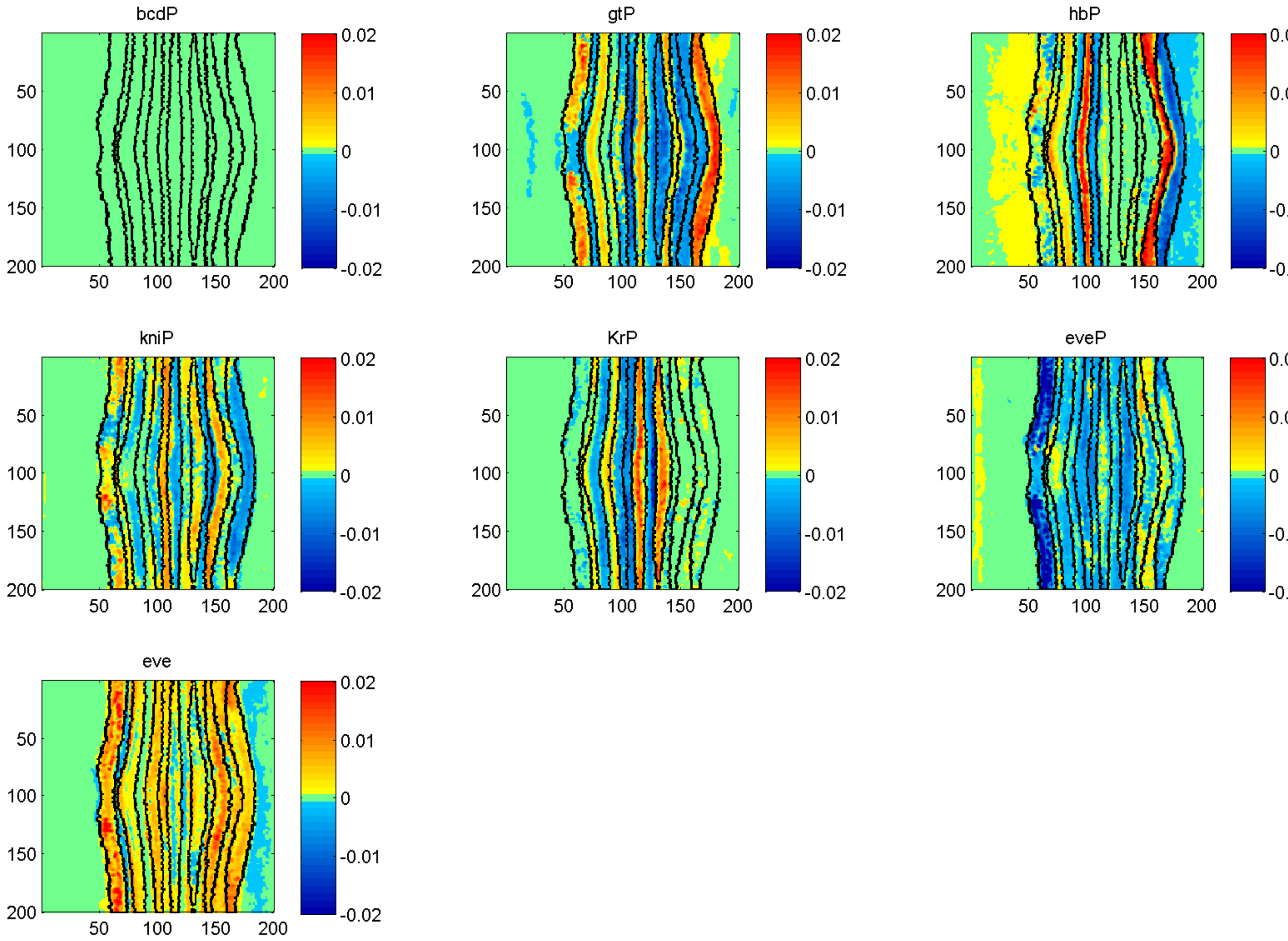
FIT: [4, 8]



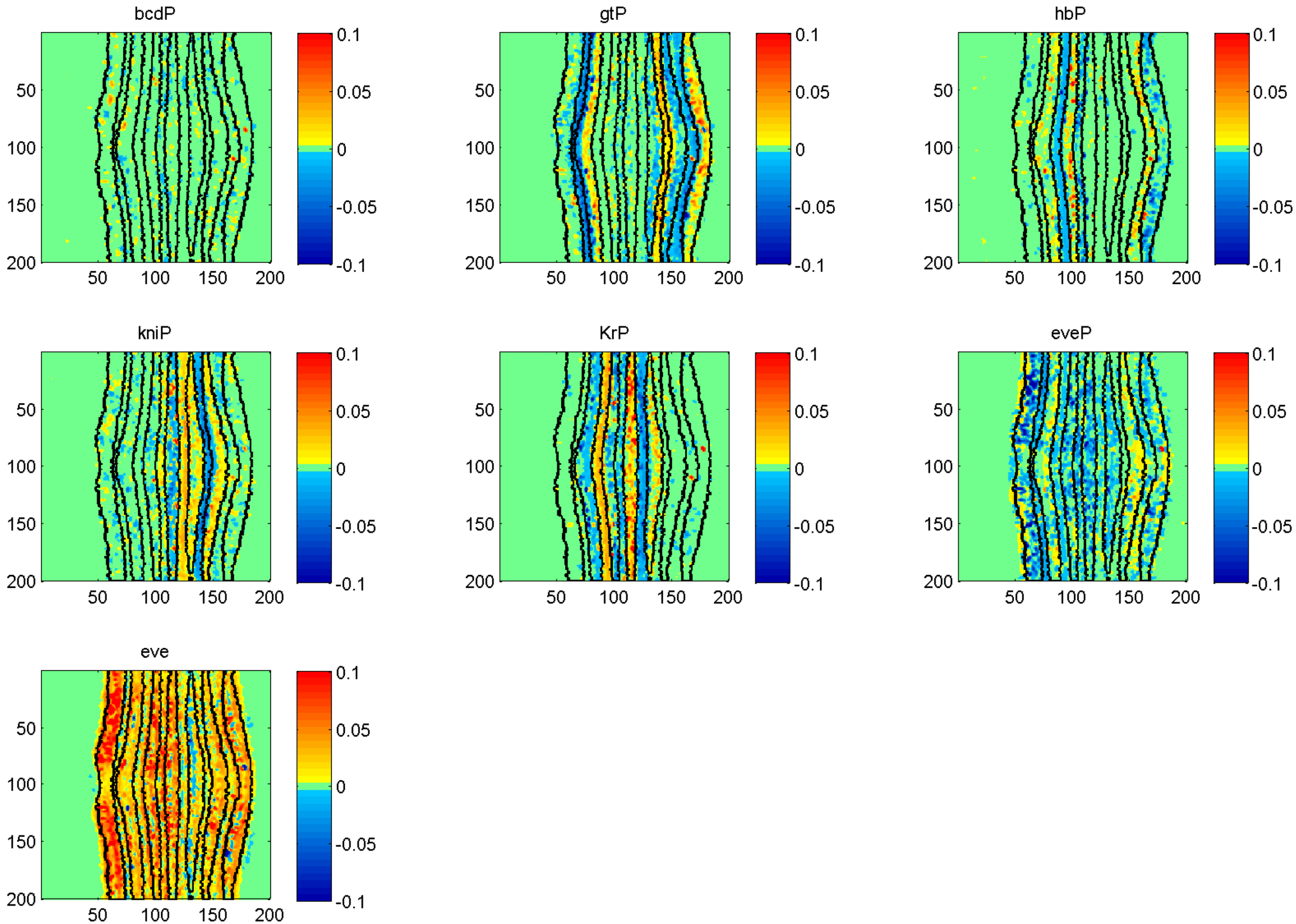
FIT: [0, 35]

RESULTS:
Heatmap of Coefficients Times Factor Concentrations on Eve Stripes at Stage 5:4-8 with Changing Window Size

In general need to explain the weakening of repression etc.

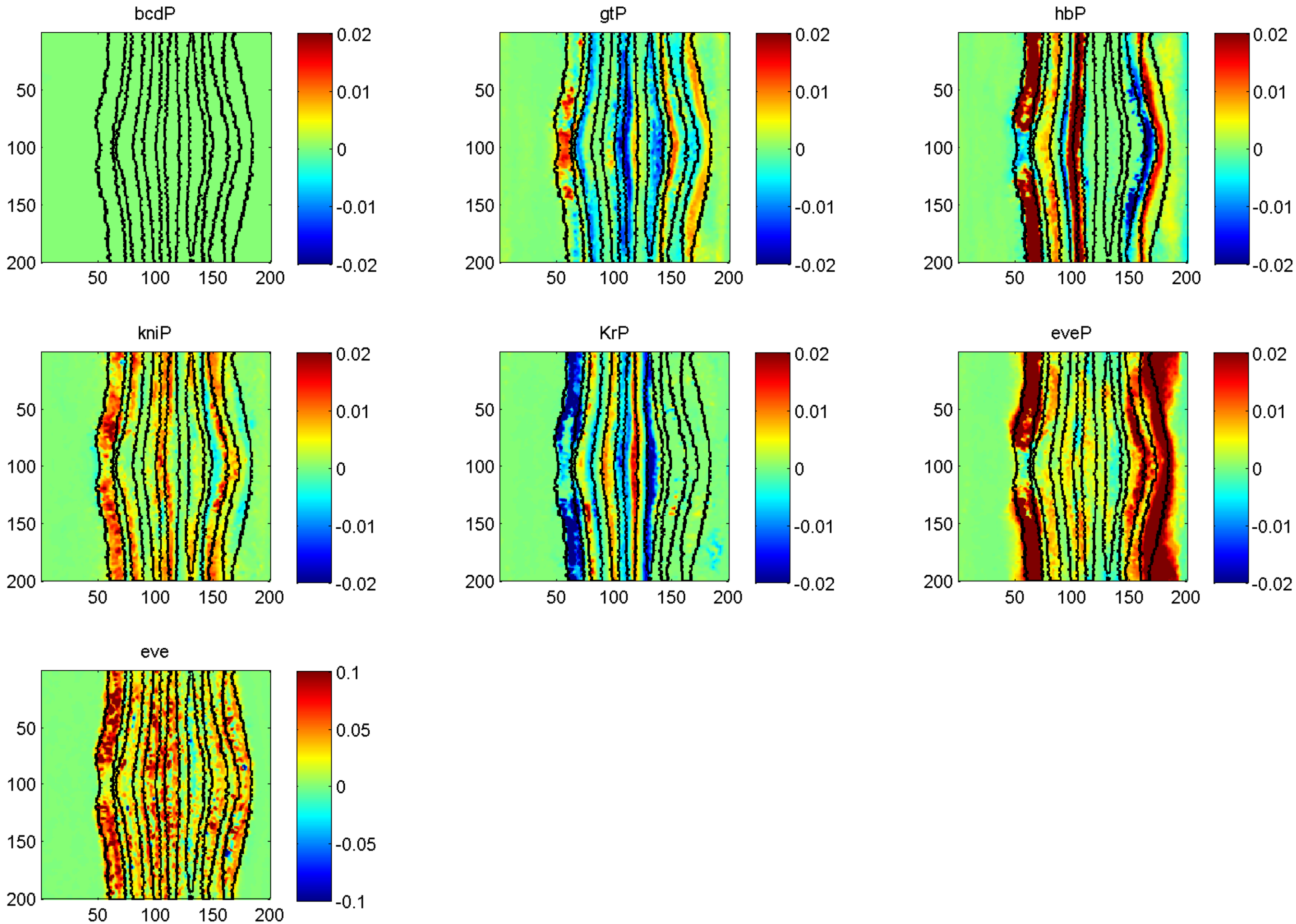

bcdP
gtP
hbP
kniP
KrP
eveP
eve
50
100
150
200
0.02
0.01
0
-0.01
-0.02

Heatmap of Coefficients Times Factor Concentrations on Eve Stripes at Stage 5:4-8 with Fixed Window Size of Circle with Width of 6 Cells



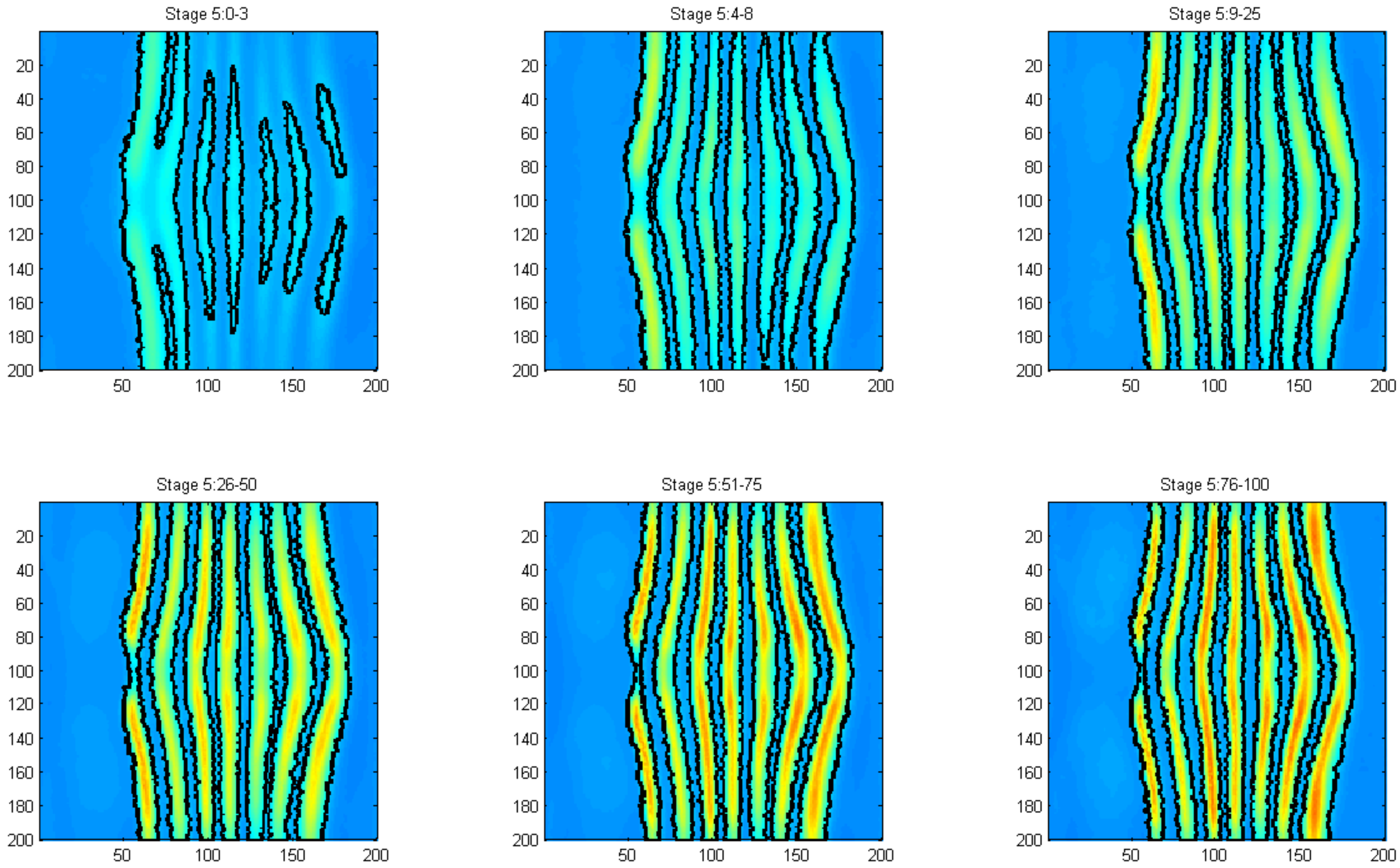
bcdP
gtP
hbP
kniP
KrP
eveP
eve
0.1
0.05
0
-0.05
-0.1
50
100
150
200

Heatmap of Correlation Between Factor Concentration and Eve Stripes at Stage 5:4-8

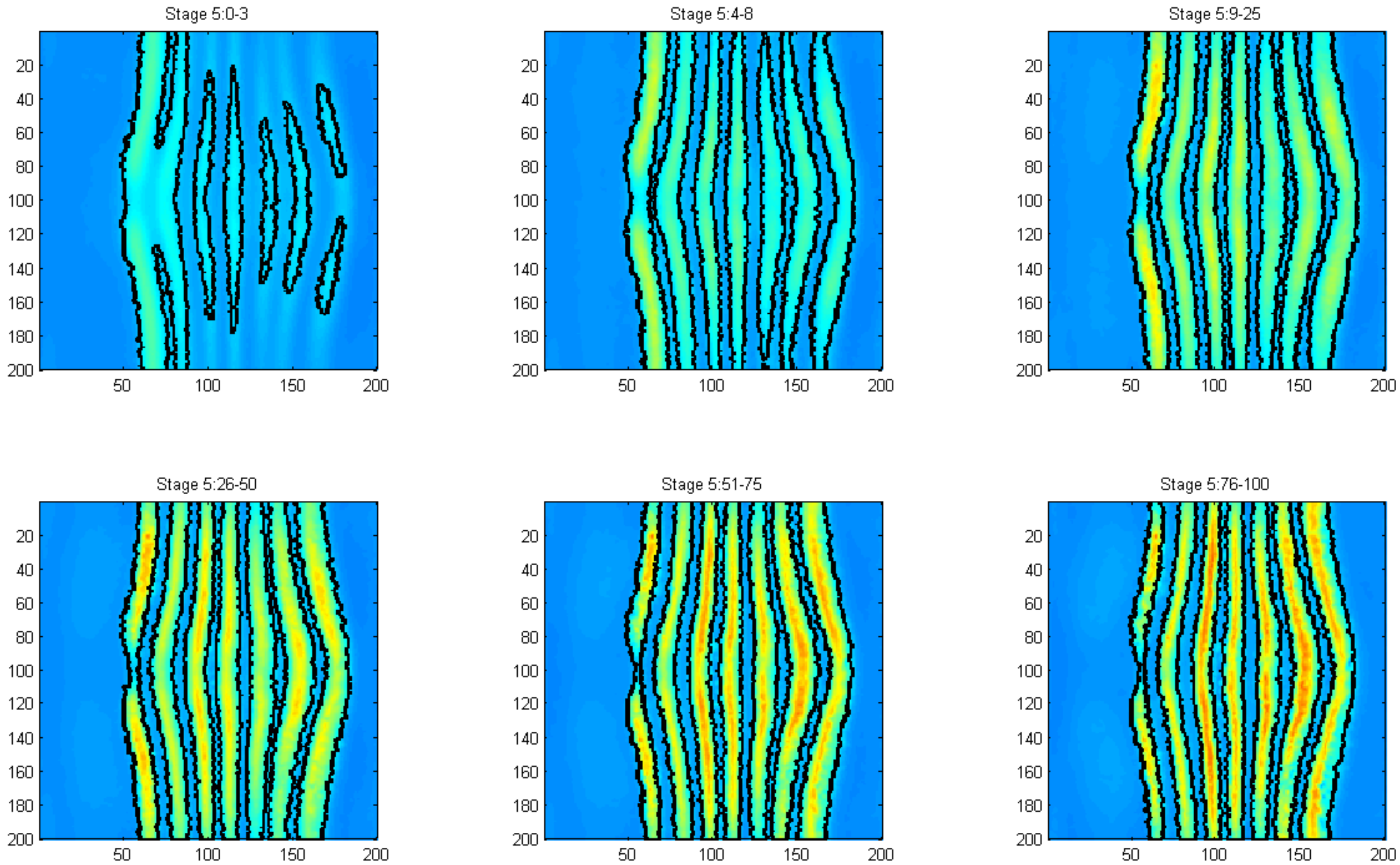


bcdP
gtP
hbP
kniP
KrP
eveP
eve
50
100
150
200
0.02
0.01
0
-0.01
-0.02
0.1
0.05
-0.05
-0.1

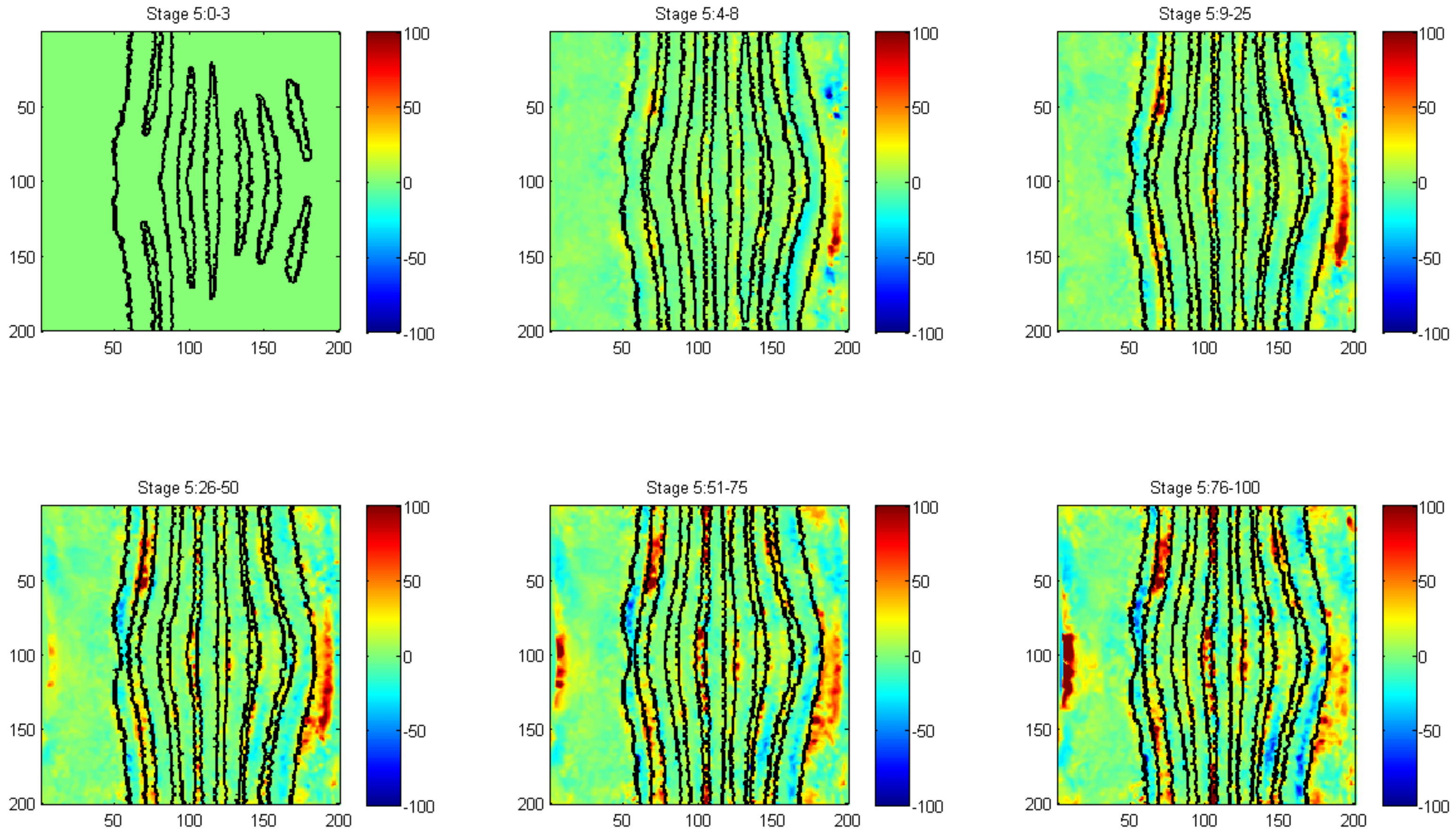
Experimental eve mRNA Patterns



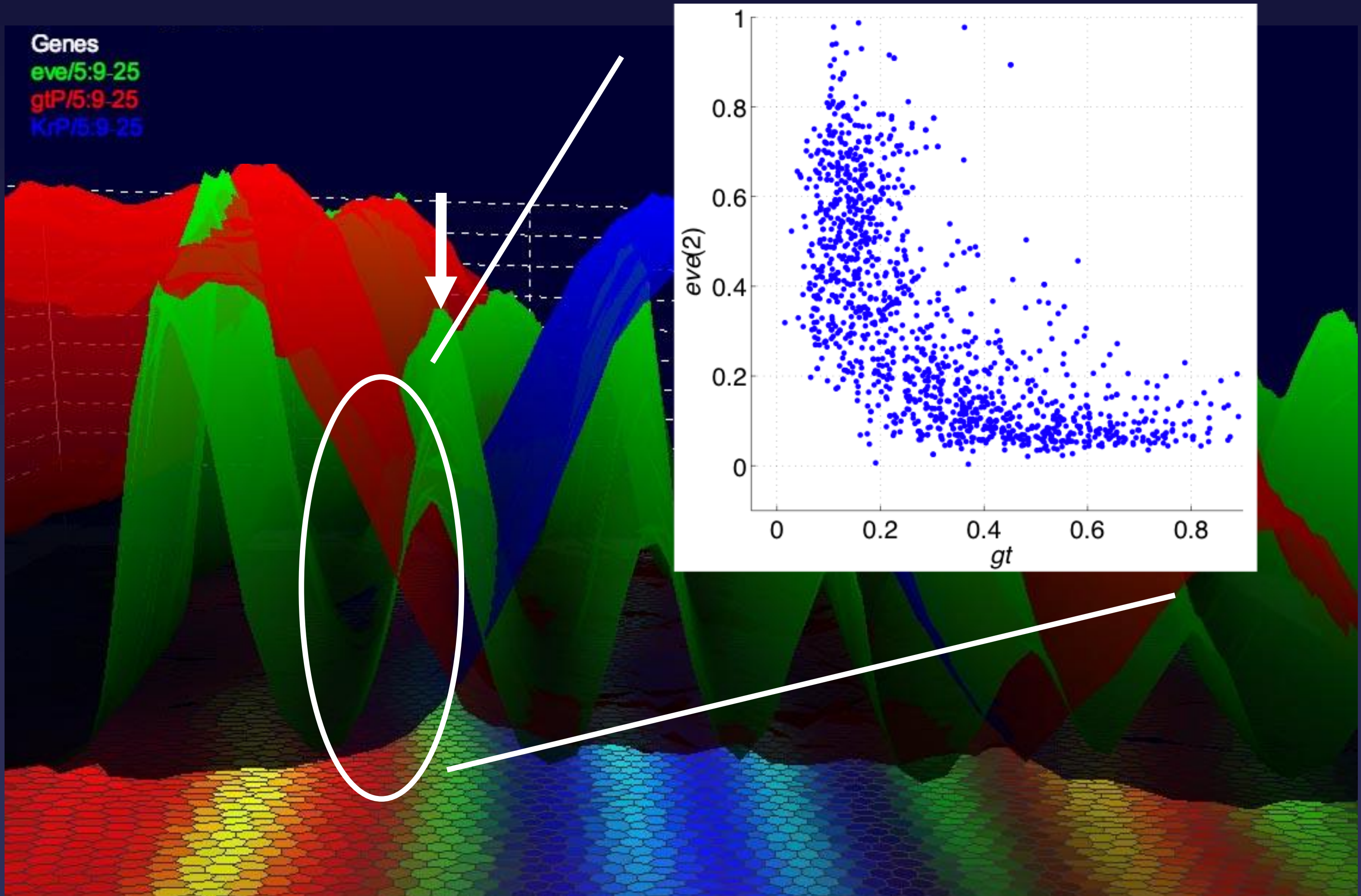
Simulated eve mRNA Patterns



Percent Error



Regulation is often associated with correlations in expression

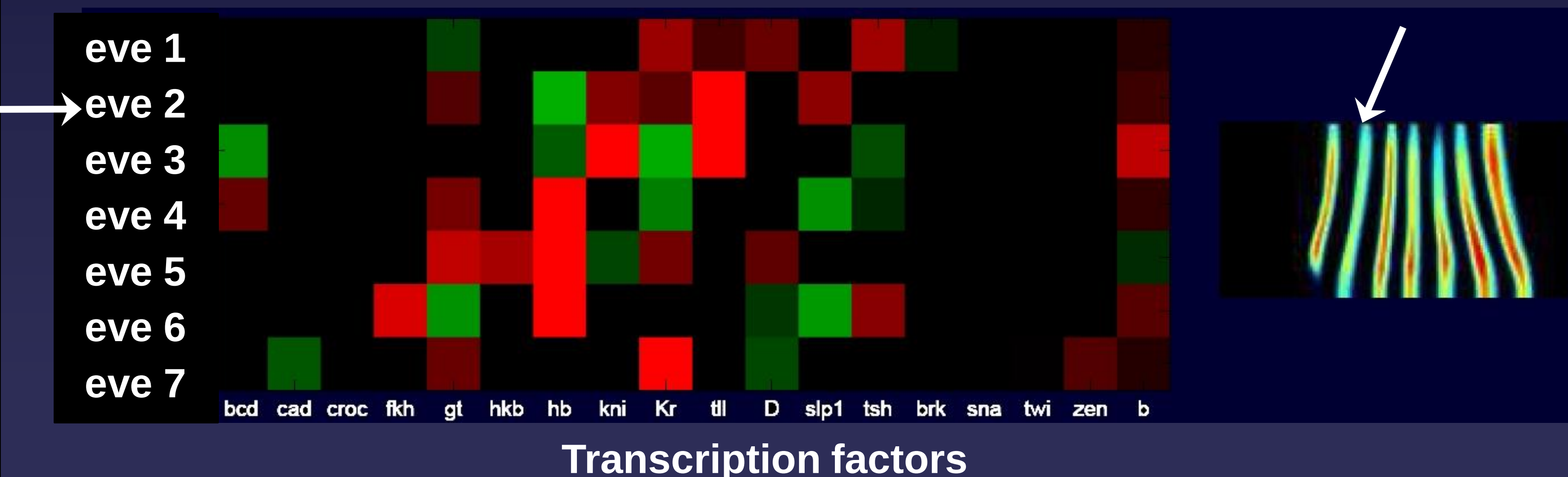


Regression analysis detects known regulatory interactions

$$M(x,t) = F\{\ P_i(x,t)\ \}$$

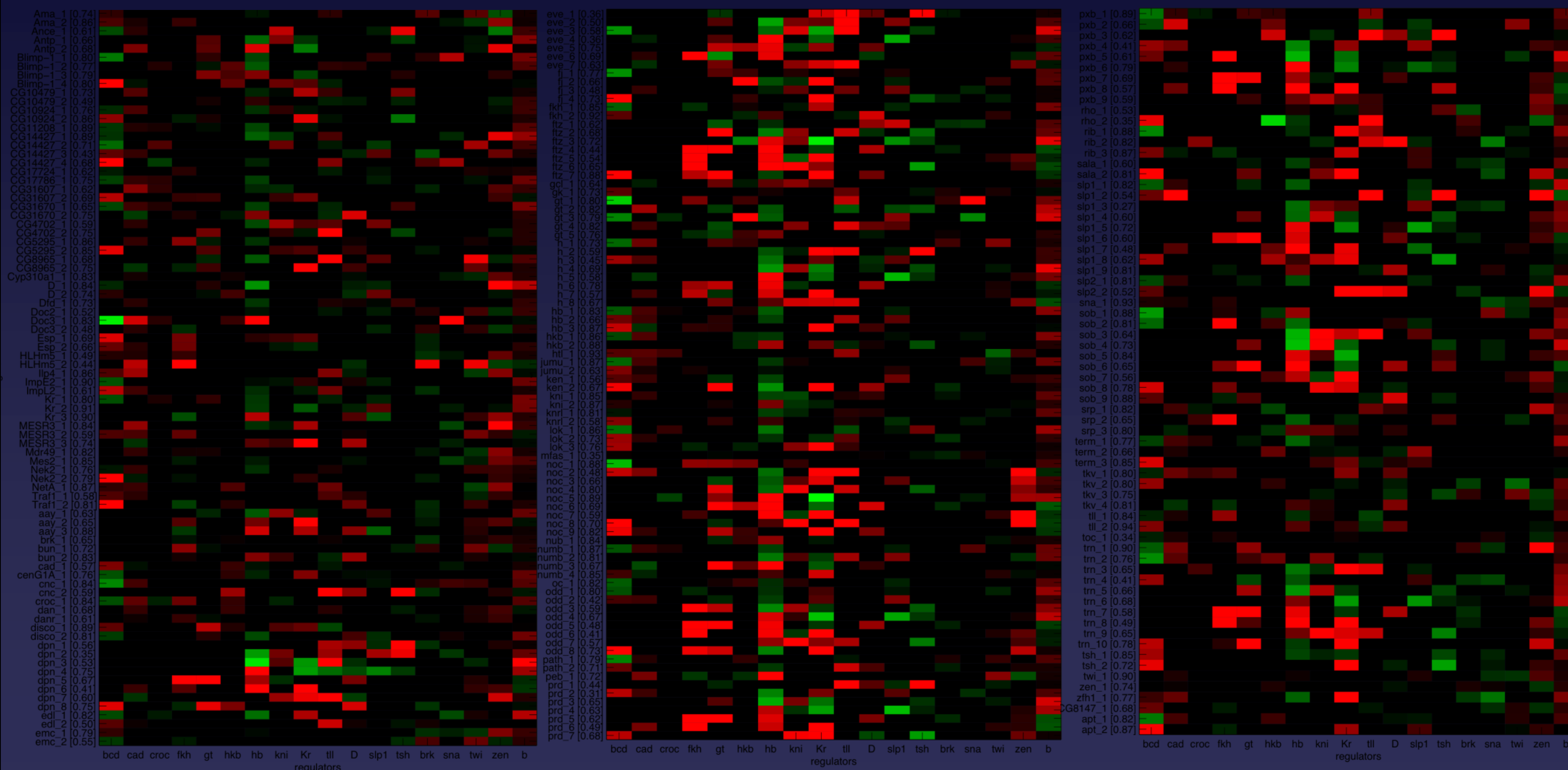
Activator

Repressor



Fowlkes et al, 2008

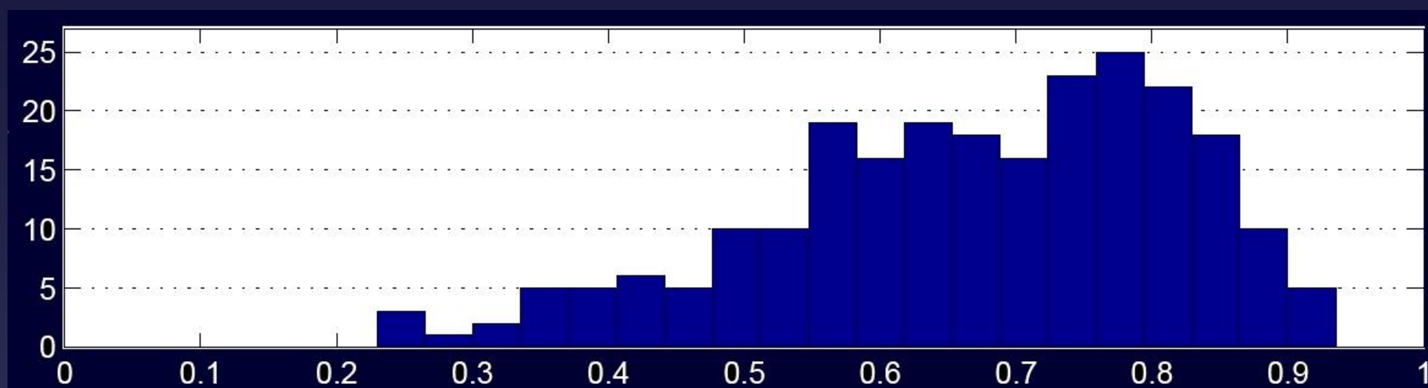
The method can be rapidly applied to any large quantitative dataset



100s of expression stripes from 95 genes

The measured expression correlates well with that predicted by the regression

Frequency



Correlation coefficient

Most expression stripes r > 0.6

Talk outline

- One slide on PCP – use as motivation (here, we assumed a structure – given to us from Jeff, before modeling). What if we didn’t have, or weren’t confident with, the structure?
- Simple pendulum example
- Mark’s system
- Local linear regression – justify, as a basis for identifying a potentially nonlinear system
  - Method gives the regions of best fit, so there is a higher density of models in “very nonlinear” regions
  - Key: protect against overfitting. If the system dynamic lies on a lower dimensional manifold, find it. (you can use the hb kr example here if you want)
  - Sparsity, high dimensionality(?), non-parametric
- Results

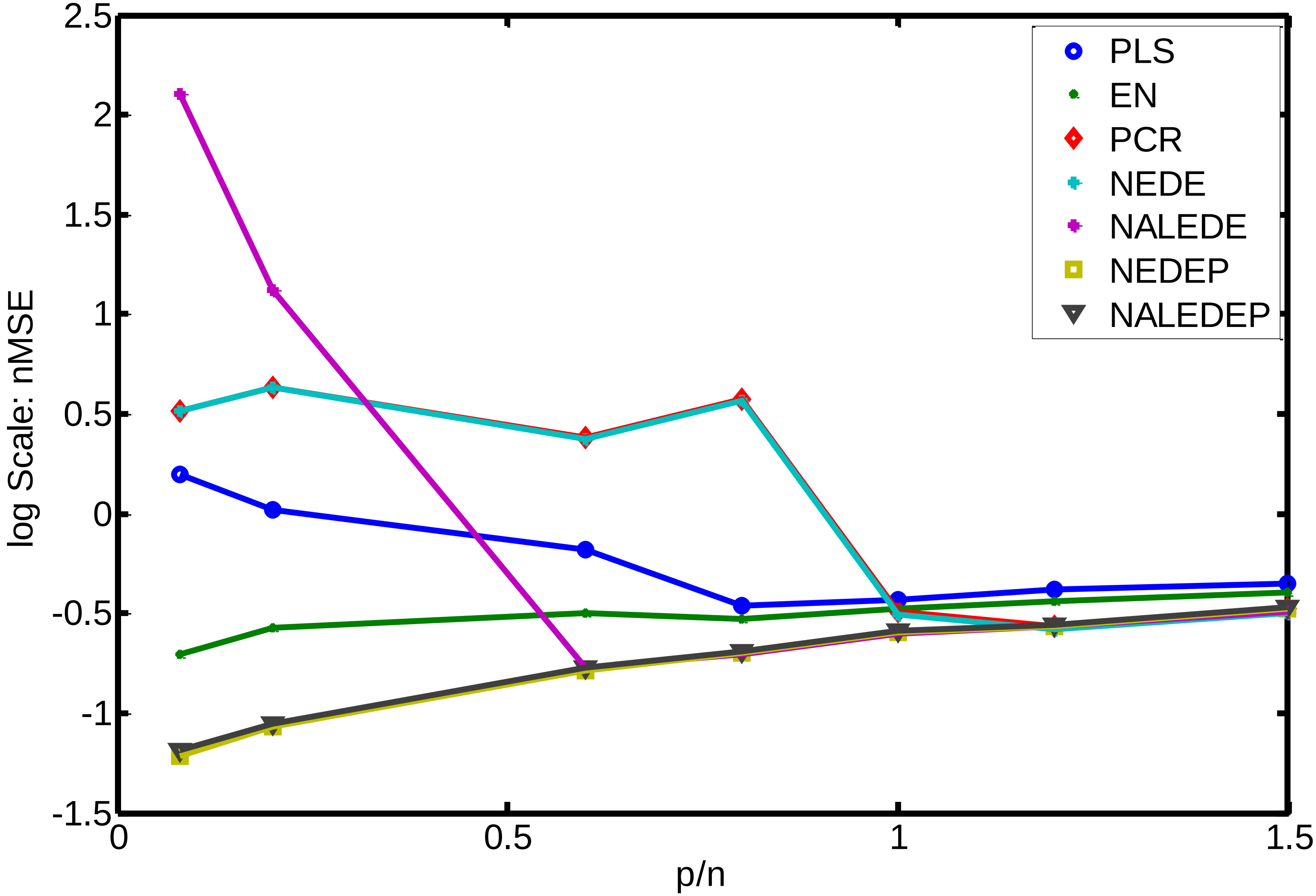
Questions

- Anil – what is the diagram on slide 74 of quals pres?
- Mark – hb, kr well known interaction?
- Anil: NEDE is equivalent to an optimization formulation of principal components regression; Elastic net is equivalent to NALEDE in which the data is pure noise (no manifold) – explain clearly what is different

Comparison to Previous Work

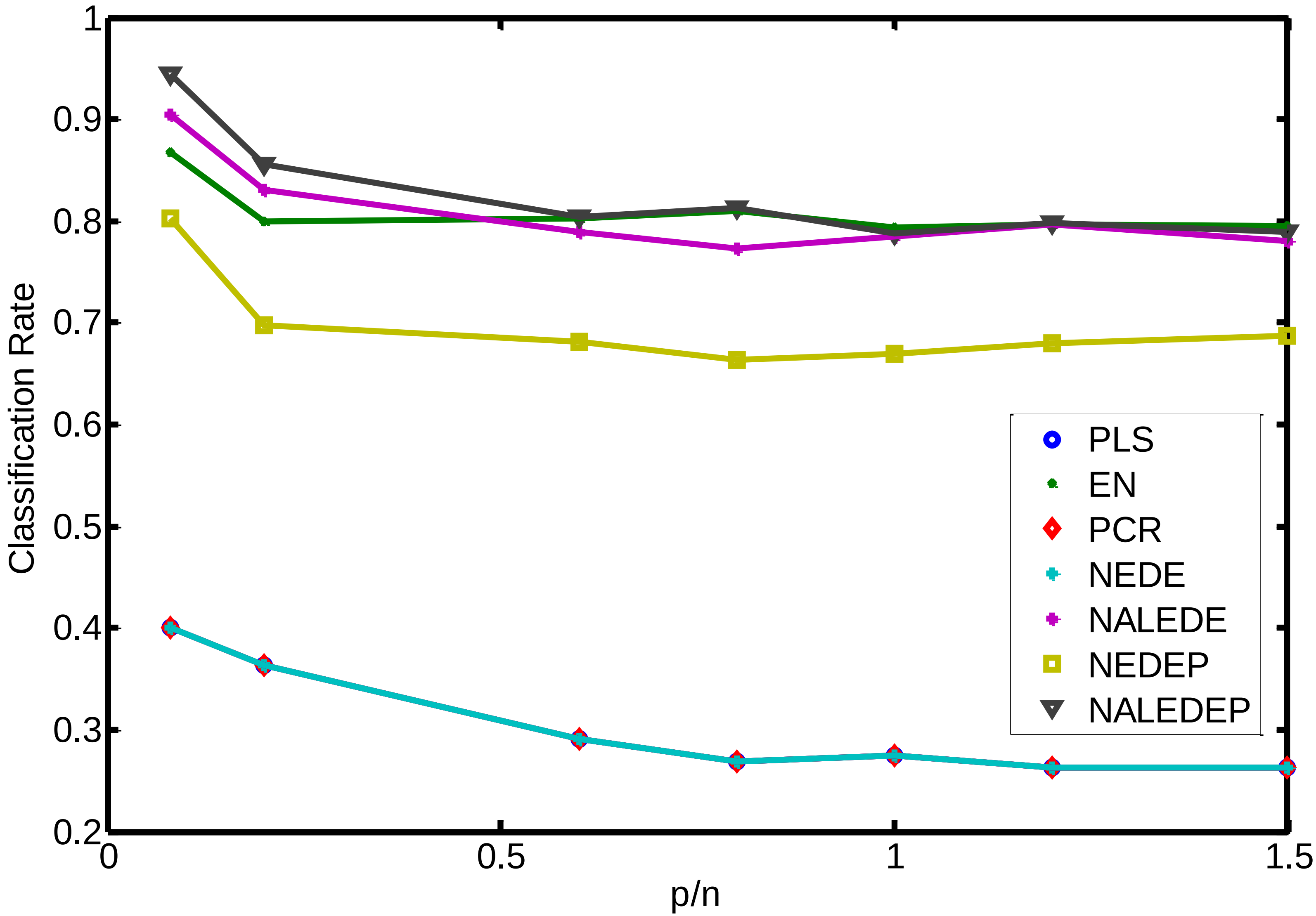
- NEDE is equivalent to an optimization formulation of principal components regression
- Elastic net is equivalent to NALEDE in which the data is pure noise (no manifold)
- Combines positive aspects of different estimators
- Computational effort comparable to that of existing estimators

Simulation Results – Normalized Mean Squared Error



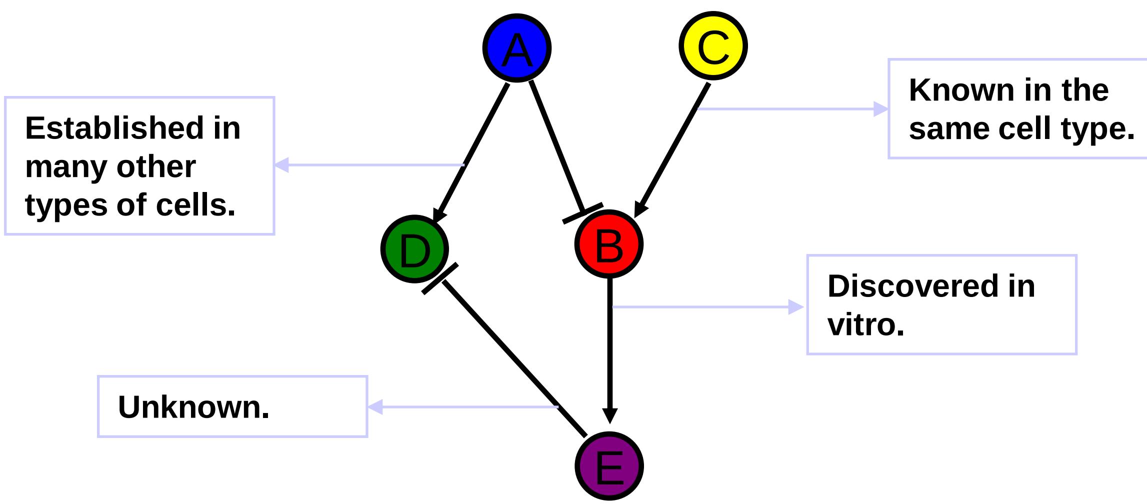
n = 50

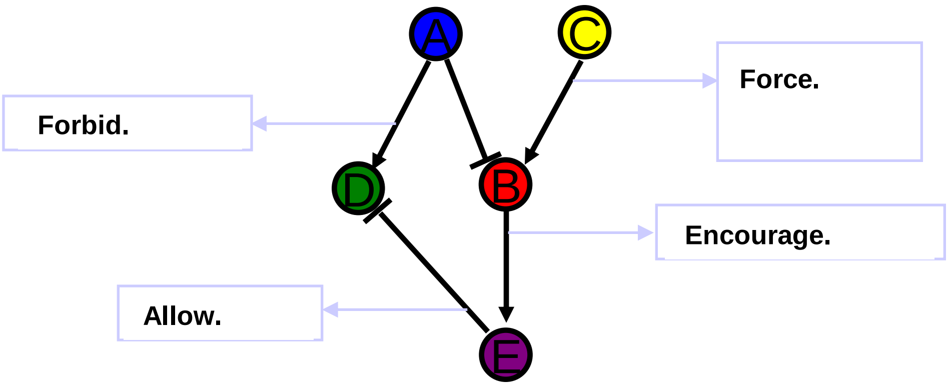
Simulation Results – Classification Rate



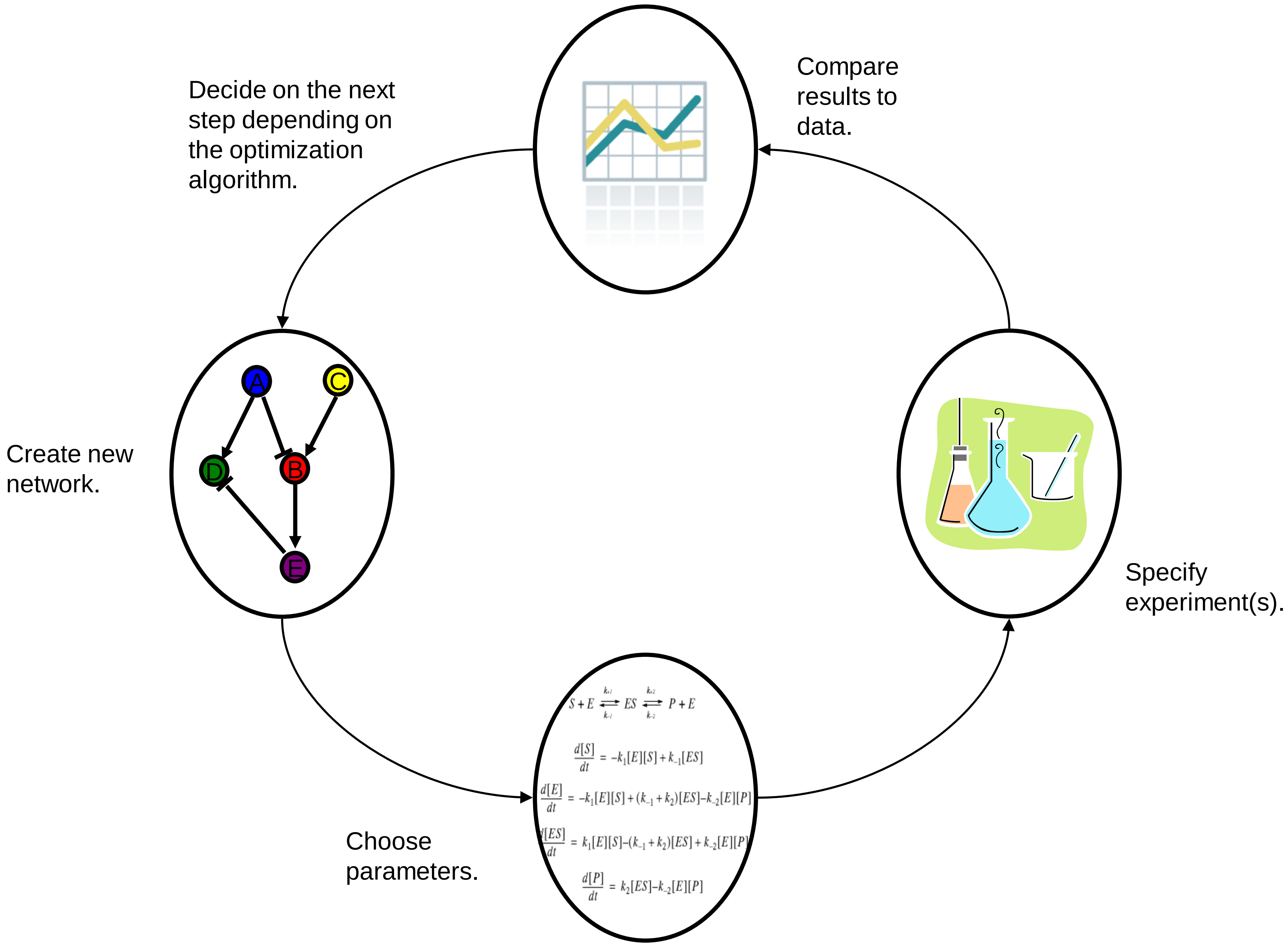
n = 50

- In previous models of the HER2/3 signaling pathway, the structure was fixed a priori (from biological knowledge.)
  - Structure from different cell types, animals, and in vitro experiments was used. These do not necessarily hold.
  - Some parts of the structure might not have been discovered to date.
- We therefore need to search for the correct structure, and not only the parameters given a certain structure.





- To include structure modifications in the optimization, we introduce a module that creates possible networks, in a controlled fashion, and a module that creates different experiments.
  - Connections in a network can be forced, prohibited, encouraged, or discouraged.


Decide on the next step depending on the optimization algorithm.
Compare results to data.
Create new network.
A
C
D
B
E
Specify experiment(s).
Choose parameters.