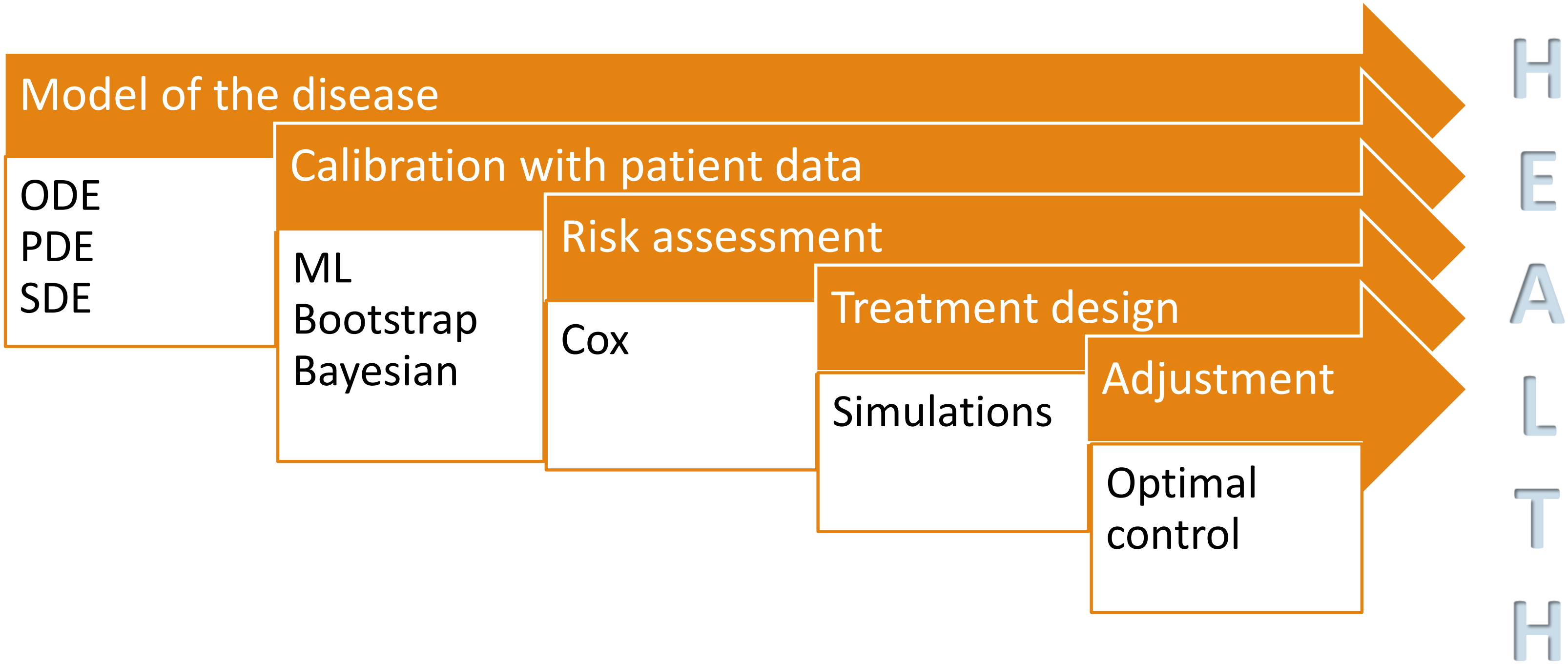
Ensemble Modeling

DAVID SWIGON

UNIVERSITY OF PITTSBURGH

Biomedical systems

GOAL: individualized medicine – designing patient-specific treatment


Model of the disease

ODE
PDE
SDE

Calibration with patient data

ML
Bootstrap
Bayesian

Risk assessment

Cox

Treatment design

Simulations

Adjustment

Optimal control

Problems with biomedical model calibration

Model complexity

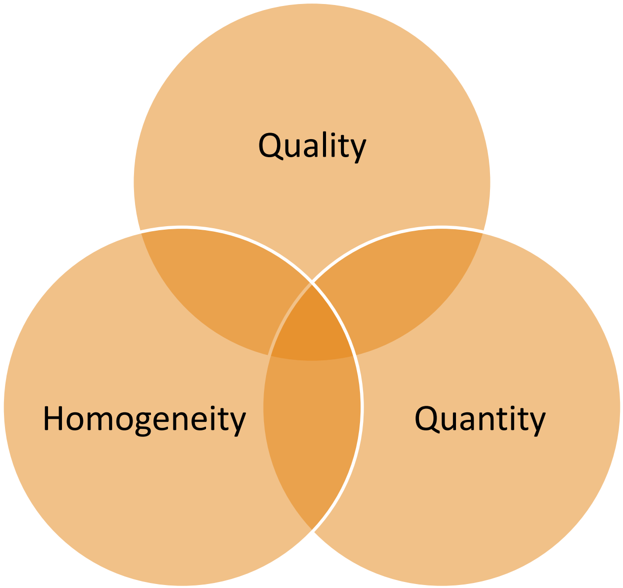
- Large number of variables and parameters, no symmetries
- Complex or unknown interactions

Data scarcity

- Unobserved variables
- Insufficient time resolution
- Small number of trajectories

Data inhomogeneity

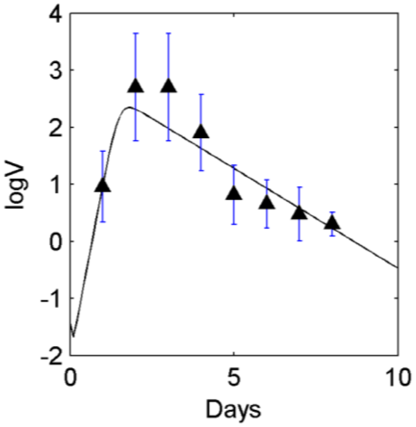
- Variability of parameters across subjects
- Multiple subjects combined in one data set
- Data aggregation



DATA TRADEOFF

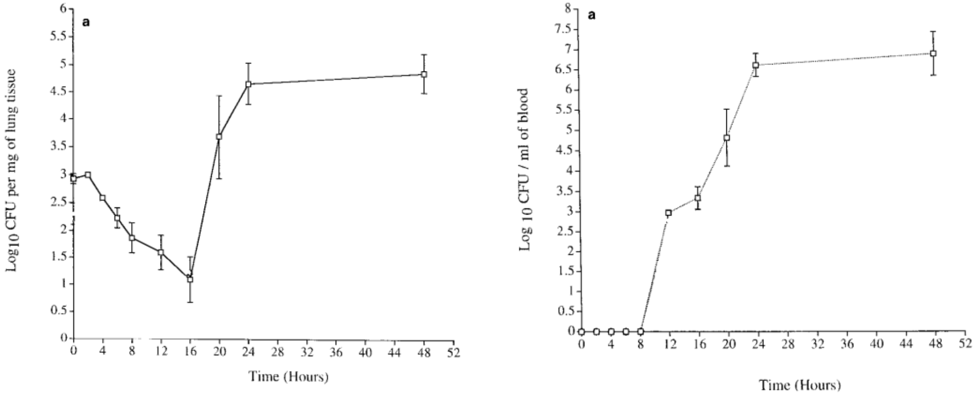
Examples

INFLUENZA



Time courses for 16 human volunteers infected by Influenza A/Texas/91 (H1N1) [Hayden , JAMA, 96]

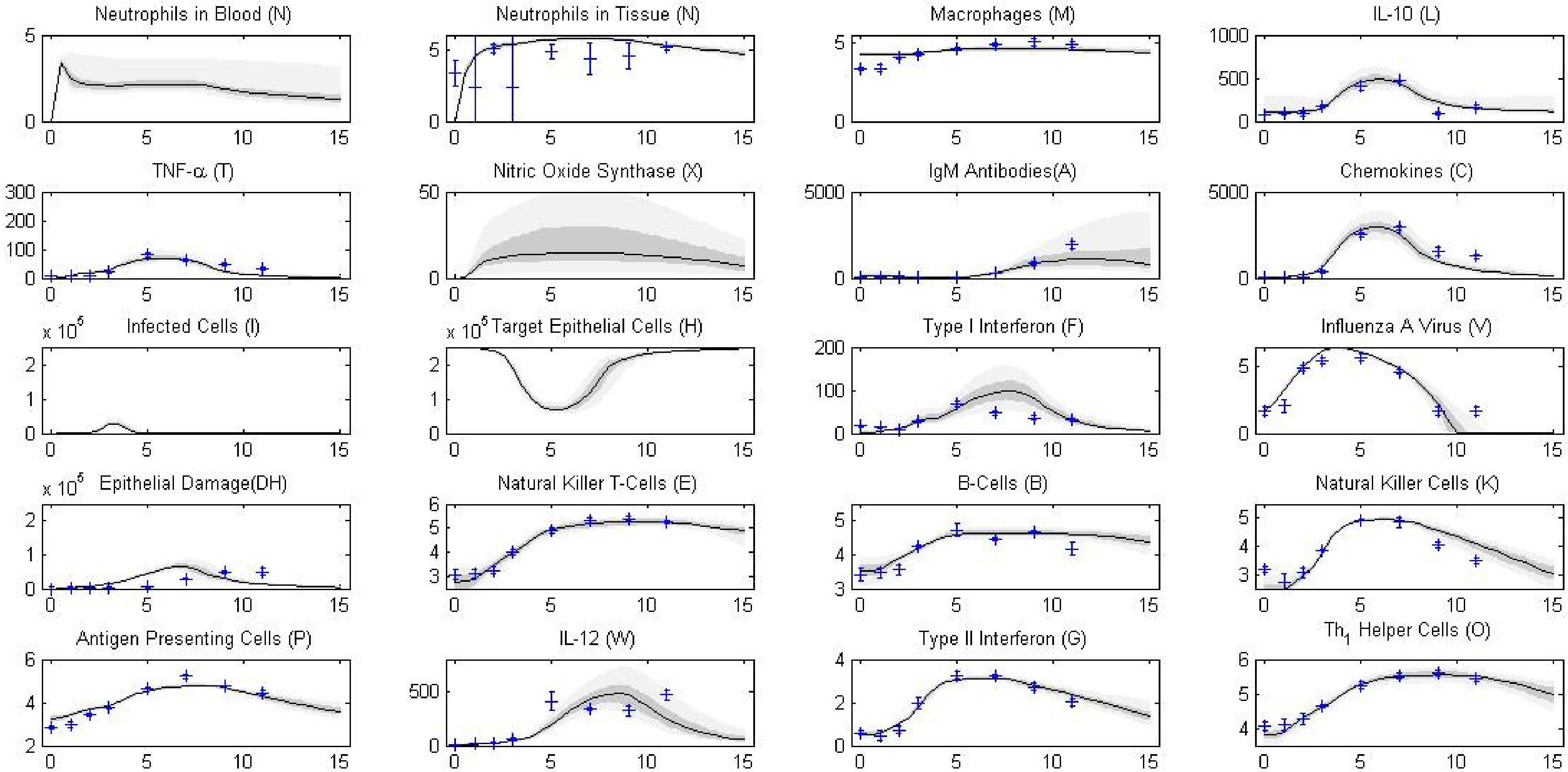
BACTERIAL PNEUMONIA



Data from MF1 mice subjects infected by D39 strain of pneumococcus, data are available for various other mice and pneumonia strains
[Kadioglu et al., Inf. and Immun., 2000]:

INFLUENZA

20 variables, 97 parameters, ODE model



Intranasal infection of 136 Balb/c mice by Influenza virus A/PR/8/34 (H1N1)
[Toapanta, Ross, Morel]

Traditional modeling

Dynamical model

$$\dot{V} = pI - cV \qquad V(0) = V_0$$

$$\dot{H} = -\beta HV \qquad H(0) = 4\times 10^8$$

$$\dot{I} = \beta HV - \delta I \qquad I(0) = 0$$

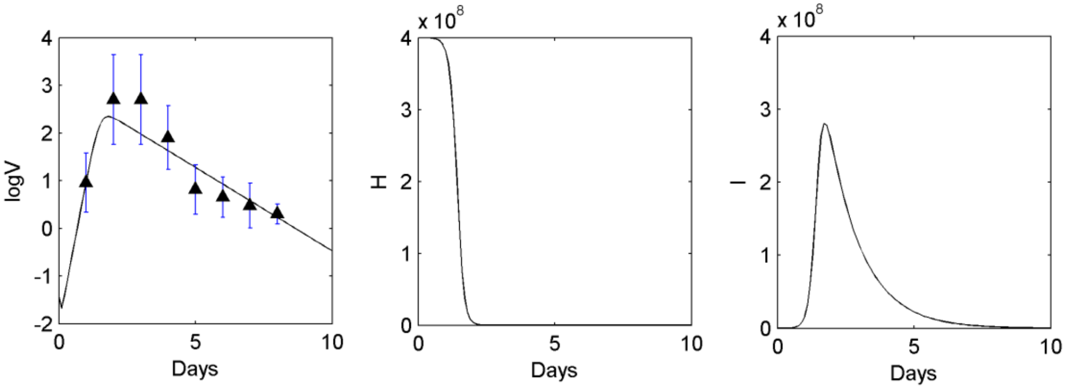
Data

$$V(k) = V_k, \quad \sigma_V(k) = s_k$$

Error minimization

$$\min_{V_0,\beta,p,c,\delta} \sum_{k=1}^{8} \frac{(V(k)-V_k)^2}{s_k^2}$$

Parameter estimates $(V_0, \beta, p, c, \delta)_{opt} = (0.0373, 0.0347, 1.33\times 10^{-5}, 16.7, 0.806)$



Questions

- Is there a parameter set/model that reproduces the data exactly?
- Is the optimal parameter set unique?
- How sensitive is the optimal parameter set to data, their uncertainty, and assumptions about error distribution?
- How well does the model fit the data? Is it even appropriate? Can the model be ruled out at the given data uncertainties?
- How large is the region in data space for which the model is appropriate?
- Since the data consist of multiple subjects, is the model appropriate and how do we know?

Answer

- Analyze more closely the relation between the system (parameters) and its solutions.
- Use probabilistic methods to describe the cohort of subjects or the parameter estimation uncertainty

Ensemble modeling

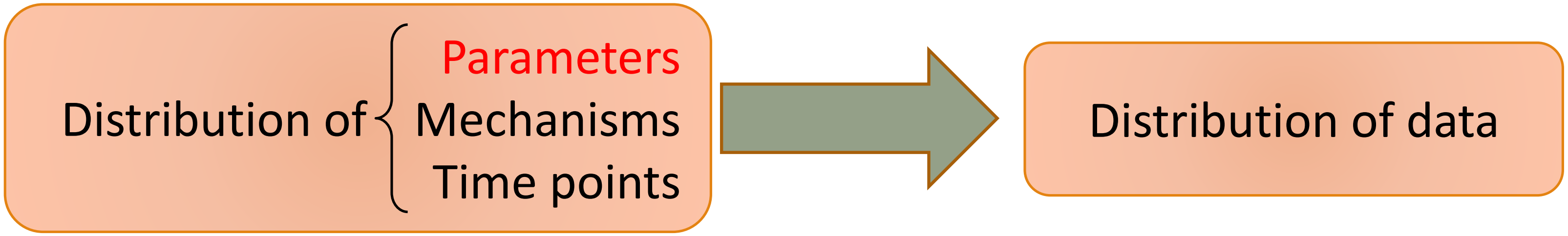
Deterministic model

Parameters
Mechanisms
Time points

Solution map

Trajectory data

Ensemble model



Distribution of { Parameters
Mechanisms
Time points

Distribution of data

Distribution describes

- Variability of parameters across cohort of subjects
- Uncertainty in determination of parameters from noisy data

ODE ensemble model

Initial value problem:

$$\dot{x} = f(x, a), \qquad x(0) = b$$

Solution: $x = x(t; a, b)$

Deterministic solution map: $y = F(a) = (x(t_1, a), \ldots, x(t_n, a))$

Parameter distribution - random variable A, density $\rho(a)$

Data distribution - random variable $Y = F(A)$, density $\eta(y)$

Value of a variable at fixed time: $X_t = x(t; A)$

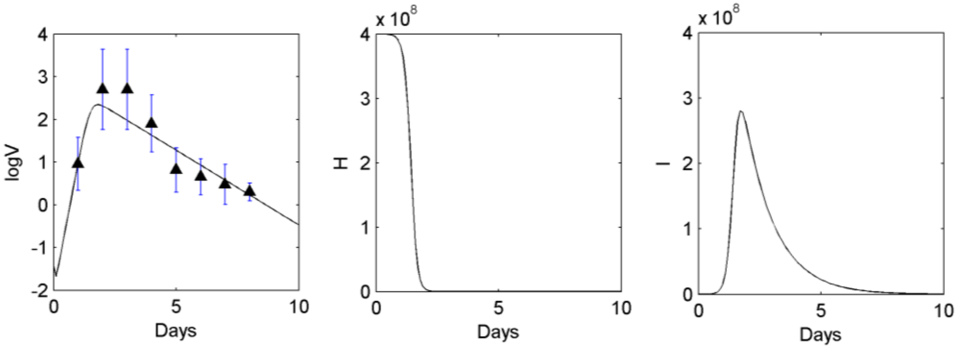
Influenza model

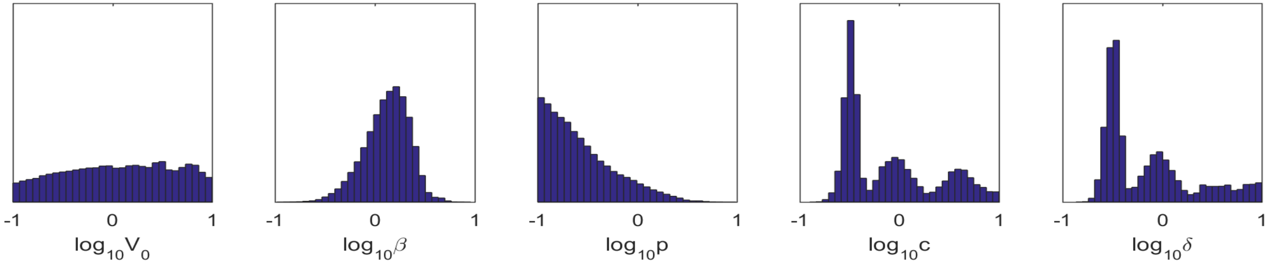
$$\dot{V} = pI - cV$$
$$\dot{H} = -\beta HV$$
$$\dot{I} = \beta HV - \delta I$$

$$a = (V_0, \beta, p, c, \delta)$$
$$y = (V(1), \dots, V(8))$$

$$(V_0, \beta, p, c, \delta)_{opt} = (0.0373, 0.0347, 1.33\times10^{-5}, 16.7, 0.806)$$

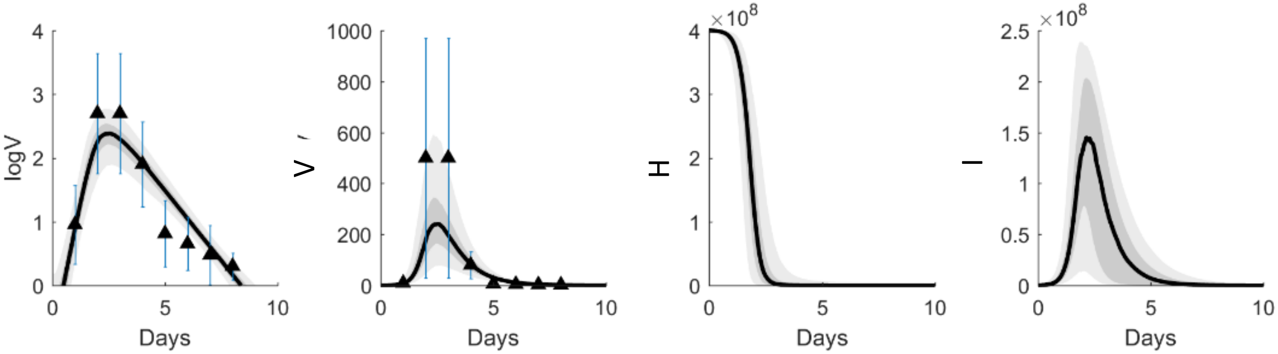
deterministic



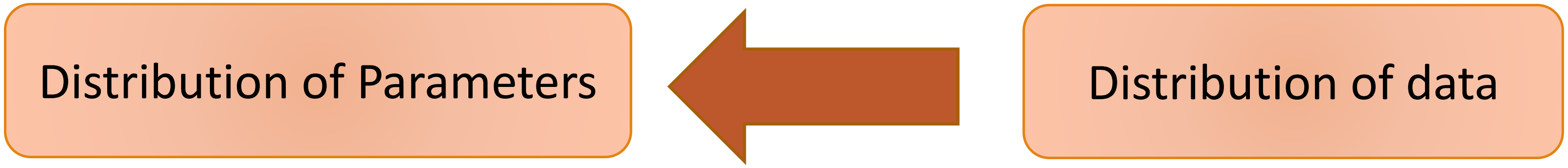


$$(V_0, \beta, p, c, \delta)_{base} = (0.25, 0.014, 2.7\times10^{-5}, 3.2, 3.2)$$

ensemble



Constructing ensemble models

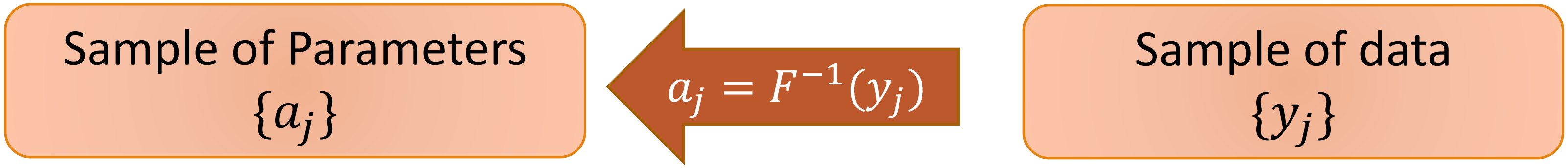


Distribution of Parameters

Distribution of data

(1) Inverse map of data sample

- Requires computation of F^{-1}, solution of the **ODE inverse problem**



Sample of Parameters
$\{a_j\}$

$a_j = F^{-1}(y_j)$

Sample of data
$\{y_j\}$

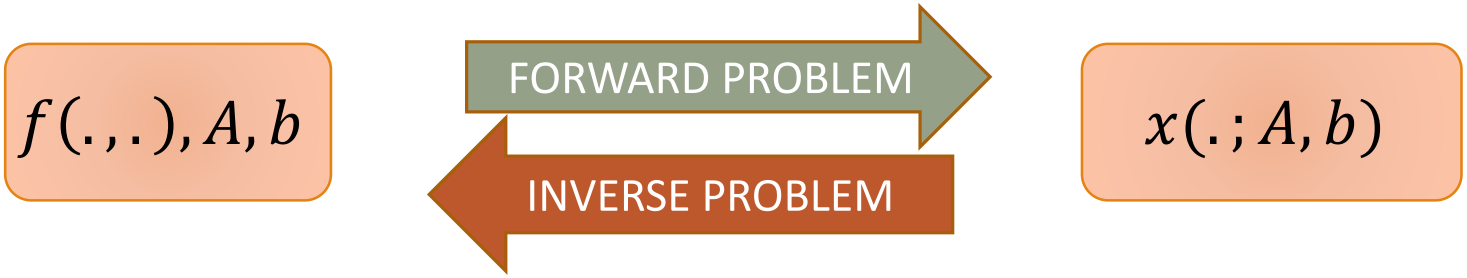
(2) Direct sampling of parameter distribution

- Requires formula for parameter density

ODE models

Initial value problem with matrix parameter A :

$$\dot{x} = f(x, A), \qquad x(0) = b$$



INVERSE PROBLEM: Determine the function $f(.,.)$ and/or parameters A, b from data about trajectory(ies) $x(.; A, b)$.

Inverse problem

$f(.,.), A, b$

INVERSE PROBLEM

$x(.;A,b)$

Mathematical analysis

- Existence, uniqueness/identifiability
- Robustness (existence/uniqueness on open sets)
- Construction of the inverse map (numerical algorithm)
- Maximal permissible uncertainty
- Prediction uncertainty

Existing results on ODE inverse problem

Observability theory

- Search for conditions that guarantee identifiability of the system state (initial conditions) from observed trajectory (can be extended to parameters)
- Kalman (1963) formulated observability conditions for linear dynamical systems
- Griffith & Kumar (1971), Kou, Elliot, & Tarn (1973) formulated sufficient conditions for observability of nonlinear dynamical systems
- Identification of analytic systems with p parameters requires at most $2p + 1$ observations [Aeyels 1981; Sontag 2002]

Polynomial systems

- Reconstruction of polynomial fields from known algebraic attractors [Sverdlove 1980,81]
- Analysis of all possible topologically invariant planar quadratic dynamical systems [Perko textbook]

Existing results on inverse problem

Algorithms for finding inverse

- Numerical simulation combined with nonlinear least squares fitting
- Multiple shooting combined with Gauss-Newton minimization [Bock 1983; Lee]
- Collocation methods using basis function expansion, combined with NLS fitting [Varah 1982; Ramsey et al 2007]
- Contraction map and ‘collage method’ [Kunze et al. 2004]
- Parameter-free approach (using Takens map) [Sauer, Hamilton, …]
- Compressed sensing

Probabilistic approaches

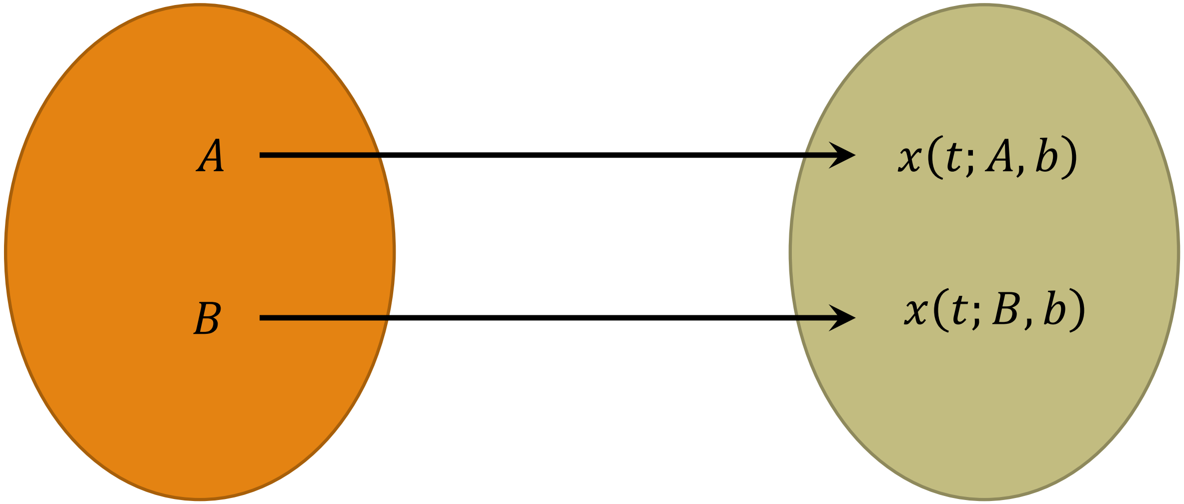
- Transformation of measure
- Numerical simulation combined with Bayesian inference
- Bayesian integration of ODEs [Chkrebtii et al.]

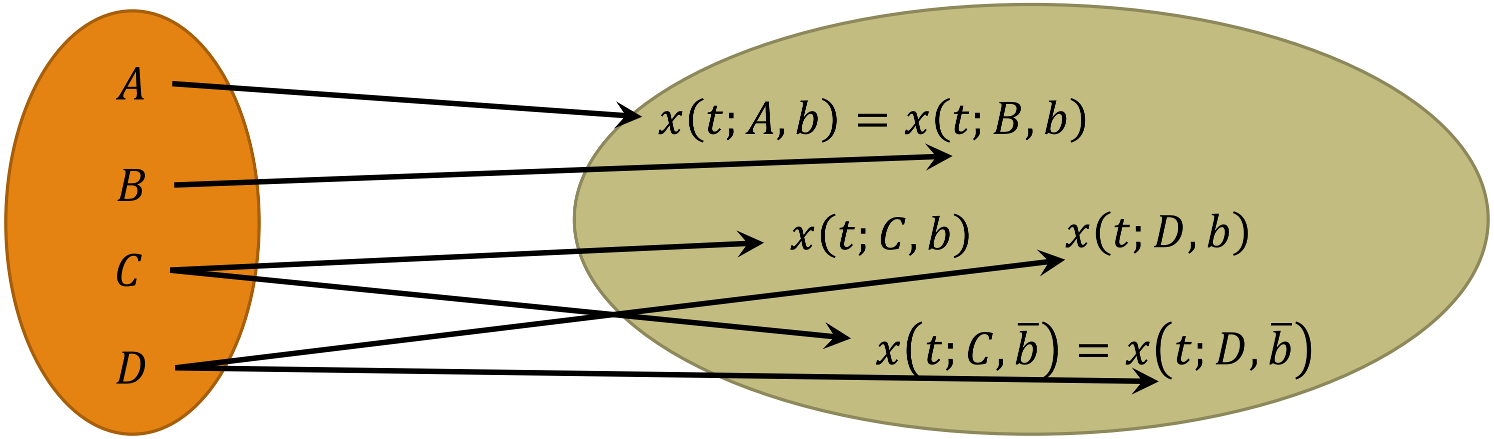
Identifiability from single complete trajectory

System: $\dot{x} = f(x, A)$

Initial condition: $x(0) = b$

Solution: $x(t; A, b)$





A

B

C

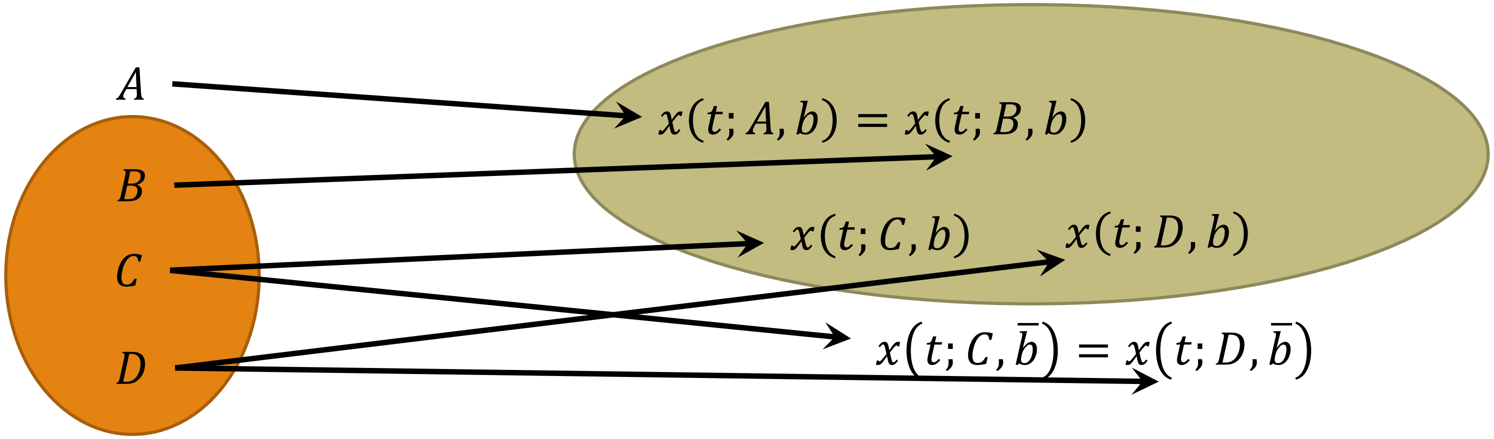
D

$x(t; A, b) = x(t; B, b)$

$x(t; C, b)$

$x(t; D, b)$

$x(t; C, \bar{b}) = x(t; D, \bar{b})$



A

B

C

D

$x(t; A, b) = x(t; B, b)$

$x(t; C, b)$

$x(t; D, b)$

$x(t; C, \bar{b}) = x(t; D, \bar{b})$

Definitions

System $\dot{x} = f(x, A)$ is

Identifiable in $\Omega \subseteq \mathbb{R}^{n \times m}$ iff for all $A, B \in \Omega$ with $A \neq B$ there exists $b \in \mathbb{R}^n$ such that $x(t; A, b) \neq x(t; B, b)$ for some $t > 0$.

Identifiable in $\Omega \subseteq \mathbb{R}^{n \times m}$ **from** b iff for all $A, B \in \Omega$ with $A \neq B$ it holds that $x(t; A, b) \neq x(t; B, b)$ for some $t > 0$.

Identifiable in $\Omega \subseteq \mathbb{R}^{n \times m}$ **from** $x(t; A, b)$ iff there exists no $B \in \Omega$ with $A \neq B$ such that $x(t; A, b) = x(t; B, b)$ for all t.

Unconditionally identifiable in $\Omega \subseteq \mathbb{R}^{n \times m}$ iff for all $A, B \in \Omega$ with $A \neq B$ and all $b \in \mathbb{R}^n$ it holds that $x(t; A, b) \neq x(t; B, b)$ for some $t > 0$.

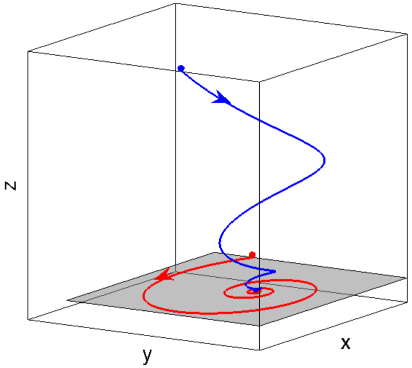
Results for linear system

Theorem [SRS14]:
System $\dot{x} = Ax$ *is* ***identifiable in*** $\mathbb{R}^{n \times m}$ **from** $x(t; A, b)$ *if and only if the orbit* $\gamma(A, b) = \{x(t; A, b), t \in \mathbb{R}\}$ *is not confined to a proper subspace of* $\mathbb{R}^n$*.*

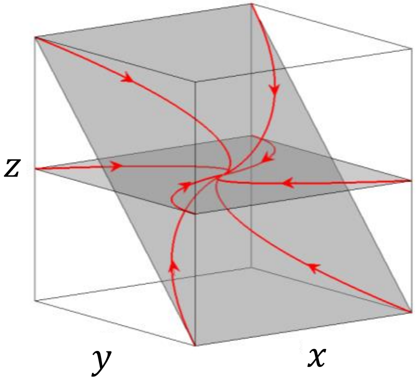
Proof sketch: By theorem of Astrom, identifiability from *b* is equivalent to $\det[b|Ab| \ldots |A^{n-1}b] \neq 0$. It suffices to show that this condition is equivalent to the orbit $\gamma(A, b)$ not being confined to a proper subspace of $\mathbb{R}^n$.

Examples: Inverse problem solution is unique only for certain trajectories/data.

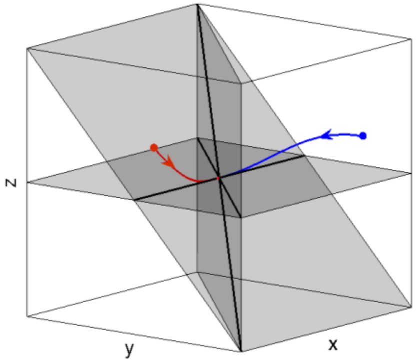
$$\begin{bmatrix}\dot{x}\\ \dot{y}\\ \dot{z}\end{bmatrix} = \begin{bmatrix}-0.2 & -1 & 0\\ 1 & -0.2 & 0\\ 0 & 0 & -0.3\end{bmatrix}\begin{bmatrix}x\\ y\\ z\end{bmatrix}$$



$$\begin{bmatrix}\dot{x}\\ \dot{y}\\ \dot{z}\end{bmatrix} = \begin{bmatrix}-1 & 1 & 0\\ 0 & -1 & 0\\ 0 & 0 & -1\end{bmatrix}\begin{bmatrix}x\\ y\\ z\end{bmatrix}$$



$$\begin{bmatrix}\dot{x}\\ \dot{y}\\ \dot{z}\end{bmatrix} = \begin{bmatrix}-1 & -3 & 2\\ 0 & -4 & 2\\ 0 & 0 & -2\end{bmatrix}\begin{bmatrix}x\\ y\\ z\end{bmatrix}$$



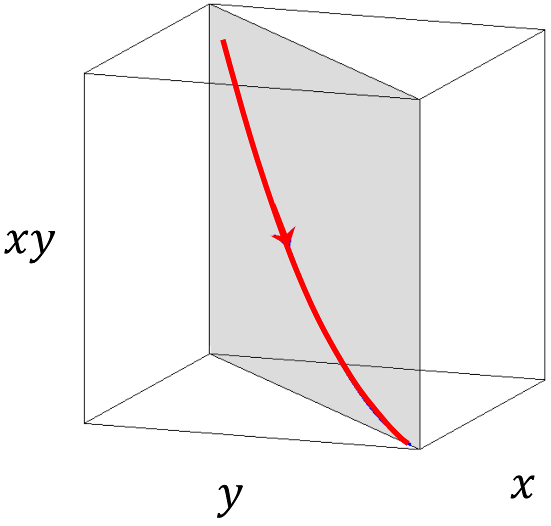
Red trajectories are **confined to linear subspaces of the flux space** => they **do not identify** the system

[SRS14]

Result for linear-in-parameter system

Theorem [SRS14]:
System $\dot{x} = Ag(x)$ *is* ***identifiable in*** $\mathbb{R}^{n\times m}$ ***from*** $x(t; A, b)$ *if and only if the image* $g(\gamma(A,b))$ *of the orbit* $\gamma(A,b)$ *is not confined to a proper subspace of* $\mathbb{R}^m$*.*

$$\begin{bmatrix}\dot{x}\\ \dot{y}\end{bmatrix} = \begin{bmatrix}-2 & 1 & 0\\ 0 & 1 & -3\end{bmatrix}\begin{bmatrix}x\\ xy\\ y\end{bmatrix} \qquad b = \begin{bmatrix}1\\ 1\end{bmatrix}$$



Trajectory is **confined** => System is **not identifiable**!

System with the same trajectory:

$$\begin{bmatrix}\dot{x}\\ \dot{y}\end{bmatrix} = \begin{bmatrix}-2+\alpha & 1 & -\alpha\\ \beta & 1 & -3-\beta\end{bmatrix}\begin{bmatrix}x\\ xy\\ y\end{bmatrix}$$

Result for confined trajectory

Theorem [SRS14]:
Suppose that V is a proper linear subspace $\mathbb{R}^n$ invariant under A. The following are equivalent

(i) V is the minimal A-invariant subspace such that $b \in V$.
(ii) The orbit of $x(t; A, b)$ is not confined to a proper subspace of V
(iii) There exists no $B \in \mathbb{R}^{n\times n}$ such that $B|_V \neq A|_V$ and $x(t; A, b) = x(t; B, b)$.

$$A = \begin{bmatrix} -1 & -3 & 2 \\ 0 & -4 & 2 \\ 0 & 0 & -2 \end{bmatrix} \qquad b = \begin{bmatrix} -1 \\ -1 \\ 1 \end{bmatrix} \qquad V = span\left\{\begin{bmatrix} 1 \\ 1 \\ 0 \end{bmatrix}, \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix}\right\} \qquad \mathscr{B} = \left\{\begin{bmatrix} 1 \\ 1 \\ 0 \end{bmatrix}, \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix}, \begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix}\right\}$$

$$[A]_{\mathscr{B}} = \left[\begin{array}{cc} \overset{A|_V}{\boxed{\begin{matrix} -4 & 2 \\ 0 & -2 \\ 0 & 0 \end{matrix}}} & \begin{matrix} -3 \\ 0 \\ -1 \end{matrix} \end{array}\right] \qquad [B]_{\mathscr{B}} = \begin{bmatrix} -4 & 2 & \beta_{11} \\ 0 & -2 & \beta_{12} \\ 0 & 0 & \beta_{13} \end{bmatrix} \qquad B = \begin{bmatrix} -\beta_{13} - 4 & \beta_{13} & 2 \\ \beta_{13} - \beta_{33} - 4 & \beta_{13} + \beta_{33} & 2 \\ -\beta_{23} & \beta_{23} & -2 \end{bmatrix}$$

Numerical determination of confinement

Numerical errors (roundoff) can obscure the confinement of trajectory represented as a collection of data points

Use Singular Value Decomposition

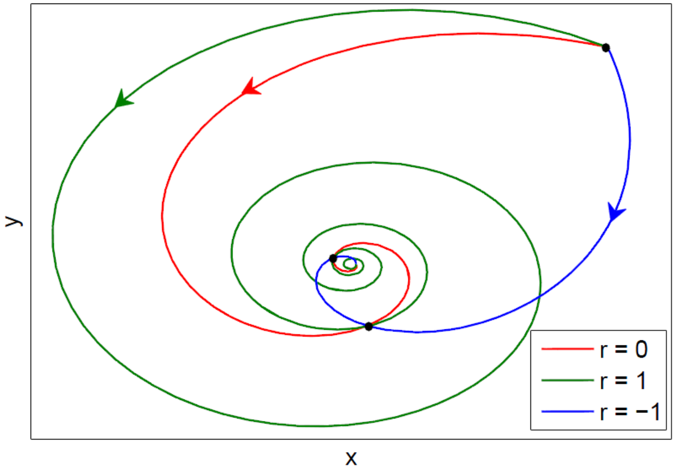
$$X = USV^T$$

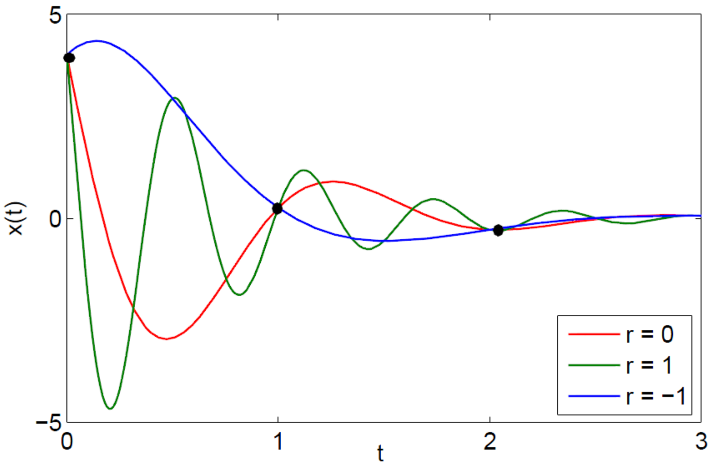
- Dimension of confining subspace = $rank(X) = \#\{S_{jj} > 0\}$
- Numerical rank: $rank_{num}(X) = \#\{S_{jj} / \max_j S_{jj} > \varepsilon_{tol}\}$

Identification from discrete data on a single trajectory

Discrete data: $X = (x_0, x_1, \dots, x_n), \quad x_i = x(t_i; A, b)$

Identifiability from complete trajectory does not imply identifiability from discrete data.





$$A = \begin{bmatrix} -3/2 & -4 \\ 4 & -3/2 \end{bmatrix} + 2\pi r \begin{bmatrix} 0 & -1 \\ 1 & 0 \end{bmatrix}$$

$$e^A = \begin{bmatrix} 0.1458 & -0.1689 \\ 0.1689 & 0.1458 \end{bmatrix}$$

Identification from discrete data on a single trajectory

GOAL:

Partition the data space into domains in which

- the system is **identifiable** from data
- the system is **robustly identifiable** (on an open neighborhood)
- the system is **identifiable to within countable number** of alternatives
- the system has a **continuous family** of compatible parameters
- there is **no parameter set** corresponding to the data
- the system has a **specific dynamical behavior** (stable, saddle, …)

Inverse map for linear systems

Discrete data (uniformly spaced): $d = (x_0, x_1, \dots, x_n)$

Solution map: $x_k = e^{Ak}b$

ALGORITHM

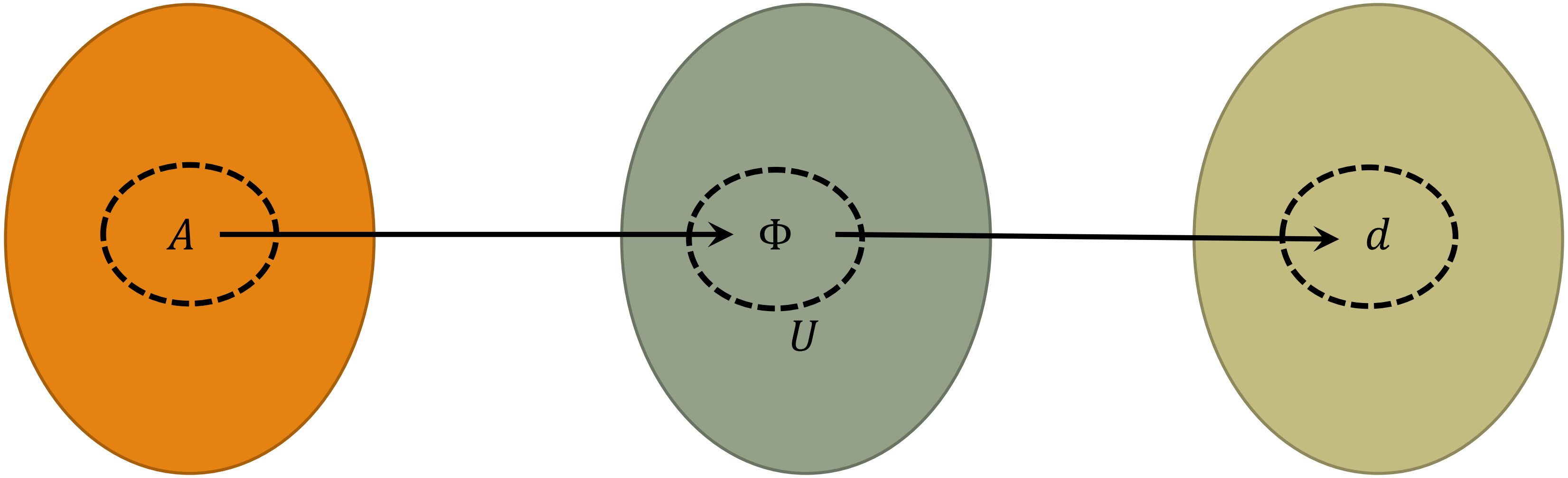
Step 1: Let $X_k = [x_k | x_{k+1} | \dots | x_{n+k-1}], \quad k = 0,1$

Step 2: Compute $\Phi = X_1 X_0^{-1}$

Step 3: Solve $e^A = \Phi$

Existence and uniqueness of A derives from conditions for existence and uniqueness of real matrix logarithm [Culver 1966].

Robust existence & uniqueness



Robust existence & uniqueness

Theorem [SRS17]: *There exists an open set* $U \subset \mathbb{R}^{n\times n}$ *containing* Φ *such that for any* $\Psi \in U$ *the equation* $e^A = \Psi$ ***has a solution*** $A \in \mathbb{R}^{n\times n}$ *iff*
(a) $\{x_0, x_1, \dots, x_{n-1}\}$ *are linearly independent and*
(b) Φ *has* ***only positive real or complex eigenvalues****.*

Theorem [SRS17]: *There exists an open set* $U \subset \mathbb{R}^{n\times n}$ *containing* Φ *such that for any* $\Psi \in U$ *the equation* $e^A = \Psi$ ***has a unique solution*** $A \in \mathbb{R}^{n\times n}$ *iff*
(a) $\{x_0, x_1, \dots, x_{n-1}\}$ *are linearly independent and*
(b) Φ *has* n ***distinct positive real eigenvalues****.*

Theorem [SRS17]: *There exists an open set* $U \subset \mathbb{R}^{n\times n}$ *containing* Φ *such that for any* $\Psi \in U$ *the equation* $e^A = \Psi$ ***does not have a solution*** $A \in \mathbb{R}^{n\times n}$ *iff*
(a) $\{x_0, x_1, \dots, x_{n-1}\}$ *are linearly independent and*
(b) Φ *has* ***at least one negative real eigenvalue of odd multiplicity****.*

Region of unique inverses

Let

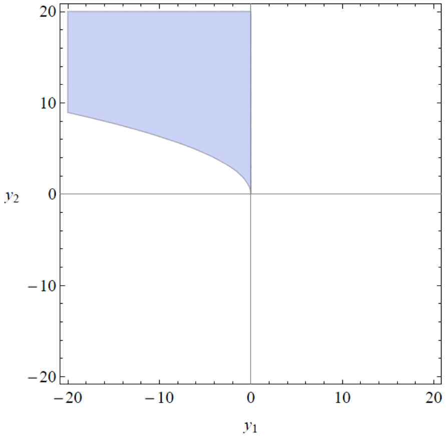
$$\hat{\Phi} = X_0^{-1}\Phi X_0 = X_0^{-1}X_1$$

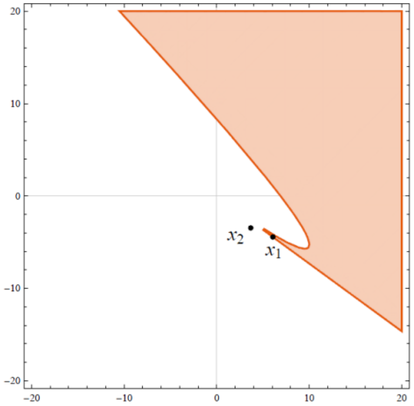
$$\hat{\Phi} = \begin{bmatrix} 0 & 0 & 0 & \cdots & y_1 \\ 1 & 0 & 0 & \cdots & y_2 \\ 0 & 1 & 0 & \cdots & y_3 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \cdots & y_n \end{bmatrix}$$

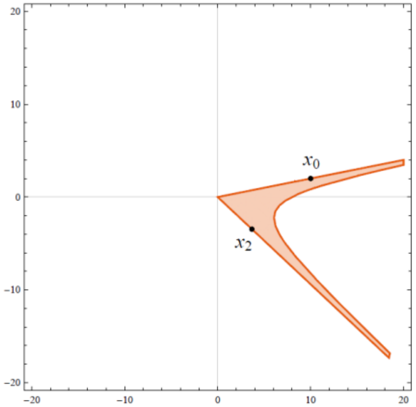
Eigenvalues of Φ are determined by $y = X_0^{-1}x_n$

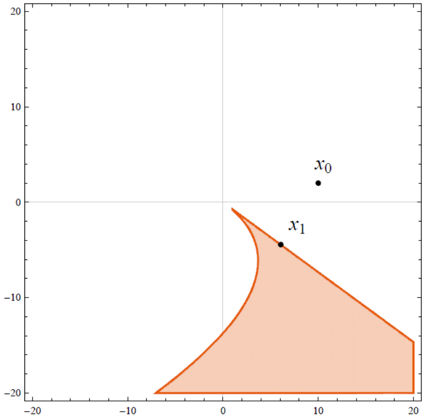
Conditions on *y* that guarantee positive real eigenvalues have been worked out for arbitrary dimension [Gantmacher; Yang]

Example in 2D (regions of data which give unique inverse)

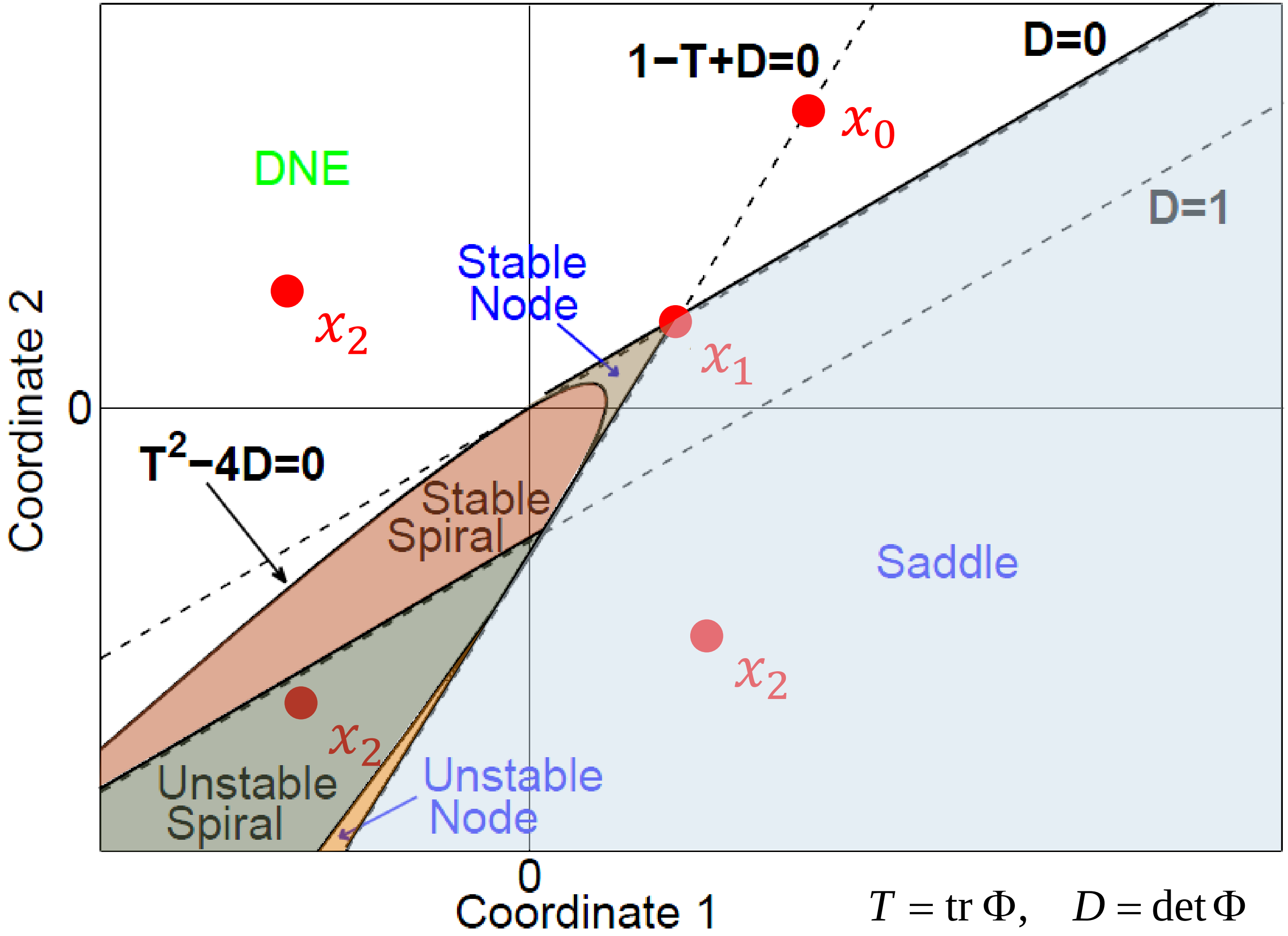








Regions of stability for 2x2 systems



$T = \operatorname{tr}\Phi, \quad D = \det\Phi$

Numerical error of identification

Exact formula

$$\Phi = X_1 X_0^{-1}$$

Numerical sensitivity of the inverse to errors in the data.

$$\frac{\|\delta\Phi\|}{\|\Phi\|} \leq \kappa(X_0)\left(\frac{\|\delta X_0\|}{\|X_0\|} + \frac{\|\delta X_1\|}{\|X_1\|}\right)$$

- Here $\|\cdot\|$ is any matrix norm and $\kappa(C) = \|C\|\|C^{-1}\|$ is the condition number of the matrix C.
- Note that collinearity of vectors in X_0 leads to a loss of accuracy.

Identification from non-uniformly spaced data

Discrete data (non-uniformly spaced)

$$d = \{x_{j_0}, x_{j_1}, \ldots, x_{j_n}\} \qquad x_{j_k} = x(j_k; A, b) = e^{Aj_k} b \qquad x_{k+1} = e^{A} x_k$$

Lemma: *Let* $x_{j_k} = \Phi^{j_k} b$ *where* j_k *are integers such that* $0 = j_0 < j_1 < \cdots < j_n$. *Let* $X_0 = [x_{j_0}, x_{j_1}, \ldots, x_{j_{n-1}}]$ *be invertible. Let* $y = X_0^{-1} x_{j_n}$ *be a vector with entries* $y_1, y_2, \ldots y_n$. *If* Φ *is diagonalizable, then* λ *is an eigenvalue of* Φ *only if it is a root of the polynomial*

$$p_y(\lambda) = \lambda^{j_n} - y_n \lambda^{j_{n-1}} - \cdots - y_2 \lambda^{j_1} - y_1$$

Identification from non-uniformly spaced data

ALGORITHM

Step 1: Get $y = X_0^{-1} x_{j_n}$

Step 2: Find all roots of $p_y(\lambda)$

Step 3: Choose a combination of *n* distinct roots and form $\widehat{\Phi}$.

Step 4: Compute vectors $z_{j_0}, z_{j_1}, \ldots, z_{j_{n-1}}$ as $z_{j_0} = e_1$ and $z_{j_{k+1}} = \widehat{\Phi}^{j_{k+1} - j_k} z_{j_k}$.

Step 5: Compute $\Phi = P^{-1} \widehat{\Phi} P$ where *P* is such that $P x_{j_k} = z_{j_k}$ for $k = 0,1, \ldots, n-1$.

Step 6: Solve $e^A = \Phi$.

Note: The variety of choices for $\widehat{\Phi}$ may lead to non-uniqueness of parameter estimates.

Uncertainty analysis

Uncertain data: $\tilde{d} = (\tilde{x}_0, \tilde{x}_1, \dots, \tilde{x}_n)$

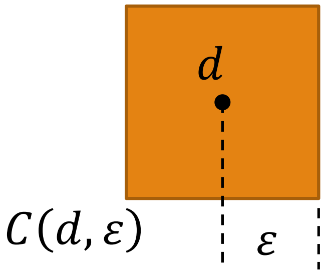
Bounds on uncertainty: $C(d, \varepsilon) = \{\tilde{d}: \max_{i,j} |\Delta x_{i,j}| < \varepsilon\}$

Neighborhood $C(d, \varepsilon)$ is called ***permissible*** for property P iff P is shared by all systems corresponding to the data $\tilde{d} \in C(d, \varepsilon)$.

Value $\varepsilon_P > 0$ is ***maximal permissible uncertainty*** for property P iff $C(d, \varepsilon)$ is permissible for P for all $0 < \varepsilon < \varepsilon_P$, and $C(d, \tilde{\varepsilon})$ is not permissible for P for any $\tilde{\varepsilon} > \varepsilon_P$

P can be

- Existence of inverse (unique or nonunique)
- Unique inverse
- Unique stable inverse



Results for linear systems

Theorem (lower bound): *Let d be such that* Φ *has n distinct positive eigenvalues* $\lambda_1, \lambda_2, \ldots, \lambda_n$. *Let* $m_1 = \frac{1}{2} \min_{i<j} \left|\lambda_i - \lambda_j\right|$, $m_1 = \min_i \lambda_i$, *and* $\delta_U = \min\{m_1, m_2\}$. *If* $0 < \varepsilon \leq \underline{\varepsilon}_U$ *where*

$$\underline{\varepsilon}_U = \frac{\delta_U}{n(\delta_U + 1 + \|\Lambda\|)\|S^{-1}\|\left\|X_0^{-1}S\right\|}$$

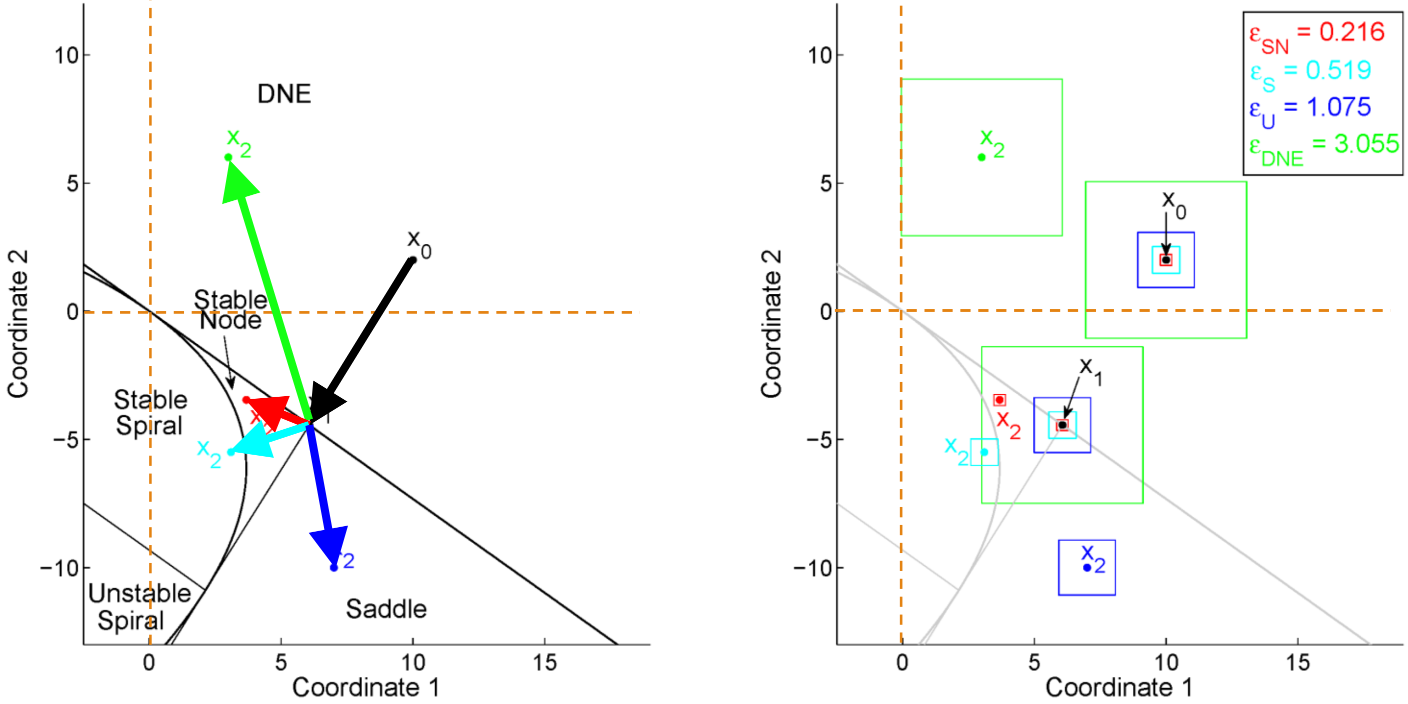
with $\Phi = S\Lambda S^{-1}$, *then for any* $\tilde{d} \in C(d, \varepsilon)$, $\widetilde{\Phi}$ *has n distinct positive eigenvalues.*

Results for linear systems

Theorem (upper bound): *Let d be such that* Φ *has n distinct positive eigenvalues* $\lambda_1, \lambda_2, \dots, \lambda_n$*, and let y be the last column vector of the companion matrix form* $\widehat{\Phi}$ *of* Φ*. The smallest neighborhood* $C(d, \varepsilon)$ *that contains data* $\tilde{d}$ *for which companion matrix* $\widehat{\widetilde{\Phi}}$ *has last column* $\tilde{y}$ *has* $\varepsilon = \bar{\varepsilon}_U$ *where*

$$\bar{\varepsilon}_U = \frac{\|X_0(y - \tilde{y})\|_\infty}{\|\tilde{y}\|_1 + 1}$$

Permissible data uncertainty: Amount of measurement error that can be tolerated without altering inference qualitatively



$$\dot{x} = Ax, \qquad x(0) = b$$

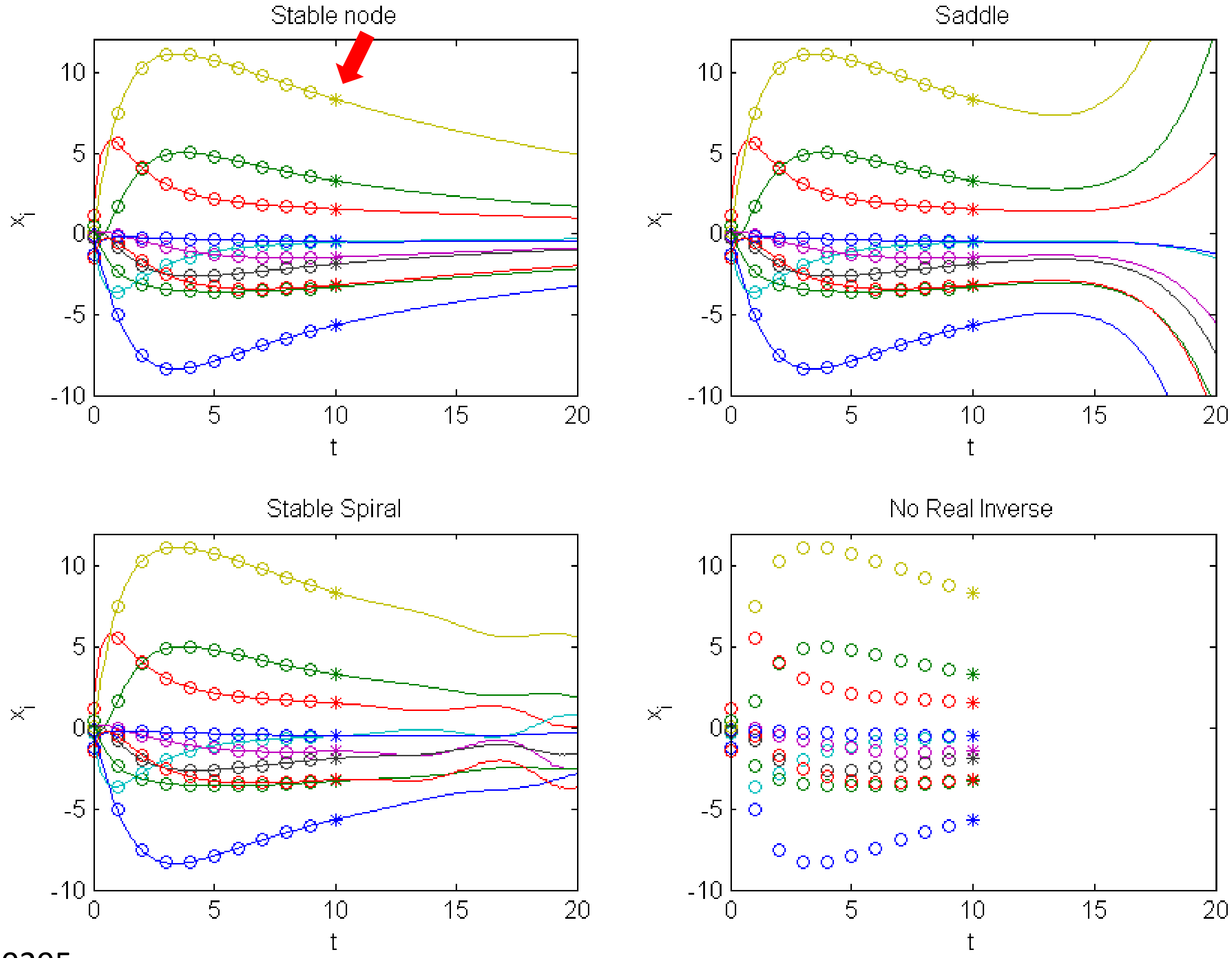
Property (X)	SN	U	S	DNE
x_2	$(3.68, -3.46)$	$(7, -10)$	$(3.1, -5.5)$	$(3, 6)$
$\tilde{\epsilon}_X$ Numerical estimate	0.216	1.075	0.519	3.055
$\underline{\epsilon}_X$	0.059	0.149	0.207	0.609
$\overline{\epsilon}_X$	0.217^*	$1.078^\dagger$	$0.519^\ddagger$	$3.059^\#$

$^*\lambda_1 = \lambda_2$, $^\dagger\lambda_1 = 0$, $^\ddagger\lambda_2 = 1$, $^\#\lambda_2 \to \infty$

Upper and lower bounds for linear systems.

[SRS17]

Prediction Uncertainty: In high dimensions inverse problem solution is very sensitive to small changes in data.



$||\Delta x_n||$=0.0049

$||\Delta x_n||$=0.0295

$||\Delta x_n||$=0.0005

[SRS17]

SUMMARY

Existence and identifiability analysis provides regions on which the inverse problem for ODEs is well posed

Additional solution attributes (e.g., stability, spirality) can be included

OPEN PROBLEMS

Existence of inverse for linear-in-parameter systems

Maximal permissible uncertainty for linear-in-parameter systems

ODE Models with random effects

Initial value problem: $\dot{x} = f(x, a), \quad x(0) = b$

Solution: $x(t; a, b) \in C^1(\mathbb{R}^n)$

Data vector: $y = H(x(t_0), x(t_1), \dots, x(t_n))$

Solution map: $\boldsymbol{y} = \boldsymbol{F}(\boldsymbol{a})$ **invertible for** $\boldsymbol{y} \in \boldsymbol{C}(\overline{\boldsymbol{y}}, \boldsymbol{\varepsilon})$

Random parameter model (RPM):

$Y = F(A)$ A is a r. v. with density $\rho(a)$

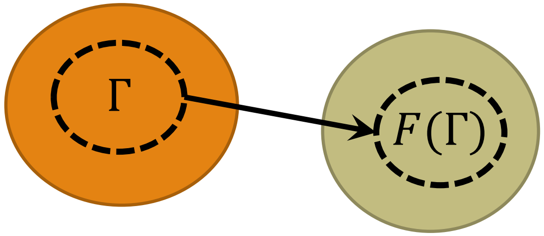
Random measurement error model (REM):

$Y = F(a) + G$ G is a r. v. with density $\gamma(g)$

Random-parameter model

$$Y = F(A)$$

Inverse problem: Find parameter density $\rho(a)$ from the knowledge of data density $\eta(y)$ and F.



Change of variables formula: for any set Γ

$$\int_{\Gamma} \rho(a)da = \int_{F(\Gamma)} \eta(y)dy = \int_{\Gamma} \eta\big(F(a)\big)J(a)da$$

$$J(a) = |\det D_a F(a)|$$

Parameter density: $\rho(a) = \eta\big(F(a)\big)J(a)$

Distribution of parameters: represents subject variability

Computational aspects

Jacobian is expensive to compute.

Methods for computing $J(a)$

- Exact formula for linear models

  For 2 × 2 diagonalizable matrix $A = WMW^{-1}$:

$$J(a) = \frac{(b_2 w_{11} - b_1 w_{21})^2 (b_2 w_{12} - b_2 w_{22})^2 e^{\mu_1 + \mu_2} (e^{\mu_1} - e^{\mu_2})^4}{(\det W)^2 (\mu_1 - \mu_2)^2}$$

- Numerical differentiation

$$\frac{\partial F}{\partial a_j}(\hat{a}) \approx \frac{1}{\varepsilon}\left(F\left(\hat{a} + \varepsilon e_j\right) - F(\hat{a})\right)$$

Requires *n* additional integrations of the system in each step

[SSZR19]

Computational aspects

- Numerical integration of the first variational equations (sensitivity coefficients) for ODE models

Let $s_{a_j}(t) = \frac{\partial x(t;a,b)}{\partial a_j}$, $s_{b_j}(t) = \frac{\partial x(t;a,b)}{\partial b_j}$

$$\dot{x} = f(x,a) \qquad\qquad x(0) = b$$

$$\dot{s}_{a_j}(t) = \left.\frac{\partial f(z,a)}{\partial z}\right|_{z=x(t)} s_{a_j}(t) + \left.\frac{\partial f(z,a)}{\partial a_j}\right|_{z=x(t)} \qquad\qquad s_{a_j}(0) = 0$$

$$\dot{s}_{b_j}(t) = \left.\frac{\partial f(z,a)}{\partial z}\right|_{z=x(t)} s_{b_j}(t) \qquad\qquad s_{b_j}(0) = e_j$$

Dimension of the ODE system increases from n to $n(p+n+1)$

[SSZR19]

Computational aspects

- Broyden's method (rank-1 update)

$$D_a F(\hat{a}) \approx D_a F\left(a^k\right) + \frac{F(\hat{a}) - F\left(a^k\right) - D_a F\left(a^k\right)\Delta a^k}{(\Delta a^k)^T \Delta a^k}\left(\Delta a^k\right)^T$$

Only one computation of $F(\hat{a})$ (integration of the system) needed at every step

Update every 10 computations keeps relative error of $D_a F(\hat{a})$ below 4%

Accuracy of the posterior is lower than the accuracy of Jacobian

[SSZR19]

Examples

Common problem:

- Jacobian calculation is expensive, so can we replace $J(a)$ by a simpler distribution $\pi(a)$?

$$\boxed{\rho(a)} \longleftarrow \boxed{\eta(y)}$$

Linear dynamical system, uniform data distribution $U(\bar{x}_{ij} - \varepsilon, \bar{x}_{ij} + \varepsilon)$

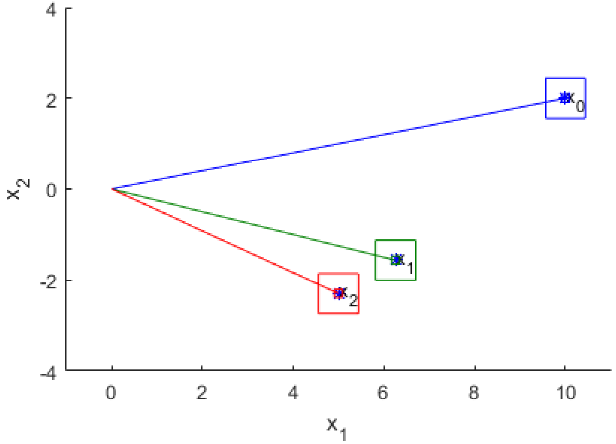
$$\boxed{\dot{x} = \Lambda x}$$

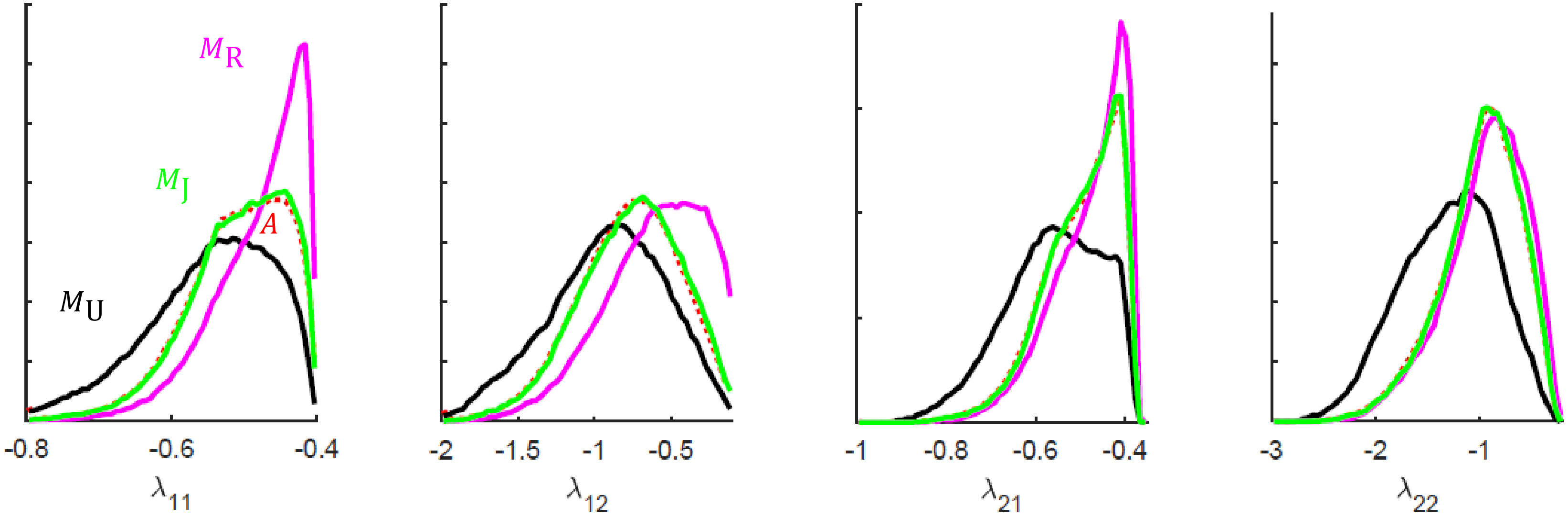
$$d = \left\{ \begin{bmatrix} 10 \\ 2 \end{bmatrix}, \begin{bmatrix} 6.2530 \\ -1.5624 \end{bmatrix}, \begin{bmatrix} 5 \\ -2.3 \end{bmatrix} \right\}$$

$$\varepsilon = 0.436$$

$$\Lambda = \begin{bmatrix} -0.5024 & -0.7491 \\ -0.4839 & -0.9519 \end{bmatrix}$$

$$a = (\lambda_{11}, \lambda_{12}, \lambda_{21}, \lambda_{22})$$





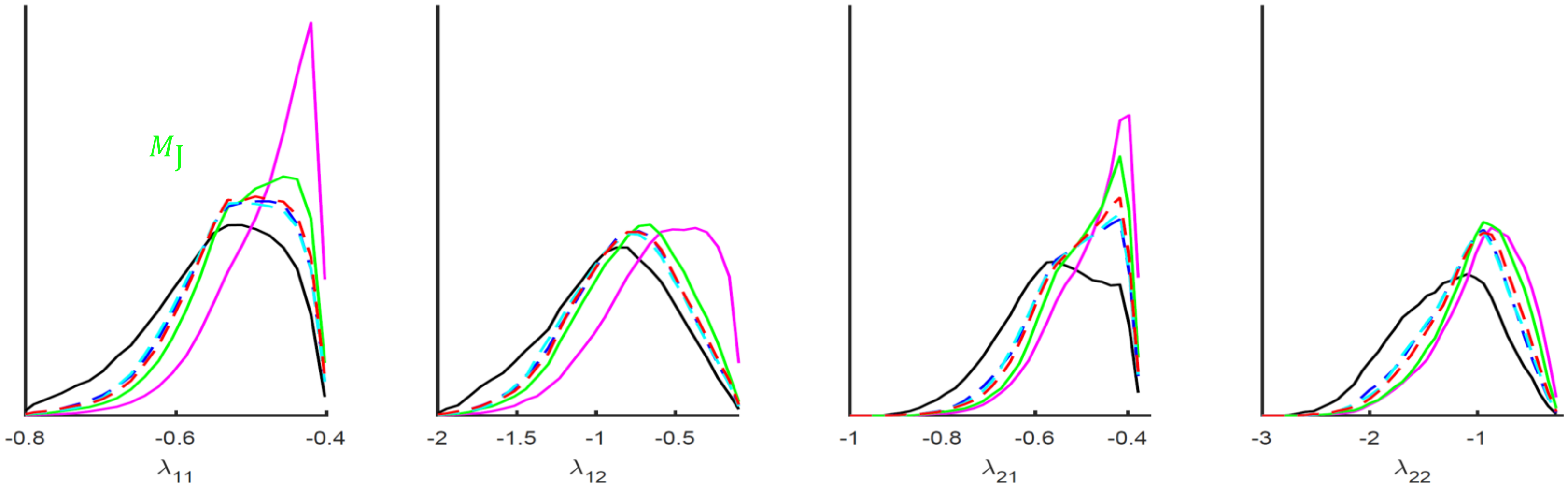
A is a sample of exact inverse density, M_π is a sample of $\rho(a) = \eta\big(F(a)\big)\pi(a)$, where $U(a) = 1, R(a) = \prod_i |a_i|^{-1}$

Linear dynamical system, uniform data distribution $U(\bar{x}_{ij}-\varepsilon,\bar{x}_{ij}+\varepsilon)$

Broyden's method (rank-1 update)

$$D_aF(\hat{a}) \approx D_aF\left(a^k\right) + \frac{F(\hat{a}) - F\left(a^k\right) - D_aF\left(a^k\right)\Delta a^k}{(\Delta a^k)^T\Delta a^k}\left(\Delta a^k\right)^T$$

- $D_aF(\hat{a})$ recomputed exactly every N steps (101, 11, 2)



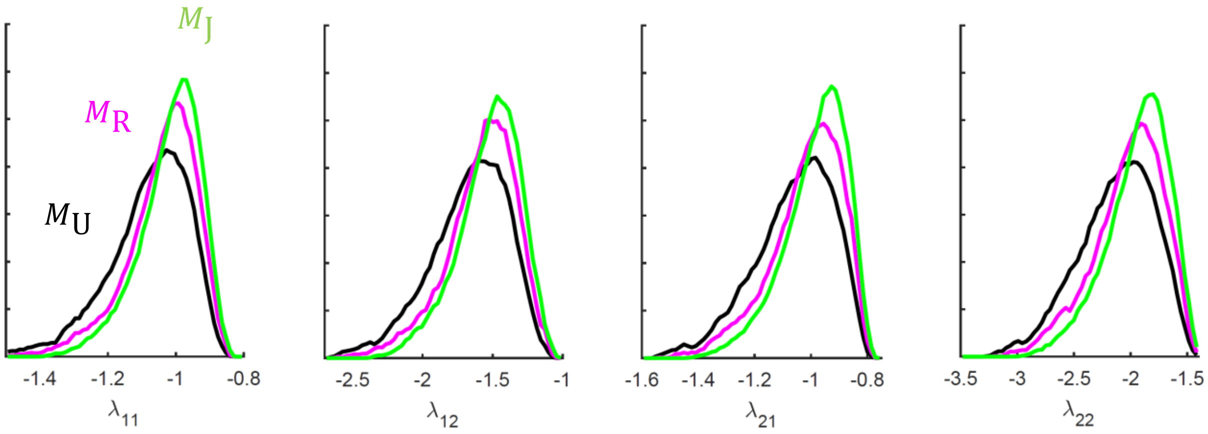
Jacobian alternatives sometimes work well

$$\boxed{\dot{x} = \Lambda x}$$

$$d = \left\{ \begin{bmatrix} 10 \\ 2 \end{bmatrix}, \begin{bmatrix} 5 \\ -2.3 \end{bmatrix}, \begin{bmatrix} 4 \\ -2.2 \end{bmatrix} \right\}$$

$$\varepsilon_U = 0.134$$

$$\Lambda = \begin{bmatrix} -1 & -1.5 \\ -1 & -2 \end{bmatrix}$$



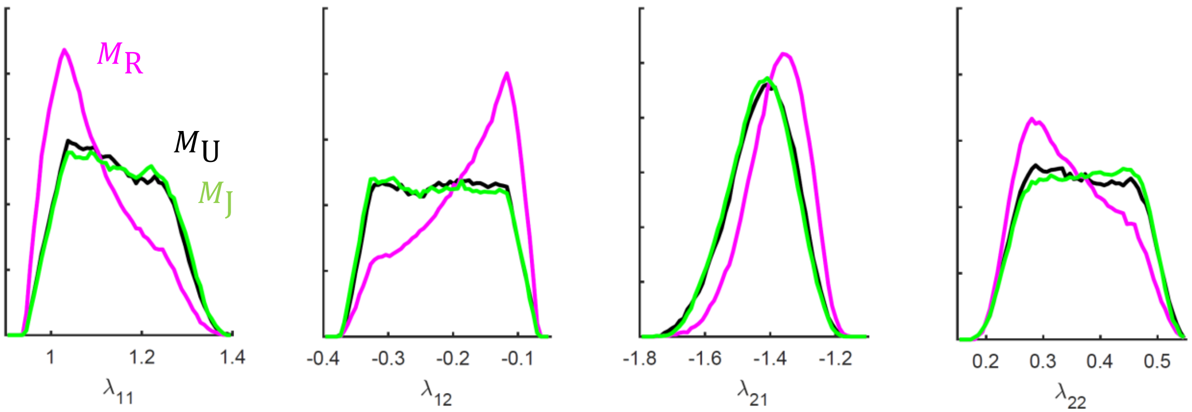
Reciprocal works well

$$\boxed{\dot{x} = \Lambda x}$$

$$d = \left\{ \begin{bmatrix} 5 \\ 15 \end{bmatrix}, \begin{bmatrix} 10 \\ 10 \end{bmatrix}, \begin{bmatrix} 30 \\ -15 \end{bmatrix} \right\}$$

$$\varepsilon_U = 1.272$$

$$\Lambda = \begin{bmatrix} 1.1347 & -0.2187 \\ -1.4213 & 0.3694 \end{bmatrix}$$

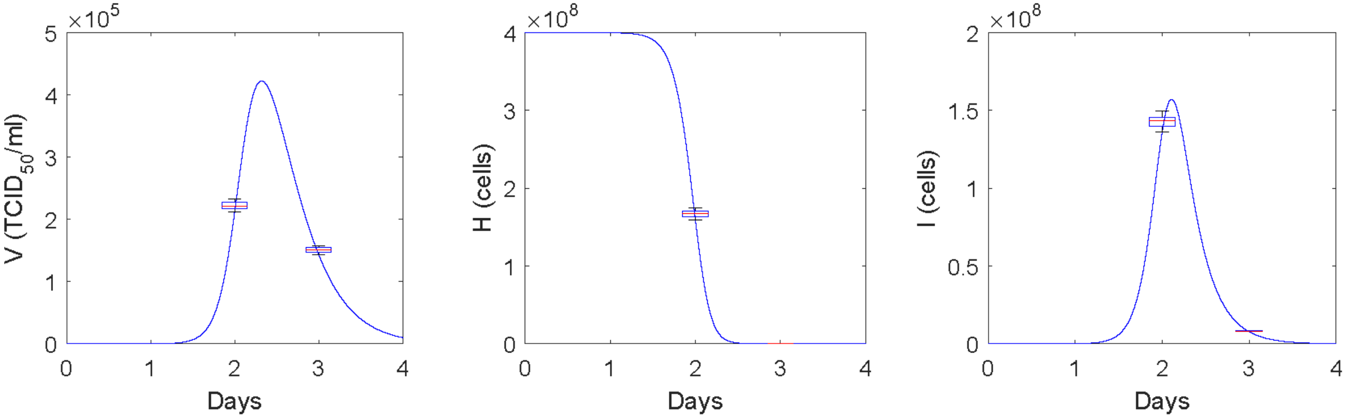


Uniform works well

[SSZR19]

Nonlinear influenza dynamics, uniform data distribution $U((1-\sigma)\bar{x}_{ij},(1+\sigma)\bar{x}_{ij})$

$$\begin{aligned}\dot{V} &= rI - cV\\ \dot{H} &= -\beta HV\\ \dot{I} &= \beta HV - \delta I\end{aligned}$$

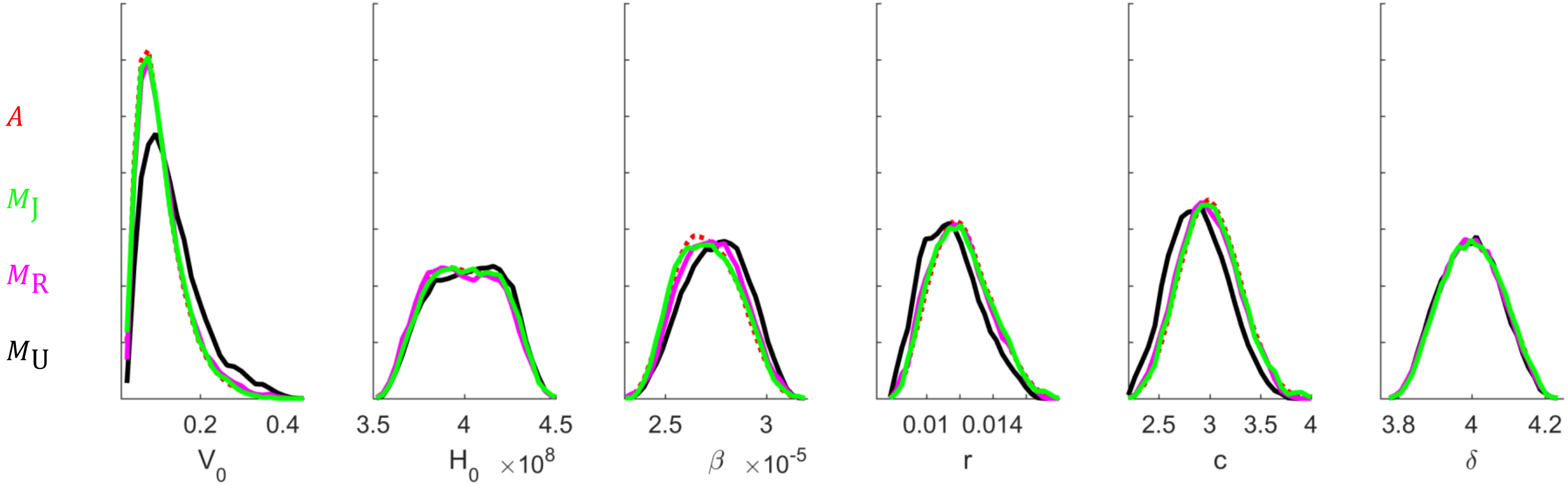


$$\sigma = 0.1$$

$$t = \{2, 3\}$$

$$a = (V_0, H_0, \beta, r, c, \delta) = (0.093,\, 4\times10^8,\, 2.7\times10^{-5},\, 0.012,\, 3.0,\, 4.0)$$

Narrow data distribution, all $\pi(a)$ work well



A is a sample of exact inverse density, M_π is a sample of $\rho(a) = \eta\big(F(a)\big)\pi(a)$, where $U(a) = 1, R(a) = \prod_i |a_i|^{-1}$

[SSZR19]

Examples

Common problem:

- Can $\eta(y)$ be approximated by aggregate Gaussian density $\tilde{\eta}(y)$?

$$\begin{tikzcd}
\rho(a) \arrow[r] & \eta(y) \arrow[d] \\
\tilde{\rho}(a) & \tilde{\eta}(y) \arrow[l]
\end{tikzcd}$$

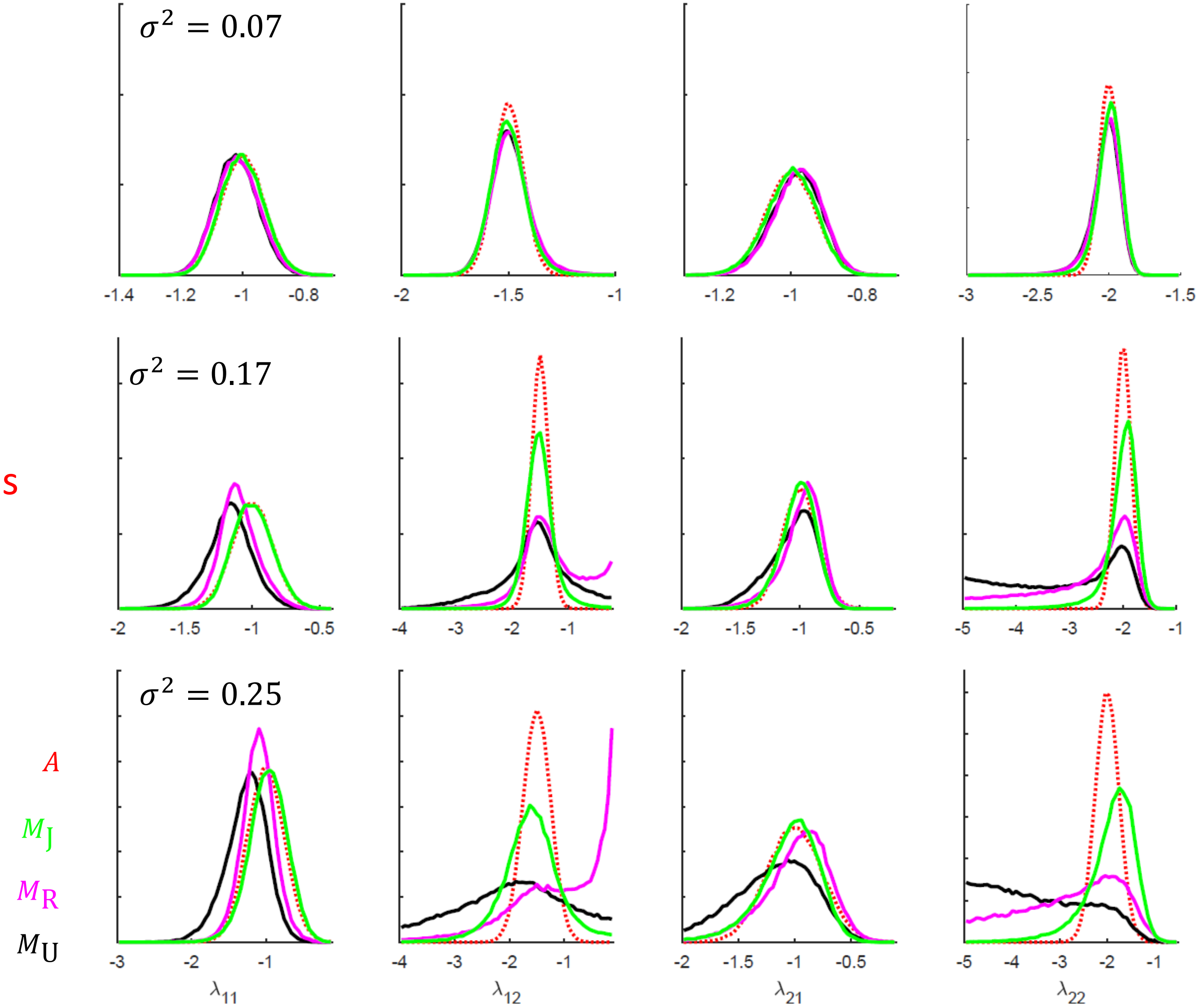
Linear dynamical system, Gaussian parameter distribution, aggregate Gaussian data distribution.

$$\boxed{\dot{x} = \Lambda x}$$

$$\overline{\Lambda} = \begin{bmatrix} -1 & -1.5 \\ -1 & -2 \end{bmatrix}$$

Only Jacobian produces localized distribution

A is the exact source density, M_π is a sample of $\rho(a) = \tilde{\eta}\big(F(a)\big)\pi(a)$

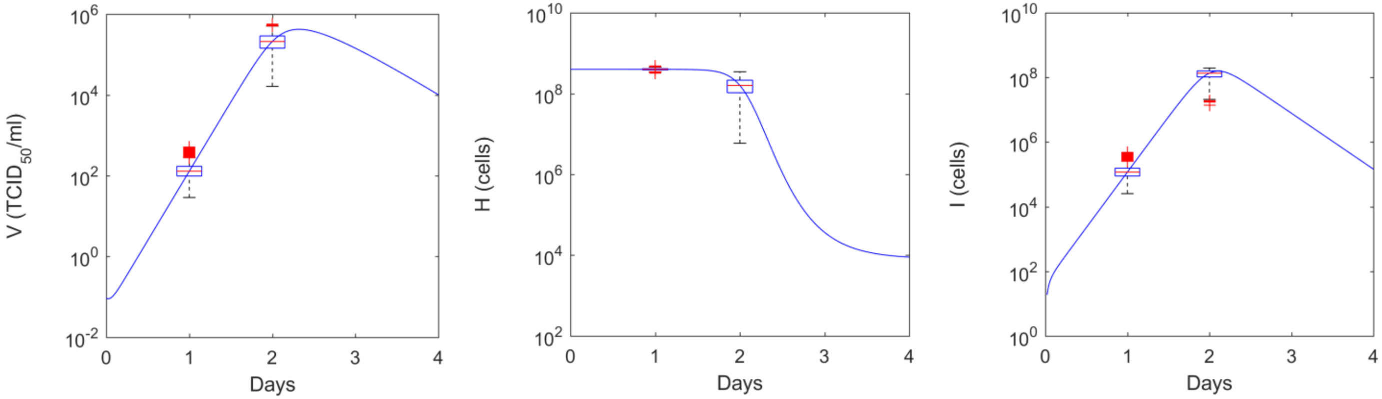


[SSZR19]

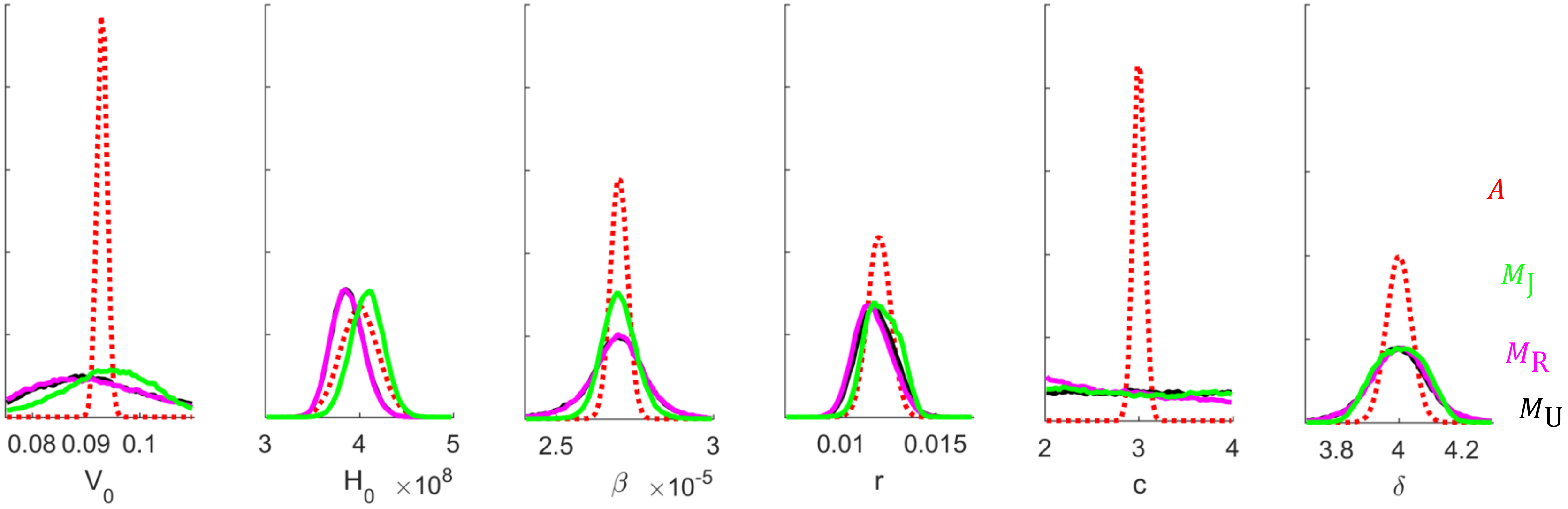
Nonlinear influenza dynamics, Gaussian parameter distribution, aggregate Gaussian data distribution

$$\begin{aligned} \dot{V} &= rI - cV \\ \dot{H} &= -\beta HV \\ \dot{I} &= \beta HV - \delta I \end{aligned}$$

$$t = \{1, 2\}$$

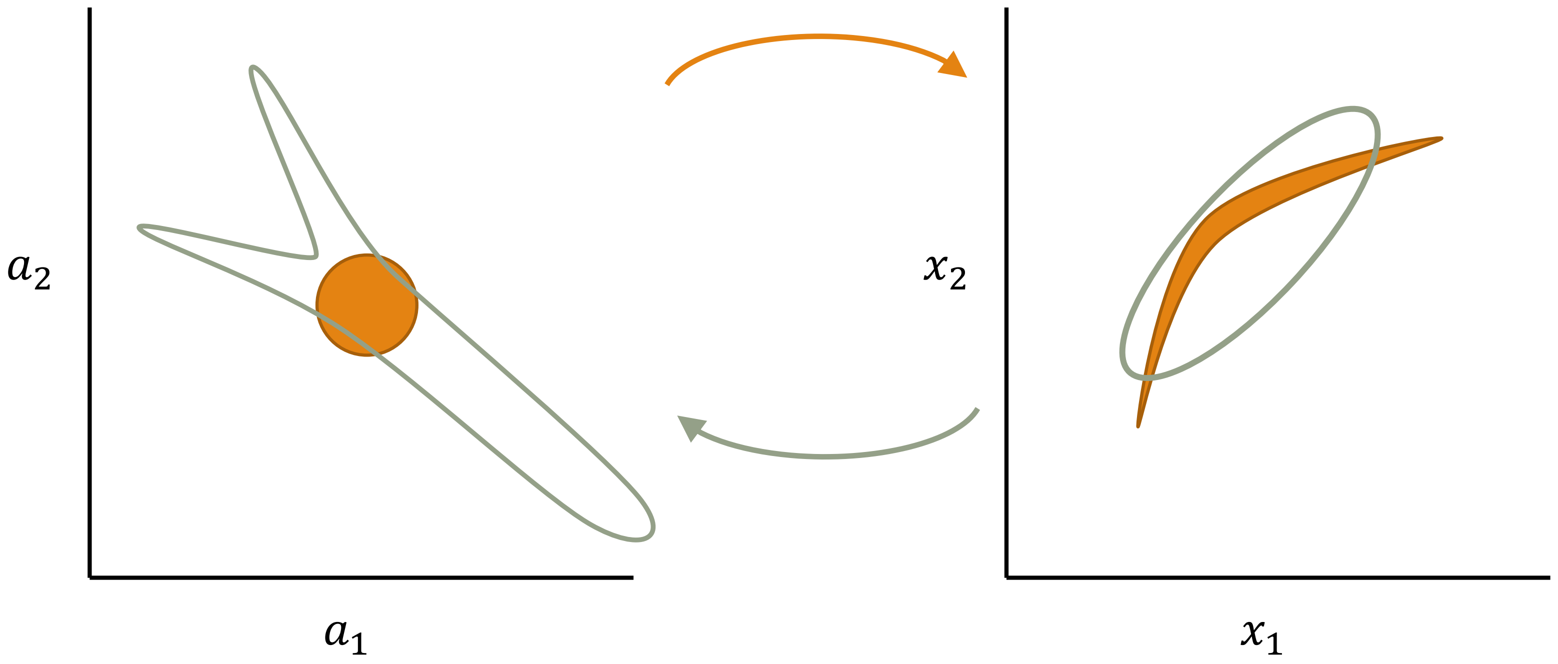


$$\bar{a} = (V_0, H_0, \beta, r, c, \delta) = (0.093,\ 4\times 10^{8},\ 2.7\times 10^{-5},\ 0.012,\ 3.0,\ 4.0)$$



[SSZR19]

Domain deformation



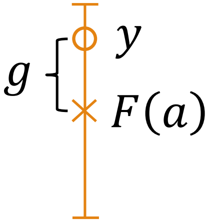
Summary

- Accurate parameter inference of ensemble models requires determination of the domain of invertibility of the solution map and the computation of Jacobian
- Further work is required to design efficient computational procedures for Jacobian calculations

Random measurement error model

$$Y = F(a) + G$$

Inverse problem: Find a from the knowledge of data $D = (y_1, \dots, y_N)$, $F(a)$, and the error density $\gamma(g)$.



Likelihood

$$L(a|y) = \rho(y|a) = \gamma(g) = \gamma\big(y - F(a)\big)$$

Bayesian posterior

$$\rho(a|y) \propto L(a|y)\pi(a)$$

- Prior $\pi(a)$ contains all information about parameters known before data are taken into account

Distribution of parameters: represents uncertainty in parameter inference

Priors

Uniform prior: $\pi(a) = 1$

Reciprocal prior: $\pi(a) = \prod_i |a_i|^{-1}$

Jeffreys invariant prior, based on the Fisher information matrix:

$$\pi(a) = \sqrt{\det I(a)}$$

$$I(a)_{ij} = \int \left(\frac{\partial}{\partial a_i} \ln L(a|y)\right)\left(\frac{\partial}{\partial a_j} \ln L(a|y)\right) L(a|y)$$

Jacobian prior: For model with likelihood $L(a|y) = \gamma\big(y - F(a)\big)$,

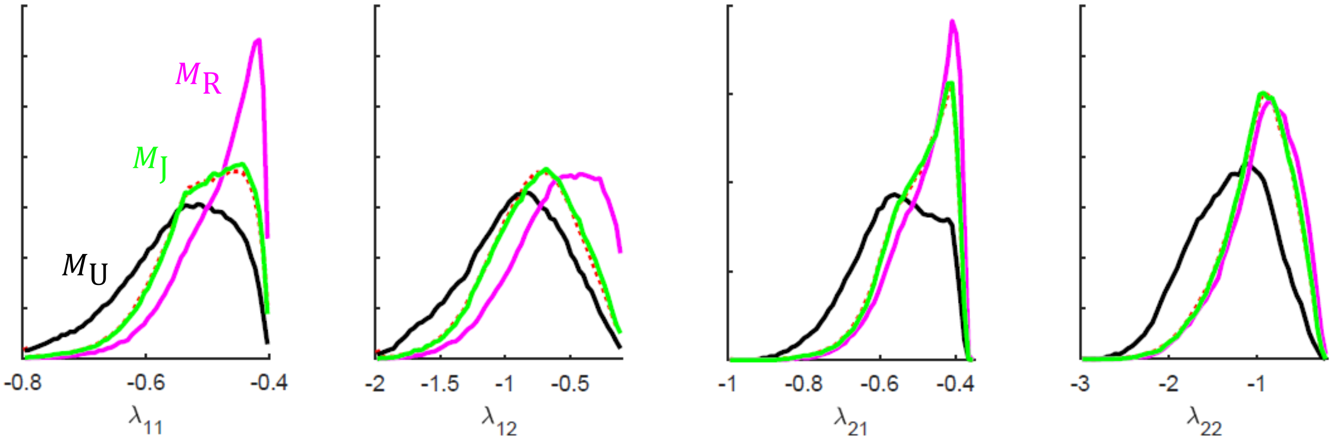
$$\pi(a) = J(a)\sqrt{\det W} \qquad J(a) = |\det D_a F(a)|$$

Relation between Bayesian posterior for REM and parameter density for RPM

Let $\tilde{\eta}(y) = \tilde{\gamma}(y - \bar{y})$ be the aggregate density obtained from data sample $Y = \{y^1, y^2, \ldots y^N\}$ and let $\tilde{\rho}(a) = \tilde{\eta}\big(F(a)\big)J(a)$

Let $\tilde{\sigma}(a|\bar{y})$ be the Bayesian posterior with likelihood $\tilde{L}(a|\bar{y}) = \tilde{\gamma}(\bar{y} - F(a))$ and prior $\pi(a)$

Theorem: $\tilde{\rho}(a) = \tilde{\sigma}(a|\bar{y})$ *whenever* $\tilde{\gamma}(y)$ *is symmetric and* $\pi(a) = J(a)$



M_π is a sample of Bayesian posterior $\sigma(a|\bar{y}) = \gamma\big(\bar{y} - F(a)\big)\pi(a)$, where $U(a) = 1, R(a) = \prod_i |a_i|^{-1}$

[SSZR19]

Jacobian prior revisited

$$Y = F(a) + G$$

Likelihood: $L(a|y) = \gamma\big(y - F(a)\big)$

Posterior: $\sigma(a|y) = L(a|y)\pi(a)/q(y)$

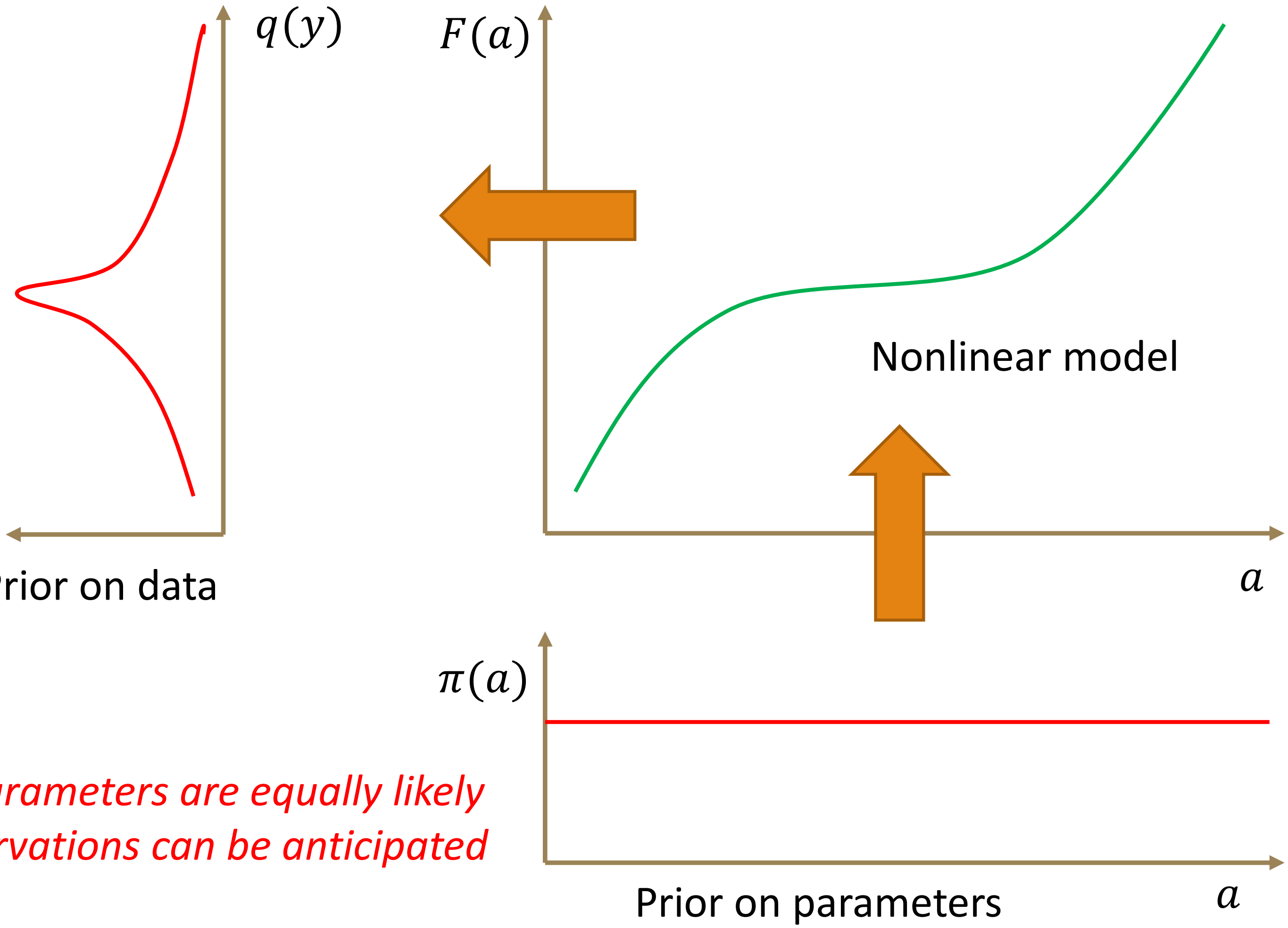
With Jacobian prior on parameters, the prior on data y is uniform

$$q(y) = \int L(a|y)\pi(a)da$$
$$= \int \gamma\big(y - F(a)\big)J(a)da = \int \gamma(y - z)dz = 1$$

The Jacobian prior is ***noninformative*** *in the* ***original sense proposed by Bayes*** *– it does not give preference to any observed data [Stigler, 1982].*

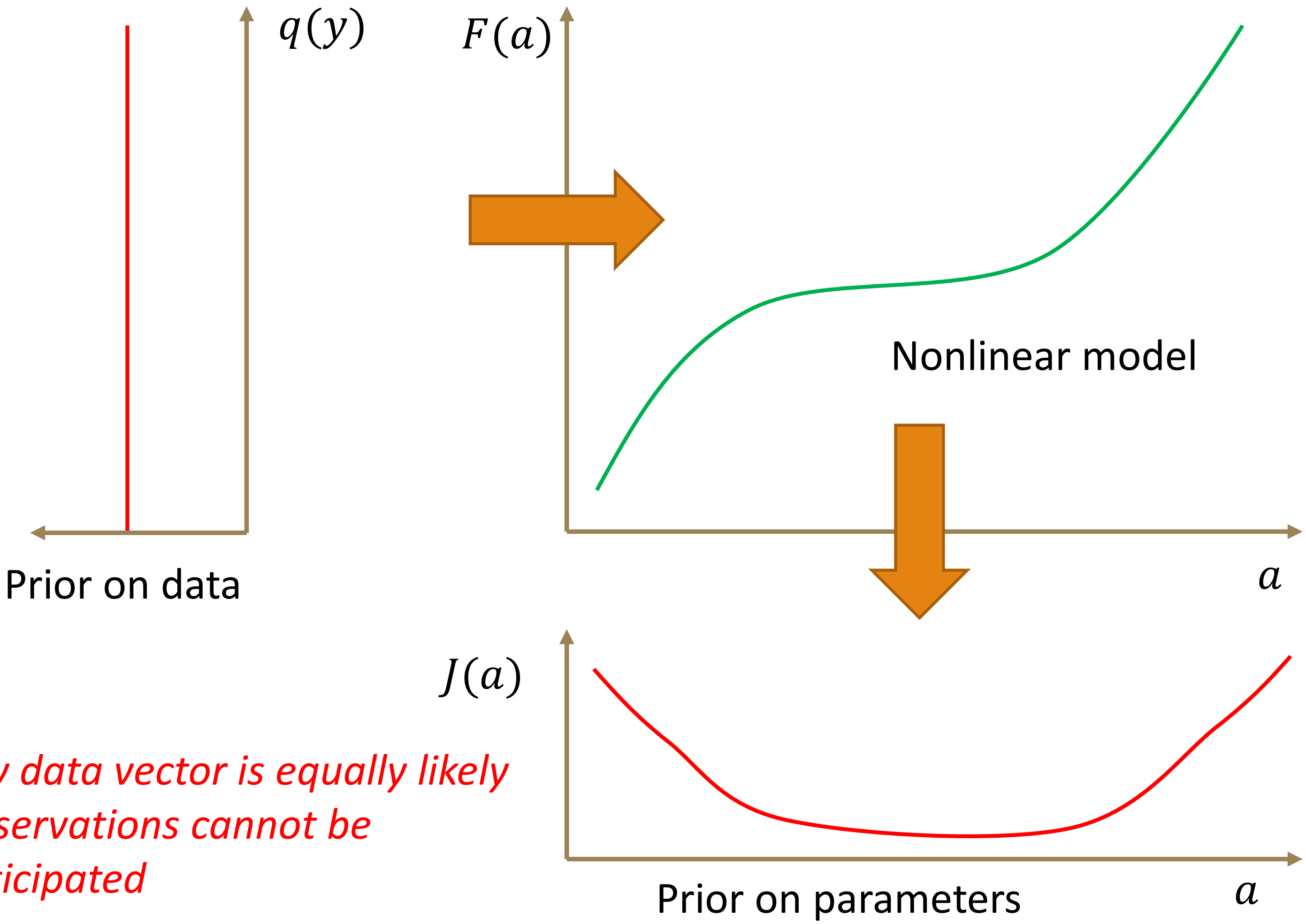
[SSZR19]

Commonly used interpretation of noninformative prior



- *All parameters are equally likely*
- *Observations can be anticipated*

Bayes’ understanding of noninformative prior



Prior on data

Nonlinear model

Prior on parameters

- *Any data vector is equally likely*
- *Observations cannot be anticipated*

Relation to least squares

Objective function minimization

$$a^* = \arg\min_a \sum_j \left(x_j - x\left(t_j; a\right)\right)^2$$

Maximum likelihood estimate

$$a_{MLE} = \arg\max_a L(a|y) = \arg\max_a \gamma\left(y - F(a)\right) = a^*$$

The data y is the most likely for model with the parameter a_{MLE}

Maximum a posteriori estimate

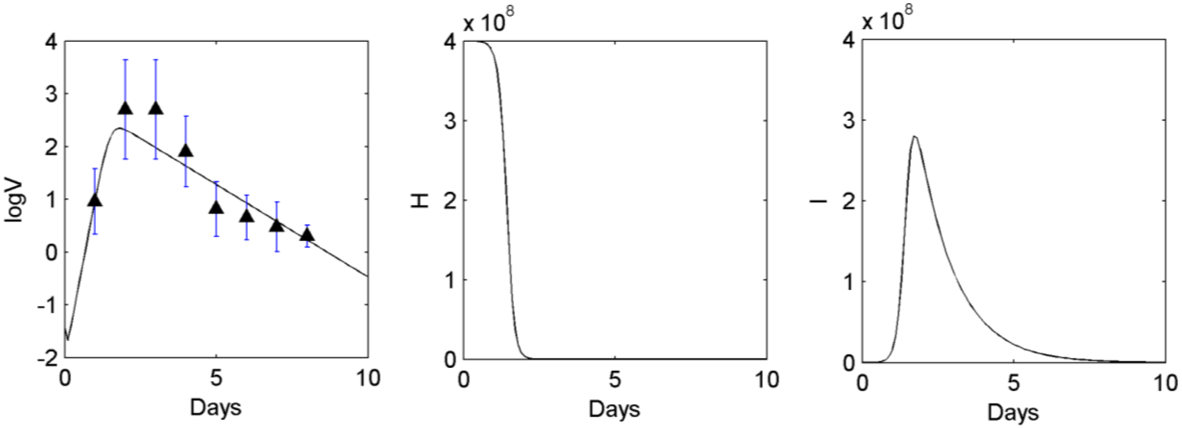
$$a_{MAP} = \arg\max_a \rho(a|y) = \arg\max_a \gamma\left(y - F(a)\right)J(a)$$

The model with parameter a_{MAP} is the most likely given observed data y

Applications to immunology

Data on time courses of 16 volunteers infected by Influenza A/Texas/91 (H1N1) [Hayden , JAMA, 96]

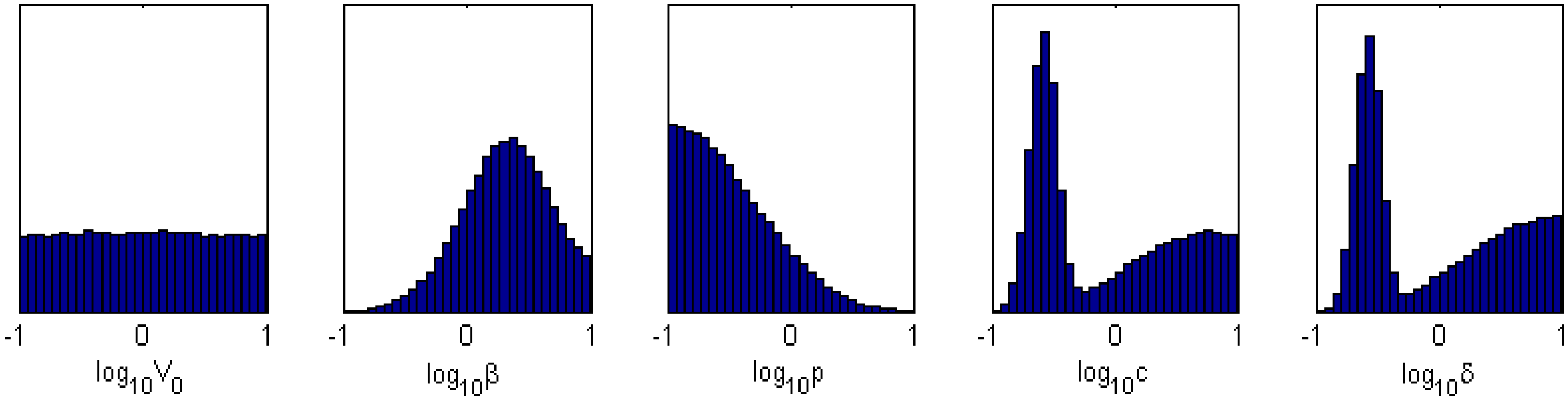
$$\dot{V} = pI - cV$$
$$\dot{H} = -\beta HV$$
$$\dot{I} = \beta HV - \delta I$$



Classical parameter estimates

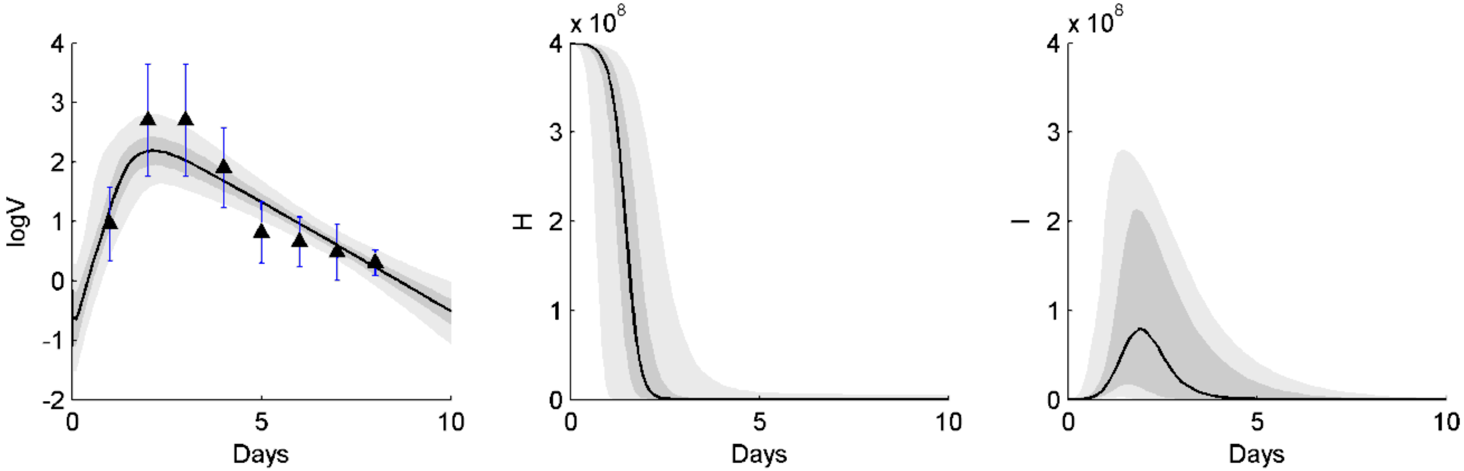
$$(V_0, \beta, p, c, \delta)_{opt} = (0.0373, 0.0347, 1.33\times10^{-5}, 16.7, 0.806)$$

Ensemble parameter distribution

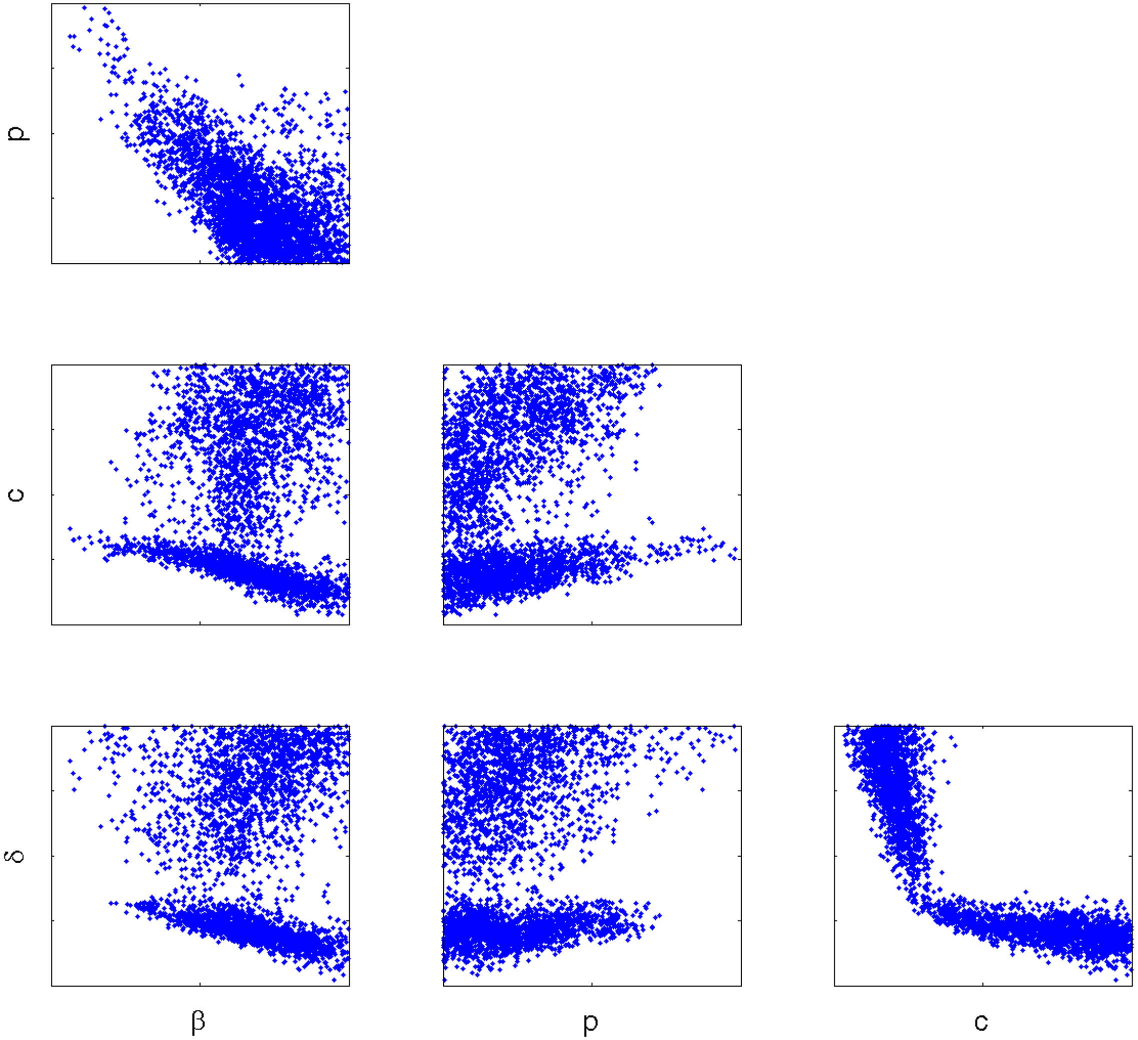


$$(V_0, \beta, p, c, \delta)_{base} = (0.25, 0.014, 2.7\times10^{-5}, 3.2, 3.2)$$

Ensemble trajectory prediction



Parameter correlations – parameter distribution is organized in clusters

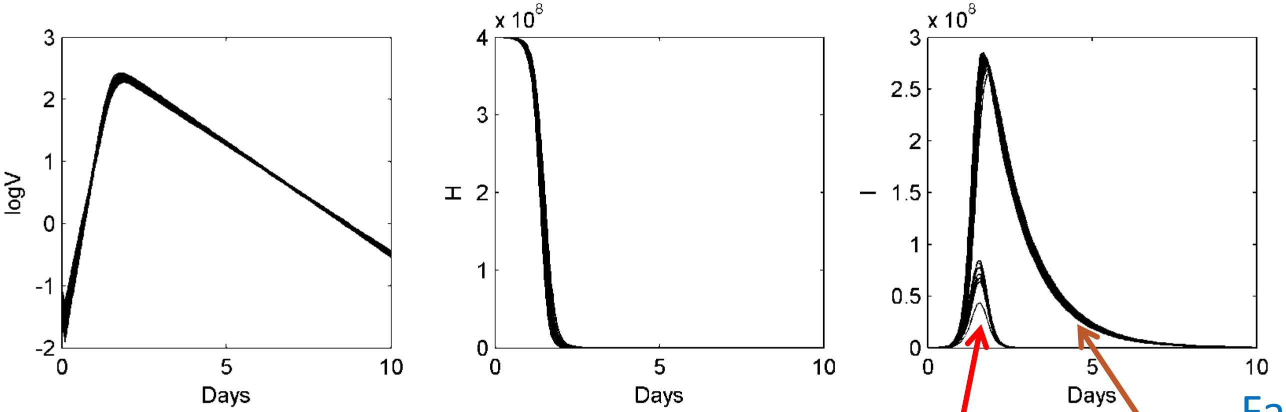


Optimal trajectories reflect bimodality of the posterior

$$\dot{V} = pI - cV$$

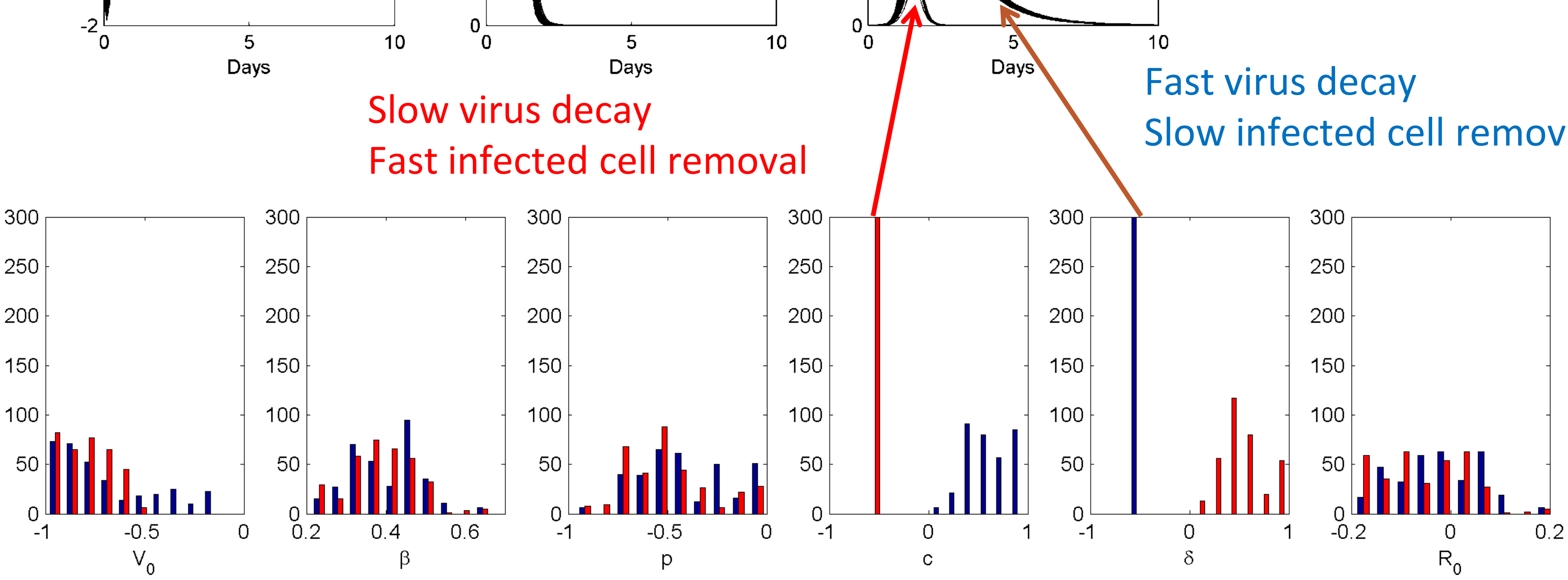
$$\dot{H} = -\beta HV$$

$$\dot{I} = \beta HV - \delta I$$



Slow virus decay
Fast infected cell removal

Fast virus decay
Slow infected cell removal



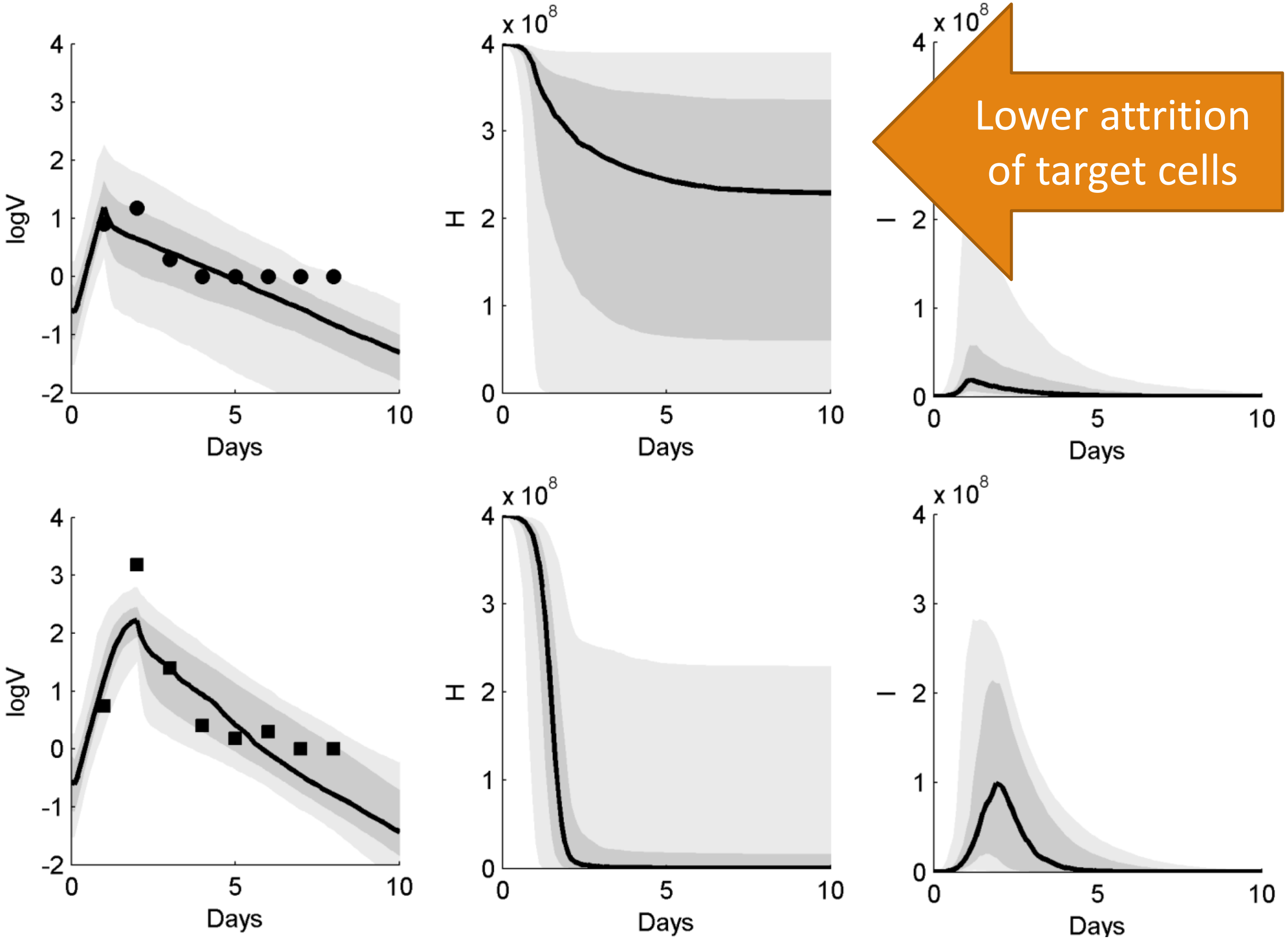
Uncertainty quantification of antiviral treatment

The effect of neuraminidase inhibitor GG167 is simulated by decreasing *p* to 1/20th of its value

Treatment starting at 26 hrs

Treatment starting at 50 hrs

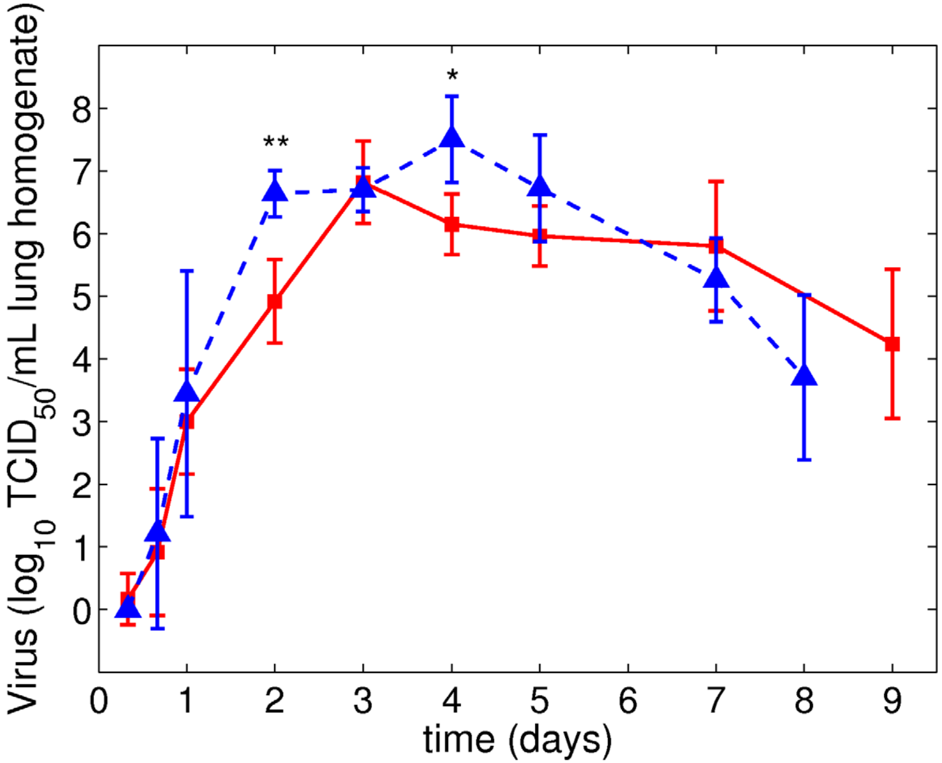
Data from [Hayden , JAMA, 96]



Lower attrition of target cells

Viral phenotype characterization

PB1-F2 protein occurring in 1918 pandemic H1N1 strain increases apoptosis in monocytes, alters viral polymerase activity in vitro, enhances inflammation and increases secondary pneumonia in vivo.

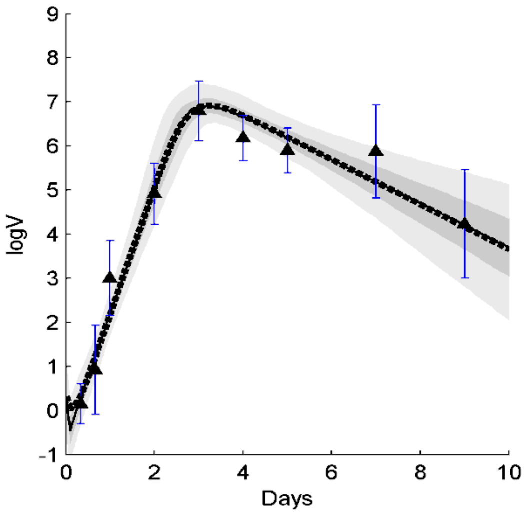


Mice infected intranasally with 100 TCID50 of influenza A virus PR8 (squares) or PR8-PB1-F2(1918) (triangles).

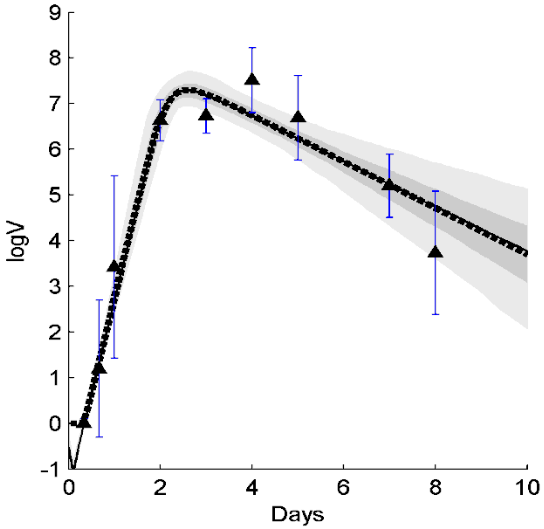
[Smith et al., PloS ONE, 2011]

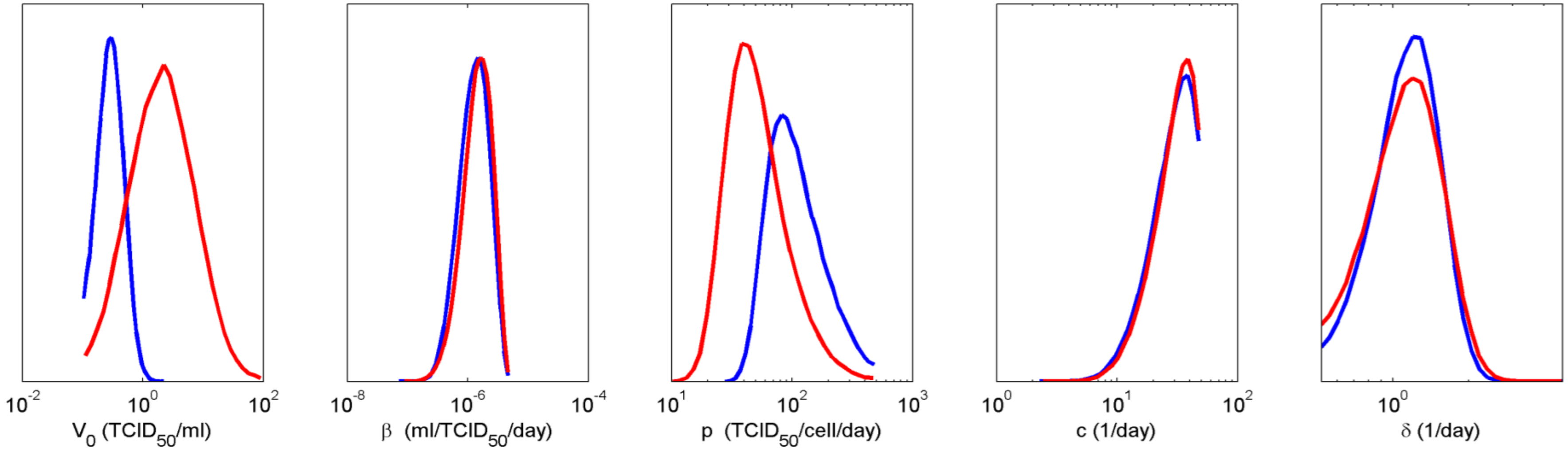
Comparison of model predictions and parameter distributions obtained using ensemble modeling.

PR8

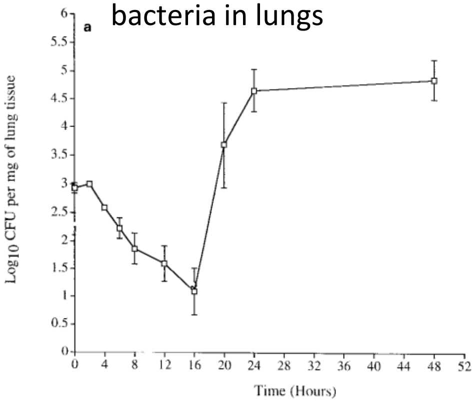


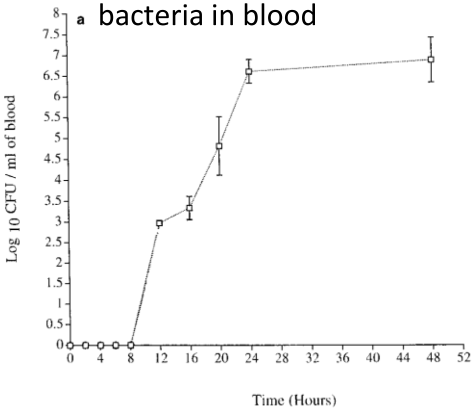
PR8-PB1-F2



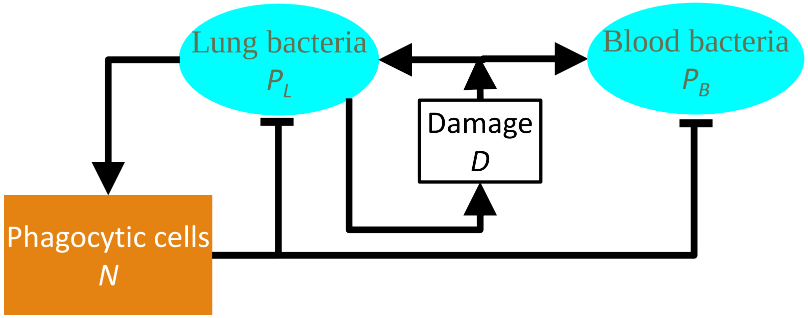


Bacterial pneumonia


bacteria in lungs


bacteria in blood

Data from MF1 mice subjects infected by D39 pneumococcus strain [Kadioglu et al., Inf. and Immun., 2000]:


Lung bacteria P_L

Blood bacteria P_B

Damage D

Phagocytic cells N

$$\frac{dP_L}{dt} = k_l P_L - \frac{\boxed{\nu} P_L}{\mu + P_L} - \frac{\boxed{\xi_{nl}} N P_L}{1 + \xi_2 P_L} + f\left[(bD + a)P_B - bDP_L\right]$$

$$\frac{dP_B}{dt} = k_b P_B \left(\frac{P_B - \varepsilon}{P_B + \varepsilon}\right)\left(1 - \frac{P_B}{K}\right) - \frac{\boxed{\xi_{nb}} N P_B}{1 + \xi_2 P_B} - (bD + a)P_B + bDP_L$$

$$\frac{dD}{dt} = qP_L(1 - D) - cD$$

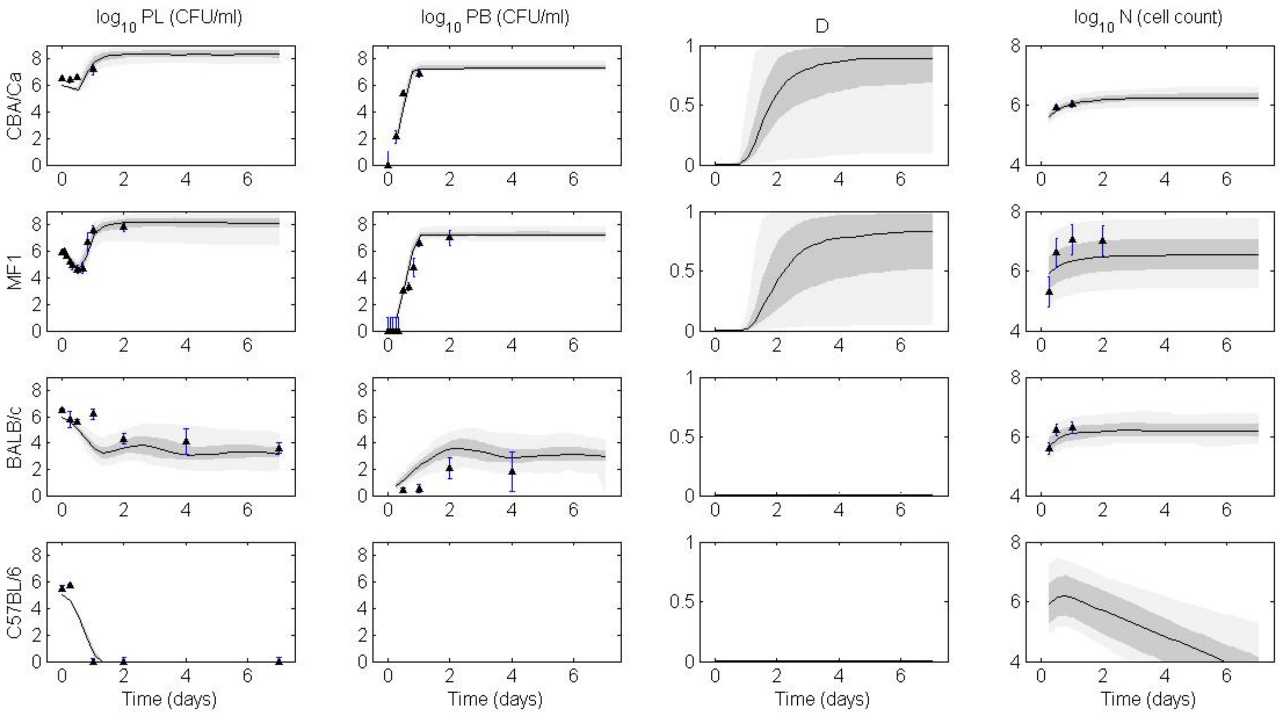
$$\frac{dN}{dt} = \left(-N + \frac{\boxed{h} P_L}{n_h + P_L}\right)\left(\frac{1}{\tau_N}\right)$$

[MSELC14]

Bacterial pneumonia - strain dependence

Response to D39 pneumococcus infection in four mice strains

[Data from Kadioglu at al, I&I, 2000; Gingles et al., I&I, 2001; Kadioglu et al., JID, 2011]



severity

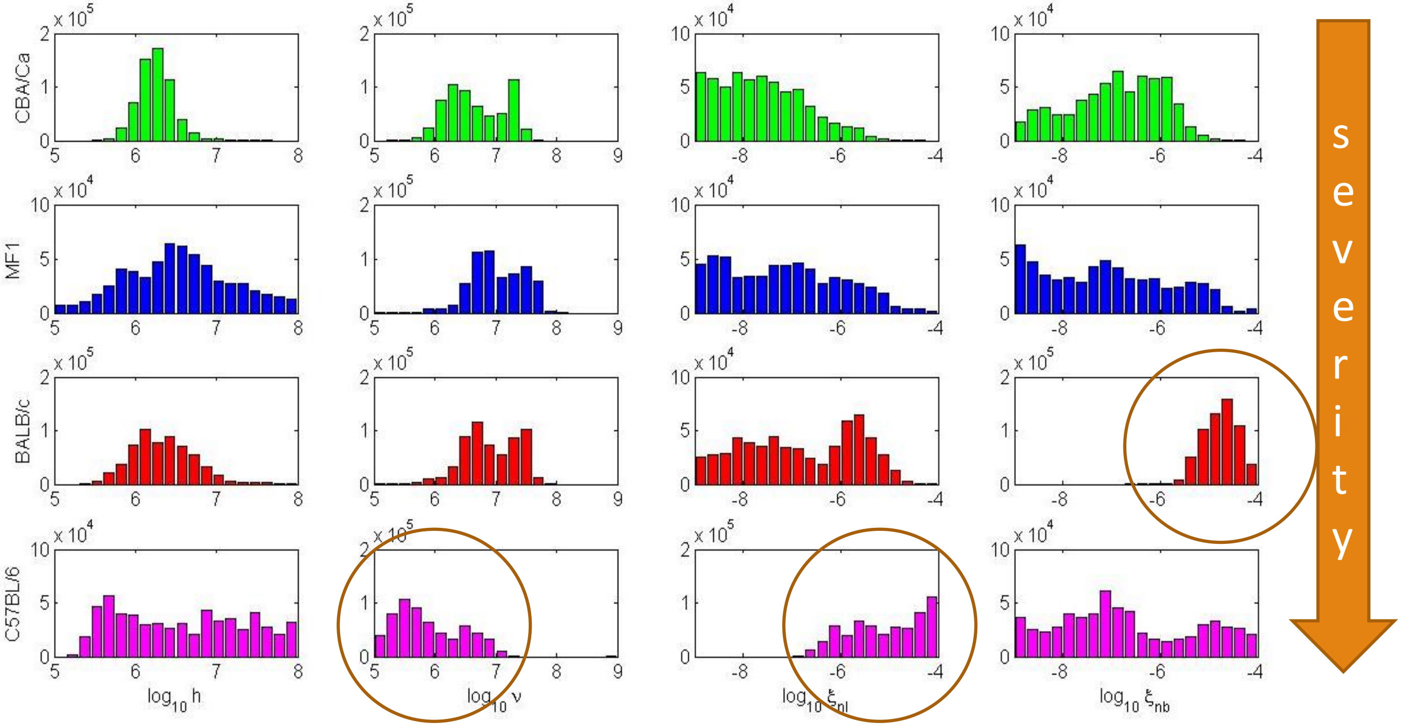
[MSELC14]

Bacterial pneumonia - strain dependence

Response to D39 pneumococcus infection in four mice strains

[Data from Kadioglu at al, I&I, 2000; Gingles et al., I&I, 2001; Kadioglu et al., JID, 2011]

Strain dependent parameters $h, \nu, \xi_{nl}, \xi_{nb}$



[MSELC14]

References

- [MSELC14] Mochan, E., Swigon, D., Ermentrout, G.B., Lukens, S., & Clermont, G., A mathematical model of intrahost pneumococcal pneumonia infection dynamics in murine strains, J. Theor. Biol, 353, 44-54 (2014).
- [SRS14] Stanhope, S., Rubin, J.E., & Swigon, D., Identifiability of linear dynamical systems from a single trajectory, *SIADS*, 13, 1792-1815 (2014).
- [SRS17] Stanhope, S., Rubin, J. E., & Swigon, D., Robustness of solutions of the inverse problem for linear dynamical systems with uncertain data, *SIAM/ASA JUQ,* 5, 572–597 (2017).
- [S17] Swigon, D. Parameter estimation for nonsymmetric matrix Riccati differential equations, *Proceedings of ENOC 2017, (*2017).
- [SSZR19] Swigon, D., Stanhope, S., Zenker, S., & Rubin, J. E., On the importance of the Jacobian determinant in parameter inference for random parameter and random measurement error models, *to appear in SIAM JUQ.*
- [DRS19] Duan, X., Rubin, J.E., Swigon, D., Identification of Affine Dynamical Systems from Single Experiment, *in preparation.*
- [S19] Swigon, D., Solution of inverse problem for linear-in-parameters dynamical systems with equally spaced data, *in preparation.*

Credits

Inverse problems

- Jon Rubin
- Shelby Stanhope
- Xiaoyu Duan

Influenza/pneumonia modeling

- Gilles Clermont
- Bard Ermentrout
- Ericka Mochan
- Sarah Lukens

Funding

- NIGMS
- NSF-RTG

UNIVERSITY OF PITTSBURGH 1787

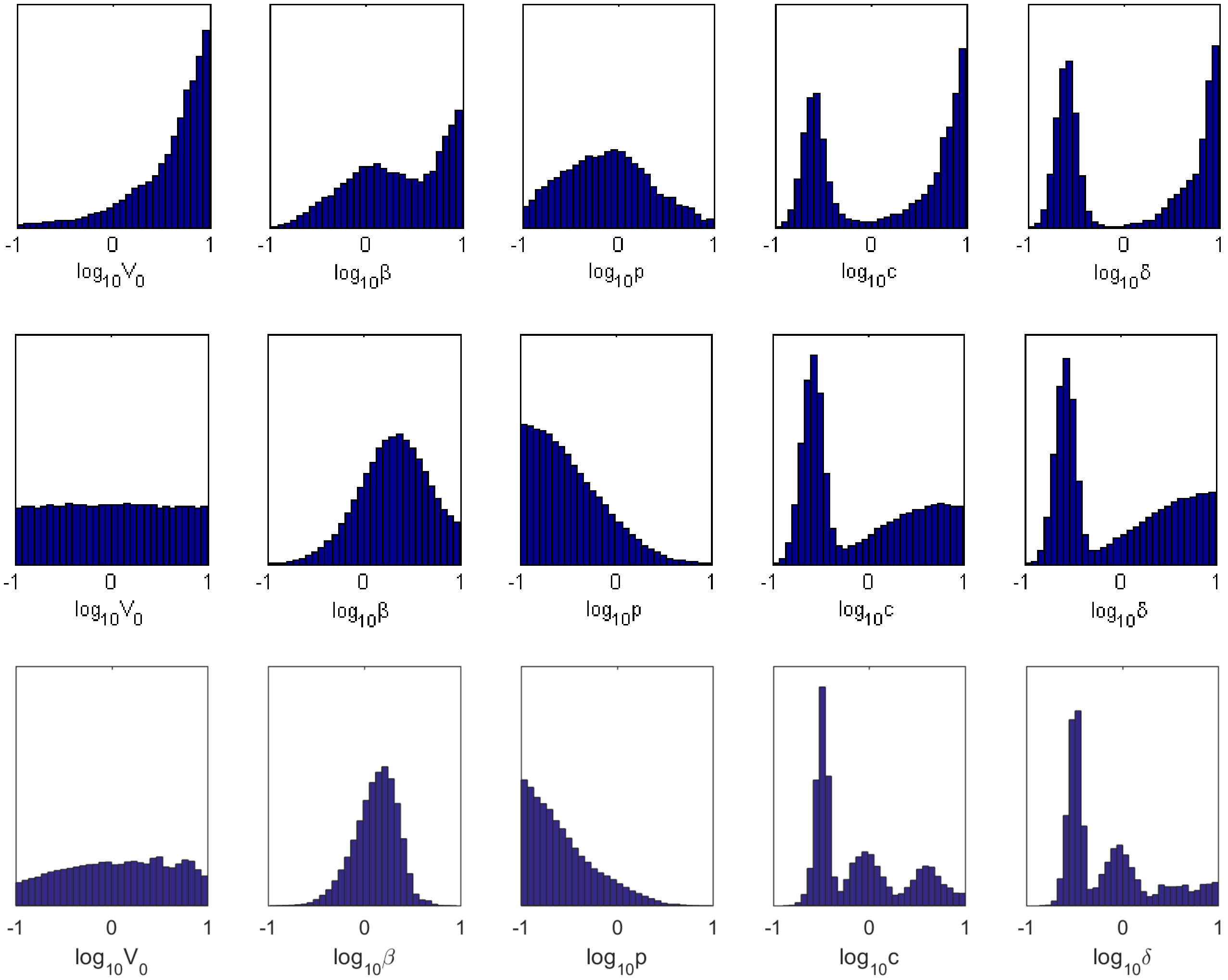
Appendix

Posterior distributions

Uniform prior

Reciprocal prior

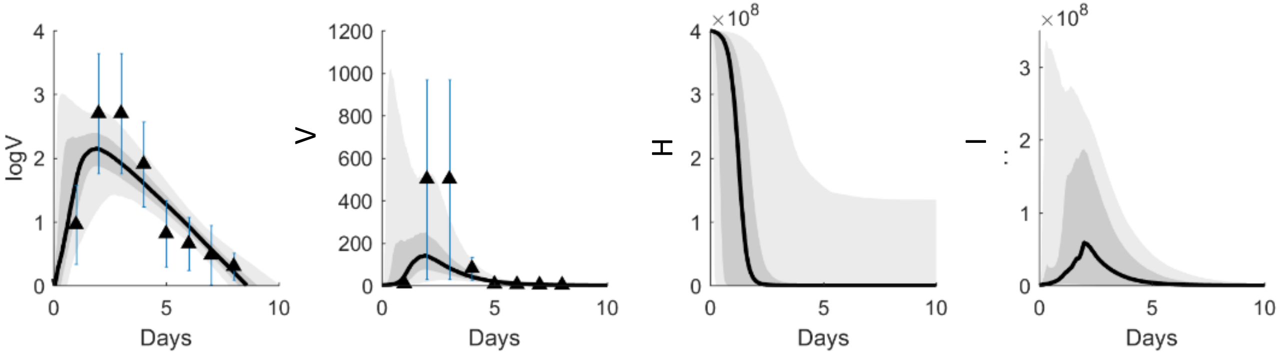
Jacobian prior



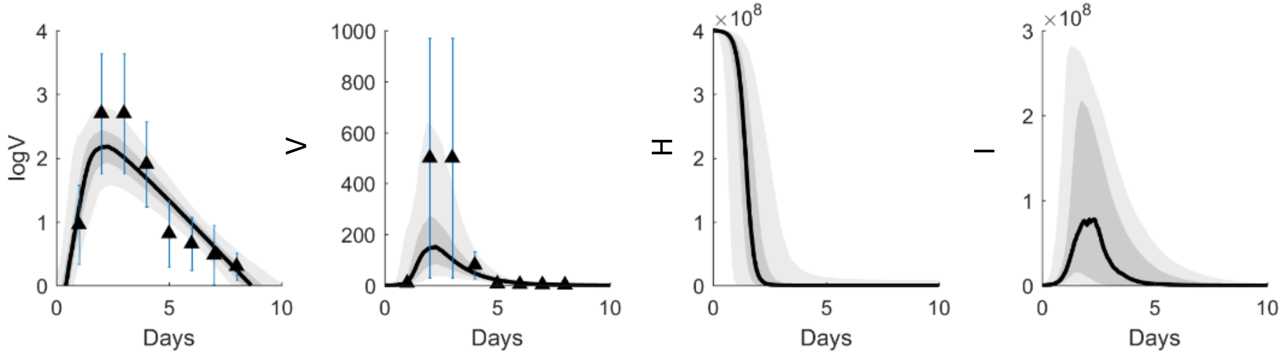
$$(V_0, \beta, p, c, \delta)_{base} = (0.25, 0.014, 2.7\times10^{-5}, 3.2, 3.2)$$

Probabilistic trajectory predictions

Uniform prior



Reciprocal prior



Jacobian prior