Model Validation: A Probabilistic Formulation

Abhishek Halder and Raktim Bhattacharya

Department of Aerospace Engineering
Texas A&M University
College Station, TX 77843-3141

50th IEEE CDC-ECC, Orlando, Florida
December 12, 2011

Model validation problem: introduction

Given (i) a candidate model, and (ii) experimentally observed measurements of the physical system at times $\{t_j\}_{j=1}^{M}$, how well does the model replicate the experimental measurements?

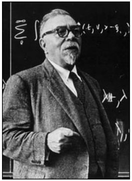
Model validation problem: introduction

Given (i) a candidate model, and (ii) experimentally observed measurements of the physical system at times $\{t_j\}_{j=1}^{M}$, how well does the model replicate the experimental measurements?

- Model invalidation
  [Smith and Doyle, 1992; Poolla *et. al.*, 1994; Prajna, 2006]

"The best model of a cat is another cat,
or better yet, the cat itself".

– *Norbert Wiener*



- Binary invalidation oracle

Q1. Is this overly conservative?
Q2. Can we compute the "degree of (in)validation"?

Model validation problem: state-of-the-art

Linear Model Validation

- Robust control framework
  - Time domain
    [Poolla *et. al.*, 1994; Chen and Wang, 1996]
  - Frequency domain
    [Smith and Doyle, 1992; Steele and Vinnicombe, 2001; Gevers *et. al.*, 2003]
  - Mixed domain
    [Xu *et. al.*, 1999]
- Statistical setting
  - Correlation analysis
    [Ljung and Guo, 1997]
  - Bayesian conditioning
    [Lee and Poolla, 1996]

Nonlinear Model Validation

- Barrier certificate method
  [Prajna, 2006]
- Polynomial chaos method
  [Ghanem *et. al.*, 2008]

Model validation problem: state-of-the-art

Linear Model Validation

- Robust control framework
  - Time domain
    [Poolla *et. al.*, 1994; Chen and Wang, 1996]
  - Frequency domain
    [Smith and Doyle, 1992; Steele and Vinnicombe, 2001; Gevers *et. al.*, 2003]
  - Mixed domain
    [Xu *et. al.*, 1999]
- Statistical setting
  - Correlation analysis
    [Ljung and Guo, 1997]
  - Bayesian conditioning
    [Lee and Poolla, 1996]

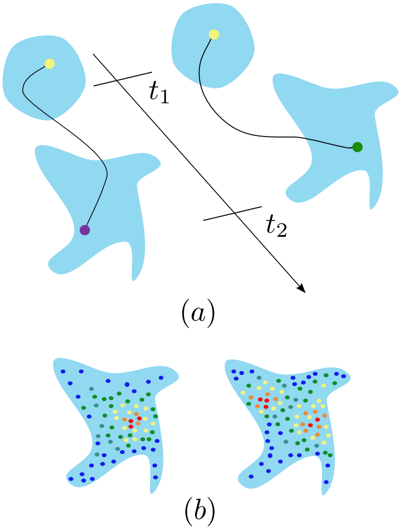
Nonlinear Model Validation

- Barrier certificate method
  [Prajna, 2006]
- Polynomial chaos method
  [Ghanem *et. al.*, 2008]

"For the general case of nonparametric (uncertainty) models, the situation is significantly more complicated"
– *[Lee and Poolla, 1996]*

Q3. Nonlinear model validation in the sense of nonparametric statistics (aleatoric uncertainty)?

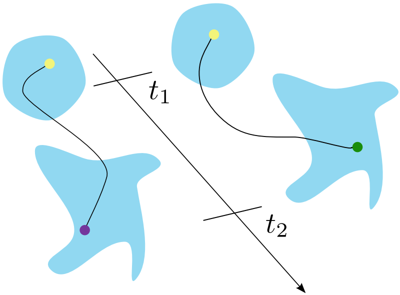
Our approach: intuitive idea



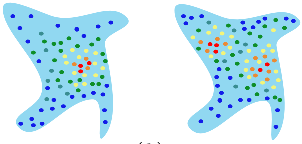
What to compare for nonlinear systems?

- Our proposal: compare shapes of the output PDFs at $\{t_j\}_{j=1}^{M}$
- Why PDFs instead of
  - trajectories?
  - supports?
  - moments?
- Why shapes?

Our approach: intuitive idea



(a)



(b)

What to compare for nonlinear systems?

- Our proposal: compare shapes of the output PDFs at $\{t_j\}_{j=1}^{M}$
- Why PDFs instead of
  - trajectories?
  - supports?
  - moments?
- Why shapes?

Should work for

- any nonlinearity
- any uncertainty
- both discrete and continuous time
- computationally tractable
- validation certificate

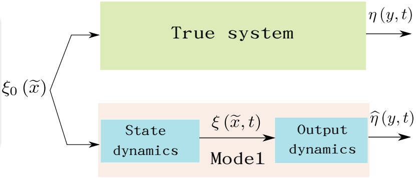
Outline

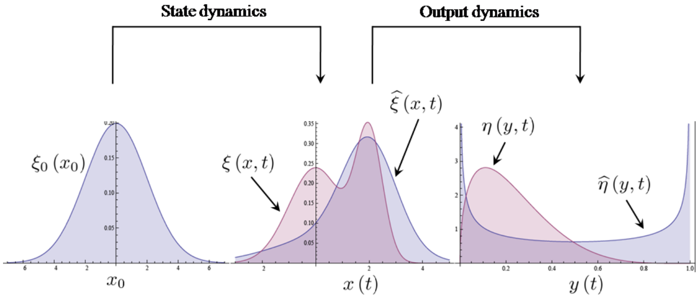
- Introduction
- State-of-the-art
- Intuitive idea
- Problem formulation
- Uncertainty propagation
- Distributional comparison
- Construction of validation certificates
- Examples
- Conclusions

Problem formulation

Proposed framework

Step 1. Uncertainty propagation
Step 2. Distributional comparison
Step 3. Construction of validation certificates

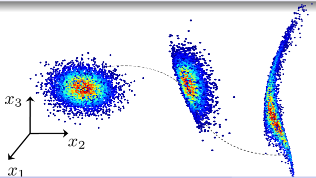




Uncertainty propagation

Continuous-time deterministic model

- Model
  $\dot{x} = f(x,t,p) \Rightarrow \dot{\widetilde{x}} = \widetilde{f}(\widetilde{x},t)$,
  $y = h(\widetilde{x},t)$
- Liouville equation
  $$\frac{\partial \widehat{\xi}}{\partial t} = -\sum_{i=1}^{n_s} \frac{\partial}{\partial x_i}\left(\widehat{\xi} f_i\right),$$
  $$\widehat{\eta}(y,t) = \sum_{j=1}^{\nu} \frac{\widehat{\xi}\left(\widetilde{x}_j^{\star},t\right)}{\left|\det\left(\mathcal{J}_h\left(\widetilde{x}_j^{\star},t\right)\right)\right|}$$
- Method-of-characteristics
  $$\frac{d\widehat{\xi}}{dt} = -\widehat{\xi}\,\nabla.f,\ \widehat{\xi}\left(\widetilde{x}(0),0\right) = \xi_0$$



Continuous-time stochastic model

- Model
  $d\widetilde{x} = \widetilde{f}(\widetilde{x},t)\ dt + g(\widetilde{x},t)\ dW$,
  $y = h(\widetilde{x},t) + V$
- Fokker-Planck equation
  $$\frac{\partial \widehat{\xi}}{\partial t} = -\sum_{i=1}^{n_s} \frac{\partial}{\partial x_i}\left(\widehat{\xi} f_i\right) + \sum_{i=1}^{n_s}\sum_{j=1}^{n_s} \frac{\partial^2}{\partial x_i \partial x_j}\left(\left(gQg^T\right)_{ij}\widehat{\xi}\right),$$
  $$\widehat{\eta}(y,t) = \left(\sum_{j=1}^{\nu} \frac{\widehat{\xi}\left(\widetilde{x}_j^{\star},t\right)}{\left|\det\left(\mathcal{J}_h\left(\widetilde{x}_j^{\star},t\right)\right)\right|}\right) * \phi_V$$
- Karhunen-Loève + MOC
  $\dot{\widetilde{x}} = \widetilde{f}(\widetilde{x},t) + g(\widetilde{x},t)\,\mathsf{KL}_N$
  $$\mathsf{KL}_\infty \overset{\text{m.s.}}{=} \sqrt{2}\sum_{i=1}^{\infty} \zeta_i(\omega)\cos\left(\left(i-\frac{1}{2}\right)\frac{\pi t}{T}\right)$$

Uncertainty propagation

Discrete-time deterministic model

- Model
  $\widetilde{x}_{k+1} = \mathcal{T}\left(\widetilde{x}_k\right),\ y_k = h\left(\widetilde{x}_k\right)$
- Perron-Frobenius opearator

$$\widehat{\xi}_{k+1} = \mathcal{L}\,\widehat{\xi}_k = \frac{\widehat{\xi}_k\left(\mathcal{T}^{-1}\left(x_{k+1}\right)\right)}{\left|\det\left(\mathcal{J}_{\mathcal{T}}\left(x_{k+1}\right)\right)\right|},$$

$$\widehat{\eta}_k = \sum_{j=1}^{\nu} \frac{\widehat{\xi}_k\left(\widetilde{x}_j^{\star}, t\right)}{\left|\det\left(\mathcal{J}_h\left(\widetilde{x}_j^{\star}, t\right)\right)\right|}$$

- Cell-to-cell mapping
  Transition probability matrix
  $P_{ij} := \dfrac{n_{ij}}{n}$

Discrete-time stochastic model

- Model
  $\widetilde{x}_{k+1} = \mathcal{S}\left(\widetilde{x}_k\right) + w_k,$
  $\widetilde{x}_{k+1} = w_k \mathcal{S}\left(\widetilde{x}_k\right),$
  $y_k = h\left(\widetilde{x}_k\right) + v_k$
- Stochastic transfer operator

$$\widehat{\xi}_{k+1} = \mathcal{L}_{\text{add}}\,\widehat{\xi}_k = \int_{\mathbb{R}^{n_s}} \widehat{\xi}_k(y)\,\phi_w\left(x_{k+1} - \mathcal{S}(y)\right) dy,$$

$$\widehat{\xi}_{k+1} = \mathcal{L}_{\text{mul}}\,\widehat{\xi}_k = \int_{\mathbb{R}^{n_s}} \widehat{\xi}_k(y)\,\frac{1}{\mathcal{S}(y)}\phi_w\left(\frac{x_{k+1}}{\mathcal{S}(y)}\right) dy,$$

$$\widehat{\eta}_k = \left(\sum_{j=1}^{\nu} \frac{\widehat{\xi}_k\left(\widetilde{x}_j^{\star}, t\right)}{\left|\det\left(\mathcal{J}_h\left(\widetilde{x}_j^{\star}, t\right)\right)\right|}\right) * \phi_v$$

Distributional comparison: axiomatic approach

Candidates for validation distance

- Kullback-Leibler divergence $D_{KL}\left(\rho_1 || \rho_2\right) := \int_{\mathbb{R}^d} \rho_1\left(x\right) \log\left(\frac{\rho_1\left(x\right)}{\rho_2\left(x\right)}\right) dx$
- Symmetric KL divergence $D_{KL}^{\text{symm}}\left(\rho_1 || \rho_2\right) := \frac{1}{2}\left(D_{KL}\left(\rho_1 || \rho_2\right) + D_{KL}\left(\rho_2 || \rho_1\right)\right)$
- Wasserstein distance ${}_pW_q\left(\mu_1, \mu_2\right) := \left[\inf_{\mu \in \mathcal{M}_2(\mu_1,\mu_2)} \int_{\Omega} \parallel \underline{x} - \underline{y} \parallel_p^q \, d\mu\left(\underline{x}, \underline{y}\right)\right]^{1/q}$

What we want	D_{KL}	D_{KL}^{symm}	W
$\geqslant 0$	✓	✓	✓
Symmetry	×	✓	✓
Triangle inequality	×	×	✓
supp $(\eta) \neq$ supp $(\widehat{\eta})$	×	×	✓
dim(supp (η)) $\neq$ dim(supp $(\widehat{\eta})$)	×	×	✓
#sample $(\eta) \neq$ #sample $(\widehat{\eta})$	×	×	✓
Convexity	✓	✓	✓
Finite range	$[0, \infty)$	$[0, \infty)$	$[0,$ diam $(\Omega)]$

Distributional comparison: axiomatic approach

Wasserstein distance in validation context

- $_{p}W_{q}\left(\mu_1, \mu_2\right) = \left(\inf\limits_{\mu \in \mathcal{M}_2(\mu_1,\mu_2)} \mathbb{E}\left[\| \underline{x} - \underline{y} \|_p^q\right]\right)^{1/q}$
- Minimum effort required to convert one shape to another
- We choose $p = q = 2$, and denote $_{2}W_{2}$ as W
- Parametric interpretation: W depends on shape difference but not on shape i.e. for $e_r := \| m_r - \widehat{m}_r \|_2$, $W = W\left(\{e_r\}_{r \geqslant 1}\right)$

When can we write W in closed-form

- Single output case:
  $$_{p}W_{q}^{q}\left(\eta, \widehat{\eta}\right) = \int_{\mathbb{R}} \| F(x) - G(x) \|_p^q \, dx = \int_0^1 \| F^{-1}(u) - G^{-1}(u) \|_p^q \, du$$
- Multivariate Normal case (comparing Linear Gaussian systems):
  $$W\left((A, C); \left(\widehat{A}, \widehat{C}\right)\right) = W\left(\eta, \widehat{\eta}\right) = W\left(\mathcal{N}\left(\mu_1, \Sigma_1\right), \mathcal{N}\left(\mu_2, \Sigma_2\right)\right) =$$
  $$\sqrt{\| \mu_1 - \mu_2 \|_2^2 + \text{tr}\left(\Sigma_1\right) + \text{tr}\left(\Sigma_2\right) - 2\,\text{tr}\left(\left(\sqrt{\Sigma_1}\Sigma_2\sqrt{\Sigma_1}\right)^{1/2}\right)}$$

Distributional comparison: computing Wasserstein distance

W computation $\rightsquigarrow$ Monge-Kantorovich optimal transportation plan

- At each time $\{t_j\}_{j=1}^{M}$, we have two sets of colored scattered data
- Construct complete, weighted, directed bipartite graph $K_{m,n}\left(U \cup V, E\right)$ with $\#\left(U\right) = m$ and $\#\left(V\right) = n$
- Assign edge weight $c_{ij} := \parallel u_i - v_j \parallel_{\ell_2}^2$, $u_i \in U$, $v_j \in V$
- minimize $\displaystyle\sum_{i=1}^{m}\sum_{j=1}^{n} c_{ij}\,\varphi_{ij}$ subject to

$$\sum_{j=1}^{n} \varphi_{ij} = \alpha_i, \qquad \forall\, u_i \in U, \qquad \text{(C1)}$$

$$\sum_{i=1}^{m} \varphi_{ij} = \beta_j, \qquad \forall\, v_j \in V, \qquad \text{(C2)}$$

$$\varphi_{ij} \geqslant 0, \qquad \forall\, \left(u_i, v_j\right) \in U \times V. \qquad \text{(C3)}$$

- Necessary feasibility condition: $\displaystyle\sum_{i=1}^{m} \alpha_i = \sum_{j=1}^{n} \beta_j$

Distributional comparison: computing Wasserstein distance

Sample complexity

- Rate-of-convergence of empirical Wasserstein estimate
$$\mathbb{P}\left(\left|W\left(\eta_m, \widehat{\eta}_n\right) - W\left(\eta, \widehat{\eta}\right)\right| > \epsilon\right) \leqslant K_1 \exp\left(-\frac{m\epsilon^2}{32C_1}\right) + K_2 \exp\left(-\frac{n\epsilon^2}{32C_2}\right)$$

Runtime complexity

- An LP with mn unknowns and $(m+n+mn)$ constraints
- For $m=n$, runtime is $\mathcal{O}\left(dn^{2.5}\log n\right)$

Storage complexity

- For $m=n$, constraint is a binary matrix of size $2n \times n^2$
- Each row has n ones. Total # of ones $= 2n^2$
- At a given snapshot, sparse storage complexity is $2n\left(3n+d+1\right) = \mathcal{O}\left(n^2\right)$
- Non-sparse storage complexity is $2n\left(n^2+d+1\right) = \mathcal{O}\left(n^3\right)$

Construction of validation certificates: PRVC

How robust is the inference?

- Set of admissible initial densities: $\Psi := \{\xi_0^{(1)}, \xi_0^{(2)}, \ldots, \xi_0^{(N)}\}$
- At time step k, validation probability is $p\left(\gamma_k\right) := \mathbb{P}\left(W\left(\eta_k, \widehat{\eta}_k\right) \leqslant \gamma_k\right)$
- Let $V_k^i := \{\widehat{\eta}_k^{(i)}\left(y\right) \ : W\left(\eta_k^i, \widehat{\eta}_k^i\right) \leqslant \gamma_k\}$
- Empirical validation probability is $\widehat{p}_N\left(\gamma_k\right) := \dfrac{1}{N}\displaystyle\sum_1^N \mathbf{1}_{V_k^i}$
- (Chernoff bound) For any $\epsilon, \delta \in (0,1)$, if $N \geqslant N_{\text{ch}} := \dfrac{1}{2\epsilon^2}\log\dfrac{2}{\delta}$, then $\mathbb{P}\left(\left|p\left(\gamma_k\right) - \widehat{p}\left(\gamma_k\right)\right| < \epsilon\right) > 1 - \delta$

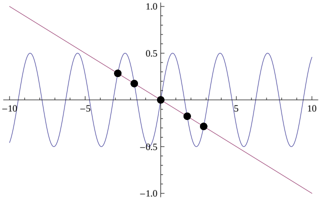
Construction of validation certificates: PRVC

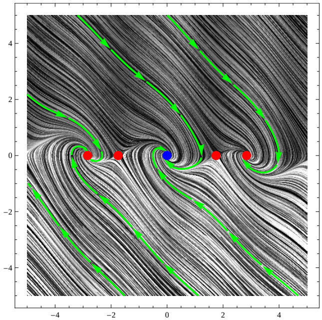
Algorithm 1 Construct PRVC

Require: $\epsilon, \delta \in (0,1)$, n, experimental data $\{\eta_k(y)\}_{k=1}^{M}$, model, tolerance vector $\{\gamma_k\}_{k=1}^{M}$

$N \leftarrow N_{\text{ch}}(\epsilon, \delta)$
Draw N random functions $\xi_0^{(1)}(\widetilde{x}), \xi_0^{(2)}(\widetilde{x}), \ldots, \xi_0^{(N)}(\widetilde{x})$
for $k = 1$ to T **do** ▷ Index for time step
  for $i = 1$ to N **do** ▷ Index for initial density
    for $j = 1$ to ν **do** ▷ Index for samples in extended state space, drawn from $\xi_0^{(i)}(\widetilde{x})$
      Propagate states using dynamics
      Propagate measurements
    end for
    Propagate state PDF
    Compute instantaneous output PDF
    Compute ${}_2W_2\left(\eta_k^{(i)}(y), \widehat{\eta}_k^{(i)}(y)\right)$ ▷ Distributional comparison by solving LP
    sum $\leftarrow 0$ ▷ Initialize
    if ${}_2W_2\left(\eta_k^{(i)}(y), \widehat{\eta}_k^{(i)}(y)\right) \leqslant \gamma_k$ **then** ▷ Check if valid
      sum $\leftarrow$ sum $+ 1$
    else
      do nothing
    end if
  end for
  $\widehat{p}_N(\gamma_k) \leftarrow \frac{\text{sum}}{N}$ ▷ Construct PRVC vector of length $M \times 1$
end for

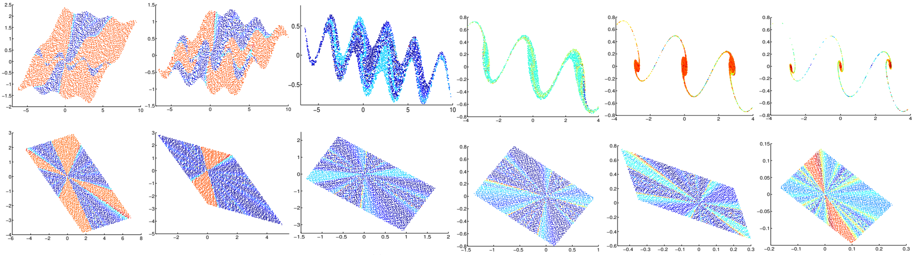
Example: Continuous-time model

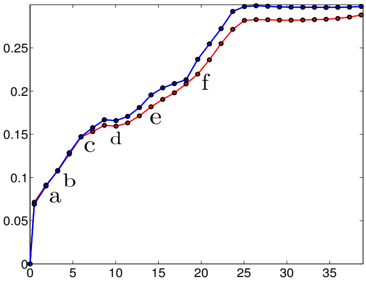




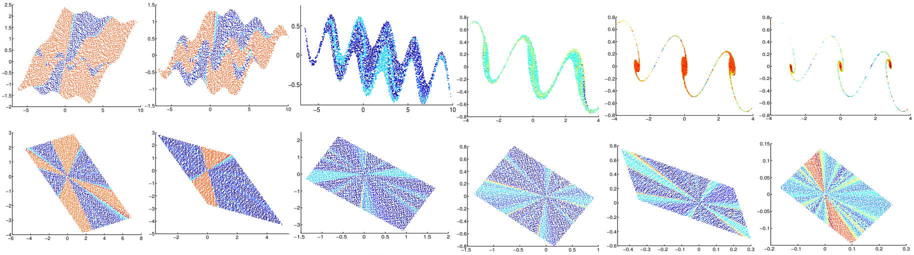
- Truth: $\ddot{x} = -ax - b\sin 2x - c\dot{x}$, $a = 0.1, b = 0.5, c = 1$.
- Five equilibria
- Model: Linearization about origin
- $\xi_0 = \mathcal{U}\left([-4,6] \times [-4,6]\right)$
- We plot time history of $\overline{W} := \dfrac{W\left(\eta_k, \widehat{\eta}_k\right)}{\text{diam}\left(\Omega_k\right)} \in [0,1]$

Example: Continuous-time model: $\overline{W}$ vs. t

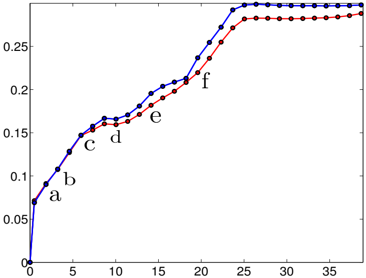

(a) (b) (c) (d) (e) (f)

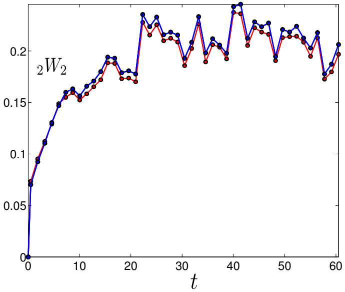


Example: Continuous-time model: $\overline{W}$ vs. t

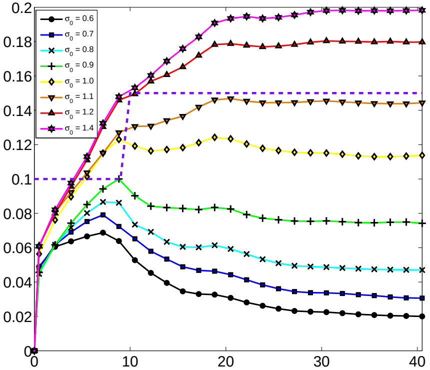


(a) (b) (c) (d) (e) (f)





Example: Continuous-time model: $\overline{W}$ vs. t



- $\xi_0^{(i)} = \mathcal{N}\left(0, \sigma_{0i}^2 \mathbf{I}\right)$
- $\text{PRVC}_{25\times 1} = \left[1, 1, 1, \frac{1}{2}, \frac{1}{2}, \frac{1}{2}, \frac{7}{8}, \underbrace{\frac{3}{4}, \ldots, \frac{3}{4}}_{18\,\text{times}}\right]^T$

Conclusions

- Unifying framework for nonlinear model validation
- Transport-theoretic Wasserstein distance as (in)validation measure
- Computable probabilistic validation certificate
- Can guide to model refinement