2011 50th IEEE Conference on Decision and Control and
European Control Conference (CDC-ECC)
Orlando, FL, USA, December 12-15, 2011

Model Validation: A Probabilistic Formulation

Abhishek Halder and Raktim Bhattacharya

***Abstract*— This paper presents a probabilistic formulation of the model validation problem. The proposed validation framework is simple, intuitive, and can account both deterministic and stochastic nonlinear systems in presence of parametric and nonparametric uncertainties. Contrary to the hard invalidation methods proposed in the literature, our formulation allows a relaxed notion of validation in probability. The construction of probabilistically robust validation certificates provides provably correct guarantees. Computational complexities and numerical examples are given to illustrate the method.**

I. Introduction

Model validation problem refers to the quantification of reliability to which a given model is consistent with the physical observations. It has been argued in the literature [1], [2] that the term 'model validation' is a misnomer since it would take infinite number of experimental observations to do so. Hence the term 'model invalidation' is preferred. In this paper, instead of hard invalidation, we will consider the validation/invalidation problem in a probabilistically relaxed sense.

Given the experimental measurements of the physical system in the form of a distribution, we compare the shape of this measured output distribution with that predicted by the model. At every instant of time, if the model-predicted distribution matches with the experimental one "reasonably well", we conclude that the model is validated to be a good candidate with some quantification of such "goodness".

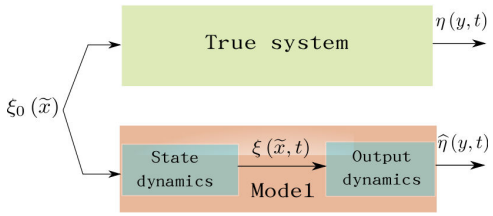

Fig. 1. The proposed model validation framework compares experimentally observed output PDF $\eta(y,t)$ with the model-predicted one $\widehat{\eta}(y,t)$, the comparison being made with respect to some suitable metric $J(\eta,\widehat{\eta})$.

Fig. 1 shows the outline of the proposed model validation framework. The experiment is carried out with the physical plant taking some initial distribution $\xi_0(\widetilde{x})$. Given the data for experimentally observed output distribution $\eta(y,t)$, one starts propagating the same initial state probability density function (PDF) through the proposed model's state dynamics, thereby computing instantaneous state PDF $\xi(\widetilde{x},t)$ and from it, obtains the output PDF $\widehat{\eta}(y,t)$ using the output dynamics prescribed by the model. If the output PDFs $\eta(y,t)$ and $\widehat{\eta}(y,t)$ are "close" in the sense that a suitable distance metric on the space of probability densities, $J(\eta,\widehat{\eta})$ remains small (within the specified tolerance level) at all times t when the experimental data are available, then it will be concluded that the model is "close" to the physical plant with some quantitative measure.

Since the basic idea relies on comparing the concentration of output trajectories at each instant of experimental observation, one can think of three distinct segments of such a model validation framework. These are

1) **Uncertainty propagation:** evolving state and output PDF using the proposed model.
2) **Distributional comparison:** measuring distance between the experimentally observed and model-predicted output PDFs and computing the margin by which the model-prediction obeys/violates the specified tolerance level.
3) **Construction of validation certificates:** probabilistic quantification of provably correct inference in this framework and providing sample complexity bounds for the same.

With respect to the literature, the contributions of this paper are as follows.

1) Instead of interval-valued structured uncertainty (as in H_∞ control framework [1]–[3]) or moment based uncertainty (as in parametric statistics framework [4]), we deal with model validation in the sense of nonparametric statistics by considering aleatoric uncertainty. In other words, the uncertainty in the model is quantified in terms of the PDFs of the associated random variables. We argue that such a formulation offers some advantages. *Firstly*, we show that model uncertainty in the parameters and initial states can be propagated accurately by spatio-temporally evolving their joint pdf. Since experimental data usually come in the form of histograms, it's a more natural quantification of uncertainty than specifying sets [5] to which the trajectories are contained at each instant of time. However, if needed, such sets can be recovered from the supports of the instantaneous pdfs. *Secondly*, as we'll see in Section IV, instead of simply invalidating a model, our methodology allows to estimate the probability that a proposed model is valid or invalid. This can help to decide which specific aspects of the model need further refinement. Hard invalidation methods don't cater such constructive information. *Thirdly*, the framework can

Abhishek Halder and Raktim Bhattacharya are with the Department of Aerospace Engineering, Texas A&M University, College Station, TX 77843, USA, {ahalder,raktim}@tamu.edu

handle both discrete-time and continuous-time nonlinear models which need not be polynomial. Previous work like [5] dealt with nonlinearities specified by semialgebraic sets and relied on sum of squares (SOS) decomposition [6] for computational tractability. From an implementation point of view, the approach presented in this paper doesn't require such conservatism.

2) Due to the uncertainties in initial conditions, parameters, process and measurement noise, one needs to compare output ensembles instead of comparing individual output realizations. This requires a metric to quantify closeness between the experimental data and the model in the sense of distribution. We use *Wasserstein distance* to compare the output pdfs and argue why common information-theoretic quantities like *Kullback-Leibler (KL) divergence* are not appropriate for this purpose.
3) We show that the uncertainty propagation through continuous or discrete-time dynamics can be done in a numerically efficient way, even when the model is high-dimensional and strongly nonlinear. Moreover, we outline how to compute the Wasserstein distance from scattered data, in this multivariate setting. Further, borrowing ideas from the analysis of randomized algorithms, we give sample-complexity bounds for probabilistically robust model validation.

The paper is organized as follows. Given a model, we outline the uncertainty propagation methodologies in section II. Section III describes how to compare the joint output PDFs for model validation. Next, section IV provides a constructive algorithm to compute probabilistically robust validation certificates to guarantee provably correct decisions. A numerical example is provided in section V to illustrate the efficacy of the proposed formulation. Section VI concludes the paper.

II. Uncertainty Propagation

A. Deterministic System

1) Continuous-time models: Consider the continuous-time nonlinear system with state dynamics given by the ODE

$$\dot{x} = f(x,p), \tag{1}$$

where $x(t) \in \mathcal{X} \subseteq \mathbb{R}^n$ is the state vector , $p \in \mathcal{P} \subseteq \mathbb{R}^p$ is the parameter vector, the dynamics $f(.,p) : \mathcal{X} \mapsto \mathbb{R}^n$ $\forall\, p \in \mathcal{P}$ and is at least locally Lipschitz . It can be put in an extended state space form

$$\dot{\widetilde{x}} = \widetilde{f}(\widetilde{x}), \tag{2}$$

by introducing $\widetilde{x} := \begin{Bmatrix} x \\ p \end{Bmatrix} \in \mathcal{X} \times \mathcal{P} \subseteq \mathbb{R}^{n+p}$, and $\widetilde{f} = \begin{Bmatrix} f_{n\times 1} \\ \mathbf{0}_{p\times 1} \end{Bmatrix}$. The output dynamics can be written as

$$y = h(\widetilde{x}), \tag{3}$$

where $y(t) \in \mathcal{Y} \subseteq \mathbb{R}^\ell$ is the output vector and $h : \mathcal{X} \times \mathcal{P} \mapsto \mathcal{Y}$ is a surjection. If uncertainties in the initial conditions $(x_0 := x(0))$ and parameters (p) are specified by the initial joint PDF $\xi_0(\widetilde{x})$, then the evolution of uncertainties subject to the dynamics (1), can be described by evolving the joint PDF $\xi(\widetilde{x},t)$ over the extended state space. Such spatio-temporal evolution of $\xi(\widetilde{x},t)$ is governed by the *Liouville equation* given by

$$\frac{\partial \xi}{\partial t} = \mathscr{L}\xi = D_1\xi = -\nabla.(\xi f) = -\sum_{i=1}^{n} \frac{\partial}{\partial x_i}(\xi f_i), \tag{4}$$

which is a quasi-linear partial differential equation (PDE), first order in both space and time. Notice that, the spatial operator $\mathscr{L}$ is a drift operator D_1 that describes the advection of the PDF in extended state space. The output PDF $\widehat{\eta}(y,t)$ can be computed from the state PDF as

$$\widehat{\eta}(y,t) = \sum_{j=1}^{\nu} \frac{\xi(\widetilde{x}_j^\star)}{|\det(\mathcal{J}(\widetilde{x}_j^\star))|} \tag{5}$$

where $\widetilde{x}_j^\star$ is the j^{th} root of the inverse transformation of (3) with $j = 1, 2, \ldots, \nu$. $\mathcal{J}$ is the Jacobian of this inverse transformation and det(.) stands for the determinant.

2) Discrete-time models: We start with the following two definitions.

Definition 1: Let $\mathcal{X} \times \mathcal{P} \subseteq \mathbb{R}^{n+p}$ be a compact set and let $\mathcal{B}(\mathcal{X} \times \mathcal{P})$ be the Borel-σ algebra defined on it. With respect to the measure space $(\mathcal{X} \times \mathcal{P}, \mathcal{B}, \mu)$, a transformation $\mathcal{T} : \mathcal{X} \times \mathcal{P} \mapsto \mathcal{X} \times \mathcal{P}$ is called measurable if $\mathcal{T}^{-1}(B) \in \mathcal{B}$, $\forall B \in \mathcal{B}$.

Definition 2: A measurable transformation $\mathcal{T} : \mathcal{X} \times \mathcal{P} \mapsto \mathcal{X} \times \mathcal{P}$ is said to be nonsingular on the measure space $(\mathcal{X} \times \mathcal{P}, \mathcal{B}, \mu)$, if $\mu\left(\mathcal{T}^{-1}(B)\right) = 0 \quad \forall B \in \mathcal{B}$ such that $\mu(B) = 0$.

Consider the discrete-time nonlinear system with state dynamics given by the vector recurrence relation

$$\widetilde{x}_{k+1} = \mathcal{T}(\widetilde{x}_k), \tag{6}$$

where $\mathcal{T} : \mathcal{X} \times \mathcal{P} \mapsto \mathcal{X} \times \mathcal{P}$ is a measurable nonsingular transformation and the time index k takes values from the ordered index set of non-negative integers $\{0, 1, 2, \ldots\}$. Then the evolution of the joint pdf $\xi(\widetilde{x}_k)$ is dictated by the *Perron-Frobenius operator* $\mathscr{P}$, given by

$$\int_B \mathscr{P}\xi(\widetilde{x}_k)\ \mu(d\widetilde{x}_k) = \int_{\mathcal{T}^{-1}(B)} \xi(\widetilde{x}_k)\ \mu(d\widetilde{x}_k) \tag{7}$$

for $B \in \mathcal{B}$. Equation (7) ensures that the evolution of $\xi(\widetilde{x}_k)$ is Markov and it respects the properties of a PDF. Further, assuming the output dynamics as $y_k = h(\widetilde{x}_k)$, one can derive $\widehat{\eta}(y_k)$ from $\xi(\widetilde{x}_k)$ using the discrete analogue of the transformation rule (5).

B. Stochastic System

1) Continuous-time models: Consider the continuous-time nonlinear system with state dynamics given by the Itô SDE

$$d\widetilde{x} = \widetilde{f}(\widetilde{x})\ dt + g(\widetilde{x})\ dW, \tag{8}$$

where $W(t) \in \mathbb{R}^{\omega}$ is the ω-dimensional Wiener process at time t, and the noise coupling $g : \mathcal{X} \times \mathcal{P} \mapsto \mathbb{R}^{n \times \omega}$. For the Wiener process $W(t)$, at all times

$$\mathbb{E}[dW_i] = 0,\ \mathbb{E}[dW_i dW_j] = Q_{ij} = \alpha_i\, \delta_{ij}\ \forall\, i, j = 1, \ldots, \omega,$$

where $\mathbb{E}[.]$ stands for the expectation operator and δ_{ij} is the Kronecker delta. Thus $Q \in \mathbb{R}^{\omega \times \omega}$ with $\alpha_i > 0\ \forall\, i = 1, 2, \ldots, \omega$, being the noise strength. The output dynamics is still assumed to be given by (3). In such a setting, the evolution of the state PDF $\xi(\widetilde{x}, t)$ subject to (8) is governed by the *Fokker-Planck equation*, also known as *forward Kolmogorov equation*

$$\frac{\partial \xi}{\partial t} = \mathscr{L}\xi = (D_1 + D_2)\,\xi$$

$$= -\sum_{i=1}^{n} \frac{\partial}{\partial x_i}(\xi f_i) + \sum_{i=1}^{n}\sum_{j=1}^{n} \frac{\partial^2}{\partial x_i \partial x_j}\left(\left(gQg^T\right)_{ij}\xi\right), \quad (9)$$

which is a homogeneous parabolic PDE, second order in space and first order in time. In this case, the spatial operator $\mathscr{L}$ can be written as a sum of a *drift operator* (D_1) and a *diffusion operator* (D_2). The diffusion term accounts for the smearing of the PDF due to process noise. Once the state PDF is computed through (9), the output PDF can again be obtained from (5).

2) Discrete-time models: In this case, we consider the nonlinear state space representation given by the stochastic maps of general form

$$\widetilde{x}_{k+1} = \mathcal{T}(\widetilde{x}_k, \zeta_k), \qquad \widetilde{y}_k = h(\widetilde{x}_k, \zeta_k), \quad (10)$$

where $\zeta_k \in \mathbb{R}^{\omega}$ is the i.i.d. sample drawn from a known distribution for the noise (stochastic perturbations). Here, the dynamics $\mathcal{T}$ is not required to be a non-singular transformation. Since $\mathcal{T}$ defines a Markov chain on $\mathcal{X} \times \mathcal{P}$, it can be shown that [7] evolution of the joint PDFs follow

$$\xi_{k+1} := \xi_{\widetilde{x}_{k+1}}(\widetilde{x}) = \int_{\mathcal{X}\times\mathcal{P}} \mathcal{K}_{\mathcal{T}}(\widetilde{x}|z)\,\xi_{\widetilde{x}_k}(z)\,dz, \quad (11)$$

$$\widehat{\eta}_k := \widehat{\eta}_{y_k}(y) = \int_{\mathcal{X}\times\mathcal{P}} \mathcal{K}_h(y|z)\,\xi_{\widetilde{x}_k}(z)\,dz, \quad (12)$$

where $\mathcal{K}_{\mathcal{T}}(\widetilde{x}|z)$ and $\mathcal{K}_h(y|z)$ are known as the *stochastic kernels* for the maps $\mathcal{T}$ and h respectively. (12) can be seen as a special case of the Chapman-Kolmogorov equation [8].

C. Computational Aspects

For deterministic flow, the Liouville PDE (4) can be solved in exact arithmetic [9] via method-of-characteristics (MOC) [10]. Since the characteristic curves for (4) are the trajectories in the extended state space, $\xi(\widetilde{x}, t)$ and hence $\widehat{\eta}(y, t)$ can be computed directly along these characteristics. Unlike Monte-Carlo, this is an "on-the-fly" computation and does not involve any approximation, and hence offers a superior performance [11]–[13] than Monte-Carlo in high dimensions. For deterministic maps, cell-to-cell mapping [14] achieves a finite dimensional approximation of the Perron-Frobenius operator.

TABLE I
COMPARISON OF KL DIVERGENCE AND WASSERSTEIN DISTANCE FOR MODEL VALIDATION

Validation requirements	KL divergence	Wasserstein distance
Shape comparison	No	Yes
Metric	No	Yes
Support robustness	No	Yes
Sampling robustness	No	Yes
Consistency	Yes	Yes
Rate of convergence	Arbitrarily slow	Fast

For stochastic flow, solving Fokker-Planck PDE (9) is numerically challenging [15] but has seen some recent success [16] in moderate (4 to 5) dimensions. For stochastic maps, discretizations for stochastic kernels (11) and (12), can be done through cell-to-cell mapping [14] resulting a random transition probability matrix [17].

III. COMPARING OUTPUT PROBABILITY DENSITY FUNCTIONS

Once the observed and model-predicted output PDFs $\eta(y, t)$ and $\widehat{\eta}(y, t)$, are obtained, we need a metric to compare the shapes of these two PDFs at each time t, where the measurement PDF $\widehat{\eta}(y, t)$ is available. We argue that the suitable metric for this purpose is Wasserstein distance.

A. Choice of Metric

Distances on the space of probability distributions can be broadly categorized into two classes, viz. Csisźar's ϕ-divergence [18] and integral probability metrics [19]. The first includes well-known distances like Kullback-Leibler (KL) divergence, Hellinger distance, χ^2 divergence etc. while the latter includes Wasserstein distance, Dudley metric, maximum mean discrepancy etc. Total variation distance belongs to both of these classes.

The choice of a suitable metric depends on application. In our context of model validation, the objective is to measure the shape difference between two instantaneous output PDFs. This is because a good model must emulate similar concentration of trajectories as observed in the measurement space. In other words, the respective joint PDFs $\eta(y, t)$ and $\widehat{\eta}(y, t)$, over the time-varying output supports, must match at times whenever measurements are available. We formally introduce the Wasserstein distance below.

Let M be a complete, separable metric (Polish) space with a p^{th} order distance metric d_p. For simplicity, we let M to be $\mathbb{R}^n$ and take d_p as the L^p norm. Then the Wasserstein distance of order q, denoted as ${}_pW_q$, between two Borel probability measures μ_1 and μ_2 on $\mathbb{R}^n$ is defined as

$${}_pW_q^q(\mu_1, \mu_2) := \inf_{\mu \in \mathcal{M}(\mu_1, \mu_2)} \int_{\mathbb{R}^{2n}} \|\underline{x} - \underline{y}\|_p^q\ d\mu(\underline{x}, \underline{y}) \quad (13)$$

where $\mathcal{M}(\mu_1, \mu_2)$ is the set of all probability measures on $\mathbb{R}^{2n}$ with first marginal μ_1 and second marginal μ_2. It's well known [20] that on the set of Borel measures on $\mathbb{R}^n$ having finite second moments, ${}_pW_q$ defines a metric. The

advantages of transport-theoretic Wasserstein distance over information-theoretic KL divergence, for validation purpose, is summarized in Table I.

B. Computing Quadratic Wasserstein Distance: LP Formulation

Computing Wasserstein distance from (13) calls for solving *Monge-Kantorovich optimal transportation plan* [20]. In this formulation, the difference in shape between two statistical distributions is quantified by the minimum amount of work required to convert a shape to the other. The ensuing optimization, known as *Hitchcock-Koopmans problem* [21]–[23], can be cast as a linear program (LP), as described next.

Consider a complete, weighted, directed bipartite graph $K_{m,n}\left(U \cup V, E\right)$ with $\#\left(U\right)=m$ and $\#\left(V\right)=n$. If $u_i \in U, i=1,\ldots,m$, and $v_j \in V, j=1,\ldots,n$, then the edge weight $c_{ij} :=\parallel u_i - v_j \parallel_{\ell_2}^2$ denotes the cost of transporting unit mass from vertex u_i to v_j. Then, according to (13), computing ${}_2W_2^2$ translates to

$$\text{minimize} \sum_{i=1}^{m}\sum_{j=1}^{n} c_{ij}\,\varphi_{ij} \tag{14}$$

subject to the constraints

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

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

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

The objective of the LP is to come up with an optimal mass transportation policy $\varphi_{ij} := \varphi\left(u_i \rightarrow v_j\right)$ associated with cost c_{ij}. Clearly, in addition to constraints (C1)–(C3), (14) must respect the necessary feasibility condition

$$\sum_{i=1}^{m}\alpha_i=\sum_{j=1}^{n}\beta_j \tag{C0}$$

denoting the conservation of mass. In our context of measuring the shape difference between two PDFs $\eta\left(y,t\right)$ and $\widehat{\eta}\left(y,t\right)$, we treat their joint probability mass function (PMF) vectors α_i and β_j to be the marginals of some unknown joint PMF φ_{ij} supported over the product space $U \times V$. Since determining joint with given marginals is not unique, (14) strives to find that particular joint PMF which minimizes the total cost for transporting the probability mass while respecting the normality condition.

C. Computational Complexity

For $m=n$, the LP formulation (14), (C1)–(C3) requires solving for n^2 unknowns subject to $\left(n^2+2n\right)$ constraints. Since the LP has only linear dependence on dimensions d and for a fixed dimension, it can be solved [24] in $O\left(n^{2.5}\log n\right)$ operations, the total runtime complexity for (14) is $O\left(d\,n^{2.5}\log n\right)$.

IV. Probabilistically Robust Validation Certificates

Often in practice, the exact initial density is not known to facilitate our model validation framework; instead a class of densities may be known. For example, it may be known that the initial density is symmetric unimodal but it's exact shape (e.g. normal, semi-circular etc.) may not be known. Even when the distribution-type is known (e.g. normal), it's often difficult to pinpoint the parameter values describing the initial density function. To account such scenarios, consider a random variable $\Delta : \Omega \rightarrow E$, that induces a probability triplet $(\Omega, \mathcal{F}, \mathbb{P})$ on the space of initial densities. Here $E \subset \Omega$ and $\#\left(E\right)=1$. The random variable Δ can be thought of as a categorical random variable which picks up an initial density from the collection of admissible initial densities $\Omega := \{\xi_0^{(1)}\left(\widetilde{x}\right), \xi_0^{(2)}\left(\widetilde{x}\right), \ldots\}$ according to the law of Δ. For example, if we know $\xi_0 \sim \mathcal{N}\left(\mu, \sigma^2\right)$ with a given joint distribution over the $\mu\ \sigma^2$ space, then in our model validation framework, one sample from this space will return one distance measure between the instantaneous output PDFs. How many such $\left(\mu, \sigma^2\right)$ samples are necessary to guarantee the robustness of the model validation oracle? The Chernoff bound provides such an estimate for finite sample complexity.

At time step k, let the *validation probability* be $p\left(\gamma_k\right) := \mathbb{P}\left({}_2W_2\left(\eta_k\left(y\right), \widehat{\eta}_k\left(y\right)\right) \leqslant \gamma_k\right)$. Here $\gamma_k \in \mathbb{R}^+$ is the prescribed instantaneous tolerance level. If the model validation is performed by drawing N samples from Ω, then the *empirical validation probability* is $\widehat{p}_N\left(\gamma_k\right) := \dfrac{1}{N}\displaystyle\sum_{i=1}^{N}\chi_{V_k^{(i)}}$ where $V_k^{(i)} := \{\widehat{\eta}_k^{(i)}\left(y\right) : {}_2W_2\left(\eta_k^{(i)}\left(y\right), \widehat{\eta}_k^{(i)}\left(y\right)\right) \leqslant \gamma_k\}$. Consider $\epsilon, \delta \in (0,1)$ as the desired accuracy and confidence, respectively.

Lemma 1: (**Chernoff bound**) [25] For any $\epsilon, \delta \in (0,1)$, if $N \geqslant N_{ch} := \dfrac{1}{2\epsilon^2}\log\dfrac{2}{\delta}$, then $\mathbb{P}\left(|p\left(\gamma_k\right) - \widehat{p}_N\left(\gamma_k\right)| < \epsilon\right) > 1-\delta$.

The above lemma allows us to construct *probabilistically robust validation certificate* (PRVC) $\widehat{p}_N\left(\gamma_k\right)$ through the algorithm below. The PRVC vector, with ϵ accuracy, returns the probability that the model is valid at time k, in the sense that the instantaneous output PDFs are no distant than the required tolerance level γ_k. Lemma 1 lets the user control the accuracy ϵ and the confidence δ, with which the preceding statement can be made. In practice, $\{\gamma_k\}_{k=1}^{T}$ is often specified as percentage tolerance. Since ${}_2W_2\left(\eta_k^{(i)}\left(y\right), \widehat{\eta}_k^{(i)}\left(y\right)\right) \in [0, \operatorname{diam}\left(\mathcal{D}_k\right)]$, where $\mathcal{D}_k := \mathcal{D}_k^{\text{experiment}} \times \mathcal{D}_k^{\text{model}}$, is the product of the experimental and model output space at time step k, a normalized Wasserstein distance $\dfrac{{}_2W_2}{\operatorname{diam}\left(\mathcal{D}_k\right)}$ can be employed for comparison purposes. Thus the framework enables us to compute a provably correct validation certificate on the face of uncertainty with finite sample complexity.

Algorithm 1 Construct PRVC

Require: $\epsilon, \delta \in (0,1)$, T, ν, law of Δ, experimental data $\{\eta_k(y)\}_{k=1}^{T}$, model, tolerance vector $\{\gamma_k\}_{k=1}^{T}$

$N \leftarrow N_{\text{ch}}(\epsilon, \delta)$ ▷ Using lemma 1
Draw random functions $\xi_0^{(1)}(\widetilde{x}), \xi_0^{(2)}(\widetilde{x}), \ldots, \xi_0^{(N)}(\widetilde{x})$ according to the law of Δ
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** ▷ Samples drawn from $\xi_0^{(i)}(\widetilde{x})$
      Propagate states using dynamics
      Propagate measurements
    end for
    Propagate state PDF ▷ Use (4), (9), (7) or (12)
    Compute instantaneous output PDF
    Compute ${}_2W_2\left(\eta_k^{(i)}(y), \widehat{\eta}_k^{(i)}(y)\right)$ ▷ Distributional comparison by solving LP (14) subject to (C0)–(C3)
    sum $\leftarrow 0$ ▷ Initialize
    if ${}_2W_2\left(\eta_k^{(i)}(y), \widehat{\eta}_k^{(i)}(y)\right) \leqslant \gamma_k$ **then**
      sum $\leftarrow$ sum + 1
    else
      do nothing
    end if
  end for
  $\widehat{p}_N(\gamma_k) \leftarrow \dfrac{\text{sum}}{N}$ ▷ Construct PRVC vector
end for

V. Numerical Example

Consider a nonlinear system originally governed by

$$\dot{x}_1 = -x_2, \qquad \dot{x}_2 = \sin x_1, \tag{15}$$

with outputs $y_1 = x_1$, $y_2 = x_2$. However, this true dynamics is not known to the modeler; only the y_1 and y_2 data are observed over time and recorded as the joint histogram over the y_1y_2 space. Suppose the following is proposed as a candidate model.

$$\dot{x}_1 = -x_2, \qquad \dot{x}_2 = x_1, \tag{16}$$

with outputs same as the states. Given this proposed model, our job is to assess the goodness of it against the measured histogram over the output space. We emphasize here that in this example, the purpose of (15) is only to create the synthetic data and to demonstrate the proof-of-concept. In a realistic model validation, the data arrives from experimental measurements, not from another model. To illustrate the formulation, we consider another proposed model to be validated, given by

$$\dot{x}_1 = -x_2, \qquad \dot{x}_2 = x_1 - \frac{x_1^3}{3!} + \frac{x_1^5}{5!}, \tag{17}$$

with outputs same as states. The phase portraits for (15), (16) and (17) are shown in Fig. 2 and we intuitively expect (17) to be closer to (15) than (16). Fig. V and V indeed confirm these intuitions. The initial joint PDF $\xi_0(x_1, x_2)$ is taken to be uniform over $[-1, 3]^2$.

Fig. V and V show the discrepancy in the mean vectors between the measured and model-predicted output PDFs for (15) and (16), and for (15) and (17), respectively. Similar plots are shown in Fig. V and V, for the discrepancy in measured and model-predicted covariance matrices. These plots reveal that moment-based parametric analysis may not capture the non-parametric shape discrepancy between the PDFs. All moments are computed from the scattered data evolving under nonlinear dynamics, using quasi-Monte Carlo (QMC) integration [26]. Given a tolerance vector $\{\gamma_k\}_{k=1}^{T}$, one can immediately obtain the PRVC following algorithm 1. Such a sample computation is omitted for brevity.

VI. Conclusions

A probabilistic notion of model validation is introduced in this paper that can account deterministic and stochastic nonlinear systems on the face of uncertainties. The theoretical underpinnings and implementation details are outlined along with a simple example. An algorithm is provided to construct a probabilistically robust validation certificate. The authors plan to apply the framework to the validation of large-scale physics-based models against experimental data and controller V&V.

References

[1] R.S. Smith, and J.C. Doyle, "Model Validation: A Connection Between Robust Control and Identification", *IEEE Transactions on Automatic Control*, Vol. 37, No. 7, 1992, pp. 942–952.

[2] K. Poolla, P. Khargonekar, A. Tikku, J. Krause, and K. Nagpal, "A Time-domain Approach to Model Validation", *IEEE Transactions on Automatic Control*, Vol. 39, No. 5, 1994, pp. 951–959.

[3] D. Xu, Z. Ren, G. Gu, J. Chen, "LFT Uncertain Model Validation with Time and Frequency Domain Measurements", *IEEE Transactions on Automatic Control*, Vol. 44, No. 7, 1999, pp. 1435–1441.

[4] R.G. Ghanem, A. Doostan, and J. Red-Horse, "A Probabilistic Construction of Model Validation", *Computer Methods in Applied Mechanics and Engineering*, Vol. 197, No. 29-32, 2008, pp. 2585–2595.

[5] S. Prajna, "Barrier Certificates for Nonlinear Model Validation", *Automatica*, Vol. 42, No. 1, 2006, pp. 117–126.

[6] P.A. Parrilo, *Structured Semidefinite Programs and Semialgebraic Geometry Methods in Robustness and Optimization*, PhD thesis, California Institute of Technology, 2000, Pasadena; CA.

[7] S. Meyn, and R.L. Tweedie, *Markov Chains and Stochastic Stability*, Cambridge University Press, Cambridge; 2009.

[8] A. Papoulis, *Random Variables and Stochastic Processes*, McGraw-Hill, NY; 1984.

[9] A. Halder, and R. Bhattacharya, "Dispersion Analysis in Hypersonic Flight During Planetary Entry Using Stochastic Liouville Equation", *Journal of Guidance, Control and Dynamics*, Vol. 34, No. 2, 2011, pp. 459–474.

[10] L.C. Evans, *Partial Differential Equations*, Graduate Studies in Mathematics, Vol. 19, American Mathematical Society, RI; 1998.

[11] A. Halder, and R. Bhattacharya, "Beyond Monte Carlo: A Computational Framework for Uncertainty Propagation in Planetary Entry, Descent and Landing", *AIAA Guidance, Navigation and Control Conference*, Toronto, ON, 2010.

[12] A. Halder, and R. Bhattacharya, "Performance Bounds for Dispersion Analysis: A Comparison Between Monte Carlo and Perron-Frobenius Operator Approach", Preprint, available at http://people.tamu.edu/~ahalder/PerformanceBoundsvSubmitted.pdf.

[13] P. Dutta, and R. Bhattacharya, "Hypersonic State Estimation using Frobenius-Perron Operator", *Journal of Guidance, Control and Dynamics*, Vol. 34, No. 2, 2011, pp. 325–344.

[14] C.S. Hsu, *Cell-to-Cell Mapping: A Method of Global Analysis for Nonlinear Systems*, Applied Mathematical Sciences, Vol. 64, Springer-Verlag, NY; 1987.

[15] H. Risken, *The Fokker-Planck Equation: Methods of Solution and Applications*, Springer-Verlag, NY; 1989.

[16] M. Kumar, S. Chakravorty, and J.L. Junkins, "A Semianalytic Meshless Approach to the Transient Fokker-Planck Equation", *Probabilistic Engineering Mechanics*, Vol. 25, No. 3, 2010, pp. 323–331.

[17] A. Edelman, and B.D. Sutton, "From Random Matrices to Stochastic Operators", *Journal of Statistical Physics*, Vol. 127, No. 6, 2007, pp. 1121–1165.

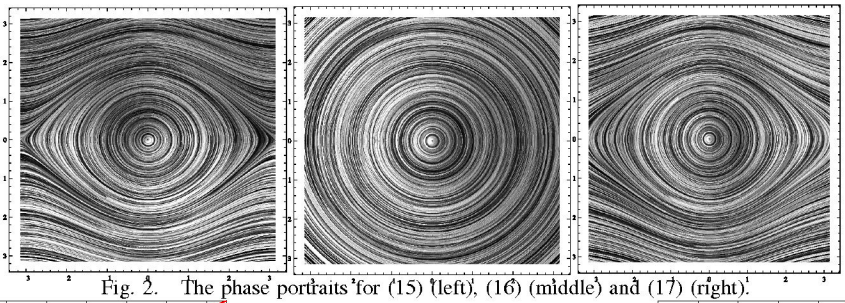

Fig. 2. The phase portraits for (15) (left), (16) (middle) and (17) (right).

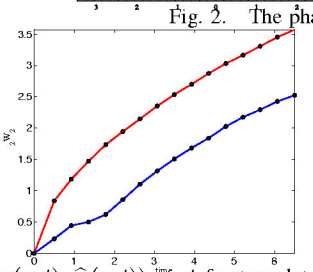

Fig. 3. $_2W_2\left(\eta\left(y,t\right),\widehat{\eta}\left(y,t\right)\right)$ vs. t for true data from (15) and model-prediction from (16). The *red* curve results when the joint PMFs in LP formulation (14), (C1)–(C3), are computed from the Liouville PDFs obtained from a MOC implementation of (4). The *blue* curve results when the joint PMFs are computed from MC histograms.

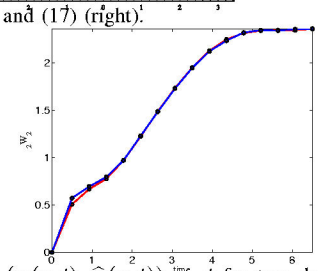

Fig. 4. $_2W_2\left(\eta\left(y,t\right),\widehat{\eta}\left(y,t\right)\right)$ vs. t for true data from (15) and model-prediction from (17). The color conventions are same as Fig. 3. Notice that $_2W_2$ for this case stays lower than that of Fig. 3, as we expect intuitively.

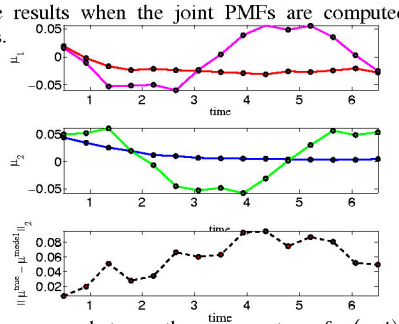

Fig. 5. Discrepancy between the mean vectors of $\eta\left(y,t\right)$ and $\widehat{\eta}\left(y,t\right)$ for (15) and (16), as function of t. The color codes are μ_1^{true} (red), μ_1^{model} (magenta), μ_2^{true} (blue), μ_2^{model} (green).

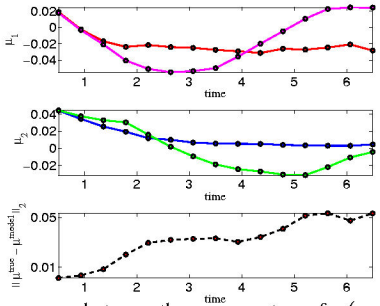

Fig. 6. Discrepancy between the mean vectors of $\eta\left(y,t\right)$ and $\widehat{\eta}\left(y,t\right)$ for (15) and (17), as function of t. The color codes are μ_1^{true} (red), μ_1^{model} (magenta), μ_2^{true} (blue), μ_2^{model} (green).

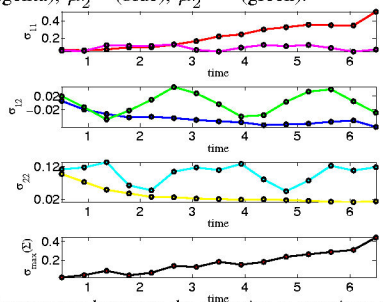

Fig. 7. Discrepancy between the covariance matrices of $\eta\left(y,t\right)$ and $\widehat{\eta}\left(y,t\right)$ for (15) and (16), as function of t. The color codes are $\sigma_{11}^{\text{true}}$ (red), $\sigma_{11}^{\text{model}}$ (magenta), $\sigma_{12}^{\text{true}}$ (blue), $\sigma_{12}^{\text{model}}$ (green), $\sigma_{22}^{\text{true}}$ (yellow), $\sigma_{22}^{\text{model}}$ (cyan). The bottom row shows the trajectory of maximum singular value of $\Sigma := \Sigma^{\text{true}} - \Sigma^{\text{model}}$, reflecting the maximum dispersion in covariance discrepancy.

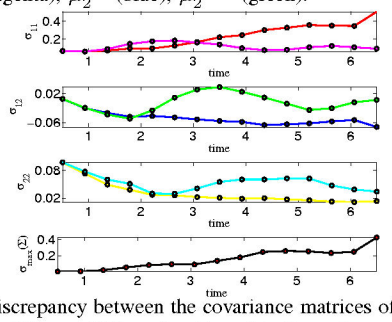

Fig. 8. Discrepancy between the covariance matrices of $\eta\left(y,t\right)$ and $\widehat{\eta}\left(y,t\right)$ for (15) and (17), as function of t. The color codes are same as in Fig. 7.

[18] I. Csiszár, "Information-type Measures of Difference of Probability Distributions and Indirect Observations", *Studia Scientiarium Mathematicarum Hungarica*, Vol. 2, 1967, pp. 299–318.

[19] A. Müller, "Integral Probability Metrics and Their Generating Classes of Functions", *Advances in Applied Probability*, Vol. 29, 1997, pp. 429–443.

[20] S.T. Rachev, "The Monge–Kantorovich Mass Transference Problem and Its Stochastic Applications", *Theory of Probability and its Applications*, Vol. 29, 1985, pp. 647–676.

[21] F. Hitchcock, "The Distribution of a Product from Several Sources to Numerous Localities", *Journal of Mathematics and Physics*, Vol. 20, No. 2, 1941, pp. 224–230.

[22] T.C. Koopmans, "Optimum Utilization of the Transportation System", *Econometrica: Journal of the Econometric Society*, Vol. 17, 1949, pp. 136–146.

[23] T.C. Koopmans, "Efficient Allocation of Resources", *Econometrica: Journal of the Econometric Society*, Vol. 19, No. 4, 1951, pp. 455–465.

[24] R.E. Burkard, M. Dell'Amico, and S. Martello, *Assignment Problems*, SIAM, PA; 2009.

[25] R. Tempo, G. Calafiore, and F. Dabbene, *Randomized Algorithms for Analysis and Control of Uncertain Systems*, Springer, NY; 2004.

[26] H. Niederreiter, *Random Number Generation and Quasi-Monte Carlo Methods*, CBMS-NSF Regional Conference Series in Applied Mathematics; 1992.