INPUT-OUTPUT REDUCED-ORDER MODELING FOR PUBLIC HEALTH INTERVENTION EVALUATION

ALEX VIGUERIE[1]**, CHIARA PIAZZOLA**[2]**, MD HAFIZUL ISLAM**[3]**, AND EVIN UZUN JACOBSON**[4]

[1,3,4] Division of HIV Prevention
Centers for Disease Control and Prevention
1600 Clifton Road
Atlanta, GA (USA) 30333
[1]xjm9@cdc.gov, [3]qla0@cdc.gov, [4]wqm4@cdc.gov

[2] Department of Mathematics
Technical University of Munich
Boltzmannstraße 3
85748, Garching bei München, Germany
chiara.piazzola@tum.de

Disclaimer: *The views and findings expressed in this document are strictly those of the authors and do not necessarily reflect those of the Centers for Disease Control and Prevention.*

Key words: Reduced-order modeling, dynamical systems, infectious disease.

Abstract. In recent years, mathematical models have become an indispensable tool in the planning, evaluation, and implementation of public health interventions. Models must often provide detailed information for many levels of population stratification. Such detail comes at a price: in addition to the computational costs, the number of considered input parameters can be large, making effective study design difficult. To address these difficulties, we propose a novel technique to reduce the dimension of the model input space to simplify model-informed intervention planning. The method works by first applying a dimension reduction technique on the model output space. We then develop a method which allows us to map each reduced output to a corresponding vector in the input space, thereby reducing its dimension. We apply the method to the HIV Optimization and Prevention Economics (HOPE) model, to validate the approach and establish proof of concept.

1 INTRODUCTION

In recent years, mathematical models have become an indispensable tool in the planning, evaluation, and implementation of public health interventions. Models often need to provide detailed information for many distinct levels of population stratification. Such detail comes at a price: in addition to computational costs, the number of considered input parameters can be

large. This leads to difficulty at all stages of the modeling workflow. For example, effective study designs can be difficult to develop and computationally demanding, and optimization algorithms are slow to converge and may exhibit (perhaps infinitely) many local solutions. Post-processing of model results can also be challenging, as the large number of varied parameters can make drawing concrete, actionable conclusions from simulation results difficult.

Recent years have seen significant progress in *reduced-order modeling* (ROM) techniques to reduce the dimension of mathematical models. Particularly, popular techniques include approaches based on the Proper Orthogonal Decomposition (POD)[1;2], surrogate modeling[3], and Dynamic Mode Decomposition (DMD) and related methods based on Koopman theory[4;5].

Such methods have been applied in public health for short-term forecasting and diagnostic applications[6]. However, intervention planning generally requires multiple model evaluations for different input configurations. Hence, ROM techniques that are useful for forecasting purposes, may not be applicable. Instead, intervention planning involves observing how model outputs *change* in response to variation in model inputs. Thus, using ROMs for intervention planning requires an approach that can incorporate parameter-dependence. While ROM techniques have been developed for parameterized problems, many of these applications are focused on cases in which the primary difficulty arises from the high dimension of the *output* space[1;2;5;7].

However, in intervention planning, most of the difficulty arises from the dimension of the *input space*, a problem which has received less attention. A recent work[8] explored the issue of extracting surrogate models from problems with high-dimensional input spaces, using a combination of dimension reduction techniques and Bayesian distribution fitting. Other notable works in this area focused on constructing surrogate models[9], and input-output ROM schemes for experimental data[10]. Active subspace (AS) and related methods have also been used to reduce input dimensionality for scalar-valued problems[11], and have been applied to problems in public health[12]. However, AS methods are applicable to scalar-valued functions, and applications to vector-valued problems typically require the use of a scalar-valued surrogate; for example, using the basic reproduction number as a surrogate for a model of infectious disease transmission[12].

We propose herein a novel method for input parameter-space reduction. The method works by applying dimension-reduction techniques over an aggregation of model outputs, generated over a range of input parameters. Through surrogate modeling over the reduced-order space, we construct a mapping which maps each reduced-order output to a vector of input parameters in a reduced-order input space. The reduced-order inputs can then be used directly to identify intervention plans based on program targets and constraints.

The paper is outlined as follows. In Section 2, we briefly introduce our model problem. In Section 3, we explain the proposed method for input/output model order reduction. In Section 4, we demonstrate the potential utility of the method by applying it on a large-scale model of HIV transmission, the HIV Optimization and Prevention Economics (HOPE) model. We conclude by summarizing our findings and discussing directions for future work.

2 BASIC MODEL AND PRELIMINARIES

Let us consider a nonlinear continuous-time dynamical system consisting of n unknowns and depending on an m-dimensional vector $\widehat{\boldsymbol{\theta}} = [\widehat{\theta}_1, \ldots, \widehat{\theta}_m] \in \mathbb{R}^m$ of parameters, assumed constant in time. Denote as $\Gamma = \Gamma_1 \times \Gamma_2 \times \ldots \times \Gamma_m$ the set of all possible values of $\widehat{\boldsymbol{\theta}}$. Each Γ_i is assumed to be closed and bounded herein, though this assumption is not strictly necessary.

We may then write:

$$
\begin{aligned}
\frac{d\boldsymbol{x}}{dt} &= F(t, \boldsymbol{x}, \widehat{\boldsymbol{\theta}}) && \text{for } t_0 < t \le t_{end}, \\
\boldsymbol{x} &= \boldsymbol{x}_0 && \text{at } t = t_0,
\end{aligned}
\tag{1}
$$

with F a nonlinear function depending on $\widehat{\boldsymbol{\theta}}$, as well the solution state $\boldsymbol{x}$ and time t. We assume that the system (1) is sufficiently regular such that the solution $\boldsymbol{x}$ is time-continuous for all $\widehat{\boldsymbol{\theta}} \in \Gamma$ and that the solution is continuous in $\widehat{\boldsymbol{\theta}}$ at all t for all $\widehat{\boldsymbol{\theta}} \in \Gamma$.

Let us assume that $\widehat{\theta}_i$, $i = 1, \ldots, m$ are independently distributed random variables taking values in Γ_i and $\widehat{\rho}_i(\widehat{\theta}_i) : \Gamma_i \to [0, \infty)$ be the corresponding probability density functions. Then the joint PDF $\widehat{\rho} : \Gamma \to [0, \infty)$ of $\widehat{\boldsymbol{\theta}}$ is given by:

$$
\widehat{\rho}(\widehat{\boldsymbol{\theta}}) := \prod_{i=1}^{m} \widehat{\rho}_i(\widehat{\theta}_i).
\tag{2}
$$

For reasons that will be clear later, we want to consider all of the input parameters as defined over a uniform interval. Let $\boldsymbol{\varphi} : \Gamma \to [0, 1]^m$ be a bijective map. Hence each $\widehat{\boldsymbol{\theta}} \in \Gamma$ can be associated with a unique $\boldsymbol{\theta} \in [0, 1]^m$ such that:

$$
\widehat{\boldsymbol{\theta}} = \boldsymbol{\varphi}^{-1}(\boldsymbol{\theta}).
\tag{3}
$$

Combining (2) and (3), the joint PDF ρ of $\boldsymbol{\theta}$ is given by:

$$
\rho(\boldsymbol{\theta}) = \left(\widehat{\rho} \circ \boldsymbol{\varphi}^{-1}\right)(\boldsymbol{\theta}) = \prod_{i=1}^{m} \left(\widehat{\rho}_i \circ \varphi_i^{-1}\right)(\theta_i).
\tag{4}
$$

Let $\{\boldsymbol{\theta}^j\}_{j=1}^k$ be k samples of $\boldsymbol{\theta}$. Define the *vector-valued function* $\boldsymbol{y}(\boldsymbol{\theta}) : [0, 1]^m \to \mathbb{R}^d$ to account for outcomes of interest of (1), whose i-th entry is given by:

$$
y_i(\boldsymbol{\theta}) = J_i\left(\boldsymbol{x} \circ \boldsymbol{\varphi}^{-1}\right)(\boldsymbol{\theta}),
\tag{5}
$$

with J_i a bounded and continuous functional, possibly nonlinear. For example, the J_i may represent the annual disease incidence or disease-related mortality among an intervention population in a given time period. Note that $\boldsymbol{y}(\boldsymbol{\theta})$ does not depend on time t directly. One may introduce time-dependent quantities through the definition of J_i, for example two distinct J_i may describe the same output considered over different time periods.

3 INPUT-OUTPUT MODEL ORDER REDUCTION

3.1 Reduction of the output space

Given k samples of $\boldsymbol{\theta}$, we solve (1) for each $\widehat{\boldsymbol{\theta}}^j = \boldsymbol{\varphi}^{-1}(\boldsymbol{\theta}^j)$. Denote $\boldsymbol{y}^j = \boldsymbol{y}(\boldsymbol{\theta}^j)$. Let Y be the $d \times k$ *snapshot matrix*:

$$Y = \begin{bmatrix} \boldsymbol{y}^1 | & \boldsymbol{y}^2 | & ... & | \boldsymbol{y}^k \end{bmatrix}. \tag{6}$$

We can then employ some dimension-reduction technique to extract a lower-dimensional representation of Y. Let $\Psi : \mathbb{R}^d \to \mathbb{R}^p$, with $p \leq d$ be a mapping defining the chosen dimension-reduction algorithm. The reduced representation of Y is then given by $\overline{Y}$:

$$\overline{Y} = \Psi(Y) = \begin{bmatrix} \Psi(\boldsymbol{y}^1) | & \Psi(\boldsymbol{y}^2) | & ... & | & \Psi(\boldsymbol{y}^k) \end{bmatrix} \tag{7}$$

Different algorithms can be used to define Ψ, including standard Principal Component Analysis (PCA) and variants such as kernel PCA[13], among others. The algorithm discussed in the present work is not restricted to any particular technique. For simplicity, we will assume a standard PCA, which we define using the Singular Value Decomposition (SVD). In this case, we assume (6) to have rows with zero mean and normalized variance.

Taking the SVD of $Y = USV^T$, we consider only the first p singular values/vectors. Denoting the i-th column of U as U_i, the first p *left* singular vectors $U_{1:p}$ are the *principal components* of Y, and define a set of orthonormal axes in the reduced space whose directions capture the majority of the variance observed in Y. We can project Y onto the reduced space as:

$$\overline{Y} = U_{1:p}^T Y. \tag{8}$$

$\overline{Y}$ is a $p \times k$ matrix whose j-th column expresses the j-th column of Y in the coordinate system defined by $U_{1:p}$. Let us then denote by $\widetilde{\boldsymbol{y}} : [0,1]^m \to \mathbb{R}^p$ the mapping that associates to $\boldsymbol{\theta}$ an output vector in the reduced space of dimension p, and such that $\widetilde{\boldsymbol{y}}(\boldsymbol{\theta}^j) = \overline{Y}_j = U_{1:p}^T \boldsymbol{y}^j$ for $j = 1, \ldots, k$. We can then introduce a surrogate model $\widetilde{\boldsymbol{\Phi}} : [0,1]^m \to \mathbb{R}^p$ of $\boldsymbol{y}$ over the reduced output space as follows. We consider the following modal expansion of $\widetilde{\boldsymbol{y}}$:

$$\widetilde{\boldsymbol{\Phi}}(\boldsymbol{\theta}) = \sum_{i=1}^s \langle \widetilde{\boldsymbol{y}}(\boldsymbol{\theta}) | \mathcal{L}_i(\boldsymbol{\theta}) \rangle_\rho \mathcal{L}_i(\boldsymbol{\theta}), \tag{9}$$

where $\langle \cdot | \cdot \rangle_\rho$ denotes the ρ-inner product over $[0,1]^m$ and $\{\mathcal{L}_i\}_{i=1}^s$ are a set of ρ-orthogonal polynomials. This is the so-called *generalized Polynomial Chaos Expansion (gPCE)*[14] truncated over a finite number of terms. Since in the present work, we assume that θ_i are uniformly-distributed random variables, the $\mathcal{L}_i$ are Legendre polynomials[14]. In practice, the computation of gPCE coefficients is done using the Sparse Grids Matlab Kit[15;16], which requires only the reduced output data $\overline{Y}$ obtained by the PCA as described above, using a sparse-grid collocation strategy to select the $\boldsymbol{\theta}^j$. We refer the reader to[15] for more details. Note that defining $\widetilde{\boldsymbol{\Phi}}(\boldsymbol{\theta})$ in this way allows for the fast calculation of its partial derivatives in θ_i via recurrence formulas[17].

3.2   Reduction of the input space

Observe that $\widetilde{\boldsymbol{\Phi}}(\boldsymbol{\theta})$ defines a map from the m-dimensional *full input* space to the p-dimensional reduced output space. We now seek to obtain a corresponding p-dimensional input space $\widetilde{\Gamma}$. Let $\boldsymbol{e}_j$, $j = 1, 2, ..., p$ denote the p-dimensional standard basis vectors; in the coordinate system defined by $U_{1:p}$, $\boldsymbol{e}_j$ corresponds to the direction spanned by U_j.

We aim to reduce the dimension of the input space by finding directions in $\boldsymbol{\tau}_j \in \mathbb{R}^m$ such that a perturbation of $\boldsymbol{\theta}$ along $\boldsymbol{\tau}_j$ results in a perturbation of $\boldsymbol{y}(\boldsymbol{\theta})$ along U_j. Assuming such a $\boldsymbol{\tau}_j$ exists, fixing $\boldsymbol{\theta}^0$ in the interior of $[0, 1]^m$:

$$\boldsymbol{y}(\boldsymbol{\theta}^0 + \delta\boldsymbol{\tau}_j) = \boldsymbol{y}(\boldsymbol{\theta}^0) + \varepsilon U_j, \tag{10}$$

for $\varepsilon > 0$, $\delta > 0$. Observe that:

$$U_{1:p}^T \boldsymbol{y}(\boldsymbol{\theta}^0 + \delta\boldsymbol{\tau}_j) = U_{1:p}^T \boldsymbol{y}(\boldsymbol{\theta}^0) + \varepsilon U_{1:p}^T U_j \rightarrow \widetilde{\boldsymbol{y}}(\boldsymbol{\theta}^0 + \delta\boldsymbol{\tau}_j) - \widetilde{\boldsymbol{y}}(\boldsymbol{\theta}^0) \approx \varepsilon \boldsymbol{e}_j. \tag{11}$$

Replacing $\widetilde{\boldsymbol{y}}$ with its gPCE (9) and exploiting its differentiability, the limit:

$$\lim_{\delta \to 0} \frac{\widetilde{\boldsymbol{\Phi}}(\boldsymbol{\theta}^0 + \delta\boldsymbol{\tau}_j) - \widetilde{\boldsymbol{\Phi}}(\boldsymbol{\theta}^0)}{\delta} = \nabla\widetilde{\boldsymbol{\Phi}}\left(\boldsymbol{\theta}^0\right)\boldsymbol{\tau}_j \tag{12}$$

exists, which, together with (11), suggests:

$$\nabla\widetilde{\boldsymbol{\Phi}}\left(\boldsymbol{\theta}^0\right)\boldsymbol{\tau}_j \approx \boldsymbol{e}_j. \tag{13}$$

Recall that:

$$\nabla\widetilde{\boldsymbol{\Phi}}(\boldsymbol{\theta}^0) = \begin{bmatrix} \partial\widetilde{\Phi}_1/\partial\theta_1 & \partial\widetilde{\Phi}_1/\partial\theta_2 & .. & \partial\widetilde{\Phi}_1/\partial\theta_m \\ \partial\widetilde{\Phi}_2/\partial\theta_1 & \partial\widetilde{\Phi}_2/\partial\theta_2 & .. & \partial\widetilde{\Phi}_2/\partial\theta_m \\ \vdots & \vdots & \ddots & \vdots \\ \partial\widetilde{\Phi}_p/\partial\theta_1 & \partial\widetilde{\Phi}_p/\partial\theta_2 & .. & \partial\widetilde{\Phi}_p/\partial\theta_m \end{bmatrix}, \tag{14}$$

with the dependence of each $\widetilde{\Phi}_i$ on $\boldsymbol{\theta}^0$ understood. From the above, we observe that $\nabla\widetilde{\boldsymbol{\Phi}}(\boldsymbol{\theta}^0)$ is a $p \times m$ matrix, and $p \neq m$ in general. Therefore, $\nabla\widetilde{\boldsymbol{\Phi}}(\boldsymbol{\theta}^0)^{-1}$ will not typically exist, and, since in most cases we expect $m > p$, the system (13) will have infinitely many solutions in general.

As an alternative, let $U_\Phi S_\Phi V_\Phi^T$ denote the SVD factorization of $\nabla\widetilde{\boldsymbol{\Phi}}(\boldsymbol{\theta}^0)$ such that $\nabla\widetilde{\boldsymbol{\Phi}}(\boldsymbol{\theta}^0) = U_\Phi S_\Phi V_\Phi^T$ and define:

$$\nabla\widetilde{\boldsymbol{\Phi}}(\boldsymbol{\theta}^0)^\dagger = V_\Phi \begin{pmatrix} (S_{\Phi, 1:p \times 1:p})^{-1} \\ 0 \end{pmatrix} U_\Phi^T. \tag{15}$$

The expression (15) is the *Moore-Penrose pseudoinverse* of $\nabla\widetilde{\boldsymbol{\Phi}}(\boldsymbol{\theta})$ [18].

Definition: Reduced-order input space.   We define the p-dimensional *reduced-order input space* $\widetilde{\Gamma}_p$ as:

$$\widetilde{\Gamma}_p = \text{span}\{\boldsymbol{\tau}_j\}_{j=1}^p, \quad \text{where:} \quad \boldsymbol{\tau}_j = \nabla\widetilde{\boldsymbol{\Phi}}(\boldsymbol{\theta}^0)^\dagger \boldsymbol{e}_j. \tag{16}$$

3.2.1 Moore-Penrose pseudoinverse and the need for a transformed parameter space

Let $Z_j = \{\boldsymbol{z} \in \mathbb{R}^m \mid \nabla \widetilde{\Phi}(\boldsymbol{\theta}^0)\boldsymbol{z} = \boldsymbol{e}_j\}$. An important consequence of defining the reduced-input space as (16) is that[18]:

$$\|\boldsymbol{\tau}_j\|_2 \leq \|\boldsymbol{z}\|_2 \quad \forall \boldsymbol{z} \in Z_j. \tag{17}$$

Although the system (13) has infinitely many solutions (the set Z_j) in general, $\boldsymbol{\tau}_j$ gives the *minimum-norm solution*. This ensures the reduced-order input directions $\boldsymbol{\tau}_j$ are well-defined.

The use of the Moore-Penrose pseudoinverse is the motivation for considering the transformed parameter vector $\boldsymbol{\theta} = \boldsymbol{\varphi}(\widehat{\boldsymbol{\theta}})$, rather than $\widehat{\boldsymbol{\theta}}$ directly. Recall that the $\widehat{\theta}_i$ refer to physical parameters, defined over intervals Γ_i, which have different numeric ranges in general. Thus, the notion of a 'minimum-norm' solution in Γ may not be physically meaningful. By considering the transformed parameter vector $\boldsymbol{\theta}$, we ensure that each θ_i is defined over a uniform interval, and consequently, that the uniqueness-condition (17) makes sense.

4 NUMERICAL STUDY OF HIV-PREVENTION INTERVENTIONS

To demonstrate our introduced methodology, we consider a series of HIV prevention interventions intended to prevent HIV acquisition and transmission consisting of expanding HIV testing, increasing the uptake and adherence of pre-exposure prophylaxis (PrEP) among persons without HIV, and increasing the uptake and adherence of antiretroviral therapy (ART) among persons with HIV (PWH). Since the burden of both HIV, as well associated social and structural inequities, vary by population, prevention efforts may also need to vary by population. We consider several distinct high priority transmission groups: men who have sex with men (MSM), heterosexual females (HETF), heterosexual males (HETM), and persons who inject drugs (PWID). We use HOPE, a compartmental model of HIV transmission in the United States, to illustrate the methods' ability to reduce the dimension of the input parameter space. Furthermore, we show the reduced input space can be used to rapidly develop and assess different intervention strategies.

4.1 HOPE Model

HOPE is a dynamic compartmental model that simulates the sexually active U.S. population. HOPE model includes 273 compartments corresponding to movement from the population without HIV to persons with HIV, across various disease progression states and stages of HIV care. Each model compartment is stratified by age, sex, race/ethnicity, risk level, transmission group and circumcision status. For inputs for which data were limited or uncertain, we obtained input distributions (assumed uniform) through a calibration process by varying those values within expected bounds. These inputs were calibrated so that key model outputs matched the most recent data available at the time of the analysis. Further details can be found in the technical report of the latest publication[19].

4.2 Reduced-order inputs/outputs and interpretation

We consider as model inputs $\boldsymbol{\theta}$, consisting of $m = 10$ intervention strategies corresponding to increased levels of HIV-related prevention and care services reaching different transmission groups. Our model outputs $\boldsymbol{y}$ consist of $d = 6$ HIV-related outcomes. Together, $\boldsymbol{\theta}$ and $\boldsymbol{y}$ are defined as:

$$\boldsymbol{\theta} = \begin{pmatrix} \text{PrEP (MSM)} \\ \text{PrEP (HETF)} \\ \text{PrEP (HETM)} \\ \text{PrEP (PWID)} \\ \text{ART (MSM)} \\ \text{ART (HET)} \\ \text{ART (PWID)} \\ \text{Testing (MSM)} \\ \text{Testing (HET)} \\ \text{Testing (PWID)} \end{pmatrix}, \ \boldsymbol{y} = \begin{pmatrix} \text{New infections (MSM)} \\ \text{New infections (HETF)} \\ \text{New infections (HETM)} \\ \text{New infections (PWID)} \\ \text{New infections (total)} \\ \text{Total HIV-related spending (billions of 2022 USD)} \end{pmatrix}. \tag{18}$$

The intervention strategies are defined as increases compared to a baseline delivery level by transmission group, beginning in the model in year 2023 and simulated through year-end 2030. The outputs refer to the cumulative totals of each outcome quantity over 2023-30. At baseline intervention levels, over the period 2023-30, our key outputs of interest are simulated as:

$$\boldsymbol{y}_{\text{Baseline}} = \begin{pmatrix} 200267 & \text{New infections (MSM)} \\ 28590 & \text{New infections (HETF)} \\ 15689 & \text{New infections (HETM)} \\ 16167 & \text{New infections (PWID)} \\ 261000 & \text{New infections (total)} \\ 367 & \text{Total HIV-related spending (billions of 2022 USD)} \end{pmatrix}. \tag{19}$$

Using the Sparse Grids Matlab Kit [15;16], we generate a total of 221 points $\boldsymbol{\theta}^j$, corresponding to different levels of the ten interventions (18). We collect each corresponding $\boldsymbol{y}^j$ into the 6×221 matrix Y, see (6). From the SVD of Y, we found that considering only the first $p = 4$ singular vectors/values retained over 95% of the variance in Y. We then construct the map $\widetilde{\boldsymbol{\Phi}}$ from the intervention space to the reduced, 4-dimensional output space, following (9).

The *principal component loadings* for each of the 4 reduced-order output directions $U_{1:4}$ are plotted in Fig. 1. These are the entries of each vector U_i, weighted by the corresponding singular value $S_{i,i}$; each entry indicates how much information of each physical variable is contained in the principal component. The first principal component is strongly associated with incidence in all groups except PWID. The second component is most associated with incidence among the HETM and PWID groups. The third principal component only reflects variations in spending, while the fourth component is predominantly associated with incidence among PWID.

We then reduce the dimension of the input space from 10 to 4, following (16) (note that $\boldsymbol{\theta}^0 = \frac{1}{2}\vec{\mathbf{1}}$ for convenience). The entries of each reduced-order input direction $\boldsymbol{\tau}_j$ are plotted in Fig. 2. The first reduced-order input direction $\boldsymbol{\tau}_1$ is dominated by PrEP and ART for MSM. The second and fourth reduced-order input directions $\boldsymbol{\tau}_{2,4}$ are primarily driven by ART among

HETM and PWID ($\boldsymbol{\tau}_2$) and HETF and PWID ($\boldsymbol{\tau}_4$), respectively. Finally, the third reduced-order input direction $\boldsymbol{\tau}_3$ corresponds primarily to PrEP use among heterosexuals.

To confirm that the reduced-order input directions calculated by (16) are correct, we compute $\boldsymbol{y}_{\boldsymbol{\tau}_j} = \boldsymbol{y}\left((\boldsymbol{x} \circ \varphi^{-1})\left(\boldsymbol{\theta}^0 + \boldsymbol{\tau}_j\right)\right)$ for each $\boldsymbol{\tau}_j$. If the calculated $\boldsymbol{\tau}_j$ are correct, we expect that (denoting $\boldsymbol{y}_{\boldsymbol{\theta}^0} = \boldsymbol{y}\left((\boldsymbol{x} \circ \varphi^{-1})\left(\boldsymbol{\theta}^0\right)\right)$):

$$U_{1:4}^T\left(\boldsymbol{y}_{\boldsymbol{\tau}_j} - \boldsymbol{y}_{\boldsymbol{\theta}^0}\right) \approx \boldsymbol{e}_j. \tag{20}$$

Since we used a truncated SVD and a PCE, we do not expect the relation (20) to hold exactly, but approximately. In fact, we find:

$$
\begin{aligned}
U_{1:4}^T\left(\boldsymbol{y}_{\boldsymbol{\tau}_1} - \boldsymbol{y}_{\boldsymbol{\theta}^0}\right) &= \begin{pmatrix} 1.0073 \\ 0.0048 \\ 0.008 \\ 0.0024 \end{pmatrix}, \ U_{1:4}^T\left(\boldsymbol{y}_{\boldsymbol{\tau}_2} - \boldsymbol{y}_{\boldsymbol{\theta}^0}\right) = \begin{pmatrix} 0.0031 \\ 1.0146 \\ 0.0022 \\ 0.0004 \end{pmatrix}, \\
U_{1:4}^T\left(\boldsymbol{y}_{\boldsymbol{\tau}_3} - \boldsymbol{y}_{\boldsymbol{\theta}^0}\right) &= \begin{pmatrix} 0.0007 \\ 0.0011 \\ 1.0088 \\ 0.0002 \end{pmatrix}, \ U_{1:4}^T\left(\boldsymbol{y}_{\boldsymbol{\tau}_4} - \boldsymbol{y}_{\boldsymbol{\theta}^0}\right) = \begin{pmatrix} -0.0001 \\ 0.0176 \\ 0.0035 \\ 1.0076 \end{pmatrix}.
\end{aligned} \tag{21}
$$

4.3 Direct use in intervention planning

We now apply the method to a hypothetical problem of prioritizing intervention strategies. We wish to find combinations of intervention strategies that meet the following hypothetical output targets (targets on HIV-related outcomes, units as in (18)):

$$
\boldsymbol{y}_{\text{Target1}} = \begin{pmatrix} 150000 \\ 24000 \\ 14000 \\ 12000 \\ 200000 \\ 425 \end{pmatrix}, \boldsymbol{y}_{\text{Target2}} = \begin{pmatrix} 150000 \\ 24000 \\ 14000 \\ 12000 \\ 200000 \\ 380 \end{pmatrix}, \boldsymbol{y}_{\text{Target3}} = \begin{pmatrix} 130000 \\ 22000 \\ 12000 \\ 11000 \\ 175000 \\ 475 \end{pmatrix}, \boldsymbol{y}_{\text{Target4}} = \begin{pmatrix} 130000 \\ 22000 \\ 12000 \\ 11000 \\ 175000 \\ 425 \end{pmatrix}. \tag{22}
$$

We can identify distinct intervention combinations that achieve these targets through with $\nabla\widetilde{\boldsymbol{\Phi}}(\boldsymbol{\theta}^0)^{\dagger}$. Note $\boldsymbol{y}_{\text{Target}j}$ can be expressed approximately as a linear combination of the U_i. From (16):

$$\boldsymbol{y}_{\text{Target}j} \approx \sum_{i=1}^4 a_i U_i \rightarrow U_{1:4}^T \boldsymbol{y}_{\text{Target}j} \approx \sum_{i=1}^4 a_i \boldsymbol{e}_i \rightarrow \nabla\widetilde{\boldsymbol{\Phi}}(\boldsymbol{\theta}^0)^{\dagger} U_{1:4}^T \boldsymbol{y}_{\text{Target}j} \approx \sum_{i=1}^4 a_i \boldsymbol{\tau}_i. \tag{23}$$

Evaluating:

$$\boldsymbol{y}_{\text{Eval } j} = \boldsymbol{y}\left((\boldsymbol{x} \circ \varphi^{-1})\left(\boldsymbol{\theta}^0 + \nabla\widetilde{\boldsymbol{\Phi}}(\boldsymbol{\theta}^0)^{\dagger} U_{1:4}^T \boldsymbol{y}_{\text{Target } j}\right)\right) = \boldsymbol{y}\left((\boldsymbol{x} \circ \varphi^{-1})\left(\boldsymbol{\theta}^0 + \Sigma_{i=1}^4 a_i \boldsymbol{\tau}_i\right)\right) \tag{24}$$
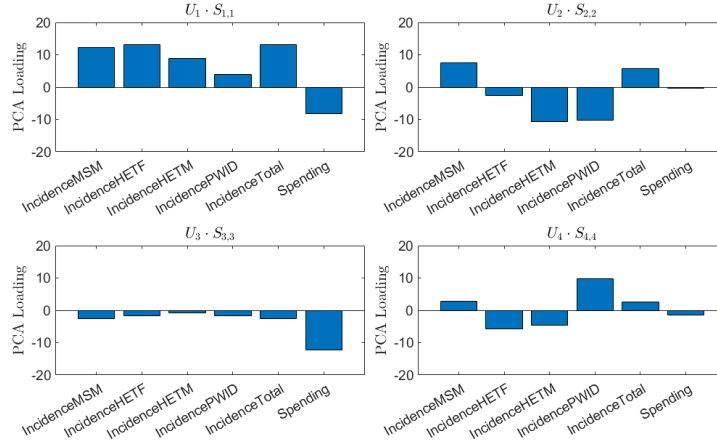
Figure 1: The four identified reduced-order outputs for the HIV prevention test case.

for each $\boldsymbol{y}_{\text{Target } j}$ (22) gives:

$$\boldsymbol{y}_{\text{Eval 1}} = \begin{pmatrix} 147039 \\ 24733 \\ 13447 \\ 12037 \\ 197257 \\ 425 \end{pmatrix}, \boldsymbol{y}_{\text{Eval 2}} = \begin{pmatrix} 145933 \\ 25105 \\ 13312 \\ 11918 \\ 196629 \\ 386 \end{pmatrix}, \boldsymbol{y}_{\text{Eval 3}} = \begin{pmatrix} 133822 \\ 22062 \\ 11801 \\ 10659 \\ 178343 \\ 472 \end{pmatrix}, \boldsymbol{y}_{\text{Eval 4}} = \begin{pmatrix} 128827 \\ 21780 \\ 11682 \\ 10454 \\ 172742 \\ 425 \end{pmatrix}. \quad (25)$$

All intervention strategy combinations come close to achieving the targets (22), and are plotted in Fig. 3. Note that $\boldsymbol{y}_{\text{Target 1, Target 2}}$ have the same HIV outcome targets, but different HIV-related spending targets. For $\boldsymbol{y}_{\text{Target 1}}$, PrEP is significantly increased from baseline levels. However, to reach the same incidence targets with a lower spending target $\boldsymbol{y}_{\text{Target 2}}$, de-emphasizes PrEP and allocates more resources toward ART. Similarly, $\boldsymbol{y}_{\text{Target 3, Target 4}}$ have identical HIV outcome targets and different spending targets. In $\boldsymbol{y}_{\text{Target 3}}$, PrEP is heavily emphasized; $\boldsymbol{y}_{\text{Target 4}}$ allocates comparatively fewer resources towards PrEP, and more towards ART. Nevertheless, PrEP remains an important component of $\boldsymbol{y}_{\text{Target 4}}$. We note that these results are qualitatively similar to those found in other allocation analyses [20;21]. We emphasize that these are hypothetical intervention targets used for illustrative purposes to demonstrate the introduced methodology, and should not be interpreted as specific program or policy recommendations.

This example demonstrates the potential utility of input/output reduced-order modeling for prioritizing intervention strategies. Different intervention combinations were identified with minimal computational cost, and did not require any sort of iterative and/or optimization procedure. Despite the low-dimension of the reduced input space, the intervention combinations identified by the input-output reduced-order model are able to account for important considerations, including identifying significantly different intervention strategies in response to changes in target spending levels.
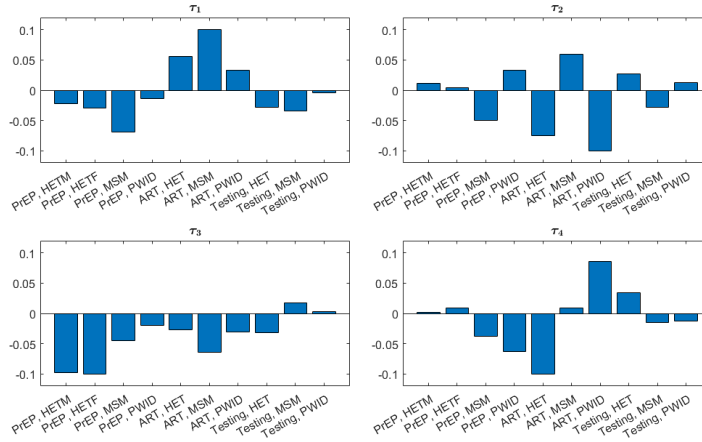
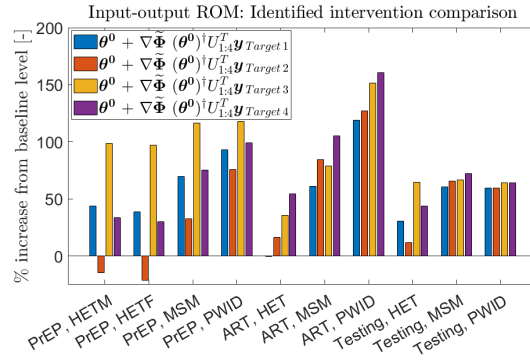**Figure 2**: Reduced-order inputs corresponding to each reduced-order output.

Figure 3: Distinct interventions identified through input-output ROM procedure.

5   CONCLUSIONS

We have introduced a novel method for reducing the dimension of the input space for models with high input dimensionality. The technique works through sampling the model over a set of sparse-grid points, aggregating the model outputs of interest for each sampled point, and performing a dimension reduction on the aggregated outputs. We then build a surrogate model over the reduced output space, which maps the input parameters to the reduced outputs. Subsequently, we reduce the dimension of the input space via the construction of a mapping which associates each axis in the reduced output space with a corresponding vector in the input space. We used a national-level model of HIV transmission to demonstrate the utility of the proposed method, and found that the approach could be applied directly to construct intervention combination strategies based on hypothetical output targets, including targets related to both HIV outcomes and HIV-related spending. Furthermore, despite its reduced dimension, sophisticated dynamics, including significant changes to intervention combination strategies in response to changes in target spending levels, were captured by the reduced-order input-output space.

The present work is intended as a proof-of-concept, and much future work remains to develop the introduced methodology. For simplicity, we used a standard PCA to reduce the order of the output space. However, other dimension-reduction techniques exist, and may be more appropriate in certain situations. In principle, the proposed method should work with any such technique, however, further investigation is needed. To emphasize the intuition behind the approach, the method was not introduced in a mathematically rigorous manner; a more rigorous treatment, establishing existence conditions and other important properties, is necessary. Furthermore, while we emphasized public health applications herein, due to the ubiquity of models with high-dimensional input spaces in the field, the introduced technique is general in nature, and can be applied to any parameterized model satisfying very basic continuity conditions. Using the proposed method in other application areas is also an important next step.

References

[1] Z. Luo and G. Chen. *Proper orthogonal decomposition methods for partial differential equations*. Academic Press, 2018.

[2] A Quarteroni, G Rozza, et al. *Reduced order methods for modeling and computational reduction*, volume 9. Springer, 2014.

[3] J.D. Jakeman, S. Friedman, M.S. Eldred, L. Tamellini, A.A. Gorodetsky, and D. Allaire. Adaptive experimental design for multi-fidelity surrogate modeling of multi-disciplinary systems. *Int J Num Methods Eng*, 123(12):2760–2790, 2022.

[4] P.J. Baddoo, B. Herrmann, B.J. McKeon, J. Nathan Kutz, and S.L. Brunton. Physics-informed dynamic mode decomposition. *P Roy Soc A-Math Phy*, 479(2271):20220576, 2023.

[5] M.W. Hess, A. Quaini, and G. Rozza. A data-driven surrogate modeling approach for time-dependent incompressible Navier-Stokes equations with dynamic mode decomposition and manifold interpolation. *Adv Comput Math*, 49(2):22, 2023.

[6] A. Viguerie, G.F. Barros, M. Grave, A. Reali, and A.L.G.A. Coutinho. Coupled and uncoupled dynamic mode decomposition in multi-compartmental systems with applications to epidemiological and additive manufacturing problems. *Comput Method Appl M*, 391:114600, 2022.

[7] F. Andreuzzi, N. Demo, and G. Rozza. A dynamic mode decomposition extension for the forecasting of parametric dynamical systems. *SIAM J Appl Dyn Syst*, 22(3):2432–2458, 2023.

[8] J. Kim, S. Yi, and Z. Wang. Dimensionality reduction can be used as a surrogate model for high-dimensional forward uncertainty quantification. *arXiv preprint arXiv:2402.04582*, 2024.

[9] C. Gadd, W. Xing, M.M. Nezhad, and A.A. Shah. A surrogate modelling approach based on nonlinear dimension reduction for uncertainty quantification in groundwater flow models. *Transport Porous Med*, 126:39–77, 2019.

[10] X. Ma and N. Zabaras. Kernel principal component analysis for stochastic input model generation. *J Comput Phys*, 230(19):7311–7331, 2011.

[11] P. G. Constantine. *Active subspaces: Emerging ideas for dimension reduction in parameter studies*. SIAM, 2015.

[12] F. Romor, M. Tezzele, A. Lario, and G. Rozza. Kernel-based active subspaces with application to computational fluid dynamics parametric problems using the discontinuous Galerkin method. *Int J Numer Meth Eng*, 123(23):6000–6027, 2022.

[13] B. Schölkopf, A. Smola, and K. Müller. Kernel principal component analysis. In *International conference on artificial neural networks*, pages 583–588. Springer, 1997.

[14] B. Sudret. Global sensitivity analysis using polynomial chaos expansions. *Reliab Eng Syst Safe*, 93(7):964–979, 2008.

[15] C. Piazzola and L. Tamellini. Algorithm 1040: The Sparse Grids Matlab Kit - a Matlab implementation of sparse grids for high-dimensional function approximation and uncertainty quantification. *ACM T Math Software*, 50(1), 2024.

[16] C. Piazzola and L. Tamellini. The Sparse Grids Matlab Kit - v. 23-5 Robert, 2023. https://github.com/lorenzo-tamellini/sparse-grids-matlab-kit.

[17] T.M. Dunster. Legendre and related functions, 2010.

[18] G.H. Golub and C. F. Van Loan. *Matrix computations*. JHU press, 2013.

[19] A. Viguerie, E.U. Jacobson, K.A. Hicks, L. Bates, J. Carrico, A. Honeycutt, C. Lyles, and P.G. Farnham. Assessing the impact of COVID-19 on HIV Outcomes in the United States: A modeling study. *Sex Transim Dis*, pages 10–1097, 2023.

[20] S.L. Sansom, K.A. Hicks, J. Carrico, E.U. Jacobson, R.K. Shrestha, T.A. Green, and D.W. Purcell. Optimal allocation of societal HIV prevention resources to reduce HIV incidence in the United States. *Am J Public Health*, 111(1):150–158, 2021.

[21] D. Hamilton, K. Hoover, D. Smith, K. Delaney, L. Wang, J. Li, T. Hoyte, S. Jenness, and S. Goodreau. Achieving the "Ending the HIV Epidemic in the US" incidence reduction goals among at-risk populations in the South. *BMC Public Health*, 23(1):716, 2023.