Modeling disease transmission near eradication: An equation free approach

Matthew O. Williams [a,*], Joshua L. Proctor [b], J. Nathan Kutz [c]

[a] *Program in Applied and Computational Mathematics, Princeton University, NJ 08544, USA*
[b] *Institute for Disease Modeling, Intellectual Ventures Laboratory, Bellevue, WA 98005, USA*
[c] *Department of Applied Mathematics, University of Washington, WA 98195, USA*

HIGHLIGHTS

- The Equation Free (EF) framework is applied to a disease eradication problem.
- An effective set of coarse variables and their related operators is proposed.
- Good agreement between the EF and full models is observed for many parameter values.
- Using the EF approach results in a factor of two speedup over the full model.

ARTICLE INFO

ABSTRACT

Although disease transmission in the near eradication regime is inherently stochastic, deterministic quantities such as the probability of eradication are of interest to policy makers and researchers. Rather than running large ensembles of discrete stochastic simulations over long intervals in time to compute these deterministic quantities, we create a data-driven and deterministic "coarse" model for them using the Equation Free (EF) framework. In lieu of deriving an explicit coarse model, the EF framework approximates any needed information, such as coarse time derivatives, by running short computational experiments. However, the choice of the coarse variables (i.e., the state of the coarse system) is critical if the resulting model is to be accurate. In this manuscript, we propose a set of coarse variables that result in an accurate model in the endemic and near eradication regimes, and demonstrate this on a compartmental model representing the spread of *Poliomyelitis*. When combined with adaptive time-stepping coarse projective integrators, this approach can yield over a factor of two speedup compared to direct simulation, and due to its lower dimensionality, could be beneficial when conducting systems level tasks such as designing eradication or monitoring campaigns.

1. Introduction

The eradication of diseases has long been a focus of the global health community and researchers in epidemiology and other related fields. Although smallpox has been eradicated [1,2], this achievement required substantial financial support, worldwide coordination, and implementation time. The Global Polio Eradication Initiative (GPEI) is a modern-day eradication campaign focused on eliminating *Poliomyelitis* (polio); since the initiative's creation in 1988, the worldwide incidence of wild poliovirus (WPV) has decreased by 99% [3], but it has not yet been eradicated. In 2012,

transmission of the WPV serotype 1 and 3 occurred in Afghanistan, Nigeria, and Pakistan [3–5], and efforts toward the eradication of the disease are currently ongoing [51].

The mathematical modeling of disease transmission and campaign implementation can play a supporting role in the final thrust toward eradication [58], and has become more important than ever due to the rarity of symptomatic cases of *Poliomyelitis*. Among their many other uses, mathematical models have been used to determine policies for the efficient distribution of a limited quantity of vaccine [6], to reveal the factors that can jeopardize an eradication campaign [7], and to quantify the impact that uncertainty in the model parameters will have on disease transmission as well as the optimal response [8–10]. As such, highly accurate models that capture a wide array of behaviors and exploit all data measurements taken in the field are of particular interest. However,

the concomitant increase in the dimensionality of these models makes detailed mathematical analysis computationally challenging. More tractable "coarse" models, which have a reduced set of possible dynamics, may be contentious due to various modeling adherents, competing modeling viewpoints, and philosophical paradigms. The benefit of coarse models is their potential to: (1) provide an enhanced level of understanding about the nature of the underlying system [11–15] such as identifying which reactions occur "slowly", and (2) their (generally) lower computational costs [16–19]. Our focus in this manuscript is on the latter task: we seek to use coarse models as a tool to reduce the computational cost of system evolution.

When the underlying governing equations (e.g., the reaction pathways and rates of a discrete stochastic system) and expected behaviors are known, then a number of highly efficient "coarse graining" techniques are already available. Methods such as Computational Singular Perturbation (CSP) [11,18] generate a lower order model by identifying a reduced set of reactants, which is an approach taken by many techniques [20–22]. Another widely used set of methods are based on the master equation associated with a given discrete stochastic system, and are valid even when the eradication of a population could occur [14,13,15,23,24,68]. These techniques can produce analytical coarse models that capture the transition probability between a set of coarse states (e.g., states where eradication has occurred, and states where it has not), and also allow meaningful statistical quantities, such as the mean time until eradication, to be computed.

Another category of model reduction or coarse graining methods can be referred to as "data driven" methods, which seek to identify a lower dimensional description of a dynamical behavior without complete knowledge of the underlying dynamical system. Data driven methods include but are not limited to the Proper Orthogonal Decomposition [25], which is also frequently referred to as Principal Component Analysis (PCA), Dynamic Mode Decomposition (DMD) [26,27,66], which has its roots in Koopman spectral analysis, and the Eigensystem Realization Algorithm (ERA) [28]. These methods use "snapshots" of the system state to generate a reduced order approximation of the underlying dynamical system. In many applications, the accuracy of these methods cannot be guaranteed, and therefore, the resulting reduced order model must be validated against a pre-existing set of experimental or numerical data (see Ref. [29] for an exception). Nonetheless, these methods are commonly used for model reduction when the evolution operator is a (system of) partial differential equations [26,66,65,30].

In this manuscript, we assume that we are provided with a mechanism for identifying disease free realizations, but that the underlying evolution equation (i.e., the detailed system) is otherwise a "black box". For problems with these constraints, the Equation Free (EF) framework is a widely used, data driven method for generating an implicit coarse model. In particular, our objective is to demonstrate that the EF framework can be used to produce coarse models for problems involving the eradication of a disease, and that the resulting models are accurate enough to reproduce the probability of eradication and other useful quantities such as the leading statistical moments of the marginal population distributions. Although we will illustrate the EF approach on a compartmental model for the transmission of *Poliomyelitis*, the variables and techniques we present could, with minor modification, also be applied to systems modeled by stochastic differential equations or individual models.

Regardless of the precise nature of the underlying dynamical system, the EF framework assumes that it possesses a "slow" manifold that solution trajectories (or, here, the statistics of an ensemble of solution trajectories) will approach after some initial, "fast" transient has elapsed. To exploit the existence of the slow manifold, a set of "coarse" variables that parameterize that manifold

must be defined. Accompanying these variables is a restriction operator, which maps from an ensemble of detailed states to a single coarse state, and a lifting operator that maps from a coarse state to a "compatible" set of detailed states. Algorithmic methods for selecting these variables is the focus of current research, but in many EF applications, the choice of coarse variables and these operators are user defined. The EF framework is mature and has been used in a number of contexts [62,31,32], including epidemiological modeling [33], but this is the first time it has been applied in the near eradication case. Therefore, there is no pre-existing "right" choice of coarse variables or operators to use in this regime. As a result, part of our contribution in this paper is to showcase the efficiency of several "common sense" choices of coarse variables that have been used in other contexts [34,33] for eradication problems. Furthermore, we also demonstrate that the evolution of "summary statistics", such as the probability of eradication, can be effectively accomplished by including them as part of the coarse state.

In lieu of defining an explicit evolution operator for the coarse system, the EF method uses short bursts of judiciously initialized detailed simulations to approximate time derivatives of the coarse variables. These time derivatives in conjunction with (slightly modified versions of) the integrators developed for systems of Ordinary Differential Equations (ODEs) [35–37], are then used to advance the state of the coarse system in a process called coarse projective integration (CPI). The benefit of this procedure is that the time steps taken by these integrators are on the "slow" timescale, and can be very large when compared to the timescales of the detailed system. This can result in significant computational savings as detailed simulations only need to be performed in short bursts to estimate coarse time derivatives rather than in a single, large simulation to advance the detailed state over the same time interval. In many applications of EF, a simple projective Euler method is enough to yield accurate results. Our second contribution in this paper is to demonstrate that more sophisticated coarse projective integrators should be used in this context; they produce more accurate approximations of meaningful quantities such as the probability of eradication, and are not significantly more computationally expensive than projective Euler.

The remainder of this manuscript is outlined as follows: in Section 2 we give a brief review of compartmental models as they appear in epidemiology. In Section 3, we provide an outline of the Equation Free framework and its application to compartmental models. In Section 4, we define several candidates for coarse variables, and demonstrate their efficacy. In Section 5, we do the same with coarse projective integrators using the distribution-based coarse variables defined in the previous section. Finally, in Section 6 a brief discussion and some concluding remarks are given.

2. Compartmental methods in epidemiology

The overarching goal of many numerical methods in epidemiology is to predict the future state (i.e., the spread, migration, or eradication) of a disease either in its endemic state or in the face of external forcing such as seasonal rainfall [52] or man made events such as routine or supplemental immunization activities [53,58,38,39]. This task can be accomplished via many fundamentally different types of models including individual (agent-based) models, where an ensemble of unique agents are generated and evolved, or compartmental models, where individuals are lumped into compartments that describe common subsets of the population [60]. For brevity, only compartmental models are considered here, but similar results exist for the agent-based models.

Compartmental models for disease transmission are often formulated as systems of ODEs, and can be efficiently solved for fixed