TRAJECTORY-ORIENTED OPTIMIZATION OF STOCHASTIC EPIDEMIOLOGICAL MODELS

Arindam Fadikar
Nicholson Collier
Abby Stevens
Jonathan Ozik

Argonne National Laboratory
Decision and Infrastructure Sciences
9700 South Cass Avenue
Lemont, IL, U.S.A.

Mickaël Binois

Inria centre at Université Côte d'Azur
Acumes project-team
2004 Route des Lucioles
06902, Sophia Antipolis, FRANCE

Kok Ben Toh

Northwestern University
Department of Preventive Medicine
Institute of Global Health
680 North Lake Shore Dr
Chicago, IL, U.S.A.

ABSTRACT

Epidemiological models must be calibrated to ground truth for downstream tasks such as producing forward projections or running what-if scenarios. The meaning of calibration changes in case of a stochastic model since output from such a model is generally described via an ensemble or a distribution. Each member of the ensemble is usually mapped to a random number seed (explicitly or implicitly). With the goal of finding not only the input parameter settings but also the random seeds that are consistent with the ground truth, we propose a class of Gaussian process (GP) surrogates along with an optimization strategy based on Thompson sampling. This *Trajectory Oriented Optimization* (TOO) approach produces actual trajectories close to the empirical observations instead of a set of parameter settings where only the mean simulation behavior matches with the ground truth.

1 INTRODUCTION

Stochastic epidemiological models have been shown to be a critical tool to support public health decision-making in times of crisis (see, e.g., Ozik et al. (2021)). These models are parameterized by calibrating their output trajectories against observed data, a crucial step for their use in policy-making. Calibration, however, can be computationally infeasible, particularly for complicated simulators and large populations. Rather than requiring simulations to be re-run with all possible parameter combinations, many have proposed using surrogate models for calibration, which replaces the expensive simulation with a simpler model such as a Gaussian process (GP) (Gramacy 2020; Carmassi et al. 2019; Higdon et al. 2004; Kennedy and O'Hagan 2001).

A typical GP surrogate models the *mean* behavior of a stochastic simulator. Given some initialization, repeated runs of the stochastic simulator will result in an ensemble of random realizations of output trajectories from a distribution (as opposed to a fixed response in the case of deterministic models). In the case of a noisy epidemic model, the stochasticity in the output can safely be represented by a mean and variance (assuming white noise) and a GP surrogate can be produced to predict the mean behavior of this process for each simulation input that, once calibrated, is consistent with the ground truth. In other modeling settings, however, each replicate may represent something more structured than white noise. For example, in the case of an agent-based epidemiological model, each replicate might define a different population mixing, and averaging outputs from all replicates may not correspond to a feasible mixing pattern.

In such cases, the random seed governing the stochasticity of each replicate can be taken as an additional input to the model, and the calibration procedure is searching not for the input parameters that most closely match the empirical data on average, but rather for the best *trajectory* (parameter-seed pair). This trajectory-oriented optimization (TOO) approach seeks to capture specific mixing patterns in stochastic models, enabling more empirically informed scenario modeling and forward projections. In this paper, we propose a class of trajectory-oriented GP surrogates with an associated optimization strategy. We leverage the common random number framework detailed by, e.g., Chen et al. (2012) and account for non-stationarities with a local formulation of the model (Edwards and Gramacy 2021). We develop a flexible Bayesian optimization technique tailored for this specific case using a Thompson sampling approach that can be generalized to multiple targets. The main contribution of this paper is the combined surrogate model and optimization strategy framework for trajectory like response. We apply our method to an Susceptible-Exposed-Infectious-Recovered (SEIR) epidemiological model of COVID-19 infections in Chicago, IL, U.S.A., and demonstrate its benefits over standard GP surrogates.

2   METHOD

Let us denote the stochastic computer simulation as $f : \mathbb{R}^d \to \mathbb{R}$ that maps a d dimensional input $\boldsymbol{x} \in \mathscr{X}^d \subseteq \mathbb{R}^d$ to a vector output. For stochastic models, f is not one-to-one and the output $f(\boldsymbol{x})$ can be indexed via replicate indices if repeated runs are available at each input. For the purpose of this paper and the trajectory-oriented approach, we augment the d-dimensional input $\boldsymbol{x}$ with a trajectory identifier r that can also be associated with the random seed used in the simulation. Later, this additional input dimension is used as a categorical input and is exploited in our proposed GP surrogate. In summary, an output from the stochastic computer simulation f would be characterized by the tuple $(\boldsymbol{x}, r)$. At this level of specification, f is treated as a deterministic mapping between the tuple $(\boldsymbol{x}, r)$ and its output. We also slightly abuse notation in adding the time variable t in $\boldsymbol{x}$ to denote time-indexed scalar output from f.

A standard GP (Rasmussen and Williams 2005; Gramacy 2020) prior characterized by a mean function $\mu(\cdot)$ and a covariance function $k(\cdot, \cdot | \Phi)$ on a set of output of size N, $\boldsymbol{Y} = (y_1, \ldots, y_N)^{\mathrm{T}}$ at input locations $\boldsymbol{X} = (\boldsymbol{x}_1, \ldots, \boldsymbol{x}_N)^{\mathrm{T}}$ provides a likelihood given by the multivariate Gaussian distribution. The covariance kernel $k : \mathbb{R}^{d+1} \times \mathbb{R}^{d+1} \to \mathbb{R}$ is positive definite and is often parameterized by additional parameter set Φ that controls various aspects of the input dependency. Inference on this additional unknown parameter is typically done via likelihood maximization or MCMC. A common practice in the computer experiments literature is to replace the actual simulation f by a fitted GP on N experiments for downstream tasks such as optimization, calibration or sensitivity analysis (see e.g., (Kleijnen 2018; Gramacy 2020)). When the observations $\boldsymbol{Y}$ from a computer experiment are noisy, white noise with fixed variance τ^2 is added to the covariance matrix such that $\boldsymbol{Y} \sim MVN(\mu, \boldsymbol{K}_N + \boldsymbol{\Sigma}_N)$, where $\boldsymbol{K}_N$ is the $N \times N$ matrix with ij coordinates $k(\boldsymbol{x}_i, \boldsymbol{x}_j)$ and $\boldsymbol{\Sigma}_N = \tau^2 \mathbb{I}$. Prediction at a new input $\boldsymbol{x}$ conditioning on the observations, i.e., $Y(\boldsymbol{x}) | \boldsymbol{Y}$ also follows a Gaussian distribution with mean and variance given by:

$$\mu(\boldsymbol{x}) = \boldsymbol{k}(\boldsymbol{x})^{\mathrm{T}} (\boldsymbol{K}_N + \boldsymbol{\Sigma}_N)^{-1} \boldsymbol{Y}, \;\; \text{where } \boldsymbol{k}(\boldsymbol{x}) = (k(\boldsymbol{x}, \boldsymbol{x}_1), \ldots, k(\boldsymbol{x}, \boldsymbol{x}_N))^{\mathrm{T}}; \quad (1)$$
$$\sigma^2(\boldsymbol{x}) = k(\boldsymbol{X}, \boldsymbol{X}) + \tau^2 - \boldsymbol{k}(\boldsymbol{x})^{\mathrm{T}} (\boldsymbol{K}_N + \boldsymbol{\Sigma}_N)^{-1} \boldsymbol{k}(\boldsymbol{x}),$$

Without loss of generality, we assume $\mu = 0$ for the rest of the description.

When replicates (repeated experimental runs at the same design point) are available, the above GP formulation remains valid. However, we can achieve computational gain without sacrificing statistical properties by aggregating replicates across inputs and modeling their mean with an updated covariance matrix to account for non-constant variance across inputs. This is often called the stochastic kriging model (Ankenman et al. 2010), while other options to model input-dependent noise can be found in Goldberg et al. (1997), Kersting et al. (2007), Binois et al. (2018), Fadikar et al. (2018).

In this work, however, we are explicitly interested in modeling individual *trajectories* rather than the mean behavior, a desirable task when the random seed encodes some essential structural information into the simulation. During calibration, rather than finding the parameter x that best matches observational data in expectation, we search for the optimal parameter-seed pair (x, r).

2.1 Common Random Number Gaussian Process (CRNGP)

We now define a new class of GP surrogates for stochastic computer simulations that models each replicate by exploiting the common random numbers (Chen et al. 2012; Pearce et al. 2019; Pearce et al. 2022). Referring back to the augmented input tuple (x, r) for stochastic f, we formally define f to be a mapping $\mathbb{R}^d \times \mathscr{S} \to \mathbb{R}$, where $\mathscr{S}$ is the set of random seeds. In general, $\mathscr{S}$ can be arbitrary, but the exact specification is not necessary for its use. Without loss of generality, we assume $\mathscr{S}$ to be countable. Now that f is deterministic, we fall back to the standard GP prior with a carefully drafted covariance kernel k to deal with this augmented input. We assume a two-component separable covariance structure here such that, $k((x_i, r), (x_j, r')) = k(x_i, x_j) \times k_s(r, r')$. The novelty is in the kernel k_s where we put our assumption(s) on replicate similarity while the kernel k for x can be set to a common family of covariance kernels such as Gaussian or Matérn.

Without any prior information on how replicates are related in a given computer simulation, we put a simplified assumption that the marginal covariance between responses at two random seeds r and r' is constant, $k_s(r, r') = \rho$, $0 < \rho < 1$. With this new specification, we write the computer simulation output vector $Y = (y_1^{(1)}, \ldots, y_1^{(r_1)}, \ldots, y_n^{(1)}, \ldots, y_n^{(r_n)})$ and the unique input locations $X = (x_1, \ldots, x_n)$, where, r_i denotes the number of replicates at x_i with $\sum_{i=1}^n r_i = N$. Without loss of generality, we also denote the random seeds by the natural numbers since they are treated as categorical in the GP formulation. The elements in covariance matrix K_N is expressed as:

$$k((x_i, r), (x_j, r')) = \begin{cases} k(x_i, x_j) & \text{if } r = r' \\ k(x_i, x_j) \times \rho & \text{otherwise,} \end{cases} \tag{2}$$

where ρ can be interpreted as an adjustment to the similarity between responses at x_i and x_j if different seeds are used at each x. Moreover, this specification of ρ results in a valid covariance kernel. A constant nugget τ^2 (typically a small positive number) is added to the diagonal of the matrix K_N, i.e., $\Sigma_N = \tau^2 \mathbb{I}$. With this formulation, the kriging formula in Equation (1) still holds true with a little modification of the new predicting location x which now actually means prediction of a particular replicate (x, r). Obtaining the mean behavior (ignoring the seed information) is equivalent to making a prediction at a new seed (i.e., for which no observation has been made) (Pearce et al. 2022). Figure 1 shows realizations from CRNGP for three different values of ρ where k is set to a Gaussian covariance kernel with length-scale and nugget at 0.01 and 0.001 respectively. The notable observation from this figure is the role of ρ in controlling the similarities among three replicated trajectories. As ρ increases, the trajectories tend to behave similarly at each x. The bottom-right panel shows a modified version of the proposed CRNGP (hereafter local CRNGP) where the value of ρ depends on the input. We describe this new family of models in detail below.

Figure 2 provides a glimpse of the fundamental difference between CRNGP and other standard GP approaches, we choose heteroskedastic GP (Binois et al. 2018) here for demonstration. When the random seed information is available, CRNGP is perfectly able to interpolate the response surface at the replicate level (for $x = 0$ and $x = 1$), whereas, a random seed agnostic GP simply tracks the mean behavior of the

response at each (x, T). The estimated value of ρ in CRNGP is 0.56. An optimization algorithm based on just the mean behavior may result in a trajectory that would be otherwise infeasible at any $\boldsymbol{x}$.
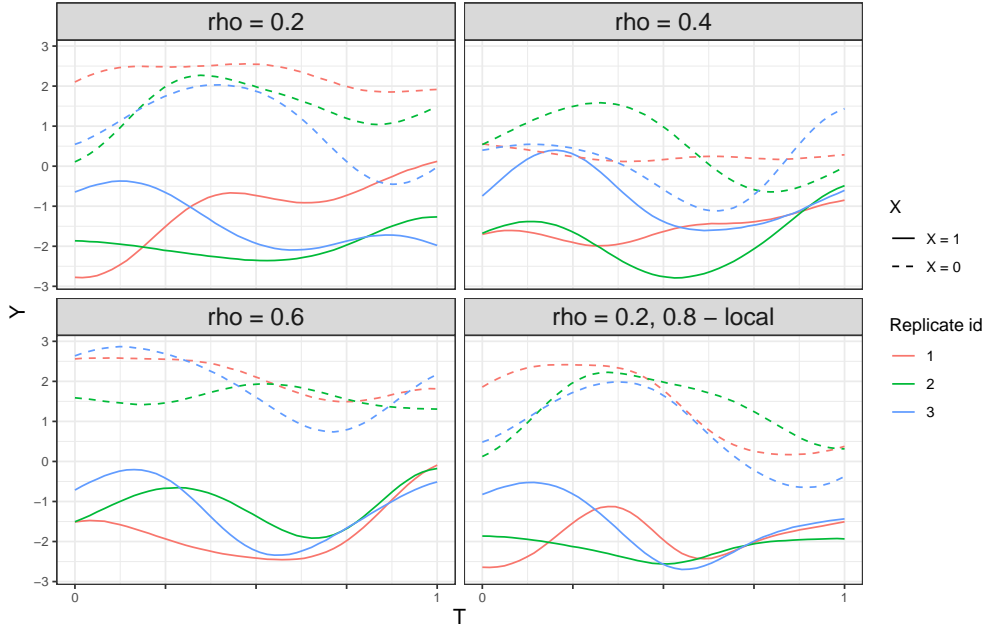

Figure 1: Realizations from CRNGP for different values of ρ (keeping the other GP hyperparameters fixed) except for the bottom-right panel which shows realization from a local CRNGP model. Each line corresponds to a random realization from the CRNGP which is vector output (trajectory) indexed by variable T (scaled to [0, 1]). In practice, T can be thought of as discrete time indices where the simulations are evaluated. Solid and dashed lines correspond to realizations at $x = 1$ and $x = 0$ respectively. The local CRNGP consists of a mixture of two CRNGPs with different values of ρs on the partition on T-axis $\{(0, 0.5), (0.5, 1)\}$.

Locally approximated CRNGP Note that the modified covariance kernel in CRNGP still maintains the stationarity and hence can be restrictive in real applications. Following common approaches for dealing with non-stationarity via local approximations (Gramacy and Apley (2015); Chipman et al. (2010)), we propose a similar recipe to extend the stationary CRNGP to support non-stationary realizations. To relax the assumption that realizations at any $\boldsymbol{x}$ has a marginal covariance of ρ that does not depend on the actual value of $\boldsymbol{x}$, we start with L partition of the input space $\mathscr{X}^d = \bigcup_{l=1}^{L} \mathscr{X}_l^d$ with $\mathscr{X}_l^d \cap \mathscr{X}_{l'}^d = \emptyset$ for any $l \neq l'$. An independent CRNGP is fitted to each subset of the data $(\boldsymbol{X}_l, \boldsymbol{Y}_l)$ with its own hyperparameters according to its membership into one of the L bins given the partition. Prediction at a new $\boldsymbol{x}^* \in \mathscr{X}_{l^*}^d$ is followed from the model in l^* partition. While this strategy provides an easy and cheap alternative to a full input-dependent GP hyperparameter estimation, a realization from this locally approximated CRNGP is not smooth and can show unrealistic behaviors at the partition boundaries. Moreover, prediction at a boundary should be influenced by all partitions sharing the boundary than just one partition. For the downstream tasks (such as optimization using Thompson sampling, discussed in the next section), a non-smooth realization on the whole input space may give rise to inaccuracies.

Instead, we define the predictive distribution at any $\boldsymbol{x}^*$ to be a weighted prediction across all L local models where the weights are proportional to the inverse of the distance between $\boldsymbol{x}^*$ and $\mathscr{X}_l^d$ for all $l = 1, \cdots, L$. Aggregating local models is also the idea behind product of experts or Bayesian committee machines, which come in many flavors, see, e.g., Rullière et al. (2018), Edwards and Gramacy (2021) and references therein. For the purpose of generating realizations from this construction, one can bypass
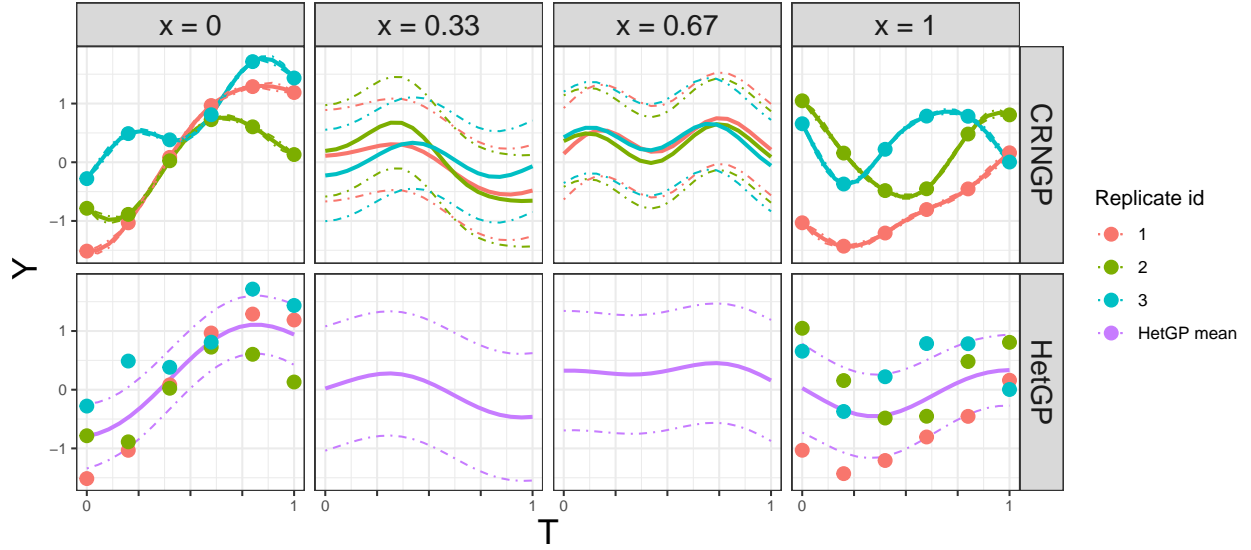
Figure 2: Comparison between fitted CRNGP and Heteroskedastic GP (HetGP) on vector-valued response at 4 different values of input *x*. Each color corresponds to a random seed value except the purple which represents the predicted mean surface from the HetGP model. The top panel shows the prediction from CRNGP which occurs at trajectory level and the bottom panel shows one mean prediction at each *x* by a heterogeneous GP.

the exact analytic form of the mean and variance of the predictive distribution at $\boldsymbol{x}^*$ and directly apply the averaging on realizations from individual models. Let $\boldsymbol{Y}^{(l)}(\boldsymbol{x}^*,s)$ be the realization from the predictive distribution of $\boldsymbol{Y}|(\boldsymbol{X}_l,\boldsymbol{Y}_l)$, then the combined realization is given by,

$$\boldsymbol{Y}(\boldsymbol{x}^*,s) = \sum_{l=1}^{L} \alpha(\boldsymbol{x}^*,\boldsymbol{X}_l)\boldsymbol{Y}^{(l)}(\boldsymbol{x}^*,s), \tag{3}$$

where the α function depicts the distance between the prediction location and the training data in each partition.

2.2 Optimization using CRNGP Surrogate

Gaussian processes are the workhorse of Bayesian optimization (Garnett 2023), also called efficient global optimization (Jones et al. 1998). The principle is to rely on the prediction and uncertainty quantification capabilities of a GP surrogate to balance between exploitation of pertinent areas where the mean is promising and exploration of unknown regions of the parameter space where the variance is large. This task is handled by the acquisition function or infill criterion, such as the expected improvement (Mockus et al. 1978) for global optimization. We refer to Garnett (2023) for a more detailed presentation of BO.

In Bayesian optimization, scalar outputs are the most typical use case, with many possible variations. Related to our TOO approach, Astudillo and Frazier (2021) solve similar trajectory problems by exploiting the composite aspect of the cost function to optimize: $g(\boldsymbol{y}(\boldsymbol{x})) = \sum_j (y_j(\boldsymbol{x}) - y_j^{obs})^2$ where y_j^{obs} are the actual observations that the simulator trajectory should match. But they solely focus on the mean behavior and not on the actual trajectories. This is also the case of Pearce et al. (2019), Pearce et al. (2022) where they exploit common random numbers with an enriched CRNGP model.

Our goal here is to identify

$$\left\{ (\boldsymbol{x}^{(1)},r^{(1)}),\dots,(\boldsymbol{x}^{(p)},r^{(p)}) \text{ s.t. } g(y(\boldsymbol{x}^{(i)},r^{(i)})) \le g(y(\boldsymbol{x},r^{(j)})) \forall \boldsymbol{x} \in \mathscr{X} \forall 1 \le i \le p, 1 \le j \le |\mathscr{S}| \right\}$$

where p is the number of trajectories needed. To do so, we look at the following optimization problem as a proxy:

$$\underset{\boldsymbol{x}\in\mathscr{X}, r\in\mathscr{S}}{\arg\min} \sum_j \left(f(\boldsymbol{x}, t_j, r) - y_j^{(obs)} \right)^2 \tag{4}$$

where the set $\mathscr{S}$ is countable (in practice, we set $\mathscr{S}$ to be a large finite set of random seeds). Compared to the more usual goal of solving $\underset{\boldsymbol{x}\in\mathscr{X}}{\arg\min} \mathbb{E}_r \left[\sum_j \left(f(\boldsymbol{x}, t_j, r) - y_j^{(obs)} \right)^2 \right] = g(\boldsymbol{y}(\boldsymbol{x}))$, we are not focusing on finding the parameters $\boldsymbol{x}$ that perform best on average but on finding good trajectories even if the mean behavior does not correspond to the data. This would allow us to provide a more diverse set of inputs than when focusing solely on the mean. Note that this is also a different approach than focusing on risk-averse (or robust) optimization.

For the trajectory-oriented setup, we propose to use Thompson sampling (TS) (Thompson 1933), a simple and versatile approach still balancing exploration and exploitation. TS amounts to sampling one realization from the GP posterior distribution $\tilde{\boldsymbol{Y}} \sim \mathcal{N} \left(\mu(\tilde{\boldsymbol{X}}, S), \sigma^2(\tilde{\boldsymbol{X}}, S) \right)$ over discrete sets $\tilde{\boldsymbol{X}}$ and S, then computing $g(\tilde{\boldsymbol{Y}})$ before selecting $(\boldsymbol{x}^*, r^*) \in \underset{(\boldsymbol{x}\in\tilde{\boldsymbol{X}}, r\in S)}{\arg\min} g(\tilde{\boldsymbol{Y}})$ as next input to evaluate. Extension to a continuous input space can be performed by using varying random discrete sets or optimizing a continuous approximation of the realization (Mutny and Krause 2018). In practice, as in Pearce et al. (2022), the set S only needs to contain seed values that have already been evaluated, plus one additional seed value accounting for the mean behavior over all unobserved seeds. This way, a new seed is evaluated only if it does better than all existing ones while promising seeds can be kept to improve over $\boldsymbol{x}$ (line 4 in the Algorithm 1). The general pseudo-code of the approach is illustrated in Algorithm 1. The batch extension of this Thompson sampling scheme is performed by sampling several posterior samples and using their respective best solutions as batch candidates (Kandasamy et al. 2018). The choice of the number of initial simulations–N_0 and total simulations–$N_{\max}$ would generally be dictated by the total computational budget available.

Algorithm 1 Pseudo-code for CRNGP based trajectory search

Require: N_0 (initial budget), $N_{\max}$ (total budget)
1: Create initial design of experiments, $(\boldsymbol{X}_n, S_n)$ of size $n = N_0$ and evaluate it $\boldsymbol{Y}_n = f(\boldsymbol{X}_n, S_n)$
2: Train the GP model on $(\boldsymbol{X}_n, S_n, \boldsymbol{Y}_n)$
3: **while** $n \leq N_{max}$ **do**
4:    $S_n \leftarrow \{S_n, (\max S_n) + 1\}$
5:    Sample from the GP posterior: $\tilde{\boldsymbol{Y}} \sim \mathcal{N} \left(\mu(\tilde{\boldsymbol{X}}, S_n), \sigma^2(\tilde{\boldsymbol{X}}, S_n) \right)$
6:    Choose $(\boldsymbol{x}_{n+1}, s_{n+1}) \in \arg\max_{\boldsymbol{x}\in\tilde{\boldsymbol{X}}, s\in S_n} g(\tilde{\boldsymbol{Y}})$
7:    Update the GP model by conditioning on $(\boldsymbol{x}_{n+1}, s_{n+1})$.
8:    $n \leftarrow n + 1$
9: **end while**

The calibration task may have $p \geq 2$ targets, corresponding to various observables matching outputs of the simulators. In such a case, there is in general no solution perfectly matching all outputs simultaneously, and the solution is defined based on compromise. That is, a solution is said to be Pareto non-dominated if there is no other one that is at least as good on all objectives and strictly better in at least one. The set of non-dominated solutions is called the Pareto set in the input space, its image in the output space being the Pareto front. In addition to fitting one GP to each output and sampling from their respective posteriors, the sequential selection procedure must be modified as well. For instance, the hypervolume contribution, e.g., (Beume et al. 2007) of each element of the GP posterior, i.e., $\left(g(\tilde{\boldsymbol{Y}}_1^{(i)}), \ldots, g(\tilde{\boldsymbol{Y}}_p^{(i)}) \right)$, is the volume added to the current Pareto formed by the $(g(y_1^{(i)}), \ldots, g(y_p^{(i)}))_{1\leq i\leq n}$.
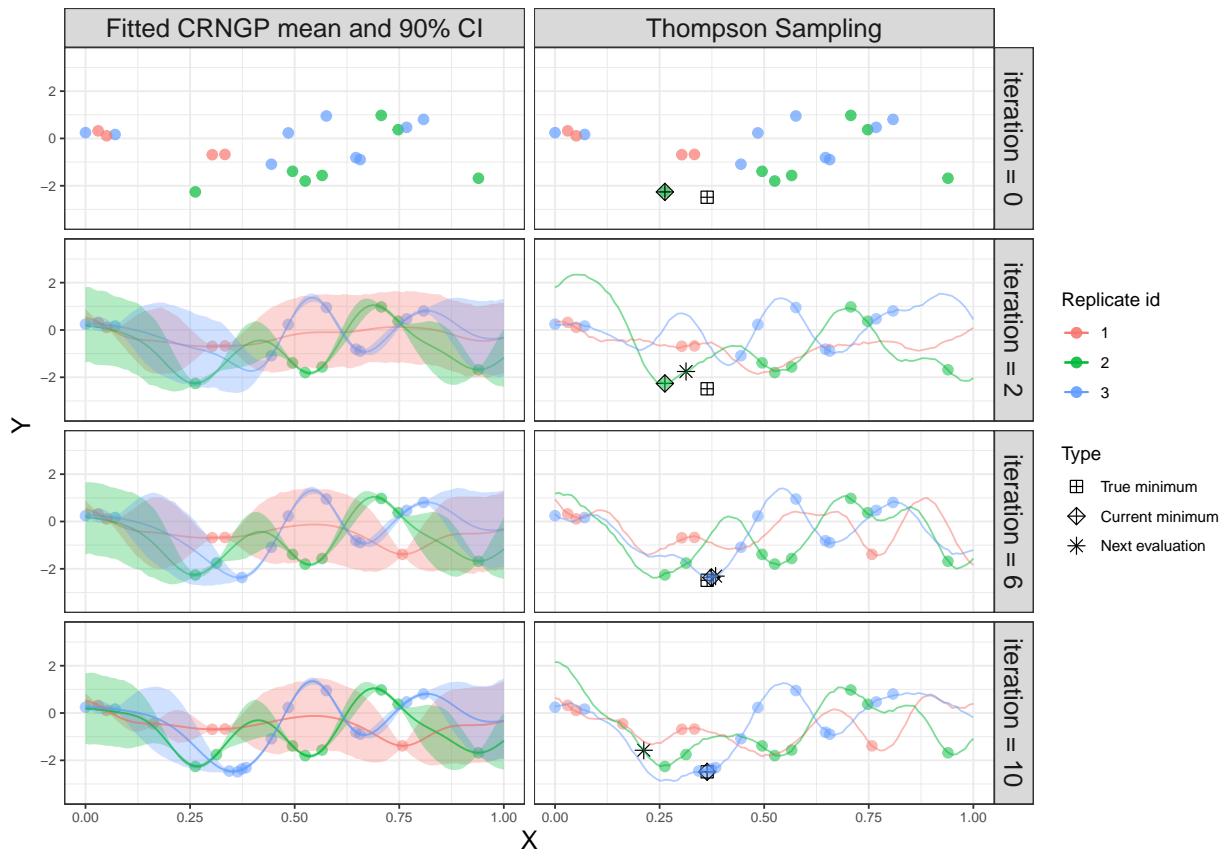
Figure 3: Thompson sampling (TS) based single objective optimization using CRNGP surrogate. Figures in the left show the predicted mean surface by CRNGP conditioning on the observations (represented by dots) and the figures in the right show random realizations from the fitted CRNGP based on which the new evaluations are obtained in the TS algorithm. Replicates are represented by different colors. Rows correspond to different iterations in the TS. In each iteration a new *x* for next evaluation is set at the arg min of a realization (right column). The global minimum is found in 11 iterations.

3 EMPIRICAL EVALUATION

3.1 Stochastic Epidemic Model

We used a published stochastic SEIR compartmental model (Runge et al. 2022) to simulate SARS-CoV-2 transmission and COVID-19 disease states. The model included multiple symptom statuses (asymptomatic, presymptomatic, mild, and severe), and severe disease outcomes (requiring hospitalization, critical illness requiring intensive care unit ICU admission, and deaths).

At the start of the epidemic, everyone is susceptible (S). The susceptible population is exposed and infected by the virus at a rate controlled by the transmission rate parameter, the number of infectious and susceptible people. After a latent period of a few days, exposed individuals (E) become infectious without symptoms: they are either completely asymptomatic (As) or presymptomatic (P). At the end of the incubation period, presymptomatic people proceed to develop either mild (Sm) or severe (Ss) symptoms. Asymptomatic or people with mild symptoms recovered (R) eventually, while those with severe symptoms would require hospital care (H) after a few days in the Ss compartment. Whether or not they eventually sought care, people who need hospital care either recover without complications or develop critical illness (C) requiring ICU care. The latter could eventually recover from the illness by first moving back to the

post-ICU hospitalization state (Hp) or die (D) from the illness. All exposed people are undetected. When a person enters the asymptomatic, presymptomatic, symptomatic mild, and symptomatic severe state, a fraction of them would be detected after a few days, which leads to isolation and reduced infectiousness. A schematic representation of the model is provided in Figure 4.

Model parameters such as incubation period, disease symptomaticity and severity, and hospital and ICU lengths of stay, are informed by literature and observed data in Chicago. The transmission rate, probability of detection, and the disease recovery rate are varied over time to better fit the daily ICU census, hospital census, and reported deaths in Chicago. A detailed description of the model and the parameters along with the source code can be found at this github repository (Runge).
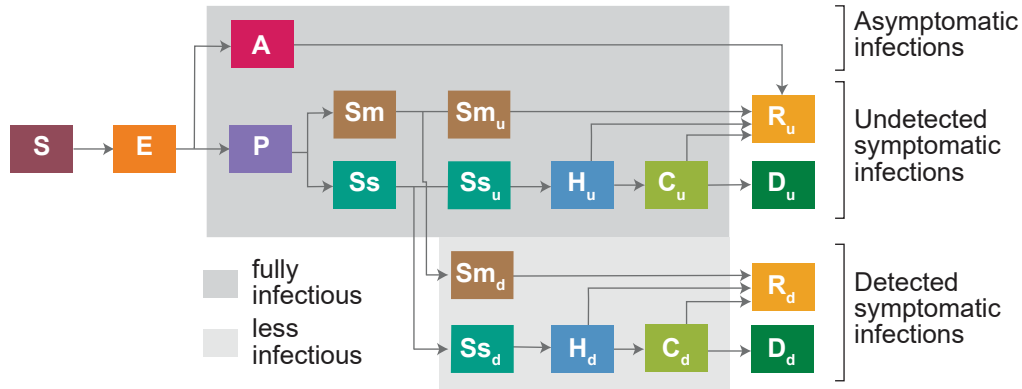

Figure 4: **SEIR model structure to simulate SARS-CoV-2 transmission.** S=Susceptible, E=Exposed, A=Asymptomatic, P=Presymptomatic, Sm=Mild symptomatic, Ss=Severe symptomatic, H=Requires hospitalization, C=Critically ill (ICU), D=Death, R=Recovered. Subscripts u and d denote if the infection is undetected or detected. Asymptomatic and presymptomatic infections may also be detected, which would result in less infectiousness.

3.2 Single and Bi-objective Optimization

To illustrate our method, we use a three parameter version of the stochastic compartmental model described above. The three input parameters are the initial disease transmissibility rate and the rate of exposed individuals becoming infectious for both asymptomatic and symptomatic cases. Other disease parameters in the model are kept fixed at values informed by literature and the observed data in Chicago. Each simulation is run for the first 100 days of the disease propagation and output is obtained at 5 discretized time values, essentially resulting in a vector-valued output of size 5. For the output of interest, we choose cumulative hospital admissions in the single objective case along with cumulative deaths in the bi-objective case. A trajectory corresponding to a random (input, random seed) pair is treated as the ground truth and we set our goal to discover all trajectories that are consistent with this ground truth. In each case we set our initial simulation budget $N_0 = 50$, which is uniformly distributed on the three dimensional input space according to a Latin hypercube design (McKay et al. 2000) and maximum budget of $N_{max} = 200$.

We also use the local version of the CRNGP surrogate, based on a fixed partition on the time component of the input. Essentially our CRNGP surrogate for vector valued output operates on five dimensional input space – three disease model parameters, discretized time indices, and the random seed. We choose only the time dimension for constituting a partition on which local surrogates would be more appropriate. This is further evident from Figure 5 and Figure 6 as to how trajectories become more diverse with time. The other aspect of a useful partition for local surrogates has to do with the number of actual simulations in each partition. A surrogate trained on only a few simulations in each partition in a high input dimension setting may defeat the goal of creating an accurate surrogate.

Figure 5 shows the 50 best trajectories that match with the ground truth (denoted by solid black dots) after the total simulation budget is exhausted. Searching for optimal input parameter values using HetGP surrogate in TS results in settings where the mean behavior of the stochastic simulation is expected to match the ground truth, whereas, the TOO approach facilitates finding not only the optimum input parameter values but also the replicates, i.e., an actual trajectory that matches with the ground truth. When calibrating to two objectives – hospitalizations and deaths, Figure 6 shows the best trajectories with respect to both quantity of interest.
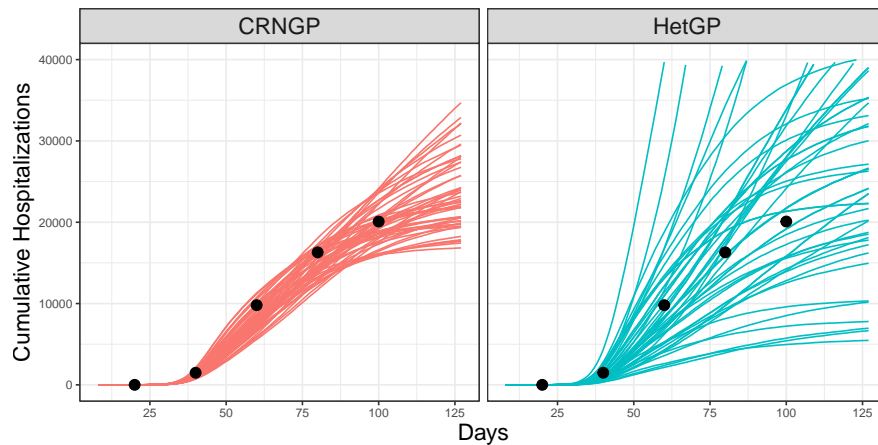

Figure 5: Best 50 trajectories of cumulative hospitalizations that are consistent with the observations denoted by black dots when using CRNGP and HetGP surrogates in Algorithm 1.
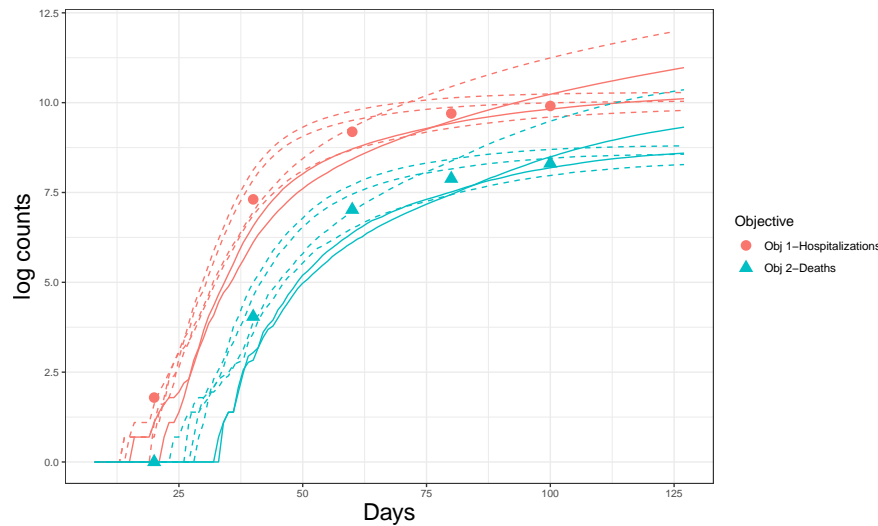

Figure 6: Best trajectories with respect to two objectives–cumulative hospitalizations and deaths simultaneously. Solid and broken lines represent CRNGP and HetGP surrogates respectively.

4   CONCLUSION

In the expensive simulation context, see, e.g., Baker et al. (2022), where simulation time is not negligible, the use of surrogate models such as Gaussian processes is required to efficiently obtain a solution of an inverse problem within a limited computational budget. For a simulation that produces a stochastic

response, often characterized by a random seed, we have proposed the CRNGP surrogate that can map each individual simulation response to a unique (input, random seed) pair. Furthermore, we have devised an optimization scheme that when paired with CRNGP can facilitate finding actual simulation outputs that are consistent with observations. The proposed method has been developed for handling vector-valued output, hence we use the term *trajectory-oriented optimization*.

In our specific epidemic application, the end goal of calibrating the simulation model is to collect simulated trajectories of disease outcomes that are consistent with empirical data trajectories. For a stochastic epidemic model (such as an ABM), each output of the simulation model may correspond to a specific instantiation of population mixing, which is often not controlled by the input parameters, but by the random seeds. Thus, in order to capture mixing patterns consistent with empirical data, one needs to find both optimal parameter settings and their corresponding random seeds. Specific mixing patterns are important to try to capture since they govern the pathways through which an epidemic will transmit, affecting how public health interventions may mitigate that transmission. Separately, in order to implement data assimilation for an ABM without rerunning the model from the very beginning of an epidemic, the state of the model, including the complete state of the agents in the model, needs to be captured at the end of each assimilating time period, so that it can be restarted from that time. The TOO approach explicitly seeks to find the best trajectories to keep in this case.

We have shown that using a CRNGP, we are not only able to find good trajectories, but we can do so within a smaller simulation budget. This fast time-to-solution aspect is particularly critical for producing actionable insights for decision makers and has been highlighted in Binois et al. (2021). Further improvement to time-to-solution is possible by considering the simulation time as an additional variable to control (see Snoek et al. (2012), Picheny and Ginsbourger (2013)) and by developing open science platforms (Collier et al. 2023) to enable these analyses. The local CRNGP surrogate can also be extended to handle input-dependent replicate similarity by considering a more complex prior model for ρ.

5 ACKNOWLEDGEMENTS

REFERENCES

Ankenman, B., B. L. Nelson, and J. Staum. 2010. "Stochastic Kriging for Simulation Metamodeling". *Operations Research* 58(2):371–382.

Astudillo, R., and P. I. Frazier. 2021. "Thinking Inside the Box: A Tutorial on Grey-Box Bayesian Optimization". In *Proceedings of the 2021 Winter Simulation Conference*, edited by S. Kim, B. Feng, K. Smith, S. Masoud, Z. Zheng, C. Szabo, and M. Loper, 1–15. Piscataway, New Jersey: Institute of Electrical and Electronics Engineers.

Baker, E., P. Barbillon, A. Fadikar, R. B. Gramacy, R. Herbei, D. Higdon, J. Huang, L. R. Johnson, P. Ma, A. Mondal et al. 2022. "Analyzing Stochastic Computer Models: A Review With Opportunities". *Statistical Science* 37(1):64–89.

Beume, N., B. Naujoks, and M. Emmerich. 2007. "SMS-EMOA: Multiobjective Selection Based on Dominated Hypervolume". *European Journal of Operational Research* 181(3):1653–1669.

Binois, M., N. Collier, and J. Ozik. 2021, 10. "A Portfolio Approach to Massively Parallel Bayesian Optimization". *arXiv:2110.09334 [math, stat]*. arXiv: 2110.09334.

Binois, M., R. B. Gramacy, and M. Ludkovski. 2018. "Practical Heteroscedastic Gaussian Process Modeling for Large Simulation Experiments". *Journal of Computational and Graphical Statistics* 27(4):808–821.

Carmassi, M., P. Barbillon, M. Chiodetti, M. Keller, and E. Parent. 2019. "Bayesian Calibration of a Numerical Code for Prediction". *Journal de la Société Française de Statistique* 160(1):1–30.

Chen, X., B. E. Ankenman, and B. L. Nelson. 2012. "The Effects of Common Random Numbers on Stochastic Kriging Metamodels". *ACM Transactions on Modeling and Computer Simulation (TOMACS)* 22(2):1–20.

Chipman, H. A., E. I. George, and R. E. McCulloch. 2010. "BART: Bayesian Additive Regression Trees". *The Annals of Applied Statistics* 4(1):266 – 298.

Collier, N., J. M. Wozniak, A. Stevens, Y. Babuji, M. Binois, A. Fadikar, A. Würth, K. Chard, and J. Ozik. 2023. "Developing Distributed High-performance Computing Capabilities of an Open Science Platform for Robust Epidemic Analysis". In *2023 IEEE International Parallel and Distributed Processing Symposium Workshops (IPDPSW)*, 868–877.

Edwards, A. M., and R. B. Gramacy. 2021. "Precision Aggregated Local Models". *Statistical Analysis and Data Mining: The ASA Data Science Journal* 14(6):676–697.

Fadikar, A., D. Higdon, J. Chen, B. Lewis, S. Venkatramanan, and M. Marathe. 2018. "Calibrating a Stochastic, Agent-Based Model Using Quantile-Based Emulation". *SIAM/ASA Journal on Uncertainty Quantification* 6(4):1685–1706.

Garnett, R. 2023. *Bayesian Optimization*. Cambridge University Press.

Goldberg, P., C. Williams, and C. Bishop. 1997. "Regression with Input-Dependent Noise: A Gaussian Process Treatment". *Advances in Neural Information Processing Systems* 10:493–499.

Gramacy, R. B. 2020. *Surrogates: Gaussian Process Modeling, Design and Optimization for the Applied Sciences*. Boca Raton, Florida: Chapman Hall/CRC.

Gramacy, R. B., and D. W. Apley. 2015. "Local Gaussian Process Approximation for Large Computer Experiments". *Journal of Computational and Graphical Statistics* 24(2):561–578.

Higdon, D., M. Kennedy, J. C. Cavendish, J. A. Cafeo, and R. D. Ryne. 2004. "Combining Field Data and Computer Simulations for Calibration and Prediction". *SIAM Journal on Scientific Computing* 26(2):448–466.

Jones, D. R., M. Schonlau, and W. J. Welch. 1998. "Efficient Global Optimization of Expensive Black-Box Functions". *Journal of Global Optimization* 13(4):455–492.

Kandasamy, K., A. Krishnamurthy, J. Schneider, and B. Póczos. 2018. "Parallelised Bayesian Optimisation via Thompson Sampling". In *International Conference on Artificial Intelligence and Statistics*, 133–142. PMLR.

Kennedy, M., and A. O'Hagan. 2001. "Bayesian Calibration of Computer Models (with discussion)". *Journal of the Royal Statistical Society, Series B* 63:425–464.

Kersting, K., C. Plagemann, P. Pfaff, and W. Burgard. 2007. "Most Likely Heteroscedastic Gaussian Process Regression". In *Proceedings of the 24th International Conference on Machine Learning*, 393–400. Association for Computing Machinery.

Kleijnen, J. P. 2018. *Design and Analysis of Simulation Experiments*. Springer.

McKay, M. D., R. J. Beckman, and W. J. Conover. 2000. "A Comparison of Three Methods for Selecting Values of Input Variables in the Analysis of Output from a Computer Code". *Technometrics* 42(1):55–61.

Mockus, J., V. Tiešis, and A. Žilinskas. 1978. "The Application of Bayesian Methods for Seeking the Extremum". *Towards Global Optimization* 2(117-129):2.

Mutny, M., and A. Krause. 2018. "Efficient High Dimensional Bayesian Optimization with Additivity and Quadrature Fourier Features". In *Advances in Neural Information Processing Systems 31*, edited by S. Bengio, H. Wallach, H. Larochelle, K. Grauman, N. Cesa-Bianchi, and R. Garnett, 9005–9016. Curran Associates, Inc.

Ozik, J., J. M. Wozniak, N. Collier, C. M. Macal, and M. Binois. 2021. "A Population Data-driven Workflow for COVID-19 Modeling and Learning". *The International Journal of High Performance Computing Applications* 35(5):483–499.

Pearce, M., M. Poloczek, and J. Branke. 2019. "Bayesian Simulation Optimization with Common Random Numbers". In *2019 Winter Simulation Conference*, 3492–3503. IEEE.

Pearce, M. A. L., M. Poloczek, and J. Branke. 2022. "Bayesian Optimization Allowing for Common Random Numbers". *Operations Research* 70(6):3457–3472.

Picheny, V., and D. Ginsbourger. 2013. "A Nonstationary Space-time Gaussian Process Model for Partially Converged Simulations". *SIAM/ASA Journal on Uncertainty Quantification* 1(1):57–78.

Rasmussen, C. E., and C. K. I. Williams. 2005. *Gaussian Processes for Machine Learning*. Cambridge, MA: MIT Press.

Rullière, D., N. Durrande, F. Bachoc, and C. Chevalier. 2018. "Nested Kriging Predictions for Datasets with a Large Number of Observations". *Statistics and Computing* 28:849–867.

Runge, M., R. A. Richardson, P. A. Clay, A. Bell, T. M. Holden, M. Singam, N. Tsuboyama, P. Arevalo, J. Fornoff, S. Patrick et al. 2022. "Modeling Robust COVID-19 Intensive Care Unit Occupancy Thresholds for Imposing Mitigation to Prevent Exceeding Capacities". *PLOS Global Public Health* 2(5):e0000308.

Runge, Manuela et. al. Covid-Chicago. https://github.com/numalariamodeling/covid-chicago, accessed 8th July.

Snoek, J., H. Larochelle, J. Snoek, and R. P. Adams. 2012. "Practical Bayesian Optimization of Machine Learning Algorithms". In *Neural Information Processing Systems*.

Thompson, W. R. 1933. "On the Likelihood That One Unknown Probability Exceeds Another in View of the Evidence of Two Samples". *Biometrika* 25(3-4):285–294.

AUTHOR BIOGRAPHIES

ARINDAM FADIKAR is an assistant computational statistician in the Decision and Infrastructure Sciences division at Argonne National Laboratory. His primary research interest is in the area of design and analysis, calibration, uncertainty quantification of computer models under input dependent noise with application in epidemiology, material science, urban traffic network systems and cosmology. His email address is afadikar@anl.gov.

MICKAËL BIONOIS is a researcher at Inria Sophia An- tipolis in the Acumes project-team. His research interests include surrogate models (especially Gaussian processes), computer experiments, Bayesian optimization (constrained, single or multi-objective), and uncertainty quantification. His email address is mickael.binois@inria.fr.

NICHOLSON COLLIER is a Senior Software Engineer at Argonne National Laboratory, and Staff Software Engineer in the Consortium for Advanced Science and Engineering at the University of Chicago. As the lead developer for the Repast project for agent-based modeling toolkits and co-developer of the Extreme-scale Model Exploration with Swift (EMEWS) framework, he develops, architects, and implements large-scale agent-based models and frameworks in a variety of application areas, and large-scale model exploration workflows across various domains, including agent-based modeling, microsimulation and machine/deep learning. His e-mail address is ncollier@anl.gov.

ABBY STEVENS is a computational data scientist at Argonne National Laboratory. Her research interests include interpretable machine learning, surrogate modeling, and data assimilation with applications in epidemiology and climate science. Her email address is stevensa@anl.gov.

KOK BEN TOH is a postdoctoral researcher at the Department of Preventive Medicine and Institute of Global Health at Northwestern University. He is broadly interested in using mathematical and statistical models to understand ecological problems and infectious disease dynamics. His current research focuses on modeling malaria and COVID-19 transmission using agent-based and stochastic compartmental models. His email address is bentoh@northwestern.edu.

JONATHAN OZIK is a Principal Computational Scientist at Argonne National Laboratory, Senior Scientist with Department of Public Health Sciences affiliation in the Consortium for Advanced Science and Engineering (CASE) at the University of Chicago, and Senior Institute Fellow in the Northwestern Argonne Institute of Science and Engineering (NAISE) at North-western University. He applies large-scale agent-based models and large-scale model exploration across research domains and computational modeling methods. He leads multiple federally funded projects in healthcare domains, and is the lead of the Repast project for agent-based modeling toolkits and the Extreme-scale Model Exploration with Swift (EMEWS) framework for large-scale model exploration on high-performance computing resources. His e-mail address is jozik@anl.gov.