



# Using data-driven agent-based models for forecasting emerging infectious diseases

Srinivasan Venkatramanan<sup>a,\*</sup>, Bryan Lewis<sup>a</sup>, Jiangzhuo Chen<sup>a</sup>, Dave Higdon<sup>b,c</sup>,  
Anil Vullikanti<sup>a,d</sup>, Madhav Marathe<sup>a,d</sup>

<sup>a</sup> Network Dynamics and Simulation Science Laboratory, Biocomplexity Institute of Virginia Tech, United States

<sup>b</sup> Social and Decision Analytics Laboratory, Biocomplexity Institute of Virginia Tech, United States

<sup>c</sup> Department of Statistics, Virginia Tech, United States

<sup>d</sup> Department of Computer Science, Virginia Tech, United States

## ARTICLE INFO

### Article history:

Received 2 July 2016

Received in revised form 30 January 2017

Accepted 17 February 2017

Available online 22 February 2017

### Keywords:

Emerging infectious diseases

Agent-based models

Simulation optimization

Bayesian calibration

Ebola

## ABSTRACT

Producing timely, well-informed and reliable forecasts for an ongoing epidemic of an emerging infectious disease is a huge challenge. Epidemiologists and policy makers have to deal with poor data quality, limited understanding of the disease dynamics, rapidly changing social environment and the uncertainty on effects of various interventions in place. Under this setting, detailed computational models provide a comprehensive framework for integrating diverse data sources into a well-defined model of disease dynamics and social behavior, potentially leading to better understanding and actions. In this paper, we describe one such agent-based model framework developed for forecasting the 2014–2015 Ebola epidemic in Liberia, and subsequently used during the Ebola forecasting challenge. We describe the various components of the model, the calibration process and summarize the forecast performance across scenarios of the challenge. We conclude by highlighting how such a data-driven approach can be refined and adapted for future epidemics, and share the lessons learned over the course of the challenge.

© 2017 The Author(s). Published by Elsevier B.V. This is an open access article under the CC BY-NC-ND license (<http://creativecommons.org/licenses/by-nc-nd/4.0/>).

## 1. Introduction

In the latter half of 20th century, there was a prevailing optimism about humanity's preparedness against infectious diseases. Nobel prize winning virologist F.M. Burnet echoed the opinion of researchers and laymen alike, when he said "*the most likely forecast about the future of infectious disease is that it will be very dull*" (Burnet and White, 1972). The confidence seemed justified given the development of antibiotics and vaccines, recent successes against polio and smallpox, etc. However, it was short-lived, with the emergence of HIV/AIDS in Africa, which has since grown into one of the greatest global health concerns of our time. Less than 20 years into the 21st century, we have been impacted by a series of infectious diseases such as SARS (2003), H1N1 (2009), Ebola (2014) and more recently Zika (2016).

With increased global connectivity, intermixing of human and animal habitats, and the looming threat of climate change, such

threats will become more common (Daszak et al., 2000). As in the case of Ebola, healthcare infrastructure, especially in densely populated and developing countries, will be severely stressed (Fauci, 2014). Generating reliable spatio-temporal forecasts, both short term and long term, will help stimulate and guide global efforts where and when required. Owing to limited understanding of the emerging infectious disease, vaccines development may not be swift, thus initial responses still need to rely on traditional epidemic control measures like isolation of cases and social distancing to curb the epidemic. The ability to evaluate and compare different behavioral interventions will be immensely valuable.

Purely data-driven approaches based on the epidemic time series and other surrogate sources from social media have gained traction in recent times (Cook et al., 2011) for epidemic forecasting. However, as Lazer et al. (2014) points out, such big data approaches often fail to account for the actual disease and social dynamics on the ground. Their performance is further marred by poor data quality (especially in developing countries) and mismatch between social media sentiment and actual prevalence, as is the case for an emerging infectious disease. This reiterates the need for a model based approach to epidemic forecasting.

Standard compartmental models in epidemiology (Anderson et al., 1992) help produce early forecasts for an ongoing epidemic

\* Corresponding author.

E-mail addresses: [vsriniv@vbi.vt.edu](mailto:vsriniv@vbi.vt.edu) (S. Venkatramanan), [blewis@vbi.vt.edu](mailto:blewis@vbi.vt.edu) (B. Lewis), [chenj@vbi.vt.edu](mailto:chenj@vbi.vt.edu) (J. Chen), [dhigdon@vbi.vt.edu](mailto:dhigdon@vbi.vt.edu) (D. Higdon), [akumar@vbi.vt.edu](mailto:akumar@vbi.vt.edu) (A. Vullikanti), [mmarathe@vbi.vt.edu](mailto:mmarathe@vbi.vt.edu) (M. Marathe).

thanks to their short setup and running time. However, they do not capture the spatial and social heterogeneity in the real world epidemic, and lack the resolution for testing individual level behavioral interventions. Further, they cannot take complete advantage of the diverse data sources and surveillance streams available during the crisis.

Detailed computational models, on the other hand, place equal emphasis on modeling and real data and are thus increasingly preferred to their statistical and compartmental counterparts. The framework of agent-based modeling allows the policymaker to define behaviors at the individual and societal level, describe the characteristics of the disease pathogen, and simulate the infectious disease evolution on a realistic synthetic population (Eubank et al., 2004).

Building, testing and refining such detailed models is time consuming, and is seldom possible during *war time* efforts against an ongoing epidemic. Recently, organizations like CDC and NIH have conducted forecasting challenges for infectious diseases such as Flu, Dengue, etc. which serve as a valuable practice ground during the *peace time*. Ebola challenge was the most recent such endeavor, organized under the Research and Policy for Infectious Disease Dynamics (RAPIDD) program at NIH. Competing teams were required to provide epidemic forecasts for synthetic Ebola disease datasets under four different scenarios generated by a previously published agent-based model (Merler et al., 2015) calibrated for Liberia. The scenarios represented varying epidemiological conditions, behavioral changes, intervention measures and data availability. The teams were asked to submit county and country level forecasts of the disease, both short-term and long-term, along with estimates of epidemiological parameters at different timepoints of the simulated epidemic.

In this paper, we describe the efforts of our team from Virginia Tech in the Ebola forecasting challenge, wherein we used an agent-based model that was built during the Ebola 2014 outbreak to provide policy support to decision makers. We begin by briefly describing the construction of the agent-based model and define the disease model parameters that were used during the challenge (Section 2). The models were calibrated using a combination of simulation optimization and Bayesian calibration based approaches. We explain our calibration methodology, highlight how we used the provided synthetic datasets (Section 3) and summarize our performance statistics (Section 4). We finally conclude with discussions (Section 5) on the limitations and potentials for refinement of our approach, and share the lessons learned over the course of the challenge.

## 2. Model description

Agent-based network models designed for infectious diseases have three key components: a realistic synthetic population, social contact network among the individuals in the population, and an appropriate disease model (Eubank et al., 2004).

In our methodology, the synthetic population is generated with demographic attributes and household structure consistent with the census data. Each individual is then assigned an activity sequence with geo-locations per activity over the course of a 24-h period. The social contact network is obtained by considering co-location of individuals, and the edges in the network are weighted by the duration of co-location (see Fig. 1). The disease model is appropriately chosen, which translates the edge weights in the social contact network into infection probability of the edge over the course of a single day.

We now briefly describe the specific case of Liberia synthetic population and disease model as used during the course of the challenge. More details on the synthetic population generation process are provided in Mortveit et al. (2015).

### 2.1. Synthetic population

During the Ebola 2014 outbreak, our group had produced synthetic populations for the affected West African countries using methods based on the best available data. In the case of Liberia, we have released four different versions of the synthetic population available at <https://www.bi.vt.edu/ndssl/projects/ebola>.

We began with the distribution of number of workplaces, households and schools of various sizes in Liberia. Whenever the data was not available, we used the data from proxy countries (Nigeria in this case), which exhibit similar social demographics. ORNL LandScan data was used to obtain population densities, which were then used to map the households and workplaces to spatial cells. A statistically accurate base population, consistent with the Census data was constructed with suitable attributes (age and gender).<sup>1</sup> These individuals were then grouped into household units, which were then assigned to residence locations.

### 2.2. Social contact network

The activity locations were restricted to Home, Work and School/College.<sup>2</sup> Daytime location by age distribution was constructed using 2008 National Population and Housing Census and each individual was assigned a daytime location, either Home (same as his household location), Work, School/College. We also used the 2010 Liberia Labor Force Survey for generating a collection of activity sequence templates.

For each individual, a compatible activity sequence was sampled which chronologically lists the activities over a single day, with start and end times. Individuals were assigned matching locations based on their activity sequence. The bipartite person-location graph  $G_{PL}$  thus formed, was used to obtain the social contact network  $G_S$ . In addition to daily activity patterns, in order to mimic the inter-regional travel, we used the FlowMinder data to match the rates of movement between different counties. In addition to the base synthetic population, the open release <https://www.bi.vt.edu/ndssl/projects/ebola/ebola-data/synthetic-liberia> includes the synthetic geolocated locations, activities and the person-person contact graph for a typical day.

### 2.3. Simulation model

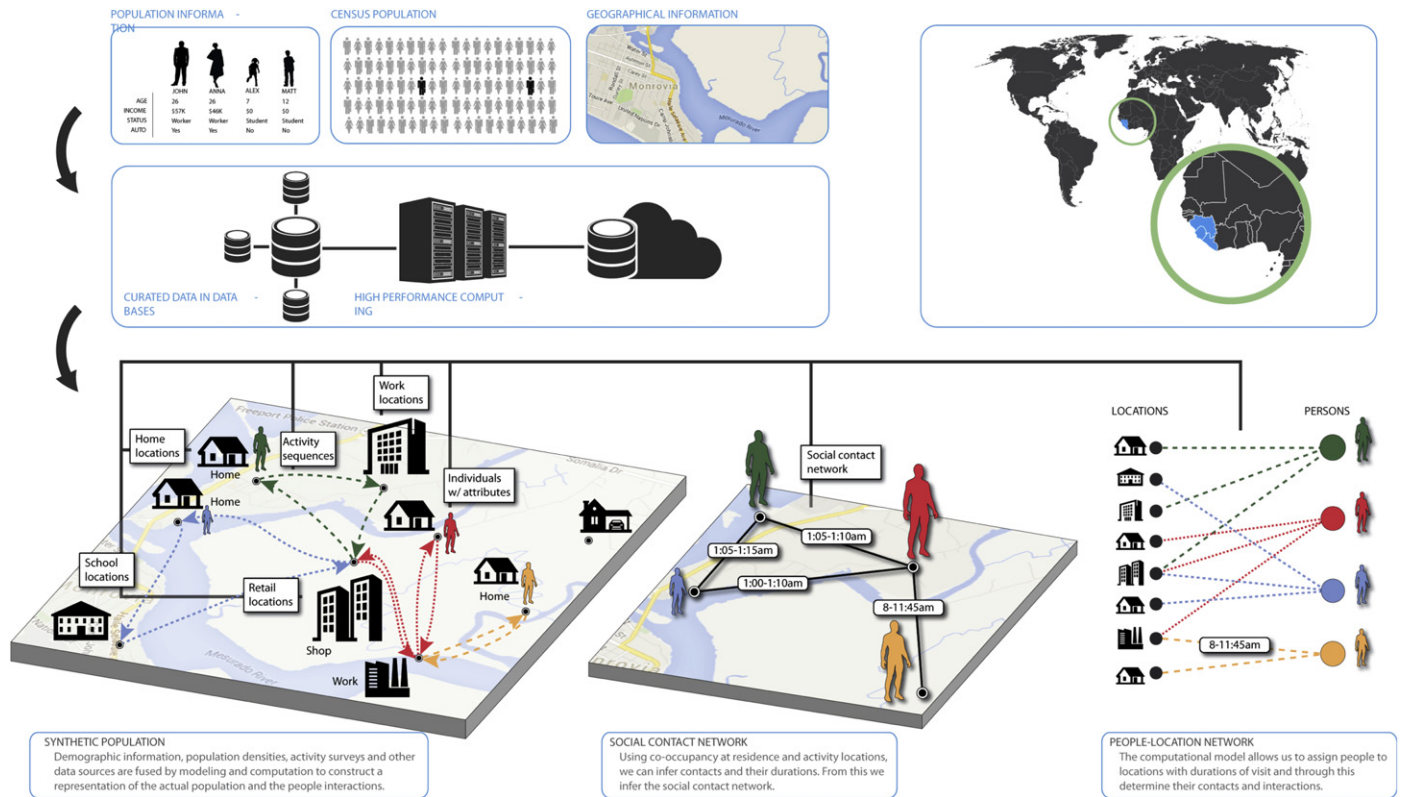
**Basic disease model:** The basic disease model used for the challenge is based on the standard stochastic SEIR (Susceptible → Exposed → Infected → Recovered) transmission dynamics. We used EpiFast (Bisset et al., 2009), a parallel algorithm and implementation for simulating the SEIR model. Key parameters of the model are Initial Infections, Transmissibility, Incubation Period Distribution and Infectious Period Distribution.

Initial Infections are fixed based on the situation report, and the cases are seeded in the appropriate county. From the second timepoint, we calibrated Initial Infections around the given value, to account for early reporting errors. Transmissibility is defined as the probability of infection for a susceptible individual per unit time of contact. Across all scenarios and timepoints, Transmissibility was calibrated by searching over the range  $[3 \times 10^{-5}, 8 \times 10^{-5}]$ .

Incubation Period Distribution and Infectious Period Distribution were initially estimated from the patient report database provided for the test scenario made available before the challenge. When the actual patient report database was

<sup>1</sup> Due to lack of sufficient data, healthcare workers were not explicitly identified.

<sup>2</sup> Funeral and Hospital locations were not explicitly modeled.



**Fig. 1.** Synthetic population generation process. The key inputs to this pipeline are the census demographics, activity patterns and the geospatial information corresponding to the region. The pipeline outputs (a) synthetic population of individuals and activity locations with explicit spatial embedding (b) a person–location bipartite graph based on activity sequences (c) an edge-weighted social contact network for disease propagation. (Courtesy: Mortveit et al. (2015).)

provided (Scenario 1, all timepoints), and (Scenario 3, timepoint 5), we used them instead of that of the test scenario. Figure 6 in the Supplementary material shows a comparison of the different distributions estimated during the course of the challenge.

**Hospitalization and Funeral:** The basic disease model does not have explicit states to model *Hospitalization* and *Funeral*. In order to incorporate the H(Hospitalized) and F(Funeral) states as described in the SEIHF model proposed by Legrand et al. (2007), we implemented *interventions* that reduce the transmissibility of an individual (modeled by *Efficacy*) for those who comply (modeled by *Compliance*) after an appropriate *Delay* and applicable for an appropriate *Duration*. For each scenario, for early timepoints when safe burials were not practiced, the *Compliance* for Funeral was set to 0 (no reduction in transmissibility). When not being calibrated, the *Delay* of each intervention was fixed to mean time from symptom onset to death/hospital from Team (2014).

**Social Mobility:** We included two other parameters which were largely fixed through the course of the challenge to model how the intensity on different social edges were weighted. *Natural Isolation* was used to reduce the transmissibility of outgoing non-household edges of infectious individuals. *Travel Reduction* was used to reduce the transmissibility (or completely cut off) long range interactions. The latter was done to better mimic the interactions in Merler et al.'s (2015) synthetic population, where individuals' long range mobility is limited.

**Behavioral interventions:** For the challenge, we did not explicitly model Ebola Treatment Units (ETU) or contact tracing. Instead, our model included two abstract behavioral interventions, intended to reduce the transmissibility of edges within different counties. The *Primary Intervention* was triggered when the number of new cases on any given day exceeded a particular threshold (modeled by *Threshold Fraction*). The

*Secondary Intervention* was triggered after an appropriate *Action Delay*. Both these interventions also included *Efficacy* and *Compliance* parameters.

**Data usage:** Across the scenarios, national epicurves were used for calibration. Epidemics were seeded in the appropriate county for each scenario. For the data rich scenario, though the regional (county-wise) epicurves were made available, due to spatial variations in our social contact network, we were unable to use it effectively to improve our results. Incubation and Infectious period distributions were inferred from the patient database whenever available. As for interventions, since our model did not explicitly model safe burials, ETUs or contact tracing, we could not use any of the quantitative information provided in the situation report. We however used it qualitatively, to turn on and tune the *Efficacy* ranges of our behavioral interventions.

Not all parameters were calibrated for all timepoints. However, within each timepoint, we used identical configurations for all scenarios (i.e., which parameters were searched/fixed/estimated). Table 6 in the Supplementary Material shows the configurations used for each timepoint.

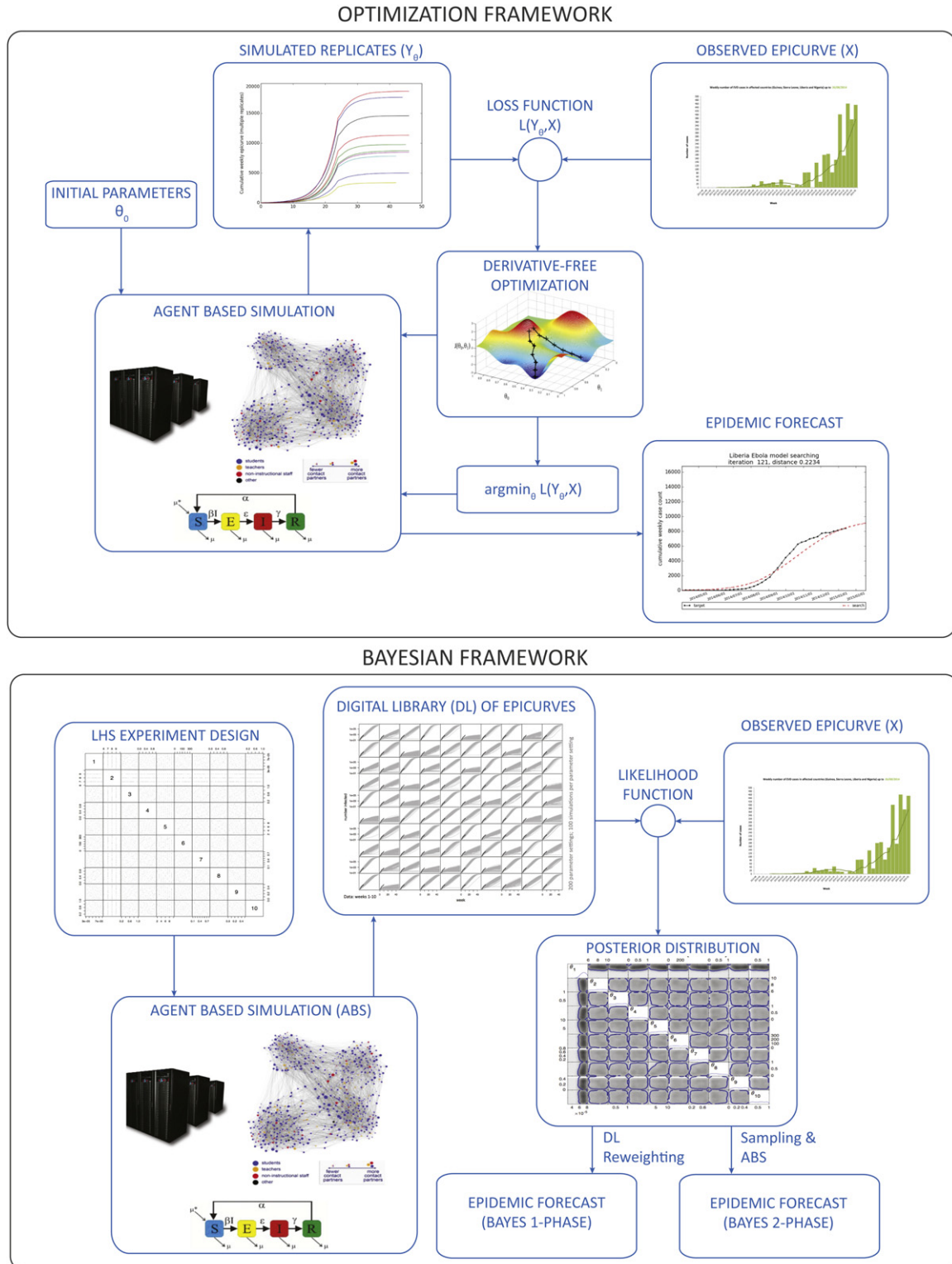
### 3. Model calibration

Model calibration is the process of identifying the parameter configurations for the model that best explain the observed ground truth. While simple models with fewer parameters can be calibrated by a naive parameter sweep, calibration of complex models require extensive computational effort. In this section, we briefly describe the different methodologies that formed part of our calibration and forecasting pipeline. The methods used for calibration follow two standard philosophies: (a) simulation optimization and (b) Bayesian calibration.

In the former approach, one poses calibration as an optimization problem, by trying to minimize the difference between the simulated and observed epicurves (Nsoesie et al., 2013). Since the function being optimized does not have an analytical form, and can be evaluated only through a simulation oracle, these methods are also known as *black-box optimization*. The task becomes one of searching for the minimizer of the loss function in the parameter

space, starting from an initial guess, also referred to as *direct search* (Audet, 2014). The obtained minimizer is then used to produce forecasts.

In the latter approach, one begins with a prior distribution over the parameter space, for the possible configurations that may produce the observed epicurve. In order to update the belief, one uses the Bayesian approach (Kennedy and O'Hagan, 2001). This is done



**Fig. 2.** Two approaches to model calibration (a) simulation optimization and (b) Bayesian calibration. The calibration framework is described in more detail in Venkatramanan et al. (2016).



by first constructing a response surface based on sampled simulation runs, and using it compute the likelihood function. The obtained posterior distribution is then used for generating forecasts. Depending on whether new runs were used for forecasting or earlier runs were re-weighted, we had two flavors of the Bayesian calibration, namely 1-phase and 2-phase.

Fig. 2 shows the different steps involved in the calibration process for both the methodologies.

### 3.1. Simulation optimization

We use the standard Nelder–Mead (NM) direct search method (Nelder and Mead, 1965; Python, 2016) for calibrating the output of our agent-based model to the given national epidemic curve. A crucial design choice for the optimization approach is the loss function that measures the error between the simulated replicates and the observed epicurve. We tested various objective functions, such as the normalized versions of L1 norm, L2 norm and weighted-L1 norm (with different exponential weights). If  $\mathbf{Y} = \{Y_i\}_1^T$  denotes the target time series until week  $T$ , and  $\mathbf{X} = \{X_i\}_1^T$  denotes the corresponding output of our simulation, then these respectively can be written down as follows:

$$\begin{aligned} \text{L1 norm: } d(\mathbf{X}, \mathbf{Y}) &= \frac{\sum_{i=1}^T |X_i - Y_i|}{\sum_{i=1}^T |X_i|}; \quad \text{L2 norm: } d(\mathbf{X}, \mathbf{Y}) = \frac{\sum_{i=1}^T (X_i - Y_i)^2}{\sum_{i=1}^T X_i^2} \\ \text{weighted-L1 norm with decay factor } \alpha: d(\mathbf{X}, \mathbf{Y}) &= \frac{\sum_{i=1}^T \alpha^{T-i} |X_i - Y_i|}{\sum_{i=1}^T \alpha^{T-i} |X_i|} \end{aligned}$$

This methodology was tested both using the incremental and cumulative version of the epidemic curves. From our experience, we observe that using the incremental epidemic curve, with a weighted-L1 norm and reasonably high decay ( $\alpha = 0.4$ ) performed very well during calibration. The mean dissimilarity for a given iteration of search was obtained by averaging across several stochastic replicates (ranging from 10 to 30). The search algorithm terminates if the mean dissimilarity falls below a tolerance value or if it exceeds the maximum number of iterations (set to 300). Once we get a convergent parameter configuration, we run the simulation forward (for 400 days) with 100 stochastic replicates and use it to generate the required predictions.

### 3.2. Bayesian calibration

While the simulation optimization approach is faster (requiring fewer simulation runs), given the dimensions of the search space it has two primary drawbacks: (a) the search may converge to a local minimum of the loss function (b) the forecasts only capture the stochasticity of the simulation and not the parameter uncertainty. To account for these, we also experimented with a Bayesian setup following the procedure outlined in Higdon et al. (2008).

To begin with, a Latin Hypercube Sample (McKay and Beckman, 2000) of size 100 is taken over the parameter space, and simulations are carried out (with  $100 \times$  replication) to produce a digital library (DL). The prior distribution of the parameter vector  $\theta$  is assumed to be uniform over a pre-specified rectangle. A response surface is fit to this DL producing a mapping from the model parameters to the simulated epidemic curve, giving the number of infected by week. Using this response surface, the likelihood of the observed epicurve  $Y_i$  can be evaluated as a function of the model parameter setting. Here the data are assumed to be normal, with a standard deviation of 10%, and independent over each time period. Thus the log-likelihood is given by

$$L(\theta|Y) = \text{const} - \frac{1}{2} \sum_{i=1}^T \frac{|X_i(\theta) - Y_i|^2}{0.1 \cdot X_i(\theta)}$$

**Table 1**

Calibration methods used during the challenge.

Timepoint	Method used
1	Optimization (Nelder–Mead)
2	Bayesian (1-phase)
3	Bayesian (2-phase)
4	Optimization (Nelder–Mead)
5	Bayesian (1-phase)

**Bayes 1-phase vs 2-phase:** In order convert the obtained posterior distribution into an ensemble of epicurves to be used for forecasting, we tested out two different approaches: (a) use the posterior distribution to appropriately reweight the curves present in the Digital Library. (b) Sample parameter configurations from the posterior distribution and run fresh simulations to produce the epicurves. We call these approaches based on DL reweighting and sampling & running the agent-based simulation as 1-phase and 2-phase respectively.

As a learning exercise, we tested out the different methods during the course of the challenge. Table 1 shows which calibration method was used to generate the submitted forecasts for various timepoints.

## 4. Model forecasts

### 4.1. Generating predictions

Both our calibration techniques finally produce an ensemble of 100 epicurves, which are then used to calculate various predictions. We follow the same procedure for short-term forecasts, long-term forecasts and estimation of epidemic parameters. For instance, consider the quantity of interest to be peak size. The peak size is calculated for each of the epicurves in the ensemble, and the predictions are produced as the median, 25th and 75th percentiles of the vector of peak sizes.

While predictions based directly on the epicurve such as short-term incidence estimates, peak size and timing, and final size are straightforward to calculate, estimation of epidemic parameters is highly dependent on the underlying model. In fact, the definitions and usage of parameters such as case fatality rate (CFR), reproduction number  $R_0$  and serial interval vary across models. We briefly describe below how these were estimated for our submissions.

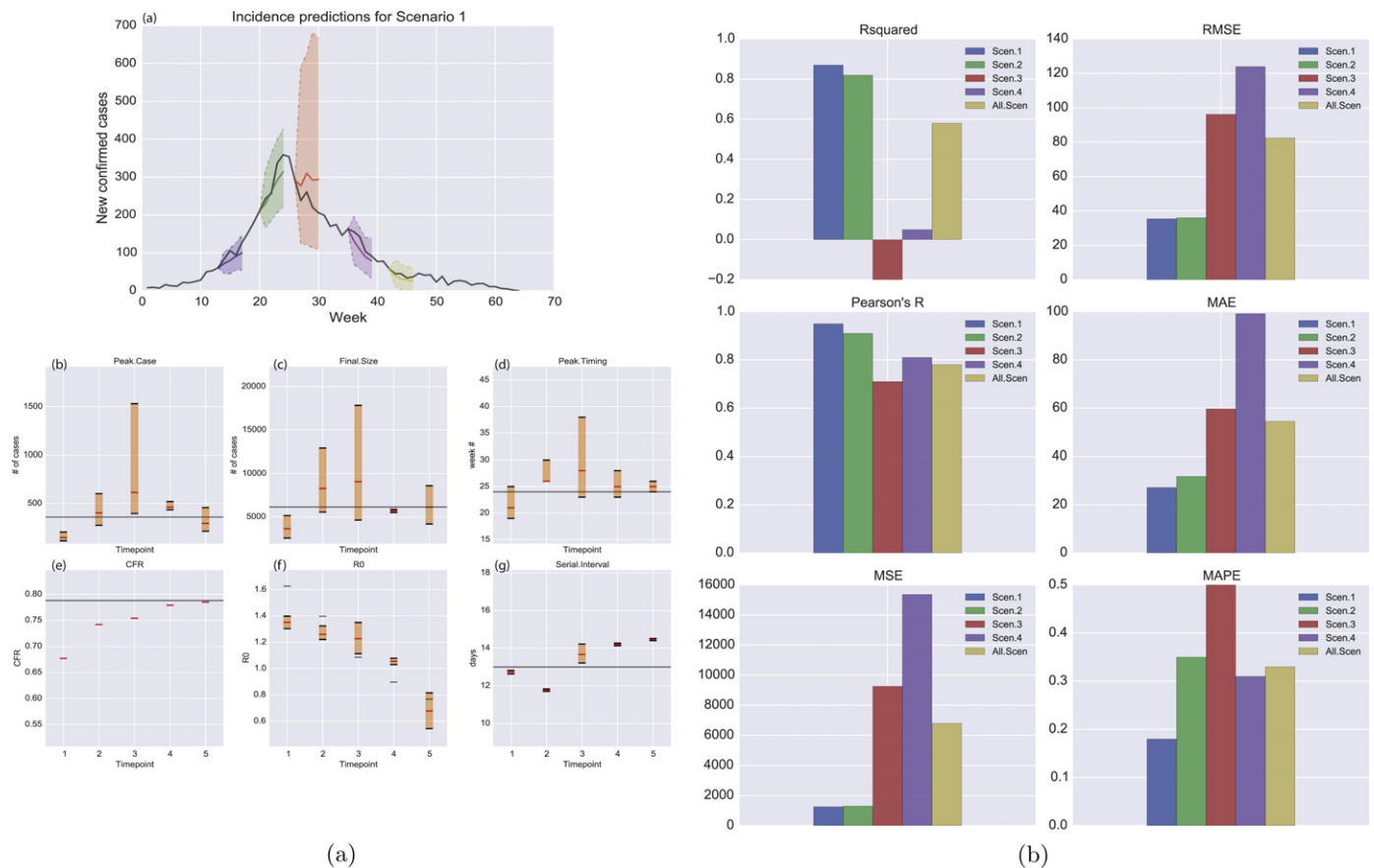
**Reproduction number  $R_0$ :** The basic reproduction number is defined as the number of secondary infections produced by a primary case in a completely susceptible population. In our case, for each of the observed epicurves, we extracted the corresponding disease transmission tree. We reported the effective reproduction number at week  $k$  as the average out-degree of individuals infected in the four weeks prior to week  $k$ . Let  $\mathcal{I}_{k,4}$  correspond to the set of nodes infected in weeks  $[k-3, k]$ , and  $d_i^{\text{out}}$  be the out-degree of node  $i$ :

$$R_0 = \frac{\sum_{i \in \mathcal{I}_{k,4}} d_i^{\text{out}}}{|\mathcal{I}_{k,4}|}$$

**Serial interval:** Serial Interval is defined as the time between successive cases in a chain of transmission. Similar to  $R_0$ , serial interval is estimated across the last four weeks prior to week  $k$ , by averaging the difference between infection times across the edges in the transmission tree  $\mathcal{T}$ . If  $T_i$  denotes the time of infection of node  $i$ ,

$$S.I. = \frac{\sum_{i \in \mathcal{I}_{k,4}, (i,j) \in \mathcal{T}} (T_j - T_i)}{\sum_{i \in \mathcal{I}_{k,4}} d_i^{\text{out}}}$$

**Case fatality rate (CFR):** CFR measures the proportion of deaths among the reported confirmed cases. Since our epidemic model



**Fig. 3.** Forecast Performance Summary. (a) Submitted predictions for Scenario 1: (top to bottom, left to right) short-term forecasts, peak case count, final epidemic size, peak timing, case fatality rate (CFR), reproduction number ( $R_0$ ), serial interval. The black/grey lines represent the ground truth, and submissions with confidence interval are shown for each of the five timepoints. (b) Incidence error measures for all scenarios (top to bottom, left to right): R-squared, Root Mean Square Error (RMSE), Pearson's R, Mean Absolute Error (MAE), Mean Square Error (MSE) and Mean Absolute Percentage Error (MAPE). Scores shown for each scenario (in order) across all timepoints, and finally the score across all scenarios.

does not explicitly differentiate between individuals who die or recover with immunity, CFR was estimated from the given time series until the current week  $k$ .

$$CFR = \frac{\# \text{ of deaths until week } k}{\# \text{ of confirmed cases until week } k - 1}$$

The one week lag between deaths and confirmed cases was heuristically chosen based on estimated mean infectious duration.

#### 4.2. Performance statistics

Fig. 3 depicts the submitted predictions for the data-rich scenario (Scenario 1). The corresponding figures for Scenario 2–4 are provided in the Supplementary Material. Of the four scenarios, the national epicurve of Scenario 1 exhibited the most “standard” trajectory, followed by Scenario 2. While Scenario 3 exhibited a sharp decline of cases after reaching the peak, Scenario 4 had the epidemic increasing even beyond the 5th timepoint. As a result, the best performance of our predictive model was observed in Scenario 1.

This can also be seen from the various error measures computed for incidence predictions across scenarios (Fig. 3). Scenario 1 predictions have the highest score for  $R^2$  and Pearson's  $R$ , both of which measure correlation between the submitted and the observed curve, and are thus *positively oriented* (higher the score, the better). Further, Scenario 1 predictions have the lowest value for the *negatively oriented* error metrics (lower the score, the better) such as RMSE (Root Mean Squared Error), MAE (Mean Absolute Error), Mean Squared Error (MSE) and MAPE (Mean Absolute Percentage Error).

The greater the difference between RMSE and MAE implies greater variance in the individual errors in the sample. By this metric, Scenario 2 submissions have lower variance in the error in comparison to Scenario 1. Scenario 3 performance ranks the lowest when considering correlation metrics such as  $R^2$  and Pearson's  $R$ . While Scenario 4 has the worst performance according to most error metrics (RMSE, MSE and MAE), in terms of percentage error (MAPE), Scenario 3 ranks the lowest. It is interesting to observe that even for incidence predictions from the same model, varying the error metrics leads to different performance rankings. *Fairly evaluating and ranking predictive models across epidemic measures and error metrics while accounting for the diversity among the models is an interesting open challenge.*

Considering the long-term forecasts, we observe from Fig. 3 that the median predictions of our model are quite close to the ground truth, at least for scenario 1. For the peak predictions, we predict within a margin of 200 cases (peak case count) and 2–3 weeks (peak timing) as early as timepoint 1. Though our estimates for final size alternate between under and over-estimation, they are centered around the observed final size. This is remarkable considering that we tested various calibration techniques across timepoints, and used several variants of the disease model. *This demonstrates the predictive power of a well-calibrated detailed computational model, and its utility in aiding early decision-making.*

*Note on post-peak predictions:* One may observe that the model continues to miss the exact peak week, even after having observed the peak in the epidemic curve. This is because the peaks of the calibrated epidemic curves will not generally coincide with the observed epidemic curve (or with each other). *Since the observed*

epidemic curves are inherently noisy, we must seek a plausible ensemble of epidemic trajectories on which to evaluate actions for combating and managing the epidemic. Note on confidence intervals: Across timepoints, the submissions for timepoint 3 show a wider confidence interval. We believe this was primarily due to the 2-phase variant of Bayesian calibration method. This was further exacerbated by the fact that short-term evolution at timepoint 3 (being close to the peak in three of the scenarios) is often tough to predict well. *Producing reliable confidence intervals that account for noisy data, model uncertainty, parameter uncertainty and inherent stochasticity of the system requires further systematic investigation.*

## 5. Discussion

In conclusion, we have described a data-driven agent-based modeling framework targeted at epidemic forecasting. We have provided an overview of the model construction, calibration and forecasting stages, and summarized its performance in the Ebola forecasting challenge.

Despite the strong performance, we still see lots of scope for refining and improving our approach. For instance, we found that despite (or perhaps, due to) the inherent spatial heterogeneity, reproducing the county-wise epicurves was found to be difficult. Calibrating an agent-based model across spatial scales remains an open question to be addressed. We also continue to refine the synthetic population by adding more attributes, updating the data sources and incorporating aspects specific to the dynamics of an emerging epidemic.

It must be noted that the two agent-based modeling approaches (ours and Merler et al.'s (2015) used by the organizers) have differences in the underlying population, social network construction and the details of the disease model. Each captures different aspects of real world dynamics better. For instance, healthcare workers and ETU are explicitly modeled in Merler et al. (2015), while our synthetic population has a better representation of work/school activities and long range mobility. While they use an MCMC approach for calibration, we employ approaches based on simulation optimization and Bayesian inference. We find that, in general, qualitative and quantitative comparison of two different agent-based modeling frameworks is in itself quite a challenge, and is outside the scope of this paper.

An interesting hurdle one faces when employing a highly-detailed model as ours is the huge number of parameters that need estimation or calibration. Calibrating a complex model in the presence of limited data will lead to parameter identifiability issues. A rigorous method to address this issue would be to perform a thorough sensitivity analysis of the model, and calibrating only the important parameters. Due to the evolving nature of the challenge, we instead resorted to calibrating different subset of parameters across timepoints, and fixing the parameters which did not vary much during the calibration process.

However, there is no denying that, an advantage of the structural agent-based modeling approach is the incorporation of qualitative data into the modeling framework. Oftentimes, when data is limited, these pieces of qualitative information is all decision makers have. Including this information into a quantitative simulation platform is a big step forward for enabling real-time simulation support for epidemic response.

The practical usage of a detailed computational model also speaks to the ever-increasing role played by high performance computing, efficient parallel algorithms and large scale data management in real-time critical decision making. Such model excels both in its descriptive, predictive and prescriptive capability, and show how a rigorous approach to data-driven modeling can benefit both modelers and policymakers.

## Acknowledgments

We extend our thanks to the organizers of the Ebola Forecasting Challenge for their thoughtful design and efficient execution. We would like to thank the members of Network Dynamics and Simulation Science Laboratory at the Biocomplexity Institute of Virginia Tech for contributing to the entire effort.

The work has been partially supported by DTRA Grant HDTRA1-11-1-0016, DTRA CNIMS Contract HDTRA1-11-D-0016-0001, NSF NetSE Grant CNS-1011769 and NIH MIDAS Grant 5U01GM070694.

## Appendix A. Supplementary data

Supplementary data associated with this article can be found, in the online version, at <http://dx.doi.org/10.1016/j.epidem.2017.02.010>.

## References

- Anderson, R.M., May, R.M., Anderson, B., 1992. Infectious Diseases of Humans: Dynamics and Control, vol. 28. Wiley Online Library.
- Audet, C., 2014. A survey on direct search methods for blackbox optimization and their applications. In: Mathematics Without Boundaries. Springer, pp. 31–56.
- Bisset, K.R., Chen, J., Feng, X., Kumar, V., Marathe, M.V., 2009. Epifast: a fast algorithm for large scale realistic epidemic simulations on distributed memory systems. In: Proceedings of the 23rd International Conference on Supercomputing. ACM, pp. 430–439.
- Bright, E.A., Rose, A.N., Urban, M.L., 2012. Landscan.
- Burnet, M., White, D.O., 1972. Natural History of Infectious Disease. CUP Archive.
- Cook, S., Conrad, C., Fowlkes, A.L., Mohebbi, M.H., 2011. Assessing google flu trends performance in the United States during the 2009 influenza virus (a/h1n1) pandemic. *PLoS ONE* 6 (8), e23610.
- Daszak, P., Cunningham, A.A., Hyatt, A.D., 2000. Emerging infectious diseases of wildlife—threats to biodiversity and human health. *Science* 287 (5452), 443–449.
- 2015. Dengue Forecasting Challenge. <http://dengueforecasting.noaa.gov/>.
- Eubank, S., Guclu, H., Kumar, V.A., Marathe, M.V., Srinivasan, A., Toroczkai, Z., Wang, N., 2004. Modelling disease outbreaks in realistic urban social networks. *Nature* 429 (6988), 180–184.
- Fauci, A.S., 2014. Ebola—underscoring the global disparities in health care resources. *N. Engl. J. Med.* 371 (12), 1084–1086.
- 2014. Flowminder.org. <http://www.flowminder.org>.
- 2015. FluSight: Seasonal Influenza Forecasting. <http://predict.phiresearchlab.org/>.
- Higdon, D., Gattiker, J., Williams, B., Rightley, M., 2008. Computer model calibration using high-dimensional output. *J. Am. Stat. Assoc.*
- Kennedy, M.C., O'Hagan, A., 2001. Bayesian calibration of computer models. *J. R. Stat. Soc. Ser. B Stat. Methodol.* 63 (3), 425–464.
- Lazer, D., Kennedy, R., King, G., Vespignani, A., 2014. The parable of google flu: traps in big data analysis. *Science* 343 (6176), 1203–1205.
- Legrand, J., Grais, R.F., Boelle, P.-Y., Valleron, A.-J., Flahault, A., 2007. Understanding the dynamics of Ebola epidemics. *Epidemiol. Infect.* 135 (4), 610–621.
- McKay, M., Beckman, R., 2000. A comparison of three methods for selecting values of input variables in the analysis of output from a computer code. *TECHNOMETRICS* 21 (2).
- Ajelli, M., Fumanelli, L., Gomes, M.F., y Piontti, A.P., Rossi, L., Chao, D.L., Longini, I.M., Halloran, M.E., Vespignani, A., 2015. Spatiotemporal spread of the 2014 outbreak of ebola virus disease in Liberia and the effectiveness of non-pharmaceutical interventions: a computational modelling analysis. *Lancet Infect. Dis.* 15 (2), 204–211.
- Mortveit, H., Adiga, A., Agashe, A., Alam, M., Alexander, K., Arifuzzaman, S., Barrett, C., Beckman, R., Bisset, K., Chen, J., Chungbaek, Y., Eubank, S., Gupta, S., Khan, M., Kuhlman, C., Lofgren, E., Lewis, B., Marathe, A., Nordberg, E., Rivers, C., Stretz, P., Swarup, S., Wilson, A., Xie, D., Marathe, M., 2015. Synthetic Populations and Interaction Networks for Guinea, Liberia and Sierra Leone. Tech. Rep. NDSSL.
- Nelder, J.A., Mead, R., 1965. A simplex method for function minimization. *Comput. J.* 7 (4), 308–313.
- Nsoesie, E.O., Beckman, R.J., Shashaani, S., Nagaraj, K.S., Marathe, M.V., 2013. A simulation optimization approach to epidemic forecasting. *PLOS ONE* 8 (6), e67164.
- 2016. Python SciPy v0.17.0 Reference Guide: Optimization and root finding. <http://docs.scipy.org/doc/scipy-0.17.0/reference/optimize.html>.
- 2015. RAPIDD Ebola Challenge. <http://ebola-challenge.org>.
- Team, W.E.R., 2014. Ebola virus disease in West Africa—the first 9 months of the epidemic and forward projections. *N. Engl. J. Med.* 371 (1), 1481–1495.
- Venkatramanan, S., Lewis, B., Chen, J., Higdon, D., Vullikanti, A., Marathe, M., 2016. Using Optimization and Bayesian Approaches for Calibrating Agent-Based Models of Infectious Diseases. Tech. Rep. NDSSL.