# Epidemic Forecasting Framework Combining Agent-Based Models and Smart Beam Particle Filtering

Farzaneh S. Tabataba*, Bryan Lewis*, Milad Hosseinipour†, Foroogh S. Tabataba‡,
Srinivasan Venkatramanan*, Jiangzhuo Chen*, Dave Higdon§, and Madhav Marathe*
* Network Dynamics and Simulation Science Laboratory, Virginia Tech, Blacksburg, Virginia 24060, Email: fstaba2@vt.edu
† Virginia Tech
‡ECE Department, Isfahan University of Technology
§Social Decision Analytics Laboratory, Virginia Tech

*Abstract*—**Over the past decades, numerous techniques have been developed to forecast the temporal evolution of epidemic outbreaks. This paper proposes an approach that combines high resolution agent-based models using realistic social contact networks for simulating epidemic evolution with a particle filter based method for assimilation based forecasting. Agent-based modeling using realistic social contact networks provides two key advantages: (*i*) they capture the causal processes underlying the epidemic and hence are useful to understand the role of interventions on the course of the epidemics – typically time series models cannot capture this and as a result often do not perform well in such situations; (*ii*) they provide detailed forecast information – this allows us to produce forecast at high levels of temporal, spatial and social granularity. We also propose a new variation of particle filter technique called *beam search particle filtering*. The modification allows us to more efficiently search the parameter space which is necessitated by the fact that agent-based techniques are computationally expensive.**

**We illustrate our methodology on the synthetic dataset of Ebola provided as a part of the NSF/NIH Ebola forecasting challenge. Our results show the efficacy of the proposed approach and suggest that agent-based causal models can be combined with filtering techniques to yield a new class of assimilation models for infectious disease forecasting.**

## I. INTRODUCTION

Epidemics cause significant social, economic, and health impact on societies [1]. More than 60 million cases of H1N1 were reported between April 2009 and April 2010 in the United States. Infectious disease outbreaks of Ebola in West Africa (2014), Cholera in Bangladesh (2011), and Zika in Brazil (2015) remind us that despite of significant progress by authorities around the world, infectious diseases continue to be an important global societal challenge.

Reliable and granular forecasting of infectious diseases facilitates intervention planning and targeted resource distribution to alleviate the negative impacts of pandemics. There have been multiple efforts to forecast seasonal as well as emerging disease outbreaks. The choice of a reliable model to emulate the disease's behavior remains a challenge. Recent efforts relied on mathematical compartment models like SEIR and SIR [2] as the disease dynamics simulator. Some studies improved model fidelity by considering the individual-level interactions across social networks [3]–[6]. However, these individual-based models impose an additional burden on computational resources by increasing the number of unknown parameters that should be adjusted through data assimilation. Moreover, the choice of calibration method can drastically affect the accuracy of the epidemic model and thus forecast performance. A variety of filtering techniques have been utilized in the context of epidemic forecasting [7], [8] and particle filters stand out among other methods because of their promising capabilities in dealing with non-linear and multi-modal systems [8].

**Our Contributions and significance.** In this paper, we propose a new methodology for forecasting infectious disease dynamics. The basic approach combines two powerful techniques: (*i*) we use a high resolution agent-based model to simulate disease evolution and (*ii*) we use a variant of particle filtering technique (called *smart beam particle filtering technique*) for assimilation-based forecasting. Together the approach provides a general technique to forecast a range of infectious diseases. We illustrate the method by carrying out detailed computational experiments using Ebola data which was generated as a part of the National Institutes of Health (NIH) (RAPIDD Ebola Challenge) [9]. We use realistic social contact networks for cities in the west Africa. These high quality representations of social contact networks provide a unique starting point. Additionally, we use a novel epidemic simulator called *Epifast* [3], that simulates not only disease dynamics but also can account for dynamic interventions that occur during the course of an outbreak.

To the best of our knowledge, no prior effort has been made towards combining a network agent-based model with a particle filter framework as the data assimilation method. Recent papers have used the particle filter as the data assimilator in the context of outbreak forecasting. These efforts, use compartmental models (SEIR models or their extensions) [8], [10]–[16] to simulate disease dynamics. Although computationally efficient, the models make simplifying assumption regarding well mixed populations that are homogeneous. Recent models

extended this by using a contact factor that determines a percentage, or fixed number of contacts between susceptible and infected individuals [17], [18].

Simulating disease propagation by using realistic agent-based models impose significant computational costs and hence impose restrictions on the number of particles as well as the number of running iterations of the particle filter. Most particle filters used in the context of epidemic forecasting generate 5,000 to one million particles to establish converged best results. Running agent-based methods with such a large number of particles is not practical. To solve this problem, we propose a smart diffuser embedded in the state dynamics of the particle filter. The smart diffuser examines the predicted and observed data based on machine learning methods and determines the direction and size of the perturbation for each parameter of the state vectors. The smart diffuser enables the particle filter to find more accurate results with a lower number of particles and fewer running iterations.

## II. NETWORKED EPIDEMIOLOGY

Networked epidemiology studies epidemic processes over networks; over the last decade these models have become popular owing to their ability to incorporate spatial and individual level heterogeneity and complex interventions. See [4], [19] for additional discussion. We briefly describe the terminology here. Let $G(V, E)$ denote a social contact graph on a population $V$ – each edge $e = (u, v) \in E$ denotes that the individuals (also referred to as nodes) $u, v \in V$ come into contact. Let $N(v)$ denote the set of neighbors of $v$.

For the SEIR model on the graph $G$, we have a dynamical process with each node being in $S$, $E$, $I$ or $R$ states. Infection can potentially spread from $u$ to $v$ along edge $e = (u, v)$ with a probability of $\beta(e, t)$ at time instant $t$ after $u$ becomes infected, conditional on node $v$ remaining uninfected until time $t$.

Let $\tau(u)$ denote the time that node $u$ would remain in the infected state, and let $\tau = \max\{\tau(u) : u \in V\}$. If a node $u \in V$ gets infected at time $t_u$, it attempts to infect each susceptible neighbor $v$ with probability $\beta((u, v), t - t_u)$ for $t = t_u + 1, \ldots, t_u + 1/\lambda(u)$. After $1/\lambda(u)$ steps (called the period of infection), node $u$ switches to state $R$. If the susceptible node $v$ contracts the infection it moves to Exposed state $E$ and stays in this state for $1/\gamma(v)$ (called the incubation period) steps and then moves to state $I$. A node in state $E$ cannot infect other nodes. We let $I(t)$ denote the set of nodes that become infected at time $t$. The sequence $I(t)$, along with the (random) subset of edges on which the infections spread, represent a disease outcome, also referred to as a *dendogram*. The time series $(|I(t)|, t = 0, 1, \ldots)$ is referred to as an *epidemic curve* corresponding to a stochastic outcome. This dynamical system starts with a configuration in which there are one or more nodes in state $I$ and reaches a fixed point in which all nodes are in states $S$ or $R$.

### A. Agent-Based Network Model

We use a high performance computing based modeling environment called *Epifast* for simulating epidemic and as-

TABLE I
PARAMETERS TO CONTROL THE EPIDEMIC MODEL AND INTERVENTIONS.

| | Parameter | Calibrated/Fixed |
|---|---|---|
| $\beta$ | Transmission Rate | Calibrated |
| $\gamma$ | Incubation period distribution | Fixed |
| $\lambda$ | Infectious period distribution | Fixed |
| $I_0$ | No. of initial infections | Calibrated |
| $\omega_{nsb}$ | Natural isolation efficacy | Fixed |
| $\beta_{HT}$ | Hospitalization efficacy | Calibrated |
| $\beta_{HD}$ | Hospitalization delay | Calibrated |
| $\omega_{tr}$ | Travel reduction ratio | Calibrated |
| $beta_{ETU}$ | Ebola treatment units' efficacy | Fixed |

sociated interventions over social contact networks [3]. Epifast simulates the spatiotemporal propagation of the disease through social interactions between individuals. Epifast provides the capability of simulating a broad range of policy-based, as well as individual-based, interventions. Interventions could be pharmaceutical (PI) or non-pharmaceutical (NPI). PIs include dispensing of antivirals, vaccines, and antibiotics, whereas NPIs refer to any change in individual interactions or social network structure aside from using medicine. More information about mathematical modeling and the details of such networks can be found in [6], [19].

### B. Agent-Based Model Output

Epifast provides detailed outputs as compared with the compartment SEIR model. The outputs of Epifast are as follows: ($i$) The daily health status of each individual; ($ii$) Dendogram that provides detailed temporal description of who infects whom and at what time; ($iii$) Health status and statistics for different subpopulations. A complete demographic data for each individual makes it possible to obtain different epidemic statistics for subpopulations in various categories based on age, gender, regions, etc.

### C. Simulating Ebola Propagation by Epifast

Epifast provides extensive freedom and capabilities for modeling disease dynamics and interventions. Those capabilities are controlled by different parameters that are mostly unknown. When Epifast uses the SEIR model to simulate Ebola dynamics, the minimum-required parameters are those that control SEIR dynamics like transmission rate ($\beta$), mean incubation period ($1/\gamma$), and mean infectious period ($1/\lambda$). Depending on the available information about disease characteristics and type of applied interventions to control the epidemic, additional parameters could show up. Table I presents the parameters used to characterize the Ebola models. Some of these parameters are estimated and fixed according to prior knowledge, and others are calibrated through a data-driven particle filter framework.

## III. PARTICLE FILTER

Particle Filter (PF), which is best known as bootstrap filtering [20], is a sequential Monte Carlo (SMC) approach to estimate the posterior distribution defined in a discrete Bayesian filter. The basic idea of this method is to represent

the posterior density function of a state vector by a set of weighted random particles. Each particle is a sample of the state vector and is shown as $x_i$. The PF algorithm consists of three main steps.

*Initialization.* The initial values of particles are set randomly by sampling from a prior distribution defined for the state vector. The PF has two fundamental parts: state dynamics and observation equations. The state dynamics (system model) describes the change of the state vector over time and is assumed to be a function of the state vector in the preceding step: $X_k = f_{k-1}(X_{k-1}, V_{k-1})$, where $X_k$ is the state vector to be estimated, $f_{k-1}$ is the known possible non-linear function, and $V_{k-1}$ is the system error. This equation corresponds to transition probability density function (pdf) $p(x_k|x_{k-1})$ in the probabilistic description of the state evolution in a Bayesian filter. The observation (measurement) equation shows the relation between the observed data and the state vector: $y_k = h_k(X_k, W_k)$, where $W_k$ is the observation error. This function corresponds to the likelihood function $p(y_k|x_k)$ in a Bayesian model. The initial condition of this model is defined as the prior pdf $p(x_0)$ for the state vector in which no measurement has been received. After the initialization step is complete, each iteration comprises of two steps.

*Prediction step.* PF uses $N$ particles to estimate the posterior distribution of the state vector. The initial particles in step $k$ are the samples from the posterior pdf $p(x_{k-1}|Y_{k-1})$ generated in the previous cycle, where $Y_{k-1}$ is all observations received up to time $k-1$ and including $y_{k-1}$ : $Y_{k-1} = \{y_i, i = 1, ..., k-1\}$. These particles are denoted by $\{x_{k-1}^{i*}\}_{i=1}^{N}$. In the prediction phase, the samples from step $k-1$ are passed through the state dynamics to generate the new set of samples, which are prior particles in step $k$:

$$x_k^i = f_{k-1}(x_{k-1}^{i*}, v_{k-1}^i)$$

This new set of particles $\{x_k^i\}$ represents the samples of the prior pdf $p(x_k|Y_{k-1})$, that is the prediction of the new state vector based on the previous observed data.

*Update Step.* In the update step, the prior samples $\{x_k^i\}$ are updated based on the new measurement received at step $k$ $(Y_k)$. A weight $\tilde{w}_k^i$ is calculated for each particle based on these measured values. This weight defines the likelihood of the observed value based on the prior sample: $\tilde{w}_k^i = p(y_k|x_k^i)$. The weights are normalized and calculated with the following equation: $w_k^i = \tilde{w}_k^i / \sum_{j=1}^{N} \tilde{w}_k^j$. The particles are resampled according to the normalized weights to generate the new set of particles, denoted by $\{x_k^{i*}\}_{i=1}^{N}$. This set of particles is considered as the samples of posterior function $p(x_k^*|Y_k)$. Therefore, the aggregation of these weighted samples constructs the required pdf $p(x_k^*|Y_k)$:

$$p(x_k^*|Y_k) = \sum_{i=1}^{N} w_k^i \times \delta(x_k - x_k^i) = \sum_{i=1}^{N} (1/N) \times \delta(x_k - x_k^{i*})$$

This completes one iteration of the algorithm. This procedure is repeated when a new observation is received.

## IV. SMART BEAM-PARTICLE FILTER

Smart Beam Particle Filter (SBPF) is inspired from the Beam Search (BS) method [21], to explore the search space more actively, alleviate the impoverishment problem, and to escape from local optima traps. The SBPF epidemic calibrator solves a dynamic estimation problem to figure out the model's parameters that best describe the trend of the observed epidemic. As mentioned earlier, a PF uses a set of particles to estimate the posterior distribution of the state vector. The state vector for our model contains the unknown parameters that control the dynamics of the epidemic model (Table I).

The mechanism of our SBPF epidemic calibrator is shown in Fig 1. In the first cycle, the state vectors $\{x_0^{i*}\}_{i=1}^{N_2}$ are initialized by sampling from a prior distribution, where $N_2 >> N_1$. Initial particles are passed unchanged to the simulator, i.e., Epifast as the first prediction for particles. Epifast receives the particles and uses them as base parameters to run the simulation of the agent-based model and generates various epidemic outputs as well as the corresponding epidemic curves.

In the other cycles, the initialization step is replaced by the prediction process in which $N_1$ sampled particles are received from the previous cycle, and the new states of the particles are predicted based on their previous values and past observations $(p(x_k|Y_{k-1}))$. The SBPF framework expands and diffuses each old sample, $x_{k-1}$, to multiple new state vectors, $(x_k)$, in different directions. This process increases the number of particles from $N_1$ particles to $N_3$, where $N_3 >> N_1$, resulting in more diverse solutions in each round. The size and direction of perturbation are determined based on predictive decisions made in the smart state dynamics component. Smart state dynamics is discussed in detail in the section IV-B. New predicted particles are passed to Epifast to simulate the disease propagation and generate corresponding epidemic curves.

The next step of the particle filter is calculating the likelihood of observations given the state vectors. Epidemic curves that are generated by epidemic simulator are compared to the observed curve to calculate the likelihood of the observed data given the model's parameters. The details of the likelihood function are provided in Section IV-A.

In the update procedure, the likelihood score can be used as the weights of the particles after normalization. The SBPF performs the non-replacement Monte Carlo sampling to resample the particles based on their weights and selects $N_1$ particles among $N_2/N_3$ ones that are the most promising state-vectors to generate the better forecast. The resampled particles are treated as the (approximate) samples from the posterior distribution with density $p(x_k^*|Y_k)$. Therefore, the aggregation of these samples could be used to produce an empirical estimate of pdf $p(x_k^*|Y_k)$. At this stage, one cycle has been completed, and the results of this cycle are fed to the next cycle. Usually, the PF cycle is repeated whenever a new measurement (observation) is received. However, the SBPF repeats the cycle more than once for the same observed data to converge to a better solution.
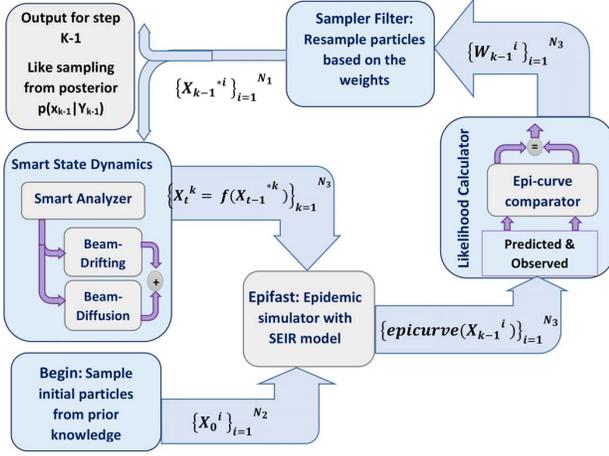
Fig. 1. Functional architecture of the smart beam particle filter. The width of arrows indicates the number of particles transferred between different units.

### A. SBPF Components: Likelihood Function

The likelihood function represents the possibility of observed epidemic data given the hidden parameters of the epidemic simulator. Although Epifast generates various outputs, the available surveillance observed data are usually limited to the time series of the weekly number of new infected cases. We use similarity/dissimilarity functions as the criteria of closeness between the observed and simulated curves for assessing the likelihood.

Examining various dissimilarity functions, we have selected wMAPE (window-MAPE) which is the MAPE function in which recent data points of the time-series are valued more than older ones. In other words, we calculate the likelihood based on the latest $\varpi$ recent data points of the time-series and discard the older ones:

$$wMAPE = 1/\varpi \sum_{i=T-\varpi}^{T} |y_i - I_i|/y_i$$

where $\{y_i\}_{i=1}^{T}$ denotes the target time-series of newly infected case counts until week $T$, $\{I_i\}_{i=1}^{T}$ represents the predicted case counts, and $\varpi$ denotes the width of the window. For comparison purposes, the similarity score is defined as the inverse of distance score. The scores are normalized over all particles and reported as the likelihood scores of the observed data given the state vector of parameters.

### B. SBPF Component: Smart State Dynamics

State dynamics models the changes in the state vector and predicts the value of the state vector in the next cycle based on its current value ($p(x_k|x_{k-1})$). We have designed a smart state dynamics utility that examines the predicted epidemic curves and the observed ones to determine the perturbation's size and direction for each parameter of the state vector. This has five basic components: ($i$) a *feature selector*, ($ii$) a *classifier*, ($iii$) a *smart director*, ($iv$) an *adaptive tuner*, and a ($v$) *diffuser*; see Fig 2. The *feature selector* picks different features from the observed and predicted epidemic curves

TABLE II
POSSIBLE CATEGORIES OF PREDICTED EPIDEMIC CURVE IN COMPARISON WITH THE OBSERVED CURVE. NEGLIGIBLE MEANS CLOSE TO OBSERVED DATA IN TERMS OF TIME OR CASE COUNTS.

| State Number | Amplitude Deviation | Time Deviation |
|---|---|---|
| 1-UL | Underestimate | Late prediction |
| 2-UE | Underestimate | Early prediction |
| 3-UN | Underestimate | Negligible |
| 4-OL | Overestimate | Late prediction |
| 5-OE | Overestimate | Early prediction |
| 6-ON | Overestimate | Negligible |
| 7-NL | Negligible | Late prediction |
| 8-NE | Negligible | Early prediction |
| 9-NN | Negligible | Negligible |

including ($a$) *peak value*, ($b$) *peak time*, ($c$) *length of season*, and ($d$) wwL1 error for the predicted curves. The rule-based *classifier* receives the selected features of both the predicted and observed epidemic curves and determines the state vector's category as summarized in Table II. The tree structure and rules of classifier is constructed based on the relative values of the predicted and observed features.

Knowing the class of each state vector (particle), the *smart director* decides about the drifting direction of each parameter. For example, if the classifier labels a predicted curve with Overestimate_Early (OE) tag, the smart director decides to change the epidemic parameters in ways that slow down the epidemic dynamics, like decreasing the transmission rate and/or infectious period. The *director* is trained based on the epidemiologist's opinion about the effect of each epidemic parameter on the trend of the pandemic. In other words, epidemiologist should simply specify the positive, negative, or neutral effect of parameters, that should get calibrated, on slowing down or speeding up the epidemic. Refer to Table I for the specified parameters. *Smart director* expands each particle into multiple ones by perturbing one or more parameters in the desired directions and generating new combinations of them. Another feature of the smart director is that it determines the level of certainty for the smart decisions. More uncertainty results in generating more random particles that are moved in erratic directions. Deterministic decisions are also combined with little randomness in a *random diffuser* to explore the search space and avoid local optima traps.

The *adaptive tuner* receives the history of the particle's evolution and determines the amount of perturbation (step size) for each parameter to help converging to the optimal solution. The inputs of the adaptive tuner are the *categories* and some *features* of both current and parent particles like perturbation directions, the distance of the current particle from the ideal solution (relative error), etc. When the relative error does not change significantly by perturbing a parameter, it implies that the step size of the parameter is too small. Adjusting the step size results in larger steps in flat areas of the search space to move faster toward good solutions. On the other hand, if perturbation direction of a parameter alternates frequently, it indicates that the step size is too large in that area of the search space. Therefore, adaptive tuner decreases the step size
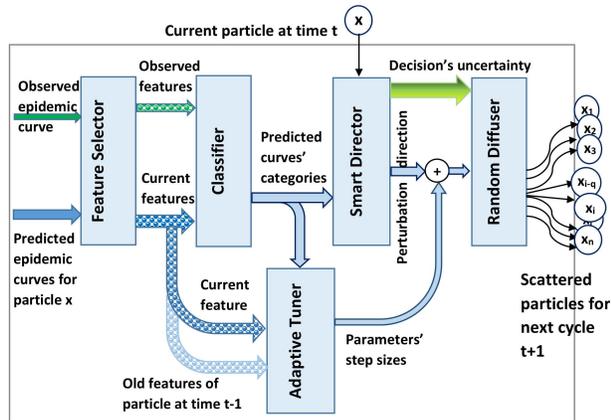
Fig. 2. Smart state dynamics framework. The feature selector chooses different features from the epidemic curves. Base on the selected features of both predicted and observed epidemic curves, the classifier determines the category to which the state vector (particle) belongs. The smart director makes the directional adjustment for each parameter. The adaptive tuner adjusts the step size for each parameter.

to avoid particle filter from fluctuating in near-optimal-solution space. The aforementioned smart analysis results in non-blind searches to achieve optimum values for the parameters in a faster way.

## V. EXPERIMENTAL RESULTS

In this section, we evaluate the accuracy and robustness of our agent-based model in simulating the propagation of Ebola as well as the performance of the proposed particle filter approach. It is intended to demonstrate how the SBPF algorithm can adjust multiple unknown parameters of the agent-based model for forecasting Ebola pandemics given independent data. The dataset we used for modeling and calibrating Ebola disease parameters was provided for the Ebola Challenge [9], organized under the Research and Policy for Infectious Disease Dynamics (RAPIDD) program at the NIH. This Ebola disease dataset was generated by a previously published agent-based model [22] calibrated with real data of 2014 Ebola epidemic for Liberia under different scenarios. Epidemic data for each scenario was released in five time points, and each time point contained outbreak situation reports, weekly reported new Ebola Virus Disease (EVD) cases at the county and country level, and current/future intervention plans for preventing and fighting the outbreak. New EVD cases were forecasted by each team at one, two, three and four weeks after each time point as the short-term prediction. We have used the country-level incidence time-series data and have predicted the short-term epidemic curve. Fig 3 represents the short-term predicted epidemic curves for the first and second time points of the first scenario. The weighted mean and weighted standard deviation of new cases of EVD are calculated based on the best weighted particles among all particles generated in the repetitive search iterations. Fig 3 demonstrates that the SBPF could perfectly calibrate the configuration parameters of the epidemic model such that the weighted-mean of predicted

curves matches the ground truth curve in the training part. The short-term prediction shows a little over-estimation, especially for the second time point. The over-estimation originates from incomplete information that is provided at each time point regarding future intervention plans. For example, in the second time point at week 20, there is no plan for contact tracing and isolating suspicious cases. However, in the next report for time point three, it is announced that 10292 individuals had been contact traced between week 20 and 26 to prevent the disease propagation. This shows the fidelity and consistency of our model with regard to the real-life intervention plans.
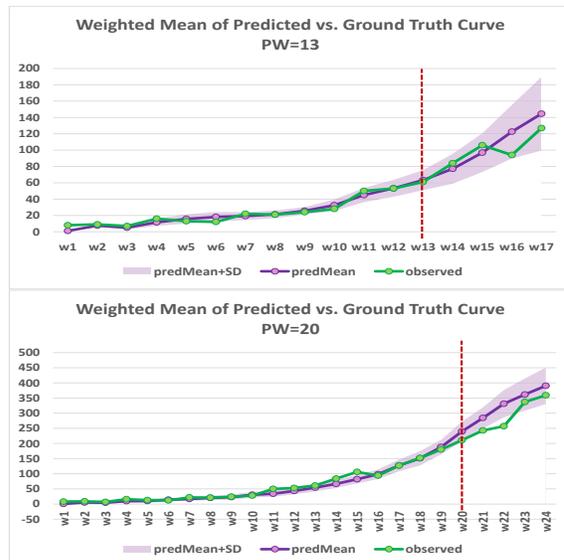


Fig. 3. Results for forecasting Ebola epidemic: The weighted mean of predicted versus ground truth curve - The weighted mean of best particles after 20 iterations of search. a) Demonstrates the short-term prediction for time point 1 (PW=13). b) Shows the prediction for time point 2 (PW=20).

We calculated several error measures [23] over the short-term predicted curves. The predefined set of error measures included: Pearson's correlation coefficient, mean absolute error (MAE), the mean absolute percentage error (MAPE), and root mean square error (RMS). We compared our results with the available output of other teams participated in the 2015 Ebola challenge. At the time of this writing, only one team had published the numerical results of their methods [24]. We compared the aforementioned error-measures with those provided by Pell et al. [24] for their two phenomenological models. Their first method, named Logistic Equation (LGM), consistently underestimated the epidemic curve and its negative value of Pearson's correlation coefficient demonstrates that the LGM's epidemic-curve does not follow the trend of the epidemic in the correct direction [24]. Pell et al. proposed another method, named Generalized Richards Model (GRM), that generates positive values for Pearson's correlation coefficient for most of the data points. Our SBPF approach demonstrates Pearson value of 0.80 and 0.94 for the first and second data points that shows agreement between the output and the trend of incidence data. SBPF generates lower MAPE

and RMS errors comparing both GRM and LGM approaches. (see Fig 4). The current results are achieved by running SBPF for only 20 iterations, $N_2 = 20$, $N_1 = 10$, and $N3 \approx 140$ as initial, resampled, and expanded particles, respectively.
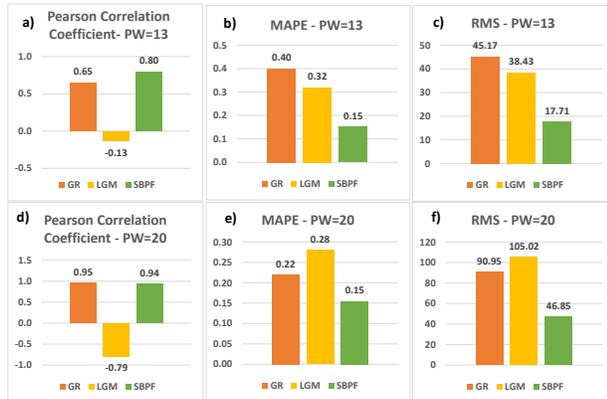


Fig. 4. Error measures on forecasting results of Ebola epidemic: Comparing the performance of SBPF with two other methods: GRM, LGM. Figures a, b, and c correspond to the first time point. Figures d, e, and f are associated with the second time point.

## VI. CONCLUSIONS

We present a data-driven causal modeling based methodology forecasting epidemic dynamics. It uses an agent-based modeling framework in conjunction with a modified particle filter approach that yields ensemble-based forecasts. The methodology was demonstrated on large regions represented by a realistic social contact network with millions of nodes/edges. The smart beam particle filter framework overcomes the computational burden of running agent-based models over large instances. The proposed particle filter approach implements a smart state dynamics unit that regulates the direction and perturbation of the particles. The smart director diffuses and scatters the particle in targeted directions to yield a non-blind search that swiftly achieves optimum values for parameters. Therefore, the smart particle filter approach is able to find near-optimal results with a lower number of particles and fewer running iterations.

## FUNDING

## REFERENCES

[1] N. A. M. Molinari, I. R. Ortega-Sanchez, M. L. Messonnier, W. W. Thompson, P. M. Wortley, E. Weintraub, and C. B. Bridges, "The annual impact of seasonal influenza in the us: Measuring disease burden and costs," *Vaccine*, vol. 25, no. 27, pp. 5086–5096, 2007.

[2] H. W. Hethcote, "The mathematics of infectious diseases," *SIAM Review*, vol. 42, no. 4, pp. 599–653, 2000.

[3] K. Bisset, J. Chen, and X. Feng, "Epifast: a fast algorithm for large scale realistic epidemic simulations on distributed memory systems," *23rd international conference on Supercomputing*, pp. 430–439, 2009.

[4] L. A. Meyers, "Contact network epidemiology: Bond percolation applied to infectious disease prediction and control," *Bulletin of the American Mathematical Society*, vol. 44, no. 1, pp. 63–86, 2007.

[5] C. Barrett, K. R. Bisset, S. G. Eubank, X. Feng, and M. Marathe, "Episimdemics: An efficient and scalable framework for simulating the spread of infectious disease on large social networks," *International Conference for High Performance Computing, Networking, Storage and Analysis (SC08)*, no. November, 2008.

[6] K. R. Bisset, J. Chen, S. Deodhar, X. Feng, Y. Ma, and M. V. Marathe, "Indemics : An interactive high-performance computing framework for data intensive epidemic modeling," *ACM Transactions on Modeling and Computer Simulation*, vol. 24, no. 1, pp. 4:1–4:32, 2014.

[7] J. Shaman, A. Karspeck, W. Yang, J. Tamerius, and M. Lipsitch, "Real-time influenza forecasts during the 2012-2013 season." *Nature communications*, vol. 4, p. 2837, 2013.

[8] W. Yang, A. Karspeck, and J. Shaman, "Comparison of filtering methods for the modeling and retrospective forecasting of influenza epidemics." *PLoS computational biology*, vol. 10, no. 4, p. e1003583, 2014.

[9] (2015, Aug.) Rapidd ebola challenge. [Online]. Available: http://ebola-challenge.org

[10] P. Loganathan, S. Ho Chee, L. Hao Ran, and S. Lakshminarayanan, "Towards forecasting flu dynamics using a regionalized state space model," *Advanced Control of Industrial Processes (ADCONIP), 2011 International Symposium on*, no. 1, pp. 175–180, 2011.

[11] J. B. S. Ong, M. I.-C. Chen, A. R. Cook, H. C. Lee, V. J. Lee, R. T. P. Lin, P. A. Tambyah, and L. G. Goh, "Real-time epidemic monitoring and forecasting of h1n1-2009 using influenza-like illness from general practice and family doctor clinics in singapore." *PloS one*, vol. 5, no. 4, p. e10036, 2010.

[12] V. Dukic, H. F. Lopes, and N. G. Polson, "Tracking epidemics with google flu trends data and a state-space seir model," *Journal of the American Statistical Association*, vol. 107, no. January, pp. 1410–1426, 2012.

[13] P. Dawson, R. Gailis, and A. Meehan, "Detecting disease outbreaks using a combined bayesian network and particle filter approach," *Journal of Theoretical Biology*, vol. 370, pp. 171–183, 2015.

[14] A. A. Koepke, I. M. Longini, M. E. Halloran, J. Wakefield, and V. N. Minin, "Predictive modeling of cholera outbreaks in bangladesh," *Annals of Applied Statistics*, vol. 10, no. 2, pp. 575–595, 2016.

[15] A. Skvortsov and B. Ristic, "Monitoring and prediction of an epidemic outbreak using syndromic observations," *Mathematical Biosciences*, vol. 240, no. 1, pp. 12–19, 2012.

[16] A. Skvortsov, B. Ristic, and C. Woodruff, "Predicting an epidemic based on syndromic surveillance," *2010 13th International Conference on Information Fusion*, pp. 1–8, 2010.

[17] N. Osgood and J. Liu, "Towards closed loop modeling: Evaluating the prospects for creating recurrently regrounded aggregate simulation models using particle filtering," in *Proceedings of the Winter Simulation Conference*, Savanah, GA, 2014, pp. 829–841.

[18] D. M. Sheinson, J. Niemi, and W. Meiring, "Comparison of the performance of particle filter algorithms applied to tracking of a disease epidemic," *Mathematical Biosciences*, vol. 255, no. 1, pp. 21–32, 2014.

[19] S. Eubank, H. Guclu, V. A. Kumar, M. V. Marathe, A. Srinivasan, Z. Toroczkai, and N. Wang, "Modelling disease outbreaks in realistic urban social networks," *Nature*, vol. 429, no. 6988, pp. 180–184, 2004.

[20] N. Gordon, D. Salmond, and a.F.M. Smith, "Novel approach to nonlinear/non-gaussian bayesian state estimation," *IEE Proceedings F Radar and Signal Processing*, vol. 140, no. 2, p. 107, 1993.

[21] S. R. Mousavi, F. Bahri, and F. S. Tabataba, "An enhanced beam search algorithm for the shortest common supersequence problem," *Eng. Appl. Artif. Intell.*, vol. 25, no. 3, pp. 457–467, Apr. 2012.

[22] S. Merler, M. Ajelli, L. Fumanelli, M. F. C. Gomes, A. P. y. Piontti, L. Rossi, D. L. Chao, I. M. Longini, M. E. Halloran, and A. Vespignani, "Spatiotemporal spread of the 2014 outbreak of ebola virus disease in liberia and the effectiveness of non-pharmaceutical interventions: A computational modelling analysis," *The Lancet Infectious Diseases*, vol. 15, no. 2, pp. 204–211, 2015.

[23] F. S. Tabataba, P. Chakraborty, N. Ramakrishnan, S. Venkatramanan, J. Chen, B. Lewis, and M. Marathe, "A framework for evaluating epidemic forecasts," *BMC Infectious Diseases*, vol. 17, no. 1, p. 345, 2017.

[24] B. Pell, Y. Kuang, C. Viboud, and G. Chowell, "Using phenomenological models for forecasting the 2015 ebola challenge." *Epidemics*, 2016.