



Systematic biases in disease forecasting – The role of behavior change

Ceyhun Eksin^{a,*}, Keith Paarporn^b, Joshua S. Weitz^{c,d,e}

^a Industrial and Systems Engineering Department, Texas A&M, College Station, TX, USA

^b Electrical and Computer Engineering Department, University of California, Santa Barbara, CA, USA

^c School of Biological Sciences, Georgia Institute of Technology, Atlanta, GA, USA

^d School of Physics, Georgia Institute of Technology, Atlanta, GA, USA

^e School of Electrical and Computer Engineering, Georgia Institute of Technology, Atlanta, GA, USA



ARTICLE INFO

Keywords:

Epidemiology
Disease forecasting
Nonlinear dynamics
Social distancing

ABSTRACT

In a simple susceptible-infected-recovered (SIR) model, the initial speed at which infected cases increase is indicative of the long-term trajectory of the outbreak. Yet during real-world outbreaks, individuals may modify their behavior and take preventative steps to reduce infection risk. As a consequence, the relationship between the initial rate of spread and the final case count may become tenuous. Here, we evaluate this hypothesis by comparing the dynamics arising from a simple SIR epidemic model with those from a modified SIR model in which individuals reduce contacts as a function of the current or cumulative number of cases. Dynamics with behavior change exhibit significantly reduced final case counts even though the initial speed of disease spread is nearly identical for both of the models. We show that this difference in final size projections depends critically in the behavior change of individuals. These results also provide a rationale for integrating behavior change into iterative forecast models. Hence, we propose to use a Kalman filter to update models with and without behavior change as part of iterative forecasts. When the ground truth outbreak includes behavior change, sequential predictions using a simple SIR model perform poorly despite repeated observations while predictions using the modified SIR model are able to correct for initial forecast errors. These findings highlight the value of incorporating behavior change into baseline epidemic and dynamic forecast models.

Infectious disease outbreaks catalyze public health responses. The scope of responses are increasingly tied to disease “forecasts”. Disease forecasts include estimates of infected cases, and associated mortality and morbidity (Yang et al., 2015). Accurate forecasts can help specify when, where, and how much response is needed to control an outbreak, e.g., whether for SARS, Ebola or Zika. However, inaccurate forecasts can lead to a mismatch between resources needed vs. those deployed. Inaccurate forecasts can also diminish public support for future interventions (Butler, 2014; Rivers, 2014).

Dynamic models underlie the science of disease forecasting (Keeling and Rohani, 2008; Anderson et al., 1992). The forecasting methodology involves fitting a dynamic model to available data and then extrapolating forward in time (and sometimes in space and in time) (Pastor-Satorras and Vespignani, 2001; Pei and Shaman, 2017; Ajelli et al., 2016). The extrapolation step depends on assumptions concerning subsequent disease transmission and recovery. These assumptions are informed, in part, by the initial fitting. Individual models may differ with respect to quantitative details, e.g., the probability of infection given contact, and even qualitative details, e.g., whether certain modes

of transmission are possible. Such differences can underlie variation in predictions, particularly for emerging outbreaks in which prior information on transmission may be limited (Biggerstaff et al., 2016; Viboud et al., 2017).

Despite their differences, leading model-based forecasting efforts often base their predictions on the assumption that individuals continue to behave in the same way irrespective of disease prevalence (Viboud et al., 2017). Given this assumption, it follows that the initial “speed” at which cases increase is informative with respect to the cumulative “size” of the outbreak. For example, if a disease spreads faster initially in outbreak A than in outbreak B, then more people will be infected in A who themselves continue to spread the disease at a faster rate than in B, all else being equal in the two outbreaks. In the long-term this should lead to a strong correlation between the initial speed and the final size (high in A, low in B).

However, the speed–size relationship may become tenuous when accounting for changes in individual behavior during an outbreak (Fenichel et al., 2011). Changes can include social distancing and other protective measures. For example, during the 2009 H1N1 pandemic,

* Corresponding author.

E-mail address: eksinc@tamu.edu (C. Eksin).

<https://doi.org/10.1016/j.epidem.2019.02.004>

Received 19 June 2018; Received in revised form 10 January 2019; Accepted 19 February 2019

Available online 28 February 2019

1755-4365/ © 2019 The Authors. 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/>).

59% to 67% of Americans washed their hands more frequently, more than 50% were prepared to stay at home if they or a member of their family were to get sick, and 25% avoided crowded areas (Steelfisher et al., 2010). As another example, decisions to forego traditional burial ceremonies during the Ebola virus disease (EVD) outbreak in West Africa reduced the rate of post-death transmission of EVD (Fang et al., 2016; Nielsen et al., 2015; Weitz and Dushoff, 2015). Such behavior change can significantly reduce disease transmission even in the absence of pharmaceutical interventions (Bauch and Galvani, 2013; Funk et al., 2010, 2014, 2015). Yet, the consequences of changing behavior are generally not accounted for in the science of disease forecasting. In this paper, we provide a model-based rationale for monitoring behavior change during the initial stages of an outbreak to avoid predictable pitfalls of forecasts.

1. Methods

1.1. SIR model with social distancing

We modify the standard SIR model to include a social distancing term $a(I, R)$ which is a function of the number of infected I and recovered R ,

$$\frac{dS}{dt} = -\beta S \frac{I}{N} a(I, R), \quad (1)$$

$$\frac{dI}{dt} = -\delta I + \beta S \frac{I}{N} a(I, R), \quad (2)$$

$$\frac{dR}{dt} = \delta I, \quad (3)$$

where β and δ are the disease infection and healing rates respectively, and $N = S + I + R$ is the population size. The social distancing term ($a(I, R): \mathbb{R}^2 \rightarrow [0, 1]$) scales the transition rate from a susceptible state to an infected state, that is given by $a\beta SI/N$ where β is the infection rate, S is the number of susceptible individuals, I is the number of infectious individuals and N is the population size.

We consider two types of social distancing models. In the first model, individuals reduce their interaction with others proportionally with the cumulative percentage of affected (infectious and recovered) individuals,

$$a = \left(1 - \frac{I + R}{N}\right)^k. \quad (4)$$

We term this “long-term awareness”.

In the second model, the reduction is proportional to the percentage of infectious individuals at a given moment,

$$a = \left(1 - \frac{I}{N}\right)^k. \quad (5)$$

We term this “short-term awareness”.

When the average infection in the population is high, the social distancing action is close to zero implying individuals are taking the utmost preemptive measures. In both distancing models, we raise the reduction term to the k th power for some behavior parameter $k \geq 0$. When $k = 0$, we recover the benchmark SIR model with *no distancing* response to disease prevalence. The distancing model is *linear* when $k = 1$. For $k > 1$, the social distancing model is *nonlinear*. As the nonlinearity (k) increases, individuals become more sensitive to disease prevalence (Funk et al., 2009; Yu et al., 2017).

In numerical simulations, we use outbreaks generated using the SIR model (1)–(3) with long-term (4) or short-term (5) awareness as ground truth. We will fix infection and healing rates to $\beta = 0.5$ and $\delta = 0.4$ respectively. Throughout the paper, the initial state of the synthetic outbreak will be $S_0/N = 99.5\%$, $I_0/N = 0.5\%$, and $R_0/N = 0\%$. We consider both linear and nonlinear distancing models. In the nonlinear distancing model, the behavior change parameter will be $k = 3$.

1.2. Sequential forecasting with ensemble-based Kalman filter

Sequential forecasting refers to repeated prediction and refinement of state and parameter estimates of the dynamic model based on new information collected about the ground truth. Denote the n dimensional vector of unknown state and parameters of a system at time t by x_t . For instance, in the SIR model for a fixed total population, we can have $x_t = (S_t/N, I_t/N, N, \beta, \delta, k)$ where S_t/N and I_t/N represent the fraction of susceptible and infected individuals at time t , respectively. Suppose we make noisy observations about the system's state at times $t = 1, 2, \dots$. The observation at time t denoted by y_t is an m dimensional vector. For instance, weekly noisy measurements may be available for the fraction of infected in the SIRS model, i.e., $y_t = I_t/N + \omega_t$ where ω_t is an additive measurement noise. Estimates of the latent states can be updated by using Bayes' rule, that is, $P(x_t | \{y_s\}_{s=1, \dots, t}) \propto P(x_t | \{y_s\}_{s=1, \dots, t-1}) P(y_t | x_t)$ where $P(x_t | \{y_s\}_{s=1, \dots, t-1})$ is the posterior distribution, is the prior distribution, and $P(y_t | x_t)$ is the likelihood of the observation given the prior state. The Bayes' update rule quickly becomes computationally intractable when the cardinality of the state increases for most prior and noise distributions and nonlinear system dynamics. The computational intractability of the Bayes' rule has led to development of filtering algorithms that approximate the posterior computation of the Bayes' rule such as particle filtering (Arulampalam et al., 2002), the simple Kalman filter (Kalman, 1960) or the ensemble-based Kalman filter (EnKF) (Evensen, 1994; Anderson, 2001).

The ensemble-based Kalman filter (EnKF) is a Monte-Carlo sampling based implementation of two steps: prediction and correction. Specifically, EnKF begins with M realizations of the state space x_t . We represent realization i of the n dimensional state at time $t - 1$ by the n dimensional vector $x_{t-1|t-1}^i$, and the ensemble vector array at time $t - 1$ by $x_{t-1|t-1} = \{x_{t-1|t-1}^i\}_{i=1}^M$. In the prediction step, we integrate each state realization $x_{t-1|t-1}^i$ using the model dynamics, e.g., SIR dynamics (1)–(3), to get the prediction $x_{t|t-1}^i$. The predicted ensemble mean and covariance are given by

$$\bar{x}_{t|t-1} = \frac{1}{M} \sum_i x_{t|t-1}^i, \quad (6)$$

$$\bar{\Sigma}_{t|t-1} = \frac{1}{M-1} \sum_i (x_{t|t-1}^i - \bar{x}_{t|t-1})(x_{t|t-1}^i - \bar{x}_{t|t-1})^T. \quad (7)$$

In the correction step, we use $\bar{x}_{t|t-1}$ and $\bar{\Sigma}_{t|t-1}$ to compute the mean and covariance estimates of the observation y_t , denoted by \hat{y}_t and \hat{O}_t respectively. For instance, if $x_t = (S_t/N, I_t/N, N, \beta, \delta, k)$ and $y_t = I_t/N + \omega_t$ where ω_t is additive white noise term, the second element of $\bar{x}_{t|t-1}$ is the observation prediction \hat{y}_t . We obtain the corrected ensemble by adjusting the ensemble predictions as follows

$$x_{t|t}^i = x_{t|t-1}^i + K_t (y_t - \hat{y}_t) \quad (8)$$

where K_t is an $n \times m$ gain matrix. The gain matrix K_t , also known as the Kalman gain, gauges the importance of the measurement relative to the prediction. Formally, it minimizes the mean square error of the estimate and can be solved by solving a set of linear equations—see Appendix B. The steps described above are repeated after each observation.

In the initialization of EnKF, we select the ensemble size M , and sample M values from a prior distribution of the state space. We denote the i th realization of the state space by $x_{0|0}^i$. Further, in our implementation of the EnKF, we add zero-mean noise to the predictions of the observation. That is, instead of a single prediction of the observation \hat{y}_t for the entire ensemble, we have a prediction for each ensemble member \hat{y}_t^i for all $i = 1, \dots, M$ —see Appendix B for details. Adding an artificial noise to our predictions prevents ensemble values from getting too close to each other after repeated correction steps (Roth et al., 2017).

1.3. Numerical implementation of the ensemble-based Kalman filter

We assume the initial fraction of susceptible and infected (S_0/N and I_0/N), and the behavior parameter k are unknown. In this paper, our focus is on comparing the accuracy of sequential forecasting with and without behavior change. Hence, we assume N , β and δ are known. Then the state space of EnKF is given by the 3×1 vector $x_t = (S_t/N, I_t/N, k)$. We will have $M = 20$ samples in an ensemble, unless explicitly stated otherwise. We denote the i th realization of the ensemble by $x_{010}^i = (1 - \hat{I}_0^i, \hat{I}_0^i, \hat{k}_0^i)$. Initial state realizations of the fraction of infected come from a Gaussian prior distribution with mean I_0/N and variance 0.05, that is, let $\hat{I}_0^i \sim \mathcal{N}(I_0/N, \sigma)$. If the prior value realized is negative, we discard it, and redraw from the Gaussian prior distribution until all realizations of \hat{I}_0^i are non-negative. We will specify the prior distribution on \hat{k}_0^i based on the specific scenarios considered in Section 2. We assume there are weekly noisy measurements of the fraction of infected, $y_t = I_t/N + w_t$, where $t = 1, 2, \dots$ weeks and w_t is a Gaussian random variable with zero mean and variance 10^{-3} that indicates measurement noise level. Given the prior ensemble vector $x_{t-1|t-1}$, we construct the predicted ensemble vector $x_{t|t-1}$ by integrating the SIR dynamics in (1)–(3) for 1 week. We obtain the mean measurement prediction by taking the mean of the second element of the ensemble members, that is, $\hat{y}_t = \sum_{i=1}^M \hat{I}_t^i/M$. We add prediction noise to \hat{y}_t during the update of each ensemble member in (8). We select prediction noise to be zero-mean Gaussian with a variance that we specify in Results section based on the scenario.

2. Results

2.1. The influence of behavior change on final outbreak size

The endemic level of the disease can be identified by finding the equilibrium of the model. We consider the long-term awareness model where distancing term is proportional to the fraction of affected. Assuming $R(0) = 0$ and $S(0) \approx N$, and noting that $R(\infty) = N - S(\infty)$ since $I(\infty) = 0$, we obtain the following infinite time relationship for susceptible individual population,

$$\frac{S(\infty)}{N} - \left[k \frac{\beta}{\delta} \left[1 - \frac{S(\infty)}{N} \right] + 1 \right]^{-1/k} = 0 \tag{9}$$

for $k > 0$ (see Appendix A for the derivation). Solving this equation when distancing is linear ($k = 1$), we have the fraction of recovered at equilibrium as $\frac{R(\infty)}{N} = 1 - \frac{1}{\mathcal{R}_0}$ where $\mathcal{R}_0 := \beta/\delta$ is the basic reproduction number. It is not possible to obtain a generic closed form solution for arbitrary k . In Fig. 1, we plot the equation for $k = \{1, 2, 3\}$, and compare the solution to this equation for $S(\infty)$ with the infinite time relationship of the SIR model with no distancing ($k = 0$) (Hethcote, 2000). As is evident, social distancing has the potential to significantly increase the fraction of uninfected individuals at the end of the epidemic relative to the SIR model without behavior change. We cannot provide a similar derivation of equilibrium points for the short-term awareness model, and turn instead to numerical simulations to measure asymptotic case counts for different models of short-term awareness.

2.2. The influence of behavior change on speed–size relationships

We define the speed of the disease outbreak in an SIR model with social distancing as the initial increase in infectious cases given a single infected individual in an otherwise susceptible population. Given the baseline dynamical systems, this speed is the infection rate (β) minus the healing rate (δ), $r = \beta - \delta$, for both social distancing models because initially the number of susceptible individuals is approximately equal to the population size ($S(0) \approx N$, $I(0) \approx 0$ and $R(0) = 0$). The speed r has units of inverse time and can be rewritten as where we recall $\mathcal{R}_0 = \beta/\delta$ is the basic reproduction number.

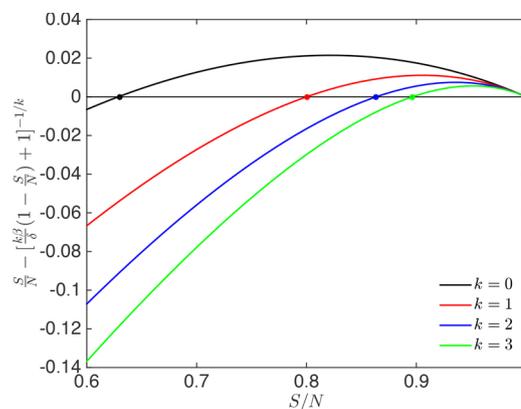


Fig. 1. Proportion of susceptible individuals at equilibrium given SIR dynamics with behavior change. We let $\mathcal{R}_0 = \beta/\delta = 1.25$. For $k = 0$, we use the well known infinite time relationship for $S(\infty)$ of the standard SIR dynamics (Hethcote, 2000) (see Appendix A). The non-trivial zero intersection of the SIR model with no distancing is at $S(\infty)/N = 0.63$ and it is the equilibrium point. For $k = \{1, 2, 3\}$, we plot the left hand side of infinite time relationship for $S(\infty)/N$. The non-trivial zero intersection of the function is at 0.8, 0.86, and 0.9 when k is equal to 1, 2, and 3, respectively.

To compare the effects of short- and long-term awareness of speed–size relationships, we simulate the SIR dynamics with social distancing, given $\mathcal{R}_0 = \beta/\delta = 1.25$. We have $k = 0$ for the benchmark model without social distancing. For both short- and long-term awareness, we let $k = 1$ and $k = 3$ for the linear and nonlinear social distancing models, respectively. We plot the percentage affected ($(I(t) + R(t))/N$) from the disease with respect to time in Fig. 2. Note that since the eventual average number of infected is zero ($I(\infty) = 0$), the eventual percentage affected is equal to $R(\infty)/N = 1 - S(\infty)/N$.

Fig. 2 demonstrates the potential pitfall of disease forecasting without accounting for changes in social behavior during an outbreak. The epidemic dynamics given distancing with long-term awareness is shown in Fig. 2 (left). The size of the epidemic is significantly lower in the social distancing model than in the benchmark model. Counter-intuitively, there is no early signal that would inform *a priori* forecast of such a reduction. The initial speed at which cases increase at the outset of the outbreak is nearly the same in both scenarios. That is to say, the number of cases increases as fast in the social distancing model as in the benchmark model. A forecast that leverages the initial speed to gain traction on disease parameters but assumes no behavior change would overestimate the final size by 85%.

The basis for this discrepancy in Fig. 2 (left) can be understood by recognizing that the fraction of infectious individuals is small at the early stages of an outbreak. In a linear social distancing model, such a fraction is too small to substantively change individual behavior. As the outbreak grows in size, the fraction of infectious individuals is large enough to influence individual behavior. When individuals decrease their infectious contacts, either via reducing interactions or by taking protective measures during risky interactions, the disease spreads slower than the benchmark model. Eventually, far fewer individuals are infected than would have been predicted via a standard extrapolation approach. In Fig. 2 (left) we also show that the over-estimation of spread can be much higher (250% in the example) when the individual response to disease prevalence is nonlinear.

When individuals respond to the prevalence of the infected, they respond to the current threat only. Fig. 2 (right) shows the resulting epidemic dynamics given distancing with short-term awareness. As is evident, the error in over-estimation is smaller (5%) when individuals distance with respect to the present threat only rather than to the present and past threat. Furthermore, over-estimation of epidemic size increases modestly (to 14%) when the social distancing response is nonlinear given short-term awareness.

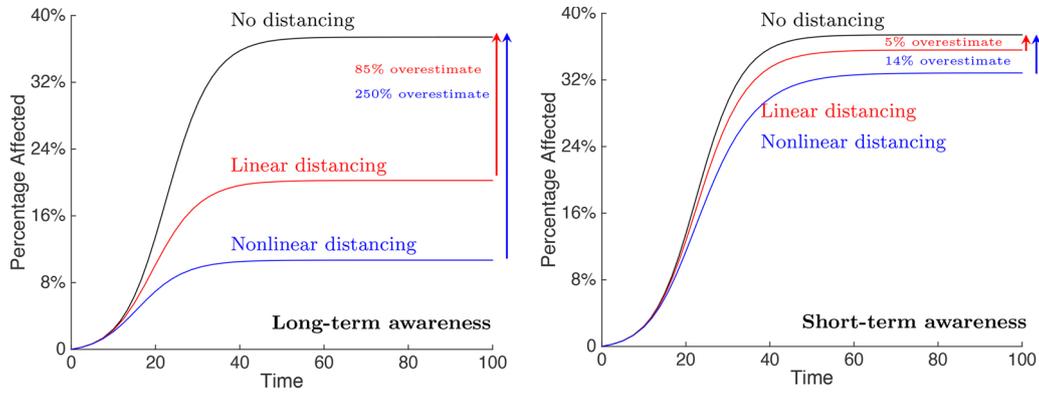


Fig. 2. Effect of social distancing on disease severity forecasts. We show the spread of disease for each model with respect to time given a basic reproduction number of $\mathcal{R}_0 = 1.25$. We consider two social distancing models including long-term and short-term awareness in the left and right panels, respectively. (left) The percentage affected when individuals distance with respect to total percentage of infected and recovered. (right) The percentage affected when individuals distance with respect to percentage of infected. At the beginning all models have comparable outbreak speeds. The simulation time horizon is 100. The benchmark model is the same in both figures given model parameters.

2.3. Sequential forecasting of a SIR outbreak with behavior modification

The prior results suggest that early forecasts that do not account for behavior change may overestimate final epidemic size. Further, these forecasts are prone to further errors given uncertainty in disease spread parameters and initial state of the disease. In practice, early forecasts are updated during an outbreak, given measurements of incidence and iterative model updating (Pei and Shaman, 2017). In this section, we use an ensemble-based Kalman filter (EnKF), described in the Methods, to analyze the expected error in sequential forecasts that leverage new information on disease spread given a baseline outbreak with behavior change.

We consider two sequential forecasting scenarios based on noisy measurements of the fraction of infected. In the first scenario, EnKF predictions do not account for behavior change making predictions based on SIR dynamics with the underlying behavior response $a = 1$. That is, the behavior parameter is assumed to be zero ($\hat{k}_t^i = 0$) and the state estimate by the i th ensemble member is given by $x_{t|t}^i = (\hat{S}_t^i, \hat{I}_t^i, 0)$ for all t where we recall $x_t = (S_t/N, I_t/N, k)$. In the second scenario, EnKF predictions treat behavior parameter k as a random variable. Specifically, state realizations of k come from a Gaussian prior distribution with mean 1 and variance equal to 0.5, i.e., we let $\hat{k}_0^i \sim \mathcal{N}(1, 0.5)$. The i th state realization is given by $x_{0|0}^i = (1 - \hat{I}_0^i, \hat{I}_0^i, \hat{k}_0^i)$. Then the behavior parameter estimate is updated after each observation.

We first generate an epidemic outbreak given long-term awareness (Fig. 3(a) and (b)). Then, we use the EnKF method to predict trajectories given weekly updates assuming an underlying model without (first scenario) and with behavior (second scenario) modification (Fig. 3(a) and (b), respectively). Both models start with inaccurate predicted trajectories where the initial mean trajectories overestimate the ground truth trajectory. Further, the predictions using the model with variability in k has a higher initial cone of uncertainty. The predictions using the model with variability quickly increase in accuracy, i.e., the mean predictions become close to ground truth and cone of uncertainty shrinks, whereas the model without behavior continues to have inaccurate mean predicted trajectories, and does not give indication that they are inaccurate, i.e., the cone of uncertainty disappears.

Next, we generate an epidemic outbreak given short-term awareness (Fig. 3(c) and (d)). Then, we use the EnKF method to predict trajectories given weekly updates for the first and second scenarios (Fig. 3(c) and (d), respectively). The predicted trajectories follow a similar trend as in the long-term case. That is, the model with behavior changes provide accurate forecasts starting from the second month while the model without behavior consistently overestimate the ground truth

trajectory. Here, forecast errors for the model without behavior changes are smaller because the short-term awareness model has a weaker influence on the ground truth trajectory than the long-term awareness model.

Fig. 4(a) and (b) shows the weekly final size predictions for the disease spread dynamics with long- and short-term awareness, respectively. The final size predictions using models with behavior eventually get close to the true final size value while the model without behavior fails to predict the final size correctly. We also compare the performance of the SIR model with behavior with a susceptible-exposed-infected-recovered (SEIR) model that has the same number of parameters but that does not have the same form as the ground truth model. We find that despite having additional flexibility, the SEIR model forecasts fail to approach true final size values (see Fig. S1 and Appendix C).

2.4. Sequential learning of behavioral response during an outbreak

The results above show that in both long-term and short-term awareness models, sequential forecasts with an underlying model that consider behavior change are able to predict the future trajectory of the disease after enough observations. Here, we investigate whether the accurate predictions are a consequence of learning the true value of the behavior parameter. Fig. 4(c) and (d) shows the time evolution of the estimate of the behavior parameter ($\hat{k}_{t|t} = \sum_i \hat{k}_{t|t}^i / M$) for the long-term (Fig. 3(b)) and short-term awareness (Fig. 3(d)) distancing scenarios, respectively. We observe that both the mean estimates $\hat{k}_{t|t}$ and the range of ensemble values for k approach the true behavior parameter value ($k = 3$). This shows that the sequential forecasts with EnKF are able to learn the behavior parameter, and hence accurately forecast the future trajectory. We note that if the ground truth is such that the behavior parameter is zero ($k = 0$), then the SIR model without behavior will perform best. However, the SIR model that treats behavior parameter k as unknown will eventually learn the value of k , including $k = 0$.

2.5. Robustness of sequential forecasting to noise

The prior information on the state (\hat{I}_0^i, \hat{k}_0^i) and observation noise level (ω_t) affect the performance of the EnKF. In particular as the prior mean of the behavior parameter \bar{k} (set to 1 in Fig. 3) gets close to the actual value of the behavior parameter ($k = 3$) forecast error improves (see Fig. S2 and Appendix D). As expected, we would like the prior variance (σ) of the behavior parameter to be small if the prior mean is close to the actual mean. Otherwise, a large prior variance is better for the EnKF because it allows for initial ensemble sample to have realizations close to the ground truth. If the prior mean is far from ground

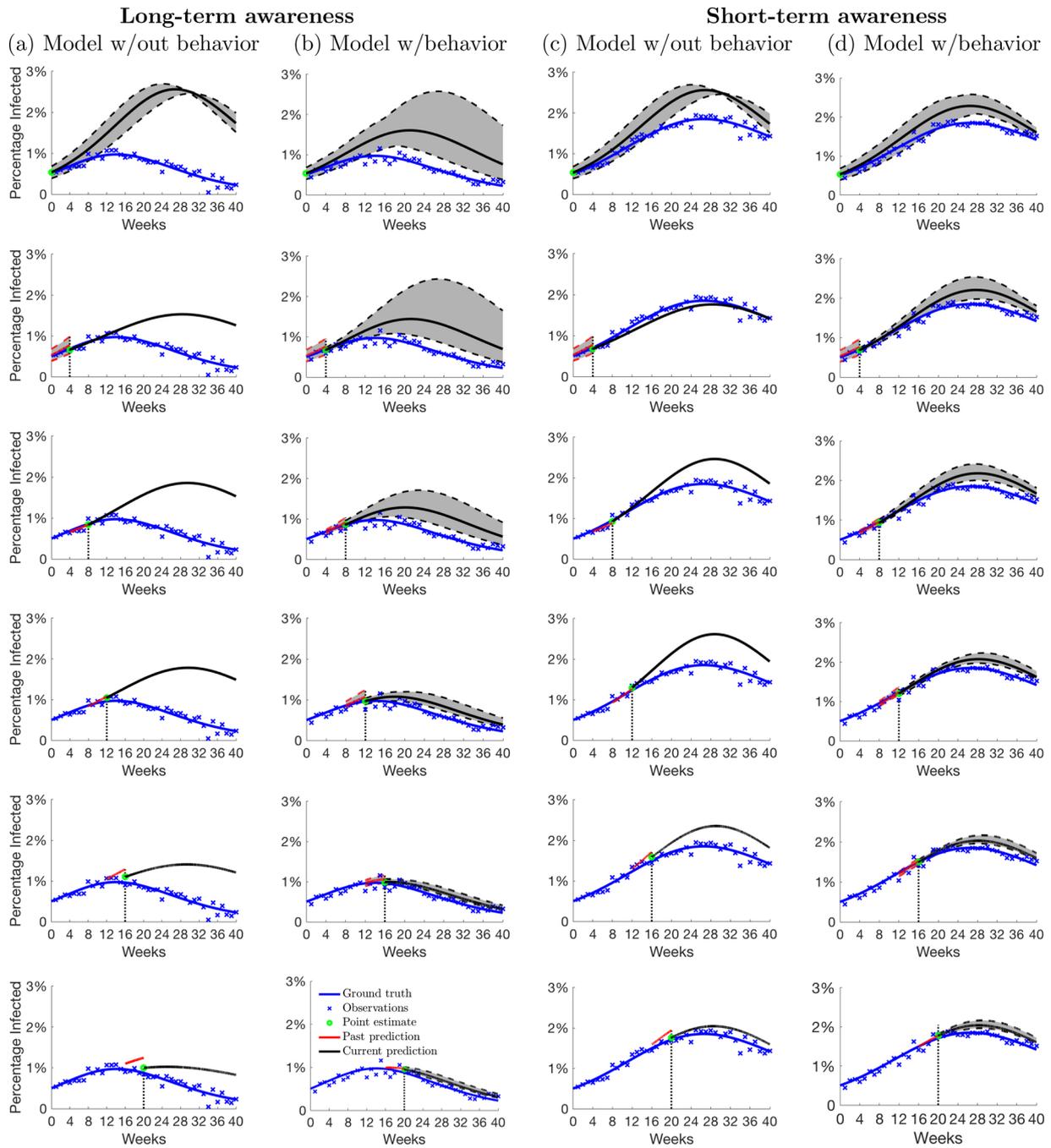


Fig. 3. Sequential prediction of a synthetic SIR outbreak with EnKF. The blue lines show the course of the synthetic outbreak. Ground truth outbreak is generated using SIR dynamics with nonlinear distancing ($k = 3$). In columns (a) and (b), distancing model is long-term awareness. In columns (c) and (d), distancing model is short-term awareness. We mark the measurements of fraction of infected with ‘x’. The green circles show the posterior mean estimates of the fraction of infected $\bar{x}_{t|t}$ for the month in consideration. (a) and (c) show monthly predictions based on an inaccurate SIR model ($k = 0$), i.e., assuming $a = 1$. (b) and (d) show monthly predictions based on an accurate SIR model that accounts for behavior change. For (a) and (c), prediction noise distribution is zero mean Gaussian with variance 10^{-10} . For (b) and (d), prediction noise distribution is zero mean Gaussian with variance 10^{-3} —see Appendix D for details on how to select the variance. In all figures, black lines show the predicted trajectories of the EnKF starting from the current week. Red lines show mean predictions of the previous month. Shaded areas show the range of the trajectories in the ensemble. EnKF predictions that do not account for behavior change ((a) and (c)) perform poorly in comparison to the EnKF predictions that do ((b) and (d)). (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

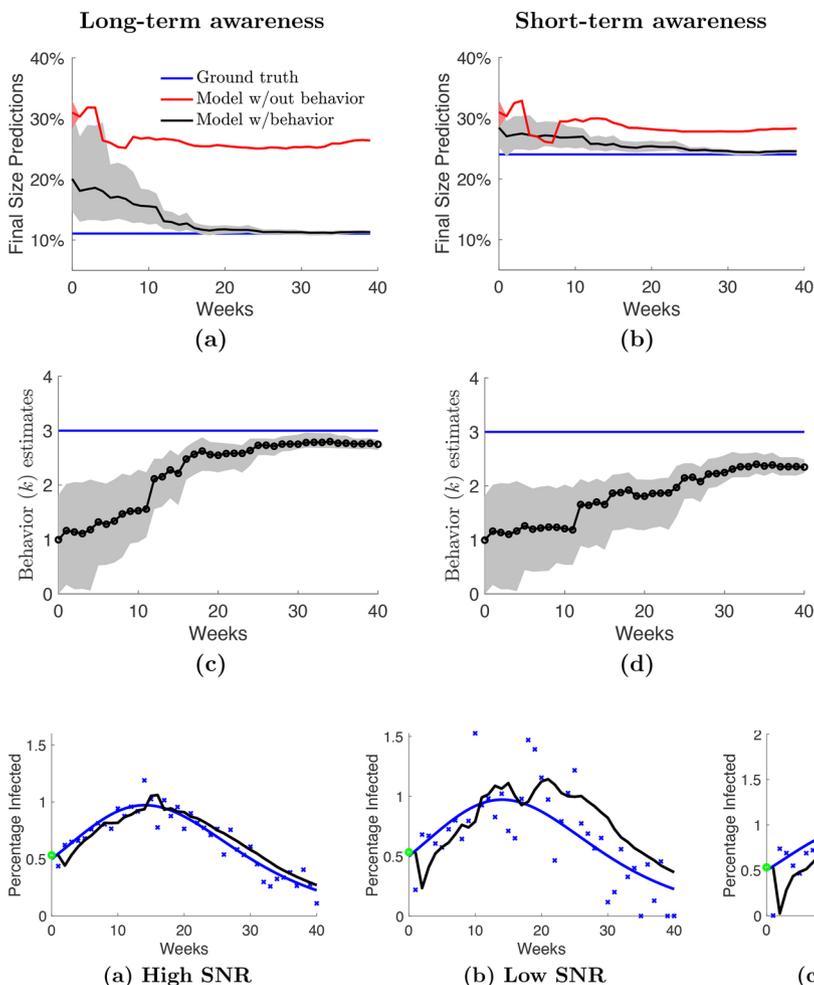


Fig. 4. Weekly final size (a and b) and behavior parameter (k) (c and d) estimates. Straight (blue) line shows the ground truth value of the behavior parameter $k = 3$. Left and right figures consider the long-term and short-term awareness models, respectively. The setups in long-term and short-term awareness models are as given in Fig. 3(b) and (d), respectively. The shaded regions show the range of values in the ensemble. While the EnKF forecasts that use SIR model with behavior converge to the actual final size, the forecasts using SIR model without behavior remain erroneous (a and b). The ensemble values for k converge toward each other as more measurements are collected and approach the ground truth $k = 3$ (c and d). (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

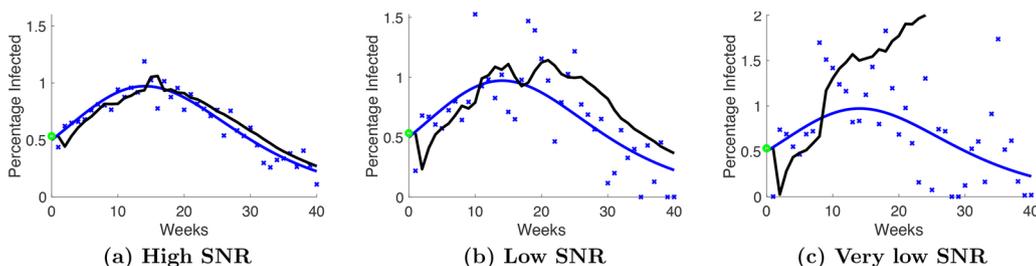


Fig. 5. Weekly infected ratio estimates for (a) high, (b) low, and (c) very low SNR for the same numerical setup as in Fig. 3(b) except we set the variance of the prediction noise to be 10^{-4} . Noise variance equals 0.1, 0.3, and 0.5, respectively for (a)–(c). The logarithm of the SNR values are given by 3.2, 1.6, and 1.2 for (a)–(c) (see Appendix D for details on SNR). When the SNR is very low, EnKF fails to track the disease progress.

truth and prior variance is small, then EnKF may fail to converge to the ground truth behavior parameter value. The other critical noise parameter is the observation noise (ω_o) which determines the signal-to-noise ratio (SNR) of the observations. In particular, when the observation noise level (variance of the observation noise) increases, EnKF performance declines. In particular, Fig. 5 shows that at very low SNR values, EnKF will not be able to track the disease progress. This failure can be avoided by appropriate selection of the EnKF parameters (ensemble size and prediction noise)—see Appendix D and Figs. S3 and S4 where we discuss the effects of design parameters.

3. Discussion

Our results point to a potential pitfall in the science of disease forecasting. Despite successes for respiratory diseases (Shaman et al., 2013; Shaman and Karspeck, 2012; Biggerstaff et al., 2016; Grenfell et al., 2002), the accuracy of disease forecasts has been questioned - particularly in light of the gap between the realized and predicted number of cases for the EVD outbreak in West Africa in 2014–2015 (Butler, 2014). On the one hand, disease forecasts can provide a road map for prevention of a bad outcome through a targeted intervention (Rivers, 2014). Nonetheless, the road map should include an accurate baseline of what might occur if no interventions were implemented (Funk et al., 2015). Ensuring that resources deployed include a buffer to achieve desired outcomes is essential. Prevention and control campaigns must also weigh the costs of over-allocation for resource management and logistics. The substantive differences in outcomes given social distancing based on long- vs. short-term awareness also present

an opportunity to improve predictions. We suggest there is a critical need to systematically study and collect behavioral information on the “stickiness” of factors – including present and past infections – that inform individual decisions to modify behavior during an outbreak.

Our results point to potential opportunities in accounting for behavior change during sequential forecasts. In particular, data assimilation methods can help us learn about behavior change during an outbreak based on seemingly unrelated data such as number of infected, or new incidences. Indeed, including behavior change in our forecasting models can prove valuable in making sense of new data. In contrast, neglecting behavior change, even when including additional complexity in disease progression, can result in inaccurate forecasts. Despite the promise, learning behavior-related parameters requires accurate measurements, particularly during early stages of an outbreak which are hard to collect in real-world outbreaks (see Figs. 3 and 4, and Appendix). In addition, the success of the EnKF method depends on selection of parameters which were pre-assigned given the synthetically generated outbreak data. In the case of real-world outbreak data, how to select the EnKF parameters and the observation model is an open research question (see Appendix D for additional details).

Here, we have provided theoretical and numerical evidence in support of reconsidering the accuracy of disease forecasts at the outset of an outbreak. We argued that comparative predictions of final sizes should not necessarily be predicated on estimates of initial speeds. Instead, accurate extrapolations from speed to size depend critically on changes in behavior. Via a sequential forecasting analysis, we showed that models that do not account for behavior do not perform well against a model that has the correct form of the behavior response with

some uncertainty about the behavior parameter. There are limitations to this approach, e.g., in this analysis we assumed the behavior parameter is unknown but fixed. In reality, behavioral responses to disease prevalence can change depending on communication campaigns, news, or other exogenous factors (Ajelli et al., 2017). Estimates and awareness of disease prevalence is also likely to be noisy. During real-world outbreaks, e.g., Ebola, underreporting, reporting delays and data collection will violate the assumptions of the observation model considered here. Moving forward it will be important to incorporate additional noise models, e.g., based on sampling variation and under-counting, in efforts to couple behavior change with sequential forecasting.

While our focus here was on individual behavior response to disease prevalence, individual behavior responses can take different forms during an outbreak, e.g., replacement of infected essential personnel with healthy ones (Scarpino et al., 2016), and they can depend on seemingly exogenous phenomena, e.g., communication campaigns, weather or co-infections (flu and pneumonia) (Hébert-Dufresne and Althouse, 2015). An interdisciplinary research effort is necessary to understand the coupling between intrinsic changes in behavior and public health efforts to modify individual behavior before, during, and after an outbreak.

Appendix A. The final outbreak size of the long-term awareness model

In the long-term awareness model we can simplify the distancing term as $a = \left(\frac{S}{N}\right)^k$. The ratio of change in susceptible individuals to the change in recovered individuals is given by

$$\frac{dS}{dR} = -\frac{\beta}{\delta} \left(\frac{S}{N}\right)^{k+1}. \tag{A.1}$$

For no social distancing ($k = 0$), we recover the well-known infinite time relationship for susceptible individuals (Hethcote, 2000):

$$\frac{S(\infty)}{N} - \exp\left[-\frac{\beta}{\delta} \left[1 - \frac{S(\infty)}{N}\right]\right] = 0, \tag{A.2}$$

where we simplified the result of integration by assuming $R(0) = 0$ and $S(0) \approx N$, and noting that $R(\infty) = N - S(\infty)$. There does not exist a closed form solution for $S(\infty)$ (Brauer, 2011). In Fig. 1, we plot the above function on the left hand side and show the critical $S(\infty)$ that equates it to zero (see case $k = 0$).

For $k > 0$ where individuals modify their behavior, we obtain the following equation by integrating (A.1) from time zero to infinity,

$$S^{-k}(\infty) - S^{-k}(0) = k \frac{\beta}{\delta N^{k+1}} [R(\infty) - R(0)]. \tag{A.3}$$

Assuming $R(0) = 0$ and $S(0) \approx N$, and noting that $R(\infty) = N - S(\infty)$ since $I(\infty) = 0$, we have

$$\frac{S(\infty)}{N} - \left[k \frac{\beta}{\delta} \left[1 - \frac{S(\infty)}{N}\right] + 1 \right]^{-1/k} = 0. \tag{A.4}$$

Solving the above equation when $k = 1$ gives the closed form solution that

$$\frac{S(\infty)}{N} = \frac{1}{\mathcal{R}_0}, \quad I(\infty) = 0, \quad \frac{R(\infty)}{N} = 1 - \frac{1}{\mathcal{R}_0} \tag{A.5}$$

where $\mathcal{R}_0 = \frac{\beta}{\delta}$.

Appendix B. Observation prediction and Kalman gain

Let the observation at time t , y_t , an $m \times 1$ vector be given by

$$y_t = Hx_t + \omega_t \tag{B.1}$$

where H is the $m \times n$ observation matrix and ω_t is the measurement noise vector of size $m \times 1$. The observation prediction of the i th member of the ensemble is given by

$$\hat{y}_t^i = Hx_{t|t-1}^i + \xi_t^i, \tag{B.2}$$

where ξ_t^i is the zero-mean prediction noise added to each ensemble prediction. We can compute the predicted mean and covariance of the observation as follows

$$\hat{y}_t = \frac{1}{M} \sum_{i=1}^M \hat{y}_t^i \tag{B.3}$$

Author contributions statement

CE, JSW, and KP designed research. CE implemented, and KP and JSW reviewed the code. CE and JSW wrote the manuscript. All authors reviewed the manuscript.

Competing interests

Authors declare no competing interests.

Materials

The code used in simulations is available online at https://github.com/ceyhuneksin/biases_disease_forecasting/tree/v1.0 with <https://doi.org/10.5281/zenodo.1281220>.

Acknowledgements

This work was supported by Army Research Office grant #W911NF-14-1-0402. The authors thank J. Dushoff and S. Scarpino for their suggestions and comments.

$$\hat{O}_t = \frac{1}{M-1} \sum_{i=1}^M (\hat{y}_t^i - \hat{y}_t)(\hat{y}_t^i - \hat{y}_t)^T, \tag{B.4}$$

where \hat{O}_t is an $m \times m$ matrix. Next, we compute the predicted cross-covariance between the state and observation, an $n \times m$ matrix, which is also known as the innovation covariance,

$$\hat{S}_t = \frac{1}{M-1} \sum_{i=1}^M (x_{t|t-1}^i - \bar{x}_{t|t-1})(\hat{y}_{t|t-1}^i - \hat{y}_t)^T = \bar{\Sigma}_{t|t-1} H^T \tag{B.5}$$

where $\bar{x}_{t|t-1}$ is defined in Eq. (6). The Kalman gain K_t , an $n \times m$ matrix, minimizes the variance by solving the following set of equations,

$$K_t \hat{O}_t = \hat{S}_t, \tag{B.6}$$

which means we can compute $K_t = \hat{S}_t \hat{O}_t^{-1}$.

Appendix C. Sequential forecasts using a SEIR model

We consider the SEIR model,

$$\frac{dS}{dt} = -\beta S \frac{I}{N}, \tag{C.1}$$

$$\frac{dE}{dt} = +\beta S \frac{I}{N} - \mu E, \tag{C.2}$$

$$\frac{dI}{dt} = -\delta I + \mu E, \tag{C.3}$$

$$\frac{dR}{dt} = \delta I, \tag{C.4}$$

where we have the exposed population (E) as the additional compartment. In this model, we assume S_t, E_t, I_t and μ are unknown. That is, the state space has cardinality four.

Fig. S1 shows the monthly EnKF forecasts using the SEIR model when the ground truth is generated using the SIR model with behavior ($k = 3$). In numerical experiments we use the same initializations for S_t and I_t as in the SIR models. We assume β and δ are known. The state realizations of μ come from a Gaussian distribution with mean 0.75 and variance 0.2. We assume the prior distribution of the fraction of exposed is identical to the fraction of infected.

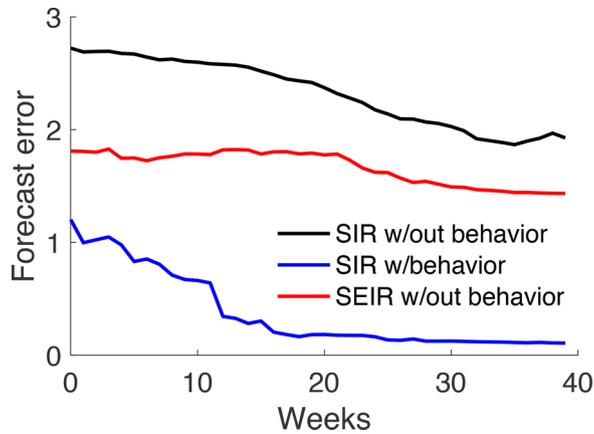


Fig. S1. Forecast error of SIR without behavior, SIR with behavior and SEIR without behavior. Forecast error is defined in Appendix D. Ground truth is the SIR model with long-term awareness $k = 3$. Simulation setup for SIR models is as in Fig. 3(a) and (b). Simulation setup for SEIR model is given in Appendix C. SEIR model reduces the forecast error in comparison to the SIR model without behavior because it is a more flexible model than the SIR model without behavior.

Appendix D. Effect of external and design parameters on sequential forecasts

The external parameters are the prior distributions of the state space, e.g., k and I_0 prior estimates, the distribution of the measurement noise (ω), and other model parameters. We base the analysis on the forecast error of the ensemble which we define as follows. For the estimated ensemble at time t , $\{x_{t|t}^i\}_{i=1}^M$, we construct trajectories $\hat{I}_{[0,T]|t}^i = (\hat{I}_{0|t}^i, \dots, \hat{I}_{T|t}^i)$ for the number of infected by integrating $x_{t|t}^i$ forward and backward. The mean trajectory of the ensemble is then the $\hat{I}_{[0,T]|t} = \sum_i \hat{I}_{[0,T]|t}^i / M$. The forecast error of the ensemble is the norm of the difference between the ensemble trajectory ($\hat{I}_{[0,T]|t}$) and the ground truth trajectory ($I_{[0,T]|t}$), i.e., $\|\hat{I}_{[0,T]|t} - I_{[0,T]|t}\|$.

Fig. S2 analyzes the effect of the mean (\bar{k}) and the variance (σ) of the prior distribution on the behavior parameter. Given the ground truth is $k = 3$, when \bar{k} changes from 2 to 3, the forecast error reduces more than a third. After the 20th week, forecast errors are small for all prior means. Fig. 5 shows that if the observation noise level is high (very low SNR), then EnKF may fail accurately track the disease trajectory. In particular, when we increase the variance of the observation noise from 0.1 to 0.5, we decrease the logarithm of the signal-to-noise ratio, $\text{SNR} := \frac{1}{T} \sum_{t=1}^T I_t / |y_t - I_t|$,

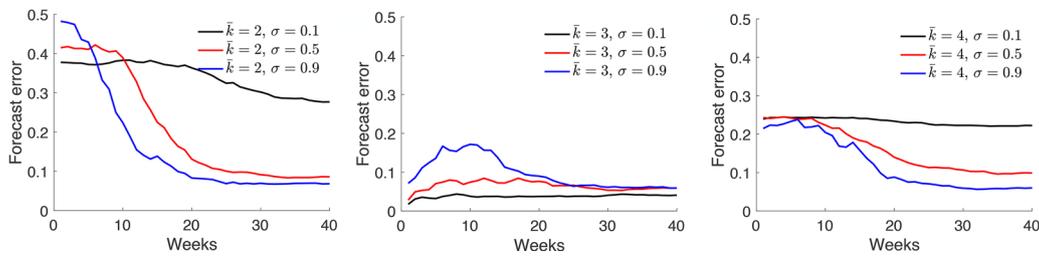


Fig. S2. Forecast error with respect to prior distribution on behavior parameter. The ground truth behavior parameter is $k = 3$. We consider the numerical setup in Fig. 3(b) with long-term awareness model. We sample $M = 20$ initial ensemble values of the behavior parameter \hat{k}_0^i according to a Gaussian distribution with mean \bar{k} and variance σ . Initial forecast errors show the value of the prior mean estimate. When the prior mean is correct, i.e., when $\bar{k} = k$, the initial forecast error is the smallest. Small variance implies high certainty on the mean estimate. When the prior mean estimate is not correct $\bar{k} \neq k$, small variance worsens performance—see left ($\bar{k} = 2$) and right ($\bar{k} = 4$) with $\sigma = 0.1$. When the prior mean estimate is correct $\bar{k} = k$, small variance improves performance—see center ($\bar{k} = 3$).

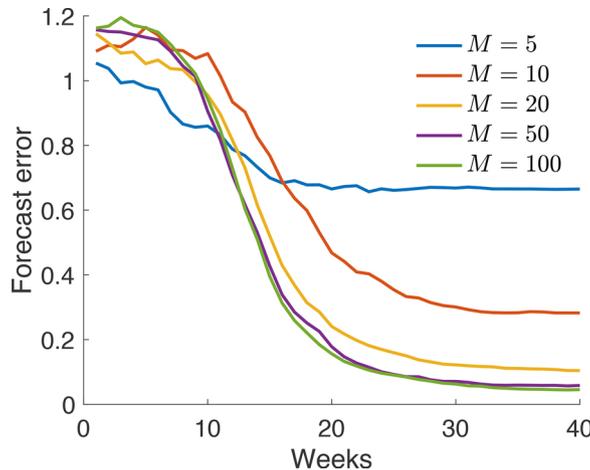


Fig. S3. Effect of ensemble size on forecast errors. We consider the numerical setup in Fig. 3(d). We measure the forecast error as defined in Appendix D. We observe diminishing improvement in forecast errors as we increase the ensemble size. As ensemble size (M) increases from 5 to 20, the accuracy of forecasts improve significantly. As ensemble size increases from 20 to 50 and 100, the improvement is less significant.

from 3.2 to 1.2.

We can tune the parameters of the EnKF in order to avoid divergence of the estimates. The EnKF parameters include the sample size M , the sampling distribution of the initial ensemble, and the distribution of the prediction noise. When the ensemble size M is small ($M \approx 10$), the sequential forecasting with EnKF performs poorly for high SNR scenario case (variance of the noise is equal to 0.1)—see Fig. S3. Numerical experiments show that in low SNR cases, we can lower the forecast error by increasing the ensemble size. In general, if the noise in the system or the cardinality of the state space increases, we need to make the ensemble size large to maintain a reasonable performance.

Finally, recall that we add artificial prediction noise to each sample to avoid the ensemble values getting too close to each other early. In the case very low SNR, setting the variance of the to 10^{-3} (was 10^{-4} in Fig. 5) makes sure that EnKF does not diverge for the given set of measurements. In general, higher variance for the prediction noise reduces the chance of divergence but may increase the forecast error. In our experiments (Fig. 3), we select the prediction noise variance value that minimizes the forecast error. The best prediction noise variance value depends on the model used. Fig. S4 (left) shows that for the SIR model without behavior, forecast error is minimized at the prediction noise variance value set around 10^{-10} . Fig. S4

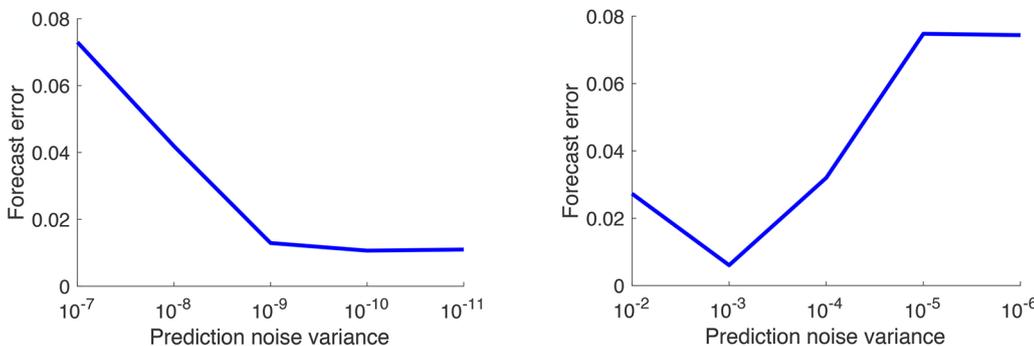


Fig. S4. Forecast error with respect to prediction noise variance. Prediction noise in (B.2) is distributed according to zero mean Gaussian with a given variance. For each variance value, we run the EnKF algorithm for multiple realizations of the epidemic. For each run, we compute the forecast error of the weekly estimates. The forecast error value shown here for each variance value is an average of the forecast errors of the multiple runs. (Left) EnKF predictions that do not account for behavior change—first scenario. (Right) EnKF predictions that account for behavior change—second scenario. The minimum forecast error is 10^{-10} and 10^{-3} respectively for Left and Right.

(right) shows that for the SIR model with behavior, forecast error is minimized at the prediction noise variance value set around 10^{-3} . In general, performance of the EnKF method is sensitive to the selection of prediction noise and the best value depends on measurement noise and other external parameters.

References

- Ajelli, M., Merler, S., Fumanelli, L., Piontti, A.P., Dean, N.E., Longini, I.M., et al., 2016. Spatiotemporal dynamics of the Ebola epidemic in Guinea and implications for vaccination and disease elimination: a computational modeling analysis. *BMC Med.* 14 (1), 130.
- Ajelli, M., Zhang, Q., Sun, K., Merler, S., Fumanelli, L., Chowell, G., et al., 2017. The RAPIDD Ebola forecasting challenge: model description and synthetic data generation. *Epidemics*.
- Anderson, R.M., May, R.M., Anderson, B., 1992. *Infectious Diseases of Humans: Dynamics and Control*, vol. 28. Oxford University Press.
- Anderson, J.L., 2001. An ensemble adjustment Kalman filter for data assimilation. *Mon. Weather Rev.* 129 (12), 2884–2903.
- Arulampalam, M.S., Maskell, S., Gordon, N., Clapp, T., 2002. A tutorial on particle filters for online nonlinear/non-Gaussian Bayesian tracking. *IEEE Trans. Signal Process.* 50 (2), 174–188.
- Bauch, C.T., Galvani, A.P., 2013. Epidemiology. Social factors in epidemiology. *Science (New York, NY)* 342 (6154), 47–49.
- Biggerstaff, M., Alper, D., Dredze, M., Fox, S., Fung, I.C.H., Hickmann, K.S., et al., 2016. Results from the Centers for Disease Control and Prevention's predict the 2013–2014 influenza season challenge. *BMC Infect. Dis.* 16 (1), 1.
- Brauer, F., 2011. A simple model for behaviour change in epidemics. *BMC Public Health* 11 (1), 1.
- Butler, D., 2014. Models overestimate Ebola cases. *Nature* 515 (7525), 18.
- Evensen, G., 1994. Sequential data assimilation with a nonlinear quasi-geostrophic model using Monte Carlo methods to forecast error statistics. *J. Geophys. Res. Oceans* 99 (C5), 10143–10162.
- Fang, L.Q., Yang, Y., Jiang, J.F., Yao, H.W., Kargbo, D., Li, X.L., et al., 2016. Transmission dynamics of Ebola virus disease and intervention effectiveness in Sierra Leone. *Proc. Natl. Acad. Sci.* 113 (16), 4488–4493.
- Fenichel, E.P., Castillo-Chavez, C., Ceddia, M.G., Chowell, G., Parra, P.A.G., Hickling, G.J., et al., 2011. Adaptive human behavior in epidemiological models. *Proc. Natl. Acad. Sci.* 108 (15), 6306–6311.
- Funk, S., Gilad, E., Watkins, C., Jansen, V.A., 2009. The spread of awareness and its impact on epidemic outbreaks. *Proc. Natl. Acad. Sci.* 106 (16), 6872–6877.
- Funk, S., Salathé, M., Jansen, V.A.A., 2010. Modeling the influence of human behaviour on the spread of infectious diseases: a review. *J. Royal Soc. Interface*.
- Funk, S., Knight, G.M., Jansen, V.A., 2014. Ebola: the power of behaviour change. *Nature* 515 (7528), 492.
- Funk, S., Bansal, S., Bauch, C.T., Eames, K.T., Edmunds, W.J., Galvani, A.P., et al., 2015. Nine challenges in incorporating the dynamics of behaviour in infectious diseases models. *Epidemics* 10, 21–25.
- Grenfell, B.T., Bjørnstad, O.N., Finkenstädt, B.F., 2002. Dynamics of measles epidemics: scaling noise, determinism, and predictability with the TSIR model. *Ecol. Monogr.* 72 (2), 185–202.
- Hethcote, H.W., 2000. The mathematics of infectious diseases. *SIAM Rev.* 42 (4), 599–653.
- Hébert-Dufresne, L., Althouse, B.M., 2015. Complex dynamics of synergistic coinfections on realistically clustered networks. *Proc. Natl. Acad. Sci.* 112 (33), 10551–10556.
- Kalman, R.E., 1960. A new approach to linear filtering and prediction problems. *J. Basic Eng.* 82 (1), 35–45.
- Keeling, M.J., Rohani, P., 2008. *Modeling Infectious Diseases in Humans and Animals*. Princeton University Press.
- Nielsen, C.F., Kidd, S., Sillah, A., Davis, E., Mermin, J., Kilmarx, P.H., et al., 2015. Improving burial practices and cemetery management during an Ebola virus disease epidemic – Sierra Leone, 2014. *MMWR Morb. Mortal. Wkly. Rep.* 64 (1), 20–27.
- Pastor-Satorras, R., Vespignani, A., 2001. Epidemic spreading in scale-free networks. *Phys. Rev. Lett.* 86 (14), 3200.
- Pei, S., Shaman, J., 2017. Counteracting structural errors in ensemble forecast of influenza outbreaks. *Nat. Commun.* 8 (1), 925.
- Rivers, C., 2014. Ebola: models do more than forecast. *Nature* 515 (7528), 492.
- Roth, M., Hendeby, G., Fritsche, C., Gustafsson, F., 2017. The ensemble Kalman filter: a signal processing perspective. *EURASIP J. Adv. Signal Process.* 2017 (1), 56.
- Scarpino, S.V., Allard, A., Hébert-Dufresne, L., 2016. The effect of a prudent adaptive behaviour on disease transmission. *Nat. Phys.*
- Shaman, J., Karspeck, A., 2012. Forecasting seasonal outbreaks of influenza. *Proc. Natl. Acad. Sci.* 109 (50), 20425–20430.
- Shaman, J., Karspeck, A., Yang, W., Tamerius, J., Lipsitch, M., 2013. Real-time influenza forecasts during the 2012–2013 season. *Nat. Commun.* 4.
- Steffens, G.K., Blendon, R.J., Bekheit, M.M., Lubell, K., 2010. The public's response to the 2009 H1N1 influenza pandemic. *N. Engl. J. Med.* 362 (22), e65.
- Viboud, C., Sun, K., Gaffey, R., Ajelli, M., Fumanelli, L., Merler, S., et al., 2017. The RAPIDD ebola forecasting challenge: synthesis and lessons learnt. *Epidemics*.
- Weitz, J.S., Dushoff, J., 2015. Modeling post-death transmission of Ebola: challenges for inference and opportunities for control. *Sci. Rep.* 5, 8751.
- Yang, W., Lipsitch, M., Shaman, J., 2015. Inference of seasonal and pandemic influenza transmission dynamics. *Proc. Natl. Acad. Sci.* 112 (9), 2723–2728.
- Yu, D., Lin, Q., Chiu, A.P., He, D., 2017. Effects of reactive social distancing on the 1918 influenza pandemic. *PLoS One* 12 (7), e0180545.