A TRANSFER ENTROPY BASED REGIONAL SIR MODEL

KELLIN RUMSEY

1. Introduction

Millions of people living in the United States are affected every year by the influenza virus, an infectious pathogen which causes respiratory illness resulting in hundreds of thousands of hospitalized individuals and significant economic impact. The Centers for Disease Control and Prevention (CDC), with the help of government and private healthcare professionals, closely monitors and reports influenza statistics. This data is critical for the adoption of flu prevention strategies such as vaccination campaigns, education programs and geographic allocation of PSA resources. Another important use area for this data is in *forecasting*, the prediction of illness into the (usually) near future. To a decision maker, this information is invaluable to improve the effectiveness of healthcare systems and to understand and prepare for potential economic burdens.

The benefits of forecasting are amplified if accurate forecasts can be provided with some local level of spatial resolution. Accurate forecasts at the city or state level, compared to a national scale, allow for policy makers to tailor prevention strategies, and evaluate public safety and economic risks in a more meaningful way. It is believed that spatial dynamics, especially at the city level, play an essential role in the transmission of infectious diseases, but these mechanics remain poorly understood in part due to the lack of fine-grain spatial data for infectious diseases. Geographic information for this type of data is usually collected at a higher levels, such as State or regional, and the spatial dynamics are largely hidden (Charu et al., 2017).

To understand how one might account for these dependencies, we first examine the wide variety of modeling choices which we can make. Agent based models are based on examining the dynamics at the level of the individual by simulating a population that mimics a real population. Often these are simple models based on elementary stochastic processes but they also include ultra-massive and computer intensive models such as the Chicago Social Interaction Model (chiSIM) which attempts to model dozens of daily interactions between millions of simulated Chicago residents. Agent based models have proven to be useful in certain situations, such as the evaluation of prevention and crisis recovery strategies (Macal et al., 2018). These models are often plagued by a large number of parameters leading to computational challenges, large uncertainty in the output and "unknown unknowns". Other models are based on *machine learning* algorithms which are capable of providing excellent predictions. The primary drawback of these models, is that they often ignore the obviously useful physics or biology of the problem. In some instances they also suffer from a lack of interpretability and computational burdens of their own (Osthus et al., 2017).

Mechanistic models, on the other hand, are driven by simple mathematical relationships usually specified as a system of differential equations. This class of models can also get arbitrarily complex, but many of the more popular models are both simple and intuitive. The SIR model, for instance, is a simple and effective tool which places biological constraints on the way in which individuals transfer from susceptible to infectious to recovered. These biologically inspired methods are often have low dimensional parameter-spaces and are easy to fit with modern computational methods. Although these models provide useful insight, they can be too simple to capture real-world dynamics and uncertainty in forecasts can be hard to quantify.

Finally, there are dynamic modeling approaches which are based on Time Series analysis in the field of statistics. Dynamic models are the gold standard of forecasting in fields such as finance, and are rarely beaten when the data is highly volatile. Parameter inference and forecast uncertainty are natural and flexible with this type of model, although it may come at a significant computational cost.

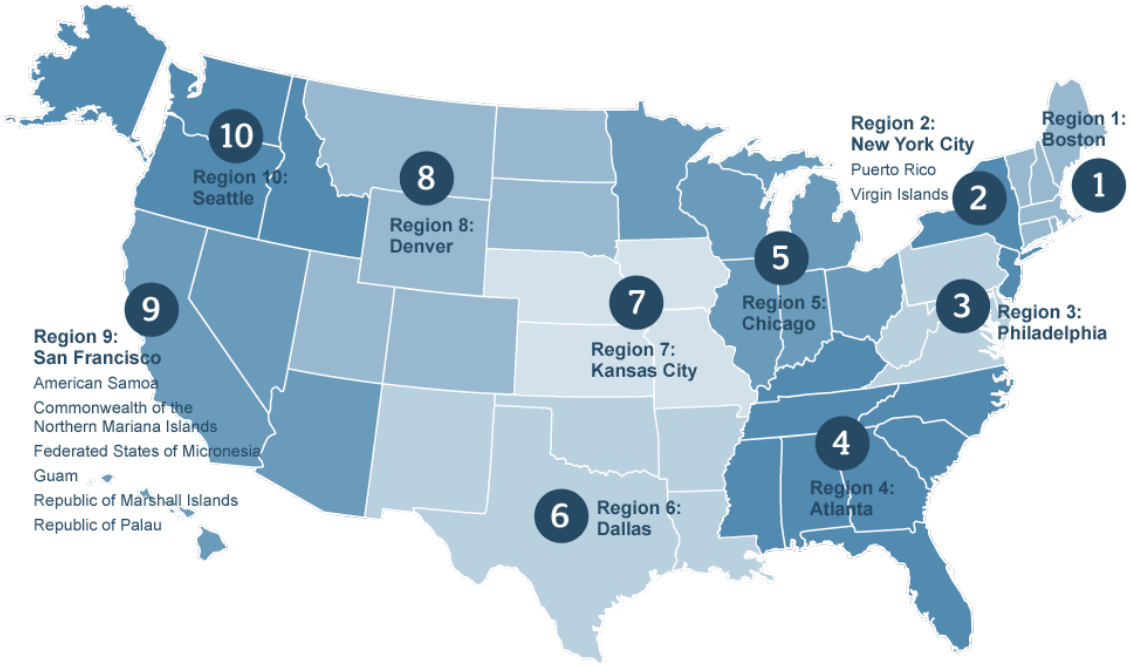

FIGURE 1. US Department of Health & Human Services Regions

Recent work at Los Alamos National Laboratories focuses on harnessing and combining the strengths of the latter two approaches for forecasting the CDCs influenza data (Osthus et al., 2019). A simple deterministic SIR model is used as a "first pass". Once the SIR dynamics have been accounted for, the remaining variability is referred to as *model discrepancy* (Kennedy & O'Hagan, 2001). Taking a Bayesian model calibration approach, this discrepancy is itself modeled using a reverse time random walk dynamic model. In aggregate, this model was shown to be extremely successful for influenza forecasting, outperforming all competition from the 2015-16 and 2016-17 CDC flu forecasting challenges.

A unique feature of this work, was the effortless inclusion of seasonal structure into the model via the dynamic process. The model discrepancy was separated into three components:

- discrepancy common to all flu seasons,
- season specific discrepancy
- and unexplainable uncertainty between the latent and observable process.

This structure is important because it allows for a particular flu-season to give information about *every* season, while still maintaining a flexible model which can accounts for annual differences between seasons.

Although the *Dynamic Bayesian* (DB) model discussed above is a state of the art forecasting model, it is executed exclusively at the national level, ignoring the regional granularity of the CDCs data. One straightforward extension of the DB model would be to partition the model discrepancy even further. This approach not only suffers from major identifiability issues, but it is also unclear how to achieve valuable spatial structure. In the remaining sections of this paper, we propose an extension to this model at the mechanistic stage which allows information to flow from one region to another, while maintaining the usefulness of the DB model. Section 2 introduces the ideas of *directional information flow* and a mathematical formalization of this idea known as *Transfer Entropy*. In section 3, we present the proposed model, emphasizing how transfer entropy can be incorporated directly into the mechanistic portion of the model. Section 4 discusses choices of prior distributions for a Bayesian implementation of the model and Section 5 illustrates the model behavior on the 2017 flu season.

2. Information Flow and Transfer Entropy

The CDC influenza data is collected through a surveillance network known as ILINet and is reported at the level of the HHS regions shown in Figure 1. This type of spatial data is known as *Areal data*, for which spatial structure is relayed through a *neighborhood*. For example, we might say that a region x is in the neighborhood of a region y if the two regions share a border. This implies that x and y are *first-order spatial neighbors*, but the general definition of a neighborhood can be far more flexible than this. It has been suggested that the spatial transmission of influenza is dominated by local traffic between cities (Charu et al., 2017). Unfortunately, the CDC data comes at a regional scale which is too coarse to capture the dynamics, and modeling approaches relying on first order neighborhoods provides little or no additional information.

Although this essentially means that the large geographic size of the regions limits the usefulness of standard spatial neighborhoods, it seems reasonable to assume that an increase in influenza activity in region x could provide information the upcoming influenza activity in region y. Thus we abandon the traditional notion of spatial neighborhoods and focus on building neighborhoods based on *information flow*. The concept of *Transfer Entropy* provides us with precisely the mathematical tool needed for this analysis (Schreiber, 2000). We start by considering two dynamic processes $X = (x_1, x_2, \cdots x_n)$ and $Y = (y_1, y_2, \cdots y_n)$. The transfer entropy from X to Y is an asymmetric measure of the amount of information *from* X to Y. Let $H(X|Y)$ denote the *conditional Shannon entropy* and let $X_{t,k}$ be the *lag k history* of X consisting of $x_{t-1}, x_{t-2}, \cdots x_{t-1}$.

Definition: *The transfer entropy from X to Y at time t with history length k is:*

$$T_{X\to Y}(t) = H(X_t|X_{t,k}) - H(X_t|X_{t,k}, Y_{t,k})$$

The transfer entropy from X to Y is taken to be the maximum over all time and history lengths.

Transfer entropy (TE) can be calculated using a "Kraskov estimator", which works be rewriting TE as an expansion of joint and conditional *mutual information* terms and applying a bias correction using a K-nearest neighbor algorithm (Moon 1995). The R package `RTransferEntropy` offers an efficient implementation of this estimator and a Bootstrap procedure which assesses the statistical significance of the resulting estimate. This allows us to define a *information flow neighborhood* (IFN) of a region y. We say that region x is inside the IFN of y if the hypothesis $T_{x\to y} \leq 0$ can be rejected for some predetermined significance level. Mathematically, we write $x \in N_y$ indicating that information flows from x to y.

For the 2017 CDC influenza data, we consider the time series y_j representing percentage of (weighted) influenza like illness for each region j. The transfer entropy between two regions is defined as $T_{i\to j} \equiv T_{y_i\to y_j}$. We find the information flow neighborhoods for each of the ten regions using $1,000$ bootstrap re-samples and a 10% significance level. The information flow and first order spatial neighborhoods are reported in Table 1.

Table 1. Information Flow and First Order Spatial Neighborhoods

HHS Region	IF-Neighborhood	First-Order Spatial
Region 1	3, 4, 6	2
Region 2	4	1, 3
Region 3	1, 2	2, 4, 5
Region 4	1, 3, 5, 7, 9, 10	3, 5, 6, 7
Region 5	1, 2, 3, 6, 7, 9	3, 4, 7, 8
Region 6	5, 7, 10	4, 7, 8, 9
Region 7	1, 2, 3, 4, 6, 9	4, 5, 6, 8
Region 8	3, 6, 7, 9	5, 6, 7, 9, 10
Region 9	1, 3, 5, 7, 8, 10	6, 8, 10
Region 10	1, 3, 6, 9	8, 9

3. The Model

We start by describing the mechanistic portion of the model. We first consider the standard SIR model, consisting of three continuous-time processes $S(t), I(t)$ and $R(t)$. For simplicity, we ignore demography and assume that $S(t) + I(t) + R(t) = 1$ for all $t > 0$. For modeling at the regional level, we assume that for each region $j = 1, 2, \cdots J$ we have

$$S_j(t) = \omega_j S(t), \qquad I_j(t) = \omega_j I(t), \qquad R_j(t) = \omega_j R(t)$$

where $\omega_j > 0$ and $\sum_{j=1}^{J} \omega_j = 1$. The main idea of this approach is write down SIR models at the regional level, allowing for regions to share information via the information flow neighborhoods discussed above. Let N_j be the IFN of region j, the set of all regions i which provide information about j.

$$\frac{dS_j}{dt} = -S_j\beta_j \left(I_j + \sum_{i \in N_j} \alpha_{ij} I_i \right)$$

$$\frac{dI_j}{dt} = S_j\beta_j \left(I_j + \sum_{i \in N_j} \alpha_{ij} I_i \right) - \gamma_j I_j$$

$$\frac{dR}{dt} = \sum_{j=1}^{J} \gamma_j I_j$$

We refer to this system of $2J + 1$ ODEs as a regional SIR with information flow (SIRwIF). As usual β_j represents a regional transmission rate and γ_j a regional recovery rate. A basic assumption of the SIR model is that the rate at which new infections occur is proportional to the product of the infectious and susceptible populations. The effective infectious population of region j is modeled as I_j plus the partial contribution of the region i infectious population, where $i \in N_j$ indicates that i gives valuable information about j. The $\alpha_{ij} \in c(0, 1)$ are *information contribution parameters*, describing the contribution of region i to the "effective" infectious population of region j.

To help motivate this model, imagine a country with two hypothetical regions x and y, such that information flows from x to y (but not necessarily from y to x). When region x experiences a significant increase in infections, there is a good chance that region y will soon exhibit similar dynamics. This model is particularly useful if it can be shown that the IF-Neighborhood structure is roughly time-dependent. This would allow us to estimate α parameters using past influenza seasons, without having to worry about time-evolution of these parameters. If these parameters are assumed to be static, then the SIRwIF model can fit easily into the BD framework of Osthus et al.

3.1. Specification of Full Model.

In order to use this model for forecasting influenza data at the regional level, we would like to specify a full probability model for the data. We borrow heavily from the DB model framework, making altercations to handle regional data when necessary and a few simplifications for convenience. We begin by making several modeling constraints. Without additional information, the initial conditions $S(0)$, $I(0)$ and $R(0)$ are inherently not jointly-identifiable, so we fix $S(0) = 0.9$ for all $t > 0$. Next, we fix the regional population sizes at each time t by treating the weights ω_j as fixed, known and proportional to the population of the region. Alternatively, an informative Dirichlet prior could be assigned the $\boldsymbol{\omega}$ vector, but we do not explore this here.

The free-parameters of the mechanistic model are now

- the global-level initial condition $I(0)$,
- the transmission parameters β_j,
- the recovery parameters γ_j
- and the information contribution parameters α_{ij} for all $i \in N_j$, $j = 1, 2, \cdots J$.

In the Bayesian framework, we must specify a prior distribution for each of these parameters. Specific choices, often based on *empirical Bayes*, will be discussed in the application section. For now we make comments for the general setting. One possible reduction of the parameter space involves fixing $\gamma_j = \gamma$ for all j. This assumption *seems* reasonable, but we find that this constraint removes too much flexibility from the model. We also note that there can be as many as $J(J-1)$ information parameters, which can be extremely challenging from a model fitting perspective. Fortunately, the matrix A of α_{ij} parameters is often quite sparse, especially when the transfer entropy significance level is small. In our application, for instance, 41 out of 90 entries of A are active at a 10% significance level, with active entries reducing to 33 and 14 for significance levels of 5% and 1% respectively. To account for false positives, we can regularize these parameters further by using priors with high density near zero, although a heavy right tail will still be desirable (Bhattacharya et al., 2015). Additionally, we normalize the *maximum total contribution* of region j's IF-neighbors by truncating α_{ij} above by the value $c|N_j|^{-1} \leq 1$ where c controls the proportion of region j's effective infectious population coming from the IF-neighbors. We set $c = 1$ for the rest of this paper.

As we discussed in the introduction, these simple mechanistic models are a useful starting point, but require extra mechanics in order to capture the volatility of the data. The observed data, denoted $y_j(t)$ at region j and time t, is interpreted as a proportion of the total population. Thus we must take care to respect the constraint $0 < y_j(t) < 1$ for all $t > 0$. Not all persons with influenza see a doctor and there are many illnesses with flu-like symptoms, therefore the true influenza statistics are an unobservable *latent process* $z_j(t)$. We model this relationship for all j and t as follows:

$$y_j(t) \overset{iid}{\sim} Be(\lambda z_j(t), \lambda(1 - z_j(t)))$$

so that $E(y_j(t)) = z_j(t)$ and $V(y_j(t)) = z_j(t)(1-z_j(t))/(\lambda+1)^{-1}$, setting $\lambda = 4500$ as in previous work. The next step accounts for model discrepancy, using a logit transformation to respect the proportion constraints on $y_j(t)$.

$$\log\left[\frac{z_j(t)}{1 - z_j(t)}\right] = \log\left[\frac{I_j(t)}{1 - I_j(t)}\right] + \delta_j(t)$$

where $\delta_j(t)$ is a region specific discrepancy term. Primarily as a complexity simplification, we choose to model the discrepancy using a *Gaussian process* with nugget τ_j as in Kennedy & O'Hagan, rather than the reverse time random walk of Osthus et al. This choice is not particularly important for our method since we do not propose using an additional discrepancy term within the DB framework. In particular, we specify that $\delta_j(t)$ is a Guassian process with mean function $\mu(t) = 0$ and covariance structure

$$\Sigma_j(\Delta t) = \phi_j R_j(\Delta t) + \tau I$$

where I is the identity and $R_j(t)$ is a Gaussian correlation matrix with range parameter κ_j.

$$R_j(\Delta t) = \exp(-\Delta t^2/\kappa_j)$$

. Therefore, in addition to the SIRwIF parameters $\theta_1 = (I(0), \boldsymbol{\beta}, \boldsymbol{\gamma}, \boldsymbol{\alpha}, \boldsymbol{\omega})$, there is a high-dimensional set of model hyper-parameters including $\theta_2 = (\lambda, \boldsymbol{\tau}, \boldsymbol{\phi}, \boldsymbol{\kappa})$. In the next section, we discuss specific choices and strategies for prior specification of these parameters.

4. Application to Influenza Data

The use of information flow neighborhoods allows us to model and forecast at the regional level, while borrowing strength from surrounding regions. The primary contribution is in the deterministic portion of the model, so that this model can be used in conjunction with the Dynamic Bayesian (DB) model which uses many influenza seasons to harness power from the data. Although this is the ideal, we focus on a single season (2017) for simplicity. The observed data $y_j(t)$ represents *weighted influenza like illness* (wILI), which is an estimate of the proportion of the population with influenza. The data is observed at the 10 HHS regions

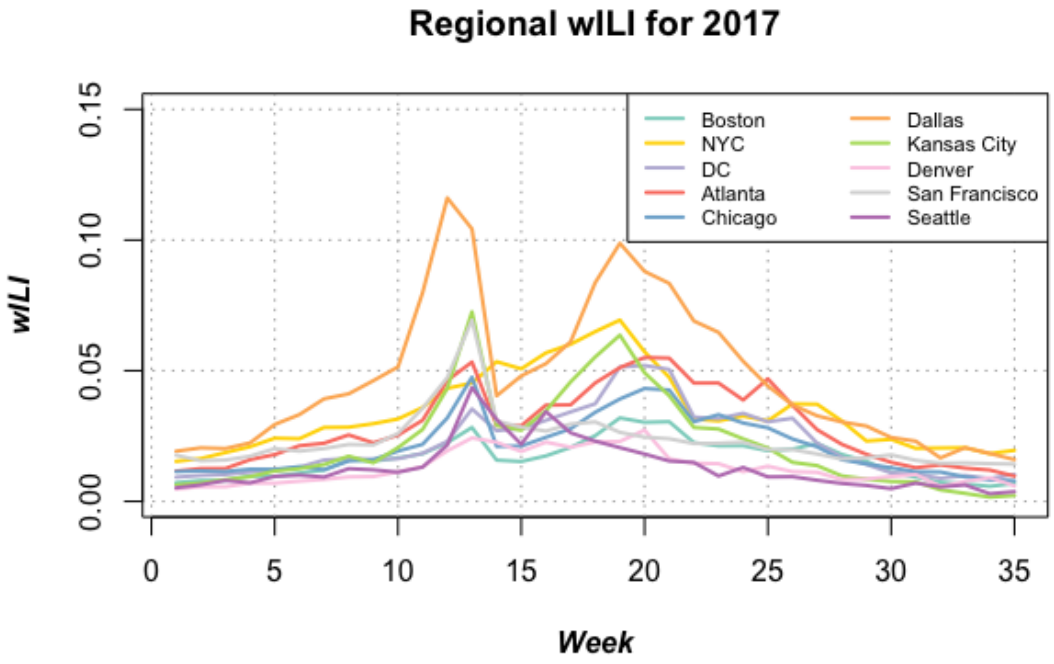

FIGURE 2. Weighted ILI for 10 regions throughout the 2017 season.

shown in Figure 1, for $t = 1, 2, \cdots 35$ weeks corresponding to roughly the first week of October through the last week of May. Figure 2 shows this data, using a large city in each region as a label.

4.1. **Bayesian Inference and Model Calibration.** To complete the specification of the full model, we must assign prior distributions to the SIRwIF parameters $\theta_1 = (I(0), \boldsymbol{\beta}, \boldsymbol{\gamma}, \boldsymbol{\alpha}, \boldsymbol{\omega})$ and the probability hyper-parameters $\theta_2 = (\lambda, \boldsymbol{\tau}, \boldsymbol{\phi}, \boldsymbol{\kappa})$. The full set of parameters is large compared to the number of observations so we will use empirical Bayes, MAP estimation and other heuristics to improve the identifiability of the problem (Brjynarsdottir & O'Hagan). We begin by fixing the values of $\boldsymbol{\omega}$ and λ as we discussed in the previous section.

Our empirical Bayes strategy begins by obtaining ordinary *least squares* estimates to the SIRwIF parameters by minimizing the equation

$$\sum_j \sum_t \left(\text{logit}\left(y_j(t)\right) - \text{logit}\left(\hat{I}_j(t, \theta_1)\right)\right)^2,$$

where $\hat{I}_j(t)$ is the solution to the SIRwIF system of differential equations. This constrained optimization problem involves searching a 62 dimensional space rendering it infeasible. Our solution is to fix $\alpha_{ij} = 0.2|N_j|^{-1}$, indicating that j's contribution to its effective infectious population is 5 times that of its IF-neighbors, and optimize over the remaining 21 parameters. The 10 estimated pairs $(\hat{\beta}_j, \hat{\gamma}_j)$ are used to fit a truncated bivariate Normal distribution, which we use as a joint prior for (β_j, γ_j). The truncation constraint is given by $0 < \gamma < 0.9\beta$ ensuring that the parameters are positive and the dynamics likely represent an epidemic. The constraint $\gamma < 0.9\beta$, which ensures an epidemic ($r_0 > 1$) for the SIR model, is not sufficient to guarantee an epidemic for the SIRwIF model for reasons we will illustrate shortly. The estimated $(\hat{\beta}_j, \hat{\gamma}_j)$ pairs (black) and draws from the fitted prior (gray) are shown in the left panel of Figure 3. To construct an empirical Bayes prior for the initial condition $I(0)$, we perform leave-one-out Jackknifing of the data to

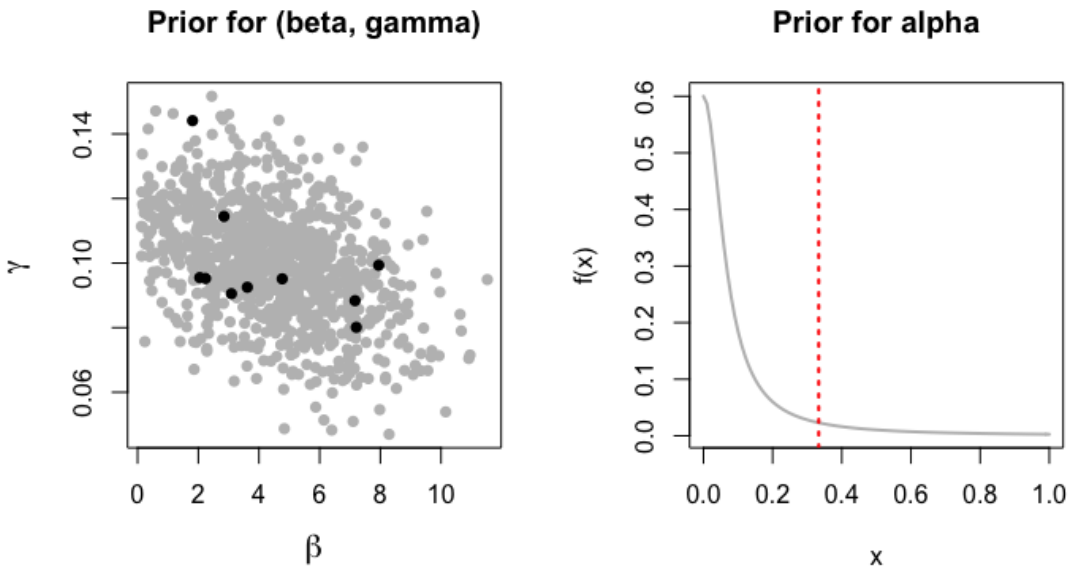

FIGURE 3. Left: Least squares fits to $(\hat{\beta}_j, \hat{\gamma}_j)$ (black) and draws from the fitted prior (gray). Right: The truncated half-Cauchy prior for the information contribution parameter α_{ij} when $|N_j| = 3$. Vertical line is upper truncation boundary for $c = 1$.

obtain a Normal approximation estimate of the $I(0)$ posterior under a flat prior. This approximation was chosen for the prior of $I(0)$ after doubling the estimated standard deviation and truncating at zero.

The information contribution parameters α_{ij} are constrained to a compact region, but the high dimensionality makes good prior specification important. Since the transfer entropy significance level was 10%, this is a large set with lots of potential for false positives suggesting the use of a sparsity-inducing prior. As shown in the right panel of Figure 3, we suggest the use of independent half-Cauchy priors with upper truncation at $c_1|N_j|^{-1}$ and a scale parameter of $c_2|N_j|^{-1}$. We choose $c_1 = 1$ and $c_2 = 0.2$ indicating maximum total contribution ratio of $1:1$ and a (pre-truncation) prior median contribution ratio of $5:1$ (section 3.1 for details). A simple alternative is a Beta$(\epsilon^{-1}, \epsilon)$ prior with truncation.

Next we consider the hyper-parameters $\theta_2 = (\boldsymbol{\tau}, \boldsymbol{\phi}, \boldsymbol{\kappa})$. To inform our prior distributions, we examine the logit scale residuals $e_j(t) = \left(\text{logit}\left(y_j(t)\right) - \text{logit}\left(\hat{I}_j(t, \hat{\theta}_1)\right)\right)$ which can be treated as the "observed discrepancy". Gaussian processes are fit to these residuals using maximum likelihood giving estimates of $\hat{\theta}_2$. In the left panel of Figure 4, we see the observed discrepancy for each of the 10 regions using the same color scheme as in Figure 2. The peak at week 13 is not being captured almost systematically, indicating that a global discrepancy function (as in DB) could be useful until at least week 16 when the discrepancy becomes more volatile.

Gaussian processes are poorly identifiable when a nugget is involved, so a manual tuning process was used to select $\tau_j = 0.02$ for all regions j. The Gaussian process fits under this constraint led to estimates $\hat{\phi}_j$ ranging from 0.024 to 0.290 and estimates of $\tilde{\kappa}_j$ ranging from 2.6 to 4.7 with an outlier at 10.2 for region 3. The range estimates $\tilde{\kappa}_j$ estimates were shrunk towards the median $(\hat{\kappa}_m)$ by setting $\hat{\kappa}_j = (\tilde{\kappa}_j + \hat{\kappa}_m)/2$ and fixed for the remainder of the analysis. Finally, the $\hat{\phi}_j$ parameters were treated with empirical Bayes, specifying

$$\hat{\phi}_j \sim \text{Gamma}(2, 2/\hat{\phi}_j - 2)$$

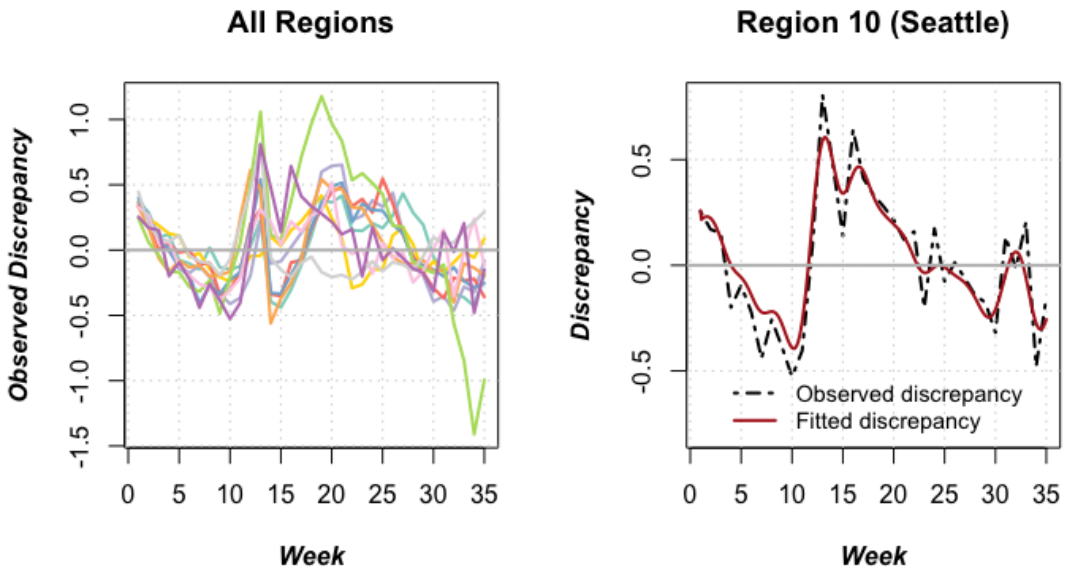

FIGURE 4. Left: Least squares fits to $(\hat{\beta}_j, \hat{\gamma}_j)$ (black) and draws from the fitted prior (gray). Right: The truncated half-Cauchy prior for the information contribution parameter α_{ij} when $|N_j| = 3$. Vertical line is upper truncation boundary for $c = 1$.

for each j, leading to prior expectation $\hat{\phi}_j$ and prior standard deviations between 0.0003 and 0.08. The right panel of Figure 4 shows the observed and estimated discrepancy functions for region 10. Notice that the second order residuals (i.e. $e_j(t) - \hat{\delta}_j(t)$) are mostly uncorrelated with a variance of $\tau = 0.02$.

The parameters were estimated using a $5,000$ iterations of a Gibbs sampling scheme with Metropolis-Hastings steps. Proposal distributions were based around a 2D random walk for (β_j, γ_j), log-random walk for $I(0)$ and $\boldsymbol{\phi}$ parameters and a Beta proposal for the $\boldsymbol{\alpha}$ parameters.

4.2. **Model Prediction.** Once the model has been fit, we can move on to different forms of inference using the full joint posterior distribution of the model parameters (θ_0, θ_1). The model can now be used to make predictions, obtain forecasts or explore general hypotheses at the regional level. For instance, we may want to estimate the basic reproductive number (r_0) for the influenza season in each region. We find r_0 analytically by examining the behavior of $\frac{dI_j}{dt}$.

$$\begin{aligned}
\frac{dI_j}{dt}|_{t=0} &= S_j(0)\beta_j \left(I_j + \sum_i \alpha_{ij} I_i \right) - \gamma_j I_j \\
&= \omega_j S(0)\beta_j \left(w_j + I(0) \sum_i \alpha_{ij}\omega_i \right) - \gamma_j \omega_j I(0) \\
&= S(0)\beta_j \left(w_j + \sum_i \alpha_{ij}\omega_i \right) - \gamma_j
\end{aligned}$$

The dynamics of the system will be epidemic as long as this quantity is positive. Defining the term in the parentheses as W_j, we see that the basic reproductive number for the SIRwIF model is

$$r_0 = \frac{S(0)\beta_j W_j}{\gamma_j}$$

TABLE 2. Basic reproductive numbers by region

	Est.(r_0)	95% LB	95% UB
Boston	2.354	1.869	3.044
NYC	3.219	3.047	3.454
DC	3.033	2.566	3.289
Atlanta	3.426	3.215	3.597
Chicago	1.545	1.102	1.925
Dallas	3.688	3.497	3.885
Kansas City	5.464	4.975	5.943
Denver	4.553	3.629	4.982
San Francisco	1.688	1.233	1.969
Seattle	4.272	3.937	4.606

We note that this looks very much like the basic regional SIR model with $S_j(0)$ getting replaced by $\tilde{S}_j(0) = W_j S(0)$, which can be interpreted as the *effective* susceptible population. Based on a sample of $1,000$ posterior draws, Table 2 gives estimated (posterior median) regional r_0 values along with a 95% credible interval. We see that regions 7, 8, and 10 showed the highest reproductive numbers for the 2017 season, while regions 5 and 9 showed the lowest with lower credible bounds just slightly above 1.

Next, we compare the dynamics of the SIR and SIRwIF models evaluated at their respective posterior means. Figure 5 shows the model fits at for two carefully selected regions. The flu season in region 6 (Dallas) got off to a fast start, with a peak season which was 1 to 2 weeks earlier than the other seasons. Region 1 (Boston) had a slow start to the flu season with a later peak wILI time. Notice how the SIRwIF model prevents over-correction, using information from other regions to pull the estimate back towards the mean. This leads to a 7% improvement in the sum of squares error in the first plot (left) and 18% improvement in the second (right). In many cases, estimates of the prediction uncertainty may be desired. Figure 6 gives 95% posterior prediction intervals for four regions with a variety of estimated reproductive numbers. These predictive intervals are comprehensive in their characterization of uncertainty, including

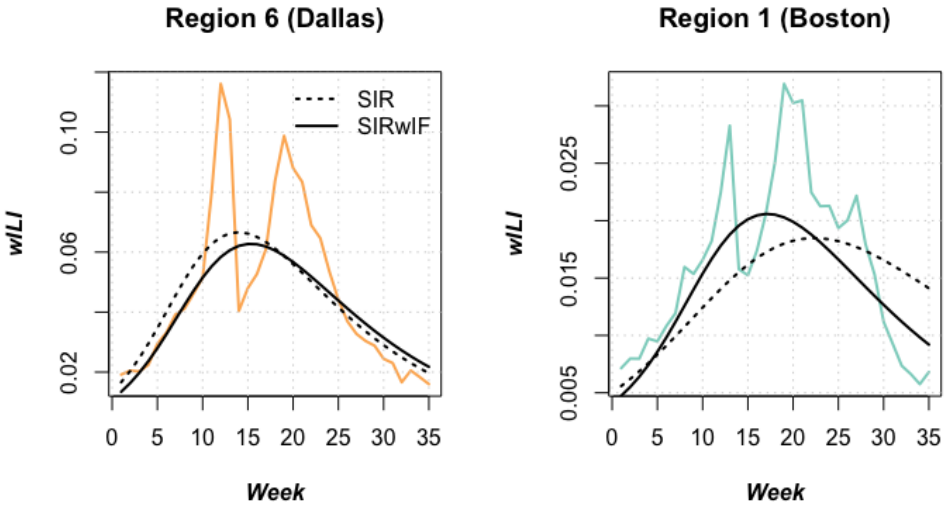

FIGURE 5. Comparison of SIR and SIRwIF mechanistic models evaluated at the posterior mean ($\hat{\theta}_1$) for regions with a fast (left) and slow (right) start to the flu season.

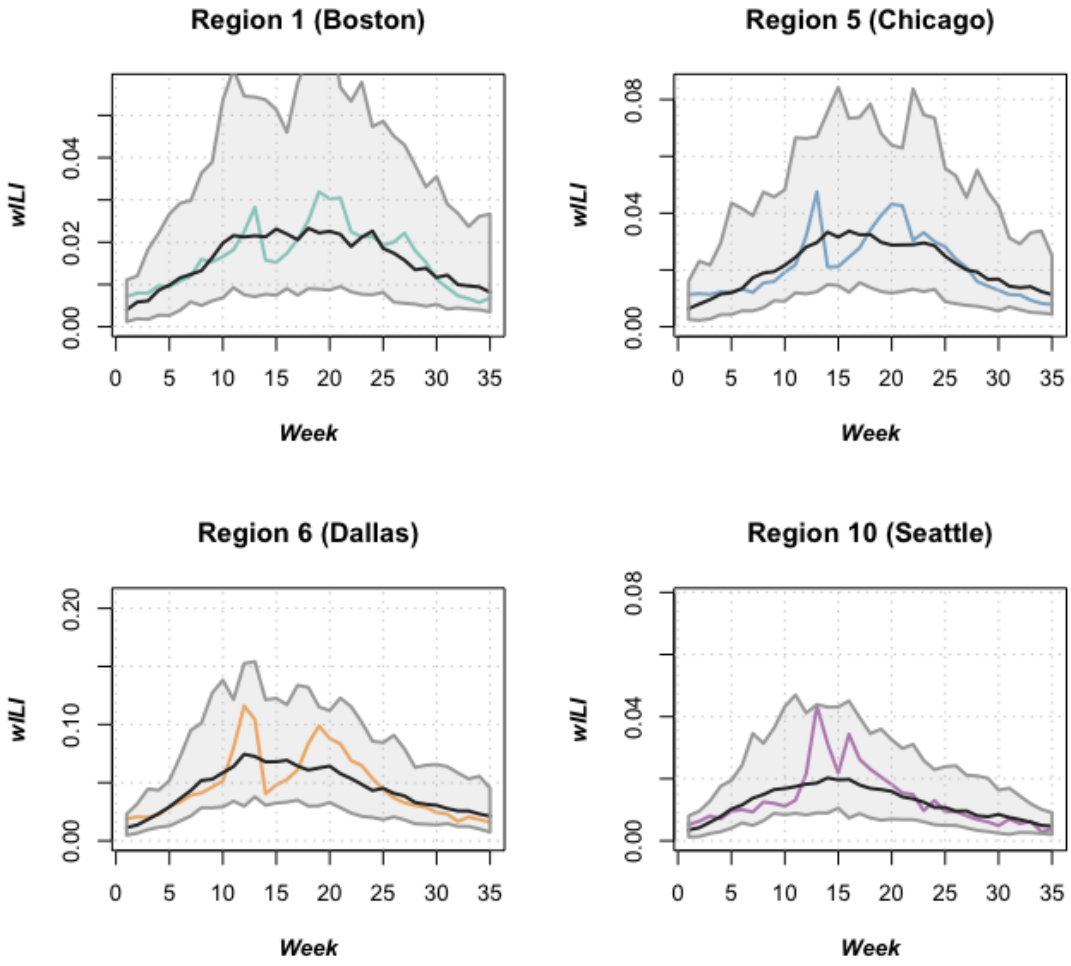

FIGURE 6. Posterior predictive distributions including 95% pointwise credible regions.

parameters, discrepancy and observation error. The black curve represents the pointwise posterior median of the prediction distribution, and gives an improved fit over the point-estimate procedure of Figure 5.

5. Conclusion

When modeling at the regional level, it can be useful to incorporate structural dependence between regions. When data is collected and reported with sufficient local resolution, the spatial structure can be harnessed. When the areal regions are coarse, such as with the CDCs regional influenza data, spatial analysis may be unable to uncover meaningful patterns. In such cases, we can generalize the idea of a neighborhood set to account for the flow of *information* between regions. Transfer entropy provides the mathematical framework necessary for the development of these neighborhoods. We propose the SIR model with information flow as a simple and efficient mechanistic procedure. which fits seamlessly with more comprehensive procedures such as the Dynamic Bayesian model of Osthus et al.

Future work would involve using the model in a forecast setting, conducting a thorough comparison with other existing models. A complete extension of the DB model to the regional level is also an interesting and worthwhile venture. Finally, we suggest further exploration of the prior specification and posterior sampling procedure as there is substantial room for improvement.

6. References

Bhattacharya, Anirban, et al. "Dirichlet?Laplace priors for optimal shrinkage." Journal of the American Statistical Association 110.512 (2015): 1479-1490.

Brynjarsdóttir, Jenný, and Anthony O?Hagan. "Learning about physical parameters: The importance of model discrepancy." Inverse Problems 30.11 (2014): 114007.

Charu, Vivek, et al. "Human mobility and the spatial transmission of influenza in the United States." PLoS computational biology 13.2 (2017): e1005382.

Kennedy, Marc C., and Anthony O'Hagan. "Bayesian calibration of computer models." Journal of the Royal Statistical Society: Series B (Statistical Methodology) 63.3 (2001): 425-464.

Macal, Charles M., et al. "chiSIM: an agent-based simulation model of social interactions in a large urban area." Proceedings of the 2018 Winter Simulation Conference. IEEE Press, 2018.

Moon, Young-Il, Balaji Rajagopalan, and Upmanu Lall. "Estimation of mutual information using kernel density estimators." Physical Review E 52.3 (1995): 2318.

Osthus, Dave, et al. "Forecasting seasonal influenza with a state-space SIR model." The annals of applied statistics 11.1 (2017): 202.

Osthus, Dave, et al. "Dynamic Bayesian Influenza Forecasting in the United States with Hierarchical Discrepancy (with Discussion)." Bayesian Analysis 14.1 (2019): 261-312.

Schreiber, Thomas. "Measuring information transfer." Physical review letters 85.2 (2000): 461.