STA 531 Final Paper

Eduardo Coronado

April 26th, 2019

Abstract

Early detection of the influenza outbreaks is one of the biggest challenges of outbreak surveillance systems. In this paper, a finite, homogeneous two-state Hidden Markov Model (HMM) was developed to determine the epidemic and non-epidemic dynamics influenza-like illnesses (ILI) in a differenced time series. These dynamics were further modeled via a first-order auto-regressive process based on the state, and parameters estimated via the Baum-Welch algorithm. The model was evaluated with US ILI data from 1998-2018.

Introduction

Every year influenza epidemics occur during winter months and result in a significant number of disease cases, deaths and public health expenses. According to the CDC, approximately 14-16% of the US population shows symptoms of influenza-like ilnesses (ILI) which causes 79,000 deaths each year. This has led public health officials to seek new ways of tracking potential outbreak situations via automated, routine influenza surveillance methods. This has traditionally been done via temporal models that have to be constantly refitted, which has practical implications, to quickly detect emerging outbreaks. Thus, it becomes especially challenging when using large datasets. Therefore, there exists a need for conceptually simpler, yet powerful methods, that can be easily automated for disease surveillance and that provide fast and accurate early detection of epidemic onsets. Therefore, the main goal of this paper is to implement a two-state Hidden Markov Model (HMM) for the early detection of influenza epidemics. HMM provides a natural way to way to model temporal data (i.e. epidemic and non-epidemic periods) by using a probability distribution for each state, and the Markov property that greatly simplifies the time dependency of the data and hidden states (Le Strat, 1). This model was tested using the US CDC's Influenza Surveillance network yearly national ILI incidence rates weighted by the states' population.

Previous Work

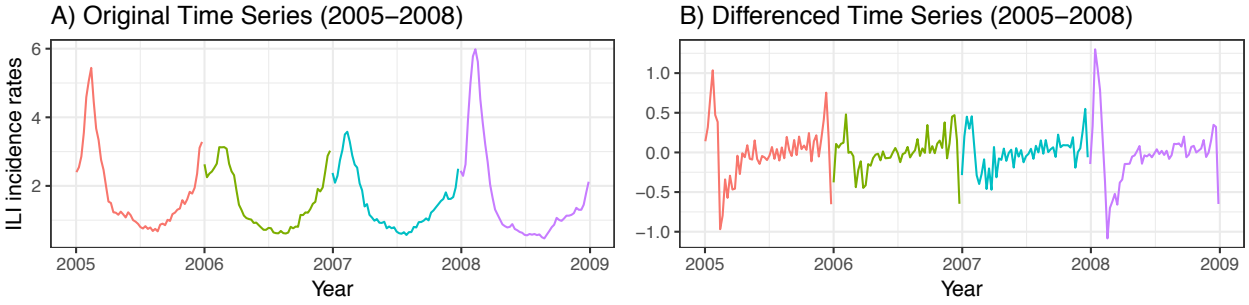
Over the last 20 years various HMMs have been proposed as temporal models for influenza and ILI surveillance methods. Le Strat and Carrat were the first to apply a two-state (epidemic and non-epidemic), homogenous HMM to ILI data with each state having a normally distributed cyclical pattern over a year (i.e. 52 week cycle) (1). Following this, others have proposed using similar HMMs but with different configurations, such as Rath *et al* which use a Poisson rather than a Gaussian model to account for potential non-zero probabilities for negative rates (2). More recently, Martínez-Beneito *et al* implemented an HMM using Bayesian Gaussian and Poisson hierarchical models on a differenced time series data (3; Conesa, 4). In 2012, Nunes *et al* further implemented a nowcasting model via a non-homogenous HMM to help shorten the time to detect early onsets of epidemics and better model the time dependency of these dynamics using Bayesian heirarchical structure (5).

Methods

Data

National ILI weekly incidence rates data were obtained from the US CDC's Influenza Surveillance network via the `cdcfluview` package for the years 1997-present. For simplicity, years 1997 and 2019 were excluded from the analysis given these didn't contain complete sets for the specified year (i.e. $T \neq 52$). Subsequently, data was preprocessed following Martínez-Beneito *et al*'s approach by generating a differenced time-series out of ILI rates (3). This procedure maintains the differences between epidemic versus non-epidemic dynamics as seen in Figure 1 below, with the epidemic states having higher variability, which helps simplify the model specification.

Fig1. Original vs Differenced ILI Time Series



HMM Model

The epidemic and non-epidemic trends in the differenced ILI data ($\mathbf{Y}$) were modeled via a finite, time-homogeneous two-state HMM based on Martínez-Beneito *et al*'s approach (3) as follows,

$$Y_{i,t} \mid Z_{i,t} = k, \psi \sim N(\mu_k, \sigma^2_{k,i}) \qquad \begin{aligned} &i = 1, .., N \quad \text{years with} \;\; N = 21 \\ &t = 1, .., T \quad \text{weeks with} \;\; T = 52 \\ &k = 0, K \quad \text{states with} \;\; K = 1 \end{aligned}$$

$$Z_{i,t} \sim Cat(a_{0,k}, a_{1,k}) \quad \text{transition probabilities } a_{0,k}, a_{1,k}$$

where $Z_{i,t}$ are the hidden components for epidemic ($k = 1$) and non-epidemic ($k = 0$) states, ψ are model parameters based on μ_k, and $a_{j,k}$ are the transition probabilities between the epidemic and non-epidemic states. As an attempt to account for the epidemic dynamics having a larger variational structure than non-epidemic ones, μ_k was set to follow an AR(1) process for epidemic states such that

$$\mu_k = \begin{cases} \rho \, Y_{i,t-1} & k = 1, \; t > 1 \\ 0 & k = 0, \; t \geq 1; \; k = 1, \; t = 1 \end{cases}$$

where ρ is the auto-regressive parameter in ψ. Similarly, $\sigma^2_{k,i}$ was set such that the variance for the epidemic state was larger than the non-epidemic one with each year being different $\sigma^2_{1,i} > \sigma^2_{0,i}$ Thus, the model is defined by the following complete data log-likelihood,

$$\ell(y, z; \psi) = \sum_{i=1}^{N} \left[log \; f(z_{1,i}) + \sum_{t=1}^{T-1} log \; p(z_{t+1,i} \mid z_{t,i}) + \sum_{t=1}^{T} log \; q(y_{t,i} \mid z_{t,i}, \psi) \right]$$

where $f(\cdot)$ specifies the initial state at $t = 1$, while $p(\cdot \mid \cdot)$ and $q(\cdot \mid \cdot)$ determine the transition and emission probabilities, respectively.

However, since the model parameters are unknown it is necessary to learn these from the data. This was achieved via the *Baum-Welch* algorithm, which leverages a type of Expectation-Maximization (EM) procedure for HMMs, on the following expected complete data log-likelihood expression (Bishop, 6)

$$
\begin{aligned}
Q(\psi, \psi^{(p+1)}) &= \sum_{i=1}^{N} \Big[\sum_{k=0}^{K} E[Z_{1,i}^{k}] \; log \; \pi_k + \sum_{t=1}^{T-1} \sum_{j,k=0}^{K} E[Z_{t,i}^{j} \; Z_{t+1,i}^{k}] log \; a_{jk} + \sum_{t=1}^{T} \sum_{k=0}^{K} E[Z_{t,i}^{k}] log \; N(\mu_k, \sigma_{k,i}^{2})] \Big] \\
&= \sum_{i=1}^{N} \Big[\sum_{k=0}^{K} \gamma_{k,i}(1) \; log \pi_k + \sum_{t=1}^{T-1} \sum_{k=0}^{K} \xi_{jk,i}(t) log \; a_{jk} + \sum_{t=1}^{T} \sum_{k=0}^{K} \gamma_{k,i}(t) \; log \; N(\mu_k, \sigma_{k,i}^{2}) \Big]
\end{aligned}
$$

where k is an indicator for $E[Z_{t,i}^{k}]$ being part of state k^{th} and

$$
\gamma_{k,i}(t) = E[Z_{t,i}^{k}] = p(Z_{t,i}^{k} = 1 \mid Y_{1:T,i}) \quad and \quad \xi_{jk,i}(t) = E[Z_{t,i}^{j} \; Z_{t+1,i}^{k}] = p(Z_{t,i}^{j} \; Z_{t+1,i}^{k} \mid Y_{1:T,i})
$$

HMM Training via Baum-Welch

E-Step - Normalized Forward-Backward algorithm

The expectations were computed via a normalized version of the forward-backward algorithm given numerical underflow is a common problem of the regular forward-backward procedure (Bishiop, 6; Zhai, 7). In summary, it normalizes the values at each step of the forward-backward procedure such that the $\sum_k \hat{\alpha}_{i,k}(t) = 1$ constraint is met in the forward step only. Thus, we obtain the following expressions for the E-Step

$$
\gamma_{k,i}(t) = \frac{\hat{\alpha}_{k,i}(t)\hat{\beta}_{k,i}(t)}{\sum_{l=0}^{K} \hat{\alpha}_{l,i}(t)\hat{\beta}_{l,i}(t)}
$$

$$
\xi_{jk,i}(t) = \frac{\gamma_{j,i}(t) \; a_{jk} \; q(y_{t+1,i} \mid z_{t+1,i} = k) \; \eta_{t+1,i} \; \hat{\beta}_{k,i}(t+1)}{\hat{\beta}_{j,i}(t)}
$$

where $\hat{\alpha}_{k,i}(t)$ and $\hat{\beta}_{k,i}(t)$ are the normalized α, β quantities, and $\eta_{t,i} = \frac{1}{\sum_{k=0}^{K} \alpha_{k,i}(t)}$ is the normalizing constant computed at each forward step (**Note:** this constant cancels out when calculating $\gamma_{k,i}(t)$)

M-Step

Subsequently, the updates were determined by finding the maximum likelihood estimates (MLEs) for ρ, π_k (initial states), a_{jk}, $\sigma_{0,i}^{2}$, and $\sigma_{1,i}^{2}$ via the arg max of $Q(\psi, \psi^{(p+1)})$ such that

$$
\pi_k^{(p+1)} = \frac{\sum_{i=1}^{N} \gamma_{k,i}(1)}{\sum_{i=0}^{N} \sum_{l=0}^{K} \gamma_{l,i}(1)} \qquad a_{jk}^{(p+1)} = \frac{\sum_{i=1}^{N} \sum_{t=1}^{T-1} \xi_{jk,i}(t)}{\sum_{i=1}^{N} \sum_{t=1}^{T-1} \sum_{l=0}^{K} \xi_{jl,i}(t)}
$$

$$
(\sigma_{0,i}^{2})^{(p+1)} = \frac{\sum_{t=1}^{T} \gamma_{0,i}(t) y_{i,t}^{2}}{\sum_{t=1}^{T} \gamma_{0,i}(t)} \qquad \rho^{(p+1)} = \frac{\sum_{i=1}^{N} \sum_{t=2}^{T} \gamma_{1,i}(t) \; y_{i,t} \; y_{i,t-1}}{\sum_{i=1}^{N} \sum_{t=2}^{T} y_{i,t-1}^{2}}
$$

$$
(\sigma_{1,i}^{2})^{(p+1)} = \frac{\sum_{t=2}^{T} \gamma_{1,i}(t)(y_{i,t} - \rho^{(p+1)} y_{i,t-1})^{2}}{\sum_{t=2}^{T} \gamma_{1,i}(t)} \qquad \left(\text{for } t = 1; \quad (\sigma_{1,i}^{2})^{(p+1)} = \frac{\gamma_{1,i}(1) y_{i,1}^{2}}{\gamma_{1,i}(1)} \right)
$$

Parameter Inference

Parameters were estimated via the Baum-Welch (BW) procedure noted above by initializing the parameters and iterating the E- and M-steps until convergence. However, given the EM algorithm optimizes for the local maxima/minima the parameters were initialized at different starting values to assess convergence (Bishop, 6). This process was repeated 7 times with two different sets of initial state probabilities each time

$$\text{Fixed Initial States} : \pi_0 = 0;\ \pi_1 = 1 \qquad \text{Random Initial States} : \pi_0 \sim Unif[0,1];\ \pi_1 \sim Unif[0,1]$$

and rest of the parameters started at different random points following Martínez-Beneito *et al*'s approach (1)

$$a_{0,0} \sim Beta(0.5, 0, 5) \quad \text{thus } a_{0,1} = 1 - a_{0,0}$$

$$a_{1,1} \sim Beta(0.5, 0, 5) \quad \text{thus } a_{1,0} = 1 - a_{1,1}$$

$$\rho \sim Unif(-1, 1)$$

$$\sigma_{0,i} \sim Unif[\theta_{low}, \theta_{mid1}] \qquad \sigma_{1,i} \sim Unif[\theta_{mid2}, \theta_{high}]$$

$$\theta_{low} \sim Unif[a, b] \qquad \theta_{mid1} \sim Unif[\theta_{low}, b]$$

$$\theta_{mid2} \sim Unif[\theta_{mid1}, b] \qquad \theta_{high} \sim Unif[\theta_{mid2}, b]$$

where $a = 1$ so that $\sigma_{0,i}$ wouldn't converge to 0, and $b = 5$ given the largest difference observed in the data was 4.5.

Decoding - Viterbi Algorithm

The averages of the learned parameters from the trained model were used to find the most likely path for each year i using the Viterbi algorithm, which implements recursive steps similar to the forward algorithm, on the same data (Bishop, 6). Here, the algorithm looked at the max over all previous steps ($v_{k,i}$) and stored the path indexes in a back-trace table ($b_{k,i}$) to easily identify the most probably path. The recursive steps were constructed via the following

$$v_{k,i}(t) = max_{k=1}^{N}\ v_{j,i}(t-1)\ a_{jk}\ q(Y_{t,i} \mid Z_{t,i} = k) \quad 1 \le k \le N, 1 < t \le T$$

$$b_{k,i}(t) = argmax_{k=1}^{N}\ v_{j,i}(t-1)\ a_{jk}\ q(Y_{t,i} \mid Z_{t,i} = k) \quad 1 \le k \le N, 1 < t \le T$$

and the most most probable path is found via $P = max_{k=1}^{N} v_{k,i}(T)$.

Parameter Inference and Decoding Comparisons

Consequently, comparative results for the parameter inference and decoding were generated via the JAGS model provided in Martínez-Beneito *et al* and the `depmixS4` R-package (3). The JAGS model was used to compare inference estimates and implemented as outlined in the paper with a slight modification - setting the standard deviation of hyperparameters to $a = 1$ and $b = 5$. Additionally, a Gelman-Rubin diagnostic was used to assess the convergence of the chains for each parameter (Bishop, 6). The `depmixS4` package was used to compare the decoding results. It was used to fit a HMM on the the same data with default settings (i.e. EM algorithm) and then to decode the most probable states based on depmixS4 trained model.

Results

Table 1 below shows the parameter inference results π_0, π_1, ρ, $a_{k,0}$, and $a_{k,1}$ for each of the 7 Baum-Welch iterations under the two states initial state conditions outlined above. Here we can see that on average the convergence of all parameters with a tolerance level of 10^{-8} is lower for the random initial states compare to the fixed ones (71 <93), with random states converging to higher probability of starting in a non-epidemic state π_0. Overall, the transition probabilites were similar except $a_{1,0}$ and $a_{1,1}$ with $a_{1,0}$ being higher and $a_{1,1}$ being lower the fixed initial state case. Similarly, *rho* was higher for the fixed initialized states. Similar tables were computed for $\sigma_{0,i}$ and $\sigma_{1,i}$ (tables not shown, see Suppl. File 1) with averages from these tables shown in Table 3.

Table 1: HMM Parameter Estimates via Baum-Welch

	Fixed Init Epidemic State								Random Init States							
T	iter	pi_0	pi_1	P0.0	P1.0	P0.1	P1.1	rho	iter	pi_0	pi_1	P0.0	P1.0	P0.1	P1.1	rho
1	97	0	1	0.977	0.226	0.023	0.774	0.193	73	0.597	0.403	0.978	0.156	0.022	0.844	0.175
2	97	0	1	0.977	0.226	0.023	0.774	0.193	73	0.597	0.403	0.978	0.156	0.022	0.844	0.175
3	81	0	1	0.970	0.214	0.030	0.786	0.232	67	0.607	0.393	0.973	0.158	0.027	0.842	0.179
4	102	0	1	0.972	0.218	0.028	0.782	0.228	69	0.598	0.402	0.975	0.158	0.025	0.842	0.177
5	96	0	1	0.977	0.226	0.023	0.774	0.193	74	0.597	0.403	0.978	0.156	0.022	0.844	0.175
6	95	0	1	0.977	0.226	0.023	0.774	0.193	67	0.598	0.402	0.975	0.158	0.025	0.842	0.177
7	86	0	1	0.968	0.201	0.032	0.799	0.240	80	0.607	0.393	0.973	0.158	0.027	0.842	0.179
Avg	**93**	**0**	**1**	**0.974**	**0.219**	**0.026**	**0.781**	**0.210**	**71**	**0.600**	**0.400**	**0.975**	**0.157**	**0.025**	**0.843**	**0.177**

Given these two initial conditions provided similar, yet slightly different results, Martínez-Beneito *et al*'s JAGS model was run to generate the parameter estimates via a Gibbs sampler (1). Tables 2 -4 below show the results from the BW and JAGS parameter inferences. From the tables we see a clear difference between the methods, with JAGS inferences having a significantly higher transition probability from epidemic to non-epidemic states. Thus, a Gelman-Rubin diagnostic was run for each parameter to assess MCMC convergence (results not shown). This led to potential scale reduction factors ranging from $1.1 - 2$ which signify a lack of convergence of the chains. No further analysis was done on this at the moment.

Table 2: Avg HMM Parameter Estimates

	P0.0	P0.1	P1.0	P1.1	rho
BW Fixed	0.974	0.219	0.026	0.781	0.210
BW Rand	0.975	0.157	0.025	0.843	0.177
MCMC	0.893	0.107	0.853	0.147	0.090

The `depmixS4` package was used instead to compare how effective the BW method delineated above was at classifying the epidemic vs non-epidemic weeks given JAGS couldn't verify the parameter estimates. However, the BW fixed initial state average parameters were used for decoding given it improved the model's sensitivity. This makes sense given overall the year starts at an epidemic state with lots of ILI cases due to weather. Figure 2 below shows the weekly classification results of the BW and `depmixS4` methods for all the years in the dataset. Overall, the BW method correctly classified the same amount of epidemic weeks as `depmixS4` (blue), but tended to miss epidemic weeks in the 2000s and late 1990s (green). However, the BW method was more sensitive at classifying the tails of the epidemic seasons in some cases (red).

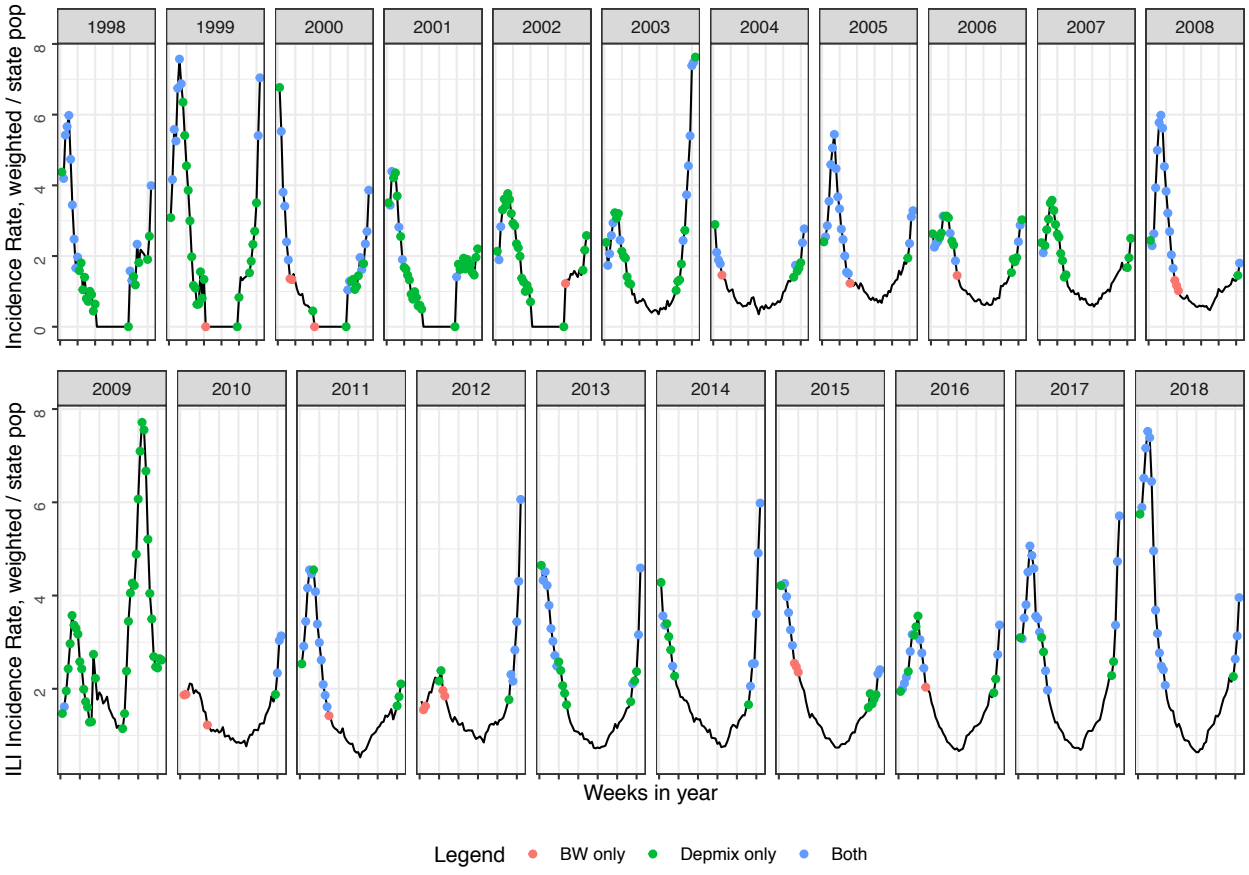
Table 3: Avg HMM Sigma 0 Param Estimates

Fixed Init	Random Init	MCMC
0.452	0.554	1.140
0.601	0.655	1.139
0.173	0.166	1.139
0.183	0.182	1.139
0.238	0.245	1.139
0.242	0.259	1.139
0.128	0.127	1.139
0.120	0.122	1.139
0.194	0.202	1.139
0.188	0.206	1.139
0.086	0.086	1.139
0.558	0.554	1.139
0.105	0.104	1.139
0.127	0.127	1.139
0.118	0.121	1.140
0.157	0.177	1.139
0.133	0.133	1.139
0.139	0.133	1.139
0.160	0.157	1.139
0.160	0.166	1.139
0.148	0.146	1.139

Table 4: Avg HMM Sigma 1 Param Estimates

Fixed Init	Random Init	MCMC
4.113	4.089	2.774
4.563	4.597	2.773
2.978	3.006	2.778
2.403	2.405	2.773
2.241	2.244	2.768
3.792	3.813	2.771
1.384	1.386	2.793
2.256	2.289	2.791
1.464	1.473	2.793
1.404	1.414	2.794
2.281	2.342	2.782
3.595	3.653	2.786
1.212	1.217	2.791
1.548	1.577	2.764
2.809	2.795	2.777
1.950	1.964	2.764
2.207	2.235	2.780
1.130	1.143	2.775
1.374	1.390	2.778
2.272	2.318	2.783
2.476	2.530	2.796

Fig 2. Classification Comparison of BW vs depmixS4



Conclusions

The goal of this project was to build a disease surveillance model focused to Influenza that could detect early onset of epidemic states. The parameters in the model, based on Martínez-Beneito *et al*'s previous approach, were successfully trained using a different parameter inference approach (BW) to that of Martínez-Beneito, which used a MCMC approach (3). When comparing these approaches the MCMC showed convergence problems and incorrect parameter estimations thus showing the limitations of their approach when applied to new data. Additionally, this approach takes a long time to generate the parameter inferences. In comparison, the EM-based approaches (BW and `depmixS4`) seemed to be successful at inferring the parameters and classifying the epidemic/non-epidemic dynamics with minor computational cost. Compared to `depmixS4`, the BW's AR(1) process specification seems to increases the classification sensitivity for the tails of the epidemics. Overall, these methods provided significantly faster inference results with the small caveat that the BW method misses multiple epidemic weeks. However, these approach does have its own limitations as it will find the local maxima/minima which highly depend on where parameters are initialized as shown above - specifically in the BW approach. Similarly, its effectiveness is dependent the correct derivations of the EM updates and subsequent implementation of said steps.
Further work in the parameter inference and modeling can be done. First, the BW implementation could be further improved the classification of epidemic dynamics similar to `depmixS4` via other optimization techniques such as Newton-Raphson or Monte Carlo EM (MCEM). Additionally, the model specification could be improved via a similar approach in Martínez-Beneito *et al*'s subsequent paper that uses a Poisson-based hierarchical approach.

References

[1] Strat, Yann Le, and Fabrice Carrat. "Monitoring Epidemiologic Surveillance Data Using Hidden Markov Models." Statistics in Medicine, vol. 18, no. 24, 1999, pp. 3463–3478., doi:10.1002/(sici)1097-0258(19991230)18:24<3463::aid-sim409>3.3.co;2-9.

[2] Rath, Toni M., et al. "Automated Detection of Influenza Epidemics with Hidden Markov Models." Advances in Intelligent Data Analysis V Lecture Notes in Computer Science, 2003, pp. 521–532.

[3] Martínez-Beneito, Miguel A., et al. "Bayesian Markov Switching Models for the Early Detection of Influenza Epidemics." Statistics in Medicine, vol. 27, no. 22, 2008, pp. 4455–4468., doi:10.1002/sim.3320.

[4] Conesa, D, et al. "Bayesian Hierarchical Poisson Models with a Hidden Markov Structure for the Detection of Influenza Epidemic Outbreaks." Statistical Methods in Medical Research, vol. 24, no. 2, 2011, pp. 206–223.,doi:10.1177/0962280211414853.

[5] Nunes, Baltazar, et al. "Nowcasting Influenza Epidemics Using Non-Homogeneous Hidden Markov Models." Statistics in Medicine, vol. 32, no. 15, 2012, pp. 2643–2660., doi:10.1002/sim.5670.

[6] BISHOP, CHRISTOPHER M. PATTERN RECOGNITION AND MACHINE LEARNING. SPRINGER-VERLAG NEW YORK, 2016.

[7] Zhai, ChengXiang. A Brief Note on the Hidden Markov Models (HMMs). University of Illinois at Urbana-Champaign, 13 Mar. 2003, pdfs.semanticscholar.org/54dc/c2a758e7fa34b8c2ef19826f39f16c4d1731.pdf.