VNIVERSITAT DE VALÈNCIA

Facultat de Ciències Matemàtiques

Departament d'Estadística i
Investigació Operativa

Programa de doctorat en Estadística i Optimització

Bayesian Temporal and Spatio-Temporal Markov Switching Models for the Detection of Influenza Outbreaks

PhD Thesis by

Rubén Amorós Salvador

supervised by

Dr. David V. Conesa Guillén

Dr. Miguel Ángel Martínez Beneito

March 2017

A mos pares,
perquè el vostre esforç també està ací.

This thesis was partially founded by the research grants MTM2013–42323–P from the Spanish Ministry of Economy and Competitiveness and ACOMP/2015/2012 from Generalitat Valenciana. Part of the research in this thesis was carried out during the visit of the author to the Centro de Investigação de Matemática of the Universidade do Minho in 2016, with the colaboration of Dr. Raquel Menezes.

Agraïments

Papà, Mamà, moltes gràcies. Ho heu fet prou bé, diria jo, ja que he arribat hasta ací. Després de llegir el treball final de màster mon pare es va sorprendre de lo bé que m'havia defensat davant de les preguntes del tribunal, mentres que ma mare estava segura de que ho anava a fer així de bé. Crec jo que ha sigut la combinació perfecta per a impulsar-me. Mon pare alçant la cella i dient "Ai... Mira que si no aplegues a temps..." (em va molt bé que m'espavilen pa que no m'encante) i ma mare fent-me sentir que confia en mí, que tot anirá bé i dient-me que disfrute del camí. També vos he d'agrair el donar-me un per qué. M'heu ensenyat a fer les coses per el plaer de medrar personalment i també per a ajudar als demés sent una part positiva de la societat, sense necessitats d'heroismes. Eixos son els meus objectius en la investigació i la docéncia, i els he heretat de vosaltres. També he d'agrair-vos, com no, per la qüestió logística. M'heu mantés durant tota la meua etapa d'estudis, permitint-me viure una vida d'universitari interessant i agradable, donant-me inclós la opció de canviar de camí, encara que suposara més temps. Heu sigut la meua beca quan m'ha fet falta i, quan no m'ha fet falta, heu sigut la red que m'ha permés fer funambulisme amb sueldos o paros que, algunes voltes, me donaven lo comido por lo servido, sense por de que em puguera quedar al carrer. Sé que no ha sigut sense sacrificis, com el retrasar alguns anys el començar a descansar del treball, així que vos ho agraïsc altra volta. Gràcies als dos per criar-me bé, per donar-me soport i per voler-me. Vos vuic molt.

David i Migue, moltes gràcies a vosaltres també. Amb David vaig començar el meu camí a la ciència i, poc temps després, el vaig contin-

uar amb Migue. David, m'has guiat i espentat quan ho he necessitat, i has sigut comprensiu amb els meus rodejos profesionals i vitals. Migue, m'has estimulat per a millorar com a investigador. Moltes gràcies als dos, he aprés molt amb vosaltres i també, clar, amb Antonio, sempre disposat a donar consell (sempre encertat). Gràcies també a Gonzalo, el teu consell, sempre amable, a l'hora de tractar amb la part teòrica de l'estadística ha sigut molt valuós. Gràcies en general a tota la gent de GEeitEma, VaBar i/o el departament d'Estad'istica i Investigació operativa. Sabeu fer sentir a la gent acollida i es genial saber que sempre es pot comptar amb la vostra ajuda.

Muito obrigado às pessoas no Centro de Investigação de Matemática da Universidade do Minho, especialmente Raquel. Fizeram minha estadia muito agradável. Obrigado por compartilhar voso conhecimento e aconselhamento. Gracias, obrigado, grazie, efcharistó, shkra, dhanyavaad, thanks, to all the nice people I met in Guimarães. I keep you all in my memory.

Vull agrair també als meus companys predocs i, en particular, a aquells amb qui vaig compartir despatx primerament: Facu, Anabel i Hèctor (el meu etern company, que ara se'n va un poc lluny i a qui trobaré a faltar). Les discussions (científiques o no) al despatx m'han ensenyat i m'han oferit noves perspectives. A més, el trobar amics en vosaltres ha fet molt més plaenter este ardu procés. Aprofite també per a agrair a Silvia i Fran. L'aventura de Bayestats amb tots vosaltres em va permetre conèixer parts de l'estadística que, d'altra manera, no haguera visitat. Vau ser uns excel·lents companys de viatge.

Gràcies als meus companys de treball al llarg d'estos anys a la EVES, al departament d'EiO, CSISP/FISABIO, l'oficina d'estadística de l'Ajuntament i el CEU. La gent es sol queixar dels seus llocs de treball, però jo sols he trobat gent amable i dispossada a ajudar que, en molts casos, s'han comportat més com a amics que com a companys de treball.

Gràcies també a tots els que m'heu ajudat a descansar, desconectar i a dur l'estrès. Els amics del poble, de la universitat, dels scouts i d'altres llocs. En particular, gràcies a (per ordre alfabètic) Bele, Desi, Gabo, Imeta, Inma, Manu, Pab, Santiago, Sergi, Thais, Victor i la Xiqueta (i respectius) per aguantar-me quan parlava d'investigació (i inclòs aportar la vostra part

de conversa sobre el tema). Segur que m'he deixat a algú que deuria estar també a aquesta llista. No m'ho tingueu en compte (i maneu-me un Whatsapp i quedem a fer una cervesseta, que ja fa temps que no ens veem segur).

Gràcies al Ministeri d'Economia i Competitivitat i a la Generalitat Valenciana per la seua contribució econòmica (relativament poca, la majoria se la he d'agrair a mos pares i als meus empleadors, jeje). Gràcies a Freepik per la imatge per a la portada (http://www.freepik.es/vector-gratis/fondo-de-acuarela-naranja_846463.htm).

Si m'he deixat algun agraïment, podeu dir-m'ho i ho afegiré a boli. Per a compensar, vos convidaré a una cervessa o un sopar, depenent de la gravetat del meu oblit.

Salut a totes i tots! Espere que disfruteu de la lectura, jejeje.

Resum

In compliance with the *Regulations for assessment and deposit of doctoral theses, Amended CG 28-VI-2016*, we present an abstract of this thesis written in Valencian, one of the official languages of the University of Valencia.

En compliment del *Reglament sobre depòsit, avaluació i defensa de tesi doctoral, Modificat CG 28-VI-2016*, presentem un resum d'aquesta tesi escrit en Valencià, un dels idiomes oficials de la Universitat de València.

Grip i models de detecció de brots

La grip és una malaltia que afecta a milions de persones i causa cents de morts cada any. Aquesta malaltia es també la causa de gran quantitats de despeses directes i indirectes degut als costos de l'atenció sanitària, absentisme i altres efectes de l'epidèmia. Per aquest raó, la vigilància d'aquesta infecció viral té un notable interès per a aquells que fan polítiques en salut ja que, entre d'altres beneficis de fer vigilància, detectar el moment exacte quan l'epidèmia està començant permet fer un millor ús dels recursos disponibles.

La epidèmia de grip té un comportament particular que dona forma als mètodes estadístics dedicats a la seua detecció. En aquest treball posem l'atenció a la detecció de la grip a les parts temperades del planeta, donat que el comportament de la malaltia és prou distint a les bandes equatorial

i tropicals, on la malaltia ocorre durant tot l'any.

La grip estacionària als països temperats ocorre durant els mesos freds de l'any. Per tant, certa estacionalitat és apreciada a l'epidèmia. En qualsevol cas, els creixements acusats i els pics d'incidència ocorren a diferents setmanes cada any, de tal manera que algunes temporades poden presentar l'inici de l'epidèmia al voltant d'any nou mentre que altres voltes, el brot es retarda fins a febrer. Inclús alguns anys es presenten sense un brot apreciable de grip estacionària. A banda de la grip estacionària, altres epidèmies de grip no estacionària poden ocórrer a qualsevol temps de l'any, normalment causades per ceps de virus que boten la barrera entre espècies dels animals als humans, com ha passat a les epidèmies anomenades 'grip porcina' o 'grip aviar'.

Independentment del seu origen i dada d'inici, les epidèmies de grip sempre duren varies setmanes (a algunes temporades l'epidèmia pot durar 5 setmanes, mentre que a altres aquesta es pot estendre fins als 3 o 4 mesos). Espacialment, les epidèmies de grip s'estenen començant per un o diversos punts, propagant-se amb el temps extensivament per països o inclús per continents sencers. Totes aquestes característiques de l'epidèmia es solen prendre en compte en diversos graus a l'hora de construir un model estadístic per a la detecció de brots de grip.

Cal tenir en compte que, amb gran freqüència, els sistemes de detecció d'epidèmies utilitzen dades sindròmiques per a vigilar la grip. D'aquesta manera, les dades que es fan servir són incidències de malaltia de tipus gripal (ILI, per l'acrònim anglès 'Influenza-Like Illness'), que es defineix com a cert conjunt de símptomes, usualment febre i alguna afecció del sistema respiratori superior com tòs i/o mal de gola. Aquesta aproximació per a la incidència de grip està lluny de ser precisa i comporta una quantitat de soroll considerable. Pel costat positiu, fer vigilància d'aquestes dades permet detectar no solament els brots de grip però també d'altres malalties com ara la síndrome respiratòria aguda greu o el potencial inici de certs atacs bio-terroristes. A partir d'ara, usarem les expressions 'vigilància de grip' i 'vigilància d'ILI' de forma indistinta entenent que, en general, es refereixen al mateixos sistemes i processos.

Diversos models estadístics han sigut proposats per a la detecció de

brots de malalties i, en particular, per als de grip, cadascú amb els seus punts forts i febles. Una revisió abreujada de la revisió de models per a la detecció d'epidèmies de grip del present treball de tesis ha sigut publicada a la revista *REVSTAT-Statistical Journal* (Amorós et al., 2015).

Una ferramenta interessant per a la modelització de mètodes per a la detecció d'epidèmies és l'ús de variables latents associades a cada temps (i localització, al cas de les propostes espai-temporals) que determinen distints models als quals les dades s'adapten millor depenent de l'estadi epidèmic al qual aquestes dades s'han observat. Per tant, cada submodel es construeix usant el coneixement que prèviament es té del diferent comportament de la malaltia durant les fases epidèmica i endèmica. Com les setmanes epidèmiques solen ser seguides per altres setmanes epidèmiques i les no epidèmiques per altres setmanes no epidèmiques, una manera comú de modelar aquestes variables latents és enllaçant-les mitjançant una cadena de Markov. Un avantatge de les variables latents per a la presa de decisions és que el seu valor esperat pot ser interpretat com la probabilitat estimada d'estar en fase epidèmica per a cada temps (i localització), de tal manera que la resposta és més rica que un simple 'sí' o 'no'.

El paradigma Bayesià ofereix un marc per a la inferència de models amb varies característiques interessants. Una d'elles és la interpretabilitat directa dels resultats dels models com a distribucions de probabilitat. Un altre avantatge és la flexibilitat dels models jeràrquics, sobre els qual es pot realitzar inferència gràcies a les ferramentes informàtiques de simulació. Això permet que varies estructures estadístiques puguen ser combinades a un sol model més complexe de manera relativament senzilla.

En aquest treball proposem dos extensions del model de Martinez-Beneito et al. (2008a) per a la detecció de brots de grip, una sobre dades temporals i una altra sobre dades espai-temporals. Ambdós combinen l'ús de ferramentes de sèries temporals i l'ús de variables latents sota un marc Bayesià. En particular, les dos propostes són models jeràrquics Bayesians amb un conjunt de variables latents associades a cada temps (i localització) distribuïdes com a una cadena de Markov oculta sobre el temps que indiquen l'estat epidèmic (fase epidèmica o no epidèmica). Depenent del valor estimat per a aquestes variables latents, les dades per a cadascun

dels temps (i localitzacions) es modelitzen mitjançant un de dos distints regressors. Tant la proposta temporal com la espai-temporal usen estructures temporals autoregressives per a reflectir el creixement i decaïment de la incidència durant l'epidèmia. El model espai-temporal inclou també un terme autoregressiu condicional el qual modela la propagació espacial de la infecció. Les estructures com la descrita anteriorment, a les quals la variable resposta depèn d'una cadena de Markov oculta i de la mateixa variable observada al temps anterior, i a les quals la relació entre la variable resposta a temps correlatius depèn del valor de la variable oculta, s'anomenen models de commutació de Markov (MSM per les sigles de l'anglés Markov Switching Model).

Proposta temporal

Una versió abreujada del treball presentat en aquesta tesis sobre la proposta temporal ha sigut ja publicada a la revista *Statistical Methods in Medical Research* (Conesa et al., 2015). Aquesta proposta consisteix en un marc de models Poisson temporals de commutació de Markov per a la detecció d'epidèmies de grip. Les dades modelitzades són dades que es solen trobar a la majoria dels sistemes de vigilància, que són series temporals de recomptes diàries o setmanals, tals com nombre d'admissions a hospital, casos d'incidència d'ILI, trucades rebudes als serveis d'emergència, etc. Passem ara a introduir la modelització de les dades del marc de models proposat.

Al model de Martinez-Beneito et al. (2008a) les taxes diferenciades de grip es modelen segons una distribució normal amb paràmetres dependents de l'estat epidèmic, de tal manera que la dinàmica no epidèmica es caracteritza per canvis aleatoris menuts i la epidèmica per canvis grans. Encara que la variabilitat permet distingir fins a cert punt les dos dinàmiques, incorporar la magnitud de la incidència pot ser avantatjós, ja que la magnitud pot també informar sobre l'estat de la malaltia, sabent que una alta incidència està clarament relacionada amb la fase epidèmica. Per açò incorporem a la proposta les taxes crues (sense diferenciar). Altra novetat

del marc de models temporals és el considerar les taxes no com a valors fixes sinó estocàstics, modelitzant-les mitjançant els recomptes observats i la població subjacent de la qual han sigut reportats. Anomenem Y_{ts} al nombre de casos observats de grip, ILI o altra característica usada per a rastrejar la grip (com ara l'absentisme o la venda de medicaments) durant la setmana (o dia) t a la temporada s. Modelitzem Y_{ts} mitjançant una distribució de Poisson, el paràmetre de la qual és funció de la taxa d'incidència r_{ts} de la setmana (o dia) t a la temporada s mitjançant la següent estructura jeràrquica:

$$\begin{aligned} Y_{ts} &\sim \mathrm{Po}(\nu_{ts})\,, \\ \nu_{ts} &= \frac{r_{ts}\mathrm{Pop}_{ts}}{100000}\,, \\ r_{ts} &\sim \mathrm{N}\left(R_{tsZ_{ts}}, \sigma^2_{sZ_{ts}}\right)\,, \end{aligned}$$

on $\mathrm{Pop_{ts}}$ representa la població vigilada a la unitat de temps corresponent. Noteu que el denominador depèn de la manera en que està definida la taxa. En aquest cas, ho hem expressat considerant que r_{ts} està definida sobre 100 000 habitants. Una volta definida la funció que uneix el nombre esperat de recomptes i la taxa d'incidència, modelem aquestes taxes mitjançant una distribució normal en la qual tant la mitjana com la variància depenen de l'estat epidèmic, determinat per la variable Z_{ts}, que identifica l'estat epidèmic de cada setmana.

Modelitzem la variable Z_{ts} com una variable latent no observada que segueix una cadena de Markov amb dos possibles estats, 1 per a les setmanes epidèmiques i 0 per a les endèmiques, amb probabilitats de transició:

$$p(Z_{t+1s} = l | Z_{ts} = k) = p_{kl}\,, \qquad\qquad k, l \in \{0, 1\}\,.$$

Sabent que $p_{k0} + p_{k1} = 1$, tenim que $p_{01} = 1 - p_{00}$ i $p_{10} = 1 - p_{11}$, per tant sols necessitem fer inferència sobre p_{00} i p_{11} per a establir tota la matriu de probabilitats de transicions entre estats. La cadena de Markov es reinicia cada temporada s, per tant també s'ha de fer inferència sobre les probabilitats inicials:

$$p(Z_{1s} = k) = p_k\,, \qquad\qquad k \in \{0, 1\}\,.$$

Amb l'objectiu d'expressar el nostre coneixement vague sobre aquestes probabilitats, usem la distribució a priori no informativa de Jeffreys usual per a assajos de Bernoulli per a p_{00}, p_{11} i p_0:

$$p_{00}, p_{11}, p_0 \sim \text{Beta}\left(\frac{1}{2}, \frac{1}{2}\right) ,$$

i obtenim la resta de probabilitats per complementarietat. La probabilitat a posteriori de que la variable Z_{ts} prenga valor 1 és la probabilitat estimada pel model de que la setmana t de la temporada s estiga en fase epidèmica.

Vejam la distribució proposta per al paràmetre de variància $\sigma^2_{sZ_{st}}$. La dinàmica endèmica es caracteritza per xicotets canvis aleatoris, mentre que la dinàmica epidèmica mostra fluctuacions majors a les taxes d'incidència. Per aquesta raó, proposem diferents valors de la variància per a cada fase de cada temporada, constrenyent una variància menor a la fase no epidèmica. Per a aconseguir açò, usem part de l'estructura jeràrquica present al model de Martinez-Beneito et al. (2008a):

$$\begin{aligned} \sigma_{s0} &\sim \text{Unif}(\theta_{[1]}, \theta_{[2]}) , \\ \sigma_{s1} &\sim \text{Unif}(\theta_{[3]}, \theta_{[4]}) , \\ \theta_m &\sim \text{Unif}(0, a) , \qquad\qquad m = 1, ..., 4 , \end{aligned}$$

on $\{\theta_{[1]}, \theta_{[2]}, \theta_{[3]}, \theta_{[4]}\}$ correspon a la seqüència ordenada de les variables $\{\theta_1, \theta_2, \theta_3, \theta_4\}$, i a és un hiperparàmetre fixat pel modelador, que típicament expressa un coneixement a priori vague. Aquesta manera de definir les variàncies evita problemes d'intercanviabilitat i falta d'identificabilitat.

El següent pas és modelar les mitjanes de les taxes per a ambdós estats. Aquesta és una de les majors novetats respecte del model de Martinez-Beneito et al. (2008a). A la seua proposta modelaven directament les taxes diferenciades mentre que a la present proposta modelem la incidència de recomptes mitjançant la distribució de les taxes crues (no diferenciades). R_{ts0} i R_{ts1} representen la magnitud esperada de les taxes d'incidència r_{ts} a cadascuna de les fases i podem prendre avantatge d'açò per a distingir les dos dinàmiques. Degut a la natura temporal de les dades, proposem una

estructura autoregressiva per a cadascun d'ells. Així apareixen diverses opcions depenent de l'ordre del procés autoregressiu elegit.

El model més senzill és considerar les mitjanes com a dos constants distintes, les quals es poden considerar com a processos autoregressius d'ordre zero:

$$\begin{aligned} R_{ts0} &= \mu_0 \,, \\ R_{ts1} &= \mu_1 \,. \end{aligned}$$

Forcem la mitjana de les taxes no epidèmiques a ser menor, $\mu_0 < \mu_1$, mitjançant la definició de les distribucions a priori:

$$\begin{aligned} \mu_0 &= \lambda_{[1]} \,, \\ \mu_1 &= \lambda_{[2]} \,, \\ \lambda_m &\sim \mathrm{Unif}(0, b) \,, \qquad\qquad m = 1, 2 \,, \end{aligned}$$

on $\{\lambda_{[1]}, \lambda_{[2]}\}$ correspon a la seqüència ordenada de les variables $\{\lambda_1, \lambda_2\}$, i b és un hiperparàmetre elegit per a fer la distribució a priori de λ_m vaga. Anomenem a aquest model AR0-AR0, el primer terme referit a l'estructura de la fase endèmica i el segon al de la epidèmica.

Es esperable certa estructura temporal de la incidència de grip o ILI. Per tant, una segona opció és considerar que la mitjana de les taxes (a un o ambdós estats epidèmics) és depenent de l'observació anterior. A aquesta configuració, les taxes es distribueixen al voltant d'un valor desconegut com abans (μ_0 i μ_1 respectivament), però amb l'afegit de que si l'anterior taxa estava per davall d'aquest valor mitjà, el següent és més probable que també ho estiga i viceversa. En aquest cas, la mitjana de les taxes segueix un procés autoregressiu d'ordre un. Tres models més poden ser considerats combinant AR0 i AR1: AR0-AR1, AR1-AR0 i AR1-AR1. Per simplicitat, mostrem solament AR1-AR1:

$$\begin{aligned} R_{1s0} &= \mu_0 \,, \\ R_{1s1} &= \mu_1 \,, \\ R_{ts0} &= \mu_0 + \rho_0 (r_{t-1s} - \mu_0) \,, \qquad\qquad t > 1 \,, \\ R_{ts1} &= \mu_1 + \rho_1 (r_{t-1s} - \mu_1) \,, \qquad\qquad t > 1 \,. \end{aligned}$$

Definim distribucions a priori uniformes per als paràmetres dels processos autoregressius de la fase endèmica i epidèmica a la regió que assegura que aquests processos són estacionaris:

$$\rho_0, \rho_1 \sim \text{Unif}(-1, 1)\,.$$

Una tercera opció per a modelitzar les taxes és considerar-les relacionades amb taxes de dos o més setmanes anteriors. Així, podem considerar processos autoregressius d'ordres majors a un. En qualsevol cas, en aquest treball treballem amb un ordre màxim de 2. Com a resultat, es poden considerar cinc nous models: AR2-AR0, AR2-AR1, AR2-AR2, AR1-AR2 i AR0-AR2. Per simplicitat, presentem el model on les dos mitjanes segueixen processos autoregressius d'ordre 2 (AR2-AR2):

$$\begin{aligned}
R_{1s0} &= \mu_0\,,\\
R_{1s1} &= \mu_1\,,\\
R_{2s0} &= \mu_0 + \frac{\rho_{10}}{1-\rho_{20}}(r_{1s} - \mu_0)\,,\\
R_{2s1} &= \mu_1 + \frac{\rho_{11}}{1-\rho_{21}}(r_{1s} - \mu_1)\,,\\
R_{ts0} &= \mu_0 + \rho_{10}(r_{t-1s} - \mu_0) + \rho_{20}(r_{t-2s} - \mu_0)\,, && t > 1\,,\\
R_{ts1} &= \mu_1 + \rho_{11}(r_{t-1s} - \mu_1) + \rho_{21}(r_{t-2s} - \mu_1)\,, && t > 1\,.
\end{aligned}$$

Definim els paràmetres dels processos autoregressius d'ordre 2 a la regió on aquests processos són estacionaris:

$$\begin{aligned}
\rho_{2k} + \rho_{1k} &< 1\,,\\
\rho_{2k} - \rho_{1k} &< 1\,,\\
-1 < \rho_{2k} &< 1\,.
\end{aligned}$$

Totes les expressions prèvies contenen tot el coneixement sobre el sistema però, com és habitual als models jeràrquics, no disposem d'expressió analítica de la distribució a posteriori dels paràmetres. Per tant utilitzem simulació de Monte Carlo en cadenes de Markov (MCMC, per les sigles angleses de Markov Chain Monte Carlo) mitjançant el programa WinBUGS per tal d'aproximar-les.

Proposta espai-temporal

La idea de la proposta espai-temporal és usar les diferents dinàmiques de les taxes diferenciades d'ILI (els salts d'una setmana a la següent) per a diferenciar entre les setmanes epidèmiques i endèmiques de les diverses localitzacions estudiades. Centrar l'atenció a les taxes diferenciades permet detectar canvis a les dinàmiques que denoten un brot sense importar que aquests canvis siguen observats a taxes baixes de grip, com és usualment el cas al començament d'una epidèmia. Amés, el fet de modelitzar directament les taxes diferenciades simplifica el model, el qual pot fer la inferència més ràpida.

Per simplicitat a la notació, donada una localització i i una temporada s, denominem y_{its} a la taxa en temps t (r_{its}) menys la taxa en temps $t-1$ (r_{it-1s}):

$$y_{its} = r_{its} - r_{it-1s}\,.$$

Comencem els índexs t de les taxes crues r_{its} en 0 de tal manera que els índexs t per a les taxes diferenciades y_{its} comencen en 1. La variable Z_{its} indica l'estat latent epidèmic o endèmic per a cada lloc i temps, amb valor 1 per a l'estat epidèmic i valor 0 per a l'estat no epidèmic. Considerem que les taxes diferenciades segueixen una distribució normal amb mitjana i variància depenent de l'estat epidèmic:

$$y_{its} \sim \mathrm{N}(R_{itsZ_{its}}, \sigma^2_{Z_{its}})\,.$$

Cal notar que la notació $R_{itsZ_{its}}$ a aquest model no té el mateix significat que al marc de models temporals proposat, on $R_{tsZ_{ts}}$ era el valor esperat de les taxes crues r_{ts} donada Z_{ts}. En aquest cas, és el valor esperat de les taxes diferenciades $r_{it+1s} - r_{its}$ donada Z_{ts}.

Com a la proposta temporal, modelitzem les variables latents Z_{its} com una cadena de Markov oculta per a cada localització. Així, la distribució de Z_{its} condicionada a Z_{it-1s} segueix una distribució de Bernoulli amb probabilitats de transició comuns a tots els temps, localitzacions i temporades:

$$Z_{its} \sim \mathrm{Ber}(p_{Z_{it-1s}1})\,.$$

Definim distribucions a priori no informatives de Jeffreys per a assajos de Bernoulli per a les probabilitats de transició de la mateixa manera que les descrites per a la proposta temporal:

$$p_{00}, p_{11}, p_0 \sim \text{Beta}\left(\frac{1}{2}, \frac{1}{2}\right) ,$$

on

$$\begin{aligned} p_{kl} &= P(Z_{its} = l | Z_{it-1s} = k) , \\ p_k &= P(Z_{i1s} = k) , \qquad\qquad k, l \in \{0, 1\} , \end{aligned}$$

i la resta de probabilitats p_{10}, p_{01} i p_1 s'obtenen per complementarietat. La probabilitat a posteriori de que la variable Z_{its} prenga valor 1 és la probabilitat estimada pel model de que la localització i a la setmana t de la temporada s estiga en fase epidèmica.

La fase no epidèmica es caracteritza per salts de les taxes propers al zero. Una primera aproximació per a modelitzar aquests salts no epidèmics pot ser deixar la mitjana d'aquests salts igual a 0, però les dades mostren que hi ha setmanes no epidèmiques amb creixements o decaïments de les taxes comuns a totes les localitzacions. Açò és probablement una manifestació de certa estacionarietat de les dades endèmiques, amb suaus decreixements després de l'estació freda i suaus creixements després de l'estació càlida. En qualsevol cas, aquests increments i decrements no ocorren a les mateixes setmanes tots els anys, i epidèmies no estacionàries poden trencar aquesta dinàmica. Per aquesta raó, considerem que cada setmana t de la temporada s té una mitjana μ_{ts0} comú a totes les localitzacions per a dita setmana, però diferent a la mitjana d'altres setmanes.

La fase epidèmica es modela amb una estructura espai-temporal més complexa per a la mitjana de les taxes diferenciades. És d'esperar a una localització en estat epidèmic el tenir diversos salts (taxes diferenciades) positius fins a aplegar al pic de l'epidèmia i, després, que els salts es tornen negatius fins a aplegar al nivell endèmic de les taxes d'incidència. Per tant, és d'esperar que les taxes diferenciades per a cada localització siguen temporalment dependents. Altre comportament esperable degut al caràcter

contagiós de la grip és que si una regió te creixements epidèmics de les taxes d'incidència, les regions veïnes poden infectar-se i tenir creixements semblants. Com a conseqüència d'açò, modelem la mitjana de les taxes diferenciades mitjançant un terme μ_{ts1} comú a totes les localitzacions però diferent per a cada temps, més una estructura autoregressiva temporal d'ordre 1 per a cada localització amb paràmetre ρ, més un model espacial autoregressiu condicional intrínsec (ICAR, per les sigles angleses d'Intrinsic Conditional Auto-Regressive model) per a cada temps. Les expressions matemàtiques per a la mitjana dels periodes endèmic i epidèmic son:

$$\begin{aligned} R_{its0} &= \mu_{ts0}\,, \\ R_{its1} &= \mu_{ts1} + \rho y_{it-1s} + \psi_{its}\,, \end{aligned}$$

amb ψ_{its} el terme ICAR. Per a assegurar l'estacionarietat del procés autoregressiu cal delimitar els valors del paràmetre ρ a l'interval $[-1, 1]$. De tota manera, si assumim que la correlació entre creixements subsegüents a la fase epidèmica és positiva, a la pràctica podem restringir a l'interval $[0, 1]$:

$$\rho \sim \text{Unif}\,(0, 1)\ .$$

En quant a μ_{ts0} i μ_{ts1}, els considerem com a dos efectes aleatoris sobre el temps, amb major variabilitat per a μ_{ts1}, donat que és esperable que els creixements i decaïments epidèmics siguen majors que aquells de la fase endèmica. Per a evitar problemes d'identificabilitat amb la variabilitat no estructurada $\sigma_{Z_{its}}$ de la distribució principal, establim les desviacions típiques dels dos efectes aleatoris com a proporcionals a aquelles de la distribució principal. Podem expressar la modelització dels termes temporals comuns així:

$$\begin{aligned} &\mu_{ts0} \sim \text{N}(0, \sigma^2_{\mu 0})\,, && \sigma_{\mu 0} = \lambda\sigma_0\,, && \\ &\mu_{ts1} \sim \text{N}(0, \sigma^2_{\mu 1})\,, && \sigma_{\mu 1} = \lambda\sigma_1\,, && \lambda \sim \text{Unif}(0, a)\,, \end{aligned}$$

on λ és el factor de proporció estimat i a és un hiperparàmetre fixat pel modelador de tal manera que expresse un coneixement vague.

En general, les taxes diferenciades estan centrades a zero, per tant, encara que el comportament de la mitjana de la variable resposta aporta informació important per a distingir entre la fase epidèmica i endèmica, el comportament de la variància és d'importància crítica per a aquesta tasca. Els salts de les taxes no epidèmiques són relativament menuts en valor absolut, mentre que els creixements o decreixements de les taxes a l'estat epidèmic són usualment més grans. Per tant, modelitzem la variància no estructurada de les dos fases obtenint dos desviacions típiques d'una distribució a priori uniforme i ordenant-les de manera que la variabilitat epidèmica siga major que la endèmica:

$$\begin{aligned} \sigma_0 &= \theta_{[1]} \\ \sigma_1 &= \theta_{[2]} \\ \theta_m &\sim \text{Unif}(0, c) \qquad\qquad m = 1, 2\,, \end{aligned}$$

amb c un hiperparàmetre a fixar pel modelador de manera que expresse un coneixement a priori vague.

Com ocorre amb la proposta temporal, no disposem d'expressió analítica per a la distribució a posteriori dels paràmetres del model jeràrquic a aquesta proposta espai-temporal. Per tant, utilitzem simulació MCMC mitjançant el programa WinBUGS per tal d'aproximar-les.

Resultats i Conclusions

La revisió dels models de detecció d'epidèmies de grip temporals i espai-temporal ha servit per a justificar la necessitat de crear noves metodologies capaces de franquejar limitacions d'anteriors mètodes i per a informar la creació de les noves propostes exposades abans. Vejam algunes de les limitacions a la literatura i cóm les nostres propostes les franquegen.

Una proposta usada a la detecció d'epidèmies és construir llindars amb dades històriques i donar una alarma quan les dades els sobrepassen. Aquestes propostes tenen l'inconvenient de necessitar una predefinició de què és considerat com a epidèmic i no epidèmic a les dades històriques, la qual

cosa és un problema de detecció d'epidèmies en ella mateixa. Un avantatge d'usar cadenes de Markov a la modelització és el fet de que no es requereix una diferenciació prèvia de quines setmanes estan en fase epidèmica o endèmica per a les dades usades per a entrenar el model.

Els models basats en gràfics de control cumulatius poden detectar com a epidèmic un increment relativament menut però persistent del nivell mitjà de les taxes que s'estenga durant varis dies o setmanes. Però aquest és un comportament habitual a la fase endèmica (no epidèmica) de les taxes d'incidència d'ILI. Les nostres propostes consideren el comportament de les taxes diferenciades (el tamany i direcció dels salts d'una setmana a altra) com un tret diferenciador, i no sols la grandària mitjana de les taxes d'incidència. D'aquesta manera, un creixement brusc o diversos creixements suaus contigus són necessaris per a donar l'alarma.

Una alternativa a l'ús de llindars que trobem a la literatura és l'ús de l'anomenat 'scan statistic', que realitza contrastos d'hipòtesis per a determinar si certa regió temporal, espacial o espai-temporal té una quantitat inusual de casos comparat amb la resta. Aquesta metodologia no sembla molt adequada per a la detecció espai-temporal d'epidèmies de grip, ja que requereix que l'extensió espacial de les epidèmies siga menor que la meitat de la regió d'estudi, mentre que les epidèmies de grip solen estendre's a regions més amplies. La nostra proposta espai-temporal, sent un MSM, és capaç de classificar qualsevol quantitat de localitzacions com a no epidèmiques o com a epidèmiques, ja que no hi ha cap restricció a la quantitat de les variables Z_{its} que poden prendre valor 0 o 1 cada setmana.

Respecte a la proposta temporal abans descrita, diverses aplicacions sobre dades reals ens han permès discutir l'actuació del marc de models, oferir algunes directrius per a elegir un model dintre del marc i discutir els problemes de les ferramentes dedicades a avaluar i comparar els mètodes de detecció de brots.

La presència o absència del terme autoregressiu a la modelització de les mitjanes de les taxes a la proposta canvia en gran mesura la classificació com a fase epidèmica o endèmica de les setmanes analitzades. No obstant, si l'estructura autoregressiva està present, el canvi en ordre d'aquesta afecta sols de manera lleu als resultats inferits. La natura de les dades de la grip

i alguns resultats observats suggereixen que és important l'ús d'estructura autoregressiva. Per raons de parsimònia, també es recomana seleccionar entre els models amb mateix ordre autoregressiu a ambdós regressors (AR1-AR1 i AR2-AR2).

Diversos mètodes es poden fer servir per a seleccionar entre diversos models de detecció d'epidèmies. La primera qüestió a considerar és si fer l'avaluació sobre resultats en línia (aplicant el model tantes voltes com setmanes a avaluar usant cada volta sols la informació actual i prèvia per a obtenir l'estimació de cada setmana) o sobre resultats retrospectius (aplicant el model amb totes les dades una sola volta i obtenir estimacions per a totes les setmanes). L'aplicació en línia és realista, però computacionalment costosa. Les probabilitats estimades amb les dos metodologies són similars, però l'aplicació en línia pot declarar alguns brots una o dos setmanes més endarrerits comparada amb l'aplicació retrospectiva (comprensible, ja que l'aplicació retrospectiva pot usar informació posterior per a inferir resultats i la en línia no).

La segona qüestió a tractar és la selecció d'alguna mesura sobre el rendiment dels models de detecció. El criteri d'informació de la desviació (DIC, per les sigles angleses de Deviance Information Criterion) és poc costós, ja que s'obté de l'aplicació retrospectiva dels models. Mesura la bondat d'ajust del model penalitzant per la seua complexitat, però no mesura directament el que es tracta d'avaluar, la qualitat de la detecció. Les mesures ROC (sigles de l'anglès Receiving Operative Characteristic, o característica operativa del receptor) ponderades avaluen la qualitat de la detecció i la seua oportunitat (rapidesa a senyalar un brot emergent), però existeixen dos dificultats. La primera és seleccionar cóm reduir una informació tridimensional (sensitivitat, especificitat i oportunitat) a una sola dimensió, donat que diferents formes de combinar resulten en distints criteris.

El segon problema és que les mesures ROC ponderades requereixen de la definició d'un estàndard d'or o prova de referència, un criteri extern que definisca quines setmanes són epidèmiques i quines endèmiques. Nosaltres hem usat dades d'aïllament del virus en laboratori per a construir l'estàndard d'or, però diverses formes de construir l'estàndard d'or són possibles amb aquest tipus de dades, i altres dades o mètodes poden també

ser usats. Vegem així que la definició d'un estàndard d'or no és en absolut trivial, i els resultats reflecteixen que les mesures ROC ponderades són considerablement sensibles a aquesta definició. Les correlacions entre DIC i mesures ROC ponderades observades a una de les aplicacions del model estan lluny de ser perfectes, però són majors al 0.5, el que indica que quan existeix dificultat per a calcular les mesures ROC ponderades, el DIC es pot usar com a un proxy d'elles.

Els models de la proposta temporal han obtingut millors puntuacions de les mesures ROC ponderades que altres models de la literatura i que el model de Martinez-Beneito et al. (2008a) del qual és una extensió. Açò indica que els models de la nostra proposta ofereixen una millor i més ràpida detecció dels brots de grip.

Un apropament distint ha sigut pres per a l'avaluació de l'aplicació de la proposta espai-temporal. Açò ha sigut degut a considerar els problemes de les mesures ROC ponderades observades a les aplicacions de la proposta temporal i a que no es disposava de dades de laboratori per a construir un estàndard d'or per a les dades espai-temporals. Per tant, s'ha fet avaluació de validació creuada aproximada usant la puntuació de probabilitat de rang continu (CRPS, per les sigles angleses de Continuous Rank Probability Score) com a mesura de la discrepància entre la distribució predictiva i les dades observades. Al fer-ho, assumim que un model que és capaç de predir les dades de manera correcta també ha d'inferir correctament els valors de les variables latents que indiquen l'estat epidèmic.

El CRPS ha mostrat ser útil a l'hora d'avaluar la rellevància de dos termes del model. Un d'ells és el terme ICAR espacial al regressor epidèmic, i l'altre es μ_{ts0}, l'efecte aleatori sobre la mitjana de la taxa diferenciada no epidèmica. Ambdós han provat que són termes de la proposta que milloren de forma notable el CRPS. Açò demostra que estos termes són necessaris per a estimar correctament les dades i, si l'assumpció que em expressat al paràgraf anterior és correcta, per a determinar millor l'estat epidèmic.

L'estructura espacial proposada per Leroux et al. (2000) també ha sigut provada com a una alternativa a l'estructura ICAR. Aquesta modelització ofereix certa flexibilització del model jeràrquic proposat que pot ser interessant teòricament. De tota manera, les probabilitats d'epidèmia estimades

són quasi idèntiques, i les puntuacions de CRPS són lleugerament pitjors.

El model de Martinez-Beneito et al. (2008a) ofereix estimacions de la probabilitat d'epidèmia prou distintes a aquelles oferides per la nova proposta en diverses localitzacions i temps. Els valors de CRPS són notablement pitjors que els de la nova proposta i les variacions abans comentades. Açò indica el notable impacte i rellevància d'usar un model espai-temporal en lloc d'usar un model temporal per a cada regió de manera aïllada, un enfocament del problema que no permet la compartició d'informació al llarg de l'espai.

Per a resumir, a aquesta investigació hem mostrat una revisió de mètodes de detecció de brots de grip i hem proposat un marc de models temporal i un model espai-temporal per a la detecció de brots de grip. Aquestes propostes tenen la capacitat de detectar els brots sense la necessitat de definir prèviament les fases epidèmiques i endèmiques a les dades històriques, no assumeixen una localització temporal fixa de les epidèmies i poden modelitzar la difusió espacial extensa usual a les epidèmies de grip. L'ús de cadenes ocultes de Markov i d'estructures temporals i espai-temporals de correlació condicional baix el paradigma Bayesià ha sigut crític per a construir les noves propostes i per a definir de manera senzilla un criteri per a donar alarmes en forma de probabilitat d'epidèmia. També, la variabilitat de les taxes i el comportament de les taxes diferenciades han demostrat ser trets importants de les dades de grip per a distingir entre fases epidèmica i no epidèmica. Hem vist també cóm l'avaluació i comparació del rendiment de models no és trivial, i hem discutit distintes maneres de fer-ho.

Contents

List of Figures

List of Tables

Introduction

Influenza is a disease that affects millions of people and causes hundreds of thousands of deaths each year. This disease is also the cause of large amounts of direct and indirect expenses due to health care costs, absenteeism and other effects of the epidemic. For this reason, the surveillance of this viral infection has notorious interest for health policy makers as, among other benefits of doing surveillance, detecting the exact moment when the epidemic is starting allows for a better use of resources.

The epidemic of influenza has a particular behavior that shapes the statistical methods dedicated to its detection. In this work we focus the attention on the detection of influenza epidemics at temperate parts of the planet, as the behavior of the disease is quite different in the equatorial and tropical bands, where the disease happens during all the year.

Seasonal influenza in temperate countries occurs during the cold months of the year, so certain seasonality is appreciated in the epidemic. Anyhow, the steep growths and the peaks of each epidemic happen in different weeks every year, so some seasons may present the start of the epidemic around New Year's Eve while some other times the outbreak waits until February to happen. There are even some years without an appreciable outbreak of seasonal influenza. Besides the seasonal outbreaks, other non-stationary influenza epidemics can happen at any time of the year, usually caused by strains of virus that jump species from animals to humans, as happened with the so called 'swine flu' or 'bird flu' epidemics.

Regardless of their origin and starting date, influenza epidemics always last several weeks (some seasons might last around 5 weeks while others

can extend during 3 or 4 months), and spread spatially, starting from some points and usually extending over whole countries or even continents. All of these epidemic characteristics are usually taken into account in varying degrees when building a statistical model for the detection of influenza outbreaks.

Several statistical models have been proposed for the detection of diseases outbreaks and, in particular, for those of influenza. Building thresholds with historical data and trigger an alarm when data surpass them, finding the break point when the mean behavior changes its monotony or doing hypothesis testing about an abnormal quantity of cases inside a region are some of the approaches that have been taken. All of them have their strengths and weak points, that will be discussed in this work.

Another alternative which is able to overcome some of the drawbacks of previous methodologies is the use of latent variables associated to each time (and location, in the case of spatio-temporal proposals) which determine different models to which data adapt better depending on the epidemic phase data are collected from. Therefore, each sub-model is built using the knowledge we have about the behavior of the disease during the endemic and the epidemic season. As epidemic weeks are usually followed by epidemic weeks and non-epidemic weeks are usually followed by non-epidemic weeks, a common way to model these latent variables is to link them with a Markov chain. One advantage of these latent variables for decision making is that its expected value can be interpreted as the estimated probability of being in epidemic phase for each time (and location), so the answer is richer than a simple 'yes' or 'no' answer.

The Bayesian paradigm offers a framework for the inference of models with several interesting characteristics. One of them is the direct interpretability of the outcome as probability distributions. Another advantage is the flexibility of the hierarchical models, on which inference can be done thanks to simulation tools. That allows for several statistical structures to be joined in a single more complex model.

In this work we propose two models for the detection of influenza outbreaks, one on temporal data and one on spatio-temporal data. The two of them combine the use of time series tools and the use of latent variables

in a Bayesian framework. In particular, both proposals are Bayesian hierarchical models with a set of latent variables associated to each time (and location) distributed as hidden Markov chains over time that indicate the epidemic state (epidemic or non-epidemic phase). Depending on the estimated value for these latent variables, data for each time (and location) are modeled by two different regressors. Both proposals use temporal autoregressive structures to reflect the growth and decay of the incidence during the epidemic, and the spatio-temporal proposal also includes a conditional autoregressive term that models the spatial spread of the infection.

The temporal proposal is not only one model but a framework of hierarchical models, with several common layers. The first layer of the hierarchical structure models the counts of incident cases through a Poisson distribution. A second layer models the rates as normally distributed, with several possible combinations of regressors for the epidemic and non-epidemic mean rates with different degrees of temporal conditional dependence. The spatio-temporal proposal is one model where the differentiated rates (the increases or decreases on the weekly incidence) are modeled with a Gaussian distribution, with temporal and spatial terms in the regressors of the mean differentiated rate. In both of the proposals, the variances are a key point for the distinction of the epidemic phase, so all the proposed models consider that data will have less variability during the non-epidemic phase.

The rest of the thesis is structured as follows: In Chapter 1 the problem is introduced by briefly discussing influenza epidemics and surveillance systems. After that, a review of temporal statistical methods for the detection of influenza outbreaks is performed. This review pays special attention to the models that involve any of the features which are present in our proposals, and concludes with the detailed explanation of the model proposed by Martinez-Beneito et al. (2008a), in which the new proposals in this thesis are based.

In Chapter 2, a framework of temporal Poisson Markov switching models on the differentiated rates for the detection of influenza outbreaks is presented. A detailed description of this framework of models is offered, defining the models through the conditional distributions of each one of the levels of the hierarchical structure. After that, several applications on real

data allow us to discuss the performance of the framework of models, give some guidelines to choose the most appropriate method among the models in the proposed framework and discuss the problematic of the tools for the evaluation of outbreak detection methods.

Chapter 3 starts with a review of spatio-temporal statistical methods for the detection of outbreaks, specially to those which deal with influenza epidemics. Special attention is paid to the models and statistical tools that deal with spatial lattice structured data, because any other type of spatial data is usually easily translatable to it. The following section shows in detail the proposed spatio-temporal Markov switching model on the differentiated rates for the detection of influenza outbreaks through the detailed definition of the conditional distributions that form the hierarchical model. An application on real data is used to test the performance of the new proposal and to compare it with several modifications and simplifications in order to asses the relevance of the model.

Chapter 4 presents the conclusions of the work and proposes several future lines of investigation.

The appendices include a review of methods for the selection of statistical algorithms for the detection of outbreaks, which describes several measures devoted to this purpose; a discussion about a specific framework for the detection of outbreaks proposed by Banks et al. (2012); WinBUGS codes of the new proposals; and some figures which would hinder the reading if they were to be put along with the text of the thesis.

Chapter 1

Methods for the detection of influenza epidemic outbreaks

1.1 Introduction

In this chapter we take a brief look at the reasons to do influenza surveillance and how this surveillance is done. To do so, we briefly explain what influenza is, how does it spread and its impact in public health and economy. We also describe what a surveillance system is, paying special attention to the processes of data collection and analysis of the data. The main part of this chapter consists of a review of temporal methods for the detection of outbreaks, focusing on those for the detection of influenza epidemics. The review visits the classic methods and then will pay special attention to Bayesian models and the use of hidden Markov chains for the detection of epidemics. Finally we describe the proposal of Martinez-Beneito et al. (2008a), a Bayesian model with a hidden Markov structure for the detection of influenza outbreaks in which the new proposals of the present work are based on.

1.2 Influenza epidemics

In order to know how to deal with detection of influenza epidemic outbreaks we must first know about the disease itself and its dynamics, as well as the reasons to monitor this specific disease.

Influenza is an infectious disease that affects the upper and/or lower parts of the respiratory tract and is caused by the influenza virus. This disease spreads all around the world; through seasonal epidemics at temperate climates and at any time of the year at tropical regions. The continuous prevalence in the central belt of the planet supposedly serve as a reservoir to the seasonal epidemics (Rambaut et al., 2009). Influenza epidemics are related to climate variables such as humidity, temperature and radiation, though the mechanisms of this relation are not clear, as stated by Tang et al. (2010).

There are 3 types of seasonal influenza viruses –A, B and C– and many subtypes of each according to the combination of various virus surface proteins. There also exist strains of influenza that affect animals and may eventually cross the border of species and infect humans. These strains may cause epidemics that have different temporal behavior from seasonal influenza, appearing not only during the cold moths of the year but at any time, so a good outbreak detection method will have to take this into account. Well known examples of this cross-species infections are the Avian influenza strains A(H5N1) and A(H7N9) or the swine influenza A(H1N1)pdm09 virus, which spread worldwide. This last epidemic was monitored with special attention and led to the analysis and re-evaluation of several influenza surveillance strategies, as shown in the works of Ortiz et al. (2009), Tilston et al. (2010), Cook et al. (2011), Kavanagh et al. (2012), De Lange et al. (2013), Mulpuru et al. (2013) and Gomez-Barroso et al. (2014).

Some of the possible symptoms of influenza are fever, cough, sore throat, runny nose, muscle and joint pain and severe malaise, which appear in about 2 days from infection. Healthy people usually recover from fever and other symptoms within a week without requiring medical attention, but people at risk, like young children, elders, and people with certain

medical conditions have higher risk of complications, which can lead to hospitalization and death. Some of the symptoms are used as a proxy to influenza, constituting what is known as Influenza-Like Illness (ILI) which is defined as a subset of the symptoms, usually fever and some upper-respiratory affection as cough and/or sore throat. Anyhow, as the work of Navarro-Marí et al. (2005) shows, this proxy is far from accurate and carries a considerable amount of noise with it. A diagnose of Influenza requires laboratory confirmation, which highly improves the specificity, but it is costly and delays the diagnose.

Other respiratory infections may have similar symptoms in some of their stadiums, such as SARS, respiratory syncytial virus, pertussis, rhinovirus, enterovirus or other potential infections not contemplated now by several public health institutions (as it was with SARS before 2003). Outbreaks of these illnesses may also be detected by influenza surveillance systems, specially those based on symptomatology. Also some potential bio-terrorist attacks have similar initial symptoms. The usefulness of syndromic surveillance systems in cases of bio-terrorism is uncertain for developed countries (Buehler et al., 2003), but some authors like Paterson and Durrheim (2013) consider that it might be appropriate for developing regions where laboratory confirmations are difficult to obtain in a timely manner.

Influenza and ILI causes a high impact in terms of mortality, absenteeism and use of health care resources, with their associated economic burden. According to the World Health Organization (a) (WHO), the annual incidence rate is estimated at 5%–10% in adults and 20%–30% in children, with 3 to 5 million cases of estimated severe illness and about 250 000 to 500 000 deaths per year. The economic cost of influenza is reviewed in detail by Gasparini et al. (2012), who indicated that each case of influenza in adults aged 18-64 years old costs several hundreds of euros in direct or indirect expenses in the US and EU.

The early detection of influenza and other ILI epidemics helps taking public health actions that save resources, money and lives. In particular, knowing when and where the epidemic starts helps optimizing health resources such as health workers, hospital facilities, medicines, publicity to encourages prevention measures, etc., by use them at the times and loca-

tions where they will be most useful. A good planning avoids the misspend of resources, but it also rises the prevention of infection, which results in less monetary spend due to work absenteeism and cost of treatments. The quality of treatment to the sector of the population at special risk, for which the infection is an actual threat to life, is also improved thanks to a good use of the resources due to a timely detection of the epidemics.

More information about the disease can be retrieved from the World Health Organization (a)'s web page.

1.3 Surveillance systems

Detection of outbreaks methods are one of the possible ways of analyzing data in surveillance systems. According to the World Health Organization (b),

> public health surveillance is the continuous, systematic collection, analysis and interpretation of health-related data needed for the planning, implementation, and evaluation of public health practice. Such surveillance can: serve as an early warning system for impending public health emergencies; document the impact of an intervention, or track progress towards specified goals; and monitor and clarify the epidemiology of health problems, to allow priorities to be set and to inform public health policy and strategies.

A surveillance system in public health is the set of resources, structures and procedures used to achieve these objective. Two main parts can be distinguished in a surveillance system: the collection of data and their analysis and interpretation.

1.3.1 Data collection

In surveillance systems, data are collected in many ways and from many sources, and each of the existing data sources and each way of collecting data from them have different specificity, sensitivity and timeliness because of their different natures. As reflected on the graph from Cheng et al.

(2009), shown in Figure 1.1, there is not such thing as the best data source. Data sources where data are collected faster and cover a wider amount of population are usually associated with lower specificity. A useful data source will balance these three characteristics, taking into consideration that the suitability of a data source for a surveillance system depends on the particular relevance of each of these three factors for the purpose of the surveillance. The type and format of the available data, as well as their expected sensitivity, specificity and timeliness will shape the analysis methods used on them.

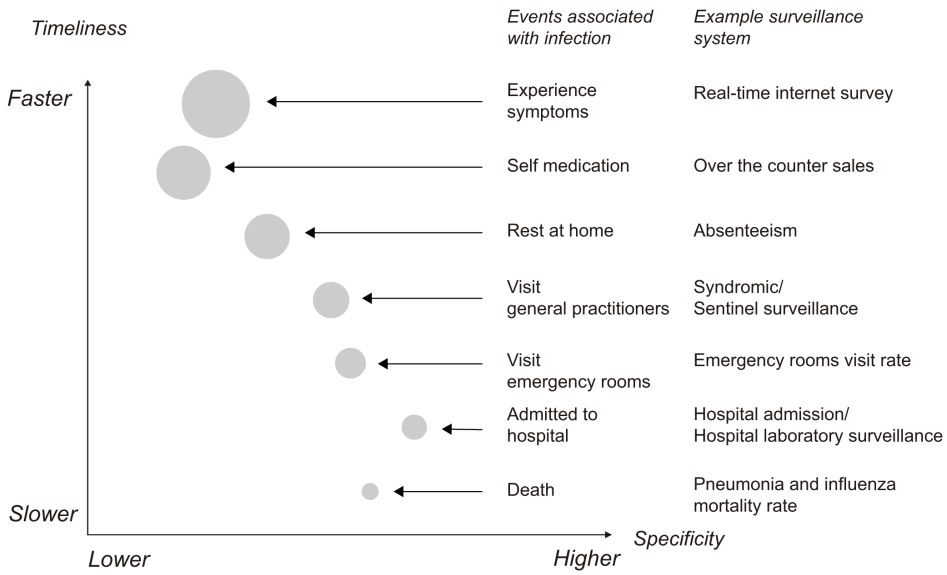

Figure 1.1: Schematic diagram of the course of illness and clinical iceberg of upper respiratory infections in a population, and examples of surveillance systems targeting each stage. Taken from Cheng et al. (2009).

Let us review some of the most common data sources for influenza surveillance systems:

Mortality registers. The register of every patient is coded by trained

coders. The specificity is usually high, but the time to process the information until it is accessible is usually long. In any case, the proportion of people dying from influenza or its complications among its incidence is very low. Some authors like Serfling (1963) used pneumonia mortality rates, while others like Sebastiani et al. (2006) focused on influenza mortality for their research.

ER and hospital admissions. Influenza has mild symptoms in healthy young adults. Hospitalization or admission to emergency room usually happens only in children, elders and patients with complications or comorbidity. The proportion of incident patients that actually are admitted to one of these health services is low, and sometimes it does not happen until several days after the start of the first symptoms. Laboratory tests are frequently available. Several authors like Wieland et al. (2007) or Corberán-Vallet and Lawson (2014) use emergency room visits for the surveillance of respiratory related disorders.

General practitioners. There are two main ways in which data from influenza incidence may be gathered from general practitioners; notifiable diseases mandatory report or sentinel networks. Influenza is listed as a mandatory notifiable disease in some countries as Australia, India or Spain, while in others, like the United States, only certain types of influenza are notifiable. In other countries like Malaysia or United Kingdom influenza is not notifiable at all. Sentinel Networks (present in countries like Switzerland, United States and Spain) are a different structure in which general practitioners voluntarily enroll and report observed cases. In this system, practitioners may be more involved in the process and give higher quality reports, but the people assigned to the enrolled physicians cover just a fraction of the total population that is being studied. In both cases notifications are usually syndrome based, reporting influenza-like illness (which introduces noise to the data), as laboratory confirmation is usually not necessary for the sake of the patient's health. Time of reporting may depend on the way the practitioners communicate the cases. Traditional systems

usually collect data once every week, but new computerized systems allow to shorten this time and collect data each day.

Laboratory isolation. All previous data sources may be laboratory confirmed or not. It is frequent that blood or sputum samples are analyzed when a patient visits the emergency room or is admitted to a hospital, but it is more uncommon when visiting a general practitioner. In some cases, a fraction of the patients may be analyzed in a sentinel network for epidemiological reasons, as it is the case of the Valencian Sentinel Network (which we will use further on in an exemplification of performance of one of our proposals). Laboratory confirmations rise the specificity of data sources but in some cases may take several days. In Nunes et al. (2013) one can see an example of a detection algorithm for an influenza surveillance system that combines data rapidly retrieved twice a week from a sentinel network and laboratory estimations, more specific but with one week of delay.

Absenteeism. School and workplace absenteeism may be used as a data source to detect the outbreak of an influenza epidemic. The cause of the absence is not always available, and when it is, it may be self reported or medically certified ILI. Besides, unemployed or retired people and people attending study or work centers where attendance is not registered are not covered by this kind of data source. In any case, the coverage is wider than for the previous mentioned data sources, and the timeliness might be higher. Its specificity, though, is lower, as medical confirmation is not always required. Some examples of the use of absenteeism for influenza surveillance may be found in Paterson et al. (2011), Cheng et al. (2012) and Kom Mogto et al. (2012).

Drugs sales. Influenza may have mild symptoms and is a sufficiently well known illness that is not appreciated as a threat by healthy young or middle aged adults. For that reason many infected people might not attend the health centers and keep going to work or school, but many of them self medicate to cope with the symptoms. Over-the-counter

sales is a data source with higher timeliness and coverage but lower specificity, as it is often based in self-diagnosis. The association of over-the-counter sales data and several respiratory illnesses has been studied in several works like Goldenberg et al. (2002), Liu et al. (2013) and Magruder (2003).

Internet sources. Information technologies have opened new ways to obtain data susceptible to be used in surveillance systems. It is true that internet data sources usually present a bias of selection, as they mostly represent internet users, but the quantity of available information and its immediateness make internet a really interesting resource for surveillance.

One way of retrieving information from internet are online surveys. Influenzanet –see also Debin et al. (2013) and Vandendijck et al. (2013)– is a remarkable example in Europe, being a network of 9 national partners with a web page each. In them, volunteer people can register and, during the cold months of the year, are asked to weekly answer a review which takes only a few seconds if the individual has no symptoms. One advantage of this kind of data is that many aspects of the symptoms and possible risk factors may be surveyed. On the other hand, it depends on the commitment of the individuals to obtain the information.

Of special interest is the automated collection of data from social networks or search engines, which provides large amounts of data from large quantities of people without their purposeful collaboration. In this way, one can obtain almost immediate data from a wide territory and a vast amount of individuals.

Internet queries do not require any special commitment from the people that use them. Some search engines offer data from the use of some words by the users of the engine, like Google Trends. Based on this engine, a specific algorithm was created using queries about topics like influenza complication, cold/flu remedy, general influenza symptoms, etc., to estimate influenza rates of the official agencies of

different countries. This particular tool is Google Flu Trends and its algorithm is discussed in Ginsberg et al. (2009). But, after the A(H1N1)pdm09 epidemic on 2009, it became clear that the behavior of the search engine users is not stable and the algorithm needed to be reevaluated and adjusted, as reflected by the discussions of Cook et al. (2011). In fact, the algorithm for the USA was reassessed in 2009, 2013 and 2014, as shown in the work of Olson et al. (2013), among others. Nowadays data are available at the website but the algorithm does not offer new estimations since August 19th, 2015.

Another source of massive instant information is the social media. Social networks such as Twitter have been used to track the presence of influenza (see, for example, Broniatowski et al., 2013; Gesualdo et al., 2013; Li and Cardie, 2013 or Grover and Aujla, 2014). These and other internet-based data sources –and its associated surveillance systems– are discussed in Milinovich et al. (2014).

There are some drawbacks of search engines and social media as data sources compared to some other more traditional ones. One is the necessity of an algorithm that transforms a wide amount of raw data into something manageable and meaningful that actually provides information related to the incidence of the disease. This is usually done by statistical methods that try to predict data from other data sources of greater specificity but with a higher time of collection. Another disadvantage, deduced from the previous one, is the lower specificity of the data compared to that of some other traditional data sources. In any case, the advantages of the timeliness, the automaticity, the wide coverage of population and the low cost of collection make this kind of data sources attractive for an epidemic detection system.

Each data source will allow to achieve different objectives in a surveillance system. Retrospective studies will rely on data sources with high specificity, regardless of the speed of retrieval of data. On the other side, systems devoted to triggering epidemic alarms with recent data will focus on the timeliness, often sacrificing specificity and, sometimes, doing it willingly to broaden the scope of illnesses which epidemics could be detected.

1.3.2 Analysis methods

Once data are collected, they must be analyzed in order to extract relevant information that may be interpreted to achieve the goals of the surveillance system. The analysis may be just a qualitative analysis of the cases for a disease with very low prevalence, a simple numerical or graphical description, a mathematical or statistical algorithm or a combination of all of them. The nature and form of the analytic tools depend on the data and on the objectives of the surveillance system. Let us review some of the possible kinds of analysis that are performed in influenza surveillance systems:

Assessment of risk factors. One of the main objectives of public health is to detect the risk and protective factors for the diseases in order to be able to set health politics in accordance. This kind of analysis is the only one in the list that requires covariates other than space and time.

Disease mapping. The objective of this analysis is to plot maps where the incidence, prevalence or mortality of the illness is represented along a certain territory, in order to observe the differences between different regions. Statistical methods are usually applied in order to cope with variability of the data in regions with low population, with lack of data or when plotting maps in a continuous way instead of using a lattice. Some methods worth mentioning are kriging, an interpolation method for continuous data proposed by Krige (1951), and Besag, York, and Mollié (1991) (BYM) model for lattice data. Temporal dimension can also be taken into account in disease mapping and thus observe the evolution of the disease. Some examples of influenza disease mapping can be seen in Carrat and Valleron (1992), where kriging is applied on ILI data in France, or the work of Hu et al. (2015), who use the BYM model for influenza A(H7N9) in Shanghai.

Anomalous regions detection This analysis takes a step further from disease mapping. Besides plotting the behavior of the illness through

space (and maybe time), it also sets some rule or algorithm to detect particular locations with different behavior from the rest. By doing so, the method is able to detect locations with a dysfunctional reporting system, local outbreaks or locations exposed to some particular risk factors. An example of this kind of analysis is presented by Li et al. (2012), who present a methodology that detects locations with different temporal patterns from the global pattern of mortality, caused by chronic obstructive pulmonary disease (though the methodology may be used on several other diseases).

Clusters detection A step further in sophistication from anomalous regions detection is cluster detection, where sets of neighboring locations with a common differentiated behavior are located. It may help to detect epidemics or risk factors that spread among a region larger than the spatial unit of observation. Some extended methods for cluster detection are scan statistics, a methodology that has been improved in many ways since proposed by Naus (1965a,b). Several of the extensions are implemented in the software SaTScan™. Kulldorff et al. (2005) show an application of this methodology on syndromic data that is capable of detecting an unusually early influenza season in New York City.

Outbreak detection Outbreak detection's scope is to detect the beginning of an outbreak as soon as possible, maximizing sensitivity and specificity. Outbreak detection models always deal with temporal data, but can also work with spatio-temporal data. Unlike the three previous analysis, outbreak detection is necessarily prospective, so timely data sources are often used for them. Because of the prospective nature of this kind of analysis, the algorithms must be applied in an 'online' basis, that is, every time new data are obtained (usually every week or day) the algorithm has to be run anew to identify the possible outbreak as soon as possible. The new proposals for influenza outbreak detection presented in Chapters 2 and 3 are based on the model proposed by Martinez-Beneito et al. (2008a), though several

other methodologies for triggering epidemic alarms of influenza may be found in the literature.

Prediction of the severity of epidemics A different temporal approach is to try to predict the behavior of the epidemics, focusing not as much in when and where it will happen but in how large and virulent will it be. Predictions can be short term (one time unit ahead) or try to predict the whole epidemic curve. For example, Yang et al. (2015) forecast the peak timing and peak magnitude for individual influenza strains as well as for the aggregate epidemics in Hong Kong.

Many times one system performs several of these analyses at once. All of these analyses may involve covariates to determine which are the risk factors for the disease in study, to isolate spatio-temporal behavior not attributable to other factors and to improve detection and prediction.

This thesis is set within the framework of outbreak detection methods. In particular we propose two novel methods for the detection of influenza or ILI outbreaks. The one presented in Chapter 2 deals with temporal data, trying to detect as soon as possible the arise of influenza epidemics. The proposal in Chapter 3 is applied on spatio-temporal data in order to signal not only when, but also where the epidemic starts, and when and where it spreads.

Before going any further, it is interesting to point out why sometimes we talk about influenza surveillance and some other times about ILI (influenza-like illness) surveillance. Most of the influenza surveillance methods presented in the literature actually do surveillance on ILI data, as most data sources are based on syndromic diagnose and laboratory confirmation of the actual influenza virus is seldom performed. Because of this, one may wonder if these are influenza or ILI surveillance methods. On one hand, the main goal of the surveillance system usually is to detect influenza epidemics and ILI data are used only as a proxy of the disease. On the other hand, ILI data usually are used instead of confirmed influenza data and the detection of other ILI is a welcomed (or even intended) byproduct. Because of that, most authors speak of influenza surveillance or ILI surveillance almost in-

distinctly. From now on, we will refer to influenza or ILI surveillance in this sense.

1.4 Review of temporal analysis methods for the detection of influenza outbreaks

In order to understand how the new proposals of influenza outbreak detection methods presented in this thesis have been built and why have we constructed them as they are, it is interesting to understand previous methods present in the literature. This will allow the reader to appreciate which of their statistical tools have inspired or been used in our proposals and which limitations found in other methods are attempted to be solved by the proposals in Chapters 2 and 3. In this section we are going to review different temporal analysis methods of surveillance systems present in the literature, centering our view in those focused on the detection of outbreaks of influenza or ILI. We will be paying special attention to the evolution from the simplest alarm triggering algorithms to the Bayesian Markov switching models.

Some extended reviews of methods for the detection of outbreaks can be found in the works of Buckeridge et al. (2005), Le Strat (2005), Burkom (2007) and Unkel et al. (2012). Several of the models are implemented in the R package `surveillance` (see Höhle, 2007 for documentation about the whole package and Salmon et al., 2016 for more details about its temporal methods). Spreco and Timpka (2016) offer a brief metanarrative review of online algorithms for detecting and predicting influenza outbreaks. A particular review centered in the use of Markov switching models for the detection of influenza outbreaks can be found in Amorós et al. (2015).

1.4.1 Methods based on historical limits

Methods based on historical limits are the most widely used for detecting the onset of influenza epidemics and with longer tradition in the epidemiological literature. These methods consist of estimating a parametric

behavior of the non-epidemic observations in previous observations (from the present and/or past seasons) and using it to calculate an upper threshold. These models are based on the so-called 'control charts', proposed by Shewhart (1931) and originally used for quality control in industrial production. In control charts, a warning is triggered when the difference between the current observation and a theoretical mean of the process surpasses the mentioned threshold. This threshold is usually set using the estimated standard error of the observations under control (in the context of outbreak detection, non-epidemic observations). Methods based on historical limits are those which follow this idea, but may be more sophisticated when modeling the theoretical mean and threshold. One way to determine this theoretical mean and threshold is to consider a window of observations of times $t-m,\ldots,t-1$ from the present year and/or $t-m,\ldots,t+m$ times from previous years and compute some central estimator and standard error for the observations in these windows, as Stroup et al. (1989), Farrington et al. (1996), Fricker et al. (2008) or Boyle et al. (2011) do. Another option would be using all non-epidemic data as training and fitting a regression model which includes time trend and Fourier periodical terms as proposed by Serfling (1963):

$$Y_t = \mu + \alpha t + \sum \beta_i \cos\theta_i + \sum \gamma_i \sin\theta_i\,, \tag{1.1}$$

with θ_i a linear function of t. This harmonic cyclic regression method has been used for influenza surveillance by the Center for Disease Control and Prevention (CDC) of the United States (Muscatello et al., 2008). This approach or some variations of it are also used in other works like Costagliola et al. (1991), Costagliola (1994) and Simonsen et al. (1997).

Other works use generalized linear models instead of linear regressions, like Chan et al. (2015), where a Negative Binomial regression is applied on all the historical data, including several covariates like climate or day of the week and applying a simple threshold on the standardized residuals. Others use this idea in a more complex methodology, like Noufaily et al. (2013), who propose a decision algorithm based on the number of actual and previous observations that are missing, zero or other. Depending on

the observations, this algorithm directly indicates the epidemic or non-epidemic state or decides to apply a quasi-Poisson model based on a window of observations to estimate a threshold. Some other simple approaches assuming a non-Gaussian distribution of the data can also be applied, as in the work of Closas et al. (2012), where all previous non-epidemic data are assumed to follow an exponential distribution. For each new datum, the Kolmogorov-Smirnov test is used to test whether it comes from the same exponential distribution or not (this test could also be expressed as a threshold for the new datum).

These approaches have some drawbacks in practice, as pointed by Rath et al. (2003):

- First, most of these methods need a predefinition of epidemic and non-epidemic periods on the historic data, as they use only the non-epidemic observations to calculate thresholds. But that differentiation between phases is precisely the final outcome that detection methods want to achieve. Several approaches have been taken to distinguish the epidemic and non-epidemic states of the historic data. The works of Costagliola et al. (1991) and Costagliola (1994) declare as epidemic weeks in the previous seasons those weeks with more than 3 patients per doctor. Some methods use only data from the present season, making sure that data start from non-epidemic and stay like that for several weeks, as done in Cowling et al. (2006). Others, like Muscatello et al. (2008), use an arbitrary criterion on the laboratory confirmations. Vega et al. (2013) propose a method that searches for the length of the epidemic of each past season as the one that, if is made larger by one week, doesn't increment the proportional amount of influenza rate inside the period more than an arbitrary percentage between 2% and 4%. The variety of possible criteria found in the literature suggest arbitrariness in their definition. Because of that, it would be convenient to perform a sensitivity study to these criteria in all detection methods which require a predefinition of non-epidemic periods, or use detection methods that do not require this predefinition.

- In second place, contiguous observations in time are treated as conditionally independent values, while we would expect that their temporal arrangement could show some kind of dependence –such as streaks– that could be used to better distinguish between epidemic and non-epidemic stages.

- Thirdly, the baseline (non-epidemic) period is sometimes estimated with national data that maybe do not properly fit if we are mostly interested in a local influenza surveillance system.

Goddard et al. (2003) also point out as a fourth drawback that the use of temporally fixed threshold values to describe the levels of influenza activity can be misleading due to long-term time trends in consultations for influenza. Specifically, they pointed out a decline in the number of influenza-related consultations in recent years that could reduce the sensitivity of these methods.

The detection methods proposed in this thesis try to overcome these drawbacks, usually present in methods based on historical limits. In particular, the novel proposals do not require a predefinition of non-epidemic periods and use time series tools to seize the temporal dependency of influenza data. Both proposals also consider possible different behaviors for each year, which can assimilate long-term time trends as well as specific years with particularly different behavior on the epidemic and/or endemic incidence rates.

1.4.2 Cumulative control charts

Two different evolutions to the control chart proposed by Shewhart (1931) are the cumulative sum chart (CUSUM chart) and the exponentially weighted moving average chart (EWMA chart). Both of them accumulate information from previous weeks and the value that is tested against a threshold is not just the single observation for each week but the statistic that accumulates information from the present and previous weeks. In that fashion, growths that are not punctually sharp but a streak of mild upward

increases may also be detected, although on the other hand the power to detect sudden outbreak starts could be diminished.

CUSUM was proposed by Page (1954), and consists of assigning a statistic S_t to each time that can be expressed recursively as

$$S_t = \max(0, S_{t-1} + Y_t - \mu)\,, \tag{1.2}$$

with Y_t being the observed values and μ a theoretical mean value the process is supposed to have. When the value of S_t exceeds certain threshold, the alarm is triggered and usually the process is reset to zero. Certain variations on the way of calculating the statistic or the threshold may be done to improve the performance of the method. Some of its applications in influenza surveillance and epidemic detection may be found in Höhle and Paul (2008), Griffin et al. (2009), Höhle (2010), Sparks et al. (2010) or Boyle et al. (2011).

EWMA consists of the calculation of the moving average of the observations with exponential weights and can be expressed recursively as follows:

$$S_t = (1 - \rho)S_{t-1} + \rho(Y_t - \mu)\,, \tag{1.3}$$

with $\rho \in [0, 1]$. The threshold widens with t approaching to a limit set by k times the sample standard deviation, with k arbitrary. The use of EWMA, proposed by Roberts (1959), is less frequent than the use of CUSUM, but we can find some applications in surveillance like the dengue fever outbreak detection in Meynard et al. (2008). Woodall et al. (2006) provide a good review of the use of control charts in health-care and public surveillance that includes the use of cumulative control charts for the detection of epidemics.

Nevertheless, cumulative control charts can detect a relatively small shift of the mean behavior of the rates that extends during several days or weeks as epidemic weeks. The endemic behavior of influenza epidemics could have these shifts with the arrival of the cold months of the year, but they do not imply the beginning of the epidemic. A way to avoid this misclassification is using models which consider the behavior of the differentiated rates (the size and direction of the jumps from one week to the next one) as a differentiating feature and not only the size of the mean

value of the incidence rates. Our proposals take this approach so that a sharp increase or several contiguous mild increases are needed to trigger an alarm. More details on how this is done will be presented in Chapter 2 and Chapter 3.

1.4.3 ARIMA

In 1981, Choi and Thacker published a paper on the forecast of pneumonia and influenza using time series analysis instead of regression analysis, showing the power of this kind of models in the surveillance of influenza. The most popular methods in time series analysis are the autoregressive integrated moving average (ARIMA) models, which are linear models for observations on a discrete time support in which the predictors for each observation consist of previous (lagged) values of the variable concerned and/or previous forecast errors. There are three components that define an ARIMA model, let us see each of them separately.

Autoregressive model of order *p*. This model expresses each observation as a regression on the previous p observations with parameters ρ_i:

$$Y_t = c + \sum_{i=1}^{p} \rho_i Y_{t-i} + \varepsilon_t\,, \qquad \varepsilon_t \sim \mathrm{N}(0, \sigma^2)\,. \tag{1.4}$$

Moving average of order *q*. This model expresses each observation as the sum of a normal error on time t plus the previous q normal errors associated to the previous times multiplied by q parameters γ_j:

$$Y_t = \mu + \sum_{j=1}^{q} \gamma_j \varepsilon_{t-j} + \varepsilon_t\,, \qquad \varepsilon_t \sim \mathrm{N}(0, \sigma^2)\,. \tag{1.5}$$

Integrated process of order *d*. Also known as random walk of order d, in this model the d times differentiated observations are expressed as a normal error:

$$\nabla^d Y_t = \varepsilon_t\,, \qquad \varepsilon_t \sim \mathrm{N}(0, \sigma^2)\,, \tag{1.6}$$

where

$$\nabla Y_t = Y_t - Y_{t-1}\,, \qquad (1.7)$$
$$\nabla^d Y_t = \nabla(\nabla^{d-1} Y_t) = \nabla^{d-1} Y_t - \nabla^{d-1} Y_{t-1}\,.$$

The possible combinations of these three models are called $ARIMA(p,d,q)$ and can be expressed as follows:

$$\nabla^d Y_t = c + \sum_{i=1}^{p} \rho_i \nabla^d Y_{t-i} + \sum_{j=1}^{q} \gamma_j \varepsilon_{t-j} + \varepsilon_t\,, \qquad \varepsilon_t \sim \mathrm{N}(0,\sigma^2)\,, \qquad (1.8)$$

though usually the complexity of the models used in practice is low ($p,d,q \leq 2$).

Cowling et al. (2006), for example, do automated monitoring of influenza sentinel surveillance data by applying a random walk of order one with predefined variance on the means of the normal distribution that models the observations:

$$Y_t \sim \mathrm{N}(r_t,\sigma_1^2)\,, \qquad r_t = r_{t-1} + \varepsilon\,, \qquad \varepsilon \sim \mathrm{N}(0,\sigma_2^2)\,, \qquad (1.9)$$

with σ_1^2 fixed and σ_2^2 to be estimated by the model. Reis and Mandl (2003) opt for a model with weekly and yearly pattern and select a combination of autoregressive and moving average components for the residuals to monitor respiratory symptoms reported at an emergency department. Williamson and Hudson (1999) describe a two-stage monitoring system consisting of univariate ARIMA models and subsequent tracking signals from several statistical process control charts (Shewhart, moving average and EWMA) and applies it on data of several notifiable diseases. The integer autoregressive model, a variation of the autoregressive model for discrete data, firstly introduced by Al-Osh and Alzaid (1987) and McKenzie (1985), is used for influenza outbreak detection in Rao and McCabe (2016). In that work, the authors propose a rule of decision based on the p-value of the forecast which can also be applied with any other model that provides a probability distribution for the forecast. In general, the use of ARIMA techniques is frequently observed in combination with other statistical tools.

Both proposals presented in this thesis use autorregressive structures over time to capture the intrinsic temporal correlation of influenza rates. Different parameters and orders for the autoregressive structures are set on the epidemic and the non-epidemic phases, so that the two different temporal behaviors can be distinguished.

1.4.4 Change point

The detection of a lingering epidemic –as it is the case of influenza epidemics– implies the identification of a moment where the behavior of the time series changes. Frisén (2003) studies how a shift from the basal mean of a process μ_0 to a higher mean μ_1 is detected by different methods (including Shewhart, CUSUM and EWMA). Later on, Frisén and Andersson (2009) and Frisén et al. (2010) propose a different approach, where the change happens from a constant basal mean to a set of increasing means at a certain time j, so that $\mu_{t=1} = \cdots = \mu_{t=j-1} < \mu_{t=j} \leq \mu_{t=j+1} \leq \cdots \leq \mu_{t=T}$. Several ways to estimate this change point are discussed in the aforementioned works.

The methods proposed in this work take the idea of modeling two different behaviors and detecting the transition between them, but embody it in a different manner. First of all, Markov switching models (which will be explained in Section 1.6.2) are used to model the two different structures –epidemic and endemic– and the transitions among them. Besides that, the distinction between phases do not only consider the change from a low mean of the incidence rates to a higher or increasing one. Other features are also considered, as are the change in the variance of the data or the spatial correlation.

1.4.5 Temporal scan statistic

A scan statistic is a tool to detect clusters of occurrences in space and/or time. The basic idea is to propose several windows that cover a connected subset of the whole region and/or of the whole time period, and test which of these possible windows shows a behavior most opposed to the null hy-

pothesis of random observations by comparing the number of observations inside and outside the window. Several possible distributions for the null hypothesis can be considered, like Poisson or Binomial among others.

Though mostly used in spatial or spatio-temporal surveillance, the simplest versions of the scan statistic, proposed by Naus (1965a) and improved in Naus and Wallenstein (2006), is purely temporal. In order to use a scan statistic for detection of outbreaks, it has to be applied in an online basis, forcing the potential cluster windows to include observations from the last week or day, as done by Ismail et al. (2003). As the authors state, *the window size and the number of events expected to occur in an interval have to be predetermined and this can be done using historical information or externally set standards.* This arises again the problem of using historical data, already discussed in Section 1.4.1, or using external information, which we tried to avoid in our proposals.

1.5 The Bayesian approach

All the references reviewed in the previous section are framed in the frequentist paradigm, though the general methodologies are not necessarily bounded to this paradigm. In this section we will discuss the role of Bayesian statistics in epidemic surveillance.

1.5.1 The Bayesian paradigm

Every statistical model tries to express the behavior of some observable variable or variables through one or several parameters $\boldsymbol{\theta}$. The Bayesian paradigm considers every parameter as a non directly observable variable with a probability distribution. The current knowledge about the parameters is expressed by a prior distribution $p(\boldsymbol{\theta})$, that is set by the researcher and expresses some previous knowledge or lack of knowledge about the parameters. Given a set of observed data $\boldsymbol{Y}$ from the observable variable or variables and a model that relates them with the parameters, the knowledge about these parameters can be updated using the Bayes' Theorem. So, if the likelihood of the observed data given the parameters is denoted

as $p(\boldsymbol{Y}|\boldsymbol{\theta})$, we have that the posterior probability distribution for the parameters is

$$p(\boldsymbol{\theta}|\boldsymbol{Y}) = \frac{p(\boldsymbol{Y}|\boldsymbol{\theta})p(\boldsymbol{\theta})}{\int_{\boldsymbol{\Theta}} p(\boldsymbol{Y}|\boldsymbol{\theta})p(\boldsymbol{\theta})\mathrm{d}\boldsymbol{\theta}}\,. \tag{1.10}$$

As the denominator is just a constant that assures that the distribution integrates to 1, it is usual to express this relation as proportionality, without regards to this integration constant: $p(\boldsymbol{\theta}|\boldsymbol{Y}) \propto p(\boldsymbol{\theta})p(\boldsymbol{Y}|\boldsymbol{\theta})$. This is known as the 'learning process', and is the basic methodology to obtain information about the parameters from the observations.

An advantage of the Bayesian paradigm is the capability of doing inference on hierarchical models with ease. Suppose that the parameters in one given model can be split in two sets, $\boldsymbol{\theta_1}$ and $\boldsymbol{\theta_2}$, so that the distribution of $\boldsymbol{\theta_1}|\boldsymbol{\theta_2} \sim p(\boldsymbol{\theta_1}|\boldsymbol{\theta_2})$ is explicit in the model, and the likelihood of the observable data is dependent only on $\boldsymbol{\theta_1}$ given $\boldsymbol{\theta_2}$, that is $p(\boldsymbol{Y}|\boldsymbol{\theta_1}, \boldsymbol{\theta_2}) = p(\boldsymbol{Y}|\boldsymbol{\theta_1})$. Then, and knowing that $p(\boldsymbol{\theta_1}, \boldsymbol{\theta_2}) = p(\boldsymbol{\theta_1}|\boldsymbol{\theta_2})p(\boldsymbol{\theta_2})$, we can express the posterior distribution of the parameters as

$$\begin{aligned} p(\boldsymbol{\theta_1}, \boldsymbol{\theta_2}|\boldsymbol{Y}) &\propto p(\boldsymbol{Y}|\boldsymbol{\theta_1}, \boldsymbol{\theta_2})p(\boldsymbol{\theta_1}, \boldsymbol{\theta_2}) \\ &= p(\boldsymbol{Y}|\boldsymbol{\theta_1})p(\boldsymbol{\theta_1}|\boldsymbol{\theta_2})p(\boldsymbol{\theta_2})\,. \end{aligned} \tag{1.11}$$

This is the basis of the Bayesian Hierarchical modeling, that allows us to make inference about parameters within complex models in a relatively simple way, as a product of simpler probability distributions.

Bayesian methodology provides a unified theory for handling uncertainty, which makes it a very advisable tool for the decision-making process of a surveillance system. Specifically, Bayesian analyses enable to quantify whichever feature of interest of any variable in the model by means of its posterior distribution. This makes the Bayesian methodology to be perfectly suited for quantifying the probability of being in an epidemic phase at any given moment. Also, the way of doing Bayesian inference ensures that all the possible sources of variability considered in the model are integrated in the output of the posterior distribution of the parameters or the predictions.

Another advantage of the Bayesian approach is the ability to cope with inference on the parameters on models with complex structures (as are spatial statistics, hierarchical models or hidden Markov models, among others) in a relatively easy way. This is done thanks to the capability of splitting complex models in multiplication of several simpler probability distributions, as seen in Expression (1.11). The proposals presented in this thesis are hierarchical models with hidden Markov structures, and one of them has also spatial structure, so they bear the kind of complexity we have just mentioned. Thus, by being defined under the Bayesian paradigm, the proposals enjoy of the relative simplicity of inference as well as the other characteristics we have previously discussed.

1.5.2 Inference tools in Bayesian statistics

As analytic inference is often impracticable or impossible, simulation and/or approximation tools are required to obtain posterior probabilities. Let us review some of them:

Markov Chain Monte Carlo (MCMC). This type of algorithms are explained in Gilks et al., 1996. The idea is to construct a Markov chain whose limiting distributions corresponds to the posterior distribution of interest, and then iterate the chain to generate samples from it. Softwares like WinBUGS and OpenBUGS (Lunn et al., 2000), JAGS (Plummer, 2003), STAN (Carpenter et al., 2017), NIMBLE or BayesX use MCMC techniques to numerically approximate the posterior distribution of the parameters of a model. This numerical algorithms can also be programmed in computational languages such as R (R Core Team, 2016) or C++ (Stroustrup, 2013).

Sequential Monte Carlo (SMC). SMC, also called particle filters (see Del Moral, 1996), is an algorithm that allows to update the inference on the posterior distribution of the parameters of a Bayesian model when new data are added to the existent ones without having to redo the inference with all the data (as is the case when using MCMC or other methods).

Integrated nested Laplace approximation (INLA). Proposed by Rue et al. (2009), INLA is an approach for hierarchical latent Gaussian models which computes very accurate approximations to the posterior marginals using the Laplace approximation. INLA can overcome convergence problems that MCMC may show and significantly diminishes the computational time in general. Anyhow, models that can not be expressed as latent Gaussian models, like hidden Markov chains, for example, cannot currently be analyzed by INLA. A wide range of statistical models are fitted with the INLA approach in the R-INLA package in R.

Aproximate Bayesian computation (ABC). ABC (see, for example, Biau et al., 2013) is a relatively recent computational technique to approximate the posterior distributions of the parameters that only requires of being able to sample from the model (likelihood) $p(\cdot|\boldsymbol{\theta})$. The algorithm consists of simulating jointly

$$\boldsymbol{\theta}^* \sim p(\boldsymbol{\theta}) \qquad \text{and} \qquad \boldsymbol{z} \sim p(\boldsymbol{z}|\boldsymbol{\theta}^*) \tag{1.12}$$

until the distance from the simulated $\boldsymbol{z}$ to the observed data $\boldsymbol{Y}$ is less than a certain threshold. Repeating this process several times provides of simulations of the posterior distribution $p(\boldsymbol{\theta}|\boldsymbol{Y})$.

1.5.3 Bayesian models for epidemic surveillance

Bayesian studies are not new in surveillance literature, but in recent years there has been increasing interest in them. An example of the motivation for this change of paradigm can be seen in Charland et al. (2009), who work on the assessment of environmental risk factors to the peak week of influenza epidemic, using Bayesian inference because

> generalized mixed models can be particularly sensitive to the point estimates of the variance parameters [...]. Bayesian hierarchical models average over the uncertainty of the parameters of the model rather than using point estimates [...].

A first approach we may consider to do surveillance from the Bayesian perspective are Bayesian networks, models where variables are conditionally related in a way that their conditional distributions can be represented in a directed graph without directed cycles. In these models, the principle of hierarchical Bayesian models presented in Expression (1.11) applies, so that the joint distribution of the model can be expressed as the product of several relatively simple conditional probability distributions. Inference can be made in a Bayesian or frequentist way, depending on the definition or not of prior distributions for the parameters of the model which are not conditionally dependent on other variables. Some examples of Bayesian networks applied on influenza surveillance are the work of Cooper et al. (2004), who proposed a Bayesian hierarchical network for the detection of CDC category A outbreak diseases (including influenza), and the work of Sebastiani et al. (2006), which integrates information from several data sources through Bayesian networks to forecast the starting and peak of epidemics.

Statistical structures already discussed in previous sections can also be considered in terms of Bayesian hierarchical models, which allow for more complex structures that can adapt better to the nature of the data. In that way, Held et al. (2006) propose a Bayesian method where the observations Y_t are the sum of two components, the endemic one (En_t) and the epidemic one (Ep_t), as follows:

$$Y_t = En_t + Ep_t\,, \quad En_t \sim \text{Po}(\nu_t)\,, \quad Ep_t|En_{t-1} \sim \text{Po}(\lambda_t En_{t-1})\,. \tag{1.13}$$

The logarithm of the parameter of the endemic component ($\log(\nu_t)$) is modeled by a Serfling regression, a linear regression with sinusoidal periodical terms. The expected value of the epidemic component is proportional to the endemic one by a factor λ_t that may change at several changepoints. λ_t rising over 1 determines the outbreak.

Manitz and Höhle (2013) fit a Bayesian Negative Binomial regression as follows:

$$Y_t \sim \text{NB}(\mu_t, \nu)\,, \qquad \log(\mu_t) = \alpha_t + \beta \cdot t + \gamma_t + \boldsymbol{\delta}'\boldsymbol{X_t} + \zeta z_t\,, \tag{1.14}$$

where α_t has a temporal random walk structure, $\beta \cdot t$ captures the long term tendency, γ_t captures the seasonal pattern of length S (with $\gamma_t = \gamma_{t+S}$) and $\boldsymbol{\delta' X_t}$ represents the covariates effect. A threshold is computed from this regression and the moment $t+1$ is declared as epidemic if Y_{t+1} surpasses it. The extra covariate z_t is equal to 1 if the moment t was declared as epidemic using data from time 1 to time $t-1$, and 0 otherwise. In this way, ζz_t captures the amount of epidemic cases in time t. A model based on those of Manitz and Höhle (2013) and of Noufaily et al. (2013) (discussed in Section 1.4.1) can be found in Salmon et al. (2015), which deals with delay of notification. In this kind of data, Y_t is not completely notified at time t, but only a fraction of the total cases are reported, and one has to wait several weeks (or days) for this underestimation to be corrected by adding the cases that were not notified in time. Thresholds are corrected in this work to take the delay of notification into account.

In this section we have discussed the potential of the Bayesian paradigm in terms of handling uncertainty, obtaining results in form of probability distributions and doing inference on complex models. We also have reviewed some of its applications in epidemic surveillance. In the next section we will review the use of hidden Markov chains in the detection of epidemics. This tool has been used under the frequentist paradigm in simple models, but is mainly addressed through the Bayesian paradigm when constructing more complex hierarchical models. The following section shows how hidden Markov chains offer a powerful and intuitive way of addressing the epidemic detection issue.

1.6 Hidden Markov chains in epidemic surveillance

There are two main forms of hidden Markov chains used in epidemic detection; hidden Markov models (HMM; see, for example, Cappé et al., 2005) and Markov switching models (MSM; see, for example, Douc et al., 2004). We will consider them with a discrete support for time and taking a finite number of possible states (commonly two, representing the epidemic and

non-epidemic states), as they are usually seen in surveillance.

A hidden Markov chain is a stochastic model that considers a set of non observed variables Z_t (hidden states) and a set of observed values Y_t (observations), one for each time unit $t \in \{1, \ldots, T\}$, so that $\boldsymbol{Z} = \{Z_t\}$ is a Markov chain. For any time t the conditional distribution of the present state given the past states depends on just the previous state

$$p(Z_t|Z_1, \ldots, Z_{t-1}) = p(Z_t|Z_{t-1}) \tag{1.15}$$

and Y_t is conditionally dependent on Z_t.

The most common use of hidden Markov chains in the surveillance field assumes two possible values for the non observed variables, also called states: $Z_t = 0$ indicating that there is no epidemic in time t and $Z_t = 1$ when an epidemic is occurring. Since Z_t depends on just Z_{t-1}, which is also a binary variable, a 2×2 transition matrix is estimated with the probabilities p_{kl} of going from epidemic state k to l;

$$p_{kl} = P(Z_t = l|Z_{t-1} = k)\,, \qquad k, l \in \{0, 1\}\,, \tag{1.16}$$

that is, the probabilities of staying or changing the epidemic state, from any time $t-1$ to the next time t. The transition matrix is therefore expressed as follows:

$$\boldsymbol{P} = \begin{pmatrix} p_{00} & p_{01} \\ p_{10} & p_{11} \end{pmatrix} \qquad \text{with} \qquad p_{kl} \in [0, 1]\,. \tag{1.17}$$

HMM are particular cases of hidden Markov chains with an added restriction, so that the value of the observed variable at each time Y_t is only dependent on the hidden state for that time, given the past observations and the present and past states

$$p(Y_t|Z_1, \ldots, Z_t, Y_1, \ldots, Y_{t-1}) = p(Y_t|Z_t)\,. \tag{1.18}$$

In a MSM, the states are also a Markov chain so they comply with the Expression (1.15) but, in this case, the restriction in Expression (1.18) is not accomplished. The observations Y_t are dependent on previous observations in $\boldsymbol{Y}$ instead, usually through an autoregressive process, and the

present state Z_t affects both the present observation and the relation between the present and the past observations. As Lu et al. (2010) state, *this setting makes the Markov switching model more suitable for time-series-related problems.* In Figure 1.2 there are two illustrations of the conditional dependencies for these two structures that may be enlightening.

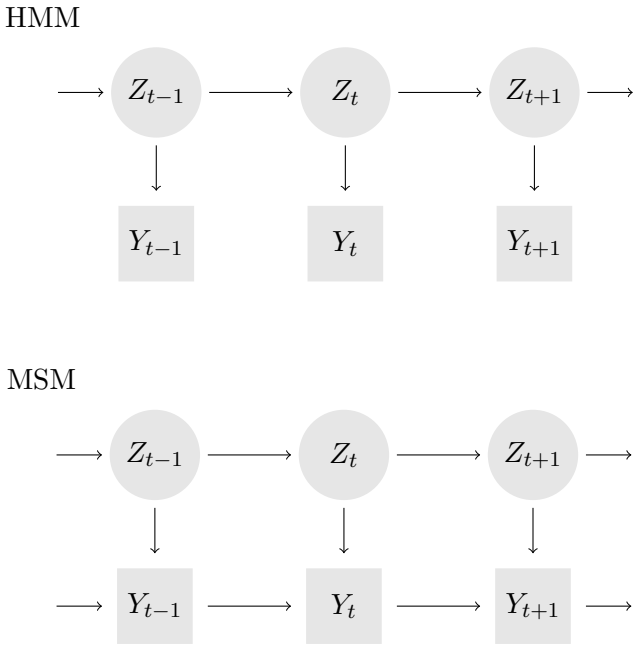

Figure 1.2: Diagrams of the conditional dependencies in a hidden Markov model (HMM) and a Markov switching model (MSM).

Once the model is determined, the values of the states and all other parameters have to be computed. In the frequentist paradigm, Viterbi algorithm (1967) is often used for computing the most likely state for each Z_t in HMM, while the transition matrix is usually estimated by the Baum-Welch algorithm (see Baum et al. 1970). Posterior probabilities for the states and the transition matrix are usually estimated by MCMC in the Bayesian approach, both for HMM and MSM. Due to their complexity

MSM models are usually fitted under the Bayesian paradigm.

An advantage of using hidden Markov chains for the detection of outbreaks is that the model classifies all past incidence rates or counts as epidemic or endemic, and use all that information for the decision in the present time. In that way, it avoids the problem of having to define which part of the historical data is endemic and which is not with an external criterion, and also avoids to dismiss the information of the epidemic historical data. Besides that, a model with a hidden Markov structure does not model just the non-epidemic behavior, as many other models do, but also specifies how epidemic data behave. In that way, the classification as endemic or epidemic is not done by accepting or refusing that data have certain structure but by discerning which of the two structures data adapt better to.

1.6.1 Hidden Markov models for oubreak detection

Le Strat and Carrat (1999) propose the use of HMM for disease surveillance in 1999 under the frequentist paradigm. An implementation of this model is available in the R package `surveillance`. Variations of the model are proposed in the same article with a Serfling cyclic regression and tested on Influenza-like illness (ILI) and Poliomyelitis data. Two examples with Gaussian or Poisson errors are described where the observations are modeled in one of these two ways:

$$Y_t \sim \mathrm{N}(\mu_{tZ_t}, \sigma^2) \qquad \text{or} \qquad Y_t \sim \mathrm{Po}(\lambda_{tZ_t})\,, \tag{1.19}$$

and where the parameters $\eta_{tk} = \mu_{tk}$ or $\eta_{tk} = log(\lambda_{tk})$ respectively are expressed by an harmonic regression

$$\eta_{kt} = \alpha_k + \beta_k \cdot t + \gamma_k \cos\left(\frac{2\pi t}{r}\right) + \delta_k \sin\left(\frac{2\pi t}{r}\right), \quad k \in \{0, 1\}\,, \tag{1.20}$$

with r the seasonal period. This kind of structure is not only used in influenza surveillance, as can be seen in Rafei et al. (2012) where a simplification of the same Poisson model is applied on tuberculosis patient registries

data. In this simplification, all the parameters of the regression are equal for η_{0t} and η_{1t} except for α_1, which is $\alpha_0 + \alpha_e$, with α_e the estimated shift on the expected counts caused by the epidemic.

A seasonal structure is not always considered to be the best baseline regression in HMM, as is reflected by Rath et al. (2003). They argue that *non-epidemic rates are always positive and approximately exponentially distributed*, as is appreciated in an exploratory histogram of non-epidemic ILI they perform. Following this idea, they propose a 2-state HMM with an exponential distribution for non-epidemic rates and a Gaussian distribution for the epidemic state, without linear or seasonal trend:

$$Y_t \sim \text{Exp}(\lambda) \quad \text{when } Z_t = 0 \quad \text{and} \quad Y_t \sim \text{N}(\mu, \sigma^2) \quad \text{when } Z_t = 1\,. \tag{1.21}$$

A comparison with the seasonal Gaussian model by Le Strat and Carrat shows that Rath et al.'s algorithm is more cautious in the sense that does not trigger the alarm too soon, does not maintain it too long when the epidemic has ended, and avoids false alarms in small peaks better.

Concerned on the problem of multiple testing under dependence, as happens when deciding whether each week is epidemic or not in a HMM, Sun and Cai (2009) propose a decision tool based on a new statistic called local index of significance instead of using p-values. They apply it on a HMM with a Gaussian distribution for the non-epidemic state and a sum of several Gaussian distributions for the epidemic state:

$$\begin{aligned} Y_t &\sim \text{N}(\mu_0, \sigma_0^2) && \text{when } Z_t = 0\,, \text{ and} \\ Y_t &\sim \sum_{l=1}^{L} \text{N}(\mu_l, \sigma_l^2) && \text{when } Z_t = 1\,. \end{aligned} \tag{1.22}$$

The amount of Gaussian distributions to be added for the epidemic state (L) is choosen by the Bayesian information criterion (BIC), proposed by Schwarz (1978). This model is tested on ILI sentinel data of France.

Some Bayesian HMM for influenza surveillance are simple adaptations of frequentist approaches. That is the case discussed by Madigan (2005), where Le Strat and Carrat's model is completed with non-informative prior

distributions for the parameters. A three states extension of this model is also proposed, where an intermediate state is considered. Madigan also proposes some variations of Rath et al. (2003) modelization, considering both epidemic and non-epidemic states with the same distribution. He compares the Gaussian, lognormal, gamma and exponential modelings using the deviance information criterion (DIC, introduced by Spiegelhalter et al., 2002). The DIC compares the goodness-of-fit of Bayesian nested models penalizing for overparametrization (see Section A.2 of the appendices for more details). The comparison finds the lognormal model scoring markedly better. In any case, Madigan himself states that he has not included Rath et al.'s model in the comparison process, as only models with the same distribution for the epidemic and the non-epidemic phase were taken into account.

1.6.2 Markov switching models for the detection of influenza outbreaks

In order to adapt better to the nature of epidemic data, it is a common practice to assume conditional dependency between close observations in time that have the same epidemic state. To do so, one must disregard the condition in Expression (1.18) of conditional independence of the observations in a HMM, which allows to open the toolbox of MSM. A recent work of Rafei et al. (2015) compares, under the frequentist paradigm, the performance on detection of tuberculosis epidemic outbreaks of several combinations of Gaussian Serfling harmonic cyclic regression, harmonic regression with temporal autoregressive dependence of first order of the observations and HMM. R^2 and BIC measures indicate the model with the three components (actually a MSM for including the autoregressive structure with changing parameter for each epidemic state) to be the one that better adjusts to the data.

Models with higher complexity are usually addressed under the Bayesian paradigm as they usually require of hierarchical structures of higher order. This is the case of Lu et al. (2010), who present what they called a Markov switching with jumps model for the detection of epidemic out-

breaks which includes several terms: the median of the means of a window of size 7 (a week size) for the 3 previous years $g(\mathbf{Y^{t-1}})$, a random jump that captures sporadic extreme values with Bernoulli probability $\xi_t J_t$, ($J_t \sim$ Bernoulli, $\xi_t \sim$ Normal), an autoregressive process X_t which depends on the epidemic state and includes day of the week $\sum_{i=1}^{6} w_i D_{ti}$, exogenous covariates effect $\sum_{j=1}^{K} b_j V_{tj}$ and a Gaussian error $\varepsilon_t \sim N(0, \sigma_\varepsilon^2)$:

$$Y_t = g(\mathbf{Y^{t-1}}) + \xi_t J_t + X_t\,, \tag{1.23}$$

$$X_t = a_{00} + a_{01} Z_t + (a_{10} + a_{11} Z_t) x_{t-1} + \sum_{i=1}^{6} w_i D_{ti} + \sum_{j=1}^{K} b_j V_{tj} + \varepsilon_t\,.$$

Conjugated priors are used for the model fitting. The model is tested on simulated data of over-the-counter sales for gastrointestinal disorders and on real-world clinic visits for respiratory syndrome with an added simulated anthrax outbreak. Each observation Y_t in this model is dependent of Y_{t-1} through the autoregressive structure. The state variables Z_t, which form a hidden Markov chain, modify the dependence between Y_t and Y_{t-1} by modifying the autoregressive parameter. This makes this model a perfect example of Markov switching model applied on ILI-related data.

A particular line of development is the creation of innovative structures that adapt to non conventional data sources, for example to sources of multivariate data. Usually these models are built ad hoc and would take considerable effort to adapt them to be applied on different data sources than those they were built for. Another novelty that may be introduced in hidden Markov chains for surveillance is time-changing transition matrices, that is, transition matrices which vary depending on time and on the observations. An example that incorporates these two characteristics is the model proposed by Nunes et al. (2013). This work is based on the work developed by Paroli and Spezia (2008), using in an algorithm that involves partial sentinel ILI observations $Y_{t(t)}$, complete sentinel ILI observations taken half a week after $Y_{t(t+1)}$ and laboratory influenza isolations taken half a week before $\nu_{t-1(t)}$.

The model on the observations taken half a week after has cyclic terms, a Gaussian error with higher variance in the epidemic state plus a polynomial

function of the partial observations which takes a different form according to the epidemic state:

$$Y_{t(t+1)} = \mu + \beta_1 \cos\left(\frac{2\pi t}{52}\right) + \beta_2 \sin\left(\frac{2\pi t}{52}\right) + \xi_{Z_t} + \varepsilon_{tZ_t}\,, \tag{1.24}$$

with $\varepsilon_{tZ_t} \sim \mathrm{N}(0, \sigma^2_{Z_t})$ and $\sigma_0 < \sigma_1$. ε_{tZ_t} is the polynomial function of the partial observations $Y_{t(t)}$, which is linear for the non-epidemic state and quadratic for the epidemic state. The later quadratic structure can better describe the step increase or decrease of rates during the epidemic. The polynomial term is set to zero for the non-epidemic state when the number of laboratory confirmations is high, that is, when they surpass an empirically set value ν_0 (set to 20 for the application in the article). In this way, when laboratory data grant that the process is in the epidemic state, only the epidemic option of the polynomial function is available to capture the relation between the partial observations $Y_{t(t)}$ and the complete observations $Y_{t(t+1)}$. The epidemic state of this term is also set zero when the number of laboratory confirmations is below another threshold ν_1 (in the article, when $\nu_{t-1(t)} < 1$). This ensures that when laboratory data show no evidence of epidemic whatsoever, only the endemic polynomial structure is available:

$$\begin{aligned} \xi_0 &= \theta_0 Y_{t(t)} \mathrm{I}(\nu_{t-1(t)} \leq \nu_0),, \\ \xi_1 &= (\theta_1 Y_{t(t)} + \theta_2 Y^2_{t(t)}) \mathrm{I}(\nu_{t-1(t)} \geq \nu_1)\,, \end{aligned} \tag{1.25}$$

with $I(\cdot)$ the indicator function that takes value 1 if its argument is true and 0 otherwise. Three different models are proposed for the transition matrix:

1. p_{kl} (the usual one, constant for all t),
2. $\mathrm{logit}(p_{t,kl}) = \alpha_{kl} + \beta_{kl} Y_{t(t)}$,
3. $\mathrm{logit}(p_{t,kl}) = \alpha_{kl} + \beta_{kl} Y_{t(t)} + \gamma_{kl} \nu_{t-1(t)}$,

with $k, l \in \{0, 1\}$ and $k \neq l$ (p_{kk} is set by complementarity). Epidemics are shown to be detected 2 weeks earlier in models with changing transition matrix (proposals 2 and 3).

Other ways of adding complexity on hidden Markov chain models to take advantage of a particular case of multivariate data can be found in spatio-temporal applications. We will review this kind of models in Section 3.3.4.

In the next section we will do a deeper analysis of a particular MSM by Martinez-Beneito et al. (2008a). The reason to discuss this model in detail is because in the present work we will present two extensions of this model, one for temporal data and one for spatio-temporal data. Knowing the structure of the model as well as its performance, advantages and limitations is key to be able to properly develop and understand these new proposals.

1.7 A Bayesian Markov switching model for the early detection of influenza epidemics

In 2008, Martinez-Beneito et al. introduced a Bayesian MSM for the detection of influenza outbreaks based on the analysis of temporal data of disease rates. One of the particular characteristics of this model is the fact that the data used on the model are not rates themselves, as those on Figure 1.3, but the difference of rates from time $t-1$ to time t, as seen in Figure 1.4. As the authors state, for the first-order differentiated series:

> The non-epidemic dynamic is characterized by small random changes around zero, while in the epidemic dynamic changes are greater and inter-related (positive and negative values are usually followed by positive and negative values, respectively). But more importantly, the fact that this new series has a zero mean allows us to restrict our study to its variability at each moment, which is expected to confer on our model a reasonable discriminatory capability, while in the former [non-differentiated] series we had differences not only in the variability but also in the means, thus making its analysis more complex. This comment is in line with that of Baron (2002), who also used differenced series to distinguish between epidemic and non-epidemic phases with

the final aim of early detection of the start of an influenza epidemic. To do so, he proposes the use of hierarchical Bayesian change-point models instead of Markov switching models.

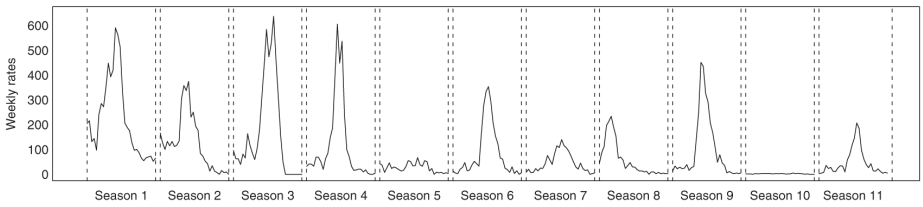

Figure 1.3: Weekly ILI incidence rates per 100 000 inhabitants during seasons from 1996–1997 to 2006–2007 of the Valencian Sentinel Network, as they appear in Martinez-Beneito et al. (2008a).

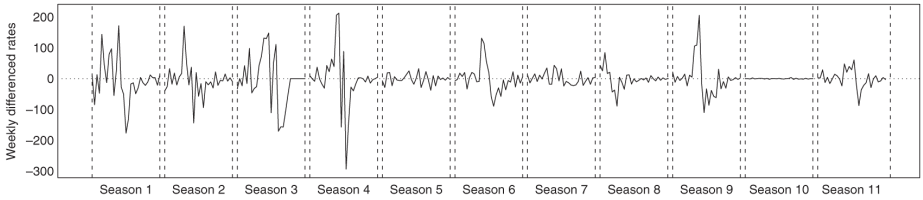

Figure 1.4: Time series of the differentiated rates per 100 000 inhabitants during seasons from 1996–1997 to 2006–2007 of the Valencian Sentinel Network, as they appear in Martinez-Beneito et al. (2008a).

Let us take a look at the model by Martinez-Beneito et al. (2008a). For convenience of notation, we will set the index for the first week of each season as $t = 0$. Let $\boldsymbol{y} = \{y_{ts}, t = 1, \ldots, T; s = 1, \ldots, S\}$ denote the set of differences between the rates of weeks t and $t - 1$ in season s:

$$y_{ts} = Y_{ts} - Y_{t-1s}\,. \tag{1.26}$$

A season may be a whole year or part of it, but starting and ending from dates that are expected to be in non-epidemic state, for example, in summer. In that way, all weeks from an epidemic share the same particular

parameters of the season it is in, and epidemics in different seasons can have some different parameters to capture the differences from season to season. For that reason, in the northern hemisphere, seasons are usually defined to start in summer or autumn of one year and end the next year after the winter has past.

The authors model the behavior of the differentiated rates time series for both epidemic and non-epidemic periods with two different structures depending on the epidemic state, indicated by a hidden variable Z_{ts}. This variable takes value 1 for epidemic and 0 for non-epidemic and, as it has been explained in Section 1.6, is a hidden variable in the sense that it is never directly observed. It seems reasonable to assume no underlying process beyond Gaussian noise for the non-epidemic period since, supposedly, no underlying mechanism should be inducing dependence among the observations. Meanwhile, the epidemic phase should show greater variability and, expectably, dependent observations in time. Therefore, the conditional distribution of y_{ts} is modeled either as a Gaussian white noise process or as an autoregressive process of order 1 depending on whether the system is in non-epidemic or epidemic phase. As the distribution of y_{ts} is conditional on Z_{ts} we will notate as $y_{tsZ_{ts}}$ this conditional distribution $y_{ts}|Z_{ts}$. The conditional distributions are made explicit as follows:

$$\begin{aligned} &y_{ts0} \sim \mathrm{N}(0, \sigma^2_{s0}), && t \in \{1, \dots, T-1\}, && s \in \{1, \dots, S\}, \\ &y_{1s1} \sim \mathrm{N}(0, \sigma^2_{s1}), && && \\ &y_{ts1} \sim \mathrm{N}(\rho y_{t-1s}, \sigma^2_{s1}), && t \in \{2, \dots, T-1\}, && s \in \{1, \dots, S\}. \end{aligned} \tag{1.27}$$

This model assumes a different variance for each season in order to reflect that the variability in any of the phases is not necessarily the same in different years, as a consequence of differences in the shape of the corresponding epidemic waves. Note also that the conditional distribution of the first difference of rates cannot be modeled as an autoregressive process as there is no previous value to condition on.

Taking into account that influenza epidemics last several weeks or months, it is sensible to consider a temporal dependency among epidemic states. The unobserved sequence of Z_{ts} are modeled to follow a two-state

Markov chain with transition probabilities:

$$p(Z_{ts} = l | Z_{t-1s} = k) = p_{kl}\,, \tag{1.28}$$

where $k, l \in \{0, 1\}$. This Markovian feature enables epidemic (respectively non-epidemic) weeks to be followed by epidemic (respectively non-epidemic) weeks with a high probability if data do not indicate otherwise. This performance could not be achieved with an independent modeling of the Z_{ts}'s and the Markovian structure makes the non-epidemic state to be more robust to sudden, although slight, changes in the differenced series. In summary, the distribution of y_{ts} depends on Z_{ts}, and the dependency structure of y_{ts} with previous observations is also dependent on Z_{ts}. Having that Z_{ts} follows a hidden Markov chain, this model is a MSM, as it has been described in Section 1.6.

Once the model is determined, the following step is to estimate its parameters. Martinez-Beneito et al. (2008a) propose using the following prior distributions for the parameters involved in the model:

$$\begin{aligned} \rho &\sim \text{Unif}(-1, 1)\,, \\ p_{11} &\sim \text{Beta}(0.5, 0.5)\,, \\ p_{00} &\sim \text{Beta}(0.5, 0.5)\,, \\ \sigma_{s0} &\sim \text{Unif}(\theta_{[1]}, \theta_{[2]})\,, \\ \sigma_{s1} &\sim \text{Unif}(\theta_{[3]}, \theta_{[4]})\,, \end{aligned} \tag{1.29}$$

where $\{\theta_{[1]}, \theta_{[2]}, \theta_{[3]}, \theta_{[4]}\}$ corresponds to the ordered sequence of the variables $\{\theta_1, \theta_2, \theta_3, \theta_4\}$ with prior distributions defined as follows:

$$\theta_j \sim \text{Unif}(a, b)\,, \qquad\qquad j = 1, ..., 4\,, \tag{1.30}$$

where a and b are hyperparameters to be fixed by the modeler, typically expressing vague prior knowledge.

Expressions (1.27) and (1.29) contain all the knowledge about the system but they do not yield analytical estimates. Therefore, computational methods like MCMC are necessary. WinBUGS software (see Spiegelhalter

et al., 2003) was used by the authors to compute the posterior distribution of the parameters of the model. See Martinez-Beneito et al. (2008a) for more details on the specific implementation of this model.

This model can be applied to any temporal data with real or estimated influenza or ILI incidence rates in temperate climate locations collected in a daily or weekly basis during several seasons. The authors show an application of the model on weekly influenza incidence rates per 100 000 inhabitants collected by the Valencian Sentinel Network from the Comunitat Valenciana, one of the 17 autonomous regions of Spain. These data were reported by around 30 volunteering practitioners from seasons covering from the 42nd week of one year to the 19th week of the next one. Data, shown in Figure 1.3, was available from 1996 to 2007. In order to check this model's performance and also to be able to compare it with other methods from the literature, Martinez-Beneito et al. (2008a) constructed a gold standard. It consisted of considering as epidemic weeks all of the weeks in a season from the first to the last one with at least one laboratory confirmation of influenza infection from the Valencian Sentinel Network. As shown in Figure 1.5, the method returns a probability between 0 and 1 of being in an epidemic phase for each week (the posterior means of the hidden variable Z_{ts}). It can be appreciated that weeks considered as epidemic by the gold standard are considered to be epidemic with a high probability by the detection method. Several rules of action can be adopted by health authorities thanks to the quantification of the probability of being in the epidemic state, but if a simple rule of thumb is required, one can always set the threshold of probability 0.5, splitting the weeks in two sets; those which are most likely non-epidemic and those which are most likely epidemic, as shown in Figure 1.6.

The relevance of all the components of the model is tested by comparing the deviance information criterion (DIC, see Spiegelhalter et al., 2002) of the complete model and three simplifications of it; same standard deviations for all seasons, avoiding autoregressive structure in epidemic phase and using a non-Markovian model. The complete model got notably better scores than the simplifications.

The first version of the area under the weighted ROC curve

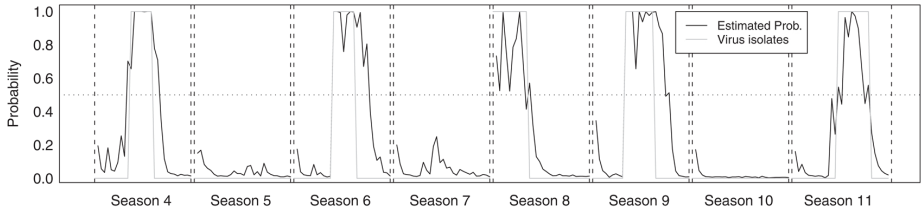

Figure 1.5: Estimated epidemic probabilities for seasons 4th to 11th and laboratory confirmations gold standard as they appear in Martinez-Beneito et al. (2008a).

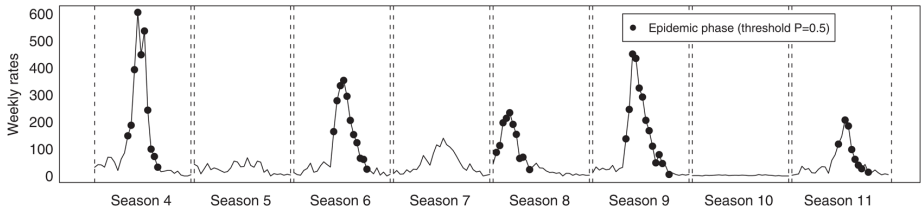

Figure 1.6: Influenza incidente rates for seasons 4th to 11th, as they appear in Martinez-Beneito et al. (2008a). Weeks with probability of being in epidemic state greater than 0.5 indicated with dots.

(AUWROC1) measure proposed by Kleinman and Abrams (2006) is used to compare in a unidimensional scale the sensitivity, specificity and timeliness of the methods in comparison. The aforementioned gold standard based in the laboratory virus isolations was considered for the construction of this score. More details on AUWROC1 measure can be found in Section A.3.2 of the appendices. The methods considered for the comparison are:

- the method proposed by Serfling (1963),
- a simplification of Le Strat and Carrat (1999) without temporal trend, implemented in the R package `depmix` (Visser, 2007),
- the Stroup et al. (1989) method without taking into account data from previous seasons and

- a Gaussian CUSUM method.

Martinez-Beneito et al. (2008a) method outperformed all other four methods. An implementation of this method is detailed in Conesa et al. (2009), where the architecture of systems of the web-based application FluDetWeb is explicit. This platform allows to upload data from any surveillance system and to obtain estimations of the probability of being in the epidemic state for each week without requiring advance knowledge on statistics, informatics or programming.

To summarize, in this chapter we have stressed the reasons to do influenza surveillance and how it is done. We have paid attention to the different data sources and reviewed the temporal analysis methods for the detection of influenza outbreaks present in the literature. Paying attention to their strengths and limitations has been key to developing the two new proposals for the temporal and spatio-temporal detection of influenza outbreaks that are discussed in detail in Chapter 2 and Chapter 3, respectively.

Chapter 2

A framework of temporal Poisson Markov switching models for the detection of influenza outbreaks

2.1 Introduction

In this chapter we present a framework of Bayesian hierarchical Poisson models with a hidden Markov structure for the detection of influenza outbreaks that generalizes the proposal of Martinez-Beneito et al. (2008a). A reduced version of this work has already been published in the journal *Statistical Methods in Medical Research* (Conesa et al., 2015).

As seen in the previous chapter, a large variety of statistical algorithms for the automated monitoring of influenza surveillance have been proposed. The most widely used approaches are based on historical limit methods or on Serfling's method, but these proposals have certain drawbacks: the need for a predefinition of epidemic and non-epidemic periods necessary to model the baseline distribution (which could condition the final results), and the fact that rates are treated as independent and identically distributed, ig-

noring temporal dependence inherent to this kind of data. It would be desirable, then, to use statistical models that are able to capture the temporal correlation of the data (by using ARIMA structures, for example) and which do not need a previous definition of the epidemic state of previous seasons (as is the case in hidden Markov chains).

The Bayesian paradigm may help to avoid these problems because of the capability of Bayesian hierarchical models of integrating in one model several complex structures like ARIMA, HMM, MSM or others and being able to do inference on these complex models in a relatively easy way. Also, the Bayesian methodology provides a unified theory for the handling of uncertainty, in which all the parameters are considered as variables with an associated probability distribution. This makes Bayesian models perfectly suited when quantifying the probability of being in an epidemic phase at any given moment. Following a Bayesian approach, Martinez-Beneito et al. (2008a) proposed a method that avoids the above-mentioned disadvantages while being able to quantify the probability of epidemic for each time point.

Two characteristics of the model by Martinez-Beneito et al. (2008a) proved to be very convenient in the surveillance context. The first one was the use of HMMs to segment the time series of influenza into epidemic and non-epidemic phases. HMMs are especially suited here as they can be applied to historical data without the need to make a previous ad hoc segmentation between epidemic and non-epidemic phases for a training period. The second issue was the use of the data variability to distinguish between both epidemic and non-epidemic phases, with non-epidemic dynamics characterized by small random changes and greater changes during the epidemic.

Nevertheless, although the variability allows both dynamics to be distinguished, incorporating the magnitude of the incidence can be very advantageous because this magnitude would also inform about the state of the illness, with high incidence clearly related to the epidemic phase. This would increase the capability of the method to distinguish between both epidemic and non-epidemic phases and therefore to determine the start of the epidemics. As a result, our main objective in this chapter is to introduce a model based on the proposal of Martinez-Beneito et al. (2008a) in which

we incorporate the magnitude of the incidence rates into the surveillance system. But this is not the only novelty; we are also considering rates as stochastic quantities instead of fixed and known quantities. In our opinion, this is a more realistic assumption, as rates are not data themselves but the quotient between a stochastic quantity (the observed cases) and the population. Therefore, rates are stochastic quantities and their variability should also be taken into account in the model. Our new proposal will emphasize the discrete nature of the weekly observed cases by giving them an appropriate discrete distribution.

The remainder of this chapter is organized as follows. First, a novel proposed framework of models for the detection of influenza outbreaks will be described. Secondly, the framework will be applied to a North Carolina Sentinel Network data set. The issue of the selection of the best model among those in the framework will be discussed and a variation to the proposal will be compared to the original one. The application of the model on a Valencia Sentinel Network data set will allow us to further discuss the problematic of the gold standard definition for the selection of the best model. In a last section, the framework of models will be applied on three Google Flu Trends data sets for further exemplification of the application of the new proposal when a gold standard is not available.

2.2 Modeling influenza time series for the detection of outbreaks

As discussed in Section 1.3.1, there are several types of surveillance data, and their format will vary according to their nature and how their collection is performed. Some statistical methods have been conceived for being applied to a specific type and format of data (although some of them could also be adapted to be used in other contexts). However, the adaptability of a statistical method to different surveillance system should be considered as an important feature of its potential usefulness. As a result, our intention is to develop a method that could easily be adapted to most kinds of surveillance data.

Therefore, our proposal uses only the type of data that can be obtained from most surveillance systems, which is a weekly/daily time series of counts, such as number of deaths due to influenza, number of hospital admissions, influenza-like illness (ILI) incidence cases, emergency phone calls received, etc. From them one can also obtain the rates by dividing the number of observed cases by the population of the region at each moment and multiplying it by a standardizing factor which is typically a power of 10.

An example of these series of surveillance data is shown in Figure 2.1, which displays ILI incidence rates obtained from the Valencian Sentinel Network (VSN), the same data source used by Martinez-Beneito et al. (2008a) in the application of their proposed detection method. Note that in this case we display two more seasons of data, and that data from season 10 have been corrected. The displayed rates for that season were divided by 10 by mistake on their article, an error of no major importance, as that particular season presented no influenza epidemic. Note that these series, as is usual in the context of influenza data, show a mixture of two dynamics: a non-epidemic dynamic in which the number of observed cases does not present big changes and varies around small values and an epidemic dynamic with higher rates in which the number of observed cases increases and decreases sharply.

One may also notice that this time series do not contain data from all the weeks of the season, but only from the 42nd week of one year to the 19th week of the following. Several surveillance systems, like this one, focus their attention and resources only on the cold months of the year, when temperate climate countries usually suffer of seasonal influenza epidemics. It is true that since the apparition of some non-seasonal strains of influenza, like the avian or swine flu, or some other diseases which share several symptoms with influenza, like SARS, some surveillance systems have changed their policies and started monitoring all year round, as previously discussed in Section 1.2. That has not been the case in many other ones though, which keep on interrupting the collection of data during the warm moths of the year. To take this possible feature of the nature of the data in consideration, our proposal will reinitiate all temporal dependencies the

first week of each season, including its hidden Markov chain.

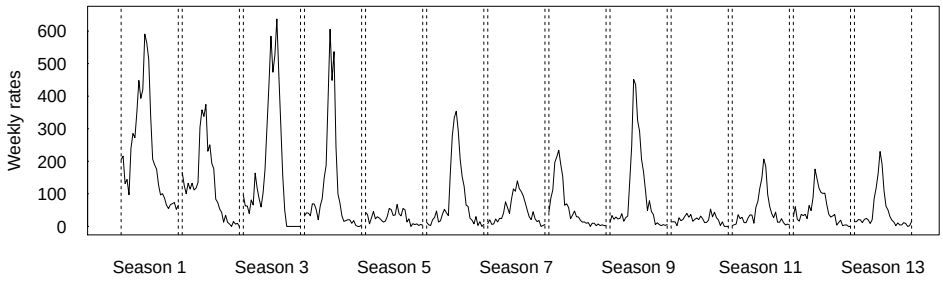

Figure 2.1: Weekly influenza incidence rates per 100 000 inhabitants during seasons from 1966–1997 to 2008–2009 in the VSN.

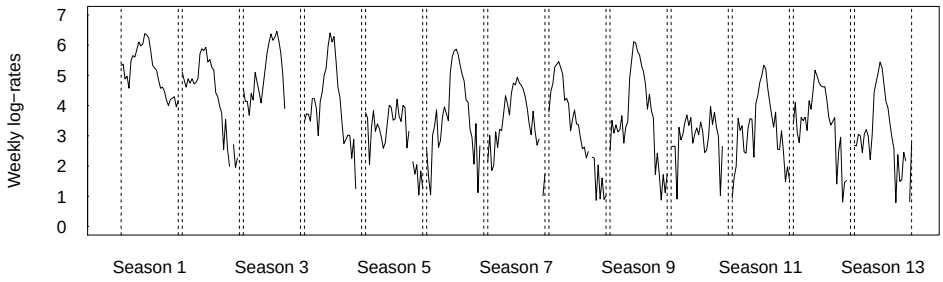

Figure 2.2: Weekly logarithm of influenza incidence rates per 100 000 inhabitants during seasons from 1996–1997 to 2008–2009 in the VSN.

2.2.1 First layers: Modeling the observations and rates

We introduce now the modeling of the data that we pursue along this chapter. As said in the introduction of this chapter, rates should be considered as stochastic quantities so we will not directly modeling the rates, which are subject to sampling variation that should be taken into account in our

proposal. Instead of that, we model them via the number of cases observed and the underlying population from which the observed cases have been reported. Let Y_{ts} denote the number of observed cases of influenza, ILI or another characteristic that could be used to track influenza, e.g. absenteeism during week (or day) t in season s. We model Y_{ts} by means of a Poisson distribution whose parameter is a function of the incidence rate r_{ts} of the week (or day) t in season s via the following hierarchical structure:

$$\begin{aligned} Y_{ts} &\sim \text{Po}(\nu_{ts})\,, \\ \nu_{ts} &= \frac{r_{ts}\text{Pop}_{ts}}{100000}\,, \\ r_{ts} &\sim \text{N}\left(R_{tsZ_{ts}}, \sigma^2_{sZ_{ts}}\right)\,, \end{aligned} \tag{2.1}$$

where Pop_{ts} represents the population under surveillance at the corresponding unit of time and r_{ts} is the corresponding influenza incidence rate. Note that the denominator in Expression (2.1) depends on the way the rate is defined. In this case we have expressed it considering the rates r_{ts} over 100 000 inhabitants. Note also that population can take different values for different weeks and seasons. For example, if data represent a whole region, population estimates are usually modified every year. If we were to deal with a sentinel network, population would refer to the number of patients assigned to the practitioners who are enrolled each week, and this can vary week to week, as practitioners can join or leave the network an any time. In any case, other expressions are also possible for ν_{ts}. Once the function which links the expected number of counts and the incidence rates is defined, we model these rates as a normal distribution in which both mean and variance depend on the epidemic state, determined by a variable Z_{ts}.

It is common when modeling the temporal behavior of a disease to model the logarithm of the rates as a normal distribution. Given the positive nature of the rates, the logarithm is able to transform the data allowing them to take values in the whole real line. Also, the growth of the incidence rate during an epidemic tends to be similar to an exponential growth. In this case we choose not to use the logarithmic transformation because it hinders the detection of the beginning of the epidemic. As can be appreciated comparing Figures 2.1 and 2.2, the values and the variability of the log-rates

are much more similar among them than those of the rates. The different variances between the two possible epidemic phases are a key point to distinguish between epidemic and endemic weeks, and the logarithmic scale drastically diminishes the difference in the variance of the data between these two phases.

It may be argued that the rates are bounded to be positive or zero, and that using a normal distribution is not adequate in this case. Strictly speaking, this is true, but temporal methods of detection as the one we are presenting are often used in aggregated data from a wide region, so the number of cases is high enough to use the normal approximation, even during the endemic phase. When applying the model to small regions, one may assume the approximation that has been proposed or alternatively opt by the use of some spatio-temporal method, which share information among small regions and thus would be more robust to variability of low rates. This will be further elaborated in Chapter 3.

2.2.2 Modeling the epidemic phase through a hidden Markov chain

To identify the epidemic state of each week, we model the variable Z_{ts} as an unobserved latent variable following a Markov chain –as described in Section 1.6– with two possible states, 1 for epidemic weeks and 0 for endemic (non-epidemic) weeks, and transition probabilities:

$$p(Z_{t+1s} = l | Z_{ts} = k) = p_{kl}\,, \qquad k, l \in \{0, 1\}\,. \tag{2.2}$$

Knowing that $p_{k0} + p_{k1} = 1$, we have that $p_{01} = 1 - p_{00}$ and $p_{10} = 1 - p_{11}$ so we only need to make inference over p_{00} and p_{11} for setting up the whole transition matrix.

As said before, for each season s, the Markov chain is reset, so inference must also be done on the initial probabilities:

$$p(Z_{1s} = k) = p_k\,, \qquad k \in \{0, 1\}\,. \tag{2.3}$$

We will consider these probabilities equal for every season. Knorr-Held and Richardson (2003) set the value of p_0 and p_1 to be equal to the stationary

distribution of the Markov chain:

$$p_0 = \frac{p_{10}}{p_{01} + p_{10}}\,, \qquad p_1 = \frac{p_{01}}{p_{01} + p_{10}}\,. \tag{2.4}$$

This is a general assumption in Bayesian analysis of hidden Markov models suggested in works like that of Robert et al. (2000). This assumption would make sense if the first week of each season was picked randomly among all the possible weeks. The truth is that most surveillance systems for influenza or ILI start tracking the disease in a time of the year where influenza is highly unlikely. Therefore, the probability p_0 of starting in a non-epidemic week will be way higher than the stationary distribution of the endemic state for the hidden Markov chain. Instead of using the stationary distributions, we prefer an independent prior distribution for p_0 and obtain p_1 by complementarity; $p_1 = 1 - p_0$.

With the objective of expressing our initial vague knowledge about all these probabilities, we use the usual Jeffreys non-informative prior distribution for Bernoulli trials for p_{00}, p_{11} and p_0:

$$p_{00}, p_{11}, p_0 \sim \text{Beta}\left(\frac{1}{2}, \frac{1}{2}\right)\,. \tag{2.5}$$

2.2.3 Modeling the variance of the rates

The value of the latent variable Z_{ts}, which we just described in the previous section, determines the distribution of the parameters of the Gaussian distribution of the rates. Let us expose now the proposal for the variance parameter $\sigma^2_{sZ_{st}}$. Non-epidemic dynamics are characterized by small random changes while epidemic dynamics show stronger fluctuations of the incidence rates. For this reason, we propose different values of the variance for each phase of each season, setting a lower variance in the endemic phase. Including the difference of variability of the observed data as a classifying feature of the model has proved to be very helpful for distinguishing

between epidemic and non-epidemic phases in the proposals of Martinez-Beneito et al. (2008a) and Nunes et al. (2013).

Different variances for each season have been assumed in order to reflect behaviors like the one observed in Figure 2.1, where it can be appreciated that the width of the variations is not the same for different seasons: some of them have higher and steeper peaks in contrast to others, with flatter epidemic waves. In any case, we constrain them all to ensure that all non-epidemic variances are lower than epidemic variances, regardless of being in different seasons. To achieve this, we have used part of the hierarchic structure used in the model by Martinez-Beneito et al. (2008a):

$$\begin{aligned} \sigma_{s0} &\sim \text{Unif}(\theta_{[1]}, \theta_{[2]}) \,, \\ \sigma_{s1} &\sim \text{Unif}(\theta_{[3]}, \theta_{[4]}) \,, \\ \theta_m &\sim \text{Unif}(0, a) \,, \qquad\qquad m = 1, ..., 4 \,, \end{aligned} \tag{2.6}$$

where $\{\theta_{[1]}, \theta_{[2]}, \theta_{[3]}, \theta_{[4]}\}$ corresponds to the ordered sequence of the variables $\{\theta_1, \theta_2, \theta_3, \theta_4\}$, and a is a hyperparameter to be fixed by the modeler, typically expressing vague prior knowledge. A way to ensure a vague enough prior distribution would be to set a as a higher than all the observed rates of the data value (it is slightly data-driven). Given that all the rates are positive, any sensible standard deviation will be lower than that value. Sensitivity analysis show that rising a from this value does not affect the posterior distribution of the parameters, and that it can even be lowered to facilitate convergence of the MCMC.

By defining the variances as we have just done, two problems are avoided; interchangeability and lack of identifiability. The indices 0 for endemic and 1 for epidemic will be correctly assigned to the two variances for each season and can not be interchanged thanks to the restriction of $\sigma^2_{s0} < \sigma^2_{s1}$. The variance parameters are also well identified even in the first weeks of the season because of the way θ_m parameters are defined. They determine two ordered intervals, and are informed by all the data from all the seasons. In that way, the first data for each season can be properly classified as epidemic or non-epidemic, as their variance will be closer to one of the intervals and farther from the other. If, for example,

the observed variance of the data falls closer to the interval $[\theta_{[1]}, \theta_{[2]}]$, this gives information suggesting that the data are more likely to come from the distribution with variance σ^2_{s0} and not so likely to come from that one with σ^2_{s1}. If no information were to be shared among seasons, no information would be available the first weeks of each season to determine whether data are supposed to come from a model with σ^2_{s0} or with σ^2_{s1}.

2.2.4 Modeling the mean of the rates

The next step is to model the mean of the rates in both states. This is one of the main novelties with respect to the model by Martinez-Beneito et al. (2008a). They modeled directly the differentiated rates while in this proposal we model the incident counts by means of the distribution of the raw (non-differentiated) rates. Note that R_{ts0} and R_{ts1} in Expression (2.1) represent the expected magnitude of the incidence rate r_{ts} at each phase and so we can take advantage of this to distinguish between both dynamics. Due to the temporal nature of the data, we propose an autoregressive structure for each one of them. Several options arise depending on the order of the autoregressive structure we pick.

Mean of the rates as white noise

The first and easiest way to model both means is to consider them as two different constants. We denote this modeling as AR0-AR0, the first term making reference to the structure of the non-epidemic rates and the second term being the one representing the epidemic phase. The AR0 notation stands for a zero-order autoregressive process, that is, a collection of normal independent variables. Therefore, for this model we define:

$$\begin{aligned} R_{ts0} &= \mu_0\,, \\ R_{ts1} &= \mu_1\,. \end{aligned} \tag{2.7}$$

We force the mean of the non-epidemic rates to be lower, $\mu_0 < \mu_1$, by the way we define their prior distributions:

$$\begin{aligned} \mu_0 &= \lambda_{[1]}\,, \\ \mu_1 &= \lambda_{[2]}\,, \\ \lambda_m &\sim \text{Unif}(0, b)\,, \qquad\qquad m = 1, 2\,, \end{aligned} \tag{2.8}$$

where again $\{\lambda_{[1]}, \lambda_{[2]}\}$ corresponds to the ordered sequence of the variables $\{\lambda_1, \lambda_2\}$, and b is a hyperparameter chosen to make the prior distribution for λ_m vague. Just as we recommended in Section 2.2.3, this hyperparameter may be set to be a superior to the highest rate in the data value, as the mean of both process will obviously be below that value.

Taking into account that μ_0 and μ_1 represent, respectively, the mean value of the rate in the endemic and epidemic phases, the phase with higher variability is modeled to be the same as the phase with higher mean. This makes sense as low incidences clearly indicate an endemic (and therefore less volatile) phase, while high incidences represent an epidemic (and consequently more volatile) phase. As a result, we now have two different features for distinguishing between epidemic and non-epidemic phases instead of just variability as in Martinez-Beneito et al. (2008a).

Note that the means for both phases in the proposal above do not change among seasons. If they were to be different, no information on the mean level for each period would be available to distinguish between phases at the beginning of a new season. This would often cause the triggering of false alarms as the model would not know the range of values for the mean of the endemic and epidemic rates and it would be easier to misclassify the first weeks of the season.

Means of the rates as autoregressive models of order 1

A second option is to consider the mean of the rates (in one or both periods) to be conditionally dependent on the previous observation. In this setting, the rates (of either the endemic or epidemic period) would be distributed around an unknown value as before (μ_0 and μ_1, respectively), but with the

addition that if the previous rate was below the mean value, the following would also be more likely to be below and vice versa. In that case, we model the mean of the rates as an autoregressive process of order one. As a result, three more models could be considered combining the AR0 and AR1 proposals, that is: AR0-AR1, AR1-AR0 and AR1-AR1. For simplicity, we present the model for AR1-AR1 in which both means are autoregressive processes of order 1:

$$\begin{aligned} R_{1s0} &= \mu_0\,, \\ R_{1s1} &= \mu_1\,, \\ R_{ts0} &= \mu_0 + \rho_0(r_{t-1s} - \mu_0)\,, && t > 1\,, \\ R_{ts1} &= \mu_1 + \rho_1(r_{t-1s} - \mu_1)\,, && t > 1\,. \end{aligned} \tag{2.9}$$

To define for example AR0-AR1, we would parametrize the non-epidemic rate as a white noise and the epidemic phase as a first order autoregressive process. No matter what combination of autoregressive orders for epidemic and endemic phases we use, the prior distribution for μ_0 and μ_1 remains the same as the one shown in Expression (2.8).

We propose flat prior distributions for the parameters of the autoregressive processes. As the non-stationary behavior of the series is intended to be modeled with the change between endemic and epidemic phases, we choose uniform distributions in the region where the processes are stationary in each of the epidemic and endemic modelings. For example, our selection in the AR1-AR1 model is:

$$\rho_0, \rho_1 \sim \text{Unif}(-1, 1)\,. \tag{2.10}$$

In order to ensure that the variance of Y_{1s} is also equal to the stationary variance of the series $\{Y_{ts}\}_{t=1}^{\infty}$, a correction is introduced on the standard deviations of each season's first week, so that the last line in Expression (2.1) for the first week is as follows:

$$r_{1s} \sim \text{N}\left(R_{1sZ_{1s}}, \frac{\sigma^2_{sZ_{1s}}}{1 - \rho_{Z_{1s}}}\right)\,. \tag{2.11}$$

This type of correction is seen, for example, in works like that of Martinez-Beneito et al. (2008b) in the context of spatio-temporal disease mapping. This correction makes the marginal variance for all the rates to be equal to $\frac{\sigma^2_{sZ_{1s}}}{1-\rho_{Z_{1s}}}$.

Means of the rates as autoregressive models of order 2

As a third option for the modeling of the rates, we may consider them (in one or both phases) to be related to rates from two or more previous weeks (or days). Then, a suitable option would clearly be to consider the mean of the rates as an autoregressive process of a higher order. Nevertheless, in this work, the order of the autoregressive processes is limited to be at most 2. This decision has been taken because the complexity of the models increases with the order of the processes and because, in the real applications considered in Section 2.3, the differences between AR0 and AR1 modelings are substantially higher than those between AR1 and AR2 (in terms of DIC and the classification of weeks into either epidemic or endemic periods). Nevertheless, there is no theoretical inconvenience in considering autoregressive processes of order higher than 2.

As a result, five more models can be considered: AR2-AR0, AR2-AR1, AR2-AR2, AR1-AR2 and AR0-AR2. For simplicity, we present the model in which both means are second-order autoregressive processes (AR2-AR2):

$$\begin{aligned}
R_{1s0} &= \mu_0\,, \\
R_{1s1} &= \mu_1\,, \\
R_{2s0} &= \mu_0 + \frac{\rho_{10}}{1-\rho_{20}}(r_{1s}-\mu_0)\,, \\
R_{2s1} &= \mu_1 + \frac{\rho_{11}}{1-\rho_{21}}(r_{1s}-\mu_1)\,, \\
R_{ts0} &= \mu_0 + \rho_{10}(r_{t-1s}-\mu_0) + \rho_{20}(r_{t-2s}-\mu_0)\,, \qquad t>1\,, \\
R_{ts1} &= \mu_1 + \rho_{11}(r_{t-1s}-\mu_1) + \rho_{21}(r_{t-2s}-\mu_1)\,, \qquad t>1\,,
\end{aligned} \tag{2.12}$$

with the same prior distribution for μ_0 and μ_1 shown in Expression (2.8). The distribution for the rates of the first and second weeks are defined so

that the autoregressive model is stationary. Any combination other than AR2-AR2 would be defined by giving to the endemic or the epidemic mean rates the AR0 or AR1 structure presented in Expressions (2.7) and (2.9).

Similarly to our choice for the autoregressive modeling of order 1, we choose a uniform joint prior distribution for (ρ_{1k}, ρ_{2k}) (for $k = 0$ or $k = 1$) on the set of values for this pair that guarantees that the autoregressive process of order 2 is stationary. As shown, for example, in Box et al. (1994), this region is bounded by the following constraints:

$$\begin{aligned} \rho_{2k} + \rho_{1k} &< 1\,, \\ \rho_{2k} - \rho_{1k} &< 1\,, \\ -1 < \rho_{2k} &< 1\,, \end{aligned} \tag{2.13}$$

which restrict the values of (ρ_{1k}, ρ_{2k}) to the triangle represented in Figure 2.3.

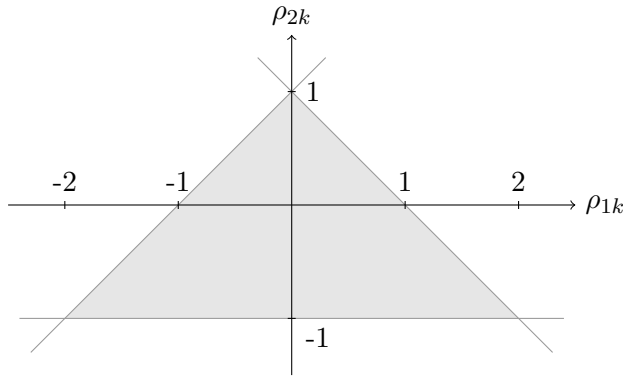

Figure 2.3: Region for the parameters where the autoregressive process of order 2 is stationary.

Again, the variances for the first and second weeks of each season are corrected so that they coincide with that of the stationary autoregressive

model (see, for example, page 62 of Box et al. (1994)). Thus,

$$\begin{aligned} r_{1s} &\sim \mathrm{N}\left(R_{1sk}, \gamma_k \sigma_{sk}^2\right) , \\ r_{2s} &\sim \mathrm{N}\left(R_{2sk}, \gamma_k \frac{(1-\rho_{2k})^2 - \rho_{1k}^2}{(1-\rho_{2k})^2} \sigma_{sk}^2\right) , \\ \gamma_k &= \frac{(1-\rho_{2k})}{(1+\rho_{2k})((1-\rho_{2k})^2 - \rho_{1k}^2)} , \end{aligned} \tag{2.14}$$

with k, which takes values in $\{0, 1\}$, being an abbreviation for Z_{1s} or Z_{2s} respectively.

It is worth to note that the selection of the best of all the modelings for the rates presented above becomes an important and sensitive issue throughout this framework. Clearly, this will depend on the kind of data being analyzed, and we will discuss it further in Sections 2.3, 2.4 and 2.5.

2.2.5 Inference on the model

All previous expressions in this chapter contain all our knowledge of the system but, as usual with hierarchical models, there is no analytic expression for the posterior distribution of the parameters. Hence, in order to make inference about the proposed models, we have to resort to numerical methods. Namely, we have resorted to MCMC using the software WinBUGS. The code of the model can be found in Appendix D.

The output obtained for each variable and parameter of interest from the MCMC is a set of 1000 simulations of the posterior probability. From them we can extract any statistical estimator that we want, including posterior means and 95% credible intervals. In particular, we can obtain the posterior mean of all the state variables Z_{ts}, which are estimations of the posterior probabilities of being in an epidemic for each week of the analyzed period.

Nevertheless, our real interest when we apply the detection method is not knowing which phase the system has been in previous weeks, but to know instead which phase the system is in during the last week of the available data, the week in which we perform the analysis. Because of that,

every week we have to rerun the model adding data from the new week. In order to compare the performance of a model in terms of detection in the most realistic way one must mimic this way of applying the model. To do so, we must run the model as many times as weeks we want to classify as epidemic or endemic using each time only data from the same and previous (but not subsequent) weeks we are trying to classify. That is what is called applying the model in an 'online' basis, and it is intended to reproduce the detection of the epidemic's beginning in the very precise week it starts, the real setting where these algorithms are supposed to be useful.

2.2.6 Some considerations

One important issue that arises when dealing with HMMs and autoregressive structures is how to model the first weeks after a change between the two phases in models where both the endemic and epidemic mean rates are modeled with autoregressive structures. There are two possible options: using information from the previous weeks (although they are from another phase) or not using it. In other words, we may maintain the conditional distributions of the autoregressive structures or make the rates conditionally independent when there is a change in epidemic phase. We have tested both kinds of models, and the ones using information from previous weeks perform better in terms of sensitivity, specificity and timeliness (in particular, using the measures of performance described in Section A.3.2 of the appendices). Consequently, we have decided to use information from previous weeks when the epidemic phase changes.

Nevertheless, the cost of this is that the method sometimes gives small probabilities of being in an epidemic phase after the peak of the season. This happens because of the strong role the autoregressive structure plays in the model. In terms of the mean of the rates, reinitiating the autoregressive structure considers the first week after the change of epidemic phase as having the mean of the new epidemic phase, without taking into account the size of the previous rate. In the weeks after the peak of the epidemic, the rates are much higher than μ_0, so the model remains in the epidemic phase until the rates are lowered close enough to the endemic mean. However,

the model which does not reinitiate the autoregressive structure tends to classify a rate as non-epidemic if it is descending towards (getting closer to) the non-epidemic mean and has similar magnitude to the previous rate (even when it is not a low rate). That is the reason of the lack of sensitivity in some weeks after the peak.

The good side of this choice of modeling is that the same phenomenon happens when going from non-epidemic to epidemic phase. The change from endemic to epidemic phase usually happens in low rates, and the model which reinitiates the autoregressive structure requires the rates to be close to μ_1 to switch the epidemic phase of the HMM. However, the model which does not reinitiate the autoregressive structure is capable of correctly detect the change from non-epidemic to epidemic in a week which is similar in magnitude to the previous non-epidemic rate but which shows a clear uprising tendency towards the μ_1 value of the epidemic process.

Changing the subject, it is also worth noting that the model of Martinez-Beneito et al. (2008a) can be considered as a particular case of this framework, if we remove the first Poisson layer of the hierarchical model. More specifically, it corresponds to a limit situation of an AR1-AR2 model (first-order autoregressive process for the endemic phase and second order for the epidemic). In their model they considered the differentiated rates as their data, that is, they model $r_{ts} - r_{t-1s}$ in the first layer of their hierarchical model. They also considered these differentiated rates to follow a first order autoregressive model for the epidemic phase and a Gaussian white noise for the non-epidemic phase. Let us see that for the epidemic phase, the first order autoregressive model of the differentiated rates is equivalent to a second order autoregressive model of the raw rates. The AR1 hypothesis on the differenced rates mean:

$$\begin{aligned} r_{ts} - r_{t-1s} &\sim \mathrm{N}(\rho(r_{t-1s} - r_{t-2s}), \sigma^2_{s1}) \leftrightarrow \\ r_{ts} &\sim \mathrm{N}((1+\rho)r_{t-1s} - \rho\, r_{t-2s}, \sigma^2_{s1}) \leftrightarrow \\ r_{ts} &\sim \mathrm{N}(\mu_1 + (1+\rho)r_{t-1s} - \rho\, r_{t-2s} - (1+\rho-\rho)\mu_1, \sigma^2_{s1}) \leftrightarrow \\ r_{ts} &\sim \mathrm{N}(\mu_1 + (1+\rho)(r_{t-1s} - \mu_1) - \rho(r_{t-2s} - \mu_1), \sigma^2_{s1}) \leftrightarrow \\ R_{ts} &= \mu_1 + (1+\rho)(r_{t-1s} - \mu_1) - \rho(r_{t-2s} - \mu_1)\,, \end{aligned} \tag{2.15}$$

therefore, this is a limit case or the AR2 modeling for the epidemic phase, with $\rho_{11} = 1 + \rho$ and $\rho_{21} = -\rho$, which lies on the limit of the parametric domain of the AR2 process. For the non-epidemic, the equivalency is quite straightforward. The limit case of the AR1 when $\rho = 1$ is a first order random walk and, as we saw in Section 1.4.3, a first order random walk for the rates corresponds to a white noise for the differentiated rates by definition. Let us show it explicitly:

$$\begin{aligned} r_{ts} - r_{t-1s} \sim \mathrm{N}(0, \sigma^2_{s0}) \leftrightarrow \\ r_{ts} \sim \mathrm{N}(1 \cdot r_{t-1s}, \sigma^2_{s0}) \leftrightarrow \\ R_{ts} = \mu_0 + 1 \cdot (r_{t-1s} - \mu_0)\,. \end{aligned} \tag{2.16}$$

It has to be taken into consideration that in Martinez-Beneito et al. (2008a) the values μ_0 and μ_1 are not present, as they are lost in the differentiation of the rates, so the effect of these two estimated quantities on the discrimination of the epidemic state is a novelty in the framework of models we are presenting in this chapter.

Thus far, we have introduced a methodological framework that allows us to distinguish between epidemic and non-epidemic phases. In the following sections, we show how this framework of models can be used as an early warning method for influenza outbreaks, applying it on real data sets.

2.3 Application on North Carolina Sentinel Network data

In this section we show the application of the proposed framework of models for the temporal detection of influenza outbreaks. We apply the models on North Carolina Sentinel Network data, compare its performance with other models in the literature and discuss the problematic of the tools for comparing them.

2.3.1 The North Carolina Sentinel Network data

The Department of Health and Human Services of North Carolina, in the United States, publishes every year a report about the surveillance of influenza in the NC DHHS Influenza (Flu) Information website. The report includes the weekly ILI reports by the North Carolina Sentinel Network (NCSN) and the Influenza Virus Isolates identified by the State Laboratory of Public Health.

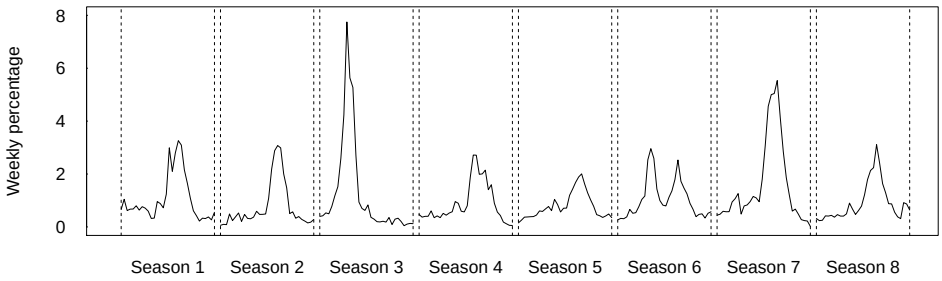

Figure 2.4: Weekly influenza incidence percentages during seasons from 2001–2002 to 2008–2009 of the NCSN.

ILI data on this surveillance system consist of the total amount of patients who visit the Sentinel Network practitioners and the counts of ILI diagnosed patients among them. We use this ILI counts as the model observations and the total amount of patients as the population. The rate of ILI incidence over visitors to the practitioners is reflected as a percentage, as shown in Figure 2.4. We modify the denominator in Expression (2.1) and divide by 100 to be coherent with that way of presenting information and because in doing so the model works with a nice scale of numbers with easy interpretation (values oscillating between 0.02 and 7.75 % for the rates). Virus Isolates are also presented as counts, split by virus strain. We use the aggregated counts of virus isolates for each week to build a gold standard of when the epidemic is likely to be occurring, which is useful to evaluate the performance of the models. Isolates data are missing for seasons 2001-2002

and 2004-2005. For each season, data are presented weekly starting from week 40 of one year and finishing in week 20 of the following year. For this illustration we are going to work with data from seasons 2001–2002 to 2008–2009.

2.3.2 Online and retrospective application of the model

As previously introduced, when applying the model on the available data one can take two approaches. The easiest way to do it, and which gives the best estimates is to apply the model in a retrospective basis. That is, to apply the model with all data only once, and get the estimation of the epidemic phase for all the weeks of all the seasons in the data. The more realistic approach is to mimic the way in which we would have applied the model if we were actually running the surveillance system in a real setting. In a realistic way we run the model once a week, adding each week the counts of ILI for the current week to the data we had from previous weeks and seasons. As we said before, that is what is called an online application of the model. In this second setting we focus on the estimated probability of the epidemic phase only for the latest available week. We are not specially concerned about the modified estimation of the epidemic state for previous weeks, as the main use of the model is knowing if the epidemic is starting now, according to the newest information available, or not. Of course, estimates obtained from a retrospective application of the model for other weeks than the current one are better, as both future and past data are used for estimating the epidemic states, but this way of applying the models is unrealistic.

In terms of computation time, at the time we ran the models they took between 10 and 20 minutes when applying the model in a retrospective basis, with the lowest times for the AR0-AR0 proposal and the highest times for the AR2-AR2 model. Calculations were performed in an Intel® Xeon® CPU E5530 with 16 cores at 2.40GHz and 48Gb of RAM, with OS Linux ubuntu 12.04 LTS. 3 parallel WinBUGS calls with 1 chain each were run in 3 different cores with 15 000 iterations of burning and 30 000 subsequent iterations. After thinning, 1002 iterations were kept, 334 from

each chain. Convergence was checked by observing the equivalent number of sample size, the $\hat{R}$ statistic (see, for example, Gelman et al. (2013)) and visual check of the chains of simulations.

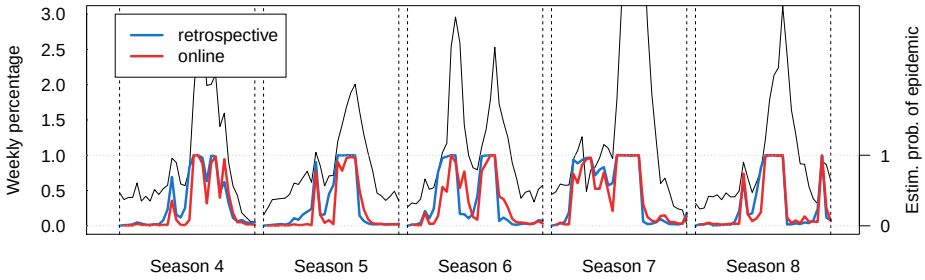

Figure 2.5: Retrospective and online estimated probability of being in epidemic phase according to the AR2-AR2 model on NCSN data. In black: weekly influenza incidence percentages during seasons from 2004–2005 to 2008–2009.

In Figure 2.5 we show the percentage of ILI and the estimates for the probability of being in the epidemic phase obtained from the AR2-AR2 model in retrospective and online basis. One can observe that the realistic way of applying the model results in more spiky estimates which also reach probabilities of 1 or 0 later, while the retrospective way shows smoother curves of estimates which rise and decay earlier thanks to using information from the past and the future for the estimation for each week.

2.3.3 Comparing models with DIC

The Deviance Information Criterion (DIC) is a typical choice for comparing models, therefore we use it as a first approach for the evaluation of the proposed models in our framework. This criterion measures the goodness-of-fit of a Bayesian model but penalizing by its complexity, having that the lower the score, the better the performance of the model. More details about DIC may be found in Section A.2 of the appendices.

Using DIC as a first evaluation tool is driven by the idea that models

which fit better the data might be estimating better the values of the variables Z_{ts} which determine the epidemic state. Therefore, DIC is only an indirect measure of the quality of detection, but it is an easy-to-use tool and computationally cheap, as it is applied to the retrospective version of the models. To compare the performance of several Bayesian models, they must be applied on the same data set. In Table 2.1 we show the DIC for the 9 different modeling proposals on the whole data set. We can observe that the best values are in those models with temporal autoregressive structure both in the epidemic and endemic phases.

	R_{ts1}		
R_{ts0}	AR0	AR1	AR2
AR0	2125.8	2105.7	2112.6
AR1	2116.6	2097.6	2100.2
AR2	2118.1	2092.9	2095.9

Table 2.1: DIC of the 9 proposals applied on the NCSN data. Higher values colored in red, lower values colored in blue.

2.3.4 Timeliness weighted ROC measures for the comparison of the models

Though DIC is an easy-to-apply tool for comparing models, it is not a direct measure of the epidemic outbreak detection performance. As any test that tries to classify entities into two possible categories, the detection of epidemics can be evaluated by measures of sensitivity (true positive rate) and specificity (true negative rate).

As those criteria usually depend on a certain arbitrary threshold that distinguishes negative from positive observations, the receiving operating curve (ROC) and the area under the ROC curve (AUROC) summarize the information of both measures for different values for that threshold. The ROC curve is drawn by joining the two dimensional points of sensitivity and 1-specificity that result in choosing all possible thresholds for the detection method (see Lusted, 1971, Egan, 1975 or Metz, 1978 for some classic

references). AUROC is the integral of that curve which would be ideally 1 for perfect methods assigning correctly all observations to either positive or negative for all the values of the threshold.

There is a third relevant measure in the field of outbreak detection, as not all the true positives are equally important in public health. Correctly classifying the first weeks of an epidemic is crucial for a good outbreak detection method, so the measure of timeliness (time since the beginning of the epidemic until it is firstly detected) is also relevant. Several ways of combining sensitivity, specificity and timeliness in a single criterion have been suggested. In this work we use two measures proposed by Kleinman and Abrams (2006). The first one, AUWROC1, is the area under a ROC curve weighted by the mean time saved. The time saved is the proportion of time an alarm is triggered before an arbitrarily given maximum delay from the outbreak. The second measure is the volume under the timeliness-ROC surface (VUTROS1). This 3-dimensions surface is constructed by creating a 3D curve in a cube $[0,1] \times [0,1] \times [0,1]$ whose axis are the sensitivity and 1-specificity, as in ROC curves, plus a third axis, which is the mean time saved. The surface is created by joining the points of this 3D curve to an upper vertex of the cube. Values near to 1 of these measures indicate good performance of the detection method in terms of sensitivity, specificity and timeliness.

We also use VUTROCS, presented in the work of Cowling et al. (2006), which is the mean of several AUROCs. Each ROC curve is constructed using a season based specificity instead of a weekly specificity. The season based specificity is defined by considering as true positives the amount of seasons where an alarm has been triggered before a certain amount of weeks from the outbreak, instead of the amount epidemic weeks correctly classified. Several ROC curves are built according to that amount of weeks, from 0 (perfect detection of the outbreak week) to the maximum delay. More details on these measures can be found in Section A.3.2 of the appendices.

Sensitivity, specificity and timeliness (and therefore the other measures discussed in this section) are computed using the estimated classification given by the detection method and the true classification of the data. In absence of a unequivocal classification of what is a true epidemic week,

some criterion is used as a reference. This reference is called gold standard and can use different data from the one used by the detection method being evaluated. The gold standard criterion usually requires higher amounts of resources in terms of time, money, computation or human effort. In Section A.3.1 of the appendices we discuss this matter in more detail, indicating the difficulties in finding a method which unequivocally locates the starting and ending of the epidemics in real data (and also some of the problems of using simulated data).

Establishing a gold standard for North Carolina

In the North Carolina example, we use the same approach for constructing a gold standard as in Cowling et al. (2006), who consider as 'true' epidemic weeks those which surpass 30% of the maximum number of weekly virus isolates for each season. Anyhow, we are aware that the gold standard defined by this procedure does not necessarily correspond to the real epidemic phase, if it can actually be defined.

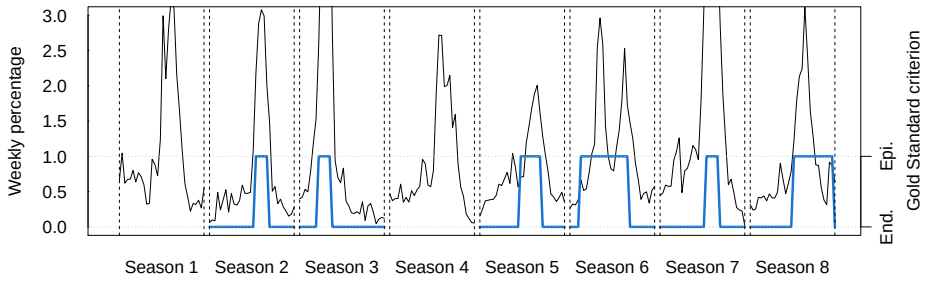

Figure 2.6: Gold standard (in blue) based on virus isolates following the criterion on Cowling et al. (2006) of surpassing 30% of the maximum number of isolates for each season (End –Endemic phase–, Epi –Epidemic phase–). In black: weekly influenza incidence percentages during seasons from 2001–2002 to 2008–2009 of the NCSN.

In Figure 2.6 we present the gold standard based on the Influenza Virus Isolates identified by the State Laboratory of Public Health and compare

it with the NCSN data. The gold standard is not defined on seasons 1 and 4, as there is no registry of the weekly virus isolation for these seasons. Comparing the gold standard criterion and the NCSN data, one may find some 'disagreements' between them. For example, one might consider that the gold standard epidemic period extends too much in time in season 8 because it comprises weeks after the ending of the descent from the epidemic peak, but not enough in seasons 5, 6 and 7, because the descent is not completely classified as epidemic. One may also consider that the gold standard does not start soon enough in seasons 3 and 7, as the rates begin to ascend towards the epidemic peak earlier. Anyhow, these are subjective considerations, and it is difficult to assure that the problem of concordance is caused by the gold standard criterion and not because of the quality of the NCSN data. We have expressed these inconsistences as the gold standard failing to adapt to the NCSN data, but one could also think that it is the NCSN data which do not completely follow the pattern of epidemic/endemic phases the gold standard gives. In any case, the correspondence between them is sensible enough and so far we lack of further tools to validate how adequate the gold standard is or how adequate the NCSN data are for the detection of influenza epidemics.

Comparing weighted ROC measures

We have selected some measures to evaluate the sensitivity, specificity and timeliness of the detection of outbreaks and have determined a gold standard, which is required by these measures. Let us use them to compare the performance of our proposal and other models in the literature on the NCSN data. In Table 2.2 we show the values of AUWROC1, VUTROS1 and VUTROCS for all the models of the framework and compare their performance with the proposal of Martinez-Beneito et al. (2008a) and other models from the literature: Serfling (1963) method, Le Strat and Carrat (1999) method without temporal trend (implemented in the R package `depmix`, see Visser, 2007) and Stroup et al. (1989) method without taking into account data from previous seasons. To test the sensitivity of the weighted ROC measures to the choice of maximum delay, we present the

measures with 1, 2 and 3 weeks after the beginning of the epidemic as maximum admissible delay. All methods are applied in an online basis (run anew every week adding the datum from that week) to emulate how they would be applied in a real surveillance system. Only seasons 5 to 8 are used to calculate these measures, as the first three seasons are used as training data for the tuning of all the models and there is no gold standard available for the fourth season.

	AUWROC1			VUTROS1			VUTROCS		
Max delay	1	2	3	1	2	3	1	2	3
M-B 2008	0.601	0.617	0.625	0.544	0.557	0.654	0.713	0.736	0.748
AR0-AR0	0.561	0.592	0.623	0.537	0.555	0.573	0.638	0.677	0.716
AR0-AR1	0.673	0.684	0.698	0.578	0.592	0.602	0.724	0.739	0.756
AR0-AR2	0.674	0.685	0.700	0.575	0.589	0.599	0.723	0.738	0.754
AR1-AR0	0.707	0.734	0.748	0.628	0.650	0.664	0.817	0.855	0.875
AR1-AR1	0.676	0.698	0.709	0.595	0.612	0.623	0.824	0.857	0.875
AR1-AR2	0.669	0.701	0.717	0.608	0.632	0.645	0.785	0.830	0.854
AR2-AR0	0.741	0.761	0.770	0.647	0.671	0.683	0.846	0.876	0.890
AR2-AR1	0.717	0.733	0.740	0.643	0.660	0.669	0.824	0.851	0.865
AR2-AR2	0.726	0.741	0.749	0.649	0.667	0.676	0.840	0.866	0.879
Stroup	0.540	0.559	0.568	0.517	0.536	0.545	0.807	0.846	0.865
HMM	0.608	0.640	0.665	0.556	0.576	0.592	0.682	0.722	0.755
Serfling	0.612	0.657	0.682	0.553	0.581	0.598	0.698	0.758	0.793

Table 2.2: Comparison of weighted ROC measures with different maximum delays on the NCSN data for the models of the proposed framework, Martinez-Beneito et al. (2008a) (M-B 2008), Stroup et al. (1989), Le Strat and Carrat (1999) (HMM) and Serfling (1963) methods, applied online. The lowest (worst) values of each column are colored in red and the highest (best) values are colored in blue.

In general, the results of the measures are quite insensitive to the election of maximum delay, as models are similarly ordered in terms of the measures for all three choices. One particular discrepancy can be observed, as is the relatively good measure of VUTROS1 with a maximum delay of 3 weeks for the model of Martinez-Beneito et al. (2008a) compared to the same measure with 1 or 2 weeks of maximum delay. In any case, given this general lack of sensitivity to the maximum delay, from now on we fix

the value of maximum delay to 1 week for any further use of the weighted ROC measures. This implies considering that an outbreak is successfully detected if an alarm is triggered the week of the beginning of the epidemic or the next week.

One can observe that, in general, the models presented in our framework outperform the other methods, with the exception of the AR0-AR0 proposal, which shows lower scores of the weighted ROC measures. In the particular case of VUTROCS, which considers seasonal sensitivity instead of weekly sensitivity, the model of Stroup et al. (1989) also shows a competitive performance. In contrast, the values for the measures that consider weekly sensitivity (AUWROC1 and VUTROS1) for this same model are the lowest among the compared models. If we focus our attention on the models on the framework, those with higher structure in the non-epidemic phase outperform the models with lower or non temporal structure in that same phase. This suggests that the temporal similarity of the endemic rates is a feature of the NCSN data which notably helps classifying these rates as non-epidemic ones. This high importance of the non-epidemic temporal structure is caused by the seasonal pattern of the endemic rates, which causes the consecutive rates and the consecutive jumps (differentiated rates) to be alike. The epidemic phase also presents similarities between consecutive jumps, during the rise of the epidemic they tend to be positive and negative during the fall. But the consecutive rates themselves are less alike because of the fast grows and decays of rates which are characteristic of this phase. Unlike the proposal of this chapter, the model of Martinez-Beneito et al. (2008a) do not give any temporal structure to the endemic phase and therefore can not take advantage of this feature to help distinguishing epidemic and endemic weeks. This is reflected in the observed measures of sensitivity, specificity and timeliness. The model with highest values for almost all measures is AR2-AR0, followed by AR2-AR2.

Relationship between DIC and weighted ROC measures

So far, we have used two tools for the comparison of methods for the detection of influenza outbreaks. DIC is a tool that evaluates the goodness-of-fit,

which is not exactly the goal of a detection method, but which is easy to obtain. On the other hand, weighted ROC measures do give a measure of sensitivity, specificity and timeliness, which directly evaluate the quality of the detection given by the methods. But, for them to correctly evaluate these classifications among epidemic and endemic weeks and the promptness of the detection, the models should be applied in an online basis. In that way, the classification being evaluated would be that obtained by a realistic application of the method in a surveillance system, using only the available information to estimate the epidemic phase each week. This can involve to run the model several dozens or even hundreds of times, which can be computationally expensive. Also, weighted ROC measures require of a reliable gold standard. If a correspondence between these two measures could be established, we could use DIC as a proxy of the more difficult to obtain weighted ROC measures.

To check the correspondence between DIC and weighted ROC measures we show the correlation between them for the nine models of the proposed framework in Table 2.3. The correlation is negative, as expected, as lower values of DIC and higher values of weighted ROC measures are associated to good performance. The correlation is far from perfect, but at least shows that DIC is a reasonable approximation to evaluate the detection power of the models. On the contrary to weighted ROC measures, DIC does not use any gold standard to be computed. As said before, the definition of a gold standard is not without problems and, as we will see in the next section, the weighted ROC measures considerably variate when the definition of the criterion to calculate the gold standard is modified. This could explain in part the discrepancies between DIC and weighted ROC measures. We observe that the best correspondence is shown between DIC and VUTROCS measure, which focuses not in the general sensitivity but specifically in the sensitivity of the detection of the starting of the epidemics, which usually is the main objective of outbreak detection methods.

	AUWROC1			VUTROS1			VUTROCS		
Max delay	1	2	3	1	2	3	1	2	3
Correlation	-0.54	-0.60	-0.58	-0.66	-0.67	-0.65	-0.77	-0.80	-0.80

Table 2.3: Correlation between DIC and weighted ROC measures calculated with a maximum acceptable delay of 1, 2 and 3 weeks after the outbreak, for the nine models of the proposed framework applied on the NCSN data.

	AUWROC1			VUTROS1			VUTROCS		
Threshold	20%	30%	40%	20%	30%	40%	20%	30%	40%
M-B 2008	0.628	0.601	0.748	0.562	0.544	0.693	0.736	0.713	0.843
AR1-AR1	0.697	0.676	0.714	0.610	0.595	0.648	0.840	0.824	0.894
AR2-AR2	0.746	0.726	0.746	0.662	0.649	0.714	0.856	0.840	0.886
Stroup	0.503	0.540	0.527	0.489	0.517	0.508	0.811	0.807	0.845
HMM	0.639	0.608	0.781	0.582	0.556	0.681	0.704	0.682	0.845
Serfling	0.711	0.612	0.622	0.595	0.553	0.596	0.798	0.698	0.738

Table 2.4: Comparison of weighted ROC measures with different definitions of gold standard on the NCSN data for the models: AR1-AR1, AR2-AR2, Martinez-Beneito et al. (2008a) (M-B 2008), Stroup et al. (1989), Le Strat and Carrat (1999) (HMM) and Serfling (1963) methods, applied online. The lowest (worst) values of each column are colored in red and the highest (best) values are colored in blue.

2.3.5 Comparing different thresholds for the gold standard

We have discussed before that a gold standard based on the data from the State Laboratory of Public Heath may not be completely well related to the data or to the real epidemic. But another problem that arises is the arbitrariness of the threshold chosen to construct this gold standard. In order to assess the sensitivity of the selection of the best model to this threshold, we compare the results for three different thresholds for building the gold standard; 20%, 30% and 40% of the maximum number of influenza virus isolates for each season (depicted in Figure 2.7). In Table 2.4 we present the scores of the weighted ROC measures for these three gold standards. For simplicity, only balanced models with temporal structure from the proposed framework of models are shown (AR1-AR1 and AR2-AR2).

Figure 2.7 shows that the gold standard may substantially vary de-

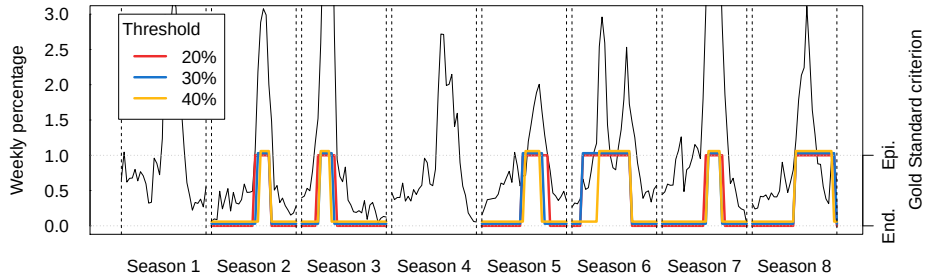

Figure 2.7: Gold standard based on virus isolates following the criterion on Cowling et al. (2006) of surpassing different percentages of the maximum number of isolates for each seasong (End –Endemic phase–, Epi –Epidemic phase–). In black: weekly influenza incidence percentages during seasons from 2001–2002 to 2008–2009 of the NCSN.

pending on the chosen threshold, as it happens in season 6. It also can be observed that the orderings of the models in Table 2.4 for the different weighted ROC scores are not always completely consistent when the criterion for defining the gold standard varies. Notice for example how the HMM method obtains better scores than Serfling (1963) method for a threshold of 40% but it gets worse scores for a threshold of 20%. It is advised, therefore, to be cautious when defining a gold standard and maybe to carry out sensitivity analyses for different values of the threshold. When deciding which method to choose among several, a good policy is to consider several complementary criteria, as are the weighted ROC measures, the DIC and others.

2.3.6 Selecting a model from the framework

In the previous sections we have discussed the performance of two tools for comparing the performance of models for the detection of influenza outbreaks and have stressed that they both have their issues. Let us discuss how to select one method among the proposed framework using DIC, weighted ROC measures and direct observation of the outcome of the mo-

dels. To start with, we will take a look at the classifications of epidemic and endemic weeks for the 9 models of the framework. We will asses whether and how the degree of autoregressive dependency for the endemic and for the epidemic phases affects this outcome and, in particular, how does it reflects on the sensitivity and specificity of the method.

In Figure 2.8 the estimated probability of being in epidemic phase by all 9 models in the framework applied on NCSN data is shown. The first graph depicts the estimated value of Z_{ts} for the models with no autoregressive structure for the non-epidemic phase. We can see there that the amount of weeks with high probability of being in epidemic phase is larger than in the other two, which represent the models with AR1 and AR2 for the endemic phase. In the third graph, which represents the models with AR2 for the endemic phase, the probabilities are the lowest. This shows that the greater the structure in the non-epidemic phase is, the more inclined the model is to signal weeks as non-epidemic. In a similar way, the red lines, which represent AR0 for the epidemic phase, are usually bellow, while the yellow ones, representing the highest autoregressive structure for the epidemic phase, are above. This indicates that the higher the temporal structure is in the epidemic phase, the more likely is the model to categorize weeks as epidemic. The discrepancy is remarkable between heterogeneous (AR0) and autoregressive (AR1 or AR2) structures, while the difference between the two possibilities for the autoregressive structures are much milder.

One problem of presenting a framework of models is that the choice of the best model for each surveillance system can be tricky. DIC is not completely related to the measures of sensitivity, specificity and timeliness, and these measures require each model to be applied in an online basis, which is computationally expensive. Besides that, as the definition of a gold standard can sometimes be doubtful, the weighted ROC measures should not be taken as the only tool for decision. To choose between the models we suggest to observe the DIC, weighted ROC measures and the assessment by an expert of the correspondence between the inferred epidemic phases and the data. But when the comparison of all the models causes troubles, we can reduce the amount of models for theoretical and observational reasons.

For an easier decision, we recommend to avoid lack of temporal structure

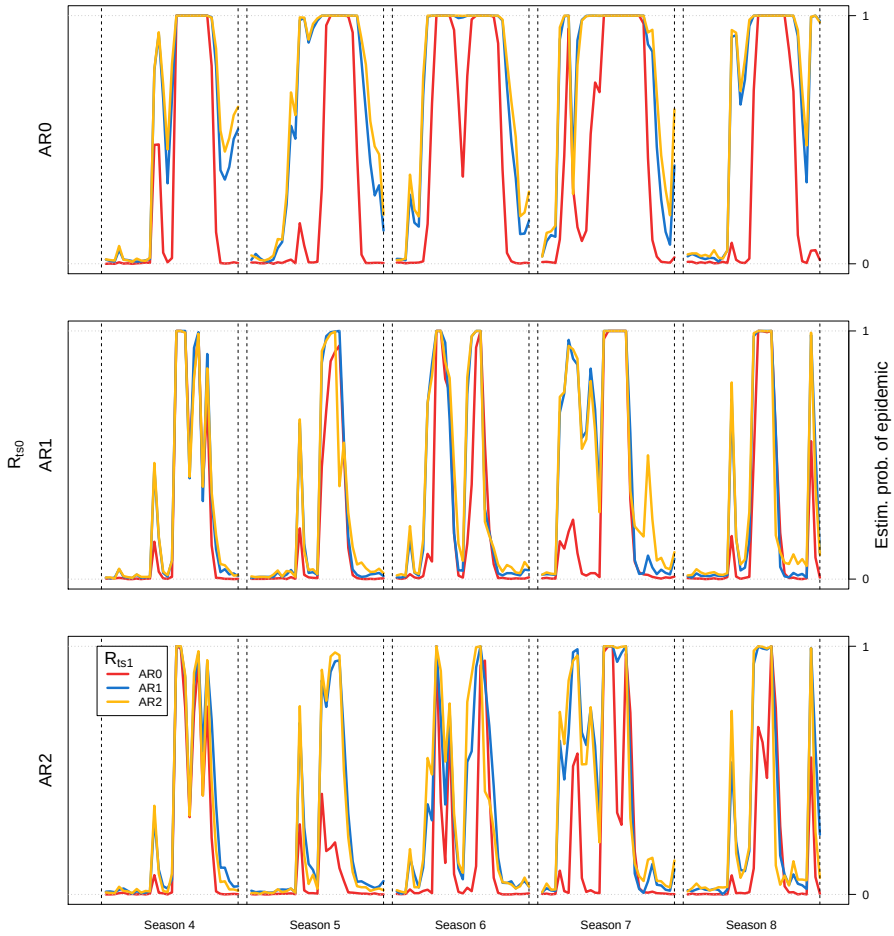

Figure 2.8: Online estimated probability of being in epidemic phase by all 9 models in the framework on NCSN data.

both in the non-epidemic and the epidemic phases. The AR0-AR0 seems like an inappropriate choice, as the nature of the data (ILI incidence rates) is qualitatively time dependent. The models where the temporal white

noise is set only in one of the phases shall also be avoided because the lack of temporal structure drastically pulls the estimations of Z_{ts} of the model to the opposite phase when the other phase is modeled with temporal structure. With the same idea, we can simplify the decision by using only balanced models (AR1-AR1 or AR2-AR2), as unbalanced models pull the classification towards the phase with higher structure. We can not assure that one of these two models will be the best choice, but it is expected to be a good one, with comparable performance to the best option of the discarded models.

In this particular case, AR2-AR2 seems like a good choice. It has good DIC score, good weighted ROC scores (and also consistent regardless the threshold of the gold standard), the temporal structures of endemic and epidemic phases are well balanced and the predictions observed in Figure 2.5 (red line) are visually consistent with what we would expect as a good detection of the outbreaks for each season.

2.4 Application on Valencian Sentinel Network data

In this section we are going to show the application of AR1-AR1 and AR2-AR2 models on the Valencian Sentinel Network data, we are going to evaluate their performance compared to other methods in the literature and to test some variations on the definition of the gold standard.

2.4.1 The Valencian Sentinel Network data

Being part of the Valencian Network of Surveillance in Public Health (Red Valenciana de Vigilancia en Salud Pública), the Valencian Sentinel Network (Red Centinela Sanitaria, VSN) tracks the incidence of several diseases, including influenza like illnesses. ILI and other diseases, are weekly reported by volunteer practitioners of all the Valencian Region, one of the 17 administrative regions of Spain. This data source reports ILI counts, the population covered by the volunteer practitioners (not only those who at-

tend the doctor's office each week but all the patients which are assigned to the doctor) and the rates associated to these two quantities over 100 000 inhabitants. The string of data we use in this application is displayed in Figure 2.1. The data consists of the reports of ILI from season 1996–1997 to season 2008–2009 during 30 weeks each season, from the 42nd week of one year to the 19th week of the following. As previously mentioned, this data is the same as that one used in Martinez-Beneito et al. (2008a) but adding two new seasons and correcting an error on the data of season 2005–2006, where the rates of this season were divided by 10. This was a minor error, as this season presents no appreciable epidemic phase. The VSN also seeks for virus isolation for a small fraction of the patients diagnosed with ILI, which we will be using to build a gold standard for the 'true' epidemic phase.

2.4.2 Evaluating model performance on VSN data

In this section we are going to select a model for the detection of influenza epidemics on the VSN data following the guidelines presented in Section 2.3.6. As stated before and for the sake of simplicity, we restrict our selection among the balanced models with temporal structure; AR1-AR1 and AR2-AR2.

Evaluating the goodness-of-fit with DIC

By running each model only once on the whole data set, we can obtain the DIC score, which offers a measure of the goodness-of-fit of the model to the data penalized by the effective number of parameters of each model. To double check the inadequacy of the model without temporal structure AR0-AR0, we also include it in the comparison of DIC. In Table 2.5 we observe that DIC for AR0-AR0 is much worse than that for the two other proposals, with a distance greater than 100 points. The best proposal in terms of DIC is AR2-AR2 with a score 10 points lower than that of AR1-AR1. In Section 2.3.6 we remarked the relatively high discrepancy in the estimation of the latent variable of the models with heterogeneous structures (AR0), both

for the endemic and epidemic phase, with respect to these models with autoregressive structure. DIC corroborates again that this estimation is likely to be worse, as the model fits worse the data. Therefore the AR0-AR0 model is not further considered for the model selection.

	AR0-AR0	AR1-AR1	AR2-AR2
DIC	2487.9	2365.8	2355.5

Table 2.5: DIC for the balanced models of the proposed framework for the VSN data.

Determining a gold standard

To use weighted ROC measures, which give an idea of the sensitivity, specificity and timeliness achieved by the models, we need a gold standard. When considering to build a gold standard from the laboratory isolation data, the approach of Cowling et al. (2006) seems not appropriate due to the low amount of isolates during the year. In the US laboratory data used by Cowling et al. (2006), 1300 samples are tested each month in median, arriving to achieve rates of influenza isolation over 0.30. VSN analyzes a much lower number of samples and the weekly number of virus isolates never surpasses 11 in our available data. Because of that, we use the gold standard as defined on Martinez-Beneito et al. (2008a) instead, considering as being in epidemic phase those weeks between the first and last influenza virus isolates of the season.

There is one correction that can be done to this gold standard, as in the first week of season 12 there is an isolate and then there is no other isolate until 3 months later. It might be sensible, therefore, to ignore this separated virus isolate and declare the epidemic to start at the 9th week of season 12, instead of at the first week. In Figure 2.9 the corrected gold standard is depicted over the rates. The non-corrected gold standard is the same but for season 12, where the epidemic would start at the beginning of the season. In the following part of the section we are going compare the effect of using the original or the corrected gold standard on the selection

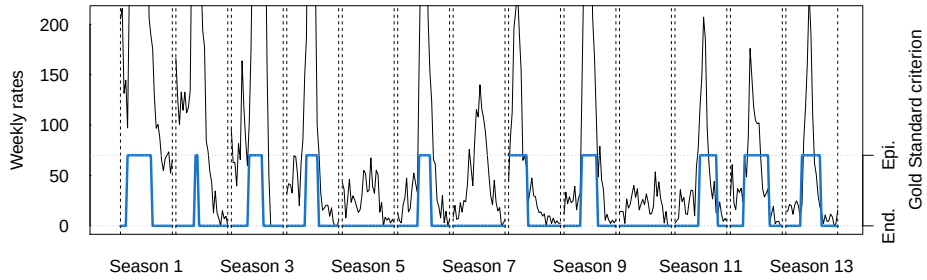

Figure 2.9: Corrected gold standard (in blue, End –Endemic phase–, Epi – Epidemic phase–). In black: weekly ILI incidence per 100 000 inhabitants from season 1996–1997 to season 2008–2009 of the VSN.

of models.

Weighted ROC measures

Weighted ROC measures are shown in Table 2.6 to evaluate the performance of AR1-AR1 and AR2-AR2 models and compare them to other models in the literature. As previously done in the NCSN analysis, the first 3 seasons are not taken into consideration for the evaluation of the weighted ROC measures, because they are considered as training data. For each measure, two columns are shown; the first one is the measure using the gold standard with the strict definition in Martinez-Beneito et al. (2008a), while the second column is the same measure with the correction on the 12th season to the gold standard.

In general, AR2-AR2 model shows the best scores in all three measures and the two gold standard definitions, but there are some discrepancies between the order of the scores for the two gold standards. One discrepancy is the notoriously better score of the model of Le Strat and Carrat (1999) (HMM) on AUWROC1 and VUTROCS. The other difference is the worse score of the method by Martinez-Beneito et al. (2008a) in VUTROCS with the corrected gold standard. Once again, this shows the sensitivity of the

	AUWROC1		VUTROS1		VUTROCS	
Gold st.	Orig.	Correc.	Orig.	Correc.	Orig.	Correc.
M-B 2008	0.871	0.876	0.760	0.754	0.921	0.911
AR1-AR1	0.806	0.847	0.751	0.786	0.916	0.935
AR2-AR2	0.871	0.905	0.794	0.828	0.953	0.968
Stroup	0.527	0.541	0.489	0.506	0.891	0.949
HMM	0.813	0.905	0.707	0.752	0.867	0.936
Serfling	0.729	0.752	0.645	0.657	0.855	0.866

Table 2.6: Comparison of weighted ROC measures with original and corrected gold standard on the VSN data for the model of Martinez-Beneito et al. (2008a) (M-B 2008), the proposed models AR1-AR1, AR2-AR2, and the models of Stroup et al. (1989), Le Strat and Carrat (1999) (HMM) and Serfling (1963) methods, applied online. The lowest (worst) values of each column are colored in red and the highest (best) values are colored in blue.

weighted ROC scores to the definition of the gold standard.

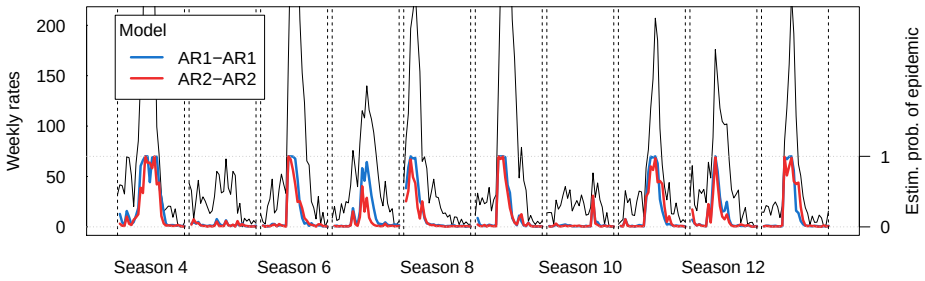

Figure 2.10: Estimated probability of being in epidemic phase by AR1-AR1 and AR2-AR2 models on VSN data. In black: weekly ILI incidence per 100 000 inhabitants from season 1999–2000 to season 2008–2009.

DIC and weighted ROC measures indicate a better goodness-of-fit and performance on AR2-AR2 model than on AR1-AR1, therefore AR2-AR2 would be a sensible choice to use in surveillance. The online estimates for

the probability of being in the epidemic state are shown in Figure 2.10. One can observe that, in some seasons, the estimates from the AR1-AR1 model linger less in intermediate values and tend to give scores closer to 1 or 0 than the AR2-AR2 model. Some public health authorities might prefer therefore the AR1-AR1 model, as it still has good DIC and weighted ROC scores and shows slightly more determination when indicating if a week is in the endemic or epidemic phase, so the decisions would be less threshold dependent.

2.4.3 Parameters of the model AR2-AR2

In Table 2.7 we present the parameters of the retrospective application of AR2-AR2 model on the complete VSN data (which are also the parameters for the estimation of the last week of the online application). Of particular interest are the values of ρ_{10}, ρ_{20}, ρ_{11} and ρ_{21}, all different from 0, which seem to support the selection of second-order autoregressive processes in both periods. It is also interesting to note that the distance between the posterior distribution of the means and variances for the endemic and epidemic phases indicates clearly different dynamics between phases.

2.5 Application on Google Flu Trends data sets

As we have discussed before, this framework of models is designed so that it can be applied to data obtained from different kinds of data sources. So far we have shown applications on data chains obtained from sentinel networks. In this section we are going to show the performance of our proposal on data obtained from a less traditional data source, Google Flu Trends, for three different countries.

2.5.1 The Google Flu Trends data

For these applications we use data from Google Flu Trends (GFT) for the countries of Spain, Japan and Netherlands from the 40th week of 2003 to the 18th week of 2010, dividing the seasons every 52 weeks. This results

Parameter	Posterior mean	95% Credible interval
p_{00}	0.96	[0.72 , 0.92]
p_{11}	0.85	[0.93 , 0.98]
$\theta_{[1]}$	3.23	[0.82 , 6.11]
$\theta_{[2]}$	7.25	[4.43 , 11.33]
$\theta_{[3]}$	28.63	[8.98 , 50.45]
$\theta_{[4]}$	77.80	[49.48 , 128.09]
μ_0	20.27	[15.43 , 25.69]
μ_1	207.72	[133.31 , 288.16]
ρ_{10}	1.23	[1.06 , 1.36]
ρ_{20}	−0.38	[−0.49 , −0.25]
ρ_{11}	1.50	[1.26 , 1.70]
ρ_{21}	−0.67	[−0.86 , −0.44]

Table 2.7: Estimated mean and 95% credible interval for the parameters of the AR2-AR2 model applied on the VSN data in a retrospective basis.

in 6 seasons of 52 weeks and one incomplete season of 32 weeks. As commented in Section 1.3.1 these data chains are estimates based on internet queries of influenza related topics that try to reproduce the incidence rates (expressed over 100 000 inhabitants) provided by official institutions of each country. The last season includes the wave of swine influenza, that might not be well estimated by GFT because data are not estimated with the last modifications of the algorithm. Also, media putted much attention on this epidemic, which most certainly affected the behavior of search engine users about the topic and, therefore, on the estimations by GFT. The estimated influenza incidence rates for the three countries are shown in Figure 2.11.

2.5.2 Online application of the AR1-AR1 and AR2-AR2 models

In this case no gold standard is available for any of the three data sets, so no measures of specificity, sensitivity and timeliness can be provided that could help in the model selection. To simplify the model selection and

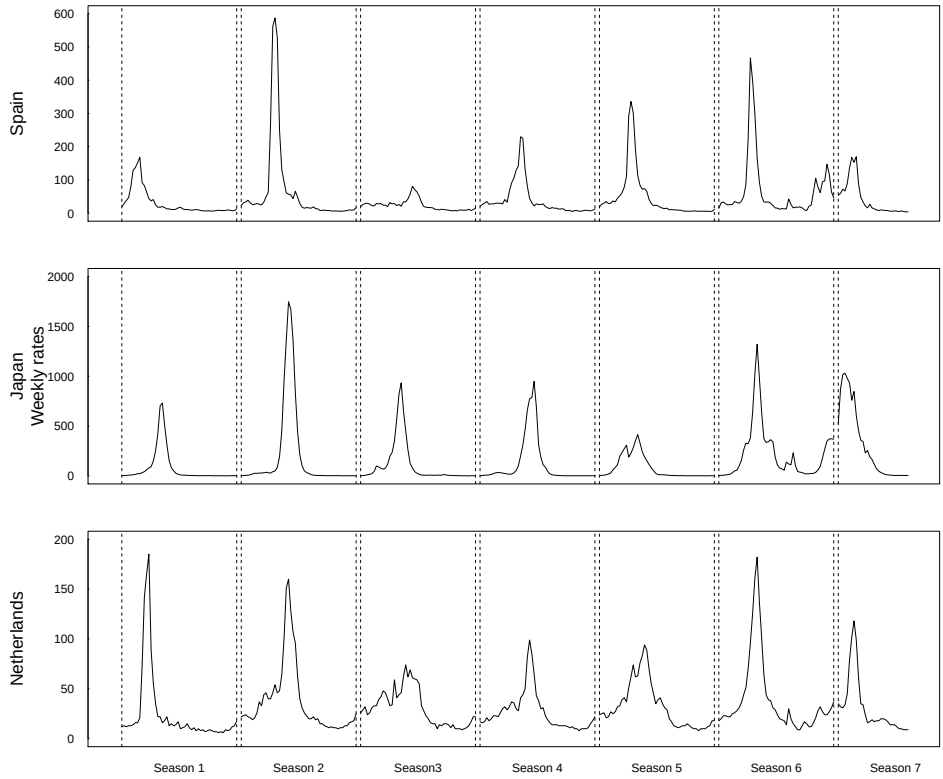

Figure 2.11: Weekly estimated influenza incidence per 100 000 inhabitants during seasons from 2003–2004 to 2009–2010 by GFT for Spain, Japan and Netherlands.

following the guidelines presented in Section 2.3.6 we do not consider the whole framework of models but only AR1-AR1 and AR2-AR2 models. In Table 2.8 we observe that, for all data sets, the model with higher temporal autoregressive structure offers better DIC scores than the model with autoregressive structure of order one, as it also was the case with both VSN and NCSN data (see Tables 2.1 and 2.5). In any case, we should remember that DIC is not a direct evaluation of the quality of detection of the models and we may consider other aspects for taking the decision. For example,

when taking a look at the estimates of the epidemic phase presented in Figure 2.12 we observe that AR1-AR1 tends to give higher estimates for Z_{ts}. In the case we were to seek for a model that favors sensitivity over specificity we might choose to select the AR1-AR1 model even if it offers a worse DIC score.

	AR1-AR1	AR2-AR2
Spain	2163.5	2107.3
Japan	2345.7	2306.9
Netherlands	2108.9	2030.8

Table 2.8: DIC for the AR1-AR1 and AR2-AR2 models of the proposed framework on the GFT data for Spain, Japan and Netherlands.

One might notice that the Japan data set presents estimated incidence of zero or near to zero during the non-epidemic periods, with extremely low variability. This behavior is captured by the non-epidemic structure of the models and whenever the estimated incidence values rise, even slightly, the estimated probability of epidemic rises so that the epidemic structure can fit the higher variability and higher values of the response variable. For this reason, our proposal could have problems to deal with data with long strings of zeroes in the non-epidemic phase. Several variations might be considered in further investigations to deal with this kind of data, as using more than two hidden states or eliminating connected sections of data that are all zeroes. On the other hand, the data sets of Spain and Netherlands do present higher than zero incidence rates and variances during the non-epidemic periods. In this situation the higher estimates of probability of epidemic tend to be located only on the steep rises and decays around the peaks of the epidemics, as it is expected from a sensible detection method.

To summarize, in this chapter we have presented a framework of Bayesian hierarchical Poisson models with a hidden Markov structure for the detection of influenza outbreaks based on the proposal of Martinez-Beneito et al. (2008a). We have described the framework of models and applied it on two surveillance data sets and three data sets from internet

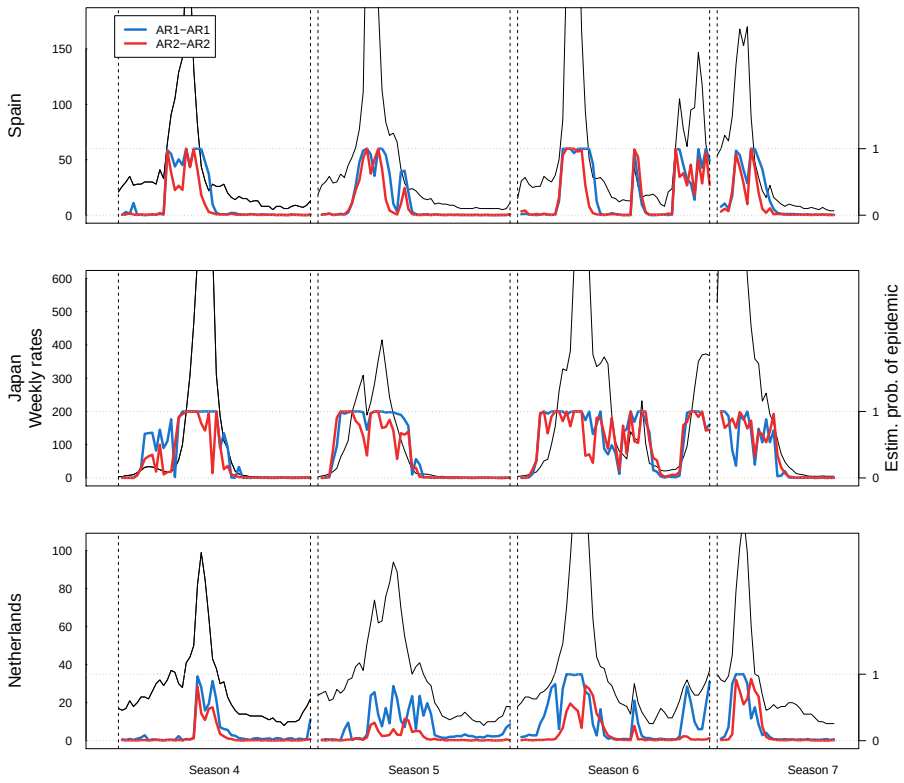

Figure 2.12: Estimated probability of being in epidemic phase by AR1-AR1 and AR2-AR2 models on GFT data for Spain, Japan and Netherlands. In black: weekly estimated influenza incidence per 100 000 inhabitants during seasons from 2006–2007 to 2009–2010.

query based sources. These applications have allowed us to discuss the issue of model selection, in particular the use of gold standards, weighted ROC curves, DIC and observation of the outcomes for selecting the best method of detection. The observed results suggest to restrict the selection of models to those with balanced autoregressive structures for both endemic and epidemic phases whenever a full comparison is impracticable. We have

also shown that several models within the proposed framework outperform other proposals in the literature in terms of weighted ROC curves based measures. The following chapter deals with a more complex problem: the detection of influenza outbreaks using spatio-temporal data. Spatial and spatio-temporal structures allows neighboring regions to share information about influenza incidence and epidemic state, therefore producing models that give a more prompt and consistent detection of influenza outbreaks than those which do not share information in space. For this reason, in the next chapter we propose a spatio-temporal extension of the proposal of Martinez-Beneito et al. (2008a) for the detection of influenza outbreaks in time and space.

Chapter 3

Spatio-temporal detection of influenza outbreaks

3.1 Introduction

Our review of temporal methods for the detection of influenza epidemics presented in Chapter 1 has highlighted the advantages of several statistical structures for this task: time series structures, hidden Markov chains or Bayesian hierarchical models, are some of the statistical tools that can help to better distinguish between epidemic and endemic phases. In Chapter 2 we have used them to construct a framework of models for the temporal detection of influenza outbreaks, which has proven to be capable of timely detect the outbreaks of influenza in different temporal data sets.

In the present chapter we present a spatio-temporal Bayesian hierarchical model on the differentiated rates with a hidden Markov structure for the detection of influenza outbreaks. This proposal is also based on the one of Martinez-Beneito et al. (2008a) and also uses the statistical structures previously mentioned (time series, hidden Markov chains and Bayesian hierarchical models). Unlike in Chapter 2, the spatio-temporal proposal has not an additional Poisson first layer on the observations. Instead, we directly model the differentiated rates as Martinez-Beneito et al. (2008a) originally

do. Some of the reasons for this choice are the improvement in simplicity of the model (and the subsequent reduction in computational time) and the fact that for some data only rates are available, but not counts. One could estimate the number of cases from the rates if population data for the regions were available, but those estimated cases would not be actual observations, so modeling them would introduce a spurious source of variability. We also believe that observing only the differentiated rates and not the actual dimension of the rates can diminish the sensibility in high rates, but can also improve the timeliness, as the outbreaks of influenza are usually located at weeks with low raw rates but relatively large increases of the rates.

In Chapter 2 we have presented a framework of models among which an expert should choose the best fitted one for their surveillance system. Many health policy makers find that capability of choosing among several models as a problem instead of as an advantage: sometimes the tools to choose between models do not offer a clear winner and there is not always a trained expert to do the choice. For that reason, we propose only one model for the spatio-temporal proposal, having in mind that this method will have to be able to fit diverse data sets with their peculiar characteristics.

There are several reasons for using spatial information to improve the performance of outbreak detection methods. One reason is that dealing with aggregated information from a large territory with temporal data can hinder the detection of the onset of outbreaks, as the beginning of influenza outbreaks often happens in small regions. Even though the increase of cases might be noticeable in these small regions, it may become unnoticed if merged with the endemic cases of the larger territory. Another advantage is to consider the contagious nature of the disease so, if an outbreak starts a certain week in a certain small region, it is more likely for neighbor regions to get into epidemic stage in the same or nearby weeks. Last but not least, a method that offers spatially demarcated alarms can be more useful when devising health policies, allowing to focus the use of resources where and when they are most needed.

The remainder of this chapter is organized as follows. First we will review some general spatial models which set the foundations for the build-

ing of several spatio-temporal models. Then we will do a review of spatio-temporal methods for the detection of outbreaks, paying special attention to those methods used for influenza outbreaks detection. This will help us putting the new proposal in context and we will also review some statistical tools that we will use but which were not described in previous chapters. We will go over spatio-temporal extensions of: ARIMA, cumulative control charts, scan statistics, Bayesian methods and hidden Markov chain models. After this contextualization, the newly proposed spatio-temporal model for the detection of influenza outbreaks will be described. The application of this model on Google Flu Trends data from USA will exemplify the performance of this new proposal, and the comparison with some variations of the model and the proposal of Martinez-Beneito et al. (2008a) will show the pertinence and relevance of this new spatio-temporal proposal.

3.2 Review of spatial models

In Section 1.4.3 we reviewed ARIMA methods, which allow to model the existing correlation between consecutive observations of influenza rates and therefore are frequently used tools in temporal statistical models for the detection of influenza outbreaks. ARIMA, as well as most of the methods presented in Chapter 1, are applied on data referenced to discrete (and usually equally spaced) time points, which can be represented as a linear graph, where each time point is connected to the previous and next ones. When having to deal with spatial data, a natural evolution is to use a discrete spatial support for the data, where the information is displayed as a lattice process. In a lattice process, the set of counts or rates of the disease $\{Y_i\}$ must be observed at a set of fixed locations $i \in \{1, \cdots, I\}$ with a certain neighboring structure. This neighborhood can be defined by an $I \times I$ matrix $\boldsymbol{W} = (w_{ij})$ with $w_{ij} = 1$ if locations i and j are neighbors and zero otherwise. We will consider that a region i is not a neighbor of itself. Usually, the fixed locations which constitute a lattice are administrative regions and the neighboring relation is defined by sharing a common border. Nevertheless, other ways of setting the lattice can be

defined, and even non binary neighboring weights could be set.

Spatial structures can share information among neighboring regions in a similar way as ARIMA structures share information among contiguous time points. This allows, for example, to better estimate effects in small regions with low population, where the variance of the estimation can be too large or affected by zero counts which offer spuriously low or high estimates of the incidence or the risk. In this section we review the intrinsic conditional autoregressive model, the proper conditional autoregressive model and the spatial moving average risk smoothing model, which are the spatial equivalents of the ARIMA structures. For simplicity we only review the structures of order 1 except for the extension of the moving average. We also review two other related spatial models, the Besag, York and Mollié model and the Leroux model. Let us describe them:

Intrinsic conditional autoregressive model. Also known as ICAR model, it is the spatial equivalent to a random walk of order 1. Was presented in Besag et al. (1991) and since then has been widely used in dissease mapping and other issues of disease surveillance. The ICAR model is conditionally expressed as follows:

$$Y_i|\boldsymbol{Y_{-i}} \sim \mathrm{N}\left(\frac{\sum_{j\sim i} Y_j}{n_i}, \frac{\sigma^2}{n_i}\right), \tag{3.1}$$

with $j \sim i$ meaning locations i and j are neighbors, $\boldsymbol{Y_{-i}} = \{Y_j : j \neq i\}$ and n_i being the number of neighbors of location i. In this way, the conditional distribution for each location follows a Gaussian distribution whose expected value is the mean value of the neighboring locations and its variance is inversely proportional to the number of neighbors. The set of all the conditional distributions is intrinsic, in the sense that one extra constraint is needed for its proper definition. Usually, a constraint on the sum of the variables $\sum_{i=1}^{I} Y_i$ is set, generally constraining it to zero. That is the case, for example, for the software WinBUGS, which we use in the application of the proposals presented in this thesis.

In order to cope with data that have mean different to zero, a general mean μ is estimated in addition to the ICAR term. The model can be also expressed as a multivariate normal distribution in the following way:

$$\boldsymbol{Y} \sim \mathrm{N}\left(\boldsymbol{0}, \sigma^2 \left(\boldsymbol{D} - \boldsymbol{W}\right)^{+}\right) , \tag{3.2}$$

where $\boldsymbol{D}$ is a diagonal matrix with n_i as the diagonal terms and $(\cdot)^+$ denotes the Moore-Penrose pseudoinverse of a matrix (see, for example, Gentle (2007)). One has to take into account that the structure matrix $(\boldsymbol{D} - \boldsymbol{W})$ is singular (having each row and column n_i in the diagonal and n_i negative ones in the rest of the positions, the sum of the rows is always zero). The range of the structure matrix is at most $I - 1$, which is why the restriction about the sum of all of the observations (or any other restriction) is necessary to make of this a proper distribution. The ICAR distribution models strong spatial dependencies, where the mean of the process in one location is highly determined by the value of the observations in the neighboring regions, as described by Botella-Rocamora et al. (2013). These authors also show that this structure implies a negative correlation of the spatial effect between locations that are far away in the lattice.

Besag, York and Mollié model. When used for disease mapping, the ICAR model is frequently combined with an independent random effect, what is known as the Besag, York, and Mollié (1991) (BYM) model. This random effect is able to capture particular behaviors of the regions that are not determined by the neighboring structure, relaxing the strong spatial dependency defined by the ICAR model. Both spatial and independent random effects are usually combined, for defining more flexible variables, as:

$$\begin{aligned} Y_i &= \mu + u_i + v_i \,, \\ u_i | u_{-i} &\sim \mathrm{N}\left(\frac{\sum_{j\sim i} Y_j}{n_i}, \frac{\sigma_u^2}{n_i}\right) , \\ v_i &\sim \mathrm{N}(0, \sigma_v) \,. \end{aligned} \tag{3.3}$$

In vectorial notation, it can be expressed as a multivariate normal distribution (as a result of the sum of two multivariate normal distributions):

$$\boldsymbol{Y} \sim \mathrm{N}\left(\mu\mathbf{1}, \sigma_u^2 \left(\boldsymbol{D} - \boldsymbol{W}\right)^+ + \sigma_v^2 \boldsymbol{I}\right) . \tag{3.4}$$

This combination allows to distribute the variability of the model in two spatially structured and unstructured terms, so that both the spatial effect of the neighbors and the particular effect of each location can be taken into account and distinguished.

Leroux model. BYM model distributes the spatially structured and unstructured variability of the data by using two Gaussian additive terms. Another way of distributing this variability is through the definition of the structure matrix (proportional to the inverse of the variance matrix). In that way Leroux et al. (2000) introduced a weight parameter ϕ (bound to be between 0 and 1) to estimate the strength of the spatially structured and unstructured variability in a unique multivariate Gaussian distribution with the following definition:

$$\boldsymbol{Y} \sim \mathrm{N}\left(0, \sigma^2 \left(\phi \left(\boldsymbol{D} - \boldsymbol{W}\right) + (1 - \phi)\, \boldsymbol{I}\right)^{-1}\right) , \tag{3.5}$$

so that the structure matrix gives more weight to the spatially structured or the unstructured variability depending on the estimated value of ϕ. This model can also be expressed with conditional marginal distributions as follows:

$$Y_i | Y_{-i} \sim \mathrm{N}\left(\frac{\phi}{1 - \phi + \phi n_i} \sum_{j \sim i} Y_j, \frac{\sigma_\psi^2}{1 - \phi + \phi n_i}\right) . \tag{3.6}$$

The use of ICAR, BYM and Leroux models for disease mapping is discussed in Riebler et al. (2016).

Proper conditional autoregressive model. Also called proper CAR model, is the spatial equivalent of the temporal autoregressive structure of order 1. The definition by conditional distributions is quite

similar to that one of the ICAR model, but adding a correlation factor ϕ to the mean:

$$Y_i|Y_{-i} \sim \mathrm{N}\left(\phi\frac{\sum_{j\sim i} Y_j}{n_i}, \frac{\sigma^2}{n_i}\right) . \tag{3.7}$$

The estimated value of this correlation factor indicates the strength of the spatial correlation of the data; $\phi > 0$ indicates that neighbor regions have similar values, while $\phi < 0$ arises when neighbor locations have opposite values. When $\phi = 0$, no spatial structure is observed in the data, but for the fact that the variance of each observation depends on its number of neighbors. The joint distribution of the proper CAR distribution is as follows:

$$\boldsymbol{Y} \sim \mathrm{N}\left(\mu\mathbf{1}, \sigma^2\left(\boldsymbol{D} - \phi\boldsymbol{W}\right)^{-1}\right) . \tag{3.8}$$

As shown in Gelfand and Vounatsou (2003), this distribution is proper as long as the parameter ϕ is strictly between the inverse values of the smallest and largest eigenvalue of $\boldsymbol{D}^{-\frac{1}{2}}\boldsymbol{W}\boldsymbol{D}^{-\frac{1}{2}}$, with the smallest eigenvalue being negative and the largest one being 1. In practice, in many fields it is common to assume that, in case of existing spatial correlation, it is positive, so ϕ is usually parameterized between 0 and 1. In the limit values of the parameter ϕ proper CAR reduces to a non-spatial Gaussian random effect for $\phi = 0$ and the ICAR model for $\phi = 1$. It should be taken into account that the restriction on the mean is not necessary in this case, as $(\boldsymbol{D} - \phi\boldsymbol{W})$ is proper in the range assumed for ϕ. Therefore, this distribution in WinBUGS includes an specific parameter μ modeling its mean.

Spatial moving average risk smoothing model. Also called SMARS, this model was presented by Botella-Rocamora et al. (2013) and is the spatial extension of the temporal moving average process of unfixed order. The conditional definition of the model is shown in the article as a hierarchical model with a Poisson distribution for the likelihood

and the moving average structure for the logarithm of the relative risk:

$$\begin{aligned} O_i &\sim \mathrm{Po}(E_i \exp(\nu_i)) \,, \\ \nu_i &= \mu + \lambda_i^{-1} \left(\gamma_0 \varepsilon_i + \gamma_1 \left(\sum_{j \sim_1 i} \varepsilon_j \right) + \cdots + \gamma_m \left(\sum_{j \sim_m i} \varepsilon_j \right) \right) \,, \\ \varepsilon_i &\sim \mathrm{N}(0, \sigma^2) \,, \end{aligned} \tag{3.9}$$

where m is the order of the SMARS (usually higher than 1) to be estimated by the model and λ_i are weighting factors set so that the components of the vector $\boldsymbol{\nu}$ have all the same variance σ^2 regardless of their location in the region of study. The notation $j \sim_v i$ denotes j being a neighbor of i in v steps or less, that is, there exists a path in the graph of neighborhoods connecting j and i through v edges or less. This model has the advantage over the ICAR model of not causing a negative correlation of the spatial effect between locations that are far away in the lattice.

3.3 Review of spatio-temporal methods for the detection of outbreaks

In this section we present a brief review of spatio-temporal methods for the detection of outbreaks with special attention to methods for the detection of influenza outbreaks. Visiting all these models of the literature provides the necessary background for the development of our spatio-temporal proposal.

3.3.1 Spatio-temporal extensions of ARIMA and spatial models

There are several spatio-temporal extensions of the temporal ARIMA models and the spatial models presented in the previous section. One example is to consider a spatio-temporal autoregression based on the mean of the

values of the response variable for the neighbors but on the previous unit of time instead of on the present time. In this fashion, Held and Paul (2012) present a model where the number of observed cases Y_{it} is set to follow a negative binomial distribution with mean μ_{it}, which is modeled as the sum of three components:

$$\mu_{it} = \rho_{it} Y_{it-1} + \phi_{it} \frac{\sum_{j \sim i} Y_{it-1}}{n_i} + \nu_{it}\,, \tag{3.10}$$

the first one being a temporal autoregressive model, the second one similar to a proper CAR but with data from week $t-1$ and the third one representing the endemic proportion of cases. To add a twist, the parameters ρ_{it}, ϕ_{it} and ν_{it} are not necessarily constants but the authors propose the option of model their logarithms through harmonic regressions in the fashion of Serfling (1963). This model is implemented in the R package `surveillance` and discussed in Meyer et al. (2014). In the same package, two other versions of the same model are presented for individual events in continuous space-time (point process data) and for individual Susceptible-Infectious-Removed event history of fixed population data. Several other spatio-temporal extensions of spatial models can be done but, due to their complexity (among other possible reasons), they are usually proposed under the Bayesian paradigm. We will discuss them later in Sections 3.3.4 and 3.3.5.

3.3.2 Spatio-temporal cumulative control charts

Several temporal models (not necessarily using ARIMA structures) in disease surveillance have spatio-temporal extensions. That is the case of the cumulative control charts, among others. Rogerson and Yamada (2004) discuss some approaches to spatio-temporal surveillance using different multivariate CUSUM extensions. As they list, the CUSUM extensions include:

1. separate monitoring of each regional disease count (Woodall and Ncube, 1985),

2. monitoring outliers among vectors of the distance between observed and expected standardized counts (e.g. a multivariate Shewhart chart),
3. monitoring a univariate measure of the distance between regional vectors of observed and expected standardized counts, and
4. monitoring the multivariate vector of cumulated differences between observed and expected standardized counts (Pignatiello and Runger, 1990; Crosier, 1988).

Another cumulative control chart type, as is the exponentially weighted moving average chart (EWMA), is also used in spatio-temporal modeling of the spread of diseases, as seen in Zhou and Lawson (2008), who perform a spatio-temporal prediction of excess of Relative Risk for a simulated generic disease, combining a Bayesian BYM model with a multivariate EWMA.

As discussed in Section 1.4.2, cumulative control charts rise the alarm of outbreak when they find mild but persistent shifts of the rate mean. But this shift is a common endemic behavior of ILI on the arrival of the cold months of the year. This behavior is also observed in the spatio-temporal extensions of cumulative control charts. As we will see later, our spatio-temporal proposal avoids this misclassification by not paying attention to the mean of the process of rates, but using the differentiated rates to distinguish the epidemic and non-epidemic phases. In that way, a sharp increase or several contiguous mild increases in the rates can trigger an alarm, while one mild increase in the mean of the process does not.

3.3.3 Spatial and spatio-temporal scan statistics

In Section 1.4.5 the use of scan statistic for temporal data has been reviewed, but its main use is found when applied on spatial and spatio-temporal data. Let us remember that the basic idea to detect clusters using this type of algorithm implies proposing several windows that cover a connected subset of the spatio-temporal space. The counts in these windows are tested against the null hypothesis of random observations, checking if there is an unusual amount of observations on their inside compared to the

outside given the window size. The detected clusters are those windows whose observations are less likely given the null hypothesis.

This tool started detecting clusters of one or two dimensions on point processes with only one possible cluster as an output and a fixed windows size with segment or square shape (Naus, 1965a,b), but it has evolved and become more complex since then. Some of the examples of this sophistication are: the use of circular clusters with different sizes on lattice spatial data with primary and secondary clusters detection by Kulldorff (1997), eliptical windows by Kulldorff et al. (2006), arbitrarily shaped windows by Assunção et al. (2006) or Takahashi et al. (2008) and spatio-temporal cylinders by Kulldorff et al. (2005).

The use of scan statistic for diseases surveillance is common. Some examples of it are the aforementioned works of Kulldorff (1997) for spatial data and Kulldorff et al. (2005) for spatio-temporal data. In Section 1.4.2 we have pointed that Ismail et al. (2003) applied statistic for temporal detection of outbreaks in an online basis forcing the potential cluster windows to include observations from the last week or day. In the same way, Kulldorff (2001) does it in a spatio-temporal basis. Several of these algorithms are implemented in the computer program SaTScan™.

Scan statistics have also been addressed under the Bayesian paradigm, as seen in the work of Neill et al. (2006) with a Bayesian spatial scan statistic, and in Neill and Cooper (2010), where a multivariate Bayesian scan statistic is used to compute the posterior probability of several types of events (for example, avian flu, bio-terrorist atacks, etc.) in each space-time region under study. The data used to detect these different events are a multivariate vector; for example, different chief complaints (respiratory, fever, etc.) to the Emergency Department or different over-the-counter medication sales for several items. Detected clusters are therefore characterized as one type of event or another depending on which complaints or which sold items are appreciated to cluster together.

A problem one finds when trying to do spatio-temporal detection of influenza outbreaks is the fact that influenza epidemics usually cover huge extensions of terrain (whole countries or even continents). The windows of scan statistics though, can cover at most only half of the region being

analyzed. Other statistical tools which can cope with this wide spread are therefore advisable for the detection of this disease.

3.3.4 Spatio-temporal Bayesian methods

Bayesian models are quite prolific in the spatio-temporal detection of outbreaks due to their capacity to make inference on models with complexly correlated data (as spatio-temporal data tend to be) in a relatively simple way. In Sections 3.3.2 and 3.3.3 we have already seen that control charts and scan statistic have been applied under the Bayesian paradigm. In this section we review several other Bayesian models for the surveillance of diseases, taking special attention to the models devoted to detect outbreaks of influenza and ILI.

Many spatio-temporal extensions of the spatial models discussed in Section 3.2 are usually embedded in Bayesian hierarchical models, due to the difficulty of implementing them under the frequentist paradigm, among other possible reasons. In this fashion, Mugglin et al. (2002) use a Poisson vectorial auto-regressive model with a spatio-temporal term (ST_{it}). The value of the spatio-temporal term for a certain location i and time t is a linear combination of the value of the spatio-temporal term in the same region at the previous time and the terms for the first and second order neighbors also in time $t-1$. Neighbors of order n of a location i are defined as those locations which are connected to region i by a path of n edges of the neighborhood matrix and are not neighbors of a lower order. The authors also add to the spatio-temporal term a proper CAR structure for each time unit ($CAR_{it}(\mu_{Z_t})$) with three possible mean values μ_{Z_t} corresponding to three stages. The three stages, ($Z_t = 1, 2$ or 3), correspond to non-epidemic, growing and decaying phases and their temporal locations and extents are not estimated by the model, but previously set by the researchers. Because of that, this model is not suitable for outbreak

detection. The proposed hierarchical model can be written as follows:

$$\begin{aligned} Y_{it} &\sim \text{Po}(E_i R_{it})\,, \\ \log(R_{it}) &= x_{it}^{'}\alpha + ST_{it}\,, \\ ST_{it} &= \rho_0 ST_{it-1} + \rho_1 ST_{jt-1} + \rho_2 ST_{kt-1} + CAR_{it}(\mu_{Z_t})\,, \end{aligned} \tag{3.11}$$

where j stands for the first degree neighbors of i, k stands for the second degree neighbors of i, E_i is the expected counts for region i and $x_{it}^{'}\alpha$ captures the covariates effect. Over this model, the measures presented in Cressie and Mugglin (2000) are applied by the authors. They allow to detect which regions don't follow the general pattern. In this case, the authors don't try to detect the onset of an epidemic, but a location or locations where the evolution of the risk of infection develops in a different way than everywhere else.

With the same objective of detecting regions with different behavior, Li et al. (2012) propose a Poisson spatio-temporal model for the detection of unusual temporal patterns. In this model, the logarithm of the relative risk for the majority of the regions is modeled as a general mean plus a common BYM spatial structure (BYM_i) plus a common temporal Random Walk (RW_t). For the unusual regions, the logarithm of the relative risk is modeled as a particular mean plus an individual temporal Random Walk (RW_{it}):

$$\begin{aligned} Y_{it} &\sim \text{Po}(E_{it} R_{it}) \\ \log(R_{it}) &= \mu_0 + BYM_i + RW_t \qquad \text{with } Z_i = 0\,, \\ \log(R_{it}) &= \mu_i + RW_{it} \qquad \text{with } Z_i = 1\,. \end{aligned} \tag{3.12}$$

This is a clear example of how data adapting better to a certain structure or another can distinguish two different epidemic behaviors by means of hidden variables. One shall notice that, in this case, the hidden variables Z_i do not follow a Markov chain, but they are conditionally independent.

Another example of spatio-temporal Bayesian model for disease surveillance is found in Corberán-Vallet and Lawson (2014). In this proposal, the argument of the Poisson model is the sum of an endemic and an epidemic

parameters. The logarithm of the relative risk for the endemic part follows a linear regressor with ICAR ($ICAR_i$) and unstructured (ξ_i) spatial random effects and a spatio-temporal random effect (ε_{it}):

$$\begin{aligned} Y_{it} &\sim \text{Po}(E_{it}R_{it} + \lambda_{it}) \,, \\ \log(R_{it}) &= \mu + ICAR_i + \xi_i + \varepsilon_{it} \,, \end{aligned} \tag{3.13}$$

with zero-mean Gaussian priors for μ, ξ_i and ε_{it}. The epidemic parameter is proportional to the previous counts on the same and neighbor locations, with the particularity that the exponential of this proportional parameter has a temporal structure that reflects the seasonality of the epidemic:

$$\begin{aligned} \lambda_{it} &= \rho_{it} \left(Y_{it-1} + \phi_i \sum_{j \sim i} Y_{jt-1} \right) \\ \rho_{it} &= \exp\left(\text{Serfling harmonics}_i + RW_{it}\right) \,, \end{aligned} \tag{3.14}$$

with a beta prior distribution for ϕ_i. A multivariate model is also proposed where the incidence of a syndromic disease helps estimating the epidemic component of the disease of interest. Both diseases are modeled with the equivalent structures, except for the addition of a term $\psi\lambda^S_{it-1}$ to λ_{it}, where λ^S_{it} is the epidemic term of the syndromic disease and λ_{it} is the equivalent term for the disease of interest.

Several other Bayesian models are used in the spatio-temporal modelization and surveillance of diseases but, when trying to do outbreak detection, most of the Bayesian models involve a latent variable to distinguish between the two possible epidemic states, as we review in the next subsection.

3.3.5 HMM and MSM for spatio-temporal outbreak detection

The value of HMM and MSM as tools for the detection of outbreaks in temporal data has been discussed in Chapters 1 and 2. Similar strategies are also applied in spatio-temporal data, using one hidden Markov chain for each of the locations in consideration. Also, some spatio-temporal models

condition the values of the hidden variables Z_{it} not only on the value of Z_{it-1} (same region and previous time), but also on the spatial neighbors at the same or previous times (Z_{jt} or Z_{jt-1}, with j neighbor of i), forming a Markov random field. In Markov random fields the graph of conditional dependencies of the hidden Markov variables is no longer a chain, but a general graph where the temporal, spatial and spatio-temporal conditional dependencies can be represented.

The description and prediction of meningococcal infection incidence (but not explicitly the detection of outbreaks) is addressed by Knorr-Held and Richardson (2003), who propose a framework of spatio-temporal Poisson MSM. In this framework, the logarithm of the relative risk of infection follows an overall temporal random walk of order 2 ($RW(2)_t$) plus an independent random effect on the month (γ_t) plus a spatial ICAR term ($ICAR_i$) plus an epidemic term ($\boldsymbol{\beta}' \boldsymbol{X_{it}}$) that is activated only when the hidden variable Z_{it} of the MSM is estimated as 1:

$$\begin{aligned} Y_{it} &= \text{Po}(E_{it} R_{it}) \\ \log(R_{it}) &= RW(2)_t + \gamma_t + ICAR_i + Z_{it} \boldsymbol{\beta}' \boldsymbol{X_{it}} \,. \end{aligned} \tag{3.15}$$

Six forms of the epidemic term are considered, plus the option of not having any epidemic term:

Model	$\boldsymbol{\beta}' \boldsymbol{X_{it}}$
0	0
I	$\beta_1 \text{I}(Y_{it-1} > 0)$
II	$\beta_1 \text{I}(Y_{it-1} > 0$ or $Y_{jt-1} > 0$ for at least one $j \sim i)$
III	$\beta_1 \text{I}(Y_{it-1} > 0) + \beta_2 \text{I}(Y_{jt-1} > 0$ for at least one $j \sim i)$
IV	$\beta_1 \log(Y_{it-1} + 1)$
V	$\beta_1 \log(Y_{it-1} + \Sigma_{j \sim i} Y_{jt-1} + 1)$
VI	$\beta_1 \log(Y_{it-1} + 1) + \beta_2 \log(\Sigma_{j \sim i} Y_{jt-1} + 1)$

(3.16)

with $\text{I}(\cdot)$ being the indicator function which is equal to 1 if its argument is true and 0 otherwise. DIC (see Section A.2 of the appendices) is used by the authors as a tool for selecting among the seven possible modelings.

The modeling of Li and Cardie (2013) has some peculiar features that are worth discussing. This model (also used in Sun et al., 2014) works with twitter data, using the counts of influenza related tweets in one region and day as raw data, though the observations in the model are taken to be not the raw counts but the relative increase or decrease on the counts, that is, $(Y_{it} - Y_{it-1})/Y_{it-1}$. The second particular feature is the use of four Markovian stages; non-epidemic, rising epidemic, stationary epidemic and declining epidemic phases. A third particular feature is that the hidden Markov variables Z_{it} are not only conditionally dependent on Z_{it-1} but also on Z_{it+1} and on Z_{jt}, $j \sim i$ through a generalized linear model. All the stages have a day of the week term modeled as a Gaussian process, and the rising epidemic and declining epidemic phases also have a temporal autoregressive structure.

Another example of a model where the conditional dependence of the hidden variable Z_{it} is not only temporal but also spatial is found in Banks et al. (2012). The authors present a theoretical multivariate Bayesian framework for syndromic surveillance on multiple data streams where, for each type of event, a Poisson model with three components is set:

$$Y_{it} \sim \text{Po}(\nu_{it} + Z_{it}\lambda_{it}) \,, \tag{3.17}$$

where $\log(\nu_{it})$ is modeled by a linear regression with periodic terms and several covariates plus a spatio-temporal term to be chosen among several options involving spatially unstructured random effects, spatial CAR distributions, temporal autoregressive structures and joint Gaussian spatio-temporal distributions. The ν_{it} terms represent the endemic part of the observed cases. $\log(\lambda_{it})$ is modeled by another regressor similar to the previous one but representing the epidemic part of the cases. For the hidden variables Z_{it} indicating the endemic or epidemic phase, the logit of the probability of being in epidemic phase $\text{logit}(p_{it})$ are linearly modeled by means of the counts on neighbors. Regrettably, the application presented in the article is the one firstly presented in Niemi et al. (2008) about drug abuse, which is actually a different more simple Bernoulli model with a CAR structure in its regressor and without the Z_{it} variable.

Two works which actually do present models within the theoretical framework presented in Banks et al. (2012) are those of Zou et al. (2012) and Heaton et al. (2012). The first layer of both models is a Poisson distribution for the counts of influenza cases with a non-epidemic term ν_{it} and an epidemic term λ_{it} which is only present when the hidden variable Z_{it} is equal to 1. In these two works, the conditional distribution of Z_{it} depends on Z_{it-1} through a matrix of transition probabilities which also depends on Z_{jt}, $j \sim i$. The probability of going from the non-epidemic state to the epidemic state is thus the probability of spontaneous epidemic p_s if no neighbor is in the epidemic phase. When at least one of the neighbors is in the epidemic state, the transition probability from endemic to epidemic is the combined probability of being infected by each neighbor:

$$p_{01,it} = p_s \mathrm{I}(\Sigma_{i\sim j} Z_{jt} = 0) + 1 - (1 - p_c)^{\sum_{i\sim j} Z_{jt}}\,, \tag{3.18}$$

with p_c the probability of being infected by one neighbor. The model focuses only on detecting the onset of the epidemic so the transition matrix is defined to be absorbent, that is, the probability of going from the epidemic state to the non-epidemic state is null: $p_{10} = 0$ and $p_{11} = 1$.

In the proposal of Zou et al. (2012) the logarithm of the non-epidemic term is modeled as a general mean plus covariates effects plus a white noise term:

$$\log(\nu_{it}) = \mu_\nu + \boldsymbol{\beta}_{\boldsymbol{\nu}}^{'} \boldsymbol{X}_{\boldsymbol{\nu it}} + \varepsilon_{it}\,, \qquad \varepsilon_{it} \sim \mathrm{N}(0, \sigma_\nu^2)\,. \tag{3.19}$$

The logarithm of the epidemic effect is modeled in a similar manner but with a spatio-temporal Gaussian term ST_{it} instead of the white noise term:

$$\log(\lambda_{it}) = \mu_\lambda + \boldsymbol{\beta}_{\boldsymbol{\lambda}}^{'} \boldsymbol{X}_{\boldsymbol{\lambda it}} + ST_{it}\,. \tag{3.20}$$

This spatio-temporal term is a fusion between a temporal autoregressive model and a spatial proper CAR:

$$ST_{it} \sim \mathrm{N}\left(\phi \frac{\sum_{j\sim i} ST_{jt}}{n_i} + \rho ST_{it-1}, \frac{\sigma_\lambda^2}{n_i}\right)\,, \tag{3.21}$$

with n_i the number of neighbors of the location i. In the proposal of Heaton et al. (2012) they opt for a general ICAR distribution common to all times in the non-epidemic phase and a spatio-temporal Gaussian term which is a fusion of a temporal autoregressive structure and a spatial ICAR with a slightly different variance term:

$$\begin{aligned} \log(\nu_{it}) &= \mu_\nu + \boldsymbol{\beta}'_{\boldsymbol{\nu}} \boldsymbol{X}_{\boldsymbol{\nu it}} + ICAR_i \\ \log(\lambda_{it}) &= \mu_\lambda + \boldsymbol{\beta}'_{\boldsymbol{\lambda}} \boldsymbol{X}_{\boldsymbol{\lambda it}} + ST_{it}\,, \\ ST_{it} &\sim \mathrm{N}\left(\frac{\sum_{j\sim i} ST_{jt}}{n_i} + \rho ST_{it-1}, \frac{\sigma^2_\lambda}{n_i+1}\right). \end{aligned} \tag{3.22}$$

In this case, the variance is scaled by n_i+1, reflecting the information given by the n_i spatial neighbors plus one temporal neighbor: the same region in the previous time period.

Both models are tested on data simulated from the models themselves. An application of the model of Heaton et al. (2012) on real data of weekly influenza mortality in 121 cities of USA can be found in the same article. The real data example of Zou et al. (2012) can be found in Zou et al. (2014), where the authors use the model with a day of the week term for daily ILI and respiratory syndrome data from 11 regions in Indiana.

Regarding prior distributions of the parameters, Zou et al. (2012) state that *the choices of hyperparameters represent vague prior information and ensure posterior propriety.* In the case of Heaton et al. (2012), the value of ρ is fixed at 0.5, and as p_s and p_c are expected to be small, an informative beta distribution Beta$(1, 30)$ is used. In Appendix B a discussion about a simplification of the general model of Banks et al. (2012) (of which the models of Zou et al. (2012) and Heaton et al. (2012) are particular cases or modifications, as we stated before) can be found. In this discussion it is proven that a simplification of these models, where no spatial or spatio-temporal terms are taken into account, and where improper non-informative priors are set for the parameters, has improper posterior distributions for certain parameters of the model. This suggests that the more complex alternatives of Zou et al. (2012) and Heaton et al. (2012) with the same main structure might also present problems with the inference process, forcing the elec-

tion of informative priors that might influence the results of the inference. An alternative is also discussed in this appendix, where the same Poisson distribution for the counts seen in Expression (3.17) is substituted by a 'switch' Poisson distribution, where the expected number of counts λ_{it} is not added to ν_{it} when the epidemic phase is active, but it substitutes the endemic term instead. Unlike the previously discussed model, the 'switch' model leads to a proper posterior distribution for the parameters of the model.

The model in Zou et al. (2014) is the most similar (though far from being equal) to the one we present in this chapter and is also one of the few models specially designed for spatio-temporal detection of the start of influenza outbreaks. Because of this, we wanted to compare the performance of our model with theirs. We found, though, several problems while trying to implement and run it on WinBUGS. We got in contact with the authors but they could not facilitate the code.

3.4 Modeling spatio-temporal influenza data for the detection of outbreaks

In this section we present a novel proposal for the spatio-temporal detection of influenza outbreaks. This method is an extension of the temporal method presented by Martinez-Beneito et al. (2008a). The idea for building this model is to use the different dynamics of the differentiated ILI rates (the jumps from one week to the next one) to distinguish between epidemic and endemic weeks of several location. This allows to detect changes in the dynamics that denote an outbreak regardless of being observed in low rates of influenza, as is usually the case at the start of an epidemic.

Influenza epidemics start in one or several geographical foci, which can appear in different weeks of the year. We intend to construct a model that can deal with the emergence of these multiple outbreaks. We also want to capture and use the spatio-temporal structure of the data in order to improve the detection power, because the way influenza epidemics spread suggests that neighboring regions tend to infect each other, and correctly

modeling this will improve the classification.

Seasonal epidemic outbreaks can occur in different weeks of the year. Some outbreaks of seasonal influenza can happen in December for one seasons while for the next season the start can be delayed till February. It can also be the case that one season has two different outbreaks or that there is no epidemic that year. In addition to that, non-seasonal influenza outbreaks can happen at any time of the year. Therefore our proposal does not make assumptions about the temporal location of the epidemics or the amount of them.

The proposed method makes use of data sources that provide information of the rates of influenza or ILI referenced both spatially and temporally. Spatial locations of the data can be expressed in several ways: individual cases may be associated with their home address, individual or aggregated cases may be assigned to a health care facility –hospitals, emergency rooms, pharmacies, etc.–, or they can be associated to an administrative region. All the possibilities can, in general, be translated to aggregated rates in administrative regions, which results in spatially discrete support of the information structured as a lattice, where a neighboring rule can be established. This last format may in some cases be coarser, but ensures that almost every spatial data of any surveillance system can be translated to it. Therefore, with the intention of making our method as versatile as possible, we model it to deal with lattice data. For the time dimension, we also choose discrete and equally spaced temporal information (daily or weekly) divided in seasons, which is the usual way of reporting data to surveillance systems.

An example of influenza rates over spatio-temporal discrete support is shown in Figure 3.1. This figure displays the original map of administrative regions (in this case, states) of the USA and the transformation of this map into a lattice support with a neighboring network defined by sharing a border. Below the spatial depiction of the map and the spatial lattice, the temporal chains for each location of estimated influenza rate by Google Flu Trends (with 4 of the 49 chains highlighted in different colors as examples) are shown. In the bottom of the figure the differentiated estimated rates are plotted. These temporal chains are the data used by our proposal.

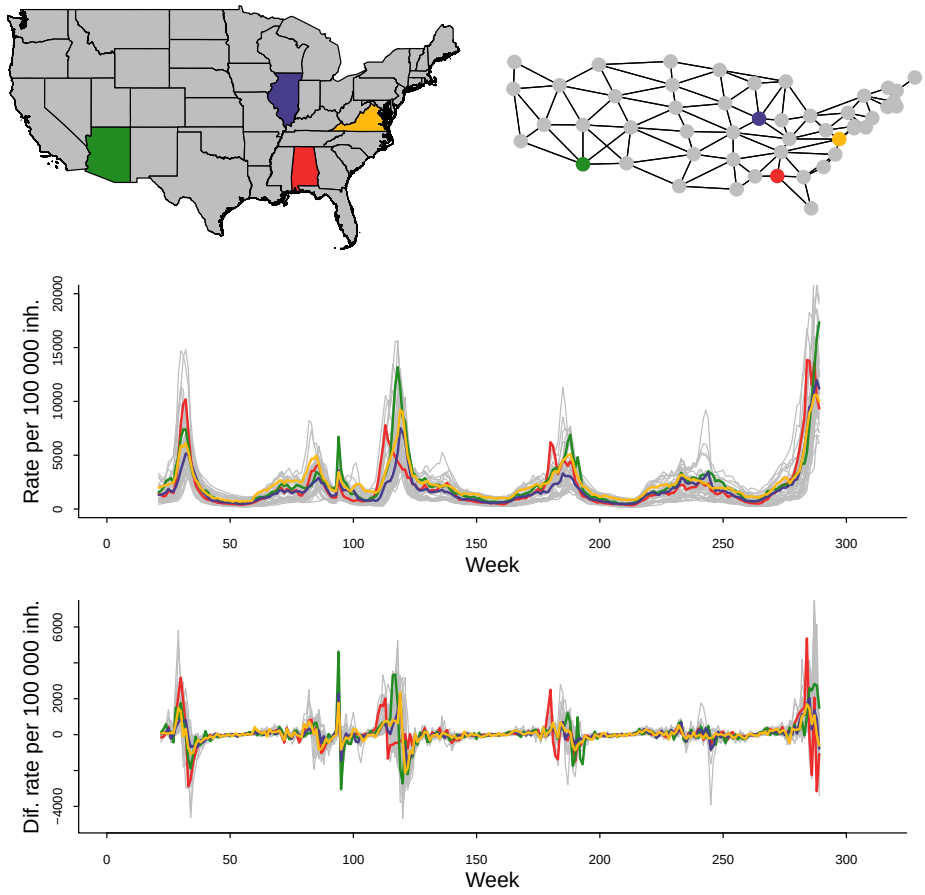

Figure 3.1: Map of USA, graph of neighborhood defined by sharing a border, estimated influenza rates by Google Flu Trends between 2007 and 2013 and the corresponding differentiated rates. Highlighted regions: Alabama in red, Arizona in green, Illinois in blue and Virginia in yellow

3.4.1 Modeling the differentiated rates

In the introduction of this chapter we have stated that for our spatio-temporal proposal we model the differentiated rates instead of modeling the counts, which was done on our proposed temporal framework of models in Chapter 2. This results in a simpler and faster model (an important feature, given the complexity that provides the spatial data and structure) which is more adaptable to different data sources and do not focus the attention on the magnitude of the rates. This can be an advantage because in some cases using the magnitude of the rates can difficult the detection of the beginning of an outbreak, as several outbreaks happen when influenza rates are still low.

For simplicity in the notation, given a location i and season s, we name y_{its} to the rate in time t (r_{its}), minus rate in time $t-1$ (r_{it-1s}):

$$y_{its} = r_{its} - r_{it-1s}\,. \tag{3.23}$$

We start the indices t of the raw rates r_{its} at 0 so that the indices t of the differentiated rates y_{its} used in the model start at 1. The variable Z_{its} indicates the latent epidemic state for each time and location, with a value of 1 for the epidemic state and a value of 0 for the non-epidemic state. We consider that the differentiated rates follow a normal distribution with mean and variance depending on the epidemic state:

$$y_{its} \sim \mathrm{N}(R_{itsZ_{its}}, \sigma^2_{Z_{its}})\,. \tag{3.24}$$

Note that in this model, the notation $R_{itsZ_{its}}$ has not the same meaning as in the model presented in Chapter 2, where $R_{tsZ_{ts}}$ was the expected value of the raw rate r_{ts} given Z_{ts}. In this case, it is the expected value of the differentiated rate $r_{it+1s} - r_{its}$ given Z_{ts}.

In Section 2.2.1 we have discussed that the common practice in disease surveillance of modeling the logarithm of the rates with a Gaussian distribution may difficult the detection of the onset of epidemics. This procedure is also common in disease mapping and other branches of spatial and spatio-temporal disease surveillance. In this case we do not take logarithms of the raw rates before differentiating them either. If we were to

do so and then differentiated them, we would have that our data would be $\log(r_{it+1s}) - \log(r_{its}) = \log(\frac{r_{it+1s}}{r_{its}})$. In this situation the model would only consider the percentage of the growth and not its magnitude. Doubling the rate would result in the same value for the response variable, whether the jump on incidence was from a rate of 20 cases per 100 000 inhab. to 40 (reasonable jump during the non-epidemic phase) or it was from 2000 to 4000 (more likely during the epidemic phase). We do not take logarithms to the differentiated rates either, as this makes no sense because approximately half of them are negative values.

3.4.2 The hidden Markov structure for the epidemic phase

In order to model the two possible phases in which each location and time can be classified, we model a set of lattice variables Z_{its} (equal to 1 to indicate epidemic phase and equal to 0 to indicate non-epidemic phase) as a hidden Markov chain for each location. In an equivalent way as described in Section 1.6, the distribution of the latent variable Z_{its} for each location i and season s conditioned to Z_{it-1s} follows a Bernoulli distribution with transition probabilities common for all times, locations and seasons:

$$Z_{its} \sim \text{Ber}(p_{Z_{it-1s}1})\,. \tag{3.25}$$

Jeffreys non-informative prior densities for Bernoulli trials are set for the transition probabilities in the same way as those described in Section 2.2.2 for the temporal proposal:

$$p_{00}, p_{11}, p_0 \sim \text{Beta}\left(\frac{1}{2}, \frac{1}{2}\right)\,, \tag{3.26}$$

where

$$\begin{aligned} p_{kl} &= P(Z_{its} = l | Z_{it-1s} = k)\,, \\ p_k &= P(Z_{i1s} = k)\,, \qquad\qquad k, l \in \{0, 1\}\,, \end{aligned} \tag{3.27}$$

and the rest of probabilities p_{10}, p_{01} and p_1 are obtained by complementarity. As we have described in the previous chapter, the initial probabilities

for each season p_0 and p_1 are not set as the stationary distribution for the transition probabilities because the initial week of each season is not randomly chosen, but is usually an endemic week.

The posterior probabilities of these Z_{its} are estimations of the probability of each location in each time to be in the epidemic state, and are used for declaring epidemic alarms. This response is not a simple yes or no so it can be used by the public health authorities to make better informed decisions which actually take into account the uncertainty of the estimation.

3.4.3 Modeling the mean of the differentiated rates

Some features used in this proposal that help characterizing the epidemic phase are the behavior of the expected value of the differentiated rates, its spatial and temporal correlation and the higher structured variability for the epidemic state. A higher temporal autoregressive structure in the epidemic phase helps distinguishing the consecutive epidemic growths and decays of the rates from the less correlated endemic differentiated rates. The spatial structure captures the contagion effect, and the higher structured variability helps distinguishing the sharper epidemic jumps from the milder endemic ones.

The non-epidemic phase is characterized for jumps on the rates close to zero. A first approach could be just leaving the mean to be 0, but a close look at the data (see Figure 3.2) shows that there are non-epidemic weeks where small growths or decreases happen on the rates in all locations at the same time. This may be manifesting a stationary behavior of the endemic weeks, with some weeks of common small decreases after the cold season (winter in the Northern Hemisphere) and some weeks of common small increases after the hot season (summer in the Northern Hemisphere). In any case, these increases or decreases do not happen at the same exact weeks in general, and non-stationary epidemics can break this dynamic. For this reason, we consider each week t of each season s to have a mean value μ_{ts0} which is common to all locations at that week, but different from the mean of other weeks. No further spatial or temporal structure is considered for the endemic phase.

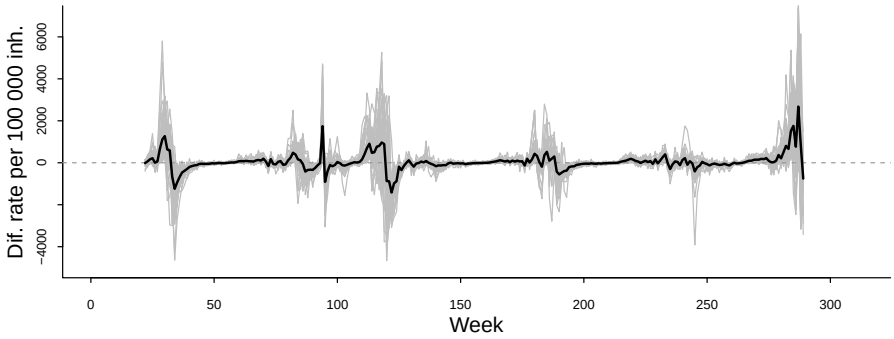

Figure 3.2: Estimated influenza differentiated rates by Google Flu Trends between 2007 and 2013 in USA for all the regions in gray and the mean of all the regions in black.

The epidemic phase is modeled with a more complex spatio-temporal structure than that of the endemic phase. It is expected for a region in epidemic state to have several positive jumps (differentiated rates) until reaching the peak of the epidemic and then the jumps are to become negative whereas the rates decrease again to the endemic level. Therefore, it is expected that the differentiated rates for each location in the epidemic state are temporally dependent.

Another expected behavior caused by the contagious nature of influenza (corroborated by observation of real data) is that if one region has epidemic growths on its incidence rates, neighbor regions may become infected and show similar growths. As a consequence, we model the mean of the differentiated rates by a term μ_{ts1}, common for all the locations (the overall epidemic rate for that week) but different for each time, plus a temporal auto-regressive structure of order 1 on the observations for each location with parameter ρ, plus a spatial intrinsic conditional auto-regressive (ICAR) model for each time. The mathematical expressions for the mean

of both endemic and epidemic periods are:

$$\begin{aligned} R_{its0} &= \mu_{ts0}\,, \\ R_{its1} &= \mu_{ts1} + \rho y_{it-1s} + \psi_{its}\,, \end{aligned} \tag{3.28}$$

with ψ_{its} being the ICAR term. Let us describe each of these three components:

The common term for each time unit

As stated before, the common term of the endemic phase μ_{ts0} can capture the mild common seasonality of non-epidemic ILI. In the case of the epidemic season, μ_{ts1} models the common rise or fall of the rates during the epidemic. In both cases, these terms model the weekly consensus along all the states in the same epidemic phase, which in general is positive in ascending phases (before the epidemic peak) and negative in descending phases (after the epidemic peak). These factors do not appear in the temporal model of Martinez-Beneito et al. (2008a), as only one region is analyzed and, therefore, no inference on the weekly consensus for all the regions can be done.

We consider μ_{ts0} and μ_{ts1} as two random effects over time, with larger variability for μ_{ts1}, as common rises or falls of the epidemic phase are expected to be larger than those of the endemic phase. In order to avoid identifiability problems with the main unstructured variability $\sigma_{Z_{its}}$ of the main distribution, we set the variances of the two random effects to be proportional to those of the main distribution. We can express the modeling of the common temporal terms as follows:

$$\begin{aligned} &\mu_{ts0} \sim \mathrm{N}(0, \sigma^2_{\mu 0})\,, && \sigma_{\mu 0} = \lambda \sigma_0\,, && \\ &\mu_{ts1} \sim \mathrm{N}(0, \sigma^2_{\mu 1})\,, && \sigma_{\mu 1} = \lambda \sigma_1\,, && \lambda \sim \mathrm{Unif}(0, a)\,, \end{aligned} \tag{3.29}$$

where λ is the estimated proportion factor and a is a hyperparameter to be set by the modeler expressing a vague prior knowledge. For example, $a = 100$ would be a possible choice, as it is really unlikely that this part of the structured variability is so much larger than the unstructured one.

The autoregressive structure in the epidemic mean

During the epidemic state of influenza, it is common to observe a rise on the rates during several weeks and, after reaching the peak, a descend of the rates that also lasts several weeks. In contrast to the lack of temporal structure of the non-epidemic mean, the autoregressive structure on the growths is able to fit this behavior. This is one of the features that allow to distinguish between epidemic and non-epidemic states.

To ensure the stationarity of the autoregressive process we must bound the parameter ρ to the interval $[-1, 1]$. Anyhow, assuming that the correlation between subsequent growths in the epidemic phase will be positive, in practice we can restrict the interval to $[0, 1]$:

$$\rho \sim \text{Unif}\,(0, 1)\ . \tag{3.30}$$

In the same fashion as in Section 2.2.4, and to ensure that the variance of y_{i1s} is equal to the stationary variance of the series $\{y_{its}\}_{t=1}^{\infty}$, the Expression (3.24) for the first week of each season is modified as follows:

$$y_{i1s} \sim \text{N}\left(R_{i1sZ_{i1s}}, \frac{\sigma^2_{Z_{its}}}{1 - Z_{i1s}\rho}\right)\ . \tag{3.31}$$

As mentioned in Section 2.2.4, this type of correction is introduced, for example, in works like that of Martinez-Beneito et al. (2008b) in the context of disease mapping.

The ICAR structure in the epidemic mean

We have seen how the autoregressive structure models similar jumps for consecutive times for each location during the epidemic phase. In an analogous way, the ICAR structure captures how neighbor locations in the epidemic phase have similar behavior in the growth of rates which differ from the general common behavior. Clusters of locations where the influenza is spreading show growths on rates that adapt better to this modeling, while neighboring regions where the disease has not extended yet adapt better

to the absence of any spatial structure (endemic state). This behavior is theoretically expected due to the nature of influenza data, but it has also been empirically suggested in the work of Fox and Dunson (2015). The authors propose a Bayesian nonparametric covariance regression, which allows the variance matrix in a multivariate regression model to vary with the predictors among time. The model is exemplified with an application on USA Google Flu Trends data between 2003 and 2009. The application does not assume any prior spatial structure but the estimated posterior correlation matrices among regions for each week show it. In addition, the estimated correlation values are higher during the epidemic weeks and lower on non-epidemic weeks.

The terms ϕ_{its} for a given week t and season s follow a joint ICAR distribution as described in Section 3.2. We set a non-informative prior for conditional standard deviation of ψ_{its}:

$$\sigma_\psi \sim \text{Unif}(0, b)\,, \tag{3.32}$$

with b a hyperparameter to be fixed by the modeler expressing a vague prior knowledge. In a similar way as in Section 2.2.3 for the temporal framework of models, the hyperparameter b may be chosen to be any value above the highest differentiated rate in absolute value, to ensure that the prior distribution is really vague.

This structure is additional to the common parameterization of the rise and fall of the rates during the epidemic phase, expressed in Expression (3.29). The existence of this μ_{t1} is necessary as the ICAR distribution is constrained to sum zero and therefore an additional component is needed to model the mean of the epidemic period at each week.

3.4.4 Modeling the variance of the differentiated rates

We have seen that one important aspect for the discrimination between epidemic and non-epidemic phases in our proposal is the behavior of the expected value of the differentiated rates, its spatial and temporal correlation and the higher structured variability for the epidemic state. Another

feature that helps characterizing the epidemic phase will be the unstructured variability.

In general, differentiated rates are centered in zero, therefore, though the behavior of the mean of the response variable gives important information to distinguish between epidemic phases, the behavior of the variance of the data is critical for this task. The jumps of the rates on the non-epidemic state are relatively small in absolute value, while the growths or decreases of the rates on the epidemic state are usually larger. A wider variance on the differentiated rates is therefore expected for the epidemic state.

In Section 2.2.3 we have already seen that characterizing the epidemic state by a higher variance has been successfully done in works like those of Martinez-Beneito et al. (2008a) and Nunes et al. (2013). The first work proposes a model on the differentiated rates and the second proposal models the raw rates. In some spatio-temporal proposals, like those of Zou et al. (2012) and Heaton et al. (2012), the extra variance of the data on the epidemic phase is modeled by the extra variance of the added spatio-temporal structure. In our case, besides the structured noise that is fitted by the spatio-temporal term for the mean, we also consider an unstructured noise in the epidemic phase, getting a combination of structured and unstructured noise in the fashion of a BYM model Besag et al., 1991. We also ensure that the unstructured noise of the epidemic phase is higher than that of the epidemic phase, emphasizing the importance of the variance as a tool for classification and avoiding the interchangeability of these terms.

Therefore, we model the unstructured variability of the two phases by obtaining both standard deviations from a uniform prior and ensuring that they are ordered:

$$\begin{aligned} \sigma_0 &= \theta_{[1]} \\ \sigma_1 &= \theta_{[2]} \\ \theta_m &\sim \text{Unif}(0, c) \qquad\qquad m = 1, 2\,, \end{aligned} \tag{3.33}$$

with c a hyperparameter to be set by the modeler expressing vague prior knowledge. Setting c as a value above the largest differentiated rate in absolute value is a sensible choice, as also suggested in Section 3.4.3 for the

standard deviation of the ICAR structure.

Unlike the temporal proposal, the spatio-temporal model does not consider different variances for each season. Peaks of raw rates may be larger or smaller depending on the season, and thus, the variance of the epidemic rates may be larger or smaller. That is why we did this distinction in the temporal model, but when dealing with the differentiated rates one is modeling the velocity with which the incidence grows or decays. We consider this velocity of growth and decay as a characteristic of the epidemic and endemic phases that remains similar over the seasons. Also, sharing the same epidemic and endemic variance parameters across years helps to better distinguish between phases during the first weeks of each season, where not enough information is available to estimate a new variability parameter.

3.4.5 The complete model

For clarity in the understanding of the proposed method, we present again all the previously introduced equations which make up the hierarchical model:

$$
\begin{aligned}
y_{its} &\sim \mathrm{N}(R_{itsZ_{its}}, \sigma^2_{Z_{its}})\,, \\
R_{its0} &= \mu_{ts0}\,, & \mu_{ts0} &\sim \mathrm{N}(0, \sigma^2_{\mu 0})\,, \\
R_{its1} &= \mu_{ts1} + \rho y_{it-1s} + \psi_{its}\,, & \mu_{ts1} &\sim \mathrm{N}(0, \sigma^2_{\mu 1})\,, \\
\psi_{its} &\sim \mathrm{N}\left(\frac{\sum_{j\sim i}\psi_{jts}}{n_i}, \frac{\sigma^2_\psi}{n_i}\right) \\
Z_{its} &\sim \mathrm{Ber}(p_{Z_{it-1s}1})\,, \\
\sigma_0 &= \theta_{[1]}\,, \\
\sigma_1 &= \theta_{[2]}\,, & \theta_m &\sim \mathrm{Unif}(0, c)\,, \quad m = 1, 2\,, \\
\sigma_{\mu 0} &= \lambda\sigma_0\,, \\
\sigma_{\mu 1} &= \lambda\sigma_1\,, & \lambda &\sim \mathrm{Unif}(0, a)\,, \\
\rho &\sim \mathrm{Unif}\,(0, 1)\,, \\
\sigma_\psi &\sim \mathrm{Unif}(0, b)\,, \\
p_{00}, p_{11}, p_0 &\sim \mathrm{Beta}\left(\frac{1}{2}, \frac{1}{2}\right)\,.
\end{aligned}
$$

As it occurs with the temporal proposal in Chapter 2, there is no analytic expression for the posterior distribution of the parameters of the hierarchical model presented in this spatio-temporal proposal. Therefore, MCMC simulation has been carried out using the software WinBUGS. The code of the model can be found in Appendix D. As done in the previous chapter, the posterior probability of the hidden variables Z_{its} are the estimations of the probability of being in epidemic for each region, time and season. These estimations for the last week available are the tool used to trigger the epidemic alarms by a surveillance system.

3.5 Application on United States Google Flu Trends data

In the past section we have defined our spatio-temporal proposal of a Markov Switching model over the differentiated rates for the detection of influenza outbreaks. In the present section we show an application of the proposed spatio-temporal model on United States Google Flu Trends data. We also compare the proposal with some simplifications of the model to stress the pertinence of using several of the statistical structures which conform this proposal.

3.5.1 The USA Google Flu Trends data

In Section 1.3.1 it has been discussed that there are several kind of data sources and several ways to retrieve data, with different specificity and timeliness. We took special interest in the automated collection of data from search engines, like Google Flu Trends (GFT), which provide almost immediately large amounts of data from vast populations and territories without the purposeful collaboration of the individuals. It has also been discussed that, over the years, the original algorithm ceased to be suited to estimate the influenza incidence, probably because some conditions and some users' behaviors changed (for example, with the appearance of a non-seasonal strain of influenza like the swine flu). For this reason the algorithm has been revised several times, specifically attending to the prediction of influenza incidence rates from the CDC for the USA, its states and its main cities. In any case, in the following sections do not pay attention to whether the data are narrowly accurate when estimating the CDC published incidence rates but we just consider that it is a realistic data source, regardless of its quality.

The data we use in this chapter, represented in Figure 3.1, consist of weekly estimates of influenza incidence for the 48 spatially connected states of the USA plus Washington, D.C., between 2007-12-02 (48th week of the year) and 2013-01-20 (3rd week of the year). We separate the seasons on the middle of summer, that is, on week 26 of each year. Data used by the

model are the differentiated rates for each state. No virus isolates or other more reliable data segregated by states that could be used to provide a gold standard were available.

3.5.2 Retrospective estimates of the epidemic phase in space and time

As a first approach to evaluate the performance of the spatio-temporal model on the GFT USA data described above, we show the results of the retrospective application of the model. The real use of a detection method is prospective, but the retrospective application is computationally cheap and can offer a proxy of its performance which can be used for its evaluation. In Figure 3.3 one can observe the estimated posterior probability of the epidemic phase for four randomly chosen states of the 49 possible ones which we use as an example for the rest of the chapter (Alabama, Arizona, Illinois and Virginia). In Section C.1 of the appendices, all estimates for the 49 states are displayed. The retrospective detection of the epidemic phase shows a sensible behavior that appears to give high probability of epidemic to those weeks with step growth or decay around the peak of the epidemic for each season.

Parameter	Posterior mean	95% Credible interval
p_{00}	0.95	[0.95 , 0.96]
p_{11}	0.89	[0.87 , 0.90]
σ_0	85.00	[82.99 , 87.01]
σ_1	430.89	[404.70 , 458.70]
$\sigma_{\mu 0}$	127.61	[116.60 , 138.40]
$\sigma_{\mu 1}$	646.69	[583.28 , 701.00]
σ_ψ	770.10	[716.90 , 822.70]
ρ	0.35	[0.31 , 0.38]
λ	1.50	[1.38 , 1.62]

Table 3.1: Posterior mean and 95% credible interval for the parameters of the spatio-temporal model applied on the GFT USA data in a retrospective basis.

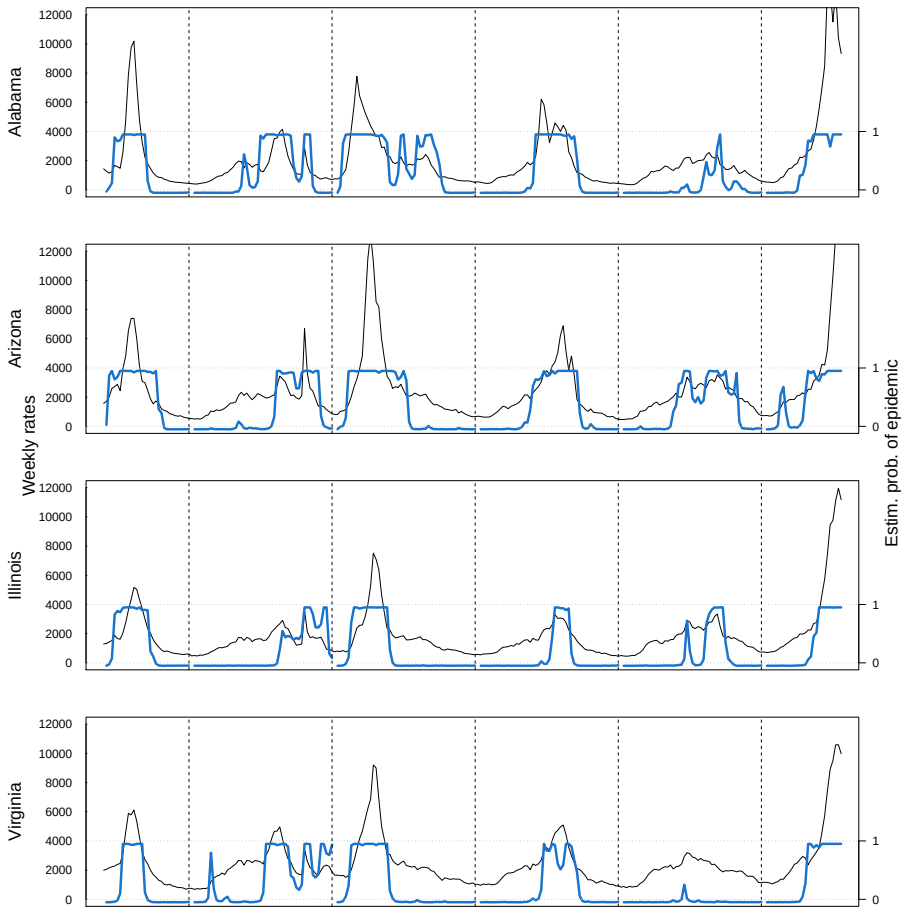

Figure 3.3: Retrospective estimated probability of being in epidemic phase by the spatio-temporal model on GFT USA data for 4 states. In black: weekly estimated influenza incidence per 100 000 inhabitants during seasons from 2007–2008 to 2012–2013.

In Table 3.1 the posterior mean and 95% credible interval for the most relevant parameters of the model are shown. The high values of p_{00} and

p_{11} indicate that weeks in a certain phase tend to be followed by weeks of the same phase. The typical season has one peak, so usually there is one transition from non-epidemic to epidemic phase and one transition from epidemic to non-epidemic phase per season. Roughly speaking, p_{00} takes a value around 1 over the average quantity of endemic weeks per season, while p_{11} takes a value around 1 over the average quantity of epidemic weeks per season. The fact that $p_{00} > p_{11}$ comes therefore because the mean length of the epidemic phase is shorter than the mean length of the endemic phase. The standard deviations associated to the endemic phase, σ_0 and $\sigma_{\mu 0}$, take notably lower values than those of the epidemic phase, σ_1, $\sigma_{\mu 1}$ and σ_ψ. This indicates that the variance of the differentiated rates is a key point for the classification of the two states of the Markov chain. All the credible intervals are quite narrow, which is an indicator of the identifiability of all the parameters of the model.

3.5.3 Online versus retrospective application of the spatio-temporal model

For a real life application of the spatio-temporal proposal in a surveillance system, the estimates of the epidemic phase would have been obtained week by week. Therefore, for each week only data from that week and previous weeks (but not following ones) could have been used to estimate the epidemic state. To reproduce this realistic behavior one must apply the model in an online basis. Figure 3.4 compares the posterior probability of epidemic for the last season of the GFT USA data for the 4 states used as example previously mentioned. In Section C.2 of the appendices, the comparison for all 49 states is displayed.

The results of both ways of applying the model are similar, though the online estimates tend to take a little bit longer to give high posterior probability to the epidemic phase. This behavior is expected, as the retrospective way of applying the method can use information from posterior weeks when estimating the phase, while the online application can not.

Paradoxically, one can also observe that the online outcome is more consistent about the estimated phase along time, in the sense that not many

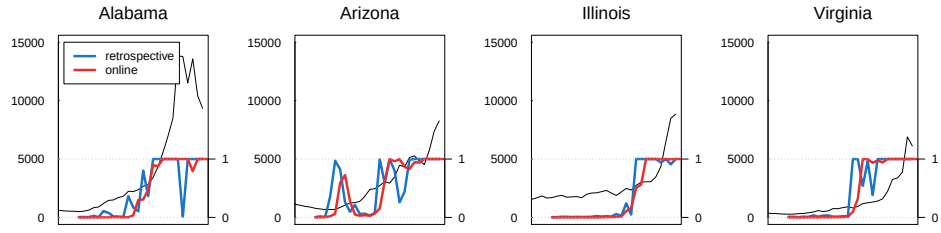

Figure 3.4: Comparison of the online and retrospective estimated probability of being in epidemic phase by the spatio-temporal model on GFT USA data for 4 states. In black: weekly estimated influenza incidence per 100 000 inhabitants during season 2012–2013.

ups and downs are observed in the evolution of the posterior probability of epidemic through the weeks. In contrasts, the retrospective estimate shows a more erratic line, jumping from high and low posterior probabilities of epidemic.

Figures 3.5, 3.6 and 3.7 show maps of the estimated posterior probability of being in epidemic phase for some of the weeks of the last season when applying the model in an online basis. One can appreciate how the model makes a sensible estimation of the spread of the outbreak, which starts from the south and east of the United States and ends up covering all the country.

3.5.4 Comparison with alternative proposals

In order to evaluate the relevance of the proposed spatio-temporal model, in this section we show a comparison of the new proposal with two simplifications of the proposal, one variation and the model of Martinez-Beneito et al. (2008a) in which it is based.

Simplifications and variations of our proposals

Here we describe the two simplifications and the variation we are going to compare the spatio-temporal proposal with. We also explain how we apply

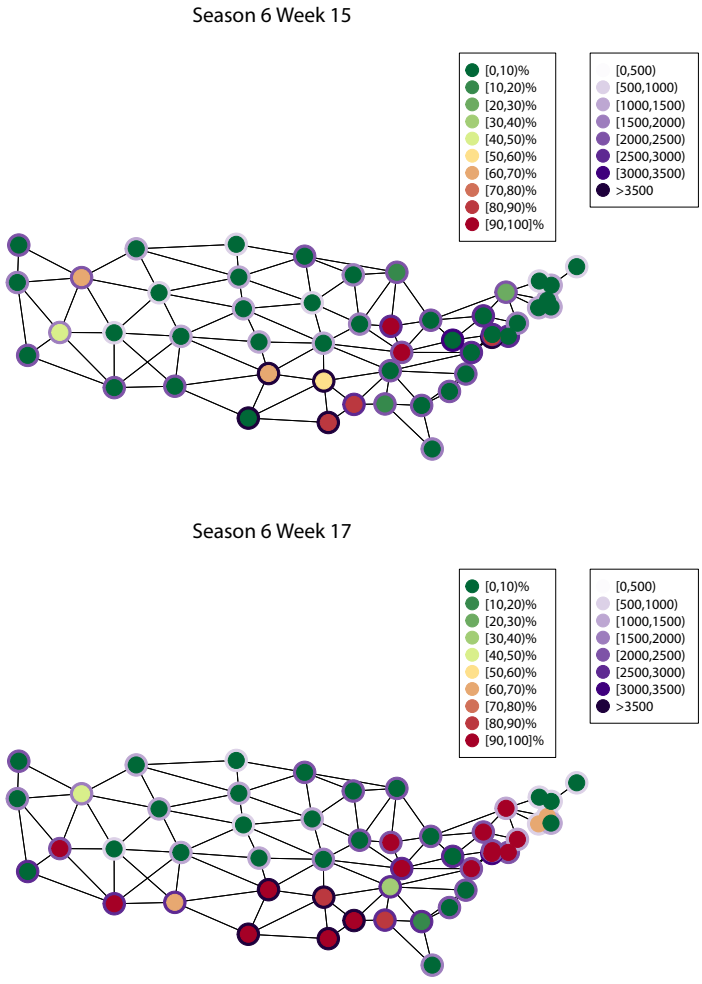

Figure 3.5: Online estimated probability of being in epidemic phase by the spatio-temporal model on GFT USA data (green-red scale). Weeks 15 and 17. Contours in purples: estimated influenza incidence per 100 000 inhabitants.

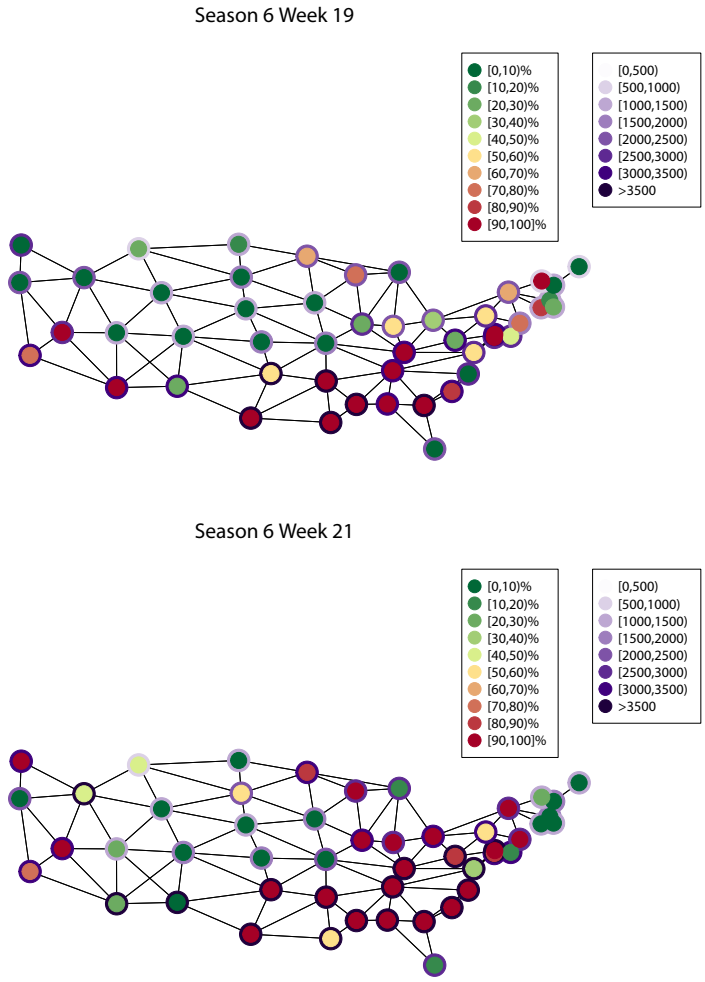

Figure 3.6: Online estimated probability of being in epidemic phase by the spatio-temporal model on GFT USA data (green-red scale). Weeks 19 and 21. Contours in purples: estimated influenza incidence per 100 000 inhabitants.

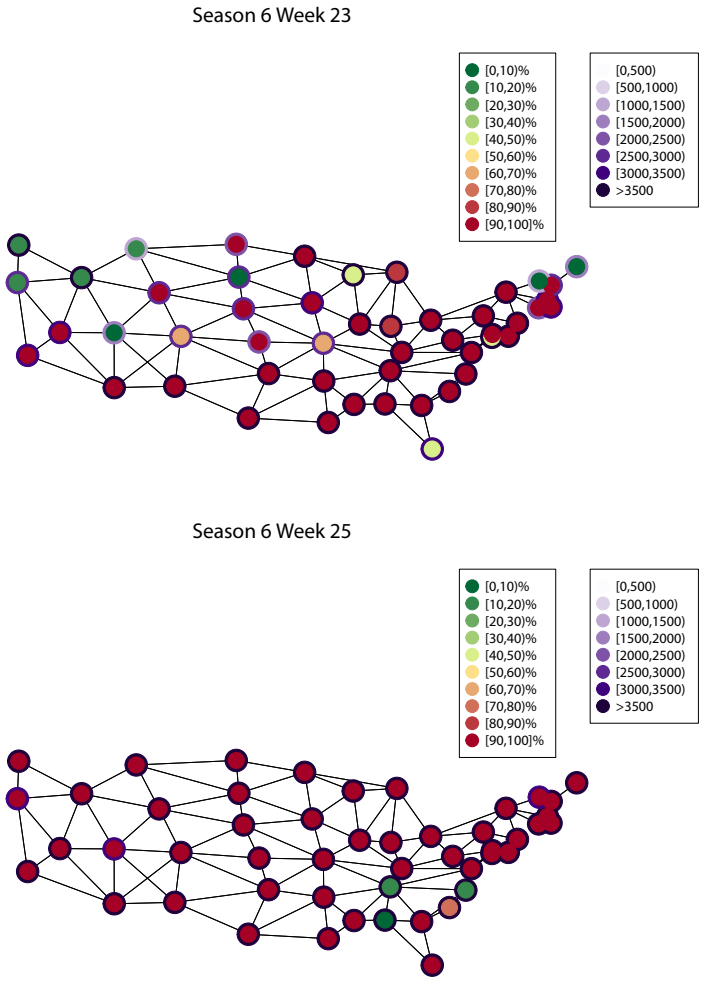

Figure 3.7: Online estimated probability of being in epidemic phase by the spatio-temporal model on GFT USA data (green-red scale). Weeks 23 and 25. Contours in purples: estimated influenza incidence per 100 000 inhabitants.

the model of Martinez-Beneito et al. (2008a) over spatio-temporal data.

Removing the spatial component. The first proposed simplification is removing the structured spatial random effect, the intrinsic CAR component, from the modeling of the epidemic phase. To do so, the lower equation in Expression (3.28) loses the ψ_{its} term. The modified expression is as follows:

$$\begin{aligned} R_{its0} &= \mu_{ts0} \\ R_{its1} &= \mu_{ts1} + \rho y_{it-1s}\,. \end{aligned} \tag{3.34}$$

By comparing this simplification with the original new proposal one can assess how the assumption of similar growths of incidence among neighbors in epidemic phase affects the fitting and the phase classification.

Removing $\boldsymbol{\mu_{ts0}}$. As it is indicated in Section 3.4.3, the term μ_{ts1} for the epidemic phase is necessary because of the constraint to zero of the ICAR term. Hence the term μ_{ts0} is added to the model to capture the mild common jumps on the endemic phase, but also thinking that it brings balance of complexity among the regressors of the epidemic and endemic phases, preventing an excessive sensitivity of the model. To check that this choice is appropriate, the second simplification we compare is the removal of this random effect on the common mean of the differentiated rates on the non-epidemic phase for each time, μ_{ts0}. By doing so, the mean of the non-epidemic phase is set equal to 0. Expression (3.28) for this modification becomes now as follows:

$$\begin{aligned} R_{its0} &= 0 \\ R_{its1} &= \mu_{ts1} + \rho y_{it-1s} + \psi_{its}\,, \end{aligned} \tag{3.35}$$

changing the original term μ_{ts0} for a 0. Also, Expression (3.29) is substituted for this modification by the following, so that μ_{ts1} has a variance that is conditionally independent from σ_0 and σ_1:

$$\mu_{ts1} \sim \mathrm{N}(0, \sigma^2_{\mu 1}) \qquad\qquad \sigma_{\mu 1} \sim \mathrm{Unif}(0, d)\,. \tag{3.36}$$

The hyperparameter d is set to a value above the largest differentiated rate (in absolute value), as suggested in Sections 3.4.3 and 3.4.4.

Leroux. The variation of the model we compare the new proposal with is the substitution of the ICAR term on the epidemic regressor by the spatial structure proposed by Leroux et al. (2000), previously mentioned in Section 3.2. Expression (3.28) remains the same, but in this case, the definition of the conditional distribution of the terms $\psi_{its}|\boldsymbol{\psi_{-its}}$ is set as a normal distribution with conditional mean and variance as those expressed in the work of Leroux et al. (2000).

Observing the conditional distribution of the Leroux effect as described in Expression (3.6), and taking into account that ϕ is bounded between 0 and 1, one may notice the following: the mean of the conditional distribution monotonically increases with ϕ, while the variance decreases as ϕ gets bigger (except when $n_i = 1$, where the variance is constantly σ^2_ψ). In the limit cases, the distribution is that of an unstructured random noise with 0 mean and variance equal to σ^2_ψ when $\phi = 0$ and an ICAR distribution, with mean equal to $\frac{1}{n_i}\sum_{j\sim i}\psi_{jts}$ and variance equal to $\frac{\sigma^2_\psi}{n_i}$, when $\phi = 1$. All intermediate values of ϕ result in a Gaussian distribution with mean and variance values in between. In that way, the parameter ϕ distributes the variability of the Leroux term between spatially structured and unstructured variability. We compare our proposal with this variation to check if the ICAR term forces a too strong spatial relation which Leroux model could be able to soften.

M-B 2008. M-B 2008 is the abbreviation we use in this chapter for the model of Martinez-Beneito et al. (2008a). It has already been explained in detail in Section 1.7 and is the model in which our proposal is based. The model is run for each of the 49 locations of the GFT USA data set independently, so no information (whether spatial or of any other nature) has been shared among them. The comparison with this model indicates the relevance of the spatio-temporal

modeling against the purely temporal modeling which does not share information among different geographic units at all.

Comparison of computational cost

Table 3.2 shows computation time in minutes for the spatio-temporal proposal and its variations, run on all the GFT USA data set in a retrospective basis. Calculations were performed in an Intel® Core^{TM} CPU I7–3770 with 4 cores at 3.40GHz and 8Gb of RAM, with OS Windows 7 Professional 64 bits. A WinBUGS process with 2 chains was run for each model, with 1 000 iterations of burning and 3 000 subsequent iterations. After thinning, 1 000 iterations were kept, 500 from each chain. Convergence was checked by observing the effective sample size, the $\hat{R}$ statistic (see, for example, Gelman et al. (2013)) and visual check of the chains of simulations.

Reference	removing ICAR	removing μ_{ts0}	Leroux	M-B 2008
64	40	38	55	19

Table 3.2: Computational cost in minutes for the spatio-temporal model applied on the GFT USA data in a retrospective basis, its variations and Martinez-Beneito et al. (2008a) model. Reference denotes the model proposed in Section 3.4.

The greatest difference in terms of computational time is observed between the new proposal and the M-B 2008 model, which runs over three times faster. The simplifications removing the ICAR or the μ_{ts0} terms of the proposal take around 60% of the time the spatio-temporal proposal takes, and the Leroux variation reduces the time of computation in about 15%. All observed computational costs are affordable if the model were to be run once a week in a real surveillance system.

Comparison of estimated parameters and epidemic phases

In order to give an insight into how the different versions alter the detection of the epidemic phase, Figure 3.8 displays the estimated posterior probability of being in the epidemic phase for the retrospective application of the

spatio-temporal proposal, its variations and the M-B 2008 model for four of the locations. The same graph for all the locations is depicted in Section C.3 of the appendices. The estimated mean for the principal parameters of the models are shown in Table 3.3.

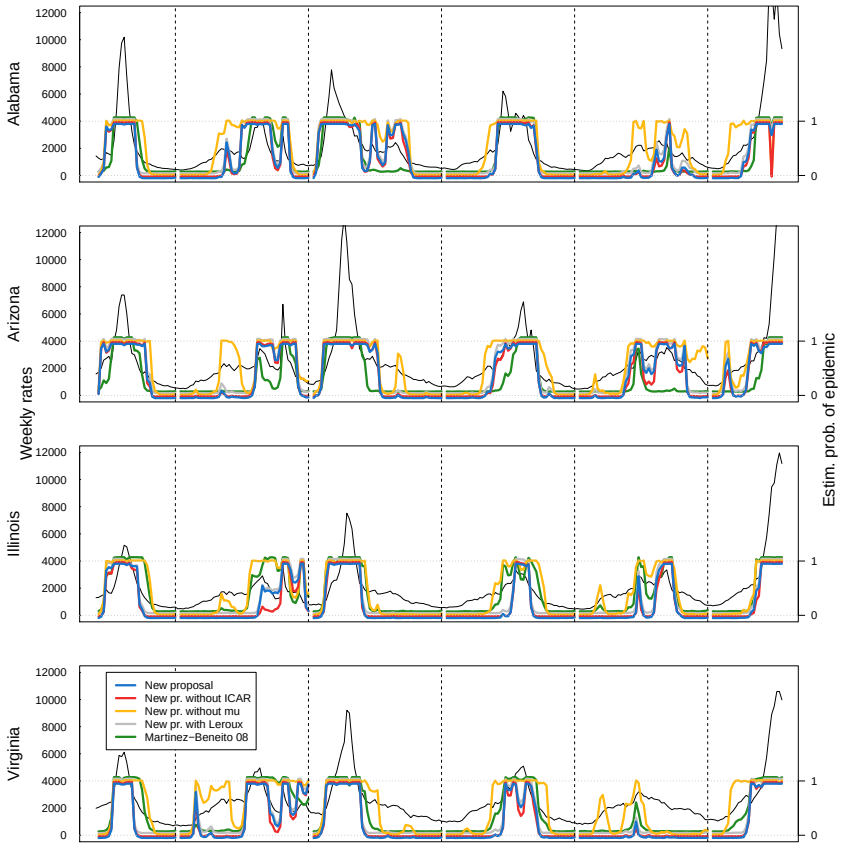

Figure 3.8: Estimated probability of being in epidemic phase by the spatio-temporal model, its variations and M-B 2008 on GFT USA data for 4 states. In black: weekly estimated influenza incidence per 100 000 inhabitants during seasons from 2007–2008 to 2012–2013.

Param.	Reference	without ICAR	without μ_{ts0}	Leroux	M-B 2008
p_{00}	0.95	0.96	0.94	0.95	0.95
p_{11}	0.89	0.88	0.92	0.87	0.89
σ_0	85.00	88.27	92.75	84.45	126.78
σ_1	430.89	712.33	390.94	259.51	935.40
$\sigma_{\mu 0}$	127.61	118.72	–	153.39	–
$\sigma_{\mu 1}$	646.69	958.09	418.72	469.81	–
σ_ψ	770.10	–	582.51	930.34	–
ρ	0.35	0.39	0.36	0.36	0.51
λ	1.50	1.34	–	1.82	–
ϕ	–	–	–	0.59	–

Table 3.3: Comparison of the estimated mean for the parameters of the spatio-temporal model applied on the GFT USA data in a retrospective basis and its variations, and the average value of the mean for the parameters of Martinez-Beneito et al. (2008a) model applied separately on the 49 temporal strains of data of each state. Reference denotes the model proposed in Section 3.4.

Taking a look at the yellow line in Figure 3.8, one can observe how the detection of the simplification without μ_{ts0} is consistently higher than the rest. The higher posterior estimated mean for p_{11}, shown in Table 3.3, also indicates the higher sensibility and lower specificity of this simplification. As it was conjectured, not having a random effect on the endemic mean of all locations depending on t gives advantage to the epidemic structure over the non-epidemic structure to be able to adapt to the data, and this rises the amount of weeks detected as epidemic. In any case, the yellow lines show some wiggly unexpected movements at moments without particular evidence of epidemic phase that advice us to discard this simplification in the original model.

The red line represents the detection when removing the spatial dependence, that is, when not considering an ICAR component for the epidemic phase. The layout of this red line is similar to that of the blue line, which represents the novel spatio-temporal proposal, but with lower values in certain weeks. This lower values tend to be located in weeks next to others

which have been classified as epidemic weeks. It seems that the addition of the spatially structured component helps with the correct classification of these weeks as epidemic ones thanks to the sharing of the information with neighbors. Observing the estimates for the means of the parameters in Table 3.3, one can see how the variability that σ_ψ can not capture (because it is absent) is split among the other two variance terms of the epidemic phase; σ_1 and $\sigma_{\mu 1}$.

The detection performance of the Leroux model is almost the same as that of the original proposal, as it is shown by the almost complete superposition of the blue and gray lines. The estimated mean of the parameters is quite similar with the notorious exception of the variances for the epidemic phase. As can be appreciated in Table 3.3, part of the non-spatial variability modeled in σ_1 in the spatio-temporal proposal moves to σ_ψ in the Leroux model, which is a parameter that captures both spatially structured and unstructured variability. The parameter ϕ indicates that part of the variability expressed by this term is spatially structured and the rest is unstructured. One slight advantage of the Leroux model is that it offers more freedom to decide the variance of the random effects for the mean of the epidemic and non-epidemic regressors. This is so because the unstructured variability is split among σ_1, which directly affects the value of $\sigma_{\mu 1}$, and σ_ψ, which does not.

The model proposed by Martinez-Beneito et al. (2008a), represented in green, offers different periods detected as epidemic than the spatio-temporal proposal or its variations, classifying as in the epidemic phase some weeks that are considered as non-epidemic by the other models and vice versa. This shows the effect of sharing information among various regions, even if the neighboring structure is not taken into account (as is the case of the model without the ICAR component). It is also worth mentioning that M-B 2008 tends to estimate a higher temporal correlation in the autoregressive structure, given that the average mean value of ρ is 0.51, compared with the 0.35 value of the new proposal, as shown in Table 3.3. This is probably because of the absence of other terms, as are μ_{1ts} or the spatial terms, to capture the structure of the epidemic data.

Comparison of the predictive power

When comparing detection methods, a first approach one can consider is assessing the sensitivity, specificity and timeliness of the models and, to do so, a gold standard is needed. When a gold standard is not available, as in our case, other approaches should be taken. One option used in some works like that of Boyle et al. (2011) is the comparison of the predictive power of the models. All models in the comparison have the latent variables Z_{its} in their formulation. The posterior probability for these latent variables in all the models are the estimated probability of being in the epidemic phase. It seems sensible to assume that a model that gives better prediction of the data will also give better estimates of the latent variable, from which the observed values directly depend. Therefore the predictive power assessment can be sensibly used as an indirect measure of the quality of detection.

Approximate cross-validatory predictive assessment, proposed by Marshall and Spiegelhalter (2003), was performed to evaluate the predictive power of the methods. This is a computationally cheaper alternative to the full leave-one-out cross-validatory assessment where the model is run only once instead of once per observation. Prediction for each region i was calculated ignoring the estimate of the parameter ψ_{its} but taking into account the parameters of the neighbors ψ_{jts} with $j \sim i$. This process was done for each week in an online basis, doing prediction using only data from the same or previous weeks. In order to do the approximate cross-validatory predictive assessment, a measure of the discrepancy between predictions and observed values is needed. As all the models in the comparison are defined under the Bayesian paradigm, predictions are expressed as probability distributions and not as punctual estimates. For this reason, in order to evaluate the goodness of the predictions in comparison to the observations, the Continuous Rank Probability Score (CRPS, see Gneiting and Raftery, 2007, for example) was used. CRPS, described in detail in Section A.4 of the appendices, is a measure of discrepancy between a probability distributions and a point. This measure considers not only the posterior expected value of a distribution but also its precision and shape to calculate the distance between the prediction and the observation, giving lower scores for

better predictions. One advantage of using approximate cross-validatory predictive assessment over using DIC (also a model selection tool based in predictive criterions) is that the former is able to evaluate the online performance of the model. DIC, instead, considers all the retrospective estimates, not only the last week and, because of that, it can not be applied to adequately evaluate the application of the models in an online basis.

Figure 3.9 shows the average CRPS for all 49 locations, for each week of season 2012–2013 calculated in the cross-validatory predictive assessment of the new spatio-temporal proposal, its variations and the M-B 2008 model, which requires applying each one of the models in an online basis for all the weeks of this last season. The graph on the bottom of that same figure shows a detail of the first 15 weeks, where the majority of the states are classified as non-epidemic. We can observe that the new proposal, depicted in blue, offers the best (lowest) scores of CRPS both in the first weeks, where the majority of the states are in the endemic phase and in the last weeks, where the majority of locations are in the epidemic phase. The red line, representing the model without spatial structure, shows almost equivalent scores in the first weeks, as should be expected, as for the non-epidemic observations the spatial component is not present. The suppression of the ICAR term in the epidemic linear regressor, though, results in worse scores in the last weeks. The same phenomenon, but much milder, happens with the model that substitutes the ICAR structure for a Leroux structure, represented by the gray line. The quality of the prediction is close to that of the new proposal, but somehow worse for the epidemic weeks. The model represented by the yellow line differs from the original novel proposal only in the non-epidemic regressor, which is modeled without the μ_{ts0} term. For this reason the most visible differences are in the first weeks, as their behavior is not well captured by the simplified model. The M-B 2008 model, shown in green, gives the worst values among all the compared models both in the first and last weeks. This indicates that sharing information among locations is important to correctly model both endemic and epidemic weeks.

To summarize, in this chapter we have presented a review on spatio-temporal methods for the detection of outbreaks, paying special attention to those dedicated to influenza surveillance. This review has set the theo-

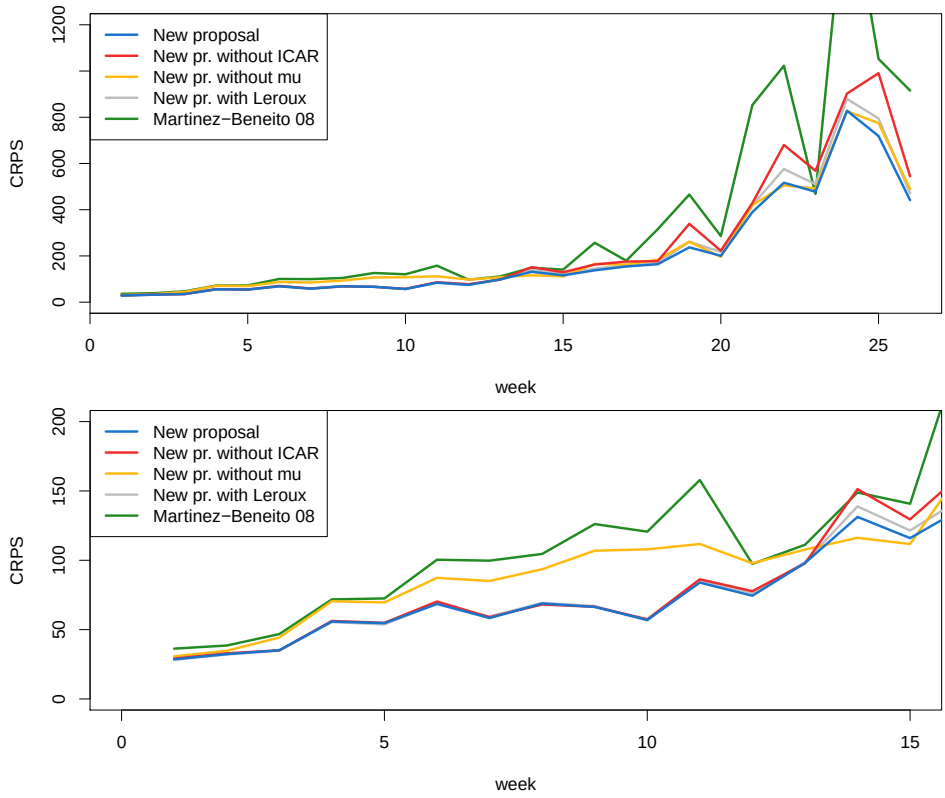

Figure 3.9: Cross-validatory predictive assessment using CRPS as measure of discrepancy for the new proposal, its variations and M-B 2008. Average CRPS for the prediction of the 49 states is shown for all the weeks of the season 2012–2013. Graph in the bottom shows a zoom for the first 15 weeks of the season so the differences may be better appreciated.

retical framework for our proposal, also presented in this chapter, a spatio-temporal Markov switching model on the differentiated rates for the detection of influenza outbreaks. This novel proposal has shown to be able to detect the onset of influenza outbreaks in a set of neighboring regions where incidence data is available along time. It has shown no problem cap-

turing the behavior of the spread of influenza, which can start in several focuses and spread along entire countries. The comparison with several simplifications and variations has shown the pertinence of the components of its hierarchical structure. In the next chapter we present the conclusions to the work in the first three chapters and some possible future lines of research are discussed.

Chapter 4

Conclussions and future lines

4.1 Conclussions

In this thesis we have intended to offer two specific contributions to the field of the statistical methods for the detection of influenza outbreaks by proposing two extensions of the model presented in the work of Martinez-Beneito et al. (2008a). These new proposals are Bayesian hierarchical statistical models with a set of latent variables structured as hidden Markov chains. A review of statistical methods for the detection of influenza outbreaks has justified the necessity to develop these new methodologies capable of avoiding limitations present in previous methods and has informed their creation.

Our proposals consider the behavior of the differentiated rate as an important discriminating feature. By paying attention to the differentiated rates and not only to the raw rate, several mild increases or a sharp one are able to rise the alarm. This happens even when the size of the raw rates are relatively small, as they usually are at the beginning of an epidemic period, allowing our proposals to trigger timely alarms without the need to wait for the rates to rise high. Also, with this approach, an isolated but persistent mild increase of the mean rate (typical of the endemic rates during the cold season of the year) does not trigger an alarm, as methods

based in cumulative charts would.

A core feature of both of our proposals is the use of MSM. The first advantage seen in the use of MSM is the fact that one is not restricted to model the endemic phase and test if data fit the model, and when they do not, declare an alarm. MSM, instead can model both the endemic and epidemic phase, and by doing so, the knowledge about the temporal and spatio-temporal behavior of influenza epidemics is also helpful for determining the epidemic and endemic phases. Models using these hidden variables to discriminate between epidemic and non-epidemic phases do not require of a previous distinction of what is epidemic and what is endemic in the training data, so one of the main problems of methods based on historic limits is avoided. Influenza epidemics spreading along huge territories is not a problem when using this hidden indicators of epidemic, as there is no problem for these variables to take the same epidemic value across all the regions for a given time. Another advantage is the fact that the expected value of these latent variables offer a decision tool that is not dichotomous, but a continuous measure between 0 and 1 for the probability of presence of epidemic.

Once again we point out that using Bayesian inference has allowed us to construct and do inference with relative ease on hierarchical models. They are capable of modeling complex behaviors by combining several components (hidden Markov chains, temporal, spatial and spatio-temporal structures of conditional dependence or any others). This paradigm also offers interpretability of the results as probability distributions, which allows us to speak of the expected probability of being in epidemic phase for each time (and location).

In the temporal proposal, the use of a Poisson distribution to model the counts of incident cases in the first layer of the framework of hierarchical models has allowed the models to capture part of the intrinsic variability of the data that directly modeling the rates does not. An important point for the classification of the epidemic or non-epidemic phase in this framework of models has been the variability, as non-epidemic phases show much less variance. Another key point has been the temporal structure of the data. The presence or absence of the autoregressive terms in the

regressors has highly changed the estimated classifications of the epidemic phase. However, if the autoregressive term had been present, the degree of this temporal structure had affected only in a mild way to the estimation of the epidemic state. The nature of the data and some of the observed results have suggested that it is important to use the autoregressive structure in both regressors. For reasons parsimony, it is also advisable to use models of the framework with the same degree of structure for both regressors. Anyhow, the four models with autoregressive terms in both regressors (AR1-AR1, AR1-AR2, AR2-AR1 and AR2-AR2) offer, in general, similar enough outcomes.

When trying to decide which model better detects influenza epidemics, several proposals have been discussed, as none of them offers a definitive criterion. The first issue to consider has been whether to use the online or the retrospective performance of the models to make this decision. The online basis (the one used in real surveillance systems for outbreak detection) offers realistic estimates, but it is computationally expensive. The estimated probabilities of epidemic have been appreciated to be similar in both fashions, but the online application of the model have declared some outbreaks one or two weeks later than the retrospective application (which is understandable, as no information about the future behavior of data is available).

A second issue has been the selection of a measure of the performance of the detection model. The DIC is a cheap to obtain measure, as it only requires a retrospective application of the model. It measures the goodness-of-fit of the model penalizing its complexity, but does not directly measure what has been tried to be assessed. Weighted ROC measures do evaluate how well a method detects and how fast it does it, but two difficulties have arisen. The first one is to determine the best way to summarize a three dimensional information (sensitivity, specificity and timeliness) in a one dimensional measure. Our results have shown that measures which define sensitivity in terms of number of correctly classified weeks and measures which define sensitivity in terms of correctly detected seasons show discrepancies in the order of the models. We have also noticed that the election of the maximum delay for a successful detection does not influence much

in the outcome of these measures.

The second problem is that weighted ROC measures require a gold standard. We have chosen to construct a gold standard based on the laboratory isolates of the virus, though several other ways are possible. The outcome of the weighted measures has proven to be quite dependent on the way the gold standard was defined. Observed correlation between DIC and weighted ROC measures in one of the applications of the model have been far from perfect, but higher than 0.5 in all cases, which indicates that DIC can be used as a proxy of weighted ROC measures when they are not available.

The proposed temporal models have shown better weighted ROC scores with respect to other methods in the literature. This indicates that the models of our proposal give better and faster detection of influenza outbreaks.

For the evaluation of performance of the spatio-temporal proposal, approximate cross-validatory predictive assessment using CRPS has been used. While doing so, we have assumed that correctly predicting the data indicates that we are doing a good inference on the value of the latent variables which indicate the epidemic state. CRPS has shown to be useful to evaluate the relevance of two terms in the model. One of the terms is the spatial ICAR term in the epidemic regressor, and the other is μ_{ts0}, the random effect on the mean of the non-epidemic differentiated rate. Both of them have proven to be terms of the proposal that notably improve CRPS. This shows that these terms are necessary for the model to correctly estimate the data and, by the assumption, to better determine the epidemic state.

The spatial structure proposed by Leroux et al. (2000) has also been tested as an alternative to the ICAR structure. This structure makes the hierarchical model more flexible which, in theory, can be interesting. Anyhow, the estimated epidemic probabilities have been almost the same and the obtained CRPS values have been slightly worse.

The model of Martinez-Beneito et al. (2008a) offers estimated probabilities of detection that in some regions and times are quite different to those of the novel proposal. The CRPS values are notably worse than those of the new proposal or its variations. This indicates the impact and importance of using a spatio-temporal model, like the one we have proposed, instead of

using a temporal model for each region, an approach that does not allow information to be shared along space.

To sum up, in this research we have shown a review on methods for the detection of influenza outbreaks and have proposed a temporal framework of models and a spatio-temporal model for the detection of influenza outbreaks. They are able to detect outbreaks without the need of previous definition of epidemic and endemic phases on historic data, do not assume temporal location of the epidemics and can model the extensive spatial spread of usual influenza epidemics. The use of hidden Markov chains and temporal and spatio-temporal structures of conditional correlation under the Bayesian paradigm have been critical to build the novel proposals and to easily set a criterion for the triggering of alarms in the form of probability of epidemic. Also, the variability of rates and the behavior of the differentiated rates have shown to be important features of the influenza data which can be used to distinguish between epidemic and non-epidemic phases. We have also seen how the evaluation and comparison of the performance of the models is not trivial, and have discussed several ways to do it.

4.2 Future lines

As a conclusion to this work, we present possible future lines of investigation based on the research presented here. The proposals presented in this work have set the temporal and spatio-temporal structures on the mean behavior of the raw or differentiated rates. Another approach is to model the matrix of transition probabilities of the hidden Markov chain, so the probability of changing to epidemic or endemic phase is not the same for all the weeks. The temporal model presented by Nunes et al. (2013) uses logit regressions with partial data and laboratory isolates to model the transition probabilites. The works of Zou et al. (2012) and Heaton et al. (2012) modifies the non-epidemic to epidemic transition probability according to the amount of neighboring regions in epidemic phase. Further models with spatial and spatio-temporal structures can be proposed to model the transition

matrix.

The edges of the neighborhood structure for the spatio-temporal application on this thesis have been defined as sharing borders and they all have the same weight. Other neighboring graphs can be considered, for example, the movements of people, using main flight connections and other modes of transport as edges of the graph. Following this idea, one may consider a non-dichotomous graph, where weights of the graph edges may take continuous values others than 0 or 1 based in some convenient criterion, as could be the observed movement of people between regions based on data from traffic, flights, etc. This would require to use a more general definition of the ICAR structure, where these weights are considered and, depending on the way the matrix is defined, lack of sparsity problems may arise. A way more challenging idea is to let the influenza data estimate the values of the neighboring matrix, whether they are assumed dichotomous or weights between 0 and 1. But this would require to deeply refurbish the model or use a different model for the estimation of the neighboring structure. A further step could be inspired by the work of Fox and Dunson (2015), who propose a Bayesian nonparametric covariance regression. This model allows the variance matrix in a multivariate regression model to vary with the predictors among time. This idea could be taken as a starting point to develop a method for the estimation of neighboring structures which would stem from the data and would dynamically change among time.

Another possible extension to the models presented in this thesis would be the use of weather covariates like humidity, temperature or solar radiation on each region and time, which may be important factors that affect the behavior of influenza incidence (see, for example, Tang et al., 2010 or Charland et al., 2009).

In our opinion, the proposals in this thesis are effective tools for the detection of influenza outbreaks which offer interesting insights on useful ways to model influenza data for outbreak detection. Thanks to them, new paths to improve detection power are opened on the field of influenza outbreaks detection.

Appendices

Appendix A

Review of methods for the selection of statistical algorithms for the detection of outbreaks

The selection of the best model within a set of models for the detection of epidemic outbreaks is not trivial. Different researches use different ways of comparing and selecting outbreak detection methods and, in many occasions, several tools are used in the same work while the authors stress the issues of each tool. A review of evaluation of detection models can be found in Watkins et al. (2006), which explains the general approaches taken and their issues. In this appendix we describe several tools used for the evaluation and selection of outbreak detection methods which have been used in this thesis: the description of the outcome, the assessment of goodness-of-fit through the Deviance Information Criterion, methods based on specificity, sensitivity and timeliness and the continuous ranked probability score used to assess the predictive power.

A.1 Description of the outcome

The first simplest method consists in graphically describing the outcome and qualitatively evaluating the quality of performance of the models without any numeric measures. One can describe how the alarms coincide with the weeks with highest incidence rates or compare the date of starting and ending of alarms of two different methods. This approach is usually presented together, but not always, with other approaches which do imply numeric measures.

Some works that use this approach as a selection method are those of Kulldorff (2001), which compares a spatio-temporal scan statistic with a purely spatial one, Knorr-Held and Richardson (2003) (actually a method for prediction, not detection), Frisén and Andersson (2009), Boyle et al. (2011), Zou et al. (2014) and Salmon et al. (2015). All of them show an application of the method on real data only.

A.2 Model fit and Deviance Information Criterion

A general approach when doing variable selection or when choosing among nested or similar models is to evaluate the goodness-of-fit of the model, for example, by means of R^2. Using measures which involve goodness-of-fit have several drawbacks which should be considered. Goodness-of-fit on nested models is always better for the more complex ones, so some way of penalizing the complexity is usually taken into consideration. Several ways of penalizing have been proposed, like the Akaike information criterion (AIC, see Akaike, 1974), the Bayesian information criterion (BIC, see Schwarz, 1978) or the Deviance Information Criterion (DIC, see Spiegelhalter et al., 2002), this last one commonly used for Bayesian hierarchical models. There is no consensus on the minimum distance in these measures which indicate that a model is qualitatively better than another. A limitation of these measures is that they do not pay attention to specific aspects of the performance of the models that may be of particularly interesting in

some contexts, as is the detection power in our case.

Sun and Cai (2009) use BIC to select between the different options that they offer. Rafei et al. (2012, 2015) also use BIC plus R^2. DIC is used to select between different proposals in the Bayesian works of Knorr-Held and Richardson (2003) and Li et al. (2012). DIC has also been used in this thesis to help choosing between the models of the proposed temporal framework and, for this reason, we present this selection method in more detail.

DIC is based on the deviance, a goodness-of-fit statistic defined as follows:

$$D(\theta) = -2\log\left(p(Y|\theta)\right) + 2\log\left(f(Y)\right) , \tag{A.1}$$

with $p(Y|\theta)$ the likelihood of the model and $f(Y)$ a fully specified standardizing term that is a function of the data alone. The $2\log\left(f(Y)\right)$ term is actually a constant that cancels out in all calculations that compare different models, so is irrelevant for comparison. In fact, several other works like Gelman et al. (2013) refer to the deviance just as

$$D(\theta) = -2\log\left(p(Y|\theta)\right) , \tag{A.2}$$

and this form is often used for comparison of models. The deviance is lower when the log-likelihood is higher and therefore when the goodness-of-fit is higher.

Another statistic that arises from the deviance is the number of effective parameters, which is defined as the mean of the deviance minus the deviance at the posterior mean of each parameter in the model:

$$p_D = \overline{D(\theta)} - D(\bar{\theta}) . \tag{A.3}$$

As stated in Gelman et al. (2013):

> p_D can be thought of as the number of 'unconstrained' parameters in the model, where a parameter counts as: 1 if it is estimated with no constraints or prior information; 0 if it is fully constrained or if all the information about the parameter comes from the prior distribution; or an intermediate value if both the data and prior distributions are informative.

A higher p_D is associated with more complex models and can indicate overparametrization. Both quantities $\overline{D(\theta)}$ and $D(\bar{\theta})$ are easily computed from MCMC simulations provided that a closed form for $D(\theta)$ is available. $\overline{D(\theta)}$ is estimated as the mean of the simulations of $D(\theta)$, and $D(\bar{\theta})$ can be calculated by estimating $\bar{\theta}$ as the mean of the iterations of the posterior distribution $p(\theta|Y)$ and plugging it into the $D(\theta)$ function.

DIC is defined in Spiegelhalter et al. (2002) as *a classical estimate of fit, plus twice the effective number of parameters.* Equivalently, given the definition of p_D in Expression (A.3), it is also the expectation of the deviance plus the effective number of parameters, which shows that this measure rewards goodness-of-fit while penalizing complexity of the models:

$$\begin{aligned} DIC &= D(\bar{\theta}) + 2p_D \\ &= \overline{D(\theta)} + p_D \,. \end{aligned} \tag{A.4}$$

Another equivalent expression that facilitates calculation is:

$$DIC = 2\overline{D(\theta)} - D(\bar{\theta}) \,. \tag{A.5}$$

Regarding the minimum distance in DIC that shows an important difference between models, Spiegelhalter et al. (2002) point out that:

> Burnham and Anderson (1998) suggested models receiving AIC within 1–2 of the 'best' deserve consideration, and 3–7 have considerably less support: these rules of thumb appear to work reasonably well for DIC. Certainly we would like to ensure that differences are not due to Monte Carlo error: although this is straightforward for $\overline{D(\theta)}$, Zhu and Carlin (2000) have explored the difficulty of assessing the Monte Carlo error on DIC.

A.3 Specificity, sensitivity and timeliness based methods

Outbreak detection is a classification problem, where one has to classify each time (and location, for spatio-temporal applications) unit as in epidemic or endemic phase. Two important measures for assessing the quality

of classifications are sensitivity and specificity. Sensitivity is the proportion of outbreak that is actually detected by the method, specificity is the proportion of non-epidemic weeks (and regions) where no false alarm is triggered. Of particular importance for our problem is the capability of the methods to correctly classify the first weeks or days of epidemic, so there is also a third target that one tries to optimize, timeliness. Timeliness concerns the time it takes for a method to flag an alarm from the beginning of an outbreak. The computation of this three measures always needs information about the 'true' classification of each unit. For real data, a gold standard is needed, that is, an external method that determines which weeks (and locations) truly are in epidemic state and which are not. The 'true' epidemic state in simulated data is determined by the way the data is simulated.

A.3.1 Epidemic state determination

There are several ways of determining which weeks (and locations) of a set of data are in epidemic phase, but all of these methods have some drawbacks. Now we discuss the most commonly used:

Gold standard from laboratory confirmations. Detection methods are usually applied on data from fast sources like sentinel networks, admissions to health care or internet based data. When laboratory isolation data are available for the same population at the same time of study, this slower but more reliable data source can be used as a gold standard. One problem is to determine which should be the threshold on the counts of laboratory confirmations for the definition of the gold standard. Cowling et al. (2006), for example, use an arbitrary 30% of maximum laboratory confirmations per week for each season, checking also 20% and 40% thresholds to test the sensitivity of the gold standard to this percentage. Martinez-Beneito et al. (2008a) opt for a presence/absence approach, declaring as gold standard epidemic phase the period between the first and the last laboratory confirmation for each season. This approach has also been used in the present work for the temporal proposal.

Gold standard from other detection method. Some authors choose to use classical methods in the literature as the providers of a gold standard to calculate sensitivity, specificity and timeliness. For example, Rath et al. (2003) and Rafei et al. (2012) use Serfling (1963) to create this gold standard. The obvious drawback is that one shall use a method to set a gold standard that does a better detection than the one being proposed. That would only make sense if the proposed method had other advantages, like the time of computation or the ease of use.

Gold standard from the proposed model with complete data. When applied in a realistic way, a detection method is run every week (or day) with data from only previous weeks available; that is what we call 'online' detection. Some models, like those based in the estimation of a hidden variable Z_{it} can give better estimations of the epidemic state in time t if data from $t+1, t+2, \ldots$ are available. With this in mind, the works of Nunes et al. (2013), Heaton et al. (2012) or Zou et al. (2012) calculate sensitivity, specificity and timeliness of the online application of the model using the estimation of the epidemic states with all the available data as a gold standard. In any case, comparing the performance of one model against itself just guarantees an intern coherence, but is a dubious way to do comparison with other methods.

Simulated data. A way to ensure a known classification of the data among epidemic and endemic phase is to simulate the epidemic data. In this fashion, one can take known endemic cases (though the decision of what is endemic is an outbreak detection problem itself) and inject artificial extra cases during one or several weeks, as authors like Lu et al. (2010) do. The other option is to create all the data from scratch, simulating both endemic and epidemic cases, as seen in Frisén and Andersson (2009), Lu et al. (2010), Salmon et al. (2015) or Rao and McCabe (2016). An issue would be to ensure that the way of simulating data is realistic, even more with diseases with complex

spatio-temporal behavior, as influenza is. Some authors, like Heaton et al. (2012) or Zou et al. (2012) choose to use the detection model itself, with some fixed parameters, to simulate the data, which allows only for intern validation of the model.

A.3.2 Combined measures of specificity, sensitivity and timeliness

Once a gold standard for the epidemics has been set and the model has been run, one can calculate several measures about the quality of the detection of the model. Specificity and sensitivity measures are broadly used in many works like Li et al. (2012) and Rafei et al. (2012). But besides the usual measures of the classification problem it is also common to use other measures that take into account the temporal nature of the data by involving time until the trigger of an alarm, for example:

Average run length. Is the expected time until the first false alarm is triggered when the model is run on endemic data.

Timeliness. Also known as conditional expected delay, is the expected time since the start of an outbreak until an alarm is triggered.

Probability of successful detection. Is the probability of an alarm to be triggered before a certain maximum delay.

This kind of measures are seen, among others, in the works of Frisén and Andersson (2009), Heaton et al. (2012), Salmon et al. (2015) and Rao and McCabe (2016).

When a detection method depends on a threshold, the measures of sensitivity, specificity and timeliness also depend on it. In order to offer measures that are not dependent on the threshold, one can build a receiver operating characteristic (ROC) curve. This graph plots the sensitivity against 1-specificity depending on the threshold (see Lusted, 1971, Egan, 1975 or Metz, 1978 for some classic references). The simple observation of the curves or the area under the ROC curves (AUROC) can be used as decision tools, as done by Rath et al. (2003) or Rafei et al. (2015).

Similar curves can be built combining other measures, like timeliness. In this fashion, Lu et al. (2010), Jiang and Cooper (2010) and Zou et al. (2012) combine the specificity and timeliness in ROC like curves. Timeliness can also be added to ROC curves as a weighting factor or as a third dimension, obtaining measures of the area under the weighted ROC curve (AUWROC) or volumes under the timeliness-ROC surface (VUTROS or VUTROCS).

Now we are going to discuss in more detail the proposals that we have used in the present work as decision tools for the temporal model. The first two measures we are going to discuss are proposed in Kleinman and Abrams (2006), while the last one is proposed in Cowling et al. (2006). For all of them there must be defined a maximum delay (MD) so that an alarm is considered of no practical use if it is triggered after that maximum delay from the beginning of the outbreak. The mean time saved $\overline{TS}_k$ is also needed, where the time saved for season s with a certain threshold labeled as k is defined as:

$$TS_{ks} = \max\left(1 - \frac{delay\,of\,the\,alarm_{ks}}{MD+1}, 0\right), \tag{A.6}$$

which is a measure with value 1 when there is no delay in the alarm and value 0 when the delay is greater than MD and indicates the mean proportion of time the alarm is triggered before the maximum delay. The three measures used in Chapter 2 which combine sensitivity, specificity and timeliness are defined as follow:

AUWROC1. The simplest one; is the area under a weighted ROC curve which is constructed using as y axis the sensitivity weighted by $\overline{TS}_k$ instead of just the sensitivity. As $\overline{TS}_k \in [0,1]$, this measure is always lower than AUROC. Figure A.1 top-right graph illustrates AUWROC1. It can be compared to AUROC depicted in top-left graph.

VUTROS1. In this case, $\overline{TS}_k$ is set as a third dimension z axis, and the points of the 3D curve that is formed are joined to the point $(0,1,1)$ forming a surface. The volume under this surface is our measure. If a

discrete set of thresholds $k \in 1, \cdots, K$ is considered, the approximate measure can be calculated as:

$$VUTROS1 = \sum_{k=1}^{K-1} \frac{1}{6} \left(\overline{TS}_k + \overline{TS}_{k+1} + 1 \right) \left| se_k sp_{k+1} - se_{k+1} sp_k \right|, \tag{A.7}$$

with se_k and sp_k the sensibility and specificity with threshold k. Figure A.1 bottom-left graph illustrates VUTROS1.

VUTROCS. This measure consists of calculating several $(MD+1)$ ROC curves and average the area under them. Each ROC curve has a different maximum delay md going from 0 to MD. The sensitivity in these curves is not measured per weeks, as done in the previous measures, but as the proportion of seasons where there is an alarm with a delay less or equal to md. VUTROCS is the mean of the areas under these ROC curves. Figure A.1 bottom-right graph illustrates VUTROCS.

This measure is almost equivalent to VUTROS3 in Kleinman and Abrams (2006) but easier to compute and more coherent. VUTROS3 calculates the volume under the surface that results of joining the ROC curves in the x, z planes and $\frac{md}{MD}$ in the y axis. Calculating each section of the volume between ROC curves in this way is equivalent to do the mean of the two adjacent areas under the curves, multiply by $\frac{1}{MD}$ each and add them all. In this fashion, when $MD > 1$, all the curves are used twice for the calculation except for those corresponding with delay 0 and maximum delay, which are used only once, so these two cases influence less to the final measure.

Both AUWROC1 and VUTROS1 are built using the same three components (weekly sensitivity, specificity and timeliness) but combining them in different ways. Because of that, they give similar though not equal results when comparing several models. Instead of weekly sensitivity, VUTROCS uses seasonal sensitivity and, because of that, the ordering of compared

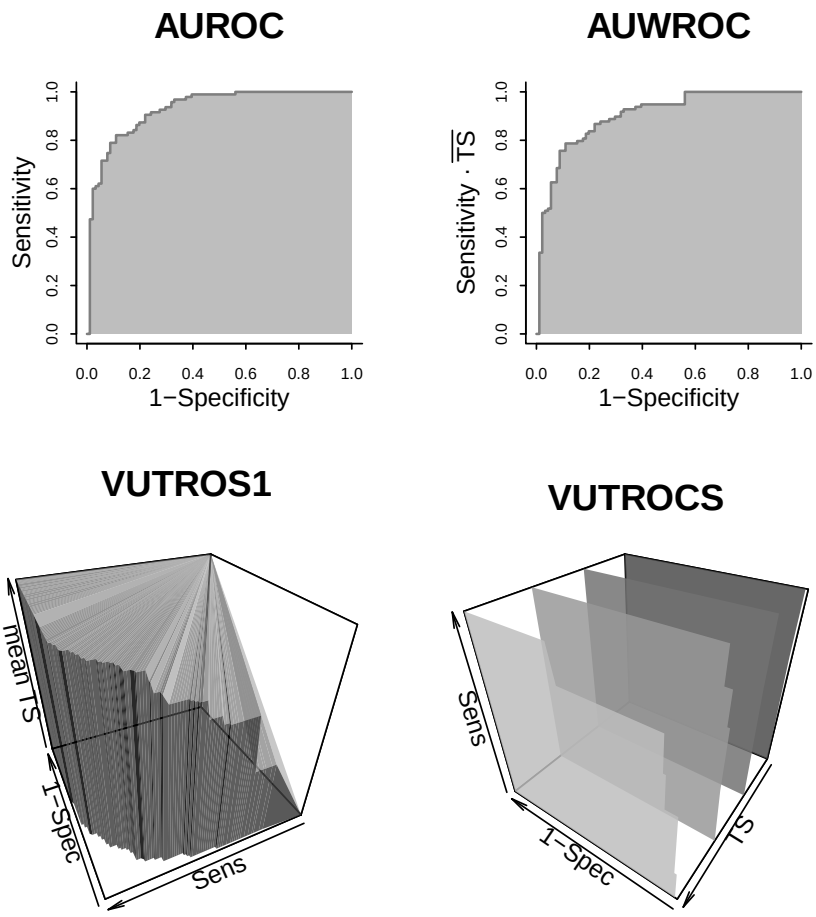

Figure A.1: Illustration of measures combining ROC curves and timeliness. Top-left: area under classic ROC curve. Top-right: AUWROC1 (slightly lower than AUROC). Bottom-left: VUTROS1. Bottom-right: Weighted ROC curves for different maximum delays; the average of the areas is VUTROCS.

models presents substantial differences with those from the two previous measures. For all of the measures, the absolute value has no direct interpretation, and only the relative value when comparing models has a meaning.

A.4 Continuous ranked probability score to assess predictive power

As seen in the previous section, the definition of a gold standard is not trivial. Also, sometimes its quality is doubtful and, without it, measures of sensitivity, specificity and timeliness can not be computed. A different approach is to evaluate the predictive power of the detection method, as seen in some works like that of Boyle et al. (2011). It is true that the predictive power is not a direct measure of the quality of detection, but if the causal relations of the model are sensible, it is expected that a good prediction also indicates a good estimation of the components of the model. Therefore, in models based in hidden variables to distinguish between epidemic and non-epidemic phases, the assessment of the predictive power indirectly evaluates the quality of estimation of those hidden variables. In doing so, the detection power is indirectly evaluated.

The predictive power of a model is usually measured by correlation or distance between the prediction and the data, but when the response of the model is not a punctual estimation but a probability distribution -as happens in Bayesian models- other measures can be considered. In this way not only the location but also the precision of the prediction are taken into account. The continuous ranked probability score (CRPS; see, for example, Gneiting and Raftery (2007)) is a measure that evaluates how close a continuous probability distribution is to a value. In the present work we have used this measure to evaluate our spatio-temporal proposal. We explain it with more detail here.

Given a cumulative distribution function $F(\cdot)$ (in our case, the cumulative predictive distribution) and a point Y (the data point which is being

predicted), CRPS is defined as follows:

$$CRPS(F,Y) = \int_{-\infty}^{\infty} \left(F(x) - \mathrm{I}(x \geq Y)\right)^2 dx\,. \tag{A.8}$$

A lower score of CRPS indicates that the predictive distribution is closer to the real value. Hersbach (2000) explains how CRPS can be estimated when the available data about the probability distribution are in the form of samples (as is the case when using MCMC for the inference). If the samples of the predictive distribution are $X_i, i \in 1,\ldots,N$ and taking into account that $\frac{i}{N}$ is the estimated cumulative probability distribution between X_i and X_{i+1}), then:

$$\begin{aligned} CRPS(F,Y) &= \sum_{i=1}^{N} \int_{X_i}^{X_{i+1}} \left(\frac{i}{N} - \mathrm{I}(x \geq Y)\right)^2 dx \\ &= \alpha_i \left(\frac{i}{N}\right)^2 + \beta_i \left(1 - \frac{i}{N}\right)^2, \end{aligned} \tag{A.9}$$

with α_i and β_i defined as follows:

if	α_i	β_i
$Y > X_{i+1}$	$X_{i+1} - X_i$	0
$X_{i+1} > Y > X_i$	$Y - X_i$	$X_{i+1} - Y$
$X_i > Y$	0	$X_{i+1} - X_i$.

(A.10)

A graphic illustration of this calculation is depicted in Figure A.2 for a sample of size 6 of a predictive distribution. In Figure A.3, another illustration with a more realistic sample of size 1000 is depicted.

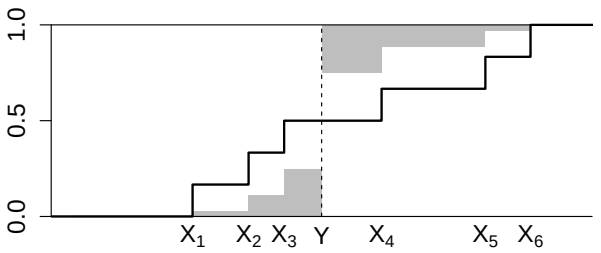

Figure A.2: Illustration of CRPS (shaded in gray) from a sample of size 6. $X_1, \ldots, X_6$ represent the sampled values. Estimated cumulative distribution function in thick line. Y represents the real datum being predicted.

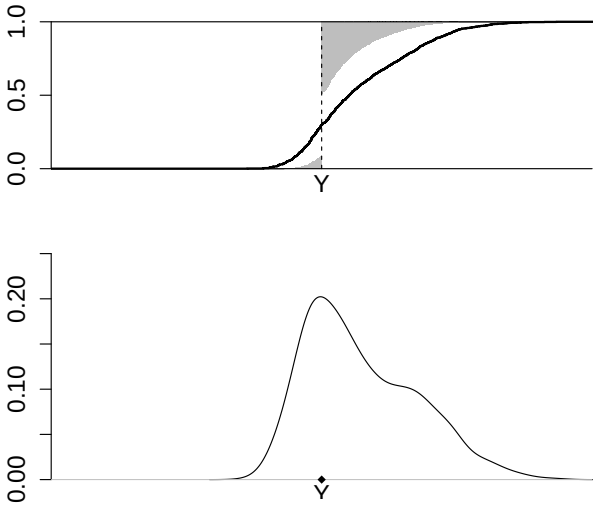

Figure A.3: Illustration of CRPS (shaded in gray, graph above) for a sample of size 1000. Estimated cumulative distribution function in thick line. Y represents the real datum being predicted. Corresponding predictive probability density function in the graph below.

Appendix B

Discussion about Banks et al.'s (2012) model

This appendix presents a discussion about the general structure of models for the spatio-temporal surveillance of diseases presented by Banks et al. (2012), in which are based the models of Heaton et al. (2012) and Zou et al. (2012). The first layer of the general model is as follows:

$$Y_{it} \sim \text{Po}(\nu_{it} + Z_{it}\lambda_{it}) \,, \tag{B.1}$$

where ν_{it} and λ_{it} represent the endemic and epidemic expected cases and Z_{it} is a latent dichotomous variable which takes value 1 when and where the system is in the epidemic state and 0 otherwise. The logarithms of ν_{it} and λ_{it} are modeled by a linear regression with covariates and a spatio-temporal term each:

$$\begin{aligned} \log(\nu_{it}) &= \boldsymbol{\beta}_{\boldsymbol{\nu}}^{\prime}\boldsymbol{X}_{\boldsymbol{it}} + \xi_{\nu it} \,, \\ \log(\lambda_{it}) &= \boldsymbol{\beta}_{\boldsymbol{\lambda}}^{\prime}\boldsymbol{U}_{\boldsymbol{it}} + \xi_{\lambda it} \,, \end{aligned} \tag{B.2}$$

where $\boldsymbol{X}_{\boldsymbol{it}}$ and $\boldsymbol{U}_{\boldsymbol{it}}$ are vectors of covariates for each of the regressors and $\xi_{\nu it}$ and $\xi_{\lambda it}$ are some spatio-temporal terms.

In this appendix we prove that a simplification of this model with only an intercept term in the regressors and with improper flat priors for these

intercept parameters has an improper posterior distribution. This suggests that a more complex structure as that of Banks et al. (2012) might also have problems with the inference process as long as an intercept term is included in the linear term, which is an extremely common setting. We also prove the propriety of the posterior distribution for an alternative model, the 'switch' model, where the parameter of the Poisson distribution is defined as $(1 - Z_{it})\nu_{it} + Z_{it}\lambda_{it}$.

B.1 Impropriety of the posterior distribution for the simplification of Banks et al., 2012

Let us consider the model above with only the intercept term for the regressors and no spatio-temporal interactions, which can be formulated as follows:

$$\begin{aligned} \log(\nu_{it}) &= \beta_\nu\,, & \pi(\beta_\nu) &\propto 1\,, \\ \log(\lambda_{it}) &= \beta_\lambda\,, & \pi(\beta_\lambda) &\propto 1\,. \end{aligned} \tag{B.3}$$

Therefore, we have that:

$$\nu_{it} = e^{\beta_\nu}\,, \qquad \lambda_{it} = e^{\beta_\lambda}\,, \tag{B.4}$$

which do not depend on i or t. Therefore, we can lose the subindices in the notation for ν and λ. Because of the Change of Variable theorem, the prior distributions of ν and λ are as follow:

$$\pi(\nu) \propto \frac{1}{\nu}\,, \qquad \pi(\lambda) \propto \frac{1}{\lambda}\,. \tag{B.5}$$

We are going to demonstrate that the posterior distribution is improper.

For simplicity in the notation, let us denote Ω to the set of pairs of indices (i, t) for all times and locations considered and A to the subset of Ω where the latent variable is in the epidemic phase ($Z_{it} = 1$):

$$\begin{aligned} \Omega &:= \{(i,t) : i \in \{1,\dots,I\}, t \in \{1,\dots,T\}\}\,, \\ A &:= \{(i,t) \in \Omega : Z_{it} = 1\}\,. \end{aligned} \tag{B.6}$$

We also notate n and n_A to the number of elements in Ω and A respectively.

A lower bound for the likelihood of the Poisson model can be expressed as follows:

$$\begin{aligned} l(\nu,\lambda) &\propto \prod_{\Omega} e^{-(\nu+Z_{it}\lambda)}\ (\nu+Z_{it}\lambda)^{Y_{it}} \\ &= e^{-\sum_{\Omega}\nu}\, e^{-\sum_{\Omega} Z_{it}\lambda} \prod_{\Omega} (\nu+Z_{it}\lambda)^{Y_{it}} \\ &= e^{-\sum_{\Omega}\nu}\, e^{-\sum_{A}\lambda} \prod_{\Omega} (\nu+Z_{it}\lambda)^{Y_{it}} \\ &\geq e^{-\sum_{\Omega}\nu}\, e^{-\sum_{A}\lambda} \prod_{\Omega} \nu^{Y_{it}} \\ &= e^{-n\nu}\, e^{-n_A\lambda}\, \nu^{\sum_{\Omega} Y_{it}}\,. \end{aligned} \tag{B.7}$$

Then, a lower bound for the joint posterior distribution of (ν,λ) is expressed as follows:

$$\begin{aligned} p(\nu,\lambda|\boldsymbol{Y}) &= \pi(\nu)\pi(\lambda)l(\nu,\lambda) \\ &\geq \nu^{-1}\,\lambda^{-1}\, e^{-n\nu}\, e^{-n_A\lambda}\, \nu^{\sum_{\Omega} Y_{it}} \\ &= e^{-n\nu}\, e^{-n_A\lambda}\, \nu^{(\sum_{\Omega} Y_{it})-1}\,\lambda^{-1}\,. \end{aligned} \tag{B.8}$$

Now we integrate λ to obtain a lower bound for the marginal posterior distribution of ν:

$$\begin{aligned} p(\nu|\boldsymbol{Y}) &= \int_0^{\infty} p(\nu,\lambda|\boldsymbol{Y})d\lambda \\ &\geq \int_0^{\infty} e^{-n\nu}\, e^{-n_A\lambda}\, \nu^{(\sum_{\Omega} Y_{it})-1}\,\lambda^{-1} d\lambda \\ &= e^{-n\nu}\, \nu^{(\sum_{\Omega} Y_{it})-1} \int_0^{\infty} e^{-n_A\lambda}\,\lambda^{-1} d\lambda\,. \end{aligned} \tag{B.9}$$

And $\int_0^{\infty} e^{-n_A\lambda}\,\lambda^{-1}d\lambda$ integrates infinity, as is proportional to the improper distribution $Ga(0,n_A)$. Therefore, the posterior distribution $p(\nu,\lambda|\boldsymbol{Y})$ is improper.

B.2 Propriety of the posterior distribution of the switch model

To avoid the impropriety of the posterior distribution of the parameters, an alternative modeling can be considered. In this alternative modeling, the epidemic regressor is not added to the non-epidemic regressor when there is an epidemic. Instead of that, the model switches between two different regressors with different parameters. That is done, for example in the work of Li et al. (2012) (though it is not used for detection of outbreaks, but for detection of unusual behavior of particular locations). This 'switching' Poisson distribution on the first layer of the hierarchical modeling has also been used in the temporal proposal of Chapter 2, also published in Conesa et al. (2015).

We add $\bar{A}$ to the notation used in the previous section, denoting the subset of Ω where the latent variable is in the non-epidemic phase ($Z_{it} = 0$):

$$\bar{A} := \{(i,t) \in \Omega : Z_{it} = 0\}\,, \tag{B.10}$$

and $n_{\bar{A}}$ as the number of elements in $\bar{A}$.

The first layer of the model, then, is

$$Y_{it} \sim \mathrm{Po}\left((1 - Z_{it})\nu_{it} + Z_{it}\lambda_{it}\right)\,. \tag{B.11}$$

Let us assume that the epidemic period A and the non-epidemic period $\bar{A}$ are not void and there is some observed case for each of them. Let us prove then that the posterior distribution of the parameters in this model is proper for the same improper prior distributions that those proposed in the previous section for the simplification of the model of Banks et al. (2012). Considering the same definition of the regressors as in Expression

(B.3) with only the interaction terms, the likelihood is expressed as follows:

$$\begin{aligned} l(\nu,\lambda) &\propto \prod_{\Omega} e^{-(1-Z_{it})\nu}\, e^{-Z_{it}\lambda}\, ((1-Z_{it})\nu + Z_{it}\lambda)^{Y_{it}} \\ &= e^{-\sum_{\Omega}(1-Z_{it})\nu}\, e^{-\sum_{\Omega} Z_{it}\lambda} \prod_{\Omega} ((1-Z_{it})\nu + Z_{it}\lambda)^{Y_{it}} \\ &= e^{-n_{\bar{A}}\nu}\, e^{-n_A\lambda} \prod_{\bar{A}} \nu^{Y_{it}} \prod_{A} \lambda^{Y_{it}} \\ &= e^{-n_{\bar{A}}\nu}\, e^{-n_A\lambda}\, \nu^{\sum_{\bar{A}} Y_{it}}\, \lambda^{\sum_A Y_{it}}\,, \end{aligned} \tag{B.12}$$

and the posterior distribution of (ν,λ) is:

$$\begin{aligned} p(\nu,\lambda|\boldsymbol{Y}) &= \pi(\nu)\pi(\lambda) l(\nu,\lambda) \\ &\propto \nu^{-1}\,\lambda^{-1}\, e^{-n_{\bar{A}}\nu}\, e^{-n_A\lambda}\, \nu^{\sum_{\bar{A}} Y_{it}}\, \lambda^{\sum_A Y_{it}} \\ &= e^{-n_{\bar{A}}\nu}\, e^{-n_A\lambda}\, \nu^{(\sum_{\bar{A}} Y_{it})-1}\, \lambda^{(\sum_A Y_{it})-1}\,. \end{aligned} \tag{B.13}$$

This distribution is proper, as long as the epidemic A and non-epidemic periods $\bar{A}$ are not void and with some observed case for each of them, as it is the product of the following two conditionally independent Gamma distributions:

$$\begin{aligned} p(\nu|\boldsymbol{Y}) &= Ga(\textstyle\sum_{\bar{A}} Y_{it}, n_{\bar{A}})\,, \\ p(\lambda|\boldsymbol{Y}) &= Ga(\textstyle\sum_{A} Y_{it}, n_{A})\,. \end{aligned} \tag{B.14}$$

Appendix C

Additional Figures

In this appendix we present several figures which show some results of Chapter 3 for all 49 states of the USA and which are not presented along the text for the sake of simplicity. In the text of that chapter, figures for only four randomly chosen states have been shown to facilitate readability.

C.1 Estimates of the probability of epidemic in all 49 states by spatio-temporal model

In this section we present the estimated posterior probability of being in the epidemic phase for all 49 states of the USA by our spatio-temporal proposal applied on Google Flu Trends (GFT) USA data.

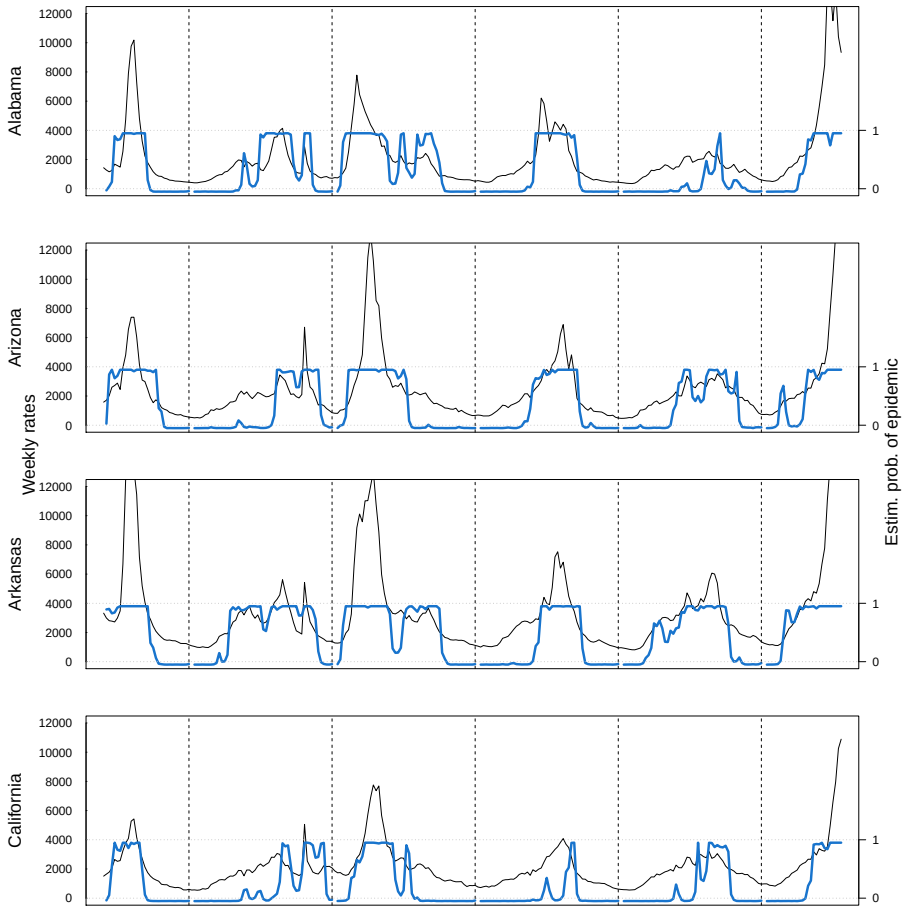

Figure C.1: Estimated probability of being in epidemic phase by the spatio-temporal model on GFT USA data for Alabama, Arizona, Arkansas and California. In black: weekly estimated influenza incidence per 100 000 inhabitants during seasons from 2007–2008 to 2012–2013.

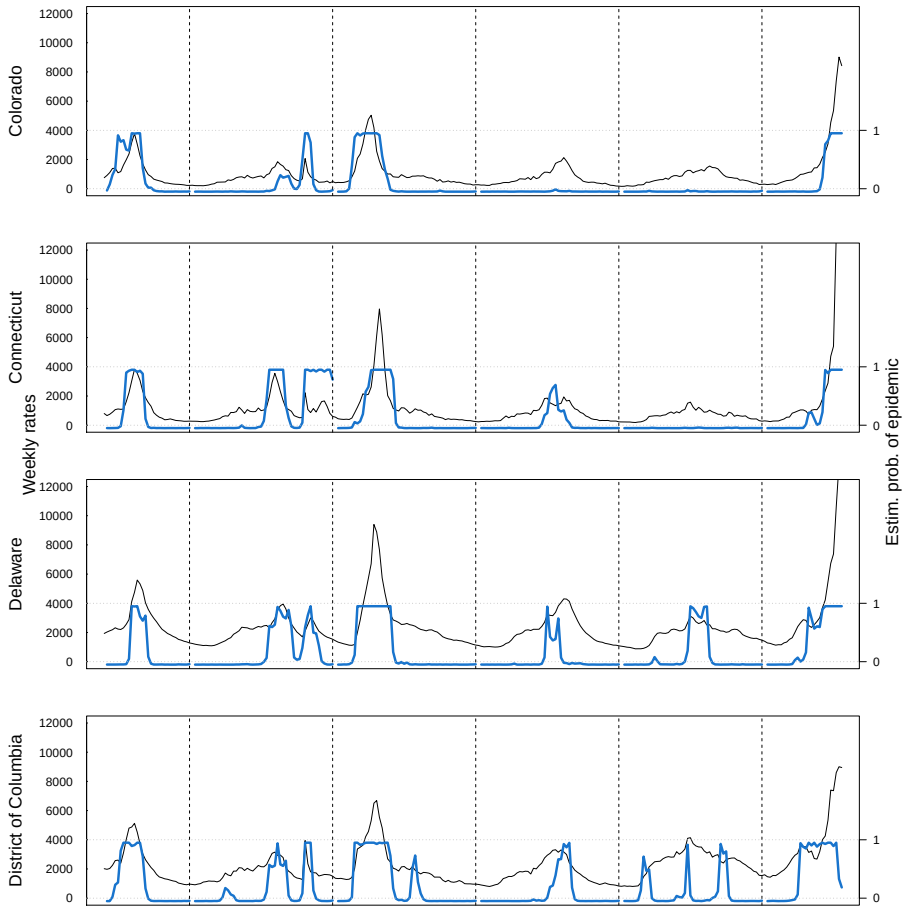

Figure C.2: Estimated probability of being in epidemic phase by the spatio-temporal model on GFT USA data for Colorado, Connecticut, Delaware and District of Columbia. In black: weekly estimated influenza incidence per 100 000 inhabitants during seasons from 2007–2008 to 2012–2013.

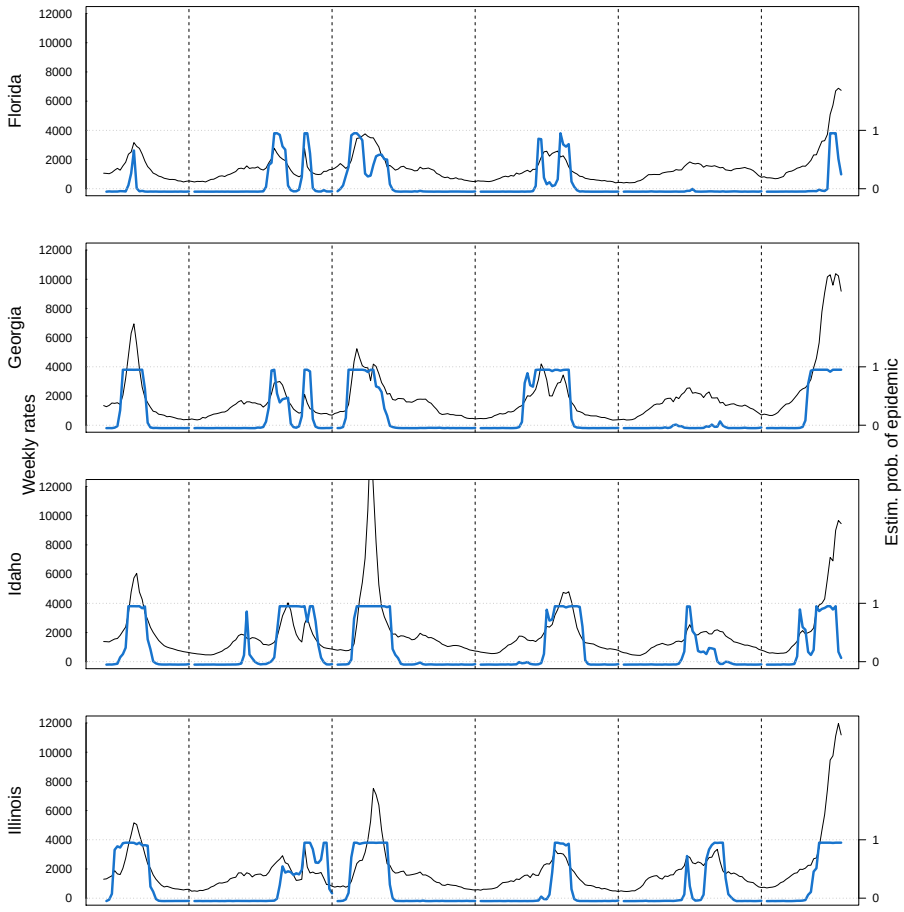

Figure C.3: Estimated probability of being in epidemic phase by the spatio-temporal model on GFT USA data for Florida, Georgia, Idaho and Illinois. In black: weekly estimated influenza incidence per 100 000 inhabitants during seasons from 2007–2008 to 2012–2013.

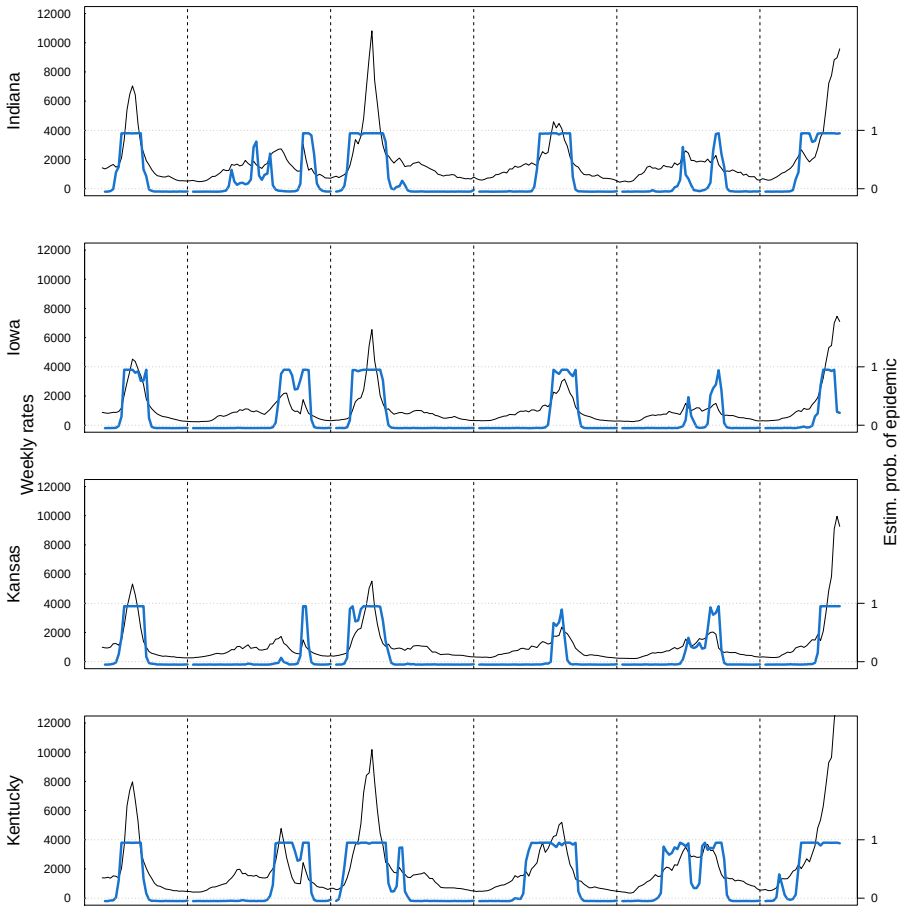

Figure C.4: Estimated probability of being in epidemic phase by the spatio-temporal model on GFT USA data for Indiana, Iowa, Kansas and Kentucky. In black: weekly estimated influenza incidence per 100 000 inhabitants during seasons from 2007–2008 to 2012–2013 of GFT USA.

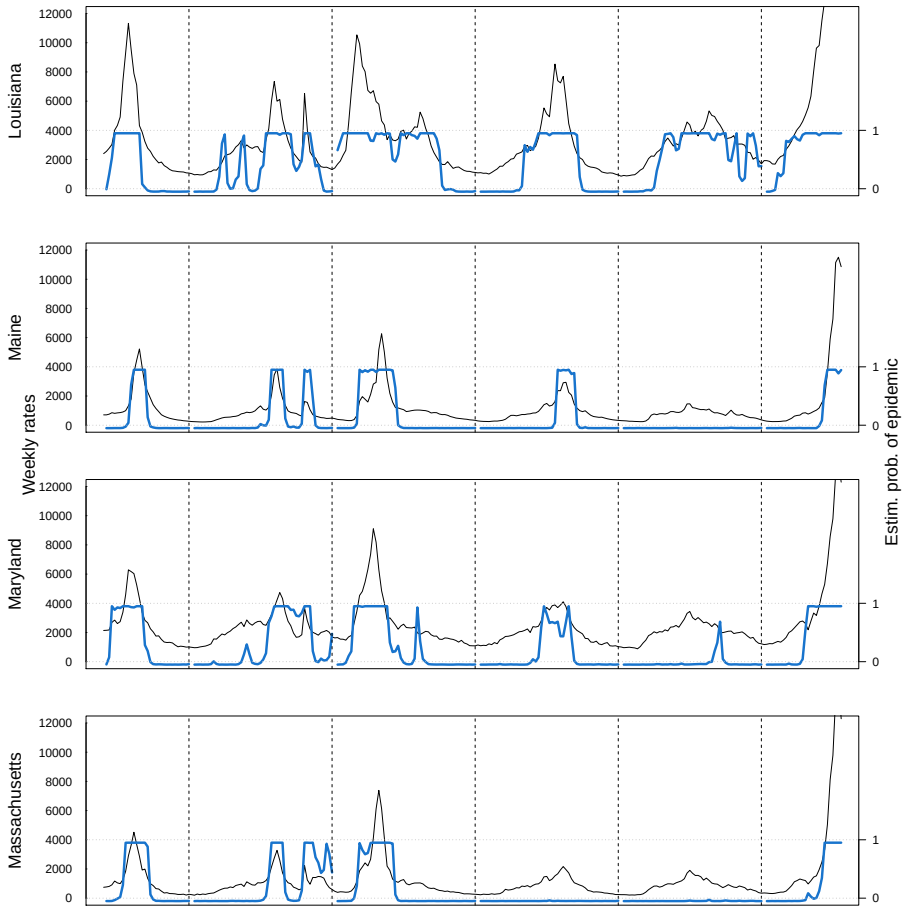

Figure C.5: Estimated probability of being in epidemic phase by the spatio-temporal model on GFT USA data for Louisiana, Maine, Maryland and Massachusetts. In black: weekly estimated influenza incidence per 100 000 inhabitants during seasons from 2007–2008 to 2012–2013.

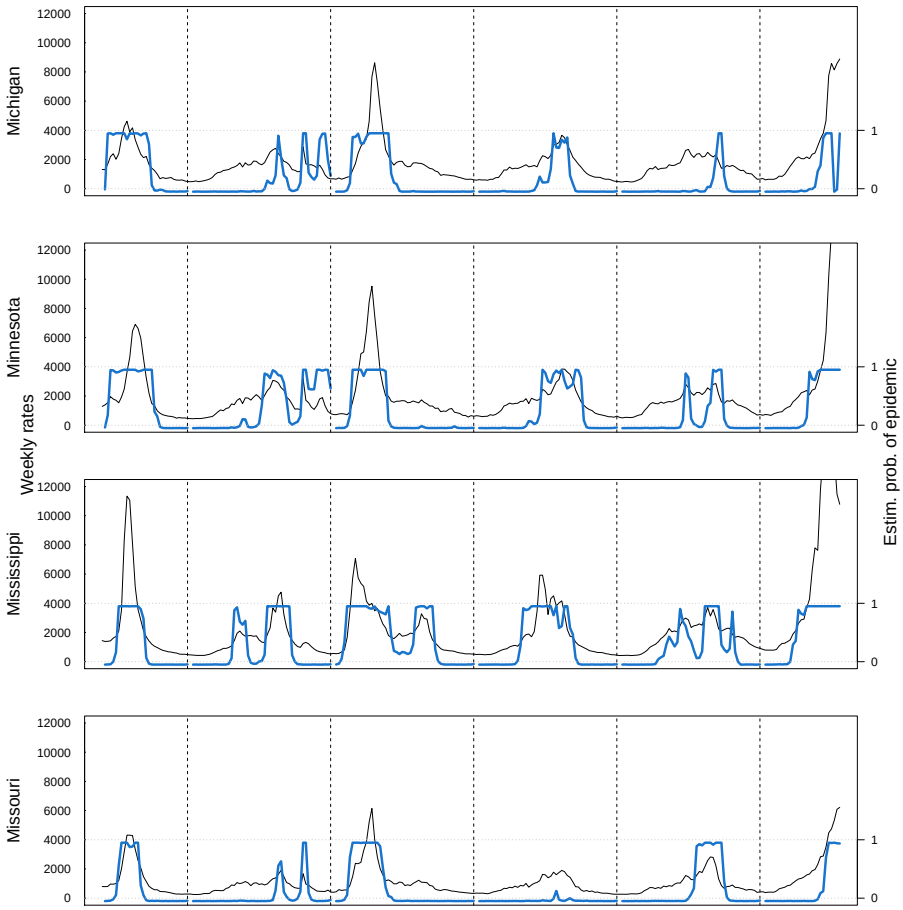

Figure C.6: Estimated probability of being in epidemic phase by the spatio-temporal model on GFT USA data for Michigan, Minnesota, Mississippi and Missouri. In black: weekly estimated influenza incidence per 100 000 inhabitants during seasons from 2007–2008 to 2012–2013.

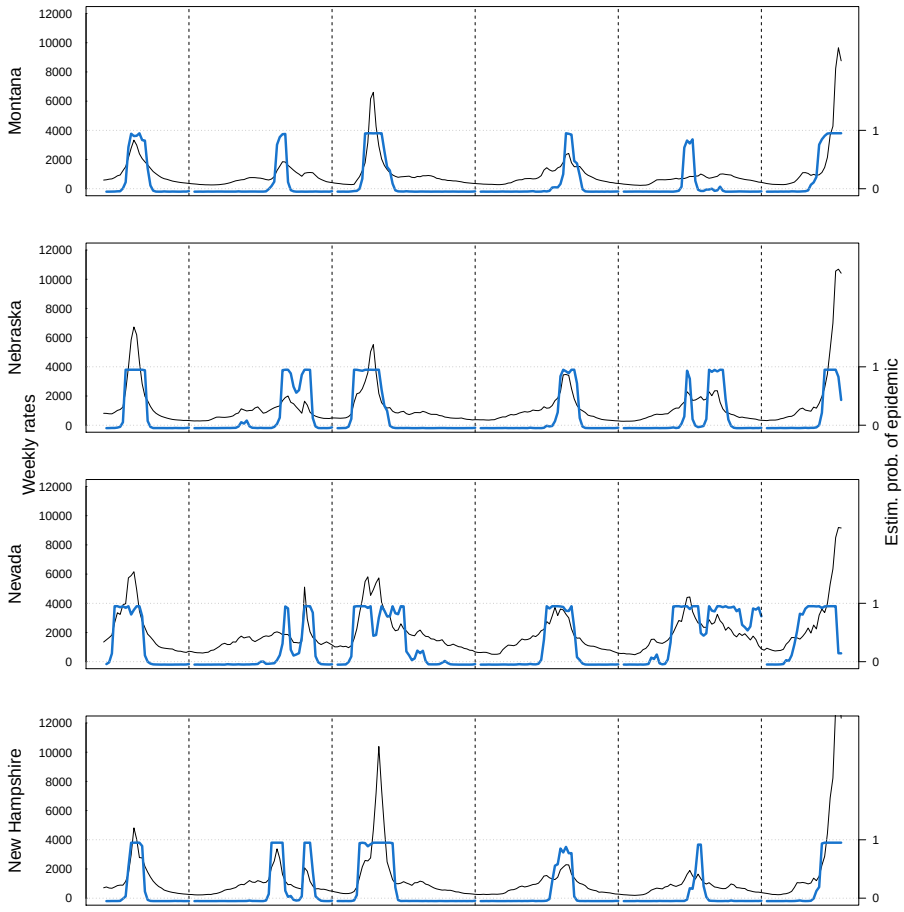

Figure C.7: Estimated probability of being in epidemic phase by the spatio-temporal model on GFT USA data for Montana, Nebraska, Nevada and New Hampshire. In black: weekly estimated influenza incidence per 100 000 inhabitants during seasons from 2007–2008 to 2012–2013.

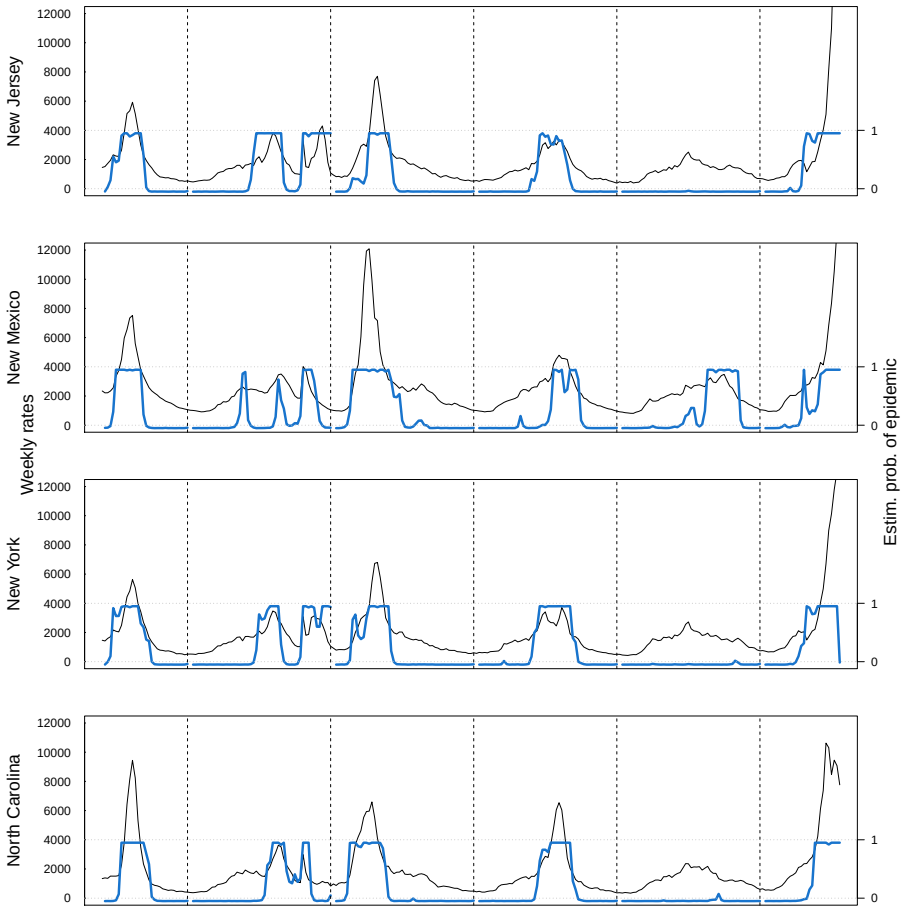

Figure C.8: Estimated probability of being in epidemic phase by the spatio-temporal model on GFT USA data for New Jersey, New Mexico, New York and North Carolina. In black: weekly estimated influenza incidence per 100 000 inhabitants during seasons from 2007–2008 to 2012–2013.

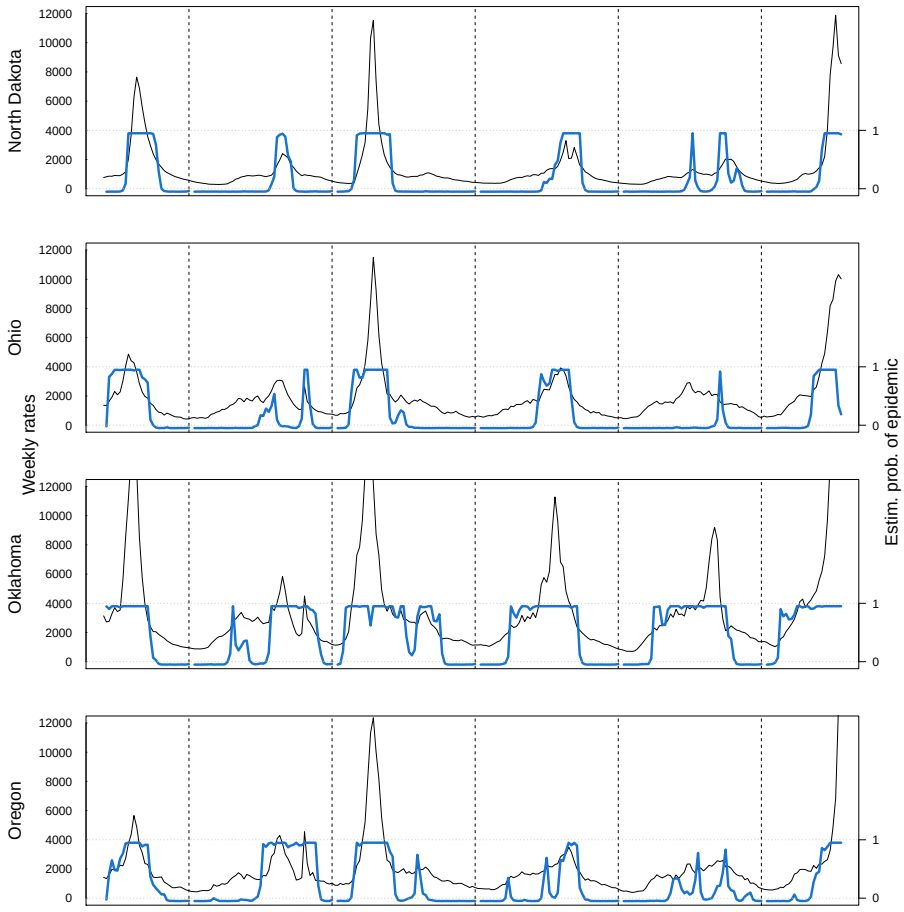

Figure C.9: Estimated probability of being in epidemic phase by the spatio-temporal model on GFT USA data for North Dakota, Ohio, Oklahoma and Oregon. In black: weekly estimated influenza incidence per 100 000 inhabitants during seasons from 2007–2008 to 2012–2013.

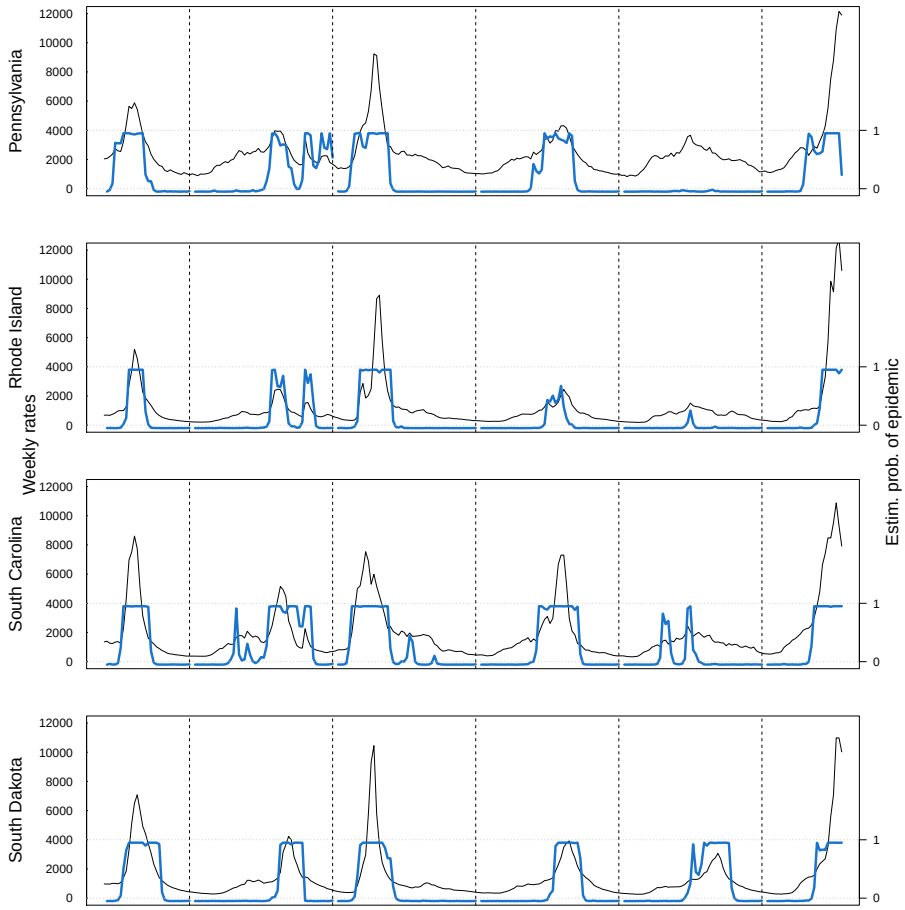

Figure C.10: Estimated probability of being in epidemic phase by the spatio-temporal model on GFT USA data for Pennsylvania, Rhode Island, South Carolina and South Dakota. In black: weekly estimated influenza incidence per 100 000 inhabitants during seasons from 2007–2008 to 2012–2013.

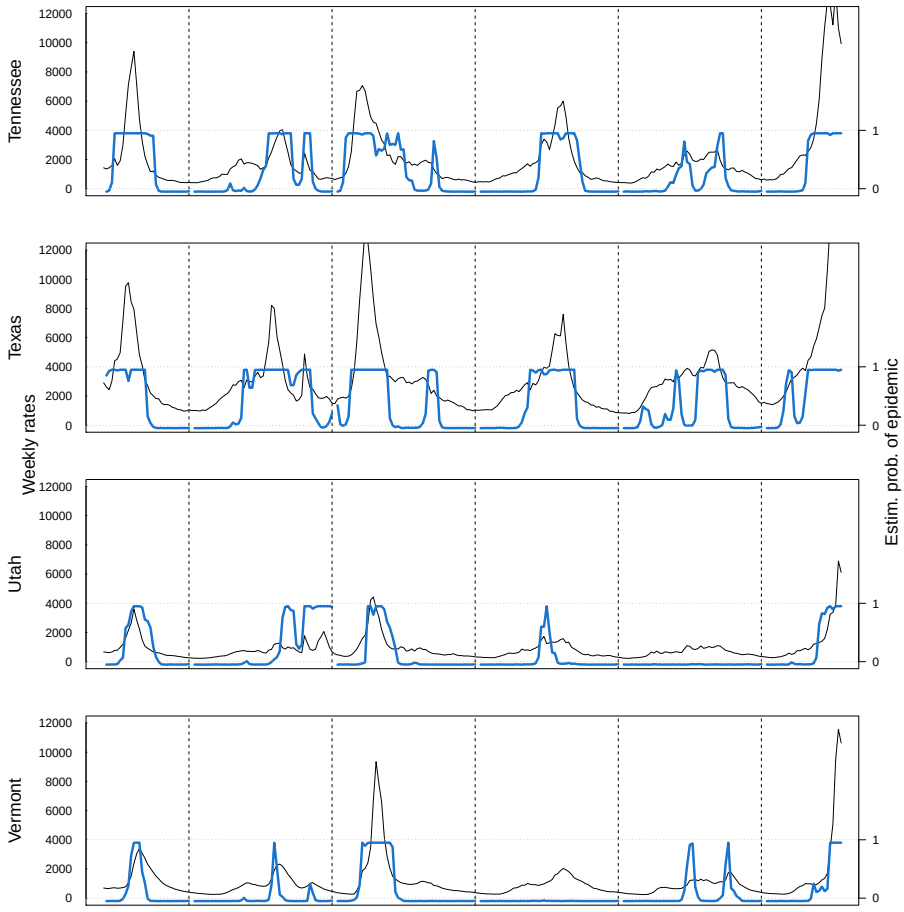

Figure C.11: Estimated probability of being in epidemic phase by the spatio-temporal model on GFT USA data for Tennessee, Texas, Utah and Vermont. In black: weekly estimated influenza incidence per 100 000 inhabitants during seasons from 2007–2008 to 2012–2013.

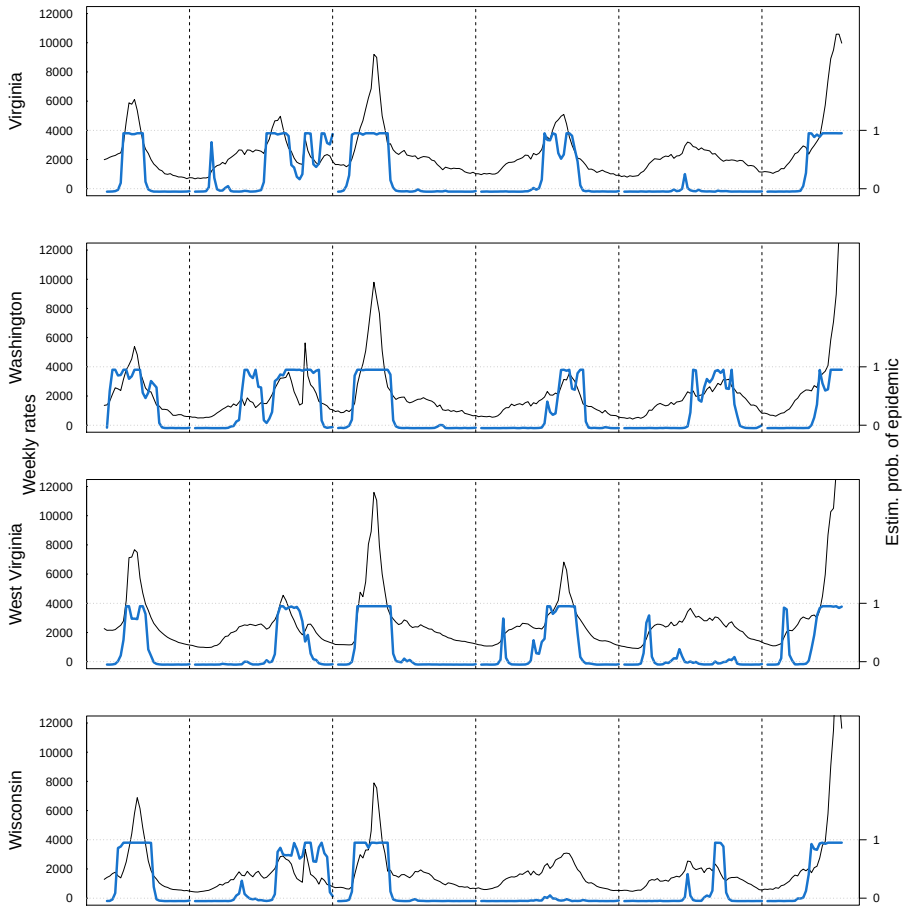

Figure C.12: Estimated probability of being in epidemic phase by the spatio-temporal model on GFT USA data for Virginia, Washington, West Virginia and Wisconsin. In black: weekly estimated influenza incidence per 100 000 inhabitants during seasons from 2007–2008 to 2012–2013.

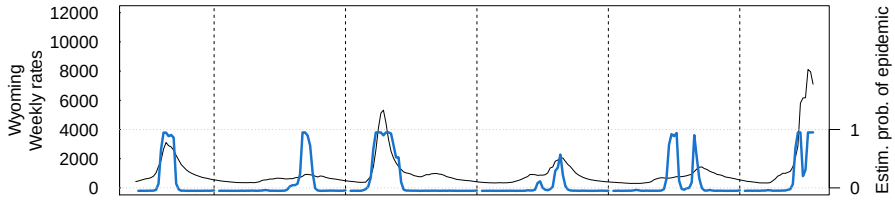

Figure C.13: Estimated probability of being in epidemic phase by the spatio-temporal model on GFT USA data for Wyoming. In black: weekly estimated influenza incidence per 100 000 inhabitants during seasons from 2007–2008 to 2012–2013.

C.2 Comparison of retrospective and online estimation of probability of epidemic phase

In this section we present the comparison of the retrospective and online estimated posterior probability of being in the epidemic phase for all 49 states of the USA during season 2012-2013 by our spatio-temporal proposal applied on GFT USA data.

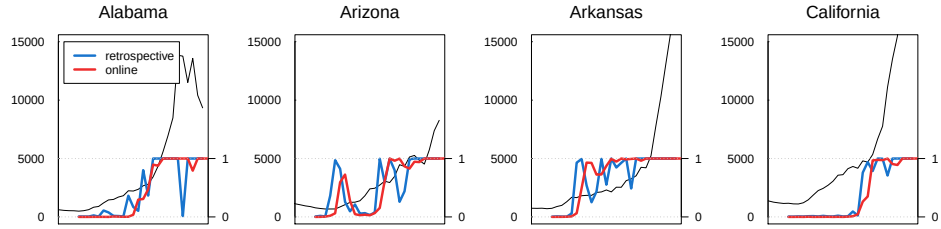

Figure C.14: Comparison of the online and retrospective estimated probability of being in epidemic phase by the spatio-temporal model on GFT USA data for Alabama–California states. In black: weekly estimated influenza incidence per 100 000 inhabitants during season 2012–2013.

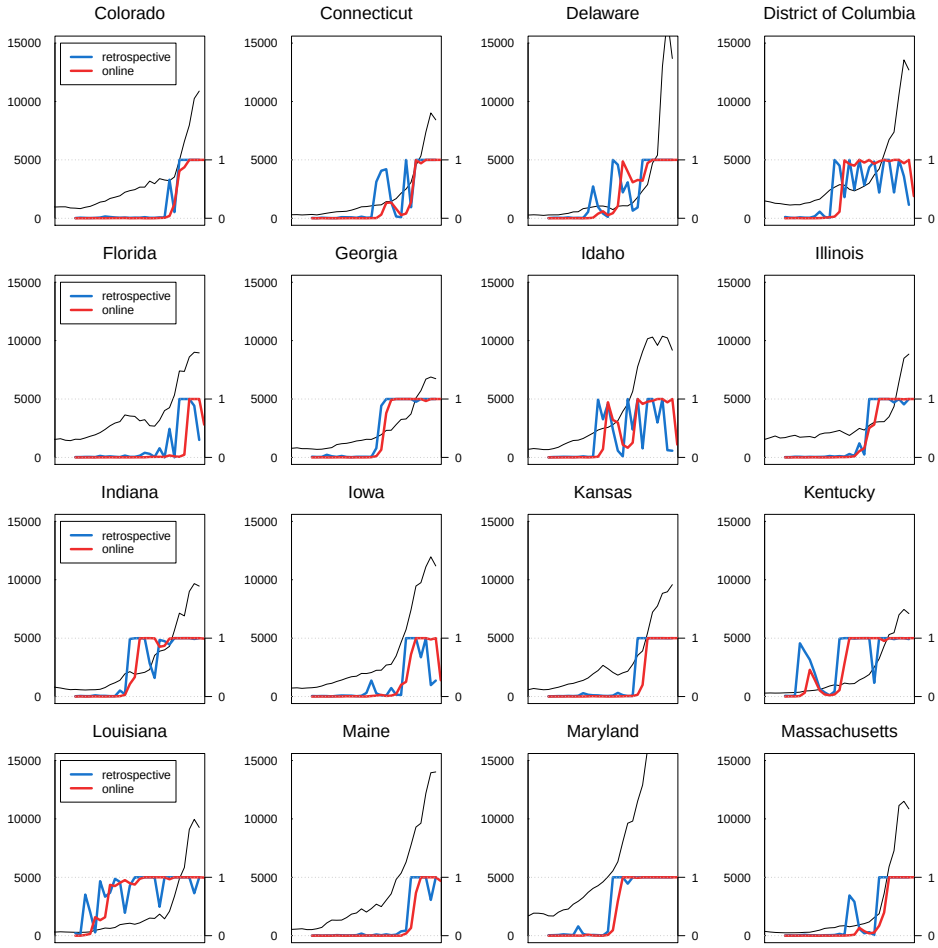

Figure C.15: Comparison of the online and retrospective estimated probability of being in epidemic phase by the spatio-temporal model for Colorado–Massachusetts states. In black: weekly estimated influenza incidence per 100 000 inhabitants during season 2012–2013.

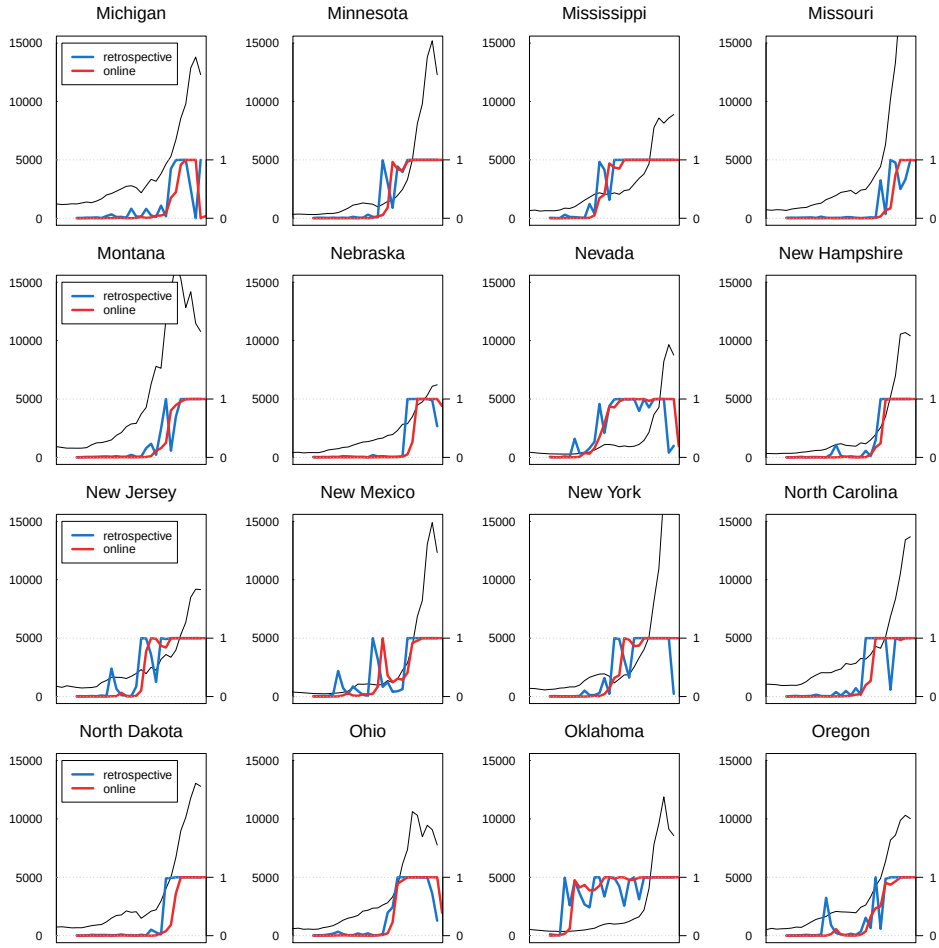

Figure C.16: Comparison of the online and retrospective estimated probability of being in epidemic phase by the spatio-temporal model for Michigan–Oregon states. In black: weekly estimated influenza incidence per 100 000 inhabitants during season 2012–2013.

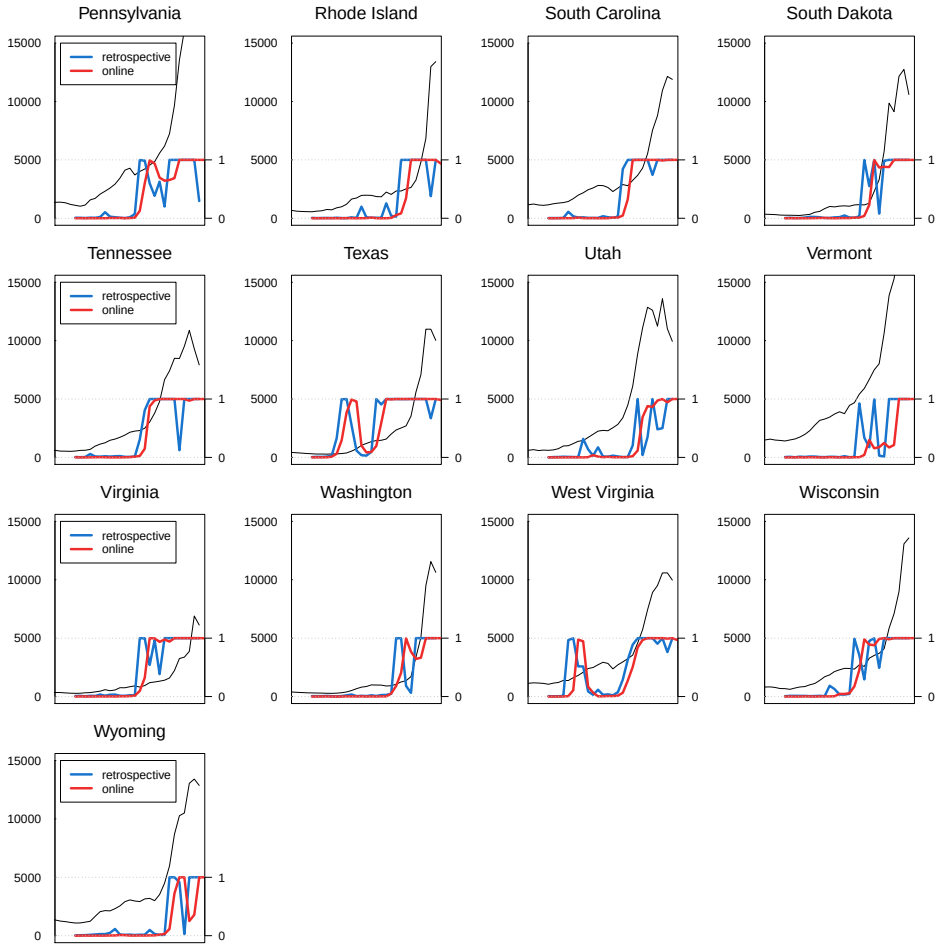

Figure C.17: Comparison of the online and retrospective estimated probability of being in epidemic phase by the spatio-temporal model for Pennsylvania–Wisconsin states. In black: weekly estimated influenza incidence per 100 000 inhabitants during season 2012–2013.

C.3 Estimates of the probability of epidemic in all 49 states by the spatio-temporal model and some alternative proposals

In this section we present the estimated posterior probability of being in the epidemic phase for all 49 states of the USA by our spatio-temporal proposal applied on Google Flu Trends (GFT) USA data, the simplification without ICAR, the simplification without μ_{ts0}, the Leroux variation and the model of Martinez-Beneito et al. (2008a).

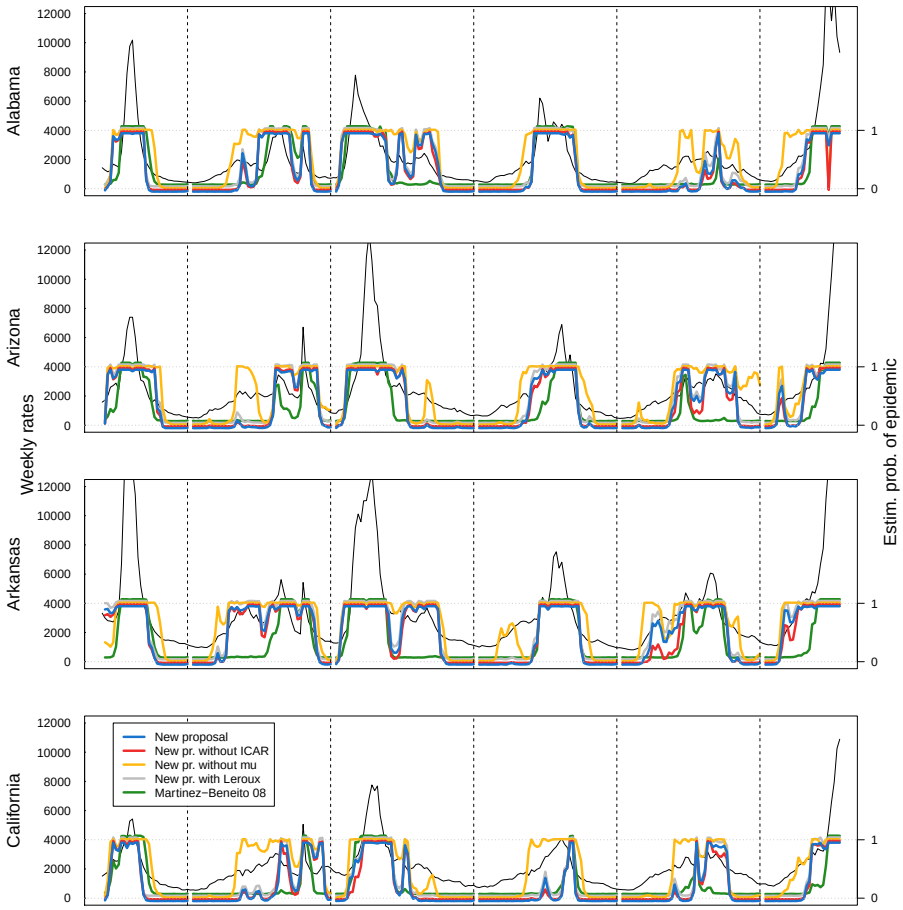

Figure C.18: Estimated probability of being in epidemic phase by the spatio-temporal model its variations and M-B 2008 on GFT USA data for Alabama, Arizona, Arkansas and California. In black: weekly estimated influenza incidence per 100 000 inhabitants during seasons from 2007–2008 to 2012–2013.

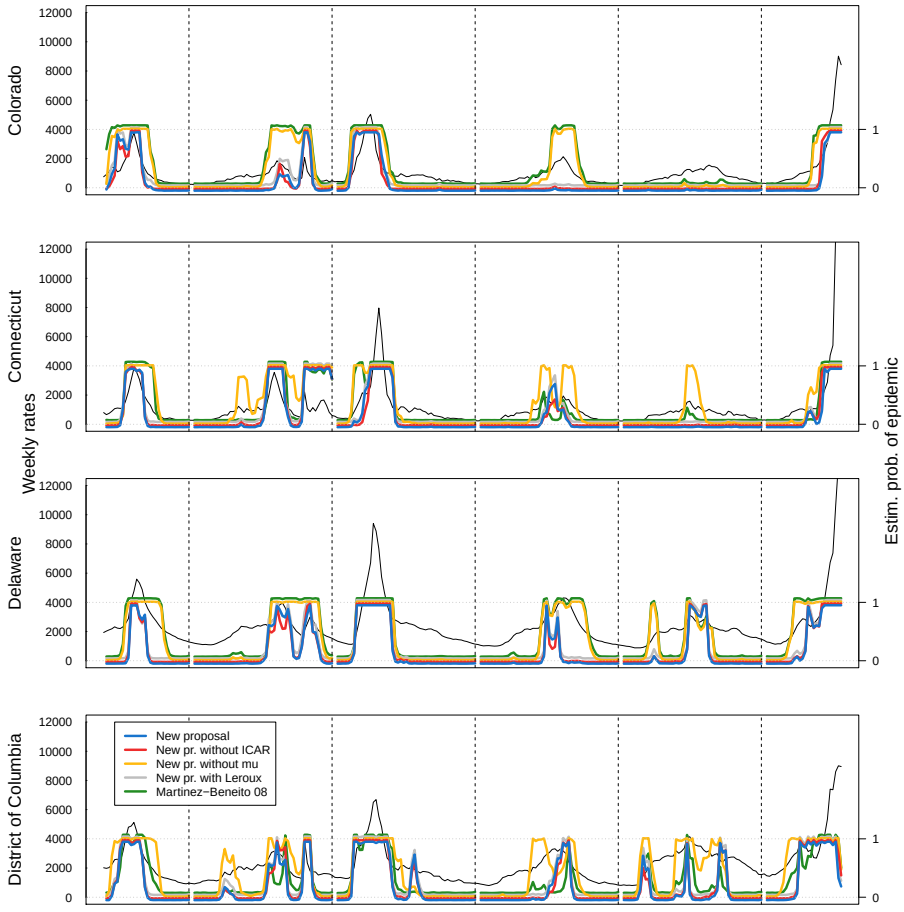

Figure C.19: Estimated probability of being in epidemic phase by the spatio-temporal model its variations and M-B 2008 on GFT USA data for Colorado, Connecticut, Delaware and District of Columbia. In black: weekly estimated influenza incidence per 100 000 inhabitants during seasons from 2007–2008 to 2012–2013.

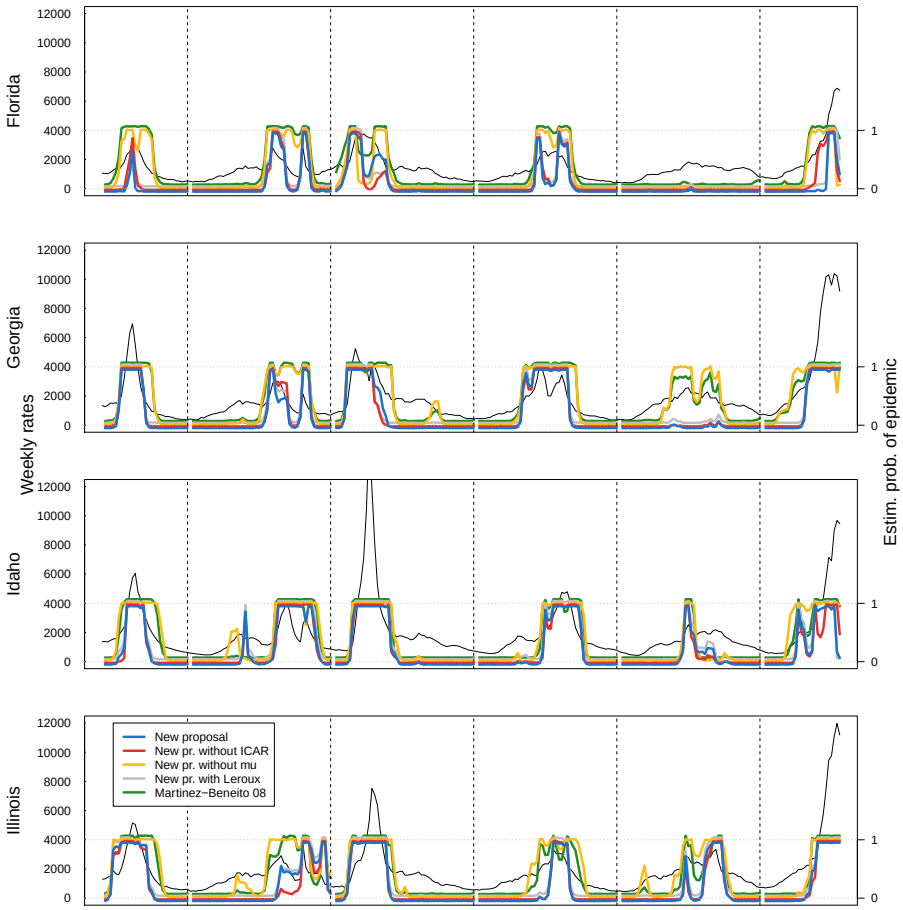

Figure C.20: Estimated probability of being in epidemic phase by the spatio-temporal model its variations and M-B 2008 on GFT USA data for Florida, Georgia, Idaho and Illinois. In black: weekly estimated influenza incidence per 100 000 inhabitants during seasons from 2007–2008 to 2012–2013.

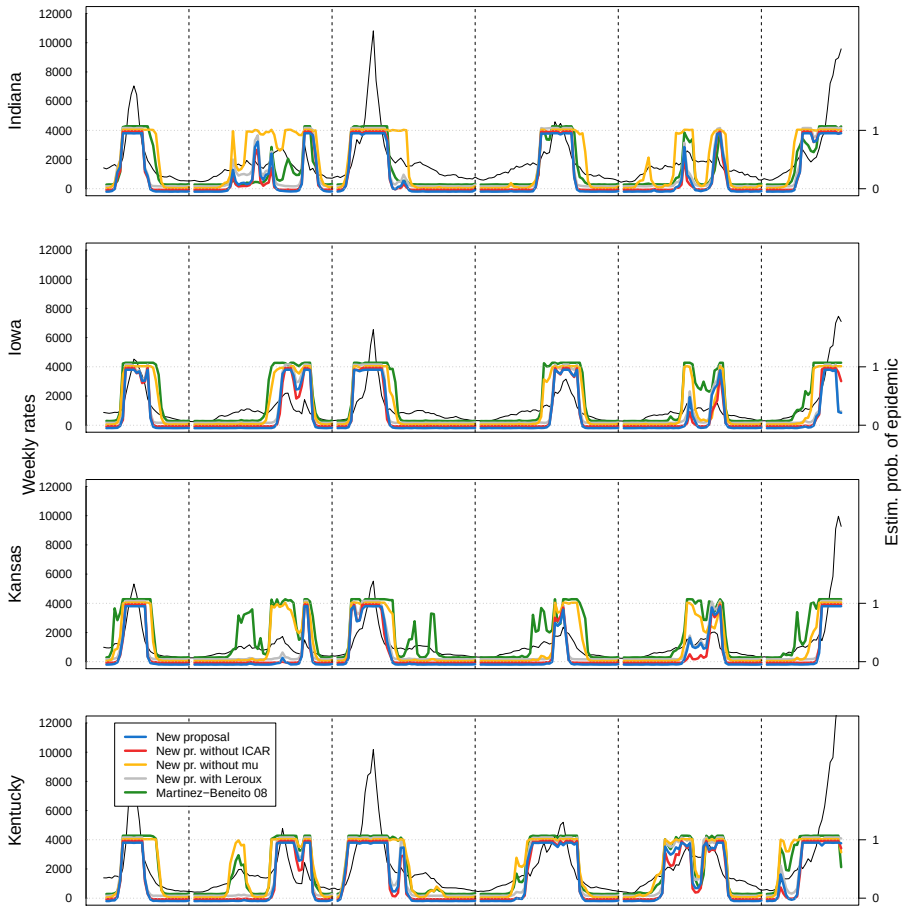

Figure C.21: Estimated probability of being in epidemic phase by the spatio-temporal model its variations and M-B 2008 on GFT USA data for Indiana, Iowa, Kansas and Kentucky. In black: weekly estimated influenza incidence per 100 000 inhabitants during seasons from 2007–2008 to 2012–2013.

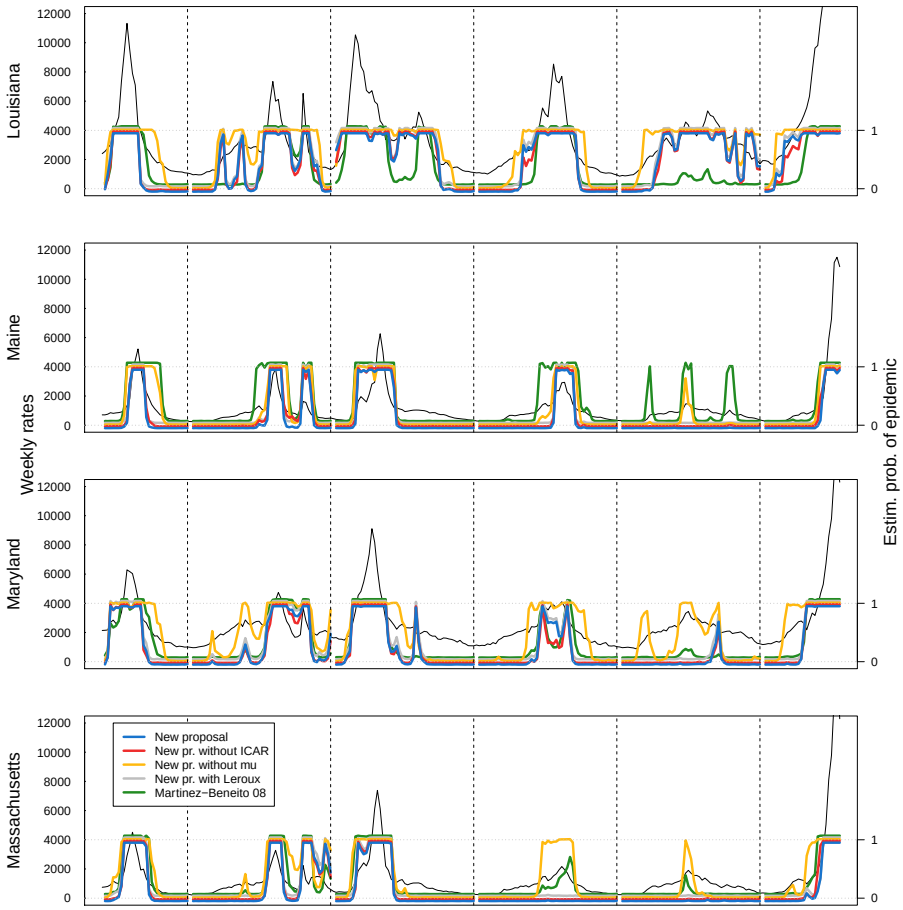

Figure C.22: Estimated probability of being in epidemic phase by the spatio-temporal model its variations and M-B 2008 on GFT USA data for Louisiana, Maine, Maryland and Massachusetts. In black: weekly estimated influenza incidence per 100 000 inhabitants during seasons from 2007–2008 to 2012–2013.

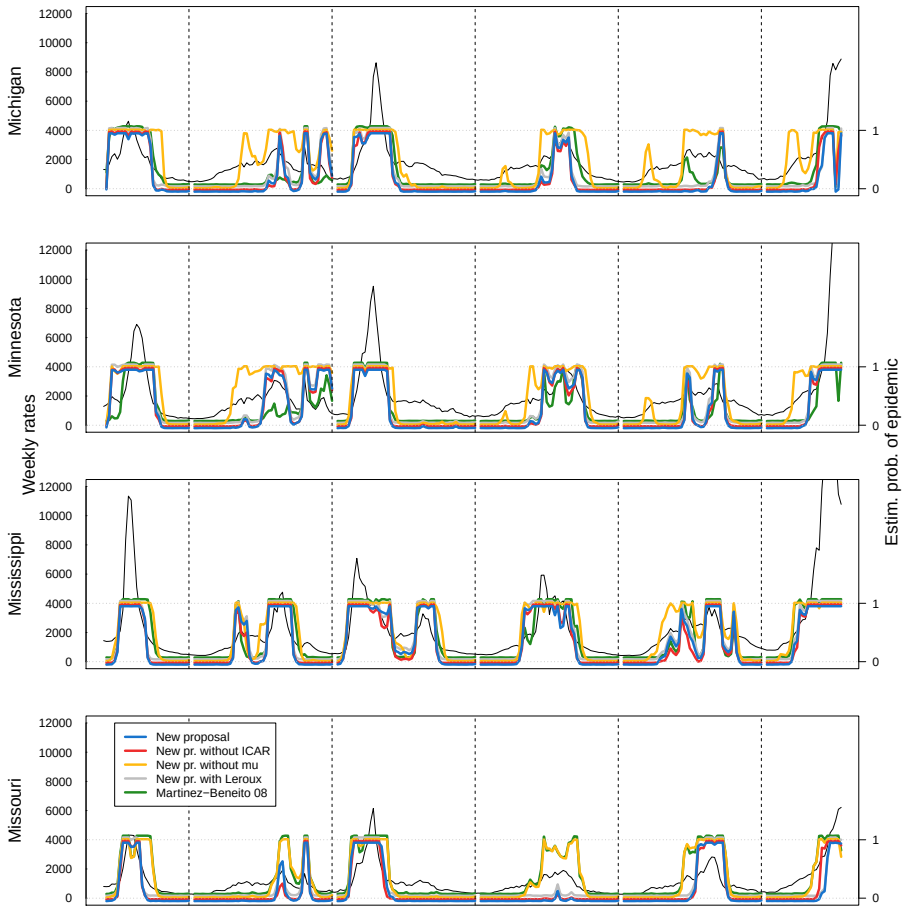

Figure C.23: Estimated probability of being in epidemic phase by the spatio-temporal model its variations and M-B 2008 on GFT USA data for Michigan, Minnesota, Mississippi and Missouri. In black: weekly estimated influenza incidence per 100 000 inhabitants during seasons from 2007–2008 to 2012–2013.

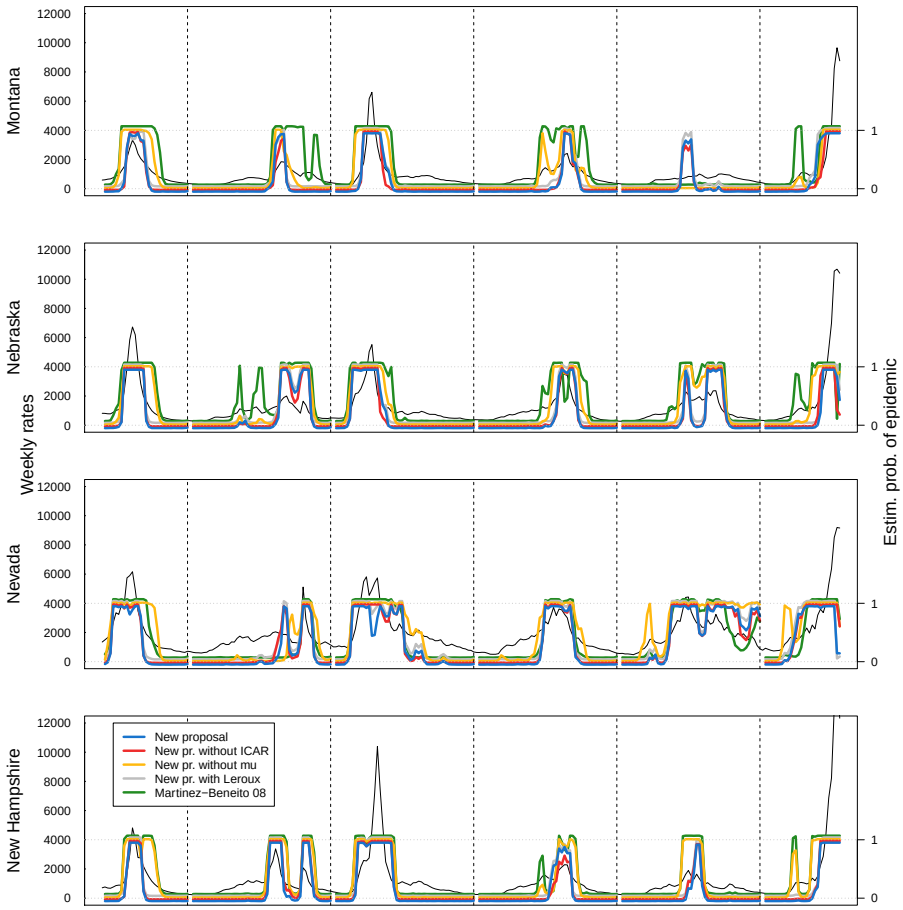

Figure C.24: Estimated probability of being in epidemic phase by the spatio-temporal model its variations and M-B 2008 on GFT USA data for Montana, Nebraska, Nevada and New Hampshire. In black: weekly estimated influenza incidence per 100 000 inhabitants during seasons from 2007–2008 to 2012–2013.

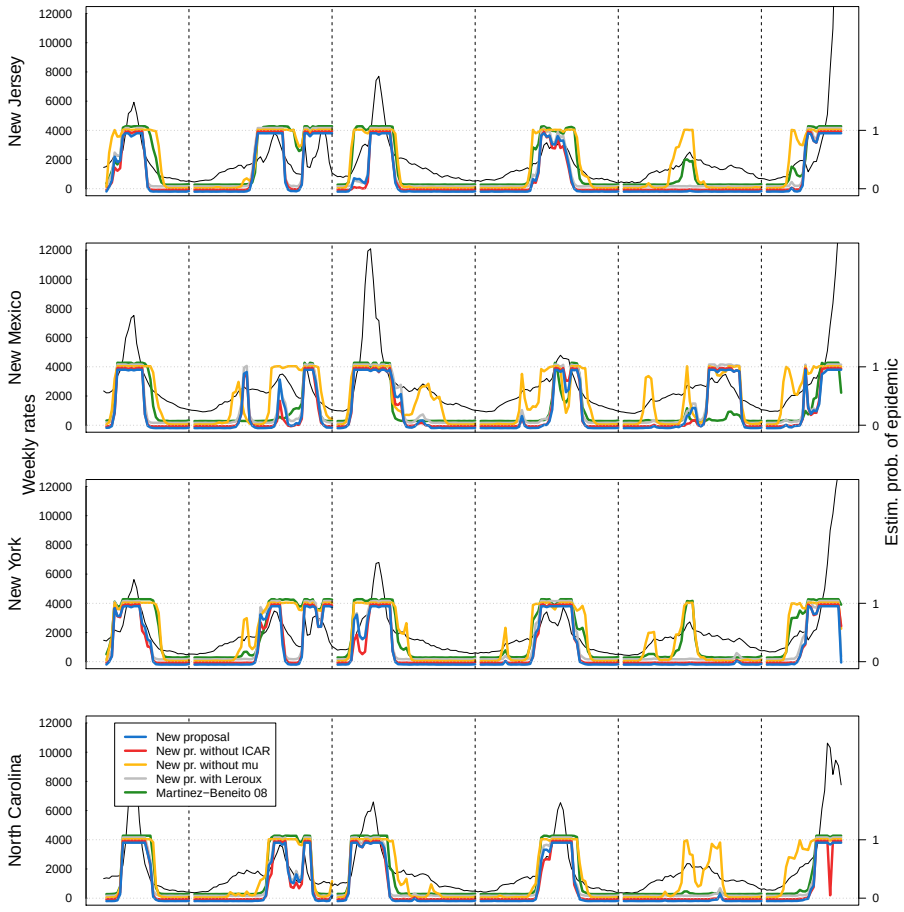

Figure C.25: Estimated probability of being in epidemic phase by the spatio-temporal model its variations and M-B 2008 on GFT USA data for New Jersey, New Mexico, New York and North Carolina. In black: weekly estimated influenza incidence per 100 000 inhabitants during seasons from 2007–2008 to 2012–2013.

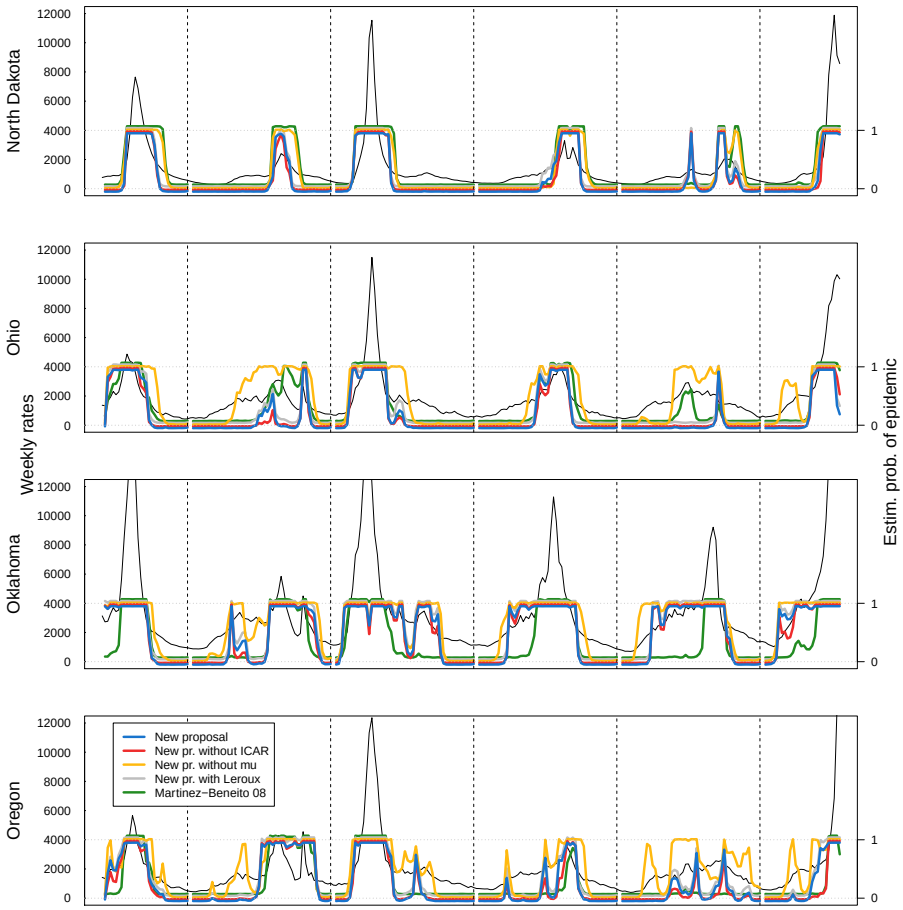

Figure C.26: Estimated probability of being in epidemic phase by the spatio-temporal model its variations and M-B 2008 on GFT USA data for North Dakota, Ohio, Oklahoma and Oregon. In black: weekly estimated influenza incidence per 100 000 inhabitants during seasons from 2007–2008 to 2012–2013.

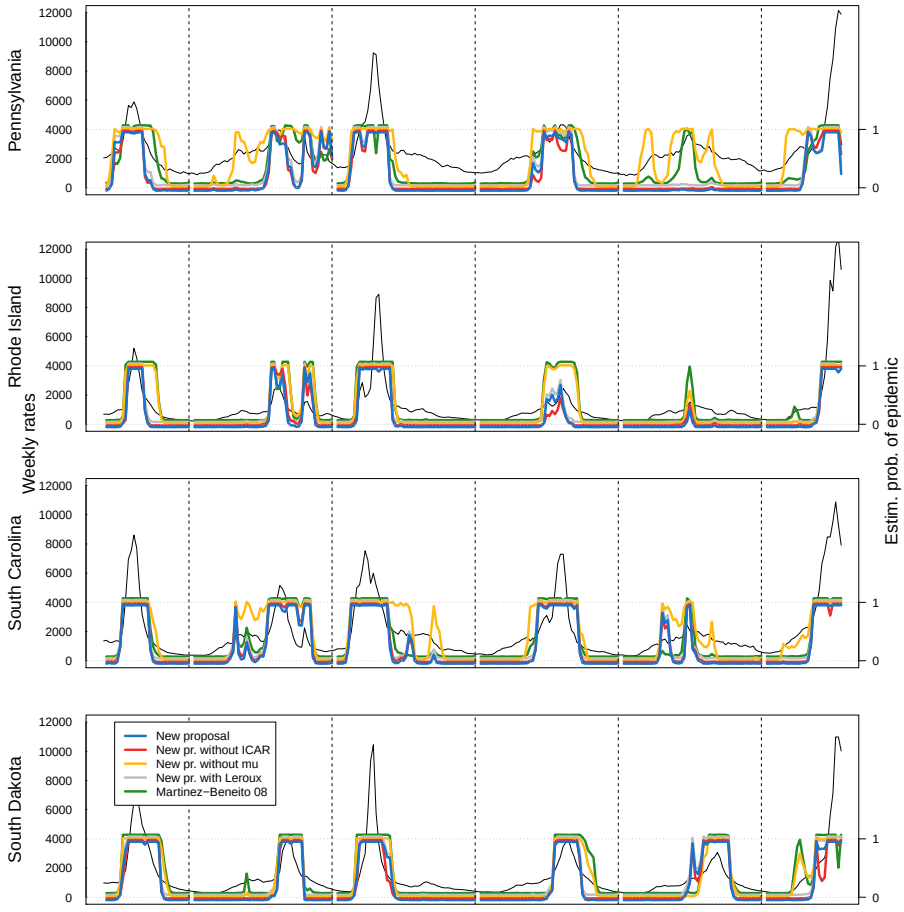

Figure C.27: Estimated probability of being in epidemic phase by the spatio-temporal model its variations and M-B 2008 on GFT USA data for Pennsylvania, Rhode Island, South Carolina and South Dakota. In black: weekly estimated influenza incidence per 100 000 inhabitants during seasons from 2007–2008 to 2012–2013.

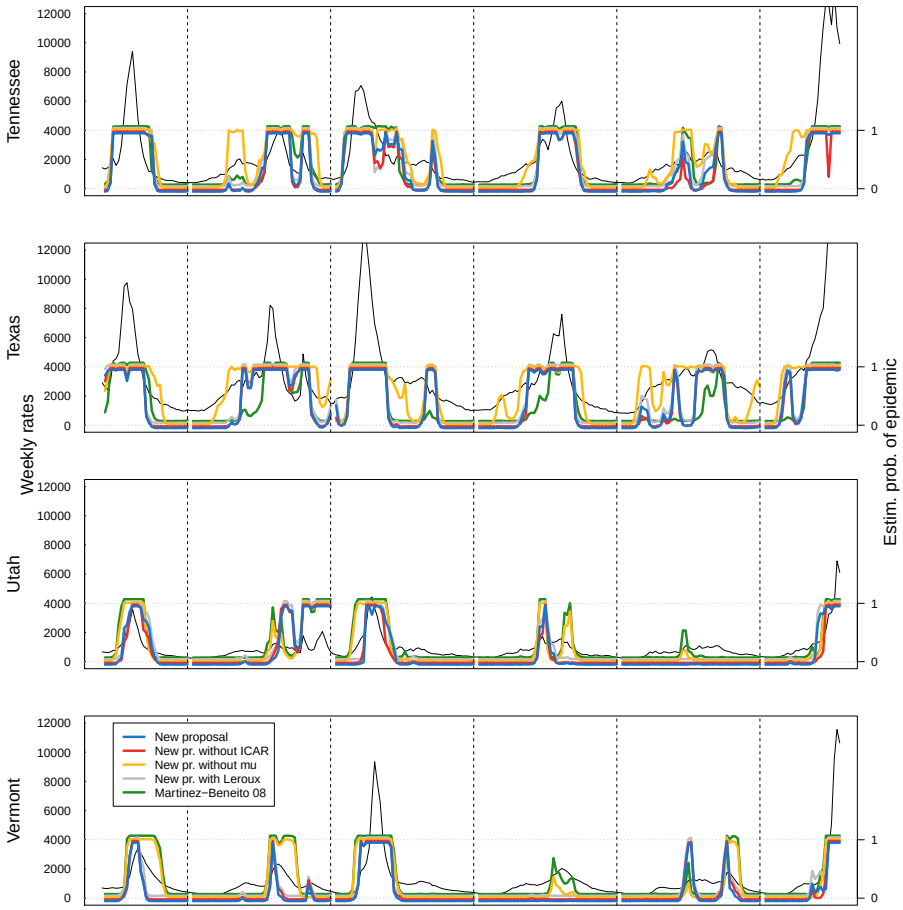

Figure C.28: Estimated probability of being in epidemic phase by the spatio-temporal model its variations and M-B 2008 on GFT USA data for Tennessee, Texas, Utah and Vermont. In black: weekly estimated influenza incidence per 100 000 inhabitants during seasons from 2007–2008 to 2012–2013.

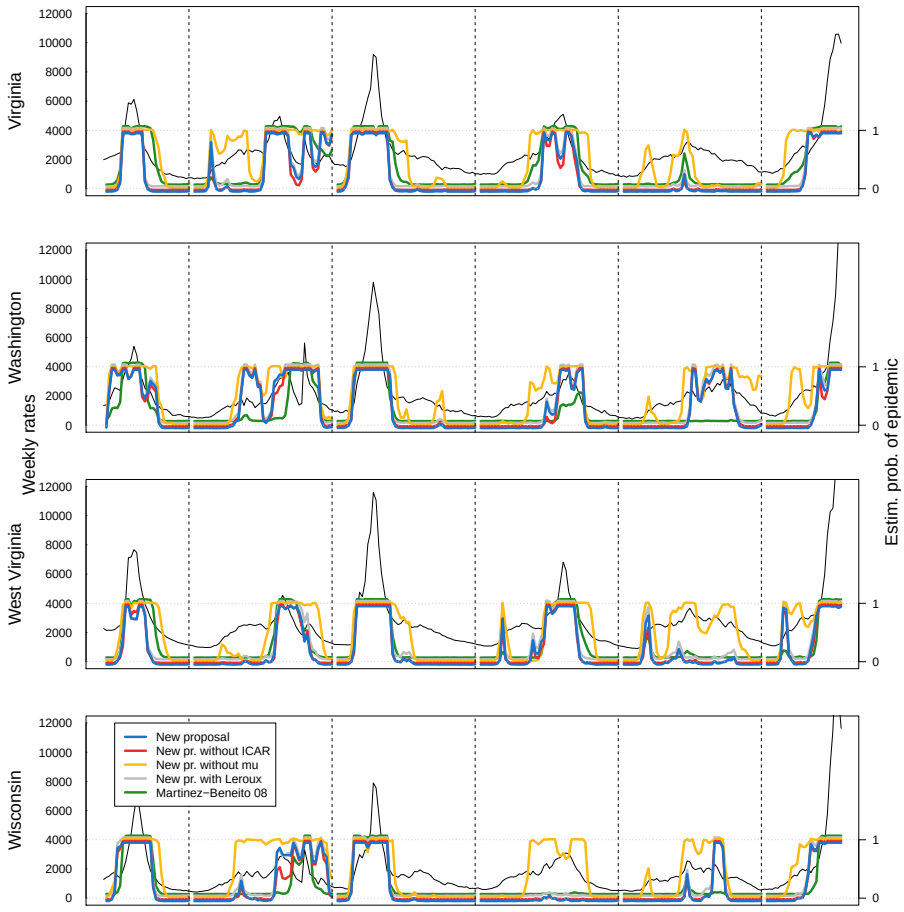

Figure C.29: Estimated probability of being in epidemic phase by the spatio-temporal model its variations and M-B 2008 on GFT USA data for Virginia, Washington, West Virginia and Wisconsin. In black: weekly estimated influenza incidence per 100 000 inhabitants during seasons from 2007–2008 to 2012–2013.

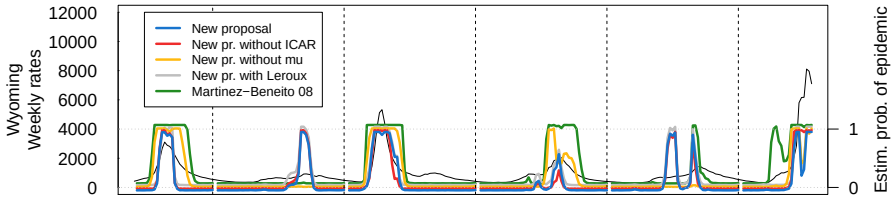

Figure C.30: Estimated probability of being in epidemic phase by the spatio-temporal model its variations and M-B 2008 on GFT USA data for Wyoming. In black: weekly estimated influenza incidence per 100 000 inhabitants during seasons from 2007–2008 to 2012–2013.

Appendix D

Codes

In this appendix we present the WinBUGS codes for the balanced models (AR0-AR0, AR1-AR1 and AR2-AR2) of the framework of temporal models presented in Chapter 2 and the model of the spatio-temporal proposal of Chapter 3. The codes for the unbalanced models of the temporal framework can be easily composed from the codes of the balanced ones.

D.1 Temporal Proposal

D.1.1 AR0-AR0 model

```
model {

  # All weeks
  for (j in 1:nyear) {
    for (i in 1:nweek[j]){
      Obs[i, j] ~ dpois(nu[i,j])
      nu[i,j] <- pob[i,j]*rate[i,j]/100000
      rate[i, j] ~ dnorm(meanrate[i,comp[i,j],j],tau[i, j])
      tau[i, j] <- pow(lambda[comp[i, j], j],-2)
      meanrate[i,1, j] <- mu[1,j]
      meanrate[i,2, j] <- mu[2,j]
```

```
  }
}

# Standard deviations, means and hidden states
for (j in 1:nyear) {
  comp[1, j] ~ dcat(P0[])

  lambda[1, j] ~ dunif(linf,lmed1)
  lambda[2, j] ~ dunif(lmed2,lsup)
  mu[1, j] <- mu1[1]
  mu[2, j] <- mu1[2]
}

mu1[1] <- ranked(thetamu[],1)
mu1[2] <- ranked(thetamu[],2)
for(i in 1:2){thetamu[i] ~ dunif(c,d)}

linf <- ranked(theta[],1)
lmed1 <- ranked(theta[],2)
lmed2 <- ranked(theta[],3)
lsup <- ranked(theta[],4)

for(i in 1:4){theta[i] ~ dunif(a,b)}

# Markovian modeling
for (j in 1:nyear) {
  for (i in 2:nweek[j]) {
    comp[i, j] ~ dcat(P.mat[comp[i-1, j], ])
  }
}

P0[1] ~ dbeta(0.5,0.5)
P0[2] <- 1-P0[1]
P.mat[1,1] ~ dbeta(0.5,0.5)
```

```
  P.mat[2,2] ~ dbeta(0.5,0.5)
  P.mat[1,2] <- 1-P.mat[1,1]
  P.mat[2,1] <- 1-P.mat[2,2]


  #NEW DATA

  # All weeks
  for (i in 1:nobs.new){
    Obs.new[i] ~ dpois(nu.new[i])
    nu.new[i] <- pob.new[i]*rate.new[i]/100000
    rate.new[i] ~ dnorm(meanrate.new[i,comp.new[i]],
                        tau.new[i])
    tau.new[i] <- pow(lambda.new[comp.new[i]],-2)
    meanrate.new[i,2] <- mu.new[2]
    meanrate.new[i,1] <- mu.new[1]
  }

  comp.new[1] ~ dcat(P0[])

  for (i in 2:nobs.new) {
    comp.new[i] ~ dcat(P.mat[comp.new[i-1], ])
  }

  lambda.new[2] ~ dunif(lmed2,lsup)
  lambda.new[1] ~ dunif(linf,lmed1)
  mu.new[1] <- mu1[1]
  mu.new[2] <- mu1[2]

}
```

D.1.2 AR1-AR1 model

```
model {

  # Week 1
  for (j in 1:nyear){
    Obs[1,j] ~ dpois(nu[1,j])
    nu[1,j] <- pob[1,j]*rate[1,j]/100000
    rate[1, j] ~ dnorm(meanrate[1,comp[1,j],j],tau[1, j])
    tau[1, j] <- (1-pow(ro[comp[1, j]],2))*
                 pow(lambda[comp[1, j], j],-2)
    meanrate[1,2, j] <- mu[2,j]
    meanrate[1,1, j] <- mu[1,j]
  }

  # Week >1
  for (j in 1:nyear) {
    for (i in 2:nweek[j]){
      Obs[i, j] ~ dpois(nu[i,j])
      nu[i,j] <- pob[i,j]*rate[i,j]/100000
      rate[i, j] ~ dnorm(meanrate[i,comp[i,j],j],
                         tau[i, j])
      tau[i, j] <- pow(lambda[comp[i, j], j],-2)
      meanrate[i,2, j] <- mu[2,j]+ro[1]*(rate[i-1, j]-mu[2,j])
      meanrate[i,1, j] <- mu[1,j]+ro[2]*(rate[i-1, j]-mu[1,j])
    }
  }

  # Parameters of the autoregressive processes
  ro[1] ~ dunif(-1,1)
  ro[2] ~ dunif(-1,1)

  # Standard deviations and means
  for (j in 1:nyear) {
```

```
    comp[1, j] ~ dcat(P0[])

    lambda[1, j] ~ dunif(linf,lmed1)
    lambda[2, j] ~ dunif(lmed2,lsup)
    mu[1, j] <- mu1[1]
    mu[2, j] <- mu1[2]
  }

  mu1[1] <- ranked(thetamu[],1)
  mu1[2] <- ranked(thetamu[],2)
  for(i in 1:2){thetamu[i] ~ dunif(c,d)}

  linf <- ranked(theta[],1)
  lmed1 <- ranked(theta[],2)
  lmed2 <- ranked(theta[],3)
  lsup <- ranked(theta[],4)

  for(i in 1:4){theta[i] ~ dunif(a,b)}

  # Markovian modeling
  for (j in 1:nyear) {
    for (i in 2:nweek[j]) {
      comp[i, j] ~ dcat(P.mat[comp[i-1, j], ])
    }
  }

  P0[1] ~ dbeta(0.5,0.5)
  P0[2] <- 1-P0[1]
  P.mat[1,1] ~ dbeta(0.5,0.5)
  P.mat[2,2] ~ dbeta(0.5,0.5)
  P.mat[1,2] <- 1-P.mat[1,1]
  P.mat[2,1] <- 1-P.mat[2,2]
```

```
#NEW DATA

# Week 1
Obs.new[1] ~ dpois(nu.new[1])
nu.new[1] <- pob.new[1]*rate.new[1]/100000
rate.new[1] ~ dnorm(meanrate.new[1,comp.new[1]],tau.new[1])
tau.new[1] <- (1-pow(ro[comp.new[1]],2))*
              pow(lambda.new[comp.new[1]],-2)
meanrate.new[1,2] <- mu.new[2]
meanrate.new[1,1] <- mu.new[1]

# Week >1
for (i in 2:nobs.new){
  Obs.new[i] ~ dpois(nu.new[i])
  nu.new[i] <- pob.new[i]*rate.new[i]/100000
  rate.new[i] ~ dnorm(meanrate.new[i,comp.new[i]],
                      tau.new[i])
  tau.new[i] <- pow(lambda.new[comp.new[i]],-2)
  meanrate.new[i,2] <- mu.new[2]+
                       ro[1]*(rate.new[i-1]-mu.new[2])
  meanrate.new[i,1] <- mu.new[1]+
                       ro[2]*(rate.new[i-1]-mu.new[1])
}

comp.new[1] ~ dcat(P0[])

for (i in 2:nobs.new) {
  comp.new[i] ~ dcat(P.mat[comp.new[i-1], ])
}

lambda.new[2] ~ dunif(lmed2,lsup)
lambda.new[1] ~ dunif(linf,lmed1)
mu.new[1] <- mu1[1]
mu.new[2] <- mu1[2]
```

```
}
```

D.1.3 AR2-AR2 model

```
model {

  # Week 1
  for (j in 1:nyear){
    Obs[1,j] ~ dpois(nu[1,j])
    nu[1,j] <- pob[1,j]*rate[1,j]/100000
    rate[1, j] ~ dnorm(meanrate[1,comp[1,j],j],tau[1, j])
    tau[1, j] <- invsigma[comp[1, j]]*
                 pow(lambda[comp[1, j],j],-2)
    meanrate[1,1, j] <- mu[1,j]
    meanrate[1,2, j] <- mu[2,j]
  }

  # Week 2
  for (j in 1:nyear){
    Obs[2,j] ~ dpois(nu[2,j])
    nu[2,j] <- pob[2,j]*rate[2,j]/100000
    rate[2, j] ~ dnorm(meanrate[2,comp[2,j],j],tau[2, j])
    tau[2, j] <- pow(lambda[comp[2, j],j],-2)*
                 invsigma[comp[2,j]]*
                 pow(pow(1-ro1[2],2)/(pow(1-ro1[2],2)-
                     pow(ro1[1],2)),equals(comp[2,j],1))*
                 pow(pow(1-ro2[2],2)/(pow(1-ro2[2],2)-
                     pow(ro2[1],2)),equals(comp[2,j],2))
    meanrate[2,1, j] <- mu[1,j]+(ro1[1]*(rate[1, j]-mu[1,j])/
                                 (1-ro1[2]))
    meanrate[2,2, j] <- mu[2,j]+(ro2[1]*(rate[1, j]-mu[2,j])/
                                 (1-ro2[2]))
  }
```

```
# Week >2
for (j in 1:nyear) {
  for (i in 3:nweek[j]){
    Obs[i, j] ~ dpois(nu[i,j])
    nu[i,j] <- pob[i,j]*rate[i,j]/100000
    rate[i, j] ~ dnorm(meanrate[i,comp[i,j],j],tau[i, j])
    tau[i, j] <- pow(lambda[comp[i, j], j],-2)
    meanrate[i,1, j] <- mu[1,j]+ro1[1]*
                        (rate[i-1, j]-mu[1,j])+ro1[2]*
                        (rate[i-2, j]-mu[1,j])
    meanrate[i,2, j] <- mu[2,j]+ro2[1]*
                        (rate[i-1, j]-mu[2,j])+ro2[2]*
                        (rate[i-2, j]-mu[2,j])
  }
}

invsigma[1] <- (1-pow(ro1[1],2)-pow(ro1[1],2)*ro1[2]-ro1[2]-
                pow(ro1[2],2)+pow(ro1[2],3))/(1-ro1[2])
invsigma[2] <- (1-pow(ro2[1],2)-pow(ro2[1],2)*ro2[2]-ro2[2]-
                pow(ro2[2],2)+pow(ro2[2],3))/(1-ro2[2])

# Parameters of the autoregressive processes
ro1[1] ~ dunif(-2,2)
ro1[2] ~ dunif(-1,1)
aux1 <- 1
aux1 ~ dbern(cond1)
cond1 <- step(1-ro1[1]-ro1[2])*step(1+ro1[1]-ro1[2])

ro2[1] ~ dunif(-2,2)
ro2[2] ~ dunif(-1,1)
aux2 <- 1
aux2 ~ dbern(cond2)
cond2 <- step(1-ro2[1]-ro2[2])*step(1+ro2[1]-ro2[2])
```

```
# Standard deviations and means
for (j in 1:nyear) {
  comp[1, j] ~ dcat(P0[])
  lambda[1, j] ~ dunif(linf,lmed1)
  lambda[2, j] ~ dunif(lmed2,lsup)
  mu[1, j] <- mu1[1]
  mu[2, j] <- mu1[2]
}
mu1[1] <- ranked(thetamu[],1)
mu1[2] <- ranked(thetamu[],2)
for(i in 1:2){thetamu[i] ~ dunif(c,d)}

linf <- ranked(theta[],1)
lmed1 <- ranked(theta[],2)
lmed2 <- ranked(theta[],3)
lsup <- ranked(theta[],4)

for(i in 1:4){theta[i] ~ dunif(a,b)}

# Markovian modeling
for (j in 1:nyear) {
  for (i in 2:nweek[j]) {
    comp[i, j] ~ dcat(P.mat[comp[i-1, j], ])
  }
}

P0[1] ~ dbeta(0.5,0.5)
P0[2] <- 1-P0[1]
P.mat[1,1] ~ dbeta(0.5,0.5)
P.mat[2,2] ~ dbeta(0.5,0.5)
P.mat[1,2] <- 1-P.mat[1,1]
P.mat[2,1] <- 1-P.mat[2,2]
```

```
#NEW DATA

# Week 1
Obs.new[1] ~ dpois(nu.new[1])
nu.new[1] <- pob.new[1]*rate.new[1]/100000
rate.new[1] ~ dnorm(meanrate.new[1,comp.new[1]],tau.new[1])
tau.new[1] <- invsigma[comp.new[1]]*
              pow(lambda.new[comp.new[1]],-2)
meanrate.new[1,1] <- mu.new[1]
meanrate.new[1,2] <- mu.new[2]

# Week 2
Obs.new[2] ~ dpois(nu.new[2])
nu.new[2] <- pob.new[2]*rate.new[2]/100000
rate.new[2] ~ dnorm(meanrate.new[2,comp.new[2]],tau.new[2])
tau.new[2] <- pow(lambda.new[comp.new[2]],-2)*
              invsigma[comp.new[2]]*
              pow(pow(1-ro1[2],2)/(pow(1-ro1[2],2)-
                  pow(ro1[1],2)),equals(comp.new[2],1))*
              pow(pow(1-ro2[2],2)/(pow(1-ro2[2],2)-
                  pow(ro2[1],2)),equals(comp.new[2],2))
meanrate.new[2,1] <- mu.new[1]+
                     (ro1[1]*(rate.new[1]-mu.new[1])/
                      (1-ro1[2]))
meanrate.new[2,2] <- mu.new[2]+
                     (ro2[1]*(rate.new[1]-mu.new[2])/
                      (1-ro2[2]))

# Week >2
for (i in 3:nobs.new){
  Obs.new[i] ~ dpois(nu.new[i])
  nu.new[i] <- pob.new[i]*rate.new[i]/100000
  rate.new[i] ~ dnorm(meanrate.new[i,comp.new[i]],
```

```
                        tau.new[i])
    tau.new[i] <- pow(lambda.new[comp.new[i]],-2)
    meanrate.new[i,1] <- mu.new[1]+ro1[1]*
                         (rate.new[i-1]-mu.new[1])+ro1[2]*
                         (rate.new[i-2]-mu.new[1])
    meanrate.new[i,2] <- mu.new[2]+ro2[1]*
                         (rate.new[i-1]-mu.new[2])+ro2[2]*
                         (rate.new[i-2]-mu.new[2])
  }

  comp.new[1] ~ dcat(P0[])

  for (i in 2:nobs.new) {
    comp.new[i] ~ dcat(P.mat[comp.new[i-1], ])
  }
  lambda.new[2] ~ dunif(lmed2,lsup)
  lambda.new[1] ~ dunif(linf,lmed1)
  mu.new[1] <- mu1[1]
  mu.new[2] <- mu1[2]

}
```

D.2 Spatio-temporal model

```
model{

  # First year
  ##############

  for (i in 1:nreg){
    Y[t0+1,i,1]  ~  dnorm(R[t0+1,i,1,Z[t0+1,i,1]],
                          tau[t0+1,i,1])
    tau[t0+1,i,1] <- pow(sigma[Z[t0+1,i,1]],-2)*
```

```
                    (1-(Z[t0+1,i,1]-1)*rho*rho)
  R[t0+1,i,1,1] <- mufirst0[1]*sigmamu[1]
  R[t0+1,i,1,2] <- mufirst1[1]*sigmamu[2]+
                    psifirst[1,i]*sigmapsi
}

for (t in 2:nobs.first){
  for (i in 1:nreg){
    Y[t+t0,i,1] ~ dnorm(R[t+t0,i,1,Z[t+t0,i,1]],
                        tau[t+t0,i,1])
    tau[t+t0,i,1] <- pow(sigma[Z[t+t0,i,1]],-2)
    R[t+t0,i,1,1] <- mufirst0[t]*sigmamu[1]
    R[t+t0,i,1,2] <- rho*Y[t+t0-1,i,1]+
                     mufirst1[t]*sigmamu[2]+
                     psifirst[t,i]*sigmapsi
  }
}

# Spatio-temporal structure
for(t in 1:nobs.first){
  mufirst0[t] ~ dnorm(0,1)
  mufirst1[t] ~ dnorm(0,1)
  psifirst[t,1:nreg] ~ car.normal(adj[],weights[],num[],1)
}

# Markovian modeling
for (t in 2:nobs.first){
  for (i in 1:nreg){
    Z[t+t0,i,1] ~ dcat(P.mat[Z[t+t0-1,i,1], ])
  }
}

for (i in 1:nreg){
  Z[1+t0,i,1] ~ dcat(P0[])
```

```
}


# Second year till second last
################################

for (i in 1:nreg){
  for (j in 2:nhist){
    Y[1,i,j] ~ dnorm(R[1,i,j,Z[1,i,j]],tau[1,i,j])
    tau[1,i,j] <- pow(sigma[Z[1,i,j]],-2)*
                  (1-(Z[1,i,j]-1)*rho*rho)
    R[1,i,j,1] <- mu0[1,j-1]*sigmamu[1]
    R[1,i,j,2] <- mu1[1,j-1]*sigmamu[2]+
                  psi[1,j-1,i]*sigmapsi
  }
}

for (t in 2:weeks){
  for (i in 1:nreg){
    for (j in 2:nhist){
      Y[t,i,j]  ~  dnorm(R[t,i,j,Z[t,i,j]],tau[t,i,j])
      tau[t,i,j] <- pow(sigma[Z[t,i,j]],-2)
      R[t,i,j,1] <- mu0[t,j-1]*sigmamu[1]
      R[t,i,j,2] <- rho*Y[t-1,i,j]+mu1[t,j-1]*sigmamu[2]+
                    psi[t,j-1,i]*sigmapsi
    }
  }
}

# Spatio-temporal structure
for(t in 1:weeks){
  for(j in 1:(nhist-1)){
    mu0[t,j] ~ dnorm(0,1)
    mu1[t,j] ~ dnorm(0,1)
```

```
    psi[t,j,1:nreg] ~ car.normal(adj[],weights[],num[],1)
  }
}

# Markovian modeling
for (t in 2:weeks){
  for (i in 1:nreg){
    for (j in 2:nhist){
      Z[t,i,j] ~ dcat(P.mat[Z[t-1,i,j], ])
    }
  }
}

for (i in 1:nreg){
  for (j in 2:nhist){
    Z[1,i,j] ~ dcat(P0[])
  }
}

# Prior distributions of parameters
rho ~ dunif(0,1)

sigmamu[1] <- sigma[1]*lambda
sigmamu[2] <- sigma[2]*lambda

lambda ~ dunif(g,h)

sigmapsi ~ dunif(e,f)

sigma[1] <- ranked(theta[],1)
sigma[2] <- ranked(theta[],2)

for(k in 1:2){
  theta[k] ~ dunif(a,b)
```

```
}

P0[1] ~ dbeta(0.5,0.5)
P0[2] <-1-P0[1]
P.mat[1,1] ~ dbeta(0.5,0.5)
P.mat[2,2] ~ dbeta(0.5,0.5)
P.mat[1,2] <- 1-P.mat[1,1]
P.mat[2,1] <- 1-P.mat[2,2]


# Last year
#############

for (i in 1:nreg){
  Y[1,i,nhist+1] ~ dnorm(R[1,i,nhist+1,Z[1,i,nhist+1]],
                         tau[1,i,nhist+1])
  tau[1,i,nhist+1] <- pow(sigma[Z[1,i,nhist+1]],-2)*
                      (1-(Z[1,i,nhist+1]-1)*rho*rho)
  R[1,i,nhist+1,1] <- munew0[1]*sigmamu[1]
  R[1,i,nhist+1,2] <- munew1[1]*sigmamu[2]+
                      psinew[1,i]*sigmapsi
}

for (t in 2:nobs.new){
  for (i in 1:nreg){
    Y[t,i,nhist+1]  ~  dnorm(R[t,i,nhist+1,Z[t,i,nhist+1]],
                             tau[t,i,nhist+1])
    tau[t,i,nhist+1]<- pow(sigma[Z[t,i,nhist+1]],-2)
    R[t,i,nhist+1,1]<- munew0[t]*sigmamu[1]
    R[t,i,nhist+1,2]<- rho*Y[t-1,i,nhist+1] +
                       munew1[t]*sigmamu[2] +
                       psinew[t,i]*sigmapsi
  }
}
```

```
  # Spatio-temporal structure
  for(t in 1:nobs.new){
    munew0[t] ~ dnorm(0,1)
    munew1[t] ~ dnorm(0,1)
    psinew[t,1:nreg] ~ car.normal(adj[],weights[],num[],1)
  }

  # Markovian modeling
  for (t in 2:nobs.new){
    for (i in 1:nreg){
      Z[t,i,nhist+1] ~ dcat(P.mat[Z[t-1,i,nhist+1], ])
    }
  }

  for (i in 1:nreg){
    Z[1,i,nhist+1] ~ dcat(P0[])
  }

  for (i in 1:nreg){
    Zlast[i]<-Z[nobs.new,i,nhist+1]-1
  }

  for (t in 1:nobs.new){
    for (i in 1:nreg){
      Zlastarray[t,i]<-Z[t,i,nhist+1]-1
    }
  }

}
```

Bibliography

Akaike H (**1974**). A new look at the statistical model identification. *IEEE Transactions on Automatic Control*, 19(6):716–723. 144

Al-Osh M. A and Alzaid A. A (**1987**). First-order integer-valued autoregressive (INAR(1)) process. *Journal of Time Series Analysis*, 8(3): 261–275. 19

Amorós R, Conesa D, Martinez-beneito M. A, and López-Quílez A (**2015**). Statistical methods for detecting the onset of influenza outbreaks: A review. *REVSTAT-Statistical Journal*, 13(1):41–62. xiii, 13

Assunção R, Costa M, Tavares A, and Ferreira S (**2006**). Fast detection of arbitrarily shaped disease clusters. *Statistics in medicine*, 25 (5):723–742. 95

Banks D, Datta G, Karr A, Lynch J, Niemi J, and Vera F (**2012**). Bayesian CAR models for syndromic surveillance on multiple data streams: Theory and practice. *Information Fusion*, 13(2):105–116. xlii, 100, 101, 102, 157, 158, 160

Baron M (**2002**). Bayes and asymptotically pointwise optimal stopping rules for the detection of influenza epidemics. In Gatsonis C, Kass R. E, Carriquiry A, Gelman A, Higdon D, Pauler D. K, and Verdinelli I, editors, *Case studies in Bayesian statistics, Vol. VI*, pages 153–163. Springer, New York. 34

Baum L. E, Petrie T, Soules G, and Weiss N (1970). A maximization technique occurring in the statistical analysis of probabilistic functions of Markov chains. *The Annals of Mathematical Statistics*, 41(1):164–171. 28

BayesX. Bayesian inference in structured additive regression models. http://www.bayesx.org. Accessed: 2017-02-19. 23

Besag J, York J, and Mollié A (1991). Bayesian image restoration, with two applications in spatial statistics. *Annals of the Institute of Statistical Mathematics*, 43(1):1–20. 10, 88, 89, 113

Biau G, Cérou F, and Guyader A (2013). New insights into approximate Bayesian computation. *Annales de l'Institut Henri Poincaré, Probabilités et Statistiques*, 51(1):376–403. 24

Botella-Rocamora P, López-Quílez A, and Martinez-Beneito M. A (2013). Spatial moving average risk smoothing. *Statistics in Medicine*, 32(15):2595–2612. 89, 91

Box G. E. P, Jenkins G. M, and Reinsel G. C (1994). *Time series analysis: Forecasting & control (3rd edition)*. Prentice Hall. 54, 55

Boyle J. R, Sparks R. S, Keijzers G. B, Crilly J. L, Lind J. F, and Ryan L. M (2011). Prediction and surveillance of influenza epidemics. *Medical Journal of Australia*, 194(4):S28–33. 14, 17, 130, 144, 153

Broniatowski D. A, Paul M. J, and Dredze M (2013). National and local influenza surveillance through Twitter: An analysis of the 2012-2013 influenza epidemic. *PloS ONE*, 8(12):e83672. 9

Buckeridge D. L, Burkom H, Campbell M, Hogan W. R, and Moore A. W (2005). Algorithms for rapid outbreak detection: A research synthesis. *Journal of Biomedical Informatics*, 38(2):99–113. 13

Buehler J. W, Berkelman R. L, Hartley D. M, and Peters C. J (2003). Syndromic surveillance and bioterrorism-related epidemics. *Emerging Infectious Diseases*, 9(10):1197–1204. 3

Burkom H (**2007**). Alerting algorithms for biosurveillance. In Lombardo J. S and Buckeridge D. L, editors, *Disease Surveillance: A Public Health Informatics Approach*, chapter 4, pages 143–192. John Wiley & Sons, Inc. 13

Burnham K. P and Anderson D. R (**1998**). *Model selection and inference: A practical information-theoretic approach*. Springer, New York. 146

Cappé O, Moulines E, and Rydén T (**2005**). *Inference in hidden Markov models*. Springer, New York. 26

Carpenter B, Gelman A, Hoffman M. D, Lee D, Goodrich B, Bentancourt M, Brubaker M, Guo J, and Riddell A (**2017**). Stan: A probabilistic programming language. *Journal of Statistical Software*, 76(1). 23

Carrat F and Valleron A. J (**1992**). Epidemiologic mapping using the "kriging" method: Application to an influenza-like illness epidemic in France. *American Journal of Epidemiology*, 135(11):1293–300. 10

Chan T. C, Teng Y. C, and Hwang J. S (**2015**). Detection of influenza-like illness aberrations by directly monitoring Pearson residuals of fitted negative binomial regression models. *BMC Public Health*, 15:168. 14

Charland K. M, Buckeridge D. L, Sturtevant J. L, Melton F, Reis B. Y, Mandl K. D, and Brownstein J. S (**2009**). Effect of environmental factors on the spatio-temporal patterns of influenza spread. *Epidemiology and Infection*, 137(10):1377–1387. 24, 140

Cheng C. K. Y, Lau E. H. Y, Ip D. K. M, Yeung A. S. Y, Ho L. M, and Cowling B. J (**2009**). A profile of the online dissemination of national influenza surveillance data. *BMC Public Health*, 9:339. 4, 5

Cheng C. K. Y, Cowling B. J, Lau E, Ho L, Leung G. M, and Ip D (**2012**). Electronic school absenteeism monitoring and influenza

surveillance, Hong Kong. *Emerging Infectious Diseases*, 18(5):885–887. 7

Choi K and Thacker S. B (**1981**). An evaluation of influenza mortality surveillance, 1962-1979. I. Time series forecasts of expected pneumonia and influenza deaths. *American Journal of Epidemiology*, 113(3):215–226. 18

Closas P, Coma E, and Méndez L (**2012**). Sequential detection of influenza epidemics by the Kolmogorov-Smirnov test. *BMC Medical Informatics and Decision Making*, 12:112. 15

Conesa D, López-Quílez A, Martinez-Beneito M. A, Miralles M. T, and Verdejo F (**2009**). FluDetWeb: An interactive web-based system for the early detection of the onset of influenza epidemics. *BMC Medical Informatics and Decision Making*, 9:36. 40

Conesa D, Martinez-Beneito M. A, Amorós R, and López-Quílez A (**2015**). Bayesian hierarchical Poisson models with a hidden Markov structure for the detection of influenza epidemic outbreaks. *Statistical Methods in Medical Research*, 24(2):206–223. xiv, 41, 160

Cook S, Conrad C, Fowlkes A. L, and Mohebbi M. H (**2011**). Assessing Google Flu Trends performance in the United States during the 2009 influenza virus A (H1N1) pandemic. *PLOS ONE*, 6(8):e23610. 2, 9

Cooper G. F, Dash D. H, Levander J. D, Wong W. K, Hogan W. R, and Wagner M. M (**2004**). Bayesian biosurveillance of disease outbreaks. In *UAI'04: Proceedings of the 20th Conference on Uncertainty in Artificial Intelligence*, pages 94–103. 25

Corberán-Vallet A and Lawson A. B (**2014**). Prospective analysis of infectious disease surveillance data using syndromic information. *Statistical Methods in Medical Research*, 23(6):572–590. 6, 97

Costagliola D (**1994**). When is the epidemic warning cut-off point exceeded? *European Journal of Epidemiology*, 10(4):475–476. 14, 15

Costagliola D, Flahault A, Galinec D, Garnerin P, Menares J, Valleron A. J, and Gamenn P (**1991**). A routine tool for detection and assessment of epidemics of influenza-like syndromes in France. *American Journal of Public Health*, 81(1):97–99. 14, 15

Cowling B. J, Wong I. O. L, Ho L. M, Riley S, and Leung G. M (**2006**). Methods for monitoring influenza surveillance data. *International Journal of Epidemiology*, 35(5):1314–1321. 15, 19, 63, 64, 70, 75, 147, 150

Cressie N and Mugglin A. S (**2000**). Spatio-temporal hierarchical modeling of an infectious disease from (simulated) count data. In *CoMPSTAT: Proceedings in computational statistics*, pages 41–52. 97

Crosier R. B (**1988**). Multivariate generalizations of cumulative sum quality-control schemes. *Technometrics*, 30(3):291–303. 94

De Lange M. M. A, Meijer A, Friesema I. H. M, Donker G. A, Koppeschaar C. E, Hooiveld M, Ruigrok N, and van der Hoek W (**2013**). Comparison of five influenza surveillance systems during the 2009 pandemic and their association with media attention. *BMC Public Health*, 13:881. 2

Debin M, Turbelin C, Blanchon T, Bonmarin I, Falchi A, Hanslik T, Levy-Bruhl D, Poletto C, and Colizza V (**2013**). Evaluating the feasibility and participants' representativeness of an online nationwide surveillance system for influenza in France. *PloS one*, 8(9):e73675. 8

Del Moral P (**1996**). Nonlinear filtering: Interacting particle resolution. *Markov Processes and Related Fields*, 2(4):555–580. 23

Douc R, Moulines É, and Rydén T (**2004**). Asymptotic properties of the maximum likelihood estimator in autoregressive models with Markov regime. *The Annals of Statistics*, 32(5):2254–2304. 26

Egan J. P (**1975**). *Signal detection theory and ROC analysis*. Academic Press, Inc., New York. 62, 149

Farrington C. P, Andrews N. J, Beale A. D, and Catchpole M. A (**1996**). A statistical algorithm for the early detection of outbreaks of infectious disease. *Journal of the Royal Statistical Society: Series A (Statistics in Society)*, 159(3):547–563. 14

FluDetWeb. http://www.geeitema.org/meviepi/fludetweb/. Accessed: 2017-02-19. 40

Fox E. B and Dunson D. B (**2015**). Bayesian nonparametric covariance regression. *Journal of Machine Learning Research*, 16:2501–2542. 112, 140

Fricker R. D Jr., Hegler B. J, and Dunfee D. A (**2008**). Comparing syndromic surveillance detection methods: EARS' versus a CUSUM-based methodology. *Statistics in Medicine*, 27(17):3407–3429. 14

Frisén M (**2003**). Statistical surveillance. Optimality and methods. *International Statistical Review*, 71(2):403–434. 20

Frisén M and Andersson E (**2009**). Semiparametric surveillance of monotonic changes. *Sequential Analysis*, 28(4):434–454. 20, 144, 148, 149

Frisén M, Andersson E, and Pettersson K (**2010**). Semiparametric estimation of outbreak regression. *Statistics*, 44(2):107–117. 20

Gasparini R, Amicizia D, Lai P. L, and Panatto D (**2012**). Clinical and socioeconomic impact of seasonal and pandemic influenza in adults and the elderly. *Human Vaccines & Immunotherapeutics*, 8(1):21–28. 3

Gelfand A. E and Vounatsou P (**2003**). Proper multivariate conditional autoregressive models for spatial data analysis. *Biostatistics*, 4(1):11–25. 91

Gelman A, Carlin J. B, Stern H. S, Dunson D. B, Vehtari A, and Rubin D. B (**2013**). *Bayesian data analysis (3rd edition)*. Chapman and Hall/CRC. 61, 126, 145

Gentle J. E (**2007**). *Matrix algebra: Theory, computations, and applications in statistics*. Springer, New York. 89

Gesualdo F, Stilo G, Agricola E, Gonfiantini M. V, Pandolfi E, Velardi P, and Tozzi A. E (**2013**). Influenza-like illness surveillance on Twitter through automated learning of naïve language. *PloS ONE*, 8 (12):e82489. 9

Gilks W. R, Richardson S, and Spiegelhalter D. J (**1996**). *Markov chain Monte Carlo in practice*. Chapman & Hall/CRC. 23

Ginsberg J, Mohebbi M. H, Patel R. S, Brammer L, Smolinski M. S, and Brilliant L (**2009**). Detecting influenza epidemics using search engine query data. *Nature*, 457:1012–1014. 9

Gneiting T and Raftery A. E (**2007**). Strictly proper scoring rules, prediction, and estimation. *Journal of the American Statistical Association*, 102(477):359–378. 130, 153

Goddard N. L, Kyncl J, and Watson J. M (**2003**). Appropriateness of thresholds currently used to describe influenza activity in England. *Communicable Disease and Public Health / PHLS*, 6(3):238–245. 16

Goldenberg A, Shmueli G, Caruana R. A, and Fienberg S. E (**2002**). Early statistical detection of anthrax outbreaks by tracking over-the-counter medication sales. *Proceedings of the National Academy of Sciences of the United States of America*, 99(8):5237–5240. 8

Gomez-Barroso D, Martinez-Beneito M. A, Flores V, Amorós R, Delgado C, Botella P, Zurriaga O, and Larrauri A (**2014**). Geographical spread of influenza incidence in Spain during the 2009 A(H1N1) pandemic wave and the two succeeding influenza seasons. *Epidemiology and Infection*, 142(12):2629–2641. 2

Google Flu Trends. http://www.google.org/flutrends/about/. Accessed: 2017-02-19. 9, 43, 78, 87, 104, 105, 109, 112, 116, 163, 180

Google Trends. https://www.google.com/trends/. Accessed: 2017-02-19. 8

Griffin B. A, Jain A. K, Davies-Cole J, Glymph C, Lum G, Washington S. C, and Stoto M. A (**2009**). Early detection of influenza outbreaks using the DC Department of Health's syndromic surveillance system. *BMC Public Health*, 9:483. 17

Grover S and Aujla G. S (**2014**). Prediction model for influenza epidemic based on Twitter data. *International Journal of Advanced Research in Computer and Communication Engineering*, 3(7):7541–7545. 9

Heaton M. J, Banks D. L, Zou J, Karr A. F, Datta G, Lynch J, and Vera F (**2012**). A spatio-temporal absorbing state model for disease and syndromic surveillance. *Statistics in Medicine*, 31(19):2123–2136. 101, 102, 113, 139, 148, 149, 157

Held L and Paul M (**2012**). Modeling seasonality in space-time infectious disease surveillance data. *Biometrical Journal*, 54(6):824–843. 93

Held L, Hofmann M, Höhle M, and Schmid V (**2006**). A two-component model for counts of infectious diseases. *Biostatistics*, 7(3): 422–437. 25

Hersbach H (**2000**). Decomposition of the continuous ranked probability score for ensemble prediction systems. *Weather and Forecasting*, 15(5): 559–570. 154

Höhle M (**2007**). `surveillance`: An R package for the monitoring of infectious diseases. *Computational Statistics*, 22(4):571–582. 13

Höhle M (**2010**). Online change-point detection in categorical time series. In Kneib T and Tutz G, editors, *Statistical Modelling and Regression Structures*, pages 377–397. Physica-Verlag HD. 17

Höhle M and Paul M (**2008**). Count data regression charts for the monitoring of surveillance time series. *Computational Statistics & Data Analysis*, 52(9):4357–4368. 17

Hu W, Zhang W, Huang X, Clements A, Mengersen K, and Tong S (**2015**). Weather variability and influenza A (H7N9) transmission in Shanghai, China: A Bayesian spatial analysis. *Environmental Research*, 136:405–412. 10

Influenzanet. A network of European citizens fighting against influenza. https://www.influenzanet.eu/. Accessed: 2017-02-19. 8

Ismail N. A, Pettitt A. N, and Webster R. A (**2003**). 'Online' monitoring and retrospective analysis of hospital outcomes based on a scan statistic. *Statistics in Medicine*, 22(18):2861–2876. 21, 95

Jiang X and Cooper G. F (**2010**). A Bayesian spatio-temporal method for disease outbreak detection. *Journal of the American Medical Informatics Association*, 17(4):462–471. 150

Kavanagh K, Robertson C, Murdoch H, Crooks G, and McMenamin J (**2012**). Syndromic surveillance of influenza-like illness in Scotland during the influenza A H1N1v pandemic and beyond. *Journal of the Royal Statistical Society: Series A (Statistics in Society)*, 175(4): 939–958. 2

Kleinman K. P and Abrams A. M (**2006**). Assessing surveillance using sensitivity, specificity and timeliness. *Statistical Methods in Medical Research*, 15(5):445–464. 39, 63, 150, 151

Knorr-Held L and Richardson S (**2003**). A hierarchical model for space-time surveillance data on meningococcal disease incidence. *Journal of the Royal Statistical Society: Series C (Applied Statistics)*, 52(2):169–183. 47, 99, 144, 145

Kom Mogto C. A, De Serres G, Douville Fradet M, Lebel G, Toutant S, Gilca R, Ouakki M, Janjua N. Z, and Skowronski

D. M (**2012**). School absenteeism as an adjunct surveillance indicator: Experience during the second wave of the 2009 H1N1 pandemic in Quebec, Canada. *PLOS ONE*, 7(3):e34084. 7

Krige D. G (**1951**). *A statistical approach to some mine valuation and allied problems on the Witwatersrand*. Master Thesis. 10

Kulldorff M (**1997**). A spatial scan statistic. *Communications in Statistics - Theory and Methods*, 26(6):1481–1496. 95

Kulldorff M (**2001**). Prospective time periodic geographical disease surveillance using a scan statistic. *Journal of the Royal Statistical Society: Series A (Statistics in Society)*, 164(1):61–72. 95, 144

Kulldorff M, Heffernan R, Hartman J, Assunção R, and Mostashari F (**2005**). A space-time permutation scan statistic for disease outbreak detection. *PLOS medicine*, 2(3):e59. 11, 95

Kulldorff M, Huang L, Pickle L, and Duczmal L (**2006**). An elliptic spatial scan statistic. *Statistics in Medicine*, 25(22):3929–3943. 95

Le Strat Y (**2005**). Overview of temporal surveillance. In Lawson A. B and Kleinman K, editors, *Spatial and Syndromic Surveillance for Public Health*, chapter 2, pages 13–29. John Wiley & Sons, Ltd. 13

Le Strat Y and Carrat F (**1999**). Monitoring epidemiologic surveillance data using hidden Markov models. *Statistics in Medicine*, 18(24):3463–3478. 29, 30, 39, 65, 66, 69, 76, 77

Leroux B. G, Lei X, and Breslow N (**2000**). Estimation of disease rates in small areas: A new mixed model for spatial dependence. In Halloran M. E and Berry D, editors, *Statistical Models in Epidemiology, the Environment, and Clinical Trials*, pages 179–191. Springer, New York. xxv, 90, 125, 138

Li G, Best N, Hansell A. L, Ahmed I, and Richardson S (**2012**). BaySTDetect: Detecting unusual temporal patterns in small area data

via Bayesian model choice. *Biostatistics*, 13(4):695–710. 11, 97, 145, 149, 160

Li J and Cardie C (**2013**). Early stage influenza detection from Twitter. *ArXiv*, 1309.7340v3. 9, 100

Liu T. Y, Sanders J. L, Tsui F. C, Espino J. U, Dato V. M, and Suyama J (**2013**). Association of over-the-counter pharmaceutical sales with influenza-like-illnesses to patient volume in an urgent care setting. *PLOS ONE*, 8(3):e59273. 8

Lu H, Zeng D, and Chen H (**2010**). Prospective infectious disease outbreak detection using Markov switching models. *IEEE Transactions on Knowledge and Data Engineering*, 22(4):565–577. 28, 31, 148, 150

Lunn D. J, Thomas A, Best N, and Spiegelhalter D (**2000**). WinBUGS - A Bayesian modelling framework: Concepts, structure, and extensibility. *Statistics and Computing*, 10(4):325–337. 23

Lusted L. B (**1971**). Signal detectability and medical decision-making. *Science*, 171(3977):1217–1219. 62, 149

Madigan D (**2005**). Bayesian data mining for health surveillance. In Lawson A. B and Kleinman K, editors, *Spatial & syndromic surveillance for public health*, chapter 12, pages 203–221. John Wiley & Sons, Inc., Chichester. 30, 31

Magruder S. F (**2003**). Evaluation of over-the-counter pharmaceutical sales as a possible early warning indicator of human disease. *Johns Hopkins Applied Physics Laboratory Technical Digest*, 24(4):349–353. 8

Manitz J and Höhle M (**2013**). Bayesian outbreak detection algorithm for monitoring reported cases of campylobacteriosis in Germany. *Biometrical Journal*, 55(4):509–526. 25, 26

Marshall E. C and Spiegelhalter D. J (**2003**). Approximate cross-validatory predictive checks in disease mapping models. *Statistics in Medicine*, 22(10):1649–1660. 130

Martinez-Beneito M. A, Conesa D, López-Quílez A, and López-Maside A (**2008**a). Bayesian Markov switching models for the early detection of influenza epidemics. *Statistics in Medicine*, 27(22):4455–4468. xiii, xiv, xvi, xxv, xxvi, xli, 1, 11, 34, 35, 37, 38, 39, 40, 41, 42, 44, 49, 50, 51, 57, 58, 65, 66, 67, 69, 74, 75, 76, 77, 81, 83, 85, 87, 103, 110, 113, 120, 124, 125, 126, 128, 129, 135, 138, 147, 180

Martinez-Beneito M. A, López-Quílez A, and Botella-Rocamora P (**2008**b). An autoregressive approach to spatio-temporal disease mapping. *Statistics in Medicine*, 27(15):2874–2889. 53, 111

McKenzie E (**1985**). Some simple models for discrete variate time series. *Journal of the American Water Resources Association*, 21(4):645–650. 19

Metz C. E (**1978**). Basic principles of ROC analysis. *Seminars in Nuclear Medicine*, 8(4):283–298. 62, 149

Meyer S, Held L, and Höhle M (**2014**). Spatio-temporal analysis of epidemic phenomena using the R package `surveillance`. *ArXiv*, 1411.0416v1. 93

Meynard J. B, Chaudet H, Texier G, Ardillon V, Ravachol F, Deparis X, Jefferson H, Dussart P, Morvan J, and Boutin J.-P (**2008**). Value of syndromic surveillance within the Armed Forces for early warning during a dengue fever outbreak in French Guiana in 2006. *BMC Medical Informatics and Decision Making*, 8:29. 17

Milinovich G. J, Williams G. M, Clements A. C. A, and Hu W (**2014**). Internet-based surveillance systems for monitoring emerging infectious diseases. *The Lancet Infectious Diseases*, 14(2):160–168. 9

Mugglin A. S, Cressie N, and Gemmell I (**2002**). Hierarchical statistical modelling of influenza epidemic dynamics in space and time. *Statistics in Medicine*, 21(18):2703–2721. 96

Mulpuru S, Smith T, Lawrence N, Wilson K, and Forster A. J (**2013**). Evaluation of 3 electronic methods used to detect influenza diagnoses during 2009 pandemic. *Emerging Infectious Diseases*, 19(12): 2062–2063. 2

Muscatello D. J, Morton P. M, Evans I, and Gilmour R (**2008**). Prospective surveillance of excess mortality due to influenza in New South Wales: feasibility and statistical approach. *Communicable Diseases Intelligence Quarterly Report*, 32(4):435–442. 14, 15

Naus J and Wallenstein S (**2006**). Temporal surveillance using scan statistics. *Statistics in Medicine*, 25(2):311–324. 21

Naus J. I (**1965**a). The distribution of the size of the maximum cluster of points on a line. *Journal of the American Statistical Association*, 60 (310):532–538. 11, 21, 95

Naus J. I (**1965**b). Clustering of random points in two dimensions. *Biometrika*, 52(1/2):263–267. 11, 95

Navarro-Marí J. M, Pérez-Ruiz M, Cantudo-Muñoz P, Petit-Gancedo C, Jiménez-Valera M, and Rosa-Fraile M (**2005**). Influenza-like illness criteria were poorly related to laboratory-confirmed influenza in a sentinel surveillance study. *Journal of Clinical Epidemiology*, 58(3):275–279. 3

NC DHHS Influenza (Flu) Information. http://www.flu.nc.gov/. Accessed: 2017-02-19. 59

Neill D. B and Cooper G. F (**2010**). A multivariate Bayesian scan statistic for early event detection and characterization. *Machine Learning*, 79(3):261–282. 95

Neill D. B, Moore A. W, and Cooper G. F (**2006**). A Bayesian spatial scan statistic. In *Advances in neural information processing systems 18 (proceedings of the Neural Information Processing Systems 2005)*, page 1003. 95

Niemi J, Smith M, and Banks D (**2008**). Test power for drug abuse surveillance. In *Proceedings of the 2008 International Workshop on Biosurveillance and Biosecurity*, pages 131–142. Springer. 100

NIMBLE. An R package for programming with BUGS models and compiling parts of R. http://r-nimble.org. Accessed: 2017-02-19. 23

Noufaily A, Enki D. G, Farrington P, Garthwaite P, Andrews N, and Charlett A (**2013**). An improved algorithm for outbreak detection in multiple surveillance systems. *Statistics in Medicine*, 32(7):1206–1222. 14, 26

Nunes B, Natário I, and Carvalho M. L (**2013**). Nowcasting influenza epidemics using non-homogeneous hidden Markov models. *Statistics in Medicine*, 32(15):2643–2660. 7, 32, 49, 113, 139, 148

Olson D. R, Konty K. J, Paladini M, Viboud C, and Simonsen L (**2013**). Reassessing Google Flu Trends data for detection of seasonal and pandemic influenza: A comparative epidemiological study at three geographic scales. *PLOS Computational Biology*, 9(10):e1003256. 9

Ortiz J. R, Sotomayor V, Uez O. C, Oliva O, Bettels D, McCarron M, Bresee J. S, and Mounts A. W (**2009**). Strategy to enhance influenza surveillance worldwide. *Emerging Infectious Diseases*, 15(8): 1271–1278. 2

Page E. S (**1954**). Continuous inspection schemes. *Biometrika*, 41(1/2): 100–115. 17

Paroli R and Spezia L (**2008**). Bayesian inference in non-homogeneous Markov mixtures of periodic autoregressions with state-dependent exogenous variables. *Computational Statistics & Data Analysis*, 52(5):2311–2330. 32

Paterson B, Caddis R, and Durrheim D (**2011**). Use of workplace absenteeism surveillance data for outbreak detection. *Emerging infectious diseases*, 17(10):1963–1964. 7

Paterson B. J and Durrheim D. N (**2013**). The remarkable adaptability of syndromic surveillance to meet public health needs. *Journal of Epidemiology and Global Health*, 3(1):41–47. 3

Pignatiello J. J Jr. and Runger G. C (**1990**). Comparisons of multivariate CUSUM charts. *Journal of Quality Technology*, 22(3):173–186. 94

Plummer M (**2003**). JAGS: A program for analysis of Bayesian graphical models using Gibbs sampling. In *Proceedings of the 3rd International Workshop on Distributed Statistical Computing*. 23

R Core Team (**2016**). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing, Vienna, Austria. URL `https://www.R-project.org`. 23

R-INLA. Bayesian computing with INLA! `http://www.r-inla.org/`. Accessed: 2017-02-19. 24

Rafei A, Pasha E, and Jamshidi Orak R (**2012**). Tuberculosis surveillance using a hidden Markov model. *Iranian J Publ Health*, 41(10):87–96. 29, 145, 148, 149

Rafei A, Pasha E, and Jamshidi Orak R (**2015**). A warning threshold for monitoring tuberculosis surveillance data: an alternative to hidden Markov model. *Tropical Medicine & International Health*, 20(7):919–929. 31, 145, 149

Rambaut A, Pybus O. G, Nelson M. I, Viboud C, Taubenberger J. K, Holmes E. C, Jeffery K, Politics A, and Reform H. I (**2009**). The genomic and epidemiological dynamics of human influenza A virus. *Nature*, 453:615–619. 2

Rao Y and McCabe B (**2016**). Real-time surveillance for abnormal events: The case of influenza outbreaks. *Statistics in Medicine*, 35(13): 2206–2220. 19, 148, 149

Rath T. M, Carreras M, and Sebastiani P (**2003**). Automated detection of influenza epidemics with hidden Markov hodels. In Berthold M. R, Lenz H, Bradley E, Kruse R, and Borgelt C, editors, *Advances in intelligent data analysis V*, pages 521–532. Springer, Berlin. 15, 30, 31, 148, 149

Red Centinela Sanitaria. http://www.sp.san.gva.es/red_centinela/. Accessed: 2017-02-19. 73

Reis B. Y and Mandl K. D (**2003**). Time series modeling for syndromic surveillance. *BMC Medical Informatics and Decision Making*, 3:2. 19

Riebler A, Sørbye S. H, Simpson D, and Rue H (**2016**). An intuitive Bayesian spatial model for disease mapping that accounts for scaling. *Statistical Methods in Medical Research*, 25(4):1145–1165. 90

Robert C. P, Rydén T, and Titterington D. M (**2000**). Bayesian inference in hidden Markov models through the reversible jump Markov chain Monte Carlo method. *Journal of the Royal Statistical Society, Series B (Statistical Methodology)*, 62:57–75. 48

Roberts S. W (**1959**). Control charts tests based on geometric moving averages. *Technometrics*, 1(3):239–250. 17

Rogerson P. A and Yamada I (**2004**). Monitoring change in spatial patterns of disease: Comparing univariate and multivariate cumulative sum approaches. *Statistics in Medicine*, 23(14):2195–2214. 93

Rue H, Martino S, and Chopin N (**2009**). Approximate Bayesian inference for latent Gaussian models by using integrated nested Laplace approximations. *Journal of the Royal Statistical Society. Series B: (Statistical Methodology)*, 71(2):319–392. 24

Salmon M, Schumacher D, Stark K, and Höhle M (**2015**). Bayesian outbreak detection in the presence of reporting delays. *Biometrical Journal*, 57(6):1051–1067. 26, 144, 148, 149

Salmon M, Schumacher D, and Höhle M (**2016**). Monitoring count time series in R: Aberration detection in public health surveillance. *Journal of Statistical Software*, 70(10). 13

SaTScan™. Software for the spatial, temporal and space-time scan statistics. http://satscan.org/. Accessed: 2017-02-19. 11, 95

Schwarz G (**1978**). Estimating the dimension of a model. *The Annals of Statistics*, 6(2):461–464. 30, 144

Sebastiani P, Mandl K. D, Szolovits P, Kohane I. S, and Ramoni M. F (**2006**). A Bayesian dynamic model for influenza surveillance. *Statistics in Medicine*, 25(11):1803–1825. 6, 25

Serfling R. E (**1963**). Methods for current statistical analysis of excess pneumonia-influenza deaths. *Public Health Reports*, 78(6):494–506. 6, 14, 25, 29, 31, 39, 41, 65, 66, 69, 70, 77, 93, 148

Shewhart W. A (**1931**). *Economic control of quality of manufactured product*. MacMillan and Co., London. 14, 16

Simonsen L, Clarke M. J, Williamson G. D, Stroup D. F, Arden N. H, and Schonberger L. B (**1997**). The impact of influenza epidemics on mortality: Introducing a severity index. *American Journal of Public Health*, 87(12):1944–1950. 14

Sparks R. S, Keighley T, and Muscatello D (**2010**). Early warning CUSUM plans for surveillance of negative binomial daily disease counts. *Journal of Applied Statistics*, 37(11):1911–1929. 17

Spiegelhalter D. J, Best N. G, Carlin B. P, and van der Linde A (**2002**). Bayesian measures of model complexity and fit. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 64(4): 583–639. 31, 38, 144, 146

Spiegelhalter D. J, Thomas A, Best N, and Lunn D. J (**2003**). *WinBUGS user manual, version 1.4*. MRC Biostatistics Unit. 37

Spreco A and Timpka T (**2016**). Algorithms for detecting and predicting influenza outbreaks: Metanarrative review of prospective evaluations. *BMJ Open*, 6(5):e010683. 13

Stroup D. F, Williamson G. D, Herndon J. L, and Karon J. M (**1989**). Detection of aberrations in the occurrence of notifiable diseases surveillance data. *Statistics in Medicine*, 8(3):323–329. 14, 39, 65, 66, 67, 69, 77

Stroustrup B (**2013**). *The C++ programming language (4th edition)*. Addison-Wesley. 23

Sun W and Cai T. T (**2009**). Large-scale multiple testing under dependence. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 71(2):393–424. 30, 145

Sun X, Ye J, Ren F, Labora A. P. K, and Intelligent A (**2014**). Real time early-stage influenza detection with emotion factors from Sina microblog. In *Proceedings of the Fifth Workshop on South and Southeast Asian Natural Language Processing*, pages 80–84. 100

Takahashi K, Kulldorff M, Tango T, and Yih K (**2008**). A flexibly shaped space-time scan statistic for disease outbreak detection and monitoring. *International Journal of Health Geographics*, 7:14. 95

Tang J. W, Lai F. Y. L, Nymadawa P, Deng Y, Ratnamohan M, Petric M, Loh T. P, Tee N. W. S, Dwyer D. E, Barr I. G, and Wong F. Y. W (**2010**). Comparison of the incidence of influenza in relation to climate factors during 2000–2007 in five countries. *Journal of Medical Virology*, 82(11):1958–1965. 2, 140

Tilston N. L, Eames K. T. D, Paolotti D, Ealden T, and Edmunds W. J (**2010**). Internet-based surveillance of Influenza-like-illness in the UK during the 2009 H1N1 influenza pandemic. *BMC Public Health*, 10 (1):650. 2

Unkel S, Farrington C. P, Garthwaite P. H, Robertson C, and Andrews N (**2012**). Statistical methods for the prospective detection of infectious disease outbreaks: A review. *Journal of the Royal Statistical Society: Series A (Statistics in Society)*, 175(1):49–82. 13

Vandendijck Y, Faes C, and Hens N (**2013**). Eight years of the Great Influenza Survey to monitor influenza-like illness in Flanders. *PLOS ONE*, 8(5):e64156. 8

Vega T, Lozano J. E, Meerhoff T, Snacken R, Mott J, Ortiz de Lejarazu R, and Nunes B (**2013**). Influenza surveillance in Europe: Establishing epidemic thresholds by the Moving Epidemic Method. *Influenza and Other Respiratory Viruses*, 7(4):546–558. 15

Visser I (**2007**). `depmix`: An R-package for fitting mixture models on mixed multivariate data with Markov dependencies. *R-package manual*. 39, 65

Viterbi A (**1967**). Error bounds for convolutional codes and an asymptotically optimum decoding algorithm. *IEEE Transactions on Information Theory*, 13(2):260–269. 28

Watkins R. E, Eagleson S, Hall R. G, Dailey L, and Plant A. J (**2006**). Approaches to the evaluation of outbreak detection methods. *BMC Public Health*, 6:263. 143

Wieland S. C, Brownstein J. S, Berger B, and Mandl K. D (**2007**). Automated real time constant-specificity surveillance for disease outbreaks. *BMC Medical Informatics and Decision Making*, 7:15. 6

Williamson G. D and Hudson G. W (**1999**). A monitoring system for detecting aberrations in public health surveillance reports. *Statistics in Medicine*, 18(23):3283–3298. 19

Woodall W. H and Ncube M. M (**1985**). Multivariate CUSUM quality-control procedures. *Technometrics*, 27(3):285–292. 93

Woodall W. H, Mohammed M. A, Lucas J. M, Watkins R, Steiner S. H, Benneyan J. C, Grigg O. A, Spiegelhalter D. J, and Burkom H (**2006**). The use of control charts in health-care and public-health surveillance (With Discussion and Rejoinder). *Journal of Quality Technology*, 38(2):89–104. 17

World Health Organization. Influenza (seasonal). Fact sheet n°211. http://www.who.int/mediacentre/factsheets/fs211/en/. Accessed: 2016-02-12. 3, 4

World Health Organization. Health topics. public health surveillance. http://www.who.int/topics/public_health_surveillance/en/. Accessed: 2017-02-12. 4

Yang W, Cowling B. J, Lau E. H. Y, and Shaman J (**2015**). Forecasting influenza epidemics in Hong Kong. *PLOS Computational Biology*, 11(7):e1004383. 12

Zhou H and Lawson A. B (**2008**). EWMA smoothing and Bayesian spatial modeling for health surveillance. *Statistics in Medicine*, 27(28): 5907–5928. 94

Zhu L and Carlin B. P (**2000**). Comparing hierarchical models for spatio-temporally misaligned data using the deviance information criterion. *Statistics in Medicine*, 19(17/18):2265–2278. 146

Zou J, Karr A. F, Banks D, Heaton M. J, Datta G, Lynch J, and Vera F (**2012**). Bayesian methodology for the analysis of spatial-temporal surveillance data. *Statistical Analysis and Data Mining*, 5(3): 194–204. 101, 102, 113, 139, 148, 149, 150, 157

Zou J, Karr A. F, Datta G, Lynch J, and Grannis S (**2014**). A Bayesian spatio-temporal approach for real-time detection of disease outbreaks: a case study. *BMC Medical Informatics and Decision Making*, 14:108. 102, 103, 144