Zero-state coupled Markov switching count models for spatio-temporal infectious disease spread

Dirk Douwes-Schultz | Alexandra M. Schmidt

Abstract
Spatio-temporal counts of infectious disease cases often contain an excess of zeros. With existing zero-inflated count models applied to such data it is difficult to quantify space-time heterogeneity in the effects of disease spread between areas. Also, existing methods do not allow for separate dynamics to affect the reemergence and persistence of the disease. As an alternative, we develop a new zero-state coupled Markov switching negative binomial model, under which the disease switches between periods of presence and absence in each area through a series of partially hidden nonhomogeneous Markov chains coupled between neighbouring locations. When the disease is present, an autoregressive negative binomial model generates the cases with a possible zero representing the disease being undetected. Bayesian inference and prediction is illustrated using spatio-temporal counts of dengue fever cases in Rio de Janeiro, Brazil.

KEYWORDS
Bayesian paradigm, dengue fever, forward filtering backward sampling, hidden Markov model, state space model, zero-inflation

1 | INTRODUCTION
In epidemiology, counts of infectious disease cases are being increasingly reported in several related areal units across time. A common issue encountered when modelling these counts
is the presence of excess zeros (Arab, 2015). That is, there are often many more zeros in the
counts than can be predicted by the usual Poisson and negative binomial count models. Lambert (1992)
proposed the zero-inflated Poisson (ZIP) model to deal with excess zeros in count
data, and this approach, with various extensions (see Young et al., 2020), has been applied in
many fields including epidemiology. We will refer to any model that incorporates zero-inflation
into a count distribution as a zero-inflated count (ZIC) model (e.g. ZIP, ZINB, ZIGP, ZICMP
models, etc. (Young et al., 2020)). In a disease mapping application of the ZIC model, the
presence/absence of the disease is generated through a Bernoulli process and then, when the
disease is present, the number of reported cases is generated through a count process, typically
negative binomial or Poisson (Fernandes et al., 2009). A zero coming from the count process
represents the disease being undetected, while a zero from the Bernoulli process represents the
true absence of the disease. A ZIC model can help account for the excess zeros, through the
additional Bernoulli zero generation, and also allows us to investigate factors related to the presen-
tce/absence of the disease while considering imperfect detection of disease presence (Vergne
etal., 2016). ZIC models have become increasingly popular in epidemiology for modelling excess
zeros in spatio-temporal infectious disease counts (Aktekin & Musal, 2015; Fernandes et al., 2009;
Wangdi et al., 2018).

However, in a spatio-temporal setting, an important issue with existing ZIC approaches is
that they only model the probability of overall disease presence. In a spatio-temporal setting we
can separate the presence of a disease into two events of epidemiological interest, persistence
(presence to presence) and reemergence (absence to presence). Some important epidemiological
covariates can affect the reemergence and persistence of an infectious disease quite differently.
An example from our application of dengue fever is temperature, as temperature affects both the
hatching rate of vertically infected mosquito eggs, which is more important for the reemergence
of the disease, and the life cycle of the infected mosquitoes, which is more important for the
persistence of the disease (Coutinho et al., 2006). A second issue concerns the use of random effects in the Bernoulli process of a ZIC model,
which can be used to account for spatio-temporal correlations in the presence of the disease
(Giorgi et al., 2018; Hoef & Jansen, 2007; Torabi, 2017). Splines can also be used for this pur-
pose (Ghosal et al., 2020), but are functionally similar so we will refer in general to random effect
models. Random effect models are built by specifying relationships between probabilities of dis-
 ease presence between areas and across time (Torabi, 2017). In contrast, it is more traditional in
epidemiology to build models of infectious disease spread by specifying the probability that the
disease will be present in an area given its actual presence in the area and neighbouring areas
previously (Hooten & Wikle, 2010; Keeling et al., 2001; Smith et al., 2002). That is, random effects
models condition on the probability of disease presence in neighbouring areas while epidemi-
ological models usually condition on the actual presence of the disease in neighbouring areas. If a
goal of the analysis is to quantify associations between various space-time factors and the effects
of disease spread between areas, like in Smith et al. (2002) who examined whether rivers block
the spread of rabies, then the traditional epidemiological approach is more appropriate. This is
because the disease can only spread from an area if it is present there, while the probability of dis-
ease presence does not tell you if the disease is spreading or not. For example, the disease could
be present in a neighbouring area and spread from there even if its probability of presence is
low. Therefore, it is difficult to quantify space-time heterogeneity in the effects of disease spread
between areas with existing ZIC models.

To help address these issues we propose a zero-state Markov switching approach to mod-
elling the zero inflation. Markov switching models, first introduced by Goldfeld and Quandt
(1973) and extensively developed by Hamilton (1989), allow a time series to be described by several submodels (states) where switching between submodels is governed by a first-order Markov chain. The motivation behind the Markov switching model, over a finite mixture model, is that the states are often dependent across time and occur consecutively (Goldfeld & Quandt, 1973). Markov switching models that switch between a zero state and a count state were considered in a temporal setting by Wang (2001) and in a spatio-temporal setting by Malyshkina and Mannering (2010). However, Malyshkina and Mannering (2010) assumed the states were independent between spatial units and neither of these applications dealt with infectious disease data.

In our framework, following traditional disease mapping ZIC models (Fernandes et al., 2009), we assume the existence of a zero state, representing the absence of the disease, and a negative binomial state, representing the presence of the disease. Then we allow the disease to switch between the presence and absence state in each area through a nonhomogeneous (Diebold et al., 1994) Markov chain. The Markov chains allow for switching between long periods of disease presence and absence, and allow for each covariate to have a separate effect on the reemergence of the disease compared to the persistence. As the disease can spread between areas, we extend the zero-state Markov switching count model to a coupled Markov switching model (Pohle et al., 2021), by coupling the Markov chains between neighbouring areas. In a coupled Markov switching model, the transition probabilities of a Markov chain can depend on the states of other Markov chains. We use a collection of coupling parameters to account for heterogeneity in the effects of neighbouring disease spread across space and time. The coupling parameters allow us to quantify associations between the effects of neighbouring disease spread and various space-time factors related to either area.

This paper is structured as follows. In Section 1.1 we introduce our motivating example of dengue fever cases in Rio de Janeiro and lay out the goals for our analysis. In Section 2 we introduce our proposed model, the zero-state coupled Markov switching negative binomial model. In Section 3 we describe Bayesian inference using data augmentation through Markov chain Monte Carlo (MCMC) methods. Here we discuss efficient sampling strategies for the parameters and unknown state indicators. In Section 4 we apply the model to the Rio’s dengue fever data. Finally, we close with a general discussion in Section 5.

1.1 Motivating example: dengue fever in Rio de Janeiro

Dengue fever is endemic in Rio de Janeiro, Brazil, and there have been several major epidemics there since 1987. We focus on modelling monthly cases of dengue fever in the 160 districts of Rio de Janeiro between 2011 and 2017, as reported by the Health Secretary of the city (http://www.rio.rj.gov.br/web/sms/exibeconteudo?id=2815389). Figure 1 shows monthly dengue fever cases for two districts, a relatively small district (a) and relatively large district (b). In the smaller district, there are both long periods of dengue presence (9 months longest period) and long periods of dengue absence (14 months longest period). Additionally, in the smaller district there is clearly a recurring seasonal pattern to the persistence and reemergence of the disease. The disease has a lower chance of persisting during the winter, where it often goes extinct, and then dengue often reemerges in the summer. This could be due to changes in rainfall and temperature, which often lead to large fluctuations in the mosquito population (Beaty & Marquardt, 1996). Additionally, temperature and rainfall play an important role in the life cycle of vertically infected eggs, which are important for the reemergence of the disease (Coutinho et al., 2006). The same
patters in disease presence are not seen in the larger district in Figure 1b, where dengue persists for much longer on average compared to the smaller district, rarely goes extinct and reemerges quickly. Figure 2 similarly shows that there is significant spatial variation in the district level average, across time, probabilities of dengue reemergence and persistence (although these are based on reported cases and do not account for imperfect detection of disease presence). This could be explained by differences in population between districts (Bartlett, 1957), but also possibly differences in socioeconomic factors as mosquitoes often use human made water containers to lay eggs (Schmidt et al., 2011).

A statistical model can help quantify some of the above patterns. Specifically, the goal of our analysis is to investigate how certain risk factors (such as rainfall, temperature, socioeconomic factors and population size) are related to the reemergence and persistence of the disease, and if any effects are different for the reemergence compared to the persistence. Additionally, as Stoddard et al. (2013) showed that many individuals are infected by dengue outside their homes, we want to quantify the risk of dengue spreading between neighbouring districts. We want to investigate whether certain space-time factors are associated with an increased risk of dengue spreading from a neighbouring district. The number of reported cases in the neighbouring district should be important in determining the risk of spread, but also, individuals are known to move more through high population areas (Grenfell et al., 2001). Finally, we wish to design some useful warning/forecasting systems for policy makers.
Average Monthly Probability of Dengue Reemergence

0.0 to 0.2
0.2 to 0.4
0.4 to 0.6
0.6 to 0.8
0.8 to 1.0
Never absent

Average Monthly Probability of Dengue Persistence

0.5 to 0.6
0.6 to 0.7
0.7 to 0.8
0.8 to 0.9
0.9 to 1.0
Never present

FIGURE 2 (top) Average monthly probability of dengue reemergence based on reported cases, that is,
\[
\frac{\#(\text{cases}=0 \rightarrow \text{cases}>0)}{\#(\text{cases}=0 \rightarrow \text{cases}>0)+\#(\text{cases}=0 \rightarrow \text{cases}=0)}
\]
(bottom) Average monthly probability of dengue persistence based on reported cases, that is,
\[
\frac{\#(\text{cases}>0 \rightarrow \text{cases}>0)+\#(\text{cases}>0 \rightarrow \text{cases}=0)}{\#(\text{cases}>0 \rightarrow \text{cases}>0)}
\]
[Colour figure can be viewed at wileyonlinelibrary.com]

2 | A ZERO-STATE COUPLED MARKOV SWITCHING NEGATIVE BINOMIAL MODEL

Assume we have areal data with \( i = 1, \ldots, N \) areas across \( t = 1, \ldots, T \) time periods. Let \( y_{it} \) be the reported case count from area \( i \) at time \( t \). Let \( S_{it} \) be a binary indicator for the true presence of the disease, meaning \( S_{it} = 1 \) if the disease is present in area \( i \) during time \( t \) and \( S_{it} = 0 \) if the disease is absent. Finally, let \( S_{(-i)(t-1)} = (S_{1(t-1)}, \ldots, S_{(i-1)(t-1)}, S_{(i+1)(t-1)}, \ldots, S_{N(t-1)})^T \) be the vector containing the state indicators of all areas excluding area \( i \) at time \( t - 1 \); and let \( y^{(t-1)} = (y_1, \ldots, y_{t-1})^T \) be the vector of all case counts up to time \( t - 1 \), where \( y_t = (y_{1t}, \ldots, y_{Nt})^T \).
We assume that when the disease is present the reported cases are generated by a negative binomial distribution and that when the disease is absent no cases will be reported, that is,

$$y_{it} | S_{it}, y^{t-1} \sim \begin{cases} 0, & \text{if } S_{it} = 0 \\ NB(\lambda_{it}, r_{it}), & \text{if } S_{it} = 1, \end{cases}$$

(1)

where $\lambda_{it}$ is the expected number of reported cases given the disease is present and $r_{it}$ is an overdispersion parameter such that $\text{Var} \ (y_{it} | S_{it} = 1, y^{t-1}) = \lambda_{it} (1 + \lambda_{it} / r_{it})$. Zero reported cases may arise when the disease is present due to the disease being undetected by the surveillance system. Therefore, when $y_{it} = 0$ we do not observe $S_{it}$ as the disease could be absent or present but undetected. In general we assume the following forms for $\lambda_{it}$ and $r_{it}$,

$$\lambda_{it} = g(\delta, x_{it}, y^{t-1}) \quad \text{and} \quad r_{it} = h(\gamma, w_{it}, y^{t-1}),$$

(2)

where $g(\cdot)$ and $h(\cdot)$ are positive valued functions, $\delta$ and $\gamma$ are vectors of unknown parameters (including possibly random effects), and $x_{it}$ and $w_{it}$ are vectors of covariates. Therefore, $\lambda_{it}$ and $r_{it}$ can depend on past values of the case counts, to possibly account for temporal autocorrelation due to transmission of the disease, but not on past values of the state indicators, which is a common assumption in Markov switching models that greatly simplifies model fitting (Frühwirth-Schnatter, 2006).

To model the switching between periods of disease presence and absence in area $i$, we assume that $S_{it}$ follows a two state nonhomogeneous Markov chain conditional on $S_{(i)(t-1)}$ and $y^{t-1}$. We propose the following conditional transition matrix for the Markov chain, for $t = 1, \ldots, T$,

$$\Gamma \left( S_{it} | S_{(i)(t-1)}, y^{(t-1)} \right) = \begin{bmatrix} s_{it=0 | 0 \text{ (absence)}} & s_{it=1 | 0 \text{ (absence)}} \\ s_{it=0 | 1 \text{ (presence)}} & s_{it=1 | 1 \text{ (presence)}} \end{bmatrix} = \begin{bmatrix} 1 - p01_{it} & p01_{it} \\ 1 - p11_{it} & p11_{it} \end{bmatrix},$$

(3)

where,

$$p01_{it} = P \left( S_{it} = 1 | S_{it-1} = 0, S_{(i)(t-1)}, y^{t-1} \right) \quad \text{(probability of disease reemergence)},$$

$$p11_{it} = P \left( S_{it} = 1 | S_{it-1} = 1, S_{(i)(t-1)}, y^{t-1} \right) \quad \text{(probability of disease persistence)}.$$

The probability of disease reemergence in area $i$ during time $t$, $p01_{it}$, is allowed to depend on a $K$-dimensional vector of space-time covariates $z_{it} = (z_{it1}, \ldots, z_{itK})^T$ as well as the presence of the disease in neighbouring areas during the previous time period,

$$\logit(p01_{it}) = \zeta_0 + z_{it}^T \zeta + \sum_{j \in \text{NE}(i)} \phi_{01j \rightarrow i}^{t-1} S_{j(t-1)},$$

(4)

where $\text{NE}(i)$ is the set of all neighbouring locations of location $i$ and $\phi_{01j \rightarrow i}^{t-1}$ is a coupling parameter that represents the effect that the disease spreading from neighbouring area $j$ has on the reemergence of the disease in area $i$ at time $t$, the specification of which we will give shortly. The probability of disease persistence, $p11_{it}$, is defined similarly but all parameters can differ,

$$\logit(p11_{it}) = \eta_0 + z_{it}^T \eta + \sum_{j \in \text{NE}(i)} \phi_{11j \rightarrow i}^{t-1} S_{j(t-1)}.$$

(5)
The transition probabilities can depend on $\mathbf{y}^{t-1}$ through $\mathbf{z}_{it}$, which may contain transformed lagged values of the reported cases (e.g. $\log(y_{it-1} + 1)$). This can be justified by the fact that the disease will only go extinct when there is a small number of infectious individuals. We also need to specify an initial state distribution for the Markov chain in each area, that is, $p(S_{i0})$ for $i = 1, \ldots, N$.

In order to account for space-time heterogeneity in the effects of neighbouring disease spread and to quantify associations between the effects of neighbouring disease spread and various space-time factors we use linear models for the coupling parameters (Smith et al., 2002),

$$
\begin{align*}
\phi_{01,j-i}^{(t-1)-t} &= \beta_{01}^{(c)} + \mathbf{z}_{01,jit}^{(c)} \mathbf{S}^{(c)} \\
\phi_{11,j-i}^{(t-1)-t} &= \beta_{11}^{(c)} + \mathbf{z}_{11,jit}^{(c)} \mathbf{H}^{(c)},
\end{align*}
$$

where $\mathbf{z}_{01,jit}^{(c)}$ and $\mathbf{z}_{11,jit}^{(c)}$ are vectors of space-time covariates, possibly different, related to either area, for example, $\mathbf{z}_{01,jit}^{(c)} = \log(y_{jt} / \text{prevalence in the neighbouring area } j)$ or $\mathbf{z}_{01,jit}^{(c)} = \log(\text{prevalence in area } j) \times \text{prevalence in area } j$ (a gravity term (Tuite et al., 2011)). This is justified as effects of disease spread between areas often vary across space and time with observed factors, some examples being: rabies is less likely to spread between areas separated by rivers (Smith et al., 2002); hand, foot and mouth disease is more likely to spread between farms with many cattle (Keeling et al., 2001); and measles is more likely to spread from large cities (Grenfell et al., 2001). We allow the effect of neighbouring disease spread on the reemergence of the disease to be different from its effect on the persistence of the disease as these represent two distinct epidemiological processes.

The disease may spread to non-infected areas from neighbouring areas (modelled through $\phi_{01,j-i}^{(t-1)-t}$ in Equation (4)); but also, neighbouring areas can supply infectious individuals needed to maintain the persistence of the disease in infected areas (modelled through $\phi_{11,j-i}^{(t-1)-t}$ in Equation (5)) (Okano et al., 2020).

As the coupling terms in Equation (4) and (5) account for disease spread from neighbouring areas, the $\zeta_0 + \zeta_0^{(c)} \mathbf{z}$ and $\eta_0 + \mathbf{z}^{(c)} \mathbf{H}$ terms account for long distance dispersal of the disease as well as within area disease dynamics, that is, spread from human and/or vector populations within the area. We assume the effect of covariate $\zeta_{ik}$ on the reemergence of the disease, represented by $\zeta_k$ in Equation (4), can be different from its effect on the persistence of the disease, represented by $\eta_k$ in Equation (5). The advantage of considering differing effects for each covariate on the reemergence and persistence of the disease is that many important epidemiological covariates can interact with the reemergence and persistence of an infectious disease quite differently, like temperature in our motivating example.

A final important motivation of the Markov chain model (3) for $S_{it}$ is that it can account for both many consecutive periods of disease presence and absence. The probability the disease will be absent in location $i$ during time $t$ given it was absent at time $t - 1$ is given by $1 - p1_{it}$, that is, $1 - p01_{it}$ is the probability of having a consecutive period of disease absence. Therefore, as the probability of disease reemergence, $p01_{it}$, approaches 0, consecutive periods of disease absence become more likely. Similarly, as the probability of disease persistence, $p11_{it}$, approaches 1, consecutive periods of disease presence become more likely. Therefore, when $p01_{it} \ll p11_{it}$ the Markov chain can model switching between long periods of disease absence and long periods of disease presence which is often observed for infectious diseases in smaller areas, see Figure 1a and Adams and Boots (2010).
We will refer to the model defined by Equations (1)–(6) as the zero-state coupled Markov switching negative binomial (ZS-CMSNB) model. Finally, we note that a Poisson version of the ZS-CMSNB model is a special case as \( r_{it} \) goes to infinity, which we will refer to as a zero-state coupled Markov switching Poisson (ZS-CMSP) model.

### 3 | INFERENTIAL PROCEDURE

Let \( S_t = (S_{1t}, \ldots, S_{Nt})^T \) be the vector of all state indicators at time \( t \) and let \( S = (S_0, \ldots, S_T)^T \) be the vector of all state indicators. The vector \( S_t \) forms a first-order Markov chain, conditional on \( \mathbf{y}^{(t-1)} \), with state space \( \{0, 1\}^N \) and \( 2^N \times 2^N \) transition matrix \( \Gamma(S_t | \mathbf{y}^{(t-1)}) \). An element of \( \Gamma(S_t | \mathbf{y}^{(t-1)}) \) is given by,

\[
P(S_{1t} = s_{1t}, \ldots, S_{Nt} = s_{Nt} | S_{1(t-1)} = s_{1(t-1)}, \ldots, S_{N(t-1)} = s_{N(t-1)}, \mathbf{y}^{(t-1)}) = \prod_{i=1}^N P(S_{it} = s_{it} | S_{i(t-1)} = s_{i(t-1)}, \mathbf{y}^{(t-1)})
\]  

(7)

Therefore, rewriting the ZS-CMSNB model in terms of \( \Gamma(S_t | \mathbf{y}^{(t-1)}) \) and \( p(\mathbf{y}_t | S_t, \mathbf{y}^{(t-1)}) = \prod_{i=1}^N p(y_{it} | S_{it}, \mathbf{y}^{(t-1)}) \) shows that it is a Markov switching model as defined by Frühwirth-Schnatter (2006). However, the transition matrix has an exceptionally large dimension \( (2^N \times 2^N) \) which necessitates some changes to the inferential procedure compared to more traditional Markov switching models (Frühwirth-Schnatter, 2006), as we will discuss next.

The likelihood of \( \mathbf{v} = (\theta, \beta)^T \), where \( \theta = (\zeta_0, \eta_0, \xi, \zeta^{(c)}, \eta^{(c)}, \xi^{(c)})^T \) and \( \beta = (\delta, \gamma)^T \), given \( \mathbf{y} = (y_1, \ldots, y_T)^T \) and \( S \) is given by,

\[
L(\mathbf{y}, S | \mathbf{v}) = \prod_{i=1}^N \prod_{t=1}^T p(y_{it} | S_{it}, \mathbf{y}^{(t-1)}, \theta) \prod_{i=1}^N p(S_0) \prod_{t=1}^T p(S_{it} | S_{i(t-1)}, \mathbf{y}^{(t-1)}, \theta).
\]  

(8)

When \( N \) is large it is not possible to marginalize out \( S \) from Equation (8), as doing so requires matrix multiplication with \( \Gamma(S_t | \mathbf{y}^{(t-1)}) \) (Frühwirth-Schnatter, 2006). Therefore, we estimate the unknown elements of \( S \) along with \( \mathbf{v} \) by sampling both from their joint posterior distribution which, from Bayes’ theorem, is proportional to,

\[
p(S, \mathbf{v} | \mathbf{y}) \propto L(\mathbf{y}, S | \mathbf{v})p(\mathbf{v}),
\]  

(9)

where \( p(\mathbf{v}) \) is the prior distribution of \( \mathbf{v} \). We specify independent uninformative normal and gamma priors for all low-level parameters. As the joint posterior is not available in closed form, we resort to Markov chain Monte Carlo methods, in particular, we used a hybrid Gibbs sampling algorithm with some steps of the Metropolis–Hastings algorithm to sample from it. We sampled most elements of \( \mathbf{v} \) individually, using an adaptive random walk Metropolis step (Shaby & Wells, 2010). The parameter vectors \( (\eta_0, \eta^{(c)}, \xi) \) and \( (\zeta_0, \zeta^{(c)} \xi^{(c)}) \) showed high posterior correlations and, therefore, we jointly sampled them using automated factor slice sampling (Tibbits et al., 2014). This doubled the efficiency (minimum effective sample size per hour) of the Gibbs sampler in our application.

It is straightforward to sample each unknown element of \( S \) one at a time. When \( y_{it} = 0 \), the full conditional of \( S_{it} \) is given by,
\[ P \left( S_{it} = 1 | \mathbf{y}, \mathbf{v}, \{ S_{jt} | j \neq i, t \in [0, \ldots, T] \} \right) = \frac{P(\mathbf{y}_{it} = 0 | S_{it} = 1)P(S_{it} = 1 | S_{i(t-1)}) \prod_{j=t}^{N} P(S_{j(t+1)} | S_{jt} = 1, S_{(−i)(t)})}{\sum_{S_{it} = 0}^{1} P(\mathbf{y}_{it} = 0 | S_{it} = s_{it})P(S_{it} = s_{it} | S_{i(t-1)}) \prod_{j=t}^{N} P(S_{j(t+1)} | S_{jt} = s_{jt}, S_{(−i)(t)})}, \tag{10} \]

where \( P(\mathbf{y}_{it} = 0 | S_{it} = s_{it}) = \left( \frac{r_{a} - \beta^{2} + s_{it}}{r_{a}} \right)^{r_{a}s_{it}} \) and the dependence of the densities on \( \mathbf{y}^{(t-1)} \), \( \beta \) and \( \theta \) is suppressed to reduce the size of the equation. However, one at a time sampling is known to lead to poor mixing in the Markov switching literature due to the strong posterior correlations between the unknown state indicators (Scott, 2002). Better mixing can be achieved by sampling all of \( S \) jointly from \( p(\mathbf{S} | \mathbf{y}, \mathbf{v}) \) (Chib, 1996). However, this is not computationally feasible with our model when \( N \) is large, as it again involves matrix multiplication with \( \Gamma(\mathbf{S}_{i} | \mathbf{y}^{(t-1)}) \) (Frühwirth-Schnatter, 2006). As an alternative, we propose to block sample \( S \) with each block containing all the state indicators in a different collection of locations. More specifically, assume we have put the locations into \( k = 1, \ldots, B \) disjoint blocks and let \( \mathbf{b}_{k} = (b_{k1}, \ldots, b_{kn_{k}})^{T} \) denote the vector of locations in block \( k \) ordered numerically. We assume \( \sum_{k=1}^{B} n_{k} = N \) so that every location is contained in some block. We will adopt the following notation, let \( \mathbf{S}_{(\mathbf{b}_{k})t} = (S_{(b_{k1})t}, \ldots, S_{(b_{kn_{k}})t})^{T} \) be the vector of all state indicators at time \( t \) whose location is in block \( k \), let \( \mathbf{S}_{(\mathbf{b}_{k})t} = (S_{(b_{k1})t}, \ldots, S_{(b_{kn_{k}})t})^{T} \), let \( \mathbf{S}_{(\mathbf{b}_{k})} = \mathbf{S}_{(\mathbf{b}_{k})(0:T)} \) be the vector of all state indicators whose location is in block \( k \), let \( \mathbf{S}_{(−\mathbf{b}_{k})} \) be \( \mathbf{S}_{t} \) with \( \mathbf{S}_{(\mathbf{b}_{k})t} \) removed, let \( \mathbf{S}_{(−\mathbf{b}_{k})(t_{1}:t_{2})} \) be \( (\mathbf{S}_{1}, \ldots, \mathbf{S}_{t})^{T} \) with \( \mathbf{S}_{(\mathbf{b}_{k})(t_{1}:t_{2})} \) removed and let \( \mathbf{S}_{(−\mathbf{b}_{k})} = \mathbf{S}_{(−\mathbf{b}_{k})(0:T)} \) so that \( \mathbf{S}_{(−\mathbf{b}_{k})} \) is the vector of all state indicators whose location is outside of block \( k \). The idea is to jointly sample \( \mathbf{S}_{(\mathbf{b}_{k})} \) from its full conditional distribution, given by,

\[ p(\mathbf{S}_{(\mathbf{b}_{k})} | \mathbf{S}_{(−\mathbf{b}_{k})}, \mathbf{y}, \mathbf{v}) = p(\mathbf{S}_{(\mathbf{b}_{k})t} | \mathbf{S}_{(−\mathbf{b}_{k})(0:T)}, \mathbf{y}, \mathbf{v}) \times \prod_{t=0}^{T-1} p(\mathbf{S}_{(\mathbf{b}_{k})t+1} | \mathbf{S}_{(−\mathbf{b}_{k})(0:t+1)}, \mathbf{y}^{(t)}, \mathbf{v}). \tag{11} \]

A blocked forward filtering backward sampling (bFFBS) algorithm is needed to sample from Equation (11), which we provide in the Supplementary Material (SM) Section 1. The bFFBS algorithm only requires matrix multiplication with a \( 2^{n_{k}} \times 2^{n_{k}} \) matrix. Our algorithm can be seen as an extension of the individual forward filtering backward sampling (iFFBS) algorithm recently proposed by Touloupou et al. (2020), who considered \( n_{k} = 1 \) for all \( k \).

Our hybrid Gibbs sampler was implemented using the R package Nimble (de Valpine et al., 2017). Nimble comes with built in Metropolis–Hastings, automated factor slice and binary (equivalent to one at a time sampling for the unknown presence/absence indicators) samplers. The bFFBS samplers were implemented using Nimble’s custom sampler feature. All Nimble R code, including for the custom bFFBS samplers, are provided on github (https://github.com/Dirk-Douwes-Schultz/ZS_CMSP_code). Nimble was chosen as it is extremely fast (C++ compiled) and only requires the coding of new samplers. In the SM Section 2, we provide a simulation study, which shows that our proposed Gibbs sampler can recover the true parameters of the ZS-CMSNB model.

### 3.1 Temporal prediction

For arbitrary \( K \) step ahead temporal predictions, we use a simulation procedure (Frühwirth-Schnatter, 2006) to draw samples from the posterior predictive distributions.
Algorithm 1 in the SM will obtain realizations from the posterior predictive distribution of the cases, $y_{i(T+k)}^{[m]} \sim p(y_{i(T+k)}|y)$, and the presence of the disease, $S_{i(T+k)}^{[m]} \sim p(S_{i(T+k)}|y)$, for $i = 1, \ldots, N$, $k = 1, \ldots, K$ and $m = M + 1, \ldots, Q$, where the superscript $[m]$ denotes a draw from the posterior distribution of the cases, $M$ is the size of the burn-in sample and $Q$ is the total MCMC sample size. However, as $S_{i(T+k)}^{[m]}$ can only take two values, 0 or 1, it is difficult to interpret the uncertainty around this prediction for the presence of the disease. Therefore, instead of using summaries of $S_{i(T+k)}^{[m]}$ we use summaries of $P(S_{i(T+k)} = 1|S_{i(T-k)}^{[m]}, S_{i(T-k-1)}^{[m]}, \ldots, S_{i}^{[m]}, \theta^{[m]})$. See the SM Section 3 for the Monte Carlo approximations of the posterior predictive distributions.

### 3.2 Fitted values

Comparing the predictions, for a fixed $K$, to the observed values is not practical, as it requires fitting the model to multiple time points. This would result in long computational times, with MCMC methods, and unstable estimates for earlier times. Assume in this section $t \in \{1, \ldots, T\}$ and $i \in \{1, \ldots, N\}$. There are three types of fitted values for a Markov switching model: one step ahead, filtered and smoothed (Hamilton, 1993). In our Bayesian setting, the smoothed fitted values are given by,

$$S_{it}^{[m]} = S_{it}^{[m]} \text{ drawn from } p(S_{it}|y),$$

$$y_{it}^{[m]} \text{ drawn from } p(y_{it}|S_{it}^{[m]}, y^{(i-1)}, \beta^{[m]}),$$

for $m = M + 1, \ldots, Q$. Note, $y_{it}^{[m]} \sim p(y_{it}^{*}|y)$, where $y_{it}^{*}|S_{it}, y^{(i-1)}, \beta \sim NB(S_{it} \hat{\lambda}_{it}, r_{it})$. That is, $y_{it}^{*}$ represents a new case count generated by the same parameters, state and past counts that generated $y_{it}$. The smoothed fitted value of the state is drawn automatically by the Gibbs sampler used to fit the model, which is an advantage in using data augmentation to fit a Markov switching model.

Note that $P(S_{it} = 1|y) \approx \frac{1}{Q-M} \sum_{m=M+1}^{Q} S_{it}^{[m]}$. When $y_{it} = 0$, the $P(S_{it} = 1|y)$ represents the posterior probability the disease is present and thus undetected. Therefore, a map of the posterior mean of $S_{it}$ for $i$ where $y_{it} = 0$ can be used by policy makers as a warning system to identify areas reporting zero cases that have a high chance of the disease being undetected.

Clearly, smoothed fitted values represent fits given all information in the data about the states, including from present and future counts. If the model is not good at predicting the state one step ahead, which is largely based on the estimated transition probabilities of the Markov chain, then the smoothed fitted values may not reveal that as much of the information about $S_{it}$ can come from $y_{it}$. The one step ahead distribution of the states is given by $p(S_{i}|y^{(i-1)}, v)$, and of the cases is given by $p(y_{i}|y^{(i-1)}, v) = \sum_{s_{i} \in \{0,1\}} p(y_{i}|s_{i}, y^{(i-1)}, \beta)p(S_{i} = s_{i}|y^{(i-1)}, v)$ (Frühwirth-Schnatter, 2006). As $v$ is unknown, we can draw from the posterior of the one step ahead distributions. That is, the one step ahead fitted value of the state is given by, $S_{it}^{[1*[m]}$ drawn from $p(S_{it}|y^{(i-1)}, v^{[m]})$, and of the counts is given by, $y_{it}^{[1*[m]}$ drawn from $p(y_{it}|S_{it}^{[1*[m]}, y^{(i-1)}, v^{[m]})$, for $m = M + 1, \ldots, Q$, where we use the superscript $1*[m]$ to denote a draw from the posterior of the one step ahead distributions. The advantage of the one step ahead fitted values is that information from $(y_{i}, \ldots, y_{T})^T$ is only used to learn about $v$ and not the states. Therefore, they should more accurately reflect the predictions compared to the smoothed values and should be able to diagnosis a lack of fit in the Markov chain component of the model. The distribution $p(S_{i}|y^{(i-1)}, v^{[m]})$ can be calculated using a Hamiltonian forward filter algorithm (Frühwirth-Schnatter, 2006). However, the filter requires...
matrix multiplication with \( \Gamma(S_t | y^{(t-1)}) \) and so is not computationally feasible to run when \( N \) is large.

As an alternative for large \( N \), we propose coupled one step ahead fitted values. Assume we wish to calculate the one step ahead fitted values for location \( i \) when \( N \) is large. We will put location \( i \) in a block with locations \( b_{NE(i)} = (i, i_1,..., i_d)^T \) where \( i_1,..., i_d \in NE(i) \). Then, borrowing notation from Equation (11), the coupled one step ahead fitted values are given by,

\[
S_{it}^{cl*[m]} \text{ drawn from } p(S_{it} | \mathbf{S}_{it}^{[m]}(-b_{NE(i)}(0:1)), y^{(t-1)}, \mathbf{v}^{[m]}),
\]

\[
y_{it}^{cl*[m]} \text{ drawn from } p(y_{it} | S_{it}^{cl*[m]}, y^{(t-1)}, \beta^{[m]}),
\]

for \( m = M + 1, ..., Q \), where we use the superscript \( cl*[m] \) to denote a draw from the posterior of the coupled one step ahead distributions. The distribution \( P(S_{it} = 1 | \mathbf{S}_{it}^{[m]}(-b_{NE(i)}(0:1)), y^{(t-1)}, \mathbf{v}^{[m]}) \) can be calculated using the forward filtering part of the bFFBS algorithm given in the supplementary material. We do not use quantiles of \( S_{it}^{cl*[m]} \) as it can only take the two values 0 or 1. Instead we use quantiles of \( P(S_{it} = 1 | \mathbf{S}_{it}^{[m]}(-b_{NE(i)}(0:1)), y^{(t-1)}, \mathbf{v}^{[m]}), \) which represents a draw from the posterior distribution of the coupled one step ahead fitted probability of disease presence. We only use information from \( (y_t,..., y_T)^T \) to learn about \( \mathbf{S}_{(-b_{NE(i)}(0:1)} \). Therefore, if we block neighbouring locations together there should be minimal information from \( (y_t,..., y_T)^T \) used in the state estimation.

4 | ANALYSIS OF THE DENGUE FEVER DATA IN RIO DE JANEIRO

4.1 | Model specification and fitting

The ZS-CMSNB model has two components, a nonhomogeneous Markov chain which switches dengue between periods of presence and absence through modelling the reemergence and persistence of the disease, and a negative binomial component that generates the reported cases when dengue is present. For the negative binomial component, we use an endemic/epidemic specification for \( \lambda_{it} \) (Bauer & Wakefield, 2018), as it can account for temporal autocorrelation due to the onward transmission of the disease,

\[
\lambda_{it} = \lambda_{it}^{AR} y_{it}^{(t-1)} + \lambda_{it}^{EN}.
\]

The autoregressive rate \( \lambda_{it}^{AR} \) is meant to represent the transmission intensity of dengue in district \( i \) during month \( t \). We model \( \lambda_{it}^{AR} \) as, \( \log (\lambda_{it}^{AR}) = \beta_{it}^{AR} + \beta_1 \text{Rain}_{t-1} + \beta_2 \text{Temp}_{t-1} \), where \( \text{Temp}_{t-1} \) is the maximum temperature in Rio during the previous month, \( \text{Rain}_{t-1} \) is the millimetres of rainfall and \( \beta_{it}^{AR} \sim N(\beta_0^{AR}, \sigma_{AR}) \) are random intercepts accounting for between district differences in transmission intensity, where \( \sigma_{AR} \) is the standard deviation of the random effects. Rainfall and temperature are entered in a lagged manner since there is usually a delay separating changes in mosquito population and dengue incidence (Coutinho et al., 2006). The endemic risk \( \lambda_{it}^{EN} \) accounts for incidence not due to within area transmission. We compared both space-time and space varying endemic risk using the widely applicable information criterion (WAIC) (Gelman et al., 2014) (results not shown) and settled on a purely spatial varying risk, \( \log (\lambda_{it}^{EN}) = \beta_{it}^{EN} \), where
\( \beta_{01}^{EN} \sim N(\beta_0^{EN}, \sigma_{EN}) \). Examining the reported dengue counts it appears the overdispersion varies by district, therefore, we specified \( r_{it} = r_1 \). We used the weakly informative log-normal prior from Bauer and Wakefield (2018) for \( r_1 \) which assumes the overdispersion is likely between 1.5 and 5 times the mean count in the district.

We considered \( z_{it} = (\text{pop}_i, \text{HDI}_i, \text{Rain}_{i-1}, \text{Temp}_{i-1}, \log(y_{i(t-1)} + 1)) \), where \( \text{pop}_i \) is the population of district \( i \) and \( \text{HDI}_i \) is the human development index of district \( i \), as covariates possibly affecting the reemergence and persistence (see Equations (4) and (5)) of dengue. The monthly rainfall and temperature data were obtained from the National Institute for Space Research (http://bancodedados.cptec.inpe.br/) and the district level human development indexes were obtained from ipeadata (http://www.ipeadata.gov.br/Default.aspx). Since we condition on the first observation, we specified \( p(S_{1i}) \sim \text{Bern}(\pi) \) if \( y_{1i} = 0 \) and \( p(S_{1i}) \) is degenerate 1 if \( y_{1i} > 0 \), as the initial state distributions.

As for covariates potentially associated with the effects of dengue spreading from a neighbouring district, that is, \( z_{01ijt}^{(c)} \) and \( z_{11ijt}^{(c)} \) in Equation (6), we considered; \( \log(y_{j(t-1)}/\text{pop}_j + 1) \), the prevalence of the disease in the neighbouring district; \( \log(\text{pop}_i \times \text{pop}_j) \), a gravity term reflecting the fact that individuals are more likely to move between high population areas; \( |\text{NE}(j)| \), the number of neighbours of area \( j \) which can account for the effects of disease spread potentially being proportioned among neighbours; and, finally, \( \log(y_{i(t-1)} + 1) \) for the persistence only, to potentially account for crowding out effects. Due to the high computational cost per covariate added to the coupling parameters we started with a simple homogeneous model of disease spread and added coupling covariates sequentially, removing them if they did not improve the WAIC substantially (by 10 units); this process is illustrated in Table 1. Based on Table 1 we decided on \( z_{01ijt}^{(c)} = (\log(y_{j(t-1)}/\text{pop}_j + 1), \log(\text{pop}_i \times \text{pop}_j)) \) and a homogeneous effect of neighbouring disease spread for the persistence of the disease. That is, we let \( \phi_{01ijt} = \phi_{01ijt}^{(c)} + \phi_{11ijt}^{(c)} \log(y_{i(t-1)}/\text{pop}_j + 1) + \phi_{11ijt}^{(c)} \log(\text{pop}_i \times \text{pop}_j) \) and \( \phi_{11ijt} = \eta_0^{(c)} \) for the final model.

Each ZS-CMSNB model in Table 1, and the ZS-CMSP model as it is a special case, was fitted using our proposed Gibbs sampler from Section 3, to the monthly Rio dengue cases for 2011-2017 (\( t = 1, \ldots, 84 \)) and all districts in the city (\( i = 1, \ldots, 160 \)). We ran the Gibbs sampler for 80,000 iterations on three chains with an initial burn in of 30,000 iterations. All sampling was started from random values in the parameter space to avoid convergence to local modes. Convergence was checked using the Gelman–Rubin statistic (all estimated parameters <1.05) and the minimum effective sample size (>1000) (Plummer et al., 2006). For the final bolded model in Table 1, we compared the efficiency (minimum effective sample size per hour) of three candidate samplers for the state indicators: a one at time sampler (binary) and two bFFBS samplers with block sizes 1 (iFFBS) and 2 (bFFBS2). For the bFFBS2 sampler, we blocked 71 neighbouring locations together leaving 15 locations that could not be matched to a neighbour that were put in single location blocks (remaining three locations had all positive observations and did not need to be sampled). The iFFBS sampler was 10% more efficient than the binary sampler and the bFFBS2 sampler was 5% more efficient. Therefore, there does not appear to be much if any gain to joint sampling the hidden states in our application.

We compared the fit of our model to a nested model without the Markov chain, that is, a model for which \( \zeta_0 = \eta_0 \), \( \zeta_1 = \eta_1 \), \( \phi_0 = \eta_0^{(c)} = 0 \) and \( \phi_1 = \eta_1^{(c)} = 0 \) in Equations (4)–(6). Note that this is a standard zero-inflated negative binomial (ZINB) model (Greene, 1994). However, the ZINB model does not account for spatio-temporal correlations in the presence of the disease. Therefore, we also compared to a model like the one used by Hoef and Jansen (2007).
TABLE 1 Model comparison using widely applicable information criterion (WAIC) for models fitted to the dengue data. Note, $\emptyset$ refers to no covariates, or a homogeneous model of disease spread, that its, $\phi_{01j-i}^{(1-1)-t} = \sigma_{0}^{c}$ or $\phi_{11j-i}^{(1-1)-t} = \theta_{0}^{c}$. The bolded model is our final ZS-CMSNB model as we do not accept models that are more complex and do not improve the WAIC substantially (by 10 units). NP = Neighbouring prevalence $=$ log($y_{it-1}/pop_{i} + 1$), G = Gravity term $=$ log($pop_{i} \times pop_{j}$), LC = Local cases $=$ log($y_{it-1} + 1$), NN = Number of neighbours $=$ |$NE(j)$|

| Model              | $z_{01,ijt}^{(c)}$ | $z_{11,ijt}^{(c)}$ | WAIC  |
|--------------------|--------------------|--------------------|-------|
| endemic/epidemic   | $\emptyset$        | $\emptyset$        | 69,939|
| ZINB               | $\emptyset$        | $\emptyset$        | 67,798|
| ZINBRE             | $\emptyset$        | $\emptyset$        | 67,782|
| ZS-CMSNB           | $\emptyset$        | $\emptyset$        | 67,729|
| $(NP)^T$           | $\emptyset$        |                   | 67,652|
| $(NP, G)^T$        | $\emptyset$        |                   | 67,632|
| $(NP, G)^T$        | $(NP)^T$           |                   | 67,631|
| $(NP, G)^T$        | $(NP, LC)^T$       |                   | 67,628|
| $(NP, G, NN)^T$    | $\emptyset$        |                   | 67,648|
| ZS-CMSP            | $\emptyset$        | $\emptyset$        | 90,458|

Let $p_{it} = P(S_{it} = 1|\theta_{RE})$, where $\theta_{RE}$ is a vector of random effects, then Hoef and Jansen (2007) fit a model where logit($p_{it}$) = $a_{0} + q_{it}^{T}\alpha + \eta_{i} + \delta_{it}$, where $q_{it}$ is a vector of space-time covariates, $\eta_{i}$ is an overall time trend which followed an autoregressive (AR) process and $\{\delta_{it}\}_{i=1}^{N}$, followed a proper conditional autoregressive (pCAR) distribution for each time period that was independent across time. As most excess variation in spatio-temporal infectious disease counts occurs temporally and not spatially (Bauer et al., 2016), we flip their proposal so that logit($p_{it}$) = $a_{0} + q_{it}^{T}\alpha + \eta_{i} + \delta_{i}(t)$, where $\eta_{i}$ is an overall spatial effect and $\delta_{i}(t)$ represents a time trend for each area. We assume a pCAR distribution for $\{\eta_{i}\}_{i=1}^{N}$. Torabi (2017) suggests using an AR process or splines to model $\delta_{i}(t)$. However, we prefer using autoregression for the spatial–temporal interaction, combining the proposals of Fernandes et al. (2009) and Yang et al. (2013), as it is computationally much simpler and also epidemiologically motivated. That is, we let $\delta_{i}(t) = \rho_{1}\text{log}(y_{i(t-1)} + 1) + \rho_{2}S_{i(t-1)}$. The motivation being that for the disease to go extinct all previous cases must fail to pass on the disease but since $y_{i(t-1)}$ represents reported cases there might still be an effect at $y_{i(t-1)} = 0$ due to the disease being undetected. We take $q_{it} = \{z_{it}\}\backslash\{\text{log}(y_{i(t-1)} + 1)\}$ as $\text{log}(y_{i(t-1)} + 1)$ is already included in the model. We call this model a zero-inflated negative binomial random effect (ZINBRE) model.

Finally, we also compared the fit of our model to a model without zero inflation (endemic/epidemic) and to a Poisson version of our model (ZS-CMSP). Table 1 gives the results of the model comparison using WAIC for the five classes of models. The endemic/epidemic, ZINB and ZINBRE models in Table 1 were fit using standard MCMC methods in Nimble, the code is available on github (https://github.com/Dirk-Douwes-Schultz/ZS_CMSP_code). From Table 1, incorporation of zero inflation and overdispersion significantly improves model fit, and the ZS-CMSNB model is the superior model for the zero inflation.
4.2 Results

From Table 1, the most relevant factors for determining the strength of dengue spread from a neighbouring area are, the reported prevalence of dengue in the neighbouring area, and the population of the neighbouring area and the home area. To quantify the association between the effects of neighbouring dengue spread and these two factors we can use the odds ratio (OR) of dengue reemergence given disease spread from a neighbouring area, or mathematically,

\[
\Omega(S_t = 1 | S_{(t-1)} = 0, S_{(t-1)} = 1) = \frac{\exp \left( \zeta_0^{(c)} + \zeta_1^{(c)} \log(y_{(t-1)} / \text{pop}_j + 1) + \zeta_2^{(c)} \log(\text{pop}_i \times \text{pop}_j) \right)}{\Omega(S_t = 1 | S_{(t-1)} = 0, S_{(t-1)} = 0)},
\]

(17)

for \( j \in NE(i) \), where \( \Omega(A) = P(A)/(1 - P(A)) \). From Table 2, which shows estimates from the Markov chain component of the model, the posterior mean and 95% posterior credible interval of \( \zeta_1^{(c)} \) is 5.63 [3.33–8.67] and of \( \zeta_2^{(c)} \) is 0.22 [0.09–0.36]. Therefore, the effect of dengue spreading from a neighbouring area is positively associated with the reported prevalence of dengue in the neighbouring area and the population of both areas, especially with the prevalence. To help visualize these associations we have plotted, in Figure 3, (17) versus the population of the neighbouring area (top) and the reported cases of dengue in the neighbouring area (bottom). In the plots, we fixed the population of the home area at 10,000, to represent a small area, and for the top graph we fixed prevalence in the neighbouring area at its median value (.08 per 1000), and for the bottom graph we fixed the population in the neighbouring area at its median value (26,000) (exact

| Covariate | Probability or odds ratio |
|-----------|--------------------------|
|           | Reemergence (absence to presence) | Persistence (presence to presence) |
| Intercept (shifted) | 0.33 [0.22–0.46] | 0.84 [0.78–0.90] |
| pop (1000s) | 1.06 [1.02–1.11] | 1.02 [1.01–1.03] |
| HDI (0.1) | 1.07 [0.71–1.52] | 0.82 [0.62–1.09] |
| Temp_{t-1} (Celc.) | 1.24 [1.11–1.39] | 1.21 [1.13–1.31] |
| Rain_{t-1} (10 mm) | 1.04 [1–1.09] | 1.13 [1.09–1.17] |
| log(y_{it-1}) + 1 | – | 5.07 [4.17–6.19] |
| Neighbourhood Presence | 1.35 [1.14–1.64]† | 1.12 [1.01–1.23] |
| Intercept | –1.38 [−2.28−−0.51] | 0.12 [0.01–0.23] |
| log (y_{it-1}/\text{pop}_j + 1) | 5.63 [3.33–8.67] | – |
| log (\text{pop}_i \times \text{pop}_j) | 0.22 [0.09–0.36] | – |

† Average effect in a small area.
Posterior means (solid lines) and 95% posterior credible intervals (dashed lines) of (top) $\exp\left(\xi_0 + \xi_1 \log(0.08 + 1) + \xi_2 \log(10 \times \text{pop}_j)\right)$ versus $\text{pop}_j$ and (bottom) $\exp\left(\xi_0 + \xi_1 \log(\text{cases}_j/26 + 1) + \xi_2 \log(10 \times 26)\right)$ versus $\text{cases}_j$ (see (17)), where, 10 is the population of a small area, 26 is the median population and 0.08 (per 1000) is the median prevalence of dengue.

From Figure 3 (top), the effect of dengue spreading from a neighbouring area increases gradually with the population of the neighbouring area and there is likely not a large effect of disease spread from low population areas, at least at median prevalence. These estimates are reasonable as there will be much more travel to high population neighbours compared to low population neighbours. From Figure 3 (bottom), the effect of dengue spreading from a neighbouring area increases rapidly with the reported prevalence of the disease in the neighbouring area. At a prevalence of $10/26 = 0.38$, where we have cut off the graph to improve visualization, the odds ratio is 5.95 [2.97-12.62], meaning the odds of dengue reemergence is increased six (likely lowest 3) times due to the disease spread, a large effect. A prevalence of 0.38 is about at epidemic levels, 75th percentile, and, therefore, the figure implies that if an epidemic occurs in an area there is a good chance that dengue will spread to neighbouring areas where it is absent.
We also wanted to investigate whether dengue spreads between areas undetected. Note that the OR of dengue reemergence given disease spread from a neighbouring area where dengue is undetected is given by \( \exp \left( \beta_0 + \beta_2 \log(\text{pop}_i \times \text{pop}_j) \right) \). Through repeated calculations we found that, \( a \ posteriori \), this quantity has a 75% chance of being greater than one if \( \text{pop}_j > 685.4 / \text{pop}_i \). Setting \( \text{pop}_i = 10 \) (1000) gives \( \text{pop}_j > 68.53 \), which represents the top 15 per cent of districts in terms of population size. Therefore, the disease likely spreads undetected only from the larger districts.

The estimated coefficients for the Markov chain part of the fitted ZS-CMSNB model are given in Table 2. From the estimated intercepts, in small areas consecutive periods of disease presence are heavily favoured on average, while consecutive periods of disease absence are only moderately favoured. This highlights the importance of preventing dengue reemergence in small areas (one effective way would be to prevent disease spread from neighbouring epidemic areas as shown in Figure 3), as when the disease does reemerge it will likely persist for some time. We shift the intercepts to small districts (\( \text{pop} = 10,000 \)) as that is where most of the 0s occur (there are not many 0s in large districts). Population size has a significantly larger effect on dengue reemergence compared to persistence (diff = 0.04 [0–0.08]), especially considering population increases the effects of dengue spreading from neighbouring areas on the reemergence of the disease, see Figure 3. This can explain why, in Figure 2, there is more spatial variation observed in the average probabilities of dengue reemergence compared to persistence. Additionally, the strong positive association between population size and dengue reemergence means we would expect much longer periods of disease absence in the smaller districts compared to larger districts, which follows the well-known theories from Bartlett (1957). Interestingly, we found no association between human development index and the risk of dengue reemergence or persistence. Both rainfall and temperature have a strong positive association with dengue reemergence and persistence. For example, we estimated that a one degree rise in maximum temperature during the previous month is associated with a 24 [11-39] per cent increase in the odds of dengue reemerging and a 21 [13-31] per cent increase in the odds of dengue persisting. Therefore, we will have longer periods of disease absence during the winter (at least in the smaller districts) and longer periods of disease presence during the summer, which follows patterns in the mosquito population. Rainfall has a significantly larger effect on the persistence of the disease (diff = 0.08 [0.02–0.15]), which clearly manifests in Figure 4, and could be due to the role played by rainfall in the lifecycle of the mosquito egg (Coutinho et al., 2006). When the disease is present water is needed for egg laying and so rainfall helps the mosquito population persist, however, if the disease is absent there are not many mosquitoes and rainfall could wash away vertically infected mosquito eggs. The effect of the previous months cases, \( \log(Y_{t-1}) + 1 \), on the persistence of the disease is quite large and reflects the fact that dengue will only go extinct when there is a small number of infected individuals.

From the Neighbourhood Presence row of Table 2, we estimated that dengue being present in a neighbouring district during the previous month is associated with a 35 [14-64] per cent increase in the odds of dengue reemerging in a small area on average and a 12 [1-23] per cent increase in the odds of dengue persisting. The average effect of neighbouring dengue spread on the reemergence of the disease is much higher than the homogeneous effect on the persistence. Therefore, efforts to restrict cross infection between districts will typically be more effective if one district is not already infected. A possible explanation is a 'spark effect', like forest fires, in that only a small amount of disease spread could be needed to start the disease being present in an area but will not contribute much to, say, a large epidemic. Note, however, that the Neighbourhood Presence row
FIGURE 4  Posterior means (solid lines) and 95% posterior credible intervals (dashed lines) of the estimated monthly probabilities of dengue reemergence and persistence in a small district (pop = 10,000). We assumed average monthly temperature and rainfall values, and two cases (median) reported in the previous month for the persistence probabilities of Table 2 only shows the average effect of neighbouring dengue spread on the reemergence of the disease in a small area, that is, \( \exp \left( \gamma_0^{(c)} + \gamma_1^{(c)} \log(.08 + 1) + \gamma_2^{(c)} \log(10 \times 26) \right) \), and the actually effect varies quite a lot with space and time as shown in Figure 3, making broad comparisons somewhat challenging.

Figure 4 shows posterior summaries of the estimated monthly probabilities of dengue reemergence and persistence for a small district (pop = 10,000) without neighbouring disease spread. The temporal evolution of the two probabilities is clearly different, which mainly reflects the increased effect of rainfall on the persistence of the disease. Additionally, the figure illustrates how the model can recreate the seasonal switching of dengue between long periods of disease absence in the winter and long periods of disease presence in the summer (Adams & Boots, 2010).

Table 3 gives the estimated parameters from the negative binomial part of the model. When dengue is present in a district, increases in temperature and rainfall lead to increased transmission of the disease. There is significant between district differences in the transmission rate...
TABLE 3. Posterior means and 95% posterior credible intervals (in squared brackets) from the negative binomial part of the fitted ZS-CMSNB model. \( \bar{r} = 1/N \sum_{i=1}^{N} r_i \) and \( \sigma_r = \sqrt{1/(N-1) \sum_{i=1}^{N} (r_i - \bar{r})^2} \)

| Covariate Parameter | Estimate |
|---------------------|----------|
| Intercept AR \( \beta_0^{AR} \) | -0.37 [-0.4–−0.33] |
| Rain, \( t-1 \) (10 mm) AR \( \beta_1 \) | 0.036 [0.03–0.04] |
| Temp, \( t-1 \) (C) AR \( \beta_2 \) | 0.23 [0.22–0.24] |
| Std. dev AR \( \sigma_{AR} \) | 0.12 [0.09–0.14] |
| Intercept EN \( \beta_0^{EN} \) | 0.37 [0.29–0.45] |
| Std. dev EN \( \sigma_{EN} \) | 0.42 [0.35–0.5] |
| Avg. overdispersion \( \bar{r} \) | 2.27 [2.16–2.39] |
| Std. dev overdispersion \( \sigma_r \) | 1.08 [0.92–1.37] |

and endemic risk. The differences in the endemic risk could be driven by either differences in the environment or reporting rates. There is significant overdispersion on average (variance is 11 times the mean on average) and there is significant between district differences in the overdispersion.

4.3 Fitted values and predictions

Figure 5 shows a map of the posterior probability that dengue is present in the districts of Rio de Janeiro during December 2017 (time \( T \)). If \( y_{iT} > 0 \) then the posterior probability that dengue is present in district \( i \) during time \( T \) is 1 (grey districts on the map). Otherwise, if \( y_{iT} = 0 \) the disease may be undetected and we can approximate \( P(S_{iT} = 1 | y) \approx \frac{1}{Q-M} \sum_{m=M+1}^{Q} \sum_{i \in NE} S_{iT}^{[m]} \). A blue arrow is drawn from \( j \in NE(i) \) to \( i \) if \( \Omega(S_{iT+1}^{[1]} = 1 | S_{iT} = 1) = S_{iT} \exp(\phi_{T→(T+1)}^{11,j→i}) + (1 - S_{iT}) \exp(\phi_{01,j→i}^{T→(T+1)}) \) is greater than 1.2 with probability 0.75, \textit{a posteriori}, where \( \Omega(A) = P(A)/(1 - P(A)) \). That is, if the current effect of disease spread from the neighbouring area is likely greater than the effect of a one degree rise in temperature (see Table 2). The map shows where dengue currently is in the city and where it is spreading to and, therefore, would be of interest to policy makers. For example, the Santa Teresa district has three arrows coming out of it indicating it is an important source of dengue spread in the city currently.

Panels of Figure 6 show posterior summaries of the coupled 1 month ahead fitted values, see Equations (14) and (15), of dengue cases (top graphs) and dengue presence (bottom graphs), for four districts. These were constructed by running the blocked forward filter in blocks with two neighbouring locations, with each MCMC draw, as explained in Section 3.2. For Figure 6, we chose mostly small districts that had a good mix of zeros and positive counts to illustrate both the zero inflated and count components of the model. Generally, the model is able to predict, in sample, the presence of the disease 1 month ahead well when the disease is observed to be present. It is difficult to assess the presence fits when the disease is reportedly absent, as dengue could be undetected and actually present. The fit to the cases is also reasonable, however, there does appear to be some overestimation of uncertainty at high counts. This is due to the quadratic mean variance relationship of the negative binomial distribution. There is an advantage to this mean variance relationship, however, for estimating the Markov chain part of
December 2017, Posterior Probability Dengue is Present

\[
(j \text{ to } i) = \frac{\Omega(S_{(T+1)} = 1|S_{jT} = 1)}{\Omega(S_{(T+1)} = 1|S_{jT} = 0)} > 1.2
\]

with probability .75, a posteriori.

**Figure 5** Posterior probability that dengue is present for December 2017 (time \(T\)). Districts that have reported cases are in grey to distinguish them from districts where dengue may be undetected. A blue arrow is drawn from \(j \in NE(i)\) to \(i\) if \(\frac{\Omega(S_{(T+1)} = 1|S_{jT} = 1)}{\Omega(S_{(T+1)} = 1|S_{jT} = 0)}\) is greater than 1.2 with probability 0.75, a posteriori, where \(\Omega(A) = P(A)/(1 - P(A))\) [Colour figure can be viewed at wileyonlinelibrary.com]

the model as most of the weight is placed at low counts which is where the reemergence and persistence of the disease occurs. As an additional goodness of fit measure, we plot the autocorrelation function of the Pearson residuals in the SM Section 4 for the four districts of Figure 6. There are not many significant autocorrelations, suggesting that there is no structure left in the residuals.

Panels of Figure 7 show summaries of the 1–12 month ahead posterior predictive distributions of dengue cases (top graphs) and dengue presence risk (bottom graphs) in two districts. These were calculated using Algorithm 1 in the SM with \(K = 12\). There was very little dengue activity in 2018 and we chose to show Bonsucesso as it was one of the few districts that experienced a small epidemic while Saúde represents a more typical district. We assumed average monthly temperature and rainfall values in the calculations, although other forecasting scenarios could be considered, such as high temperature scenarios, or forecasts of the meteorological variables could be used. Both predictions of dengue presence and the case counts show a clear seasonal pattern, peaking in the summer and declining during the winter. The uncertainty in the predictions of dengue presence is large in the Saúde district. A reasonable explanation for this large uncertainty is that Saúde reported zero cases for the last time point, meaning there is uncertainty around whether dengue is present in Saúde at time \(T = 84\). As Saúde is a small district (pop = 2,000) there is a large difference in the probabilities of disease reemergence and persistence (due to the larger effect of population size on the reemergence of the disease), and so uncertainty around the presence of the disease previously leads to wide confidence intervals for the future predictions of disease presence.
**FIGURE 6** Posterior summaries of the coupled 1 month ahead fitted values for four districts. (top graphs) Coupled 1 month ahead fitted values of cases versus observed cases. (bottom graphs) Coupled 1 month ahead fitted values of dengue presence risk versus observed presence (0=0 reported cases which may not correspond to the actual absence of the disease). Posterior means (solid lines), 95% posterior credible intervals (dashed lines) and observed (points). Summer/winter seasons highlighted in red/light blue in the bottom graphs [Colour figure can be viewed at wileyonlinelibrary.com]

**5 | CONCLUDING REMARKS**

We have proposed a zero-state coupled Markov switching negative binomial (ZS-CMSNB) model for general spatio-temporal infectious disease counts that contain an excess of zeros. Our approach can model nonhomogeneous switching between long periods of disease presence and absence while accounting for space-time heterogeneity in the effects of disease spread between areas. The main difference with existing ZIC models used in spatio-temporal disease mapping is that the disease switches between periods of presence and absence in each area through a series of nonhomogeneous Markov chains coupled between neighbouring locations, as opposed to the presence of the disease forming a series of conditionally independent Bernoulli random variables, which is presently the most popular approach (Young et al., 2020). The difference is essentially between a finite mixture and finite Markov mixture model (Frühwirth-Schnatter, 2006), although we also introduce some spatial dependence in the states by coupling the partially hidden Markov chains between neighbouring areas.

There are several alternatives to our coupled Markov chains, within the framework of finite mixture ZIC models (Young et al., 2020), for accounting for spatio-temporal correlations in the presence of the disease, mainly: random effects (Hoef & Jansen, 2007), splines (Ghosal et al.,
FIGURE 7  Summaries of the 1–12 month ahead posterior predictive distributions of the cases (top graphs) and dengue presence risk (bottom graphs) for two districts. Posterior predictive means (solid lines) and 95% posterior predictive credible intervals (dashed lines). Solid circles are from the last 7 years (2011–2017) used to fit the model and open circles are future observed values.

2020) and autoregression (Fernandes et al., 2009; Yang et al., 2013). The main advantages of our approach are (1) we allow for each covariate, and between area disease spread, to have a separate effect on the reemergence of the disease compared to the persistence, which is epidemiologically justified in many instances and (2) it is much easier to quantify space-time heterogeneity in the effects of disease spread between areas with our approach as we allow the effects of neighbouring disease spread to depend on a vector of space-time covariates related to either area. In our application of dengue fever our model allowed for several interesting insights into the epidemiology of the disease beyond what existing ZIC methods can provide, some examples being: dengue is more likely to spread from high population areas controlling for prevalence, if an epidemic of dengue occurs in an area the disease is very likely to spread to neighbouring areas, there is typically a larger effect of neighbouring dengue spread on the reemergence of the disease compared to the persistence, and rainfall has a larger effect on the persistence of dengue which leads to clear differences in the temporal evolutions of the two probabilities (see Section 4 for these and more examples). Additionally, our model fit the dengue data better than a reasonably specified alternative existing ZIC model that combined random effects and autoregression, although there are many possible specifications for such a model.
Although we have applied the ZS-CMSNB model to spatio-temporal infectious disease counts, it could be applied to data in other fields as well. For example, Malyshkina and Mannering (2010) considered a zero-state Markov switching count model, without coupling, for modelling traffic accidents across 335 highway segments in Indiana between 1995 and 1999. In their application, the zero state represented a low risk of accidents and the count state represented a high risk of accidents. It could be that a highway segment in the high risk state signals that nearby highway segments are also unsafe, and by coupling the chains between neighbouring highways we could borrow strength between them to help determine the states.

There are also some limitations with our approach. We use a negative binomial distribution to account for overdispersion, but other count distributions could be used, such as the generalized Poisson distribution (Joe & Zhu, 2005). The negative binomial distribution places a lot of weight on low counts, which is helpful when modelling the reemergence and persistence of a disease, as these events mainly occur at low counts, but has the disadvantage of leading to an overestimation of uncertainty during epidemic periods. Additionally, our approach relies heavily on having good covariates that reflect knowledge of the disease. We could incorporate random effects or splines into logit($p_{0|t}$) or logit($p_{1|t}$) to account for additionally heterogeneity when known covariates are not sufficient. Finally, we are only explicitly considering disease spread between neighbouring areas and, especially in a city, people often move around far outside their neighbouring areas. We could allow for disease spread from all areas and incorporate distance into the coupling parameters, although this would likely come at a great computational cost that would need to be overcome to make it a viable approach.

**ACKNOWLEDGEMENTS**

This work is part of the PhD thesis of D. Douwes-Schultz under the supervision of A. M. Schmidt in the Graduate Program of Biostatistics at McGill University, Canada. Schmidt is grateful for financial support from the Natural Sciences and Engineering Research Council (NSERC) of Canada (Discovery Grant RGPIN-2017-04999). Douwes-Schultz is grateful for financial support from IVADO and the Canada First Research Excellence Fund/Apogée (PhD Excellence Scholarship 2021-907035349). This research was enabled in part by support provided by Calcul Québec (www.calculquebec.ca) and Compute Canada (www.computecanada.ca). The authors thank the Editor, Associate Editor and two reviewers whose comments improved the presentation of the paper.

**DATA AVAILABILITY STATEMENT**

All data is publicly available and we include it in the GitHub repo for convenience.

**ORCID**

Dirk Douwes-Schultz https://orcid.org/0000-0002-6186-2275
Alexandra M. Schmidt https://orcid.org/0000-0002-6448-6367

**REFERENCES**

Adams, B. & Boots, M. (2010) How important is vertical transmission in mosquitoes for the persistence of dengue? Insights from a mathematical model. *Epidemics*, 2, 1–10.

Aktekin, T. & Musal, M. (2015) Analysis of income inequality measures on human immunodeficiency virus mortality: a spatiotemporal Bayesian perspective. *Journal of the Royal Statistical Society: Series A (Statistics in Society)*, 178, 383–403.

Arab, A. (2015) Spatial and spatio-temporal models for modeling epidemiological data with excess zeros. *International Journal of Environmental Research and Public Health*, 12, 10536–10548.
Bartlett, M.S. (1957) Measles periodicity and community size. *Journal of the Royal Statistical Society: Series A (General)*, 120, 48–60.

Bauer, C. & Wakefield, J. (2018) Stratified space–time infectious disease modelling, with an application to hand, foot and mouth disease in China. *Journal of the Royal Statistical Society: Series C (Applied Statistics)*, 67, 1379–1398.

Bauer, C., Wakefield, J., Rue, H., Self, S., Feng, Z. & Wang, Y. (2016) Bayesian penalized spline models for the analysis of spatio-temporal count data. *Statistics in Medicine*, 35, 1848–1865.

Beaty, B.J. & Marquardt, W.C. (1996) *The Biology of Disease Vectors*. Niwot, Colorado: University Press of Colorado.

Chib, S. (1996) Calculating posterior distributions and modal estimates in Markov mixture models. *Journal of Econometrics*, 75, 79–97.

Coutinho, F.A.B., Burattinia, M.N., Lopez, L.F. & Massada, E. (2006) Threshold conditions for a non-autonomous epidemic system describing the population dynamics of dengue. *Bulletin of Mathematical Biology*, 68, 2263–2282.

Diebold, F., Lee, J.-H. & Weinbach, G. (1994) Regime switching with time-varying transition probabilities. In: Granger, C.W.J. & Mizon, G. (Eds.) *Nonstationary Time Series Analysis and Cointegration*, Oxford: Oxford University Press, pp. 282–302.

Fernandes, M.V.M., Schmidt, A.M. & Migon, H.S. (2009) Modelling zero-inflated spatio-temporal processes. *Statistical Modelling: An International Journal*, 9, 3–25.

Frühwirth-Schnatter, S. (2006) *Finite Mixture and Markov Switching Models*. Springer Series in Statistics. New York: Springer-Verlag.

Gelman, A., Hwang, J. & Vehtari, A. (2014) Understanding predictive information criteria for Bayesian models. *Statistics and Computing*, 24, 997–1016.

Ghosal, S., Lau, T.S., Gaskins, J. & Kong, M. (2020) A hierarchical mixed effect hurdle model for spatiotemporal count data and its application to identifying factors impacting health professional shortages. *Journal of the Royal Statistical Society: Series C (Applied Statistics)*, 69, 1121–1144.

Giorgi, E., Schlüter, D.K. & Diggle, P.J. (2018) Bivariate geostatistical modelling of the relationship between Loa loa prevalence and intensity of infection. *Environmetrics*, 29, e2447.

Goldfeld, S.M. & Quandt, R.E. (1973) A Markov model for switching regressions. *Journal of Econometrics*, 1, 3–15.

Greene, W.H. (1994) Accounting for excess zeros and sample selection in Poisson and negative binomial regression models. *NYU Working Paper No. EC-94-10*, New York University, Rochester, NY.

Grenfell, B.T., Bjørnstad, O.N. & Kappes, J. (2001) Travelling waves and spatial hierarchies in measles epidemics. *Nature*, 414, 716–723.

Hamilton, J.D. (1989) A new approach to the economic analysis of nonstationary time series and the business cycle. *Econometrica*, 57, 357–384.

Hamilton, J.D. (1993) 9 Estimation, inference and forecasting of time series subject to changes in regime. In: Mad-dala, G.S., Rao, C.R. & Vinod, H.D. (Eds.) *Handbook of Statistics vol. 11 of Econometrics*, Amsterdam: Elsevier, pp. 231–260.

Hoef, J.M.V. & Jansen, J.K. (2007) Space—time zero-inflated count models of harbor seals. *Environmetrics*, 18, 697–712.

Hooten, M.B. & Wikle, C.K. (2010) Statistical agent-based models for discrete spatio-temporal systems. *Journal of the American Statistical Association*, 105, 236–248.

Joe, H. & Zhu, R. (2005) Generalized Poisson distribution: the property of mixture of Poisson and comparison with negative binomial distribution. *Biometrical Journal*, 47, 219–229.

Keeling, M.J., Woolhouse, M.E., Shaw, D.J., Matthews, L., Chase-Topping, M., Haydon, D.T. et al. (2001) Dynamics of the 2001 UK foot and mouth epidemic: stochastic dispersal in a heterogeneous landscape. *Science*, 294, 813–817.

Lambert, D. (1992) Zero-inflated Poisson regression, with an application to defects in manufacturing. *Technometrics*, 34, 1–14.

Malyskina, N.V. & Mannering, F.L. (2010) Zero-state Markov switching count-data models: an empirical assessment. *Accident Analysis & Prevention*, 42, 122–130.

Okano, J.T., Sharp, K., Valdano, E., Palk, L. & Blower, S. (2020) HIV transmission and source–sink dynamics in sub-Saharan Africa. *The Lancet HIV*, 7, e209–e214.
Plummer, M., Best, N., Cowles, K. & Vines, K. (2006) CODA: convergence diagnosis and output analysis for MCMC. R News, 6, 7–11.

Pohle, J., Langrock, R., der Schaar, M., King, R. & Jensen, F.H. (2021) A primer on coupled state-switching models for multiple interacting time series. Statistical Modelling, 21, 264–285.

Schmidt, W.-P., Suzuki, M., Thiem, V.D., White, R.G., Tsuzuki, A., Yoshida, L.-M. et al. (2011) Population density, water supply, and the risk of dengue fever in Vietnam: cohort study and spatial analysis. PLOS Medicine, 8, e1001082.

Scott, S.L. (2002) Bayesian methods for hidden Markov models. Journal of the American Statistical Association, 97, 337–351.

Shaby, B.A. & Wells, M.T. (2010) Exploring an Adaptive Metropolis Algorithm. Department of Statistics, Duke University.

Smith, D.L., Lucey, B., Waller, L.A., Childs, J.E. & Real, L.A. (2002) Predicting the spatial dynamics of rabies epemics on heterogeneous landscapes. Proceedings of the National Academy of Sciences, 99, 3668–3672.

Stoddard, S.T., Forshay, B.M., Morrison, A.C., Paz-Soldan, V.A., Vazquez-Prokopec, G.M., Astete, H. et al. (2013) House-to-house human movement drives dengue virus transmission. Proceedings of the National Academy of Sciences of the United States of America, 110, 994–999.

Tibbits, M.M., Groendyke, C., Haran, M. & Liechty, J.C. (2014) Automated factor slice sampling. Journal of Computational and Graphical Statistics, 23, 543–563.

Torabi, M. (2017) Zero-inflated spatio-temporal models for disease mapping. Biometrical Journal, 59, 430–444.

Touloupou, P., Finkenstädt, B. & Spencer, S.E.F. (2020) Scalable Bayesian inference for coupled hidden Markov and semi-Markov models. Journal of Computational and Graphical Statistics, 29, 238–249.

Tuite, A.R., Tien, J., Eisenberg, M., Earn, D.J., Ma, J. & Fisman, D.N. (2011) Cholera epidemic in Haiti, 2010: using a transmission model to explain spatial spread of disease and identify optimal control interventions. Annals of Internal Medicine, 154, 593–601.

de Valpine, P., Turek, D., Paciorek, C.J., Anderson-Bergman, C., Lang, D.T. & Bodik, R. (2017) Programming with models: writing statistical algorithms for general model structures with NIMBLE. Journal of Computational and Graphical Statistics, 26, 403–413.

Vergne, T., Korennoy, F., Combelles, L., Gogin, A. & Pfeiffer, D.U. (2016) Modelling African swine fever presence and reported abundance in the Russian Federation using national surveillance data from 2007 to 2014. Spatial and Spatiotemporal Epidemiology, 19, 70–77.

Wang, P. (2001) Markov zero-inflated Poisson regression models for a time series of counts with excess zeros. Journal of Applied Statistics, 28, 623–632.

Wangdi, K., Clements, A.C.A., Du, T. & Nery, S.V. (2018) Spatial and temporal patterns of dengue infections in Timor-Leste, 2005–2013. Parasites & Vectors, 11, 9.

Yang, M., Zamba, G.K.D. & Cavanaugh, J.E. (2013) Markov regression models for count time series with excess zeros: a partial likelihood approach. Statistical Methodology, 14, 26–38.

Young, D.S., Roememe, E.S. & Shi, X. (2020) Zero-inflated modeling part II: zero-inflated models for complex data structures. WIREs Computational Statistics, e1540.

**SUPPORTING INFORMATION**

Additional supporting information may be found in the online version of the article at the publisher’s website.

**How to cite this article:** Douwes-Schultz, D. & Schmidt, A.M. (2022) Zero-state coupled Markov switching count models for spatio-temporal infectious disease spread. Journal of the Royal Statistical Society: Series C (Applied Statistics), 71(3), 589–612. Available from: https://doi.org/10.1111/rssc.12547