

# Comparing the spatial attractiveness of hospitals using zero-inflated spatial models

I. Saley<sup>1</sup> · N. Molinari<sup>2</sup> · M. Ribatet<sup>1</sup>

Received: 5 October 2016 / Revised: 26 October 2017 / Accepted: 1 March 2018  
© Springer Science+Business Media, LLC, part of Springer Nature 2018

**Abstract** Policy makers increasingly rely on hospital competition to incentivize patients to choose high-value care. Amongst all possible drivers, the travel distance without any doubt is one of the most important. In this paper we propose the use of a spatial Bayesian hierarchical model to assess the impact of distance on the number of patient admissions in hospitals, and thereby, compare hospital attractiveness. To this aim a MCMC sampler has been designed. We apply our methodology to patient admissions for asthma in four hospitals located in the Hérault department of France. Results indicate that the most attractive hospital is the CHU Montpellier.

**Keywords** Hospital attractiveness · Spatial Bayesian hierarchical model · Zero-inflated data

## 1 Introduction

Competition between hospitals for the provision of health care has been increasingly advocated for its potential to improve efficiency as well as the quality of care (Propper and Leckie 2011). As a consequence of this increasing competition, it is hoped that hospitals decrease production costs and improve the quality of health care delivered so as to attract new patients. The success of this strategy depends on the patients' propensity to switch hospitals in order to choose high-value care. Understanding how patients choose hospitals is, therefore, of major importance for health care providers and policy makers. Hospital attractiveness factors include price, quality, distance (or travel time), waiting time, provider network and others. Early studies identified distance or travel time as the major

---

✉ I. Saley  
saleyissa@yahoo.fr

<sup>1</sup> IMAG,UMR CNRS 5149, Université de Montpellier, 4 place Eugène Bataillon, 34095 Montpellier Cedex 5, France

<sup>2</sup> Service DIM, CHRU de Montpellier, 39 avenue Charles Flahaut, 34295 Montpellier Cedex 5, France

factor negatively affecting hospital choice, including in metropolitan areas where there are many hospitals within short distances (Victoor et al. 2012), even though the sensitivity to distance varies with patient characteristics (age, ethnicity, income and religion) and the type of admission.

As in many European countries, French hospitals are financed through Diagnosis Related Groups based on a prospective payment system. Regarding acute care, this system was fully implemented in 2005 for private-for-profit hospital budgets and in 2008 for public hospital budgets. The reform was intended to improve efficiency and fairness in financing and also to increase competition between and within the public and private sectors (Chevreul et al. 2010).

Information on hospital performance and quality in France hasn't been published yet. Hospital selection is made by the patients themselves, following advice from their general practitioners. Since general practitioners do not face any financial incentives to refer their patients to a given hospital, we can assume that they take into account patient preferences, among other factors. Thus, we can hypothesize that reputation (Jung et al. 2011), as perceived by the patients and the general practitioners, partially reflects the quality of care. In these circumstances, travel distance certainly plays a key role in patient decisions (Varkvisser et al. 2010).

In this context, spatial models describing the number of patient admissions may be useful for hospital classification (from most to least attractive). Indeed in regions where several hospitals can be chosen by the patients, it seems reasonable to think that the selected one meets the patient's needs better than the other hospitals, e.g., price, quality of health care, ease of access, accommodation... In particular, when patients tend to visit a hospital regardless of the distance traveled, it seems reasonable to assume that this hospital can be tagged as "very attractive". From a policy maker standpoint, such conclusions could offer a framework for regulation.

In this paper, we introduce a spatial Bayesian hierarchical model to assess the impact of distance on the number of patient admissions. In the first layer of the model, i.e., *the data layer*, the number of patients admitted to a specific hospital at a given location is modelled by a Poisson distribution whose mean drives how attractive is this hospital. Note however that it is possible that no admission occurs. Typically two different cases may cause such behaviour:

- All patients have decided to go to other hospitals—and hence indicates poor hospital attractiveness;
- All potential subjects are "healthy" or have already been healed by primary health facilities, so that in both cases, there is no need for an "admission process". In such situation, evaluation of the hospital attractiveness is impossible.

To cope with this problem, a common practice is to consider zero inflated models, e.g., Zero Inflated Poisson (ZIP), where a mass  $p$  is considered to capture these extra-zeros (Cohen 1963; Johnson and Kotz 1969). (Lambert 1992) used a ZIP model where both the  $p$  probability and the mean  $\lambda$  of the Poisson distribution can depend on covariates. In this paper we propose to analyse in addition to covariate dependent ZIP model, Zero Inflated Negative Binomial (ZINB) models (Greene 1994).

Another key feature when modelling patients' admission is that the data are intrinsically spatial as a given disease is likely to impact a town as well as its neighbouring cities. (Agarwal et al. 2002) used a spatial zero-inflated Poisson model for areal data where the Poisson mean is modelled by a standard log-linear model. They modelled the spatial

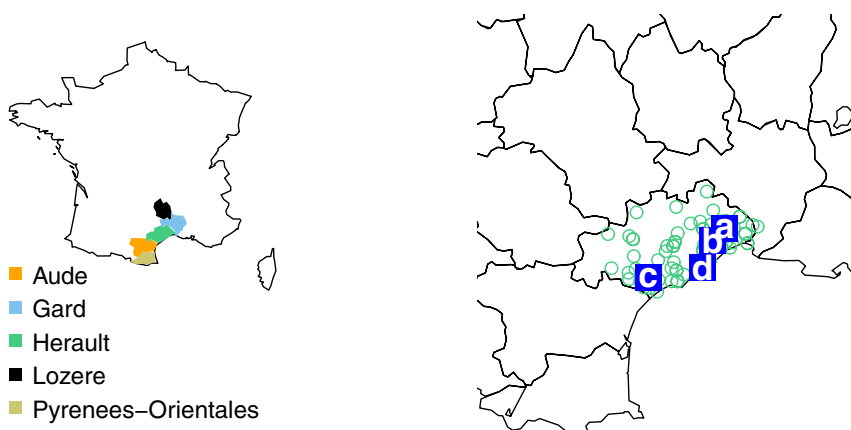
dependence of the data within the random effects that they imposed in the standard log-linear model by a Gaussian Markov field. In this paper, we suppose that the mean of the Poisson distribution,  $\lambda(x)$ , depends on patient residence location  $x$ . In the second layer, i.e., *the process layer*, we model  $\lambda(x)$  as random and assume a multivariate log-normal distribution whose mean vector depends on relevant covariates, e.g., distance between the patient location  $x$  and the hospital, and covariance matrix is derived from an exponential correlation function family. Finally as we are working within a Bayesian framework prior distributions are assumed to each parameters of the model. As often with Bayesian hierarchical models, inference is based on the posterior distribution which can be obtained using Markov Chain Monte Carlo (MCMC) methods.

Section 2 describes the data set analysed in this paper. Section 3 introduces the statistical models under study while Sect. 4 describes the implementation details for our MCMC sampler. Section 5 presents results obtained with an illustrative data set. The paper ends with a brief discussion and supplementary material about the MCMC sampler can be found in Sect. 4.

## 2 The asthma data set

The principal data source used in this paper is the *Programme de Médicalisation des Systèmes d'Information* (PMSI). The PMSI is an administrative data set recording all patient admissions in French private and public hospitals covering all social health insurance programs. In this data set, limited information about the patient and the hospital stay, such as the disease, place of residence, location of the hospital, type of hospital, and travelling distance, are recorded. Therefore, for every region or department in France, we can obtain the number of patients per locality admitted in a given hospital for a given disease.

In this study we focus on the Languedoc–Roussillon region of France. The left panel of Fig. 1 highlights this region which is composed of five departments: Aude, Gard, Hérault, Lozere, Pyrenees–Orientales,



**Fig. 1** Mapping of the illustration data. The left map shows where the Languedoc-Roussillon Region is located in France and subsequently the Hérault department. The picture on the right displays hospital locations (*a*: CHU Montpellier, *b*: Clinique le Millénaire, *c*: CH Béziers and *d*: CH Bassin de Thau) and localities (green points) where at least one resident admission is reported in these hospitals in the study (Color figure online)

Lozère and Pyrénées–Orientales. In 2009, 407 hospital patient stays were recorded for asthma in Hérault hospitals (see PMSI database). The department consists of 343 localities called “communes” (municipalities) spread over 6224 km<sup>2</sup> with a population of approximately 1.1 million, including both rural and metropolitan areas. The Hérault department is known to have a high competitive health care market in France.

In this paper we focus on the main 4 hospitals of Hérault, i.e., with the largest number of asthma patient stays. Among these four medical institutions, only one is private-for-profit while the remaining ones are not-for-profit. The right panel of Fig. 1 plots the locations of these four hospitals as well as the spatial locations of localities where at least one resident admission in these hospitals is reported. Table 1 displays the distribution of the number of admission per hospital.

### 3 Models

As a reminder, the aim of this paper is to provide a method for measuring the effect of distance on the spatial number of patient admissions for a given hospital by using a spatial Bayesian hierarchical model, and to subsequently compare their attractiveness. We suppose that a patient can choose any of the hospitals in competition without any restriction than the way he or she perceives the attractiveness of the hospital.

For the purpose of this study, we suppose that for each localities and each hospital, we have access to the number of admissions. The number of residents per locality as well as the pairwise distance matrix between localities are available. Note however that for the distance between localities, the Euclidean distance was used instead of road distance. Indeed as our application involves a large number of localities, obtaining pairwise road distances between localities is difficult.

#### 3.1 Data layer

Our first model is based on a Poisson distribution to model the number of admissions in a given hospital at a given location. At first sight the use of the Poisson distribution seems particularly relevant as the probability that a single person at risk gets ill and decides to go to a given hospital is rather small and hence justifies considering, at least theoretically, the so-called “law of small numbers”. Let  $Y_h(x_i)$  denotes the number of admissions registered for hospital  $h \in \{1, \dots, H\}$  at the  $i$ th locality, i.e., whose geographic coordinates are  $x_i$ . We assume that  $Y_h(x_i)$  follows a Poisson distribution with mean  $\lambda_h$  where  $\lambda_h$  expresses how attractive the hospital is for patients. Since it is sensible to assume that the function  $x \mapsto \lambda_h(x)$  varies smoothly in space, we assume that

$$\lambda_h : x \mapsto C(R_i)\lambda_{0,h}(\|x_i - x_h\|),$$

**Table 1** Number of admission per hospital for the Asthma data set

Hospital	Number of asthma patient stays
CH Bassin de Thau	71
CH Béziers	79
CHU Montpellier	127
Clinique Le Millénaire	23

for some positive and non-decreasing function  $C$  and where  $R_i$  denotes the number of resident for the  $i$ th locality. One possibility for  $C$  is to take  $C(R_i) = KR_i$ ,  $K > 0$ , where  $K$  is can be thought as the overall probability to be ill, i.e., the so called prevalence ratio, so that  $C(R_i)$  would represent the expected number of resident affected by asthma independently of the geographic location. The function  $d \mapsto \log \lambda_{0,h}(d)$  is typically taken to be a polynomial.

As stated in the introduction, the number of admissions,  $Y_h(x_i)$ , can be equal to zero in two different cases: a lack of attractiveness of hospital  $h$  (all patients have chosen to go to other hospitals); nobody is affected by asthma in the  $i$ th locality or have been healed by primary health facilities so that there is no need for an “admission process”. This feature induces an extra probability mass at 0 that needs to be taken into account in our modelling strategy. Consequently, it makes sense to use a mass  $p(x_i)$  depending on locality  $i$  for zeros not related to attractiveness so that the first layer of our model becomes

$$Y_h(x_i) \mid p(x_i), \lambda_h(x_i) \sim \begin{cases} p(x_i) + \{1 - p(x_i)\} \exp\{-\lambda_h(x_i)\}, & Y_h(x_i) = 0 \\ \{1 - p(x_i)\} \frac{\lambda_h(x_i)^{Y_h(x_i)}}{Y_h(x_i)!} \exp\{-\lambda_h(x_i)\}, & Y_h(x_i) > 0, \end{cases} \quad (1)$$

where  $k \mapsto k!$  corresponds to the factorial function.

A widely used alternative to the zero inflated Poisson model is the zero inflated negative binomial model. As a mixture of Poisson and Gamma distributions, there are some situations where the negative binomial distribution is better suited than the Poisson distribution. Keeping the same notations as in (1), the data layer for this model is

$$Y_h(x_i) \mid p(x_i), r, \lambda_h(x_i) \sim \begin{cases} p(x_i) + \{1 - p(x_i)\} \left\{1 + \frac{\lambda_h(x_i)}{r}\right\}^{-r}, & Y_h(x_i) = 0 \\ \{1 - p(x_i)\} \frac{\Gamma\{Y_h(x_i) + r\}}{Y_h(x_i)! \Gamma(r)} \left\{1 + \frac{\lambda_h(x_i)}{r}\right\}^{-r} \left\{\frac{\lambda_h(x_i)}{r + \lambda_h(x_i)}\right\}^{Y_h(x_i)}, & Y_h(x_i) > 0, \end{cases} \quad (2)$$

where  $\Gamma(\cdot)$  is the Gamma function. One recover the zero inflated Poisson model as  $r \rightarrow \infty$ .

### 3.2 Process layer

It is sensible to assume that residents from neighbouring localities will have and share the same habits and decisions. For instance, persons can take into consideration what their neighbours opinions/experience before choosing a hospital. It is therefore reasonable to assume that the random vector  $\lambda_h(\mathbf{x}) = \{\lambda_h(x_1), \dots, \lambda_h(x_n)\}^\top$  exhibits some kind dependence. Mathematically speaking we assume that  $\lambda_h(\mathbf{x})$  follows a multivariate log-normal distribution so that one can benefit of the so-called *borrowing strength* of hierarchical models to characterize  $\lambda_h(x_i)$  for some  $i \in \{1, \dots, n\}$ . We then have

$$\lambda_h(\mathbf{x}) \mid \mu_h(\mathbf{x}), \gamma_h(\mathbf{x}), C(R) \sim (2\pi)^{-n/2} |\gamma_h(\mathbf{x})|^{-1/2} \times \prod_{i=1}^n \lambda_h(x_i)^{-1} \exp \left\{ -\frac{1}{2} [\log \lambda_h(\mathbf{x}) - \mu_h(\mathbf{x})]^\top \gamma_h(\mathbf{x})^{-1} [\log \lambda_h(\mathbf{x}) - \mu_h(\mathbf{x})] \right\}, \quad (3)$$

where  $\mu_h(\mathbf{x})$  is the mean vector of the log-normal distribution and  $\gamma_h(\mathbf{x}) = \{\gamma_h(x_i - x_j)\}_{i,j=1,\dots,n}$  its covariance matrix.

Since  $\log \lambda_h(x_i) = \log C(R_i) + \log \lambda_{0,h}(x_i)$ , we have conditionally on  $C(R_i)$ ,  $\log \lambda_{0,h}(\mathbf{x}) \sim N\{\mu'(\mathbf{x}), \gamma_h(\mathbf{x})\}$  with  $\mu(\mathbf{x}_i) = \mu'(\mathbf{x}_i) + \log C(R_i)$ .

Throughout this study, we assume that

$$\mu'_h(x_i) = \beta'_{0,h} + \beta_{1,h}\|x_i - x_h\|, \quad i = 1, \dots, n, \quad h = 1, \dots, H,$$

where  $x_h$  denotes the geographic coordinates of hospital  $h \in \{1, \dots, H\}$ . This implies that, for any  $i \in \{1, \dots, n\}$ ,

$$\begin{aligned} \mu_h(x_i) &= \beta'_{0,h} + \log R_i + \log K + \beta_{1,h}\|x_i - x_h\| \\ &= \beta_{0,h} + \log R_i + \beta_{1,h}\|x_i - x_h\|, \end{aligned}$$

where  $\beta_{0,h} = \beta'_{0,h} + \log K$  so that there is no need to impose any distribution on the unknown prevalence ratio  $K$ .

Among all parameter,  $\beta_{1,h}$  is of primary importance for our study as it controls how the log-mean of the Poisson/Negative binomial distribution varies with the distance from the hospital and therefore summarize to some extent how attractive is this hospital. As a consequence it is expected that the larger the  $\beta_{1,h}$ , the more attractive is hospital  $h$ . We can rank hospital attractiveness by ranking their respective summary of the posterior distribution of  $\beta_{1,h}$ , e.g., posterior mean.

To avoid over pasteurization, it is often more convenient to add parametric structure for the covariance function. In this study we assume an isotropic exponential covariance function family for  $\gamma_h(\mathbf{x})$ , i.e.,

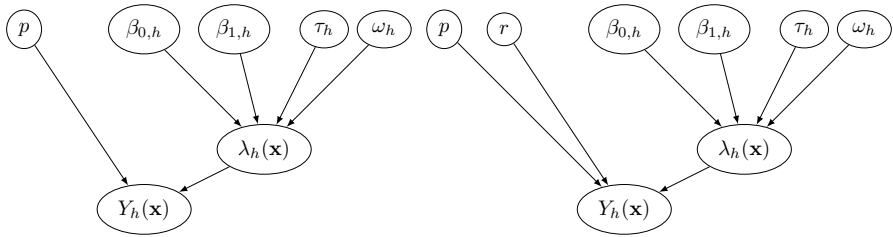
$$\gamma_h(x_i - x_j) = \tau_h \exp\left(-\frac{\|x_i - x_j\|}{\omega_h}\right),$$

where  $\tau_h > 0$  is the variance and  $\omega_h > 0$  the scale parameter. It is well known that considering covariance function families with a shape parameter such as the powered exponential or Whittle–Matérn induces some non-identifiability issues (Zhang 2004; Sang and Gelfand 2010; Davison et al. 2012). The exponential covariance family appear to be a good trade off between identifiability problems and flexibility.

Although a spatial correlation could be assumed on  $p(x_i)$  and to avoid over parametrized models, we treat  $p$  as spatially constant and independent of the hospital so that we will write  $p$  instead of  $p(x_i)$ . A preliminary study, not presented here, showed that the  $r$  parameter was mainly the same for all 4 hospitals. Consequently throughout this paper we will consider a zero inflated negative binomial model where the  $r$  parameter is constant across hospitals. To have a complete picture of the proposed models, Fig. 2 gives the direct acyclic graph for the zero inflated Poisson/negative binomial models.

### 3.3 Parameters prior distributions

Since we are working in a Bayesian framework, prior distributions have to be assumed on each parameters of the model. In this study we assign independent priors for each parameters, i.e.,  $p$ ,  $r$ ,  $\beta_{0,h}$ ,  $\beta_{1,h}$ ,  $\tau_h$  and  $\omega_h$ . For numerical reasons we use a reparametrization for  $p$  by letting  $p = 1/(1 + e^{-\theta})$  and thus assign a Gaussian prior distribution on  $\theta$  instead of  $p$ . Whenever possible we use conjugate priors to speed up our MCMC sampler. More precisely a non informative Gaussian prior distribution is assumed for  $\beta_{0,h}$  and  $\beta_{1,h}$ . For the zero inflated negative binomial model a uniform distribution on the interval  $(0, 100)$  was assumed for  $r$ . More care is need when defining prior distributions for  $\tau_h$  and  $\omega_h$ . Indeed



**Fig. 2** Direct acyclic graph of the Bayesian hierarchical models. Left: Zero inflated Poisson model. Right: Zero inflated negative binomial model

(Berger et al. 2001) and (Banerjee et al. 2004) showed that improper priors for these parameters, can, in general, result in improper posterior distributions. Following (Banerjee et al. 2004) who suggested using informative prior distributions, we assumed an Inverse-Gamma distribution with shape and scale parameters equal to 2 for both  $\tau_h$  and  $\omega_h$ . As this specific choice for the prior distribution is somehow arbitrary, we conducted a sensitivity analysis about the impact of such prior distribution in Sect. 5.2.

### 4 Full conditional distribution

To ease the notation let  $\lambda(\mathbf{x}) = \{\lambda_1(\mathbf{x}), \dots, \lambda_H(\mathbf{x})\}$ ,  $\beta_0 = (\beta_{0,1}, \dots, \beta_{0,H})$ ,  $\beta_1 = (\beta_{1,1}, \dots, \beta_{1,H})$ ,  $\tau = (\tau_1, \dots, \tau_H)$  and  $\omega = (\omega_1, \dots, \omega_H)$ . Similarly we write  $\lambda_{-h}(\mathbf{x})$ ,  $\beta_{0,-h}$ ,  $\beta_{1,-h}$ ,  $\tau_{-h}$  and  $\omega_{-h}$  to denote the corresponding vector without the  $h$ th element. As often with hierarchical models, the likelihood involves multiple integrals so that likelihood-based inference is not possible. To bypass this hurdle, a widely used strategy is to have resort to Monte Carlo Markov Chain algorithms whose aim is to generate a Markov chain whose stationary distribution is the posterior distribution.

For our purposes, a particularly well suited MCMC algorithm is the Gibbs sampler (Gelfand and Smith 1990). This sampler is based on the full conditional distributions which turn out to be for the zero inflated Poisson model

$$\begin{aligned} \pi\{\theta \mid y(\mathbf{x}), \lambda(\mathbf{x}), \beta_0, \beta_1, \tau, \omega\} &\propto \pi\{y(\mathbf{x}) \mid \theta, \lambda(\mathbf{x})\}\pi(\theta) \\ \pi\{\lambda_h(\mathbf{x}) \mid y(\mathbf{x}), \lambda_{-h}(\mathbf{x}), \theta, \beta_0, \beta_1, \tau, \omega\} &\propto \pi\{y(\mathbf{x}) \mid \theta, \lambda(\mathbf{x})\}\pi\{\lambda_h(\mathbf{x}) \mid \beta_{0,h}, \beta_{1,h}, \tau_h, \omega_h\} \\ \pi\{\beta_{0,h} \mid y(\mathbf{x}), \theta, \lambda(\mathbf{x}), \beta_{0,-h}, \beta_1, \tau, \omega\} &\propto \pi\{\lambda_h(\mathbf{x}) \mid \beta_{0,h}, \beta_{1,h}, \tau_h, \omega_h\}\pi(\beta_{0,h}) \\ \pi\{\beta_{1,h} \mid y(\mathbf{x}), \theta, \lambda(\mathbf{x}), \beta_0, \beta_{1,-h}, \tau, \omega\} &\propto \pi\{\lambda_h(\mathbf{x}) \mid \beta_{0,h}, \beta_{1,h}, \tau_h, \omega_h\}\pi(\beta_{1,h}) \\ \pi\{\tau_h \mid y(\mathbf{x}), \theta, \lambda(\mathbf{x}), \beta_0, \beta_1, \tau_{-h}, \omega\} &\propto \pi\{\lambda_h(\mathbf{x}) \mid \beta_{0,h}, \beta_{1,h}, \tau_h, \omega_h\}\pi(\tau_h) \\ \pi\{\omega_h \mid y(\mathbf{x}), \theta, \lambda(\mathbf{x}), \beta_0, \beta_1, \tau, \omega_{-h}\} &\propto \pi\{\lambda_h(\mathbf{x}) \mid \beta_{0,h}, \beta_{1,h}, \tau_h, \omega_h\}\pi(\omega_h). \end{aligned}$$

Similar expressions are readily obtained for the zero inflated negative binomial model.

Note that since many of the above full conditional distributions are not related to well identified distribution such as Gaussian or Gamma distributions, Metropolis-Hasting updating scheme might be used in various places (Gelman et al. 2004). When this is the case a random walk, eventually of the log-scale for non-negative parameters, with Gaussian innovation was used. This Metropolis-Hastings within Gibbs algorithm was written in R (R Core Team 2017) and could be provided upon request to the authors.

## 5 Application to the asthma data set

In this section, we apply the methodology introduced in Sects. 3 and 4 to the asthma data of Sect. 2. Recall that our analysis focuses on 4 hospitals: CH Bassin de Thau in the city of Sète, CH Béziers in the city of Béziers and both CHU Montpellier and Clinique le Millénaire in the city of Montpellier. The localities used for estimation are all the 343 localities within the Hérault department of France.

### 5.1 Results

#### 5.1.1 MCMC chains and estimates

For convergence control and parameter estimation, 10 Markov chains of length 50,000 were generated. For each chain, the first 30,000 iterations were treated as burn-in-period and a thinning lag of 10 iterations was used to mitigate serial dependence in the simulated Markov chains. At the end of this post-processing of our MCMC sampler, a chain of length 2000 that could be reasonably considered as sampled from the posterior distribution and showing very weak serial dependence was used. To double check if convergence to the stationary distribution was reached, the  $\hat{R}$  statistics (Aho 2015) was used. Results indicates that convergence to the stationary distribution, i.e., the posterior distribution, can be safely assumed.

Table 2 reports the estimated posterior means and its standard error obtained from these 10 Markov chains.

#### 5.1.2 Model selection

One first question of interest is to determine if one should prefer the zero inflated negative binomial model over its Poisson counterpart. To choose the model that better fits our data from among the ZIP and ZINB models, we used the Deviance Information Criterion (DIC), (Spiegelhalter et al. 2002). The smaller the DIC, the better the model fits the data. These DIC values are reported in Table 2 as well as their respective standard errors. Although the

**Table 2** Posterior means for the ZIP/ZINB parameters and associated Standard Errors (SE) obtained from 10 MCMC chains for the asthma data set. The Deviance Information Criterion (DIC) for each model is also reported

Hospital	$\beta_{0,h}$	$\beta_{1,h}$	$\tau_h$	$\omega_h$	$\tau_h/\omega_h$
ZIP Model: $p = 0.58$ (0.02), DIC = 653 (6)					
CH Bassin de Thau	- 8.18 (0.26)	- 0.058 (0.009)	1.03 (0.32)	2.5 (2.0)	0.91 (0.14)
CH Béziers	- 6.99 (0.15)	- 0.052 (0.004)	1.09 (0.29)	2.5 (2.5)	0.93 (0.15)
CHU Montpellier	- 8.21 (0.14)	- 0.015 (0.002)	0.92 (0.15)	2.4 (1.4)	0.95 (0.16)
Clinique Le Millénaire	- 9.56 (0.11)	- 0.019 (0.002)	0.85 (0.20)	1.4 (0.4)	0.92 (0.18)
ZINB Model: $p = 0.59$ (0.02), $r = 47.2$ (2.7), DIC = 669 (8)					
CH Bassin de Thau	- 8.04 (0.14)	- 0.055 (0.010)	1.04 (0.29)	2.76 (1.23)	0.804 (0.17)
CH Béziers	- 7.05 (0.10)	- 0.055 (0.004)	1.04 (0.23)	1.98 (0.63)	0.95 (0.19)
CHU Montpellier	- 8.34 (0.05)	- 0.012 (0.002)	0.94 (0.09)	2.16 (1.20)	0.89 (0.17)
Clinique Le Millénaire	- 9.76 (0.11)	- 0.018 (0.002)	0.77 (0.08)	1.61 (0.90)	0.82 (0.20)



difference is small, we can see that the DIC value tends to be in favour of the ZIP model. This conclusion could be expected since the posterior mean for  $r$  in the ZINB model is large and it is well known that the negative binomial distribution converges to the Poisson distribution as  $r \rightarrow \infty$ . Consequently and in accordance with Ockam’s razor, the DIC tends to prefer the simpler model, i.e., Poisson.

To confirm this conclusion we also analyse prediction performance for each model based on the Mean Squared Error criterion

$$MSE_h = \frac{1}{343} \sum_{i=1}^{343} \{y_h(x_i) - \hat{y}_h(x_i)\}^2, \quad h = 1, \dots, H.$$

where  $y_h(x_i)$  is the observed value and  $\hat{y}_h(x_i)$  the computed posterior mean at locality  $i$ . Table 3 shows the computed MSE for both models. Although both models show good performance, the MSE criteria tends to be in favour our the ZIP model and is in agreement with our previous conclusions. Based on these results we will therefore consider only the ZIP model for the rest of this study.

### 5.1.3 Comparing hospital attractiveness

As stated in Sect. 3, hospital attractiveness can be achieved by inspecting the posterior distribution of  $\beta_{0,h}$  and  $\beta_{1,h}$ . Of primary importance is  $\beta_{1,h}$  as it controls how varies the log-mean of the process as one goes away to the hospital while  $\beta_{0,h}$  characterizes attractiveness at “home city”, i.e., when  $\|x - x_h\| = 0$ . For example, from Table 2 one can see that Clinique le Millénaire is the less attractive at home city, due probably to the presence of CHU Montpellier in the same city. CH Béziers seems to be more attractive at home city; it is practically the only well-known center in the city of Béziers.

Table 2 shows that the posterior mean of  $\beta_{1,h}$  are negative for all hospitals. This indicates that the mean of the Poisson distribution, or equivalently hospital attractiveness, decreases as one moves away from the hospital. Such behaviour could be expected since, as identified by prior studies (Victoor et al. 2012), distance is a factor that negatively affects hospital choice. Although the Clinique Le Millénaire is slightly less attractive, the CHU Montpellier appears to be the most attractive hospital. Such findings could be expected since the CHU Montpellier is a well-known, large public health center and has the largest accommodation capacity in the region. It is also known for delivering high quality care: the faculty of medicine from which it depends, is one of the oldest in Europe (founded in the 13th century). The city of Montpellier also has easy access: highway A9, railways (TGV and TER) and many other roads for access. The less attractive hospital is the CH Bassin de Thau for which more than half of admissions are from the home city (Sète) and only five localities out of 343 do not have zero admissions. This behaviour can be explained by the

**Table 3** Mean squared error (MSE) and associated standard errors (in bracket) for the ZIP and ZINB models on the asthma data set

	ZIP	ZINB
CH Bassin de Thau	1.71 (1.09)	2.87 (1.31)
CH Béziers	0.85 (0.20)	1.02 (0.28)
CHU Montpellier	0.39 (0.10)	0.44 (0.08)
Clinique Le Millénaire	0.10 (0.02)	0.13 (0.04)

fact that the CH Bassin de Thau is a small public center and is located between Montpellier and Béziers where the 2 largest public centers in the department are found. Sète also lacks the transportation facilities available in the other 2 cities : no TGV or highway. Despite its 22 admissions, the Clinique le Millénaire seems to draw patients from farther away than CH Bassin de Thau and CH Béziers. The Clinique Le Millénaire is a well-known private center in Montpellier and its neighbourhood. It has a good reputation. The small number of admissions can be explained by a smaller accommodation capacity and it is, like other private centers, more specialized in acute health care such as surgery. It is also in the same city with the biggest well-known center: CHU Montpellier. Globally, the proposed center ranking seems sensible.

## 5.2 Sensitivity analysis

As stated in Sect. 3.3, to avoid improper posterior distribution one has to assume informative prior distribution on  $\tau_h$  and  $\omega_h$ . Since no external prior information is available for these parameter, the definition of informative prior distribution is somehow artificial. In this section we therefore investigate if posterior inference is affected by a change in these prior distributions.

To this aim, we used more and less informative priors for  $\tau_h$  and  $\omega_h$  compared to that used in Sect. 3.3 and check whether our conclusions remains valid. Table 4 reports the evolution of the posterior means when the prior distribution for  $\tau_h$  and  $\omega_h$  changes to an Inverse-Gamma(3,3) to an Inverse-Gamma (1,1)—when the prior distribution is less informative. While the posterior means for  $\beta_{0,h}$  and  $\beta_{1,h}$  appears to be fairly robust to the prior choice, the posterior means for  $\omega_h$  and  $\tau_h$  appear to be strongly affected.

It is well known the the parameter  $\tau_h$  and  $\omega_h$  are typically negatively correlated (Zhang 2004). For instance, a large variance  $\tau_h$  combined with a large scale parameter  $\omega_h$  may lead to the same overall variability than a smaller variance and small scale parameter. It is therefore more appropriate to analyse the sensitivity of the ratio  $\tau_h/\omega_h$  to the prior distribution definition. Table 5 shows the evolution of the posterior mean of this ratio as the prior distributions for  $\tau_h$  and  $\omega_h$  vary. Although at first sight some differences are noticeable, taking into account the associated uncertainties we can see

**Table 4** Posterior means for ZIP parameters and associated Standard Errors over 10 MCMC chains for the asthma data set as the prior distributions for  $\tau_h$  and  $\omega_h$  vary

	$\beta_{0,h}$	$\beta_{1,h}$	$\tau_h$	$\omega_h$
Prior: Inverse-Gamma (3,3), $p = 0.574$ (0.021)				
CH Bassin de Thau	- 8.07 (0.14)	- 0.057 (0.003)	0.99 (0.25)	1.68 (0.63)
CH Béziers	- 7.04 (0.03)	- 0.054 (0.002)	1.21 (0.15)	1.75 (0.41)
CHU Montpellier	- 8.42 (0.03)	- 0.016 (0.002)	0.95 (0.08)	1.27 (0.22)
Clinique Le Millénaire	- 9.73 (0.11)	- 0.016 (0.003)	0.94 (0.08)	1.25 (0.17)
Prior: Inverse-Gamma (1,1), $p = 0.576$ (0.013)				
CH Bassin de Thau	- 8.13 (0.17)	- 0.056 (0.004)	1.00 (0.35)	37.24 (4.69)
CH Béziers	- 7.11 (0.07)	- 0.053 (0.003)	1.11 (0.22)	6.92 (4.26)
CHU Montpellier	- 8.24 (0.04)	- 0.015 (0.002)	0.96 (0.21)	4.38 (3.68)
Clinique Le Millénaire	- 9.82 (0.05)	- 0.016 (0.005)	0.84 (0.26)	2.61 (2.18)

**Table 5** Table of posterior means for the ratio  $\tau_h/\omega_h$  and associated Standard Errors (SE) over 10 MCMC chains as the prior distributions vary

	Inverse-Gamma (1,1)	Inverse-Gamma (2,2)	Inverse-Gamma (3,3)
CH Bassin de Thau	0.55 (0.35)	0.91 (0.14)	0.86 (0.13)
CH Béziers	0.71 (0.32)	0.93 (0.15)	1.06 (0.07)
CHU Montpellier	0.91 (0.48)	0.95 (0.16)	1.01 (0.08)
Clinique le Millénaire	0.88 (0.12)	0.92 (0.18)	1.00 (0.10)

that the ratio  $\tau_h/\omega_h$  appears to be fairly robust to the prior definition—the ratio tends to be larger as the prior distribution gets increasingly more informative though.

## 6 Discussion and conclusions

In this paper, we have presented a spatial Bayesian hierarchical model based on zero inflated distribution to model count data. The two proposed models were applied to hospital admission for asthma in the Hérault department of France. To our knowledge, it is the first numerical method to measure a distance effect on patient admissions or other similar cases, more importantly in the case of the diagnosis related groups as defined in France. For a given locality or in a relatively homogeneous area, these models can be used to compare hospital spatial attractiveness and thus to classify them from the least to the most attractive. Such ranking can be used to understand the impact of implemented reforms. Plausible ranks are obtained when comparing CHU Montpellier (first), Clinique le Millénaire (second), CH Béziers (third), and CH Bassin de Thau (least) according to the admissions due to asthma in 2009 in the Hérault. However, certain issues associated with the method are:

- Run-time: as it often the case for MCMC modules, run-times can be long. For example, for our own R software codes, with 4 hospitals and 343 localities, a chain of length 50,000 can take over 60 h for our computer with a 2.60 GHz Intel (R) Core (TM) i-3230M CPU and 4 Go of memory;
- Our model does not take into consideration the presence of other hospitals. The processes are only linked with the mass of extra-zeros ( $p$ ) and  $r$ . It would be good to also consider covariates that distinguish hospitals such as the size or type of hospital;
- We used a linear distance function, which might be very restrictive. Further studies including both travelling time as a covariate and other functional relations such as step functions, polynomial, ...should be performed.

### Compliance with ethical standards

**Conflict of interest** We, the authors, declare that we have no conflicts of interest.

**Human and animals rights** That this article does not contain any studies with human participants or animals performed by any of us.

## Appendix A

In this section, we give the derivation of conditional parameters distributions for the ZIP model. To obtain the derivation for the ZINB model, the distributions used in the ZIP model have to be replaced by those used in the ZINB model. As a reminder, (a) Bayes formula is

$$\pi(\theta | x) = \frac{\pi(x | \theta)\pi(\theta)}{\int \pi(x | \theta)\pi(\theta) d\theta} \propto \pi(x | \theta)\pi(\theta),$$

where  $\pi(\theta | x)$  is called the posterior distribution,  $\pi(\theta)$  the prior distribution and  $\pi(x | \theta)$ , the likelihood. b) and the prior distributions used in this case study for the ZIP model are

$$\begin{aligned} \beta_{0,h} &\sim \frac{1}{10^3\sqrt{2\pi}} \exp\left\{-\frac{\beta_{0,h}^2}{2*10^6}\right\} & \beta_{1,h} &\sim \frac{1}{10^3\sqrt{2\pi}} \exp\left\{-\frac{\beta_{1,h}^2}{2*10^6}\right\} \\ \tau_h &\sim \frac{1}{\Gamma(1)}(1/\tau_h)^2 \exp\{-1/\tau_h\} & \omega_h &\sim \frac{1}{\Gamma(1)}(1/\omega_h)^2 \exp\{-1/\omega_h\} \\ \theta &\sim \frac{1}{10^3\sqrt{2\pi}} \exp\left\{-\frac{\theta^2}{2*10^6}\right\} \end{aligned}$$

Let's call  $\pi\{y(\mathbf{x}), \theta, \lambda(\mathbf{x}), \beta_0, \beta_1, \tau, \omega\}$  (with  $\lambda(\mathbf{x}), \beta_0, \beta_1, \tau$  and  $\omega$  as defined in Sect. 4), the joint distribution for model parameters and the data. Using the definition of conditional probabilities and assuming parameters independence within their prior distributions, we show that

$$\begin{aligned} \pi\{y(\mathbf{x}), \theta, \lambda(\mathbf{x}), \beta_0, \beta_1, \tau, \omega\} &= \pi\{y(\mathbf{x}) | \theta, \lambda(\mathbf{x})\} \pi\{\theta\} \\ &\times \prod_{h=1}^4 \pi\{\lambda_h(\mathbf{x}) | \beta_{0,h}, \beta_{1,h}, \tau_h, \omega_h\} \pi\{\beta_{0,h}\} \pi\{\beta_{1,h}\} \pi\{\tau_h\} \pi\{\omega_h\} \end{aligned} \tag{4}$$

Then using Bayes formula and Eq. 4, we derive the conditional distributions

$$\begin{aligned} \pi\{\theta | y(\mathbf{x}), \lambda(\mathbf{x}), \beta_0, \beta_1, \tau, \omega\} &\propto \pi\{y(\mathbf{x}) | \theta, \lambda(\mathbf{x})\} \pi(\theta) \\ \pi\{\lambda_h(\mathbf{x}) | y(\mathbf{x}), \lambda_{-h}(\mathbf{x}), \theta, \beta_0, \beta_1, \tau, \omega\} &\propto \pi\{y(\mathbf{x}) | \theta, \lambda(\mathbf{x})\} \times \\ &\pi\{\lambda_h(\mathbf{x}) | \beta_{0,h}, \beta_{1,h}, \tau_h, \omega_h\} \\ \pi\{\beta_{0,h} | y(\mathbf{x}), \theta, \lambda(\mathbf{x}), \beta_{0,-h}, \beta_1, \tau, \omega\} &\propto \pi\{\lambda_h(\mathbf{x}) | \beta_{0,h}, \beta_{1,h}, \tau_h, \omega_h\} \pi(\beta_{0,h}) \\ \pi\{\beta_{1,h} | y(\mathbf{x}), \theta, \lambda(\mathbf{x}), \beta_0, \beta_{1,-h}, \tau, \omega\} &\propto \pi\{\lambda_h(\mathbf{x}) | \beta_{0,h}, \beta_{1,h}, \tau_h, \omega_h\} \pi(\beta_{1,h}) \\ \pi\{\tau_h | y(\mathbf{x}), \theta, \lambda(\mathbf{x}), \beta_0, \beta_1, \tau_{-h}, \omega\} &\propto \pi\{\lambda_h(\mathbf{x}) | \beta_{0,h}, \beta_{1,h}, \tau_h, \omega\} \pi(\tau_h) \\ \pi\{\omega_h | y(\mathbf{x}), \theta, \lambda(\mathbf{x}), \beta_0, \beta_1, \tau, \omega_{-h}\} &\propto \pi\{\lambda_h(\mathbf{x}) | \beta_{0,h}, \beta_{1,h}, \tau_h, \omega_h\} \pi(\omega_h) \end{aligned}$$

which have to be used in the Gibbs sampling. To obtain the explicit form of these distributions, every term in the right members has to be replaced by its mathematical expression. Thus, let's denote  $n_{1,h}$  the number of locations where we count zero stays for a hospital

$h$  and  $n_{2,h}$  the number of locations where we have at least one stay in the hospital. From Eq. 1, Sect. 3.1,

$$\begin{aligned} \pi(y(\mathbf{x}) \mid \theta, \lambda(\mathbf{x})) &= \prod_{h=1}^4 \prod_{i=1}^{n_{1,h}} \left[ \frac{1}{1 + \exp\{-\theta\}} + \left( 1 - \frac{1}{1 + \exp\{-\theta\}} \right) \exp\{-\lambda_h(x_i)\} \right] \\ &\times \prod_{j=1}^{n_{2,h}} \left[ \left( 1 - \frac{1}{1 + \exp\{-\theta\}} \right) \frac{\lambda_h(x_j)^{y(x_j)}}{y(x_j)!} \exp\{-\lambda_h(x_j)\} \right] \end{aligned} \tag{5}$$

and by definition in Sect. 3.2

$$\begin{aligned} \pi\{\lambda_h(\mathbf{x}) \mid \beta_{0,h}, \beta_{1,h}, \tau_h, \omega_h\} &= (2\pi)^{-n/2} |\gamma_h(\mathbf{x})|^{-1/2} \\ &\times \prod_{i=1}^n \lambda_h^{-1}(x_i) \exp\left\{ -\frac{1}{2} [\log \lambda_h(\mathbf{x}) - \mu_h(\mathbf{x})]^T \gamma_h(\mathbf{x})^{-1} [\log \lambda_h(\mathbf{x}) - \mu_h(\mathbf{x})] \right\} \end{aligned} \tag{6}$$

By replacing every term by its explicit expression, we obtain

$$\begin{aligned} \pi(\theta \mid y(\mathbf{x}), \lambda(\mathbf{x}), \beta_0, \beta_1, \tau, \omega) &\propto \prod_{h=1}^4 \prod_{i=1}^{n_{1,h}} \left[ \frac{1}{1 + \exp\{-\theta\}} + \frac{\exp\{-\theta\}}{1 + \exp\{-\theta\}} \exp\{-\lambda_h(x_i)\} \right] \\ &\times \left[ \frac{\exp\{-\theta\}}{1 + \exp\{-\theta\}} \right]^{n_{2,h}} \exp\left\{ -\frac{\theta^2}{2 * 10^6} \right\} \\ \pi(\beta_{0,h} \mid y(\mathbf{x}), \lambda(\mathbf{x}), \beta_{0,-h}, \beta_1, \tau, \omega, \theta) &\propto \exp\left\{ -\frac{1}{2} [\log \lambda_h(\mathbf{x}) - \mu_h(\mathbf{x})]' \gamma_h(\mathbf{x})^{-1} [\log \lambda_h(\mathbf{x}) - \mu_h(\mathbf{x})] \right\} \\ &\times \exp\left\{ -\frac{\beta_{0,h}^2}{2 * 10^6} \right\} \\ \pi(\beta_{1,h} \mid y(\mathbf{x}), \lambda(\mathbf{x}), \beta_0, \beta_{1,-h}, \tau, \omega, \theta) &\propto \exp\left\{ -\frac{1}{2} [\log \lambda_h(\mathbf{x}) - \mu_h(\mathbf{x})]' \gamma_h(\mathbf{x})^{-1} [\log \lambda_h(\mathbf{x}) - \mu_h(\mathbf{x})] \right\} \\ &\times \exp\left\{ -\frac{\beta_{1,h}^2}{2 * 10^6} \right\} \\ \pi(\tau_h \mid y(\mathbf{x}), \theta, \lambda(\mathbf{x}), \beta_0, \beta_1, \tau_{-h}, \omega) &\propto \exp\left\{ -\frac{1}{2} [\log \lambda_h(\mathbf{x}) - \mu_h(\mathbf{x})]' \gamma_h(\mathbf{x})^{-1} [\log \lambda_h(\mathbf{x}) - \mu_h(\mathbf{x})] \right\} \\ &\times |\gamma_h(\mathbf{x})|^{-1/2} (1/\tau_h)^2 \exp\{-1/\tau_h\} \\ \pi(\omega_h \mid y(\mathbf{x}), \theta, \lambda(\mathbf{x}), \beta_0, \beta_1, \tau, \omega_{-h}) &\propto \exp\left\{ -\frac{1}{2} [\log \lambda_h(\mathbf{x}) - \mu_h(\mathbf{x})]' \gamma_h(\mathbf{x})^{-1} [\log \lambda_h(\mathbf{x}) - \mu_h(\mathbf{x})] \right\} \\ &\times |\gamma_h(\mathbf{x})|^{-1/2} (1/\omega_h)^2 \exp\{-1/\omega_h\} \\ \pi(\lambda_h(\mathbf{x}) \mid y(\mathbf{x}), \theta, \lambda_{-h}(\mathbf{x}), \beta_0, \beta_1, \tau, \omega) &\propto \exp\left\{ -\frac{1}{2} [\log \lambda_h(\mathbf{x}) - \mu_h(\mathbf{x})]' \gamma_h(\mathbf{x})^{-1} [\log \lambda(\mathbf{x}) - \mu(\mathbf{x})] \right\} \\ &\times \prod_{k=1}^n \lambda_h^{-1}(x_k) \prod_{j=1}^{n_{2,h}} \left[ \frac{\lambda_h(x_j)^{y(x_j)}}{y(x_j)!} \exp\{-\lambda_h(x_j)\} \right] \\ &\times \prod_{i=1}^{n_{1,h}} \left[ \frac{1}{1 + \exp\{-\theta\}} + \frac{\exp\{-\theta\}}{1 + \exp\{-\theta\}} \exp\{-\lambda_h(x_i)\} \right] \end{aligned}$$

## References

- Agarwal, D.K., Gelfand, A.E., Citron-Pousty, S.: Zero-inflated models with application to spatial count data. *Environ. Ecol. Stat.* **9**(4), 341–355 (2002)
- Aho, K.: *asbio: a collection of statistical tools for biologists*. R Package Version 1. 1–5 (2015)
- Banerjee, S., Carlin, B., Gelfand, A.: *Hierarchical modeling and analysis for spatial statistics*. Chapman & Hall/CRC, New York (2004)
- Berger, J.O., De Oliveira, V., Sansó, B.: Objective bayesian analysis of spatially correlated data. *J. Am. Stat. Assoc.* **96**(456), 1361–1374 (2001)
- Chevreur, K., Durand-Zaleski, I., Bahrami, S., Hernández-Quevedo, C., Mladovsky, P.: Health system review. *Erasmus* **12**, 93–103 (2010)
- Cohen, A.C.: *Estimation in mixtures of discrete distributions*. Statistical Pub. Society (1963)
- Davison, A., Padoan, S., Ribatet, M.: Statistical modelling of spatial extremes. *Stat. Sci.* **7**(2), 161–186 (2012)
- Gelfand, A.E., Smith, A.F.: Sampling-based approaches to calculating marginal densities. *J. Am. Stat. Assoc.* **85**(410), 398–409 (1990)
- Gelman, A., Carlin, J.B., Stern, H.S., Rubin, D.B.: *Bayesian data analysis*. In: *Texts in Statistical Science Series* (2004)
- Greene, W.H.: *Accounting for excess zeros and sample selection in poisson and negative binomial regression models* (1994)
- Johnson, N.L., Kotz, S.: *Distributions in Statistics*. Wiley, Hoboken (1969)
- Jung, K., Feldman, R., Scanlon, D.: Where would you go for your next hospitalization? *J. Health Econ.* **30**(4), 832–841 (2011)
- Lambert, D.: Zero-inflated poisson regression, with an application to defects in manufacturing. *Technometrics* **34**(1), 1–14 (1992)
- Propper, C., Leckie, G.: *Increasing competition between providers in health care markets: the economic evidence* (2011)
- R Core Team: *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna (2017)
- Sang, H., Gelfand, A.: Continuous spatial process models for spatial extreme values. *J. Agric. Biol. Environ. Stat.* **15**(1), 49–65 (2010)
- Spiegelhalter, D.J., Best, N.G., Carlin, B.P., Van Der Linde, A.: Bayesian measures of model complexity and fit. *J. R. Stat. Soc. Ser. B (Stat. Methodol.)* **64**(4), 583–639 (2002)
- Varkevisser, M., van der Geest, S.A., Schut, F.T.: Assessing hospital competition when prices don't matter to patients: the use of time-elasticities. *Int. J. Health Care Finance Econ.* **10**(1), 43–60 (2010)
- Victoor, A., Delnoij, D.M., Friele, R.D., Rademakers, J.J.: Determinants of patient choice of healthcare providers: a scoping review. *BMC Health Serv. Res.* **12**(1), 272 (2012)
- Zhang, H.: Inconsistent estimation and asymptotically equal interpolations in model-based geostatistics. *J. Am. Stat. Assoc.* **99**(465), 250–261 (2004)