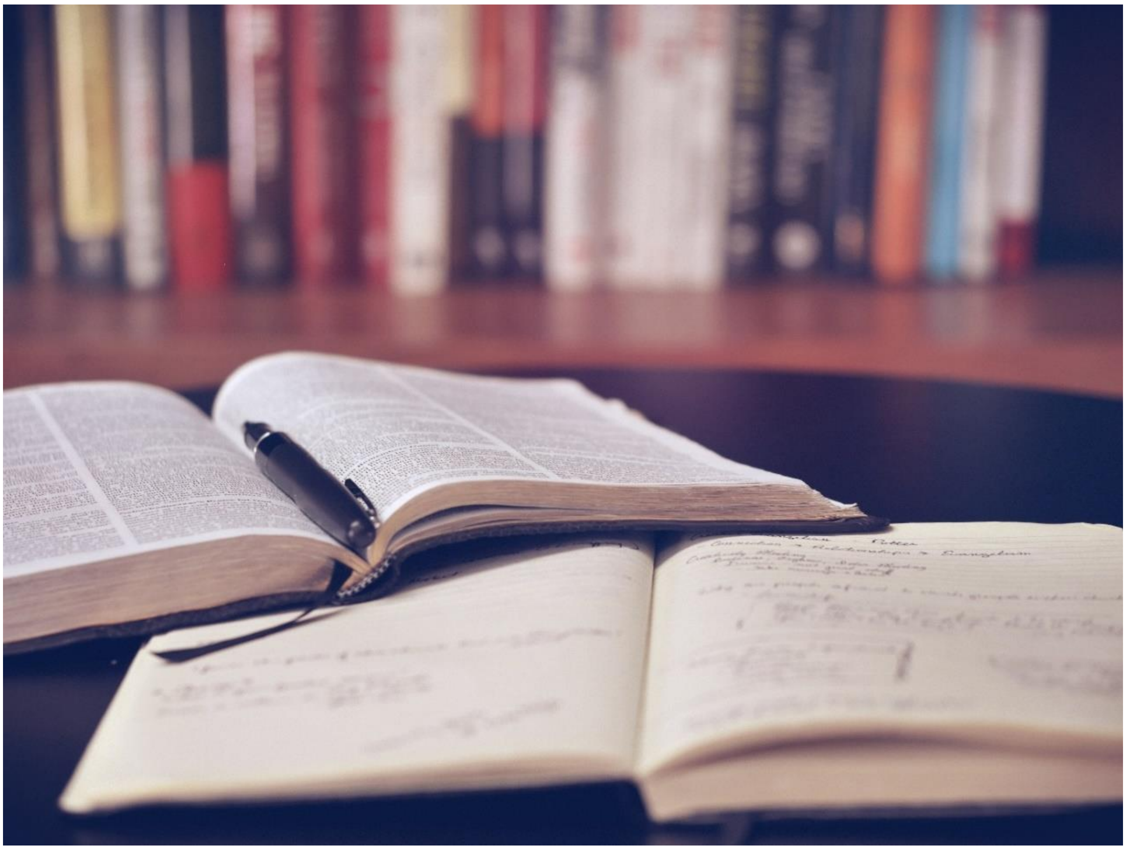
UMR CNRS 6240 LISA

Working Paper TerRA n°14

03-02-2020



Spatio-Temporal Modeling of second homes dynamics in Corsica

Yuheng Ling, Claudio Detotto, Dominique Prunetti

Spatio-Temporal Modeling of second homes dynamics in Corsica

LING Yuheng*[1], Claudio Detotto[1,2], Dominique Prunetti[1]

February 1, 2020

Abstract

Second home development is an important issue to a housing market and local land use. Understanding second home dynamics at the small area scale will help policymakers to conduct effective interventions to meet the concern of local inhabitants. Conventional approaches do not consider second homes varied across space and over time. Hence, spatial models, such as the Besag, York and Mollie (BYM) model, are useful in terms of offering a feasible way to model both covariates and spatial, temporal dependence. This tool is then applied to analyse the second home ratio at a county level in Corsica, France, over the period 2007-2016. We intend to investigate the impact of spatially referenced covariates and the spatio-temporal dependencies of the second home ratio in each Corsican county. Binomial regression models with spatial and temporal random effects are implemented via Integrated Nested Laplace Approximations pertaining to a Bayesian paradigm. Results show that the model with spatial, temporal random components and a spatiotemporal interaction term produces the most realistic estimates and the best forecast. Then, physical landscape counts, coastal counties, and mountainous counties are positively associated with second home rates, while cultural landscape counts, number of households and interest rates are negatively associated with the second home rates. We also identify the "hot spot" and "cold spot" of the second homes. While the temporally structured component reveals that there was merely a gradual increase on the second home ratio over the past 10 years.
Keywords: Bayesian hierarchical modelling, Besag–York–Mollie model, Corsican second home ratio, Integrated Nested Laplace Approximations, spatio-temporal dynamics
JEL classifications: C11, C23, R31

*Email: ling_y@univ-corse.fr.
[1]UMR CNRS 6240 LISA; University of Corsica - Pasquale Paoli; Campus Mariani, BP52, 20250 Corte, France.
[2]CRENoS; Via Muroni 23, 07100 Sassari, Italy.

1 Introduction

Academic discussions of second home phenomena have never ceased over the past decades (Coppock, 1977; Müller, Hall, and Keen, 2004). Topics include regional development, rural planning, tourism promotion and local governance (Farstad and Rye, 2013; Hall, 2015; Müller, Hall, and Keen, 2004).

Corsican context is worthy of studying for several reasons. In 2014, The second home ratio in Corsica equalled 35%, which was 26% higher than French national level. In some villages, this proportion even reached 80% (Maupertuis, Tafani, and Poggioli, 2017). With the growth of tourists, this ratio might go up. Further, local public agencies (CdC, 2019) and economists (Giannoni, Beaumais, and Tafani, 2017; Caudill, Detotto, and Prunetti, 2019; Ling, 2019) have identified the growth of second home properties as a crucial issue to urban planning and rural development. The regional council even voted to build a so-called "statut de résident", which restricted the right to acquire residences in Corsica.

Perhaps surprisingly, even though Corsica is a famous tourist destination in the Mediterranean and most inhabitants think that second homes affect their lives, there appears to be little work on discussing the second homes, especially, the spatial and temporal nature of second home occurrence. Hence, all suggest that a detailed examination of Corsican second home development and change patterns is timely.

Apart from the under-explored Corsican case, past research on second homes is dominated by qualitative approaches or case study designs. Few papers employ basic quantitative approaches, *e.g.*, mapping volumes, identifying spatial clusters and applying classical linear regression (Back and Marjavaara, 2017; Barke, 1991, 2007; Marjavaara and Müller, 2007). However, these methods may lead to biased results and unrealistic inference due to the presence of spatial or temporal dependence in data.

Facing the above-mentioned research gaps, this paper makes two main contributions. Firstly, we attempt to explain the second home dynamics using suitable Binomial regression with the inclusion of latent spatial and temporal effects. More specifically, the main idea of the proposed spatiotemporal models is that after preserving the fixed covariates effects, the residual is decomposed into spatial shared components, temporal shared components and a space-time component. These complex components account for the spatio-temporal dependency on the second home ratio in each Corsican county, since observations tend to be similar if they are geographically close (Tobler, 1970) or temporally adjacent. Additionally, the specifications of the components are flexible, both a parametric and a nonparametric structure for temporal shared components are considered. Moreover, through borrowing strength from neighboring areas and adjacent periods, the proposed models produce reliable estimates. Another key point is that the models can be used within a Binomial response setting. So, they overcome the limitation of a Gaussian response variable in classical spatial econometric models, *e.g.*, the spatial autoregressive model (SAR) and the spatial error model (SEM) (Anselin, 1988). Consequently, we can directly apply the proposed models for count data, such as the number of second homes, rather than implement data transformation to meet the Gaussian response variable.

Incorporating many random effect parameters significantly increases model complexity, and therefore estimation is carried out either by penalized quasi-likelihood (PQL) (Breslow and Clayton, 1993) or Markov Chain Monte Carlo (MCMC)[1] (Gilks, Richardson, and Spiegelhalter, 1995). Nevertheless, MCMC may outperform PQL due to absorbing prior information, but it usually requires a lot of computational resources. Hence, we apply a novel statistical tool called Integrated Nested Laplace Approximations (INLA) (Rue, Martino, and Chopin, 2009) pertaining to a Bayesian paradigm. As an alternative of MCMC, INLA take advantages of numerical integrations and Laplace approximations, which could significantly reduce computational time and return the reliable posterior probability distributions of model parameters. Hence, all models in our work are fitted by the R-INLA package (Martins, Simpson, Lindgren, and Rue, 2013).

Briefly, this research investigates the spatiotemporal patterns of second homes at the county level in Corsica, France. Moreover, we attempt to:
(i) illustrate a novel approach to modeling the second home ratio;
(ii) examine which factor affects second home occurrence in Corsica;
(iii) identify the overall temporal trend in second home occurrence in Corsica from 2007 to 2016.

The remainder of this paper is structured as follows. In Section 2, we provide an overview of second home literature, especially the factors influenced second home occurrence in different regions. In Section 3, we describe the Bayesian hierarchical Binomial model specification, in particular, how to incorporate space, time and space-time random effects in the hierarchical Binomial models. Section 4 details our dataset, outlines empirical strategies and displays the estimation results. We draw a conclusion in Section 5.

2 Theoretical aspects and empirical evidence

Second homes in different regions have been investigated from economic, social and environmental perspectives. We find that there are two key factors affecting second homes, *i.e.*,

[1] PQL belongs to a frequentist approach, while MCMC belongs to a Bayesian approach.

amenities and socio-economic factors.

Many researchers (Coppock, 1977; Hall and Müller, 2004; Müller, 2006; Müller and Hoogendoorn, 2013) hold the view that one of the key determinants for selecting second home locations is high amenity value and extensive research analyses the impact of various amenities in different regions.

Müller (2002) finds that in Northern Sweden, more and more second homes concentrate around coastlines and uplands due to scenic qualities and recreation opportunities. Barnett (2007) conducts a survey of the second home market in Central and Eastern Europe. He concludes that ideal second home locations should satisfy the elements, such as weather, infrastructure, views, history and nature. Norris and Winston (2009) analyze the Ireland case and indicate that second home usually locate in the areas with amenity-rich landscapes or proximity to sea, rivers, lakes and mountains. Kaltenborn, Andersen, Nellemann, Bjerke, and Thrane (2008) show that Norwegian second homes increasingly occur closely around mountain and coast tourism resorts. This phenomenon is related to Norwegian culture, where nature plays an important role in human life and second homes provide a link between human life and nature. Additionally, some second homes locate in areas with historical or social meaning (Kaltenborn, Andersen, and Nellemann, 2007).

Regarding social-economic factors, Norris and Winston (2009) conclude that the growth of second homes is likely related to local factors (Hall, 2015; Orueta, 2012).

Keane and Garvey (2006) find that the increasing growth of second homes in Courttown and Drumshanbo is due to local tax incentive. According to Norris and Shiels (2007), some local governments pass tax incentive schemes to promote second home development because associated construction creates job opportunities and tourism provides income for the local government.

In addition, Barke (2007) analyses the second home changes in Spain from 1981 to 2001. He clearly shows that the number of second homes in a given province is related to the provincial population size. A statistical analysis provides evidence that the second home rate of a province with small population size is usually high, while the province with large population size has a low second home rate. He concludes that depopulation is an important factor to create second homes.

Finally, Dower (1977) holds the view that high cost of borrowing money could slacken the pace of second home growth because of additional capital demand.

Although the aforementioned factors affect second home occurrence in different regions, their impact on Corsican second homes is still unknown. In addition, many scholars are likely to overlook the spatial distribution of second home data, which could result in a biased estimation. The fact is that the relationship between a response variable and covariates seems to vary over space and even time. Further, there seems to be too restrictive to impose a linear relation for some regressors. To address these issues, Bayesian small area estimation methods (Rao, 2003) are integrated into second home analyses. More precisely, we gauge spatial, temporal correlation in second home data by means of Bayesian hierarchical binomial models.

3 Spatiotemporal modeling

This section is divided into two subsections. The first outline spatial data types, in particular, areal data structures. The second describes Bayesian hierarchical Binomial models.

3.1 Spatial data

Following the classification from Cressie (1992), there are three types of spatial data: point-referenced data, spatial point patterns and areal data. The first two types refer to individual-level data, while the last one is based on aggerated-level data. Typically, areal data own two core features. They are built on a given region which is decomposed into many non-overlapped sub-regions. Each sub-region is known as an areal unit. In addition, data are collected through aggregated counts within each areal unit. In our study, we center around areal data, due to the fact that, on the one hand, we are interested in investigating second home rates. On the other hand, the available dataset does not provide exact coordinates of second home properties. Instead, the dataset provides information of second home counts for small areas[2] over years.

3.2 Bayesian Hierarchical Binomial Regression Models

To capture all information from areal data, the proposed models should handle both count data and latent spatial processes. Further, in some cases, it is plausible to consider temporal patterns in the sense that data are collected over years. Temporal correlation may appear and is likely to be stronger for adjacent years than for several years apart.

Facing such issues, a family of Bayesian hierarchical models (BHM) (Banerjee, Carlin, and Gelfand, 2014) is introduced. The advantage of a BHM over a classical linear model is that hidden spatial or temporal processes can be involved into the BHM via conditional distributions, resulting in an additional stage, known as a process model. Therefore, the true state of a phenomenon is modeled. For example, spatial autocorrelation and heterogeneity in the areal data are incorporated in this stage to avoid unrealistic independent and identically

[2] *e.g.,* administrative units.

distributed assumptions. The hierarchical structure is usually specified as follows (Gelfand, 2012):

- Stage 1 – Data model: [data — processes, parameters]
- Stage 2 – Process model: [process — parameters]
- Stage 3 – Parameter model: [parameters]

The first stage describes the distribution of observed data, also known as likelihood. This level conditions with a latent process and the parameters of the data model. The second stage specifies the true latent process, given process parameters. For example, the latent process is delineated by spatial and temporal random effects with precision parameters[3] in our application. At the third level of the hierarchy, we assign hyperpriors to all parameters in the previous stages. Finally, together with the likelihood, process and prior distribution of all parameters, we can estimate the posterior distribution of the model parameters via Bayes theorem.

In particular, to model Binomial-distributed[4], spatial data, a Binomial model is adapted to Bayesian hierarchical modelling framework. The dependent variable Y_{it}, also known as the number of successes, is assumed to follow a Binomial distribution, given n_{it} trials (Ferrari and Comelli, 2016).

$$Y_{it} \sim \text{Binomial}\left(n_{it}, \pi_{it}\right) \tag{1}$$

where the sub-index $i = \{1, \ldots, N\}$ denotes a set of spatial units and $t = \{1, \ldots, T\}$ denotes consecutive periods. π_{it} is the probability of the successes in area i at time t. Further, via a logit link function, the π is linked to a series of structured additive predictors.

$$logit\left(\pi_{ij}\right) = \mathbf{X}_{it}^{\top}\beta + \omega_{it} \tag{2}$$

where $\mathbf{X}_{it}$ are fixed covariate effects and $\beta' = (\beta_1, \ldots, \beta_p)$ is the vector of corresponding parameters. Latent processes are modelled via the random effect component ω_{it}. Consider the case of latent spatiotemporal processes, ω_{it} may include one or more sets of spatiotemporally autocorrelated random effects[5], such as $\omega_i{}' = (\omega_1, \ldots, \omega_N)$, $\omega_t{}' = (\omega_1, \ldots, \omega_T)$ and $\omega_{it}{}' = (\omega_{11}, \ldots, \omega_{NT})$.

[3] A precision parameter is the reciprocal of a variance parameter.
[4] *e.g.,* count data.
[5] Different spatiotemporal random effects represent different spatiotemporal structures.

Note that there are various applications using spatiotemporal Bayesian hierarchical Binomial models, including modelling disease risks (MacNab, 2003; Staubach, Schmid, Knorr-Held, and Ziller, 2002), violent crimes (Zhu, Gorman, and Horel, 2006), public health intervention (Viola, Arno, Maroko, Schechter, Sohler, Rundle, Neckerman, and Maantay, 2013), presidential elections (Linzer, 2013) and public confidence on police (Williams, Haworth, Blangiardo, and Cheng, 2019).

3.3 Modelling spatial dependence

Regarding the spatial random effect, ω_i', we specify its prior distribution by an intrinsic conditional autoregressive model (ICAR) (Besag, York, and Mollié, 1991), which is a special case of conditional autoregressive model (CAR).

Given a set of areal units at sites $i \in \{1, 2, \ldots, n\}$, the spatial interaction between a pair of unit i and unit j can be modelled through a conditional univariate Gaussian distribution. More specifically, this is done by assuming that the conditional distribution depends only on the sites that are neighbors of site i. Formally, the full conditional distribution is:

$$\upsilon_i | \upsilon_j \sim N\left(\frac{\sum_{j\in\partial_i} w_{ij}\upsilon_j}{w_{i+}}, \frac{\sigma^2}{w_{i+}}\right) \tag{3}$$ [6]

where W is an adjacency matrix describing a neighborhood graph. Its diagonal elements are all zero, $w_{ii} = 0$ and the off-diagonal entities, w_{ij}, show the proximity of regions n_i and n_j. w_{i+} is the row sum of the ith row, $w_{i+} = \Sigma_i w_{ij}$ and ∂_i denotes the index of neighbors of i. The core features of an ICAR model include (i) the mean of the conditional distribution equals the average of its neighbouring values; (ii) the variance is given by the global variance divided by the number of its neighbors. Therefore, as the number of the neighbours (N_i) increases, the variance of the conditional distribution ($\frac{\sigma^2}{N_i}$) will decrease. In small area estimation literature (Rao, 2003), this setting implies that an area "borrows strength" from its neighboring areas (see A.1 for more details).

In the last decade, many spatial models for areal data have been proposed. Most of them are extensions of CAR models, with examples including the popular Besag-York-Mollié (BYM) model. BYM model can be considered as an ICAR component (υ) coupled with an

[6] Under first-order contiguity, the full conditional distribution can be simplified as:

$$\upsilon_i | \upsilon_{j, j\in\partial_i} \sim N\left(\frac{1}{N_i}\sum_{j\in\partial_i} \upsilon_j, \frac{\sigma^2}{N_i}\right)$$

where j is a neighbor of i, N_i is the sum of neighbors of the unit i, $N_i = w_{i+}$.

exchangeable random effect (IID) component[7] (ν). An advantage of BYM over ICAR is that BYM could address the issue of heterogeneity (Harris, 2019).

3.4 Modelling temporal dependence and spatio-temporal variability

As was mentioned in the previous section, it is necessary to take temporal correlation into account if data are collected over years, since an ICAR or BYM model captures spatial information only. Hence, on the basis of the ordinary BYM, several spatio-temporal specifications are proposed, involving spatial, temporal components and even space-time interaction terms. For temporal components, parametric and non-parametric smoothing functions of time are considered. Further, a space-time interaction term gauges the specific temporal trend for each unit (López-Quilez and Munoz, 2009).

The two prominent specifications for the temporal and spatiotemporal components are the linear trend model proposed by Bernardinelli, Clayton, Pascutto, Montomoli, Ghislandi, and Songini (1995) and the dynamic trend model proposed by Knorr-Held and Besag (1998).

Regarding the linear trend model, the linear parametric formulation of time is written as:

$$(\xi + \varphi_i)\, t^{8} \tag{4}$$

where $\xi \times t$ is a global linear function of time representing a mean trend in time and $\varphi_i \times t$ an area-specific linear function gauging an area-specific trend to depart from the mean trend. We specify a Gaussian exchangeable prior to φ_i.

$$\varphi_i \sim \text{ Normal } \left(0, \sigma_\varphi^2\right) \tag{5}$$

Seen differently, this model is a variant of a random slope model with a spatially varying slope ($\xi + \varphi_i$). In a word, this model is simple, straightforward, but the linear assumption seems to be restrictive.

To relax the linear restriction, it is possible to model time by a dynamic nonparametric formulation (Knorr-Held and Besag, 1998):

$$\zeta_t + \gamma_t + \delta_{it} \tag{6}$$

[7] In BYM literature, the ICAR component is also known as a spatially structured random effect and the IID component is known as a spatially unstructured random effect.

[8] Note that t is normalized.

where there is a structured random effect ζ_t, an unstructured random effect γ_t and a space-time interaction term δ_{it}.

ζ_t is usually modelled by a first order of random walk:

$$\zeta_t|\zeta_{t-1} \sim \text{ Normal } \left(\zeta_{t-1}, \sigma^2_\zeta\right) \tag{7}$$

where ζ_t gauges temporal dependence in the sense that the parameter for each time period depends on the previous one.

γ_t denotes the temporally unstructured random effects. A Gaussian exchangeable prior with 0 and σ^2_γ is assigned to this parameter.

$$\gamma_t \sim \text{ Normal } \left(0, \sigma^2_\gamma\right) \tag{8}$$

For the space-time interaction component δ_{it}, a Type I unstructured interaction is used (Knorr-Held, 2000). The Type I interaction is the cross product of spatially and temporally unstructured terms[9], which can be considered as a random intercept based on all observations. Seen differently, this component represents global space-time heterogeneity and measures the deviation from the spatial and temporal main effects. Again, a Gaussian exchangeable prior with 0 and σ^2_δ is specified.

$$\delta_{it} \sim \text{ Normal } \left(0, \sigma^2_\delta\right) \tag{9}$$

The aforementioned spatial and spatiotemporal models pertain to a class of structured additive regression (STAR) models (Brezger, Kneib, and Lang, 2005; Klein, Kneib, Lang, and Sohn, 2015). On the one hand, a STAR model has a link function and structured additive predictors. The latter uses different smooth functions to gauge various types of effects, including nonlinear effects, spatially and temporally correlated effects. On the other hand, STAR formulates within a hierarchical Bayesian framework, so that different random effects describe different processes[10].

A common way to fit a STAR model is applying MCMC simulation. However, MCMC may experience long computational time. In particular, the spatial and spatiotemporal models usually require more computational resources[11], but these models can formulate into latent Gaussian Markov Random Field (GMRF) models (Rue and Held, 2005; Schrödle and Held, 2011). Therefore, we can apply INLA, which are designed for the latent GMRF model and return fast and accuracy Bayesian inference.

[9] ν_i and γ_t interact.

[10] *e.g.,* random walk, first order autoregressive, intrinsic conditional autoregressive.

[11] *e.g.,* time and memory.

4 Empirical analysis

4.1 Study region

As explained in the Introduction, we attempt to investigate the dynamics and impact factors of second home rates in Corsica, France. Corsica is one of the 18 French administrative regions located in the Mediterranean Sea. Additionally, Corsica is divided into 360 counties. As "the pearl of Mediterranean", Corsica is famous for its rich tourism resources (Vogiatzakis, Pungetti, and Mannion, 2008). A single mountain range crosses the center of the island with mountainous and alpine landscapes. While, beautiful beaches and seaward cliff predominate coastal areas. For instance, Calvi and Porto Vecchio, located in the northwest and southeast, are famous for their sandy beaches; Corte with many historical sites is suited in the center of Corsica; Ajaccio and Bastia are the main cities in the island.

4.2 Data

In the study, data are collected over 360 counties from 2007 to 2016, representing 10 periods. As such, this setting results in a total of 3600 space-time units without any missing values.

Data for each year concerning the number of the second home and the number of the house in each county are obtained from INSEE[12] through the French population census from 2007 to 2016. It also provides population information in each county. Interest rate data are acquired from the "Banque de France"[13] via the item "Narrowly defined interest rate median for real-estate loans for individuals". Since the time measure of data is the quarter, we then convert the quarterly interest rate to the annual interest rate. Council tax data are obtained from the "Ministère de l'action et des comptes publics". Unemployment rates at the "zone d'emploi" level are also collected from INSEE. Lastly, the research center "UMR CNRS 6240 LISA" provides data for the rest of variables[14].

Descriptive statistics for the second homes and houses till 2016 are displayed in Table 1. In the final year, there are 91,622 second homes distributed over 360 counties, meaning that the second homes occupy 37.2% of the total houses at the end of the study period.

Fig 1 gives the annual second home rates[15] among Corsican 360 counties.

Temporally, a gradual but steady augmentation of second home proportion is observed. More precisely, regarding the median of second home proportion, it keeps stable during the first two years. Then, the proportion slightly increases in the subsequent five years.

[12] INSEE stands for "The National Institute of Statistics and Economic Studies".

[13] French central bank.

[14] Table 7 in the appendix A.2 illustrates the variables and corresponding data sources.

[15] The raw counts of the second homes can be found in Table 8.

Table 1: Descriptive statistics for second homes and total houses by county (period: 2007 - 2016)

	Study region		Dissemination	Area	
	Current Total Count	Mean	Min	Max	SD*
Second home counts	91,622	222.88	1	6,748	461.52
Total house counts	245,851	619.34	15	33,895	2,132.65
Ratio	0.372	0.51	0.022	0.83	0.16

* SD stands for Standard Deviation.

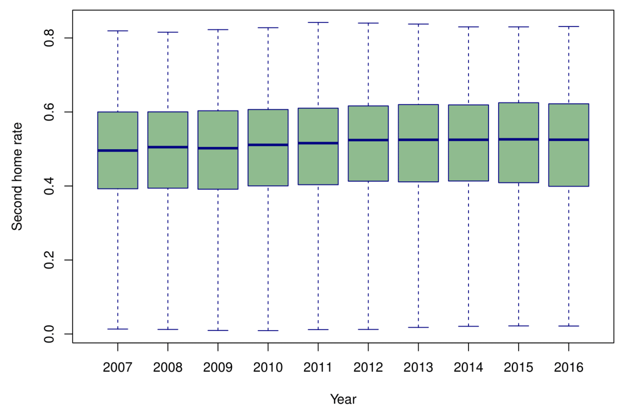


Figure 1: Boxplot of the temporal trend in the raw rate of second homes as a proportion of the total number of houses.

The proportion of the latter three years keeps stable again. The interquartile range of the proportion shows the same step-change to the median proportion and always falls into [0.4, 0.6].

Fig 2 displays the distribution of the ratio of the second home counts over the total house counts in Corsican 360 counties during 2007-2016.

There seems to be an annual change of the spatial variation in some areas. For instance, several counties[16] suited in the Corsica northern tip have a low second home rate in 2007, whereas the rate upturns in 2008 and keeps stable in the subsequent years. The second home rate in Bonifacio[17] is relatively lower until 2010 and then it goes up. This phenomenon suggests that once the second home rate in an area goes up, it may never drop off.

We also observe that counties with high second home rate cluster within the north tip area, Saint-Florent area, Balagne area. On the contrary, low rate zones involve the northern

[16] *e.g.,* Rogliano, Luri county.
[17] A county locates in the south corner of the island.

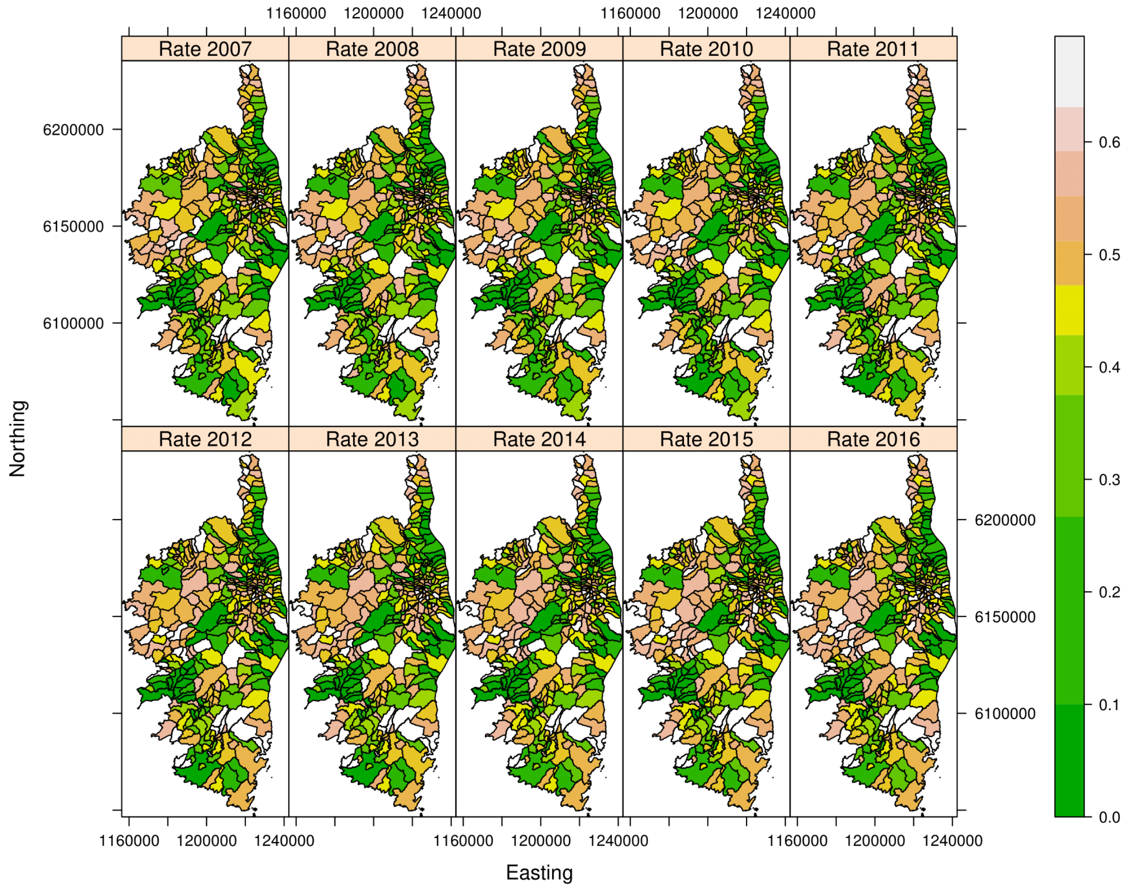


Figure 2: Geographic distribution of second home ratio from 2007 to 2016.

east coastal area[18], the Corsica center and the Ajaccio area. These high and low rate zones likely persist during the whole study period. In contrast, other high and low rate zones disperse throughout the whole island.

Instead of ocular judgements, we employ Moran's I tests[19] (Moran, 1950) to further investigate the spatial dependence structure in the data. The results in Table 2 provide evidence that there is positive spatial dependence in the second home ratios with an average of 0.334 (P-value < 0.05) over the 10 years. Hence, when we model the Corsican second home ratio, spatial dependence should not be neglected.

Following the review of the second home literature and considering the data availability, two types of explanatory variables are considered, such as amenity variables and socio-economic variables. Descriptive statistics for these variables are shown in Table 3.

[18] Bastia and its neighbour counties.

[19] We calculate Moran's I statistics separately for each year. In this case, the adjacent matrix is defined by queen contiguity weights shown in Fig. 6.

Table 2: Annual Moran's I statistics on second home rates

Year	Moran's I	P-value
2007	0.304	<0.05
2008	0.325	<0.05
2009	0.337	<0.05
2010	0.334	<0.05
2011	0.335	<0.05
2012	0.322	<0.05
2013	0.324	<0.05
2014	0.343	<0.05
2015	0.359	<0.05
2016	0.362	<0.05

Table 3: Descriptive statistics of independent variables

Statistic	Mean	SD	Min	Pctl(25)	Pctl(75)	Max
physical landscapes	0.719	1.830	0	0	1	19
cultural landscapes	0.747	2.419	0	0	1	28
$log_2(household)^*$	6.567	1.972	1.585	5.129	7.728	14.822
household growth	0.005	0.037	-0.24	-0.012	0.025	0.22
log_2(interest rate)*	-4.778	0.307	-5.479	-4.968	-4.556	-4.353
log_2(council tax)*	-2.801	0.650	-5.703	-3.249	-2.307	-0.712
log_2(unemployment rate)*	-3.412	0.240	-4.053	-3.556	-3.231	-2.905

* To facilitate interpretation, a base 2 logarithm transformation is applied to the variables.

Table 4: Candidate Models for the second stage

Model	Type	Equation
0	Classical Binomial without any latent structures	$\text{logit}(\pi_{it}) = x_{it}\beta + \sum_{j=1}^{p} f\left(x_{jt}\right)$ (11)
1	Purely spatial	$\text{logit}(\pi_{it}) = x_{it}\beta + \sum_{j=1}^{p} f\left(x_{jt}\right) + \upsilon_i + \nu_i$(12)
2	Spatial and temporal jointly	$\text{logit}(\pi_{it}) = x_{it}\beta + \sum_{j=1}^{p} f\left(x_{jt}\right) + \upsilon_i + \nu_i + \xi t$(13)
3	Spatial and temporal jointly	$\text{logit}(\pi_{it}) = xit\beta + \sum_{j=1}^{p} f\left(x_{jt}\right) + \upsilon_i + \nu_i + \zeta_t + \gamma_t$(14)
4	Spatiotemporal with interactions	$\text{logit}(\pi_{it}) = x_{it}\beta + \sum_{j=1}^{p} f\left(x_{jt}\right) + \upsilon_i + \nu_i + \left(\xi + \varphi_i\right) t$(15)
5	Spatiotemporal with interactions	$\text{logit}(\pi_{it}) = x_{it}\beta + \sum_{j=1}^{p} f\left(x_{jt}\right) + \upsilon_i + \nu_i + \zeta_t + \gamma_t + \delta_{it}$(16)

4.3 Candidate models

To model the second home data, the BYM model and its spatiotemporal extensions are considered. In this study, we specify five latent processes with increasingly complex representation of space and time. As discussed above, Bayesian spatial and spatiotemporal models can be written as a three-level hierarchical model.

At the first level, we assume that the observed counts for second homes (y_{ij}) are Binomial-distributed[20], where the rate of second homes (π_{it}) is a proportion of the total number of houses (n_{it}). $i = \{1, \ldots, 360\}$ denotes the county and $t = \{1, \ldots, 10\}$ is the temporal unit, year.

At the first stage, a binomial likelihood for second home counts is specified.

$$Y_{it} \sim \text{Binomial}\left(n_{it}, \pi_{it}\right) \tag{10}$$

At the second stage, a logit transformation is applied to π_{ij}, representing the estimated rate of second homes, and then the transformed π_{it} equals the sum of all linear, non-linear predictors and different random effect components.

The candidate specifications for the second stage include the models described in table 4. In equations (13) to (16), υ_i denotes the spatially structured random effects capturing spatial dependence, ν_i represents the spatially unstructured random effects capturing spatial heterogeneity, as explained in Subsection 3.3.

Thus far, we have known the candidate specifications for the latently spatial, temporal process. At the third stage, we assign hyperprior distributions to all parameters appeared

[20] In our case, the annual housing survey is considered a Bernouilli trial where a second home is treated a success. The fact is that housing transactions are not frequent in Corsica. In particular, a second home could hardly sell more than once in a year, meaning that second home sales probably constitute a Bernouilli trials. Hence, we prefer a Binomial model to a Poisson model. Moreover, the second home ratio is relatively high, which also provides evidence that a Poisson model may be not adequate. Typically, the fit of the Binomial model could be close to the fit of the Poisson model with a small number of successes but a large number of trials.

in the previous stages. In this study, we assign a $Gamma(1, 0.00005)$ distribution to to all precision parameters[21]. The list of all precision parameters is shown as follows.

$$\frac{1}{\sigma^2_{x_{jt}}}, \frac{1}{\sigma^2_{\upsilon}}, \frac{1}{\sigma^2_{\nu}}, \frac{1}{\sigma^2_{\zeta}}, \frac{1}{\sigma^2_{\gamma}}, \frac{1}{\sigma^2_{\delta}} \tag{17}$$

4.4 Modeling strategy

In the last section, we introduce the candidate models, starting from the classical Binomial model and ending with the Binomial model with spatial, temporal and type I spatiotemporal interaction components. Model 0, referred to as a benchmark, is a classical Binomial model without any spatial effects. However, we detect the spatial autocorrelation in the data, it is natural to consider a model in the CAR family and then to involve temporal components for capturing possible temporal trends.

Different spatial or temporal components will result in different smoothing in the posterior estimates. Regarding Model 1, estimated values are likely to vary across neighbors but not over the test period. For Model 2, a single value is estimated for each neighbor and time period, but the parametric formulation seems restrictive. The linear assumption is relaxed in Model 3 and the model can gauge temporal autocorrelation in the data. Comparing with Model 2, Model 4 incorporates a spatiotemporal interaction term, which allows for a specific temporal trend for each observation while retaining the main, linear temporal trend. As such, Model 4 should perform better than Model 2. Concerning Model 5, the inclusion of the spatiotemporal interaction component should additionally improve the goodness of fit. The empirical data will be more accurately represented.

4.5 Model implementation and assessment

All models are carried out by using R 3.5.3 (R Core Team, 2013) and R-INLA 19.04.16.

The Deviance Information Criterion (DIC), the mean logarithmic conditional predictive ordinate (LCPO) and the holdout method in cross-validation are used for assessing model fit and prediction capability. DIC is a widely-used criterion to evaluate goodness of fit in Bayesian hierarchical models (Spiegelhalter, Best, Carlin, and Van Der Linde, 2002). DIC is based on the posterior distribution of the deviance statistic. Model fit is computed by the posterior expectation of the deviance $\overline{D} = E_{\theta|y_{it}}(D)$[22], while model complexity is summarized by the effective number of parameters $p_D = \overline{D} - D(\overline{\theta})$. Combining the two parts and we then obtain DIC:

[21] In the INLA approach, we have $\log(precision) \sim \log - \text{Gamma}(1, 0.00005)$.

[22] y represents observed data and θ is a parameter vector.

$$\text{DIC} = \overline{D} + p_D = 2\overline{D} - D(\bar{\theta}) \tag{18}$$

For DIC, smaller values correspond to better-fitting models.

Alternatively, CPO (Geisser, 1993) is used as a predictive measure of a model. CPO pertains to leave-one-out cross-validation and is defined as

$$\text{CPO}_{it} = \pi\left(y_{it,obs}|y_{-(it),obs}\right) \tag{19}$$

where $\pi\left(y_{it,obs}|y_{-(it),obs}\right)$ is the cross-validated predictive density at the omitted observation it given all the other data. Following the suggestion from Roos and Held (2011), we calculate the mean logarithmic CPO (LCPO)[23] as follows:

$$\text{LCPO} = -\frac{1}{N \times T}\sum_{t=1}^{T}\sum_{i=1}^{N}\log\left(\text{CPO}_{it}\right) \tag{20}$$

Here, lower LCPO scores also indicate better models.

Predictive performance is also tested by holding out the data for the most recent most year[24]. We then computed predicted values for the holdout units through the models trained by a training dataset. The root mean square error (RMSE) is considered to measure the closeness between the predicted second home ratio $\hat{\pi_{it}}$ and the observed second home ratio π_{it}, and defined by

$$\text{RMSE} = \sqrt{\frac{1}{N \times T}\sum_{t=1}^{T}\sum_{i=1}^{N}\left(y\left(s_i, t\right) - y\,\widehat{\left(s_i, t\right)}\right)^2} \tag{21}$$

Again, lower RMSE values represent better predictive power.

4.6 Results

DIC, LCPO and RMSE values are displayed in Table 5.

DIC scores indicate that the classical Binomial model is poorly fitted. Comparing Model 0 with Model 1, the fit is much improved ($\Delta_{DIC} = -174133.40$). The marked decrease of DIC also provides evidence that the ill-fitting Model 0 is a result of omitting unobserved spatial patterns. In addition, adding a linear, parametric temporal component (Model 2) improves the model fit ($\Delta_{DIC} = -12.58$) slightly. The fit is further improved ($\Delta_{DIC} = -20.77$) in

[23] LCPO is also known as the cross-validated logarithmic score.

[24] We chose to test the model on the data of the most recent year rather than a random holdout dataset in order to simulate practical application of the model, predicting future distribution of the second home ratio in a year for which the model is naive.

Table 5: Model assessment via DIC, LCPO and RMSE

	DIC	LCPO	RMSE
Model 0	199,824.40	97,781.77	0.10852
Model 1	25,691.02	13,255.66	0.04115
Model 2	25,678.44	13,251.93	0.04113
Model 3	25,657.67	13,234.20	0.04093
Model 4	25,345.78	13,068.86	0.03670
Model 5	25,257.19	12,978.69	0.03439

Model 3 with the relaxation of the linear restriction in Model 2, using instead a dynamic nonparametric temporal component. Comparing Model 2 with Model 4, the DIC scores are largely reduced ($\Delta_{DIC} = -332.66$) because of the inclusion of a space-time interaction term. After incorporating the Type-I space-time interaction term, the DIC scores are further reduced in Model 5. Hence, Model 5 has the best fit overall.

Regarding model predictive performance, the LCPO and RMSE values show in the same sequence as the DIC score, meaning that LCPO and the holdout method favour Model 5 as well.

For these reasons, Model 5 is the best model to perform both estimation and prediction among all candidate models. Subsequently, this model will be used to making an inference.

4.7 Discussion

One aim of this study is to estimate the fixed effect coefficients. In addition to the fixed effects, the spatial, temporal and spatiotemporal random effects are also investigated to provide possible explanations for the space-time pattern of the second home ratio.

For each covariate, the upper part of Table 6 presents the log odds of the probability being second homes versus the probability being other types of houses associated with a 1-unit or percentage increase and its associated 95% credible interval (CI).

In general, most covariates are significant except for the household growth, council tax and unemployment rate.

It can be seen that the relative log odds of the probability being second homes increases 0.094 ($95\%CI, 0.042; 0.147$) times with a 1-unit increase in physical landscape counts, given all else is equal. In contrast, a unit increase in the cultural landscape count decreases 0.051 ($95\%CI, -0.090; -0.0012$) times in the log odds of the probability being second homes. Hence, both landscape variables are informative. We may infer that the physical landscape count significantly increases the log odds, while the cultural landscape count negatively con-

Table 6: Posterior estimates of the covariates in Model 5

	Mean	St.Dev	0.025 quant	0.975 quant
Intercept	-1.786*	0.263	-2.314	-1.280
physical landscapes	0.094*	0.027	0.042	0.147
cultural landscapes	-0.051*	0.020	-0.090	-0.012
coastal county	0.759*	0.109	0.545	0.974
mountainous county	0.172*	0.079	0.016	0.327
$\sigma^2_{log_2(household)}$	0.267*	0.046	0.188	0.367
household growth	-0.140	0.101	-0.338	0.059
log_2(interest rate)	-0.124*	0.025	-0.175	-0.077
log_2(council tax)	-0.019	0.044	-0.106	0.067
log_2(unemployment rate)	-0.050	0.044	-0.136	0.036
σ^2_{υ}	0.134623	0.055502	0.053763	0.268292
σ^2_{ν}	0.261115	0.035488	0.200101	0.339167
σ^2_{ζ}	0.00031	0.00024	0.00004	0.00091
σ^2_{γ}	0.00019	0.00033	0.00001	0.00092
σ^2_{δ}	0.00738	0.00055	0.00635	0.00851

* Indicates the significance of independent variables.

tributes to the log odds. It indicates that second home buyers likely prefer natural scenery to artificial elements. A possible explanation for this finding is that since cultural landscapes usually locate in a town with good accessibility, it may bring overcrowding issues (Goodwin, 2017; Milano, Cheer, and Novelli, 2018). Second home buyers often look for an area with beautiful scenery, silence and low population density, so they probably consider the overcrowding as a dis-amenity.

The posterior mean coefficient 0.172 ($95\% CI, 0.016; 0.327$) indicates that the odds of the probability being second homes versus other house types increases by $exp(0.172) \approx 1.188$ times as a mountainous county changes to a flat county, given all else equal. In addition, there would be a 2.136[25] times increase in the odds of the probability being second homes as a coastal county changes to an inland county. One possible implication of this is that Corsican second home buyers prefer living near coasts to living in mountainous areas.

After analysing plots of each covariate against the model residuals[26], a non-linear trend for the logged households[27] is detected. Hence, the default first-order random walk (RW1)

[25] The coefficient of coastal counties has a posterior mean of 0.759 ($95\% CI, 0.545; 0.974$)

[26] LOESS smoothing is applied to fit the points in each plot.

[27] The household variable may experience an endogenity issue due to reverse causality. Since it lacks appropriate instruments in the context and the instrumental variable method within the Bayesian framework is still under investigated, we test strict exogenity of the household variable via the Wooldridge's approach

smoother in R-INLA is applied to the logged households. Figure 3 displays the non-linear relation between the base 2 logarithm of households and the probability being second homes, showing on the log odds scale. We initially observe a downward trend. The log odds of the probability being second homes are at their highest in areas with the lowest value of logged households and decrease as the value of logged households increases. The decrease is non-linear, with a descent reaches around -0.5, followed by a leap of the slope. The leap occurs when the logged households reaches the interval $(11, 13)$. A possible explanation for the leap is that these mid-size counties[28] locate close to the main cities or are the capital of cantons, and some of them are not far from national parks. In a words, these counties are easy to reach and people can find different services and also public facilities. Then, the slope decreases again, but the width of the 95% CI is relatively large because of limit observations.

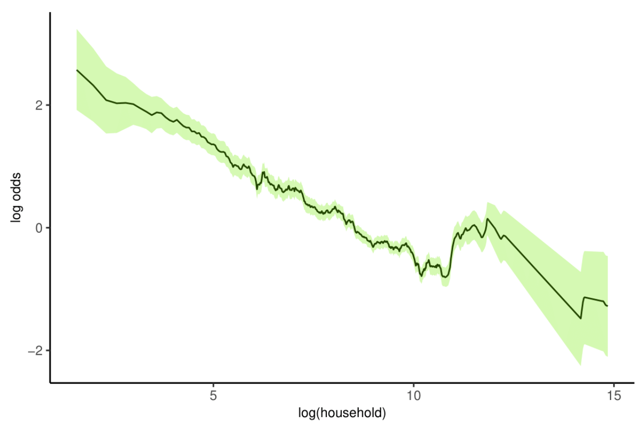


Figure 3: Log odds relations between $log_2(household)$ and the probability being second homes. The black line indicates the posterior mean log odds, while green represents the corresponding 95% CI.

The coefficient on the base 2 logarithm of interest rates has a posterior mean of -0.124 $(95\%CI, -0.175; -0.077)$, meaning that a doubling of interest rates translates to a 11.66% decrement for the odds, given all else equal. The coefficient reported here appears to support the assumption that a low interest rate will encourage house buyers to enter the market. There is a likely explanation for this result, house buyers are more likely to take out a home loan and low interest rate means that obtaining home finance is more affordable (Paris, 2009).

The following part moves on to describe the posterior estimates of spatial and temporal

(see Wooldridge, 2010, pg.285; pg.490). The lead household variable is included in the model additionally. We initially run the Model 5 including the logged households as a linear predictor. Then, the lead-1 or lead-3 households is included in the Model 5 additionally. The posterior estimates for the two additional variables are 3.369×10^{-5} $(95\%CI, -1.587 \times 10^{-6}; 6.936 \times 10^{-5})$ and 5.114×10^{-8} $(95\%CI, -3.517 \times 10^{-5}; 3.547 \times 10^{-5})$. Such result shows that there is not any endogenity issue (See Table 10 for more details).

[28] The mid-size counties refer to counties with around 2,000 to 8,000 households.

random effects.

Residual components provide a description of baseline log odds of being second homes in each county. More specifically, Fig 4 reveals the contribution of the spatially structured and unstructured residuals to log odds in each county after controlling all fixed covariate effects and temporal variation.

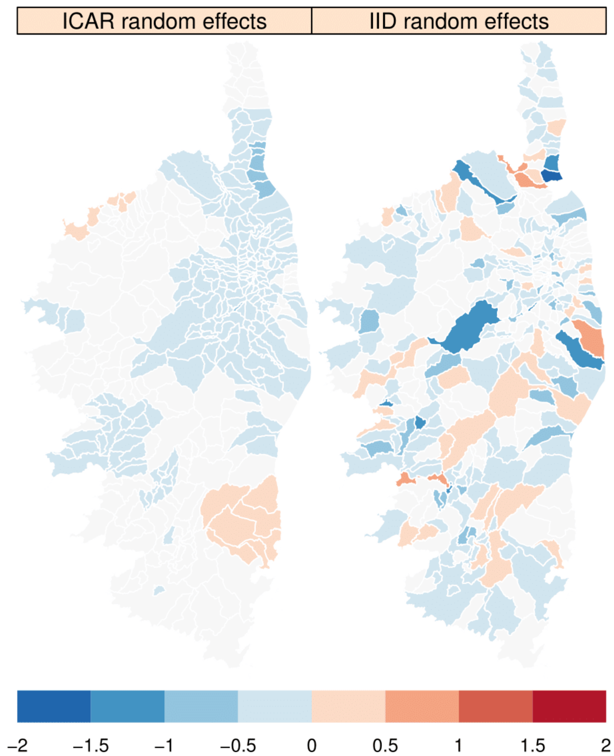


Figure 4: Posterior mean estimates of spatial random effects for Model 5. Left panel: spatially structured random effects. Right panel: spatially structured random effects.

To distinguish the positive and negative contribution, we define "hot spots" and "cold spots". The "hot spot" means the area for which there is evidence that its location positively contributes to the log odds most, given all covariates. More precisely, regarding the spatially structured residual, five main "hot spots" are Lecci, Algajola, Lumio, Zonza and Conca county. These counties pertain to the two spatial clusters, Balagne area and Southeastern area. Policymakers should be aware of these danger clusters, which show higher than average level of second homes. Introducing tightening measures by local governments to prevent the growth of second homes in these counties is urgent. For example, the local government may apply a so-called "taxe d'habitation sur les logements vacants" according to French law. In contrast, Bastia, Furiani, Biguglia, Ville-Di-Pietrabugno and San-Martino-Di-Lota are marked as "cold spots". All these counties are merely located in the Bastia area. That is to say, the Bastia housing market is less affected by the presence of second homes comparing with the Balagne and Southeastern housing markets. Concerning the spatially unstructured residual, the "hot spots" include Linguizzetta, Grosseto-Prugna, Saint-Florent, Oletta and Aullene, while Furiani, Biguglia, San-Gavino-Di-Tenda, Ambiegna and Alzi are "cold spot" counties. Spatially, the "hot spot" and "cold spot" counties derived from spatially unstructured residual are quite dispersed. For the counties within the "hot spot" derived from the spatially unstructured residual, we suggest that the local government introduce tightening measures as well.

Fig 5 shows the empirical temporal trend for the log odds of the probability being second homes. Notably, the scale of the vertical axis of the red line range from -0.025 to 0.025 approximately, which suggests a slightly positive temporal trend over the past 10 years. Hence, we believe that nowadays, the high second home rate is not due to the second home increment during 2007-2016. Moreover, we do not observe sharp fluctuations in the graph, which also provides evidence that there is not any booming in Corsica during 2007-2016.

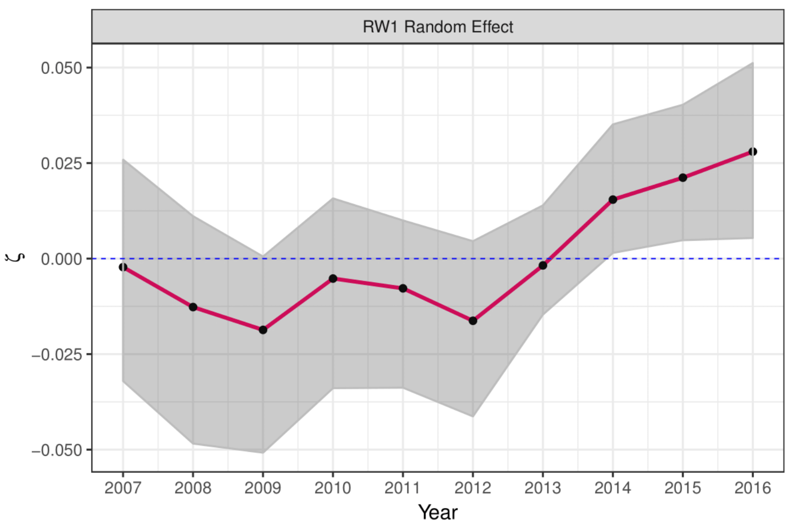


Figure 5: Posterior structured temporal trend for second home ratio in Corsica. The red line indicates the posterior mean trend and grey represents the corresponding 95% CI.

Investigating the type-I space-time interaction term provides further insight into the spatiotemporal pattern of the second home ratio, as shown in Fig 8 in the A.3. Instead of interpreting the spatiotemporal pattern directly, we calculate the proportion of marginal variance explained by each component, given by

$$p_i = \sigma_i^{-1} / \left(\sigma_\upsilon^2 + \sigma_\nu^2 + \sigma_\zeta^2 + \sigma_\gamma^2 + \sigma_\delta^2\right) \times 100\%, i = \{\upsilon, \nu, \zeta, \gamma, \delta\}$$

The corresponding proportions are 33.35%, 64.68%, 0.081%, 0.06%, 1.83%. These results suggest that a huge part of the variability is explained by the spatial structures. More precisely, the spatially unstructured component contributes to the variability most and the spatially structured component explains a sizable proportion of the remaining second home ratio variance. Differently, the temporal components explain very little of the variability. Finally, the space-time interaction term explains a small proportion of the remaining variance, but the proportion is relatively larger than the proportion of any temporal component.

5 Robustness Check

Matters of concern may arise from two parts, the sensitivity to the priors and the necessity of including covariates.

Regarding the prior sensitivity, different priors are tested to assess the change in the posterior distribution of all covariates and variance parameters. The tested priors (Simpson, Rue, Riebler, Martins, and Sørbye, 2017) are shown in Table 9 in the A.2. Note that in this study, the change in the posterior distribution of the covariates is a main criterion to decide whether Model 5 is sensitive to priors.

For the fixed effects shown in Table 11 in the A.2, the posterior distribution of all covariates obtained from the tested priors is almost the same as the posterior distribution of covariates using the default prior. These results suggest that Model 5 should be not sensitive to priors.

To evaluate the need to include all covariates, we rerun the Model 5 without any covariates, named as a convolution model[29]. From Model 5 to the convolution model, the decrease of DIC can be clearly seen in Table 12 in the A.2. Rao (2003) holds the view that incorporating covariates to small areas estimation models can increase the model predictive power and our finding provides evidence for this point of view.

6 Concluding Remarks

In this study, we propose using a Bayesian spatiotemporal approach to identifying the spatial and temporal variation in the second home ratio in Corsica, France, from 2007 to 2016. We also investigate the impact of fixed effects, including amenity factors and socioeconomic factors on the second home ratio. These fixed effects provide additional insight into the dynamics of Corsican second home ratio. To our knowledge, the approach in this study has previously not been used for analysing second homes and offer a useful tool for practitioners to investigate fixed effects and to identify spatial clusters and temporal trends.

[29] The convolution model consists solely of an intercept term with all random effects components.

Regarding the statistical modelling, the models are initially motivated by spatiotemporal data indexed at medium geographical and temporal resolution. Further, we intend to point out the importance of space and time in second home analyses. To gauge latent spatial and temporal information, we introduce 6 candidate models with a range of spatial and temporal representation structures, which still formulate into Bayesian hierarchical models. Note that all models are implemented by the R-INLA package in the statistical programming language R. As a result, the models are widely applicable and highly reproducible. INLA do provide a feasible way to hand large, spatiotemporal datasets and return a considerably accurate result for practical applications. Alternatively, MCMC methods can be used to fit the models, but many scholars (Huang, Malone, Minasny, McBratney, and Triantafilis, 2017; Modrák, 2018) hold the view that INLA require less computationally resource than the MCMC methods. Furthermore, INLA currently become the default spatial analytic tool. Lastly, based on the information criteria, the best fitting model owns spatial, temporal and spatiotemporal random components. Hence, the stable spatial variation and gradual temporal trend in the second home ratio are uncovered.

Some key findings are related to fixed effects. Most amenity covariates are associated with an increase in the second home ratio. To be more specific, a 1-unit increase in the physical landscape is found to significantly increase the odds of the probability being second homes, by approximately 1.098 times. The coastal and mountainous county is also found to significantly increase the odds comparing with the inland and flat county, by around 2.136 and 1.188 times respectively. While the cultural landscape significantly decreases the odds. In terms of socio-economic covariates, an increment in the log-transformed household is associated with a considerable reduction in the odds. Furthermore, an increment in the log-transformed interest rate is associated with a moderate decline in the odds.

Regarding the random effect, the "hot spot" and "cold spot" areas of second home ratio in Corsica are well identified[30]. In addition, the gradual temporal tend for the second home ratio is also recognized throughout the 10-year study period. The spatiotemporal dynamics of the second home ratio are finally described by the space-time interaction random effect. We gain some further insight into the spatiotemporal dynamics, rather than lose the information in the noise term. Therefore, the proposed approach can be viewed as a decent tool to analyse second home dynamics and the local government may perform additional interventions and activities to the identified areas in the findings.

In conclusion, this work contributes to the existing research in second homes in two broad ways. The study assesses the second home in Corsica, an island in the Mediterranean, at

[30] *e.g.,* the Balagne region, the South-eastern region, the Ajaccio city and its neighbour counties, as well as Bastia and its neighbour counties.

the medium geographical and temporal resolution. Since this area has not been analysed quantitatively, our findings probably provide valuable information for the intervention from the local government. From a methodological perspective, we underline the importance of space and time. Spatial, temporal and spatiotemporal information is very likely missed in many second home analyses. Furthermore, to our knowledge, the approach has not been used for investigating second homes in a given area. In addition to latent spatial and temporal information, the inclusion of the amenity and socio-economic factors offers additional insights in the Corsican second home ratio. However, applications of quantitative techniques in second home analyses are still in their infancy and much remains to be done. The limitation firstly comes from the availability of data. Second, the quantitative method should deal with both spatial and temporal dynamics in data. But addressing these issues substantially complicates model specification or requires more flexible models. Finally, endogeneity of independent variable is still under investigated in Bayesian spatial modelling.

A Appendices

A.1 Intrinsic conditional autoregressive model

The conditional distribution in Eq.3 allows us to derive a joint distribution. With the help of the Brook's lemma (1964), Besag (1974) demonstrates that the joint distribution for random vector, $\Upsilon = (\upsilon_1, \ldots, \upsilon_n)$, is multivariate normal with a mean of 0 and a precision matrix $\sum$ [31].

$$\begin{aligned} &\Upsilon \sim \mathrm{N}\left(0, Q^{-1}\right) \\ &Q = [\tau(D - W)]^{-1} \end{aligned}$$

where D is a diagonal matrix, where diagonal entity d_{ii} is the number of neighbors for region n_i. τ is a precision parameter, $\tau = \frac{1}{\sigma^2}$. Under the condition of the precision parameter of 1 with a fully connected neighborhood graph, the joint distribution can be simplified and then rewritten a pairwise difference formulation:

$$p(\Upsilon) \propto \exp\left(-\frac{1}{2}\sum_{i\sim j}\left(\upsilon_i - \upsilon_j\right)^2\right) \text{[32]} \tag{22}$$

where $-\frac{1}{2}\sum_{i\sim j}\left(\phi_i - \phi_j\right)^2$ is considered as a penalty term. As spatial proximity decreases, penalty strength increases. In the small area estimation literature, it means that a unit borrows more strength from adjacent units than those further apart (See Banerjee, Carlin, and Gelfand (see 2014, chap. 4); Schmidt and Nobre (2018) for more details.).

A.2 Tables

[31] The precision matrix is an inverse of the covariance matrix, $\sum = Q^{-1}$.

[32] p represents the probability density function.

Table 7: Data description and sources

Variable	Desciption	Source
Physical landscape	Physical landscape counts within a county e.g., lakes, mountains, alpine rocks, estuary	Corsica Recreational Areas database, UMR LISA
Cultural landscape	Cultural landscape counts within a county e.g., castles, city walls, towers, churches	Corsica Recreational Areas database, UMR LISA
Coastal county	Dummy variable, 1: coastal county; 0 otherwise	Corsica GIS database, UMR LISA
Mountainous county	Dummy variable, 1: the average elevation $\geq$ 500m; 0 otherwise	Corsica GIS database, UMR LISA
Population	Measured by household	www.insee.fr/fr/information/2008354
Household growth	$(household_{it} - household_{it-1}) / household_{it-1}$	www.insee.fr/fr/information/2008354
Interest rate		webstat.banque-france.fr/en/ quickview.do?SERIES_KEY=243.MIR1.Q. FR.R.A22FRX.A.R.A.2254FR.EUR.C50
Council tax	"Taxe d'habitation"	www.impots.gouv.fr/portail/statistiques
Unemployment rate	At "zone d'emploi" level 360 counties are divided into 7 "zone d'emplois".	www.insee.fr/fr/statistiques/1893230

Table 8: Descriptive statistics for the temporal variation of second home counts

	Mean	Min	Max	St. Dev.
2007	197.336	6	4,465	389.425
2008	203.175	4	4,906	413.219
2009	207.750	4	5,077	426.535
2010	212.553	4	5,115	432.917
2011	220.594	4	5,375	450.401
2012	229.789	4	6,025	482.729
2013	236.519	4	6,025	486.182
2014	244.769	4	6,748	524.319
2015	250.450	4	6,539	524.530
2016	254.512	4	6,581	529.661

Table 9: Tested hyperpriors in the prior sensitivity analysis

	Default	Test 1	Test 2
Spatially Structured Component	$\log\tau = \log\text{Gamma}(1, 5\times10^{-5})$	$\log\tau = \log\text{Gamma}(1, 1\times10^{-4})$	$\text{Prob}(\tau > 0.3/0.31) = 0.01$
Spatially Unstructured Component	$\log\tau = \log\text{Gamma}(1, 5\times10^{-5})$	$\log\tau = \log\text{Gamma}(1, 1\times10^{-4})$	$\text{Prob}(\tau > 0.3/0.31) = 0.01$
Spatially Structured Component	$\log\tau = \log\text{Gamma}(1, 5\times10^{-5})$	$\log\tau = \log\text{Gamma}(1, 1\times10^{-4})$	$\text{Prob}(\tau > 0.3/0.31) = 0.01$
Spatially Unstructured Component	$\log\tau = \log\text{Gamma}(1, 5\times10^{-5})$	$\log\tau = \log\text{Gamma}(1, 1\times10^{-4})$	$\text{Prob}(\tau > 0.3/0.31) = 0.01$
Space-time interaction term	$\log\tau = \log\text{Gamma}(1, 5\times10^{-5})$	$\log\tau = \log\text{Gamma}(1, 1\times10^{-4})$	$\text{Prob}(\tau > 0.3/0.31) = 0.01$

Table 10: Estimated posterior mean and quantiles of the covariates for the strict exogenity test

	Test 1			Test 2 lead-1			Test 3 lead-3		
	mean	0.025quant	0.975quant	mean	0.025quant	0.975quant	mean	0.025quant	0.975quant
Intercept	0.878	0.303	1.432	0.813	0.090	1.506	0.570	-0.516	1.586
physical landscapes	0.093	0.046	0.141	0.099	0.050	0.148	0.093	0.044	0.141
cultural landscapes	-0.033	-0.067	-0.0001	-0.051	-0.088	-0.014	-0.043	-0.080	-0.006
coastal county	0.766	0.573	0.960	0.745	0.547	0.943	0.705	0.506	0.903
mountainous county	0.202	0.059	0.344	0.202	0.056	0.347	0.221	0.077	0.366
$log_2(household_t)$	-0.329	-0.365	-0.293	-0.330	-0.369	-0.292	-0.301	-0.341	-0.261
$log_2(household_{t+1})$				3.369×10^{-5}	-1.587×10^{-6}	6.936×10^{-5}			
$log_2(household_{t+3})$							5.114×10^{-8}	-3.517×10^{-5}	3.547×10^{-5}
household growth	-0.123	-0.336	0.090	-0.188	-0.410	0.033	-0.251	-0.496	-0.007
log(interest rate)	-0.147	-0.203	-0.096	-0.162	-0.242	-0.089	-0.125	-0.282	0.023
log(council tax)	-0.006	-0.029	0.017	-0.0003	-0.024	0.023	0.003	-0.021	0.027
log(unemployment rate)	-0.051	-0.145	0.042	-0.057	-0.166	0.050	-0.129	-0.276	0.017

Table 11: Estimated posterior mean, standard deviation and quantiles of the parameters for different hyperpriors

Prior	Test 1				Test 2			
	mean	St.Dev	0.025quant	0.975quant	mean	St.Dev	0.025quant	0.975quant
Intercept	-1.686	0.252	-2.188	-1.200	-1.744	0.266	-2.274	-1.226
physical landscapes	0.095	0.027	0.042	0.148	0.094	0.027	0.041	0.147
cultural landscapes	-0.052	0.020	-0.091	-0.013	-0.051	0.020	-0.090	-0.012
coastal county	0.745	0.108	0.533	0.957	0.748	0.109	0.533	0.962
mountainous county	0.166	0.079	0.011	0.321	0.166	0.079	0.010	0.321
$\sigma^2_{log_2(household)}$	0.267	0.046	0.188	0.368	0.268	0.045	0.189	0.367
household growth	-0.140	0.101	-0.339	0.059	-0.137	0.101	-0.336	0.062
log(interest rate)	-0.117	0.024	-0.166	-0.070	-0.118	0.031	-0.179	-0.056
log(council tax)	-0.009	0.011	-0.030	0.012	-0.012	0.011	-0.034	0.009
log(unemployment rate)	-0.004	0.043	-0.088	0.079	-0.015	0.043	-0.100	0.068
σ^2_θ	0.13397	0.05550	0.05363	0.26817	0.14581	0.05658	0.06349	0.28274
σ^2_ν	0.26127	0.03539	0.20029	0.33898	0.25719	0.03433	0.19712	0.33171
σ^2_ζ	0.00032	0.00034	0.00002	0.00120	0.00121	0.00185	0.00004	0.00566
σ^2_γ	0.00020	0.00023	0.00002	0.00078	0.00031	0.00054	0.000001	0.00158
σ^2_δ	0.00739	0.00055	0.00634	0.00850	0.00743	0.00055	0.00639	0.00855

Table 12: Refitted models assessment

	DIC	LCPO	RMSE
Model 5	25,257.31	12,978.80	0.0344
Convolution Model5	25,615.81	13,305.67	0.0353

A.3 Graphs

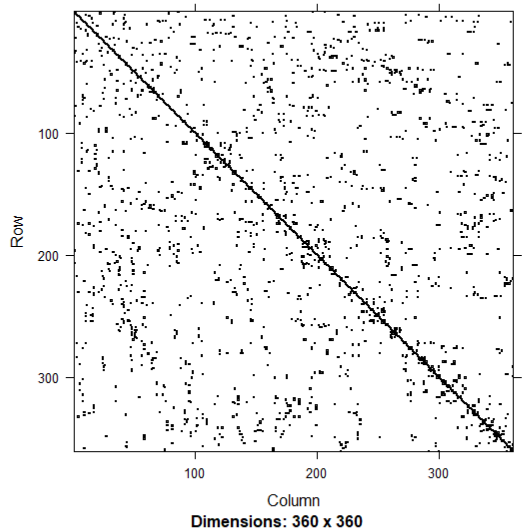


Figure 6: Adjacency matrix: rows and columns identify areas; squares identify neighbors.

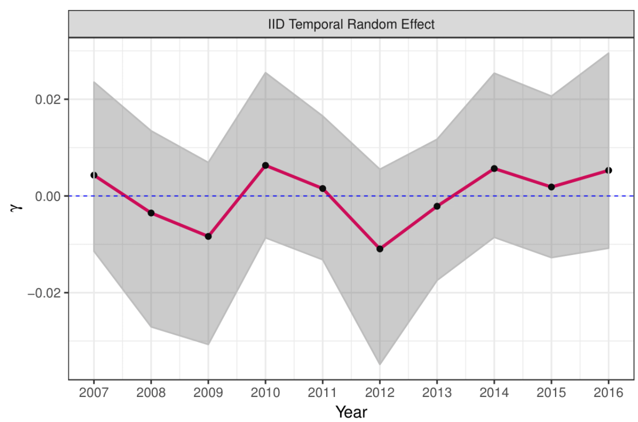


Figure 7: Posterior unstructured temporal trend for second home ratio π_{it} in Corsica.

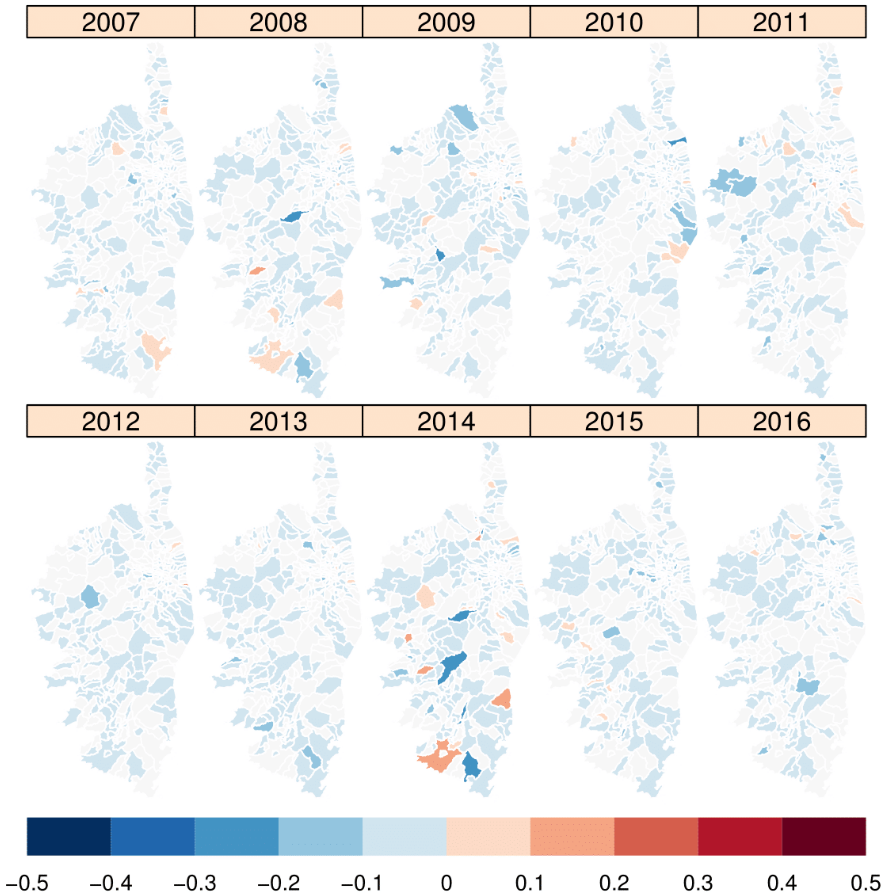


Figure 8: Posterior mean of the type-I spatio-temporal interaction δ_{it} for the log odds of the prbability being second homes.

References

Anselin, L. (1988): *Spatial Econometrics: Methods and Models*, vol. 4. Springer Netherlands.

Back, A., and R. Marjavaara (2017): "Mapping an invisible population: the uneven geography of second-home tourism," *Tourism Geographies*, 19(4), 595–611.

Banerjee, S., B. P. Carlin, and A. E. Gelfand (2014): *Hierarchical modeling and analysis for spatial data*. Chapman and Hall/CRC, Boca Raton.

Barke, M. (1991): "The growth and changing pattern of second homes in Spain in the 1970s," *Scottish Geographical Magazine*, 107(1), 12–21.

——— (2007): "Second Homes in Spain: An analysis of change at the provincial level, 1981-2001," *Geography*, pp. 195–207.

Barnett, R. (2007): "Central and Eastern Europe: Real estate development within the second and holiday home markets," *Journal of Retail & Leisure Property*, 6(2), 137–142.

Bernardinelli, L., D. Clayton, C. Pascutto, C. Montomoli, M. Ghislandi, and M. Songini (1995): "Bayesian analysis of space—time variation in disease risk," *Statistics in medicine*, 14(21-22), 2433–2443.

Besag, J. (1974): "Spatial interaction and the statistical analysis of lattice systems," *Journal of the Royal Statistical Society: Series B (Methodological)*, 36(2), 192–225.

Besag, J., J. York, and A. Mollié (1991): "Bayesian image restoration, with two applications in spatial statistics," *Annals of the Institute of Statistical Mathematics*, 43(1), 1–20.

Breslow, N. E., and D. G. Clayton (1993): "Approximate inference in generalized linear mixed models," *Journal of the American statistical Association*, 88(421), 9–25.

Brezger, A., T. Kneib, and S. Lang (2003): "BayesX: Analysing Bayesian structured additive regression models," Discussion paper.

——— (2005): "BayesX: Analyzing Bayesian Structural Additive Regression Models," *Journal of Statistical Software*, 1(11).

Brook, D. (1964): "On the Distinction Between the Conditional Probability and the Joint Probability Approaches in the Specification of Nearest-Neighbour Systems," *Biometrika*, 51(3/4), 481–483.

Caudill, S., C. Detotto, and D. Prunetti (2019): "Bargaining power in apartment sales in Corsica: A latent class approach," *Urban Studies*.

CdC (2019): "Rapport d'information sur les marchés fonciers et immobiliers," in *Rapport no 2019/E3/241*. Collectivité de Corse.

Coppock, J. T. (1977): *Second homes: Curse or blessing?* Pergamon.

Cressie, N. (1992): "Statistics for spatial data," *Terra Nova*, 4(5), 613–617.

Dower, M. (1977): "Planning aspects of second homes. Di dalam Coppock JT," *Second homes: Curse or blessing*, pp. 210–237.

Farstad, M., and J. F. Rye (2013): "Second home owners, locals and their perspectives on rural development," *Journal of Rural Studies*, 30, 41–51.

Ferrari, A., and M. Comelli (2016): "A comparison of methods for the analysis of binomial clustered outcomes in behavioral research," *Journal of Neuroscience Methods*, 274, 131–140.

Geisser, S. (1993): *Predictive inference : an introduction*. Chapman & Hall.

Geisser, S. (2017): *Predictive inference*. Routledge.

Gelfand, A. E. (2012): "Hierarchical Modeling for Spatial Data Problems," *Spatial statistics*, 1, 30–39.

Giannoni, S., O. Beaumais, and C. Tafani (2017): "Foreign buyers, price dispersion and local buyers eviction in bargaining markets," *TerRA Working Papers No.5.*

Gilks, W. R., S. Richardson, and D. Spiegelhalter (1995): *Markov chain Monte Carlo in practice*. Chapman and Hall/CRC.

Goodwin, H. (2017): "The challenge of overtourism," *Responsible Tourism Partnership.*

Hall, C. M. (2015): "Second homes planning, policy and governance," *Journal of Policy Research in Tourism, Leisure and Events*, 7(1), 1–14.

Hall, C. M., and D. K. Müller (2004): *Tourism, mobility, and second homes: between elite landscape and common ground*, vol. 15. Channel View Publications.

Harris, P. (2019): "A Simulation Study on Specifying a Regression Model for Spatial Data: Choosing between Autocorrelation and Heterogeneity Effects," *Geographical Analysis*, 51(2), 151–181.

Huang, J., B. P. Malone, B. Minasny, A. B. McBratney, and J. Triantafilis (2017): "Evaluating a Bayesian modelling approach (INLA-SPDE) for environmental mapping," *Science of the Total Environment*, 609, 621–632.

Kaltenborn, B. P., O. Andersen, and C. Nellemann (2007): "Second home development in the Norwegian mountains: Is it outgrowing the planning capability?," *The International Journal of Biodiversity Science and Management*, 3(1), 1–11.

Kaltenborn, B. r. P., O. Andersen, C. Nellemann, T. Bjerke, and C. Thrane (2008): "Resident attitudes towards mountain second-home tourism development in Norway: The effects of environmental attitudes," *Journal of Sustainable Tourism*, 16(6), 664–680.

Keane, M. J., and E. Garvey (2006): "Measuring the employment effects of the rural renewal tax scheme," *Regional Studies*, 40(03), 359–374.

Klein, N., T. Kneib, S. Lang, and A. Sohn (2015): "Bayesian structured additive distributional regression with an application to regional income inequality in Germany," *The Annals of Applied Statistics*, 9(2), 1024–1052.

Knorr-Held, L. (2000): "Bayesian modelling of inseparable space-time variation in disease risk," *Statistics in medicine*, 19(17-18), 2555–2567.

Knorr-Held, L., and J. Besag (1998): "Modelling risk from a disease in time and space," *Statistics in medicine*, 17(18), 2045–2060.

Leroux, B. G., X. Lei, and N. Breslow (2000): "Estimation of disease rates in small areas: a new mixed model for spatial dependence," in *Statistical models in epidemiology, the environment, and clinical trials*, pp. 179–191. Springer.

Lindgren, F., H. Rue, and J. Lindström (2011): "An explicit link between gaussian fields and gaussian markov random fields: The stochastic partial differential equation approach," *Journal of the Royal Statistical Society. Series B: Statistical Methodology*, 73(4), 423–498.

Ling, Y. (2019): "Time, space and hedonic prediction accuracy: evidence from Corsican apartment markets," *Annals of Regional Science*.

Linzer, D. A. (2013): "Dynamic Bayesian forecasting of presidential elections in the states," *Journal of the American Statistical Association*, 108(501), 124–134.

López-Quilez, A., and F. Munoz (2009): "Review of spatio-temporal models for disease mapping," Discussion paper.

MacNab, Y. C. (2003): "Hierarchical Bayesian spatial modelling of small-area rates of non-rare disease," *Statistics in Medicine*, 22(10), 1761–1773.

Marjavaara, R., and D. K. Müller (2007): "The development of second homes' assessed property values in Sweden 1991–2001," *Scandinavian Journal of Hospitality and Tourism*, 7(3), 202–222.

Martins, T. G., D. Simpson, F. Lindgren, and H. Rue (2013): "Bayesian computing with INLA: New features," *Computational Statistics and Data Analysis*, 67, 68–83.

Maupertuis, M.-A., C. Tafani, and A. Poggioli (2017): "Les résidents secondaires en Corse: différentes façons de s' ancrer au territoire. L'exemple de la commune de Lumiu," *Téoros. Revue de recherche en tourisme*, 36(36, 2).

Milano, C., J. M. Cheer, and M. Novelli (2018): "Overtourism: a growing global problem," *The Conversation*.

Modrák, M. (2018): "A Gentle Stan vs. INLA Comparison," .

Moran, P. A. P. (1950): "Notes on continuous stochastic phenomena," *Biometrika*, 37(1/2), 17–23.

Müller, D. K. (2002): "Second home ownership and sustainable development in Northern Sweden," *Tourism and hospitality research*, 3(4), 343–355.

——— (2006): "The attractiveness of second home areas in Sweden: A quantitative analysis," *Current Issues in Tourism*, 9(4-5), 335–350.

Müller, D. K., C. M. Hall, and D. Keen (2004): "Second home tourism impact, planning and management," in *Tourism, mobility and second homes: Between elite landscape and common ground*, pp. 15–32. Channel View Publications.

Müller, D. K., and G. Hoogendoorn (2013): "Second homes: Curse or blessing? A review 36 years later," *Scandinavian Journal of Hospitality and Tourism*, 13(4), 353–369.

Norris, M., and P. Shiels (2007): "Housing affordability in the Republic of Ireland: is planning part of the problem or part of the solution?," *Housing Studies*, 22(1), 45–62.

Norris, M., and N. Winston (2009): "Rising second home numbers in rural Ireland: Distribution, drivers and implications," *European Planning Studies*, 17(9), 1303–1322.

Orueta, F. D. (2012): "Second homes," in *International Encyclopedia of Housing and Home*, ed. by S. J. Smith, pp. 237–242. Elsevier.

Paris, C. (2009): "Re-positioning second homes within housing studies: Household investment, gentrification, multiple residence, mobility and Hyper-consumption," *Housing, Theory and Society*, 26(4), 292–310.

R Core Team (2013): "R: A language and environment for statistical computing," .

Rao, J. N. K. (2003): *Small Area Estimation*. John Wiley & Sons, Inc., Hoboken, NJ, USA.

Riebler, A., S. H. Sørbye, D. Simpson, and H. Rue (2016): "An intuitive Bayesian spatial model for disease mapping that accounts for scaling," *Statistical methods in medical research*, 25(4), 1145–1165.

Roos, M., and L. Held (2011): "Sensitivity analysis in Bayesian generalized linear mixed models for binary data," *Bayesian Analysis*, 6(2), 259–278.

Rue, H., and L. Held (2005): *Gaussian Markov random fields : theory and applications*. Chapman & Hall/CRC.

Rue, H., S. Martino, and N. Chopin (2009): "Approximate Bayesian inference for latent Gaussian models by using integrated nested Laplace approximations," *Journal of the Royal Statistical Society. Series B: Statistical Methodology*, 71(2), 319–392.

Schmidt, A. M., and W. S. Nobre (2018): *Conditional Autoregressive (CAR) Model*pp. 1–11. American Cancer Society.

Schrödle, B., and L. Held (2011): "Spatio-temporal disease mapping using INLA," *Environmetrics*, 22(6), 725–734.

Simpson, D., H. Rue, A. Riebler, T. G. Martins, and S. H. Sørbye (2017): "Penalising model component complexity: A principled, practical approach to constructing priors," *Statistical science*, 32(1), 1–28.

Spiegelhalter, D. J., N. G. Best, B. P. Carlin, and A. Van Der Linde (2002): "Bayesian measures of model complexity and fit," *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 64(4), 583–639.

Staubach, C., V. Schmid, L. Knorr-Held, and M. Ziller (2002): "A Bayesian model for spatial wildlife disease prevalence data," *Preventive Veterinary Medicine*, 56(1), 75–87.

Tobler, W. R. (1970): "A Computer Movie Simulating Urban Growth in the Detroit Region," *Economic Geography*, 46, 234–240.

Viola, D., P. S. Arno, A. R. Maroko, C. B. Schechter, N. Sohler, A. Rundle, K. M. Neckerman, and J. Maantay (2013): "Overweight and obesity: Can we reconcile evidence about supermarkets and fast food retailers for public health policy?," *Journal of Public Health Policy*, 34(3), 424–438.

Vogiatzakis, I. N., G. Pungetti, and A. M. Mannion (2008): "Corsica," in *Mediterranean island landscapes: natural and cultural approaches*, vol. 9, chap. 10, pp. 220–244. Springer, Dordrecht.

Williams, D., J. Haworth, M. Blangiardo, and T. Cheng (2019): "A spatiotemporal Bayesian hierarchical approach to investigating patterns of confidence in the police at the neighborhood level," *Geographical Analysis*, 51(1), 90–110.

Wooldridge, J. M. (2010): *Econometric analysis of cross section and panel data*. MIT press.

Zhu, L., D. M. Gorman, and S. Horel (2006): "Hierarchical Bayesian spatial models for alcohol availability, drug "hot spots" and violent crime," *International Journal of Health Geographics*, 5(1), 54.