



HAL
open science

Corsican housing market analysis: Applications of bayesian hierarchical model

Yuheng Ling

► **To cite this version:**

Yuheng Ling. Corsican housing market analysis: Applications of bayesian hierarchical model. Economics and Finance. Université Pascal Paoli, 2020. English. NNT: 2020CORT0011 . tel-03721267

HAL Id: tel-03721267

<https://theses.hal.science/tel-03721267v1>

Submitted on 12 Jul 2022

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



UNIVERSITE DE CORSE-PASCAL PAOLI
ECOLE DOCTORALE ENVIRONNEMENT ET SOCIETE
UMR CNRS 6240 (LISA)



Thèse présentée pour l'obtention du grade de
DOCTEUR EN SCIENCES ECONOMIQUES

Soutenue publiquement par
Yuheng LING
Le 4 Décembre 2020 à 9h

CORSICAN HOUSING MARKET ANALYSIS:
APPLICATIONS OF BAYESIAN HIERARCHICAL MODELS

Directeurs :

M. Dominique Prunetti, Dr-HDR, Université de Corse
M. Claudio Detotto, Dr, Université de Corse

Rapporteurs :

Mme Julie Le Gallo, Pr, Université de Bourgogne Franche-Comté
M. Florent Sari, Pr, Université Paris-Est Créteil

Jury

M. Olivier Beaumais, Pr, Université de Rouen
Mme Julie Le Gallo, Pr, Université de Bourgogne Franche-Comté
Mme Marie-Antoinette Maupertuis, Pr, Université de Corse
Mme Paula Moraga, Dr, University of Bath (UK)
M. Dominique Prunetti, Dr-HDR, Université de Corse
M. Florent Sari, Pr, Université Paris-Est Créteil

Membres invités

M. Claudio Detotto, Dr, Université de Corse
Mme Emanuela Dreassi, Dr, Università Degli Studi Firenze (IT)

Corsican housing market analysis:

Applications of Bayesian hierarchical models

Yuheng Ling

UMR CNRS 6240 LISA

University of Corsica - Pascal Paoli; Campus Mariani, BP52, 20250 Corte,
France.

Contents

Abbreviations	xi
1 Introduction	1
1.1 Introduction	1
1.2 Corsica context	3
1.3 A glance at the literature	8
1.4 Research gaps.	10
1.5 Aims of the thesis	13
1.6 Research approaches	15
1.6.1 Data source	15
1.6.2 Statistical models.	16
1.6.3 Bayes inference.	17
1.7 Thesis organization	17
2 Literature review on housing analysis	19
2.1 Introduction	19
2.2 Housing and wealthy	20
2.3 Specific characteristics of housing	22
2.4 Methods of housing analysis	23
2.5 Hedonic Price Model (HPM)	28
2.5.1 A brief history	28
2.5.2 Theoretical foundation.	30
2.6 Location and housing prices	32
2.6.1 Background	32
2.6.2 How is location measured in housing valuation?	34
2.7 Location and housing analysis: a regional perspective.	37
2.7.1 Ripple effects	37
2.7.2 Second homes.	38

2.8	Summary	39
3	Quantitative models for housing analysis	41
3.1	Introduction	41
3.2	Spatial econometrics and housing analysis	42
3.2.1	Spatial autocorrelation and spatial heterogeneity.	42
3.2.2	Basic spatial econometric models	44
3.2.3	Some alternatives	45
3.3	Spatial data	47
3.4	Analysis of lattice data.	48
3.4.1	Markov random fields	48
3.4.2	Conditional Autoregressive (CAR) model family	49
3.4.3	Besag, York and Mollie (BYM) model and Leroux's model.	52
3.5	Analysis of geostatistical data	53
3.5.1	Gaussian process	54
3.5.2	Matérn covariance function.	54
3.5.3	Stochastic partial differential equation approach.	55
3.6	Spatiotemporal extensions	56
3.7	Structured additive regression models	60
3.7.1	Latent Gaussian models.	61
3.7.2	Spatial components in latent Gaussian models	62
3.8	Bayesian statistics	63
3.8.1	Bayes' rule.	64
3.8.2	Integrated nested Laplace approximations comput- ing scheme	66
3.9	Model assessment	69
3.10	Summary	71
4	Time, space and hedonic prediction accuracy: evidence from Corsican apartment market	73
4.1	Introduction	73
4.2	Literature review: research trends in the HPM-based prop- erty valuation	75

4.3	Spatiotemporal modeling for geostatistical data.	78
4.3.1	Spatiotemporal data	78
4.3.2	Bayesian hierarchical models with space-time random effects	79
4.3.3	INLA-SPDE approach	81
4.4	The case of Corsican apartment markets.	82
4.4.1	Data	82
4.4.2	Model discussion	82
4.4.3	Implement details	86
4.4.4	Model assessment criteria	86
4.4.5	Results and discussion	88
4.5	Concluding remarks	94
5	Unveiling spatial and temporal patterns of second home dynamics: a Bayesian spatiotemporal analysis for a Mediterranean island	97
5.1	Introduction	97
5.2	Second homes in Corsica: An exploratory spatial data analysis	100
5.3	Spatiotemporal modeling for areal data.	105
5.3.1	Spatiotemporal data	105
5.3.2	Modelling spatial dependence	106
5.3.3	Modelling temporal dependence and spatio-temporal variability	107
5.3.4	Candidate model specifications	108
5.4	Empirical results	110
5.4.1	Model assessment results.	110
5.4.2	Fixed effects and interpretation	111
5.4.3	Spatial and temporal random effects and interpretation	113
5.5	Robustness Check	116
5.6	Concluding remarks	120

6	The value of sea views in a Mediterranean island: Hedonic analysis with spatial multilevel modelling	123
6.1	Introduction	123
6.2	Literature review	125
6.2.1	What is in a view?	125
6.2.2	How to measure a view?	126
6.2.3	Spatial effects at different spatial scales	127
6.3	Spatiotemporal modeling for spatial data with multilevel structures	128
6.3.1	Leroux's conditional autoregressive model	128
6.3.2	spatial multilevel models	129
6.3.3	Space-time extensions	131
6.4	Empirical analysis	132
6.4.1	Generating sea view Index	132
6.4.2	Measuring beach access	133
6.4.3	Data	133
6.4.4	Econometric strategies	134
6.5	Results	138
6.5.1	Model assessment	138
6.5.2	Interpreting covariate effects	139
6.5.3	Interpreting random effects and visualization	142
6.5.4	Policy implications	144
6.6	Concluding remarks	145
7	Summary and Future works	147
7.1	Main Findings	148
7.2	Main contributions	151
7.3	Strengths and Weaknesses of Each Model	152
7.4	Future works	153
	Appendix	157
7.5	Tables	158
7.6	Graphs	160

8 Bibliography	163
-----------------------	------------

Acknowledgements	191
-------------------------	------------

Abbreviations

AR(1)	First-order autoregressive process
HPM	Hedonic price method
CAR	Conditional autoregressive process
CI	Credible interval
GMRF	Gaussian Markov random field
GRF	Gaussian random field
ICAR	Intrinsic conditional autoregressive
IID	Identically independent distribution
INLA	Integrated nested Laplace approximation
LCAR	Leroux's conditional autoregressive
HSAR	Hierarchical spatial autoregressive
MAE	Mean absolute error
McMC	Markov chain Monte Carlo
ML	Maximum likelihood
PCC	Pearson correlation coefficient
REML	Restricted Maximum likelihood
RMSE	Root mean square error
RW1	First-order random walk
SAR	Spatial lag model
SPDE	Stochastic partial differential equation
St.Dev	Standard deviation

1

Introduction

1.1. Introduction

Housing plays a crucial role in both family assets and local/national economies. From a consumer or investor's view, housing is an essential asset for a family, thus accurately assessing residential housing values is important for making investment decisions (Li, 2016). From a local authority perspective, understanding local housing markets could enable planners to allow land to develop and then promote the local economy (Mitchell *et al.*, 2015). For national policymakers, housing relates to macroeconomic stability, e.g., economic growth, inflation rates, unemployment (Reichert, 1990). As a complex matter, housing is usually affected by several social and economic factors. Nowadays, with the increasing availability of spatial data and advances in geographical information systems (GIS), more and more economists notice that locational/geographical factors play a crucial role in determining housing prices (Anselin, 2010) and in provoking regional market comovement (Abbott and De Vita, 2012; Holmes *et al.*, 2017a; Holmes *et al.*, 2017b). As spatial econometric techniques took a huge step forward over the past thirty years, it became a useful instrument in the economists' toolbox that helps economists investigate the locational/geographical factor, including spatial dependence and spatial hetero-

geneity.

Corsica is one of the 18 French administrative regions, whose housing market shows some specific phenomena. As a beautiful island in the Mediterranean Sea, Corsica attracts not only thousands of tourists (Torre, 2018) but also foreign real estate investors (Touzani, 2018). With the augmentation of the population (Bretel, 2018), the need for housing has never ceased. Additionally, as an island, land supply is always limited. This probably leads to high housing prices. All suggest that a detailed examination of the Corsican housing market is timely.

Furthermore, only a small number of studies have focused on the issues relating to the Corsican housing market. Kessler and Tafani's monograph (Kessler and Tafani, 2015) contained several chapters investigating the Corsican housing and land market from economic and geographic perspectives. For example, economists clearly showed that there existed two distinct property markets, i.e. a market for local buyers and a market for external investors. They also emphasized that the second home phenomenon should be further studied. Recently, new empirical studies (Caudill *et al.*, 2019; Giannoni *et al.*, 2017) on these topics arise with new datasets.

Along with these recent studies, many relevant questions still remain to be answered. Further, the increasing availability of large geo-referenced datasets and high resolution maps, as well as novel statistical tools can help economists better analyze these questions. This thesis, therefore, offers a twofold contribution to the existing literature. The first contribution relates to methodological issues. It is known that real estate and spatial models naturally complement each other. The traditional spatial econometric models¹ like spatial fixed effects (dummy variables) models, spatial lag models and spatial error models may experience some limitations if they are used for analyzing cross-sectional housing sales data and data collected over multiple periods (Dubé and Legros, 2013b). To overcome these drawbacks, I propose several new approaches in this thesis, and such approaches adapt to spatial and temporal correlation in housing data and provide accurate predictions of the dependent variable and marginal effect estimates. More importantly, how locational/geographical

¹"...spatial econometrics consists of a sub-set of econometric methods that is concerned with spatial aspects present in cross-sectional and space-time observations." (Anselin, 2010)[p.1213]

features (spatial patterns) affect all aspects of the Corsican housing market is the key question throughout this thesis. Additionally, I focus on forecasting Corsican apartment prices and identifying geographical “hot spots” and “cold spots” in terms of housing prices. I also explore second home phenomena in Corsica. I demonstrate that the “ripple effects” exist in the Corsican second home market. The findings clearly show the county belonging to “hot spots” in terms of second home rates. Lastly, I estimate the implicit price of Mediterranean Sea views with the consideration of the Corsican regional context.

The remainder of this chapter is structured as follows. In Section 2, I detail the Corsican context, in particular the background information about the regional housing market. In Section 3, I briefly introduce relevant literature. I highlight research gaps in Section 4. Section 5 discusses research questions and the aim of the thesis. Section 6 glances at research approaches, and Section 7 details the thesis organization.

1.2. Corsica context

Corsica is an island located in the Mediterranean Sea. It is the fourth largest island in the Mediterranean Sea. It is about 183 km in length (north-south) and 83 km in width (east-west). It is located south-east of the French mainland, north of Sardinia and west of the Italian Peninsula.

As “the pearl of Mediterranean”, Corsica is famous for its rich tourism resources (Vogiatzakis *et al.*, 2008). A single mountain range crosses the center of the island with alpine landscapes, while beautiful beaches and seaward cliff dominate coastal areas. In particular, Calvi and Porto Vecchio (ID. 73 and ID. 249), located in the northwest and southeast, are famous for their sandy beaches. Corti (ID. 195) with many mountains and historical sites is suited in the center of Corsica.

Corsica is one of the 18 French administrative regions consisted of 360 counties (called “communes”). Bastia and Ajaccio are the two main cities (“communes”) located in the north part and the south part of the island, respectively. Other cities include Corti, Calvi, Bonifacio, Porto Vecchio, etc. In contrast, most small villages (“communes”) are often hidden in the mountains

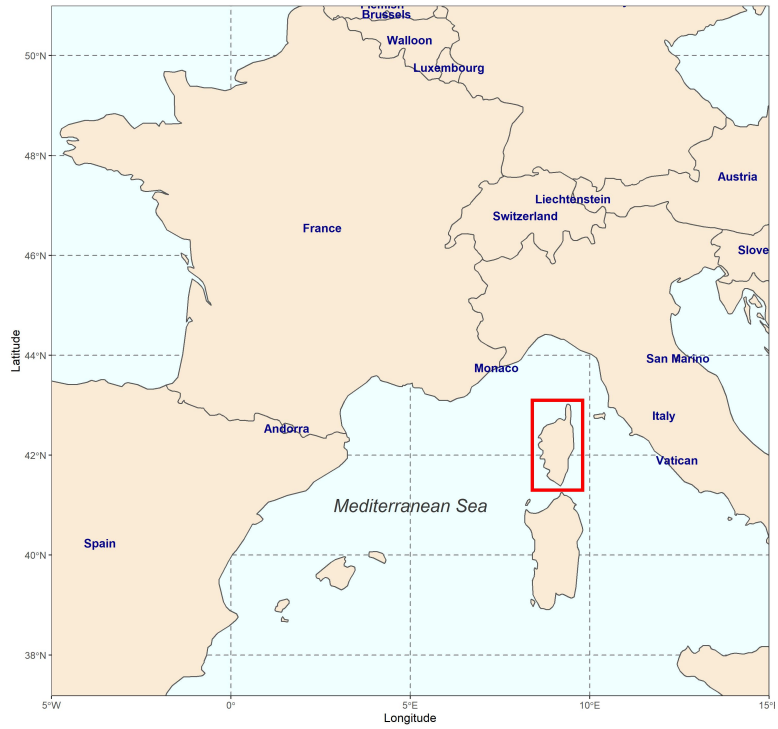


Figure 1.1: The location of Corsica (within the red rectangle).

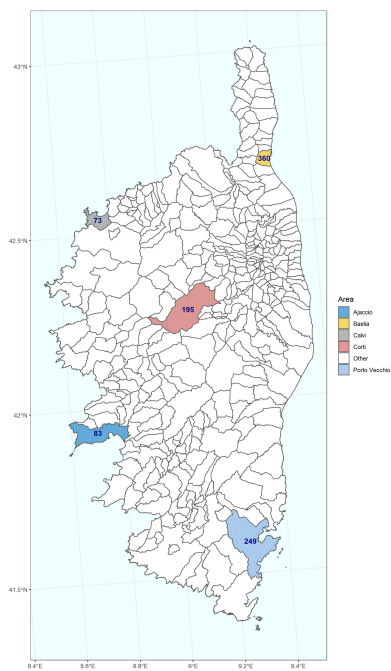


Figure 1.2: The above-mentioned Corsican counties.

such as Soveria, Lama, Olmo, etc.

Due to the particular topography, Corsica owns rural, semi-urban and urban landscapes and inhabitant activities are widely affected. Along with the growing population over the past four decades, population distribution and densities express significant spatial heterogeneity.

INSEE (National Institute of Statistics and Economic Studies) reported that from 1982 to 2011, Corsican inhabitants increased by 74 000 (Tempier, 2014). The corresponding annual growth rate was 0.9%, nearly double the French national level (0.54%). Concerning the geographical distribution of population, most inhabitants lived in the two large urban zones (called “grandes aires urbaines”), the Bastia area and the Ajaccio area. Over the past thirty years, the two zones accounted for about 67% of the additional population, and most additional population was received in the peripheral counties of the two cities, rather than in the city center. Conversely, the population in small rural counties (less than 250 inhabitants) decreased. In 2011, 61.4% of the total inhabitants live the two large urban zones. Other urban areas received 17.4% of the total population in 2011. These urban areas include Porto-Vecchio, Calvi, Corti, Penta-di-Casinca, I’île-Rousse, Propriano, Ghisonaccia. The rest population (21.2% of the population) dispersed in rural areas. In addition to the urban-rural distinction in the population distribution, a coastal and inland difference was also observed. In 2011, 79.3% of the population lived in coastal counties, whereas the rest lived in inland counties. The gradual growth of the population continued in Corsica. In 2013, the total population in Corsica reached 320 200 (Bretel, 2016). This number rose to 324 200 in 2014, to 327 000 in 2015, and to 330 500 in 2016 (Bretel, 2017a; Bretel, 2017b; Bretel, 2018). Nevertheless, the geographical distribution of the population did not change significantly.

Apart from the continuous population growth, INSEE also highlighted that significant residential migration has emerged in Corsica. In 2013, there were approximately 8 000 immigrants but 4 000 emigrants. The immigration rate defined by immigrants per 1 000 inhabitants was around 25.7, ranking the Corsica (upper Corsica and southern Corsica) 83th out of 99 French departments. Moreover, the emigration rate (13.3 emigrants per 1 000 inhabitants approximately) made Corsica the last one among 99 French departments. In

contrast to these low rates, with 12.5 net migration rate (12.5 net immigrants per 1 000 inhabitants), Corsica ranked 3rd out of all French departments. It should be noted that the negative net migration rate was found for people between the ages of 22 and 25. This situation continued for the following years. In 2016 there were 7 100 immigrants but 4 800 emigrants (Tirroloni, 2017; Luciani, 2020). Corsica ranked the 2nd out of 13 French metropolitan regions in terms of the immigration rate and 5th in terms of the emigration rate. As the net migration rate was positive in 2016, the immigrants of different ages are evenly-distributed. All these immigrants needed housing.

As previously stated, Corsica is well-known for its rich tourism resources. Nearly 9.5 million overnight stays were recorded at Corsican hotels in 2015 (Luquet, 2016). The regional tourist office also reported that the overnight stays recorded in the Ajaccio area occupied 24% of the total amount recorded in the region. The overnight stays recorded in the Bastia area, Balagne area and Extreme Sud area occupied 31.4%, 18.8%, and 25.8% respectively. The recorded overnight stays reached 9.7 million (Tirroloni *et al.*, 2017) in 2016, with the proportions being 41.2%, 23.3%, 18.4% and 17.1% for the Ajaccio area, Bastia area, Balagne area and Extreme Sud area, respectively. The increase in overnight stays continued. In 2017, this number was nearly 10.8 million (Genovese-Bolleyn and Torre, 2018), and the proportions were 42.3%, 23.3%, 16.9%, 17.5%. In a word, a growing number of inhabitants, ceaseless immigration and growing tourists put intense pressure on the local housing market. The strong need for housing remains.

On the other hand, the supply of housing has indeed increased. In 1999, there were 177 460 houses irrespective of housing types. This number reached 222 460 in 2011 and 242 100 in 2015 (Pedielli and Bretel, 2015; Touzani, 2018). From 2011 to 2015, the annual growth in housing supply reached 5 000 units. In response to the population growth, the government planned to boost housing supply. From 2010 to 2030, around 70 000 new houses would be built (Balzer and Murati, 2015). Nevertheless, a number of houses were bought and converted to second homes. In 2015, Corsica had the highest second home rate among all French regions. This rate equaled 37.2%, while the national second home rate was 9.6%. Additionally, low-rental housing was rare in Corsica. There was no low-rental housing in several counties. At

the regional level, the rate of low-rental housing was around 26%, which was much lower than the national level (42.1%) in 2009 (Insee Corse, 2010).

Concerning housing prices in Corsica, the average price of a property in an apartment building was €3 599 per m^2 in 2015. Such a high price made the Corsica become the fifth most expensive region in terms of housing prices. Further, housing appreciation has never ceased. The price increased by 5% compared with that in 2014, and the price augmented 3.5% from 2012 to 2013. There was also an intra-regional difference. The average price reached €4 024 per m^2 in Southern Corsica, which was higher than that in Upper Corsica (€3 152 per m^2).

As is well known, housing prices are often correlated with land prices. According to an official report (Préfet de Corse, 2017), from 2006 to 2011, Corsican land prices experienced a surge. Within this period, the average land price rose from 30% to 50% depending on areas. The residential land price was around €60 per m^2 , which was higher than that in other French regions, holding the demographic increase constant.

In a word, after decades of appreciation, housing prices and residential land prices attained a very high level in Corsica. Even though local authorities attempted to boost housing supply, the phenomena such as the rapid growth of second homes and the low rate of low-rental houses remained in the Corsican real estate market.

1.3. A glance at the literature

Although the old cliché “location, location, location” has been used in the housing business for a century, it has not been abandoned. The cliché was firstly underpinned by the monocentric model derived from the work of Alonso (1964), Muth (1969) and Mills (1972). In simple, the monocentric model shows how land prices, land types and population density changes relate to the distance to the city center. The development of the monocentric model is also considered as the birth of modern urban economics. More recently, the economic underpinning of the cliché has moved to the polycentric models proposed by Fujita and Ogawa (1982), but the location is still the key determinant of property values. Not surprisingly, within the polycentric framework, the lo-

cation is still associated with remunerative economic opportunities, scenic and desirable landscapes and amenities, and transport networks, etc. In a word, virtually every attribute of a house beyond its physical dimensions can be reduced to locations.

After briefly exploring the economic foundation of the location, my interest now moves to assessing the location. The assessment is usually based on the so-called hedonic price method. Proposed by Rosen (1974) in a seminal paper, the hedonic approach has enjoyed a great deal of success in the empirical literature investigating housing markets, e.g., housing appraisals, calculating the implicit price of non-market goods.

Hedonic housing price models have evolved alongside the development of spatial econometric models. In the early stages, housing location was often represented by accessibility (Heyman *et al.*, 2018). The fact is that researchers considered the location of a house as the geographical position in relation to its neighbours, and accessibility provided a simple and straightforward measure of it. More recently, some researchers noticed that although accessibility attributes were taken into account, there were still spatial correlations in residuals (Koschinsky *et al.*, 2012). This led to the application of spatial econometric models, since spatial econometrics has provided a feasible way to account for spatial effects in models of housing values. Additionally, in the early stages, the computational burden made researchers implement linear regression models, rather than spatial econometric models. The development of the spatial two-stage least squares estimator (Kelejian and Prucha, 1998) and the acceptance of maximum likelihood methods hence accelerated the use of spatial econometrics.

To my knowledge, the combination of spatial econometrics and the hedonic approach may date back to the late 80s and the early 90s. Can (1990) and Can (1992) published a series of articles discussing the relationship between housing price differentials and housing locations. In these articles, traditional hedonic price models were extended to incorporate spatial neighborhood dynamics, and this led to the use of spatial econometrics. She also discussed the advantages of using these models. After then, empirical applications of spatial econometrics and the hedonic approach have increased dramatically. Geoghegan *et al.* (1997) studied the implicit price of landscapes

through spatial econometrics within the hedonic price framework. More recently, Anselin and Lozano-Gracia (2009) published a paper discussing the fusion of spatial econometrics and the hedonic pricing method, namely “spatial hedonic models”. Several scholars (Anselin, 2010; LeSage, 2014) argued that the spatial hedonic model has become the mainstream of quantitative methods in analyzing housing markets. In recent years, due to the diversity and complexity of regional affairs and contexts, traditional spatial econometric models have faced criticism and more and more sophisticated spatial econometric/statistical models have been developed (Dong and Harris, 2015; Shi and Lee, 2018). In summary, the monocentric, polycentric model and the hedonic price model lay the microeconomic foundation of housing prices.

Regarding regional housing markets, Meen (1996) introduced the so-called “ripple effect”, which describes a market phenomenon, where housing price shocks in the Greater London region spread out their influence to neighbouring regions and eventually to the rest of the country over time. Chiang (2014) specified the ripple effect as a kind of local housing price comovements. In particular, the ripple effect is presented in a lead-lag manner, begins with the spillover of housing prices in a specific region to its neighbouring regions and it eventually leads to comovement among local housing prices. Meen (1999) also stated that the economic mechanism behind ripple effects was still unclear, but he proposed four factors that may facilitate the ripple effect, including migrations, equity transfer, spatial arbitrage and spatial patterns in the determinants of housing prices. Many scholars (Balcilar *et al.*, 2013; Chiang, 2014; Kyriazakou and Panagiotidis, 2018; Teye and Ahelegbey, 2017) have provided empirical evidence that the ripple effect exists in the housing market in many different counties and regions.

1.4. Research gaps

The existing literature on evaluating housing prices and on investigating regional housing markets was extensive. However, regarding the Corsican housing market, empirical analyses were scarcer. A few economic studies centered on the impact of second dwellings purchased by foreign investors.

For example, Maupertuis *et al.* (2017) looked into Corsican coastal counties

and second home buyers who lived there. Taking Lumio as an example, they identified four factors that might explain the purchase, including previous connection with the county, physical characteristics of the county, local culture and social network, and the relationship between buyer and the local society. Giannoni *et al.* (2017) proposed using a bargaining framework to analyze foreign investors in the Corsican land/housing market. They clearly showed that there is the so-called non-local buyer premium, which drove local buyers away from the market. However, they illustrated this mechanism by analyzing farmland transaction data. Caudill *et al.* (2019) investigated the segmentation of Corsican housing markets through Harding, Rosenthal, *et al.*'s (2003) bargaining model. They finally found that there were two distinct submarkets in the Corsican housing market. In the first submarket, asymmetric bargaining power arose for local French and non-French residents. In contrast, similar bargaining power arose for local French, Corsican and non-French residents in the second submarket. Although these studies offer some enlightenment about the Corsican housing market, many relevant phenomena remain unexplained and questions remain unanswered.

In addition to the research gap related to the Corsican housing market, some methodological drawbacks exist in the aforementioned studies and also in the real estate valuing literature. For example, none of the aforementioned studies account for unexplained spatial effects in empirical analyses. To improve these works, spatial econometrics may provide a feasible way. However, as indicated previously, spatial econometrics experiences some criticism and much of it relate to the specification of spatial weighting matrices. Corrado and Fingleton (2012) argued that in empirical practice, it is common to build a spatial weighting matrix based on a rule of thumb or the prior knowledge of data. The simplified weighting matrix does reduce computational workloads, but it neglects economic implications. Corrado and Fingleton thus suggested that economic theories, rather than a rule of thumb should serve as a guide to generate an appropriate spatial weighting matrix. McMillen (2012) criticized misspecifying spatial weighting matrices in empirical analyses, and concluded that spatial econometric models seemed over-used. He found that nonparametric models (locally weighted regression) offered accurate prediction and marginal effect estimates relative to spatial econometric

models with misspecified spatial weighting matrices. Brady and Irwin (2011) argued that spatial econometric models were used in the context, in which the primary interest was to estimate the relationship of neighboring values of the dependent variable to itself. Furthermore, these models were sensitive to the design of weighting matrices. In most hedonic housing applications, the goal was probably not to identify that relationship, but instead accurate parameter estimates that were robust to model misspecification and unobserved spatially correlated variables. They also suggested that nonparametric and semiparametric models such as spline regression, Fourier analysis, locally weighted regression, and kernel methods were alternatives of spatial econometric models in the context, in which spatial weighting matrices were difficult to build or select.

Another methodological gap concerns the temporal dimension of housing data. The spatial effects within the hedonic housing price model have been intensively discussed over the past decades, but temporal dimensions have attracted very little attention. In many hedonic price applications, housing sales data consist of cross sections pooled over time (known as repeated cross sections), and only a few applications (Kiel and Zabel, 2008) have panel data, i.e., repeated sales data. Additionally, the time dimension is sometimes neglected in applications, where transactions are effected within a short period. Conversely, in the case of fairly long periods, many researchers prefer the time dummy variable approach, which is simple and straightforward. The time dummies are mainly used for capturing temporal heterogeneity. Waltl (2016) believed that time should be treated as a continuous variable, and the application of time dummies is a compromise between temporal dimensions and model complexity. However, repeated cross sections make things more complicated, since data contain continuous-space information together with continuous-time information. Moreover, the temporal relation generates a unidirectional effect, which usually requires the inclusion of an autoregressive component in models. In general, there are two ways to tackle both spatial and temporal dimensions continuously. Waltl (2016) modeled the housing price variations by means of two individual nonparametric functions, meaning that there is no interaction between space and time. In other words, housing prices have a common evolving pattern in space, and also a common pat-

tern in time. The house price variation in space-time is expected to be much more flexible. Cohen *et al.* (2017) combined the time dimension with the continuous-space through the nonparametric part of local polynomial regression, and this allowed the housing price variation across the space to vary over time. Even though this approach can capture the spatiotemporal variation in housing prices, there is a flaw in the sense that continuous-time information is overlooked again. Consequently, it is necessary to propose a spatiotemporal model for filling this research gap.

The last gap in the literature that I would like to highlight is the data type. Generally speaking, economists distinguish three basic types of economic data, i.e., cross-sectional data, time series data and Panel data. This classification may seem inappropriate for analyzing housing markets, as housing and spatial models complement each other. Housing data could be recognized in two basic types, the geostatistical (point-referenced) data and areal data. More precisely, individual housing sales collected over space and time or micro-data² related to housing pertain to geostatistical data. In contrast, data coming in an aggregated form with locational information often pertain to areal data. Different data types thus correspond to different models in terms of tackling spatial effects. For example, simultaneous and conditional autoregressive processes are apt to handle areal data, while Gaussian processes are apt to tackle geostatistical data. For an economist, it is better to recognize the data type before starting statistical modeling.

1.5. Aims of the thesis

This thesis makes a twofold contribution to the literature. I attempt to fill the above-mentioned methodological gaps. Then, I provide empirical evidence on how dispersal mechanisms impact the Corsican housing market. Towards these goals, a number of issues related to the Corsican housing market are investigated through spatiotemporal Bayesian hierarchical models. These issues roughly include finding the determinant of housing prices, identifying second home rate growth spatially and temporally, and recognizing the economic benefits of some environmental amenities.

²Dubé and Legros defined the micro-data as

“Observations that are points on a geographical projection...”. (Dubé and Legros, 2014)[p.xi]

The subject of the thesis belongs to the housing economics (Kingsbury, 1941), and these empirical analyses are rooted in the Corsican context. More precisely, the first empirical analysis mainly looks into apartment price predictions. In order to assess the apartment prices across the entire region, we need models to predict housing prices, where there are no recorded sales. To address this problem, I focus on the spatial attributes of housing, since the housing location is an important determinant of its price. On the other hand, the spatial autocorrelation usually occurs in apartment prices. Additionally, housing transaction time cannot be ignored either. To incorporate both fixed covariate effects and spatial or spatiotemporal random effects, I consider the class of Bayesian hierarchical models. I initially propose four candidate model with increasing complexity, and I attempt to select the most effective model that that is satisfactory with respect to the goodness of fit, predictive power and computational costs.

In the second empirical, I intend to demonstrate the existence of ripple effects in the Corsican second home rates. In doing so, I investigate spatial patterns for second home rates at the county level in Corsica and how these patterns change over time, after gauging covariate effects that may affect the second home rate. Possible explanations for the ripple effect include that the rapid growth of second home prices could shift the second home demand to the surrounding areas. In addition, second home spillovers from one county to another may coincide with the expectation of housing sellers in surrounding counties. In the context of the high demand for surrounding counties, property developers would like to build new houses for second home buyers and second home buyers may also choose second-hand houses. This, in turn, accelerates the second home rate among the surrounding counties. It is important to note that second home rates data come in an aggregated form, which implies that second home counts and total house counts within each county are collected. To handle such data and to demonstrate the existence of ripple effects, a spatiotemporal model within the Bayesian hierarchical framework is proposed, where spatial patterns are modeled using the conditional autoregressive process.

The previous chapters use point-referenced data and areal data respectively. One of the motivations for the third empirical analysis comes from the

methodological aspect. How do we model spatially misaligned housing data? Such data refer to the combination of the abovementioned point-referenced data and areal data. Concerning the misalignment, I propose a spatial multi-level model within the Bayesian hierarchical framework. Furthermore, I intend to estimate the implicit price of sea views and closeness to beaches based on the proposed multilevel model. It is widely accepted that living close to environmental amenities like coasts provides a large amount of welfare to households. Coastal areas offer not only aesthetic elements for local landscapes, but also space for recreation, leisure and viewing. As an island in the Mediterranean Sea, Corsican coastal areas attract the attention of house buyers. For example, households living within the viewing zone of the Mediterranean Sea could benefit sea views. In this context, understanding the environmental amenity value is useful for planners and local policymakers, because the better recognition of the economic impacts of environmental amenities could provide justification to preserve them, and minimize negative impacts associated with urbanization.

To summarize, this thesis contains three empirical analyses with different topics about the Corsican housing market. A crucial issue throughout the thesis is how locational/geographical features affect the Corsican housing market, i.e., housing prices, sub-regional second home rates and the implicit price of some amenities. I also look into the impact of the time dimension. In doing so, several dynamic spatiotemporal models are proposed. Spatial and temporal components are investigated jointly and separately within these models.

1.6. Research approaches

1.6.1. Data source

In this thesis, data are obtained from different sources. In the first empirical chapter, I use the "PERVAL" dataset obtained from "Notaire de France". It is important to note that the "PERVAL" dataset registered all the apartment sale contracts. Further, the dataset provides information from housing transaction prices to apartment characteristics. Over 10 000 transactions are registered spanning from 2006 to 2017, and I, therefore, have a repeated cross-sectional dataset. From a geostatistician's point of view, these data belong to geosta-

tistical (point-referenced) data (Cressie, 2015).

In the second empirical chapter, data are acquired mostly from open sources, e.g. the website of INSEE and “Banque de France” (see Table 7.1 in Appendix 7.5). Research center “LISA” provides data about Corsican tourist attractions. Based on these data, a panel dataset is built. The data contain 360 spatial units and cover the period from 2006 to 2016. Again, these data are areal data.

In the third empirical chapter, the above-mentioned two datasets are combined. I, therefore, obtain misalignment spatial data, meaning that the spatial data at different scales of spatial resolution do not coincide (Gelfand *et al.*, 2010). The misaligned dataset contains 6 377 observations from 2006 to 2016.

1.6.2. Statistical models

A classical linear model is inappropriate to model spatial data if spatial effects emerge. Additionally, classical spatial econometric models have some intrinsic properties, which severely limit its use. I, therefore, keep my eyes on the so-called geoaddivitive models proposed by Kammann and Wand (2003). Some researchers (Fahrmeir, Kneib, *et al.*, 2013) classified the geoaddivitive model into the latent approach relative to the direct approach such as spatial lag models. To be more specific, in a geoaddivitive model, the mean and variance of a response variable are functions of both fixed covariate effects and the spatially structured random effects. There are several specifications for capturing the spatial effects, including Markov random field (MRF) models³, Gaussian random field (GRF) models (Cressie, 2015), two-dimension (2D) splines (Toraichi *et al.*, 1988). Further, according to Rue, Martino, *et al.* (2009), some geoaddivitive models can be rewritten into a latent Gaussian model⁴ and hence adapt to Integrated Nested Laplace Approximations (INLA) based inference.

As stated previously, there are several candidate specifications for the spa-

³e.g., Conditional Autoregressive (CAR) models (Besag, 1974); Besag-York-Mollie (BYM) models (Besag *et al.*, 1991)

⁴Geoaddivitive models, along with generalised linear models, generalised additive models, smoothing-spline models, state-space models, semiparametric regression, spatial and spatio-temporal models, log-Gaussian Cox-processes belong to a so-called structured additive regression (STAR) family. Latent Gaussian model is a subset of the STAR family.

tial effects. In the Bayesian paradigm, these specifications typically work as priors. The conditional autoregressive or Besag-York-Mollie (BYM) model is used for capturing discrete spatial structures and often induces sparsity in a precision matrix. While continuous spatial structures can be modeled by a GRF represented by a Matérn covariance field. Then, via a Stochastic Partial Differential Equation (SPDE) approach, the GRF is approximated by a Gaussian Markov random field (GMRF) and then generates a sparse precision matrix (Lindgren *et al.*, 2011). The sparsity produces significant computational gains for INLA-based inference.

1.6.3. Bayes inference

All models used in this thesis are fitted by the Integrated Nested Laplace Approximation within the Bayesian paradigm. The change from frequentist approaches to Bayesian approaches is motivated in different ways. Common reasons include accounting for uncertainty from data, processes and parameters, incorporating prior knowledge, avoiding interpretation of confidence intervals and p-values. As such, the Bayesian approach should provide a more reasonable interpretation of parameters and adapt to fit various complex spatial models. More details about the research approach can be found in Chapter 2.

1.7. Thesis organization

This thesis is organized into seven chapters as follows: Chapter 1: General Introduction gives a brief overview of this thesis, including study areas, a short literature review, research gaps, research aims and research approaches.

Chapter 2: Literature review discusses the economic basis of the hedonic price method, housing locations and ripple effects.

Chapter 3: Quantitative models for housing analysis provides a detailed examination of different econometric/statistical tools. In particular, I look into the latent Gaussian model with different spatial components and estimation techniques (integrated nested Laplace approximations for Bayesian analysis).

Chapter 4 corresponds to the first empirical chapter. A set of Bayesian hierarchical spatio-temporal models coupled with the INLA-SPDE approach are

specified. Our aim is to select the most effective model that accounts for the goodness of fit, the prediction capability and the computational costs. This model is then used to make precise housing price predictions. Moreover, the spatiotemporal random effects are displayed on maps to analyse their impacts on housing prices.

Chapter 5 corresponds to the second empirical chapter. A Bayesian generalized additive regression model is developed to investigate the spatial and temporal patterns of Corsican second home rates. I identify several counties as “hot spots” in terms of second home rates and policymakers should be aware of the second home issue in these counties. Methodologically, I highlight the usefulness of CAR models working as priors.

Chapter 6 corresponds to the third empirical chapter. I review the literature on scenic amenities and then generate a subjective, continuous measurement of views. I also proposed a spatial multilevel hedonic model to estimate the implicit price of sea views on Corsican apartment prices in the presence of nested data structures.

Chapter 7: Conclusions present overall conclusion remarks. The contribution of this thesis to the existing literature. I also discuss the limitation of each empirical study and the potential directions for future work.

2

Literature review on housing analysis

2.1. Introduction

As previously stated, housing has attracted the attention of researchers, policy makers, planners, homeowners and investors. High-quality data on housing prices at a disaggregated level can be used to make investment or buying decisions (Jayantha and Lau, 2016). For researchers, the availability of such data can help them to build housing indices and to accurately assess non-market goods (Taylor, 2003). On the other hand, analyzing dynamics in a housing market on aggregated levels (e.g., regions, counties, etc.) provides valuable information about the linkage of a regional/sub-regional housing market to the rest of the country (Jansen *et al.*, 2011). Therefore, I intend to provide an overview of the economic literature associated with housing analysis in this chapter.

I initially point out the importance of housing for both personal finance and the national economy (Hill, 2013). Further, housing also has some special features (i.e., heterogeneity, spatial fixity and durability) that make it different from other goods (Galster, 1996). Based on these features, I introduce an

analytical framework, that is, the hedonic price method (Rosen, 1974). I discuss the application of the hedonic price method and its alternatives in the housing analysis literature. The discussion illustrates that the hedonic price method is advantageous over its alternatives. The economic cornerstone of the hedonic price model is also reviewed.

More importantly, I highlight that the location is an important determinant of housing prices, as well as regional second home markets in the literature. In the early stages, most scholars considered the location of a dwelling as the geographical position in relation to its surroundings (Marcus, 2010), and therefore accessibility was a straightforward and feasible way to measure it. The meaning of locations was further developed. Beyond the geographical position, the location is considered the source of spillovers. In the early 90s, Can (1992) demonstrated that for houses, locations were represented not only by accessibility, but also by adjacent effects. Regarding the regional market, Meen (1999) proposed using so-called ripple effects to describe the comovement among certain regions that were spatially close to each other.

Once all of these have been examined, I hope that readers can understand why we use the hedonic price method, the importance of the location for both housing and sub-regions.

2.2. Housing and wealthy

Housing markets are an integral part of an economy. On a global scale, Syz (2008) stated that almost one third of global total wealth (\$21.6 trillion approximately) was tied up in residential housing. Real estate also forms a dominant component of the total wealth of many countries and regions. Liu, Park, *et al.* (2002) clearly showed that the entire real estate sector accounted for 30% of the gross domestic product (GDP) growth of China in 2001, and local governments gained considerable revenue from real estate (Liu, 2018). Wagenmakers *et al.* (2008) demonstrated that in 2006, the real estate value in Sweden was nearly three times higher than its GDP.

Furthermore, housing is a major component of wealth for most households. According to a report from European Central Bank (2003), the esti-

mates of households' housing wealth¹ in the euro area augmented from €3.7 trillion in 1980 to €19.6 trillion in 2003. In Australia, land and housing assets accounted for 55% of household assets and 70% of household net worth in 2015. For several OECD countries (the US, UK, Canada and Germany), the ratio of housing prices relative to the disposable income of households kept rising over the past two decades. In France, this rise was around 60%, and housing capital measured by housing prices experienced a surge in the 1990s (Bonnet *et al.*, 2014).

Not only housing wealth in general, but the associated housing prices have received attention from economists as well. Many economists believe that the linkage between housing markets and the rest of the economy is mainly through house price fluctuations. Case, Quigley, and Shiller (2005) indicated that changes in housing prices had a significant impact on household consumption. Case and Quigley (2008) noted that the impact of the downturn in housing prices might be reflected via income effects. Moreover, the global financial crisis in 2008 was an obvious example, since the collapse of the U.S. subprime mortgage market followed by the fall in housing prices almost triggered that crisis (Shiller, 2007).

From a microeconomics perspective, much of the housing wealth is spread across households. Households act not only as homeowners but also as consumers and investors. For example, among homebuyers, housing market prices are usually several times the household's annual income, and a house is typically the largest single item purchased by a household. The purchase of a house possibly represents the largest part of household savings (Goodman and Muth, 2013). If house prices go up, homebuyers may wait and see. When homebuyers become homeowners, the house is the major asset in their portfolio. In this context, if house prices increase, they will become better off. Some homeowners will borrow more to spend on other goods and services. Conversely, if house prices go down, homeowners risk that their house will be worth less than mortgages. As such, they will probably cut down on spending and will hold off on investments. In short, housing prices have a considerable impact on households' consumption, investment and savings (Case, Quigley,

¹"Households' housing wealth is defined as the value at current market prices of all residential dwellings, including the value of land on which the buildings are built, owned by households." (European Central Bank, 2003)[P. 47]

and Shiller, 2005).

In addition to the above-mentioned channels, other channels include: housing price fluctuations may impact personal savings ratio (Green and Hadjimatheou, 1990); housing price fluctuations are likely to affect the distribution of wealth (Giussani and Hadjimatheou, 1991); and housing price fluctuations may have impacts on labour mobility among cities or regions (Alun, 1993; Karahan and Rhee, 2019).

2.3. Specific characteristics of housing

Housing is often referred to as a shelter and therefore meets basic human needs. Moreover, in most cases, a house implies a home and is therefore important for individuals (Bourne, 1981). Bourne (1981) and Dieleman (1996b) also highlighted that housing brought psychological satisfaction such as comfort, privacy and security. Jansen *et al.* (2011) mentioned that housing also accompanied with social activities, such as talking to the neighbours family, and friends. For economists, housing is a complex, composite good, and it can act not only as a consumer good, but also as a capital good and an investment good. Even though housing has triple status, according to economists, housing is often analyzed through conventional market theory. On the other hand, housing is set apart as a special market due to some particular features involving heterogeneity, spatial fixity, and durability (Galster, 1996; Sheppard, 1999). These three features are explained in more detail below.

According to Taylor (2003), heterogeneous goods refer to the products with characteristics that are significantly different, even though the products are sold in one market. It is evident that housing is a heterogeneous good (Łaszek, 2013). Its intrinsic characteristics vary across houses, such as types, size, age, building materials, accessibilities, decoration and etc. The degree of heterogeneity also varies among housing units. Housing is a package or a bundle of many features in terms of quality and quantity. The associated value for consumers is the sum of these features weighted by the consumer's preferences.

Spatial fixity or spatially immobile features indicate that housing occupies a fixed place, and that the location is its intrinsic attribute. For homeowners,

regardless of whether they come or leave, the house remains there. More importantly, spatial fixity also generates physical and social externalities. If a household purchases a house, it obtains not only housing structural features, but also the socio-economic status of the neighbourhood and the accessibility to any desired places, such as central business districts (CBD), private goods or public facilities. These externalities provide additional utility and implicitly improve housing prices. It should be noted that no two houses can occupy the same location, no matter how similar their structural features are. As such, the location also provides evidence that housing is a heterogeneous good. In a word, the location is a crucial determinant of housing quality and household welfare.

Durability means that housing, in general, is a durable good and has a very long lifespan. Unlike single-use or short-term consumer goods, housing can serve for many years (Dieleman, 1996a). Moreover, in the market, a large share of dwellings is provided by houses or apartment buildings constructed in the past.

Combining the durability, heterogeneity, and spatial fixity contribute to differences in housing quality and housing prices across locations. Further, because of these particular features, housing can be analyzed using the hedonic price method. In the following section, I will examine the major approaches to housing analysis.

2.4. Methods of housing analysis

Housing has been extensively studied. Bourne (1981) concluded several commonly seen topics in housing studies. From the demand side perspective, topics included household location decisions, land use and neighbourhood change, environmental (non-market goods) valuation. By contrast, the topics on the supply side involve investment and capital markets, institutional behavior. More recently, Jansen *et al.* (2011) provided a summary of analytical techniques for measuring housing preferences and choices. Among all the techniques, they mentioned the hedonic price method. As an analytical method within the neoclassical economics framework, this method is familiar to economists. The hedonic price method uses information on housing

prices and corresponding features, and then provides monetary estimates of the willingness to pay for and equilibrium of alternatives (De Haan and Diewert, 2013). That is to say, the hedonic price function reveals the maximum amount that buyers would like to pay for a housing attribute. They also stated that the key idea of the hedonic price method was that price differences depended on a set of features associated with housing. This emphasized the heterogeneous feature of housing.

Malpezzi (2008) also pointed out the heterogeneous nature of housing, and therefore the application of the hedonic price method was appropriate. He stated

“The method of hedonic equations is one way that expenditures on housing can be decomposed into measurable prices and quantities, so that rents for different dwellings or for identical dwellings in different places can be predicted and compared.” (Malpezzi, 2008)[p. 68]

In practice, a hedonic price function is expressed by multiple regression of transaction prices on housing features. Regression coefficients are usually interpreted as the implicit price of the housing features.

It is natural to use the hedonic price method for assessing non-market goods², since some environmental benefits can be measured via features that are related to housing prices. For instance, if non-environmental factors are controlled, any remaining differences in price can be attributed to differences in environmental quality. Taylor (2003) stated that the hedonic method could be used for assessing non-market goods, and the assessment was based on housing transactions to determine the implicit value of a feature. By observing the difference in prices between two product varieties that have only one distinct feature, researchers can indirectly observe the monetary trade-offs that agents are willing to make in terms of the difference in that feature. Since researchers cannot directly observe the value of the feature, but infer that value from transaction information, the hedonic price method thus is an “indirect” valuation method.

A rival to the hedonic price method is the contingent valuation (Ready *et al.*, 1997). The contingent valuation has been used to study the external bene-

²For example, environmental amenities (aesthetic views or proximity to recreational sites), and environmental quality (e.g., air pollution, water pollution, or noise).

fits, such as improved air quality (Belhaj, 2003; Ligus, 2018), newly-built local railway systems (Utsunomiya, 2018) and increased green space (Chen and Qi, 2018; Latinopoulos *et al.*, 2016). As a “direct” valuation method, the contingent valuation relies on surveys, where individual respondents clearly show their preferences for a hypothetical environmental change. The aggregate individual valuations can then help to make inferences about a community-wide valuation for the landscape in question. Bateman (1994) concluded that two kinds of questions often appeared in surveys. The first one was about the respondent’s willingness to pay to conserve a landscape. That is to say, how much people are willing to pay to prevent welfare losses due to the changing landscape. Another one is about the respondent’s willingness to pay to improve landscapes. In other words, how much people are willing to pay to achieve welfare gains due to the changing landscape.

Relying on surveys, the contingent valuation can be used for measuring nonuse values³, whereas the hedonic price method can not.

Furthermore, in conducting the survey, three types of bidding formats have been used to elicit the willingness to pay for a non-market good (Cuccia, 2020). They are open-ended, iterative, and dichotomous choices.

The open-ended bidding implies that people’s willingness to pay is induced by open questions. The respondents are asked to make a decision (i.e., he/she is willing to pay for a non-market good.) in a hypothetical situation. However, this bidding format has been critiqued. For example, some situations are set to be very hypothetical and dissimilar to real-world market transactions (Hoyos and Mariel, 2010). Further, some respondents may not be sensitive to the quantity and quality of changes (e.g., landscape changes). As such, the willingness to pay is likely inaccurate (Brander and Koetse, 2011).

The iterative bidding means that interviewers directly introduce a starting bid, and then increments or decrements are given. This procedure is repeated several times until an acceptable bid is reached. This iterative bidding may be the most frequently used format in the contingent valuation literature (Mousumi *et al.*, 2007; Hoyos and Mariel, 2010).

The dichotomous-choice bidding refers to respondents indicating whether

³Mendelsohn and Olmstead (2009) noted that the difference between the use and nonuse value of the environment was whether there was any real interaction between an individual and the environment.

they are willing to pay a specific amount for a non-market good on the basis of closed questions. This bidding format is simple and straightforward, but it is affected by many factors, such as payment methods. Bateman *et al.* (1993) found that changing the payment method from donation to tax almost doubled the willingness to pay. As such, the willingness to pay may be inconsistent.

In summary, as a survey-based method, the contingent valuation has specific characteristics, like investigating nonuse values, but it requires complex survey designs, and the associated questions are based on a hypothetical situation, rather than on observed economic choices (Brander and Koetse, 2011). Moreover, this method typically required a high budget and long operating time.

In addition to acting as a valid valuation method, the hedonic price method is also a crucial tool in identifying the determinant of housing prices, and further predicting house prices. Belke and Keil (2018) summarized several commonly-used approaches to identify the determinants of housing prices in empirical studies. These approaches include: the asset pricing approach, the hedonic price method, and the analysis based on aggregate housing price indices.

The asset price approach attempts to demonstrate the relationship between housing prices and expected future discounted earnings derived from properties (Himmelberg *et al.*, 2005; Hott and Monnin, 2008; Weeken, 2004). In particular, Hott and Monnin (2008) employed a user cost model and showed that housing prices mainly depended on mortgage rates, depreciation, maintenance costs, taxes or capital gains.

The hedonic price method can also be employed to identify the determinant of housing prices. The fact is that it can measure each structural or neighbourhood feature that has contributed to the overall housing price. Empirical studies use data on individual housing transactions with corresponding prices and various characteristics. Rich literature can be found in this field, from the early applications (Can, 1992; Freeman, 1979; Li and Brown, 1980; Rothenberg *et al.*, 1991) to the recent studies (Delgado and Wences, 2019; Noh, 2019; Yang *et al.*, 2019).

The last approach refers to some econometric models that directly estimate the housing price determinants based on some form of aggregate (i.e.,

national, regional or local level) housing price indices. In contrast to the hedonic price method using individual properties, these models use national, regional or local level economic variables as determinants. For example, Sutton (2002) investigated how shocks in incomes, interest rates and equity prices affect housing prices in six OECD countries by means of a vector autoregressive (VAR) model. The results clearly showed that positive shocks to income or equity prices and negative shocks to interest rates resulted in housing price appreciation.

Thus far, I have looked into various approaches to investigating housing price determinants. It should be noted that the application of these approaches depends on specific research questions and the availability of data.

Aside from the abovementioned applications, the last application of the hedonic price method that I intend to discuss is building price indices (Fik *et al.*, 2003). In general, there are three types of housing price indices (Rappaport, 2007), the median index, the repeat-sales index and the hedonic index.

As the simplest, straightforward measure, the median index lists the temporal changes in the median housing price of a region from one period to the next. However, it experiences several drawbacks. First, this index confounds changes in prices with quality differences. As a result, the information provided by this index is very noisy, and we cannot distinguish the pure housing price change from the price change due to quality changes. A variation of this median index is the mix-adjustment (stratification) index. Even though the mix-adjustment index takes an arithmetic or geometric mean for the median index of all sub-regions, it does not overcome the same issue in constructing a median index (Gan and Hill, 2009).

The repeat-sales index is calculated using repeat-sales housing data. Attributed to Bailey *et al.* (1963), this index is then developed and promoted by Case and Shiller (1989). Nevertheless, this index also experiences some limitations. First, it drops a lot of data, for instance, the property that only sells once has to be removed to build this index. Further, dropping these observations may give rise to sample selection biases. Second, the repeat-sales index may have the same problem as the median index. If an observation was renovated or extended between two sales, the repeated sales index may also confound changes in prices and quality differences (Prasad and Richards,

2008).

The hedonic index, like the previous hedonic applications, considers housing as a bundle of attributes with their implicit prices (Goodman, 1978; Ong *et al.*, 2003). Using regression techniques, the hedonic index can account for the extreme heterogeneity of housing, as well as provides quality-adjusting measures of price changes.

In summary, the hedonic price method has been used in a number of ways in housing contexts. First, as previously stated, it can be used to measure the implicit price of certain non-market goods. Second, it is used to identify the determinants of housing prices and corresponding impacts, and also to provide general appraisals of houses. Third, it is used to build housing price indexes. In these three situations, the hedonic price method often has advantages over its alternatives. For example, it is multitasking and does not restrict data types. Considering the availability of data in the following empirical analyses and research aims, I will look into the economic underpinning of the hedonic price method in the following sections.

2.5. Hedonic Price Model (HPM)

As one of the most popular methods, the hedonic price method provides a theoretical basis for many housing studies. In short, the idea of the hedonic price method can be expressed as follows: an entire good can be reduced to its constituent attributes, and the implicit value for these attributes is then calculated based on market transaction prices. In the following subsections, I will review the evolution of the hedonic price method, followed by an examination of its economic underpinnings.

2.5.1. A brief history

It is widely accepted that Andrew Court is a pioneer in applying the hedonic price method (Goodman, 1998; Herath and Maier, 2010; Malpezzi, 2008). Working as an economist in General Motors, Court intended to develop automobile price indices in the 1930s. In his paper, he explained the reason to name the method "hedonic" and gave an example of the newly-born method. He noted:

“Utilitarianism, seeking the good in the greatest happiness of the community as a whole, is the chief hedonistic doctrine. Thus, Hedonic price comparisons are those which recognize the potential contribution of any commodity, a motor car in this instance, to the welfare and happiness of its purchasers and the community... Passenger cars serve so many diverse purposes that such a single, most important specification cannot be found. The simple method is inapplicable, but why not combine several specifications to form a single composite measure?” (Court, 1939)[p. 107]

The phrases contain some key ideas of the hedonic price method, including the heterogeneous attributes of products, the heterogeneities of consumers’ demands and the composite measurement.

Following Court’s study, Lancaster (1966) introduced a new consumer theory, which provides the underpinning of the hedonic price method. Lancaster’s new theory states that utility comes from the intrinsic properties of goods, rather than a good itself. He wrote:

1. The good, per se, does not give utility to the consumer; it possesses characteristics , and these characteristics give rise to utility.
2. In general, a good will possess more than one characteristic, and many characteristics will be shared by more than one good.
3. Goods in combination may possess characteristics different from those pertaining to the goods separately. (Lancaster, 1966)[p. 134]

In other words, consumers purchase relevant characteristics of goods rather than goods themselves, the composite goods are heterogeneous due to different “characteristics”. Thus, people purchase goods based on the number of good characteristics and per unit cost of each characteristic. Lancaster’s new consumer theory is the microeconomic foundation for analyzing utility bearing characteristics. However, a possible limitation of his theory is that it considers the demand side of the market only.

The complete theoretical foundation of the hedonic price method is laid by Rosen (1974). Presented in a seminal paper, Rosen overcame the limitation of Lancaster’s theory. Rosen’s model integrates the earlier hedonic price model into a neo-classical economics framework⁴. Additionally, the theoretical utility framework links consumers’ preferences for characteristics goods and

⁴Rosen’s theory assumes that everything happens in a perfectly competitive market.

market equilibrium prices. Following Lancaster's theory, Rosen also argued that goods are valued by characteristics. Hence, the total price is a function of the characteristic and the implicit price of a characteristic is the derivative of the total price with respect to the amount of the characteristic⁵. Further, he demonstrated the market equilibrium and concluded that each attribute has a unique implicit price in an equilibrium market, which equals Marshallian willingness to pay (WTP).

Five years later, Freeman (1979) brought a paper named "hedonic prices, property values and measuring environmental benefits: a survey of the issues". He highly praised Rosen's model. He also underlined that a household's willingness to pay for a unit of each characteristic should equal the housing attribute prices. More precisely, treating the price of a house as a dependent variable and associated housing characteristics as independent variables, researchers can apply regression techniques. The estimated coefficients are interpreted as implicit marginal prices.

Rothenberg *et al.* (1991) summarized that the hedonic price method has two advantages over other valuing methods in measuring house prices. On the one hand, decomposing different houses into many characteristics makes an adequate simplification and avoids using multi-commodity models. On the other hand, housing characteristics can be regarded as the marginal trade-offs both supplier and demanders made on markets.

2.5.2. Theoretical foundation

In Rosen's theory, the hedonic price model is built on two stages. The first-stage analysis is used for estimating the implicit price of characteristics and reflecting consumers' preference for these characteristics through the transaction prices of a differentiated product and its characteristics (Taylor, 2017). The first-stage analysis is widely used for assessing housing prices. A possible explanation is that most regional scientists are interested in the implicit price of characteristics, and it is easy to collect data for the first-stage analysis. Housing characteristics, neighborhood and amenities attributes are usually

⁵Rosen wrote:

"...a class of differentiated products is completely described by a vector of objectively measured characteristics. Observed product prices and the specific amounts of characteristics associated with each good define a set of implicit prices." (Rosen, 1974)[p. 34]

included in an econometric model.

Conversely, few researchers (Bajari and Kahn, 2005; Bishop and Timmins, 2011) discuss the second-stage analysis in terms of estimating inverse demand functions. This gap in the literature is undoubtedly due to the identification strategy in empirical analyses. Since the prices for calculating the inverse demand function are not actually observed, the estimated implicit prices in the first stage must be put into the second-stage simultaneous estimation. As such, the hedonic price method experiences an endogeneity issue in the sense that the hedonic price function simultaneously determines both the number of characteristics and the marginal price associated with them. Researchers can use instrumental variables to correct this endogeneity issue, but it is very difficult to find a truly exogenous instrument (Bishop and Timmins, 2019).

Regarding housing, as a differentiated product, each housing unit has specific characteristics, which result in non-uniform house prices in a perfectly competitive market⁶. To be more specific, let z_j denotes a bundle of n characteristics associated to the house j , $z_j = (z_1, \dots, z_n)$. Following Rosen's theory, the equilibrium price for the house is a function of all characteristics associated with the house,

$$p_j = p(z_j) = p^j(z_1, \dots, z_n) \quad (2.1)$$

where p_j denotes the market price of house j . This function is also known as the hedonic price function or the hedonic price schedule. Since buyers in the market own heterogeneous preferences and budgets, they can only affect the price of their chosen house via a selection of attributes, but they can not affect the equilibrium price in the market. For example, the buyer h seeks to maximize his/her utility by choosing a differentiated product z and expenditure of non-housing numeric goods x , subject to a budget constraint, the buyer's income y^h .

$$\begin{aligned} \max \quad & u_j^h = U^h(z_j, x) \\ \text{s.t.} \quad & p_j + x \leq y^h, \end{aligned} \quad (2.2)$$

⁶The perfectly competitive market ensures that all combinations of housing characteristics are available on the market.

where u_j^h denotes the utility of house j for the given household h . p_j denotes a buyer's housing expenditure. To find the optimal choice for buyer h , the first-order condition is applied,

$$j^*(h) = \operatorname{argmax} U^h(z_j, x) \quad (2.3)$$

$$\frac{\partial p(z_j^*, x)}{\partial z_j} = \frac{\frac{\partial U^h}{\partial z_j}}{\frac{\partial U^h}{\partial x}} \quad (2.4)$$

where the left-hand side of Eq.2.4 is the imputed price of j^{th} attribute, and the right-hand side of Eq.2.4 represents the marginal rate of substitution between the j^{th} attribute and the numeric non-housing good x .

Thus far, I have thoroughly examined the hedonic price method from the associated applications to its economic underpinnings. In the next section, I will look into another specific feature of housing, that is, the spatial fixity.

2.6. Location and housing prices

In this section, I provide an overview of the relationship between housing locations and prices from a microeconomic perspective. In particular, special attention has been paid to measuring the impact of locations on housing prices.

2.6.1. Background

Economists have focused on urban structures for a long time. In the "isolated state" model, Von Thünen (1966) assumed that there existed a monocentric market with concentric circles for agricultural land use. Farmers intend to maximize their profits, which is the market price of farming products minus transport and production costs. Consequently, activities with the highest transport costs and the highest revenues were found near the center.

In the middle 20th century, Alonso, Mills, and Muth (Alonso, 1964; Mills, 1972; Muth, 1969) extended Von Thünen's (1966) model to adapt to a modern urban context. More precisely, a so-called Central Business District (CBD) plays a key role in the Alonso-Mills-Muth monocentric model, where land val-

ues decline with distance from the central business district of a city. Further, different types of land use, land prices, population and employment relate to the distance to the central business district (Brueckner, 2007; Wheaton, 2004). Many empirical analyses using the Alonso-Mills-Muth model investigate the relationship between distances to the central business district and observed land values. Several researchers have shown that a negative gradient exists, mostly for residential land (McMillen, 1996). That is to say, for residential land, as the distance from the central business district increases, the residential land value decreases.

The simple and analytically elegant monocentric model has dominated the urban economic theory since the middle of the 20th century. As time went by, more and more researchers found that the magnitude of the gradients decreased significantly, and this was widely interpreted as urban decentralization. Economists, therefore, attempted to explain the emergence of such complex patterns. In the late 20th century, Fujita and Ogawa (1982) proposed the polycentric model, where cities/regions owned multiple centers including the so-called primary urban centers and subcenters. Arribas-Bel and Sanz-Gracia (2014) showed that some large city-regions in North America had become polycentric, but the monocentric model could still explain the spatial structure of most North American regions. Several researchers (Romero *et al.*, 2014; Lopez and Olivera, 2005) pointed out that in Europe, formerly independent smaller cities formed larger regional conurbations with polycentric structures. In a polycentric city, the negative house pricing gradients did not maintain. Waddell *et al.* (1993) studied house prices in Dallas, USA and stated “the emergence of new nodes of regional significance has created housing price gradients that far overshadow any residential gradient with respect to the CBD.” (Waddell *et al.*, 1993)[p. 15]

That is to say, the relative location of residential land is a crucial factor in creating residential gradients. If a new node emerges in a metropolitan region, the old residential gradients will be broken and new gradients will be created.

Blotevogel (1998) noted the same phenomenon in the Rhine Ruhr metropolitan region, Germany. He showed that in the polycentric metropolitan region, house price gradients are a combination of the previous patterns of formerly

independent smaller cities.

2.6.2. How is location measured in housing valuation?

Even though the corresponding economic theory changed from the monocentric model to the polycentric model (Heikkilä *et al.*, 1989; McDonald and McMillen, 1990), the location has always been the key impact factor on housing prices, and there are a great number of empirical analyses to support this point. Bowen *et al.* (2001) reviewed the theoretical and empirical studies of the hedonic price method from a spatial analysis perspective. They advocated that empirical studies applying the hedonic price method should consider spatial effects.

After reviewing empirical analyses employing different location characteristics, I conclude that location characteristics in empirical work can be classified into two broad categories, that is, relative or absolute locations (Marcus *et al.*, 2019). The notion of the absolute location is derived directly from geography. The geographical position (location) is usually described by coordinates given longitudes and latitudes. Empirical studies considering the absolute location within the hedonic price framework may date back to the early 1990s. Dubin (1992) proposed applying Kriging⁷ (Krige, 1952). She highlighted that the Kriging method could account for spatial relationships explicitly, and thus produced a spatial trend map. Clapp and his colleagues developed the so-called local regression model (Clapp, 2003; Clapp *et al.*, 2002). This model contains both a fixed covariate effects component and a nonparametric smoother. The nonparametric smoother relies on local polynomial regression, which estimates the value of locations using a function of longitude and latitude. McMillen and Redfearn (2010) developed the so-called locally weighted regression. The locally weighted regression also involves a nonparametric component for computing the value for a given location. Furthermore, the nonparametric component is allowed to interact with other housing characteristics.

Conversely, the relative location depicts the relationship between locations. Some scholars (Marcus, 2010; Rérat, 2018) held the view that the relative

⁷Kriging is also known as Gaussian process regression (Latinopoulos *et al.*, 2016), details about Gaussian process regression are found in Sect.3.5.1.

location was a subset of the absolute location. The fact is that the absolute location also reflects the relative location, but the relative location is merely used to describe the relationship between locations.

However, in the early stages, most studies account for the relative location through accessibility to specific amenities. For example, several scholars (Colwell and Guntermann, 1984; Downes and Zabel, 2002; Dhar and Ross, 2012) investigated the accessibility to schools. Others (Anderson and West, 2006; Bolitzer and Netusil, 2000; Dehring and Dunse, 2006; Yoo and Wagner, 2016) studied the accessibility to nearby parks or other open spaces.

Accessibility in these studies is measured by Euclidean distance, and acts as a continuous or categorical independent variable in a classical linear regression model. Heyman *et al.* (2018) reviewed 52 hedonic housing studies between 2000 and mid-2016, which used accessibility as a measure of the location. They concluded that accessibility was mainly measured in four ways in hedonic studies, including Euclidean distance, network distance, travel cost, and zone. The Euclidean distance refers to the straight line distance between two points, and is usually the shortest distance in space. The network distance is the distance measured by a network, such as road networks. The travel time, as its name suggests, often accounts for both traveling speed and distance. Lastly, the zone is also a way to specify accessibility.

However, externalities generated by housing imply the relative location. An economist, Can (1992) gave another explanation. Instead of using the accessibility or the spillover, he used the term “adjacent effects” and “neighborhood effects” to describe some phenomena on housing markets.

For the “neighborhood effects”,

“The common approach with respect to the measurement of neighborhood effects is the inclusion of a set of characteristics pertaining to the socioeconomic and physical make-up of the neighborhood, accessibility to urban amenities, and the level of public services.” (Can, 1992)[p. 456]

For the “adjacent effects”,

“Similarity in the prices of nearby houses can be partially explained by externalities due to shared neighborhood characteristics. In addition, however, there will be a set of absolute spatial spillover effects on a given residential structure of the physical quality as well as the uses associated with the

surrounding neighboring structures.” (Can, 1992)[p. 457]

In a word, the “adjacent effects” emphasize the role of spillover effects in a housing market, but the “neighborhood effects” do not have this meaning.

2

Regarding the adjacent effect, externalities generated by a house in a place are capitalized into the nearby house prices. As the distance from that house increases, the strength of the externality decreases. Moreover, houses in the same block, built at the same time and having similar intrinsic characteristics may reinforce such externalities. Moreover, housing buyers and sellers often use similar sales between neighbours as a reference for assessing transaction prices, which is an efficient way to reduce uncertainties due to lack of information. Further, Páez (2009) noted that in the real estate literature, the adjacent effect was widely accepted to be interpreted as the pattern of similarities observed in housing prices due to locations, after gauging housing structural attributes.

As indicated previously, in practice, the hedonic price method is often carried out through classical linear regression. If the adjacent effect appears in the housing data, the estimation technique like ordinary least square (OLS) is not appropriate (Pace and LeSage, 2004; Wilhelmsson, 2009). The fact is that an accurate OLS (i.e., unbiasedness, efficiency and consistency) should follow several assumptions, such as constant variance (homoscedasticity) and no correlations in error terms (iid). Since the adjacent effect often occurs in cross-sectional or pooled housing data, these assumptions are seldom fulfilled. If the adjacent effect is presented, but not modeled, OLS will be biased and will lead to unrealistic inference and prediction. To correct these problems, spatial analysis methods such as spatial statistics and spatial econometrics are required, since they explicitly account for the influence of space, and significantly improve model quality and prediction accuracy (Osland, 2010).

This section uncovers one of the determinants of housing, that is, its location. I also give particular attention to the so-called adjacent effect derived from the location. It is important to note that here the observation refers to transacted properties individually, which lead to a micro dataset. In the next section, I move to investigate regional/sub-regional housing markets based on aggregate data.

2.7. Location and housing analysis: a regional perspective

As previously stated, analysing the temporal behaviour of aggregate housing market fluctuations remains the interest of many economists (Belke and Keil, 2018). Further, many scholars (Kiel and Zabel, 2008; Pace and LeSage, 2004) accepted housing to be an inherently spatial phenomenon, so we need to account for the impact from space if we are to investigate aggregate housing market fluctuations. Additionally, the literature on second home analysis often overlooks these spatial and temporal patterns. To fill in this gap, I will focus on the so-called ripple effects.

2.7.1. Ripple effects

In the late 1980s, British researchers (Giussani and Hadjimatheou, 1991; Meen, 1996) noticed that regional house prices showed a distinct spatial pattern. For instance, housing price appreciation first occurred in the Greater London region, then gradually spread out to neighbouring regions, and eventually spread out to the rest of the country over time.

To depict this phenomenon, Meen (1996) used the term “ripple effects”. Further, the ripple effect occurs in a lead-lag manner. That is to say, house price appreciation was first observed in a specific region, and then transmitted to its contiguous regions and then beyond. This also implies that increased comovement in closer, rather than distant, regions is significantly observed (Meen, 1999). This phenomenon has attracted increasing attention and various techniques have been proposed, including cointegration (Kyriazakou and Panagiotidis, 2018) and spatial econometrics (Meen, 1996).

Meen (1996) employed the standard spatial econometric framework, and he believed that spatial effects, referring to both spatial autocorrelation (dependence) and spatial heterogeneity played an important role in capturing the ripple effect. Regarding spatial dependence, the British housing market is probably characterized by a series of interconnected local markets rather than a single national market. On the other hand, spatial heterogeneity shows that the British housing market may exhibit structural differences. More importantly, spatial dependence cannot properly explain the entire ripple effect.

The fact is that even though spatial heterogeneity reflects differences in the structure of British housing markets, it is represented by random coefficients exhibiting non-random spatial patterns.

He also argued that in the British case, the economic mechanism leading to the ripple effect was still not entirely clear. Possible explanations include migration, equity transfer, spatial arbitrage and spatial patterns in the determinants of house prices. Although it is hard to demonstrate these explanations, the first two are the most convincing explanations. The migration implies that if average housing prices are higher in a region than those in neighbouring regions, home buyers may be expected to migrate to the neighbouring region benefiting from low housing prices. The equity transfer refers to lagged housing prices from neighbouring regions affecting the current house prices in a region.

2.7.2. Second homes

Second homes have attracted researchers' attention for a long time (Coppock, 1977). Even though second dwellings are an integral part of the overall housing market, they have conventionally been considered to be a particular sub-market, which is distinct from primary dwellings (Barke, 2008). As such, topics on second dwellings are dominated by investigating the internal feature of the second home sub-market. For example, the socio-economic profile of the second home owner, the seasonal migration or tourists relating to the second homes, the conflicts between second home buyers and local inhabitants and the locational feature of second homes (Paris, 2009).

Much rarer is the consideration of links between second homes and the overall housing market. For example, the fluctuation of second home rates in certain sub-regions and subsequent ramifications for the rest of the housing market is still unknown in the regional housing market context.

To fill this gap in the literature, I borrow from the idea of the ripple effect. Second home rates initially rise in a sub-region, and then spread out to the rest of the region over time. The upswing in second home rates is probably due to certain landscapes. Furthermore, the ripple effect in terms of second home rates may also imply the migration or the equity transfer. For example, second home buyers seek out certain landscapes, like beaches in a county, and then

the second home rate in the county goes up. As second home supply is limited in a short period, price appreciation occurs. This price appreciation could drive some second home buyers to surrounding counties with fairly low housing prices. Consequently, the second home rates in neighbouring counties go up. In addition, second home spillovers from one county to another may coincide with the expectation of housing sellers in surrounding counties. In the context of the upswing demand for surrounding counties, property developers would like to build new houses for second home buyers. Further, second home buyers may also choose second-hand houses. This, in turn, accelerates the second home rate among the surrounding counties.

The existence of the ripple effect will result in some practical issues (e.g., spatial autocorrelation) if classical linear regression is applied. In this context, spatial econometrics or spatial statistics could be applied, as it can gauge the spatial effects properly.

The last two sections uncover the role of locations for both individual housing and regional housing markets. I highlight the existence and importance of adjacent effects for properties and ripple effects for sub-regional second home markets. These two effects are translated into spatial autocorrelation or spatial heterogeneity in the residuals of classical linear regression. As a result, spatial analysis is required, and its associated notions and techniques will be reviewed in the next chapter.

2.8. Summary

This chapter is divided into three parts. First, I indicate the considerable impact on individuals and countries from housing, and the specific characteristics of housing. Second, I thoroughly review the major approach (the hedonic price method) that is used to analyze housing from different angles. Third, I examine how the location affects housing and regional housing markets.

More precisely, this chapter starts with an emphasis on the role of housing. I highlight that housing is important for both the national economy and personal finances. Further, I clearly show that as a good, housing has some special features that most goods do not have. These features include heterogeneity, spatial fixity and durability. Based on these features, several re-

search topics associated with the hedonic price method are presented, such as assessing non-market goods, identifying price determinants and predicting prices, and building price indices. I then thoroughly examine the theoretical underpinnings of the hedonic price method. Regarding the spatial immobile feature of housing, it relates to some economic notions, such as adjacent effects. From a regional perspective, I review the so-called ripple effects, and I look into regional second home markets. I also suggest borrowing the idea from the ripple effects to describe the comovement of second home rates within a region. Additionally, I briefly indicate that the adjacent effects and the ripple effects often give rise to series correlation (spatial autocorrelation) and heteroskedasticity (spatial heterogeneity) in the residuals of classical linear regression, and thus leads to the failure of OLS. To overcome these issues, scholars prefer to apply spatial econometrics, which gauges the influence from space explicitly. All these statistical tools will be examined in the next chapter.

In summary, the literature review chapter examines the role of housing in the economy, the cornerstone hedonic price method, and the location that acts as a crucial determinant for housing and regional markets. All of these contribute to the theoretical basis of the following empirical analyses.

3

Quantitative models for housing analysis

3.1. Introduction

Spatial modeling and the Bayesian paradigm are used throughout this thesis. In this chapter, I initially review some notions and basic models in spatial econometrics, since these methods are often considered a classical way to handle spatial data. On the other hand, I offer an overview of some specific models (i.e. latent Gaussian models) in spatial statistics and the key concepts in the Bayesian paradigm.

This chapter is divided into two parts. In the first part, I give an overview of spatial econometrics, such as spatial lag models and spatial error models. More importantly, I point out some drawbacks of these models. In the second part, I focus on spatial statistical models, which play an alternative role to spatial econometric models. I start with an introduction to spatial data types. Then, I investigate the spatial models corresponding to the spatial data types. Lastly, I give a detailed examination of the integrated nested Laplace approximation (INLA), which is a deterministic approximation for Bayesian inference. Following this organization (data, model and estimation), I hope that readers

can understand these data-driven models and INLA-based Bayesian inference, since these methods provide the methodological cornerstone of the following empirical analysis.

3.2. Spatial econometrics and housing analysis

In the last chapter, I mention that the spatial dependence of housing characteristics, prices and sub-regional second home rates make the bias towards OLS almost inevitable. Ignoring these would result in severe statistical errors, and thus spatial analysis methods would be required. Spatial econometrics¹ explicitly gauges the influence of space in houses, urban areas and regions (Pace and LeSage, 2004). The key concepts in spatial analysis are spatial autocorrelation and spatial heterogeneity (Anselin, 1988).

3.2.1. Spatial autocorrelation and spatial heterogeneity

According to Anselin and Lozano-Gracia (2009), spatial autocorrelation is a subset of cross-sectional dependence, where the structure of the covariance between observations is related to their location termed as spatial ordering. The spatial ordering is defined by the spatial arrangement of the observations including relative positioning or relative distance. Spatial autocorrelation is derived from Tobler's First Law of Geography whose statement is "everything is related to everything else, but near things are more related than distant things" (Tobler, 1970).

For housing, the most straightforward reflection of spatial autocorrelation is the dependence on housing prices, whose strength depends on the relative location between properties, and may diminish as the distance between the properties increases. The auto-correlated housing prices, in turn, provide evidence for the need of implementing spatial modelling. Other reasons to implement spatial analysis include omitting spatially correlated variables and misspecifying functional forms (Anselin and Lozano-Gracia, 2009; Wilhelmsson, 2009).

Another key concept in spatial analysis is spatial heterogeneity. Anselin

¹According to Anselin and Lozano-Gracia (2009), spatial econometrics is defined as "a sub-set of econometric methods that is concerned with spatial aspects present in cross-sectional and space-time observations."

and Lozano-Gracia (2009) noted:

“Spatial heterogeneity is a special instance of structural instability, which can be observed or unobserved. The spatial aspect of this issue is that spatial structure provides the basis for the specification of the heterogeneity. This may inform models for spatial structural change (referred to as spatial regimes), heteroskedasticity, or spatially varying and random coefficients.” (Anselin and Lozano-Gracia, 2009)[p.1227-p.1228]

The housing submarket is a reflection of spatial heterogeneity. The existence of housing submarkets may result from spatial arbitrage (Grigsby, 1963). For example, ideally, houses within an area can be substituted for each other. However, search costs and information constraints may impose limits on the degree of spatial substitutability. Spatial characteristics like accessibility and neighbourhood characteristics also limit spatial substitutability (Jones *et al.*, 2004). These elements eventually induce some barriers that segment the entire area. If spatial submarkets persist, structural differences should exist between local markets, which are reflected by the spatial heterogeneity of implicit prices (Bourassa, Hoesli, and Peng, 2003). Geographically weighted regression (GWR) (Brunsdon *et al.*, 1998) is widely used to gauge spatial heterogeneity. Other approaches include classical linear regression combined with submarket-specific dummy variables (Goodman and Thibodeau, 2003), multilevel modelling (Orford, 2000) and finite mixture models (Belasco *et al.*, 2012).

Several scholars believed that a mixture of spatial autocorrelation and spatial heterogeneity was expected to be present in cross-sectional housing data. De Graaff *et al.* (2001) argued that scholars should tackle both spatial autocorrelation and heterogeneity simultaneously for three reasons. First, there might be no significant difference between spatial heterogeneity and autocorrelation in some geo-referenced datasets. Second, spatial autocorrelation might lead to a particular form of heteroskedasticity. Lastly, in some empirical studies, it was difficult to distinguish spatial autocorrelation from spatial heterogeneity. More recently, Anselin (2010) argued that both spatial autocorrelation and spatial heterogeneity occurred simultaneously, and it was difficult to distinguish spatial heterogeneity from spatial autocorrelation in most cross-sectional settings.

3.2.2. Basic spatial econometric models

Promoted by Anselin (1988), classical spatial econometric models have two basic specifications, the spatial lag model and the spatial error model.

The spatial lag model² explicitly incorporates a lagged dependent variable on the right-hand side. This model is also known as a direct approach (Fahrmeir and Kneib, 2011), in the sense that spatial autocorrelations are assumed to be gauged via a part of the response:

$$y = \rho W y + X \beta + \epsilon \quad (3.1)$$

where y is the dependent variable. W is the $n \times n$ spatial weights matrix, and ρ is the spatial autoregressive coefficient. X is an $n \times k$ matrix of independent variables, and β is a $k \times 1$ vector of coefficients. ϵ is the error term.

The spatial weights matrix W is a simplification of connection between an object and its neighbours. Considering a contiguity-based spatial weight matrix, the element in the matrix is assigned to 0 or 1 following the rule below:

$$w_{ij} = \begin{cases} 1 & \text{sharing boundary or edge} \\ 0 & \text{otherwise} \end{cases} \quad (3.2)$$

where if observation i and j are neighbours, the element w_{ij} , which shows the interaction between i and j , will equal 1 and 0 otherwise. Conventionally, diagonal elements are set to 0 since an object can not be its own neighbours. Row-standardization is often implemented on W in spatial econometrics. The fact is that on one hand, the lagged dependent variable $W y$ represents a weighted average of the neighbouring values, given the row-standardized W . On the other hand, row-standardization ensures the largest and smallest eigenvalues, which facilitate maximum likelihood estimation or draws of the autocorrelation coefficient ρ (LeSage and Pace, 2009). The spatial autoregressive coefficient ρ measures the strength of spatial spillovers. Even though many hedonic price applications choose the spatial lag specification, Anselin and Lozano-Gracia (2009) argued that spatial spillovers among housing units were difficult to maintain. Therefore, a spatial error model may be

²Simultaneous autoregressive model on dependent variables.

appropriate, since unobserved neighbourhood effects are probably shared by housing units in the same blocks and then result in non-spherical residuals. By assuming a spatial autoregressive process with coefficient λ in residuals, we have

$$\begin{aligned} y &= X\beta + \epsilon \\ \epsilon &= \lambda W\epsilon + u \end{aligned} \tag{3.3}$$

where ϵ is the spatially correlated residual term and the term u is identically distributed (iid). The corresponding variance-covariance matrix reads

$$E[\epsilon\epsilon'] = \sigma^2 [(I - \lambda W)(I - \lambda W')]^{-1}. \tag{3.4}$$

The interpretation of λ is different from ρ in the spatial lag specification, since the aim of the term $\lambda W\epsilon$ is to mitigate biases and to produce reliable inference.

3.2.3. Some alternatives

Spatial econometric methods have been criticized, and most criticisms center on spatial weighting matrices³.

As shown previously, in a spatial econometric model, spatial effects are assumed to be captured by a predefined spatial weighting matrix (contiguity, nearest neighbours, inverse distance weighting, etc.). This also implies that the nature of the unobserved spatial effects is known a priori.

Several researchers (Corrado and Fingleton, 2012) argued that building a weighting matrix following the rule of thumb was not appropriate, and economic theories should serve as a guide to predefine that weighting matrix. Others (Brady and Irwin, 2011) stated that spatial econometric models were used in the context, in which the primary interest was to estimate the relationship of neighboring values of the dependent variable to itself. In most hedonic housing applications, however, the goal was probably not to identify that relationship, but instead accurate parameter estimates that were robust

³The spatial weighting matrix is also known as the spatial weights matrix, the spatial adjacency matrix.

to model misspecification and unobserved spatially correlated variables. In some cases, selecting an appropriate spatial weight matrix for unobserved spatial effects is not easy. For example, in a polycentric city-region, there are complex spatial effects, which cannot be captured by a simplified weighting matrix based on the contiguity or nearest neighbours.

To address the weakness of standard econometric models, several scholars (Brunauer *et al.*, 2010; Marcelo and Sebastian, 2018; Razen *et al.*, 2015) have proposed using geoaddivitive models for hedonic house price valuation. The reason is that geoaddivitive models do not require any assumptions about the structure of the omitted spatial effects in the housing market, and thus produce more accurate predicted values and marginal effect estimates compared with spatial econometric models.

Introduced by Kammann and Wand (2003), the geoaddivitive models family is the mixture of a generalized additive model and a geostatistical component. There are several candidate specifications for the geostatistical component, such as Gaussian random fields, 2-dimension thin-plate splines and tensor products of longitude and latitude. These semiparametric/nonparametric components are treated as a hedonic covariate.

Geniaux and Napoléone (2008) compared a geoaddivitive model with a tensor product of longitude and latitude and a geographically weighted regression model empirically. Their geoaddivitive model reads,

$$y = X\beta + f(\text{long, lat}) + \epsilon. \quad (3.5)$$

They concluded that geoaddivitive models had advantages over geographically weighted regression since the geoaddivitive model provides better fits, processes large datasets, and handle nonlinearity. McMillen (2012) demonstrated that nonparametric and semiparametric models are generally a preferable approach for more descriptive spatial analysis. To be more specific, local weighted regression models outperform spatial econometric models in the presence of misspecified weighting matrices. Von Graevenitz and Panduro (2015) presented a geoaddivitive model with a 2D thin-plate splines to gauge the spatial effects. The thin plate spline bases are a series of known polynomials of increasing complexity based on the longitude and latitude of a

property. They clearly showed that the geoaddivitive model fits better than a-spatial models and spatial fixed effects models. These geoaddivitive models can be estimated either by the frequentist approach, e.g, restricted maximum likelihood (REML)⁴ or by the Bayesian approach, e.g. Markov chain Monte Carlo (McMC) simulation.

In this section, I focus mainly on spatial econometrics models, which is the preference of economists. I initially detail the fundamental notions in spatial analysis, referring to spatial autocorrelations and spatial heterogeneity. I describe the basic spatial econometric models that can handle the spatial effects. More importantly, I list the drawbacks of spatial econometrics in handling housing data, and alternative approaches. From the next section, I will turn to spatial statistical models. This begins with an introduction of spatial data types, since different spatial data types correspond to different spatial statistical models.

3.3. Spatial data

Spatial data are data from known locations in space. They are defined as the realization of a stochastic process indexed by space,

$$Y(\mathbf{s}) \equiv \{y(\mathbf{s}), \mathbf{s} \in D \subset \mathbb{R}^2\}. \quad (3.6)$$

According to Cressie (2015), spatial data are often classified into three major types: areal data, point-referenced (geostatistical) data and spatial point patterns. In this thesis, my interest is centered around areal data and point-referenced data.

Typically, areal data are built on a given region, which is decomposed into many non-overlapped sub-regions with clear boundaries, and observations are grouped according to the sub-regions. This implies that areal data are composed of a finite number of areal units with well-defined boundaries, irrespective of their regular or irregular shapes. Furthermore, areal data do not provide exact coordinate information on observations, but the observations are assigned to areal units with a specific spatial alignment. Hence, I simply

⁴More detail on the fitting geoaddivitive models with thin-plate splines or tensor products can be found in Wood (2017) and the vignette of mgcv R-package (Wood, 2015).

identify the s distinct areas with their indices, leading to $\mathbf{s} = \{1, \dots, n\}$. The most common example is the census data. For example, a city is divided into many administrative districts due to practical or confidentiality reasons. This implies the discrete nature of areal data.

Concerning point-referenced (geostatistical) data $Y(\mathbf{s}) = \{y(s_1), \dots, y(s_n)\}$, $y(s_i)$ is a univariate random variable at location s_i , where $\mathbf{s} = \{s_1, \dots, s_n\}$ denotes the locations on the continuous domain D . Theoretically, the number of locations in D is infinite. More importantly, location information is represented by coordinates $s_i = (long_i, lat_i)$.

3

3.4. Analysis of lattice data

I start with the discussion of modeling lattice data, because the relevant models are straightforward and establish the basis for analyzing geostatistical data.

When data collected in nearby areas are more similar than those further apart due to similar contexts, spatial information can be represented by neighborhood structures, which leads to a large class of spatial autoregressive models. The two most common spatial autoregressive models are the simultaneous autoregressive and conditional autoregressive models. Promoted by Anselin (1988), the simultaneous autoregressive model has received much attention from economists, while the conditional autoregressive model (Besag *et al.*, 1991) may seem overlooked. Conditional autoregressive models, however, offer computational advantages and insight on fully conditional random effects compared to simultaneous autoregressive models within the hierarchical modeling framework (Arab *et al.*, 2017; Ver Hoef, Peterson, *et al.*, 2018). On the other hand, conditionals autoregressive models are flexible in terms of handling spatiotemporal data. For these reasons, I focus on conditional autoregressive models, and I attempt to apply these models to analyze economic data indexed over space and time.

3.4.1. Markov random fields

Most spatial models for lattice data are special cases of Markov Random Fields (MRFs) (Tyagi, 2017; Won and Gray, 2004). A Markov random field is based on conditional independence assumptions applying to a random vector. Given

my interest in spatial applications, I assume the random vector $U = (u_1, \dots, u_s)$ of spatial random effects for areas $\mathbf{s} = \{1, \dots, s\}$. The conditional independence assumption relies on neighbourhood structures, where two spatial random effects are assumed to be conditionally independent if they are not neighbours. Formally,

$$u_i \perp\!\!\!\perp U_{-\partial_i} | U, \quad (3.7)$$

where u_i is the random effect of the i^{th} area. ∂_i denotes the neighbours of the i^{th} area, and hence $U_{-\partial_i}$ means the random effects of the complement areas of the neighbours of area i . This leads to the sparse precision matrix Q of U :

$$Q_{ij} = \begin{cases} \partial_i & \text{if } i = j \\ 1 & \text{if } j \in \partial_i \\ 0 & \text{otherwise} \end{cases} . \quad (3.8)$$

More importantly, the sparseness of the precision matrix results in efficient computations (Rue and Held, 2005).

3.4.2. Conditional Autoregressive (CAR) model family

Again, one fundamental property of spatial data is spatial dependence derived from Tobler's first law of geography (Tobler, 2004; Tobler, 1970), which implies that areas tend to have similar values if they are spatially close.

Concerning the discrete characteristic of lattice data, the spatial closeness is defined by adequate neighbourhood structures (spatial adjacency/neighbour matrices), rather than the distance between two locations⁵. Relying on the predefined spatial adjacency matrix, spatial dependence can be modelled by the so-called proper conditional autoregressive (PCAR) model. More precisely, a Gaussian random variable is used for conditionally modelling the spatial dependence between a pair of area i and area j ⁶,

⁵It is possible to define the neighbourhood structure based on the distance between the centroid of areal units. However, contiguity-based neighbourhood structures are commonly used in the conditional autoregressive literature.

⁶I simplify the index of locations (s_1, \dots, s_n) by $(1, \dots, n)$ in a lattice system.

$$u_i | u_{j, i \neq j} \sim N \left(\rho \sum_{j \in \partial_i} \frac{w_{ij} u_j}{w_{i+}}, \frac{\sigma^2}{w_{i+}} \right) \quad (3.9)$$

where u_i represents an area-specific random variable. W denotes an $n \times n$ binary adjacency matrix describing a neighbourhood structure. Notably, W is firmly symmetric ($w_{ij} = w_{ji}$). Its off-diagonal entities (w_{ij}) show the proximity of units n_i and n_j . All diagonal elements remain zero $w_{ii} = 0$. w_{i+} is the row sum of the i^{th} row $w_{i+} = \sum_j w_{ij}$, and ∂_i denotes the index of neighbours of i . The conditional variance parameter $\frac{\sigma^2}{w_{i+}}$ is spatially varying.

The formula Eq.3.9 is called the full conditional distribution. Since there are n given areas, we can derive a joint distribution via the n full conditional distributions $U = (u_1, \dots, u_n)$. With the help of the Brook's lemma (Brook, 1964) and Hammersley-Clifford theorem (Hammersley and Clifford, 1971), Besag (1974) proved that the joint distribution for the random vector U is multivariate normal with a mean of 0 and a precision matrix Q ,

$$\begin{aligned} \mathbf{U}_{PCAR} &\sim MVN(0, \mathbf{Q}_{PCAR}^{-1}) \\ \mathbf{Q}_{PCAR} &= \tau(D - \rho W) \end{aligned} \quad (3.10)$$

where D denotes the $n \times n$ diagonal matrix, whose diagonal entries d_{ii} equal the number of neighbours of unit n_i and off-diagonal entries are all zero. τ is the precision parameter, $\tau = \frac{1}{\sigma^2}$. The spatial dependence parameter ρ is between 0 and 1. The term ρ shows the amount of spatial dependence. On the other hand, the term τ ensures the positive defined precision matrix. The log probability density of \mathbf{U}_{PCAR} reads,

$$\begin{aligned} \pi(\mathbf{U}_{PCAR}) &= -\frac{n}{2} \log(2\pi) + \frac{1}{2} \log(\det(\mathbf{Q}_{PCAR})) - \frac{1}{2} \mathbf{U}_{PCAR}^T \mathbf{Q}_{PCAR} \mathbf{U}_{PCAR} \\ &\propto \frac{n}{2} \log(\tau) + \frac{1}{2} \log(\det(D - \rho W)) - \frac{1}{2} \mathbf{U}_{PCAR}^T \tau(D - \rho W) \mathbf{U}_{PCAR} \end{aligned} \quad (3.11)$$

where n is the number of given areas. However, computing the determi-

nant of \mathbf{Q}_{PCAR} is expensive, usually requiring N^3 operations⁷ (Morris *et al.*, 2019).

Apart from the proper conditional autoregressive model, the intrinsic conditional autoregressive (ICAR) model remains popular in literature. In the intrinsic conditional autoregressive model, the spatial dependence parameter ρ is simply set to 1, implying strong, positive spatial autocorrelation in the data. Moreover, the intrinsic conditional autoregressive model often acts as a prior model to impose spatial dependence structures to the given areas. The full conditional specification of the intrinsic conditional autoregressive model reads,

$$u_i | u_{j, i \neq j} \sim N \left(\sum_{j \in \partial_i} \frac{w_{ij} u_j}{w_{i+}}, \frac{\sigma^2}{w_{i+}} \right) \quad (3.12)$$

where the conditional mean parameter $\sum_{j \in \partial_i} \frac{w_{ij} u_j}{w_{i+}}$ is proportional to the weighted average of the neighbouring values of area i , indicating that the conditional mean parameter is similar locally. The conditional variance parameter $\frac{\sigma^2}{w_{i+}}$ is inversely proportional to the number of the neighbours of unit i , meaning that areas with few neighbours tend to vary more strongly comparing with units with many neighbours. Thus, Both conditional mean and conditional variance parameters highlight that the intrinsic conditional autoregressive model imposes smoothing over local dependence structure, rather than global dependence structures⁸.

As previously stated, the intrinsic conditional autoregressive model is often used as a prior distribution for the area-specific variables. Two elements should affect the shrinkage of the intrinsic conditional autoregressive prior. The exogenous element is the sample size. If there is a large dataset and the likelihood dominates the posterior distribution, the intrinsic conditional autoregressive prior plays a minimal role. The intrinsic element relates to the neighbourhood structure, i.e., the number of neighbours of area i . The denominator w_{i+} in Eq.3.12 ensures that the more neighbours an area has, the more local smoothing it will receive. Considering an extreme case, where the

⁷For example, if $N = 100$, calculating $\det(\mathbf{Q}_{PCAR})$ will require 100^3 operations.

⁸The latter refers that the conditional mean of u_i would depend on the complement area sets to u_j with $j \neq i$.

complement areas are all neighbours of area i , the conditional mean equals the globally weighted average value and the conditional variance tends to be small. This is equivalent to placing a strong prior. If strong, positive spatial autocorrelation occurs, the strong prior works well, because more neighbours mean that more information can be borrowed properly. Nevertheless, if there is weak or no spatial autocorrelation, the “global” neighbours setting should be inappropriate.

3.4.3. Besag, York and Mollie (BYM) model and Leroux’s model

Thus far, we have reviewed the conditional autoregressive model family. The key idea of conditional autoregressive models is to impose a spatial dependence structure on the area-specific random variables. However, in the presence of weak spatial dependence, the intrinsic conditional autoregressive model probably produces over-smoothing issues. Moreover, there are some elements of randomness in terms of similarity, e.g., gentrification or spatial mismatches in regional science. To mitigate spatial over-smoothing and the impact of spatial outliers, an additional area-specific random vector $\mathbf{V} = \{v_1, \dots, v_n\}$ with an exchangeable prior $MVN(0, \sigma_v^2)$ is introduced. \mathbf{V} is known as the spatially-unstructured component in the literature. This results in the so-called Besag-York-Mollie (BYM) model (Besag *et al.*, 1991),

$$\mathbf{U}_{ICAR} + \mathbf{V}. \quad (3.13)$$

The two additive random effect components in the BYM model are assumed to be independent. Again, the spatially-structured component is developed for spatial dependence, whereas the spatially-unstructured component captures spatial heterogeneity (Harris, 2019).

Some researchers (Riebler *et al.*, 2016) argued that the spatially structured and unstructured components in the BYM model could not be seen independently from each other. The separation of the two components should lead to identifiability issues and make prior definitions challenging. Alternatively, Leroux *et al.* (2000) proposed the specification (LCAR), where the spatially structured component in the BYM model is replaced by a mixing parameter ρ ,

which shows the compromise between unstructured and structured components. Its full conditional distribution reads,

$$\begin{aligned} E(u_i | u_{j \neq i}) &= \frac{\rho}{1 - \rho + \rho w_{ii}} \sum_{j \sim i} u_j \\ \text{Var}(u_i | u_{j \neq i}) &= \frac{\sigma_u^2}{1 - \rho + \rho w_{ii}} \\ \rho &\in (0, 1), \end{aligned} \quad (3.14)$$

which leads to the unique joint distribution,

$$\begin{aligned} U_{LCAR} &\sim \text{MVN}(0, \mathbf{Q}_{PCAR}^{-1}) \\ \mathbf{Q}_{PCAR} &= \tau [\rho(D - W) + (1 - \rho)I]. \end{aligned} \quad (3.15)$$

If $\rho = 1$, Leroux's model becomes the intrinsic autoregressive model, whereas Leroux's model turns to an independent and identically distributed component if $\rho = 0$.

Thus far, I have introduced several major models for analyzing lattice data. In subsequent sections, I will move to a model designed specifically for geostatistical data.

3.5. Analysis of geostatistical data

As explained earlier, geostatistic data are measurements of a continuously spatial process, which are collected at particular sites. Although I have a finite number of observations, I attempt to estimate the characteristics of the entire spatial process e.g., the mean and variance of the process over the study region. These characteristics are useful for predicting the process at unobserved locations and building a continuous surface.

A typical example is the spatial distribution of housing prices in a region. In this case, point-referenced prices data are represented as the partial realization of a spatial process,

$$Y(\mathbf{s}) \equiv \{y(\mathbf{s}), \mathbf{s} \in D \subset \mathbb{R}^2\}, \quad (3.16)$$

where $Y(\mathbf{s})$ is a collection of housing prices at the location set \mathbf{s} , and D denotes the study region. Such a stochastic process is often assumed to follow a Gaussian distribution, and thus $Y(\mathbf{s})$ is a Gaussian Process (GP) (Cressie, 2015) (a special case of Gaussian random fields (GRFs)), where each collection of housing prices follows a multivariate normal distribution.

3

3.5.1. Gaussian process

A Gaussian process is often assumed to fulfill two important properties: stationarity and isotropy. A Gaussian process $Y(\cdot)$ is assumed to be strictly stationary under the following conditions:

$$E[Y(\mathbf{s})] = \mu, \forall \mathbf{s} \in D, \quad (3.17)$$

meaning that the Gaussian process has a constant mean. In addition, the covariance matrix of the Gaussian process depends on the difference between locations only,

$$\text{Cov}(Y(\mathbf{s}), Y(\mathbf{s} + \mathbf{h})) = C(\mathbf{h}), \forall \mathbf{s} \in D, \forall \mathbf{h} \in \mathbb{R}^2 \quad (3.18)$$

where \mathbf{h} is a shift in space. Furthermore, the Gaussian process $Y(\cdot)$ is assumed to be isotropic if the covariance between two observations only depends on the distance between the two locations regardless of directions. Otherwise, it is called anisotropic.

3.5.2. Matérn covariance function

Delving into the covariance matrix more deeply, the dependence structure among observations is captured by the covariance matrix. There are several specifications of covariance functions, but a commonly used covariance function is the Matérn family (Gelfand *et al.*, 2010),

$$\text{Cov}(y(s_i), y(s_j)) = \frac{\sigma^2}{2^{\nu-1}\Gamma(\nu)} (\kappa \|s_i - s_j\|)^{\nu} K_{\nu}(\kappa \|s_i - s_j\|). \quad (3.19)$$

Here, σ^2 denotes the marginal variance of the process. $\Gamma(\nu)$ is the gamma function. $\|s_i - s_j\|$ is the Euclidean distance between locations s_i and s_j .

K_ν is the modified Bessel function of the second kind, and the order parameter $\nu > 0$. For the special case of $\nu = \frac{1}{2}$, the Matérn covariance function is equivalent to the exponential covariance function ($\text{Cov}(Y(s_i), Y(s_j)) = \sigma^2 \exp(-\kappa ||s_i - s_j||)$). The non-negative parameter κ has the relationship: $\rho = \sqrt{8\nu}/\kappa$, indicating the distance at which the correlation between two locations declines to 0.1 (Simpson *et al.*, 2012).

3.5.3. Stochastic partial differential equation approach

To model geostatistical data, the spatial effect component $f_{geo}(s)$ in latent Gaussian models is specified by a spatially continuous variable underlying the observations and the variable follows a Gaussian random field.

However, fitting the Gaussian random field with a full covariance matrix in latent Gaussian models is computationally expensive when sample size N is large (the so-called “big n problem” (Banerjee *et al.*, 2014)). More precisely, the Gaussian random field has a dense covariance matrix built on all locations. Performing factorization on such a large, dense matrix requires N^3 operations.

One approach to reducing computational complexity is to introduce the Markovian property. Hence, random variables at observed locations are conditionally independent. Based on the conditional independence, the Gaussian random field can be converted to a Gaussian Markov Random Field with a sparse precision matrix. The trick here is called the Stochastic Partial Differential Equation approach (SPDE) introduced by Lindgren *et al.* (2011) The point of this approach is to use a discretely indexed random process, a Gaussian Markov random field, approximating the continuously indexed Gaussian random field with the Matérn covariance function. Formally,

$$(\kappa^2 - \Delta)^{\alpha/2} \times \omega(s) = W(s) \quad (3.20)$$

$$\alpha = \nu + \frac{d}{2}, \quad \kappa > 0, \quad \nu > 0, \quad s \in \mathbb{R}^d \quad (3.21)$$

where Δ is a Laplace operator. α is an integer with a default setting of 2 in R-INLA. d indicates the dimension of data. $W(s)$ stands for the spatial Gaussian white noise process.

To implement the stochastic partial differential equation approach, Lind-

gren *et al.* (2011) propose a two-step approach. At first, the entire study area is disassembled by a large number of non-overlapped triangles⁹. Then, based on the discrete triangles, the so-called piecewise linear approach is applied.

$$\omega(s) = \sum_{l=1}^n \psi_l(s) w_l \quad (3.22)$$

where $\omega(s)$ is a Gaussian random field. $\psi_l(x)$ are called the basis functions. w_l are Gaussian-distributed weights. In fact, Eq.3.22 uses a weighted sum of basis functions to approximate a spatial process (Simpson *et al.*, 2012).

To summarize, given the discrete approximation and conditional independence, we obtain a considerable computational gain, where the computational complexity dramatically reduces to $N^{3/2}$ flops for Gaussian Markov random fields. Again, INLA takes over the following task.

In this section, I have reviewed the Gaussian process regression for geostatistical data. In particular, I have looked into the Matérn kernel (covariance) function and the fast calculation method, the SPDE approach. In the next section, these spatial models will be extended to handling spatiotemporal data.

3.6. Spatiotemporal extensions

Investigating the spatial effect as described in the previous sections seems insufficient, temporal variation is also important and interesting. For example, housing transactions naturally happen at a specific location and at a specific time point. As such, housing observations are indexed both in time and space. Due to limited modelling approaches and estimation methods, housing transaction data are collected and aggregated into different time scales, e.g., months, quarters and years. Considering a one-year housing transaction dataset, if we aggregate data over the year, we can model the spatial pattern only; but if we aggregate data over months or quarters, we may observe a temporal trend over spatial patterns. That is to say, the temporal trend is represented either by the same spatial patterns or by different spatial patterns across the entire time frame. From another angle, ignoring time dimensions

⁹The process is also called triangulation and the aggregate of the non-overlapped triangles is called a mesh.

(i.e., we aggregate data over the year) is equivalent to assuming that future observations can influence present and even past observations. This assumption does not seem to be true if the time dimension is large enough. For these reasons, it is necessary to consider the time dimension in our analyses.

The spatial process introduced in the previous sections can be easily extended to spatiotemporal processes with the inclusion of time dimensions, and the data are defined by a process indexed by space and time (Gneiting and Guttorp, 2010),

$$Y(\mathbf{s}, t) \equiv \{y(\mathbf{s}, t), (\mathbf{s}, t) \in D \subset \mathbb{R}^2 \times \mathbb{R}\}. \quad (3.23)$$

Here, \mathbf{s} has the same definition as in the pure spatial case, but we have $t = \{1, \dots, T\}$.

Again, we start with the areal data model. The simplest and most straightforward way to capture temporal variation is to directly add a temporal component γ based on the temporal indicator t , and γ_t denotes γ measured in time point t . That is to say, the spatial effect and the temporal variation are gauged separately.

There are many specifications for the temporal component (Schrödle *et al.*, 2011). In the economics literature, the temporal component is often specified by dummy variables (a kind of fixed effect), and thus estimating the effect of each time point independently from others. Some researchers believe that time dummies are mainly used for capturing temporal heterogeneity (Füss and Koller, 2016). In contrast to more restrictive approaches such as time dummies or linear parametric models (i.e., $\gamma = \beta \times t$), models with fewer restrictions or allowing for flexible shapes in the evolution curve are preferred, especially if there are long periods of time. Most of these models use a structured random effect component, and thus ensuring that adjacent periods are likely to be similar. Examples include first-order random walks (RW(1)), first-order autoregressive (AR(1)) processes (Held and Besag, 1998), and splines (Ugarte *et al.*, 2012). A random walk of the first order is defined as,

$$\gamma_t - \gamma_{t-1} \sim N(0, \sigma_\gamma^2) \quad (3.24)$$

And the first-order autoregressive reads,

$$\begin{aligned}\gamma_1 &\sim N\left(0, \frac{\sigma_\gamma^2}{1-\rho^2}\right) \\ \gamma_t &= \rho\gamma_{t-1} + \epsilon_t \\ \epsilon_t &\sim N(0, \tau_\epsilon)\end{aligned}\tag{3.25}$$

Modeling spatial and temporal effects separately has received several critiques. Several researchers (Nobre *et al.*, 2005; Sun *et al.*, 2000; Xia and Carlin, 1998) have argued that these approaches do not account for the spatiotemporal interactions. As such, Bernardinelli *et al.* (1995) developed a parametric model, which was an extension of a random coefficient model ($\gamma + \Phi = (\beta + \theta_i) \times t$). The spatiotemporal component is written as,

$$\Phi = \theta_i \times t\tag{3.26}$$

where θ_i is an area-specific coefficient. Since θ_i is spatially structured, neighbouring areas can evolve similarly over time. θ_i is often interpreted as the deviation from the global trend $\beta \times t$. $\theta_i < 0$ implies that some areas evolve slower than the global trend, while $\theta_i > 0$ implies that some areas evolve more rapidly than the global trend. This specification is simple and straightforward, but the assumption of a linear evolution in time for the area-specific component seems too restrictive and inappropriate for modeling long periods.

To overcome these issues, Held (2000) established non-parametric models on the basis of the interaction between different spatial and temporal random effects. Applying the Kronecker product on the covariance matrices of different spatial and temporal components¹⁰, Held obtains four possible specifications.

Type I interaction is the product of a spatially unstructured component (i.e., the iid term in the BYM model) and a temporally unstructured component (i.e., a temporal iid term that captures temporal heterogeneity), 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 global spatial and temporal effects. It also represents all

¹⁰Here we present a generalization by using abstract types of interaction, but not restricted to any specific form of spatial and temporal effects.

kinds of area-specific but non-persistent factors that can give rise to a slight increase or decrease in the dependent variable.

Type II interaction combines a spatially unstructured component with a temporally structured term (i.e., the random walk of the first order). To be more specific, this type of interaction assumes that once the global temporal trend is fixed, each areal unit still follows a random walk temporally, differing from neighbouring units. This does not mean that each area has a temporal evolution that differs from its neighbours, since they may share a global temporal effect.

Type III interaction is composed of a spatially structured component (i.e., the ICAR term in the BYM model) and a temporally unstructured component. As such, there is spatial autocorrelation among units for each period, but no temporal correlation between adjacent periods. This implies that each unit may have a slight deviation from the global spatial trend and that deviation tends to be spatially correlated within the same period. In practice, this interaction shows the context, in which unobserved areal factors affect an area and its neighbouring areas, but the impact does not persist over time.

Lastly, the Type IV interaction is based on the Kronecker product between a spatially structured component and a temporally structured component. This type of interaction describes that the departures from the global spatial and temporal trends are likely to be correlated with their neighbours both in space and time.

Regarding the application of the abovementioned temporal component and spatiotemporal components, the simplest approach is to use time dummy variables or a parametric function of time. Both approaches can be replaced by a more flexible random effect component such as a first-order random walk or autoregressive process, or a spline. These approaches, however, assume that the temporal parameters are spatially independent. Considering the context in which neighbouring areas may have a similar temporal evolution, space-time interaction terms are introduced. The complexity increases from Bernardinelli *et al.*'s (1995) parametric model to Held's (2000) interaction model based on random effects.

On the other hand, we need to specify a spatiotemporal covariance function (Finkenstadt *et al.*, 2006) for spatiotemporal geostatistical data. Scholars

(Cressie and Huang, 1999; Gneiting, 2002) have proposed several valid non-separable spatiotemporal covariance functions. A commonly-seen spatiotemporal covariance function reads,

$$\text{Cov}(\omega(s_i, t), \omega(s_j, t')) = \sigma_\omega^2 C_\theta(h, l) \quad \forall i \neq j, t \neq t' \quad (3.27)$$

where σ_ω^2 is the constant variance of the spatiotemporal process. C_θ is the spatiotemporal correlation function parameterized by θ . $h = \|s_i - s_j\|$ is the Euclidean distance between location i and j . $l = \|t - t'\|$ denotes the temporal lag between time knot t and t' . In practice, the estimation of a model employing this spatiotemporal covariance function is computationally expensive. To reduce the computational complexity, researchers often simplify the correlation function by assuming separability, where $C_\theta(\cdot)$ consists of a purely spatial component $C_S(h)$ and a purely temporal component $C_T(l)$. Through the Kronecker product, we have $C_\theta(h, l) = C_S(h) \otimes C_T(l)$ (Gneiting, 2002).

Another approach is to assume the constant spatial correlation in time when $t = t'$, and is zero otherwise (Harvill, 2010). This assumption induces the following correlation function,

$$C_\theta(h, l) = \begin{cases} 0 & \text{if } t \neq t' \\ C_\theta(h) & \text{if } t = t' \end{cases} \quad (3.28)$$

In this case, the temporal evolution can also be considered by assuming the spatial process evolved in time through an autoregressive dynamics, i.e., the first-order autoregressive process. This is done by applying the Kronecker product of the above-mentioned spatial correlation function and the temporal correlation function of the first-order autoregressive process (Sahu, Yip, *et al.*, 2009; Sahu and Bakar, 2012).

Thus far, I have reviewed spatial data types, the corresponding spatial models and their spatiotemporal extensions. In the following section, I will look into latent Gaussian models, in which spatial models can be embedded.

3.7. Structured additive regression models

Structured Additive Regression (STAR) (Fahrmeir, Kneib, *et al.*, 2013) is a unified and general framework covering a series of regression models, including

generalized additive models (GAM) (Hastie and Tibshirani, 1987), generalized additive mixed models (GAMM) (Lin and Zhang, 1999), geoadditive models (Kammann and Wand, 2003) and varying coefficient models (VCM) (Hastie and Tibshirani, 1993). Structured Additive Regression models incorporate flexibly structured additive predictors. For example, the response variable y is assumed to follow a Gaussian distribution, where the mean $E(y)$ is linked to structured additive predictors η through an identity link function, so that

$$\begin{aligned}
 y &= E(y) \\
 E(y) &= \eta \\
 \eta &= f_1(v_1) + \dots + f_q(v_q) + \beta_0 + \beta_1 x_1 + \dots + \beta_k x_k + \epsilon
 \end{aligned}
 \tag{3.29}$$

where x_1, \dots, x_k are standard covariates with linear effects β_0, \dots, β_k . v_1, \dots, v_q are nonlinear modelled covariates, and f_j are different smooth functions capturing various types of effects:

$f_1(v_1) = f_1(z_1)$	$v_1 = z_1$	nonlinear effect of z_1 ;
$f_2(v_2) = f_{geo}(s)$	$v_2 = s$	spatial effects of location variable s ;
$f_3(v_3) = f_{long,lat}(s_{long}, s_{lat})$	$v_3 = (s_{long}, s_{lat})$	two-dimensional surfaces s_{long} with s_{lat} ;
$f_4(v_4) = \gamma_i$	$v_4 = i$	individual-specific random intercept;
$f_5(v_5) = \gamma_i u$	$v_5 = (u, i)$	individual-specific random slope of u .

3.7.1. Latent Gaussian models

According to Rue, Martino, *et al.* (2009), Latent Gaussian Models (LGMs) are a subset of structured additive regression models within the Bayesian framework. For example, a simplified Eq.3.29 is written as,

$$\eta_i = \alpha + \sum_j \beta_j z_{ij} + \sum_q w_q f^q v_{iq}
 \tag{3.30}$$

where α is an intercept. $\sum_j \beta_j z_{ij}$ are linear, fixed covariate effects with the known covariate z_{ij} and the coefficient β_j . $\sum_q w_q f^q v_{iq}$ is used for modelling nonlinear, random effects, where w_k is a vector of known weights, f^q are basis functions or smooth functions and v_{iq} are the covariates.

The key feature of a latent Gaussian model is that a Gaussian prior is as-

signed to all latent variables, which generates a latent Gaussian field denoting $\mathbf{x} = \{\alpha, \beta_0, \dots, \beta_k, f^1, \dots, f^q\}$. The latent Gaussian field is controlled by a vector of hyperparameters θ . Lastly, observed data \mathbf{y} are assumed to be conditionally independent given the latent Gaussian field \mathbf{x} and hyperparameters θ , and hence the univariate likelihood model describes the marginal distribution of the data. Formally, the three-level latent Gaussian model is written as:

3

$$\begin{aligned} \mathbf{y}|\mathbf{x}, \theta &\sim \prod \pi(y_i|\eta_i, \theta) && \text{likelihood} \\ \mathbf{x}|\theta &\sim \mathcal{N}(\mathbf{0}, Q^{-1}(\theta)) && \text{latent Gaussian field} \\ \theta &\sim \pi(\theta). && \text{hyperparameters} \end{aligned} \quad (3.31)$$

Here $Q(\theta)$ is the precision matrix (the inverse of the covariance matrix) of the latent Gaussian field.

3.7.2. Spatial components in latent Gaussian models

The aforementioned conditional autoregressive component (model) or the Gaussian process component (model) act as the (nonlinear) spatial random effect component, and then is easily embedded in the latent Gaussian model. As such, the spatial effects can follow the specifications: (i) $f_{geo}(i) = \mathbf{U}_{PCAR}i$; (ii) $f_{geo}(i) = \mathbf{U}_{ICAR} + \mathbf{V}$; (iii) $f_{geo}(i) = \mathbf{U}_{\text{Gaussian process}}$.

Thus, a three-level latent Gaussian model with a Gaussian distributed dependent variable \mathbf{Y} is given by:

$$\begin{aligned} \mathbf{Y} &\sim [\mathbf{Y}|\mu, \theta_Y] \quad \epsilon \sim N(0, \sigma_\epsilon^2) \\ \mu &= \eta = X\beta + \mathbf{U}_{(\cdot)} \\ \mathbf{U}_{(\cdot)} &\sim MVN(0, \mathbf{Q}_{(\cdot)}^{-1}) \\ \mathbf{Q}_{(\cdot)} &\equiv F(W, D, \rho, \kappa, \nu, \dots) \end{aligned} \quad (3.32)$$

where $[\mathbf{Y}|\mu, \theta_Y]$ is the data model. It depends on the mean parameter μ , where $E(Y) = \mu$ and other parameters θ_Y . ϵ is a measurement error. Regarding the process model $[\mu|\theta_\mu]$, η represents the structured additive predictors. X is a design matrix involving all covariates, β are unknown coefficients. $\mathbf{U}_{(\cdot)}$ denotes the latent spatial random vector with the above-mentioned conditional autoregressive specifications. F represents a function that describes the relationship between a series parameters and the precision matrix $\mathbf{Q}_{(\cdot)}$.

θ_μ and $\theta_{\mathbf{Y}}$ denotes hyperparameters appeared in the process model and data model respectively. Lastly, I can impose priors on the hyperparameter vector $\theta = [\theta_{\mathbf{Y}}, \theta_\mu]$.

Relying on the three levels, we obtain the posterior distribution of the process and parameters via the Bayesian theorem $[\mu, \theta | \mathbf{Y}] \propto [\mathbf{Y} | \mu, \theta_{\mathbf{Y}}] \times [\mu | \theta_\mu] \times [\theta]$.

3.8. Bayesian statistics

After reviewing spatial data and latent Gaussian models, our aim is to estimate the parameters in the models. There are two predominant schools in the statistic field, the frequentist and the Bayesian (Efron, 2005). They hold different points of view in terms of probability and further inferential procedures (Wagenmakers *et al.*, 2008).

Frequentists, dominating in the 20th century, own a frequency view of probability, meaning that probability is long-run frequencies, where an event occurs in an identical experiment repeated over infinite times. That is to say, the probability is a measurement of uncertainty derived from randomness in a repeated event. Further, frequentist inference is often aimed at providing point estimation, the best guess of an unknown parameter, and interval estimation, a range for unknown parameters (Samaniego, 2010). It is important to note that frequentists do not allow probability statements about the unknown parameters, therefore unknown parameters as considered as fixed quantities. For example, a 95% confidence interval for a parameter α is $[0.1, 0.2]$. It does not imply that there is a 95% probability that α is in $[0.1, 0.2]$. By contrast, in the context of repeating the same procedure to construct confidence intervals over infinite times, 95% of the cases would the true α drop into the 95% confidence interval. In a word, frequentist inference is evaluated under hypothetical repeated sampling of the data. Such a point of view receives criticisms, since it is sometimes infeasible to repeat the same procedure over infinite times for constructing the 95% confidence interval (Baker, 2016; Gelman and Carlin, 2017).

Conversely, Bayesians have a subjective view of probability (subjective probability and thereafter), where individuals give their evaluation to the odds

that an event will occur or not. Subjective probability also follows the rules of the probability calculus. In contrast to the fixed quantities of parameters in frequentist inference, both the degree of belief and unknown parameters are considered as random variables, and then updated by observed data in Bayesian inference. Therefore, the Bayesian approach is more natural than the traditional frequentist approach, since it allows tackling the uncertainty in the model and its parameters. More precisely, priori degrees of beliefs to an event (prior and thereafter) have an influence on the uncertainty of the model, and the total uncertainty can be represented by a probability distribution. In a word, the Bayesian approach starts with priori knowledge of parameters and data, and subsequently extrapolates backward to the true probability distribution of the parameters. In Bayesian statistics, this inversion process is carried out by the so-called Bayes' rule (Lambert, 2018). Finally and perhaps the most important, the three-level latent Gaussian model is a good reflection of Bayesian thinking, as it contains hyperpriors (prior distributions are imposed to hyperparameters), a latent process model (prior distributions are also imposed to the parameters of latent variables) and a likelihood (i.e. data or observations)¹¹.

3.8.1. Bayes' rule

Bayes' rule is naturally derived from conditional probability:

$$Pr(A|B) = \frac{Pr(A \cap B)}{Pr(B)} \quad (3.33)$$

where $Pr(A|B)$ corresponds to the probability that event A will occur given that event B has already been observed. The numerator of the right-hand side equation ($Pr(A \cap B)$) represents the probability that event A and B occur jointly, while the denominator ($Pr(B)$) means that the probability of observing event B only. Eq.3.33 can be rewritten as follows:

$$Pr(A \cap B) = Pr(A|B) \times Pr(B). \quad (3.34)$$

To obtain Bayes rule', we initially calculate the probability of the event B

¹¹ $\frac{[process, parameters|data]}{[process|parameters] \times [parameters]} \propto [data|process, parameters] \times$

given that A occurs,

$$Pr(B|A) = \frac{Pr(A \cap B)}{Pr(A)}. \quad (3.35)$$

Then, the numerator in Eq.3.35 is substituted by Eq.3.34, and Bayes theorem is written as,

$$Pr(B|A) = \frac{Pr(A|B) \times Pr(B)}{Pr(A)}. \quad (3.36)$$

Specifically, $Pr(B)$ is known. The probability of event A given B is used to update $Pr(B)$, so that $Pr(B|A)$ is obtained.

For example, given a model¹², Bayesian inference is achieved by applying Bayes' rule,

$$Pr(\theta|y) = \frac{Pr(y|\theta) \times Pr(\theta)}{Pr(y)} \quad (3.37)$$

where the prior distribution $Pr(\theta)$ for the parameter θ is updated by observed data $Pr(y|\theta)$ to yield a posterior distribution $Pr(\theta|y)$. Concerning on the numerator of Eq.3.37,

$$Pr(\theta|y) \propto Pr(y|\theta) \times Pr(\theta) \quad (3.38)$$

meaning that the posterior distribution $Pr(\theta)$ is proportional to the product of the likelihood of data $Pr(y|\theta)$ and prior distribution $Pr(\theta)$.

While for the denominator of Eq.3.37, we obtain

$$Pr(y) = \int Pr(y, \theta) d\theta = \int Pr(\theta) Pr(y|\theta) d\theta. \quad (3.39)$$

This equation clearly shows that $Pr(y)$ is a weighted average likelihood, where the weight is defined by the prior $Pr(\theta)$. Further, $Pr(y)$ is a number, which is conveniently ignored in estimating the parameter θ .

¹²Models are the mathematical formulation of the observed events, which are affected by model parameters.

3.8.2. Integrated nested Laplace approximations computing scheme

To fit a latent Gaussian model, the McMC method is widely applied. However, McMC may experience poor mixing or convergence issues. Moreover, for large models or big data sets, McMC takes a lot of computational resources (e.g., time and memory). Given my particular interest in spatial models, spatial/spatiotemporal dependence arises in data, and then McMC requires additional computational resources (De Smedt *et al.*, 2015; Taylor and Diggle, 2014). These drawbacks show that McMC may not be an appropriate tool for fitting spatial models.

Rue, Martino, *et al.* (2009) therefore developed a deterministic approximation approach, namely Integrated Nested Laplace Approximations (INLA) to estimate the posterior distributions of the parameters in latent Gaussian models. INLA is built on the combination of analytical approximations and efficient numerical integration. The main advantages of INLA over McMC can be concluded in the following points:

(i) INLA has significantly computational gains for large, complex models under several conditions¹³;

(ii) INLA is carried out by the mature R-INLA package (Martins *et al.*, 2013) for the R programming language (R Core Team, 2019), which provides a user-friendly interface for making Bayesian inference (Bakka *et al.*, 2018).

Having reviewed the highlights of the INLA approach, I will move on to introduce the INLA computing scheme. Following Rue, Martino, *et al.* (2009), the joint posterior distribution of the latent field \mathbf{x} and hyperparameters θ in Eq.3.31 is written as:

¹³E.g., the dimension of hyperparameter vectors requires small and the precision matrix should be sparse.

$$\pi(\mathbf{x}, \theta | \mathbf{y}) \propto \pi(\theta) \pi(\mathbf{x} | \theta) \prod_{i=1}^n \pi(y_i | x_i, \theta) \propto \pi(\theta) |\mathbf{Q}(\theta)|^{1/2} \exp \left\{ -\frac{1}{2} \mathbf{x}^T \mathbf{Q}(\theta) \mathbf{x} + \sum_{i=1}^n \log(\pi(y_i | x_i, \theta)) \right\}. \quad (3.40)$$

Notably, the size of the hyperparameter vector θ should be small (less than < 15 empirically), while the dimension of data \mathbf{y} and latent Gaussian fields \mathbf{x} can be large. The fact is that numerical integration is implemented over the θ space. Furthermore, the dependence structure of data is usually stored in the precision matrix θ . Since the size of the latent Gaussian field is fairly large ($n = 10^2 - 10^5$), Rue, Martino, *et al.* (2009) impose conditional independent (Markovian) properties. Consequently, the latent Gaussian field \mathbf{x} becomes a Gaussian Markov random field (GMRF) with a sparse precision matrix. This step offers additional computational gains, since numerical methods for sparse matrix calculation is much faster than dense matrix calculation (Martino and Riebler, 2019).

The joint posterior distribution for the latent Gaussian field and the hyperparameters has been derived. However, in practice, researchers are interested in the marginal posterior distributions of the hyperparameter $\pi(\theta_j | \mathbf{y})$ and the latent variable $\pi(x_i | \mathbf{y})$, rather than their joint posterior distribution. In doing so, integration is applied.

$$\pi(x_i | \mathbf{y}) = \int \pi(x_i | \theta, \mathbf{y}) \pi(\theta | \mathbf{y}) d\theta, \quad (3.41)$$

$$\pi(\theta_j | \mathbf{y}) = \int \pi(\theta | \mathbf{y}) d\theta_{-j} \quad (3.42)$$

where x_i is the i^{th} latent variable. θ_j is the j^{th} hyperparameter. θ_{-j} is the complement hyperparameter vector to θ_j . INLA applies a three-step procedure to achieve the integration. The first step aims to calculate $\pi(\theta | \mathbf{y})$. Because this term serves to calculate the marginal distribution of both the hyperparameter and the latent variable. Through conditional probability rules,

the relation is obtained:

$$\pi(\theta|\mathbf{y}) = \frac{\pi(\mathbf{x}, \theta|\mathbf{y})}{\pi(\mathbf{x}|\theta, \mathbf{y})} \propto \frac{\pi(\mathbf{y}|\mathbf{x}, \theta)\pi(\mathbf{x}|\theta)\pi(\theta)}{\pi(\mathbf{x}|\theta, \mathbf{y})}. \quad (3.43)$$

The numerator is easy to compute. While INLA approximates the denominator through a Gaussian approximation $\tilde{\pi}_G(\mathbf{x}|\theta, \mathbf{y})$ ¹⁴ by matching the mode of the full conditional of \mathbf{x}^* for a given θ ,

3

$$\frac{\pi(\mathbf{y}|\mathbf{x}, \theta)\pi(\mathbf{x}|\theta)\pi(\theta)}{\pi(\mathbf{x}|\theta, \mathbf{y})} \approx \frac{\pi(\mathbf{y}|\mathbf{x}, \theta)\pi(\mathbf{x}|\theta)\pi(\theta)}{\tilde{\pi}_G(\mathbf{x}|\theta, \mathbf{y})} \Big|_{\mathbf{x}=\mathbf{x}^*(\theta)} =: \tilde{\pi}(\theta|\mathbf{y}). \quad (3.44)$$

The second step aims to compute $\tilde{\pi}(x_i|\mathbf{y})$, an approximation of $\pi(x_i|\mathbf{y})$. This can be done by following the same logic for calculating $\pi(\theta|\mathbf{y})$. However, it is more complex since \mathbf{x} is high-dimensional. Rue, Martino, *et al.* (2009) offer three possible alternatives: a Gaussian approximation, a Laplace approximation and a simplified Laplace approximation. The most obvious one is the Gaussian approximation, because $\tilde{\pi}_G(\mathbf{x}|\theta, \mathbf{y})$ has already computed in the first step. This approach is very fast, but it may provide inaccurate results if the conditional distribution is skewed. The second approach is built on the Laplace approximation directly, namely the full Laplace approximation. In doing so, Rue, Martino, *et al.* (2009) firstly partition the latent field into $\mathbf{x} = [x_i, x_{-i}]$, and apply the Laplace approximation to $\pi(x_{-i}|x_i, \theta, \mathbf{y})$ around its model $\mathbf{x}_{-i}^*(x_i, \theta)$:

$$\pi(x_i|\theta, \mathbf{y}) = \frac{\pi((x_i, x_{-i})|\theta, \mathbf{y})}{\pi(x_{-i}|x_i, \theta, \mathbf{y})} \approx \frac{\pi(\mathbf{x}, \theta|\mathbf{y})}{\tilde{\pi}(x_{-i}|x_i, \theta, \mathbf{y})} \Big|_{x_{-i}=\mathbf{x}_{-i}^*(x_i, \theta)} =: \tilde{\pi}(x_i|\theta, \mathbf{y}). \quad (3.45)$$

This approach returns very precise estimates, but it is computationally expensive. The last approach is a compromise of accuracy and computational costs, namely the simplified Laplace approximation. It addresses the drawback of the Gaussian approximation by implementing a Taylor series expansion up to the third order to both the numerator and the denominator of Eq.3.45.

¹⁴Tierney and Kadane (1986) proved that this equation is equivalent to the Laplace approximation of a marginal posterior distribution.

This is also the default option in the R-INLA package.

The last step is to integrate θ in Eq.3.41 by applying numerical integration. Concerning the uncertainty of θ , it is important to find suitable integration points θ^k of the hyperparameters vector. Therefore, Rue, Martino, *et al.* (2009) provide three alternatives depending on the number of hyperparameters, i.e., the grid approach, the central composite design and the approach ignoring the hyperparameter variability. The first approach is very precise but it is appropriate for one or two hyperparameters only. The second approach is the default option of the R-INLA if there are more than two hyperparameters.

Finally, we have the following equation,

$$\tilde{\pi}(x_i|\mathbf{y}) \approx \sum_{k=1}^K \tilde{\pi}(x_i|\theta^k, \mathbf{y})\tilde{\pi}(\theta^k|\mathbf{y})\Delta_k \quad (3.46)$$

where k refers to $1, \dots, K$ integration points and Δ_k are corresponding weights. Thus, the meaning of INLA is clear in the three-step procedure. Integrated refers to applying numerical integration. Nested means that to calculate the marginal posterior distribution of latent variables, researchers have to estimate the marginal posterior distribution for the hyperparameters first. Laplace approximation indicates that researchers use Laplace approximations several times during the calculation (Blangiardo and Cameletti, 2015; Faraway *et al.*, 2018).

3.9. Model assessment

Thus far, we have known the spatial data, the geoaddivitive model family and the black-box of INLA. A number of Bayesian criteria for model assessment are also provided in the R-INLA package.

The Deviance Information Criterion (DIC) (Spiegelhalter *et al.*, 2002) is a well-known Bayesian model choice criterion for comparing complex hierarchical models. This criterion relies on the deviance,

$$D(x, \theta) = -2\log\pi(y|x, \theta) \quad (3.47)$$

where x, θ are latent effects and hyperparameters, respectively.

The Deviance Information criterion incorporate the posterior expectation of deviance ($\bar{D}()$) and an estimate the number of effective parameters (pD) of a model,

$$DIC = \bar{D}(x, \theta) + pD. \quad (3.48)$$

where $\bar{D}()$ is considered as a Bayesian measure of model fit, and defined as

$$\bar{D}(x, \theta) = E_{x, \theta|y}[D(x, \theta)] = E_{x, \theta|y}[-2 \log \pi(y|x, \theta)]. \quad (3.49)$$

Here, the better the model fits the data, the larger the log-likelihood value, but the smaller the posterior expectation of the deviance. pD works as a penalty term indicating model complexity, and is defined as the difference between the posterior expectation of the deviance and the deviance evaluated at the posterior expectation of the latent effects and hyperparameters,

$$pD = \bar{D}(x, \theta) - D(\bar{x}, \bar{\theta}), \quad (3.50)$$

and a higher pD indicates more complex models. Finally, a lower DIC value reflects a better model in terms of fitting data.

An alternative model comparison tool is the conditional predictive ordinate (CPO) (Pettit, 1990). It is a predictive measure since the calculation of CPO is depended on cross-validation. For each observation, the conditional predictive ordinate is given by

$$CPO_i = \pi(y_i|y_{-i}). \quad (3.51)$$

Here, the conditional predictive ordinate value is defined as the cross-validated predictive density $\pi(y_i|y_{-i})$ at the observation y_i , y_{-i} denotes the excluded observations. A larger value indicates a better fit of the model to the data, whereas smaller values indicate an ill-fitting model. Roos and Held (2011) suggested calculating the so-called cross-validated logarithmic score, which is the mean of all CPO values,

$$\text{LCPO} = -\frac{1}{N} \sum_{i=1}^N \log(\text{CPO}_i). \quad (3.52)$$

Lower LCPO scores indicate better models.

3.10. Summary

In this chapter, I initially offer an overview of spatial econometrics and relevant notions. I also list the drawback of these models. To overcome these drawbacks, spatial statistical models are introduced. I review different types of spatial data, areal data models, geostatistical models and their related concepts. I also look into spatiotemporal processes, and some extensions of the aforementioned spatial models for fitting the spatiotemporal process. As these spatial or spatiotemporal models can be incorporated into the hierarchical modeling framework, it is natural to estimate the parameters via the Bayesian approach. I, therefore, investigate the Bayesian framework and the integrated nested Laplace approximation technique. I also provide several model choice criteria for choosing different Bayesian models. This chapter provides the cornerstone of statistical modeling for this thesis. I will use the aforementioned models in later chapters of this thesis.

4

Time, space and hedonic prediction accuracy: evidence from Corsican apartment market

4.1. Introduction

It is widely agreed that housing locations affect housing prices. To explain this phenomenon, Can (1992) identified two effects: (1) "neighbourhood effects", and (2) "adjacent effects". The first refers to sharing a series of location-specific amenities and public goods, whereas the second refers to a sort of spillover.

In addition to the theoretical explanation, economists have attempted to construct appropriate econometric models for evaluating properties. In particular, via the hedonic price method (HPM), researchers can evaluate "neighbourhood effects" and interpret them as the marginal willingness to pay for corresponding attributes, even if these attributes cannot be observed directly

on the market. Nowadays, with the help of spatial econometrics (Anselin, 2010), economists can also estimate “adjacent effects”.

However, arguments arise since researchers may overlook the impact from a temporal dimension if data are collected over time (Dubé and Legros, 2013a). Specifically, researchers usually have no information on repeated sales in most cases. As such, housing transactions data pertain to repeated cross sections, meaning that the data are composed of the different observations from given populations following a chronological order. According to Dubé and Legros,

“The structure of real estate databases is different from conventional panels, since the same observation is not necessarily repeated. Instead, real estate data are collections of many cross-sectional data pooled over time.” (Dubé and Legros, 2013a, p. 5)

The most straightforward approach to process such data is to build a large pooled cross section and then apply pooled OLS regression with space or time dummy variables. However, this approach has an explicit limitation. It fails to capture the correlation in space and over time. As such, the estimated coefficients may be biased and prediction may be unrealistic.

Consequently, the aim of this chapter is threefold. I suggest that particular attentions should be paid to both spatial and temporal attributes of housing sales data. As such, the repeated cross sections belong to spatiotemporal point-referenced data. Secondly, a Bayesian hierarchical spatiotemporal model is initially introduced. Spatiotemporal correlation is gauged via the latent random effect component in the model. I illustrate how this model is fitted with two new techniques, namely integrated nested Laplace approximations (INLA) (Rue, Martino, *et al.*, 2009) and the stochastic partial differential equation (SPDE) approach (Lindgren *et al.*, 2011). INLA relies on direct numerical integration and is designed for latent Gaussian models. In addition, the SPDE approach makes use of Matérn covariance structures and Delaunay triangles to yield a Gaussian Markov random field (GMRF), which is the good approximation of a Gaussian random field (GRF). Lastly, a set of Bayesian hierarchical models are used for studying Corsican apartment markets. I compare the data fitting result, the prediction accuracy and the computational complexity of all candidate models, which involve spatial and temporal random components individually and jointly. Based on the result of the comparison, I finally illustrate

the most effective model for prediction.

This chapter is structured as follows. In Section 4.2, I briefly review literature, in particular, certain new methods to gauge spatial and temporal correlation in property prices. In Section 4.3, I describe Bayesian hierarchical spatiotemporal models and briefly introduce the INLA method and the SPDE approach. In Section 4.4, I detail the used dataset, the model specifications and the results of estimation and prediction in the study. The conclusions are presented in Section 4.5.

4.2. Literature review: research trends in the HPM-based property valuation

4

Half a century ago, Lancaster (1966) introduced his utility theory. He stated that goods per se did not give utilities, but the involved characteristics provided utilities. As such, the utilities of a good equalled the sum of the utilities of all characteristics. Rosen (1974) integrated Lancaster theory into a market equilibrium framework and developed the hedonic price method.

In the context of house valuation, the hedonic price method ensures that a house buyer's utility is a function of housing characteristics involved in the purchased house. Moreover, since the house buyer's utility can be expressed by the market equilibrium price of the purchased house (Deaton and Muellbauer, 1980), the house price is a function of the characteristics associated with the house. Following Malpezzi's (2008) definition, the empirical representation of a house price is given as:

$$P = f(S, N, L, C, T, \beta)$$

where P is the house price, S is the structural characteristics of the house, N represents the neighbourhood characteristics, L denotes the locational characteristics, C describes the contract conditions and T is time. β is the vector of the parameters to be estimated.

Although the hedonic price method provides the theoretical basis to investigate how housing characteristics affect prices, a crucial challenge is estimation. More specifically, this method merely provides a general analytical

framework rather than guidance on a specified case (Moreira de Aguiar *et al.*, 2014). Researchers must face issues ranging from variable selection to model specifications.

We have observed some current trends (Sheppard, 1999) in empirical studies addressing these issues, e.g., applications of semi-parametric or nonparametric methods and applications of spatial econometrics (Anselin and Lozano-Gracia, 2009).

The pioneering application of semi-parametric models on property valuation comes from Pace (1995). To address model specification problems, he suggested so-called semi-parametric index regression. He also demonstrated how this model avoided misspecification and controls spatial trends.

Clapp (2003) developed local polynomial regression belonging to semi-parametric models. The regression contained a nonlinear term based on latitudes and longitudes to calculate housing location values.

Kammann and Wand (2003) then introduced a geadditive models family, which was a mixture of additive modelling (Buja *et al.*, 1989) and a geostatistical component. There are several candidate specifications for the geostatistical component, such as a Kriging component or a smooth spatial trend component based on the tensor product of longitude and latitude.

Basile *et al.* (2013) applied a geadditive model incorporating a two-dimensional tensor product smoother for space to investigate European industrial locations. They indicated that their model outperformed other parametric models since it allowed to control unobserved spatial patterns, to reduce misspecification and to point out inward foreign direct investment clusters simultaneously. Further, it is possible to model space-time interactions via a three-dimensional tensor product smoother (Augustin *et al.*, 2009), and therefore, a model incorporating such tensor product smoother can handle repeated cross sections.

In short, the semi-parametric and geadditive models have advantages such as flexibility, handling repeated cross sections, avoiding model misspecification and of course, mitigating estimation biases.

Another trend related to widely used spatial econometrics (Anselin, 2010). Within the hedonic price framework, spatial econometric models were used for measuring location-specific amenities and spillover effects. However, applying spatial econometric models on repeated cross sections was rarely seen in the

existing literature, most empirical studies focused on cross-sectional or panel data Anselin (2010) and Shi and Lee (2018).

After exploring spatial econometrics literature, I found two outstanding applications. Dubé and Legros (2013a) said that tools for analysing point-referenced house transactions data were very limited. Pooling data over time and applying time dummy variables might lead to biased estimates, because the time dummy variables merely captured temporal variability, but neglect temporal correlation (Palmquist, 2005). Hence, they initially proposed a so-called spatiotemporal autoregressive (STAR) model. Certainly, the spatiotemporal autoregressive model is an extension of a spatial lag model, where the spatial weighting matrix in the spatial lag model is replaced by a spatiotemporal weighting matrix. The author also showed that the spatiotemporal autoregressive model performed better than the spatial lag model in their case. However, due to its specification, the spatiotemporal autoregressive model may have limited use.

Another approach called unbalanced spatial lag pseudo-panel models with nested random effects came from Baltagi *et al.* (2015). The authors were also confronted with a point-referenced house transaction dataset, but they considered the hierarchical structure of the data. As a result, the data took the structure of an unbalanced pseudo-panel. Regarding the model, the spatial correlation was captured by time-varying spatial weighting matrices and temporal variability was captured by time fixed dummy variables. Nevertheless, the temporal correlation was not considered in the model.

Notably, the above-mentioned models are estimated via frequentist approaches including maximum likelihood (ML) or restricted maximum likelihood (REML). However, regarding uncertainty and model complexity, Bayesian inference should be a better choice (Weller, 2016; Browne and Draper, 2006; Mathew *et al.*, 2015). To be more precise, model parameters are random variables defined by a probability distribution in the Bayesian approach. Hence, both information in data and prior knowledge are absorbed into the posterior distribution of the model parameter.

Weller (2016) held the view that Markov chain Monte Carlo (MCMC) methods had advantages over likelihood-based methods in fitting a linear mixed model because of considering the uncertainty of parameters of interest. Ac-

According to Browne and Draper (2006), both McMC and likelihood-based methods provided unbiased point estimates, if they fitted a two-level variance components model with a large dataset. Recent research (Mathew *et al.*, 2015; Huang *et al.*, 2017) suggested that INLA was a good alternative of the McMC and REML methods in fitting a linear mixed model with multiple random components. Moreover, Martino and Riebler (2019) showed that INLA was remarkably efficient comparing with McMC for fitting large, complex models, and it provided accurate point estimates. Besides, INLA did not experience convergence and mixing issues.

After investigating the advantages and drawbacks of the aforementioned models and estimation, I introduce the Bayesian hierarchical spatiotemporal model, which is deemed as the extension of a geoadditive model. The fact is that there is a GRF component evolving over time in the hierarchical spatiotemporal model, so that we can gauge spatiotemporal random effects and handle repeated cross sections. Additionally, model estimation is done via INLA-SPDE approach.

4

4.3. Spatiotemporal modeling for geostatistical data

4.3.1. Spatiotemporal data

Before reviewing Bayesian hierarchical spatiotemporal models, I concentrate on data types. Palmquist (2005) stated that most hedonic property valuation studies used micro-data. Dubé and Legros defined the micro-data as

“Observations that are points on a geographical projection...”. (Dubé and Legros, 2014, p. xi)

The definition implies that the geographical coordinates of properties are stored in the micro-data. From a geostatistical perspective, the micro-data pertain to geostatistical data. It is a widely held view that such data can be considered as a stochastic process indexed on a continuous plane (Arbia, 1989; Ver Hoef, 2002). Further, regarding a time dimension, the stochastic process can be indexed both in space and time (see Section 3.6).

Hence, I assume that $y(s, t)$ denotes the realization of the stochastic process, which describes the observed log-transformed transaction price of

the apartment $i = \{1, \dots, n\}$ at location s_i with instant $t = \{1, \dots, T\}$.

4.3.2. Bayesian hierarchical models with space-time random effects

The proposed Bayesian hierarchical spatiotemporal model is derived from the modified hierarchical autoregressive model (Sahu and Bakar, 2012). The key point is that the underlying spatiotemporal dependence is modelled by first-order autoregressive (AR(1)) dynamics with spatially correlated innovations. This specification is known as an AR(1) group model, where the group refers to the spatial group $\omega(s_i)$. As such, the dependence in space and the temporal dependence between spatial groups are captured. This proposed model is also formulated in a latent Gaussian model, which is written as,

$$Y \sim [Y|\mu, \theta_Y], \quad (4.1)$$

$$g(\mu) = \eta(s_i, t) = z(s_i, t)\beta + \xi(s_i, t), \quad (4.2)$$

$$\theta = \{\theta_Y, \theta_\mu\} = \{\sigma_\omega^2, a, \kappa, \sigma_\epsilon^2\}. \quad (4.3)$$

Here, $z(s_i, t)$ ¹ is the vector of covariates referring to the fixed effects. $\epsilon(s_i, t)$ is the measurement error following a Gaussian distribution. ξ is the spatiotemporal random component. ξ and β consists of the latent Gaussian field $\mathbf{x} = \{\xi, \beta\}$. More importantly, $\xi(s_i, t)$ is used for describing spatiotemporal process, which is assumed to be a Gaussian random field evolved over time. It reads,

$$\xi(s_i, t) = a\xi(s_i, t-1) + \omega(s_i, t), \quad (4.4)$$

where $\omega(s_i, t)$ is a time-independent Gaussian random field, whose spatiotemporal covariance function is described by,

$$Cov(\omega(s_i, t), \omega(s_j, t')) = \begin{cases} 0 & \text{if } t \neq t' \\ \sigma_\omega^2 C(\|s_i - s_j\|) & \text{if } t = t'. \end{cases} \quad (4.5)$$

¹ $z(s_i, t) = (z_1(s_i, t), \dots, z_p(s_i, t))$

Here $C(\|s_i - s_j\|)$ denotes the Matérn correlation function, which depends on $\|s_j - s_i\|$.² $\|s_j - s_i\|$ is the Euclidean distance between the observation i and j . Notably, the Matérn correlation function implies that the spatial process is second-stationary and isotropic (Cressie, 2015). Subsequently, the Matérn covariance function reads:

$$\sigma_\omega^2 C(\|s_i - s_j\|) = \sigma_\omega^2 \times \frac{2^{1-\nu}}{\Gamma(\nu)} \times (\kappa \times \|s_i - s_j\|)^\nu \times K_\nu(\kappa \times \|s_i - s_j\|), \quad (4.6)$$

where Γ is the gamma function. K_ν is the modified Bessel function of the second kind with order ν . Generally, the ν is the smoothness parameter and is a non-negative number³. κ is the scaling parameter and is also a non-negative number. Based on the empirically derived definition (Lindgren *et al.*, 2011), the relation among κ , ν and r is expressed as:

$$r = \frac{\sqrt{8\nu}}{\kappa},$$

where r indicates the distance where spatial correlation diminishes to 0.1.

Concerning the former part of Eq. 4.4, a is the autoregressive parameter with $|a| < 1$. More importantly, to implement the model in the R-INLA package (Martins *et al.*, 2013), we actually add a $\sqrt{1-a^2}$ term before $\omega(s_i, t)$ to ensure the stationary of the AR(1) process. Further, the time-dependent Gaussian random field $\xi(s_i, t)$ has a space-time separable covariance function, which is characterized by:

$$\Sigma_\xi = \Sigma_T \otimes \Sigma_S \quad (4.7)$$

where Σ_T is the covariance function of the temporal process and Σ_S is the Matérn covariance function of the spatial process. \otimes is the Kronecker product. Eq. 4.7 is further written as:

² $i \neq j$

³Based on Eq. 3.20 and Eq. 3.21, we have $\nu = \alpha - \frac{d}{2}$. In Eq. 4.6, d equals 2 and hence $\nu = 1$.

$$\begin{aligned}
\Sigma_{\xi} &= \text{Cov} \left(\sum_{k=0}^{\infty} \rho^k \omega(s_i, t - k), \sum_{k'=0}^{\infty} \rho^{k'} \omega(s_j, t' - k') \right) \\
&= \sum_{k, k'=0}^{\infty} \left(\rho^{k+k'} \text{Cov}(\omega(s_i, t - k), \omega(s_j, t' - k')) \right). \quad (4.8) \\
&= \sigma_{\omega}^2 \frac{a^{|t-t'|}}{1-a^2} \otimes \mathcal{C}(\|s_i - s_j\|)
\end{aligned}$$

As such, the correlation over space and time is gauged simultaneously. Lastly, I can assign hyperpriors to the hyperparameter vector θ in Eq. 4.3.

4

4.3.3. INLA-SPDE approach

Fitting the proposed Bayesian hierarchical spatiotemporal model is challenging, in particular, the “big n problem” (Banerjee *et al.*, 2014) often occurs.

Several solutions have been proposed to overcome this issue. Bakar and Sahu (2015) developed the “spTimer” package, where they employ MCMC with a low-rank approximation. A recent solution is to apply INLA coupled with the SPDE approach. Apart from the advantages such as accounting for uncertainty and low computational costs, this approach can estimate spatial range and other hyperparameters automatically. More precisely, the SPDE approach is devoted to tackling complex spatial dependence. Relying on the Matérn covariance function for the the spatial random component and subsequently introducing Markovian properties, the value at each location is conditional independent and then the dense Matérn covariance matrix is substituted by a sparse precision matrix. To conclude, the GRF is approximated by a GMRF to speed up the calculation. Currently, the difficulty becomes how to define that GMRF, which is the best substitution for the GRF, given local neighbourhoods and the sparse precision matrix. Lindgren *et al.* (2011) proposed using the SPDE approach. The numerical resolution of a stochastic partial differential equation given by the piecewise linear basis functions based on a mesh (the aggregate of a large number of non-overlapped triangles over the study area is called a mesh) can provide a good approximation to the Matérn covariance. INLA will take over the following task.

4.4. The case of Corsican apartment markets

4.4.1. Data

Our study focuses on Corsica, and we use the dataset extracted from the “PERVAL” database. This dataset is a collection of Corsican real-estates conveyances registered by real-estate lawyers. It includes all transaction information such as sale prices, point-referenced ownership, structural attributes, and locations (longitude and latitude). The raw dataset contains more than 10 000 apartments sold from 2006 to 2017. Data are then screened to remove the apartments with incomplete or questionable attributes data. Hence, the final dataset is made up of 7 634 sales of apartments spanning from 2006 to 2017. Actual sales prices were available for all apartments in the final dataset with the mean €149 467,08 after adjusting for inflation (The base year is 2006.). Further, apartment sales prices ranged from €57 445,76 to €325 431,67. Other key variables like apartment structural characteristics and accessibility variables are listed in Table 4.1 along with detailed descriptions.

For spatial prediction purposes in the following subsections, the cleaned dataset is split into a training set and a validation set. I randomly place 80% of the data in the training set for fitting the models, and the remaining 1 527 observations are assigned in the validation set for testing the models. The summary statistics of continuous variables in the training set are presented in Table 4.2.

4.4.2. Model discussion

One objective of this study is to compare the prediction performance of different models that delineate the impacts of space and time in different ways. To meet this goal, I examine four candidate models with different specifications of spatial and temporal components. Starting from a linear additive model, I then add different random components, *e.g.*, a spatial random component, spatial and temporal random components jointly and a spatiotemporal random component. The summary of the random components in the hierarchical models is displayed in Table 4.3.

Let us explore these models in more detail. The dependent variable is the transaction price of an apartment in a logarithmic scale. The application of

Table 4.1: List of hedonic variables with description

Variable	Abbreviation	Description/Unit
STRUCTURAL ATTRIBUTES		
Room	Ro	Number of rooms
Bathroom	Bat	Number of bathrooms
Parking	Par	Number of parkings
Floor	Sto	Number of floors
Living area	Surf	Square meters (m^2)
Apartment type		
Standard apartment (reference)	SA	
Duplex apartment	DU	
Studio apartment	ST	
Construction period		
1850-1913 (reference)	Contr	
1914-1947	Contr 1850-1913	
1948-1969	Contr 1914-1947	
1970-1980	Contr 1948-1969	
1981-1991	Contr 1970-1980	
1992-2000	Contr 1981-1991	
2001-2010	Contr 1992-2000	
2011-2017	Contr 2001-2010	
	Contr 2011-2017	
ACCESSIBILITY VARIABLES		
Distance to the nearest beach	Bea	kilometres (km)
Distance to the nearest health facility	Hosp	kilometres (km)
Distance to the nearest public primary school	PSch	kilometres (km)

the logarithmic transformation is not only intended to stabilize the variance of the dependent variable, but also to make the distribution of price data approximately normal.

The base model (M0) is the linear additive model (Eq. 4.9). You can find all structural variables and accessibility variables in the model, but there are not any spatial or temporal random components.

$$\begin{aligned}
 \log(\text{TransactionPrice}) = & \beta_0 + \beta_1 Ro + \beta_2 Bat + \beta_3 Par + \beta_4 Sto \\
 & + \beta_5 Typ + \beta_6 Contr + \beta_7 Bea \\
 & + \beta_8 Hosp + \beta_9 PSch \\
 & + f(\text{Surf})
 \end{aligned}
 \tag{4.9}$$

Table 4.2: Descriptive statistics about housing attributes for the training set

	Mean	St. Dev.	Min	Pctl(25)	Pctl(75)	Max
Transaction Price	149,440.60	58,789.31	57,445.76	102,921.40	185,406.50	325,431.70
log(Transaction Price)	11.84	0.39	10.96	11.54	12.13	12.69
Ro	2.68	0.97	0	2	3	8
Bat	1.05	0.26	0	1	1	3
Par	0.79	0.72	0	0	1	8
Sto	1.85	1.73	-3*	1	3	12
Surf	59.38	22.46	6	42	73	197
Bea	3.75	7.07	0.00	1.02	3.41	50.09
Hosp	10.43	12.11	0.05	1.64	16.51	72.24
PSch	1.34	1.74	0.000	0.47	1.46	39.51

N=6107; St. Dev.=standard deviation;

Pctl(25)=25% quantile; Pctl(75)=75% quantile;

*The negative number appears because there are "souplex" apartments.

Table 4.3: Summary of random components in the hierarchical models

Model Identifier	M0	M1	M2	M3
Spatial	None	$\omega(s_i)^*$	$\omega(s_i)^*$	None
Temporal	None	None	$\mu(t)$	None
Spatiotemporal	None	None	None	$\xi(s_i, t)$

* $\omega(s_i)$ is a Gaussian random field that is temporally constant.

where the living area is modelled as a nonlinear covariate regarding residuals and literature (Baltagi *et al.*, 2015; Shimizu *et al.*, 2014). Concerning the nonlinear specification $f(\cdot)$, the default first-order random walk (RW1) smoother in the R-INLA package is used.

The second model (M1) is the combination of the first model (M0) with the spatial random component ($\omega(s_i)$). This specification is similar to the geoaddivitive model proposed by Kammann and Wand (2003). In this context, I do not consider impacts from time. That is to say, the temporal dimension is collapsed to "zero thickness". The corresponding spatial covariance function reads $Cov(\omega(s_i), \omega(s_j)) = \sigma_\omega^2 C(\|s_i - s_j\|)$, and $C(\|s_i - s_j\|)$ is the aforementioned Matérn covariance function.

Based on the model M1, the third model (M2) involves the additional temporal random component $\mu(t)$. Therefore, spatial and temporal effects are investigated jointly in this model. Furthermore, the purely spatial ($\omega(s_i)$) and the purely temporal ($\mu(t)$) random component are uncorrelated. The purely temporal component is specified via an AR(1) process on the ordinal quarters.

$$\begin{aligned}\mu(1) &= N\left(0, (\tau_{ar1} (1 - \rho^2))^{-1}\right), \quad t = 1 \\ \mu(t) &= \rho \times \mu(t - 1) + \varepsilon(t), \quad \varepsilon(t) \sim N\left(0, \tau_{ar1}^{-1}\right)\end{aligned}\quad (4.10)$$

where ρ is the autoregressive parameter and ε is the measurement error with the precision parameter τ_{ar1} (The corresponding variance parameter is σ_{ar1}^2). In the Corsican case, there are 48 quarters spanning from the first quarter of 2006 to the fourth quarter of 2017. As a result, the spatiotemporal covariance function consists of the purely spatial and temporal covariance functions, which reads,

$$\text{Cov}(\mu(t), \mu(t')) + \text{Cov}(\omega(s_i), \omega(s_j)) = \rho^{|t-t'|} \frac{\sigma_{ar1}^2}{1 - \rho^2} + \sigma_{\omega}^2 C(\|s_i - s_j\|)\quad (4.11)$$

Lastly, M3 is the hierarchical spatiotemporal model introduced in Subsection 4.3.2. This can be considered as an improvement over M2 since the temporally-constant Gaussian random field together with the AR(1) temporal process is replaced by a temporally-dynamic Gaussian random field only.

Regarding the computational complexity of each model, M2 incorporates the purely temporal AR(1) process (see Eq. 4.10), which is assumed to be constant in space. That is to say, the temporal latent process is built on the temporal knots and is unidimensional. However, M3 incorporates an AR(1) structure with spatially correlated but temporally independent innovations (see Eq. 4.4). This latent process is defined by the number of spatial observations. From a computational perspective, M3 requires more computational resources than M2, even if M2 has an additional parameter (σ_{ar1}^2) to estimate.

4.4.3. Implement details

Before training all candidate models, I highlight two key points, the used mesh and prior distributions. I construct a convex hull mesh, as shown in Figure 4.1. Considering distances among housing locations and Corsican city sizes, I set the maximum triangle length to 10 km in the interior domain, while the length is specified to 50 km in the outer extension to avoid boundary effects⁴. To reach a compromise between dense housing locations in urban areas and too many tiny triangles, minimum distances among points are specified to 0.1 km. Lastly, I obtain the mesh containing 3 060 vertices of triangles.

One advantage of Bayesian modelling is the inclusion of prior knowledge. I select penalized-complexity priors (PC priors) for all hyperparameters. The list of prior distributions is shown in Table 4.4.

Table 4.4: Priors for hyperparameters of the hierarchical model

Parameters in the models	Prior specification*
τ_{rw1} for the precision of Surface	$\text{Prob}(\frac{1}{\sqrt{\tau_{rw1}}} > 1) = 0.5$
r for the spatial range of the random field	$\text{Prob}(r < 20km) = 0.8$
σ_ω for the standard deviation of the random field	$\text{Prob}(\sigma_\omega > 0.4) = 0.2$
a for the AR(1) parameter in Eq. 4.4	$\text{Prob}(a > 0.5) = 0.7$
τ_{ar1} for the precision of AR(1) in Eq. 4.10	$\text{Prob}(\frac{1}{\sqrt{\tau_{ar1}}} > 5) = 0.1$
ρ for the AR(1) parameter in Eq. 4.10	$\text{Prob}(\rho > 0.5) = 0.7$

* All priors are PC priors.

4.4.4. Model assessment criteria

In this subsection, I describe how to assess the predictive performance of the models. As previously stated, we are interested in determining which is the most effective model for spatial predicting. In doing so, I consider a set of metrics, including the deviance information criterion (DIC), the cross-validated logarithmic scores (LCPO), the mean absolute error (MAE) and the root mean square error (RMSE) and the Pearson correlation coefficient (PCC).

The deviance information criterion (DIC) is calculated based on the whole

⁴Source: Bakka Haakon (30 January 2018). "Mesh Creation including Coastlines", <https://haakonbakka.bitbucket.io/btopic104.html>

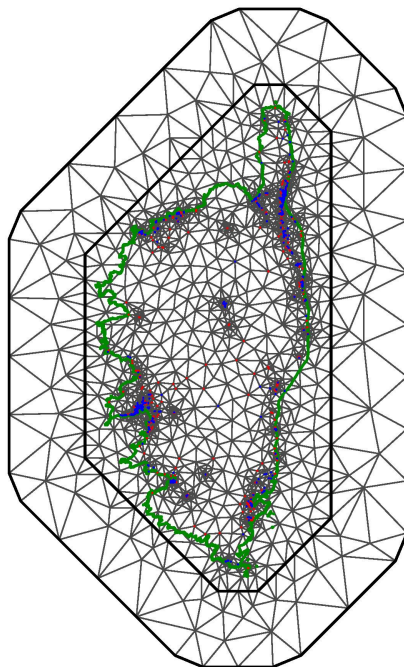


Figure 4.1: The mesh. The boundary of Corsica is delineated in green. The observations in the training set are dotted in red, whereas the observations in the validation set are dotted in blue. The inner black line distinguishes the inner mesh and the outer mesh.

dataset. The cross-validated logarithmic scores (LCPO) are also calculated based on the training dataset. Both are directly obtained from R-INLA outputs. The rest indices are calculated through the difference between the predicted housing price on the logarithmic scale and the observed measurements on the validation set. More precisely, the mean absolute error is defined by $MAE = \frac{1}{n} \sum_{j=1}^n |y_j - \hat{y}_j|$, and the root mean square error is computed as $RMSE =$

$\sqrt{\frac{1}{n} \sum_{j=1}^n (y_j - \hat{y}_j)^2}$. The Pearson correlation coefficient between observations and predictions is also computed. These three metrics quantify prediction errors, where lower values of MAE and RMSE, but higher values of PCC indicate more similarity among observed measurements and predicted values.

4.4.5. Results and discussion

The posterior statistics, including the mean, 0.025 and 0.975 quantiles of the fixed effect coefficients for the candidate models are displayed in Table 4.5.

Table 4.5: Regression results of all candidate models

	M0			M1			M2			M3		
	Mean	0.025quant	0.975quant	Mean	0.025quant	0.975quant	Mean	0.025quant	0.975quant	Mean	0.025quant	0.975quant
Intercept	11.703	11.595	11.811	11.745	11.635	11.855	11.715	11.590	11.840	11.742	11.638	11.847
ROOM	0.009	-0.005	0.024	0.008	-0.002	0.019	0.008	-0.002	0.019	0.009	-0.001	0.020
BATHROOM	0.096	0.066	0.127	0.063	0.042	0.084	0.063	0.042	0.084	0.059	0.039	0.080
GAR	0.020	0.008	0.031	0.043	0.034	0.053	0.048	0.039	0.058	0.048	0.039	0.058
FLOOR	0.009	0.005	0.013	0.019	0.015	0.022	0.018	0.015	0.022	0.019	0.016	0.022
DU	0.055	0.020	0.089	0.035	0.007	0.063	0.035	0.007	0.063	0.036	0.008	0.063
ST	-0.076	-0.116	-0.037	-0.100	-0.128	-0.072	-0.099	-0.127	-0.071	-0.100	-0.128	-0.072
PERIOD B	0.091	-0.010	0.191	0.008	-0.063	0.079	0.003	-0.068	0.073	0.001	-0.068	0.070
PERIOD C	0.085	-0.005	0.175	0.003	-0.064	0.070	-0.004	-0.071	0.062	-0.005	-0.071	0.061
PERIOD D	0.150	0.060	0.240	0.032	-0.035	0.100	0.023	-0.044	0.090	0.023	-0.044	0.089
PERIOD E	0.163	0.073	0.254	0.057	-0.010	0.125	0.051	-0.016	0.118	0.051	-0.015	0.118
PERIOD F	0.184	0.085	0.282	0.124	0.050	0.198	0.121	0.047	0.194	0.111	0.038	0.184
PERIOD G	0.237	0.148	0.327	0.232	0.163	0.301	0.238	0.170	0.307	0.226	0.158	0.295
PERIOD H	0.233	0.144	0.322	0.243	0.174	0.313	0.228	0.159	0.297	0.239	0.171	0.308
DBEAD	-0.012	-0.014	-0.010	-0.016	-0.022	-0.011	-0.016	-0.022	-0.011	-0.016	-0.021	-0.011
DHealFac	0.019	0.015	0.024	0.007	-0.001	0.015	0.007	-0.001	0.015	0.007	-0.001	0.015
DPuPriSch	-0.004	-0.005	-0.003	-0.002	-0.004	0.001	-0.002	-0.004	0.001	-0.001	-0.004	0.001
SURF σ_{rw1}^2	0.051	0.029	0.084	0.040	0.022	0.066	0.041	0.022	0.067	0.041	0.023	0.069
σ_ϵ^2	0.074	0.072	0.077	0.030	0.029	0.031	0.029	0.028	0.030	0.026	0.025	0.028
σ_ω^2				0.1	0.083	0.119	0.103	0.085	0.123	0.097	0.082	0.114
σ_{ar1}^2							0.003	0.0007	0.011			
Spatial Range (km)				1.880	1.606	2.244	1.881	1.598	2.196	1.741	1.502	2.021
ρ							0.939	0.828	0.990			
α										0.991	0.988	0.994

We find that the estimates for the fixed covariate effects are fairly robust with respect to M1, M2 and M3. This finding implies that the introduction of different spatial, temporal and spatiotemporal components does not significantly affect these estimates. More precisely, the posterior mean of the intercept is around 11.7, meaning that the average apartment price is approximately €120 571. Moreover, most covariates measuring the structural characteristics of apartments are significant with the expected signs. As expected, everything else remains the same, an additional room, bathroom, parking, floor and surface area generally improve the price of an apartment. A duplex apartment is likely to be more expensive than a standard apartment, but a studio apartment is likely to be cheaper. The non-linear relationship between apartment prices and living area in each model is displayed in Figure 4.2. Even though these curves are shown in the graph, we observe overall increasing trends over the whole range, which follow our expectation. That is to say, a positive coefficient for the living area would be expected. A significantly positive coefficient for the proximity to the nearest beach is also observed. Insignificant relations are found between the distance to the nearest public high school and apartment prices, between the distance to the nearest health facility and apartment prices.

As seen in the lower part of Table 4.5, we find that the majority of the remaining variance is due to the variance of the spatiotemporal process σ_{ω}^2 , rather than the variance of the measurement error, σ_{ϵ}^2 . In particular, for M1, M2 and M3, the estimate of σ_{ω}^2 is around four times greater than the estimate of σ_{ϵ}^2 . Such large variance is derived from the complex spatiotemporal structure, and is useful for explaining variability of the spatio-temporal process⁵. Further, we may conclude that the Corsican apartment price is partly determined by the spatiotemporal structure.

The posterior estimates of correlation parameters and spatial ranges are also reported in the lower part of Table 4.5. Regarding M1, the posterior mean estimate of the spatial range is 1.880 km (95% *CI*, 1.606; 2.244), meaning that the distance at which mean spatial correlation declines to 0.1 is 1.880 km. For M2, I estimate a range of about 1.881 km (95% *CI*, 1.598; 2.196). Further,

⁵The prior and posterior distributions for the hyperparameters in M3 are displayed in Figure 7.1 in Appendix 7.6.

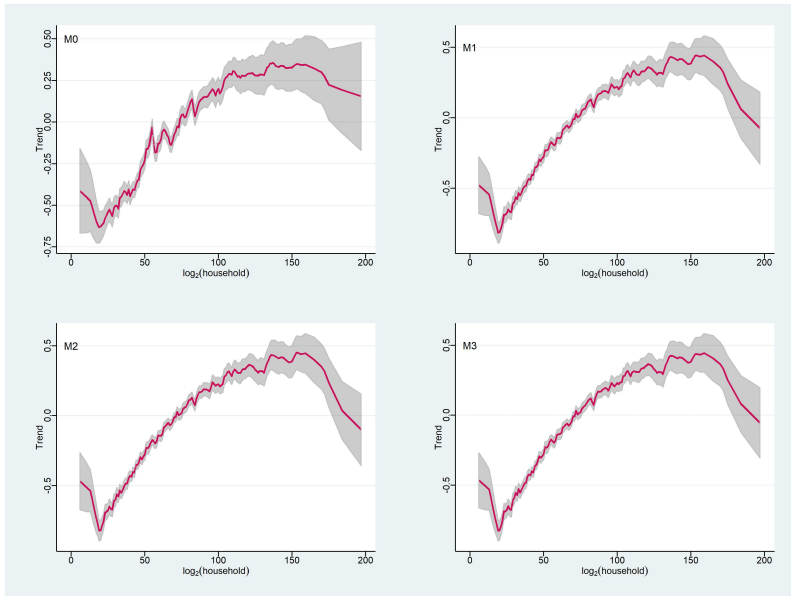


Figure 4.2: The empirical mean of the living area is shown as a black line. The 95% CI for the living area is shaded green. The model identifier is shown in the upper-left corner of each panel.

the posterior estimate of the AR(1) temporal correlation coefficient ρ is 0.939 (95% CI, 0.828; 0.990), showing the strong quarterly persistence of housing prices. M3 has the smallest spatial range, with 1.741 km (95% CI, 1.502; 2.021). This implies that the spatial correlation decreases rapidly, and that at 1.741 km there is a very weak correlation. On the other hand, its temporal correlation is strong. The posterior estimate is 0.991 (95% CI, 0.988; 0.992), indicating that the spatial random effects change quite slowly from quarter to quarter.

Table 4.6 displays the metrics for choosing the most effective model for our case study.

Regarding the DIC scores, the models involving any random effect components (M1, M2, M3) outperform the referenced model (M0). The improvement of the DIC scores could be evidence of the usefulness of considering spatial and temporal correlation in modelling. A possible reason for the well fitting mixed models (M1, M2, M3) is that through random components, observations can borrow strength from their neighbours in space and over time. Typically, adding the spatial random component to the base model results in a consider-

Table 4.6: Summary of model assessment

Model	DIC	LCPO	MAE	RMSE	PCC	Elapsed Time (Second)
M0	1,817.71	0.119	0.209	0.264	0.725	17.80
M1	-4,561.66	-0.278	0.132	0.186	0.875	78.36
M2	-4,710.98	-0.286	0.131	0.185	0.876	131.32
M3	-5,197.43	-0.304	0.129	0.184	0.878	19,744.39

4

able improvement in model fitting ($\Delta = -6,379.37$). The model (M2) includes an additional temporal random component that leads to further improvements ($\Delta = -149.32$). Lastly, M3 has the smallest DIC values ($\Delta = -486.45$ relative to M2), meaning that it provides the best fit to the data. The LCPO scores display in the same sequence as the DIC values. For M1, the LCPO score reaches -0.278 . With the inclusion of different spatial, temporal and spatiotemporal components, this score diminishes continuously, and reaches -0.304 for M3. This result shows that M3 should offer the strongest predictive power among the above-mentioned models. The descending order is observed for the MAE and RMSE values, and the ascending order is observed for the PCC coefficients. Note that compared with M2, M3 shows tiny increments or decrements with respect to the MAE, RMSE and PCC value ($\Delta_{MAE} = -0.002$, $\Delta_{RMSE} = -0.001$, $\Delta_{PCC} = 0.002$), but these increments or decrements cannot be ignored.

Overall, M0 offers the lowest prediction performance, with the largest DIC value, LCPO score, RMSE value and the smallest PCC value. Conversely, M3 should be the best model in terms of the predictive power. M2 seems to have similar predictive power compared with M3, but tiny differences still exist.

In addition to the results of model assessment, the information about the estimation time (in seconds) is displayed in the last column in Table 4.6. As expected, the running time during the estimation increases with the number of parameters and the complexity of random effect components. M0 without any random component is obviously the least time consuming. Compared with M1, M2 is characterized by an extra component for gauging temporal dynamics. The inclusion of the temporal component almost doubles the running time. M3 is particularly most time-consuming one. The running time of M3 is 150 times greater than that of M2. As previously indicated, M2 has a one-dimensional

vector for the AR(1) equation, while M3 has multidimensional matrices for the the AR(1) equation.

Thus far, I demonstrate that M3 has a better prediction capability, and such capability should come from a high computational cost. To further explore the spatial prediction, I decide to draw the posterior mean of the spatial random fields $\xi(s_i, t)$ in some quarters in Figure 4.3.

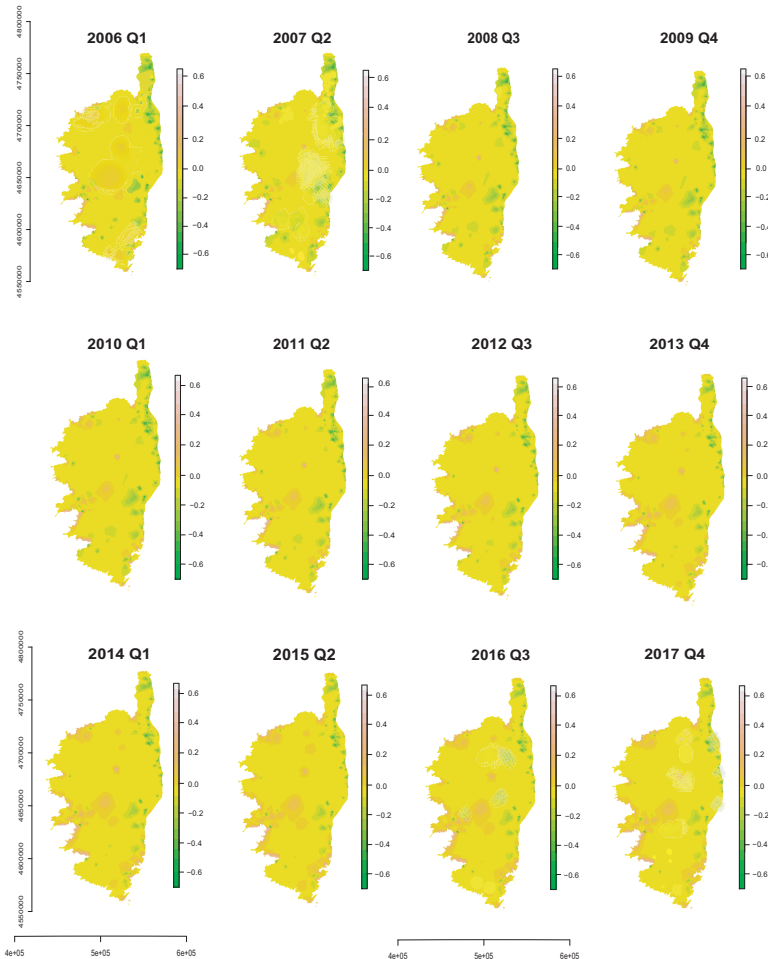


Figure 4.3: Time-sliced plots displaying the posterior mean of spatiotemporal random effects. 2006 Q1 refers to the first quarter of the year 2006. We observe that the clusters near the island center change over time regarding the colour and surface.

This figure gives us the first impression that the spatial pattern across the island is relatively constant from one quarter to the next. It is thought that the constant spatial pattern is also a reflection of the high posterior mean of the AR(1) parameter. We also observe several clusters. Apartment prices are significantly affected by their location inside these clusters. In general, most clusters are situated on coastal plains. A few clusters are dotted in the junction between the coastal plain and the inland area. More specifically, the cluster in which apartment prices are positively affected by their location is called a "hot spot". If an apartment is located in a hotspot, the location will generate an additional increase in its price. A cluster in warm colour is the "hot spot" in these figures. In contrast to the "hot spot", the cluster in which apartment prices are negatively affected by their location is deemed a "cold spot". The "cold spot" is plotted in cold colours. Spatially, the "cold spots" include the eastern coastline and the whole northern tip of the island, whereas the "hot spots" include the north-western area and the western coastline. Furthermore, "cold spots" and "hot spots" near the coast evolve slowly over time. By contrast, "hots pots" located in the inland area evolve relatively fast.

To gain further insight into the impact of location on apartment prices, I take an anti-logarithm. Concerning the scale of Figure 4.3, it varies from -0.6 to 0.6 approximately, which means that in some "hot spots", the location may increase the expected apartment price up to 82.21% ($\exp^{(0.6)} - 1 = 0.8221$). However, in some "cold spot" zones, the location probably causes a 45.12% ($\exp^{(-0.6)} - 1 = -0.4511$) reduction in the expected apartment price. We may conclude that the location can be a crucial factor for apartment prices.

4.5. Concluding remarks

In this chapter, I propose a powerful framework to predict apartment prices. This framework is based on the flexible Bayesian hierarchical models and two novel data fitting techniques, INLA and SPDE.

To illustrate this framework, I investigate the Corsican apartment prices using a unique dataset on the apartment transactions from the first quarter of 2006 to the fourth quarter of 2017. Further, I propose a set of Bayesian hierarchical models incorporating different spatial, temporal and spatiotemporal

random components. The specifications of these components are demonstrated with increasing complexity, from a pure spatial process to the sum of a pure spatial process and a pure temporal process, and finally to a spatial process evolving in time according to an AR(1) specification with spatially correlated innovations. All models are fitted by INLA coupled with the SPDE approach. Fitting these models usually requires large computational resources and long running time, but the INLA-SPDE approach addresses these problems by Laplace approximation, numerical integration and GMRFs. Thus, this approach provides fast and reliable Bayesian inference under affordable computing power.

The model comparison is based on criteria that account for data fitting, spatial predictive power and computational costs. Moreover, the estimates of the covariates for the candidate models with random effect components are fairly robust as the specification changes. This allows us to compare models solely based on their random effects components. M1 is discarded since it does not take into account the temporal process. M2 requires less computational resources, but its prediction capability is slightly weaker than that of M3. M3 offers the best predictive power. Even though M3 is time consuming, the computational resources are still affordable.

According to our findings, it can be confirmed that most apartment attributes and the proximity to the nearest beach significantly affect the apartment price. I also point out the nonlinear relationship between the living area and apartment price. Additionally, regarding the spatial relation, I specify several "hot spots" and "cold spots", where the apartment prices are significantly affected by their locations. Moreover, in general, the location of the "hot spots" and "cold spots" remains stable over time. This phenomenon reflects the high posterior mean value of the AR(1) parameter.

Currently, the INLA-SPDE approach is not widely seen in hedonic pricing literature, I suggest that researchers consider both the Bayesian hierarchical spatiotemporal model and the INLA-SPDE approach as instruments in their toolbox when they investigate point-referenced housing data and real estate economics. In the next chapter, a Bayesian hierarchical spatiotemporal model for areal data is proposed. This model is used for studying second home rates in Corsica. I also look into different spatial, temporal and spatiotemporal latent

effects, but the spatial component is built on the conditional autoregressive model family.

5

Unveiling spatial and temporal patterns of second home dynamics: a Bayesian spatiotemporal analysis for a Mediterranean island

5.1. Introduction

With the rapid evolution of land cover/land use change in the past decades, many second homes were built in the European countries around the Mediterranean, e.g., Spain, Italy and France (Hof and Blázquez-Salom, 2013).

Second homes have often been considered as a threat to rural/coastal areas, rather than as an opportunity (Coppock, 1977; Gallent, Mace, *et al.*,

2017). The impact of such threat spreads from social, environmental aspects to economic aspects (Hall and Müller, 2018). Regional scientists (Gallent, Mace, *et al.*, 2003) argued that second home development could undermine local culture, languages in rural areas. Others (Stettler and Giovanelli, 2008) stated that second homes decrease the quality of landscapes around tourist attractions. More importantly, economists (Paris, 2009) demonstrated that second homes relate to gentrification and the instability of local housing markets. Thus, several scholars (Gallent and Tewdwr-Jones, 2001) employed the term “epidemic” or “endemic” to describe the effect of second homes on local communities.

On a European scale, the falling cost of travel, the progressing of information and communications technology and the growing of wealthy households in affluent countries have contributed to the investment in second homes (Paris, 2009). In addition, these purchases are recognized as the external demand for a local housing market (Égert and Mihaljek, 2007; Wheaton and Nechayev, 2008). Following basic economic theories, the housing price should be a function of supply and demand in a perfectly competitive market. If the demand for houses increases but the supply remains, there will be a shortage of houses in the market and therefore house prices will go up.

Despite the Europe-wide nature of second homes, to my knowledge, no spatiotemporal analysis has been performed on the second home growth yet. As such, the objective of this study is to investigate second home rates by means of identifying their spatiotemporal patterns at a county level. In the study, second homes in Corsica (a French region) are chosen as a case study.

The Corsican context is worthy of studying for several reasons. Corsica is a famous tourist destination in the Mediterranean, whereas most inhabitants think that second homes affect their lives. In 2014, the second home rate in Corsica equalled 35%, which was 26% higher than that at the French national level. In some villages, this proportion even reached 80% (Maupertuis *et al.*, 2017). With the growth of tourism, this rate might further increase. As an island, Corsica merely provides limited residential lands. Local public agencies and economists (Giannoni *et al.*, 2017; Caudill *et al.*, 2019) have identified the growth of second home properties as a crucial issue to urban planning and rural development. They, 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 was dominated by qualitative approaches or case study designs. Few papers employed basic quantitative approaches, *e.g.*, mapping volumes, identifying spatial clusters and applying classical linear regression (Back and Marjavaara, 2017; Barke, 2007). These methods, however, may lead to biased results and unrealistic inference due to the presence of spatial and/or temporal dependence in data.

Facing the above-mentioned research gaps, this chapter makes two main contributions. Firstly, I 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 components, temporal components and a space-time component. These complex components account for the spatio-temporal dependency on the second home rate 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, therefore both a parametric and a nonparametric structure for temporal components are considered. Through borrowing strength from neighboring areas and adjacent periods, the proposed models provide reliable estimates. Another key point is that the models can be used within a binomial response setting. In this study, they overcome the limitation of a Gaussian response variable in classical spatial econometric models, *e.g.*, the spatial lag model (SAR) and the spatial error model (SEM) (Anselin, 1988). Consequently, I 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 components significantly increases model complexity. Therefore estimation is carried out by the Integrated Nested Laplace Approximation (INLA) approach (Rue, Martino, *et al.*, 2009). This approach takes 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 this work are fitted by the R-INLA package (Martins *et al.*, 2013).

The primary objective of this chapter is to reveal the spatiotemporal patterns of second home rates at the county level in Corsica. I examine potential factors, which may affect the second home rate. Different spatial and temporal priors are investigated to give accurate estimates. The remainder of this paper is structured as follows. In Section 5.2, an exploratory spatial data analysis is implemented for Corsican second home data. In Section 5.3, I describe the Bayesian hierarchical binomial model specification, in particular, how to incorporate space, time and space-time random effects in the Bayesian hierarchical model. Section 5.4 details empirical results and interpretation. A robustness check is conducted in Section 5.5, and I draw a conclusion in Section 5.6.

5.2. Second homes in Corsica: An exploratory spatial data analysis

The Corsican second home counts data are collected over 360 counties from 2006 to 2016, representing 11 years. As such, this setting results in a total of 3960 space-time units without any missing values. These data at the county level are obtained from “French National Institute of Statistics and Economic Studies” (INSEE) through the French population census.

Figure 5.1 gives the annual second home rates among the 360 counties, where the raw counts can be found in Table 7.2 in Appendix 7.5. A gradual and steady augmentation of second home proportion is observed. I notice that the median of second home proportion remains stable during the first two years and slightly increases during the subsequent five years. The proportion of the last four years remains fairly stable. The interquartile range of the proportion shows similar step-changes to the median proportion, falling into $[0.4, 0.6]$.

Figure 5.2 displays the distribution of the rate of the second homes over the total house counts in the 360 counties during 2006-2016. Several counties (ID. 125, 128) situated in the northern tip of Corsica had a low second home rate in 2006 and 2007, which turned up in 2008 and remained stable in the subsequent years. The second home rate in Bonifacio (ID. 80) was relatively low until 2010 and then increased. Moreover, I observed that counties with

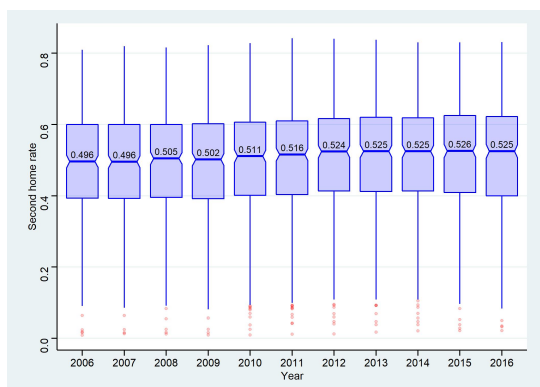


Figure 5.1: Boxplot of the temporal trend in the raw rate of second homes as a proportion of the total number of houses. The marked number indicates the median of the annual second home rate among Corsican counties.

a high second home rate clustered in the Cap Corse area (in pink in Figure 5.3), Saint-Florent area (in olive in Figure 5.3) and Balagne area (in yellow in Figure 5.3). Low rate zones involve the Grand Bastia area (in orange in Figure 5.3), the Corsica center (in navy in Figure 5.3) and the Ajaccio area (in green in Figure 5.3). These high and low rate zones likely persisted during the whole study period. As the temporal trend of the second home rate by county evolves differently, it is too restrictive to impose a linear relationship. Hence, space-time interaction terms should be considered for modelling the second home rate.

I employ Moran's I tests (Moran, 1950) to further investigate the spatial dependence structure in the data. To do so, the adjacency matrix is defined by queen contiguity weights (The visualisation of the spatial weights matrix is shown in Figure 7.2 in Appendix 7.6). The results in Figure 5.4 provide evidence that there is positive spatial dependence in the second home rates with an average of 0.334 ($p < 0.05$) over the 11 years. Hence, when modelling the Corsican second home rate, I take the spatial dependence into account.

After reviewing second home literature from economic, social and environmental perspectives, I find that there are two key factors affecting second homes, i.e., amenities and socio-economic factors. Several researchers (Hall and Müller, 2004; Müller and Hoogendoorn, 2013) showed that a key de-

5. Unveiling spatial and temporal patterns of second home dynamics:
102 a Bayesian spatiotemporal analysis for a Mediterranean island

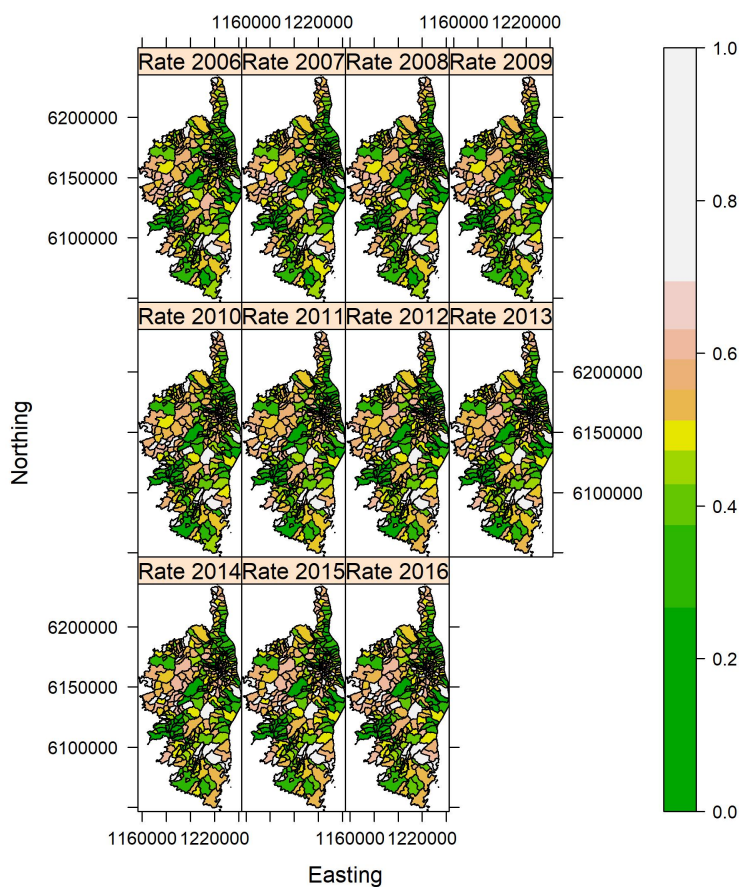


Figure 5.2: Geographic distribution of second home rate from 2007 to 2016.

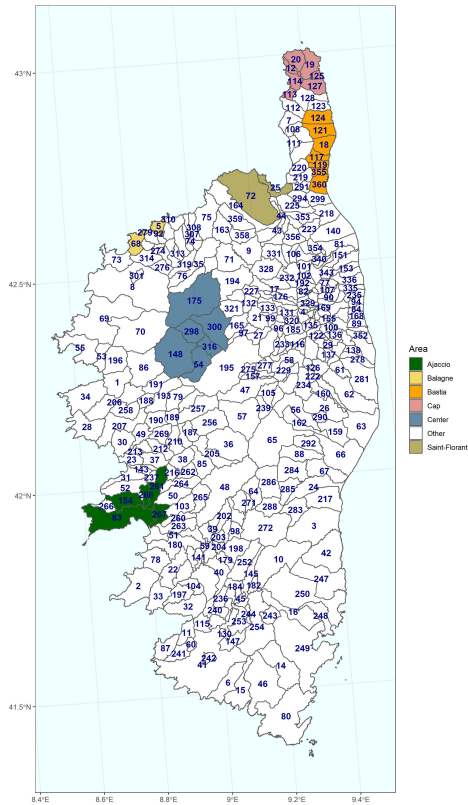


Figure 5.3: 360 counties of Corsica. County identifiers are shown on the map.

5. Unveiling spatial and temporal patterns of second home dynamics: 104 a Bayesian spatiotemporal analysis for a Mediterranean island

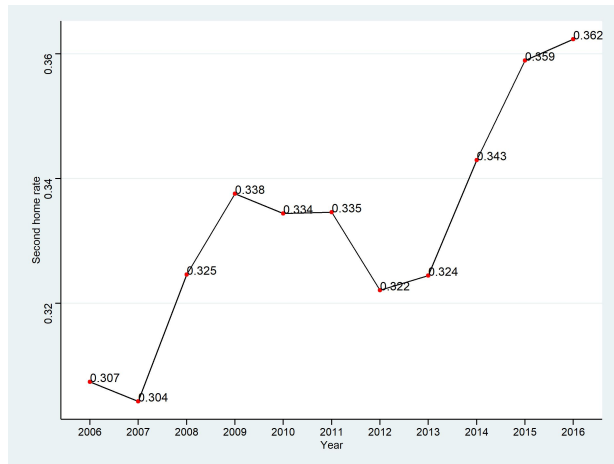


Figure 5.4: Annual Moran's I statistics on second home rates. All are significant at $p = 0.05$ level.

5

terminant for selecting second home locations was their high amenity value. Extensive research has focused on the impact of various amenities in different regions. Müller (2002) found that more and more second homes were concentrated around coastlines and uplands due to scenic qualities and recreation opportunities in Northern Sweden. Barnett (2007) concluded that ideal second home locations should satisfy weather, infrastructure, views, history and nature in Central and Eastern Europe. Norris and Winston (2009) indicated that second homes were usually located in the areas with amenity-rich landscapes or proximity to sea, rivers, lakes and mountains in Ireland. Kaltenborn, Andersen, Nellemann, *et al.* (2008) showed that Norwegian second homes increasingly occurred closely around mountain and coast tourism resorts. Additionally, some second homes located in areas with historical or social meaning (Kaltenborn, Andersen, and Nellemann, 2007).

Regarding social-economic factors, Norris and Winston (2009) concluded that the growth of second homes was likely related to local factors (Hall, 2015). According to Norris and Shiels (2007), some local governments passed tax incentive schemes to promote second home development. In addition, Barke (2007) showed that the number of second homes in a given province was related to the provincial population size in Spain, and concluded that depopulation was an important factor to create second homes. Lastly, Dower

(1977) showed that high costs of borrowing money could slacken the pace of second home growth because of additional capital demand.

According to the existing literature and data availability, I introduce three types of variables in my study. The first type is the amenity. I consider four variables: the physical landscape counts within a county, e.g., the number of lakes, alpine rocks, estuaries; the cultural landscape counts within a county, e.g., the number of castles, city walls, towers, churches; the coastal county, indicating a county locates on the coast or not; and the mountainous county, indicating a county locates in the mountainous areas or not, where the threshold of the mountainous area is based on the 500 m elevation. These four variables are time-invariant, meaning that these variables are mainly used for gauging spatial variation.

The second type of variables are social-economic variables. These include the number of households of a county; the annualized interest rate; the council tax; and the unemployment rate. The third type of variables are accessibility variables. I consider distances to the nearest "gate" of Corsica, and distances to the regional university. The "gates" contain all commercial ports and airports on the island. To avoid the impact of each variable measured in different units, I rescale all continuous variables by means of a logarithm transformation.

I use the number of households of a county as a proxy for county population. The household, council tax and land tax covariate are spatiotemporal, but the interest rate is time-varying only. The unemployment rate is spatiotemporal, based upon the "zone d'emploi" level. Hence, it captures limited spatial variability. Descriptive statistics for these variables are shown in Table 5.1.

5.3. Spatiotemporal modeling for areal data

5.3.1. Spatiotemporal data

Before introducing Bayesian hierarchical spatiotemporal models, we look into data types. As previously stated, second homes data are areal data, which are collected over 360 spatial units (counties) for 11 years. This, however, implies a balanced spatial panel setting.

Table 5.1: Descriptive statistics of independent variables

Type	Statistic	Mean	St.Dev	Min	Pctl(25)	Pctl(75)	Max
count	physical landscapes			0			19
	cultural landscapes			0			28
dummy	mountainous county			0			1
	coastal county			0			1
continuous	$\log_2(\text{household})^*$	6.575	1.976	1.585	5.129	7.748	14.846
	$\log_2(\text{interest rate})^*$	-4.865	0.402	-5.733	-5.058	-4.556	-4.353
	$\log_2(\text{council tax})^*$	-2.801	0.645	-5.702	-3.206	-2.281	-0.712
	$\log_2(\text{unemployment rate})^*$	-3.412	0.240	-4.053	-3.556	-3.231	-2.905
	$\log_2(\text{distance to "gate"})^*$	3.915	0.927	-0.160	3.429	4.587	5.411
	$\log_2(\text{distance to university})^*$	4.985	0.840	2.288	4.501	5.575	6.594

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

To capture all information from the data, the proposed models should handle both count data and latent spatial, temporal and spatiotemporal processes. Therefore, I turn to the Bayesian hierarchical spatiotemporal model again (Banerjee *et al.*, 2014, chap. 6).

5

5.3.2. Modelling spatial dependence

I apply the LCAR model to gauge spatial dependence among spatial units. Therefore, the random vector for joint spatial effects (also known as a joint spatial component) reads, $\Gamma = (\gamma_1, \dots, \gamma_S)$, and together with a mixing parameter λ in the precision matrix. The mixing parameter indicates the joint spatial effect explained by spatially structured and unstructured components, as well as measures the strength of spatial dependence among units. The random vector Γ is assumed to follow a multivariate normal distribution with mean 0 and a designed precision matrix Q_Γ ,

$$\Gamma \sim MVN(0, \sigma_\Gamma^2 Q_\Gamma^{-1}) \quad (5.1)$$

$$Q_\Gamma = [\lambda W^* + (1 - \lambda) I_S]. \quad (5.2)$$

Here I_S denotes the identity matrix of dimension S . W^* denotes the variant of a $S \times S$ spatial adjacency matrix. Its diagonal elements w_{ii}^* are equal to the number of neighbours of each unit n_{i+} , the off-diagonal elements w_{ij}^* are equal to -1 if unit i and unit j are neighbours, and 0 otherwise. The mixing parameter ensures that the precision matrix Q_Γ is a weighted average of the

W^* and I_S matrices with $0 \leq \lambda \leq 1$.

5.3.3. Modelling temporal dependence and spatio-temporal variability

In addition to the spatial latent process (dependence), the temporal latent process (dependence) should be taken into account if data are collected over the years. Hence, several temporal and spatiotemporal specifications are proposed. Further, both parametric and non-parametric smoothing functions of time are considered.

The two prominent specifications for the temporal and spatiotemporal components are the linear trend model proposed by Bernardinelli *et al.* (1995) and the dynamic trend model proposed by Held (2000).

Regarding the linear trend model,

$$(\xi + \varphi_i) \cdot t \quad (5.3)$$

where $\xi \cdot t$ denotes a linear function of time representing a global trend in time and $\varphi_i \cdot t$ is an area-specific linear function gauging an area-specific trend to depart from the global trend. I specify a Gaussian exchangeable prior $\text{Normal}(0, \sigma_\varphi^2)$ to φ_i . This model is a simple, straightforward variant of a random slope model, where the slope $(\xi + \varphi_i)$ is spatially varying.

The linearity assumption, however, is restrictive. To relax the linear restriction, I model time as a dynamic nonparametric formulation:

$$\zeta_t + \delta_{it} \quad (5.4)$$

where ζ_t ($Z = (\zeta_1 \dots \zeta_T)$) is the temporally structured random effect (the temporal main effect) and δ_{it} denotes the space-time interaction term. The parameter ζ_t is usually modelled by a first-order of random walk $\zeta_t | \zeta_{t-1} \sim \text{Normal}(\zeta_{t-1}, \sigma_\zeta^2)$.

For the space-time interaction component $\Delta = (\delta, \dots, \delta_{S1}, \dots, \delta_{1T}, \dots, \delta_{ST})$, a Type I unstructured interaction is used (Held, 2000). The Type I interaction is the Kronecker product of spatially and temporally unstructured components, which can be considered as a random intercept based on all observations. Seen differently, this component represents global space-time het-

erogeneity and measures the deviation from the spatial and temporal main effects. This component is assigned a multivariate normal distribution $\Delta \sim MVN(0, \sigma_{\Delta}^2 (I_S \otimes I_T))$.

These spatial and spatiotemporal models can be formulated as latent Gaussian models (Rue and Held, 2005). I apply INLA for these models, as it is designed for latent Gaussian models, and hence return fast and accuracy Bayesian inference.

5.3.4. Candidate model specifications

To model the second home data, the LCAR model and its spatiotemporal extensions are considered. In this study, I specify five latent processes with increasingly complex representation of space and time. As discussed above, Bayesian spatial and spatiotemporal models can be written as latent Gaussian models. I assume that the observed counts for second homes (y_{ij}), conditionally on the rate of second homes (π_{it}) and the total number of houses (n_{it}), are binomial-distributed (Ferrari and Comelli, 2016),

$$Y_{it} \sim \text{Binomial}(n_{it}, \pi_{it}). \quad (5.5)$$

Counties are denoted by $i = \{1, \dots, S\}$ ($S = 360$ here) and years are labelled as $t = \{1, \dots, T\}$ ($T = 11$). Housing transactions are not frequent in Corsica, and hence a second home could hardly sell more than once in a year, meaning that an annual housing survey is considered a Bernoulli trial, where the occurrence of a second home is treated a success. Various applications using spatiotemporal Bayesian hierarchical binomial models include modelling disease risks (MacNab, 2003), violent crimes (Zhu *et al.*, 2006), public health intervention (Viola *et al.*, 2013), presidential elections (Linzer, 2013) and public confidence on police (Williams *et al.*, 2019).

In the process model, a logit transformation is applied to π_{ij} , and thus the transformed π_{it} equals the sum of all structured additive predictors. Table 5.2 shows the candidate specifications for the second level.

Table 5.2: Candidate Models for the second level

Model	Type	Equation
0	Classical Binomial without any latent structures	$\text{logit}(\pi_{it}) = x_{it}\beta + \sum_{j=1}^p f(x_{jt})$ (5.6)
1	Purely spatial	$\text{logit}(\pi_{it}) = x_{it}\beta + \sum_{j=1}^p f(x_{jt}) + \Gamma$ (5.7)
2	Spatial and temporal jointly	$\text{logit}(\pi_{it}) = x_{it}\beta + \sum_{j=1}^p f(x_{jt}) + \Gamma + \xi \cdot t$ (5.8)
3	Spatial and temporal jointly	$\text{logit}(\pi_{it}) = x_{it}\beta + \sum_{j=1}^p f(x_{jt}) + \Gamma + Z$ (5.9)
4	Spatiotemporal with interactions	$\text{logit}(\pi_{it}) = x_{it}\beta + \sum_{j=1}^p f(x_{jt}) + \Gamma + (\xi + \varphi_i) \cdot t$ (5.10)
5	Spatiotemporal with interactions	$\text{logit}(\pi_{it}) = x_{it}\beta + \sum_{j=1}^p f(x_{jt}) + \Gamma + Z + \Delta$ (5.11)

Thus far, we have known the candidate specifications for latent spatial, temporal and spatiotemporal processes. At the third level, I assign hyperprior distributions to all parameters appeared in the previous levels. The used priors are listed below:

$$\lambda \sim \text{Uniform}(0, 1),$$

$$\log\left(\frac{1}{\sigma_{x_{jt}}^2}\right), \log\left(\frac{1}{\sigma_{\Gamma}^2}\right), \log\left(\frac{1}{\sigma_{\xi}^2}\right), \log\left(\frac{1}{\sigma_{\Delta}^2}\right), \log\left(\frac{1}{\sigma_{\varphi}^2}\right) \sim \text{logGamma}(1, 0.00005).$$

(5.12)

I maintain the following strategy. Model 0, serves as the benchmark. It is a classical Binomial model without any spatial effects. However, I detect the spatial autocorrelation in the data, it is natural to consider a model in the conditional autoregressive family and then to involve temporal components for capturing 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 neighbours but not over the test period. For Model 2, a single value is estimated for each neighbour and period, but the parametric formulation seems restrictive. The linear assumption is relaxed in Model 3 and the model can gauge temporal dependency 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.

5.4. Empirical results

5.4.1. Model assessment results

DIC, LCPO and RMSE values are displayed in Table 5.3.

Table 5.3: Model assessment via DIC, LCPO and RMSE

	DIC	LCPO	RMSE
Model 0	210,055.60	26.2451	0.14298
Model 1	28,633.33	3.7611	0.04227
Model 2	28,572.06	3.7517	0.04226
Model 3	28,550.95	3.7466	0.04205
Model 4	27,939.49	3.6524	0.04045
Model 5	27,857.43	3.6292	0.03412

The DIC score indicates that the classical Binomial model is poorly fitted. Comparing Model 0 with Model 1, the fit is much improved ($\Delta_{DIC} = -181,422.27$). The marked decrease of the DIC score 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 ($\Delta_{DIC} = -61.27$) slightly. The fit is further improved ($\Delta_{DIC} = -82.38$) in 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 score is largely reduced ($\Delta_{DIC} = -693.84$) because of the inclusion of a space-time interaction term. After incorporating the Type-I space-time interaction term, the DIC score is further reduced in Model 5. 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 will be used to showing the estimated coefficients of covariates, as well as the estimated spatial, temporal and spatiotem-

Table 5.4: Posterior estimates of the covariates in Model 5

	Mean	St.Dev	0.025 quant	0.975 quant
Intercept	-3.154*	0.500	-4.157	-2.190
physical landscapes	0.082*	0.027	0.030	0.135
cultural landscapes	-0.047*	0.019	-0.085	-0.010
coastal county	0.720*	0.109	0.506	0.935
mountainous county	0.217*	0.083	0.054	0.379
$\sigma_{\log_2(\text{household})}^2$	0.217	0.039	0.151	0.304
$\log_2(\text{interest rate})$	-0.161*	0.028	-0.217	-0.108
$\log_2(\text{council tax})$	-0.016	0.011	-0.037	0.006
$\log_2(\text{unemployment rate})$	-0.049	0.044	-0.136	0.036
$\log_2(\text{dis_gates})$	0.051	0.057	-0.059	0.164
$\log_2(\text{dis_university})$	0.170*	0.069	0.038	0.308
σ_F^2	0.7344	0.1290	0.5202	1.0253
λ	0.2980	0.0903	0.1502	0.4999
σ_Z^2	0.0015	0.0009	0.0004	0.0038
σ_Δ^2	0.0093	0.0006	0.0082	0.0105

* Indicates the significance of independent variables.

poral patterns of the second home rate.

5.4.2. Fixed effects and interpretation

For each covariate, the upper part of Table 5.4 presents the log odds of the second home rate associated with a 1-unit or percentage increase and its associated 95% credible interval (CI).

Most covariates are significant except for the council tax, the unemployment rate and the distance to the nearest "gates".

The relative log odds of the second home rate (π_{it}) increases 0.082 (95%CI, 0.030; 0.135) times with a 1-unit increase in the physical landscape count, given all else is equal. In contrast, a unit increase in the cultural landscape count decreases 0.047 (95%CI, -0.085; -0.010) times in the log odds of the second home rate. Hence, both landscape variables are informative. The result suggests 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 (Cheer *et al.*, 2019). 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.

There would be a 0.720 (95%CI, 0.506; 0.935) increase in the log odds

of the second home rate, as an inland county changes to a coastal county. One possible implication of this is that Corsican second home buyers prefer living near coasts to living in mountainous areas. In addition, the posterior mean coefficient 0.217 (95%CI, 0.054; 0.379) indicates that the log odds of the second home rate increases by 0.217 approximately, as a mountainous county compares with a flat county, given all else equal.

The positive association of the log odds of second home rate with the distance to the university is also observed $\beta_{\log_2(dis_university)} = 0.170$ (95%CI, 0.038; 0.308). Consequently, a county far away from the university tends to have a high second home rate.

After analysing plots of each covariate against the model residuals, a non-linear relation between the logged households and the second home rate is detected. Hence, the default first-order random walk (RW1) smoother in the R-INLA package is applied to the logged households. Figure 5.5 displays the non-linear relation between the base 2 logarithm of households and the second home rate, showing on the log odds scale. We initially observe a downward trend. The log odds of the second home rate are at their highest in areas with the lowest value of logged households and decrease as the value of the 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 (around 2 000 to 8 000 households) locate close to the main cities or are the capital of cantons, and some of them are not far from national parks. In a word, 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.

The coefficient on the base 2 logarithm of interest rates has a posterior mean of -0.161 (95%CI, -0.217 ; -0.108), meaning that a doubling of interest rates translates to a 19.51% 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 the low interest rate means that obtaining home finance is more

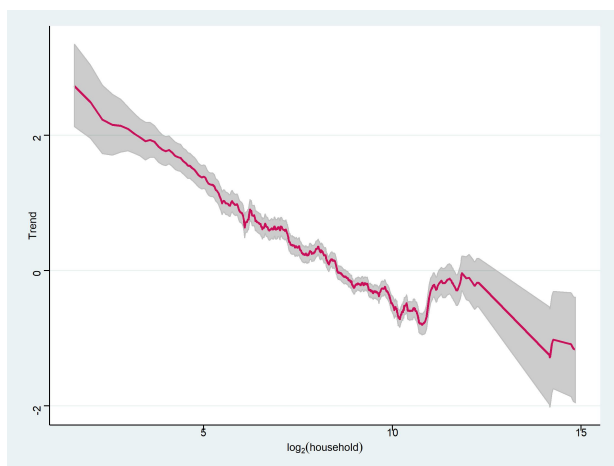


Figure 5.5: Log odds relations between $\log_2(\text{household})$ and the second home rate. The red line indicates the posterior mean log odds, while grey represents the corresponding 95% CI.

affordable (Paris, 2009).

5.4.3. Spatial and temporal random effects and interpretation

Once the covariates effects are gauged properly, I can investigate the spatial and temporal random effects, and the spatiotemporal interaction. The spatial random effect component shows the log odds of the second home rate associating with each county. The temporal random effect component indicates the overall temporal trend relating to unobserved temporal covariates, and finally the spatiotemporal interaction term captures the idiosyncrasy of a county. Initially, I calculate the proportion of marginal variance explained by each component, given by

$$p_i = \sigma_i^2 / (\sigma_\Gamma^2 + \sigma_Z^2 + \sigma_\Delta^2) \times 100\%, i = \{\Gamma, Z, \Delta\} \quad (5.13)$$

The corresponding proportions are 98.56%, 0.20%, 1.24%. These results suggest that a huge part of the variability is explained by the spatial patterns. According to the posterior distribution of the mixing parameter $\lambda = 0.2980$ (95%CI, 0.1502; 0.4999), the spatially unstructured component

contributes to the spatial variability most (69.18%), while the spatially structured component explains a sizable proportion (29.38%). Differently, the temporal component explains 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 the temporal component.

Figure 5.6 reveals the spatial pattern. In general, southeastern, north-western and some central counties show a greater second home rate than other counties. Moreover, a county (ID. 281) on the eastern coast presents the highest second home rates. In contrast, the Bastia area, the Ajaccio area and Corti are relatively low in terms of the second home rate. To distinguish the positive and negative contributions, I 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 of the second home rate most, given all covariates. More precisely, five main “hot spots” are Linguizzetta (ID. 281), Lumio (ID. 68), Grosseto-Prugna (ID. 51), Lecci (ID. 248) and Quenza (ID.10) county. In contrast, Biguglia (ID. 218), Furiani (ID. 299), Bastia (ID. 360), Alzi (ID.116) and Ambiegna (ID. 213) are marked as “cold spots”.

Figure 5.7 shows the overall temporal trend for the log odds of the second home rate. Two marked falls are recognized, including 2007-2009 and 2011-2012. The first period may be related to the global financial crisis of 2007-2008, and the second period may be related to the European debt crisis¹ during 2011-2012 (Pérez, 2019). We further notice that the scale of the vertical axis of the red line ranges from -0.05 to 0.04 approximately, which suggests a slightly positive increase in the second home rate over the past 11 years.

Investigating the type-I space-time interaction term provides further insight into the spatiotemporal pattern of the second home rate, as shown in Figure 5.8. In the first three years, the spatial pattern seemed to stabilize. From 2009 to 2013, the spatial pattern seemed unstable, and the map grew darker from 2014 to 2017. Moreover, some counties (ID. 8, 10, 28, 34, 301) were colored in light red at the beginning of the study period, while these

¹Source: Seth W. Feaster; Nelson D. Schwartz; Tom Kuntz (22 October 2011). “It’s All Connected: A Spectators Guide to the Euro Crisis”. The New York Times. https://archive.nytimes.com/www.nytimes.com/imagepages/2011/10/22/opinion/20111023_DATAPOINTS.html?ref=sunday-review.

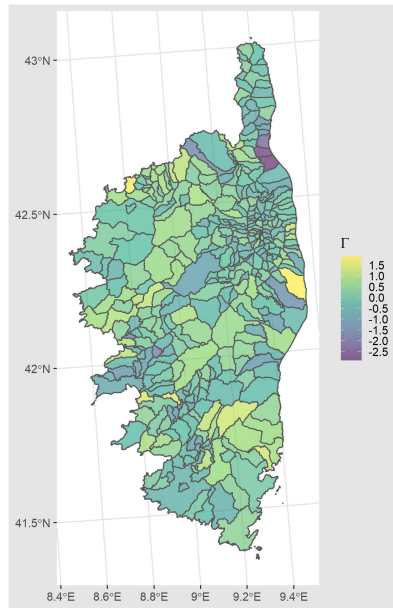


Figure 5.6: Posterior mean estimates of spatially joint random effects.

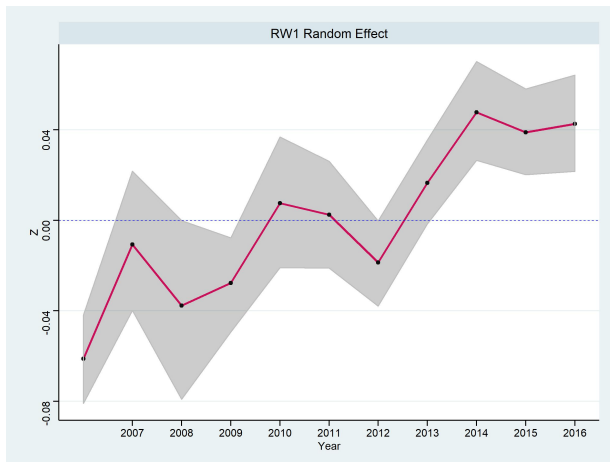


Figure 5.7: Posterior structured temporal trend for second home rate in Corsica. The red line indicates the posterior mean trend and grey represents the corresponding 95% CI.

5. Unveiling spatial and temporal patterns of second home dynamics: 116 a Bayesian spatiotemporal analysis for a Mediterranean island

Table 5.5: Tested hyperpriors in the prior sensitivity analysis

Component	Default	Test 1	Test 2
Spatially joint component (τ)	$\log \tau \sim \log \text{Gamma}(1, 5 \times 10^{-5})$	$\log \tau \sim \log \text{Gamma}(1, 1 \times 10^{-4})$	$\sigma \sim \text{Uniform}(0, \infty)$
Temporally Structured component (Z)	$\log \tau \sim \log \text{Gamma}(1, 5 \times 10^{-5})$	$\log \tau \sim \log \text{Gamma}(1, 1 \times 10^{-4})$	$\sigma \sim \text{Uniform}(0, \infty)$
Space-time interaction term (Δ)	$\log \tau \sim \log \text{Gamma}(1, 5 \times 10^{-5})$	$\log \tau \sim \log \text{Gamma}(1, 1 \times 10^{-4})$	$\sigma \sim \text{Uniform}(0, \infty)$

* $\sigma = \frac{1}{\sqrt{\tau}}$

Table 5.6: Estimated posterior mean, standard deviation and quantiles of the parameters for different hyperpriors

Prior	$\log \tau \sim \log \text{Gamma}(1, 1 \times 10^{-4})$				$\sigma \sim \text{Uniform}(0, \infty)$			
	Mean	St.Dev	0.025quant	0.975quant	Mean	St.Dev	0.025quant	0.975quant
Intercept	-3.153	0.500	-4.156	-2.188	-3.223	0.522	-4.272	-2.217
physical landscapes	0.082	0.027	0.030	0.135	0.081	0.027	0.028	0.133
cultural landscapes	-0.047	0.019	-0.085	-0.010	-0.046	0.019	-0.084	-0.009
coastal county	0.720	0.109	0.506	0.935	0.722	0.110	0.505	0.939
mountainous county	0.217	0.083	0.054	0.379	0.218	0.083	0.054	0.381
$\sigma_{\log_2(\text{household})}^2$	0.217	0.039	0.151	0.304	0.227	0.040	0.157	0.315
$\log_2(\text{interest rate})$	-0.161	0.027	-0.217	-0.109	-0.165	0.033	-0.232	-0.102
$\log_2(\text{council tax})$	-0.015	0.011	-0.037	0.006	-0.018	0.011	-0.039	0.003
$\log_2(\text{unemployment rate})$	-0.049	0.044	-0.136	0.036	-0.056	0.044	-0.144	0.031
$\log_2(\text{dis_gates})$	0.051	0.057	-0.059	0.164	0.051	0.058	-0.062	0.168
$\log_2(\text{dis_university})$	0.170	0.069	0.038	0.308	0.173	0.071	0.036	0.316
σ_{τ}^2	0.7355	0.1291	0.5207	1.0263	0.7756	0.1371	0.5403	1.0771
λ	0.2980	0.0903	0.1501	0.4998	0.3204	0.0937	0.1592	0.5216
Z	0.0015	0.0009	0.0004	0.0039	0.0026	0.0018	0.0006	0.0072
Δ	0.0093	0.0006	0.0082	0.0105	0.0093	0.0006	0.0082	0.0105

colored counties turned in blue after 2010. There were few counties (ID. 190, 191, 328) with light red color from 2011 to 2016.

5.5. Robustness Check

Concern may arise from three major issues: sensitivity to the priors, necessity of including covariates and endogeneity.

Regarding the prior sensitivity, I test different prior distributions to assess the change in the posterior distribution of all covariates and variance parameters. The tested priors and the fixed effects are shown in Table 5.5 and Table 5.6 respectively. 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. In addition, the posterior distribution of the variance parameters is quite similar for the different priors. These results suggest that Model 5 should not be sensitive to priors.

To evaluate the need to include all covariates, I rerun Model 5 without any covariates, named as the convolution model. From Model 5 to the con-

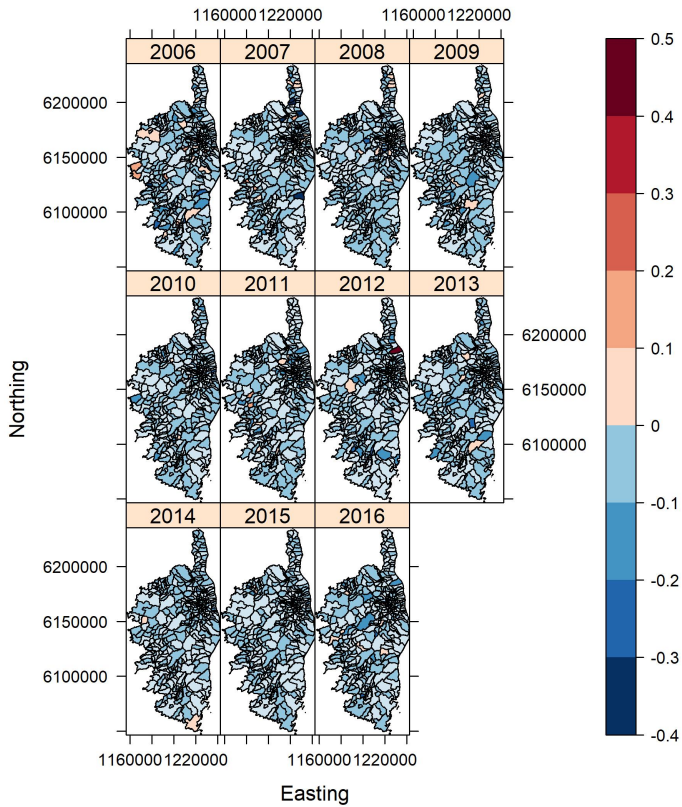


Figure 5.8: Posterior mean of the type-I spatio-temporal interaction δ_{it} for the log odds of the probability being second homes.

volution model, the decrease of the DIC score can be clearly seen in Table 5.7. We also notice that the posterior mean of the mixing parameter λ equals 0.542 (95%CI, 0.346; 0.741) in the convolution model. The increase of the posterior mean of the mixing parameter demonstrates that spatially-referenced covariates capture some spatial variability. Rao (2003) stated that incorporating covariates into small area estimation models could increase the model predictive power. The finding provides evidence for this point of view.

Table 5.7: Refitted models assessment

	DIC	LCPO	RMSE
Model 5	27,857.44	3.6292	0.0341
Convolution Model5	28,226.61	3.7106	0.0351

5

Lastly, the household variable may experience an endogeneity 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 investigation, I test strict exogeneity of the household variable via the Wooldridge’s approach (Wooldridge, 2010, pg. 285, pg. 490). The lead household variable is included in the model additionally. I initially run Model 5 including the logged households as a linear predictor. Then, the lead-1 or lead-3 households is included in Model 5 additionally. The posterior estimates for the two additional variables are 0.024 (95%CI, -0.132; 0.179) and 0.072 (95%CI, -0.024; 0.168) (See Table 5.8). Such a result shows that there are not any endogeneity issues.

Table 5.8: Estimated posterior mean and quantiles of the covariates for the strict exogeneity test

			Test 1				Test 2 lead-1				Test 3 lead-3	
	Mean	St.Dev	0.025quant	0.975quant	Mean	St.Dev	0.025quant	0.975quant	Mean	St.Dev	0.025quant	0.975quant
Intercept	0.088	0.483	-0.870	1.030	-0.184	0.518	-1.214	0.828	-0.025	0.624	-1.275	1.180
physical landscapes	0.091	0.024	0.043	0.138	0.090	0.024	0.043	0.137	0.088	0.024	0.040	0.135
cultural landscapes	-0.028	0.017	-0.061	0.005	-0.031	0.017	-0.064	0.002	-0.039	0.017	-0.072	-0.006
coastal county	0.733	0.099	0.538	0.929	0.719	0.100	0.523	0.916	0.675	0.102	0.474	0.876
mountainous county	0.243	0.076	0.093	0.392	0.245	0.076	0.095	0.395	0.257	0.077	0.106	0.408
$\log_2(\text{household}_i)$	-0.342	0.018	-0.378	-0.307	-0.350	0.055	-0.458	-0.241	-0.358	0.034	-0.425	-0.291
$\log_2(\text{household}_{i+1})$					0.016	0.055	-0.091	0.124				
$\log_2(\text{household}_{i+3})$									0.051	0.034	-0.016	0.118
$\log_2(\text{interest rate})$	-0.179	0.028	-0.237	-0.125	-0.211	0.037	-0.288	-0.140	-0.103	0.064	-0.231	0.022
$\log_2(\text{council tax})$	-0.009	0.011	-0.031	0.013	-0.004	0.011	-0.027	0.018	0.0003	0.012	-0.024	0.024
$\log_2(\text{unemployment rate})$	-0.066	0.046	-0.157	0.023	-0.085	0.052	-0.188	0.018	-0.128	0.067	-0.261	0.002
$\log_2(\text{dis_gates})$	-0.001	0.051	-0.103	0.099	0.005	0.074	-0.141	0.151	0.029	0.077	-0.121	0.180
$\log_2(\text{dis_university})$	0.130	0.062	0.007	0.255	0.186	0.091	0.008	0.367	0.184	0.094	0.0004	0.371

5.6. Concluding remarks

In this study, I propose a Bayesian spatiotemporal approach to identify the spatial and temporal dynamics on the second home rate in Corsica, France, from 2006 to 2016. I also investigate the impact of fixed effects, including amenity, socioeconomic factors and accessibility factors. These fixed effects provide additional insight into the dynamics of Corsican second home rates. To my knowledge, the approach in this study has previously not been used for analysing second homes and has offered a useful tool for practitioners to discover underlying covariates, to localize spatial clusters and to identify temporal trends.

The choice of models is initially motivated by spatiotemporal data indexed at medium geographical and temporal resolution. Further, I point out the importance of space and time in second home analyses. To gauge latent spatial and temporal information, I 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 programming language R. As a result, the models are widely applicable and well reproducible. INLA provides a feasible way to handle large, spatiotemporal datasets and return a considerably accurate result for practical applications. Based on the information criteria, the best fitting model owns spatial, temporal and spatiotemporal random components. Hence, the stable spatial pattern and the gradual temporal trend on the second home rate are uncovered.

Some key findings are related to the fixed effects. Most amenity covariates are associated with an increase in the log odds of the second home rate, except the cultural landscape. While, the negative association between the log odds of the second home rate and the households and the interest rate is currently detected. Moreover, there is not a significant relationship between the council tax, the unemployment rate and the log odds. Concerning accessibility covariates, the distance to the nearest "gate" is not statistically significant, but I find a positive association between the distance to the university and the log odds of the second home rate.

Regarding the random effects, the "hot spot" and "cold spot" areas of second home rates in Corsica are well identified. In addition, the gradual temporal

tend with two falls for the second home rate is also recognized throughout the 11-year study period. The spatiotemporal dynamics of the second home rate are finally described by the space-time interaction random effect. We gain some further insight into the spatiotemporal dynamics, rather than lose the information. Therefore, the proposed approach can be viewed as a decent tool to analyze second home dynamics, and the findings offer useful insights for formulating effective policies to manage second homes in Corsica, and then to meet the concern of inhabitants.

First, the outcomes of the proposed Bayesian hierarchical model provide convincing evidence that interest rates negatively contribute to Corsican second home rates. This finding justifies a potential policy intervention. The regional governors can propose the central authority to pass a law that forces local inhabitants to declare their properties. According to the deceleration, second home owners are not allowed to negotiate interest rates with their banks, and to benefit any tax incentives. As a result, the loan costs will increase, and investment purchases should be restrained.

In addition to the financial tool, interventions on second home hot spots may achieve better results than focusing on the financial tool alone. In doing so, all levels of government are advised to prioritize these zones, and there is a great need for coherent interventions to curb the high second home rate and to prevent second home buyers spilling over to adjacent counties. For example, second home purchasing restrictions should be implemented strictly in the hot spot zones and neighbouring counties (E.g., “taxe d’habitation sur les logements vacants” according to French law.). The regional authority may also require to regulate the real estate developers and to provide more low-rental houses in the hot spot zones.

While, for other counties, local governments should decide whether to apply the second home purchasing restriction. This enables the local government to be more responsive to the inhabitants’ needs. In other words, after recognizing the unique context of each county, a diversified, county-specific second home regulation is urgently needed. Therefore, upgrading the existing institutional framework for improving coordination between urban planning and housing policies, between the regional government and local government are essential elements to achieve balanced economic development and sus-

tainable urban planning, and hence to improve the quality of people's lives.

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 the medium geographical and temporal resolution. Since this area has not been analysed quantitatively, my findings probably provide valuable information for the intervention from the local government. From a methodological perspective, I underline the importance of space and time. Spatial, temporal and spatiotemporal information is very likely missed in many second home analyses. In addition to latent spatial and temporal information, the inclusion of the amenity, accessibility and socio-economic factors offers additional insights in the Corsican second home rate. Moreover, Bayesian hierarchical models provide an alternative way out of classical spatial econometric models to handle spatial and spatiotemporal data. After investigating Corsican apartment prices and second home rates, a question arises, how to handle both point-referenced data and areal data simultaneously? In the next chapter, I will estimate the implicit prices of some environmental amenities via a Bayesian hierarchical model in the presence of so-called spatially misaligned data.

6

The value of sea views in a Mediterranean island: Hedonic analysis with spatial multilevel modelling

6.1. Introduction

Environmental amenities usually play an important role in people's residential property choices (Chasco and Le Gallo, 2013; Irwin *et al.*, 2014; Liu, Hu, *et al.*, 2020; Liu, 2020). Recognition of the economic value of these amenities helps policymakers better understand the decisions of homebuyers. This knowledge could also contribute to urban development planning and differentiating compensation policies (Carrión-Flores and Irwin, 2017; Troy and Romm, 2004). Corsica is a beautiful island in the Mediterranean Sea, its coastal zones are aesthetically desirable and economically important. For example, roughly 70% of the population lives in shore-adjacent counties. Most cities and tourist attractions distribute along coasts (Vogiatzakis *et al.*, 2008). Even though local

researchers (Kessler and Tafani, 2015) have paid attention to the coast for a long time, no studies have been conducted to quantify the value of sea views. In the existing literature, the main approach to assessing amenities is to analyze their implicit prices, via the hedonic pricing method (Rosen, 1974). In this chapter, I assess the implicit price of sea views on Corsican apartment prices within the hedonic pricing framework.

Scenic amenities, however, are usually perceived through the human eyes. Views, therefore, are difficult to quantify. Earlier studies attempt to value views using dummy variables, categorical variables within the hedonic pricing framework (Benson *et al.*, 1998; Bourassa, Hoesli, and Peng, 2003; Michael *et al.*, 2002). While the utilization of these approaches has some limitations. Nowadays, with the help of geographic information systems (GIS) and point-referenced data, recent studies incorporate a more precise and objective measure of views (Hindsley *et al.*, 2013; Paterson and Boyle, 2002; Yu *et al.*, 2007).

In addition to the difficulty of measuring view, another important question arises: what is the implicit price of sea views in the presence of spatial effects? To answer this question, spatial econometrics (Anselin, 1988) seems to provide a promising way. We have observed a large number of hedonic price applications incorporating spatial econometrics (Yoo and Wagner, 2016), since it became the mainstream in regional science methodology over the last few decades. Further, point-referenced housing data are sometimes organized hierarchically (Bivand *et al.*, 2017; Dong, Harris, *et al.*, 2015; Dong and Harris, 2015). For example, individual houses are located in neighborhoods, which are nested within districts, and then are situated in even higher or more aggregated scales such as municipalities and regions. Such multilevel structures are reinforced if we investigate regional or national data, as higher spatial scales lead to more heterogeneous factors. Ignoring or mis-specifying the multilevel structure in a hedonic housing price model is likely to produce biased inference, and of course, incorrect estimates of the implicit price for sea views. As such, the question is further specified, what is the implicit price of sea views in the presence of both the spatial effect and the multilevel structure.

To account for the spatial effect and multilevel structure simultaneously, single-level regression seems inappropriate. I, therefore, develop a Bayesian

hierarchical spatial model (Arab *et al.*, 2017) containing two Leroux's conditional autoregressive (LCAR) components (Leroux *et al.*, 2000) (spatial multilevel models hereafter). The LCAR process is used for modelling spatial effects at both the apartment and the higher scale. More precisely, I consider a context, in which the apartment-level spatial effect is probably induced by omitting spatial covariates or misspecifying spatially delineated variables; further, the high-level spatial effect may be derived from spatial spillovers. The proposed spatial multilevel model allows spatial autocorrelation, heterogeneity and nested structures to be captured simultaneously, and hence offers precise estimates of the implicit price for sea views.

The rest of this chapter is organized as follows. Section 6.2 reviews previous hedonic analyses on the economic valuation of views, and examines spatial hedonic models with different spatial scales. Section 6.3 gives a detailed introduction of statistical models, including the LCAR model, the proposed spatial multilevel models and their space-time extensions. Section 6.4 introduces the procedure to generate the variable of interest, followed by the empirical investigation of the value of sea views. Section 6.5 presents the results and policy implications, and I draw conclusions in Section 6.6.

6.2. Literature review

6.2.1. What is in a view?

Researchers have recognized the economic benefits of views on property values for a long time.

Bourassa, Hoesli, and Peng (2003) believed that the benefit of views came from aesthetic characteristics in the environment. More importantly, it is necessary to distinguish the view from the accessibility to an amenity, since good views usually imply the closeness to an amenity. Bin, Crawford, *et al.* (2008) confirmed their and demonstrated that view variables could capture the rest residual amenity values, after controlling for the access to an amenity. Bourassa, Hoesli, and Sun (2004) also reviewed 35 earlier studies focusing on the economic benefit of views and concluded that views often induce positive impacts on property values. The magnitude of the impact ranges from 0.01% to 147% according to different types and definitions of views.

Benson *et al.* (1998) investigated different types and scope of views and found that both ocean and lake views can attract premiums. A poor partial ocean view attracted the lowest premium of 8%, while a full ocean view commanded the highest premium of 59%. However, mountain views did not affect property values significantly. Tse (2002) estimated the impact of a sea view on Hong Kong residential properties. Their results demonstrated that the sea view had significantly positive impacts on property values, irrespective of different model specifications. Sea views were found to increase the value by 6% in the selected model, while linear regression omitting spatial effects overestimated positive impacts. Jim and Chen (2009) studied the harbour view of Hong Kong and showed that the harbour view premium was around 3%. Moreover, the mountain view of Hong Kong was found to be insignificant. Sander and Polasky (2009) investigated how a single-family house's viewshed affects residential property values. They concluded that properties with larger viewsheds often had higher transaction prices compared with smaller viewsheds.

6

6.2.2. How to measure a view?

Many studies accessing views in housing prices fall into the hedonic pricing model. Thus, views act as an explanatory variable in hedonic pricing regression. In the existing literature, researchers have attempted different ways to generate view variables, including discrete measures in the early stages and recent continuous measures. For example, a view-related dummy variable equals 1 if a property has a view of an amenity and 0 otherwise. The dummy variable approach has clear limitations, since it neglects the quality of the view. To address this critique, multiply dummy variables or a polytomous variable is applied. In the study of Benson *et al.* (1998), four dummy variables (e.g., a full ocean view dummy, a superior partial ocean view, a good partial ocean view and a poor partial ocean view) are implemented to describe the quality of ocean view. Bourassa, Hoesli, and Sun (2004) applied a polytomous variable including narrow, medium and wide levels to measure the quality of lake views. However, the polytomous variable approach may experience subjective classification, and fail to properly capture the impact of views.

Recent advances in GIS allow us to carry out a viewshed analysis, which provides a precise, objective and continuous measurement of views compared with the previous discrete indicators. In simple, the viewshed analysis quantifies a two/three-dimensional view by examining whether each point on the surface is visible from an observation point or not. Advances in GIS have promoted viewshed analyse. Bin, Crawford, *et al.* (2008) created an ocean view variable relying on the two-dimensional viewshed analysis. Their result showed that homebuyers would like to pay \$995 for a one-degree increase in the view of the Atlantic Ocean. Hamilton and Morgan (2010) investigated the willingness to pay for ocean views using data from Florida coastal communities. The variable was also built on the two-dimensional viewshed analysis. It was found that a one-degree increase in the viewshed would command a \$1 228 premium. Hindsley *et al.* (2013) analyzed the impact of Ocean views on residential property prices using data from Pinellas County, Florida. To capture the impact of ocean view, they generated three continuous, ocean view indicators, e.g., total visibility, maximum visibility and mean visibility. All indicators were statistically significant and the authors concluded that ocean view would command premiums.

6.2.3. Spatial effects at different spatial scales

Standard hedonic regression, however, does not account for spatial effects. Regional scientists then apply spatial hedonic regression, which often offers satisfying results compared with standard hedonic regression. For example, the above-mentioned studies from Bin, Crawford, *et al.* (2008) and Hindsley *et al.* (2013) implemented spatial lag models. More recently, researchers are aware of the importance of the multilevel characteristic of spatial data, since different spatial scales probably induce different spatial interactions.

To the best of my knowledge, few researchers give solutions to account for both multilevel structures and spatial effects (Harris, 2019). Glaesener and Caruso (2015) assessed the effect of green space amenities and neighbourhood services on residential land prices in Luxembourg by means of a cross regressive multilevel model containing an additional spatial lag of independent variables (Halleck Vega and Elhorst, 2015). Dong and Harris (2015) suggested a more sophisticated, spatial multilevel model, namely the Hierar-

chical Spatial Autoregressive (HSAR) model. It is designed for the context, in which spatial spillovers appear in both high-level units and low-level units. Hence, the model incorporates two spatial weighting matrices, one is used for gauging the spatial spillover among all low-level units, another is used to assessing the spatial spillover among the high-level units. This model is used for investigating the impact of schools on property prices (Dong and Wu, 2016). Liu, Hu, *et al.* (2020) proposed a spatial multilevel model to investigate the spillover effects of ecological lands. Studies estimating the impact of views with a consideration of multilevel characteristics and spatial effects are even rare in the literature. Yamagata *et al.* (2016) recently proposed a so-called spatial multilevel additive regression model, which employs a variance component to gauge building-wise difference and a two-dimensional thin-plate spline smoother to capture spatial effects among buildings. Their results showed that only very nice ocean views attracted premiums in Yokohama city, Japan.

My study mainly assesses the apartment's sea views, specifically the view of the Mediterranean Sea. To achieve this goal, I employ a more specific and meaningful measure of sea views. In addition, this study is different from others regarding two significant features. First, the study area is a French region, where housing and regional characteristics show both spatial autocorrelation and heterogeneity. To address this issue, I account for both spatial effects and multilevel characteristics through the proposed spatial multilevel model by incorporating the LCAR component. Second, to fill in the gap that transaction periods may affect estimation accuracy within the multilevel modelling, spatiotemporal extensions are also investigated to tackle repeated cross-sectional housing data from 2006 to 2016.

6

6.3. Spatiotemporal modeling for spatial data with multilevel structures

6.3.1. Leroux's conditional autoregressive model

The key part of the proposed multilevel model (Schrödle *et al.*, 2011) is the so-called LCAR components (Leroux *et al.*, 2000). With the LCAR component, the proposed model can be fitted by Integrated Nested Laplace Approximation (INLA) approach (Rue, Martino, *et al.*, 2009), which is more computational

efficient than the Markov chain Monte Carlo (MCMC) method in fitting HSAR models. More importantly, the motivation of applying the two-level (LCAR) spatial model is different from the HSAR model, and the difference mainly comes from the application of the LCAR process or the simultaneous autoregressive process for low level units. In the HSAR model, the simultaneous autoregressive process for low level units aims to capture autocorrelation in the presence of a strong spatial trend. However, as Anselin and Lozano-Gracia stated,

“From a theoretical viewpoint, a spatial error specification is the more natural way to include spatial effects in a hedonic model. Unobserved neighbourhood effects will be shared by housing units in the same area and naturally lead to spatially correlated error terms”. (Anselin and Lozano-Gracia, 2009, pg. 1221)

I believe that spatial autocorrelation or heterogeneity in residuals due to omitted unobserved neighbourhood effects should be expected, and this leads to the application of LCAR components. On the other hand, the proposed model has an advantage over the Osland *et al.*'s (2016) conditional autoregressive model, that is, employing the Moran's I test for high-level random effects is not necessary. Because the LCAR component incorporates an additional parameter, which compromises between spatially structured and unstructured random effects. If spatially structured random effects dominate, the LCAR component will turn into an ICAR component. In contrast, if spatially unstructured random effects dominate, the LCAR component will become an IID component. In an extreme case, where spatially unstructured random effects dominate the two levels, the proposed model will become a standard multilevel model (Goldstein, 2011).

6.3.2. spatial multilevel models

The LCAR component is often considered as a prior specification in Bayesian paradigms and easily embedded in a hierarchical generalized linear model, whose three levels are:

- Level 1 – Data model: $[Y|\mu, \theta_Y]$
- Level 2 – Process model: $[\mu|\theta_\mu]$

- Level 3 – Parameter model: $[\theta_Y, \theta_\mu]$.

For example, apartments are located at n non-overlapping points $i = \{1, \dots, n\}$ and nested into m non-overlapping areas (corresponding to the high-level units) labeled as $j = \{1, \dots, m\}$, which are linked to a $n \times 1$ vector of responses $Y = (y_1, \dots, y_n)$. The structured additive predictor contains a $n \times p$ housing covariate matrix X , a $n \times q$ areal covariate matrix Z , an LCAR component for apartments $\Gamma_i = (\gamma_1, \dots, \gamma_n)$, and an LCAR component for areas $\Gamma_j = (\gamma_{1(1)}, \dots, \gamma_{m(n)})$. $m(n)$ denotes the area m to which apartment n belongs. For areal covariates, apartments within an area are assigned the same areal covariate value¹. It is worth noting that Γ_i is a n length random vector, while Γ_j is a m length vector with $m \ll n$, and Δ is a $n \times j$ block diagonal design matrix linking apartments to areas. The two LCAR components are used for gauging spatial autocorrelation at different levels that remain in the data after the covariate effects have been accounted for. Lastly, β and ξ correspond to $p \times 1$ and $q \times 1$ vectors of unknown coefficients, respectively.

Finally, the hierarchical model reads,

$$\begin{aligned}
 Y &\sim [Y|\mu, \theta_Y] \\
 g(\mu) &= \eta_i = X\beta + Z\xi + \Gamma_i + \Delta\Gamma_j \\
 \Gamma_i &\sim MVN\left(0, \sigma_i^2 Q_{\Gamma_i}^{-1}\right) \\
 \Gamma_j &\sim MVN\left(0, \sigma_j^2 Q_{\Gamma_j}^{-1}\right) \\
 \theta_\mu &= \{\lambda_i, \lambda_j, \sigma_i^2, \sigma_j^2\}.
 \end{aligned} \tag{6.1}$$

The data model $[Y|\mu, \theta_Y]$ depends on the process μ and other parameters θ_Y . Different likelihood functions (Poisson, Binomial, Gaussian, etc.) can be used for the data model. Regarding the process model, $g(\cdot)$ denotes the given link function, η_i is the corresponding additive linear predictor. In the empirical analysis of this chapter, since the log-transformed house prices are closely approximated by the normal distribution. I, therefore, have $\ln(Y) \sim [\ln(Y)|\mu, \theta_Y]$ and $\theta_Y = \{\sigma_\epsilon^2\}$ for the data model, and an identity link function is specified for the process model,

¹That is to say, Z is derived from two matrices, a $n \times j$ matrix illustrating the relation between apartments and areas, and a $j \times q$ matrix showing areal covariates.

$$\begin{aligned}
 E[\ln(Y)] &= \ln(Y) = g(\mu) \\
 g(\mu) &= \eta_i = X\beta + Z\xi + \Gamma_i + \Delta\Gamma_j.
 \end{aligned}
 \tag{6.2}$$

Here, $Q_{\Gamma_{(i)}}$ is the corresponding precision matrix for the LCAR component $\Gamma_{(i)}$,

$$Q_{\Gamma_{(i)}} = [\lambda W^* + (1 - \lambda) I_S]
 \tag{6.3}$$

where I_S denotes the identity matrix of dimension S^2 . W^* denotes the variant of a $S \times S$ spatial adjacency matrix W . The mixing parameter λ ($0 \leq \lambda \leq 1$) ensures the separation of spatially structured and unstructured variability.

Moreover, θ_μ contains all parameters that appeared in the process model. Lastly, I can impose prior distributions on the parameter model $[\theta_Y, \theta_\mu]$. Following the three levels, we obtain the posterior distribution of the process and parameters via the Bayesian theorem $[\mu, \theta|Y] \propto [Y|\mu, \theta_Y] \times [\mu|\theta_\mu] \times [\theta]$.

6.3.3. Space-time extensions

On the basis of the two-level spatial model, it is possible to extend this model to a space-time model.

A simple, computationally convenient and widely used space-time extension on random effects is the parametric model proposed by Bernardinelli *et al.* (1995). Their parametric time trends model consists of a global linear time trend term and a so-called differential time trend component,

$$(\phi + \delta_i) \times t
 \tag{6.4}$$

where $\phi \times t$ captures the global time trend. The differential time trend component $\delta_i \times t$ identifies an interaction between the unit i and period t , implying a unit-specific temporal regime (e.g., increase, decrease or stability), departing from the global time trend. For example, $\delta_i \times t < 0$ implies that the time trend of unit i is less steep than the global time trend.

It is possible to extend this unit-specific interaction to an interaction be-

²Here, s equals either m or n depending on the level.

tween area j and period t , implying an area-specific time trend $\delta_{j(i)} \times t$ ($\delta_{j(i)} = (\delta_{1(1)}, \dots, \delta_{m(n)})$) to depart from the global time trend for all units $\phi \times t$. Thus, the linear predictor is written as,

$$g(\mu) = \eta_{it} = X_{it}\beta + Z_{it}\xi + \Gamma_i + \Delta\Gamma_j + (\phi + \delta_{j(i)}) \times t \quad (6.5)$$

It is important to note that all random effects components $\Gamma_i, \Gamma_j, \delta_i, \delta_{j(i)}$ must be centered around zero to avoid identifiability issues of the models (Goicoa *et al.*, 2018).

6.4. Empirical analysis

6.4.1. Generating sea view Index

As previously stated, more and more studies employ continuous metrics to measure view. The most common are variants of viewable areas or view angles based on a radius. For my study, the sea view index relies on an individual property's viewable areas adapted with minor modifications from Hamilton and Morgan (2010). I detail the procedure to create viewable areas into three steps that involved:

- (i) Positioning a virtual "observer". To reduce heavy computational loads and to facilitate following calculation, the observer point for each apartment is set to the centroid of the parcel where an apartment locates. I believe that the centroid of a parcel is a good approximation of apartment location, because the parcel is the smallest unit of land and usually corresponds to one apartment building in Corsica. Moreover, to determine the observer's height, I first calculate the elevation at the centroid of a parcel. Then, a point 1.8-meters off the ground is added.
- (ii) Creating apartments' potential view areas at a 10-km distance from the "observer" point. It is important to note that the view area should not exceed half of the circle given the "observer" point and the 10-km radius.
- (iii) Summarizing the field, where the view area overlaps with the Mediterranean Sea. The unit of the view area is in square meters.

Finally, for the 6 377 sea view scores, 8.2% of the scores are zeros, 0.9% of the scores are duplicate with non-zero values.

6.4.2. Measuring beach access

In the hedonic analysis, another crucial factor relating to the sea view index is the accessibility of the nearest beach, e.g., actual access distances from each property to the nearest beach, to enable the implicit value of living close to the beach to be quantified (Bin and Kruse, 2006). Previous studies typically used the Euclidean distance from each property to the centroid of the nearest beach (Parsons and Noailly, 2004; Pompe, 2008). This simple measurement could be a good proxy of access, but it fails to truly account for the actual access to the nearest beach. Because we believe that beaches are likely to be accessed using road networks, rather than cutting across other places in a roughly as-the-crow-flies manner. For example, many sandy beaches in Corsica are located on the other side of hills, access is not always easy. People have to follow signed beach access points and drive along crooked roads.

I, therefore, argue that the road distance to the nearest beach from each property should offer a more accurate measurement of beach access than the Euclidean distance. Hence, the road distance to each beach from properties is calculated on the basis of local road networks in a GIS environment. Based on the average driving speed on different types of roads (average driving speeds are identified by a vehicle GPS database provided by Navteq), I further give the driving time to the nearest beach from each apartment. These data with sea view information will be joined back to apartment sales data.

6.4.3. Data

Housing transaction data are obtained from the "PERVAL" dataset used in the Chapter 4, and the data associated with "EPCI" are an aggregation of the Corsican county-level data used in the Chapter 5. After a combination, the final dataset is made up of 6 377 sales of apartments.

Actual sales prices are available for all apartments in the final dataset with the mean €149 685 and median €138 042 after adjusting for inflation (the base year is 2006). Further, apartment sale prices range from €57 446 to €325 432. This variable is transformed into a logarithmic scale and serves as

the dependent variable.

Housing structure covariates can be found in the dataset directly, while neighbourhood covariates and view variables for each apartment are calculated via a geographic information system (GIS). All independent variables are summarized in Table 6.1 along with the expected impact of each variable on apartment sale prices, and Table 6.2 provides the corresponding descriptive statistics. Concerning accessibility variables, as access to certain amenities probably affects housing sale prices, road distances from each apartment to the closest pharmacy, hospital, primary school, and beach are calculated.

In addition to the individual housing data, contextual data are collected from different sources depending on variables. Three contextual covariates are the population density of each "EPCI"³, the averaged property tax ("taxe sur le foncier bâti"⁴) in an "EPCI", and physical landscape counts within each "EPCI" offered by Research Center "LISA".

Housing sales data, GIS-based view/accessibility data, and contextual data are merged together and we finally obtain a sample of 6 377 apartments for use in estimation. The variance inflation factor (VIF) scores for all covariates are calculated to examine the potential problem of multicollinearity. There is no evidence of such a problem. Most variance inflation factor scores are around 1.3, below the normal threshold of 10.

6

6.4.4. Econometric strategies

I begin with the classical hedonic regression (Eq. 6.6), labelled as Model 0,

$$\begin{aligned}
 \eta_{it} = & \alpha + f(\text{Surf}) + \beta_2\text{Sto} + \beta_3\text{Bat} + \beta_4\text{Par} + \beta_5\text{Typ} + \beta_6\text{Contr} \\
 & + \beta_7\text{Pha} + \beta_8\text{Hosp} + \beta_9\text{PSch} + \beta_{10}\text{Bea} \\
 & + \beta_{11}\text{Sea} \\
 & + \xi_1\text{PopDen} + \xi_2\text{CTax} + \xi_3\text{Nat} \\
 & + \epsilon
 \end{aligned}
 \tag{6.6}$$

Subsequently, I examine spatial effects and multilevel structures in data.

³Source: <https://www.insee.fr/fr/information/2008354>

⁴Source: <https://www.service-public.fr/particuliers/vosdroits/F59>

Table 6.1: Definitions for study variables and expected relationship to apartment sale values.

Variable name	Definition	Expected signs
Structural variables		
Price	Sales price adjusted to 2006 (Euros)	
Surf	Total area (m^2)	Positive
Sto	Floor	Positive
Bat	Number of bathrooms	Positive
Par	Number of parkings	Positive
Typ	Types of apartments. Base: Apartment	Studio Negative; Duplex Positive
Contr	Construction periods. Baseline: 1850-1913. 1914-1947;1948-1969;1970-1980;1981-1991 1992-2000;2001-2010;2011-2017	Positive
Accessibility variables		
Pha	Road distance to closest pharmacy in km	Negative*
Hosp	Road distance to closest health facilities in km	Negative*
PSch	Road distance to closest primary school in km	Negative*
Bea	Driving time to closest beach in minute	Negative*
View variable		
Sea	Sea view index from 0 to 100	Positive
Contextual variables		
PopDen	Population density in each "EPCI" ($people/km^2$)	Negative
Tax	Averaged property tax rate in each "EPCI"	Negative
Nat	Physical landscape counts in each "EPCI"	Positive

* For these distance-based variables, a negative expected relationship means that home buyers would like to pay more to live near an amenity.

Table 6.2: Descriptive statistics for quantitative variables.

Variable name	Mean	SD*	Min	Pct(25)	Pct(75)	Max
Structural variables						
Price	149,685.400	58,602.020	57,446	103,829.9	185,952.0	325,432
Surf	59.537	21.983	13	43	73	197
Sto	1.863	1.733	-3	1	3	12
Bat	1.056	0.265	0	1	1	3
Par	0.821	0.717	0	0	1	8
Accessibility variables						
Pha	1.520	2.237	0.000	0.396	1.448	38.662
Hosp	10.674	12.250	0.051	1.697	16.612	72.244
PSch	27.702	37.821	0.063	2.753	33.074	182.492
Bea	5.472	6.452	0.001	2.360	6.626	67.706
View variable						
Sea	12.060	11.303	0.000	1.670	20.801	47.907
Contextual variables						
Popden	273.249	307.849	6.888	21.946	307.539	828.975
CTax	0.115	0.026	0.066	0.094	0.140	0.156
Nat	11.842	9.059	1	6	14	40

* Standard deviation

As indicated in Chapter 1, Corsica owns rural, semi-urban and urban landscapes (Vogiatzakis *et al.*, 2008), and public services, infrastructures and human activities are mainly concentrated in the urban area of Corsica. Cities and

peripheral counties, therefore, consist of a cluster in terms of sharing public or private services, e.g., schools, hospitals, gyms, parks, shopping malls, and this cluster probably produces delimitation to others. Hence, in addition to attributes associated with an apartment, the factors associated with clusters (hereafter contextual factors) probably affect the apartment price as well.

In this context, I take the so-called "Etablissement public de cooperation intercommunale" ("EPCI") as the cluster. There are 19 "EPCI" zones in Corsica. The spatial configuration of Corsica "EPCI" zones are displayed in Figure 6.1 (Table 7.3 in Appendix 7.5 shows the corresponding names). According to an official report (Nicolai, 2019), in 2016, there were 130 200 active employees in Corsica, and most of them commuted within the EPCI where they lived. Since apartments are nested into "EPCIs", spatial dependence caused by omitting unobserved neighbourhood effects may occur among apartments that are spatially close within an "EPCI". Further, spatial dependence between "EPCIs" seems to be expected. More precisely, "EPCIs" located further apart should be less similar than "EPCIs" located closer to each other. For example, we may observe the transition from a highly urbanized area to semi-urban and even rural areas, if we move from an "EPCI" to its neighbours and then further "EPCIs". As such, LCAR components for the different levels are incorporated in Eq. 6.6, leading to the same specification of Eq. 6.2 (labelled as Model 1).

Lastly, I investigate the impact of sale periods on the basis of the spatial effect and multilevel structures. In the hedonic pricing literature, the temporal dimension has attracted little attention. In some hedonic applications, sales data are collected over short periods, e.g., one year or two years, and researchers prefer to collapse the few periods to a single time point. In the case of long periods, some scholars intend to capture temporal heterogeneity via time dummies. Following the same logic, I develop Model 2. On the other hand, other researchers attempt to capture temporal trends via a linear or spline function. I extend a linear trend function by incorporating additional "EPCI"-specific and unit-specific components to gauge both temporal trend and spatiotemporal variation, like the specifications shown in Section 6.3.3. These models are labelled as Model 3 and Model 4 respectively. A summary of model specifications are shown in Table 6.3.

Note that different prior distributions are assigned to the hyperparameters

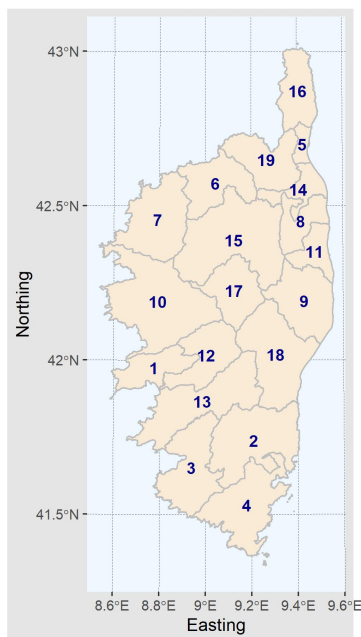


Figure 6.1: Sale counts within each “EPCI”.

Table 6.3: Summary of random components in the spatial multilevel models

Model Identifier	Spatial	Temporal
M0	No control	No control
M1	2 LCAR components	No control
M2	2 LCAR components	Time dummy
M3	2 LCAR components	A Linear temporal trend with the “EPCI”-specific component
M4	2 LCAR components	A Linear temporal trend with the unit-specific component

below:

$$\lambda_i, \lambda_j \sim \text{Uniform}(0, 1),$$

$$\log\left(\frac{1}{\sigma_{\Gamma_i}^2}\right), \log\left(\frac{1}{\sigma_{\Gamma_j}^2}\right), \log\left(\frac{1}{\sigma_{\delta_{j(i)}}^2}\right), \log\left(\frac{1}{\sigma_{\delta_i}^2}\right) \sim \text{logGamma}(1, 0.00005).$$

(6.7)

Furthermore, the high-level spatial adjacency matrix W^E is built on the queen contiguity of the 19 "EPCIs". In contrast, an element in the low-level spatial adjacency matrix W^A is specified as:

$$w_{ij}^A = \begin{cases} 1 & d_{ij} < d \\ 0 & \text{otherwise,} \end{cases} \quad (6.8)$$

where d is the distance threshold. Considering the Corsican context (the 95%CI for the spatial range is (1.291; 1.698)), this threshold is set to 2.5 km. d_{ij} is the Euclidian distance between apartment i and j . Figure 6.2 and Figure 6.3 show the high level and low level spatial adjacency matrices, respectively.

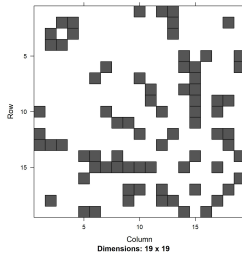


Figure 6.2: Spatial adjacency matrix for "EPCIs", rows and columns represent "EPCI" units.

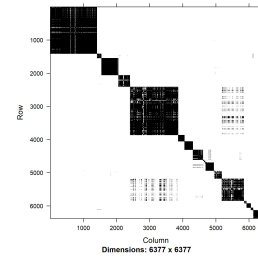


Figure 6.3: Spatial adjacency matrix for apartments, rows and columns identify apartment units.

6

6.5. Results

6.5.1. Model assessment

Model comparison results for the four models are provided in Table 6.4.

Table 6.4: Model assessment via DIC and LCPO

	DIC	LCPO
M0	-585.683	-0.0449
M1	-2,314.626	-0.1006
M2	-2,726.109	-0.1073
M3	-2,616.119	-0.1047
M4	-1,894.402	-0.1053

As shown in Table 6.4, Model 2 produces the best model fit according to both the DIC (Spiegelhalter *et al.*, 2002) and LCPO (Pettit, 1990) scores, compared with these of the nonspatial model (Model 0), spatial model (Model 1), and the model with spatial and linear time trend components (Model 3 & 4). A better model fit of Model 1 against Model 0 indicates the benefit of account for spatial effects, and a better model fit of Model 2 against Model 3 and 4 indicates that the dummy variable approach captures more temporal variation compared with the linear temporal trend approach.

6.5.2. Interpreting covariate effects

The estimation results from Model 2 are provided in Table 6.5.

Most structural variables are statistically significant, and their signs are as expected. Several construction periods, including the period 1914-1947, 1948-1969, 1970-1980 and 1981-1991 are not significant. In contrast, the price of apartments built over the period 1992-2000, 2001-2010, 2011-2017 has a significant increase, compared with the apartment built over the period 1850-1913. Further, the signs for the apartment type variable are as expected. This result indicates that studio apartments should have lower values compared with normal apartments, while duplex apartments should have higher values, and the associated marginal implicit prices are -0.106 (95%CI, $-0.135; -0.077$) and 0.047 (95%CI, $0.020; 0.074$), respectively.

These results clearly show that most accessibility covariates influence apartment sale prices, and the corresponding signs are as expected. In general, increasing the distance to the nearest hospital decreases home sale prices, and increasing the distance to the nearest hospital by 1 km produces a 0.3% (95%CI, $-0.005; -0.001$) decrease in apartment sale prices. Increasing the distance to the nearest primary school by 1 km also suggests an apartment sale price decrease of 0.3% (95%CI, $-0.005; -0.002$). In contrast, we do not observe a significant impact of the distance from the nearest pharmacy on apartment prices.

Increasing the time to the nearest beach by 1 minute produces a -1.3% (95%CI, $-0.015; -0.011$) decrease in apartment sale prices. The coefficient for the variable of interest, the sea view, is 0.005 (95%CI, $0.004; 0.005$). It means that increasing sea view by 1 m^2 should steam a 0.5% increase in

Table 6.5: Posterior estimates of the covariate effects

	Mean	St.Dev	0.025quant	0.975quant
Structural variables				
Intercept*	11.754	0.104	11.555	11.961
Sto*	0.013	0.002	0.010	0.017
Bat*	0.069	0.012	0.046	0.092
Par*	0.044	0.005	0.034	0.053
Contr 1914-1947	0.075	0.046	-0.014	0.165
Contr 1948-1969	-0.006	0.041	-0.086	0.074
Contr 1970-1980	0.045	0.041	-0.034	0.125
Contr 1981-1991	0.079	0.041	-0.0004	0.159
Contr 1992-2000*	0.184	0.043	0.099	0.269
Contr 2001-2010*	0.230	0.040	0.151	0.309
Contr 2011-2017*	0.214	0.040	0.135	0.293
Duplex*	0.047	0.014	0.020	0.074
Studio*	-0.106	0.015	-0.135	-0.077
σ_{Surf}^2 *	0.00098	0.00026	0.00056	0.00158
Accessibility variables				
Fha*	0.001	0.002	-0.002	0.004
Hosp*	-0.003	0.001	-0.005	-0.001
PSch*	-0.003	0.001	-0.005	-0.002
Bea*	-0.013	0.001	-0.015	-0.011
View variable				
Sea*	0.005	0.0004	0.004	0.005
Contextual variables				
PopDen	-0.0003	0.0002	-0.001	0.0001
CTax	-0.458	0.665	-1.766	0.846
Nat*	0.013	0.005	0.004	0.023
Time dummies				
2007*	0.088	0.017	0.054	0.121
2008*	0.133	0.017	0.099	0.166
2009*	0.076	0.017	0.042	0.110
2010*	0.063	0.018	0.029	0.098
2011*	0.133	0.019	0.096	0.170
2012*	0.129	0.018	0.093	0.165
2013*	0.122	0.018	0.085	0.158
2014*	0.113	0.019	0.076	0.149
2015*	0.109	0.019	0.071	0.147
2016*	0.144	0.020	0.104	0.184

* statistical significance

apartment sale prices. This illustrates the preference of homebuyers for apartments with large sea views, even though the magnitude is small. Since both the time to beaches and the sea view have a significant impact on apartment prices, I confirm Bin, Crawford, *et al.*'s (2008) point, where view variables

could capture additional amenity values, even though researchers have already controlled for the distance to that amenity.

Figure 6.4 shows the nonlinear relationship between apartment surfaces and sale prices, and we observe an overall increasing trend. This finding confirms that apartment surfaces have a positive impact on sale prices. Moreover, the shape of the curve is similar to that in Chapter 4.

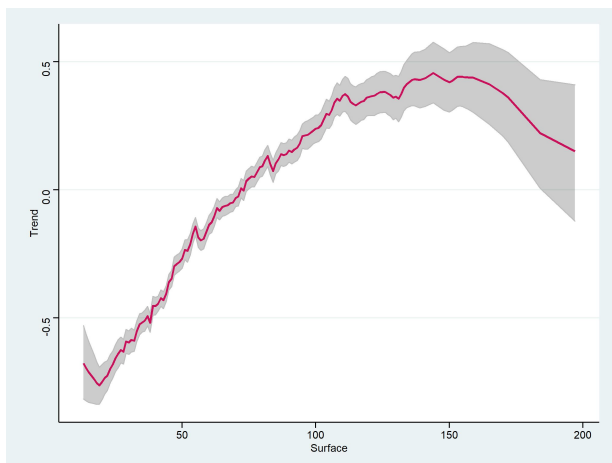


Figure 6.4: Mean (the red line) and a 95% credible interval (the grey area) of the Surface.

For the contextual covariates, the coefficient for the population density is negative, but this variable does not significantly impact apartment sales prices, indicating that population density is not a particular consideration for homebuyers. The averaged council tax rate within an “EPCI” has a large negative coefficient, but it is not statistically significant either. On the other hand, the physical landscape counts have a significant impact on apartment prices. The estimated coefficient is 0.013 (95%CI, 0.004; 0.023), meaning that the marginal implicit price for increasing the physical landscape count by 1 evaluated at the mean apartment sale price indicates a 1.3% price increase. This illustrates the preference of homebuyers for physical landscapes, e.g., lakes, mountains, alpine rocks, estuaries.

Lastly, all time dummies are statistically significant and positive, compared with the baseline year (2006). The evolution of their coefficients is shown in Figure 6.5. Time dummies show that apartment prices quickly reached their

peak (the year 2008) and then experienced a decrease till 2010. After the rebound in 2010, apartment prices have tended to stable since 2011.

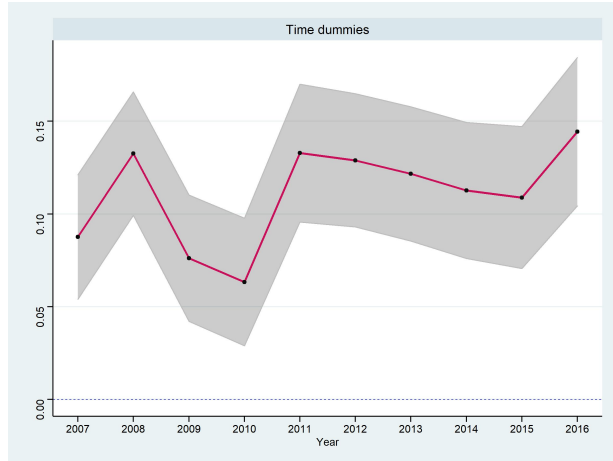


Figure 6.5: The time dummies and their 95% credible intervals. The dash line shows the baseline.

6

6.5.3. Interpreting random effects and visualization

The estimates of random effect components are shown in Table 6.6.

Table 6.6: Posterior estimates of the random effects

	Mean	St.Dev	0.025 quantile	0.975 quantile
σ_ϵ^2	0.0240	0.0025	0.0192	0.0290
λ_i	0.0026	0.0005	0.0017	0.0038
σ_i^2	0.0421	0.0033	0.0361	0.0489
λ_j	0.2795	0.2171	0.0186	0.7947
σ_j^2	0.0545	0.0311	0.0166	0.1350

I initially calculate the proportion of marginal variance explained by each component, given by

$$p_\varphi = \sigma_\varphi^2 / (\sigma_j^2 + \sigma_i^2 + \sigma_\epsilon^2) \times 100\%, \quad \varphi = \{j, i, \epsilon\}. \tag{6.9}$$

Here, the corresponding proportions are 45.19%, 34.91% and 19.90% respectively. This indicates that the spatial pattern at the "EPCI" level explains

a huge part of variability, then the spatial pattern at the apartment level also explains a considerable part of variability. Lastly, the Gaussian residual component explains the smallest part of variability.

In addition to the variance components, the spatial mixing parameters at both the apartment level and the "EPCI" level are statistically significant at the 95% credible level. That is to say, the proposed spatial multilevel hedonic model, rather than a standard nonspatial multilevel model, is the most effective model. The spatial mixing parameter at the "EPCI" level equals 0.2795 (95%CI, 0.0186; 0.7947), confirming that a considerable part of variability is due to spatial autocorrelation. Regarding σ_j^2 and λ_j jointly, I can conclude that apartment prices are affected by spatial effects from their own "EPCI" as well as by effects from surrounding "EPCIs".

The spatial mixing parameter at the apartment level is also significant, but its magnitude is fairly low 0.0026 (95%CI, 0.0017; 0.0038). The low value indicates that heterogeneity dominates variability among apartments. These results conform to the motivation of applying the proposed spatial multilevel models.

Figure 6.6 maps the estimated posterior means of "EPCI" level random effects from Model 2. Lighter colours represent stronger positive effects and darker colours indicate stronger negative effects. Overall, there is a clear spatial pattern: "EPCIs" with strong negative effects cluster around the northeastern coast (zone 16, 14, 8), and "EPCIs" with strong positive effects are located on the southeastern coast (zones 2, 4). These may be a reflection of the significant and positive spatial mixing parameter. Further, the strongest positive effects are found in zone 17, alongside the "EPCI" (zone 10) with a strong negative effect. I believe that this is a reflection of heterogeneity. More precisely, the strongest positive random effects found in zone 17 may be related to the presence of the University of Corsica. The strong negative random effects of the northeastern cluster are likely associated with counties dominated by agricultural and industrial activities. Concerning the southeastern cluster, the high positive random effects are probably related to cities (Porto Vecchio, Bonifacio) and two commercial ports. There are many residential communities in this cluster, along with many commercial facilities such as shopping malls and many service-related job opportunities.

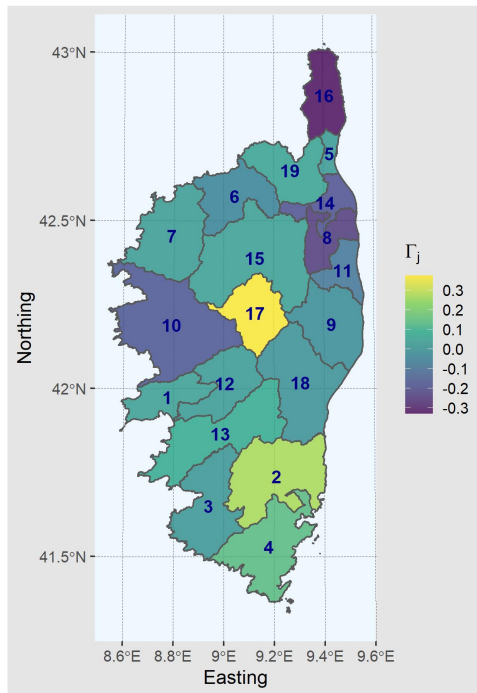


Figure 6.6: The “EPCI” level random effects from Model 2

6.5.4. Policy implications

As the regional council (“Collectivité Territoriale de Corse”) requires local authorities to release new zoning plans and regulations soon, planners or policy-makers will make important decisions about developing or preserving different types of lands in each county. These decisions will determine future land uses. This, in turn, will influence the availability of environmental amenities both locally and regionally.

To formulate relevant planning, a question should be addressed primarily, “what is the implicit price of each amenity?” Once the accurate information on residents’ preferences is estimated, the planner can formulate appropriate planning and policies to meet the community expectations and to minimize the negative economic, social, and environmental impacts associated with land use changes. In this study, the estimates serve as a starting point for the investigation of residents’ preference and expectation for coastal views,

beach access, etc.

Further, this study suggests that sea views and beach access jointly should be considered as key elements in future planning. Since the beach access significantly affects apartment sale prices, a potential policy intervention is that local authorities should improve the accessibility, e.g., by providing better transportation infrastructure or public transport to beaches. This intervention seems more attractive for non-coastal counties than for coastal counties. Concerning the sea view, planners should preserve this amenity, for example, levying council tax in coastal counties and limiting building height. Policy-makers could enhance the benefit of the sea view by releasing amenity-led growth strategies. Local authorities should also provide more public goods like hospitals and schools.

Lastly, my findings also suggest the regional council to establish differentiated planning depending on the "EPCI". For example, zone 17 has the strongest positive random effects, which probably relate to the housing demand from college students. The regional authority should conduct surveys to collect detailed information on this area, so that planners could decide whether to increase land supply or not, e.g., to convert agricultural or industrial lands for residential uses.

6.6. Concluding remarks

Employing availability indicators, including housing structural variables, accessibility variables, view variables (sea view index) and contextual variables, and a spatial multilevel hedonic model, in which spatial effects and nested housing data structures are considered, I mainly quantify the impact of sea view on apartment prices in Corsica, France. The inclusion of both the sea view indicator and the accessibility measure to the nearest beach overcomes the possible deviations in traditional hedonic regression, which accounts for either an accessibility indicator or a visibility indicator. Further, multilevel modelling can avoid misalignment, referring to mismatches when transferring the relationship from contextual variables to normal variables and vice versa. In the end, incorporating LCAR components to multilevel models addresses spatial effects induced by unobserved neighbourhood effects among apartments and

geographical closeness of “EPCIs”.

The first basic problem I focused on in this study is building a sea view index, which can quantify sea views objectively and appropriately. In doing so, I generate a continuous index based on the viewshed analysis. This index can enhance the judgement of planners, as compared with the traditional use of subjective rules of thumb.

Subsequently, I apply this index with other covariates to assess the implicit price of sea views via the hedonic price method. The results, like those of previous studies, clearly indicate that a preference exists for living near prime schools, pharmacies. More importantly, I find that the sea view has a small but significant impact on housing prices. My results also confirm that homebuyers are willing to pay a premium for living close to coasts, and that premium consists of both accessibility values and the value of sea views. As such, the value calculated for either the beach access or the sea view does not represent the real values for living near to coasts.

Lastly, I focus on the rationality of the proposed spatial multilevel hedonic model. The proportion of each marginal variance component clearly shows that the nested structure in housing data is important. The estimates of mixing parameters at each geographic scale demonstrate that spatial effects play a crucial role in housing prices as well. These results indicate that any regression analysis conducted on a single level that ignores spatial effects is not appropriate, especially in handling point-referenced data with nested structures.

This study, however, has several limitations. First, the sea view index can be built in a more sophisticated way. The index used in this study does not account for apartment height and window location. It is necessary to mitigate these assumptions with the help of available Light Detection and Ranging (LIDAR) data. Moreover, I do not consider the factors affecting human perception, such as trees, obscured properties, angle of depression. With the development of computer graphics and virtual reality tools, view indices should be quantified in more detail in future studies, taking into account those factors. Lastly, this study does not investigate the impact of different prior distributions on posterior estimates for the spatial multilevel model, and this should be explored in future studies.

7

Summary and Future works

This thesis provides an investigation of the housing market in Corsica, France. The present work primarily focuses on the determinants, especially how the spatial and temporal dimensions (i.e., location and time) affect different aspects of the Corsican housing market. More precisely, I deal with various issues associated with the Corsican housing market. These issues include identifying housing price determinants and making predictions, revealing spatial and temporal patterns of second home rates among Corsican counties, and recognizing the implicit prices of sea views in the presence of nested spatial data structures.

The hedonic price method, the so-called adjacent effect and ripple effect offer the economic underpinnings for analyzing such complex issues. The use of the "PERVAL" dataset and other open source datasets are also crucial factors in addressing these issues. More importantly, the corresponding empirical analyses are based on the Bayesian hierarchical modeling, which has advantages over classical spatial econometrics in terms of handling data indexed both spatially and temporally. In particular, I specify Bayesian hierarchical

models that contain different spatial, temporal and spatiotemporal components depending on spatial data types. Our findings indicate that housing location and transaction time are the key elements in determining housing prices, the ripple effects do exist among Corsican counties and consumers clearly show their preference for sea views, but the magnitude is small.

Section 7.1 attempts to pull together the findings from the empirical chapters of this thesis. I indicate the general contribution of these chapters to the existing literature in Section 7.2. Section 7.3 presents the strengths and weaknesses of each model. Section 7.4 sets out possible directions for future research.

7.1. Main Findings

The first empirical chapter (Chapter 4), "*Time, space and hedonic prediction accuracy: evidence from Corsican apartment market*"¹ mainly evaluates three Bayesian hierarchical models with different spatial, temporal and spatiotemporal components. I attempt to select the most effective model that satisfies the goodness of fit, the predictive power and the computational costs. Our results demonstrated that the modified hierarchical autoregressive model (Sahu and Bakar, 2012) is the most effective. In particular, this model involves a space-time random effect specified by an AR(1) structure with spatially correlated but temporally independent innovations. Based on this particular component, spatial and temporal autocorrelation across housing units is gauged. In other words, an observation directly borrows strength from its neighbours in space and time.

Moreover, I find that most housing structural attributes and accessibility attributes significantly affect housing prices. More importantly, the spatiotemporal patterns explain a large portion of the variance. I also draw a map showing housing price gradients. I identify the hot spots and cold spots in terms of housing prices on the map. Some hot spots may lead to price appreciation by up to 82%, whereas some cold spots may result in a 45% price reduction. Regarding the time dimension, the location of these hot spots and cold spots has not changed dramatically between the first quarter of 2006 and

¹The extracts of Chapter 4 are published in *The Annals of Regional Science*, 64, 367–388 (2020), <https://doi.org/10.1007/s00168-019-00967-2>.

the fourth quarter of 2017.

The second empirical chapter (Chapter 5), "*Unveiling spatial and temporal patterns of second home dynamics: a Bayesian spatiotemporal analysis for a Mediterranean island*"² demonstrates that the so-called ripple effect exists in the Corsican second home rates. Promoted by Meen (1996), the ripple effect is initially used to describing regional/sub-regional housing prices comovement in the United Kingdom. This comovement happens on both the spatial and temporal scales. I, therefore, intend to study the comovement of Corsican second home rates at the county level.

In Corsica, second homes have been considered a major development issue, but few empirical studies discuss this issue at an island scale (Maupertuis *et al.*, 2017). To study the comovement of second home rates, I look into the spatial and temporal dimensions of second home rate changes among Corsican counties. Borrowing ideas from small area estimation, I propose a Bayesian hierarchical model built on the Besag-York-Mollie (BYM) model and its variants (i.e. LCAR model (Leroux *et al.*, 2000)).

Through model comparison criteria, I demonstrate that the model involving spatial, temporal components (i.e., the LCAR component and the first-order random walk component) and a spatiotemporal interaction term produces the most precise estimates and the best forecasts compared with other models involving different combinations of spatial and temporal components. This also implies the existence of the comovement of the second home rates in Corsica. In addition to the spatial and temporal components, the spatiotemporal component is used for capturing the deviation beyond the main spatial and temporal trends.

To interpret the spatial comovement, I also identify hot spots and cold spots, where specific counties exhibit higher second home rates or lower second home rates respectively. For instance, the Bastia area (e.g. Biguglia, Furiani, Bastia) is a significant cold spot, whereas some counties in the Balagne area (e.g., Lumio) and the Southern area (e.g., Lecci) belong to hot spots. The temporal comovement reveals a gradual increase in the second home rate over the past 11 years. Furthermore, the study found that physical landscape counts, coastal counties, and mountainous counties are positively

²The extracts of Chapter 5 have been required major revision in Spatial Economic Analysis.

associated with second home rates, while cultural landscape counts, number of households and interest rates show a negative association.

The results of this study should be useful for policy makers. They can intervene in hot spots to curb the high second home rate and to prevent second home buyers spilling over to adjacent counties. As a result, the inhabitants' needs are met.

The third empirical chapter (Chapter 6), "*The value of sea views in a Mediterranean island: Hedonic analysis with the spatial multilevel model*" mainly assesses the implicit price of sea views on Corsica apartment prices in the presence of nested spatial data structures. This study is based on the hedonic price method, as it is the main approach for estimating the implicit price of environmental attributes.

As a scenic amenity, views are usually perceived through human eyes and are difficult to measure and quantify for valuation purposes. After reviewing the literature, I find that the viewshed analysis implemented on geographic information systems provides a feasible way to generate an objective, continuous measurement of views. Then, estimating sea views relies on the hedonic price method. However, I notice that the nested data structure may affect the results if I account for additional contextual variables. To address this issue, I propose a spatial multilevel model, where spatial effects and the nested structure of data are considered simultaneously. Furthermore, this model is extended by including a temporal component to account for temporal dynamics.

Our findings demonstrate that after controlling for all covariates, the sea view has a small but statistically significant impact on apartment prices. That is to say, home buyers have a higher marginal willingness to pay for larger sea views, even if the magnitude is small. More importantly, I find that proximity to beaches also produces positive impacts on apartment sale prices as well. We may conclude that the implicit price of living near to beaches consists of two parts, the implicit price of sea views and the implicit price of proximity to beaches. Further, most apartment characteristics are statistically significant, but variables associated with "EPCIs" are not. I also find that strong spatial spillovers occur at the EPCI scale, but unobserved heterogeneity dominates at the apartment level.

The results of this chapter provide valuable information for planners. Based on the estimated implicit prices, planners can prepare appropriate planning to meet the expectations of the inhabitants and to minimize the negative impacts on the economy, society, and environment during the land use changes.

7.2. Main contributions

In addition to the above-mentioned results, this thesis contributes to the existing literature on two broad aspects, the methodological proposition and the empirical findings.

Methodologically, I apply several ad-hoc geoaddivitive models combined with the conditional autoregressive and Gaussian processes, and the Bayesian approach to analyze repeated cross-sectional and panel data that exhibit clear spatial and temporal dimensions. As alternatives to classical spatial econometrics models, the proposed models are used for analyzing housing markets and land use changes. To my knowledge, these models have not been considered in the existing literature.

It is common knowledge that in a residential property market, if buyers face many unobservable local factors, public amenities and unreasonable asking prices of a house from a property agency, they often choose to take a look at the prices of nearby properties. As such, they can obtain some information and reduce associated uncertainty. The application of geoaddivitive models with the conditional autoregressive process and the Gaussian process preserves the idea of spatial spillovers, but the spillovers are shown in another manner. Both the conditional autoregressive process and the Gaussian process borrow strength from the neighbours of an observation.

Lastly, spatial econometrics does provide a powerful tool for examining spatial patterns across different properties. There has been a significant amount of research on spatial effects over the last few decades, but temporal dimensions have received limited attention in empirical analyses of housing. Since classical spatial econometrics may not be flexible enough to handle spatiotemporal data, I turn to geoaddivitive models, which can gauge spatial and temporal dynamics in different ways. I also highlight that the integrated nested Laplace approximation (INLA) technique is used for model fitting and

inference throughout the thesis.

Empirically, I focus on housing prices and second home rate changes in Corsica (a beautiful island in the Mediterranean), France. The economic growth in the last decade has given rise to concerns about growing housing prices and the number of second homes. Residents have experienced high housing prices and land use conflicts. Further, few researchers focused on the local housing market. Therefore, the inference and the prediction of housing prices and second home dynamics are timely and needed. They are also imperative for planners and policymakers to develop sustainable development strategies.

Moreover, the three empirical chapters of this thesis (from Chapter 4 to Chapter 6) investigate the determinants of housing prices, the spatial pattern and temporal evolution of second home rate changes, and the implicit price of sea views on apartment prices. The research questions in Chapter 4 and Chapter 6 are addressed via the hedonic price method, since I attempt to investigate the determinants of housing prices and to evaluate non-market goods. The adjacent effect is also an important economic underpinning for these two studies. Our findings, such as location being an important determinant, the sea view and the accessibility to beaches having significant impacts, can provide useful information for planners and policymakers. Regarding Chapter 5, I intend to investigate whether the comovement of second home rates among all Corsican counties exists. This is done by analyzing the spatial and temporal patterns of second home rates. Our results clearly show the existence of the comovement. In addition, I identify the hot spots and cold spots of second home rates, meaning that additional interventions need to be applied to these areas.

7

7.3. Strengths and Weaknesses of Each Model

The proposed model in Chapter 4 has several advantages over classical spatial econometrics models. For example, it does not require a pre-defined spatial adjacency matrix, but it calculates distance (known as “spatial ranges”) where spatial dependence diminishes to the threshold automatically. Moreover, it offers a more flexible way to handle repeated cross-sectional data compared

with the spatiotemporal autoregressive model (Dubé and Legros, 2013a) and the pseudo-panel model (Baltagi *et al.*, 2015).

The spatiotemporal Bayesian binomial regression described in Chapter 5 is a new approach to handling second home rate changes. The model accounts for both potential covariates and spatiotemporal trends, which outperforms classical linear models and original spatiotemporal Bayesian binomial models that do not consider covariate effects. Further, this model contains a Leroux's conditional autoregressive component relying on spatial adjacency matrices. The selection of an appropriate spatial adjacency matrix is still challenging.

The multilevel spatial model in Chapter 6 is also a new approach to handle point-referenced housing data. Since the proposed model is different from the spatial multilevel from Dong and Harris (2015), it enriches the spatial econometrics literature. However, from an economic perspective, endogenous variables may exist in the model, i.e., property tax. Moreover, for a multilevel model, prior selection is challenging. An in-depth analysis of various prior distributions should be necessary.

7.4. Future works

In this thesis, particular attention has been paid to addressing the spatial and/or temporal dimensions in hedonic housing prices models and mapping second home rate dynamics. In this section, we suggest some possible directions for future research.

Concerning the models in chapter 4, future studies may focus on the predictive power of different spatiotemporal components, including the tensor product approach, thin-plate splines and the SPDE approach. Moreover, it is possible to extend the proposed Bayesian hierarchical model to study the bargaining power associated with both home buyers and sellers in the market. For example, weak buyers probably pay higher prices for a house, but strong sellers may receive higher prices (Harding, Rosenthal, *et al.*, 2003; Harding, Knight, *et al.*, 2003).

In Section 7.3, I have mentioned that the proposed model in Chapter 5 relies on a spatial adjacency matrix, which should be selected prudently. A recent study on conditional autoregressive models focuses on the so-called

stochastic spatial adjacency matrix (Corpas-Burgos and Martinez-Beneito, 2020). This matrix creates an additional level in Bayesian hierarchical models for estimating the elements in the stochastic matrix. For example, Lu and Carlin (2005) and Lu, Reilly, *et al.* (2007) proposed a stochastic spatial adjacency matrix approach, where the element in the stochastic matrix is assumed to be a binary stochastic value following a Bernoulli distribution. I would like to extend the model described in Chapter 4 to adapt to this flexible spatial matrix selection approach.

On the other hand, I would also like to compare the model described in Chapter 5 with spatial lag models with spatiotemporal trends. A few scholars (Ver Hoef, Hanks, *et al.*, 2018; Wall, 2004) compare conditional autoregressive models and spatial lag models empirically and theoretically. To my knowledge, Simões *et al.* (2017) found that the spatial lag model outperform the conditional autoregressive model when they analyzed Portuguese national health line data. I intend to know, which model performs better in my case.

Concerning Chapter 6, future work may focus on the two aspects.

The first aspect centres on spatial adjacency matrices. Anselin and Arribas-Bel (2013) investigated a special case where spatial dependence is group-wise, with all observations in the same group as neighbours of each other. The group-wise setting is very interesting in the multilevel modelling context. As previously stated, EPCIs may have some impacts on individual units, known as group dependence. Therefore, I intend to know whether a multilevel model with a group-wise adjacency matrix outperforms other multilevel models with traditional adjacency matrices in terms of fitting data.

Another issue is about endogenous variables. Ross and Yinger (1999) indicated that early studies on the relationship between housing prices and property taxes may experience endogeneity of tax rates, therefore the two stages least square method should be used. Sirmans *et al.* (2008) examined the existing literature on property tax capitalization and pointed out that one third of the examined studies applied the two stages least square method, others used the ordinary least square method, and some researchers applied random coefficients models. I believe that the endogeneity of tax rates should be investigated in the Corsican case as well. Combining the spatial multilevel regression model with the instrumental variable approach should also be in-

vestigated in the future.

Appendix

7.5. Tables

Table 7.1: Data sources

Variable	Source
Physical landscape	Corsica Recreational Areas database, UMR LISA
Cultural landscape	Corsica Recreational Areas database, UMR LISA
Coastal county	Corsica GIS database, UMR LISA
Mountainous county	Corsica GIS database, UMR LISA
Population	www.insee.fr/fr/information/2008354
Household growth	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	www.impots.gouv.fr/portail/statistiques
Unemployment rate	www.insee.fr/fr/statistiques/1893230

7

Table 7.2: Descriptive statistics for the temporal variation of second home counts

	Mean	Min	Max	St. Dev.
2006	194.183	1	4474	386.796
2007	197.336	6	4465	389.425
2008	203.175	4	4906	413.219
2009	207.750	4	5077	426.535
2010	212.553	4	5115	432.917
2011	220.594	4	5375	450.401
2012	229.789	4	6025	482.729
2013	236.519	4	6025	486.182
2014	244.769	4	6748	524.319
2015	250.450	4	6539	524.530
2016	254.512	4	6581	529.661

Table 7.3: "EPCI" identifier and corresponding names

ID	Name
1	Communauté d'Agglomération du Pays Ajaccien
2	Communauté de communes de l'Alta - Rocca
3	Communauté de communes du Sartonais - Valinco
4	Communauté de communes du Sud Corse
5	Communauté d'Agglomération de Bastia
6	Communauté de communes d'île-Rousse - Balagne
7	Communauté de communes de Calvi - Balagne
8	Communauté de communes de Castagniccia - Casinca
9	Communauté de communes de l'Orient
10	Communauté de communes de l'Ouest Corse
11	Communauté de communes de la Costa Verde
12	Communauté de communes de la Haute Vallée de la Gravona
13	Communauté de communes de la Pieve d'Ornano
14	Communauté de communes de Marana - Golo
15	Communauté de communes des quatre territoires
16	Communauté de communes du Cap Corse
17	Communauté de communes du Centre Corse
18	Communauté de communes du Fium'Orbu - Castellu
19	Communauté de communes du Nebbiu Conca d'Oru

7.6. Graphs

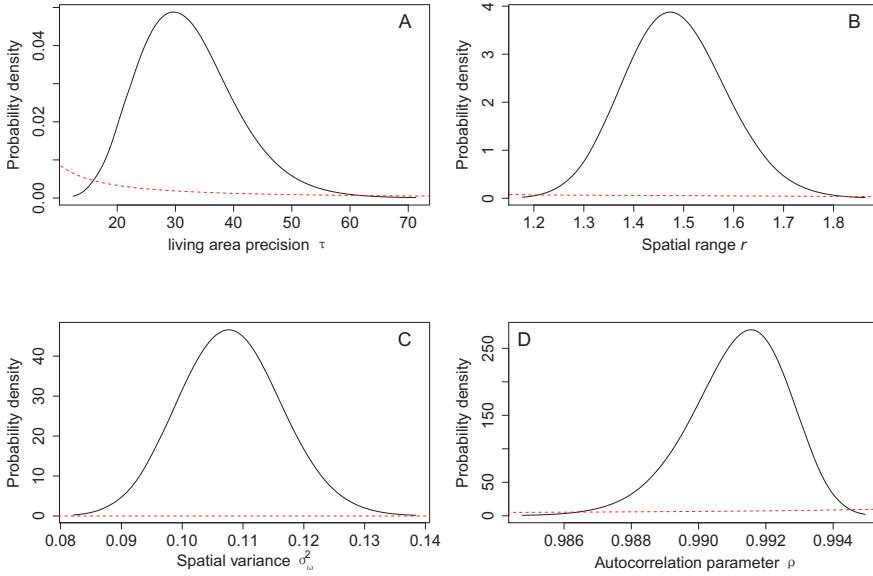


Figure 7.1: Prior (red dash line) and posterior (solid line) distribution for hyperparameters in M3. A. Precision parameter $\tau_{r_{w1}}$ for living area. B. Spatial range r in kilometres. C. Spatial variance σ_ω^2 . D. Autocorrelation parameter α .

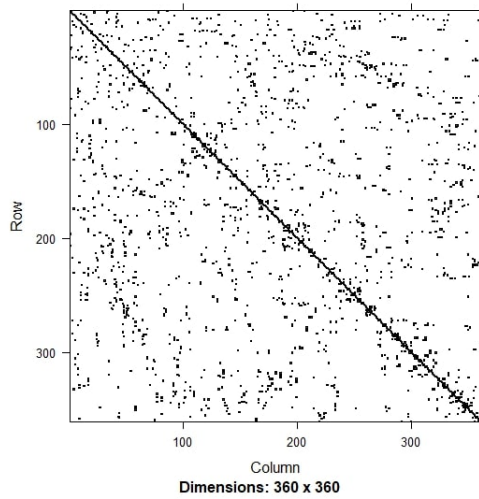


Figure 7.2: Adjacency matrix: rows and columns identify areas; squares identify neighbors.

8

Bibliography

- Abbott, A. and De Vita, G. (2012). "Pairwise convergence of district-level house prices in London". In: *Urban Studies* 49.4, pp. 721–740.
- Alonso, W. (1964). *Location and Land Use: Toward a General Theory of Land Rest.* Harvard University Press, p. 279. ISBN: 9780674730854.
- Alun, T. (1993). "The influence of wages and house prices on British interregional migration decisions". In: *Applied Economics* 25.9, pp. 1261–1268.
- Anderson, S. and West, S. (2006). "Open space, residential property values, and spatial context". In: *Regional Science and Urban Economics* 36.6, pp. 773–789.
- Anselin, L. (1988). *Spatial Econometrics: Methods and Models.* Vol. 4. Springer Netherlands, p. 289. ISBN: 9024737354.
- (2010). "Thirty years of spatial econometrics". In: *Papers in Regional Science* 89.1, pp. 3–25.
- Anselin, L. and Arribas-Bel, D. (2013). "Spatial fixed effects and spatial dependence in a single cross-section". In: *Papers in Regional Science* 92.1, pp. 3–17.

- Anselin, L. and Lozano-Gracia, N. (2009). "Spatial Hedonic Models". In: *Palgrave Handbook of Econometrics*. Palgrave Macmillan. Chap. 26, pp. 1213–1250.
- Arab, A., Hooten, M., and Wikle, C. (2017). "Hierarchical Spatial Models". In: *Encyclopedia of GIS*. Ed. by S. Shekhar, H. Xiong, and X. Zhou. Springer International Publishing, pp. 837–846.
- Arbia, G. (1989). "Stochastic spatial processes". In: *Spatial data configuration in statistical analysis of regional economic and related problems*. Kluwer Academic Publishers. Chap. 4, pp. 43–92.
- Arribas-Bel, D. and Sanz-Gracia, F. (2014). "The validity of the monocentric city model in a polycentric age: US metropolitan areas in 1990, 2000 and 2010". In: *Urban Geography* 35.7, pp. 980–997.
- Augustin, N., Musio, M., Von Wilpert, K., Kublin, E., Wood, S., and Schumacher, M. (2009). "Modeling spatiotemporal forest health monitoring data". In: *Journal of the American Statistical Association* 104.487, pp. 899–911.
- Back, A. and Marjavaara, R. (2017). "Mapping an invisible population: the uneven geography of second-home tourism". In: *Tourism Geographies* 19.4, pp. 595–611.
- Bailey, M., Muth, R., and Nourse, H. (1963). "A regression method for real estate price index construction". In: *Journal of the American Statistical Association* 58.304, pp. 933–942.
- Bajari, P. and Kahn, M. (2005). "Estimating Housing Demand With an Application to Explaining Racial Segregation in Cities". In: *Journal of Business & Economic Statistics* 23.1, pp. 20–33.
- Bakar, S. and Sahu, S. (2015). "spTimer: Spatio-temporal bayesian modelling using R". In: *Journal of Statistical Software* 63.15, pp. 1–32.
- Baker, M. (2016). "Statisticians issue warning over misuse of P values". In: *Nature News* 531.7593, p. 151.
- Bakka, H., Rue, H., Fuglstad, G.-A., Riebler, A., Bolin, D., Illian, J., Krainski, E., Simpson, D., and Lindgren, F. (2018). "Spatial modeling with R-INLA: A review". In: *WIREs Computational Statistics* 10.6. URL: <https://doi.org/10.1002/wics.1443>.
- Balcilar, M., Beyene, A., Gupta, R., and Seleteng, M. (2013). "'Ripple' effects in South African house prices". In: *Urban Studies* 50.5, pp. 876–894.

- Baltagi, B., Bresson, G., and Etienne, J. (2015). "Hedonic Housing Prices in Paris: An Unbalanced Spatial Lag Pseudo-Panel Model with Nested Random Effects". In: *Journal of Applied Econometrics* 30.3, pp. 509–528.
- Balzer, S. and Murati, D. (2015). *70000 logements à construire d'ici 2030*. [Online]. Available from: <https://www.insee.fr/fr/statistiques/1379785>. INSEE ANALYSES CORSE 7.
- Banerjee, S., Carlin, B., and Gelfand, A. (2014). *Hierarchical modeling and analysis for spatial data*. Chapman and Hall/CRC, p. 562. ISBN: 9781439819173.
- Barke, M. (2007). "Second Homes in Spain: An analysis of change at the provincial level, 1981-2001". In: *Geography* 92.3, pp. 195–207.
- (2008). "Second homes in the Spanish housing market: one market or two?" In: *Journal of Housing and the Built Environment* 23.4, pp. 277–295.
- Barnett, R. (2007). "Central and Eastern Europe: Real estate development within the second and holiday home markets". In: *Journal of Retail & Leisure Property* 6.2, pp. 137–142.
- Basile, R., Benfratello, L., and Castellani, D. (2013). "Geoadditive models for regional count data: an application to industrial location". In: *Geographical Analysis* 45.1, pp. 28–48.
- Bateman, I. (1994). "Contingent valuation and hedonic pricing: problems and possibilities". In: *Landscape Research* 19.1, pp. 30–32.
- Bateman, I., Langford, I., and Rasbash, J. (1993). *The impacts of changing willingness to pay question format in contingent valuation studies: An analysis of open-ended, iterative bidding and dichotomous choice formats*. Working Paper 93-05. The Centre for Social and Economic Research on the Global Environment (CSERGE).
- Belasco, E., Farmer, M., and Lipscomb, C. (2012). "Using a finite mixture model of heterogeneous households to delineate housing submarkets". In: *Journal of Real Estate Research* 34.4, pp. 577–594.
- Belhaj, M. (2003). "Estimating the benefits of clean air contingent valuation and hedonic price methods". In: *International Journal of Global Environmental Issues* 3.1, pp. 30–46.

- Belke, A. and Keil, J. (2018). "Fundamental Determinants of Real Estate Prices: A Panel Study of German Regions". In: *International Advances in Economic Research* 24.1, pp. 25–45.
- Benson, E. D., Hansen, J. L., Schwartz, A. L., and Smersh, G. T. (1998). "Pricing residential amenities: the value of a view". In: *The Journal of Real Estate Finance and Economics* 16.1, pp. 55–73.
- Bernardinelli, L., Clayton, D., Pascutto, C., Montomoli, C., Ghislandi, M., and Songini, M. (1995). "Bayesian analysis of space-time variation in disease risk". In: *Statistics in medicine* 14.21-22, pp. 2433–2443.
- Besag, J. (1974). "Spatial interaction and the statistical analysis of lattice systems (with discussion)". In: *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* 36.2, pp. 192–236.
- Besag, J., York, J., and Mollié, A. (1991). "Bayesian image restoration, with two applications in spatial statistics". In: *Annals of the Institute of Statistical Mathematics* 43.1, pp. 1–20.
- Bin, O., Crawford, T., Kruse, J., and Landry, C. (2008). "Viewscales and flood hazard: Coastal housing market response to amenities and risk". In: *Land economics* 84.3, pp. 434–448.
- Bin, O. and Kruse, J. (2006). "Real estate market response to coastal flood hazards". In: *Natural Hazards Review* 7.4, pp. 137–144.
- Bishop, K. and Timmins, C. (2011). *Hedonic prices and implicit markets: Estimating marginal willingness to pay for differentiated products without instrumental variables*. Working Paper 17611. National Bureau of Economic Research, pp. 1–44. URL: <https://www.nber.org/papers/w17611>.
- (2019). "Estimating the marginal willingness to pay function without instrumental variables". In: *Journal of Urban Economics* 109, pp. 66–83.
- Bivand, R., Sha, Z., Osland, L., and Thorsen, I. (2017). "A comparison of estimation methods for multilevel models of spatially structured data". In: *Spatial Statistics* 21, pp. 440–459.
- Blangiardo, M. and Cameletti, M. (2015). *Spatial and Spatio-temporal Bayesian Models with R - INLA*. John Wiley & Sons, p. 323. ISBN: 9781118950210.
- Blotevogel, H. (1998). "The Rhine-Ruhr metropolitan region: Reality and discourse". In: *European Planning Studies* 6.4, pp. 395–410.

- Bolitzer, B. and Netusil, N. (2000). "The impact of open spaces on property values in Portland, Oregon". In: *Journal of Environmental Management* 59.3, pp. 185–193.
- Bonnet, O., Bono, P.-H., Chapelle, G., and Wasmer, E. (2014). *Does housing capital contribute to inequality? A comment on Thomas Piketty's Capital in the 21st Century*. Working Paper 2014-07. Sciences Po. URL: <https://spire.sciencespo.fr/hdl:/2441/30nstiku669glbr6616n7mc2oq/resources/2014-07.pdf>.
- Bourassa, S., Hoesli, M., and Peng, V. (2003). "Do housing submarkets really matter?" In: *Journal of Housing Economics* 12.1, pp. 12–28.
- Bourassa, S., Hoesli, M., and Sun, J. (2004). "What's in a view?" In: *Environment and Planning A* 36.8, pp. 1427–1450.
- Bourne, L. (1981). *The Geography of Housing*. Scripta series in geography. Wiley, p. 288. ISBN: 9780470270585.
- Bowen, W. M., Mikelbank, B. A., and Prestegaard, D. M. (2001). "Theoretical and empirical considerations regarding space in hedonic housing price model applications". In: *Growth and Change* 32.4, pp. 466–490.
- Brady, M. and Irwin, E. (2011). "Accounting for spatial effects in economic models of land use: Recent developments and challenges ahead". In: *Environmental and Resource Economics* 48.3, pp. 487–509.
- Brander, L. and Koetse, M. (2011). "The value of urban open space: Meta-analyses of contingent valuation and hedonic pricing results". In: *Journal of environmental management* 92.10, pp. 2763–2773.
- Bretel, A. (2016). *320200 habitants en Corse au 1er janvier 2013*. [Online]. Available from: <https://www.insee.fr/fr/statistiques/1895102>. INSEE FLASH CORSE 10.
- (2017a). *324200 habitants en Corse au 1er janvier 2014*. [Online]. Available from: <https://www.insee.fr/fr/statistiques/2539230>. INSEE FLASH CORSE 20.
- (2017b). *327000 habitants en Corse au 1er janvier 2015*. [Online]. Available from: <https://www.insee.fr/fr/statistiques/3290785>. INSEE FLASH CORSE 29.

- Bretel, A. (2018). *330500 habitants en Corse au 1er janvier 2016*. [Online]. Available from: <https://www.insee.fr/fr/statistiques/3680694>. INSEE FLASH CORSE 37.
- Brook, D. (1964). "On the distinction between the conditional probability and the joint probability approaches in the specification of nearest-neighbour systems". In: *Biometrika* 51.3/4, pp. 481–483.
- Browne, W. J. and Draper, D. (2006). "A comparison of Bayesian and likelihood-based methods for fitting multilevel models". In: *Bayesian analysis* 1.3, pp. 473–514.
- Brueckner, J. (2007). "Urban growth boundaries: An effective second-best remedy for unpriced traffic congestion?" In: *Journal of Housing Economics* 16.3-4, pp. 263–273.
- Brunauer, W. A., Lang, S., Wechselberger, P., and Bienert, S. (2010). "Additive Hedonic Regression Models with Spatial Scaling Factors: An Application for Rents in Vienna". In: *The Journal of Real Estate Finance and Economics* 41.4, pp. 390–411.
- Brunsdon, C., Fotheringham, S., and Charlton, M. (1998). "Geographically weighted regression". In: *Journal of the Royal Statistical Society: Series D (The Statistician)* 47.3, pp. 431–443.
- Buja, A., Hastie, T., and Tibshirani, R. (1989). "Linear smoothers and additive models". In: *The Annals of Statistics* 17.2, pp. 453–510.
- Can, A. (1990). "The Measurement of Neighborhood Dynamics in Urban House Prices". In: *Economic Geography* 66.3, pp. 254–272.
- (1992). "Specification and estimation of hedonic housing price models". In: *Regional Science and Urban Economics* 22.3, pp. 453–474.
- Carrión-Flores, C. and Irwin, E. (2017). "A fixed effects logit model of rural land conversion and zoning". In: *The Annals of Regional Science* 58.1, pp. 181–208.
- Case, K. and Quigley, J. (2008). "How housing booms unwind: income effects, wealth effects, and feedbacks through financial markets". In: *European Journal of Housing Policy* 8.2, pp. 161–180.
- Case, K., Quigley, J., and Shiller, R. (2005). "Comparing wealth effects: the stock market versus the housing market". In: *The BE Journal of Macroeconomics* 5.1, pp. 1–30.

- Case, K. and Shiller, R. (1989). "The Efficiency of the Market for Single-Family Homes". In: *The American Economic Review* 79.1, pp. 125–137.
- Caudill, S., Detotto, C., and Prunetti, D. (2019). "Bargaining power in apartment sales in Corsica: A latent class approach". In: *Urban Studies* forthcoming. DOI: [10.1177/0042098019886738](https://doi.org/10.1177/0042098019886738).
- Chasco, C. and Le Gallo, J. (2013). "The Impact of Objective and Subjective Measures of Air Quality and Noise on House Prices: A Multilevel Approach for Downtown Madrid". In: *Economic Geography* 89.2, pp. 127–148.
- Cheer, J., Milano, C., and Novelli, M. (2019). "Tourism and community resilience in the Anthropocene: accentuating temporal overtourism". In: *Journal of Sustainable Tourism* 27.4, pp. 554–572.
- Chen, B. and Qi, X. (2018). "Protest response and contingent valuation of an urban forest park in Fuzhou City, China". In: *Urban Forestry & Urban Greening* 29, pp. 68–76.
- Chiang, S.-h. (2014). "Housing Markets in China and Policy Implications: Co-movement or Ripple Effect". In: *China & World Economy* 22.6, pp. 103–120.
- Clapp, J. (2003). "A Semiparametric Method for Valuing Residential Locations: Application to Automated Valuation". In: *The Journal of Real Estate Finance and Economics* 27.3, pp. 303–320.
- Clapp, J., Kim, H., and Gelfand, A. (2002). "Predicting Spatial Patterns of House Prices Using LPR and Bayesian Smoothing". In: *Real Estate Economics* 30.4, pp. 505–532.
- Cohen, J. P., Coughlin, C. C., and Clapp, J. M. (2017). "Local Polynomial Regressions versus OLS for Generating Location Value Estimates". In: *The Journal of Real Estate Finance and Economics* 54.3, pp. 365–385.
- Colwell, P. and Guntermann, K. (1984). "The value of neighborhood schools". In: *Economics of Education Review* 3.3, pp. 177–182.
- Coppock, J. T. (1977). *Second homes: Curse or blessing?* Pergamon, p. 229. ISBN: 0080213715.
- Corpas-Burgos, F. and Martinez-Beneito, M. A. (2020). "On the use of adaptive spatial weight matrices from disease mapping multivariate analyses". In: *Stochastic Environmental Research and Risk Assessment* 34.3, pp. 531–544.

- Corrado, L. and Fingleton, B. (2012). "Where is the economics in spatial econometrics?" In: *Journal of Regional Science* 52.2, pp. 210–239.
- Court, A. T. (1939). "Hedonic price indexes with automotive examples". In: *The dynamics of automobile demand*. Ed. by New York: General Motors Corporation., pp. 98–117.
- Cressie, N. (2015). *Statistics for Spatial Data. Revised*. John Wiley & Sons, p. 928. ISBN: 9780471002550.
- Cressie, N. and Huang, H. (1999). "Classes of Nonseparable, Spatio-Temporal Stationary Covariance Functions". In: *Journal of the American Statistical Association* 94.448, pp. 1330–1339.
- Cuccia, T. (2020). "Contingent valuation". In: *Handbook of Cultural Economics*. Edward Elgar Publishing. Chap. 10, pp. 95–105.
- De Graaff, T., Florax, R., Nijkamp, P., and Reggiani, A. (2001). "A general misspecification test for spatial regression models: dependence, heterogeneity, and nonlinearity". In: *Journal of Regional Science* 41.2, pp. 255–276.
- De Haan, J. and Diewert, E. (2013). "Hedonic regression methods". In: *Handbook on Residential Property Price Indices*. OECD. Chap. 5, pp. 49–64.
- De Smedt, T., Simons, K., Van Nieuwenhuysse, A., and Molenberghs, G. (2015). "Comparing MCMC and INLA for disease mapping with Bayesian hierarchical models". In: *Archives of Public Health* 73.1, O2.
- Deaton, A. and Muellbauer, J. (1980). *Economics and consumer behavior*. Cambridge university press, p. 450. ISBN: 0521296765.
- Dehring, C. and Dunse, N. (2006). "Housing Density and the Effect of Proximity to Public Open Space in Aberdeen, Scotland". In: *Real Estate Economics* 34.4, pp. 553–566.
- Delgado, J. and Wences, G. (2019). "A hedonic approach to the valuation of the effect of criminal violence on housing prices in Acapulco City". In: *Empirical Economics*, pp. 1–20.
- Dhar, P. and Ross, S. (2012). "School district quality and property values: Examining differences along school district boundaries". In: *Journal of Urban Economics* 71.1, pp. 18–25.
- Dieleman, F. (1996a). *Households and Housing: Choice and Outcomes in the Housing Market*. Routledge, p. 275. ISBN: 9780882851563.

- (1996b). "Modeling housing choice". In: *Netherlands journal of housing and the built environment* 11.3, pp. 201–207.
- Dong, G. and Harris, R. (2015). "Spatial Autoregressive Models for Geographically Hierarchical Data Structures". In: *Geographical Analysis* 47.2, pp. 173–191.
- Dong, G., Harris, R., Jones, K., and Yu, J. (2015). "Multilevel modelling with spatial interaction effects with application to an emerging land market in Beijing, China". In: *PloS one* 10.6, e0130761.
- Dong, G. and Wu, W. (2016). "Schools, land markets and spatial effects". In: *Land Use Policy* 59, pp. 366–374.
- Dower, M. (1977). "Planning aspects of second homes". In: *Second homes: Curse or blessing*. Elsevier Science & Technology. Chap. 10, pp. 210–237.
- Downes, T. and Zabel, J. (2002). "The impact of school characteristics on house prices: Chicago 1987–1991". In: *Journal of Urban Economics* 52.1, pp. 1–25.
- Dubé, J. and Legros, D. (2013a). "A spatio-temporal measure of spatial dependence: An example using real estate data". In: *Papers in Regional Science* 92.1, pp. 19–30.
- Dubé, J. and Legros, D. (2013b). "Dealing with spatial data pooled over time in statistical models". In: *Letters in Spatial and Resource Sciences* 6.1, pp. 1–18.
- (2014). *Spatial econometrics using microdata*. Wiley, p. 228. ISBN: 9781119008767.
- Dubin, R. (1992). "Spatial autocorrelation and neighborhood quality". In: *Regional Science and Urban Economics* 22.3, pp. 433–452.
- Efron, B. (2005). "Bayesians, frequentists, and scientists". In: *Journal of the American Statistical Association* 100.469, pp. 1–5.
- Égert, B. and Mihaljek, D. (2007). "Determinants of house prices in central and eastern Europe". In: *Comparative economic studies* 49.3, pp. 367–388.
- European Central Bank (2003). *Structural factors in the EU housing markets*. Tech. rep. Bank European Central, pp. 1–55. URL: <https://www.ecb.europa.eu/pub/pdf/other/euhousingmarketsen.pdf>.
- Fahrmeir, L. and Kneib, T. (2011). "Spatial smoothing, interactions and geoadaptive regression". In: *Bayesian smoothing and regression for longitudinal*,

- spatial and event history data*. Oxford University Press. Chap. 5, pp. 307–414.
- Fahrmeir, L., Kneib, T., Lang, S., and Marx, B. (2013). "Structured Additive Regression". In: *Regression: Models, methods and applications*. Springer Berlin Heidelberg. Chap. 9, pp. 535–595.
- Faraway, J., Wang, X., and Ryan, Y. (2018). *Bayesian Regression Modeling with INLA*. Chapman and Hall/CRC, p. 312. ISBN: 9781351165747.
- Ferrari, A. and Comelli, M. (2016). "A comparison of methods for the analysis of binomial clustered outcomes in behavioral research". In: *Journal of Neuroscience Methods* 274, pp. 131–140.
- Fik, T., Ling, D., and Mulligan, G. (2003). "Modeling Spatial Variation in Housing Prices: A Variable Interaction Approach". In: *Real Estate Economics* 31.4, pp. 623–646.
- Finkenstadt, B., Held, L., and Isham, V. (2006). *Statistical Methods for Spatio-Temporal Systems*. CRC Press, p. 286. ISBN: 9781420011050.
- Freeman, M. (1979). "Hedonic Prices, Property Values and Measuring Environmental Benefits: A Survey of the Issues". In: *The Scandinavian Journal of Economics* 81.2, pp. 154–173.
- Fujita, M. and Ogawa, H. (1982). "Multiple equilibria and structural transition of non-monocentric urban configurations". In: *Regional science and urban economics* 12.2, pp. 161–196.
- Füss, R. and Koller, J. (2016). "The role of spatial and temporal structure for residential rent predictions". In: *International Journal of Forecasting* 32.4, pp. 1352–1368.
- Gallent, N., Mace, A., and Tewdwr-Jones, M. (2003). "Dispelling a myth? Second homes in rural Wales". In: *Area* 35.3, pp. 271–284.
- (2017). *Second homes: European perspectives and UK policies*. Routledge, p. 272. ISBN: 1351901486.
- Gallent, N. and Tewdwr-Jones, M. (2001). "Second homes and the UK planning system". In: *Planning practice and research* 16.1, pp. 59–69.
- Galster, G. (1996). "William Grigsby and the analysis of housing sub-markets and filtering". In: *Urban Studies* 33.10, pp. 1797–1805.
- Gan, Q. and Hill, R. (2009). "Measuring housing affordability: Looking beyond the median". In: *Journal of Housing economics* 18.2, pp. 115–125.

- Gelfand, A., Diggle, P., Guttorp, P., and Fuentes, M. (2010). *Handbook of spatial statistics*. CRC press, p. 619. ISBN: 0429136501.
- Gelman, A. and Carlin, J. (2017). "Some natural solutions to the p-value communication problem—and why they won't work". In: *Journal of the American Statistical Association* 112.519, pp. 899–901.
- Geniaux, G. and Napoléone, C. (2008). "Semi-parametric tools for spatial hedonic models: an introduction to mixed geographically weighted regression and geoaddivitive models". In: *Hedonic methods in housing markets*. Springer. Chap. 5, pp. 101–127.
- Genovese-Bolleyn, A. and Torre, V. (2018). *Bilan annuel du tourisme - 2017*. [Online]. Available from: <https://www.insee.fr/fr/statistiques/3545896>. INSEE DOSSIER CORSE 9.
- Geoghegan, J., Wainger, L., and Bockstael, N. (1997). "Spatial landscape indices in a hedonic framework: an ecological economics analysis using GIS". In: *Ecological Economics* 23.3, pp. 251–264.
- Giannoni, S., Beaumais, O., and Tafani, C. (2017). *Foreign buyers, price dispersion and local buyers eviction in bargaining markets*. Working Paper 005. Laboratoire Lieux, Identités, eSpaces et Activités (LISA). URL: <https://umrlisa.univ-corse.fr/wp-content/uploads/2019/11/Working-Paper-TerRA-n%C2%B005.pdf>.
- Giussani, B. and Hadjimatheou, G. (1991). "Modeling regional house prices in the United Kingdom". In: *Papers in Regional Science* 70.2, pp. 201–219.
- Glaesener, M.-L. and Caruso, G. (2015). "Neighborhood green and services diversity effects on land prices: Evidence from a multilevel hedonic analysis in Luxembourg". In: *Landscape and Urban Planning* 143, pp. 100–111.
- Gneiting, T. (2002). "Nonseparable, Stationary Covariance Functions for Space–Time Data". In: *Journal of the American Statistical Association* 97.458, pp. 590–600.
- Gneiting, T. and Guttorp, P. (2010). "Continuous Parameter Spatio-Temporal Processes". In: *Handbook of Spatial Statistics*. Vol. 97. CRC Press. Chap. 23, pp. 427–436.
- Goicoa, T., Adin, A., Ugarte, M., and Hodges, J. (2018). "In spatio-temporal disease mapping models, identifiability constraints affect PQL and INLA

- results". In: *Stochastic Environmental Research and Risk Assessment* 32.3, pp. 749–770.
- Goldstein, H. (2011). *Multilevel statistical models*. Vol. 922. John Wiley & Sons, p. 382. ISBN: 9781119956822.
- Goodman, A. (1978). "Hedonic prices, price indices and housing markets". In: *Journal of Urban Economics* 5.4, pp. 471–484.
- (1998). "Andrew Court and the Invention of Hedonic Price Analysis". In: *Journal of Urban Economics* 44.2, pp. 291–298.
- Goodman, A. and Muth, R. (2013). *The economics of housing markets*. Taylor & Francis, p. 164. ISBN: 113646106X.
- Goodman, A. and Thibodeau, T. (2003). "Housing market segmentation and hedonic prediction accuracy". In: *Journal of Housing Economics* 12.3, pp. 181–201.
- Green, F. and Hadjimatheou, G. (1990). "Regional differences in personal savings". In: *Applied Economics* 22.7, pp. 933–945.
- Grigsby, W. (1963). *Housing markets and public policy*. University of Pennsylvania Press, p. 346. ISBN: 1512816477.
- Hall, M. (2015). "Second homes planning, policy and governance". In: *Journal of Policy Research in Tourism, Leisure and Events* 7.1, pp. 1–14.
- Hall, M. and Müller, D. (2004). *Tourism, mobility, and second homes: between elite landscape and common ground*. Vol. 15. Channel View Publications, p. 304. ISBN: 1873150814.
- (2018). *The Routledge Handbook of Second Home Tourism and Mobilities*. Routledge, p. 366. ISBN: 9781317193531.
- Halleck Vega, S. and Elhorst, P. (2015). "The SLX model". In: *Journal of Regional Science* 55.3, pp. 339–363.
- Hamilton, S. and Morgan, A. (2010). "Integrating lidar, GIS and hedonic price modeling to measure amenity values in urban beach residential property markets". In: *Computers, Environment and Urban Systems* 34.2, pp. 133–141.
- Hammersley, J. and Clifford, P. (1971). "Markov fields on finite graphs and lattices". URL: <http://www.statslab.cam.ac.uk/~grg/books/hammfest/hamm-cliff.pdf>.

- Harding, J. P., Knight, J. R., and Sirmans, C. F. (2003). "Estimating bargaining effects in hedonic models: Evidence from the housing market". In: *Real estate economics* 31.4, pp. 601–622.
- Harding, J. P., Rosenthal, S. S., and Sirmans, C. F. (2003). "Estimating bargaining power in the market for existing homes". In: *Review of Economics and statistics* 85.1, pp. 178–188.
- Harris, P. (2019). "A Simulation Study on Specifying a Regression Model for Spatial Data: Choosing between Autocorrelation and Heterogeneity Effects". In: *Geographical Analysis* 51.2, pp. 151–181.
- Harvill, J. (2010). "Spatio-temporal processes". In: *WIREs Computational Statistics* 2.3, pp. 375–382. URL: <https://onlinelibrary.wiley.com/doi/abs/10.1002/wics.88>.
- Hastie, T. and Tibshirani, R. (1987). "Generalized additive models: some applications". In: *Journal of the American Statistical Association* 82.398, pp. 371–386.
- (1993). "Varying-coefficient models". In: *Journal of the Royal Statistical Society: Series B (Methodological)* 55.4, pp. 757–779.
- Heikkila, E., Gordon, P., Kim, J. I., Peiser, R. B., Richardson, H. W., and Dale-Johnson, D. (1989). "What happened to the CBD-distance gradient? land values in a poliocentric city". In: *Environment and Planning A* 21.2, pp. 221–232.
- Held, L. (2000). "Bayesian modelling of inseparable space-time variation in disease risk". In: *Statistics in medicine* 19.17-18, pp. 2555–2567.
- Held, L. and Besag, J. (1998). "Modelling risk from a disease in time and space". In: *Statistics in medicine* 17.18, pp. 2045–2060.
- Herath, S. and Maier, G. (2010). *The hedonic price method in real estate and housing market research: a review of the literature*. Working Paper. WU Vienna University of Economics and Business. URL: https://epub.wu.ac.at/588/1/sre-disc-2010_03.pdf.
- Heyman, A., Law, S., and Berghauser Pont, M. (2018). "How is Location Measured in Housing Valuation? A Systematic Review of Accessibility Specifications in Hedonic Price Models". In: *Urban Science* 3.1, p. 3.
- Hill, R. (2013). "Hedonic price indexes for residential housing: A survey, evaluation and taxonomy". In: *Journal of economic surveys* 27.5, pp. 879–914.

- Himmelberg, C., Mayer, C., and Sinai, T. (2005). "Assessing high house prices: Bubbles, fundamentals and misperceptions". In: *Journal of Economic Perspectives* 19.4, pp. 67–92.
- Hindsley, P., Hamilton, S., and Morgan, A. (2013). "Gulf views: toward a better understanding of viewshed scope in hedonic property models". In: *The Journal of Real Estate Finance and Economics* 47.3, pp. 489–505.
- Hof, A. and Blázquez-Salom, M. (2013). "The linkages between real estate tourism and urban sprawl in Majorca (Balearic Islands, Spain)". In: *Land* 2.2, pp. 252–277.
- Holmes, M. J., Otero, J., and Panagiotidis, T. (2017a). "A Pair-wise Analysis of Intra-city Price Convergence Within the Paris Housing Market". In: *The Journal of Real Estate Finance and Economics* 54.1, pp. 1–16.
- (2017b). "Climbing the property ladder: An analysis of market integration in London property prices". In: *Urban Studies* 55.12, pp. 2660–2681.
- Hott, C. and Monnin, P. (2008). "Fundamental real estate prices: an empirical estimation with international data". In: *The Journal of Real Estate Finance and Economics* 36.4, pp. 427–450.
- Hoyos, D. and Mariel, P. (2010). "The Contingent and Hedonic Valuation Methods: Techniques for Valuing a Community's Resources". In: *Contingent valuation: Past, present and future* 4, pp. 329–343.
- Huang, J., Malone, B., Minasny, B., McBratney, A., and Triantafyllis, J. (2017). "Evaluating a Bayesian modelling approach (INLA-SPDE) for environmental mapping". In: *Science of the Total Environment* 609, pp. 621–632.
- Insee Corse (2010). *Atlas social de la Corse: logement*. [Online]. Available from: <https://www.insee.fr/fr/statistiques/1291502>. LES DOSSIERS DE L'ÉCONOMIE CORSE.
- Irwin, E., Jeanty, W., and Partridge, M. (2014). "Amenity Values versus Land Constraints: The Spatial Effects of Natural Landscape Features on Housing Values". In: *Land Economics* 90.1, pp. 61–78.
- Jansen, S., Coolen, H., and Goetgeluk, R. (2011). *The measurement and analysis of housing preference and choice*. Springer Science & Business Media, p. 272. ISBN: 9789048188949.
- Jayantha, W. M. and Lau, J. M. (2016). "Buyers' property asset purchase decisions: an empirical study on the high-end residential property market in

- Hong Kong". In: *International Journal of Strategic Property Management* 20.1, pp. 1–16.
- Jim, C. and Chen, W. (2009). "Value of scenic views: Hedonic assessment of private housing in Hong Kong". In: *Landscape and urban planning* 91.4, pp. 226–234.
- Jones, C., Leishman, C., and Watkins, C. (2004). "Intra-urban migration and housing submarkets: Theory and evidence". In: *Housing studies* 19.2, pp. 269–283.
- Kaltenborn, B., Andersen, O., and Nellemann, C. (2007). "Second home development in the Norwegian mountains: Is it outgrowing the planning capability?" In: *The International Journal of Biodiversity Science and Management* 3.1, pp. 1–11.
- Kaltenborn, B., Andersen, O., Nellemann, C., Bjerke, T., and Thrane, C. (2008). "Resident attitudes towards mountain second-home tourism development in Norway: The effects of environmental attitudes". In: *Journal of Sustainable Tourism* 16.6, pp. 664–680.
- Kammann, E. and Wand, M. (2003). "Geoadditive models". In: *Journal of the Royal Statistical Society: Series C (Applied Statistics)* 52.1, pp. 1–18.
- Karahan, F. and Rhee, S. (2019). "Geographic reallocation and unemployment during the Great Recession: The role of the housing bust". In: *Journal of Economic Dynamics and Control* 100, pp. 47–69.
- Kelejian, H. H. and Prucha, I. R. (1998). "A Generalized Spatial Two-Stage Least Squares Procedure for Estimating a Spatial Autoregressive Model with Autoregressive Disturbances". In: *Journal of Real Estate Finance and Economics* 171.1, pp. 99–121.
- Kessler, G. and Tafani, C. (2015). *Le double marché foncier et immobilier en Corse. Itinéraires géographiques*. Editions L'Harmattan, p. 190. ISBN: 9782343053141.
- Kiel, K. and Zabel, J. (2008). "Location, location, location: The 3L Approach to house price determination". In: *Journal of Housing Economics* 17.2, pp. 175–190.
- Kingsbury, L. M. (1941). "What Is Housing Economics?" In: *The Journal of Land & Public Utility Economics* 17.3, pp. 354–357.

- Koschinsky, J., Lozano-Gracia, N., and Piras, G. (2012). "The welfare benefit of a home's location: An empirical comparison of spatial and non-spatial model estimates". In: *Journal of Geographical Systems* 14.3, pp. 319–356.
- Krige, D. G. (1952). "A statistical analysis of some of the borehole values in the Orange Free State goldfield". In: *Journal of the Southern African Institute of Mining and Metallurgy* 53.3, pp. 47–64.
- Kyriazakou, E. and Panagiotidis, T. (2018). "A nonlinear pairwise approach for the convergence of UK regional house prices". In: *International Economics and Economic Policy* 15.2, pp. 467–481.
- Lambert, B. (2018). *A Student's Guide to Bayesian Statistics*. SAGE Publications, p. 520. ISBN: 9781526418289.
- Lancaster, K. (1966). "A New Approach to Consumer Theory". In: *Journal of Political Economy* 74.2, pp. 132–157.
- Łaszek, J. (2013). "Housing in consumer's theory". Working Paper. URL: <https://mpra.ub.uni-muenchen.de/52365/>.
- Latinopoulos, D., Mallios, Z., and Latinopoulos, P. (2016). "Valuing the benefits of an urban park project: A contingent valuation study in Thessaloniki, Greece". In: *Land Use Policy* 55, pp. 130–141.
- Leroux, B., Lei, X., and Breslow, N. (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*. Springer. Chap. 4, pp. 179–191.
- LeSage, J. (2014). "What regional scientists need to know about spatial econometrics". In: *SSRN Electronic Journal*, pp. 1–31. URL: <https://ssrn.com/abstract=2420725>.
- LeSage, J. and Pace, K. (2009). *Introduction to spatial econometrics*. Chapman and Hall/CRC, p. 340. ISBN: 9781420064254.
- Li, J. (2016). *Housing Policies in Hong Kong, China and the People's Republic of China*. Working Paper 566. Asian Development Bank Institute, pp. 1–34. URL: <https://www.adb.org/sites/default/files/publication/183031/adbi-wp566.pdf>.
- Li, M. and Brown, J. (1980). "Micro-neighborhood externalities and hedonic housing prices". In: *Land economics* 56.2, pp. 125–141.

- Ligus, M. (2018). "Measuring the Willingness to Pay for Improved Air Quality: A Contingent Valuation Survey." In: *Polish Journal of Environmental Studies* 27.2, pp. 763–771.
- Lin, X. and Zhang, D. (1999). "Inference in generalized additive mixed models by using smoothing splines". In: *Journal of the royal statistical society: Series b (statistical methodology)* 61.2, pp. 381–400.
- Lindgren, F., Rue, H., and Lindström, J. (2011). "An explicit link between gaussian fields and gaussian markov random fields: The stochastic partial differential equation approach". In: *Journal of the Royal Statistical Society. Series B: Statistical Methodology* 73.4, pp. 423–498.
- Linzer, D. (2013). "Dynamic Bayesian forecasting of presidential elections in the states". In: *Journal of the American Statistical Association* 108.501, pp. 124–134.
- Liu, H., Park, Y., and Zheng, S. (2002). "The interaction between housing investment and economic growth in China". In: *International real estate review* 5.1, pp. 40–60.
- Liu, M. (2018). "Land Financing-led Urbanization in China: Evolution, Scale and Lessons". In: *United Nations Economic Commission for Africa (UNECA)*, pp. 1–16.
- Liu, T., Hu, W., Song, Y., and Zhang, A. (2020). "Exploring spillover effects of ecological lands: A spatial multilevel hedonic price model of the housing market in Wuhan, China". In: *Ecological Economics* 170, p. 106568.
- Liu, W. (2020). "Valuation of Water Level: A Spatial Hedonic Analysis on Lakeshore Properties". In: *Journal of Agricultural and Resource Economics* forthcoming. URL: <http://ageconsearch.umn.edu/record/298432>.
- Lopez, M. A. G. and Olivera, I. M. (2005). *Employment decentralisation: polycentric compaction or sprawl? The case of the Barcelona Metropolitan Region 1986-1996*. Working Paper wpdea0511. Department of Applied Economics at Universitat Autònoma of Barcelona.
- Lu, H. and Carlin, B. (2005). "Bayesian areal wombling for geographical boundary analysis". In: *Geographical Analysis* 37.3, pp. 265–285.
- Lu, H., Reilly, C. S., Banerjee, S., and Carlin, B. (2007). "Bayesian areal wombling via adjacency modeling". In: *Environmental and Ecological Statistics* 14.4, pp. 433–452.

- Luciani, A. (2020). *Migrations résidentielles: un solde élevé composé majoritairement d'actifs*. [Online]. Available from: <https://www.insee.fr/fr/statistiques/4283384>. INSEE ANALYSES CORSE 28.
- Luquet, B. (2016). *Résidences de tourisme et hôtelières, villages de vacances...Hausse de fréquentation durant la saison 2015*. [Online]. Available from: <https://www.insee.fr/fr/statistiques/1895116>. INSEE FLASH CORSE 12.
- MacNab, Y. (2003). "Hierarchical Bayesian spatial modelling of small-area rates of non-rare disease". In: *Statistics in Medicine* 22.10, pp. 1761–1773.
- Malpezzi, S. (2008). "Hedonic Pricing Models: A Selective and Applied Review". In: *Housing Economics and Public Policy*. John Wiley & Sons. Chap. 5, pp. 67–89.
- Marcelo, C. and Sebastian, E. (2018). "Spatial effects and non-linearity in hedonic modeling: Will large data sets change our assumptions?" In: *Journal of Property Investment & Finance* 36.1, pp. 32–49.
- Marcus, L. (2010). "Spatial capital". In: *The Journal of Space Syntax* 1.1, pp. 30–40.
- Marcus, L., Heyman, A., Hellervik, A., and Stavroulaki, G. (2019). "Empirical support for a theory of spatial capital: Housing prices in oslo and land values in gothenburg". In: *12th International Space Syntax Symposium*, pp. 1–16.
- Martino, S. and Riebler, A. (2019). "Integrated nested laplace approximations". In: *arXiv preprint arXiv:1907.01248*.
- Martins, T., Simpson, D., Lindgren, F., and Rue, H. (2013). "Bayesian computing with INLA: New features". In: *Computational Statistics and Data Analysis* 67, pp. 68–83.
- Mathew, B., Léon, J., and Sillanpää, M. J. (2015). "Integrated nested Laplace approximation inference and cross-validation to tune variance components in estimation of breeding value". In: *Molecular Breeding* 35.3, p. 99.
- Maupertuis, M., Tafani, C., and Poggioli, A. (2017). "Les résidents secondaires en Corse: différentes façons de s'ancrer au territoire. L'exemple de la commune de Lumiu". In: *Téoros. Revue de recherche en tourisme* 36.36, 2.

- McDonald, J. and McMillen, D. (1990). "Employment subcenters and land values in a polycentric urban area: the case of Chicago". In: *Environment and Planning A* 22.12, pp. 1561–1574.
- McMillen, D. (1996). "One hundred fifty years of land values in Chicago: A nonparametric approach". In: *Journal of Urban Economics* 40.1, pp. 100–124.
- (2012). "Perspectives on spatial econometrics: linear smoothing with structured models". In: *Journal of Regional Science* 52.2, pp. 192–209.
- McMillen, D. and Redfearn, C. (2010). "Estimation and hypothesis testing for nonparametric hedonic house price functions". In: *Journal of Regional Science* 50, pp. 712–733.
- Meen, G. (1996). "Spatial aggregation, spatial dependence and predictability in the UK housing market". In: *Housing Studies* 11.3, pp. 345–372.
- (1999). "Regional House Prices and the Ripple Effect: A New Interpretation". In: *Housing Studies* 14.6, pp. 733–753.
- Mendelsohn, R. and Olmstead, S. (2009). "The Economic Valuation of Environmental Amenities and Disamenities: Methods and Applications". In: *Annual Review of Environment and Resources* 34.1, pp. 325–347.
- Michael, B., Vicky, S., and Michael, S. (2002). "Residential real estate prices: a room with a view". In: *Journal of Real Estate Research* 23.1-2, pp. 129–138.
- Mills, E. (1972). *Studies in the Structure of the Urban Economy*. Published for Resources for the Future by Johns Hopkins Press, p. 151. ISBN: 0801813670.
- Mitchell, D., Enemark, S., and Van der Molen, P. (2015). "Climate resilient urban development: Why responsible land governance is important". In: *Land Use Policy* 48, pp. 190–198.
- Moran, P. (1950). "Notes on continuous stochastic phenomena". In: *Biometrika* 37.1/2, pp. 17–23.
- Moreira de Aguiar, M., Simões, R., and Braz Golgher, A. (2014). "Housing market analysis using a hierarchical-spatial approach: the case of Belo Horizonte, Minas Gerais, Brazil". In: *Regional Studies, Regional Science* 1.1, pp. 116–137.
- Morris, M., Wheeler-Martin, K., Simpson, D., Mooney, S., Gelman, A., and DiMaggio, C. (2019). "Bayesian hierarchical spatial models: Implement-

- ing the Besag York Mollié model in stan". In: *Spatial and Spatio-temporal Epidemiology* 31, p. 100301.
- Mousumi, D., Sarmila, B., and Zakir, H. (2007). "Untapped demand for heritage: A contingent valuation study of Prinsep Ghat, Calcutta". In: *Tourism Management* 28.1, pp. 83–95.
- Müller, D. (2002). "Second home ownership and sustainable development in Northern Sweden". In: *Tourism and hospitality research* 3.4, pp. 343–355.
- Müller, D. and Hoogendoorn, G. (2013). "Second homes: Curse or blessing? A review 36 years later". In: *Scandinavian Journal of Hospitality and Tourism* 13.4, pp. 353–369.
- Muth, R. (1969). *Cities and Housing: The Spatial Pattern of Urban Residential Land Use*. University of Chicago Press, p. 355. ISBN: 9780226554136.
- Nicolai, M. (2019). *Les trois quarts des Corses travaillent dans leur intercommunalité de résidence*. [Online]. Available from: <https://www.insee.fr/fr/statistiques/4197169>. INSEE FLASH CORSE 42.
- Nobre, A., Schmidt, A., and Lopes, H. (2005). "Spatio-temporal models for mapping the incidence of malaria in Pará". In: *Environmetrics* 16.3, pp. 291–304.
- Noh, Y. (2019). "Does converting abandoned railways to greenways impact neighboring housing prices?" In: *Landscape and urban planning* 183, pp. 157–166.
- Norris, M. and Shiels, P. (2007). "Housing affordability in the Republic of Ireland: is planning part of the problem or part of the solution?" In: *Housing Studies* 22.1, pp. 45–62.
- Norris, M. and Winston, N. (2009). "Rising second home numbers in rural Ireland: Distribution, drivers and implications". In: *European Planning Studies* 17.9, pp. 1303–1322.
- Ong, S. E., Ho, K. H. D., and Lim, C. H. (2003). "A Constant-quality Price Index for Resale Public Housing Flats in Singapore". In: *Urban Studies* 40.13, pp. 2705–2729.
- Orford, S. (2000). "Modelling spatial structures in local housing market dynamics: A multilevel perspective". In: *Urban Studies* 37.9, pp. 1643–1671.

- Osland, L. (2010). "An application of spatial econometrics in relation to hedonic house price modeling". In: *Journal of Real Estate Research* 32.3, pp. 289–320.
- Osland, L., Thorsen, I., and Thorsen, I. (2016). "Accounting for local spatial heterogeneities in housing market studies". In: *Journal of Regional Science* 56.5, pp. 895–920.
- Pace, K. (1995). "Parametric, semiparametric, and nonparametric estimation of characteristic values within mass assessment and hedonic pricing models". In: *The Journal of Real Estate Finance and Economics* 11.3, pp. 195–217.
- Pace, K. and LeSage, J. (2004). "Spatial Statistics and Real Estate". In: *The Journal of Real Estate Finance and Economics* 29.1, pp. 147–148.
- Páez, A. (2009). "Recent research in spatial real estate hedonic analysis". In: *Journal of Geographical Systems* 11.4, pp. 311–316.
- Palmquist, R. (2005). "Property Value Models". In: *Handbook of Environmental Economics*. Vol. 2. Elsevier. Chap. 16, pp. 763–819.
- Paris, C. (2009). "Repositioning second homes within housing studies: Household investment, gentrification, multiple residence, mobility and Hyperconsumption". In: *Housing, Theory and Society* 26.4, pp. 292–310.
- Parsons, G. and Noailly, J. (2004). "A value capture property tax for financing beach nourishment projects: an application to Delaware's ocean beaches". In: *Ocean & Coastal Management* 47.1-2, pp. 49–61.
- Paterson, R. and Boyle, K. (2002). "Out of sight, out of mind? Using GIS to incorporate visibility in hedonic property value models". In: *Land economics* 78.3, pp. 417–425.
- Pedinielli, E. and Bretel, A. (2015). *Forte progression du nombre de logements*. [Online]. Available from: <https://www.insee.fr/fr/statistiques/1285516>. INSEE FLASH CORSE 5.
- Pérez, S. (2019). "A Europe of creditor and debtor states: explaining the north/south divide in the Eurozone". In: *West European Politics* 42.5, pp. 989–1014.
- Pettit, L. (1990). "The conditional predictive ordinate for the normal distribution". In: *Journal of the Royal Statistical Society: Series B (Methodological)* 52.1, pp. 175–184.

- Pompe, J. (2008). "The effect of a gated community on property and beach amenity valuation". In: *Land Economics* 84.3, pp. 423–433.
- Prasad, N. and Richards, A. (2008). "Improving median housing price indexes through stratification". In: *Journal of Real Estate Research* 30.1, pp. 45–72.
- Préfet de Corse (2017). *La pression foncière en Corse*. [Online]. Available from: <https://www.prefectures-regions.gouv.fr/corse/Grands-dossiers/Urbanisme-et-pression-fonciere>. Urbanisme et pression foncière.
- R Core Team (2019). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria. URL: <https://www.R-project.org/>.
- Rao, J. N. K. (2003). *Small Area Estimation*. John Wiley & Sons, p. 315. ISBN: 9780471722182.
- Rappaport, J. (2007). "A guide to aggregate house price measures". In: *Economic Review* 92.2, pp. 41–71.
- Razen, A., Brunauer, W., Klein, N., Lang, S., and Umlauf, N. (2015). "Hedonic House Price Modeling Based on Multilevel Structured Additive Regression". In: *Computational Approaches for Urban Environments*. Springer International Publishing. Chap. 5, pp. 97–122.
- Ready, R., Berger, M., and Blomquist, G. (1997). "Measuring Amenity Benefits from Farmland: Hedonic Pricing vs. Contingent Valuation". In: *Growth and Change* 28.4, pp. 438–458.
- Reichert, A. K. (1990). "The impact of interest rates, income, and employment upon regional housing prices". In: *The Journal of Real Estate Finance and Economics* 3.4, pp. 373–391.
- Rérat, P. (2018). "Spatial capital and planetary gentrification: residential location, mobility and social inequalities". In: *Handbook of Gentrification Studies*. Edward Elgar Publishing. Chap. 7, pp. 103–118.
- Riebler, A., Sørbye, S., Simpson, D., and Rue, H. (2016). "An intuitive Bayesian spatial model for disease mapping that accounts for scaling". In: *arXiv preprint arXiv:1601.01180*.

- Romero, V., Solís, E., and De Urena, J. (2014). "Beyond the metropolis: New employment centers and historic administrative cities in the Madrid global city region". In: *Urban geography* 35.6, pp. 889–915.
- Roos, M. and Held, L. (2011). "Sensitivity analysis in Bayesian generalized linear mixed models for binary data". In: *Bayesian Analysis* 6.2, pp. 259–278.
- Rosen, S. (1974). "Hedonic Prices and Implicit Markets: Product Differentiation in Pure Competition". In: *Journal of Political Economy* 82.1, pp. 34–55.
- Ross, S. and Yinger, J. (1999). "Sorting and voting: A review of the literature on urban public finance". In: *Handbook of regional and urban economics*. Vol. 3. Elsevier. Chap. 47, pp. 2001–2060.
- Rothenberg, J. R., Galster, G. C., Batler, R. V., and Pitkin, J. (1991). *The maze of urban housing markets*. University of Chicago Press, p. 549. ISBN: 9780226729510.
- Rue, H. and Held, L. (2005). *Gaussian Markov random fields : theory and applications*. Chapman & Hall/CRC, p. 263. ISBN: 9781584884323.
- Rue, H., Martino, S., and Chopin, N. (2009). "Approximate Bayesian inference for latent Gaussian models by using integrated nested Laplace approximations". In: *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* 71.2, pp. 319–392.
- Sahu, S. and Bakar, S. (2012). "Hierarchical Bayesian autoregressive models for large space-time data with applications to ozone concentration modelling". In: *Applied Stochastic Models in Business and Industry* 28.5, pp. 395–415.
- Sahu, S., Yip, S., and Holland, D. (2009). "Improved space-time forecasting of next day ozone concentrations in the eastern US". In: *Atmospheric Environment* 43.3, pp. 494–501.
- Samaniego, F. (2010). *A comparison of the Bayesian and frequentist approaches to estimation*. Springer Science & Business Media, p. 235. ISBN: 1441959416.
- Sander, H. A. and Polasky, S. (2009). "The value of views and open space: Estimates from a hedonic pricing model for Ramsey County, Minnesota, USA". In: *Land Use Policy* 26.3, pp. 837–845.

- Schrödle, B., Held, L., Riebler, A., and Danuser, J. (2011). "Using integrated nested Laplace approximations for the evaluation of veterinary surveillance data from Switzerland: a case study". In: *Journal of the Royal Statistical Society: Series C (Applied Statistics)* 60.2, pp. 261–279.
- Sheppard, S. (1999). "Hedonic analysis of housing markets". In: *Handbook of regional and urban economics* Vol3, pp. 1595–1635.
- Shi, W. and Lee, L. (2018). "A spatial panel data model with time varying endogenous weights matrices and common factors". In: *Regional Science and Urban Economics* 72, pp. 6–34.
- Shiller, R. (2007). *Understanding recent trends in house prices and home ownership*. Working Paper 13553. National Bureau of Economic Research, pp. 1–46. URL: <https://www.nber.org/papers/w13553>.
- Shimizu, C., Karato, K., and Nishimura, K. (2014). "Nonlinearity of housing price structure: Assessment of three approaches to nonlinearity in the previously owned condominium market of Tokyo". In: *International Journal of Housing Markets and Analysis* 7.4, pp. 459–488.
- Simões, P., Carvalho, M., Aleixo, S., Gomes, S., and Natário, I. (2017). "A spatial econometric analysis of the calls to the portuguese national health line". In: *Econometrics* 5.2, p. 24.
- Simpson, D., Lindgren, F., and Rue, H. (2012). "Think continuous: Markovian Gaussian models in spatial statistics". In: *Spatial Statistics* 1, pp. 16–29.
- Sirmans, S., Gatzlaff, D., and Macpherson, D. (2008). "The history of property tax capitalization in real estate". In: *Journal of Real Estate Literature* 16.3, pp. 327–344.
- Spiegelhalter, D., Best, N., Carlin, B., and Van Der Linde, A. (2002). "Bayesian measures of model complexity and fit". In: *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* 64.4, pp. 583–639.
- Stettler, J. and Giovanelli, D. (2008). "Image, truth and illusion in tourism promotion: The problem of the rapid spread of second homes in Switzerland and planning strategies". In: *Real Estate and Destination Development in Tourism. Successful Strategies and Instruments*. Erich Schmidt Verlag. Chap. 7, pp. 249–266.

- Sun, D., Tsutakawa, R., Kim, H., and He, Z. (2000). "Spatio-temporal interaction with disease mapping". In: *Statistics in Medicine* 19.15, pp. 2015–2035.
- Sutton, G. (2002). "Explaining changes in house prices". In: *BIS quarterly review* 32.1, pp. 46–55.
- Syz, J. M. (2008). *Property derivatives: Pricing, hedging and applications*. Vol. 440. John Wiley & Sons. ISBN: 0470694556.
- Taylor, B. and Diggle, P. (2014). "INLA or MCMC? A tutorial and comparative evaluation for spatial prediction in log-Gaussian Cox processes". In: *Journal of Statistical Computation and Simulation* 84.10, pp. 2266–2284.
- Taylor, L. (2003). "The Hedonic Method". In: *A Primer on Nonmarket Valuation*. Springer Netherlands. Chap. 10, pp. 331–393.
- (2017). "Hedonics". In: *The Economics of Non-Market Goods and Resources*. Springer. Chap. 7, pp. 235–292.
- Tempier, A. (2014). *1982-2011: trente ans de démographie en Corse*. [Online]. Available from: <https://www.insee.fr/fr/statistiques/1292178>. QUANTÎLE 26.
- Teye, A. L. and Ahelegbey, D. F. (2017). "Detecting spatial and temporal house price diffusion in the Netherlands: A Bayesian network approach". In: *Regional Science and Urban Economics* 65, pp. 56–64.
- Tierney, L. and Kadane, J. (1986). "Accurate approximations for posterior moments and marginal densities". In: *Journal of the American Statistical Association* 81.393, pp. 82–86.
- Tirroloni, A. (2017). *La croissance de la population portée par l'excédent migratoire*. [Online]. Available from: <https://www.insee.fr/fr/statistiques/2553590>. INSEE ANALYSES CORSE 11.
- Tirroloni, A., Lépidi, A., and Torre, V. (2017). *Bilan annuel du tourisme - 2016*. [Online]. Available from: <https://www.insee.fr/fr/statistiques/2854071>. INSEE DOSSIER CORSE 7.
- Tobler, W. (1970). "A Computer Movie Simulating Urban Growth in the Detroit Region". In: *Economic Geography* 46, pp. 234–240.
- (2004). "On the first law of geography: A reply". In: *Annals of the Association of American Geographers* 94.2, pp. 304–310.

- Toraichi, K., Yang, S., Kamada, M., and Mori, R. (1988). "Two-dimensional spline interpolation for image reconstruction". In: *Pattern Recognition* 21.3, pp. 275–284.
- Torre, V. (2018). *Une hausse modérée soutenue par les touristes étrangers*. [Online]. Available from: <https://www.insee.fr/fr/statistiques/3650486>. INSEE FLASH CORSE 34.
- Touzani, H. (2018). *Des résidences principales et secondaires en forte croissance*. [Online]. Available from: <https://www.insee.fr/fr/statistiques/3571002>. INSEE FLASH CORSE 32.
- Troy, A. and Romm, J. (2004). "Assessing the price effects of flood hazard disclosure under the California natural hazard disclosure law (AB 1195)". In: *Journal of Environmental Planning and Management* 47.1, pp. 137–162.
- Tse, R. (2002). "Estimating Neighbourhood Effects in House Prices". In: *Urban Studies* 39.7, pp. 1165–1180.
- Tyagi, V. (2017). "Content-Based Image Retrieval: An Introduction". In: *Content-Based Image Retrieval: Ideas, Influences, and Current Trends*. Springer. Chap. 1, pp. 1–27.
- Ugarte, M. D., Goicoa, T., Etxeberria, J., and Militino, A. F. (2012). "A P-spline ANOVA type model in space-time disease mapping". In: *Stochastic Environmental Research and Risk Assessment* 26.6, pp. 835–845.
- Utsunomiya, K. (2018). "The value of local railways: An approach using the contingent valuation method". In: *Research in Transportation Economics* 69, pp. 554–559.
- Ver Hoef, J. (2002). "Sampling and geostatistics for spatial data". In: *Eco-science* 9.2, pp. 152–161.
- Ver Hoef, J., Hanks, E., and Hooten, M. (2018). "On the relationship between conditional (CAR) and simultaneous (SAR) autoregressive models". In: *Spatial statistics* 25, pp. 68–85.
- Ver Hoef, J., Peterson, E., Hooten, M., Hanks, E., and Fortin, M. (2018). "Spatial autoregressive models for statistical inference from ecological data". In: *Ecological Monographs* 88.1, pp. 36–59.
- Viola, D., Arno, P., Maroko, A., Schechter, C., Sohler, N., Rundle, A., Necker-man, K., and Maantay, J. (2013). "Overweight and obesity: Can we recon-

- cile evidence about supermarkets and fast food retailers for public health policy?" In: *Journal of Public Health Policy* 34.3, pp. 424–438.
- Vogiatzakis, I., Pungetti, G., and Mannion, A. (2008). "Corsica". In: *Mediterranean island landscapes: natural and cultural approaches*. Vol. 9. Springer. Chap. 10, pp. 220–244.
- Von Graevenitz, K. and Panduro, T. (2015). "An Alternative to the Standard Spatial Econometric Approaches in Hedonic House Price Models". In: *Land Economics* 91.2, pp. 386–410.
- Von Thünen, J. H. (1966). *The isolated state*. Wartenberg, CM trans. Translation of: *Der isolierte Staat (1826)*. Pergamon Press.
- Waddell, P., Berry, B. J. L., and Hoch, I. (1993). "Housing price gradients: The intersection of space and built form". In: *Geographical Analysis* 25.1, pp. 5–19.
- Wagenmakers, E., Lee, M., Lodewyckx, T., and Iverson, G. (2008). "Bayesian versus frequentist inference". In: *Bayesian evaluation of informative hypotheses*. Springer. Chap. 9, pp. 181–207.
- Wall, M. (2004). "A close look at the spatial structure implied by the CAR and SAR models". In: *Journal of statistical planning and inference* 121.2, pp. 311–324.
- Waltl, S. R. (2016). "A hedonic house price index in continuous time". In: *International Journal of Housing Markets and Analysis* 9.4, pp. 648–670.
- Weeken, O. (2004). "Asset pricing and the housing market". In: *Bank of England Quarterly Bulletin*, Spring, pp. 32–41.
- Weller, J. (2016). "Least Squares, Maximum Likelihood, and Bayesian Parameter Estimation". In: *Genomic Selection in Animals*. John Wiley & Sons. Chap. 5, pp. 21–29.
- Wheaton, W. (2004). "Commuting, congestion, and employment dispersal in cities with mixed land use". In: *Journal of Urban Economics* 55.3, pp. 417–438.
- Wheaton, W. and Nechayev, G. (2008). "The 1998-2005 Housing "Bubble" and the Current "Correction": What's Different This Time?" In: *Journal of real Estate research* 30.1, pp. 1–26.
- Wilhelmsson, M. (2009). "Construction and updating of property price index series". In: *Property Management* 27.2, pp. 119–137.

- Williams, D., Haworth, J., Blangiardo, M., and Cheng, T. (2019). "A spatiotemporal Bayesian hierarchical approach to investigating patterns of confidence in the police at the neighborhood level". In: *Geographical Analysis* 51.1, pp. 90–110.
- Won, C. and Gray, R. (2004). "Noncausal Markov Random Fields". In: *Stochastic Image Processing*. Springer Science & Business Media. Chap. 2, pp. 11–66.
- Wood, S. (2015). "Package 'mgcv'". In: *R package version 1*, p. 29.
- (2017). *Generalized additive models: An introduction with R*. Chapman and Hall/CRC, p. 476. ISBN: 9781498728331.
- Wooldridge, J. (2010). *Econometric Analysis of Cross Section and Panel Data*. Vol. 1. MIT Press Books. The MIT Press, p. 1064. ISBN: 9780262232586.
- Xia, H. and Carlin, B. (1998). "Spatio-temporal models with errors in covariates: mapping Ohio lung cancer mortality". In: *Statistics in Medicine* 17.18, pp. 2025–2043.
- Yamagata, Y., Murakami, D., Yoshida, T., Seya, H., and Kuroda, S. (2016). "Value of urban views in a bay city: Hedonic analysis with the spatial multilevel additive regression (SMAR) model". In: *Landscape and Urban Planning* 151, pp. 89–102.
- Yang, L., Zhou, J., and Shyr, O. F. (2019). "Does bus accessibility affect property prices?" In: *Cities* 84, pp. 56–65.
- Yoo, S. and Wagner, J. (2016). "A review of the hedonic literatures in environmental amenities from open space: a traditional econometric vs. spatial econometric model". In: *International Journal of Urban Sciences* 20.1, pp. 141–166.
- Yu, S.-M., Han, S.-S., and Chai, C.-H. (2007). "Modeling the value of view in high-rise apartments: a 3D GIS approach". In: *Environment and Planning B: Planning and Design* 34.1, pp. 139–153.
- Zhu, L., Gorman, D. M., and Horel, S. (2006). "Hierarchical Bayesian spatial models for alcohol availability, drug "hot spots" and violent crime". In: *International Journal of Health Geographics* 5.1, p. 54.

Acknowledgements

This thesis would not have been possible without the assistance of a number of people and institutions.

First of all, I would like to express my sincere gratitude to Dr. Dominique Prunetti and Dr. Claudio Detotto who co-supervised me throughout this research. I express my thanks to them for their advice and support.

I would also like to thank the research center (UMR CNRS LISA 6240) and people who provided access to data for my thesis and who kindly answered all my related questions: Mr. Johan Jouve and Mr. Florian Gueniot.

I am very grateful to the R-INLA project, maintained by Dr. Havard Rue and his team members, who answered questions associated with R-INLA during these years.

My special thanks go to Dr. Ben Lambert, who has shared a series of courses on econometrics and Bayesian statistics online.

I would also like to thank to my friend, who have helped me during this journey: Dr. Romain Justinien Razafindravaosolonirina, Miss. Santiana Diaz etc.

Last but not least, my warmest thanks to all my family, particularly to my mother, who supports me all the time.