TITLE: *Are Indigenous Peoples more vulnerable to irregular migration than other groups? A Bayesian spatial analysis.*

M. Joaquín Lopez-Huertas

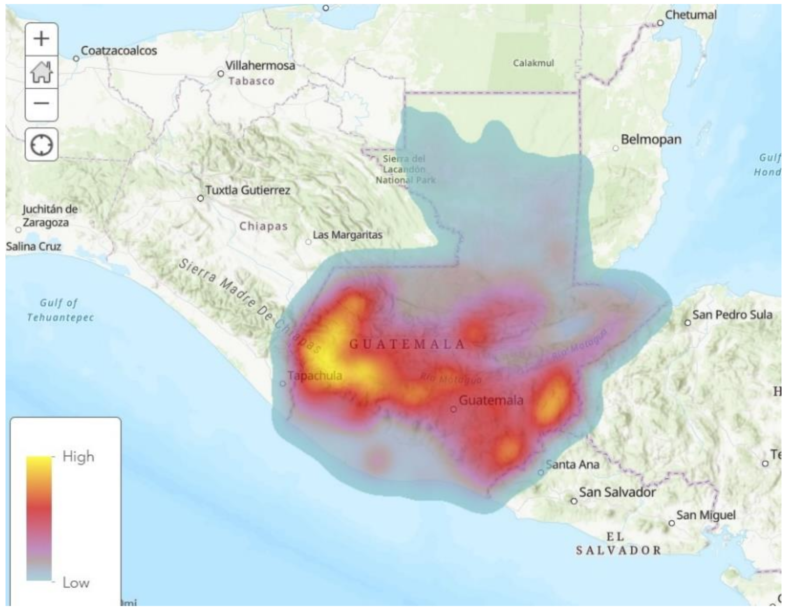
ABSTRACT

Irregular migration represents a phenomenon that requires an exhaustive analysis beyond conventional methods. In the case of Guatemala, although an increasing flow of irregular migrants to the US every year, little attention has been paid to the study of the factors that influence these events. Bayesian spatial analysis represents an alternative to studying this phenomenon by estimating the probability of events like irregular migration. Instead of focusing on accuracy in defining relationships among variables like frequentist models, Bayesian estimates trends that may arise based on available data. In this project, I conduct a Bayesian spatial analysis to understand the incidence of five socio-ecological components (insecurity, cultural, economic, governance, and environmental) with the number of irregular migrants to the US from Guatemala. I conduct the study on the Municipio (Towns) territorial scale and pay particular attention to the Indigenous population, the most vulnerable group in the country due to the inequalities and limitations experienced every day. By considering an initial number of 15 variables from the five components, I conduct a Principal Component Analysis to prioritize the elements that explain the variance. Then, I apply Integrated Nested Laplace Approximations in R-Software to develop a model that describes the trends, spatial effects, and exceedance rates of migrants in the following years. The results suggest the importance of accounting for a spatial location where migration flows might be more prevalent and provide trends on the probabilities of increasing rates in some Municipios compared with others, including the differences between urban and rural populations. The comparison between Indigenous and non-Indigenous Peoples suggests the need for a cultural understanding of Indigenous ways of being where racism and coloniality are embedded within Guatemalan society. This study offers a methodological and practical application for similar cases and allows to inform planning strategies in order to mitigate the effects of irregular migration in non-developed countries.

INTRODUCTION

Guatemala is one of the world's most culturally and naturally diverse countries. The indigenous population is the remnant of the 35 centuries of legacy of the Mayan civilization. Currently, more than 45% of the population is considered indigenous, and most live in rural areas. With more than 63% of the total population living in poverty and 25% in extreme poverty, Indigenous communities represent the most vulnerable group (Arriola, 2022; Menkos Zeissig & Medina Bermejo, 2020). The majority based their economy on practicing subsistence agriculture in a scattered distribution in rural areas. However, with climate change, severe droughts, and ecological challenges, irregular migration of the Indigenous population from rural areas to the US has increased. There is a social cost of irregular migration, considering the risk that immigrants experience. They leave their homeland, relation to family, land, and culture to be subject to a category where they are not considered individual with rights anymore. In addition, the impact of those who stay is also evident with the exacerbation of their limitations when there is a loss of connection between their relatives. In this view, it is essential to understand the conditions that influence Indigenous Guatemalans to leave their homeland. Particularly, I am interested in understanding the difficulties that Indigenous people face in the rural areas of Guatemala.

This project assesses the spatial conditions of Municipios in Guatemala with socioeconomic data to understand irregular migration patterns. Municipios represent a second-level political territorial division in Guatemala: the country is divided into Departamentos, like states or provinces, and every Departamento is divided into Municipios, like towns. In this analysis, I focus on the possible predictors, moderators, or mediators, including periods of drought, cropland, forest, violence, food insecurity, poverty, and urban-rural conditions, among others. Beyond estimating significant relationships among variables in a frequentist model, I apply Bayesian spatial analysis to understand the complexities of irregular migration phenomena among the variables. This method of analysis allows the estimation of posterior spatial distributions and identifies the risks or probabilities of increases in irregular immigration in Guatemala on a Municipal scale. As an introductory context, the image to the right presents the concentration of individuals who have migrated illegally to the US and have been returned to Guatemala, which overlaps with the highest concentration of Indigenous people. In this study, I intend to understand the interaction of those variables with spatial configurations and estimate probabilities of migration for the following years.



LITERATURE REVIEW

Studying immigration in Guatemala

Because of particular socioeconomic conditions and geographic location, scholars have studied the complexities of irregular migration in Guatemala to the US. International organizations like UN Migration have analyzed triggers that include extorsion, socio-natural disasters, and the impact of remittances on Guatemalan families (International Organization for Migration & Sweden Government, 2018). UN Migration monitors the number of apprehended or illegal immigrants caught in the US and sent back by Immigration, Customs, and Enforcement agents (ICE). The US Agency for International Development (USAID) is probably the most concerned international agency on illegal immigration to the US. With the investment in several projects across Guatemala, including a campaign titled “Quedate Aqui” (“Stay here”), they have studied the inequalities and the conditions that might lead to increased immigration (US Agency for International Development (USAID), 2020). These studies have been resourceful in data generation. I include some of this data in my analysis.

When reviewing articles concerned with irregular migration, scholars like Najera (2017) have conducted analyses of cross-border spaces between Guatemala and Mexico and the economic benefits migration entails. Others have focused on the inequalities of those who stay or are left behind, particularly with the challenges experienced by young and old women (Landry, 2011). Ruiz Marrujo (2001) studied an archive of interviews and observations made by institutions in contact with immigrants to build the concept of risk maps and understand population flows hotspots. Probably one of the most comprehensive analyses of immigration in Guatemala is the book "Guatemala – US Migration: Transforming Regions" (Jonas & Rodríguez, 2015). Through a chronological process of documenting historical changes, the authors conduct a collaborative research project with case studies and qualitative analysis in Guatemalan immigrant communities in the US. Their approach to the transitions from living to the periphery to the semi-periphery contextualizes the impacts of capitalism on different levels. They also observe differences experienced by the Guatemalan Mayan population compared to other immigrant groups due to varying forms of racism, discrimination, and violence. Besides diverse qualitative methods applied to studying immigration in Guatemala, including those who claim causality without considering empiric studies (Smolarek, 2007), there is little attention on understanding the problem from a spatial or quantitative approach. Paredes Orozco (2009) conducted a binary logistic regression to understand the differences between groups in Mexico and the US coming from Guatemala. His model included socioeconomic variables as possible predictors in three different "Logit" models. Although the results are inconclusive, he suggests a more thorough approach to understanding more than socioeconomic conditions. Finally, no spatial and quantitative analysis exists on the triggers, conditions, or influences of irregular migration from Guatemala to the US. In this project, I intend to develop a model that accounts for those spatial complexities.

Bayesian Spatial Data Analysis

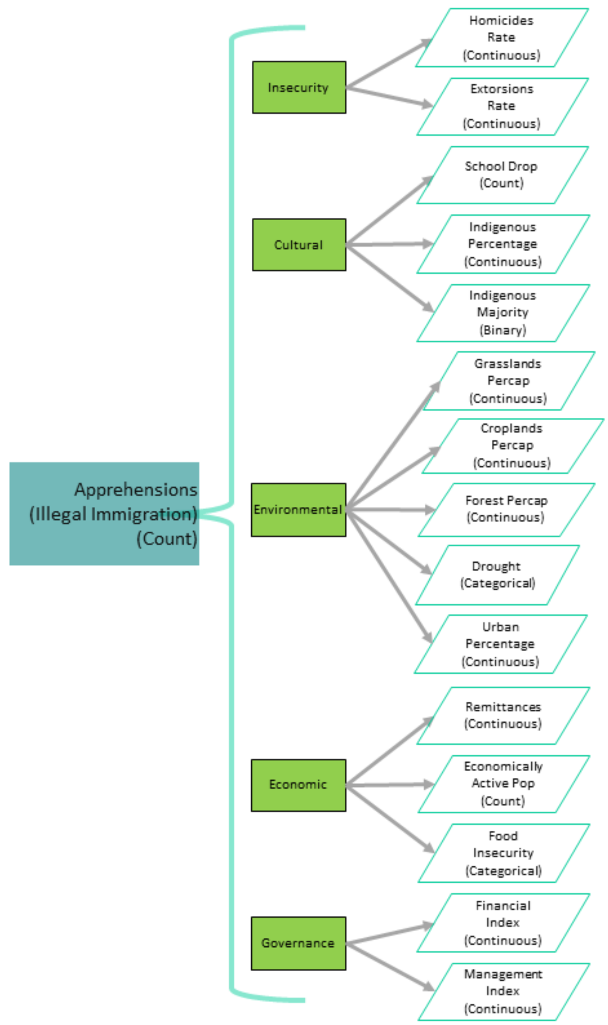
Bayesian spatial analysis allows the estimation of a probability distribution, also known as likelihood, with data observation. With prior information added to the model, it enables the re-allocation of credibility by eliminating possible reasons for the outcome and creates a posterior distribution that increases the credibility of the results (Kruschke, 2015). Credibility ranges from 0 to 1, where 0 is the lowest credible result, and one is the most credible. Unlike frequentist models that seek to calculate a significant relationship among variables, Bayesian analysis works with beliefs and interpretations based on available information. In Bayesian analysis, hypotheses are not tested, and instead of working with confidence intervals, it works with credibility intervals. In addition, Bayesian models also account for the challenges of fitting classical models with incomplete data, duplicated measures, and others (Moraga, 2019; Wang et al., 2018).

Bayesian Analysis in Immigration

Scholars apply Bayesian analysis in a myriad of contexts to understand behaviors or trends in data (Faubet et al., 2007; Faubet & Gaggiotti, 2008; Schaub & Fletcher, 2015). In irregular migration, some analyses have focused on the host immigrant countries to estimate future population flows. In Europe, scholars have studied the acceptance levels of immigration based on surveys and demographic data (Dalla Valle et al., 2020), including the estimation from selected countries with the implementation of priors with Delphi surveys (Bijak & Wiśniowski, 2010). Azose & Raftery (2015) conducted a joint probabilistic projection in the US for immigrants coming from different country. However, most of these cases are non-spatial, which suggests little attention to spatial autocorrelation. One of the few examples of Bayesian analysis applied with a spatial model was conducted by Núñez & López (2020) in the State of Chiapas, Mexico. Chiapas borders Guatemala in the west, which implies a constant flow of irregular migration. Their study applies a Monte Carlo Markov Chain to estimate the relative risks of immigration in Chiapas' 124 Municipios.

METHODS

The estimation of posterior distributions in Bayesian analysis represents one of the most significant challenges of this method. Alternatives have emerged with methods like the Monte Carlo Markov Chain (MCMC) that generates a value of sample chains while running iterations of possible results. This requires deciding at what level of iterations the model has reached the posterior distribution (Moraga, 2019). Although significant advances in Bayesian analysis with MCMC, the method requires high computational capacity and entails inconvenience when interpreting results. In this view, the Integrated Nested Laplace Approximations (INLA) approach has emerged as an alternative to applying Bayesian spatial analysis (Krainski et al., 2018; Moraga, 2019). Conceptualized by Laplace, the model estimates posterior distributions based on a latent model (Penny et al., 2007). The INLA package from the R software creates Laplace approximations of the data once hyperparameters have been introduced. Results provide posterior distributions, posterior marginals, and posterior density of the hyperparameters that allow an estimation of probabilities for a phenomenon to occur (Blangiardo & Cameletti, 2015; Gómez-Rubio, 2020).



Variables

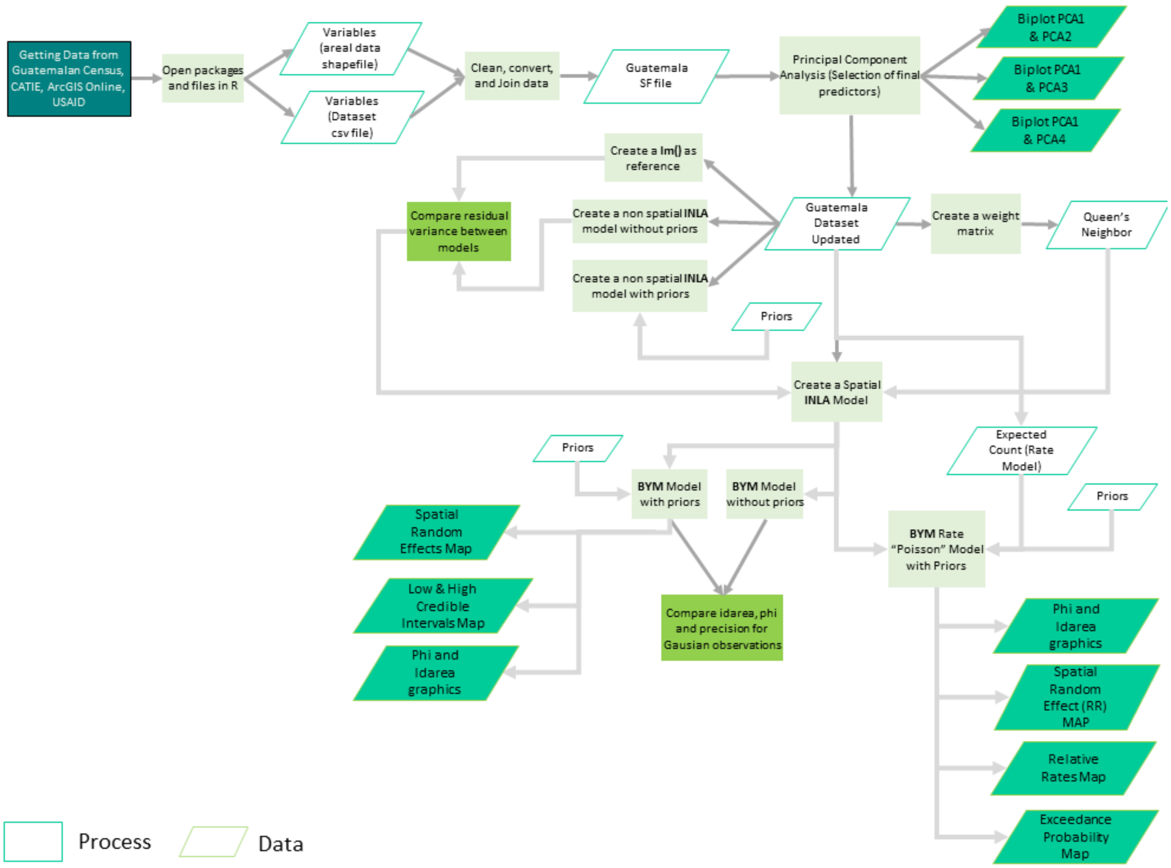
This section includes the definition of the variables applied to model irregular migration in each Municipio in Guatemala. All of them belong to 2018. The variable I intend to model represents the number of migrations as a count variable. Because an accurate number of individuals who migrate to the US from Guatemala is challenging to track, a variable that can be considered for this measure represents the number of individuals apprehended by the Income, Customs, and Enforcement agents. This variable is monitored by the International Organization for Migration. Here, I will use the data from 2018. In order to understand the conditions that trigger irregular migration, I focused on five components that influence Guatemalan culture and represent elements based on the literature that affect migration. The image to the right presents the variables included by component. I describe each component below:

- Insecurity: social factors that influence human safety and contextualize different levels of violence
    - **Homicides**: count of homicides as a result of a violent event. (Rate / continuous)
    - **Extorsions**: reported acts of coercion for money or any other type of property, particularly by organized crime. (Rate / continuous)

- Cultural: intrinsic cultural conditions that evidence population stratification and education levels
    - **School drop**: number of students that dropped school during the studied period. (# of school drop / count)
    - **Maya Population**: I analyzed two variables with this information. The first is a dichotomous (dummy) variable where 1 represents the majority of the Mayan population in each Municipio and 0 is not the majority. Second, the percentage of the Mayan population as a proportion of the total population in each Municipio in a continuous value from 0 to 1.
- Environmental: conditions of the built environment that influence human development in social and economic issues.
    - **Grasslands**: availability of grasslands that serve as food for cattle or other economic activities influencing income generation. (Area per capita / continuous)
    - **Croplands**: total arable land available for self-consumption or commerce. (Area per capita / continuous)
    - **Forest**: native forest area providing ecosystem services such as firewood, non-timber forest products, etc. (Area per capita / continuous)
    - **Drought**: Drought threat levels based on region and climate change adaptation (Levels of severity / categorical)
    - **Urban/Rural population**: proportion of the number of the urban and rural population. Although most of the population lives in urban areas, Guatemala is still one of the most rural countries in America. (Percentage / continuous)
- Economic: economic variables that influence food security and its relationships with migration reasons such as remittances.
    - **Remittances**: Number of US $ dollars sent to those who stay by immigrants from the US (value per capita / continuous).
    - **Economically active population**: number of individuals over 14 years old with possibility to engage in any formal or informal income generation activity (# of individuals economically active / count).
    - **Food insecurity**: Average among all the families in each Municipio with the capacity of covering nutritional demands every month (Levels of insecurity / categorical).
- Governance: Government structures and services performance influence the opportunities of the population for human development, including economic and social factors.
    - **Financial**: ranking of efficiency on financial performance in comparison with the other Municipios (index / continuous).
    - **Management**: ranking of efficiency on multiple activities that include access to information, public services, strategic planning, financial, public participation, and administrative management (index / continuous).

Workflow

The image below conceptualizes the process of analyzing the information in R software. They are described in the following paragraphs.



Gathering and transforming data

Because of the challenges of working with updated data in Guatemala, I focused on information from 2018 for all 352 Municipios. This considers the year of the last Guatemalan census. Therefore, all the variables applied in the study belong to the same year. I obtained some variables in areal data from ArcGIS Online in reliable sources such as the Centro Agronomico Tropical de Investigacion y Enseñanza (CATIE) from Costa Rica and the US Agency for International Development (USAID). To analyze the data, I used R-Software with packages that include dplyr, plyr, lme4, sf, spdep, ggplot2, ggpubr, tmap, and INLA. I modified and transformed the variables, normalized the distributions (rates instead of counts and per capita values instead of general territorial units), and applied left and spatial joins between SF files and CSV data frames to clean and elaborate a new file dataset for my analysis.

Principal Component Analysis

In order to find the most efficient model and check for high collinearity among the 18 predictors, I conducted a Principal Component Analysis of all the variables. I used prcomp() function in R-software because of its accuracy with singular values decomposition.

INLA non-spatial models

Comparing Integrated Nested Laplace (INLA) non-spatial models with frequentists linear models (lm) allows understanding if differences among residual errors are relevant and provide credibility to the study. In this case, by including the variables described in the previous section, I developed a simple linear model:

$$y = E(y) + e$$

Where y = observation

$E(y)$ = expected value

e = disturbance (deviation of any observation from the mean)

Then, I developed an INLA Linear Model and compare both of them.

$$yi = E(y) + ei$$

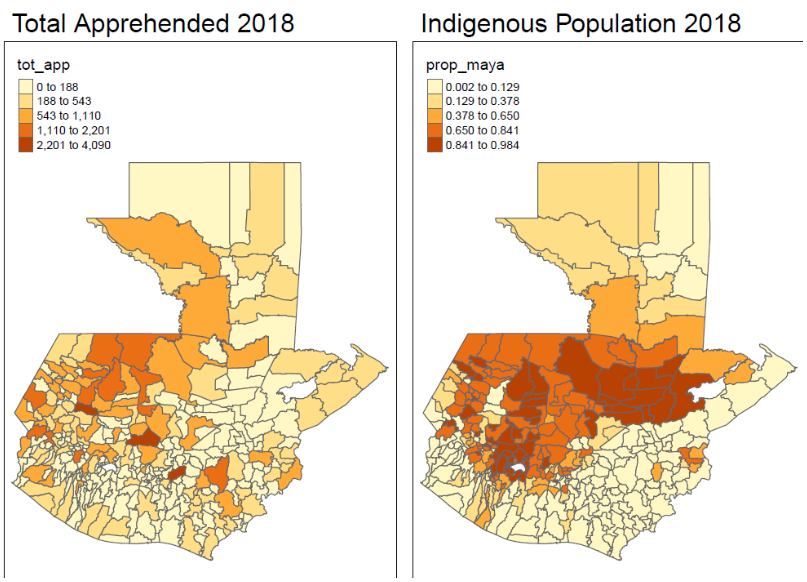
Where y = observation

$E(y) \sim \beta Xi$ = default prior for β

$e \sim \mathrm{N}(0, \sigma 2)$ = default prior for $\sigma 2$

INLA spatial models

Laplace approximations allow an understanding of the incidence of autocorrelation in spatial models (Wang et al., 2018). After comparing the non-spatial linear and INLA models, I developed a model to understand the spatial patterns of irregular migration to the US in each Municipio in Guatemala. As an overview, the map below presents a possible relationship between the number of Apprehensions during 2018 and the Indigenous Population as a percentage of the total population from the 352 Municipios. This includes a possible spatial dependence of Apprehensions in the north-central region where some of the highest Mayan population live.



Borrowing from disease modeling for health data (Moraga, 2019), autocorrelation can be analyzed with a model that relies on Conditional Autoregressive (CAR) distribution. This is the case for the Besag, York, and Mollie (BYM) model that adjust the data based on a neighborhood structure. I applied the queen's contiguities in a matrix to define proximity between neighbors. In the case of INLA, this translates to introducing a new random effect with the *f()* function. The new random effect requires an index of the number of areas (*idarea*), a model (*bym2*), and a spatial weight matrix (queen's neighbor). The model looks like:

$$yi = E(y) + ui + vi$$

Where yi = observations

$E(y) \sim \beta Xi, \beta \sim N(0, 1 / \tau\beta)$

$vi \sim N(0, \sigma 2v), 1/ \sigma 2v = \tau\beta \sim log\text{-}gamma(a,b)$

ui = Spatial error for any location, which is conditional based on its neighbors

In order to increase the validity of the model, it was necessary to run a version of the model without priors and another with priors. In this model, I specified a penalized complexity prior to increase phi value precision (spatial + non-spatial effect). This value is given by $u\alpha$ (spatial error), where α is the probability of hyperparameters exceeding u (Moraga, 2019). In addition, having the Maya population in two different formats allowed to test results with a dichotomous or a proportional (continuous) variable.

By applying the BYM model, the output of the model includes hyperparameters values described below:

- Precision for the *Gaussian Observations*: this represents the precision of the non-spatial residuals or, in other words, what the model cannot explain.
- The precision of the *Idarea*: precision on the spatial field effect. The value describes the error term divided by u = size of the spatial variance and v = size of the residual variance.
- *Phi* for the *Idarea*: a mixing parameter that decomposes the precision of the *Idarea* into a spatial (*u*) and non-spatial (*e*) effect. If the value is one, all the leftover information is entirely spatially autocorrelated; if the value is 0 is not autocorrelated.

For a better understanding of spatial effects and enable a more suitable analysis for the outcome variable (count), I modified the last INLA version into a Poisson model. Unlike the previous models, the results were modeled as rates rather than counts. A moderate expectation of the count represents a measurement that depends on the size of the studied population. In this model, the number of Apprehensions is correlated with the total population of each Municipio. With a rate of Apprehensions per Municipio, the model can estimate if a rate on a per capita basis is higher or lower than the expected number of cases. We can assume this also as the relative risks for irregular migration per Municipio. With the R-INLA package, I estimated a relative risk model based on Moraga's application for the case of lung cancer in Ohio (2019):

$$yi \sim Po(Ei\ \theta i),$$

Where yi = observation

Ei = Expected number of cases

$\theta i\ (log(\theta ij) = \alpha + ui + vi + (\beta + \delta i)$ = a sum of multiple components that consider spatial configurations which spatial autocorrelation.

In addition, I selected again penalized complexity priors (prob of theta > 1 is 0.01 & prob of theta > 0.5 is 3/4) to produce a relative smooth effect. The last model in R-Studio with the INLA packages is included below:

```
# Priors definition
Priors_Guate <- list(
  prec = list(prior = "pc.prec",
              param = c(0.5 / 0.31, 0.01)),
  phi = list(prior = "pc",
             param = c(0.5, 2 / 3))
)
# Formula with all the variables
Formula <- tot_app ~ 1 + prop_maya + SCHOOl_DRO + FoodInsVal + remitt_percap +
  PEA + crop_percap + tree_percap + CY_EXTORSI +
  FinanceInd + Urb_percent + f(idarea, model ="bym2", graph = gt_w,
                               hyper = prior_bmy3)
# INLA Model
Guate_rr <- inla(formula5, data = guate, family = "poisson", E = Ei,
                 control.compute = list(dic = TRUE, waic = TRUE,
                                        return.marginals.predictor = TRUE),
                 control.predictor = list(compute = TRUE),
                 control.inla = list(strategy = "adaptive",
                                     int.strategy = "eb"),
                 verbose = TRUE)
```

Where **Priors_Guate** = configuration of penalized complexity priors

Formula = Number of apprehensions as a function of all selected variables. This includes the f() with the idarea, type of model (bym2), weight matrix, and priors

Guate_rr = INLA model with the formula, priors and the definition of the Poisson family + expected count, and the call for the model to return different values including DIC, WAIC or the marginal distribution.

RESULTS

Principal Component Analysis

With the first four components, the proportion of the variance reached 0.62 (or 62% of the variance explained by the first four components). The image below provides the results of the 18 PCA and highlights the first four.

Importance of components:

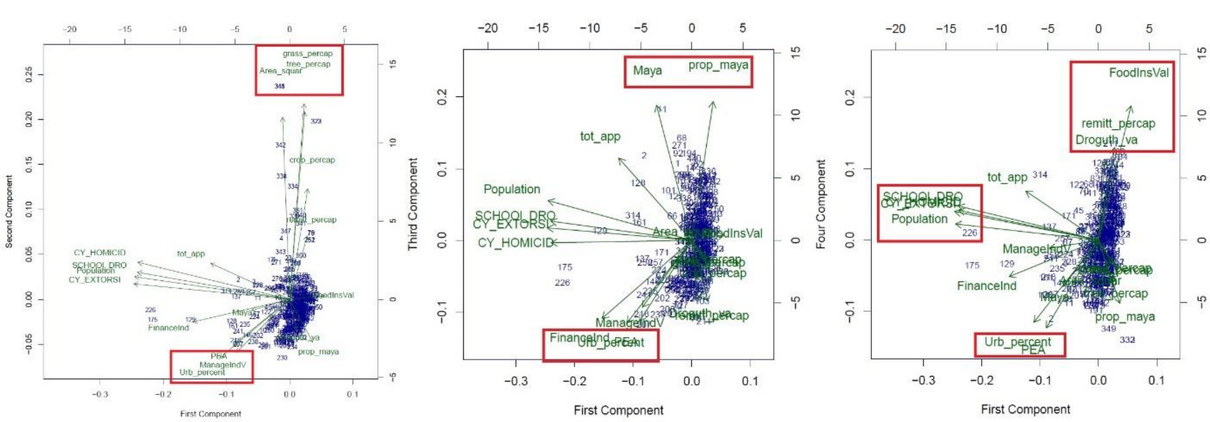
	PC1	PC2	PC3	PC4	PC5	PC6	PC7	PC8
Standard deviation	2.1450	1.5856	1.4996	1.3623	1.08215	1.01631	0.9420	0.84193
Proportion of Variance	0.2556	0.1397	0.1249	0.1031	0.06506	0.05738	0.0493	0.03938
Cumulative Proportion	0.2556	0.3953	0.5202	0.6233	0.68836	0.74574	0.7950	0.83442

	PC9	PC10	PC11	PC12	PC13	PC14	PC15	PC16
Standard deviation	0.77822	0.72038	0.66307	0.61928	0.59991	0.52185	0.51723	0.29006
Proportion of Variance	0.03365	0.02883	0.02443	0.02131	0.01999	0.01513	0.01486	0.00467
Cumulative Proportion	0.86807	0.89690	0.92132	0.94263	0.96262	0.97775	0.99261	0.99729

	PC17	PC18
Standard deviation	0.1848	0.12120
Proportion of Variance	0.0019	0.00082
Cumulative Proportion	0.9992	1.00000

The loadings represent associations between the original variables and the model's components. Higher values make the associations stronger, and they can be positive or negative. When comparing the first four components, it shows that Homicides, Extorsions, and Population increase with negative values of PC1 while Grasslands and Tree Coverage increase with positive values on PC2, suggesting that PC1 relates with safety while PC2 to the natural environment. By extracting the scores of every Municipio, I created a gradient from negative to positive values. Biplots between the first four components provide the gradients and facilitate the selection of the final variables.

The images below represent the biplot between the four PCA. The chart on the left is a biplot between PCA1 and PCA2, which suggest that PCA2 splits between urban population and grassland and tree coverage. This assumes a relationship between urban centers and the natural environment. The chart at the center indicates that PCA3 is divided between the Maya proportion and the Urban population, which suggests the relationship between Indigenous people's location and the built configuration. The image at the right presents PCA1 with PCA4 suggesting a socio-cultural component where PC1 is explained partially with safety (homicides and extorsions) and school drop variables. PC4 is divided between economic conditions (economically active population) and location (urban/rural) and food security (drought, food insecurity).



With this analysis, I eliminated Homicides because Extorsions provided a higher value for the variance. In the same way, I removed grasslands because of the high collinearity with forest. Moreover, forest represents a better variable for understanding migration triggers because of the ecosystem services that it provides. Drought severity was also removed because of high collinearity with food insecurity. In this case, Food Insecurity represents a more direct effect than Drought on possible triggers for migration. Between governance variables, the Finance index represents a better measurement considering that it focuses on elements similar to the Management index but does not relate to economic conditions (PCA4). Although School Drop and Extorsions were correlated, they represent different measures, and both were included in the final model. I selected ten variables as predictors to understand the spatial incidence of irregular migration including: Extorsions, Maya Population, School Drop, Croplands, Forests, Food Insecurity, Urban/Rural, Economically Active Population, Remittances, and Finance Index for each Municipio.

INLA non-spatial models

The image to the right provides the results of the simple linear model. Although the outcome variable is a count and a Poisson model would be more suitable, comparing the residual error facilitated the model design. The linear model suggests an increase in Apprehensions with variables like Mayan Population, School Drop, and Remittances while a decrease with the Economically Active Population. All of them are statistically significant. The remaining variables do not have statistically significant relationships. In addition, the R-squared is 0.28 (28% of the variance of Apprehensions is explained by the predictors), which is significant and a Residual Standard Error of 399.4.

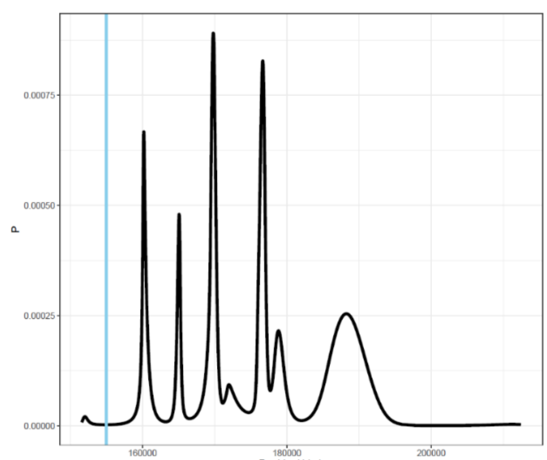
```
Call:
lm(formula = formula, data = guate)

Residuals:
   Min     1Q Median     3Q    Max
-625.6 -191.3  -95.9   37.2 3587.7

Coefficients:
               Estimate Std. Error t value Pr(>|t|)
(Intercept)   4.911e+02  2.108e+02   2.329 0.020423 *
prop_maya     1.975e+02  6.721e+01   2.939 0.003520 **
SCHOOl_DRO    5.350e-01  2.641e-01   2.026 0.043581 *
FoodInsVal    6.555e+01  5.049e+01   1.298 0.195082
remitt_percap 1.432e-01  8.325e-02   1.720 0.086382 .
PEA          -1.207e+03  3.128e+02  -3.858 0.000137 ***
crop_percap  -3.005e-02  1.935e-02  -1.553 0.121327
tree_percap   5.498e-04  9.675e-04   0.568 0.570200
CY_EXTORSI    4.397e-01  8.634e-01   0.509 0.610870
FinanceInd    9.072e+01  3.737e+02   0.243 0.808323
Urb_percent   9.487e+01  1.004e+02   0.945 0.345322
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 399.4 on 341 degrees of freedom
Multiple R-squared:  0.2888,    Adjusted R-squared:  0.268
F-statistic: 13.85 on 10 and 341 DF,  p-value: < 2.2e-16
```

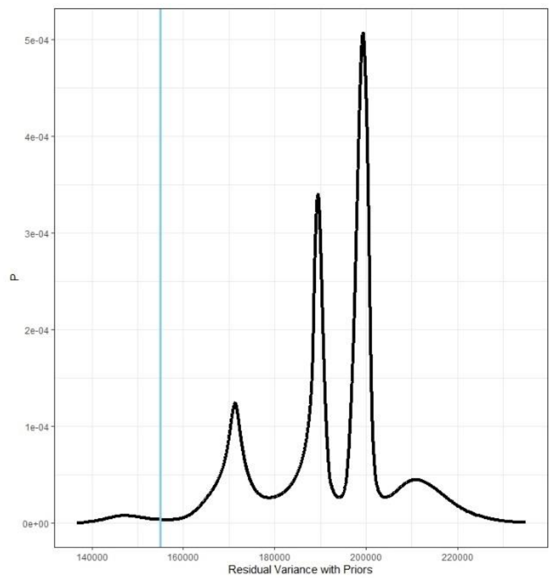
The first INLA model did not include priors. The results suggest that 0.025 and 0.975 quantiles are only partially credible at those levels. However, they might be credible to lower levels, such as 80%. This model allows an understanding of the limitations when ignoring spatial autocorrelation. Although Bayesian models do not get a standard error and p-value, we can describe the posterior distribution of each coefficient. The plot to the right represents a posterior distribution of the residual variance. The blue line is the estimation for the linear model.



The INLA model provides the precision for the Gaussian observation values. Comparing the mean of this precision value (415.5) with the residual variance of the previous linear model (399.4) suggests a credible model that I further developed. In addition, the low symmetric Kullback-Leibler divergence (kld) values also provide credibility to the model (Krainski et al., 2018).

One of the principles of Bayesian analysis, as described previously, is the addition of priors in order to reallocate posterior distributions and increase credibility. The second version of the INLA Linear Model included the informed priors to the slope and intercept coefficients. This allows for reducing the variance or range values effectively. The mean of these priors was 0.

The image to the left presents the posterior distribution of the residual variance after adding the informed priors, while the blue line is the estimation for the linear model. Here we see that the model improved with the "log gamma" priors. As a rule of thumb, the Deviance Information Criterion (DIC) and the Watanabe-Akaike Information Criterion (WAIC) must improve and reduce their values to improve the model. In this case, none of the coefficients improved substantially. Although the intercept improves its credibility in the 0.025 and 0.975 quantiles, it was not the case for relevant variables such as the Maya population, Food Insecurity, Remittances, and other variables. This required account for possible autocorrelation, which is included in the following paragraphs.

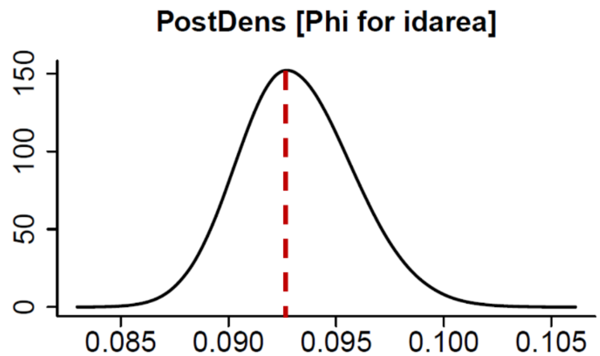


INLA spatial models

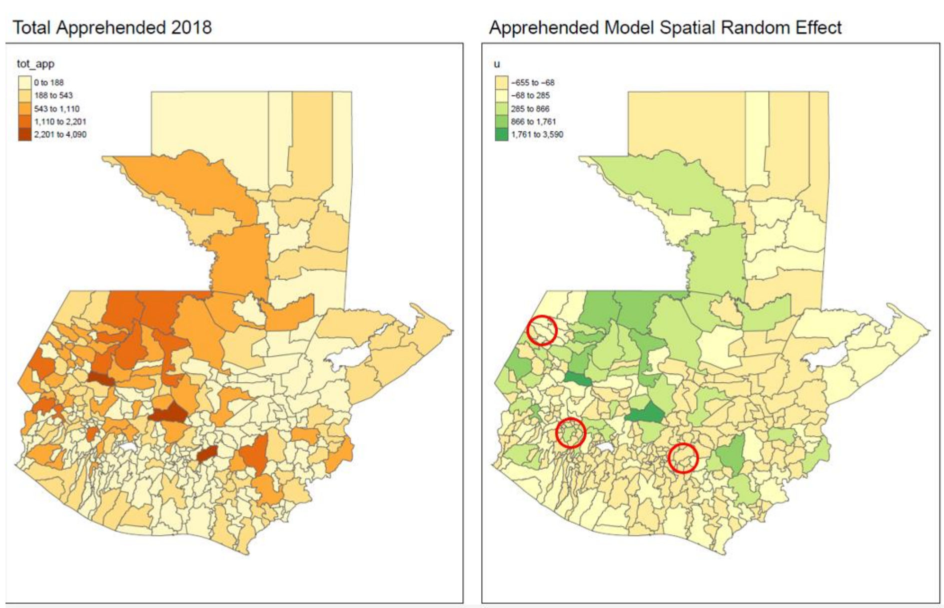
The chart below describes the changes in coefficients and quantiles based on applying two different variables as predictors. The case of the Mayan population as a proportion of the total population (instead of the Mayan population as a binary dichotomous variable) provides more consistent results and compared with the WAIC, DIC, and Quantile values of the first non-spatial INLA analysis, the model has improved credibility.

	Without Priors			**With Penalized Priors**		
Mayan Pop	WAIC	DIC	Quantiles	WAIC	DIC	Quantiles
Binary (0,1)	1595	1673	Only Forest and Finance are not credible at 0.975	5223	5242	Only Intercept and School drop are credible at 0.975
Proportion (% of total)	-909	-793	Only Finance is not credible at 0.975	-3092	-2984	Only Finance is not credible at 0.975

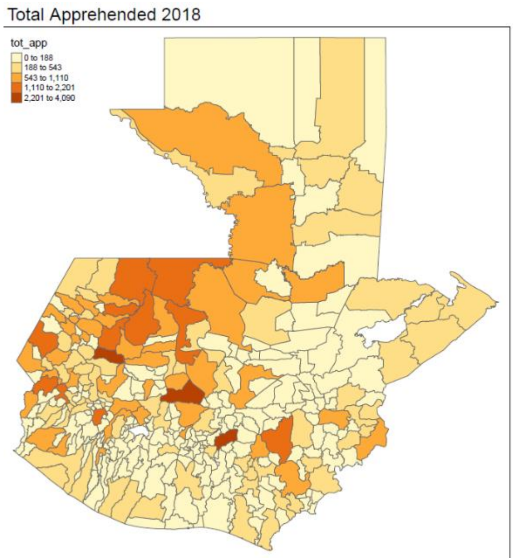
In the results of the model with the Proportion of the Mayan population, the Phi for the Idarea is 0.093, which means that a spatially autocorrelated error explains 9.3% of the remaining unexplained variation. In other words, the model has accounted for 90.7% of the spatially autocorrelated error. The chart below shows the posterior distribution of the Phi value.

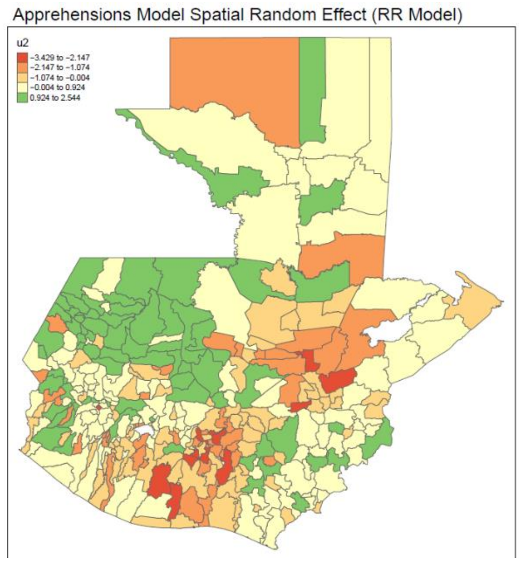


With the Precision of the *Idarea* value, it was possible to create a map of the spatial effect. This represents the random effect in space modeled by BYM for the 352 Municipios. The maps below present the estimation as the adjustment of the model for each Municipio (left) and is compared with the actual number of Apprehensions (left). With the spatial random effect map, the low values (oranges and yellows) tell us that the cases on each Municipio from the original model with the spatial effect will be overestimated, while the high values (greens) tell us that the spatial effect will be underestimated. In this case, there is still an apparent congruency with the original count number except for a few Municipios where there is a higher concentration of urban population (red circles in the right map).

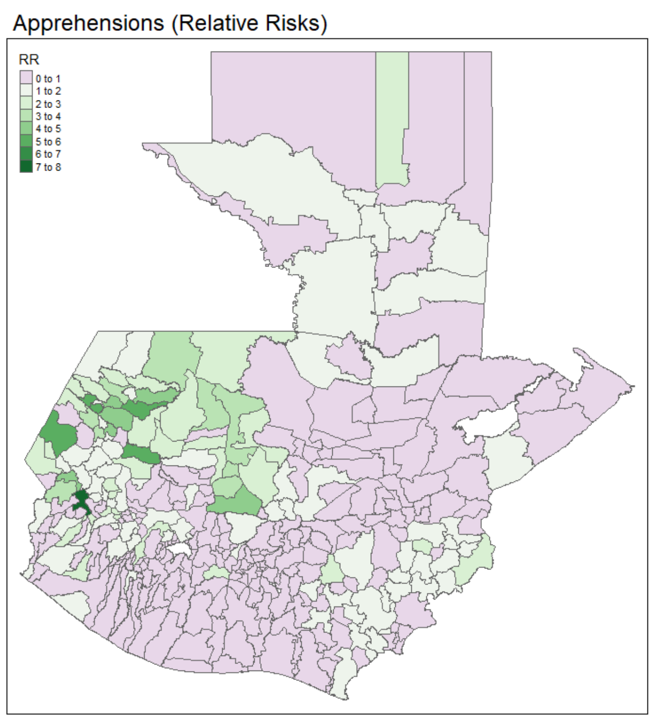


After improving the model under a Poisson family and accounting for expected rates, the spatial random effect was compared again with the total number of apprehensions as an adjustment for each Municipio. This is included in the map below. Where green, the original model without the spatial effect, will be underestimated, and where red, the model will be overestimated. The comparison suggests that the original model underestimates some Municipios of the central west where higher cases could be expected. In contrast, it overestimates the metropolitan area (center) and the central east, where we would expect fewer cases.



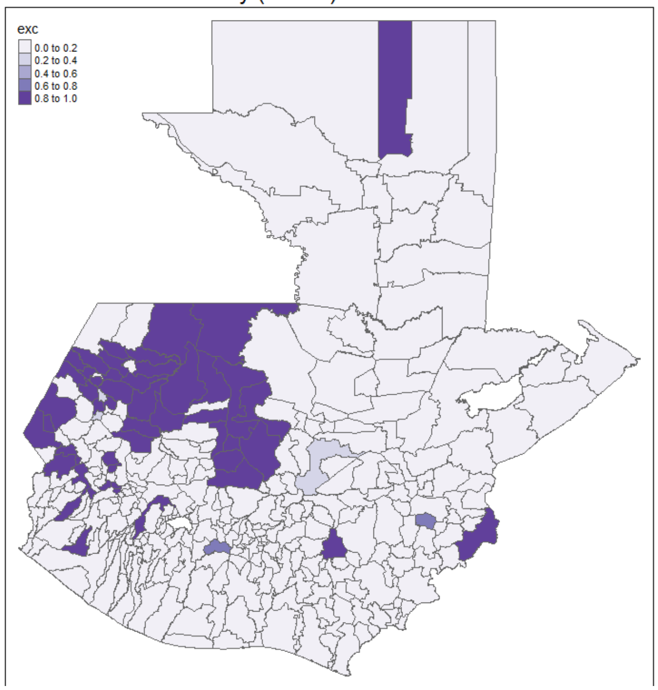


The INLA model generates an average of the posterior distribution based on relative rates (or risk for irregular migration) by fitted values. The map to the right presents the estimated relative values for each Municipio. This means the probability of having more cases than expected. The scale goes from purple with values less than 1, suggesting Municipios will have fewer cases than expected. Higher values than 1 indicate more cases as a rate of Apprehensions on a per capita basis. The variation tells us how high or low is the relative risk. Here, Municipios that were not explicitly high in number Apprehension appear as being subject to increase in posterior years to 2018.



Finally, because irregular migration flows experience dramatic changes over the years based on several conditions, the estimation of relative risks was analyzed with INLA to predict the probability of exceeding a specific threshold. This is also known as exceedance probability. I estimated the probability of having twice the average apprehensions on each Municipio. This threshold allows a better estimation of certain areas compared to others. The results are provided in the map below. Sites with a probability close to 1 are likely to have a relative risk that exceeds twice the average, while areas close to 0 are unlikely to exceed the threshold. Municipios with probabilities of 0.5 are the most unpredictable since it is complex to suggest if their relative risks are above or below 2 with equal probabilities.

Exceedance Probability (RR >2)



DISCUSSION

Applying principal component analysis allowed to prioritize the relevant variables and reduce the possibilities of collinearity in creating the most suitable model. In the same way, comparisons among simple linear and non-spatial INLA models revealed credibility on the residual variance, and low kld convergences but not an improvement on the WAIC, DIC, and quantile values even after introducing priors to reallocate posterior distributions.

The development of two INLA spatial models based on the different measures for the Mayan population (Dichotomous or continuous) provided better results with the continuous variable. This was evident with the reduction of the WAIC, DIC, and higher credibility of the quantile values of the predictors. The consideration of this variable also included a low value of Phi for the Idarea, where 9.3% was the remaining unexplained variation of autocorrelated errors. With the Besag, York, and Mollie approach in a Poisson family, the model improved when comparing the spatial random effect of the Apprehensions in relation to the total count. This suggests the need to review some Municipios where their numbers are underestimated but can exceed their expected annual accounts. In contrast, in Municipios where cases have been overestimated, more rigorous monitoring is required to understand better the interaction with insecurity, cultural, environmental, governance, and economic components.

The last two maps represent the most relevant findings in terms of practical application for the different institutions that work with irregular migration inside and outside Guatemala. The map of Relative Risks highlights the Municipios that will experience (in different scales) more or fewer cases than they expect.

Although some Municipios with the highest values are similar to the count map, it is only the case for some of them, which is a matter of concern, like in the cases of the north of Guatemala and the southwest. Ultimately, the map of exceedance probability emphasizes these risks in a policy-oriented matter. Based on the irregular migration changes triggered by many factors, not only those included in this study, some Municipios might increase their cases and exceed expected thresholds. In this view, estimating exceedance probabilities is necessary to prioritize some Municipios above others.

CONCLUSION

This project provides an understanding of the influence of spatial components and their incidence in the irregular migration of Indigenous people in Guatemala. By defining a process to develop a model that accounts for several factors, the application of Bayesian spatial analysis with the Integrated Nested Laplace Approximations allows estimating the probability to exceed a threshold in some Municipios where most Indigenous populations live. More probabilistic than deterministic, this experience represents an effort to understand the prioritization of areas that experience social inequalities in Guatemala and can be applied to inform strategic policies that seek to mitigate or reduce irregular migration.

BIBLIOGRAPHY

Arriola, G. (2022). Desafíos y oportunidades para Guatemala: Hacia una agenda de futuro. La celeridad del cambio, una mirada territorial del desarrollo humano 2002—2019 (p. 424). Programa de las Naciones Unidas para el Desarrollo (PNUD). https://www.undp.org/sites/g/files/zskgke326/files/2022-07/PNUD_GT_INDH%20Guatemala_WEB_0.pdf

Azose, J. J., & Raftery, A. E. (2015). Bayesian Probabilistic Projection of International Migration. Demography, 52(5), 1627–1650. https://doi.org/10.1007/s13524-015-0415-0

Bijak, J., & Wiśniowski, A. (2010). Bayesian forecasting of immigration to selected European countries by using expert knowledge: Bayesian Forecasting of Immigration. Journal of the Royal Statistical Society: Series A (Statistics in Society), 173(4), 775–796. https://doi.org/10.1111/j.1467-985X.2009.00635.x

Blangiardo, M., & Cameletti, M. (2015). Spatial and spatio-temporal Bayesian models with R-INLA. John Wiley and Sons, Inc.

Dalla Valle, L., Leisen, F., Rossini, L., & Zhu, W. (2020). Bayesian analysis of immigration in Europe with generalized logistic regression. Journal of Applied Statistics, 47(3), 424–438. https://doi.org/10.1080/02664763.2019.1642310

Faubet, P., & Gaggiotti, O. E. (2008). A New Bayesian Method to Identify the Environmental Factors That Influence Recent Migration. Genetics, 178(3), 1491–1504. https://doi.org/10.1534/genetics.107.082560

Faubet, P., Waples, R. S., & Gaggiotti, O. E. (2007). Evaluating the performance of a multilocus Bayesian method for the estimation of migration rates: EVALUATING THE PERFORMANCE OF BAYESIAN METHODS. Molecular Ecology, 16(6), 1149–1166. https://doi.org/10.1111/j.1365-294X.2007.03218.x

Gómez-Rubio, V. (2020). Bayesian Inference with INLA (1st ed.). Chapman and Hall/CRC. https://doi.org/10.1201/9781315175584

International Organization for Migration, & Sweden Government. (2018). Perfil de Gobernanza Migratoria: OIM Guatemala (p. 32). https://mic.iom.int/webntmi/descargas/2018/MGI-GT2018.pdf

Jonas, S., & Rodríguez, N. (2015). Guatemala-U.S. Migration: Transforming Regions. University of Texas Press. https://doi.org/10.7560/760608

Krainski, E., Gómez-Rubio, V., Bakka, H., Lenzi, A., Castro-Camilo, D., Simpson, D., Lindgren, F., & Rue, H. (2018). Advanced Spatial Modeling with Stochastic Partial Differential Equations Using R and INLA (0 ed.). Chapman and Hall/CRC. https://doi.org/10.1201/9780429031892

Kruschke, J. K. (2015). Doing Bayesian data analysis: A tutorial with R, JAGS, and Stan (Edition 2). Academic Press.

Landry, V. (2011). Migración y Cambios Sociales en Guatemala: Familia transnacional y mujer. Revista Sociedad Y Equidad, 1, 1–8. https://doi.org/10.5354/rse.v0i1.10603

Menkos Zeissig, J., & Medina Bermejo, A. (2020). La crisis del Covid-19, su impacto sobre la agenda ODS y el cumplimiento del Plan Nacional de Dessarollo. Instituto Centroamericano de Estudios Fiscales (ICEFI). https://mail.icefi.org/sites/default/files/icefi_-_guatemala_-_la_crisis_del_covid-19_y_su_impacto.pdf

Moraga, P. (2019). Geospatial Health Data: Modeling and Visualization with R-INLA and Shiny (1st ed.). Chapman and Hall/CRC. https://doi.org/10.1201/9780429341823

Nájera Aguirre, J. N. (2017). Migración, fuerza de trabajo y familia, elementos en la definición del espacio transfronterizo México-Guatemala. Entre Diversidades. Revista de Ciencias Sociales y Humanidades, 1(8), 119–150. https://doi.org/10.31644/ED.8.2017.a04

Núñez Medina, G., & López Arévalo, J. (2020). Modelación espacial bayesiana del riesgo de inmigración municipal en Chiapas | RIEM. Revista internacional de estudios migratorios. Revista Internacional de Estudios Migratorios, 9(2), 308–328.

Paredes Orozco, G. (2009). Migración de guatemaltecos a México y Estados Unidos a partir de la Encuesta sobre migración en la frontera Guatemala-México 2004: Un análisis de estrategias migratorias. Migraciones Internacionales, 5(1), 93–124.

Penny, W., Kiebel, S., & Friston, K. (2007). Variational Bayes. In Statistical Parametric Mapping (pp. 303–312). Elsevier. https://doi.org/10.1016/B978-012372560-8/50024-3

Ruiz Marrujo, O. (2001). Los riesgos de cruzar: La migración centroamericana en la frontera México-Guatemala. Frontera norte, 13(25), 7–41.

Schaub, M., & Fletcher, D. (2015). Estimating immigration using a Bayesian integrated population model: Choice of parametrization and priors. Environmental and Ecological Statistics, 22(3), 535–549. https://doi.org/10.1007/s10651-015-0309-8

Smolarek, B. (2007). Causes and Effects of Guatemalan Immigration to the United States. Journal of Undergraduate Research X. https://scholar.googleusercontent.com/scholar?q=cache:-VcG5-6duGgJ:scholar.google.com/+immigration+guatemala&hl=en&as_sdt=0,45

US Agency for International Development (USAID). (2020). Guatemala Country Development Cooperation Strategy (CDCS). https://www.usaid.gov/sites/default/files/2022-05/CDCS-Guatemala-2025-Amended-30Aug2021-Public.pdf

Wang, X., Yue, Y., & Faraway, J. J. (2018). Bayesian regression modeling with INLA. CRC Press.