Table of contents

word count: 2200

Candidate Number: 26684

Introduction

This report presents an analysis of the trends in Emergency Room (ER) visits across the state of California from 2011 to 2019. Utilising Bayesian hierarchical models implemented in R-INLA, I examine the fluctuations in ER visitation rates over the specified period, incorporating both spatial and spatio-temporal dynamics. My analytical framework allows us to account for regional variability and temporal trends simultaneously, providing a robust model of ER visit determinants.

A key focus of my study is the exploration of the impact of environmental and demographic variables on ER visit counts. Specifically, I incorporate the atmospheric concentration of fine particulate matter (PM2.5), a known health hazard, to assess its correlation with the frequency of ER visits. Additionally, I explore the demographic composition of counties, particularly the proportion of the white population, to determine its influence on healthcare utilisation patterns.

By integrating these variables into the models, I aim to uncover influences that drive changes in ER visit counts, offering insights for policymakers, healthcare providers, and public health professionals aiming to optimise emergency medical services and address public health challenges effectively.

Through the analysis, I discovered that incorporating a structured spatial component did not significantly enhance the model's performance. However, by including a temporal component in our models, I was able to identify specific counties exhibiting significant changes in ER visit counts over time.

Summary of the data

Our dataset encompasses a collection of variables recorded across all 58 counties in California from the years 2011 to 2019. This data provides an overview of Emergency Room (ER) visits, alongside key environmental and demographic variables. The primary components of the dataset are outlined as follows:

1. Counts of ER Visits: The dataset includes counts of ER visits segmented into two demographic groups: individuals under 18 years of age and all age groups combined. This allows for analysis of ER usage patterns both by children and the general population.

2. Proportion of White Population: Each county's demographic profile is captured through the variable 'white,' which indicates the proportion of the population that is White. This demographic variable is important for analysing patterns of healthcare access and utilisation across different racial groups.

3. PM2.5 Levels: The dataset measures the amount of PM2.5 in the air in each county. PM2.5 is fine particulate matter with a diameter of less than 2.5 micrometers, known to pose significant health risks and is a major concern in air quality and public health studies.

Exploratory Data Analysis

My exploratory data analysis focused on the distribution and characteristics of ER visit counts for different age groups across California counties. Key findings include:

1. Skewness in Data: Both subgroups - under 18s and all ages - showed a significant right skew in ER visit counts. To address this, I applied a log transformation in our spatial models to normalise the distribution.

2. Adjustments in Spatio-Temporal Models: Instead of transforming counts for spatio-temporal models, I subtracted 15,000 from extreme outliers to reduce their influence, ensuring a more accurate representation of general trends.

3. Overall Trends: Histograms of total counts indicated a downward trend in ER visits over the study period, suggesting potential improvements in public health or changes in healthcare access.

Area Level Spatial Models

For this analysis I will focus only on ER asthma counts for all ages group.

Random Intercept Model

$$R_{ij} = \beta_0 + \beta_1 Year_i + \beta_2 white_j + \beta_3 PM2.5_j + \mu_j + \epsilon_{ij}$$

R_{ij} : the rate of emergency asthma admissions for year i and county j

i : the index representing the year

j : the index representing the county

μ_j : County specific random effect

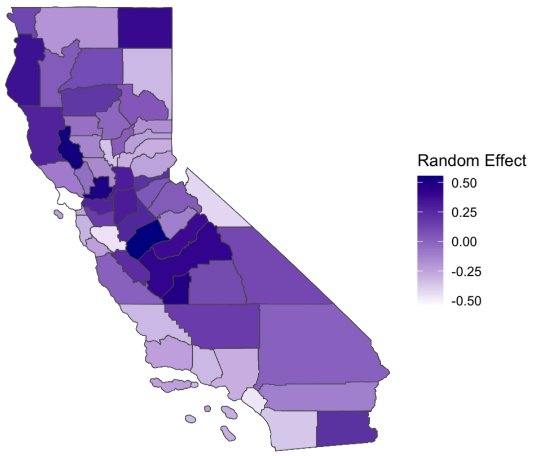
For this model I have used Gaussian link with log population as offset to account for various sizes of counties. I have also modelled the variable $Year$ as a factor.

Table of fixed effects

Term	Mean	SD	2.5%	97.5%
Intercept	-4.357	0.720	-5.773	-2.942
2012	0.071	0.052	-0.032	0.173
2013	0.002	0.052	-0.100	0.104
2014	0.031	0.052	-0.071	0.134
2015	0.044	0.052	-0.058	0.147
2016	-0.067	0.052	-0.170	0.035
2017	0.074	0.052	-0.029	0.176
2018	-0.085	0.052	-0.188	0.017
2019	-0.123	0.052	-0.225	-0.021
White	-0.012	0.009	-0.030	0.006
PM2.5	-0.006	0.013	-0.031	0.019

From the table of fixed effects we can see that the most of the variables are not significant since they include 0 in their credible interval and therefore are not reliable predictors. The model selected the year 2011 as its baseline Year. The rate of emergency asthma admissions appears to be lower in the years 2018 and 2019 relative to 2011.

Random Intercept Model Heat Map



The heat map provided above illustrates that counties coloured in dark blue have baselines that exceed the overall average, whereas those in lighter shades have baselines below the overall average. This discrepancy may be attributed to various factors not accounted for by the model, potentially including socio-economic or agricultural influences.

Analysis of the random effects reveals that Merced, Lake, and Solano counties, which are among the poorest in California, exhibit the highest random effects. Conversely, Marin, Orange, and Santa Clara counties, which are among the wealthiest, display the lowest random effects. This pattern underscores the potential influence of economic status on the observed variations.

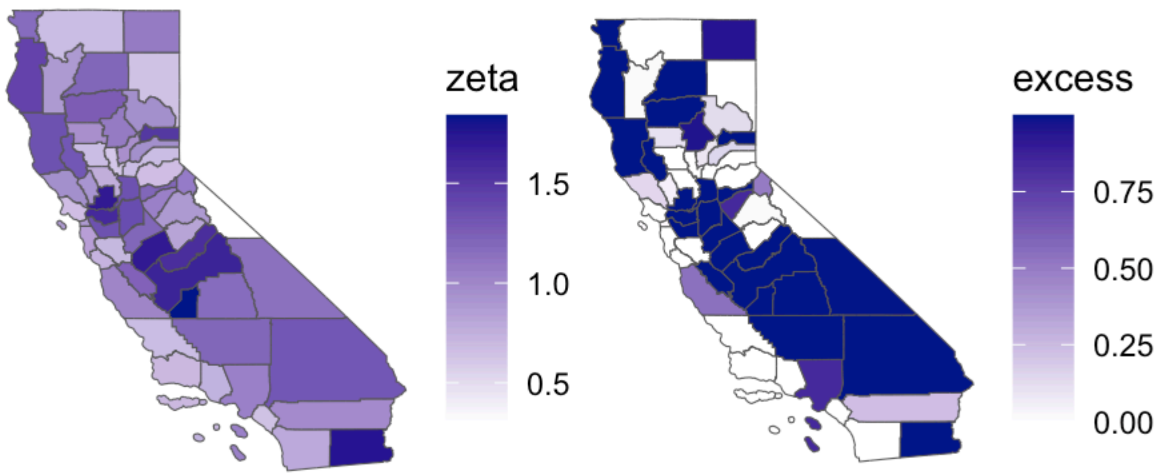
Spatial Random Effects: Disease Mapping

$R_j = \beta_0 + \mu_j + v_j + \epsilon_j$

R_j: Rate of emergency asthma admissions for county j
μ_j: Random effect for county j
v_j: Spatial structure for county j

For this model I have also used Gaussian link with log population as offset to account for various sizes of counties. I have selected the data only for the year 2011 as this was also the year chosen by the previous model.

Disease Mapping Heat Maps



From the heat maps above, we can observe that incorporating a structured spatial component into our model does not significantly change the variation of random effects, if at all. This is confirmed by a very low fractional variation of 0.0097. This intuitively makes sense since asthma is not a contagious disease and is unlikely to affect neighbouring counties. Factors contributing to an increase in ER asthma visits appear to be localised within counties and do not spill over to surrounding counties.

From the heat maps, we can also observe that the southern part of California has a higher excess risk of ER asthma visits. As mentioned before, Southern California has a higher proportion of relatively poor counties compared to the northern part, which could be a factor in why the excess risk is higher for southern counties.

Spatial Random Effects: Ecological Regression

$R_j = \beta_0 + \beta_1 white_j + \beta_2 PM2.5_j + \mu_j + v_j + \epsilon_j$

R_j: Rate of emergency asthma admissions for county j
μ_j: Random effect for county j
v_j: Spatial structure for county j

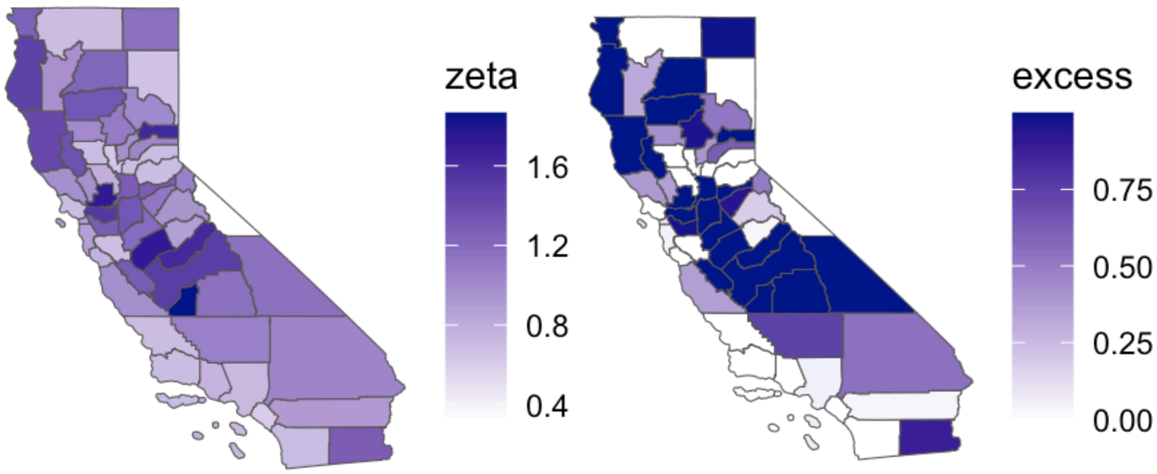
For this model I have also used Gaussian link with log population as offset to account for various sizes of counties. I have selected the data only for the year 2011 as this was also the year chosen by the previous model.

Table of fixed effects

Term	Mean	SD	2.5%	97.5%
Intercept	-5.0405895	0.3171031	-5.6646753	-4.4165073
White	-0.0061654	0.0039495	-0.0139379	0.0016081
PM2.5	0.0088576	0.0057306	-0.0023867	0.0201791

From the table of fixed effect we can see that the variables White and PM2.5 are not significant and therefore not very reliable.

Ecological Regression Heat Maps



By adding predictors to our model that incorporates structured spatial component, the fractional variation has slightly increased to 0.0126 but is still very low. This indicates that spatial component does not improve our model.

The notable difference is in lower excess risk in counties Los Angeles, San Bernardino and Kern which all are in the southern part of California.

Model Comparison

Model	WAIC
Random Intercept (Non Spatial)	-591.1063
Disease Mapping	-331.4855
Ecological Regression	-333.7209

For my analysis, I used the WAIC (Widely Applicable Information Criterion) to compare various models. It is a statistical measure used to compare the fit of various statistical models while accounting for model complexity. WAIC is particularly useful in Bayesian model selection, as it provides a balance between model fit and simplicity by penalising models that are overly complex. The lower the criterion, the better the balance between the two.

From the table of WAIC values above we can see that the model with the lowest WAIC is the Random Intercept model without spatial component. This indicates the the spatial component is not very informative for this particular analysis.

Factors contributing to an increase in ER asthma visits appear to be localised within counties and do not spill over to surrounding counties, making the spatial component uninformative.

Spatio-Temporal Models

In this analysis I have chosen to focus on ER asthma visits for under 18 ages. This is because children are more susceptible to environmental factors such as indoor allergens, pollution, and secondhand smoke, which are significant triggers for asthma episodes.

For these models, I have used a Negative Binomial link because the distribution of counts for ages under 18 is highly right-skewed and takes only positive counts. It is also more appropriate than the Poisson distribution, as the variance is larger than its mean, which addresses the problem of overdispersion. I have used log population as an offset to account for different population sizes between counties.

Random walk of order 2

y_{it}~NegBin(μ_{it}, θ)

$\mu_{it} = E_{it} p_{it}$

$\log(p_{it}) = \eta_{it}$

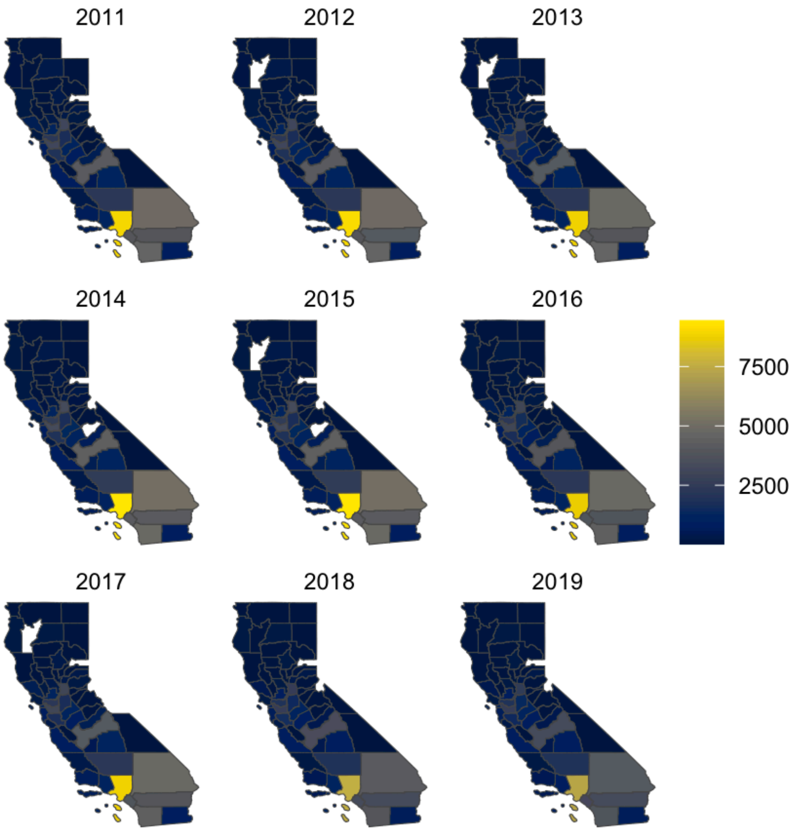
$\eta_{it} = \beta_0 + \beta_1 white + \beta_2 PM2.5 + u_i + v_i + \gamma_t + \phi_t$

u_i: unstructured spatial component
v_i: structured spatial component
γ_i: unstructured temporal component
ϕ_i: structured temporal component

Table of fixed effects

	mean	sd	0.025quant	0.5quant	0.975quant
(Intercept)	-5.90036	0.41576	-6.71826	-5.90038	-5.08235
white	-0.00876	0.00520	-0.01901	-0.00876	0.00147
PM2.5	0.00935	0.00727	-0.00495	0.00935	0.02365

RW2 Heat maps



From the heat maps above, we can see that some data are missing, which explains the missing patches in the heat maps. This is not concerning since the missing counties do not seem to change over time and therefore do not limit my analysis.

The most significant decrease in ER visits appears to be in Los Angeles County between the years 2011 and 2019. There also appears to be a decrease in San Bernardino and San Diego counties. All of these counties are in the southern part of California. This decrease in ER visits might be attributed to improved access to healthcare or to better air quality.

Random walk of order 2 with TYPE I interaction

$y_{it} \sim \text{NegBin}(\mu_{it}, \theta)$

$\mu_{it} = E_{it} p_{it}$

$\log(p_{it}) = \eta_{it}$

$\eta_{it} = \beta_0 + \beta_1 white + \beta_2 PM2.5 + u_i + v_i + \gamma_t + \phi_t + \delta_{it}$

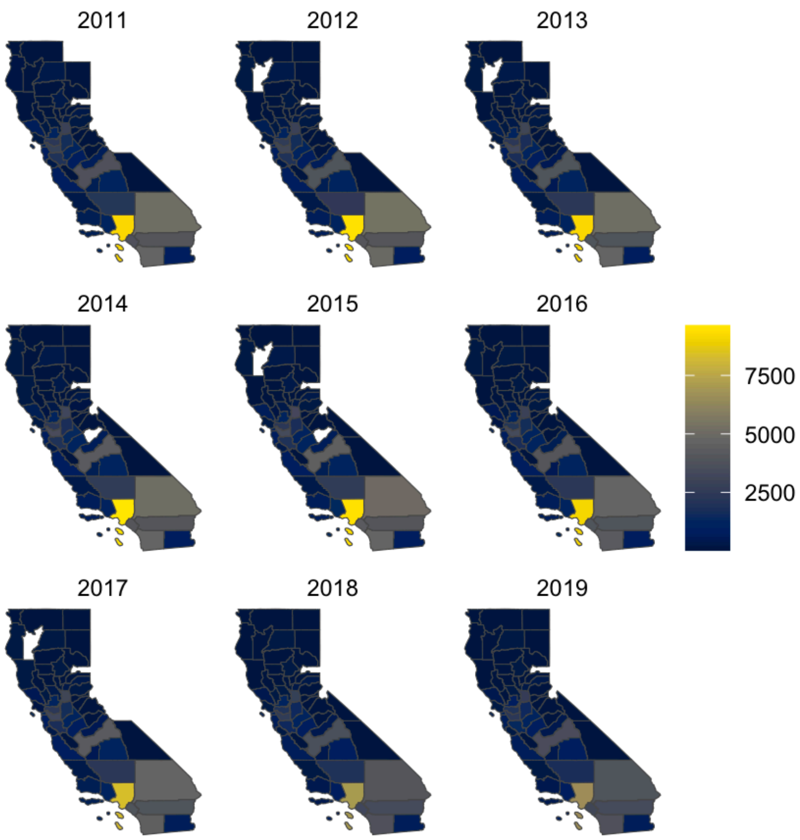
u_i: unstructured spatial component
v_i: structured spatial component
γ_i: unstructured temporal component
ϕ_i: structured temporal component
δ_{it}: interaction term where the unstructured parts of time and space interact

Table of fixed effects

	mean	sd	0.025quant	0.5quant	0.975quant
(Intercept)	-5.90217	0.41427	-6.71774	-5.90219	-5.08647
white	-0.00877	0.00519	-0.01898	-0.00877	0.00144
PM2.5	0.00938	0.00725	-0.00489	0.00938	0.02365

From the table of fixed effects we can see that the estimates do not differ much from estimates of the model without interaction.

RW2 + TYPE I Heat maps



Linear Parametric Trend

$y_{it} \sim \text{NegBin}(\mu_{it}, \theta)$

$\mu_{it} = E_{it} p_{it}$

$\log(p_{it}) = \eta_{it}$

$\eta_{it} = \beta_0 + \beta_1 Year + \beta_2 white + \beta_3 PM2.5 + u_i + v_i + \delta_i Year$

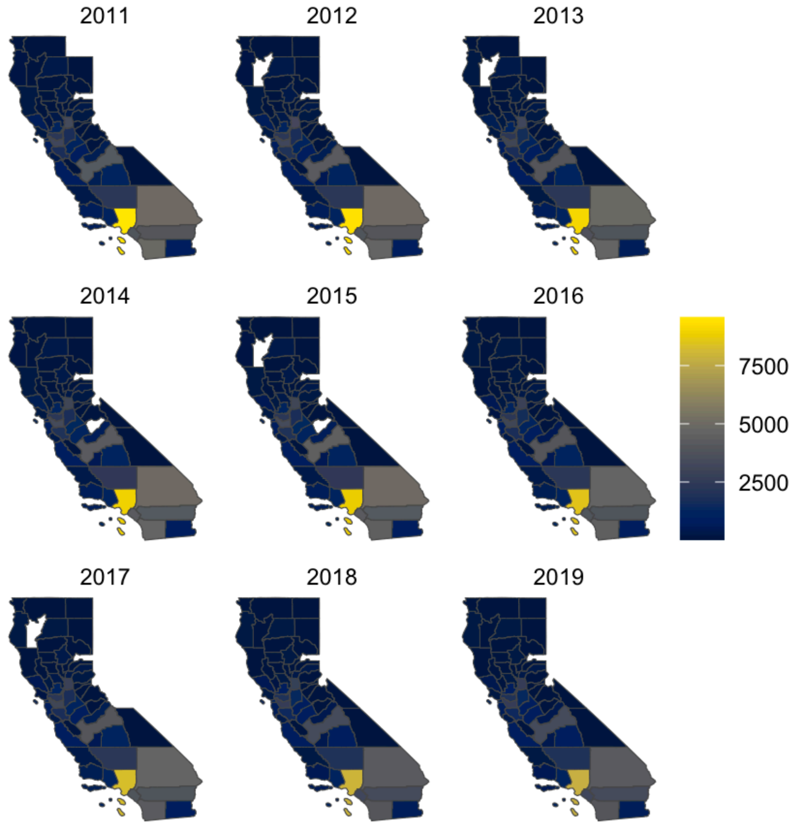
u_i: unstructured spatial component
v_i: structured spatial component
$\delta_i Year$: differential trend

Table of fixed effects

	mean	sd	0.025quant	0.5quant	0.975quant
(Intercept)	48.09097	7.99755	32.37982	48.09709	63.76727
Year	-0.02679	0.00396	-0.03456	-0.02680	-0.01901
PM2.5	0.00930	0.00731	-0.00508	0.00930	0.02368
white	-0.00879	0.00523	-0.01909	-0.00879	0.00150

From the table we can see the Year trend is significant and negative, indicating that the rate of ER asthma visits for under 18s in California was decreasing between 2011 and 2019.

Linear Trend Heat maps



Linear Parametric Trend with TYPE I interaction

$y_{it} \sim \text{NegBin}(\mu_{it}, \theta)$

$\mu_{it} = E_{it} p_{it}$

$\log(p_{it}) = \eta_{it}$

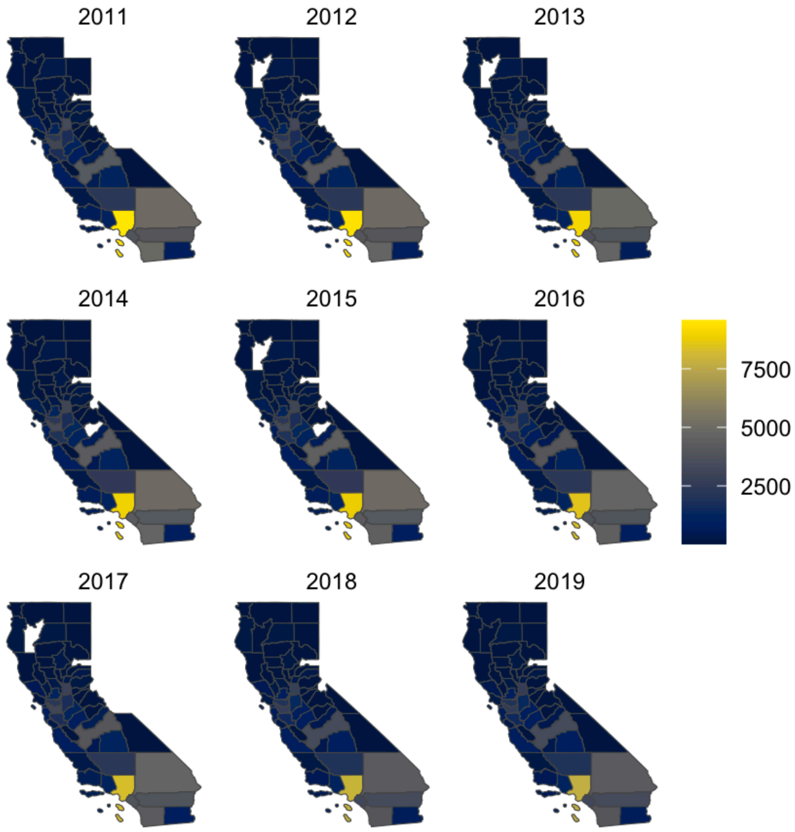
$\eta_{it} = \beta_0 + \beta_1 Year + \beta_2 white + \beta_3 PM2.5 + u_i + v_i + \gamma_t + \phi_t + \delta_i Year + \alpha_{it}$

u_i: unstructured spatial component
v_i: structured spatial component
$\delta_i Year$: differential trend
γ_i: unstructured temporal component
ϕ_i: structured temporal component
α_{it}: interaction term where the unstructured parts of time and space interact

Table of fixed effects

	mean	sd	0.025quant	0.5quant	0.975quant
(Intercept)	48.08923	8.01989	32.33337	48.09548	63.80964
Year	-0.02679	0.00397	-0.03458	-0.02679	-0.01898
PM2.5	0.00919	0.00720	-0.00504	0.00920	0.02332
white	-0.00881	0.00511	-0.01887	-0.00880	0.00124

Linear Trend + TYPE I Heat maps



Model comparison

Table of WAIC values

	Linear Trend	Random walk of order 2
WITH TYPE I interaction	5112.575	4951.999
WITHOUT TYPE I interaction	5112.836	5081.994

From the table of WAIC values, we can see that Random Walk of Order 2 models perform better than Linear models. This was expected as the decline in ER asthma visits does not change much from 2011 to 2015 and then falls quite sharply from 2015 to 2019. Adding a Type I interaction into the model further improves its performance.

Conclusion

In this report, I have applied Bayesian spatial and spatio-temporal models using R-INLA to analyse ER asthma visits among the under-18 age group and all age groups, incorporating predictors such as the proportion of white residents per county and PM2.5 air pollution levels. My spatial analysis indicates that ER visits for asthma are predominantly higher in the southern parts of California, particularly in poorer counties. Interestingly, the minimal fractional variation suggests that the spatial component has limited influence, indicating that factors leading to increased ER visits are localised within individual counties rather than affecting neighbouring areas.

From the spatio-temporal models, I observed a notable decrease in ER visits over time, especially in the southern region, with Los Angeles showing the most significant reduction from 2011 to 2019. This trend may be attributable to improved healthcare access or improvements in air quality, potentially driven by specific health policies implemented during this period aimed at addressing the disproportionately high ER visit rates in these areas. Given these findings, it is reasonable to investigate the specific health policies or infrastructure developments undertaken in these areas during the specified years to better understand their impact on asthma-related ER visits.

My analysis also highlighted that the predictors 'white' and 'PM2.5' did not serve as reliable indicators of ER visit trends, suggesting the need for exploring additional or alternative predictors in future studies. Additionally, while seasonal variations were not examined due to the annual interval of the data, acquiring monthly data could provide insights into how different seasons affect asthma cases, offering a richer temporal resolution for future analyses.

Attempts to integrate Type II and Type IV interactions into the models were hindered by computational constraints, preventing a deeper exploration of these potentially informative aspects. Overcoming these limitations could be a focus for subsequent research efforts, aiming to refine and enhance the predictive power of our models.

In conclusion, my findings underscore the significance of localised health interventions and the potential benefits of targeted environmental and healthcare policies. Further research is needed to understand the effects of specific interventions and to expand the analytical framework, potentially incorporating more granular temporal data and overcoming computational barriers to include more complex model interactions. This will enable a more comprehensive understanding of the dynamics influencing asthma-related ER visits and inform more effective public health strategies.