Predicción espacial de ventas mediante enfoques bayesianos y aprendizaje automático utilizando datos de área

Jhair Santiago Martinez Osorio

Universidad Nacional de Colombia
Escuela de Estadística
Medellín, Colombia
2023

Predicción espacial de ventas mediante enfoques bayesianos y aprendizaje automático utilizando datos de área

Jhair Santiago Martinez Osorio

Trabajo final de maestría presentado para optar al título de:
Magister en Ciencias - Estadística

Director:
Dr. Francisco Javier Rodríguez Cortés

Universidad Nacional de Colombia
Facultad de Ciencias, Escuela de Estadística
Medellín, Colombia
2023

Dedicado a mi familia por su cariño incondicional.

Agradecimientos

Mi más sincero agradecimiento al director de este trabajo, el Dr. Francisco Javier Rodríguez Cortés y a Erick Orozco Acosta, distinguido profesor de la Universidad Simón Bolívar, por su invaluable orientación y apoyo en la realización de mi trabajo final de Maestría.

El Dr. Rodríguez Cortés y el Prof. Orozco Acosta desempeñaron un papel fundamental en mi trayectoria académica, brindándome su experiencia, conocimientos y dedicación. Su guía experta y su disposición para responder a mis preguntas contribuyeron significativamente al éxito de este proyecto.

Resumen

Predicción espacial de ventas mediante enfoques bayesianos y aprendizaje automático utilizando datos de área

Las ventas son un indicador crítico del desempeño de una compañía y las proyecciones acertadas de ventas futuras permiten la toma de decisiones sobre su presupuesto, producción, inventario y expansión. La incorporación de herramientas estadísticas que permitan entender la variabilidad espacio-temporal de las ventas y que apoye la programación logística de la compañía es de gran interés para directivos y administradores. La obtención de pronósticos de ventas a través de la implementación de modelos que tengan en cuenta la autocorrelación espacial y que evolucionan en el tiempo, permite a la compañía definir estrategias de mercadeo y producción con características particulares sobre los lugares en los cuales presta sus servicios.

Este trabajo se centra en la comparación de diferentes técnicas estadísticas y la Aproximación de Laplace Anidada e Integrada (INLA) para la inferencia Bayesiana aproximada en el análisis y pronóstico temporal, espacial y espacio-temporal de las ventas textiles de una compañía en la costa este de los Estados Unidos desde el 2017 hasta el 2022. Se analizan datos de las ventas durante un período cercano a 6 años, incluyendo como covariable demográfica la población de habitantes en los estados involucrados en el estudio. Inicialmente, se aborda el análisis de las ventas desde el punto de vista de las series de tiempo, la cual es forma usual en la literatura como se modela este tipo de datos. En un segundo escenario se considera la componente geográfica de los datos para estimar espacialmente la probabilidad de venta en toda la región de estudio. Finalmente, se estudia la dinámica de la evolución espacial de la probabilidad de ventas a través del tiempo, ajustando un modelo de aprendizaje automático.

Para el estudio del comportamiento temporal de las ventas en la compañía textil, se ajustaron modelos autorregresivos clásicos y su aproximación Bayesiana con INLA, se compararon el ajuste y capacidad de pronóstico sin encontrar diferencias significativas entre los resultados obtenidos con ambas metodologías. Para el análisis espacial de las ventas textiles se compararon los modelos autorregresivo condicional (CAR) y la aproximación Bayesiana basada en ecuaciones diferenciales parciales estocásticas con INLA, el cual presentó mejor rendimiento en el ajuste del modelo de pronóstico espacial en comparación con el modelo CAR tradicional.

Los modelos espacio-temporales están relacionados con problemas en los que se quiere analizar y predecir cómo algo varía en el espacio y/o el tiempo. La metodología Bayesiana de aproximación INLA permite la fácil extensión de los modelos espaciales al caso espacio-temporal, posibilitando la inserción de efectos aleatorios espaciales y temporales, así como efectos de interacción entre el espacio y el tiempo. De otro lado, la relación entre los métodos Bayesia-

nos y aprendizaje automático estadístico es intrínseca, puesto que el análisis Bayesiano es un método de aprendizaje automático estadístico natural. Para el análisis espacio-temporal de las ventas textiles comparamos la metodología Bayesiana de aproximación INLA y su nueva propuesta de aprendizaje automático estadístico basada en la corrección variacional de Bayes de rango bajo (VBC), que utiliza el método de Laplace y, posteriormente, una corrección variacional de Bayes a la media posterior, la cual mostró igual precisión en los ajustes de los modelos y sus pronósticos. Los análisis se realizarán utilizando el lenguaje de programación estadístico R y la librería R-INLA.

Palabras clave: Aproximación de Laplace anidada e integrada, modelos bayesianos, Estadística Espacial, Series de tiempo, Pronósticos.

Abstract

Spatial Sales Prediction Using Bayesian Approaches and Machine Learning Using Area Data

Sales are a critical indicator of a company's performance and accurate projections of future sales allow decision making about its budget, production, inventory and expansion. The incorporation of statistical tools that allow understand the spatio-temporal variability of the sales and that support the company's logistics programming is of great interest to managers and administrators. Obtaining sales forecasts through the implementation of models that will take spatial autocorrelation into account and that will evolve over time, allows the company to define marketing and production strategies with particular characteristics on the places in which it provides its services.

This work focuses on the comparison of different statistical techniques and the Integrated Nested Laplace Approximation (INLA) for approximate Bayesian inference for the temporal, spatial and spatio-temporal analysis and forecasting of textile sales of a company on the East Coast of the United States from 2016 to 2023. Sales data over a 8 year period are analyzed, including as a demographic covariate the population of the states involved in the study. Initially, the analysis of sales is approached from the point of view of time series, which is the usual way in the literature how this type of data is modeled. In a second scenario, the geographic component of the data is considered in order to spatially estimate the probability of sales throughout the study region. Finally, the dynamics of the spatial evolution of the sales probability over time is studied by fitting a machine learning model.

For the study of the temporal behaviour of the sales in the textile company, classic autoregressive models and their Bayesian approximation with INLA were fitted, the adjustment and forecasting capacity were compared without finding significant differences between the results obtained with both methodologies. For the spatial analysis of textile sales, the conditional autoregressive (CAR) models and the Bayesian approximation based on stochastic partial differential equations with INLA were compared, which presented better performance for spatial forecasting compared to the traditional CAR model.

Spatio-temporals are related to problems in which one wants to analyze and predict how something varies in space and/or time models. The easy Bayesian INLA approximation methodology allows the extension of spatial models to the spatio-temporal case, enabling the insertion of random spatial and temporal effects, as well as interaction effects between space and time. On the other hand, the relationship between Bayesian methods and statistical machine learning is intrinsic, since Bayesian analysis is a natural statistical machine

learning method. For the spatio-temporal analysis of textile sales, we compared the INLA Bayesian approximation methodology and its new machine learning proposal based on low-range Bayes variational correction (VBC), which uses the Laplace method and, subsequently, a Bayes variational correction to the posterior mean, which demonstrated greater accuracy in forecasts. The analyzes are carried out using the statistical programming language R and the R-INLA library.

Keywords: Integrated Nested Laplace Approximation, Bayesian modeling, space-time, prediction)

Índice general

Lista de Figuras

Lista de Tablas

1 Introducción

Las ventas son un indicador crítico del desempeño de una compañía. Las proyecciones precisas de las ventas futuras le permiten a una compañía tomar decisiones informadas sobre su presupuesto, producción, inventario y expansión entre otras. En la última década, la predicción de las ventas ha sido objeto de investigación y es de interés tanto en el campo de la estadística como en la ciencia de datos. La mayoría de los estudios han utilizado la teoría de las series de tiempo y técnicas de regresión para modelar los datos históricos de ventas y hacer proyecciones futuras [Shakti et al., 2017, Catal et al., 2019, Chatfield, 2000]. Sin embargo, estas técnicas a menudo no tienen en cuenta la influencia del espacio geográfico y la dependencia espacial en las ventas de una compañía.

Para el modelado de datos de ventas, formas funcionales simples como la regresión lineal, logarítmica o las series de tiempo, pueden ser inadecuadas para capturar una posible no linealidad de los datos, esta última entendida como la relación entre la variable respuesta y las variables predictoras. Sin embargo, en la selección del modelo es crucial examinar el alcance en la precisión de la estimación y la predicción. Los modelos autorregresivos (AR) [Wei, 2018] y los modelos basados en INLA (aproximación de Laplace anidada e integrada) [Lindgren and Rue, 2015], son dos enfoques comunes para modelar series de tiempo. Sin embargo, nuestra intención es identificar ¿cuál es la mejor opción para ajustar un modelo en términos de precisión de la predicción y capacidad de ajuste?

Los modelos AR se han utilizado ampliamente para modelar series de tiempo debido a su simplicidad y eficacia. No obstante, estos modelos pueden ser limitados en términos de capacidad de ajuste y pueden ser inadecuados para datos no estacionarios. Por otro lado, los modelos temporales basados en INLA son más flexibles y pueden manejar datos no estacionarios y modelos jerárquicos complejos. Varios estudios han comparado los modelos AR y los modelos basados en INLA en términos de su capacidad de modelado y precisión predictiva. Por ejemplo, [Zuur et al., 2017] compararon modelos lineales generalizados (GLM) y modelos lineales generalizados de efectos mixtos (GLMM) sobre series de tiempo y modelos bayesianos basados en la metodología INLA con el fin de determinar el modelo que mejor se ajuste a los datos ecológicos utilizados. Esta investigación reveló que los modelos basados en INLA superaron a los modelos lineales generalizados (GLM) y modelos lineales generalizados de efectos mixtos (GLMM) en términos de precisión predictiva y capacidad de ajuste. [Dutta et al., 2022] describe el análisis estadístico bayesiano de series temporales multivariadas de

valores positivos, con aplicación a flujos de datos financieros utilizando la aproximación de Laplace anidada e integrada (INLA) para el modelado bayesiano aproximado rápido. Es por tanto que los modelos basados en INLA tienen ventajas sobre los modelos clásicos de series de tiempo en términos de capacidad de modelado y precisión predictiva para. Sin embargo, la elección del modelo adecuado depende de la naturaleza de los datos y el objetivo de modelado.

El desarrollo de nuevos modelos matemático-estadísticos que se ajusten a los retos científicos actuales y cuya implementación computacional sea eficiente, es uno de los constantes desafíos para las ciencias aplicadas y la industria. En la actualidad se experimenta una acelerada evolución tecnológica, la cual trae consigo la generación de grandes volúmenes de información. Esta información puede contar de manera intrínseca con una etiqueta geoespacial, por ejemplo, las transacciones en un cajero automático, la compra de un producto en una tienda física o por internet, la utilización de una red social en un dispositivo móvil para promocionar un producto, entre otros. La necesidad de obtener conclusiones relevantes provenientes de este tipo de datos ha dado lugar a lo que hoy se conoce como estadística espacial y desde sus inicios se ha evidenciado su gran potencial teórico y aplicado. La evolución de sus métodos, así como algunos de sus resultados, ha permitido crear vínculos e implicaciones con otras áreas del conocimiento [Gelfand et al., 2010]. Para el análisis de datos con características espaciales y que estén estructurados como conteos por regiones geográficas, en la estadística espacial es usual utilizar modelos como los autorregresivos condicionales CAR de sus siglas en inglés (Conditional Autoregressive) [Fotheringham et al., 2003, Bivand et al., 2008], los cuales tienen en cuenta la estructura de vecindad entre las regiones que conforman el área de estudio espacial, mientras que el modelo INLA se basa en la integración numérica y la aproximación Laplaciana. Ambos enfoques tienen ventajas y desventajas en términos de eficiencia y precisión.

En un estudio reciente, [Lawson, 2021] compara los modelos INLA y CAR para la estimación de la mortalidad infantil en un conjunto de datos espaciales en el Reino Unido. Los autores encontraron que los modelos basados en INLA eran más precisos en la estimación de la mortalidad infantil, mientras que los modelos CAR eran más interpretables y permitían una mejor visualización de los patrones espaciales. Otro estudio de [Bivand et al., 2008] comparó los modelos INLA y CAR para ayudar a explicar parte de la variabilidad espacial del riesgo de nacimientos en el condado de Carolina del norte de Estados Unidos. Los autores encontraron que ambos enfoques eran igualmente eficientes, pero que los modelos CAR eran más fáciles de interpretar debido a la naturaleza intuitiva de las matrices de vecindad. En resumen, tanto los modelos espaciales basados en INLA como los modelos CAR tienen ventajas y desventajas en términos de eficiencia y precisión. La elección entre estos dos enfoques depende de las necesidades específicas del análisis y de la complejidad de los datos espaciales.

En relación con modelos espaciales asociados a las ventas [Jank and Kannan, 2005], propusie-

ron un modelo multinomial espacial de la elección del cliente e ilustraron cómo el modelado espacial de las elecciones de los clientes en línea en los mercados geográficos proporciona información útil en el contexto de una combinación de productos y una decisión de precios de un editor de libros en línea. El modelo espacial multinomial explica específicamente las correlaciones espaciales entre las opciones de los clientes entre diferentes formas de productos: impresos y PDF. Los resultados de la estimación obtenidos utilizando datos generados a partir de un experimento en línea muestran que el modelo espacial explica la variación geográfica en muchos de los efectos no observados posiblemente debido a diferencias de ubicación y sensibilidades de precios.

En términos de eficiencia, los modelos basados en INLA han sido destacados por su rapidez computacional y su capacidad para manejar grandes conjuntos de datos espaciales en diferentes trabajos de investigación como [Núñez Medina et al., 2019]. Por otro lado, [Bivand et al., 2008] dice que los modelos CAR son más computacionalmente intensivos que los modelos INLA, ya que requieren de un mayor tiempo de procesamiento debido a la complejidad de las matrices de vecindad. En términos de generales se puede decir que los modelos INLA pueden ser más precisos en la estimación de los parámetros de interés y más eficientes computacionalmente, especialmente cuando se utilizan modelos de datos complejos. Por otro lado, los modelos CAR son más interpretables y permiten una mejor visualización de los patrones espaciales de los datos.

Durante las últimas tres décadas, los métodos bayesianos se han desarrollado demasiado en el campo de las ciencias aplicadas y la industria. Sus principales desafíos se centran en el costo computacional, la complejidad de los modelos y la dimensión de las bases de datos, los cuales siguen siendo una limitación para el desarrollo de la teoría. Recientemente, el uso de campos aleatorios Gaussianos se ha vuelto cada vez más popular, ya que muy a menudo los datos se caracterizan por una estructura espacial y/o temporal que debe tenerse en cuenta en el proceso inferencial. El enfoque INLA se ha desarrollado como una alternativa computacionalmente eficiente a los métodos clásicos y la disponibilidad de software libre permite a los investigadores aplicar fácilmente este método [Blangiardo and Cameletti, 2015]. Los modelos bayesianos espacio-temporal con INLA presentan los paradigmas básicos del enfoque bayesiano y simplifican los problemas computacionales asociados. Además, esta metodología bayesiana de aproximación permite la inserción de efectos aleatorios espaciales, temporales y los efectos de interacción entre el espacio y el tiempo.

Los modelos de área espacio-temporales en particular pueden ser una herramienta valiosa para investigar la distribución espacial y temporal de diversos fenómenos que suceden en nuestra vida diariamente. Aunque los modelos espacio-temporales se han aplicado principalmente en epidemiología para analizar enfermedades crónicas como el cáncer, algunas investigaciones usan estos modelos para buscar la distribución espacial de ciertos delitos como el robo a per-

sonas [Li et al., 2014]. Otro tipo de escenarios en los que se puede adaptar esta metodología es por ejemplo el propuesto en [Blangiardo and Cameletti, 2015], el cual ajusta un modelo bayesiano espacio-temporal implementado bajo INLA para investigar la mortalidad por suicidio en los municipios de Londres durante el período del 1989 al 1993, donde pudieron identificar áreas caracterizadas por un riesgo relativo inusualmente alto o bajo de cometer suicidio en esa población. En el contexto de análisis de datos ecológicos, [Zuur et al., 2017] discute sobre las herramientas frecuentistas que están disponibles para el análisis de datos temporales, espaciales y espacio-temporales, concluyendo que su aplicación es bastante limitada, especialmente si se requieren distribuciones no Gaussianas. Por lo tanto, se sugiere considerar análisis utilizando modelos alternativos, pero estos requieren técnicas bayesianas. Adicionalmente, [Zuur et al., 2017] muestra de manera practica cómo es posible incluir dependencia espacial y espacio-temporal en los modelos de regresión a través de efectos aleatorios correlacionados espaciales (y/o temporales) implementados en R-INLA [Lindgren and Rue, 2015].

Es fácil ver que el aprendizaje automático (ML) se ha convertido en un enfoque ampliamente utilizado en casi todas las disciplinas para resolver una extensa gama de tareas y problemas con datos estructurados y no estructurados, incluidos, entre otros, regresión, agrupación, clasificación y predicción. El aprendizaje automático ha demostrado ser una herramienta poderosa y eficaz en varios campos de aplicación donde los aspectos espaciales son esenciales, incluidos los siguientes: clasificación del uso y la cobertura del suelo [Zhang et al., 2018], caracterización transversal [Law et al., 2020] y cambio longitudinal [Hagenauer et al., 2019], crecimiento urbano [Guan et al., 2005], gestión de desastres [Resch et al., 2018], predicción del rendimiento de cultivos [Masjedi and Crawford, 2020], aparición y propagación de enfermedades infecciosas [Adhikari et al., 2019], análisis de accidentes y transporte [Effati et al., 2015], mapeo de hábitats [Chegoonian et al., 2017] y predicción espacio-temporal [Deng et al., 2017, Deng et al., 2018].

Los datos espaciales exhiben ciertas propiedades distintivas como la escala, la dependencia y heterogeneidad espacial [Nikparvar and Thill, 2021], es por tanto que al igual que en otros enfoques de modelado, debemos ser conscientes de las especificidades que implican estas propiedades cuando realizamos aprendizaje automático en datos espaciales. De hecho, el manejo explícito de estas propiedades espaciales puede mejorar el rendimiento del modelo de aprendizaje automático o agregar información significativa sobre el proceso de aprendizaje de una tarea. Al mismo tiempo, no incluir adecuadamente estas propiedades en el modelo de aprendizaje automático puede afectar negativamente el aprendizaje [Nikparvar and Thill, 2021]. Con el propósito de cuantificar la incertidumbre en los modelos de aprendizaje automático espaciales, se han desarrollado redes neuronales bayesianas [Niraula et al., 2022] y sus aplicaciones se han extendido en varios problemas espacio-temporales [McDermott and Wikle, 2019]. Sin embargo, en el campo del modelado y comprensión de la dinámica de las ventas, el uso de redes neuronales en combinación con la estadística espacial y la inferencia

bayesiana es realmente limitado.

En este trabajo se comparan diferentes técnicas estadísticas y la metodología bayesiana de aproximación en INLA para el análisis y pronóstico temporal, espacial y espacio-temporal de las ventas textiles de una compañía en la costa este de los Estados Unidos desde el 2017 hasta el 2022. Se analizan datos de las ventas durante un período de 6 años, incluyendo como covariable demográfica la población de los estados involucrados en el estudio, la cual es relevante para la interpretación de los resultados. El documento está organizado de la siguiente manera. En el Capítulo 2 se hace una descripción del conjunto de datos de ventas y del proceso de minería (limpieza e imputación). En el Capítulo 3 se presenta una corta descripción sobre la teoría clásica de series de tiempo y la aproximación bayesiana con INLA para inferencia desde un enfoque jerárquico para modelar la serie de tiempo. Se compararon la capacidad de ajuste de ambas metodologías y se realizaron sus pronósticos. En el Capítulo 4 se analiza la variabilidad espacial de las ventas textiles en la costa este y se comparan los ajustes espaciales de los modelos autorregresivo condicional (CAR) y la aproximación bayesiana basada en ecuaciones diferenciales parciales estocásticas con INLA. En el Capítulo 5 se comparan los riesgos relativos de las ventas textiles y se comparan los ajustes de un modelo espacio-temporal clásico implementado con la metodología bayesiana basada en ecuaciones diferenciales parciales estocásticas con INLA y su nueva propuesta de aprendizaje automático estadístico basada en la corrección variacional de Bayes de rango bajo (VBC) [van Niekerk and Rue, 2021], que utiliza el método de Laplace y, posteriormente, una corrección variacional de Bayes a la media posterior. El documento finaliza en el Capítulo 6 con una discusión general y conclusiones.

2 Descripción y preparación de los datos

En este capítulo se describe el proceso de limpieza, depuración e imputación de la base de datos de las ventas diarias de una compañía textil en la costa este de los Estados Unidos. Esta compañía cuenta inicialmente con registros desde el año 2016 hasta el mes de Marzo del año 2023.

Este proceso de limpieza, depuración e imputación comienza con una consulta estructurada en SQL (por sus siglas en inglés Structured Query Language) [Perez, 2011] desde el lenguaje de programación para cómputo estadístico R [R Core Team, 2023], con el propósito de extraer las transacciones que contienen información particular de los clientes como sus direcciones, nombres, fecha en la cual hizo la compra, tipo de producto, marca, referencia, unidades compradas, devoluciones, talla y color como se ve en la Figura **2-1**.

ventas	marca	talla	color	fecha	estado	condado	codigo_postal
1	Fajas Salome	M	Black	2017-05-05	FL	Lake County	32724
1	Lowla	11	Black	2019-06-12	NEW YORK	Westchester County	10604
1	LT Rose	L	Beige	2021-07-24	NY	Orange County	10940-3808
2	Diane & Geordi	L	Black	2023-02-23	Va	Spotsylvania County	22553
2	Diane & Geordi	L	Black	2023-02-23	Va	Spotsylvania County	22553
1	Sonryse	M	Mocha	2020-05-08	NC	Kosciusko County	28398
1	Diane & Geordi	M	Black	2023-03-20	TX	Tarrant County	76051-2523
1	Maria E	L	Beige	2020-08-22	PA	*NA*	18974-3655
1	LT Rose	2XS	Cocoa	2021-05-07	PR	*NA*	00705-3423
1	Sonryse	L	Black	2022-07-03	NV	Clark County	89122-4631
1	Sonryse	L	Black	2022-07-03	NV	Clark County	89122-4631
1	Sonryse	M	Black	2022-10-23	TX	Navarro County	75110-4908
1	Sonryse	L	Black	2021-03-28	TX	Harris County	77396-3814
1	Fajas Salome	XL	Beige	2019-11-28	LA	Orleans Parish	70125
1	Fajas Salome	L	Black	2019-04-26	KY	Henrico County	40475
1	Sonryse	S	Mocha	2023-01-07	FL	Hillsborough County	33566-0562
1	Sonryse	2XL	Cocoa	2020-08-10	FL	*NA*	34208-1442

Figura 2-1: Resultados de la consulta en la base de datos de la compañía sobre las ventas en los diferentes estados de los Estados Unidos desde el 2016 hasta el 2023.

Para este trabajo se tuvieron en cuenta las unidades devueltas debido a que estas fueron despachadas por la compañía por lo tanto cuenta como una venta. También es importante mencionar que se utilizó información poblacional total, por año, de los estados de la costa este de Estados Unidos desde el año 2017 hasta el año 2022, como covariable para los modelos espaciales y espacio-temporales. Esta información se obtuvo de la página Web oficial del censo de Estados Unidos [Census, 2022].

La minería de datos y las técnicas de imputación de datos son herramientas esenciales para el

análisis de los diferentes tipos de información que se generan en una compañía. El almacenamiento, procesamiento y análisis estadístico de los datos, permite a las compañías desvelar patrones y tendencias que ayuden a tomar decisiones informadas sobre sus estrategias de venta, producción y marketing. En el contexto de ventas, la minería de datos puede ayudar a las empresas a identificar patrones de compra de los clientes, así como las tendencias y factores que influyen en las ventas. Sin embargo, en algunos casos, los datos pueden estar incompletos o tener información faltante, lo que puede dificultar el procesamiento y análisis, es en este tipo de situaciones donde las técnicas de imputación son fundamentales. Según un estudio realizado por [Pérez López, 2007], la minería de datos se define como un proceso de descubrimiento de nuevas y significativas relaciones, patrones y tendencias al examinar grandes cantidades de datos. La utilización de esta amplia variedad de técnicas en las compañías, pueden ser aprovechadas para incrementar sus ingresos, recortar costos, mejorar sus relaciones con clientes, reducir riesgos, entre otras.

Para [Medina, 2007] la presencia de datos faltantes es la situación a la que permanentemente se enfrentan investigadores y tomadores de decisiones. Disponer de un archivo de datos completos es ideal, pero aplicar métodos de imputación inapropiados para lograrlo puede generar más problemas de los que resuelve. De allí deriva la necesidad de escoger una buena metodología para abordar las diferentes situaciones que presentan las bases de datos. Los registros de ventas que se utilizarán en este trabajo fueron proporcionados por una compañía textil que tiene su operación centralizada en el territorio de Estados Unidos y la base de datos utilizada contienen información detallada sobre las transacciones de venta de sus productos tales como el número de la orden, el número asociado a cada referencia, marca, estado, longitud, latitud, cantidades vendidas, cantidades devueltas, información del cliente como nombre, dirección, edad y fecha de la compra como se muestra en la Figura **2-1**. Como se mostrará más adelante, la base de datos que se utilizará en este trabajo no cuenta con datos de ventas diarios, semanales e incluso mensuales para algunos condados de Estados Unidos, es por esto la relevancia de estudiar técnicas de imputación y en general técnicas de minería de datos que nos faciliten la utilización de la base de datos de ventas para nuestros análisis.

2.1. Extracción de datos utilizando SQL y R

Para extraer los datos de ventas, se utilizó una consulta de SQL [Perez, 2011] que selecciona las columnas relevantes de la tabla de ventas alojada en un servidor remoto el cual es un servicio de Amazon Web Services (AWS). AWS es una plataforma integral de computación en la nube que incluye ofertas de infraestructura como servicio (IaaS) que es un modelo de servicio en la nube que ofrece recursos de infraestructura bajo demanda, como computación, almacenamiento, redes y virtualización, a empresas y particulares a través de la nube y de plataforma como servicio (PaaS) que es un tipo de modelo de servicio de computación en la nube que ofrece una plataforma flexible y escalable para desarrollar, desplegar, ejecutar y

gestionar aplicaciones [Amazon, 2023].

Mediante el lenguaje de programación R se ejecutó las consultas necesarias para extraer la información de las ventas hechas por la compañía durante los últimos 7 años y se almacenaron en un formato "Data Frame", como se muestra en la Figura **2-1**. Esta consulta de SQL se realizó utilizando el paquete *odbc* "Connect to ODBC Compatible Databases (using the DBI Interface)" de R [Hester and Wickham, 2023], que permite conectarse a una base de datos y obtener, alimentar o transformar las tablas según la necesidad. La base de datos contiene 2.218.104 registros con características de la compra como marca, estado, condado, información geográfica como la longitud y latitud donde fue hecha la compra y además de la fecha en que se hizo la compra y la cantidad.

2.2. Limpieza de datos

Cuando una compañía usa datos para impulsar la toma de decisiones, es fundamental que use datos relevantes, completos y precisos. Sin embargo, los conjuntos de datos suelen contener errores que tienen que eliminarse antes del análisis estadístico. Pueden incluir errores de formato, como fechas, cantidades monetarias y otras unidades de medida escritas incorrectamente que puedan repercutir de forma significativa en las predicciones.

Los valores atípicos son una preocupación particular, pues siempre distorsionan los resultados. Otros errores que suelen verse en los datos son la información duplicada, información faltante y errores tipográficos. Los datos irrelevantes en un algoritmo influyen en el resultado y afectan tanto a la precisión, como la tasa de éxito de las estrategias de "marketing" que contribuyen a los buenos resultados la compañía. Por lo tanto, eliminar los datos irrelevantes es esencial para aportar eficacia al resultado.

La limpieza de datos se refiere al proceso mediante el cual se modifican o eliminan registros de una base de datos que son calificados como incorrectos o poco relevantes para el análisis. El proceso que elimina la inexactitud de los resultados al quitar los datos no deseados ayuda a que la herramienta de análisis ahorre tiempo considerable al aprovechar la depuración de datos antes de ponerla en marcha. También garantiza que los datos sean coherentes, correctos y utilizables. Para esta tarea se utilizó un protocolo de tres pasos como se presenta a continuación.

2.2.1. Identificación de los errores

Antes de procesar los datos y aportar precisión al resultado de los análisis estadísticos, es necesario identificar los errores, lo cual permitirá encontrar la solución óptima en un tiempo mínimo a los problemas presentados en los estudios desarrollados por la compañía en

función de sus ventas. En nuestro caso el error más común que se pudo identificar en la base de datos fue información faltante en las variables de "estado", "longitud", "latitud" y "cantidades". Después de inspeccionar la base de datos se pudo establecer que la causa más común de tener transacciones cuya información del "estado", "longitud" y "latitud" estuvieran vacíos era debido a que la dirección que asocia el cliente pertenece a un "PO Box" (apartado postal) que según el "Servicio Postal de los Estados Unidos" (USPS) es una caja postal que te permite tener una dirección en una oficina de correos en Estados Unidos, para recibir paquetes o documentos. Estás direcciones son actualizadas cada que se realiza un censo territorial por lo que puede existir información desactualizada de las direcciones. Debido a esto, las transacciones que no cuentan con esta información fueron excluidas del análisis, ya que no permite asociar la venta a un estado o condado en particular.

Por otro lado, la razón más común por la cual el campo de "ventas" que hace referencia a las unidades vendidas, se encontrara en cero era debido a la cancelación de la orden por parte del cliente. Al momento de realizar la cancelación, esta orden cambia de estatus y pasa de "Vendida" a "Cancelada". Las unidades que fueron canceladas y cuentan con cantidades en cero fueron excluidas del análisis ya que no fueron despachadas por la compañía como por el contrario las devoluciones las cuales se dijo anteriormente que si serían tomadas en cuenta. De los 2.218.104 registros se excluyeron del análisis en total 8.482 registros que presentaban los problemas que mencionamos anteriormente. Dejando 2.209.622 para continuar con el proceso de limpieza e imputación de los datos.

2.2.2. Homogenización de los registros

Otro aspecto importante en la minería de los datos es verificar si los errores se deben a la falta de homogeneidad en los registros. Cada valor de los datos debe tener un formato estandarizado para su registro y posterior procesamiento. En nuestro caso, se estandarizaron los registros por columna unificando para cada categoría la misma etiqueta. Por ejemplo, en la variable "estado" se encontraron registros como "New York", "New Y", "NY" entre otros, para referirse al mismo estado como se muestra en la Figura **2-2**. Este mismo caso se presentó para la variable "condado", donde se pudieron encontrar registros como "Orange County", "O.C", "OC", "Condado de Orange" para referirse al mismo condado

Dado que las variables pueden contener diversos tipos de datos, según sean, dichos tipos pueden tener un formato y tratamiento diferenciando. Para esto se creó un diccionario y se declararon las variables de tipo carácter, numérica y fechas, como se ve en la Figura **2-3** y así garantizar que el tipo de dato que contiene cada variable fuera el correcto.

Entrada	Estado
NewYork	NY
New York	NY
New Y	NY
NY	NY
Nueva York	NY
nuevayork	NY
nueva york	NY
Nueva Y	NY

Figura 2-2: Homogenización de los nombres de los estados.

Columna	Clase	Tipo
Marca	chr	Character
State	chr	Character
County	chr	Character
Longitude	num	Float
Latitude	num	Float
Fecha	Date	Date
Cantidad	int	Integer

Figura 2-3: Tipos de variables y formatos.

2.2.3. Filtrar de datos duplicados

Los datos duplicados pueden no causar ningún tipo de error, pero generan costo computacional y en algunos estudios específicos pueden introducir sesgo o violar supuestos sobre los modelos utilizados para el análisis. Sin embargo, esto se puede resolver identificando los duplicados antes del análisis de datos. En nuestro caso se utilizó una función *unique* del paquete *base* de R [R Core Team, 2023], el cual permite detectar los registros duplicados teniendo en cuenta las columnas con el identificador, el cual es único de cada transacción, como el número de la orden, el número asociado a cada referencia y la fecha. Con estas tres variables se pudo determinar que no había registros duplicados en la tabla anteriormente extraída mediante una consulta con la herramienta SQL.

2.3. Imputación de datos

El proceso anteriormente descrito de limpieza de datos y manipulación de la base de datos, generó la necesidad de trabajar con grupos de información de las ventas discriminadas por estados, condados y los diferentes años en los que se han recopilado las transacciones de esta compañía. Esta necesidad surge debido a la estructura necesaria para la implementación de los modelos temporales, espaciales y espacio-temporales considerados en este trabajo. Reunir las transacciones por año, por estado y por condado, permitió identificar la existencia de

estados y condados que en algunos periodos de tiempo no tuvieron información de ventas. La causa de estos datos faltantes es que la compañía no tuvo presencia en ciertos periodos de tiempo en algunos estados y condados de Estados Unidos.

Hay muchas técnicas de imputación de datos que se pueden utilizar [Zhang, 2016], entre las más comunes se encuentran imputación de media o mediana, la imputación de valores aleatorios, imputación de vecinos más cercanos, imputación por regresión, imputación de múltiples variables e imputación basada en valores extremos entre otros. Debido a este nuevo panorama se trazó una estrategia para aprovechar la mayor cantidad de datos reales de ventas en todo el territorio de los Estados Unidos, la cual consiste en:

- Filtrar las ventas por estados y determinar cuáles de esos estados tenían información completa de todos sus condados. Entendiendo como información completa a los condados que hayan tenido ventas en cada uno de los años comprendidos entre el 2016 y el 2023.
- Identificar los estados cuya información de ventas en sus condados sea mayor al 70 % en los años analizados.
- Eliminar el año con información de ventas más antigua dado que constituye el conjunto de transacciones con más errores y datos faltantes por estado, por lo tanto, se eliminó el año 2016.

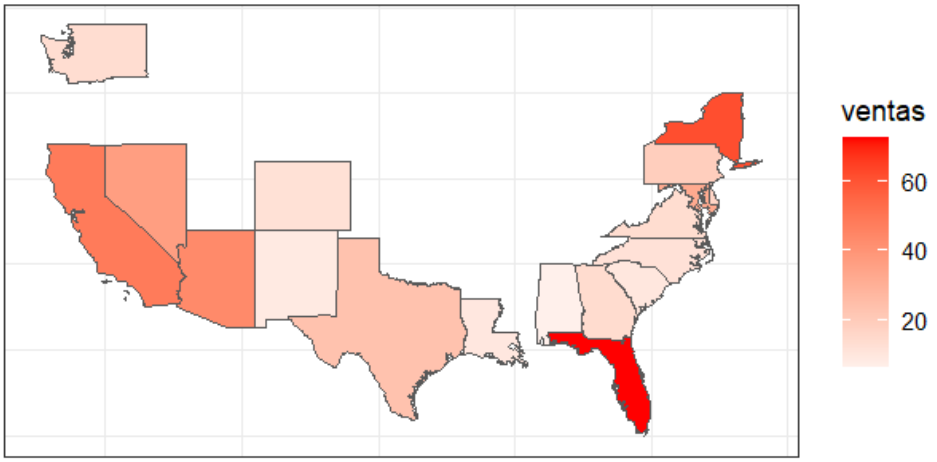


Figura 2-4: Estados de Estados Unidos cuya información de las ventas en sus condados es superior al 70 % desde 2017 hasta 2022.

Después de agrupar la base de datos de ventas por estados y años, se puede notar que se forman dos grandes grupos en las costas que podríamos referirnos a ellos como el "grupo de la costa Este" conformado por los estados de (Florida, New Hampshire, Massachusetts, New

York, Pennsylvania, Connecticut, Rhode Island, New Jersey, Delaware, Maryland, Virginia, North Carolina, Georgia, South Carolina), y el grupo de la costa Oeste conformado por los estados de (California, Nevada, Arizona, Colorado, Utah), como se muestra en la Figura **2-4**.

estado	condado	serie	porcentaje
FL	Hall County	Incompleto	60
FL	Hardee County	Incompleto	60
FL	Loudoun County	Incompleto	60
FL	Norfolk County	Incompleto	60
FL	Onslow County	Incompleto	60
FL	Strafford County	Incompleto	60
FL	Suwannee County	Incompleto	60
GA	Albany County	Incompleto	60
GA	Burlington County	Incompleto	60
GA	Camden County	Incompleto	60
GA	Cumberland County	Incompleto	60
GA	Hampton city	Incompleto	60

Figura 2-5: Filtrado de la información sobre transacciones de ventas superior al 70 %.

Con la intención de mantener la mayor cantidad de datos reales en los modelos, se decidió trabajar con el grupo de estados de la costa Este, ya que tiene un mayor número de estados contiguos lo cual permite construir una matriz de vecindad espacial bien definida y nos proporcionan un escenario ideal para los análisis espaciales y espacio-temporales que se realizarán posteriormente en los capítulos 4 y 5.

año	estado	condado	serie_anual
2016	FL	Loudoun County	Completa
2017	FL	Loudoun County	Completa
2018	FL	Loudoun County	Completa
2019	FL	Loudoun County	Completa
2020	FL	Loudoun County	Completa
2021	FL	Loudoun County	Completa
2022	FL	Loudoun County	Incompleta
2016	FL	Putnam County	Incompleta
2017	FL	Putnam County	Incompleta
2018	FL	Putnam County	Incompleta
2019	FL	Putnam County	Completa
2020	FL	Putnam County	Incompleta

Figura 2-6: Imputación de datos faltantes por condado en la costa Este.

A pesar de que estos estados de la costa Este conforman un gran grupo con una buena cantidad de información de ventas por condado, es necesario imputar la información de las

ventas en los condados donde su cantidad de transacciones no llega al 70 % cómo se ve en las Figuras **2-5** y **2-6**. La técnica de imputación utilizada para llenar los datos faltantes fue la técnica de imputación de vecinos más cercanos. Este es un método para imputar los datos faltantes, que consiste en asignar a cada dato perdido en un individuo un valor obtenido a partir de la información disponible de los k individuos más cercanos o parecidos a este (donantes o vecinos) [Chiapella, 2020].

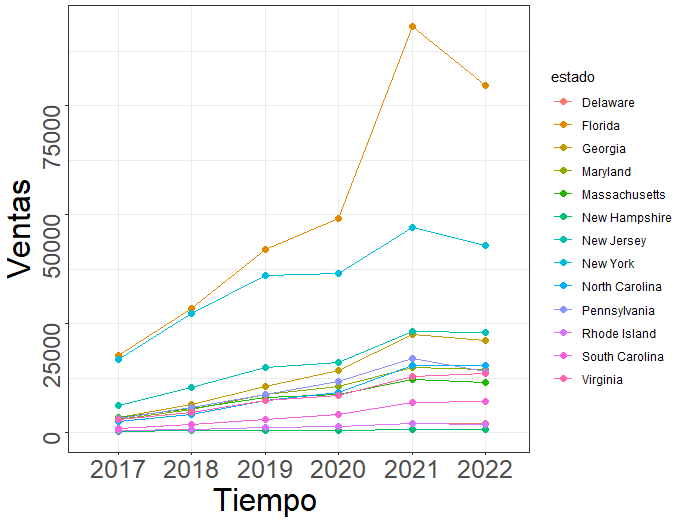


Figura 2-7: Serie de tiempo del número de ventas en costa Este de Estados Unidos por estado desde el año 2017 hasta el año 2022.

Como resultado final, se obtuvieron los datos de las ventas en una compañía en los estados que conforman la costa Este de los Estados Unidos desde el año 2017 hasta el año 2022. Esta nueva base de datos cuenta con 1218 registros de ventas agregados por estado, año y mes en los estados de la costa Este de Estados Unidos como se muestra en la Figura **2-8**. También se contó con información poblacional agregada a nivel anual de estos estados que se utilizó como covariable en la parte del análisis espacial y espacio-temporal de este trabajo, la Figura **2-9** muestra la población agregada por año en los estados de la costa Este de Estados Unidos. En las Figuras **2-7** y **2-8** se muestra la base de datos de ventas depurada e imputada en sus componentes temporal y espacial.

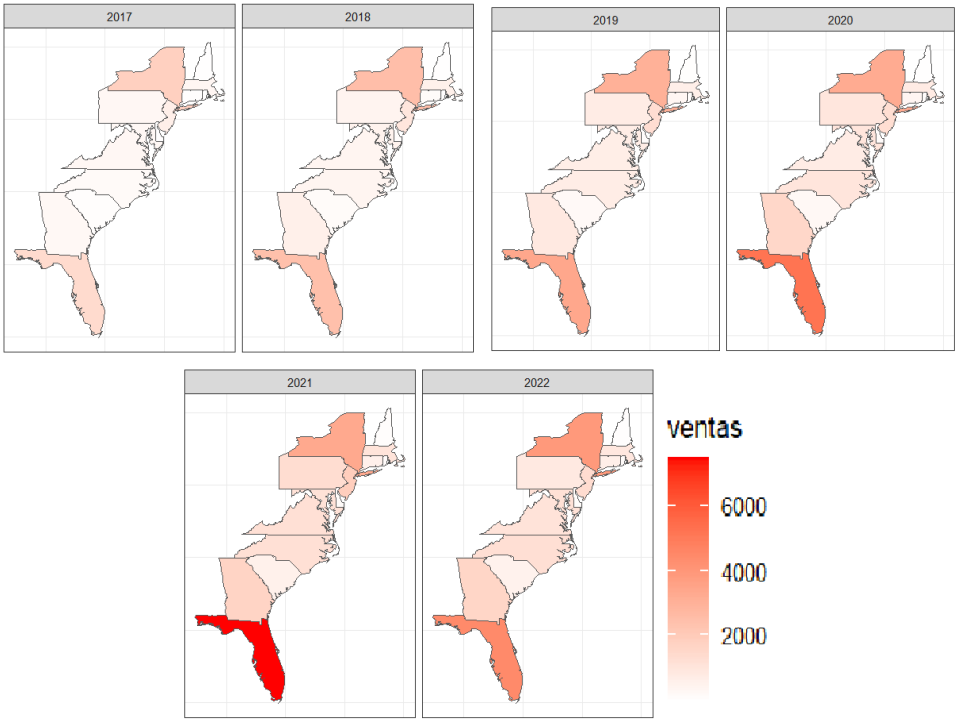


Figura 2-8: Representación espacio-temporal de las ventas por estado en costa Este de Estados Unidos desde el año 2017 hasta el año 2022.

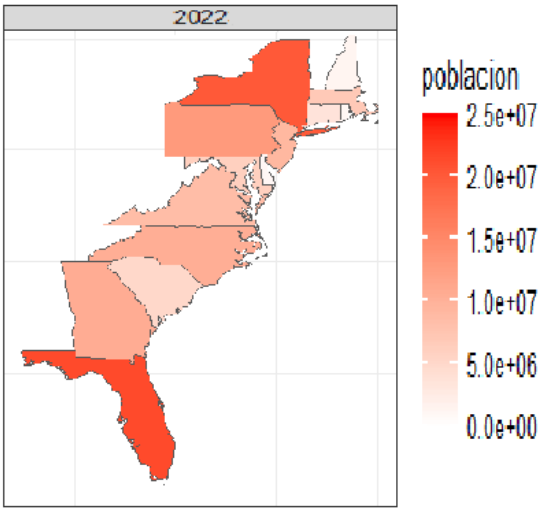


Figura 2-9: Población anual por estados en la costa este de Estados Unidos para el año 2022.

3 Comparación de pronósticos de ventas entre modelos AR y su aproximación con INLA en series de tiempo

En este capítulo se compara uno de los modelos clásicos para series de tiempo, el modelo autorregresivo de orden uno $AR(1)$, contra los modelos autorregresivos de orden uno $RW1$ implementado mediante la aproximación INLA, mediante el ajuste de las ventas textiles de la compañía en la costa este de los Estados Unidos desde el 2017 hasta el 2022. Se evaluó la capacidad de cada modelo para capturar la variabilidad, tendencia en los datos y se compraron las medidas de rendimiento con el ánimo de determinar cuál de los modelos es el más adecuado para la predicción de las ventas. Para este análisis se sumaron las ventas mensuales para cada estado de la costa este de Estados Unidos durante los años comprendidos entre el 2017 y el 2022 como se ve en las Figuras **3-1** y **3-2**.

	Jan	Feb	Mar	Apr	May	Jun	Jul	Aug	Sep	Oct	Nov	Dec
2017	9144	4842	9112	8970	10340	7082	6478	6546	6024	5574	4062	8090
2018	4292	5108	8316	9946	12654	12112	12040	12280	10146	12228	14690	13362
2019	14480	14670	19086	17846	18810	18710	17988	16404	15632	16796	18400	20584
2020	19982	21316	27716	26190	27976	25272	26128	25252	24732	28994	27352	27830
2021	27790	30190	26602	20466	27716	28730	24218	30370	32688	36010	34230	34518
2022	40724	39542	61948	53622	49832	42078	46478	44996	41498	45114	42664	44742

Figura 3-1: Datos mensuales de las ventas textiles de la compañía en la costa este de los Estados Unidos desde el 2017 hasta el 2022.

3.1. Modelo autorregresivo

Una serie tiempo es una secuencia de observaciones, medidos en determinados momentos del tiempo, ordenados cronológicamente y espaciados entre sí de manera uniforme, de esta manera los datos usualmente son dependientes entre sí.

3.1.1. Componentes de una serie temporal

Según [Villavicencio, 2010], el análisis clásico de las series temporales puede descomponerse en diferentes partes que describen diferentes aspectos del proceso subyacente. Los principales componentes son:

- Tendencia: La tendencia representa la dirección general o el patrón de crecimiento o declive a largo plazo en la serie temporal. Puede ser lineal o no lineal.
- Estacionalidad: La estacionalidad se refiere a patrones que se repiten en ciclos de tiempo fijos, como estaciones del año, días de la semana o períodos mensuales.
- Ciclos: Los ciclos son patrones de oscilación que no tienen una duración fija y pueden extenderse más allá del período estacional.
- Componente irregular: También conocido como ruido, este componente representa la variación aleatoria o estocástica que no puede explicarse mediante la tendencia, la estacionalidad o los ciclos.

De estos componentes los dos primeros son componentes determinísticos, mientras que los últimos son aleatorios [Brockwell and Davis, 2002]. Así se puede denotar la serie de tiempo como

$$Y_t = T_t + E_t + I_t, \tag{3-1}$$

donde T_t es la tendencia, E_t es la componente estacional e I_t es el componente aleatorio.

3.1.2. Clasificación descriptiva de las series temporales

Las series de tiempo pueden clasificarse en función de sus características y propiedades. Algunas categorías comunes son:

- Estacionarias: Una serie es estacionaria cuando es estable a lo largo del tiempo, es decir, cuando la media y varianza son constantes en el tiempo. Esto se refleja gráficamente en que los valores de la serie tienden a oscilar alrededor de una media constante y la variabilidad con respecto a esa media también permanece constante en el tiempo.
- No estacionarias: Son series en las cuales la tendencia y/o variabilidad cambian en el tiempo. Los cambios en la media determinan una tendencia a crecer o decrecer a largo plazo, por lo que la serie no oscila alrededor de un valor constante.
- Determinísticas y estocásticas: Las series de tiempo determinísticas siguen patrones predecibles, mientras que las estocásticas tienen una componente aleatoria.
- Univariadas y multivariadas: Las series de tiempo univariadas tienen una sola variable temporal, mientras que las multivariadas involucran múltiples variables interdependientes a lo largo del tiempo.

3.1.3. Procesos estocásticos

Los modelos de series de tiempo se basan en procesos estocásticos, que son secuencias de variables aleatorias ordenadas en el tiempo. Estos procesos pueden caracterizarse mediante sus propiedades estadísticas, como la media, la varianza y la función de autocorrelación [Brockwell and Davis, 2002]. Algunos de los procesos estocásticos comúnmente utilizados son:

- Proceso de media móvil (MA): Un proceso estocástico donde cada observación es una combinación lineal de términos de ruido blanco o "shocks" aleatorios y sus valores pasados.
- Proceso autorregresivo (AR): Un proceso estocástico donde cada observación es una combinación lineal de sus valores pasados y términos de ruido blanco.
- Proceso autorregresivo de media móvil ($ARMA$): Una combinación de procesos AR y MA.
- Proceso de diferenciación integrada autorregresiva ($ARIMA$): Un proceso que incorpora operaciones de diferenciación para hacer estacionarias a las series de tiempo no estacionarias, combinando AR y MA.
- Proceso autorregresivo de componente integrada y media móvil ($ARIMA$): Una combinación de $ARIMA$ y términos estacionales para modelar series de tiempo con estacionalidad.

Proceso estocástico estacionario

Un proceso estocástico se dice que es estacionario si su media y su varianza son constantes en el tiempo y si el valor de la covarianza entre dos periodos depende solamente de la distancia o rezago entre estos dos periodos de tiempo y no del tiempo en el cual se ha calculado la covarianza [Villavicencio, 2010]. Y_t es una serie de tiempo estacionaria si satisface las siguientes propiedades:

$$\mathrm{E}(Y_t) = \mathrm{E}(Y_{t+k}) = \mu, \tag{3-2}$$

$$\mathrm{Var}(Y_t) = \mathrm{Var}(Y_{t+k}) = \sigma^2, \tag{3-3}$$

$$\mathrm{Cov}(Y_t) = \mathrm{E}[(Y_t - \mu)(Y_{t+k} - \mu)] = \gamma_k, \tag{3-4}$$

donde γ_k, la covarianza (o autocovarianza) al rezago k, es la distancia entre los valores de Y_t y Y_{t+k}, que están separados k periodos. En resumen, si una serie de tiempo es estacionaria, su media, su varianza y su autocovarianza (en diferentes rezagos) permanecen iguales sin importar el momento en el cual se midan; es decir, son invariantes respecto al tiempo [Villavicencio, 2010].

Ruido blanco

Un ruido blanco es un caso simple de los procesos estocásticos, donde los valores son independientes e idénticamente distribuidos a lo largo del tiempo con media cero e igual varianza, el cual se denota por ε_t.

$$\varepsilon_t \sim N(0, \sigma^2) \quad \text{con} \quad \text{Cov}(\varepsilon_{t_i}, \varepsilon_{t_j}) = 0. \tag{3-5}$$

Camino aleatorio

El término camino aleatorio se refiere a un proceso estocástico Y_t en el cual la primera diferencia de este proceso estocástico es un ruido blanco, es decir $\nabla Y_t = \varepsilon_t$.

Autocorrelación

La autocorrelación es una medida de dependencia entre variables en la cual los valores que toma la serie en el tiempo no son independientes entre sí, sino que un valor determinado depende de los valores anteriores [Villavicencio, 2010]. Se utilizan dos enfoques para medir la autocorrelación: la función autocorrelación y la función autocorrelación parcial.

La función de autocorrelación mide la correlación entre dos variables separadas por k periodos y se define como

$$\rho_j = \text{Corr}(Y_j, Y_{j-k}) = \frac{\text{Cov}(Y_j, Y_{j-k})}{\sqrt{\text{Var}(Y_j)}\sqrt{\text{Var}(Y_{j-k})}}, \tag{3-6}$$

la cual cuenta con las propiedades de $\rho = 1$, $-1 < \rho_j, < 1$ y $\rho_j = \rho_{-j}$.

Donde el valor de ρ indica el grado de correlación entre las variables, es decir $\rho = 1$ indica que una variable depende de la otra completamente y $\rho = -1$ indica que una variable no depende de la otra variable.

La función autocorrelación parcial mide la correlación entre dos variables separadas por k periodos cuando no se considera la dependencia creada por los retardos intermedios existentes entre ambas y se define como

$$\pi_j = \text{Corr}(Y_j, Y_{j-k} | Y_{j-1}, Y_{j-2}, \ldots, Y_{j-k+1}), \tag{3-7}$$

o de forma equivalente

$$\pi_j = \frac{\text{Cov}(Y_j - \hat{Y}_j, Y_{j-k} - \hat{Y}_{j-k})}{\sqrt{\text{Var}(Y_j - \hat{Y}_j)}\sqrt{\text{Var}(Y_{j-k} - \hat{Y}_{j-k})}}. \tag{3-8}$$

Los modelos autorregresivos se basan en la idea de que el valor actual de la serie Y_t, puede explicarse en función de p valores pasados $Y_{t-1}, Y_{t-2}, \ldots, Y_{t-p}$, donde p determina el número de rezagos necesarios para pronosticar un valor actual [Villavicencio, 2010]. El modelo

autorregresivo de orden p está dado por

$$Y_t = \phi_0 + \phi_1 Y_{t-1} + \phi_2 Y_{t-2} + \cdots + \phi_p Y_{t-p} + \varepsilon_t, \tag{3-9}$$

expresado en términos del operador de retardos,

$$(1 - \phi_1 L - \phi_2 L^2 - \cdots - \phi_p L^p) Y_t = \varepsilon_t, \tag{3-10}$$

por tanto

$$\phi_p(L) Y_t = \varepsilon_t, \tag{3-11}$$

donde ε_t es un proceso de ruido blanco y $\phi_0, \phi_1, \phi_2, \ldots, \phi_p$ son los parámetros del modelo.

Diferenciación

Según [Villavicencio, 2010] diferenciar una serie temporal Y_t en tiempo discreto, consiste en transformar Y_t en una nueva serie $D_t^{(1)}$ definido como:

$$D_t^{(1)} = D(Y_t) = Y_t - Y_{t-1}. \tag{3-12}$$

El procedimiento de diferenciación puede volver a aplicarse sobre una serie previamente diferenciada; obtenemos así las diferencias de segundo orden:

$$D_t^{(2)} = D(D_t^{(1)}) = D_t^{(1)} - D_{t-1}^{(1)}. \tag{3-13}$$

En general, la diferencia de orden m se obtiene como:

$$D_t^{(m)} = D(D_t^{(m-1)}) = D_t^{(m-1)} - D_{t-1}^{(m-1)}. \tag{3-14}$$

En general la diferenciación es una técnica utilizada habitualmente para eliminar la tendencia de una serie temporal [Villavicencio, 2010].

Para la parte descriptiva de los modelos de series de tiempo y con el fin de determinar el modelo más adecuado para desarrollar el análisis estadístico, se realizó un análisis exploratorio tradicional de series de tiempo como se presenta a continuación. La Figura **3-2** muestra la serie mensual de las ventas textiles de la compañía en la costa este de los Estados Unidos desde el 2017 hasta el 2022.

Inicialmente se puede observar en la Figura **3-2** que en la serie mensual de ventas no hay estacionariedad ya que al parecer tiene una tendencia estocástica creciente, pero hay indicios que permiten pensar que la serie de tiempo tiene un componente estacional dado que los primeros meses de cada año presenta un crecimiento en sus ventas.

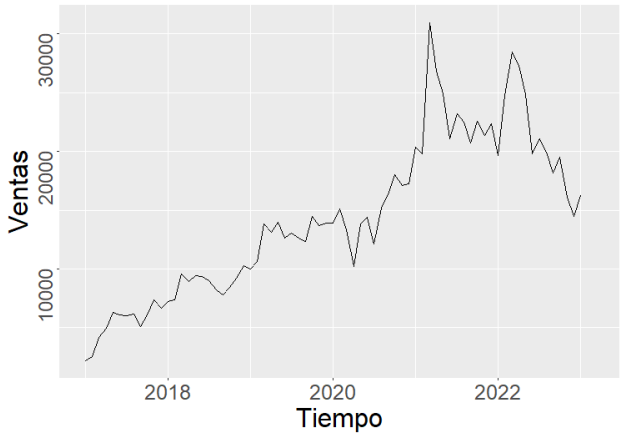


Figura 3-2: Serie mensual de las ventas textiles de la compañía en la costa este de los Estados Unidos desde el 2017 hasta el 2022.

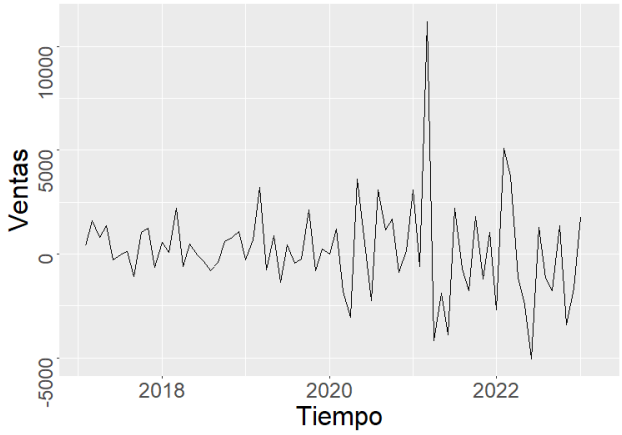


Figura 3-3: Serie mensual de las ventas diferenciada.

Al diferenciar la serie de tiempo original para analizar el comportamiento de la varianza como se muestra en la Figura **3-3**, se observa que la varianza es muy inestable particularmente al final de la serie, por lo tanto, es necesario proponer una transformación de varianza. Utilizamos la prueba de Box-Cox [Sakia, 1992] con el fin de estabilizar la varianza como se puede ver en la figura **3-4**.

Este método consiste en un grupo de transformaciones potenciales usadas en estadística para corregir sesgos en la distribución de errores, para corregir varianzas desiguales y principalmente para corregir la no linealidad en la relación. La transformación potencial está definida como una función continua que varía con respecto a la potencia lambda λ. Para los datos

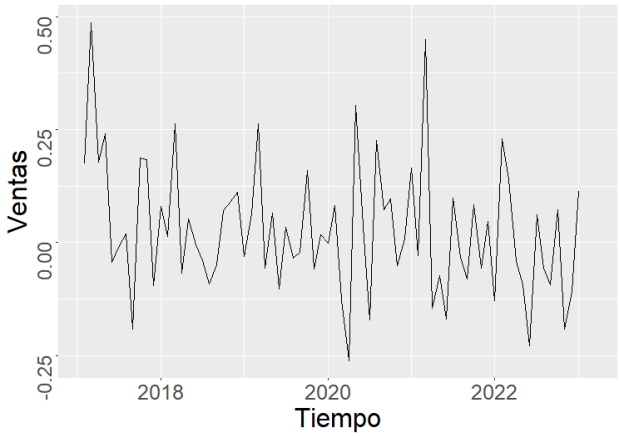


Figura 3-4: Serie mensual de las ventas transformada.

(Y_i'=Y_i^λ) se realiza la transformación de la siguiente manera.

$$Y_i^\lambda = \begin{cases} K_1(Y_i^\lambda - 1), & \text{si } \lambda \neq 0, \\ K_2 \log(Y_i), & \text{si } \lambda = 0, \end{cases} \tag{3-15}$$

donde K_2 es la media geométrica de los valores $Y_1, \ldots, Y_n$,

$$K_2 = (\prod_{i=1}^{n} Y_i)^{1/n} = (Y_1 \cdot Y_1 \cdots Y_1)^{1/n}, \tag{3-16}$$

con K_1 es el parámetro que depende de K_2 y de λ, así

$$K_1 = \frac{1}{\lambda \cdot K_2^{\lambda-1}}. \tag{3-17}$$

Como resultado de la prueba de Box-Cox se propuso una transformación cuadrática de la serie mensual, como se ve en la Figura **3-4**, con el fin de normalizar su comportamiento y observar más fácilmente la tendencia y la estacionalidad de la serie.

Posteriormente se graficaron las funciones de autocorrelación para buscar evidencia del modelo adecuado para ajustar la serie de tiempo como se puede ver en las Figuras **3-6** y **3-7**.

Teniendo en cuenta el hecho de que la función de autocorrelación ρ_k (ACF) decaiga rápidamente como se ve en la Figura **3-6** y se observe un pico alto en el primer rezago de la función de autocorrelación parcial (PACF), Figura **3-7**; constituye una señal de que el mejor modelo para analizar esta serie de tiempo es un AR1 y además es posible que no sea necesario diferenciar la serie de tiempo original a fin de volverla estacionaria. Para corroborar esto último, también podemos utilizar la prueba de raíces unitarias.

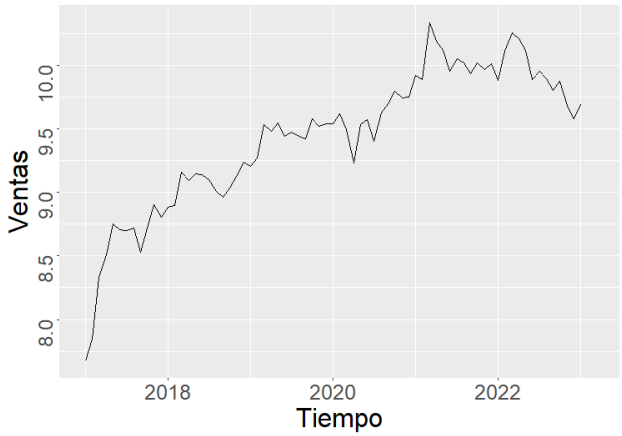


Figura 3-5: Serie mensual de las ventas transformada.

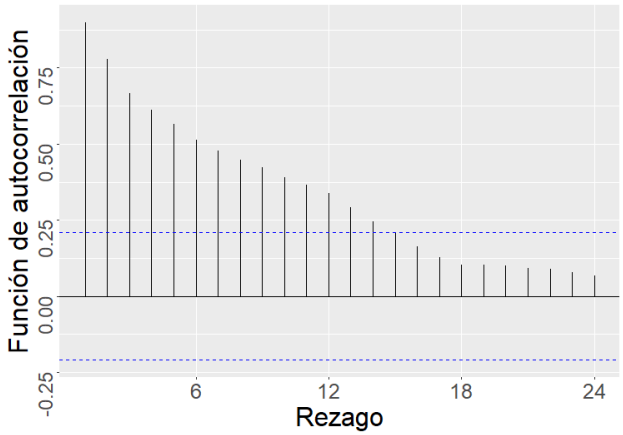


Figura 3-6: Función de autocorrelación de la serie mensual de ventas.

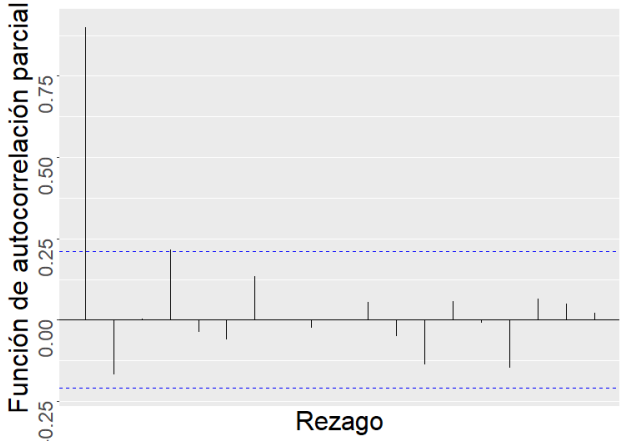


Figura 3-7: Función de autocorrelación parcial de la serie mensual de ventas.

En nuestro caso no es necesario utilizar la prueba de raíces unitarias ya que para efectos de este trabajo el análisis exploratorio anterior es evidencia suficiente para utilizar los modelos autorregresivos de orden uno en el modelamiento de la serie mensual de las ventas textiles de la compañía en la costa este de los Estados Unidos desde el 2017 hasta el 2022.

Validación de supuestos

Uno de los supuestos más importantes de las series de tiempo consiste en probar que los errores del modelo siguen una distribución normal, o no se desvían de la distribución normal. Según [Villavicencio, 2010], los errores en estadística son una medida de la falta de ajuste entre los valores observados y los valores predichos por un modelo estadístico. Son una herramienta esencial para evaluar la precisión del modelo y detectar posibles puntos atípicos o influencias en los datos.

```
        Shapiro-Wilk normality test

data:  res1_est
W = 0.97305, p-value = 0.1246
```

Figura 3-8: Prueba de Shapiro-Wilk sobre los errores de la serie mensual de ventas.

La Figura **3-9** permite pensar que los errores estandarizados del modelo $AR1$ se ajustan a la distribución normal dado que están dentro de la campana de Gauss de la distribución normal [Krithikadatta, 2014] y la Figura **3-10** que representa el qq-plot que sirve para comprobar la normalidad de los errores cuando los puntos se ajustan a la línea recta. Sin embargo, con el fin de corroborar esta hipótesis se realizó la prueba de normalidad de Shapiro [Brzezinski, 2012], que plantea como hipótesis nula que los errores del modelo provienen de una distribución normal.

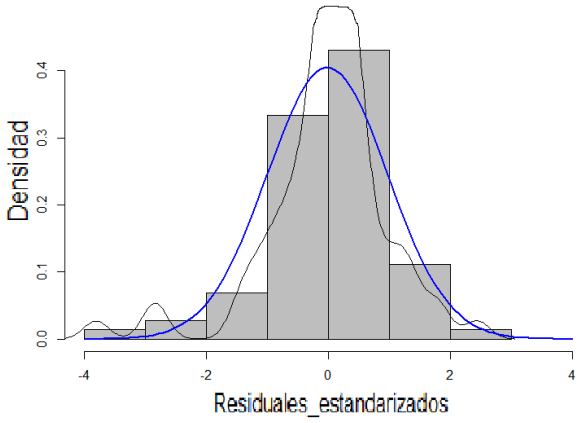


Figura 3-9: Histograma de los errores de la serie mensual de ventas.

Para contrastar esta hipótesis, se eligió un nivel de significancia del 0.05. Dado que esta prueba arrojó un $p-valor$ mayor a 0.05 se pudo aceptar la hipótesis nula, es decir, se puede concluir que los errores del modelo siguen una distribución normal como se ve en la figura **3-8**.

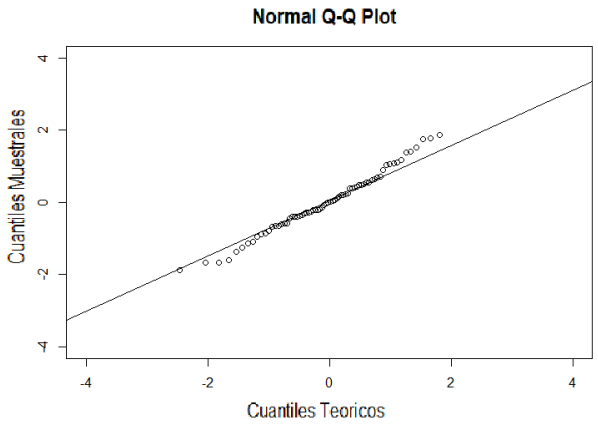


Figura 3-10: QQ-plot de los errores de la serie mensual de ventas.

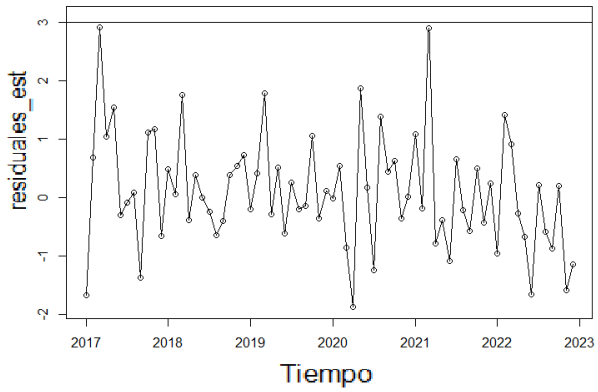


Figura 3-11: Prueba de varianza constante sobre los residuales de la serie mensual de ventas.

En los procesos $AR(1)$ la variable Y_t está determinado únicamente por el valor pasado, esto es Y_{t-1},

$$Y_t = \phi Y_{t-1} + \varepsilon_t, \tag{3-18}$$

donde ε_t es un proceso de ruido blanco con media 0 y varianza constante σ^2 como se puede comprobar con la prueba de varianza constante de los residuales presentada en la Figura **3-11**, ϕ es el parámetro del modelo. Adicionalmente, se supone que el proceso es no anticipante, es decir, el futuro no incluye en el pasado [Wei, 2018]. Una forma alterna de escribir el modelo

$AR(1)$ es la siguiente:

$$Y_t = \phi Y_{t-1} + \varepsilon_t, \tag{3-19}$$

y por tanto

$$(1 - \phi L)Y_t = \varepsilon_t, \tag{3-20}$$

de donde Y_t puede ser expresado como

$$Y_t = \frac{1}{1 - \phi L}\varepsilon_t. \tag{3-21}$$

3.2. Modelo temporal INLA

3.2.1. Inferencia Bayesiana

La aproximación a través de la metodología INLA se basa en la inferencia Bayesiana y utiliza un enfoque jerárquico para modelar la serie de tiempo. La inferencia Bayesiana como se describe en [Blangiardo and Cameletti, 2015], al considerar una variable aleatoria Y, esta puede ser modelada utilizando su función de probabilidad o densidad indexada por un parámetro genérico θ. Sea

$$L(\theta) = p(Y = y|\theta),$$

conocida como función de verosimilitud que especifica la distribución de los datos y bajo un modelo definido por θ. Donde $p(\cdot)$ se utiliza para indicar la función de probabilidad o densidad (dependiendo si se considera una variable aleatoria discreta o continua respectivamente), nos referiremos a esta función como $p(y|\theta)$. El parámetro θ es una variable aleatoria que puede ser modelado a través de una distribución de probabilidad a priori para $p(\theta)$ que refleja el comportamiento que se cree que tiene el parámetro. En nuestro caso Dadas estas dos componentes (verosimilitud y a priori), desde una perspectiva Bayesiana para realizar inferencias se utiliza el teorema de Bayes, por lo tanto

$$p(\theta|y) = \frac{p(y|\theta)p(\theta)}{p(y)}, \tag{3-22}$$

de este modo se obtiene la distribución a posteriori $p(\theta|y)$, que representa el aprendizaje sobre la creencia inicial del parámetro de interés θ después de haber observado los datos. De esta manera $p(y)$ es la distribución marginal de los datos, esta puede ser obtenida como

$$p(y) = \sum_{\theta \in \Theta} p(y|\theta)p(\theta), \tag{3-23}$$

en el caso de que θ sea una variable aleatoria discreta, mientras que en caso de que sea una variable aleatoria continua se puede obtener como

$$p(y) = \int_{\theta \in \Theta} p(y|\theta)p(\theta)\mathrm{d}\theta. \tag{3-24}$$

3.2.2. Aproximación de Laplace Anidada e Integrada

Basado en la definición dada por [Blangiardo and Cameletti, 2015], INLA (Aproximación de Laplace Anidada e Integrada) es un algoritmo determinista para inferencia Bayesiana propuesto por [Rue et al., 2009], diseñado especialmente para modelos Gaussianos latentes y comparado con las cadenas de Markov-Montecarlo (MCMC) [Blangiardo and Cameletti, 2015], proporciona resultados precisos en un tiempo de cálculo más corto. El objetivo de este algoritmo en la inferencia Bayesiana es poder aproximar las distribuciones marginales posteriores para cada elemento del vector de parámetros de interés (debido a que no siempre es posible obtenerlos de forma analítica).

La idea detrás de este método es aproximar la integral de una función de densidad de una variable aleatoria X, tal que

$$\int f(x)\mathrm{d}x = \int \exp(\log(f(x)))\mathrm{d}x, \tag{3-25}$$

a través de una expansión de la serie de Taylor evaluada en $x = x_0$,

$$\log(f(x)) \approx \log(f(x_0)) + (x - x_0) \times \frac{\partial \log(f(x))}{\partial x}\Big|_{x+x_0} + \frac{(x-x_0)^2}{2} \times \frac{\partial^2 \log(f(x))}{\partial x^2}\Big|_{x=x_0}. \tag{3-26}$$

Si consideramos que $x_0 = x^* = \arg\max_x \log(f(x))$ entonces $\frac{\partial \log(f(x))}{\partial x}\Big|_{x+x^*} = 0$, así la aproximación es

$$\log(f(x)) \approx \log(f(x^*)) + \frac{(x-x_0)^2}{2} \times \frac{\partial^2 \log(f(x))}{\partial x^2}\Big|_{x=x^*}. \tag{3-27}$$

La integral de interés es aproximada por

$$\int f(x)\mathrm{d}x \approx \int \exp\left(\log(f(x^*)) + \frac{(x-x_0)^2}{2} \times \frac{\partial^2 \log(f(x))}{\partial x^2}\Big|_{x=x^*}\right)\mathrm{d}x \tag{3-28}$$

$$= \exp(\log(f(x*))) \int \exp\left(\frac{(x-x_0)^2}{2} \times \frac{\partial^2 \log(f(x))}{\partial x^2}\Big|_{x=x^*}\right)\mathrm{d}x, \tag{3-29}$$

donde esta integral puede ser asociada a la densidad de una distribución Normal. En efecto, tomando $\sigma^{2*} = -1/\frac{\partial^2 \log(f(x))}{\partial x^2}$ se obtiene

$$\int f(x)\mathrm{d}x \approx \exp(\log(f(x*))) \int \exp\left(\frac{(x-x*)^2}{2\sigma^2*}\right)\mathrm{d}x, \tag{3-30}$$

se observa el kernel de una distribución Normal con media x^* y varianza σ^{2*}. De aquí se obtiene que la integral evaluada en el intervalo (a, b) es aproximada por

$$\int_a^b f(x)\mathrm{d}x \approx f(x^*)\sqrt{2\pi\sigma^{2*}}(\Phi(b) - \Phi(a)), \tag{3-31}$$

donde $\Phi(\cdot)$ denota la función de densidad acumulada de la distribución $N(x^*, \sigma^{2*})$.

Con este resultado es posible aproximar la distribución marginal $p(y)$ de los datos (y), lo cual a su vez implica tener acceso a una aproximación de la distribución a posteriori $p(\theta|y)$ para el vector de parámetros de interés θ.

3.2.3. Análisis bayesiano de series temporales

En muchas aplicaciones, estamos interesados en modelar una serie de tiempo de observaciones Y_t, $t = 1, \ldots, n$, en función de la información disponible [Ravishanker et al., 2022]. La información puede consistir únicamente en valores observados de y_t, o puede incluir un conjunto de observaciones pasadas y actuales sobre predictores exógenos además de la historia pasada de y_t. Las observaciones $Y^n = (y_1, y_2, \ldots, y_n)'$, recopiladas en puntos de tiempo discretos $t = 1, \ldots, n$, es una realización del proceso estocástico de tiempo discreto Y_t.

La característica más importante de las series temporales es la posible correlación serial entre las observaciones [Ravishanker et al., 2022]. Describir dicha correlación es la esencia del modelado de series temporales. Como señalo [Sakia, 1992], dicha correlación generalmente se describe mediante el uso de estrategias de modelado basadas en observación o basadas en parámetros. El primero incluye modelos como los procesos de media móvil autorregresiva (ARMA) y las cadenas de Markov, mientras que el segundo involucra modelos de estructura latente como los modelos de espacio de estados donde la correlación se crea por la dependencia temporal entre factores latentes. Los modelos controlados por parámetros, debido a su estructura dinámica latente, son capaces de capturar comportamientos no estacionarios en las medias y varianzas. En nuestro desarrollo, nos ocuparemos del análisis bayesiano de los procesos impulsados por observación y por parámetros. El enfoque bayesiano para el modelado de series temporales todavía está impulsado por la noción de "describir la incertidumbre mediante la probabilidadz el cumplimiento de las reglas de probabilidad y el principio de verosimilitud [Ravishanker et al., 2022].

El análisis bayesiano típico de un modelo de series de tiempo seguirá implicando derivar una distribución posterior y predictiva de cantidades desconocidas en el modelo especificado. Debido a la dependencia temporal de los datos de series de tiempo y la posible estructura dinámica de los parámetros del modelo, nuestras inferencias y predictores deben realizarse de forma secuencial. El procesamiento secuencial se puede manejar adecuadamente en el paradigma bayesiano, como se analizará a continuación [Ravishanker et al., 2022].

Para poder considerar los modelos basados en observación y parámetros, descompondremos nuestro vector de parámetros genérico Θ en parámetros estáticos y dinámicos denotados por

θ y $x^n = (x_1, \ldots, x_n)'$, respectivamente. Por lo tanto, definimos $\Theta = (\theta, X^n)$. En el momento t, denotamos los datos observados por $y^t = (y_1, y_2, \ldots, y_t)'$ y la distribución posterior de Θ por $\pi(\Theta|y^t)$. Se puede demostrar mediante el teorema de Bayes que

$$\pi(\Theta|y^t) \ \propto \ \pi(\Theta|y^{t-1})L(\Theta; y_t, y^{t-1}), \tag{3-32}$$

donde $\pi(\Theta|y^{t-1})$ es la distribución posterior de Θ en el momento (t-1) que puede considerarse como la anterior de Θ antes de observar y_t. El término de probabilidad $L(\Theta; y_t, y^{t-1})$ se obtiene evaluando $p(y_t|\Theta, y^{t-1})$ en el valor observado y_t en el momento t. Es importante tener en cuenta que, para los modelos basados en observación, $p(y_t|\Theta, y^{t-1}) \neq p(y_t|\Theta)$ [Ravishanker et al., 2022]. Dado y_t, la distribución predictiva de y_{t+1} viene dada por

$$p(y_{t+1}|y^t) = \int p(y^{t+1}|\Theta, y^t)\pi(\Theta|y^t)\mathrm{d}\Theta. \tag{3-33}$$

la distribución marginal posterior de todo el vector latente X^n dado y^t se obtiene por la marginalización de la distribución posterior de θ como

$$\pi(x^t|y^t) = \int \pi(x^n, \theta|y^t)\mathrm{d}\theta. \tag{3-34}$$

Las distribuciones posteriores de los componentes de x_n, es decir, $\pi(x_\tau|y^t), \tau = 1, ..., n$ se pueden obtener de manera similar. La distribución del componente latente actual x_t, $\pi(x_t|y^t)$ es denominado como la distribución de filtrado. Los marginales restantes para x_τ, donde $\tau < t$ y $\tau > t$ se conocen como distribuciones de suavizado y distribuciones de pronóstico, respectivamente.

En el análisis secuencial de datos de series temporales, a menudo el interés se centra en la distribución de filtrado de x_t que se puede obtener a partir de la actualización secuencial de x_t y θ como

$$\pi(x_t, \theta|y^t) \ \alpha \ \pi(x_t, \theta|y^{t-1}) \ L(x_t, \theta; y_t, y^{t-1}). \tag{3-35}$$

La distribución marginal posterior de x_t obtenida de $\pi(x_t, \theta|y^{t-1})$, es decir, $\pi(x_t|y^{t-1})$, es la distribución prevista de x_t.

3.2.4. Estructura INLA para series temporales

Sea y_n los datos observados (vector de respuesta), x_n un vector de variables Gaussianas latentes que describen el modelo y θ sea un vector de hiperparámetros. La dimensión del vector de estado x_n es grande, típicamente siendo n en problemas de series de tiempo, mientras que la dimensión de los hiperparámetros θ es pequeña, generalmente por debajo de diez. INLA usa un marco jerárquico para representar la subyacente estructura probabilística, como se analiza en el siguiente ejemplo.

Modelo de campo aleatorio gaussiano de Markov

Considere el modelo a nivel local

$$y_t = x_t + v_t; v_t \sim N(0, \sigma_v^2), \tag{3-36}$$
$$x_t = x_{t-1} + w_t; w_t \sim N(0, \sigma_w^2). \tag{3-37}$$

donde, v_t y w_t son errores aleatorios que se supone que siguen distribuciones normales de media cero con varianzas desconocidas σ_v^2 y σ_w^2 respectivamente.

Este modelo es uno de los modelos dinámicos más simples para una serie temporal univariada y_t y es apropiado cuando la serie temporal muestra un nivel que cambia lentamente a lo largo del tiempo. Donde las distribuciones de X_t son

Distribución a priori de x_t. Después de que lo veamos en el momento $t-1$, podemos obtener la distribución posterior de x_{t-1} dado $y^{t-1} = (y_1, ..., y_{t-1})$, es decir, $\pi(x_{t-1}|y^{t-1})$. Esto lleva a la distribución a priori de x_t antes de observar y_t; es decir, $\pi(x_t|y^{t-1})$.

Distribución a posteriori de x_t. Una vez que los datos en el momento t, es decir, y_t, están disponibles, actualizamos nuestra creencia sobre x_t, es decir, obtenemos la posterior de x_t dado y^t, denotada por $\pi(x_t|y^t)$.

La implementación original de INLA asume que el vector latente x_n está definido por un campo aleatorio Gaussiano de Markov (GMRF) [Rue and Held, 2005]. El modelo GMRF lo podemos expresar como una jerarquía de tres niveles con los datos observados y^n, el proceso latente no observado x^n y los hiperparámetros aleatorios desconocidos $\theta = (\sigma_v^2, \sigma_w^2)'$ que constituyen los tres niveles de la jerarquía. La distribución de datos puede pertenecer a la familia exponencial y podría ser normal, binomial, gamma, pero el proceso latente es siempre gaussiano. No es necesario que las distribuciones de los hiperparámetros sean gaussianas [Ravishanker et al., 2022]. Por ejemplo, en este ejemplo, podemos suponer que los hiperparámetros tienen las siguientes distribuciones:

$$\frac{1}{\sigma_v^2} \sim Gamma(a_v, b_v) \text{ y } \frac{1}{\sigma_v^2} \sim Gamma(a_w, b_w), \tag{3-38}$$

donde (a_v, b_v) y (a_w, b_w) son parámetros especificados para estas distribuciones anteriores. Podemos representar el modelo en una estructura jerárquica como

$$y^n|x^n, \theta \sim N(x^n, \sigma_v^2 I_n) \text{ modelo de datos,} \tag{3-39}$$

$$x^n|\theta \sim \prod_{t-1}^{n} \pi(x_t|x_{t-1}, \theta) \text{ a priori de Markov,} \tag{3-40}$$

$$\theta \sim \pi(\theta) \text{ hiperpriori,} \tag{3-41}$$

donde $\pi(x_t|x_{t-1}, \theta)$ es $N(x_{t-1}, \sigma_w^2)$, I_n es la matriz de identidad n-dimensional y $\pi(\theta)$ viene dado por el producto de dos densidades gamma [Ravishanker et al., 2022].

Para simular los modelos 3-37 R-INLA utiliza información en las ecuaciones de estado similarmente a la formula utilizada por la función lm() de R. Se puede especificar efectos del tiempo y en general cualquier efecto estructurado usando la función f() dentro de la formula. Entre los efectos estructurados para series de tiempo univariados son iid, rw1 y rw2.

3.3. Evaluación de los modelos

La evaluación de la precisión de los modelos de series de tiempo desempeña un papel fundamental en la toma de decisiones informadas y la mejora de las predicciones.

Entre las numerosas métricas disponibles para evaluar la calidad de los ajustes de modelos de series de tiempo, dos de las más ampliamente utilizadas son el Error Porcentual Absoluto Medio (MAPE) y el Error Cuadrático Medio (MSE). Estas métricas proporcionan una visión cuantitativa de la discrepancia entre los valores observados y los valores predichos por el modelo.

3.3.1. Error porcentual absoluto medio (MAPE)

El error porcentual absoluto medio (MAPE) es una medida que evalúa la precisión de un modelo en términos porcentuales. Esta métrica se utiliza combinada en el contexto de pronósticos y series de tiempo para evaluar cuánto se desvían las predicciones del modelo en promedio con respecto a los valores reales. El MAPE se calcula de la siguiente manera:

$$\text{MAPE} = \frac{1}{n} \sum_{t=1}^{n} \left| \frac{Y_t - \hat{Y}_t}{Y_t} \right| \times 100\,\%, \tag{3-42}$$

donde

- Y_t es el valor real en el tiempo t.
- $\hat{Y}_t$ es el valor predicho por el modelo en el tiempo t.

- n es el número total de observaciones en la serie de tiempo.

El MAPE expresa la discrepancia relativa entre los valores observados y los valores predichos como un porcentaje promedio. Cuanto más bajo sea el MAPE, mayor será la precisión del modelo. El concepto de MAPE se ha utilizado ampliamente en el campo del pronóstico y la evaluación de modelos de series de tiempo. Aunque no se le atribuye a un autor específico, su aplicación en la evaluación de pronósticos es bien conocida en la literatura.

3.3.2. Error cuadrático medio (MSE)

El Error cuadrático medio (MSE) es otra métrica ampliamente utilizada para evaluar modelos de series de tiempo. A diferencia del MAPE, el MSE mide la discrepancia en términos absolutos, lo que significa que no considera la magnitud de los valores. El MSE se calcula de la siguiente manera:

$$\text{MSE} = \frac{1}{n}\sum_{t=1}^{n}(Y_t - \hat{Y}_t)^2, \tag{3-43}$$

donde:

- Y_t es el valor real en el tiempo t.
- $\hat{Y}_t$ es el valor predicho por el modelo en el tiempo t.
- n es el número total de observaciones en la serie de tiempo.

El MSE calcula el promedio de los cuadrados de las diferencias entre los valores reales y los valores predichos. Cuanto menor sea el valor del MSE, menor será la discrepancia entre las predicciones del modelo y los datos reales. El Error cuadrático medio (MSE) es una métrica clásica que se ha utilizado en estadísticas y análisis de series de tiempo. Aunque no tiene un autor específico, su aplicación en la evaluación de modelos estadísticos es ampliamente reconocida. Tanto el MAPE como el MSE son métricas importantes para evaluar la precisión de los modelos de series de tiempo. El MAPE se destaca por expresar la discrepancia en términos porcentuales, lo que facilita la interpretación de la precisión del modelo en un contexto relativo. Por otro lado, el MSE proporciona una medida de discrepancia en términos absolutos, lo que puede ser útil en situaciones donde se requiere una métrica insensible a la magnitud de los valores.

Se compararon los resultados de estas medidas de precisión para cada modelo como muestra la Tabla **3-1** y la Tabla **3-2**. Donde $AR1$ es el resultado de la aplicación del modelo tradicional y $RW1$ el resultado de la aplicación de modelos autorregresivos de orden uno mediante la implementación de la aproximación de la metodología INLA.

MSE	
$AR1$	$RW1$
25,750	35,727

Tabla 3-1: Error cuadrático medio sobre los ajustes de la serie mensual de ventas.

Estos modelos fueron elegidos gracias al análisis descriptivo realizado en las secciones anteriores donde en las Figuras **3-6** y **3-7** se pudo determinar que el modelo que más podría ajustar los datos de ventas era el modelo AR1.

MAPE	
$AR1$	$RW1$
9,083	34,009

Tabla 3-2: Error porcentual absoluto medio sobre los ajustes de la serie mensual de ventas.

Es fácil ver que la comparación de las medidas de precisión utilizadas en esta sección para medir el desempeño en el ajuste de los modelos de las ventas mensuales tanto de los modelos clásicos de series de tiempo como los modelos bajo la implementación INLA, tienen grandes diferencias entre ambas metodologías, es posible pensar que para efectos de este trabajo el modelo clásico de series de tiempo tiene mejor rendimiento. Es posible que con un conjunto de datos más grande esta comparación arroje diferencias más significativas en la comparación de estos resultados.

3.4. Análisis de los pronósticos de ventas

Los modelos de series de tiempo son una herramienta útil y práctica para el análisis y la predicción de datos de series temporales y a su vez la modelación Bayesiana, ofrece flexibilidad y gran capacidad para manejar diferentes fuentes de incertidumbre. En este trabajo se encontraron diferencias entre los métodos utilizados para ajustar y predecir las ventas utilizando únicamente como insumo el factor temporal de los datos de las ventas en la costa este de Estados Unidos, siendo el modelo con mejor desempeño en el ajuste el modelo clásico de series de tiempo.

Estos pronósticos se obtuvieron de dos maneras diferentes, primero para el modelo clásico de series de tiempo se utilizó el paquete de R "$forecast$" [Hyndman and Khandakar, 2008], el cual automáticamente ajusta el modelo que le indicamos y pronostica el número de periodos que nosotros indiquemos, en este caso los primeros 3 meses del año 2023. Este paquete implementa pronósticos automáticos utilizando suavizado exponencial [Hyndman et al., 2002],

modelos ARIMA, el método Theta [Assimakopoulos and Nikolopoulos, 2000], splines cúbicos [Hyndman et al., 2005], así como otros métodos de pronóstico comunes.

Según [Hyndman and Khandakar, 2008] se pueden combinar las ideas anteriores para obtener un algoritmo de pronóstico automático robusto y ampliamente aplicable. Los pasos involucrados se resumen a continuación.

- Para cada serie, aplicar todos los modelos que sean adecuados, optimizando los parámetros (tanto los de suavizado como la variable de estado inicial) del modelo en cada caso.
- Seleccionar el mejor de los modelos según el AIC.
- Produzca pronósticos puntuales utilizando el mejor modelo (con parámetros optimizados) para tantos pasos adelante como sea necesario.
- Obtener intervalos de predicción para el mejor modelo utilizando los resultados analíticos de [Hyndman et al., 2005], o simulando trayectorias de muestreo futuras para $y_{n+1}, ..., y_{n+h}$ y encontrar los percentiles $\alpha/2$ y $1-\alpha/2$ de los datos simulados en cada horizonte de pronóstico.

[Hyndman et al., 2002] aplicaron esta estrategia de pronóstico automático a los datos de competencias [Makridakis et al., 1982] y a los datos de competencia IJF-M3 [Makridakis and Hibon, 2000] utilizando un conjunto restringido de modelos de suavizado exponencial, y demostraron que la metodología es particularmente bueno para pronósticos a corto plazo (hasta aproximadamente 6 períodos de anticipación), y especialmente para series estacionales a corto plazo (superando a todos los demás métodos en la competencia para estas series).

Para el modelo implementado bajo la metodología INLA, obtener el pronóstico consiste en agregar a la serie de tiempo un conjunto de fechas igualmente espaciadas en el tiempo las cuales se desean predecir, es decir, si la serie de tiempo tiene información mensual hasta el mes diciembre del año 2022 se debe agregar los meses del año 2023 que se quieren pronosticar con valores nulos y al ejecutar el algoritmo, este pronosticará automáticamente estos meses. En cuanto al tiempo de cómputo y la optimización de parámetros, los modelos funcionaron casi de la misma manera teniendo en cuenta la cantidad de datos que se utilizó para este análisis.

En las Figuras **3-12** y **3-13** podemos observar tanto los ajustes de los modelos como los pronósticos realizados bajo la metodología clásica de series de tiempo y el pronóstico bajo la implementación INLA.
Dada la estacionalidad de los primeros meses de cada año en el último par de años, es factible pensar que el pronóstico hecho bajo la metodología clásica ajusta un poco más la realidad

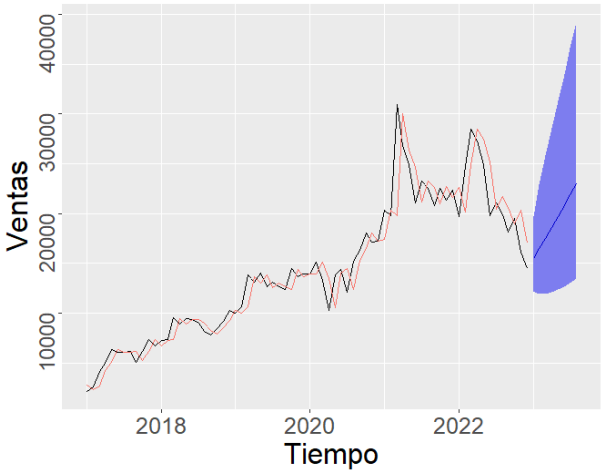


Figura 3-12: Ajuste y pronóstico de ventas mensuales de la compañía en la costa este de los Estados Unidos desde el 2017 hasta el 2022 con el modelo $AR(1)$.

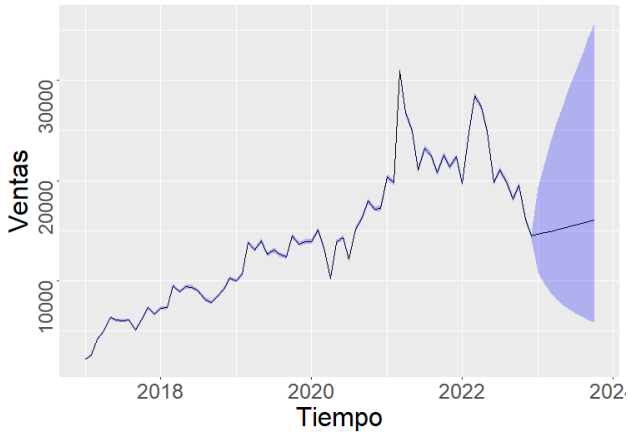


Figura 3-13: Ajuste y pronóstico de ventas mensuales de la compañía en la costa este de los Estados Unidos desde el 2017 hasta el 2022 con el modelo INLA.

de la creciente demanda de este producto textil en la costa este de Estados Unidos, pero ampliando la mirada un par de años más atrás hasta el inicio del 2018 es posible pensar que este comportamiento de aumento de la demanda en estos primeros meses es un efecto provocado por el fin de la pandemia del COVID-19 y cada vez está mermando sus picos de demanda, es por esta razón que el ajuste y pronósticos hechos bajo la metodología INLA también parece ajustar bien la realidad.

3.5. Conclusiones

Para predecir el comportamiento de las ventas de una compañía textil ubicada en Estados Unidos, dada la información de sus ventas hechas en los últimos siete años agrupadas mensualmente. Es posible concluir que la metodología que mejor ajusta y pronostica las ventas mensuales de una compañía en la costa este de Estados Unidos son los modelos clásicos de series de tiempo como el modelo autorregresivo de orden uno.

Dado el análisis hecho en la sección anterior de los resultados de los pronósticos observados en las Figuras **3-12** y **3-13**. Es importante para la compañía tener en cuenta todos los escenarios posibles a la hora de elegir una metodología para predecir el número de unidades que necesitará para cubrir la demanda del próximo año.

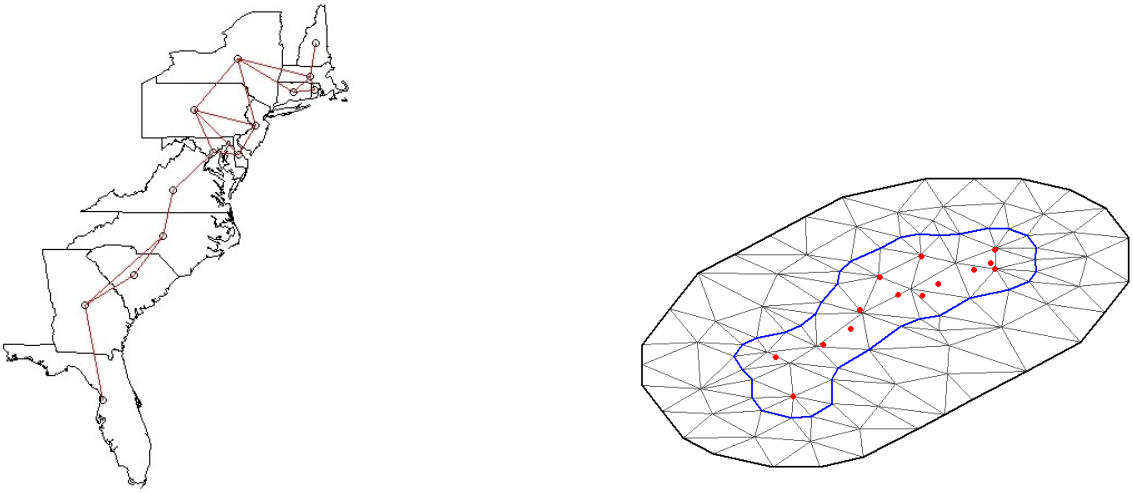
4 Comparación entre los modelos espaciales CAR y la aproximación INLA aplicados a datos de las ventas

En este capítulo se presenta una comparación entre el modelo Conditional Autoregressive (CAR) y un modelo espacial dado por la implementación del paquete INLA aplicados a los datos de las ventas con información espacial. El objetivo es determinar cuál de estos modelos presenta una mayor capacidad de ajuste y generar pronosticos sobre la variable de interés (ventas) en función de la información espacial disponible.

4.1. Descripción de los modelos

En muchos casos, es común suponer que las observaciones son independientes e idénticamente distribuidas, pero esto puede no ser el caso cuando se trabaja con datos espaciales ya que puede existir alguna correlación entre áreas vecinas. También puede ser difícil separar el impacto de la autocorrelación y las diferencias espaciales en la distribución de la observación. En esta parte del análisis se agruparon las observaciones de las ventas para cada estado de la costa este de Estados Unidos para el año 2022, en un centroide que corresponde a una ubicación espacial centrada dentro del polígono que hace referencia a cada estado, como se ve en la Figura **4-1**, esta agrupación es más conocida en la literatura como datos de área. También se agregó como covariable del modelo la población de cada uno de estos estados extraída de [Census, 2022] para el año 2022.

Se define como datos de área cuando $y(s)$ es un valor aleatorio definido sobre una unidad de área s con límites bien definidos en D, el cual es definido como una colección finita de unidades espaciales de dimensión d. Ejemplos de este tipo de datos son los límites administrativos de comunas, provincias, regiones, etc. [Blangiardo and Cameletti, 2015]. La Figura **4-1**a muestra las ubicaciones espaciales centradas dentro del polígono. La Figura **4-1**b muestra la triangulación para la costa este de los Estados Unidos con puntos rojos que indican las ubicaciones de los estados donde se encuentra disponible la información de ventas utilizada por INLA. La línea interna define el casco no convexo.



(a) Grafo de conectividad basado en centroides.

(b) Triangulación refinada de la ventana espacial.

Figura 4-1: Centroides asignados a los estados de la costa este de Estados Unidos y triangulación refinada restringida de la ventana de observación.

Siguiendo la definición dada por [Blangiardo and Cameletti, 2015]. Cuando se trabaja con este tipo de datos, la dependencia espacial es tomada en cuenta a través de una estructura de vecindades. Simplificando la notación introducida previamente así $(s_1, \ldots, s_n)$ pasa a ser $(1, \ldots, n)$, donde típicamente dada el área i sus vecinos $N(i)$ son definidos como áreas las cuales comparten sus bordes con el (vecinos de primer orden) o que comparten sus bordes con él y con sus vecinos de primer orden (vecinos de segundo orden) cómo se ve en la Figura **4-2**.

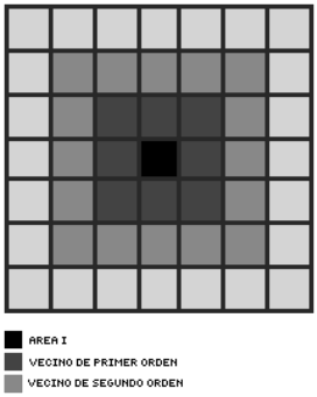


Figura 4-2: Estructura espacial de vecindades [Blangiardo and Cameletti, 2015].

La dependencia espacial tiende a reducir la cantidad de información contenida en las observaciones mientras más alejada se esta del centro como se ve en la Figura **4-2**, dado que las observaciones próximas pueden usarse en parte para predecirse entre sí. Un método comúnmente utilizado para probar si existe o no correlación espacial es aplicando la prueba de moran [Griffith, 1989] a los datos, como se ve a continuación. En términos simples, la prueba I de Moran es una forma de cuantificar qué tan cerca se agrupan los valores en un espacio. La fórmula para calcular el I Moran es,

$$I = \left(\frac{N}{S_0}\right) * \frac{\sum_{i=1}^{n}\sum_{j=1}^{n} W_{ij} Z_i Z_j}{\sum_{i=1}^{n} Z_i^2}, \tag{4-1}$$

donde $S_0 = \sum_{i=1}^{n}\sum_{j=1}^{n} W_{ij}$ o dicho de una manera más sencilla la suma de elementos de la matriz de pesos. Las observaciones Z son las desviaciones a la media $(x_i - \overline{x})$ ó $(x_j - \overline{x})$ donde x_i es el valor de la variable en una unidad espacial determinada y x_j es el valor de la variable en otra localización, normalmente las vecinas a x_i. Habitualmente, en la matriz, se le asigna el valor de 1 a los vecinos de cada unidad espacial y 0 al resto.

El valor de I Moran puede variar de -1 a 1 donde -1 indica que la variable de interés está perfectamente dispersa, 0 indica que la variable de interés se dispersa aleatoriamente y 1 indica que la variable de interés esté perfectamente agrupada. La línea azul ajustada añadida la Figura **4-3**, es el resultado de un modelo de regresión OLS y la pendiente de esta recta es el coeficiente I de Moran.

El coeficiente I de Moran es 0.086. La pendiente positiva (hacia arriba) sugiere que a medida que aumenta la cantidad de las ventas de dicho polígono, también lo hacen los de sus polígonos vecinos. Si la pendiente fuera negativa (es decir, con pendiente hacia abajo), esto sugeriría una relación negativa en la que los valores crecientes en dicho polígono estarían rodeados por polígonos con valores de ventas decrecientes.

Podemos concluir que en nuestros datos existe autocorrelación espacial o que nuestros datos no están aleatoriamente dispersos. Este resultado permite la aplicación y el correcto análisis del modelo CAR sobre nuestros datos. Por otro lado, la creación de ponderaciones espaciales entre estos centroides es un paso necesario en el uso de datos de área, tal vez solo para verificar que no queden patrones espaciales en los residuos. El primer paso es definir a qué relaciones entre las observaciones se les dará un peso distinto de cero, es decir, elegir el criterio vecino que se utilizará; el segundo es asignar pesos a los enlaces vecinos identificados. Para esta tarea se utilizó la librería *spdep* [Thisted, 1998] con la función *poly2nb* [Roger Bivand, 2022] que asigna pesos de cero a los centroides que no comparten frontera con otro.

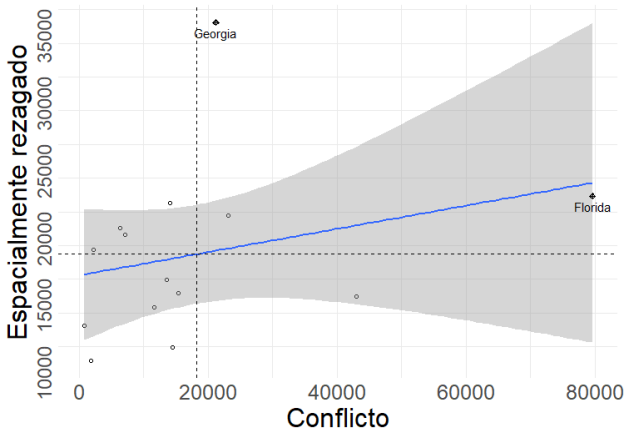


Figura 4-3: Gráfico de Moran I.

4.2. Modelo espacial autorregresivo condicional

Considere una región espacial que está dividida en n subregiones indexadas por enteros $1, 2, \ldots, n$. Suponga que esta colección de subregiones está denotada de un sistema de vecindad $V_i : i : 1, \ldots, n$, donde V_i denota la colección de subregiones que, en un sentido bien definido, son vecinos de la subregión i. En términos espaciales,

$$V_i = j : \text{subregiones } i \text{ y } j \text{ que comparten borde, para } i \in 1, 2, \ldots, n. \tag{4-2}$$

Los modelos autorregresivos condicionales (CAR), que son un tipo de modelo de campo aleatorio de Markov [Lee, 2011]. Estos modelos se especifican mediante un conjunto de n distribuciones condicionales completas para efectos aleatorios correspondientes a las variables de respuesta para cada $y_i, \ldots, y_n$ definidas como $\theta_1, \ldots, \theta_n$ y dadas por $\theta = \theta_i|\theta_{-i}, i = 1, \ldots, n$ [Cruz Reyes, 2020]. Suponiendo que la distribución condicional completa de θ_i dada θ_{-i} para todos $i = 1, \ldots, n$ es una distribución normal con media y precisión dadas respectivamente por

$$E(\theta_i|\theta_{-1}) = \mu_i - \sum_{j:j\sim i} \beta_{ij}(\theta_j - \mu_j), \tag{4-3}$$

y

$$\text{Prec}(\theta_i|\theta_{-j}) = k_i, \tag{4-4}$$

los componentes del vector θ son nodos espaciales. Supongamos que $\theta_1, \ldots, \theta_n$ sean las observaciones realizadas en las áreas $1, \ldots, n$, denotemos por $j \sim i$ que el nodo j es vecino del nodo i.

Suponiendo que $\beta_{ij} = \rho_\xi / d_i^\xi$, si $i \sim j$, y $\beta_{ij} = 0$ caso contrario en que $k_i = d_i^\xi / \Sigma_\xi^2$, esto es,

$$\theta_i|\theta_{-i} \sim N\left(\mu_i + \rho_\xi \bar{\theta}_i, \frac{\sum_\xi^2}{d_i^\xi}\right), \tag{4-5}$$

donde $\sum_{\xi}^{2}/d_i^{\xi}$ es la variable condicional de $\theta_i|\theta_{-i}$, ρ_{ξ} es una constante de proporcionalidad, d_i^{ξ} es el número de vecinos del nodo i en el grafo ξ, la media de los vecinos del nodo i es $\bar{\theta}_i = \sum_{\wp^{\xi}}(d_i^{\xi})^{-}1(\theta_j)$ y $\wp^{\xi} = [(i,j) \in E(\xi) : j \sim i]$ es el conjunto de aristas que pertenece al grafo ξ. Si dos áreas se definen como vecinas, sus efectos aleatorios están correlacionados, mientras que los efectos aleatorios en áreas no vecinas se modelan como condicionalmente independientes dados los elementos restantes de θ [Cruz Reyes, 2020]. El término condicional, en el modelo CAR se usa porque cada elemento del proceso aleatorio se especifica condicionalmente en los valores de los nodos vecinos [Cruz Reyes, 2020].

En resumen, el modelo CAR está definido por una estructura de correlación inducida por el grafo de la siguiente forma: la region de estudio V se divide en n unidades de área sobre el conjunto de regiones $V_i : i : 1, \ldots, n$ que están vinculados a un conjunto correspondiente de respuesta $y = (y_1, y_2, \ldots, y_n)$. Para la implementación del modelo CAR, se utilizó el paquete *spdep* [Roger Bivand, 2022] en R, que permite ajustar modelos espaciales para datos de áreas pequeñas. Se ajustó un modelo lineal generalizado (GLM) [Nelder and Wedderburn, 1972] con estructura CAR, asumiendo una distribución normal para los residuos y con una estructura de vecinos definida por la proximidad espacial de los puntos de venta.

4.3. Modelo espacial INLA

Para la implementación del modelo espacial bajo la metodología INLA, como define [Blangiardo and Cameletti, 2015], varias estructuras de área pueden ser especificadas para $u = u_1, \ldots, u_n$, nos centraremos en la condicional autorregresiva propuesta por [Besag and Kooperberg, 1995], dado que viene implementada en R-INLA.

Considerando n áreas, cada una caracterizada por un conjunto de vecinos $N(i)$ asumamos que u_i es la siguiente variable aleatoria

$$u_i|U_i \sim N\left(\mu_i + \sum_{j=1} r_{ij}(u_j - \mu_j), s_i^2\right), \tag{4-6}$$

donde μ_i es la media del área i y $s_i^2 = \sigma_u^2/N_i$ es la varianza para la misma área, la cual depende de su número de vecinos ($N_i = N(i)$), es decir, si el área tiene muchos vecinos entonces su varianza será menor. Esta estructura de la varianza reconoce el hecho de que existe una fuerte correlación espacial, mientras más vecinos tenga el área entonces más información hay en los datos sobre el valor del efecto aleatorio, mientras que el parámetro de la varianza σ_u^2 controla la cantidad de variación entre los efectos aleatorios estructurados espacialmente [Blangiardo and Cameletti, 2015].

La cantidad r_{ij} indica la proximidad espacial y puede ser calculada como $\rho\mathbf{W}_{ij}$, donde $\mathbf{W}_{ij} = a_{ij}/N_i$, a_{ij} es 1 si las áreas i y j son vecinos y 0 en otro caso. Considerando $\mathbf{W}$

como una matriz con elementos genéricos $\mathbf{W}_{ij}$ y $\mathbf{S} = \text{diag}(s_1, \ldots, s_n)$, para asegurar que la distribución para los efectos aleatorios espacialmente estructurados es adecuada, la matriz de covarianza $(\mathbf{I} - \rho\mathbf{W})^{-1}\mathbf{S}^2$ debe ser definida positiva; entonces los valores de ρ deben estar entre $1|\min_{i=1,\ldots,n} x_i$ y $1|\max_{i=1,\ldots,n} x_i$, con x_i siendo un valor propio genérico de $\mathbf{W}$. Entonces la especificación adecuada condicional autorregresiva (CAR) u es una variable aleatoria Normal multivariada,

$$U \sim MVN\left(\mu, (\mathbf{I} - \rho\mathbf{W})^{-1}\mathbf{S}^2\right), \tag{4-7}$$

donde $\mu = \mu_1, \ldots, \mu_n$ es el vector de medias, $\mathbf{I}$ es la matriz identidad y $\mathbf{S}^2$ fue definida anteriormente. Entonces la distribución condicional de $u_i|U_i$ es

$$u_i|U_i \sim N\left(\mu_i + \rho\frac{1}{N_i}\sum_{j=1} a_{ij}(u_j - \mu_j), s_i^2\right), \tag{4-8}$$

y la relación entre áreas i y j depende solo de ρ y $\mathbf{W}$ y es dada por

$$\frac{\sqrt{N_i}}{N_i} \times \frac{(I - \rho\mathbf{W})_{ij}^{-1}}{\sqrt{(I - \rho\mathbf{W})_{ii}^{-1}(I - \rho\mathbf{W})_{jj}^{-1}}}. \tag{4-9}$$

Una particularidad de INLA en la estimación de estructuras espaciales es que resulta eficiente para modelar estructuras ralas, es decir con gran presencia de valores ceros, en la inversa de la matriz de variancias y covariancias (matriz de precisión). La estructura rala de la matriz de precisión se debe a la no dependencia de las variables aleatorias en la distribución multivariada conjunta [Rue and Held, 2005].

En INLA particularmente se logran matrices de precisión ralas utilizando aproximaciones por ecuaciones diferenciales parciales estocásticas (SPDE) [Lindgren et al., 2011, Lindgren and Rue, 2015]. Bajo este enfoque la grilla de predicción se abarca a través de una malla construida a partir de triángulos que cubren el dominio entero, cada vértice de los triángulos representa un nodo sobre los que se predice por interpolación [Blangiardo and Cameletti, 2015]. Además de las ventajas computacionales que el algoritmo ofrece, permite trabajar con límites y bordes complejos [Bakka et al., 2018].

SPDE consiste en representar un proceso espacial continuo como un campo gaussiano (GF) utilizando un proceso aleatorio espacial discretamente indexado como un campo aleatorio gaussiano de Markov (GMRF). Los modelos autorregresivos condicionales (CAR) conducen a algunos resultados contrarios a la intuición o poco prácticos cuando se utilizan redes irregulares y/o las "celdas" tienen un área muy diferente [Wall, 2004]. Según [Bakka et al., 2018] cualquier parametrización del modelo CAR debe dar matrices de precisión definidas positivas. Además, establecer prioridades en los parámetros CAR necesita lidiar con los límites entre los modelos propios e intrínsecos [Bakka et al., 2018].

El enfoque SPDE, por otro lado, genera matrices de precisión con las buenas propiedades computacionales de los modelos CAR y es aplicable a cualquier conjunto de ubicaciones de observación. Por lo tanto, hemos utilizado la técnica SPDE que efectivamente permite que INLA calcule de manera eficiente la estructura de autocorrelación espacial del conjunto de datos en los vértices de la malla.

4.4. Evaluación de los modelos

Se agruparon las observaciones de las ventas para cada estado de la costa este de Estados unidos para el año 2022. Estos pronósticos se obtuvieron de dos maneras diferentes, primero para el modelo espacial CAR se utilizó la función *spautolm* del paquete de R "*spdep*" [Roger Bivand, 2022], el cual automáticamente ajusta el modelo que le indicamos y pronostica el número de periodos que nosotros indiquemos, para el modelo implementado bajo la metodología INLA, obtener el pronóstico consiste en agregar a los datos un conjunto de datos nulos igual a la cantidad de estados los cuales se desean predecir, es decir, si la base de datos cuenta con información de 10 estados para determinado año, entonces se deben agregar 10 espacios nulos que corresponden a los 10 estados que se quieren pronosticar y al ejecutar el algoritmo, este pronosticará automáticamente estos valores.

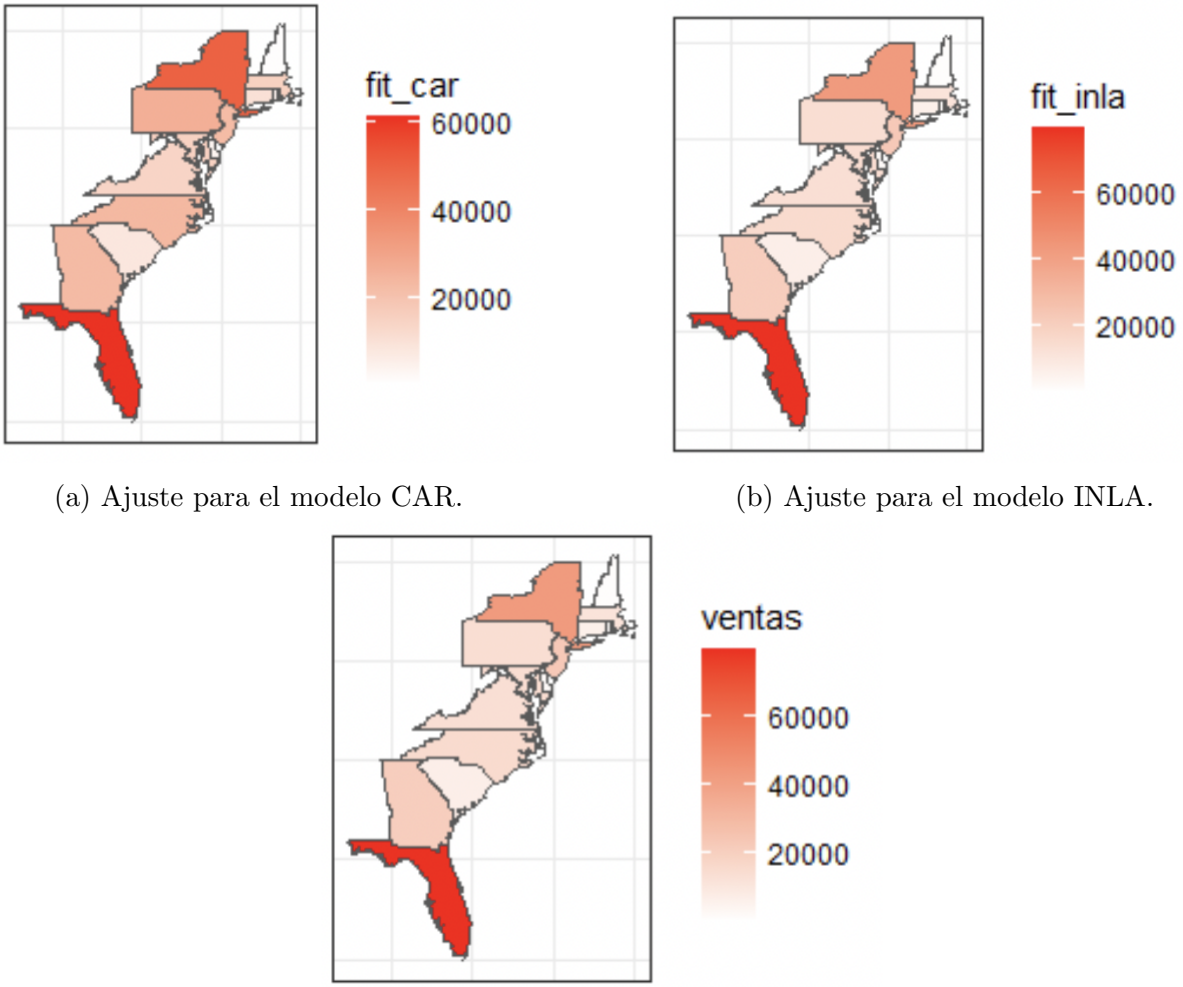
Se evaluó el rendimiento de ambos modelos utilizando medidas de ajuste tradicionales como el Error cuadrático medio (MSE) [Hodson et al., 2021] el cual mide el promedio de los errores al cuadrado, es decir, la diferencia entre el estimador y lo que se estima. Los resultados de las medidas de ajuste muestran que el modelo implementado con el paquete INLA presenta un mejor rendimiento en el ajuste y la predicción de las ventas en comparación con el modelo CAR como se detalla en la siguiente sección.

La mayor precisión en la estimación de los efectos espaciales que presenta el modelo implementado con el paquete INLA se debe en gran medida a su capacidad para modelar la estructura de dependencia espacial en los datos. Mientras que el modelo CAR asume una dependencia espacial homogénea entre todas las unidades espaciales. El modelo implementado con INLA permite modelar la dependencia espacial de manera más flexible, utilizando diferentes estructuras de dependencia para diferentes áreas o regiones [Blangiardo and Cameletti, 2015].

Es importante destacar que la implementación del paquete INLA puede requerir un mayor conocimiento técnico y de programación en R en comparación con el modelo CAR tradicional. Sin embargo, los resultados muestran que vale la pena la inversión de tiempo y recursos para obtener una mayor precisión en la modelación y predicción de los datos de ventas con información espacial.

4.5. Análisis de resultados

En la Tabla **4-1** se presentan los resultados de los dos modelos ajustados. Se puede observar que el modelo ajustado con el paquete INLA presenta un menor Error cuadrático medio (MSE) [Hodson et al., 2021] en el conjunto de prueba, lo que indica que este modelo tiene una mejor capacidad predictiva que el modelo CAR.



(a) Ajuste para el modelo CAR.

(b) Ajuste para el modelo INLA.

(c) Ventas por estado.

Figura 4-4: Ajuste espacial de las ventas textiles de la compañía en la costa este de los Estados Unidos desde el 2017 hasta el 2022.

Modelos	Error cuadrático medio (MSE)	Error absoluto medio (MAE)
CAR	8,34	2,37
INLA	5,82	1,98

Tabla 4-1: Resultados del ajuste de los modelos espaciales CAR e INLA.

Del mismo modo en la Figura **4-4** es fácil ver que el mejor ajuste lo tuvo el modelo espacial implementado con la metodología INLA en la Figura **4-4**b ya el ajuste espacial del modelo CAR en la Figura**4-4**a está subestimando gran parte de las ventas en la Figura **4-4**c en los estados de la costa este de los Estados Unidos.

Los resultados indican que la implementación del paquete INLA es una mejor opción que el modelo CAR para modelar y predecir las ventas en función de la información espacial disponible. Es posible que esto se deba a la capacidad del paquete INLA para modelar la incertidumbre y la complejidad de los datos, lo que permite una mejor capacidad predictiva.

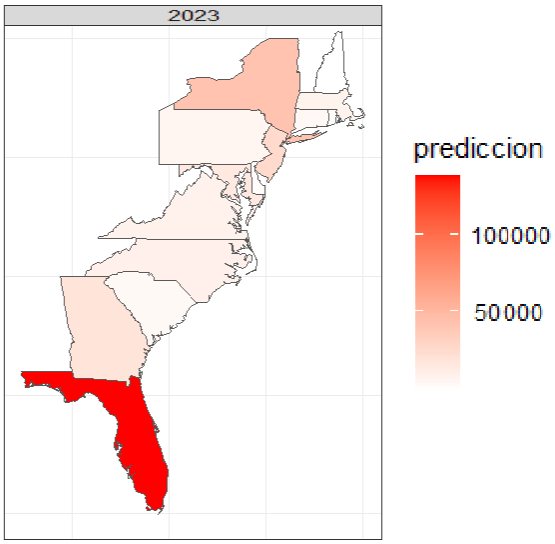


Figura 4-5: Predicción espacial de las ventas para el año 2023.

4.6. Conclusiones

En conclusión, la implementación del paquete INLA para el análisis de datos de ventas con información espacial presenta un mejor rendimiento en comparación con el modelo CAR tradicional. En ambos modelos se incluyó como covariable la población anual por estado. El modelo implementado con INLA permite modelar de manera más precisa la estructura de

dependencia espacial en los datos de ventas, lo que se traduce en una mejor capacidad de ajuste y, por lo tanto, puede ser una herramienta valiosa para la toma de decisiones empresariales basada en datos.

Cómo se puede ver en la Figura **4-5**, se espera que el estado de Florida el año 2023 venda más de 100.000 unidades. Se puede pensar que este crecimiento es debido al crecimiento poblacional de este estado en los últimos anos presuntamente debido al alto flujo de migrantes a sus principales ciudades como Miami. Del mismo modo otro hallazgo interesante es que el estado de Nueva York no presenta crecimiento en sus unidades para el año 2023 y esto se puede atribuir de nuevo a la población que se ha mantenido estable en los últimos años y a la disminución de las ventas totales presentadas el último año cómo es posible ver en la Figura **3-2**. Ambos estados han presentado un constante crecimiento en el tiempo como se ve en la Figura **2-8**.

5 Comparación entre los modelos espacio-temporales INLA y un enfoque de aprendizaje automático Bayesiano

En este capítulo se presenta la comparación de dos modelos espacio-temporales dados por la implementación del paquete INLA aplicados a datos de las ventas con información espacial, siendo uno de estos un modelo de aprendizaje automático. El objetivo es determinar cuál de estos modelos presenta una mayor capacidad de ajuste y generar pronósticos sobre la variable de interés (ventas) en función de la información espacial disponible.

5.1. Descripción de los modelos

Investigar únicamente el patrón espacial como se presentó en el capítulo anterior no nos permite decir nada acerca de su variación temporal que podría ser igualmente importante e interesante.

Cuando se estudia la tendencia temporal, por lo general, pueden ocurrir dos situaciones, como se presenta en [Abellan et al., 2008], la tasa es constante en el tiempo y solo varía espacialmente o cambia con el tiempo en todas o algunas de las áreas. Las mismas situaciones pueden surgir cuando estimamos un fenómeno que varía espacialmente, por ejemplo, la cantidad de lluvia en un período particular, por ejemplo, un año; si consideramos los datos agregados a lo largo del tiempo, solo podemos modelar el patrón espacial, pero si desagregamos los datos por día, semana o mes, podemos investigar una tendencia temporal y permitir que sea igual en todas las áreas o diferente para algunas de ellas.

5.2. Modelo espacio-temporal INLA

El proceso espacial presentado en el capítulo anterior puede extenderse fácilmente al caso espacio-temporal incluyendo una dimensión temporal; los datos ahora están definidos por un proceso indexado por espacio y tiempo. De acuerdo a [Blangiardo and Cameletti, 2015], un

proceso espacio-temporal indexado por espacio (s) y tiempo (t) se define como

$$Y(s,t) \equiv y(s,t), (s,t) \in D \subset \mathbb{R}^2 \times \mathbb{R}, \tag{5-1}$$

donde Y es observado en n localidades espaciales o áreas y en T tiempos. Los modelos espacio-temporales como menciona [Blangiardo and Cameletti, 2015] son ampliamente utilizados en estudios de vigilancia de enfermedades. Cuando es de interés identificar el patrón espacio-temporal de una enfermedad.

Dado que en nuestro caso al ser similar a los problemas espacio-temporales de mapeo o vigilancia de enfermedades se asumió la misma distribución apriori $Poisson$, ya que el pronóstico de ventas en los estados de la costa este de Estados Unidos puede ser abordado como un problema de conteo. En la práctica, es la extensión del modelo de ventas espacial definido previamente agregándole una componente temporal

- $y_{it} \sim Poisson(\lambda_{it})$,
- $\lambda_{it} = E_{it}\rho_{it}$,
- $log(\rho_{it}) = \eta_{it}$,
- $\eta_{it} = b_0 + u_i + v_i + \text{Temporal}_t$,

con $t = 1, \ldots, T$; y Temporal_t con una estructura paramétrica o no paramétrica especificada.

5.2.1. Tendencia paramétrica

Como menciona [Blangiardo and Cameletti, 2015] se puede asumir que la tendencia de la componente temporal es paramétrica y se asume que el predictor lineal puede ser escrito como

$$\eta_{it} = b_0 + u_i + v_i + (\beta + \delta_i) \cdot t, \tag{5-2}$$

esta formulación incluye el efecto espacial principal del caso espacial u_i, la tendencia lineal principal v_i, β que representa el efecto global temporal, y una tendencia diferencial δ_i, que identifica la interacción entre espacio y tiempo.

Ya que para la identificabilidad se propone una restricción suma igual a cero para $\delta = \delta_1, ..., \delta_n$, los términos δ_i representa la diferencia entre la tendencia global β y la tendencia de un área en específico. Si $\delta_i < 0$ entonces la tendencia de una área en específico es menos empinada que la tendencia media, mientras que $\delta_i > 0$ implica que la tendencia de un área en específico es más empinada que la tendencia promedio.

5.2.2. Tendencia Dinámica no paramétrica

En el modelo especificado anteriormente, se impone una restricción lineal en la tendencia diferencial temporal δ_i; sin embargo, como sugiere [Blangiardo and Cameletti, 2015] es posible eliminarla utilizando una formulación no paramétrica dinámica para el predictor lineal como

$$\eta_{it} = b_0 + u_i + v_i + \gamma_t + \phi_t. \tag{5-3}$$

Aquí b_0, u_i y v_i tienen la misma parametrización que en el caso anterior, sin embargo el termino γ_t representa la estructura del efecto temporal, modelado dinámicamente (por ejemplo, como un camino aleatorio de orden 1 o 2).

5.2.3. Interacciones espacio-tiempo

Como se menciona en [Blangiardo and Cameletti, 2015], se puede seguir expandiendo el modelo anterior para agregar interacciones entre espacio y tiempo utilizando la siguiente especificación.

$$\eta_{it} = b_0 + u_i + v_i + \gamma_t + \phi_t + \delta_{it}, \tag{5-4}$$

el vector de parámetros δ sigue una distribución normal con matriz de precisión dada por $\tau_\delta \mathbf{R}_\delta$, donde τ_δ es un escalar desconocido, mientras que $\mathbf{R}_\delta$ es una matriz estructural identificando la dependencia si es de tipo espacial y/o temporal entre los elementos de δ.

$\mathbf{R}_\delta$ puede ser factorizada como el producto de Kronecker de la matriz de estructura de los efectos principales correspondientes que interactúan. Hay cuatro formas de definir la matriz de estructura.

- **Interacciones de tipo I:** Se asume que los dos efectos sin estructura v_i y ϕ_t interactúan entre sí. Podemos escribir la matriz de estructura como

  $$R_\delta = R_u \otimes R_\phi = I \otimes I = I, \tag{5-5}$$

  la última igualdad se debe a que v y ϕ no tienen una estructura espacial o temporal. Por lo tanto, asumimos que no hay una estructura temporal y/o espacial en la interacción entre ambas.

- **Interacciones de tipo II:** En este caso se combina la estructura temporal del efecto principal γ_i y el efecto espacial sin estructura v_i. Podemos escribir la matriz de estructura como

  $$R_\delta = R_u \otimes R_\gamma, \tag{5-6}$$

  donde $R_v = I$ y R_γ es la estructura de vecindad especificada por ejemplo como un camino aleatorio de primer o segundo orden. Esto conlleva a asumir que para la i-ésima área el vector de parámetros $\delta_{i1}, ..., \delta_{iT}$ tiene una estructura autorregresiva en la componente temporal, que es independiente de las demás áreas.

- **Interacciones de tipo III:** Este caso combina el efecto temporal sin estructura ϕ_t y la estructura del efecto principal u_i. Escribimos la matriz de estructura como

$$R_\delta = R_\phi \otimes R_u, \tag{5-7}$$

  donde $R_\phi = I$ y R_u es una estructura de vecindad definida a través de la especificación CAR. Esto conlleva a asumir que los parámetros en el t-ésimo tiempo $\delta_1, ..., \delta_n$ tiene una estructura espacial independiente de los demás tiempos.

- **Interacciones de tipo IV:** Este es el caso más complejo de interacción, asumiendo que los efectos estructurados espaciales y temporales u_i y γ_t interactúan entre sí. La matriz de estructura puede ser escrita como

$$R_\delta = R_u \otimes R_\gamma, \tag{5-8}$$

  la cual básicamente asume que la estructura de dependencia temporal para cada área no es independiente de las demás áreas, pero depende del patrón temporal así como de las áreas vecinas.

5.3. Modelo de corrección media Bayesiana variacional de rango bajo (VBC)

De acuerdo con lo señalado por [van Niekerk and Rue, 2021], los métodos bayesianos se fundamentan en la existencia de una creencia previa acerca de un modelo particular y la capacidad de aprender de los datos disponibles para llegar a una nueva creencia, conocida como creencia posterior. Matemáticamente, la creencia posterior se deriva a partir de la creencia previa y la evidencia empírica proporcionada por los datos, utilizando la regla de Bayes. En este sentido, el análisis bayesiano emerge como un método de aprendizaje automático estadístico natural y particularmente poderoso en situaciones donde los conjuntos de datos son pequeños, existen datos faltantes o se emplean modelos altamente complejos.

Aprendizaje automático

El aprendizaje automático, también conocido como machine learning en inglés, es un campo de la inteligencia artificial que se enfoca en desarrollar algoritmos y modelos computacionales que permiten a las máquinas aprender y mejorar su desempeño en tareas específicas a través de la experiencia y la interacción con datos [James et al., 2013].

En lugar de utilizar reglas de programación tradicionales, los algoritmos de aprendizaje automático analizan datos y reconocen patrones que les permiten realizar predicciones o tomar decisiones. Estos patrones se extraen de un conjunto de datos de entrenamiento, que

consiste en ejemplos previos de la tarea que se desea aprender [James et al., 2013]. El proceso de aprendizaje automático implica

- Recopilación de Datos: Se reúnen datos relevantes para la tarea, que pueden incluir imágenes, texto, números u otros tipos de información.
- Preprocesamiento de Datos: Los datos se limpian y transforman para eliminar ruido, outliers y asegurar que sean adecuados para el modelo.
- Selección del Modelo: Se elige un algoritmo de aprendizaje automático adecuado para la tarea, como regresión, clasificación, clustering, redes neuronales, entre otros.
- Entrenamiento del Modelo: Se alimenta al modelo con el conjunto de datos de entrenamiento para que aprenda patrones y relaciones entre los datos.
- Validación y Evaluación: Se evalúa el rendimiento del modelo utilizando un conjunto de datos de prueba para determinar su precisión y generalización.
- Ajuste y Optimización: Si es necesario, se ajustan los hiperparámetros del modelo para mejorar su rendimiento.
- Despliegue y Uso: Una vez que se ha entrenado y validado el modelo, se puede utilizar para realizar predicciones o automatizar decisiones en situaciones del mundo real.
- Mejora Continua: El modelo se puede mejorar continuamente a medida que se recopilan más datos y se obtienen nuevos conocimientos.

El aprendizaje automático tiene aplicaciones en una amplia variedad de campos, desde reconocimiento de voz y visión por computadora hasta recomendaciones personalizadas en línea y medicina. Es un área en constante evolución que impulsa avances significativos en la automatización y la toma de decisiones basadas en datos [James et al., 2013].

Cuando la inferencia exacta se vuelve impracticable debido a la complejidad del modelo o a la gran cantidad de datos involucrados, los métodos de inferencia aproximada, como el método de Laplace, las aproximaciones de Laplace y los métodos variacionales, se convierten en alternativas populares. La Corrección Variacional Bayesiana de Bajo Rango (VBC), en particular, utiliza inicialmente el método de Laplace y luego aplica una corrección variacional de Bayes a la media posterior.

Esto se lleva a cabo con un costo computacional esencialmente equivalente al del método de Laplace, lo que garantiza la escalabilidad de este enfoque. Supongamos que tenemos datos y, de tamaño n con covariables X y efectos aleatorios u, entonces

$$Y_i|\beta, f \sim Poisson(\exp(\eta_i)), \tag{5-9}$$

con

$$\eta_i = \beta_0 + \beta X_i + \sum k = 1 f^k(u_k). \tag{5-10}$$

Asumimos una distribución a priori normal centrado para el campo latente (aumentado) con matriz de precisión Q_π. Podemos corregir la parte lineal de la expansión de la serie de Taylor como

$$\mu^* = \mu + \delta^*, \tag{5-11}$$

donde $\delta^* = Q_j^{-1}\lambda^*$ para corregir principalmente las medias posteriores de β_0 y β y luego propagar el efecto de estas correcciones a la media posterior del predictor lineal [van Niekerk and Rue, 2021].

Con el método variacional de Bayes, una de las principales desventajas es la suposición de una factorización simple de la articulación posterior, generalmente en posteriores marginales independientes, ignorando así la dependencia posterior del espacio de parámetros, o la necesidad de una descomposición de bajo rango de la matriz de covarianza si la dependencia en el espacio de parámetros está permitido. Por otro lado, se sabe que la propagación de expectativas (EP) sobreestima las varianzas posteriores, mientras que (VB) subestima estas varianzas.

5.4. Evaluación de los modelos

Al conjunto de datos de ventas que fue utilizado en la sección anterior, que contiene información sobre la ubicación geográfica de cada punto de venta, se le agrego la parte temporal de los datos con el fin de analizar el comportamiento de las ventas en el tiempo a lo largo del territorio de la costa este de los Estados Unidos. La definición del mejor modelo se basó en un análisis de coeficientes de la estimación como el DIC y el WAIC.

Cuando se dispone de información temporal, es posible construir modelos espaciotemporales que incluyen efectos aleatorios espaciales y temporales, así como efectos de interacción entre el espacio y el tiempo. Un modelo espacio-temporal resulta separable cuando la estructura de covarianza espacio-temporal puede descomponerse como un término espacial y temporal. Para efectos de este estudio, se procede a adaptar parte de la metodología de análisis seguida por [Rodrigues, 2017] y [Briz-Redón, 2021], que son estudios que a su vez adoptan la metodología de estudios previos como [Knorr-Held, 2000] y [Blangiardo et al., 2013].

La inclusión de efectos espacio-temporales ayuda a comprender cómo las ventas se ha extendido por el territorio en estudio. Específicamente, las estimaciones de efectos espaciales y temporales aleatorios y su interacción permiten asignar un riesgo relativo a cada unidad espacial, temporal o espacio-temporal bajo análisis. Estos riesgos relativos se obtienen exponenciando los parámetros espacio-temporales del modelo y captura la evolución de las ventas

en la costa Este de Estados Unidos.

Inicialmente se consideró la distribución Poisson como la más adecuada basado en los datos de bajos conteos como lo más adecuada. Investigaciones como [Gayawan et al., 2020]; [Jalilian and Mateu, 2021]; [Jalilian and Mateu, 2021]; [Sartorius et al., 2021], han aplicado modelos espaciotemporales con distribución Poisson en sus estudios. El costo de ambos modelos es esencialmente el del método de Laplace que asegura la escalabilidad de los métodos.

Para evaluar los modelos, se utilizaron los criterios de información DIC y $WAIC$ dado que si se consideran dos modelos candidatos $M1$ y $M2$ para describir los datos dados y_n. Usando las probabilidades del modelo posterior $\pi(M_k|y_n)$ para $k = 1, 2$, podemos escribir las probabilidades posteriores como,

$$\frac{\pi(M_1|Y^n)}{\pi(M_2|Y^n)} = \frac{m(Y^n; M_1)}{m(Y^n; M_2)} \times \frac{\pi(M_1)}{\pi(M_2)}, \tag{5-12}$$

donde $m(y_n; M_k)$ y $\pi_(M_k)$ denota la probabilidad marginal y la probabilidad previa para el modelo M_k, respectivamente. El primer término del lado derecho se conoce como el factor de Bayes (BF) a favor del modelo $M1$ cuando se compara con $M2$ [Kass and Raftery, 1995], es decir, el BF puede interpretarse como el resumen del apoyo proporcionado por los datos a un modelo en contraposición a uno alternativo.

$$BF_{12} = \frac{m(Y^n; M_1)}{m(Y^n; M_2)}. \tag{5-13}$$

Como señalaron [Kass and Raftery, 1995], el Bayes Factor (BF) se puede aproximar utilizando el criterio de Schwarz, que también se conoce como el criterio de información Bayesiano (BIC). Especialmente para los modelos Bayesianos jerárquicos, no es fácil evaluar BIC ya que no se conocerá el número "efectivo" de parámetros en el modelo. Derivado del BIC, el criterio de información de desviación (DIC) fue propuesto por [Spiegelhalter et al., 2002]. Para un vector de parámetros genérico θ, el DIC se define como,

$$DIC = \overline{D} + r_D, \tag{5-14}$$

donde

$$D = -2 \log L(\theta; Y^n),\ D = E_{\theta|Y^n(D)},\ \text{ y }\ r_D = \overline{D} - D(\hat{\theta}), \tag{5-15}$$

donde θ estimado es la media posterior, $\overline{D}$ representa la bondad del ajuste del modelo, mientras que el término r_D representa una penalización por complejidad reflejada por el número efectivo de parámetros del modelo. El número efectivo desconocido de parámetros del modelo se estima como la diferencia entre la media posterior de la desviación y la desviación evaluada en la media posterior de θ.

Una alternativa al DIC es el criterio de información de Watanabe Akaike ($WAIC$), que se define como en [Watanabe and Opper, 2010] y [Gelman et al., 2014],

$$WAIC = lppd + p_{WAIC}, \tag{5-16}$$

donde $lppd$ es el logaritmo de la densidad posterior puntual y p_{WAIC} es un término de corrección para el número efectivo de parámetros para ajustar por sobreajuste. Los resultados de la comparación entre los coeficientes de ajuste no muestran diferencia entre el modelo de aprendizaje automático implementado con el paquete INLA en comparación con el modelo clásico de INLA como se detalla en la siguiente sección.

5.5. Análisis de resultados

En la Tabla **5-1** se presentan los resultados de los dos modelos ajustados. Se puede observar que ambos modelos presentan un ajuste idéntico debido a que la metodología que utilizan ambos son muy similares, lo que indica que en nuestro caso de estudio ambos modelos tienen la misma capacidad predictiva.

Modelos	ML INLA	INLA
WAIC	32038,68	32038,68
DIC	33180,08	33180,08

Tabla 5-1: Coeficientes de ajuste de los modelos ML-INLA e INLA.

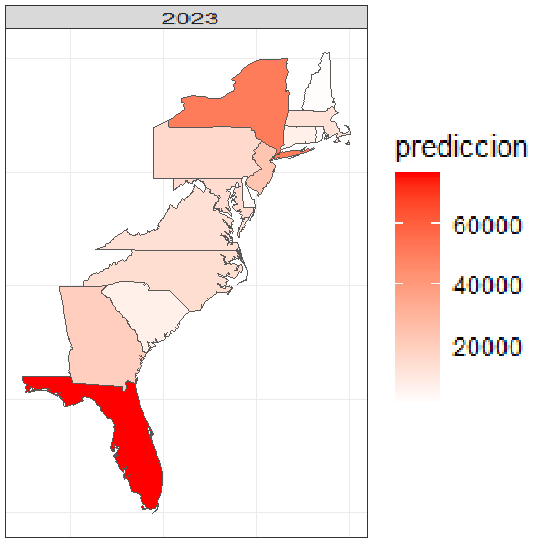


Figura 5-1: Predicción espacio-temporal de las ventas.

5.6. Conclusiones

En conclusión, la implementación del paquete INLA para el análisis de datos de ventas con información espacial y temporal presentan igual rendimiento en el ajuste de los datos de ventas como se ve en la tabla **5-1**, entre el modelo de aprendizaje automático (ML INLA) y el modelo clásico de INLA. Esto puede atribuirse principalmente por la cantidad de datos espacio-temporales utilizados. Estos modelos implementados con INLA permiten modelar de manera más precisa la estructura de dependencia espacial y temporal en los datos de ventas, lo que se traduce en una mejor capacidad predictiva y, por lo tanto, puede ser una herramienta valiosa para la toma de decisiones empresariales basada en datos.

Cómo se puede ver en la Figura **5-1**, el comportamiento de las ventas para el año 2023 mantiene un comportamiento similar al comportamiento que han tenido las ventas en los años inmediatamente anteriores cómo es posible ver en la Figura **2-8**. Lastimosamente en este trabajo no se graficaron los ajustes mensuales por estado debido a la alta complejidad a la hora de especificar estos modelos.

6 Conclusiones y recomendaciones

6.1. Conclusiones

En esta última sección se presentan las conclusiones del trabajo. Se incluyen las reflexiones finales, se discuten brevemente los resultados y se sugieren lineamientos para futuras investigaciones del tema.

En este trabajo de grado, se llevó a cabo una comparación exhaustiva entre métodos de estimación temporal, espacial y espacio-temporal clásicos y sus homólogos basados en la metodología R-INLA. El objetivo principal fue estimar las ventas de un producto y determinar cuál de las metodologías proporciona resultados más precisos.

En la introducción, se realizó una revisión bibliográfica de los modelos temporales, espaciales y espacio-temporales clásicos, así como de la metodología INLA. Se identificaron las principales ventajas y desventajas de cada método y se estableció la intención de este trabajo, que es compararlos en términos de precisión en la estimación de ventas y generar un pronóstico.

En el capítulo 3, se comparó un modelo AR1 de series de tiempo con un modelo temporal basado en la metodología INLA. Se encontró que el modelo temporal INLA y el modelo AR1 son significativamente similares en términos de precisión y capacidad de capturar la variabilidad temporal de las ventas a pesar de la flexibilidad de INLA para incorporar efectos aleatorios y autocorrelación temporal. Este resultado puede estar sujeto a la poca cantidad de datos utilizados en este análisis.

En el capítulo 4, se comparó un modelo espacial CAR con un modelo espacial INLA. Se encontró que el modelo espacial INLA ofrece una mejor precisión en la estimación de ventas en comparación con el modelo espacial CAR. La capacidad de INLA para incorporar efectos aleatorios espaciales y modelar la autocorrelación espacial resultó ser determinante en esta comparación.

Finalmente, en el capítulo 5, se implementaron dos modelos espacio-temporales en INLA que combina las ventajas de los modelos temporales y espaciales INLA. Los resultados demostraron que no había diferencias en el método de aprendizaje automático y el modelo clásico de INLA. Ambos modelos proporcionan igual precisión en la estimación de ventas.

Este resultado puede estar sujeto a la poca cantidad de datos utilizados en este análisis.

En general, las comparaciones realizadas en este trabajo favorecieron la metodología INLA en términos de precisión y capacidad para capturar la variabilidad tanto temporal como espacial en las ventas. INLA demostró ser una herramienta poderosa y eficiente para el análisis de datos espaciales y temporales complejos, permitiendo incorporar efectos aleatorios y autocorrelación de manera sencilla y eficaz.

Un problema importantes que se presentan en el ajuste de nuestros modelos es la dificultad en la especificación de los modelos bajo la metodología INLA, dado que es necesario un alto nivel técnico para el despliegue de los modelos temporales, espaciales y espacio-temporales.

En esta investigación, no se emplearon, modelos iniciales para determinar la mejor covariable principal. Para explicar las variaciones espaciales y temporales de las ventas en los estados de la costa Este de Estados Unidos, en el futuro, se sugiere contemplar la posibilidad considerar otras covariables regionales, ambientales y socioeconómicas.

Este trabajo adaptó modelos de estudios que emplearon INLA, como el de [Briz-Redón, 2021]. [Rodrigues, 2017] también emplea INLA, así como el uso de covariables para ajustar el modelo. Más que predecir, en la mayoría de estudios espacio-temporales se ha procurado ajustar los modelos, a fin de estimar riesgos relativos en el período de análisis.

La literatura existente es muy amplia sobre aplicaciones espacio-temporales a problemas epidemiológicos más no se ha evidenciado hasta la fecha de entrega de este trabajo antecedentes de aplicaciones similares a la presentada en este estudio. Esta metodología ofrece diversas vías para investigar las procesos bayesianos espacio-temporales aplicados a las ventas.

En estudios futuros, sería recomendable analizar las diferencias entre las cadenas de Márkov y el INLA en modelos espacio-temporales. Un ejemplo relevante es el caso de [Knorr-Held, 2000], que propone varias formas de distribución previa en los modelos espacio-temporales.

En conclusión, este trabajo confirma que la metodología INLA es una opción altamente recomendada para la estimación de ventas y, en general, para el análisis de datos espaciales y temporales. Su flexibilidad, precisión y capacidad para modelar efectos complejos la convierten en una herramienta valiosa para la investigación y el análisis de datos en diversas disciplinas. Se sugiere que futuras investigaciones exploren aún más las capacidades de INLA en diferentes contextos y aplicaciones para obtener una visión más completa de su potencial en el análisis de datos espaciales y temporales.

Bibliografía

[Abellan et al., 2008] Abellan, J. J., Richardson, S., and Best, N. (2008). Use of space–time models to investigate the stability of patterns of disease. *Environmental health perspectives*, 116(8):1111–1119.

[Adhikari et al., 2019] Adhikari, B., Xu, X., Ramakrishnan, N., and Prakash, B. A. (2019). Epideep: Exploiting embeddings for epidemic forecasting. In *Proceedings of the 25th ACM SIGKDD international conference on knowledge discovery & data mining*, pages 577–586.

[Amazon, 2023] Amazon (2023). Amazon web services. `https://aws.amazon.com/es/what-is-aws/`.

[Assimakopoulos and Nikolopoulos, 2000] Assimakopoulos, V. and Nikolopoulos, K. (2000). The theta model: a decomposition approach to forecasting. *International journal of forecasting*, 16(4):521–530.

[Bakka et al., 2018] 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. *Wiley Interdisciplinary Reviews: Computational Statistics*, 10(6):e1443.

[Besag and Kooperberg, 1995] Besag, J. and Kooperberg, C. (1995). On conditional and intrinsic autoregressions. *Biometrika*, 82(4):733–746.

[Bivand et al., 2008] Bivand, R. S., Pebesma, E. J., Gómez-Rubio, V., and Pebesma, E. J. (2008). *Applied spatial data analysis with R*. Springer, New York.

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

[Blangiardo et al., 2013] Blangiardo, M., Cameletti, M., Baio, G., and Rue, H. (2013). Spatial and spatio-temporal models with r-inla. *Spatial and spatio-temporal epidemiology*, 7:39–55.

[Briz-Redón, 2021] Briz-Redón, Á. (2021). The impact of modelling choices on modelling outcomes: a spatio-temporal study of the association between covid-19 spread and environmental conditions in catalonia (spain). *Stochastic Environmental Research and Risk Assessment*, 35(8):1701–1713.

[Brockwell and Davis, 2002] Brockwell, P. J. and Davis, R. A. (2002). *Introduction to time series and forecasting*. Springer.

[Brzezinski, 2012] Brzezinski, M. (2012). The chen–shapiro test for normality. *The Stata Journal*, 12(3):368–374.

[Catal et al., 2019] Catal, C., Kaan, E., Arslan, B., and Akbulut, A. (2019). Benchmarking of regression algorithms and time series analysis techniques for sales forecasting. *Balkan Journal of Electrical and Computer Engineering*, 7(1):20–26.

[Census, 2022] Census, U. S. (2022). Total population. `https://www.census.gov/data/tables/time-series/demo/popest/2020s-state-total.html`.

[Chatfield, 2000] Chatfield, C. (2000). *Time-series forecasting*. CRC press, Boca Raton.

[Chegoonian et al., 2017] Chegoonian, A., Mokhtarzade, M., and Valadan Zoej, M. (2017). A comprehensive evaluation of classification algorithms for coral reef habitat mapping: challenges related to quantity, quality, and impurity of training samples. *International Journal of Remote Sensing*, 38(14):4224–4243.

[Chiapella, 2020] Chiapella, L. (2020). Impacto de estrategias para el tratamiento de información faltante sobre la estimación de modelos de regresión de cox.

[Cruz Reyes, 2020] Cruz Reyes, D. L. (2020). Modelos gráficos probabilísticos aplicados al análisis espacial en r: Hurtos de celulares en bogotá. *Tecciencia*, 15(29):9–22.

[Deng et al., 2017] Deng, M., Yang, W., and Liu, Q. (2017). Geographically weighted extreme learning machine: A method for space–time prediction. *Geographical Analysis*, 49(4):433–450.

[Deng et al., 2018] Deng, M., Yang, W., Liu, Q., Jin, R., Xu, F., and Zhang, Y. (2018). Heterogeneous space–time artificial neural networks for space–time series prediction. *Transactions in GIS*, 22(1):183–201.

[Dutta et al., 2022] Dutta, C., Ravishanker, N., and Basu, S. (2022). Modeling multivariate positive-valued time series using r-inla. *arXiv preprint arXiv:2206.05374*.

[Effati et al., 2015] Effati, M., Thill, J.-C., and Shabani, S. (2015). Geospatial and machine learning techniques for wicked social science problems: analysis of crash severity on a regional highway corridor. *Journal of Geographical Systems*, 17:107–135.

[Fotheringham et al., 2003] Fotheringham, A. S., Brunsdon, C., and Charlton, M. (2003). *Geographically weighted regression: the analysis of spatially varying relationships*. John Wiley & Sons, Chichester.

[Gayawan et al., 2020] Gayawan, E., Awe, O. O., Oseni, B. M., Uzochukwu, I. C., Adekunle, A., Samuel, G., Eisen, D. P., and Adegboye, O. A. (2020). The spatio-temporal epidemic dynamics of covid-19 outbreak in africa. *Epidemiology & Infection*, 148:e212.

[Gelfand et al., 2010] Gelfand, A. E., Diggle, P., Guttorp, P., and Fuentes, M. (2010). *Handbook of spatial statistics*. CRC press, Boca Raton.

[Gelman et al., 2014] Gelman, A., Hwang, J., and Vehtari, A. (2014). Understanding predictive information criteria for bayesian models. *Statistics and computing*, 24(6):997–1016.

[Griffith, 1989] Griffith, D. A. (1989). Spatial econometrics: Methods and models. *Journal of the Royal Statistical Society Series A: Statistics in Society*, 135(5):370–384.

[Guan et al., 2005] Guan, Q., Wang, L., and Clarke, K. C. (2005). An artificial-neural-network-based, constrained ca model for simulating urban growth. *Cartography and Geographic Information Science*, 32(4):369–380.

[Hagenauer et al., 2019] Hagenauer, J., Omrani, H., and Helbich, M. (2019). Assessing the performance of 38 machine learning models: the case of land consumption rates in bavaria, germany. *International Journal of Geographical Information Science*, 33(7):1399–1419.

[Hester and Wickham, 2023] Hester, J. and Wickham, H. (2023). *odbc: Connect to ODBC Compatible Databases (using the DBI Interface)*. R package version 1.3.4.

[Hodson et al., 2021] Hodson, T. O., Over, T. M., and Foks, S. S. (2021). Mean squared error, deconstructed. *Journal of Advances in Modeling Earth Systems*, 13(12):e2021MS002681.

[Hyndman and Khandakar, 2008] Hyndman, R. J. and Khandakar, Y. (2008). Automatic time series forecasting: the forecast package for r. *Journal of statistical software*, 27:1–22.

[Hyndman et al., 2005] Hyndman, R. J., King, M. L., Pitrun, I., and Billah, B. (2005). Local linear forecasts using cubic smoothing splines. *Australian & New Zealand Journal of Statistics*, 47(1):87–99.

[Hyndman et al., 2002] Hyndman, R. J., Koehler, A. B., Snyder, R. D., and Grose, S. (2002). A state space framework for automatic forecasting using exponential smoothing methods. *International Journal of forecasting*, 18(3):439–454.

[Jalilian and Mateu, 2021] Jalilian, A. and Mateu, J. (2021). A hierarchical spatio-temporal model to analyze relative risk variations of covid-19: a focus on spain, italy and germany. *Stochastic Environmental Research and Risk Assessment*, 35:797–812.

[James et al., 2013] James, G., Witten, D., Hastie, T., Tibshirani, R., et al. (2013). *An introduction to statistical learning*, volume 112. Springer.

[Jank and Kannan, 2005] Jank, W. and Kannan, P. (2005). Understanding geographical markets of online firms using spatial models of customer choice. *Marketing Science*, 24(4):623–634.

[Kass and Raftery, 1995] Kass, R. E. and Raftery, A. E. (1995). Bayes factors. *Journal of the american statistical association*, 90(430):773–795.

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

[Krithikadatta, 2014] Krithikadatta, J. (2014). Normal distribution. *Journal of conservative dentistry: JCD*, 17(1):96.

[Law et al., 2020] Law, S., Seresinhe, C. I., Shen, Y., and Gutierrez-Roig, M. (2020). Street-frontage-net: urban image classification using deep convolutional neural networks. *International Journal of Geographical Information Science*, 34(4):681–707.

[Lawson, 2021] Lawson, A. B. (2021). *Using R for Bayesian spatial and spatio-temporal health modeling*. CRC Press, Boca Raton.

[Lee, 2011] Lee, D. (2011). A comparison of conditional autoregressive models used in bayesian disease mapping. *Spatial and spatio-temporal epidemiology*, 2(2):79–89.

[Li et al., 2014] Li, G., Haining, R., Richardson, S., and Best, N. (2014). Space–time variability in burglary risk: a bayesian spatio-temporal modelling approach. *Spatial Statistics*, 9:180–191.

[Lindgren and Rue, 2015] Lindgren, F. and Rue, H. (2015). Bayesian spatial modelling with R-INLA. *Journal of statistical software*, 63(19):1–25.

[Lindgren et al., 2011] 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. *Journal of the Royal Statistical Society Series B: Statistical Methodology*, 73(4):423–498.

[Makridakis et al., 1982] Makridakis, S., Andersen, A., Carbone, R., Fildes, R., Hibon, M., Lewandowski, R., Newton, J., Parzen, E., and Winkler, R. (1982). The accuracy of extrapolation (time series) methods: Results of a forecasting competition. *Journal of forecasting*, 1(2):111–153.

[Makridakis and Hibon, 2000] Makridakis, S. and Hibon, M. (2000). The m3-competition: results, conclusions and implications. *International journal of forecasting*, 16(4):451–476.

[Masjedi and Crawford, 2020] Masjedi, A. and Crawford, M. M. (2020). Prediction of sorghum biomass using time series uav-based hyperspectral and lidar data. In *IGARSS 2020-2020 IEEE International Geoscience and Remote Sensing Symposium*, pages 3912–3915. IEEE.

[McDermott and Wikle, 2019] McDermott, P. L. and Wikle, C. K. (2019). Bayesian recurrent neural network models for forecasting and quantifying uncertainty in spatial-temporal data. *Entropy*, 21(2):184.

[Medina, 2007] Medina, Fernando y Galván, M. (2007). *Imputación de datos: teoría y práctica*. Cepal.

[Nelder and Wedderburn, 1972] Nelder, J. A. and Wedderburn, R. W. (1972). Generalized linear models. *Journal of the Royal Statistical Society Series A: Statistics in Society*, 135(3):370–384.

[Nikparvar and Thill, 2021] Nikparvar, B. and Thill, J.-C. (2021). Machine learning of spatial data. *ISPRS International Journal of Geo-Information*, 10(9):600.

[Niraula et al., 2022] Niraula, P., Mateu, J., and Chaudhuri, S. (2022). A bayesian machine learning approach for spatio-temporal prediction of covid-19 cases. *Stochastic Environmental Research and Risk Assessment*, 36(8):2265–2283.

[Núñez Medina et al., 2019] Núñez Medina, G., López Arévalo, J., et al. (2019). Modelación espacial bayesiana del riesgo de inmigración municipal en chiapas.

[Perez, 2011] Perez, M. (2011). *Microsoft SQL Server 2008 R2. Motor de base de datos y administración*. RC Libros.

[Pérez López, 2007] Pérez López, César y Santin González, D. (2007). *Minería de datos. Técnicas y herramientas: técnicas y herramientas*. Ediciones Paraninfo, SA.

[R Core Team, 2023] R Core Team (2023). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria.

[Ravishanker et al., 2022] Ravishanker, N., Raman, B., and Soyer, R. (2022). *Dynamic Time Series Models Using R-INLA: An Applied Perspective*. CRC Press.

[Resch et al., 2018] Resch, B., Usländer, F., and Havas, C. (2018). Combining machine-learning topic models and spatiotemporal analysis of social media data for disaster footprint and damage assessment. *Cartography and geographic information science*, 45(4):362–376.

[Rodrigues, 2017] Rodrigues, Â. A. (2017). *Spatio-temporal modelling of tornados with R-INLA, at the county-level in Texas and Ocklahona*. PhD thesis.

[Roger Bivand, 2022] Roger Bivand (2022). R packages for analyzing spatial data: A comparative case study with areal data. *Geographical Analysis*, 54(3):488–518.

[Rue and Held, 2005] Rue, H. and Held, L. (2005). *Gaussian Markov random fields: theory and applications*. CRC press, New York.

[Rue et al., 2009] Rue, H., Martino, S., and Chopin, N. (2009). Approximate bayesian inference for latent gaussian models by using integrated nested laplace approximations. *Journal of the Royal Statistical Society Series B: Statistical Methodology*, 71(2):319–392.

[Sakia, 1992] Sakia, R. M. (1992). The box-cox transformation technique: a review. *Journal of the Royal Statistical Society Series D: The Statistician*, 41(2):169–178.

[Sartorius et al., 2021] Sartorius, B., Lawson, A., and Pullan, R. (2021). Modelling and predicting the spatio-temporal spread of covid-19, associated deaths and impact of key risk factors in england. *Scientific reports*, 11(1):5378.

[Shakti et al., 2017] Shakti, S. P., Hassan, M. K., Zhenning, Y., Caytiles, R. D., and Iyenger, N. (2017). Annual automobile sales prediction using arima model. *International Journal of Hybrid Information Technology*, 10(6):13–22.

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

[Thisted, 1998] Thisted, R. A. (1998). What is a p-value. *Departments of Statistics and Health Studies*, 20(6):7–16.

[van Niekerk and Rue, 2021] van Niekerk, J. and Rue, H. (2021). Correcting the laplace method with variational bayes. *arXiv preprint arXiv:2111.12945*, pages 1–18.

[Villavicencio, 2010] Villavicencio, J. (2010). Introducción a series de tiempo. *Puerto Rico*, pages 1–33.

[Wall, 2004] Wall, M. M. (2004). A close look at the spatial structure implied by the car and sar models. *Journal of statistical planning and inference*, 121(2):311–324.

[Watanabe and Opper, 2010] Watanabe, S. and Opper, M. (2010). Asymptotic equivalence of bayes cross validation and widely applicable information criterion in singular learning theory. *Journal of machine learning research*, 11(12):5–15.

[Wei, 2018] Wei, W. W. (2018). *Multivariate time series analysis and applications*. John Wiley & Sons, New York.

[Zhang et al., 2018] Zhang, C., Sargent, I., Pan, X., Li, H., Gardiner, A., Hare, J., and Atkinson, P. M. (2018). An object-based convolutional neural network (ocnn) for urban land use classification. *Remote sensing of environment*, 216:57–70.

[Zhang, 2016] Zhang, Z. (2016). Missing data imputation: focusing on single imputation. *Annals of translational medicine*, 4(1).

[Zuur et al., 2017] Zuur, A., Elena, N., and Anatoly, A. (2017). Beginner's guide to spatial, temporal, and spatial-temporal ecological data analysis with r-inla volume i: using glm and glmm. highland statistics ltd. *Newburgh United Kingdom*.