Santiago de Cali, junio 13 de 2023

Doctor

Andrés Felipe Amador Rodríguez

Director de la carrera de Matemáticas Aplicadas

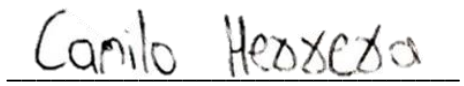
PONTIFICIA UNIVERSIDAD JAVERIANA CALI

Cordial Saludo,

Por medio de la presente me permito hacer la entrega a usted de mi proyecto de grado titulado **"COMPARACIÓN DEL MÉTODO DE COMPONENTES DEMOGRÁFICAS Y PROPUESTA ALTERNATIVA PARA LAS PROYECCIONES POBLACIONALES DE JAMUNDÍ"**, para ser evaluado por la facultad.

Espero que este proyecto cumpla con los requisitos estipulados para su aprobación.

Atentamente,



Juan Camilo Herrera Palacio
CC. 1112498660
Código estudiantil: 8936380

Santiago de Cali, junio 13 de 2023

Doctor

Andrés Felipe Amador Rodríguez

Director de la carrera de Matemáticas Aplicadas

PONTIFICIA UNIVERSIDAD JAVERIANA CALI

Cordial Saludo,

Por medio de la presente me permito informarle que el estudiante del programa de Matemáticas Aplicadas, Juan Camilo Herrera Palacio con código 8936380, trabajó y finalizó bajo mi dirección, el proyecto de grado denominado **"COMPARACIÓN DEL MÉTODO DE COMPONENTES DEMOGRÁFICAS Y PROPUESTA ALTERNATIVA PARA LAS PROYECCIONES POBLACIONALES DE JAMUNDÍ"**, el cual considero se encuentra en condiciones para ser sometido a evaluación.

Atentamente,

Isabel C. García A.

Isabel Cristina García Arboleda
Profesora
Departamento de Ciencias Naturales y Matemáticas
Pontifica Universidad Javeriana Cali

Comparación del método de componentes demográficas y propuesta alternativa para las proyecciones poblacionales de Jamundí

Juan Camilo Herrera Palacio

Trabajo de grado

Dirigido por:
Isabel Cristina García Arboleda, PhD

Pontificia Universidad Javeriana Cali
Facultad de Ingeniería y Ciencias
Programa de Matemáticas Aplicadas
2023

Índice general

1. Introducción

El estudio acerca de los niveles poblacionales en ciertas áreas geográficas resulta imprescindible hoy día para la toma de decisiones, tal como lo resltan Jacob & Swanson (2004) [14] en los últimos años ha incrementado el uso de la demografía en materia de salud pública, planeación gubernamental local, planeación de los recursos tanto humanos como empresariales, medio ambiente, y tráfico. Ante esto, el Departamento Nacional de Planeación (DNP) también ha expuesto la importancia y ha incursionado en proyectos que permitan acentuar las bases para el entendimiento de las relaciones entre población, desarrollo y territorio [9]; al respecto, ha mencionado el principio de "... considerar a la población como el centro del desarrollo y por ende la planeación y el diseño de la política pública deben estar orientadas a la población y tomarla en cuenta como actor indispensable".

En Colombia varios procesos y planes de índole pública se ven supeditados ante las dinámicas poblacionales. Por ejemplo, el decreto 1232 de 2020[1] que reglamenta la Ley 388 de 1997[2], propone que "... para la definición del suelo de expansión urbana se debe considerar: (i) Las previsiones de crecimiento de la ciudad en función de las dinámicas demográficas y poblacionales. ..."; otro ejemplo radica en el decreto 1652 del 2021[3], en el cual se evidencia la importancia del tamaño poblacional para la determinación de las tasas impositivas, específicamente en el Artículo 1.2.1.28.2.15 se expresa que el Departamento Administrativo Nacional de Estadística (DANE) es el ente que debe certificar oficialmente la población municipal en Colombia, para cumplir con este propósito.

Si bien el estudio de la población concierne a la demografía, Swanson & Siegel (2004) [14] afirman que "... La necesidad de gestionar la incertidumbre en las estimaciones y proyecciones poblacionales ha conducido a la aplicación de la teoría de la decisión, el análisis de series temporales y la teoría de la probabilidad a métodos para fijar los límites de confianza de las estimaciones y proyecciones - un proceso denominado estimación y proyección demográfica estocástica". Esto es, la necesidad de conocer con mayor precisión el comportamiento futuro de la población, ha incrementado la aplicación de modelos matemáticos y su incidencia en este tema, específicamente el análisis de series temporales y procesos estocásticos.

Pese a que Colombia y México han sido reconocidos por el Banco Interamericano de Desarrollo (BID) (2019) [5], como los países de América Latina y el Caribe (ALC) que

[1]Decreto 1232 del 2020. "Por medio del cual se adiciona y modifica el artículo 2.2. 1.1 del Título 1, se modifica la Sección 2 del Capítulo 1 del Título 2 y se adiciona al artículo 2.2.4.1.2.2 de la sección 2 del capítulo 1 del Título 4, de la Parte 2 del Libro 2 del Decreto 1077 de 2015 Único Reglamentario del Sector Vivienda, Ciudad y Territorio, en lo relacionado con la planeación del ordenamiento territorial". Recuperado de: *https://www.funcionpublica.gov.co/eva/gestornormativo/norma.php?i=142020.*

[2]Ley 388 de 1997 (Ley de ordenamiento territorial). "Por la cual se modifica la Ley 9 de 1989, y la Ley 2 de 1991 y se dictan otras disposiciones". Recuperada de: *https://www.funcionpublica.gov.co/eva/gestornormativo/norma.php?i=339.*

[3]Decreto 1652 del 2021. "Por el cual se reglamentan los literales a), b), c), d), f), g), h), e i) del parágrafo 5 del artículo 240 del Estatuto Tributario, modificado por el artículo 41 de la Ley 2068 de 2020 y se adicionan las Secciones 2 y 3 al Capítulo 28 del Título 1 de la Parte 2 del Libro 1del Decreto 1625 de 2016, Único Reglamentario en Materia Tributaria". Recuperado de: *https://www.dane.gov.co/files/acerca/Normatividad/decretos/DECRETO-1652-DE-2021.pdf*.

empiezan a actualizar sus métodos de proyecciones poblacionales, el estudio del BID resalta la debilidad general de la región en esta materia, pues existen diversos motivos que han desembocado en una constante aplicación de metodologías que no están a la vanguardia estadística y no responden a las particularidades socioeconómicas de las áreas geográficas para las cuales se ofrecen proyecciones de población. Esto, destaca el BID (2019) [5] y otros estudios, es muy delicado y peligroso debido a la importancia de esta variable en la toma de decisiones públicas y privadas; especialmente, la evidencia empírica que muestran algunas investigaciones como la de Wilson, Grossman & Temple (2021) [17] destacan que las metodologías de proyecciones poblacionales suelen fallar más a medida que el área geográfica en la cual se proyecta es menor, es decir, específicamente en áreas subnacionales (entiéndase departamentos, municipios, localidades, comunas, barrios, etc).

Esta investigación toma el caso particular del municipio de Jamundí (Valle del Cauca - Colombia), para estudiar diferentes métodos de proyección poblacional y comparar los resultados obtenidos con las publicaciones del DANE. En adelante, el anteproyecto de grado se divide de la siguiente manera: definición del problema (2); objetivos generales y específicos (3); alcance (4); estado del arte (5); marco teórico (6); metodología (7).

2. Definición del problema

Convencionalmente, las proyecciones poblacionales en Colombia se han realizado mediante el método de componentes demográficas, es decir, se estiman las variables determinantes del crecimiento poblacional: fecundidad, mortalidad y migración. Para la componente de migración, a diferencia de las componentes de la dinámica vegetativa, no existe información que ayude a verificar las estimaciones y modelaciones matemáticas, como sí ocurre al nivel nacional mediante los registros de fronteras; es por esto que para esta estimación se suele utilizar la información consolidada mediante los censos poblacionales.

Así, queda en evidencia la dificultad de realizar proyecciones poblacionales a nivel sub-nacional (situación que ocurre también en toda la región de ALC); esto se debe a que, pese a la información recolectada en los censos, estos no se realizan de manera continua cada año, particularmente en nuestro país los últimos cuatro censos se realizaron en los años 1993, 2005 y 2018, es decir, ha habido un espaciado temporal de aproximadamente 10 años, impidiendo tener un conocimiento aproximado a la realidad entre los periodos intercensales mediante esta operación estadística.

Con lo anterior, el problema de esta investigación consiste en encontrar modelos alternativos al utilizado por el DANE para la realización de proyecciones poblacionales a nivel municipal, puntualmente aplicado al caso de Jamundí.

3. Objetivos

3.1. Objetivo general

Estudiar diferentes alternativas de proyecciones poblacionales que permitan tener estimaciones más cercanas a la realidad.

3.2. Objetivos específicos

- Encontrar, entre la información disponible para el municipio de Jamundí, las variables que permitan aproximar el nivel poblacional actual.
- Proyectar la población del municipio de Jamundí con diferentes métodos.
- Comparar los resultados obtenidos por los diferentes métodos estimados con los resultados entregados por el DANE.

4. Alcance

Debido a las limitaciones de información existente para los niveles subnacionales del país, en este trabajo de grado no se podrá conocer con certeza si los resultados obtenidos con los diferentes métodos de proyecciones poblacionales realmente se ajustan a la realidad actual del municipio de Jamundí; sin embargo, sí se pretende realizar un estudio comparativo y plantear alternativas que pueden resultar de interés para la futura consideración del DANE, especialmente para las áreas geográficas menores.

5. Estado del arte

Una de las instituciones que publica resultados sobre la dinámica, el comportamiento y las proyecciones de población, a nivel internacional, es la Organización de las Naciones Unidas (ONU). Recientemente frente a su publicación sobre las ”Perspectivas de población mundial 2019”, la Comisión Económica para América Latina y el Caribe (CEPAL) (2020) [16] se encargó de traducir el documento sobre la metodología de la ONU para las estimaciones y proyecciones poblacionales. En este documento, se exhibe que el método estimado corresponde al de Componentes por Cohortes, este a su vez fue publicado por la misma ONU desde 1956 [2], convirtiéndose en el referente metodológico para las proyecciones de población de los países de la región [1]. A escala regional para los países de América Latina, según la CEPAL [1] el Centro Latinoamericano y Caribeño de Demografía (CELADE) y la División de Población de las Naciones Unidas (DPNU) son las entidades encargadas de producir las estimaciones y proyecciones poblacionales de veinte países, los trece países restantes de la región son suplidos por las estimaciones de la DPNU.

De esta manera, si bien como lo afirma la CEPAL [1], la metodología ha tenido actualizaciones desde el momento de su publicación, especialmente debido a los progresos tecnológicos y en materia de computación, queda en evidencia que esta metodología lleva más de sesenta años (64 años) implementándose. Sin embargo, se debe destacar que una de las razones que podría influir en la permanencia de este método a día de hoy, resulta en que, por su naturaleza y definición, genera proyecciones desagregadas por sexo y edad (esto hace alusión a su nombre), pero también permite la elaboración de diferentes indicadores demográficos, que son de gran utilidad para la toma de decisiones; además, involucra también los tres factores de crecimiento poblacional desde una perspectiva demográfica, natalidad, mortalidad y migración [1].

A pesar de las facilidades que permite el modelaje mediante componentes por cohortes, el BID (2019) [5] se pregunta por "¿qué pueden hacer los países de ALC para mejorar la calidad de sus proyecciones de población a fin de maximizar la eficiencia de la implementación de las políticas públicas?", lo cual apunta a la carencia observada en las proyecciones de población publicadas por las Oficinas Nacionales de Estadística (ONE) de los países de la región. El estudio intenta responder a esta pregunta mediante el análisis de las diferentes metodologías de proyección poblacional, pero también considera aspectos institucionales. Para este fin, el BID realizó entrevistas a funcionarios de los 15 países inmersos en el estudio que se encargan de la sproyecciones de población, a 9 expertos demográficos e investigadores de las estadísticas de ALC, y a 7 expertos demográficos por fuera de ALC que usan estas proyecciones; además, realizó un análisis de la bibliografía académica donde se norman, discuten, describen y critican proyecciones poblacionales; sumado a lo anterior, se realizó un comparativo entre los 15 países incluidos y un grupo de países por fuera de ALC que están a la vanguardía en métodos estadísticos para el fin de interés. En conclusión, el estudio resalta que Colombia y México son los únicos dos países que empiezan a actualizar sus metodologías, pero de igual manera, en general ALC se encuentra distante en las metodologías frente a países que cuentan con métodos a la vanguardía, y entre otras, resalta también la debilidad en materia institucional de los productores de proyecciones poblacionales.

En continuidad con la línea anterior frente a la crítica existente con la producción de proyecciones poblacionales, el artículo de Wilson, Grossman & Temple (2021) [17] evalua diferentes métodos de pronóstico poblacional para áreas pequeñas, esto debido al histórico desacierto de las proyecciones para zonas subnacionales al ser contrastadas con la realidad; frente a esto, resaltan que los principales motivos pueden ser la escaza calidad de datos en estas áreas, y el gran componente aleatorio en comunidades pequeñas, lo cual dificulta enmarcar los patrones demográficos, pero también mencionan la carencia de investigación en áreas subnacionales, la cual se ve disminuida por la atención prestada a proyecciones nacionales. EL estudio concluye que es relevante incorporar a los modelos las limitaciones locales de la zona, para evitar crecimientos irrazonables en la proyección de cada zona; también se menciona las diferencias encontradas entre las proyecciones nacionales y la suma de las proyecciones para áreas subnacionales. Los investigadores no logran seleccionar alguno de los métodos evaluados, pues, todos producen errores inaceptables en algunas de las áreas pequeñas, y menciona la necesidad de mayor investigación en el tema y el contraste frente a metodologías recientes.

Por otro lado, recientemente Raftery et al. (2021) [13] publicaron un estudio donde extienden la metodología publicada por la (ONU) desde el 2015, donde se utiliza un enfoque probabilístico bayesiano para las proyecciones de población en el largo plazo. El estudio se encuentra motivado principalmente por las emisiones de carbono, que dependen de las previsiones demográficas, y por el uso de la metodología bayesiana reciente, pues, destacan el uso tradicional de métodos deterministas para este tipo de propósitos. Mencionan que la ciencia aún no llega a un concenso frente a la modelización detallada, y se debe continuar en la investigación al respecto.

Referente a las proyecciones poblacionales en Colombia, al revisar la metodología

publicada por el DANE (2009) [8] se encuentra el método de componentes demográficas, donde se estiman las tasas de natalidad, mortalidad y migración, y también el método de relación por cohortes, es decir, la metodología propuesta por la ONU desde 1956 para las estimaciones y proyecciones de población en los países de la región. También se destaca que el DANE incorpora en cierta medida las limitaciones locales de cada área subnacional a proyectar (departamental, municipal, clase[4].) mediante el punto de partida de su enmarque suprazonal, es decir, las proyecciones departamentales están supeditadas al pronóstico nacional, las municipales a la departamental, y las de clase a la municipal.

Finalmente, en el ámbito de las proyecciones poblacionales en pequeñas delimitaciones geográficas, recientemente el DANE (2020) [6] realizó las proyecciones de población, hogares y vivienda, para las localidades de la ciudad capital (Bogotá), desde 2019 hasta 2035, y de las Unidades de Planificación Zonal (UPZ) en el periodo 2019 - 2024. Para generar estas proyecciones el DANE sigue variaciones de la metodología publicada originalmente en 2005 [8], pero con la información desagregadas para las áreas geográficas mencionadas. Esto es, la aplicación de la estimación individual de los componentes demográficos, natalidad, mortalidad y migración, en conjunción de la división por cohortes. Además, el DANE hace especial énfasis en la distinción que se debe hacer entre población, hogares y viviendas, y la estimación individual de estas variables, debido a que si bien se relacionan entre ellas y son interdependientes, mencionando que es riesgoso utilizar factores como personas por hogar, u hogares por vivienda, que permanezcan constantes en el tiempo para la determinanción de una de las variables con base en la otra.

6. Marco teórico

Las series de tiempo, como son conocidas habitualmente, corresponden a listados de datos que guardan un orden (son indexados) mediante una frecuencia temporal, por ejemplo, minutos, horas, días, semanas, meses, años, entre otros. La fecuencia con que se presenta la marca temporal en una variable de tiempo, suele depender de la complejidad o naturaleza de esta, así, variables macroeconómicas como el PIB, el desempleo, y la inflación suelen tener medidas trimestrales, semestrales o anuales, pues, su medida requiere de la captura de grandes volúmenes de datos para muchas unidades de observación.

Por ejemplo, la estimación de la tasa de desempleo de un país, requiere de un muestreo en ciertas ciudades. En el caso colombiano, esta medida se presenta mensualmente para trece ciudades y áreas metropolitanas, mediante la captura de información con la Gran Encuesta Integrada de Hogares (GEIH) desarrollada por el DANE. Si bien esta información da cuenta de una idea general acerca del desempleo en la nación, no puede ser tomada como realidad para cada uno de los 1.101 municipios del país. Lo anterior da cuenta de la complejidad de este tipo de medidas, y que por ende, requiere un gran esfuerzo en términos presupuestales, técnicos y de tiempo.

Aunque el enfoque del análisis de series de tiempo para prever el futuro parece reciente,

[4]Clase: cabeceras municipales, centros poblados y rural disperso

lo actual son los métodos y herramientas matemáticas que se han desarrollado, pues, según Hyndman (2008) [11] los pronósticos han fascinado a la humanidad durante miles de años. Ante este escenario, desde los tiempos babilónicos se han desarrollado diversos métodos tanto cuantitativos como cualitativos, para prever el futuro. En este apartado no se pretende brindar una reseña histórica acerca del desarrollo de las metodologías, ni de la evolución de los conceptos, en cambio, se centra en aclarar a manera introductoria los conceptos clave y los métodos que serán implementados para alcanzar el objetivo del trabajo (1); estos serán de utilidad para entender los procesos que se desarrollan en la metodología (7), y en la exposición de resultados.

A continuación, el capítulo se desarrolla en las siguientes secciones: introducción a los conceptos clave de series de tiempo, para lo cual se sigue la exposición de Box & Jenkins (2016) [3]; método de suavizado exponencial, desarrollado originalmente entre 1950 y 1960 por Holt (1957) [10], Brown (1959) [4], Winters (1960) [18], para lo cual se toma la explicación dada por Hyndman (2008) [12]; conceptualización de los modelos ARIMA, para esto se sigue a Hyndman (2018) [11] y Box & Jenkins (2016) [3].

6.1. Introducción a las series de tiempo

Los dos conceptos fundamentales para entender el marco de este trabajo son los de serie de tiempo y proceso estocástico. Estos se presentan a continuación:

- Una **serie de tiempo** es un conjunto de observaciones generado secuencialmente a través del tiempo. Si el conjunto es continuo, se dice que la serie de tiempo es continua; análogamente se define **serie de tiempo discreta**. El marco de este trabajo está bajo la base de las series de tiempo discretas donde las observaciones se presentan en intervalos de tiempo equidistantes.

  Esto último significa que, si consideramos la serie de tiempo $u = u_1, u_2, \cdots, u_t, \cdots, u_N$, la diferencia en el momento de tiempo entre dos observaciones consecutivas es siempre fija. Llamemos h a esta distancia en el tiempo, entonces se concluye que $u_2 = u_{1+h}$, $u_3 = u_{2+h} = u_{1+2h}$, y así sucesivamente hasta que $u_N = u_{1+h(N-1)}$.

  Además, existen también otros dos tipos de series de tiempo, las deterministas y las estadísticas. Cuando la serie de tiempo puede ser descrita por una función matemática exacta para cada instante del tiempo, se dice que esta es determinista; por otro lado, si los términos de la serie de tiempo solo pueden ser descritos por distribuciones de probabilidad, entonces la serie es estadística. Nuevamente, el marco de este trabajo está direccionado por las series de tiempo estadísticas.

- Un **proceso estocástico** es un fenómeno estadístico que evoluciona en el tiempo acorde a leyes probabilísticas.

  La importancia de este concepto para esta investigación, subyace en que la serie de tiempo debe ser pensada como una realización particular, producida por el mecanismo de probabilidad inmerso en el sistema de estudio. Es decir, al analizar

una *serie de tiempo*, se debe tener en mente que esta es la realización de un *proceso estocástico*.

Una clase muy especial de procesos estocásticos, son los **procesos estacionarios**. Estos, asumen que el proceso está en un estado particular de equilibrio estadístico. Se dice que un proceso estocástico es **estrictamente estacionariono** si sus propiedades no se ven afectadas por un cambio en el origen del tiempo, es decir, si la distribución de probabilidad conjunta asociada con n observaciones del proceso $z_{t_1}, z_{t_2}, \cdots, z_{t_n}$, realizadas en un conjunto cualquiera de tiempo $t_1, t_2, \cdots, t_n$, es la misma que la asociada con n observaciones $z_{t_1+k}, z_{t_2+k}, \cdots, z_{t_n+k}$. Esto quiere decir que, para garantizar que un proceso discreto sea estrictamente estacionario, se debe cumplir que la distribución de probabilidad conjunta de cualquier conjunto de observaciones, no debe ser afectada ante cambios en el tiempo hacía atrás o adelante en una cantidad entera cualquiera.

Un concepto asociado al anterior es el de **estacionariedad débil de orden** f. Se dice que, por ejemplo, un proceso z_t es debilmente estacionario de orden 2, o estacionario de orden 2, si la media $E[z_t] = \mu$ es una constante fija para todo t y las autocovarianzas $cov[z_t, z_t + k] = \gamma_k$ dependen solo en la diferencia de tiempo k para todo tiempo t.

Para ejemplificar el concepto anterior se introduce el concepto de **ruido blanco**. Los procesos de ruido blanco consisten en secuencias de variables aleatorias independientes e identicamente distribuidas, para las cuales también se asume media cero y varianza σ^2. Si se denota este proceso como a_t, entonces se obtiene que debido a su definición, la función de autocovarianza es

$$\gamma_k = E[a_t a_{t+k}] = \begin{cases} \sigma_a^2 & k = 0 \\ 0 & k \neq 0 \end{cases} \tag{1}$$

6.2. Modelos de suavizado exponencial

Como se ha mencionado, tal como lo resalta Hyndman (2008) [12] aunque los modelos de suavizado exponencial fueron desarrollados desde cerca de 1950, un marco de trabajo completo acerca del modelamiento estocástico, calculos de probabilidad, intervalos de predicción y procesos para la selección de modelos, no fue desarrollado hasta hace pocos años, con los trabajos de Ord et al. (1997) y Hyndman et al. (2002).

Para entender a grandes rasgos las ideas generales de esta clase de modelos, se debe abordar en primera instancia, lo que se conoce como descomposición de series de tiempo. En este sentido, a continuación se aborda primero la temática en mención, y se finaliza con la definición de suavizado exponencial, y la clasificación de estos modelos.

Descomposición de series de tiempo

En aplicaciones prácticas como la economía y los negocios, es normal pensar en las series de tiempo como la combinación de diferentes componentes, estos son: la tendencia (T), el ciclo (C), la estacionalidad (S), y un componente irregular o aleatorio (E). Estos se definen brevemente a continuación

- **Tendencia (T):** la dirección en el largo plazo de la serie.

- **Cycle (C):** patrón que se repite con alguna regularidad pero con una periodicidad desconocida.
- **Estacionalidad (S):** patrón que se repite con una periodicidad conocida.
- **Irregular (E):** parte impredescible de la serie.

Es común suponer que la componente cíclica se encuentra inmersa en el comportamiento tendencial de la serie de tiempo, por tanto, es habitual reducir el estudio en los modelos de suavizado exponencial a solo 3 componentes: tendencia, estacional e irregular. Además, estos tres componentes pueden ser combinados de manera multiplicativa o aditiva, como se muestra en las ecuaciones (2 y 3), respectivamente, denotando a y como una serie de tiempo cualquiera.

$$y = T * S * E \tag{2}$$

$$y = T + S + E \tag{3}$$

También es posible combinar de manera mixta los componentes, es decir, considerando agrupaciones de manera multiplicativa y otras aditivas.

Definición de los modelos de suavizado exponencial

Los modelos de suavizado exponencial son una familia de modelos con una propiedad en común, los pronósticos corresponden a combinaciones ponderadas de observaciones pasadas, donde las observaciones más cercanas al período de pronóstico tienen mayor peso que las pasadas. En este sentido, el nombre hace referencia a que la ponderación decrece exponencialmente a medida que la observación es más antigua.

En esta clase de modelos siempre se empieza con la modelación del componente de tendencia, el cual se define como una combinación del término de nivel (l) y el término de crecimiento (b). El nivel y la tendencia se pueden combinar de diversas formas. Si T_h denota la tendencia proyectada en los próximos h periodos de tiempo, y $\phi \in]0,1[$ denota un parámetro de amortiguación, entonces los cinco tipos de tendencia o patrones de crecimiento se definen de la siguiente manera

- **Ninguno:** $T_h = l$
- **Aditivo:** $T_h = l + bh$
- **Amortiguación aditiva:** $T_h = l + (\phi + \phi^2 + \cdots + \phi^h)b$
- **Multiplicativo:** $T_h = lb^h$
- **Amortiguación multiplicativa:** $T_h = lb(\phi + \phi^2 + \cdots + \phi^h)$

Las tendencias factor amortiguador resultan de gran ayuda práctica cuando se conoce la tendencia en la serie de tiempo, pero se asumen que esta será distinta al final de la serie frente al inicio. En este sentido, el nombre de amortiguación corresponde con la función del factor, pues amortigua la tendencia de la serie a medida que el horizonte de tiempo incrementa.

Una vez se ha escogido el componente de tendencial, se elige la componente estacional y de error, ambas también de manera multiplicativa o aditiva. Históricamente, la naturaleza (aditiva o multiplicativa) del factor de error ha sido ignorada, debido a que la elección de esta naturaleza no presenta grandes variaciones en los resultados finales de pronósticos puntuales.

Para cerrar esta sección, se debe aclarar la diferencia entre los modelos de suavizado exponencial y los de espacio estado (no abordados en este trabajo). El método de **suavizado exponencial** es un algoritmo que solo produce pronósticos puntuales. Los modelos de espacio estado producen los mismos pronósticos puntuales, pero también producen un marco para encontrar intervalos de predicción y otras propiedades.

6.3. Modelos ARIMA

Antes de introducir generalmente los modelos ARIMA, se deben entender tres procesos fundamentales: procesos generales lineales; procesos autorregresivos; procesos de media móvil. Estos tres se introducen a continuación a manera conceptual, sin profundizar en detalles, demostraciones y propiedades.

Procesos generales lineales

Los **Procesos Generales Lineales** (GLP) son representaciones de procesos estocásticos como resultado de un filtro lineal, donde el argumento de este filtro es un proceso ruido blanco. Matemáticamente se representa como

$$\begin{aligned} \tilde{z}_t &= a_t + \Psi_1 a_{t-1} + \Psi_2 a_{t-2} + \cdots \\ &= a_t + \sum_{j=1}^{\infty} \Psi_j a_{t-j} \end{aligned} \tag{4}$$

donde $\tilde{z}_t = z_t - \mu$ es la desviación del proceso de algún origen, o de la media si el proceso es estacionario. Es decir, la importancia del Proceso General Lineal (4) radica en que permite representar $\tilde{z}_t$ como una suma ponderada de valores pasados y presentes del ruido blanco a_t.

Para que $\tilde{z}_t$ represente un proceso estacionario, es necesario que los coeficientes Ψ_j sean absolutamente sumables, es decir, se debe cumplir $\sum_{j=0}^{\infty} |\Psi_j| < \infty$. Ante condiciones adecuadas, $\tilde{z}_t$ también es también una suma ponderada de $\tilde{z}_t$'s y un ruido blanco adicional a_t, como se ilustra acontinuación

$$\begin{aligned} \tilde{z}_t &= \pi_1 \tilde{z}_{t-1} + \pi_2 \tilde{z}_{t-2} + \cdots + a_t \\ &= \sum_{j=1}^{\infty} \pi_j \tilde{z}_{t-j} + a_t \end{aligned} \tag{5}$$

Es claro que, si la representación (4 & 5) del proceso general lineal tiene infinitos términos, no resulta útil en la práctica. Por tanto, a continuación se desarrollan los procesos autorregresivos y de media móvil, los cuales pretenden simplificar los PGL.

Procesos autorregresivos

Un **proceso autorregresivo** de orden p, o brevemente proceso AR(p), consiste en considerar el caso especial de (5) en el que solo los primeros p ponderadores son diferentes de 0. Es decir,

$$\tilde{z}_t = \phi_1 \tilde{z}_{t-1} + \phi_2 \tilde{z}_{t-2} + \cdots + \phi_p \tilde{z}_{t-p} + a_t \tag{6}$$

Los casos particulares AR(1) y AR(2) son de gran importancia y uso práctico en la modelación de series de tiempo, pues, con estos suelen estar inscritos en muchos de los modelos que intentan describir diversas variables en la realidad.

Procesos de media móvil

Los **procesos de media móvil** de orden q, o más brevemente MA(q), son casos especiales de la ecuación (4), donde solo los primeros q términos de los ponderadores Φ son distintos de cero. Matemáticamente, esto es

$$\tilde{z}_t = a_t - \theta_1 a_{t-1} - \theta_2 a_{t-2} - \cdots - \theta_q a_{t-q} \tag{7}$$

Tal como en el caso de los AR(p), los procesos MA(1) y MA(2), son de gran importancia en la práctica.

Procesos mixtos autorregresivos y de media móvil

Los **procesos mixtos autorregresivos y de media móvil** (ARMA, por sus siglas en inglés), son modelos que resultan de combinar tanto los procesos de media móvil como los de autocorrelación. El resultado se presenta a continuación

$$\tilde{z}_t = \phi_1 \tilde{z}_{t-1} + \cdots + \phi_p \tilde{z}_{t-p} + a_t - \theta_1 a_{t-1} - \cdots - \theta_q a_{t-q} \tag{8}$$

o, de otra manera

$$\phi(B)\tilde{z}_t = \theta(B)a_t \tag{9}$$

Los procesos con la forma de las ecuaciones (8 y 9), se denominan ARMA(p, q), guardando congruencia con el orden de su componente autorregresiva y de media móvil, respectivamente. Además, dado que $\tilde{z}_t = z_t - \mu$, donde $\mu = E[z_t]$ es la media del proceso en el caso estacionario, en general el proceso ARMA(p, q) puede escribirse también en términos del proceso original z_t, de la siguiente forma

$$\phi(B)z_t = \theta_0 + \theta(B)a_t \tag{10}$$

donde el término cosntante θ_0 es

$$\theta_0 = (1 - \phi_1 - \phi_2 - \cdots - \phi_p)\mu \tag{11}$$

Procesos integrados mixtos autorregresivos y de media móvil

Finalmente, dentro de la conceptualización de los modelos ARIMA, se termina con el **proceso integrado mixto autorregresivo y de media móvil**, los cuales corresponden a series de tiempo que no tienen media fija a través del tiempo, aunque estos exhiben cierta homogeneidad y guardan tendencias a nivel local.

En términos sencillos, estos modelos corresponden a modelos ARMA no estacionarios, que pueden alcanzar la condición de estacionariedad mediante la aplicación del operador diferenciador a la componente autorregresiva. En términos matemáticos, el modelo se expresa de la siguiente manera

$$\phi(B) \bigtriangledown^d z_t = \theta(B) a_t \tag{12}$$

donde el operador diferenciador $\bigtriangledown = 1 - B$, y el orden de su potencia (d) corresponde al orden de la diferenciación. De esta manera, estos modelos se suelen escribir como ARIMA(p, d, q) guardando coherencia con el orden de la componente autorregresiva, de diferenciación del modelo, y de media móvil.

6.4. Método Propeht

Pese a que la traducción del método que se expone en esta sección sería ”Profeta”, en adelante se designará como ”Prophet”, acorde a su nombre original en inglés. Además, se aclara que, si bien este método se podría enmarcar rigurosamente dentro de modelos de regresión, aquí no se hará énfasis en la explicación de esos modelos en general, sino que se describirá puntualmente el marco general de esta metodología, para mayor detalle se recomienda consultar [15].

Una de las principales características del método consiste en que permite incorporar el conocimiento intuitivo del usuario frente a los datos a modelar, sin necesidad de conocer explícitamente el modelo. La base del método es la combinación aditiva de tres componentes principales: tendencia, estacionalidad y días festivos; tal como se muestra en la ecuación 13.

$$y(t) = g(t) + s(t) + h(t) + \varepsilon_t \tag{13}$$

donde:

$y(t)$: valor de la serie de tiempo
$g(t)$: representa la tendencia, estos son, cambios no periódicos
$s(t)$: representa cambios periódicos
$h(t)$: representa el efecto de los días festivos
ε_t : representa cualquier cambio que no es capturado por el modelo.

Este modelo, como se mencionó al inicio, se puede clasificarse como un modelo de regresión con suavizadores en sus variables regresoras. Note que, de igual forma, el modelar la estacionalidad como un componente aditivo, genera ciertas simiitudes con el método de suavizado exponencial expuesto anteriormente. Además, se debe recalcar que este método se originó frente a las necesidades de proyección con base en los datos de

Facebook, por tanto, la modelación de cada uno de los componentes está diseñada para que se ajuste de la mejor manera a los datos que suele utilizar la compañía.

En esencia, este modelo busca un ajuste de la curva a modelar frente a los datos, lo que lo diferencia de los modelos tradicionales de series de tiempo que utilizan explícitamente la dependencia temporal en la estructura de la serie. En términos prácticos, pese a que el método no permite realizar diversos procesos de inferencia estadística, los autores proponen las siguientes ventajas de este método frente a otros:

- La flexibilidad de que el usuraio haga varios ajustes sobre la tendencia, mientras el modelo ajusta el efecto estacional a diferentes períodos de tiempo.
- El método no requiere de períodos de tiempo regularmente espaciados; además, tampoco requiere interpolar datos perdidos.
- El ajuste del modelo a los datos es muy rápido.
- Los parámetros del modelo son fácilmente interpretables y esto permite que se puedan modificar para evaluar distintos supuestos.

Para el componente tendencial se utilizan dos modelos, uno de crecimiento con saturación y otro de crecimiento lineal por partes; siendo el primero muy común en modelaciones de ecosistemas y crecimiento demográfico. En aras de que el modelo permita capturar los efectos de estacionalidad para diferentes períodos de tiempo, se utilizan funciones periódicas del tiempo para lograr este objetivo, puntualmente se hace uso de las series de fuerier. Con el propósito de capturar el efecto de los días festivos, el método implementa una lista de los días festivos de cada país y captura, con la información histórica, el factor de cambio para estos días puntuales, tomando así, esta información como una modelación independiente; además, también se tiene en cuenta el comportamiento de algunos días alrededor del día festivo, pues los patrones podrían verse afectado en algunos días alrededor del festivo.

6.5. Método de componentes demográficas

El método de componentes demográficas, utilizado por el DANE, se expone en este trabajo a manera de referencia mas no se implementará en aras de obtener estimaciones, pues, se toman como referencia y punto de comparación los resultados obtenidos por el DANE. De esta manera, la exposición de esta metodología se sustenta directamente en los documentos metodológicos que presenta el DANE en su publicación de proyecciones poblacionales [7], para más detalle se recomienda su revisión.

La base del presente método es la definición de las componentes que determinan el cambio demográfico. En ese sentido, los tres componentes demográficos son: i) fecundidad, que proporciona la información para los nacimientos en el periodo de análisis; ii) mortalidad, que permite estimar la población que fallecería en el período de análisis; iii) migración, que da cuenta de las personas que se van a vivir fuera del área de referencia (emigrantes), o las personas que llegan al área de referencia (inmigrantes), durante el período de análisis.

$$N^{t+a} = N^t + B^{t,t+a} - D^{t,t+a} + M^{t,t+a} \tag{14}$$

donde:

N^t : población en el instante t

$B^{t,t+a}$: nacimientos ocurridos en el periodo intercensal

$D^{t,t+a}$: defunciones ocurridas en el periodo intercensal

$M^{t,t+a}$: migrantes netos en el período intercensal.

La Tasa Global de Fecundidad (TGF) es la estimación principal para proyectar los nacimientos. El DANE utiliza la ecuación 15 para este cálculo. Esta ecuación parte de proyectar el cambio de cada departamento con base en su cambio relativo frente al cambio nacional; esto se realiza para los diferentes quinquenios y edades. Es importante mencionar que el parámetro constante se obtiene de la asintota inferior de la estimación nacional de la TGF.

$$TGF^{t,dpt} = 1{,}4 - (1{,}4 - TGF^{\text{base, dpt}}) \left(\frac{1{,}4 - TGF^{\text{nacional}}}{1{,}4 - TGF^{\text{nacional}-1}} \right) \tag{15}$$

La mortalidad se estima a partir del registro de Estadísticas Vitales (EEVV), por sexo y edad para cada departamento. Además del insumo de la EEVV se utiliza el módulo de fallecidos en el CNPV 2018 para corregir las estimaciones además de algunos métodos de suavizado para ajustar los resultados, de las diferentes correcciones en esta estimación, finalmente se utiliza el resultado que genere la mínima suma de errores al cuadrado.

Para la estimación de los migrantes netos en el período intercensal se elabora el Índice Sintético de Migración, el cual tiene como principales insumos los módulos de migración de los cuatro censos mencionados; puntualmente se analiza la información de migración departamental en el mediano y largo plazo, información que se captura mediante la pregunta de lugar de residencia hace 5 años y hace un año, respectivamente. Estos datos en combinación con la probabilidad de supervivencia, permiten obtener los flujos migratorios en los departamentos de la nación.

Frente a la metodología seguida por el DANE es relevante detallar la forma en que se logra desagregar los resultados en distintas áreas geográficas, pasando de la nación a los departamentos y municipios. El método se aplica inicialmente para la nación, partiendo de información representativa en grandes áreas geográficas que cuentan con series de tiempo robustas en la captura de la información requerida, puntualmente de los departamentos y Bogotá. Posteriormente, se extrapola la participación poblacional de los municipios en cada departamento por edad y sexo, con base en funciones logísticas que se formulan a partir de los cortes censales de 1985, 1993, 2005 y 2018; con esto, a partir de la previa estimación departamental, se obtienen las estimaciones para cada uno de los municipios.

En este método se debe hacer énfasis en dos aspectos. El primero, reconocido por el DANE, es que en Colombia el método de componentes demográficas no se recomienda para la proyección poblacional a nivel municipal, especialmente por la carencia de información para estas áreas; el segundo, derivado del punto anterior, es que las estimaciones que se presentan para cada municipio no parte la información propia de los municipios, sino de su participación en los departamentos, lo cual genera que el error se

acumule, debido a que se agrega el error obtenido por las estimaciones propias de los departamentos y el de las estimaciones de la participación de cada municipio en el departamento.

7. Metodología

La metodología a desarrollar se divide en tres pasos: i) identificar la información existente y disponible públicamente para el municipio de Jamundí, lo cual permite identificar los datos que se utilizan para proyectar la población en los próximos años; ii) procesar la información que se utiliza para proyectar la población; iii) utilizar la información identificada y procesada en los diferentes modelos que se utilizaran para contrastar los resultados obtenidos con los del DANE; las estimaciones obtenidas en este último paso se presentan en el apartado de resultados y los códigos de programación utilizados para los diferentes modelos pueden ser consultados en el Anexo 1.

A continuación, se presenta el análisis de la información existente. La información que se muestra de las proyecciones y retroproyecciones poblacionales del DANE se obtuvieron del sitio web oficial de la entidad[5]; en esta se hace énfasis en las proyecciones con base en los censos de los años 2005 y 2018.

Por otro lado, la información correspondiente a la venta de vivienda en el municipio de Jamundí se obtiene de los reportes de venta de vivienda para Cali y su área metropolitana que realiza periodicamente la Cámara Regional de la Construcción - CAMACOL Valle -, quienes generan esta información en el marco de un estudio de mercado del sector donde se analiza la oferta y venta de vivienda en la zona[6].

Finalmente, los datos de suscriptores al servicio de energía se obtienen del Sistema Único de Información (SUI), puntualmente de la herramienta O3[7] en donde se pueden consultar diversas variables del servicio de energía, puntualmente el número de suscriptores, así como desagregar la información a nivel de áreas geográficas, divisiones político-administrativas, estratos, entre otras. La razón por la que se decide utilizar este servicio público frente a otros es la cobertura, pues, los censos del 2005 y 2018 reportan respectivamente una cobertura del 96.9 % y 98.6 %, siendo esta la más alta de todos los servicios públicos en ambos periodos.[8]

[5] Sitio web oficial de las proyecciones poblacionales: https://www.dane.gov.co/index.php/estadisticas-por-tema/demografia-y-poblacion/proyecciones-de-poblacion

[6] Estudios económicos de CAMACOL Valle: https://camacolvalle.org.co/estudios-economicos/

[7] Energía eléctrica en el portal del SUI: http://sui.superservicios.gov.co/Reportes-del-Sector/Energia

[8] Reporte del DANE, tomado de:https://www.dane.gov.co/files/investigaciones/planes-desarrollo-territorial/100320-Info-Alcaldias-Candelaria-Yumbo-Jamund-Palmira.pdf

7.1. Análisis de la información existente

Censos, proyecciones poblacionales del DANE y suscriptores de servicios públicos

Acorde al DANE [9], por lo antecedentes de censos realizados en Colombia, a la fecha se han realizado 10 de estas operaciones estadísticas; sin embargo, las más recientes datan del 2005 y 2018, siendo estas las de mejor robustez metodológica.

Como se puede apreciar, existen diferencias destacables entre las tres series de tiempo presentadas en la Figura 1, incluso en el comportamiento de la serie, pues, pese a las discrepancias, las proyecciones con base en los censos 2005 y 2018 conservan un comportamiento lineal, mientras que, la que se basa en los suscriptores de servicios públicos recuerda a una función logística.

Por otro lado, se aprecia que en el 2018 hay un momento de bastante cercanía entre la proyección con base en el CNPV 2018 y los suscriptores de servicios públicos. Esto se debe a que la metodología del DANE, ajusta sus proyecciones a la información que ayuda a tener un estimado de la población que habita actualmente en el territorio, tal como los suscriptores de servicios públicos o la afiliación al sistema de salud; pese a esto, este ajuste se presenta puntualmente para ese periodo de tiempo, pero no se evidencia en el resto de la serie.

A continuación, se presentan las proyecciones poblacionales con base en el censo 2005, así como las retroproyecciones y proyecciones poblacionales con base en el censo 2018, al igual que los suscriptores a servicios públicos, para el municipio de Jamundí en el periodo 2005-2020.

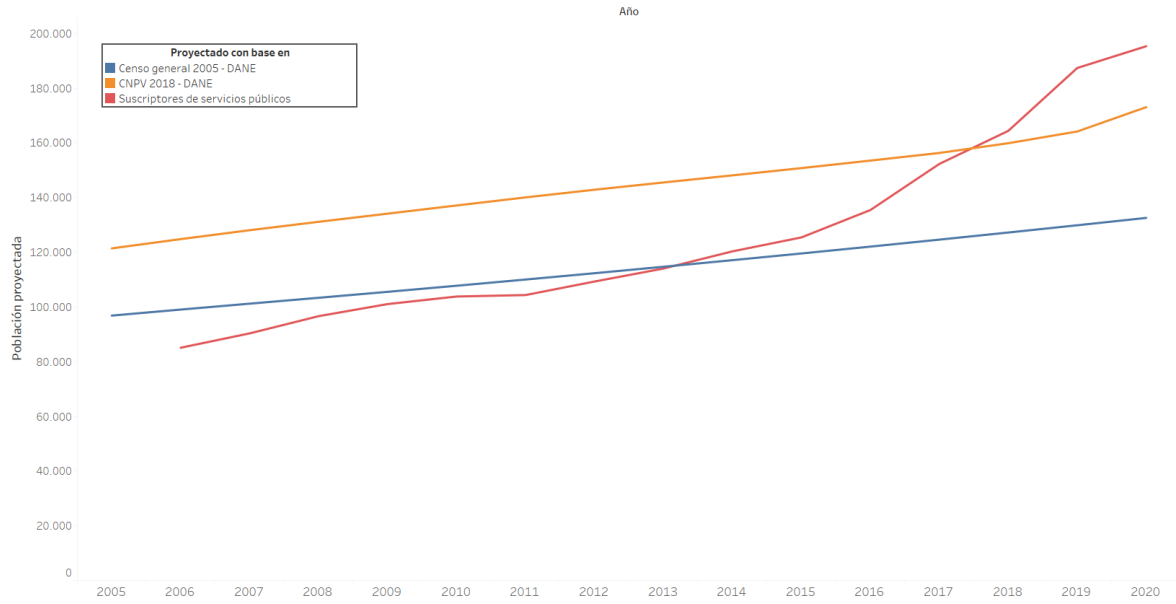


Figura 1: Proyecciones poblacionales con base en el censo general del 2005, el CNPV 2018 y los suscriptores de servicios públicos

[9] Obtenido de la ficha metodológica del CNPV 2018: https://microdatos.dane.gov.co/index.php/catalog/643/related-materials

Una de las principales razones por las que se escogió a Jamundí como caso a evaluar en el presente estudio, se debe al precipitado crecimiento poblacional que exhibe el municipio aún en las proyecciones poblacionales presentadas por el DANE y que se resalta aún más al revisar los suscriptores de servicios públicos. Mediante la interpretación gráfica de la Figura 2 se concluye que el crecimiento en el periodo intercensal de Jamundí es atípico frente a los municipios del Valle del Cauca; además, en comparación con todos los municipios del país con más de 100 mil habitantes, se encuentra por encima del 50 % de los datos, estando entre los 20 primeros municipios con mayor crecimiento intercensal.

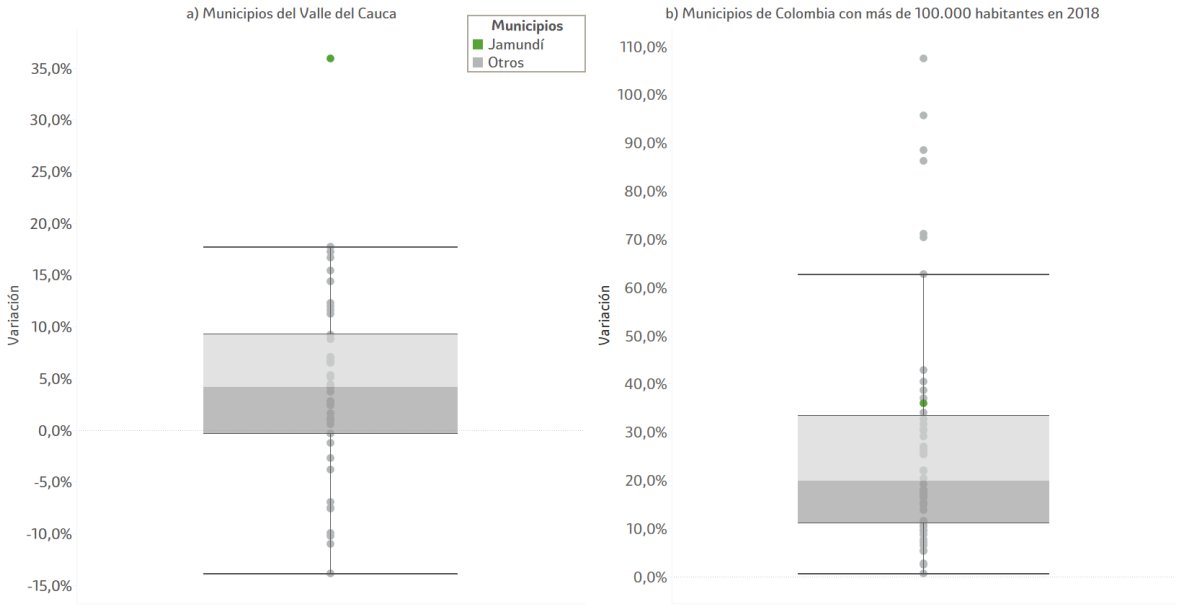


Figura 2: Variación poblacional en el periodo intercensal

Viviendas y suscriptores de servicios públicos

Una fuente pública que puede utilizarse para aproximar el nivel poblacional son las viviendas, así como los suscriptores a servicios públicos. Evidentemente, estas dos variables guardan una estrecha relación, pues, los servicios públicos son esenciales para la habitabilidad de cualquier vivienda, además, como ya se mostró la cobertura de servicios como energía supera el 95 %. A continuación, en la Tabla 1 se presentan los nuevos suscriptores a servicios públicos, entendidos como la diferencia entre los suscriptores de un año y el siguiente, y las viviendas vendidas, entre 2008 y 2021.

Como se puede apreciar en la Tabla 1 existe una estrecha relación entre ambas variables, relación que ya se ha explicado. Pese a esto, se debe mencionar que en general suelen ser más las viviendas vendidas que los nuevos suscriptores a energía, esto se puede explicar debido a que en muchas ocasiones las viviendas se venden y aún no cuentan con el servicio conectado, incluso muchas veces ni siquiera se ha construido aún al momento de la venta, bajo esta figura funcionan, por ejemplo, las casas modelo. En ese sentido, la información que mejor puede reflejar el comportamiento presente de la población es la de suscriptores a servicios públicos, además, esta información es de obligatorio reporte,

Año	Nuevos suscriptores	Viviendas vendidas
2008	1.881	943
2009	1.440	825
2010	1.051	760
2011	502	782
2012	1.775	1.188
2013	1.797	1.405
2014	2.305	3.411
2015	2.053	6.354
2016	3.587	7.388
2017	5.809	5.555
2018	4.553	5.967
2019	7.905	5.798
2020	3.379	7.109
2021	8.956	9.559
Total	46.992	57.044

Cuadro 1: Nuevos suscriptores de servicios públicos y viviendas vendidas por año

su reporte se realiza de manera mensual por lo que se cuenta con más momentos de tiempo en el análisis y también permite diferentes desagregaciones adicional al municipio.

7.2. Procesamiento de la información que se utiliza para proyectar la población

El procesamiento de la información se suscriptores a servicios públicos consiste en descargar los archivos de la herramienta O3 del SUI, estos se desagregaron para el municipio de Jamundí por área (urbano, rural), estratos, en el periodo 2005-2021, utilizando solamente información de suscriptores residenciales, dado que son los que coinciden con habitantes. La información original para la zona urbana se presenta en la Figura 3.

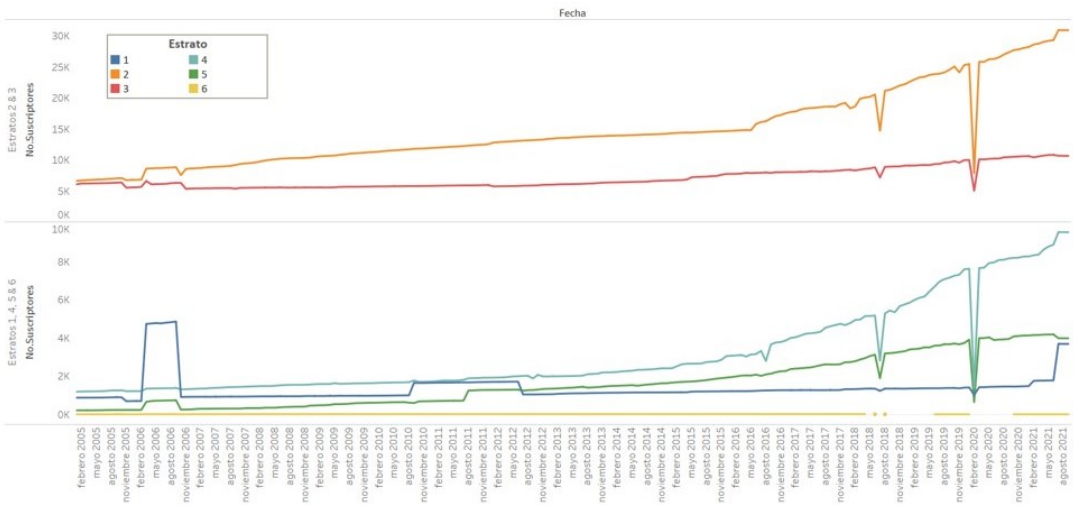


Figura 3: Suscriptores urbanos por estrato - serie original (2005 - 2021)

Tal como muestra la Figura 3, las series temporales tienen tendencia creciente, pese a esto, dado que esta información se reporta de manera mensual, se evidencian algunos cambios estructurales en su comportamiento, debidos probablemente a errores en la digitación y reporte, por lo que esta se corrige mediante medias móviles, donde la longitud del promedio se establece dependiendo del número de registros consecutivos que no siguen la tendencia general que presenta la serie histórica. En la Figura 4 se puede apreciar la serie corregida mediante medias móviles.

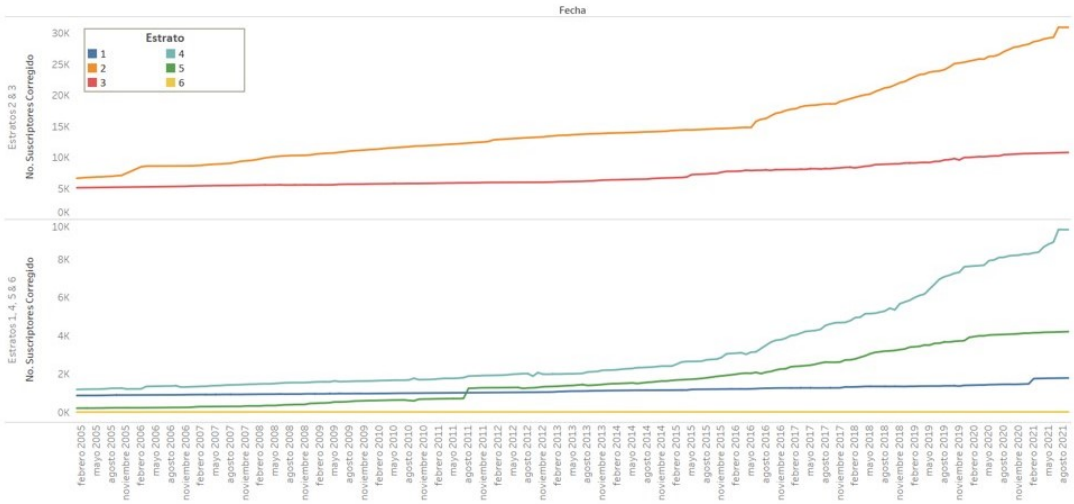


Figura 4: Suscriptores urbanos por estrato - serie ajustada (2005 - 2021)

Por otro lado, en las Figuras 5 y 6 se presentan respectivamente los datos originales y ajustados mediante medias móviles para la zona rural. Los datos se desagregaron en urbano y rural debido a las diferentes dinámicas poblacionales que se presentan entre ambas áreas, pese a esto, se evidencia igualmente una tendencia creciente en el número de suscriptores de la zona rural.

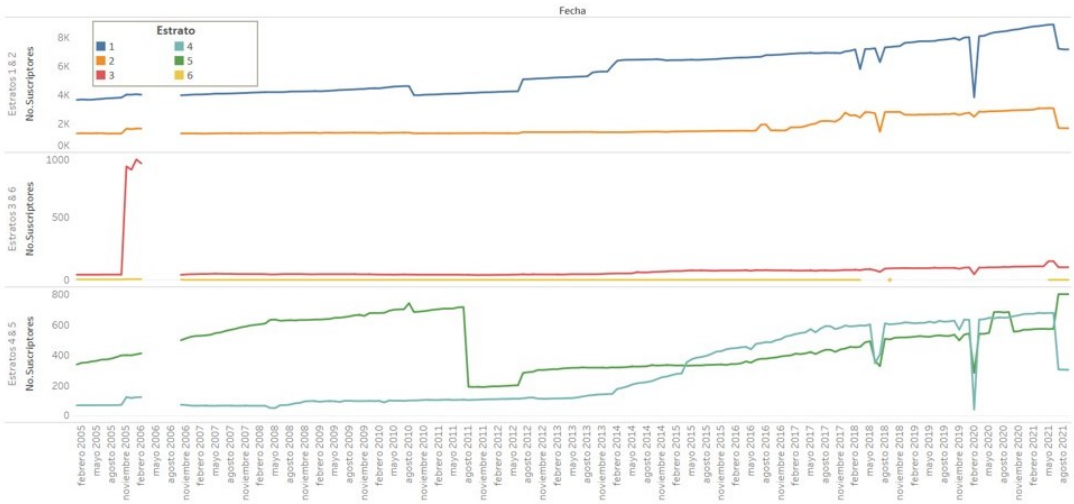


Figura 5: Suscriptores rurales por estrato - serie original (2005 - 2021)

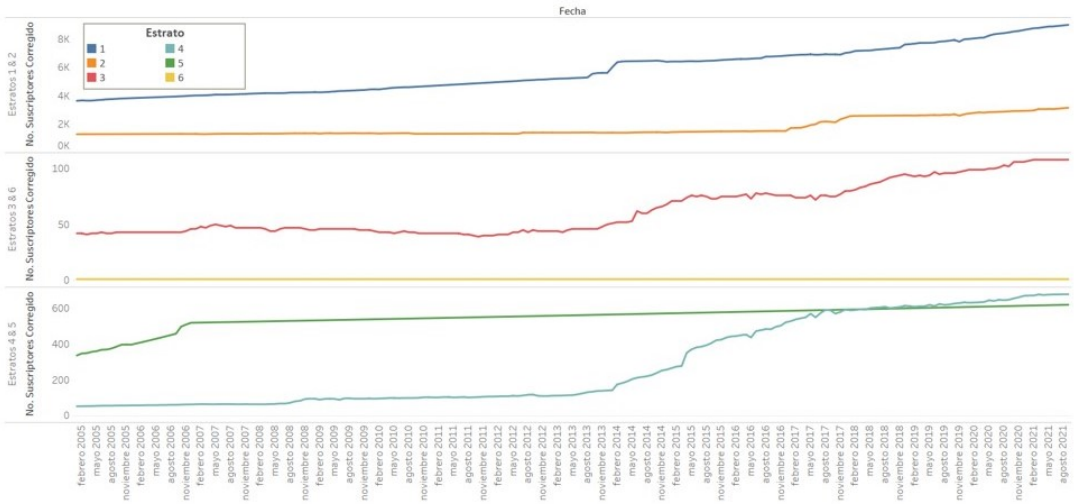


Figura 6: Suscriptores rurales por estrato - serie ajustada (2005 - 2021)

8. Resultados

A continuación se presentan los resultados obtenidos por los modelos de suavizado exponencial, ARIMA y Prophet, respectivamente. Al finalizar la sección se muestra un comparativo general entre los resultados de los modelos implementados y los publicados por el DANE.

8.1. Suavizado exponencial

Para suavizado exponencial se implementó el modelo de Holt aditivo con tendencia amortiguada, acorde a la estrucutra que se visualiza en los datos, presentada en la sección anterior. Los resultados se estimaron para cada estrato en la zona urbana y rural, los resultados de esta estimación se presentan en las Figuras 7 y 9. Estos resultados proyectan el número de suscriptores de energía hasta el 2035, con un error absoluto promedio inferior a 1 en los suscriptores de estratos 3, 5 y 6, mientras que en los estratos 1, 2 y 4, el error abosluto promedio no supera los 25 suscriptores.

Con los resultados obtenidos, se aplica el factor de personas por hogar estimado por estrato y área mediante los microdatos del CNPV 2018 de dos maneras: i) manteniendo constante este factor a lo largo de los años; ii) variando este factor con base en la disminución de personas por hogar proyectada por el DANE. Los resultados comparativos de estas dos estimaciones de población y la del DANE se presentan en la Figura 9, donde solo se presenta la desagregación entre urbano y rural.

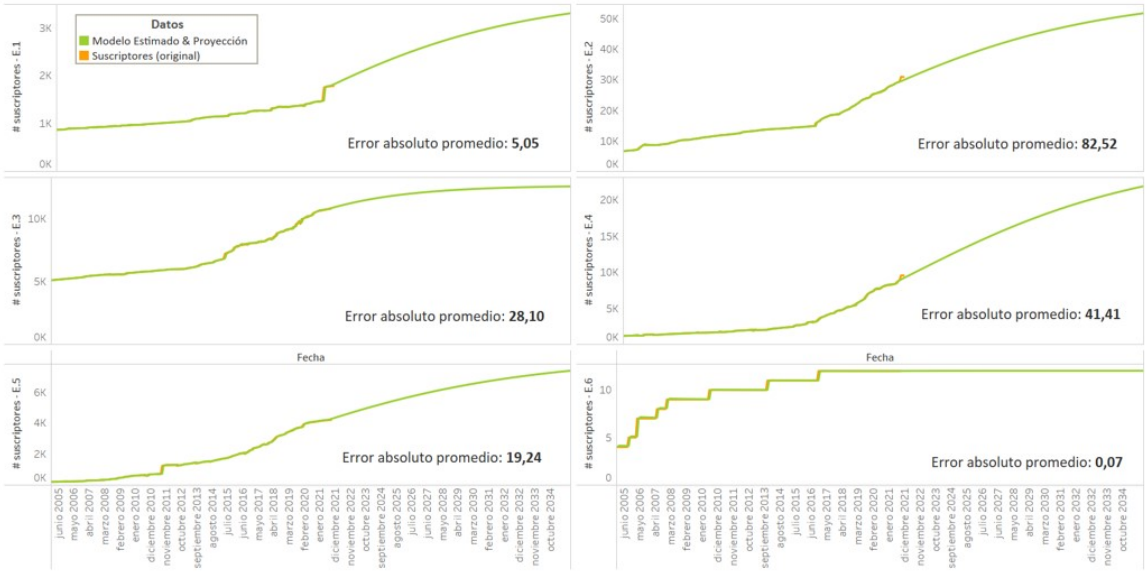


Figura 7: Suscriptores urbanos por estrato - proyección a 2035 - Modelo de suavizado exponencial

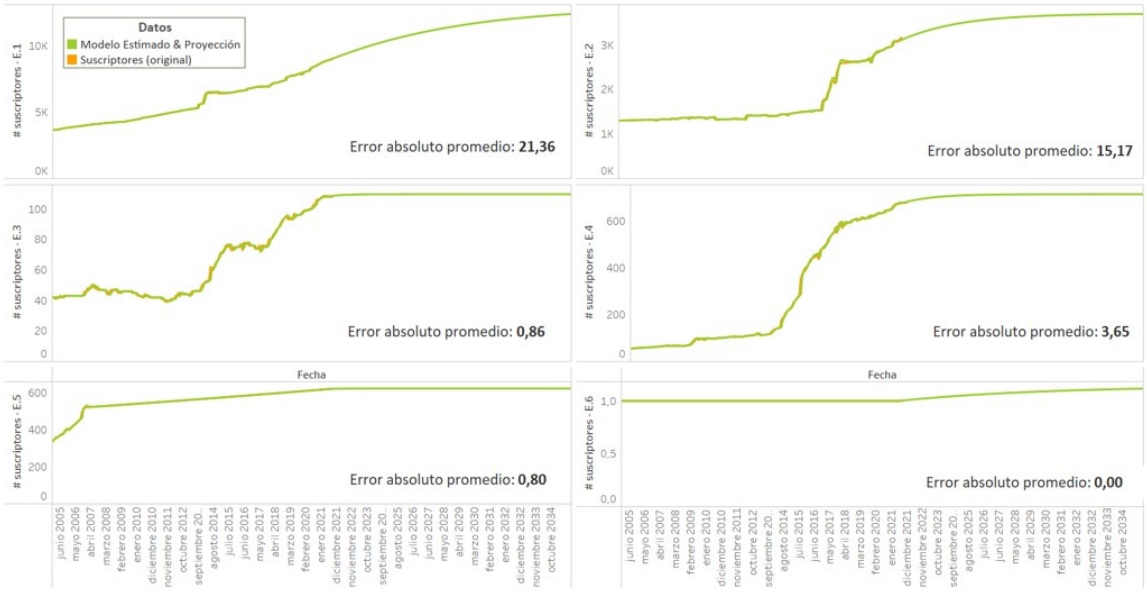


Figura 8: Suscriptores rurales por estrato - proyección a 2035 - Modelo de suavizado exponencial

8.2. ARIMA

Para la estimación de modelos ARIMA se utilizó la función de Python que ajusta automáticamente el mejor modelo a los datos brindados, aclarando la periodicidad mensual y dejando cabida a un posible factor estacional, sin embargo, los modelos ajustados no estimaron algún componente estacional, sino que solo tuvieron componentes autorregresivos y de media móvil. Los resultados estimados por estrato para las zonas urbana y rural se presentan respectivamente en las Figuras 11 y 11.

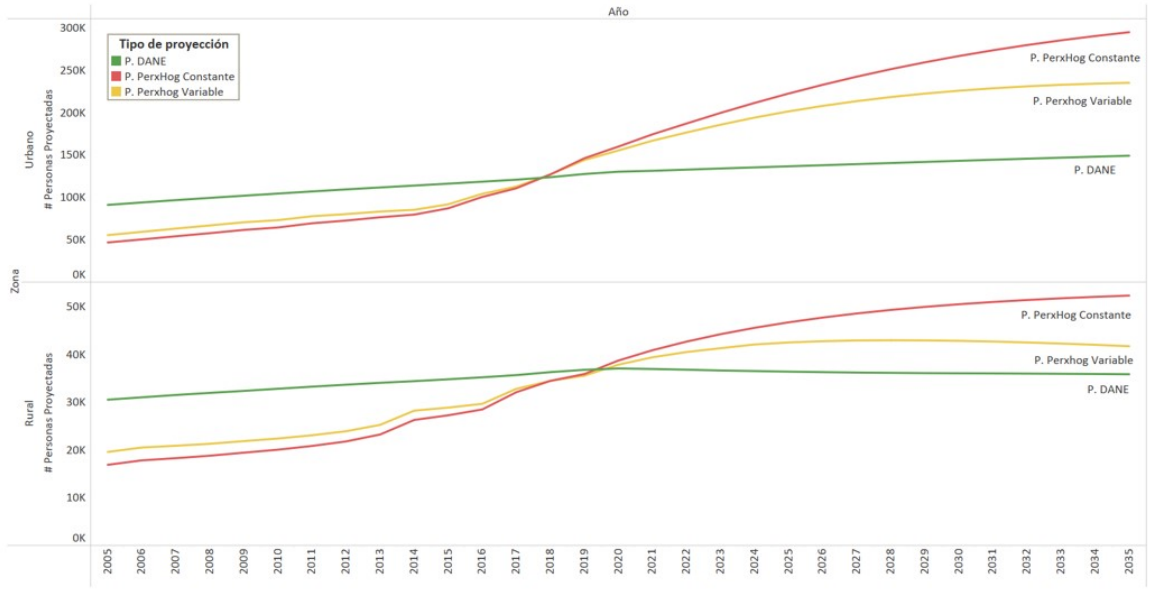
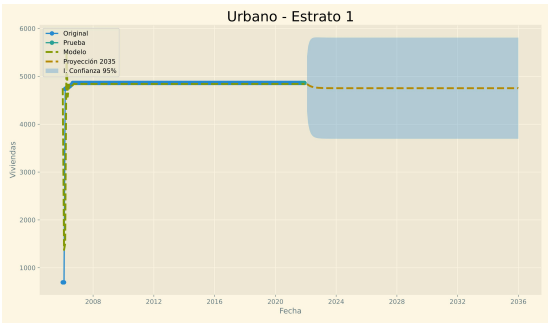
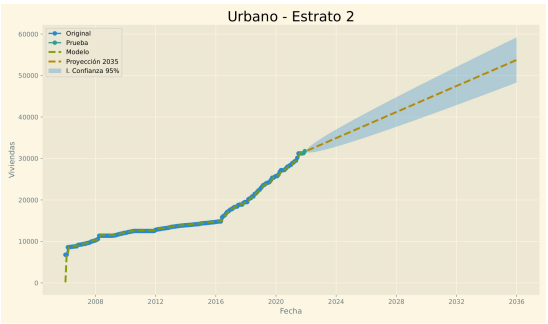


Figura 9: Suscriptores urbanos y rurales - proyección a 2035 - Modelo de suavizado exponencial

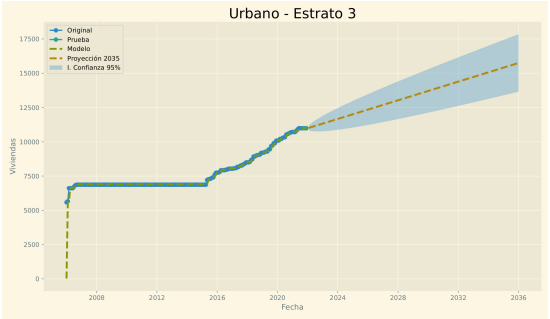
Como se puede apreciar en las figuras, una de las ventajas del modelo ARIMA frente al de suavizado exponencial es que permite estimar intervalos de confianza para las proyecciones, los cuales también se aprecian en las Figuras, destacándose cómo estos se amplían a medida que el horizonte de tiempo es mayor.



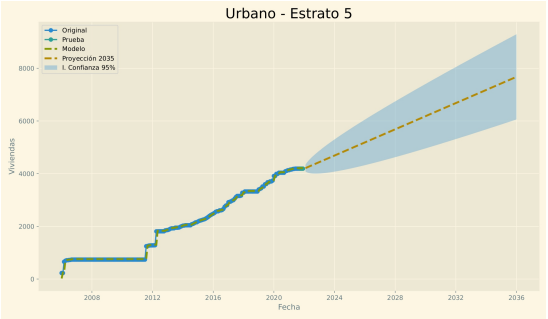
(a) Estrato 1



(b) Estrato 2



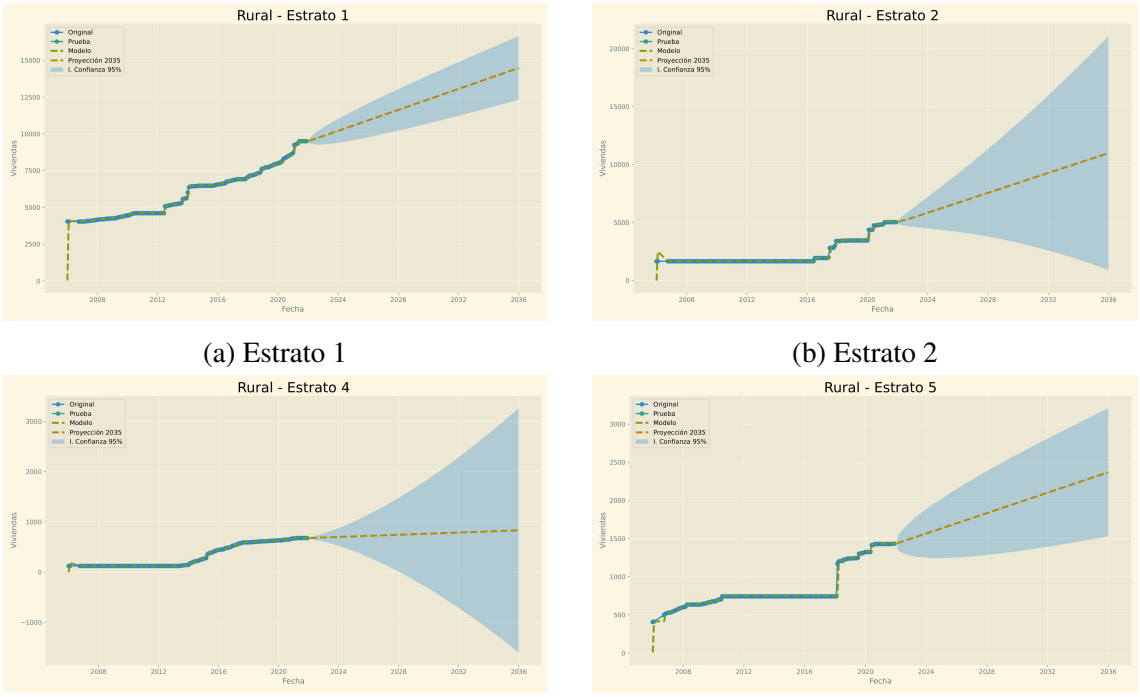
(c) Estrato 3



(d) Estrato 5

Figura 10: Proyección de suscriptores urbanos por estrato - modelo ARIMA

Por otro lado, como análisis exploratorio, también se utiliza el modelo ARIMA para



(a) Estrato 1

(b) Estrato 2

(c) Estrato 4

(d) Estrato 5

Figura 11: Proyección de suscriptores rurales por estrato - modelo ARIMA

proyectar la venta de vivienda en el municipio (ver Figura 13); en este caso el modelo sí utiliza componente estacional como se aprecia en las proyecciones. El estudio de los residuales del modelo se presenta en la Figura 13, donde se aprecia que estos alcanzan un comportamiento aparente al ruido blanco a partir del primer rezago, lo cual es muestra del buen ajuste del modelo.

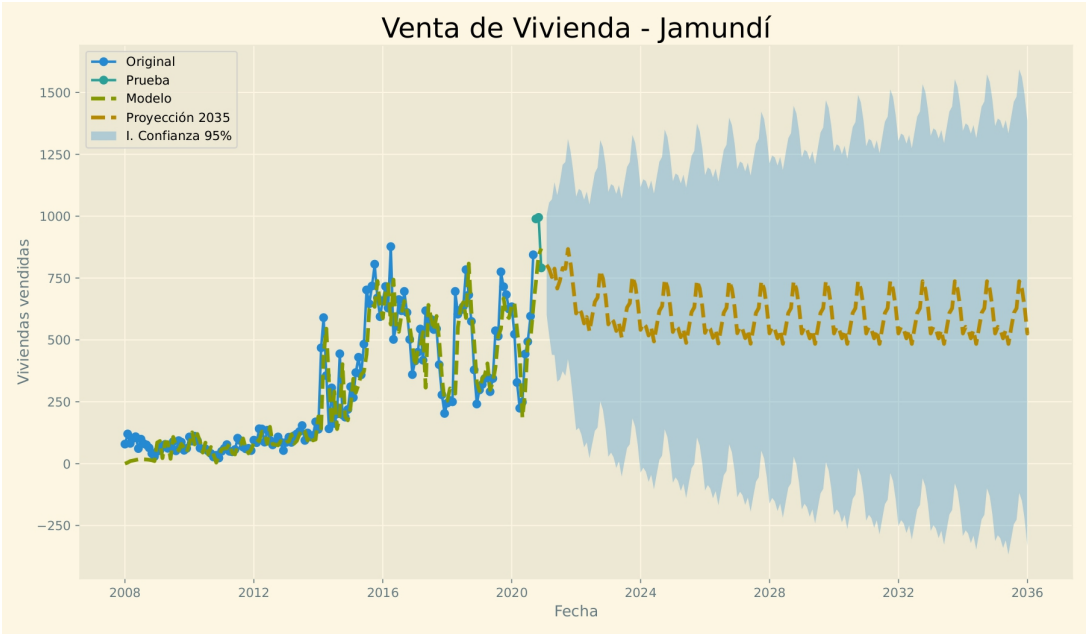


Figura 12: Proyección de vivendas vendidas - modelo ARIMA

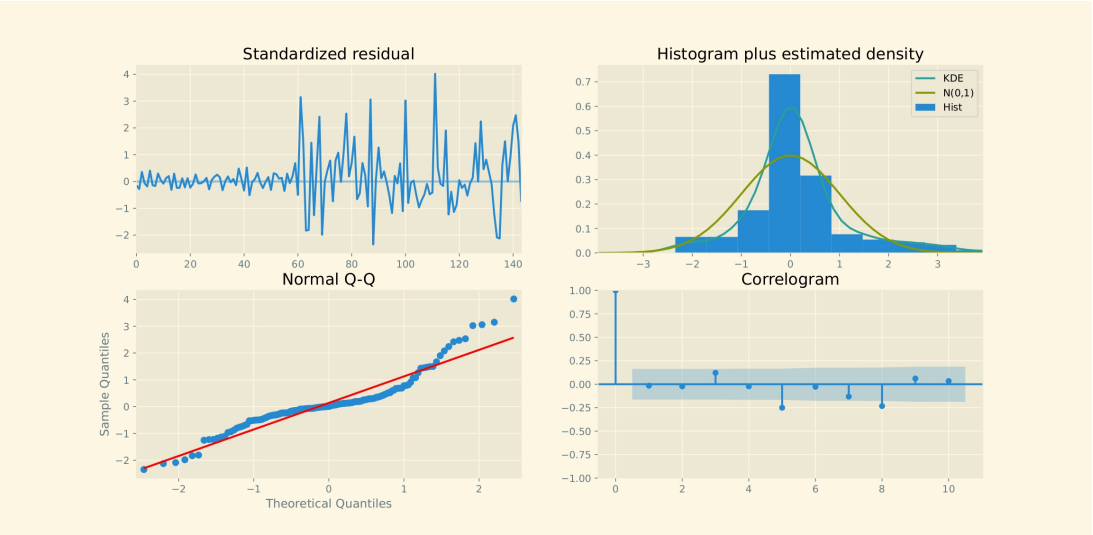


Figura 13: Análisis de residuales - modelo ARIMA

8.3. Prophet

A continuación, en la Figura 14 se presenta el pronóstico del total de suscriptores realizado mediante el algoritmo Prophet. Dada la naturaleza de los datos utilizados, el método solo estima el componente tendencial, mas no hay una estimación de los componentes estacional y de días festivos.

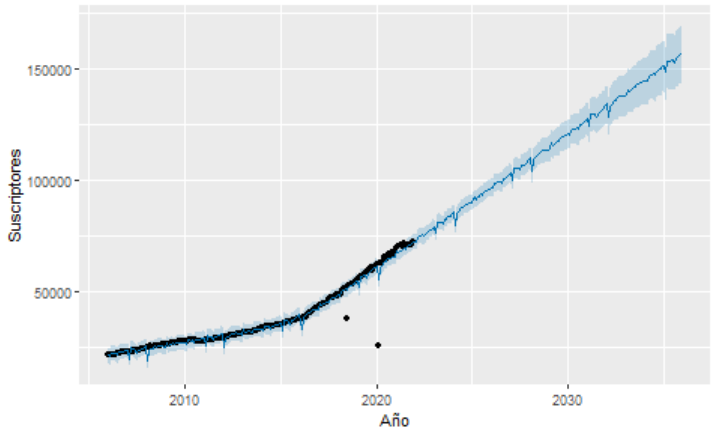


Figura 14: Suscriptores - proyección a 2035 - Algoritmo Prophet

8.4. Comparación de resultados obtenidos y estimaciones del DANE

Para finalizar esta sección, la Tabla 2 muestra los diferentes resultados obtenidos en las proyecciones de suscriptores realizadas por los diversos métodos, agregadas por cada año desde el 2021 hasta el 2035, así como las estimaciones del DANE. Para obtener las personas proyectadas se utilizó el factor de personas por hogar aplicado a los suscriptores estimados en el respectivo año.

DANE			Métodos tradicionales				Métodos modernos	
			Suavizado exponencial		ARIMA		Prophet	
Año	P. x H.	Personas	Suscriptores	Personas	Suscriptores	Personas	Suscriptores	Personas
2.021	3,21	168.079	70.841	227.156	79.377	254.528	72.227	231.600
2.022	3,16	169.280	75.764	239.336	82.667	261.144	78.813	248.969
2.023	3,11	170.492	80.391	250.224	87.061	270.986	85.395	265.798
2.024	3,07	171.644	84.717	260.295	91.460	281.014	89.823	275.984
2.025	3,03	172.821	88.740	268.748	95.859	290.307	96.400	291.946
2.026	2,99	173.997	92.464	276.069	100.258	299.338	102.986	307.485
2.027	2,94	175.209	95.896	282.328	104.657	308.119	109.568	322.578
2.028	2,90	176.434	99.047	287.596	109.055	316.658	113.996	331.004
2.029	2,86	177.672	101.927	291.978	113.454	324.998	120.573	345.390
2.030	2,83	178.929	104.553	295.585	117.853	333.187	127.160	359.497
2.031	2,79	180.160	106.939	298.473	122.253	341.214	133.741	373.279
2.032	2,76	181.376	109.101	300.711	126.652	349.085	138.169	380.831
2.033	2,72	182.548	111.055	302.386	131.051	356.831	144.746	394.122
2.034	2,69	183.675	112.817	303.543	135.450	364.438	151.333	407.173
2.035	2,66	184.786	114.403	304.236	139.849	371.903	157.914	419.946

Cuadro 2: Comparación entre resultados del DANE y resultados por métodos de series de tiempo

9. Conclusiones y recomendaciones

En este estudio se realizaron proyecciones poblacionales para el municipio de Jamundí hasta el año 2035, partiendo de información histórica de suscriptores a servicios públicos; se encontraron diferencias significativas frente a las proyecciones poblacionales realizadas por el DANE con base en los censos poblacionales y registros administrativos de las administraciones municipales y departamentales. Las diferencias se ven explicadas principalmente por la fuente de información utilizada en las estimaciones propias y los diversos métodos alternativos al del DANE.

Los resultados presentados dejan en evidencia que existen variables alternativas a las que utiliza el método de componentes demográficas para proyectar población, como lo son los suscriptores de servicios públicos y la venta de viviendas, teniendo predilección para este trabajo por la primera, debido a su periodicidad, disponibilidad y menor rezago en el tiempo en guardar relación con niveles poblacionales.

Las proyecciones poblacionales al año 2035 llegan a presentar diferencias de hasta más del doble frente a la que proyecta el DANE para el mismo año. Los modelos utilizados consisten en métodos tradicionales de series de tiempo, como los modelos de suavizado exponencial y ARIMA, así como de un método desarrollado recientemente, Prophet.

De los métodos utilizados en este trabajo se debe resaltar la practicidad en su implementación. Esta se caracteriza por dos bondades principalmente: i) la existencia de librerías en software de libre acceso para su ejecución; ii) la escaza exigencia en cantidad de información requerida. De la información utilizada también se aprecian tres características propicias para su utilización: i) acceso público sin ningún tipo de registro o restricción por usuario; ii) desagregación para áreas subnacionales (departamentos y municipios), así como también para estratos y áreas geográficas (urbano y rural); iii) frecuencia mensual, lo cual permite tener un histórico de información con un volumen de datos aceptable para la implementación de diferentes métodos.

Pese a las facultades a favor que se mencionaron anteriormente, se debe resaltar que el método utilizado por el DANE permite generar resultados que no se pueden alcanzar con la metodología desarrollada en esta investigación. i) como su nombre lo indica, las componentes demográficas estiman las componentes que determinan el cambio demográfico en una población, por tanto, para obtner la estimación poblacional se debe calcular previamente la natalidad, mortalidad y migración, información que puede ser de relevancia para la toma de decisiones; ii) permite estimaciones a nivel de rangos de edad y sexo, que también puede ser de interés para la concepción de políticas públicas o proyectos de inversión específicos; iii) la estimación se realiza por niveles, primero nacional, luego departamental y finalmente municipal, esto permite mantener supeditada la información de áreas más pequeñas a las estimaciones previas en áreas mayores, lo cual permite guardar congruencia en las estimaciones.

Debido al contraste entre los métodos dado por sus bondades y limitaciones, se recomienda que las estimaciones realizadas por el DANE, que utilizan los territorios para la toma de decisiones y la nación para transferir recursos, entre otras funciones, sean contrastadas por métodos alternativos como los presentados en esta investigación; espcialmente en las áreas geográficas más pequeñas (municipios), pues, como lo reconoce el propio DANE y lo muestra la evidencia nacional e internacional, la metodología implementada actualmente no es adecuada para los municipios. Este contraste y evaluación continua podría permitir que la toma de decisiones de cara a la política pública sea más acertada a los territorios, puesto que la población es el centro de toda decisión pública, además de las eficiencias que se podrían proveer igualmente al sector privado.

Bibliografía

Libros

[3] George E. Box et al. «Time series analysis: forecasting and control». En: 2016. ISBN: 978-1-118-67502-1.

[11] Rob Hyndman y George Athanasopoulos. «Forecasting: principles and practice». En: 2018. DOI: 10.1007/978-0-9875071-1--2.

[12] Rob Hyndman et al. «Forecasting with exponential smoothing. The state space approach». En: ene. de 2008. DOI: 10.1007/978-3-540-71918-2.

[14] David Swanson y Jacob Siegel. «The Methods and Materials of Demography, 2nd Edition». En: ene. de 2004. ISBN: 0-12-641955-8.

Artículos

[4] Robert G. Brown. «Statistical forecasting inventory control». En: (1959).

[10] Charles C. Holt. «Forecasting seasonals and trends by exponentially weighted averages (O.N.R. Memorandum No. 52).» En: (1957).

[13] Adrian E. Raftery y Hana Ševčíková. «Probabilistic population forecasting: Short to very long-term». En: *International Journal of Forecasting* (2021). ISSN: 0169-2070. DOI: https://doi.org/10.1016/j.ijforecast.2021.09.001. URL: https://www.sciencedirect.com/science/article/pii/S0169207021001394.

[15] Sean Taylor y Benjamin Letham. «Forecasting at scale». En: (sep. de 2017). DOI: 10.7287/peerj.preprints.3190v2.

[17] Tom Wilson, Irina Grossman y Jeromey Temple. «Evaluation of the best M4 competition methods for small area population forecasting». En: *International Journal of Forecasting* (2021). ISSN: 0169-2070. DOI: https://doi.org/10.1016/j.ijforecast.2021.09.005. URL: https://www.sciencedirect.com/science/article/pii/S0169207021001497.

[18] Peter R. Winters. «Forecasting sales by exponentially weighted moving averages». En: (1960).

Guías

[2] ONU. Departamento de Asuntos Económicos y Sociales. *Manual III: métodos para preparar proyecciones de población por sexo y edad*. 1956.

[7] Dirección de censos y demografía. *Metodología General de Proyecciones de Población y Estudios Demográficos (PPED)*. Mayo de 2021. URL: https://www.dane.gov.co/index.php/estadisticas-por-tema/demografia-y-poblacion/proyecciones-de-poblacion.

[8] Dirección de censos y demografía. *Metodología Proyecciones de Población y Estudios Demográficos (PPED)*. Mayo de 2009. URL: https://www.dane.gov.co/files/investigaciones/fichas/Proyecciones_poblacion.pdf.

[9] *Guía No. 1: Elementos poblacionales básicos para la planeación*. Mar. de 2006. ISBN: 958-97393-7-7.

Reportes técnicos

[5] Villacis Byron. *Proyecciones de población - Un estudio comparativo*. Banco Interamericano de Desarrollo, ago. de 2019.

[6] Dirección de censos y demografía. *Documento Metodológico de elaboración de las proyecciones de población de Bogotá, D.C., a nivel de localidad hasta el año 2035 y de Unidad de Planeamiento Zonal – UPZ hasta el año 2024*. Departamento Administrativo Nacional de Estadística (DANE), dic. de 2020. URL: https://www.dane.gov.co/files/censo2018/proyecciones-de-poblacion/Bogota/proyecciones-bogota-metodologia-desagreacion-loc-2018-2035-UPZ-2018-2024.pdf.

[16] Naciones Unidas. *Perspectivas de la población mundial 2019: metodología de las Naciones Unidas para las estimaciones y proyecciones de población*. N° 132 (LC/TS.2020/95). Santiago: Comisión Económica para América Latina y el Caribe (CEPAL), 2020.

10. Anexo 1

10.1. Código en Python para el método de suavizado exponencial

```
1
2  import pandas as pd
3  import matplotlib.pyplot as plt
4  from statsmodels.tsa.api import Holt
5
6  ## Ruta
7  path_BD = "C:/Users/camil/Desktop/Datos/BD Procesadas/CELSIA/Suscriptores SUI/"
8  path_saveImage = "C:/Users/camil/Desktop/Camilo/Trabajo/Alcaldía de Jamundí/POT/Di
9  ## Nombre del archivo
10 BD_nom = "Compilado Arreglado - 1985-2021"
11 ## Importamos la BD
12 BD = pd.read_excel(path_BD + BD_nom + ".xlsx")
13
14 Lista_Zona = ["Urbano", "Rural"] ## Lista de zona
15
16 fecha_pronostico = []
17
18 for j in range(2022,2036):
19         fecha_pronostico.append(str(j))
20
21 fecha_pronostico = pd.DatetimeIndex(fecha_pronostico)
22
23 for zona in Lista_Zona:
24     ## Filtramos para la zona
25     datos = BD.loc[BD['Zona'] == zona]
26     ## Creamos la marca temporal
27     fecha_index = pd.DatetimeIndex(datos.iloc[:,1])
28     ## Creamos la serie
29     serie_tiempo = pd.Series(datos['Población'].values,fecha_index.values)
30
31     ##Aproximamos con el método de Holt
32     aprox = Holt(serie_tiempo, damped_trend = True,
33                  exponential = "add" ,
34                  initialization_method = "estimated").fit(
35         smoothing_trend = 0.075,use_brute = True ,optimized = True)
36
37     Pronos = aprox.forecast(14).rename("Pronóstico")
38
39     plt.figure(figsize=(20, 9))
40     (line1,) = plt.plot(serie_tiempo, marker="*", color="black")
41     (line2,) = plt.plot(aprox.fittedvalues, marker = "*", color="blue")
42     (line3,) = plt.plot(Pronos, marker="s", color="green")
43     plt.legend([line1, line2, line3],
44                ["Serie original - Zona " + " " + zona, "Modelo",Pronos.name])
```

```
45
46      plt.savefig(path_saveImage + "Pronóstico - Zona " + zona + ".png")
47
48      BD_serie_pronóstico = pd.DataFrame(data = {'Año':fecha_pronostico,
49                                                 'Personas':Pronos})
50      BD_serie_modelo = pd.DataFrame(data = {'Año':fecha_index.values,
51                                             'Modelo':aprox.fittedvalues})
52      BD_serie_original = pd.DataFrame(data = {'Año':fecha_index.values,
53                                               'Personas':datos['Población'].values})
54
55      BD_final = BD_serie_original.merge(BD_serie_modelo,
56                                         left_on = "Año",
57                                         right_on = "Año", how = "inner")
58      BD_final = BD_final.append(BD_serie_pronóstico, ignore_index = True)
59
60      zona_list = []
61      for contador in range(0,len(BD_final)):
62          zona_list.append(zona)
63
64      BD_final["Zona"] = zona_list
65
66      BD_final.to_excel(path_BD + "BD_Pronostico - Zona " + zona + ".xlsx",
67                        index = False)
68
69  for j in Lista_Zona:
70          BD_pegar = pd.read_excel(path_BD + "BD_Pronostico - Zona " + j + ".xlsx")
71
72          if ((j == "Urbano")):
73              BD_consolidada = BD_pegar
74              next
75
76          BD_consolidada = BD_consolidada.append(BD_pegar, ignore_index = True)
77
78  BD_consolidada.to_excel(path_BD + "Proyección V4 Personas-2035 -AÑo&Zona.xlsx",
79                          index = False)
80
```

10.2. Código en Python para modelo ARIMA

[mathescape, linenos]python

```
1
2  ######### Librerias
3
4  # Operaciones Matemáticas
5  import math
6  import scipy as sp
7  import numpy as np
8
```

```
# Manejo de Datos
import pandas as pd
import json
import requests

# Graficar
import matplotlib as mpl
import matplotlib.pyplot as plt
import matplotlib.dates as mdates
#matplotlib inline
plt.style.use('Solarize_Light2')
from statsmodels.graphics.tsaplots import plot_acf
from statsmodels.graphics.tsaplots import plot_pacf
import seaborn as sns



# Machine Learning & Métodos estadísticos
import pmdarima as pm
import statsmodels.api as sm
from statsmodels.tsa.seasonal import seasonal_decompose
from statsmodels.tsa.stattools import adfuller

# Manejo de fechas
import datetime

# Métodos de evaluación
from sklearn.preprocessing import StandardScaler
from statsmodels.tsa.stattools import kpss
from sklearn.metrics import mean_absolute_error
from sklearn.metrics import mean_squared_error

# Función para graficar
def plot_train_test_pred(train,test,pred_df, forecast, conf_int, title):
    fig,ax = plt.subplots(figsize=(12,7))
    kws = dict(marker='o')

    ax.plot(train,label='Original',**kws)
    ax.plot(test,label='Prueba',**kws)
    ax.plot(pred_df['Viviendas'],label='Modelo',ls='--',linewidth=3)
    ax.plot(forecast['Viviendas'],
            label='Proyección 2035',ls='--',linewidth=3)

    ax.fill_between(x=conf_int.index,
                    y1=conf_int['Lim_inf'],y2=conf_int['Lim_sup'],
                    alpha=0.3, label = "I. Confianza 95%")
    ax.set_title(title, fontsize=22)
    plt.xlabel("Fecha")
```

```
    plt.ylabel("Viviendas")
    ax.legend(loc='upper left')
    fig.tight_layout()

    plt.show()
    plt.savefig(file_path + title + ".png")

############

## Declaramos la ruta y el nombre del archivo
file_path = "C:/Users/camil/Desktop/Camilo/Trabajo/Alcaldía de Jamundí/POT/Formulaci
file_name = "Suscriptores - Residenciales - 2006_2021.xlsx"

## Importamos la Base Original
BD_O = pd.read_excel(file_path + file_name)
BD_O = pd.DataFrame(BD_O)

## Obtenemos los valores únicos de cada vari. diferenciadora
Ubicaciones = BD_O["UBICACION"].unique()
Estratos = BD_O["ESTRATO"].unique()

## Nos quedamos solo con los residenciales y Estratos 1, 2, 3, 4, 5 & 6
Ubicaciones = [Ubicacion for Ubicacion in Ubicaciones if Ubicacion in ["Urbano",
                                                                       "Rural"]]
Estratos = [Estrato for Estrato in Estratos if Estrato in ["Estrato 1",
                                                           "Estrato 2",
                                                           "Estrato 3",
                                                           "Estrato 4",
                                                           "Estrato 5",
                                                           "Estrato 6"]]

## Creamos el vector de filtro para hacer cada una de las proyecciones
Vect_filt = []

for Ubicacion in Ubicaciones:
    for Estrato in Estratos:
        Vect_filt.append([Ubicacion, Estrato])

Vect_filt.remove(["Rural", "Estrato 6"])

BD_Proyectada_export = pd.DataFrame(columns = ['FECHA', 'Viviendas' ,
                                               'UBICACION', 'ESTRATO'])
for Filt in Vect_filt:

    BD_F = BD_O[(BD_O["UBICACION"] == Filt[0]) & (BD_O["ESTRATO"] == Filt[1])]
    BD_F.sort_values(by = ["UBICACION", "ESTRATO", "AÑO"],
                     ascending = True)
    BD_F["FECHA"]= BD_F["FECHA"].astype("datetime64")
```

```
105     BD_F = BD_F.set_index("FECHA")
106     BD_F = BD_F.sort_index()
107 
108     if (len(BD_F) == 0):
109         continue
110     else:
111 
112         for i in range(1,len(BD_F)-1):
113             valor1 = BD_F.iloc[i-1,4]
114             valor3 = BD_F.iloc[i+1,4]
115             if ((BD_F.iloc[i,4] == np.nan) or (BD_F.iloc[i,4] == np.inf) or (BD_F.
116                 BD_F.iloc[i,4] = np.mean([valor1,valor3])
117 
118         for i in range(1,len(BD_F)):
119             valor1 = BD_F.iloc[i-1,4]
120             if (BD_F.iloc[i,4] < valor1):
121                 BD_F.iloc[i,4] = valor1
122 
123 
124         BD_Train,BD_Test = BD_F[:-3], BD_F[-3:]
125         BD_Train = BD_Train[["Viviendas"]]
126         BD_Test = BD_Test[["Viviendas"]]
127         BD_Train.sort_index(inplace = True)
128         BD_Test.sort_index(inplace = True)
129 
130         ## Descomponemos la serie en tendencia, estacionalidad, ruido
131         Descomp = seasonal_decompose(BD_Train['Viviendas'],
132                                      model = 'multiplicable',
133                                      period = 12)
134 
135         Descomp.plot()
136         plt.show()
137 
138 
139         BD_Train.plot(figsize=(12,3));
140         BD_Train['Log_Dif_data'] = (np.log(BD_Train['Viviendas']))
141 
142 
143         ## Test de dickey fuller
144 
145         print(" > Is the data stationary ?")
146         dftest1 = adfuller(BD_Train.Viviendas, autolag='AIC')
147         print("Test statistic = {:.3f}".format(dftest1[0]))
148         print("P-value = {:.3f}".format(dftest1[1]))
149         print("Critical values :")
150         for k, v in dftest1[4].items():
151             print("\t{}: {} - The data is {} stationary with {}% confidence".forma
152 
```

```
153
154
155
156         print("\n > Is the log data stationary ?")
157         dftest4 = adfuller(BD_Train.Log_Dif_data.dropna(), autolag='AIC')
158         print("Test statistic = {:.3f}".format(dftest4[0]))
159         print("P-value = {:.3f}".format(dftest4[1]))
160         print("Critical values :")
161         for k, v in dftest4[4].items():
162             print("\t{}: {} - The data is {} stationary with {}% confidence".format(
163
164         ## Ajuste del modelo
165
166         Modelo = pm.auto_arima(BD_Train['Log_Dif_data'].dropna(),
167                                seasonal = True, m=12)
168         BD_Fit =BD_F['Viviendas']
169         Modelo.fit(BD_Fit.values)
170         Predicciones = Modelo.predict_in_sample(BD_Fit.values)
171         Predicciones = pd.DataFrame(Predicciones,
172                                     index = BD_Fit.index, columns = ['Viviendas'])
173
174         print(Modelo.summary())
175
176         Modelo.plot_diagnostics(figsize=(16, 8))
177         plt.show()
178         plt.savefig(file_path + "Diagnóstico Modelo-" + Filt[0] + "-" + Filt[1] + ".
179
180         # Proyección
181         Tiempo_proyectado = pd.date_range(start="2022-01", periods = 12*14, freq = '
182         Proyectados_35, conf_int = Modelo.predict(n_periods = 12*14, return_conf_int
183         Proyectados_35 = pd.DataFrame(Proyectados_35,
184                                       index = Tiempo_proyectado, columns=['Viviendas
185         conf_int = pd.DataFrame(conf_int,
186                                 index = Tiempo_proyectado, columns=['Lim_inf','Lim_s
187
188         # Llamamos a la función para graficar los resultados finales
189         title = Filt[0] + " - " + Filt[1]
190         plot_train_test_pred(BD_Train['Viviendas'],
191                              BD_Test['Viviendas'],Predicciones,
192                              Proyectados_35,conf_int, title)
193
194         ## Creamos la Base de datos
195         BD_Proyectada = BD_Train['Viviendas'].append(BD_Test['Viviendas'])
196         BD_Proyectada = BD_Proyectada.append(Proyectados_35['Viviendas'])
197
198         BD_Proyectada['UBICACION'] = Filt[0]
199         BD_Proyectada['ESTRATO'] = Filt[1]
200         BD_Proyectada['FECHA'] = BD_Proyectada.index
```

```
        BD_Proyectada_export = BD_Proyectada_export.append(BD_Proyectada)

BD_Proyectada_export.to_excel(file_path + "Proyección-" + file_name)
```

10.3. Código en R para el método Prophet

```
library("prophet")
library("readxl")
library("dplyr")
library("writexl")

dir_datos <- paste0(getwd(),"/Datos/")

f_name_viv <- "Viviendas"
f_name_sus <- "Suscriptores"

BD_Viv <- read_excel(paste0(dir_datos,f_name_viv,".xlsx"))

BD_Viv$Mes <- as.numeric(BD_Viv$Mes)
BD_Viv$Año <- as.numeric(BD_Viv$Año)
BD_Viv$Dia <- 1

BD_Viv$ds <- with(BD_Viv, Date <- sprintf("%d-%02d-%02d", Año, Mes,Dia))

BD_Viv <- BD_Viv[c("ds","ventas")]
colnames(BD_Viv)[2] <- "y"

fit <- prophet(BD_Viv)

future = make_future_dataframe(fit,periods=180,freq = "month")

# Print predictions
cat("\nPredictions:\n")
tail(future)

# Forecast
forecast <- predict(fit, future)
tail(forecast[c('ds', 'yhat',
                'yhat_lower', 'yhat_upper')])

# Output to be present
# As PNG file
png(file = "facebookprophetGFG.png")

# Plot
```

```r
plot(fit, forecast)

# Saving the file
dev.off()

########################################

## SUscriptores

BD_Sus <- read_excel(paste0(dir_datos,f_name_sus,".xlsx"))

BD_Sus <- BD_Sus %>% group_by(AÑO, MES) %>%
                     summarise(Viviendas = sum(Viviendas))

BD_Sus$MES <- as.numeric(BD_Sus$MES)
BD_Sus$AÑO <- as.numeric(BD_Sus$AÑO)
BD_Sus$Dia <- 1

BD_Sus$ds <- with(BD_Sus, Date <- sprintf("%d-%02d-%02d", AÑO, MES,Dia))

BD_Sus <- BD_Sus[c("ds","Viviendas")]
colnames(BD_Sus)[2] <- "y"

fit <- prophet(BD_Sus)

future = make_future_dataframe(fit,periods=168,freq = "month")

# Print predictions
cat("\nPredictions:\n")
tail(future)

# Forecast
forecast <- predict(fit, future)
tail(forecast[c('ds', 'yhat',
                'yhat_lower', 'yhat_upper')])

# Output to be present
# As PNG file
png(file = "ProphetViviendas.png")

# Plot
plot(fit, forecast, xlabel = "Año", ylabel = "Suscriptores")

# Saving the file
dev.off()

prophet_plot_components(fit, forecast)

```

```
89 write_xlsx(forecast, "Pronosticos_SUs.xlsx")
90
```