



## RESULTADOS sTEMO

Smart Territory Modelling

En Gijón , a 31 de Diciembre de 2020

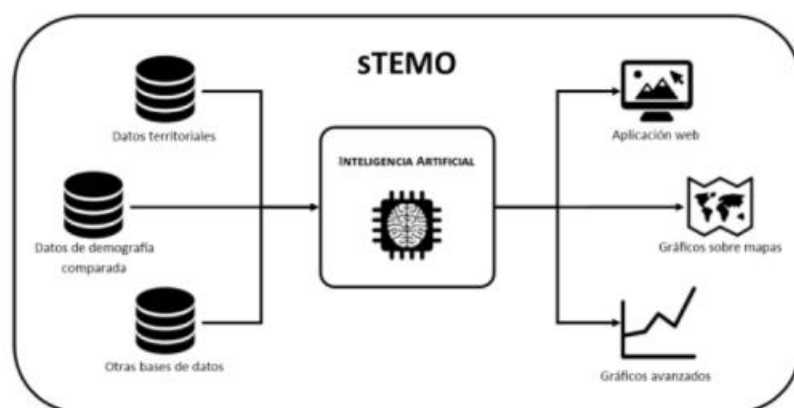
## 1. MEMORIA TÉCNICA

### Introducción

El proyecto sTEMO (Smart Territory Modelling) es ejecutado por Fundación CTIC – Centro Tecnológico entre el 1 de julio de 2018 y el 31 de diciembre de 2020.

El proyecto sTEMO se plantea con el objetivo de ofrecer un apoyo en el desarrollo de políticas activas que permitan abordar el cambio demográfico y el despoblamiento, especialmente en el medio rural, a través de una Plataforma para la Simulación de Escenarios Territoriales, aplicada a la planificación y el ordenamiento urbano y territorial. Esta plataforma se apoya en la filosofía de los SDSS (Sistemas de Soporte de Decisión Espacial) y se plantea como una plataforma escalable e interoperable bajo estándares W3C. Las simulaciones se basan en una combinación de técnicas de inteligencia artificial y análisis de decisión multicriterio sobre datos espaciales.

El proyecto plantea un caso de estudio aplicado a la problemática del despoblamiento y envejecimiento poblacional del Principado de Asturias.



*Figura 1: Esquema conceptual de la plataforma sTEMO.*

La ejecución del proyecto sTEMO se ha dividido en cuatro Hitos claramente diferenciados:

- Hito 1: Análisis de los aspectos generales en demografía y población.
- Hito 2: Modelización del sistema formado por el conjunto de variables territoriales.
- Hito 3: Diseño y desarrollo del prototipo e integración de componentes.
- Hito 4: Desarrollo del sistema completo y prueba en entorno piloto.

## **JUSTIFICACIÓN 2018**

Durante el periodo que cubre este documento, se ha realizado la ejecución completa de las tareas correspondientes al Hito 1 y se ha iniciado la ejecución del Hito 2.

A continuación, se describen las tareas realizadas en cada uno de los hitos de trabajo mencionados anteriormente y los resultados conseguidos.

### **Hito 1: Análisis de los aspectos generales en demografía y población**

El objetivo de este hito ha consistido en analizar todos aquellos aspectos generales que condicionan cualquier proceso de dinámica territorial, como puedan ser la demografía, la ordenación del territorio, u otras variables que tengan incidencia en los procesos de despoblamiento. El fin de este estudio es definir escenarios con demografía comparada para completar los datos disponibles con otros de regiones similares. Este hito se ha desarrollado a través de 4 tareas:

- T1.1: Análisis de aspectos generales en demografía y territorio.
- T1.2: Análisis de procesos y dinámicas asociados a la dinámica de despoblamiento.
- T1.3: Definición de escenarios con demografía comparada.
- T1.4: Definición de variables de incidencia en procesos de despoblamiento.

#### **T1.1: Análisis de aspectos generales en demografía y territorio**

Los asentamientos humanos se pueden definir como un conjunto de personas que interactúan entre sí y con su medio ambiente, para alcanzar sus objetivos de reproducción, producción y bienestar social. Para lograr estos hitos, se establecen organizaciones sociales que se regulan y definen los contextos de normas que rigen sus actividades. Esta interacción entre las poblaciones y su entorno en diferentes escalas interconectadas, se adapta al territorio donde se desarrolla y de forma biunívoca, adapta el territorio a sus necesidades en la medida de que su nivel desarrollo socioeconómico, cultural y tecnológico se lo permite.

Por tanto, cada población constituye una entidad con características diferenciales, pero con rasgos comunes, cuyas superposiciones espacio-temporales conforman la idiosincrasia de los pueblos y definen su futuro en función de su presente y su pasado.

Como resultado de las interacciones que las poblaciones y sus gentes establecen entre sus unidades demográficas y el territorio, los asentamientos van adquiriendo determinadas características, ritmos y tendencias de crecimiento que influyen en la composición y la distribución de la población en el territorio a lo largo del tiempo.

Un paso decisivo para la comprensión de la dinámica de una población y de sus relaciones con el territorio es la caracterización de su dinámica demográfica. Ésta permite reconocer las características básicas de la población objeto de análisis, cuántas personas la componen, dónde están y quiénes son en términos de su edad, sexo, condición étnica y cómo se desplazan en el tiempo y el espacio.

La forma en la que se comportan y evolucionan la población y sus agregaciones sólo puede ser adecuadamente comprendido si se tiene en cuenta su ritmo de transformación y su continuo cambio,

influenciado tanto por factores intrínsecos (recursos naturales, condiciones territoriales, rasgos poblacionales, etc.) y factores extrínsecos (influencia de otras áreas poblacionales, fenómenos naturales, etc.).

La demografía estudia el tipo de relaciones y forma de evolución de los ecosistemas humanos y su entorno, que asimila a modelos teóricos y conceptuales para su mejor comprensión. De este modo, se pueden establecer un conjunto de variables de los que emanan indicadores que permiten explorar, ponderar y prever las pautas de un territorio y su población en un periodo concreto.

Estos indicadores se refieren a los elementos que definen el tamaño, crecimiento, estructura y distribución de la población en un territorio, los cuales a su vez se nutren del comportamiento de la natalidad y la mortalidad, y del comportamiento de la inmigración y la emigración, procesos que periódicamente adicionan o sustraen población del territorio.

Para poder conceptualizar los componentes que intervienen en la dinámica demográfica e interacción con el territorio, se propone el siguiente esquema. En este diagrama se puede comprobar como los cambios en cada uno de sus componentes genera a su vez, cambios en el resto. Se trata, por tanto, de un sistema multivariante en el que el conjunto de variables mantiene relaciones con otras dimensiones del proceso de dinámica vital y social de la población en su conjunto, dentro de un contexto territorial concreto.

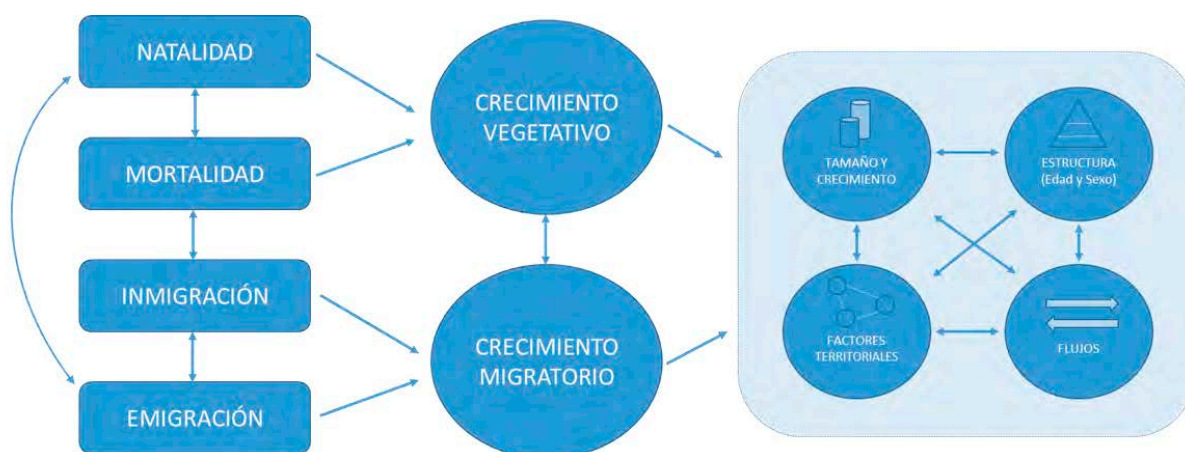


Figura 3 Diagrama sobre componentes básicos en la dinámica demográfica

Dentro de los componentes básicos de la dinámica demográfica, los más significativos serían:

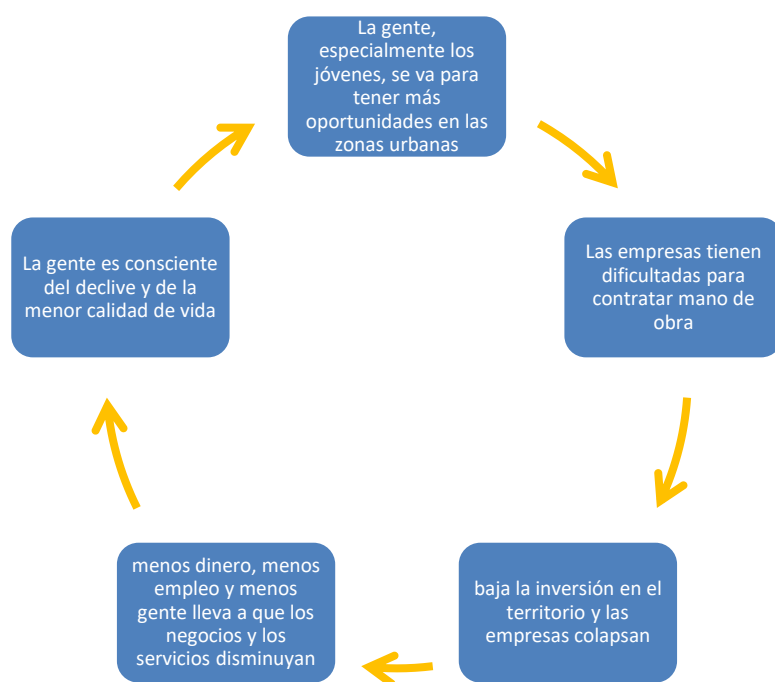
1. Su **tamaño** o número de habitantes residentes en el territorio;
2. Su **crecimiento** total o número de nuevas personas habitantes, número que se incrementa, mantiene o reduce en un periodo de tiempo considerado (generalmente un año)
3. Su **estructura** o conjunto de grupos con diferentes características de edad y sexo, étnicas y socio-culturales.
4. Su **distribución geográfica** o la concentración más o menos alta de población en determinadas zonas del territorio

### **T1.2: Análisis de procesos y dinámicas asociados a la dinámica de despoblamiento**

La despoblación no es un fenómeno espontáneo ni contemporáneo en el tiempo, sino complejo y con múltiples causas, como por ejemplo, entre otros factores, el mantenimiento de pautas de comportamiento y percepción social del espacio propias de un país como España de tardía industrialización, incluida una percepción de la oposición “ciudad-campo” desfasada y perjudicial para las áreas rurales; o el complicado paso del sistema agrario tradicional a una nueva economía agraria deficientemente acomodada en el sistema de mercado.

Especialmente preocupante es el panorama en varias comunidades autónomas, como Castilla y León, Asturias, Aragón, Galicia y Cantabria (en estas dos últimas, sobre todo en el interior) y de provincias como Soria, Teruel y Cuenca. Con otra dimensión también se manifiesta en Extremadura y en Castilla-La Mancha, al igual que en numerosos municipios de La Rioja y en otros de Navarra.

Los pueblos se han ido vaciando progresivamente, la escasa población que queda en muchos municipios está cada vez más envejecida y en muchos casos, además, la dispersión del poblamiento y la desigual accesibilidad dificultan una adecuada distribución de servicios sociales, sanitarios, educativos, etc.



**Figura 4: Ciclo de retroalimentación del despoblamiento. Fuente: MIT Center for Civic Media**

En la figura anterior, se puede comprobar cómo funciona el ciclo del despoblamiento, es decir, las causas que lo retroalimenta. Este esquema fue planteado por el MIT Center for Civic Media en el marco de una región rural despoblada de Japón y su aplicación puede ser global.

Dentro de los componentes básicos de la dinámica demográfica, los más significativos serían:

1. Su tamaño o número de habitantes residentes en el territorio;
2. Su crecimiento total o número de nuevas personas habitantes, número que se incrementa, mantiene o reduce en un periodo de tiempo considerado (generalmente un año)

3. Su estructura o conjunto de grupos con diferentes características de edad y sexo, étnicas y socio-culturales.
4. Su distribución geográfica o la concentración más o menos alta de población en determinadas zonas del territorio

Estos son los componentes con los que usualmente se caracteriza de manera rápida una población. Sin embargo, ellos resultan del juego entre dos tipos de crecimientos: El crecimiento natural o vegetativo y el crecimiento migratorio.

Crecimiento Vegetativo = + Nacimientos durante el año –  
Defunciones durante el año

$$CV = + N - D$$

El Crecimiento Vegetativo se refiere a la capacidad POTENCIAL de aumento de la población y proviene de los aportes que periódicamente hacen los nuevos Nacimientos (N), al volumen de población ya existente y de las pérdidas que periódicamente se producen por efecto de la mortalidad (Defunciones, D). El Crecimiento Vegetativo (CV), se establece como el saldo (o diferencia aritmética) entre Nacimientos, N, y Decesos, D, registrados en el territorio en un período de tiempo (por lo general de un año).

Crecimiento Migratorio = + Inmigrantes durante el año –  
Emigrantes durante el año

$$CM = + I - E$$

El Crecimiento Migratorio (CM) se refiere al intercambio entre Inmigrantes, I, y Emigrantes, E, en el territorio, durante un período de tiempo.

Cuando se dispone de información de los cuatro componentes (N, D, I, E), la dinámica demográfica puede describirse cuantitativamente mediante el uso de una sencilla función matemática que les incluye a todos y que se denomina la “ecuación de compensación”.

Su nombre se deriva del hecho de que en realidad la expresión cuantitativa de la dinámica demográfica de una población es el juego de compensaciones entre los volúmenes de nacimientos, defunciones, inmigrantes y emigrantes.

Población al final del año = Población a principio de año + Nacimientos del año – Defunciones del año +  
Inmigrantes del año – Emigrantes del año

$$Pr = Po + N - D + I - E$$

Población al final del año = Población a principio de año + Saldo vegetativo + Saldo Migratorio

$$Pr = Po + (\pm CV) + (\pm MN)$$

### **T1.3: Definición de escenarios con demografía comparada**

Una vez definidos los aspectos y procesos generales en demografía, territorio y despoblamiento, se pasa en este capítulo a la definición profunda de las variables con incidencia en procesos de despoblamiento y que conformarán la base de la definición de escenarios demográficos para realizar una serie de análisis comparativos y proponer modelos de comportamiento futuro.

Para poder definir los escenarios en los que poder estudiar las regiones menos densamente pobladas, tomando como referencia el ámbito de la Unión Europea, se han de buscar aquellas regiones con distintos valores sobre densidad de población (fundamentalmente aquellos con valores bajos o muy bajos) para poder conocer todos sus elementos demográficos y establecer escenarios *ad futurum*.

Se exponen a continuación, las variables o indicadores que se han elegido para formar parte del proyecto **STEMO** en la categoría de comparación de escenarios en demografía comparada.

CÓDIGO	INDICADOR ESPECÍFICO
IE 01	Movimiento natural de la población
IE 02	Tasa de natalidad
IE 03	Tasa de mortalidad
IE 04	Crecimiento Natural
IE 05	Migración
IE 06	Tasa de fecundidad
IE 07	Esperanza de vida
IE 08	Proporción de género
IE 09	Edad media
IE 10	Índices de dependencia
IE 11	Evolución del empleo
IE 12	Evolución del desempleo
IE 13	Tasa de empleo por grupos
IE 14	Tasa de desempleo por grupos
IE 15	Tasa de actividad económica
IE 16	Acceso a banda ancha
IE 17	Evolución de VAB por hogares
IE 18	Evolución de la renta por habitante

### **T1.4: Definición de variables de incidencia en procesos de despoblamiento**

El objetivo del proyecto STEMÓ en este sentido consiste en buscar aquellas variables que poseen relación directa con el proceso de envejecimiento y además cumplen una serie de requisitos sobre su fiabilidad, frecuencia de actualización, capacidad de comparación, escala de aplicación, etc.

El reto demográfico en las regiones Europeas requiere acciones integrales pero que, al mismo tiempo, deben adaptarse a cada situación particular. Las medidas en el mundo rural no pueden ser independientes de las adoptadas en el mundo urbano, aunque por supuesto, esto no quiere decir que la respuesta deba ser única. Incluso, la división rural-urbano está claramente superada por la elevada heterogeneidad que muestran ambos mundos.

Por tanto, para determinar aquellas variables que asociadas a los procesos de despoblamiento y que además resulten relativamente sencillas de obtener, fiables y comparables con otras regiones, se ha recurrido a una intensa búsqueda bibliográfica y documental.

Como resultado del trabajo descrito, se ha orientado la categorización de las variables relacionadas con el despoblamiento hacia aquellas que además pueden tener relación directa con la aplicación de medidas de gobernanza que permitan la aplicación de medidas para paliar o cambiar la tendencia actual, que se da fundamentalmente en el ámbito rural. De este modo, se ha considerado como más apropiado poner el foco en los Programas de Desarrollo Rural (PDR) que articulan la aplicación de la Política Agraria Común (PAC) a través del Fondo Europeo Agrícola de Desarrollo Rural (FEADER).

Se exponen a continuación, las variables o indicadores que se han elegido para formar parte del proyecto **STEMÓ** en esta categoría:

CÓDIGO	INDICADOR DE CONTEXTO
C.01	Población
C.02	Estructura de edad
C.03	Territorio
C.04	Densidad de población
C.05	Tasa de empleo (*)
C.06	Tasa de autoempleo
C.07	Tasa de desempleo
C.08	PIB per cápita (*)
C.09	Tasa de pobreza (*)

C.10	Estructura de la economía
C.11	Estructura del empleo
C.12	Productividad laboral por sector económico
C.13	Empleo por actividad económica
C.14	Productividad laboral en la agricultura
C.15	Productividad laboral en la silvicultura
C.16	Productividad laboral en la industria alimentaria
C.17	Explotaciones agrícolas (granjas)
C.18	Superficie agrícola
C.19	Superficie agrícola dedicada a la agricultura ecológica
C.20	Tierras de regadío
C.21	Unidades de ganado
C.22	Mano de obra agrícola
C.23	Estructura de edad de los administradores de las explotaciones
C.24	Formación agraria de los responsables de las explotaciones agrícolas
C.25	Renta de los factores agrarios (*)
C.26	Renta agroempresarial (*)
C.27	Productividad total de los factores en la agricultura (*)
C.28	Formación bruta de capital fijo en la agricultura
C.29	Bosques y otras tierras boscosas (FOWL)
C.30	Infraestructura turística

C.31	Cobertura del suelo
C.32	Áreas que enfrentan restricciones naturales y otras restricciones específicas - ANCs6
C.33	Intensidad agrícola
C.34	Zonas Natura 2000
C.35	Índice de aves agrícolas (FBI)
C.36	Estado de conservación de los hábitats agrícolas (pastizales)
C.37	Agricultura de alto valor natural
C.38	Bosque protegido
C.39	Extracción de agua en la agricultura
C.40	Calidad de las aguas
C.41	Carbono orgánico total del suelo (SOC) en tierras cultivables: megatoneladas (Mt);
C.42	Erosión del suelo por el agua
C.43	Producción de energía renovable a partir de la agricultura y la silvicultura
C.44	Uso de energía en la agricultura, la silvicultura y la industria alimentaria
C.45	Emisiones procedentes de la agricultura

## **Hito 2: Modelización del sistema formado por el conjunto de variables territoriales**

El objetivo de este hito consiste en parametrizar los procesos, actividades y dinámicas asociadas al territorio con el fin de identificar los escenarios y variables que tienen incidencia sobre el proceso, con el fin de crear un sistema de modelización que sea capaz de integrar todos estos parámetros. Este hito se divide en cuatro tareas, si bien durante el año 2018 se ha finalizado la primera de las tareas y apenas se ha empezado la segunda de ellas, lo cual es acorde a la planificación inicial del proyecto:

- T2.1: Descripción del proceso general de planificación, gestión y ordenación del territorio.
- T2.2: Parametrización de procesos, actividades y dinámicas asociadas a flujos.
- T2.3: Elección de variables influyentes.

- T2.4: Diseño y desarrollo del sistema de modelización.

### **T2.1: Descripción del proceso general de planificación, gestión y ordenación del territorio**

La Ordenación del Territorio, en su acepción más generalizada, se define como el marco jurídico general conformado por normas, programas, directrices y planes, que orientan y regulan las actuaciones y procesos de asentamiento sobre el territorio, integrador y coordinador de las políticas públicas con incidencia en el territorio entendido como espacio físico.

A partir de la Constitución de 1978 la ordenación del territorio es una competencia exclusiva de las Comunidades Autónomas (artículo 148.1.3), al haber asumido las mismas en sus Estatutos de Autonomía, habiendo ejercido en la actualidad prácticamente todas sus competencias legislativas en la materia.

Pueden establecerse convencionalmente dos grupos de Comunidades Autónomas en función al tratamiento normativo que efectúan de la Ordenación del Territorio en su relación con el urbanismo: las Comunidades que desarrollan la materia con carácter independiente o autónomo y aquellas que regulan la ordenación territorial de manera conjunta con el urbanismo.

La Ordenación del Territorio, en su acepción más generalizada, se define como el marco jurídico general conformado por normas, programas, directrices y planes, que orientan y regulan las actuaciones y procesos de asentamiento sobre el territorio, integrador y coordinador de las políticas públicas con incidencia en el territorio entendido como espacio físico. Supone una vertebración de las políticas específicas sectoriales que pueden afectar de algún modo al territorio.

Una de las definiciones pioneras acuñada por la doctrina científica española es la de Morell Ocaña, que entiende por ordenación del territorial, "más que una política concreta, el cuadro general que enmarca y da coherencia a una vasta serie de políticas específicas. De entre ellas, por su importancia y por su tendencia a visiones generalistas, destacan poderosamente dos: la urbanística y del desarrollo económico... En cualquier caso, la política de ordenación del territorio tiende a situarse en un plano más alto y unir en una sola planificación general las distintas planificaciones concretas, quedando éstas convertidas en planificaciones de carácter sectorial."

Es común entre la doctrina citar la definición recogida en la Carta Europea de la Ordenación del Territorio, auspiciada en el seno del Consejo de Europea (1983): "es la expresión espacial de la política económica, social, cultural y ecológica de toda la sociedad."

La Ordenación del Territorio, por tanto, es uno de los instrumentos con los que las distintas Administraciones Públicas cuentan para desarrollar sus políticas y lograr el objetivo de la cohesión de oportunidades a través de la equiparación de oportunidades en cualquier territorio soberano. En cuanto al fenómeno del despoblamiento y el envejecimiento de la población, dado que es en el medio rural en el que este fenómeno se da con mayor crudeza, se ha priorizado el estudio en este tipo de territorios.

El medio rural es uno de los elementos fundamentales que vertebran social y económicamente el territorio europeo y configuran su paisaje. En los territorios rurales se concentra un importante porcentaje de la población, siendo el soporte físico de un extenso y variado patrimonio natural y cultural que representa hoy uno de los grandes valores de la UE, además de constituir la base de muchas actividades económicas que generan importantes niveles de empleo y renta.

Tradicionalmente el medio rural se ha percibido como un espacio caracterizado más por sus diferencias respecto al medio urbano (hasta el punto de considerársele su antítesis), que por sus propios valores y atributos, habiendo sido calificado con frecuencia como un lugar donde se manifiestan carencias en infraestructuras y equipamientos y donde escasean las oportunidades de promoción social y económica para la población.

Compartiendo este planteamiento, se considera que las políticas públicas expresadas a través de la Ordenación del Territorio y sus instrumentos de desarrollo, deben abrir su campo de actuación, contemplando a todos los sectores y actores del territorio desde una perspectiva global e integral y en un contexto de gobernanza flexible en sus múltiples niveles. Señalamos, en definitiva, que las nuevas políticas deben impulsar estrategias de cooperación entre el medio rural y el medio urbano favoreciendo las sinergias necesarias entre ambas poblaciones, y todo ello con el objetivo de aprovechar los recursos tanto endógenos, como exógenos, para situarlos en la senda que lleve al desarrollo sostenible de tales territorios.

Entre los instrumentos legislativos que constituyen hoy herramientas para promover la cohesión de los territorios rurales en España, la Ley 47/2007 de Desarrollo Sostenible del Medio Rural es uno de los más destacados por su novedad en el planteamiento y por la metodología que utiliza.

Junto a otros instrumentos legislativos, como la Ley 42/2007 de Patrimonio Natural y la Biodiversidad o los programas regionales de desarrollo rural –en aplicación del reglamento europeo en las distintas Comunidades Autónomas–, esta Ley puede verse como un instrumento con potencial suficiente para aplicar en las áreas rurales españolas una política de cohesión territorial que respete la singularidad de cada territorio y que apueste por una adecuada vertebración de esas áreas con el espacio circundante, sea también rural o sea sólo urbano.

### **T2.2: Parametrización de procesos, actividades y dinámicas asociadas a flujos**

El ámbito de análisis en esta tarea, se centra en aquellas regiones que comparten un problema acuciante de despoblamiento como son **Galicia, Asturias, Castilla-León y Aragón** y sus provincias.

El territorio implicado, comparte problemáticas e intereses comunes, fruto de sus características socioeconómicas, de su relación de vecindad y de sus profundos vínculos históricos que les ha permitido desarrollar fórmulas de colaboración de vecindad entre ellas. Al mismo tiempo, se dispone de fórmulas de colaboración horizontal en el marco de una nueva política de cooperación y de cohesión del Estado Autonómico, iniciativas y actuaciones que difícilmente se podrían llevar a cabo de forma aislada por una Comunidad Autónoma. Ejemplo de esta iniciativa es la constitución de un [Foro estable de colaboración entre Comunidades Autónomas con Desafíos Demográficos \(FREDD\)](#).



**Figura 5: Ubicación geográfica de las regiones NUTS2 y NUTS3 estudiadas**

En cuanto a los procesos, actividades y dinámicas asociadas a flujos que dan lugar a fenómenos de despoblamiento, tras analizar el conjunto de regiones NUTS2 y NUTS3 que presentan los mayores índices de despoblamiento, se puede concluir con lo siguiente.

En la mayoría de los casos, se trata de **áreas del interior rural** –frecuentes en muchas zonas de la ruralidad profunda española y, en particular, en las zonas de montaña–, que suelen estar alejadas de los centros urbanos, sufrir importantes déficits estructurales, tener bajos niveles de población, estar escasamente equipadas y disponer de un sistema alimentario poco dinámico, debido a que o bien la actividad agraria no ha desarrollado un sector industrial o comercial vinculado a ella o bien a que la agricultura es poco competitiva. En cualquier caso, su sistema alimentario local no desempeña el papel de ser un motor capaz de garantizar las rentas de su población, una población que apenas consigue reproducirse social y económicamente. El resto de la actividad económica es igualmente débil, y su situación en infraestructuras tampoco permite a la población aprovechar el alto valor ambiental de estos territorios.

Son áreas con un grave problema de despoblamiento y serias probabilidades de abandono, y donde a duras penas se consigue alcanzar el umbral de población mínimo para hacer viable cualquier proyecto de desarrollo y justificar, desde un punto de vista social, las inversiones en infraestructuras y equipamientos que serían necesarias para garantizar a los que todavía allí residen unas mínimas condiciones de vida. No obstante, el valor de esos territorios es, en muchos casos, un valor intangible, no comercializable, sino vinculado a la preservación del paisaje o la conservación del espacio natural. De hecho, buena parte de su valor ambiental está vinculado a actividades de larga tradición histórica (agrosilvopastoriles, cinegéticas, pesquero-fluviales, recolección de frutos silvestres, apicultura, micología...) cuyo abandono conllevaría la pérdida de importantes elementos de biodiversidad social y cultural relacionados con el conocimiento de prácticas locales.

El reto en buena parte de estos territorios, muy heterogéneos entre sí, es cómo garantizarle a la población unas condiciones dignas de vida como base para la conservación de los espacios naturales, dado que su posible despoblamiento podría crear áreas abandonadas con un alto riesgo de incendios y un deterioro

general del entorno. Desde nuestro punto de vista, creemos que sería útil abrir un amplio debate social y político sobre el futuro y viabilidad de dichos territorios, rompiendo así el tabú a que han estado sometidos estos asuntos, un debate que debe incluir el modelo de gestión territorial que queremos y no sólo la mera descripción de las funciones ambientales que deben cumplir.

Dadas las características de estos territorios y la necesidad de impulsar en ellos la más amplia cooperación social e institucional, deberían formar parte de su sistema de gobernanza las corporaciones locales, pero también las entidades responsables de las políticas de conservación de los espacios naturales, además de las asociaciones ambientalistas presentes en dichas áreas, los grupos de acción local, las organizaciones agrarias, las asociaciones forestales, las cooperativas y, en general, las asociaciones de la industria y el pequeño comercio ubicadas en la zona.

Los actores del sistema educativo, del sistema financiero (bancos, cajas de ahorro y cajas rurales) y de los medios de comunicación local, así como las asociaciones cívicas y culturales presentes en el territorio, son también elementos importantes del sistema de gobernanza, al ser generadores de confianza y capital social entre los vecinos, y vías para reactivar la identidad comarcal y la autoestima de la población.

De este modo, y con un adecuado programa de sensibilización social, se podrían superar las carencias de participación que se han producido en la formulación y puesta en marcha de los planes de desarrollo sostenible en estos territorios, y que han conducido a un grave déficit de legitimidad, provocando el recelo, cuando no el rechazo, de la población afectada y generando serias dificultades para una eficiente gestión de los mismos.

Además de este tipo de regiones que suponen el rango más alejado de los núcleos urbanos, se encuentran otras situaciones más asimilables a cabeceras de subregión (villas, núcleos con cierto dinamismo, etc.) que suponen el nodo de conexión entre ambos y que deberían de constituir uno de los focos principales de atención para poder “rescatar” a las zonas más despobladas e invertir la tendencia de estas mismas regiones.

Se trata por tanto, de **zonas que se encuentran en una situación intermedia**, donde el sistema alimentario local, y dentro de él la agricultura, se combina con otras actividades económicas, y donde la población, gracias, entre otras cosas, a la movilidad geográfica, a su buena posición logística y a disponer de adecuadas infraestructuras, diversifica las fuentes de renta aprovechando las distintas oportunidades que les ofrece el entorno económico (empleos en el sector industrial, trabajos en el sector servicios de las ciudades y pueblos más cercanos,...).

Son áreas donde, si bien los sistemas alimentarios pueden encontrar un amplio mercado local y comarcal para sus productos, su principal dinamismo no proviene del sector agroalimentario, sino de actividades propiciadas por su cercanía con los núcleos urbanos y que favorecen la práctica del *commuting* (desplazamiento diario de trabajadores de las zonas rurales a las ciudades para desempeñar tareas en distintos servicios).

Además, por su localización estratégica y por las ventajas comparativas que ofrecen (gracias a las TIC y a la mejora de las comunicaciones viarias), algunas de estas áreas actúan como importantes polos de atracción para los profesionales autónomos e incluso para el personal que trabaja en los servicios públicos (sanitarios, educativos, asistencia social...), transformando la composición de estas comunidades rurales e introduciendo un nuevo dinamismo social y económico en ellas. Son territorios donde se produce un

flujo migratorio de doble sentido (rural-urbano y urbano-rural), que es su gran valor y fortaleza de cara a garantizar su desarrollo presente y futuro.

El desafío en estas zonas es comprender que tanto el medio rural como los núcleos urbanos forman parte del mismo territorio y que el desarrollo pasa por el respeto a los valores naturales, culturales, arquitectónicos y paisajísticos, al ser parte de un patrimonio que es de todos. Para ello es necesario intensificar las relaciones entre ambos espacios, poniendo a disposición de los actores económicos y sociales los medios e infraestructuras adecuadas para hacer posible esa interacción, sobre todo en lo que se refiere al transporte sostenible, la movilidad y la comunicación.

La ordenación del territorio, planteada, en definitiva, con un sentido integral y sostenible y en concertación con los actores locales, debe jugar una función clave a la hora de evitar el riesgo de una desenfrenada expansión urbana en estas áreas, desempeñando un importante papel en ello la figura de los “polos de desarrollo rural” que propugna el **Libro Verde de la Cohesión Territorial**.

No obstante, las acciones de tipo sectorial provenientes de los departamentos de agricultura de las administraciones públicas, son insuficientes para abordar el desarrollo de estas áreas, por mucho que puedan propiciar la mejora del sector agroalimentario y favorecer su integración en los mercados locales.

Por ello, será imprescindible una implicación real y coordinada de los diversos departamentos ministeriales y/o de la administración regional, así como de las corporaciones locales y las entidades financieras, para establecer las sinergias necesarias que garanticen las infraestructuras que demandan los agentes económicos y sociales, y todo ello en el marco de instancias de gestión y concertación que superen los restringidos ámbitos locales y se adentren en una visión territorial más amplia (supramunicipal).

Asimismo, será necesario establecer una cooperación institucional eficiente y concertada para satisfacer las demandas de los ciudadanos en materia de calidad de vida, equipamientos y servicios, y todo ello con el objetivo de generar nuevas oportunidades de empleo para la población.

Lo importante es que las políticas públicas contemplen el territorio desde una perspectiva integral, tratándolo como un amplio espacio donde confluyen actividades diversas que deben interactuar de forma eficiente, sin olvidar que las estrategias deben ser compatibles con la dinámica global de cada uno de los sectores económicos.

De lo que se trata, en definitiva, es de crear el entorno apropiado para que fluyan las iniciativas, se produzca la interacción rural/urbana, y la población pueda aprovechar las oportunidades que ofrece el propio territorio.

### **JUSTIFICACIÓN 2019**

Durante el periodo que cubre este documento, se ha finalizado la ejecución del hito 2, y se ha ejecutado en su totalidad el Hito 3.

A continuación, se describen las tareas realizadas en cada uno de los hitos de trabajo mencionados anteriormente y los resultados conseguidos.

## **Hito 2: Modelización del sistema formado por el conjunto de variables territoriales**

El objetivo de este hito consiste en parametrizar los procesos, actividades y dinámicas asociadas al territorio con el fin de identificar los escenarios y variables que tienen incidencia sobre el proceso, con el fin de crear un sistema de modelización que sea capaz de integrar todos estos parámetros. Este hito se divide en cuatro tareas, la primera de las cuales se finalizó en la anualidad 2018, terminándose los trabajos del resto de ellas en esta anualidad:

- T2.1: Descripción del proceso general de planificación, gestión y ordenación del territorio.
- T2.2: Parametrización de procesos, actividades y dinámicas asociadas a flujos.
- T2.3: Elección de variables influyentes.
- T2.4: Diseño y desarrollo del sistema de modelización.

### **T2.2: Parametrización de procesos, actividades y dinámicas asociadas a flujos**

Los trabajos de esta tarea se iniciaron en la anualidad 2018, y la actividad realizada fue reflejada en el correspondiente informe de justificación, por lo que no se repiten en el presente informe. Se resumen aquí los trabajos realizados durante la anualidad 2019.

A través del INE (Instituto Nacional de Estadística) se proporciona información detallada de cada uno de los municipios nacionales anualmente acerca de la población con respecto al género y a la edad ([link INE](#)). El histórico disponible es anual desde 1998 hasta la actualidad. Debido a diferentes cambios en la presentación de los resultados por parte del INE, ha sido necesario distinguir el pre-procesado de éstos en función del año de procedencia, quedando dos grupos de datos a analizar que se dividen en [2011-actualidad] y [1998-2010].

Los resultados más recientes, desde el 2011 en adelante, se presentan por parte del INE en conjunto para todos los municipios nacionales en un solo fichero, en formato PX.

En la siguiente figura se presenta un fragmento de un fichero PX con la información del padrón continuo del año 2018. Nótese como los valores de edad se presentan de forma agregada por intervalos. Si bien para los últimos años desde 2011 dicha información se puede obtener de forma individualizada año a año, el histórico desde 2010 hacia atrás se presenta de esta única forma, por lo que, por unicidad, se ha procedido a su recogida a través de la misma estructura.

```

CHARSET="ANSI";
AXIS-VERSION="2000";
CREATION-DATE="20150325";
SUBJECT-AREA="00.- Nacional ";
SUBJECT-CODE="000000";
MATRIX="00000001";
TITLE="Población por sexo, municipios y edad (grupos quinquenales).";
CONTENTS="Población";
UNITS="personas";
STUB="sexo","municipios";
HEADING="edad (grupos quinquenales)";
VALUES("sexo")="Ambos sexos","Hombres","Mujeres";
VALUES("edad (grupos quinquenales)")="Total"," 0-4"," 5-9"," 10-14 "," 15-19",
" 20-24"," 25-29"," 30-34"," 35-39"," 40-44",
" 45-49"," 50-54"," 55-59"," 60-64"," 65-69"," 70-74"," 75-79"," 80-84"," 85-89"," 90-94"," 95-99"," 100 y más";
VALUES("municipios")="Total",
"01051-Agurain/Salvatierra ",
"01001-Alegria-Dulantzi ",
"01002-Amurrio ",
"01049-Añana ",
"01003-Aramaio ",
...;
PRETEXT("sexo")=0;
PRETEXT("edad (grupos quinquenales)")=0;
ELIMINATION("sexo")="Ambos sexos";
ELIMINATION("municipios")="Total";
ELIMINATION("edad (grupos quinquenales)")="Total";
DECIMALS=0;
SHOWDECIMALS=0;
LAST-UPDATED="05/10/99";
SOURCE="Instituto Nacional de Estadística ";
CONTACT="INE E-mail:www.ine.es/infoine. Internet: www.ine.es. Tel: +34 91 583 91 00 Fax: +34 91 583 91 58";
MAP("municipios")="muni0109_esp";
INFOFILE="ayudal";
DATA=
46722980 2079174 2421550 2434173 2253314 2288096 2557214 2909061 3624705 3992655 3768313 3586094 3196535 2703945
2396644 2114482 1525430 1413284 948377 397134 97293 15507
5018 335 339 250 233 194 228 301 445 469 380 386 346 320 232 185 110 109 93 48 11 4
2887 159 239 230 163 106 121 123 225 343 334 257 152 114 102 86 57 38 26 11 1 0
10286 570 554 433 382 408 484 606 810 791 733 796 814 716 680 513 327 313 227 108 17 4
151 1 6 3 1 10 9 3 8 11 15 18 16 9 10 7 7 6 5 5 1 0
1483 63 79 94 80 63 51 55 88 114 109 122 137 109 77 82 53 53 37 10 7 0
219 8 14 20 6 5 4 12 16 29 28 22 14 9 10 7 5 4 4 2 0 0
717 34 44 31 14 18 20 40 53 64 68 55 46 50 42 47 25 26 25 13 1 1
982 29 47 49 68 49 34 41 54 59 92 103 94 75 79 37 19 22 17 8 4 2
1831 102 133 103 84 57 63 103 155 186 136 140 138 123 93 62 37 45 43 20 7 1
1643 71 85 84 72 73 76 71 101 121 139 163 160 105 97 65 44 56 38 17 5 0
2967 155 195 210 137 95 106 127 208 349 302 251 214 135 119 120 75 76 59 27 7 0
286 6 8 9 16 18 8 13 20 14 23 32 23 15 22 22 11 10 11 4 0 1

```

El pre-procesado de los datos ha sido realizado a través del software de programación estadística R. Para la lectura de los ficheros PX se ha recurrido al paquete diseñado para tal fin: "pxR". De este modo, es posible obtener dicha información en formato tabular.

Ha sido necesario extraer la información de cada municipio de forma codificada para evitar ambigüedades en la escritura de los nombres dependiendo de las fuentes analizadas. Para ello, se ha recurrido al código INE, el cual representa un identificador único de cada municipio de cinco dígitos, donde los dos primeros se corresponden al código de la provincia, y los tres últimos al código del propio municipio. Dicho código se presenta en la variable "municipios", donde se ha extraído de forma automática a una variable "codigo\_ine", e identificando los casos totales mediante códigos completados con ceros ("00000" para total nacional, "XX000" para total provincial).

Finalmente, se ha reestructurado la información, de modo que sea más sencillo su uso a futuro. De este modo, la variable género se ha dividido en diferentes atributos, de modo que se proporcione de forma conjunta la información de ambos géneros y totales en una misma entrada para cada municipio, año y franja de edad.

	codigo_ine	anio	edad	total	hombres	mujeres
1	00000	2011	0-4	2470560	1273539	1197021
2	00000	2011	10-14	2187263	1124459	1062804
3	00000	2011	15-19	2267904	1167727	1100177
4	00000	2011	20-24	2611257	1326358	1284899
5	00000	2011	25-29	3274267	1662639	1611628
6	00000	2011	30-34	4059412	2088902	1970510
7	00000	2011	35-39	4095793	2112291	1983502
8	00000	2011	40-44	3853445	1967008	1886437
9	00000	2011	45-49	3580521	1801840	1778681
10	00000	2011	5-9	2356355	1208743	1147612
11	00000	2011	50-54	3191434	1587776	1603658
12	00000	2011	55-59	2677814	1317232	1360582
13	00000	2011	60-64	2470911	1195059	1275852
14	00000	2011	65-69	2132734	1013458	1119276
15	00000	2011	70-74	1794140	817643	976497
16	00000	2011	75-79	1762589	760014	1002575
17	00000	2011	80-84	1306383	510779	795604
18	00000	2011	85-	1097711	347720	749991
19	00000	2012	0-4	2443703	1259518	1184185
20	00000	2012	10-14	2214708	1137976	1076732

La figura anterior muestra un fragmento de dicha reestructura, formada por un total de 1.169.388 entradas y 6 variables.

Para los resultados entre los años 1998 y 2010, no se proporciona la información agregada a nivel nacional, sino que es necesario realizar tantas descargas por año como provincias y ciudades autónomas componen el territorio nacional, es decir, 52.

Así, se han de descargar 52 ficheros con respecto a cada uno de los años en estudio entre 1998 y 2010, es decir, un total de  $52 \times 13 = 676$  ficheros. Para ello, y dado que se trata de una tarea mecánica, se ha procedido a su descarga automatizada desde R, a través de la aplicación de la función "download.file". La estructura de las URLs de descarga es la siguiente:

[http://www.ine.es/jaxi/files/\\_px/es/csv\\_sc/t20/e245/p05/aAÑO/I0/000CODIGO\\_PROVINCIA001.sc?nocab=1](http://www.ine.es/jaxi/files/_px/es/csv_sc/t20/e245/p05/aAÑO/I0/000CODIGO_PROVINCIA001.sc?nocab=1)

donde **AÑO** y **CODIGO\_PROVINCIA** son los campos variables en URL para determinar la descarga a realizar.

Debido a diferencias en la codificación de los datos entre 2006 y 2010, y entre 1998 y 2005, no es posible la lectura de todos éstos a través del formato PX previo, por lo que ha sido necesario proceder al procesado de los datos mediante CSV. En concreto, dichos ficheros presentan un formato análogo a los XLSX proporcionados por el INE, los cuales tienen entradas que no satisfacen la estructura general.

```

1 Estadística del Padrón Continuo
2 01.- Alava
3 Población por sexo, municipios y edad (grupos quinquenales)
4 Unidades: personas
5
6 ;Total;0-4;05-09;10-14;15-19;20-24;25-29;30-34;35-39;40-44;45-49;50-54;55-59;60-64;65-69;70-74;75-79;80-84;85 y más;
7 Ambos sexos;
8
9 Total;286497.0;10894.0;11135.0;13360.0;17741.0;23631.0;24565.0;24522.0;23953.0;22191.0;20187.0;20107.0;16556.0;13690.0;14250.0;1119
8.0;8508.0;5260.0;4749.0;
10 01001 Alegria-Dulantzi;1401.0;74.0;63.0;67.0;80.0;113.0;132.0;158.0;135.0;114.0;99.0;74.0;84.0;49.0;38.0;40.0;31.0;30.0;20.0;
11 01002 Amurrio;9720.0;333.0;383.0;524.0;679.0;835.0;725.0;781.0;817.0;800.0;788.0;662.0;506.0;423.0;477.0;401.0;295.0;169.0;122.0;
12 01003 Aramaior;1416.0;56.0;61.0;67.0;71.0;89.0;112.0;111.0;132.0;126.0;102.0;91.0;73.0;77.0;74.0;64.0;37.0;41.0;32.0;
13 01004 Artziniega;1338.0;50.0;67.0;81.0;83.0;96.0;84.0;111.0;136.0;124.0;104.0;74.0;64.0;45.0;70.0;52.0;47.0;26.0;24.0;
14 01006 Armiñon;146.0;2.0;3.0;3.0;7.0;10.0;7.0;18.0;13.0;12.0;9.0;5.0;8.0;8.0;16.0;7.0;8.0;5.0;5.0;
15 01008 Arzua-Ubarrundia;726.0;24.0;21.0;35.0;48.0;56.0;44.0;57.0;66.0;69.0;63.0;53.0;33.0;30.0;37.0;31.0;34.0;13.0;12.0;
16 01009 Asparrena;1578.0;71.0;60.0;58.0;72.0;100.0;118.0;143.0;149.0;127.0;87.0;88.0;62.0;75.0;95.0;103.0;79.0;46.0;45.0;
17 01010 Ayala;2064.0;75.0;85.0;83.0;101.0;134.0;168.0;197.0;194.0;143.0;110.0;122.0;106.0;104.0;125.0;105.0;94.0;65.0;53.0;
18 01011 Baños de Ebro;328.0;17.0;14.0;12.0;18.0;14.0;23.0;23.0;25.0;18.0;17.0;19.0;20.0;16.0;26.0;26.0;18.0;12.0;10.0;
19 01013 Barrundia;610.0;25.0;15.0;16.0;29.0;31.0;47.0;57.0;73.0;49.0;38.0;37.0;43.0;28.0;42.0;26.0;21.0;17.0;16.0;
20 01014 Berantevilla;430.0;12.0;8.0;11.0;22.0;45.0;40.0;42.0;33.0;13.0;21.0;35.0;32.0;33.0;31.0;15.0;17.0;8.0;12.0;
21 01016 Bernedo;547.0;12.0;12.0;13.0;15.0;31.0;46.0;39.0;46.0;31.0;29.0;37.0;35.0;30.0;45.0;49.0;40.0;22.0;15.0;
22 01017 Campezo;1060.0;31.0;29.0;37.0;49.0;68.0;59.0;91.0;95.0;75.0;71.0;64.0;70.0;60.0;56.0;69.0;54.0;48.0;34.0;
23 01018 Zigoitia;1242.0;39.0;53.0;66.0;79.0;96.0;93.0;128.0;132.0;124.0;92.0;70.0;46.0;48.0;39.0;48.0;16.0;22.0;
24 01019 Crispain;188.0;9.0;6.0;4.0;8.0;10.0;11.0;17.0;22.0;14.0;9.0;11.0;14.0;14.0;12.0;7.0;7.0;9.0;4.0;
25 01020 Cuartango;374.0;5.0;7.0;12.0;17.0;19.0;20.0;31.0;41.0;42.0;30.0;20.0;17.0;21.0;25.0;30.0;16.0;8.0;13.0;
26 01021 Elburgo;319.0;18.0;12.0;15.0;12.0;19.0;24.0;30.0;41.0;32.0;26.0;15.0;12.0;10.0;17.0;9.0;11.0;9.0;7.0;
27 01022 Elciego;925.0;31.0;30.0;35.0;53.0;57.0;62.0;67.0;70.0;71.0;56.0;56.0;51.0;48.0;65.0;56.0;53.0;33.0;31.0;
28 01023 Elvillar;363.0;11.0;11.0;19.0;14.0;20.0;23.0;26.0;30.0;22.0;18.0;28.0;19.0;23.0;28.0;23.0;24.0;11.0;13.0;
29 01027 Iruraz-Gauna;431.0;14.0;7.0;17.0;24.0;28.0;26.0;40.0;52.0;33.0;41.0;26.0;25.0;11.0;28.0;30.0;10.0;12.0;7.0;
30 01028 Labastida;1243.0;47.0;63.0;52.0;72.0;98.0;97.0;95.0;114.0;90.0;94.0;85.0;49.0;62.0;73.0;51.0;48.0;34.0;19.0;
31 01030 Lagrán;202.0;1.0;3.0;2.0;8.0;11.0;11.0;6.0;11.0;10.0;16.0;13.0;12.0;14.0;19.0;18.0;26.0;14.0;7.0;

```

las primeras cinco líneas no contienen información de interés ni estructurada de forma homogénea, del mismo modo que la línea 7, así como dos líneas intermedias correspondientes a la división de cada uno de los géneros, o líneas finales a modo de pie de página. Por esta razón, ha sido necesario realizar un paso previo para eliminar de cada fichero en su lectura las líneas no estructuradas, y obtener un *dataframe* conjunto con toda la información en estructura tabular.

De modo análogo al intervalo analizado en el apartado previo, se extrae el código INE de la variable "municipio", realizando las mismas convenciones respecto a los totales provinciales. Tras esto, se realiza una nueva reestructura, de modo que se obtenga una análoga a obtenida para el periodo 2011 en adelante.

	codigo_ine	anio	edad	total	hombres	mujeres
1	01000	1998	0-4	10665	5508	5157
2	01000	1998	10-14	14685	7521	7164
3	01000	1998	15-19	20081	10249	9832
4	01000	1998	20-24	24379	12499	11880
5	01000	1998	25-29	24069	12246	11823
6	01000	1998	30-34	24440	12260	12180
7	01000	1998	35-39	23285	11503	11782
8	01000	1998	40-44	21149	10527	10622
9	01000	1998	45-49	20396	10364	10032
10	01000	1998	5-9	11808	6042	5766
11	01000	1998	50-54	19283	9913	9370
12	01000	1998	55-59	14277	7470	6807
13	01000	1998	60-64	14911	7607	7304
14	01000	1998	65-69	13580	6665	6915
15	01000	1998	70-74	10515	4751	5764
16	01000	1998	75-79	7605	3101	4504
17	01000	1998	80-84	5123	1789	3334
18	01000	1998	85-	4347	1336	3011
19	01000	1999	0-4	10604	5459	5145
20	01000	1999	10-14	14102	7224	6878

La figura anterior presenta un fragmento de dichos datos reestructurados, formados en total por 1.896.084 entradas con 6 variables diferentes.

Una vez obtenida una estructura homogénea de ambos intervalos de años, se procede a su unión, obteniendo así un total de 3.065.472 entradas, con un total de 7 variables, al dividir la variable “codigo\_ine” tanto en “codigo\_provincia” como en “codigo\_municipio”, a través de los primeros dos dígitos y los tres últimos, respectivamente. En la **¡Error! No se encuentra el origen de la referencia.** se presenta un pequeño fragmento de dichos datos finales tras todo el proceso.

codigo_provincia	codigo_municipio	año	edad	total	hombres	mujeres
1 00	000	2011	0-4	2470560	1273539	1197021
2 00	000	2011	10-14	2187263	1124459	1062804
3 00	000	2011	15-19	2267904	1167727	1100177
4 00	000	2011	20-24	2611257	1326358	1284899
5 00	000	2011	25-29	3274267	1662639	1611628
6 00	000	2011	30-34	4059412	2088902	1970510
7 00	000	2011	35-39	4095793	2112291	1983502
8 00	000	2011	40-44	3853445	1967008	1886437
9 00	000	2011	45-49	3580521	1801840	1778681
10 00	000	2011	5-9	2356355	1208743	1147612
11 00	000	2011	50-54	3191434	1587776	1603658
12 00	000	2011	55-59	2677814	1317232	1360582
13 00	000	2011	60-64	2470911	1195059	1275852
14 00	000	2011	65-69	2132734	1013458	1119276
15 00	000	2011	70-74	1794140	817643	976497
16 00	000	2011	75-79	1762589	760014	1002575
17 00	000	2011	80-84	1306383	510779	795604
18 00	000	2011	85-	1097711	347720	749991
19 00	000	2012	0-4	2443703	1259518	1184185
20 00	000	2012	10-14	2214708	1137976	1076732

A través del CNIG (Centro Nacional de Información Geográfica) se ha obtenido el nomenclátor geográfico de municipios y entidades de población, proporcionado en formato ODB (<http://centrodedescargas.cnig.es/CentroDescargas/catalogo.do?Serie=NGMEN>). Dicha base de datos está compuesta por cinco tablas, siendo la denominada “MUNICIPIOS” la de interés para el desarrollo de este proyecto.

Para su procesado, se ha exportado dicha tabla en formato CSV. Posteriormente, se procede a estructurarla y eliminar información innecesaria mediante el software R. En concreto, se ha mantenido la información de identificación de los municipios (“codigo\_provincia”, “codigo\_municipio”), y características como la población total a 2018, la superficie (en hectáreas), el perímetro (en metros) y la altitud (en metros). En la figura siguiente se muestra un fragmento de dichos datos tras el procesado mencionado.

codigo_provincia	codigo_municipio	poblacion	superficie	perimetro	altitud
1 01	001	2887	1994.58	35069	568
2 01	002	10286	9635.28	65134	219
3 01	003	1483	7327.40	41818	333
4 01	004	1831	2728.73	22886	210
5 01	006	219	1294.43	24606	467
6 01	008	982	5740.98	66639	538
7 01	009	1643	6533.64	53712	602
8 01	010	2967	14074.15	64072	306
9 01	011	286	945.97	18610	421
10 01	013	883	9742.34	57432	575
11 01	014	443	3573.00	38711	471
12 01	016	511	13044.29	85973	710
13 01	017	1040	8536.08	53461	578
14 01	018	1753	10206.93	48596	585
15 01	019	181	1251.59	15980	691
16 01	020	381	8403.47	44173	594
17 01	021	603	3209.30	58514	540
18 01	022	1025	1632.31	21825	451
19 01	023	348	1750.22	21190	581
20 01	027	517	4712.86	81139	634

Posteriormente ha sido necesario establecer la relación de los municipios con los códigos. Desde el INE, se proporciona un glosario de los diferentes municipios, asociando a cada uno de ellos sus códigos de provincia y municipio, así como los nombres oficiales de éstos (<http://www.ine.es/daco/daco42/codmun/codmun19/19codmun.xlsx>). Tras su procesado, se obtiene una tabla con un total de 8.168 entradas con 3 variables, incluyendo la información tanto de los municipios como de las diferentes provincias y total nacional

	↑	↻	↻	
	codigo_provincia	codigo_municipio	municipio	
1	1	1	Alegría-Dulanzi	
2	1	2	Amurrio	
3	1	49	Añana	
4	1	3	Aramaio	
5	1	6	Armiñón	
6	1	37	Arraia-Maeztu	
7	1	8	Arrazua-Ubarrundia	
8	1	4	Artziniega	
9	1	9	Asparrena	
10	1	10	Ayala/Aiara	
11	1	11	Baños de Ebro/Mañueta	
12	1	13	Barrundia	
13	1	14	Berantevilla	
14	1	16	Bernedo	
15	1	17	Campezo/Kanpezu	
16	1	21	Elburgo/Burgelu	
17	1	22	Elciego	
18	1	23	Elvillar/Bilar	
19	1	46	Erriberagoitia/Ribera Alta	
20	1	56	Harana/Valle de Arana	

Dicha tabla permitirá ofrecer en los distintos procesos del proyecto información homogénea en cuanto a la nomenclatura de los municipios y provincias nacionales, tal y como lo estipula el INE a través de dichas fuentes.

### **T2.3: Elección de variables influyentes**

Se han diseñado dos ramas independientes que permitan realizar el análisis de la evolución temporal de las pirámides poblacionales de los distintos municipios y provincias del territorio nacional. Por un lado, un modelo que prediga la estructura de la pirámide poblacional a partir de los datos de años previos, sin tener en cuenta la población total (con población relativa para cada tramo de edad y género con respecto al total). Por otro lado, un modelo de regresión que permita predecir la población total del municipio a partir del histórico previo de población disponible. De este modo, mediante la unificación de ambos modelos, se podrá realizar una predicción final de las pirámides poblacionales, generando las formas de las pirámides a partir del primero, y dándole las dimensiones correctas mediante el segundo.

En ambos casos, el entrenamiento de los modelos requiere del histórico de la población previa, incluyendo en ambos casos los datos de cada una de las combinaciones de tramo de edad y género, con respecto a cada uno de los cinco años previos al año en estudio. Dicha información procede de los datos del padrón continuo nacional.

Por un lado, se han de generar los datos correspondientes al padrón continuo nacional con la proporción con respecto al total de dicho caso particular, de modo que sean dichos datos los utilizados para la generación del conjunto de entrenamiento para el modelo de estructura de la pirámide poblacional. En la siguiente figura se presenta un fragmento de éstos.

codigo_provincia	codigo_municipio	año	edad	total	hombres	mujeres
1 00	000	2011	0-4	0.052352918	0.026987194	0.025365724
2 00	000	2011	10-14	0.046349654	0.023828083	0.022521570
3 00	000	2011	15-19	0.048058493	0.024744963	0.023313531
4 00	000	2011	20-24	0.055334387	0.028106466	0.027227921
5 00	000	2011	25-29	0.069384039	0.035232499	0.034151540
6 00	000	2011	30-34	0.086021818	0.044265314	0.041756504
7 00	000	2011	35-39	0.086792757	0.044760944	0.042031813
8 00	000	2011	40-44	0.081657231	0.041682294	0.039974937
9 00	000	2011	45-49	0.075873778	0.038182267	0.037691511
10 00	000	2011	5-9	0.049932833	0.025614121	0.024318712
11 00	000	2011	50-54	0.067628749	0.033646099	0.033982650
12 00	000	2011	55-59	0.056744777	0.027913080	0.028831697
13 00	000	2011	60-64	0.052360356	0.025324147	0.027036208
14 00	000	2011	65-69	0.045194145	0.021475893	0.023718252
15 00	000	2011	70-74	0.038019098	0.017326435	0.020692664
16 00	000	2011	75-79	0.037350510	0.016105235	0.021245275
17 00	000	2011	80-84	0.027683182	0.010823769	0.016859413
18 00	000	2011	85-	0.023261274	0.007368433	0.015892841
19 00	000	2012	0-4	0.051701817	0.026647825	0.025053993
20 00	000	2012	10-14	0.046856933	0.024076341	0.022780592

A la hora del entrenamiento de los modelos, es necesario que dicha información venga resumida por vectores, donde para cada uno de los casos con valor real conocido, se le asocie una serie de atributos utilizados como base de aprendizaje del modelo. De este modo, ha sido necesario generar una nueva tabla de datos para cada uno de los dos modelos a entrenar, partiendo de los datos del padrón continuo.

En particular, para los datos a estudiar de un año concreto, se recogerá la información asociada a los cinco años anteriores, con respecto a las diferentes combinaciones de franjas de edad y género. Es decir, el conjunto inicial de datos a utilizar para el entrenamiento consistirá en un total de  $5 \times 2 \times 18 = 180$  variables. Para codificar dichas variables de forma intuitiva, se han estructurado sus nombres del siguiente modo:

**G\_E\_A**

donde:

- **G** representa el género, y tomará valores "H" (hombre) o "M" (mujer);
- **E** representa la franja de edad y tomará valores: "0-4", "5-9", "10-14", "15-19", "20-24", "25-29", "30-34", "35-39", "40-44", "45-49", "50-54", "55-59", "60-64", "65-69", "70-74", "75-79", "80-84", "85-";

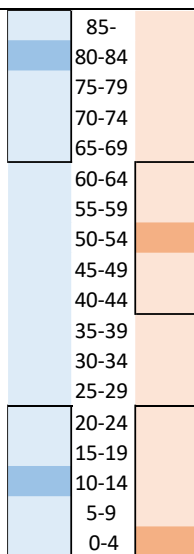
- A representa la antigüedad del año del dato con respecto al año en estudio y tomará valores: 1, 2, 3, 4 o 5.

Dado el gran volumen de datos generables a partir del histórico disponible, se ha restringido su obtención a los años entre 2013 y 2018, obteniendo así un total de 1.755.864 diferentes entradas brutas de entrenamiento. En la siguiente figura se muestra un fragmento de dichos datos de entrenamiento para el caso relativo (recortado tanto por filas como por columnas). El caso para la población total será análogo, pero con respecto a valores totales y no relativos.

año	codigo_ine	genero	edad	real	H_0-4_1	H_0-4_2	H_0-4_3	H_0-4_4	H_0-4_5	H_5-9_1	H_5-9_2
2013	13005	mujeres	10-14	0.0245519657210772	0.027138885	0.026819205	0.027210459	0.026919431	0.025040775	0.029097687	0.028132033
2016	09152	mujeres	65-69	0.0195530726256903	0.008264463	0.014880952	0.012658228	0.020346837	0.022346369	0.019283747	0.011904762
2015	47131	hombres	40-44	0.0153846153846154	0.000000000	0.000000000	0.005128205	0.005347594	0.005434783	0.024390244	0.025252525
2017	37180	mujeres	10-14	0	0.000000000	0.000000000	0.000000000	0.000000000	0.000000000	0.000000000	0.000000000
2014	45004	hombres	25-29	0.0183098591549296	0.027322404	0.023943662	0.030555556	0.024113475	0.018978102	0.020491803	0.026760563
2018	06111	hombres	40-44	0.0273311897106109	0.018126888	0.020926756	0.023916293	0.031818182	0.038167939	0.037764350	0.034379671
2014	26179	hombres	65-69	0.06	0.000000000	0.000000000	0.000000000	0.000000000	0.000000000	0.000000000	0.018518519
2013	38016	mujeres	70-74	0.0303951367781155	0.012091898	0.010942249	0.011742892	0.011320755	0.014312383	0.013301088	0.015197568
2016	37369	mujeres	30-34	0.00699300699300699	0.005882353	0.005780347	0.006451613	0.000000000	0.006578947	0.000000000	0.005780347
2018	23030	mujeres	85-	0.0281014393420151	0.016938519	0.018714910	0.016991819	0.017737003	0.020912548	0.025721455	0.018714910
2015	34140	hombres	55-59	0.0825688073394495	0.000000000	0.000000000	0.000000000	0.000000000	0.009259259	0.000000000	0.000000000
2018	34152	mujeres	40-44	0.0198675496688742	0.012048193	0.017964072	0.018518519	0.018633540	0.000000000	0.018077289	0.011976048
2014	37176	hombres	50-54	0.0700636942675159	0.025316456	0.018867925	0.019108280	0.018867925	0.013071895	0.012658228	0.018867925
2017	40903	mujeres	30-34	0.0380952380952381	0.016260163	0.008333333	0.017094017	0.026315789	0.018348624	0.016260163	0.008333333
2016	18022	hombres	60-64	0.02282769179748	0.040966574	0.037804095	0.037450199	0.037769784	0.037751492	0.032120169	0.033020244
2013	29061	hombres	80-84	0.0135908440629471	0.024899058	0.026973684	0.026744945	0.017167382	0.018518519	0.013458950	0.012500000
2013	47226	mujeres	55-59	0.0351758793969849	0.005025126	0.000000000	0.005524862	0.011764706	0.011560694	0.020100503	0.015873016
2014	36001	hombres	20-24	0.025137741046832	0.012830794	0.015042306	0.013481631	0.014118457	0.013352073	0.015503876	0.016295832
2015	16199	mujeres	5-9	0	0.000000000	0.000000000	0.034482759	0.035714286	0.035714286	0.000000000	0.000000000
2013	43169	mujeres	25-29	0.0364500792393027	0.011400651	0.012678288	0.017214397	0.015923567	0.014262074	0.009771987	0.014263074
2013	23076	mujeres	85-	0.0123711340206186	0.020202020	0.020618557	0.018981019	0.022312373	0.021739130	0.022222222	0.018556701
2016	49071	hombres	75-79	0.0328068043742406	0.002166847	0.006315789	0.003257329	0.002322880	0.002372479	0.009750813	0.010526316
2015	08132	mujeres	60-64	0.0399159663865546	0.023305085	0.022587269	0.022774327	0.018947368	0.023255814	0.021186441	0.016427105
2017	16074	hombres	0-4	0	0.000000000	0.000000000	0.000000000	0.019230769	0.000000000	0.000000000	0.000000000
2017	11031	hombres	80-84	0.00896040483882772	0.032299434	0.031735711	0.030566693	0.029379106	0.028162143	0.029517318	0.030652053
2014	45088	hombres	85-	0.00709876543209877	0.041918176	0.047557425	0.045999352	0.043354430	0.043518519	0.037558685	0.037528308
2018	49242	hombres	55-59	0.0393374741200828	0.006980803	0.007054674	0.005535055	0.003809524	0.003952569	0.015706806	0.014109347
2017	09113	hombres	50-54	0.0357142857142857	0.020766773	0.018242123	0.019966722	0.018803419	0.020761246	0.012779553	0.016583748
2018	37031	mujeres	30-34	0.0161290322580645	0.006928406	0.009638554	0.009803922	0.007575758	0.005102041	0.009237875	0.004819277
2017	16094	hombres	85-	0.0177514792899408	0.031914894	0.046948357	0.028037383	0.010810811	0.000000000	0.031914894	0.032863850

En la obtención del modelo de estructura de la pirámide poblacional, se generarán modelos independientes para cada una de las combinaciones de género y tramo de edad, es decir, 36 modelos diferentes. En función de la combinación seleccionada, se determinará un subconjunto distinto de variables para el entrenamiento de los modelos.

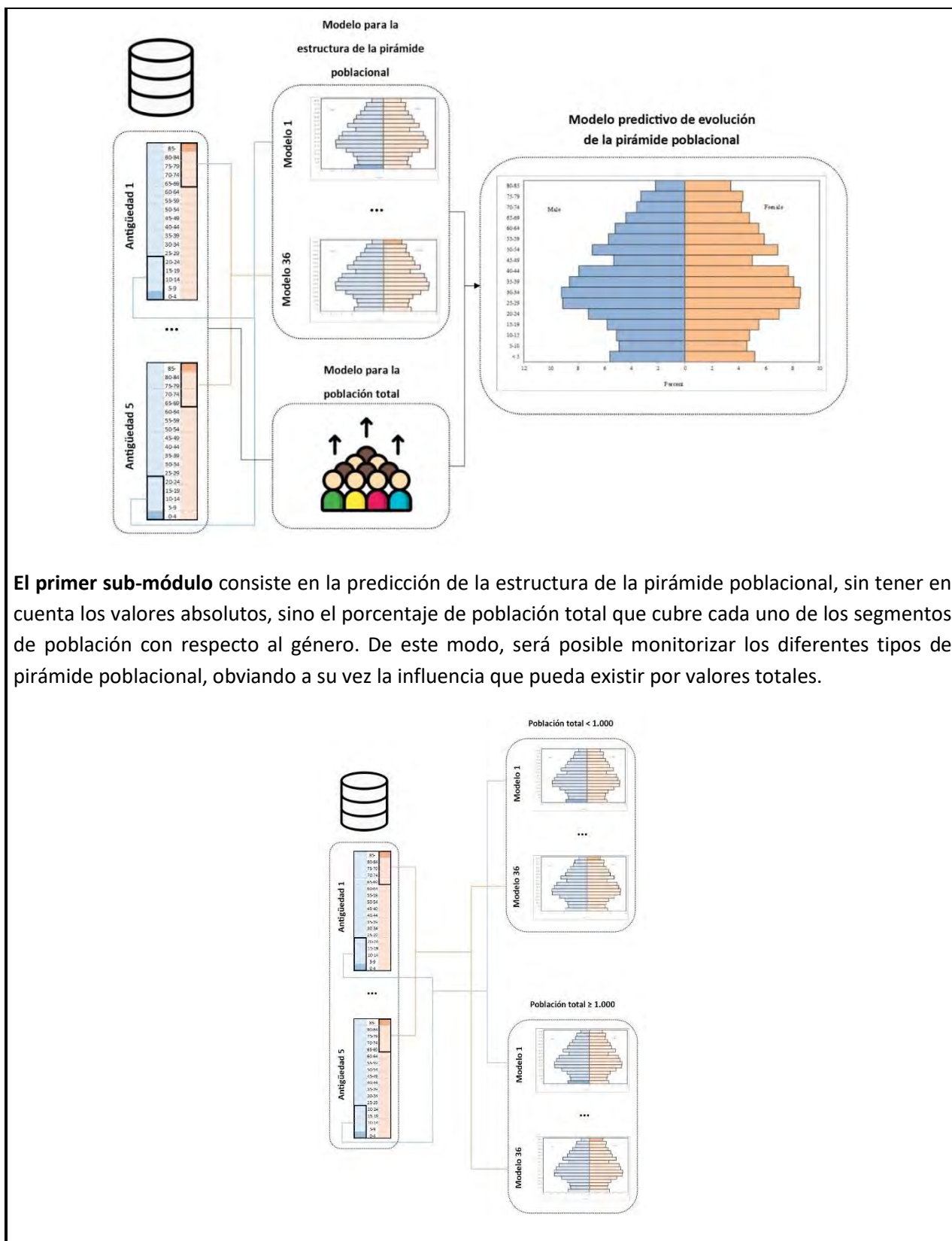
Se seleccionarán aquellas variables relativas a los cinco tramos de edad más cercanos al analizado para el género en estudio y con respecto a los cinco años de antigüedad seleccionados. Es decir, el subconjunto de variables seleccionadas vendrá dado por un total de  $1 \times 5 \times 5 = 25$  atributos. En la siguiente figura se presenta un diagrama con la selección de las variables a utilizar en función del tramo de edad y el género a predecir en cuatro ejemplos diferentes. El caso en estudio está destacado con un color más oscuro, siendo los casos recuadrados los correspondientes a su análisis para los 5 años de antigüedad disponibles. De este modo, el modelo dispondrá de información acerca de los datos alrededor del tramo en estudio para los anteriores años que permita estimar la tendencia.



En el caso del modelo para la predicción de la población total, se han utilizado todas las variables disponibles para los 5 años de antigüedad en estudio, los dos géneros y los 18 tramos de edad, es decir, 180 atributos de entrada.

#### **T2.4: Diseño y desarrollo del sistema de modelización**

El modelo propuesto está diseñado para la predicción de la evolución temporal de las pirámides poblacionales de los distintos municipios y provincias del territorio nacional. Debido a la complejidad que dicho análisis presenta, se ha estructurado su obtención mediante la generación de módulos más sencillos, con una unión final para obtener los resultados finales deseados.



Esta figura presenta el esquema de dicho sub-módulo. Los datos de entrada para el entrenamiento de cada uno de los modelos corresponden a los descritos anteriormente en este informe y que se detallan en el entregable “E2.1: Catálogo de parámetros a considerar en el modelo y variables relevantes”. Del mismo modo, la selección de variables para cada uno de los modelos se corresponde a la explicada en dicho documento.

En concreto, se distinguirán dos casos generales, donde el procedimiento individualizado será análogo. Dichos casos se determinarán por el tamaño poblacional del municipio o provincia en análisis, separando aquellos que no superen los 1.000 habitantes de los que sí.

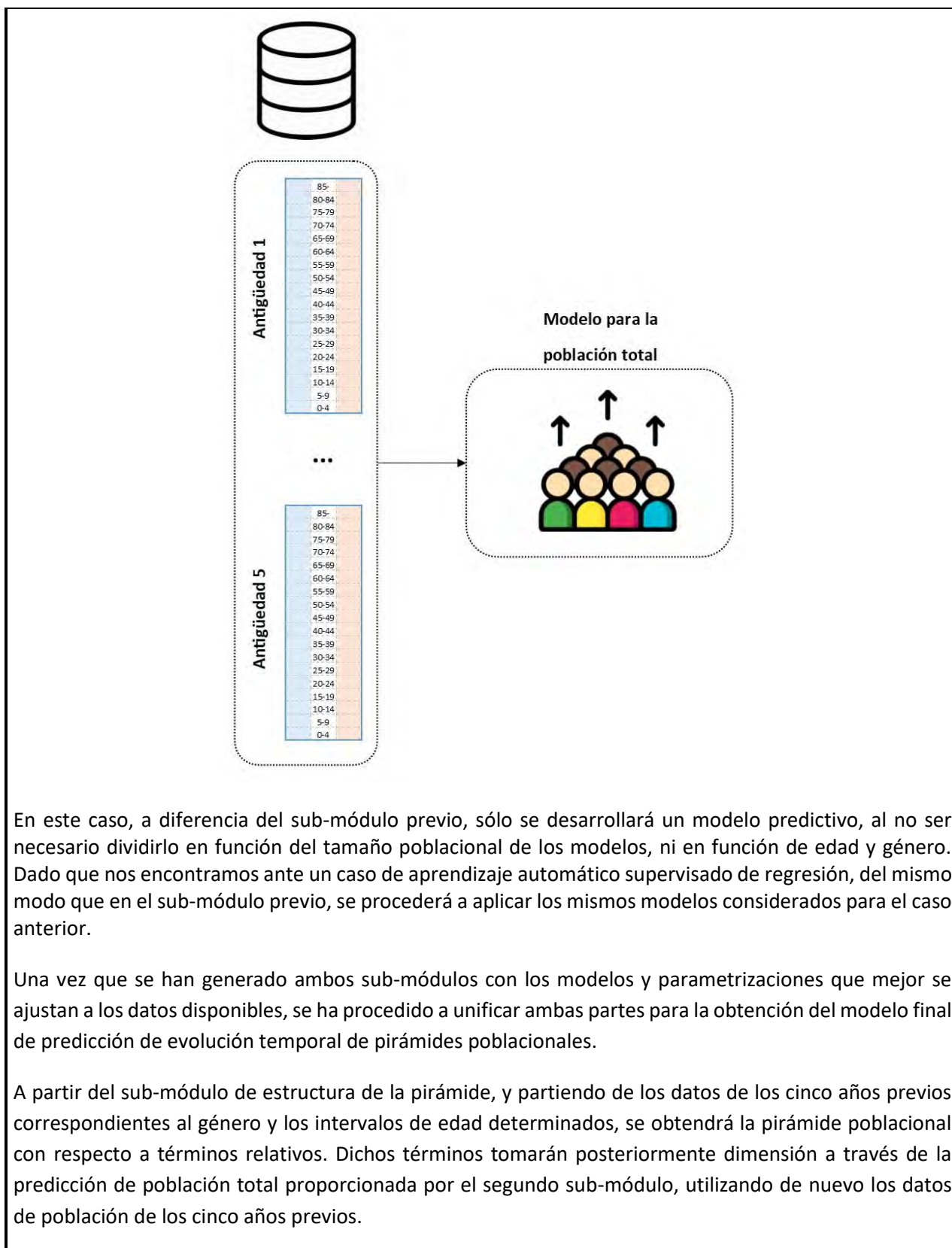
En cada uno de los casos, se generará un modelo independiente para cada una de las combinaciones de tramos de edad y género. De este modo, el sub-módulo de predicción de estructura de la pirámide poblacional estará formado finalmente por un total de  $2 \times 18 \times 2 = 72$  modelos.

Para generarlos, se considerarán diferentes modelos, comparándolos experimentalmente, y realizando un ajuste de parámetros detallado de la tipología seleccionada.

A continuación, se describen los diferentes modelos de regresión considerados para su entrenamiento:

- Regresión lineal. Método de aprendizaje automático supervisado de regresión ampliamente extendido y conocido, siendo uno de los métodos más básicos y que mejor funcionan en situaciones relativamente sencillas.
- SVM (Support Vector Machines). Si bien los modelos SVM son enfocados originalmente para problemas de aprendizaje automático supervisado de clasificación mediante la generación de hiperplanos separadores de espacios para las clases en cuestión, existe una adaptación al caso de regresión. Dicha versión mantiene la base de funcionamiento del modelo, trabajando con los márgenes entre diferentes hiperplanos, y buscando minimizar los errores resultantes a través de la tolerancia del modelo.
- Árboles de regresión. Se trata de modelos basados en la concatenación de diferentes reglas con respecto a las variables independientes. El árbol de regresión seleccionado para este proyecto ha sido el árbol de regresión M5P.
- Random Forest. Modelo basado en la combinación de varios árboles de regresión independientes, de modo que el resultado final se obtenga de la agregación de los resultados de los diferentes árboles que lo componga.
- XGBoost. Método del descenso del gradiente ampliamente aplicado en proyectos recientes por sus buenos resultados. Está basado en el *gradient boosting*, modelo que produce predicciones a través de la unión de modelos de predicción débiles, como pueden ser árboles de decisión.

**El segundo sub-módulo** corresponde al desarrollo de un modelo predictivo para la población total de cada uno de los municipios y provincias. Para ello, el conjunto de variables a utilizar como entrenamiento consiste en todos los datos de población asociados a los cinco años previos del municipio o provincia en estudio, incluyendo la separación asociada a cada uno de los tramos de edad y al género.

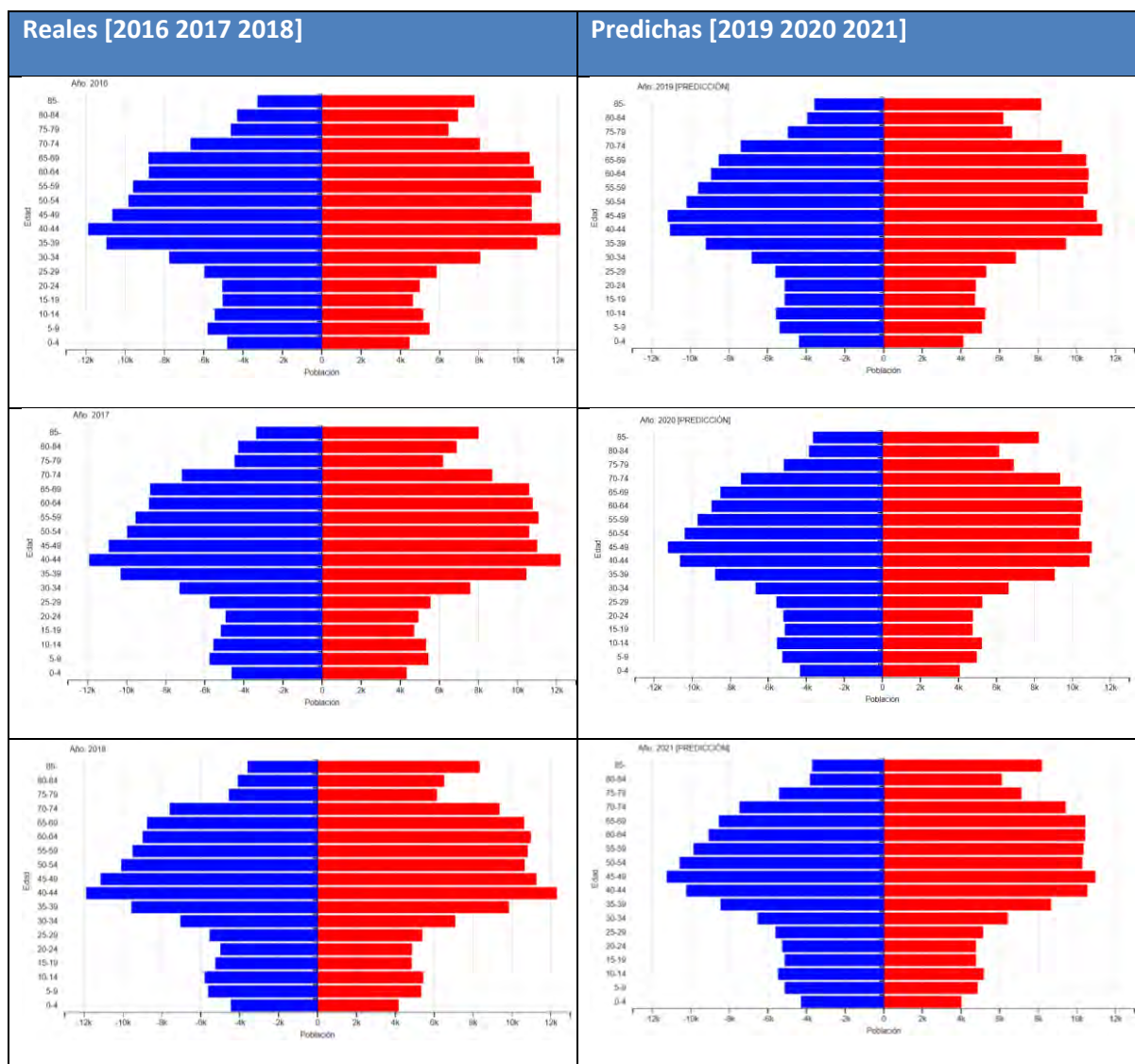


En este caso, a diferencia del sub-módulo previo, sólo se desarrollará un modelo predictivo, al no ser necesario dividirlo en función del tamaño poblacional de los modelos, ni en función de edad y género. Dado que nos encontramos ante un caso de aprendizaje automático supervisado de regresión, del mismo modo que en el sub-módulo previo, se procederá a aplicar los mismos modelos considerados para el caso anterior.

Una vez que se han generado ambos sub-módulos con los modelos y parametrizaciones que mejor se ajustan a los datos disponibles, se ha procedido a unificar ambas partes para la obtención del modelo final de predicción de evolución temporal de pirámides poblacionales.

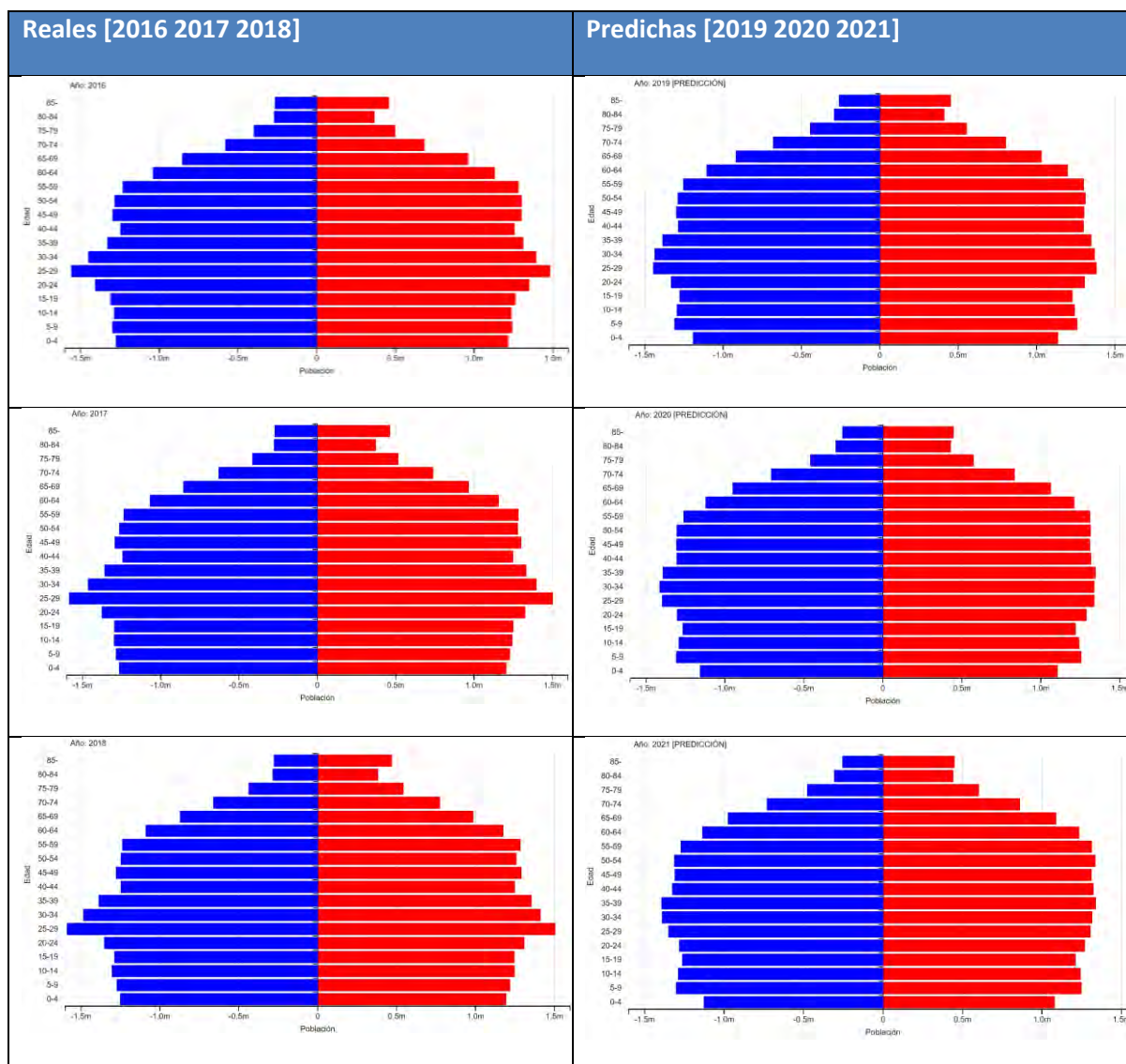
A partir del sub-módulo de estructura de la pirámide, y partiendo de los datos de los cinco años previos correspondientes al género y los intervalos de edad determinados, se obtendrá la pirámide poblacional con respecto a términos relativos. Dichos términos tomarán posteriormente dimensión a través de la predicción de población total proporcionada por el segundo sub-módulo, utilizando de nuevo los datos de población de los cinco años previos.

A modo ilustrativo, se presentan en la siguiente las pirámides poblacionales del municipio de Gijón, mostrando las tres últimas con datos conocidos (2016, 2017, 2018) y las predicciones obtenidas para los tres años consecutivos (2019, 2020, 2021).



Para demostrar su adaptabilidad a otros territorios, se ha aplicado dicho modelo a los diferentes estados de Estados Unidos, a partir de datos proporcionados por parte del *United States Census Bureau*. Los datos obtenidos ofrecen los datos poblacionales para cada uno de los estados, entre los años 2010 y 2018, con respecto a ambos sexos y edades individualizadas desde los 0 años hasta los 84, agrupando los mayores de dicha edad en 85.

De nuevo, a modo ilustrativo, se presentan en la siguiente figura las pirámides poblacionales del estado de California, incluyendo tanto las tres últimas con datos conocidos (2016, 2017, 2018), como las predicciones obtenidas para los tres años consecutivos (2019, 2020, 2021).



### H3: Diseño y desarrollo del prototipo e integración de componentes

El objetivo que se persigue con la ejecución del presente hito es la obtención de un prototipo funcional para el análisis de la dinámica de un territorio. Para ello se realizó un diseño conceptual con el fin de determinar los módulos necesarios para el correcto funcionamiento del prototipo. Posteriormente se diseñó la interfaz de usuario, de forma que sea manejable y fácilmente interpretable. Por último se implementó el prototipo a través de la integración del sistema de modelización y la interfaz de usuario.

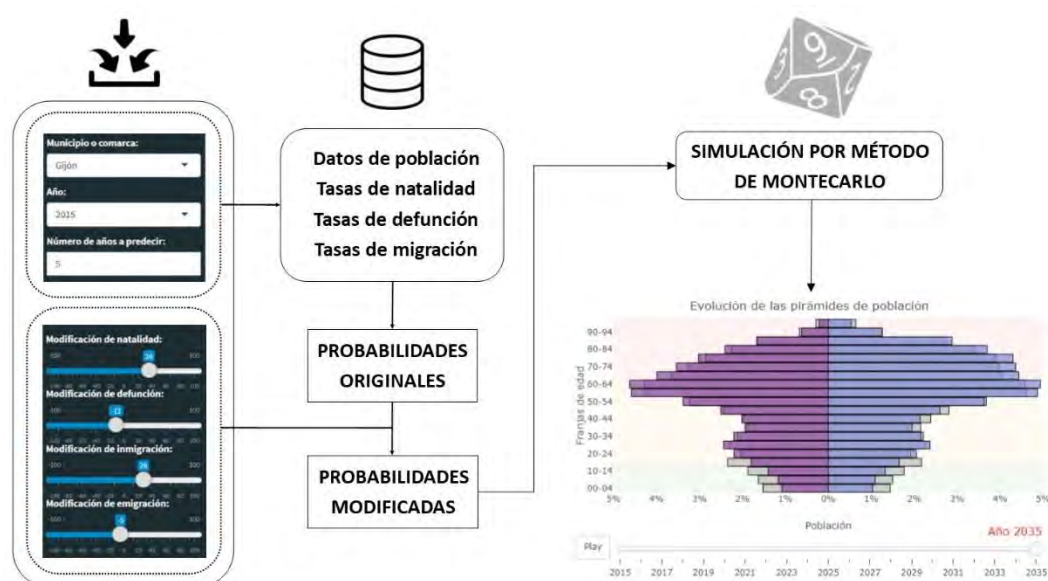
Estos trabajos se dividen en tres tareas que se han ejecutado íntegramente a lo largo de la presente anualidad 2019:

- T3.1: Diseño conceptual del prototipo.
- T3.2: Desarrollo del interfaz de usuario para el prototipo.
- T3.3: Integración del sistema de modelización con el interfaz de usuario.

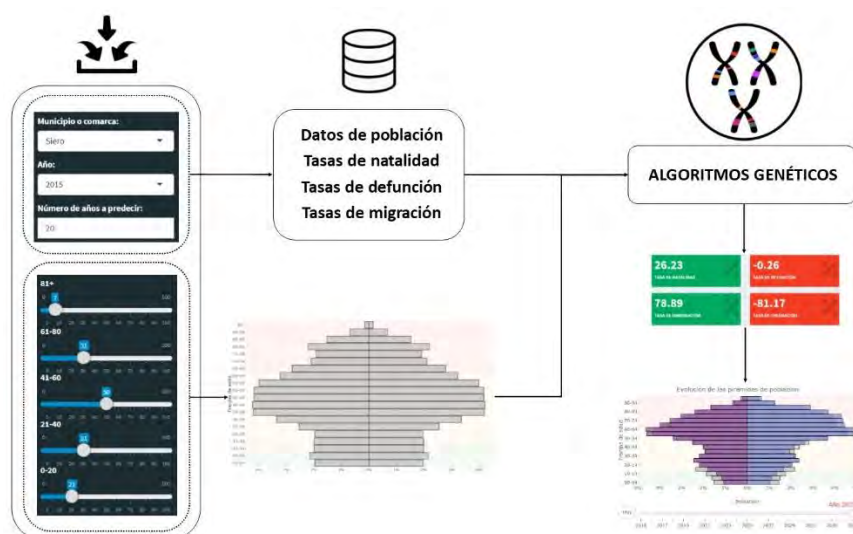
### T3.1: Diseño conceptual del prototipo

Una vez realizado el modelado descrito a lo largo de los entregables correspondientes al Hito 2 “H2: Modelización del sistema formado por el conjunto de variables territoriales”, y utilizando como base para ello los datos recogidos a través de las tareas del Hito 1 “H1: Análisis de los aspectos generales en demografía y población se han diseñado tres herramientas complementarias:

- Herramienta comparativa. Se diseñó un módulo de visualización que facilita la comparación de las pirámides poblacionales de cada municipio, generadas de manera previa mediante métodos de inteligencia artificial. Este módulo facilita el análisis exploratorio de datos, presentando componentes interactivos como un mapa de selección de municipios con información agregada sobre cada región.
- Herramienta de simulación. Se diseñó un módulo capaz de ejecutar simulaciones de la evolución de la población de los diferentes municipios asturianos a través del método de Montecarlo. Dicha simulación está basada en la repetición de un escenario concreto un número determinado de veces, agregando los resultados obtenidos en cada uno de ellas para obtener el escenario final, pudiendo predecir así la evolución que va produciéndose en las variables analizadas a lo largo de los años.



- Herramienta de estimación. Se diseñó un módulo que tiene como objetivo realizar el proceso inverso a la herramienta de simulación. Partiendo de los mismos parámetros de entrada iniciales (municipio/comarca, año inicial, y años a predecir), el usuario introducirá, del modo proporcionado, una estructura de pirámide de población, la cual se buscará replicar con la selección de los cuatro parámetros originales (natalidad, defunción, inmigración, emigración) de forma que la pirámide final resultante sea lo más cercana posible a la proporcionada, mediante la aplicación de algoritmos genéticos.



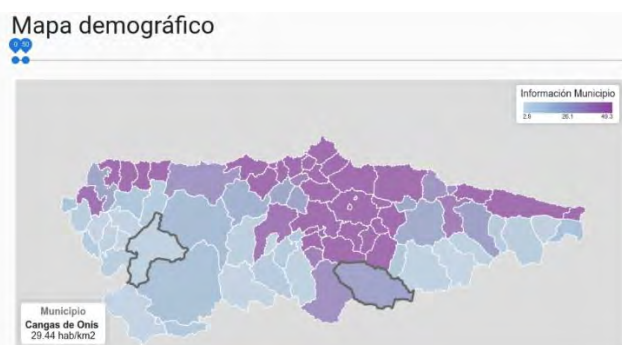
### T3.2: Desarrollo del interfaz de usuario para el prototipo

De acuerdo al diseño de las 3 herramientas descritas en la tarea "T3.1: Diseño conceptual del prototipo", se han generado 3 interfaces de usuario distintos:

#### Herramienta comparativa

Está formada por un componente web para las labores de visualización y por una API remota en la que se centraliza el acceso a los datos, almacenados en una base de datos MySQL.

La interfaz web fue implementada utilizando el framework Vue.js, siendo las principales librerías utilizadas chart.js para la visualización de gráficos y leaflet.js para la gestión de mapas. La API fue implementada en node.js utilizando librerías como express.js.

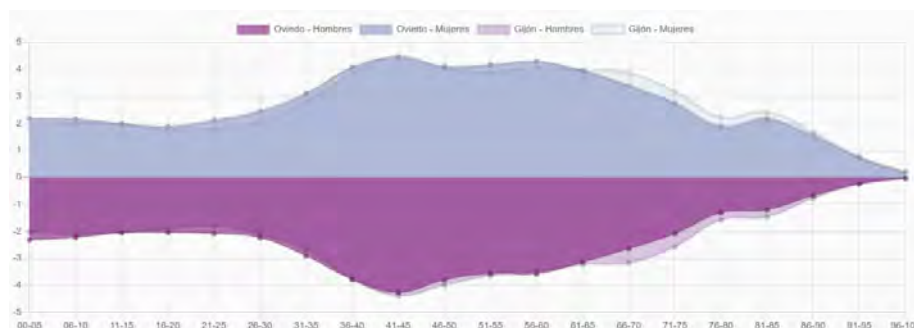


La sección dedicada a los grupos de edad presenta diferentes configuraciones de pirámides de población. El primer gráfico superpone las pirámides de población de dos municipios en un año determinado, siendo estos parámetros configurables mediante el menú o el mapa interactivo. El valor absoluto de cada una de las franjas se muestra al superponer el cursor sobre cada franja.

## Grupos de edad

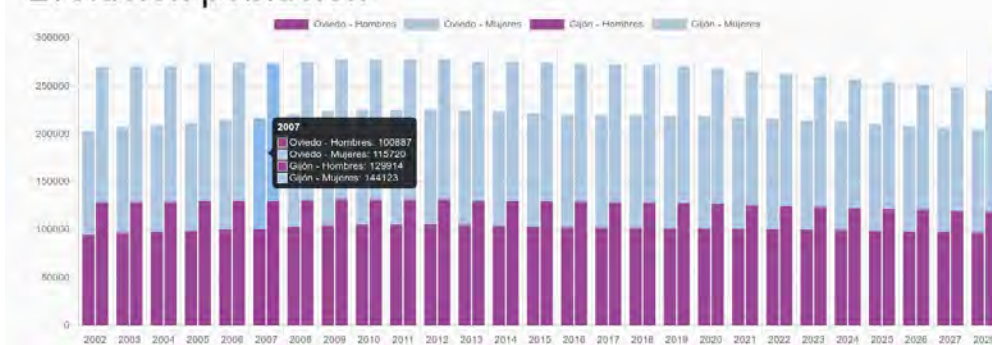


El siguiente gráfico presenta otra configuración de las pirámides de población de dos regiones superponiendo el área bajo la línea de los valores absolutos de cada grupo de edad.



Se implementó un gráfico adicional con la evolución de la población a lo largo del tiempo en términos absolutos, agregando la información de cada grupo de edad e incluyendo las predicciones realizadas por los métodos de inteligencia artificial.

## Evolución población



## Herramienta de simulación

Está formada por una interfaz desarrollada en R e integrada en la herramienta comparativa. La interfaz se divide en la parte de introducción de datos que permiten modificar las probabilidades de ocurrencia de la natalidad, mortalidad, inmigración y emigración según desee el usuario. En concreto, se dispondrá de cuatro sliders que permitan variarlas, tomando éstos, valores entre -100 y 100, siendo dicho valor el porcentaje de modificación de las probabilidades asociadas (es decir, un valor 20 para la slider de natalidad indica que las probabilidades asociadas a dicha variable se verán incrementadas en un 20%).

**Municipio o comarca:**

**Año:**

**Número de años a predecir:**

---

**Modificación de natalidad:**

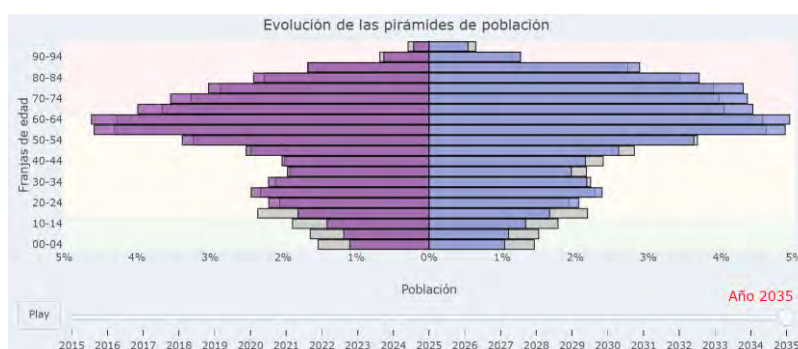
**Modificación de defunción:**

**Modificación de inmigración:**

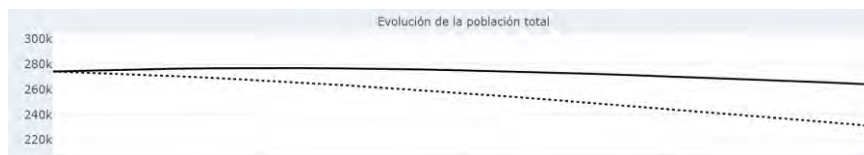
**Modificación de emigración:**

La segunda parte del interfaz se corresponde con la pantalla de visualización de resultados:

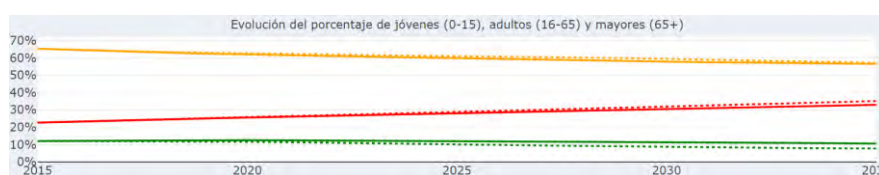
- Evolución de las pirámides de población a través de una gráfica con animación integrada, donde se superponen la evolución de la pirámide generada con las probabilidades sin modificar (en color) y la pirámide generada con las probabilidades modificadas (en gris).



- Serie temporal con la evolución de la población total tanto en el caso modificado (línea continua) como en el caso original (línea discontinua).



- Serie temporal con la evolución del porcentaje de población por franjas amplias de edad, 0-15 en verde, 16-65 en naranja, 66 en adelante en rojo, para el caso modificado (línea continua) y para el caso original (línea discontinua).



- Tabla resumen de los resultados, con comparación con datos reales y sus errores asociados en caso de existir simulación de datos conocidos.

Año	Pob. sin modificar	Pob. modificada	Pob. real	Edad sin modificar	Edad modificada	Edad real	Error sin modificar	Error modificada
2015	274290	274290	274290	46.59	46.59	46.59	0.00%	0.00%
2016	273193	275189	273422	46.96	46.85	46.83	0.75%	1.09%
2017	271982	275953	272365	47.32	47.12	47.09	1.07%	1.83%
2018	270572	276533		47.69	47.38			
2019	269067	276860		48.06	47.64			
2020	267358	277034		48.43	47.91			
2021	265599	276947		48.8	48.18			
2022	263590	276761		49.16	48.44			
2023	261467	276435		49.52	48.7			
2024	259321	275845		49.89	48.97			

### Herramienta de estimación

Está formada por una interfaz desarrollada en R e integrada en la herramienta comparativa. La interfaz se divide en la parte de introducción de datos que permiten que introducir todos los valores de la pirámide resultaría excesivamente costoso para el usuario, se realiza dicha introducción mediante cinco sliders, correspondientes a cinco intervalos de edad (0-20, 21-40, 41-60, 61-80, 81+). Cuanto mayor sea el valor asignado a la slider, más peso tendrá dicho intervalo en la pirámide (no es necesario que las cinco sliders sumen un total de 100%, ya que se realiza una normalización posterior que evite dicho proceso por parte del usuario).

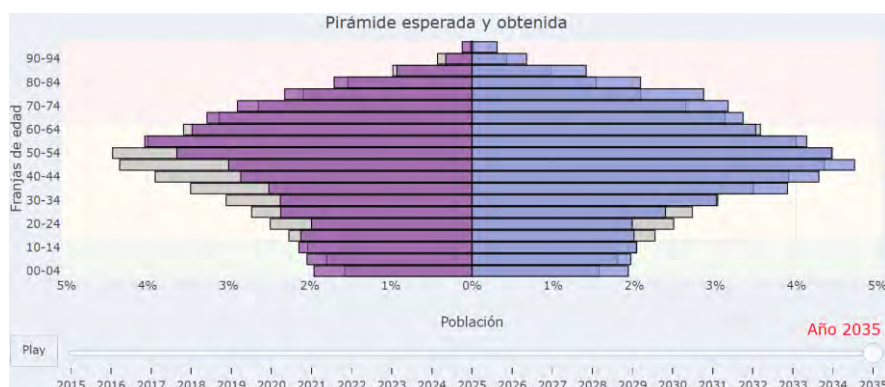


La segunda parte del interfaz se corresponde con la pantalla de visualización de resultados:

- Parámetros determinados como mejores por parte del algoritmo genético aplicado, dando los valores porcentuales de modificación de cada uno de ellos.



- Evolución de la pirámide de población a través de una gráfica con animación integrada, donde se superponen la evolución de la pirámide generada con las probabilidades obtenidas (en color) y la pirámide esperada de forma estática (en gris).



- Tablas con las probabilidades iniciales asociadas a cada parámetro por franja de edad y género (salvo en natalidad), así como las probabilidades modificadas por los parámetros obtenidos.

Probabilidades de defunción				
Show 10 entries		Search: <input type="text"/>		
Edad	Género	Probabilidad original	Probabilidad modificada	
0	Hombres	0.00133	0.00133	
0	Mujeres	0.00107	0.00107	
1	Hombres	0	0	
1	Mujeres	0	0	
2	Hombres	0	0	
2	Mujeres	0	0	
3	Hombres	0	0	
3	Mujeres	0	0	
4	Hombres	0	0	
4	Mujeres	0	0	

Showing 1 to 10 of 202 entries. Previous 1 2 3 4 5 ... 21 Next

### T3.3: Integración del sistema de modelización con el interfaz de usuario

Las interfaces de usuario se han ido mostrando en la tarea anterior en la descripción de las diferentes herramientas que integran STEMO. Por ello en esta tarea se resume cómo ha sido la integración del sistema de modelización con dichas interfaces.

La integración del sistema de modelización con cada uno de los interfaces de usuario ha sido diferente en cada caso:

#### **Herramienta comparativa**

Se entrenó un sistema de “*machine learning*” con los datos recogidos en el “*Hito 2: Modelización del sistema formado por el conjunto de variables territoriales*”, integrando a su vez datos históricos de la evolución de la población con aquellas predicciones generadas mediante los desarrollos de visión artificial.

#### **Herramienta de simulación**

Esta herramienta se basa en la aplicación del método de Montecarlo aplicado a un escenario determinado, marcado por:

- Municipio o comarca.
- Año de inicio.
- Años a simular.

A partir de estos datos, se recoge la información asociada a cuatro parámetros determinantes en la evolución de la población del municipio en análisis:

- Tasa de natalidad.
- Tasa de defunción.

- Tasa de inmigración.
- Tasa de emigración.

Dichos datos se obtienen a través de diferentes conjuntos de datos disponibles de forma abierta a través del correspondiente instituto estadístico, ya sea del INE (Instituto Nacional de Estadística) o de la SADEI (Sociedad Asturiana de Estudios Económicos e Industriales). A partir de dicha información, extrayendo la correspondiente al escenario estudiado (municipio, año inicial), se genera un conjunto de probabilidades asociadas a cada una de las cuatro tasas, asociadas a cada uno de las edades entre 0 y 100 años, y para ambos géneros. Además, se complementa con una probabilidad asociada al género de los nuevos nacimientos en función de la natalidad conocida de años previos en dicho municipio.

edad	genero	nacimiento	muerte	inmigracion	emigracion
0	hombres	0.000000000	1.331394e-03	0.012665133	0.010516870
0	mujeres	0.000000000	1.069175e-03	0.012736598	0.010294883
1	hombres	0.000000000	0.000000e+00	0.012665133	0.010516870
1	mujeres	0.000000000	0.000000e+00	0.012736598	0.010294883
2	hombres	0.000000000	0.000000e+00	0.012665133	0.010516870
2	mujeres	0.000000000	0.000000e+00	0.012736598	0.010294883
3	hombres	0.000000000	0.000000e+00	0.012665133	0.010516870
3	mujeres	0.000000000	0.000000e+00	0.012736598	0.010294883
4	hombres	0.000000000	0.000000e+00	0.012665133	0.010516870
4	mujeres	0.000000000	0.000000e+00	0.012736598	0.010294883
5	hombres	0.000000000	0.000000e+00	0.009397254	0.008385342
5	mujeres	0.000000000	0.000000e+00	0.009615959	0.008456174
6	hombres	0.000000000	0.000000e+00	0.009397254	0.008385342
6	mujeres	0.000000000	0.000000e+00	0.009615959	0.008456174
7	hombres	0.000000000	0.000000e+00	0.009397254	0.008385342
7	mujeres	0.000000000	0.000000e+00	0.009615959	0.008456174
8	hombres	0.000000000	0.000000e+00	0.009397254	0.008385342
8	mujeres	0.000000000	0.000000e+00	0.009615959	0.008456174
9	hombres	0.000000000	0.000000e+00	0.009397254	0.008385342
9	mujeres	0.000000000	0.000000e+00	0.009615959	0.008456174
10	hombres	0.000000000	0.000000e+00	0.008217774	0.007265753
10	mujeres	0.000000000	0.000000e+00	0.008285337	0.007184065
11	hombres	0.000000000	0.000000e+00	0.008217774	0.007265753
11	mujeres	0.000000000	0.000000e+00	0.008285337	0.007184065

### Herramienta de estimación

Esta herramienta se basa en la aplicación de algoritmos genéticos y se ha recurrido a los mismos datos de entrada iniciales previamente descritos para la herramienta de simulación. Además, se accede a los mismos datos públicos para la generación de las probabilidades iniciales (provenientes igualmente de INE y SADEI).

### JUSTIFICACIÓN 2020

Durante el periodo que cubre este documento, se ha ejecutado en su totalidad el Hito 4. A continuación, se describen las tareas realizadas en este hito de trabajo y los resultados conseguidos.

#### **Hito 4: Desarrollo del sistema completo y prueba en entorno piloto**

El objetivo de este hito consiste en parametrizar los procesos, actividades y dinámicas asociadas al territorio con el fin de identificar los escenarios y variables que tienen incidencia sobre el proceso, con el fin de crear un sistema de modelización que sea capaz de integrar todos estos parámetros. Este hito se divide en cuatro tareas, la primera de las cuales se finalizó en la anualidad 2018, terminándose los trabajos del resto de ellas en esta anualidad:

- T4.1: Descripción del proceso concreto de dinámica poblacional y socioeconómica.
- T4.2: Parametrización de acciones dentro de flujo de actuaciones en caso piloto.
- T4.3: Ejecución del prototipo de modelización con los datos del entorno piloto.
- T4.4: Análisis de resultados sobre entorno piloto y reajuste de parametrización.

##### **T4.1: Descripción del proceso concreto de dinámica poblacional y socioeconómica.**

Mediante los modelos realizados a lo largo del Hito 2 “H2: Modelización del sistema formado por el conjunto de variables territoriales”, basados a su vez en los datos iniciales recabados y analizados a lo largo del primer hito “H1: Análisis de los aspectos generales en demografía y población”, se han desarrollado diferentes herramientas tanto de visualización como de análisis más avanzado, enmarcadas dentro del hito 3 “H3: Diseño y desarrollo del prototipo e integración de componentes”.

Así, a lo largo de este último hito, “H4: Desarrollo del sistema completo y prueba en entorno piloto”, se ha procedido a la aplicación de dichas herramientas a un caso de uso particular, en concreto, a los municipios asturianos de mayor población, analizándolos de forma más pormenorizada, recurriendo a unidades de población más pequeñas, tales como distritos.

A lo largo de esta tarea, se presenta un informe de los diferentes datos utilizados para dicho caso de uso, incluyendo la información tanto demográfica como territorial que los componen.

##### **DATOS UTILIZADOS**

Tal y como se ha indicado previamente, en el hito cuarto “H4: Desarrollo del sistema completo y prueba del entorno piloto” se ha procedido a la aplicación de las herramientas desarrolladas a lo largo del proyecto a un caso de uso concreto.

En particular, el caso de uso seleccionado está enmarcado en la región de Asturias, abordando los municipios más poblados de la región, y aplicando los enfoques respecto a unidades de población más pequeñas que las municipales. En concreto, se han recabado los datos a nivel de sección censal, si bien posteriormente se han agrupado a nivel distrito.

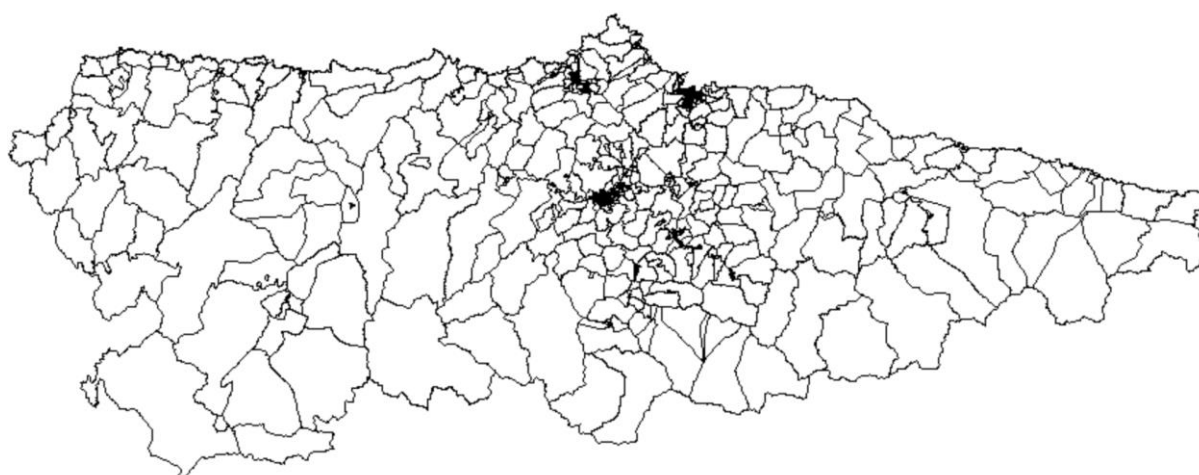
A lo largo de los siguientes subapartados procedemos a detallar los diferentes datos que se han utilizado.

### Cartografía de secciones censales

El primer conjunto de datos necesario para los análisis posteriores es la cartografía de las secciones censales de los municipios asturianos. Para ello, se ha recurrido al INE, el cual proporciona dicha información de forma detallada mediante ficheros *shapefile*.

Dichos datos son actualizados anualmente, siendo los disponibles del año 2019 los que se pueden descargar en el siguiente enlace: [https://www.ine.es/prodyser/cartografia/seccionado\\_2019.zip](https://www.ine.es/prodyser/cartografia/seccionado_2019.zip). Dicho fichero comprimido contiene los diferentes componentes de un archivo *shapefile*, el cual es posible leer con diferentes herramientas. Este fichero contiene información tanto descriptiva de las diferentes secciones censales, como gráfica.

De dicha información resulta relevante, de cara al proyecto, la información asociada a la identificación de cada sección censal, junto a la superficie en metros cuadrados. Por otro lado, también proporciona información gráfica de dichos territorios, mostrándose las 850 secciones censales asturianas gráficamente en la Figura 1.



**Figura 1 – Cartografía de las secciones censales asturianas.**

Como se puede observar, la mayor densidad de éstas se encuentra en los núcleos de población más densos: Oviedo, Gijón y Avilés.

### Población por sección censal

Por otro lado, dado que el análisis se hará de forma pormenorizada a nivel sección censal, será necesario disponer de información de población de dichos territorios, abordándolos a través de población tanto por género como por franja de edad.

Dicha información es proporcionada anualmente por el INE. En concreto, se puede obtener mediante el siguiente enlace los datos más recientes correspondientes a 2019: [https://www.ine.es/jaxi/files/\\_px/es/xls/t20/e245/p07/a2019/l0/3301.xls?nocab=1](https://www.ine.es/jaxi/files/_px/es/xls/t20/e245/p07/a2019/l0/3301.xls?nocab=1). Sin embargo, está

disponible con mayor antigüedad, habiendo seleccionado para nuestro proyecto los datos desde 2.011 en adelante.

Estos datos proporcionan la información con respecto a tres factores: el código de la sección censal, la franja de edad en grupos de edad quinquenales, y el género.

Si bien existe información previa al 2.011, no dispone de la misma partición de edad que la actual, por lo que se ha procedido a descartarlos. En concreto, los datos hasta 2.010 agrupan información de edad de 85 años en adelante, mientras que los más recientes lo hace a partir de los 100 años.

#### Población por municipio

De cara a distintos pasos de procesado posterior, ha sido necesario recurrir de nuevo a la población por municipio, previamente descrita a lo largo del entregable “E2.1: Catálogo de parámetros a considerar en el modelo y variables relevantes”. Dicha información ha sido procesada en hitos previos de este proyecto.

#### Tasas de natalidad por municipio

Tal y como se indica en el hito previo en el entregable “E3.1: Prototipo de modelización del territorio”, se recurren a las tasas de natalidad (y las analizadas en los apartados posteriores) para el desarrollo de la herramienta de simulación correspondiente. Dichos datos son proporcionados por la SADEI anualmente, pudiendo acceder a los más recientes, correspondientes al año 2.018, a través del enlace <http://www.sadei.es/datos/indicetematico/cuadros/02/06/04/06/02060406122018c.xls>. Se presenta el número de nacimientos según el municipio y el grupo de edad de la madre.

#### Tasas de mortalidad por municipio

De modo análogo a la tasa previa, se procede a la recopilación de la información asociada a la mortalidad de la población por municipio. En este caso, los datos de 2.018 se pueden obtener en el enlace <http://www.sadei.es/datos/indicetematico/cuadros/02/06/08/020608102018c.xls>, de nuevo a través de la SADEI.

En este caso, dicho fichero Excel está compuesto por tres páginas, una para el número total, y las otras dos correspondientes a las defunciones de hombres y mujeres por separado. Cada una de ellas tiene dicho valor con respecto al municipio y a la franja de edad.

#### Tasas de migración por municipio

Finalmente, las últimas tasas previamente utilizadas y necesarias en este hito se corresponden con las tasas migratorias por municipio. En este caso, estos datos son proporcionados por el INE, separando las tasas de altas y bajas en dos ficheros independientes, pudiendo obtenerse los correspondientes a 2.018 en los enlaces [http://www.ine.es/jaxi/files/px/es/xls/t20/p307/a2018/I0/a4\\_1.xls?nocab=1](http://www.ine.es/jaxi/files/px/es/xls/t20/p307/a2018/I0/a4_1.xls?nocab=1) y [http://www.ine.es/jaxi/files/px/es/xls/t20/p307/a2018/I0/a4\\_7.xls?nocab=1](http://www.ine.es/jaxi/files/px/es/xls/t20/p307/a2018/I0/a4_7.xls?nocab=1), respectivamente. En este caso, se proporcionan conjuntamente a nivel nacional, pasando posteriormente a la utilización exclusiva de los datos de los municipios asturianos.

#### **T4.2: Parametrización de acciones dentro de flujo de actuaciones en caso piloto.**

En esta segunda tarea, se incluyen diferentes pasos de preprocesado necesarios a partir de los datos recabados tal y como se ha realizado a lo largo de la tarea previa, de modo que sea posible obtener la información necesaria para la aplicación de las herramientas. En concreto, se tendrán en cuenta cuatro variables determinantes a parametrizar en el proceso: natalidad, mortalidad, inmigración y emigración.

##### PREPROCESADO

A lo largo de este apartado se detallarán las diferentes fases de preprocesado que permiten adecuar los datos previamente mencionados a la estructura necesaria para la aplicación de los desarrollos generados a lo largo de los hitos previos de este proyecto.

##### Reestructura de población por sección censal

Dados los datos presentados en la tarea previa, éstos han de modificarse para generar una estructura tabular única, adecuada para los análisis posteriores. Por esta razón, se ha procedido a modificarlos y unificar todos los datos desde el año 2.011 hasta 2.019. Se tratan de datos asociados a 164.304 entradas y 7 variables.

##### Población por distrito año a año

Tal y como se puede observar en el subapartado previo, la población por sección censal se proporciona en origen por intervalos de edad quinquenales. Sin embargo, para el proceso de simulación presentado en el hito 3 de este proyecto, es necesario disponer de dicha información año a año, y transformada a información de distrito.

En un primer nivel, se ha procedido a agrupar la información de nivel sección censal a nivel distrito, de modo que los conjuntos poblacionales sean representativos de un mayor número de personas, fusionando así conjuntos demasiado reducidos. Se trata de un proceso sencillo, al tener la codificación de cada sección censal implícita la de su distrito, correspondiente a sus primeras 7 cifras en lugar de las 10 de los datos de sección censal.

Posteriormente, para generar dichos datos año a año, es necesario procesarlos de forma que, a partir de la información del grupo quinquenal, sea posible extrapolar la información individual de cada uno de los cinco años que representa.

Para ello, se procederá a trasladar la distribución de cada tramo de edad a sus individuales teniendo en cuenta la información de su municipio correspondiente, y trasladando dicha información proporcionalmente. Dicha proporcionalidad se obtendrá teniendo en cuenta la distribución de dichos tramos individualmente, con respecto a los diferentes municipios asturianos con población superior total a 10.000 habitantes. Dicha tabla presenta la información necesaria para trasladar cada intervalo a los años individuales. De este modo, se aplicará dichas proporcionalidades, obteniendo los datos estimados año a año. En total, están compuestos por 24.780 entradas con 7 variables asociadas.

#### Probabilidades asociadas a las tasas por municipio año a año

Una vez obtenidas las tasas de natalidad, mortalidad y movimientos migratorios, se procede a generar una tabla de probabilidades asociadas a cada una de dichas cuatro tasas, y con respecto tanto al género como la edad de la población.

Tal y como se indica en los apartados correspondientes, dicha información es proporcionada por tramos de edad, por lo que se lleva a cabo una imputación año a año a través de la misma metodología detallada en el entregable correspondiente al hito 3, "E3.1: Prototipo de modelización del territorio". Dichos datos constan de un total de 127.664 entradas con 8 variables asociadas.

#### Probabilidades asociadas a las tasas por distrito año a año

Una vez obtenidas las probabilidades por municipio, ha de obtenerse para los objetivos de este proyecto dicha información asociada a la unidad poblacional en estudio. En este caso, a nivel de distrito. Por esa razón, será necesario generar una metodología que permita transformar la información municipal en información de distrito para las cuatro tasas analizadas.

Para ello, se ha llevado a cabo un procedimiento que asocia a la información por distrito una modificación de los datos municipales correspondientes. Dicha variación se genera a través de la correspondencia entre las densidades de población del distrito y del municipio correspondiente. Para ello, se consideran las correspondencias entre la provincia de Asturias y sus 78 municipios con respecto a su densidad, seleccionando para cada distrito la correspondencia asociada al municipio con una proporción de densidades más cercana a la propia entre el distrito y su municipio.

Formalmente, dado un distrito y su municipio,  $DIS$  y  $MUN$ , y sus correspondientes densidades poblacionales,  $d_{DIS}$  y  $d_{MUN}$ , se selecciona el municipio más afín del siguiente modo:

$$\arg \min_{i \in \{1, \dots, 78\}} \left| \frac{d_{DIS}}{d_{MUN}} - \frac{d_{MUN_i}}{d_{AST}} \right|,$$

donde  $d_{AST}$  representa la densidad poblacional de Asturias, y  $d_{MUN_i}$  la densidad poblacional del municipio  $i$ -ésimo de entre los 78 que componen Asturias.

Así, una vez seleccionado dicho municipio, se utilizará su proporcionalidad en las tasas analizadas de Asturias a éste, trasladándola en el paso del distrito analizado y su municipio asociado.

De este modo, se obtienen las probabilidades asociadas a los datos disponibles para dichas tasas para cada uno de los distritos. En concreto, está formado por un total de 210.080 entradas con 9 variables.

#### **T4.3: Ejecución del prototipo de modelización con los datos del entorno piloto.**

##### HERRAMIENTA DE SIMULACIÓN

Tal y como se detalla en el entregable del hito 3 “E3.1: Prototipo de modelización del territorio”, se ha desarrollado una herramienta de simulación que permite obtener, a partir de diferentes argumentos de entrada, cuál será la evolución de la población de una región, tanto siguiendo las tendencias actuales, como mediante modificaciones especificadas por el usuario. En la Figura 2 se presenta un esquema de su funcionamiento, con ciertas modificaciones respecto al correspondiente al caso general del hito 3.

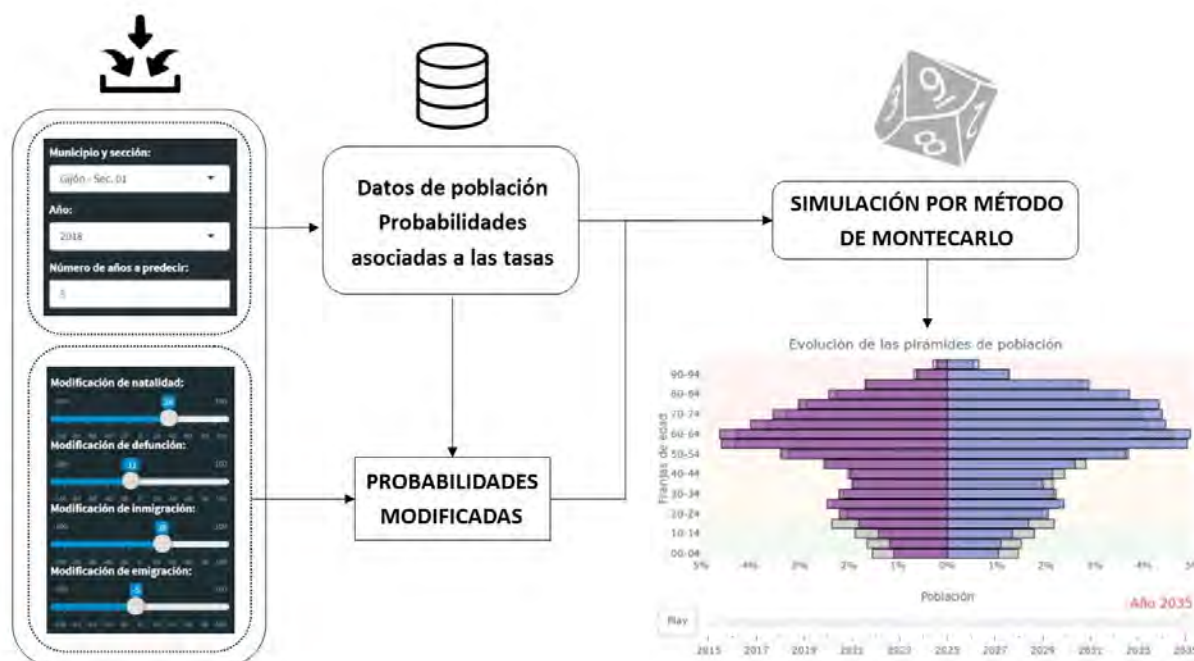


Figura 2 – Esquema de la herramienta de simulación adaptada a distritos.

Así, se han integrado las nuevas fuentes de datos en la herramienta, de modo que sea posible aplicarlo para los distritos de los diferentes municipios asturianos.

Tal y como se indica en las tareas previas, los datos descargados tienen en común información entre los años 2.011 y 2.018, por lo que la herramienta permitirá tomar como punto de partida cualquiera de dichos años.

#### Errores individuales

De cara al contraste de la bondad de la herramienta, se ha procedido a generar las predicciones de todos los distritos asturianos, partiendo del año 2.011, y prediciendo hasta 2.018, es decir, un total de 7 años. En la Figura 3 se presenta un fragmento de dichos errores.

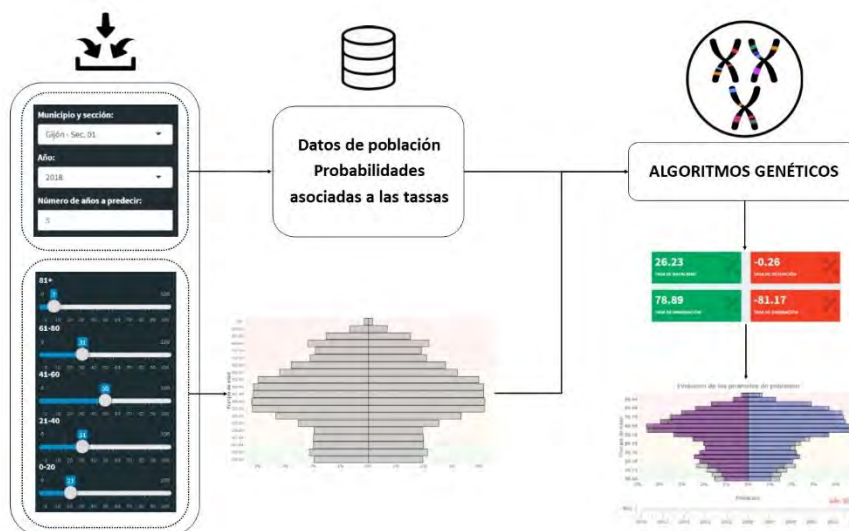
cod_sc	año	edad	genero	value.x	value.y	error_abs	error_rel	cod_municipio
3300401	2012	50	mujeres	740	713	27	0.036486486	4
3300401	2013	50	mujeres	737	715	22	0.029850746	4
3300401	2017	55	mujeres	736	711	25	0.033967391	4
3300401	2012	51	mujeres	734	726	8	0.010899183	4
3300401	2018	55	mujeres	734	711	23	0.031335150	4
3300401	2012	52	mujeres	731	724	7	0.009575923	4
3300401	2016	55	mujeres	730	722	8	0.010958904	4
3300401	2013	51	mujeres	730	711	19	0.026027397	4
3300401	2011	50	mujeres	728	728	0	0.000000000	4
3300401	2013	52	mujeres	727	723	4	0.005502063	4
3300401	2017	56	mujeres	727	720	7	0.009628611	4
3300401	2012	53	mujeres	726	719	7	0.009641873	4
3300401	2018	56	mujeres	724	707	17	0.023480663	4
3300401	2015	50	mujeres	723	713	10	0.013831259	4
3300401	2011	51	mujeres	722	722	0	0.000000000	4
3300401	2013	53	mujeres	722	722	0	0.000000000	4
3300401	2014	50	mujeres	721	709	12	0.016643551	4
3300401	2016	56	mujeres	720	718	2	0.002777778	4
3300401	2011	52	mujeres	719	719	0	0.000000000	4
3300401	2011	45	mujeres	717	717	0	0.000000000	4
3300401	2016	50	mujeres	717	716	1	0.001394700	4

Figura 3 – Fragmento de los errores generados para todos los distritos asturianos sin agregar.

En el caso del error absoluto, se obtiene a través de la diferencia absoluta entre el valor real y el estimado. En el caso del error relativo, se genera mediante el cociente del error absoluto y el valor real (considerando las excepciones donde el valor real sea nulo). De forma desagregada, dichos errores no son suficientemente informativos, por lo que se procede a agregar en un primer nivel respecto a diferentes parámetros, correspondiente a la siguiente tarea.

### HERRAMIENTA DE ESTIMACIÓN

La segunda herramienta desarrollada a lo largo de este proyecto, y a la cual se le ha dado aplicación a lo largo de este caso de uso es la correspondiente a la herramienta de estimación basada en la herramienta de simulación, utilizando algoritmos genéticos para tal fin. En la Figura 4 se presenta un esquema de dicha herramienta adaptado al caso de uso concreto.



**Figura 4 – Esquema de la herramienta de estimación adaptada a distritos.**

Toda la estructura de la aplicación es idéntica a la explicada a lo largo del hito 3 “H3: Diseño y desarrollo del prototipo e integración de componentes”, en el entregable “E3.1: Prototipo de modelización del territorio”. En este caso, se ha modificado de modo que el proceso se genere contra distritos asturianos en lugar de municipios, basándose en la herramienta de simulación adaptada en este hito, y explicada a lo largo del subapartado previo.

Dado que el sistema busca ajustar las modificaciones de las tasas iniciales en función de las características de la pirámide deseada, dadas como argumento de entrada, pueden existir distintas soluciones que lleven a una pirámide con la misma o parecida semejanza. Por esta razón, no es posible contrastar formalmente los resultados, si bien es posible ver la cercanía de los resultados entre lo esperado y lo obtenido.

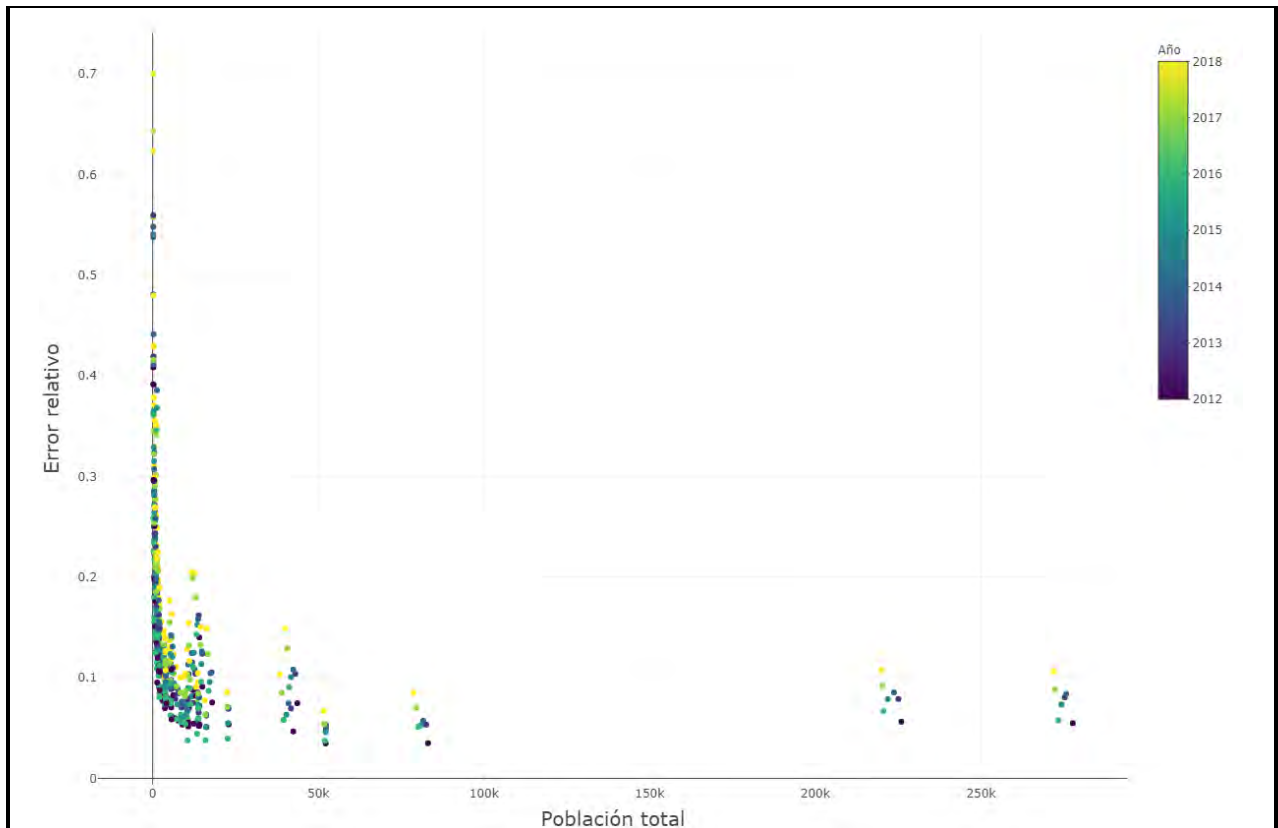
#### **T4.4: Análisis de resultados sobre entorno piloto y reajuste de parametrización.**

##### HERRAMIENTA DE SIMULACIÓN

##### Errores por municipio y año

Dichos datos están formados por 546 entradas respecto a 5 variables, siendo las mismas que en el caso general mostrado en la tarea previa, donde la población total real pasa a estar definida mediante la variable “pob\_total”, y se sustituye el código de distrito por el código de municipio “cod\_municipio”. En este caso, se puede apreciar como los errores mejoran si se analizan municipios con mayor población, mientras que los más pequeños presentan errores más amplios.

En la Figura 5 se muestra una representación gráfica de los diferentes municipios analizados, donde se presenta en el eje X la población total real, en el eje Y el error relativo, y el color representando el año de análisis.



**Figura 5 – Representación gráfica de errores relativos de cada municipio con respecto a la población total y el año analizado.**

Se puede apreciar fácilmente que los municipios con población total inferior a 20.000 habitantes tienen un comportamiento mucho más marcado en cuanto al error relativo, siendo éste más alto. Esto se debe al hecho que un sistema predictivo (ya sea con métodos de *machine learning* o a través de metodologías de simulación estocástica) es más sensible a variaciones no controladas en casos con menor representación. Nótese que estos errores se obtienen de estimaciones individuales por año y género con respecto a cada uno de los distritos, es decir, el número de habitantes se reparte en un total de 40 segmentos por cada distrito que componga el municipio, por lo que el total de habitantes en la simulación individual es mucho más reducido.

También se puede apreciar la influencia del año en análisis, y como cuanto más cercano sea el año de análisis al año inicial, menor es el error relativo. Resulta coherente, teniendo en cuenta que dichas simulaciones basan su estimación en el año previo, y en este caso, a excepción de la simulación del 2.012, todos basan su estimación en estimaciones anteriores, acumulando errores según se vaya avanzando temporalmente.

#### Errores por distrito y año

Por otro lado, se puede realizar un análisis análogo, pero pasando a estudiar los propios distritos en lugar de los municipios. Dichos datos están formados por 910 entradas respecto a 5 variables, siendo éstas las mismas que en el caso municipal, incluyendo el código de distrito “cod\_sc” en lugar del de municipio. De

nuevo, los errores mejoran si se analizan distritos con mayor población. En la Figura 6 se muestra la representación análoga a nivel distrito en lugar de municipal.



**Figura 6 – Representación gráfica de errores relativos de cada distrito con respecto a la población total y el año analizado.**

Las conclusiones son semejantes al caso municipal, donde se ve que distritos inferiores a 5.000 habitantes pasan a tener mayores errores por el mismo razonamiento previamente expuesto. Del mismo modo, se detecta la misma influencia del año en análisis respecto a la cercanía de una base de datos reales y no simulados.

#### Errores de distritos de Gijón

En este subapartado y los siguientes, analizaremos individualmente los casos de interés de este último hito del proyecto, comenzando por el municipio más poblado de Asturias: Gijón.

Para ello, tomando los datos de los errores por distrito y año y filtrando a los casos del municipio de interés, se puede proceder al análisis individual de los propios distritos. En este caso, Gijón consta de 9 distritos diferentes. En la Figura 7 se muestra gráficamente los distritos que componen el municipio de Gijón.

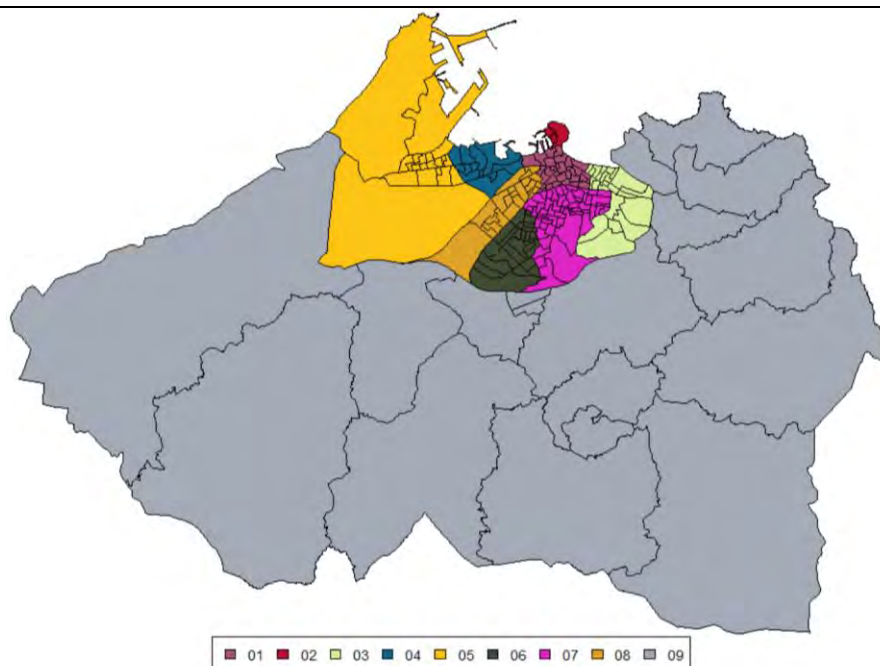


Figura 7 – Mapa de los distritos de Gijón.

Como se puede ver, 8 distritos engloban la parte urbana del municipio, mientras que el último hace lo propio con el resto. Nótese como las zonas de la parte urbana están bien definidas.

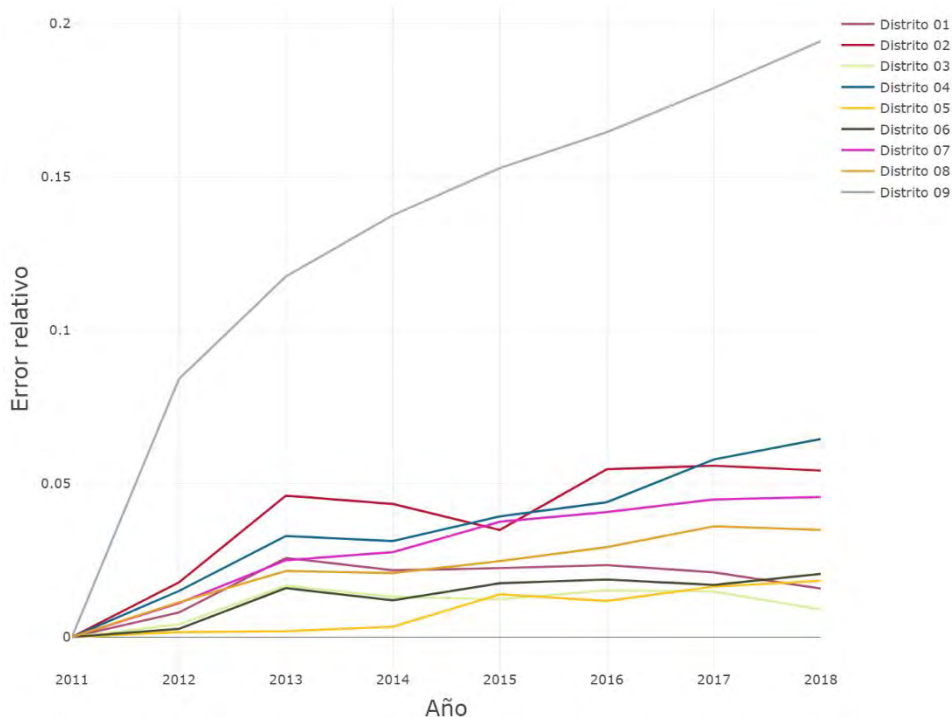
Con respecto a la población a 2.019 de cada distrito, se presenta en la Figura 8 su distribución tanto en términos totales, como por hombres y mujeres separados.

cod_sc	anio	total	hombres	mujeres
All	All	All	All	All
3302401	2019	32712	14551	18161
3302402	2019	2636	1273	1363
3302403	2019	32615	15267	17348
3302404	2019	17590	8456	9134
3302405	2019	31304	15107	16197
3302406	2019	33186	15902	17284
3302407	2019	64724	30266	34458
3302408	2019	30907	14513	16394
3302409	2019	26106	12793	13313

Figura 8 – Datos de población de los distritos de Gijón a 2.019.

Como se puede apreciar, solo el distrito 02 tiene una población sensiblemente más baja que el resto, correspondiente a Cimadevilla. En el polo opuesto, nos encontramos con el distrito 07, el cual engloba a El Llano, Ceaes, Contrueces y El Coto.

En la Figura 9 se presenta la evolución del error relativo de cada uno de los 9 distritos de Gijón previamente definidos, a lo largo de los años inicial (2.011) y de simulación (2.012-2.018).



**Figura 9 – Evolución del error relativo con respecto al año de simulación para los distritos de Gijón.**

Tal y como se puede observar en dicha figura, el error relativo de todos los distritos aumenta con los años, dado que la base de una nueva simulación es la simulación previa, y a mayor distancia con el origen, mayor error acumulado de paso a paso. Por otro lado, es destacable como los errores de todos los distritos se mantienen por debajo del 5% o lo superan ligeramente, a excepción del distrito 09.

Sin embargo, dicho caso concreto se puede entender al ver la evolución de la población de dicho distrito particular, tal y como se muestra en la Figura 10.

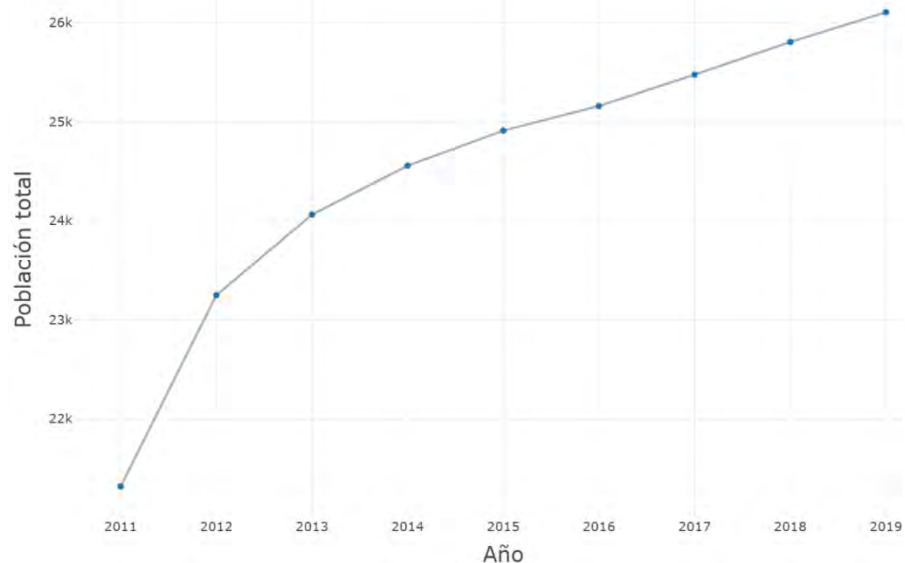


Figura 10 – Población del Distrito 09 de Gijón a lo largo de los últimos años.

Tal y como se aprecia, existe un salto de población brusco desde el 2011 hasta el 2012, lo que indica la existencia de una situación atípica en dicho distrito en ese instante temporal concreto, y cuyo comportamiento, como es de esperar, un sistema de simulación como el aquí presentado no tiene capacidad de predecir o reproducir.

#### Errores de distritos de Oviedo

Una vez estudiado el caso previo, se procede a realizar un análisis exhaustivo del segundo municipio más poblado de la región: Oviedo. De modo análogo, en la Figura 11 se muestran gráficamente los distritos que componen el municipio de Oviedo.

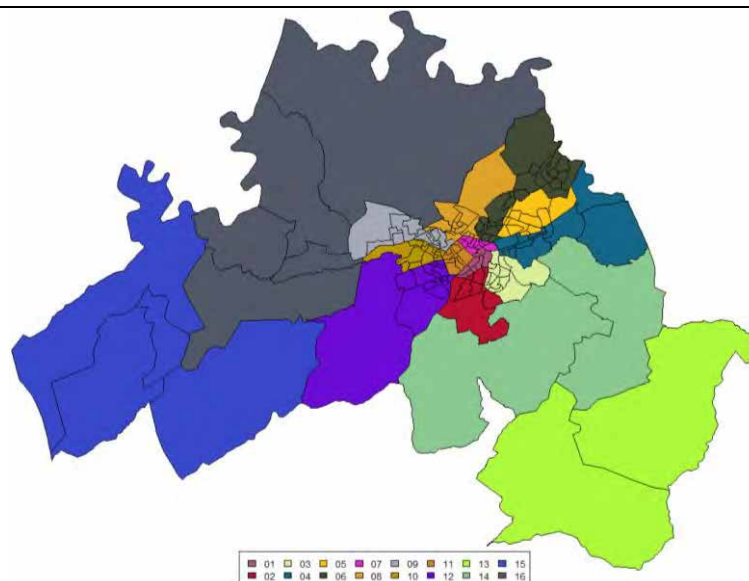


Figura 11 – Mapa de los distritos de Oviedo.

En este caso, Oviedo está compuesto por un total de 16 distritos diferentes, correspondientes a diferentes zonas del municipio, de forma similar a Gijón, agrupando por un lado barrios urbanos, y por otro, zonas exteriores del municipio.

Con respecto a la población a 2.019 de cada distrito, se presenta en la Figura 12 la distribución de la población tanto en términos totales, como por hombres y mujeres.

cod_sc	año	total	hombres	mujeres
All	All	All	All	All
3304401	2019	7623	3293	4330
3304402	2019	10240	4769	5471
3304403	2019	16681	7755	8926
3304404	2019	19263	9010	10253
3304405	2019	17131	8047	9084
3304406	2019	36899	17445	19454
3304407	2019	10455	4556	5899
3304408	2019	19490	8957	10533
3304409	2019	21065	9946	11119
3304410	2019	19281	8735	10546
3304411	2019	8797	3875	4922
3304412	2019	19148	8765	10383
3304413	2019	2493	1223	1270
3304414	2019	3032	1491	1541
3304415	2019	3514	1672	1842
3304416	2019	4574	2193	2381

Figura 12 – Datos de población de los distritos de Oviedo a 2.019.

En este caso, no existen distritos con un tamaño poblacional demasiado marcado respecto al resto, más allá del distrito 06, el cual tiene una población superior al resto.

En la Figura 13 se presenta de nuevo la evolución del error relativo para cada distrito de Oviedo, para los años 2.011 hasta 2.018.

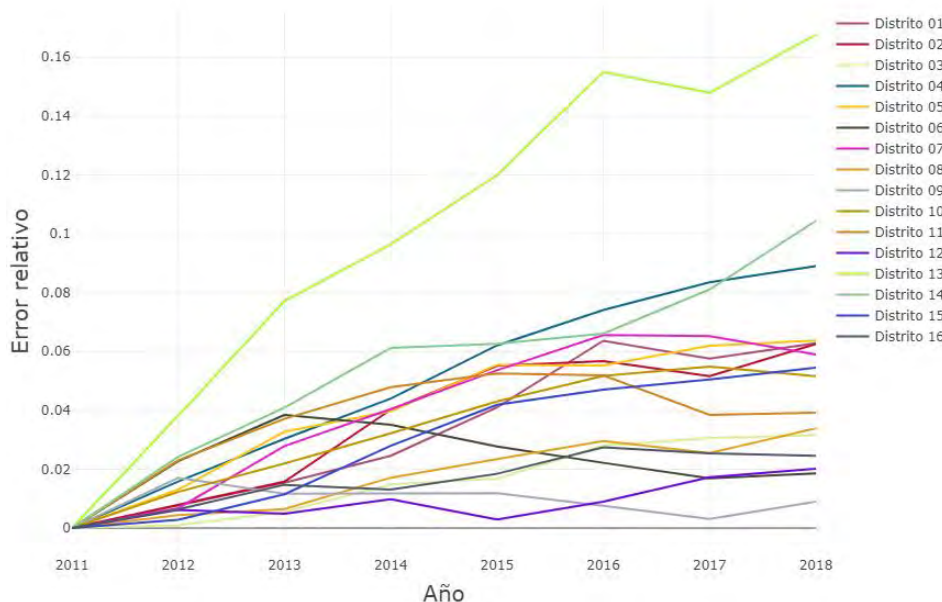


Figura 13 – Evolución del error relativo con respecto al año de simulación para los distritos de Oviedo.

En este caso, existe un distrito que obtiene errores relativos más marcados (distrito 13) al igual que en el caso de Gijón, si bien no lo hace de una forma tan pronunciada. Analizando dicho distrito de forma aislada, se puede ver la evolución de su población en la Figura 14.

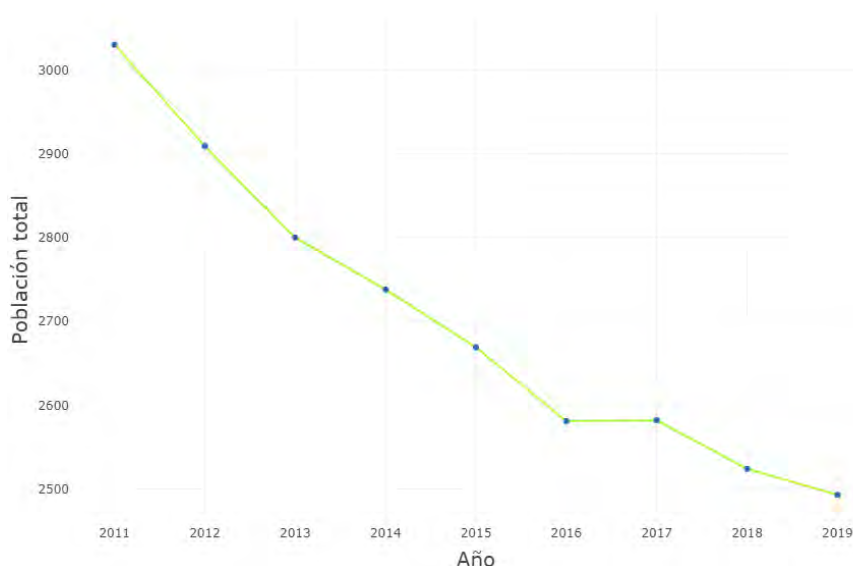


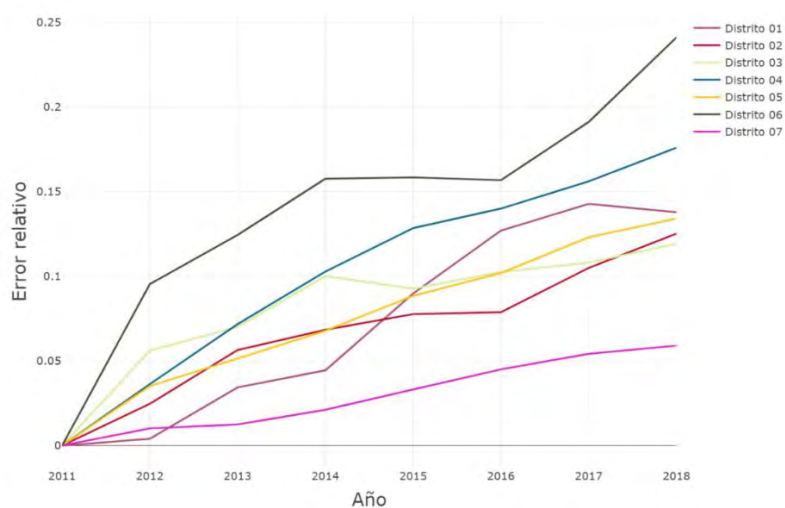
Figura 14 – Población del Distrito 13 de Oviedo a lo largo de los últimos años.

De modo opuesto al caso del distrito 09 de Gijón, existe un declive muy brusco en población en el inicio, lo cual, al igual que en el caso previo, lastra los resultados de las simulaciones del resto de años.

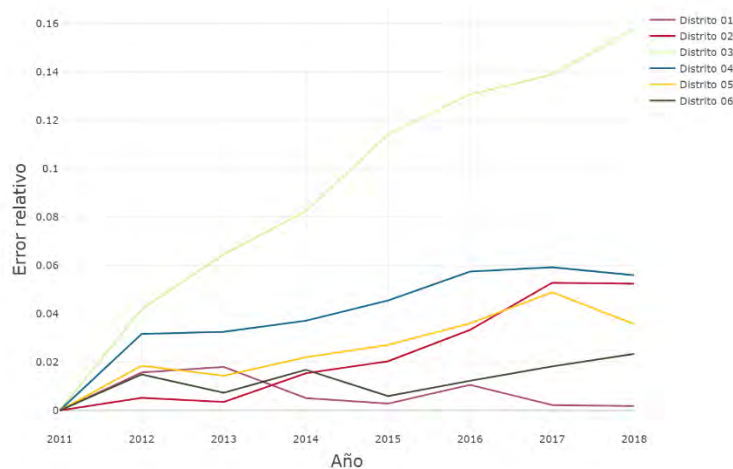
### Errores de distritos de otros municipios

Una vez analizados los dos municipios con mayor población y número de distritos, se realiza un análisis similar de otros municipios con un alto número de distritos, incluyendo menos detalle que en los casos previos. Se incluye tanto el mapa de distritos, como la serie temporal de los errores relativos por año y distrito.

#### Langreo



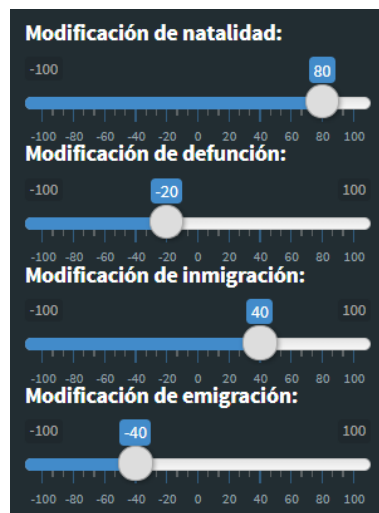
#### Villaviciosa



HERRAMIENTA DE ESTIMACIÓN

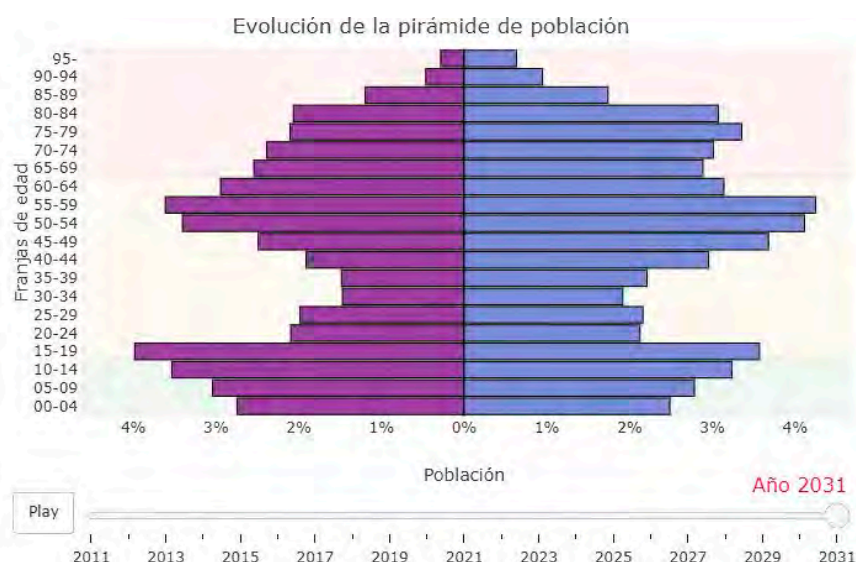
**Gijón: distrito 03**

En el primer caso de uso, se han seleccionado como argumentos de entrada el distrito 3 de Gijón, partiendo del año 2.011 como inicial, y prediciendo a 20 años vista. Previo a la aplicación del método de estimación, se ha generado una simulación a través de la herramienta previa, mediante los modificadores de entrada presentados en la Figura 15.



**Figura 15 – Argumentos de entrada para los modificadores de la aplicación de la herramienta de simulación previa para el caso de uso del distrito 03 de Gijón.**

Así, los resultados obtenidos en el último año en estudio son los mostrados en la pirámide de población de la Figura 16.



**Figura 16 – Pirámide obtenida mediante la herramienta de simulación a 20 años vista para el caso de uso del distrito 03 de Gijón.**

De este modo, dada dicha pirámide, se ha incluido en el sistema de estimación como argumento de entrada, de modo que se busquen obtener los 4 parámetros que lleven de la forma más cercana a unos resultados semejantes a los deseados. Consecuentemente, los resultados deberían ajustarse en la medida de lo posible a los parámetros dados de entrada en la herramienta de simulación. En la Figura 17 se muestra la pirámide obtenida mediante el proceso de estimación.

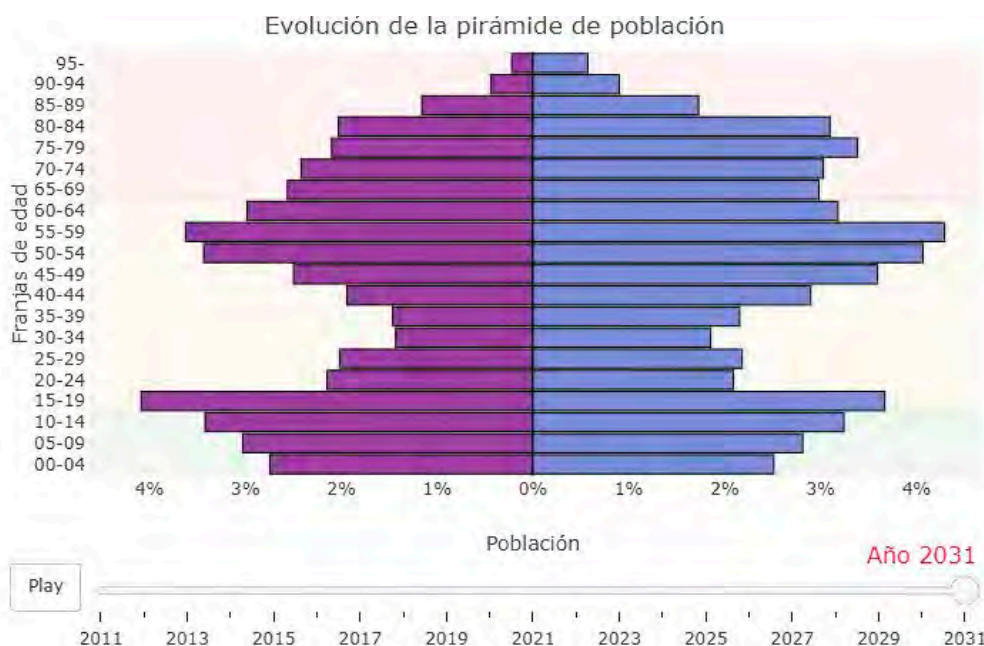


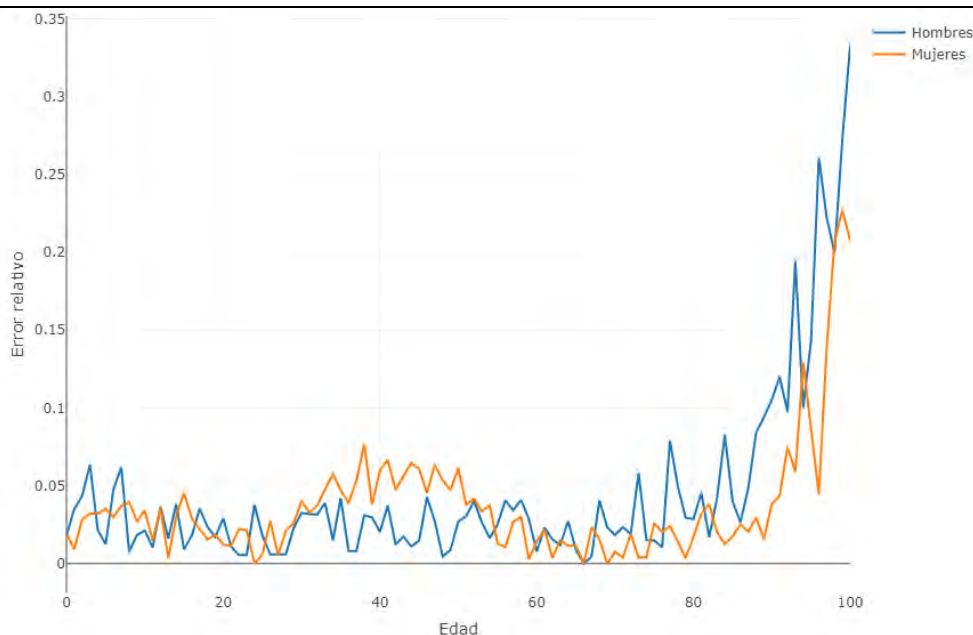
Figura 17 – Pirámide obtenida mediante los argumentos seleccionados para el caso de uso del distrito 03 de Gijón.

Se puede observar que el ajuste es casi perfecto en todos los tramos de edad y género. En la Tabla 1 se presentan los parámetros obtenidos, junto a los iniciales, y su diferencia en términos absolutos.

Tabla 1 – Comparativa de parámetros iniciales y estimados para el caso de uso del distrito 03 de Gijón.

Tasa	Inicial	Estimado	Diferencia
Natalidad	80%	81,27%	1.27%
Mortalidad	-20%	-13,04%	6.96%
Inmigración	40%	27,89%	12.11%
Emigración	-40%	-44,6%	4.6%

Tal y como se puede apreciar, las diferencias son muy reducidas en todos los parámetros, solo notándose una ligera divergencia en la tasa de inmigración. En cuanto a errores individuales respecto a edad y género, se presenta el error relativo de cada combinación de ambas variables en la Figura 18.



**Figura 18 – Error relativo por edad y género para el caso de uso del distrito 03 de Gijón.**

Como se puede apreciar, los errores están por debajo del 5% en la mayor parte de los casos, solo disparándose a partir de los 90 años, debido al menor volumen de población, donde una pequeña desviación reporta un error relativo más amplio.

#### Oviedo: distrito 10

Por otro lado, se ha aplicado la herramienta al distrito 10 de Oviedo, seleccionando como argumentos de entrada como inicio el año 2.011, y prediciendo a 20 años vista. De nuevo, se ha realizado una simulación a través de la herramienta correspondiente, utilizando los mismos parámetros de entrada que en el caso de uso anterior, mostrados en la Figura 15. De este modo, el último año predicho mediante dicha parametrización se presenta gráficamente mediante la pirámide mostrada en la Figura 19.

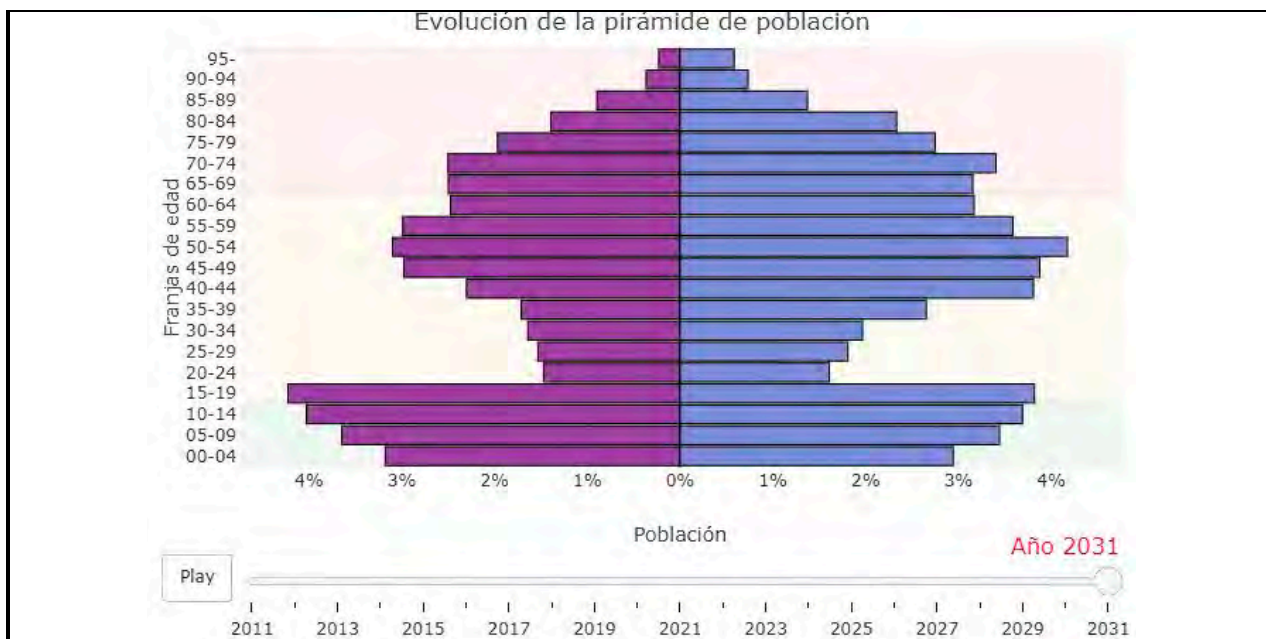


Figura 19 – Pirámide obtenida mediante la herramienta de simulación a 20 años vista para el caso de uso del distrito 10 de Oviedo.

De este modo, dada dicha pirámide, se ha incluido en el sistema de estimación como argumento de entrada, de modo que se busquen obtener los 4 parámetros que lleven de la forma más cercana a unos resultados semejantes a los deseados. Consecuentemente, los resultados deberían ajustarse en la medida de lo posible a los parámetros dados de entrada en la herramienta de simulación. En la Figura 20 se muestra la pirámide obtenida mediante el proceso de estimación.

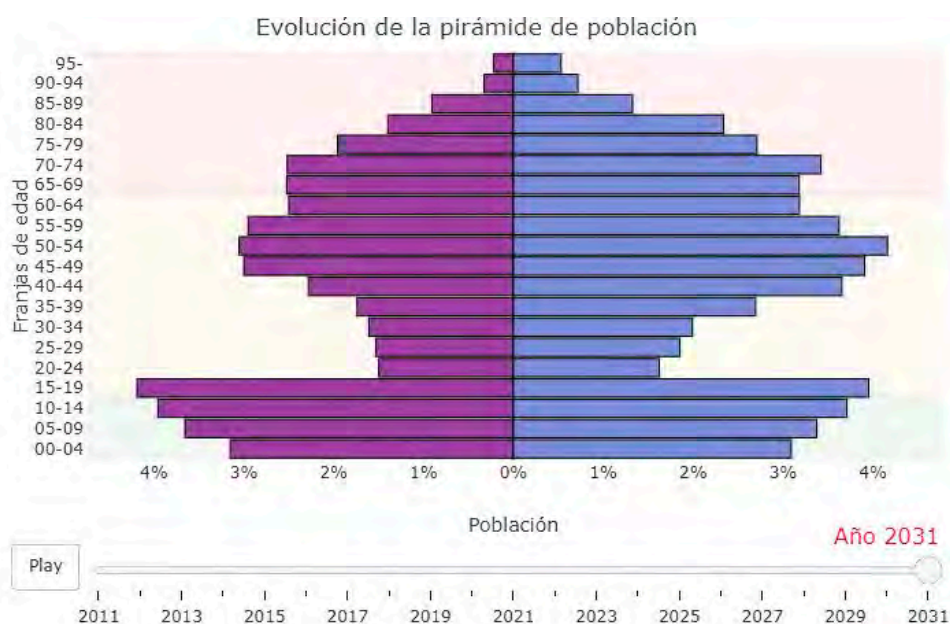


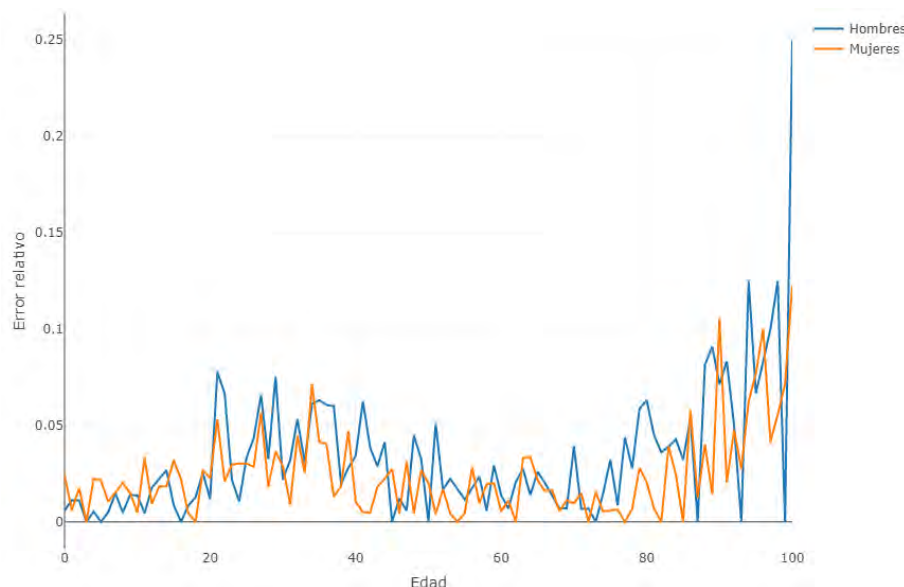
Figura 20 – Pirámide obtenida mediante los argumentos seleccionados para el caso de uso del distrito 10 de Oviedo.

Al igual que en el caso de uso anterior, se aprecia un buen ajuste de todos los tramos tratados por edad y género. En la Tabla 2 se presenta una comparación análoga de los parámetros reales y obtenidos.

**Tabla 2 – Comparativa de parámetros iniciales y estimados para el caso de uso del distrito 10 de Oviedo.**

Tasa	Inicial	Estimado	Diferencia
Natalidad	80%	78,99%	1.01%
Mortalidad	-20%	-15,86%	4.14%
Inmigración	40%	63,82%	23.82%
Emigración	-40%	-14,42%	25.38%

En este caso, el ajuste de natalidad y mortalidad es aún mejor que en el caso del distrito 03 de Gijón, si bien existe una mayor divergencia en la tasa migratoria obtenida a partir de las tasas individuales de inmigración y emigración. Del mismo modo, se presentan los errores relativos individuales respecto a edad y género en la Figura 21.



**Figura 21 – Error relativo por edad y género para el caso de uso del distrito 10 de Oviedo.**

Se identifica un comportamiento idéntico al del caso del distrito 03 de Gijón, donde la mayor parte de errores siguen por debajo del 5%, solo superándolo en las edades más avanzadas debido a la repercusión de variaciones en poblaciones más reducidas.

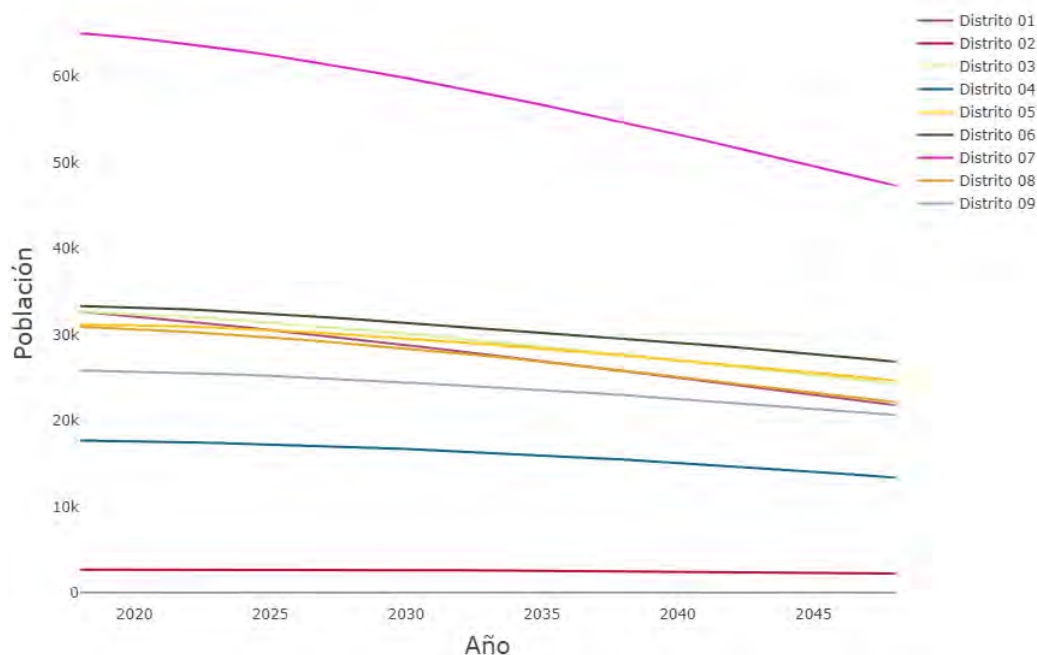
### PROYECCIONES FUTURAS

En este último apartado, procedemos a obtener a través de la herramienta de simulación las proyecciones futuras de los diferentes casos de uso aplicados previamente, de modo que sea posible visualizar las expectativas de evolución de dichos escenarios con respecto a las variables utilizadas.

En los diferentes casos de uso, se tomará como punto de partida el último año completo de datos, es decir, 2.018, y se procederá a realizar predicciones a 30 años vista para todos los distritos que compongan el municipio en estudio.

### Distritos de Gijón

En el primero de los casos de uso, se realizan las predicciones correspondientes a los 9 distritos que componen el municipio de Gijón. En la Figura 22 se presenta dicha evolución por distrito de forma agregada en cuanto a población total.



**Figura 22 – Evolución de la población de los distritos de Gijón entre 2.018 y 2.048.**

Se puede apreciar como la mayoría de los distritos presentan una tendencia descendente de forma homogénea entre distritos, si bien dicho movimiento podría verse más marcado en algunos distritos más que en otros a lo largo de la gráfica dada la escala global. Para contrastar dichas variaciones numéricamente, en la Tabla 3 se presentan las variaciones porcentuales de la población de cada distrito entre el año inicial y el año final.

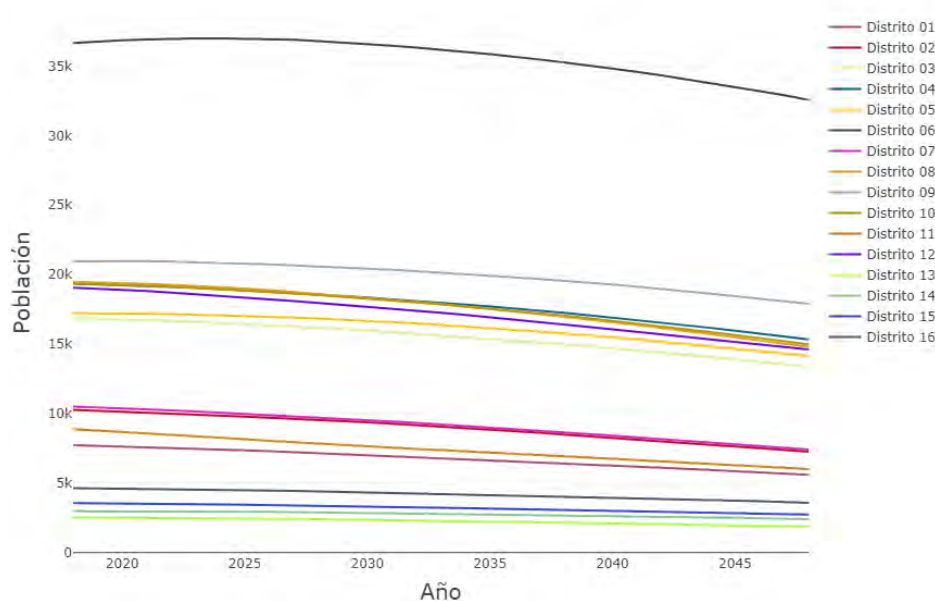
**Tabla 3 – Variación de la población total de los distritos de Gijón entre 2.018 y 2.048.**

Distrito	Variación
01	-33.32%
02	-17.03%
03	-25.7%
04	-24.24%
05	-20.89%
06	-19.45%
07	-27.23%
08	-28.38%
09	-19.97%
<b>TOTAL</b>	<b>-25.24%</b>

Dichos datos muestran que las oscilaciones son siempre negativas, siendo el distrito 02 (zona centro) el que menos pérdida porcentual sufre, y siendo el distrito 01 (Cimadevilla) el que presenta una bajada más pronunciada. A nivel global del municipio, la perspectiva es de una disminución de la población de algo más de un cuarto de la población.

### Distritos de Oviedo

De nuevo, el segundo caso de uso se centra en el municipio de Oviedo y sus 16 distritos. En la Figura 23 se muestra la evolución de la población total correspondiente a cada uno de ellos.



**Figura 23 – Evolución de la población de los distritos de Oviedo entre 2.018 y 2.048.**

Al igual que en el caso de Gijón, se detecta gráficamente la tendencia negativa, si bien no es posible apreciar la pendiente real de cada una de las curvas. Por dicha razón, se ha vuelto a generar los datos individuales a nivel porcentual de descenso desde 2.018 hasta 2.048, presentados en la Tabla 4.

**Tabla 4 – Variación de la población total de los distritos de Oviedo entre 2.018 y 2.048.**

Distrito	Variación
01	-27.71%
02	-29.29%
03	-20.71%
04	-20.73%
05	-17.84%
06	-11.16%
07	-29.4%
08	-23.97%
09	-14.58%

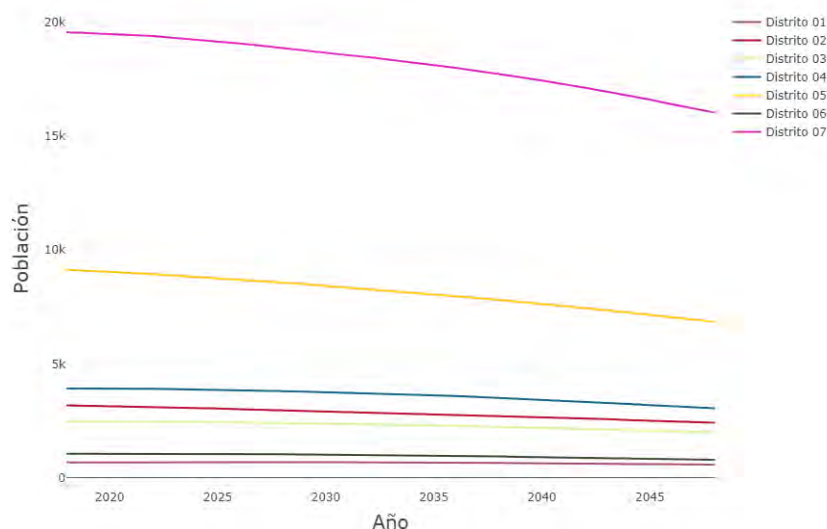
10	-22.61%
11	-32.38%
12	-23.37%
13	-26.7%
14	-19.12%
15	-23.17%
16	-22.74%
<b>TOTAL</b>	<b>-20.66%</b>

Se puede apreciar como todas las variaciones son negativas, con unas tendencias semejantes entre los diferentes distritos, incluyendo distritos con casos menos pronunciados como el distrito 06, o con una mayor caída, como el distrito 11. A nivel municipal, el descenso es de un quinto de la población.

### Otros distritos

De nuevo, al igual que en el apartado correspondiente a la herramienta de simulación, se ha procedido a analizar otros dos municipios con un número suficientemente amplio de distritos que los compongan: Langreo y Villaviciosa. En ambos casos, se muestran los mismos datos que en los previos, adaptados a dichos municipios, sin incluir conclusiones, al ser semejantes a los previos.

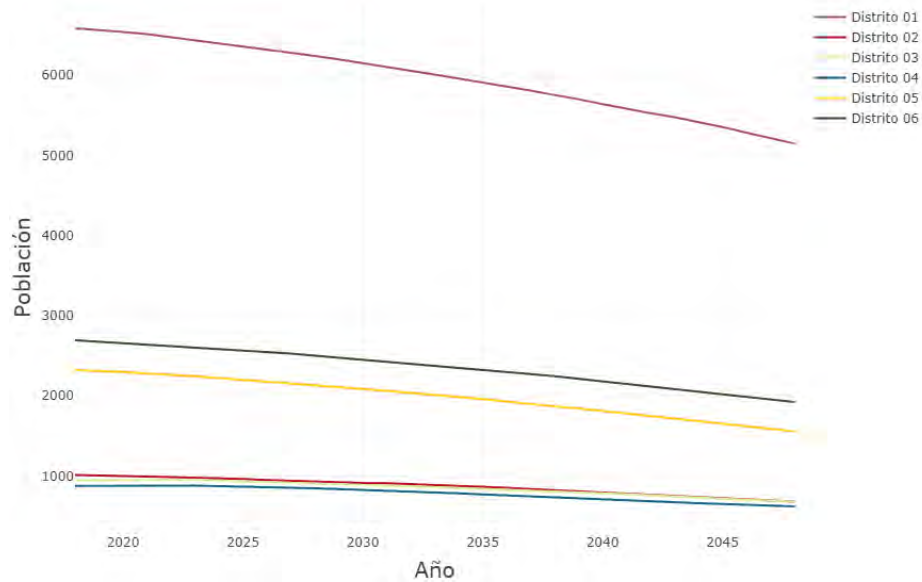
### Langreo



Districto	Variación
01	-14.24%
02	-24.05%
03	-19.69%
04	-22.18%
05	-25.1%
06	-25.83%
07	-18.08%

**TOTAL -20.8%**

Villaviciosa



Distrito	Variación
01	-21.83%
02	-32.84%
03	-27.86%
04	-29.34%
05	-33.03%
06	-28.65%
<b>TOTAL</b>	<b>-26.52%</b>

## 2. RESULTADOS CONSEGUIDOS

### JUSTIFICACIÓN 2018

Durante el periodo de tiempo que cubre este informe, se han conseguido todos los resultados planificados, tal y como se había planteado en la memoria inicial. Se detallan a continuación estos resultados, relacionándolos con los objetivos específicos del proyecto:

- Obtención de un catálogo de procesos y escenarios tipo para el modelo de datos con el conjunto de datos multivariante

Se han definido un conjunto muy amplio de procesos, escenarios y variables sobre demografía aplicados a despoblamiento y envejecimiento, superando siempre los valores óptimos previstos y ampliando muy considerablemente tanto el conjunto de datos que formarán parte del sistema, como los marcos de aplicación de la herramienta, que podrá dar respuesta a un número más amplio de modelos de análisis para alcanzar de una manera más óptima los objetivos marcados.

- Modelización del comportamiento del conjunto multivariante sobre un núcleo de Inteligencia Artificial sobre algoritmos de Aprendizaje Automático.

### JUSTIFICACIÓN 2019

Durante el periodo de tiempo que cubre este informe, se han conseguido todos los resultados planificados, tal y como se había planteado en la memoria inicial. Se detallan a continuación estos resultados, relacionándolos con los objetivos específicos del proyecto:

- Modelización del comportamiento del conjunto multivariante sobre un núcleo de Inteligencia Artificial sobre algoritmos de Aprendizaje Automático.
  - Se ha definido un conjunto muy amplio de variables recogidas de varias fuentes oficiales de datos. Una vez recogidas, se han procesado para eliminar la información superflua y unificar los agregados de datos de tal forma que puedan ser utilizados por los modelos de predicción que se han creado
  - Se ha creado un sistema de modelización funcional basado en diversas técnicas y algoritmos que son capaces de operar con las variables y datos demográficos recogidos para crear el sistema.
- Desarrollo de un prototipo funcional con arquitectura web aplicable a cualquier territorio y problemática.:
  - Se ha creado un prototipo formado por 3 herramientas que permite, usando modelos matemáticos y datos demográficos, representar simulaciones de la evolución demográfica poblacional en un tiempo futuro dado.

### JUSTIFICACIÓN 2020

Durante el periodo de tiempo que cubre este informe, se han conseguido todos los resultados planificados, tal y como se había planteado en la memoria inicial. Se detallan a continuación estos resultados, relacionándolos con los objetivos específicos del proyecto:

- Desarrollo de una plataforma cloud interoperable con capacidad para tratamiento de grandes volúmenes de información y con capacidades para la visualización dinámica de escenarios demográficos específicos:

Se ha desarrollado una plataforma web completamente funcional con los diferentes desarrollos mostrados a lo largo de todos los hitos que componen el proyecto, donde se pueden llevar a cabo comparaciones de municipios asturianos, la simulación a futuro de éstos, así como la estimación inversa basada en algoritmos genéticos.

- Tal y como se indica en el anterior punto, dicha plataforma incluye las diferentes aplicaciones desarrolladas a lo largo del proyecto y, por tanto, los algoritmos que las componen. Así, la misma plataforma preparará los resultados de salida para que puedan consultarse en el momento en que la ejecución termine, con la información correspondiente al análisis realizado.