



CENTRO DE ESTUDIOS DEMOGRÁFICOS, URBANOS Y AMBIENTALES

**Dinámica demográfica de las subpoblaciones a nivel municipal
2015-2020: el caso del Corredor Metropolitano Centro País.**

Tesis presentada por

Julio César Martínez Sánchez

Para optar por el grado de

Doctor en Estudios de Población

Codirectores de tesis

Dr. Manuel Ordorica Mellado

Dr. José Luis Silván Cárdenas

CIUDAD DE MÉXICO, AGOSTO 2022



CENTRO DE ESTUDIOS DEMOGRÁFICOS, URBANOS Y AMBIENTALES

Doctorado en Estudios de Población

Constancia de aprobación de tesis

Ciudad de México, 5 de agosto de 2022

Directores de Tesis:

Dr. Manuel Ordorica Mellado
Dr. José Luis Silván Cárdenas

Aprobada por el Jurado Examinador:

Sinodales propietarios

Presidente:

Dr. Manuel Ordorica Mellado

Primer Vocal:

Dr. Juan Guillermo Figueroa Perea

Vocal Secretario:

Dr. José Luis Silván Cárdenas

Suplente:

Dra. Olga Lorena Rojas Martínez

Agradecimientos

La culminación de este trabajo de tesis es un esfuerzo conjunto de muchas y maravillosas personas que participaron de una forma u otra, ya sea escuchando, cuestionando o dando un consejo, el cual detonó en alguna idea que fue incorporada en algún lugar de este trabajo. Reconozco que los aciertos de esta tesis son el resultado del apoyo de todas aquellas personas que estuvieron a lo largo del camino.

El Colegio de México, el CEDUA y CONACyT fueron las instituciones que me permitieron llegar hasta este punto y a las cuales les estoy muy agradecido. Me faltan palabras para agradecer a mi comité de tesis formado por los doctores Manuel Ordorica, José Luis Silvan y Juan Guillermo Figueroa, quienes en todo momento me apoyaron en cuestiones técnicas, metodológicas, conceptuales y anímicas; son un ejemplo a seguir. Tengo muy presentes nuestras reuniones de comité que se volvían muy amenas cuando jugaban “los pumas”. Los aciertos de esta tesis se deben a algún consejo que me dieron; los errores son solo míos.

Agradezco a mi familia conformada por Bernardo, Clara, Angélica, Héctor y Byanka quienes estuvieron presentes en los momentos más difíciles y cuando los ánimos decaían. También reconozco el aporte de mis amigos quienes hicieron posible la culminación de este proyecto, me parece injusto tratar de nombrar a todos y todas porque no deseo olvidar a alguien, pero quiero que sepan que su apoyo se reflejó a lo largo de este trabajo. Asimismo, estoy en deuda con *mi tribu*, por hacer que estos cuatro años fueran más llevaderos.

RESUMEN

Debido a que la *población* como categoría analítica anula la diversidad, asume que todas las personas son iguales, concentra el interés en los agregados y toma la forma de un estimador, para analizar a los conglomerados humanos se debe tomar como referencia el *espacio* y el *tiempo*, de acuerdo con Vieira (1973). En consecuencia, estos elementos deberían estar presentes al momento de generar los datos poblacionales. Sin embargo, en la realidad existen dos problemas: 1) el análisis de los agregados oculta comportamientos particulares y se cae en lo denominado como “la falacia de los promedios” (Savage, Danzier, y Markowitz, 2012), y 2) aunque se reconoce la importancia del *espacio* en la manera de concebir algún fenómeno social, ha pasado a segundo término en el quehacer demográfico al momento de estudiar a la población (Matthews y Parker, 2013).

Ante tal escenario, esta investigación examina la dinámica demográfica de las subpoblaciones y hace una propuesta metodológica para estudiar su comportamiento incorporando el *espacio*. Para ello, las preguntas guía para esta investigación son: 1) ¿Cómo aprehender la dinámica demográfica de las subpoblaciones?, y 2) ¿Cómo incorporar al *espacio* en las estimaciones de población?

Para responderlas, se seleccionó un área de estudio y se generó un *estimador* a partir de incorporar los datos de Encuesta Intercensal 2015, los datos de las proyecciones de población, las imágenes satelitales de estas zonas, y aplicar un algoritmo matemático que sirve para modelar fenómenos que cambian con el tiempo, mejor conocido como el filtro de Kalman. El área de estudio está delimitada por la zona que comprende el Corredor Metropolitano Centro País compuesto por 105 municipios de Aguascalientes, Querétaro, San Luis Potosí y Guanajuato, zona de gran importancia económica cuya distribución poblacional es variable pues combina municipios con alta y baja densidad poblacional.

Debido a que existen algunos indicios de que la precisión de los datos censales (Censo 2020) es menor a la de otros ejercicios debido al impacto de la COVID-19, se vuelve más visible la importancia de llevar a cabo este tipo de ejercicios en el que se realiza una proyección de la población. Como contribución, a partir del presente estudio se desarrolló una capa dentro de un SIG que muestra la distribución de la población en la zona de interés.

ÍNDICE GENERAL

Introducción general	7
Capítulo 1	
De la población a las subpoblaciones	9
1.1 Introducción	9
1.2 La <i>población</i> como categoría analítica	10
1.3 Teorías poblacionales	15
1.4 Subpoblaciones	19
1.5 Estudio formal de la población	21
1.6 Conclusiones	23
Capítulo 2	
La demografía y el espacio	25
2.1 Introducción	25
2.2 La incorporación del espacio	26
2.3 Enfoque teórico	28
2.4 Enfoque metodológico	30
2.5 Falacias referentes a la extensión y magnitud	32
2.5.1 El caso de las islas	35
2.6 Subpoblaciones, áreas pequeñas o subgrupos	36
2.7 Conclusiones	40
Capítulo 3	
La investigación demográfica-espacial	41
3.1 Introducción	41
3.2 Los datos demográficos	42
3.3 Las fuentes de información	45
3.4 Los métodos demográficos	48
3.5 Demografía espacial	50
3.6 Conclusiones	54
Capítulo 4	
Desagregación de la información	55

4.1	Introducción	55
4.2	Apertura a la información	57
4.3	Utilidad de la información	59
4.3.1	El interés público por desagregar la información	60
4.3.2	¿Para qué desagregar la información?	63
4.3.4	Problemas con la apertura de los datos	71
4.4	Dilemas éticos en el uso de datos geoespaciales	73
4.5	Acuerdos sobre la ciencia	83
4.6	Conclusiones	85
Capítulo 5		
Metodología para estimar el total de población en áreas pequeñas		87
5.1	Introducción	87
5.2	Enfoque demográfico	88
5.3	Enfoque demográfico-estadístico	90
5.4	Enfoque estadístico	93
5.4.1	Estimaciones directas	93
5.4.2	Estimaciones indirectas	95
5.5	Propuesta de análisis: Filtro de Kalman	99
5.6	Conclusiones	106
Capítulo 6		
Datos y área de estudio		107
6.1	Introducción	107
6.2	Datos espaciales	108
6.2.1	Percepción remota	108
6.2.2	Interpretación	112
6.2.3	Vínculo entre imágenes e información sociodemográfica	114
6.2.4	Del MAUP al MSTUP	116
6.2.5	Limitaciones al vínculo entre imágenes e información sociodemográfica	121
6.2.5	Máscaras urbanas	125
6.3	Datos sociodemográficos	128
6.3.1	Encuesta Intercensal 2015	128
6.3.2	Proyecciones de población	128
6.4	Área de estudio	130
6.5	Unidad de análisis	132
6.6	Conclusiones	133

Capítulo 7	
Resultados	135
7.1 Introducción	135
7.2 Estimaciones de población	136
7.3 Comparación de las estimaciones	141
7.3.1 Con los datos censales	142
7.3.2 Error Porcentual Absoluto Medio	144
7.3.3 Comparación temporal (proyecciones de población)	146
7.4 Visualización de los resultados	148
Reflexiones finales	153
Principales resultados de la tesis	157
Limitaciones de la tesis	158
Futuras investigaciones	159
Reflexiones sobre el Censo 2020	161
REFERENCIAS	163
ANEXO	173

ÍNDICE DE CUADROS

Cuadro 1 Resultados del modelo.....	136
Cuadro 2 Validación longitudinal de las estimaciones	147

ÍNDICE DE MAPAS

Mapa 1 Luces nocturnas.....	53
Mapa 2 Máscaras urbanas	127
Mapa 3 Municipios seleccionados	132
Mapa 4 Estimación 2020.....	141
Mapa 5 Datos censales	143
Mapa 6 Diferencia entre las estimaciones y los datos censales	144
Mapa 7 Distribución de la población de acuerdo con los resultados del filtro de Kalman	149

ÍNDICE DE DIAGRAMAS

Diagrama 1 Población y percepción de la población	15
Diagrama 2 Diagrama del filtro de Kalman	100
Diagrama 3 Sistema básico de percepción remota.....	109
Diagrama 4 Archivos raster.....	111
Diagrama 5 Interpretación de los archivos raster.....	113
Diagrama 6 Unión de los archivos raster y los datos poblacionales	115
Diagrama 7 Cambios a través del tiempo.....	117
Diagrama 8 Cambios de los elementos a través del tiempo.....	118
Diagrama 9 Composición de imágenes a través del tiempo.....	119
Diagrama 10 Niveles de análisis de las imágenes satelitales	121
Diagrama 11 Datos utilizados para el modelo y validación.....	130

Introducción general

Población es una categoría analítica que sirve para nombrar a un grupo de personas; al igual que todas las categorías su significado ha evolucionado a lo largo del tiempo e incluso se ha convertido en una palabra de uso cotidiano. Al hacer referencia a la *población* se anulan las diferencias entre las personas y se incurre en un problema: se asume que todos los individuos se comportan de la misma manera y viven en un mismo lugar. Este escenario es irreal, por lo que investigadores como Viera (1973) argumentan que para estudiar a la *población* como categoría analítica se debe de tomar en cuenta, no solo a los grupos de personas, sino incorporar dos términos más a la ecuación, a los que el autor denomina ejes coordinados: el *tiempo* y el *espacio*.

Para estudiar a la *población*, la demografía cuenta con una serie de métodos y procedimientos que se han desarrollado desde principios del siglo XVII (Caldwell, 1996). Pero analizando con detenimiento los supuestos en los que se sustentan estos métodos, el *espacio* carece de un sitio preponderante para estudiar la *población* y prueba de ello es que la Ecuación compensadora, considerada el pilar del análisis demográfico, privilegia la dimensión temporal. Debido al privilegio de la categoría temporal, algunos autores como Matthews y Parker (2013) y Chen (2017), consideran que el *espacio* desempeña un papel secundario; sin embargo, Viera (1973) sostiene que es uno de los ejes para el estudio que la *población*.

Desde esta perspectiva, estos métodos que se han consolidado a lo largo de los años y que hoy en día forman parte del quehacer sociodemográfico, limitan el estudio de la *población* porque: 1) ocultan comportamientos particulares que solo son visibles al fijar la atención en grupos de menor tamaño (Savage, Danzier, y Markowitz, 2012), y 2) restringen la capacidad de análisis, ya que solo se pueden llevar a cabo investigaciones que sigan los métodos y datos que son aceptados por la comunidad académica (González, 2018).

Ante tal escenario, el objetivo de esta investigación es examinar la dinámica demográfica de las subpoblaciones de 105 municipios que se ubican en el Corredor Metropolitano Centro País de 2015 a 2020 mediante el uso de imágenes satelitales. Las preguntas de investigación que sirven de guía son las siguientes: ¿cómo aprehender la dinámica demográfica de las subpoblaciones?, y ¿cómo incorporar fuentes de información no sociodemográfica en las estimaciones de población?

Para responder estas preguntas, en el Capítulo 1 se discute el origen de la categoría analítica denominada *población*, las teorías que explican su comportamiento y cómo, en el marco qué hacer socio demográfico podemos abordar el estudio de los grupos de menor tamaño denominados subpoblaciones. En el Capítulo 2, se reflexiona sobre la incorporación del espacio; tanto desde el punto de vista teórico como desde el metodológico y se discute sobre las falacias relacionadas con las delimitaciones político-administrativas. Así, en el Capítulo 3 se profundiza en las fuentes de información y los métodos demográficos y se pone de manifiesto que debido a las limitaciones existentes existe la necesidad de incorporar nuevas fuentes de información. Ahora bien, dado que el objetivo de esta investigación es fijar la tensión en poblaciones específicas, el Capítulo 4 reflexiona sobre la utilidad de tener información granular y los problemas éticos que surgen al fijar la atención en grupos específicos y al incorporar datos geoespaciales.

Una vez expuestos los temas de la incorporación del *espacio* y la necesidad de utilizar nuevas fuentes de información, en el Capítulo 5 se desarrolla una propuesta metodológica para vincular los datos espaciales y sociodemográficos mediante el filtro de Kalman. Posteriormente, en el Capítulo 6 se describen los datos que van a ingresar al modelo y se pone énfasis en cómo analizar las imágenes satelitales para crear un estimador mixto que se configure con la parte sociodemográfica de fuentes tradicionales y la parte espacial con imágenes satelitales. En este capítulo también se describe el área de estudio, el Corredor Metropolitano Centro País y la unidad político-administrativa que sirve de referencia para unir estas dos fuentes de información.

De esta forma, en el último capítulo se presentan los resultados de las estimaciones para los 105 municipios seleccionados y se discute sobre la pertinencia de emplear este tipo de estimadores en la investigación sociodemográfica, con sustento en el hecho de que cumple con la incorporación del *tiempo* el *espacio* y los agregados poblacionales que son los ejes analíticos que proponen autores como Viera (1973), Almanza (1988) y Canales y Lerner (2003). Para cerrar, se muestra el desempeño de las estimaciones y se comparan los resultados con los datos censales, y se advierte que estos últimos deben de tomarse con cautela porque el levantamiento censal coincidió con el inicio de la pandemia y esto produjo un aumento de la no respuesta, reconocido por el Inegi.

Capítulo 1

De la población a las subpoblaciones

1.1 Introducción

El término *población* es un constructo teórico que ha ido evolucionando largo del tiempo, el significado que ahora se tiene, y con el cual se analiza a los grupos humanos, no es el mismo al que se tenía tiempo atrás. El objetivo de esta sección es examinar el origen de la palabra *población* y cuáles son los elementos que le permiten posicionarse como categoría analítica, la cual se usa como referencia para estudiar a los grupos humanos. Para lograrlo, primero se examina cuál es su origen etimológico, el cual se remonta en las culturas grecolatinas porque fueron los primeros en reflexionar sobre la conformación de los grupos humanos.

Esta reflexión sobre cómo concebir a los grupos humanos se formalizó con las diversas teorías que aparecieron en el siglo XVIII, en donde se cimentaron las primeras reflexiones sobre el análisis formal del constructo teórico denominado población. Estas teorías sirvieron de referencia para el desarrollo de la demografía y los estudios de población, disciplinas encargadas de estudiar las características de los grupos humanos. Sin embargo, pese a todo este desarrollo aún existen algunos vacíos en estas disciplinas y uno de ellos es el estudio de las áreas pequeñas o subgrupos. En este capítulo se discuten algunas de las definiciones y se enfatiza que existen discrepancias en cuanto a cómo abordar este constructo teórico porque depende del enfoque.

Esta reflexión teórica y de los vacíos conceptuales que aún existen se enmarcan en el quehacer demográfico y en las formas de cómo se obtiene una legitimidad de que el conocimiento es válido. Por este motivo, el capítulo concluye con una reflexión en torno a cuáles son aquellos elementos que se consideran válidos y cuáles son los acuerdos explícitos o implícitos que rigen la manera de abordar los fenómenos y en particular a las pequeñas.

1.2 La *población* como categoría analítica

La etimología de la palabra *población* se remonta a época de los griegos, los primeros en reflexionar sobre el hombre y su relación con el estado. En este sentido, Platón fue uno de los primeros pensadores en tratar de construir el término *población*; el filósofo griego teorizó sobre la reproducción de los seres humanos (a los que denominó *polis*) y su efecto en la sociedad. Además, no únicamente se limitó a observar y describir su comportamiento, sino que llegó a concluir que era importante proponer un número ideal de ciudadanos, el cual fijaba en 5 040 (Kreager, 1991; Petit, 2019). Aunque no está del todo claro cómo calculó esa cifra, de acuerdo con Kreager (1991), este pensador estaba influenciado por la escuela pitagórica y lo más probable es que la cifra propuesta tenga un origen divino. Aunque a primera vista esta propuesta parece radical porque supone una manipulación en el tamaño, estructura o distribución de los individuos, en su momento estuvo respaldada por argumentos convincentes como la búsqueda de un equilibrio y por cuestiones más generales como saber *¿cuántos somos?*

Ese interés por examinar las características y la dinámica de los grupos humanos también fue abordado por Aristóteles, quien fijó su atención en torno a qué es la población y cómo se configura. A diferencia de su precursor, él se interesó en la forma en que las personas se agrupan para formar un cuerpo único; por lo que indagó sobre los cambios que ocurren en su comportamiento para facilitar dicha unión y, que, al mismo tiempo, restringen su voluntad individual. A pesar de que se apoyó en algunas ideas de Platón, por ejemplo, que ciertos derechos estaban reservados para la élite, su inclinación por estudiar a las minorías le permitió reconocer a grupos excluidos como esclavos y extranjeros (Kreager, 2008).

En el recuento de la construcción del término *población*, etimológicamente se tiene que mencionar a los romanos. Ellos crearon el término *civis* para hacer referencia a las personas que compartían un hábitat en común y que, al mismo tiempo, gozaban de derechos políticos. Sin embargo, al igual que *polis*, no se trata de un concepto incluyente, ya que dejaba fuera a aquellos que habían nacido en otros lugares, a los esclavos y a las mujeres. Durante un largo periodo, las diversas formas de nombrar a los grupos con características diferentes se mantuvieron sin cambios. En la Edad Media se le llamaba *populus* a los miembros de una comunidad, este término hacía referencia, de nuevo, a hombres adultos, libres y con la capacidad de poseer bienes y participar en las decisiones del Estado. Más adelante, en la época bizantina surgió la noción de pueblo (*demos*),

aunque este concepto tenía un cambio importante al considerar a personas que tenían la encomienda de aclamar a los gobernantes (Kreager, 1991). A partir de esta línea es posible observar un factor común en todas estas denominaciones: aludían a grupos específicos; tuvo que pasar mucho tiempo para que se pudiera construir un concepto que hiciera referencia a todas las personas que habitan, comparten y hacen funcional un espacio en común.

El término *población* siguió ganando popularidad para nombrar genéricamente a un grupo de personas, a tal punto que, para el siglo XVII, ya era un término que se usaba de manera cotidiana. Sin embargo, a lo largo de los años dejó de ser solamente un vocablo y se convirtió en una herramienta de categorización para visualizar a los seres humanos (Petit, 2013). En la sección siguiente se examina cómo fue dicho proceso y qué elementos intervinieron.

De acuerdo con Canales (2003), el primer requisito para que el concepto *población* se convirtiera en una categoría analítica es su capacidad de quitar las diferencias entre los grupos humanos. Pero ¿cómo se llegó a suprimir las características individuales? Para explicarlo existen dos hipótesis: la primera fue planteada por Astorga (1988), quien sugiere que el cristianismo, y su búsqueda por convertirse en la religión universal, difundió la idea de igualdad entre todos los hombres y mujeres, de ahí que uno de sus dogmas más importantes es que: “ante la ley de Dios todos son iguales”. Sin embargo, este argumento resulta insuficiente para visualizar las diferencias que existen entre los individuos, sobre todo si se toma en cuenta que un gran número de ellos no profesaban la fe cristiana.

La segunda hipótesis fue planteada por Petit (2013), quien apunta a que fue la consolidación de los estados nación lo que impulsó la idea de crear una “identidad común”. Aunque este término tiene diversos significados y para algunos pueblos hace referencia a su herencia histórica, pero en general, todos ellos trataron de difundir la idea que los habitantes de un mismo territorio son libres e iguales Mansilla (2011). Aunque ambas hipótesis no dejan de ser discutibles y la falta de claridad sobre cómo el término *población* llegó a incluir a todas las personas sin distinción, la tendencia a nombrar igual a un grupo de individuos siguió ganando popularidad a través de los años.

Sin embargo, el concepto *población* como categoría de análisis exige un cambio a nivel cognitivo porque su uso actual trae dos implicaciones: 1) se deja de pensar en las personas y se concentra el interés en los agregados, los individuos pierden importancia y solamente adquieren notoriedad si forman parte de un conglomerado, dicho de otra manera, se privilegia a los grupos

antes que a los individuos (Astorga, 1988:150); 2) |Al ser un concepto genérico e integrador, anula la diversidad y asume que todas las personas que viven en un determinado territorio se comportan de la misma manera (Anselin, Sridharan y Gholston, 2007; Vieira, 1973). Aun y cuando estos supuestos se pueden desestimar, han resultado ventajoso para los estudios demográficos porque gracias a ellos se tiene la capacidad de hacer una simplificación de la realidad. Así, un sinónimo teórico como es *población* se materializa en algo tangible, y con ello, se abre la posibilidad de establecer mecanismos para describir y comprender su comportamiento (Vallin, 1991). Además, un argumento a favor de la unificación es el siguiente: los seres humanos no se desarrollan en el vacío y únicamente el conglomerado tiene la capacidad de modificar un territorio para garantizar su sustento. Una sola persona es incapaz de lograr un cambio significativo, y aunque lo hiciera, esto no garantiza su supervivencia, de ahí que su pertenencia a un conglomerado tenga sus ventajas (Hauser y Duncan, 1959; Vieira, 1973).

Ahora, conviene preguntarse ¿es suficiente tener un constructo que haga alusión a un conjunto de personas para aprehender el comportamiento de los seres humanos? De acuerdo con Vieira (1973), la respuesta es no, ya que es necesario tomar en cuenta dos conceptos clave: el espacio y el tiempo; estos proporcionan el andamiaje teórico y metodológico para analizar a los conglomerados, lo que el autor denomina como “coordenadas demográficas”. Esto supone que cualquier intento por estudiar a un grupo de personas exige realizar delimitaciones espaciales y temporales porque de otra manera se dificulta su estudio (Hauser y Duncan, 1959). Por este motivo, conviene dedicar algunas líneas a analizar los componentes que dan forma a la tríada población-espacio-tiempo y sus implicaciones en el ámbito demográfico.

La relación que existe entre población y espacio tiene un impacto bidireccional. Por un lado, las características de los territorios pueden presentar restricciones que impidan el asentamiento humano o condicionen su comportamiento; por el otro, si dicha aprehensión se lleva a cabo, entonces el territorio deja de ser un espacio físico y adquiere un significado social para las personas que habitan en él (Chen, 2017; Matthews y Parker, 2013). Lo anterior puede ser comprobado al hacer una revisión de acontecimientos que han sucedido a lo largo de la historia, en donde muchas de las disputas armadas tenían el objetivo de que un grupo se apropiase de un territorio para obtener beneficios (Livi-Bacci, 2009; Sánchez, 2014). Para tener una mayor claridad sobre los conceptos, Vieira (1973) lleva a cabo una reflexión en torno a qué significa el espacio y observa una marcada diferencia entre lo que es un espacio físico y un espacio demográfico. El primero se refiere a un

lugar en la superficie terrestre, mientras que el segundo va más allá y hace referencia a un lugar que, en palabras de Vieira: “es aprehendido por un conjunto de seres humanos para su supervivencia y reproducción” (Vieira, 1973:65).

Ante esta afirmación, existen diversas posturas en torno a cómo se pueden vincular el espacio y los asentamientos humanos. Algunos autores como Livi-Bacci (2009) consideran que el entorno natural es el factor clave que favorece o limita los asentamientos humanos, esto explica por qué gran parte de las ciudades se asentaron cerca de los ríos, los cuales les proporcionaban un suministro constante de líquido. En contraparte, otros autores señalan que el espacio se construye a partir de las necesidades de los seres humanos y que son ellos quienes modifican la naturaleza, por lo que los criterios administrativos son los que importan, ya que mediante ellos se imponen y determinan las formas en cómo se establece la población (Vieira, 1973; Canales, 2003).

En este sentido, la siguiente coordenada sociodemográfica a la que hace referencia Vieira (1973) es el tiempo. El argumento principal del autor es que la población es un cuerpo dinámico que evoluciona en cada momento y nunca permanece estático, a pesar de estar ubicada en un territorio determinado. De ahí que la dimensión temporal es un factor clave para el estudio de las poblaciones humanas, los cuales únicamente se hacen evidentes cuando se compara el mismo fenómeno en dos momentos distintos, de otra manera, no hay parámetro de referencia para identificar las transformaciones (Astorga, 1988). Un ejemplo de ello es el crecimiento poblacional, el cual toma en cuenta a la cantidad de individuos que viven en una misma zona en dos momentos distintos en el tiempo y a partir de ello identifica las variaciones.

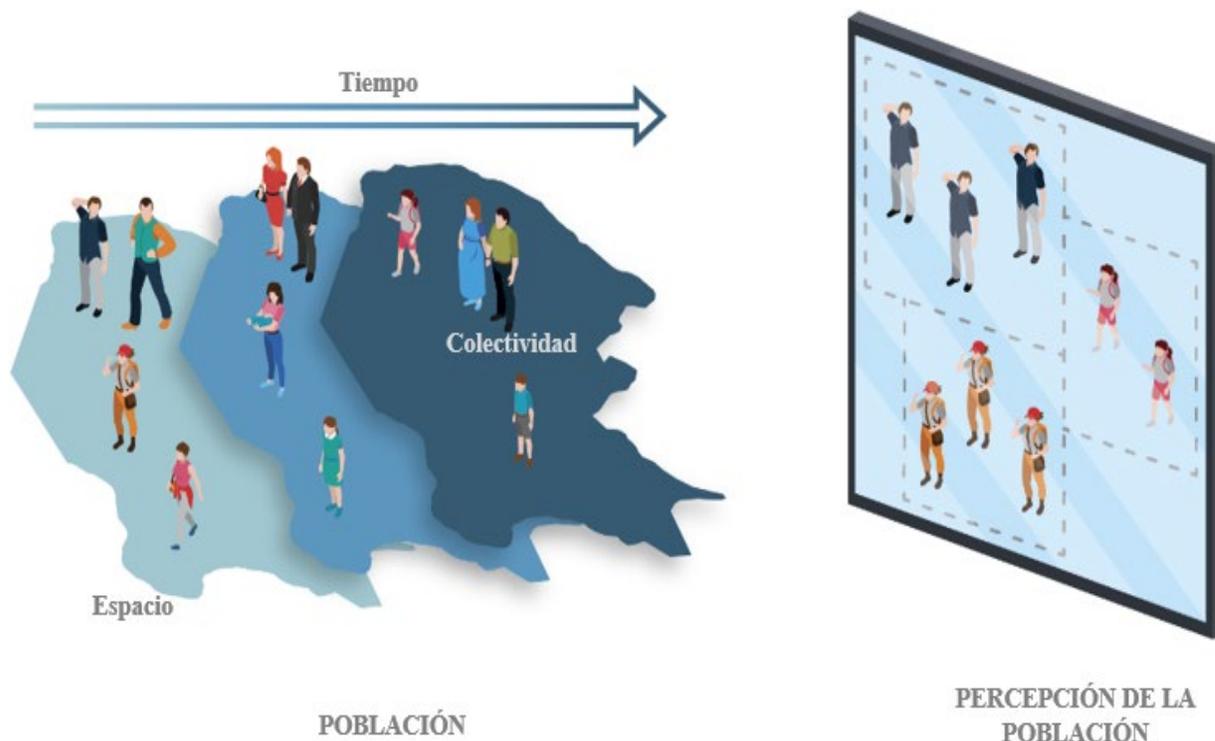
Es difícil ahondar en una definición precisa de lo que es tiempo, este tipo de reflexiones competen a los estudios filosóficos (Vieira, 1973). No obstante, si se analiza su interacción con la población como categoría analítica, se presenta una paradoja: ¿si en todo momento la cantidad de seres humanos cambia, ya sea por nuevos nacimientos o defunciones, entonces sería imposible llegar a conocer este objeto de estudio? De ahí que es necesario asumir cierta estabilidad para tratar de conocer su crecimiento o cambios en el volumen, sobre todo al momento de recoger los datos, ya que de otra manera sería imposible obtener información (Hauser y Duncan, 1959). Ante esta necesidad de hacer visible el efecto temporal han surgido algunos paradigmas, como el denominado curso de vida, cuyo eje analítico de referencia es el estudio de los fenómenos demográficos de manera diacrónica (Courgeau, 2001). Con esta perspectiva se asume que las personas están

inmersas en un tiempo y lugar determinado, de modo que los cambios sociodemográficos interaccionan unos con otros y configuran su trayectoria de vida (Blanco, 2011).

Sin embargo, un inconveniente que se tiene en el uso de *población* como categoría analítica es lo borroso de saber en qué momento un grupo de personas ya se considera como un cuerpo único y susceptible de ser analizado. Esta situación se puede ejemplificar con la siguiente paradoja: una sola persona en un momento y lugar determinado no es una población; si se considera a diez, aún podría ser discutible; pero si se incorporan más individuos, entonces se convertirán en población. De este planteamiento es importante recuperar que no existe una forma de distinguir en qué momento un grupo de individuos adquiere la cualidad de ser considerados como *población* (Vieira, 1973; Vallin, 1994).

Con estos tres elementos se puede determinar que la población es más que un conjunto de personas, pues se define a partir de su colectividad y sus dimensiones espaciales y temporales. En este sentido, Canales (2008) reflexionó en torno a ¿cómo un observador puede aprehender a la población como categoría analítica? La conclusión a la que llegó es que no puede debido a que los investigadores simplifican la realidad y pasan por alto algunas características porque están limitados por su propia percepción del fenómeno. En el Diagrama 1 se compara la diferencia entre lo que es la población como categoría analítica y lo que se logra captar con las herramientas metodológicas de la demografía. En particular, se distinguen tres situaciones: primero, se simplifica la dimensión temporal a un solo momento en el tiempo; segundo, se sustituye el espacio geográfico real por delimitaciones político-administrativas; tercero, se anula la diversidad. Esto último es el resultado de comportamientos particulares, los cuales son ocultados cuando se toma un dato agregado, o lo que Savage *et al.* (1999) denominan “la falacia de los promedios” (Savage, Danzier, y Markowitz, 2012). Sin embargo, la incapacidad de aprehender dichas características condiciona la forma de entender y analizar los fenómenos (Canales, 2019). Esta limitación no significa que se deba de omitir su estudio, sino estar consiente al momento de hacer inferencias sobre su comportamiento.

Diagrama 1 Población y percepción de la población



Fuente: elaboración propia con base en Canales (2019) y Vieira (1973)

Una vez que se hizo el recuento sobre la etimología de la palabra población y cómo se posicionó como categoría analítica, el siguiente paso es examinar las teorías demográficas que se han desarrollado para estudiar su estructura, evolución y características generales.

1.3 Teorías poblacionales

Alrededor del siglo XVIII se generó una disputa intelectual en torno al comportamiento que debían seguir las poblaciones para garantizar su sobrevivencia. William Godwin y Nicolás de Condorcet consideraron que los seres humanos poseían el control sobre su comportamiento reproductivo y contaban con la capacidad de producir alimento suficiente, de tal forma que la permanencia quedaría garantizada con una correcta redistribución de los recursos. Sin embargo, Robert Malthus opinaba lo contrario; él sostenía que la población no tenía pleno control en sí misma, por lo que, si

continuaba creciendo, se generaría un escenario catastrófico y la sobrevivencia se vería comprometida (Vieira, 1973).

Las ideas de Malthus se plasmaron en su libro *An Essay on the Principle of Population* (Ensayo sobre el principio de la población), la cual es reconocida como la primera obra con enfoque demográfico. En ella postuló una relación matemática entre el crecimiento poblacional y los alimentos disponibles de la siguiente manera: si los seres humanos se reproducen exponencialmente y producen los alimentos de manera geométrica, llegará el momento en el cual esta cantidad sea insuficiente. Debido a este postulado, el crecimiento poblacional adquirió una connotación negativa y se convirtió en un problema social, sin embargo, no deja de ser paradójico que, en la búsqueda por la sobrevivencia de la especie, la reproducción humana sea una amenaza (Malthus, 1966). Además, la necesidad de establecer control subyace a la búsqueda de limitar el crecimiento poblacional y es el origen de las políticas de población que se materializaron años después. El trabajo de Malthus fue el primero en enfocarse en el estudio de la población, por lo que es pionero en describir su crecimiento. Aunque no utilizó datos reales y sus deducciones se basaron en un ejercicio de razonamiento matemático, sus ideas fueron aceptadas en la comunidad académica y se difundieron en todo el mundo.

Su legado sobre el estudio matemático de la población tuvo un auge importante a finales del siglo XVII, con en los estudios de Alfred James Lotka, quien es considerado padre de la demografía matemática. Su aportación principal fue demostrar que las poblaciones cerradas a la migración en donde permanecen constantes las tasas de fecundidad y mortalidad, modifican su comportamiento, de lo que resulta que la estructura por edad y la tasa de crecimiento permanezcan constantes a lo largo del tiempo. Estas ideas se engloban en lo que se denominó Teoría analítica de las asociaciones biológicas y su publicación ayudó a consolidar el enfoque matemático con el cual se estudia a la población. Sin embargo, también hubo avances que privilegiaron el enfoque teórico y a partir de ello establecieron criterios que ayudaron a sustentar los cambios en el crecimiento de la población (Ordorica-Mellado *et al.*, 2019).

Ahora bien, la teoría de la transición demográfica sostiene que las tasas de fecundidad y mortalidad en los países en desarrollo son altas e inestables, pero disminuyen y se estabilizan cuando las naciones experimentan un proceso de modernización. Durante este proceso surgen etapas intermedias en donde existe un crecimiento poblacional importante denominado “bono demográfico”. Se trata de un periodo en el cual los países cuentan con un gran número de personas

en edades laborales y pocos individuos en edad avanzada, lo que se traduce como una baja relación de dependencia (de Cosío, 2014). Si bien, es cierto que esta ventana de oportunidad puede ser usada con el propósito de impulsar las actividades económicas y reducir la pobreza; tiene fecha de caducidad porque con el paso del tiempo los jóvenes envejecen, se reduce la población económicamente activa y aumenta la necesidad de atención médica. Así, luego de este crecimiento existe una etapa de recesión y después se genera un periodo de estabilización en donde el crecimiento se mantiene, aunque en algunos casos disminuye (Coale, 1984; Lesthaeghe, 2014). Según esta teoría, la mortalidad se reduce antes que la fecundidad debido a las mejoras en el sistema de salud e infraestructura que, a su vez, son el resultado de un avance tecnológico.

A Warren Thompson se le atribuye el origen de esta teoría de inicios del siglo XX, aunque existen algunas controversias debido a que otros autores también la reconocen como propia. Cien años después es difícil saber quién fue el autor original, e incluso, de reconocerse, las controversias seguirían. Lo más importante es que durante todo este tiempo se ha acumulado evidencia que pone en duda su validez. Por este motivo, algunos autores sugieren que no se trata de una teoría, sino de una descripción del cambio demográfico que ocurrió en los países europeos, por lo que no sirve de parámetro para estudiar el comportamiento de otras naciones, un ejemplo de ello es el caso mexicano (de Cosío, 2014).

El inicio de la transición demográfica en México ocurrió en la década de 1930, su primer indicio fue la baja de la mortalidad vinculada a las mejoras sanitarias e implementación de programas de vacunación. Situación por demás relevante en un país que había experimentado una revolución armada y estaba en un proceso de estabilización. Sin embargo, este comportamiento de la mortalidad se explica con los postulados de Meslé y Vallin (1995), quienes argumentan que muchos de los avances médicos se adoptaron de naciones avanzadas. Gracias a estas condiciones favorables, la esperanza de vida comenzó a aumentar cada vez más y dio como resultado que la población creciera rápido debido a que las familias solían tener varios hijos.

Casi después de 50 años, en la década de 1980, el ritmo de crecimiento disminuyó debido a una baja en la tasa del número de hijos por mujer. En esta época se presentaron las condiciones descritas en el modelo de Coale (1973), que llevaron al descenso de la fecundidad: se tuvo acceso a los nuevos métodos anticonceptivos; se incentivó el deseo de tener menos hijos y; las familias pequeñas se empezaron a ver como una forma mejor de vivir. En estas condiciones dio inicio la reducción de las tasas de fecundidad, el rejuvenecimiento de la población femenina y se mantuvo

la tendencia de tener hijos antes de cumplir 30 años. Esto podría pasar desapercibido de no ser porque en otros países con niveles de fecundidad similares a México se observó un efecto contrario, es decir, el calendario de la fecundidad se retrasó (Figuroa Perea, 2010; García y Ordorica, 2010; Léridon, 2014).

Todos estos cambios generaron eco en la esfera social y cultural al interior de las naciones, y provocaron un cambio en el sistema de valores de las sociedades. Algunas de las transformaciones con mayor impacto fueron el aumento de la participación de las mujeres en la esfera económica y su grado de escolaridad, además se generaron cambios en los modos de relación entre las familias. Este aumento en la esperanza de vida generó nuevas perspectivas de desarrollo y privilegió la realización personal. Sin embargo, no todas las personas los llevaron a cabo. La población que habita en zonas urbanas comenzó a incorporarlos en su comportamiento; luego, se fueron difundiendo a otros lugares. Estas nuevas modalidades de vida no se desarrollaron por sí mismas y siempre están acompañadas de políticas públicas orientadas a la planificación familiar y el uso de anticonceptivos (Caldwell, 2006; Coale, 1984; Lesthaeghe, 2014).

En el caso mexicano, el Estado asumió un papel protagónico, no solo en materia de salud, sino a través de la promulgación de las Leyes Generales de Población (Miró, 1971; Miró, 2015). La primera de ellas fue de corte pro natalista y buscaba incentivar el aumento de la población; la segunda tenía un sentido inverso y su objetivo era frenar la expansión poblacional. Para lograrlo, el estado estableció programas de planificación familiar y metas de crecimiento que generaron algunos incentivos perversos como el de establecer metas de crecimiento sin claridad en los derechos humanos, lo que conllevó a abusos a los derechos reproductivos de las mujeres debido a la feminización de la reproducción (Figuroa, 2010; Guerrero, 2014).

Un aspecto positivo de la teoría de la transición demográfica es su contribución al desarrollo de otras teorías que explican aspectos específicos vinculados con el descenso de la mortalidad y la fecundidad. Por ejemplo, la transición epidemiológica surgió para explicar el cambio en los patrones de mortalidad. Su argumento principal es que las muertes por enfermedades infecciosas que solían ocupar el primer lugar se sustituyeron por aquellas ocasionadas por las crónico-degenerativas. Esto se explica porque muchos de los padecimientos que ocasionaban gran parte de los decesos en la población infantil se lograron controlar mediante la expansión de los esquemas de vacunación. De esta forma, la población logró sobrevivir por más tiempo, por lo que sus causas de muerte empezaron a ser consecuencia de enfermedades propias de la vejez o producidas por su

estilo de vida (Horiuchi, 1999; Omran, 1971). Sin embargo, esto no significa que el cambio se diera de manera uniforme, puesto que había poblaciones específicas que tuvieron un comportamiento distinto. Por ejemplo, Omran (1971) argumenta que las mujeres no pasaron de las enfermedades infecciosas a las crónico-degenerativas en un solo paso, pues al inicio de la transición epidemiológica los mayores índices de mortalidad se debían a complicaciones en la maternidad, pero al mejorar los sistemas de salud se logró controlar esta situación. Al igual que sucede en el caso de la transición demográfica, en esta teoría se establecen tres modelos que buscan ser el referente para clasificar a cada uno de los países según su grado de industrialización.

Hasta este momento ya se ha definido que es la *población* como categoría analítica y cuáles son las teorías con las que se intenta explicar su comportamiento. Sin embargo, en ambos casos siempre se alude a los agregados poblacionales, esto es, la misma teoría de la transición demográfica está pensada para explicar el cambio en la estructura poblacional de los países, pero no es útil para estudiar el comportamiento de grupos más pequeños. Por este motivo, en la siguiente sección se hace una reflexión en torno a qué es un área pequeña, cómo se puede definir y si es sinónimo de una subpoblación.

1.4 Subpoblaciones

Existe un vacío conceptual en cuanto a cómo definir a los subgrupos que se unen para dar forma a la población, ¿se trata de subpoblaciones o áreas pequeñas? En la literatura demográfica no existe una definición unánime de qué es una *subpoblación*, aún y cuando el término se suele utilizar de manera regular para hacer delimitaciones conceptuales. De acuerdo con Bryant *et al.* (2019), las subpoblaciones son “una combinación única de categorías... [que] forman parte de un grupo más amplio” (Bryant *et al.*, 2019:45). Cuando se somete a escrutinio esta propuesta, la primera observación es que omite la referencia a la ubicación geográfica, si no a la pertenencia a un grupo de mayor tamaño. Esto significa que los individuos que conforman una subpoblación pueden estar dispersos en un territorio y no compartir una zona en común, pero si este es el caso, ¿es posible analizar su comportamiento y, sobre todo, asegurar que tienen características similares? Probablemente no, esto se observa en los estudios de la migración, en donde las personas suelen modificar sus actitudes para adaptarse a su nuevo contexto y dejan de lado las características que los identificaban como parte de un grupo en particular. Por ejemplo, muchos de los migrantes que van a las grandes ciudades suelen dejar de hablar su lengua para evitar discriminación y garantizar

su inclusión en su nuevo lugar de residencia, lo que significa que pierden las características que los definían como parte de un subgrupo (Arroba, 2009).

Sin embargo, existe otro concepto que toma en cuenta tanto a la población como a la dimensión geográfica y se denomina “áreas pequeñas”. De acuerdo con Molina (2019), se trata de una combinación de delimitaciones político-administrativas en las cuales no se tienen datos de la población en particular. Este concepto presenta el inconveniente que no es un enfoque, sino una forma de nombrar a un conjunto de técnicas estadísticas para llevar a cabo la estimación de un valor. Además, no hace referencia al número de personas sino a la falta de información, por lo que es factible asumir que países enteros pueden ser susceptibles de ser catalogados como “áreas pequeñas” si se cumple el criterio de la falta de información (Rao y Molina, 2015).

Otra observación en torno a esta suerte de conceptualizar a los subgrupos es que presupone una evaluación previa de los datos disponibles. Esto va en dirección opuesta a lo que suponen varios autores sobre cómo llevar a cabo una investigación mediante datos empíricos, en un escenario ideal “la teoría precede a la ratificación” (Prada, 1987:318). Esto significa que es necesario contar con un marco conceptual que establezca las pautas de lo que capturar, cómo hacerlo y, sobre todo, cuáles serán las delimitaciones que se tendrán al momento de aprehender algún fenómeno. Si bien es cierto que este camino parece razonable porque permite dar sustento a los datos, en el caso de las *áreas pequeñas* se propone lo contrario, es decir, partir de los datos disponibles, evaluarlos y luego determinar si se aplica el concepto o no (Molina, 2019; Rao y Molina, 2015).

En lo que están de acuerdo ambas propuestas es que los subgrupos se configuran a partir de ciertas características específicas, sin embargo, bajo el criterio de que *población* es una categoría analítica que incluye a la dimensión geográfica y espacial, esta última está ausente en las definiciones anteriores; de acuerdo con Salles (2003:106) “no es solo un parámetro, [sino] una propiedad misma de los fenómenos”. Así, se podría asumir que una *cohorte de nacimiento* es una subpoblación por sí misma, en donde el criterio de selección es el tiempo y lo que comparten cada uno de los miembros es que tienen un origen similar, sin importar que el resto de las características sean distintas.

Existe otro enfoque demográfico que algunos autores denominan *estimaciones de población subnacionales* (Namboodiri, 1972; Swanson, 2012; Wardrop *et al.*, 2018). A diferencia de los enfoques anteriores en donde la delimitación geográfica se establece en función del

fenómeno que se analiza, en este caso se fijan las delimitaciones administrativas. Aunque no es una regla, sí obedece a la forma en la que se han llevado a cabo diversos estudios; uno de los motivos para operar de esta manera es que la información disponible se colecta a partir de dichas delimitaciones administrativas.

Luego de poner a debate estas dos vertientes y analizar sus características, resulta positivo hablar de subpoblaciones por dos motivos: primero, porque al pensar en *áreas pequeñas* se traslada el problema inicial que supone la homogeneidad de las poblaciones a una escala menor, por ejemplo, si en lugar de considerar el análisis de un nivel nacional, se considera un nivel local, el investigador podría pensar que los resultados serán más precisos, sin embargo, la realidad es que aún en ese nivel, la falacia de los promedios sigue operando. Esto no significa descartar la dimensión geográfica, sino incorporarla como parte de la definición misma, por lo tanto, para hablar de subpoblaciones es necesario identificarlas como un grupo de personas que comparten coordenadas espaciotemporales, es decir, características demográficas y espaciales similares en un momento y lugar determinado.

Para cerrar esta sección conviene hacer una reflexión en torno al quehacer demográfico, muchas de las prácticas que se llevan a cabo hoy en día responden a los criterios que el gremio acepta como válidos. De acuerdo con González (2018), estas comunidades del conocimiento, a las que denomina epistémicas, están constituidas por un conjunto de expertos que comparten conceptos, teorías e ideologías. De esta manera, para que una investigación sea tomada como válida debe apearse a los principios que establecen estas comunidades (Villoro, 1996).

1.5 Estudio formal de la población

La Demografía es la disciplina social cuyo objetivo es estudiar a la población en términos de su estructura, volumen, distribución y cambios en el tiempo. El primero en hacer referencia a esta disciplina es Achille Guillard en 1855, sin embargo, sus bases científicas datan del siglo XVII con los trabajos de Graunt y Halley (Caldwell, 1996). Este último es mejor conocido por sus aportaciones a la astronomía y en particular al cálculo de la órbita del cometa que hoy en día lleva su nombre; más allá que esta observación resulta peculiar, sirve de contexto para reconocer el origen de la disciplina y su relación con las ciencias naturales. En aquella época ya se hablaba de la *aritmética política* para nombrar el estudio de los seres humanos, por lo que algunos autores,

como Vallin (1994:3), refieren que con la Demografía hubo “un bautizo”, o renombramiento, más que el nacimiento de un nuevo campo científico.

Algunos detractores de la Demografía señalan que no es una disciplina científica, dado su origen difuso, y argumentan que adolece de teoría y recurre a otras ciencias para lograr explicar los fenómenos que le competen. En parte esto resulta cierto, debido a que mucha de su metodología y conceptos se retoman de disciplinas como la Economía y la Estadística, lo cual le impide conceptualizar a su objeto de estudio según sus propios criterios y únicamente tiene la capacidad de hacer una descripción matemática de sus características. Sin embargo, el recibir apoyo por parte de otras disciplinas puede considerarse como una fortaleza y no como una debilidad. Por ejemplo, la vinculación que logra le permite establecer un diálogo más amplio con otras áreas de conocimiento y enriquecerse de su acervo (Petit, 2019). De esta forma se puede explicar que existen varios subcampos que operan con criterios y metodologías particulares, aunque todos comparten la premisa que es posible analizar, y predecir, el comportamiento de los grupos humanos.

Si se deja de lado a la demografía como disciplina científica y se analiza como una profesión, su carácter multidisciplinario resulta desfavorable. El primer punto a considerar es que se trata de un gremio heterogéneo, en donde es usual que sus integrantes se formaron en áreas distintas. Esto los hace privilegiar enfoques diferentes, lo que crea ciertas discrepancias, quizás la más sonada es aquella que gira en torno a cómo analizar los fenómenos poblacionales, de forma cuantitativa o cualitativa.

Ahora, resulta conveniente llevar la discusión a un plano distinto y reflexionar cómo obtiene legitimidad el conocimiento que generan los demógrafos. Al igual que sucede en otras áreas, este conocimiento es validado por un grupo de expertos quienes determinan su autenticidad y pertinencia, sin embargo, en el gremio demográfico existe una particularidad más, como comunidad epistémica tiene un peso importante (González, 2018; Villoro, 1996). Este factor se hace visible en las Conferencias Internacionales sobre Población y Desarrollo, convocadas por la ONU, en donde se reúnen expertos, líderes políticos y organizaciones de la sociedad civil para discutir sobre los retos y necesidades que enfrenta la humanidad, y cuyo resultado se traduce en una serie de recomendaciones sobre los programas y políticas de población.

Dichas recomendaciones son explícitas y se materializan en objetivos concretos, en los que se exhorta a los países miembros a desarrollar estrategias que coadyuven a su logro. Un primer conjunto de recomendaciones se planteó en el año 2000 y se nombraron los Objetivos de Desarrollo

del Milenio (ODM), los cuales son ocho propósitos encaminados a reducir algunos de los problemas que más aquejan a la humanidad, de acuerdo con la visión de la ONU, por ejemplo: la pobreza extrema, la igualdad de género y la necesidad de aumentar la matrícula escolar en beneficio de la educación de cada región (de Rentería *et al.*, 2019; Gil, 2018).

Para 2015, se buscó dar continuidad a la iniciativa de los ODM y se implementó la Agenda 2030 en donde se plantearon 17 objetivos, denominados Objetivos de Desarrollo Sostenible (ODS), en ellos el foco de atención estaba en el desarrollo de las naciones, pero en armonía con el medioambiente. Además de contribuir a cambios en la esfera política, estos objetivos influyen en el rumbo que toman algunas investigaciones demográficas, puesto que el componente central de la Agenda es el mismo que el de la demografía: la población (Gil, 2018).

Si bien es cierto que la visión de la ONU, vía los ODS, consiste en promover objetivos comunes que permitan el desarrollo global, existen algunas críticas que ponen en duda su viabilidad a largo plazo. Una de ellas gira en torno a que las metas planteadas son idealistas e irrealizables, pues se fundamentan en acuerdos incumplidos en años anteriores. Aunque en el discurso retórico se habla de dar continuidad a dichos acuerdos, es válido asumir que en realidad el objetivo es tratar de dar seguimiento a los compromisos incumplidos. Otro punto que genera controversia, y que no es una crítica, refiere a que los ODS fueron generados con una visión amplia y no únicamente se enfocan en analizar el comportamiento de los grandes agregados poblacionales. Su interés radica en dar cuenta de las diversas realidades que se encuentran al interior de los países, por lo que la falta de datos desagregados presenta un desafío importante. Aunque el reto parece ser metodológico, los primeros diagnósticos que se hicieron sobre la disponibilidad de información muestran que el problema va más allá de cómo se recolectan los datos, por lo que atañe a la esfera teórica y política (Gil, 2018).

1.6 Conclusiones

A lo largo de este capítulo se discutió el concepto de *población* como categoría analítica. Para ello, se analizó su origen y cómo ha evolucionado a lo largo del tiempo hasta llegar al concepto que se utiliza hoy en día y que es la referencia para estudiar a los grupos humanos. Una de las situaciones que se enfatizó a lo largo de este capítulo y que forma parte de la reflexión final es que esta categoría

es útil para quitar las diferencias que existen entre los individuos y considerarlos como una sola entidad que puede ser analizada. Lo cual es favorable para llevar a cabo el análisis estadístico, pero resulta contraproducente, ya que oculta comportamientos particulares que no son revelados porque se privilegia el comportamiento grupal. Sin embargo, una de las situaciones que aún sigue en duda es hasta qué momento un grupo de personas ya puede ser considerado como un cuerpo único y susceptible de ser analizado.

Este recuento histórico de lo que es el constructor denominado como población sirvió de referencia para introducir las teorías poblacionales. Estas son un conjunto de principios y criterios que establecen la forma en que se deben analizar a los grupos humanos, lo que más adelante dio paso a la consolidación de la demografía como ciencia encargada de estudiar a la población. Sin embargo, uno de los vacíos conceptuales que aún existen es determinar que son las subpoblaciones o las áreas pequeñas, ya que dependiendo el enfoque se tendrá diversas definiciones. Un argumento para explicar esta situación es el hecho de que las prácticas que se desarrollan en el marco de la demografía responden a criterios que el gremio acepta como válidos. Uno de ellos es entrar énfasis a la evolución temporal de la población. Por este motivo, el capítulo concluye con una reflexión en torno al quehacer demográfico y algunas de las controversias que existen al respecto, como el hecho de que se privilegia la técnica, pero se adolece de teorías, las cuales se retoman de otras disciplinas como la sociología.

Capítulo 2

La demografía y el espacio

2.1 Introducción

La complejidad del mundo contemporáneo requiere de acercamientos multidisciplinarios que ayuden a la explicación de ciertos fenómenos, en los que, por sí solo, un campo temático quedaría limitado a sus paradigmas teóricos y metodológicos. Por ello, ante la necesidad de contar con mayores elementos explicativos, los científicos sociales dialogan y desdibujan las fronteras de grandes áreas temáticas. En este contexto, las siguientes páginas ofrecen una reflexión de la relación entre la demografía y el espacio. La pregunta que guía la exposición de ideas es cuestionar ¿qué resultados se pueden obtener de las investigaciones demográficas que dan mayor peso a la dimensión espacial? La intención es centrar el análisis en la particular manera en que la demografía, explica los fenómenos de su interés, sin considerar al espacio.

Para hacer visible la importancia entre demografía y espacio, el capítulo inicia con la incorporación de la perspectiva del espacio, vista como una concepción de utilidad para dar contexto a los datos poblacionales. Se argumenta como el énfasis en predicciones recurre a un enfoque temporal que desdeña a la dimensión espacial. Revela, a través de ejemplos, el beneficio de incorporar la dimensión geográfica. Y agrega que, el uso de este enfoque también conlleva contrariedades que subyacen a la teoría de la transición demográfica.

Con respecto al enfoque teórico, analiza que la transición demográfica ha sido la principal perspectiva para estudiar la composición y los cambios en las estructuras poblacionales, sin embargo, se destaca el menoscabo de referentes contextuales que son inherentes a las poblaciones humanas, paradójicamente, siendo este último, el principal interés de la demografía. Así que retoma, el avance de algunos científicos para mostrar que el acompañamiento del espacio a la fecundidad permite darle continuación a un proceso humano.

Con respecto al enfoque metodológico, el apartado postula que las conciliaciones entre demografía y espacio debe estar presente al abordar problemas relacionados con las escalas, lo cual implica un cambio en la forma en la que se capta y analiza algún fenómeno demográfico. También sostiene que la controversia sobre las escalas y sus características tiene como objetivo hacer inferencias válidas sobre el comportamiento de la población.

Con respecto a las falacias referentes a la extensión y magnitud, se plantea que uno de los problemas relacionados con el inferir comportamientos usando datos georreferenciados es la posibilidad de falsear los resultados, ya que se puede transferir propiedades de los agregados a propiedades individuales. También señala que las delimitaciones geográficas distorsionan la percepción de los fenómenos, toda vez que son constructos delimitados a partir de lo político-administrativo y no de la población.

Para seguir con esta discusión, el texto utiliza las islas como un ejemplo a la delimitación natural, en la que la intervención humana para demarcar un territorio es inexistente y que, por lo tanto, su comportamiento demográfico es singular, debido a que no corresponde a versiones reducidas de grandes Estados, sino que la población está influenciada por características sociales y económicas propias a su contexto y necesidades.

Finalmente, el capítulo concluye afirmando que la admisión del espacio al quehacer demográfico posibilita que las delimitaciones geográficas, sociodemográficas y contextuales acoten el comportamiento de grupos particulares o subpoblaciones, en lugar de aspirar al estudio de la población total. No obstante, en la disciplina demográfica, el concepto subpoblación figura muy poco, pese a que es un término usado de manera frecuente para limitar los análisis. Así que sugiere cruzar los límites tradicionales de la demografía para llenar de contenido a este concepto que sí está presente, con un mayor desarrollo, en otras disciplinas como la sociología, incluso en la estadística bajo el nombre de áreas pequeñas, que, aunque no es concepto en sí mismo, hace referencia a un criterio. De tal suerte que, bajo la idea común de que los subgrupos se configuran a partir de ciertas características específicas, es posible que al incorporar al espacio como una herramienta más de análisis en la agenda demográfica para que se actualice y amplíe.

2.2 La incorporación del espacio

La concepción de *espacio* que se emplea en la forma de concebir algún fenómeno social resulta crucial para cualquier investigación, sin embargo, pocas veces se reflexiona sobre su naturaleza. En las ciencias sociales se tiende a interpretar que *espacio* es un área o proporción de la superficie terrestre que sirve de contenedor para el desarrollo de las personas y se puede localizar mediante coordenadas geográficas (Ramírez; López, 2015). Sin embargo, esta definición ha sido cuestionada

por disciplinas como la sociología o la geografía misma, que han argumentado que el espacio es, además, una categoría llena de símbolos y significados.

A partir de estas posturas, algunos investigadores han establecido que *espacio* es una estructura que afecta el comportamiento humano y además es un escenario donde ocurren los fenómenos (Ortega, 2000). Sin embargo, pese a estos argumentos a favor de hacer más específica y útil esta categoría, algunas disciplinas, como la Demografía, se han mantenido al margen de la concepción tradicional de *espacio* y consideran que es una dimensión homogénea y plana (Ramírez; López, 2015).

Esta tendencia y carencia de visión proviene del momento mismo en que la Demografía como disciplina comenzó a hacerse preguntas de investigación. Al igual que la ciencia, que sigue un método que contempla el momento histórico, el lugar desde donde se la piensa y la ideología de quienes la conciben, la Demografía, como disciplina científica, sigue un enfoque positivista desde el que analiza la evolución de las poblaciones humanas y, con base en datos, trata de predecir su comportamiento. Esta preocupación por generar pronósticos del comportamiento poblacional en las investigaciones, fomentó el privilegio del enfoque temporal encima de la dimensión espacial, pese a su importancia (Voss, 2007). En este sentido, Alho (2005), señala que el gremio académico considera como *demógrafo* a la persona que usa el diagrama de Lexis, el cual sirve para mostrar los cambios temporales que le ocurren a un grupo de personas. Bajo este supuesto, los profesionales de esta disciplina deben ser capaces de hacer predicciones sobre el comportamiento de los fenómenos (Pressat, 1967; Véron *et al.*, 2002).

Ante esta situación conviene preguntarse ¿qué resultados pueden obtener las investigaciones demográficas que dan mayor peso a la dimensión espacial? De acuerdo con Bell (2015), Castro (2007), Matthews y Parker (2013), la incorporación del componente geográfico coadyuva a comprender la difusión de los fenómenos a lo largo de un territorio. Por ejemplo, al observar los patrones reproductivos se hace visible que la ubicación geográfica de una comunidad favorece la adopción de determinadas ideas que se traducen en cambios en los niveles de fecundidad. Además, el espacio también es de utilidad para dar contexto a los datos poblacionales; por ejemplo, una alta escolaridad puede estar ligada a la presencia o ausencia de planteles educativos. Por lo que, la dimensión geográfica favorece las investigaciones al incorporar nuevas variables explicativas a los modelos econométricos.

Sin embargo, la inclusión de la dimensión espacial en el quehacer demográfico conlleva problemas de diversa índole. En el terreno teórico, demanda de una revisión de los postulados que subyacen a la teoría de la transición demográfica, principal referente para estudiar los cambios en la estructura y composición de las poblaciones humanas. En la cuestión metodológica, los argumentos apuntan en varias direcciones, a saber: 1) los problemas relacionados con las escalas (Manson, 2008); 2) las falacias al momento de inferir comportamientos usando datos georreferenciados (Guerra, 1977); y 3) reconocer que las delimitaciones geográficas distorsionan la percepción de los fenómenos (Bell, 2015).

Por estas razones, la discusión teórica y el punto de vista metodológico serán puestos sobre la mesa de análisis en las siguientes páginas.

2.3 Enfoque teórico

Cuando dentro de algún cuerpo de conocimiento previo o estado de la cuestión de una disciplina surgen problemas de investigación, estos deben evidenciarse dentro de los parámetros de la comunidad científica a la cual está adherida, debido a que esta le brinda legitimidad y validez (Kuhn, 1993; Güell, 1985). Cuando el resultado postula teorías que están inscritas en un contexto social e histórico, se puede generar un paradigma, el cual modela la forma de pensar de sus agremiados. Sin embargo, si una teoría amparada en un paradigma no satisface los niveles explicativos, se producen cambios radicales en las maneras explicativas y deviene una revolución científica (Kuhn, 1993). En el caso de la Demografía, esta teoría es la de la transición demográfica porque es el referente inmediato con que cuenta un investigador para estudiar la composición y los cambios en las estructuras poblacionales.

Aunque la teoría de la transición demográfica ha dominado el pensamiento académico durante el último siglo porque en ella se postula que las poblaciones humanas experimentan una transición como resultado de la modernización y las mejoras sanitarias que las conduce de niveles altos de mortalidad y fecundidad a niveles bajos y controlados (Bongaarts, 2006; Overbeek, 1974, Caldwell, 2006; Coale, 1984); en los últimos años, estos planteamientos han generado gran controversia y se ha puesto en duda su validez, sobre todo cuando se fija la atención en lugares diferentes a Europa (Cosio, 1992).

Debido a que las poblaciones humanas no solo dependen de la interacción entre mortalidad y fecundidad, sino también de “los contextos históricos y espaciales” (Cosio, 1992:14); adquirir un pensamiento espacial permite ahondar en cuestiones relacionadas con la distribución de los fenómenos y, en el caso de los cambios, observar y analizar los procesos de difusión que se desarrollaron en un determinado territorio.

Para hacer visible la importancia de la dimensión espacial en la teoría de la transición demográfica, conviene retomar lo que han avanzado algunos investigadores, por ejemplo, la Primera Ley de Geografía de Tobler, fórmula que “todas las cosas están relacionadas entre sí, pero las cosas más próximas en el espacio tienen una relación mayor que las distantes” (Waters, 2018). Esta idea de que la distancia entre una región y otra influye en la difusión de los comportamientos conecta con los postulados de Cosio (1975), quien sugiere que la reducción de la fecundidad se dio primero en las zonas urbanas y luego se propagó a otros lugares.

A este respecto, Bocquet-Appel y Jakobi (1996) también siguieron esta idea de difusión espacial y analizaron el comportamiento de la fecundidad en Europa occidental. Usaron los datos de la tasa de fecundidad de mujeres casadas para evaluar si hubo barreras que evitaran la difusión de los comportamientos reproductivos. Sus resultados muestran que en las regiones con mayoría católica los cambios se produjeron más rápido en comparación con sus pares protestantes.

En el caso de América Latina, Bocquet-Appel y Jakobi (1996) hacen referencia a las metrópolis como los puntos de difusión (Cosio, 1997). Esto explica por qué los niveles de fecundidad en estas zonas fueron similares a los de los países europeos, mientras que en las zonas rurales la fecundidad permaneció elevada (Schkonik y Chackiel, 1997). Como una forma de ilustrar esta situación es la evolución de la tasa de fecundidad en México, de acuerdo con Cosio, (1997) no fue unánime y que en aquellas regiones en donde los niveles eran altos, la reducción fue paulatina. De esta manera, la idea de una reducción unánime se complementa con una noción de difusión de los procesos, de tal manera que se demuestra que la cercanía —o lejanía— de los lugares tuvo un impacto en el comportamiento de la fecundidad.

Ante lo anterior es necesario señalar que pese a que la transición demográfica tiene un sitio de privilegio en la Demografía para estudiar a la población, su enfoque es temporal y muchos de los análisis realizados a partir de ella buscan predecir los cambios y anticipar la evolución de los fenómenos poblacionales. Para enriquecer este enfoque, dar importancia al espacio y reconocer su presencia en el desarrollo de los fenómenos, la transición demográfica también puede verse como

“un proceso de difusión continuo de uno o varios puntos localizados no muy lejos uno de otro” (Bocquet-Appel y Jakobi, 1996:127), ya que la cercanía dentro de una población desempeña un papel fundamental para la propagación de las ideas, lo que puede propiciar la adopción de nuevos comportamientos.

2.4 Enfoque metodológico

La incorporación del espacio en la metodología demográfica modifica la forma en la que un fenómeno se capta y se analiza. Un primer elemento a tomar en cuenta son las *escalas* que, en términos generales, se definen como “constructos teóricos que sirven para medir y estudiar los fenómenos sociales” (Ruiz y Galicia, 2016). Para algunos autores, las escalas son naturales e independientes de las observaciones, mientras que otros sostienen que son construcciones forzadas, social e inevitablemente sesgadas. Manson (2008) contrasta ambos argumentos y concluye que, más allá de las controversias en torno a su naturaleza, las escalas son entidades epistemológicas que tienen un efecto en la investigación científica porque son capaces de construir objetos de estudio (Gibson, 2000).

El cambio de escala tiene un efecto en la manera de aprehender y analizar los fenómenos sociales. De acuerdo con Ruiz y Galicia (2016), el primer elemento de las escalas es la resolución, la cual representa la unidad de observación más pequeña, es decir, “la unidad de medida por debajo de la cual no hay una mayor heterogeneidad, o la que existe no es de interés” (Ruiz y Galicia, 2016:140). El segundo componente es la extensión, definida como la dimensión o tamaño del área de estudio. El tercero es el nivel, categoría que usa el observador geográfico para ubicar y clasificar un fenómeno dentro de una jerarquía. Este último hace referencia a “un sistema de agrupación de objetos o procesos (conceptual o causalmente ligados) en una escala analítica” (Gibson *et al.*, 2000: 218).

Un ejemplo del impacto que tienen las escalas geográficas en la construcción de los objetos de estudio se muestra con la ecuación compensadora. Esta relación matemática es el pilar de muchos cálculos demográficos, debido a que expresa el cambio de la población en el tiempo (Keyfitz y Caswell, 2005; Pressat, 1967; Véron *et al.*, 2002):

$$P_t = P_0 + N(0, t) - D(0, t) + I(0, t) - E(0, t) + \varepsilon$$

Al analizar con detenimiento esta expresión, queda claro que se rige por un criterio temporal, pues se compara el total de población en dos momentos distintos, al inicio del periodo P_0 y al final P_t , con ello se pretende hacer visibles los cambios en la estructura poblacional. A la diferencia entre los nacimientos (N) y las defunciones (D) se le suele denominar crecimiento natural, mientras que a la resta de los inmigrantes (I) y emigrantes (E) se les considera migración neta. ¿Qué sucede si se cambia la escala?

Rogers (2015) propone una forma de introducir la dimensión espacial mediante una desagregación regional. Para este autor es insuficiente hacer referencia a los movimientos migratorios en la ecuación compensadora e incluir el componente geográfico de manera explícita a menor escala, para lo cual propone modificar la ecuación anterior de la forma siguiente:

$$P_u^{(t)} = [1 + (N_u - D_u) - o_{uv}]P_u^{(0)} + o_{vu}P_v^{(0)} + \varepsilon$$

En esta fórmula se mantiene el elemento temporal y se compara a la población en dos momentos en el tiempo $P_u^{(t)}$ y $P_u^{(0)}$, pero se incorpora la dimensión espacial denotada por los subíndices u y v . De esta manera, los nacimientos (N) y las defunciones (D) también toman en cuenta la ubicación geográfica, dando como resultado que el componente migratorio sufra una variación. Ahora, se trata de una razón que expresa el porcentaje de población que cambia de la región u a la v y viceversa, con lo cual se tiene una visión más acotada.

Al pensar en cómo afectan las escalas, la manera de captar los fenómenos sociodemográficos es inevitable preguntarse cuál es la mejor escala para aprehender los cambios en la población. El debate sobre las escalas y sus características tiene como objetivo hacer inferencias válidas sobre el comportamiento de la población. Saber cuál es el nivel y desde qué posición se deben estudiar los fenómenos es una de las discusiones más añejas de las ciencias sociales y también está presente en los estudios demográficos (Salles, 2001). La cuestión central es determinar si el comportamiento social corresponde a la estructura o a las decisiones individuales.

Retomando el ejemplo de la transición demográfica, lo anterior es equivalente a preguntarse si la decisión de limitar el número de hijos depende del contexto histórico y social o es el resultado

de las decisiones individuales (Cosío, 1995). El inconveniente aquí radica en que la población es, por definición, un agregado de individuos que constituyen entidades en un nivel macro, por lo que privilegiar un enfoque micro modifica la esencia de la demografía (Billari, 2006; Lee, 2001).

2.5 Falacias referentes a la extensión y magnitud

El hecho de considerar un nivel u otro conlleva a incurrir en falacias y poner en duda la validez de las inferencias. Una de ellas es la denominada falacia ecológica, la cual consiste en transferir propiedades de los agregados a propiedades individuales. Brown y Guinnane (2007) analizan la transición demográfica en Europa e ilustran esta situación de la siguiente manera: “[si analizamos] la proporción de la población que es católica y la proporción urbana, las estimaciones no nos dirían nada sobre si los habitantes católicos de las ciudades tienen una fertilidad mayor o menor que los habitantes de las ciudades protestantes [ya que] no podemos afirmar que los datos agregados puedan recuperar efectos individuales (traducción propia)” (Brown y Guinnane, 2007:580).

Otro ejemplo que ilustra de buena manera esta situación es la interpretación de la esperanza de vida al nacimiento. Este es uno de los indicadores más importantes que aporta la demografía a la sociedad y al desarrollo de políticas públicas (García y Ordorica, 2010; 2012; Partida, 2018). Si se analiza con detenimiento, el dato que se ofrece es a una escala nacional, por lo que no significa que los individuos vivan un número de años en particular, sino que, como grupo, esa es la edad a la que pueden llegar (Partida, 2015). Una investigación de este tipo también puede arrojar el caso opuesto y propiciar que se infieran comportamientos grupales a partir de un análisis a nivel individual (Landman, 2003). Esta situación la advierte Lee (2001) y señala que al fijar la atención en el nivel individual se pierde de vista la generalización de los procesos y su dinámica, propiciando una falacia.

Otra falacia en la que se puede incurrir es la contextual, la cual se refiere a un tipo de análisis que pretende explicar características o comportamientos individuales a partir del contexto o entorno en el que se presentan (López, 1977). Esto significa que se asumen comportamientos generales a partir de las características históricas o sociales de una unidad territorial. Nuevamente, un ejemplo de ello lo proporcionan Brown y Guinnane (2007), quienes advierten que es incorrecto asumir que

a partir de la segunda mitad del siglo XX se inició el descenso unánime de la fecundidad en Europa. En algunas regiones esto no sucedió, pero el contexto indicaba que era una realidad incuestionable para todas las regiones.

Además de los problemas metodológicos anteriores, Brown y Guinnane (2007) plantean dos, relacionados con el espacio al momento de demostrar empíricamente la teoría de la transición demográfica. El primero se vincula con el problema del área modificable (MAUP por sus siglas en inglés), el cual se refiere al efecto del tamaño de las unidades de estudio en la correlación que tienen los datos. Cuanto mayor sea la serie de unidades de estudio, mayor será el coeficiente de correlación y en consecuencia se puede concluir que dos fenómenos están ligados aun y cuando en realidad no lo estén. De hecho, este es el argumento central que usan Brown y Guinnane (2007) para poner en duda los resultados del proyecto de Coale (1986), que buscó dar sustento empírico a la teoría de la transición demográfica planteada por Notestein en la década de 1930. Los investigadores encontraron que las mediciones de la fecundidad son sensibles en su construcción porque dependen de los niveles de agregación y de las combinaciones de unidades contiguas (Brown y Guinnane, 2007; Wong, 2003).

Autores como Openshow (1979), Wong (2004), Griffith *et al.* (2003) o Flowerdew (2011) refuerzan la idea de que los resultados que se obtienen al analizar los datos dependen de la forma en que se agrega la información. Este último examinó el impacto de la agregación en los resultados del censo de Inglaterra en 2001 a través de correlaciones y encontró que pocas variables se ven afectadas por el cambio de escalas. Sin embargo, este efecto crece si se consideran todas las combinaciones que se pueden hacer con los datos que sí muestran un efecto negativo. Una vez más queda claro que los resultados están condicionados por el tamaño de las zonas geográficas que se toman en cuenta (Chen, 2017). Flowerdew (2011) ofrece una serie de recomendaciones al final de su estudio que se deben de tomar en cuenta al momento de analizar los datos censales. Una de ellas es cuestionarse si “¿los resultados son afectados por las delimitaciones de las áreas de estudio?” (Flowerdew, 2011:117). Esta pregunta es de utilidad al momento de analizar los datos. No obstante, quizá antes deba responderse una pregunta más general: ¿cómo se establecen los límites?

Algunas investigaciones sostienen que los límites se definen por criterios administrativos del Estado (ONU, 2010), y hay quienes consideran que lo correcto es que las mismas características de la población sean las que fijen dichas delimitaciones (Chávez *et al.*, 2016). Sin embargo, tal y como apunta Logan (2016), es frecuente que en la mayoría de los estudios sociodemográficos se

retomen los límites político-administrativos instituidos, ya que suelen ser las unidades de referencia para recolectar la información. Esta situación no está exenta de críticas; por ejemplo, Lichter y Ziliak (2017) observan las desventajas de dichos límites, pues consideran que en la mayoría de los casos están sujetos a las intenciones de quienes los crean, por lo que tienen un fin político y no demográfico.

El estudio de Chávez *et al.* (2016) es otro ejemplo del impacto de las delimitaciones de las unidades administrativas en los fenómenos sociodemográficos. En él, los investigadores analizan el flujo migratorio que se tiene en 18 ciudades de 6 países latinoamericanos, y encuentran que no existe un comportamiento homogéneo en todas estas regiones. Si bien es cierto que este resultado parecería contundente, enfatizan que en los países estudiados el saldo neto migratorio depende de los límites de las regiones. Esto significa que al modificar dichas fronteras el valor cambia, lo que pone en duda el comportamiento del fenómeno migratorio. Ante la incapacidad de solventar esta situación, los autores toman como referencia el área metropolitana en un intento por considerar las características demográficas de la población.

Ante lo anterior, Chávez *et al.* (2016) y otros investigadores concluyen que, al mantener las delimitaciones administrativas establecidas por los países, es posible hacer comparaciones a lo largo del tiempo (Silvey, 2004). Sin embargo, de acuerdo con la ONU (2014), si existen modificaciones en los límites, se puede generar ambigüedad en los datos registrados. En este sentido, optar por límites propios impide que los ejercicios se puedan replicar en los siguientes años, pese a que estas delimitaciones sean más congruentes con el desarrollo de los fenómenos (Chávez *et al.*, 2016).

No obstante, las delimitaciones administrativas cambian a partir de decisiones políticas y esto genera consecuencias negativas para la comparación a largo plazo. Por ejemplo, en el estudio de Silván-Cárdenas *et al.* (2015) se trató de comparar el crecimiento de las AGEBS urbanas de la Ciudad de México, tomando como referencia la información censal de 2000 y 2010. En el proceso se encontró que los límites se habían modificado y, aunque los autores intentaron unir los datos para hacerlos comparables, sin embargo, en algunos casos fue imposible lograr una armonía, por lo que optaron por descartar regiones.

El establecimiento de los límites geográficos para un levantamiento de datos está relacionado con el objetivo principal: obtener información. Por lo tanto, un mayor nivel de desagregación coadyuva a mejorar la comprensión de los fenómenos, tal y como lo manifiestan las

recomendaciones de las políticas internacionales y las recomendaciones de la ONU (2010) para la generación de estadísticas poblacionales, como la Agenda 2030 que dispone que uno de los criterios para la generación de datos debe ser el desglose de estadísticas a nivel subnacional. El argumento es que únicamente de esta manera se logra dar cuenta de grupos con características particulares (Desa, 2016).

2.5.1 El caso de las islas

¿El tamaño de las regiones tiene un efecto en el comportamiento demográfico de las personas? Para responder esta interrogante, conviene poner el foco en un tipo particular de regiones consideradas pequeñas y con una delimitación natural, lo que disipa la discusión acerca de cómo establecer los límites geográficos. Estas regiones son las islas, cuyas delimitaciones territoriales están dadas de forma natural, sin intervención humana. Algunas de ellas son colonias o su administración política depende de otras naciones, por lo que pueden considerarse como subpoblaciones, ya que sus habitantes son miembros de un estado más grande.

Desde los estudios llevados a cabo en los años treinta y cincuenta, ya se advertía que la población que vive en las islas tiene un comportamiento demográfico particular. Un ejemplo de ello lo muestran Mauldin y Berelson (1978), quienes señalan que las tasas de fecundidad se redujeron en las regiones insulares antes que en los países en desarrollo. Esto sugiere que en las islas existe una mayor preocupación por encontrar un equilibrio entre los recursos disponibles y el tamaño poblacional, tal y como Malthus indicó cuatrocientos años atrás. De acuerdo con estos autores, los habitantes de las islas son más conscientes de su aislamiento, del tamaño de la población y de que la migración es una opción para buscar nuevas oportunidades de desarrollo.

También los estudios demográficos de Caldwell *et al.* (1980), abordan el comportamiento de las poblaciones en las islas como microestados. De acuerdo con estos autores, un cuarto de los miembros de la Organización de las Naciones Unidas son países con menos de medio millón de habitantes y en su mayoría son islas. Si se toma como referencia el número de personas que las habitan y la extensión de su territorio, las islas se pueden considerar como microestados que, en esencia, tienen una escala menor si se comparan con sus pares que se ubican en algún continente. Sin embargo, su comportamiento demográfico es particular y no corresponde a versiones reducidas de los grandes estados. Por ejemplo, sus niveles de migración son altos y en dirección hacia naciones desarrolladas o a regiones de las cuales alguna vez fueron colonia. Además, dentro de

estos territorios, la migración rural-urbana, tal y como se concibe en los grandes estados, es difícil y difusa de captar. Esto se debe a que no hay grandes diferencias entre lo que se podría considerar una zona rural y una urbana.

Estos microestados también presentan baja mortalidad, pues, debido a que cuentan con poca población, la infraestructura médica es accesible para una mayor cantidad de habitantes. Así, la mayoría de las islas del Pacífico tienen una esperanza de vida al nacimiento alrededor de los 60 años, un nivel encontrado en Europa occidental después de la Segunda Guerra Mundial y que está trece años por arriba del sureste asiático. Otro fenómeno que destacan Caldwell *et al.* (1980) es la disminución de la tasa de fecundidad. Su hipótesis supone que este comportamiento obedece al cambio económico y social a diferencia de lo que ocurre en otras partes del mundo, lo cual reduce la presión para que las mujeres contraigan matrimonio a temprana edad.

2.6 Subpoblaciones, áreas pequeñas o subgrupos

Ahondar en la incorporación del espacio en el quehacer demográfico tiene como objetivo poner de manifiesto que no se puede estudiar a la población en su totalidad. Por lo tanto, es necesario establecer delimitaciones geográficas y sociodemográficas para explorar el comportamiento de grupos particulares, también llamados subpoblaciones. Sin embargo, no existe en la literatura demográfica una definición unánime y explícita del concepto subpoblación, aun cuando el término se suele utilizar de manera regular para hacer delimitaciones o acotar los análisis.

De acuerdo con Bryant *et al.* (2019), las subpoblaciones son “una combinación única de categorías... [que] forman parte de un grupo más amplio” (Bryant *et al.*, 2019:45). Conviene hacer dos observaciones al respecto: primero, como la Demografía es una disciplina empírica que usa constructos medibles para llevar a cabo sus investigaciones, dichas categorías hacen referencia a todos aquellos rasgos que se captan en una fuente de información, es decir: edad, sexo, nivel de escolaridad (Lee, 2001; Pressat, 1967; Véron *et al.*, 2002). Por lo que la clasificación de las poblaciones se puede hacer de manera teórica sin ahondar en características subjetivas de las personas o la conexión de unas con otras.

El segundo punto está ligado con el hecho de que la Demografía busca dar cuenta de los macroprocesos, con la intención de analizar los cambios en la estructura, volumen y composición de la población (Billari, 2006; Lee, 2001). Bajo esta idea se pueden combinar dos categorías (edad

y sexo), para definir un subgrupo de la población; por ejemplo, el grupo de mujeres entre 15 y 49 años (el cual se usa como referencia para la tasa de fecundidad). En este caso se asume que el comportamiento es el mismo para todos los individuos del subgrupo, sin tomar en cuenta que en realidad pueden tener comportamientos distintos, lo que representa un problema al momento de analizar e interpretar los datos. Cosio (1997) y Yan (1986), en su estudio sobre el descenso de la fecundidad en Europa, demostraron que subpoblaciones similares tienen comportamientos diferentes y el origen de esas discrepancias se debe a su ubicación geográfica.

En la definición Bryant *et al.* (2019) existe otro problema: la omisión del componente geográfico de las subpoblaciones. Su ausencia significa que los individuos pueden estar dispersos en un territorio, pero, aun así, compartir las categorías de referencia que los podrían ubicar como parte de una subpoblación. En este escenario cabe preguntarnos si ¿es posible analizar su comportamiento y, sobre todo, relacionarlo con su pertenencia a una subpoblación? En los estudios de Caldwell *et al.*, (1980), sobre el comportamiento demográfico de las personas que habitan en islas, llegan a la conclusión de que la ubicación geográfica es el elemento clave para comprender su comportamiento, incluso cuando existen algunos casos particulares en donde los habitantes mantienen un vínculo político-administrativo e histórico con naciones más grandes.

Esto significa que resulta insuficiente la combinación específica de categorías que se proponen desde el enfoque demográfico para estudiar a estos grupos particulares. Con los ejemplos que presentan Cosio (1978), Caldwell *et al.*, (1980) está claro que para poder estudiar a grupos particulares es necesario incorporar el componente geográfico, lo que refuerza el argumento de Matthews (2016) en cuanto a que la “demografía es una disciplina que ha llegado tarde a explotar la dimensión espacial”, lo cual llama la atención debido a que el concepto mismo de población se configura a partir de una base geográfica, tal y como lo expone Hauser y Duncan (1959).

Como algunos autores como Pressat (1967) definen a la Demografía como el estudio estadístico de la población, para ahondar en la definición de las subpoblaciones, es menester acercarse a la Sociología y la Estadística, disciplinas muy ligadas. El vínculo es tan estrecho entre estas tres que hoy en día existen los estudios de población, una vertiente de los estudios demográficos, cuyo objetivo está dirigido a captar información sobre los comportamientos sociales.

En la Sociología se reconoce que los individuos se integran a través de grupos humanos que, a su vez, constituyen el pilar fundamental de cualquier sociedad. Dentro de estos grupos, los individuos adquieren el aprendizaje social necesario para participar en la vida colectiva y a través

de ellos configuran su identidad; además, el grupo al que se pertenece transfiere normas y valores para socializar y regular su comportamiento frente a sus pares (Ibáñez, 2004). El grupo social se constituye con relaciones recíprocas, conciencia de pertenencia, valores y creencias. Además, establece reglas que especifican cuál es o debería ser la conducta en el interior del grupo, las posiciones de sus miembros y su relación con las posiciones de otras personas. De acuerdo con Hare (1962), para que un grupo social se forme tiene que existir interacción entre cada uno de sus miembros y los elementos que generen cohesión y una estructura o sistema de jerarquías (Shaw, 1983). Esta última característica es la que para algunos teóricos da lugar a la existencia de subgrupos, los cuales forman parte de un conglomerado mayor, pero cuentan con alguna creencia o cohesión particular que los une y diferencia del resto.

La Sociología privilegia las relaciones que existen entre los individuos. Su foco de atención está en encontrar aquellos elementos que sirven de cohesión para configurar a los grupos humanos, aunque no existe un acuerdo general de cómo medir a los grupos y subgrupos. Ante esta discrepancia, muchas veces se emplean algunos enfoques sistemáticos bajo criterios subjetivos y objetivos, lo que da como resultado diferentes subgrupos que se configuran a partir de relaciones íntimas, personales o afectuosas, y son fundamentales para formar la naturaleza e ideas sociales de individuos. Esta forma de entender a los subgrupos, donde se toman en cuenta aspectos subjetivos, como los sistemas de valores y creencias, permite incluir a miembros con características muy diversas y que no están condicionados por los marcadores tradicionales de la demografía, como la edad y el sexo.

En la Estadística también existe un concepto que toma en cuenta a grupos particulares, privilegiando la ubicación geográfica: las *áreas pequeñas*. De acuerdo con Molina (2019), se trata de una combinación de delimitaciones político-administrativas que no contemplan datos de la población en particular. Este concepto presenta el inconveniente de que no es un enfoque, sino una suerte de nombrar a un conjunto de técnicas estadísticas para llevar a cabo la estimación de un valor. Además, no hace referencia al número de personas sino a la falta de información, por lo que es factible asumir que países enteros pueden ser susceptibles de ser catalogados como *áreas pequeñas* si se cumple el criterio de la falta de información (Rao y Molina, 2015).

Esta manera de conceptualizar a los subgrupos merece algunas observaciones más, por ejemplo que presupone una evaluación previa de los datos disponibles. Situación que va en dirección opuesta a lo que plantean varios autores sobre cómo llevar a cabo una investigación

mediante datos empíricos; en un escenario ideal, “la teoría precede a la ratificación” (Prada, 1987:318). Esto significa que es necesario contar con un marco conceptual que especifique las pautas de qué capturar, cómo hacerlo y, sobre todo, cuáles son las delimitaciones que se tendrán al momento de aprehender algún fenómeno. Si bien es cierto que este camino parece razonable porque permite dar sustento a los datos, en el caso de las *áreas pequeñas* se propone lo contrario, es decir, partir de los datos disponibles, realizar una evaluación y a partir de ellos determinar si se aplica el concepto o no (Molina, 2019; Rao y Molina, 2015).

En lo que están de acuerdo estas disciplinas es en que los subgrupos se configuran a partir de ciertas características específicas. Sin embargo, incluso bajo el criterio de que *población* es una categoría analítica que incluye la dimensión geográfica y espacial, la ubicación geográfica está ausente en las definiciones anteriores. Quizá debido a lo que Salles (2003:106) ha observado, que “no es solo un parámetro, [sino] una propiedad misma de los fenómenos”.

En la agenda demográfica el *espacio* es una herramienta que permite estudiar agrupaciones de fenómenos complejos que se encuentran en la superficie terrestre y su desarrollo. Iracheta (2002:73) habla de la región como un espacio subnacional o subestatal, cuyo objetivo es dividir el territorio de acuerdo con determinados propósitos, obedeciendo a la combinación de determinadas características geográficas y socioeconómicas.

Ahora bien, y recapitulando, ante el problema de que no existe en la literatura demográfica una definición unánime y explícita del concepto *subpoblación*, y a partir de lo expuesto anteriormente, para este estudio resulta positivo hablar de *subpoblaciones en áreas pequeñas* las cuales definiremos como: “grupo poblacional con características sociodemográficas específicas que se ubican en un determinado territorio para el cual se carece de información”. Con esta definición se busca hacer notar que se trata de un grupo de personas que comparten coordenadas, espacio-temporales, es decir, características demográficas y espaciales similares en un momento y lugar determinado.

De esta forma evitaremos continuar operando en las investigaciones con una definición en la que solo se hable de *áreas pequeñas* y evitaremos trasladar el problema inicial que supone la homogeneidad de las poblaciones a una escala menor y los problemas de falacias como la de los falsos promedios que puede dar como resultado situaciones en las que el investigador podría pensar

que los resultados serán más precisos con tan únicamente cambiar la escala de análisis de un nivel nacional a uno local.

2.7 Conclusiones

Lo planteado en este apartado implicó avances considerables, por una parte, la introducción del espacio hace ruido a los modelos analíticos de la demografía, ya que agrega una dimensión real y más tangible a los fenómenos. Por otro parte, se demostró que la perspectiva espacial está incrustada en todas las ciencias sociales y sus respectivos métodos, así que cada disciplina puede construir su concepto del espacio a partir de sus premisas básicas. La principal conclusión de este capítulo es reconocer que la incorporación del espacio en la tarea demográfica, además de evidenciar las limitaciones que tiene en sus principales y muy usuales instrumentos analíticos, permite admitir que ninguna disciplina, por sí sola, es capaz de estudiar a la población en su totalidad, por lo cual, es más auténtico utilizar el término de subpoblaciones en áreas pequeñas.

La apuesta por esta locución responde a que se trata de una noción que hace referencia a un grupo de personas que comparten coordenadas espacio temporales, que dotan de características demográficas y espaciales similares en un momento y lugar determinado. De tal manera que, para el objetivo de esta investigación, esta perspectiva nos es más útil porque sitúa en un tiempo y momento determinado a nuestro objeto de estudio, como si se tomara una fotografía.

El interés por la inclusión de la perspectiva espacial pretende superar el postulado de localización que se reduce a la pregunta ¿dónde?, al mismo tiempo que se busca dejar de mantener o seguir reproduciendo cartografías detalladas de todos los puntos de este planeta. Por el contrario, se busca empatar el quehacer demográfico con contextos de espacio y tiempo definidos. Especializar a la demografía sin duda es una aspiración mayúscula, no obstante, en este particular caso de investigación, es ineludible por lo expuesto en páginas previas. Hay que reconocer que el estudio de cualquier fenómeno social, que ignore analíticamente la espacialidad, concibe, de alguna forma, una mutilación al fenómeno de interés. También implica someterse a una investigación que tendrá una comprensión limitada de lo social. En particular, para la demografía, este reconocimiento abre un abanico de posibilidades para actualizar sus áreas de oportunidad como las ya mencionadas. Y brinda la oportunidad, de experimentar otros enfoques y otras herramientas analíticas que permitan el nacimiento de nuevos paradigmas.

Capítulo 3

La investigación demográfica-espacial

3.1 Introducción

Este capítulo pone en el centro de interés en ¿cómo se construye el dato demográfico?, y para hacerlo describe a grandes rasgos la historia de la demografía. Se cuestiona lo poco probable de generar leyes universales para la transición demográfica y se observa a la demografía como una disciplina criticada por tener diversos métodos, pero adolecer de teoría. En cuanto a la construcción de datos demográficos, las siguientes páginas apuntan el recorrido teórico metodológico para su elaboración: concepto teórico, indicador, dato. Y se destaca que, los datos demográficos son el resultado de observaciones o mediciones del comportamiento de las personas, por consiguiente, están sujetos a error y pueden ser cuestionados tan pronto como se obtienen.

También se analizan las fortalezas y debilidades de las fuentes de información más utilizadas en la demografía. Se sostiene que estos datos presentan al menos dos problemas: todas tienen un cierto grado de error en sus estimaciones y sus resultados únicamente son válidos hasta cierto nivel de desagregación; además de los sesgos que existen en el diseño de los instrumentos de captación y el propio levantamiento.

Se argumenta que la disponibilidad de información y su nivel de desagregación influye en la selección de la investigación demográfica, ya que el abanico de posibilidades para abordar un tema depende del acceso a los datos. Por otra parte, se observa que los métodos estadísticos de estas fuentes cuentan con una delimitación geográfica que se extiende en el diseño mismo, así como sus métodos y operativos de campo. Pero también, se refleja en el análisis de los resultados, por lo cual, es ineludible revisar el concepto de espacio desde una perspectiva más amplia.

Mediante referencias concretas, el texto indica que la dimensión espacial se asume como una característica más de la población, toda vez que su objetivo se centra en estudiar el efecto de ciertas condiciones a partir de una delimitación concreta de lugares que aglutinan poblaciones. Aunque, en años recientes, el desarrollo de otros métodos asigna al espacio un sitio con mayor

relevancia y se asume como el factor clave que determina el comportamiento de las personas, de forma concreta los conceptos que se verifican son la dependencia y la heterogeneidad espacial.

Hoy día se habla de la demografía espacial como una perspectiva que permite estudiar el crecimiento de la población utilizando la gran cantidad de información, mucha de ella disponible debido a la tecnología aeroespacial que se ha incrementado de manera exponencial. Aunque sin olvidar la diferencia que existe entre la información obtenida por imágenes satelitales y los datos demográficos tradicionales. De esta manera, las imágenes satelitales de alta resolución presentan una fuente de información inexplorada y la percepción remota se convierte en la posibilidad de hacer análisis con diferentes niveles de desagregación.

3.2 Los datos demográficos

Uno de los principales debates en el siglo XX en torno a la ciencia es el método que se debía emplear para realizar un estudio. Los partidarios del positivismo simpatizaban con el uso de una lógica hipotético-deductiva cuyo objetivo era comprobar las hipótesis con base en los datos disponibles. Los que preferían los estudios fenomenológicos sostuvieron que los asuntos sociales deben estudiarse con un método inductivo asociado a la metodología cualitativa. Pese a esta diferencia tan importante entre ambos, y entre el tipo de resultados que pudieran arrojar en una investigación, el positivismo preponderó en la manera de estudiar fenómenos demográficos (Leyva, 2001), debido a que las actividades científicas deben efectuarse en el marco del empirismo y tomar como referencia todo aquello que es verificable por un sistema de medición (Adler; 1964).

El positivismo parte de un conocimiento previo, sus preguntas de investigación deben ser accesibles a la observación empírica y resueltas mediante el método experimental. Uno de sus objetivos es encontrar leyes universales que permitan comprender el entorno, tal y como sucede con los fenómenos naturales. Este enfoque provocó que las llamadas “ciencias del espíritu” fueran criticadas debido a que sus afirmaciones no devienen de un método experimental con base en la cuantificación. Sin embargo, debido a que algunos fenómenos no son tan estáticos, continuos, uniformes y estables, sino contradictorios y en continua recomposición, un dato específico pierde vigencia de un momento a otro y únicamente es válido en un espacio y tiempo determinados (Habermas, 1990).

Los detractores del positivismo cuestionaron la idea de un método científico único y aplicable a todos los campos de estudio. Cuestionan el hecho de que todo el conocimiento producido esté libre del sesgo inherente a la actividad humana (Bachelard, 1987), uno de los argumentos para sostener esta crítica es que el positivismo se basa en la observación fragmentada de la realidad, por lo que sería poco probable que se generen leyes universales (Popper, 1992). Un ejemplo de ello es la teoría de la transición demográfica que se abordó en el capítulo anterior, la cual trata de establecer una serie de principios generales sobre cómo opera el comportamiento de la fecundidad y mortalidad, pero es cuestionada al momento de observar casos en donde no se cumple.

Theodor W. Adorno y Max Horkheimer argumentan que la visión positivista es inadecuada para estudiar a la población porque no se pueden hacer observaciones totalmente fiables, por lo que su estudio no es equiparable a las ciencias de la naturaleza en donde el método analítico deductivo resulta confiable. También criticaron el privilegio que se le ha dado de la observación empírica como medio para el conocimiento de la realidad, pues en esta existen aspectos que no son visibles en lo inmediato y requieren de una exploración más detallada (Adorno, 2001).

El enfoque positivista supone que el dato está dado y es directamente observable por parte de los investigadores, bajo este supuesto el conocimiento se reduce a coleccionar y analizar dichos datos (Bunge, 1997; Villoro, 1996). De acuerdo con Parta (1970), el dato es una herramienta que sirve para conocer las características que se le atribuyen a un objeto de estudio y es una forma de aprehender la realidad (Bunge, 1997). Sin embargo, algunos autores han refutado esta posición al considerar que el dato depende de la teoría (Feyerabend), el instrumento (Bachelard, 1987) e incluso el lenguaje (Hughes, 1999) con el que se construye; y que en su construcción intervienen relaciones de poder y cultura que subyacen los acuerdos de la comunidad académica en torno a lo que es verdadero (González, 2018).

En este contexto surge una interrogante ¿cómo se construye el dato demográfico?

El primer aspecto para tomar en cuenta es que el concepto “dato” es distinto al de “número”, ya que el dato es el resultado de la observación y es producto de la aplicación de un método científico. En el escenario ideal todo parte de las hipótesis sobre el comportamiento observable de los individuos y sobre el significado de sus acciones, sí con estas hipótesis se ha captado qué es lo que ocurre, entonces las observaciones se condensan en datos y con ellos se pueden efectuar

investigaciones generales de tipo inductivo o hipotético deductivo (Klimovsky; Hidalgo; 1998: 216).

Para la construcción del dato se usa un puente intermedio: la operacionalización de conceptos, es decir, expresar los conceptos en términos de índices empíricos capaces de medir aspectos importantes del problema estudiado. Esto significa bajar de los conceptos teóricos y abstractos a la realidad empírica, descomponer el concepto y volverlo cuantificable por medio de instrumentos, por ejemplo los cuestionarios, con el fin de transformar una cualidad empírica en algo observable que es imperceptible en primera instancia (Lazarsfeld; 1985). Por ejemplo, para lograr captar la reducción en el número de hijos e hijas, los investigadores del proyecto de Princeton tuvieron que conceptualizar la “tasa de fecundidad” con la intención de observar si era cierto que había una disminución en los nacimientos (Coale, 1984; Lesthaeghe y Neels, 2002).

El dato demográfico es un fragmento de conocimiento que requiere de un contexto para cobrar sentido. También requiere de hipótesis que guíen su búsqueda y de teorías que sean adecuadas para interpretarlos. Para Bunge (1999) un problema de investigación surge de preguntas que pueden responderse a través de los datos, pero estos deben filtrarse para identificar la información que es pertinente. Los mecanismos de filtración son la filosofía y la teoría, que direcciona qué tipo de datos deberíamos recolectar o producir para el problema de investigación. Sin embargo, la demografía es una disciplina criticada y colocada como “rica en métodos, pero carente de teoría” (Hauser y Duncan, 1959:10).

Aunque los datos demográficos son el resultado de observaciones o mediciones del comportamiento de las personas, estos se recolectan bajo la ruta: concepto teórico, indicador, dato; por consiguiente, están sujetos a error y pueden ser cuestionados tan pronto como se obtienen; esta situación ha generado controversia entre los investigadores. Aunado a lo anterior, las hipótesis dependen, en última instancia, de la capacidad para recolectar datos (Bunge 1999: 141), son susceptibles de ser cuestionadas y tarde o temprano se les somete a pruebas mediante otras observaciones o a la luz de alguna otra teoría.

Ante este panorama, Mayer-Schönberger y Cukier (2013) consideran que la humanidad se encuentra en un periodo de datificación de los fenómenos, lo que significa que se busca “plasmarse en un formato cuantificable sus características de modo que sea posible su análisis” (Mayer-Schönberger y Cukier, 2013:100). Debido a lo anterior, la cantidad de datos disponibles ha crecido de manera exponencial en los últimos años, así como su diversidad temática que va a más allá de

la esfera demográfica. Este auge de información ha llevado a que algunos autores como Greengard (2015) escriban sobre la “cuantificación del mundo” que, sumado al desarrollo tecnológico, trae consigo lo que podría ser una nueva revolución, comparable al surgimiento de la máquina de vapor (Mayer-Schönberger y Cukier, 2013).

Con esta sobreproducción de datos la posibilidad del error pone en riesgo su adecuada interpretación. Esto se debe a que los grandes avances tecnológicos en materia de recolección de información ofrecen resultados a gran velocidad, pero muchas veces se realiza sin una herramienta adecuada o sin rastrear la construcción del dato en sí mismo (López, 1998: 34). No obstante, y debido a que esa información es el punto de partida en muchas investigaciones, antes de incorporar todos los datos al cuerpo de la investigación se debe tener en cuenta qué aportan las fuentes, así como qué limitaciones y potencialidades tienen (Paquette, 2008:775).

Este auge de información genera la necesidad de nuevos métodos de análisis cuyo objetivo está en desarrollar técnicas que permitan detectar patrones a partir de los datos disponibles (Greengard, 2015; Mayer-Schönberger y Cukier, 2013). Lo que significa que ya no se privilegia a la teoría, sino a los datos y, sobre todo, a la capacidad computacional para analizar y encontrar patrones. Esta forma de operar tampoco resulta ajena al quehacer demográfico, hoy en día se cuenta con una gran cantidad de fuentes de información (encuestas, censos, registros administrativos) a partir de las cuales se llevan a cabo las investigaciones. De hecho, y aunque se reconoce poco, muchas de las preguntas de investigación demográfica se plantean según la disponibilidad de información y, sobre todo, su nivel de desagregación. Sin embargo, ante la ausencia de datos o ante un escenario en el que su nivel de desagregación sea inadecuado, en la investigación se opta por enfoques diferentes, debido a que coleccionar información desde cero no es una tarea sencilla y requiere de cierto grado de especialización para llevarla a cabo de manera adecuada (Hauser y Duncan, 1959).

3.3 Las fuentes de información

La Demografía es una ciencia que estudia a la población a la distancia, por lo que un sistema de recolección de datos en donde se puedan identificar las características de las personas puede servir como fuente de información (Hauser y Duncan, 1959). Lo ideal sería que este conjunto de datos se encuentre de un marco de referencia adecuado y brinde la posibilidad de construir conocimiento (Villoro, 1996). Gran parte de las fuentes demográficas tradicionales obtienen sus datos mediante

un proceso de interlocución entre las personas que brindan su información y quienes la recolectan (Vieira, 1973). Los instrumentos más utilizados en la literatura demográfica son tres: 1) registros administrativos, 2) encuestas, y 3) Censos de Población y Vivienda.

Las estadísticas vitales son un conjunto de datos que se configuran a partir del registro de acontecimientos en donde se expide algún tipo de certificado o acta para dar certeza jurídica a las personas y se remontan a la Edad Media. Durante mucho tiempo la Iglesia era la encargada de coleccionar información sobre acontecimientos como muertes o nacimientos, pero con el tiempo esta responsabilidad fue asumida y controlada por las autoridades civiles (Livi-Bacci, 1993: 27). En estas estadísticas quedan registrados los eventos más usuales de la sociedad y que son objeto de estudio para la Demografía, los nacimientos, matrimonios, divorcios y defunciones (Pressat, 1967). Este tipo de información tiene algunas ventajas, por ejemplo, que se suelen publicar con más frecuencia (Rodríguez, 2018); además, ofrecen una opción asequible para captar fenómenos que son difíciles de observar como la migración internacional (Bush y Herrera, 2007).

Las encuestas, por otro lado, son instrumentos estadísticos que permiten recolectar datos con un alto grado de profundidad sobre algún tema en particular (Cochran, 2007; Johnson y Elliott, 1998; Lumley, 2010). A diferencia de los censos en donde se trata de obtener información sobre todas las personas, en las encuestas se elige a un subconjunto según sus características sociodemográficas y su distribución espacial (Cochran, 2007). Esto las hace menos costosas, permite una especialización mayor en el personal que las lleva a cabo y las hace fáciles de levantar de manera frecuente. Algunas de ellas se llevan a cabo de manera continua y permiten dar seguimiento a las personas por un periodo determinado, como la Encuesta Continua de Población (*Current Population Survey* en inglés) o la Encuesta Nacional de Ocupación y Empleo (Inegi, 2007).

Este tipo de instrumentos tuvo un gran impulso durante toda la primera mitad del siglo XX con el desarrollo de las técnicas de muestreo probabilístico (ONU, 2008). Estas técnicas mejoraron los métodos para hacer inferencias a partir de un subconjunto de personas e incrementó su precisión. Sin embargo, los datos de encuestas presentan dos grandes problemas: las observaciones no se obtienen de forma aleatoria ni son independientes (Cochran, 2007) y todas tienen un cierto grado de error en sus estimaciones, por lo que sus resultados únicamente son válidos hasta cierto nivel de desagregación, usualmente para las unidades administrativas más grandes.

Los Censos de Población y Vivienda son los instrumentos más conocidos dentro y fuera del ámbito demográfico (Inegi, 2010; ONU, 2010), enumeran a la población y permiten obtener datos sobre sus características socioeconómicas y culturales, así como sobre la infraestructura de la vivienda y activos de los hogares. De acuerdo con la visión de la ONU, este tipo de instrumentos son útiles para la administración pública y sus resultados permiten garantizar la equidad y distribución de los servicios gubernamentales (ONU, 2010). Este tipo de ejercicios se han llevado a cabo durante muchos años con la finalidad de recabar los datos de las personas para saber sus características y necesidades. Spiegelman (1972) señala que las primeras enumeraciones se hicieron en Babilonia y China, alrededor de 2500 a.C.

Sin embargo, este tipo de instrumentos son costosos, demandan de una fuerte inversión pública y los países requieren de cierta estabilidad social para desarrollarlos. Por este motivo, algunas naciones son incapaces de llevar a cabo un operativo de esta naturaleza, por lo que la falta de eventos censales es una constante que se presenta en muchas regiones del mundo (ONU, 2010). Wardrop *et al.*, (2018) hacen un recuento de los eventos censales de los últimos años y encuentran que en la mayor parte de Latinoamérica se tiene una brecha que oscila entre los 6 y los 10 años, mientras que en otras regiones existen casos extremos en donde el último censo se llevó a cabo hace más de medio siglo. Esto significa que durante todo este tiempo se han desarrollado e implementado políticas públicas sin saber a ciencia cierta cuántas personas existen, ni sus características.

Existe otra situación a tomar en cuenta, se asume que los datos censales son confiables. Si bien, esta afirmación puede ser verdadera, existen algunos casos, o circunstancias, en donde esto no se cumple. De acuerdo con Baffour y Valente (2012) “es bien sabido que, a pesar de todos los esfuerzos de los censos, siempre habrá algunas personas que no sean contadas”. Además, hay que añadir los sesgos que existen en el diseño de los instrumentos de captación y la omisión o falsificación de información que se genera en los operativos de campo, la cual pasa inadvertida en las oficinas de estadísticas (ONU, 2010).

Al hacer un balance de estas dos últimas fuentes de información se debe mencionar que son, en esencia, métodos estadísticos en donde las delimitaciones geográficas tienen una función fundamental que se refleja desde el diseño mismo de los instrumentos hasta los métodos implementados por los operativos de campo. Por ejemplo, para levantar cualquier tipo de encuesta se utilizan las denominadas Unidades Primarias de Muestreo y los estratos, los cuales son

particiones de la superficie terrestre según las características físicas del terreno, las viviendas y los rasgos sociodemográficos de las personas que en ellas habitan (Inegi, 2011). Sin embargo, el impacto de las delimitaciones geográficas también se refleja al momento de analizar los resultados, si la encuesta no se había planificado para estimar a un nivel tan desagregado, el número de observaciones que se tienen en algunas áreas podría ser demasiado pequeño, y en consecuencia los errores de muestreo grande (Molina, 2019).

3.4 Los métodos demográficos

La Demografía es una disciplina que combina métodos propios y técnicas estadísticas que son de uso general para todas las ciencias (Véron *et al.*, 2002). Muchas de las técnicas que se han desarrollado están ligadas a la disponibilidad de información (Hauser y Duncan, 1959), por lo que es apremiante favorecer el desarrollo de nuevas técnicas de análisis. Este estrecho vínculo puede ser observado en la tabla de vida, un método conocido y cuya aplicación se extiende más allá del ámbito demográfico. Para su construcción se utilizan los registros administrativos de los nacimientos y las defunciones, así como los datos poblacionales que se obtienen en censos o en encuestas. La combinación de estas fuentes de información permite construir la tabla de vida y el grado de especificidad de la información permite mejorar la calidad de las estimaciones (Mina, 2011).

El quehacer demográfico estudia el comportamiento de la mortalidad, natalidad y migración; pero, además, examina la forma en que ocurren otros fenómenos sociales. Para ello, se vale de diversas técnicas econométricas, por ejemplo, Puig-Barrachina *et al.*, (2019), utilizan modelos de regresión lineal ajustados para calcular las tasas de fecundidad y describir su descenso en los últimos años; Hernández (2008) se plantea como objetivo dar seguimiento a los movimientos migratorios y para ello usa modelos de ecuaciones estructurales que le permite identificar las variables que más influyen en la decisión de migrar. Otros estudios con un enfoque más social también han empleado diversas técnicas econométricas, tal es el caso de Román (2013), quien estudia los niveles de precariedad laboral que sufren los jóvenes que se incorporan al mercado de trabajo mediante regresiones logísticas.

El común denominador en todas estas investigaciones en donde se emplean modelos econométricos es que su intención es identificar los factores que favorecen o inhiben la ocurrencia de cierto fenómeno. Esta búsqueda por encontrar la causalidad de los eventos es una constante en

la investigación demográfica (Hauser y Duncan, 1959), situación que puede estar permeada por la necesidad de hacer predicciones sobre el comportamiento de la población. Cuando se detecta que la dimensión geográfica tiene un rol significativo para la explicación de los fenómenos, lo usual es distinguir el lugar de origen de las personas. Por ejemplo, Giorguli *et al.* (2010) analizan el vínculo entre los procesos educativos y la dinámica demográfica, por lo que emplean cuatro modelos de regresión lineal, uno para cada grado de urbanización del país. En el estudio que lleva a cabo Solís (2013), hay otro ejemplo en donde se incorpora la dimensión espacial. Como su objetivo es estudiar el efecto que tienen las condiciones socioeconómicas de la familia de origen sobre los niveles de escolaridad de la población, el autor utiliza regresiones logísticas multinomiales y dentro de las variables predictoras considera al tamaño de las localidades de nacimiento. En ambos ejemplos la dimensión espacial se asume como una característica más de la población.

En últimas fechas se han desarrollado otros métodos en donde el espacio cobra una mayor relevancia y se asume como el factor clave que determina el comportamiento de las personas. Para entrar en materia, conviene retomar dos conceptos clave que están conectados con el análisis de dichos patrones y son: la dependencia y la heterogeneidad espaciales. El primer concepto se retoma de la Ley de Tobler en la cual se establece que las cosas más cercanas están relacionadas unas con otras porque están cercanas (Waters, 2018), de ahí que la proximidad es un predictor de los comportamientos sociodemográficos. Un ejemplo de este tipo de análisis lo llevan a cabo Asuad y Quintana,

(2010) quienes profundizan en el crecimiento económico de los estados en el periodo de 1970 a 2008 y encuentran que existen núcleos en donde la bonanza económica se mantiene a lo largo de los años y que los estados beneficiarios son aquellos colindantes a los polos de mayor desarrollo (Asuad y Quintana, 2010). El segundo concepto es la heterogeneidad espacial el cual refiere situaciones en donde las asociaciones entre variables desempeñan un rol diferente al que se les asigna, dependiendo de la ubicación geográfica de las personas. Esta situación la ilustra Lobao (*et al.*, 2007), quienes examinan la relación que existe entre la fecundidad y la educación, en donde ocurre que las mujeres cuyo grado educativo es mayor suelen tener menos hijos; sin embargo, esta relación no siempre se presenta, por ejemplo, en África subsahariana las mujeres con altos niveles educativos suelen tener más hijos.

Sin embargo, contar con métodos estadísticos que tengan un enfoque espacial no es suficiente, hace falta desarrollar una forma diferente de pensar los fenómenos. Para crear un marco

de referencia general que involucre tanto a la Demografía como a la distribución de la población; algunos autores como Vos (2016) hablan de una demografía espacial. En la siguiente sección se discute sobre qué y cuáles son los elementos que la diferencian de la demografía tradicional.

3.5 Demografía espacial

Este término fue desarrollado por el británico Robert y hace referencia a la búsqueda de los aspectos demográficos que sirvan para responder las preguntas alrededor del problema: por qué se tienen patrones espaciales determinados en un sector en particular. Esta subrama de la Demografía difiere de la Geografía de la población que se enfoca en la movilidad de las personas, mientras la demografía espacial pretende encontrar las relaciones entre los aspectos espaciales y los fenómenos demográficos (Anselin, 1989; Haining, 2003; Holt y Lo, 2008; Lobao *et al.*, 2007).

El vínculo entre la Demografía y el espacio se fortaleció con el auge de los Sistemas de Información Geográfica (SIG), los cuales se definen como un conjunto de técnicas computacionales que permiten obtener información geográfica digital que se convierte en datos que pueden vincularse con otras disciplinas (Congalton, 1991). Los alcances de esta tecnología van desde la construcción de acervos, hasta el empleo de modelos estadísticos que reflejen la interacción de los fenómenos con su dimensión territorial y ambiental. En un inicio, este tipo de sistemas se enfocaban en georreferenciar espacialmente a las unidades de análisis (Tomlinson, 2007), ahora existen propuestas que van más allá de ello y proponen estudiar los cambios en la estructura por edad de la población mediante el uso de imágenes satelitales (Hillson *et al.*, 2019; Iisaka y Hegedus, 1982; Li y Weng, 2010; Ordorica-Mellado *et al.*, 2019; Silván-Cárdenas *et al.*, 2010). El análisis de la población mediante fotografías aéreas no es nuevo, los primeros intentos se llevaron a cabo el siglo pasado, pero su vinculación con los estudios demográficos se ha dado de manera paulatina.

Para estudiar el crecimiento de la población existe una gran cantidad de información que podría ser de utilidad, alguna ha sido posible gracias a la tecnología aeroespacial que se ha incrementado de manera exponencial. Un ejemplo de ello es la denominada Capa Global de Asentamientos Humanos (GHSL por sus siglas en inglés), desarrollada por la Unión Europea, que pretende producir un archivo raster con la cantidad de personas que habitan en un espacio de 250 m² (Florczyk *et al.*, 2019). Otros proyectos como el de Población Cuadrículada del Mundo

(GPWv4) van un paso más allá y ofrecen datos sociodemográficos básicos: rango de edad y sexo. Esto resulta atractivo no únicamente para las naciones que carecen de información estadística, sino también para aquellas que producen información de manera regular, así los datos censales se pueden comparar y evaluar.

Sin embargo, existe una diferencia importante entre la información que se puede obtener vía imágenes satelitales y los datos demográficos tradicionales. Las primeras se basan en la caracterización de los píxeles por lo que se hace a la distancia y no se tiene contacto directo con el objeto de estudio, que en este caso es la población (Cárdenas y Ramírez, 2012). El principal insumo de las segundas se obtiene mediante contacto directo con el informante, ya sea que se le aplique un cuestionario (como en el caso de las encuestas y los censos) o se haga un registro de algún evento que le haya ocurrido. Esta aparente disyuntiva en los enfoques conlleva a la siguiente pregunta ¿se pueden captar los fenómenos sociales desde el espacio? Para responderla, a continuación se hace un recuento de algunos estudios donde se abordan temas demográficos. Primero se tratará de los más generales y luego de aquellos que son más específicos.

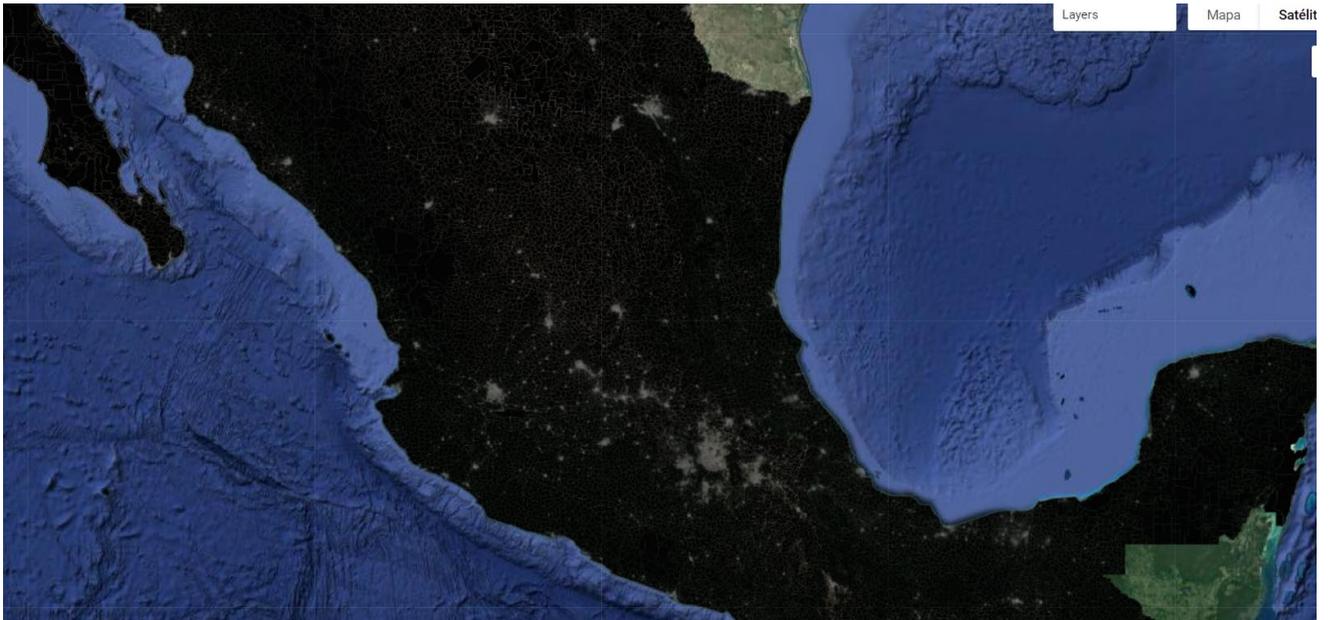
Uno de los primeros estudios que tuvo como objetivo estimar la población utilizando técnicas de percepción remota fue el realizado por Iisaka y Hegedus (1982). Ellos emplearon imágenes satelitales para analizar la radiancia de cuatro bandas espectrales y calcular la densidad de población en la zona metropolitana de Tokio. En esta misma línea, Hillson *et al.* (2019), usaron modelos estadísticos, a partir de los datos del Landsat, para estudiar el comportamiento demográfico de la población urbana de Bo, en Sierra Leona. Esta región carece de información censal, por lo que sus resultados adquirieron mayor relevancia.

Por otro lado, Li y Weng (2010) también llevaron a cabo un ejercicio similar, pero se plantearon hacer una aproximación más detallada, por lo que utilizaron una capa de mayor resolución que les permitiera implementar algoritmos que identificaran a las viviendas. En el caso de América Latina, Rodríguez *et al.* (2010) desarrollaron un índice de calidad ambiental urbana para la ciudad de Cali e hicieron diversas recomendaciones para la implementación de políticas públicas. Silván-Cárdenas *et al.* (2010) utilizaron sensores lidar (LIDAR por sus siglas en inglés) de alta resolución para hacer un mapeo tridimensional del espacio y estimar poblaciones en áreas pequeñas. Todas estas investigaciones mostraron que las técnicas de percepción remota son una alternativa viable para enriquecer los estudios demográficos. Además, tienen el potencial para

hacer un contraste de los datos censales, tal y como lo reconoció la ONU en 2012, así como para reducir el rango de omisiones, lo que permite visibilizar poblaciones que no están siendo incluidas.

Chen y Nordhaus (2011), desarrollan otro uso de imágenes satelitales que merece ser tomado como ejemplo; los investigadores hacen un análisis de cómo emplear el nivel de luminosidad de las regiones como aproximaciones a variables sociales (ver Mapa 1); como una respuesta a la falta de desagregación o ante la ausencia de información que enfrentan diversos países. En este sentido, las imágenes satelitales de alta resolución presentan una fuente de información inexplorada. La metodología para utilizar estos instrumentos consiste en contabilizar los números digitales de cada uno de los píxeles que contienen los archivos raster. Si bien, el grado de luminosidad de una región está ligado con el desarrollo económico y al tamaño de la población, los resultados de su estudio han mostrado que los datos satelitales son más confiables en los países en desarrollo. Además, un aspecto que enfatizan es que este tipo de datos sirven como complemento a las variables estadísticas tradicionales, por lo que la percepción remota se convierte en la posibilidad de hacer análisis con diferentes niveles de desagregación, con lo que se ofrece un contexto y se permite generar una visión diferente de cómo analizar los fenómenos sociales.

Mapa 1 Luces nocturnas



Fuente: elaboración propia

3.6 Conclusiones

La Demografía, como otras ciencias es una disciplina que combina métodos propios y técnicas estadísticas que son de uso general para todas las ciencias, por lo cual es preciso promover el desarrollo de nuevas técnicas de análisis. Sin embargo, contar con métodos estadísticos que tengan un enfoque espacial es insuficiente, ya que hace falta desarrollar una forma diferente de pensar los fenómenos, para crear un marco de referencia general que involucre tanto a la Demografía como a la distribución de la población.

Incluir el contexto espacial en el estudio de los temas demográficos permite generar una visión diferente de cómo analizar los fenómenos sociales. Esto implica una apertura en la agenda de las disciplinas y una revisión de la terminología básica que utilizan los demógrafos para sus análisis, principalmente porque se habla de una datificación de los fenómenos y la diversidad temática trasciende en muchos de los casos a la esfera demográfica.

La demografía puede ser complementaria con la geografía espacial, esto obedece a que ambas analizan poblaciones, donde resulta importante identificar la cantidad de habitantes y los cambios en su comportamiento, no obstante, se debe recordar que la población estudiada está siempre formada por individuos y que muchas veces el estudio deberá emplear técnicas estadísticas. La Estadística permite explicar las condiciones regulares en fenómenos de tipo aleatorio, los cuales, se componen en buena parte de nuestro mundo.

Es muy tentador acudir al auge de información, sin embargo, sin métodos de análisis cuyo objetivo sea desarrollar técnicas que permitan detectar patrones a partir de los datos disponibles, es muy probable que la demografía se mantenga en un círculo vicioso que le impida adaptarse a la velocidad de los datos. Sin olvidar que el vínculo entre la Demografía y el espacio se fortaleció con el auge de los Sistemas de Información Geográfica.

El paradigma que hoy se asiste es el digital, teniendo en cuenta lo anterior vale la pena advertir que la datificación amenaza el uso de las muestras, un pilar fundamental en el análisis demográfico, y que, por lo tanto, es necesario revisar los aspectos epistemológicos y éticos de la Demografía como ciencia, así como las repercusiones en el uso de los macrodatos. Estos macrodatos, es más que una bolsa de datos, ya que se complementa de las diversas técnicas para extraerlos, hacerlos modelos estadísticos y almacenarlos, incluso trabajarlos en tiempo real, por lo que la explosión de datos ofrece convertirlos en conocimiento, apostamos en que la Demografía puede contribuir a esta gran oportunidad.

Capítulo 4

Desagregación de la información

4.1 Introducción

Este capítulo reflexiona sobre la oferta y disponibilidad de información para la elaboración de investigaciones, y un aspecto clave es analizar las desventajas que genera tener que modificar protocolos y objetivos de investigación para adaptarlos a los datos disponibles. También recomienda equiparar la calidad y la cantidad de información que se genera en los estudios demográficos, así como los datos y herramientas.

A propósito de lo anterior, se revisan las características operativas de algunas oficinas estadísticas, haciendo énfasis en la mexicana y se discurre sobre la unilateralidad que prevalece en los Institutos en cuanto qué medir y cómo hacerlo, lo cual implica, entre otras cosas, un gran control para acceder a los datos, pese a que la información es “accesible” por internet. Esto se puede notar en el control de la información que trasciende a esferas de la vida pública de un país, debido a que la información obtenida nutre políticas públicas y toma de decisiones.

Teniendo en cuenta lo anterior, en las siguientes páginas se apunta el rezago en la actualización de temas e instrumentos de, por lo menos, la oficina estadística mexicana y alude a que el manejo de la desagregación de datos es casi inexistente pese a que existen leyes que impiden la violación de la confidencialidad. Con ello revela un antagonismo entre el derecho a la información pública y el derecho a salvaguardar los datos confidenciales. El tema de la confidencialidad es abordado a nivel nacional e internacional. En el primer caso, se menciona que el principal proveedor de datos en el país monopoliza esta información y en consecuencia, se advierte de firmas privadas, consultorías o institutos de investigación que producen datos que contrastan los oficiales y esto podría restar credibilidad a alguna de las fuentes; en el segundo caso, se describen diversas prácticas internacionales con el objetivo de ejemplificar la viabilidad de operar y compartir en cierta medida datos confidenciales.

Con respecto al interés por desagregar la información, la pregunta que guía los argumentos es cuál podría ser el límite que debe existir para utilizar estos datos, la respuesta proviene de una perspectiva ética, en la que por el hecho de que esta información esté disponible al público, ¿esto hace que pueda ser explotada?, la respuesta está en debate, lo cierto es que en este capítulo se sostiene que es necesario contar con un conjunto mínimo de protecciones que garanticen la confidencialidad, se anticipen los posibles daños y se busquen alternativas para minimizarlos, de forma concreta, así que el consentimiento informado, escrito o verbal, recobra relevancia.

Esta noción se entiende mejor como un intento de construir marcos de confianza que hagan posible la minimización de exposición de datos confidenciales, ya que el consentimiento informado, introduce una perspectiva ética para investigadores y participantes. Y de manera general se recupera la idea de que, al realizar cualquier investigación, así sea cualitativa o cuantitativa, hay decisiones que tomar que afectan el comportamiento de las personas y conlleva a situaciones de riesgo. Así que el documento apela a una metodología demográfica con principios éticos desde la perspectiva cualitativa, principalmente porque, aunque la investigación sea cuantitativa, la pregunta que queda de fondo es, ¿cuáles serían los beneficios justificables de exponer de una manera u otra, a las personas participantes en la investigación?

Las posibilidades para la Demografía en cuanto al uso y apertura de datos sugieren la utilización de estos de manera responsable y con cautela, ya que podría ser que la información no sea totalmente fidedigna principalmente por el arribo de una gran cantidad de datos. Los datos geoespaciales vienen deconstruyendo y reformando los principios tradicionales de la estadística, la demografía y el resto de las disciplinas, porque disminuye la privacidad e incrementa el riesgo de exposición vía geolocalización.

El capítulo cierra apuntando que, lo mejor que se puede hacer en el quehacer de cualquier área temática, incluyendo a la demografía, es definir una serie de principios éticos fundamentales para la investigación y las estadísticas, que incluya derechos y obligaciones de los participantes, así como el seguimiento y la vigilancia, tal como ocurre con los principios éticos de la estrategia geoespacial del Reino Unido y la OCDE que las siguientes páginas describen.

De esta forma, este capítulo sugiere que el uso de datos geoespaciales tiene el potencial de perjudicar un área geográfica, en cuanto al uso y la interpretación de estos, y que hace falta regular la relación de poder entre quien “produce” y cómo es el contexto de esta información, así que,

siguiendo la perspectiva ética, se sugiere utilizar una codificación visual, con una gama de colores que permita contrastar la información y presentar los resultados obtenidos de la manera más objetiva.

4.2 Apertura a la información

Con respecto al tema paradigmático de si debe o no haber una apertura de datos, a mitad del siglo XX, el acceso a la información era limitado y los gobiernos lo restringían justificados en argumentos éticos: era obligación del poseedor de los datos salvaguardarlos y ser leal a la confianza de las personas que proporcionaban su información en encuestas y censos de población. Aunque las cláusulas de privacidad no estaban muy claras en la legislación, si los individuos advertían que su privacidad había sido violentada, perdían la confianza en las instituciones que llevaban a cabo la colecta de información y la sancionaban de forma personal emitiendo respuestas falsas en las futuras encuestas o negándose a contestar los cuestionarios para evitar que su información se compartiera. Sin embargo, esta restricción en el acceso a los datos, pese a que pareciera estar justificada en asuntos éticos, tenía su lado negativo: 1) limitaba los datos que los investigadores podían consultar y 2) podían poner en duda las decisiones que se fundamentaban en el análisis de los datos disponibles (Ruggles *et al.*, 2003).

Sin embargo, desde finales del siglo XX y principios del XXI, se gestó un cambio en el paradigma en torno a la necesidad de facilitar el acceso a los datos que captan las características sociodemográficas y a hacer transparentes las decisiones gubernamentales. Este cambio abrió la posibilidad de mayor independencia de los investigadores y mejores resultados en las investigaciones. Debido a ese camino, en las últimas décadas se ha generado un cambio en la forma de analizar y entender los fenómenos demográficos a partir de la mayor disponibilidad de fuentes de información y nivel de precisión avanzado que ofrecen. Por ejemplo, el cálculo de la pobreza que realiza el Consejo Nacional de Evaluación de la Política de Desarrollo Social (CONEVAL), ofrece resultados tabulados y además proporciona las bases de datos con las cuales llevó a cabo las mediciones y su procesamiento. De esta manera, los investigadores e interesados cuentan con la posibilidad de replicar los cálculos paso por paso y corroborar que la información presentada es correcta. Esta apertura y mayor disponibilidad de datos permite que los investigadores cuenten con

un abanico de datos espaciales para ejecutar sus investigaciones y le permite a la Demografía utilizar otras fuentes de información para auxiliarse.

En esta investigación nos enfocamos en el uso de datos satelitales, explícitamente de imágenes como fuentes de información demográfica y abordamos algunas cuestiones sobre los usos éticos que se les da a las imágenes, a través de una revisión del estado del debate sobre el tema actualmente.

A saber, al aumento en la cantidad de fuentes de información disponibles, Ruggles *et al.* (2013) lo denomina “Big microdata”, y observa que es posible tener mayor independencia para abordar los fenómenos desde diferente perspectiva. Para ejemplificar este cambio surgió el proyecto denominado Integrated Public Use Microdata Series (IPUMS, por sus siglas en inglés), cuya intención es concentrar los diferentes proyectos estadísticos (censos o encuestas) y homologarlos de tal forma que sean legibles para su análisis. En este proyecto se pretende democratizar el acceso a las fuentes de información de tal suerte que resulte útil para la formulación y desarrollo de políticas públicas a lo largo del tiempo y el espacio (Ruggles *et al.*, 2003).

Ahora bien, al inicio del proyecto IPUMS los investigadores asociados incluyeron únicamente los datos que producían las diferentes oficinas estadísticas a nivel internacional. Esta fuente de información tiene características específicas: se trata de información estructurada que se genera con propósitos específicos, algunos autores como Cochran (2007), Lahiri (2003), Rao y Molina (2015) los denominan datos de diseño. Además, su unidad de análisis son las personas, por lo que en principio su estructura está pensada en una matriz de modo que las filas representen a cada uno de los individuos que son entrevistados, mientras que las columnas hacen referencia a las diferentes dimensiones o variables que se captan como parte de la investigación (Lumley, 2010). Por estas razones, para llevar a cabo este proceso se requiere de un cuestionario o formulario con el objetivo específico de obtener información de temas particulares (Molina, 2019). Lo que deja a este proyecto con un gran reto: la comparabilidad a nivel internacional, motivo por el cual gran parte del esfuerzo del IPUMS se centra en hacer una integración de datos que facilite la comparación entre diversos países.

Cada vez es más evidente la necesidad de complementar la información oficial con la que ofrecen los microdatos poblacionales (Zaki *et al.*, 1995). Por este motivo, en las últimas décadas en las investigaciones se comenzaron a incluir datos sobre el clima, el uso de la tierra y la cobertura de suelo (Wijedasa *et al.*, 2012). Esta cruzada por la integración responde a la necesidad de fuentes

de información capaces de proveer un contexto y servir de referencia para comprender la dinámica demográfica de los individuos. Ya no es suficiente con armonizar u homologar las fuentes de información tradicionales, ya es necesario complementar la información con otras fuentes que aporten, complementen y permitan observar. En el caso IPUMS, se creó el proyecto denominado Terra Populous cuyo objetivo es la conservación y preservación de la información poblacional en un contexto espacio temporal. De acuerdo con Kugler *et al.* (2015), para lograrlo, tuvieron que recurrir a información espacial, específicamente a imágenes satelitales que les permitieran obtener una cartografía actualizada de la superficie terrestre.

Aunque la idea de fotografiar la superficie terrestre no es nueva. En 1965, por iniciativa del servicio geológico de los Estados Unidos de América (USGS, por sus siglas en inglés) se utilizaron aviones para llevar a cabo este proceso, y tan solo unos años más tarde se empezaron a emplear satélites (Guzmán *et al.*, 2013). De esta forma nacieron diversos programas científicos y militares cuyo objetivo consistía en fotografiar la tierra de manera constante (Zaki *et al.*, 1995). Debido a la importancia de los datos buscados y obtenidos, los satélites de observación terrestre comenzaron a incluir escáneres y ampliaron el espectro electromagnético para analizar el suelo terrestre con mayor detalle y precisión. Al inicio de las operaciones de la USGS la información obtenida no era pública, estaba restringida y era usada únicamente con fines militares; la seguridad de los datos estaba garantizada. Sin embargo, en 2008, debido al cambio en el paradigma en el tema del acceso a la información, se aprobó que todos los datos estuvieran disponibles para su descarga gratuita (Woodcock *et al.*, 2008), pero al mismo tiempo deja abierto dilemas éticos.

4.3 Utilidad de la información

Cualquier nación o territorio que quiera ser administrado de manera óptima requiere estar al tanto de la información sobre las características de su población, los recursos con los que cuentan, cómo es su economía y qué ventajas o desventajas tiene el territorio que habitan, con el fin (ideal) de que el gobierno y la sociedad organizada planeen proyectos que los beneficien y tomen decisiones respaldadas en información objetiva. Por lo tanto, resulta indispensable uniformar la calidad y la cantidad de información que se genera en los estudios demográficos, así como los datos y herramientas informativas que surjan del análisis de estos, con el fin de no estar en un escenario de incertidumbre (DOF, 2008)..

Pero, ¿qué ocurre cuando en un país el instituto de estadística encargado de generar no está ofreciendo los resultados esperados o la información que presenta es insuficiente? ¿Qué desventajas genera en el investigador tener que modificar sus protocolos o los objetivos mismos de la investigación para adaptarlos a los datos disponibles y poco precisos?

En el caso de México, el Instituto Nacional de Información Estadística y Geografía (Inegi) es el órgano encargado de captar, procesar y difundir información estadística oficial. Conocerlo, saber cómo funciona y poner sobre la mesa sus protocolos y las leyes mediante las que se rige nos permitirá comprender su postura ante la desagregación de datos y facilitará comunicar las ventajas metodológicas que propone esta investigación.

Las principales facultades que tiene el Inegi es producir información estadística nacional, lo cual implica realizar censos y encuestas para conocer características específicas de la población mexicana. Estas funciones se encuentran enunciadas por la *Constitución Política de los Estados Unidos Mexicanos* (CPEUM), así como de la ley que rige a este instituto. Con ello, el Inegi adquiere relevancia por ser la única fuente de producción y difusión de datos de carácter oficial bajo una perspectiva de procurar un adecuado procedimiento estadístico y geográfico (DOF, 2008).

El carácter oficial de los datos que genera el Sistema Nacional de Información Estadística y Geográfica (SNIEG) adquiere el grado de Información de Interés Nacional (IIN). De acuerdo con DOF (2008) esta Ley establece:

1. El funcionamiento del Sistema Nacional de Información Estadística y Geográfica.
2. Los procedimientos que guían el actuar de los informantes, tanto en sus obligaciones como en sus derechos.
3. La reglamentación y solución de las posibles controversias relacionadas con el acceso a la información.
4. Los criterios metodológicos que se deben cumplir al momento de recolectar la información sociodemográfica.

4.3.1 El interés público por desagregar la información

En los últimos años, vivimos una era del algoritmo, en la época en la que el *big data* permite el acceso a recursos e información que difícilmente se podría acceder, el uso de múltiples herramientas y dispositivos como teléfonos, tabletas, a través de plataformas y canales de audio, video, con almacenamiento en la nube o alguno de los dispositivos antes mencionados, son

considerados instrumentos con gran potencial para transformar a la sociedad. Por ejemplo, la tecnología combinada con características de localización geográfica en teléfonos inteligentes ha mejorado la eficacia de correspondencia entre los trabajadores de aplicaciones (*Uber*, *DiDi*) y los clientes, que para el contexto de pandemia experimentado en 2020 y 2021, han cobrado relevancia por su importante actividad (Mayer-Schönberger y Cukier, 2013). No obstante, detrás de todo hay un supuesto cuestionable ¿hasta dónde es posible el uso de la información? Para los interesados, la obtención de fuentes de datos es una necesidad, mientras que para los proveedores de información se trata de poder y control. Es cierto que los datos “bien utilizados” podrían ayudar a georeferenciar regiones para dar respuesta a datos de inseguridad, darle seguimiento a actividades como la apertura de negocios, trámites entre otras cosas, sin embargo, ¿cuál es el límite que debe existir para su utilización?

Un supuesto universal es que la información, los datos y descubrimientos tienen un valor que le permite la humanidad dotarla de sentido y transformar el entorno. Hasta antes de esta revolución tecnológica basada en chips, georeferenciación y uso de apps (aplicaciones) existía una menor exposición a datos confidenciales como información personal, gustos y necesidades (Ruggles *et al.*, 2013). De tal forma que la ciencia se encuentra en una encrucijada que descansa en la ética, por un lado, se argumenta que entre más se desagregue la información, se tendrá un mejor impacto en las intervenciones o aproximaciones al recorte de la realidad de interés, pero al mismo tiempo, se advierte que, a una mayor desagregación existe mayor riesgo en causar efectos colaterales, en donde se pueden modificar dinámicas y estilos de vida que probablemente no quedarían exhibidos antes de formar parte de una encuesta o investigación (O’Farrell y Montagnier, 2020).

Como el conocimiento es un proceso de construcción constante, la mente genera estructuras de conocimientos a partir de los datos que toma del mundo exterior, interpretándolos y organizándolos. Los seres humanos desde su nacimiento heredan toda una evolución genética, pero el desarrollo final está en función de las características del contexto en el que vive, por lo que para acceder a ese conocimiento es necesaria la investigación. El punto de partida del investigador comienza al cuestionarse sobre la naturaleza del problema de investigación y continúa con la consideración del acceso a los datos y la plena consciencia de la incidencia o el resultado que derivará de esta investigación (Figueroa, 2006; O’Farrell y Montagnier, 2020).

Actualmente, el tema ético ha ganado relevancia; en particular, el uso de la información demográfica con cuestiones políticas e ideológicas, como ha quedado registrado en congresos en donde se ha discutido y reflexionado sobre lo comprometido que quedaría disponer de información desagregada por zonas geográficas bien delimitadas (Figueroa, 2006), como es de suponerse, esto implica exponer una realidad y cotidianidad que puede ser utilizada para cualquier fin. A este respecto, se alertó sobre las consecuencias metodológicas, de confiabilidad que repercutirían en la confianza de las oficinas de estadísticas, y a la larga generarán datos imprecisos en los casos en que las personas deliberadamente quieran ocultar información una vez que esta ha llegado a ser empleada en su contra, a pesar de la insistencia y promesa de apegarse a la confidencialidad (Figueroa, 2006; Dwork *et al.*, 2019).

Si bien el problema ético ha estado más cercano al ámbito clínico o biomédico, en el que se involucran los seres humanos, en la actualidad urge acercar esta problemática a las ciencias sociales y, para el caso que nos ocupa, a la demografía, debido a la indisoluble relación del quehacer de la demografía (y otras disciplinas) y los requisitos éticos de cualquier investigación, ya que se trabaja con seres humanos que pueden verse perjudicados por su participación en una investigación (biomédicos o sociales). Razón por la cual se sustenta que debe existir un comité ético que regule las actividades de todos los científicos sociales, para que al mismo tiempo que se sigue el camino de la búsqueda del conocimiento se considere al ser humano como un fin y no como un medio para el levantamiento de datos, de tal suerte que, los investigadores adquieran la responsabilidad de la investigación, la sociedad en general y su propio quehacer (Florencia, 2016).

En este punto, la discusión de si es necesario contar con un estándar ético “universal” o globalizado, se puede dejar para un segundo plano, ya que lo realmente relevante es contar con un conjunto mínimo de protecciones que garanticen a los participantes preservar su integridad, anticipar los posibles daños y buscar alternativas para minimizarlos. En este sentido, el consentimiento informado, escrito o verbal, adquiere relevancia porque siempre se comparte información confidencial (Figueroa, 2006; Florencia, 2016).

Este escenario produce una paradoja: actuar éticamente resulta contraproducente porque, al menos para el caso de México, se tiene la creencia que si no se publican los resultados obtenidos, la investigación es de poca calidad o deficiente; difícil encrucijada a la que se enfrentan los científicos sociales. Ante tal momento, surgen argumentos en contra de la evaluación ética: tener

presente la relevancia social de las investigaciones sociales y justificar la realización de la investigación (Flores, 2016).

Si bien es cierto que muchas investigaciones sociales son de gran importancia por los temas que abordan y debido a que los resultados son utilizados para hacer política pública, también es cierto que detrás de ello existen intereses y sesgos subjetivos.¹ Aunque hay mucho por reflexionar, los problemas éticos que por ahora interesan refieren a las estrategias metodológicas que se utilizan en las ciencias sociales, las personas y grupos con los que se investiga, el contexto y el tipo de vínculo que se establece con los participantes y otras cuestiones que pueden afectar a los participantes, ponerlos en riesgo físico o psicológico, poner en riesgo la privacidad o contribuir a la pérdida de la misma y los daños que se pueden suscitar; asuntos que deben considerar directamente el o los encargados de diseñar e implementar la investigación (Flores, 2016).

Para el caso de los métodos con los que se recaba información demográfica, existen sugerencias (aunque en todo el mundo estas se concentran en entrevistas y registros administrativos) en donde la confidencialidad de la información es necesaria, aunado a la advertencia que siempre debe existir del alcance de la información. Con respecto al daño que puede ser producido a una persona física o moral, vale la pena detenerse a reflexionar sobre los riesgos que implicaría hacer públicos datos que pudieran revelar la identidad de los individuos. En el caso de las entrevistas, usualmente semiestructuradas, para el caso de censos o las encuestas, aún con la planeación que se requiere se deben considerarse situaciones posibles como el trauma al que se lleva a los informantes cuando se abordan temas delicados. En estos casos se puede argumentar que existe cierto ejercicio de poder de parte de los entrevistadores, ya que los informantes responden las preguntas sin posibilidad de comentar algo más o detener la entrevista (Héran, 2013; Figueroa, 2006).

4.3.2 ¿Para qué desagregar la información?

Por todo lo anterior, partiendo de un aspecto en común entre las ciencias biomédicas y las ciencias sociales, referente a la relación especial que se establece entre los participantes, resulta pertinente

¹ En este orden de ideas, otro tema a tema interesante que por ahora no se quedará en el tintero es el del poder que ejercen las comunidades científicas investidas en los comités científicos encabezados por investigadores de universidades o centros de investigación de gran prestigio e incluso por la élite burocrática de institutos como el Inegi.

responder a esta pregunta. Una primera respuesta puede ser que la ciencia es autónoma y su desarrollo abona al bien de la humanidad, a este argumento podemos añadirle la idea del valor social de sus resultados, puesto que ya en innumerables ocasiones se ha comprobado que la información oportuna representa el insumo fundamental para atender problemáticas y tomar decisiones (Héran, 2013).

Para abonar a esta idea, tomemos como referente algunas situaciones que ocurren en la sociedad: día a día los aspectos de la vida cotidiana son referidos a datos; los medios de comunicación citan cifras todo el tiempo; los políticos utilizan a ultranza porcentajes para apuntalar sus posiciones; los ciudadanos obtienen datos de baja calidad y pocas veces son capaces de distinguir datos falsos, en algunos casos se vuelven escépticos de cualquier dato. En un contexto como este, la exigencia de datos confiables parece justificar el aumento en la demanda de información cada vez más desagregada (Héran, 2013).

La desagregación de datos se refiere a separar la información en unidades pequeñas, en universos reducidos para elucidar tendencias y patrones subyacentes, así que todo parece indicar que desagregar información es una prioridad, en el caso de los Objetivos de Desarrollo Sostenible (ODS) 2030, se han pensado en estos términos², siendo así, es necesario introducir una perspectiva ética para evitar la identificación directa o indirecta de las personas y salvaguardar los derechos de los informantes.

Canales (2004), considera que de igual forma que otras ciencias sociales, la demografía afronta un desafío que trasciende a las reformulaciones teóricas o metodológicas y dicho reto se centra en el “reposicionamiento de la disciplina, de nuevas formas de mirar y comprender los fenómenos demográficos” (Canales, 2004:48) y, particularmente, de un “cambio radical en la pregunta original que dio origen al pensamiento demográfico” (Canales, 2004:48) de ahí que, sin dar respuesta a estos interrogantes, la inquietud de Canales se suma a la reflexión ética que debe asumir. Es incuestionable su importancia como disciplina científica y no se somete a duda la

² Para estos objetivos, la desagregación es el desglose de las observaciones dentro de una rama común, de una jerarquía, a un nivel más específico al que se llevan las observaciones detalladas y sus indicadores deben desglosarse, siempre que fuera pertinente, por ingreso, sexo, edad, raza, etnicidad, estado migratorio, discapacidad, ubicación geográfica u otras características, de conformidad con los Principios Fundamentales de las Estadísticas Oficiales (resolución 68/261 de la Asamblea General). Consultar: https://www.cepal.org/sites/default/files/courses/files/los-ods-demanda-datos-desagregados-cepal-pauline_stockins.pdf

consolidación de esta, basadas en sus herramientas y técnicas específicas capaces de estimar el volumen, la estructura y las transformaciones originadas por la interacción de la fecundidad, la mortalidad y los movimientos migratorios de las poblaciones humanas (Welti-Chanes, 2011). Lo que sí debemos cuestionarnos es, qué tanto puede incidir el quehacer del demógrafo en la vida de las personas que conforman su fuente de datos o los insumos con los que la Demografía se nutre.

Una respuesta ingenua sería que el investigador no es capaz de transgredir esa cotidianidad porque, en la mayoría de los casos, los demógrafos utilizan bases de datos que han pasado por una “limpieza” que eliminan nombres, lugares, fechas que permitan la identificación del informante o de personas a las que se refieren y que la información utilizada por estos científicos sociales en todo caso se convierte en códigos capaces de realizar cruces, gráficas y así analizar a gran escala. También se argumentará que en general hay un énfasis en la perspectiva cuantitativa y que la interacción con los participantes de las encuestas es inexistente, ya que, para ser considerado un estudio demográfico, usualmente, se debe eliminar la cualidad y dar prioridad a la cantidad, sin embargo, la demografía tiene una perspectiva que supera lo puramente cuantitativo en la explicación de los fenómenos que son de su interés, aunque la privilegia (Welti-Chanes, 2011).

A causa de esto, las cuestiones éticas se han olvidado de la esfera demográfica, bajo la creencia de que el demógrafo nunca está en posición de tomar decisiones que afectan el comportamiento de las personas y conlleve a situaciones de riesgo o en donde se vea vulnerada su integridad física y se llega a comparar con disciplinas como la psicología o la medicina donde la interacción es más evidente, la demografía no interviene de manera directa, sino que lo hace a la distancia y en un momento en donde los dilemas morales ya han sido establecidos o descartados (Héran, 2013). Este autor argumenta que esta concepción conlleva al pensamiento de que la demografía es inocua, por trabajar con datos agregados y asumir que todos los individuos tienen un mismo peso y son contabilizados de manera similar. Aunque como sostiene Figueroa (2006), esta forma de analizar a la población a la distancia es el elemento clave para suponer que el quehacer demográfico está exento de dilemas éticos. Además, en términos de la producción de conocimiento el fin último de la ciencia es aumentarla, por lo tanto, entre mayor especificidad en la información (desagregación), se supondría que los datos tendrían que ser de mejor calidad.

Sin embargo, no se debe perder de vista que la fuente de información primaria siempre son las personas cuyas declaraciones son conectadas, ya sea en modo de datos administrativos o como

respuestas a los cuestionarios que se hacen en censos de población y encuestas. Así, la responsabilidad de tomar en cuenta a los individuos, cuyas respuestas son codificadas en una base de datos, plantea la necesidad de ahondar en el costo que tiene la producción de conocimiento (Mayer-Schönberger y Cukier, 2013).

Para que se cree una fila en una base de datos fue necesario que un individuo estuviera de acuerdo con responder una encuesta, tuvo que haber pasado un proceso en el cual fue cuestionado sobre su vida, de tal manera que el vínculo que existe entre la ciencia y la sociedad adquiere, inevitablemente, una dimensión ética. Herán (2013) revela un ejemplo que ocurrió durante el Censo del 2000 en Estados Unidos: la oficina encargada de llevar a cabo este ejercicio había hecho la sugerencia que se efectuaran entrevistas en profundidad en las áreas urbanas pobres, sin embargo, los miembros del Partido Republicano objetaron esta decisión porque temieron que hubiera un incremento en las preferencias del partido contrincante, la consecuencia de esta polémica fue la renuncia del director del operativo debido a que fue acusado de tener inclinaciones políticas, por lo que quedó claro que la neutralidad estadística había sido violada, y con ello, los datos agregados tuvieron repercusiones de índole éticas.

Otro ejemplo adicional que pone en tela de juicio la neutralidad estadística en el quehacer demográfico es el enfoque maltusiano, en donde el argumento central es que la población puede crecer desmesuradamente y crear un caos en la sociedad (Malthus, 1966). A raíz de que este postulado se ha tomado como un hecho, se han generado planes y programas para evitar que la población crezca, o al menos se han puesto frenos en algunos lugares específicos como en México, en donde a partir de 1970 se adoptó la idea de que el crecimiento poblacional traería consigo un desequilibrio económico y se originó el Consejo Nacional de Población (Conapo), como órgano encargado de regular el crecimiento poblacional (Welti, 2014). Hoy en día, sus planes más importantes siguen siendo los mismos que hace 40 años: promover una planificación familiar para evitar el crecimiento desmesurado de la población, sobre todo en zonas de bajos recursos económicos. Con esto queda claro que los ejercicios estadísticos, como las proyecciones de población, sí tienen eco y repercusiones en la política nacional.

En esta misma línea, en México, la reducción de la mortalidad en el primer tercio del siglo XX tuvo como consecuencia la modificación de las estructuras por edades y dio como resultado el rejuvenecimiento de la fecundidad. Es decir, la fecundidad se controló a partir de la Ley General

de Población proclamada en diciembre de 1974 y a partir de ello, a la fecha se sostiene una planificación familiar aún vigente. Con la Ley se creó el Consejo Nacional de Población que tiene a su cargo la planeación demográfica del país sin contar con las herramientas para cumplir con su tarea, de los 123 artículos que comprenden dicha ley, solo 6 refieren a acciones de población, mientras que el resto refieren a asuntos migratorios (Welti, 2014). Para este caso, es evidente que la intervención y el ejercicio del poder de los tomadores de decisiones implicaron el control de la fecundidad. Las discrepancias o acuerdos que están detrás son difíciles de conocer, lo que es innegable es que repercutió e incidió en la vida de las mujeres, dejó huella y generó un cambio en las conductas sociales (García y Ordorica, 2010).

Todo lo planteado hasta ahora refiere la necesidad de una mirada ética para nuestro quehacer como científicos sociales, en el caso de la Demografía, es necesaria una reconciliación con disciplinas como la Psicología y Sociología para aprender de ellas el camino andado que tienen con el fin de tratar de ajustarlo y abrir la posibilidad a otras opciones. En estas disciplinas la interacción con las personas (informantes) es más cercana, por ello sus investigaciones han sido etiquetadas como de corte cualitativo, pues utilizan estrategias metodológicas que se centran en prácticas muy distintas a las de la Demografía (Ordorica, 2011). Porque en esas metodologías y herramientas está quizá la forma de lograr información de mejor calidad, por ejemplo, antes de comenzar la entrevista se crea un vínculo que va más allá de la aplicación de un cuestionario, pues se trata de obtener la confianza, que en caso de lograrse, mejora la participación de los implicados en la investigación. Con lo que el consentimiento informado puede firmarse en un documento físico o no, porque el vínculo de confianza permite esta modalidad. En ocasiones, este vínculo trasciende el momento de la investigación y se obtienen resultados tangibles porque en la medida de lo posible se da seguimiento a la problemática de interés (Héran, 2013; González, 2018).

Una premisa en estas disciplinas es que todas las personas son distintas, algunas más susceptibles que otras, el tiempo se relativiza y se apuesta por la calidad de la información más que por la cantidad. Se apuesta a más de una técnica para la obtención de la información, lo cual implica que además de realizar entrevistas (estructuradas, semiestructuradas o en profundidad) también se escriben diarios de campo o análisis documental y fotográfico con el objetivo de tener un contexto, porque si se es muy sistemático estos agregados se convierten en un dato cualitativo (Ordorica, 2011). Otro recurso que es muy eficaz consiste en no pedir información personalizada, como nombre y dirección, usualmente se manejan seudónimos y en el tratamiento de la información, no

hay registro del nombre real de los informantes. La base de datos o el cúmulo de información construida no es compartida para otras investigaciones de interés similar, en estos casos se parte de los hallazgos de la investigación, pues se supondría que, lo publicable ha pasado por un proceso “objetivo” de análisis (Héran, 2013).

Una de las preguntas más importantes al momento de llevar a cabo cualquier investigación desde el área demográfica o de los estudios de población consiste en cuestionarse: ¿los beneficios esperados justifican la exposición a la que se ven sometidas las personas al participar en una investigación?

La respuesta a esta interrogante es ardua, ya que la cuestión ética frecuentemente pasa inadvertida, ya sea por la ausencia de una normatividad en las instituciones, o porque en el método científico están ausentes los límites de integridad. En todo caso, es necesario incluir la ética en todos los ámbitos, sobre todo cuando la investigación tiene intromisión en la vida privada de personas. No obstante, el acercamiento al ámbito ético comienza en la selección del tema de investigación, se consolida con la obligación que tiene el investigador con sus colegas y la comunidad científica a la que pertenece (universidad, institución, consultoría) de manifestar y reconocer sus desventajas técnicas, fallas y debilidades haciéndolas del conocimiento a sus pares y sus superiores (González, 2018). Una vez recabada la información lo ideal sería decidir qué datos se pueden difundir o utilizar para cumplir con los objetivos planteados aunque también entra en juego la disyuntiva moral del crédito o prestigio que puede atraer los resultados obtenidos (Ordorica, 2011).

La perspectiva ética continua al perpetuar la idea de que para validar la investigación los hallazgos deben ser publicados, omitiendo los retos y paradojas a los que se enfrentó durante el proceso. Por otra parte, y en concordancia con quien realiza la investigación hay una perspectiva ética siempre presente al momento de recabar la información, en donde de manera implícita está latente la posibilidad de que se puedan revelar datos o indicios de las personas y quien realiza la investigación deberá asegurar confidencialidad.

En México, existe una norma técnica para unidades de Estado y para el Inegi que se debe seguir al manipular o adquirir un domicilio geográfico y refiere a las funciones a cumplir cuando se tiene un cargo en la captación, actualización e integración de registros administrativos ejecutada

por sí mismas o por terceros. De acuerdo DOF (2008), cuando se localiza la posición exacta de las personas se debe de omitir esta información.

Es un riesgo que esta información esté concentrada en una base de datos accesible a personas que podrían utilizar dicha información para cualquier fin. Así, valdría la pena reflexionar sobre los “beneficios” personales, indirectos que supone la idea de participar en una investigación, estudio o sondeo. Disociar la idea del beneficio mutuo puede ayudar a entender para qué necesitamos esta información y reflexionar si más allá de los beneficios personales de quien emprende la pesquisa es necesario compartir los resultados y hasta dónde se puede socializar esta información.

Nunca se es realmente anónimo tal y como lo demuestran Kearns y Roth (2019) en su libro “el algoritmo ético”, ya que quien realiza la entrevista o indagatoria ha reunido la información necesaria que hace a la persona indistinguible. Por lo tanto, se ha preferido invocar a la confidencialidad que entre otras cosas es el compromiso de no hacer de uso público la información, desde este enfoque se incorporan connotaciones legales en quien solicita la información y en quien la proporciona. Un ejemplo de ello son los datos censales en donde se marca con un asterisco (*) el total de hombres o mujeres cuando dicho número es menor a 5, pero con la información total y distribución por edad, sería factible indagar el dato omitido.

Los datos personales incluyen la información que nos hace identificables (identidad, descripción y precisión) así como nuestra forma de pensar, estado de salud, preferencia sexual y demás rasgos que nos definen como persona en la sociedad. Prueba de ello es el auge de encuestas electorales previo a las elecciones. En los últimos años, esta información es cada vez más compartida y se ha convertido en un requisito indispensable para cualquier trámite, en el año 2010 se publicó la Ley de Protección de Datos Personales en Posesión de Sujetos Obligados de la Ciudad de México (LPDPPSOCDMX) para regular la posesión y el tratamiento que los particulares deben atender al recabar información tan delicada (GCDMX, 2021).

La ley de protección de datos aplica a todos aquellos que recaben información para realizar trámites o compras. Por otro lado, la ley que resguarda la participación para obtener información y nutrir las políticas públicas es la tan citada ley LSNIEG del Inegi. En el Artículo 103 reseña sobre las infracciones que se pueden cometer quienes, en calidad de informantes del Sistema (entrevistado), se niegan a proporcionar datos, exhibir documentos, proporcionen datos falsos e incongruentes, a quienes omitan inscribirse a registros establecidos por la Ley, y a quienes se

opongan a inspecciones a la verificación en cumplimiento de las disposiciones de esta ley. Todo esto engloba una idea que merece ser cuestionada, ¿estamos obligados a participar en encuestas o censos?, todo parece que sí, incluso se ha fijado una multa, la sanción es determinada por el departamento jurídico y depende de la gravedad del asunto (DOF, 2008).

Desde esta visión ¿cuáles serían las opciones que tenemos como ciudadanos al tener que entregar nuestra información?, porque a la luz de esta perspectiva los tomadores de decisiones, que pueden ser quienes usan la información y quienes recaban y administran nuestros datos adquieren una posición poderosa con pocas posibilidades de defensa para quienes no desean entregar datos que los definen. Las obligaciones de los informantes parecen justificarse por el supuesto beneficio que puede atraer participar o cumplir con la obligación de generar información, una inquietud recurrente es que el Instituto puede ser muy ético, no obstante, ¿quién garantiza que el encuestador es confiable?, es sabido que se trata de personal contratado por tiempo determinado, por evento y tienen el incentivo de obtener datos a toda costa, pues de ello depende su ingreso. A mayor número de encuestas, mayor beneficio económico. Aunado a lo anterior, la premisa bajo la que descansa la investigación o la ciencia es el espíritu voluntario. Sin embargo, bajo lo que se ha expuesto, es evidente que los sujetos de estudio o participantes no tienen la oportunidad ni de ser voluntarios ni de rehusarse a participar (Figuroa, 2006; Ordorica, 2011, O'Farrell y Montagnier, 2020).

La situación poco cambia al pedir a nuestros informantes u objetos de estudio que participen en una investigación académica, salvo la obligatoriedad bajo amenaza de ser multado. Quienes deciden hacerlo, también entregan datos valiosos a quienes, por solidaridad, empatía, engaño, por sacar algún provecho o por sentirse útiles responden algún instrumento (encuesta, entrevista). De esta forma, también se debería reflexionar sobre la posibilidad de incluir al informante y al investigador en un contexto que vaya más allá de cómo se desarrolla la obtención de la información. De esta manera, podemos concentrarnos en si debemos seguir haciéndolo de esta manera, lo cual implica una responsabilidad moral en la investigación (Figuroa, 2006; Ordorica, 2011). Hasta ahora, algunas reflexiones sugieren que quienes utilizamos datos lo hagamos con cautela, ya que podría ser que la información no sea totalmente fidedigna, y al ser llevar la batuta de una investigación, nos preguntemos antes de invitar o buscar informantes ¿participarías sin miedo a ser perjudicado en una investigación como esta?, y finalmente, preguntarnos ¿cuáles serían los límites a los que toda investigación (pública y privada) debería atender?, estas interrogantes encuentran su mejor reflejo en la disyuntiva ética que plantea el desarrollo de la inteligencia artificial (IA), ya

que muestra de manera clara las ventajas y desventajas que traería. Cabe recordar que en la historia de la ciencia los científicos siempre se han enfrentado a consideraciones éticas (Russell y Norvig, 2004).

4.3.4 Problemas con la apertura de los datos

Con el auge de nuevas y mejores formas de conectar información, los datos se han empezado a considerar como un bien público; esto significa que ha habido un reconocimiento de su potencial para el diseño de políticas públicas y ha prosperado la filosofía *Open Government Data* (OGD), la cual promueve rendición de cuentas y transparencia en torno a los datos que generan los diferentes gobiernos. Durante este proceso se promueve que la información esté disponible para que las instituciones públicas sean más transparentes al momento de tomar sus decisiones (Jetzek *et al.*, 2014). Para apoyar esta iniciativa, la OECD genera un índice en donde evalúa a cada una de las naciones en torno a 3 criterios importantes: apertura, utilidad y reutilización de los datos gubernamentales. De acuerdo con el último estudio disponible para 2019, el país mejor evaluado fue Corea del Sur y el último lugar Turquía.

Aunque en primera instancia el hecho de tener una apertura total a los datos es favorable para analizar el comportamiento de la población ¿cuáles son los problemas de hacer públicos los datos?. Actualmente y debido al uso ilícito que se ha hecho de los datos de la población existe la preocupación en torno a la protección de los datos personales y la posibilidad de que, los nuevos algoritmos computacionales, puedan revelar la identidad de los informantes. Para evidenciar esta preocupación pondremos de ejemplo el Censo de Estados Unidos de América en 2020, en el cual se utilizó una herramienta denominada privacidad diferencial, que consiste en aplicar un algoritmo matemático que intencionalmente genera “ruido” en cualquier conjunto de datos. El objetivo es garantizar la privacidad de los informantes y evitar que puedan ser identificadas por sus características demográficas o por medio de las respuestas que están codificadas en el Censo. La implementación de este algoritmo responde a un episodio durante el Censo de 2010; en ese momento se logró identificar a cerca de 50 millones de individuos a partir del análisis de los datos anónimos. Esta situación de violación de la privacidad generó pánico y se puso sobre la mesa de discusión el tema de que existían posibilidades de vulnerar la información y hacer pública la identidad de quienes respondieron el Censo 2020 (Evans *et al.*, n.d.).

Aunque la seguridad de los datos debe ser prioridad, la aplicación de este algoritmo en los datos censales produce resultados incongruentes. De acuerdo con Garfinkel *et al.* (2018) dicha inconsistencia puede observarse en que actualmente en el Censo se pueden identificar hogares encabezados por menores de edad.

Ciertamente, esto confunde y pone en duda la interpretación de los datos, además de que frena la posibilidad de indagar con mayor detalle los fenómenos que atañen al gremio demográfico. Por este motivo, diversos investigadores que utilizan los censos de población como su fuente de información primaria, redactaron una carta en donde pidan transparencia respecto a la aplicación que los algoritmos para generar los datos censales (Garfinkel y Leclerc, 2020). Como parte de su petición, solicitaron un archivo en donde se omitiera la aplicación del algoritmo para tener acceso a los datos brutos. Esta petición estaba respaldada en el argumento de que es la única manera de detectar los cambios socio demográficos que se desarrollan en escalas pequeñas. Al momento que se escriben estas líneas sigue en debate la pertinencia de crear un archivo con los datos en bruto y no existe una respuesta por parte de las oficinas del Censo debido a que la disponibilidad y apertura de los datos podría poner en evidencia errores u omisiones de parte de la oficina censal (Inegi, 2021).

Aunque en el caso mexicano se experimenta una situación distinta y al momento no hay controversias en torno a los datos censales de 2020, en 2018 el gobierno emitió un acuerdo mediante el cual la Secretaría de Comunicaciones y Transportes (SCT) estableció una serie de lineamientos para fortalecer los servicios satelitales y el uso eficiente de sus capacidades. A partir de esta nueva disposición se reconoce el potencial que tiene el empleo de las nuevas tecnologías y se fomenta el desarrollo de las capacidades satelitales propias (DOF, 2018).

Para el presente texto, este acuerdo tiene varios puntos destacables en materia de investigación y posibilidades para la Demografía. Por ejemplo, reconoce el papel de la educación como catalizador del avance tecnológico que se ha vivido en los últimos tiempos, así como la necesidad y la importancia de una actualización constante. Además, promueve la cooperación internacional y sugiere aprovechar las alianzas existentes para extender la cooperación con América Latina y el mundo.

Por este motivo una de sus prioridades consiste en fortalecer la Oficina de Asuntos Internacionales del Gobierno en Materia Satelital y fomentar la colaboración con otros países (DOF, 2018). La iniciativa parte de la idea de que México cuenta con un sector satelital maduro y

promueve la creación de un comité de internet satelital en el cual pueden participar tanto actores públicos como privados, así como instituciones académicas y de investigación. Este comité tiene el propósito de definir indicadores que midan la eficacia de las políticas satelitales del Gobierno federal, brindar asesoría en estos temas e impulsar el desarrollo de las capacidades nacionales, con lo que el gobierno mexicano reconoce el potencial que posee la información obtenida en los sistemas de comunicación satelitales y promueve su desarrollo.

4.4 Dilemas éticos en el uso de datos geoespaciales

Como ya se ha mencionado, el uso de los datos conlleva dilemas éticos. Sin embargo, en la era digital estos temas no se han puesto sobre la mesa de manera suficiente. En la investigación realizada y presentada en estas páginas únicamente se encontraron cuatro documentos que tratan el punto de la ética en el empleo de datos geoespaciales. Asunto que debería ser debatido con mayor predominancia en los institutos de investigación y en las empresas privadas dedicadas al análisis de los datos.

En el presente, el análisis estadístico asiste a un profundo proceso de reconversión marcado por la era digital en la que cada vez es más frecuente el empleo de técnicas distintas a las tradicionales para construir objetos de investigación que resulten de interés para la sociedad. El reciente problema mundial de la pandemia por COVID-19 mostró la importancia de contar con datos oportunos (CEPAL, 2018). El alcance de estas técnicas exigen la modernización de las estadísticas tradicionales y, por ende, plantean nuevos desafíos técnicos y éticos; debido a que las nuevas fuentes de datos y técnicas como la vinculación de datos y los datos geoespaciales presentan disyuntivas en el ámbito de la ética (Golden *et al.*, 2005).

Los datos geoespaciales han puesto en jaque a la estadística tradicional y sus respectivos principios éticos, entre otros temas porque su utilización es masiva y relativamente fácil acercarse o construir una base de datos de este tipo. Estos datos geoespaciales tienen mucha flexibilidad y permiten moldear y recortar la realidad bajo ciertos criterios como el comportamiento de ciertos elementos. De hecho, es justo en esta flexibilidad de recorte de la realidad para dar cuenta de un problema de investigación es donde se encuentra la discusión más álgida sobre la utilización que se debería dar a esta valiosa fuente de datos (UK Statistics, n.d.).

Hay una razón para adentrarse en los dilemas éticos al poner a disposición de todos los datos satelitales; la información puede utilizarse para predecir, planear y verificar un asunto, fenómeno o política pública, pero también puede ser empleada para vigilar, registrar y restar privacidad a las personas, las sociedades y la humanidad entera. Actualmente, el manejo y empleo de esta novedosa fuente de datos derivada de la tecnología es de dominio público y se ha convertido en parte de la vida cotidiana de la población, basta con buscar una dirección desde alguna aplicación como la Street View de Google Earth, recibir y pedir productos a domicilio por medio de Amazon, compartir tu ubicación como un sustituto de proveer tu dirección, y realizar una transacción financiera en la banca electrónica. En todos estos casos el algoritmo puede rastrear tu ubicación y hacer cualquiera puede ver un historial de los movimientos más frecuentes (Mayer-Schönberger y Cukier, 2013).

Aunque el registro espacial se ha practicado desde hace tiempo, en marzo del año 2020, cuando las oficinas estadísticas suspendieron operaciones, la tecnología geoespacial se apuntaló como una herramienta usual para dar cuenta del comportamiento de contagios, y en general, ofrece una experiencia personalizada a los usuarios al otorgar mayor precisión en la ubicación y ofrecer productos o servicios cada vez más vinculados con sus necesidades más inmediatas (CEPAL, 2021, 2018).

Al mismo tiempo, la utilización de herramientas geoespaciales disminuye la privacidad e incrementa el riesgo de todos los usuarios a ver comprometida su información de geolocalización que se almacena, dejando serias dudas de si los datos se destruyen o pueden ser transferibles. En este sentido, el uso público de los datos tiene la disyuntiva de poner lo privado al servicio de la ciencia, e implica que quien los recaba tendría que hacer un tratamiento especial y mantenerlos en resguardo. Ante este panorama los usuarios deben mantener cautela con el manejo de la información recopilada, ya que, puesto que los datos son de todos, la privacidad pende de un hilo delgado, principalmente por la aparición de delincuentes cibernéticos y el acoso mercadotécnico de algunas aplicaciones (CEPAL, 2018).

Partiendo de este contexto es innegable la utilidad de este tipo de información. Sin embargo, los datos de los Sistemas de Posicionamiento Global (GPS) y las etiquetas de geolocalización, combinadas con información personal o privada puede ser utilizada para distintos fines. Ante este panorama resulta indispensable hacer una pregunta ¿quién regula las medidas indispensables para el manejo de esta información? Algunos de los actores involucrados en el uso de los datos

argumentan que es responsabilidad del usuario de aplicaciones vigilar las que emplean datos de geolocalización, lo cual puede ser parcialmente cierto, sin embargo, qué sucede cuando dicha geolocalización es involuntaria, al compartir imágenes y videos que contienen metadatos de localización (UK Statistics, n.d.). Por tanto, vale la pena reflexionar en los siguientes tópicos, ¿quién regula la producción de los datos? ¿Quién puede acceder a ellos? ¿Cómo puede prevenirse el uso indebido de los datos geoespaciales?

La utilización de datos geoespaciales conlleva un dilema central que va de los beneficios y los riesgos que conlleva su aplicación en la investigación científica, hasta los principios morales que deben ser observados por quienes aplican la ciencia. La ética, entonces, es una poderosa disciplina que puede dirigir el comportamiento humano en la producción, uso y consumo de datos georeferenciados (Marshall, 1991; Smith y Wise, 2007).

Si bien la cartografía (croquis, planos y mapas) se realizaba a mano alzada durante el siglo pasado, visitando el espacio para dibujar planos topográficos elementales en un sistema coordenado cartesiano, en el que la reseña de una casa, un edificio, una plaza, permitía situar de manera general a la persona, vivienda, evento. Esta información se mantenía en resguardo del investigador quien les garantizaba a los informantes no utilizarla en asuntos que los comprometieran o que los expusiera públicamente. Es decir, el investigador tenía la obligación de evitar causar daño a otros y buscar el consentimiento informado (Evans *et al.*, n.d.; Garfinkel *et al.*, 2018).

Actualmente, la cartografía es digital y la gran mayoría de los datos que utilizamos para producir estadísticas se pueden ubicar al vincular a direcciones, códigos postales o áreas locales, o con GPS en ubicaciones más precisas, lo cual lo hace de dominio público y pone en riesgo un pilar central en la estadística: la confidencialidad o el anonimato. Esta situación ha puesto focos rojos a las buenas prácticas estadísticas que podrían privilegiar el uso extendido de métricas populares en detrimento de la confidencialidad, seguridad y calidad de los datos (Kearns y Roth, 2019). Este problema implica diversas amenazas a la ética de por lo menos tres actores, 1) a las oficinas estadísticas como generador y proveedor de datos oportunos y oficiales; 2) al usuario, quien puede decidir cómo y en qué casos explotar o aprovechar los datos disponibles y, finalmente, 3) al informante el actor más vulnerable en por lo menos dos vertientes: a) al participar en el levantamiento de los datos (sea consensuado o no) y b) al difundirse los resultados del estudio, tesis, diagnóstico, o cualquier otro dictamen que coloque su condición en el conocimiento público

o que pueda ser descubierto en la localidad o muestra debido a la gran precisión que hoy día nos arroja un dato georreferenciado (Golden *et al.*, 2005).

Los informantes, pese a ser el centro de los levantamientos de datos, pocas veces son considerados como pieza esencial para la construcción de la estadística y su participación conlleva dos temas éticos fundamentales: la confidencialidad y el consentimiento. Estos conflictos se refieren a la privacidad, la seguridad de la información o la manipulación de los datos, ya que el hecho de que los datos con los que trabaja el investigador sean cuantificables no supone de manera automática que equivalgan a la verdad absoluta, en especial si esos datos no se contextualizan adecuadamente.

En los últimos años, se ha realizado un esfuerzo por definir una serie de principios éticos importantes para la investigación y las estadísticas que entre otras cosas plantean un marco basado en obligaciones que deben regir cualquier investigación científica que implique la presencia de seres humanos como la privacidad o la protección de las fuentes, así como la advertencia de los posibles perjuicios y daños que se puedan causar cuando se difunda la información. Estas cuestiones éticas son lo suficientemente amplias como para abordarlas de manera específica, así que en el presente texto se acotará a los datos geoespaciales y a las recomendaciones desarrolladas por el Centro de Ética de Datos Aplicados de Estadísticas del Reino Unido (UK Statistics, n.d.). A continuación, y a partir de la revisión de los cuatro documentos existentes que se mencionaron al inicio de estos apartados, se abordarán las recomendaciones más recientes en esta materia.

La ética en la investigación proporciona una estructura para evaluar los riesgos y beneficios potenciales para la comunidad en que se realiza la investigación, determina un curso apropiado de acción tomando en cuenta normas internacionalmente aceptadas y valores culturales; se aleja de prescribir una serie de normas o políticas específicas y conduce a través de acciones de seguimiento y vigilancia. Los principios éticos proporcionan pautas para tratar con todo tipo de datos y auxilian en la planificación de la recopilación de datos, la elección de métodos y las decisiones sobre cómo se presentarán los resultados (Golden *et al.*, 2005).

Los principios éticos que a continuación se reseñaran se derivan de la estrategia geoespacial del Reino Unido y servirán de guía para asegurar el bien público de la investigación y las estadísticas, mantener la confidencialidad de los datos, comprender los riesgos y limitaciones potenciales en los nuevos métodos y tecnologías de investigación, así como cumplir los requisitos legales y la transparencia en la recopilación, uso e intercambio de datos. Los datos de ubicación o

datos geoespaciales tienen consideraciones éticas muy particulares porque además de ser estáticos y ubicar a personas y direcciones en tiempo y espacio, pueden ser dinámicos y brindar información en tiempo real, por esta razón son considerados datos íntimos y sensibles; por ello es necesario contar con una serie de principios éticos que ayuden durante la investigación. Debido a su naturaleza, los datos geoespaciales alertan especial cuidado en ámbitos como la confidencialidad, la inclusión, los prejuicios y la discriminación; y la búsqueda de garantizar su justa representatividad. Por lo tanto, desarrollaremos los aspectos éticos de cada uno de los casos de cuidado especial en el tratamiento de datos (Kearns y Roth, 2019; Marshall, 1991).

En relación con la confidencialidad, los datos geoespaciales pueden ser reveladores en los siguientes casos: 1) si se es muy específico en la caracterización de una población; 2) si el área espacial seleccionada es tan pequeña que impida mantener un bajo perfil o el anonimato del grupo estudiado; 3) si al combinar distintas fuentes de datos la población es claramente identificable. Ante estos escenarios se deben proteger los datos delicados, principalmente si se trata de la ubicación de sitios que comprometan la seguridad de personas refugiadas, en investigación o resguardo (DOF, 2018). Para ello es necesario que quien investiga (acompañado por los comités de ética que vigila la investigación) elija cuidadosamente la representación gráfica que usará en presentaciones, avances de tesis e incluso en la publicación de sus resultados.

En el ejercicio de la ciencia y con el avance de la tecnología, el uso de datos geoespaciales puede ser utilizado como un elemento de veracidad en la investigación. Y, aunque mucho se ha debatido en torno a si la fuente de datos es un peligro a la confidencialidad, el riesgo durante la etapa de divulgación se concentra en cómo se emplea dicha información (González, 2018). En estos casos el cuidado debe concentrarse en la revisión de la privacidad en imágenes obtenidas; a este respecto, las recomendaciones del Reino Unido para mitigar los riesgos enunciados anteriormente se tratarán en los siguientes párrafos.

Al elegir un mapa, se debe reflexionar si la información que se presenta es reveladora; también se deben considerar distintas opciones al seleccionar mapas impresos o digitales y el usuario debe elegir hasta qué nivel la información estadística será representada. Se sugiere proteger los datos y desvincularlos de la identidad de las poblaciones o entidades analizadas, para ello se puede usar la liberación limitada de datos transformados, datos sintéticos y respuestas aleatorias. También se recomienda elaborar esquemas de acceso que sean seguros para la protección de los datos y a los que tengan acceso solo quienes validan la información (UN, 2000; UK Statistics, n.d.).

En el caso de investigación para estudiantes, los comités de ética y los directores de tesis son quienes van resguardando la información (Figuroa, 2006). Para investigadores y casas consultoras, la información queda resguardada bajo contratos o convenios de confidencialidad que los obliga a resguardar datos críticos. Esto se relaciona con el consentimiento otorgado para el uso de los datos personales que siempre debe buscarse. En caso de utilizar datos, se debe reflexionar sobre cómo se ha obtenido el consentimiento, si las personas aceptaron participar en la investigación y se les informó todo lo que implica (recabar y otorgar datos confidenciales); si la información recabada es transparente y lo suficientemente responsable para ser compartida aún entre comunidades científicas o colegas que garanticen responsablemente a su utilización (Ordorica, 2011).

Sin ser la intención final, al seleccionar una unidad de análisis se excluye a otras unidades, esto es, si se analiza a la población urbana, se excluye a la rural, o si se realiza una investigación cuantitativa, el análisis cualitativo queda descartado. Lo mismo sucede con la selección de la muestra y la fuente de datos, su selección delimita y exceptúa de un gran número de posibilidades (Inegi, 2021). Por lo anterior, la recopilación de datos debe tener en cuenta cualquier sesgo en su análisis y en la medida de lo posible informarlo para conocer los alcances y limitaciones.

El uso de datos geoespaciales, no es la excepción. Al seleccionarlos, se descarta a otro grupo de datos, aunque tienen una ventaja sobre los datos tradicionales: se producen en grandes cantidades, son fáciles de asegurar, codificar y recopilar. Sin embargo, pese a lo extendido que pueda ser su utilización, y su recolección, tienen dos grandes desventajas, son más vulnerables (por aportar datos confidenciales) y no reflejan a toda la población, por más masificado que sea el instrumento (UK Statistics, n.d.). Por lo anterior, el control en la recopilación de datos debe ser impecable y, aun así, debe tener en cuenta que siempre habrá sesgos, incluso cuando se trabaja con datos censales, que en principio, y por un corto tiempo, se puede sostener que el levantamiento fue aplicado a toda la población. Las recomendaciones para aminorar los sesgos es cuestionar si el método de recolección excluye a algún grupo y si la muestra es realmente representativa. Estas interrogantes deberán traducirse como contexto de análisis y si es posible en un apartado metodológico que advierta lo que se puede esperarse de la investigación (Golden *et al.*, 2005).

El camino metodológico que se recorre para la realización de la investigación tiene el potencial de advertir las posibles contrariedades y planear cómo salir de ellas. El hecho de que los datos geoespaciales excluyan a ciertos grupos debe documentarse y comunicarse, este apartado es

el idóneo para hacerlo. Por otro lado, se puede incluir el uso de fuentes de datos adicionales para la mejor comprensión de las poblaciones o comunidades, así como para minimizar la exclusión de cualquier grupo. La representatividad debe quedar explícitamente mencionada, porque los resultados obtenidos podrían ser generalizados para toda la población (Bell, 2015).

Siempre que un investigador realiza una investigación busca conducirse objetivamente y dejar de lado lo subjetivo. Sin embargo, mucho se ha documentado que dicha separación es imposible porque al seleccionar un tema o unidad de análisis lo objetivo y subjetivo se conjugan (Canales y Lerner, 2003). En este tenor, los datos geoespaciales, son más propensos a poner a trabajar los aspectos subjetivos y de percepción al dar a conocer una ubicación específica.

Por ejemplo, Vieira (1973) argumenta como las áreas geográficas reflejan condiciones socioeconómicas particulares, son más factibles a las suposiciones basadas en preconcepciones. Estos sesgos se pueden presentar en quien ejecuta la investigación, en quien evalúa los resultados y en quienes utilizan los resultados de la investigación. Esto nos remite a una cuestión epistemológica de la objetividad de la ciencia, así que se debe poner atención en la forma en que se recopilan o emplean los datos espaciales. En todo caso, se debe apelar en la construcción de los instrumentos para recabar información y evitar dar por hechas las hipótesis con las que se trabaja de manera inicial.

Es un error muy recurrente en las investigaciones que los investigadores busquen confirmar teorías y no se preocupen por confrontar con los datos obtenidos, también es frecuente que los analistas se vean influenciados por comunidades científicas para avalar ciertas tendencias teóricas de moda y encuentren en los datos geoespaciales un semillero ideal para confirmarlas (UK Statistics, n.d.).

En este sentido, en México muchas veces ocurre que al hablar de teorías y datos sobre la pobreza se emplean ejemplos de Estados del sur del país para confirmar posturas ideológicas y teóricas sobre el tema, desconociendo, con o sin intención, que la pobreza también se encuentra en otras regiones. Es decir, como el uso de datos geoespaciales tiene el potencial de perjudicar o arruinar injustamente un área geográfica o una comunidad, el uso y la interpretación de los mismos dependerá de cuestiones éticas, del contexto, del propósito del proyecto y de la naturaleza de las fuentes de datos usadas (Czarnecki, 2013).

Cuando se identifica una ubicación y se publican los resultados, se debe poner atención en cómo son presentados dichos datos. Las visualizaciones nos ayudan a aprovechar nuestra capacidad

para percibir patrones a través del sentido de la vista, por lo tanto, si se divulgan las ubicaciones y a su vez se vinculan con comunidades o temas socialmente conflictivos, el resultado puede provocar un impacto a largo plazo e incluso podría dejar en la memoria colectiva repercusiones que deberían ser reflexionadas (UN, 2014; Eckman *et al.*, 2018).

Por ejemplo, cuando se quieren observar fenómenos como la delincuencia, el solo hecho de resaltar determinadas áreas en un mapa de inmediato la estigmatiza; pero si, además, son resaltadas en color rojo es muy probable que visualmente se deje una impresión negativa de la zona, entidad o área geográfica. Así que ante datos como el anterior (en el que muchas veces no hay un uso adecuado de los colores) el papel del analista consiste en ser imparcial y comunicar de forma clara y abierta cualquier riesgo o debilidad en el análisis (Vieira, 1973; Campos *et al.*, 2020.).

Las sugerencias para mitigar la discriminación y los prejuicios consisten en seleccionar las fuentes de información que se utilizarán y considerar sus sesgos para aminorar los vacíos. Reportar en cada etapa cualquier sesgo en la selección de los datos para dar cuenta de las distorsiones que se presentan y advertir a futuros usuarios de la información, la salvedad que deben considerar (Kearns y Roth, 2019). Considerar e incluir como parte de su justificación los prejuicios y la discriminación que se tienen sobre una área o tema seleccionado y reflexionar como podrían influir en el enfoque estas suposiciones. Valorar si la presentación de ubicación tendría un impacto negativo en la población o las entidades analizadas, y de ser así, plantearse la posibilidad de usar otras maneras de presentación de datos para salvaguardar la ubicación real. De ser posible se sugiere incluir la participación de miembros de la comunidad o expertos en el tema en cuestión para comprender y transmitir el contexto social y cultural de manera adecuada (UN, 2014).

La oferta de fuentes de información y mapas es cada vez mayor; por lo que el acceso y la sofisticación de la tecnología nos acerca cada vez más a mejores fuentes. Sin embargo, ante esta abundancia debe prevalecer la premisa de que la elección y el diseño de mapas debe tener una visión justa de la realidad. Es decir, el recorte de la realidad a través de un mapa debe contar con una dimensión que busque representar el fenómeno de interés. Dadas las características de un mapa, este puede ser empleado de manera tendenciosa, para mostrar una parte de la realidad o para resaltar una característica que nos interese (con intención o no) (UN, 2000; Wijedasa *et al.*, 2012; Xu *et al.*, 2020).

A partir del análisis realizado en el estado de la cuestión en este tema, también se sugiere emplear una codificación visual objetiva, con una gama de colores que permita contrastar la

información y respaldar el mapeo, y recordar que, la información y los resultados obtenidos se pueden replicar fuera del contexto de la investigación (UK Statistics, n.d.). Para conseguir tal cometido en una investigación conviene preguntarse ¿la información es parcial o tiene una carga subjetiva?, ¿el mapa que se empleará falsea la realidad?

La cartografía ha dado a la historia importantes aportes, su uso permite conocer la evolución de la humanidad, nos remite al pasado y es capaz de revelar estilos de vida y características que definen y delimitan a la sociedad. En la actualidad, el empleo de los datos geoespaciales tiene el potencial de conocer en tiempo real dichas características y puede ayudar a pronosticar con mayor certeza ciertos fenómenos sociales. Los datos que atraviesa y recauda permiten tocar tramas que van desde el cambio climático, epidemias, conflictos territoriales, geopolítica y economía, feminismo, desapariciones forzadas, y demás temáticas que tienen la capacidad de generar cambios sociales influyentes. Debido a estas características, la utilización de los datos geoespaciales enfrenta dos ofensivas: los riesgos y los beneficios (Campos *et al.*, 2020; Eckman *et al.*, 2018).

Desde la perspectiva de los beneficios, existe una gama de oportunidades que la sociedad tiene al contar con datos abiertos, pues su acceso a la información se vuelve plural y democrático, por lo tanto, dejan de estar concentrados en unas cuantas “manos”, enriquece los análisis y permite tomar decisiones de forma más responsable ya y basada en “evidencia”. En cuanto al riesgo, este se concentra en la amenaza a la eliminación de la privacidad, debido a que son datos detallados que vulneran la identidad de las personas, quienes muchas veces participan en investigaciones privadas y públicas. En este sentido, la privacidad puede ser considerada el bien más preciado para dar continuidad a los beneficios de los datos geoespaciales y en todo caso contrarrestar la relación de poder que se pueda ejercer (Vieira, 1973).

Los datos abiertos geoespaciales, como un bien público, pueden ser utilizados para cualquier beneficio, así que hace falta regular por lo menos dos aspectos: a) la relación de poder entre quien “produce” y b) cómo se produce esta información, es decir, el contexto de producción. Debido a que la producción del conocimiento occidental moderno tiene como base a la ciencia, que a su vez reúne a comunidades científicas que avalan métodos para obtener información, también supone aceptar la herencia de teorías, metodologías, conceptos sin cuestionar si son vigentes o no al contexto social específico, lo cual es importante porque se perpetúan directrices que van más allá de ser simples marcos de actuación e influyen notablemente en las prácticas y estilos de visualizaciones geográficas (Yang, 2021).

Por lo anterior, de forma concatenada, quién produce los datos está afiliado o pertenece, por convicción o para acreditar su investigación, a este estilo de hacer datos geoespaciales, por lo tanto, replicará su concepción del mundo al representar un mapa, al elegir qué componentes se incluirán en él. En conjunto, es un modo de ejercer el poder, toda vez que se trata de una manera de actuar sobre otros, y ello mientras estos actúan o son susceptibles de actuar. En este caso, el poder se centra sobre quienes no cuentan con la información para indicar que los datos geoespaciales de una investigación deban ser realizados de alguna otra manera, o que no deban incluir ciertas características que vulneren la privacidad de las personas. Algunos ejemplos serían, los ciudadanos pocas veces tienen injerencia en el levantamiento censal de un país y de cómo se presentarán los resultados (UK Statistics, n.d.).

En otro ejemplo, el ejercicio del poder también se encuentra aún con la firma de consentimiento para participar en una investigación que vaya a producir datos geoespaciales, pues no existe garantía de eliminar el riesgo de la pérdida de la privacidad, finalmente, los datos personales han sido recabados y entregados, así que ya no se tiene control de ellos. Esto aumenta el peligro de la privacidad aun en países en los que el uso de los datos personales está protegido legalmente, debido a que dichas jurisdicciones son bastante específicas y en Internet los datos son de todos. Sin duda, ha habido muchos esfuerzos para controlar los riesgos de perder la privacidad, uno muy común es controlar dichos datos a través de solicitudes individuales para evitar que los datos públicos se conviertan en datos abiertos, sin embargo, esto solo sería aplicable en los casos en los que, dichos datos sean gestionados por oficinas gubernamentales u oficinas estadísticas (Kearns y Roth, 2019; Canales y Lerner, 2003; Campos *et al.*, 2020).

En particular, la Organización para la Cooperación y el Desarrollo Económicos (OCDE) ha trazado ocho principios básicos de privacidad que deben tenerse en cuenta cuando un investigador, una institución gubernamental u otro actor recopila y retiene información personal. Estos principios incluyen: 1) limitar la recopilación de información personal; 2) garantizar la precisión y la integridad; 3) especificar el propósito para el que se recopilan los datos; 4) usar los datos solo como se especifica; 5) proteger los datos con una seguridad razonable; 6) apertura en las prácticas de datos y la identidad del administrador de datos; 7) participación individual en sus datos; y 8) responsabilidad de los propietarios de los datos con los principios anteriores. Algunas sugerencias adicionales para mejorar la privacidad descansan en evaluar los objetivos de la publicación de los datos, los datos abiertos, no tendrían que publicarse a menos de que fuera necesario hacerlo. Si se

publicasen se deberá adoptar un proceso para que puedan llegar a asociarse con personas (OCDE, 1980).

Ahora bien, un elemento que puede ser clave para mantener la privacidad es incluir herramientas técnicas- tecnológicas que resguarden la privacidad al publicar datos. Además, otra forma de resguardar la privacidad es utilizar seudónimos para las personas, bajo el entendido que aún con ello, en un análisis profundo se podrían asociar e identificar a las personas. Aunado a lo anterior, se debe contemplar la anonimización de los datos personales, pese a que se podría identificar los datos si se ocultan en las bases y si se mantiene información adicional que haría un cruce de información, capaz de detectar a quien se busca ocultar (Garfinkel *et al.*, 2018).

Los metadatos constituyen, entonces, una herramienta de uso común debido a que los datos son agrupados en cantidades muy específicas. También se podría restringir la manera que se reutilizan los datos, dichas restricciones pueden derivar en licencias en las cuales, el propietario de los datos tiene la capacidad de limitar su distribución. Finalmente, una manera recomendable para mantener la privacidad es restringir el acceso a los datos a cierto número de personas. Esta medida es considerada un término intermedio entre la total apertura y la restricción completa. Para objeto de esta investigación cabe mencionar que hay cuatro tipos de restricciones de acceso; datos gestionados, basados en atributos, públicos y sin procesar. Cada tipo de restricción de acceso proporciona un nivel diferente de protección de la privacidad. No obstante, el riesgo de perder la privacidad sigue latente, ya que aún existe la posibilidad de que quien tenga el acceso pueda hacer un uso indebido de ellos (Dwork *et al.*, 2019).

4.5 Acuerdos sobre la ciencia

Actualmente, es común hablar de las nuevas tecnologías y las capacidades tecnológicas disponibles, sin embargo, conviene reconocer que este fenómeno se enmarca en la investigación académica fundamentada en el método científico. Este método implica varias fases, pero una de las más importantes es la delimitación del tema. Sobre este asunto, de acuerdo con Bunge (1997), Fucaraccio, (1975), Galison y Stump (1996) existen algunas restricciones como:

- i. El tema, ya que debe cumplir con una perspectiva o enfoque particular
- ii. Temporal, pues debe adscribirse a una época o momento histórico
- iii. Espacial, ya que estar ubicado en un lugar específico

- iv. Recursos, pues es innegable las limitaciones presupuestales o de tiempo
- v. Riesgo de la privacidad, ya que en busca de generar un conocimiento se puede revelar información sensible. De ahí la necesidad de tener conciencia ética al momento de usar los datos (Figueroa, 2012).

De forma generalizada, bajo este esquema operan todos los centros de investigación, universidades y sitios que producen “ciencia”. No obstante, delimitar un tema de investigación supone recortar la realidad a partir de un foco de interés propio y ajeno, pues es bien conocido que el conocimiento se construye en comunidades científicas (Galison y Stump, 1996).

De esta forma, los esquemas presentados anteriormente están respaldados por comunidades científicas que pueden subdividirse en subcomunidades de acuerdo con campos temáticos particulares de la ciencia. La comunidad científica se sustenta en un paradigma científico imperante, cuya vigencia dura hasta que se presente una evidencia sustancial que contradiga o invalide dicho paradigma. Por ejemplo, tal y como lo reconocen Peuquet y Duan (1995), el enfoque de tomar en cuenta a los eventos y explorarlos a través del tiempo y el espacio puede ser ideal, pero aún sigue en el ámbito teórico, ya que no existe una metodología unánime que permita llevar a cabo este procedimiento. Esto muestra cómo el paradigma científico es un conjunto de teorías, modelos o esquemas que dan estructura a la manera en la que los científicos hacen ciencia.

La perpetuidad del paradigma vigente y la comunicación entre miembros se legitima a través de la publicación de trabajos de investigación en revistas revisadas y acreditadas por miembros de la comunidad científica. También, mediante la asistencia de los investigadores a congresos y su participación en mesas donde se intercambian, debaten y divulgan los resultados de las investigaciones realizadas (González, 2018). Ahora bien, el acercamiento que efectúan las investigaciones a la realidad a la que tienen acceso conlleva otro supuesto que la epistemología de la ciencia ha venido discutiendo: ¿los datos están dados o se construyen?

Para responder esa interrogante es indispensable mirar el panorama o contexto donde se generan los datos. En cada aproximación que un investigador hace al objeto de estudio, se propone adquirir un conocimiento cada vez más profundo y completo sobre el fenómeno de interés, pero quien investiga es un sujeto que carga con su subjetividad y que no puede dejarla. Esta subjetividad modifica la generación de los datos y, aunque, pese a que cada vez es más frecuente la utilización de apoyos tecnológicos para la investigación (estos tienen funciones acotadas para recoger datos

cada vez más precisos, ordenarlos y procesarlos y muchas veces) no cuentan con la capacidad de hacer operaciones analíticas y de interpretación por lo que la intervención del analista continúa siendo indispensable (Bunge, 1997; Canales y Lerner, 2003).

De esta forma se generan los datos que el sujeto adquiere del objeto, se sistematizan y se analizan. La aproximación al tema o realidad de interés depende de la inmediatez con la que obtengan los datos y de las herramientas de aproximación que se usen (directas o indirectas). En este sentido, en cualquier trabajo de investigación es inevitable la inclusión de la subjetividad, incluso es indispensable para mostrar el interés de cualquier recorte de la realidad, debido a que es necesario partir de un interés personal y colectivo (comunidades científicas) (González, 2018). Para ser más concreto, en el ámbito de la demografía, implica una reflexión sobre las ventajas y desventajas de la vinculación de imágenes satelitales con los datos sociodemográficos disponibles.

4.6 Conclusiones

De este capítulo se desprenden más interrogantes que conclusiones, una de ellas es si debiésemos aceptar la relación de desigualdad creada por una asimetría o monopolio en la producción de datos en México. De ser afirmativa la respuesta, implicaría que dichas desigualdades trasciendan a la esfera de producción de conocimiento nuevo y necesario para la realidad que hoy se vive. Otra incógnita que se desprende de este capítulo es que la gestión de los datos personales en la era de los macrodatos y su relación con el uso demográfico pone en la mesa la diferencia entre el consentimiento y la privacidad, además de colocarnos en la encrucijada entre datos voluntarios, datos observados y datos inferidos, lo cual es una disección que quedaría pendiente por analizar.

Una inquietud más a manera de conclusión es que ante el arribo de la datificación, se debe estar alerta para discernir una situación “real” y una situación real generada a partir de los macrodatos. Hoy en día parece que nuestra vida cotidiana y acciones más superfluas en ella, pueden convertirse en un registro más de una base de datos, y ante la oferta de esta información, se presente una demanda. Así que es conveniente cuestionarse si desde la ciencia debemos abandonar la obsesión por la causalidad de las cosas y entregarnos a meras correlaciones sin conocer el sentido profundo de los hechos sociales.

Es cierto que la amplia gama de información es cada vez mayor, y que la tecnología nos permite allegarnos a fuentes de mayor calidad, particularmente me refiero al uso de los mapas como un instrumento de gran alcance para el análisis demográfico-espacial. No obstante, se debe privilegiar la selección y el diseño lo más cercano posible a la realidad, lo cual significa que los componentes éticos deben regir dicha selección.

La mayor parte de las instituciones creadoras y proveedoras de datos, han sido creadas bajo la presunción de que las decisiones subjetivas se dejan de lado, y que, por lo tanto, ofrecen información exacta y de naturaleza causal, sin embargo, se observó que no es así, y que la situación cambia cuando los datos tienen un gran volumen, y deben ser procesados por paquetes estadísticos o modelos sofisticados, en estos casos, las decisiones de cómo clasificarlos y ordenarlos no las toman del todo los humanos, sino que comparten esta responsabilidad con dichas herramientas computacionales y estadísticas.

Finalmente, a partir de todo el contenido expuesto, una gran conclusión es que el uso de los macrodatos en la demografía cambia nuestra definición y gestión de conocimiento, ya que con la asistencia automatizada de paquetes estadísticos para el manejo de esta gran variedad de datos, cambia nuestra definición de conocimiento, de cómo hacer ciencia, de cómo investigar y pone en el centro de interés diferentes fundamentos relacionados con decisiones humanas, decisiones técnicas y un híbrido entre ambas.

Capítulo 5

Metodología para estimar el total de población en áreas pequeñas

5.1 Introducción

¿Es posible saber el número real de personas que existen en un momento y lugar determinado? A pesar de todos los esfuerzos metodológicos que se hagan para conocer ese dato, resulta imposible saberlo debido al menos a dos factores: 1) no hay forma de tener un registro infalible porque la población está en constante movimiento, crecimiento y cambio, y 2) existen omisiones e imperfecciones en los instrumentos de captación y rezagos en el procesamiento de la información, fenómeno que ocurre incluso en las fuentes más confiables. En otras palabras, cuando miramos los datos de un sitio estamos viendo una fotografía imperfecta de cómo era el pasado, porque no todos los miembros de una población salen en la foto y no todos fueron tomados en cuenta en los levantamientos de información. Es decir, no se hicieron estimaciones de población en áreas pequeñas y se omitió mucha información.

Sin embargo, pese a estos sesgos y a que no es posible llegar al *dato real* sobre una población, el objetivo de los diferentes métodos demográficos es generar estimaciones que permitan hacer inferencias sobre el comportamiento de una determinada población de estudio. Estas estimaciones constituyen aproximaciones a los datos reales y se construyen con base en la información disponible. Su objetivo no es llegar al dato real, porque eso resulta imposible, sino generar un valor que sea lo más confiable posible y con el cual se pueda operar. Por lo tanto, cuando se generan técnicas que minimizan el error en las estimaciones se considera que existe un avance metodológico (Lomax y Smith, 2017; Molina, 2019; Paredes y Silva, 2017; Schmertmann *et al.*, 2013).

Los resultados que se obtienen con estos métodos tradicionales toman la forma de promedios, razones, totales o relaciones entre dos o más características como en la diferencia de medias. En un escenario ideal, estos resultados y relaciones reflejan el comportamiento de la población de estudio, sin embargo, esto depende de la metodología utilizada y, sobre todo, de la calidad de los datos.

En el caso de las estimaciones que se hacen en Demografía existe una peculiaridad, estas se producen tanto para el presente como para el pasado, lo que significa que podrían sustituir valores que se habían considerado como reales. En el mejor escenario, cuando esto ocurre, la diferencia entre el valor con el que se contaba y el que se aportó es mínima, pero existe una gran posibilidad de que los cambios sean significativos.

Las técnicas que existen para llevar a cabo estimaciones en áreas pequeñas, consideradas como unidades administrativas que están por debajo del nivel nacional y estatal, se ubican en alguna de las siguientes categorías: extrapolación, razón, sintomático, regresión, componente, basado en muestras y otros métodos (Swanson, 2012). No obstante, desde la esfera estadística, Rao y Molina (2015) condensan la metodología en dos grandes grupos: métodos directos e indirectos. En esta sección se pretende hacer un recuento histórico y representativo de cómo han evolucionado estas metodologías y las diferencias entre ellas.

5.2 Enfoque demográfico

Para calcular el total de población en áreas pequeñas, existen varias metodologías que se pueden agrupar según la disciplina que intervenga: 1) las de enfoque demográfico, 2) las de enfoque demográfico-estadístico y 3) las de enfoque estadístico. Los métodos con enfoque demográfico, toman como referencia a la ecuación compensadora. Si bien es cierto que es una ecuación sencilla que relaciona crecimiento natural y social, su capacidad para predecir el comportamiento de la población se mantiene. Por ejemplo, Davis (1995) ofrece una guía para llevar a cabo estimaciones de población en áreas pequeñas a partir de los componentes básicos de la ecuación compensadora. Su idea consiste en hacer una interpolación a partir de los datos disponibles y tratar de descubrir si el comportamiento de la población adopta una forma lineal, exponencial o polinómica. Es decir, quiere ver hacia dónde se mueven los números, qué nos indican y si es posible predecir el comportamiento de una población.

Para aventurar esta parametrización y poder calcular los valores desconocidos que facilitan su aplicación es importante tomar en cuenta los supuestos: 1) la población tiene un comportamiento homogéneo y 2) las áreas pequeñas son una proporción del total poblacional. Aunque aparentemente y en términos matemáticos, este supuesto es válido porque cumple con la idea de

que el total es equivalente a la suma de sus partes, desde el punto de vista demográfico este supuesto es falible. Ya en el capítulo anterior se puso el ejemplo de Caldwell *et al.*, (1980) quienes mostraron que una población que está restringida a una zona pequeña puede adoptar comportamientos totalmente diferentes al resto de la población, debidos a su aislamiento y también a que la dinámica demográfica se comporta de distinta manera.

Otro de los problemas que se presentan al aplicar la ecuación compensadora tiene que ver con uno de sus componentes: la migración. Este fenómeno suele ocurrir de una manera muy particular, como sostienen ambos autores, Caldwell *et al.*, (1980) y Davis (1995), quienes aseguran que modelar el comportamiento migratorio de las personas es muy complicado. Esto se debe a que en estos movimientos no se presentan patrones regulares y las fuentes de información tradicionales carecen del grado de precisión necesario para capturar lo que ocurre y alimentar los distintos modelos para generar datos. La migración no es un indicador en una página sino un fenómeno que depende de muchas variantes, se presenta de manera irregular y existen momentos donde se intensifica, por ejemplo, cuando ocurren conflictos sociales o desastres medioambientales.

Por ejemplo, en el caso de Alexander *et al.*, (2019), los investigadores analizan la migración de puertorriqueños como consecuencia de los desastres ocasionados por el huracán María. Su análisis utilizó como fuente la información de Facebook y lograron estimar el porcentaje de personas en los nuevos asentamientos. Este uso de datos alternativos para dar cuenta de un fenómeno migratorio se ha convertido en una tendencia; porque los datos ya estaban ahí y se iban actualizando solos en tiempo real y directamente por cada uno de los actores.

También Deville *et al.*, (1995) emplearon datos alternativos e hicieron un mapeo de cómo se distribuye la población migrante en Francia. Para hacerlo, los autores analizaron datos telefónicos y pusieron su atención en la distribución de las llamadas de acuerdo con su geolocalización, de este modo estimaron la densidad de población a nivel subnacional. Además, llegaron a la misma conclusión de Davis (1995), de que estudiar la migración en áreas pequeñas resulta complicado y es necesario usar fuentes de información alternativas que no tienen una naturaleza demográfica.

La identificación de que había un problema en la calidad de los datos y que esta era más deficiente en la medida en que se reducía el área de estudio, se puso sobre la mesa a partir de los años setenta. Uno de los estudios pioneros en analizar este fenómeno fue el realizado por Ericksen (1973), él utilizó los datos del censo de 1960 en Estados Unidos de América (EUA) en conjunto

con registros administrativos e incluyó otras variables que, sin estar presentes en los censos, podían reflejar la densidad de población. Lo novedoso de su estudio en aquel momento fue la intención de modelar el comportamiento de la población a partir de variables sintomáticas, y de paso quitar la dependencia de la enumeración censal que ocurre, en el mejor de los casos, cada 10 años. Gracias a ello se incentivó la generación y uso de los registros administrativos para complementar la información deficiente (Martin, 2006). De este modo, los censos de población no logran tener un recuento preciso de la población, sin embargo, tal y como lo reconoce Martin (2006), lo recomendable es que los países mantengan esos eventos, ya que brindan información esencial, además de que su aporte también está en la esfera geográfica y permite hacer un mapeo de los asentamientos urbanos. Por lo tanto, desde la década de los años setenta ya se había notado que para estudiar las áreas pequeñas se requiere de distintas fuentes de información.

5.3 Enfoque demográfico-estadístico

A mediados del siglo XX, Bogue (1950) propuso complementar la información censal con datos del registro de nacimientos o listas de votantes. De esta conjunción de disciplinas nació uno de los primeros métodos demográficos estadísticos para estimar poblaciones en áreas pequeñas. La intención del investigador era retomar las técnicas de extrapolación que se utilizaban en otras áreas, aplicarlas al ámbito demográfico para predecir comportamientos futuros y complementar la información faltante o deficiente.

El empleo de información complementaria para estimar el total de población abrió la posibilidad de dar seguimiento a lo largo del tiempo. Schmitt y Crosetti (1954) fueron pioneros en usar este tipo de modelos en la esfera demográfica bajo la lógica de que es factible modelar los fenómenos sociales mediante técnicas econométricas. En general, todos estos modelos toman como variable dependiente tomaron alguna característica poblacional como la densidad o el total de personas, y como variables independientes usaban las que pudieran dar cuenta de los cambios. A partir de ese momento se empezaron a implementar modelos cada vez más sofisticados, aunque la idea básica de tener una relación lineal se mantuvo:

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \dots + \beta_n X_n$$

Desde esos años hasta ahora, y con el avance de la tecnología, lo que ha cambiado en la ecuación son las variables predictoras, ya que se ha vuelto más fácil introducir nuevos indicadores

que anteriormente no estaban disponibles. En ese sentido, Cai y Tippett (2015), sumaron a esta idea de modelar a la población e introdujeron lo que denominan como método de correlación de razón. Su idea consistía en utilizar como variable próxima el número de viviendas que existen en un lugar determinado y considerarla como la variable dependiente. De esta forma, el modelo se complementa con otras dimensiones relacionadas con la ocupación habitacional como el padrón vehicular o el número de menores que asisten a las escuelas.

En diversos estudios es común asumir que la densidad de población se modela de manera lineal a partir de un conjunto de covariables. Sin embargo, en el caso mencionado los autores introducen una variación a la manera en que se ingresan las variables al modelo, ya que todas entran como razones. Para ello construyen los indicadores en dos niveles: primero obtienen el dato de la localidad en proporción al estatal; luego repiten este proceso, pero ahora introduciendo una variante temporal y comparan la diferencia que hay entre la información que se obtiene directamente del censo y los datos estimados en los años subsecuentes. Este tratamiento que les dan a las covariables es uno de los puntos a favor de este método, puesto que incorporan la dimensión temporal y espacial para modelar la ocupación habitacional en las diferentes delimitaciones geográficas.

Hauer *et al.*, (2015) también estiman la población que existe en zonas particulares, su argumento para justificar el análisis se sustenta en el hecho de que las proyecciones de áreas pequeñas que se obtienen a partir de los datos nacionales son erróneas. Ellos toman en cuenta que puede haber un incremento o decremento sustancial del número de individuos, así como cambios significativos en la composición por sexo y edad; según argumentan “es bien sabido que las metodologías de proyección para unidades de áreas pequeñas (es decir, sub condados) tienden a ser menos robustas que las metodologías de proyección con unidades de área más grandes (traducción propia)” (Hauer *et al.*, 2015:46). Matemáticamente, su idea se formaliza de la siguiente manera:

$$P = H * PPHU + GQ$$

Donde P es el total de población, H es el número de viviendas habitacionales, $PPHU$ es la proporción de personas que existen en cada una de ellas y el término GQ representa a los lugares en donde se presume que viven personas. De cierta manera, este último término funciona como un parámetro de ajuste para hacer equiparable sus estimaciones con los datos censales, los cuales

toman esa referencia para validar sus resultados. Los investigadores optan por un enfoque en donde se retoman las estimaciones históricas disponibles para generar una estimación del total de viviendas (H), y esto lo hacen de la siguiente forma:

$$\hat{H}_{ij}^t = \left(\frac{C_j^t}{\sum_{1939}^{t=1} H_j^t} \right) * \sum_{1939}^{t=1} H_j^t$$

Así, el número de viviendas estimado (H_{ij}) está dado por el total de viviendas que se tiene de acuerdo con los datos censales (C) para el grupo de bloques i en el condado j . Una de las ventajas que tiene este método es que retoma una de las preguntas del censo de Estados Unidos que es el “año de construcción de la estructura”. Murdock y Hoque (1995) continuaron en la idea de contabilizar el número de viviendas bajo el supuesto de que esta es una variable cercana al tamaño real de la población. Este método fue propuesto en los años ochenta y al momento en que escribieron el artículo había sido aplicado con éxito en varios estados de la unión americana (Smith, 1986).

Schmertmann y Gonzaga (2018) realizaron, también, otro tipo de estudios para analizar poblaciones pequeñas. Dejaron de lado la densidad de población e intentaron modelar dos fenómenos de vital importancia para la Demografía: la mortalidad y la esperanza de vida. Para su empresa, partieron del hecho de que el registro civil está incompleto por lo que una proporción importante de las muertes no registran en las estadísticas oficiales. En consecuencia, existe una alta variabilidad de las muertes que se resume de la siguiente manera: entre más pequeñas sean las zonas de interés, mayor será la imprecisión de las estimaciones.

Para desarrollar el modelo, los autores partieron del supuesto que la mortalidad real (Dx) es mayor a la registrada (Rx) y cuyo registro es independiente de la edad de las personas. Este supuesto es válido cuando se tienen deficiencias institucionales que omiten el registro de los fenómenos sin importar a quién le ocurre. Para llevar a cabo el modelo suponen que los comportamientos se pueden parametrizar y tienen un comportamiento Poisson, lo que significa que, los eventos se producen en un intervalo de tiempo de forma aleatoria, pero con algunas restricciones que tienen que ver con el hecho de que son procesos dicotómicos (la persona muere o no). Al asumir una parametrización de este estilo se obtienen ventajas, por ejemplo, se conocen algunas características de cómo se comportan las variables, así como la media y desviación estándar. Optar por un enfoque

bayesiano, en lugar de uno frecuentista permite librar el escaño de que este último tiene un pobre desempeño cuando los datos son deficientes. De esta manera, el modelo queda definido de la siguiente manera en donde λ^* representa el calendario constante que incorpora los patrones básicos de mortalidad; mientras que la segunda parte contempla una parte lineal compuesta por segmentos de línea recta entre edades designadas que representa las diferencias entre el programa estándar y logarítmico de mortalidad en la población de interés.

$$\lambda = \lambda^* + B\alpha$$

5.4 Enfoque estadístico

Desde el punto de vista estadístico, una de las preocupaciones más importantes es calcular el estimador del parámetro de interés con un nivel de confianza aceptable y con el menor sesgo posible. Dicho parámetro concentra la información de características demográficas y espaciales que tiene una población en particular. En general, las estimaciones se concentran en dos grandes grupos: por un lado, están las directas: aquellas que se obtienen con una sola fuente de información; por el otro lado están las indirectas: aquellas que requieren de fuentes de información adicional.

5.4.1 Estimaciones directas

En la literatura demográfica y estadística se considera que un estimador es directo si se calcula utilizando una sola fuente de información. Gran parte de los institutos nacionales de estadística usan este tipo de estimadores bajo el argumento de que las fuentes de información son de buena calidad (Álvarez, 2001). Un grupo de técnicas que ejemplifican los estimadores directos son los métodos de extrapolación (Duchesne, 1988; Guerrero, 2014). Estos se guían bajo el supuesto de que el comportamiento de la población sigue una forma particular, por lo que es factible saber el comportamiento de valores desconocidos. Parten de la hipótesis de que las poblaciones tienen una cierta estabilidad histórica porque han estado presentes a lo largo del tiempo y es posible recabar información de períodos anteriores. Otro supuesto implícito en esa afirmación es que suponen un comportamiento homogéneo que, como ya se revisó en la sección anterior, no siempre se cumple

debido a la manera en que se captan los datos y a las escalas que se utilizan (Flowerdew, 2011; Wong, 2003).

También existen otros métodos más complejos que tienen un enfoque longitudinal; uno de ellos consiste en los ajustes que se llevan a cabo mediante series de tiempo. Con estos métodos que asumen que la población crece a un ritmo regular y tiende a estabilizarse con el tiempo, por lo que es factible ajustar una curva polinómica que adecúe su comportamiento. Así, es posible analizar el comportamiento de la curva y calcular valores para poblaciones particulares (Lee y Tuljapurkar, 1994; Hakkert, 1985). Este tipo de modelos ofrecen una mejor capacidad predictiva y conservan la idea de una estructura y un comportamiento estocástico que cambian a lo largo del tiempo. Con el desarrollo de las herramientas computacionales su eficiencia ha mejorado.

Existen otros métodos que han tenido mucho auge en los últimos años, por ejemplo los de razón de correlación. En ellos se supone que existe una población (de gran tamaño) que contiene a su vez a otros grupos particulares, lo que significa que la suma de los subgrupos es igual al total de la población (Swanson y Tayman, 2012). Si bien es cierto que este supuesto parece razonable desde el punto de vista matemático, en la demografía enfrenta un problema muy importante: no se trabaja con datos fijos, sino con estimaciones. Esto quiere decir que un valor (total, razón o promedio) oscila en un rango o intervalo de confianza debido a que los instrumentos para coleccionar la información son falibles, tal y como se ha evidenciado en la sección anterior.

Además de este problema, estos métodos tienen otras particularidades que vician la información obtenida; por ejemplo, los mismos autores señalan que el comportamiento que generan no siempre corresponde con casos particulares. Para ilustrar esta situación apuntan que en muchas ocasiones las series históricas muestran que la población crece en todo momento, sin embargo, existen regiones en donde ocurre lo contrario debido a fenómenos relacionados con la migración.

Continuando con esta revisión metodológica, Duchesne (1988) propone usar el método de relación de cohorte para hacer estimaciones de población en áreas pequeñas. El objetivo central de este modelo consiste en hacer una proyección sociodemográfica a partir del comportamiento de una población mayor en la cual se encuentran contenidos los subgrupos a analizar. Este método, al igual que los anteriores, supone que el comportamiento general es igual al de los casos particulares, pero como ya se ha discutido a lo largo de este trabajo, ese supuesto es falible.

Para ilustrar el hecho de que la suma de las partes no siempre es igual al total, se tiene la siguiente situación: supongamos que tenemos tres estimaciones de grupos particulares y una que

corresponde al total de la población, en cada uno de los casos se asume que el valor se encuentra en un intervalo de confianza cuyo coeficiente de variación es igual a 1. Si el valor de los estimadores es el punto medio, tal y como se muestra en el ejemplo, entonces el total de población equivale a la suma de las partes. Sin embargo, si el valor de los estimadores es 5.9, 4.8 y 1.8 se tiene un valor 12.5, el cual queda totalmente fuera del intervalo que se había considerado para el total de la población. Es decir, la suma de las partes no siempre es igual al total de población.

5.4.2 Estimaciones indirectas

En el polo opuesto están los métodos indirectos, los cuales usan varias fuentes de información para llevar a cabo las estimaciones y combinan diversas variables que están relacionadas con el dato que es de interés. Estos últimos han tenido gran auge en América Latina debido al aumento en la periodicidad de las fuentes de información, la mejora en la recolección de los datos y la disponibilidad de medios computacionales de bajo costo (González y Torres, 2012). En estos métodos, las estimaciones se consideran como variables dependientes y están sujetas al comportamiento de las variables independientes o sintomáticas. Estas últimas variables suelen medir de manera indirecta los cambios poblacionales y adoptan la forma de registros administrativos u otras fuentes de información como los censos de población y vivienda.

De acuerdo con Chi y Vos (2011), los métodos indirectos suelen ser poco utilizados por parte de la comunidad académica con el argumento de que no encuentran diferencias significativas entre aplicar este tipo de modelos y utilizar una extrapolación simple. Además del hecho de que en las regresiones, las unidades de análisis y las variables involucradas en el modelo asumen independencia, mientras que en el caso de las poblaciones esto no sucede, tal y como se ha venido mostrando con los ejemplos en los que los asentamientos humanos que están cerca unos de otros generan y reciben influencia mutua.

En este grupo de los estimadores indirectos más empleados para poblaciones pequeñas, desde el punto de vista estadístico, de acuerdo con González y Torres (2012), y Rao y Molina (2015), los estimadores generalizados de regresión (GREG por sus siglas en inglés) están a la vanguardia. Estos estimadores toman como referencia un área d para la cual se quiere conocer algún valor que es de interés que, como se ha mencionado anteriormente, adopta la forma de totales $X_d = \sum_{i=1}^{N_d} x_{di}$ o promedios $\bar{X}_d = N_d^{-1} \sum_{i=1}^{N_d} x_{di}$, de alguna característica particular x . Cuando se

analiza dicha característica (sociodemográfica) en el área d se genera el vector x_{di} , el cual representa los valores que se obtienen de cada uno de los individuos.

En el caso de las encuestas, que son una de las fuentes de información más utilizadas para la investigación demográfica, cada uno de los individuos tiene un cierto peso poblacional, es decir, representa a un número determinado de personas en la realidad. Eso significa que los estimadores adoptan la siguiente forma $\widehat{X}_d = \sum_{i=1}^{N_d} w_{di} x_{di}$ (estimador de Horvitz-Thompson) en donde el ponderador w_{di} se obtiene con base en la ubicación geográfica de las personas y su probabilidad de inserción (π), es decir, la probabilidad que tienen los individuos de ser seleccionados.

Cuando la información es insuficiente para dar cuenta de la característica poblacional x del área d , se recurre a información externa que se obtiene de otras fuentes. De esta suerte se generan p variables auxiliares que sirven para ajustar a la estimación de la característica poblacional que es de interés. Así, los estimadores GREG buscan calibrar a los ponderadores w_{di} que son los que determinan el peso que tiene cada una de las observaciones y generan los valores poblacionales. La nueva estimación \widehat{Y}_d^{GREG} no solo considera los datos de una sola fuente de información, sino que incorpora las p variables auxiliares del área d que sirven para mejorar la estimación de la característica x que es de interés. Su fórmula está dada por:

$$\widehat{Y}_d^{GREG} = \widehat{Y}_d + (\bar{X}_d - \widehat{X}_d)' \widehat{B}_d \quad (1)$$

En esta ecuación \widehat{B}_d es el estimador de mínimos cuadrados ponderados del vector de coeficientes de la regresión lineal y se calcula de la siguiente forma:

$$\widehat{B}_d = \left(\sum_{i \in S_d} w_{di} x_{di} x_{di}' / c_{di} \right)^{-1} \sum_{i \in S_d} w_{di} x_{di} x_{di} / c_{di}$$

Cabe resaltar que esta ecuación cobra sentido en el caso de las encuestas porque se construyen a partir de datos muestrales y dada su construcción, en particular \widehat{X}_d , toma como referencia el peso inicial. Como se había mencionado anteriormente, dicho ponderador es elemento clave para desarrollar la estimación correspondiente y es en donde se lleva a cabo el ajuste del modelo. De esta forma Y_{di} se construye con base en las variables auxiliares x'_{di} que corresponden al área que es de interés:

$$Y_{di} = x'_{di}\beta_d + \epsilon_{di}, \quad i = 1, \dots, N_d$$

De esta forma existen tres situaciones a tener en cuenta: 1) se asume que hay información en todas las áreas d que se están analizando, lo cual puede no corresponder a la realidad, por ejemplo, en el caso de que el área sea muy pequeña, es usual que no se cuente con información; 2) el modelo asume una relación lineal entre las variables, el cual es un supuesto que puede ser difícil de cumplir, sobre todo si se considera que el número de observaciones es pequeño. En todo caso, el modelo supone que existe un error de ajuste ϵ_{di} con esperanza cero y varianza σ_{di}^2 en donde el último término se asume mayor a 0 y constante, lo que representa la posible heteroscedasticidad del modelo. 3) tal y como se discutió en los acápites anteriores, se puede incorporar información geoespacial como parte de las variables auxiliares para mejorar las estimaciones. Una virtud que tienen este tipo de datos es que no dependen de la interacción de las personas cómo se hace en el caso de las encuestas o los censos, sino de una fuente externa: los datos satelitales.

Las técnicas anteriormente descritas tienen un enfoque frecuentista, sin embargo, también se han desarrollado aproximaciones desde otras perspectivas. De acuerdo con Assunção *et al.*, (2005), una forma de abordar el problema de las estimaciones en áreas pequeñas y su alta variabilidad es mediante técnicas empíricas de Bayes (EB). Su objetivo es tomar en cuenta la distancia que existe entre las unidades geográficas y en la medida en que estén más cerca unas de otras, la información suele ser similar. Uno de los primeros que implementó estas técnicas fue Marshall (1991), cuyo razonamiento es que las áreas cercanas suelen tener comportamientos iguales y lo interesante es que, aún y cuando en el texto no se hace explícito, esta idea sigue el principio de la ley de Tobler en donde los elementos que están cercanos tienen mayor influencia entre sí (Waters, 2018). Un aspecto clave en la propuesta de Assunção *et al.*, (2005) es que la información se obtiene a partir de coleccionar datos en las zonas que son de interés y cuando la muestra es grande, entonces las estimaciones locales son más precisas y se minimiza el error, esto se observa matemáticamente en la fórmula:

$$\widetilde{\theta}_a = \widehat{\theta}_a + S_a[E(\theta_a) - \widehat{\theta}_a]$$

En donde la estimación final de alguna característica poblacional $\widetilde{\theta}_a$ en un área específica a depende del factor de ajuste S_a y la diferencia entre $E(\theta_a) - \widehat{\theta}_a$, se vuelve igual a cero en la

medida en que exista una menor variación en las estimaciones y esto se logra si el número de observaciones crece.

También el uso de las imágenes satelitales para estimar el total de población es otra propuesta que parte de la estadística. En el estudio de Ito *et al.*, (2020) adoptan esta perspectiva y para ello consideran técnicas de *aprendizaje profundo* para calcular el número de viviendas en zonas específicas. La clasificación de objetos a partir de datos satelitales no es nuevo, otros autores como Silván-Cárdenas *et al.*, (2010) o Lu (2006) son ejemplos de investigaciones en donde se utilizan fotografías aéreas para clasificar objetos. No obstante, el equipo de Ito (2020) aplicó una metodología de reconocimiento de patrones implementado en una red neuronal con una arquitectura compleja que le permite diferenciar entre viviendas con un solo piso y las que tienen dos o más. Una ventaja que presenta este estudio es que el modelo de redes convolucionales puede ser usado en el futuro para entrenar redes. Esto es lo que se conoce como transferencia de aprendizaje y sirve para tomar las experiencias pasadas e ir mejorando los modelos de manera significativa. Además, estos investigadores encontraron que es necesario llevar un procedimiento de reconocimiento en campo debido a que las fotografías son incapaces de garantizar que se trate viviendas.

En el texto de Robinson *et al.*, (2017), los autores retoman la idea de usar modelos de aprendizaje profundo para hacer estimaciones de población. A diferencia del estudio pasado, en este caso los investigadores no se plantean identificar a las viviendas, sino generar cuadrículas de alta resolución a partir de los datos satelitales en donde sea plausible identificar las zonas con una alta densidad de población para lo cual establecen 13 categorías. De esta forma, cuando ingresan nueva información de una zona diferente, por ejemplo, una cuadrícula nueva, se tiene la capacidad de hacer una clasificación y así predecir la densidad de población que existe.

En un intento de emular el funcionamiento de las redes neuronales que requieren de una base de entrenamiento con la cual se ajustan las ponderaciones, los investigadores utilizaron los datos del censo del año 2000 para entrenar el modelo. Luego, con los datos del censo de 2010 fue posible hacer una evaluación del desempeño del modelo, hicieron los ajustes necesarios y fueron capaces de modificar los pesos sinápticos para tener un modelo mejor entrenado. Esta idea de usar los datos de censos consecutivos garantiza que el modelo tenga un buen desempeño, además, complementaron su entrenamiento de redes neuronales con los datos de Landsat y siguieron una

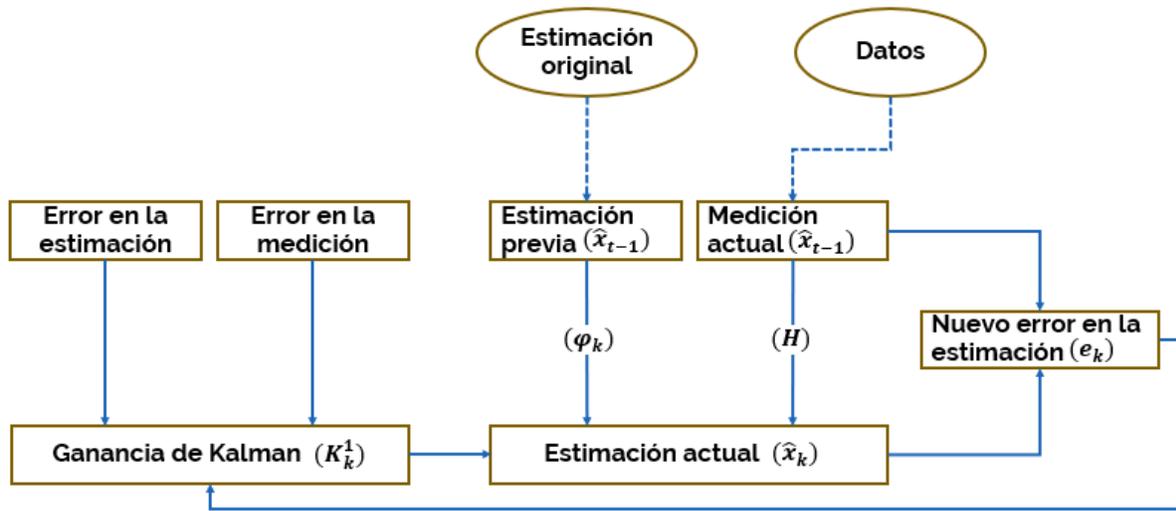
lógica retrospectiva, así, el modelo final tenía una combinación de datos censales e imágenes satelitales.

El estudio de Campos *et al.*, (2020) es otro ejemplo de cómo se pueden emplear las imágenes satelitales de alta resolución para producir estimaciones de población en años no censales y a nivel municipal. El uso del satélite Landsat + ETM está justificado debido a que las imágenes que ofrece son de alta resolución, son gratuitas y tienen una larga periodicidad que permite obtener imágenes de años anteriores. En este estudio la densidad de población está asociada con las características que se ofrecen en las diversas imágenes satelitales y para ello se emplean modelos de regresión que incorpora la información de los años 2000 y 2010. La idea central es que la variable dependiente es la población, mientras que las covariables son la reflectancia de las bandas, así como otras variables que miden el área urbana o rural.

5.5 Propuesta de análisis: Filtro de Kalman

El filtro de Kalman es un algoritmo matemático que sirve para modelar fenómenos que cambian con el tiempo. Una de sus virtudes es que combina los valores observados y las estimaciones que genera el propio modelo, con ello es factible analizar sus diferencias y tomar esta información para generar un nuevo valor. En la medida en que las iteraciones avanzan el error se reduce y con ello se mejora la predicción del modelo, el siguiente diagrama se muestra cómo opera este algoritmo y la combinación de las observaciones y las estimaciones.

Diagrama 2 Diagrama del filtro de Kalman



Fuente: elaboración propia con base en Spiegel *et al.* (2013).

Un estimador \hat{x}_k es un valor aproximado al dato real que se calcula con la información disponible y con relaciones matemáticas, en el escenario ideal debe ser insesgado lo que significa que su promedio debe ser igual al parámetro real, es decir $E(\hat{x}_k) = x_k$ (Spiegel *et al.*, 2013). Por otro lado, las mediciones $\{z_1, z_2, \dots, z_{t-1}\}$ son observaciones que se hacen a los fenómenos directamente por lo que se considera que son datos reales en la esfera matemática. Sin embargo, de acuerdo con Canales (2001), en el área demográfica nunca es posible conocer el valor real de dichas mediciones, un ejemplo de ello son las proyecciones de población (estimaciones), las cuales se asumen como válidas hasta que se levanta el censo de población y se obtienen los datos confiables. Por este motivo, es mejor usar el enfoque de Rao y Molina (2015) y asumir que se trata de estimaciones directas, las cuales son definidas como aproximaciones que utilizan una sola fuente de información. De esta forma, el Filtro de Kalman tiene como objetivo encontrar una estimación \hat{x}_k del vector x_k a partir de un conjunto de mediciones $\{z_1, z_2, \dots, z_{t-1}\}$ y tratar de minimizar el error.

El Filtro de Kalman parte de la idea que la medición de los fenómenos en un momento determinado en el tiempo z_k son combinaciones lineales de un valor estimado \hat{x}_k más un cierto ajuste v_t , cuya forma está dada por la expresión:

$$(1)$$

$$z_t = H\hat{x}_k + v_t$$

Esto quiere decir que es factible conocer la medición real en un momento específico en el tiempo z_t si se cuenta con el valor de la estimación, el ruido o ajuste, así como la sensibilidad de la medición que se denota con la letra H . Además, de esta igualdad se desprende otra propiedad importante, dado que se trata de un valor que se mueve con el tiempo, entonces al contar un valor en un tiempo k determinado, se infiere el valor siguiente:

$$\hat{x}_{k+1} = \varphi_k \hat{x}_k + w_t \quad (2)$$

A partir de las propiedades anteriormente mencionadas, se deduce que existe una relación temporal entre los valores estimados a priori \hat{x}_{t-1} , a posteriori \hat{x}_{t+1} y las mediciones z_k que se puede modelar con la siguiente relación (Grewal y Andrews, 2014):

$$\hat{x}_{t+1} = K_k^1 \hat{x}_{t-1} + \bar{K}_k^1 z_k \quad (3)$$

Entonces, el objetivo es encontrar las matrices K_k^1 y \bar{K}_k^1 con las cuales se logre satisfacer la igualdad, las cuales representan los valores numéricos que permiten hacer las transiciones de los diferentes estados en el tiempo. De acuerdo con Grewal y Andrews (2014), esto significa que se debe de cumplir con el principio de ortogonalidad que se representa con la siguiente relación:

$$E([x_k - \hat{x}_{t-1}]z_k^T) = 0 \quad (4)$$

Si se sustituyen las ecuaciones 2 y 3 en la ecuación 4, entonces se tiene lo siguiente:

$$E([\varphi_{k-1}x_{k-1} + w_t - K_k^1 \hat{x}_{t-1} - \bar{K}_k^1 z_k]z_k^T) = 0 \quad (5)$$

Otro término que se conoce del planteamiento inicial es z_k , por lo que se puede hacer la sustitución de la ecuación 1 en esta última y se tiene:

$$\begin{aligned} E([\varphi_{k-1}x_{k-1} + w_t - K_k^1 \hat{x}_{t-1} - \bar{K}_k^1 (H\hat{x}_k + v_t)]z_k^T) &= 0 \\ E([\varphi_{k-1}x_{k-1} + w_t - K_k^1 \hat{x}_{t-1} - \bar{K}_k^1 H\hat{x}_k - \bar{K}_k^1 v_t)]z_k^T &= 0 \end{aligned} \quad (6)$$

Dado que una de las propiedades de la esperanza matemática es que se puede obtener el valor individual de cada término. Además, se había supuesto que $E(w_t) = 0$ con lo que $E(w_t z_k^T) = 0$, la expresión anterior se reduce a (Grewal y Andrews, 2014):

$$E([\varphi_{k-1} x_{k-1} - K_k^1 \hat{x}_{t-1} - \bar{K}_k^1 H x_k - \bar{K}_k v_k] z_k^T) = 0 \quad (6)$$

Una vez que se ha hecho esta reducción a partir de las propiedades propuestas y que se sustituyeron los valores conocidos, conviene desglosar la ecuación anterior a partir de reconocer que los argumentos H , φ_{k-1} , K_k^1 , y \bar{K}_k^1 son valores numéricos o constantes y no variables aleatorias, por lo cual, al aplicar las propiedades de la esperanza, la expresión anterior se reduce a (Grewal y Andrews, 2014):

$$\Phi_{k-1} E(x_{k-1} z_i^T) - K_k^1 E(\hat{x}_{k-1} z_i^T) - \bar{K}_k H_k \Phi_{k-1} E(x_{k-1} z_i^T) - \bar{K}_k E(v_k z_i^T) = 0 \quad (7)$$

Al igual que para llegar a la ecuación 6, se había supuesto que $E(v_t) = 0$ con lo que $E(v_t z_k^T) = 0$, por lo que la ecuación anterior se reduce a:

$$\begin{aligned} \Phi_{k-1} E(x_{k-1} z_i^T) - K_k^1 E(\hat{x}_{k-1} z_i^T) - \bar{K}_k H_k \Phi_{k-1} E(x_{k-1} z_i^T) &= 0 \\ E([x_k - \bar{K}_k H_k x_k - K_k^1 x_k] - K_k^1 (\hat{x}_{k-1} - x_k) z_i^T) &= 0, \\ [I - K_k^1 - \bar{K}_k H_k] E(x_{k-1} z_i^T) &= 0. \end{aligned}$$

De acuerdo con (Grewal y Andrews, 2014), esta última igualdad se satisface si alguno de los dos términos que están multiplicando son iguales a 0, en particular, si se fija la atención en el primer término $I - K_k^1 - \bar{K}_k H_k$, entonces se tiene que para alguna x_k se tiene que $K_k^1 = I - \bar{K}_k H_k$. Esto significa que la elección de K_k^1 juega un papel determinante para lograr la igualdad.

Ahora bien, el hecho de tener una estimación a priori \tilde{x}_{k-1} y posterior \tilde{x}_{k+1} , así como una medición z_k en el mismo tiempo da lugar a que se puedan calcular los errores. Esto es:

$$e_k = H_k \hat{x}_{k-1} - Z_k.$$

Debido a que \hat{x}_k depende linealmente de x_k quien, a su vez, está relacionado con z_k , entonces se retoma la ecuación 6 y se sustituye a z_k con lo cual se obtiene:

$$E([\Phi_{k-1}x_{k-1} + w_{k-1} - K_{k-1}^1 - \bar{K}_k z_k][H_k \hat{x}_{k-1} - z_k]^T) = 0$$

No obstante, dado que $E(w_k z_k^T) = E(w_k \hat{x}_{k+1}^T) = 0$ entonces la expresión se reduce a:

$$E([\Phi_{k-1}x_{k-1} - K_k^1 \hat{x}_{k-1} - \bar{K}_k z_k][H_k \hat{x}_{k-1} - z_k]^T) = 0.$$

Y ahora, se sustituye K_k^1 , z_k y \hat{x}_{k+1} , además de la propiedad antes mencionada que $E(w_k z_k^T) = 0$ para obtener:

$$\begin{aligned} 0 &= E([\Phi_{k-1}x_{k-1} - \hat{x}_{k-1} + \bar{K}_k H_k \hat{x}_{k-1} - \bar{K}_k H_k x_k - \bar{K}_k v_k][H_k \hat{x}_{k-1} - H_k x_k - v_k]^T) \\ &= E([(x_k - \hat{x}_{k-1}) - \bar{K}_k H_k (x_k - \hat{x}_{k-1}) - \bar{K}_k v_k][H_k \hat{x}_{k-1} - H_k x_k - v_k]^T) \\ &= E([(-\hat{x}_{k-1} + \bar{K}_k H_k \hat{x}_{k-1} - \bar{K}_k v_k)[H_k \hat{x}_{k-1} - H_k x_k - v_k]^T). \end{aligned}$$

Por definición, la covarianza a priori es $P_{k-1} = E(\tilde{x}_{k-1} \tilde{x}_{k-1}^T)$ y satisface la ecuación:

$$[I - \bar{K}_k H_k] P_{k-1} H_k^T - \bar{K}_k R_k = 0$$

Por lo tanto, la ganancia de Kalman se puede expresar como:

$$\bar{K}_k = P_{k-1} H_k^T [H_k P_{k-1} H_k^T + R_k]^{-1}$$

De acuerdo con Grewal y Andrews (2014) esta es la solución para la ganancia de Kalman y está en función de la covarianza de las estimaciones a priori. Un hecho interesante es que se puede obtener una fórmula similar para los valores a posteriori, los cuales están definidos como:

$$P_{k+1} = E([\tilde{x}_{k+1} \tilde{x}_{k+1}^T])$$

Al sustituir las ecuaciones anteriores y restar x_k de ambos lados de la ecuación y uniendo la ecuación $\tilde{x}_{k+1} = (I - \bar{K}_k H_k) \hat{x}_{k-1} + \bar{K}_k z_k$ con la expresión $\tilde{x}_{k+1} = \hat{x}_{k-1} + \bar{K}_k [Z_k - H_k \hat{x}_{k-1}]$ se tiene que:

$$\begin{aligned}
\tilde{x}_{k+1} - x_k &= \hat{x}_{k-1} + \bar{K}_k H_k x_k + \bar{K}_k v_k - \bar{K}_k H_k \hat{x}_{k-1} - x_k, \\
\tilde{x}_{k+1} &= \tilde{x}_{k-1} + \bar{K}_k H_k \tilde{x}_{k-1} + \bar{K}_k v_k, \\
\tilde{x}_{k+1} &= (I - \bar{K}_k H_k) \tilde{x}_{k-1} + \bar{K}_k v_k,
\end{aligned}$$

Por lo tanto, se tiene que:

$$\begin{aligned}
P_{k+1} &= E([I - \bar{K}_k H_k] \tilde{x}_{k-1} \tilde{x}_{k-1}^T [I - \bar{K}_k H_k]^T + \bar{K}_k v_k v_k^T \bar{K}_k^T). \\
&= (I - \bar{K}_k H_k) P_{k-1} (I - \bar{K}_k H_k)^T + \bar{K}_k R_k \bar{K}_k^T.
\end{aligned}$$

A esta última expresión se denomina “Joseph” de la covarianza y fue propuesta originalmente por Bucy y Joseph. De ella se puede sustituir k_k y arreglar el resultado como sigue:

$$\begin{aligned}
P_{k+1} &= P_{k-1} - \bar{K}_k H_k P_{k-1} - P_{k-1} H_k^T \bar{K}_k^T + \bar{K}_k H_k P_{k-1} H_k^T \bar{K}_k^T + \bar{K}_k R_k \bar{K}_k^T \\
&= (I - \bar{K}_k H_k) P_{k-1} - P_{k-1} H_k^T \bar{K}_k^T + \bar{K}_k (\bar{K}_k H_k H_k^T + R_k) \bar{K}_k^T \\
&= (I - \bar{K}_k H_k) P_{k-1}
\end{aligned}$$

De esta forma se tiene que $P_{k+1} = (I - \bar{K}_k H_k) P_{k-1}$, el cual se toma como referencia para implementar el filtro de Kalman en los diferentes paquetes estadísticos.

En resumen, las ecuaciones más reconocidas del Filtro de Kalman son las siguientes:

$$\begin{aligned}
z_t &= c x_t + v_t \\
x_{t+1} &= H x_t + w_t
\end{aligned}$$

La primera se conoce como ecuación de medición y significa que se puede llegar a conocer una estimación en el tiempo si se conoce el valor anterior. Esto es equivalente a un modelo de series de tiempo en donde el valor futuro se estima a partir de los valores que lo preceden. La segunda ecuación significa que cualquier valor que sea factible de medir se puede expresar como una combinación lineal del valor estimado y un ruido o error de ajuste que se denota con la letra v_t . En palabras de Rao y Molina (2015: 244) el Filtro de Kalman “permite suavizar las estimaciones pasadas a medida que se dispone de nuevos datos”.

Como se hacía notar en los capítulos anteriores, en la demografía siempre se trabaja con valores estimados y no con los datos reales. Aun y cuando la literatura señala que los censos de población y vivienda son los instrumentos que proporcionan el valor real de la población, en el

caso mexicano solamente se tiene acceso a la muestra censal, por lo que significa que las estimaciones tienen un cierto margen de error (Cochran, 2007).

En los capítulos anteriores se mencionó que en la literatura estadística existen dos tipos de estimadores, los que se denominan directos y los indirectos, el primer grupo tiene la característica de que son insesgados, pero pueden llegar a tener problemas con la varianza cuando el tamaño muestra es pequeño. El segundo grupo lo constituyen los estimadores indirectos que, a diferencia de los directos, su varianza es pequeña, pero existen casos en donde pueden llegar a estar sesgados.

Existe un tercer grupo de estimadores los cuales se denominan compuestos y de acuerdo con Molina (2019: 40), su objetivo es “disminuir la varianza de los estimadores directos a cambio de una porción del sesgo de los sintéticos”, esto significa que se busca hacer una compensación en donde se aumente la eficiencia que tienen los estimadores directos y se reduzca el sesgo de los estimadores indirectos.

$$\hat{Y}_d^C = \Phi_d \hat{x}_d^{DIR} + (1 - \Phi_d) \hat{x}_d^{SYN}$$

En donde la ponderación φ_d se establece para minimizar el error cuadrático medio, oscila entre 0 y 1, además, en el caso en el que su valor sea igual a 0.5 se trata de un promedio simple. Esta idea es la que retoman García y Ordorica (2012) y Ordorica *et al.* (2019), quienes plantean que una forma de obtener estimadores demográficos combinando los datos poblacionales con el espacio y el tiempo es mediante la siguiente relación:

$$\hat{x}_{k_{est}} = r_k(x'_{k,\varphi}) + (1 - r_k)(x''_{k,s})$$

Para el ejercicio que desarrollaron estos autores, el valor de la ponderación r_k tiene un comportamiento dinámico que se obtiene a partir de la implementación del filtro de kalman, el cual se actualiza ya sea con las estimaciones o con las mediciones. Además, es importante mencionar que el estimador $x'_{k,\varphi}$, el cual bajo la lógica de Rao y Molina (2015) sería el directo, se construye a partir de los datos poblacionales, mientras que el estimador $x''_{k,s}$ se construye con los datos que se obtienen de las imágenes satelitales. Desde el punto de vista teórico, esta combinación de población, tiempo y espacio son las coordenadas demográficas a las cuales hace referencia Almanza (1988), Canales y Lerner (2003).

5.6 Conclusiones

La pregunta que sirvió de guía en este capítulo es cuestionarse sobre la posibilidad de conocer el número real de personas que existen en un momento en lugar determinado. Esta pregunta engloba tanto en la parte sociodemográfica que es la encargada de dar cuenta del número de individuos que existen y de la parte geográfica que es la que enfatiza el hecho de que estas personas se ubican en un punto en la superficie terrestre. Para responder a esta pregunta, en la primera parte de este capítulo se llevó a cabo un recuento de los principales métodos y cuáles son los supuestos que les dan origen, se pudo llevar a cabo una comparación entre las bondades y limitaciones que tienen estos modelos.

A lo largo de esta discusión se enfatizó en la necesidad de que los datos retomarán tanto la perspectiva sociodemográfica como espacial y esto se logra por medio del filtro de Kalman. Este es un algoritmo cuya principal característica es modelar el comportamiento de fenómenos que cambian a lo largo del tiempo y es la referencia para llevar a cabo las estimaciones ya aquí es una combinación de los datos poblacionales, el tiempo y el espacio. Todos estos elementos se combinan en un único estimado que sirve para calcular el volumen total de población y es el referente para llevar a cabo los cálculos de la población. Por este motivo, en la segunda parte de este capítulo se describe con precisión el desarrollo del filtro de Kalman:

$$\hat{x}_{k_{est}} = r_k(x'_{k,\varphi}) + (1 - r_k)(x''_{k,S})$$

Así, el estimador resultante $\hat{x}_{k_{est}}$ es una combinación lineal de los datos poblacionales $x'_{k,\varphi}$ y las imágenes satelitales $x''_{k,S}$. Sin embargo, uno de los retos que subyace en esta propuesta es el hecho de analizar las imágenes satelitales, lo cual es un área inexplorada desde los estudios sociodemográficos, por lo que en el siguiente capítulo se describe como usar estas fuentes de información.

Capítulo 6

Datos y área de estudio

6.1 Introducción

En este capítulo se presentan los datos que se utilizan para llevar a cabo el análisis sociodemográfico de un área pequeña. En la primera parte del capítulo se lleva a cabo una introducción teórica de cómo se construyen las imágenes satelitales, ya que se considera que es un tema inexplorado en la demografía, dado que en el quehacer demográfico no se suele emplear este tipo de información.

Además, se ahonda en el asunto de los datos contenidos en las imágenes satelitales y el acceso al público general que se tiene garantizado desde la década de 1960, cuyos beneficios se han aprovechado en diversas áreas del conocimiento y ha permitido poner a discusión cuál es la mejor forma de captar los datos y cómo se deben de interpretar los resultados. Por este motivo, durante el desarrollo de este capítulo se tratará de dar respuesta a algunas interrogantes: ¿cómo se construyen las imágenes?, ¿cuál es el proceso mediante el cual se interpretan los datos?, ¿cómo se vinculan las imágenes con la información sociodemográfica?, ¿cuáles son los problemas relacionados con el área modificable temporal?, y ¿cómo tomar en cuenta el vínculo que tienen los datos con el tiempo?

También se presentan las fuentes de información socio demográfica que complementan al análisis y que son la Encuesta Intercensal de 2015 y las Proyecciones de Población. La primera es una fuente de información que se generan de manera regular, pero tiene la característica de que tienen un enfoque demográfico, lo que significa que para conectar su información se recurre a preguntar a las personas sobre sus datos. La segunda se trata de un ejercicio de estimación en donde se aplican diversos modelos matemáticos.

Para cerrar el capítulo se invita a una reflexión en torno a los acuerdos tácitos que existen en las comunidades académicas para determinar qué tipo de conocimiento es válido. Se presentan algunas limitaciones que existen en el momento de tratar de vincular los datos sociodemográficos

con la información geoespacial; y se muestran algunas de las ventajas y desventajas (pros y contras) que resultan de este vínculo.

6.2 Datos espaciales

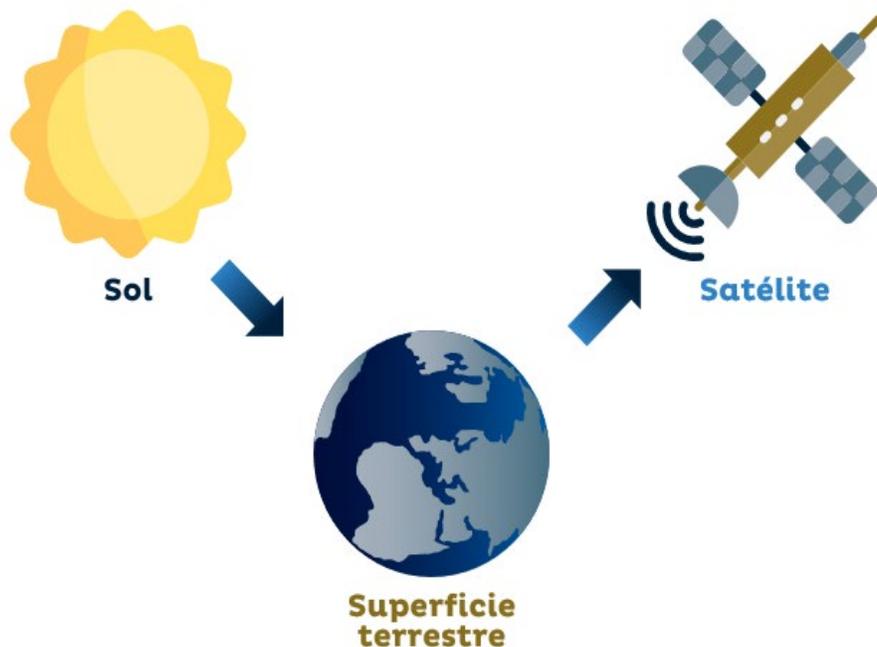
En esta sección se describen los datos que se van a utilizar para el estudio y que provienen de las imágenes satelitales. Sin embargo, dado que se trata de una fuente de información nueva en el ámbito socio demográfico, primero se lleva a cabo una reflexión sobre cómo se generan estos datos, su interpretación y el vínculo que se tiene con la información socio demográfica.

6.2.1 Percepción remota

En este apartado se dará respuesta a la interrogante ¿cómo se construyen las imágenes? Pues bien, la historia del modelado espacio temporal se puede resumir de la siguiente manera. En la década de 1960 se abrió la información satelital. Posteriormente, de los años sesenta hasta los años ochenta tuvo auge la programación orientada a objetos. En la década de 1990 los Marcos conceptuales colocaron el espacio y el tiempo al mismo nivel. A partir de la primera década del siglo XXI se tuvo un gran auge en la modelación teórica de cómo tenían que operar los modelos espacio temporales. Es en esta época cuando comenzó el uso de ontologías para dar forma al conocimiento en las Ciencias geográficas, esto resultó esencial para la semántica y el razonamiento. Finalmente, a partir del 2010 hubo un punto de inflexión en materia del procesamiento de los datos, ya que aumentó de manera exponencial la cantidad de información disponible y de la cual se podía echar mano. Esto generó una evolución de los sistemas de información geográfica (Carré y Hamdani, 2021). La apertura en los datos permitió acceder a una gran cantidad de fotografías de la superficie terrestre, pero generó la necesidad de desarrollar estudios y herramientas que pudieran obtener información mediante el análisis de esos datos. De acuerdo con Lillesand *et al.* (2015), la percepción remota se define como “la ciencia que permite obtener información sobre objetos, áreas o fenómenos mediante el análisis de los datos que son adquiridos a través de dispositivos que no están en contacto físico con ellos” (Lillesand *et al.*, 2015: 1).

Para poder captar la información sobre los fenómenos que ocurren en la superficie terrestre se requieren tres elementos: 1) una fuente de energía que emita radiación; 2) la interacción que existe entre la energía y la materia que se encuentra en la superficie terrestre; 3) un instrumento (sensor) que capte la radiación electromagnética que emiten las diferentes superficies. En el siguiente diagrama se muestra la interacción de estos tres elementos que constituyen un sistema básico de percepción remota.

Diagrama 3 Sistema básico de percepción remota



Fuente: elaboración propia con base en Lillesand *et al.*, (2015)

Uno de los programas más efectivos para cartografiar la superficie terrestre es el denominado Landsat, cuyo inicio en la década de 1960 arrancó con la puesta en órbita del satélite ERTS. A inicio de la década de 1990 este programa tuvo un viraje importante porque se había puesto en duda su continuidad debido a los recortes presupuestarios. En 1992 el Gobierno de los Estados Unidos de América recibió presiones por parte de la comunidad académica y profesional, y se comprometió a tener un financiamiento estable del Landsat. Con esta acción se puso en marcha el lanzamiento de otros satélites con mejoras técnicas. Cada nuevo modelo mejoraba las capacidades de los sensores y la calidad de las imágenes (Masek *et al.*, 2020).

Pese a que el equipo que se requiere y el proceso que se sigue para captar imágenes de la superficie de la Tierra es complejo, este esfuerzo representa solamente la mitad del camino. La otra mitad consiste en analizar, corregir y, eventualmente, interpretar los datos para generar conocimiento, por este motivo conviene reflexionar en torno a ¿qué es una fotografía aérea?, ¿qué tipo de información proporciona?, o ¿qué límites territoriales muestran?

De acuerdo con Câmara *et al.* (2001:2), una imagen es “una función bidimensional, que surge de la respuesta muestreada de una región de la Tierra a una fuente de energía externa (el sol o un rayo de radar) medida por un sensor pasivo o activo, respectivamente”. Las fotografías aéreas captan objetos que se encuentran ubicados en la superficie terrestre y que pueden ser de dos tipos, según Smith y Mark: aquellos que corresponden a objetos que existen solamente en la parte cognitiva de los seres humanos como escuelas, zonas recreativas o teatros; y aquellos que tienen límites independientemente de los actos humanos como las montañas, los lagos o los ríos. Cuando se captura la información mediante fotografías aéreas o imágenes satelitales, todos estos objetos son codificados en píxeles, los cuales se agrupan en forma de malla para representar una imagen digital.

Cada una de las celdas, o píxeles, tiene un valor que codifica algún atributo o algún fenómeno de la superficie terrestre. Estas celdas únicamente pueden tener un valor único, sin embargo, dado que los objetos pueden tener diversas propiedades que son de interés, se requieren varias capas para todos los atributos. Por ejemplo, cuando el objetivo consiste en identificar a una vivienda, entonces se le asigna un valor particular a todo el polígono que ocupa en una banda terminada, cómo podría ser el carro, y para poder tener una cobertura distinta se debe agregar una capa nueva. A este tipo de archivos satelitales de imagen se les denomina datos raster (ver Diagrama 4).

Diagrama 4 Archivos ráster



Fuente: elaboración propia con base en Zaki *et al.* (1995) y Campos *et al.* (2020)

Un archivo ráster (un archivo satelital) se configura a partir de la unión de un grupo de celdas en filas y columnas, en donde cada una de ellas se considera como un pixel y tiene un valor que hace referencia a una información en particular. Estos archivos pueden almacenar información de entidades como montañas o lagos, también de fenómenos como la temperatura de la superficie terrestre e información de estructuras (edificios, construcciones) que han sido creadas por los seres humanos (Campos *et al.*, 2020; Zaki *et al.*, 1995). Este tipo de archivos suelen formar parte de un sistema de información geográfica (SIG), el cual permite tener una visualización de los objetos de la superficie terrestre que son capturados en las fotografías. Este tipo de formatos tiene la capacidad de representar y almacenar superficies mediante una diversidad de formas (puntos o polígonos), además puede hacer superposiciones, lo que posibilita combinar archivos raster (Congalton, 1991).

Las celdas o píxeles del raster representan un área específica de la superficie terrestre. La dimensión de las celdas puede variar, sin embargo, en la medida en la que su tamaño es menor, el nivel de precisión aumenta porque cada pixel representa un área de menor tamaño en la superficie terrestre (Cámara *et al.*, 2001). Sin embargo, al tener una representación de una dimensión menor se requiere mayor capacidad de procesamiento y el espacio requerido para su almacenamiento es mayor.

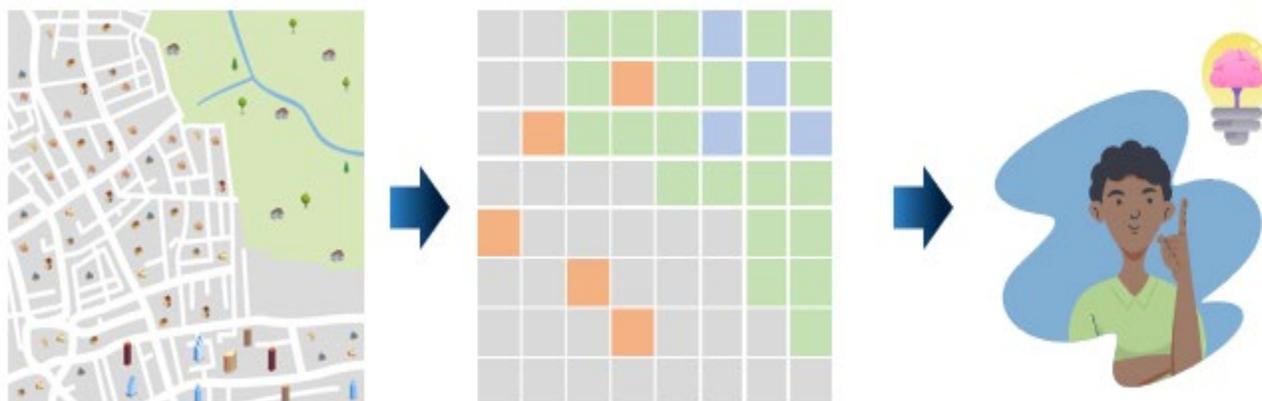
6.2.2 Interpretación

Al inicio del capítulo se planteó la interrogante ¿cuál es el proceso mediante el cual se interpretan los datos?, en este apartado se abordan los asuntos relacionados con este tema. De acuerdo con Lillesand *et al.* (2015) los archivos raster contienen un cúmulo de datos que deben de ser procesados para que se conviertan en información, por lo que se requiere un proceso de interpretación de imágenes que parta de identificar algunos elementos que sean familiares o conocidos, así como la escala y la resolución de la superficie terrestre que se está analizando (Campbell y Wynne, 2011). Si bien es cierto que este proceso dejó de hacerse a mano y comenzó a automatizarse por medio de los programas especializados en análisis espacial, existen algunos elementos que desempeñan un papel fundamental más allá de la forma en cómo se analicen los datos. Olson (1960) y Lillesand *et al.* (2015) son algunos de los autores que pretenden fijar la atención en características como la manera y tamaño de los objetos o los patrones y los colores que se identifican en una imagen.

También existen otros elementos como el grado o la frecuencia con que aparece un cierto tono en la imagen, así como los patrones de sombras que pueden estar presentes debido a la interacción de los objetos fotografiados con otros elementos de la superficie terrestre. Además, un factor clave que constituye el éxito o fracaso de una buena interpretación es el hecho de que se depende de la resolución que tienen las imágenes. Algunos autores como Robinson *et al.*, (2017), Silván *et al.* (2010) o Wijedasa *et al.* (2012) han mostrado que la identificación de los objetos varía de acuerdo con el número de píxeles que se tienen en la imagen.

En la década de 1960, una de las estrategias más utilizadas para tomar fotografías aéreas era la de ubicar objetos cuyo tamaño y forma eran conocidos, con el fin de llevar a cabo una correcta interpretación de los elementos capturados. Hoy en día, con el avance tecnológico, gran parte de este procesamiento se lleva a cabo con la ayuda de diversos programas, sin embargo, tal y como lo reconocen Gutman *et al.* (2013) y Schmertmann *et al.* (2011) se genera la sensación de que las imágenes son perfectas, cuando en realidad existe toda una serie de distorsiones que pueden afectar su calidad (ver Diagrama 5).

Diagrama 5 Interpretación de los archivos raster



Fuente: elaboración propia con base en Câmara *et al.* (2001)

Tras capturar las imágenes y acceder a los archivos raster, igual que con otras fuentes de información, existen algunos problemas al momento de codificar los datos. Uno de los aspectos que señala Câmara *et al.* (2001:2) es que extraer información a partir de las imágenes involucra un componente de incertidumbre, por ejemplo, si existieron complicaciones al momento de captura de una imagen debido a efectos atmosféricos, se dificulta la codificación de la información en píxeles y su posterior análisis. Uno de los casos más recurrentes es que las fotografías aéreas son vulnerables a afectaciones por la presencia de nubes que impiden tener una visión completa de la superficie terrestre (Gutman *et al.*, 2013; Zaki *et al.*, 1995).

En este sentido, la calidad de una imagen está relacionada con sus propiedades intrínsecas como la periodicidad de adquisición y la resolución espectral. Además de ser instrumentos capturados para propósitos específicos y capturan respuestas de diferente tipo de objetos y fenómenos geográficos, por lo que están relacionadas con los procesos de medición, lo cual se hace más evidente cuando se comparan datos adquiridos por diferentes sensores (Câmara *et al.*, 2001; Cárdenas *et al.*, 2010; Masek *et al.*, 2020). Al igual que en la demografía, en la percepción remota, lo que captan las imágenes son objetos que deben ser interpretados por personas con sesgos y preferencias al momento de llevar a cabo un análisis.

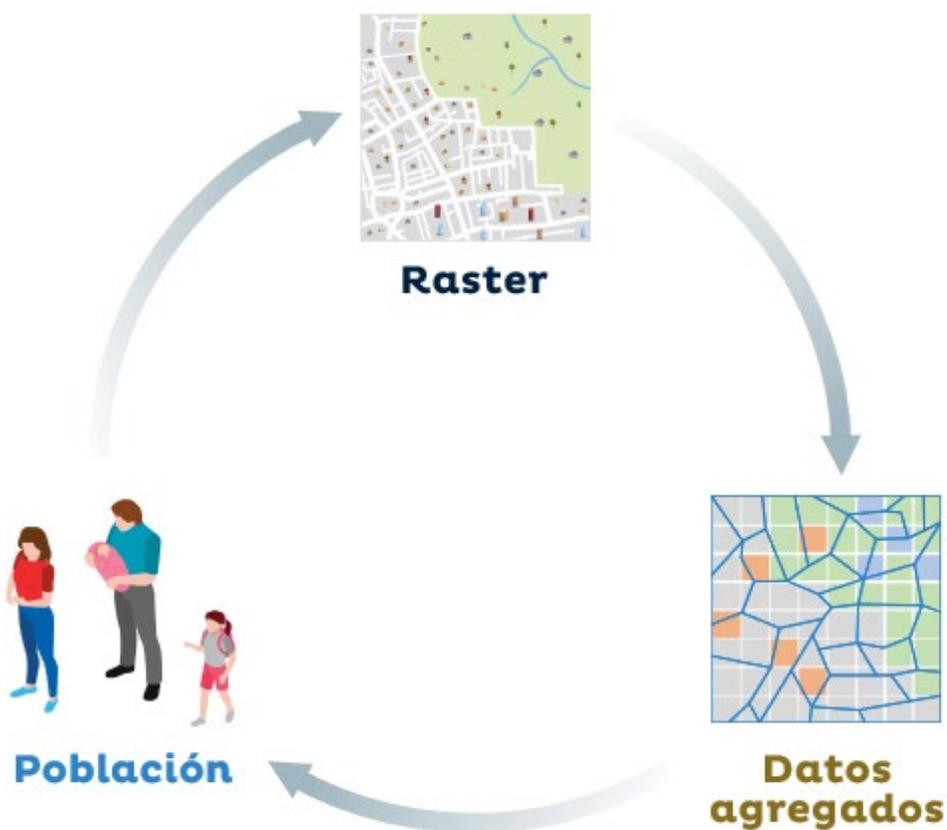
En la percepción remota, el campo instantáneo de vista de los sensores es lo suficientemente amplio para captar algunos objetos como las casas, árboles e incluso los cuerpos de agua. Sin embargo, el problema que declaran (Cárdenas *et al.*, 2010) ocurre al inferir la proporción de cobertura de cada clase dentro del sensor, lo cual se suele denominar clasificación; sumado a la

cobertura en fracciones de los píxeles. Este problema había sido pasado por alto en años anteriores, pero con la llegada de satélites comerciales de alta resolución y programas especializados se ha vuelto factible evaluar los píxeles e incluso establecer diferentes resoluciones para predecir la distribución espacial de las clases dentro de cada archivo raster.

6.2.3 Vínculo entre imágenes e información sociodemográfica

Para encontrar el vínculo entre las imágenes satelitales y la información sociodemográfica vale la pena hacernos la pregunta ¿cómo se puede hacer la vinculación de datos codificados a través de píxeles con aquellos cuya unidad de análisis son las personas? Kugler *et al.*, (2015) proponen que el elemento clave que sirve para vincular a los archivos raster y las fuentes de información tradicionales es la ubicación geográfica. Para hacerlo es necesario ubicar las unidades administrativas en las cuales se colecta la información de las personas y tomarla como referencia para analizar las imágenes satelitales, de esta forma los límites proporcionan nuevas unidades de análisis que sirven para vincular el espacio con la población. En el siguiente diagrama se muestra este proceso en el que se puede observar en el modo en que las imágenes satelitales de la Tierra se pueden vincular con la población a través de las unidades administrativas, es decir, de los límites que permiten ligar a las personas con su ubicación en el espacio.

Diagrama 6 Unión de los archivos raster y los datos poblacionales



Fuente: elaboración propia con base en Kugler *et al.*, (2015)

Este proceso se entorpece con algunas dificultades técnicas, la más común es que no existe una misma identificación de las unidades administrativas que sea común entre las fuentes de información tradicionales y las imágenes, sobre todo cuando dichas unidades son pequeñas. En este orden de ideas, Kugler *et al.*, (2015) describen otro de los problemas que es que las delimitaciones geográficas que son establecidas en las unidades administrativas no son fijas, sino que cambian a lo largo del tiempo y esto dificulta un seguimiento temporal. Esta situación se puede observar en los censos de población que se llevan a cabo cada 10 años, en donde suele pasar que las unidades administrativas cambian o se modifican entre un levantamiento y otro.

Los límites de las unidades administrativas de mayor tamaño suelen tener menos variación, tal y como es el caso de los países. Pero en los límites de menor tamaño se enfrenta el problema de que los límites cambian de dimensión o incluso desaparecen las unidades administrativas. Para

salvar ese problema, actualmente existen algunos proyectos cuyo objetivo es coleccionar y distribuir la información acerca de los límites administrativos de menor tamaño, cómo podrían ser los Estados, municipios o condados; por ejemplo, el Grupo de Trabajo de Información Geográfica de Naciones Unidas (UNSALB) o el proyecto GADM que se lleva a cabo en la Universidad de Berkeley. En este sentido, Brigham *et al.*, (2011) llevaron a cabo una comparación para medir el nivel de precisión que tenían estos dos proyectos y encontraron diferencias en el número de países que proporcionaron información y aquellos que lo hicieron de manera completa. De esta forma, concluyeron que el proyecto UNSALB contaba con una cobertura mayor, pero su precisión era menor al momento de comparar los datos con las fuentes de información oficiales.

Esta situación pone en evidencia las dificultades al momento de considerar los límites administrativos y, que no son elemento fijo en el territorio, sino que responde a acuerdos políticos que pueden cambiar a lo largo del tiempo. Por ejemplo, esta situación reportó una incidencia en el Censo de Población y Vivienda 2020 que llevó a cabo el Inegi. De acuerdo con la presentación de resultados, en esta nueva edición se consignaron 13 municipios de nueva creación que no aparecían en la versión de 2010 (Inegi, 2020). Esto no solamente significa un cambio en la manera de presentar los resultados, sino un impacto en las políticas públicas, ya que los recursos económicos se distribuyen con base en criterios poblacionales.

6.2.4 Del MAUP al MSTUP

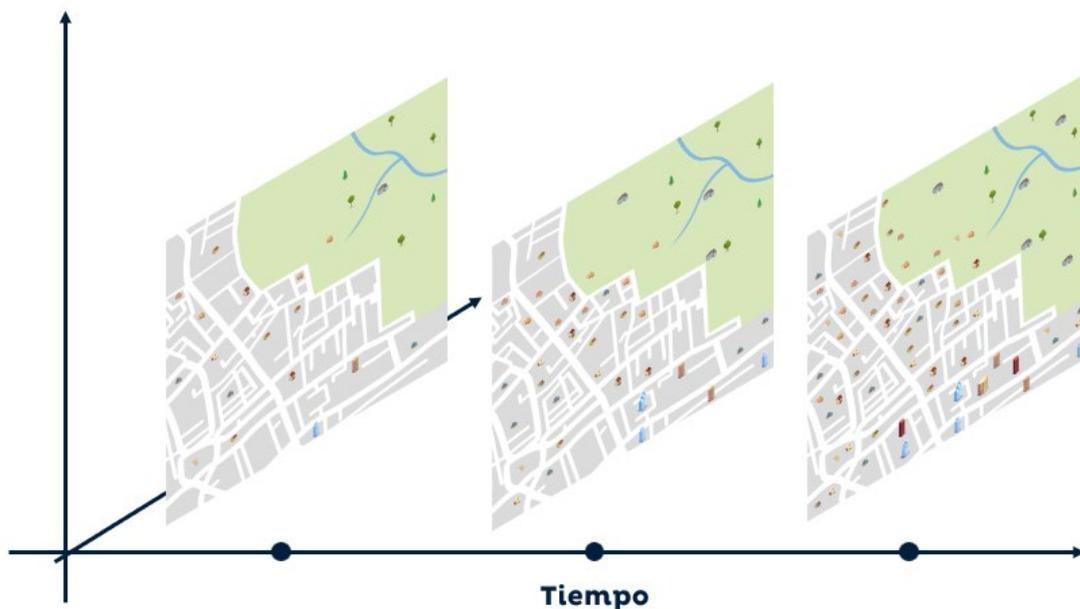
Ahora bien, existen distorsiones que se generan en las estadísticas resumen (como el promedio) que ocurren cuando varía la escala geográfica, conocidas en Geografía como el problema del área modificable. Estas distorsiones se deben al hecho de que una menor cantidad de unidades, ubicadas en un área en particular, provoca que la varianza de las estimaciones aumente. Mientras que si se tiene mayor cantidad de observaciones, ocurre lo opuesto, es decir, disminuye la distorsión de los datos.

De acuerdo con Martin *et al.* (2015) la variación que existe en la escala geográfica también se presenta cuando los fenómenos tienen una variación a través del tiempo. Al igual que sucede con el MAUP, la varianza de las estimaciones que provienen de una serie de tiempo fluctúa y aumenta en la medida en que la duración del tiempo se reduce. Uno de los ejemplos que describe este autor para mostrar este efecto son los ciclos de temperatura que suelen ser más bajos por la

noche y más altos durante el día. Entonces, si solo se considera un periodo de 24 horas y se toma como fecha de referencia el mediodía, el resultado está condicionado por el ciclo temporal o dicho de otra manera por la hora específica en la cual se llevan a cabo las observaciones. El problema lo resume Yuan (2007) cuando afirma que la información que tiene un sustento temporal sólido, carece de un soporte espacial adecuado y viceversa, cuando se privilegia la información geográfica se omite la información temporal.

Ante esta necesidad existen varios modelos teóricos que incorporan la dimensión temporal y espacial a las observaciones geográficas. Uno de los primeros acercamientos lo sugiere Yuan (2007) quien propone pensar en *objetos*, los cuales poseen características tanto espaciales como temporales (ver Diagrama 7). Así, lo que se toma como referencia son entidades a partir de las cuales se les puede dar seguimiento a las modificaciones espaciales que ocurren en un periodo determinado. Este seguimiento abre la posibilidad de dar cuenta de las modificaciones a la apariencia del objeto, sin embargo, también se puede dar el caso de que desaparezca o mantenga una estabilidad a lo largo del tiempo. Uno de los problemas que se tiene al llevar a cabo este proceso es que el tiempo se vuelve discreto, ya que no es posible dar un seguimiento continuo, por más que las mediciones se tomen una tras otra.

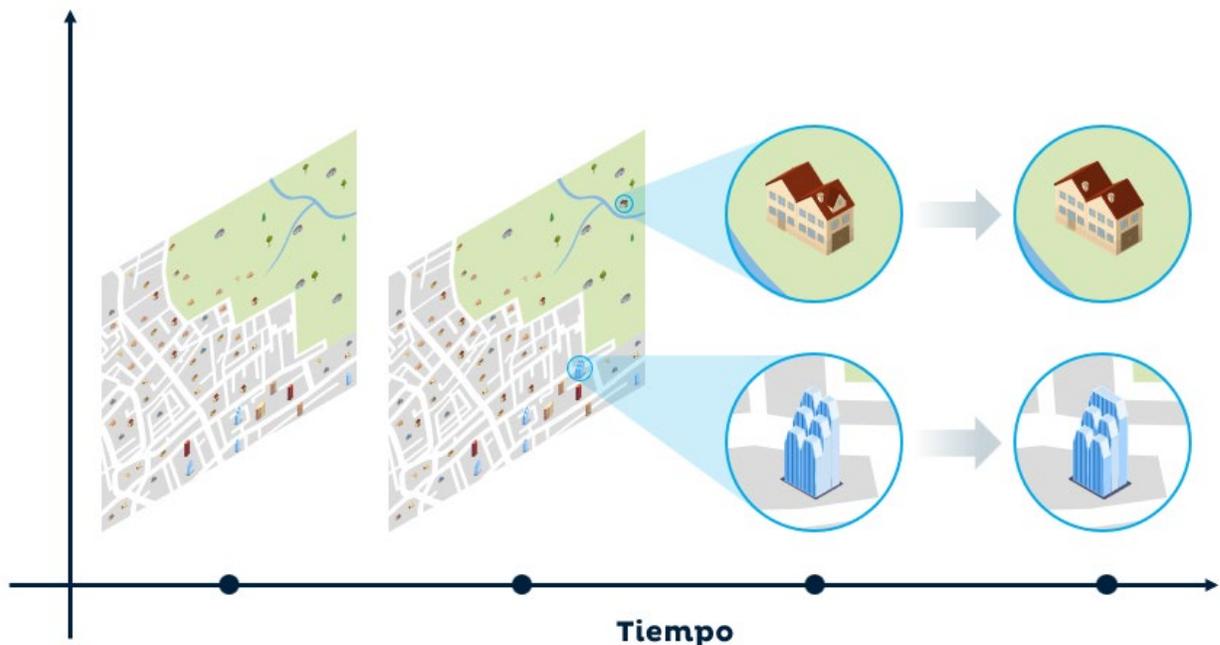
Diagrama 7 Cambios a través del tiempo



Fuente: elaboración propia con base en Yuan (2007)

Otro de los modelos para capturar los cambios que ocurren a lo largo del tiempo en las unidades gráficas es una lista de una virtud variable en donde cada uno de los términos corresponde a un objeto bajo la premisa básica de registrar los cambios que ocurren en intervalos temporales. Y aunque, en la teoría parece que el modelo ofrece buenos resultados, de acuerdo con Ma *et al.*, (2020) existen restricciones computacionales: buscar un estado del pasado lejano requiere leer toda la historia y esto puede resultar ineficiente para determinado tipo de investigaciones (ver Diagrama 8).

Diagrama 8 Cambios de los elementos a través del tiempo



Fuente: elaboración propia con base en Yuan (2007)

Ante este panorama, los autores sugieren la posibilidad de llevar a cabo diversos tipos de interacciones, a largo de un periodo determinado, que modifican o determinan el estado actual. En

la práctica, esto es ya ocurre cuando se toma el promedio de los píxeles que están contenidos en diferentes fotografías instantáneas y se asume que esa imagen representa “el periodo evaluado”. Por ejemplo, ante la necesidad de observar el crecimiento de una ciudad, se recurre a la toma de instantáneas en diferentes momentos y estas fotografías se unifican en una sola imagen, ante la cual se asume que representa el comportamiento de ese periodo en particular (un semestre, un año, una década) (ver Diagrama 9).

Diagrama 9 Composición de imágenes a través del tiempo



Fuente: elaboración propia con base en Yuan (2007)

Para representar la interacción que existe entre el espacio y el tiempo existen otros modelos, por ejemplo, Peuquet y Duan (1995), sugieren que el punto clave es fijar la atención en los *eventos* que dan origen a los cambios y se reflejan en los diferentes estados. Esta propuesta trae consigo un cambio de paradigma en torno a cómo se mide el tiempo, ya que no se considera una métrica

preestablecida, sino que está sujeta a la ocurrencia de un hecho en particular, el cual se toma como referencia para llevar a cabo las mediciones.

Carré y Hamdani (2021) proponen niveles que funcionan como una pirámide. En el primero de ellos el mundo se percibe a través de fotos instantáneas que son ordenadas a través del tiempo y aunque esto parece muy simple, sirve para responder a la pregunta de si hay cambios en lo que se está observando. En el segundo nivel, cambia la perspectiva y el foco de atención son los objetos, hay seguimiento a través del tiempo y, al igual que en el nivel anterior, es posible evaluar cuáles son esos cambios. En el tercer nivel, en esencia, se comparan los dos niveles anteriores y se introducen los conceptos de evento y proceso, de esta manera los fenómenos espaciales se tratan como una colección de acontecimientos. En el cuarto nivel se toma en cuenta la identidad, esto quiere decir que la preocupación está en saber si el objeto de referencia se mantiene de forma coherente a pesar de que existan cambios. En el quinto nivel el énfasis está en estudiar los fenómenos geográficos para darles sentido, lo que significa que se trata de responder por qué se han producido cambios. Finalmente, en el último nivel se fija la atención en la interpretación y se trata de dar un sentido a los cambios, con base en la experiencia y el campo de especialización de las personas (ver Diagrama 10).

Diagrama 10 Niveles de análisis de las imágenes satelitales



Fuente: elaboración propia con base en Carré y Hamdani (2021)

6.2.5 Limitaciones al vínculo entre imágenes e información sociodemográfica

Hasta aquí se ha podido observar que las imágenes satelitales tienen alcances y limitaciones. En cuanto a sus limitaciones, se enlistan tres: 1) la gran vulnerabilidad climática a la que está expuesta su recolección porque impacta en la calidad de la imagen; 2) las imágenes satelitales responden a objetivos e intereses específicos, lo cual representa una oferta de información que no siempre empata con intereses específicos; y, 3) se debe atender el tema de si es o no es posible equiparar personas con objetos para realizar un análisis demográfico.

Ahora bien, en la lista de los alcances vale la pena destacar que este instrumento permite obtener información sin tener que estar en contacto físico con los objetos de estudio, lo cual abre un abanico de posibilidades a cuestiones en las que ya se cuenta información. Este enlace tendría un camino de ida y vuelta, ya que afianzar el vínculo del espacio con la población requiere de información proveniente de fuentes de datos tradicionales, en este caso, el censo de población y vivienda. Lo que supondría la complementación de las unidades de análisis, el problema aparece

si no existe una misma identificación entre las unidades administrativas, sobre todo cuando estas son pequeñas y cambian cada diez años. Esto nos coloca en un dilema, pues (a falta de información) implicaría utilizar unidades administrativas de mayor tamaño, propiciaría que la valiosa oferta enunciada en párrafos anteriores sea cuestionada y enfrentaría el problema de dar preferencia a una mayor cobertura sacrificando la precisión de los resultados. ¿Esta situación es la que queremos reproducir en las investigaciones?

La respuesta es complicada, pues entre otros temas implica que las dos perspectivas analíticas son instrumentos necesarios para poder allegarse al conocimiento más completo de un fenómeno. Para el caso de México, existe un retraso en la infraestructura para el análisis socioespacial de la información y por ende, del acceso público o institucional que se tiene. En la medida que estos se fortalezcan, será posible realizar ejercicios que ayuden a futuros investigadores e interesados a emparejar y complementar temas de este tipo.

A partir de todo lo expuesto anteriormente, la disponibilidad de datos o fuentes de información se ha convertido en un campo de disputa en el que convergen los beneficios y las desventajas de tener acceso a ellos.

Debido a que las fuentes de información pueden contener datos cuya relevancia es importante para los resultados de una investigación, es necesario cuestionar algunos detalles antes de su uso: ¿la información o la fuente es fidedigna?, ¿la información o los datos son obsoletos?, y ¿la fuente, la información o los datos son pertinentes? Sin duda, la veracidad de una fuente debe ser cuestionada para conocer si la información obtenida procede de una metodología adecuada o ha sido recabada bajo una perspectiva ética y honesta, y determinar si la información que nos brinda es confiable y, por lo tanto, se adapta a nuestros objetivos.

En el presente, con el acceso a medios digitales, la confiabilidad se ha vuelto cada vez más relevante debido a que comunicar y publicar información está al alcance de cualquier persona. A partir de esta apertura, las diferentes fuentes de información se pueden clasificar según su formato (impresas, electrónicas, virtuales o en línea) y según origen de la información (personales, institucionales, documentales).

Las fuentes de información personal, se caracterizan por su difícil acceso, ya que ofrecen datos relativos a una persona física, identificable y, bajo este argumento, cualquier persona o grupo

de personas puede constituir una fuente. Un medio usual para la obtención de este tipo de información es la transmisión oral que posteriormente puede quedar plasmada en documentos; otra forma de obtención, es a la inversa, bajo el registro de forma personal, a través de un registro en cualquier documento que posteriormente se sistematiza en una fuente de datos personales. El uso de datos personales en investigaciones ha causado debates intensos con relación al uso potencial que se puede obtener al conjugar dichos datos con otras fuentes de información, por ejemplo, las imágenes satelitales. En tales casos, entra en juego un asunto ético relacionado con los límites y con la posibilidad de que la información pueda ser usada para identificar directa o indirectamente a una persona y traspasar territorios como la privacidad y la confidencialidad (por privacidad entiéndase "el derecho del individuo a decidir si divulgará al gobierno sus pensamientos, opiniones, sentimientos y hechos de su vida personal y en qué medida" [Comisión Presidencial, 1971: 197]). Esta delimitación hace referencia a la privacidad de la información e implica la posibilidad de que una gran cantidad de información se difunda rápidamente por medios electrónicos, por lo que, el sujeto tendría cierta libertad en elegir qué información compartirá y cuál reservará.

Por otro lado, la confidencialidad se refiere a "la transmisión de información personal por parte de otra persona que no sea la persona identificada". Esto aplica cuando la información que nos identifica es entregada a otra persona, entidad o institución, y es esta quien, bajo restricciones, transforma, edita o recorta la información para que no sea identificada la persona; es el caso de las encuestadoras o institutos que recaban información con fines estadísticos.

Ante el panorama presentado, existen leyes y regulaciones que exigen la protección a la privacidad de los ciudadanos, sin embargo, estas reglamentaciones suelen ser diferentes en cada país y los habitantes pocas veces conocen este derecho y, por lo tanto, desconocen los límites de dicho aseguramiento. Además de lo anterior y pese a las normas, vale la pena reflexionar sobre el poder que adquieren las oficinas estadísticas al recabar y reservar información confidencial de cientos de ciudadanos, sobre todo porque se trata de información que una vez transferida pierde la capacidad del libre albedrío porque se convierten en datos electrónicos. El gran debate, entonces, consiste en que se contraponen el derecho a la privacidad con el derecho al acceso de la información. De esta manera, las principales tensiones en el tema son ¿hasta qué punto las agencias estadísticas deben respetar y proteger la información personal que se les confía?, y, al hacerlo, ¿se le resta confiabilidad a los datos?

Un acercamiento permite observar que, por un lado, la ética de las agencias da garantía de confidencialidad que facilita el reclutamiento, la honestidad y la confiabilidad de los sujetos de estudio que entregan datos confidenciales; pero, por otra parte, la responsabilidad ética de mantener la confidencialidad de las personas no siempre se comparte y en ocasiones el derecho al acceso a los datos se sobrepone. Aunque hasta ahora, esta tensión se había mantenido controlada, el conflicto se profundizó con el acceso a imágenes satelitales, en donde es posible alcanzar un grado muy fino de detalle que posibilita resultados sobresalientes que pocos investigadores están dispuestos a despreciar, por lo cual es necesario explorar los beneficios y riesgos del uso de datos geoespaciales, y revisar las leyes de protección de datos.

A estas alturas del presente texto, el asunto de la ética se complica profundamente porque las restricciones de la privacidad pueden limitar el acceso a las bases de datos que posibilitan el desarrollo de una investigación; además existe el riesgo de que la información confidencial pueda ser revelada, lo que pondría en riesgo las prácticas científicas como levantamientos de censos, encuestas o registros administrativos. La encrucijada es tal que en el primer caso se estaría vulnerando el derecho a la información; en el segundo, el derecho a la privacidad; pero en ambos el ejercicio científico sufriría afectaciones.

Ante este fenómeno y sin lugar a duda, el respeto por la privacidad debe imperar antes que cualquier ejercicio, así que una forma, entre otras posibilidades, de resolver el conflicto entre el derecho de usuarios y de los proveedores de servicios puede ser establecer un sistema de registro que identifique a las personas que acceden a datos confidenciales, al proyecto para el que se solicitaron los datos, y un monitoreo del cumplimiento de esta regla. Además de considerar la heterogeneidad de datos, su complejidad y variaciones en los sistemas de información, así como las diferencias en la gestión en los múltiples sitios y los objetivos de las agencias de datos.

Actualmente, las innovaciones en el acceso a los datos, los vínculos y las búsquedas han aumentado el riesgo de divulgación, poniendo en tela de juicio métodos utilizados para aplicaciones y productos de datos híbridos mediante la combinación de datos geoespaciales con otros datos demográficos y socioeconómicos (Martin y Higgs, 1997). Por tal motivo, se sugiere la protección de datos; una manera consiste en clasificar en datos restringidos y datos de acceso restringido, Wolf describe estos enfoques. El primero se refiere a restringir “el contenido de los datos antes de su divulgación al público en general” y el segundo a limitar “las condiciones bajo las cuales se puede

acceder a los datos, es decir, quién puede tener acceso, en qué lugares, para qué fines”(De Wolf, 2003: 66).

Usualmente, las oficinas estadísticas cuidan la confidencialidad mediante protección de datos en el análisis estadístico y mediante diversas modalidades de acceso, la más extendida es la modificación de datos eliminando información que podría revelar directa o indirectamente la identidad de una persona. En este tenor, también se puede realizar un análisis estadístico propio de los datos y con ello se publica información específica sobre las tendencias observadas para grupos de encuestados en lugar de para individuos.

Algunas oficinas estadísticas controlan el acceso de datos mediante licencias, acceso remoto, sitios seguros y ofrecen una variedad de acceso a los usuarios al tiempo que protegen los datos confidenciales. Y, aunque, todas las medidas mencionadas son funcionales, solamente aplican para la difusión de microdatos y la creciente demanda pública y la disponibilidad de datos detallados asociados con unidades geográficas de áreas pequeñas desafían la consideración tradicional del riesgo de divulgación (Steel y Sperling, 2001).

En cualquiera de los casos, es necesario tomar en cuenta en cualquier investigación que al utilizar aplicaciones de los datos satelitales se debe considerar cómo recopilar y usar esos datos de manera ética, lo cual involucra a quien recaba dicha información y a quien la utiliza (las comunidades científicas, podrían ser quienes regulen dicho acceso).

6.2.5 Máscaras urbanas

Las máscaras urbanas se retoman del análisis y monitoreo de crecimiento urbano del corredor Metropolitano Centro País³, el cual cuenta con la participación de especialistas del Centro de Investigación en Ciencias de Información Geoespacial. Su objetivo es mapear el grado de urbanización de 109 municipios del centro del país mediante imágenes satelitales y técnicas de percepción remota.

Para llevar a cabo este proceso los investigadores tomaron como referencia las imágenes satelitales del satélite Landsat 8, el cual tiene la característica que toma una fotografía de la misma zona de la superficie terrestre cada 16 días. Esto significa que cada mes se llegan a tener hasta dos imágenes de una misma región, por lo que este mapeo constante de la superficie terrestre permite

³ Proyecto FORDECyT no. 297259 2da. Etapa

visualizar cambios como el aumento de la cobertura urbana, la cual está relacionada con el crecimiento de la población. Con todas las imágenes recopiladas, los investigadores construyeron 29 mosaicos en donde los píxeles tenían el menor índice de nubosidad posible, con lo cual se garantizaba que su contenido reflejaba las características de la superficie terrestre.

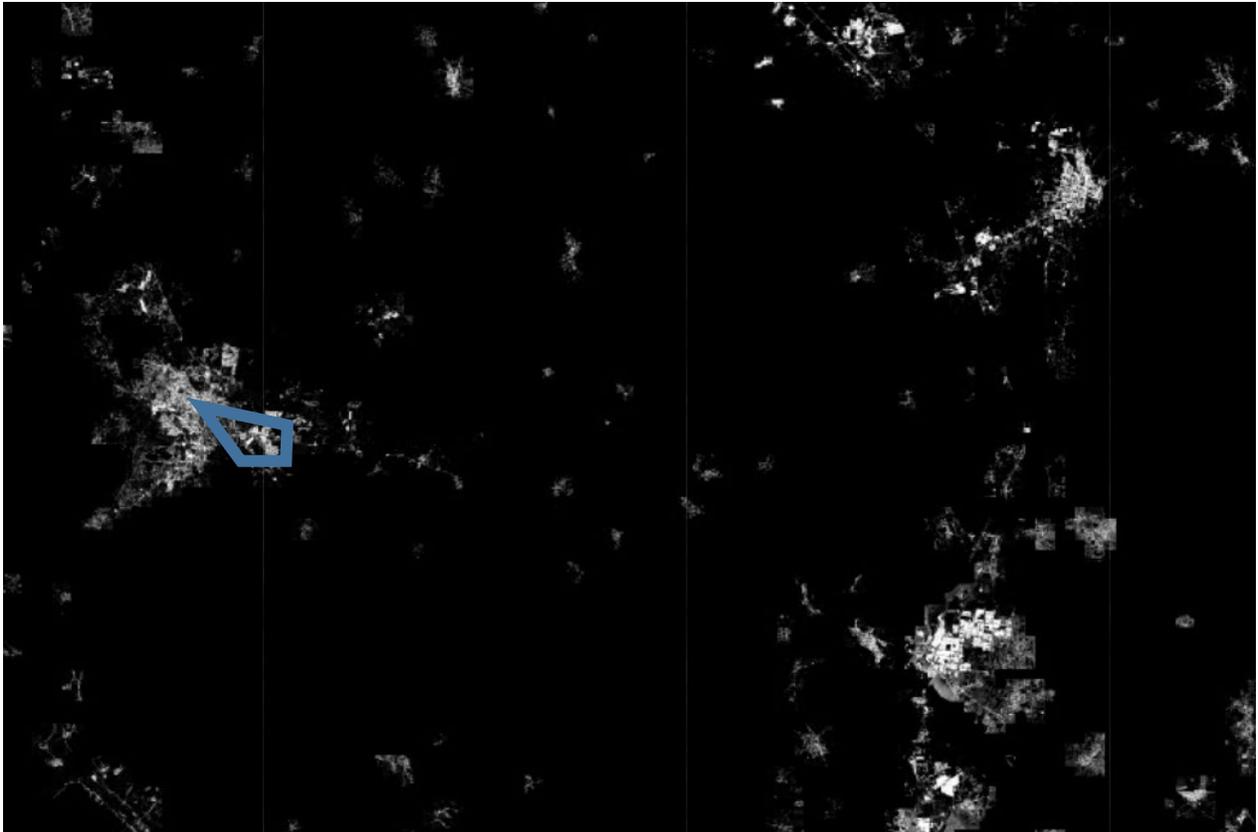
La técnica que utilizan para asignar una clasificación a los píxeles es lo que se conoce como de mezclado espectral lineal (Keshava y Mustard, 2002; Silván-Cárdenas *et al.*, 2010), el cual toma como referencia las firmas espectrales de materiales sintéticos que se suelen utilizar en los asentamientos urbanos. Por ejemplo, algunos materiales como el cemento reflejan un patrón particular de firmas espectrales que son captadas por los sensores y gracias a ello son reconocibles. De esta manera, cuando se tome una fotografía en otra zona y se encuentra en los mismos valores se puede concluir que se trata de un lugar en donde se encuentran asentamientos urbanos o se tiene algún grado de urbanización.

El proceso consiste en seleccionar las firmas espectrales de tres materiales que se suelen encontrar en las superficies urbanas, a saber: techos de lámina, ladrillo y cemento. Con estas firmas espectrales se calculó el porcentaje de cobertura urbana como la suma de las tres abundancias asociadas a cada una de ellas multiplicado por 100. Esto significa que cada uno de los píxeles obtuvo un valor numérico que representaba el grado de cobertura urbana y en donde los mayores porcentajes se encontraron en zonas céntricas y disminuyeron en la medida en la que se llegaba a las zonas periféricas de los asentamientos urbanos. Un aspecto que señalan los investigadores es que los caminos no se detectan o en todo caso presentan una cobertura urbana baja, ya que su firma espectral no estuvo representada en los 3 *endmembers* urbanos que fueron elegidos (Silván-Cárdenas *et al.*, 2021).

Una de las limitaciones que se tienen al llevar a cabo esta técnica es que se suelen identificar suelos rocosos y arbustos como parte de la cobertura urbana. Para corregir estos errores se utiliza información auxiliar, por ejemplo, las imágenes de luces nocturnas, las cuales captan el grado de luminosidad en las diversas zonas por lo que son un buen indicador de los asentamientos humanos, a pesar de que su resolución es de 465 metros. Esta última capa se comparó con la capa de cobertura urbana y en aquellos casos en donde no se traslapaban los píxeles se le asignó un valor de cero al porcentaje de cobertura urbana, ya que no representaba un asentamiento humano como tal si no era el resultado de una mala clasificación de los píxeles. Esta información auxiliar sirvió para calibrar la máscara de cobertura urbana y detectar inconsistencias (Silván-Cárdenas *et al.*, 2021).

Para llevar a cabo este proceso los investigadores tomaron una imagen representativa para todo el periodo de análisis y encontraron que el porcentaje de cobertura urbana se calculó como la suma en las 3 abundancias asociadas con superficies urbanas. En el mapa siguiente se muestra un ejemplo de los archivos raster con los cuales se construye el modelo y que corresponde al año.

Mapa 2 Máscaras urbanas



Fuente: Silván-Cárdenas *et al.* (2021)

Para incorporar estos datos se llevó a cabo una estadística zonal, lo que significa que se contabilizó el número de píxeles y su valor dentro de cada una de las delimitaciones político-administrativas, por ejemplo, en el mapa anterior esto corresponde a la zona azul. De esta forma, bueno se tiene un valor para cada una de las regiones que representa el grado de urbanización y con ello es factible llevar a cabo una comparación a través del tiempo ya que el valor que se desprende del número de píxeles es diferente.

6.3 Datos sociodemográficos

En esta sección se presentan las dos fuentes de información socio demográficas que se utilizaron para llevar a cabo las estimaciones. Por un lado, se usó la encuesta Intercensal 2015 y aquí es el ejercicio de mayor relevancia que se tiene a mitad de cada década y sirve para actualizar los datos poblacionales y como referencia para el levantamiento censal. Por el otro, se encuentran las proyecciones de población, ya que son los datos temporales que permiten tener una idea de la evolución poblacional con distintos niveles de desagregación entre los cuales se encuentran los municipios.

6.3.1 Encuesta Intercensal 2015

El objetivo de la encuesta Intercensal 2015 fue actualizar la información socio demográfica de México. Por este motivo se llevó a cabo cinco años después del censo de 2010, lo cual le permitió servir como referencia para iniciar los preparativos del levantamiento censal en 2020. Además, se trata de la encuesta de mayor tamaño que se aplica en los hogares, su información es útil para llevar a cabo un diagnóstico y seguimiento de los Objetivos de Desarrollo Sostenible (ODS)(Gil, 2018).

Con la información de esta encuesta si analiza la distribución poblacional a nivel nacional, estatal, municipal y para localidades que tiene más de 50,000 habitantes. Esto significa que el nivel de confianza de los estimadores es alto, ya que su distribución y tamaño permiten tener suficientes registros para dar cuenta de las características sociodemográficas del país (Inegi, 2018).

6.3.2 Proyecciones de población

De acuerdo con CONAPO (2018: 3), “las proyecciones de la población son un instrumento de política pública que permite construir y evaluar posibles escenarios a mediano y largo plazo que sirven de base para calcular las futuras demandas de empleo, educación, salud y vivienda, entre otras.” En el papel, son un elemento para la planeación local que permite a los gobiernos de diversos niveles diseñar estrategias con las cuales se atienden las necesidades básicas de la población. Para su desarrollo se consideran diversas fuentes de información como son los centros de población y conteos, así como los resultados de la encuesta Intercensal de 2015. De acuerdo con la nota

metodológica que presentó CONAPO en el último ejercicio, se analizaron diversos métodos para hacer los cálculos y se optó por un enfoque temporal debido a las inconsistencias en las fuentes de información tradicional. Por ejemplo, en el caso de los nacimientos, las estadísticas vitales tienen un subregistro con registro extemporáneo de los eventos, esto mismo sucede en el caso de las defunciones porque existe un alto nivel de omisión en muchos de los municipios del país. Debido a estos problemas en la ecuación compensadora para llevar a cabo el ejercicio de proyección fue necesario recurrir a modelos de series de tiempo con el supuesto de que los factores que afectan el comportamiento de la población se mantienen a lo largo del tiempo.

Si bien es cierto que este supuesto fue válido cuando se llevó a cabo el ejercicio en 2018, la aparición de la pandemia en 2020 significó un cambio en los patrones que hasta ese momento se habían observado y por ende el supuesto de continuidad de los factores asociados al cambio en el volumen y composición de la población es falible. Aun así, los investigadores dejan claro que se trata de estimaciones, por lo que tienen un cierto nivel de confianza y es factible suponer que para algunos casos se sobrepasen los niveles de confianza deseados y por ende la información no tenga la calidad necesaria para ser considerada como un insumo en el diseño de políticas públicas. Sin embargo, a pesar de esta precisión no existe una medida de confiabilidad; en cada una de las estimaciones solamente se mencionan algunos casos específicos.

En este trabajo se propone complementar la información sociodemográfica con datos espaciales. Para desarrollar las estimaciones temporales con base en el filtro de Kalman se toman como referencia los datos de la Encuesta Intercensal 2015 y las proyecciones de población sobre el total de población que existe en los municipios seleccionados y su correspondiente varianza. Posteriormente, se incorpora la información de la máscara urbana (ver Diagrama 11), la cual proviene de la información satelital y del proyecto del Centro de Investigación en Ciencias de Información Geoespacial, A.C. (Silván *et al.*, 2021).

Esta idea de combinar fuentes sociodemográficas y espaciales inició en la década de 1960 con la intención de tener una mejor aproximación al total de población y ha sido de bastante utilidad, sobre todo cuando en los lugares se carece de información. Un ejemplo de ellos lo describen Hillson *et al.* (2019), quienes usaron imágenes satelitales en Bo (Sierra Leona) a partir de los datos del Landsat, para estimar el total de población en regiones alejadas a la capital donde se carece de información censal que permita identificar a los asentamientos urbanos. Otro ejercicio

similar lo llevaron a cabo Iisaka y Hegedus (1982) quienes utilizaron imágenes satelitales para calcular la densidad de población en la zona metropolitana de Tokio.

Finalmente, para validar las estimaciones se comparan los resultados obtenidos con la información censal, la cual se emplea como referencia para la toma de decisiones. A pesar de que estos son los datos que se consideran como válidos para conocer el perfil sociodemográfico de México, estos se levantaron al inicio de la pandemia de 2020, por lo que conviene hacer una reflexión en torno a sus alcances y limitaciones.

Diagrama 11 Datos utilizados para el modelo y validación



Fuente: elaboración propia.

6.4 Área de estudio

Para llevar a cabo el estudio se consideran 105 municipios que se ubican en Aguascalientes, Guanajuato, Querétaro y San Luis Potosí, las cuales forman parte del *Corredor Metropolitano Centro* e integran la región centro-norte del país (ver Mapa 3). Las características de esta región la hacen de gran importancia para su estudio, ya que es una de las zonas industriales más importantes del país y es donde se concentran algunas de las plantas automotrices de mayor relevancia para la economía mexicana (Yuridia, 2014). De acuerdo con Pasalagua y Sánchez (2022), esto se debe a que desde la década de los años 60 empezó una reordenación de las empresas automotrices, lo que dio lugar a corredores industriales en donde se concentran gran parte de las empresas y esta región fue una de las que se vio favorecida. Para el objeto de estudio que aquí se realizará, se

presenta una descripción general de las cuatro entidades con el fin de evidenciar esas características que nos servirán más adelante en el análisis de datos.

Aguascalientes

Calientes es una de las entidades federativas por su ubicación geográfica es una zona estratégica que sirve de enlace entre las actividades que desarrollan en el centro del país y el norte. Una de sus principales características es que cuenta cuando un sector automotriz de gran importancia y esto ha generado con importante impulso económico (Yuridia, 2014). De acuerdo con los datos de la Encuesta Intercensal de 2015 su población era de 1.3 millones de habitantes y desde ese mismo año ha experimentado una reducción en sus niveles de pobreza⁴.

Guanajuato

De acuerdo con Hidalgo *et al.* (2019) el estado de Guanajuato se caracteriza por ser una entidad especializada en la producción del calzado y en particular el municipio de León se ha distinguido por ser una de las regiones con mayor auge económico en las últimas décadas. Además, con base en los datos de la Encuesta Intercensal de 2015 su población era de 5.8 millones de habitantes y al igual que Aguascalientes tuvo una relevante reducción de la pobreza según las estimaciones de Consejo Nacional de Evaluación de la Política de Desarrollo Social (CONEVAL).

San Luis Potosí

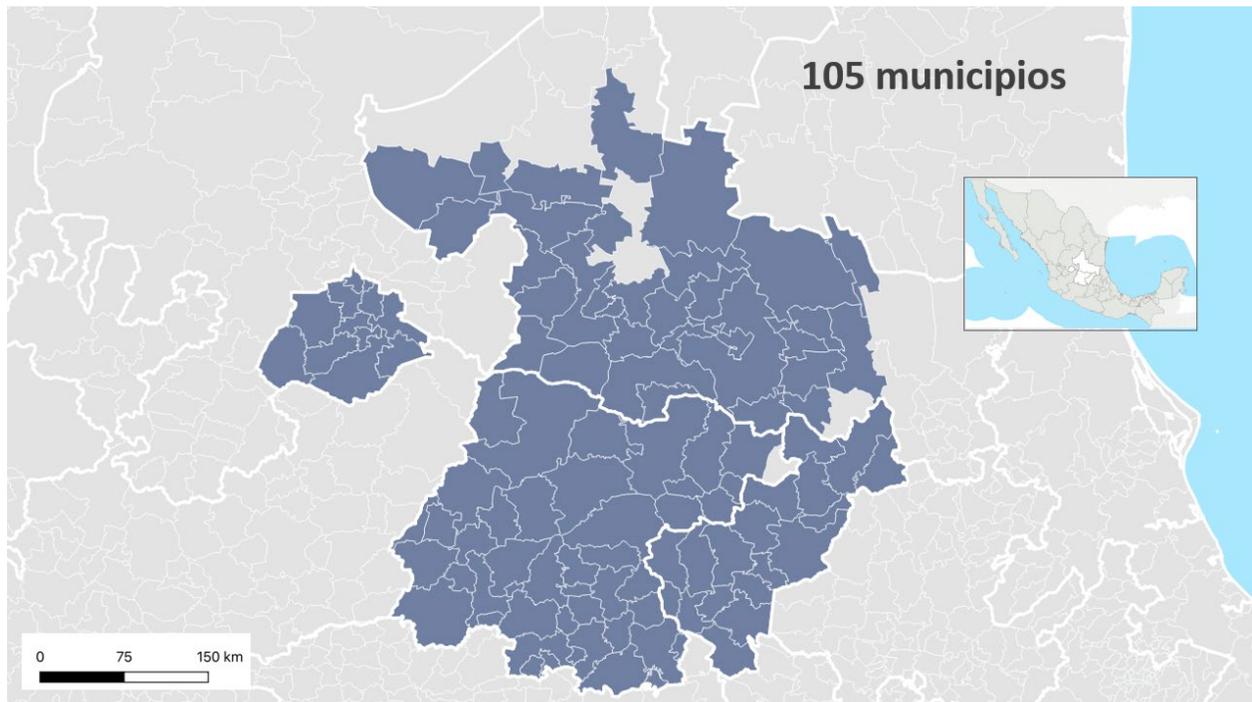
El estado de San Luis Potosí (SLP) se compone de 58 municipios y se ubica en medio de tres grandes ciudades: CDMX, Monterrey y Guadalajara. En 2015 su población era de 2 millones de habitantes y aunque un descenso en sus niveles de pobreza de 2016 a 2018, en 2020 experimentó un incremento de cuatro puntos porcentuales. De acuerdo con Pasalagua y Sánchez (2022), SLP que forma parte de la región del centro del país en donde se concentra una de las industrias más relevantes para la economía mexicana que es un sector automotriz.

⁴ Ver https://www.coneval.org.mx/Medicion/MP/Paginas/Pobreza_2020.aspx

Querétaro

Al igual que los casos anteriores, Querétaro es una de las zonas industriales de mayor importancia económica tanto para la región como para el país. Sin embargo, una de las características que distingue a esta entidad federativa es se ha convertido en uno de los polos de desarrollo del sector aeronáutico. De acuerdo con Salinas-García (2012), este auge ha provocado el impulso de otros sectores como el educativo, ya que en 2007 se creó la Universidad aeronáutica de Querétaro con la intención de expandir el desarrollo de esta industria y contar con una mano de obra altamente calificada. En cuanto a los datos poblacionales, se trata de una entidad que tenía 2.7 millones de habitantes de acuerdo con los datos de la Encuesta Intercensal 2015 y tuvo una reducción en los niveles de pobreza de 2016 a 2018, sin embargo, en 2020 registró un ligero aumento.

Mapa 3 Municipios seleccionados



Fuente: elaboración propia.

6.5 Unidad de análisis

Tal y como se discutió en los capítulos anteriores, la unidad de análisis juega un papel determinante al momento de hacer las estimaciones. Sin embargo, una de las restricciones que se tiene es que la información disponible para llevar a cabo los ejercicios es escasa, por este motivo se optó por

utilizar como referencia el municipio. Esta decisión obedece a que este nivel (o dominio) es utilizado para ponderar las encuestas y con base en ello se generan las estimaciones de las diferentes características sociodemográficas de la población. Por este motivo, resulta relevante llevar a cabo un ejercicio de esta naturaleza, ya que si se obtienen información de buena calidad entonces es factible ocuparla para calibrar los factores de expansión las encuestas.

En contraparte, si se toma una escala distinta tendrán datos aislados que no permitirán ser usados más allá de obtener un recuento poblacional. Por ejemplo, y se toma como referencia a las AGEBS, entonces la información será de utilidad únicamente para conocer el total de población, pero no tendrá una utilidad más allá de eso. El motivo es que el diseño muestral de la gran mayoría de las encuestas no permite tener una desagregación a este nivel.

6.6 Conclusiones

A lo largo de este capítulo se llevó a cabo una descripción de los datos que se van a utilizar para aplicar el filtro de Kalman y poder llevar a cabo las estimaciones de población de los 105 municipios seleccionados que forman parte del Corredor Centro País. los datos socio demográficos que se utilizan son las proyecciones de población y los datos de la Encuesta Intercensal de 2015, de esta última también se toma la varianza de los datos ya que, como se explicó en el capítulo número cinco, el modelo requiere de una medición que cuantifique el error de los datos que se toman como punto de partida para las estimaciones.

En cuánto a las imágenes satelitales, se utilizan las máscaras urbanas que fueron desarrolladas por el Centro Geo y que muestran los asentamientos humanos con un alto grado de resolución. Por este motivo, en la segunda parte del texto se describe con detalle cuál es el procedimiento para elaborar y, sobre todo, interpretar las imágenes satelitales. Además, un punto que es de vital importancia es que en el capítulo se destaca la unión de ambas fuentes de información que, si bien es cierto ofrecen un panorama de la misma área de estudio, tienen estructuras diferentes: la información sociodemográfica está codificada en una forma de base de datos por medio de filas y columnas; mientras que la información satelital son píxeles. Así, la forma que unir la información es mediante las delimitaciones de los municipios, ya que es el elemento común y que sirve de vínculo para unir estas dos fuentes de información.

Capítulo 7

Resultados

7.1 Introducción

A continuación, se describen los resultados de las estimaciones a nivel municipal con base en el filtro de Kalman. Si bien, en la primera parte del capítulo se presentan los resultados para los 105 municipios seleccionados que constituyen el Corredor urbano centro país, también se tomaron como referencia los datos censales para medir el desempeño de las estimaciones a partir de dos procedimientos. En el primero, es necesario obtener la diferencia porcentual de cada uno de los municipios para observar qué tanto varían los datos y generar una herramienta gráfica que permita visualizar los casos en donde existe una mayor discrepancia. En el segundo procedimiento, se obtiene el error absoluto medio, medida global utilizada para saber qué tan bueno fue el desempeño de las estimaciones. Esto se usa principalmente en casos como este en los que se hace una predicción.

Tras realizar los dos procedimientos y obtener esta medida se encontró que el error absoluto medio es de 21 %; y dado que no existe un criterio unánime sobre cuáles son los rangos en los que una información se considera confiable, para analizar este dato se retomó el criterio que emplea el Inegi para evaluar la confiabilidad de las cifras que se generan a partir de encuestas o censos. Dicho criterio que mide el desempeño de las cifras establece que si la variación es menor a 15 % la calidad del dato se considera excelente; si se encuentra entre 15 % y 25 % se considera bueno, y si existe una variación superior a 26 % se le estima como deficiente. Tomando este criterio como referencia, se determinó que la calidad de las estimaciones hechas con base en los datos poblacionales y las imágenes satelitales mediante el uso del filtro de Kalman tienen un buen desempeño.

Es importante resaltar que este estimador se construye con base en la dimensión espacio temporal, ejes fundamentales del constructo que denominamos población. También vale la pena tomar en cuenta que el estimado se construye con dos fuentes de información distintas: los datos de la encuesta intersexual de 2015 y la información de las imágenes satelitales.

7.2 Estimaciones de población

Para llevar a cabo este ejercicio se tomaron como referencia los datos de la encuesta Intercensal de 2015 y su variación correspondiente (error estándar) de cada uno de los estimadores que reflejan el total de población a nivel municipal. A partir de estos datos se analizó la evolución de los resultados durante un periodo de 5 años utilizando las imágenes satelitales que reflejan los cambios en la mancha urbana. Ya en el Capítulo 5 se desarrolló el modelo siguiente que sirve de referencia para las estimaciones:

$$\hat{x}_{k_{est}} = r_k(x'_{k,\varphi}) + (1 - r_k)(x''_{k,s})$$

En esta ecuación, el estimador $x'_{k,\varphi}$ se construye a partir de los datos poblacionales y toma como punto de partida el total de población de la Encuesta Intercensal 2015, mientras que el estimador $x''_{k,s}$ se construye con los datos que se obtienen de las imágenes satelitales y que corresponde al promedio de píxeles que existe en cada una de las delimitaciones administrativas. Tal y como se mencionó a lo largo de la discusión teórica, esta combinación de población, tiempo y espacio son las coordenadas demográficas a las cuales hace referencia Almanza (1988), Canales y Lerner (2003) y es el ideal bajo el cual se deberían de construir los estimadores poblacionales.

En el cuadro siguiente se muestran los resultados de las estimaciones para cada uno de los municipios seleccionados, en la primera columna se consigna el valor inicial que se tenía en 2015 y en la segunda columna se refiere el valor estimado para 2020.

Cuadro 1 Resultados del modelo

Municipio	Estimación 2020 con el filtro de Kalman
Aguascalientes	1 039 577
Asientos	58 596
Calvillo	69 183
Cosío	19 874

Jesús María	150 615
Pabellón de Arteaga	55 756
Rincón de Romos	67 916
San José de Gracia	11 049
Tepezalá	25 410
El Llano	25 332
San Francisco de los Romo	59 976
Abasolo	111 781
Acámbaro	138 159
San Miguel de Allende	218 535
Apaseo el Alto	80 103
Apaseo el Grande	114 859
Celaya	594 085
Manuel Doblado	45 837
Comonfort	101 644
Coroneo	15 124
Cortazar	120 801
Cuerámbaro	34 802
Doctor Mora	32 464
Dolores Hidalgo Cuna de la Independencia Nacional	187 957
Guanajuato	226 908
Huanímaro	26 691
Irapuato	715 871
Jaral del Progreso	47 464
Jerécuaro	59 206
León	1 876 137
Moroleón	59 190
Ocampo	28 589
Pénjamo	186 028

Pueblo Nuevo	15 405
Purísima del Rincón	98 873
Romita	73 886
Salamanca	342 723
Salvatierra	126 307
San Diego de la Unión	49 327
San Felipe	140 176
San Francisco del Rincón	143 137
San José Iturbide	101 087
San Luis de la Paz	149 568
Santa Catarina	6 845
Santa Cruz de Juventino Rosas	104 407
Santiago Maravatío	8 707
Silao de la Victoria	240 185
Tarandacuao	15 483
Tarimoro	43 062
Tierra Blanca	24 567
Uriangato	76 280
Valle de Santiago	175 571
Victoria	26 609
Villagrán	76 871
Xichú	14 924
Yuriria	87 154
Amealco de Bonfil	84 376
Pinal de Amoles	35 214
Arroyo Seco	18 942
Cadereyta de Montes	85 904
Colón	83 825
Corregidora	222 605

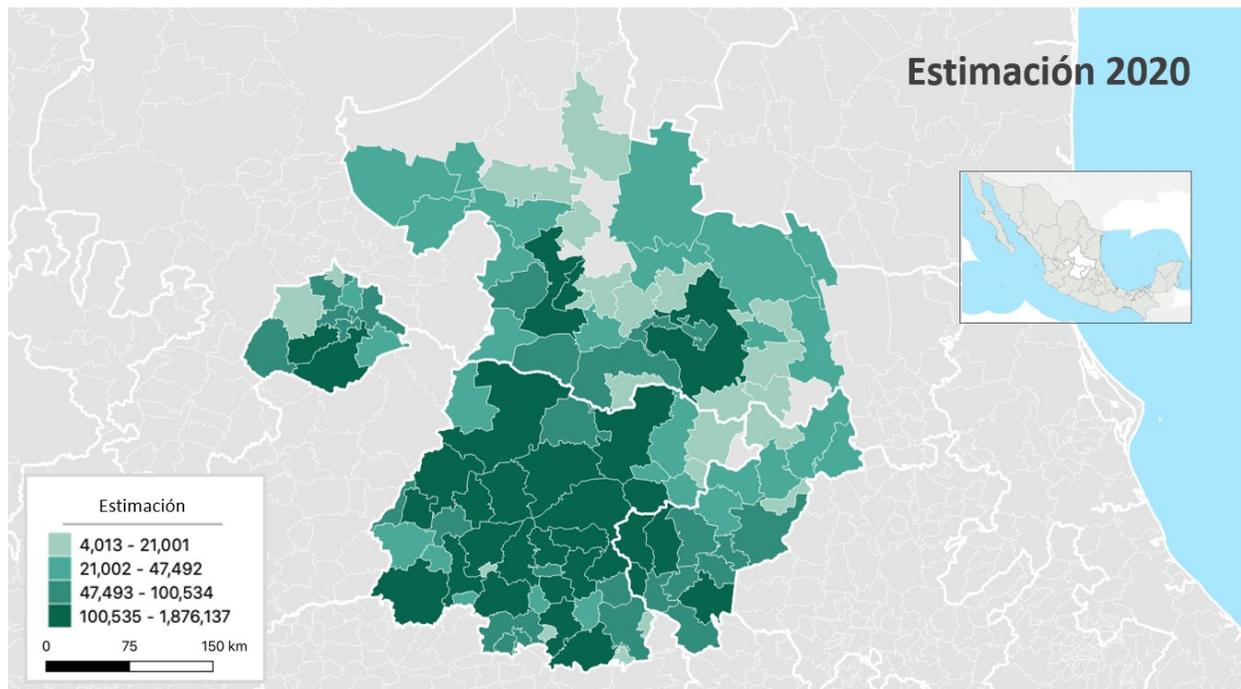
Ezequiel Montes	52 401
Huimilpan	48 762
Jalpan de Serra	34 169
Landa de Matamoros	23 015
El Marqués	197 221
Pedro Escobedo	90 369
Peñamiller	24 186
Querétaro	1 029 235
San Joaquín	12 164
San Juan del Río	341 835
Tequisquiapan	90 212
Tolimán	34 807
Ahualulco	22 686
Alaquines	10 307
Armadillo de los Infante	4 817
Cárdenas	22 748
Cerritos	26 014
Cerro de San Pedro	5 992
Ciudad del Maíz	40 634
Ciudad Fernández	62 976
Guadalcázar	31 863
Lagunillas	7 384
Mexquitic de Carmona	74 894
Moctezuma	24 059
Rayón	20 802
Rioverde	123 342
Salinas	36 658
San Ciro de Acosta	12 917
San Luis Potosí	967 891
San Nicolás Tolentino	6 576

Santa María del Río	47 520
Soledad de Graciano Sánchez	363 395
Tamasopo	40 390
Tierra Nueva	11 019
Venado	17 627
Villa de Arriaga	21 721
Villa de Guadalupe	11 234
Villa de Ramos	45 853
Villa de Reyes	63 273
Villa Juárez	13 843
Zaragoza	33 017
Villa de Arista	18 498
El Naranjo	26 686

Fuente: elaboración propia

Para tener una mejor idea de los resultados, en el Mapa 4 que se ofrece a continuación, se grafican los resultados y se observa que en los municipios de León, Aguascalientes y San Luis Potosí es en donde se concentra la mayor cantidad de personas y fueron también los que tuvieron un mayor crecimiento al comparar las cifras de 2015 y la estimación de 2020.

Mapa 4 Estimación 2020



Fuente: elaboración propia.

7.3 Comparación de las estimaciones

En la sección pasada se mostraron los resultados de la estimación a 2020, en la que, como ya se mencionó, el estimador combina información de las personas, información espacial e información temporal, las tres dimensiones teóricas que dan sustento a lo que se denomina como *población*. En esta sección se muestra la importancia de que se lleve a cabo el proceso de validación en las estimaciones, mediante el que se logra tener una idea de que tan bueno o malo fue el desempeño del modelo propuesto.

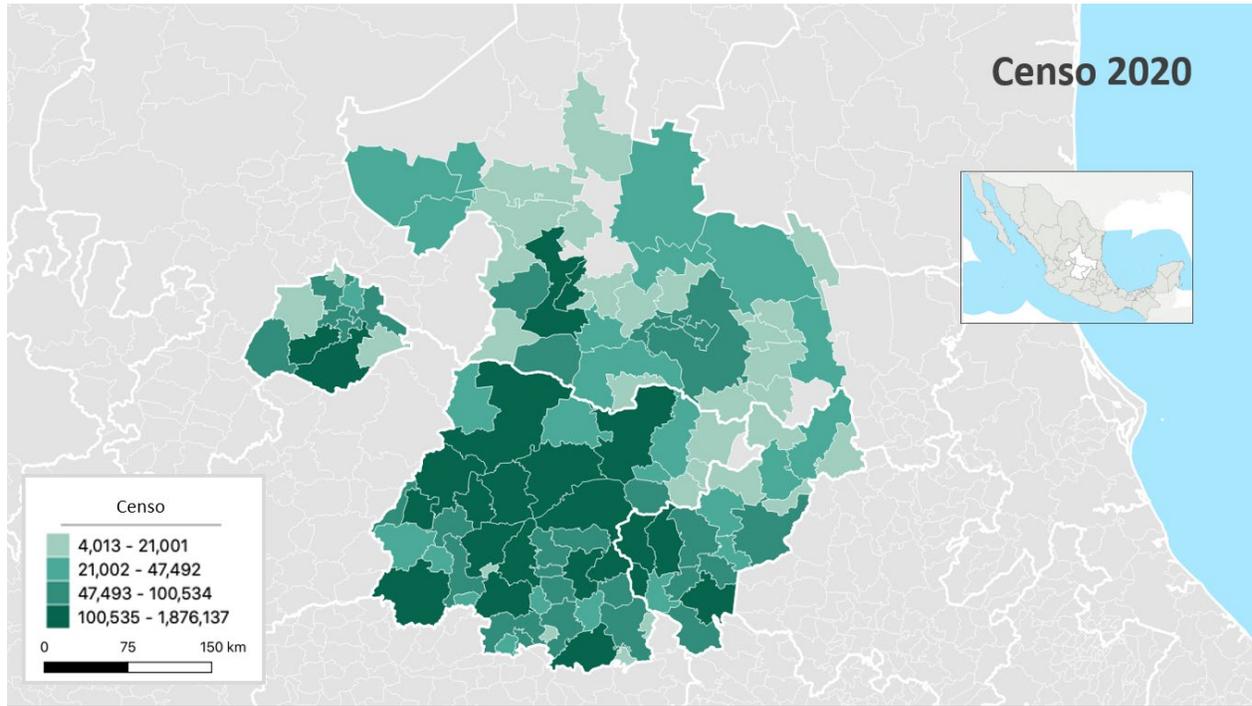
7.3.1 Con los datos censales

Para llevar a cabo el cotejo de las cifras se utilizan los datos del Censo de Población y Vivienda de 2020, el proyecto estadístico más importante desarrollado por los países para conocer la distribución y el tamaño de su población. Una de sus características es que los datos que arroja en el levantamiento censal son considerados de interés nacional y constituyen la referencia que emplean los gobiernos estatales y el Gobierno federal para el desarrollo de las políticas públicas. En México, los levantamientos censales se llevan a cabo cada 10 años en cumplimiento de las recomendaciones internacionales que sugieren contar con información actualizada de manera regular.

A diferencia de lo que ocurrió en los levantamientos censales anteriores, en 2020 el mundo enfrentó la pandemia por COVID-19. Esta situación provocó el aumento de la no respuesta porque los informantes se negaron a responder el Censo por temor a contagiarse y a pesar de que en algún momento se tuvo la propuesta de interrumpir o postergar el Censo, las autoridades decidieron continuar con el operativo. Estos factores son de suma importancia, pues el hecho de que su levantamiento coincidió con el inicio de la pandemia supone que los problemas se incrementaron. En la última sección de este capítulo se lleva a cabo una reflexión en torno al Censo, por el momento conviene tenerlo como referencia para comparar los resultados de las estimaciones y las cifras oficiales.

En el Mapa 5 se muestran los resultados oficiales del Censo de Población y Vivienda 2020 y en términos generales, se puede observar que ambos mapas tienen el mismo comportamiento en donde la mayor concentración de población se encuentra en los municipios ubicados en la región sureste. Esta región corresponde a los municipios que se encuentran en la zona de Guanajuato, en donde hay un gran desarrollo industrial que se ha venido potenciando en los últimos años y que forma parte del crecimiento de toda esta zona geográfica.

Mapa 5 Datos censales



Fuente: elaboración propia.

Para comparar las estimaciones que se obtuvieron con base en el modelo y los datos censales se obtiene un indicador de la diferencia que se configura de la siguiente forma:

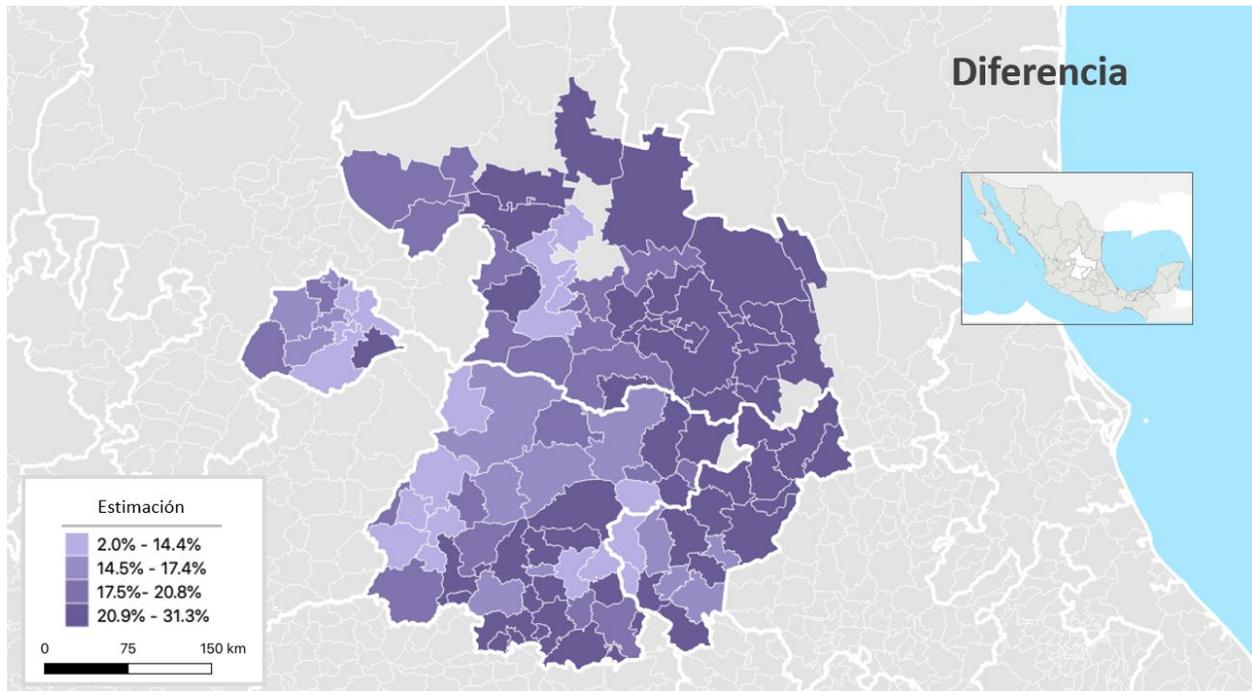
$$d_k = \frac{|T_{k,CENSO} - T_{k,Estimación2020}|}{T_{k,CENSO}}$$

En donde $T_{k,CENSO}$ es el total de población del municipio k de acuerdo con los datos censales y $T_{k,Estimación2020}$ son los resultados del modelo. Resulta importante mencionar que el dato de la diferencia se toma en valor absoluto porque existe la posibilidad de que en algunos casos sea negativo y en otros positivos, sin embargo, la intención es poner el foco de atención en la magnitud de esta diferencia y así obtener una medida porcentual que sirva como referencia para evaluar el desempeño de las estimaciones.

Para cerrar este apartado, en el siguiente mapa se muestra la distribución de d_k , los colores oscuros representan aquellas zonas en donde la diferencia entre un estimador y otro es mayor, mientras que en las zonas tenues sugieren que los valores son muy parecidos y cercanos a cero. Al

hacer esta comparación (ver Mapa 6) se puede observar que las mayores diferencias se encuentran en aquellos municipios en donde se tiene una menor cantidad de población.

Mapa 6 Diferencia entre las estimaciones y los datos censales



Fuente: elaboración propia.

7.3.2 Error Porcentual Absoluto Medio

En la sección anterior se llevó a cabo una comparación de los resultados que se obtuvieron con las estimaciones y los datos censales, para ello se obtuvo un indicador que mide la diferencia porcentual que existe entre ambos valores, sin embargo, con la intención de tener una medida del desempeño global se calculó un indicador que ofrece una medición del error absoluto en términos porcentuales. La lógica es que entre más pequeño es su valor, se tiene un mejor desempeño de los datos, ya que la variación entre lo que se asume como el dato real y lo estimado varía poco. La expresión para calcular el *MAPE* es la siguiente:

$$MAPE = \frac{100\%}{n} \sum_{t=1}^n \left| \frac{C_t - E_t}{C_t} \right|$$

Donde C_t es el valor que se asume como real y que en este caso se compone de los datos del Censo de Población y Vivienda 2020; E_t representa a las estimaciones que se obtuvieron con la aplicación del modelo para cada uno de los m municipios analizados. En términos generales, este indicador compara, uno a uno, la variación que se tiene entre el dato real y el estimado para que se obtenga un promedio y genere una medida global.

A partir de este estimador se generó un valor de **21 %**, lo que significa que las estimaciones tienen una variación de 21 % con respecto al dato real. Sin embargo, al examinar con mayor detalle el comportamiento y las características de los municipios que se están analizando existen algunas situaciones que se deben de tomar en cuenta y que a continuación se discuten:

1. No existe un criterio unánime para evaluar el desempeño de los datos, ni mucho menos en las estimaciones que se generan en el ámbito socio demográfico. En los últimos años, el Inegi ha hecho la sugerencia de que todos aquellos datos que provienen de encuestas son susceptibles de presentar errores debido a la cobertura, debido a lo cual se estima el coeficiente de variación para tener una medida de cuánto oscilan las estimaciones. Con este criterio se considera que una oscilación menor a 15 % quiere decir que los datos tienen una muy buena calidad, si se encuentran entre 15 % y 25 % quiere decir que la calidad es buena y si supera 25 % quiere decir que su desempeño es deficiente. Si se utilizan estos criterios de las encuestas para evaluar las estimaciones que se llevaron a cabo durante este proyecto de investigación, entonces la variación porcentual de 21 % ubicaría a las estimaciones en el rubro de buenas.
2. De los 105 municipios que fueron analizados, la mediana de la población es de 40 000 habitantes, que si se considera que en promedio una vivienda lo constituyen cuatro personas, entonces se tienen alrededor de 10 000 viviendas habitadas y, de acuerdo con los criterios con los que se levantó la muestra censal, entrarían dentro del rubro de municipios seleccionados con certeza o lo que es lo mismo aquellos en donde el tamaño de la población

los hace susceptibles de ser analizados con mayor precisión porque son los lugares en donde se tiene una menor información.

3. Para llevar a cabo estas estimaciones están tomando como referencia la información de la Encuesta Intercensal 2015 y las imágenes satelitales que mide la cobertura urbana y, con base en ellas, se lleva a cabo la estimación de población. Esto significa que el modelo tiene una exactitud de 80 % para predecir el total de población cuando su fuente principal no son datos poblacionales sino geospaciales. Lo que significa que se puede tener un acercamiento a las estimaciones de población de manera indirecta, sin la necesidad de depender de las fuentes sociodemográficas tradicionales como Censo de población o las encuestas.
4. Además, esta forma de estimar los datos es una nueva referencia para discutir con las proyecciones de población, las cuales sirven para planear los programas sociales que se llevan a cabo en México. Hasta este punto no existe otra fuente de información que genere datos a nivel municipal a través del tiempo, este es uno de los primeros ejercicios que se planteó como objetivo tener datos espaciales y temporales. Por este motivo representan una oportunidad para evaluar las proyecciones de población.

7.3.3 Comparación temporal (proyecciones de población)

En la sección anterior se llevó a cabo una evaluación de las estimaciones que se obtuvieron con base en las imágenes satelitales y el filtro de Kalman utilizando los datos del censo población y vivienda 2020. La razón para llevar a cabo este ejercicio de validación es si los datos censales tienen una mayor precisión en comparación con las demás fuentes de información y son las referencias para el desarrollo de los planes y políticas públicas que se llevan a cabo en México. Sin embargo, estos datos ofrecen una fotografía de la población a nivel nacional, de hecho, se tomó como referencia el 3 de marzo de 2020 como la fecha de referencia para llevar a cabo los cálculos. Las estimaciones que se desarrollaron en esta investigación tienen una perspectiva longitudinal,

por lo que no solamente se tiene el dato de 2020, también se cuenta con información para 2016, 2017, 2018 y 2019. Estas cuatro observaciones permiten analizar cómo evoluciona la población a lo largo del tiempo.

Para llevar a cabo una evaluación de los datos a nivel temporal se recurre a las proyecciones de población que genera el CONAPO. Al igual que los datos censales, las proyecciones de población son los datos oficiales que se utilizan para llevar a cabo un seguimiento a través del tiempo de la evolución poblacional en los diferentes municipios y son el resultado de una estimación matemática y lo que se denomina la conciliación demográfica, la cual consiste en la revisión hecha por expertos en cada una de las áreas (migración, fecundidad) lo que permite ajustar las cifras. De esta forma, las proyecciones permiten dar un seguimiento a lo largo del tiempo y se toman como referencia para llevar a cabo una comparación con las estimaciones que se obtuvieron en esta investigación.

Si se toma en cuenta el número de municipios y los años para los cuales estimó el total de población, entonces tienes alrededor de 500 datos estimados, conseguidos de esta manera, para llevar a cabo una comparación con las proyecciones de población se obtiene el **error porcentual absoluto medio (MAPE)** para cada uno de los años y esto se observa en el cuadro siguiente.

Cuadro 2 Validación longitudinal de las estimaciones

2016	2017	2018	2019
4.1 %	9.0 %	12.4 %	14.9 %

Fuente: elaboración propia

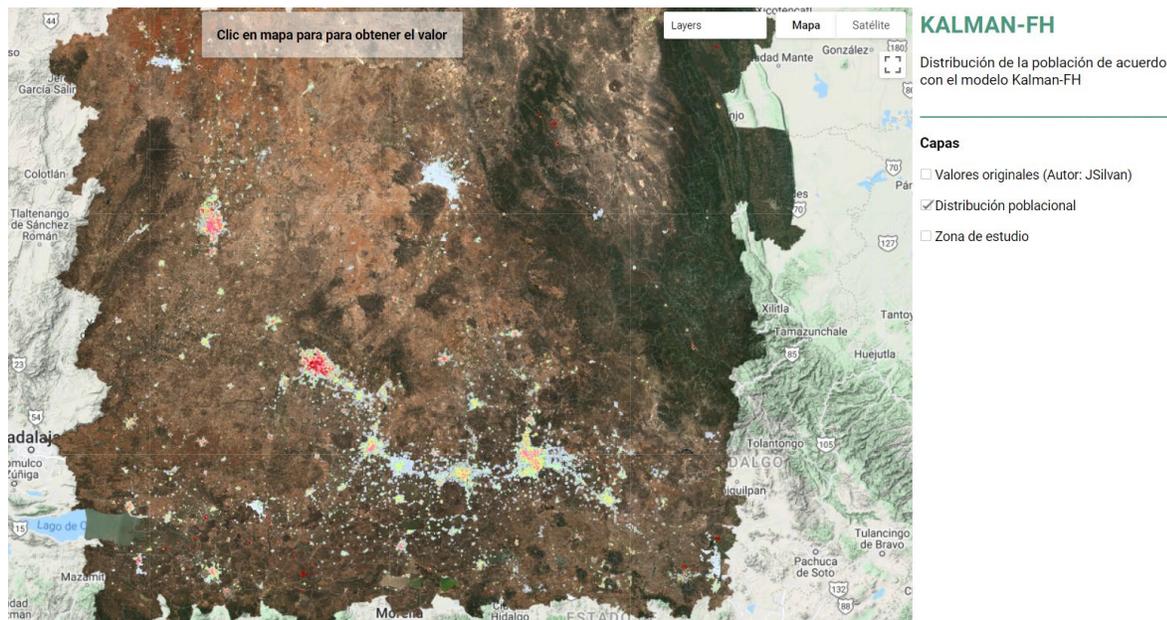
En 2016, los resultados de las estimaciones hechas con el filtro de Kalman y las proyecciones de población son muy parecidas y esto se refleja en el valor de 4.1 %. Aunque a medida que transcurren los años su valor se incrementa hasta llegar a 14.9 %. Tal y como se mencionaba en la sección anterior, no existe un criterio unánime para determinar qué porcentaje es adecuado y a partir de qué rango las observaciones son deficientes. Sin embargo, al igual que se planteó en la sección pasada, se puede tomar como referencia el hecho de que un indicador que tenga una variación menor a 15 % se considera como un dato de buena calidad.

Bajo este criterio, las cuatro estimaciones posteriores a la fecha de referencia (2015) tienen un comportamiento adecuado y serían consideradas como confiables. Aunque no se debe perder de vista que las proyecciones de población solo consideran datos sociodemográficos, pero no incluyen información referente a la distribución de la población en el territorio, mientras que las estimaciones que se proponen en esta investigación toman la información demográfica y la combinan con datos geoespaciales. Esta combinación es un cambio de paradigma, ya que incluye la información de las personas, pero también los datos que se captan de manera indirecta, porque en las fotografías aéreas no involucran directamente a las personas, pero se toman en cuenta cómo se ubican en el espacio y los cambios que sufre este a través de los años.

7.4 Visualización de los resultados

Una vez que se tienen las estimaciones de cada uno de los municipios, el siguiente paso es mostrarlas en un mapa. A diferencia de los casos anteriores En donde el municipio entero se colorea de acuerdo con su valor, acá se propone hacer una distribución del total de la población en cada uno de los píxeles. Con esta idea se trata de enfatizar que la distribución poblacional no está sujeta a las delimitaciones político-administrativas de cada una de estas regiones, ya que existen zonas en donde se concentra la mayor parte de la población, mientras que en otras zonas el número de personas es escasa. Para llevar a cabo este procedimiento se utiliza la plataforma de Google Earth Engine en donde se generó un visualizador interactivo (ver Mapa 7) que permite a los usuarios seleccionar las distintas capas en la superficie terrestre:

Mapa 7 Distribución de la población de acuerdo con los resultados del filtro de Kalman



Fuente: elaboración propia.

Herramienta digital disponible en <https://julioccolmex.users.earthengine.app/view/app-1>

En este capítulo se presentaron los resultados de las estimaciones para los 105 municipios que forman parte del área de estudio. Dado que se trata de un cálculo que se llevó a cabo con base en la información de la encuesta intercensal 2015 y las imágenes satelitales, se llevó a cabo un proceso de validación de los datos con base en la información del Censo 2020. Sin embargo, dado el antecedente del levantamiento durante la COVID-19, se compararon los datos censales y las estimaciones por medio del cálculo del **error porcentual absoluto medio (MAPE)**, que compara las dos cifras para cada uno de los municipios y arroja una medida que sirve como referencia para medir la calidad de las estimaciones (se obtuvo un valor de 21 %). Dado que no existe un criterio unánime para saber si este valor es alto bajo, se toma como referencia a los rangos que utiliza el Inegi cuando genera las estimaciones sociodemográficas y de acuerdo con este criterio la calidad de las estimaciones es buena.

Por otro lado, estas estimaciones se generaron anualmente entre 2016 y 2020, por lo que se llevó a cabo un ejercicio similar al anterior, pero ahora comparando las cifras con los resultados de las proyecciones de población que genera el CONAPO. Al igual que los valores que se obtienen en esta investigación, las cifras de las proyecciones son cálculos obtenidos a partir de incorporar

diversas fuentes de información sociodemográfica por lo que son susceptibles de tener algún grado de error. Por este motivo, se obtuvo el *MAPE* de 2016 a 2019 y así se pudo observar que la variación de todos estos años es menor a 15 %, lo cual como ya se mencionó pediría dentro del rubro de estimaciones confiables.

Algunas preguntas importantes para esta investigación son ¿para qué obtener el total de población?, ¿cuál es la utilidad de hacer un ejercicio de estimación? En el contexto construido hasta el momento, una estimación de este estilo no únicamente es útil para conocer la cantidad de personas que viven en un lugar determinado, sino que sirve de referencia como información auxiliar que puede ser utilizada para ajustar los factores de expansión de las encuestas. Este proceso se conoce formalmente como calibración de los ponderadores y consiste en tomar como referencia los datos obtenidos en los levantamientos en campo y expandir la información de tal modo que se tenga una representatividad estadística.

Para llevar a cabo este proceso existen diversas técnicas, sin embargo, en todas ellas es necesario contar con información auxiliar que permita llevar a cabo este proceso de calibración de los datos. De acuerdo con el Inegi (2020), uno de los datos más empleados son los totales de población los cuales se retoman a partir las proyecciones poblacionales que genera CONAPO. Con esta información se logra que las encuestas reflejen el comportamiento real de la población, ya que los datos que se obtienen en campo no son totalmente fiables, sino que tienen que pasar por un proceso de depuración para garantizar dicha representatividad.

Por este motivo, es muy importante tener información confiable y que sirva de referencia para llevar a cabo este proceso de calibración de las encuestas, ya que en ellas se pregunta información específica y se levantan datos sociodemográficos de toda la población. Esto significa que un cálculo acertado va más allá de solo conocer el total de población, sino que representa un insumo que puede ser utilizado para estimar otro tipo de indicadores.

Finalmente, un elemento que se debe de tomar en cuenta es que con la irrupción del COVID-19 se debe plantear la necesidad de revisar no solamente las cifras censales sino también las proyecciones de población para los siguientes años. El incremento en la tasa de no respuesta afectó las cifras del censo, pero se desconoce en qué magnitud, sin embargo, estas diferencias e inconsistencias ya están teniendo un impacto en el desarrollo de las políticas públicas y esto se puede observar con la inconformidad que manifestaron algunas entidades federativas en torno a los totales de población de la ENOE, los cuales fueron ajustados con base en el censo de población

2020. Al momento no existe un ejercicio que permita saber cómo afectó el inicio de la pandemia con el levantamiento censal, por este motivo este ejercicio en donde se analiza a la población de manera indirecta, mediante imágenes satelitales, cobra mayor importancia porque enfrenta los problemas que se tienen en la recopilación de los datos cara a cara.

Reflexiones finales

El término *población* es un constructo teórico que evolucionó a lo largo del tiempo y que se utiliza para hacer referencia a un conjunto de personas que viven en un lugar determinado. De acuerdo con Canales y Lerner (2003), una de las cualidades que tiene ese constructo teórico es que elimina las diferencias que existen entre las personas, lo que significa que suprime los rasgos individuales y privilegia las características grupales. Es decir, los individuos solamente adquieren notoriedad siempre y cuando formen parte de un grupo más grande; aunque en primera instancia esto puede ser considerado como menosprecio por los individuos, Vallin (1991) afirma que esta unificación se debe a que permite el desarrollo de los seres humanos, pues solamente el conjunto garantiza la supervivencia y tienen la capacidad de modificar el territorio.

Sin embargo, este constructo que hace referencia a un conjunto de personas es insuficiente para estudiar su comportamiento. De acuerdo con Vieira (1973), para estudiar a los grupos humanos es necesario tomar en cuenta dos elementos adicionales que sirven de coordenadas: el espacio y el tiempo. El tiempo es uno de los elementos que más se utiliza en el análisis demográfico, ya que permite hacer evidentes los cambios que surgen en los diferentes fenómenos que se estudien. Además, la teoría de la transición demográfica sostiene que existe una modificación en las tasas de fecundidad y mortalidad de los países, las cuales presentan una disminución en la medida en que las diferentes naciones experimentan un proceso de modernización (Zavala, 1992). Por otro lado, a diferencia del tiempo, el espacio ha quedado en segundo término dentro de las investigaciones demográficas porque solo se había interpretado como una proporción en la superficie terrestre, que sirve de contenedor (Vieira, 1973). Por lo que, al incorporar el componente geográfico en las investigaciones se pretende una mayor comprensión de la difusión de los fenómenos a lo largo del territorio. Uno de los ejemplos más usados para hacer evidente su impacto es el análisis del comportamiento reproductivo de la población (Weeks, 2016).

De hecho, al igual que otros fenómenos sociales, los fenómenos demográficos siguen el comportamiento de la primera ley de Tobler. En ella se establece que todas las cosas están relacionadas entre sí, pero las que se encuentran más próximas tienen una relación mayor que aquellas que se encuentran distantes (Waters, 2018). Si se toma como referencia esta idea, entonces se pueden entender algunos de los patrones que se describen en la teoría de la transición demográfica (Weeks, 2001, 2016). Por ejemplo, se explica que la reducción de la fecundidad se

dio primero en las zonas urbanas y después se fue extendiendo a otras regiones que se encuentran más alejadas. Si se analiza con detenimiento este comportamiento desde una perspectiva espacial, se puede afirmar que se trata de un efecto de contagio en donde los fenómenos empiezan en un punto en particular y después se van extendiendo a lo largo del territorio (Bell, 2015). Por este motivo, se sugiere que la teoría de la transición demográfica no únicamente debe de hacer referencia a los cambios que surgen durante el transcurso del tiempo sino también a la forma en la cual se generan dichos cambios.

Sin embargo, en las investigaciones, el hecho de privilegiar el enfoque temporal también se observa en la parte metodológica. La ecuación compensadora, que es el referente para analizar el comportamiento de la población, toma en cuenta el cambio que ocurre de un momento a otro, pero deja de lado la ubicación espacial de los individuos (Bocquet y Jakobi, 1996). A pesar de que el componente migratorio forma parte de esta expresión matemática, se privilegia la movilidad de las personas, pero se descarta hacia dónde se mueven.

Conviene mencionar, además, que en las investigaciones demográficas se suele ignorar la escala con la cual se colecta la información. De hecho, Brown y Guinnane (2007) mencionan que el proyecto llevado a cabo por Coale (1986), en donde se buscó dar sustento empírico a la teoría de la transición demográfica que fue planteada por Notestein, tiene un sesgo debido a la escala con la cual fueron conectados los datos. De manera más específica, los autores argumentan que las mediciones de la fecundidad que se llevaron a cabo son altamente sensibles a los niveles de agregación y sobre todo a las combinaciones de las unidades político-administrativas que se tomaron como referencia. En consecuencia, existe una tendencia decreciente debido al efecto que tienen las escalas; pero si se miden con otro nivel o se toman otras escalas, el comportamiento puede pasar inadvertido o ser contrario al que inicialmente se había planteado. En otras disciplinas, como la Geografía, este problema ha sido estudiado con mayor profundidad y se denota como el problema del área modificable, este plantea que los resultados son afectados por las delimitaciones de las áreas de estudio (Flowerdew, 2011; Vela, 2016; Wong, 2004).

Para visualizar a poblaciones que pasaban inadvertidas debido a que estaban inmersas en un conjunto de mayor tamaño y quedaban invisibilizadas, actualmente existe una demanda de datos para incorporar información con un mayor nivel de detalle y la precisión que permita abordar estos fenómenos (Baussola y Fabrizi, 2008; Bryant y Graham, 2013; Molina, 2019). Este deseo de obtener información granular trae consigo otro debate: la utilidad de los datos. Por un lado, hay

posturas que demandan más apertura y mayor nivel de especificidad; por el otro, hay autores como Garfinkel *et al.* (2018) que advierten sobre los peligros de hacer públicos los datos con un alto grado de desagregación. Por ejemplo, en el Censo de Estados Unidos de América se utilizó lo que se denomina privacidad diferenciada, es decir, un algoritmo que busca proteger la identidad de las personas induciendo ruido a los datos (Garfinkel y Leclerc, 2020). De esta forma, se logra tener y analizar un comportamiento grupal, pero se protege la identidad de las personas evitando la identificación de los individuos (Garfinkel *et al.*, 2018). De cierta manera, esto va en sincronía con el análisis demográfico en donde se omite el análisis individual y se privilegia el grupal (Vieira, 1973).

La estimación de población en áreas pequeñas es uno de los esfuerzos metodológicos más importantes de las últimas décadas (Ghosh y Rao, 1994; Pfeffermann, 2013; Rao, 2008; Rao y Molina, 2015). Una de las razones que motiva el desarrollo de nuevas técnicas es que los registros que contabilizan a la población son falibles debido a que los individuos se encuentran en constante movimiento (Lomax y Smith, 2017). Esta situación ha sido reconocida por los organismos que colectan información a nivel nacional, por lo que la gran mayoría de las cifras que se ofrecen son estimaciones que se hacen con un cierto nivel de confianza (Alvarez, 2001).

Las estimaciones en áreas pequeñas se pueden llevar a cabo desde dos perspectivas: 1) el enfoque demográfico (Bryant y Graham, 2013; Swanson, 2012) y 2) el enfoque estadístico (Rao y Molina, 2015). Si bien es cierto que en principio estos enfoques pertenecen a áreas diferentes, el hecho de que la Demografía tenga un diálogo constante con otras disciplinas da lugar a que se puedan compartir métodos y las generalizaciones que se llevan a cabo resultan válidas tanto en el campo demográfico como en el estadístico. En ambos casos se habla de “nivel de precisión”, tanto para los demógrafos como para los estadísticos es claro que este término hace referencia a cierto grado de incertidumbre que algunas veces es imposible de cuantificar, pero que denota que los valores no corresponden a las cifras exactas, sino que son aproximaciones.

La propuesta metodológica que se desarrolla en esta tesis plantea combinar las fuentes de datos tradicionales con las imágenes satelitales. Ambas están vinculadas por medio de su ubicación espacial, la variable clave que permite hacer una articulación entre los archivos raster y las bases de datos que se conectan de manera tradicional. En el escenario ideal, los límites político-administrativos responden al comportamiento de un fenómeno, en este caso el de la densidad de población, sin embargo, existe una dependencia en la forma en cómo se conecta la información

(Ruggles *et al.*, 2003). Si bien es cierto que esta dependencia es una limitante importante que tiene un efecto en los resultados, algunos otros autores como Kugler *et al.* (2015) advierten que es la única manera para poder hacer una integración de los datos.

El área de estudio corresponde a 105 municipios que se ubican en la región central del país, a este grupo se le conoce como el Corredor Metropolitano Centro País, el cual es de gran importancia comercial y es una de las zonas económicas de mayor crecimiento en el país. Además, en términos sociodemográficos, tiene la característica de que confluyen municipios en donde la población es de más de un millón de personas y otros en donde no superan los cien mil habitantes. Esta característica permite tener diversos escenarios en donde se puedan llevar a cabo el análisis.

Ahora bien, para integrar este análisis existe una herramienta que en la estadística se conoce como “filtro de Kalman”: un algoritmo matemático que sirve para modelar el comportamiento de los fenómenos a lo largo del tiempo. Este algoritmo tiene una característica que lo hace particularmente útil, permite una combinación entre los valores observados y los valores predichos (Chui y Chen, 2017). Lo que significa que tiene una gran flexibilidad y permite incorporar información que resulta de un cálculo de las observaciones reales, o bien, que se asumen como reales. Este algoritmo iterativo va ajustando el comportamiento del fenómeno a lo largo del tiempo y, aunque, se desarrolló en primera instancia para calcular fenómenos como el vuelo de los aviones, su gran capacidad y flexibilidad para combinar las observaciones lo ha hecho útil y aplicable en otras áreas como la Demografía y la Estadística (Ordorica-Mellado *et al.*, 2019).

Este modelo genera un estimador, es decir, una aproximación, un dato real. Aunque en primera instancia esto puede parecer poco efectivo, lo cierto es que ni en la demografía ni en los estudios de población se tolera que la información tenga un determinado nivel de confianza y resulta más importante determinar el comportamiento que siguen los fenómenos. Esto significa que es de mayor utilidad si existe un cambio creciente o decreciente, aún y cuando los valores específicos tengan una cierta variación (García Guerrero y Ordorica Mellado, 2012). Por ejemplo, cuando se generan los resultados de las encuestas se suelen hacer pruebas estadísticas para ver si la variación que existe entre los valores de años consecutivos resulta ser significativa o simplemente se trata de un comportamiento derivado de la forma en cómo fue captada la información.

Tal y como se mencionó anteriormente, existen ventajas de integrar en los estudios de población las imágenes satelitales debido a que son fuentes de información que se obtienen con

mayor periodicidad y, al igual que las fuentes de información tradicionales, abarcan diferentes temas (Gutman *et al.*, 2013; Masek *et al.*, 2020). Si bien es cierto que la lógica de todas ellas es la de codificar la información de la superficie terrestre en píxeles, dicha información se puede utilizar para generar distintas capas temáticas (Lillesand *et al.*, 2015).

Para llevar a cabo este ejercicio, se usaron las máscaras de cobertura urbana que desarrollaron los especialistas del Centro Geo. Este grupo de investigación tomó como referencia las imágenes satelitales Landsat 8 e hicieron una clasificación de los píxeles con la técnica denominada “desmezclado espectral lineal” que consiste en tomar como referencia las firmas espectrales de los elementos que se encuentran en la superficie terrestre y con esta información llevar a cabo una clasificación para saber si se trata de una zona en construcción, un cuerpo de agua o vegetación.

Ahora bien, si hacemos una analogía de este tipo de datos con la forma en cómo se levanta la información demográfica tendríamos lo siguiente: a) cada uno de los objetos que se encuentran en el territorio equivaldrían a las unidades de análisis, b) las diferentes bandas del espectro corresponderían a las variables y c) los valores que se obtienen en cada una de ellas corresponderían a los valores que tiene cada una de dichas variables. Bajo esta lógica se tienen varias unidades de análisis que corresponden a una misma clase, pero que poseen algunas diferencias o particularidades en el comportamiento de sus variables, por lo tanto, el objetivo es hacer una validación para determinar si, pese a las ligeras variaciones que tienen, su comportamiento es el mismo. Retomando la analogía de cómo se configuran las bases de datos, esto sería equivalente a cuando se genera un estimador y se establece un intervalo de confianza en el cual oscila. Se asume que el dato puede oscilar, pero si se encuentra dentro de dicho rango, entonces el comportamiento es adecuado.

Principales resultados de la tesis

A continuación se escriben las principales contribuciones de esta investigación:

1. Desde el punto de vista teórico se contribuye con la discusión en torno al sesgo que existe del concepto *población*, el cual únicamente toma como referencia el tiempo. Si bien es cierto que este es uno de los ejes que guía el quehacer socio demográfico, la dimensión espacial ha pasado a segundo término y debe de recuperarse como un eje fundamental de análisis. Al hacerlo, trae consigo algunas preguntas, por ejemplo, si el nivel de

desagregación es el adecuado para llevar a cabo los análisis socio demográficos. Además, el hecho de fijar la atención en poblaciones pequeñas trae consigo la reflexión en torno a la utilidad y disponibilidad de los datos. Esto significa retomar algunas preguntas sobre para que usar la información y no solamente fijarse en si es técnicamente posible.

2. Desde el punto de vista metodológico, se discutió sobre la manera de utilizar las imágenes satelitales e incorporarlas al quehacer socio demográfico. Para lograr esta unión se requiere de un elemento en común: las delimitaciones político-administrativas, las cuales tienen un impacto en la forma en que se captan los fenómenos, pero pocas veces se discute su efecto en las estimaciones.
3. El estimador que se obtiene con el filtro de Kalman es un indicador mixto: por un lado, incluye la dimensión geográfica; por el otro la esfera sociodemográfica. Este indicador refleja de manera empírica la unión teórica del tiempo y el espacio como ejes coordinados para estudiar a la población.
4. En cuanto a las estimaciones que se obtuvieron con el indicador anteriormente mencionado, se encontró que su nivel de confiabilidad es bueno, bajo los parámetros que considera el Inegi para dar certeza a la variabilidad de los datos. Esto significa que este modelo es una alternativa para llevar a cabo estimaciones en los próximos años sin la dependencia directa del Censo u otra fuente de información socio demográfica.
5. Para validar el estimador se usaron los datos del Censo de Población y Vivienda 2020. Sin embargo, al llevar a cabo este ejercicio se pudieron constatar algunas de las fallas que tiene el censo y que estarán impactando en las decisiones de política pública y las investigaciones que se lleven a cabo en los próximos años.

Limitaciones de la tesis

Debido a que a lo largo el desarrollo de esta tesis se encontraron algunas limitaciones relacionadas con las fuentes de información disponibles, a continuación se describen:

1. Se tomaron como referencia los datos en la Encuesta Intercensal 2015 y con ellos se llevó a cabo una proyección de los datos a 2020. Sin embargo, se debe de reconocer que esta encuesta también genera estimaciones de población por lo que es probable que estos efectos se trasladarán el cálculo que se hizo con el filtro de Kalman a pesar de que este último toma en cuenta el error asociado a dichas estimaciones.

2. Las capas que se utilizaron para el análisis fueron desarrolladas por los especialistas del Centro de Investigación en Ciencias de Información Geoespacial, A.C. para el Corredor Centro Urbano. Al momento de analizar los datos se pudo observar que, por la forma en que se procesaron las imágenes, existen zonas que corresponden a carreteras e infraestructura en donde no necesariamente existen asentamientos humanos. Por lo que es factible suponer que existe un cierto error al momento de hacer la predicción de los datos, sin embargo, hacer esta diferenciación requiere de imágenes con una mayor resolución las cuales se debe de pagar una cuota para poder generarlas.
3. De acuerdo con la evaluación de las estimaciones se observa que las predicciones que tienen una variabilidad menor son aquellas que corresponden en los primeros dos años. Luego de este tiempo los datos tienen una variación mayor y con ello se reduce su confiabilidad. Por ejemplo, el momento de hacer las validaciones se observó que en 2016 hubo una separación de 4 %, en 2017 fue de 9 % y en 2020 fue de 21 %.
4. A lo largo de esta investigación se exploraron diversas delimitaciones geográficas e incluso al inicio de la tesis se propuso generar delimitaciones propias. Para llevar a cabo un ejercicio de esta naturaleza fue necesario utilizar aquellas delimitaciones político-administrativas ya establecidas, ya que son las que cuentan con información sociodemográfica. Sin embargo, desde el punto de vista geográfico, la delimitación geográfica no es una limitante para el análisis, pero si lo es al momento de que se quieren unir tanto la información espacial como la información socio demográfica porque debe de haber un punto común para unir los datos.
5. Para desarrollar el modelo de filtro de Kalman, en un primer momento se emplearon varias capas como las luces nocturnas o la infraestructura urbana. Sin embargo, se observó que estas capas no servían como un indicador de temporalidad (solo se tenían para 2020) y al hecho de que distorsionaba los resultados.

Futuras investigaciones

En este trabajo de investigación se reflexionó sobre las nuevas formas de llevar a cabo estimaciones de población haciendo uso de imágenes satelitales, lo cual es un tema poco explorado desde el ámbito sociodemográfico. También se reflexionó sobre la apertura a nuevas fuentes de información, tal y como se mencionó en el punto anterior existieron algunas limitaciones para

llevar a cabo esta investigación y es por ello que a continuación se presentan algunas reflexiones finales para futuras investigaciones:

1. El uso de imágenes satelitales demanda de un conjunto de habilidades particulares que permitan explotar su potencial. Para hacerlo, es necesario que en los cursos de formación se brinden las herramientas necesarias, tanto teóricas como metodológicas, para explotar estas fuentes de información que, si bien es cierto que no son nuevas, su incorporación al área sociodemográfica sí lo es.
2. La irrupción de la COVID-19 puso de manifiesto dos situaciones:
 - a. Es necesario explorar fuentes de información alternativas en donde no se tenga una interacción cara a cara. A inicios del 2020 todos los proyectos fueron cancelados y el Gobierno prohibió llevar a cabo encuestas en donde hubiera una interacción entre las personas. Esto significó que por algunos meses no se contará con información sociodemográfica con el nivel de precisión deseado. Aunque algunas fuentes de información se levantaron vía remota lo cierto es que en todo momento el Inegi enfatizó que debido a este cambio los datos no eran estrictamente comparables.
 - b. Los fenómenos sociodemográficos pueden cambiar de un momento a otro. Esta velocidad con la que pueden ocurrir los fenómenos es una situación que se debe de tomar en cuenta al momento de llevar a cabo los análisis, ya que la información que se levanta generalmente proporciona una panorámica desfasada, lo que significa que en realidad se está explicando lo que sucedió, pero no lo que está sucediendo. Esta situación abona a la propuesta de usar nuevas fuentes de información que se generen de manera constante.
3. El Censo de Población y Vivienda 2020 se tomó como referencia para saber cuál es la confiabilidad de las estimaciones, sin embargo, existen algunos indicios (que se mencionan en la tesis) de que este instrumento tiene problemas debido a que su levantamiento coincidió con el inicio de la jornada de sana distancia por lo que aumentó la no respuesta. Esta incertidumbre ya ha sido puesta de manifiesto con las quejas que han llevado a cabo algunas entidades federativas en donde ponen en duda los resultados censales. Esto son los primeros signos de alerta para continuar y profundizar en los métodos de estimación de la población.

Reflexiones sobre el Censo 2020

Al igual que en otros países latinoamericanos, México estaba preparando su levantamiento censal cuando la OMS declaró la emergencia sanitaria a nivel mundial. A pesar de que algunos miembros del gobierno⁵ propusieron suspender o retrasar el Censo, el operativo de campo inició el 2 de marzo de 2020 y terminó el 27 de marzo de ese mismo mes. Sin embargo, en el mismo mes de marzo la Secretaría de Salud anunció el inicio de la Jornada Nacional de Sana Distancia (JNSD), en donde se promovió un distanciamiento social y el cierre de todos los negocios no esenciales con la intención de mitigar los contagios. Así, el levantamiento de los datos coincidió con la JNSD, lo que provocó: incremento de la no respuesta, agresiones al personal que llevó a cabo las entrevistas y una deserción de encuestadores y supervisores ante el temor de contagiarse de COVID-19 al momento de interactuar con otras personas (Inegi, 2021).

Para responder a estos problemas, el Instituto Nacional de Estadística y Geografía (Inegi) continuó con el seguimiento remoto hasta el 31 de mayo de 2020. Para ello, los encuestadores entregaron cartas en donde se invitó a la población que no había sido censada para que ingresaran a la página de Internet del instituto o llamaran por teléfono para responder el cuestionario censal. Cabe señalar que esta es la primera vez que se implementa el autoempadronamiento vía remoto en México como una forma complementaria de recopilación de datos censales y aunque esta alternativa fue apropiada en el contexto de pandemia, presenta algunos retos como: falta de cobertura telefónica y de acceso a Internet; requiere un nivel extendido de alfabetización informática; y necesita la aceptación pública para contestar información personal a través de plataformas digitales (UNFPA, 2020). El reporte de campo no está disponible para los usuarios, pero en otros proyectos que levantó el Instituto encontró que la tasa de respuesta en los operativos remotos es menor y existe un sesgo, ya que los individuos con mayor nivel de educación son quienes responden por medios electrónicos (Dutwin y Buskirk, 2021; Fowler *et al.*, 2016; Olson *et al.*, 2021).

Además, el gobierno prohibió la realización de censos o encuestas que involucraran la movilización de personas e interacción física (cara a cara) de abril a julio de 2020 (DOF, 31 de marzo 2020a, 17 de julio 2020b). Esto provocó la cancelación de la encuesta de posenumeración

⁵ Ver <https://www.pan.senado.gob.mx/2020/03/por-coronavirus-julen-rementeria-pide-suspender-el-censo-nacional-de-poblacion-y-vivienda/>

(Inegi, 2020b), la cual serviría para evaluar la congruencia de los datos censales (UN, 2010a, 2010b). El argumento para respaldar esta decisión consistió en que de acuerdo con UN (2010a), la encuesta de posenumeración debe aplicarse inmediatamente después del Censo, pero con las restricciones impuestas se hubiera levantado entre cinco o seis meses después de levantar los datos. Esto significa que México no cuenta con un instrumento específico para medir la calidad de los datos censales.

Sin embargo, las anomalías trascendieron a una inconformidad que alcanzó límites jurídicos y llegó como solicitud a la Suprema Corte de Justicia para que invalidara la información desprendida de una encuesta que tiene como referencia los datos del censo. Esta situación toma otro carácter y deja la puerta abierta para cuestionar de ahora en adelante, cualquier dato oficial, aquellos que por mucho tiempo habían sido considerados indiscutibles. Es el caso de la controversia presentada por al menos dos gobiernos de las 32 entidades mexicanas, que solicitaron impugnar el resultado de los datos de la ENOE por considerarlas inexactas. Más allá de si procede o no esta discrepancia, el ejercicio de refutar información oficial ante instancias jurídicas siembra un antecedente digno de reflexión y permite preguntarse ¿podemos dar por válida toda la información que se desprenda del censo 2020? O ¿Podemos hacer estimaciones paralelas para contrastar la información censal?

REFERENCIAS

- Alexander, M., Zagheni, E. & Polimis, K. (2019). *The impact of Hurricane Maria on out-migration from Puerto Rico: Evidence from Facebook data*. <https://doi.org/10.31235/osf.io/39s6c>
- Alho, J. (2005). *Statistical demography and forecasting* (B. D. Spencer (ed.)). New York, N.Y. : Springer, c2005.
- Almanza, L. A. A. (1988). La invención de la “población.” *Revista Mexicana de Sociología*, 50(4), 135–170. <https://doi.org/10.2307/3540586>
- Alvarez, G. (2001). Estimación de población en áreas menores mediante variables sintomáticas: una aplicación para los departamentos de la República Argentina, 1991 y 1996. En *Población y Desarrollo*. Naciones Unidas Comisión Económica para América Latina y el Caribe (CEPAL). Recuperado de: <http://pbidi.unam.mx:8080/login?url=http://search.ebscohost.com/login.aspx?direct=true&db=edsrep&AN=edsrep.p.ecr.col045.7145&lang=es&site=eds-live>
- Arroba, M. E. S. (2009). Migración y pérdida de la lengua maya en Quintana Roo. *Migración y Políticas Públicas*, 397-citation_lastpage.
- Assunção, R. M., Schmertmann, C. P., Potter, J. E., & Cavenaghi, S. M. (2005). Empirical bayes estimation of demographic schedules for small areas. *Demography*, 42(3), 537–558. <https://doi.org/10.1353/dem.2005.0022>
- Baussola, M., & Fabrizi, E. (2008). Small area estimation and economic policy. *Rivista Internazionale Di Scienze Sociali*, 116(4), 379–385. Recuperado de: <http://www.jstor.org/stable/41625215>
- Bell, M. (2015). Demography, time and space. *Journal of Population Research*, 32(3/4), 173–186. Recuperado de: <http://www.jstor.org/stable/43919975>
- Billari, F. C. (2006). *Bridging the gap between micro-demography and macro-demography* (Vol. 4). Academic Press.
- Blanco, M. (2011). El enfoque del curso de vida: orígenes y desarrollo. *Revista Latinoamericana de Población*, 5(8), 5–31.
- Bocquet-Appel, J.-P. & JAKOBI, L. (1996). Barriers to the Spatial Diffusion for the Demographic Transition in Western Europe. *Spatial Analysis of Biodemographic Data*, 16.
- Bogue, D. J. (1950). A Technique for Making Extensive Population Estimates. *Journal of the American Statistical Association*, 45(250), 149–163. <https://doi.org/10.2307/2280675>
- Brown, J., & Guinnane, T. (2007). Regions and Time in the European Fertility Transition: Problems in the Princeton Project’s Statistical Methodology. *The Economic History Review*, 60, 574–595. <https://doi.org/10.1111/j.1468-0289.2006.00371.x>
- Bryant, J. R., & Graham, P. J. (2013). Bayesian Demographic Accounts: Subnational Population Estimation Using Multiple Data Sources. *Bayesian Anal.*, 8(3), 591–622. <https://doi.org/10.1214/13-BA820>
- Bunge, M. (1997). *Epistemología : curso de actualización* (2a ed.). Siglo Veintiuno.
- Cai, Q. & Tippett, R. (2015). Housing-Unit Method in Comparison: The Virginia Case. In *Emerging Techniques in Applied Demography* (pp. 81–92). Springer.
- Caldwell, J. C., Harrison, G. E. & Quiggin, P. (1980). The demography of micro-states. *World Development*, 8(12), 953–962. [https://doi.org/https://doi.org/10.1016/0305-750X\(80\)90087-X](https://doi.org/https://doi.org/10.1016/0305-750X(80)90087-X)
- Caldwell, J. C. (2006). *Demographic transition theory* (B. K. Caldwell (ed.)). Springer.
- Campos, J., Rigotti, J. I., Baptista, E. A., Monteiro, A. M. & Reis, I. A. (2020). Population Estimates from

- Orbital Data of Medium Spatial Resolution: Applications for a Brazilian Municipality. In *Sustainability* (Vol. 12, Issue 9). <https://doi.org/10.3390/su12093565>
- Canales Cerón, A. (2003). *Demografía de la desigualdad: el discurso de la población en la era de la globalización*.
- _____. & Lerner, S. (2003). *Desafíos teórico-metodológicos en los estudios de población en el inicio del milenio* (1a ed.). El Colegio de México, Centro de Estudios Demográficos y de Desarrollo Urbano, Programa de Salud Reproductiva, Sociedad Mexicana de Demografía, Zapopan, Jalisco, México, Universidad de Guadalajara, Centro Universitario de Ciencias Económicas.
- Castro, M. (2007). Spatial Demography: An Opportunity to Improve Policy Making at Diverse Decision Levels. *Population Research and Policy Review*, 26, 477–509.
- CEPAL. (2021). *Tecnologías digitales para un nuevo futuro*. CEPAL.
- CEPAL, N. U. (2018). *Datos, algoritmos y políticas: la redefinición del mundo digital*.
- Cerón, A. I. C. (2001). Discurso demográfico y posmodernidad. Una revisión crítica del pensamiento malthusiano. *Estudios Sociológicos*, 19(56), 381–417. Recuperado de: <http://www.jstor.org/stable/40420987>
- Chávez, A. M., Galindo, C., Rodríguez, J., Acuña, M., Barquero, J., Macadar, D., Cunha, J. & Sobrino, J. (2016). Migración interna y cambios metropolitanos: ¿qué está pasando en las grandes ciudades de América Latina? Internal migration and metropolitan changes: What is going on in Latin America's large cities? *Revista Latinoamericana de Microbiología*, 10. <https://doi.org/10.31406/relap2016.v10.i1.n18.1>
- Chen, Y. (2017). Recapturing space: New middle-range theory in spatial demography. In *Journal of Regional Science* (Vol. 57, Issue 3, pp. 540–541). <https://doi.org/10.1111/jors.12344>
- Chui, C. K. & Chen, G. (2017). *Kalman filtering*. Springer.
- Coale, A. J. (1984). The Demographic Transition. *The Pakistan Development Review*, 23(4), 531–552.
- Cochran, W. G. (2007). *Sampling techniques*. John Wiley & Sons.
- Courgeau, D. (2001). *Análisis demográfico de las biografías*. Recuperado de: https://www.researchgate.net/publication/233864401_Analisis_demografico_de_las_biografias
- Czarnecki, L. (2013). El CONEVAL. La institucionalización del concepto oficial de la pobreza en México. *BARATARIA. Revista Castellano-Manchega de Ciencias Sociales*, 16, 177–190.
- Davis, H. C. (1995). *Demographic projection techniques for regions and smaller areas: a primer*. UBC.
- De Rentería, P. G., Olza, M. B. & Zabaleta, A. P. (2019). Los objetivos de desarrollo sostenible: Una agenda para transformar el mundo. *Economistas*, 162, 205–209.
- Desa, U. N. (2016). *Transforming our world: The 2030 agenda for sustainable development*. Recuperado de: <https://sustainabledevelopment.un.org/post2015/transformingourworld/publication>
- Diario Oficial de la Federación. (2018). *ACUERDO que establece la política en materia satelital del Gobierno Federal.2*. Recuperado de: https://dof.gob.mx/nota_detalle.php?codigo=5522574&fecha=15/05/2018#gsc.tab=0
- Diario Oficial de la Federación. (2020a). *Acuerdo por el que se establecen acciones extraordinarias para atender la emergencia sanitaria generada por el virus SARS-CoV2*. Recuperado de: https://www.dof.gob.mx/nota_detalle.php?codigo=5590914&fecha=31/03/2020&print=true
- _____. (2020b). *ACUERDO por el que se establecen acciones extraordinarias para atender la emergencia sanitaria generada por el virus SARS-CoV2*. 2020–2021. Recuperado de: https://www.dof.gob.mx/nota_detalle.php?codigo=5590914&fecha=31/03/2020#gsc.tab=0
- _____. (2020c). *Acuerdo por el que se reanudan todos los censos y encuestas a realizarse en el territorio*

- nacional. Recuperado de:
http://www.dof.gob.mx/nota_detalle.php?codigo=5596930&fecha=17/07/2020
- _____. (2008). *Ley del Sistema Nacional de Información Estadística y Geográfica*. México.
- Dutwin, D. & Buskirk, T. D. (2021). Telephone Sample Surveys: Dearly Beloved or Nearly Departed? Trends in Survey Errors in the Era of Declining Response Rates. *Journal of Survey Statistics and Methodology*, 9(3), 353–380. <https://doi.org/10.1093/jssam/smz044>
- Dwork, C., Kohli, N. & Mulligan, D. (2019). Differential Privacy in Practice: Expose your Epsilons! *Journal of Privacy and Confidentiality*, 9(2). <https://doi.org/10.29012/jpc.689>
- Eckman, S., Himelein, K. & Dever, J. A. (2018). Innovative Sample Designs Using GIS Technology. In *Advances in Comparative Survey Methods* (pp. 65–92). <https://doi.org/doi:10.1002/9781118884997.ch4>
- Ericksen, E. P. (1973). A method for combining sample survey data and symptomatic indicators to obtain population estimates for local areas. *Demography*, 10(2), 137–160. <https://doi.org/10.2307/2060811>
- Evans, G., King, G., Smith, A. D. & Thakurta, A. (n.d.). *Differentially Private Survey Research*.
- Figueroa Perea, J. G. (2006). Algunas reflexiones sobre ética y estudio de los varones en la investigación demográfica: una experiencia personal en la IUSSP. *Estudios Demográficos y Urbanos*, 21(1), 227–237.
- Figueroa Perea, J. G. (2010). La construcción de la titularidad para el ejercicio de los derechos reproductivos. En *Población: los grandes problemas de México. El Colegio de México* (pp. 253–289). El Colegio de México.
- Figueroa Perea, J. G. (2012). Diálogo con mi proceso de investigación sociodemográfica: un enfoque socrático. *Estudios Demográficos y Urbanos*, 27(3), 839–852. <https://doi.org/10.24201/edu.v27i3.1428>
- Florencia Santi, M. (2016). *Ética de la investigación en ciencias sociales: un análisis de la vulnerabilidad en la investigación social*. Globethics.net.
- Flowerdew, R. (2011). How serious is the Modifiable Areal Unit Problem for analysis of English census data? *Population Trends*, 145, 102–114. <https://doi.org/10.1057/pt.2011.20>
- Fowler, F. J., Roman, A. M., Mahmood, R. & Cosenza, C. A. (2016). Reducing Nonresponse and Nonresponse Error in a Telephone Survey: An Informative Case Study. *Journal of Survey Statistics and Methodology*, 4(2), 246–262. <https://doi.org/10.1093/jssam/smw004>
- Fucaraccio G., A. (1975). *Notas para una discusión acerca de la ley de población en Marx* (F. González (ed.)). Consejo Latinoamericano de Ciencias Sociales, Comisión de Población y Desarrollo, Unidad Central del Programa de Investigaciones Sociales sobre Problemas de Población Relevantes para Políticas de Población en América Latina.
- Galison, P. L. & Stump, D. J. (1996). *The disunity of science: Boundaries, contexts, and power*. Stanford University Press.
- García, B. & Ordorica, M. (2010). *Los grandes problemas de México. Población. TI*. El Colegio de México.
- García Guerrero, V. M. & Ordorica Mellado, M. (2012). Proyección estocástica de la mortalidad mexicana por medio del método de Lee-Carter. *Estudios Demográficos y Urbanos*, 27(2), 409–448.
- Garfinkel, S. L., Abowd, J. M. & Powazek, S. (2018). *Issues Encountered Deploying Differential Privacy*. <https://doi.org/10.1145/3267323.3268949>
- Garfinkel, S. L. & Leclerc, P. (2020). *Randomness Concerns When Deploying Differential Privacy*. <https://doi.org/10.1145/3411497.3420211>

- Gobierno de la Ciudad de México. (2021). *Ley de protección de datos personales en posesión de sujetos obligados de la Ciudad de México*. Cámara de Diputados del H. Congreso de la Unión. Recuperado de:
https://data.consejeria.cdmx.gob.mx/images/leyes/leyes/LEY_DE_PROTECCION_DE_DATOS_PERSONALES_EN_POSESION_DE_SUJETOS_OBLIGADOS_DE_LA_CDMX_4.pdf
- Ghosh, M. & Rao, J. N. K. (1994). Small Area Estimation: An Appraisal. *Statistical Science*, 9(1), 55–76. Recuperado de: <http://www.jstor.org/stable/2246284>
- Gil, C. G. (2018). Objetivos de Desarrollo Sostenible (ODS): una revisión crítica. *Papeles de Relaciones Ecosociales y Cambio Global*, (140), 107–118.
- Golden, M. L., Downs, R. R., Kent, D. P. & Davis-Packard, K. (2005). *Confidentiality Issues and Policies Related to the Utilization and Dissemination of Geospatial Data for Public Health Applications*. March.
- González Gutiérrez, J. (2018). *La representación de la desigualdad social en el estudio de la mortalidad infantil en México durante el periodo de 1982 a 2014* (J. G. Figueroa Perea, M. Ordorica, & C. A. Dávila-Cervantes (eds.)). El Colegio de México.
- Gorelick, N., Hancher, M., Dixon, M., Ilyushchenko, S., Thau, D. & Moore, R. (2017). Google Earth Engine: Planetary-scale geospatial analysis for everyone. *Remote Sensing of Environment*, 202, 18–27. <https://doi.org/https://doi.org/10.1016/j.rse.2017.06.031>
- Grewal, M. S. & Andrews, A. P. (2014). *Kalman filtering: Theory and Practice with MATLAB*. John Wiley & Sons.
- Griffith, D. A., Wong, D. W. S. & Whitfield, T. (2003). Exploring Relationships Between the Global and Regional Measures of Spatial Autocorrelation. *Journal of Regional Science*, 43(4), 683–710. <https://doi.org/10.1111/j.0022-4146.2003.00316.x>
- Guerra, L. L. (1977). Niveles de análisis, Falacia Ecológica y Falacia Contextual. *Revista Española de La Opinión Pública*, 48, 69–87. <https://doi.org/10.2307/40199477>
- Guerrero, V. M. G. (2014). *Proyecciones y políticas de población en México*. El Colegio de México.
- Gutman, G., Huang, C., Chander, G., Noojipady, P. & Masek, J. G. (2013). Assessment of the NASA-USGS Global Land Survey (GLS) datasets. *Remote Sensing of Environment*, 134, 249–265. <https://doi.org/10.1016/j.rse.2013.02.026>
- Hauer, M. E., Evans, J. M. & Alexander, C. R. (2015). Sea-level rise and sub-county population projections in coastal Georgia. *Population and Environment*, 37(1), 44–62. <https://doi.org/10.1007/s11111-015-0233-8>
- Hauser, P. M., & Duncan, O. D. (1959). *The study of population: An inventory and appraisal*. Chicago: University of Chicago Press, 1959.
- Héran, F. (2013). The population ageing process in global perspective. *Global Ageing in the Twenty-First Century: Challenges, Opportunities and Implications*, 13–29.
- Hidalgo, M. G. A., González, D. del C. C. & Carrillo, M. A. O. (2019). Competitividad e innovación en la industria del calzado en México. Caso: Christian Gallery en León Guanajuato. En *El papel de las empresas locales en el desarrollo regional de México* (pp. 107–128). Recuperado de:
https://www.researchgate.net/publication/334440580_El_papel_de_las_empresas_locales_en_el_desarrollo_Regional_de_Mexico
- Horiuchi, S. (1999). Epidemiological transitions in human history. *Health and Mortality: Issues of Global Concern*, 54–71.
- Inegi. (2021). *Así se contó México*. Instituto Nacional de Estadística y Geografía.

- _____. (2018). *Encuesta Intercensal 2015: marco conceptual*. Instituto Nacional de Estadística y Geografía.
- _____. (2020). *Comunicado de prensa núm. 142/20*.
<https://www.inegi.org.mx/contenidos/saladeprensa/boletines/2020/especiales/INEGI-COVID.pdf>
- Ito, R., Hara, K., Shimazaki, Y., Mori, N., Okabe, A. & Okabe, A. (2020). A Method for Estimating the Number of Households in a Region from the Number of buildings estimated by deep learning with the Adjustment of its Number using Ancillary Datasets: Case Study in Djakarta. *Isprs - International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences, XLIII-B4-2*, 673–678. <https://doi.org/10.5194/isprs-archives-XLIII-B4-2020-673-2020>
- Jetzek, T., Avital, M. & Bjorn-Andersen, N. (2014). Data-Driven Innovation through Open Government Data. *Journal of Theoretical and Applied Electronic Commerce Research* , 9(2), 100–120.
<https://doi.org/10.4067/S0718-18762014000200008>
- Keshava, N. & Mustard, J. F. (2002). Spectral unmixing. *IEEE Signal Processing Magazine*, 19(1), 44–57.
- Keyfitz, N. & Caswell, H. (2005). *Applied mathematical demography* (Vol. 47). Springer.
- Kugler, T. A., Van Riper, D. C., Manson, S. M., Haynes II, D. A., Donato, J., Stinebaugh, K., Haynes, D. A., Donato, J. & Stinebaugh, K. (2015). Terra Populus: Workflows for Integrating and Harmonizing Geospatial Population and Environmental Data. *Journal of Map & Geography Libraries*, 11(2), 180–206. <https://doi.org/10.1080/15420353.2015.1036484>
- Kumar, L. & Mutanga, O. (2018). Google Earth Engine Applications Since Inception: Usage, Trends, and Potential. *Remote Sensing* , 10(10), 1–15. <https://doi.org/10.3390/rs10101509>
- Lee, R. D. (2001). Demography abandons its core. *Remarks Made Flatpanel Discussion on Micro-Macro Issues at the Annual Meetings of the Population Association of America in*.
- Léridon, H. (2014). *Demografía: enfoque estadístico y dinámica de las poblaciones* (L. Toulemon & S. Luna Santos (eds.); Primera ed). El Colegio de México, Centro de Estudios Demográficos, Urbanos y Ambientales.
- Lesthaeghe, R. (2014). The second demographic transition: A concise overview of its development. *Proceedings of the National Academy of Sciences*, 111(51), 18112 LP – 18115.
<https://doi.org/10.1073/pnas.1420441111>
- Lichter, D. T. & Ziliak, J. P. (2017). The rural-urban interface: New patterns of spatial interdependence and inequality in America. *The ANNALS of the American Academy of Political and Social Science*, 672(1), 6–25.
- Lillesand, T., Kiefer, R. W., Chipman, J. & Congalton, R. G. (2015). Remote sensing and image interpretation. In *Photogrammetric Engineering & Remote Sensing* (Vol. 81, Issue 8). John Wiley & Sons. <https://doi.org/10.14358/pers.81.8.615>
- Livi-Bacci, M. (2009). *Historia mínima de la población mundial*. Ariel.
- Logan, J. (2016). *Challenges of Spatial Thinking* (pp. 11–36). https://doi.org/10.1007/978-3-319-22810-5_2
- Lomax, N. M. & Smith, A. P. (2017). *Microsimulation for demography*. Northern Institute, Charles Darwin University. Recuperado de:
<http://pbidi.unam.mx:8080/login?url=http://search.ebscohost.com/login.aspx?direct=true&db=edsair&AN=edsair.core.ac.uk....e0f421c270cb9b210987c088f5fdc0c0&lang=es&site=eds-live>
- Malthus, T. R. (1966). *Primer ensayo sobre la población* (J. M. Keynes & P. de Azcárate Diz (eds.)). Alianza.

- Mansilla Torres, K. (2011). De los Estados nación a los Estados plurinacionales. En *Universitas. Revista de Ciencias Sociales y Humanas* (Issue 15). Recuperado de: <https://www.redalyc.org/articulo.oa?id=476147383002>
- Manson, S. M. (2008). Does scale exist? An epistemological scale continuum for complex human–environment systems. *Geoforum*, 39(2), 776–788. <https://doi.org/https://doi.org/10.1016/j.geoforum.2006.09.010>
- Marshall, R. J. (1991). Mapping Disease and Mortality Rates Using Empirical Bayes Estimators. *Journal of the Royal Statistical Society. Series C (Applied Statistics)*, 40(2), 283–294. <https://doi.org/10.2307/2347593>
- Martin, D. (2006). Last of the Censuses? The Future of Small Area Population Data. *Transactions of the Institute of British Geographers*, 31(1), 6–18. <http://www.jstor.org/stable/3804416>
- Masek, J. G., Wulder, M. A., Markham, B., McCorkel, J., Crawford, C. J., Storey, J. & Jenstrom, D. T. (2020). Landsat 9: Empowering open science and applications through continuity. *Remote Sensing of Environment*, 248(April), 111968. <https://doi.org/https://doi.org/10.1016/j.rse.2020.111968>
- Matthews, S. A. (2016). *Instruction in Spatial Demography BT - Recapturing Space: New Middle-Range Theory in Spatial Demography* (F. M. Howell, J. R. Porter, & S. A. Matthews (eds.); pp. 355–369). Springer. https://doi.org/10.1007/978-3-319-22810-5_17
- Matthews, S. A. & Parker, D. M. (2013). Progress in Spatial Demography. *Demographic Research*, 28, 271–312. Recuperado de: <http://www.jstor.org/stable/26349953>
- Mayer-Schönberger, V. & Cukier, K. (2013). *Big data : la revolución de los datos masivos*. Turner Publicaciones S.L.
- Miró, C A. (1971). *Política de población: qué? por qué? para qué? como?* Centro Latinoamericano de Demografía. Recuperado de: <https://books.google.com.mx/books?id=N2M-AAAAYAAJ>
- Miró, Carmen A. (2015). América Latina, Población y Desarrollo. In *Journal of Chemical Information and Modeling* (Vol. 53, Issue 9). <https://doi.org/10.1017/CBO9781107415324.004>
- Molina, I. (2019). Desagregación de datos en encuestas de hogares. *CEPAL - Serie Estudios Estadísticos*, 97.
- Murdock, S. H. & Hoque, M. N. (1995). The Effect of Undercount on the Accuracy of Small-Area Population Estimates: Implications for the Use of Administrative Data for Improving Population Enumeration. *Population Research and Policy Review*, 14(2), 251–271. Recuperado de: <http://www.jstor.org/stable/40230058>
- Namoodiri, N. K. (1972). On the Ratio-Correlation and Related Methods of Subnational Population Estimation. *Demography*, 9(3), 443–453. <https://doi.org/10.2307/2060865>
- United Nations (2000). *Handbook on geographic information systems and digital mapping*. UN,. <http://digitallibrary.un.org/record/426424>
- _____ (2014). *Naciones Unidas. Los datos demográficos: Alcances, limitaciones y métodos de evaluación. Serie Manuales*.
- O’Farrell, R. & Montagnier, P. (2020). Measuring digital platform-mediated workers. *New Technology, Work and Employment*, 35(1), 130–144.
- Olson, K., Smyth, J. D., Horwitz, R., Keeter, S., Lesser, V., Marken, S., Mathiowetz, N. A., McCarthy, J. S., O’Brien, E., Opsomer, J. D., Steiger, D., Sterrett, D., Su, J., Suzer-Gurtekin, Z. T., Turakhia, C. & Wagner, J. (2021). Transitions from Telephone Surveys to Self-Administered and Mixed-Mode Surveys: AAPOR Task Force Report. *Journal of Survey Statistics and Methodology*, 9(3), 381–411. <https://doi.org/10.1093/jssam/smz062>

- Omran, A.-R. (1971). *The health theme in family planning*. Carolina Population Center, University of North Carolina at Chapel Hill.
- ONU. (2010). *Principios y recomendaciones para los censos de población y habitación. Revisión 2*. Naciones Unidas Nueva York.
- Openshow, S. (1979). A million or so correlation coefficients, three experiments on the modifiable areal unit problem. *Statistical Applications in the Spatial Science*, 127–144.
- Ordorica-Mellado, M., Silvan-Cárdenas, J. L. & Núñez, J. M. (2019). Estimaciones de la población desde la tierra como desde el cielo: el caso de Tenosique. *Papeles de Población; Vol. 24 Núm. 98 (2018): Papeles de Población*. Recuperado de: <https://rppoblacion.uaemex.mx/article/view/10160>
- Ordorica, M. (2011). Dilemas éticos en las estimaciones demográficas: dos minihistorias. *Estudios Demográficos y Urbanos*, 26(2), 495. <https://doi.org/10.24201/edu.v26i2.1392>
- Organización para la Cooperación y Desarrollo Económico (2008). Directrices de la OCDE que regulan la protección de la privacidad y el flujo transfronterizo de datos personales. Disponible en: http://www.oas.org/es/sla/ddi/docs/Directrices_OCDE_privacidad.pdf Acceso el 15 de agosto de 2020
- Partida Bush, V. (2015). *Conciliación demográfica de México 1950-2015*. Consejo Nacional de Población.
- Partida, V. (2018). Proyecciones de la población de México 2005-2050 Consejo Nacional de Población (CONAPO). Recuperado de: <http://www.conapo.gob.mx/00cifras/Proy/Proy05-50pdf>.
- Pasalagua, M. F. & Sánchez, E. T. (2022). Territory and Water in the NAFTA Automotive Corridor in San Luis Potosí, Mexico. *Norteamérica, Revista Académica Del CISAN-UNAM*, 17(1 se-dossier). <https://doi.org/10.22201/cisan.24487228e.2022.1.577>
- Perilla, G. A., Mas, J.-F., Perilla, G. A., & Mas, J.-F. (2020). Google Earth Engine (GEE): una poderosa herramienta que vincula el potencial de los datos masivos y la eficacia del procesamiento en la nube. *Investigaciones Geográficas*, 101. <https://doi.org/10.14350/ig.59929>
- Petit, V. (2013). *Counting populations, understanding societies : towards an interpretative demography*. Springer.
- Pfeffermann, D. (2013). New Important Developments in Small Area Estimation. *Statistical Science*, 28(1), 40–68. Recuperado de: <http://www.jstor.org/stable/43288411>
- Pressat, R. (1967). *El análisis demográfico: métodos, resultados, aplicaciones* (F. I. N. d'Etudes Démographiques (ed.)). Fondo de Cultura Económica, [1967].
- Rao, J. N. K. (2008). Some methods for small area estimation. *Rivista Internazionale Di Scienze Sociali*, 116(4), 387–406. <http://www.jstor.org/stable/41625216>
- Rao, J. N. K. & Molina, I. (2015). *Small area estimation*. (Second edi). Wiley. Recuperado de: <http://pbidi.unam.mx:8080/login?url=http://search.ebscohost.com/login.aspx?direct=true&db=cat02025a&AN=lib.MX001001900681&lang=es&site=eds-live>
- Robinson, C., Hohman, F., & Dilkina, B. (2017). *A Deep Learning Approach for Population Estimation from Satellite Imagery*. <https://doi.org/10.1145/3149858.3149863>
- Rogers, A. (2015). *Applied multiregional demography: migration and population redistribution*. Springer. Recuperado de: <http://pbidi.unam.mx:8080/login?url=http://search.ebscohost.com/login.aspx?direct=true&db=cat02025a&AN=lib.MX001001888473&lang=es&site=eds-live>
- Ruggles, S. (2013). Big Microdata for Population Research. *Demography*, 51(1), 287–297. <https://doi.org/10.1007/s13524-013-0240-2>

- Ruggles, S., King, M. L., Levison, D., Mccaa, R. & Sobek, M. (2003). *Historical Methods: A Journal of Quantitative and Interdisciplinary History IPUMS-International*. 36(2), 60–65.
<https://doi.org/10.1080/01615440309601215>
- Ruiz Rivera, N. & Galicia, L. (2016). La escala geográfica como concepto integrador en la comprensión de problemas socio-ambientales. *Geographical Scale as Integrative Concept for Understanding Socio-Environmental Problems (English)*, 2016(89), 137–153. Recuperado de:
<http://10.0.56.14/rig.47515>
- Salinas-García, R. J. (2012). Desarrollo industrial y formación profesional en la industria aeronautica en Querétaro. *Revista de Educación y Desarrollo*, 23, 5–14.
- Salles, V. (2001). El debate micro-macro: Dilemas y contextos. *Perfiles Latinoamericanos: Revista de La Sede Académica de México de La Facultad Latinoamericana de Ciencias Sociales*, 10, 115–151.
- Sánchez, N. (2014). Historia mínima de la población de América Latina. *El Colegio de México*.
- Schmertmann, C. P. & Gonzaga, M. R. (2018). Bayesian Estimation of Age-Specific Mortality and Life Expectancy for Small Areas With Defective Vital Records. *Demography*, 55(4), 1363–1388.
<https://doi.org/10.1007/s13524-018-0695-2>
- Schmitt, R. C. & Crosetti, A. H. (1954). Accuracy of the Ratio-Correlation Method for Estimating Postcensal Population. *Land Economics*, 30(3), 279–281. <https://doi.org/10.2307/3144384>
- Silván-Cárdenas, Jose L, Wang, L., Rogerson, P., Wu, C., Feng, T. & Kamphaus, B. D. (2010). Assessing fine-spatial-resolution remote sensing for small-area population estimation. *International Journal of Remote Sensing*, 31(21), 5605–5634. <https://doi.org/10.1080/01431161.2010.496800>
- Silván-Cárdenas, José Luis, Montejano-Escamilla, J. & Cervantes-Salas, M. (2015). Percepción remota para la estimación de población en áreas geoestadísticas básicas. *Realidad Datos y Espacios*, 6(1), 50–71.
- Silván-Cárdenas, José Luis, Valdiviezo Navarro, J. C., & Salazar Garibay, A. (2021). Mapeo de la Mancha Urbana en el Tiempo: Caso de Estudio Corredor Metropolitano del Bajío. En E. Martínez Viveros (Ed.), *Monitoreo de procesos territoriales con imágenes de satélite en escala regional* (1ª ed, pp. 55–76). Centro de Investigación en Ciencias de Información Geoespacial A.C. Recuperado de: <https://idegeo.centrogeo.org.mx/documents/4805>
- Silvey, R. (2004). On the boundaries of a subfield: social theory's incorporation into population geography. *Population Space & Place*, 10(4), 303.
<http://pbidi.unam.mx:8080/login?url=http://search.ebscohost.com/login.aspx?direct=true&db=edb&AN=20433262&lang=es&site=eds-live>
- Smith, M. J. & Wise, S. M. (2007). Problems of bias in mapping linear landforms from satellite imagery. *International Journal of Applied Earth Observation and Geoinformation*, 9(1), 65–78.
<https://doi.org/https://doi.org/10.1016/j.jag.2006.07.002>
- Spiegel, M. R., Schiller, J. J., Srinivasan, R. A., Nagore Cázares, G. & Spiegel, M. R. (2013). *Probabilidad y estadística*. (Cuarta edi). McGraw-Hill Interamericana. Recuperado de:
<http://pbidi.unam.mx:8080/login?url=http://search.ebscohost.com/login.aspx?direct=true&db=cat02025a&AN=lib.MX001001698960&lang=es&site=eds-live>
- Swanson, D. A. (2012). *Subnational population estimates* (J. Tayman (ed.)). Dordrecht : London : Springer.
- UK Statistics. (n.d.). *Ethical considerations in the use of geospatial data for research and statistics – UK Statistics Authority*. 18 May 2021. Retrieved September 15, 2021. Recuperado de:
<https://uksa.statisticsauthority.gov.uk/publication/ethical-considerations-in-the-use-of-geospatial->

- data-for-research-and-statistics/pages/1/
- Vela, H. M. S. (2016). Revisando los métodos de agregación de unidades espaciales: MAUP, algoritmos y un breve ejemplo. *Estudios Demográficos y Urbanos*, 31(2), 385–411. Recuperado de: <http://pbidi.unam.mx:8080/login?url=http://search.ebscohost.com/login.aspx?direct=true&db=edsdoj&AN=edsdoj.424b9d00c12147fc868d1f2a17653cb7&lang=es&site=eds-live>
- Véron, J., Preston, S., Heuveline, P. & Guillot, M. (2002). Demography, Measuring and Modeling Population Processes. *Population (French Edition)*, 57, 591. <https://doi.org/10.2307/1535065>
- Vieira Pinto, A. B. (1973). *El pensamiento crítico en demografía*. Centro Latinoamericano de Demografía.
- Villoro, L. (1996). *Crear, saber, conocer* (9 ed.). Siglo Veintiuno.
- Voss, P. (2007). Demography as a Spatial Social Science. *Population Research and Policy Review*, 26, 457–476. <https://doi.org/10.1007/s11113-007-9047-4>
- Wardrop, N. A., Jochem, W. C., Bird, T. J., Chamberlain, H. R., Clarke, D., Kerr, D., Bengtsson, L., Juran, S., Seaman, V. & Tatem, A. J. (2018). Spatially disaggregated population estimates in the absence of national population and housing census data. *Proceedings of the National Academy of Sciences*, 115(14), 3529 LP – 3537. <https://doi.org/10.1073/pnas.1715305115>
- Waters, N. (2018). *Tobler's First Law of Geography*. 1–15. <https://doi.org/10.1002/9781118786352.wbieg1011.pub2>
- Weeks, J. (2001). The role of spatial analysis in demographic research. *Spatially Integrated Social Science*.
- Weeks, J. R. (2016). *Demography Is an Inherently Spatial Science BT - Recapturing Space: New Middle-Range Theory in Spatial Demography* (F. M. Howell, J. R. Porter, & S. A. Matthews (eds.); pp. 99–122). Springer. https://doi.org/10.1007/978-3-319-22810-5_6
- Welti-Chanes, C. (2011). La Demografía en México, las etapas iniciales de su evolución y sus aportaciones al desarrollo nacional. *Papeles de Población*, 17(69), 9–47.
- Welti Chanes, C. (2014). El Consejo Nacional de Población a 40 años de la institucionalización de una política explícita de población en México. *Papeles de Población*, 20(81), 25–58.
- Wijedasa, L. S., Sloan, S., Michelakis, D. G., & Clements, G. R. (2012). Overcoming Limitations with Landsat Imagery for Mapping of Peat Swamp Forests in Sundaland. *Remote Sensing*, 4(9), 2595–2618. <https://doi.org/10.3390/rs4092595>
- Wong, D. W. S. (2003). Spatial Decomposition of Segregation Indices: A Framework Toward Measuring Segregation at Multiple Levels. *Geographical Analysis*, 35(3), 179–194. <https://doi.org/10.1111/j.1538-4632.2003.tb01109.x>
- Wong, D. W. S. (2004). *The Modifiable Areal Unit Problem (MAUP) BT - WorldMinds: Geographical Perspectives on 100 Problems: Commemorating the 100th Anniversary of the Association of American Geographers 1904–2004* (D. G. Janelle, B. Warf, & K. Hansen (eds.); pp. 571–575). Springer. https://doi.org/10.1007/978-1-4020-2352-1_93
- Xu, M., Cao, C. & Jia, P. (2020). Mapping Fine-Scale Urban Spatial Population Distribution Based on High-Resolution Stereo Pair Images, Points of Interest, and Land Cover Data. *Remote Sensing*, 12(4). <https://doi.org/10.3390/rs12040608>
- Yang, W., John, V. O., Zhao, X., Lu, H. & Knapp, K. R. (2016). Satellite Climate Data Records: Development, Applications, and Societal Benefits. *Remote Sensing* (Vol. 8, Issue 4). <https://doi.org/10.3390/rs8040331>
- Yuridia, C. B. (2014). *La región bajío-centro-occidente de México, ¿el nuevo polo industrial del país?* Recuperado de: http://www.saree.com.mx/unam/sites/default/files/CANEDOB8_2014.pdf

Zavala De Cosio, M. E. (1992). La transición demográfica en América Latina y en Europa. *Notas de Población*, 20(56), 11–32.

ANEXO

El objetivo de esta sección es mostrar los detalles técnicos de cómo se llevaron a cabo las estimaciones y la manera en que se presenta la información.

Tal y como se expuso a lo largo de la tesis, hoy en día se cuenta con una gran cantidad de información sociodemográfica y satelital. Esto último se debe a que a partir de la década de los años 60 el servicio geológico de los Estados Unidos dio acceso a los datos que generaban los diversos sensores que se pusieron en órbita. Esto significa que, a diferencia de lo que ocurre en otras áreas como la demografía, la disponibilidad de información precedió a la capacidad de análisis (Yang *et al.*, 2016).

A finales de 2010 Google puso a disposición de los usuarios la plataforma denominada Google Earth Engine (GEE), la cual buscó proporcionar una mayor capacidad de cómputo y permitir que los usuarios llevarán a cabo análisis cada vez más complejos. Una de sus ventajas es que todo el procesamiento se lleva a cabo en la nube o, dicho de otra manera, se ocupan los recursos informáticos de Google. Gracias a ello se puede procesar una gran cantidad de información y no se ocupan los recursos que ofrecen las computadoras personales de quienes llevan a cabo las investigaciones. Para poner en perspectiva esta situación, Gorelick *et al.* (2017) señalaron que desarrollar un mapa con 15 m de resolución le toma a GEE un par de días, para replicar este proceso en una computadora personal que se encuentran disponibles en el mercado se tendría que ocupar alrededor de 14 años.

Otro de los elementos que señalan Perilla *et al.* (2020) que tiene un gran impacto en el uso de esta plataforma es que contiene una gran cantidad de acervos que provienen de diversos sensores. Estos datos se actualizan constantemente y crean un catálogo que crece en todo momento, además cuenta con una gran diversidad temática y abarca temas referentes a los componentes geográficos de la tierra, así como información sociodemográfica. Además, el acceso a las bases de datos es gratuito y se puede hacer mediante un entorno de desarrollo integrado (IDE) denominado Code Editor. Actualmente, la información se puede obtener mediante el lenguaje de programación JavaScript y cuenta con una extensión que permite hacer el vínculo con Python.

Durante el desarrollo de esta tesis, se intentó hacer esta conexión entre ambos lenguajes de programación, sin embargo, existieron una gran cantidad de problemas debido a la incompatibilidad de versiones y al hecho de que los entornos de desarrollo utilizados para correr Python no se pueden acceder de manera directa al acervo que ya está implementado en el Code Editor. Por este motivo, la visualización de los datos se implementó con los comandos de JavaScript, aunque el procesamiento de estos se llevó a cabo en R.

Otro elemento que también debe de ser tomado en cuenta es que esta plataforma es de código abierto, lo que significa que los usuarios pueden modificar las rutinas o funciones previamente establecidas e intercambiar los resultados con otras personas. Esto abre un debate en torno a la forma en que se genera conocimiento. Por un lado, el hecho de que el código se puede compartir con la comunidad permite mejorar su implementación y servir de plataforma para desarrollar otras metodologías. Por el otro, cae dentro de las preocupaciones que algunas personas han alertado referentes a los peligros de utilizar la información con fines negativos. El hecho de que las imágenes satelitales tengan una resolución cada vez mayor y que su frecuencia de adquisición sea a instantáneo, podría representar una situación negativa ya que es un mecanismo de vigilancia. En este último punto la reflexión sería que lo importante es resaltar que el peligro no está en la herramienta, sino el uso que se le da y los fines para los cuales si utiliza (Kumar y Mutanga, 2018).

Código de Google Earth Engine

```
//0. Consideraciones iniciales

app.data = {};
Map.style().set('cursor', 'crosshair');
Map.setOptions("TERRAIN")

//1. Procesamiento

//Variables
var ColNom = 'Municipio', //Nombre de la Columna
    MpioNom = "Cosío"; //Nombre de municipio

//1.1 Filtrar el municipio a procesar
var muni_area = ee.FeatureCollection("users/julioccolmex/mpios_area_pob"),
    mpio = muni_area.filter
        (ee.Filter.eq(ColNom, MpioNom));

var empty = ee.Image().byte();
var limites = empty.paint({
  featureCollection: muni_area,
  color: 1,
  width: 1
});

Map.centerObject(muni_area, 8) //centra la vista a toda la zona de estudio

//1.2 Leer la población total de cada municipio
var PobMunSel = mpio.reduceColumns(
    ee.Reducer.sum(),
    ['Total']).values().getInfo()
var PobMuni = PobMunSel;

//1.3 Pixeles nulos y 0 (no hay cobertura urbana)
var OriginalPob = ee.Image("users/julioccolmex/FMPU_LC08_2020-1"),
    PobMask = OriginalPob.updateMask(OriginalPob.neq(0)),
    Pob = ee.Image("users/julioccolmex/FMPU_LC08_2020-1").mask(PobMask);

var imagePobMpio = Pob.clip(mpio); //pixeles > 0 en base al municipio filtrado
```

```
//1.4 Obtiene la suma de los valores de todos los pixeles del municipio
var PobTotPixel = imagePobMpio.select(["b1"]);
```

```
//1.5 Suma de los valores para el municipio filtrado
var PobSumMpio = PobTotPixel.reduceRegion( {
  reducer: ee.Reducer.sum(),
  geometry: mpio,
  scale: 30
}).values().getInfo();//convierte a valores numéricos
```

```
//1.6 Calculo
var PobOri = imagePobMpio.expression(
  'PobSumMpio / imagePobMpio', {
  PobSumMpio: PobSumMpio,           //diccionario
  imagePobMpio: imagePobMpio       //diccionario
});
```

```
var PobDistr = PobOri.expression (
  'PobMuni / PobOri', {
  PobMuni: PobMuni,
  PobOri: PobOri
});
```

```
//1.7 Visualiación
var DistVis = {"opacity":1,
  "bands":["constant"],
  "min":0.005565691473122755,
  "max":8.854409620258656,
  "palette":["e7ffff","c6cddf","baffc0","d6ffa9","fcff45",
    "ffc79f","ff81de","ff749a","ff5e5e","d63131","ab0000"]};
```

```
//1.8 Quita espacios a los nombres de los municipios
var MpioNom2 = MpioNom.replace(/ /g, "");
```

```
//2. Unir los municipios
var mos = ee.Image(ee.ImageCollection([image, image2, image3, image4, image5, image6,
image7, image8, image9, image10, image11, image12, image13, image14, image15, image16,
image17, image18, image19, image20, image21, image22, image23, image24, image25,
image26, image27, image28, image29, image30, image31, image32, image33, image34,
image35, image36, image37, image38, image39, image40, image41, image42, image43,
image44, image45, image46, image47, image48, image49, image50, image51, image52,
image53, image54, image55, image56, image57, image58, image59, image60, image61,
image62, image63, image64, image65, image67, image68, image69, image70, image71,
```

```

image72, image73, image74, image75, image76, image77, image78, image79, image80,
image81, image82, image83, image84, image85, image86, image87, image88, image89,
image90, image91, image92, image93, image94, image95, image96, image97, image98,
image99, image100, image101, image102, image103, image104, image105, image106,
image107, image108, image109, image110, image111, image112, image113, image114,
image115, image116, image117, image118, image119, image120, image121, image122,
image123, image124, image125, image126, image127, image128, image129, image130,
image131, image132, image133, image134, image135, image136, image137, image138,
image139, image140, image141, image142, image143, image144, image145, image146,
image147, image148, image149, image150, image151, image152, image153, image154,
image155, image156, image157, image158, image159, image160, image161, image162,
image163, image164, image165, image166, image167, image168, image169, image170,
image171, image172, image173, image174, image175, image176, image177, image178,
image179, image180, image181, image182, image183, image184, image185, image186,
image187, image188, image189, image190, image191, image192, image193, image194,
image195, image196, image197, image198, image199, image200, image201, image202,
image203, image204, image205, image206, image207, image208, image209, image210,
image211, image212, image213, image214, image215, image216, image217, image218,
image219, image220, image221, image222, image223, image224, image225, image226,
image227, image228, image229, image230, image231, image232, image233, image234,
image235, image236, image237, image238])

```

```

.mosaic()
.select("constant");

```

//2.1 Para visualizar la capa COPERNICUS (Parámetros iguales a Google)

```

function maskS2clouds(image) {
  var qa = image.select('QA60');
  var cloudBitMask = 1 << 10;
  var cirrusBitMask = 1 << 11;

  var mask = qa.bitwiseAnd(cloudBitMask).eq(0)
    .and(qa.bitwiseAnd(cirrusBitMask).eq(0));

  return image.updateMask(mask).divide(10000);
}

```

```

var dataset = ee.ImageCollection('COPERNICUS/S2_SR')
  .filterDate('2020-01-01', '2021-10-30')
  // Pre-filter to get less cloudy granules.
  .filter(ee.Filter.lt('CLOUDY_PIXEL_PERCENTAGE',20))
  .map(maskS2clouds);

```

```

var dataset2=dataset.mean().clip(muni_area);

```

```

var visualization = {

```

```

    min: 0.0,
    max: 0.3,
    bands: ['B4', 'B3', 'B2'],
};

```

```

Map.addLayer(dataset2, visualization);

```

```

// 3. Agregar capas con su formato correspondiente

```

```

var c1 = ui.Map.Layer(OriginalPob, {}, "Capa de valores Originales", false);
var c2 = ui.Map.Layer(mos, DistVis, 'Distribución de Población', true);
var c3 = ui.Map.Layer(limites, {palette: 'black'}, 'Municipios', false);
Map.add(c1);
Map.add(c2);
Map.add(c3);

```

```

// 4. APP

```

```

// 4.1 Encabezado

```

```

var header = ui.Label('KALMAN-FH',
    {fontSize: '25px', fontWeight: 'bold', color: '4A997E'});

```

```

// 4.2 Texto

```

```

var text = ui.Label(
    'Distribución de la población de acuerdo con el modelo Kalman-FH',
    {fontSize: '15px'});

```

```

// 4.3 Se unen el encabezado y el texto

```

```

var panel = ui.Panel({
    widgets:[header, text],//uno los puntos 4.1 y 4.2
    style: {width: '300px',position:'middle-right'}});

```

```

// 4.4 Se crea otro panel abajo

```

```

var intro = ui.Panel([
    ui.Label({
        value: '_____ ',
        style: {fontWeight: 'bold', color: '4A997E',
        },
    }),
    ui.Label({
        value:'Capas',
        style: {fontSize: '15px', fontWeight: 'bold'}
    })]);

```

```

// 5. Se unen los paneles
panel.add(intro);

// 6. Mandar el panel a la interfaz, se usa: ui.root.
ui.root.insert(1,panel);

// 6.1 Construir Casillas de Verificación
var extCheck = ui.Checkbox('Valores originales (Autor: JSilvan)').setValue(false); //false =
unchecked
var extCheck2 = ui.Checkbox('Distribución poblacional').setValue(true);//
var extCheck3 = ui.Checkbox('Zona de estudio').setValue(false);//

// 6.2 Se agregan al panel
panel.add(extCheck)
    .add(extCheck2)
    .add(extCheck3)
//    .add(br)

// 7. Funciones para activar/desactivar las capas
var doCheckbox = function() {
    extCheck.onChange(function(checked) {
        c1.setShown(checked)
    })
}
doCheckbox();

var doCheckbox2 = function() {
    extCheck2.onChange(function(checked) {
        c2.setShown(checked)
    })
}
doCheckbox2();

var doCheckbox3 = function() {
    extCheck3.onChange(function(checked) {
        c3.setShown(checked)
    })
}
doCheckbox3();

```

```
//7.1 Crea panel y agrega al mapa
var inspector = ui.Panel([ui.Label(
    {value: 'Clic en mapa para para obtener el valor',
    style: {fontWeight: 'bold',
    textAlign: 'center',
    backgroundColor: 'rgba(255, 255, 255, 0)'
    },
    })
]);
```

```
inspector.style().set('backgroundColor', 'rgba(255, 255, 255, 0.5)')
Map.add(inspector);
```

```
//7.2 Efecto al momento de darle click
Map.onClick(function(XY) {
    inspector.widgets().set(0, ui.Label({
        value: 'Calculado ...',
        style: {color: 'FF0000',
            fontWeight: 'bold',
            textAlign: 'center',
            backgroundColor: 'rgba(255, 255, 255, 0)'
        }
    }
}));
```

```
//8. Extrae el valor según las coordenadas al dar clic en mapa
var punto = ee.Geometry.Point(XY.lon, XY.lat);
var capa = mos;
var valor = capa.sample(punto, 30);
var computedValue = valor.first().get('constant');
```

```
//8.1 Regresa el valor en el servidor
computedValue.evaluate(function(result) {
    inspector.widgets().set(0, ui.Label({
        value: 'Distribución poblacional: ' + result.toFixed(2),
        style: {color: "#3498DB",
            fontWeight: 'bold',
            textAlign: 'center',
            textDecoration: 'underline',
            backgroundColor: 'rgba(255, 255, 255, 0)'
        },
    }
}));
});
});
```