

Encuesta de Movilidad Domiciliaria 2014

Base usuario

Resumen metodológico

Buenos Aires, agosto de 2015

Contenido

1. Características generales del relevamiento.....	3
2. Objetivos	4
3. Aspectos conceptuales	5
4. Selección de viviendas a encuestar y expansión de la muestra.....	7
5. Ingreso de datos.....	14
6. Recuperación de información faltante y controles de calidad.....	16
7. Tratamiento de datos faltantes	18
8. Georreferenciación de los datos relevados	20

1. Características generales del relevamiento

La Encuesta de Movilidad Domiciliaria 2014 (EnMoDo-2014) se realizó en la Región Metropolitana de Buenos Aires, entre los días 20 de agosto y 12 de diciembre de 2014, a través de una muestra probabilística, polietápica y estratificada extraída de la Muestra Maestra Urbana de Viviendas de la República Argentina¹. De esta muestra maestra, construida a partir del Censo Nacional de Población, Hogares y Viviendas 2010, fueron seleccionadas alrededor de 24.500 viviendas.

El operativo fue desarrollado por la Dirección de Estudios de Ingresos y Gastos de los Hogares, dependiente de la Dirección Nacional de Estadísticas de Condiciones de Vida del Instituto Nacional de Estadística y Censos a partir de un Plan de Acción Conjunto firmado a solicitud del Ministerio del Interior y Transporte de la Nación.

El área de estudio, denominada Región Metropolitana de Buenos Aires, está compuesta por la Ciudad Autónoma de Buenos Aires y 43 partidos de la provincia de Buenos Aires (Almirante Brown, Avellaneda, Berazategui, Berisso, Brandsen, Campana, Cañuelas, Ensenada, Escobar, Esteban Echeverría, Exaltación de la Cruz, Ezeiza, Florencio Varela, General Rodríguez, General San Martín, Hurlingham, Ituzaingó, José C. Paz, La Matanza, La Plata, Lanús, Las Heras, Lobos, Lomas de Zamora, Luján, Malvinas Argentinas, Marcos Paz, Mercedes, Merlo, Moreno, Morón, Navarro, Pilar, Presidente Perón, Quilmes, San Fernando, San Isidro, San Miguel, San Vicente, Tigre, Tres de Febrero, Vicente López y Zárate).

¹ “Muestra Maestra Urbana de Viviendas de la República Argentina”, documento de trabajo presentado en la I Reunión del Grupo de Trabajo Sobre Encuestas a Hogares de la Conferencia de Estadística de las Américas el 15 y 16 de octubre de 2012 en Buenos Aires, Argentina.

2. Objetivos

Esta encuesta tuvo como objetivo relevar información sobre el modo en que se realizan los desplazamientos (tenencia de vehículos en el hogar, características de los viajes realizados y composición de las etapas de los mismos, gastos en transporte) de la población residente en la Región Metropolitana de Buenos Aires, compuesta por la Ciudad Autónoma de Buenos Aires y 43 Partidos del Conurbano Bonaerense.

3. Aspectos conceptuales

Se consideró **viaje** a aquellas actividades que requieran un desplazamiento mayor de dos cuadras (es decir, tres o más) y que se realizan por un motivo determinado entre un *origen* y un *destino*. Los viajes relevados son los que el entrevistado realizó dentro del *período de referencia* de la encuesta, es decir, entre las 4:00 de la mañana del día anterior a la visita, y las 4:00 de la mañana del día en que se realizó la entrevista.

Todos los viajes registrados en la Encuesta tienen su origen o destino en la Región Metropolitana de Buenos Aires. Si una persona no se encontraba en el área de estudio el día anterior a la entrevista y realizó viajes fuera de ésta, se considera que no viajó y se registró el motivo de no viaje correspondiente.

Un **viaje** puede tener una o más **etapas**, según los modos de transporte utilizados. Así, cada vez que se cambie de modo de transporte o cada vez que se cambie de línea de colectivo o ferrocarril será una nueva etapa del viaje. Respecto del transporte público, la etapa puede pensarse como cada vez que una persona debe abonar el boleto. Así, bajarse de un colectivo y volver a subirse a otro para llegar del hogar al trabajo constituye un viaje con dos etapas, mientras que realizar el mismo recorrido pero combinando dos líneas de subtes sería una única etapa.

También es posible que una persona haya realizado viajes caminando. Las caminatas de tres cuadras o más serán consideradas como un viaje con una etapa. En este sentido, es importante aclarar que sólo se consideró que una etapa fue realizada a pie cuando todo el viaje haya se realizó de ese modo. Las cuadras caminadas antes y después de acceder a un modo de transporte se indagaron en otro capítulo del cuestionario y, a los efectos de la Encuesta, no tienen carácter de etapa.

Los **ingresos corrientes** de los hogares están constituidos por el conjunto de ingresos que perciben mensualmente los miembros del hogar por su participación en el proceso productivo al realizar alguna actividad económica, por la percepción de transferencias corrientes (ya sea públicas como

jubilaciones, pensiones y asignaciones familiares o privadas) y/o rentas de la propiedad.

La Encuesta relevó también información sobre las **características de las viviendas** así como las características **demográficas, educacionales y ocupacionales** de sus miembros.

4. Selección de viviendas a encuestar y expansión de la muestra

En la EnMoDo-2014 se utilizó una muestra de viviendas probabilística, estratificada y multietápica según las características determinadas por la Dirección Nacional de Planificación y Coordinación del Transporte que, de acuerdo al Plan de Acción Conjunto, actúa en representación del Ministerio del Interior y Transporte de la Nación.

Estas características se refieren fundamentalmente a la estratificación del dominio bajo estudio en macrozonas definidas a partir de una clasificación de los radios censales (tarea realizada en función del estudio previo realizado en el año 2008/09) y al número mínimo de encuestas efectivas por macrozona, fijada en 30 por la mencionada Dirección Nacional en función de los requerimientos del modelo de transporte para el cual esta encuesta será insumo.

Dominios de estudio

Los dominios para los cuales se requieren estimaciones separadas son las llamadas *macrozonas*, agrupamientos de radios elaborado por la Secretaría de Transporte. En total se definieron 347 macrozonas, de tamaño variable.

Cuadro 1

Macrozonas	Población
Con menor población	3311
Con mayor población	90848
Percentil 10	19909
Percentil 25	25852
Percentil 75	47025

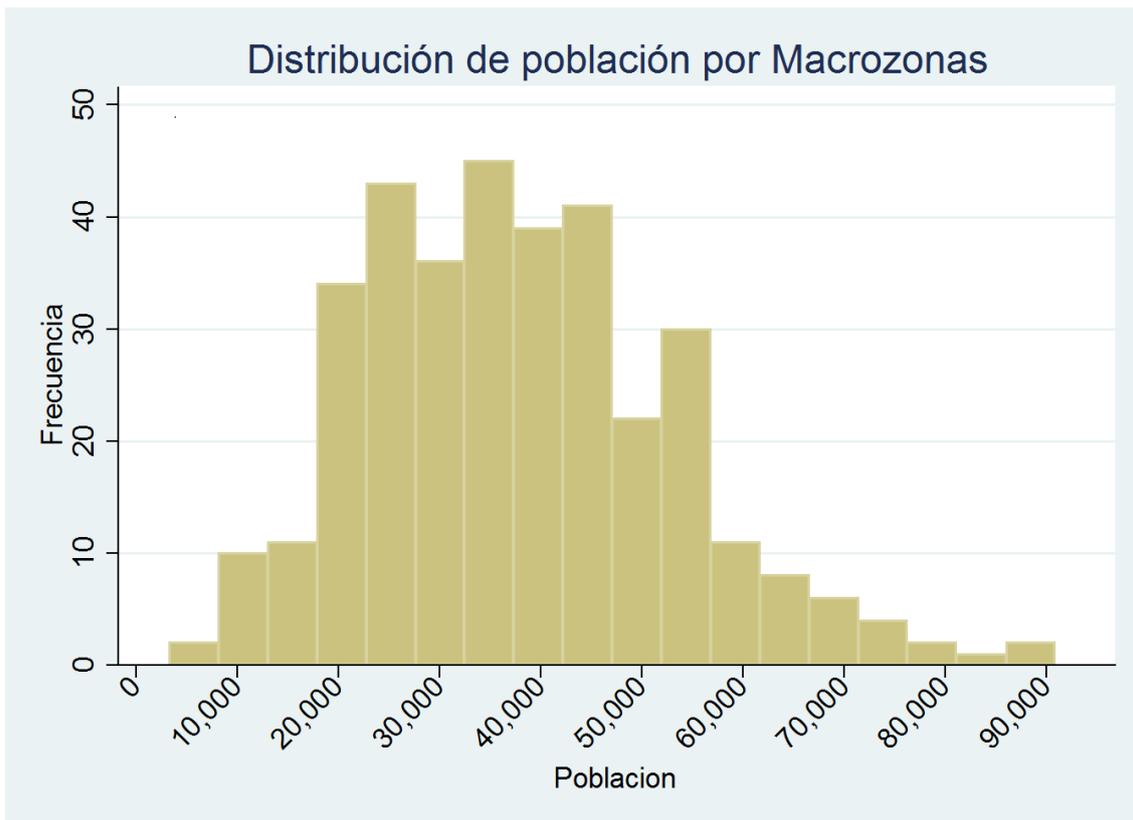


Figura 1: Distribución de población por Macrozonas.

- La Macrozona con mayor población es la nº 641203. Su población es de 90.848 personas y está constituida por 72 radios de la Provincia de Buenos Aires pertenecientes al departamento nº 412.
- La Macrozona con menor población es la nº 200106. Su población es de 3.311 personas y está constituida por 45 radios pertenecientes a la Ciudad Autónoma de Buenos Aires; 43 del ellos corresponden al departamento nº 1 y los dos restantes al departamento nº 2.

Tamaño de muestra mínimo requerido por macrozona

A los efectos de permitir estimaciones confiables, se solicitó relevar en cada macrozona un mínimo de 30 viviendas. En función de ello, el diseño permite que el rendimiento de la muestra tolere hasta un valor del 50%.

Este estudio presenta una cantidad de dominios de estimación (347) mucho mayor que las encuestas a hogares tradicionales, la EPH por ejemplo, lo que

implica la necesidad de garantizar una cantidad mínima de viviendas efectivas relevadas en cada dominio de estimación (macrozonas).

Tamaño de muestra teórico

Dada la heterogeneidad de las macrozonas en cuanto a población, se decidió otorgar un tamaño de muestra mayor según la población de las mismas.

A continuación se presenta el tamaño muestral fijado por estrato agrupado en los siguientes tramos.

Cuadro 2: Cantidad de manzanas seleccionadas por tramo de población

Tramo	Nº de Casos	n	Población
1	246	2.460	7.382.805
2	83	1.245	4.380.426
3	13	260	909.189
4	5	150	422.075
Total	347	4.115	13.094.495

Tramo 1: Macrozonas con población menor a 45000: 10 manzanas seleccionadas

Tramo 2: Macrozonas con población entre 45000 y 65000: 15 manzanas seleccionadas

Tramo 3: Macrozonas con población entre 65000 y 80000: 20 manzanas seleccionadas

Tramo 4: Macrozonas con población mayor o igual a 80000: 30 manzanas seleccionadas

Marco de muestreo

Como marco de muestreo se tomó el conjunto de Radios y Manzanas censales definidas a partir de las bases de microdatos del Censo de Población y Viviendas del 2010 de la República Argentina.

Unidades de primera etapa (UPE)

Las UPE estuvieron constituidas por manzanas. En cada UPE se seleccionaron seis viviendas contiguas para la administración de la encuesta.

Tomando en consideración las características determinadas por la Dirección Nacional de Planificación y Coordinación del Transporte, que establecía un total de 347 macrozonas, de tamaño variable, la cantidad teórica total de viviendas a seleccionar es de 24.690 viviendas.

Estratificación de las UPE

Las unidades de primera etapa se estratificaron en macrozonas, según información proporcionada por el Ministerio de Interior y Transporte. Estas macrozonas son tanto estrato de selección como dominio de estimación.

Selección de la muestra en la primera etapa

La selección de unidades primarias (manzanas) se realizó mediante un muestreo estratificado, con una selección sistemática con probabilidad de inclusión proporcional al número de viviendas por manzana del Censo 2010.

Si el número de manzanas en el estrato h es N_h y el tamaño de la muestra de manzanas a seleccionar en él es n_h , la probabilidad de selección de una

manzana i en el estrato h será $\frac{n_h v_{hi}}{\sum_{i=1}^{N_h} v_{hi}}$, donde v_{hi} es la cantidad de

viviendas en la manzana i del estrato h . El factor de expansión de la primera

etapa, inversa de esa expresión, será $\frac{\sum_{i=1}^{N_h} v_{hi}}{n_h v_{hi}}$.

Unidades de segunda etapa (USE)

Las USE estuvieron constituidas por viviendas particulares.

Selección de la muestra de segunda etapa

La selección de la segunda etapa se hizo de dos formas diferentes, según el dominio en cuestión.

Selección de unidades de segunda etapa en CABA. Manzanas con más de 300 viviendas

En estas manzanas, se seleccionó primero un lado de la manzana con probabilidad de selección proporcional al número de viviendas:

si L_{hi} es el número de lados en la manzana i en el estrato h y v_{ij} es la cantidad censal de viviendas del lado j de la manzana i en el estrato h , entonces la probabilidad de selección del lado j será $\frac{v_{ij}}{\sum_{j=1}^{L_{hi}} v_{ij}}$.

Dicho lado se listó en forma exhaustiva (la totalidad de las viviendas particulares de dicho lado) para luego seleccionar mediante un muestreo aleatorio simple un segmento de seis viviendas contiguas, con una selección eventual suplementaria en dicho lado de la manzana para lograr el objetivo solicitado por la Dirección Nacional de Planificación y Coordinación del Transporte. Con lo cual la probabilidad de selección de una vivienda k (con 6 viviendas seleccionadas), en el lado j , en la manzana i del estrato h , es igual a

$$\frac{6}{v_{ij}^*} .$$

El factor de expansión de la segunda etapa, la inversa del producto de las dos

probabilidades de selección, será entonces: $\frac{\sum_{j=1}^{L_{hi}} v_{ij}}{v_{ij}} \cdot \frac{v_{ij}^*}{6}$.

Manzanas de CABA con menos de 300 viviendas y manzanas en partidos de la Provincia de Buenos Aires

En estas manzanas se efectuó un listado exhaustivo de todas las viviendas particulares en todos los lados de cada manzana. Es a partir de este listado que se seleccionó un segmento de seis viviendas contiguas mediante un muestreo aleatorio simple, con un suplemento eventual -en caso de requerirse-, para lograr el objetivo solicitado por la Dirección Nacional de Planificación y Coordinación del Transporte. Por lo tanto, si el total de viviendas particulares del Censo 2010 en una manzana **i** en el estrato **h** es M_{hi} , la probabilidad de selección (con 6 viviendas seleccionadas) de una vivienda **k** será $\frac{6}{M_{hi}}$.

El factor de expansión de la segunda etapa es la inversa de esa expresión, $\frac{M_{hi}}{6}$. Finalmente, el factor de expansión de la vivienda es igual al producto de los factores de expansión de primera y segunda etapa.

Estadísticas de la muestra seleccionada

Cuadro 8: Muestra final seleccionada por tramo de población.

Tamaño poblacional de la macrozona	Cantidad de macrozonas	Cantidad de manzanas	Cantidad De Viviendas	Cantidad de población en la macrozona
Menor a 45.000	246	2.435	14.698	7.382.805
Entre 45.000 y 64.999	83	1.220	7.367	4.380.426
Entre 65.000 y 79.999	13	258	1.543	909.189
80.000 y más	5	144	864	422.075
Total	347	4.066	24.472	13.094.495

5. Ingreso de datos

El sistema de ingreso de datos EnMoDo-2014 fue diseñado en función de la complejidad del flujo del cuestionario y la necesidad de numerosos controles de consistencia de datos.

La interfaz gráfica fue desarrollada para el acceso a través de un navegador web (PHP, HTML y JavaScript), prescindiendo así de la instalación de aplicaciones ad-hoc en las máquinas cliente. La arquitectura cliente-servidor permitió mantener un seguimiento centralizado del avance de carga sin interrumpir el trabajo de los Ingresadores.

El almacenamiento de los datos ingresados se realizó sobre una base de datos PostgreSQL. Esta arquitectura ha dado resultados ampliamente satisfactorios en más de media docena de operativos de alcance nacional.

Los datos se ingresaron página a página, mostrando en pantalla una estructura muy similar a la del cuestionario en papel. Cada una de las páginas cargadas fue almacenada inmediatamente en la base de datos, minimizando así la pérdida de información ante eventuales problemas de infraestructura durante el ingreso. El sistema aceptó únicamente el ingreso de valores de respuesta que se muestran en el cuestionario en papel, en el orden el flujo entre preguntas.

El análisis de consistencias advirtió sobre los errores de carga o completado de cuestionario que no hayan sido capturados en etapas anteriores. En caso de encontrar datos inválidos, el Ingresador fue advertido e inhabilitado para cerrar la carga del cuestionario hasta tanto no solucionara el o los problemas detectados. El control de los saltos entre preguntas se realizó durante el ingreso y de forma automática, evitando la alteración del flujo del cuestionario y la carga de respuestas en celdas equivocadas.

Los datos ingresados pasaron por tres etapas de control antes de que cada cuestionario se considerara cerrado:

1. Detección de interrupciones y errores en el flujo del cuestionario, preguntas sin respuesta y respuestas que exceden los valores mínimos y máximos estipulados.
2. Validación de nombres de hitos, estaciones de tren/subte, jurisdicciones e intersecciones de calles.
3. Verificación de la coherencia entre preguntas del cuestionario.

Las políticas de seguridad implementadas evitaron que personas ajenas al operativo tengan acceso a la información cargada. Los usuarios del sistema, cada uno con su respectiva contraseña, estuvieron limitados por un conjunto de permisos y restricciones para leer, modificar y/o eliminar datos, extraer reportes de avance, alta y baja de usuarios, etc.

6. Recuperación de información faltante y controles de calidad

Durante todo el período de relevamiento de datos se procedió a la recuperación de manera presencial (recuperación en campo) y telefónica de información faltante en aquellos casos en que correspondiera, pudiendo desagregarla de la siguiente manera:

- 1) Información del Bloque del Hogar: los datos que tuvieran algún tipo de consistencia o donde no existiera un valor aceptado dentro de los parámetros acordados con los equipos técnicos del Ministerio del Interior y Transporte se recuperaron de las dos maneras mencionadas anteriormente.
- 2) Información del Bloque Individual: en este caso, la ausencia de información puede corresponder a dos categorías diferentes en su tratamiento:
 - a. Ausencia de información parcial: cuando un dato de algún viaje de los miembros presentaba errores de captación o no se encontraba registrado en el cuestionario, se procedió a su recuperación de manera presencial y telefónica. Cuando esto no fue posible, se registró una no respuesta (Ns/Nr) en la pregunta.
 - b. Ausencia de información total: cuando la totalidad de la información de uno o más viajes de un miembro no se encontraba registrada en el cuestionario, se indagó nuevamente a los miembros del hogar que correspondieran, teniendo en cuenta que los datos a recolectar fueran del día anterior al momento de la entrevista. El mismo tratamiento se aplicó para aquellos casos donde la ausencia de información no era total pero los datos con valores faltantes impedían que los viajes fueran considerados válidos (por ejemplo: un miembro que fue entrevistado pero no respondió los lugares de destino de uno o más viajes debía ser entrevistado nuevamente para registrar los viajes del día anterior a esta última entrevista).

Una vez relevada la información por el personal de campo, se desarrolló un mecanismo recursivo de validación de datos en gabinete que consistió en las siguientes etapas:

Recepción de los cuestionarios

Un equipo de recepcionistas, capacitados especialmente para la tarea, recibió los cuestionarios de manos de Encuestadores y Recuperadores, revisando la consistencia de los datos relevados en campo. En caso de datos faltantes o con algún error, procedieron de acuerdo a los criterios mencionados anteriormente.

Análisis de la información relevada en los cuestionarios

Personal técnico con experiencia en el análisis de consistencia de numerosas encuestas a hogares realizó un análisis en profundidad de una serie de lotes de cuestionarios elegidos de manera aleatoria entre el total de la información, detectando patrones de errores por parte de los Encuestadores y Recuperadores, así como potenciales omisiones en el control de los Recepcionistas. En los casos en los que fue necesario, se recapacitó al personal.

Ingreso de datos

Los cuestionarios se ingresaron en un sistema diseñado para cumplir con los estándares de calidad requerida en la Encuesta de Movilidad Domiciliaria, que comprende una batería de más de 150 consistencias entre aquellas consideradas advertencias, errores y fatalidades, según su importancia. El carácter recursivo del sistema de validación de datos surge a partir del regreso a gabinete de la información considerada como errónea por el sistema de ingreso, momento en el cual se procedió a su corrección utilizando los métodos mencionados anteriormente.

7. Tratamiento de datos faltantes

Los únicos datos imputados en la Encuesta fueron los correspondientes al ingreso total del hogar, para aquellos casos en los que la unidad de análisis entrevistada no declaró el monto de ingreso ni el rango en el que se encontraba dicho monto. La base cuenta con 14.516 registros, de los cuales 9.573 no registraban valor en la variable de monto. De estos, 4.780 tampoco presentaban respuesta en la variable de rango.

La imputación se realizó mediante el método Hot Deck. Todas las unidades de la muestra se clasificaron en grupos disjuntos, lo más homogéneos posible. Para cada valor faltante, se seleccionó aleatoriamente un valor observado de un registro o unidad donante: se aceptaron como donantes aquellos registros con ingresos mayores a 100. Los hogares imputados no se tomaron como donantes en ningún caso.

Para determinar el donante se tomó un conjunto de variables explicativas iniciales que coincidieran con las del caso a imputar. Dicho conjunto se fue reduciendo hasta imputar el total de valores faltante. Esta metodología se basa en la premisa de que -dentro de cada grupo de clasificación- las unidades no respondientes tienen la misma distribución que aquellas que responden.

Las variables explicativas que se usaron para esta imputación son:

- Donante (aquellos registros que declararon tener monto de ingreso mayor que 100 y no fueron imputados)
- Rango de ingreso
- Código de provincia
- Código de departamento
- Código de macrozona
- Cantidad de componentes, reagrupando en la categoría 5 aquellos hogares de 5 o más componentes
- Cantidad de personas que trabajan en el hogar

- Nivel de instrucción del jefe del hogar
- Tipo de vivienda
- Condición de ocupación del jefe de hogar
- Edad del jefe de hogar reagrupada
- Sexo del jefe de hogar

Se comenzó por imputar la variable de monto para aquellos registros que contaban con un valor en la variable de rango. A continuación, se realizó la imputación de los registros con valor faltante tanto en la variable de monto como en la de rango. Una vez imputada en su totalidad la variable de monto, se imputa la variable de rango por definición usando la siguiente tabla.

Categoría	Rango de ingreso
0	Sin ingresos
1	\$1 a \$1.000
2	\$1.001 a \$1.500
3	\$1.501 a \$2.000
4	\$2.001 a \$2.500
5	\$2.501 a \$3.000
6	\$3.001 a \$3.500
7	\$3.501 a \$4.000
8	\$4.001 a \$4.500
9	\$4.501 a \$5.000
10	\$5.001 a \$6.000
11	\$6.001 a \$8.000
12	\$8.001 a \$10.000
13	\$10.001 a \$12.000
14	\$12.001 a \$15.000
15	\$15.001 a \$18.000
16	\$18.001 a \$23.000
17	\$23.001 y más

8. Georreferenciación de los datos relevados

El proceso de asignación de coordenadas a los puntos origen y destino de viajes y etapas relevados en campo se realizó en tres etapas, que se detallan a continuación:

- a) Estandarización de los nombres de calles e hitos y trazado de nuevos nodos: en este momento, se corrigió toda la información ingresada en el sistema que no se encontrara en la matriz de calles o de hitos por error en el ingreso de la información. Del mismo modo, se incorporó a la capa de calles aquellas intersecciones no contempladas en la matriz original.
- b) Localización de alturas numéricas y con caracteres: posteriormente a la estandarización de aquellos puntos ubicables por una intersección o por el nombre de un hito, se procedió a localizar los puntos que tuvieran altura numérica o en caracteres (como, por ejemplo, el kilometraje de una ruta).
- c) Incorporación de coordenadas a la base de datos: una vez georreferenciados todos los puntos, se crearon dos variables específicas en las tablas Viajes y Etapas que permiten observar la longitud y la latitud de cada punto relevado. Los hitos y las estaciones, junto con su respectivas coordenadas geográficas, fueron provistos por el Ministerio de Interior y Transporte. Para las intersecciones se hizo uso la herramienta de OpenStreetMap. El sistema de coordenadas geográficas utilizado fue WGS 1984, EPSG: 4326, conjuntamente con la cartografía del Censo Nacional de Población, Hogares y Viviendas 2010.