Universidad Autónoma del Estado de México

Facultad de Ciencias de la Conducta
Doctorado en Ciencias de la Salud

ATLAS DE SALUD Y RIESGO AMBIENTAL EN LA CIUDAD DE MÉXICO

T E S I S

para obtener el grado de:

Doctora en Ciencias de la Salud

PRESENTA:

MTRA. KAROL BACA LÓPEZ

COMITÉ TUTORIAL:

Tutor académico:
Miguel Ángel Camacho López
Tutor interno:
Dra. Miriam Verónica Flores Merino
Tutor externo:
Dr. Enrique Hernández Lemus

Toluca, Estado de México, Mayo 2022

Índice general

Introducción

La salud de cada individuo depende de factores como su genética, entorno físico, comportamiento (i.e. alimentación y ejercicio) así como de condiciones sociales. Cada uno de estos factores o bien, la combinaciones de estos, puede ser la causa o agravamiento de enfermedades. A su vez, conjuntos de personas que comparten un espacio geográfico, son susceptibles de estar expuestos a los mismos factores de riesgo. Dado el número de factores y su compleja interacción, no es posible definir cómo van a afectar al individuo o su comunidad en el corto o largo plazo.

Si bien no es factible estudiar el efecto de cada uno de los factores de riesgo en un individuo y de la misma manera no lo es para una comunidad, un primer esfuerzo por contribuir a esta problemática es identificar cuáles de estos factores son los que afectan en mayor grado a una mayor proporción de la población. Además, es importante mencionar que no todos estos factores pueden ser alterados o controlados para mejorar la condición de salud de una persona o población. Por esta razón, las instituciones encargadas de atender las problemáticas que atentan contra la salud pública y el bienestar de la población, deben enfocarse en aquellos factores que, puedan ser atendidos a través de programas y políticas públicas.

Para desarrollar e implementar programas sociales más efectivos, los gobiernos de las grandes ciudades y metrópolis están invirtiendo cada vez más recursos en infraestructura para generar datos confiables que puedan ser utilizados en la toma de decisiones. Con este fin, herramientas como los atlas de salud y/o ambientales son muy útiles, siempre y cuando se cuente con la variedad y calidad de datos necesarios. Estas herramientas son visualizaciones de diferentes capas de información (datos) que al integrarse, permiten a los usuarios identificar regiones geográficas de mayor riesgo (ambiental y/o social) así como las posibles relaciones entre las distintas variables de su entorno.

Hoy en día, es responsabilidad y prioridad del gobierno y las distintas instancias a nivel nacional, estatal y local, tomar acción sobre problemáticas relacionadas con los efectos nocivos de la contaminación del aire. En las últimas décadas, se han estudiado ampliamente los efectos negativos a la exposición a contaminantes del aire y se ha sido vinculado a enfermedades respiratorias, cardiovasculares, cáncer, déficit cognitivo, nacimientos prematuros entre otros. Además, estas afectaciones aumentan y tienen más

impacto en grupos vulnerables como los niños y los adultos mayores. Es importante mencionar las diferentes patologías o agravamiento de las mismas, puede ser resultado de exposición a concentraciones elevadas por un corto periodo de tiempo o bien, bajas concentraciones en periodos largos de exposición.

La población más afectada por la mala calidad del aire es principalmente la que se encuentra en las grandes ciudades y Metrópolis. Estas ciudades comparten problemáticas como el incremento en la densidad poblacional, industrialización, tráfico vehicular y carecen de programas eficientes para mitigar los altos niveles de polución que éstas generan. La Ciudad de México y su zona metropolitana no es la excepción, e incluso puede considerarse uno de los mejores casos de estudio, como una gran urbe que conjunta múltiples fuentes de contaminantes y una gran heterogeneidad en las características de su población. En particular, en la última década se ha observado una tendencia a la alta en la incidencia de diferentes enfermedades crónicas como el cáncer y afecciones cardio respiratorias para los diferentes grupos etarios. A su vez, esto sugiere que el número de muertes en la población asociadas a dichas enfermedades también está en aumento.

Con la finalidad de coadyuvar a la toma de decisiones en materia de Salud Pública, en particular, disminuir la tasa de mortalidad causa específica por exposición a elevadas concentraciones de contaminantes del aire. En este proyecto se desarrolló una metodología que consiste en i) la colección de información de bases de datos públicas de forma semi-automática, ii) el procesamiento de información, iii) interpolación espacio-temporal de concentraciones de contaminantes y tasas de mortalidad y iii) visualización de las tendencias de las diferentes variables a diferentes niveles de granularidad en tiempo (trimestre y año) y espacio (colonia y delegación). A saber, se estudiaron todos los contaminantes criterio monóxido de carbono (CO), dióxido de nitrógeno (NO_2), ozono (O_3), dióxido de azufre (SO_2), partículas de radio hidrodinámico menor a 2.5 y 10 micras ($PM_{2.5}$ y PM_{10}), óxidos de nitrógeno (NO y NO_X) y partículas gruesas (PM_{CO}) que son monitoreados actualmente por la Secretaría del Medio Ambiente (SEDEMA) a través de la Red de Monitoreo Automático (REDMA) para el periodo 2009-2018. Es importante mencionar que este estudio es el más exhaustivo (a la fecha) que examina la disponibilidad de datos de contaminantes. Ya que algunos estudios sobre predicciones de calidad del aire no toman en cuenta el impacto de la falta de información tanto por estación de monitoreo como en periodos de tiempo.

Finalmente, los valores estimados de concentraciones de contaminantes y tasas de mortalidad mediante la interpolación espacio temporal, pueden ser utilizados como mapas dinámicos en los que se muestre, de forma independiente, cada una de las variables estimadas a diferentes niveles de granularidad. Este es el tipo de herramienta a la que los tomadores de decisiones deben tener acceso para desarrollar e implementar programas más precisos y oportunos en cuanto a prevención de enfermedades y muertes por contaminación ambiental.

Capítulo 1

Antecedentes

1.1. Determinantes Sociales de la Salud

Para poder mejorar la salud de los habitantes de una población, aunque sea en cierta grado, es necesario conocer cuáles son los factores en su entorno y vida diaria que afectan su salud de manera negativa. Estos factores se conocen como Determinantes Sociales de la Salud (SDoH, por sus siglas en inglés) y pueden ser condiciones genéticas, de comportamiento, ambiente, sociales, entre otras. A la fecha, no hay un listado consenso de estos factores o categorías bien definidas. El estudio más extenso, basado en fuentes como la Organización Mundial de la Salud (OMS), es el realizado por la asociación *goinvo* (1) que considera 5 categorías y el porcentaje aproximado con el que cada una contribuye a la salud de un individuo (ver Figura 1.1. De éstas categorías, aquellas sobre las que es posible tener un impacto a través de programas sociales son *Circunstancias sociales*, *Ambiente físico* y *Cuidado médico*.

Para poder crear herramientas, modelos o programas confiables que puedan implementarse y tener el impacto esperado, es necesario contar con datos y estos no suelen estar disponibles para todas las variables de interés. En particular, la contaminación del aire es uno de los factores ambientales susceptibles de ser modificado a través de políticas públicas por lo que se eligió como tema de estudio en esta investigación. En la mayoría de las grandes ciudades, el monitoreo de la calidad del aire es un tema prioritario desde hace algunas décadas y se realiza de manera continua (diaria en la mayoría de los casos). Para México, y en particular para la Ciudad de México y su Zona Metropolitana, éstos datos diarios sobre concentraciones de contaminantes están disponibles públicamente y es posible tener acceso a su histórico. Dependerá del Gobierno e instituciones de Salud, que otra información relevante esté disponible para fines de investigación y puedan incluirse en estudios más completos sobre el impacto de diferentes determinantes de salud para nuestra población.

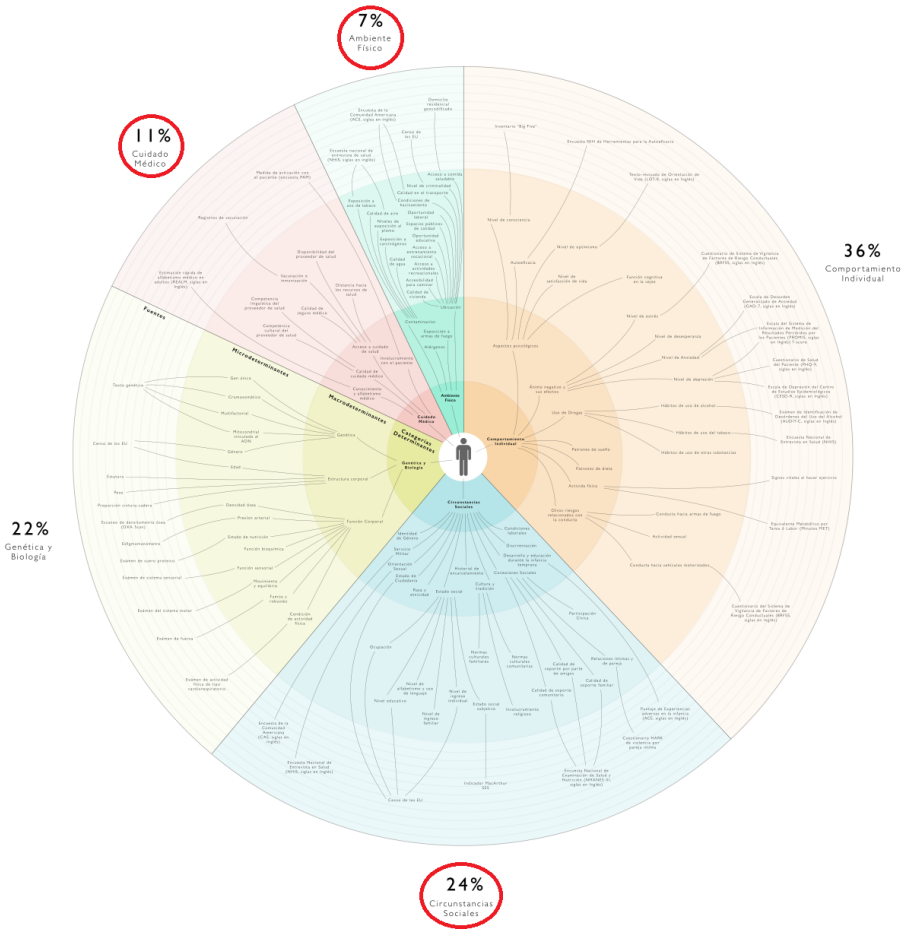


Figura 1.1: Determinantes Sociales de la Salud. Las categorías y porcentajes con el que contribuyen son: i) Genética y Biología (22 %), ii) Circunstancias sociales (24 %), iii) Comportamiento individual (36 %), iv) Ambiente Físico (7 %) y v) Cuidado médico (11 %).

1.2. Calidad del aire y Salud Pública

La exposición a niveles altos de contaminación ha sido considerada la causa principal de enfermedades respiratorias y cardiovasculares. Sin embargo, investigaciones más recientes alrededor del mundo muestran resultados consistentes que vinculan la exposición a contaminantes con el desarrollo de cáncer y complicaciones de otras enfermedades crónicas e incluso, afectaciones de tipo cognitivo. Y es importante mencionar que la contaminación ambiental amenaza principalmente a los más vulnerables como lo son los infantes, mujeres, adultos mayores, personas con padecimientos crónicos y aquellos con falta de acceso a servicios médicos (2).

En el estudio *Global Burden of Diseases 2015* se identificó a la contaminación del

aire como el principal factor de riesgo de salud principalmente en países de bajo y medio nivel económico. El estudio consistió en analizar las tendencias espaciales y temporales en las tasas de morbilidad y mortalidad atribuible a la contaminación a nivel global, regional y nacional. Se encontró que tan sólo la exposición a $PM_{2.5}$ causó 4.2 millones de muertes, representando el 7.6 % del total de muertes a nivel mundial. Mientras que, la exposición a Ozono causó 254 mil muertes más (3).

Por otra parte, la Comisión en Contaminación y Salud (Comission on Pollution and Health) de *The Lancet*, reportó que la contaminación era la causa ambiental de enfermedades y muertes prematuras más importante. En la figura 1.2 se muestran las muertes estimadas a nivel mundial (en millones) para diferentes factores de riesgo como: contaminación total, consumo de tabaco, enfermedades como el SIDA, malaria y tuberculosis, uso de alcohol entre otros. Como se puede observar, la contaminación total, que incluye diferentes tipos como lo es contaminación del aire, suelo, agua y ocupacional contribuye con una mayor cantidad de muertes (datos del año 2015). Lo que justifica el interés de la comunidad científica y de las organizaciones de salud a nivel mundial, nacional y regional. Sorprendentemente, las muertes por contaminación (incluyendo todos los tipos) son mayores que las ocasionadas por uso de alcohol y drogas, malnutrición, accidentes en carretera, guerra (asesinatos) y el ébola en su conjunto.

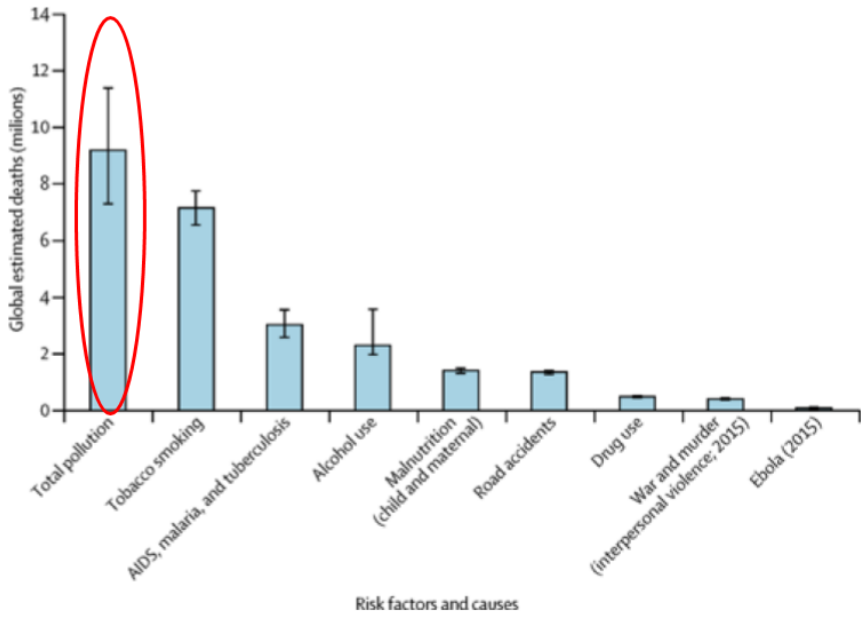


Figura 1.2: Número de muertes estimadas (millones) a nivel mundial ocasionadas por los principales factores de riesgo y causas (4).

Se estima que, tan solo en 2015, los diferentes tipos de contaminación (por agua, aire, suelo y ocupacionales) contribuyeron con el 16 % de muertes a nivel mundial. En las naciones más afectadas, se estima que la contaminación es responsable de más de

una muerte de cada 4. Resultados del mismo estudio indican que más del 90 % de muertes ocasionadas por polución ocurren en países de bajo y medio nivel económico. E incluso, en general, independientemente de la economía del país, la incidencia de enfermedades para las poblaciones vulnerables es mucho mayor.

La contribución al número de muertes a nivel mundial por tipo de contaminación se puede observar en la figura 1.3. Independiente de la fuente de base de datos (Institute for Health Metrics and Evaluations o World Health Organization) es claro que la contaminación del aire cobra más vidas que las causadas por contaminación del agua (en 2do lugar) y que la contaminación de suelo o por exposición ocupacional. Este es uno de los motivos por los cuáles es urgente atender la problemática, sin minimizar la importancia de otros factores de riesgo de morbilidad y mortalidad.

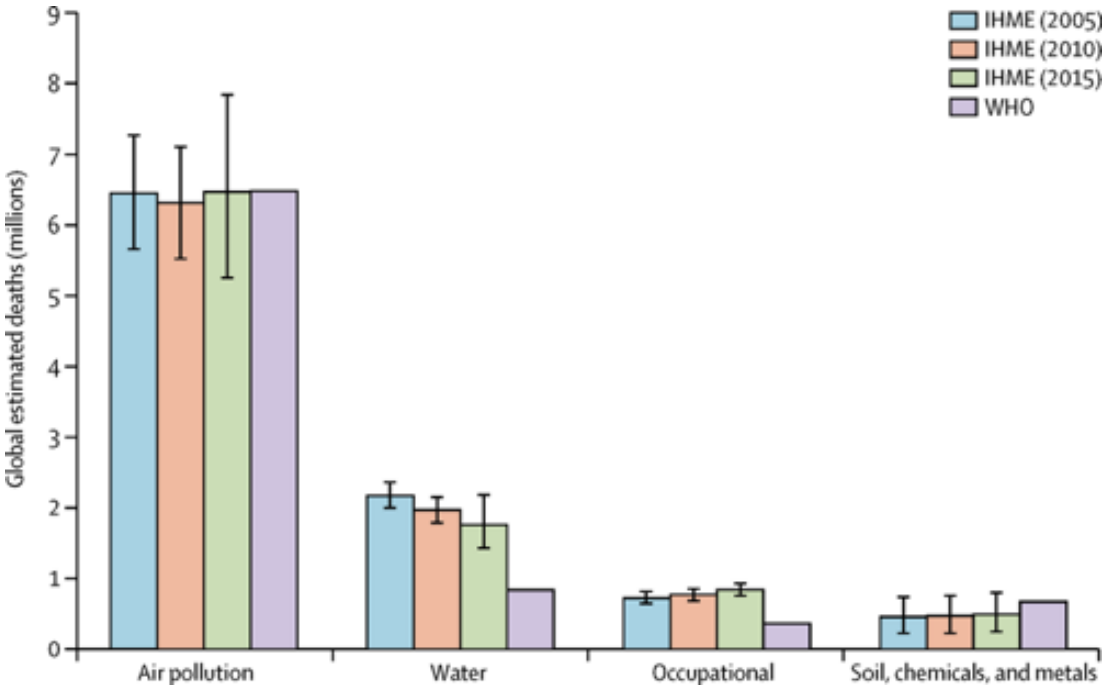


Figura 1.3: Número de muertes estimadas (millones) a nivel mundial por factores de riesgo asociados a contaminación (4).

En particular, la población infantil enfrenta un riesgo más grande ya que incluso, la exposición a muy bajas concentraciones durante ciertas ventanas de vulnerabilidad (en el útero e infancia temprana) pueden ocasionar enfermedades y muerte en la infancia (4). En la figura 1.4 se muestra el número de muertes estimadas para diferentes grupos etarios ocasionados por los diferentes tipos de contaminación. En la mayoría de los grupos de edad, la contaminación del aire es el factor de riesgo predominante, excepto en la etapa postnatal y hasta los 24 años de edad en los que predomina el riesgo por exposición ocupacional (de los padres principalmente). Además es importante observar

que las etapas de mayor riesgo por exposición a contaminantes del aire son la postnatal e incrementa de forma considerable después de los 50 años.

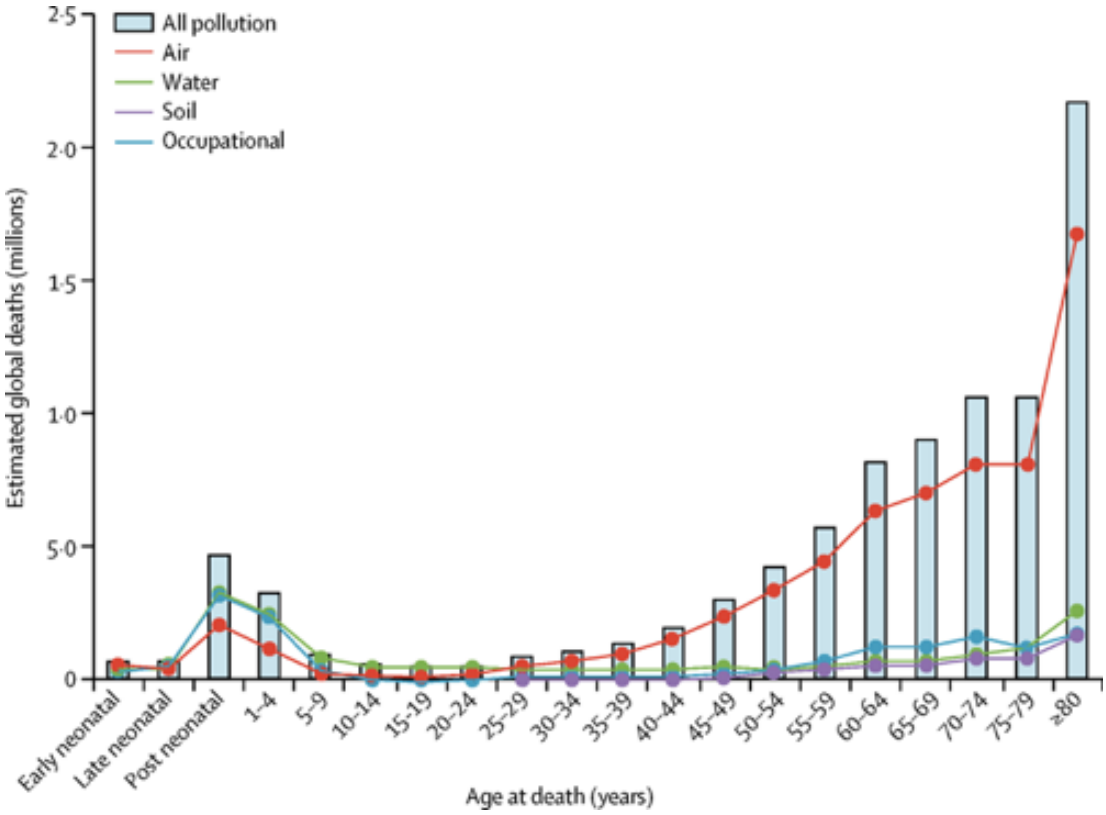


Figura 1.4: Número de muertes estimadas (millones) a nivel mundial causadas por exposición a contaminantes en aire, agua, suelo y por exposición ocupacional para diferentes grupos etarios (4).

A continuación, se explica cuáles son las fuentes predominantes de polución, los contaminantes que producen y cuáles son las afectaciones más frecuentes derivadas de la exposición a éstas en el corto y largo plazo. Estudios recientes señalan a las partículas de radio hidrodinámico menor a 2.5 micras ($PM_{2.5}$)como uno de los agentes más dañinos para la salud. Éstas partículas son generadas principalmente en el proceso de combustión de combustibles fósiles (quema de carbón y tráfico vehicular principalmente). Y la exposición prolongada a este agente ha sido asociada con el aumento de mortalidad por cardiopatía isquémica (5). Además, este estudio sugiere que la mortalidad asociada a $PM_{2.5}$ varía dependiendo la fuente, y recomienda que para atender esta problemática se debe disminuir la contaminación por tráfico vehicular.

Además se ha reportado que, principalmente entre las partículas de materia, clasificadas como gruesas (PM_{10}, $2.5 - 10\mu m$), finas ($PM_{2.5}$, $< 2.5\mu m$) o ultrafinas ($PM_{0.1}$, $< 0.1\mu m$), las $PM_{2.5}$ se han asociado con el incremento de hospitalizaciones, enfermedad

cardiovascular e infartos. Existe evidencia de que la exposición diaria a concentraciones extremadamente elevadas de PM se asocian con mortalidad cardiovascular y respiratoria. Así también la exposición por cortos períodos a concentraciones bajas de PM finas pueden ocasionar incidentes cerebrovasculares isquémicos e infarto de miocardio. Además, la exposición a monóxido de carbono (CO), dióxido de nitrógeno (NO_2) y dióxido de azufre (SO_2) también han sido asociadas con el riesgo de infarto de miocardio. Estos resultados sirven como evidencia de que periodos de exposición cortos (menores a un día o de un par de días) pueden ser suficiente para ocasionar incidentes cardíacos y en algunos casos, la muerte (6).

Por otra parte, en cuanto a la exposición a PM finas por periodos prolongados, se ha encontrado que se incrementa el riesgo de mortalidad por problemas cardiovasculares y cáncer de pulmón. Además, estudios independientes concluyeron que tanto la exposición a bajas y altas concentraciones por periodos largos, incrementan la mortalidad por problemas cardiovasculares. Y en conjunto con otros contaminantes como NO_2 y SO_2, las muertes por enfermedades respiratorias y cáncer de pulmón también se incrementaron (6).

Recientes investigaciones señalan a $PM_{2.5}$ como responsable de la mayor cantidad de muertes de entre los diferentes contaminantes en el aire. Aunque no ha sido posible determinar, cuál(es) de todos sus componentes son los más dañinos para la salud. En el estudio realizado por Y. Yang y colaboradores (7), identificaron que el carbón negro y el carbón orgánico están asociados con la mortalidad cardiovascular. Otros componentes tóxicos como el nitrato, sulfato, zinc, silicón, hierro, níquel, vanadio y potasio están asociados con problemas cardiovasculares. Mientras que nitrato, sulfato y vanadio están vinculados con problemas respiratorios.

1.3. Estadística espacio-temporal y Salud Pública

Originalmente establecida en 1950 para el estudio de reservas de oro en la industria de la minería, la estadística espacial (o geoestadística) ha sido aplicada a diversos problemas y áreas como la hidrología, calidad del agua, suelos, biología, manejo ambiental y, el área de Salud Pública no es la excepción. En particular, el uso de las herramientas de la estadística espacial en el área de Salud Pública tiene como objetivo dilucidar, por ejemplo, cómo se propaga una enfermedad, cuáles son las regiones afectadas por la enfermedad, así como pronosticar cuáles serán las próximas regiones afectadas y en qué grado. Jef Caers define la geoestadística como "*la rama de las ciencias estadísticas que estudia el fenómeno espacial/temporal y capitaliza las relaciones espaciales para modelar posibles valores de variables en lugares no observados y no muestreados*"(8). En general, son cuatro los tipos de investigaciones en los que se conjuntan la epidemiología y la salud pública:

- Mapeo de enfermedades
- Estudios ecológicos

- Estudios de agrupación de enfermedades
- Evaluación ambiental y monitoreo

Para la variedad de aplicaciones epidemiológicas y de salud que existen, un problema común en los conjuntos de datos es que no se cuenta con todos los valores de la variable de interés ya que no fueron medidos o muestreados en cierta ubicación o ventana de tiempo. Para abordar éstas limitaciones relacionadas con el análisis de bases de datos *incompletas*, se han desarrollado diferentes métodos de interpolación espacial (y que pueden ser aplicados a la componente temporal). Éstos métodos de interpolación son procedimientos que permiten obtener en valor de una variable en el espacio y/o tiempo a partir de valores conocidos en otros puntos del espacio (y tiempo).

1.3.1. Métodos de interpolación para evaluación de la calidad del aire

De entre los métodos de interpolación más utilizados basados en Sistemas de Información Geográfica (SIG) están: *Inverse distance weighted* (IDW), *Radial basis function* (RBF), *Global polynomial interpolation* (GPI) y *Local polynomial interpolation* (LPI). Los cuáles son *métodos de interpolación determinísticos* que consisten en generar una superficie a partir de las mediciones basado en el grado de similitud y suavizado (*smoothing*). Por otra parte, los *métodos geoestadísticos* generan una superficie usando las propiedades estadísticas de los valores de los puntos muestrales tomando en cuenta su configuración espacial alrededor de un nuevo punto de interés. En éstos métodos se asume que cierta diferencia espacial observada puede ser modelada como un proceso aleatorio con autocorrelación espacial. A continuación se describen algunas de las principales aplicaciones de la interpolación espacial en las ciencias ambientales y la problemática de salud pública en la que inciden.

Uno de los métodos más populares para la estimación de concentraciones de contaminantes es el modelo de regresión de uso de suelo (LUR, por sus siglas en inglés) cuya ventaja es que logra capturar variaciones en áreas pequeñas. El trabajo realizado por Giehae C. (2017) es uno de los pocos estudios comparativos entre un modelo LUR vs Nearest Location, Inverse Distance Weighting y Kriging Ordinario. El modelado se realizó en función de variables como el promedio anual de NO_2, datos de tráfico vehicular, uso de suelo y la altitud de las 13 estaciones de monitoreo en la red de monitoreo de la República de Corea. Las estimaciones por estos modelos fueron comparadas con mediciones directas (biomonitoreo) en una cohorte de 969 participantes. Los resultados sugieren que el área de transportación, distancia a áreas residenciales y las áreas de humedales fueron predictores importantes de NO_2. Además de que este método es más efectivo en regiones de estudio pequeñas y con baja disponibilidad de datos (9).

La Ciudad de México también ha sido objeto de estudio para la aplicación de modelos LUR para analizar la distribución espacial de contaminantes. En particular, en el análisis realizado por Hinojosa-Baliño, et al. se estudió la distribución de $PM_{2.5}$ utilizando datos de exposición individual. Además, se incluyeron datos meteorológicos,

demográficos y sociales. El modelo predictivo generado reveló dos áreas de alta concentración (hasta 109.3 $\mu g/m^3$) y dos más con bajas concentraciones (entre 72 y 86.5 $\mu g/m^3$) (10).

Además de la estimación de concentraciones de contaminantes, los métodos de interpolación espacial han sido utilizados para incorporar los patrones de dependencia espacial de tasas de morbilidad y mortalidad en el mapeo de valores de riesgo y en la cuantificación de la incertidumbre asociada. Un ejemplo de esto es la aplicación del método de Poisson Kriging a datos de tasas de mortalidad ajustadas por edad de cáncer de mama y cérvix. Con esto se generó una aplicación (disponible públicamente) con datasets de ejemplo para la generación de mapas de mortalidad. Una de las principales observaciones es que el método de Kriging resulta apropiado cuando existe dependencia espacial en las tasas de mortalidad (11).

Otra de las ventajas del método de Kriging es su utilidad para lidiar con la escasez de datos. Tadesse, et al. realizaron una estimación de la distribución espacial de casos de tuberculosis en Gurage, Etiopía. Esta región, como muchas otras en países subdesarrollados, carecen de recursos para la recolección, manejo de datos e implementación de programas de control. Se obtuvieron datos de 1601 pacientes, datos poblacionales y de geo-localización. Se identificaron las áreas de mayor riesgo sobre muestras de diferentes tamaños (40, 60 y 80 %) con diferentes modelos de variogramas (12). Como mencionan los autores, este tipo de análisis tienen la finalidad de coadyuvar a las autoridades locales a identificar las áreas que requieren más atención e intervención.

1.3.2. Representatividad espacial de estaciones de monitoreo

Como se ha mencionado anteriormente, los estudios para estimar la calidad del aire con fines de prevención, se basan en la medición de concentraciones obtenidas de las estaciones y redes de monitoreo ambiental. Por esta razón, es fundamental que los datos utilizados para dichas estimaciones sean representativos y confiables. Debido a esto, es crucial hacer una evaluación de la representatividad de las mediciones en las redes de monitoreo, previo al uso de los datos para la predicción de la calidad del aire y la toma de decisiones basado en ello. Brevemente, se puede entender que la concentración de un contaminante que se mide en cada punto dentro del área de representatividad no puede diferir del valor medido en la estación de monitoreo por más de un valor umbral dado. Además del umbral de concentración, otros enfoques utilizan parámetros como las fuentes de emisión, variables climáticas y/o topográficas.

La Representatividad espacial (RE) es un parámetro relevante para determinar los sitios de colocación de estaciones de monitoreo y evaluar los efectos en la población por la exposición a largo plazo a contaminantes. Asimismo, es útil para evitar redundancia en las mediciones y tener una red de monitoreo costo-efectiva (13). A la fecha, no hay una metodología consenso para evaluar el *área de representatividad espacial* de una estación de monitoreo.

A la fecha, el estudio más extenso sobre representatividad espacial es el realizado

por el Foro de Modelado de Calidad del Aire en Europa (FAIRMODE, por sus siglas en inglés) de la Comisión Europea que contempla 14 países. El ejercicio consistió en un estudio comparativo de 25 métodos de evaluación de RE para PM_{10} y NO_2 en un sitio de tráfico y para PM_{10}, NO_2 y PM_{10} y O_3 en dos sitios urbanos de fondo. Esto con la finalidad de facilitar el reporte de la RE de las redes de monitoreo de los países miembros. Los resultados muestran una variabilidad considerable entre las estimaciones realizadas con cada método tanto en la extensión y posición de las áreas de representatividad, como del input de datos utilizados de manera efectiva. Derivado de este análisis se enfatiza la necesidad de contar con un concepto sólido de área de representatividad y contar con criterios consistentes para el cálculo de la misma (14). Desde la propuesta de este estudio comparativo en 2015 y la obtención de los primeros resultados en 2017, una serie de metodologías han sido desarrolladas alrededor del mundo sin tener a la fecha un conseso sobre qué método es mejor y bajo qué criterios o características de los conjuntos de datos.

Dentro de los métodos más robustos implementados a la fecha para evaluar la RE están los modelos LUR que consideran las concentraciones promedio anuales y las características de uso de suelo; que lo hace adecuado cuando se cuenta con este tipo de información (15). Otros procedimientos que evalúan la exposición tanto a corto como largo plazo utilizan información adicional de monitoreo móvil (además del estacionario)(16). Enfoques como el Concentration Similarity Frequency (CSF) han sido probados tanto para el monitoreo de áreas rurales como de áreas industriales con alta variabilidad en concentración y tipos de contaminantes (17). Lo que es una conclusión común entre los diferentes estudios es el papel que juega la correcta evaluación de la RE y su incorporación en los análisis de exposición en el corto y largo plazo. Es decir, los estudios de exposición acumulada pueden estar sesgados si los modelos y cálculos dependen de datos de monitoreo, y las mediciones de dicho monitoreo no son representativas en espacio y tiempo para cada contaminante (18).

Debido a que las estimaciones de concentraciones utilizando datos de las redes de monitoreo de contaminantes están siendo utilizadas en investigaciones en el área de salud (i.e. se asignan a individuos en conjunto con otras variables), es prioritario determinar cuales son los mejores métodos para evaluar la representatividad de dichas redes y/o estaciones, así como para generar modelos de exposición.

1.4. Políticas Públicas en materia de salud

En las grandes urbes a nivel mundial se ha observado una considerable disminución de las tasas de morbilidad y mortalidad asociadas a la exposición a contaminantes después de la implementación de programas promovidos por los gobiernos e instituciones locales. La Evaluación de Impacto a la Salud (EIS) es un enfoque utilizado a nivel mundial por los tomadores de decisiones para evaluar el impacto económico y de salud de una intervención. En la Ciudad de México y Área Metropolitana, los programas en pro de la reducción de emisiones de contaminantes han sido administrados por el

Programa para mejorar la calidad del aire de la Zona Metropolitana del Valle de México (PROAIRE). El plan de manejo de calidad del aire se actualiza cada 10 años y provee los lineamientos y políticas públicas enfocadas a reducir la exposición de la población a elevadas concentraciones de contaminantes (19).

1.4.1. Evaluación de Impacto a la Salud en la ZMVM

En la Evaluación de Impacto a la Salud realizado por Riojas-Rodríguez, et al. para el periodo 2011-2020 se analizaron diferentes escenarios de reducción de PM_{10} y O_3 y se cuantificaron los posibles beneficios. Es importante mencionar que ésta evaluación siguió los lineamientos de la Organización Mundial de la Salud (OMS) y la Agencia de Protección Ambiental de Estados Unidos (US-EPA). Se determinó que una reducción de los niveles de concentración de PM_{10} a $20\mu g/m^3$ podría prevenir 2300 muertes por año aproximadamente. O bien 1041 muertes prevenibles para una reducción a $40\mu g/m^3$. En el caso de O_3, una reducción a 0.05 y 0.06 *ppm* de promedio anual podría prevenir aproximadamente 400 y 110 muertes por aõ respectivamente. Lo anterior beneficiaría principalmente a la población de adultos mayores (mayores de 65 aõs) reduciendo la tasa de mortalidad causadas por problemas cardio-pulmonares (19).

Las EIS son sólo un ejemplo de estudios que son relevantes para la elaboración de políticas preventivas en materia de salud pública. En la última década, los programas vigentes para control de la contaminación atmosférica (principalmente por O_3) producida por vehículos automotores, incluye: i) los Programas de Verificación Vehicular Obligatoria, ii) Hoy No Circula y iii) Contingencias Ambientales Atmosféricas. Además de las NORMAS Oficiales Mexicanas NOM-EM-167-SEMARNAT-2016 y NOM-020-SSA1-2014. Sin embargo, contingencias ambientales como las ocurridas en 2016, 2017 y 2019 ponen a discusión la eficacia y pertinencia de dichos programas. Por esto, es necesario que la comunidad científica en conjunto con las instituciones reguladoras, promuevan la generación y uso apropiado de datos que sean la base para la adaptación de programas existentes o bien, la implementación de nuevos planes de control de emisión de contaminantes.

1.4.2. Atlas ambiental y de salud

La creciente cantidad de datos que se generan hoy en día hace imperativo la necesidad de contar con herramientas que integren y resuman la información de distintas fuentes para un análisis más rápido y efectivo. Especialmente, cuando se trata de analizar conjuntos masivos de datos y múltiples variables. En particular, cuando se trata de variables cuyo valor cambia respecto a áreas geográficas, los mapas son imprescindibles para la visualización de la información en un estudio. En el área de Salud Pública, no es nuevo contar con mapas para monitorear y evaluar problemas de salud y los resultados de políticas públicas. Lo ha sido desde 1954 para describir el brote de cólera. Ramos Herrera (2016) define al *Atlas de Salud* como Úna colección de mapas relacionados a condiciones de salud, con un método único para analizar datos y describir la magnitud

de problemas de salud, identificar su relación con situaciones sociales, determinar factores condicionantes, y apoyar la toma de decisiones por parte de las autoridades de salud, gobierno, organizaciones no gubernamentales y la comunidad"(20).

Contar con un atlas de salud, permite a investigadores y tomadores de decisiones, explorar, confirmar, sintetizar y presentar información de manera más efectiva. Además, éstos atlas describen las variaciones geográficas de los problemas de salud y los resultados de las prácticas médicas en el área. Esto, con propósitos de seguridad y calidad de los servicios y prácticas de salud. Se espera que cada individuo tenga acceso a hospitales, profesionales de la salud a quienes consultar, condiciones ambientales adecuadas y entornos seguros independiente de su ubicación geográfica. Además, contar con información geográfica comparativa incluyendo variables ambientales, ayuda a identificar potenciales correlaciones entre factores sociales, ambientales y culturales. Al integrar toda esta información, un atlas puede ayudar a dilucidar si las variaciones en la magnitud de enfermedades, riesgo o falta de acceso a servicios sanitarios resulta únicamente del área geográfica o de otros factores como las preferencias de cada individuo o factores no controlables (20).

Actualmente, países como los Estados Unidos, Canadá, Inglaterra, entre otros; continúan desarrollando y actualizando plataformas disponibles públicamente en las que concentran información poblacional, ambiental y de salud con la finalidad de extraer información relevante para la toma de decisiones. Dentro de los temas de interés para la elaboración de estos mapas se encuentran principalmente la planeación urbana y la prevención de riesgos. La figura 1.5 muestra el portal web del Atlas Ambiental y de Salud para Inglaterra. Consiste en un conjunto de mapas interactivos (disponibles públicamente) para 14 *condiciones de salud* como diferentes tipos de cáncer (pulmón, mama, próstata, piel, etc), enfermedades cardiacas, etc. En ellos se muestra el riesgo relativo calculado para el periodo 1985-2009 y se muestra para hombres y mujeres de forma independiente. Así también el sitio web cuenta con mapas para *agentes ambientales* como la concentración de contaminantes (PM_{10}, NO_2, herbicidas, fungicidas)y algunas variables meteorológicas (mapas independientes) (21). Si bien éstos son mapas que se pueden explorar a diferente resolución espacial y brindan información relevante para el análisis de las variables, tienen la limitación de que cada variable se representa de forma independiente y cada mapa cuenta con diferente escala o nivel de resolución espacial y temporal. El problema de la elaboración de mapas dinámicos más completos, i.e. que presentan variables con las mismas escalas espaciales y temporales e incluso múltiples capas de información simultáneamente, no es propiamente técnico, sino que depende de la existencia y disponibilidad de los datos.

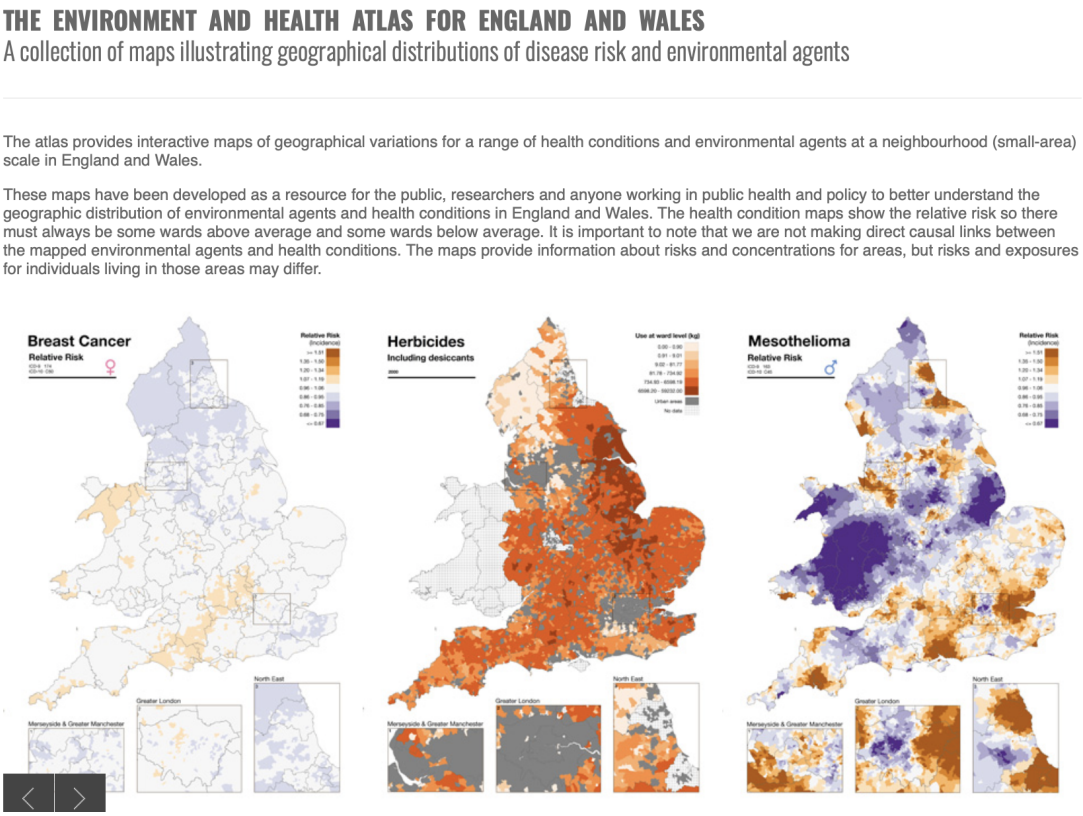


Figura 1.5: Atlas ambiental y de salud de Inglaterra y Wales. Tomado del portal web (21). Muestra la distribución espacial de la incidencia de diferentes condiciones de salud, así como el índice de contaminación por diferentes agentes.

Incluir atlas de mortalidad - USA

Capítulo 2

Planteamiento del problema

La ZMVM como otras metrópolis en el mundo, padece de problemáticas ocasionadas por la alta densidad poblacional, falta de planeación urbana y creciente polución en agua, suelo y aire a consecuencia de la industrialización. Esto ha ocasionado, en gran medida, el incremento en el número de hospitalizaciones y muertes. En la última década, en la ZMVM se han implementado programas para controlar y reducir las emisiones de contaminantes derivados del tráfico vehicular. A la fecha, programas como el Hoy no Circula y la Verificación Obligatoria no han sido completamente efectivos o bien, suficientes, para combatir el problema de los elevados índices de contaminación y la mala calidad del aire. La población está expuesta de manera constante a agentes contaminantes que amenazan su la salud, principalmente la de los grupos vulnerables como los infantes, mujeres embarazadas y los adultos mayores. Los efectos nocivos de la contaminación atmosférica van desde el dolor de cabeza, ojos irritados, problemas respiratorios leves, hasta cáncer y muerte por complicaciones cardiovasculares.

El Instituto Nacional de Salud Pública reportó en 2020 que a nivel nacional, cerca de 48 mil mexicanos mueren de forma prematura por contaminación del aire cada año. Lo que coloca a la polución del aire como el noveno factor de riesgo de muerte. Tan solo por $PM_{2.5}$ se estimaron 14 mil muertes prematuras en 15 ciudades y en el primer semestre de 2020, la Ciudad de México registró cerca de 11,000 muertes vinculadas con la contaminación del aire. Si bien es poco probable que la ZMVM disminuya su crecimiento y actividad industrial y los problemas que conlleva, es necesario contar datos, herramientas y eventualmente programas de acción más sólidos y basados en evidencia que permitan prevenir contingencias ambientales, controlar la exposición de los individuos y atender de manera pertinente las afecciones y enfermedades para reducir las tasas de mortalidad (al menos relacionada a esta problemática). Es importante proveer a los tomadores de decisiones e instituciones con herramientas eficientes que les permitan entender las potenciales causas de enfermedades y muertes en la región. Analizar la relación entre las variables sociales, económicas y ambientales. Así como planificar, desarrollar y monitorear programas orientados a mejorar la calidad de vida de la población.

Capítulo 3

Justificación

La Zona Metropolitana del Valle de México (ZMVM) es una metrópoli que, como muchas otras en el mundo, padece de muchos problemas ocasionados por la alta densidad poblacional y la falta de planeación urbana. Uno de estos problemas es el incremento de muertes causadas directa o indirectamente por la constante exposición a elevadas concentraciones de contaminantes en el aire. Así como del incremento en la incidencia de enfermedades crónicas o agravamiento de éstas. A saber, la Secretaría de Salud no reporta las muertes causadas específicamente por los efectos nocivos de contaminantes en el aire. Sin embargo, es posible asociar la pobre calidad del aire con el incremento de enfermedades crónicas como las cardiovasculares, respiratorias y cáncer, con los altos índices de mortalidad. A la fecha, es posible acceder a fuentes de datos públicas con información ambiental y de salud. Las tasas de mortalidad se encuentran disponibles por grupo etario por delegación de la Ciudad de México. Mientras que las concentraciones promedio diarias están disponibles para todos los contaminantes considerados *criterio*. El análisis de éstos conjuntos de datos es el punto de partida para identificar la disponibilidad y calidad de la información que se pretende utilizar para establecer el riesgo al que está expuesta la población.

Para que el gobierno, a través de las instituciones responsables de atender los temas ambientales y de Salud Pública, puedan coordinar esfuerzos y desarrollar programas basados en evidencia (datos), es indispensable contar con herramientas que faciliten la integración y el análisis de información de diferentes fuentes. Sin embargo, sólo la ZMVM mantiene una red de monitoreo de calidad del aire sofisticada y completa que ha estado activa por más de una década y por tanto, se cuenta con una base de datos histórica y datos al día. Asimismo, datos históricos sobre tasas de mortalidad, han sido recolectados a mayor detalle solo para las 16 delegaciones en la Ciudad de México. Por tanto, este primer esfuerzo por analizar y conjuntar información ambiental y tasas de mortalidad es posible realizarse únicamente para esta región.

Capítulo 4

Hipótesis

El desarrollo de una herramienta para visualizar información espacial y temporal de variables relacionadas con contaminación ambiental y mortalidad, puede servir como guía para la implementación de programas en materia de Salud Pública en la Ciudad de México.

Capítulo 5

Objetivos: General y específicos

Objetivo General

Contribuir a la toma de decisiones en materia de Salud Pública a través del modelado y visualización espacio-temporal de zonas de riesgo ambiental e información de salud.

Objetivos Específicos

1. Generar una base de datos espacio-temporal de mortalidad y contaminantes de la Ciudad de México.
2. Modelar la relación espacio-temporal de la base de datos, a los efectos de buscar patrones relevantes para la toma de decisiones.
3. Generar un Atlas de mortalidad edad-específica para la Ciudad de México.

Capítulo 6

Diseño metodológico

El presente trabajo es un estudio de tipo observacional descriptivo el cuál se desarrolló en tres etapas principales: creación de la base de datos, modelado espacio-temporal y elaboración del Atlas de mortalidad.

6.1. Diseño del estudio

Estudio ecológico (Observacional descriptivo).

- Variables independientes: latitud, longitud, tiempo y valores de concentraciones de contaminantes CO, NO_2, O_3, SO_2, $PM_{2.5}$, PM_{10}, NO, NO_X y PM_{CO}.
- Variables dependientes: Tasa de mortalidad edad-específica anual por delegación de la Ciudad de México.

6.2. Universo y muestra

Concentraciones de contaminantes criterio

Universo o muestra Base de datos pública del Sistema de Monitoreo de Calidad del Aire de la Ciudad de México.

Tamaño de la muestra La muestra de datos comprende los valores de concentración por hora, por contaminante, por estación, para el periodo 2009-2018.

Tasa de mortalidad edad-específica por delegación

Universo o muestra Base de datos de tasas de mortalidad edad específica de la Secretaría de Salud de la Ciudad de México.

Tamaño de la muestra Todos los registros para cada una de las 16 delegaciones en el periodo 2000-2016.

6.3. Criterios de inclusión y exclusión

Concentraciones de contaminantes criterio

Criterios de inclusión Estaciones de monitoreo que miden alguno de los contaminantes criterio (CO, NO_2, O_3, SO_2, $PM_{2.5}$, PM_{10}, NO, NO_X y PM_{CO}) y que haya estado activa durante el periodo 2009-2018.

Criterios de exclusión Estación-contaminante con menos del 70 % de los datos diarios y menos del 70 % de los datos por semana.

Criterios de eliminación No aplica.

Tasa de mortalidad edad-específica por delegación

Criterios de inclusión No aplica.

Criterios de exclusión No aplica.

Criterios de eliminación No aplica.

6.4. Procedimiento

Como se ve en la figura 6.1 el proyecto se dividió en 3 etapas para cubrir cada uno de los objetivos de esta investigación. La primera etapa consistió en construir la base de datos que incluye datos de concentraciones de los 8 contaminantes criterio (CO, NO_2, O_3, SO_2, $PM_{2.5}$, PM_{10}, NO, NO_X y PM_{CO}) obtenidos de la Red de Monitoreo Automática (REMA) para el periodo 2009-2018. Además de los datos de las tasas de mortalidad edad-específica (todas las causas) para las 16 delegaciones de la Ciudad de México obtenidas de la Secretaría de Salud para el periodo 2000-2016.

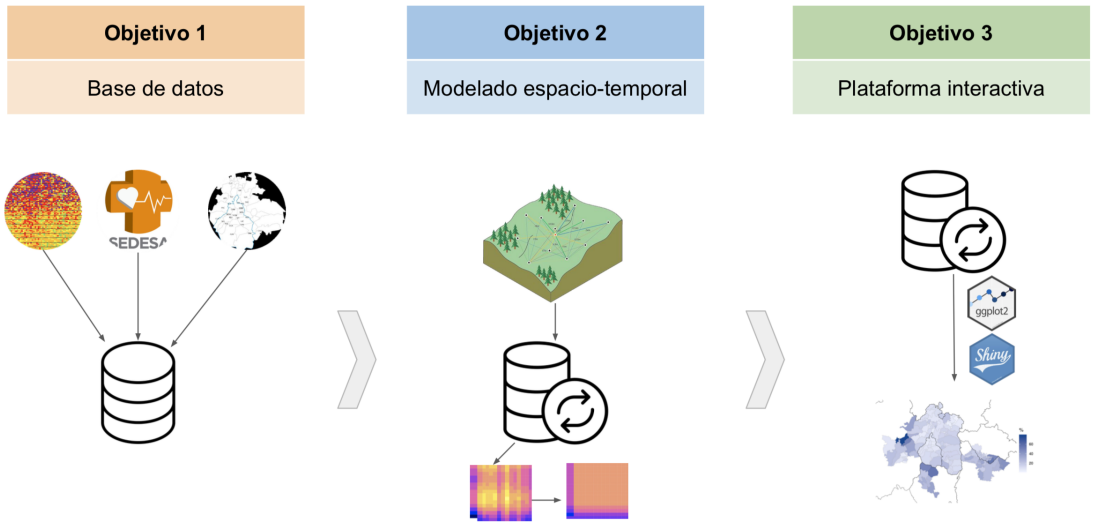


Figura 6.1: Esquema del flujo de trabajo que incluye 3 principales etapas. i) la construcción de la base de datos, ii) el modelado espacio-temporal y iii) desarrollo del prototipo de una plataforma interactiva.

En la segunda etapa se desarrolló e implementó una metodología para estudiar la distribución espacial y temporal de los diferentes contaminantes y para determinar la cobertura de la red de monitoreo en la zona metropolitana. Además, se llevó a cabo la interpolación espacial y temporal de las tasas de mortalidad edad-específica (todas las causas) para inferir valores con una granularidad más fina, de año a trimestre y de delegación a colonia.

Finalmente, en la tercera etapa se trabajó en la elaboración de un conjunto de mapas que describen espacial y temporalmente los patrones de mortalidad edad-específica en la Ciudad de México.

Objetivo 1: Construcción de la base de datos

Los datos de los valores de concentración (por hora) para cada uno de los contaminantes criterio se obtuvieron de la base datos pública del Sistema de Monitoreo de Calidad del Aire disponible en el portal Aire CDMX (40). Para descargar los datos se utilizó el paquete aire.zmvm en [R] (41). Dicha base de datos incluye la información de 70 estaciones de monitoreo que constituyen la Red Automática de monitoreo. De éstas, sólo 49 estaciones miden concentraciones de los contaminantes de interés. Es importante mencionar que no todos los contaminantes se miden en todas las estaciones de monitoreo y no todas las estaciones tienen registros completos de dichos datos. Lo anterior puede ser debido principalmente a periodos de mantenimiento en las estaciones.

Por otra parte, los datos de las tasas de mortalidad edad-específicas (todas las causas) a nivel delegación se obtuvieron de la base de datos de la Secretaría de Salud de la Ciudad de México (SEDESA) (95). Los grupos de edad que se consideraron son:

1. Infantil ($x < 1$ años de edad)
2. Pre-escolar ($1 \leq x < 4$ años de edad)
3. Escolar ($4 \leq x < 14$ años de edad)
4. Productiva ($14 \leq x < 64$ años de edad)
5. Post-productiva ($x \geq 64$ años de edad)

Para poder generar mapas del área de estudio con diferente granularidad, se utilizaron shapefiles de polígonos a nivel colonia que al ser combinados por el identificador de la delegación permite reconstruir el mapa de la Ciudad de México con su respectiva división por delegación y colonia. Los datos de polígonos se obtuvieron del Instituto Nacional de Estadística y Geografía (INEGI) y comprenden 16 delegaciones y 2,097 colonias. Para poder asignar el valor de concentración de un contaminante, así como la tasa de mortalidad a un punto en el espacio (dentro de la colonia o dentro de la delegación), se calcularon los centroides usando los paquetes rgeos, gCentroid o gPointOnSurface dependiendo de la forma del polígono (97).

Objetivo 2: Modelado espacio temporal
Concentraciones de contaminantes criterio
Dado que no todos los datos para el periodo de interés están disponibles, se consideró los siguientes criterios para incluir los datos de una estación de monitoreo. Se incluyeron los valores de una estación si existían al menos 17 hrs por día registradas y a su vez, 5 días en una semana. Los datos que cumplieron con este criterio se sumarizan en un valor promedio por semana, por estación y por contaminante.

Para determinar el grado de dependencia entre mediciones espaciales y temporales para cada contaminante, se utilizó el semivariograma. Esta función indica cómo y qué tanto varían las mediciones tomadas entre un punto y otro (en espacio y tiempo). Para el caso unidimensional, el semivariograma que depende de la distancia h tiene la forma:

$$\hat{\gamma}(h) = \frac{1}{2N(h)} \sum_{i=1}^{N(h)} [Z(x_i + h) - Z(x_i)]^2 \tag{6.1}$$

donde $Z(x_i)$ es el valor observado en la coordenada x_i en la posición i-ésima, $Z(x_i + h)$ es el valor observado en la posición $x_i + h$ y $N(h)$ el número de puntos muestrales dentro de una distancia h. Asimismo, la descripción temporal tiene una versión equivalente. El objetivo entonces es, obtener el modelo de variograma que mejor se ajuste a los valores observados (semivariograma experimental). De entre los modelos de variograma más utilizados están el exponencial, esférico y gaussiano.

$$\hat{\gamma}_{Exp}(h) = (s-n)\left(1 - e^{-\frac{h}{ra}}\right) + nH_{(0,\infty)}(h)$$
$$\hat{\gamma}_{Sph}(h) = (s-n)\left(\left(\frac{3h}{2r} - \frac{h^3}{2r^3}\right)H_{(0,r)}(h) + H_{[0,\infty)}(h)\right) + nH_{(0,\infty)}(h)$$
$$\hat{\gamma}_{Gau}(h) = (s-n)\left(1 - e^{-\frac{h^2}{r^2a}}\right) + nH_{(0,\infty)}(h) \tag{6.2}$$

donde n es el *nugget* que puede interpretarse como el residuo de la dependencia espacial o bien, $lim_{h\to 0}\hat{\gamma}(h)$. En la figura 6.2 se puede observar que el valor del nugget es el valor del intercepto con el eje y y refiere a que no existe correlación con otras mediciones excepto con el valor en ese mismo punto. Por otro lado, el *sill* $s = lim_{h\to\infty}\hat{\gamma}(h)$ que en la figura 6.2 se observa como el valor máximo del variograma cuando la distancia entre las mediciones es muy grande. El rango r es el valor de la distancia a la cual se pierde la correlación espacial entre puntos muestrales. Finalmente, $a = 1/3$ y $H_A(h)$ es la función de Heaviside la cual vale 1 si $h \in A$ y 0 de otra forma.

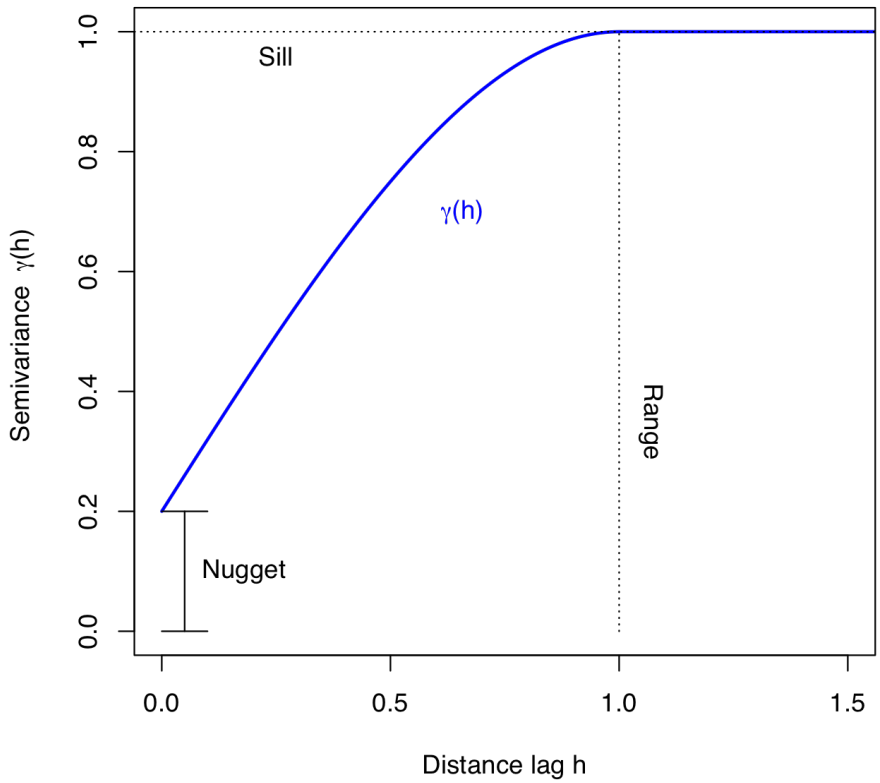


Figura 6.2: Se muestran los 3 parámetros que definen el semivariograma, también conocido como semivarianza. El *nugget* n corresponde al residuo de la dependencia espacial, el *sill* s es el valor máximo del semivariograma y el *rango* r es la distancia a la cual se pierde la correlación espacial.

El semivariograma experimental, así como los distintos modelos descritos hasta ahora en función de la distancia en el espacio (h) en la forma unidimensional, aplican de la misma manera para el análisis temporal (u). Pero, para una descripción de la correlación espacio-temporal, es necesario considerar la estructura de covarianza y el semivariograma $\gamma(h, u)$. Cada uno de los modelos que se pusieron a prueba están incluidos en el paquete gstat de [R] (**?**). A saber, los modelos: *separable*, *product sum*, *st*, *metric*, *sum metric* y *simple sum metric*.

Para el modelado, los valores iniciales de los parámetros nugget, sill y rango a partir de los datos experimentales y con los siguientes criterios:

- Para el nugget se asignó el valor de la mediana de los 3 primeros valores empíricos.
- El sill, como el valor de la mediana de los últimos 5 valores del variograma empírico.

- El rango, como un tercio del valor máximo de la distancia y para el rango temporal, el valor máximo.

Además del modelado del semivariograma y la covarianza, la anisotropía espacial y temporal se implementó de acuerdo a la metodología descrita en el paquete *gstat*. Así, para cada contaminante, se probaron los modelos de variograma exponencial, gaussiano y esférico con los respectivos modelos de covarianza: *separable*, *product sum*, *st*, *metric*, *sum metric* y *simple sum metric*. El mejor modelo se eligió en base al valor del error cuadrático medio ponderado. A partir de dicho modelo ganador, se obtuvieron los valores finales de los parámetros.

En particular, con el valor del rango espacial para cada año, fue posible asignar las áreas de representatividad de las estaciones de monitoreo. Así, se obtuvo una descripción de cómo ha evolucionado la representatividad de la red de monitoreo en el periodo de estudio. En breve, el parámetro de rango espacial se utilizó como el radio para trazar una circunferencia partiendo de la ubicación de la estación de monitoreo y definiendo así su área de cobertura. Aplicando este procedimiento para cada estación, se obtiene el área total de cobertura o representatividad espacial de la red de monitoreo. Esto se realizó para cada uno de los contaminantes. Además, respecto de la componente temporal del rango, el resultado (para cada contaminante) nos indica que tanto afectan los valores de concentración en un momento dado a valores en un tiempo posterior.

Tasas de mortalidad edad-específica por delegación

En cuanto al análisis de las tendencias de las tasas de mortalidad, se implementó un análisis de interpolación tanto espacial como temporal. Contar con una descripción más fina a nivel espacial cuando se trata de estudios de mortalidad, es relevante en una urbe como la Ciudad de México la cual presenta una gran heterogeneidad de condicionantes de salud como la exposición a fuentes contaminantes, crimen y disponibilidad de servicios sanitarios. Estos factores pueden variar de forma drástica entre puntos tan cercanos como de colonia a colonia. Análogamente, pueden haber cambios muy significativos en condicionantes de la salud cuya descripción sea más apropiada a escalas más finas. Por ejemplo, para patógenos estacionales, y condiciones climatológicas que se relacionan con los efectos de contaminantes y que tienden a ser más nocivos en determinados periodo (bimestre, trimestre, etc) y que no es posible capturar cuando se analizan únicamente las tasas anuales.

Se utilizó el método de interpolación Area-to-point Kriging, para encontrar la estructura espacio-temporal de variance-covariaza de mortalidad. Adicionalmente, el método Global Ordinary Kriging fue implementado (de acuerdo la librería gstat) para obtener las tasas de mortalidad edad específicas a nivel trimestre y nivel colonia. Los valores iniciales de los parámetros requeridos: nugget, sill y rango, se definen bajo el mismo criterio mencionado anteriormente para el análisis de contaminantes.

Para evaluar la contribución de la tasa de mortalidad por grupo de edad a nivel colonia y delegación, se implementó un modelo lineal de efectos mixtos. El modelo se implementó en el software infostat (versión 2018) y su fundamento es el siguiente:

$$\begin{aligned} y_{ijk} = \mu + \alpha_i + \beta_j + \gamma \times t_k + \delta \times t_k^2 + \zeta \times t_k^3 + \alpha_i \times \beta_j + \alpha \times \gamma \times t_k + \\ \alpha_i \times \delta \times t_k^2 + \alpha_i \times \zeta \times t_k^3 + \beta_j \times \gamma \times t_k + \beta_j \times \delta \times t_k^2 + \beta_j \times \zeta \times t_k^3 + \\ \alpha_i \times \beta_j \times \gamma \times t_k + \alpha_i \times \beta_j \times \delta \times t_k^2 + \\ \alpha_i \times \beta_j \times \zeta \times t_k^3 + \varepsilon_{ijk} \end{aligned} \tag{6.3}$$

$$\varepsilon_{ijk} = \lambda_{t_k} + \nu_{ijk} \tag{6.4}$$

$$\lambda_{t_k} = \phi_{\beta_j} \lambda_{t_{k-1}} + u_{t_k} \tag{6.5}$$

$$var(\epsilon_{ijk}) = \sigma^2 g^2(\alpha_i) \tag{6.6}$$

Donde y_{ijk} representa la tasa de mortalidad para el grupo de edad específico (α_i), a nivel delegación o colonia (β_j), para el tiempo (t_k); μ es el promedio global; γ, δ, ζ son los correspondientes coeficientes de 3er orden. ε_{ijk} es el término de error, donde λ_{t_k} denota los efectos de tiempo no observados y ν_{ijk} el término de error idiosincrático (124).

Objetivo 3: Desarrollo de mapas de mortalidad edad-específica

Para concentrar y facilitar el uso de la información de salud y ambiental, los mapas son de las herramientas más utilizadas. En este trabajo, además de llevar a cabo análisis sofisticados de una gran cantidad de información (la extensa base de datos de contaminantes) y estudios detallados de las tendencias espaciales y temporales de las tasas de mortalidad con valores interpolados, se elaboraron un conjunto de mapas geográficos detallando los niveles de mortalidad por colonia y delegación. Esto se realizó utilizando el software [R] y distintas paqueterías.

6.5. Variables de estudio

6.6. Aspectos éticos

El presente trabajo cumplió con lo establecido en: Declaración de Helsinki, Ley General de Salud. El protocolo de este trabajo fue aprobado por el Comité de Ética en Investigación del Centro de Investigación en Ciencias Médicas con el registro XXXXX del CEI del CICMED. tiene el título del proyecto anterior

Variables	**Definición operativa**	**Tipo de variable**	**Escala de medición**
Concentración de Monóxido de carbono (CO)	El valor promedio temporal detectado en el aire. La condición normal corresponde a presión de 1 atm y temperatura de 25°C	Independiente cuantitativa	ppm
Concentración de Dióxido de nitrógeno (NO_2), Dióxido de azufre (SO_2), Ozono (O_3), Monóxido de nitrógeno (NO) y otros Óxidos de nitrógeno (NO_X)	El valor promedio temporal detectado en el aire en partes por billón (ppb). La condición normal corresponde a presión de 1 atm y temperatura de 25°C.	Independiente cuantitativa	ppb
Concentración de Partículas menores a 2.5 micrómetro ($PM_{2.5}$)	El valor promedio temporal de material particulado detectado en el aire con un diámetro aerodinámico inferior a $2.5\mu m$	Independiente cuantitativa	$\mu g/m^3$
Concentración de Partículas menores a 10 micrómetros (PM_{10})	El valor promedio temporal de material particulado detectado en el aire con un diámetro aerodinámico inferior a $10\mu m$	Independiente cuantitativa	$\mu g/m^3$
Concentración de Partículas gruesas (PM_{CO})	El valor promedio temporal de material particulado detectado en el aire con un diámetro aerodinámico entre 2.5 y 10 μm	Independiente cuantitativa	$\mu g/m^3$
Tasa de mortalidad edad-específica	Es la tasa de mortalidad para un grupo de edad en particular. Se calcula como el número de muertes en dicho grupo de edad entre el número de personas de esa edad en la población	Dependiente cuantitativa	razón

Tabla 6.1: Descripción de las variables de estudio.

Capítulo 7

Resultados

7.1. Artículo publicado

7.1.1. Título del artículo publicado

Spatio-Temporal Representativeness of Air Quality Monitoring Stations in Mexico City: Implications for Public Health

7.1.2. Página frontal del manuscrito

frontiers in Public Health

ORIGINAL RESEARCH
published: 12 January 2021
doi: 10.3389/fpubh.2020.536174

Spatio-Temporal Representativeness of Air Quality Monitoring Stations in Mexico City: Implications for Public Health

Karol Baca-López[1,2†], Cristóbal Fresno[3†], Jesús Espinal-Enríquez[2,4], Mireya Martínez-García[5], Miguel Angel Camacho-López[1], Miriam V. Flores-Merino[6] and Enrique Hernández-Lemus[2,4]*

[1] School of Medicine, Autonomous University of the State of Mexico, Toluca de Lerdo, Mexico, [2] Computational Genomics Division, National Institute of Genomic Medicine, Mexico City, Mexico, [3] Technological Development Office, National Institute of Genomic Medicine, Mexico City, Mexico, [4] Centro de Ciencias de la Complejidad, Universidad Nacional Autónoma de México, Mexico City, Mexico, [5] Sociomedical Research Unit, National Institute of Cardiology 'Ignacio Chávez', Mexico City, Mexico, [6] School of Chemistry, Autonomous University of The State of Mexico, Toluca de Lerdo, Mexico

Assessment of the air quality in metropolitan areas is a major challenge in environmental sciences. Issues related include the distribution of monitoring stations, their spatial range, or missing information. In Mexico City, stations have been located spanning the entire Metropolitan zone for pollutants, such as CO, NO_2, O_3, SO_2, $PM_{2.5}$, PM_{10}, NO, NO_x, and PM_{CO}. A fundamental question is whether the number and location of such stations are adequate to optimally cover the city. By analyzing spatio-temporal correlations for pollutant measurements, we evaluated the distribution and performance of monitoring stations in Mexico City from 2009 to 2018. Based on our analysis, air quality evaluation of those contaminants is adequate to cover the 16 boroughs of Mexico City, with the exception of SO_2, since its spatial range is shorter than the one needed to cover the whole surface of the city. We observed that NO and NO_x concentrations must be taken into account since their long-range dispersion may have relevant consequences for public health. With this approach, we may be able to propose policy based on systematic criteria to locate new monitoring stations.

Keywords: public policy, air pollution, missing data, geo-temporal analysis, semivariogram

OPEN ACCESS

Edited by:
Yang Liu,
Emory University, United States

Reviewed by:
Marie-Cecile Genevieve Chalbot,
New York City College of Technology, United States
Hiep Nguyen Duc,
New South Wales Department of Planning, Industry and Environment (DPIE), Australia
Edmilson Dias De Freitas,
University of São Paulo, Brazil

***Correspondence:**
Enrique Hernández-Lemus
ehernandez@inmegen.gob.mx

†These authors have contributed equally to this work

Specialty section:
This article was submitted to Environmental Health, a section of the journal Frontiers in Public Health

Received: *18 February 2020*
Accepted: *09 November 2020*
Published: *12 January 2021*

Citation:
Baca-López K, Fresno C, Espinal-Enríquez J, Martínez-García M, Camacho-López MA, Flores-Merino MV and Hernández-Lemus E (2021) Spatio-Temporal Representativeness of Air Quality Monitoring Stations in Mexico City: Implications for Public Health. Front. Public Health 8:536174. doi: 10.3389/fpubh.2020.536174

1. INTRODUCTION

As population density, mobility, and industrial activity keep growing at an accelerated rate, air pollution has gained the attention of policy makers in urban and metropolitan areas. There is a common concern in highly polluted cities regarding the increasing mortality associated with chronic and acute diseases whose effects may be aggravated due to exposure to air contaminants (1–3).

It is well-known that different diseases or health-related effects depend on both the exposure time and concentration levels (1). As evidence suggests, not only long periods of exposure can be damaging, but exposure to high levels in short periods—even a few hours—may have an immediate negative impact (4, 5).

7.1.3. Carta de aceptación

Solo tengo email de confirmación-aceptación

Frontiers: Congratulations! Your manuscript is accepted - 536174

1 mensaje

Frontiers Public Health Editorial Office <publichealth.editorial.office@frontiersin.org> 9 de noviembre de 2020, 10:06
Responder a: Frontiers Public Health Editorial Office <publichealth.editorial.office@frontiersin.org>
Para: kbaca@inmegen.edu.mx

Dear Dr Baca-Lopez,

Frontiers Public Health Editorial Office has sent you a message. Please click 'Reply' to send a direct response

I am pleased to inform you that your manuscript Spatio-Temporal Representativeness of Air Quality Monitoring Stations in Mexico City: Implications for Public Health has been approved for production and accepted for publication in Frontiers in Public Health, section Environmental Health.
Your manuscript is currently being prepared for publication. The provisional version of the abstract or introductory section is currently available online. Please do not communicate any changes at this stage. You will be contacted as soon as the author proofs are ready for your revisions.

Manuscript title: Spatio-Temporal Representativeness of Air Quality Monitoring Stations in Mexico City: Implications for Public Health
Journal: Frontiers in Public Health, section Environmental Health
Article type: Original Research
Authors: Karol Baca-Lopez, Cristobal Fresno, Jesús Espinal-Enríquez, Mireya Martinez-Garcia, Miguel Angel Camacho-Lopez, Miriam V. Flores-Merino, Enrique Hernandez-Lemus
Manuscript ID: 536174
Edited by: Yang Liu

Due to lockdown orders in various countries you may experience a delay in the production and publication of your article but please be assured that we are working to provide them as soon as possible and ask for your patience.

You can click here to access the final review reports and manuscript: http://www.frontiersin.org/Review/EnterReviewForum.aspx?activationno=b0687c68-21de-4264-9430-ebb914b55ebb

As an author, it is important that you maintain your Frontiers research network (Loop) profile up to date, as your publication will be linked to your profile allowing you and your publications to be more discoverable. You can update profile pages (profile pictures, short bio, list of publications) using this link: http://loop.frontiersin.org/people/

Tell us what you think!

At Frontiers we are constantly trying to improve our Collaborative Review process and would like to get your feedback on how we did. Please complete our short 3-minute survey and we will donate $1 to Enfants du Monde, a Swiss non-profit organization:
https://frontiers.qualtrics.com/jfe/form/SV_8q8kYmXRvxBH5at?survey=author&aid=536174&uid=1075438

Thank you very much for taking the time to share your thoughts.

With best regards,

Your Frontiers in Public Health team

Frontiers | Editorial Office - Collaborative Peer Review Team
www.frontiersin.org
12 Moorgate,
EC2R 6DA, London, UK

7.1.4. Resumen

La evaluación de la calidad del aire en las áreas metropolitanas es un reto importante en las ciencias ambientales. Entre los problemas relacionados se encuentran la distribu-

ción de las estaciones de monitoreo, su alcance espacial o la falta de información. En la Ciudad de México se han ubicado estaciones que abarcan toda la zona metropolitana para contaminantes como CO, NO_2, O_3, SO_2, $PM_{2.5}$, PM_{10}, NO, NO_X y PM_{CO}. Una cuestión fundamental es si el número y la ubicación de dichas estaciones son adecuados para cubrir de forma óptima la ciudad. Mediante el análisis de las correlaciones espacio-temporales para las mediciones de contaminantes, evaluamos la distribución y el desempeño de las estaciones de monitoreo en la Ciudad de México de 2009 a 2018. Con base en nuestro análisis, la evaluación de la calidad del aire de esos contaminantes es adecuada para cubrir las 16 alcaldías de la Ciudad de México, con excepción del SO2, ya que su rango espacial es menor al necesario para cubrir toda la superficie de la ciudad. Observamos que las concentraciones de NO y NOx deben ser tomadas en cuenta ya que su dispersión de largo alcance puede tener consecuencias relevantes para la salud pública. Con este enfoque, podemos proponer una política basada en criterios sistemáticos para ubicar nuevas estaciones de control.

7.1.5. Apartados del artículo

Introduction

As population density, mobility and industrial activity keep growing at an accelerated rate, air pollution has gained the attention of policy makers in urban and metropolitan areas. There are a common concern in highly polluted cities regarding the increasing mortality associated to chronic and acute diseases whose effects may be aggravated due to exposure to air contaminants (3, 6, 22).

It is well-known that different diseases or health related effects depend on both, the exposure time and concentration levels (3). As evidence suggests, not only long periods of exposure can be damaging, but exposure to high levels in short periods -even few hours- hours may have immediate negative impact (5, 23).

Mexico City, as many other metropolis worldwide, has implemented strategies for urban planning, transportation and regulations of industrial activity to reduce contaminant emissions (24). As an example, the Metrobus transport system started operating in the year 2005 as an emission reduction strategy. By comparing CO, NO_X, PM_{10} and SO_2 measurements before and after the Metrobus operations, it was in fact observed a reduction ranging from 5-9 % for the different contaminants in city areas (25). Another example is driving restriction policy in Mexico City, which was originally set only for weekdays. In an attempt to improve results, the program extended this restriction to saturdays without meeting the expected results of reducing emmission by almost 15 % (26).

So far, these efforts have not been successful as planned. On the contrary, pollution levels have not decreased, which is noticeable from the continuous environmental alerts throughout the years. Some contaminants such as particulate matter PM_{10} vary seasonally and although some regulations may be effective for this type of pollutant, may not be useful for others (27).

In this regard, public policies will only be effective if they rely on the proper identification of pollution sources, the understanding of the dispersion dynamics and the adequate measurement of relevant variables.

Determining the number and distribution of air quality monitoring stations depends on the area to be covered, traffic, spatial variability due to land use, infuence of meteorological variables (temperature, wind speed and ultra violet radiation) and dispersion dynamics of each pollutant (15, 16, 28).

Environmental policy planning, needs reliable methods to assess the risk level associated with exposure to chemical and other noxous agents. This latter can be made by direct and indirect measurements of pollutants with epidemiological and toxicological dimensions.

Direct approaches require the estimation of incidence of undesired effects by considering individual exposures to contaminants. Environmental hazard of this kind often relies on the analysis of spatial data collected by environmental surveys (29).

Monitoring networks have two main purposes. First, by measuring spatial and temporal trends of pollutants concentrations, they provide air quality estimations to determine whether population is exposed to dangerous levels or not. In addition, with the use of socio-demographic, land use related variables and meteorological data, simulation models can guide to better decision making procedures.

Second, once implemented, the effectiveness of public policies and regulations can be evaluated by analysing changes in pollution levels that are caused merely by the imposed regulations. Thus, the development of a monitoring system, is a critical component for public health policy making, to decrease toxic emission and eventually prevent population from adverse contaminant effects.

A relevant emerging concept is environmental health surveillance. For this concept, the quality and completeness of information has been found variable, depending on individual hazards or exposures, even in well-developed public health surveillance systems, such as the one in Canada (29).

The Mexico City Ministry of Environment (Secretaría del Medio Ambiente, SEDEMA) is responsible for establishing measuring procedures, data gathering and reporting air quality levels (30). The estimations are based on the measurement of carbon monoxide (CO), nitrogen dioxide (NO2), ground level ozone (O_3), sulphur dioxide (SO_2), small fine particulate matter ($PM_{2.5}$ and PM_{10}), nitrogen oxide (NO), other nitrogen oxides (NO_X) and coarse particulate matter (PM_{CO}).

Although Mexico City's monitoring network meets international standards, it fails to have complete records for all the contaminants. In some cases, continuous monitoring stations stopped functioning due to technical reasons, maintenance, whereas others just stopped operating and in some cases, measurements were not registered while they were still active.

The appropriate functioning of such monitoring stations is extremely relevant for public health issues(31). It is known, for instance, that ozone and particulate matter ($PM_{2.5}$) levels have been closely associated with a number of adverse health effects, that may lead to premature mortality (32). Such effects are particularly relevant in the

context of urban environments (33).

To determine spatial representativeness of background monitoring stations from concentration measurements of air pollutants, has been a matter of intense research (13, 17, 18, 34, 35). It has been shown that size and shape of representative areas differ between pollutants and measured locations, and representative areas may range from 220 to 4500 km^2 (13).

To improve the assessment of coverage estimation in the case of a limited number of stations, detailed pollutant concentration maps at pedestrian level have been used (35). In this example, for Pamplona, Spain, authors found that approximately 18 % of total area is well represented, as most of the residential areas are included. This result states that it is possible to assess the covered area by air quality networks integrated by a limited number of stations for a small city (23 km^2).

The most complete study on the spatial representativeness of monitoring sites is the JCR Technical Report developed by the Forum for Air Quality Modelling in Europe (FAIRMODE) (14, 36). The aim was to perform an inter-comparison of 25 assessment methods from 14 different countries based on literature review of scientific journals and technical documentation.

The outcomes from this study were focused to better established a definition of spatial representativeness and to propose standard methodological procedures by country members in Europe. The different methodologies can be categorized according to their assessment criteria such as modelling, measurements, proxies, station classifications and annual concentrations. The outputs from these studies are presented as delimited areas or size parameters.

In order to have an adequate evaluation of monitoring networks effectiveness the city's complexity and spatial heterogeneities should be taken into account.

To estimate concentrations at unmeasured locations, interpolation methods such as Land-Use Regression (LUR), Inverse Distance Weighting (IDW), or Kriging, use historical data from monitoring stations and other monitoring procedures (9, 37). These estimations are mainly used for health risk assessment. Prediction of high values and trends helps to guide decisions both, locally and at city-wide levels.

Recently, Kriging geo-statistical approach has been used to analyse spatial representativeness from NO_2 preliminary concentrations in urban areas (34, 38). The Kriging methodology for spatio-temporal interpolation is based on the covariance data structure at spatial or spatio-temporal level. To achieve that task, the empirical semivariogram is modeled with a parsimonious covariance structure, through the use of different kernel functions, to determine the spatial and temporal correlation range.

In this work, we present a novel methodology based on the use of spatial and temporal variogram ranges modeling, to estimate monitoring stations representativeness. This methodology does not require estimation of pollutant concentration (full interpolation procedure).

This work aims to show the temporal evolution of spatial representativeness of monitoring stations for Mexico City, one of the most complex networks and metropolitan

areas worldwide. Additionally, temporal representativeness is shown, which is not the case for most of these type of studies. We explore these spatial coverage and temporal dependence on measurement for all pollutants currently reported in Mexico City.

In brief, two main questions are addressed here:

- What is the spatial and temporal representativeness of the air quality monitoring network in Mexico City?
- Which is the space/time range within which sample points measurements are correlated with measured values at monitoring stations?

We also discuss about the public health implications of these questions and how can we use this information to provide feedback to health and public policy makers.

It is worth mentioning that, to our knowledge, this is the first time that a spatio-temporal analysis of pollutants is made in a Mexican city. This is in part, due to the availability of public data from several pollutants, as well as an important number of monitoring stations.

Materials and Methods

Study area

Mexico City which belongs to the Valley of Mexico Metropolitan Area (VMMA) is located at $99^\circ\,21'53.64'' - 98^\circ\,56'25.08''$ West and $19^\circ\,2'53.52'' - 19^\circ\,35'34.08''$ North. By the year 2015, its total population was 8,985,339 people as reported by the National Institute of Statistics and Geography (Instituto Nacional de Estadística y Geografía, INEGI) (39). The polygon shape files of Mexico city were obtained from the National Institute of Statistic and Geography (Instituto Nacional de Estadística y Geografía, INEGI) (39). Hereafter, the geo-spatial data granularity was kept at the 16 available boroughs (municipalities).

Air pollution database

The Mexico City Air Quality Monitoring System public database was used in this study, according to the R package `aire.zmvm` implementation (40, 41). This database is part of the Automatic Network of Atmospheric Monitoring (RAMA, according to the Spanish acronym of Red Automática de Monitoreo Atmosférico).

The network was established by the Metropolitan Environmental Commission of Mexico City to monitor compliance with ambient air quality standards. The RAMA is part of the Atmospheric Monitoring System (SIMAT, Sistema de Monitoreo Atmosférico), a program responsible for ongoing measurements of the main air pollutants in Mexico City.

SIMAT comprehends 70 monitoring stations distributed along the VMMA, where only 49 of them register contaminant levels. At each monitoring station, hourly contaminant concentrations are available for: i) Particulate matter with an aerodynamic

diameter of less than 2.5 and 10 μm, $PM_{2.5}$ and PM_{10}, respectively, ii) Carbon monoxide (CO), iii) Ozone (O_3), iv) Sulphur dioxide (SO_2) and v) Nitrogen oxides, monoxide and dioxide, NO_X, NO and NO_2, respectively.

Naturally, all the monitoring stations do not collect all types of contaminants. Additionally, there exists missing records due to service maintenance or other incidents. The time period used in this work is from 2009 up to 2018, when possible. Location of monitor stations for each pollutant can be observed in Figure 7.1. Complete monitor stations data can be found in Supplementary Table S1.

Spatio-temporal statistics

Collected data was explored to get a clear picture of the pollutant monitor stations representativeness in Mexico City. Hourly contaminant data was summarized by their average into a week time-basis, if more than five days were available containing seventeen hour records or more. Data was plotted for each pollutant. Figure 7.2 shows data for nitrogen dioxide (NO_2). The rest of pollutant data can be found in Supplementary Figures S1 to S8.

Initial variogram parameter values were obtained from the empirical spatio-temporal pollutant measurements using *gstat* R package implementation as described in (42, 43, 44). Briefly, using the initial variogram parameters (see Supplementary Table S2), different spatial, temporal or joint covariance structures were tested to find the best parsimonious correlated data description, according to the available implementations in `gstat` (*metric*, *separable*, *productSum*, *sumMetric* and *simpleSumMetric*) (42, 43).

All possible single, double or triple variogram combinations (Exponential, Gaussian and Spherical) were tested according to the corresponding covariance structure, using `fit.StVariogram` function of `gstat` R library (42, 43), with the optimization method `L-BFGS-B` only with lower limits sets to 0.001 and upper limits left to its default value (infinity) . The covariance structure model selection criterion used was to minimize the weighted mean squared error (see, Supplementary Table S3). Finally, the winner spatio-temporal covariance structure was used to extract final variogram parameters (see, Supplementary Table S4). Particularly, the spatial `range` parameter was obtained out of the winner spatial variogram structure (Exponential, Gaussian or Spherical), under the winner model (*metric*, *separable*, *productSum*, *sumMetric* and *simpleSumMetric*), which can be found as the "Space" column in Supplementary Table S4. In addition, for a fixed year, the range parameter was re-calculated under the patrimonial global covariance structure constrained to the available data for the specific year.

Last but not least, the integration of both, contaminant's monitoring representativeness plots and final spatio-temporal variogram range parameters, were used to get a clear picture of Mexico City pollutant radius representativeness according to the time evolution monitor station activity.

Results

Spatio-temporal data exploration

Monitoring stations geo-localization are depicted in Fig. 7.1 along with the 16 boroughs of Mexico City. At first sight, the global picture makes it clear that not all the contaminants are acquired for each available station. Second, it seems that for all the contaminants the south of Mexico City (borough numbers 8, 11 and 12) are not as well represented as its northern counterpart.

This concern is related to rural and urban distribution, where most urban populated boroughs are located to the north of the city. Indeed, some contaminant stations are located outside Mexico City. Pollutants such as PM_{CO} and PM_{25} are almost exclusively collected inside Mexico City, whereas the rest have monitor stations outside the city.

To further explore the data completeness, the monitor station representativeness plots were generated. In Fig. 7.2 it is presented the case of NO_2, for the available stations from the beginning of 2009 until late 2018. Complete data acquisition summary for all the contaminant can be found in supplementary table S5.

It is clear that there exists block of missing data (in white), where some of them can be tracked down to the monitoring station inauguration (CUA, MON, CAM, NEZ and so on) in the mid 2011, or until 2015 for AJM and MGH stations.

As a matter of fact, there is no stations whatsoever that had not lost track of NO_2 at least for some hours (red to yellow cells) or even had stop working for a time gap of days, weeks or months. The last case, can be pictured for TLI, VIF, ATI and ACO to name some stations in the time window including the beginning of 2016.

The last time pattern can be considered as the complete station shutdown, as depicted by the block of LAG, CES, AZC, TAC and TAX, that stopped working at the end of 2010. Fortunately, these stations are not located neither in the same borough nor close each other to leave a non-measured extension (see Fig. 7.1). However, this data description level does not represent the extend covered by each monitoring station.

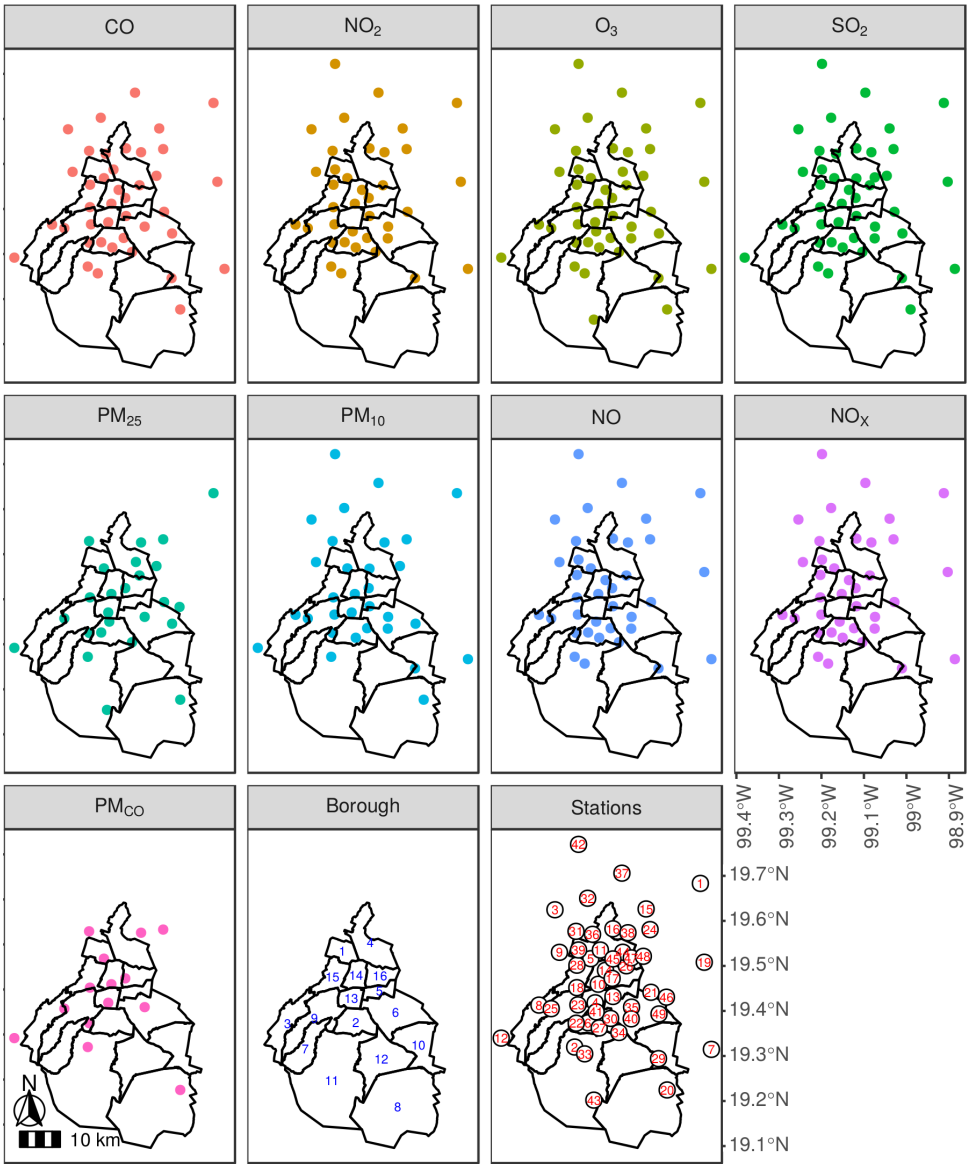


Figura 7.1: Monitor stations per contaminant in 2019. A borough level description for Mexico City is presented in the different contaminant panels, where the monitor stations are depicted in colour dots. The included contaminants are: CO, NO_2, O_3, SO_2, $PM_{2.5}$, PM_{10}, NO, NO_X and PM_{CO}. In addition, a panel shows borough names: 1. Azcapotzalco, 2. Coyoacán, 3. Cuajimalpa de Morelos, 4. Gustavo A. Madero, 5. Iztacalco, 6. Iztapalapa, 7. La Magdalena Contreras, 8. Milpa Alta, 9. Álvaro Obregón, 10. Tláhuac, 11. Tlalpan, 12. Xochimilco, 13. Benito Juárez, 14. Cuauhtémoc, 15. Miguel Hidalgo and 16. Venustiano Carranza. A similar idea is used to match the stations location to their corresponding three letter code: 1. ACO, 2. AJM, 3. ATI, 4. BJU, 5. CAM, 6. CCA, 7. CHO, 8. CUA, 9. FAC, 10. HGM, 11. IMP, 12. INN, 13. IZT, 14. LAG, 15. LLA, 16. LPR, 17. MER, 18. MGH, 19. MON, 20. MPA, 21. NEZ, 22. PED, 23. PLA, 24. SAG, 25. SFE, 26. SJA, 27. SUR, 28. TAC, 29. TAH, 30. TAX, 31. TLA, 32. TLI, 33. TPN, 34. UAX, 35. UIZ, 36. VAL, 37. VIF, 38. XAL, 39. AZC, 40. CES, 41. COY, 42. CUT, 43. AJU, 44. GAM, 45. LVI, 46. PER, 47. ARA, 48. FAR and 49. SAC. Notice that for contaminants such as PM_{CO} and

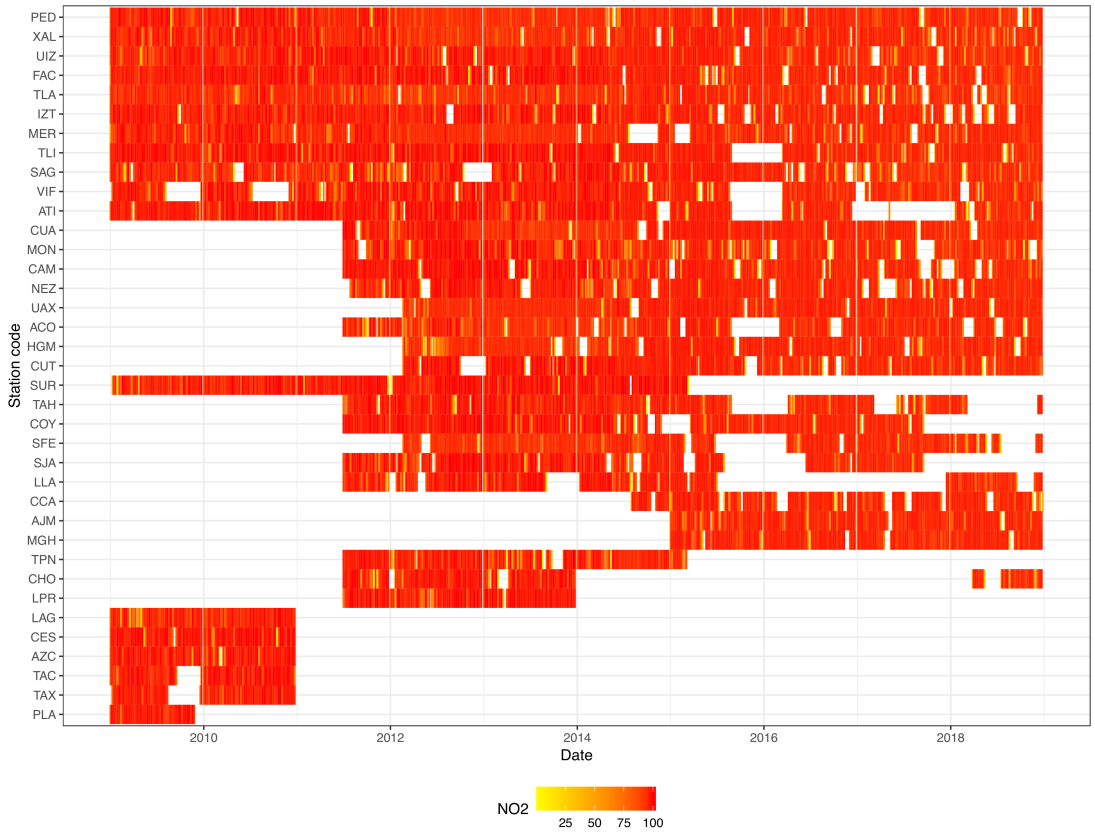


Figura 7.2: NO_2 monitor stations representativeness. The tile plot represents data completeness, on a weekly basis, for each monitor station (three letter code). The color bar indicates the percentage of complete data. Interestingly, there exists some time gaps whether the stations were not available yet, lost contaminant track or ended their monitoring life. However, in most of the recorded periods, there exists almost complete records (in red). See Supplementary Table S1 for monitor stations full description.

Spatio-temporal variogram estimation

Using the available data, sample spatio-temporal variograms were addressed. These variograms were used to generate the initial guesses values (Supplementary Table S2). Depending on the contaminant, the initial guesses are different for the nugget, range, sill and spatio-temporal anisotropy (stAni). In this context, the nugget is the model intercept attributable to measurement errors or spatial sources of variation at distances smaller than the sampling interval or both.

Interestingly, these sources of variations are negligible for CO, in contrast to the wide range of nugget values $(0.03-178.67)$. In addition, the correlation extends between

measurements, also known as range, in all cases is almost the same for all contaminants and last about 12 years for as far as 21.4 km. The variogram values obtained in the range, a.k.a the sill, are as close to the nugget only for CO and departs from it at most double its value.

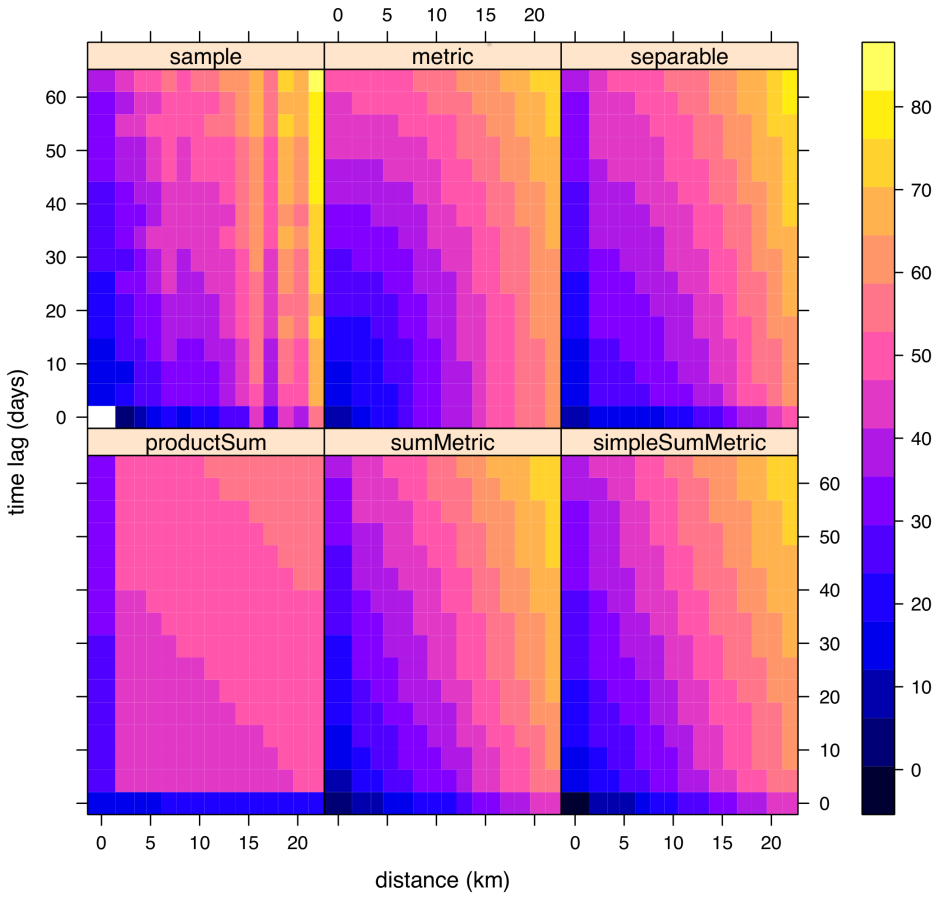


Figura 7.3: Tested spatio-temporal semivariograms for NO_2. The 2D-sample semivariogram was obtained from the collected data. In addition, the winner fitted covariance structure models (metric, separable, product sum, sum metric and simple sum metric) are also presented. In this case, the sum metric structure is the one that outperforms its competitors.

Final covariance model weighted mean square error for all the tested variogram permutations can be found in Supplementary Table S3. It is worth to mention that the lowest error for the different covariance structure methods was the one included: *sumMetric* for CO, NO_2, O_3, NO, NO_x, PM_{10} and PM_{CO}; *simpleSumMetric* SO_2 and

$PM_{2.5}$. Within these covariance models, there was no apparent pattern in the winner variogram model permutation (temporal + spatial + joint).

This is a data-driven approach that required to explore the complete permutation grid, in order to reach a parsimonious spatio-temporal correlation model. A visual comparison of each winner covariance model and sample variogram can be found in Fig. 7.3 for NO_2. The rest of variograms for the other pollutants can be found in Supplementary Figures 9-16.

Regarding ranges from winner models, the case of spatial correlation is presented in Figure 7.4A. The spatial range values measured in km are interpreted as the separation distance between two measured locations, i.e. monitoring stations, that from this value onwards, measurements are no longer correlated.

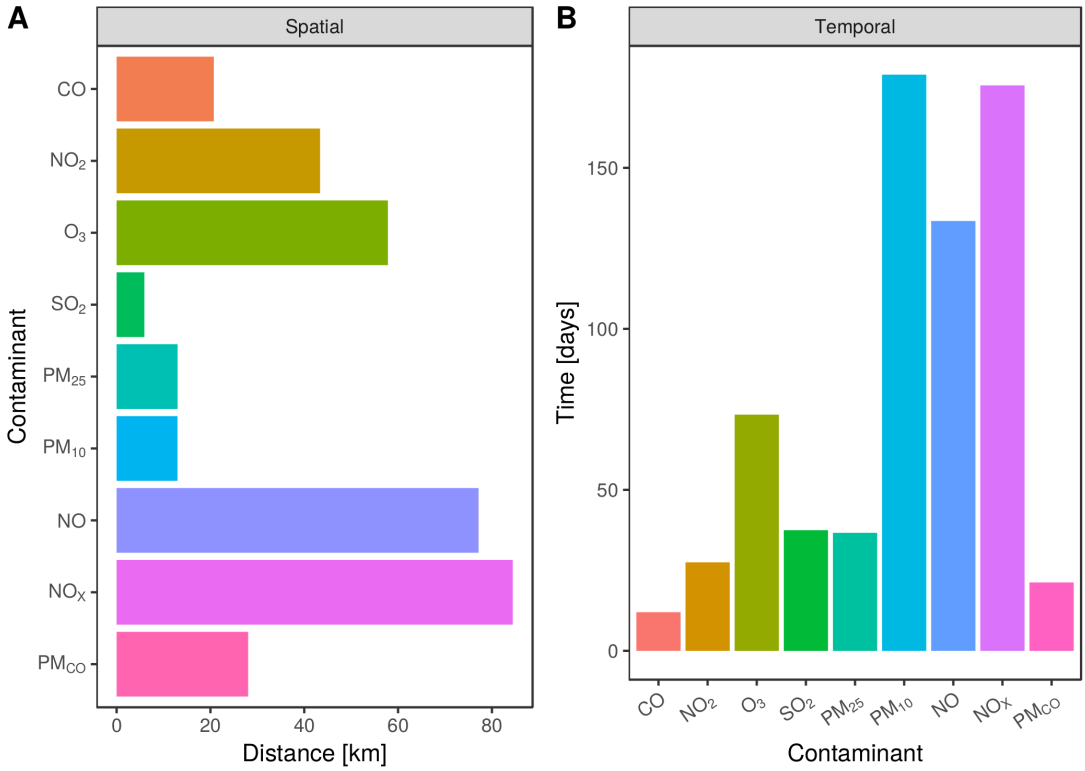


Figura 7.4: Spatio-temporal range panel. The horizontal or vertical bars stands for the winner covariance structure semivariogram range for each contaminant: CO_2, O_3 SO_2, $PM_{2.5}$, PM_{10}, NO, NO_X and PM_{CO}. **A**) Spatial range depicted by horizontal bars. The bigger the bar, the more distance it will be required to find uncorrelated data points. **B**) Temporal range in vertical bars. Notice that the two contaminants with the biggest spatial range (NO and NO_X), are almost the ones with the biggest time range but for PM_{10}.

This is, measured values in a certain monitoring station will be correlated to some other station while those that are located within this range. For instance, nitrogen oxides (NO_X) measurements between monitoring locations have the longest spatial correlation of 84.44 km, followed by nitrogen oxide (NO) with 77.15 km. On the contrary, particulate matter $PM_{2.5}$, PM_{10} and sulphur dioxide SO_2 have the shortest ranges, 13, 12.98 and 5.92 km respectively (see Supplementary Table 4).

For the case of temporal correlation, ranges are shown in Figure 7.4B. Particulate matter PM_{10} has the longest value of approximately six months (178.35 days) followed by nitrogen oxides (NO_X) and nitrogen oxide (NO) with ranges of 175.60 and 133.49 days respectively. These three contaminants differ in great extent in their correlation measurements in comparison to CO, PM_{CO}, NO_2, $PM_{2.5}$, $S0_2$ and O_3 with ranges between 12 to 73 days overall (see Supplementary Table 4).

To graphically show the spatial representativeness of each pollutant, we used spatial ranges reported in Table 7.4 that were obtained as final parameter values of the variogram models shown in Table S4.

For example, for the case of nitrogen dioxide (NO_2), a spatial range of 43.37 km was obtained from the empirical semivariogram and covariance structure modelling (see Figure 7.3). Thus, for each monitoring station that measured NO_2, the center of a circular area with radii 43.37 is matched to the station's location.

To generate a buffer area for NO_2 to represent the spatial influence for measuring this contaminant, circles traced at each location were joint to define a single border area. This processes is performed for NO_2 in the years 2009, 2012, 2015 and 2018.

As seen in Figure 7.2, for each of the selected years (for visualization purposes), there is a different number of active monitoring sites. Specifically, in 2009, only 18 sites were available and actively collecting hourly concentrations for NO_2. In 2012, 6 additional monitoring stations started collecting data for this contaminant. For the years, 2015 and 2018 the active sites were 26 and 25 respectively. In general, an increasing number of active sites can be seen for all pollutants starting with the year when measurements began (see Figures S1-S8).

In the same manner, using the calculated spatial ranges for CO, NO_2, O_3, SO_2, $PM_{2.5}$, PM_{10} NO, NO_X and PM_{CO}, temporal evolution of representativeness areas are shown in Figure 7.5. The different contaminants are color coded and displayed as columns, while rows are assigned to selected years. If we look at 2009 year panel, the most noticeable thing is that covered areas between contaminants differ widely. The biggest difference in range can be seen for SO_2 and NO_X with the smallest and largest ranges respectively (5.92 and 84.44 km).

The representativeness area for SO_2 is seen to mostly cover the north part of the city, while the south is not and for the years 2012, 2015 and 2018 similar patterns were obtained. As expected, although the number of monitored locations increased, there is not a significant increase in the covered area throughout the years, due the small range of measurement correlation. This small range depends on the intrinsic physicochemical properties of SO_2 and consequently, its diffusion in the atmosphere, as well as due to the complex traits of urban environments. However, regardless of land use,

traffic, population density and other variables involved for a selected year, it can be seen that for the same locations, the representativeness area for relatively different monitoring networks, this patterns are pollutant dependent.

Similar patterns of an increasing covered area that goes from north to south is observed for CO, $PM_{2.5}$, PM_{10} and PM_{CO}. It can be seen in the timeline, that for these pollutants in the year 2009 (except for PMCO, which was not registered at the time), the south area was not included in the network but, in the consecutive years this area was extended to almost cover the entire city. For NO_2, O_3, NO and NO_X, regardless of the amount of monitoring stations in 2009, because of the large correlation of measurements (ranges), the representativeness area of the networks accounts for th the whole city and a considerable percentage of the Valley of Mexico Metropolitan Area (VMMA). Thus, although monitoring stations were added to the network, no significant change is observed for the consecutive years.

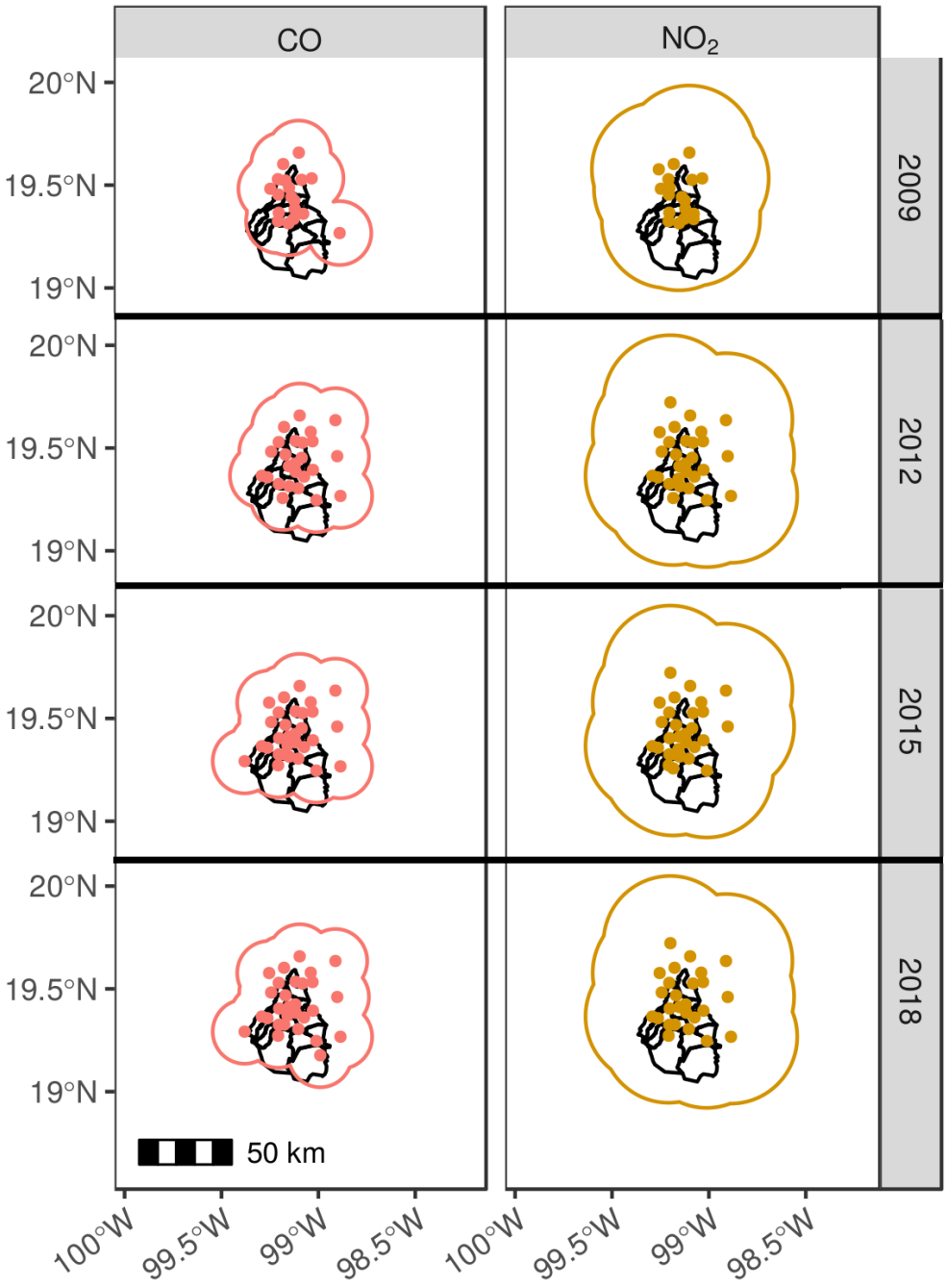


Figura 7.5: Contaminant representatives radius. The panel is arranged in a matrix style, where the columns correspond to CO and NO_2 and the rows to years of measurement. In this configuration, each cell represents the Mexico City and the corresponding working monitoring stations. In addition, the contaminant representativeness zone was created by the union of circles with the radius equivalent to the spatial range.

The complete spatio-temporal evolution for these contaminants depicted in Supplementary Figure 17, from years 2009 to 2018 makes clear that regardless of the increase in monotoring stations and their siting location, since 2009 there has been an adequate coverage of the city. In Figure 7.5 we present the case of CO and NO_2. The case of well represented monitoring networks is shown for NO and NO_X in Figure 7.6, as example of pollutants with long spatial correlation, 77.15 and 84.44 km respectively (see Figure 7.4). In other words, the amount and selected location for this monitors to construct this network can be considered as effective and even has shown improvement since throughout the years (see Figure 7.5).

Another aspect of the NO and NO_X networks is that their representative area clearly exceed the city's territory, which is beneficial for both, Mexico City and the Valley of Mexico Metropolitan Area (VMMA). A relevant issue of this extended coverage is that monitoring stations installed in one of the 16 boroughs in Mexico city are able to capture the influence of pollution from sources outside the city, as this pollutants can eventually move towards the city due to dispersion phenomena.

Additionally, this results allow us to establish neighbourhood limits for interpolation purposes. This is, in order to select the proper amount of monitoring stations required to estimate concentration values at unmeasured locations, we can refer to spatial and temporal ranges of correlation to determine which stations have to be included in the analysis.

An interesting comparison between a well represented network and one with lack of representativeness is displayed in Figure 7.7. The full time evolution of the networks coverage for SO_2 and $PM_{2.5}$ is shown in Figure 7.5, and their status in year 2018 can be seen in Figure 7.7. For the case of SO_2, the central and north part of the city are almost covered, which is not the case from the center to the south. Additional fifteen stations located outside the city, i.e. in the VMMA, are partially connected to the network in the north and four are disconnected. The current status of $PM_{2.5}$ representation area is an example of an improvement compared with previous years as seen in Figure 7.5. For the year 2018 as presented in Figure 7.7 there is a complete coverage of Mexico City's surface. All areas of individual stations are connected and contrary to SO_2, this area includes the south.

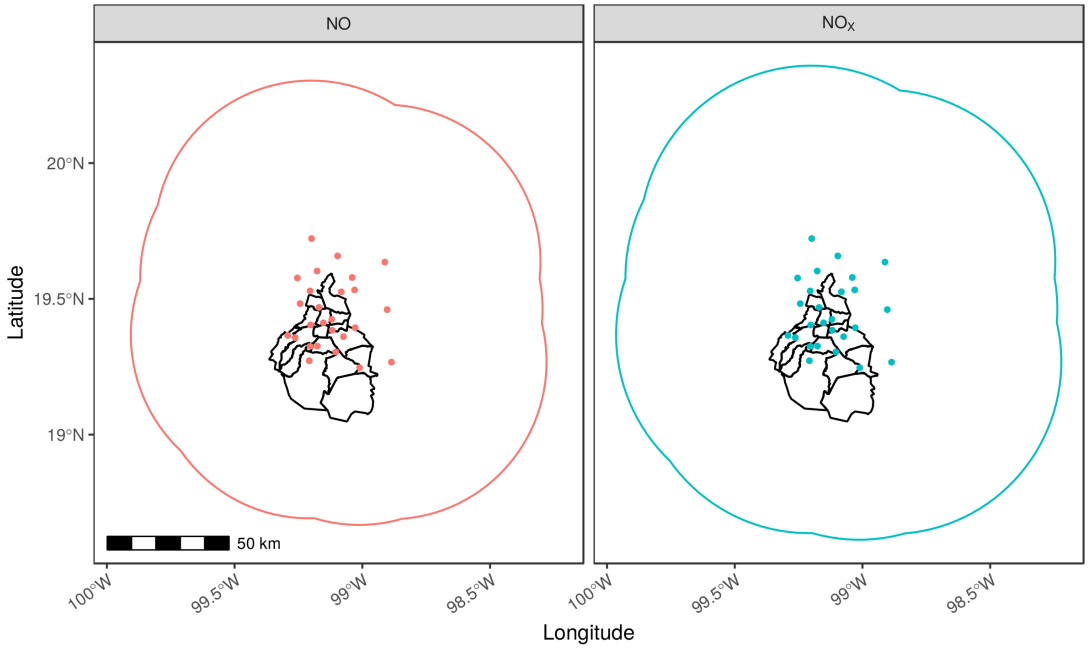


Figura 7.6: Representativeness areas for NO and NO$_X$. Covered areas for NO and NO$_X$ are shown for year 2018. Monitoring stations (colored dots) report hourly concentrations for these and other pollutants. In this case, both exhibit long spatial correlation which indicates a higher correlation between distan monitor stations.

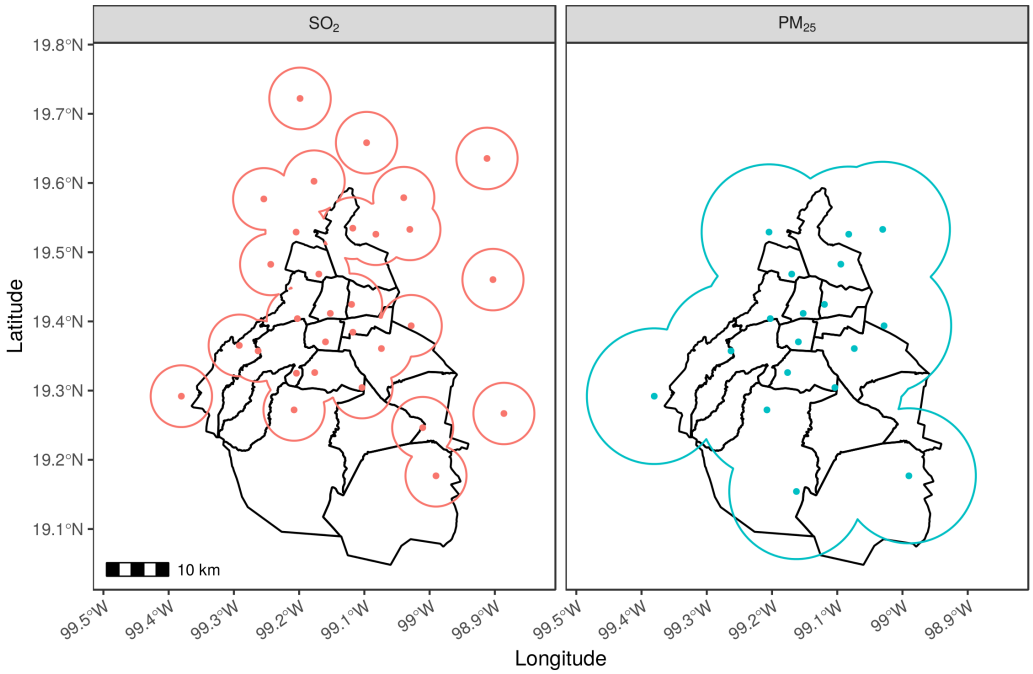


Figura 7.7: Representativeness areas for SO_2 and $PM_{2.5}$. A comparison of covered areas for SO_2 and $PM_{2.5}$ is shown for year 2018. Monitoring stations (colored dots) report hourly concentrations for these and other pollutants. Different coverage pattern which depends on the spatial range of each pollutant can be seen for SO_2 and $PM_{2.5}$. In the case of SO_2, some monitoring stations seem isolated outside the city's limits while for $PM_{2.5}$ most of the city's surface is well represented.

Discussion

Air quality assessment is crucial in terms of public health, individual and general population wellness. People's quality of life is strongly determined by the levels of contamination. Hence, to have reliable measurements for air pollutants in time and space is of outmost importance. In this work, based on spatio-temporal correlations of monitoring stations of nine air pollutants in Mexico City, we have been able to evaluate whether the location of those stations is adequate to cover the surface of the city.

The analysis showed that the distribution and amount of the monitoring stations is sufficient to evaluate the majority of pollutants, with the exception of sulphur dioxide (SO_2). The spatial range of monitoring station for SO_2 is shorter than the other pollutants. The problem does not occur in all zones of the city, it is constrained to the southern part, as it can be observed in Figures 7.1, 7.5, and 7.7A. The southern part of Mexico City has a large rural region, and concomitantly, the population density is

also short. That is possibly the reason for which there is a limited number of stations in that zone.

Regarding SO_2, this compound is mainly generated by industrial activity, which is carried into the northern side of Mexico City, and SO_2 levels in the south are likely due to the dispersion of this pollutant.

Regarding NO and NO_X, and $PM_{2.5}$, whose high levels are crucial in terms of individual and public health, the first two pollutants are well measured and estimated; however, this is not the case of $PM_{2.5}$ and SO_2 as the spatial ranges of the monitoring stations for these pollutants are short. However, in the case of $PM_{2.5}$ monitoring stations, the AJU station (43), which is the southern-most located one, it is able to measure $PM_{2.5}$, and given that, the monitoring stations able to measure this pollutant cover almost the entire surface of the city. In terms of public policies, an economic and at-hand option to increase the measured surface of SO_2 stations, is to enable the AJU station with SO_2 capacity.

It is worth to mention that the metropolitan areas are not isolated, and contamination also could move from external places. For instance, during May 2019, Mexico City experimented an unusual environmental challenge due to an elevated concentrations of $PM_{2.5}$ and O_3 (45). Several forest fires have taken place in the Tepozteco National Park and Ajusco National Park, which are both located in the southern border of Mexico City. With this in mind, we state that it is necessary to have also monitoring stations in the periphery of the city to be able to establish, based on spatio-temporal criteria, models to predict contamination indexes and have a better planning to reduce the occurrence of these episodes.

It is also important to mention that to establish a correct methodology to measure air pollutants in time and space, several factors must be taken into account. Correlation between pollutant concentration and health should be carefully evaluated in order to avoid misinterpretations. In what follows we will discuss some relevant factors that must be observed:

Depending on the type of pollutant, the resident time in the atmosphere may vary from minutes to weeks. For example, ultra-fine PM ($< 0.1\mu m$) remain suspended in the air in the range of minutes to hours. Conversely, PM10 may remain suspended from minutes to hours (Air quality criteria for particulate matter. Washington, DC, US Environmental Protection Agency, 2004 (http://cfpub.epa.gov/ncea/cfm/partmatt.cfm)).

Additionally, photochemical transformations due to sunlight radiation, meteorology, or several other factors should be taken into account in the assessment and concomitant establishment of public policies.

With this approach, we do not only provide a protocol to measure and estimate areas of representativeness for several pollutants, but also provide suggestions for public policies that are not expensive or logistically complicated and may have important impact in the evaluation of the quality of air in Mexico City and hence, to help to increase the quality of life of people living in this place.

Public health implications

The results just discussed may have important implications in the development of succesful Health Impact Assessment (HIA) programs. HIA programs are aimed at the identification, mitigation and optimization of the impacts that non-health sector policies (such as industry, society and other govermental offices) may have on public health (46). Risk quantification used to be mainly based on toxicological or biomedical studies, but more recently the scope of HIAs has broaden to incoporate more general social determinants of health (47).

The progressive incorporation of information sampling and retrieval technologies and the use of geographic information systems (GIS) to analyze the data have become a central tenet of HIA programs. The way to analyze the data however is shifting from merely transactional reports to the use of advance analytics such as the ones used in business intelligence and data science. Latin American countries have developed specialized programs to make use of GIS and computational intelligence to improve their HIA programs. Studies such as the ALBA, GeoSur or, in the case of Mexico, the Global Environmental Outlook (GEO) are aiming in this direction (29). GEO has indeed developed its own strategy within the `geotext` framework in order to use spatial analysis to provide policy makers (and even the public) with enhanced information resources, however these resources are just as good as the information they are based on (48, 49).

In the case of air pollutant monitoring stations, the WHO has actually advanced some guidelines as to what standards are desirable for the data sources to be useful in the context of HIA programs (50). Mexico City is doing partially well, according to these standards, however our results have shown that there are things that need improvement, in particular taking into account the size and urban characteristics of the metropolitan area of Mexico City as large urban areas pose particular environmental challenges (51). It has been discussed that increased risks created by urban development include unhealthy conditions which may arise from unplanned settlements or rapidly growing urban environments, environmental pollution by over-concentration of waste and other pollutants, and overcrowding, among others (52).

As it was shown here, using large scale empirical data from the monitoring network itself, some of the actual challenges have to do with the fact that the radii of coverage is actually quite different from the different pollutants (see Figure 7.5). It is worth noticing that the areas in that figure correspond to the empirical distribution of air pollutants as given by the characteristic environmental conditions of the metropolitan area of Mexico City and the particular monitoring technology available there.

These facts are indeed object of current interest, since air pollution in large Latin American cities has become a source of special concerns in recent times. According with a report from the Pan American Health Organization (PAHO) (53), the leading causes for urbain air pollution in the Americas are fossil fuels in industry and transportation. The same report states that in the case of Mexico city metropolitan area, transportation alone is responsible for some 12 % of PM_{10} particles, 30 % of $PM_{2.5}$, 5.06 % of SO_2, a staggering 98 % of CO, 79 % of Nitrogen oxides NO_X as well as 31 % of the volatile organic compounds, 51.4 % of CO_2 and up to 29.6 % of all toxic pollutants (further

details and comparison with the WHO standards and other countries in the region are presented in Table 23-1 of reference (53)).

As discussed, even if Mexico City has implemented some regulatory systems to reduce air particle concentrations, the results have not been enough to comply with national and international standards. This may be due to the fact that programs approved by policy makers have relied on inadequate air quality measurements. We argue that some of this lag may be due to a still sub-optimal monitoring approach. Indeed, PAHO has been stated that *... there is a clear need for better monitoring systems to analyze trends using more exhaustive, continuous, reliable and complex data and methodologies that are comparable between countries, so that better intervention measures could be adopted to control air pollution...* (53). Our analysis as presented here, indeed aims to diagnose some aspects of what is missing and what can be improved in terms of the spatio-temporal representativeness of the air quality monitoring stations in the metropolitan area of Mexico City.

Improving on our HIA programs and policy is extremely relevant, in particular after considering that the steady decline in, say PM_{10} particulate levels that had being observed from the early 2000s in Mexico City was overturned by a dramatic increase during the years 2008 and 2009. Even if another decline has been observed since then, we are still lagging to reach the WHO recommended levels (for further information see Figure 23-1 and Table 23-2 in reference (53)). Mexico was in fact, the country with more deaths due to outdoor air pollution than other countries in the Americas (20,496 in 2012) according with a recent survey (54).

Intervention policy and assessment

The development of analytical approaches to determine and assess environmental pollution data with the best spatio-temporal granularity is key for the design and implementation of proper intervention policy. As proposed by the *PAHO Regional Plan on Urban Air Quality and Health 2000-2009* efficient systems for air pollution health impact monitoring must include –aside from air quality management– periodic surveillance of morbidity and mortality associated with air pollution, risk assessment, effective information systems, and reliable estimation of social costs related to air pollution. In this regard, research designs such as the one advanced here will help to address some of the main concerns included in the PAHO plan and also will allow us to comply with the agreements on other initiatives such as the *Air Management Information Systems* (AMIS). To this end, Mexico (as a country) has developed a nationwide air quality monitoring program (the *Sistema Nacional de Información de la Calidad del Aire*, SINAICA https://sinaica.inecc.gob.mx/). It is worth noticing that the flagship implementation of SINAICA has been indeed the metropolitan area of Mexico City.

The information derived from the SINAICA program (in particular the one constituted in the PROAIRE initiative) has already allowed the country to develop a series of policies (the PROAIRE strategies for emission reduction) that include plans to i) encourage the reduction of emissions in industry and services, ii) preserve and restore natural resources and to prevent continued urban sprawl, iii) integrate urban development,

transport and air quality policies, iv) prevent public exposure to high levels of pollution via risk evaluation and communication, v) strengthen and enforce the already existent regulatory framework, vi) empower environmental education, research and technological development, among other. The PROAIRE website includes a quite comprehensive repository of resources useful for research and policy making that can be found at http://www.aire.cdmx.gob.mx/descargas/publicaciones/flippingbook/proaire-2011-2020-anexos/ [In Spanish].

Research efforts along these lines, although admittedly far from complete, have allowed to implement public health policy to lower the negative health impact on air pollution. Take for instance the case of Ozone, whose high levels are known to affect human health, in particular that of vulnerable or over exposed groups such as athletes, outdoor workers, asthmatics and people with respiratory illnesses and children. It has been reported that by implementing some of the recommendations in the PROAIRE initiative, average ozone levels in the Metropolitan area of Mexico City diminished from almost 0.18 parts per million (ppm) in 1991 to around 0.1 ppm in 2007 and have remained below that level ever since (53). It is expected that with such diminishing ozone levels, respiratory illnesses such as asthma, bronchial sensitivity, and lung inflammation will also present lower incidence.

It is however, complex to determine the real impact of such measures, although HIA programs have pointed out that by implementing appropriate policies up to 33,084 ozone related deaths may have been prevented in Mexico City in the period 2000-2020 (55). Another multicentric study in three of the largest cities in the Americas (Mexico City, Sao Paulo and New York) states that in the same period 2000-2020, up to 64,000 premature deaths, 65,000 cases of chronic bronchitis, 91,000 hospital admissions and a staggering 37 million person-days of lost work could be prevented just by reducing the levels of ozone and particulate matters in around 10 % by adopting greenhouse gas mitigation policies (56).

Recalling Figure 7.5, it is possible to notice that intervention policy has indeed improved the quality of monitoring stations for most (but not all) of the pollutants considered. We can notice that the metropolitan area of Mexico City has been covered well for ozone levels monitoring since 2009. This however, was not the case for $PM_{2.5}$ which was poorly covered in the southside of the city in 2009 and by 2018 has almost complete coverage. A similar case happens to CO levels monitoring which was almost uncovered in 2009 and is well covered since 2018. Other cases are still worse, most strikingly in the case of SO_2 levels which were poorly covered in the southside of the city in 2009 and still remains not covered there up to date. This is not to be disregarded since the southside of Mexico City consists mostly of residential areas with the industrial zones more widely present in the north and east sides of the metropolitan area. A closer look at Figure 7.5 and Supplementary Figure 1 would show that SO_2 monitoring station facilities have indeed improved in number and effectivity. However, due to the different coverage features of the stations for the different pollutants, these efforts have been insufficient to date. This is why spatiotemporal representation studies such as the present are relevant for public policy making.

The aforementioned facts highlight the relevance of data-driven efforts to improve health impact assessments. Aside from air quality monitoring, there are other data-centered measures that must be implemented, such is the case of exposure evaluation which is indeed essential to calculate risk levels. A number of epidemiological methodologies have been developed to assess population exposure to air pollutants. Most of them are based on the consideration of the radial distance from stations within the local monitoring networks, used as proxy to the proximity of population subjects within the study groups (19). It should be noted that besides using monitoring data for health impact assessment, modelling method using air quality models is also used to assess HIA. Since pollutant concentrations in urban environments can vary widely, geostatistical approaches to environmental epidemiology have gained even more relevance (57, 58). The present study aims to present a practical approach to this problem based on the information already gathered in the existing monitoring stations in the Mexico City metropolitan area. It is expected that data-centered studies such as this one, will motivate public policy makers to strengthen the monitoring, data gathering and data analysis strategies in large urban environments such as the metropolitan area of Mexico City. We are aware of the many challenges that effective environmental assessment has from economic, logistic and political, but also for technical and analytical reasons. However, we also believe that there are good reasons to be confident that this kind of studies will be relevant for the construction of new, more efficient models of policy making.

7.2. Artículo publicado

7.2.1. Título del artículo publicado

Metropolitan age-specific mortality trends at borough and neighborhood level: The case of Mexico City

7.2.2. Página frontal del manuscrito

PLOS ONE

RESEARCH ARTICLE

Metropolitan age-specific mortality trends at borough and neighborhood level: The case of Mexico City

Karol Baca-López[1,2☯], Cristóbal Fresno[3☯], Jesús Espinal-Enríquez[2], Miriam V. Flores-Merino[4], Miguel A. Camacho-López[1], Enrique Hernández-Lemus[2,5]*

1 School of Medicine, Autonomous University of the State of Mexico, Toluca, State of Mexico, Mexico,
2 Computational Genomics Department, National Institute of Genomic Medicine, Mexico City, Mexico,
3 Technology Development Department, National Institute of Genomic Medicine, Mexico City, Mexico,
4 School of Chemistry, Autonomous University of the State of Mexico, Toluca, State of Mexico, Mexico,
5 Centro de Ciencias de la Complejidad, Universidad Nacional Autonoma de Mexico, Mexico City, Mexico

☯ These authors contributed equally to this work.
* ehernandez@inmegen.gob.mx

OPEN ACCESS

Citation: Baca-López K, Fresno C, Espinal-Enríquez J, Flores-Merino MV, Camacho-López MA, Hernández-Lemus E (2021) Metropolitan age-specific mortality trends at borough and neighborhood level: The case of Mexico City. PLoS ONE 16(1): e0244384. https://doi.org/10.1371/journal.pone.0244384

Editor: Wenjia Zhang, Peking University Shenzhen Graduate School, CHINA

Received: April 11, 2020

Accepted: December 8, 2020

Published: January 19, 2021

Peer Review History: PLOS recognizes the benefits of transparency in the peer review process; therefore, we enable the publication of all of the content of peer review and author responses alongside final, published articles. The editorial history of this article is available here: https://doi.org/10.1371/journal.pone.0244384

Data Availability Statement: All relevant data are within the paper and its Supporting information files.

Abstract

Understanding the spatial and temporal patterns of mortality rates in a highly heterogeneous metropolis, is a matter of public policy interest. In this context, there is no, to the best of our knowledge, previous studies that correlate both spatio-temporal and age-specific mortality rates in Mexico City. Spatio-temporal Kriging modeling was used over five age-specific mortality rates (from the years 2000 to 2016 in Mexico City), to gain both spatial (borough and neighborhood) and temporal (year and trimester) data level description. Mortality age-specific patterns have been modeled using multilevel modeling for longitudinal data. Posterior tests were carried out to compare mortality averages between geo-spatial locations. Mortality correlation extends in all study groups for as long as 12 years and as far as 13.27 km. The highest mortality rate takes place in the Cuauhtémoc borough, the commercial, touristic and cultural core downtown of Mexico City. On the contrary, Tlalpan borough is the one with the lowest mortality rates in all the study groups. Post-productive mortality is the first age-specific cause of death, followed by infant, productive, pre-school and scholar groups. The combinations of spatio-temporal Kriging estimation and time-evolution linear mixed-effect models, allowed us to unveil relevant time and location trends that may be useful for public policy planning in Mexico City.

Introduction

Identifying trends in Metropolitan Mortality (MM) is a challenging problem nowadays. Systematic approaches to discriminate the relevance of social, economic, demographic, educational, environmental or criminal factors in MM are matters of intense current research [1–3].

MM can be differentiated from rural mortality since the causes and risk factors are substantially different [3]. To develop a useful model that may explain the evolution in time and space

7.2.3. Carta de aceptación

Solo tengo email de confirmación-aceptación

Notification of Formal Acceptance for PONE-D-20-10411R2 - [EMID:bc7c03a0de4a596b]

1 mensaje

PLOS ONE <em@editorialmanager.com> 21 de diciembre de 2020, 22:44
Responder a: PLOS ONE <plosone@plos.org>
Para: Karol Baca-Lopez <kbaca@inmegen.edu.mx>

You are being carbon copied ("cc:'d") on an e-mail "To" "Enrique Hernandez-Lemus" ehernandez@inmegen.gob.mx CC: "Karol Baca-Lopez" kbaca@inmegen.edu.mx, "Cristobal Fresno" cfresno@inmegen.gob.mx, "Jesús Espinal-Enriquez" jespinal@inmegen.gob.mx, "Miriam V. Flores-Merino" mvfloresm@uaemex.mx, "Miguel A. Camacho-Lopez" macamachol@uaemex.mx

PONE-D-20-10411R2
Metropolitan age-specific mortality trends at borough and neighborhood level: The case of Mexico City

Dear Dr. Hernandez-Lemus:

I'm pleased to inform you that your manuscript has been deemed suitable for publication in PLOS ONE. Congratulations! Your manuscript is now with our production department.

If your institution or institutions have a press office, please let them know about your upcoming paper now to help maximize its impact. If they'll be preparing press materials, please inform our press team within the next 48 hours. Your manuscript will remain under strict press embargo until 2 pm Eastern Time on the date of publication. For more information please contact onepress@plos.org.

If we can help with anything else, please email us at plosone@plos.org.

Thank you for submitting your work to PLOS ONE and supporting open access.

Kind regards,
PLOS ONE Editorial Office Staff

on behalf of
Dr. Wenjia Zhang
Academic Editor
PLOS ONE

7.2.4. Resumen

Entender los patrones espaciales y temporales de las tasas de mortalidad en una metrópoli altamente heterogénea, es una cuestión de interés para las políticas públicas. En este contexto, no hay, hasta donde sabemos, estudios previos que correlacionen las tasas de mortalidad espacio-temporales y por edad en la Ciudad de México. Se utilizó la modelación Kriging espacio-temporal sobre cinco tasas de mortalidad por edad (de los años 2000 a 2016 en la Ciudad de México), para obtener una descripción tanto espacial (municipio y colonia) como temporal (año y trimestre) de los datos. Los patrones específicos de mortalidad por edad se han modelado utilizando un modelo multinivel

para datos longitudinales. Se han realizado pruebas posteriores para comparar las medias de mortalidad entre ubicaciones geoespaciales. La correlación de la mortalidad se extiende en todos los grupos de estudio hasta 12 años y hasta 13,27 km. La tasa de mortalidad más alta tiene lugar en la delegación Cuauhtémoc, el núcleo comercial, turístico y cultural del centro de la Ciudad de México. Por el contrario, la delegación Tlalpan es la que presenta las tasas de mortalidad más bajas en todos los grupos de estudio. La mortalidad postproductiva es la primera causa de muerte por edad, seguida de los grupos infantil, productivo, preescolar y escolar. Las combinaciones de estimación espacio-temporal de Kriging y de modelos lineales de efectos mixtos con evolución temporal, permitieron develar tendencias temporales y de localización relevantes que pueden ser útiles para la planeación de políticas públicas en la Ciudad de México.

7.2.5. Apartados del artículo

Introduction

Identifying trends in Metropolitan Mortality (MM) is a challenging problem nowadays. Systematic approaches to discriminate the relevance of social, economic, demographic, educational, environmental or criminal factors in MM are matters of intense current research (59, 60, 61).

MM can be differentiated from rural mortality since the causes and risk factors are substantially different (61). To develop a useful model that may explain the evolution in time and space of MM, is crucial to have a comprehensive information regarding the aforementioned factors. However, missing data is a common problem in developing and emerging countries, particularly for small-scale spatial and temporal level measurements. Diverse limitations and restricted access to public health data to collect information at the neighborhood and monthly level, lead to the modifiable areal unit problem (MAUP). It implies that a bias is generated affecting statistical hypothesis testing due to the combination of two or more spatial scales. To address this problem there have been several alternatives such as: correct the variance-covariance matrix using samples from individual-level data; focus on local spatial regression rather than global regression; design areal units to maximize a particular statistical result; developing statistics that change across scales in a predictable way; Bayesian hierarchical models as a general methodology for combining aggregated and individual-level data, or, using simulated data.

One of the most useful variables to explain MM is the age group, since the causes associated to risk factors, disease incidence and other several related variables are substantially different between such groups. Seminal works regarding age-specific MM can be found elsewhere (62, 63, 64). Often, these age groups are separated into non-overlapping sets as follows:

1. Infant ($x < 1$ years old)

2. Pre-school ($1 \leq x < 4$ years old)
3. School ($4 \leq x < 14$ years old)
4. Productive ($14 \leq x < 64$ years old)
5. Post-productive ($x \geq 64$ years old)

In this work, we refer to the productive group as the economically productive population. Post-productive group is referred as economically dependent from the productive population. In addition, age-specific mortality can be influenced by social determinants at the individual level and at national or state level (65, 66). Individual mortality is influenced by personal-level characteristics such as genetics, socioeconomic status and education.

However, there is growing evidence that collective or regional disadvantages can also be good predictors of individual mortality and population level mortality (59, 67). Example of these regional disparities can be classified into social, demographic and environmental factors. In a recent study, Gavourova and Toth (2019) described how environmental factors influence changes in preventable mortality and how they impact differ from district to district in Slovakia (60). In another context, individual and regional characteristics have been simultaneously analyzed for cardiovascular disease, to elucidate the effects of their interaction with air pollution, psychosocial stress, adverse childhood experiences and neighborhood deprivation index (68, 69).

Regarding the description of spatial mortality trends, two main approaches are used: i) All-cause or ii) Cause-specific mortality studies. In the first case, spatial and temporal variability have been measured, using different levels of granularity (70, 71, 72, 73, 74). For cause-specific mortality several models have been developed for diseases such as cancer, diabetes, hypertension, chronic obstructive pulmonary disease, cardiovascular disease, hepatitis C and HIV/AIDS (64, 71, 74, 75, 76, 77, 78, 79, 80, 81, 82, 83). As an example of cause-specific mortality, Dwyer-Lindgren, et al. (2017), studied the variations in life expectancy, mortality rates and years of life lost from 152 causes of death at the county and neighborhood levels by age group and sex. They mainly conclude that county level estimations mask important local differences (between neighborhoods) (72, 84). Studies with these characteristics have been performed in developed countries which generally have large, accessible and almost complete databases, thus, there is no MAUP.

In developing and emerging countries, it has been described an exacerbated problem in health disparities due to poverty, environmental threats, inadequate access to health care and educational inequalities that may lead to higher mortality rates (62). In most cases, the conducted studies about mortality rates and other health outcomes rely on data that is not documented or incomplete. The last can be due to the lack of economic resources from public institutions or data restrictions for research purposes.

Mexico City, is one of the most populated cities worldwide. Its economy, employment rates and health services have improved over the last decades. Unfortunately, these resources are unevenly distributed, mainly affecting three risk groups: children,

the elderly and the poor (85). To worsen this situation, exposure to environmental risk factors derived from urbanization and migration has increased, with its associated negative health effects gaining attention in recent years (86, 87, 88, 89).

To set some context of mortality causes and trends in Mexico City, in a recent study, Aburto et al. (2018) analyzed lifespan and preventable mortality for Mexico City and the other 31 states of the country. They described changes in age groups. Particularly, they reported an increase in Diabetes and heart diseased-related mortality rates (90). Furthermore, beyond diseased related mortality causes, an important number of deaths in Mexico City corresponds to homicides. Aburto et al. conducted different studies for the periods 2000-2010 and 2005-2015 regarding the impact of homicides in life expectancy and lifespan inequality. The authors reported an increase in homicide mortality that surpassed positive outcomes from health care reforms oriented to promote life expectancy at the national level (91, 92).

Regarding health outcomes, Gómez-Dantés, et al. (2016) report the leading risk factors for children and adults, supporting the fact that public health reforms and interventions should vary according to the specific risk factor for each age group. To give some examples of the latter, diarrhoeal diseases, undernutrition and poor sanitation were the leading risk factors for children, meanwhile for adults, chronic diseases and violence resulted the highest risk factors (93).

Despite the above mentioned previous background, at the borough and smaller scales, health, socioeconomic, educational or environmental disparities for Mexico City have not been formally quantified. In general, quality of life indexes of the urban and rural blocks are drastically different. However, these differences may be larger between urban neighborhoods than between rural and urban blocks.

Recently, new estimations suggested a high degree of social backwardness (94). In Mexico City, health disparities such as illness and mortality rates can vary between close neighborhoods, just a few blocks apart, within the same borough. Taking this into account, assessment of health outcomes such as MM, using borough as a measurement unit, may result inadequate. A similar behavior arises from considering the temporal component: coarse-grained measurements may mask the variant behavior of MM. Once again, the MAUP is a matter to take into consideration at both spatial and temporal scales.

It is known that temporal variations in environmental factors (air pollutants for example) and socioeconomic variations (socio-economic level among boroughs and neighborhoods) can lead to biased results when not having measurements at the appropriate scales. Thus, there is an urgent need to count on fine-grained data in order to better interpret spatio-temporal correlations in health outcomes. To the best of our knowledge, there is no previous report in which time lapses for measurements were taken into smaller scales than in a year's time. Nevertheless, both spatio-temporal data are available in Mexico City for MM. These data include, at the borough level, information for general mortality, age, gender and other descriptive variables from 2000 to 2016 in a year basis.

In this work, using available data for Mexico City, a two-level spatial and temporal

description for MM was analyzed using geo-statistical interpolation. In this context, the spatial component is tackled at borough and neighborhood scale, whereas the time scale was used in a year and trimester levels. Thus, using a multilevel modeling for longitudinal data provides a finer description of these phenomena that may result in a more accurate spatio-temporal explanation of MM in Mexico City. This in turn may allow to better capture the multi-scale complexity of mortality patterns in large urban areas such as Mexico City aiming at improving integrative policy designs.

Materials and methods

Regarding the data collection in Mexico, the National Institute of Statistics, Geography and Informatics (INEGI, Instituto Nacional de Estadística, Geografía e Informática) (39), each ten years develops a country-level census of population. In that national survey, economic, demographic and social data of all citizens in Mexico are collected and stored in a public database. Hence, INEGI provides the core of socioeconomic and demographic data, used as input for those databases that we used in this work. In addition, the mortality data was obtained from the Secretariat of Health in Mexico City (Secretaría de Salud de la Ciudad de México: SEDESA) (95). Detailed description of both databases can be found in the following subsections.

Study area

The study area is the capital of Mexico, Mexico City which was known as the Federal District (Distrito Federal) until 2016 (Fig 7.8A). Mexico City belongs to the Valley of Mexico Metropolitan Area, the biggest metropolis in the central region of the country. The city is divided into 16 administrative boroughs (municipalities) and 2,097 neighborhoods, according to the geo-spatial information, i.e., polygon shapefiles, obtained from the Geostatistic framework, December 2018 (Marco Geoestadístico, Diciembre 2018) (96).

The neighborhood polygons were combined by borough identifiers to create the corresponding borough regions (Fig 7.8B), in order to have a multilevel hierarchical description of Mexico city. Borough centroids, were obtained using *rgeos* R package (97) (Fig 7.8C). In addition, neighborhood centroids were obtained by *gCentroid* or *gPointOnSurface* for convex or concave polygons respectively. All maps were created using R-software libraries: `sf`(98), `rgeos`(97), `raster`(99), `geosphere`(100), `spacetime`(101), `sp`(102), `rgdal`(103), `ggplot2` (104), `cowplot`(105), `gridExtra`(106) and `ggspatial`(107).

A mortality database was created from available open data obtained at the Secretariat of Health in Mexico City (Secretaría de Salud de la Ciudad de México: SEDESA) (95). The database comprehends a 16 borough, age-specific mortality rates follow up complete records, from the year 2000 up to 2016 and can be downloaded from (108).

The age-specific groups include the following five non-overlapping age descriptors:

1. Infant ($x < 1$ years old)
2. Pre-school ($1 \leq x < 4$ years old)

3. School ($4 \leq x < 14$ years old)
4. Productive ($14 \leq x < 64$ years old)
5. Post-productive ($x \geq 64$ years old)

In this work, we refer to the productive group as the economically productive population. Post-productive group is referred as economically dependent from the productive population. The mortality rates are expressed, in all cases, per $1,000$ inhabitants, but the Infant rate, which was calculated per $1,000$ liveborn according to the data obtained from INEGI and the Mexican National Population Council (Consejo Nacional de Población, CONAPO) (39, 108).

A five year time-step borough level evolution descriptive panel for every age-specific group is presented in Fig 7.9. Rows stand for age-specific mortality rate, and columns for a summarized time period (year) as reported by SEDESA. Naturally, with the available data at hand, the only alternative to visually explore the spatial data component, is to homogeneously color each borough with a single mortality rate for the reported year. Fortunately, there is discontinuity among boroughs, i.e., adjacent neighbor at using borough as area unit. Unfortunately, in this representation, neither time evolution nor spatial mortality age-specific rates are represented.

To overcome this drawback, a visual description at borough level is presented in Fig 7.10. It can be seen that there are different age-specific time evolution patterns (linear trend, in blue), such as a decreasing mortality rate for Infant and Post-productive groups, in contrast to the increasing trend in the Productive case. Although the Cuauhtémoc borough (in green) is the smallest one in terms of its neighborhoods (38 in total), it seems to overcome the mortality rates for most age-specific cases in comparison to the rest of the boroughs. No apparent time-dependent correlation can be observed for School and Pre-school, where several borough vibrant mortality curves are presented.

Spatio-temporal interpolation

In order to have a more robust input dataset to analyze trends in MM, the measured variables must have a temporal component as fine-grained as possible, since it is well known that several risk factors associated with mortality have a cyclic behavior, grounded on temperature, air pollution, seasonal pathogens or even individual social aspects (109, 110, 111). Several studies on MM have taken into account yearly data to associate certain variables with the response outcome (67, 77, 78, 79). Often, monthly data are not available for developing countries; in other cases, a great number of missing data results very common.

In general, although different area-level health outcomes might share a variety of explanatory variables such as socioeconomic, pollution, delinquency levels, health access, among others; by down scaling, spatio-temporal heterogeneity might arise (77, 78, 112, 113, 114). To overcome the MAUP, a variety of geo-statistical procedures have been implemented to estimate mortality rates at different granularity levels. For instance, Population-Weighted Average, local and global Empirical Bayes and Poisson Kriging

have been used to estimate disease-specific mortality rates from age-adjusted data. Accounting for spatial correlation patterns for low and high frequency rates, Poisson Kriging have shown better results (11, 115). Kriging methods allow to estimate spatial risk considering heterogeneity among small areas from poorly reported databases (12).

In this context, area-to-point (AtP) Kriging provides instantaneous estimation of the spatial regression, which is valid for each time point, i.e., it is appropriate for a cross-sectional study, nonetheless, we have both longitudinal and spatial data points. In addition, whenever a time-wise progression estimate is needed, AtP falls short of considering spatio-temporal correlations unless an appropriate time-regression procedure is considered, i.e., one compliant with Gauss-Markov theorem. In this case, a sum-integrated joint method - minimizing error via least squares regression - will give rise to a valid measure; formally a best linear unbiased estimator. Doing this, it will however be formally equivalent to a joint (sumMetric) integrated spatio-temporal estimator.

To overcome data scarcity, spatio-temporal Kriging has been used to jointly interpolate spatial missing data at the county level as well as for temporal interpolation in mortality data (64, 70, 116). Kriging estimates has been also used to describe changes in child mortality trends, evaluating between-countries and within-country sources of variation (63). However, in current literature, a standard spatio-temporal methodology that properly addresses heterogeneity and data scarcity, has not been reported yet. The spatio-temporal description of the mortality database used in this work has been constructed on a yearly basis at the borough level.

Although, both spatial and temporal description levels can be accurate for open data summarized statistics, such description fails to represent the actual statistical metric unit, e.g., trimester measurements at the neighborhood level. In addition, it is known that in this context, data points have a spatio-temporal correlation which can be used to estimate values at unmeasured locations in space and time. Among the different geo-statistical interpolation methodological alternatives, the Kriging family provide unbiased estimates that minimize the variance-covariance spatio-temporal structure using an appropriate variogram modeling to tackle the MAUP (117).

Briefly, a second data description level in both space and time, using neighborhood centroid locations and a trimester time basis, was carried out using Global Ordinary Kriging, for each age-specific mortality rate. All estimations were obtained using *gstat* R package implementation (42, 43). Initial variogram parameter values were obtained from the complete mortality empirical (s)patio-(t)emporal rates variogram matrix $\gamma(s,t)$ (**??** Table 1):

- **Nugget**: The median value of the first three empirical variogram matrix row/column means, for the spatial or temporal initial guess respectively.
- **Range**: The spatial range is one third of the lagged maximum spatial value; for the temporal case, it corresponds to the maximum value.
- **Sill**: The median value of the last five empirical variogram matrix row/column means, for the spatial or temporal initial guess respectively.

- **stAni**: The spatio-temporal anisotropy was estimated using a linear model as implemented in *gstat* R package (42, 43).
- **Joint spatio-temporal initial values**, are based on the mean of the independent spatial and temporal values, respectively.

Using the initial variogram parameters, different spatial, temporal or joint covariance structures were tested to find the best parsimonious correlated data description, according to the available implementations in *gstat* (*metric*, *separable*, *productSum*, *sumMetric* and *simpleSumMetric*) (42, 43). All possible single, double or triple variogram combinations (Exponential, Gaussian and Spherical) were tested according to the corresponding covariance structure. Hence, computational power was used in order to find the best variance-covariance spatio-temporal structure using the appropriate variogram in order to use the parsimony principle. All covariance models were fitted using a quasi-Newton box constrained method, where only the lower-bound was set to 0.001 for every parameter. The upper-bound was left to its default value (infinite), i. e., no box constraint was imposed for the maximum value. The selection criterion to choose a covariance structure model, was to minimize the weighted mean squared error. Complete age-specific mortality tested covariance fitted model results can be found in **??** Table 2. Finally, the spatio-temporal mortality rate interpolation was obtained using all the available data points, under the best covariance model description, for the neighborhood centroids at a trimester level, in the period 2000 to 2016. Under this configuration the complete estimation root mean square error for the space-time Kriging data points was 8.16×10^{-10} (see **??** Table 3 for age-specific errors).

Time-evolution modeling

In the literature, there are many alternatives to assess the mortality time-evolution patterns using different types of models such as multilevel modeling for longitudinal data, longitudinal multilevel model, and longitudinal hierarchical linear model, among others. All of them target the different ways to model the variance-covariance structure, taking advantage of the inter/intra data structure and/or modifiable area unit.

Raudenbush & Bryk 1986 (118) took advantage of observed unit (individuals) when considering inter and intra school effects. Hence, their proposal can bee seen as a hierarchical or multilevel approach, where first, the within-group model is estimated by a separate regression equation for each school (referenced as Eq 1 in the original work). Then, the between-group model uses the regression coefficients as dependent output and try to model the within-school structural relationships (Eq 2 in (118)). Finally, the Raudenbush & Bryk can be coupled into a single equation by substitution of Eq 2 into Eq 1. The resulting equation allows to model the error term with many degrees of freedoms associated with the studied variables of interest.

However, the model estimation itself is not explained in detail. On the other hand, Kwok et al. 2008 (119) modeled variable time data points, but, they were focused on time correlation, hence, they introduce first-order auto-regression structure (AR(1)) to model the variance-covariance structure using SPSS (MIXED) and SAS (PROC MI-

XED) procedures. The two previous procedures are in essence, Linear Mixed-Effects Models. Recently, Anaya & Al-Delaimy 2017 (120) and Green et al. 2019 (121) did implement multilevel modeling for longitudinal data using linear mixed-effects models with R software (122) using the `lme4` package.

In order to assess the mortality contribution of the age-specific group, the spatial unit area (borough or neighborhood) and polynomial time evolution, we used a linear mixed-effects model to account for data constraints and lack of error independence using the definitions of Eq (7.1- 7.4) (123). The model was specified using *Infostat* software version 2018, which is an R (122) front-end as follows (124):

$$\begin{aligned} y_{ijk} = \mu + \alpha_i + \beta_j + \gamma \times t_k + \delta \times t_k^2 + \zeta \times t_k^3 + \alpha_i \times \beta_j + \alpha \times \gamma \times t_k + \\ \alpha_i \times \delta \times t_k^2 + \alpha_i \times \zeta \times t_k^3 + \beta_j \times \gamma \times t_k + \beta_j \times \delta \times t_k^2 + \beta_j \times \zeta \times t_k^3 + \\ \alpha_i \times \beta_j \times \gamma \times t_k + \alpha_i \times \beta_j \times \delta \times t_k^2 + \\ \alpha_i \times \beta_j \times \zeta \times t_k^3 + \varepsilon_{ijk} \qquad (7.1) \\ \varepsilon_{ijk} = \lambda_{t_k} + \nu_{ijk} \qquad (7.2) \\ \lambda_{t_k} = \phi_{\beta_j} \lambda_{t_{k-1}} + u_{t_k} \qquad (7.3) \\ var(\epsilon_{ijk}) = \sigma^2 g^2(\alpha_i) \qquad (7.4) \end{aligned}$$

where, y_{ijk} is the mortality rate for the i-th age-specific mortality group (α_i), at the j-th borough or neighborhood level (β_j), for the k-th time (t_k); μ is the global mean; γ, δ, ζ are the corresponding third order time polynomial coefficients; in addition, the double and triple complete fixed effects model interactions and the error term ε_{ijk} were also specified. Indeed, the error term ε_{ijk} in Eq (7.1) is modeled in Eq (7.2) by a two-level model, to account for the lack of independent errors; where, λ_{t_k} denotes the unobserved time effect and ν_{ijk} is the idiosyncratic error term. The correlated errors were tackled using a first order auto-regressive model for time dependence as described by Eq (7.3), where ϕ_{β_j} is the corresponding coefficient and u_{t_k} the individual effect. Finally, the heteroscedasticity in Eq (7.4) was modeled as a multiplicative effect of the residual variance σ^2 times the variance error function $g(.)$ using a *varIdent* definition of the different age-specific mortality groups α_i (124).

The model was fitted using the R language with the `nlme` package under restricted estimation of the maximum likelihood (122, 125). When possible, back-step model selection strategy was applied to remove the least significant fixed-effect term, one at the time, until no difference was found using a maximum likelihood test between competitor models. Type III sum of squares was used to assess an Analysis of Variance (ANOVA) table for marginal hypothesis tests for the fixed effects. Posterior Fisher's Least Significance Difference (LSD) tests were applied over statistical significant terms, using a multiple comparison Bonferroni p-value correction. When possible, bilateral test was used and Type I error was set to 0.05.

Results

Spatio-temporal variogram estimation

Regarding spatio-temporal age-specific mortality estimations, the initial guesses obtained from the sample spatio-temporal variograms are shown in **??** Table 1 (see material and methods section). Depending on the age-specific mortality group, the initial guesses are different for the nugget, range, sill and spatio-temporal anisotropy (stAni).

In this context, the *nugget* is the model intercept attributable to the smallest error measurements or spatial sources of variation. Interestingly, these sources of variations are negligible for Pre-school and School, in contrast to the other age-specific groups, with a wide range of nugget values ($0.01 - 21.31$). In addition, the correlation extends between measurements, also known as *range*, in all cases is exactly the same for all age-specific groups and last about 12 years for as far as 13.27 km.

The variogram values obtained at the *range*, a.k.a. the *sill*, is as close to the *nugget* for the School and Pre-school age-specific groups, and departs from it at most to double its value for Infant group. The *anisotropy* remains the same for Post-productive, School and Infant groups, but differ for Productive and Pre-school counterparts. Final covariance model weighted mean square error for all the tested variogram permutations can be found in **??** Table 2.

It is worth to mention that the lowest error for the different covariance structure methods included the *metric* case for School and Pre-school; *sumMetric* for Productive and Infant and *simpleSumMetric* for Post-productive was the best one. Within these covariance models, there was no apparent pattern in the winner variogram model tested permutation (temporal + spatial + joint). The Gaussian + Gaussian (Gau + Gau) was the winner's choice of temporal and spatial in the Post-productive age-specific group.

Moreover, Infant mortality also followed this pattern, with an additional Gau component for the joint variogram. The Exponential + Gaussian (Exp + Gau) was the winning choice for Productive mortality and Gau joint variogram. Indeed, this was a data-driven approach that required to explore the complete permutation grid, in order to reach a parsimonious spatio-temporal correlation model. A visual comparison of each winner covariance model and sample variogram can be found in **??** File 1. Finally, under these configurations the complete root mean square error estimation for the space-time Kriging data points was 8.16×10^{-10} (see **??** Table 3 for age-specific errors).

Two-level mortality rate spatial description in Mexico City

Productive mortality rate

The spatio-temporal mortality description in Mexico City starts at the raw data presented in Fig 7.9. Let us consider the whole picture, using, for example, productive mortality rate as presented in Fig 7.11. Using the borough spatio-temporal granularity top panel in Fig 7.11, there is a clear spatial pattern, no matter the selected year (2000, 2005, 2010 or 2015). It seems that there is a global maximum mortality rate value (hotspot) at Cuauhtémoc borough (number six in Fig 7.8C). There, the mortality rate values radially decrease as long as we depart from this location to the outer

boundaries of Mexico City. Indeed, the decrease is not homogeneous, i. e., it is dictated by a spatial anisotropy where the north and northeast direction have a less marked decline in comparison to the south and south-west direction. Moreover, at this temporal description, there is also an increment in the global maximum rates as we move from 2000 to 2015, in agreement with the temporal patterns presented in Fig 7.10 for the Global case.

Moving towards a deeper data exploration, the spatio-temporal interpolation obtained by Ordinary Kriging provides a productive mortality rate smooth surface (see neighborhood level at Fig 7.11). For a fair comparison, the same time scale was used (years), whereas the spatial description considered the neighborhood centroids grid. With this zoom-in, a more realistic geographical-continuous mortality transitions can be observed, unlike the discrete phenomena at borough level, for adjacent boroughs, presented at borough level in Fig 7.11. The latter reflects a far less abrupt change in mortality rates from one borough to another, thus giving a continuity between neighborhoods.

Furthermore, the productive mortality rates increase throughout neighborhoods and boroughs consistently from 2000 to 2015 in both spatial-scales. Also, notice that the mortality rate scale bar has also decreased from a maximum of 6 (at the borough level) into a 4.2 when we move towards into the fine grain spatial description for this age-specific group.

If we zoom-in even further, we can see the neighborhoods at a single borough, e.g. Cuauhtémoc borough Fig 7.11, where we can distinguish neighborhood level trends. Here, Cuauhtémoc borough was selected, since it has the highest mortality rate. Although apparently imperceptible, there are distinguishable differences in mortality values among Cuauhtémoc neighborhoods. Similar differences are obtained for the rest of the boroughs as seen at borough level description in Fig 7.11.

School mortality rate

Analogously to Figure 7.11, in Figure 7.12 we may observe the mortality rate through 15 years in the school group in three different levels of granularity. As it can be appreciated in the figure, mortality patterns in the school age are not homogeneously distributed in the 16 boroughs of Mexico City. Instead, each borough has its own pattern.

Unlike the previous case regarding productive age, in the school group the spatial trend is not clear. However, a visible and measurable decrease is observed from 2010 to 2015 in practically the whole city.

It is worth noticing that the scales in both figures are different. In Figure 7.11 the upper value for mortality rate is close to 6, meanwhile for school age, the top value mortality rate is around 0.4.

The comparison between those different groups in Cuauhtémoc borough and its neighborhoods is also remarkable. Meanwhile, for school group, 2015 is the year with the lowest mortality rate, that year was the highest for the productive age. Additionally, the differences in mortality rate at the neighborhood level is more visible in the productive group than the schoool one.

Similar two-level mortality rate spatial description in Mexico City can be found for

the rest of the age-specific groups (Infant, Pre-school and Post-productive) in **??** Fig 1-3 respectively.

Time-evolution mortality rate modeling in Mexico City

So far, we have filled the spatio-temporal gaps for the missing data points for both scales within the different mortality age-specific groups. Now, that we have overcome the summary time period for the mortality rate constraint, which is usually in a year base, we can study temporal/seasonal trends within the temporal range (12 years). In addition, we can explore the between mortality rate variability at borough level or even within its borough, a.k.a. neighborhood level data description. The later is a must, in order to address the MAUP, as we are changing the spatial scale, from several kilometers long (borough level), to a very different area unit scale at neighborhood (couple of blocks), where the prior estimated spatial correlation range is as far as 13.27 km. Finally, the mortality rate itself can be decomposed into both fixed-effects and random variance-covariance structure contribution using Eq (7.1-7.4), according to the data level description as follows.

Borough level mortality rate contribution

Borough-level results are presented in Table 1. The ANOVA results showed that the only non-significant effect at the borough level, is the triple interaction that includes the time to the third power ($p = 0.22$). Hence, the mortality rate time-evolution pattern in Fig 7.10 can be parsimoniously captured by our methodological proposal. In addition, the auto-correlation parameter had an impact not as high as one would expect at borough level ($\phi = 0.12$). On the other hand, the variance function did address the different age-specific groups where Infant was the one with the highest value (11.59) followed by Post-productive (10.34), Pre-school (0.85), Productive (0.82) and School (0.43).

To further explore the mortality behavior, posterior Fisher's LSD tests results were obtained over the age-specific, α_i, Mexico City's borough, β_j, and double borough times age-specific interaction terms $\alpha_i \times \beta_j$ (Fig 7.13 A-C, respectively). The first remarkable result is the mortality contribution to the main effects α_i and β_j . Mexico City's model results evidence, for the time-period studied, suggests that the age-specific term (α_i) is the main responsible for the mortality rate when compared to the spatial borough contribution (β_j). Indeed, the mortality rate means are definitely empowered by age-specific groups (upper bounded at 50 [$x1000$]) rather than spatial borough locations, which is upper bounded at 17[$x1000$].

Secondly, the age-specific mortality rate groups do not overlap between each other, due to the different Fisher's LSD letters (A-E) in Fig 7.13A. Moreover, the Post-productive group (A) outruns any other group, but also doubles its following competitor, the Infant group (B). Interestingly, this trend also remains when we compare two consecutive groups, i.e., B vs. C, C vs D and so on. In this age-specific context, the school is the group with letter E, i.e, is the one with the least mortality rate. Interestingly, previous results could be, to some extend, the explanation that models the Mexico's actual population pyramid shape data (126) - pencil like figure, i.e., long bar

Tabla 1: Analysis of the Variance at two spatial data description level

	Borough level			neighborhood level		
Model Term	**Degs. of freedom**	**F-value**	**p-value**	**Degs. of freedom**	**F-value**	**p-value**
μ	1	17634.638	<0.0001	1	281297.3649	<0.0001
α_i	5	6981.1137	<0.0001	5	6929.1962	<0.0001
β_j	15	8.3045	<0.0001	33	6.5863	<0.0001
t	1	7.9808	0.0048	1	714.5003	<0.0001
t^2	1	34.1072	<0.0001	1	248.0783	<0.0001
t^3	1	33.6751	<0.0001	1	430.7367	<0.0001
$\alpha_i \times \beta_j$	75	20.0209	<0.0001	165	5.6587	<0.0001
$\beta_j \times t$	15	2.9735	0.0001	33	9.0428	<0.0001
$\beta_j \times t^2$	15	2.4779	0.0013	33	2.2043	<0.0001
$\beta_j \times t^3$	15	1.9233	0.0177			
$\alpha_i \times t$	5	4.172	0.0009	5	69.3982	<0.0001
$\alpha_i \times t^2$	5	7.2432	<0.0001	5	88.1181	<0.0001
$\alpha_i \times t^3$	5	7.4925	<0.0001	5	89.7661	<0.0001
$\alpha_i \times \beta_j \times t$	75	1.7383	0.0001	165	4.9837	<0.0001
$\alpha_i \times \beta_j \times t^2$	75	1.3908	0.0175			
$\alpha_i \times \beta_j \times t^3$	75	1.1254	0.2224			

Type III sum of squares was used to assess the model defined in Eq (7.1-7.4), where, μ, is the global mean; α_i the age-specific mortality term; β_j the borough or neighborhood term; t , t^2 and t^3 the third order time polynomial; and the double and triple interactions accordingly. These definitions were used at both spatial levels, i. e., borough and neighborhood. In addition, fixed effects, back-step model selection was carried out from the maximal to the current model at the neighborhood level. Empty cells correspond to discarded terms.

with sharp-pointed end.

Here, the Post-productive people are the most underrepresented group with a sharp-pointed shape at the tip of the population pyramid, due to the high mortality rate evidence of our borough level model results. Following this rationale, the second highest mortality rate group is B (Infant), which could be probably the reason why the typical pyramid shape is broken into a uniform/bar shape shared by the Infant, Pre-school, School and Productive age-specific mortality groups, influenced by their low mortality rate.

Thirdly, taking into consideration Fisher's LSD results for the 16 borough's mortality rate contribution β_j (Fig 7.13B), it can be seen three important aspects: i) Among group letters from A to I, Cuauhtémoc (A) and Tlalpan (I) are the boroughs with the highest and lowest model estimated mean values respectively; ii) The LSD mortality rate group mean letters cluster up to five boroughs per cluster, i.e., borough that share a single letter, are not statistically different after Bonferroni multiple test correction ($p > 0.05$), thus, belong to the same cluster; iii) The borough Fisher's LSD cluster structure is correlated to spatial proximity.

Fourthly, moving towards $\alpha_i \times \beta_j$ group results (Fig 7.13C), i.e., age-specific times borough model's interaction LSD group means, it can be seen that the original and predominant age-specific mortality rate contribution in panel A is modulated by the borough contribution of panel B. Thus, if we average LSD group mean results in panel C, by age-specific groups, we should return into panel's A results. At this level, we are dealing at "between" borough mortality rate level description, where the number of Fisher's LSD group letters has increased proportional to the number of possible $\alpha_i \times \beta_j$ levels.

Neighborhood level mortality rate contribution

In order to model "within" borough mortality rate data description, we need to change our attention into another spatial scale representation, a.k.a. neighborhood level. If we pick the borough with the highest mortality rate (Cuauhtémoc), we found that the model presented in Eq (7.1-7.4) is not well-suited for this level description data. Hence, a model selection process was considered in order to obtain the best parsimonious data description.

The current model results are presented in Table 1, where it can be seen that some cells are empty due to terms discarded from the analysis. Interestingly, the third and second order time triple interaction terms have been excluded from the analysis, in addition to the borough times time to the third power. Hence, in this context the modeling complexity has been reduced, at the expense of a higher autocorrelation ($\phi = 0.88$) and different variance function parameters and ranking, i. e., Productive (3.80), Post-productive (2.43), Pre-school (2.22), School (1.35) and Infant (0.04).

The Cuauhtémoc borough results can be found in Fig 7.14. Its 38 neighborhoods were numbered according to the high mortality rate downwards (Fig 7.14A). This result is complemented by Fig 7.14B, where the Fisher's LSD means were used to describe

the mortality rate landscape. Interestingly, at the neighborhood level, there is also a radial mortality rate decay starting at the central neighborhood with number one and the letter A, which correspond to Tabacalera neighborhood. Conversely, the two neighborhoods with the lowest mortality rate (letter D) are situated at the opposite borough extremes – south-west (Hipódromo de la Condesa) and northeast (Valle Gómez) borders. Moreover, the age-specific mortality patterns have changed from Post-productive and Infant mortality at the borough description in Fig 7.13A into Post-productive and Pre-school as seen in Fig 7.14C. Moreover, the Fisher's LSD test results in this borough, but at the neighborhood level, showed some regions with up to four possible over-lapping letters (Fig 7.14D).

Taking about the different mortality rate model description, we have to keep in mind that Fig 7.13 panel B "between borough" mortality rate data order of magnitude, is going to be fine grained ("within borough") modeled using the same framework depicted in Eq (7.1-7.4). The first result to be discussed is that, unlike borough level, neighborhood mortality rate data does not cope with third order neighborhood interaction and the only triple significant interaction has a linear time tendency for age-specific times neighborhood mortality, as shown in Table 1.

The second aspect, is the new insight of the mortality rate at the neighborhood level. Here, the first mortality rate level description (at borough) leaves Cuauhtemoc's borough near 15 [$x1000$]. Now, Fig 7.14 panels C and D, decomposed into a finer grain considering the same age-specific groups, but, now Cuauhtemoc's neighbors are included, respectively. At this data level description, mortality age-specific groups within Cuauhtemoc are closer to each other (same magnitude order, units) unlike borough level (one magnitude order, tenths). In addition, between the neighborhood mortality rate is almost shared by all the 34 neighborhoods (above 7 and below 8), is we consider the shared letters of the LSD results in Fig 7.14 panel D. Finally, the rest of the mortality rate contribution to sum up to 15 (Cuauhtemoc's borough mortality description), it is distributed upon the different neighborhood model terms of Eq (7.1-7.4).

Discussion

Understanding metropolitan mortality is not only relevant in terms of a mere descriptive statistical approach, but, to take into account variables that could determine, with the highest accuracy, this crucial health outcome. This knowledge, in due time, could lead to the establishment of appropriate public policies to improve citizen life quality in metropolitan areas.

In the aforementioned terms, by means of the development of a modeling approach, based on a systematic interpolation of missing data, one may observe spatio-temporal dynamics of mortality in the urban areas with higher precision. Kriging family methods have proven to be useful to achieve this goal.

In this work we have demonstrated how the improvement of the granularity level at both spatial and temporal definition, could explain some of the socio-demographic variables underlying the changes in mortality rates between boroughs. This task was

achieved by statistically interpolating those data points into a fine-grained level. In what follows, we will comment on the findings using spatio-temporal Kriging-based methods.

It can be easily noticed that Cuauhtémoc borough is the one with the highest mortality rate in all age groups, but for pre-school –which ranks in third place– as depicted in Fig 7.10. This borough has several particularities that should be commented, in order to unveil some hints towards plausible explanations, for the unique behavior observed there along time and space.

As observed in Fig 7.13C, Cuauhtémoc is the borough with the highest mortality mean value in all age groups, but in Pre-school, which shares places with Cuajimalpa and Iztapalapa boroughs. Except in that case, Cuauhtémoc has an outstanding mortality behavior over time. The case is particularly dramatic in the productive age (14 to 64 years old).

In addition, in the Productive time-evolution pattern in Fig 7.10 can be observed the different temporal behavior of Cuauhtémoc (solid green dashed line), compared to the rest of the boroughs. Cuauhtémoc is the economic center of Mexico City. There, the executive and legislative powers are placed, as well as the most important commerce hotspots. Indeed, Cuauhtémoc is the most densely populated borough of Mexico City.

In economic terms, Cuauhtémoc concentrates 4.6 % of the gross domestic product of the entire country (127). Around 5 million people pass through this borough every day, despite its population oscillates only around $500,000$ inhabitants. This is the place in Mexico City with the highest number of public transport stations. Cuauhtémoc also concentrates the largest markets of informal commerce of the city (Tepito Market, in the Morelos neighborhood). The high density and the flux of money and services may help to explain, to some extent, the different behavior of the mortality rate in the productive age in this borough.

Another point to take into account when mortality rate in Cuauhtémoc is observed, lies on high levels of insecurity for the aforementioned reasons regarding population density and economic concentration. Additionally, the Drug War launched at the end of 2006 by former President Felipe Calderón, affected mortality rates in a large part of the country (128), being the capital of the country also upset, in particular, the city downtown, Cuauhtémoc.

By taking into account the spatio-temporal Kriging, it has been possible to modeled mortality rates at the neighborhood level. With this spatio-temporal kriging model, the four places with the highest mortality rates were Tabacalera, Centro, Juárez and Doctores neighborhoods, part of Cuauhtémoc borough.

According to reports of the Executive Secretariat of the National System of Public Security System (Secretariado Ejecutivo del Sistema Nacional de Seguridad Pública, SESNSP), Centro is one of the most insecure boroughs of Mexico City and the homicide rate is the highest in the city (129). By integrating transit and other accident-related issues, with employment determinants and crime-associated mortality, it may be possible to present an explanation for which the productive age mortality presents a consistent increase in Cuauhtémoc, compared to the rest of the boroughs.

Possible intrinsic biases due to possible errors in the registry must be taken into account. For example, the record of a death is registered once a certificate of death has been provided. In some cases, it could last days, depending on diverse factors. However, in the case of Mexico City, that under-registry is extremely low since social and health services are guaranteed in practically the whole area of the city.

The model presented here also has caveats, since it shows an interpolation of the coarse-grained data at the borough level of description. For instance, according to the model, Tabacalera is the neighborhood with the highest mortality rate, however, by looking at the data, the homicide rate as an example is not as large as Centro or Morelos (129).

At this stage is not possible to disambiguate whether these inconsistencies are due to the interpolating strategy or indeed reflect different causes of death, such as the ones related to environmental factors like air pollution and other contaminants. Despite this caveat, the model shows a radial decrease of mortality in Mexico City, starting from Cuauhtémoc downtown, similar to the general behavior observed at the borough level (Fig 7.9).

Conclusions

The granularity enhancement in mortality and health outcomes data will further improve its usage in a variety of public policies, such as urban development, security, creation of health centres, hospitals, public transport and even water re-usage. In these terms, efforts in gathering data at the lowest possible source (time and space), are highly desirable for future works.

Finally, in order to build predictive models that allow to tailor public policy design, the spatio-temporal dynamics of MM should incorporate socio-demographic, environmental, economical factors and relevant covariates, as detailed as possible. Future directions include, but are not limited to, utilizing environmental data, such as air pollution, quality of water, disposition of water supplies among others. Some of the above mentioned variables are actually available as open data resources. The ultimate understanding of MM by using the data presented here, as well as environmental and other risk factors, will help in the searching for the improvement of life quality of the metropolitan areas in Mexico City.

Lorem ipsum dolor sit amet, consectetuer adipiscing elit. Etiam lobortis facilisis sem. Nullam nec mi et neque pharetra sollicitudin. Praesent imperdiet mi nec ante. Donec ullamcorper, felis non sodales commodo, lectus velit ultrices augue, a dignissim nibh lectus placerat pede. Vivamus nunc nunc, molestie ut, ultricies vel, semper in, velit. Ut porttitor. Praesent in sapien. Lorem ipsum dolor sit amet, consectetuer adipiscing elit. Duis fringilla tristique neque. Sed interdum libero ut metus. Pellentesque placerat. Nam rutrum augue a leo. Morbi sed elit sit amet ante lobortis sollicitudin. Praesent blandit blandit mauris. Praesent lectus tellus, aliquet aliquam, luctus a, egestas a, turpis. Mauris lacinia lorem sit amet ipsum. Nunc quis urna dictum turpis accumsan semper.

Figura 7.8: Mexico City study area. A) The map of Mexico shows the location of Mexico City (in red), formerly known as Distrito Federal, one of the 32 states of Mexico which is located in the central area. **B**) Borough level description of the 16 cases (in numbers) corresponding to: 1. Álvaro Obregón, 2. Azcapotzalco, 3. Benito Juárez, 4. Coyoacán, 5. Cuajimalpa de Morelos, 6. Cuauhtémoc, 7. Gustavo A. Madero, 8. Iztacalco, 9. Iztapalapa, 10. La Magdalena Contreras, 11. Miguel Hidalgo, 12. Milpa Alta, 13. Tláhuac, 14. Tlalpan, 15. Venustiano Carranza and 16. Xochimilco . **C**) Neighborhood level description. Blue lines describe borough limits, whereas white lines at neighborhood areas (in gray). Red dots depict borough calculated centroids. In panels A and B, the scale bar and north arrow are also included. Notice that some boroughs have a dense neighborhood description in comparison.

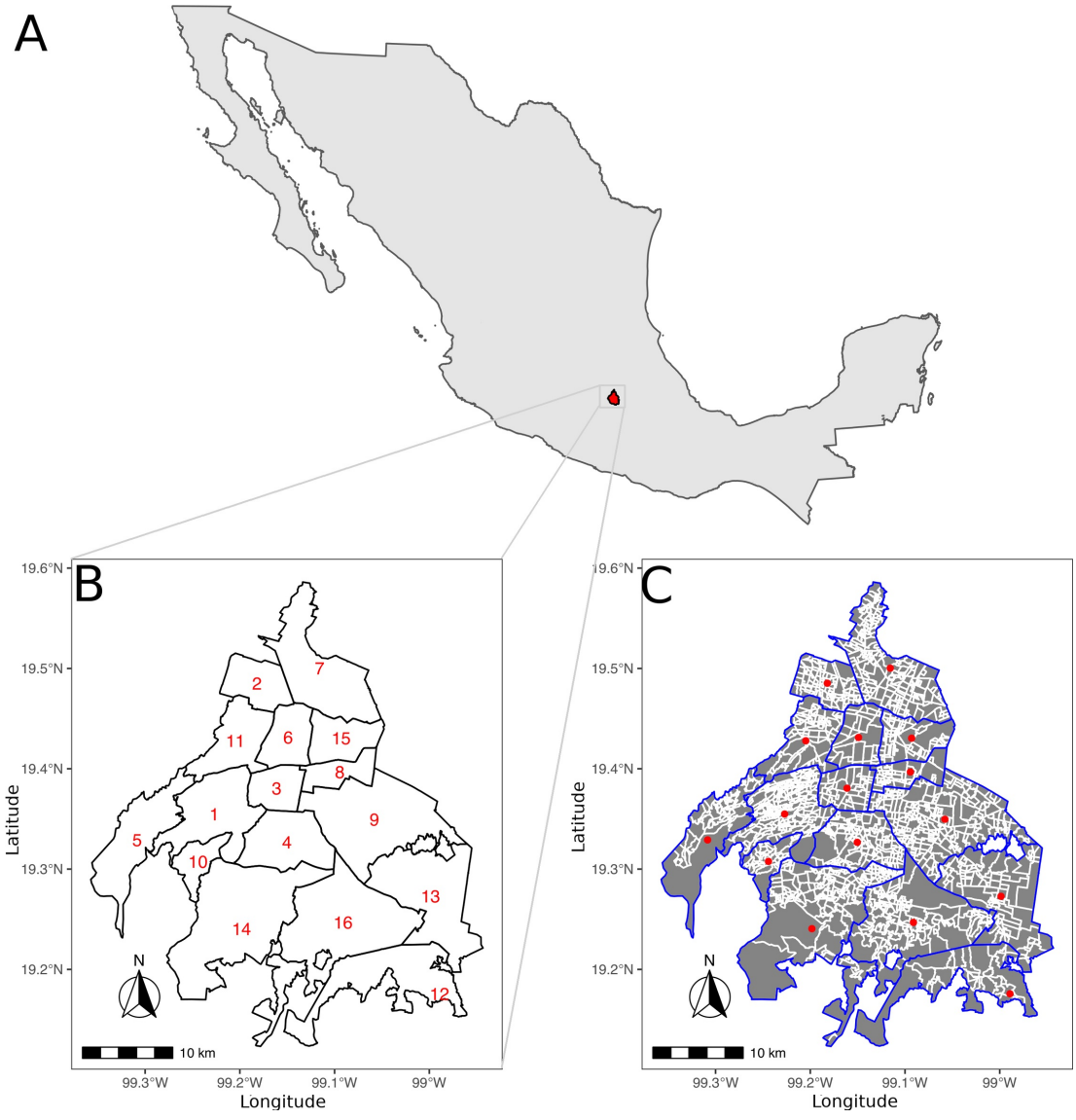


Figura 7.9: Spatial age-specific mortality rates in Mexico City's boroughs. Each row corresponds to a particular age-specific mortality rate, i.e., Post-productive ($x \geq 64$ years old), Productive ($14 \leq x < 64$ years old), School ($4 \leq x < 14$ years old), Pre-school ($1 \leq x < 4$ years old) and Infant ($x < 1$ years old). Each column stands for the selected years 2000, 2005, 2010 and 2015 from the total yearly available period 2000 to 2016. Mexico City boroughs are treated as a the unit area and color coded according to the corresponding mortality rate, which make them comparable by row. Interestingly, notice the different mortality rate ranges (color bars) depending on the age-specific group of analysis. Polygon shapefiles files can be freely downloaded at INEGI's website (96), whereas mortality data from SEDESA (108).

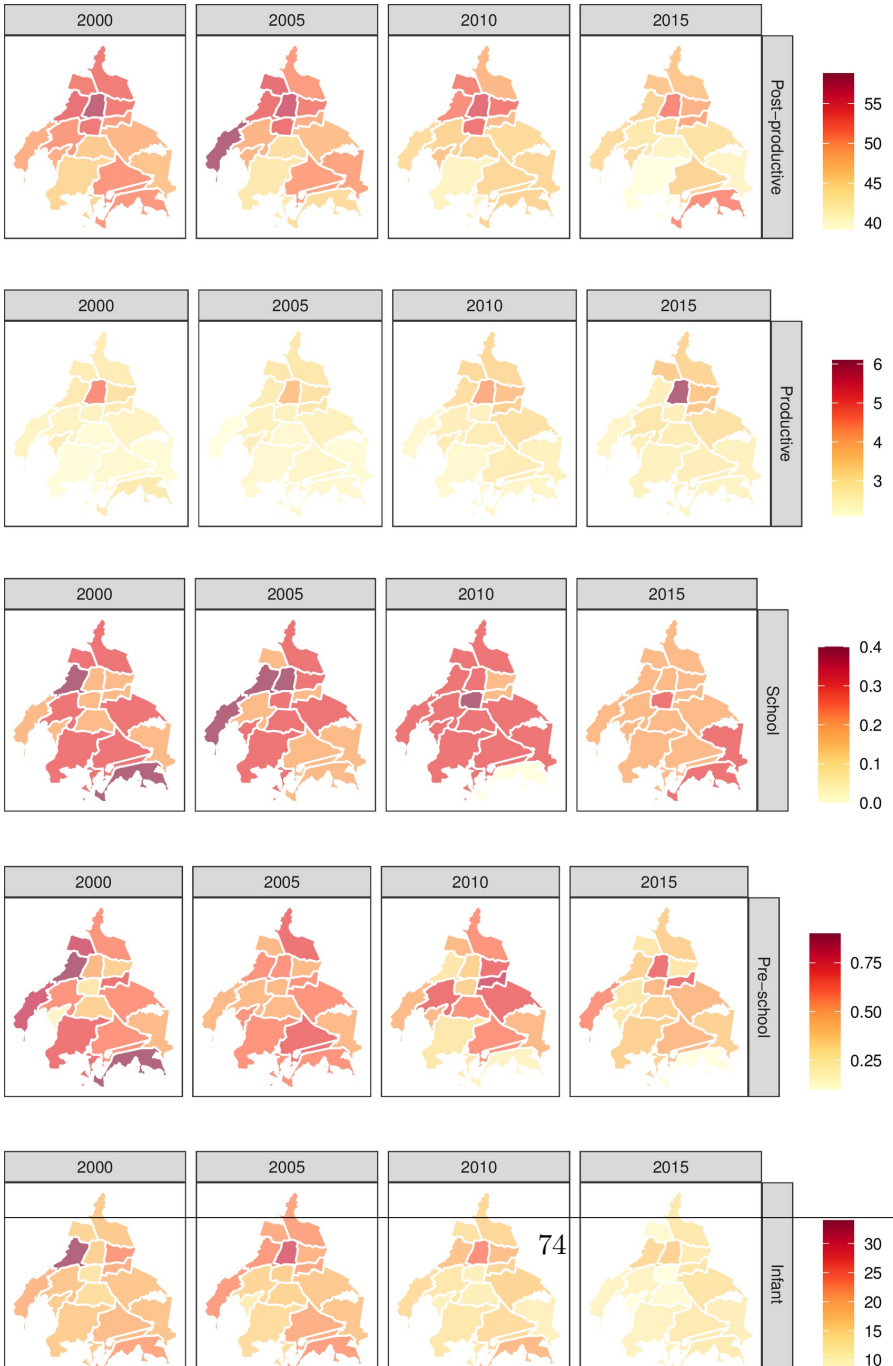


Figura 7.10: Temporal age-specific mortality rates in Mexico City's boroughs. Each panel corresponds to an age-specific mortality rate, i.e., Post-productive ($x \geq 64$ years old), Productive ($14 \leq x < 64$ years old), School ($4 \leq x < 14$ years old), Pre-school ($1 \leq x < 4$ years old) and Infant ($x < 1$ years old). All panels include complete age-specific mortality rates records by year from 2000 up to 2016 (in colour points) from SEDESA (108). Color points stand for one of the 16 boroughs in Mexico City longitudinal mortality rate data. Borough data have been joined by their corresponding color dashed lines, whereas complete age-specific panel has been modeled by a linear regression (intercept and slope, blue line) with its respective standard deviation (grey area), to get a clear picture of the time evolution pattern. Interestingly, as the years pass, the mortality rate time evolution seems to diminish for Infant and Post-productive groups, whereas the Productive age-specific counterpart tends to increase.

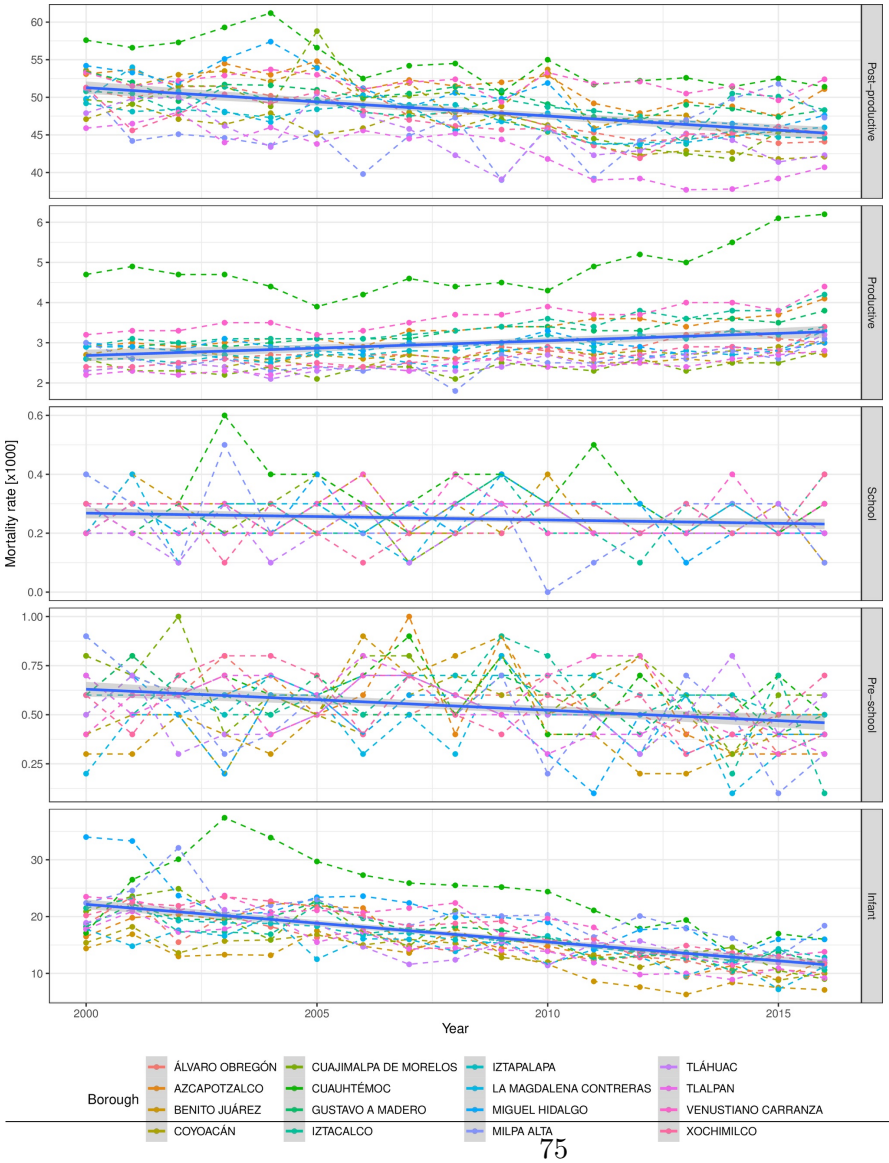


Figura 7.11: Three levels of granularity for productive mortality rate in Mexico City. First one, productive mortality rate original data description at borough level (panel **A**). Second one, productive mortality rate kriged data description at the neighborhood level is presented in panel **B**. In panel **C**, it is presented a zoom-in at a second description level for Cuauhtémoc borough. Interestingly, this is the borough with the highest productive mortality rate no matter the selected year (2000, 2005, 2010 or 2015), according to panel A and B (central borough in red). However, the mortality rate is not homogeneous at the neighborhood level, as depicted by the kriged values presented in panel C for the different years. Polygon shapefiles files can be freely downloaded at INEGI's website (96).

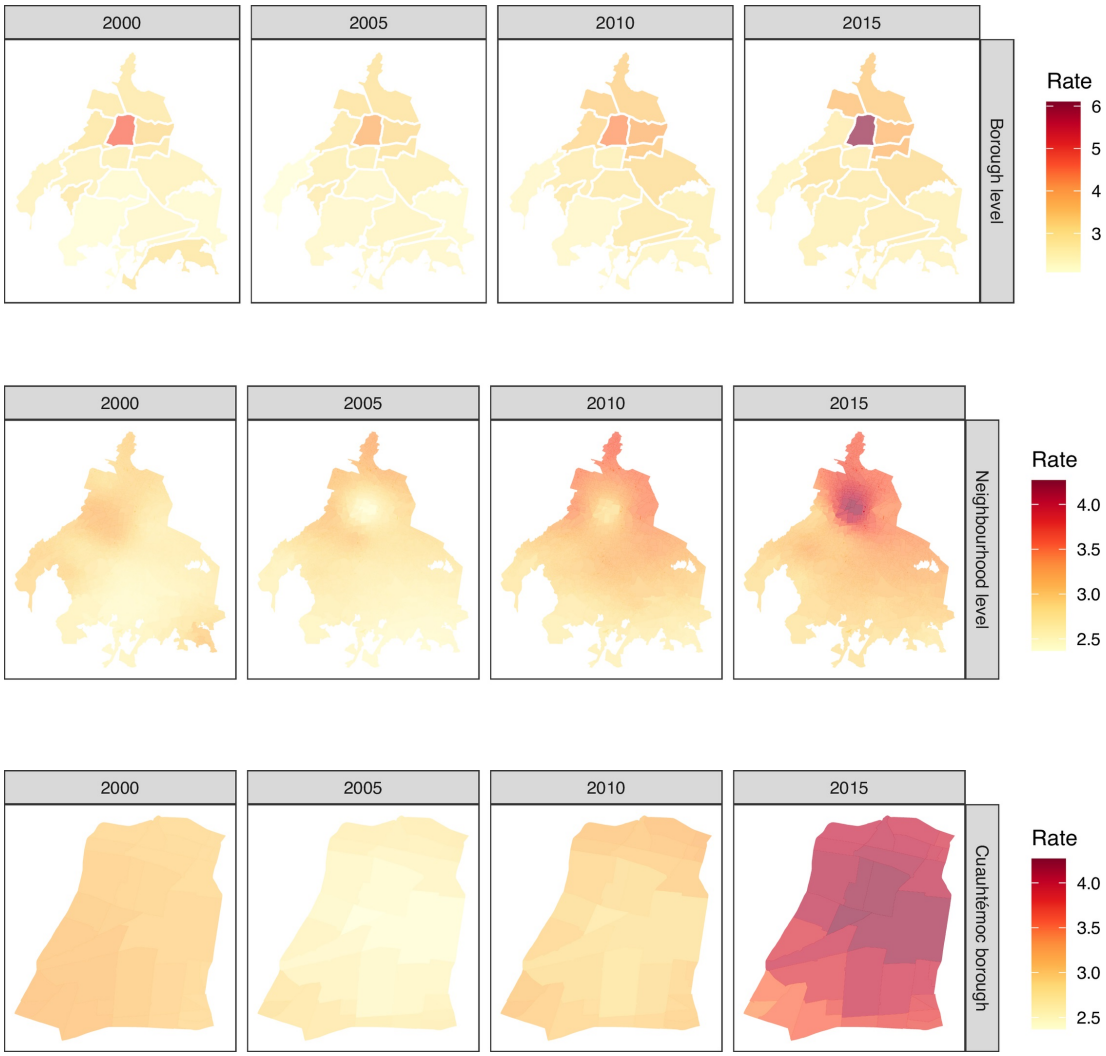


Figura 7.12: Three levels of granularity for school mortality rate in Mexico City. First, school mortality rate original data description is presented at borough level (panel **A**). Second, school mortality rate kriged data description at the neighborhood level is presented at panel **B**. In panel **C**, it is presented a zoom-in at the neighborhood level for the Cuauhtémoc borough. Interestingly, the school mortality rate for this borough increments from the year 2000 to 2005, while it reduces from 2005 to 2015 according to panels A and B, whereas the city's southeast area shows an increment from 2005 to 2015. However, the mortality rate is neither homogeneous nor constant at the neighborhood level, as depicted by the kriged values presented in panel C. Polygon shapefiles files can be freely downloaded at INEGI's website (96).

Figura 7.13: Posterior mortality rate test results at borough level for Mexico City. Fisher's Least Significant Difference (LSD) tests were performed over the estimated age-specific mean mortality rate, according to model description in Eq ((7.1-7.4)). In panels, Mexico City Fisher's LSD test results for: **A**) Age-specific mortality , **B**) Borough level, and, **C**) Borough × Age-specific interaction. The LSD group mean test results are presented as bar-plots with their corresponding standard deviation bar and mean group letters (A, B, C ...), where bars that share at least a single letter, are not statistically different after Bonferroni multiple test correction ($p > 0.05$). Interestingly, mortality rates is mainly composed by age-specific contribution when compared to Borough impact. In addition, Borough × Age-specific interaction retains the mortality age-specific rate pattern, but, it is modulated by the borough contribution.

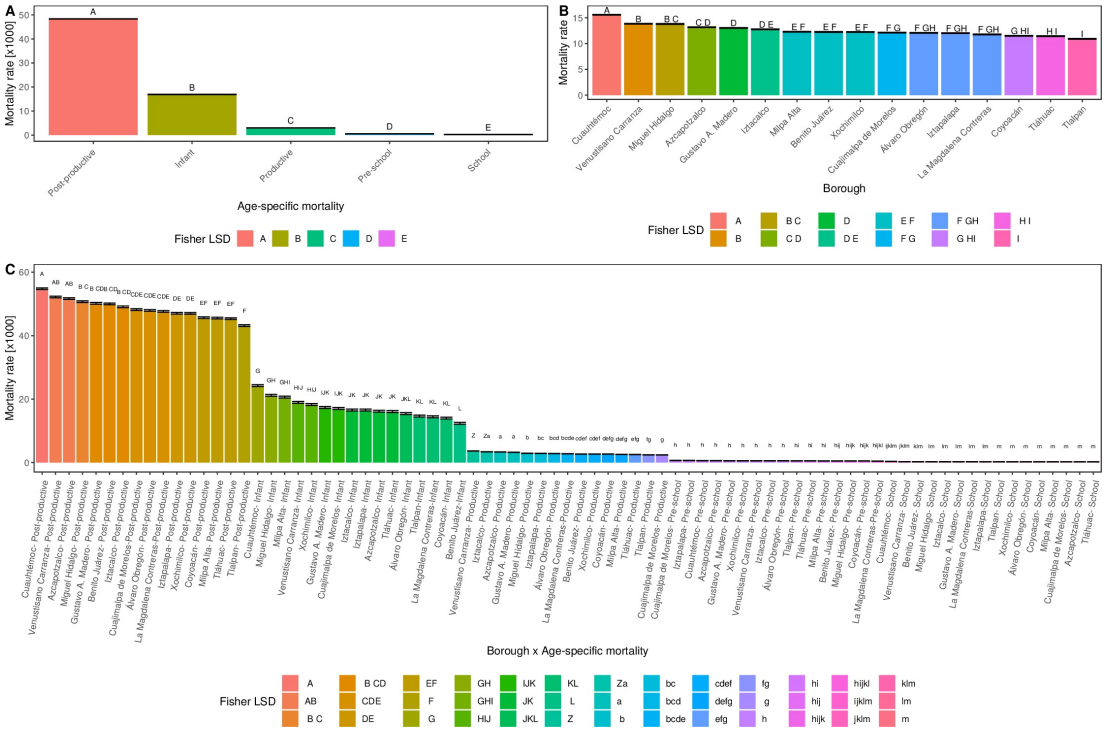
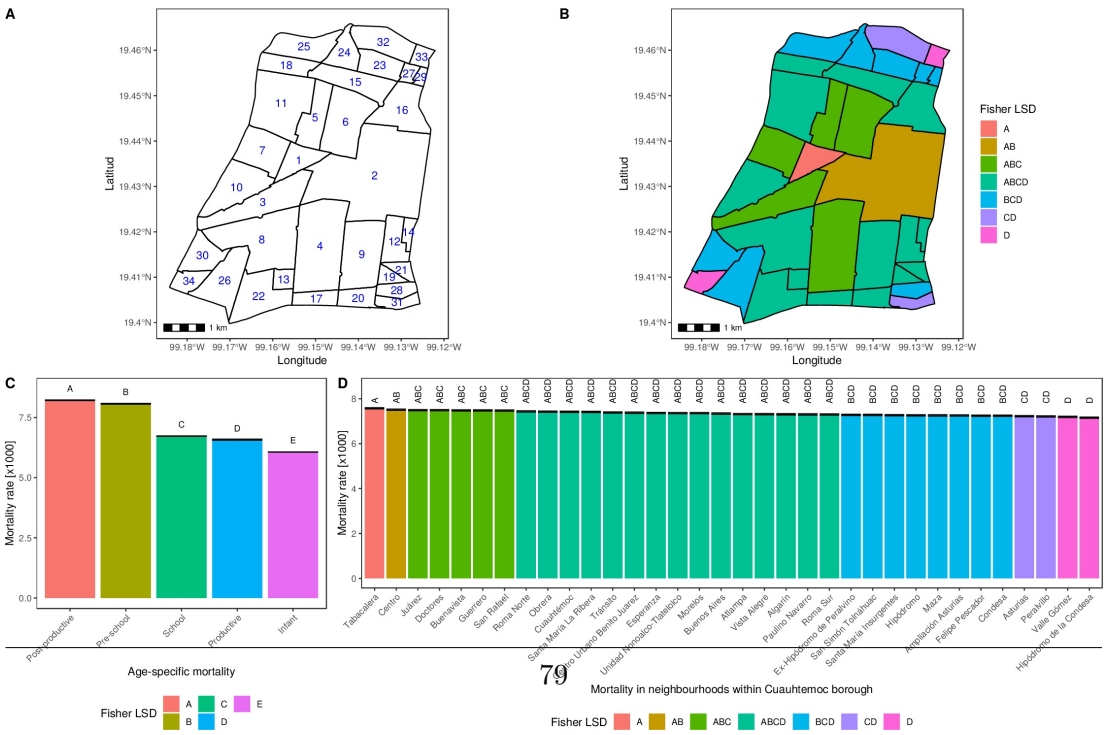


Figura 7.14: Posterior neighborhood mortality test results zoom-in at Cuauhtémoc borough in Mexico City. A) Cuauhtémoc borough is divided into its 34 neighborhoods. Numbers are ordered from the highest (1) to the lowest (34) model estimated neighborhood fixed effect mortality mean according to Eq (7.1-7.4) description: 1. Tabacalera, 2. Centro, 3. Juárez, 4. Doctores, 5. Buenavista, 6. Guerrero, 7. San Rafael, 8. Roma Norte, 9. Obrera, 10. Cuauhtémoc, 11. Santa María la Ribera, 12. Tránsito, 13. Centro Urbano Benito Juárez, 14. Esperanza, 15. Unidad Hab. Nonoalco Tlatelolco, 16. Morelos, 17. Buenos Aires, 18. Atlampa, 19. Vista Alegre, 20. Algarín, 21. Paulino Navarro, 22. Roma Sur, 23. Ex-hipódromo de Peralvillo, 24. San Simón Tolnáhuac, 25. Santa María Insurgentes, 26. Hipódromo, 27. Maza, 28. Ampl. Asturias, 29. Felipe Pescador, 30. Condesa, 31. Asturias, 32. Peralvillo, 33. Valle Gómez and 34. Hipódromo de la Condesa. **B**) Neighborhoods are filled according to the Fisher's Least Significant Difference (LSD) group mean obtained at this level of representation. **C**) Age-specific Fisher's LSD results within Cuauhtémoc borough. **D**) Fisher's LSD neighborhood contribution. In all cases, capital letters stand for Fisher's LSD groups, where bars that share at least a single letter, are not statistically different after Bonferroni multiple test correction ($p > 0.05$). Results are presented as ordered mean $\pm$ standard error estimation according to model description of Eq (7.1-7.4). Notice that at this data level decomposition, the LSD group means are at the same mortality rate order, i.e., the age-specific group mean values are comparable to the different neighborhood contribution. In addition, age-specific mortality values do not overlap whereas, most of the neighborhoods within this specific borough, share a common LSD group.



Capítulo 8

Discusión general

En este trabajo se propuso una metodología robusta para procesar y analizar información de salud y ambiental. En particular, tasas de mortalidad edad-específica (todas las causas) y concentraciones de contaminantes criterio para la Ciudad de México. Esto con la finalidad de crear un conjunto de mapas de mortalidad, también conocidos como *Atlas*, y poder generar hipótesis sobre los factores de riesgo que pudieran explicar las elevadas tasas de mortalidad para los diferentes grupos de edad en determinadas regiones de la Ciudad (i.e. colonias y/o delegaciones). En particular, en la Ciudad de México es de gran interés la morbilidad y mortalidad causada por la mala calidad del aire que es un problema constante día a día. Por esto, en esta investigación se incluyó el análisis de niveles de concentración de los contaminantes CO, NO_2, O_3, SO_2, $PM_{2.5}$, PM_{10}, NO, NO_X y PM_{CO}.

El estudio de concentraciones de contaminantes implica diferentes retos. Primero, el manejo de un gran volumen de datos aún cuando se trate de un sólo contaminante ya que este se mide cada hora y en múltiples estaciones de monitoreo. Así, el reto es mayor al analizar más de un contaminante. Segundo, que existe una gran variedad de métodos e incluso metodologías híbridas para la estimación de concentraciones y algunos de éstos métodos requieren en uso de datos meteorológicos o de uso de suelo que en la mayoría de los casos no están disponibles. Último, pero no menos importante; la base de datos de la red automática de monitoreo no está completa para todos los contaminantes en determinados periodos de tiempo debido a fallas técnicas o mantenimiento. Por esta razón, consideramos imperativo el analizar a profundidad el problema de la disponibilidad de datos de la Red de Monitoreo de la CDMX, así como determinar si la cantidad y distribución de las estaciones de monitoreo es adecuada para medir los diferentes contaminantes de forma consistente en toda la Zona Metropolitana como es su objetivo.

Con el análisis espacio-temporal de las concentraciones de contaminantes se obtuvieron correlaciones de éstos valores entre las 49 estaciones de monitoreo para los nueve contaminantes criterio. Con esto se determinó que tanto la cantidad de estaciones como

su colocación es adecuada para la medición de la mayoría de los contaminantes excepto para el dióxido de azufre (SO_2). El rango espacial (cobertura) de las estaciones para este contaminante es pequenño por lo que la red completa no tiene alcance suficiente, en particular en la zona sur de la Ciudad. Una de las explicaciones del escaso número de estaciones al sur es que el territorio es principalmente rural (menos fuentes de contaminación en la zona). Lamentablemente, una de las consecuencias de no contar con la apropiada cobertura incluso en éstas zonas es que no es posible medir y por ende, prevenir, cuando los contaminantes se dispersan y pasan por esta zona sin cobertura. Este resultado sugiere habilitar más estaciones en la zona sur para medir contaminantes que lleguen por dispersión a la zona o por causas ajenas a la industrialización como por ejemplo los incendios.

Es importante mencionar que la magnitud de los rangos espaciales y temporales, así como la representatividad de la red de monitoreo en su totalidad para cada contaminante calculados en este trabajo, es específico para la Ciudad de México. Lo anterior dadas las condiciones físico-químicas del enternos, así como la tecnología utilizada en cada estación para las mediciones. Por tanto, estos resultados no deben ser comparados con las magnitudes y representatividad de redes de monitoreo en otras regiones o otros países. Sin embargo, sí es de interés de las organizaciones internacionales, el estandarizar los dispositivos utilizados para medición así como los métodos para evaluar la representatividad, exposición y riesgo de salud de la población ocasionado por la mala calidad del aire.

Además de la contaminación ambiental, la Ciudad de México enfrenta problemáticas económicas y sociales que deterioran la salud de sus habitantes ocasionando una elevada tasa de hospitalizaciones y muerte. Lamentablemente, los programas sociales para atender dichas problemáticas no inciden en las personas o comunidades que lo necesitan. Esto se debe principalmente a que existe una amplia heterogeneidad en las condiciones económicas, de seguridad, salubridad, y de acceso a servicios entre cada comunidad (delegación, e incluso entre colonias). Por esto, es importante que la evaluación de las tasas de morbilidad y mortalidad se lleve a cabo a niveles granulares más finos comparado con las escalas a las que actualmente se recolectan datos a través de censos, encuestas o proporcionados por instituciones. En particular, en este trabajo, se implementó una metodología para interpolar valores de tasas de mortalidad edad-específica disponibles a nivel delegación, para obtener estimaciones de dichas tasas a nivel colonia. Y de este manera, vincular los valores estimados con factores determinantes de salud que son conocidos para las diferentes colonias y también delegaciones.

Con las tasas de mortalidad a nivel delegación y los valores interpolados a nivel colonia, fue posible generar mapas de mortalidad conocidos como Atlas para visualizar a dos niveles de granularidad, la distribución de las tasas de mortalidad en la Ciudad de México. Además se analizaron los patrones temporales de mortalidad a nivel trimestral y anual para el periodo 2000-2016. Dentro de los principales resultados se encontró que

la delegación Cuauhtémoc concentra un mayor porcentaje de muertes y esta tendencia es la misma durante el periodo de estudio. Además, a nivel colonia es posible observar el mismo patrón. Ésta delegación tiene una variedad de condicionantes de riesgo de salud y mortalidad que van desde las elevados índices de crimen, elevada densidad poblacional, alta movilidad vehicular (público y privado); lo que dificulta indagar en las causas de muertes mientras no existan datos colectados o bien que sean disponibles para su análisis.

Respecto de los grupos de edad, y en particular en la delegación Cuauhtémoc, la mayor tasa de mortalidad para el periodo de estudio se da en el grupo de edad productiva (14 a 64 años de edad). Mientras que la delegación Tlalpan tiene las tasas de mortalidad más bajas para todos los grupos de edad. El resultado anterior, indica que no es adecuado aplicar programas sociales de la misma manera dentro de una ciudad, cuyas delegaciones o municipios son tan heterogéneos. Y de la misma manera, se debe considerar las características intrínsecas de cada colonia dentro de una delegación cuando se trata de urbes tan complejas como lo es la Ciudad de México.

Bibliografía

[1] Determinants of Health; 2022. Accessed: 2022-04-09. https://www.goinvo.com/vision/determinants-of-health/. 1

[2] Dhimal M, Chirico F, Bista B, Sharma S, Chalise B, Dhimal ML, et al. Impact of air pollution on global burden of disease in 2019. Processes. 2021;9(10):1719. 2

[3] Cohen AJ, Brauer M, Burnett R, Anderson HR, Frostad J, Estep K, et al. Estimates and 25-year trends of the global burden of disease attributable to ambient air pollution: an analysis of data from the Global Burden of Diseases Study 2015. The Lancet. 2017;389(10082):1907–1918. 3, 34

[4] Landrigan PJ, Fuller R, Acosta NJ, Adeyi O, Arnold R, Baldé AB, et al. The Lancet Commission on pollution and health. The lancet. 2018;391(10119):462–512. 3, 4, 5

[5] Thurston GD, Burnett RT, Turner MC, Shi Y, Krewski D, Lall R, et al. Ischemic heart disease mortality and long-term exposure to source-related components of US fine particle air pollution. Environ Health Perspect. 2015;124(6):785–794. 5, 34

[6] Franklin BA, Brook R, Pope III CA. Air pollution and cardiovascular disease. Curr Probl Cardiol. 2015;40(5):207–238. 6, 34

[7] Yang Y, Ruan Z, Wang X, Yang Y, Mason TG, Lin H, et al. Short-term and long-term exposures to fine particulate matter constituents and health: a systematic review and meta-analysis. Environmental pollution. 2019;247:874–882. 6

[8] Bhunia GS, Shit PK. Spatial Statistics and Public Health Events. In: Geospatial Analysis of Public Health. Springer; 2019. p. 99–138. 6

[9] Choi G, Bell ML, Lee JT. A study on modeling nitrogen dioxide concentrations using land-use regression and conventionally used exposure assessment methods. Environ Res Lett. 2017;12(4):044003. 7, 36

[10] Hinojosa-Baliño I, Infante-Vázquez O, Vallejo M. Distribution of PM2. 5 air pollution in Mexico City: Spatial analysis with land-use regression model. Applied sciences. 2019;9(14):2936. 8

[11] Goovaerts P. Geostatistical analysis of disease data: estimation of cancer mortality risk from empirical frequencies using Poisson kriging. Int J Health Geogr. 2005;4(1):31. 8, 62

[12] Tadesse S, Enqueselassie F, Gebreyesus SH. Estimating the spatial risk of tuberculosis distribution in Gurage zone, southern Ethiopia: a geostatistical kriging approach. BMC Public Health. 2018;18(1):783. 8, 62

[13] Piersanti A, Vitali L, Righini G, Cremona G, Ciancarella L. Spatial representativeness of air quality monitoring stations: a grid model based approach. Atmospheric Pollut Res. 2015;6(6):953–960. 8, 36

[14] Martín F, Santiago JL, Kracht O, García L, Gerboles M. FAIRMODE Spatial representativeness feasibility study. Publications Office of the European Union Luxembourg, Luxembourg. 2015;. 9, 36

[15] Janssen S, Dumont G, Fierens F, Deutsch F, Maiheu B, Celis D, et al. Land use to characterize spatial representativeness of air quality monitoring stations and its relevance for model validation. Atmospheric Environ. 2012;59:492–500. 9, 35

[16] Li HZ, Gu P, Ye Q, Zimmerman N, Robinson ES, Subramanian R, et al. Spatially dense air pollutant sampling: Implications of spatial variability on the representativeness of stationary air pollutant monitors. Atmospheric Environ: X. 2019;2:100012. 9, 35

[17] Vitali L, Morabito A, Adani M, Assennato G, Ciancarella L, Cremona G, et al. A Lagrangian modelling approach to assess the representativeness area of an industrial air quality monitoring station. Atmos Pollut Res. 2016;7(6):990–1003. 9, 36

[18] Duyzer J, van den Hout D, Zandveld P, van Ratingen S. Representativeness of air quality monitoring networks. Atmospheric Environ. 2015;104:88–101. 9, 36

[19] Riojas-Rodríguez H, Álamo-Hernández U, Texcalac-Sangrador JL, Romieu I. Health impact assessment of decreases in PM10 and ozone concentrations in the Mexico City Metropolitan Area. A basis for a new air quality management program. Salud Pública de México. 2014;56(6):579–591. 10, 54

[20] Health Atlas as a Policy Tool: How to investigate geographic variation and utilize the information for decision-making; 2022. Accessed: 2022-05-12. https://documents1.worldbank.org/curated/en/539331530275649169/pdf/Colombia-Health-Atlas-Methodology.pdf. 11

[21] The Environment and Health Atlas; 2022. Accessed: 2022-05-12. https://www.envhealthatlas.co.uk/homepage/. 11, 12

[22] Lelieveld J, Evans JS, Fnais M, Giannadaki D, Pozzer A. The contribution of outdoor air pollution sources to premature mortality on a global scale. Nature. 2015;525(7569):367. 34

[23] Pope III CA, Muhlestein JB, Anderson JL, Cannon JB, Hales NM, Meredith KG, et al. Short-term exposure to fine particulate matter air pollution is preferentially associated with the risk of ST-segment elevation acute coronary events. J Am Heart Assoc. 2015;4(12):e002506. 34

[24] Programa para mejorar la calidad del aire en la Zona Metropolitana del Valle de México 2011-2020; 2019. Accessed: 2019-03-01. http://proaire.cdmx.gob.mx/PROAIRE/pagina.php. 34

[25] Bel G, Holst M. Evaluation of the impact of Bus Rapid Transit on air pollution in Mexico City. Transport Policy. 2018;63:209–220. 34

[26] Davis LW. Saturday driving restrictions fail to improve air quality in Mexico city. Sci Rep. 2017;7:41652. 34

[27] Amador-Muñoz O, Bazán-Torija S, Villa-Ferreira S, Villalobos-Pietrini R, Bravo-Cabrera JL, Munive-Colín Z, et al. Opposing seasonal trends for polycyclic aromatic hydrocarbons and PM10: health risk and sources in southwest Mexico City. Atmospheric Res. 2013;122:199–212. 34

[28] Azmi SZ, Latif MT, Ismail AS, Juneng L, Jemain AA. Trend and status of air quality at three different monitoring stations in the Klang Valley, Malaysia. Air Qual Atmos Health. 2010;3(1):53–64. 35

[29] Gasselin P, Morrison K, Lapointe S, Valcke M. Beyond risk assessment and environmental epidemiology. In: Galvao LAC, Finkelman J, Henao S, editors. Environmental and Social Determinants of Health. vol. 1. 1st ed. Washington Dc, USA: Panamerican Health Organization and World Health Organization; 2016. p. 89–122. 35, 51

[30] Secretaría del Medio Ambiente de la Ciudad de México; 2019. Accessed: 2019-03-01. https://sedema.cdmx.gob.mx/. 35

[31] Holgate ST, Koren HS, Samet JM, Maynard RL. Air pollution and health. Elsevier; 1999. 35

[32] Fann N, Risley D. The public health context for PM 2.5 and ozone air quality trends. Air Quality, Atmosphere & Health. 2013;6(1):1–11. 35

[33] Frumkin H. Urban sprawl and public health. Public health reports. 2016;. 36

[34] Beauchamp M, Malherbe L, De Fouquet C, Létinois L. A necessary distinction between spatial representativeness of an air quality monitoring station and the delimitation of exceedance areas. Environ Monit Assess. 2018;190(7):441. 36

[35] Rivas E, Santiago JL, Lechón Y, Martín F, Ariño A, Pons JJ, et al. CFD modelling of air quality in Pamplona City (Spain): Assessment, stations spatial representativeness and health impacts valuation. Sci Total Environ. 2019;649:1362–1380. 36

[36] Miglietta MM, Thunis P, Georgieva E, Pederzoli A, Bessagnet B, Terrenoire E, et al. Evaluation of WRF model performance in different European regions with the DELTA–FAIRMODE evaluation tool. International Journal of Environment and Pollution. 2012;50(1):83. 36

[37] Aguirre-Salado A, Vaquera-Huerta H, Aguirre-Salado C, Reyes-Mora S, Olvera-Cervantes A, Lancho-Romero G, et al. Developing a hierarchical model for the spatial analysis of PM10 pollution extremes in the Mexico City metropolitan area. Int J Environ Res Public Health. 2017;14(7):734. 36

[38] Berman J, Jin L, Bell M, Curriero FC. Developing a geostatistical simulation method to inform the quantity and placement of new monitors for a follow-up air sampling campaign. J Exposure Sci Environ Epidemiol. 2019;29(2):248. 36

[39] Instituto Nacional de Estadística y Geografía; 2019. Accessed: 2019-02-15. https://www.inegi.org.mx/. 37, 60, 61

[40] Calidad del aire - CDMX; 2019. Accessed: 2019-02-01. www.aire.cdmx.gob.mx. 23, 37

[41] Valle-Jones D. aire.zmvm: Download Mexico City Pollution, Wind, and Temperature Data; 2019. R package version 0.8.2. Available from: https://CRAN.R-project.org/package=aire.zmvm. 23, 37

[42] Pebesma EJ. Multivariable geostatistics in S: the gstat package. Computers & Geosciences. 2004;30:683–691. 38, 62, 63

[43] Gräler B, Pebesma E, Heuvelink G. Spatio-Temporal Interpolation using gstat. The R Journal. 2016;8:204–218. Available from: https://journal.r-project.org/archive/2016/RJ-2016-014/index.html. 38, 62, 63

[44] Baca-Lopez K, Fresno C, Espinal-Enriquez J, Flores-Merino MV, Camacho-Lopez MA, Hernandez-Lemus E. Metropolitan age-specific mortality trends at borough and neighbourhood level: The case of Mexico City. PLoS ONE, Submitted. 2020;. 38

[45] Secretaría de Salud. Sobre el impacto en la calidad del aire por los incendios en el Valle de México; 2019. Accessed: 2019-05-16.

https://www.salud.cdmx.gob.mx/comunicacion/nota/12052019-sobre-el-impacto-en-la-calidad-del-aire-por-los-incendios-en-el-valle-de-mexico. 50

[46] Cole BL, Fielding JE. Health impact assessment: a tool to help policy makers understand health beyond health care. Annu Rev Public Health. 2007;28:393–412. 51

[47] Mindell J, Boltong A, Forde I. A review of health impact assessment frameworks. Public health. 2008;122(11):1177–1187. 51

[48] Jäger J, Arreola ME, Chenje M, Pintér L, AIT PR, Abido MS. The GEO Approach to Integrated Environmental Assessment. 2009;. 51

[49] Cifuentes LA, Krupnick AJ, O'Ryan R, Toman MA, et al. Urban air quality and human health in Latin America and the Caribbean. Inter American Development Bank. 2005;. 51

[50] Organization WH, et al. WHO Air quality guidelines for particulate matter, ozone, nitrogen dioxide and sulfur dioxide: global update 2005: summary of risk assessment. Geneva: World Health Organization; 2006. 51

[51] De Sherbinin A, Schiller A, Pulsipher A. The vulnerability of global cities to climate hazards. Environment and urbanization. 2007;19(1):39–64. 51

[52] Jimenez de la Jara J, Torres-Hidalgo M, Salcedo-Hansen R. Cities and determinants of health. In: Galvao LAC, Finkelman J, Henao S, editors. Environmental and Social Determinants of Health. vol. 1. 1st ed. Washington Dc, USA: Panamerican Health Organization and World Health Organization; 2016. p. 229–248. 51

[53] Romieu I, Alamo-Hernandez U, Texcalac-Sangrador JL, Perez L, Gouveia N. Air pollution in the Americas: Impact and policies. In: Galvao LAC, Finkelman J, Henao S, editors. Environmental and Social Determinants of Health. vol. 1. 1st ed. Washington Dc, USA: Panamerican Health Organization and World Health Organization; 2016. p. 541–562. 51, 52, 53

[54] para América Latina y el Caribe) CCE. Anuario Estadístico de América Latina y el Caribe 2012. Comisión Económica para América Latina y el Caribe; 2012. 52

[55] Bell ML, Davis DL, Gouveia N, Borja-Aburto VH, Cifuentes LA. The avoidable health effects of air pollution in three Latin American cities: Santiago, Sao Paulo, and Mexico City. Environmental research. 2006;100(3):431–440. 53

[56] Cifuentes L, Borja-Aburto VH, Gouveia N, Thurston G, Davis DL. Assessing the health benefits of urban air pollution reductions associated with climate change mitigation (2000-2020): Santiago, São Paulo, México City, and New York City. Environmental Health Perspectives. 2001;109(suppl 3):419–425. 53

[57] Jerrett M, Arain A, Kanaroglou P, Beckerman B, Potoglou D, Sahsuvaroglu T, et al. A review and evaluation of intraurban air pollution exposure models. Journal of Exposure Science & Environmental Epidemiology. 2005;15(2):185–204. 54

[58] Brunekreef B, Holgate ST. Air pollution and health. The Lancet. 2002;360(9341):1233–1242. 54

[59] Gaskin DJ, Roberts ET, Chan KS, McCleary R, Buttorff C, Delarmente BA. No man is an island: the impact of neighborhood disadvantage on mortality. Int J Environ Res Public Health. 2019;16(7):1265. 57, 58

[60] Gavurova B, Toth P. Preventable Mortality in Regions of Slovakia—Quantification of Regional Disparities and Investigation of the Impact of Environmental Factors. Int J Environ Res Public Health. 2019;16(8). 57, 58

[61] Rodríguez-Sanz M, Gotsens M, dell?Olmo MM, Borrell C. Trends in mortality inequalities in an urban area: the influence of immigration. Int J Equity Health. 2019;18(1):37. 57

[62] Ayele DG, Zewotir TT. Childhood mortality spatial distribution in Ethiopia. J Appl Stat. 2016;43(15):2813–2828. 57, 58

[63] Burke M, Heft-Neal S, Bendavid E. Sources of variation in under-5 mortality across sub-Saharan Africa: a spatial analysis. Lancet Public Health. 2016;4(12):e936–e945. 57, 62

[64] Tlou B, Sartorius B, Tanser F. Space-time patterns in maternal and mother mortality in a rural South African population with high HIV prevalence (2000–2014): results from a population-based cohort. BMC Public Health. 2017;17(1):543. 57, 58, 62

[65] Montez JK, Zajacova A, Hayward MD. Explaining inequalities in women's mortality between US States. SSM Popul Health. 2016;2:561–571. 58

[66] Shin J, Cho K, Choi Y, Lee S, Park EC, Jang SI. Combined effect of individual and neighborhood socioeconomic status on mortality in patients with newly diagnosed dyslipidemia: A nationwide Korean cohort study from 2002 to 2013. Nutr Metab Cardiovasc Dis. 2016;26(3):207–215. 58

[67] Kindig DA, Cheng ER. Even as mortality fell in most US counties, female mortality nonetheless rose in 42.8 percent of counties from 1992 to 2006. Health Aff. 2013;32(3):451–458. 58, 61

[68] Hazlehurst M, Nurius P, Hajat A. Individual and Neighborhood Stressors, Air Pollution and Cardiovascular Disease. Int J Environ Res Public Health. 2018;15(3):472. 58

[69] Casas I, Delmelle E, Delmelle EC. Potential versus revealed access to care during a dengue fever outbreak. Journal of Transport & Health. 2017;4:18–29. 58

[70] Vaughan AS, Kramer MR, Waller LA, Schieb LJ, Greer S, Casper M. Comparing methods of measuring geographic patterns in temporal trends: an application to county-level heart disease mortality in the United States, 1973 to 2010. Ann Epidemiol. 2015;25(5):329–335. 58, 62

[71] Bethea TN, Palmer JR, Rosenberg L, Cozier YC. Neighborhood socioeconomic status in relation to all-cause, cancer, and cardiovascular mortality in the Black Women's Health Study. Ethn Dis. 2016;26(2):157. 58

[72] Dwyer-Lindgren L, Stubbs RW, Bertozzi-Villa A, Morozoff C, Callender C, Finegold SB, et al. Variation in life expectancy and mortality by cause among neighbourhoods in King County, WA, USA, 1990–2014: a census tract-level analysis for the Global Burden of Disease Study 2015. Lancet Public Health. 2017;2(9):e400–e410. 58

[73] Manda SO, Abdelatif N. Smoothed temporal atlases of age-gender all-cause mortality in South Africa. Int J Environ Res Public Health. 2017;14(9):1072. 58

[74] Vaughan AS, Schieb L, Quick H, Kramer MR, Casper M. Before the here and now: What we can learn from variation in spatiotemporal patterns of changing heart disease mortality by age group, time period, and birth cohort. Soc Sci Med. 2018;217:97–105. 58

[75] Parise CA, Caggiano V. Regional variation in disparities in breast cancer specific mortality due to race/ethnicity, socioeconomic status, and urbanization. J Racial Ethn Health Disparities. 2017;4(4):706–717. 58

[76] Dwyer-Lindgren L, Bertozzi-Villa A, Stubbs RW, Morozoff C, Kutz MJ, Huynh C, et al. US county-level trends in mortality rates for major causes of death, 1980-2014. JAMA. 2016;316(22):2385–2401. 58

[77] Cho KH, Lee SG, Nam CM, Lee EJ, Jang SY, Lee SH, et al. Disparities in socioeconomic status and neighborhood characteristics affect all-cause mortality in patients with newly diagnosed hypertension in Korea: a nationwide cohort study, 2002–2013. Int J Equity Health. 2016;15(1):3. 58, 61

[78] Cho KH, Nam CM, Lee EJ, Choi Y, Yoo KB, Lee SH, et al. Effects of individual and neighborhood socioeconomic status on the risk of all-cause mortality in chronic obstructive pulmonary disease: a nationwide population-based cohort study, 2002–2013. Respir Med. 2016;114:9–17. 58, 61

[79] Li C, Hu S, Yu C. All-Cause and Cancer Mortality Trends in Macheng, China (1984–2013): An Age-Period-Cohort Analysis. Int J Environ Res Public Health. 2018;15(10):2068. 58, 61

[80] Ford MM, Desai PS, Maduro G, Laraque F. Neighborhood inequalities in hepatitis c mortality: spatial and temporal patterns and associated factors. J Urban Health. 2017;94(5):746–755. 58

[81] Sifuna P, Otieno L, Andagalu B, Oyieko J, Ogutu B, Singoei V, et al. A Spatio-temporal Analysis of HIV-Associated Mortality in Rural Western Kenya 2011–2015. J Acquir Immune Defic Syndr. 2018;78(5):483. 58

[82] Roth GA, Abate D, Abate KH, Abay SM, Abbafati C, Abbasi N, et al. Global, regional, and national age-sex-specific mortality for 282 causes of death in 195 countries and territories, 1980–2017: a systematic analysis for the Global Burden of Disease Study 2017. The Lancet. 2018;392(10159):1736–1788. 58

[83] Rawla P, Sunkara T, Gaduputi V. Epidemiology of Pancreatic Cancer: Global Trends, Etiology and Risk Factors. World J Oncol. 2019;10(1):10. 58

[84] Dwyer-Lindgren L, Bertozzi-Villa A, Stubbs RW, Morozoff C, Mackenbach JP, van Lenthe FJ, et al. Inequalities in life expectancy among US counties, 1980 to 2014: temporal trends and key drivers. JAMA Intern Med. 2017;177(7):1003–1011. 58

[85] Ramos GCD, et al. Real Estate Industry as an Urban Growth Machine: A Review of the Political Economy and Political Ecology of Urban Space Production in Mexico City. Sustainability. 2019;11(7):1–24. 59

[86] Cohen AJ, Brauer M, Burnett R, Anderson HR, Frostad J, Estep K, et al. Estimates and 25-year trends of the global burden of disease attributable to ambient air pollution: an analysis of data from the Global Burden of Diseases Study 2015. The Lancet. 2017;389(10082):1907–1918. 59

[87] Lelieveld J, Evans JS, Fnais M, Giannadaki D, Pozzer A. The contribution of outdoor air pollution sources to premature mortality on a global scale. Nature. 2015;525(7569):367–371. 59

[88] Bel G, Holst M. Evaluation of the impact of bus rapid transit on air pollution in Mexico City. Transport Policy. 2018;63:209–220. 59

[89] Davis LW. Saturday driving restrictions fail to improve air quality in Mexico City. Scientific reports. 2017;7:41652. 59

[90] Aburto JM, Riffe T, Canudas-Romo V. Trends in avoidable mortality over the life course in Mexico, 1990–2015: a cross-sectional demographic analysis. BMJ Open. 2018;8(7):e022350. 59

[91] Aburto JM, Beltrán-Sánchez H, García-Guerrero VM, Canudas-Romo V. Homicides in Mexico reversed life expectancy gains for men and slowed them for women, 2000–10. Health Affairs. 2016;35(1):88–95. 59

[92] Aburto JM, Beltrán-Sánchez H. Upsurge of homicides and its impact on life expectancy and life span inequality in Mexico, 2005–2015. American journal of public health. 2019;109(3):483–489. 59

[93] Gómez-Dantés H, Fullman N, Lamadrid-Figueroa H, Cahuana-Hurtado L, Darney B, Avila-Burgos L, et al. Dissonant health transition in the states of Mexico, 1990–2013: a systematic analysis for the Global Burden of Disease Study 2013. The Lancet. 2016;388(10058):2386–2402. 59

[94] Benita F. Social backwardness in Mexico City metropolitan area. Soc Indic Res. 2016;126(1):141–160. 59

[95] Secretaría de Salud de la Ciudad de México. Secretaría de Salud de la Ciudad de México; 2019. Accessed: 2019-02-01. http://data.salud.cdmx.gob.mx/portal/. 24, 60

[96] Instituto Nacional de Estadística y Geografía. Instituto Nacional de Estadística y Geografía: Marco Geoestadístico, diciembre 2018; 2019. Accessed: 2019-02-15. http://en.www.inegi.org.mx/app/biblioteca/ficha.html?upc=889463674658. 60, 74, 76, 77

[97] Bivand R, Rundel C. rgeos: Interface to Geometry Engine - Open Source ('GEOS'); 2018. R package version 0.4-2. Available from: https://CRAN.R-project.org/package=rgeos. 24, 60

[98] Pebesma E. Simple Features for R: Standardized Support for Spatial Vector Data. The R Journal. 2018;10(1):439–446. Available from: https://doi.org/10.32614/RJ-2018-009. 60

[99] Hijmans RJ. raster: Geographic Data Analysis and Modeling; 2019. R package version 2.9-23. Available from: https://CRAN.R-project.org/package=raster. 60

[100] Hijmans RJ. geosphere: Spherical Trigonometry; 2017. R package version 1.5-7. Available from: https://CRAN.R-project.org/package=geosphere. 60

[101] Pebesma E. spacetime: Spatio-Temporal Data in R. Journal of Statistical Software. 2012;51(7):1–30. Available from: http://www.jstatsoft.org/v51/i07/. 60

[102] Pebesma EJ, Bivand RS. Classes and methods for spatial data in R. R News. 2005 November;5(2):9–13. Available from: https://CRAN.R-project.org/doc/Rnews/. 60

[103] Bivand R, Keitt T, Rowlingson B. rgdal: Bindings for the 'Geospatial' Data Abstraction Library; 2019. R package version 1.4-3. Available from: https://CRAN.R-project.org/package=rgdal. 60

[104] Wickham H. ggplot2: Elegant Graphics for Data Analysis. Springer-Verlag New York; 2016. Available from: http://ggplot2.org. 60

[105] Wilke CO. cowplot: Streamlined Plot Theme and Plot Annotations for 'ggplot2'; 2019. R package version 1.0.0. Available from: https://CRAN.R-project.org/package=cowplot. 60

[106] Auguie B. gridExtra: Miscellaneous Functions for "Grid"Graphics; 2017. R package version 2.3. Available from: https://CRAN.R-project.org/package=gridExtra. 60

[107] Dunnington D. ggspatial: Spatial Data Framework for ggplot2; 2018. R package version 1.0.3. Available from: https://CRAN.R-project.org/package=ggspatial. 60

[108] Secretaría de Salud de la Ciudad de México. Secretaría de Salud de la Ciudad de México: Mortalidad; 2019. Accessed: 2019-02-01. http://data.salud.cdmx.gob.mx/portal/index.php/informacion-en-salud/103-informacion-salud/354-mortalidad. 60, 61, 74, 75

[109] Gouveia N, Junger WL, Romieu I, Cifuentes LA, de Leon AP, Vera J, et al. Effects of air pollution on infant and children respiratory mortality in four large Latin-American cities. Environ Pollut. 2018;232:385–391. 61

[110] Salto-Quintana JN, Rivera-Alfaro G, Sánchez-Ramos EL, Gómez-Gómez A, Noyola DE. Post-pandemic influenza-associated mortality in Mexico. Pathog Glob Health. 2019;p. 1–8. 61

[111] Wang N, Mengersen K, Tong S, Kimlin MG, Zhou M, Wang L, et al. Lung Cancer Mortality in China: Spatial and Temporal Trends Among Subpopulations. Available at SSRN 3315845. 2019;. 61

[112] Wändell P, Carlsson AC, Gasevic D, Sundquist J, Sundquist K. Neighbourhood socio-economic status and all-cause mortality in adults with atrial fibrillation: a cohort study of patients treated in primary care in Sweden. Int J Cardiol. 2016;202:776–781. 61

[113] Yang W, Olson DR, Shaman J. Forecasting influenza outbreaks in boroughs and neighborhoods of New York City. PLOS Comput Biol. 2016;12(11):e1005201. 61

[114] Ryvicker M, Sridharan S. Neighborhood environment and disparities in health care access among urban Medicare beneficiaries with diabetes: A retrospective cohort study. NQUIRY J Health Car. 2018;55:0046958018771414. 61

[115] Lin WC, Lin YP, Wang YC, Chang TK, Chiang LC. Assessing and mapping spatial associations among oral cancer mortality rates, concentrations of heavy metals in soil, and land use types based on multiple scale data. Int J Environ Res Public Health. 2014;11(2):2148–2168. 62

[116] Chien LC, Yu HL, Schootman M. Efficient mapping and geographic disparities in breast cancer mortality at the county-level by race and age in the US. Spatial Spatio-temporal Epidemiol. 2013;5:27–37. 62

[117] Oliver M, Webster R. A tutorial guide to geostatistics: Computing and modelling variograms and kriging. Catena. 2014;113:56–69. 62

[118] Raudenbush S, Bryk AS. A hierarchical model for studying school effects. Sociology of education. 1986;p. 1–17. 63

[119] Kwok OM, Underhill AT, Berry JW, Luo W, Elliott TR, Yoon M. Analyzing longitudinal data with multilevel models: an example with individuals living with lower extremity intra-articular fractures. Rehabilitation psychology. 2008;53(3):370. 63

[120] Anaya G, Al-Delaimy WK. Effect of the US-Mexico border region in cardiovascular mortality: ecological time trend analysis of Mexican border and non-border municipalities from 1998 to 2012. BMC public health. 2017;17(1):400. 64

[121] Green EP, Cho H, Gallis J, Puffer ES. The impact of school support on depression among adolescent orphans: a cluster-randomized trial in Kenya. Journal of Child Psychology and Psychiatry. 2019;60(1):54–62. 64

[122] R Core Team. R: A Language and Environment for Statistical Computing. Vienna, Austria; 2018. Available from: https://www.R-project.org/. 64

[123] Pinheiro J, Bates D. Mixed-effects models in S and S-PLUS. Springer Science & Business Media; 2006. 64

[124] Di Rienzo J, Casanoves F, Balzarini M, Gonzalez L, Tablada M, Robledo C. InfoStat. Universidad Nacional de Córdoba; 2011. 28, 64

[125] Pinheiro J, Bates D, DebRoy S, Sarkar D, R Core Team. nlme: Linear and Nonlinear Mixed Effects Models; 2018. R package version 3.1-137. Available from: https://CRAN.R-project.org/package=nlme. 64

[126] Instituto Nacional de Estadística y Geografía. Instituto Nacional de Estadística y Geografía: Población 2015; 2020. Accessed: 2020-09-01. https://www.inegi.org.mx/temas/estructura/. 67

[127] Gobierno de la Ciudad de México. Programa Delegacional de Desarrollo en Cuauhtémoc 2016-2018 ; 2019. Accessed: 2019-04-01. http://www.data.seduvi.cdmx.gob.mx/portal/index.php/programas-de-desarrollo/programas-delegacionales. 71

[128] Espinal-Enríquez J, Larralde H. Analysis of Mexico's narco-war network (2007–2011). PLoS One. 2015;10(5):e0126503. 71

[129] Gobierno de la Ciudad de México. Secretariado Ejecutivo del Sistema Nacional de Seguridad Pública; 2019. Accessed: 2019-04-01. https://www.gob.mx/sesnsp. 71, 72