



INSTITUTO TECNOLÓGICO DE CULIACÁN

Ingeniería en Sistemas Computacionales

Tópicos de Inteligencia Artificial Hora: 12:00 – 1:00 PM

Tema: Aplicación de modelos de redes neuronales para el análisis y predicción de picos de contaminación atmosférica derivados de emisiones de combustión para mejorar la gestión urbana ambiental.

Autores:

Peña López Miguel Ángel Robles Rios Jacquelin

Docente:

Zuriel Dathan Mora Félix

Culiacán, Sinaloa 30/08/2025

Índice

1	Marc	co Conceptual	3
2	Intro	oducción	4
3	Obje	etivo general	4
	-	Objetivos específicos	
4		ificación	
5		nce	
6		arrollo	
	6.1 6.1.1	La contaminación atmosférica	
		Inteligencia artificial y redes neuronales: herramientas innovadoras para la ambiental	7
	6.2.1	Casos de éxito	8
	6.3 6.3.1	Arquitecturas de modelado predictivo	
	6.4 6.4.1 6.4.2	Propuesta de modelo mediante la implementación de redes neuronales Arquitectura propuesta	. 12
7	Agei	nda	12
8	Con	clusión	13
9	Refe	erencias	14

1 Marco Conceptual

1. Contaminantes

- Contaminantes Orgánicos Persistentes (COP)
- Contaminantes Atmosféricos (CA)
- Contaminantes del Agua (CAG)
- Residuos Sólidos y Peligrosos (RSP)
- Monóxido de Carbono (CO)
- Compuestos Orgánicos Volátiles (COV)
- Amoníaco (NH₃)
- Óxido Nítrico (NO)
- Partículas PM₁₀
- Partículas PM₂₋₅
- Dióxido de Azufre (SO₂)
- Ozono (O₃)

2. Inteligencia Artificial y Modelado

- Inteligencia Artificial (IA)
- Machine Learning (ML)
- Red Neuronal Artificial (RNA)
- Red Neuronal Recurrente (RNN)
- Memoria de Largo-Corto Plazo (LSTM)
- Red Neuronal Convolucional (CNN)
- Red Neuronal para Detección de Eventos (NNED)
- Promedio Móvil Integrado Autoregresivo (ARIMA)

3. Métricas de Evaluación

- Error Absoluto Medio (MAE)
- Error Cuadrático Medio (RMSE)
- Coeficiente de Determinación (R²)

4. Sistemas y Servicios de Monitoreo

- Pronóstico Numérico del Tiempo (NWP)
- Predicción Numérica de Contaminación (NPP)
- Servicio de Monitoreo Atmosférico Copernicus (CAMS)
- Red de Monitoreo de Calidad del Aire de Bogotá (RMCAB)
- Sistema Predictivo de Calidad del Aire (SOCAIRE)

5. Indicadores Socioeconómicos

- Producto Interno Bruto (PIB)
- Generación de Residuos Sólidos

2 Introducción

Uno de los múltiples problemas ambientales que tenemos hoy en día es la contaminación atmosférica. Aunque su origen no es reciente, su importancia se ha tomado mayor importancia en años recientes. Actualmente la preocupación acerca de la contaminación atmosférica, sus efectos sobre nuestro planeta y, en consecuencia, sobre todos nosotros como habitantes de este, no ha hecho más que aumentar (Kelly F,J & Fussel. 2015).

Ante este escenario, la inteligencia artificial se presenta como una herramienta innovadora para el estudio y la gestión de este fenómeno. Por lo tanto, este trabajo busca explorar y evaluar el potencial del uso de las redes neuronales para el desarrollo de modelos predictivos para el análisis de la contaminación atmosférica en México.

3 Objetivo general

Explorar el potencial de diseñar un modelo predictivo de contaminación atmosférica en México, que utilice redes neuronales e integre los factores meteorológicos y urbanos más influyentes, con la finalidad de obtener resultados precisos que permita tomar acciones preventivas y correctivas ante la contaminación.

3.1 Objetivos específicos

- Revisar y sintetizar investigaciones recientes sobre el uso de redes neuronales para la predicción de la calidad del aire.
- Identificar las variables más relevantes (meteorológicas, de tráfico, temporales) utilizadas en modelos predictivos de contaminación.
- Explorar el diseño de una arquitectura de red neuronal artificial (RNA) que integre algoritmos de aprendizaje supervisado para predecir niveles de contaminación con gran precisión.
- Analizar las limitaciones y viabilidad de implementación de estos modelos en contextos reales.

4 Justificación

La contaminación atmosférica representa uno de los desafíos ambientales y de salud pública en México es por eso que la implementación de un modelo predictivo de de

concentraciones de contaminantes atmosfericos mediante redes neuronales se justifica por la necesidad de abordar los efectos documentados de la contaminación atmosférica en la salud pública mexicana. Estudios epidemiológicos nacionales han demostrado que "por cada 10 μ g/m³ de incremento en PM₁₀, se observa un aumento de 0.96% en la mortalidad general, 1.82% en mortalidad respiratoria y 1.32% en mortalidad cardiovascular" (Rosales-Castillo et al., 2001, p. 544). Estos datos revelan el gran impacto de estos contaminantes en la población.

México cuenta con la infraestructura de monitoreo adecuada proporcionada por el SINAICA y SEMARNAT que únicamente generan datos puros, lo cual frente a otros países representa una brecha tecnológica. Por ello en esta investigación se explorará la posible propuesta de un modelo basado en redes LSTM específico para el perfil de emisiones mexicano. Esto no solo contribuirá al avance científico en la aplicación de la inteligencia artificial para problemas ambientales locales, sino que apoyará a la toma de decisiones en cuanto a políticas de gestión ambiental para ciudades sostenibles y salud pública.

5 Alcance

Este estudio tendrá un alcance exploratorio. Mediante una revisión de literatura, se analizarán y sintetizarán investigaciones recientes sobre la aplicación de arquitecturas de redes neuronales para la predicción de concentraciones de contaminación atmosférica. Este análisis se centrará en comprender las variables predictoras utilizadas y en evaluar la conveniencia de las arquitecturas disponibles, con el objetivo de fundamentar una propuesta sólida.

6 Desarrollo

6.1 La contaminación atmosférica

Con el pasar de los años la humanidad ha buscado mejorar en varios aspectos de la vida y es así como surge el desarrollo industrial el cual propicio un crecimiento económico gracias a los procesos de producción innovadores, sin embargo esto mismo llevo a una explotación de recursos; comenzando así un deterioro ambiental.

Hoy en día este fenómeno se manifiesta severamente presentándose de varias maneras como lo es la contaminación atmosférica una problemática grande y sumamente preocupante, se presenta principalmente en zonas urbanas, donde, el tráfico, incineración de residuos aunado a las fábricas que no regulan sus emisiones se generan concentraciones de contaminantes que exceden los límites establecidos por organizaciones como la OMS, volviendo que el aire se vuelva inseguro para la salud humana.

Frente a este desafío, naciones como China han implementado políticas y acciones para la regulación de este fenómeno, para ello se han comprometido en gran medida siendo así que han optimizado su infraestructura energética, cerrado calderas de carbón y una reestructuración industrial logrando reducir en gran medida la contaminación atmosférica.

En este sentido, las medidas implementadas como las de China son fundamentales y de gran avance, sin embargo eso no basta, también resulta necesario contar con herramientas que nos permitan estimar la calidad del aire en el futuro para así anticipar riesgos y diseñar mejores planes para su gestión.

6.1.1 Contexto mexicano: Perfil de emisiones y contaminantes prioritarios

México no es ajeno a la problemática de la contaminación atmosférica, en la cual la calidad del aire representa un desafío de salud pública y sostenibilidad ambiental de notable importancia. De acuerdo con la Secretaría de Medio Ambiente y Recursos Naturales (SEMARNAT), las fuentes de emisión de contaminantes atmosféricos en México se clasifican en 5 categorías principales, como se muestra en la Figura 1.

Tipo de fuente	Porcentaje
Fuentes de área	24,37 %
Fuentes fijas	10,28 %
Fuentes móviles carreteros	28,51 %
Fuentes móviles que no circulan por carretera	0,88 %
Fuentes naturales	35,96 %

Figura 1. Fuentes contaminantes en México

Un análisis detallado de estas fuentes revelan que existen tanto fuentes naturales como fuentes provenientes de actividades humanas, en la cuales las actividades humanas son las de mayor impacto en las emisiones contaminantes, destacando el sector de transporte e industrias.

Al analizar la distribución de estos contaminantes según su fuente de origen como se muestra en la Figura 2 podemos observar que el NO_x y el CO son principalmente emitidos en el sector del transporte, mientras que el COV, PM_{10} y $PM_{2\cdot5}$ tienen origen principalmente de procesos industriales y actividades agícolas.

Tipo de fuente	CO	COV	NH_3	NOx	PM 10	PM 2.5	SO2
Fuentes de área	30,10%	18,17%	98,53%	3,74%	63,21%	60,50%	1,17%
Fuentes fijas	3,90%	1,16%	0,35%	15,92%	27,05%	26,31%	96,68%
Fuentes móviles carreteros	65,24%	4,97%	1,12%	44,79%	8,95%	12,04%	1,82%
Fuentes móviles que no circulan por carretera	0,76%	0,13%	0,00%	4,24%	0,79%	1,14%	0,33%
Fuentes naturales		75,57%		31,31%			

Figura 2. Porcentaje de cada contaminante según su fuente

En cuanto al impacto porcentual de cada contaminante en México que se muestra en la figura 3 el COV es el que lidera en emisiones con más del 40%, seguido del mónoxido de carbono con un 30%. Además el NO representa aproximadamente el 12%, mientras que el SO_2 contribuye con un 6%. Estos contaminantes provienen principalmente de fuentes móviles e industriales, lo que confirma los análisis anteriores sobre el perfil contaminante del país.

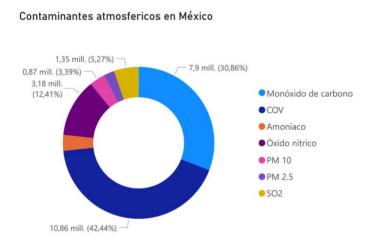


Figura 3. Porcentaje de cada contaminante

El análisis de estos perfiles de emsión dan a conocer la urgencia en la necesidad de implementar sistemas de monitoreo y predicción, para así anticipar episodios de contaminación atmosferica y poder crear politicas que ayuden a la conservación de la calidad del aire en México.

6.2 Inteligencia artificial y redes neuronales: herramientas innovadoras para la gestión ambiental

Ante la variedad de factores que repercuten en la calidad del aire y sus efectos en la salud, las herramientas tradicionales pueden resultar insuficientes es aquí donde el uso de la inteligencia artificial interviene como un enfoque innovador para la implementación de un modelo centrado en la predicción de la calidad del aire, resultando útil a para evaluar las inquietudes presentes y las futuras amenazas de la contaminación atmosférica como la salud humana.

Con un sistema basado en inteligencia artificial se podría llegar a anticipar la probabilidad de concentraciones de contaminación y peligro de estas en una zona urbana. Como señala Francisco Vargas Marcos subdirector General de Sanidad

Ambiental Francisco, V. M. (s. f.). "estar expuesto a varios contaminantes peligrosos puede incrementar el riesgo de enfermar por encima de lo esperado si uno estuviera expuesto a la acción separada de cada uno de estos factores. Por ejemplo, el amianto afecta y multiplica varias veces el riesgo de contraer cáncer en fumadores." Es por ello que es de suma importancia contar con las herramientas adecuadas que ayuden a prevenir concentraciones de estos contaminantes.

Por ende, la aplicación de redes neuronales no es únicamente la predicción numérica, también tiene importancia como una herramienta para la toma de decisiones en políticas ambientales, permitiendo así alertas tempranas y medidas correctivas antes de que causen mayores daños y que arriesguen en mayor medida a la población.

6.2.1 Casos de éxito

6.2.1.1 SOCAIRE: Forecasting and Monitoring Urban Air Quality in Madrid

Este trabajo consistió en el diseño e implementación de un sistema capaz de predecir con 48 horas de antelación los niveles de cuatro contaminantes críticos: NO_2 , O_3 y PM10 y PM2.5.

La implementación se llevó a cabo en 24 estaciones de la red de medición de Madrid, cada una con distintos contaminantes. Esto con el objetivo de anticipar momentos de alta contaminación y así justificar la activación de restricciones de movilidad vehicular y medidas preventivas de salud pública, minimizando su impacto con intervenciones tempranas.

Las variables integradas en el modelo fueron las siguientes:

- Variables endógenas: Series temporales históricas de las concentraciones de los contaminantes objetivo NO₂, O₃, PM10, PM2.5 en las 24 estaciones de monitoreo.
- Variables meteorológicas NWP: Pronósticos numéricos de altura de la capa límite, presión superficial, temperatura, precipitación, y componentes U (esteoeste) y V (norte-sur) del viento a 10m. Estas variables son críticas para modelar la dispersión y transformación química de los contaminantes.
- Variables de contaminación a macroescala NPP: Pronósticos del servicio europeo CAMS, que proporcionan una visión sinóptica de la contaminación que luego es refinada por el modelo local.
- Variables antropogénicas: Es uno de los puntos más fuertes. Incluye datos de calendario como días festivos, operaciones de salida y retorno (puentes, vacaciones) y el calendario escolar. Esto permite modelar con alta precisión los cambios en los patrones de tráfico y emisiones.

 Variable de protocolo: Información sobre la activación previa de las medidas restrictivas del protocolo de NO₂, lo que afecta directamente los niveles de emisión.

En cuanto a los resultados, la precisión reportada es excepcionalmente alta. El modelo demostró una superioridad clara frente a los baselines de comparación. Para NO_2 , reportó un RMSE de 14.9 ± 4.8, superando por amplio margen al modelo CAMS (RMSE: 23.5) y al modelo de Persistencia (RMSE: 26.4), lo que representó una mejora de aproximadamente 36% y 44% respectivamente. Los resultados fueron consistentemente mejores para los otros contaminantes O_3 , PM10, PM2.5.

Además de la precisión en la predicción puntual, su mayor fortaleza es la predicción probabilística. El sistema fue capaz de predecir la distribución completa de probabilidad (quantiles 1% al 99%) para cada contaminante, estación y hora, lo que demostró ser crucial para calcular las probabilidades de activación de protocolos.

6.2.1.2 Predicción de Contaminantes Atmosféricos en Bogotá utilizando Redes LSTM

Esta investigación consistió en el desarrollo e implementación de modelos basados en redes LSTM para predecir los niveles de contaminantes atmosféricos en Bogotá, utilizando datos de la estación meteorológica de Las Ferias, parte de la Red de Monitoreo de Calidad del Aire de Bogotá (RMCAB). Se recopilaron y procesaron datos del período entre 2021 y 2023, sumando más de 29,200 registros, que fueron empleados para entrenar, validar y probar los modelos.

El conjunto de variables consideradas en el modelo incluyó:

- Variables endógenas: Concentraciones históricas de los contaminantes individuales a predecir (PM10, PM2.5, CO, NO₂).
- Variables meteorológicas: Datos de velocidad y dirección del viento (transformados en componentes vectoriales velX y velY para mejor procesamiento), temperatura, humedad, presión barométrica y precipitación.

En cuanto al desempeño predictivo, los resultados mostraron que los modelos LSTM multivariados superaron consistentemente a los modelos univariados.

Las mejoras en el RMSE fueron significativas: 18.22% para PM2.5, 17.24% para CO, 12.7% para PM10 y 7.58% para NO₂. Los valores de RMSE absolutos (e.g., 6.94 para PM10, 5.79 para PM2.5) son difíciles de evaluar sin el contexto de los niveles de concentración típicos en Bogotá, pero la mejora relativa demuestra claramente el valor de incorporar variables meteorológicas.

6.2.1.3 Análisis y Predicción de las Emisiones de CO₂ en Bolivia a través de Redes Neuronales Artificiales

El estudio analizó la evolución y predicción de emisiones de CO_2 en Bolivia entre 2000 y 2024 utilizando modelos de regresión lineal, ARIMA y redes neuronales. La red neuronal mostró el mejor rendimiento R^2 : 0.9391, RMSE: 0.3868, identificando como variables clave la pérdida de cobertura forestal, el consumo de gas licuado y el PIB industrial. Además proyectó que las emisiones seguirán aumentando hasta 2029 sin no se implementan políticas de mitigación efectivas, destacando la necesidad de estrategias sostenibles urgentes.

En este estudio se aborda el problema desde una perspectiva macroeconómica y de uso de suelo, no desde la calidad del aire urbano en tiempo real. Las variables son anuales y agregadas a nivel nacional:

- Variable objetivo: Emisiones anuales de CO₂
- Variables predictoras: Incluye un mix de indicadores ambientales, energéticos y económicos: Pérdida de cobertura forestal por incendios, Consumo de diversos combustibles fósiles (gasolina, diésel, gas licuado - GLP, kerosene, jet fuel), PIB del sector industrial manufacturero, RR.SS, Población urbana, y Ganado bovino faenado.

La precisión del modelo de RNA fue excelente y se puede comparar favorablemente contra modelos clásicos: La RNA obtuvo un R² de 0.9391, lo que indica que el modelo explica el 93.91% de la varianza en las emisiones de CO₂.

Los errores (RMSE de 0.3868, MAE de 0.3294) fueron bajos en comparación con la escala de los datos.

Su performance fue muy superior a la del modelo ARIMA (R²: 0.88) y abrumadoramente mejor que la Regresión Lineal (R²: 0.2529), demostrando su capacidad para capturar relaciones no lineales complejas.

6.2.1.4 Modelo de Predicción de Emisiones por Beatriz Ordóñez Becker

Este proyecto se centró en el desarrollo de un modelo de predicción de la calidad del aire utilizando técnicas de aprendizaje automático. El objetivo principal fue identificar aquellas variables que contribuyen a la acumulación de contaminantes en áreas urbanas y predecir los niveles de calidad del aire en tiempo real basándose en estas variables.

El estudio implicó la recopilación y el análisis de datos sobre contaminantes atmosféricos, particularmente el óxido nítrico (NO), las partículas PM2.5 y PM10, de

varias estaciones de monitoreo en la Comunidad de Madrid. Centrandose en predecir la calidad del aire a nivel de calle, integrando variables de tráfico y meteorología:

- Variables de contaminantes: Mediciones de Óxido Nítrico (NO), PM2.5 y PM10.
- Variables de tráfico: Incluye de forma explícita el "Número de carriles" como un proxy del volumen y tipo de tráfico en la vía, una variable clave que a menudo falta en otros modelos.
- Variables meteorológicas: Velocidad y dirección del viento, temperatura, humedad, presión barométrica, radiación solar y precipitación.
- Variables espaciales y temporales: ID de la estación (ubicación), área que abarca, hora del día y día de la semana (para capturar patrones cíclicos).

El reporte de precisión se enfocó en el coeficiente de determinación R² para una variedad de modelos. Los mejores modelos (Random Forest, XGBoost, y un Ensemble) alcanzaron valores de R² muy elevados: 0.8292 para NO, 0.7547 para PM2.5 y 0.8959 para PM10. Estos valores indican una precisión de buena a excelente para un problema de regresión del mundo real.

Se confirma, una vez más, la pobre performance de los modelos lineales (Regresión Lineal, Lasso) con R² alrededor de 0.17-0.37, resaltando la naturaleza no lineal del problema.

6.3 Arquitecturas de modelado predictivo

La efectividad de un sistema de predicción en este ámbito depende mucho de la capacidad de la arquitectura en la que esta basada, esto debido a que el fenómeno conlleva la interdependencia de múltiple factores. Esta sección se centra principalmente en evaluar la capacidad de un conjunto de arquitecturas disponibles.

6.3.1 Enfoques y sus limitaciones

Existen distintos enfoques para la predición de datos como el *Machine Learning* el cual involucra los árboles de decision y sus variantes, son ampliamente usados por su gran capacidad de manejo de datos y encontrar relaciones en ellos. Sin embargo estos modelos tienen sus limitaciones ya que toman cada momento como muestras independientes, es decir que no encuentran patrones a lo largo del tiempo.

Para superar esta problemática de se han optado por usar diferentes arquitecturas como las redes neuronales recurrentes (RNN) o las arquitecturas como LSTM (Long Short-Term Memory) posicionandose como las soluciones adecuadas gracias a que permiten tener un historial de información relevante para así llegar a una mejor predicción.

6.4 Propuesta de modelo mediante la implementación de redes neuronales

El análisis de las arquitecturas y los casos de éxito revisados en secciones anteriores nos confirma el potencial que tienen los modelos de inteligencia artificial, principalmente las redes neuronales en la predicción de la calidad atmosferica, sin embargo es importante tomar en cuenta que la efectividad de esto dependerá de que se adapte correctamente utilizando las variables adecuadas. Tomando en base lo anterior y considerando el perfil de emisiones mexicano se propone un modelo de redes neuronales para la predicción picos de contaminación atmosférica derivados de emisiones de combustión.

6.4.1 Arquitectura propuesta

Se propone el desarrollo de un sistema de predicción basado en redes LSTM, especialmente para el pronostico de concentraciones de contaminantes atmosféricos. El uno de las LSTM es adecuado para este propósito, debido a su capacidad para aprender dependencias a lo largo del tiempo, permitiendo así encontrar patrones en distintos lapsos de tiempo en los datos.

6.4.2 Datos de entrenamiento

Para el desarrollo del modelo predictivo, se recopilaran datos provenientes de diversas instituciones mexicanas, obteniendo los datos de contaminantes CO, COV, NH_3 , NOx, PM_{10} , $PM_{2\cdot5}$, SO_2 desde la plataforma del Sistema Nacional de Información de la Calidad del Aire (SINAICA) y del Sistema de Monitoreo Atmosférico de la Ciudad de México (SIMAT), los cuales operan bajo regulaciones y coordinación con la Secretaría de Medio Ambiente y Recursos Naturales (SEMARNAT).

Las variables meteorológicas necesarias se adquirirán del Servicio Meteorológico Nacional (SMN) de la Comisión Nacional del Agua (CONAGUA), incluyendo datos horarios como la temperatura, humedad, velocidad y dirección del viento, precipitación y presión atmosférica. De igual forma se tomarán en cuenta variables de movilidad urbana recaudados desde el Instituto Nacional de Estadística y Geografía (INEGI), todo esto con el objetivo de hacer que el modelo sea lo mas preciso posible.

7 Agenda

ACTIVIDAD	SEMANA 1	SEMANA 2	SEMANA 3	SEMANA 4
Búsqueda de literatura	х	x		

Análisis de artículos	х	x	
Redacción de revisión		x	X
Propuesta de modelo propio			х
Elaboración de recomendaciones			Х

8 Conclusión

Esta investigación demostró el potencial de las distintas arquitecturas de redes neuronales como la LSTM junto con la integración de datos locales el modelo capturará adecuadamente dependencias temporales para la predicción de concentraciones de contaminación atmosférica en México.

Además el análisis del perfil nacional de emisiones confirmó la presencia de contaminantes como PM_{10} , $PM_{2\cdot5}$, NO_x y COV, originados principalmente por las emisiones de combustión del sector transporte e industrial, es por ello que se sustenta la urgencia de implementar estos sistemas que anticipen episodios críticos de contaminación.

Es por ello que investigación proporciona las bases para desarrollar sistemas de alerta temprana que contribuyan a la gestión ambiental y protección de la salud pública en entornos urbanos mexicanos.

9 Referencias

- Kelly, F. J., & Fussell, J. C. (2015). Air pollution and public health: emerging hazards and improved understanding of risk. *Environmental Geochemistry And Health*, *37*(4), 631-649. https://doi.org/10.1007/s10653-015-9720-1
- Francisco, V. M. (s. f.). La contaminación ambiental como factor determinante de la salud. https://scielo.isciii.es/scielo.php?script=sci_arttext&pid=S1135-57272005000200001
- Modelo de prediccion de emisiones. (2024). [Universidad Pontificia Comillas].
 https://repositorio.comillas.edu/xmlui/bitstream/handle/11531/80074/TFG-
 Ordonez%20Becker%2c%20Beatriz.pdf?sequence=1&isAllowed=y
- Sánchez, C. A. S. (2024). Predicción de Contaminantes Atmosféricos en Bogotá utilizando Redes LSTM.
 https://portal.amelica.org/ameli/journal/266/2664941006/html/
- Vista de Análisis y Predicción de las Emisiones de CO₂ en Bolivia a través de Redes
 Neuronales Artificiales. (s. f.).

 https://revistasaga.org/index.php/saga/article/view/86/147
- FORECASTING AND MONITORING URBAN AIR QUALITY IN MADRID. (2020). [Universidad Nacional de Educación a Distancia]. https://arxiv.org/pdf/2011.09741
- Rosales-Castillo, J. A., Torres-Meza, V. M., Olatz-Fernández, G., & Borja-Aburto, V. H. (2001). Los efectos agudos de la contaminación del aire en la salud de la población: evidencias de estudios epidemiológicos. *Salud Pública de México*, 43(6), 544-555.