



UNIVERSIDAD NACIONAL AUTÓNOMA DE MÉXICO

PROGRAMA DE POSGRADO EN CIENCIA E INGENIERÍA DE
LA COMPUTACIÓN

Análisis y propuesta de aceleración del Modelo de Distribución Espacial, Temporal del Inventario de Emisiones de México (DiETE).

TRABAJO FINAL
QUE PARA OPTAR POR EL GRADO DE
ESPECIALIZACIÓN EN CÓMPUTO DE ALTO RENDIMIENTO

PRESENTA:
DULCE ROSARIO HERRERA MORO

TUTOR:
DR. OSCAR ALEJANDRO ESQUIVEL FLORES

MIEMBROS DEL COMITÉ
DR. JOSÉ JESÚS CARLOS QUINTANAR SIERRA, CIENCIAS
DR. AGUSTÍN GARCÍA REYNOSO, ICAYCC.

CIUDAD UNIVERSITARIA CDMX, Abril 2022



Universidad Nacional
Autónoma de México

Dirección General de Bibliotecas de la UNAM

Biblioteca Central



UNAM – Dirección General de Bibliotecas
Tesis Digitales
Restricciones de uso

DERECHOS RESERVADOS ©
PROHIBIDA SU REPRODUCCIÓN TOTAL O PARCIAL

Todo el material contenido en esta tesis esta protegido por la Ley Federal del Derecho de Autor (LFDA) de los Estados Unidos Mexicanos (México).

El uso de imágenes, fragmentos de videos, y demás material que sea objeto de protección de los derechos de autor, será exclusivamente para fines educativos e informativos y deberá citar la fuente donde la obtuvo mencionando el autor o autores. Cualquier uso distinto como el lucro, reproducción, edición o modificación, será perseguido y sancionado por el respectivo titular de los Derechos de Autor.

Dedicado a Cali, mi hermosa luna.

Análisis y propuesta de aceleración del Modelo de Distribución Espacial, Temporal del Inventario de Emisiones de México (DiETE).

Tesis Especialidad

Dulce Rosario Herrera Moro
Especialización en Cómputo de Alto Rendimiento
Posgrado en Ciencia e Ingeniería de la Computación
Universidad Nacional Autónoma de México

Resumen

La contaminación atmosférica, fuertemente relacionada con problemas de morbilidad y mortalidad, disminuye notablemente la calidad de vida de la población, en especial la más vulnerable que corresponde a los niños menores de cinco años y los adultos mayores de una región contaminada. La modelación de la calidad del aire puede ayudar a conocer el impacto de políticas públicas en concentraciones ambientales, por lo que poder generar un inventario de emisiones para modelación, fácil de modificar y que estime las emisiones de forma expedita, es una herramienta fundamental para alcanzar el control de emisiones y disminuir la exposición de la población a contaminantes. Este trabajo presenta el análisis y propuesta de optimización de código fuente del modelo de distribución espacial, temporal y de especiación (DiETE) del Inventario de Emisiones de México para el año base 2016 desarrollado por investigadores del Instituto de Ciencias de la Atmósfera y Cambio Climático. El modelo ha sido utilizado en diferentes estudios de Calidad de Aire en la Ciudad de México, Monterrey y otros dominios de interés para definir o evaluar medidas en la gestión ambiental. Además, el DiETE se usa en el pronóstico operativo de calidad de aire que se publica a diario en la página Web del Instituto.

El modelo está implementado en Fortran y como restricción del proyecto se respetó el lenguaje de programación. Durante el análisis de la implementación, se identificó al proceso de generación de los archivos NETCDF con la distribución de las emisiones como el más tardados llegando a ocupar hasta un 60 % del tiempo total de ejecución. En el trabajo se presentan los resultados obtenidos de dos de las estrategias probadas, con los mejores resultados en la acelerar la implementación del proceso antes mencionado. En la primera aproximación, se toma el algoritmo original del proceso y se realiza una paralelización de granularidad gruesa utilizando directivas de compilador para OMP. La paralelización en la que se de las emisiones de área, móviles y puntuales logrando un factor de mejora del rendimiento de 2 utilizando tres cores. El resultado de la segunda aproximación fue el cambiar un algoritmo de orden $O(n^2)$ por otro de orden $O(n)$ con lo que se obtuvo un factor de rendimiento promedio de 28. Esta propuesta permite que el tiempo de ejecución del DiETE con la configuración más tardada pase de 1 hora 19 minutos a 34 minutos 37 segundos con la misma cantidad de recursos de cómputo (cores).

Analysis and acceleration proposal of the Spatial and Temporal Distribution Model of the Emissions Inventory of Mexico (DiETE)

Specialization Thesis

Dulce Rosario Herrera Moro

High Performance Computing Specialization
Posgrado en Ciencia e Ingeniería de la Computación
Universidad Nacional Autónoma de México

Summary

Air pollution has been strongly related to morbidity and mortality problems, notably decreasing quality of life in the general population but even more in children under five years old and elderly. Air quality modeling can help to understand the impact of public policies on environmental concentrations, so being able to generate an emissions inventory for modeling that is easy to modify and that estimates emissions quickly is a fundamental tool, in order to achieve the goals of controlling emissions and reducing the population's exposure to pollutants. This work presents an analysis and a proposal for source code optimization of the model called DiETE that makes a spatial, temporal, and speciation distribution of the Mexican Emission Inventory for the base year 2016 developed by researchers in the Institute for Atmospheric Sciences and Climate Change.

The DiETE model has been used in different Air Quality studies for Mexico City, Monterrey, and other relevant domains to define or evaluate environmental management regulations. In addition, DiETE is used in the operational air quality forecast published daily on the Institute's web page. A project constraint consists of not changing the original programming language Fortran. During the implementation analysis phase, the output NETCDF files generation process was identified as the most time-consuming, taking up to 60 % of the total execution time.

Results obtained from the two best strategies for code optimization tested are shown. In the first approach, the original algorithm is taken and a coarse-grained parallelization is implemented by OMP compile directives. Using this approach, an average speedup value of two is acquired using three cores. The second approach consists of changing an algorithm of order $O(n^2)$ for another of order $O(n)$, thus obtaining an average performance factor of 28. This proposal allows the execution time of DIETE with the slowest configuration to go from 1 hour 19 minutes to 34 minutes 37 seconds with the same amount of computational resources (cores).

Índice general

Dedicatoria	I
Resumen	II
Summary	III
1 Introducción	1
§1.1 Motivación	2
§1.2 Objetivos	3
§1.2.1 General	3
§1.2.2 Particular	3
§1.3 Metodología	3
§1.4 Organización del documento	4
2 Antecedentes	5
§2.1 Contaminación del aire	5
§2.1.1 Clasificación y efectos de los contaminantes	6
§2.1.2 Efectos de la contaminación del aire	7
§2.2 Inventarios de Emisiones	8
§2.3 Modelo DiETE	10
§2.3.1 Introducción	10
§2.3.2 Procesos del Modelo DiETE	12
§2.4 Resumen	23
3 Diagnóstico y Aceleración del código DiETE	24
§3.1 Diagnóstico del código original	24
§3.2 Estrategias de aceleración y resultados	31
§3.2.1 Paralelización OPENMP	32
§3.2.2 Cambio de algoritmo	34
§3.3 Resumen	37
4 Conclusiones y Trabajo Futuro	38
§4.1 Conclusiones	38

Apéndices	42
Códigos de clasificación de fuentes	43

Índice de figuras

2.1	Clasificación de las fuentes de emisión en el Inventario Nacional de Emisiones	9
2.2	Principales componentes de un modelo numérico de calidad de aire	11
2.3	subdominios DiETE	13
2.4	Proceso para generar las retículas en DiETE	14
2.5	Distribución de emisiones de área	16
2.6	Distribución de emisiones de área	17
2.7	Archivos con los perfiles temporales en DiETE	18
2.8	Ejemplo de las distribuciones espacial y temporal de las emisiones de DiETE	20
2.9	Ejemplos de como son almacenadas las variables en NetCDF	21
2.10	Diagrama general de los procesos del DiETE	22
3.1	Porcentaje del tiempo total de ejecución que toman los diferentes módulos del código de original del DiETE usando diferentes dominios y resoluciones.	28
3.2	Aceleración del algoritmo paralelo OMP	34
3.3	Aceleración del algoritmo propuesto	36
3.4	Aceleración del algoritmo propuesto	37

Capítulo 1

Introducción

El Centro de Ciencias de la Atmósfera (CCA) de la Universidad Nacional Autónoma de México (UNAM) se ha consolidado como un referente a nivel nacional en investigación en ciencias del ambiente en el que, de acuerdo con su informe de labores 2013-2021, se realizan el 75 por ciento de las publicaciones sobre atmósfera y ambiente a nivel nacional. El Centro en colaboración con diversas entidades públicas y privadas a nivel federal y estatal ha contribuido tanto en el monitoreo atmosférico, ambiental y climático así como a la generación de Normas Oficiales Mexicanas (NOM). Entre las diferentes entidades gubernamentales con las que el Centro ha establecido colaboraciones, se destacan pero no se limitan a: la Comisión Nacional del Agua (CONAGUA), la Comisión Ambiental de la Megalópolis (CAME), la Secretaría de Ciencia, Tecnología e Innovación de la CDMX (SECITI), el Servicio Meteorológico Nacional (SMN), la Secretaría de Salud (SSA), el Centro Nacional de Prevención de Desastres (CENAPRED), la Secretaría de Energía (SENER), la Comisión Federal de Electricidad (CFE), la Secretaría del Medio Ambiente (SEDEMA), la Secretaría de Relaciones Exteriores (SRE), el Instituto Nacional de Ecología y Cambio Climático (INECC) así como diversos gobiernos estatales. Estas colaboraciones le permitieron al Centro realizar proyectos nacionales de gran impacto social entre los que destacan: su participación en la elaboración y validación científica de Normas Oficiales Mexicanas (NOM) en materia ambiental, participación en los comités de normas ISO-14000 del Comité Técnico Nacional de Sistemas de Administración Ambiental (COTENSAAM) y la participación en la Comisión Ambiental de la Megalópolis (CAME).

En 2012 el departamento de Instrumentación diseñó y construyó una radiosonda cautiva para el estudio de la atmósfera y ozono. La radiosonda se ha utilizado para efectuar estudios de dispersión de contaminantes para la Comisión Federal de Electricidad en tres termoeléctricas del país, entre ellas, la de Tuxpan, Veracruz. En 2016 el Grupo de participó en un proyecto con la Secretaría de Ciencia, Tecnología e Innovación de la CDMX (SECITI) para estudiar la exposición a la contaminación por material particulado en la Ciudad de México que consistió en el diseño y construcción de sistemas para obtener la localización espacio-temporal en tiempo real de monóxido de carbono en áreas urbanas. En 2018, el grupo de Contaminación Ambiental inició un proyecto para la evaluación de la calidad del aire, depósito atmosférico y meteorología para desarrollar el programa para la prevención y minimización del posible deterioro ambiental significativo en el Recinto Portuario de Veracruz y otras zonas de interés. El proyecto es financiado por la Administración Portuaria Integral de Veracruz, S.A. de C.V. (APIVer). En agosto del mismo año (2018) se firma un convenio colaboración con la SEDEMA [SEDEMA, 2018] y el Gobierno de la Ciudad de México el Instituto para contribuir en la generación de información y conocimiento que sustenten la creación de políticas públicas eficaces que contribuyan a reducir la contamina-

ción atmosférica en la CDMX. En 2019 el Centro fue responsable de la generación de cinco boletines informativos y participó en una conferencia de prensa durante la contingencia ambiental declarada el 14 de mayo de ese año durante. Por todos los logros, contribuciones y la gran relevancia durante sus 44 años de existencia en septiembre de 2021 el pleno del Consejo Universitario aprobó por unanimidad la transformación de del CCA a Instituto de Ciencias de la Atmósfera y Cambio Climático (ICAYCC).

Además de los diversos proyectos y estudios científicos que se realizan en el día a día del Instituto, desde el año 2010, se corre diariamente un pronóstico de calidad del aire. El pronóstico fue implementado por el grupo de Físicoquímica Atmosférica y utiliza el modelo numérico WRF-chem (*Weather Research and Forecasting with Chemistry*) que acopla información meteorológica y de química atmosférica para estimar las emisiones biogénicas a partir de datos de fuentes de emisiones puntuales, móviles, de área y naturales. Cada día se publica el pronóstico a 73 horas de las concentraciones de monóxido de carbono (CO), óxidos de nitrógeno (NOx), *ozono*, Partículas PM_{10} y dióxido de azufre (SO_2) para la CDMX y Golfo del México en <https://www.atmosfera.unam.mx/calidad-del-aire-en-mexico>. Los datos meteorológicos utilizados en el pronóstico de calidad de aire son obtenidos a partir del pronóstico meteorológico implementado en el Instituto mientras que los datos de emisiones se basa en el inventario de emisiones de la República Mexicana del año 2016 que son procesados por el modelo de distribución espacial, temporal y de especiación (DiETE) [García-Reynoso et al., 2018] desarrollado por el Dr. José Agustín García Reynoso investigador del ICAYCC. Además del pronóstico operativo, el DiETE se ha utilizado para realizar diversos estudios de calidad del aire en Guadalajara, Monterrey y Centro de México. El DiETE está implementado en Fortran90 y toma un tiempo promedio de 1 hora 15 minutos para convertir el inventario anual municipal a un inventario horario en una retícula de 3x3km con las emisiones de un solo día de toda la República Mexicana, considerando que para el pronóstico se calculan a diario las emisiones, la ejecución del DiETE tomaría alrededor de 3 horas 45 minutos. Con la finalidad de reducir los tiempos de la versión actual del DiETE, la estrategia ha sido el no calcular las emisiones de la toda la República Mexicana sino dividir el dominio en diferentes áreas de estudio. Es por esta razón que el pronóstico diario se limita a la área metropolitana de la Ciudad de México.

Durante el desarrollo del presente trabajo se realizó un análisis del código fuente de la implementación del modelo DiETE con la finalidad de proponer e implementar una actualización que permitan reducir el tiempo de ejecución sin tener que reducir tamaño del dominio o cambiar el tamaño de las celdas para reducir su número.

1.1. Motivación

La contaminación atmosférica está fuertemente relacionada con problemas de morbilidad y mortalidad, disminuyendo notablemente la calidad de vida de la población, en especial la más vulnerable que corresponde a los niños menores de cinco años y los adultos mayores de una región contaminada. En 2012 la Organización Mundial de la Salud (OMS) estimó que a causa de la contaminación atmosférica exterior, producida tanto por fuentes urbanas como rurales, se produjeron 3,7 millones de muertes en todo el mundo. Para el

caso de la Ciudad de México, el informe anual de sobre Calidad del Aire 2018 reportó 271 días en la categoría MALA y 5 días en la categoría MUY MALA del índice de calidad del aire. Dada la relevancia que ha alcanzado el ICAYCC sus estudios en calidad tienen el potencial de influir en la creación y validación de políticas públicas que regulan las actividades y procesos en la industria, en la agricultura y otras actividades económicas vitales del país.

Un elemento indispensable en los estudios de calidad del aire son los modelos numéricos. En el ICAYCC, el modelo más utilizado en calidad del aire es el WRF-CHEM desarrollado por un grupo de expertos. Sin embargo, tanto para pre-procesamiento de entradas como el análisis de las salidas se realizan con diversas herramientas de Software que en muchas ocasiones son desarrolladas por los investigadores. El presente trabajo se enfoca en reducir el tiempo de ejecución del código fuente del modelo DiETE que se encarga convertir el inventario de emisiones oficial a un formato adecuado para el modelo WRF-CHEM. Lo anterior, se aplicando las habilidades y conocimientos técnicos adquiridos en la Especialidad del Cómputo de Alto Rendimiento.

1.2. Objetivos

1.2.1. General

El objetivo del presente trabajo reducir el tiempo de ejecución de la implementación actual el código fuente del modelo DiETE, esto se logrará mediante el análisis de la implementación actual para proponer e implementar estrategias computacionales adecuadas.

1.2.2. Particular

- Entender, dominar y documentar los módulos de la implementación actual del modelo.
- Identificar el o los subprocesos que representan un cuello de botella durante la ejecución del modelo.
- Realizar un análisis de los algoritmos implementados en para proponer e implementar estrategias computacionales adecuadas que permitan reducir el tiempo de ejecución.

1.3. Metodología

En esta sección se presenta la metodología seguida en el presente trabajo para reducir el tiempo de ejecución del modelo DiETE.

- Estudiar y analizar el código fuente de la implementación actual del modelo DIETE.

- Medir los tiempos de ejecución de los diferentes módulos identificando cuales de ellos impactan más en el tiempo de ejecución total.
- Identificar los elementos que más constituyen a crear un cuello de botella durante la ejecución.
- Proponer alternativas a los elementos problemáticos.
- Refactorizar el código con base en la solución propuesta.
- Realizar comparaciones de los tiempos de ejecución para comprobar que se ha cumplido con el objetivo de reducción de tiempo.
- Dar propuestas de trabajo futuro sobre el tema.

1.4. Organización del documento

El presente documento está organizado en cuatro capítulos entre los que se incluye esta introducción. Adicionalmente a los capítulos, se incluye una sección adicional de apéndices que complementan el trabajo. A continuación se presenta una breve descripción de cada uno de los siguientes capítulos.

En el capítulo 2, *Antecedentes*, se presentan los conceptos fundamentales relativos a Contaminación del aire los factores que influyen y el impacto que tiene en la salud de la población. En el capítulo se presenta también una descripción detallada del Modelo DiETE, de la implementación actual, el análisis de los tiempos de ejecución, y aspectos mejorables del código fuente.

En el capítulo 3, *Aceleración del código DiETE*, se presentan las modificaciones realizadas al código fuente y el impacto que tienen en los tiempos de ejecución.

Finalmente se concluye el trabajo en el capítulo 4, *Conclusiones*, que contiene los comentarios finales y sugerencias de trabajo futuro.

Capítulo 2

Antecedentes

En el presente Capítulo se exponen los conceptos básicos de Contaminación del Aire y sus efectos, qué son los Inventarios de Emisiones y el papel del Instituto de Ciencias de la Atmósfera y Cambio en los estudios de contaminación del aire. Además se describe y documenta el funcionamiento del código fuente que implementa el modelo DiETE.

2.1. Contaminación del aire

Un contaminante del aire puede definirse como cualquier sustancia emitida a la atmósfera que altera de forma negativa la composición natural del aire. [INECC, 2005]. Por lo tanto, la contaminación del aire hace referencia a la presencia de un conjunto de contaminantes, gases o partículas, de origen natural o se resultantes de actividades humanas.

Los factores que intervienen en la dispersión y acumulación de contaminantes son [Vigil et al., 2013].

- **Eventos meteorológicos.** La temperatura del aire determina el estado y movimiento de las masas de aire facilitando o dificultando la dispersión de contaminantes.
- **Características geográficas.** Influyen en la difusión o acumulación de los contaminantes, por ejemplo, las montañas frenan los vientos, favoreciendo la acumulación de contaminantes en zonas urbanas.
- **Fuentes de emisión.** Lanzan sustancias al aire por una fuente localizada o como resultado de reacciones fotoquímicas.

La calidad del aire en una región se determina por comparación entre mediciones de contaminantes criterio y los límites definidos por las normas de salud que establecen los valores tolerables de a los que los seres humanos y ecosistemas pueden estar expuestos a dichos contaminantes. Los contaminantes criterio se establecen con base en los impactos que tienen en la salud de la población y en el ambiente. Es decir, se seleccionan aquellos contaminante que provocan efectos agudos en la salud humana. Los contaminantes criterio establecidos para la CDMX son los óxidos de nitrógeno (NO_x), óxidos de azufre (SO_X), partículas suspendidas (PM), monóxido de carbono (CO) y compuestos orgánicos volátiles (COV).

Si bien los contaminantes criterio a corto plazo no son los únicos y existen contaminantes que pueden provocar efectos a largo plazo en la salud por lo que es de suma importancia realizar estudios

La contaminación atmosférica en está fuertemente relacionada con problemas de morbilidad y mortalidad, disminuyendo notablemente la calidad de vida de la población, en especial la más vulnerable que corresponde a los niños menores de cinco años y los adultos mayores de una región contamina. En 2012 la Organización Mundial de la Salud (OMS) estimó que a causa de la contaminación atmosférica exterior producida por fuentes tanto urbanas como rurales, se produjeron 3,7 millones de muertes en todo el mundo.

En [Silva et al., 2017] se predice que en los años 2030 y 2100 habrá 55,600 y 215,000 muertes asociadas una sobre exposición a partículas $PM_{2.5}$ respectivamente.

2.1.1. Clasificación y efectos de los contaminantes

En [Vigil et al., 2013] se realiza la siguiente una clasificación y caracterización de contaminantes:

- **Contaminantes primarios.** Son vertidos directamente a la atmósfera por alguna fuente de emisión como chimeneas, automóviles, entre otros. Los contaminantes que se encuentran en esta clasificación son: (SO_2), (CO), (NO_2), partículas $PM_{2.5}$, partículas PM_{10} y COV .
 - *Dióxido de azufre (SO_2).* El dióxido de azufre irrita las vías respiratorias y en altas concentraciones puede causar broncoconstricción y agravar enfermedades respiratorias y cardiovasculares. Los grupos de población vulnerables al dióxido de azufre son los niños, los adultos mayores y personas que padecen asma y enfermedades pulmonares. Se cree que la toxicidad del dióxido de azufre es modulada por la presencia de partículas.
 - *Monóxido de carbono (CO).* Es un gas incoloro no irritante sin olor o sabor que se produce por la combustión incompleta de carbón. Alrededor del 70 por ciento del CO proviene de los vehículos. En altas concentraciones inhabilita el transporte de oxígeno hacia las células. Si la exposición es prolongada puede causar daño permanente en el corazón y cerebro.
 - *Óxido de Nitrógeno (NO_2).* Son gases que se emiten en los procesos de combustión que se llevan a cabo en relación con el tráfico (en especial vehículos Diesel), así como en instalaciones industriales de alta temperatura y de generación eléctrica. En humanos pueden producir inflamación de las vías aéreas, afecciones de órganos, como hígado o bazo, o de sistemas, como el sistema circulatorio o el inmunitario, que propician a su vez infecciones pulmonares e insuficiencias respiratorias). En el medio ambiente causa acidificación y eutrofización de ecosistemas, afecciones metabólicas así como limitación del crecimiento vegetal.
 - *Partículas.* Es material respirable presente en la atmósfera en forma sólida o líquida (polvo, cenizas, hollín, partículas metálicas, cemento y polen, entre otras).

De acuerdo con su tamaño se pueden dividir en dos grupos principales: las de diámetro igual o inferior a los $10\mu\text{m}$ (PM_{10}) y las de fracción respirable más pequeña ($PM_{2.5}$). Las partículas (PM_{10}) agravan el asma y se asocian con el desarrollo de diabetes además, pueden ocasionar disminución en el tamaño del feto. Además de agravar las enfermedades respiratorias cardiovasculares, una exposición crónica a altas concentraciones de partículas ($PM_{2.5}$) pueden provocar un incremento en el riesgo de morbilidad y mortalidad.

- *Hidrocarburos (HC)*. Las emisiones de HC están asociadas a la mala combustión de derivados del petróleo. Las fuentes más importantes de emisión son el transporte por carretera, los disolventes, pinturas, vertederos y la producción de energía. Los de mayor interés son los compuestos orgánicos volátiles (COV), dioxinas, furanos, bifenilos policlorados (PCB) y los hidrocarburos policíclicos aromáticos (PAH).
- **Contaminantes secundarios**. Son los contaminantes originados en el aire como consecuencia de la transformación y reacciones químicas que sufren los contaminantes primarios en la atmósfera. (O_3), lluvia ácida y contaminación fotoquímica.
 - *Ozono (O_3)*. Forma parte de la composición de la atmósfera, sin embargo a baja altura resulta perjudicial por su carácter oxidante, reactivo corrosivo y tóxico. En humanos irrita las vías respiratorias y en altas concentraciones reduce la función pulmonar
 - *Lluvia ácida*. Es el término que se ha usado para describir el proceso por el cual ciertos ácidos se forman en la atmósfera a partir de contaminantes y luego se precipitan a la tierra. El SO_2 (dióxido de azufre) y los NO_x , causan la lluvia ácida. Estas sustancias en presencia de agua, O_2 y otros compuestos químicos forman ácido sulfúrico (H_2SO_4) y ácido nítrico (HNO_3) respectivamente, que se precipitan a la tierra en forma líquida cuando se presentan lluvias o en forma seca en presencia de nevadas o neblinas. Cuando la lluvia ácida cae sobre el suelo y cuerpos de agua, altera sus características químicas y pone en peligro el equilibrio de los ecosistemas. Además, los compuestos químicos que contiene la lluvia ácida dañan la superficie de edificios, estatuas, monumentos y automóviles.
 - *Contaminación fotoquímica*. La constituyen la luz solar y sustancias susceptibles de ser oxidadas. El smog fotoquímico es una mezcla de contaminantes que se forman por reacciones producidas por la luz solar al incidir sobre contaminantes primarios. Los efectos en la salud humana va desde empeoramiento de alergias, hasta dificultades para respirar, cáncer, etc. Además puede causar una reducción de la visibilidad por irritación ocular

2.1.2. Efectos de la contaminación del aire

En la sección anterior se presentó como afectan a la salud humana los contaminantes. Sin embargo, la contaminación del aire no solo afecta la salud de los seres humanos también tiene efectos en los ecosistemas y está íntimamente relacionada con el cambio climático.

Los ecosistemas se ven afectados principalmente por contaminantes secundarios como la lluvia ácida que obstruye y acidifica los poros de las hojas dificultando el proceso de fotosíntesis, además de contribuir a la degradación los suelos que afecta a las raíces y por lo tanto a la nutrición de las plantas. Otro de sus efectos de la lluvia ácida es la disminución de peces al acidificar lagos, lagunas, ríos y arroyos [Vigil et al., 2013]

En [Ruiz-Suárez et al., 2018] se reporta el impacto económico que tiene la exposición a altas concentraciones de ozono en cuatro cultivos importantes en el área de la Megalópolis de la Ciudad de México. El estudio reporta pérdidas en el rendimiento de los cultivos del maíz en 3 %, avena 26 %, frijol 14 % y sorgo 15 %

En el 2003 la sección de Contaminación Ambiental del CCA-UNAM publicó un estudio sobre el efecto de la lluvia ácida sobre la piedra caliza en los monumentos mayas de Tulum, Quintana Roo [Bravo-Álvarez et al., 2003]. El estudio determinó que la roca caliza sufre una disolución de cuatro a cinco micrómetros por año provocada de la lluvia ácida con el carbonato de calcio propio de la roca caliza.

La contaminación del aire y el cambio climático están interrelacionados ya que el clima, y en particular la temperatura del planeta, dependen del balance entre la energía que recibe del sol y el calor que emite. Los gases de efecto invernadero (GEI) presentes naturalmente en la atmósfera dejan pasar la radiación solar hacia la superficie terrestre, pero absorben la radiación infrarroja que esta última emite, produciendo con ello un efecto neto de calentamiento [SEMARNAT, 2018]. La quema de combustibles fósiles emiten grandes cantidades de gases y compuestos de efecto invernadero que se suman a los existentes en la atmósfera, con lo que se incrementa su concentración en la atmósfera y con ello el efecto de calentamiento.

2.2. Inventarios de Emisiones

Los inventarios de emisiones tienen como finalidad identificar y caracterizar las principales fuentes de contaminantes que afectan la calidad de aire de una región.

En la Figura 2.1 se muestra la clasificación de las Fuentes de emisiones que se utiliza en los Inventarios de Emisiones.

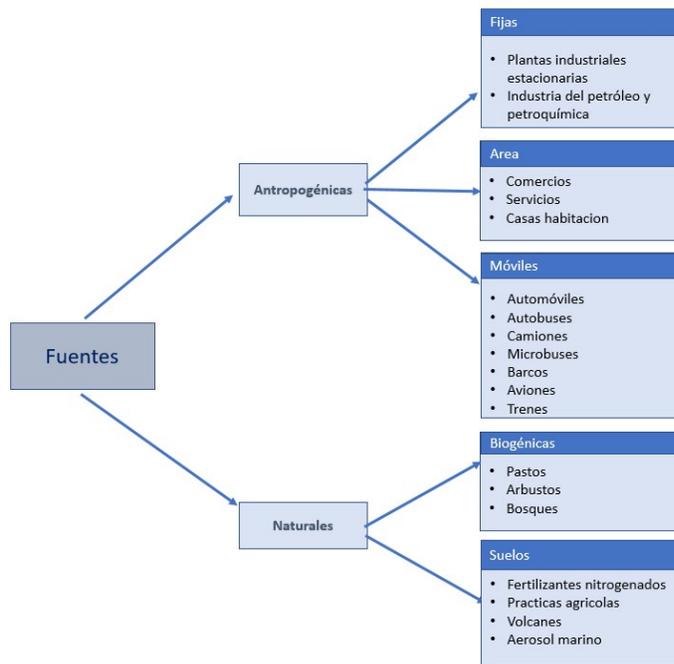


Figura 2.1: Clasificación de las fuentes de emisión en el Inventario Nacional de Emisiones

Las fuentes de emisión se clasifican en [Jimate, 2013]:

- **Fuentes fijas.** Emiten contaminantes al aire por ductos o chimeneas. Los contaminantes emitidos por este tipo de fuentes son producto de la combustión utilizada para la generación de energía. Los principales contaminantes asociados a estas fuentes son las partículas SO_2 , NO_x , CO_2 , CO e hidrocarburos.
- **Fuentes de área.** Son todas aquellas actividades que en conjunto afectan la calidad del aire, como el uso de madera, el tratamiento de aguas residuales, plantas de composteo, rellenos sanitarios, artes gráficas, panaderías, distribución y almacenamiento de gas LP, principalmente.
- **Fuentes de móviles.** Se consideran los vehículos automotores que circulan por carretera. Las emisiones que producen los vehículos son del tubo de escape, los vapores provenientes de fugas de combustible con el motor en operación, las emisiones que se producen por el aumento de la temperatura ambiente en el tanque de combustible.
- **Fuentes naturales o biogénicas.** Se refiere a la generación de emisiones producidas por volcanes, océanos, plantas, suspensión de suelos, emisiones por digestión anaerobia y aerobia de sistemas naturales. Las emisiones biogénicas incluyen óxido de nitrógeno, hidrocarburos no metanogénicos, metano, dióxido y monóxido de carbono y compuestos nitrogenados y azufrados.

En los inventarios de emisiones se estiman las aportaciones de contaminantes de cada una de las fuentes de emisión. Para la estimación de emisiones de fuentes fijas se utilizan

factores de emisión para contaminantes generados por la Agencia de Protección Ambiental de Estados Unidos (APA) y factores de emisión para gases y compuestos de efecto invernadero del Grupo Intergubernamental de Expertos sobre el Cambio Climático (IPCC). Los factores comúnmente los factores se expresan como el peso del contaminante dividido por unidad de peso, volumen, distancia o duración de la actividad. En 2.1 se presenta la ecuación general para una estimación en donde: E= emisión , A= Taza de actividad , FE=Factor de emisión

$$E = (A)(FE) \quad (2.1)$$

En la estimación de emisiones de fuentes móviles se consideran los datos de actividad son los kilómetros recorridos por los vehículos, flota vehicular , tipos de vialidades así como el tipo y calidad de los combustibles.

Entre los años 1995 y 2000 el Instituto Nacional de Ecología (ahora INECC), con el apoyo de la Agencia de Protección al Ambiente de los Estados Unidos y la Asociación de Gobernadores del Oeste de ese país, se adecuó una metodología a las condiciones particulares de México para crear por primera vez un inventario de emisiones. En total se han desarrollado cinco inventarios de emisiones con cobertura nacional (1999, 2005, 2008, 2011 y 2013). Sin embargo, los inventarios elaborados por las autoridades ambientales poseen una resolución espacial y temporal que impide su utilización directa en los modelos numéricos de calidad del aire.

2.3. Modelo DiETE

2.3.1. Introducción

Los modelos numéricos de calidad de aire se han consolidado como una herramienta indispensable en los estudios de calidad de aire. Los modelos de calidad de aire emplean un conjunto de ecuaciones diferenciales discretizadas para caracterizar los procesos físicos y químicos que se presentan por la dispersión y reacción de emisiones en la atmósfera para determinar las concentraciones de especies de contaminantes sobre un dominio determinado. Para la gestión ambiental, los modelos de calidad de aire se han consolidado como una herramienta de investigación que resulta particularmente útil en los siguientes temas (Seinfeld 1978, Seinfeld and Pandis 1998):

- Los modelos son consultados durante la planificación territorial de futuras fuentes de contaminantes atmosféricos.
- Se usan para determinar cuál es la contribución de la fuente A a la inmisión del contaminante B.
- Se utilizan para hacer simulaciones de episodios de contaminación atmosférica en la elaboración de planes de actuación.

- Determinar responsabilidades frente a los niveles actuales de contaminación atmosférica.
- Se usan como base científica para la promulgación de legislaciones sobre el control de las emisiones.
- En la evaluación de técnicas y estrategias propuestas para el control de las emisiones.
- Determinar cuál es la estrategia considerando la efectividad-costo para reducir las concentraciones de un contaminante por debajo de un criterio de calidad del aire.
- Realizar pronósticos de calidad del aire.

Las características de una modelación numérica se definen con base en el objetivo que se persigue por lo que se debe seleccionar cuidadosamente el periodo y la región a estudiar, el mecanismo químico a emplear, la configuración meteorológica seleccionada, etc. En la Figura 2.2 se presenta un esquema con los componentes principales involucrados en un modelo numérico de calidad de aire.

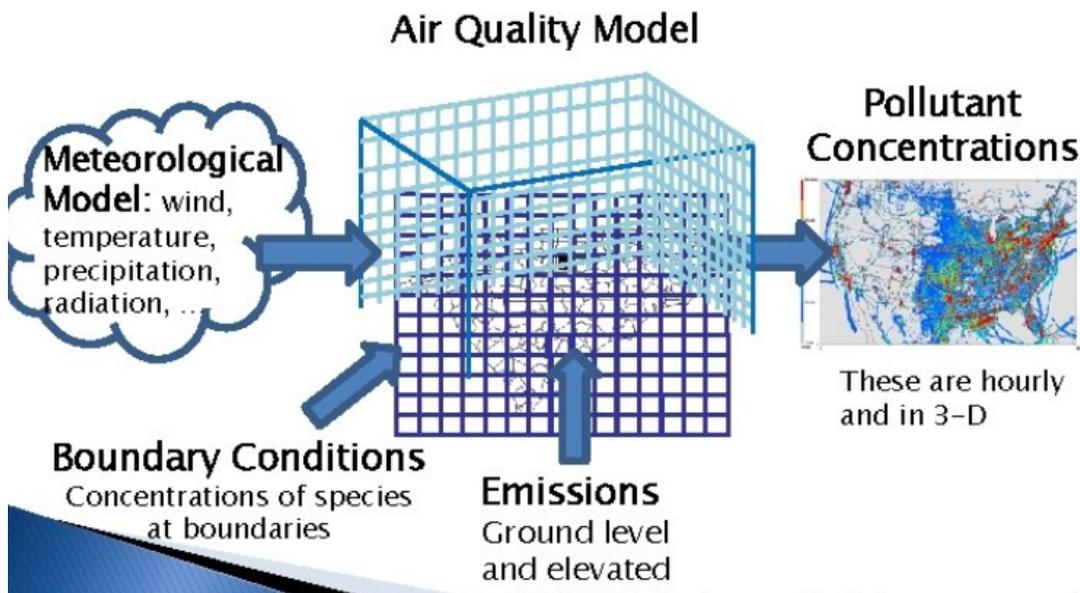


Figura 2.2: Principales componentes de un modelo numérico de calidad de aire

- **Datos meteorológicos.** Los datos de temperatura, viento, precipitación, temperatura y radiación se obtienen de modelos meteorológicos. Para el caso del pronóstico operativo de calidad de aire del ICAyCC utiliza el Global Forecast System (GFS) https://nomads.ncep.noaa.gov/txt_descriptions/GFS_doc.shtml.
- **Condiciones de frontera.** Tanto de variables meteorológicas como de concentraciones de especies químicas en los límites del dominio de estudio.

- **Emisiones de contaminantes atmosféricos** contienen la estimación de las emisiones liberadas a la atmósfera de los contaminantes de origen antropogénico como natural.

Los inventarios de emisiones son decisivos en la eficiencia y calidad de los resultados obtenidos de la modelación ya que se utilizan como condiciones iniciales y gobiernan las concentraciones ambientales en el dominio de modelación. La limitación en el uso de los Inventarios de Emisiones Nacionales en los modelos de calidad de aire radica en que, generalmente, poseen una resolución espacial y temporal que impide su utilización directa.

En [García-Reynoso et al., 2018] Investigadores del ICAYCC proponen un modelo de distribución espacial, temporal y de especiación (DiETE) que incorpora los datos oficiales de los del Inventario de Emisiones de México 2008 generado por SEMARNAT.

En [Rodríguez Zas and García Reynoso,] publicado en Revista Internacional de Contaminación Ambiental en 2021 se aplicó el modelo (DiETE) para evaluar y actualizar Inventario Nacional de Emisiones de México de 2013 a partir de una metodología no paramétrica de factores de escalamiento.

En la siguiente sección se describe a detalle las características y funciones del modelo DiETE objeto de estudio del presente trabajo.

2.3.2. Procesos del Modelo DiETE

El código del DiETE fue escrito usando programación estructurada usando el lenguaje de programación Fortran. La selección del lenguaje de programación se basó principalmente en la experiencia del desarrollador en el uso de ese lenguaje y se estableció como un requisito restrictivo en el presente trabajo. Las principales características del DiETE son:

- Tiene una cobertura nacional con resolución 9×9 km y 3×3 km.
- Coberturas de Áreas Metropolitanas con resolución 3×3 Km y 1×1 Km.
- Puede ser configurado para realizar la especiación química de contaminantes para los mecanismos de reacción SAPRC99, RADM2, RACM y CBM05 que emplea tanto el WRF-chem como los modelos CHIMERE y CMAQ.
- Utiliza ocho capas verticales para distribuir emisiones de fuentes puntuales.
- Utiliza Códigos de Clasificación de Fuentes (SCC) para realizar la distribución temporal y especiación química.
- Las emisiones de fuentes móviles se distribuyen con base en la longitud y número de carriles de las vialidades y carreteras.
- La distribución de emisiones de fuentes puntuales se realiza con base en la ubicación geográfica de las fuentes.

- Emisiones de fuentes de área distribuidas con base en Áreas Geoestadísticas Básicas (AGEB) de las zonas urbanas del país.

Con la finalidad de reducir los tiempos de ejecución en la implementación original se dividió la república mexicana en subdominios que abarcan zonas de interés. En la Figura 2.3 se muestran las diferentes zonas de estudio disponibles en DiETE.

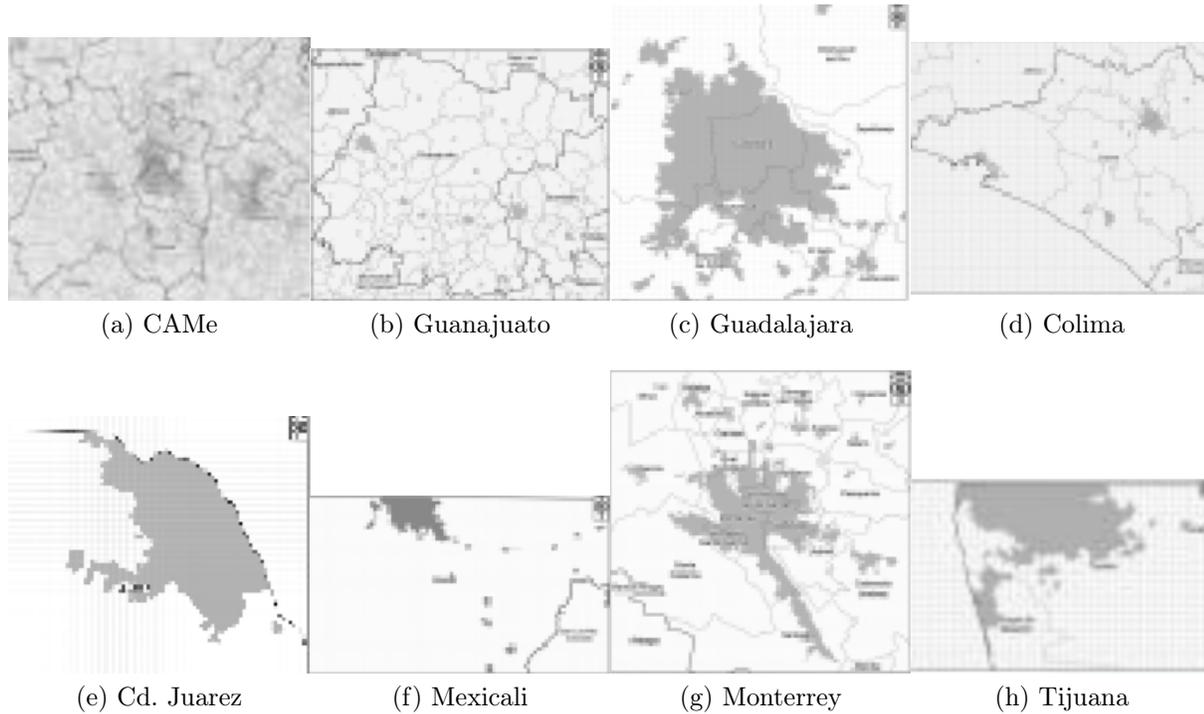


Figura 2.3: subdominios DiETE

La implementación del modelo DiETE está conformado de cuatro procesos principales: distribución espacial, distribución temporal, especiación-agrupación y formación de archivo de salida que se describen a continuación.

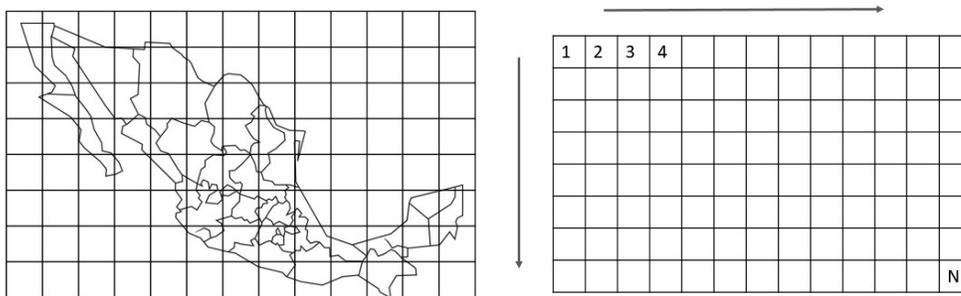
Distribución espacial

Como se vio en la sección 2.3.1 los modelos numéricos necesitan utilizar las emisiones distribuidas en cuadrículas para poder calcular las concentraciones, sin embargo, los inventarios de emisiones oficiales tienen las concentraciones anuales totales de las emisiones de contaminantes por municipio. Por lo que durante la primera etapa en el desarrollo del modelo DiETE consistió en la definición de retículas de celdas cuadradas sobre la República Mexicana (ver Figura 2.4a). Las retículas están centradas en 24.06° N, 102.03° O y tienen resoluciones de 9 km, 3 km o 1 km en cada lado. A cada una de las celdas de la retícula se le asignó un identificador numérico en la Figura 2.4b se ilustra la forma en la que se asignaron los identificadores. La información de una retícula se representa en un archivo

separados por comas (CSV) denominado `localiza.csv`. Los archivos están conformados por las siguientes columnas:

- GRIDCODE. Identificador de la celda
- LONG . Longitud correspondiente a la esquina superior izquierda de la celda.
- LATITUDE. Latitud correspondiente a la esquina superior izquierda de la celda.
- Pob. Número de habitantes en el área de la celda.
- UTMX. Coordenada X en el sistema universal UTM
- UTMY. Coordenada X en el sistema universal UTM
- UTMZ. Coordenada X en el sistema universal UTM

En la Figura 2.4c se ilustra un ejemplo del archivo `localiza.csv` del área metropolitana que abarca la CDMX y sus alrededores a la que se le denominada ECAIM.



(a) La República Mexicana se divide en bloques (b) A cada bloque se le asigna un identificador único



(c) Representación del dominio en archivos separados por comas

Figura 2.4: Proceso para generar las retículas en DiETE

Una vez generada la retícula, se utilizan herramientas de procesamiento GIS para distribuir espacialmente las fuentes de contaminantes en las celdas de la retícula. Esto se realiza trasladando la posición de las celdas de la retícula con archivos vectoriales shapefile

que contienen la posición geográfica de aeropuertos, zonas agrícolas, zonas bosques, carreteras, centrales de autobuses, ferrocarriles, caminos pavimentados, caminos de tercería, puertos y vialidades. Los archivos shapefile son proporcionados por el Instituto Nacional de Estadística y Geografía (INEGI). El resultado del proceso anterior son archivos csv, uno por cada característica geográfica considerada, que contiene la fracción del área de una celda que ocupa la característica de interés. En otras palabras, si la celda con `GRIDCODE=4000` contiene un bosque que ocupa la mitad de su área, se almacena el registro correspondiente en el archivo `bosque.csv`. La superficie marcada como *bosque* representa todo tipo de vegetación no agrícola en una celda.

Dado que en el inventario se reportan las emisiones de área y móviles por municipio, se creó el campo `CID` conformado por: `id_estado + id_municipio` el resultado se escribe en el formato `eemmm` en donde *e* corresponde a los dígitos del identificador del estado y *m* del a los dígitos del identificador de municipio.

- **Distribución espacial de emisiones de área.** Este módulo recibe como entradas los archivos con los datos descriptivos de las celdas y los archivos con los datos del inventario para emisiones de área. Se tiene un archivo de emisiones por cada uno de los contaminantes *BC*, *CO*, *CO₂*, *INH₃*, *NO_x*, *PM₁₀*, *PM_{2.5}*, *SO₂*, *VOC*. Los archivos están en formato `csv` (en la carpeta *area*) con la siguiente estructura:

CVE ESTADO	CVE MUNICIPIO	SCC1	SCC2	...	SCCN
------------	---------------	------	------	-----	------

En donde `CVE ESTADO` es el número de identificación de estado, el `CVE MUNICIPIO` es el número de identificación del municipio y las columnas de `SCC_1` de `SCC_N` contienen emisiones anuales clasificadas por el código `SCC` asociado al contaminante reportado.

Dado que las emisiones están asociadas a ciertas actividades o característica geográfica, DiETE distribuye las emisiones de área municipales reportadas en el inventario. Por ejemplo, para calcular las emisiones generadas la zona boscosa de la celda **450** de la Figura 2.5 se suman las emisiones etiquetadas con `SCC = 2`, 810,001,000 de los municipios (10043 + 01061 + 01052), el resultado se multiplica por la fracción de área de bosque calculada para la celda.

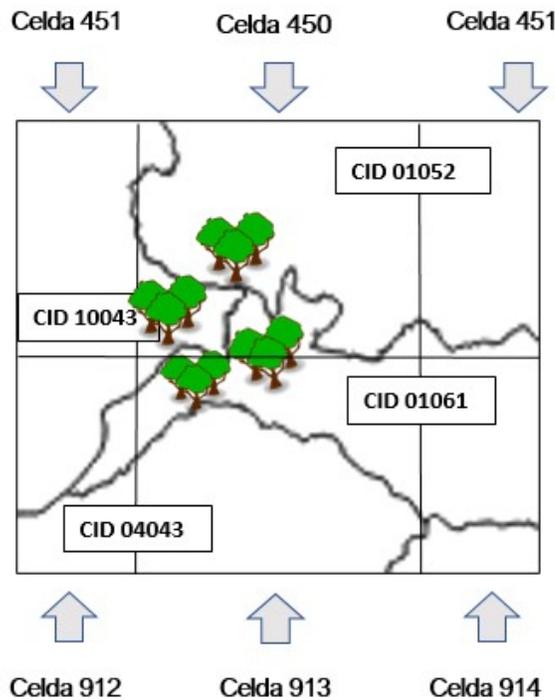


Figura 2.5: Distribución de emisiones de área

En la Tabla 2.1 se presenta como DiETE relaciona las emisiones reportadas en el inventario con base en su SCC al tipo de uso de suelo o actividad humana.

Agrícola	2801500100, 2801000002, 2267005000, 2801700000, 2805020000, 2461850000, 2801000005
Bosque	2810001000
Aeropuertos	2275000000, 2275050000
Centrales camineras	2230070310, 2270008010
Puertos Marítimos	2280000000
Ferrocarriles	2285000000, 2285002010
Vialidades	2461020000, 2401008000
Terraceria	2296000000
Población	2104007000, 30500304, 2102004000, 2102006000, 2425010000, 2102007000, 2425030000, 2103006000, 2103007000, 2104006000, 2104008000, 2104011000, 2222222222, 2302002000, 2425040000, 2302050000, 2311010000, 2401001000, 2401005000, 2465100000, 2401065000, 2401080000, 2401020000, 2401990000, 2465000000, 2415000000, 2415010000, 2420000055, 2425000000, 2465400000, 2465800000, 2465900000, 2501060000, 2610000000, 2630030000, 2850000010, 3333333333, 5555555555, 2296000000, 2294000000, 2465200000, 2465600000, 2810030000, 2265005000, 2620030000, 2270005000

Tabla 2.1: Distribución de las emisiones con base en actividades y características geográficas.

En la Figura 2.6 se ilustra el proceso que DiETE realiza para la distribución espacial se emplean coberturas de suelo comunes a las categorías de emisiones (SCC). Por ejemplo, las emisiones producidas por combustión residencial por gas se distribuyen sobre lugares donde se tenga población mientras que las emisiones relacionadas al consumo de Diesel por actividades agrícolas se distribuyen en áreas con cobertura de suelo agrícola y así sucesivamente tomando en cuenta la clasificación que se presenta en la tabla 2.1.

La salida de este proceso es un archivo csv por cada contaminante que contiene el identificador de la celda (GRIDCODE), fracción de área de la celda que pertenece a zona urbana (*Furb*), fracción de área rural (*Frural*), y los SCCs asociados al contaminante analizado. Obsérvese que en una celda pueden contener las emisiones de más de un municipio.

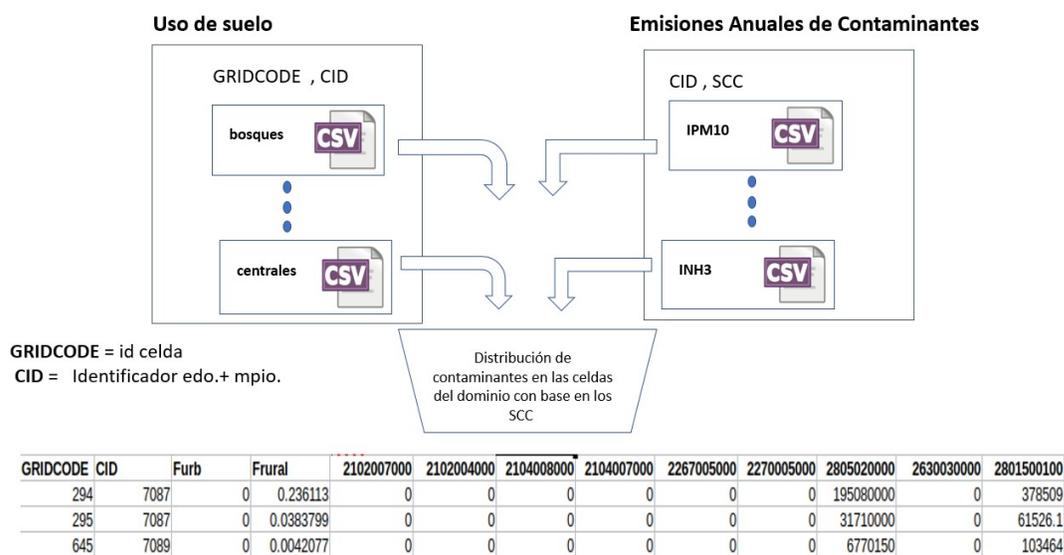


Figura 2.6: Distribución de emisiones de área

- Distribución espacial de emisiones móviles.** Las entradas a este proceso son los archivos `CARRETERA.csv` y `VIALIDADES.csv` contienen la longitud total de las vías de transporte en el municipio multiplicada por el número de carriles (dando así el área total de circulación de vehículos) y la fracción de área de vialidad correspondiente a cada una de las celdas de la retícula.

Los archivos con las emisiones móviles del inventario tiene la siguiente estructura:

CVENMUN	SCC	VOC	CO	NO	NO2	NH3	PM10
PM25	CN	CO2	S02	CH4	Categoria		

El proceso para distribuir las emisiones puntuales consiste en

- Calcular las fracciones de carreteras y vialidades siguiendo la siguiente fórmula:

$$\frac{Long_x_celda}{Long_total_x_municipio}$$

- Se combinan las fracciones calculadas de carreteras y vialidades.
- Se dividen las emisiones del municipio con base en las fracciones calculadas.

Distribución temporal

La emisión de contaminantes varía dependiendo del mes, día de la semana y hora. DiETE toma la fecha y hora configurada y selecciona los perfiles temporales generados por la EPA (2011) para obtener la distribución temporal de las emisiones. DiETE emplea 18,833 códigos SCC que se relacionan con 980 perfiles temporales mensuales, 44 semanales y 188 horarios. La distribución temporal considera las tres zonas horarias (Centro, Pacífico y Noroeste) de la República Mexicana. En la Figura 2.7 se muestra un ejemplo de los perfiles que se consultan con base en la fecha de las emisiones calculadas.

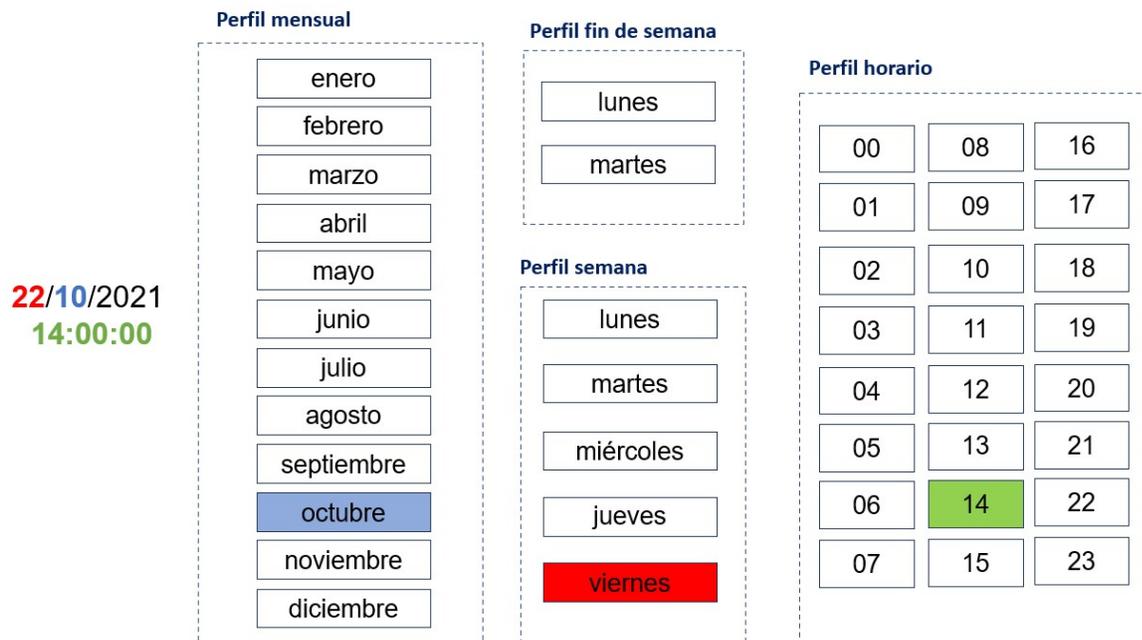


Figura 2.7: Archivos con los perfiles temporales en DiETE

Perfil temporal es un archivo *.txt* con las siguientes columnas:

SCC	MonProf	WekProf	DiuProf
-----	---------	---------	---------

SCC de emisión, MonProf identificador del perfil mensual, WekProf identificador del perfil semanal y DiuProf perfil horario.

- **Emisiones de área y móviles.** Para el proceso de distribución temporal de las emisiones de área se toman como entrada los archivos `csv` generados en el proceso de distribución espacial descrito en la sección 2.3.2.

La salida del proceso de distribución temporal son archivos `csv` con el siguiente formato para cada compuestos inorgánicos como CO, SO₂, NO_x, PM₁₀, en el caso de los compuestos inorgánicos y PM_{2.5} se incluye la columna de SCC para emplearse en pasos posteriores.

GRIDCODE	Hr1	Hr2	Hr3	. . .	Hr24
----------	-----	-----	-----	-------	------

- **Emisiones puntuales.** El archivo de entrada con las emisiones puntuales del inventario tiene las siguientes columnas:

lat	lon	SCC	ANNPM10	ANNPM25	ANNSO2	ANNCO	ANNNOX
-----	-----	-----	---------	---------	--------	-------	--------

ANNVOC	ANNCO2	ANNCH4	ANNNH3	ANNCN	Capa 6 -18	Capa 6 -18
--------	--------	--------	--------	-------	------------	------------

Especiación Química

Los compuestos orgánicos volátiles consideran un conjunto de especies químicas cuya proporción depende de la fuentes de emisión y el proceso que los genera.

La **especiación** se refiere a la degradación de un contaminante en sus componentes químicos individuales o en grupos químicos específicos [Jimate, 2013]. La especiación se puede realizar a través de perfiles que representan la fracción de las especies químicas individuales presentes en un contaminante. Este proceso es necesario para representar adecuadamente las reacciones químicas que ocurren en la atmósfera dado que en la modelación fotoquímica se asigna un nivel de reactividad fotoquímica propio para cada grupo de especie. En DiETE el proceso de especiación se aplica tanto para los compuestos orgánicos volátiles (VOC) como para partículas $PM_{2.5}$ para luego agruparlos en el mecanismo solicitado por el usuario sumando las emisiones de las especies químicas de una categoría mediante un factor de agregación (fa) (Middleton et al. 1990). Durante el proceso se convierten las emisiones de masa a mol por medio del peso molecular de la categoría. En la Tabla 2.2 se presentan los mecanismos fotoquímicos implementados en DiETE y el número de clases en los que agrupan a los compuestos.

Mecanismo químico	No. clases
MOZART (Model of Ozone and Related chemical Tracers)	39
RADM (Regional Acid Deposition Model)	36
RACM (Regional Atmospheric Chemistry Mechanism)	55
SAPRC99 (Statewide Air Pollution Research Center Mechanism)	47
SAPRC07 (Statewide Air Pollution Research Center Mechanism)	49

Tabla 2.2: Número de clases en los que se agrupan los compuestos orgánicos según el mecanismo fotoquímico.

En la Figura 2.8 se presenta un ejemplo de la distribución espacial y temporal que realiza DiETE de las emisiones de monóxido de carbono para el día 11 de mayo de 2021. Como se puede apreciar en la figura, la concentración de emisiones de carbono son más intensas en la mañana.

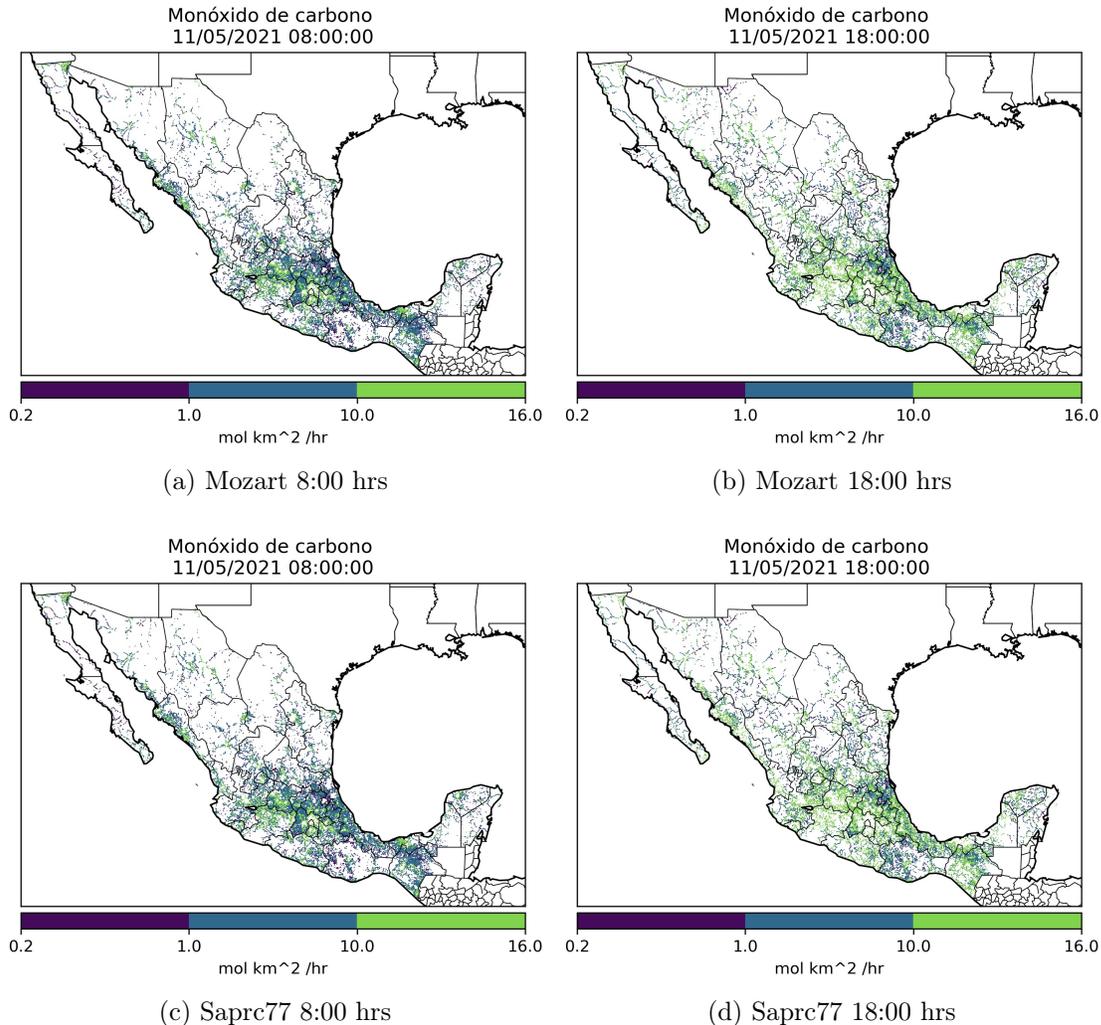


Figura 2.8: Ejemplo de las distribuciones espacial y temporal de las emisiones de DiETE

Formación de archivos de salida NetCDF

El formato de archivos Network Common Data Form (**NetCDF**) fue desarrollado por el proyecto Unidata en la Corporación Universitaria para la Investigación Atmosférica (UCAR). El formato permite almacenar variables multidimensionales y a lo largo de 30 años varias organizaciones y grupos científicos alrededor del mundo lo han adoptado como estándar para el intercambio de ciertos datos científicos. [Gaba, 2019]

Los archivos **NetCDF** está conformado por:

- **Variables:** Matrices de datos N-dimensionales que pueden ser de tipo `char`, `byte`, `short`, `int`, `float` o `double`. En NetCDF los datos unidimensionales son almacenados en arreglos, los bidimensionales en matrices y los datos de más de tres dimensiones son representados por una serie de matrices tal como se muestra en la Figura 2.9.

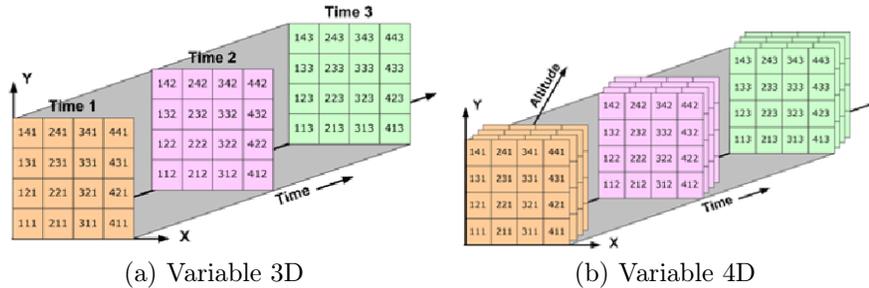


Figura 2.9: Ejemplos de como son almacenadas las variables en NetCDF

- **Dimensiones:** conformados por un nombre y una longitud, describen los ejes de las matrices de datos. Una dimensión ilimitada tiene una longitud que se puede que crecer conforme se escribe la matriz.
- **Atributos:** son valores escalares o matrices 1D que representan metadatos asociados a una variable o en el caso de atributos globales describir al archivo completo.

El producto final del DiETE corresponde a un archivo con 24 horas o dos archivos de 12 horas por cada día solicitado, en formato NetCDF, con las emisiones distribuidas en una malla de la resolución y zona de interés. En la Tabla 2.3 se presenta un resumen de las características de los archivos de salida.

Zona	Tamaño de archivo
República Mexicana	22 GB
Ecaim	2.4 GB
Bajío	1.6 GB
Colima	333 MB
Monterrey	176 MB
Mexicali	141 MB
Guadalajara	59 MB
Cd. Juárez	51 MB
Tijuana	42 MB

Tabla 2.3: Tamaño de los archivos NetCDF

En la Figura 2.10 se ilustra la interacción entre los diferentes procesos que conforman a DiETE. La comunicación entre los procesos es a través de archivos csv. Las flechas en el diagrama representan el flujo de datos entre los diferentes procesos inicia en la parte superior izquierda. Los iconos de archivos representan los archivos de configuración fijos o que se definen por el usuario antes de iniciar la ejecución del programa.

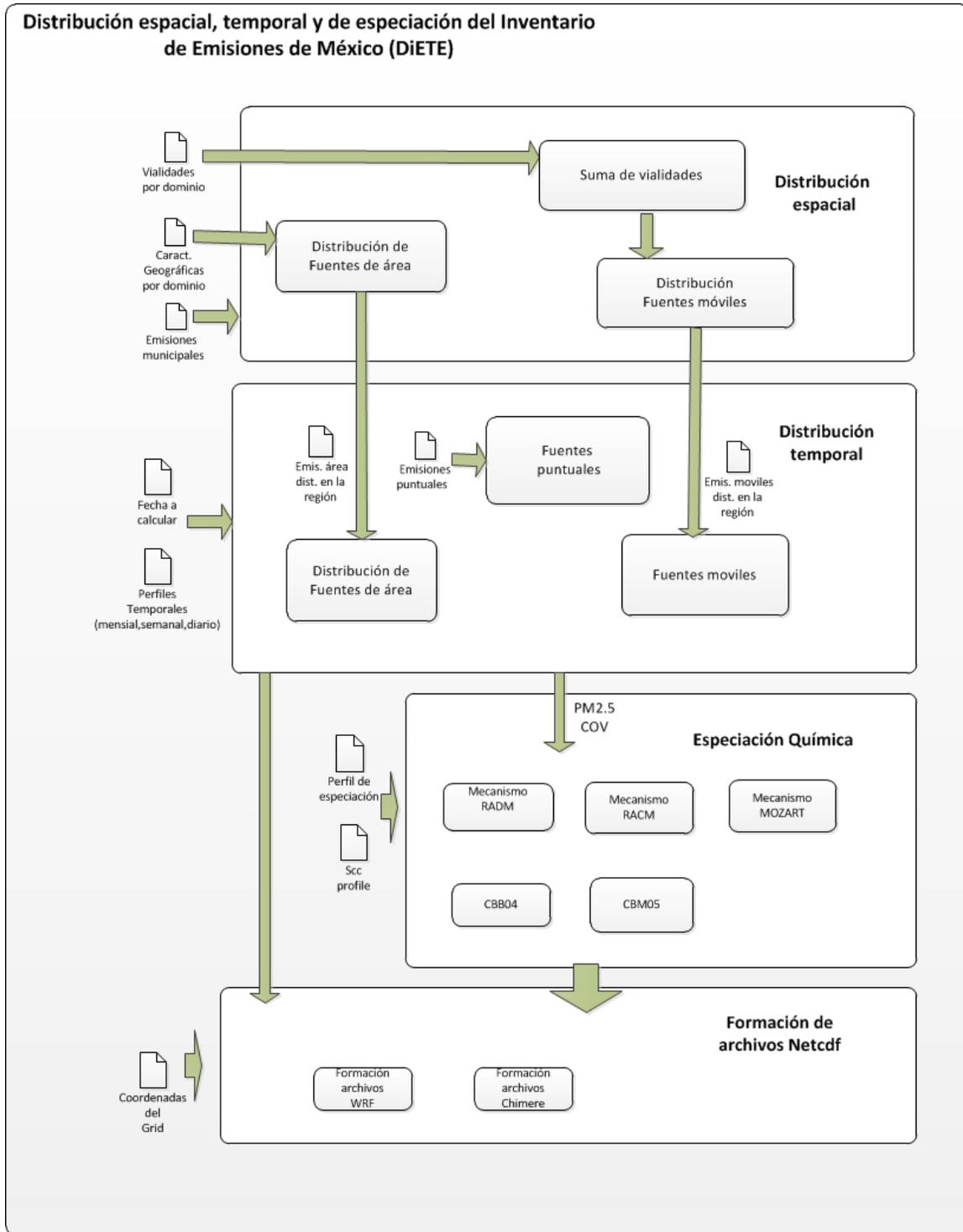


Figura 2.10: Diagrama general de los procesos del DiETE

2.4. Resumen

En este capítulo se presentó una descripción de los conceptos fundamentales en los que se encuentra enmarcado en el presente proyecto. El primer tema tratado en el capítulo fue la contaminación del aire. El segundo tema fueron los inventarios de emisiones. Finalmente se presentó una descripción de diferentes procesos que se realizan en el modelo DiETE.

Capítulo 3

Diagnóstico y Aceleración del código DiETE

En este capítulo se presenta el análisis realizado al código original de DiETE así como la descripción de los resultados obtenidos al utilizar estrategias de aceleración de código con la consiguiente reducción del tiempo ejecución.

3.1. Diagnóstico del código original

El modelo DiETE fue implementado en el lenguaje de programación Fortran. La elección del lenguaje de programación se hizo con base en tres factores:

1. La tradición del uso de este lenguaje para la implementación de modelos numéricos ampliamente usados en el área como lo son WRF y WRF-CHEM.
2. La experiencia del autor en el uso del lenguaje Fortran.
3. La expectativa de reducir el tiempo de ejecución al utilizar un lenguaje de más bajo nivel que los utilizados en los últimos años como Python.

Dada su aplicación El código fuente ha sido actualizado, siendo la última actualización publicada en octubre de 2021 que se enfocaron en hacer ajustes a los parámetros con el objetivo de mejorar los resultados en la modelación de calidad del aire. Otras modificaciones no publicadas estuvieron enfocadas en hacer el código más legible pero no hubo cambio en los algoritmos implementados por lo que no representaron un cambio sustancial los tiempos de ejecución.

El primer paso para reestructurar el código fue tomar el código original y documentar su implementación. Lo anterior debido a que, si bien el modelo está descrito en las diversas publicaciones, la documentación de la implementación se limitaba a una cabecera en los archivos de código fuente en la que se describe su función. La implementación de DiETE no consiste en un archivo con un proceso principal (**main**) en el que se llaman a los diferentes procedimientos y funciones que conforman el modelo sino en un junto de ejecutables Fortran y que se mandan a llamar mediante scripts en bash. Los archivos que conforman la implementación de DiETE se encuentran organizados en la siguiente estructura de carpetas:

- **Archivo emis_2016.sh.** Script en bash equivalente al programa principal que sirve como punto de entrada. El script contiene directivas para la ejecución de DiETE en el administrador de tareas **Slurm** y define los siguientes parámetros de ejecución:
 - **dominio.** Define el área o región para la que se calculan las emisiones. Los dominios disponibles son: bajo, bajo3, cdjuarez, colima, ecacor, ecaim, ecaim3, guadalajara, jalisco, mexicali, mexico, mexico9, monterrey, monterrey3, queretaro y tijuana
 - **HacerArea.** Puede tomar los valores 1 ó 0 para determinar si se hará o no la distribución espacial de las emisiones.
 - **MECHA.** Se usa para establecer el mecanismo de especiación química. Los valores pueden ser: `cbm04`, `cbm05`, `mozart`, `racm2`, `radm2`, `saprc99` o `saprc07`
 - **AQM_SELECT.** Si es 0 la salida es para el modelo WRF en caso de ser 1, la salida es para el modelo CHEMIRE.
 - **mes.** Especifica el mes de salida.
 - **dia.** Establece el día inicial de las emisiones calculadas.
 - **dia2.** Establece el día final de las emisiones calculadas.
 - **nyear.** Establece el año de las emisiones calculadas.
 - **nfile.** Puede tomar los valores de 1 para generar un archivo con 24 hrs. ó 2 archivos de 12 hrs. de emisiones diarias.
- **Archivo fuctions.sh.** Script bash que contiene funciones que llaman a los ejecutables Fortran correspondientes a los módulos del DiETE.
- **Carpeta 01_datos.** Contiene carpetas (`mexico`, `guadalajara`) de cada uno de los dominios disponibles.

Además, contiene las siguientes subcarpetas:

 - **dominio.** Existe una por cada dominio disponible (`mexico`, `mexico2`, `jalisco`, `mexicali`, etc.) y contiene los archivos `csv` información geográfica.
 - **chem.** Contiene los archivos de configuración, perfiles químicos, factor de escala para cada uno de los mecanismos químicos usados en el proceso de especiación descrito en la sección [2.3.2](#).
 - **emis.** Contiene las emisiones reportadas en el inventario oficial 2016. Los archivos de emisiones están organizados en subcarpetas con base en el tipo de fuente de emisión.
 - **time.** Contiene los archivos con los perfiles temporales usados en el proceso de distribución temporal, ver sección [2.3.2](#)
- **Directorio Source.** Contiene los archivos con el código fuente de las subrutinas en Fortran.
- **Directorio bin.** Contiene los ejecutables de las subrutinas en Fortran.

En la tabla 3.1 se presentan los archivos involucrados en la implementación de los diferentes procesos de DiETE.

Función bash	Ejecutable	Código fuente	Descripción
hace_area	ASpatial.exe	area_espacial.F90	Distribución espacial de emisiones de área
hace_movil	vial.exe	suma_vialidades.F90	Obtiene la fracción del área que corresponde a diferentes tipos de vialidades
	carr.exe	suma_carretera.F90	Obtiene la fracción del área que corresponde a carretera
	agrega.exe	agrega.f90	Suma las fracciones de área correspondiente a carretera y vialidades.
	MSpatial.exe	movil_spatial.F90	Distribuye espacialmente las emisiones móviles del inventario
emis_area	Atemporal.exe	atemporal.f90	Distribuye temporalmente las emisiones de área distribuidas en las celdas
	spm25a.exe	pm25_speci_a.F90	Especiación de partículas $PM_{2.5}$ para emisiones de fuentes de área
	spa.exe	agg_a.f90	Especiación de VOC para emisiones de fuentes de área
emis_fijas	Puntual.exe	atemporal.f90	Hace la distribución temporal de las emisiones fijas del inventario
	spm25p.exe	pm25_speci_p.F90	Especiación de partículas $PM_{2.5}$ para emisiones de fuentes de fijas
	spp.exe	agg_p.f90	Especiación de VOC para emisiones de fuentes fijas
emis_movil	Mtemporal.exe	atemporal.f90	Distribuye temporalmente las emisiones de móviles distribuidas en las celdas
	spm25m.exe	pm25_speci_p.F90	Especiación de partículas $PM_{2.5}$ para emisiones de fuentes de móviles
	spm.exe	agg_m.f90	Especiación de VOC para emisiones de fuentes móviles
	emiss.exe	g_emiss.F90	Genera los archivos NetCD de salida

Tabla 3.1: Distribución del código fuente del DiETE

Cada proceso está conformado por un conjunto de ejecutables (columna 2), escritos en Fortran, que se llaman desde funciones Bash (columna 1). Además, en la tabla se presentan los archivos de código fuente involucrados (columna 3) y una breve descripción (columna 4). Como se mencionó anteriormente, existe un script bash, `emis_2016`, con directivas para ejecutarse en el manejador de recursos Slurm. Tanto por el lenguaje de programación como por la forma en la que se ejecuta a través de scripts bash, la implementación de DiETE no es portable y se ejecuta sólo en sistemas basados en Unix.

Cabe aclarar que en el código analizado no tiene implementados los procesos de generación de retícula y cálculo de la fracción del área que ocupa cierta característica geográfica.

Estos procesos son realizados por un experto utilizando herramientas GIS que para generar los archivos que se encuentran en las carpetas `datos_01/dominio`.

Durante el análisis del código, el primer objetivo es determinar los factores que influyen en el tiempo de ejecución por lo que se probaron cómo influyen los diferentes parámetros de entrada.

En la Figura 3.1 se presentan gráficas de sectores que muestran el porcentaje del tiempo de ejecución que ocupan cada una de las subrutinas (descritas en la Tabla 3.1) que implementan los procesos del DiETE usando diferentes dominios y resoluciones de la retícula.

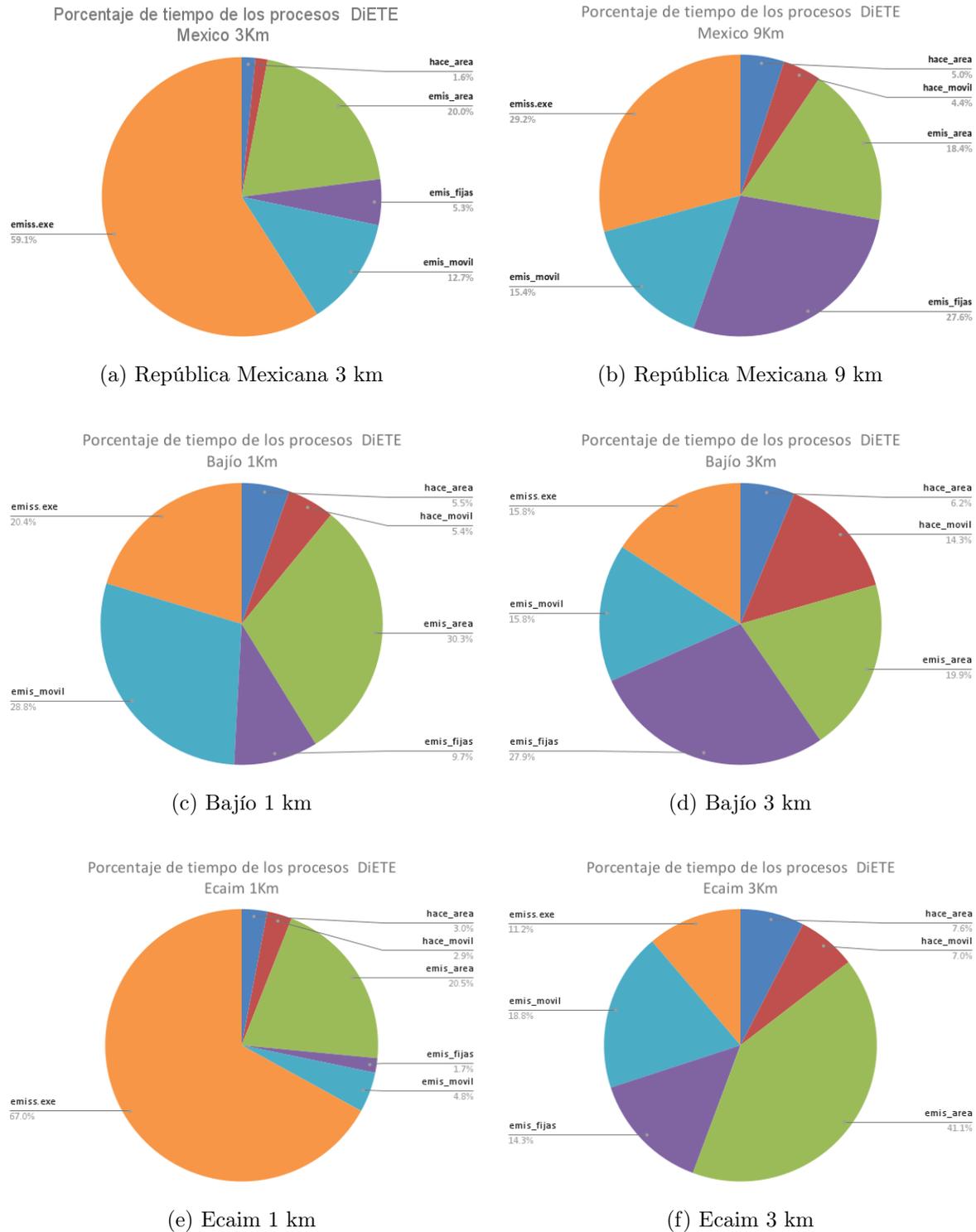


Figura 3.1: Porcentaje del tiempo total de ejecución que toman los diferentes módulos del código de original del DiETE usando diferentes dominios y resoluciones.

Para calcular estos tiempos, se ejecutó DiETE para un día de emisiones con diferentes

dominios. Los dominios seleccionados fueron los disponibles para dos diferentes resoluciones presentado los de más alta resolución la columna izquierda. Durante las pruebas se notó que los porcentajes no preservan su proporción ni entre diferentes dominios ni en dominios iguales a diferente resolución. Sin embargo, en las gráficas se aprecia que el proceso más sensible al cambio de dimensiones de las retícula es `emiss.exe` que de acuerdo a la Tabla 3.1, corresponde a la generación del archivos de salida. Esto se puede explicar debido a que DiETE no involucra cálculos complejos sino más bien cálculos sencillo aplicados a un gran número de celdas.

En la Tabla 3.2 se presentan los valores numéricos de los tiempos calculados para toda la República Mexicana. Los tiempos están expresados en segundos y fueron obtenidos al usando la función de Fortran `cpu_time()` en las regiones de interés.

<i>Proceso</i>	México 3km		México 9km	
	<i>Segundos</i>	<i>Porcentaje</i>	<i>Segundos</i>	<i>Porcentaje</i>
<code>hace area</code>	74.95 s	15.58 %	16.912 s	6 %
<code>hace móvil</code>	66.745 s	1.41 %	14.833 s	5 %
<code>emis area</code>	948.368 s	19.98 %	61.849 s	19 %
<code>emis fijas</code>	252.039 s	5.31 %	93.129 s	28 %
<code>emis movil</code>	600.424 s	12.65 %	52 s	16 %
<code>emis</code>	2,803.661 s	52 %	98.221 s	30 %
Tiempo total	4746.187 s		336.944 s	

Tabla 3.2: Tiempos de ejecución de los principales procesos del DiETE para la República Mexicana a 3km y 9 km.

Dado que el dominio correspondiente a la República Mexicana a 3 km contiene a los demás, se usó como referencia para realizar las pruebas.

En la Tabla 3.3 se presentan los tiempos de ejecución del modelo DiETE para los diferentes mecanismos de especiación química y los diferentes formatos de salida (para WRF ó CHEMIRE) que ofrece. Los resultados muestran que el mecanismo más tardado es el Sapcr77, mientras que no se encontraron diferencias en los tiempos de ejecución al cambiar el tipo de salidas.

Mecanismo	WRF	QMA
<code>Mozart</code>	4,064 s	4,071 s
<code>Sapcr77</code>	4,784 s	4,807 s
<code>Sapcr99</code>	4,770 s	4,610 s

Tabla 3.3: Tiempos de ejecución

Para entender la estructura de los archivos NETCDF de salida así como la distribución espacial y temporal se escribieron scripts en Python3 para graficar las salidas. Otro aspecto

que se probó fue averiguar las posibles diferencias entre variables usando diferentes mecanismos de especiación para lo que se realizó la resta de matrices que demostró que no había diferencia entre las variables comunes calculadas usando diferentes mecanismos de especiación.

En la Tabla 3.4 se presentan los tiempos obtenidos al calcular uno y dos días de emisiones, como se puede apreciar en la tabla, los tiempos de ejecución de los procesos encargados de realizar la distribución espacial no presentan diferencias significativas mientras que el tiempo del resto de procesos aumenta linealmente.

Subproceso	Un día	Dos días
hace_area	74.955 s	74.118 s
hace_movil	66.745 s	67.702 s
emis_area	948.638 s	1,887.262 s
emis_fijas	252.039 s	496.192 s
emis_movil	612.039 s	1,1196.794 s
emis	2803.661 s	5169.496 s

Tabla 3.4: Tiempos de ejecución de los diferentes módulos del DiETE para el cálculo de emisiones de 1 y 2 días.

Con base en los tiempos reportados, el proceso seleccionado para iniciar la optimización del código fue el módulo de formación de los archivos de salidas NetCDF. El módulo está conformado por las subrutinas:

- **lee_namelist**: Encargada de leer del archivo de configuración el mecanismo químico.
- **setup_mecha**: Se encarga de inicializar variables con base en el mecanismo químico seleccionado.
- **lee_emis**: Se encarga de leer los archivos con las emisiones calculadas en procesos anteriores y preparar las matrices que se almacenarán en el archivo Netcdf.
- **escribe_var**: Utiliza la librería Fortran NetCDF de Unidata para escribir las matrices formadas por **lee_emis**: en los archivos de salida (uno cada 12 ó 24 horas solicitadas)

En este punto del análisis, la primera impresión que se tiene es que el cuello de botella es que el proceso de escritura del archivo NETCDF (subrutina **escribe_var**) debido a que la escritura en disco representa una de las instrucciones más lentas. Sin embargo, en la Tabla 3.5 en la que se presentan los tiempos que toman cada una de las subrutinas de **emis** muestra que la subrutina más tardada es **lee_emis**.

Función	Tiempo ejecución
lee_namelsit	0.001 s
setup_mecha	0.0004 s
lee_emis	2,836.67 s
escribe_var	32.5860021 s
Tiempo total	2,869.2530 s (47.820 minutos)

Tabla 3.5: Tiempos de ejecución de las funciones que conforman el módulos *emiss*

Una vez identificada la subrutina con más tiempo de procesamiento, en la siguiente sección se presentan las modificaciones realizadas al código fuente del ejecutable *emiss.exe* que se encarga de generar los archivos de salidas NetCDF.

3.2. Estrategias de aceleración y resultados

El primer acercamiento para reestructurar el código consistió en cambiar el lenguaje de programación a Python y la representación de los datos de archivos *csv* a una base de datos relacional en MySQL y otra en base de datos orientada a documentos MogoDB. Se hicieron pruebas iniciales en una versión sin optimización en las que no se veía una ganancia en los tiempos de ejecución sin comenzar a modificar el código para su optimización.

Debido a las políticas del instituto fue necesario respetar el lenguaje original en que fue implementado el modelo DIETE, en los apartados siguientes se describen las modificaciones hechas al código en Fortran original.

En el Listado 3.1 se presenta el código original Fortran de la función `lee_emis`. La función se llama para cada una de las variables asociadas al mecanismo seleccionado por el usuario se crea una matriz en el archivo de salida en donde se suman los diferentes tipos de emisiones (área, móviles y puntuales) calculadas. El primer ciclo, en la línea 15, recorre las líneas del archivo con la distribución temporal de las emisiones. La primera línea de cada línea contiene el identificador de la celda de la retícula a la que pertenece la emisión por lo que los ciclos entre las líneas 22-33 se recorre el vector `idcg` que contiene los identificadores de celdas que pertenecen al dominio de interés para ubicar la emisión en la posición correcta dentro de la matriz `eft`.

Listing 3.1: código modificado `lee_emis`

```

1 open(newunit=iun , file=trim(ruta)//fnameA(ii) , status='OLD'
2     , action='READ')
3 read(iun , *) cdum
4 if (ii.eq.1) then
5     read(iun , *) j , current_date , cday !j number of lines in file
6     print *, current_date , '_', cday
7 else
8     read(iun , *) j , current_date

```

```

9 end if
10
11 if(ii.ge.L_PM25-1) then; is=L_PM25 ;else ;is=ii;end if
12 if(ii.eq.L_CH4_i) is=L_CH4_S
13 if(model.eq.1.and.ii.eq.1) SUPF1=SUPF1*6.022e23/1e10/3.6e3
14 constant=scala(ii)*SUPF1/WIM(is)
15 do
16   if(ii.eq.L_PM25) then
17     read(iun,*,END=100) idcf,rdum,(edum(ih),ih=1,nh)
18   else
19     read(iun,*,END=100) idcf,(edum(ih),ih=1,nh)
20   end if
21   k=0
22   busca: do j=1,ny
23     do i=1,nx
24       k=k+1
25       if(idcg(k).eq.idcf) then
26         do ih=1,nh
27           eft(i,j,1,ih)=eft(i,j,1,ih)+edum(ih)*constant
28           ! Emission from g to gmol by 1/WIM 1/(CDIM*CDIM)
29         end do
30         exit busca
31       end if
32     end do
33   end do busca
34 end do
35 100 close(iun)
36
37
38 !CODIGO SIMILAR PARA ARCHIVOS CON EMISIONES MOVILES
39
40 !COODIGO SIMILAR PARA ARCHIVOS CON EMISIONES PUNTUALES

```

3.2.1. Paralelización OPENMP

Una de las estrategias para acelerar el código de DIETE fue identificar las partes susceptibles de paralelización mediante el uso de directivas OMP del compilador. Como se observa en el listado de código 3.2, se identificó que es posible procesar en paralelo la generación de matrices de las emisiones de área, móviles y puntuales.

Listing 3.2: código modificado lee_emis

```

1 !$omp parallel sections num_threads (3) private(i,j,ih,k,idcf,
2 constant, is, iun, edum, ruta)
3

```

```

4 !$omp section
5
6     !CODIGO PARA ARCHIVOS DE EMISIONES DE AREA
7
8 !$omp section
9
10    !CODIGO PARA ARCHIVOS DE EMISIONES MOVILES
11
12 !$omp section
13
14    !CODIGO PARA ARCHIVOS DE EMISIONES PUNTUALES

```

En la Tabla 3.6 se presentan los resultados de la paralelización. El número de registros reportados en la tabla corresponde a la suma de los archivos de entrada con las emisiones temporales de las emisiones de área, móviles y temporales. Los tiempos fueron obtenidos usando las funciones para la versión serial y la función para la versión paralela. El *Speedup* o aceleración se calculó usando la ecuación 3.1

$$speedup = \frac{tiempo_algoritmo_serial}{tiempo_algoritmo_paralelo} \quad (3.1)$$

No. Registros	Algoritmo Serial	Algoritmo Paralelo	Speedup
288,575	45.74 s	22.55 s	2.03
288,648	45.83 s	22.62 s	2.03
288,712	45.79 s	22.62 s	2.02
435,802	21.38 s	10.54 s	2.03
548,057	86.48 s	40.33 s	2.14
548,411	40.87 s	17.88 s	2.29
548,583	40.88 s	17.92 s	2.28
577,153	45.79 s	22.61 s	2.03
577,287	45.83 s	22.62 s	2.03
654,850	61.41 s	31.77 s	1.93
1,306,261	152.07 s	90.36 s	1.68

Tabla 3.6: Comparativo de los tiempos de ejecución entre el código serial y paralelo de la función `lee_emis`.

En la Figura 3.2 se presenta el *Speedup* de la función `lee_emis`. Los valores de *Speedup* van de 1.68 a 3.29 utilizando 3 unidades de procesamiento o cores.

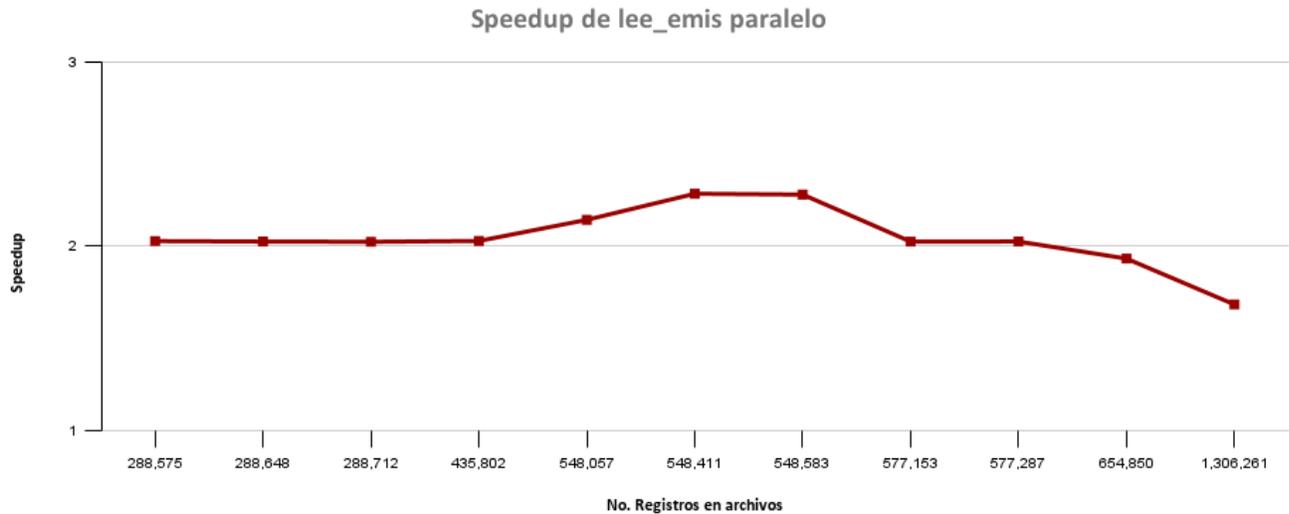


Figura 3.2: Aceleración del algoritmo paralelo OMP

3.2.2. Cambio de algoritmo

Supóngase que el dominio es de ny renglones y nx columnas y que en todas las celdas del subdominio se tienen valores de emisiones por lo tanto se tienen $ny \times nx$ registros en el archivo. El algoritmo original recorre los registros y busca su posición dentro de la matriz recorriendo uno a uno cada elemento en las filas y columnas del dominio. En el peor de los casos, el valor de que el valor se encontrará en la posición final y se tuvieron que buscar todos los elementos del dominio si tomamos $N = ny \times nx$, se tiene que las instrucciones se ejecutan $N \times N$ veces por lo que el algoritmo es de orden $O(n^2)$.

En el listado de código 3.3 se muestra la optimización del procedimiento `lee_emis` en el que se eliminaron los ciclos anidados. Entre las modificaciones se propuso la ruina `getdimension` que lee los archivos de emisiones y calcula las dimensiones de una matriz para contener el contenido completo del archivo. El arreglo `pos_A` contiene los GRIDCODE de cada uno de los registros en los archivos de emisiones. La función `get_position` toma como argumentos el GRIDCODE (`pos` en el listado) y el número de columnas de dominio (`nx` en el listado), como resultado, la función regresa el número de renglón `pren` y el número de columna `pcol` en la matriz `eft` en donde se acumulan las emisiones de cada una de las celdas de la retícula.

Listing 3.3: código modificado `lee_emis`

```

1 !      ...
2 call getdimension(filename,3,nrows,ncols)
3 print*, '*****_ ', filename, '_de dimensiones, ', nrows,ncols
4 open(unit=30,file=trim(filename),status='OLD',action='READ')
5 read(30,*)cdum !ENCABEZADO
6 read(30,*)j,current_date,cday
7 ALLOCATE(data(ncols,nrows))

```

```

8 ALLOCATE(sum_A(nh))
9 ALLOCATE(pos_A(nrows))
10 READ(30,*) data
11
12 !      ....
13 do i=1,nrows
14   pos=pos_A(i)+1
15   call get_position(pos,nx, pren, pcol)
16   if (pren<0 .or. pren>ny .or. pcol<0 .or. pcol>nx)then
17     print*, filename, 'la_posici n_: ', pren, pcol
18   endif
19   sum_A= data(i_start:ncols, i)*constant
20   eft(pcol, pren, 1,:) = eft(pcol, pren, 1,:) + sum_A
21 end do

```

En el listado 3.4 se presenta la función que mapea el Gridcode reportado en el archivo de entrada a las posiciones de renglón y columna para la matriz de salida.

Listing 3.4: código modificado *get_position*

```

1 subroutine get_position(cell, ncol, ren, col)
2   implicit none
3   integer, intent(in) :: cell, ncol
4   integer, intent(out) :: ren, col
5   ren = cell/ncol
6   col = mod(cell, ncol)
7   if (col.ne.ncol) then
8     ren=ren+1
9   end if
10 end subroutine

```

Con el cambio de algoritmo se tiene que el número de operaciones corresponden únicamente al número de emisiones en los archivos de entrada, N , por lo que el nuevo algoritmo es de orden $O(n)$.

En la Tabla 3.7 se presentan los resultados del cambio de algoritmo en `lee_emis`. El cálculo de Speedup se utilizó la ecuación 3.2.

$$speedup = \frac{tiempo_algoritmo_original}{tiempo_algoritmo_nuevo} \quad (3.2)$$

No. Registros	Algoritmo Original	Algoritmo Propuesto	Aceleración
288,575	45.74 s	1.61 s	28.34
288,648	45.83 s	1.66 s	27.59
288,712	45.79 s	1.65 s	27.71
435,802	21.38 s	0.84 s	25.39
548,057	86.48 s	3.00 s	28.82
548,411	40.87 s	1.45 s	28.09
548,583	40.88 s	1.46 s	27.99
577,153	45.79 s	1.65 s	27.67
577,287	45.83 s	1.67s	27.47
654,850	61.41 s	1.97 s	31.24
1,306,261	152.07 s	5.86s	25.96

Tabla 3.7: Comparativo de los tiempos de ejecución entre el código original y la función optimizada `lee_emis`.

En la Figura 3.3 se presenta el Speedup obtenido al cambiar el algoritmo del procedimiento `lee_emis`. Los valores de Speedup van de 25.39 a 28.82 utilizando los mismos recursos de cómputo que la implementación original.

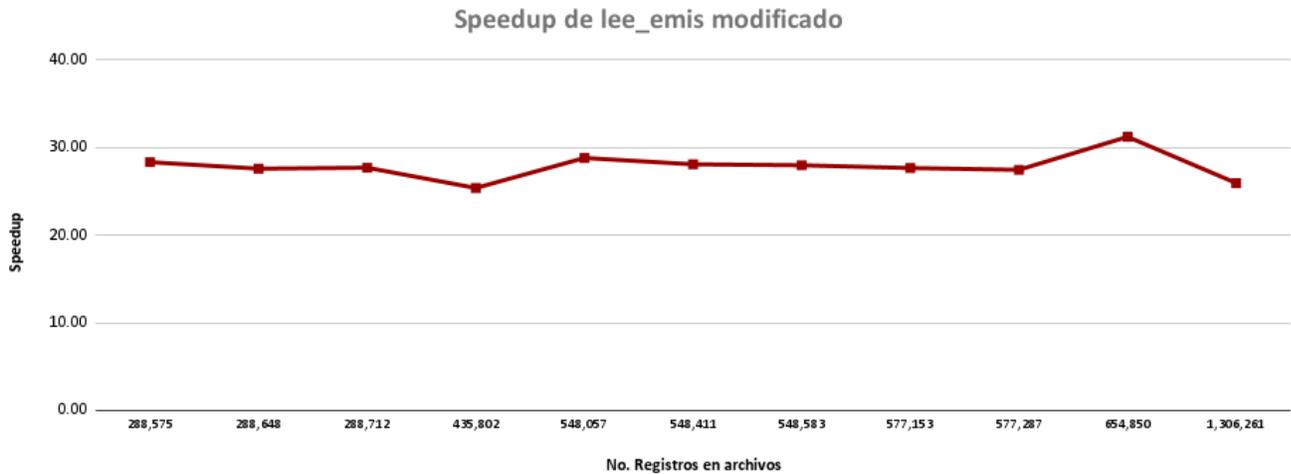


Figura 3.3: Aceleración del algoritmo propuesto

En la Figura 3.4 se presenta una gráfica comparativa entre los tiempos de ejecución de la versión serial representada por la barra verde, la ejecución del algoritmo en paralelo y algoritmo optimizado de la función `lee_emis`. Como se puede apreciar, el tiempo que se obtuvo al reducir la complejidad del algoritmo fue mucho menor que el obtenido al ejecutar el algoritmo original en paralelo.

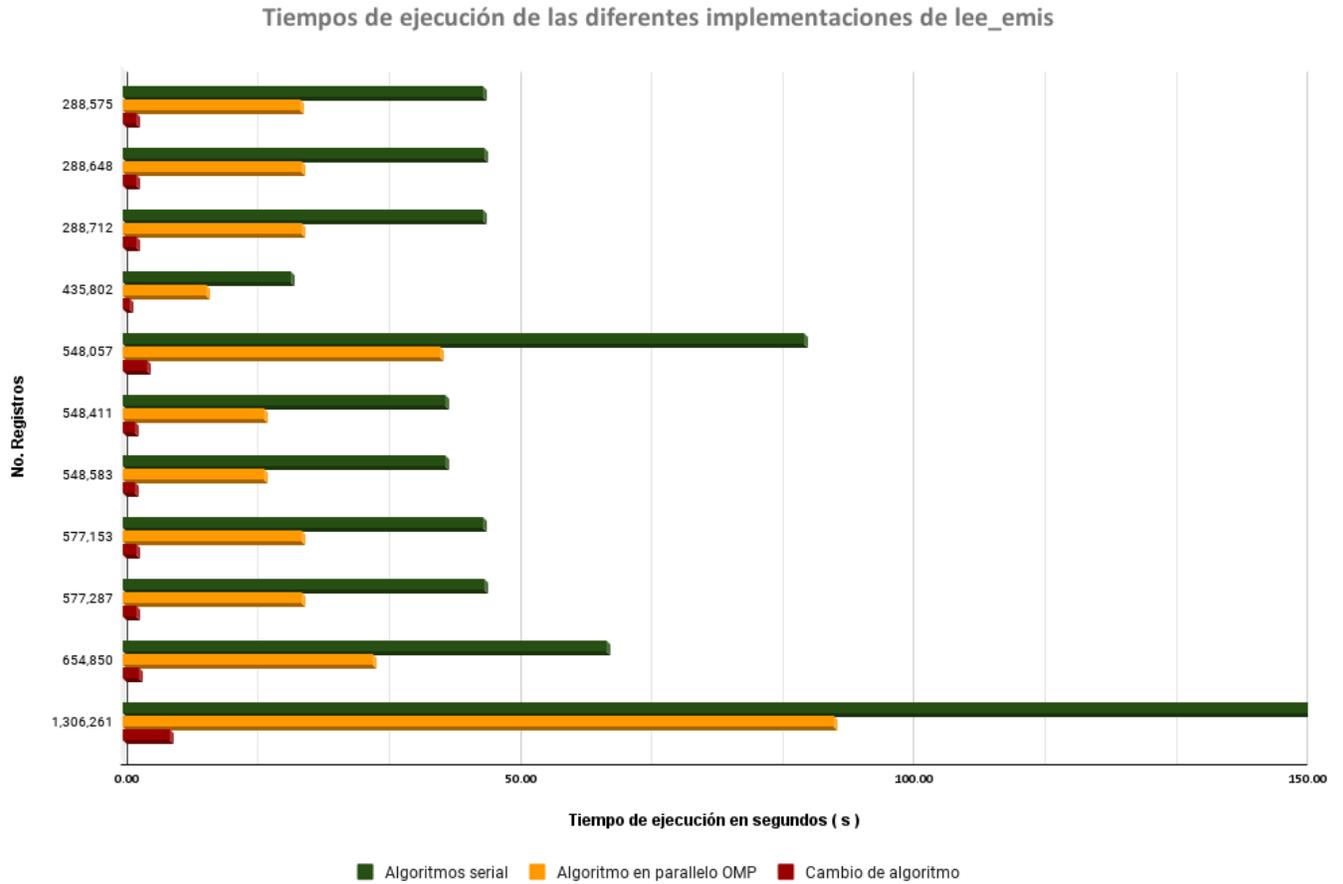


Figura 3.4: Aceleración del algoritmo propuesto

3.3. Resumen

En este capítulo se describió el análisis de los tiempos de ejecución de los procesos de la implementación original de DiETE con la que se pudo identificar el proceso y concretamente la subrutina que representa mayor tiempo de procesamiento. Se presentaron dos estrategias para optimizar el algoritmo y acelerar la ejecución del código fuente de la subrutina más tardada. La primera estrategia consistió en utilizar directivas de compilación OpenMP para procesar de manera paralela a las emisiones de área, puntuales y móviles. La segunda estrategia consistió en realizar un cambio de un algoritmo de orden $O(n^2)$ por uno de orden $O(n)$.

Capítulo 4

Conclusiones y Trabajo Futuro

4.1. Conclusiones

En este trabajo se ha realizado un análisis del código fuente del modelo DiETE con el objetivo principal de optimizarlo para reducir el tiempo de ejecución. Durante el desarrollo del proyecto se documentó el funcionamiento e implementación de cada uno de los procesos involucrados en el modelo para identificar los puntos susceptibles de optimización.

El proceso identificado como el más susceptible de optimización fue el proceso de generación de los archivos de salidas NeTCDF con la distribución de las emisiones que llega a ocupar hasta un 60% del tiempo total de ejecución del modelo. Con las modificaciones realizadas al código fuente para obtener el tiempo de ejecución a mayor detalle en una sección de interés permitió identificar que a diferencia de lo que se asumió al inicio, la subrutina de preparación (`lee_emis`) es hasta 87 veces más tardado que la subrutina encargada propiamente de escribir las matrices calculadas en los archivos de salida.

En el presente documento se presentaron dos de las estrategias de optimización que demostraron mejores resultados en la reducción en los tiempos de ejecución de la subrutina `lee_emis`. La primera estrategia consistió en identificar las áreas del código susceptibles de paralelización y mediante directivas OMP del compilador. La paralelización del algoritmo más tardado mediante directivas OMP aceleró la ejecución 2 veces si bien esta aceleración podría resultar atractiva, debe de considerarse que involucra el uso de más recursos en este caso unidades de procesamiento (cores). La segunda estrategia consistió en analizar a fondo la funcionalidad y la implementación del algoritmo para obtener el orden de complejidad y reemplazarlo con un algoritmo que realice la misma funcional con una complejidad menor. El resultado de la segunda aproximación permitió cambiar un algoritmo de orden $O(n^3)$ por uno de orden $O(n)$ con lo que se obtuvo un factor de rendimiento promedio de 28. En cuanto al tiempo global del modelo DiETE, la reducción en la complejidad del algoritmo permite que el tiempo de ejecución usando la configuración más tardada, pase de 1 hora 19 minutos a 34 minutos 37 segundos por lo que se puede afirmar que se cumplió con el objetivo general del proyecto.

Durante el estudio de código fuente se identificó que en general, se hace uso de varios ciclos anidados que podrían optimizarse utilizando la notación vectorial de Fortran. Sin embargo, al analizar la función de muchos de ese ciclos anidados se observa que se utilizan principalmente para realizar búsquedas exhaustivas sobre vectores y matrices que pueden ser sustituidas por algoritmos más eficientes. En este sentido se sugiere actualizar la implementación a un lenguaje de programación de más alto nivel que ofrecen librerías

optimizadas y en paralelo con esa funcionalidad.

Como trabajo futuro se proponen las siguientes actividades

1. Automatizar el proceso de generación para la generación de retícula.
2. Modificar el código para leer archivos de configuración que eviten recompilar el código para modificar parámetros como los utilizados en los mecanismos de especiación.
3. Cambiar los algoritmos de búsqueda exhaustiva en los módulos de distribución temporal de las emisiones.
4. Implementar el código necesario para realizar en paralelo el cálculo la distribución temporal y especiación de las emisiones en paralelo.

Bibliografía

- [Bravo-Álvarez et al., 2003] Bravo-Álvarez, H., Soto-Ayala, R., Universidad Nacional Autónoma de México, Sosa-Echeverría, R., Universidad Nacional Autónoma de México, Sánchez-Álvarez, P., and Universidad Nacional Autónoma de México (2003). Efecto de la lluvia ácida en el material constituyente de los monumentos mayas mexicanos. *Ingeniería Investigación y Tecnología*, 4(4):195–205.
- [Gaba, 2019] Gaba, C. (2019). Presentarion OF THE NETCDF FILE FORMAT AND TOOLS FOR PROCESSING AND VISUALIZATION. <http://indico.ictp.it/event/8847/session/71/contribution/246/material/slides/0.pdf>.
- [García-Reynoso et al., 2018] García-Reynoso, J. A., Mar-Morales, B. E., and Ruiz-Suárez, L. G. (2018). MODELO DE DISTRIBUCIÓN ESPACIAL, TEMPORAL Y DE ESPECIACIÓN DEL INVENTARIO DE EMISIONES DE MÉXICO (AÑO BASE 2008) PARA SU USO EN MODELIZACIÓN DE CALIDAD DEL AIRE (DiETE). *Revista Internacional de Contaminación Ambiental*, 34(4):635–649. Number: 4.
- [INECC, 2005] INECC (2005). Guía de elaboración y usos de inventarios de emisiones. Secretaría de Medio Ambiente y Recursos Naturales, Instituto Nacional de Ecología : Western Governor’s Association, México, D.F. OCLC: 639527918.
- [Jimate, 2013] Jimate, I. C. (2013). Guía metodológica para la estimación de emisiones de fuentes fijas.
- [Rodríguez Zas and García Reynoso,] Rodríguez Zas, J. and García Reynoso, J. Updating of the 2013 national emissions inventory for air quality modeling in central mexico. 37:463–487.
- [Ruiz-Suárez et al., 2018] Ruiz-Suárez, L., Mar-Morales, B., García-Reynoso, J., Andraca-Ayala, G., Torres-Jardón, R., García-Yee, J., Barrera-Huertas, H., Gavilán-García, A., and Cruz, R. (2018). Estimation of the impact of ozone on four economically important crops in the city belt of central mexico. 9(6).
- [SEDEMA, 2018] SEDEMA (2018). Firma Gobierno de la CDMX convenio con el Centro de Ciencias de la Atmósfera de la UNAM. <https://sedema.cdmx.gob.mx/comunicacion/nota/firma-gobierno-de-la-cdmx-convenio-con-el-centro-de-ciencias-de-la-atmosfera-de-la-unam>.
- [SEMARNAT, 2018] SEMARNAT (2018). Informe de la situación del medio ambiente en méxico 2018. https://apps1.semarnat.gob.mx:8443/dgeia/informe18/tema/pdf/Cap5_atmosfera.pdf.
- [Silva et al., 2017] Silva, R., West, J., Lamarque, J.-F., Shindell, D., Collins, W., Faluvegi, G., Folberth, G., Horowitz, L., Naik, V., Nagashima, T., Rumbold, S., Sudo, K., Takemura, T., Bergmann, D., Cameron-Smith, P., Doherty, R., MacKenzie, I., Stevenson,

- D., Josse, B., and Zeng, G. (2017). Future global mortality from changes in air pollution attributable to climate change. Nature Climate Change, 7(9):647–651. Publisher: Nature Publishing Group 647.
- [Vigil et al., 2013] Vigil, A. M. C., Gloria García Santiago, and Bárbara Icaza Hernández (2013). Calidad del aire: una práctica de vida. Secretaría de Medio Ambiente y Recursos Naturales, Tlalpan México, D.F.

Apéndice

Códigos de clasificación de fuentes

El código de clasificación de fuentes (Source Classification Code o SCC, por sus siglas en inglés) se utilizó para clasificar las actividades de las fuentes. En la tabla se presenta la lista de fuentes consideradas en el modelo DiETE.

Clase	Subclase	SCC	Actividad	Sustituto	Contaminantes
Procesos industriales	Productos minerales	30500302	Ladrillera	Población urbana	COV, SO ₂ , PM2.5, PM10, NO _x , CH ₄ , CO, CO ₂ , BC
	Alimentos y afines	2302002000	Asados al carbón	Población total	COV, SO ₂ , PM2.5, PM10, NO _x , CO, CO ₂ , BC
		2302050000	Panificadoras	Población total	COV
	Construcción	2311010000	Polvo de construcción	Población total	PM2.5, PM10
Fuente estacionaria combustión de combustible	industriales	2102004000	Combustión industrial de Diésel	Población total	COV, SO ₂ , PM2.5, PM10, NO _x , NH ₃ , CH ₄ , CO, CO ₂ , BC
		2102007000	Combustión industrial de Gas Licuado del Petróleo (GLP)	Población total	COV, SO ₂ , PM2.5, PM10, NO _x , CH ₄ , CO, CO ₂ , BC
		2103006000	Combustión industrial de Natural Gas (GN)	Población urbana	COV, SO ₂ , PM2.5, PM10, NO _x , NH ₃ , CH ₄ , CO, CO ₂ , BC
	Comercial	2103007000	Combustión Comercial y/o Institucional de Gas Licuado del Petróleo (GLP)	Población urbana	COV, SO ₂ , PM2.5, PM10, NO _x , NH ₃ , CH ₄ , CO, CO ₂ , BC
	Residencial	2104006000	Combustión residencial de Gas Natural (GN)	Población urbana	COV, SO ₂ , PM2.5, PM10, NO _x , NH ₃ , CH ₄ , CO, CO ₂ , BC
		2104007000	Combustión residencial de Gas Licuado del Petróleo (GLP)	Población urbana	COV, SO ₂ , PM2.5, PM10, NO _x , CH ₄ , CO, CO ₂ , BC
		2104008000	Combustión residencial de leña (cocinas de leña, chimeneas)	Población rural	COV, SO ₂ , PM2.5, PM10, NO _x , CH ₄ , CO, CO ₂ , BC
		2104011000	Combustión residencial de queroseno	Población rural	COV, SO ₂ , PM2.5, PM10, NO _x , NH ₃ , CH ₄ , CO, CO ₂ , BC

Clase	Subclase	SCC	Actividad	Sustituto	Contaminantes
Solventes	Revestimientos	2401001000	Solventes en recubrimientos arquitectónicos	Población total	COV
		2401005000	Pintado automotriz	Población total	COV
		2401008000	Solventes de los revestimientos en la señalización vial	Población total	COV
		2401020000	Recubrimientos en muebles de madera	Población total	COV
		2401065000	Disolventes en recubrimiento de superficies industriales	Población total	COV
		2401080000	Fabricación de equipo de transporte	Población total	COV
		2401990000	Industria manufacturera	Población total	COV
	Desengrasantes	2415000000	Limpieza	Población urbana	COV
		2415010000	Industrial de metales primarios	Población total	COV
	Lavado en seco	2420000055	Percloroetileno	Población urbana	COV
	Artes gráficas	2425000000	Imprenta	Población urbana	COV
		2425010000	Litografías	Población urbana	COV
		2425030000	Rotograbado	Población urbana	COV
		2425040000	Flexografía	Población urbana	COV
	Otros usos industriales	2461020000	Asfaltado	Población total	COV
		2461800000	Aplicación de pesticidas	Área agrícola	COV
2465000000		Productos en aerosol	Población total	COV	
2465100000		Productos de cuidado personal	Población total	COV	
2465200000		Productos de limpieza domésticos	Población total	COV	
2465400000		Productos de cuidado automotriz	Población urbana	COV	
2465600000		Adhesivos y selladores	Población total	COV	
2465800000		Pesticidas comerciales y residenciales	Población total	COV	
2465900000	Productos misceláneos	Población total	COV		
Almacenamiento y transporte	Almacenamiento de petróleo y productos derivados	2501060000	Gasolineras	Población total	COV
Eliminación, tratamiento y recuperación de residuos	Quemas abierta	2610000000	Quema abierta de basura	Población total	COV, SO ₂ , PM2.5, PM10, NO _x , CH ₄ , CO, CO ₂ , BC
	Tratamiento de agua	2630030000	Plantas de tratamiento de aguas	Población total	COV, CH ₄

Clase	Subclase	SCC	Actividad	Sustituto	Contaminantes
Otras fuentes de área	Producción agrícola	2801000002	Labranza	Área agrícola	PM2.5, PM10
		2801000005	Cosecha	Área agrícola	PM2.5, PM10
		2801500100	Quema de campos agrícolas	Área agrícola	COV, SO ₂ , PM2.5, PM10, NO _x , NH ₃ , CH ₄ , CO, CO ₂ , BC
		2801700000	Aplicación de fertilizantes	Área agrícola	NH ₃
		2805020000	Emisiones de desechos de ganado	Área agrícola	PM2.5, PM10, NH ₃ , CH ₄
	Combustión	2810001000	Incendios forestales	Área vegetación	COV, SO ₂ , PM2.5, PM10, NO _x , NH ₃ , CH ₄ , CO, CO ₂ , BC
		2810030000	Incendio de estructuras	Población total	COV, SO ₂ , PM2.5, PM10, NO _x , CO, BC
	Servicios de salud	2850000010	Esterilización de hospitales	Población total	COV
Distribución de Gas Licuado del Petróleo (GLP)	Distribución de Gas Licuado del Petróleo (GLP)	3333333333	Distribución de Gas Licuado del Petróleo (GLP)	Población total	COV
Amoníaco doméstico	Amoníaco doméstico	5555555555	Uso residencial (Productos de limpieza)	Población total	NH ₃
Fuentes móviles	Cruces fronterizos	2222222222	Cruces fronterizos	Población total	COV, SO ₂ , PM2.5, PM10, NO _x , NH ₃ , CO, BC
	Vehículos diésel	2230070310	Vehículos Diésel de servicio pesado	Central Camioneras	COV, SO ₂ , PM2.5, PM10, NO _x , NH ₃ , CO
		2270005000	Maquinaria agrícola	Área agrícola	COV, SO ₂ , PM2.5, PM10, NO _x , CH ₄ , CO, CO ₂ , BC
	Gas Licuado del Petróleo (GLP)	2267005000	Equipos agrícolas	Población total	COV, SO ₂ , PM2.5, PM10, NO _x , CH ₄ , CO, CO ₂ , BC
	Aeronaves	2275000000	Todos los tipos y operaciones de aeronaves	Población total	COV, SO ₂ , PM2.5, PM10, NO _x , CO, BC
		2275050000	Aviación general	Población total	COV, SO ₂ , PM2.5, PM10, NO _x , CO, BC
	Buques marinos, comerciales	2280000000	Todos los combustibles	Población total	COV, SO ₂ , PM2.5, PM10, NO _x , CO, BC
	Equipo de ferrocarril	2285000000	Locomotoras de arrastre (todos los combustibles)	Población Total	COV, SO ₂ , PM2.5, PM10, NO _x , CO, BC
		2285002010	Locomotoras de patio Diésel	Población Total	COV, SO ₂ , PM2.5, PM10, NO _x , CO
	Caminos sin pavimentar	2296000000	Polvo de camino sin pavimentar	Terracería	PM2.5, PM10
	Vehículos de gasolina	2201001000	Vehículos de servicio ligeros de gasolina	Carreteras y vialidades	CH ₄ , CN, CO, CO ₂ , GSO ₄ , NH ₃ , NO, NO ₂ , PM10, PM2.5, POA, SO ₂ , VO, otros
		2201020000	Camiones de gasolina de servicio liviano 1 y 2	Carreteras y vialidades	CH ₄ , CN, CO, CO ₂ , GSO ₄ , NH ₃ , NO, NO ₂ , PM10, PM2.5, POA, SO ₂ , VO, otros
		2201040000	Camiones de gasolina de servicio liviano 3 y 4	Carreteras y vialidades	CH ₄ , CN, CO, CO ₂ , GSO ₄ , NH ₃ , NO, NO ₂ , PM10, PM2.5, POA, SO ₂ , VO, otros
		2201070000	Terminal de camiones	Población total	CH ₄ , CN, CO, CO ₂ , GSO ₄ , NH ₃ , NO, NO ₂ , PM10, PM2.5, POA, SO ₂ , VO, otros
		2201080000	Motocicletas de gasolina	Carreteras y vialidades	CH ₄ , CN, CO, CO ₂ , GSO ₄ , NH ₃ , NO, NO ₂ , PM10, PM2.5, POA, SO ₂ , VO, otros
	Vehículos Diesel	2230001000	Vehículos ligeros de servicio diésel en carretera	Carreteras y vialidades	CH ₄ , CN, CO, CO ₂ , GSO ₄ , NH ₃ , NO, NO ₂ , PM10, PM2.5, POA, SO ₂ , VO, otros
		2230060000	Camiones diésel de servicio liviano 1 a 4 (M6)	Carreteras y vialidades	CH ₄ , CN, CO, CO ₂ , GSO ₄ , NH ₃ , NO, NO ₂ , PM10, PM2.5, POA, SO ₂ , VO, otros
		2230070000	Camiones diésel de servicio pesados	Carreteras y vialidades	CH ₄ , CN, CO, CO ₂ , GSO ₄ , NH ₃ , NO, NO ₂ , PM10, PM2.5, POA, SO ₂ , VO, otros