

# Estimación de flujos de emisión atmosférica mediante técnicas de modelado inverso y simulaciones de Montecarlo

---

## Trabajo Fin de Grado Grado en Física

### Autor

Alejandro Pujante Pérez

### Tutores

Rafael García Molina  
Departamento de Física

Antonio Jesús Jara Valera  
Director I+D en Libelium

UNIVERSIDAD DE  
MURCIA



**libelium**

Behind the change. Beyond the challenge.



**CAMPUS MARE NOSTRUM**  
de Excelencia Internacional

Este trabajo se ha realizado en el marco del «PROGRAMA ENCUENTRO TF»  
subvencionado por la Dirección General de Investigación e Universidades de la CARM



## DECLARACIÓN DE ORIGINALIDAD

D./Dña. Alejandro Pujante Pérez, estudiante del Grado en Física de la Facultad de Química de la Universidad de Murcia,

### DECLARO:

Que el Trabajo de Fin de Grado que presento para su exposición y defensa titulado y cuyo/s tutor/es es/son

D./Dña. Rafael García Molina

D./Dña. Antonio Jesús Jara Valera

es original y que todas las fuentes utilizadas para su realización han sido debidamente citadas en el mismo.

Murcia, a 21 de mayo de 202

Firma

*Genius is one percent inspiration  
and ninety-nine percent perspiration.  
Thomas Edison*



# Agradecimientos

A mi familia, mis padres y mi hermana, por no dejar que me rinda a lo largo de todo este camino y por apoyarme en cada una de las decisiones que he tomado siempre.

A Antonio Jesús Jara, por toda la confianza y apoyo que me ha mostrado durante todo este proyecto. Para mi ha sido todo un placer haber podido realizar este TFG en una empresa como Libelium.

A mi tutor, Rafael García Molina, por aceptar mi propuesta y embarcarse en un proyecto el cuál tenía un futuro bastante incierto en un principio. Gracias por la implicación durante todo el trabajo y por transmitirme toda tu experiencia.

A Eduardo Illueca, por ser mi principal apoyo moral en este trabajo y acompañarme desde el inicio hasta el final. Gracias por compartir conmigo todos tus conocimientos y por animarme a seguir en los peores momentos.



# Resumen

El estudio de la calidad del aire ha cobrado una gran importancia durante los últimos años, y evaluar el impacto que la contaminación produce a nivel social y de medio ambiente supone un gran reto científico. Por ello, se han desarrollado modelos físico-matemáticos que simulan la propagación de contaminantes en la atmósfera. Estos modelos cuentan con diversos parámetros de entrada, los cuales llevan asociada una incertidumbre que afecta a su buen rendimiento. Los flujos de emisión de contaminante en la superficie de la Tierra son uno de estos datos de entrada importantes y con gran peso en los resultados finales de las simulaciones, además, son difíciles de estimar en la gran mayoría de casos, siendo complicado establecer inventarios de emisiones precisos. Por ello, en este trabajo se calcula una estimación de los flujos de emisión de fuentes contaminantes en la atmósfera mediante técnicas de modelado inverso y simulaciones de Montecarlo, con el objetivo de que el desempeño de estos modelos de propagación se vea beneficiado. La metodología está basada en la estimación de los parámetros de entrada de un modelo a través de datos de observaciones, para ello, se generan valores de flujos de emisión aleatorios reiteradamente utilizando un mismo modelo de dispersión, en este caso, la ecuación de convección-difusión, resuelta por el método de los elementos finitos, para hacer una comparación entre los datos simulados y los observados (por ejemplo a través de sensores de calidad del aire), de forma que se escogen los parámetros de entrada que resultan de un menor error entre observación y simulación. Los resultados obtenidos muestran la validez del modelo en la estimación de dos y tres flujos de emisión de fuentes contaminantes, aproximándose de forma precisa a los valores reales -estimación obtenida es 92% del valor real con 4000 simulaciones Montecarlo-. En este estudio se realiza también un análisis de sensibilidad, permitiendo identificar los parámetros más influyentes en la desviación de las estimaciones, además de una comprobación rigurosa de la validez del método de Montecarlo. El trabajo deja abiertos nuevos objetivos y siguientes etapas en el desarrollo e investigación de este proyecto, como extender los resultados a un mayor número de fuentes.





# Abstract

The study of air quality has become increasingly important in recent years, and assessing the social and environmental impact of pollution is a major scientific challenge. For this reason, physical-mathematical models have been developed to simulate the propagation of pollutants in the atmosphere. These models have various input parameters, which are associated with uncertainties that affect their performance. The fluxes of pollutant emissions at the Earth's surface are one of these important inputs and have a large influence on the final results of the simulations. Moreover, they are difficult to estimate in the vast majority of cases, making it difficult to establish precise emission inventories. Therefore, in this work an estimation of the emission fluxes of pollutant sources in the atmosphere is calculated using inverse modelling techniques and Monte Carlo simulations, with the aim of benefiting the performance of the pollutant propagation models. The methodology is based on the estimation of the input parameters of a model through observational data, for which random emission flux values are generated repeatedly using the same dispersion model, in this case, the convection-diffusion equation, solved by the finite element method, to make a comparison between simulated and observed data (for example through air quality sensors), so that the input parameters that result from a lower error between observation and simulation are chosen. The results obtained show the validity of the model in the estimation of two and three pollutant source emission flows, accurately approaching the real values -estimation 92 % of the real value with 4000 Monte Carlo simulations-. A sensitivity analysis is also carried out in this study, allowing the identification of the most influential parameters in the deviation of the estimates and a rigorous verification of the validity of the Monte Carlo method. The work leaves open new objectives and next steps in the development and research of this project, such as extending the results to a larger number of sources or evaluating the methodology under a real case of pollution sources using air quality sensors.



# Índice

<b>1. Introducción</b>	<b>1</b>
<b>2. Metodología</b>	<b>3</b>
2.1. Ecuación de convección-difusión . . . . .	3
2.2. Resolución de la ecuación de convección-difusión . . . . .	4
2.2.1. Método de elementos finitos . . . . .	5
2.2.2. Estabilidad de la ecuación . . . . .	6
2.3. Método de Montecarlo . . . . .	8
<b>3. Resultados</b>	<b>11</b>
3.1. Caso de dos flujos de emisión . . . . .	11
3.2. Caso de tres flujos de emisión . . . . .	13
3.3. Análisis de sensibilidad . . . . .	14
3.3.1. Número $N$ de simulaciones . . . . .	14
3.3.2. Distribuciones a priori . . . . .	18
<b>4. Conclusiones</b>	<b>19</b>
<b>A. Estabilidad de la ecuación de convección-difusión</b>	<b>23</b>



# Índice de figuras

1.1. Fuentes de contaminantes. . . . .	2
2.1. Mapas de dispersión que muestran las soluciones a la ecuación de convección difusión para diferentes valores de $P_e$ . El tamaño de la fuente de emisión que se ha considerado es $L = 2r = 12\text{m}$ . . . . .	7
2.2. Esquema del problema a resolver en este TFG. Los círculos rojos denotan las fuentes de contaminación cuyos flujos se estimarán. Las cruces azules son los puntos de medición de datos de sensores de calidad del aire. . . . .	9
2.3. Esquemas del algoritmo de Montecarlo y del modelado inverso usados en este trabajo.	10
2.4. Diagrama del modelo empleado en este TFG. . . . .	10
3.1. Mapa de dispersión que muestra la solución a la ecuación de convección-difusión con fuentes de emisión en los puntos $P_1$ y $P_2$ . Los contornos en color negro muestran valores de igual concentración. . . . .	12
3.2. Representaciones 2D (a) y 3D (b) de los valores muestreados, $q_1$ y $q_2$ . En el eje $z$ el error cometido en la simulación para cada conjunto de flujos de emisión. Con el símbolo amarillo se señala el mínimo encontrado por el algoritmo. . . . .	12
3.3. Mapa de dispersión que muestra la solución a la ecuación de convección-difusión con fuentes en los puntos $P_1$ , $P_2$ y $P_3$ . Los contornos en color negro muestran valores de igual concentración. . . . .	13
3.4. Dependencia de $N$ con la desviación estándar $\sigma$ , para la fuente $q_1$ . . . . .	15
3.5. Ajuste lineal por mínimos cuadrados, con pendiente $-0.48$ . . . . .	16
3.6. Ajuste lineal por mínimos cuadrados, con pendiente $51.26$ y ordenada en el origen $0.11$ .	16
3.7. Representación de $\sigma$ en función de $N$ . Los símbolos corresponden a las simulaciones y la curva es el ajuste numérico de los datos simulados. . . . .	17
3.8. Histograma de resultados y desviaciones obtenidas en las simulaciones para la fuente $q_1$ , con $m = 1000$ . . . . .	17



# Índice de tablas

3.1.	Resultados de las estimaciones de los flujos de emisión de contaminantes para cada una de las fuentes y los parámetros $\mu$ y $\sigma$ utilizados en el muestreo de los datos iniciales.	11
3.2.	Resultados de las estimaciones con $N = 4000$ de los flujos de emisión para cada una de las fuentes, así como los parámetros $\mu$ y $\sigma$ utilizados en el muestreo de los datos iniciales. . . . .	13
3.3.	Parámetros estadísticos recogidos para la fuente $q_1$ (a) y fuente $q_2$ (b) con variaciones en el número $N$ de simulaciones Montecarlo. . . . .	15





## Introducción

En los últimos años, el control de la calidad del aire es un reto científico y social debido a la creciente concienciación sobre los efectos que la contaminación produce en la salud humana y en el medio ambiente, afectando de forma directa a millones de personas. Para abordar este problema, durante la última década se han desarrollado modelos físico-químicos cada vez más complejos, con el fin de estudiar el transporte de contaminantes, predecir eventos de contaminación que mitiguen el impacto de estos o ayudar a las autoridades y/o entidades privadas en la gestión y toma de medidas preventivas.

Estos modelos involucran sistemas de ecuaciones diferenciales, teniendo en cuenta multitud de variables, como reacciones químicas, emisiones, tráfico de vehículos o deposición (proceso mediante el cual los contaminantes suspendidos en el aire, como partículas y gases, se depositan en la superficie terrestre o en partículas de agua). Como es normal, pretenden ser lo más realistas posible en sus predicciones y, por ello, es necesario cuantificar la incertidumbre relacionada con las parametrizaciones del modelo y los datos de entrada. En lo que respecta a los modelos, constituyen aspectos fundamentales la descripción de procesos físicos, contenidos en ecuaciones o condiciones de contorno, la formulación del mecanismo químico y/o físico, así como la resolución espacial del dominio analizado. Por otro lado, existen incertidumbres asociadas a los datos de entrada, tales como los parámetros meteorológicos, los flujos de actínidos, los flujos de velocidad, y, por último, pero no menos importante, las emisiones de contaminantes.

Numerosos estudios concluyen que las emisiones procedentes de la superficie son parámetros cruciales, además de ser regularmente inciertos, para los que los resultados finales de concentración son extremadamente sensibles [1, 2]. La incertidumbre en estos datos de entrada es siempre bastante grande y difícil de estimar, además de que hay que tener en cuenta la variabilidad temporal a escala estacional, semanal o diaria. Generalmente, los inventarios de emisiones utilizados para la simulación de los modelos se elaboran siguiendo una metodología ascendente, lo que quiere decir que diferentes organismos de distintos sectores trabajan conjuntamente para elaborar estos inventarios, en los que comúnmente se termina asociando un flujo de emisión por parcela de espacio o celdilla del dominio. La referencia muestra el inventario de CAMS, realizado por el programa europeo *Copernicus* [3].

Queda de manifiesto, entonces, el interés y la importancia en los datos de entrada a la hora de hacer uso de los diferentes modelos, pues tienen gran peso en los resultados y la calidad de las simulaciones.

No obstante, existen métodos para evaluar y mejorar los inventarios de emisiones, y es en este contexto donde aparece la modelización inversa, una técnica utilizada para determinar las condiciones o parámetros desconocidos de un sistema a partir de observaciones o mediciones realizadas en el mismo. En términos generales, se trata de invertir un modelo físico-matemático para encontrar los valores de entrada que produjeron unos resultados determinados, minimizando una función de coste que describe la diferencia entre un conjunto de observaciones y las simulaciones correspondientes. Para estimar la incertidumbre, en las emisiones se utiliza un análisis de Montecarlo, que consiste en realizar un gran número de simulaciones sucesivas con un mismo modelo de dispersión, pero con un conjunto distinto de datos de entrada muestreados a partir de distribuciones de probabilidad. Se considera que la simulación que mejor se ajusta a las observaciones es la que ha producido los valores observados y, por tanto, los parámetros de entrada son los estimados, que en este caso son los valores de emisiones de contaminantes en diferentes ubicaciones.

Es, por tanto, objeto de este trabajo explorar el uso de la técnica de modelación inversa y simulaciones de Montecarlo para estimar flujos de emisión de contaminantes (figura 1.1). Para ello, se utilizará como modelo de dispersión la ecuación de convección-difusión, resuelta a través de métodos numéricos de elementos finitos en *SfePy* [4], un software ejecutado a través del lenguaje de programación Python para resolver sistemas de ecuaciones diferenciales en derivadas parciales en 1D, 2D y 3D mediante técnicas numéricas que resuelven la ecuación de convección-difusión, que incorporan prácticamente la totalidad de los modelos de dispersión y propagación de contaminantes.



Figura 1.1: Fuentes de contaminantes.

Para poner a prueba esta técnica, se ha desarrollado un análisis de la estimación de flujos de emisión bajo escenarios de contaminación generados por la resolución de la ecuación y un pequeño número de puntos emisores de contaminante que nos permitan conocer el alcance de la metodología que se lleva a cabo. Es decir, los datos de concentración observados serán originados a partir de otras simulaciones, de tal modo que se conozcan los valores de emisión reales y la metodología pueda ser validada. Estas simulaciones reciben como dato de entrada flujos de emisión conocidos para cada una de las fuentes y así validar el modelo en diferentes casos, tales como el número de simulaciones de Montecarlo o la desviación de las estimaciones respecto de los valores reales. Para ello, este TFG está estructurado como sigue. Un apartado de metodología, donde se exponen las técnicas empleadas para resolver el problema, como la descripción de la ecuación de difusión-convección y sus términos, el software utilizado *SfePy* en la resolución de la ecuación diferencial así como los problemas de inestabilidad que presenta bajos ciertas condiciones, y cómo se ha planteado el algoritmo Montecarlo que estima los flujos de emisión. A continuación, se hace un análisis de los resultados obtenidos en el caso de dos y tres flujos de emisión, al igual que un análisis de sensibilidad para predecir qué parámetros son los que más afectan al modelo y cómo influyen en el error de los resultados. Finalmente, se presentan las conclusiones de este estudio.

## Metodología

Las técnicas utilizadas en este trabajo para abordar el problema de la estimación de flujos de emisión de contaminantes están basadas en tres partes bien diferenciadas. Por un lado, la elección y, resolución de una ecuación diferencial que describa la dispersión y el movimiento de los contaminantes en el espacio, pues se requiere de un modelo o conjunto de ecuaciones capaces de proporcionarnos unos resultados de concentración sobre la superficie. Por otra parte, se requiere un software que resuelva esta ecuación, dadas unas condiciones iniciales y de contorno. Por último, es necesario un algoritmo iterativo de Montecarlo que minimice una función de coste, dados unos valores iniciales del modelo, para obtener como salida los datos que producen la mayor similitud con las observaciones. Aplicaremos la metodología expuesta anteriormente para estimar los valores de flujos de emisión de contaminantes en el caso de dos y tres flujos de emisión.

### 2.1. Ecuación de convección-difusión

Calcular la dispersión de contaminantes en la atmósfera es complejo debido a la inestabilidad atmosférica y la gran cantidad de procesos físicos y químicos involucrados, aunque es posible obtener una buena aproximación a través de la ecuación de convección-difusión [5]. Esta ecuación es una de las ecuaciones en derivadas parciales usada con profusión en ciencia e ingeniería y es la base común de todos los modelos de transporte y de CFD (Computational Fluid Dynamics) que resuelven las ecuaciones de Navier-Stokes. Esta ecuación describe cómo se transporta en el tiempo y en el espacio una cantidad escalar  $c$  dentro de un fluido, teniendo en cuenta efectos de convección y difusión,

$$\frac{\partial c}{\partial t} = \nabla \cdot (D\nabla c) - \nabla \cdot (\vec{v}c) + R \quad , \quad (2.1)$$

donde  $c(x, y, z)$  es la variable escalar de interés,  $D$  es el coeficiente de difusión del medio, que tomaremos como constante en el tiempo y el espacio,  $\vec{v}$  es un campo vectorial con el que la cantidad  $c$  se mueve y  $R$  describe fuentes o sumideros (dependiendo su signo).

El término  $\nabla \cdot (D\nabla c)$  describe la difusión del material y puede obtenerse combinando la ecuación de continuidad, ec. (2.2) y la primera ley de Fick, ec. (2.3). La primera establece que un cambio en la densidad en cualquier parte del sistema se debe a la entrada y salida de material hacia y desde esa

parte del sistema; Es decir, no se crea ni se destruye materia. Por otro lado, la primera ley de Fick sostiene que el flujo va desde una región de alta concentración a las regiones de baja concentración, con una magnitud que es proporcional al gradiente de concentración:

$$\frac{\partial c}{\partial t} = -\nabla \cdot \vec{\Gamma} \quad , \quad (2.2)$$

$$\vec{\Gamma} = -D\nabla c \quad , \quad (2.3)$$

Sustituyendo la ec. (2.3) en la ec. (2.2) obtenemos el término de difusión  $D\nabla^2 c$ .

La contribución  $-\nabla \cdot (\vec{v}c)$  se denomina convección o advección. Se deriva de la ley de conservación de la masa y representa el flujo de masa del contaminante debido al movimiento del fluido; su efecto es aumentar la tasa de transporte del contaminante a través del medio. Si la velocidad del flujo es alta, la convección puede dominar sobre la difusión, lo que significa que el transporte del contaminante se debe principalmente al movimiento del fluido.

Aunque la ec. (2.1) es una ecuación simplificada, sigue siendo muy útil para modelar una amplia gama de procesos en áreas como la dinámica de fluidos, la química atmosférica, la biología y la ingeniería ambiental. Una de las principales ventajas de utilizar la ecuación de convección-difusión como modelo de transporte es su rapidez de cómputo mediante métodos numéricos, y aunque existen ecuaciones más complejas que pueden proporcionar resultados más precisos, la resolución de estas ecuaciones puede ser mucho más lenta y requerir recursos computacionales más potentes. Son estos los motivos que han llevado a elegir esta ecuación como solución al transporte de contaminantes y cuya solución será expuesta en el siguiente apartado.

## 2.2. Resolución de la ecuación de convección-difusión

Actualmente existen procedimientos mucho más sofisticados y avanzados para el modelado del transporte de contaminantes en el aire, como CHIMERE [6], CALPUFF [7] o WRF-Chem [8] y otros modelos de complejidad similar. Estos modelos utilizan una amplia gama de herramientas y técnicas numéricas para simular la dispersión de contaminantes en la atmósfera, basándose en la dinámica de fluidos computacional, y/o incluyendo procesos químicos avanzados que tienen en cuenta todo tipo de reacciones en la atmósfera. A pesar de que estos modelos más complejos tienen la capacidad de proporcionar resultados más precisos y detallados, su coste computacional es realmente elevado comparado con otras técnicas de solución numérica, y en una metodología donde se requiere la repetición sucesiva de simulaciones lo hace aún menos adaptable. Es por esto que, tras barajar el uso de alguno de estos modelos, y aunque en paralelo a la realización este trabajo se está explorando su uso a través de técnicas de computación cuántica que permitan una ejecución exponencialmente más rápida [9], en este trabajo se ha optado por resolver la ecuación de convección-difusión a través de métodos numéricos basados en elementos finitos.

### 2.2.1. Método de elementos finitos

El método de los elementos finitos (MEF) [10] es una técnica que se utiliza para obtener una solución numérica aproximada de problemas en medios continuos. El MEF divide el cuerpo, estructura o dominio en subdominios llamados 'elementos finitos'. Cada elemento finito tiene nodos, que se utilizan para discretizar el dominio y relacionar el valor de las variables incógnitas definidas en cada nodo. Estas relaciones se escriben en forma de sistema de ecuaciones lineales y se resuelven para obtener la solución aproximada del caso estudiado, convirtiendo un problema de carácter diferencial en otro de carácter variacional o algebraico, haciendo uso de la formulación débil de la ecuación de convección-difusión. El MEF es muy utilizado debido a su generalidad y capacidad de adaptarse a problemas complejos en varias dimensiones, incluyendo problemas de mecánica de sólidos deformables, transmisión de calor, mecánica de fluidos y campo electromagnético. Para la puesta en práctica del método hacen falta, por un lado, la discretización del dominio, que se puede hacer con diferentes programas de generación de mallados, aunque en este caso se ha utilizado *gmsh* [11], unas condiciones de contorno del dominio donde se resuelve la ecuación, las cuales han sido escogidas haciendo  $c = 0$  en  $\partial\Omega$ , donde  $\Omega$  es la frontera del mallado. La razón de imponer esas condiciones de contorno es que suponemos que el contaminante tenderá a ser 0 en los puntos suficientemente alejados de la fuente. Por último, es necesario un software que calcule la solución a la ecuación diferencial que se propone sobre este mallado, que en nuestro caso ha consistido en utilizar *SfePy*.

*SfePy* es un software especialmente preparado para resolver sistemas de ecuaciones diferenciales en derivadas parciales (EDP) acopladas, por el método de los elementos finitos (MEF) en 1D, 2D y 3D, es decir, calcula la solución en el mallado de elementos finitos. Está escrito en su mayor parte en Python (85%), aunque también utiliza lenguajes como Fortran o C. El uso de *SfePy* en este proyecto se debe a su sencillez y rapidez de cómputo. Se ha considerado una buena opción para obtener la solución del problema que se plantea en el trabajo, puesto que este se centra en desarrollar una metodología basada en el modelado inverso y análisis de Montecarlo para estimar flujos de emisión. Además, *SfePy* ofrece gran flexibilidad en términos de personalización y adaptabilidad. Es una herramienta de código abierto y modular que permite al usuario personalizar y modificar fácilmente el software para adaptarse a las necesidades específicas del proyecto. También es más fácil de usar para aquellos que ya están familiarizados con el lenguaje Python, lo que reduce la curva de aprendizaje y el tiempo de implementación. *SfePy* no cuenta con ningún tipo de interfaz gráfica y todo el modelado del problema se hace a través de archivos de configuración y scripts de Python, para ser ejecutado desde la terminal de un sistema operativo Linux. El procesamiento de los datos de salida de *SfePy* se realiza en archivos de formato VTK (Visualization Toolkit Legacy. Kitware), y tanto su procesamiento como la automatización de la ejecución del modelo se ha implementado mediante los lenguajes de programación Python y Bash (Bourne-Again Shell).

### 2.2.2. Estabilidad de la ecuación

Uno de los problemas que surgen al resolver ecuaciones en derivadas parciales por métodos numéricos es la estabilidad de la solución. El estudio de la estabilidad de la ecuación diferencial que se quiere resolver es importante, pues es posible que diferentes condiciones iniciales, condiciones de contorno o parámetros del problema den soluciones que no se corresponden con el sistema físico que se pretende simular. Es por eso que, en esta sección, se estudiará bajo qué condiciones la ecuación de convección-difusión es estable.

Dada la naturaleza del problema que se quiere resolver, el término de convección resulta ser dominante, ya que el movimiento de los contaminantes en la atmósfera se debe, en gran parte, debido a la velocidad del viento. La estabilidad de la ecuación de convección-difusión radica en qué término es el dominante: el término de convección o el término de difusión.

El número de Péclet  $P_e$  es un cociente adimensional entre el término de convección y el término de difusión [12],

$$P_e = \frac{vL}{D} \quad , \quad (2.4)$$

donde  $v$  es el módulo de la velocidad del viento,  $D$  es el coeficiente de difusión y  $L$  es una longitud característica del sistema estudiado; cuyo significado se detalla más adelante. El número de Péclet se define como el producto del número de Reynolds  $R_e = \frac{vL}{\mu}$  y el número de Schmidt  $R_c = \frac{\mu}{D}$ , donde  $\mu$  es la viscosidad del fluido y  $L$  es el tamaño de la fuente que se está moviendo.

Si  $P_e > 1$ , el problema se denomina de convección dominante, lo que quiere decir que el transporte de contaminantes ocurre mayoritariamente por movimiento del viento y la difusión pasa a un segundo plano. Los trabajos realizados en [13, 14] muestran que para números de Péclet  $P_e \gg 1$ , la solución puede presentar problemas de estabilidad y de oscilaciones. La causa de la aparición de estas fluctuaciones se puede demostrar fácilmente al comprobar el comportamiento de ciertos parámetros en la solución discretizada que emplea el método numérico (apéndice A), y se debe principalmente a la formulación del problema en términos de la ecuación de Fick como ecuación constitutiva. Es objetivo de este apartado comprobar cuáles son los valores máximos de  $P_e$  para los que encontramos estabilidad de la ecuación de convección-difusión por el método de elementos finitos, bajo ciertos parámetros de mallado y de tamaño de la fuente.

Para ello se ha resuelto la ecuación en estado estacionario ( $\partial c/\partial t = 0$ ) para diversos valores de  $P_e$ . El dominio de estudio es un cuadrado de  $200 \times 200$  m, con un mallado que contiene 11827 nodos y con una fuente de emisión que posee una sección circular de  $r = 6$  m. La elección de la longitud del radio se ha elegido de forma que se tenga una buena visualización de los resultados en el mapa de dispersión, aunque estemos considerando en este ejemplo una dimensión exagerada comparada con la realidad, solamente es con un fin ilustrativo. En lo que sigue, en este TFG se han considerado siempre soluciones con un número de Péclet que de estabilidad a la ecuación.

La figura 2.1 muestra la resolución de la ecuación para varios valores de  $P_e$ , comenzando desde valores próximos a 1 hasta valores mucho más grandes, encontrando así el límite estable de la ecuación.

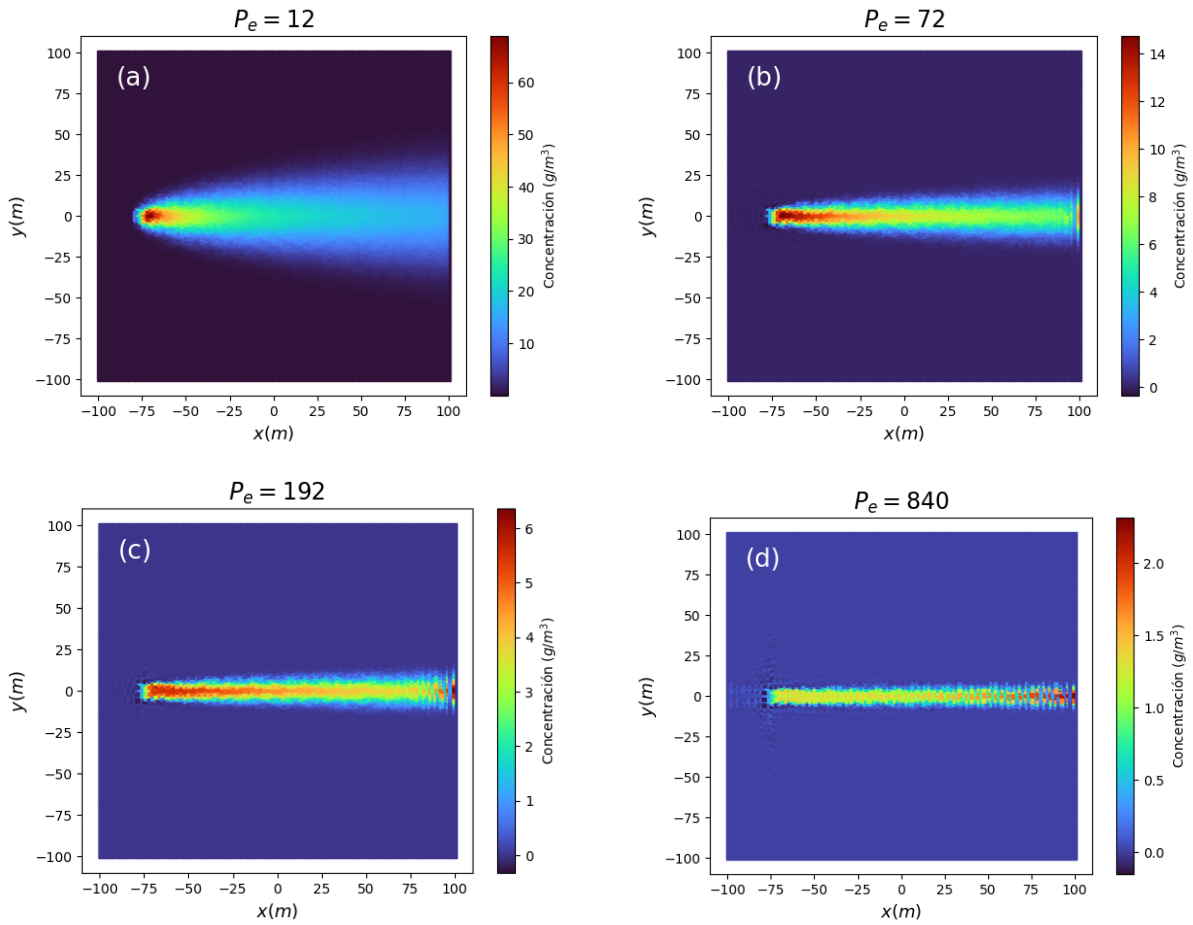


Figura 2.1: Mapas de dispersión que muestran las soluciones a la ecuación de convección difusión para diferentes valores de  $P_e$ . El tamaño de la fuente de emisión que se ha considerado es  $L = 2r = 12\text{m}$ .

En el caso de convección dominante con número de Péclet igual a 12 (figura 2.1 (a)) la solución parece precisa y no oscila en ningún momento. Si aumentamos  $P_e$  hasta rangos de 70-100 (figura 2.1 (b)), se puede observar que la solución presenta algunas oscilaciones a una distancia de 175m de la fuente. Para valores más altos de  $P_e$  (figuras 2.1(c) y 2.1(d)), deja de tener significado físico para valores muy próximos a la fuente de emisión y no es posible resolverla bajo estas condiciones. Analizar este resultado es importante porque nos permite explorar los límites del método numérico con el que estamos resolviendo la ecuación de convección-difusión y bajo qué casos es aplicable o no. Por ejemplo, si el tamaño de la fuente es constante y tuviésemos coeficientes de difusión del material muy bajos, como por ejemplo el caso del  $CO_2$  ( $0.16 \text{ cm}^2/\text{s}$ ), implicaría que la velocidad del viento tendría que ser también muy baja y reducida para que la solución sea exacta. Existen diversas soluciones para solventar este tipo de problemas, como formulaciones alternativas para discretización de la solución por elementos finitos, la agregación de nuevos términos a la ecuación de convección-difusión que la hagan estable o utilizar otro tipo de ecuaciones que sustituyan a la ecuación de Fick, como la propuesta por la referencia [15].

## 2.3. Método de Montecarlo

El método Montecarlo [16] es una técnica de simulación numérica que se utiliza para resolver problemas mediante el uso de variables elegidas aleatoriamente de acuerdo con determinadas distribuciones de probabilidad, las cuales dan cuenta de los fenómenos que se desea estudiar y en los que el comportamiento del sistema es incierto o complejo. A diferencia de los métodos analíticos tradicionales que se basan en fórmulas matemáticas, el método de Montecarlo se basa en la generación de números aleatorios para simular el comportamiento del sistema y obtener una estimación estadística del resultado deseado. La idea del método de Montecarlo reside en el principio de que, si se realizan suficientes simulaciones aleatorias, los resultados de estas simulaciones se aproximarán a una distribución de probabilidad realista. Esto significa que cuanto más grande sea el número de simulaciones que se realicen, más precisas serán las estimaciones obtenidas. Por lo tanto, este método es particularmente útil para modelar sistemas complejos y no lineales, para los cuales no existen fórmulas matemáticas exactas que puedan describir su comportamiento.

Generalmente, el proceso es simple: se elabora un programa que utiliza variables aleatorias como valores de entrada para resolver el problema deseado, teniendo en cuenta las distribuciones de probabilidad de cada variable que interviene en los cálculos. Luego se repite el algoritmo un número determinado de veces. Mientras más se repita, mayor será la precisión de los resultados obtenidos. Además, cada evaluación será independiente de las otras. Al final, se toma el valor medio de todas las evaluaciones realizadas para obtener el resultado final. En la práctica, el método de Montecarlo se aplica en una amplia gama de campos, incluyendo la física, la química, la ingeniería, las finanzas, la medicina y muchas otras áreas.

Enfocando esto al problema que nos ocupa, implementaremos el método de Montecarlo para estimar los flujos de emisión de diferentes fuentes contaminantes a través de la generación aleatoria de flujos de emisión como datos de entrada y aplicando técnicas de modelización inversa, que permiten directa o indirectamente incorporar información observada (en este caso datos de concentración monitorizados a través de sensores de calidad del aire). Se comparará en cada iteración el valor simulado con el observado para, tras varias, simulaciones escoger los flujos de emisión que den lugar a un menor error respecto a los valores observados de las concentraciones.

Para comenzar el proceso iterativo, se deberá conocer la ubicación de los puntos emisores de contaminante y de los captadores de datos de concentración (sensores), pues estas ubicaciones son constantes a lo largo de todo el algoritmo. En cada una de las simulaciones se genera un flujo de emisión para cada una de las fuentes de contaminación, elegido a través de una variable aleatoria que sigue una distribución de probabilidad normal, con media y desviación estándar determinadas mediante un análisis previo de los datos de campo o a través de inventarios de emisiones previos, que dependerán de las condiciones específicas del caso de estudio, como el tipo de gas o el carácter de la fuente (industria, naturales...). El uso de esta distribución de probabilidad para muestrear los datos de entrada no tiene ninguna preferencia respecto de otras, podría hacerse uso de una distribución



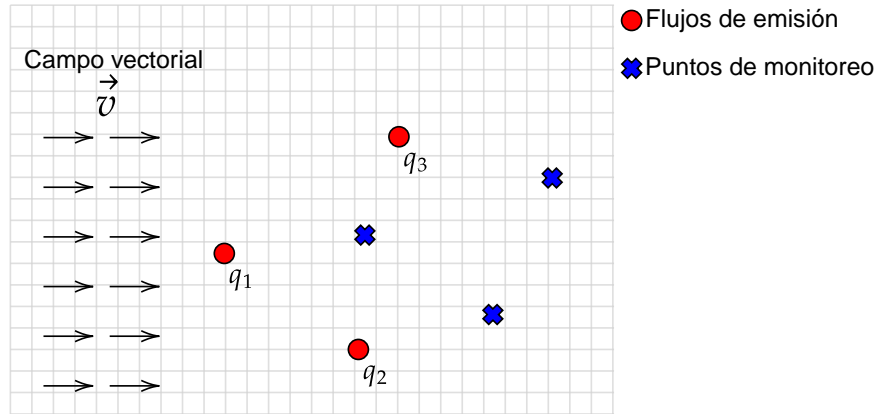


Figura 2.2: Esquema del problema a resolver en este TFG. Los círculos rojos denotan las fuentes de contaminación cuyos flujos se estimarán. Las cruces azules son los puntos de medición de datos de sensores de calidad del aire.

uniforme o log-normal como en el trabajo realizado en [1]. En cada simulación se genera un conjunto de  $k$  flujos de emisión aleatorios, siendo  $k$  el número de fuentes del sistema con el que el modelo calcula la dispersión de contaminantes. A cada uno de estos flujos aleatorios los llamaremos  $q_1, q_2, \dots, q_k$ . En cada una de las  $N$  simulaciones que se realizan, el algoritmo calcula una función de coste. En este caso, esta función corresponde al error cometido entre la concentración producida por la simulación y la concentración recogida por el sensor. Este error, que se denota como  $\epsilon$ , se calcula como la suma de las diferencias al cuadrado entre el valor simulado  $s_i$  y el observado  $o_i$  en cada uno de los puntos de observación, según la ec. (2.5). Al finalizar, se habrán generado  $N$  conjuntos de los  $k$  flujos de emisión como representa la tabla de la figura (2.3)

$$\epsilon = \sum_{i=1}^l (o_i - s_i)^2 \quad , \quad (2.5)$$

siendo  $l$  el número de sensores,  $o_i$  el dato real de concentración recogido por el sensor en la ubicación  $i$  y, por último,  $s_i$  es el dato de concentración obtenido de la simulación en la ubicación del sensor  $i$ . Tras cada una de las simulaciones se guarda el error cometido y el conjunto de flujos de emisión generados aleatoriamente para cada una de las fuentes. Al final de las iteraciones, el algoritmo escoge los valores de flujo de emisión que se han generado en la simulación que minimiza el error de la función de coste  $\epsilon$ . Tras finalizar las  $N$  simulaciones obtendremos la solución a la estimación de cada uno de los  $k$  flujos de emisión. Si para un determinado valor de  $N$  repetimos el algoritmo  $m$  veces, podremos recabar una lista de valores obtenidos para cada uno de estos flujos, es decir, se puede hacer un análisis estadístico de los resultados tras las  $m$  simulaciones hallando parámetros como media y desviación o visualizar en un histograma la distribución de resultados que se obtienen para cada una de las fuentes.

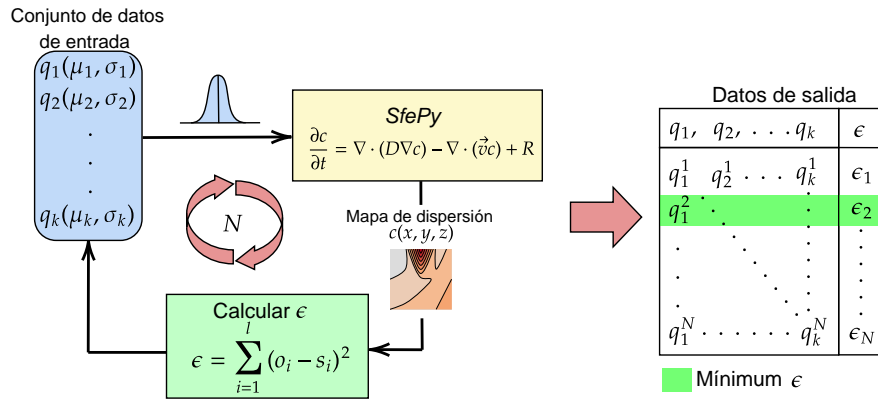


Figura 2.3: Esquemas del algoritmo de Montecarlo y del modelado inverso usados en este trabajo.

En este trabajo, para validar el algoritmo y comprobar de forma exacta cuál es la precisión con el que se obtienen los resultados, los datos de las observaciones, que corresponden a los datos reales de concentración  $o_i$ , no se han recolectado de sensores de calidad del aire reales, sino que se han sido tomados de una propia simulación de *SfePy* resolviendo la ecuación de convección-difusión, introduciendo como datos de entrada unos flujos de emisión conocidos y que posteriormente se estimarán con el algoritmo Montecarlo. En otras palabras, se crea un escenario de contaminación con *SfePy*, sobre el que medimos valores de concentración en los puntos de observación producidos por unos flujos de emisión conocidos. Ahora, con estos datos observados comienza el algoritmo de Montecarlo para estimar los flujos que hemos introducido previamente. Con esto conseguimos saber el valor real del flujo que se está estimando, pues los datos de concentración observados provienen de una simulación en la que estos flujos se conocen y por tanto podemos saber el grado de exactitud con el que el algoritmo consigue la estimación. Las distribuciones de probabilidad a priori se escogen de forma arbitraria de forma que cubran un amplio espectro de valores de entrada.

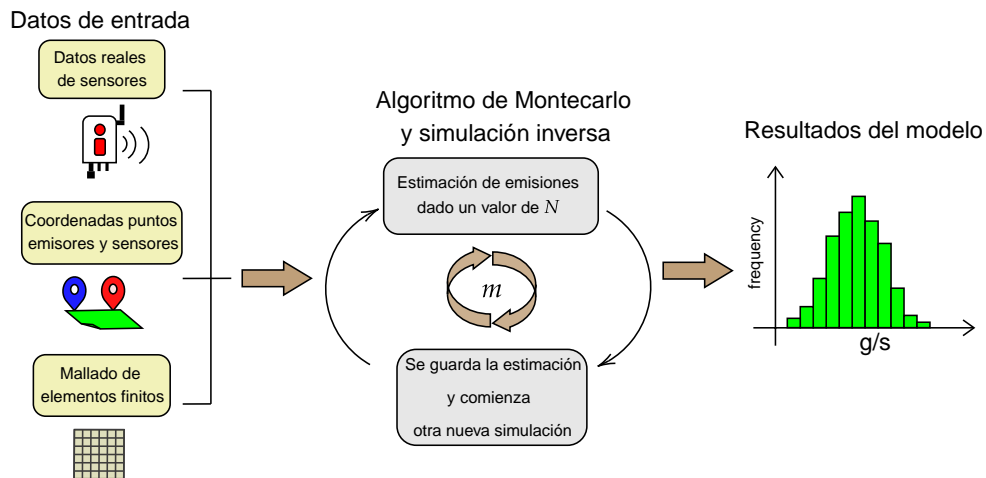


Figura 2.4: Diagrama del modelo empleado en este TFG.

## Resultados

Este capítulo comprende, por un lado, los resultados obtenidos del algoritmo de Montecarlo en el caso de dos y tres flujos de emisión así como un análisis de sensibilidad donde se evalúan cuáles son los parámetros más influyentes, como el número de simulaciones  $N$  y las distribuciones a priori con las que se generan los flujos de emisión. La resolución de la ecuación de convección-difusión se ha resuelto en estado estacionario, es decir, cuando  $\partial c/\partial t = 0$ , la concentración no varía con el tiempo y sobre un dominio de estudio de  $200 \times 200$ m con un mallado de 11827 nodos. La velocidad del viento  $\vec{v}$  (constante en módulo y en la dirección positiva del eje  $x$ ) y el coeficiente de difusión  $D$  han sido elegidos dentro de los valores de número de Péclet que dan estabilidad a la ecuación. Además, la geometría de los flujos de emisión se ha supuesto circular de radio  $r$ .

### 3.1. Caso de dos flujos de emisión

Consideremos en primer lugar un caso de estudio sencillo, en el que tenemos dos flujos de emisión situados en un plano de coordenadas  $(x, y)$  en los puntos  $P_1 = (-40, 10)$ m y  $P_2 = (-75, -10)$ m emitiendo contaminantes desde una fuente con un radio  $r = 2$ m. Contamos con dos puntos de monitoreo de concentración de estas fuentes en los puntos  $R_1 = (-25, -15)$ m y  $R_2 = (0, 20)$ m. Mediante el modelado inverso y el análisis de Montecarlo descrito en la metodología, se han estimado los flujos de emisión  $q_1$  y  $q_2$  con  $N = 4000$ . En este caso las distribuciones a priori con las que se muestrean los datos han sido distribuciones gaussianas con  $\mu_1 = \mu_2 = 30$  y  $\sigma_1 = \sigma_2 = 20$ .

	real	estimación	$\mu$ (g/s)	$\sigma$ (g/s)
$q_1$	10	10.78	30	20
$q_2$	20	19.22	30	20

Tabla 3.1: Resultados de las estimaciones de los flujos de emisión de contaminantes para cada una de las fuentes y los parámetros  $\mu$  y  $\sigma$  utilizados en el muestreo de los datos iniciales.

Los resultados de los valores reales y valores estimados presentados en la tabla 3.1 muestran que el algoritmo ha estimado con precisión los valores reales. La solución a la ecuación de convección-difusión bajo estas condiciones se muestra en la figura 3.1.

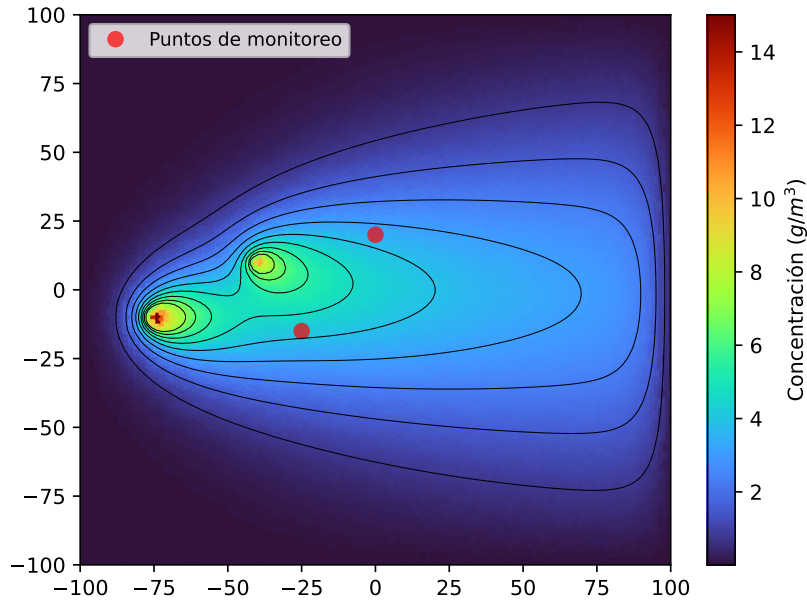
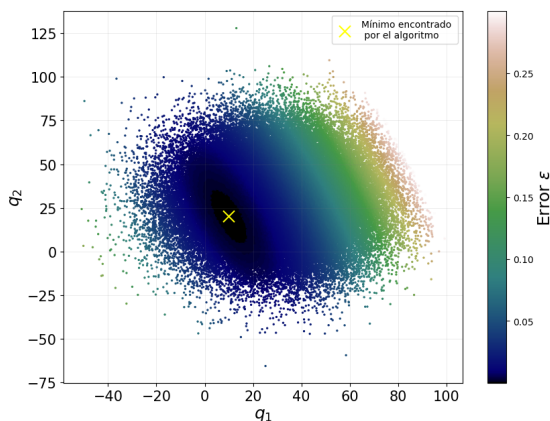
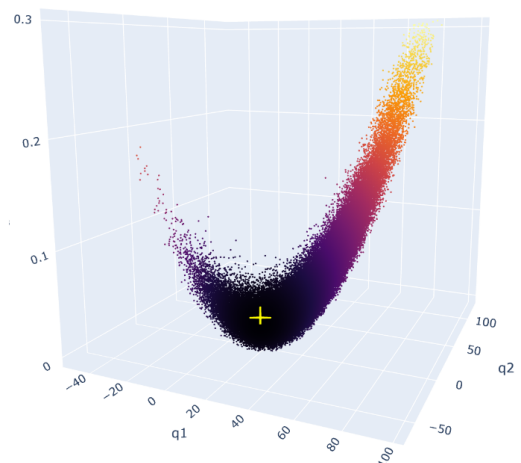


Figura 3.1: Mapa de dispersión que muestra la solución a la ecuación de convección-difusión con fuentes de emisión en los puntos  $P_1$  y  $P_2$ . Los contornos en color negro muestran valores de igual concentración.

Podríamos poner en duda la existencia de un solo valor mínimo de la función de coste  $\epsilon$  o si existen valores de entrada de flujos de emisión  $q_1$  y  $q_2$  diferentes pero que producen también un mínimo. Es decir, ¿Existe más de un conjunto de valores de flujos de emisión que se alejan de los valores reales pero que producen un mínimo de la función de coste? Para dar respuesta a esta pregunta podemos graficar un diagrama que contenga en el eje de abscisas los valores muestreados de  $q_1$ , en el eje de ordenadas los valores muestreados de  $q_2$ , y en el eje  $z$  el error  $\epsilon$  que se comete para cada par de datos de entrada  $q_1$  y  $q_2$  mediante una escala de color.



(a) Representación 2D de los valores de  $q_1$  y  $q_2$  y el error cometido para cada uno de ellos en la simulación.



(b) Representación 3D de los valores de  $q_1$  y  $q_2$  y el error cometido en el eje  $z$ .

Figura 3.2: Representaciones 2D (a) y 3D (b) de los valores muestreados,  $q_1$  y  $q_2$ . En el eje  $z$  el error cometido en la simulación para cada conjunto de flujos de emisión. Con el símbolo amarillo se señala el mínimo encontrado por el algoritmo.

En la figura 3.2 se aprecia cómo existe un único mínimo local de la función de coste  $\epsilon$ . Este resultado es importante, pues significa que el algoritmo no solo hace una estimación precisa de los flujos de emisión, sino que converge a un único mínimo cuyas coordenadas son  $(q_1, q_2)$ , correspondientes a los valores de emisión estimados.

### 3.2. Caso de tres flujos de emisión

En la búsqueda de generalizar el correcto funcionamiento del modelo, y puesto que se ha demostrado que en el caso de dos flujos de emisión los resultados son precisos, este apartado comprueba si sucede lo mismo en el caso de tres flujos de emisión y tres puntos de observación. Para ello, supongamos que tenemos tres flujos de emisión en los puntos  $P_1 = (-40, 0)\text{m}$ ,  $P_2 = (-75, -20)\text{m}$  y  $P_3 = (-70, 30)\text{m}$  y tres puntos de observación de contaminantes  $R_1 = (-25, -15)\text{m}$ ,  $R_2 = (0, 20)\text{m}$  y  $R_3 = (50, -10)\text{m}$ . Los parámetros de entrada de la ecuación diferencial y resolución del dominio a estudiar son iguales que en el caso de dos flujos de emisión. El número de simulaciones ha sido  $N = 4000$ .

	real	estimación	$\mu$ (g/s)	$\sigma$ (g/s)
$q_1$	10	9.64	30	20
$q_2$	15	15.59	30	20
$q_3$	25	23.96	30	20

Tabla 3.2: Resultados de las estimaciones con  $N = 4000$  de los flujos de emisión para cada una de las fuentes, así como los parámetros  $\mu$  y  $\sigma$  utilizados en el muestreo de los datos iniciales.

La concentración de contaminantes para este caso, y de nuevo teniendo en cuenta velocidad del viento  $\vec{v}$  constante en módulo y dirección, se representa en la figura 3.3.

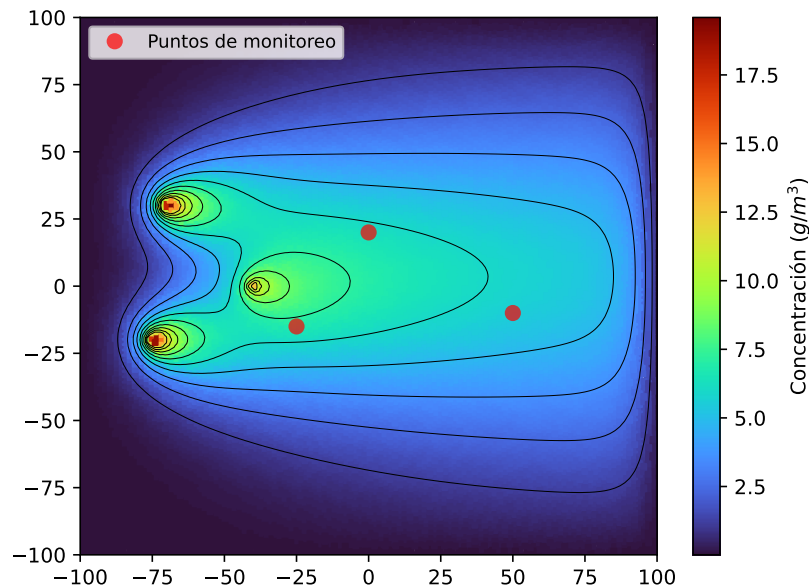


Figura 3.3: Mapa de dispersión que muestra la solución a la ecuación de convección-difusión con fuentes en los puntos  $P_1$ ,  $P_2$  y  $P_3$ . Los contornos en color negro muestran valores de igual concentración.

Los resultados se muestran en la tabla 3.2. Nuevamente, el algoritmo encuentra valores precisos para los flujos de emisión en el caso de 3 fuentes contaminantes. A primera vista los valores parecen ser un poco menos exactos que en el caso de 2 fuentes para el mismo valor de  $N$ . Aún teniendo en cuenta que los resultados nunca serán iguales debido a la componente probabilística que contiene el método de Montecarlo, parece haber una relación entre el error cometido en la estimación y el número de puntos de emisión. Un análisis riguroso se encuentra en la sección 3.3.2.

Una vez comprobado y evaluado el buen funcionamiento del algoritmo, podríamos cuestionar si podemos extender el resultado del modelo a un número más elevado de fuentes contaminantes. La respuesta es que sí, aunque con ciertas consideraciones. Por un lado, como se expondrá en la sección 3.3, requerirá de un número más elevado de simulaciones, puesto que generar el conjunto de datos de entrada adecuado será menos probable. Además, estos datos de entrada deberán ser generados de forma más precisa, lo que requiere de unas mejores distribuciones de probabilidad. El resultado de la comprobación del algoritmo sobre 5 y más flujos de emisión se emite en este TFG por espacio.

### 3.3. Análisis de sensibilidad

Si bien es cierto que se ha logrado una estimación de los flujos de emisión de cada una de las fuentes con cierta exactitud, hasta el momento no se ha realizando un análisis de sensibilidad donde se evalúen cuáles son los parámetros que más afectan al algoritmo y cuál es la relación que tienen con la incertidumbre final de los resultados. Si pensamos en qué variables pueden afectar en la precisión del algoritmo y, puesto que este está basado en el método de Montecarlo, el número de simulaciones  $N$  debe ser un parámetro crucial a estudiar, aunque también lo son otros como las distribuciones de probabilidad a priori o la relación entre número de fuentes y número de sensores de calidad del aire que proporcionan datos de observaciones. Por ejemplo, no tendríamos la misma fiabilidad en los resultados si tenemos un número muy elevado de fuentes contaminantes comparado con el número de puntos de monitoreo que de forma contraria, mayor número de puntos de medición en relación al número de puntos de emisiones.

#### 3.3.1. Número $N$ de simulaciones

El objetivo de este apartado es evaluar la precisión de los resultados en relación al número de simulaciones Montecarlo. Para ello, dado un número  $N$  de simulaciones, una vez obtenidos los flujos de emisión, se calculan de nuevo  $m$  veces reiteradamente (figura 2.4). Esto da como resultado una lista de flujos de emisión estimados para cada una de las fuentes y para cada valor de  $N$ , de la que podemos obtener parámetros como la media y la desviación estándar. Como se ha expuesto, es esperable que el algoritmo mejore los resultados con el aumento de  $N$ , es decir, que la media se vaya aproximando al valor real y que la desviación disminuya.

Para hallar el comportamiento de estos parámetros se han ejecutado simulaciones en el caso de dos fuentes de emisión y dos puntos de observación o monitoreo, con un valor de  $m = 1000$  y registrando los valores de media  $\mu$  y desviación  $\sigma$  de la lista de resultados conforme  $N$  aumenta. Los valores de flujo de emisión reales para cada una de estas dos fuentes son  $q_1 = 100$  g/s y  $q_2 = 60$  g/s.

$N$	$\mu$ (g/s)	$\sigma$ (g/s)
64	97.063	6.409
256	99.381	3.532
1024	99.806	1.767
4096	99.939	0.909
16384	99.966	0.457
65536	100.002	0.226

(a)

$N$	$\mu$ (g/s)	$\sigma$ (g/s)
64	60.954	2.805
256	60.225	1.495
1024	60.072	0.777
4096	60.004	0.376
16384	60.002	0.192
65536	60.001	0.096

(b)

Tabla 3.3: Parámetros estadísticos recogidos para la fuente  $q_1$  (a) y fuente  $q_2$  (b) con variaciones en el número  $N$  de simulaciones Montecarlo.

Los datos mostrados en la tabla 3.3 muestran que, en efecto, conforme se aumenta el valor de  $N$ , la media se va acercando de forma progresiva al valor real del flujo de emisión y la desviación estándar disminuye. Si atendemos más específicamente a la disminución de  $\sigma$  (figura 3.4), a simple vista se percibe que la relación con  $N$  no es lineal.

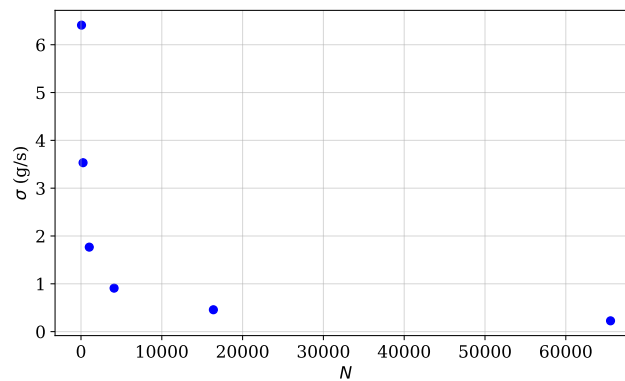


Figura 3.4: Dependencia de  $N$  con la desviación estándar  $\sigma$ , para la fuente  $q_1$ .

Para hallar la dependencia entre estos dos parámetros, supondremos que existe una relación de tipo potencia entre ellos,

$$\sigma \propto N^a \quad . \quad (3.1)$$

Para determinar el valor del parámetro  $a$  a partir de los datos que tenemos, tomaremos logaritmos a ambos lados, de forma que la ecuación quede linealizada:

$$\log \sigma = \log N^a \quad \Rightarrow \quad \log \sigma = a \log N \quad . \quad (3.2)$$

Aplicando propiedades de los logaritmos, obtenemos una relación lineal  $\log \sigma \propto \log N$ . Si hacemos un ajuste por mínimos cuadrados, (figura 3.5) la pendiente de la recta será igual al valor del exponente.

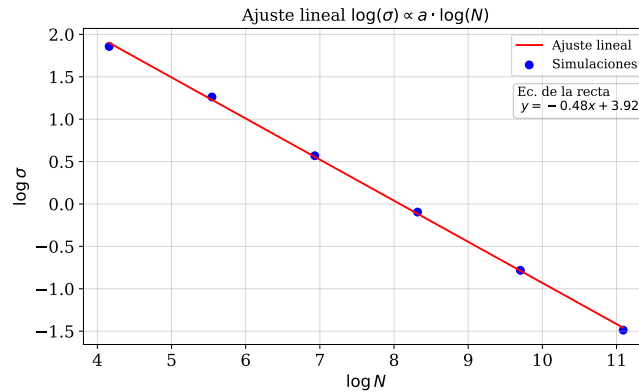


Figura 3.5: Ajuste lineal por mínimos cuadrados, con pendiente  $-0.48$ .

El ajuste lineal revela que, efectivamente, la relación entre  $\sigma$  y  $N$  es exponencial y podemos linealizarla tomando logaritmos. Además, el exponente  $a = -0.48 \simeq -0.5$ , lo que quiere decir que el decaimiento de sigma viene dado por la inversa de la raíz de  $N$  de la forma

$$\sigma \propto \frac{1}{\sqrt{N}} \quad (3.3)$$

Este resultado es importante, pues de esta forma se comprueba uno de los resultados fundamentales de la estadística física (basado en el teorema del límite central), el cual establece que la anchura de una distribución de estimadores es proporcional a  $\frac{1}{\sqrt{N}}$ , donde  $N$  es el número de puntos, simulaciones o ejecuciones utilizado para muestrear el estimador. En el método de Montecarlo, los estimadores se obtienen a partir de una muestra aleatoria de números generados por un proceso estocástico. Por lo tanto, la precisión del estimador obtenido también debe cumplir esta relación, demostrando que la precisión del algoritmo sigue las mismas leyes estadísticas que los métodos de estadística inferencial y justificando una precisión matemática rigurosa además de la validez del método. Una demostración más extensa se encuentra en el capítulo 1 de la referencia [17].

Si queremos hallar de forma exacta el comportamiento de estas dos variables, podemos hacer de nuevo un ajuste  $\sigma \propto \frac{1}{\sqrt{N}}$ , y, puesto que la relación debe ser lineal, la pendiente de la recta y la ordenada en el origen deben darnos el ajuste completo de la dependencia (figura 3.6).

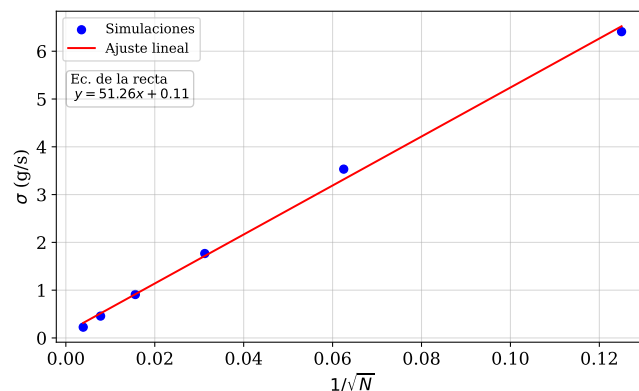


Figura 3.6: Ajuste lineal por mínimos cuadrados, con pendiente 51.26 y ordenada en el origen 0.11.



Por último, la figura 3.7 muestra el ajuste superpuesto a los datos simulados comprobando que, efectivamente, las simulaciones caen sobre la curva  $\sigma = 51.26 \frac{1}{\sqrt{N}} + 0.11$ .

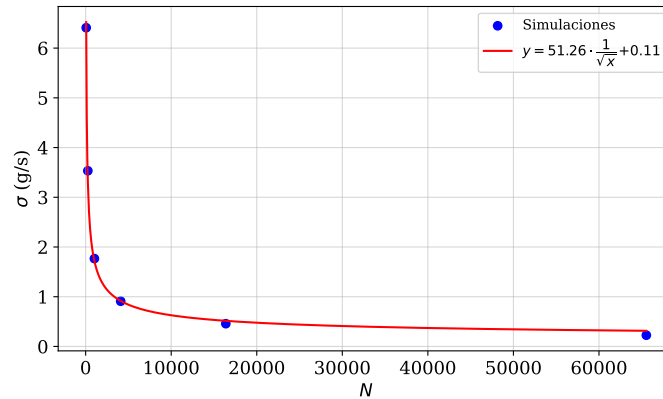


Figura 3.7: Representación de  $\sigma$  en función de  $N$ . Los símbolos corresponden a las simulaciones y la curva es el ajuste numérico de los datos simulados.

Cabe esperar entonces, dada la naturaleza estocástica del método de Montecarlo, que para un mismo valor de  $N$  no se obtengan las mismas estimaciones de los flujos de emisión. Sin embargo, vemos que cuanto mayor es el número de simulaciones que se realizan, menor es la dispersión de los resultados respecto del valor real. De forma gráfica, la figura 3.8 muestra el histograma de los resultados obtenidos para la fuente  $q_1$  en este análisis.

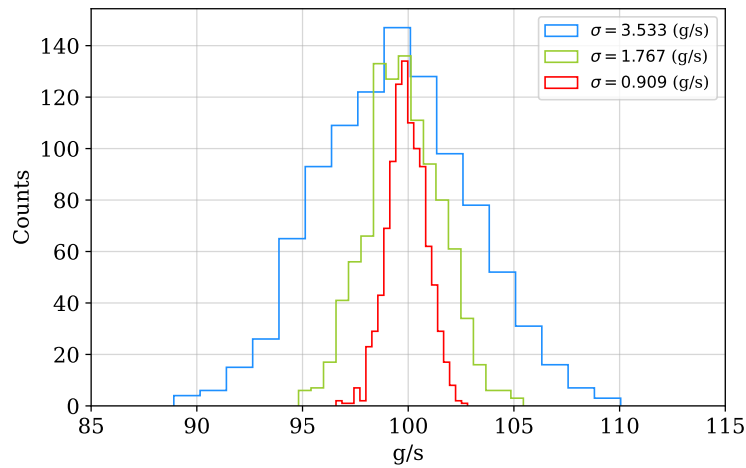


Figura 3.8: Histograma de resultados y desviaciones obtenidas en las simulaciones para la fuente  $q_1$ , con  $m = 1000$ .

La elección de los valores de  $N$  mostrados en la tabla 3.3 se ha escogido en potencias de 4, es decir, multiplicando por 4 el valor de  $N$  anterior, lo que hace que, puesto que la desviación es de la forma  $\sigma \propto \frac{1}{\sqrt{N}}$ , la anchura se reduce a la mitad cada vez y la media tiende a centrarse en el valor real del flujo de emisión, como demuestran los resultados.

### 3.3.2. Distribuciones a priori

Un análisis para dar cuenta de la importancia del número de simulaciones  $N$  y de la elección de las distribuciones a priori puede ser el siguiente: Está claro que para obtener una estimación del conjunto real de flujos de emisión, que denotaremos como  $\{v\}$ , es decir, el conjunto que contiene los flujos reales de cada uno de los puntos de emisión que se pretenden estimar  $\{v\} = \{v_1, v_2, \dots, v_k\}$  es necesario generar un conjunto de emisiones lo más parecido posible en cada una de las  $N$  simulaciones. Como se ha mencionado, cada uno de los valores de flujos de emisión se genera a través de una distribución normal en cada una de las simulaciones. La probabilidad de obtener un valor entre dos valores de una distribución normal viene dada en la ec. (3.4).

$$\mathcal{P}(a < x < b) = \int_a^b \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-\mu)^2}{2\sigma^2}} dx \quad (3.4)$$

Ahora bien, la probabilidad de generar el conjunto de flujos de emisión real dado un cierto margen que llamaremos  $\delta$ , vendrá dado por la multiplicación sucesiva de la probabilidad de generar un valor entre el flujo real y  $\pm\delta$  para cada uno de las componentes del conjunto  $\{v\}$ , o lo que es lo mismo, la probabilidad de generar el conjunto de emisiones real bajo un margen  $\delta$  puede calcularse por el teorema de la probabilidad total, calculando la probabilidad de generar el valor de interés por cada una de las fuentes. La ec. (3.5) muestra como quedaría la expresión, siendo  $x$  la variable de muestreo, en este caso tiene unidades de flujo de emisión (g/s). La probabilidad de generar el conjunto generado decae rápidamente debido a dos principales motivos. En primer lugar, si las distribuciones a priori no están bien escogidas, el valor de las integrales (probabilidad de generar un flujo entre dos valores próximos al flujo real) sería pequeño (por supuesto menor que la unidad) y disminuiría rápidamente cuando aumentamos el número de fuentes. Por otra parte, si aumentamos el número de puntos de emisión  $k$ , la multiplicación sucesiva de valores menores que la unidad hará que la probabilidad decrezca, lo que quiere decir que es un parámetro que afecta a la precisión de los resultados finales:

$$\mathcal{P}(\{v\} - \delta < \{v\} < \{v\} + \delta) = \prod_{i=1}^k \int_{v_i - \delta}^{v_i + \delta} \frac{1}{\sigma_k\sqrt{2\pi}} e^{-\frac{(x-\mu_k)^2}{2\sigma_k^2}} dx \quad (3.5)$$

Aunque este cálculo permite evaluar la importancia del número de simulaciones inicial y de la buena elección de distribuciones a priori en base a otros estudios o inventarios previos, es evidentemente que no se puede realizar si no se tienen los flujos de emisión reales antes de resolver el problema. En este trabajo se plantea el problema desde un punto de vista de validación del algoritmo que permita explorar los límites de la técnica y los parámetros más importantes a tener en cuenta a la hora de llevarlo al estudio de un caso real. Por ejemplo, los parámetros de media y desviación de las distribuciones a priori se pueden obtener de inventarios de emisiones, por ejemplo el citado en la referencia [3], obteniendo la media y la desviación para la localización del caso de estudio a lo largo de un mes o un año.

## Conclusiones

Se ha realizado un estudio con el principal objetivo de estimar los flujos de emisión atmosférica utilizando técnicas de modelado inverso y simulaciones de Montecarlo. La metodología empleada, basada en el uso de valores de entrada aleatorios de los flujos de emisión para hacer una comparación directa con los valores de concentración reales, escogiendo los valores de la simulación que minimiza la diferencia entre el valor simulado y el observado, ha mostrado la validez de esta técnica para estimar de manera precisa los flujos de emisión de dos y tres fuentes de contaminantes. Por un lado, los resultados de las estimaciones en ambos casos han sido realmente precisas, llegando a acercarse a los valores de los flujos de emisión reales. Además, se ha comprobado que en el caso de dos flujos de emisión, el algoritmo converge a una única solución, encontrando de forma unívoca el mínimo de la función de coste, que da como resultado las estimaciones de los flujos de emisión. Este resultado es extrapolable a los casos de 3 y más fuentes. Por otra parte, se ha comprobado que las simulaciones de Montecarlo siguen las mismas leyes estadísticas que los métodos de estadística inferencial, cumpliendo la relación del estimador  $\sigma \propto 1/\sqrt{N}$ , donde  $N$  es el número de simulaciones, lo que indica la validez del modelo de una forma rigurosa. Respecto a la ecuación de convección-difusión, utilizada para simular la dispersión de contaminantes en la atmósfera, se ha explorado el límite en casos de convección dominante, encontrando una limitación en la estabilidad de la ecuación dependiendo del número de Péclet al resolverla por el método de elementos finitos.

Como conclusión final, podemos definir nuevos objetivos y siguientes fases en la evolución e investigación de este estudio. Algunos de ellos son: evaluar el modelo bajo un caso real, donde los datos de las observaciones provengan de sensores o estaciones de la red de calidad del aire en las inmediaciones de fuentes de emisiones como pueden ser polígonos industriales, puertos, o fábricas, estimándose así los correspondientes flujos de emisión de contaminantes y comparándolas con las emisiones que las entidades declaran emitir. Estudiar el posible uso de otro modelo que simule el transporte de contaminantes y que no presente inestabilidades bajo ciertos parámetros de entrada. Continuar investigando la ejecución de estos sistemas de ecuaciones bajo el paradigma de la computación cuántica [9]. Integrar en el modelo la elección de los parámetros de entrada  $\mu$  y  $\sigma$  de las distribuciones a priori a partir de datos de inventarios de emisiones existentes, de manera que se parta de un valor medio y una desviación consecuente con el caso de estudio. Extender los resultados de este trabajo a un número más elevado de fuentes y estudiar la dependencia que hay entre la desviación de las estimaciones y la relación entre número de sensores y número de fuentes de emisión.



# Bibliografía

- [1] L. Deguillaume, M. Beekmann, and L. Menut, “Bayesian Monte Carlo analysis applied to regional-scale inverse emission modeling for reactive trace gases,” *Journal of Geophysical Research: Atmospheres*, vol. 112, no. D02307, 2007.
- [2] M. Beekmann and C. Derognat, “Monte Carlo uncertainty analysis of a regional-scale transport chemistry model constrained by measurements from the atmospheric pollution over the Paris area (ESQUIF) campaign,” *Journal of Geophysical Research: Atmospheres*, vol. 108, no. 8559, 2003.
- [3] M. Guevara, O. Jorba, C. Tena, H. Denier van der Gon, J. Kuenen, N. Elguindi, S. Darras, C. Granier, and C. Pérez García-Pando, “Copernicus Atmosphere Monitoring Service TEMPORal profiles (CAM5-TEMPO): global and European emission temporal profile maps for atmospheric chemistry modelling,” *Earth System Science Data*, vol. 13, no. 2, pp. 367–404, 2021.
- [4] R. Cimrman, V. Lukeš, and E. Rohan, “Multiscale finite element calculations in Python using SfePy,” *Advances in Computational Mathematics*, vol. 45, pp. 1897–1921, 2019.
- [5] K. Lakshminarayananachari, K. S. Pai, M. S. Prasad, and C. Pandurangappa, “Advection-diffusion numerical model of air pollutants emitted from an urban area source with removal mechanisms by considering point source on the boundary,” *International Journal of Application or Innovation in Engineering & Management*, vol. 2, pp. 251–268, 2013.
- [6] L. Menut, B. Bessagnet, R. Briant, A. Cholakian, F. Couvidat, S. Mailler, R. Pennel, G. Siour, P. Tuccella, S. Turquety, *et al.*, “The CHIMERE v2020r1 online chemistry-transport model,” *Geoscientific Model Development*, vol. 14, no. 11, pp. 6781–6811, 2021.
- [7] P. Holnicki, A. Kałuszko, and W. Trapp, “An urban scale application and validation of the CALPUFF model,” *Atmospheric Pollution Research*, vol. 7, no. 3, pp. 393–402, 2016.
- [8] M. Gupta and M. Mohan, “Validation of WRF/Chem model and sensitivity of chemical mechanisms to ozone simulation over megacity Delhi,” *Atmospheric Environment*, vol. 122, pp. 220–229, 2015.
- [9] A. Pujante Perez, E. Illueca Fernandez, N. Bernabe Mulero, N. Gomariz Kühne, J. T. Fernandez Breis, and A. J. Jara Valera, “QuCHIM: First steps in the design of the first quantum chemistry transport model,” *Chimere Workshop*, 2023.

- 
- [10] O. C. Zienkiewicz, R. L. Taylor, and J. Z. Zhu, *The finite element method: its basis and fundamentals*. Elsevier, 2005.
- [11] C. Geuzaine and J.-F. Remacle, “Gmsh: A 3-D finite element mesh generator with built-in pre- and post-processing facilities,” *International Journal for Numerical Methods in Engineering*, vol. 79, no. 11, pp. 1309–1331, 2009.
- [12] M. Huysmans and A. Dassargues, “Review of the use of Péclet numbers to determine the relative importance of advection and diffusion in low permeability environments,” *Hydrogeology Journal*, vol. 13, pp. 895–904, 2005.
- [13] H. C. Elman, D. J. Silvester, and A. J. Wathen, *Finite elements and fast iterative solvers: with applications in incompressible fluid dynamics*. Numerical Mathematics and Scie, 2014.
- [14] R. Codina, “A finite element formulation for the numerical solution of the convection-diffusion equation,” *Monograph CIMNE*, 2019.
- [15] H. Gómez, I. Colominas, F. Navarrina, and M. Casteleiro, “Un planteamiento generalizado para el problema de convención-difusión. Formulación numérica y aplicaciones,” 2005.
- [16] R. Y. Rubinstein and D. P. Kroese, *Simulation and the Monte Carlo method*. John Wiley & Sons, 2016.
- [17] P. Glasserman, *Monte Carlo methods in financial engineering*, vol. 53. Springer, 2004.
- [18] G. L. Guymon, V. Scott, and L. Herrmann, “A general numerical solution of the two-dimensional diffusion-convection equation by the finite element method,” *Water Resources Research*, vol. 6, no. 6, pp. 1611–1617, 1970.
- [19] M. Mulholland and J. H. Seinfeld, “Inverse air pollution modelling of urban-scale carbon monoxide emissions,” *Atmospheric Environment*, vol. 29, no. 4, pp. 497–516, 1995.
- [20] D. Pritchard and J. Currie, “Diffusion of coefficients of carbon dioxide, nitrous oxide, ethylene and ethane in air and their measurement,” *Journal of Soil Science*, vol. 33, no. 2, pp. 175–184, 1982.

## Estabilidad de la ecuación de convección-difusión

La ecuación de convección-difusión es una ecuación diferencial que se utiliza para describir la evolución temporal de una sustancia en un medio en movimiento. Esta ecuación es muy útil para modelar una amplia variedad de fenómenos físicos en ciencias aplicadas, como el transporte de sustancias en el suelo, la difusión de contaminantes en la atmósfera o el movimiento de líquidos en tuberías. Sin embargo, cuando la convección es dominante, es decir, cuando la sustancia se mueve rápidamente en el medio, la obtención de soluciones numéricas precisas y estables para la ecuación de convección-difusión se vuelve muy difícil. En estas situaciones, la ecuación se convierte en una ecuación hiperbólica, lo que significa que las soluciones numéricas pueden ser inestables y producir resultados erróneos en forma de oscilaciones. Veamos de una forma analítica bajo qué situaciones y por qué surgen estos problemas.

Comenzaremos considerando la ecuación diferencial de convección-difusión en una sola dimensión en el intervalo  $[0, L]$  con condiciones de contorno

$$D \frac{d^2 c}{dx^2} - v \frac{dc}{dx} = R(x), \quad 0 < x < 1 \quad c(0) = 0, \quad c(1) = 1 \quad . \quad (\text{A.1})$$

Para resolver la ecuación, supongamos que el dominio está dividido en  $k$  partes

$$0 = x_0 < x_1 < \dots < x_{k-1} < x_k = L \quad , \quad (\text{A.2})$$

La discretización de Galerkin para elementos finitos es de la forma

$$(1 - \gamma)c_{i+1} - 2c_i + (1 + \gamma)c_{i-1} = \frac{h^2}{D} R(x_k) \quad (\text{A.3})$$

donde  $\gamma = \frac{vh}{2D}$  es el número de Péclet y  $h$  es la distancia entre cada uno de los segmentos en los que se ha discretizado el dominio. La ec. (A.1) en el caso de  $R(x) = 0$  tiene como solución analítica la función  $c(x)$

$$c(x) = \frac{e^{\frac{v}{D}x} - 1}{e^{\frac{v}{D}} - 1} \quad (\text{A.4})$$

La solución en cada una de las particiones del dominio será por tanto

$$c(x) = \frac{e^{\frac{v}{D}x_i} - 1}{e^{\frac{v}{D}} - 1} \quad i = \{1, 2 \dots k\} \quad (\text{A.5})$$

Si  $h$  es la distancia entre cada uno de los nodos en los que hemos discretizado el dominio, entonces  $x_i = i \cdot h$  y teniendo en cuenta la definición de número de Péclet, indicada al principio de la demostración, la ec. (A.5) puede formularse como

$$c_i = \frac{e^{\frac{v}{D}ih} - 1}{e^{\frac{v}{D}} - 1}$$

$$c_i = \frac{e^{2i\gamma} - 1}{e^{\frac{v}{D}} - 1} \quad (\text{A.6})$$

Si sustituimos esta expresión en la ecuación ec. (A.3) sacando factor común el término  $(e^{\frac{v}{D}} - 1)$  y, puesto que está igualado a 0, solamente quedan los términos

$$(1 - \gamma)(e^{2(i+1)\gamma} - 1) - 2(e^{2i\gamma} - 1) + (1 + \gamma)(e^{2(i-1)\gamma} - 1)$$

Si desarrollamos los términos y sacamos factor común podemos simplificar la ecuación anterior

$$e^{2\gamma i} e^{2\gamma} - 1 - \gamma e^{2\gamma i} e^{2\gamma} + \gamma - 2e^{2\gamma i} + 2 + e^{2\gamma i} e^{-2\gamma} - 1 + \gamma e^{2\gamma i} e^{-2\gamma} - \gamma = 0$$

$$e^{2\gamma i} e^{2\gamma} - \gamma e^{2\gamma i} e^{2\gamma} - 2e^{2\gamma i} + e^{2\gamma i} e^{-2\gamma} + \gamma e^{2\gamma i} e^{-2\gamma} = 0$$

$$e^{2\gamma i} e^{2\gamma} (1 - \gamma) + e^{2\gamma i} e^{-2\gamma} (1 + \gamma) - 2e^{2\gamma i} = 0$$

Podemos sacar factor común el término  $e^{2\gamma i}$  e igualar a cero de nuevo para obtener la expresión final en términos del número de Péclet  $\gamma$

$$e^{2\gamma} (1 - \gamma) + e^{-2\gamma} (1 + \gamma) - 2 = 0 \quad (\text{A.7})$$

Para simplificar la ecuación podemos multiplicar a ambos lados por  $e^{2\gamma}$

$$e^{4\gamma} (1 - \gamma) - 2e^{2\gamma} + (1 + \gamma) = 0 \quad (\text{A.8})$$

Y haciendo  $\lambda = e^{2\gamma}$  queda una ecuación de segundo grado en  $\lambda$

$$(1 - \gamma)\lambda^2 - 2\lambda + (1 + \gamma) = 0 \quad (\text{A.9})$$



La ecuación A.9 tiene como soluciones

$$\lambda_1 = 1 \quad , \quad \lambda_2 = \frac{1 + \gamma}{1 - \gamma}$$

Si deshacemos el cambio anterior  $\lambda = e^{2\gamma}$ , podemos reescribir la solución discretizada de la ec. (A.11) en función de las soluciones y del número de Péclet  $\gamma$

$$c_i = \frac{(\lambda_2)^i - 1}{e^{\frac{\gamma}{D}} - 1} \quad (\text{A.10})$$

Se ha tomado la solución  $\lambda_2$  porque es la que depende del número de Péclet, y por tanto la que nos interesa estudiar. Si sustituimos, la solución a la concentración en los diferentes puntos discretizados queda de la forma

$$c_i = \frac{\left(\frac{1+\gamma}{1-\gamma}\right)^i - 1}{e^{\frac{\gamma}{D}} - 1} \quad (\text{A.11})$$

De esta forma se ve claramente que la solución será oscilatoria cuando  $\gamma > 1$ , pues el término del numerador se hace negativo para valores de  $i$  impares, y positivo cuando los valores de  $i$  son pares, lo que hará una solución oscilatoria conforme se van recorriendo los puntos de la división del dominio. Este resultado es extrapolable al caso de la ecuación de convección-difusión en 2D, por tanto queda entonces demostrado que la solución discretizada por elementos finitos de Galerkin presenta inestabilidad y oscilaciones para la ecuación diferencial de convección-difusión en casos de convección dominante.