

EMMANUEL GALVÁN  
HERNÁNDEZ

INFERENCIA BAYESIANA EN PROBLEMAS INVERSOS: ANÁLISIS DE UN MODELO ATMOSFÉRICO DE  
ECUACIONES DIFERENCIALES ORDINARIAS

2021



**Universidad Autónoma de Querétaro**

**Facultad de Ingeniería**

INFERENCIA BAYESIANA EN PROBLEMAS INVERSOS: ANÁLISIS DE UN  
MODELO ATMOSFÉRICO DE ECUACIONES DIFERENCIALES ORDINARIAS

**Tesis**

Que como parte de los requisitos  
para obtener el Grado de

LICENCIADO EN MATEMÁTICAS APLICADAS

Presenta:

Emmanuel Galván Hernández

Dirigido por:

Dr. Mario Santana Cibran

Querétaro, Qro., abril 2021



Universidad Autónoma de Querétaro

Facultad de Ingeniería

Licenciatura en Matemáticas Aplicadas

**Inferencia Bayesiana en problemas inversos: Análisis de un modelo atmosférico de ecuaciones diferenciales ordinarias.**

TESIS

Que como parte de los requisitos para obtener el grado de  
Licenciado en Matemáticas Aplicadas

Presenta:

**Emmanuel Galván Hernández**

Dirigido por:

**Dr. Mario Santana Cibrian**

SINODALES

Dr. Mario Santana Cibrian

Presidente

Firma

Dr. Eduardo Castaño Tostado

Secretario

Firma

M. en C. Juan Antonio Villeda Reséndiz

Vocal

Firma

M. en C. Wilfrido Jacobo Paredes García

Vocal

Firma

Dr. Manuel Toledano Ayala

Director de la Facultad de Ingeniería

Centro Universitario

Querétaro, QRO

México.

Abril 2021

© 2021 - Emmanuel Galván Hernández

All rights reserved.

Dirección General de Bibliotecas de la UAQ

Dirección General de Bibliotecas de la UAQ

*Dedicatoria*

A mis padres quienes me han apoyado en todo momento, quienes me han enseñado valores y a no frenar ante las dificultades y miedos.

Dirección General de Bibliotecas de la UAQ

Dirección General de Bibliotecas de la UAQ

# Agradecimientos

Quiero expresar de forma breve pero sincera mi gratitud a todas las personas que han creído en mí, quienes me han impulsado para llegar hasta aquí; en especial mi familia y amigos. Cada una de esas personas especiales ha aportado algo valioso en mi vida, por ello los recuerdo con cariño.

A mi asesor de tesis, el Dr. Mario Santana Cibran, por su tiempo, dedicación y paciencia; por brindarme la oportunidad de trabajar con él.

Dirección General de Bibliotecas de la UAQ

# Abstract

In this thesis, an atmospheric model of ordinary differential equations is analyzed. It describes the equivalent potential temperature of an infinitesimal air bubble called air parcel and it allows to study the clouds atmospheric instability. The aim of this work is to show the estimation process of unknown parameters involved in dynamical systems using this atmospheric model as an example. This type of problem is known as an *inverse problem* and, in the thesis, it is approached from the statistical perspective using Bayesian inference. Within this methodology, the solution to an *inverse problem* is the exploration of the posterior distribution of the parameters of interest. To construct such a distribution, weakly informative prior probability distributions are used. Also, a Gaussian likelihood function was employed to reflect the relationship between the data and the model. Two statistical models are presented which mainly differ in the stochastic term used to measure the observational error. The estimation method is illustrated with a real data set consisting of cloud temperature observations at different heights. The results show that there is enough information in the data to estimate the parameters of interest and the estimates produced by both models are consistent with each other. This work shows that the statistical estimation approach is more appropriate than classical regularization methods for this kind of problems since it allows to measure the uncertainty of the process, unlike classical approaches that only perform calibration.

Dirección General de Bibliotecas de la UAQ

# Resumen

En el presente trabajo se analiza un modelo atmosférico de ecuaciones diferenciales ordinarias, el cual describe la temperatura potencial equivalente de una burbuja de aire de dimensiones infinitesimales denominada parcela, y permite estudiar la inestabilidad atmosférica de dichas parcelas. El objetivo del trabajo es mostrar el proceso de estimación de parámetros desconocidos en sistemas dinámicos usando como ejemplo el modelo atmosférico. Este tipo de problema es conocido como *problema inverso* y se aborda desde el punto de vista estadístico por medio de inferencia Bayesiana. En este enfoque, la solución al *problema inverso* consiste en la exploración de la distribución *a posteriori* de los parámetros de interés. Para construir dicha distribución, se utilizan distribuciones de probabilidad *a priori* débilmente informativas y una función de verosimilitud Gaussiana, la cual refleja la relación entre los datos y el modelo. En el trabajo se proponen dos modelos estadísticos cuya diferencia principal radica en la forma del término estocástico que se utilizó para medir el error en los datos. El método de estimación se ejemplifica con un conjunto de datos reales que consta de observaciones de temperatura de nubes a distintas alturas. Los resultados muestran que hay suficiente información en los datos para estimar los parámetros de interés y las estimaciones producidas por ambos modelos son consistentes entre sí. El trabajo muestra que el enfoque estadístico de estimación es más apropiado que los métodos clásicos de regularización para esta clase de problemas ya que permite cuantificar la incertidumbre del proceso, a diferencia de los esquemas clásicos que sólo hacen calibración.

Dirección General de Bibliotecas de la UAQ

# Índice General

Agradecimientos	I
Abstract	III
Resumen	V
Índice General	VII
Índice de Figuras	IX
Índice de Tablas	XI
<b>1. Introducción</b>	<b>1</b>
<b>2. Metodología</b>	<b>5</b>
2.1. Inferencia Estadística	5
2.2. Inferencia Bayesiana	6
2.2.1. Distribución <i>a posteriori</i>	7
2.2.2. Distribución <i>a priori</i>	9
2.2.3. Distribución <i>a posteriori predictiva</i>	11
2.3. Markov Chain Monte Carlo (MCMC)	11
2.3.1. Algoritmo Metrópolis-Hastings	12
2.3.2. T-Walk	14
2.4. Problemas inversos	14
<b>3. Inferencia sobre la temperatura potencial</b>	<b>17</b>
3.1. Modelo teórico para el estudio de inestabilidades atmosféricas de parcelas de aire	17
3.2. Modelo Estadístico 1	19
3.2.1. Función de Máxima Verosimilitud	20

3.2.2. Función de probabilidad <i>a priori</i> . . . . .	21
3.2.3. Función de probabilidad <i>a posteriori</i> . . . . .	21
3.2.4. Distribución <i>a posteriori predictiva</i> . . . . .	22
3.3. Modelo Estadístico 2 . . . . .	23
3.3.1. Función de Máxima Verosimilitud . . . . .	25
3.3.2. Función de probabilidad <i>a priori</i> . . . . .	25
3.3.3. Función de probabilidad <i>a posteriori</i> . . . . .	25
<b>4. Resultados</b>	<b>27</b>
4.1. Resultados del modelo estadístico 1 . . . . .	27
4.2. Resultados del modelo estadístico 2 . . . . .	31
4.3. Comparación de los modelos estadísticos . . . . .	34
<b>5. Conclusiones</b>	<b>39</b>
<b>A. Conjunto de observaciones de temperaturas en las nubes</b>	<b>41</b>
<b>B. Códigos de R</b>	<b>43</b>
<b>Bibliografía</b>	<b>56</b>

# Índice de Figuras

3.1. Comparación entre el modelo estocástico (3.14) y datos reales . . . . .	24
4.1. Función de energía del modelo estadístico 1 . . . . .	28
4.2. Trayectorias de las cadenas de los parámetros desconocidos del modelo estadístico 1 . . . . .	29
4.3. Histogramas de las muestras de las distribuciones <i>a posteriori</i> de los parámetros desconocidos del modelo estadístico 1 . . . . .	30
4.4. Bandas de probabilidad del 95 % del valor esperado de la temperatura potencial equivalente de la parcela, para el modelo estadístico 1 . . . . .	31
4.5. Bandas predictivas del 95 % de la temperatura potencial equivalente de la parcela, para el modelo estadístico 1 . . . . .	32
4.6. Función de energía del modelo estadístico 2 . . . . .	33
4.7. Trayectorias de las cadenas de los parámetros desconocidos del modelo estadístico 2 . . . . .	34
4.8. Histogramas de las muestras de las distribuciones <i>a posteriori</i> de los parámetros desconocidos del modelo estadístico 2 . . . . .	35
4.9. Bandas de probabilidad del 95 % del valor esperado de la temperatura potencial equivalente de la parcela, para el modelo estadístico 2 . . . . .	36
4.10. Bandas de predictivas del 95 % del valor esperado de la temperatura potencial equivalente de la parcela, para el modelo estadístico 2 . . . . .	37
4.11. Comparación de las curvas predictivas de temperatura obtenidas con los modelos 1 y 2 . . . . .	38
4.12. Comparación entre el ajuste del modelo estadístico 2 y el ajuste presentado en Flores [2017] . . . . .	38

Dirección General de Bibliotecas de la UAQ

# Índice de Tablas

3.1. Parámetros conocidos del modelo diferencial (3.1). . . . .	19
3.2. Distribuciones <i>a priori</i> de los parámetros desconocidos del modelo estadístico 1. . . . .	21
3.3. Distribuciones <i>a priori</i> de los parámetros desconocidos del modelo estadístico 2. . . . .	26
4.1. Comparación de las estimaciones de los parámetros de los modelos 1 y 2. . . . .	36

Dirección General de Bibliotecas de la UAQ

Dirección General de Bibliotecas de la UAQ

# Introducción

En la actualidad existe una gran cantidad de modelos matemáticos que sirven como herramientas para entender, modelar y predecir diversos fenómenos naturales en distintas ramas de la ciencia. Estos fenómenos en muchas ocasiones son altamente complejos y típicamente son modelados por sistemas de ecuaciones diferenciales que dependen del tiempo y/o espacio. A pesar de que estos modelos matemáticos son sólo aproximaciones del mundo real, proporcionan una herramienta fundamental para el estudio de estos fenómenos.

Si suponemos que un modelo representa adecuadamente a un fenómeno de interés entonces, para poder reproducir la dinámica observada, los modelos dinámicos tienen un conjunto de parámetros que deben ser determinados. Algunos de estos parámetros son fáciles de obtener del conocimiento existente del fenómeno o a partir de estudios previos, mientras que otros parámetros requieren un proceso de calibración o estimación. En la literatura, a los problemas de estimación de parámetros en sistemas dinámicos se les conoce como problemas inversos. Estos se pueden encontrar en diferentes áreas de estudio como en geofísica, medicina, ingeniería, biología, etc. Por ejemplo, los procedimientos para la reconstrucción de imágenes médicas como la tomografía computarizada de rayos-X [Kaipio and Somersalo, 2006] son considerados problemas inversos. En epidemiología, la estimación de la fuerza de infección de modelos tipo SIR (Susceptible-Infectado-Recuperado), los cuales describen la propagación de una enfermedad, también son considerados problemas inversos [Acuña-Zegarra et al., 2020, Santana-Cibrian et al., 2020]. Este tipo de problemas tienen la característica de que las cantidades de interés no son directamente observables.

En la práctica, el esquema de estimación estadística de parámetros consiste en encontrar los valores de los parámetros que describen mejor a un conjunto de datos observados que están sujetos a cierta cantidad de ruido tanto ambiental como experimental. La estimación de parámetros, en el contexto de ecuaciones diferenciales, comúnmente se lleva a cabo mediante algún esquema de optimización con restricciones a través de los llamados *métodos de regularización*. Estos métodos suelen producir una estimación puntual razonable para los parámetros de interés a partir de los datos observados. Sin embargo, las técnicas de regularización no cuantifican la incertidumbre que presenta este proceso, y suelen tener problemas de

estabilidad a la hora de hallar las soluciones. En este sentido, el enfoque estadístico Bayesiano representa una mejor opción ya que considera de forma explícita la variabilidad inherente de los datos y del fenómeno. En particular, el esquema de inferencia Bayesiana considera a los parámetros de interés como variables aleatorias. Esto permite cuantificar la incertidumbre en la estimación de parámetros ya que la solución no es una estimación puntual sino una distribución de probabilidad. Estos procedimientos generan información sobre la variabilidad de los parámetros que, a su vez, permiten evaluar la variabilidad en el ajuste del modelo. Más aún, estos métodos de inferencia permiten incorporar toda la información disponible del fenómeno al proceso de estimación. Por último, la interpretación que se obtiene es mucho más clara e intuitiva que con los métodos de regularización.

El presente trabajo se enfoca en un modelo atmosférico de ecuaciones diferenciales ordinarias, presentado en la tesis *Teoría de la parcela de fluidos para el estudio de inestabilidades atmosféricas y distribución de alturas de las nubes*, [Flores, 2017], donde no se muestra evidencia de atacar la problemática de la estimación de parámetros mediante algún método. Lo anterior es relevante para la aplicación “correcta” de dicho modelo. En general, la comunidad científica pone bastante esfuerzo en estudiar las propiedades teóricas de un modelo (tal como la existencia y unicidad de la solución, etc.) dejando a un lado la aplicación del mismo, siendo este el caso presentado en [Flores, 2017]. Por tal motivo, en este trabajo se toma como caso de estudio este modelo para mostrar las ventajas de los esquemas de inferencia Bayesiana en el contexto de problemas inversos.

## Objetivos

### Objetivo General

Utilizar el método Bayesiano para determinar los parámetros de interés, involucrados en el modelo planteado por [Flores, 2017], con base en observaciones del fenómeno.

### Objetivos específicos

- Evaluar dos modelos estadísticos para dar cumplimiento al objetivo general, los cuales miden de forma distinta la incertidumbre del fenómeno.
- Evaluar los resultados propuestos en [Flores, 2017] utilizando datos reales.

## Estructura de la Tesis

En el Capítulo 2 se describen las bases de la metodología empleada para dar cumplimiento al objetivo general de esta tesis. En el Capítulo 3 se presenta el modelo teórico planteado por [Flores, 2017]; se dan algunos conceptos clave en los que dicho modelo sienta bases, y se presentan los modelos estadísticos para la estimación de los parámetros de interés. En el Capítulo 4 se muestran los resultados de las estimaciones encontradas y se da interpretación

a éstas. Finalmente, en el Capítulo 5 se dan a conocer las conclusiones obtenidas en el desarrollo de esta tesis.

Dirección General de Bibliotecas de la UAQ

Dirección General de Bibliotecas de la UAQ

---

# Metodología

En este Capítulo se explicará la teoría y metodología empleada en la tesis. La Sección 2.1 da las bases de inferencia estadística y de las dos metodologías predominantes del área. En la Sección 2.2 se describe en detalle el paradigma Bayesiano de inferencia. Aquí se explican conceptos como: el teorema de Bayes, distribución *a priori*, distribución *a posteriori* y distribución *a posteriori predictiva*. En la Sección 2.3 se da una breve introducción a los métodos de Monte Carlo (algoritmos Metrópolis-Hastings y T-Walk), los cuales permiten la implementación del esquema Bayesiano en escenarios complejos. Finalmente, la Sección 2.4 describe la noción de *problema inverso* y como se aborda este tipo de problemas a través de la inferencia Bayesiana.

## 2.1. Inferencia Estadística

La inferencia estadística se define como el proceso de utilizar un conjunto de datos para inferir la distribución de probabilidad que genera dicho conjunto de datos. A grandes rasgos, los problemas de inferencia estadística pueden ser clasificados en: estimación puntual de parámetros, estimación por intervalo o pruebas de hipótesis.

Los dos enfoques de inferencia más comunes para la estimación de parámetros son el frequentista y el Bayesiano. El rol de probabilidad en ambos enfoques es crucial, sin embargo, cada uno de ellos lo aborda de forma distinta. A continuación, se presentan los postulados de ambas formas de inferencia de acuerdo a Wasserman [2013].

### ■ Inferencia Frequentista

1. La probabilidad se refiere a un límite de frecuencias relativas. Las probabilidades son propiedades objetivas del mundo real. Para ejemplificar, supóngase que se realizan  $n$  repeticiones de un experimento y se contabilizan  $n_E$  ocurrencias del

evento  $E$ , luego la probabilidad frecuentista del evento  $E$  se define como  $P(E) = \lim_{n \rightarrow \infty} \frac{n_E}{n}$ , [Rincón, 2014]. Sin embargo, dado que no es posible realizar de forma infinita este experimento, entonces empíricamente se tiene que  $P(E) \approx \frac{n_E}{n}$ .

2. Los parámetros son constantes fijas (usualmente desconocidos). Esto implica que no se puedan hacer declaraciones de probabilidad sobre ellos.
3. Los procedimientos estadísticos deben ser diseñados para que tengan propiedades frecuentistas bien definidas de largo plazo. Por ejemplo, un intervalo de confianza del 95 % debe atrapar el verdadero valor del parámetro con una frecuencia límite de al menos el 95 %.

### ■ Inferencia Bayesiana

1. La probabilidad describe el grado de creencia acerca de la ocurrencia de un evento. Como tal, podemos hacer declaraciones de probabilidad sobre muchas cosas, no solo de datos que están sujetos a variaciones aleatorias. Por ejemplo, en un encuentro de boxeo ¿cuál es la probabilidad de que un participante en particular sea el ganador? Alguien podría afirmar que dicha probabilidad es 0.71, puesto que se sabe que ese participante está físicamente en forma. Sin embargo, un experto en el tema de boxeo puede dar una idea más certera acerca de esta probabilidad. Este ejemplo muestra que la probabilidad depende del observador y del conocimiento que éste tiene del tema; es decir, toda probabilidad es *condicional*.
2. Los parámetros a inferir se consideran variables aleatorias y se pueden hacer declaraciones de probabilidad sobre ellos.
3. La inferencia de un parámetro  $\theta$  se hace obteniendo una distribución de probabilidad, de la cual se pueden extraer estimaciones puntuales y por intervalo.

Existe una amplia discusión en la literatura sobre las ventajas y desventajas de ambos enfoques. No obstante, en este trabajo se utiliza el enfoque Bayesiano ya que se ha observado que tiene ventajas sobre el esquema frecuentista en el contexto de problemas inversos [Kaipio and Somersalo, 2006].

## 2.2. Inferencia Bayesiana

Dado un modelo de probabilidad paramétrico que describe el comportamiento de un fenómeno y dada una muestra de observaciones de dicho fenómeno, es factible realizar inferencia sobre aquellos parámetros desconocidos aplicando el paradigma Bayesiano. Este proceso no es más que incorporar la información que contiene la muestra junto con la información previa que se tiene sobre los parámetros por medio del teorema de Bayes. Esto da como resultado una distribución de probabilidad condicional para el vector de parámetros, llamada distribución *a posteriori*. A continuación, se explicará formalmente este proceso.

Sea  $\mathbf{y} = (y_1, \dots, y_n)$  una muestra de la variable aleatoria continua  $Y$  que depende del vector de parámetros desconocidos  $\boldsymbol{\theta}$  y sigue una función de densidad<sup>1</sup>. Se desea estimar el valor desconocido de cada entrada del vector  $\boldsymbol{\theta}$ . Los pasos a seguir en inferencia Bayesiana son:

1. Elegir una función de densidad *a priori*,  $f(\boldsymbol{\theta})$ , para el vector de parámetros desconocidos, la cual refleja nuestras creencias (o conocimiento) sobre  $\boldsymbol{\theta}$ .
2. Calcular la función de densidad conjunta de los datos condicional a los valores de los parámetros,  $f(\mathbf{y}|\boldsymbol{\theta})$ , es decir, la función de verosimilitud. Esta función refleja la relación entre el vector de parámetros desconocidos  $\boldsymbol{\theta}$  y el vector de observaciones  $\mathbf{y}$ .
3. Emplear el teorema de Bayes para actualizar las creencias sobre  $\boldsymbol{\theta}$  después de haber observado la información  $\mathbf{y}$ , y así determinar la función de densidad *a posteriori*  $f(\boldsymbol{\theta}|\mathbf{y})$ .

Los pasos anteriores describen el caso para variables aleatorias continuas; no obstante, el caso para variables aleatorias discretas es análogo, basta con utilizar la función de probabilidad en lugar de una función de densidad.

### 2.2.1. Distribución *a posteriori*

Una vez establecidos ambos la distribución *a priori* y el modelo estadístico (función de verosimilitud), se emplea el teorema de Bayes para actualizar el conocimiento previo de  $\boldsymbol{\theta}$  dada la información que se tiene  $\mathbf{y}$ , y así determinar la distribución *a posteriori*. Recordemos que el teorema de Bayes, publicado en 1763 por Thomas Bayes, inicialmente fue pensado para eventos:

**Teorema 1 (Teorema de Bayes).** Sean  $A_1, A_2, \dots, A_n$  eventos, cada uno con probabilidad positiva, que particiona a el espacio muestral  $\Omega$ . Si  $B$  es un evento definido en  $\Omega$ , con  $P[B] > 0$ , entonces para cualquier  $1 \leq j \leq n$

$$P[A_j | B] = \frac{P[B | A_j] P[A_j]}{\sum_{k=1}^n P[B | A_k] P[A_k]}. \quad (2.1)$$

El teorema de Bayes se puede aplicar a variables aleatorias: continuas y/o discretas. Para el caso de variables aleatorias continuas se tiene el siguiente teorema:

<sup>1</sup>Notación: Por simplicidad, en este trabajo se escribe  $f(x)$  y  $f(y)$  para denotar a las funciones de densidad de las variables aleatorias  $X$  y  $Y$  respectivamente, es decir, el argumento de  $f$  indica la variable aleatoria a la que pertenece dicha función. Esto es un abuso de notación ya que, formalmente, al hablar de la función de densidad de la variable aleatoria  $X$  (que toma valores  $x$ ), se debería escribir  $f_X(x)$ . Sin embargo, esto puede crear notaciones muy complejas al considerar distribuciones conjuntas de dos o más variables aleatorias.

**Teorema 2 (Teorema de Bayes para v.a's continuas).** Sean  $X$  y  $\Theta$  variables aleatorias continuas y  $\int f(x|\theta)f(\theta) d\theta > 0$ , entonces

$$f(\theta|x) = \frac{f(x|\theta)f(\theta)}{\int f(x|\theta)f(\theta) d\theta}, \quad (2.2)$$

donde  $f(\theta)$  representa la distribución *a priori*;  $f(\theta|x)$  es la distribución *a posteriori*, y  $f(x|\theta)$  es el modelo estadístico que refleja la relación entre  $x$  y  $\theta$ . La integral  $\int f(x|\theta)f(\theta) d\theta$  es una constante que no depende de  $\theta$ , llamada *constante de normalización*, la cual hace que  $f(\theta|x)$  sea una función de densidad bien definida.

En general, el teorema de Bayes fue diseñado para el cálculo de probabilidades condicionales. En el caso específico de inferencia estadística, es necesario considerar las siguientes modificaciones para el cálculo de la distribución *a posteriori*:

- a) Si  $\Theta$  es un vector aleatorio y se tienen  $n$  observaciones i.i.d.  $x_1, x_2, \dots, x_n$  generadas por la variable aleatoria  $X$ , entonces  $f(x|\theta)$  es reemplazada por  $f(\mathbf{x}|\theta)$ , donde

$$f(\mathbf{x}|\theta) = \prod_{i=1}^n f(x_i|\theta) = \mathcal{L}_n(\theta).$$

Aquí  $\mathbf{x} = (x_1, \dots, x_n)$ ,  $\theta = (\theta_1, \dots, \theta_k)$  y  $\mathcal{L}_n(\theta)$  es la función de máxima verosimilitud.

- b) Con base en lo anterior, la ecuación (2.2) toma la forma:

$$f(\theta|\mathbf{x}) = \frac{\mathcal{L}_n(\theta)f(\theta)}{\int \mathcal{L}_n(\theta)f(\theta) d\theta}. \quad (2.3)$$

- c) Como  $\int \mathcal{L}_n(\theta)f(\theta) d\theta$  no depende de  $\theta$  y por tanto es una constante, entonces

$$f(\theta|\mathbf{x}) \propto \mathcal{L}_n(\theta)f(\theta). \quad (2.4)$$

Con frecuencia nos referimos a la distribución *a posteriori* usando la ecuación (2.4) donde  $f(\theta|\mathbf{x})$  es proporcional al producto entre la verosimilitud  $\mathcal{L}_n(\theta)$  y la distribución *a priori* para  $\theta$ . No obstante, la constante de normalización juega un papel importante a la hora de calcular la distribución *a posteriori*. En algunos casos dicha constante se puede deducir observando la forma (2.4), por ejemplo cuando se utilizan distribuciones *a priori* conjugadas. Cuando esto no es posible, es necesario calcular la integral correspondiente ya sea de forma analítica o numérica. Encontrar la constante de normalización es más complicado, e incluso

imposible, cuando el número de parámetros a estimar aumenta, o bien, cuando la función de verosimilitud proviene de un modelo complejo. En estos casos, en lugar de calcular directamente la constante se recurre a un algoritmo de simulación para explorar las características de la distribución *a posteriori*. En la Sección 2.3 se describirán los métodos de Monte Carlo, los cuales son de gran utilidad en estos problemas.

### 2.2.2. Distribución *a priori*

La elección de la distribución de probabilidad *a priori* refleja el grado de creencia que se tiene sobre el parámetro desconocido, es decir, qué tanta información se conoce sobre éste. Sin embargo, dicha elección es subjetiva por lo que surge la interrogante: ¿cómo elegir la distribución *a priori*? Y aún más, ¿es correcta dicha elección?

Una alternativa viable, cuando no hay información previa sólida, es definir una “*a priori* no informativa” o “plana”, es decir, darle el mismo peso a todos los posibles valores de  $\theta$ . Una distribución *a priori* no informativa podría ser una función constante  $k > 0$  o la distribución Uniforme. Sin embargo, se debe tener en cuenta lo siguiente al elegir este tipo de distribución *a priori*. Si el parámetro  $\theta$  está definido en  $(-\infty, \infty)$  y la función de densidad *a priori* es  $f(\theta) \propto k$ , con  $k > 0$  constante, se observa que  $\int f(\theta) d\theta = \infty$ , por lo que  $f$  no es una función de densidad. A este tipo de distribuciones *a priori* se le conoce como distribución *a priori impropias*. No obstante, en algunos casos se puede calcular la distribución *a posteriori* utilizando el teorema de Bayes. Es importante tener en claro que las distribuciones *a priori impropias* no son un problema siempre y cuando la distribución *a posteriori* resultante sea una distribución de probabilidad bien definida.

Utilizar una *a priori* no informativa no siempre es la mejor opción. Existen situaciones en las que se desea que la distribución *a priori* cumpla la propiedad de invarianza, es decir, una distribución de probabilidad *a priori* que de la misma información sobre un parámetro y transformaciones de él. Por ejemplo, la distribución Uniforme no es invariante frente a transformaciones. Ante la discusión anterior, Jeffreys propone construir una distribución *a priori* de la forma  $f(\theta) \propto I(\theta)^{1/2}$ , donde  $I(\theta)$  es la función de información de Fisher. Esta propuesta es una transformación invariante. En general, en un problema de múltiples parámetros, la *a priori* de Jeffreys está definida por  $f(\boldsymbol{\theta}) \propto \sqrt{\det|I(\boldsymbol{\theta})|}$ , [Jeffreys, 1946].

Cuando la información previa sobre el modelo es bastante limitada, la distribución *a priori* a menudo se elige de una familia paramétrica  $\mathcal{F}$  con la siguiente característica: dada una *a priori*  $f(\theta)$  que pertenece a  $\mathcal{F}$ , la distribución *a posteriori*  $f(\theta|x)$  también pertenece a  $\mathcal{F}$ . Cuando esto ocurre, decimos que la *a priori* es conjugada [Gelman et al., 2013].

Los incisos a) y b) del siguiente ejemplo muestran la elección de una *a priori* conjugada y la *a priori* de Jeffreys, respectivamente.

**Ejemplo 1.** Sea  $X_1, \dots, X_n \sim \text{Poisson}(\lambda)$ :

a) Suponer que, *a priori*,  $\lambda \sim \text{Gamma}(\alpha, \beta)$ . Sea  $\mathbf{x} = (x_1, \dots, x_n)$ . Por teorema de Bayes,

la a posteriori toma la forma

$$\begin{aligned}
 f(\lambda|\mathbf{x}) &\propto \mathcal{L}_n(\mathbf{x}|\lambda)f(\lambda) \\
 &\propto \left[ \frac{\exp(-n\lambda) \lambda^{\sum x_i}}{\prod x_i!} \right] \left[ \frac{\beta^\alpha}{\Gamma(\alpha)} \lambda^{\alpha-1} \exp(-\beta\lambda) \right] \\
 &\propto \exp(-n\lambda) \lambda^{\sum x_i} \beta^\alpha \lambda^{\alpha-1} \exp(-\beta\lambda) \\
 &\propto \beta^\alpha \lambda^{\sum x_i + \alpha - 1} \exp(-\lambda(n + \beta)).
 \end{aligned}$$

Observando la última ecuación se deduce que la densidad a posteriori es una Gamma con parámetros  $\alpha' = \alpha + \sum x_i$  y  $\beta' = \beta + n$ . Por lo tanto,

$$f(\lambda|\mathbf{x}) = \frac{(\beta')^{\alpha'}}{\Gamma(\alpha')} \lambda^{\alpha'-1} \exp(-\beta'\lambda).$$

Como  $f(\lambda)$  y  $f(\lambda|\mathbf{x})$  pertenecen a la familia Gamma, entonces la distribución Gamma es conjugada para el modelo Poisson.

- b)** Para encontrar la a priori de Jeffrey's notar que la densidad de una Poisson( $\lambda$ ) es  $f(x; \lambda) = \frac{e^{-\lambda} \lambda^x}{x!}$ , al tomar el logaritmo de  $f(x; \lambda)$  se obtiene que  $\log(f(x; \lambda)) = -\lambda + x \log(\lambda) - \log(x!)$ . Tomando la primer derivada de lo anterior se tiene que  $\frac{\partial}{\partial \lambda} \log(f(x; \lambda)) = -1 + \frac{x}{\lambda}$ . Entonces, la información de Fisher esta dada por

$$\mathbb{I}_n(\theta) = \sum_{i=1}^n \mathbb{V}_\lambda \left[ -1 + \frac{x_i}{\lambda} \right] = \sum_{i=1}^n \left( \frac{1}{\lambda} \right)^2 \mathbb{V}_\lambda [x_i] = \frac{n}{\lambda}.$$

Por otro lado, utilizando la a priori  $f(\lambda) \propto \frac{n}{\lambda}$  y aplicando el teorema de Bayes

$$\begin{aligned}
 f(\lambda|\mathbf{x}) &\propto \mathcal{L}_n(\mathbf{x}|\lambda)f(\lambda) \\
 &\propto \left[ \frac{\exp(-n\lambda) \lambda^{\sum x_i}}{\prod x_i!} \right] \left( \frac{n}{\lambda} \right)^{1/2} \\
 &\propto \exp(-n\lambda) \lambda^{\sum x_i} \frac{1}{\lambda^{1/2}} \\
 &\propto \exp(-n\lambda) \lambda^{\sum x_i + 1/2 - 1}.
 \end{aligned}$$

Por lo tanto, la densidad a posteriori es una Gamma( $\tilde{\alpha}, \tilde{\beta}$ ) con  $\tilde{\alpha} = \sum x_i + \frac{1}{2}$  y  $\tilde{\beta} = n$ ,

$$f(\lambda|\mathbf{x}) = \frac{\tilde{\beta}^{\tilde{\alpha}}}{\Gamma(\tilde{\alpha})} \lambda^{\tilde{\alpha}-1} \exp(-\tilde{\beta}\lambda).$$

Este ejemplo muestra cómo al tomar dos *a priori* distintas, la *a posteriori* resultante pertenece a la misma familia de distribuciones (*Gamma*), pero con distintos parámetros.

Como se ha visto, la elección de la distribución *a priori* depende del conocimiento previo acerca del parámetro. Por ejemplo, si el único conocimiento acerca del parámetro es que está definido como un número real positivo, se podría elegir una distribución *Gamma*. Por otro lado, cuando  $\theta$  es un vector de parámetros desconocidos se tendría que definir una función de densidad conjunta sobre  $\theta$ ; esto en la práctica es difícil puesto que es complicado saber cómo interactúan los parámetros *a priori*. Es de importancia realizar un análisis de sensibilidad sobre la *a priori* para evaluar si los resultados se ven o no afectados por la elección de la información previa. Si la función de verosimilitud no tiene mucha información sobre los parámetros, entonces la elección de la distribución *a priori* debe hacerse con cuidado.

### 2.2.3. Distribución *a posteriori* predictiva

Cuando el objetivo de un estudio es realizar inferencia sobre datos aún no observados nos encontramos con un problema de *predicción*, es decir, se busca predecir los valores de observaciones nuevas dada la información disponible.

Sea  $\mathbf{x} = (x_1, \dots, x_n)$  una muestra de datos generada por la variable aleatoria  $X$ , cuya función de densidad está dada por  $f(\mathbf{x}|\theta)$ , que depende del vector de parámetros desconocidos  $\theta$ . Sea  $f(\theta|\mathbf{x})$  la función de densidad *a posteriori* para  $\theta$ . Ahora bien, supóngase que se desea predecir el valor de una observación nueva  $x_{n+1}$ , y supóngase que las observaciones  $\mathbf{x}$  y  $x_{n+1}$  son condicionalmente independientes dado  $\theta$ . Con base en los supuestos anteriores y aplicando el teorema de Bayes, se obtiene que la función de densidad conjunta para  $(X_{n+1}, \theta)$  condicional a las observaciones  $\mathbf{x}$  está dada por:

$$\begin{aligned} f(x_{n+1}, \theta|\mathbf{x}) &= f(x_{n+1}|\theta, \mathbf{x})f(\theta|\mathbf{x}) \\ &= f(x_{n+1}|\theta)f(\theta|\mathbf{x}). \end{aligned}$$

Luego, la densidad *a posteriori* predictiva  $f(x_{n+1}|\mathbf{x})$  está dada por la ecuación:

$$f(x_{n+1}|\mathbf{x}) = \int f(x_{n+1}|\theta)f(\theta|\mathbf{x}) d\theta. \quad (2.5)$$

## 2.3. Markov Chain Monte Carlo (MCMC)

Anteriormente se comentó que no siempre es posible encontrar de forma explícita (o analítica) la constante de normalización de la distribución *a posteriori* dada por (2.3). Esto puede deberse a la forma de la verosimilitud y/o la elección de la *a priori*. Los métodos Markov Chain Monte Carlo (MCMC) son una alternativa que permite explorar la distribución *a posteriori*  $f(\theta|\mathbf{x})$  mediante la generación de muestras de  $f(\theta|\mathbf{x})$  sin tener que simular directamente de  $f(\theta|\mathbf{x})$ .

Sea  $\pi$  una distribución de probabilidad de la cual se desea simular, llamada *distribución objetivo*, con soporte  $\mathcal{X} \subseteq \mathbb{R}^k$ . En particular, en inferencia Bayesiana, la distribución *a*

*posteriori*  $f(\theta|\mathbf{x})$  (2.3) representa la función objetivo  $\pi$ . El objetivo de un algoritmo MCMC es construir una cadena de Markov  $\{x^{(t)}\}_{t \in \mathbb{N}}$  cuya distribución ergódica sea precisamente la distribución objetivo  $\pi$ , es decir, que la distribución de la cadena converga a  $\pi$ . Si esto se garantiza, entonces al obtener muestras de la cadena se obtienen muestras de la distribución objetivo. Estos algoritmos son iterativos y requieren el uso de un kernel de transición que garantice la ergodicidad de la cadena.

La teoría detrás de las cadenas de Markov y los métodos MCMC es extensa y explicarla a detalle está fuera del objetivo de este trabajo. En las secciones siguientes nos limitamos a presentar los algoritmos que se usan en este trabajo. Para mayores referencias, el lector puede consultar [Karlin and Taylor, 1975, Resnick, 1992] y [Rincón, 2012].

### 2.3.1. Algoritmo Metrópolis-Hastings

El algoritmo Metrópolis-Hastings (MH) fue introducido en [Hastings, 1970] y actualmente es la base de la mayoría de los métodos MCMC. En términos generales, para obtener muestras de una distribución de probabilidad  $\pi$  con función de densidad  $\pi(x)$  y soporte  $\mathcal{X} \subseteq \mathbb{R}^k$ , el algoritmo MH funciona de la siguiente forma. Partiendo de un valor inicial  $x^{(0)}$  y de forma iterativa se genera una cadena de Markov donde cada iteración consta de dos pasos: 1) proponer un nuevo valor para el siguiente estado de la cadena, y 2) rechazar o aceptar dicho valor con cierta probabilidad; en caso de rechazar el nuevo valor, se mantiene el valor actual de la cadena.

Para formalizar el procedimiento anterior, definimos la distribución condicional *instrumental* o *distribución propuesta*,  $q(\cdot|x)$ , que determinará la probabilidad de transición del estado  $x^{(t)} = x$  al estado  $x^{(t+1)}$ . En la práctica, dicha distribución debe estar definida al menos en todo  $\mathcal{X}$  y se busca que sea fácil de simular de ella. Si la distribución  $q$  no es simétrica, es decir  $q(y|x) \neq q(x|y)$ , entonces es necesario que tenga una forma explícita (salvo la constante de normalización cuando es independiente de  $x$ ).

El algoritmo MH usa la distribución instrumental  $q$  para generar una cadena de Markov  $\{x^{(t)}\}_{t \in \mathbb{N}}$ , cuya distribución estacionaria es  $\pi$ , por medio del siguiente procedimiento. Dado un punto inicial  $x^{(0)} \in \mathcal{X}$ , la  $t$ -ésima iteración consiste en:

1. Simular  $x^*$  de  $q(\cdot|x^{(t)})$ ;
2. Simular  $u$  de una distribución Uniforme en  $(0, 1)$ ;
3. Hacer un cambio de estado

$$x^{(t+1)} = \begin{cases} x^* & \text{si } u \leq \rho(x^*, x^{(t)}), \\ x^{(t)} & \text{si } u > \rho(x^*, x^{(t)}), \end{cases}$$

donde  $\rho$ , llamada probabilidad de aceptación Metrópolis-Hastings, esta dada por

$$\rho(x^{(t)}, x^*) = \min \left\{ 1, \frac{\pi(x^*) q(x^{(t)}|x^*)}{\pi(x^{(t)}) q(x^*|x^{(t)})} \right\}.$$

Es importante resaltar varios aspectos del algoritmo anterior. Una gran ventaja del procedimiento es que las distribuciones  $\pi$  y  $q$  pueden tener o no forma analítica, basta que se puedan evaluar numéricamente en sus argumentos. Más importante aún, el algoritmo MH solo depende de la distribución objetivo a través del cociente

$$\frac{\pi(x^*)}{\pi(x^{(t)})},$$

por lo cual no es necesario conocer la constante de normalización de  $\pi$ . Esta es la razón fundamental por la cual estos métodos son de gran utilidad para la exploración de la distribución *a posteriori*, ya que, como se mencionó antes, la constante de normalización de dicha distribución es difícil o imposible de determinar.

Cabe mencionar que el valor inicial propuesto  $x^{(0)}$  no es relevante en primera instancia ya que la cadena es ergódica, es decir, cualquier valor que esté dentro del soporte  $\mathcal{X}$  conduce a la convergencia de la cadena. No obstante, ciertos valores iniciales hacen más rápida la convergencia a la distribución estacionaria. Por otra parte, dado que los primeros valores de la simulación no pertenecen a la distribución objetivo  $\pi$ , se descarta un cierto número  $k$  de iteraciones conocido como *burn-in*.

Las muestras de la distribución posterior  $x^{(k+1)}, x^{(k+2)}, x^{(k+3)}, \dots, x^{(k+m)}$  obtenidas mediante el algoritmo MH (después de eliminar el *burn-in*) se usan para estimar características como esperanza y varianza *a posteriori* de los parámetros. La teoría clásica estadística hace estas estimaciones a partir de una muestra independiente; no obstante, las muestras que se obtienen no son independientes ya que se generan a través de una cadena de Markov. A pesar de esto, es posible obtener una muestra pseudo-independiente al seleccionar un subconjunto de estados de la cadena cada cierto número  $\tau$  de iteraciones, esto es  $x^{(k+\tau)}, x^{(k+2\tau)}, \dots, x^{(k+n\tau)}$ . Este valor  $\tau$  es de gran utilidad para determinar el número total de iteraciones que debe tener la cadena de Markov. Existen varias alternativas para determinar cuál es un valor adecuado para  $\tau$  como, por ejemplo, el *integrated autocorrelation time* (IAT) [Roberts et al., 2001] el cual puede pensarse como el número de muestras correlacionadas con el mismo poder de reducción de varianza que una muestra independiente. Este valor está asociado a la eficiencia de los algoritmos MCMC, entre más cercano a 1 es el IAT, más eficiente es la cadena, es decir, requiere menos iteraciones.

Para garantizar que  $\pi$  es la distribución estacionaria de la cadena de Markov  $\{x^{(t)}\}_{t \in \mathbb{N}}$  generada por el algoritmo MH, se debe satisfacer como condición mínima, ver [Robert and Casella, 2004], que

$$\mathcal{X} \subset \bigcup_{x \in \mathcal{X}} \text{soporte } q(\cdot|x).$$

A grandes rasgos, esto implica que  $q$  debe estar definida en todo  $\mathcal{X}$ . Aun cuando lo anterior es fácil de garantizar para una gran cantidad de distribuciones  $q$ , la elección es crucial para que el algoritmo sea computacionalmente eficiente. Otros detalles técnicos respecto de la construcción de estos algoritmos pueden ser consultados en [Liu, 2004].

### 2.3.2. T-Walk

En este trabajo se usa un algoritmo MCMC que no requiere determinar los parámetros de la distribución propuesta  $q$ , el cual es conocido como *t-walk*.

A grandes rasgos, este algoritmo está basado en el algoritmo MH, construido sobre el espacio producto del soporte de la distribución objetivo, y diseñado de tal forma que es invariante ante la escala de los parámetros y aproximadamente invariante ante transformaciones afines del espacio de estados. El algoritmo realiza cambios en los estados de la cadena mediante cuatro propuestas Metrópolis-Hastings: *walk move* (o *scaled random walk*), *traverse move*, *hop move* y *blow move*. Las propuestas se eligen de forma aleatoria de acuerdo al vector de probabilidades (0.4918, 0.4918, 0.0082, 0.0082), respectivamente. El movimiento *walk move* es una caminata aleatoria escalada para cada parámetro. El movimiento *traverse move* genera estados de la cadena considerando la correlación entre los parámetros. Por último, los movimientos *traverse move* y *hop move* generan estados extremos que permiten explorar posibles modas de la distribución *a posteriori*.

Los detalles técnicos del algoritmo están fuera del alcance de este trabajo, por lo que referimos al lector interesado en ellos al artículo original [Christen et al., 2010].

## 2.4. Problemas inversos

Hasta ahora, se ha revisado a lo largo de este capítulo el paradigma Bayesiano y los algoritmos MCMC para explorar la distribución *a posteriori* de un parámetro o un vector de parámetros. En esta sección se describirá de forma breve como se aplica la metodología anterior a la problemática central de esta tesis.

El objetivo de este trabajo es encontrar los parámetros de la ecuación diferencial que describe la temperatura en las nubes, la cual fue propuesta en [Flores, 2017]. En la literatura, la estimación de parámetros en ecuaciones diferenciales se clasifica como un *problema inverso*. Para explicar este concepto, primero se tiene que dar la noción de lo que es un *problema directo*. Desde el punto de vista matemático, un *problema directo* consiste en encontrar el resultado (o efecto)  $y$  generado por un conjunto de valores (o causas)  $x$  sometidas a un proceso  $A$ , es decir  $A(x) = y$ . En consecuencia, un *problema inverso* consiste en encontrar los valores (o causas)  $x$  si se conoce el resultado (o efecto)  $y$  y el proceso  $A$ , es decir,  $x = A^{-1}(y)$ . Se considera que este tipo de problemas están *mal planteados* ya que cumplen con una o más de las siguientes condiciones:

- 1) no existe una solución;
- 2) la solución puede no ser única;
- 3) pequeñas variaciones en las causas  $x$  generan fuertes variaciones en los resultados  $y$ .

En la literatura clásica sobre este tipo de problemas, la solución se obtiene mediante un esquema de optimización con restricciones.

Por otro lado, desde el punto de vista estadístico, un *problema inverso* es simplemente un problema de inferencia. Usando la misma notación anterior, el resultado  $y$  representa a

los datos observados los cuales están sujetos a cierta cantidad de ruido  $\epsilon$ , mientras que los parámetros son la causa  $x$ . De esta forma, la solución del problema consiste en estimar la causa  $x$  a partir de los datos  $y$ . Esto puede pensarse como un problema de regresión de la forma:

$$y = A(x) + \epsilon. \quad (2.6)$$

En un esquema de regresión clásico,  $A(x)$  es una función lineal de  $x$ ; sin embargo, en un *problema inverso*  $A(x)$  típicamente representa la solución de un sistema de ecuaciones diferenciales ordinarias o parciales, o una transformación de dicha solución. Desde el punto de vista de inferencia Bayesiana, la solución al *problema inverso* (2.6) es la exploración de la distribución *a posteriori* de los parámetros de interés. Esto se ejemplificará de forma detallada en el Capítulo 3.

Dirección General de Bibliotecas de la UAQ

---

# Inferencia sobre la temperatura potencial

En este capítulo se describe en detalle el modelo atmosférico de ecuaciones diferenciales ordinarias planteado en [Flores, 2017] y su motivación. De la misma forma, se presentan dos modelos estadísticos, propuestos en este trabajo, para la estimación de los parámetros de interés siguiendo las ideas presentadas en el Capítulo 2.

## 3.1. Modelo teórico para el estudio de inestabilidades atmosféricas de parcelas de aire

En [Flores, 2017] se propone un modelo de ecuaciones diferenciales ordinarias para describir la temperatura potencial equivalente de una parcela de aire, basado en las ideas de [Fraedrich, 1985]. El modelo se fundamenta en la teoría de la parcela, la cual considera conceptos como procesos adiabáticos, auto-conversión rápida, no-supersaturación, entre otros. El modelo teórico presenta dos variantes, una determinista y otra estocástica. El objetivo de dicho trabajo es caracterizar las soluciones de ambos modelos, obteniendo expresiones analíticas para el caso determinista y resultados teóricos sobre la media y varianza del caso estocástico. Así mismo, parte de la motivación de dicho trabajo es la existencia de modelos atmosféricos altamente complejos y computacionalmente costosos, cuyas predicciones del tiempo (eventos atmosféricos ocurridos en escalas de tiempo corto, semanas o de menor escala) no son confiables a mediano y largo plazo. En este sentido, los resultados de [Flores, 2017] podrían utilizarse para alimentar estos modelos atmosféricos más complejos haciéndolos más confiables.

Para definir el modelo se requiere la introducción de algunos conceptos. Se denomina parcela a una región de aire húmedo (o seco) de dimensiones infinitesimales con ciertas características; por ejemplo, la parcela está térmicamente aislada de su entorno de modo que

su temperatura cambia adiabáticamente (sin intercambio de calor entre medios) al ascender o descender. Se define la *temperatura potencial* de una parcela de aire como la temperatura que la parcela tendría si se expande o se comprime mediante un proceso adiabático a una presión estándar (por lo general, 1000 *hPa*). Otro concepto clave es la *temperatura potencial equivalente* que mide como la temperatura varía con la presión bajo condiciones de saturación adiabáticas en ascenso o descenso. Si bien la teoría de la parcela supone procesos adiabáticos, el modelo estudiado en [Flores, 2017] considera el intercambio de calor latente  $L_v$  ocurrido durante la condensación y asume una mezcla  $\varepsilon$  entre la parcela y el medio ambiente por medio de una relajación de la temperatura potencial equivalente de la parcela  $\theta_e$  hacia su estado base  $\theta_{e,env}$ . La mezcla o razón de mezcla  $\varepsilon$  mide la interacción entre ambos medios y en la práctica es un valor numérico adimensional [Wallace and Hobbs, 2006]. La temperatura potencial equivalente de la parcela  $\theta_e$  y del medio ambiente  $\theta_{e,env}$  son mediciones en grados Kelvin,  $^{\circ}K$ . Este modelo es representado por la ecuación diferencial (3.1) donde  $\theta_e(0) = \theta_{e,0}$  es la condición inicial, es decir, el valor de la temperatura potencial equivalente de la parcela al nivel superficie  $z = 0$ .

$$\begin{cases} \frac{d\theta_e}{dz} = -\varepsilon [\theta_e(z) - \theta_{e,env}(z)] \\ \theta_e(0) = \theta_{e,0}. \end{cases} \quad (3.1)$$

Este modelo considera que la temperatura potencial equivalente del medio ambiente (3.2) tiene humedad del ambiente exponencial (3.3). Además, se considera una temperatura potencial del ambiente lineal dada por (3.4).

$$\theta_{e,env}(z) = \theta_{env}(z) + \frac{L_v}{C_p} q_{v,env}(z) \quad (3.2)$$

$$q_{v,env}(z) = q_{0,env} \exp(-z/z_0) \quad (3.3)$$

$$\theta_{env}(z) = \theta_{0,env} + Bz \quad (3.4)$$

Con base en los supuestos anteriores, la ecuación diferencial (3.1) puede resolverse mediante el método del factor integrante y la solución está dada de forma analítica por (3.5).

$$\begin{aligned} \theta_e(z) = & \theta_{e,env}(z) - \frac{B}{\varepsilon} + \frac{L_v}{C_p} \left( \frac{1}{\varepsilon z_0 - 1} \right) q_{v,env}(z) + \\ & \left[ \theta_{e,0} - \theta_{0,env} + \frac{B}{\varepsilon} - \frac{L_v}{C_p} \left( \frac{\varepsilon z_0}{\varepsilon z_0 - 1} \right) q_{0,env} \right] \exp(-\varepsilon z). \end{aligned} \quad (3.5)$$

En (3.5) se muestra que la temperatura potencial equivalente de la parcela  $\theta_e(z)$  queda en función de la temperatura potencial equivalente del medio ambiente  $\theta_{e,env}(z)$ , la cual, a su vez, depende de la altura de la parcela  $z$  (*km*). Asimismo,  $\theta_e(z)$  está sujeta a los parámetros: nivel de equilibrio de la parcela de aire  $z_0$  (*km*), calor latente de condensación a  $0^{\circ} C$ ,  $L_v$  ( $J kg^{-1}$ ), calor específico de aire seco a presión constante  $C_p$  ( $J kg^{-1} ^{\circ}K$ ), razón de enfriamiento  $B$  ( $^{\circ}K km^{-1}$ ), razón de mezcla  $\varepsilon$  y de los valores referenciales al nivel superficie  $z = 0$  tales como temperatura potencial del medio ambiente  $\theta_{0,env}$  ( $^{\circ}K$ ), temperatura potencial equivalente de

la parcela  $\theta_{e,0}$  ( $^{\circ}K$ ) y vapor de agua del medio ambiente  $q_{0,env}$  ( $g/kg$ ). Los valores de los parámetros conocidos se muestran en la Tabla 3.1.

Los parámetros desconocidos son  $\varepsilon$ ,  $q_{0,env}$  y  $B$ . La razón de mezcla  $\varepsilon$  mide la cantidad de vapor de agua en un cierto volumen de aire y se expresa en  $g/kg$ ; en la atmósfera la magnitud de  $\varepsilon$  varia poco en latitudes medias y llega hasta valores alrededor de  $20g/kg$  en los trópicos [Wallace and Hobbs, 2006]. La cantidad de vapor de agua del medio ambiente  $q_{0,env}$  a nivel superficie  $z = 0$  también se expresa en unidades de  $g/kg$ . Ambas medidas  $\varepsilon$  y  $q_{0,env}$  en la práctica son expresadas de forma adimensional, esto es,  $g/g$ , por lo que sus rangos de valores cambian y son menores a 1. Por otra parte, la razón de enfriamiento  $B$  en una columna de aire varia en promedio de 6 a 7  $^{\circ}K km^{-1}$  en la troposfera, sin embargo, toma una amplia gama de valores en distintas ubicaciones [Wallace and Hobbs, 2006].

Los parámetros  $\varepsilon$ ,  $q_{0,env}$  y  $B$  son el objeto de estudio de esta tesis y se pretende estimar su valor por medio de inferencia Bayesiana. En las subsecciones posteriores, veremos dos propuestas de modelos estadísticos a partir de la ecuación (3.5) para encontrar tal estimación.

Parámetro	Valor	Descripción
$z_0$	$2 km$	Nivel de equilibrio de la parcela de aire. Indica que tan rápido es el decremento en (3.3) respecto la altura.
$L_v$	$2.5 \times 10^6 J kg^{-1}$	Calor latente de condensación (o vaporización) a $0^{\circ}C$ .
$C_p$	$1,004 J kg^{-1} ^{\circ}K^{-1}$	Calor específico del aire seco a presión constante.
$\theta_{0,env}$	$300 ^{\circ}K$	Temperatura potencial del medio ambiente al nivel superficie $z = 0$ .
$\theta_{e,0}$	$345 ^{\circ}K$	Temperatura potencial equivalente de la parcela al nivel superficie $z = 0$ .

Tabla 3.1: Parámetros conocidos del modelo diferencial (3.1).

## 3.2. Modelo Estadístico 1

Como se comentó antes, la estimación de los parámetros que determinan un sistema de ecuaciones diferenciales se considera un *problema inverso*. Para resolver el problema desde el punto de vista estadístico se requiere entender la naturaleza de los datos observados y

su relación con los parámetros. En esta sección presentamos el primero de dos modelos estadísticos para realizar la tarea de estimación en el contexto descrito en la Sección 3.1.

Sean  $y_1, y_2, \dots, y_n$  observaciones de la temperatura de las nubes en la  $i$ -ésima altura  $z_i$ , y cuyas unidades son  $^\circ K$ , es decir,  $Y(z_i) = y_i$ ,  $i = 1, \dots, n$ . Note que estas observaciones provienen de un proceso estocástico desconocido por lo que no es posible considerar que son independientes. Por otro lado, sea  $A(z_i|\boldsymbol{\vartheta})$  la solución de la ecuación (3.5), donde  $\boldsymbol{\vartheta} = (\varepsilon, q_{0,env}, B)$  es el vector de parámetros desconocidos de la ecuación. Dada una altura fija  $z_i$  y condicional al valor de los parámetros  $\boldsymbol{\vartheta}$ , suponemos que

$$Y(z_i) = A(z_i|\boldsymbol{\vartheta}) + \xi_i, \quad \text{para } i = 1, \dots, n, \quad (3.6)$$

donde  $\xi_i$  es un término de error que sigue una distribución Normal con media  $\mu = 0$  y precisión desconocida  $\tau = 1/\sigma^2$  donde  $\sigma^2$  es la varianza de  $\xi_i$ . Esto implica que  $Y(z_i)|\boldsymbol{\vartheta}, \tau \sim N(A(z_i|\boldsymbol{\vartheta}), \tau)$ . De manera simple, este modelo implica que el valor observado de la temperatura es, en promedio, la solución de la ecuación diferencial más un error. Más aún, suponemos que el error  $\xi_k$  es independiente de  $\xi_j \forall k \neq j$  y que, dado el vector de parámetros  $\boldsymbol{\vartheta}$ , las observaciones  $y_1, \dots, y_n$  son condicionalmente independientes.

### 3.2.1. Función de Máxima Verosimilitud

Con base en lo anterior, la función de máxima verosimilitud está dada por

$$\begin{aligned} \mathcal{L}_n(\boldsymbol{\vartheta}, \tau) &= \prod_{i=1}^n f(y_i; A(z_i|\boldsymbol{\vartheta}), \tau) \\ &= \prod_{i=1}^n \frac{\tau^{1/2}}{\sqrt{2\pi}} \exp\left(-\frac{\tau}{2} [y_i - A(z_i|\boldsymbol{\vartheta})]^2\right) \\ &= (2\pi)^{-n/2} \tau^{n/2} \exp\left(-\frac{\tau}{2} \sum_{i=1}^n [y_i - A(z_i|\boldsymbol{\vartheta})]^2\right) \end{aligned} \quad (3.7)$$

donde  $f$  denota una distribución Normal univariada con media  $\mu_i = A(z_i|\boldsymbol{\vartheta})$  y precisión  $\tau$ . Es importante enfatizar que cada observación  $y_i$  tiene su propia media  $\mu = A(z_i|\boldsymbol{\vartheta})$  y una precisión desconocida constante  $\tau$  para cada altura  $z_i$ .

En este punto ya es posible realizar inferencia sobre  $\boldsymbol{\vartheta}$  y  $\tau$  encontrando el vector de parámetros que maximiza la función de verosimilitud, posiblemente bajo algunas restricciones que se asemejan a los modelos de regresión Ridge y Lasso [Hastie et al., 2013]. A pesar de que el parámetro  $\tau$  no es parte de los parámetros de interés ya que no está relacionado con la ecuación diferencial, su estimación es de gran relevancia ya que nos permitirá cuantificar la incertidumbre sobre el vector  $\boldsymbol{\vartheta}$ . Este hecho es el que distingue los procedimientos puramente de calibración de los procedimientos estadísticos, donde estos últimos aportan mayor información. No obstante, en este trabajo se opta por usar el enfoque Bayesiano de inferencia

por lo que es necesario proponer una distribución *a priori* conjunta para  $(\vartheta, \tau) = \psi$ .

### 3.2.2. Función de probabilidad *a priori*

Para determinar la distribución *a priori* del vector de parámetros  $\psi = (\vartheta, \tau)$  se tiene que codificar el conocimiento previo de cada parámetro y de sus interacciones en una distribución de probabilidad conjunta. En la práctica esto suele ser difícil por lo que se opta por suponer que, *a priori*, todos los parámetros son independientes. En este escenario es posible analizar individualmente los parámetros y determinar una distribución *a priori* marginal para cada uno de ellos.

Siguiendo el análisis de [Flores, 2017], los parámetros  $\varepsilon$ ,  $q_{0,env}$  y  $B$  solo toman valores positivos. Dado que los parámetros  $\varepsilon$  y  $q_{0,env}$  están acotados superiormente por uno, entonces se les asignó una distribución *a priori* no informativa Uniforme en  $(0, 1)$ . En el caso de los parámetros  $B$  y  $\tau$  se eligió una distribución *a priori* Gamma eligiendo parámetros de escala  $\alpha_k$  y forma  $\beta_k$ ,  $k = 1, 2$ , respectivamente, de forma que las *a priori* resultantes sean débilmente informativas, es decir, que tengan colas pesadas a la derecha. La información completa sobre las distribuciones *a priori* se muestra en la Tabla 3.2.

Parámetro	Función de densidad <i>a priori</i>
$\varepsilon$	U(0, 1)
$q_{0,env}$	U(0, 1)
$B$	Gamma ( $\alpha_1 = 2, \beta_1 = 14$ )
$\tau$	Gamma ( $\alpha_2 = 8/5, \beta_2 = 250$ )

Tabla 3.2: Distribuciones *a priori* de los parámetros desconocidos del modelo estadístico 1.

Debido al supuesto de independencia de los parámetros, la distribución *a priori* conjunta puede escribirse como

$$\begin{aligned}
 f(\psi) &= f(\varepsilon)f(q_{0,env})f(B)f(\tau^2) \\
 &\propto B^{\alpha_1-1} \exp\left(-\frac{B}{\beta_1}\right) (\tau^{-1})^{\alpha_2-1} \exp\left(-\frac{\tau^{-1}}{\beta_2}\right).
 \end{aligned} \tag{3.8}$$

### 3.2.3. Función de probabilidad *a posteriori*

Una vez establecidas la distribución *a priori* conjunta (3.8) y la función de máxima verosimilitud (3.7) para el modelo estadístico 1, entonces la función de densidad *a posteriori*

queda determinada por la siguiente expresión:

$$\begin{aligned}
f(\boldsymbol{\psi}|Y) &\propto \mathcal{L}_n(\boldsymbol{\psi})f(\boldsymbol{\psi}) \\
&\propto \mathcal{L}_n(\boldsymbol{\theta})f(\varepsilon)f(q_{0,env})f(B)f(\tau) \\
&\propto \tau^{n/2} \exp\left(-\frac{\tau}{2} \sum_{i=1}^n [y_i - A(z_i|\boldsymbol{\vartheta})]^2\right) \times \\
&\quad B^{\alpha_1-1} \exp\left(-\frac{B}{\beta_1}\right) (\tau^{-1})^{\alpha_2-1} \exp\left(-\frac{\tau^{-1}}{\beta_2}\right) \quad (3.9)
\end{aligned}$$

Es importante enfatizar que la distribución *a posteriori* no tiene una forma simple a pesar de que la solución de la ecuación diferencial  $A(z_i|\boldsymbol{\vartheta})$  es analítica. Debido a esto, el cálculo de la constante de normalización de la ecuación (3.9) es complicado y se requiere del uso de métodos numéricos para explorar la distribución *a posteriori*. En este trabajo se usa el algoritmo *t-walk* del cual se habló en la Sección 2.3.2.

Por cuestiones de estabilidad numérica, los métodos MCMC utilizan el logaritmo de la función de densidad *a posteriori*. De modo que el logaritmo de (3.9), denotado por  $l(\boldsymbol{\psi}) = \log(f(\boldsymbol{\theta}|Y))$ , está dado por (3.10).

$$\begin{aligned}
l(\boldsymbol{\psi}) &\propto \left(\frac{n}{2} - \alpha_2 + 1\right) \log(\tau) + (\alpha_1 - 1) \log(B) - \\
&\quad \left(\frac{B}{\beta_1} + \frac{\tau}{\beta_2} + \frac{\tau}{2} \sum_{i=1}^n [y_i - A(z_i|\boldsymbol{\vartheta})]^2\right) \quad (3.10)
\end{aligned}$$

### 3.2.4. Distribución *a posteriori* predictiva

La distribución *a posteriori* predictiva se definió en la Sección 2.2.3 y, en el contexto del modelo estadístico 1, permite predecir la distribución de la temperatura potencial a una altura nueva, digamos  $z_{n+1}$ . La función de densidad *a posteriori* predictiva estaría dada por la integral:

$$\begin{aligned}
f(y_{n+1}|\mathbf{y}) &= \int f(y_{n+1}|\boldsymbol{\psi})f(\boldsymbol{\psi}|\mathbf{y}) d\boldsymbol{\psi} \quad (3.11) \\
&\propto \int \tau^{1/2} \exp\left(-\frac{\tau}{2} [y_{n+1} - A(z_{n+1}|\boldsymbol{\vartheta})]^2\right) \times \\
&\quad \tau^{n/2} \exp\left(-\frac{\tau}{2} \sum_{i=1}^n [y_i - A(y_i|\boldsymbol{\vartheta})]^2\right) \times \\
&\quad B^{\alpha_1-1} \exp\left(-\frac{B}{\beta_1}\right) (\tau^{-1})^{\alpha_2-1} \exp\left(-\frac{\tau^{-1}}{\beta_2}\right) d\boldsymbol{\psi}.
\end{aligned}$$

Nuevamente, dado que no conocemos la constante de normalización de la distribución *a posteriori*, entonces no es posible obtener una versión analítica de la distribución *a posteriori*

*predictiva*. No obstante, es fácil obtener simulaciones de esta última. El proceso consiste iterar los siguientes dos pasos para  $j = 1, \dots, m$ :

1. Simular  $\boldsymbol{\psi}^{(j)} = \left( \varepsilon^{(j)}, q_{0,env}^{(j)}, B^{(j)}, \tau^{(j)} \right)$  de la distribución *a posteriori*.
2. Para una altura fija  $z_{n+1}$ , simular  $y_{n+1}^{(j)}$  de una distribución Gaussiana con parámetros media  $A\left(z_{n+1} | \varepsilon^{(j)}, q_{0,env}^{(j)}, B^{(j)}\right)$  y precisión  $\tau^{(j)}$ .

Dado que ya se tienen muestras de la distribución *a posteriori* obtenidas mediante los algoritmos MCMC, las muestras de la distribución *a posteriori predictiva* no representan ningún problema; no obstante, este proceso puede ser lento si la solución de la ecuación es difícil, en términos computacionales.

Ya que se tienen las muestras de la distribución *a posteriori predictiva*, entonces se puede evaluar el valor medio de la predicción utilizando, por ejemplo,  $E[y_{n+1}] \approx \frac{1}{m} \sum_{i=1}^m y_{n+1}^{(j)}$ . Más aún, los cuantiles empíricos de estas muestras pueden ser usados para estimar la incertidumbre de la predicción, es decir, permiten obtener intervalos de probabilidad predictivos.

### 3.3. Modelo Estadístico 2

El modelo estadístico 1 considera que, para cada altura  $z$ , la variabilidad de la temperatura potencial es constante y está dada por  $1/\tau$ . En la práctica, esta hipótesis es difícil de justificar. Por esta razón, el modelo estadístico 2 que se presenta en esta sección modifica esta hipótesis y permite que la variabilidad cambie con la altura.

Utilizando la misma notación de la sección anterior suponga que, para una altura fija  $z_i$ , la temperatura potencial a esa altura  $Y(z_i)$ , condicional a los parámetros  $\boldsymbol{\vartheta} = (\varepsilon, q_{0,env}, B)$  y  $d_\theta$ , puede ser descrita por

$$Y(z_i) = A(z_i | \boldsymbol{\vartheta}) + \eta(z_i | d_\theta), \quad \text{para } i = 1, \dots, n, \quad (3.12)$$

donde  $A(z_i | \boldsymbol{\vartheta})$  sigue representando la solución de la ecuación (3.1). No obstante, ahora  $\eta(z_i | d_\theta)$  representa una variable aleatoria Gaussiana con media cero y varianza dada por

$$\sigma^2(z_i) = \frac{d_\theta^2}{2\varepsilon} [1 - \exp(-2\varepsilon z_i)]; \quad (3.13)$$

es decir, la varianza depende de la altura, de la razón de mezcla  $\varepsilon$  y un parámetro  $d_\theta$  de dispersión.

Note que la diferencia principal entre los modelos estadísticos 1 y 2 radica en la forma del término estocástico de error. Mientras que en el modelo 1 se tiene un error con varianza constante, el modelo 2 permite un término de error más flexible que cambia con la altura. El segundo modelo está motivado por la versión estocástica de la ecuación diferencial (3.1) propuesta en [Flores, 2017], la cual puede verse en la ecuación (3.14). Aquí se introduce un término estocástico  $dW$  cuya amplitud está determinada por el parámetro  $d_\theta$ . Se puede

demostrar que la solución de la ecuación estocástica tiene como valor esperado la solución de la ecuación determinista  $A(\cdot|\boldsymbol{\theta})$  para cualquier altura  $z$ . Más aún, la varianza de la solución está dada por la ecuación (3.13) para cada altura  $z$ . De este último hecho se deriva el modelo estadístico 2.

$$\frac{d\theta_e}{dz} = -\varepsilon [\theta_e(z) - \theta_{e,env}(z)] + d_\theta dW \quad (3.14)$$

En este punto el lector podría argumentar que el modelo estocástico es más realista ya que considera términos de error en la dinámica. Entonces ¿por qué no se usa el modelo estocástico para la estimación de parámetros? Existen dos razones principales por las que no se siguió esta ruta. La primera es que consideramos que el modelo estocástico no añade ninguna propiedad diferente al modelo. Es decir, el término estocástico sólo introduce ruido a las soluciones, pero no crea dinámicas distintas. La Figura 3.1 muestra la solución del modelo estocástico y se aprecia claramente que la diferencia principal respecto al modelo determinista son los pequeños picos que presenta la curva. La segunda razón es que la

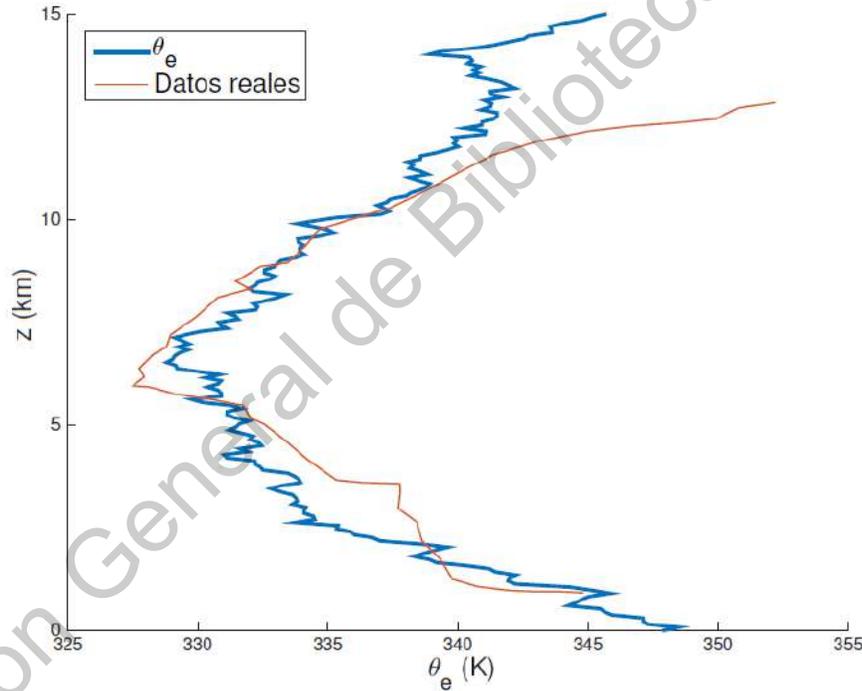


Figura 3.1: Comparación entre el modelo estocástico (3.14) (línea azul) y datos reales (línea roja). Tomado de “Teoría de la parcela de fluidos para el estudio de inestabilidades atmosféricas y distribución de alturas de nubes” (p. 63) por Flores [2017], Querétaro, Qro. México.

estimación de parámetros en ecuaciones diferenciales, desde el punto de vista estadístico, se vuelve un problema de computo super-intensivo. En una ecuación estocástica, la solución de la ecuación es una variable aleatoria con una distribución de probabilidad que no tiene forma

analítica salvo en ejemplos aislados. Esto implica que la función de verosimilitud no pueda ser evaluada ni siquiera de forma numérica. En estos casos se debe recurrir a métodos que aproximen la función de verosimilitud o métodos libres de verosimilitud como Approximated Bayesian Computation (ABC) [Turner and Van Zandt, 2012].

Para poder trabajar con el modelo 2, es necesario inferir los parámetros desconocidos  $\varepsilon$ ,  $q_{0,env}$ ,  $B$  y  $d_\theta$ . Entonces, el conjunto de parámetros a estimar se denotará por el vector aleatorio  $\boldsymbol{\nu} = (\varepsilon, q_{0,env}, B, d_\theta) = (\boldsymbol{\vartheta}, d_\theta)$ .

### 3.3.1. Función de Máxima Verosimilitud

De la ecuación (3.12) se sigue que las observaciones  $y_1, y_2, \dots, y_n$  siguen una distribución Normal con media  $A(z_i|\boldsymbol{\vartheta})$  y varianza dada por (3.13). Asumiendo independencia condicional, la función de máxima verosimilitud está determinada por la ecuación (3.15)

$$\mathcal{L}(\boldsymbol{\nu}) = (2\pi\sigma^2(z_i))^{-n/2} \exp\left(-\frac{1}{2\sigma^2(z_i)} \sum_{i=1}^n [y_i - A(z_i|\boldsymbol{\vartheta})]^2\right). \quad (3.15)$$

Observe que, a diferencia del modelo 1, (3.6), las observaciones  $Y(z_i) = y_i$  tienen su propia media y varianza en cada altura  $z_i$ .

### 3.3.2. Función de probabilidad *a priori*

Nuevamente se recurre al supuesto de independencia entre los componentes del vector  $\boldsymbol{\nu}$ , por lo que se le asignará una distribución *a priori* univariada a cada componente. Ya se han analizado las características de los parámetros  $\varepsilon$ ,  $q_{0,env}$  y  $B$ , resta entonces analizar  $d_\theta$ . De acuerdo a [Flores, 2017],  $d_\theta$  mide la amplitud del término estocástico y toma valores positivos; por ende, se le asignará una *a priori Gamma* con parámetro de forma  $\alpha_3 = 2$  y de escala  $\beta_3 = 250$ , lo cual resulta en una *a priori* poco informativa. Los detalles de las distribuciones *a priori* pueden verse en la Tabla 3.3. Así, la función de densidad *a priori* conjunta es:

$$\begin{aligned} f(\boldsymbol{\nu}) &= f(\varepsilon)f(q_{0,env})f(B)f(d_\theta) \\ &\propto B^{\alpha_1-1} \exp\left(-\frac{B}{\beta_1}\right) d_\theta^{\alpha_3-1} \exp\left(-\frac{d_\theta}{\beta_3}\right). \end{aligned} \quad (3.16)$$

### 3.3.3. Función de probabilidad *a posteriori*

La función de densidad *a posteriori* se obtiene de forma similar al modelo 1. Utilizando el teorema de Bayes, se combina la función de verosimilitud (3.15) con la densidad *a priori*

Parámetro	Función de densidad <i>a priori</i>
$\varepsilon$	U(0, 1)
$q_{0,env}$	U(0, 1)
$B$	Gamma ( $\alpha_1 = 2, \beta_1 = 14$ )
$d_\theta$	Gamma ( $\alpha_3 = 2, \beta_3 = 250$ )

Tabla 3.3: Distribuciones *a priori* de los parámetros desconocidos del modelo estadístico 2.

conjunta (3.16) obteniendo

$$\begin{aligned}
f(\boldsymbol{\nu}|\mathbf{y}) &\propto \mathcal{L}(\boldsymbol{\nu})f(\varepsilon)f(q_{0,env})f(B)f(d_\theta) \\
&\propto [\sigma^2(z_i)]^{-n/2} \exp\left(-\frac{1}{2\sigma^2(z_i)} \sum_{i=1}^n [y_i - A(z_i|\boldsymbol{\vartheta})]^2\right) \times \\
&\quad B^{\alpha_1-1} \exp\left(-\frac{B}{\beta_1}\right) d_\theta^{\alpha_3-1} \exp\left(-\frac{d_\theta}{\beta_3}\right). \tag{3.17}
\end{aligned}$$

De forma análoga al modelo 1, la constante de normalización  $\int \mathcal{L}(\boldsymbol{\nu})f(\boldsymbol{\nu}) d\boldsymbol{\nu}$  no se puede calcular de forma simple. En consecuencia, se explorará a  $f(\boldsymbol{\nu}|Y)$  por medio del algoritmo *t-walk*. Para ello se requiere calcular el logaritmo de (3.17) denotado por  $l(\boldsymbol{\nu}) = \log(f(\boldsymbol{\nu}|Y))$  y dado por (3.18).

$$\begin{aligned}
l(\boldsymbol{\nu}) &\propto (\alpha_1 - 1) \log(B) + (\alpha_3 - 1) d_\theta \\
&\quad - \left( \frac{B}{\beta_1} + \frac{d_\theta}{\beta_3} + \frac{n}{2} \log(\sigma^2(z_i)) + \frac{1}{2\sigma^2(z_i)} \sum_{i=1}^n [y_i - A(z_i|\boldsymbol{\vartheta})]^2 \right) \tag{3.18}
\end{aligned}$$

Finalmente, la función de densidad *a posteriori predictiva* del modelo 2 tampoco tiene una expresión simple, pero es fácil simular de ella utilizando el procedimiento descrito en la Sección 3.2.4.

Ahora que se han definido los dos modelos estadísticos de interés en esta tesis, se puede proceder a mostrar los resultados de la inferencia y la comparación de ambos modelos en un ejemplo con datos reales. Esto se mostrará en el capítulo siguiente.

# Resultados

En este Capítulo se aplica la metodología de estimación descrita en el Capítulo 3 al conjunto de datos reales presentado en [Flores, 2017], el cual, a su vez, fue extraído de [Fraedrich, 1985]. El conjunto de datos reales consiste de  $n = 49$  observaciones de la temperatura en las nubes ( $^{\circ}K$ ) a 49 alturas distintas ( $km$ ) y fueron tomados bajo condiciones atmosféricas específicas. Este conjunto de datos se muestra en el Apéndice A.

## 4.1. Resultados del modelo estadístico 1

Para estudiar la distribución *a posteriori* del vector de parámetros  $\psi$  del modelo 1, se usó el algoritmo *t-walk* en el software libre de computo estadístico R Core Team [2020]. Los códigos utilizados se pueden consultar en el Apéndice B. El algoritmo utiliza como función objetivo el logaritmo de la función de densidad *a posteriori* (3.10) multiplicado por -1, es decir,  $-\log f(\psi|\mathbf{y})$ . En la literatura, dicha función se conoce como función de energía. Se corrió la cadena de Markov por 1,000,000 iteraciones utilizando distintos puntos iniciales aleatorios generados a partir de la distribución *a priori* de  $\psi$ . Las cadenas generadas por el algoritmo *t-walk* convergen rápidamente en este ejemplo, típicamente 1,000 iteraciones son suficientes. No obstante, se descartan los primeros 10,000 pasos de la cadena como burn-in para asegurar la convergencia con distintos puntos iniciales. De las 990,000 iteraciones restantes, se hace un submuestreo cada 103 muestras para obtener una cadena pseudo-independiente. El valor 103 se obtiene del cálculo del IAT sobre la función de energía, el cual es 102.63.

La Figura 4.1 muestra la evolución de función de energía de la cadena resultante, después de omitir el burn-in y realizar el submuestreo. Se puede observar que la función objetivo se estabiliza entre -128 y -118. Por otro lado, se analizó la trayectoria de las cadenas de los parámetros  $\varepsilon$ ,  $q_{0,env}$ ,  $B$  y  $\tau$  mostrando que también convergen a partir de la iteración número 10,000, ver Figura 4.2.

Para visualizar las distribuciones *a posteriori* de los parámetros se realizaron histogramas utilizando 9612 muestras, las cuales pueden observarse en la Figura 4.3. Note que el paráme-

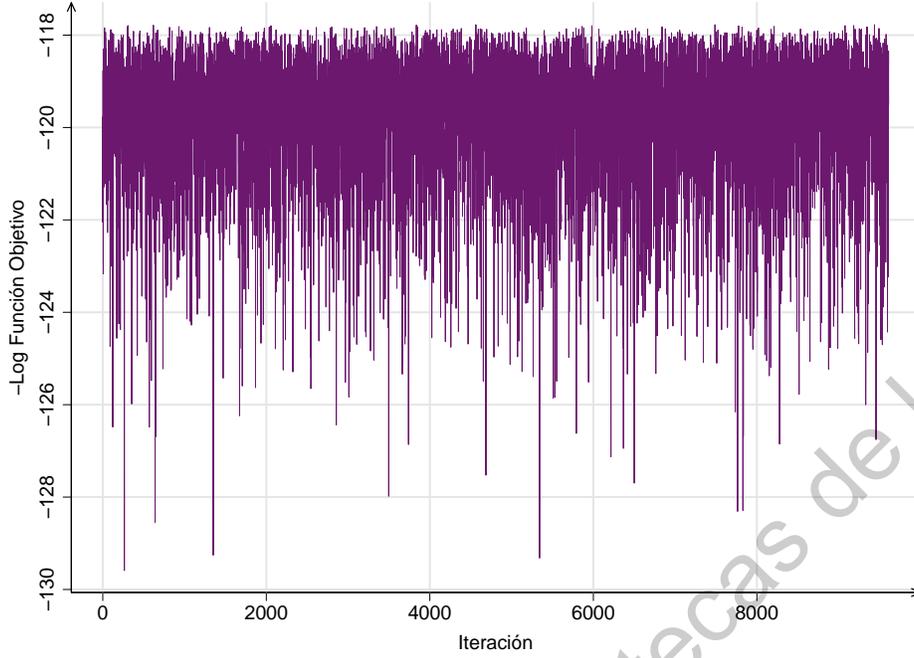


Figura 4.1: Función de energía del modelo estadístico 1 de la cadena resultante, después de haber omitido el burn-in y realizado el submuestreo.

tro de razón de mezcla  $\varepsilon$  tiene mayor probabilidad de tomar valores en el intervalo (0.28, 0.42), ver Figura 4.3a. Es importante mencionar que  $\varepsilon$  se expresa en forma adimensional; si se desea rescatar las unidades basta con multiplicar los valores anteriores por 1,000, lo cual resulta en el intervalo (280, 420) con unidades  $g/kg$ . El valor referencial de vapor de agua del medio ambiente  $q_{0,env}$  al nivel superficie  $z = 0$  toma valores con mayor probabilidad en el intervalo (0.017, 0.023), ver Figura 4.3b. Al igual que  $\varepsilon$ ,  $q_{0,env}$  es un valor adimensional que al multiplicarlo por 1,000 cambia el intervalo a (17, 23) con unidades  $g/kg$ . El parámetro razón de enfriamiento  $B$ , cuyas unidades son  $^{\circ}K km^{-1}$ , toma valores con mayor probabilidad en el intervalo (4.45, 4.80), ver Figura 4.3c. Por último, se observa en la Figura 4.3d que la precisión del modelo 1 toma valores con mayor probabilidad entre (0.15, 0.35). La Figura 4.3 también muestra las distribuciones *a priori* para cada parámetro. Claramente, la distribución *a posteriori* tiene menor variabilidad que la distribución *a priori*, lo cual indica que en efecto existe información acerca de los parámetros en los datos. Más aún, los histogramas tienen una moda claramente definida y con colas ligeras relativamente simétricas. Ahora bien, comparemos los valores puntuales propuestos en [Flores, 2017] y los estimados en esta tesis. Los valores reportados por [Flores, 2017] se obtienen mediante una búsqueda sin ningún procedimiento riguroso de regularización, mucho menos estimación. El objetivo ahí fue ver si el modelo propuesto es capaz de dar una descripción aproximada de los datos. Los valores encontrados se muestran a continuación:

$$\varepsilon = 0.4, \quad q_{0,env} = 0.015, \quad B = 4 \text{ } ^{\circ}K km^{-1}, \quad d_{\theta} = 2. \quad (4.1)$$

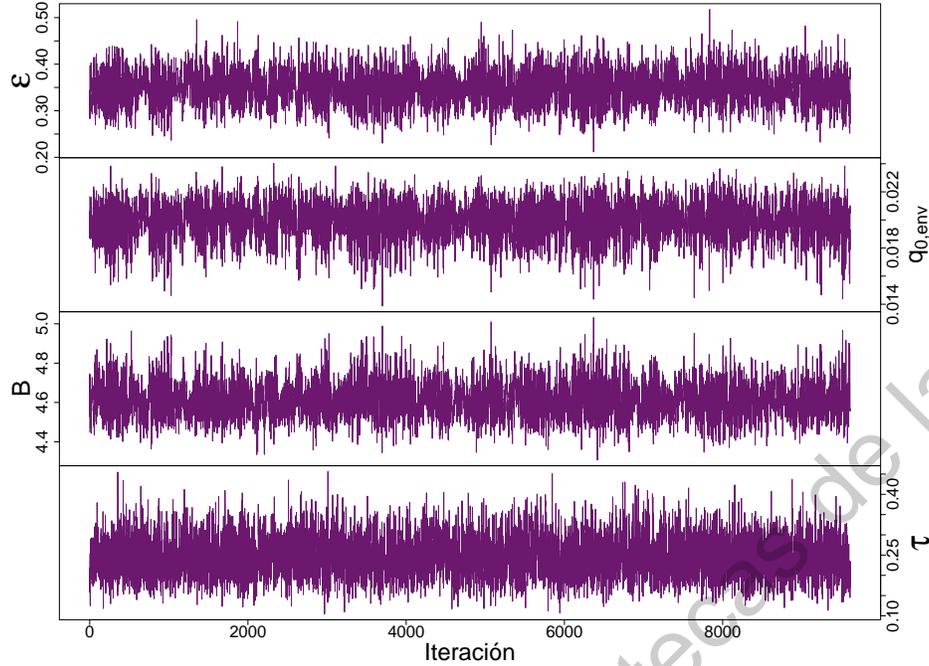
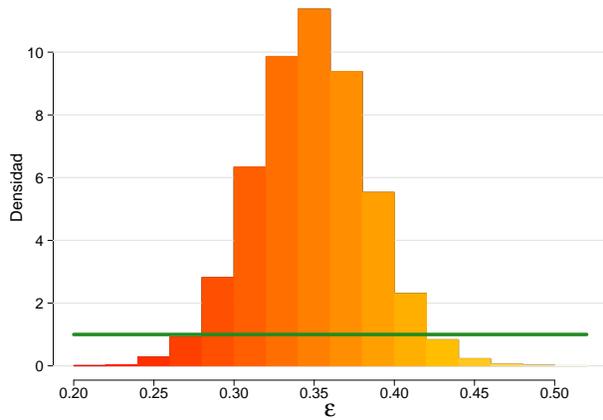


Figura 4.2: Trayectorias de las cadenas de los parámetros  $\varepsilon$ ,  $q_{0,env}$ ,  $B$  y  $\tau$  del modelo estadístico 1, después de haber omitido el burn-in y realizado el submuestreo.

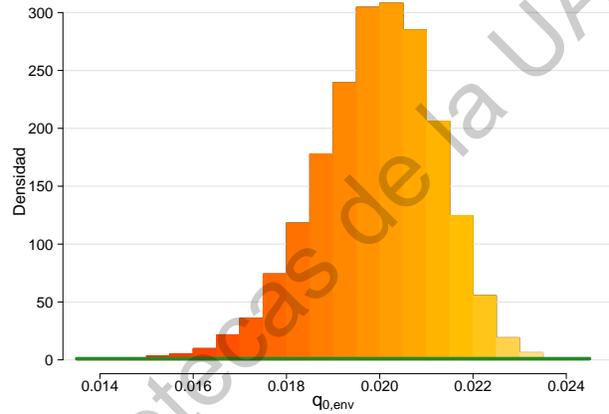
Se observa que un valor de 0.4 para el parámetro razón de mezcla  $\varepsilon$  está dentro de la región de alta probabilidad, ver Figura 4.3a. Por su parte, los valores 0.015 y  $4 \text{ }^\circ\text{K km}^{-1}$  para los parámetros cantidad de vapor de agua en el medio ambiente  $q_{0,env}$  y razón de enfriamiento  $B$ , respectivamente, no están dentro de la región de alta probabilidad, ver Figura 4.3b y 4.3c. Por tanto, dos de los valores dados en (4.1) difieren de los valores altamente probables que se determinaron con el modelo 1.

Una de las ventajas de los procedimientos Bayesianos es que es fácil cuantificar la variabilidad de cualquier función de los parámetros del modelo, en particular, es posible evaluar la variabilidad de las soluciones de la ecuación diferencial generada por la variabilidad en los parámetros estimados. Para esto, basta resolver la ecuación diferencial para cada vector de parámetros  $\vartheta^{(i)}$ ,  $i = 1 \dots, k$ , producido por el algoritmo MCMC. De esta forma se obtiene un abanico de valores de la temperatura potencial para cada altura  $z$ . A partir de estos valores, podemos construir bandas de probabilidad puntuales tomando cuantiles empíricos de las muestras. El resultado de este procedimiento puede verse en la Figura 4.4 donde se muestran cuantiles puntuales de nivel 2.5%, 50% y 97.5% para la solución media de la ecuación, es decir, bandas de probabilidad del 95% para el valor esperado de la dinámica de la temperatura potencial. Se puede ver que, en general, la curva describe la tendencia de los datos.

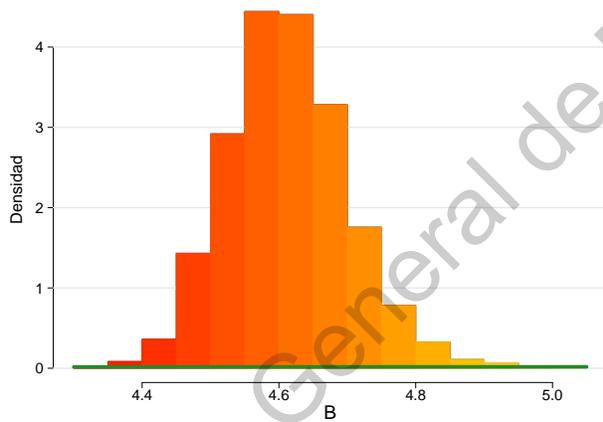
Por otro lado, si se desea crear predicciones para la temperatura en lugar del valor esperado de la temperatura, entonces se puede utilizar la distribución *a posteriori predictiva*



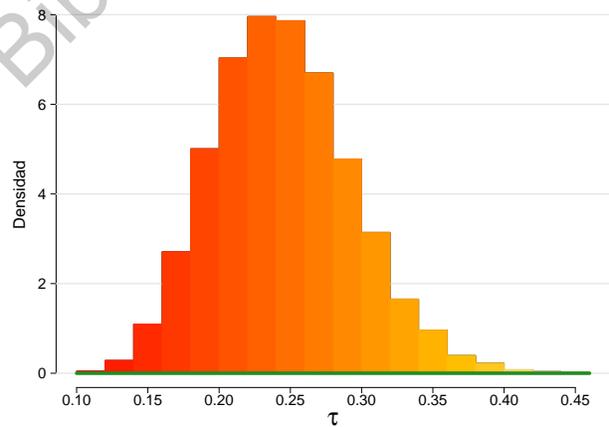
(a) Razón de mezcla.



(b) Cantidad de vapor de agua en el medio ambiente.



(c) Razón de enfriamiento.



(d) Parámetro de precisión del modelo estadístico 1.

Figura 4.3: Histogramas con 9612 muestras de las distribuciones *a posteriori* de los parámetros  $\varepsilon$ ,  $q_{0,env}$ ,  $B$  y  $\tau$  del modelo estadístico 1.

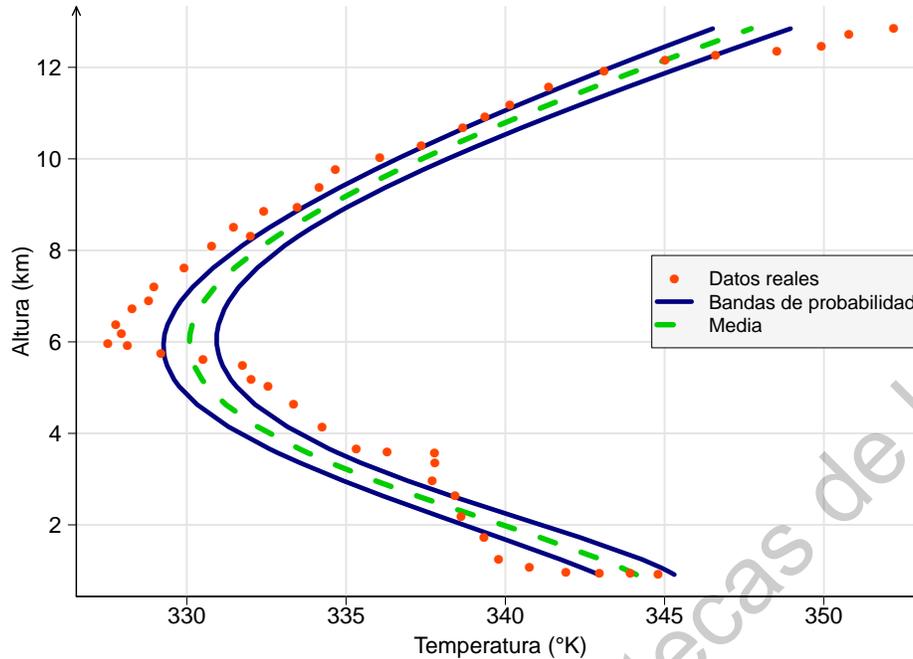


Figura 4.4: Bandas de probabilidad del 95 % del valor esperado (o media) de la temperatura potencial equivalente de la parcela, para el modelo estadístico 1. Los datos reales son representados por puntos, las bandas por líneas continuas azules (banda inferior y superior, izquierda y derecha, respectivamente) y la media por la línea punteada verde.

descrita en la Sección 3.2.4. La Figura 4.5 muestra bandas de probabilidad puntual del 95 % para el valor de la temperatura potencial en cada altura  $z$ . Puede observarse que estas curvas también describen la tendencia de los datos cubriéndolos en su totalidad. Esto quiere decir que, para cualquier altura, la predicción del modelo es razonable. Para dar una idea de la interpretación de estas bandas, considere la temperatura a una altura de 6 km. Se puede observar, de acuerdo a la distribución predictiva, que la temperatura está entre los 326 y 334 °K aproximadamente. Este intervalo tiene una amplitud de 8 °K. Si este rango es amplio o no y, más aún, si es útil o no, depende del uso que se le quiera dar al modelo. No obstante, desde el punto de vista estadístico, lo importante es que se puede cuantificar la incertidumbre en la predicción, es decir, que este intervalo tiene una probabilidad posterior del 95 %. Este es el aspecto que se desea resaltar y que no se presenta en [Flores, 2017].

## 4.2. Resultados del modelo estadístico 2

Para estudiar la distribución *a posteriori* del vector de parámetros  $\nu$  del modelo 2, se utilizó el algoritmo *t-walk* tomando como función objetivo la ecuación (3.18) multiplicada por -1. Se corrió la cadena de Markov por 1,000,000 de iteraciones utilizando distintos puntos iniciales aleatorios generados a partir de la distribución *a priori* de  $\nu$ . Se descartaron las

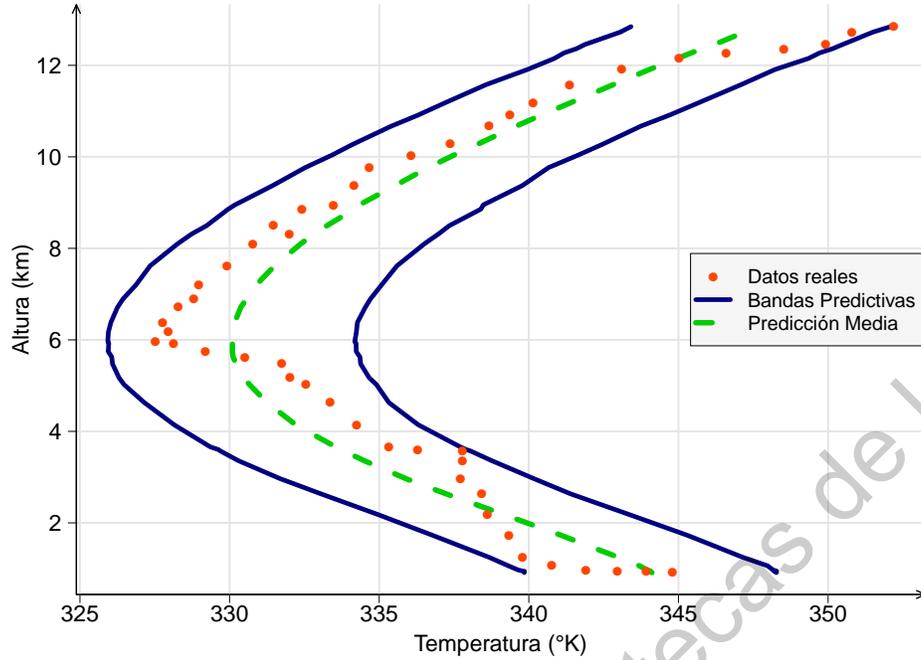


Figura 4.5: Bandas predictivas del 95 % de la temperatura potencial equivalente de la parcela, para el modelo estadístico 1. Los datos reales son representados por puntos, las bandas por líneas continuas azules (banda inferior y superior, izquierda y derecha, respectivamente) y la predicción media por la línea punteada verde.

primeras 10,000 iteraciones de la cadena como burn-in. De las 990,000 iteraciones restantes, se hace un submuestreo cada 139 muestras para obtener una cadena pseudo-independiente. El valor de 139 se obtiene del cálculo del IAT sobre la función de energía del modelo 2; siendo  $IAT = 138.31$ .

La Figura 4.6 muestra la evolución de la función de energía de la cadena resultante, después de omitir el burn-in y realizar el submuestreo. La función de energía muestra estabilizarse entre -130 y -119. Asimismo, se analizaron las trayectorias de las cadenas de los parámetros  $\varepsilon$ ,  $q_{0,env}$ ,  $B$ , y  $d_\theta$  mostrando que también convergen a partir de la iteración 10,000 (ver Figura 4.7).

Para visualizar las distribuciones *a posteriori* de los parámetros se realizaron histogramas utilizando 7123 muestras, las cuales pueden observarse en la Figura 4.8. Note que el parámetro razón de mezcla  $\varepsilon$  tiene mayor probabilidad de tomar valores en el intervalo (0.26, 0.42), ver Figura 4.8a. El valor referencial de vapor de agua del medio ambiente  $q_{0,env}$  al nivel superficie  $z = 0$  toma valores con mayor probabilidad en el intervalo (0.016, 0.022), ver Figura 4.3b. Recordemos que los parámetros  $\varepsilon$  y  $q_{0,env}$  están expresados adimensionalmente, para rescatar sus unidades ( $g/kg$ ) basta multiplicar sus valores por 1,000. El parámetro razón de enfriamiento  $B$ , con unidades  $^\circ K km^{-1}$ , toma valores con mayor probabilidad en el intervalo (4.45, 4.85), ver Figura 4.3c. Por último, se observa en la Figura 4.8d que el parámetro de

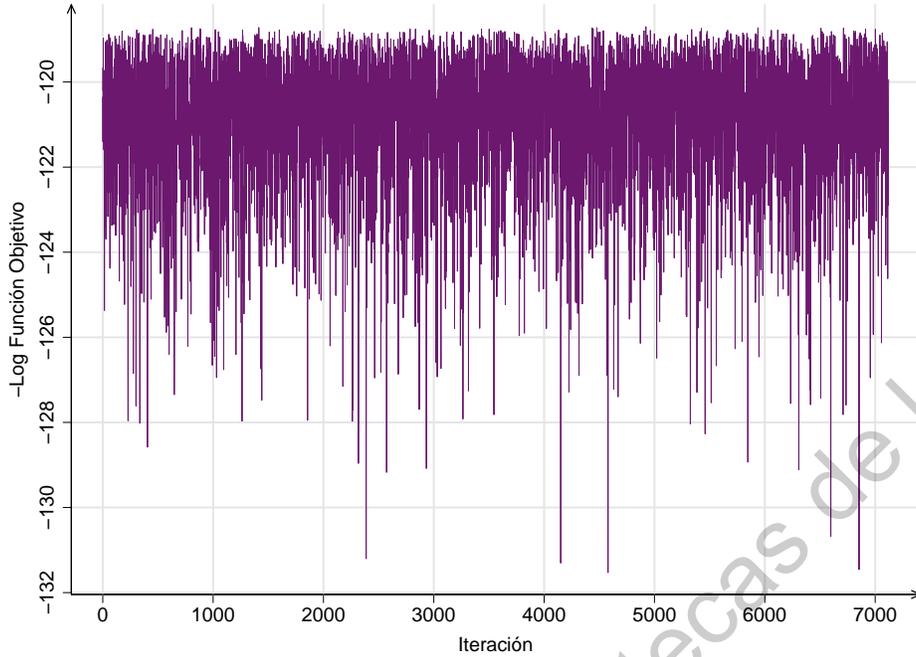


Figura 4.6: Función de energía del modelo estadístico 2 de la cadena resultante, después de haber omitido el burn-in y realizado el submuestreo.

dispersión  $d_\theta$  toma valores con mayor probabilidad entre (1.4, 2.4). La Figura 4.3 también muestra las distribuciones *a priori* para cada parámetro, mostrando que la distribución *a posteriori* (3.17) tiene menor variabilidad que la distribución *a priori* (3.16). Además, los histogramas tienen una moda claramente definida y con colas ligeras.

Nuevamente compararemos los valores puntuales de los parámetros dados por (4.1), pero ahora, con las estimaciones de los parámetros del modelo 2. Se observa que un valor de  $\varepsilon = 0.4$  y  $d_\theta = 2$  están dentro de la región de alta probabilidad de sus respectivas densidades *a posteriori*, ver Figura 4.8a y 4.8d. Por otro lado, un valor de  $q_{0,env} = 0.015$  y  $B = 4$  no están dentro de la región de alta probabilidad de sus respectivas densidades *a posteriori*, ver Figura 4.8b y 4.8c.

Ahora, vamos a analizar la variabilidad de las soluciones de la ecuación diferencial (3.5) generada por la variabilidad en los parámetros estimados por el modelo 2. Esto se hace de forma análoga a lo realizado en la Sección 4.1, es decir, se construyen bandas de probabilidad puntuales para cada altura  $z$ . La Figura 4.9 muestra los cuantiles puntuales de nivel 2.5%, 50% y 97.5% para la solución media de la ecuación (3.5). Se observa que, en general, la curva describe la tendencia de los datos, esto es, la curva decrece de 0 a 6 km e incrementa a partir de los 6 km tal como los datos observados.

Por último, a fin de crear predicciones para la temperatura utilizaremos la función de densidad *a posteriori predictiva* del modelo 2. La Figura 4.10 muestra bandas de probabilidad puntual del 95% para el valor de la temperatura potencial en cada altura  $z$ . Puede

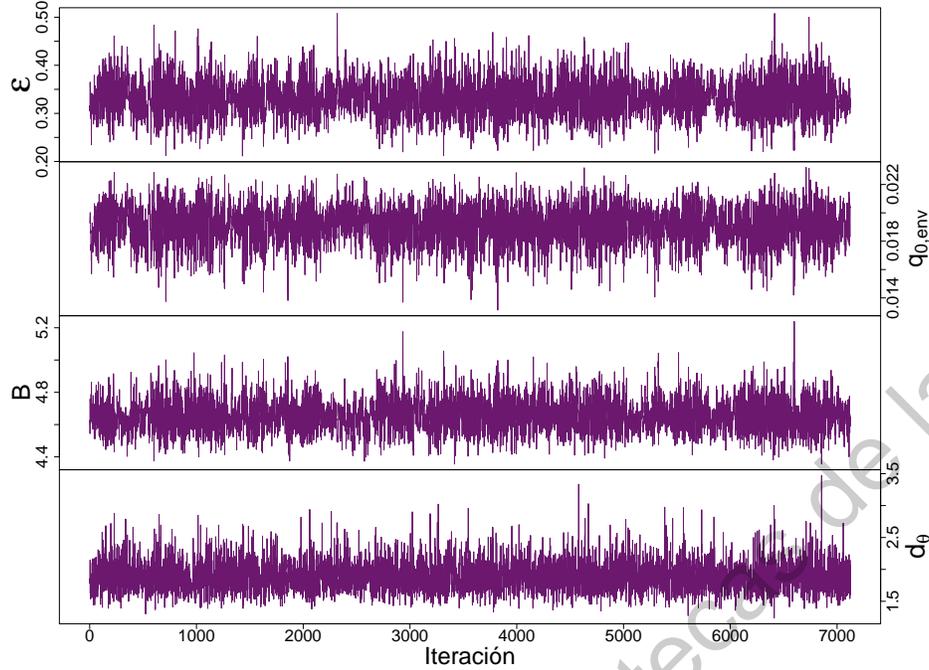


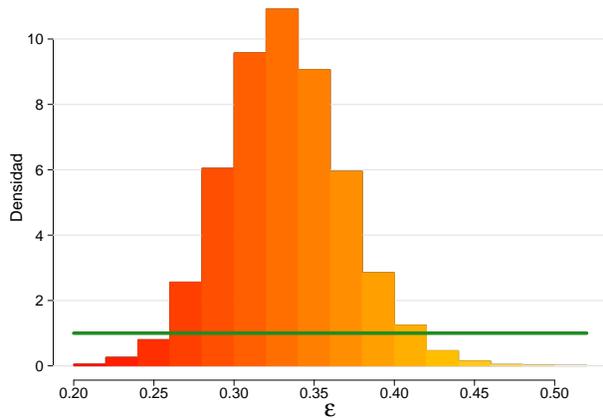
Figura 4.7: Trayectorias de las cadenas de los parámetros  $\varepsilon$ ,  $q_{0,env}$ ,  $B$  y  $d_\theta$  del modelo estadístico 2, después de haber omitido el burn-in y realizado el submuestreo.

observarse que estas curvas también describen la variabilidad de los datos, para cualquier altura, generando predicciones razonables. Para ejemplificar el uso de las bandas, considere dos temperaturas en diferentes alturas. De acuerdo a la distribución predictiva, para las alturas de 2 y 12  $km$ , sus respectivas temperaturas están entre (335, 344) y (340, 350)  $^\circ K$  aproximadamente. Además, note que la amplitud de los intervalos incrementa a mayores alturas (esto es de esperarse por la construcción del modelo). Le recordamos al lector que estos resultados permiten medir la incertidumbre del modelo teniendo una probabilidad posterior del 95 %.

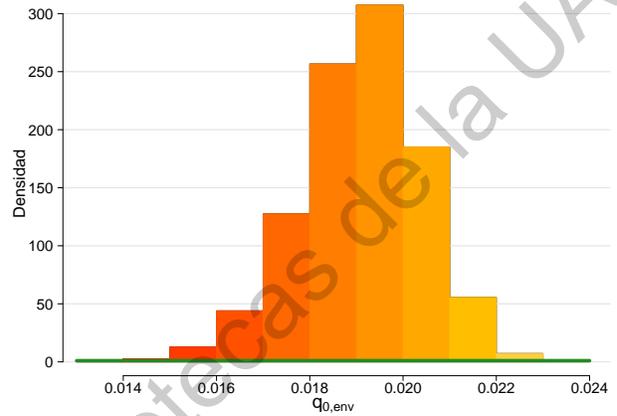
### 4.3. Comparación de los modelos estadísticos

Para comparar ambos modelos estadísticos, la Tabla 4.1 muestra las estimaciones de los parámetros de cada modelo usando la mediana posterior y un intervalo de probabilidad del 95 %. Puede verse que las estimaciones puntuales de los parámetros que participan en la ecuación diferencial,  $(\varepsilon, q_{0,env}, B)$ , son muy parecidas. Lo mismo ocurre para los intervalos de probabilidad posterior del 95 %.

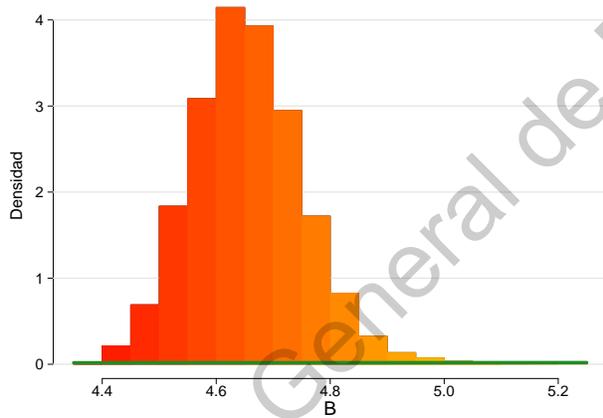
Para comparar la variabilidad en los modelos 1 y 2, la Figura 4.11 muestra la mediana de la curva esperada de temperatura y las bandas del 95 % de probabilidad posterior, que se obtienen con cada modelo. Como puede verse, ambos modelos generan estimaciones muy similares y sus respectivas bandas predictivas del 95 % contienen el total de los datos. La me-



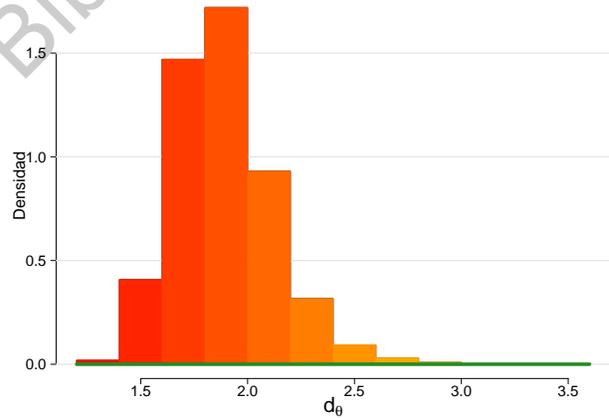
(a) Razón de mezcla.



(b) Cantidad de vapor de agua en el medio ambiente.



(c) Razón de enfriamiento.



(d) Parámetro de dispersión.

Figura 4.8: Histogramas con 7123 muestras de las distribuciones *a posteriori* de los parámetros  $\varepsilon$ ,  $q_{0,env}$ ,  $B$  y  $d_\theta$  del modelo estadístico 2.

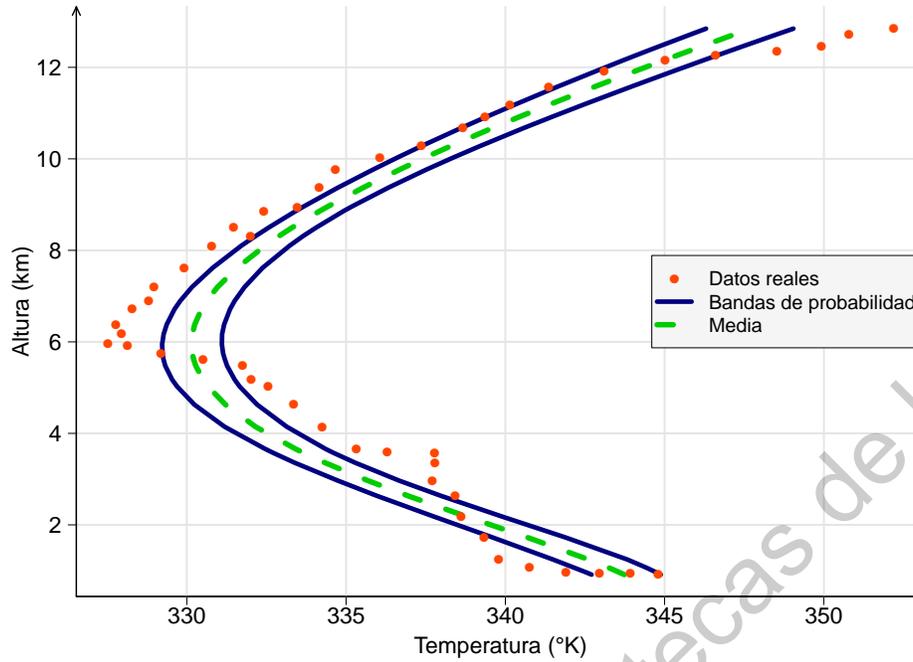


Figura 4.9: Bandas de probabilidad del 95 % del valor esperado (o media) de la temperatura potencial equivalente de la parcela, para el modelo estadístico 2. Los datos reales son representados por puntos, las bandas por líneas continuas azules (banda inferior y superior, izquierda y derecha, respectivamente) y la media por la línea punteada verde.

	Modelo 1		Modelo 2	
	Mediana	Intervalo 95 %	Mediana	Intervalo 95 %
$\varepsilon$	0.3486	(0.2801, 0.4183)	$\varepsilon$	0.3307 (0.2624, 0.4083)
$q_{0,env}$	0.0200	(0.0171, 0.0222)	$q_{0,env}$	0.0192 (0.0163, 0.0215)
$B$	4.6077	(4.4532, 4.8019)	$B$	4.6500 (4.4779, 4.8601)
$\tau$	0.2420	(0.1576, 0.3493)	$d_{\theta}$	1.8626 (1.5053, 2.4133)

Tabla 4.1: Comparación de las estimaciones de los parámetros de los modelos 1 y 2.

diana en ambos casi indistinguible. Las bandas del modelo 1 son ligeramente menos amplias que las del modelo 2, exceptuando los primeros 3 km. Por su parte, las bandas del modelo 2 comienzan siendo estrechas en la parte inferior del gráfico y éstas se abren conforme la altura crece.

Para hacer una evaluación más rigurosa del mejor modelo, se utiliza una ligera modificación del criterio AIC (*Akaike Information Criterion*), el cual es uno de los criterios de uso

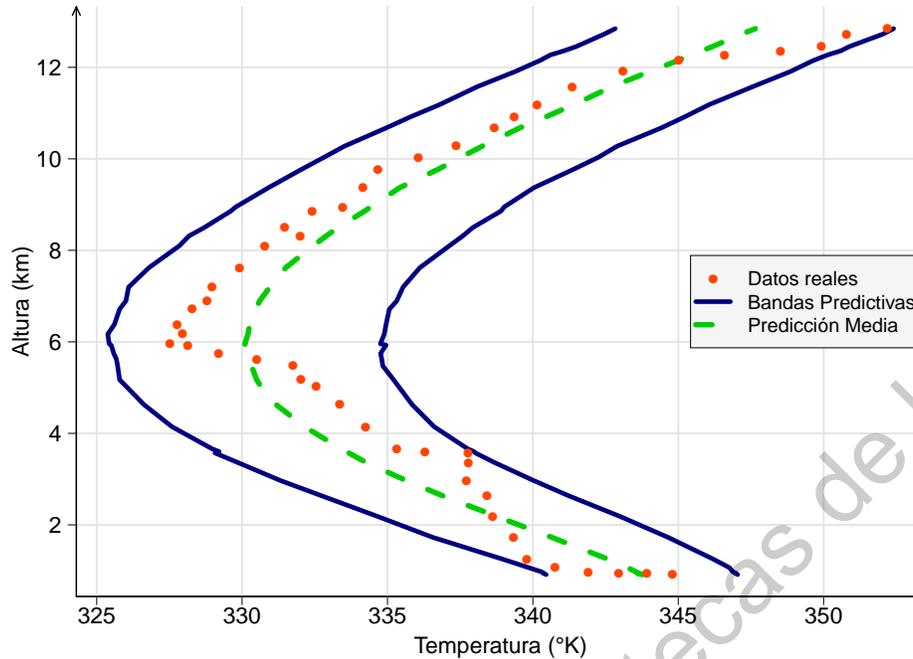


Figura 4.10: Bandas de predictivas del 95 % del valor esperado (o media) de la temperatura potencial equivalente de la parcela, para el modelo estadístico 2. Los datos reales son representados por puntos, las bandas por líneas continuas azules (banda inferior y superior, izquierda y derecha, respectivamente) y la media por la línea punteada verde.

frecuente para la selección de modelos. El AIC se calcula utilizando la fórmula

$$-2 \log \left( p \left( y | \hat{\theta} \right) \right) + 2k, \quad (4.2)$$

donde  $p(y|\hat{\theta})$  es la densidad *a posteriori predictiva* dada la media posterior  $\hat{\theta}$ , y  $k$  es el número de parámetros estimados dentro del modelo. Los detalles técnicos del AIC se pueden revisar en [Gelman et al., 2013]. Se obtuvo  $AIC_1 = 216.265$  y  $AIC_2 = 216.795$  para el modelo 1 y 2, respectivamente. Siendo  $AIC_1 < AIC_2$  indica que el modelo 1 es el mejor; aunque esta evidencia no es lo suficientemente clara.

Para terminar este capítulo, la Figura 4.12 muestra el ajuste de los datos que se presenta en [Flores, 2017] calibrando los parámetros y el ajuste que se obtiene con los modelos de inferencia presentados en este trabajo. Consideramos que la aproximación estadística provee más información ya que nos permite conocer el rango de los parámetros que pueden explicar los datos y, en general, provee un mejor ajuste.

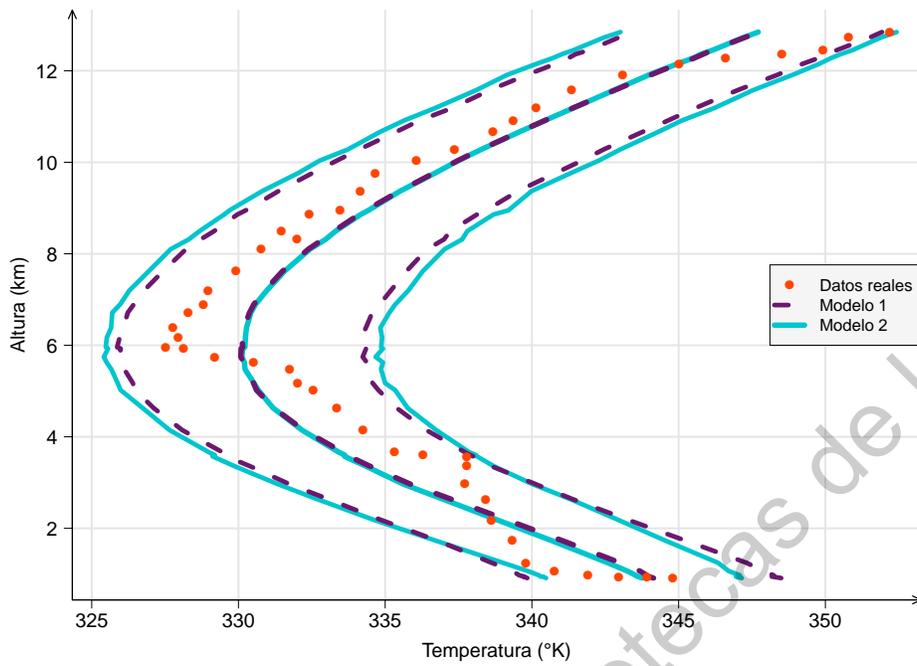
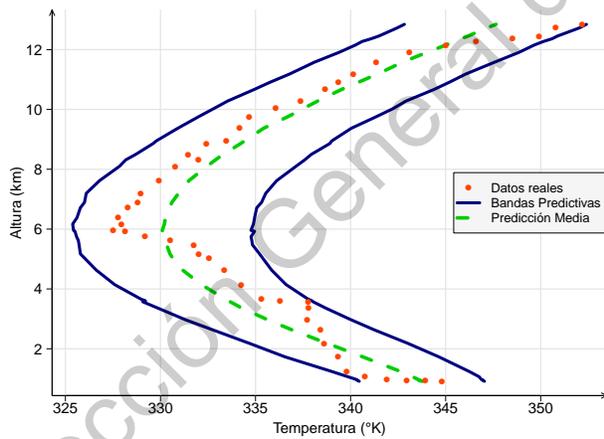
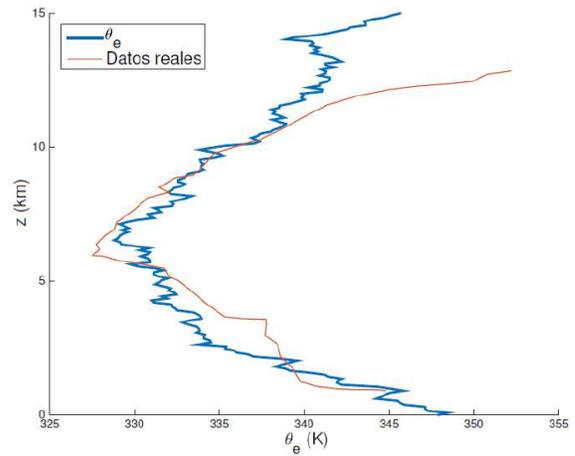


Figura 4.11: Comparación de las curvas predictivas de temperatura obtenidas con los modelos 1 y 2.



(a) Curvas predictivas de temperatura en nubes obtenidas con el modelo estadístico 2.



(b) Ajuste del modelo estocástico presentado en Flores [2017].

Figura 4.12: Comparación entre el ajuste del modelo estadístico 2 y el ajuste presentado en Flores [2017].

---

## Conclusiones

Siguiendo el objetivo de este trabajo, se implementó la metodología Bayesiana para estimar el vector de parámetros desconocidos,  $\boldsymbol{\vartheta} = (\varepsilon, q_{0,env}, B)$  del modelo atmosférico (3.1) presentado en [Flores, 2017]. Para ello, se propusieron dos modelos estadísticos que permitieron obtener estimadores puntuales e intervalos de probabilidad para cada parámetro. A partir de estos resultados, se pudo cuantificar la variabilidad de la solución del modelo diferencial a partir de las estimaciones hechas, creando bandas de probabilidad puntual del 95 % para el valor esperado de la temperatura potencial equivalente de la parcela en cada altura  $z$ . Más aún, se pudo evaluar la incertidumbre en la predicción de los modelos estadísticos al crear bandas de probabilidad puntual predictivas del 95 % para la dinámica observada.

Las curvas obtenidas por ambos modelos son muy similares. La mediana posterior de la temperatura de la parcela predicha por el modelo describe bien la tendencia observada, aunque existen patrones en los datos que no parecen ser provocados sólo por ruido. Por otra parte, las curvas predictivas mostraron cubrir la totalidad de los datos indicando que se pueden obtener, por parte de ambos modelos, predicciones razonables de la temperatura dada cualquier altura de la tropósfera.

El criterio estadístico AIC para comparar modelos indicó que el modelo 1 es mejor, aunque consideramos que la evidencia a favor del modelo 1 es mínima. A pesar de esto, es interesante que el modelo 2 no ofrezca una clara ventaja sobre el modelo 1, a pesar de que el primero considera una varianza en el término de error que cambia con la altura. Esto sugiere que modelos simples con supuestos restrictivos como el de varianza constante pueden dar buenos resultados.

Este trabajo sirve de ejemplo para resaltar las ventajas de utilizar procedimientos estadísticos en cualquier proceso de calibración de modelos, particularmente en problemas inversos. La metodología implementada en la tesis provee información de gran relevancia que no se presenta en el trabajo original [Flores, 2017], no sólo por el hecho de poder cuantificar la incertidumbre de los parámetros y el modelo, sino porque la aproximación estadística permite: 1) evaluar la identificabilidad de los parámetros, 2) medir desviaciones significativas del modelo a los datos.

A pesar de que la metodología presentada se implementó de forma exitosa produciendo buenos resultados, aún quedan cosas por estudiar. En particular, el supuesto de independencia condicional es cuestionable ya que no es fácil medir el tipo de correlación que se induce en las observaciones. Es posible que la falta de ajuste del modelo se deba a la simplificación que se hace sobre la estructura de correlación de los datos. Una posible solución es asumir que los datos se comportan como un proceso Gaussiano con una cierta correlación temporal (exponencial o Matern) y estimar algunos de los parámetros asociados a dicha estructura de correlación. Dejamos esto como investigación futura.

Dirección General de Bibliotecas de la UAQ

## Conjunto de observaciones de temperaturas en las nubes

	Temperatura ( $^{\circ}K$ )	Altura ( $km$ )
1	344.801240	0.91238671
2	343.922987	0.94259819
3	342.958075	0.94259819
4	341.904383	0.97280967
5	340.760852	1.06344411
6	339.789580	1.24471299
7	339.334022	1.72809668
8	338.616367	2.18126888
9	338.425028	2.63444109
10	337.711613	2.96676737
11	337.785552	3.35951662
12	337.778131	3.57099698
13	336.285843	3.60120846
14	335.318811	3.66163142
15	334.249218	4.14501511
16	333.355064	4.62839879
17	332.551810	5.02114804
18	332.020194	5.17220544
19	331.746436	5.47432024
20	330.513065	5.62537764
21	329.193036	5.74622357
22	328.134044	5.92749245

	<b>Temperatura (<math>^{\circ}K</math>)</b>	<b>Altura (<math>km</math>)</b>
23	327.518948	5.95770393
24	327.950125	6.16918429
25	327.767266	6.38066465
26	328.281921	6.71299094
27	328.801876	6.89425982
28	328.966714	7.19637462
29	329.916786	7.61933535
30	330.777018	8.10271903
31	331.997668	8.31419940
32	331.464992	8.49546828
33	332.417183	8.85800604
34	333.466635	8.94864048
35	334.153549	9.37160121
36	334.666084	9.76435045
37	336.060052	10.0362538
38	337.367361	10.2779456
39	338.669370	10.6706949
40	339.362644	10.9123867
41	340.142577	11.1842900
42	341.356867	11.5770393
43	343.099592	11.9093656
44	345.020936	12.1510574
45	346.595643	12.2719033
46	348.522288	12.3625378
47	349.922616	12.4531722
48	350.790269	12.7250755
49	352.189537	12.8459215

## Códigos de R

En este apéndice se encuentra el código de R el cual está basado en la metodología descrita en el Capítulo 2 y 3. El código se divide en 4 subcódigos para facilitar la lectura de estos. El código *runtwalk.R* realiza los pasos del algoritmo *t-walk* para obtener resultados (distribución *a posteriori*, estimador puntual, gráficos, etc.). *functions.R* contiene la función objetivo y el soporte de los modelos estadísticos, la función para obtener el estimador e intervalos de probabilidad, y la función para obtener el AIC. *twalkgraphics.R* contiene algunas funciones para graficar las cadenas de Markov obtenidas por el *t-walk* y para visualizar la distribución *a posteriori* marginal de los parámetros de interés. Por último *probbands.R* contiene funciones para graficar las bandas de probabilidad.

### *runtwalk.R*

```

1 #####
2 # ANTES DE CORRER EL CODIGO INDICA EL DIRECTORIO DONDE SE ENCUENTRAN LOS
3 # SUBCODIGOS Y BASE DE DATOS. ELIGE EL MODELO 1 0 2 PARA REALIZAR LA
4 # ESTIMACION DE PARAMETROS; VER LINEAS 9 - 20.
5 #####
6 rm(list = ls())                ### Limpia memoria
7 library(Rtwalk)                ### Libreria T-WALK
8 library(LaplacesDemon)        ### Libreria para calcular el IAT
9 Model = 1                      ### Modelo estadistico 1 0 2
10 tr    = 2*500000              ### No. iteraciones
11 # Directorio donde se encuentran los subcodigos y el conj de observaciones
12 directorio = "C:/Users/Usuario/OneDrive/Tesis_Final/Codigos_R_Tesis/"
13 setwd( directorio )

```

```

14 source("probbands.R")          ### Bandas de Prob. y Predictivas del 95 %
15 source("functions.R")         ### Codigos de los Modelos Estadisticos
16 source("twalkgraphics.R")     ### Graficos del t-walk
17 #####
18 #----- CONJUNTO DE ONSERVACIONES -----
19 #####
20 data = read.csv("data.csv", header = F)
21 y     = data[,1]               # TEMPERATURA EN NUBES
22 z     = data[,2]               # ALTURA DE LAS NUBES
23 #####
24 #----- PUNTOS INICIALES ALEATORIOS Y FUNCION OBJETIVO -----
25 #####
26 seed = 7                       # Es necesario fijar una semilla para reproducir los datos
27 if( Model == 1 ) {
28   par_name      = c("E", "q0", "B", "tau")
29   par_expr      = expression(epsilon, q[0][",,"][env], B, tau)
30   Objective_fun = Neg_LogPosterior_M1
31   LogLik        = Loglikelihood_M1
32   set.seed(seed+1); X0 = Random_Start_Points_of_theta(model = 1)
33   set.seed(seed-1); XP = Random_Start_Points_of_theta(model = 1)
34 } else {
35   par_name      = c("E", "q0", "B", "d0")
36   par_expr      = expression(epsilon, q[0][",,"][env], B, d[theta])
37   Objective_fun = Neg_LogPosterior_M2
38   LogLik        = Loglikelihood_M2
39   set.seed(seed+1); X0 = Random_Start_Points_of_theta(model = 2)
40   set.seed(seed-1); XP = Random_Start_Points_of_theta(model = 2)
41 }
42 names(X0) = par_name
43 names(XP) = par_name
44 #####
45 #----- A PRIORIS -----
46 #####
47 f_U = function(x) { dunif( x, min = 0 , max = 1 ) } # U(0,1)
48 f_G_B = function(x) { dgamma(x, shape = 2 , scale = 14 ) } # Gamma(2, 14)
49 if(Model == 1) {
50   f_G_x = function(x) {dgamma(x, shape = 8/5, scale = 250)}#Gamma(8/5,250)
51 } else {
52   f_G_x = function(x) {dgamma(x, shape = 2 , scale = 250)}# Gamma(2, 250)
53 }

```

```

54 #####
55 #----- EJECUTAR T-Walk -----
56 #####
57 cat("\n----- Corrida del T-walk -----\n\n")
58 set.seed(seed)
59 info = Runtwalk(dim = 4, Tr = tr, Supp = Support, Obj = Objective_fun,
60               x0 = X0, xp0 = XP, PlotLogPost = FALSE)
61 #----- REMOVIENDO BURN-IN -----
62 from = 10000
63 to   = NROW(info$output)
64 PlotLogObjective(info$Us, from = from)
65 colnames(info$output) = par_name
66 #----- RESULTADOS (GRAFICOS, ESTIMACIONES E INTERVALOS) -----
67 cat("\n\n----- Resultados -----")
68 #----- IAT, ESTIMACIONES E INTERVALOS DE PROB 95% -----
69 iat = IAT(info$Us[(1+from):to]) # Sin burn-in
70 estimadores = Estimators(info$output, from = from, by = iat)
71 cat("\n\n ") ; print(estimadores)
72 cat("\n\n El IAT de la cadena de Markov es", iat)
73 cat("\n\n El AIC del modelo", Model, " es ",
74     A.I.C(LogLik, info$output[ seq(1+from, to, ceiling(iat)), ] ) )
75 #----- TRAYECTORIAS DE LAS CADENAS -----
76 PlotLogObjective(info$Us , from =from, by =iat, cex.axis=1.4, cex.lab=1.4)
77 TSPlot(info$output, from=from, by=iat, yax.flip=T, cex.axis=1.4, ann=F)
78 for(i in 1:4) mtext(par_expr[i], cex = 1.6, side = if(i%%2) 2 else 4,
79                 adj = (.102)*(8-2*(i-1)), line = if(i%%2) 2.4 else 0.6)
80 #----- HISTOGRAMAS -----
81 Histograma(info$output, par = 1, from = from, by = iat, xlab =par_expr[1],
82            add_aprior_fun = f_U, cex.axis = 1.4, cex.lab = 1.4)
83 Histograma(info$output, par = 2, from = from, by = iat, xlab =par_expr[2],
84            add_aprior_fun = f_U, cex.axis = 1.4, cex.lab = 1.4)
85 Histograma(info$output, par = 3, from = from, by = iat, xlab =par_expr[3],
86            add_aprior_fun = f_G_B, cex.axis = 1.4, cex.lab = 1.4)
87 Histograma(info$output, par = 4, from = from, by = iat, xlab =par_expr[4],
88            add_aprior_fun = f_G_x, cex.axis = 1.4, cex.lab = 1.4)
89 #----- BANDAS DE CONFIANZA Y PREDICTIVAS -----
90 plot_bands(y, z, info$output, from = from, by = iat, cex = 1.4, lwd = 6,
91           pt.cex = 1.4, type_bands = "CB")
92 set.seed(seed-4)
93 plot_bands(y, z, info$output, from = from, by = iat, cex = 1.4, lwd = 6,

```

```

94     pt.cex = 1.4,  model = Model, type_bands = "PB")
95 # -----

```

## *functions.R*

```

1 #####
2 #----- MODELO ATMOSFERICO -----
3 #####
4 D.model = function(z, theta) {
5   L = 2.5*10^6
6   Cp = 1004
7   z0 = 2
8   E = theta[1]
9   q0 = theta[2]
10  B = theta[3]
11  #----- TEMPERATURA POTENCIAL LINEAL DEL MEDIO AMBIENTE -----
12  temp.env = function(){ return( 300 + B*z ) }
13  #----- CANTIDAD DE VAPOR DE AGUA EN EL MEDIO AMBIENTE (Humedad) -----
14  q.v.env = function(){ return( q0*exp(-z/z0) ) }
15  #----- TEMPERTURA POTENCIAL EQUIVALENTE DEL MEDIO AMBIENTE -----
16  temp.e.env = function(){ return( temp.env() + (L/Cp)*q.v.env() ) }
17  #-----
18  return( temp.e.env() - B/E + (L/Cp)*q.v.env()/(E*z0-1) +
19          exp(-E*z)*( 345 - 300 + B/E - q0*(L/Cp)*E*z0/(E*z0-1) ) )
20 }
21 #####
22 #----- SOPORTE DEL MODELO ESTADISTICO 1 Y 2 -----
23 #####
24 Support = function( theta ) {
25   ( theta[1]>0 & theta[1]<=1 & theta[2]>0 & theta[2]<=1 & theta[3]>0 &
26     theta[4]>0 ) }
27 #----- MODELO ESTADISTICO 1 -----
28 #####
29 #----- LOGARITMO DE LA FUNCION DE MAXIMA VEROSIMILITUD -----
30 Loglikelihood_M1 = function( theta ) {
31   tau = theta[4]
32   logf = dnorm(y, mean = D.model(z,theta), sd = sqrt(1/tau), log = T)
33   return( sum(logf) )

```

```

34 }
35 #----- -LogPOSTERIOR -----
36 Neg_LogPosterior_M1 = function( theta ) {
37   fE   = dunif( theta[1], min   = 0 , max   = 1)
38   fq0  = dunif( theta[2], min   = 0 , max   = 1)
39   fB   = dgamma(theta[3], shape = 2 , scale = 14)
40   ft   = dgamma(theta[4], shape = 8/5, scale = 250)
41   logL = Loglikelihood_M1(theta)
42   return( -1*( logL + log(fE) + log(fq0) + log(fB) + log(ft) ) )
43 }
44 #####
45 #----- MODELO ESTADISTICO 2 -----
46 #####
47 #----- LOGARITMO DE LA FUNCION DE MAXIMA VEROSIMILITUD -----
48 Loglikelihood_M2 = function( theta ) {
49   sigma2 = (theta[4]**2)*(1 - exp(-2*z*theta[1]))/(2*theta[1])
50   logf   = dnorm(y, mean = D.model(z,theta), sd = sqrt(sigma2), log = T)
51   return( sum(logf) )
52 }
53 #----- -LogPOSTERIOR -----
54 Neg_LogPosterior_M2 = function( theta ) {
55   fE   = dunif( theta[1], min   = 0 , max   = 1 )
56   fq0  = dunif( theta[2], min   = 0 , max   = 1 )
57   fB   = dgamma(theta[3], shape = 2 , scale = 14 )
58   fd0  = dgamma(theta[4], shape = 2 , scale = 250)
59   logL = Loglikelihood_M2(theta)
60   return( -1*( logL + log(fE) + log(fq0) + log(fB) + log(fd0) ) )
61 }
62 #####
63 #----- AKAIKE INFORMATION CRITERIA (AIC) -----
64 #####
65 A.I.C = function( Loglikelihood, data ) {
66   posterior_mean = apply(data, MARGIN = 2, FUN = mean)
67   k = NROW( posterior_mean )
68   ( -2*Loglikelihood(posterior_mean) + 2*k )
69 }
70 #####
71 #----- ESTIMADORES E INTERVALOS DE PROB. DEL 95% -----
72 #####
73 Estimators = function( data, from = 0, to = NROW(data), by = NULL,

```

```

74         q = c(0.025,0.975) ) {
75     i     = seq(1+from, to, by = if(is.null(by)) 1 else ceiling(by))
76     data = as.matrix(data)
77     data = data[i,]
78     info = matrix(0, nrow = NCOL(data), ncol = 2+length(q))
79     for(j in 1:NCOL(data))
80         info[j,] = as.numeric( c(median(data[,j]), mean(data[,j]),
81                                 quantile(data[,j], probs = q) ) )
82     colnames(info) <- c("Mediana", "Media", paste0("Cuantil_",q) )
83     rownames(info) <- colnames(data)
84     return( round(info, digits = 4) )
85 }
86 #####
87 ## PUNTOS INICIALES ALEATORIOS DE LOS PARAMETROS DESC. DEL MODELO 1 Y 2 ##
88 #####
89 Random_Start_Points_of_theta = function( model = 1 ) {
90     p     = numeric(0)
91     p[1] = runif( 1, min     = 0 , max     = 1)
92     p[2] = runif( 1, min     = 0 , max     = 1)
93     p[3] = rgamma(1, shape = 2 , scale = 14)
94     if( Model==1 )
95         p[4] = rgamma(1, shape = 8/5, scale = 250)
96     else
97         p[4] = rgamma(1, shape = 2, scale = 250)
98     return( p )
99 }
100 #####
101 #-- FUNCION PARA TOMAR UNA SUBMUESTRA DEL TOTAL DE LA CADENA DE MARKOV --#
102 #####
103 subsampling = function( data, from = 0, to = NROW(data), by = NULL,
104                         select_col = 1 ) {
105     select_row = seq( 1+from, to, by = if(is.null(by)) 1 else ceiling(by) )
106     data = as.matrix(data)    ### Evita problemas de dim (en caso de vector)
107     data = data[select_row, select_col]
108     if( !is.null(by) ) {
109         cat("\n      -Del total de la cadena, se tomara una submuestra tomando
110             espacios de ", ceiling(by),
111             "\n      -La submuestra es de tamaño", NROW(data)) }
112     invisible( data )
112 }

```

## *twalkgraphics.R*

```

1 #####
2 #----- GRAFICO DEL -LOG DE LA FUNCION OBJETIVO -----
3 #####
4 PlotLogObjective = function(data, from = 0, to = NROW(data), by = NULL,
5                             ylab = "-Log Funcion Objetivo",
6                             xlab = "Iteracion", col = "#6B186EFF",
7                             col.axis = par("col.axis"), bty = "l",
8                             las = 0, tcl = -0.55, axes = T, ...) {
9   cat("\n\n PlotLogObjective:")
10  oldpar <- par(mar = c(3.9, 4.0, 0.5, 0.55) + 0.5)
11  on.exit( par(oldpar) )
12  y = -subsampling(data, from, to, by )
13  if(is.null(by)) x = from:(to-1) else x = 0:(NROW(y)-1)
14  plot(x, y, type = "l", col = col, ylab = ylab, xlab = xlab,
15        col.axis = col.axis, bty = bty, las = las, tcl = tcl, axes = axes,
16        panel.first=if(axes) grid(lty=1, lwd=2.5, col="gray90") else NULL,
17        ...)
18  u <- par("usr")
19  arrows(u[1], u[3], u[2], u[3], code=2, xpd=T, length=0.125, angle=25,
20         col = if(axes) col.axis else "white", lwd = par("cex.axis"))
21  arrows(u[1], u[3], u[1], u[4], code=2, xpd=T, length=0.125, angle=25,
22         col = if(axes) col.axis else "white", lwd = par("cex.axis"))
23 }
24 #####
25 #----- TRAYECTORIAS DE LAS CADEMAS DE MARKOV (SERIES DE TIEMPO) -----
26 #####
27 TSPlot = function (data, pars = 1:NCOL(data), from = 0, to = NROW(data),
28                   by = NULL, xlab = "Iteracion", main = "",
29                   col = "#6B186EFF", axes = T, yax.flip = T, ...) {
30   cat("\n\n TSLOT: ")
31   x = subsampling(data, from, to, by, pars)
32   x = as.ts(x)
33   if( length(pars) <= 10 ) {
34     plot(x, xlab = xlab, main = main, col = col, axes = axes, yax.flip =
yax.flip, ...)

```

```

35 }else
36   cat("No se puede graficar series de tiempo para mas de 10 parametros,",
37       "\nselecciona un subconjunto con el argumento pars\n\n")
38 }
39 #####
40 #----- HISTOGRAMAS DE LAS MUESTRAS DE LOS PARAMETROS -----#
41 #####
42 Histograma = function(data, par = 1, from = 0, to = NROW(data), by = NULL,
43                       freq=F, col.H=NULL, border=NULL, breaks = "Sturges",
44                       las = 1, tcl = -0.55, main = "", xlab = "",
45                       ylab = if( freq )"Frecuencia" else "Densidad",
46                       densityPlot = F, lwd_density = 6,
47                       col_density = "#6B186EFF", add_aprior_fun = NULL,
48                       lwd_aprior_fun = 7, col_aprior_fun = "forestgreen",
49                       axes = T, ...) {
50   cat("\n\n Hist", colnames(data)[par], ": ")
51   oldpar <- par(mar = c(3.9, 4.1, 0.3, 0.3) + 0.6)
52   on.exit(par(oldpar))
53   data = subsampling(data, from, to, by, par)
54   data_h = hist(data, freq = freq, xlab = xlab, ylab = ylab, main = main,
55                breaks = breaks, las = las, tcl = tcl, ...)
56   if( is.null(col.H) ) col.H = colores( length(data_h$mids) )
57   if( is.null(border) ) border = col.H
58   if(axes) grid(NA, NULL, lty = 1, lwd = 2.5, col = "gray90")
59   hist(data, col = col.H, freq = freq, breaks = breaks, add = T, border =
60         border)
61   if(densityPlot && !freq) lines(density(data), lty = 2, lwd = lwd_density
62     , col = col_density)
63   if( !is.null(add_aprior_fun) && !freq ) {
64     curve(add_aprior_fun, lwd=lwd_aprior_fun, col=col_aprior_fun, add=T) }
65   invisible(data_h)
66 }
67 #####
68 #----- COLORES DEGRADADOS PARA LOS HISTOGRAMAS -----#
69 #####
70 colores = function( n, alpha = 1, rev = F, start= 1/100, end = 1/8) {
71   if ((n <- as.integer(n[1L])) > 0) {
72     j <- n%/%4
73     i <- n - j
74     cols <- c(rainbow(i, start = start, end = end, alpha = alpha),

```

```

73         if(j > 0) {
74             hsv(h = end, s = seq.int(from = 1-1/(2*j),
75                                     to = 1/(2*j), length.out=j),
76                                     v = 1, alpha = alpha) } )
77     if (rev) { cols <- rev(cols) }
78     cols
79 } else character()
80 }
81 # -----

```

## *probbands.R*

```

1 #####
2 #----- BANDAS DE PROBABILIDAD DEL 95 %-----
3 #####
4 ConfidenceBands = function(z, data, from = 0, to = NROW(data), by = NULL,
5                             q = c(0.025,0.50,0.975) ) {
6     data = subsampling(data, from, to, by, select_col = 1:NCOL(data) )
7     N = NROW(data)
8     m = length(z)
9     ### Evaluando el modelo atmosferico en cada altura z_i y para theta_i
10    ### (vector parametros) de la muestra obtenida con el algoritmo t-walk
11    solutions = matrix(0, nrow = N, ncol = m)
12    for(i in 1:N) solutions[i,] = D.model(z, data[i,])
13    ### Quantiles
14    sample_quantiles = matrix(0, nrow = m, ncol = length(q))
15    for (i in 1:m) sample_quantiles[i,] = quantile(solutions[,i], probs = q)
16    colnames(sample_quantiles) <- paste0("cuantil_", q*100, "%")
17    invisible( sample_quantiles )
18 }
19 #####
20 #----- BANDAS PREDICTIVAS DEL 95 %-----
21 #####
22 PredictBands = function(z, data, from = 0, to = NROW(data), by = NULL,
23                          model = 1, type_bands = c("CB","PB"),
24                          q = c(0.025,0.50,0.975) ) {
25     cat("\n      -Modelo", model)
26     data = subsampling(data, from, to, by, select_col = 1:NCOL(data) )
27     N = NROW(data)

```

```

28 m = length(z)
29 y.new = matrix(0, nrow = N, ncol = m)
30 if( model == 1 ) s2 = function(pars) { return(1/pars[4]) }
31 if( model == 2 ) s2 = function(pars) {
32     return( (pars[4]**2)*( 1 - exp(-2*z*pars[1]) )/(2*pars[1]) ) }
33 ### Simulaciones Y ~ Normal(m,s^2)
34 for ( i in 1:N ) y.new[i,] = rnorm(m, mean = D.model(z, data[i,]),
35     sd = sqrt(s2(data[i,])))
36 ### Quantiles
37 sample_quantiles = matrix(0, nrow = m, ncol = length(q))
38 for (i in 1:m) sample_quantiles[i,] = quantile(y.new[,i], probs = q)
39 colnames(sample_quantiles) <- paste0("cuantil_", q*100, "%")
40 invisible( sample_quantiles )
41 }
42 #####
43 #----- GRAFICO DE BANDAS DEL 95 % -----
44 #####
45 plot_bands = function(y, z, data, from = 0, to = NROW(data), by = NULL,
46     model = 1, type_bands = c("CB","PB"),
47     q = c(0.025,0.50,0.975), xlab = "Temperatura (K)",
48     ylab = "Altura (km)", font = par("font"), pt.cex = 1,
49     cex = par("cex"), lwd = par("lwd"), lty = c(0,1,2),
50     bty = "l", tcl = -0.55, las = 1, col.axis = "black",
51     colours = c("orangered", "navy", "green3","navy"),
52     legendd = if( type_bands[1]=="CB" ) {
53         c("Datos reales","Bandas de probabilidad","Media")
54     } else c("Datos reales","Bandas Predictivas",
55         "Prediccion Media"), axes = TRUE ) {
56 if( type_bands[1]=="CB" ) {
57     cat("\n\n Confidence Bands:")
58     quantiles = ConfidenceBands(z, data, from, to, by, q)
59 } else {
60     cat("\n\n Predictive Bands:")
61     quantiles = PredictBands(z, data, from,to,by, model, type_bands, q) }
62 oldpar <- par(mar = c(3.9, 4.0, 0.5, 0.55) + 0.5)
63 on.exit(par(oldpar))
64 xlim = c( range(y, quantiles[,1])[1], range(y, quantiles[,3])[2] )
65 plot(y, z, xlim = xlim, ylab = ylab, xlab = xlab, bty = bty, tcl = tcl,
66     las = las, col.axis = col.axis, font.axis = font, font.lab = font,
67     cex.axis = cex, cex.lab = cex,

```

```

68     panel.first =if(axes) grid(lty=1, lwd=2.5, col="gray90") else NULL)
69 lines(quantiles[,1], z, lty = lty[2], lwd = lwd, col = colours[2])
70 lines(quantiles[,2], z, lty = lty[3], lwd =lwd + 1, col = colours[3])
71 lines(quantiles[,3], z, lty = lty[2], lwd = lwd, col = colours[4])
72 points(          y, z, pch =16          , cex = pt.cex, col = colours[1])
73 u <- par("usr")
74 arrows(u[1], u[3], u[2], u[3], code = 2, xpd = T, length = 0.125,
75        angle = 25, col = col.axis, lwd = cex)
76 arrows(u[1], u[3], u[1], u[4], code = 2, xpd = T, length = 0.125,
77        angle = 25, col = col.axis, lwd = cex)
78 legend("right", legend = legendd, col = colours, lty = lty,
79        lwd = c(0, lwd, lwd+1), pch = c(16, NA, NA), pt.cex = pt.cex,
80        bty = "o", box.lty = 1, box.lwd = 1, bg = "gray96", merge = T,
81        cex = cex-0.25, text.font = font, inset = .001 )
82 invisible( quantiles )
83 }
84 #-----

```

Dirección General de Bibliotecas de la UAQ

---

# Bibliografía

- Manuel Adrian Acuña-Zegarra, Mario Santana-Cibrian, and Jorge X Velasco-Hernandez. Modeling behavioral change and COVID-19 containment in Mexico: A trade-off between lockdown and compliance. *Mathematical biosciences*, 325:108370, 2020. ISSN 1879-3134. URL <https://doi.org/10.1016/j.mbs.2020.108370>.
- J Andrés Christen, Colin Fox, et al. A general purpose sampling algorithm for continuous distributions (the t-walk). *Bayesian Analysis*, 5(2):263–281, 2010. URL <http://ba.stat.cmu.edu/journal/2010/vol05/issue02/christen.pdf>.
- Verónica Flores. Teoría de la parcela de fluidos para el estudio de inestabilidades atmosféricas y distribución de alturas de las nubes. Tesis de licenciatura, Universidad Autónoma de Querétaro, Facultad de Ingeniería, 2017.
- Klaus Fraedrich. On the distribution of cloud top heights based on stochastic forcing. *Tellus*, 37 A, pages 176–184, 1985.
- Andrew Gelman, John B Carlin, Hal S Stern, David B Dunson, Aki Vehtari, and Donald B Rubin. *Bayesian data analysis*. CRC press, 2013.
- Trevor Hastie, Robert Tibshirani, Gareth James, and Daniela Witten. An introduction to statistical learning with applications in r, 2013.
- W. K. Hastings. Monte carlo sampling methods using markov chains and their applications. *Biometrika*, 57(1):97–109, 1970. URL <http://www.jstor.org/stable/2334940>.
- Harold Jeffreys. An invariant form for the prior probability in estimation problems. *Proceedings of the Royal Society of London. Series A. Mathematical and Physical Sciences*, 186 (1007):453–461, 1946.
- Jari Kaipio and Erkki Somersalo. *Statistical and computational inverse problems*, volume 160. Springer Science & Business Media, 2006.
- Samuel Karlin and Howard M. Taylor. *First Course in Stochastic Processes*. Academic Press, 1975.

- J.S. Liu. *Monte Carlo Strategies in Scientific Computing*. Springer Series in Statistics. Springer, 2004.
- R Core Team. *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria, 2020. URL <https://www.R-project.org/>.
- Sidney I. Resnick. *Adventures in Stochastic Processes*. Springer Science & Business Media, 1992.
- Luis Rincón. *Introducción a los procesos estocásticos*. Universidad Nacional Autónoma de México, Facultad de Ciencias, 2012.
- Luis Rincón. *Introducción a la probabilidad*. Universidad Nacional Autónoma de México, Facultad de Ciencias, 2014.
- Christian Robert and George Casella. *Monte Carlo statistical methods*. Springer Science & Business Media, second edition edition, 2004.
- Gareth O Roberts, Jeffrey S Rosenthal, et al. Optimal scaling for various metropolis-hastings algorithms. *Statistical science*, 16(4):351–367, 2001.
- Mario Santana-Cibrian, Manuel A. Acuña-Zegarra, and Jorge X. Velasco-Hernandez. Lifting mobility restrictions and the effect of superspreading events on the short-term dynamics of COVID-19. *Mathematical Biosciences and Engineering*, 17(5):6240–6258, 2020. ISSN 1551-0018. URL <http://dx.doi.org/10.3934/mbe.2020330>.
- Brandon M. Turner and Trisha Van Zandt. A tutorial on approximate bayesian computation. *Journal of Mathematical Psychology*, 56(2):69 – 85, 2012. ISSN 0022-2496. doi: <https://doi.org/10.1016/j.jmp.2012.02.005>. URL <http://www.sciencedirect.com/science/article/pii/S0022249612000272>.
- John M Wallace and Peter V Hobbs. *Atmospheric science: an introductory survey*, volume 92. Elsevier, 2006.
- Larry Wasserman. *All of statistics: A concise course in statistical inference*. Springer Science & Business Media, 2013.