



UNIVERSIDAD NACIONAL AUTÓNOMA DE MÉXICO
PROGRAMA DE MAESTRÍA Y DOCTORADO EN CIENCIAS MATEMÁTICAS Y DE LA
ESPECIALIZACIÓN EN ESTADÍSTICA APLICADA

ASIMILACIÓN SECUENCIAL BAYESIANA DE DATOS CON SISTEMAS DINÁMICOS CAÓTICOS. UN
ENFOQUE EXPERIMENTAL.

TESINA
QUE PARA OPTAR POR EL GRADO DE:
MAESTRO (A) EN CIENCIAS

PRESENTA:
MAT. DANIEL ORNELAS DURÁN

DIRECTOR
DR. MARCOS AURELIO CAPISTRÁN OCAMPO
CENTRO DE INVESTIGACIONES EN MATEMÁTICAS, A.C.

CIUDAD DE MÉXICO, FEBRERO 2023

AGRADECIMIENTOS

Le agradezco a mi familia, novia y amigos el apoyo brindado todo este tiempo, sobre todo a María Guadalupe, Eduardo, Melchor, Stephania y Axel.

Quiero agradecer a la UNAM por darme la oportunidad de seguir mis estudios de posgrado, a mis profesores y compañeros del mismo por las excelentes clases, lecciones y pláticas, en especial a Emmanuel, Ricardo, Melissa y a los Doctores Jorge Martínez y Ramón Plaza.

También quiero agradecer al Dr. Marcos Capistrán por aceptarme como su alumno a pesar de no conocerme, por su infinita paciencia y por su apoyo en la realización de este trabajo.

Adicionalmente, quiero agradecer al CIMAT por dejarme utilizar su equipo de súper cómputo y al CONACYT por otorgarme una beca para llevar a cabo mis estudios de maestría.

ÍNDICE GENERAL

1. Introducción	1
1.1. Antecedentes	1
1.2. Trabajo Relacionado	2
1.3. Contribuciones y limitaciones.	2
2. Marco Teórico	4
2.1. Sistemas dinámicos caóticos.	4
2.2. Sistema de Lorenz.	5
2.3. El enfoque bayesiano	6
2.4. Muestreo : Parallel-Tempering, PT.	7
2.5. Problema y propuesta de un método de BSDA.	7
2.6. Casos de estudio.	10
2.7. Modelación de las distribuciones	10
3. Resultados Experimentales	16
3.1. Detalles de la implementación y resultados	16
3.2. Figuras y tablas de los experimentos I, II y III.	20
4. Conclusiones	27

1.1. Antecedentes

La asimilación de datos o **DA**, es la combinación de datos experimentales, modelos matemáticos y métodos computacionales para hacer y mejorar predicciones, selección de modelos y cálculo de cantidades de interés [1, 2]. El proceso de generación de conocimiento a través de la asimilación de datos consiste en varios pasos. Primero, se observa un fenómeno y se obtienen datos de éste a través de la experimentación. De aquí, se generan hipótesis, teorías y modelos matemáticos que intentan describir al fenómeno. A partir de este punto se suelen plantear modelos por medio de ecuaciones que surgen de las hipótesis propuestas y, por otro lado, de la estadística se obtiene toda la información que los datos nos puedan otorgar sobre el fenómeno de estudio. Aquí es donde interviene la asimilación de datos, combinando ambas fuentes de información del mismo fenómeno. Usualmente los cálculos que conllevan combinar estas fuentes de información son costosos en tiempo y energía, es por eso que se implementan en la computadora. El proceso de la Figura 1.1 se puede repetir una y otra vez, siempre y cuando se sigan teniendo a la mano datos nuevos. Hay métodos de **DA** que se pueden aplicar de forma subsecuente combinando la información de datos nuevos con la ya obtenida anteriormente por el mismo método. A la colección de estos métodos se le conoce como **asimilación secuencial de datos** o **SDA**. Si la actualización de la información se lleva a través del Teorema de Bayes, se dice que el método es de **asimilación secuencial bayesiana de datos** o de **BSDA**. Sabemos que hacer predicciones de variables de estado de un fenómeno que está gobernado por un sistema de ecuaciones diferenciales caótico es muy complicado, ya que sus soluciones tienen sensibilidad a las condiciones iniciales. Aquí es donde toma importancia la **SDA**. La idea consiste en corregir sucesivamente las predicciones hechas por un sistema dinámico caótico, utilizando la información dada por nuevos datos sobre los estados predichos. Las aplicaciones de **SDA** con sistemas dinámicos encuentran su lugar, principalmente, en el estudio de la atmósfera terrestre y en la oceanografía, debido a la dinámica turbulenta de los fluidos involucrados en algunos fenómenos que se presentan en la atmósfera y océano [12, 17].

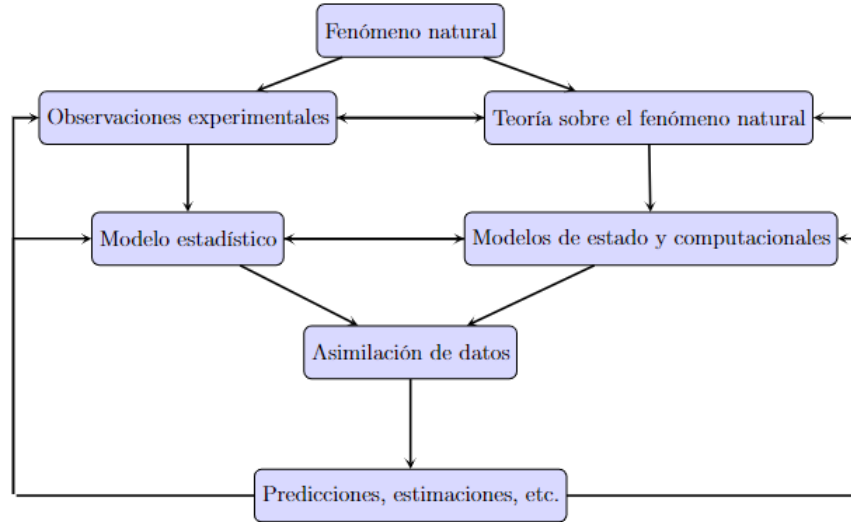


Figura 1.1: Esquema de la participación de la asimilación de datos en el proceso de predicción, estimación de parámetros o elección de modelos, así como su subsecuente repetición.

1.2. Trabajo Relacionado

El objetivo de este trabajo es presentar predicciones utilizando un método de **BSDA** presentado en [3] con el sistema de Lorenz. El método de **BSDA** presentado en [3], desarrollado por Daza-Torres, Capistrán, Capella y Christen, es el medio por el cual se hacen las predicciones y cálculos de algunas cantidades de interés referentes al desarrollo de la pandemia de COVID-19 en México, utilizando un sistema dinámico no caótico.

Con respecto a la predicción con el sistema de Lorenz, Zhang presenta en [24, 25] métodos para predecir la velocidad del viento, combinando datos reales de campos eólicos, técnicas de machine learning y series de tiempo construidas a partir de soluciones del sistema de Lorenz. En [26], Ying ZHANG presenta una aplicación de la técnica bred vectors al sistema de Lorenz.

Con respecto a dinámicas caóticas y el sistema de Lorenz, en [23] Zeng y Yang presentan un profundo estudio sobre la dinámica del sistema estocástico de Lorenz y su impacto en el entendimiento del efecto del ruido estocástico en un modelo determinista. En [22], a partir del sistema de Lorenz, Wang construye un sistema de ecuaciones diferenciales ordinarias en \mathbb{R}^4 que presenta más de un exponente de Lyapunov positivo, que sugiere un aumento en la sensibilidad a las condiciones iniciales.

1.3. Contribuciones y limitaciones.

Dentro de las contribuciones de este trabajo podemos encontrar que se aplica un método de **BSDA** para hacer predicciones utilizando el sistema de Lorenz. Es la primera vez que se experimenta con este método para estudiar el sistema de Lorenz. En consecuencia, otra contribución es justamente la presentación de diversos experimentos computacionales y sus resultados aplicando el método de **BSDA** presentado en [3]. Por otro lado, mientras la mayoría

de los estudios computacionales de predicciones que utilizan sistemas dinámicos caóticos se basan en métodos de machine learning y frecuentemente no se presentan medidas de la incertidumbre en sus predicciones, en este trabajo se da, de manera natural, una medida de la incertidumbre de las predicciones realizadas. Por último, dado que se trata de un experimento computacional nuevo, abre la pauta para futuros estudios y modificaciones en el método que mejoren las predicciones y la medida de la incertidumbre de éstas.

Dentro de las limitaciones, podemos encontrar dos principales: el poder de cómputo y el tiempo de horizonte u horizonte de predictibilidad. El método fue aplicado en el equipo de súper cómputo del CIMAT y cada paso del experimento llevó en promedio un tiempo aproximado de un día, lo cual es un indicio de que se requiere contar con un significativo poder de cómputo en caso de que se quiera aplicar este método. Por otro lado, el horizonte de predictibilidad es básicamente un tiempo límite que nos dice que tanto tardan dos soluciones de un sistema dinámico caótico en alejarse. A pesar de comenzar con dos soluciones con condiciones iniciales cercanas, inevitablemente, éstas se separarán en un cierto tiempo, usualmente corto. Debido al tiempo de horizonte, es muy complicado hacer predicciones en tiempos largos utilizando sistemas dinámicos caóticos, lo cual se ve reflejado en los resultados experimentales.

2.1. Sistemas dinámicos caóticos.

A continuación ofrecemos un listado de definiciones necesarias para recordar el significado de caos en un sistema de ecuaciones diferenciales ordinarias. Sea

$$\dot{v} = f(v) \quad (2.1)$$

un sistema de ecuaciones diferenciales ordinarias, donde $f \in C^{(1)}(\mathbb{R}^n)$. Sea

$$\dot{v} = f(v), \quad v(0) = v_0. \quad (2.2)$$

su problema de valores iniciales, con $v_0 \in \mathbb{R}^n$.

Definición 1. El **flujo** de (2.1) es la función $\varphi : \mathbb{R} \times \mathbb{R}^n \rightarrow \mathbb{R}^n$ tal que para $v_0 \in \mathbb{R}^n$, $t \rightarrow \varphi(t, v_0)$ es la solución de (2.2).

Definición 2. Sean φ el flujo de (2.1) y $A \subset \mathbb{R}^n$. Decimos que A es un **conjunto invariante bajo** φ si y sólo si para toda $t \geq 0$, $\varphi(t, A) \subset A$.

Definición 3. Decimos que $\omega_1, \dots, \omega_r \in \mathbb{R}$ son **racionalmente independientes** si y sólo si para todo $k_1, \dots, k_r \in \mathbb{Q} \setminus \{0\}$ se cumple que $\sum_{i=1}^r k_i \omega_i \neq 0$.

Definición 4. Sean $\omega_1, \dots, \omega_r \in \mathbb{R}$ racionalmente independientes. Una función $g : \mathbb{R} \rightarrow \mathbb{R}^n$ dada por $t \rightarrow g(t)$ es **cuasi-periódica con frecuencias** $(\omega_1, \dots, \omega_r)$ si y sólo si existe una función periódica $F = F(\theta_1, \dots, \theta_r)$ en $\theta_1, \theta_2, \dots$ y θ_r con periodo 1 tal que para todo $t \in \mathbb{R}$, $g(t) = F(\omega_1 t, \dots, \omega_r t)$.

Definición 5. Decimos que la solución $t \rightarrow \varphi(t, v_0)$ de (2.2) se **aproxima a** $P \subset \mathbb{R}^n$ si y sólo si

$$\lim_{t \rightarrow \infty} \inf_{p \in P} \|\varphi(t, v_0) - p\| = 0.$$

Definición 6. Decimos que la solución $t \rightarrow \varphi(t, v_0)$ de (2.2) tiene un **comportamiento aperiódico en tiempos largos** si y sólo si no se aproxima a un punto fijo, una órbita periódica u órbita cuasi-periódica de (2.1) cuando $t \rightarrow \infty$.

Definición 7. Sean φ el flujo de (2.1) y $A \subset \mathbb{R}^n$ invariante bajo φ . Decimos que φ exhibe **sensibilidad a las condiciones iniciales en A** si existe $r > 0$ tal que para cada $v \in A$ y para cada $\varepsilon > 0$ existen $y = y(v, \varepsilon) \in B(v, \varepsilon) \cap A$ y $t_h = t_h(v, \varepsilon) \geq 0$ tales que $|\varphi(t_h, v) - \varphi(t_h, y)| > 0$.

Definición 8. Sea φ el flujo de (2.1). Decimos que (2.1) es **caótico** si y sólo si existe $A \subset \mathbb{R}^n$ compacto e invariante bajo φ tal que φ exhibe sensibilidad a las condiciones iniciales en A y toda solución de (2.2) con condición inicial en A tiene un comportamiento aperiódico en tiempos largos.

2.2. Sistema de Lorenz.

En general, el sistema de Lorenz se define como

$$\begin{aligned}\dot{x} &= \sigma(y - x) \\ \dot{y} &= rx - xz - y \\ \dot{z} &= xy - bz,\end{aligned}\tag{2.3}$$

donde x, y y z son funciones dependientes del tiempo, \dot{x}, \dot{y} y \dot{z} son las derivadas de x, y y z con respecto al tiempo, σ es el número de Prandtl, r es el número de Raileigh y b es un número adimensional. Tanto σ, r y b son números reales positivos. Un resumen de las propiedades más importantes de (2.3) proveniente de [20, 18, 10] es el siguiente:

- (1) Tiene la simetría $(x, y, z) \rightarrow (-x, -y, z)$.
- (2) El eje z es invariante.
- (3) Su flujo contrae el volumen exponencialmente.
- (4) Si $0 < r < 1$, el origen es el único punto crítico y es, además, un atractor global.
- (5) Si $r > 1$, tiene tres puntos de equilibrio: el origen,

$$C_1 = (\sqrt{b(r-1)}, \sqrt{b(r-1)}, r-1) \quad \text{y}$$

$$C_2 = (-\sqrt{b(r-1)}, -\sqrt{b(r-1)}, r-1).$$

Ahora, sea

$$r_H = \frac{\sigma(\sigma + b + 3)}{\sigma - b - 1}$$

con $\sigma - b - 1 > 0$.

- (6) Si $1 < r < r_H$, tanto C_1 como C_2 son estables. El origen es inestable.
- (7) En $r = r_H$ hay una bifurcación subcrítica de Hopf.

- (8) En $r > r_H$, C_1 y C_2 pierden su estabilidad absorbiendo un ciclo límite inestable en la bifurcación subcrítica de Hopf.
- (9) Con los valores $\sigma = 10$ y $b = \frac{8}{3}$ y $r > r_H = 24.74$, la pérdida de estabilidad de C_1 y C_2 se traduce en la formación de un conjunto llamado el **atractor de Lorenz** que hace al sistema de Lorenz un sistema dinámico caótico.

2.2.1. Tiempo de horizonte en el atractor de Lorenz.

El predecir utilizando sistemas dinámicos caóticos es complicado. En la definición, no se menciona cuál es el tiempo finito en el cual se alejan las soluciones que inician con condiciones iniciales cercanas. Para el sistema (2.3), con $\sigma = 10$ y $b = \frac{8}{3}$ y $r = 28$, se tiene una aproximación de que tan rápido se alejan dos soluciones vecinas en el atractor de Lorenz. Supongamos que $(x(t), y(t), z(t))$ es un punto en el atractor de Lorenz en el tiempo t y consideremos un punto cercano en el atractor, $(x(t), y(t), z(t)) + \delta(t)$ donde δ es un vector con coordenadas no negativas para todo t y tal que $\delta(0) = \delta_0 > 0$. Se sabe que

$$\|\delta(t)\| \sim \|\delta_0\| \exp(\lambda t),$$

donde λ es aproximadamente 9×10^{-1} [20]. Ahora, consideremos que $(x(t), y(t), z(t)) + \delta(t)$ es la predicción de $(x(t), y(t), z(t))$ en un cierto intervalo de tiempo. Supongamos que $a > 0$ es la medida de nuestra tolerancia de la predicción. El tiempo que tarda δ en cumplir que $\|\delta(t)\| \geq a$ está dado por la aproximación,

$$t_h \sim O\left(\frac{1}{\lambda} \log\left(\frac{a}{\|\delta_0\|}\right)\right). \quad (2.4)$$

A t_h se le conoce como el **tiempo de horizonte del atractor de Lorenz** [19, 20].

2.3. El enfoque bayesiano

El problema de estimar parámetros es un problema clásico dentro de la estadística. Consiste en encontrar el valor de ciertos parámetros desconocidos θ a partir de un conjunto de datos w . Dentro de éste último a θ se le llama **cantidad desconocida** y la piensa como una variable aleatoria a la cual se le asigna una distribución de probabilidad que debe representar el conocimiento que se tiene sobre θ antes de tener los datos w . Este conocimiento previo sobre θ se actualiza por medio de la información dada por los datos w . Matemáticamente, el proceso de actualización de la información se da a través del Teorema de Bayes para distribuciones. Supongamos que $\tilde{\theta}$ y W son variables aleatorias con soportes Θ y \mathcal{W} , respectivamente. Denotemos por \mathcal{S} al soporte de la variable aleatoria $(\tilde{\theta}, W)$ y denotemos a la condicional de W dado $\tilde{\theta} = \theta$, con $\theta \in \Theta$, como $W|\tilde{\theta} = \theta$.

Teorema 1 (Teorema de Bayes, [11]). *Supongamos que $W|\tilde{\theta} = \theta$ tiene una densidad $f_{W|\tilde{\theta}}(w|\theta)$, que $\tilde{\theta}$ tiene densidad marginal $f_{\tilde{\theta}}(\theta)$ y que $f_W(w) > 0$. Entonces,*

$$f_{\tilde{\theta}|W}(\theta|w) = \frac{f_{W|\tilde{\theta}}(w|\theta)f_{\tilde{\theta}}(\theta)}{f_W(w)} = \frac{f_{W|\tilde{\theta}}(w|\theta)f_{\tilde{\theta}}(\theta)}{\int_{\Theta_w} f_{W|\tilde{\theta}}(w|t)f_{\tilde{\theta}}(t)dt}, \quad (2.5)$$

donde $\Theta_w := \{w \in \mathcal{W} | (\theta, w) \in \mathcal{S}\}$ y donde la integral del denominador de (2.5) será una suma o serie en caso de que $\tilde{\theta}$ sea discreta. ■

En la ecuación (2.5), a $f_{\tilde{\theta}}(\theta)$ se le conoce como la **distribución a priori** (o simplemente prior), a la distribución $f_{W|\tilde{\theta}}(w|\theta)$ se le conoce como la **verosimilitud** y a $f_{\tilde{\theta}|W}(\theta|w)$ se le llama la **distribución final** o la **distribución posterior** (o simplemente la posterior).

2.4. Muestreo : Parallel-Tempering, PT.

Parallel-Tempering o **PT** es una técnica de muestreo computacional capaz de encontrar rápidamente las regiones con mayor probabilidad en una distribución posterior objetivo. La idea del PT es utilizar varias cadenas de Markov para explorar todo el espacio de parámetros al mismo tiempo, donde cada cadena tiene una distribución objetivo a muestrear de la forma

$$\pi(\theta|d, \beta) = \pi(\theta)\pi(d|\theta)^\beta, \quad \text{con } 0 < \beta \leq 1.$$

Para cada valor de $0 < \beta \leq 1$, a $\pi(\theta|d, \beta)$ se le llama una **versión temperada de la distribución posterior**. Las idea de las versiones temperadas de la distribución posterior es regular la influencia de la verosimilitud, lo cual permite que para los valores de β más pequeños, las cadenas de Markov correspondientes puedan muestrear una distribución lo más parecida a la prior, mientras que para valores de β más cercanos a uno, las cadenas de Markov correspondientes puedan muestrear una distribución más parecida a la posterior. El muestreo de cada una de las distribuciones es llevado a cabo aplicando el algoritmo de **Metropolis-Hashtings** [15]. En particular, la cadena de interés es la que explora a la distribución $\pi(\theta|d, \beta)$ con $\beta = 1$, que es la distribución posterior. Dentro del PT se hace que las cadenas que exploran las distintas distribuciones intercambien información por medio de un intercambio de estados. En Algorithm 1 se muestra el PT aplicado en el presente trabajo y cuyo intercambio de estados se puede dar entre todas las cadenas de Markov del muestreo de manera aleatoria [9, 16]. En el pseudocódigo Algorithm 1, $N \in \mathbb{N}$ es el número de muestras a obtener, $K \in \mathbb{N}$ es el número de distribuciones temperadas, $m \in \mathbb{N}$ es la dimensión del espacio de parámetros, I la matriz identidad de $m \times m$, $N(\bar{0}, A)$ es la distribución normal con media $\bar{0}$ y matriz de covarianzas A , $U([0, 1])$ es la distribución uniforme con soporte en $[0, 1]$, $U(\{1, \dots, K\})$ es la distribución uniforme con soporte en $\{1, \dots, K\}$ y $\bar{\beta} = (\beta_1, \dots, \beta_K) \in (0, 1]^K$ tal que $1 = \beta_1 > \dots > \beta_K > 0$. También para cada $n \in \{1, \dots, N\}$ y para cada $k \in \{1, \dots, K\}$, $\theta_k^{(n)}$ es el n -ésimo estado de la k -ésima cadena de Markov. Además, para cada $k \in \{1, \dots, K\}$, $\theta_k^{(1)}$ es el estado inicial de la k -ésima cadena de Markov, $\theta_0 = (\theta_1^{(1)}, \dots, \theta_K^{(1)})$, $\bar{\rho} = (\rho_1, \dots, \rho_K) \in (0, \infty)^K$ y para cada $k \in \{1, \dots, K\}$, $\pi(\theta|d, \beta_k) = \pi(\theta)\pi(d|\theta)^{\beta_k}$.

2.5. Problema y propuesta de un método de BSDA.

Supongamos que conocemos un sistema de ecuaciones diferenciales que describe la ley por la cual se rige un fenómeno. Supongamos que cada cierto tiempo se nos entregan datos provenientes de mediciones y experimentos de las variables de estado que están presentes en el sistema de ecuaciones. El problema consiste en hacer predicciones de las variables de estado del sistema, y si es el caso, hacer estimación de las cantidades desconocidas que no se puedan medir directamente con la ayuda de las fuentes de información recién mencionadas: datos y conocimiento teórico del fenómeno. En este trabajo se utiliza el método de **BSDA** propuesto en [3] para resolver la problemática planteada. Presentamos un resumen de dicho método en el siguiente párrafo.

Algorithm 1: Parallel-Tempering con intercambios aleatorios de estados.

```

1 procedure  $PT(N, K, m, \bar{\beta}, \theta_0, \bar{\rho})$ ;
2 for  $n \leftarrow 1, N$  do
3   for  $k \leftarrow 1, K$  do
4     Draw  $\omega_k \sim N(0, \rho_k I)$ ;
5     Calculate  $\theta_k^{\text{new}} = \theta_k^{(n)} + \omega_k$ ;
6     Calculate  $\alpha = \min \left\{ 1, \frac{\pi(\theta_k^{\text{new}} | d, \beta_k)}{\pi(\theta_k^{(n)} | d, \beta_k)} \right\}$ ;
7     Draw  $u \sim U([0, 1])$ ;
8     if  $u < \alpha$  then
9       | Accept : Set  $\theta_k^{(n+1)} = \theta_k^{\text{new}}$ 
10    end
11    else
12      | Reject : Set  $\theta_k^{(n+1)} = \theta_k^{(n)}$ 
13    end
14  end
15  for  $k \leftarrow 1, K$  do
16    Draw  $p \leftarrow U(\{1, \dots, K\})$ ;
17    Draw  $q \leftarrow U(\{1, \dots, K\})$ ;
18    Calculate  $\alpha = \min \left\{ 1, \frac{\pi(\theta_p^{(n+1)} | d, \beta_q) \pi(\theta_q^{(n+1)} | d, \beta_p)}{\pi(\theta_p^{(n+1)} | d, \beta_p) \pi(\theta_q^{(n+1)} | d, \beta_q)} \right\}$ ;
19    Draw  $u \sim U([0, 1])$ ;
20    if  $u < \alpha$  then
21      | Accept : Swap states  $\theta_p^{(n+1)}$  and  $\theta_q^{(n+1)}$ 
22    end
23    else
24      | Reject : Do not swap states  $\theta_p^{(n+1)}$  and  $\theta_q^{(n+1)}$ 
25    end
26  end
27 end
28 output  $\theta = (\theta_1^{(1)}, \dots, \theta_1^{(N)})$ ;
29 end procedure

```

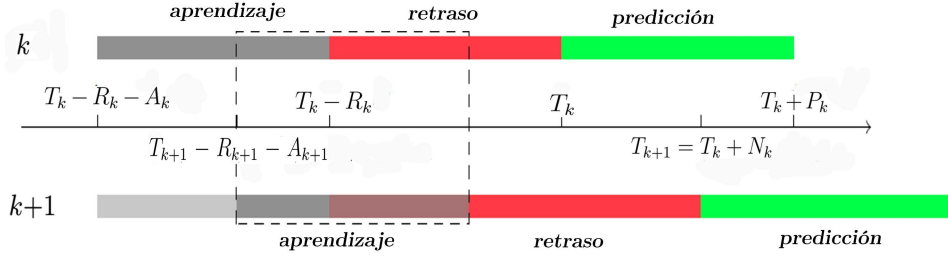


Figura 2.1: Los intervalos de aprendizaje, retraso y predicción de los pasos del algoritmo k y $k+1$, respectivamente. Ilustración tomada y modificada de [3].

Sean $k \in \mathbb{N}^0$, $A_k > 0$, $R_k \geq 0$, $P_k > 0$ y $T_k \geq 0$. Supongamos que se obtienen datos experimentales $w^{(k)}$ de algunas variables relacionadas con algún fenómeno tomados en el lapso de tiempo $[T_k - R_k - A_k, T_k - R_k]$. Al intervalo $[T_k - R_k - A_k, T_k - R_k]$ le llamamos el **k -ésimo intervalo de aprendizaje**. Consideramos que hay factores que pueden retrasar las predicciones y estimaciones, digamos que el lapso de tiempo de este retraso está dado por el intervalo $[T_k - R_k, T_k]$ y lo llamamos el **k -ésimo intervalo de retraso**. El tiempo al cual inicia la predicción es T_k y lo llamamos como el **k -ésimo tiempo de predicción**. Al intervalo $[T_k, T_k + P_k]$ lo llamamos como el **k -ésimo intervalo de predicción** y al intervalo $[T_k - R_k - A_k, T_k + P_k]$ lo llamamos la **k -ésima ventana de predicción** (ver la Figura 2.1). Sea $v(t)$ un vector de variables de estado del fenómeno de estudio dependientes del tiempo. Supongamos que conocemos un modelo de estados expresado como un problema de Cauchy con un sistema caótico de ecuaciones diferenciales ordinarias definido en $[T_k - R_k - A_k, T_k + P_k]$. Digamos que dicho problema está dado por

$$\dot{v}(t) = F(v(t), \theta_k) \quad (2.6)$$

$$v(T_k - R_k - A_k) = v_k$$

donde $T_k - A_k - R_k$, v_k y θ_k son el tiempo iniciales, el estado inicial y el vector de parámetros que calibran al sistema (2.6), respectivamente, en la ventana de predicción $[T_k - R_k - A_k, T_k + P_k]$. Supongamos que tanto v_k como θ_k cantidades desconocidas. Se define $g^{(k)} = (v_k, \theta_k)$. Aplicamos el enfoque Bayesiano y consideramos a las incógnitas como variables aleatorias. Proponemos una distribución $f_{G^{(k)}}(g^{(k)})$ para las incógnitas y una verosimilitud $f_{W^{(k)}|G^{(k)}}(w^{(k)}|g^{(k)})$ para encontrar una distribución posterior $f_{G^{(k)}|W^{(k)}}(g^{(k)}|w^{(k)})$ a través del Teorema de Bayes. Después, a través de alguna técnica de muestreo, obtenemos muestras de la distribución posterior. Con cada una de las muestras obtenidas de $f_{G^{(k)}|W^{(k)}}(g^{(k)}|w^{(k)})$ hacemos una predicción de $v(t)$ en la ventana de predicción $[T_k - R_k - A_k, T_k + P_k]$ utilizando (2.6). Ahora, supongamos que tenemos un nuevo vector de datos $w^{(k+1)}$, que $T_{k+1} = T_k + N_k$, con $N_k > 0$, es el siguiente tiempo de predicción y que $[T_{k+1} - R_{k+1} - A_{k+1}, T_{k+1} + P_{k+1}]$ es la nueva ventana de predicción, con $A_{k+1} > 0$, $R_{k+1} \geq 0$ y $P_{k+1} > 0$ y con la restricción de que $T_{k+1} - R_{k+1} - A_{k+1} \in [T_k - R_k - A_k, T_k - R_k]$. Usando los valores que se predicen para $v(t)$ en $t = T_{k+1} - R_{k+1} - A_{k+1}$ obtenidos de (2.6) y las muestras de $f_{G^{(k)}|W^{(k)}}(g^{(k)}|w^{(k)})$, proponemos una nueva distribución prior $f_{G^{(k+1)}}(g^{(k+1)})$ para $g^{(k+1)} = (v_{k+1}, \theta_{k+1})$, donde v_{k+1} es el estado inicial y θ_{k+1} es el vector de parámetros que calibran al sistema

$$\dot{v}(t) = F(v(t), \theta_{k+1}) \quad (2.7)$$

$$v(T_{k+1} - R_{k+1} - A_{k+1}) = v_{k+1}$$

definido en $[T_{k+1} - R_{k+1} - A_{k+1}, T_{k+1} + P_{k+1}]$. De ahí, simplemente hacemos $k \leftarrow k+1$ y repetimos el procedimiento para crear una nueva predicción y estimación de parámetros. Notemos que la idea primordial del procedimiento anterior es aplicar el conocimiento generado por el paso k en el algoritmo como conocimiento previo para el paso $k+1$ y actualizar la información por medio del Teorema de Bayes. Adicionalmente, suponemos que para cada $k \in \mathbb{N}^0$, $A_k = A > 0$, $R_k = R > 0$, $P_k = P > 0$ y $N_k = N > 0$ (ver Algorithm 2).

2.6. Casos de estudio.

En esta sección explicamos los casos de estudio provenientes del problema de **BSDA** presente en la sección 2.5 utilizando como modelo de estado al sistema de Lorenz. Para ello, escribamos explícitamente cómo se ve el sistema de Lorenz en cada ventana de predicción. Para cada $k \in \mathbb{N}^0$, el problema de condiciones iniciales (2.6) se convierte en

$$\begin{aligned} \dot{x} &= \sigma_k(y - x), & x(T_k - R - A) &= x_k, \\ \dot{y} &= x(r_k - z) - y, & y(T_k - R - A) &= y_k, \\ \dot{z} &= xy - b_k z, & z(T_k - R - A) &= z_k, \end{aligned} \tag{2.8}$$

donde x , y y z son las variables de estado del sistema de Lorenz, σ_k es el número de Prandtl, r_k el número de Rayleigh, b_k el parámetro geométrico y x_k , y_k y z_k son los estados iniciales de x , y y z , respectivamente, en el tiempo inicial $T_k - A - R$. El problema (2.8) está definido en la ventana de predicción k -ésima $[T_k - R - A, T_k + P]$.

2.6.1. Caso I.

Sea $K \in \mathbb{N}$. Este caso consiste en que para cada $k \in \{0, \dots, K\}$, debemos hacer predicciones de x , y y z en el k -ésimo intervalo de predicción $[T_k, T_k + P]$ a partir del problema (2.8) y el método de **BSDA** descrito en Algorithm 2, bajo el conocimiento de que $\sigma_k = 10$, $r_k = 28$ y $b_k = \frac{8}{3}$, y el desconocimiento del valor del vector de condiciones iniciales $g^{(k)} = (x_k, y_k, z_k)$.

2.6.2. Caso II.

Sea $K \in \mathbb{N}$. Este caso consiste en que para cada $k \in \{0, \dots, K\}$, debemos hacer predicciones de x , y y z en el k -ésimo intervalo de predicción $[T_k, T_k + P]$ y hacer estimaciones sobre el valor de r_k , a partir del problema (2.8) y el método de **BSDA** descrito en Algorithm 2, bajo el conocimiento de que $\sigma_k = 10$, $b_k = \frac{8}{3}$ y de que el valor de r_0 se encuentra en el intervalo $(0, 100)$, y el desconocimiento del valor del vector de cantidades desconocidas $g^{(k)} = (x_k, y_k, z_k, r_k)$.

2.7. Modelación de las distribuciones

En esta sección vamos a presentar la verosimilitud y las distribuciones prior necesarias para la aplicación del método de **BSDA** descrito en Algorithm 2 en cada uno de los casos de estudio.

Algorithm 2: Método de BSDA.

```

1 Input  $K \in \mathbb{N}$ ,  $f_{G^{(0)}}(g^{(0)})$ ,  $A > 0$ ,  $R > 0$ ,  $P > 0$ ,  $N > 0$ ,  $T_0 > 0$  y para cada  $k \in \{0, \dots, M\}$ , los datos  $w^{(k)}$ 
   tomados en el intervalo  $[T_k - R - A, T_k - R]$  y la propuesta de verosimilitud  $f_{W^{(k)}|G^{(k)}}(w^{(k)}|g^{(k)})$ .
2 Output Para cada  $k \in \{0, \dots, M\}$ , muestras de las distribuciones posteriores  $f_{G^{(k)}|W^{(k)}}(g^{(k)}|w^{(k)})$  y
   predicciones de las cantidades de interés en el intervalo  $[T_k, T_k + P]$ .
3 for  $k \leftarrow 0, K$  do
4   for  $m \leftarrow 1, 4$  do
5     if  $m = 1$  then
6       if  $k = 0$  then
7         Proponemos  $f_{G^{(0)}}(g^{(0)})$  como distribución a priori.
8       end
9       else
10        (I) Para las condiciones iniciales  $v(T_k - R - A) = v_k$  en la ventana de predicción
            $[T_k - R - A, T_k + P]$ , utilizamos las muestras de  $f_{G^{(k-1)}}(g^{(k-1)})$  para ajustar las
           predicciones  $v(T_k - R - A; g^{(k-1)})$  a una distribución conocida  $f_{V_k}(v_k)$  que usamos como
           distribución a priori para  $v_k$ .
11        (II) Para los parámetros del modelo de estados  $\theta_k$ , utilizamos los valores de  $\theta^{(k-1)}$ 
           provenientes de las muestras de  $f_{G^{(k-1)}|W^{(k-1)}}(g^{(k-1)}|w^{(k-1)})$  para ajustar una distribución
           conocida  $f_{\theta_k}(\theta_k)$  que usamos como distribución a priori para  $\theta_k$ .
12        (III) Proponemos  $f_{G^{(k)}}(g^{(k)}) = f_{\theta_k}(\theta_k)f_{V_k}(v_k)$  como prior para  $g^{(k)}$ .
13       end
14     end
15     if  $m = 2$  then
16       Obtenemos  $f_{G^{(k)}|W^{(k)}}(g^{(k)}|w^{(k)})$  a través de Teorema de Bayes y  $f_{W^{(k)}|G^{(k)}}(w^{(k)}|g^{(k)})$ ;
17       Obtenemos muestras de  $f_{G^{(k)}|W_k}(g^{(k)}|w^{(k)})$ .
18     end
19     if  $m = 3$  then
20       Obtenemos las soluciones  $v(t, g^{(k)})$  del sistema dinámico caótico para hacer las predicciones de
           las cantidades de interés hasta el tiempo  $T_k + P_k$ , donde  $g^{(k)}$  son los valores provenientes de las
           muestras de  $f_{G^{(k)}|W_k}(g^{(k)}|w^{(k)})$ .
21     end
22     if  $m = 4$  then
23       Guardamos las muestras de  $f_{G^{(k)}|W_k}(g^{(k)}|w^{(k)})$ .
24     end
25   end
26 end
27 end procedure

```

2.7.1. Verosimilitud

La verosimilitud es la distribución de los datos dada una realización de las cantidades desconocidas, así, para su construcción necesitamos una relación entre éstas. Para ello, suponemos que los datos de las variables de estado del sistema de Lorenz tienen ruido que se presenta de manera aditiva. Así, podemos pensar que tenemos la relación

$$w^i = \mathcal{H}^i(\mathcal{F}(g)) + \varepsilon^i, \quad (2.9)$$

donde $i \in \{0, \dots, n\}$ con $n \in \mathbb{N}$, $w^i = (x^i, y^i, z^i)$ son observaciones en el tiempo $t_i \in \mathbb{R}$, $\mathcal{F} : \mathcal{G} \rightarrow C(\mathbb{R}, \mathbb{R}^3)$ está dada por $\mathcal{F}(g) = S_g(t)$, con $S_g(t)$ como la solución del problema de condiciones iniciales para el sistema de Lorenz (2.3) la cual depende de la cantidad desconocida $g \in \mathcal{G}$ (en los casos considerados aquí \mathcal{G} puede ser $\mathbb{R}^2 \times [0, \infty)$ o $\mathbb{R}^2 \times [0, \infty) \times (0, 100)$) de condiciones iniciales y parámetros desconocidos, $\mathcal{H}^i : C(\mathbb{R}, \mathbb{R}^3) \rightarrow \mathbb{R}^3$ está dada por $\mathcal{H}^i(u) = u(t_i)$ y donde $\varepsilon^i \in \mathbb{R}^3$ representa el ruido de la observación i -ésima. Notemos que la función \mathcal{F} está bien definida gracias al teorema de existencia y unicidad para EDO's [13]. Sean W^i , G y \mathcal{E}^i las variables aleatorias cuyas realizaciones son w^i , g y ε^i . Por (2.9) tenemos la relación

$$W = \mathcal{H}(\mathcal{F}(G)) + \mathcal{E}, \quad (2.10)$$

donde

$$W = (W^1, \dots, W^n)^T, \quad \mathcal{H} = (\mathcal{H}^1, \dots, \mathcal{H}^n)^T \quad \text{y} \quad \mathcal{E} = (\mathcal{E}^1, \dots, \mathcal{E}^n)^T.$$

A la función $\mathcal{H} \circ \mathcal{F}$ se le conoce como el **mapeo de observación**. Suponiendo que la distribución de \mathcal{E} es $f_{\mathcal{E}}(e)$ y fijando g una realización de G , por (2.10), tenemos que la distribución de W dado que $G = g$ está dada por

$$f_{W|G}(w|g) = f_{\mathcal{E}}(w - \mathcal{H}(\mathcal{F}(g))), \quad (2.11)$$

donde $w = (w^1, \dots, w^n)^T$. Suponiendo que W^1, \dots, W^n son condicionalmente independientes con respecto al valor de G , tenemos que (2.11) se convierte en

$$f_{W|G}(w|g) = \prod_{i=1}^n f_{\mathcal{E}^i}(w^i - \mathcal{H}^i(\mathcal{F}(g))). \quad (2.12)$$

Ahora, basados en las ideas presentadas en [25], proponemos un modelo normal para el error. Es decir, suponemos que para cada $i \in \{0, \dots, n\}$, $\mathcal{E}^i \sim N_3(e|\bar{0}, \Sigma)$, con una matriz de covarianza positiva definida $\Sigma \in \mathbb{M}_{3 \times 3}(\mathbb{R})$. Así, (2.12) se convierte en

$$f_{W|G}(w|g) = \prod_{i=1}^n N_3(w^i - \mathcal{H}^i(\mathcal{F}(g))|\bar{0}, \Sigma). \quad (2.13)$$

2.7.2. Mapeo de observación.

El mapeo de observación que aparece en la expresión de la verosimilitud (2.13) es la evaluación en distintos tiempos de la solución del problema de Cauchy correspondiente del sistema de Lorenz. Como no se conoce la expresión matemática cerrada para expresar dicha solución, se opta por utilizar una versión numérica de ésta. Las soluciones numéricas que se utilizan son generadas por el método **LSODA** de la librería `scipy.solve_ivp` de la biblioteca `Scipy` de Python 3. LSODA consiste en aplicar un esquema propuesto por Linda Petzold que identifica automáticamente

si una cierta región de un problema de valores iniciales es más eficiente resolverlo con métodos diseñados para problemas **rígidos** o **no rígidos** y cambia los métodos según sea el caso. Aplica **BDF** de orden de precisión 1 al orden de precisión 5 para las partes rígidas y **Adams-Moulton** del orden de precisión 1 al orden de precisión 12 para las partes no rígidas [14, 7, 21]. Tanto Adams-Moulton como BDF fueron diseñados para resolver problemas de Cauchy con una ecuación de la forma $\dot{y} = f(y(t), t)$ y con condición inicial $y(0) = y_0$, donde $m \in \mathbf{N}$, $t \in \mathbf{R}$, $y : \mathbf{R} \rightarrow \mathbf{R}^m$, $y_0 \in \mathbf{R}^m$, y donde $f : \mathbf{R}^m \times \mathbf{R} \rightarrow \mathbf{R}^m$ es Lipschitz. Ambos métodos son **métodos numéricos lineales multipaso**. Un **método numérico lineal multipaso de m -pasos**, con $m \in \mathbf{N} \cup \{0\}$ y tamaño de paso constante h , es de la forma

$$a_0 y_{k+1} = a_1 y_k + \dots + a_m y_{k-m+1} + h(b_0 f_{k+1} + \dots + b_m f_{k-m+1}) \quad (2.14)$$

donde $a_0, a_1, \dots, a_m, b_0, \dots, b_m \in \mathbf{R}$, $|a_m| + |b_m| \neq 0$ y $f_{k+1}, \dots, f_{k-m+1}$ son aproximaciones numéricas de la función f en los tiempos de la malla $t_{k+1}, \dots, t_{k-m+1}$. La integración en este caso se hace en el intervalo $[t_{k-m+1}, t_{k+1}]$. En el caso de Adams-Moulton, el esquema numérico está dado por la ecuación (2.14) con $a_0 = a_1 = 1$, $a_2 = \dots = a_m = 0$, obteniendo una expresión de la forma

$$y_{k+1} = y_k + h(b_0 f_{k+1} + \dots + b_m f_{k-m+1}). \quad (2.15)$$

Ahora, las constantes restantes quedan determinadas por $b_j = \int_{t_k}^{t_{k+1}} \prod_{i=0, i \neq j}^m \frac{s - t_{k-i+1}}{t_{k-j+1} - t_{k-i+1}} ds$, para $j = 0, \dots, m$.

En LSODA, la solución de las ecuaciones generadas por (2.15) se hace aplicando iteración funcional [14, 7]. En el caso de BDF, primero notemos que (2.14) se puede expresar de la forma

$$c_m y_{k+1} + c_{m-1} y_k + \dots + c_0 y_{k-m+1} = h(d_m f_{k+1} + \dots + d_0 f_{k-m+1}). \quad (2.16)$$

Se definen el **primer y segundo polinomios característicos** de (2.16) como $p(z) = c_m z^m + \dots + c_1 z + c_0$ y $q(z) = d_m z^m + \dots + d_1 z + d_0$, respectivamente. Un método clásico para definir métodos multipaso es proponer los dos polinomios característicos y tomar los coeficientes de dichos polinomios característicos como los coeficientes del método multipaso. Para el método BDF, se impone la condición $c_m = 1$ y polinomios característicos $q(z) = dz^m$ y $p(z) = d \sum_{i=1}^m \frac{1}{m} z^{m-i} (z-1)^i$, donde $d = (\sum_{i=1}^m \frac{1}{i})^{-1}$ [8]. Las ecuaciones generadas por el esquema de BDF se resuelven en LSODA por medio de métodos de Newton modificados [14, 7].

2.7.3. La matriz Σ

Para especificar la matriz de covarianzas de la expresión (2.13), vamos a suponer que conocemos la relación señal a ruido SNR . La relación señal a ruido se define como

$$SNR = \frac{\text{magnitud de la señal}}{\text{desviación estándar de la señal}}$$

En todos los experimentos, tomamos $SNR = 100$ y obtenemos que las desviaciones estándar de las señales de las variables x , y y z del sistema de Lorenz son

$$\hat{\sigma}_x = \frac{\text{magnitud de la señal de } x}{100}, \quad \hat{\sigma}_y = \frac{\text{magnitud de la señal de } y}{100} \quad \text{y} \quad \hat{\sigma}_z = \frac{\text{magnitud de la señal de } z}{100}$$

respectivamente. Proponemos a la matriz de covarianzas definida positiva como

$$\Sigma = \text{diag}(\hat{\sigma}_x^2, \hat{\sigma}_y^2, \hat{\sigma}_z^2). \quad (2.17)$$

2.7.4. Prior para el caso I y $k = 0$

Aplicamos Naïve Bayes a las condiciones iniciales [6] obteniendo que

$$f_{G^{(0)}}(g^{(0)}) = f_{X_0}(x_0)f_{Y_0}(y_0)f_{Z_0}(z_0). \quad (2.18)$$

Para proponer $f_{X_0}(x_0)$, $f_{Y_0}(y_0)$ y $f_{Z_0}(z_0)$, resolvemos numéricamente el sistema (2.3), con $r = 28$, $b = 8/3$ y $\sigma = 10$, utilizando el método LSODA con condición inicial $(0, 1, 0)$, desde $t = 0$ hasta $t = 1000$. Rescatamos los valores obtenidos de las variables x , y y z , y calculamos su media y desviación estándar m_{x_0} , s_{x_0} , m_{y_0} , s_{y_0} , m_{z_0} y s_{z_0} , respectivamente. Así, las distribuciones a priori propuestas son

$$f_{X_0}(x_0) = N(x_0|m_{x_0}, s_{x_0}^2), \quad f_{Y_0}(y_0) = N(y_0|m_{y_0}, s_{y_0}^2) \quad \text{y} \quad f_{Z_0}(z_0) = N(z_0|m_{z_0}, s_{z_0}^2). \quad (2.19)$$

Sustituyendo las expresiones de (2.19) en (2.18), obtenemos que

$$f_{G^{(0)}}(g^{(0)}) = N(x_0|m_{x_0}, s_{x_0}^2)N(y_0|m_{y_0}, s_{y_0}^2)N(z_0|m_{z_0}, s_{z_0}^2). \quad (2.20)$$

2.7.5. Prior para el caso I y $k > 0$

Por Naïve Bayes obtenemos que

$$f_{G^{(k)}}(g^{(k)}) = f_{X_k}(x_k)f_{Y_k}(y_k)f_{Z_k}(z_k). \quad (2.21)$$

Ahora, sea \mathcal{M}_{k-1} el conjunto de muestras obtenidas de $f_{G^{(k-1)}|W^{(k-1)}}(g^{(k-1)}|w^{(k-1)})$. Se calcula la mediana y la desviación estándar m_{x_k} , s_{x_k} , m_{y_k} , s_{y_k} , m_{z_k} y s_{z_k} , de cada uno de los conjuntos $\mathcal{D}_{x_k} = \{x(T_k - R - A; q) \mid q \in \mathcal{M}_{k-1}\}$, $\mathcal{D}_{y_k} = \{y(T_k - R - A; q) \mid q \in \mathcal{M}_{k-1}\}$ y $\mathcal{D}_{z_k} = \{z(T_k - R - A; q) \mid q \in \mathcal{M}_{k-1}\}$, respectivamente. Las distribuciones a priori propuestas son

$$f_{X_k}(x_k) = N(x_k|m_{x_k}, s_{x_k}^2), \quad f_{Y_k}(y_k) = N(y_k|m_{y_k}, s_{y_k}^2) \quad \text{y} \quad f_{Z_k}(z_k) = N(z_k|m_{z_k}, s_{z_k}^2). \quad (2.22)$$

Sustituyendo las expresiones de (2.22) en (2.21), obtenemos que

$$f_{G^{(k)}}(g^{(k)}) = N(x_k|m_{x_k}, s_{x_k}^2)N(y_k|m_{y_k}, s_{y_k}^2)N(z_k|m_{z_k}, s_{z_k}^2). \quad (2.23)$$

2.7.6. Prior para el caso II y $k = 0$

En este caso $g^{(0)} = (x_0, y_0, z_0, r_0)$. Aplicamos Naïve Bayes a las cantidades desconocidas para obtener que

$$f_{G^{(0)}}(g^{(0)}) = f_{X_0}(x_0)f_{Y_0}(y_0)f_{Z_0}(z_0)f_{R_0}(r_0). \quad (2.24)$$

Aquí no podemos repetir lo hecho en el caso I ya que no conocemos el valor exacto de r_0 . Sin embargo, sí conocemos el rango en el que se encuentra. Para cada $r \in \{1, 26, \dots, 99\}$, resolvemos numéricamente el sistema (2.3) con condición inicial $(0, 1, 0)$, $\sigma = 10$ y $b = 8/3$, desde el tiempo $t = 0$ hasta $t = 1000$ con el método LSODA. Reunimos los valores absolutos de los datos generados de x , y y z en los conjuntos A_x , A_y y A_z , respectivamente. Sean a_x , a_y y a_z los valores máximos de A_x , A_y y A_z , respectivamente. Sean $b_x > a_x$, $b_y > a_y$ y $b_z > a_z$. Como no tenemos más

información de las condiciones iniciales o de r_0 , se proponen distribuciones uniformes como priors. Las propuestas son

$$f_{X_0}(x_0) = U_{(b_x, -b_x)}(x_0), \quad f_{Y_0}(y_0) = U_{(b_y, -b_y)}(y_0), \quad f_{Z_0}(z_0) = U_{(b_z, -b_z)}(z_0) \quad \text{y} \quad f_{R_0}(r_0) = U_{(0,100)}(r_0). \quad (2.25)$$

Sustituyendo las expresiones de (2.25) en (2.24), obtenemos que

$$f_{G^{(0)}}(g^{(0)}) = U_{(b_x, -b_x)}(x_0)U_{(b_y, -b_y)}(y_0)U_{(b_z, -b_z)}(z_0)U_{(0,100)}(r_0). \quad (2.26)$$

2.7.7. Prior para el caso II y $k > 0$

Aplicando Naïve Bayes obtenemos que

$$f_{G^{(k)}}(g^{(k)}) = f_{X_k}(x_k)f_{Y_k}(y_k)f_{Z_k}(z_k)f_{R_k}(r_k). \quad (2.27)$$

Primero, nos fijamos en las muestras del número de Rayleigh, \mathcal{R}_{k-1} , provenientes de la distribución posterior del paso $k-1$, $f_{G^{(k-1)}|W^{(k-1)}}(g^{(k-1)}|w^{(k-1)})$. Calculamos la mediana y la desviación estándar de \mathcal{R}_{k-1} , m_{r_k} y s_{r_k} , respectivamente, y le ajustamos una distribución normal, así, la distribución prior propuesta para el número de Rayleigh es

$$f_{R_k}(r_k) = N(r_k|m_{r_k}, s_{r_k}^2). \quad (2.28)$$

Ahora, sea \mathcal{M}_{k-1} el conjunto de muestras obtenidas de $f_{G^{(k-1)}|W^{(k-1)}}(g^{(k-1)}|w^{(k-1)})$. Se calcula la mediana y la desviación estándar m_{x_k} , s_{x_k} , m_{y_k} , s_{y_k} , m_{z_k} y s_{z_k} de cada uno de los conjuntos $\mathcal{D}_{x_k} = \{x(T_k - R - A; q) \mid q \in \mathcal{M}_{k-1}\}$, $\mathcal{D}_{y_k} = \{y(T_k - R - A; q) \mid q \in \mathcal{M}_{k-1}\}$ y $\mathcal{D}_{z_k} = \{z(T_k - R - A; q) \mid q \in \mathcal{M}_{k-1}\}$, respectivamente. Las distribuciones a priori propuestas para las condiciones iniciales son

$$f_{X_k}(x_k) = N(x_k|m_{x_k}, s_{x_k}^2), \quad f_{Y_k}(y_k) = N(y_k|m_{y_k}, s_{y_k}^2) \quad \text{y} \quad f_{Z_k}(z_k) = N(z_k|m_{z_k}, s_{z_k}^2). \quad (2.29)$$

Sustituyendo las expresiones de (2.28) y (2.29) en (2.27), obtenemos que

$$f_{G^{(k)}}(g^{(k)}) = N(x_k|m_{x_k}, s_{x_k}^2)N(y_k|m_{y_k}, s_{y_k}^2)N(z_k|m_{z_k}, s_{z_k}^2)N(r_k|m_{r_k}, s_{r_k}^2). \quad (2.30)$$

3.1. Detalles de la implementación y resultados

Se llevan a cabo tres experimentos computacionales utilizando el método de **BSDA** que está en Algorithm 2 bajo las hipótesis de los casos I y II. Los experimentos I y II están bajo las hipótesis del caso I, mientras que el experimento III está bajo las hipótesis del caso II.

Los datos que se utilizan en este trabajo son datos sintéticos que se construyen a través de la relación (2.9). El proceso para ello es resolver numéricamente con el método **Radau** (se encuentra en la librería `scipy.solve_ivp` de la biblioteca Scipy de Python 3 [5, 21, 14]) al sistema de Lorenz, con una cierta condición inicial (x_d, y_d, z_d) y ciertos parámetros σ , r y b , desde el tiempo 0 hasta el tiempo 10, 100. A la solución numérica del tiempo 10, 000 al tiempo 10, 100 es a la que se le adiciona el ruido blanco, es decir, le sumamos valores generados por una variable normal con media 0 y desviación estándar $\hat{\sigma}_x$, $\hat{\sigma}_y$ y $\hat{\sigma}_z$, respectivamente, que son valores que se suponen conocidos y que se utilizan para construir la matriz de covarianzas (2.17). El método Radau fue elegido ya que es un método que resuelve numéricamente el sistema de Lorenz satisfactoriamente y que es diferente a los métodos usados en LSODA. La cantidad de datos que tenemos son 10 datos por día, los cuales son equiespaciados.

Los muestreos de las distribuciones posteriores presentes en cada experimento se lleva a cabo aplicando el PT descrito en el pseudocódigo Algorithm 1. Se muestran las trazas de cada uno de los muestreos, los cuales son realizados con las características descritas a continuación. Sea N_β el número de cadenas presentes en el PT. Siguiendo [9], los datos de tamaño de salto de las cadenas ρ_s con $s = 1, \dots, N_\beta$ y la elección del vector β en el PT son propuestos buscando un índice de aceptación en cada una de las cadenas del PT de aproximadamente 234×10^{-3} . A su vez, la elección de los valores del vector β se hizo por medio de la regla de dedo propuesta en [4], la cual dice que $T_m = a^{m-1}$ para alguna constante $a > 1$ y para cada $m = 1, \dots, N_\beta$ donde $T_m = 1/\beta_m$. El código que se utiliza para hacer los muestreos es el proporcionado en [9].

Se presentan gráficas de las ventanas de predicción de cada una de las variables. Para ello, en cada experimento se

fijan los valores de A , R , P y N . En cada una de las gráficas de las ventanas de predicción, las regiones comprendidas entre las líneas punteadas negras verticales, rojas verticales y verdes verticales delimitan a los intervalos de aprendizaje, retraso y predicción, respectivamente. Los puntos son datos sintéticos. Los datos negros son los que se utilizan para el aprendizaje. Los datos rojos que están en el intervalo de retraso no se utilizan en el aprendizaje y se puede o no tener conocimiento de ellos, se piensa que se están recolectando en el tiempo en el cual está corriendo el algoritmo. Los puntos verdes que están en el intervalo de predicción representan datos totalmente desconocidos que se recolectarán en el futuro. Con las soluciones calculadas en la ventana de predicción, se hace un cálculo de los cuantiles del 5 %, 25 %, 73 %, 95 %, de la mediana. También se encuentra una versión computacional del **MAP** (recordemos que el MAP de una distribución posterior $f(\theta|w)$ está definido como el elemento $\hat{\theta} = \arg \max_{\theta \in \Theta} f(\theta|w)$) para calcular la solución con esa condición inicial. Las curvas azul fuerte y naranja corresponden a la mediana y a la solución que tiene como condición inicial el MAP de la distribución posterior, respectivamente. Las regiones azul claro comprendidas entre los cuantiles del 5 % y 95 % y las regiones azul oscuro comprendidas entre los cuantiles del 25 % y 75 % se interpretan como las **regiones con el 90 % y 50 % de probabilidad**, respectivamente, de que los datos de las respectivas variables analizadas se encuentren ahí según las predicciones generadas por el método descrito en Algorithm 2. La implementación computacional se llevó a cabo en el equipo de súper cómputo del CIMAT, se escribió en Python y los códigos se encuentran en [Git-Rep-Tesina](#).

3.1.1. Experimento I

En este experimento se considera el caso I y se aplica el algoritmo descrito en el pseudocódigo Algorithm 2 con $A = 2$, $R = 2$, $P = 4$ y $N = 1$, lo cual implica que la ventana de predicción es de 8 días. Se aplican tres pasos del algoritmo, $k = 0, 1, 2$. Los datos se construyeron con la condición inicial $(6, 6, 6)$ y con $r = 28$, $b = \frac{8}{3}$ y $\sigma = 10$. Las trazas y las distribuciones posteriores se muestran en las figuras 3.1 y 3.2, respectivamente. Los parámetros de los muestreos de éste experimento se muestran en la tabla 3.1. En las figuras 3.3, 3.4 y 3.5 se muestran las ventanas de predicción de las variables x , y y z de este experimento.

3.1.2. Análisis de resultados: Experimento I

De las figuras 3.3 (a), 3.4 (a) y 3.5 (a), se observa que para $k = 0$ las predicciones dadas por la solución con condición inicial el MAP y la mediana siguen la tendencia de los datos de x , y y z en el intervalo de aprendizaje, el intervalo de retraso y los primeros tres días del intervalo de predicción. La solución con condición inicial el MAP, sigue la tendencia de los datos en el tiempo restante de la ventana de predicción en las tres variables, x , y y z . Las regiones de probabilidad del 50 % y del 90 % se van ampliando conforme avanza el tiempo, lo cual es de esperarse debido al incremento de incertidumbre en la predicción dada por la mediana.

Los resultados obtenidos en $k = 1$ en la predicción de las variables x , y y z no son del todo satisfactorios (figuras 3.3 (b), 3.4 (b) y 3.5 (b)). A partir de la mitad de aproximadamente la mitad del intervalo de retraso, ni la solución con condición inicial el MAP ni la mediana siguen la tendencia de los datos.

Los resultados para el paso $k = 2$ mostrados en las figuras 3.3 (c), 3.4 (c) y 3.5 (c), muestran una reducción en el área de las regiones de probabilidad del 50 y 90 %, lo cual es de esperado, ya que ahí se ve reflejado el conocimiento

previo adquirido a través de los pasos $k = 0$ y $k = 1$. Por otro lado, tanto la mediana como la solución con condición inicial el MAP siguen la tendencia de los datos en los intervalos de aprendizaje, retraso y predicción. Como es de esperarse y semejante a los pasos $k = 0$ y $k = 1$, las regiones de probabilidad del 50 y 90 % se van ampliando conforme avanza el tiempo, reflejo del aumento de incertidumbre.

Es de remarcar que si bien las regiones de probabilidad del 50 y 90 % son importantes en el sentido de que en el futuro nos puede dar una idea del comportamiento de los datos y una medida de incertidumbre en la predicción, en este caso, el lapso de tiempo en el que tenemos una incertidumbre razonable es pequeño, ya que las regiones se expanden súbitamente en lapsos de tiempo muy cortos, lo cual es consecuencia de la sensibilidad a las condiciones iniciales que presenta el sistema de Lorenz. Por otro lado, las distribuciones posteriores de las condiciones iniciales de cada paso resultan ser unimodales. En cada paso, las condiciones iniciales más probables encontradas por el muestreo se concentran cada vez más, lo cual es consecuencia de que se va adquiriendo más conocimiento previo en cada paso. Aunque éste es el objetivo del algoritmo, es decir, ir aprendiendo en cada paso para hacer mejores predicciones con zonas de alta probabilidad útiles para sacar conclusiones y tomar decisiones, el muestreo se va volviendo más complicado, ya que el **PT** debe buscar en regiones cada vez más pequeñas las condiciones iniciales más probables.

3.1.3. Experimento II

En este experimento se considera el caso I y se aplica el algoritmo descrito en el pseudocódigo Algorithm 2 con $A = 4$, $R = 2$, $P = 4$ y $N = 1$, lo cual implica que la ventana de predicción es de 10 días. Se aplican dos pasos del algoritmo, $k = 0, 1$. Los datos se construyeron con la condición inicial $(6, 6, 6)$ y con $r = 28$, $b = \frac{8}{3}$ y $\sigma = 10$. Las trazas y las distribuciones posteriores se muestran en las figuras 3.6 y 3.7, respectivamente. Los parámetros de los muestreos de éste experimento se muestran en la tabla 3.2. En las figuras 3.8, 3.8 y 3.8 se muestran las ventanas de predicción de las variables x , y y z de este experimento.

3.1.4. Análisis de resultados: Experimento II

Este experimento comparte las mismas cantidades de interés que el experimento I (que son las condiciones iniciales), la variante es que tiene dos días más en el intervalo de aprendizaje. El muestreo de las distribuciones posteriores mostradas en las figuras 3.7 (a) y (b) requirió el aumento en el número de cadenas de Markov en el PT y el aumento del parámetro a con respecto al experimento I (véanse las tablas 3.2 y 3.2). Interpretamos que si aumenta el intervalo de aprendizaje, los muestreos son más complicados de hacer.

De las figuras 3.8 (a), 3.9 (a) y 3.10 (a) podemos decir que para $k = 0$ tanto la mediana como la solución con condición inicial el MAP siguen la tendencia de los datos de x y y durante el intervalo de aprendizaje, el intervalo de retraso y los primeros dos días del intervalo de predicción, y siguen la tendencia de los datos de la variable z durante todo la ventana de predicción, respectivamente. A partir del inicio del tercer día del intervalo de predicción, las regiones del 50 y 90 % de probabilidad tienen una expansión súbita y tanto la mediana como la solución con condición inicial el MAP dejan de seguir la tendencia de los datos de x y y . Comparando el paso $k = 0$ de los experimentos I y II, tenemos la mediana y la solución con condición inicial el MAP siguen por un tiempo más

prolongado las tendencias de los datos en el experimento I que en el experimento II. Sin embargo, la expansión súbita de las regiones de probabilidad del experimento I se da al inicio del intervalo de predicción, mientras que en este experimento se dan al inicio del tercer día del intervalo de predicción, es decir, hay más incertidumbre en las predicciones dadas por las medianas en el experimento I.

De la figura 3.7, notamos que las condiciones iniciales más probables ocupan un espacio más pequeño para $k = 1$ en comparación con $k = 0$, debido a que en el paso $k = 1$ el método ya está tomando en cuenta toda la información previa proveniente del paso $k = 0$. Por otro lado, de las figuras 3.8 (b), 3.9 (b) y 3.10 (b), tenemos que la solución con condición inicial el MAP y la mediana siguen la tendencia de los datos de x , y y z durante el intervalo de aprendizaje, el intervalo de retraso y los primeros tres días del intervalo de predicción, para $k = 1$. La expansión súbita de la región del 50 % de probabilidad se da hasta el tercer día del intervalo de predicción para $k = 1$, mientras que para $k = 0$ se da desde el inicio del intervalo de predicción. Interpretamos que la incertidumbre en la predicción con la mediana es mayor en la predicción cuando $k = 0$. Adicionalmente, decimos que es claramente observable que las predicciones del paso $k = 1$ son mejores en el experimento II que en el experimento I, debido a que las predicciones siguen por más tiempo la tendencia de los datos y la incertidumbre en las predicciones con la mediana es menor. Interpretamos que es necesario tener suficientes días de entrenamiento para no perder la tendencia de los datos en el intervalo de retraso, como lo que pasó en el experimento I.

3.1.5. Experimento III

En este experimento se considera el caso II y se aplica el algoritmo descrito en el pseudocódigo Algorithm 2 con $A = 4$, $R = 2$, $P = 4$ y $N = 1$, lo cual implica que la ventana de predicción es de 10 días. Se aplican dos pasos del algoritmo, $k = 0, 1$. Los datos se construyeron con la condición inicial $(4, 7, 3)$ y con $r = 70$, $b = \frac{8}{3}$ y $\sigma = 10$. Las trazas y las distribuciones posteriores se muestran en las figuras 3.11 y 3.12, respectivamente. Los parámetros de los muestreos de este experimento se muestran en la tabla 3.3. En las figuras 3.13, 3.13 y 3.13 muestran las ventanas de predicción de las variables x , y y z de este experimento.

3.1.6. Análisis de resultados: Experimento III

Éste es el único experimento presentado que se encuentra bajo las hipótesis del caso II. El caso II fue planteado buscando dos objetivos: el primero es ver si el algoritmo es capaz, a partir de distribuciones prior uniformes para $k = 0$, de recuperar información sobre las condiciones iniciales y el parámetro r ; el segundo es ver cómo son las predicciones en comparación de los experimentos del caso I. Este caso es un pequeño acercamiento a lo que pasa en la realidad, usualmente no sólo se busca hacer predicciones confiables de ciertas variables, sino que se buscan los valores de otras cantidades de interés que no se pueden medir directamente y están representadas como parámetros en los sistemas dinámicos [3].

Para $k = 0$. De la figura 3.12 (a) observamos que los valores más probables de r está en el intervalo $[64.5, 71.0]$. Como el valor real de r que es igual a 70.0, entonces se logró obtener un intervalo de medida 6.5 que contiene al valor verdadero de r . La mediana para r está próxima a 69.0 y el MAP para r está en el intervalo $[70.0, 71.0]$. En las figuras 3.13 (a) y 3.14 (a), se muestra que tanto la solución con condición inicial el MAP y la mediana siguen

la tendencia de los datos de x , y y z desde el inicio del intervalo de aprendizaje hasta la mitad del intervalo de retraso, punto a partir del cual ninguna de las dos predicciones siguen las tendencias de los datos. Las regiones de probabilidad del 50 y 90 % se expanden súbitamente a partir de la mitad del intervalo de retraso.

Para $k = 1$. Observamos a partir de la figura 3.12 (b) que los puntos más probables están más concentrados que en $k = 0$. En el caso puntual del parámetro r , tenemos que su soporte se encuentra contenido en una pequeña vecindad al rededor del 69.99, comparando con $k = 0$, se redujo el soporte para r y se mejoró la predicción del valor de r (el cual es 70.0). En las figuras 3.13 (b), 3.14 (b) y 3.15 (b) se muestran las predicciones para las variables x , y y z , respectivamente. Las regiones del 90 % de probabilidad comienzan su aumento súbito de área al final del intervalo de retraso, mientras que las regiones del 50 % comienzan su aumento súbito de área a partir del inicio segundo día del intervalo de predicción, además, se observa una disminución en sus áreas en comparación con $k = 0$, lo cual interpretamos como una disminución en la incertidumbre de las predicciones dadas por las medianas. Tanto la mediana como la solución con condición inicial el MAP siguen la tendencia de los datos en el intervalo de retraso, lo cual representa una mejora en comparación del paso $k = 0$. Las medianas siguen la tendencia de los datos hasta la mitad del primer día del intervalo de predicción, mientras que la solución con condición inicial el MAP sigue la tendencia de los datos hasta el final del segundo día del intervalo de predicción. Las predicciones aceptables de este experimento son más cortas en comparación con las predicciones dadas por los experimentos I y II, lo cual es un reflejo de que el desconocimiento del verdadero valor del parámetro r tiene gran impacto en el rendimiento de las predicciones.

3.2. Figuras y tablas de los experimentos I, II y III.

Paso k	N	Burning	N_β	a	ρ_s
0	500,000	20 %	5	12,000	15×10^{-3} , 6×10^{-2} , 15×10^{-2} , 5×10^{-1} , 2×10^0
1	900,000	20 %	5	12,000	45×10^{-4} , 2×10^{-2} , 8×10^{-2} , 25×10^{-2} , 9×10^{-1}
2	600,000	20 %	5	12,000	8×10^{-5} , 27×10^{-5} , 8×10^{-3} , 55×10^{-3} , 0.18×10^{-2}

Tabla 3.1: Datos de los muestreos para el experimento I. Los tamaños de salto ρ_s para el muestreo de cada una de las versiones temperadas de la distribución final están listadas con respecto al orden $\beta_1 > \beta_2 > \dots > \beta_{N_\beta}$.

Paso k	N	Burning	N_β	a	ρ_s
0	500,000	20 %	6	19,000	58×10^{-5} , 58×10^{-4} , 25×10^{-3} , 7×10^{-2} , 15×10^{-2} , 4×10^0
1	400,000	20 %	7	19,000	6×10^{-8} , 21×10^{-8} , 38×10^{-6} , 31×10^{-5} , 176×10^{-5} , 12×10^{-3} , 43×10^{-3}

Tabla 3.2: Datos de los muestreos para el experimento II. Los tamaños de salto ρ_s para el muestreo de cada una de las versiones temperadas de la distribución final están listadas con respecto al orden $\beta_1 > \beta_2 > \dots > \beta_{N_\beta}$.

Paso k	N	Burning	N_β	a	ρ_s
0	500,000	20 %	6	19,000	3×10^{-4} , 8×10^{-4} , 5×10^{-3} , 1×10^{-1} , 4×10^{-1} , 12×10^0
1	400,000	20 %	7	19,000	5×10^{-6} , 21×10^{-6} , 55×10^{-6} , 94×10^{-6} , 0.0005×10^{-4} , 3×10^{-2} , 179×10^{-3}

Tabla 3.3: Datos de los muestreos para el experimento III. Los tamaños de salto ρ_s para el muestreo de cada una de las versiones temperadas de la distribución final están listadas con respecto al orden $\beta_1 > \beta_2 > \dots > \beta_{N_\beta}$.

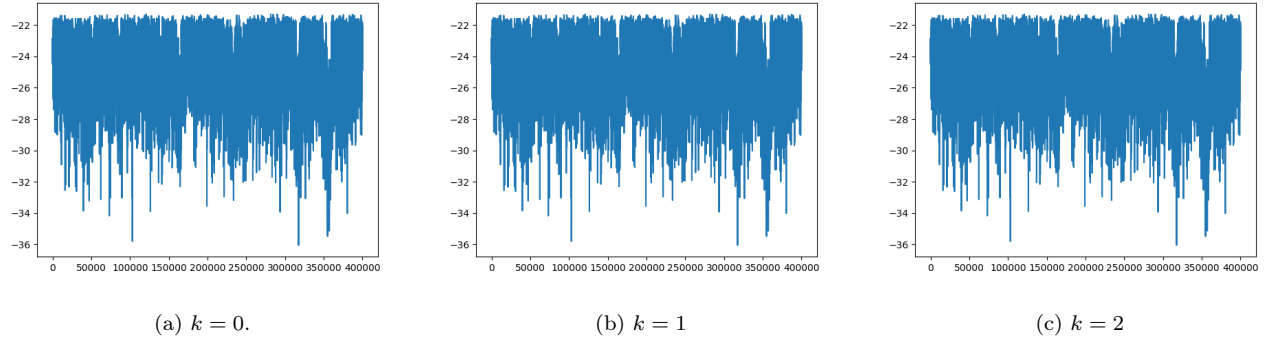


Figura 3.1: Trazas de cada uno de los pasos del Experimento I caso I.

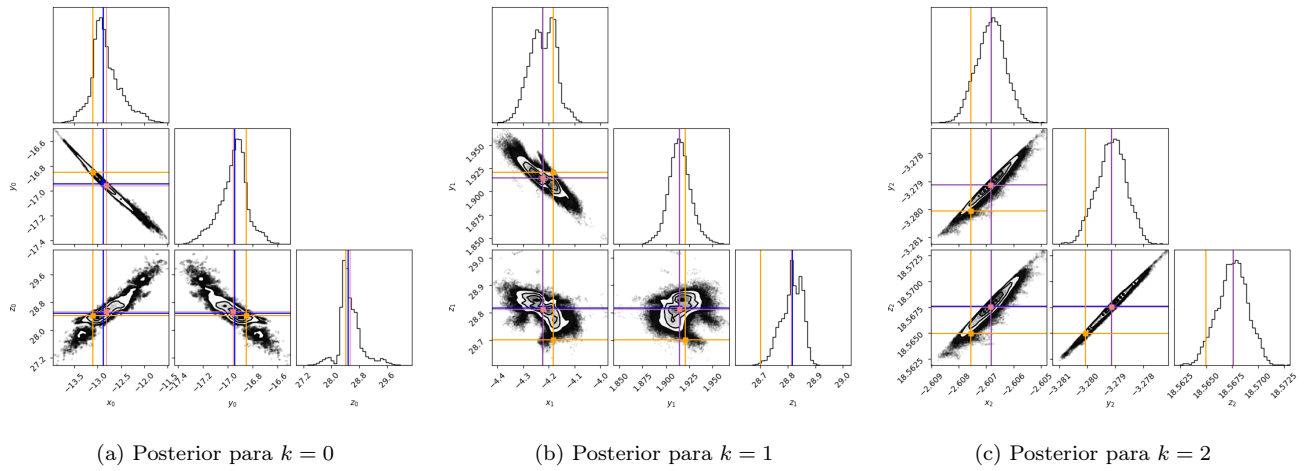


Figura 3.2: Distribuciones posteriores del experimento I del caso I. La línea rosa representa a la media, la azul a la mediana y la naranja al MAP de la distribución posterior.

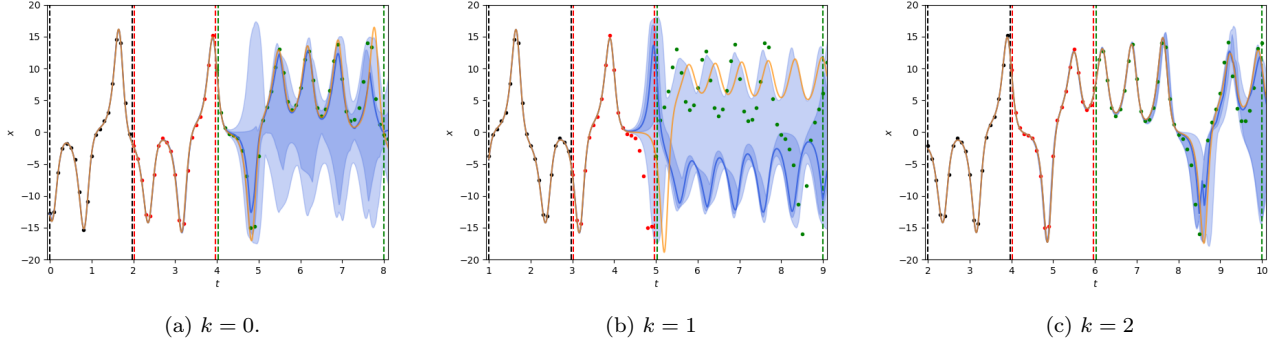


Figura 3.3: Ventana de predicción de la variable x en los pasos $k = 0, 1, 2$ del experimento I. Las líneas punteadas negras, rojas y verdes verticales comprenden a los intervalos de aprendizaje, retraso y predicción, respectivamente. La línea azul fuerte corresponde a la mediana y la naranja a la solución con condición inicial el MAP.

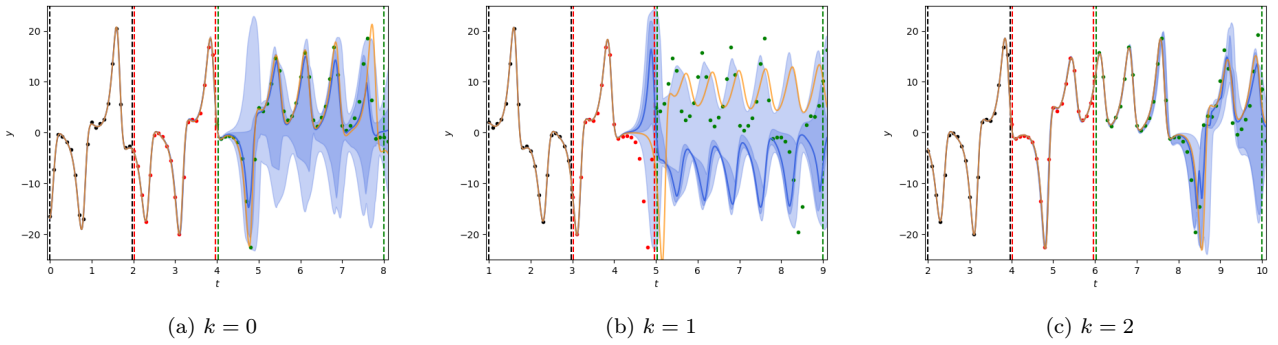


Figura 3.4: Ventana de predicción de la variable y en los pasos $k = 0, 1, 2$ del experimento I. Las líneas punteadas negras, rojas y verdes verticales comprenden a los intervalos de aprendizaje, retraso y predicción, respectivamente. La línea azul fuerte corresponde a la mediana y la naranja a la solución con condición inicial el MAP.

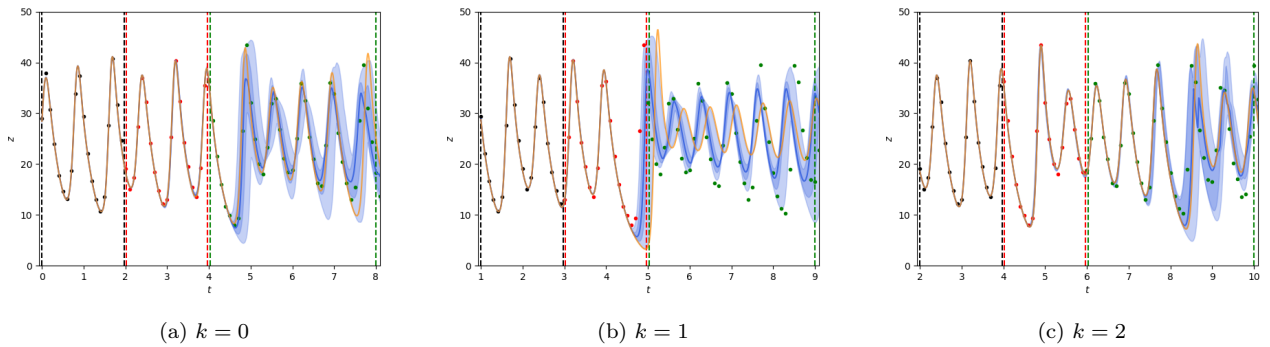


Figura 3.5: Ventana de predicción de la variable z en los pasos $k = 0, 1, 2$ del experimento I. Las líneas punteadas negras, rojas y verdes verticales comprenden a los intervalos de aprendizaje, retraso y predicción, respectivamente. La línea azul fuerte corresponde a la mediana y la naranja a la solución con condición inicial el MAP.

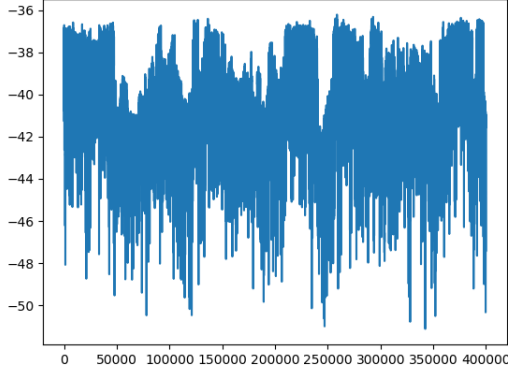
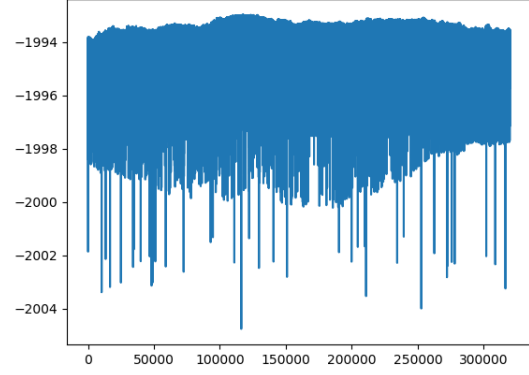

 (a) $k = 0$.

 (b) $k = 1$

Figura 3.6: Trazas de cada uno de los pasos del experimento II caso I.

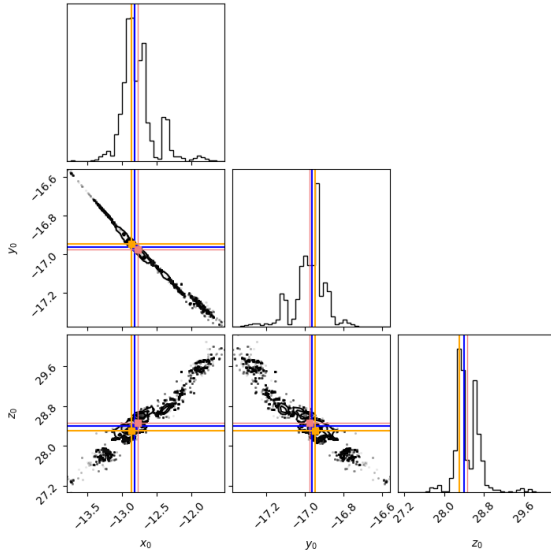
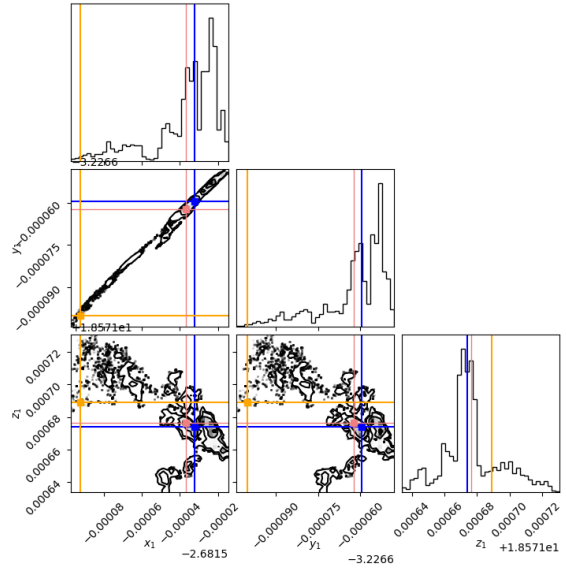

 (a) Posterior para $k = 0$

 (b) Posterior para $k = 1$

Figura 3.7: Distribuciones posteriores del experimento II del caso I. La línea rosa representa a la media, la azul a la mediana y la naranja al MAP de la distribución posterior.

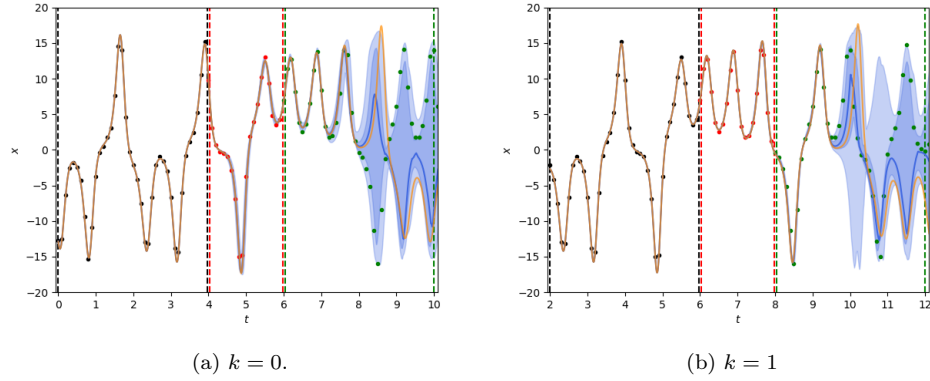


Figura 3.8: Ventanas de predicción de la variable x en el experimento II. Las líneas punteadas negras, rojas y verdes verticales comprenden a los intervalos de aprendizaje, retraso y predicción, respectivamente. La línea azul fuerte corresponde a la mediana y la naranja a la solución con condición inicial el MAP.

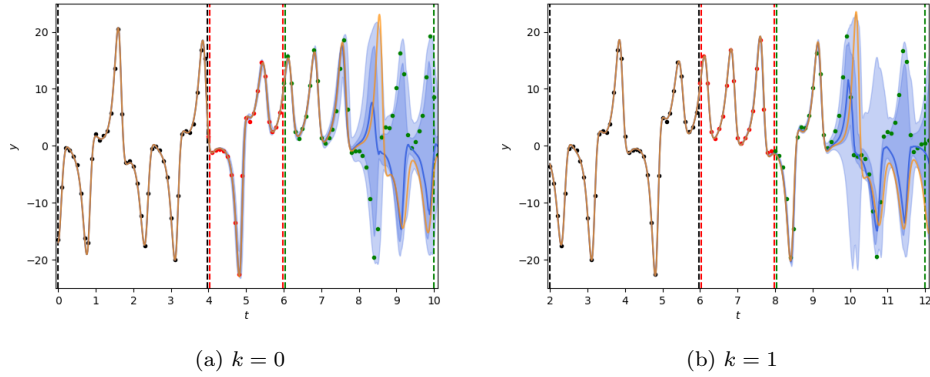


Figura 3.9: Ventanas de predicción de la variable y en el experimento II. Las líneas punteadas negras, rojas y verdes verticales comprenden a los intervalos de aprendizaje, retraso y predicción, respectivamente. La línea azul fuerte corresponde a la mediana y la naranja a la solución con condición inicial el MAP.

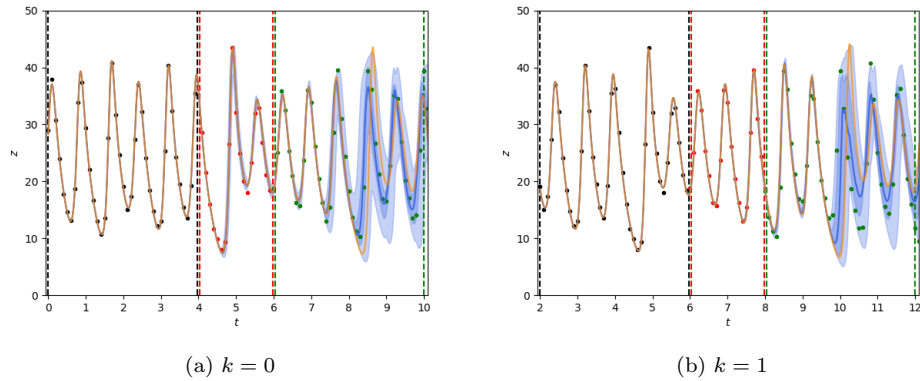


Figura 3.10: Ventanas de predicción de la variable z en el experimento II. Las líneas punteadas negras, rojas y verdes verticales comprenden a los intervalos de aprendizaje, retraso y predicción, respectivamente. La línea azul fuerte corresponde a la mediana y la naranja a la solución con condición inicial el MAP.

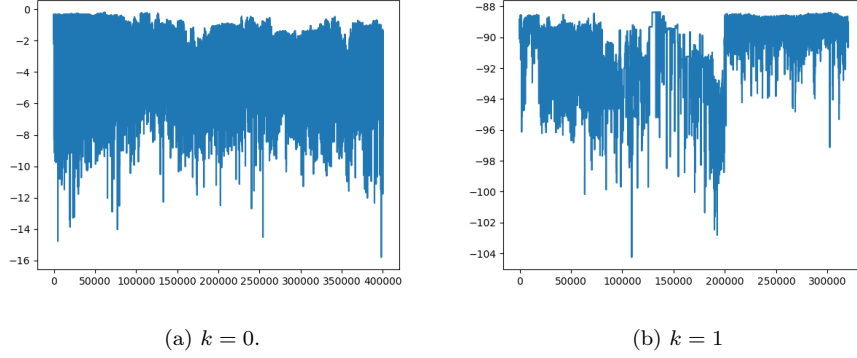


Figura 3.11: Trazas de cada uno de los pasos del experimento III del caso II.

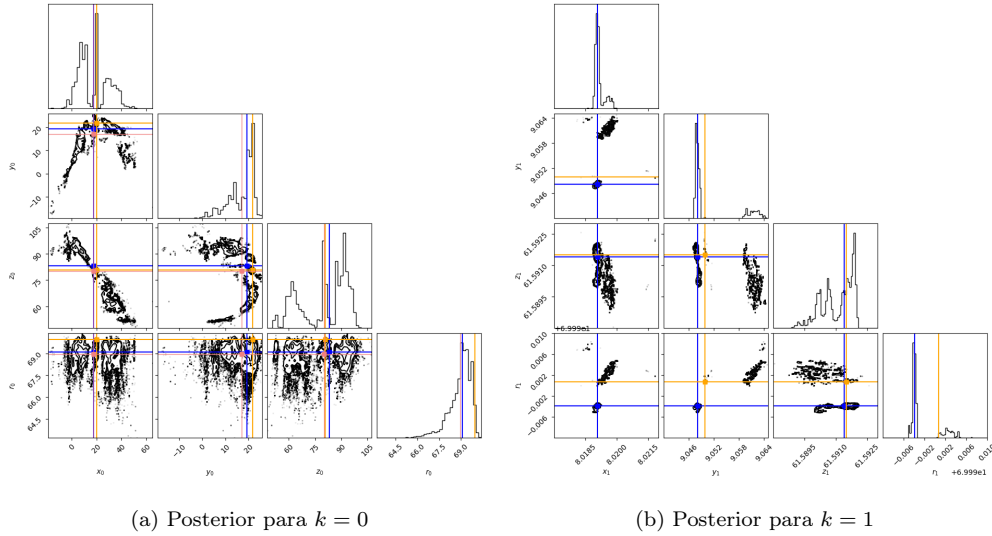


Figura 3.12: Distribuciones posteriores del experimento III del caso II. La línea rosa representa a la media, la azul a la mediana y la naranja al MAP de la distribución posterior.

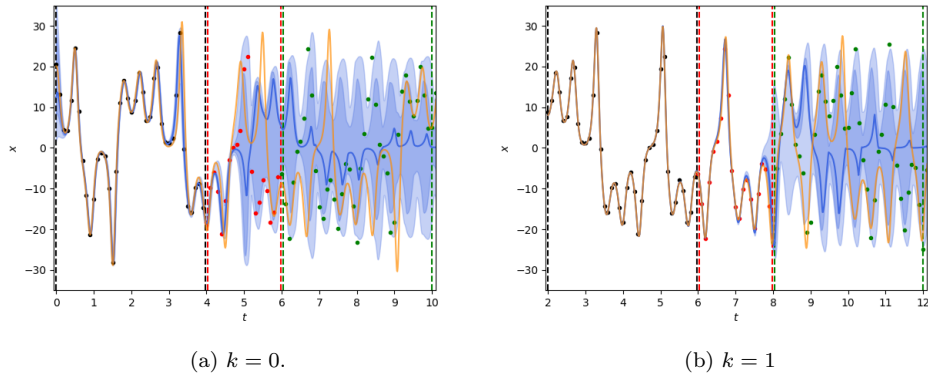


Figura 3.13: Ventanas de predicción de la variable x en el experimento III. Las líneas punteadas negras, rojas y verdes verticales comprenden a los intervalos de aprendizaje, retraso y predicción, respectivamente. La línea azul fuerte corresponde a la mediana y la naranja a la solución con condición inicial el MAP.

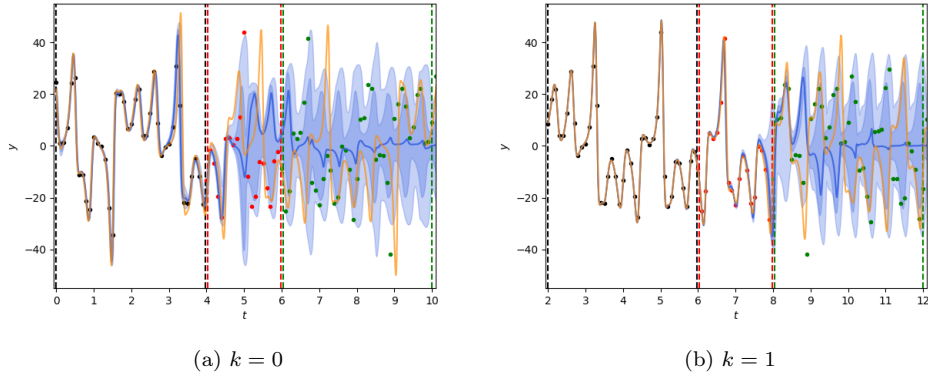


Figura 3.14: Ventanas de predicción de la variable y en el experimento III. Las líneas punteadas negras, rojas y verdes verticales comprenden a los intervalos de aprendizaje, retraso y predicción, respectivamente. La línea azul fuerte corresponde a la mediana y la naranja a la solución con condición inicial el MAP.

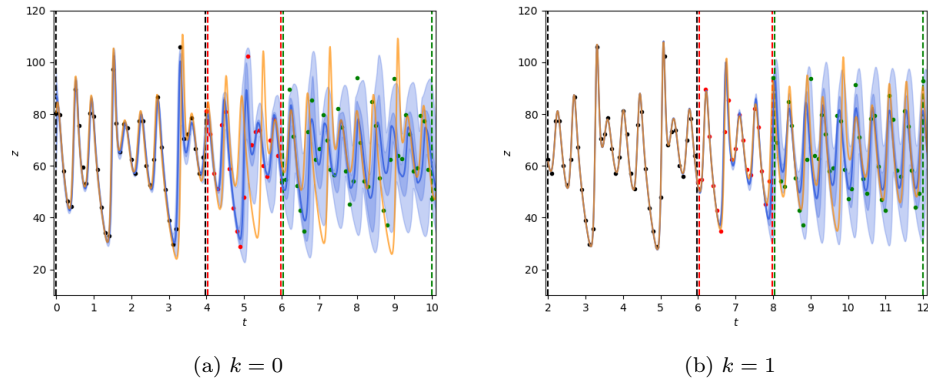


Figura 3.15: Ventanas de predicción de la variable z en el experimento III. Las líneas punteadas negras, rojas y verdes verticales comprenden a los intervalos de aprendizaje, retraso y predicción, respectivamente. La línea azul fuerte corresponde a la mediana y la naranja a la solución con condición inicial el MAP.

En esta tesina se ha aplicado un método tipo **BSDA** al sistema dinámico de Lorenz. Se ilustra a través de ejemplos que el método es capaz de predecir a corto plazo las variables de estado x , y y z . También, podemos concluir que el método tiene un excelente rendimiento al hacer los ajustes de los datos de las variables x , y y z de la etapa de entrenamiento y la etapa de retraso (nowcasting). Basándonos en el Experimento III, decimos que el método de **BSDA** es aceptable recuperando los probables valores del parámetro r en desconocimiento de las condiciones iniciales, utilizando suficiente información en el intervalo de entrenamiento y con distribuciones prior uniformes.

La contribución principal de este trabajo es ilustrar como es posible predecir secuencialmente el estado de un sistema caótico balanceando los periodos de aprendizaje, retraso y predicción. También, podemos identificar una complejidad mayor en la geometría de las distribuciones posteriores conforme se aumentan los días de aprendizaje, lo cual, dificulta enormemente el muestreo y la aplicación práctica del método. Por otro lado, en el caso del Experimento I con $k = 1$, se interpreta que aunque la distribución posterior es unimodal (relativamente sencilla de muestrear), la información en el intervalo de entrenamiento no fue suficiente para predecir de mejor manera. Es decir, interpretamos que debe existir un balance entre los días del aprendizaje utilizados y la dificultad del muestreo: más días de entrenamiento se reflejan en una predicción más confiable en tiempos cortos (con menos incertidumbre), pero en una distribución posterior más complicada de muestrear.

Este trabajo puede ser extendido ya sea proponiendo otros modelos para la verosimilitud, utilizar otras técnicas que hagan más ameno el muestreo, proponiendo distribuciones no necesariamente normales como priors, cambiar el sistema dinámico a otro más realista y utilizar datos reales.

- [1] M. Asch, M. Bocquet, and M. Nodet. *Data Assimilation: Methods, Algorithms, and Applications*. Fundamentals of Algorithms. Society for Industrial and Applied Mathematics, 2016.
- [2] Marc Bocquet et al. Introduction to the principles and methods of data assimilation in geosciences. *Notes de cours, École des Ponts ParisTech*, 2014.
- [3] Maria L. Daza-Torres, Marcos A. Capistrán, Antonio Capella, and J. Andrés Christen. Bayesian sequential data assimilation for covid-19 forecasting. *arXiv e-prints*, page arXiv:2103.06152, March 2021.
- [4] David J. Earl and Michael W. Deem. Parallel tempering: theory, applications, and new perspectives. *Physical chemistry chemical physics : PCCP*, 7 23:3910–6, 2005.
- [5] E. Hairer and G. Wanner. *Solving Ordinary Differential Equations II: Stiff and Differential-Algebraic Problems*. Springer Series in Computational Mathematics. Springer Berlin Heidelberg, 2010.
- [6] David J Hand and Keming Yu. Idiot’s bayes—not so stupid after all? *International statistical review*, 69(3):385–398, 2001.
- [7] A.C. Hindmarsh and Lawrence Livermore Laboratory. *ODEPACK, a Systematized Collection of ODE Solvers*. Lawrence Livermore National Laboratory, 1982.
- [8] Arieh Iserles. *A First Course in the Numerical Analysis of Differential Equations*. Cambridge Texts in Applied Mathematics. Cambridge University Press, 2 edition, 2008.
- [9] Jonas Latz, Juan P Madrigal-Cianci, Fabio Nobile, and Raúl Tempone. Generalized parallel tempering on bayesian inverse problems. *Statistics and Computing*, 31(5):1–26, 2021.
- [10] Stephen Lynch. *Dynamical Systems with Applications using Python*. Springer International Publishing, 2018.
- [11] John I Marden. Mathematical statistics: Old school. *University of Illinois*, 2012.
- [12] Timothy N Palmer. Extended-range atmospheric prediction and the lorenz model. *Bulletin of the American Meteorological Society*, 74(1):49–66, 1993.

-
- [13] L. Perko. *Differential Equations and Dynamical Systems*. Texts in Applied Mathematics. Springer New York, 2008.
 - [14] Linda Petzold. Automatic selection of methods for solving stiff and nonstiff systems of ordinary differential equations. *SIAM Journal on Scientific and Statistical Computing*, 4(1):136–148, 1983.
 - [15] C. Robert and G. Casella. *Monte Carlo Statistical Methods*. Springer Texts in Statistics. Springer New York, 2013.
 - [16] Malcolm Sambridge. A Parallel Tempering algorithm for probabilistic sampling and multimodal optimization. *Geophysical Journal International*, 196(1):357–374, 10 2013.
 - [17] Jörg Schwinger, Jerry F Tjiputra, Christoph Heinze, Laurent Bopp, James R Christian, Marion Gehlen, Tatiana Ilyina, Chris D Jones, David Salas-Méla, Joachim Segschneider, et al. Nonlinearity of ocean carbon cycle feedbacks in cmip5 earth system models. *Journal of Climate*, 27(11):3869–3888, 2014.
 - [18] C. Sparrow. *The Lorenz Equations: Bifurcations, Chaos, and Strange Attractors*. Applied Mathematical Sciences. Springer New York, 2012.
 - [19] Julien Clinton Sprott. *Chaos and Time-Series Analysis*, volume 69. Oxford university press Oxford, 2003.
 - [20] Steven H. Strogatz. *Nonlinear dynamics and chaos. With applications to physics, biology, chemistry, and engineering*. Boulder, CO: Westview Press, 1994.
 - [21] Pauli Virtanen, Ralf Gommers, Travis E. Oliphant, Matt Haberland, Tyler Reddy, David Cournapeau, Evgeni Burovski, Pearu Peterson, Warren Weckesser, Jonathan Bright, Stéfan J. van der Walt, Matthew Brett, Joshua Wilson, K. Jarrod Millman, Nikolay Mayorov, Andrew R. J. Nelson, Eric Jones, Robert Kern, Eric Larson, C J Carey, İlhan Polat, Yu Feng, Eric W. Moore, Jake VanderPlas, Denis Laxalde, Josef Perktold, Robert Cimrman, Ian Henriksen, E. A. Quintero, Charles R. Harris, Anne M. Archibald, Antônio H. Ribeiro, Fabian Pedregosa, Paul van Mulbregt, and SciPy 1.0 Contributors. SciPy 1.0: Fundamental Algorithms for Scientific Computing in Python. *Nature Methods*, 17:261–272, 2020.
 - [22] Xingyuan Wang and Mingjun Wang. A hyperchaos generated from lorenz system. *Physica A: Statistical Mechanics and its Applications*, 387(14):3751–3758, 2008.
 - [23] Caibin Zeng and Qigui Yang. Dynamics of the stochastic lorenz chaotic system with long memory effects. *Chaos: An Interdisciplinary Journal of Nonlinear Science*, 25(12):123114, 2015.
 - [24] Yagang Zhang, Guifang Pan, Chenhong Zhang, and Yuan Zhao. Wind speed prediction research with emd-bp based on lorenz disturbance. *Journal of Electrical Engineering*, 70:198–207, 2019.
 - [25] Yagang Zhang, Jingyun Yang, Kangcheng Wang, Zengping Wang, and Yinding Wang. Improved wind prediction based on the lorenz system. *Renewable Energy*, 81:219–226, 2015.
 - [26] Ying Zhang, Kayo Ide, and Eugenia Kalnay. Bred vectors of the lorenz63 system. *Advances in Atmospheric Sciences*, 32(11):1533–1538, 2015.