



Benemérita Universidad Autónoma de Puebla

Facultad de Ciencias Físico Matemáticas

Identificación de parámetros variables en modelos
dinámicos mediante filtros Kalman y Savitzky-Golay

Tesis presentada al

Posgrado en Física Aplicada

como requisito parcial para la obtención del grado de

MAESTRO EN CIENCIAS

por

Víctor Manuel Fuentes Mirón

Asesorado por

Dr. Jorge Velázquez Castro

Dr. Javier Miguel Hernández López

Puebla Pue.
17 de octubre de 2025



Benemérita Universidad Autónoma de Puebla

Facultad de Ciencias Físico Matemáticas

Identificación de parámetros variables en modelos
dinámicos mediante filtros Kalman y Savitzky-Golay

Tesis presentada al

Posgrado en Física Aplicada

como requisito parcial para la obtención del grado de

MAESTRO EN CIENCIAS

por

Víctor Manuel Fuentes Mirón

Asesorado por

Dr. Jorge Velázquez Castro

Dr. Javier Miguel Hernández López

Puebla Pue.
17 de octubre de 2025

Identificación de parámetros variables en modelos dinámicos mediante filtros Kalman y Savitzky-Golay

Víctor Manuel Fuentes Mirón

17 de octubre de 2025

Título: Identificación de parámetros variables en modelos dinámicos mediante filtros Kalman y Savitzky-Golay

Estudiante: VÍCTOR MANUEL FUENTES MIRÓN

COMITÉ

Dr. Benito De Celis Alonso
Presidente

Dra. Beatriz Bonilla Capilla
Secretario

Dr. José Jacobo Oliveros Oliveros
Vocal

Dr. Jorge Velázquez Castro
Asesor

Dr. Javier Miguel Hernández López
Asesor

Agradecimientos

Considero indispensable agradecer a la Secretaría de Ciencia, Humanidades, Tecnología e Innovación (SECIHTI) por el apoyo económico otorgado para la realización de mis estudios de maestría.

A mi asesor, el Dr. Jorge Velázquez Castro, por su paciencia, guía y dedicación a lo largo de este proceso. Aprecio profundamente sus consejos y el tiempo que compartimos. Sus enseñanzas trascienden este y cualquier otro trabajo académico que pueda emprender. Agradezco también al jurado calificador por su disposición y la energía invertida en la evaluación de este trabajo.

A mi madre, Eva Guadalupe Mirón Mendoza, cuyo ejemplo de determinación y fortaleza ha sido mi mayor inspiración. Este logro está completamente dedicado a ella, porque en su temple encontré la voluntad necesaria para culminar este camino.

A mis amigas Victoria y Miriam, por acompañarme más allá de la licenciatura y por estar presentes en cada momento en que las necesité, incluso a la distancia.

A mi novio Ángel, por su escucha atenta, su apoyo incondicional y su comprensión durante las etapas más difíciles de este proceso.

Finalmente, a mis hermanos, amigos y compañeros, por su compañía, aliento, apoyo moral y por los momentos compartidos que hicieron este trayecto más llevadero.

Índice general

Resumen	VII
1. Introducción	1
1.1. Objetivos	2
2. Fundamentos Teóricos	3
2.1. Redes de Regulación Genética	3
2.1.1. Modelación matemática de las GNR	4
2.2. El Represilador	6
2.3. Filtro Kalman	8
2.3.1. Estimaciones escalares	8
2.3.2. Fusión de 2 estimaciones escalares	10
2.3.3. Fusión de varias estimaciones escalares	11
2.3.4. Fusión de Incrementos	12
2.3.5. Estimaciones vectoriales	12
2.3.6. Fusión de múltiples estimaciones vectoriales	13
2.3.7. Para 2 Estimaciones Vectoriales	14
2.3.8. Best Linear Unbiased Estimator	15
2.4. Filtro Kalman para Sistemas lineales	17
2.4.1. Modelo de la Evolución del Estado y la predicción	18
2.4.2. Fusión de Observaciones Completas	18
2.4.3. Fusión de las Observaciones Parciales de un estado	19
2.5. Filtro Kalman Extendido (EKF)	21
2.6. Filtro de Savitzky-Golay	23
2.6.1. Implementación	24
3. El Represilador y sus mediciones sintéticas	25
3.1. El represilador mediante diferencias finitas de Euler	25
3.2. Simulación de datos para el filtro Kalman	27
4. Estimación del represilador mediante Filtro Kalman Extendido	29
4.1. Ecuaciones para la estimación de las variables del represilador mediante el filtro Kalman Extendido	30
4.2. Estimación de las variables suavizando mediante el filtro Savitzky-Golay después del filtro Kalman	33
4.2.1. Estimación por filtro Kalman	33
4.2.2. Suavizado de la curva mediante filtro Savitzky-Golay	35
4.3. Estimación de las variables suavizando mediante el filtro Savitzky-Golay antes del filtro Kalman	36
4.3.1. Suavizado de las mediciones usando el filtro S-G	36
4.3.2. Estimación por filtro Kalman usando mediciones suavizadas	37

5. Estimación de los parámetros del represilador	39
5.1. Interpolación	42
5.1.1. Interpolación de la estimación Kalman	42
5.1.2. Interpolación de la estimación Kalman-Savitzky-Golay (KSG)	43
5.2. Cálculo de Parámetros	44
5.2.1. Parámetros usando la estimación Kalman	44
5.2.2. Parámetros usando la estimación KSG	46
5.3. Estimación de 6 parámetros diferentes para el represilador	48
6. Conclusión	51
Bibliografía	53

Resumen

La observación de los procesos de represión y activación génica ha permitido el desarrollo de modelos matemáticos conocidos como redes de regulación genética (GRNs, por sus siglas en inglés). Estas redes dinámicas no solo ayudan a comprender mejor los mecanismos de regulación natural, sino que también permiten diseñar nuevas redes de regulación celular mediante ingeniería genética [1].

Actualmente, el estudio de las GRNs es esencial para mejorar los tratamientos contra diversas enfermedades, optimizar el diseño de experimentos y desarrollar biomarcadores útiles para el diagnóstico [2]. Los genes interactúan entre sí y con su entorno celular, activándose o reprimiéndose mutuamente. Estas interacciones pueden desencadenar condiciones que propicien el desarrollo de enfermedades. Por ello, modelar matemáticamente estas redes de señalización es una herramienta valiosa en el diseño de terapias efectivas.

Uno de los principales desafíos en esta área es la presencia de ruido en los datos experimentales y la variabilidad en las tasas de reacción de los modelos, producto de las fluctuaciones estocásticas en sus parámetros.

En este trabajo, se propone aplicar el filtro de Kalman a un sistema dinámico con el objetivo de estimar sus parámetros, asumiendo que varían en el tiempo. Este filtro, ampliamente utilizado en sistemas de navegación, conducción automática y robótica, permite estimar las variables de un sistema dinámico combinando un modelo matemático con sus observaciones ruidosas.

Dado que los parámetros del sistema desempeñan un papel crucial en la dinámica del modelo, se ajustarán utilizando las estimaciones de estado proporcionadas por el filtro Kalman, de manera que el modelo reproduzca de manera consistente la dinámica observada en los datos.

La metodología se aplicará específicamente al modelo conocido como el represilador, una red de regulación genética compuesta por tres genes que se inhiben mutuamente. Se buscará estimar los parámetros del sistema como funciones del tiempo, de forma que reproduzcan adecuadamente la dinámica estimada por el filtro a partir del modelo.

Capítulo 1

Introducción

El estudio de las redes de regulación genética tiene aplicaciones muy amplias y muchas ya se han implementado de manera práctica. En control celular, se han diseñado circuitos de retroalimentación basados en proteínas sintéticas como LOCKR, que funcionan como un equivalente biológico de los controladores PID [3], y permiten regular funciones celulares con precisión, lo cual es útil para terapias contra el cáncer o lesiones neurológicas. También se han aplicado técnicas de interferencia de ARN para reconstruir redes regulatorias de células en tejidos, encargadas de estimular respuestas inmunitarias en presencia de antígenos en su superficie, lo que permite identificar genes clave en la respuesta inmune y posibles blancos terapéuticos [4].

En biología sintética, se han construido “contadores celulares” en bacterias *E. coli*, capaces de ejecutar instrucciones genéticas en serie, lo que abre la posibilidad de programar células vivas como si fueran circuitos electrónicos [1]. Estos avances se han implementado también con la finalidad de combatir el cáncer, diseñando circuitos lógicos que integran señales de microARNs para reconocer células tumorales y activar respuestas específicas, como apoptosis o emisión de fluorescencia, aumentando la precisión del diagnóstico y la terapia [5].

Estas aplicaciones muestran que el estudio de redes de regulación genética no solo ayuda a comprender la dinámica celular, sino que además se traduce en innovaciones concretas para la salud, la industria y la biotecnología, con un enorme potencial de impacto social y científico.

Las redes de regulación genética son representadas mediante sistemas de ecuaciones diferenciales ordinarias y muchas de ellas son no lineales. Conocer los parámetros implicados en estos sistemas dinámicos es fundamental porque determinan cómo evoluciona el sistema en el tiempo. Estos parámetros incluyen tasas de transcripción y traducción, constantes de degradación, coeficientes de interacción y constantes de sensibilidad como los coeficientes de Hill. Saber sus valores permite predecir con precisión si una red alcanzará un equilibrio, generará oscilaciones, responderá de forma rápida o lenta a estímulos externos. También ayuda a identificar puntos críticos donde un pequeño cambio puede alterar el comportamiento celular, lo cual es clave en enfermedades como el cáncer. En biología sintética, los parámetros son esenciales para diseñar circuitos genéticos estables y funcionales, mientras que en biotecnología permiten optimizar microorganismos para la producción de medicamentos, enzimas o biocombustibles. En conjunto, conocerlos ofrece una herramienta poderosa para entender, controlar y aplicar el funcionamiento de las células en contextos médicos e industriales [6].

La estimación de los parámetros no es una tarea simple al estudiar sistemas biológicos puesto que cada gen y proteína está asociada a múltiples parámetros, lo que hace que incluso redes pequeñas tengan un gran número de parámetros a estimar. Además, los datos experimentales suelen ser ruidosos y limitados: medir directamente tasas de reacción o afinidades en condiciones fisiológicas es complicado, y muchas veces solo se dispone de abundancias relativas de ARNm. A esto se suma que las redes son sistemas no lineales, donde diferentes combinaciones de parámetros pueden producir comportamientos similares, lo que dificulta su identificabilidad. También influye la variabilidad celular, ya que no todas las células de un mismo tejido muestran los mismos niveles de expresión. Por esta razón, es necesario desarrollar diferentes técnicas para la estimación de parámetros en redes de regulación genética, incluso si no todas las variables del sistema son observables.

1.1. Objetivos

Objetivo principal: Estimar la evolución temporal de los parámetros del represilador mediante datos.

Objetivos Específicos:

1. Generar datos sintéticos que simulen mediciones ruidosas.
2. Estimar el estado del sistema mediante el filtro Kalman.
3. Calcular los parámetros del represilador asumiendo que cambian en el tiempo usando la estimación realizada mediante el filtro Kalman.
4. Evaluar la eficiencia del método al compararlo con ajuste de mínimos cuadrados en una ventana móvil.

Capítulo 2

Fundamentos Teóricos

2.1. Redes de Regulación Genética

La célula vive en un ambiente complejo en el cuál existen múltiples señales incluyendo variables físicas como la temperatura y la presión ósmica, o bien, señales biológicas desde otras células. A esto la célula responde produciendo múltiples proteínas a través de *circuitos genéticos* compuestos por genes, promotores y proteínas reguladoras, útiles para realizar diversas tareas, por ejemplo: la reparación celular, transportar el azúcar a la célula para utilizarla o absorber principalmente glucosa cuando existe presencia de lactosa y glucosa simultáneamente [13]. En este último ejemplo, la célula lo consigue reprimiendo el gen que codifica la lactasa (la enzima que descompone la lactosa) mientras haya glucosa disponible [18]. Por lo anterior, conocer la activación o represión de los genes se vuelve una tarea con múltiples aplicaciones en la medicina personalizada. Saber cómo varía la activación o represión génica a lo largo del tiempo ofrece información valiosa para comprender procesos como la respuesta inmune, la infección o la diferenciación celular [2].

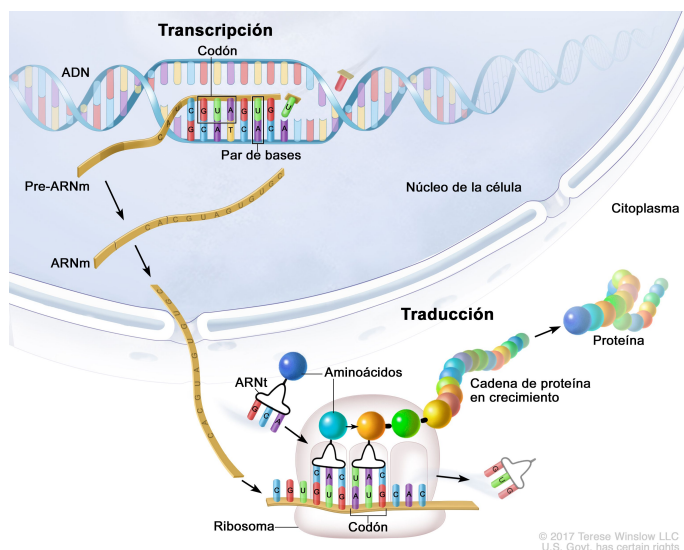


Figura 2.1: Representación gráfica de las etapas de transcripción y traducción del ADN [8].

Cuando una célula responde a señales internas o estímulos del entorno, lo hace regulando la expresión génica mediante dos procesos: la transcripción y la traducción.

Durante la transcripción, la enzima ARN polimerasa se une a una región específica del ADN conocida como *promotor*, ubicada al inicio de un gen. A partir de ahí, la ARN polimerasa sintetiza una molécula de **ARN mensajero (ARNm)**.

Posteriormente, en la etapa de traducción, el ARNm es transportado desde el núcleo hacia el citoplasma, donde es *traducido* por los ribosomas, que ensamblan aminoácidos en una secuencia específica para formar proteínas.

Algunas de estas proteínas recién sintetizadas actúan como *factores de transcripción*, los cuales retornan al núcleo, se unen a regiones reguladoras del ADN y modulan la expresión de otros genes, ya sea activando o reprimiendo su transcripción [7]. Ambos mecanismos constituyen componentes esenciales de las redes de regulación genética, representados en la figura 2.1.

Estos procesos componen las redes de regulación genética (GNR). (Figura 2.2)

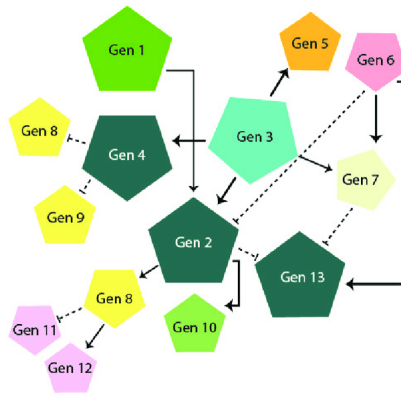


Figura 2.2: Representación gráfica de una red de regulación genética. Las flechas indican la activación de los genes y las líneas punteadas su represión [12].

2.1.1. Modelación matemática de las GNR

Las redes de regulación genética se modelan matemáticamente a través de sistemas de ecuaciones diferenciales ordinarias (ODEs). Las variables en estos sistemas son las concentraciones de moléculas de ARNm y se usan funciones de Hill para modelar la activación o represión en las expresiones de los genes.

Las funciones de Hill se han generalizado en la modelación de redes genéticas y tienen las siguientes formas para la activación y represión de genes, respectivamente:

$$f^+(p) = \frac{\alpha}{1 + \left(\frac{K}{p}\right)^n}, \quad f^-(p) = \frac{\alpha}{1 + \left(\frac{p}{K}\right)^n} \quad (2.1)$$

En estas ecuaciones α es la tasa máxima de transcripción (tasa a la que se producen copias de ARNm), p es la concentración del factor de transcripción (proteína de activación o represión), K es la constante de disociación (si tiene un valor grande entonces se necesita una alta concentración de p) y n es el coeficiente de Hill (modela qué tan *cooperativa* es la proteína p para tener efecto sobre la activación o represión del gen) [9], [10].

En el año 2000, se implementaron los primeros *circuitos biológicos* (el **Represilador** y el **Toggle Switch**) *in vivo* en bacterias *Escherichia coli* demostrando su importancia en el estudio experimental de las redes de regulación genética [18], [1]. De igual manera, se ha demostrado que las GRN que están relacionadas a la dinámica del crecimiento de tumores o al desarrollo de las enfermedades han mostrado puntos de vista importantes dentro de su evolución y han dirigido a descubrimientos de mejores períodos de dosificación de los medicamentos [11].

Un tipo particular de red de regulación génica (GRN) son las redes oscilatorias como el represilador. Están presentes en numerosos procesos biológicos como el ritmo circadiano, el metabolismo, la apoptosis y el ciclo celular. En estos casos, las oscilaciones naturales del sistema desempeñan un papel fundamental en el mantenimiento de la homeostasis. Alteraciones en estas dinámicas pueden conducir a diversas patologías, que van desde trastornos del sueño hasta distintos tipos de cáncer [19]. Es por ello que el estudio de este tipo de redes es útil para develar mecanismos en resistencia a fármacos y en la optimización de la terapia combinatoria dirigida.

2.2. El Represilador

Este fue uno de los primeros *circuitos genéticos sintéticos* diseñados para producir oscilaciones periódicas en la expresión génica dentro de una célula. Fue diseñado e implementado en *Escherichia coli* en el año 2000 por Michael Elowitz y Stanislas Leibler. Su objetivo era demostrar que es posible construir sistemas biológicos artificiales y predecibles. Este mecanismo involucra tres concentraciones de proteínas represoras p_i , y sus correspondientes concentraciones de ARN mensajero (ARNm) m_i . Cada una de las seis especies moleculares participa en la transcripción, la traducción y degradación.

El sistema está conformado por LaCI (represor lac) que inhibe al tetR (tetracycline-resistance transposon) y la proteína de tetR inhibe a su vez al gen CI-lambda, cuya proteína inhibe al represor LaCI. En el experimento se usó un *gen reportero* GFP (proteína fluorescente verde por sus siglas en inglés) para visualizar y medir la concentración de las proteínas [1].

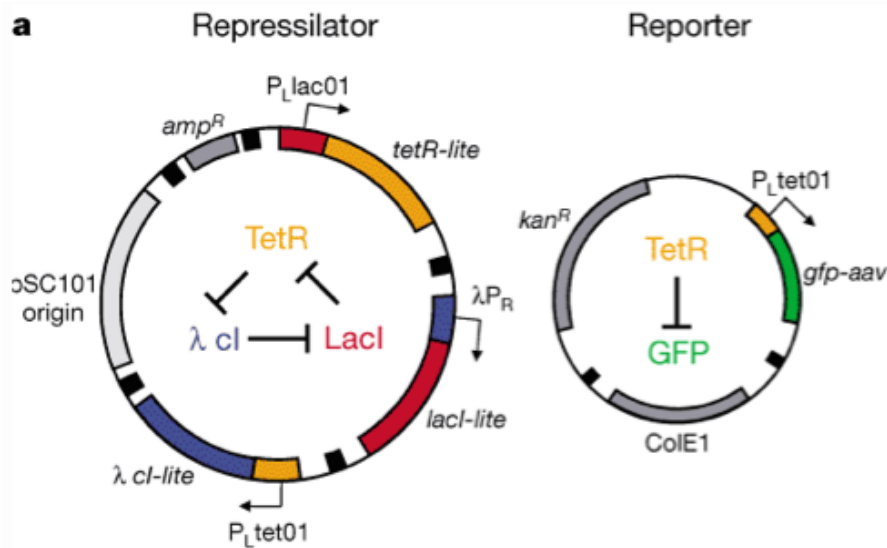


Figura 2.3: Representación del Represilador (Macmillan Publishers Ltd: Nature 403 335-338 copyright 2000).

El modelo matemático que lo describe es el siguiente.

$$\left\{ \begin{array}{l} \frac{dm_1}{dt} = -m_1 + \frac{\alpha}{(1 + p_3^n)} + \alpha_0 \\ \frac{dm_2}{dt} = -m_2 + \frac{\alpha}{(1 + p_1^n)} + \alpha_0 \\ \frac{dm_3}{dt} = -m_3 + \frac{\alpha}{(1 + p_2^n)} + \alpha_0 \\ \frac{dp_1}{dt} = -\beta(p_1 - m_1) \\ \frac{dp_2}{dt} = -\beta(p_2 - m_2) \\ \frac{dp_3}{dt} = -\beta(p_3 - m_3) \end{array} \right. \quad (2.2)$$

dónde:

- Las variables p_i son proporcionales a la concentración de las proteínas.
- Las variables m_i son proporcionales a las concentraciones del mRNA correspondiente a cada proteína.
- n es el coeficiente de Hill y representa la cooperatividad, refleja la multimerización de las proteínas necesarias para afectar al promotor.
- α_0 es el número de copias proteicas por célula producidas por un promotor dado (en presencia de la saturación del represor).
- α es la tasa máxima de transcripción.
- β denota la razón de la tasa de decaimiento de la proteína con la tasa de decaimiento del mRNA.

Los términos negativos representan la degradación de las moléculas. La función sigmoide (función de Hill) refleja la síntesis del mRNA desde el ADN, controlada por los promotores y se asume que las tres proteínas son idénticas, haciendo así que el modelo sea simétrico [1], [19].

2.3. Filtro Kalman

El filtro Kalman es un algoritmo matemático diseñado para estimar el estado de un sistema dinámico a partir de observaciones incompletas, ruidosas o inciertas. Su principal objetivo es usar la información provista por las predicciones de un modelo matemático con mediciones ruidosas para obtener una mejor estimación del estado real del sistema.

El filtro Kalman es uno de los algoritmos de estimación de estados de sistemas dinámicos más usados, ya que toma en cuenta de forma óptima los efectos del ruido en las mediciones, las perturbaciones y la incertidumbre del modelo, fusiona el Best Linear Unbiased Estimator (BLUE) y la estimación que produce la varianza mínima con los datos experimentales. Para poder aplicarlo debe hacerse sobre un sistema dinámico lineal por lo que, si no lo es, se encuentra una aproximación mediante una expansión en serie de Taylor [22].

Este filtro fue desarrollado en 1960 con la motivación de seguir y predecir trayectorias de objetos en movimiento, como misiles y vehículos espaciales. De forma típica, el filtro Kalman se usa en procesamiento y control automático de vehículos y, actualmente, también para resolver problemas computacionales como control de voltaje o frecuencia en procesadores, robots automáticos y en sistemas lineales con ruido Gaussiano [23].

2.3.1. Estimaciones escalares

¿Cómo funciona?

Partiendo de que por medio de i dispositivos de observación se toman mediciones x , entonces cada dispositivo tiene una *Densidad de Probabilidad* $p_i(x)$, no necesariamente gaussianas.

$$x_i \rightarrow p_i(\mu_i, \sigma_i^2)$$

Donde x_i representa una muestra aleatoria del conjunto de datos que forma la distribución p_i . Estas tienen cierta media μ_i y cierta varianza σ_i^2 que dependen directamente de los errores de medición.

Tipos de Error

Existen 2 tipos de error:

1. *Error Sistemático o bias (sesgo)*: Se trata de un error sistemático o un sesgo si la media μ_i de su distribución es diferente del valor REAL de la variable x . De otro modo, el instrumento no está sesgado.

$$\mu_i \neq x_{REAL}$$

2. *Error Aleatorio*: Es el error medido por medio de la varianza σ_i^2 en cada medición. Este puede ser eliminado tomando muchas mediciones con un dispositivo dado y promediando sus valores, pero esto no disminuirá el error sistemático.

Algunos de los puntos importantes que se toman en cuenta son los siguientes:

- Se debe tomar en cuenta el impacto de los errores aleatorios en cada medición. Por lo tanto, la confianza en un dispositivo se modela formalmente por la varianza de la distribución asociada. Si la varianza es baja, la confianza es mayor.
- Para formalizar el hecho de que los errores aleatorios de alguna medición no influyen en otras se dice que las variables x_i no están *correlacionadas*.
- *Variables Independientes*: Si las variables son independientes entonces No están correlacionadas, es decir su covarianza es 0. En otras palabras, conocer los valores de x_1 no aporta ninguna información sobre la variable x_2 . Es decir;

$$p(x_2|x_1) = p(x_2)$$

- *Variables No Correlacionadas*: La covarianza es 0, sin embargo no son necesariamente independientes puesto que conociendo x_1 podría darnos una idea de los posibles valores para x_2 , es decir se restringen algunas posibilidades sin cambiar el valor de la media μ_2 .

$$E[x_2|x_1] = E[x_2] \implies E[(x_1 - \mu_1)(x_2 - \mu_2)] = 0$$

- *Estimador no sesgado*: Es aquel cuya media es igual al valor desconocido de la variable que queremos estimar.
- **MSE**: Mean Square Error (error cuadrático medio), es la varianza de un estimador escalar no sesgado.

Por lo tanto, se establece el siguiente lema [23].

Lema 2.1

Sea $x_i \rightarrow p_i(\mu_i, \sigma_i^2)$ un conjunto de variables aleatorias no correlacionadas. Tenemos la combinación lineal;

$$y = \sum_{i=1}^n \alpha_i x_i \tag{2.3}$$

1. La media y su varianza son

$$\mu_y = \sum_{i=1}^n \alpha_i \mu_i \tag{2.4}$$

$$\sigma_y^2 = \sum_{i=1}^n \alpha_i^2 \sigma_i^2 \tag{2.5}$$

2. Si x_{n+1} no está correlacionada con x_1, \dots, x_n , entonces tampoco con y

2.3.2. Fusión de 2 estimaciones escalares

En este caso, el sistema a estudiar es unidimensional y únicamente se tienen 2 aparatos de medición.

Sabiendo que la suma de todos los coeficientes es igual a 1

$$\sum_{i=1}^n \alpha_i = 1$$

Al realizar la fusión de únicamente dos estimaciones escalares $n = 2$:

$$1 = \alpha' + \alpha \implies \alpha' = 1 - \alpha$$

$$y_\alpha(x_1, x_2) = (1 - \alpha)x_1 + \alpha x_2$$

La idea fundamental del filtro Kalman es encontrar los parámetros α y α' que minimicen la varianza de y . Por lo tanto, usando la expresión (2.5):

$$\frac{d}{d\alpha} \sigma_y^2 = 0 \implies \frac{d}{d\alpha} ((1 - \alpha)^2 \sigma_1^2 + \alpha^2 \sigma_2^2) = 0$$

$$\implies \alpha = \frac{\sigma_1^2}{\sigma_1^2 + \sigma_2^2} = K = \frac{\frac{1}{\sigma_2^2}}{\frac{1}{\sigma_1^2} + \frac{1}{\sigma_2^2}} = \frac{\nu_2}{\nu_1 + \nu_2}$$

K es el parámetro óptimo en la fusión de 2 estimaciones. Este valor es conocido como la ganancia Kalman.

La estimación y corregida está dada por la siguiente expresión:

$$y(x_1, x_2) = \frac{\nu_1}{\nu_1 + \nu_2} x_1 + \frac{\nu_2}{\nu_1 + \nu_2} x_2$$

La varianza de la distribución de probabilidad asociada a la estimación calculada por el filtro es la siguiente:

$$\sigma_y^2 = (1 - K)^2 \sigma_1^2 + K^2 \sigma_2^2 = \left(\frac{\frac{1}{\sigma_1^2}}{\frac{1}{\sigma_1^2} + \frac{1}{\sigma_2^2}} \right)^2 \sigma_1^2 + \left(\frac{\frac{1}{\sigma_2^2}}{\frac{1}{\sigma_1^2} + \frac{1}{\sigma_2^2}} \right)^2 \sigma_2^2$$

$$\sigma_y^2 = \frac{1}{\frac{1}{\sigma_1^2} + \frac{1}{\sigma_2^2}} \implies \frac{1}{\sigma_y^2} = \frac{1}{\sigma_1^2} + \frac{1}{\sigma_2^2}$$

$$\nu_y = \nu_1 + \nu_2$$

Las expresiones en términos de la ganancia Kalman resultan como:

$$\begin{aligned}
 K &= \frac{\sigma_1^2}{\sigma_1^2 + \sigma_2^2} \\
 y_\alpha(x_1, x_2) &= x_1 + K(x_2 - x_1) \\
 \sigma_y^2 &= (1 - K)\sigma_1^2
 \end{aligned}
 \tag{2.6}$$

2.3.3. Fusión de varias estimaciones escalares

Para el caso en que se tengan n aparatos, cada uno con su medición escalar x_i .

Teorema 3.2

Sea $x_i \rightarrow p_i(\mu_1, \sigma_i^2)$ un conjunto de variables aleatorias correlacionadas. Se considera la estimación:

$$y_{n,\alpha}(x_1, \dots, x_n) = \sum_{i=1}^n \alpha_i x_i, \quad \sum \alpha_i = 1$$

La varianza de la estimación se minimiza usando para cada coeficiente α_i :

$$\alpha_i = \frac{\frac{1}{\sigma_i^2}}{\sum \frac{1}{\alpha_j}}$$

Usando multiplicadores de Lagrange se calculan la varianza y la estimación óptima;

$$y_n(x_1, \dots, x_n) = \sum_{i=1}^n \frac{\nu_i}{\sum \nu_j} x_i \tag{2.7}$$

$$\nu_{y_n} = \sum_{i=1}^n \nu_i \tag{2.8}$$

2.3.4. Fusión de Incrementos

En muchas aplicaciones, los valores x_i corresponden a observaciones de una serie de tiempo que describe la evolución de un sistema. Es posible usar las ecuaciones 2.7 y 2.8 para fusionar esos valores y estimar el estado del sistema en cualquier momento, obteniendo así una predicción de la evolución futura de sus variables.

Por lo tanto, para el tiempo t_n

$$y_n(x_1, \dots, x_n) = \frac{\nu_{y_{n-1}}}{\nu_{y_{n-1}} + \nu_n} y_{n-1}(x_1, \dots, x_{n-1}) + \frac{\nu_n}{\nu_{y_{n-1}} + \nu_n} x_n \quad (2.9)$$

$$\nu_{y_n} = \nu_{y_{n-1}} + \nu_n \quad (2.10)$$

2.3.5. Estimaciones vectoriales

Al tener mediciones con más de una dimensión, se tienen expresiones que son escritas como sistemas de ecuaciones, por lo tanto, conviene desarrollar expresiones para sistemas de n dimensiones. Es por ello que se establecen las siguientes expresiones.

Esta es la *Matriz de Covarianza*;

$$\Sigma_{\mathbf{x}\mathbf{x}} = E[(\mathbf{x} - \boldsymbol{\mu}_{\mathbf{x}})(\mathbf{x} - \boldsymbol{\mu}_{\mathbf{x}})^T] \quad (2.11)$$

Siendo el primer término un vector columna, y el segundo un vector fila. Esta será una matriz de la correlación de la variable x_i y x_j .

Nota: Si las variables no están correlacionadas entonces las entradas (i, j) serán 0.

Lema 2.2

Sea $\mathbf{x}_i \rightarrow p_i(\boldsymbol{\mu}_i, \Sigma_i)$ un conjunto de variables vectoriales (con m entradas aleatorias) no correlacionadas.

Se tiene la combinación lineal;

$$\mathbf{y} = \sum_{i=1}^n A_i \mathbf{x}_i \quad (2.12)$$

Donde A_i es una matriz

1. La media y matriz de covarianza de \mathbf{y} son:

$$\boldsymbol{\mu}_{\mathbf{y}} = \sum_{i=1}^n A_i \boldsymbol{\mu}_i \quad (2.13)$$

$$\Sigma_{\mathbf{y}\mathbf{y}} = \sum_{i=1}^n A_i \Sigma_i A_i^T \quad (2.14)$$

2. Si \mathbf{x}_{n+1} no está correlacionada con $\mathbf{x}_1, \dots, \mathbf{x}_n$, entonces tampoco con \mathbf{y}

La *MSE* de un estimador no sesgado \mathbf{y} es la traza de su matriz de correlación.

$$MSE = E[(\mathbf{y} - \boldsymbol{\mu}_y)^T(\mathbf{y} - \boldsymbol{\mu}_y)] = Tr(\Sigma_{yy})$$

2.3.6. Fusión de múltiples estimaciones vectoriales

Para calcular la mejor estimación del estado de un sistema de muchas dimensiones usando las observaciones provistas por n dispositivos, la combinación lineal para la estimación es la siguiente:

$$\mathbf{y}_{n,A}(\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_n) = \sum_{i=1}^n A_i \mathbf{x}_i, \quad \sum_{i=1}^n A_i = I \quad (2.15)$$

Aquí A es un vector específico donde cada entrada es una matriz A_i . Para calcular la estimación, es fundamental hallar la colección de matrices $A = (A_1, A_2, \dots, A_n)$ que minimizan el error cuadrático medio de $\mathbf{y}_{n,A}$, es decir; $MSE(\mathbf{y}_A) = E[(\mathbf{y}_A - \boldsymbol{\mu}_{y_A})^T(\mathbf{y}_A - \boldsymbol{\mu}_{y_A})]$.

Teorema 4.1

Sea $\mathbf{x}_i \rightarrow p_i(\boldsymbol{\mu}_i, \Sigma_i)$ un conjunto de variables vectoriales aleatorias no correlacionadas. Considerando la estimación;

$$\mathbf{y}_{n,A}(\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_n) = \sum_{i=1}^n A_i \mathbf{x}_i, \quad \sum_{i=1}^n A_i = I \quad (2.16)$$

El valor del error cuadrático medio es minimizado por;

$$A_i = \left[\sum_{j=1}^n \Sigma_j^{-1} \right]^{-1} \Sigma_i^{-1} \quad (2.17)$$

Por lo tanto, la estimación óptima es

$$\mathbf{y}_{n,A}(\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_n) = \left[\sum_{j=1}^n \Sigma_j^{-1} \right]^{-1} \sum_{i=1}^n \Sigma_i^{-1} \mathbf{x}_i \quad (2.18)$$

La matriz de covarianza de \mathbf{y} puede calcularse usando la expresión siguiente:

$$\Sigma_{yy} = \left[\sum_{j=1}^n \Sigma_j^{-1} \right]^{-1} \quad (2.19)$$

La matriz inversa de la covarianza es denotada como la matriz N , por lo tanto la estimación está dada por:

$$\mathbf{y}(\mathbf{x}_1, \dots, \mathbf{x}_n) = N_y^{-1} \sum_{i=1}^n N_i \mathbf{x}_i \quad (2.20)$$

$$N_y = \sum_{j=1}^n N_j \quad (2.21)$$

2.3.7. Para 2 Estimaciones Vectoriales

El caso particular del uso de 2 estimaciones vectoriales es principalmente estudiado de forma que una de ellas es la *predicción* realizada por un modelo matemático y la otra es la *medición* realizada por algún dispositivo. Es por ello que en términos de la ganancia Kalman se obtienen las siguientes expresiones.

$$\begin{aligned} K &= \Sigma_1(\Sigma_1 + \Sigma_2)^{-1} \\ \mathbf{y}_\alpha(\mathbf{x}_1, \mathbf{x}_2) &= \mathbf{x}_1 + K(\mathbf{x}_2 - \mathbf{x}_1) \\ \Sigma_{yy} &= (1 - K)\Sigma_1 \end{aligned} \quad (2.22)$$

Estas son las expresiones finales para fusionar dos estimaciones vectoriales. Posteriormente, serán usadas para el presente trabajo.

2.3.8. Best Linear Unbiased Estimator

Del Teorema de Gauss-Markov se sabe que en determinadas condiciones, el estimador por mínimos cuadrados ordinarios (MCO) de los coeficientes de un modelo de regresión lineal es el mejor estimador lineal insesgado (BLUE) [20].

Partiendo de que el estado de un sistema representado por un vector \vec{y} no está dado de forma completa por las mediciones, entonces es necesario estimar la parte faltante. Imaginando que se tienen diferentes mediciones \vec{x} y \vec{y} , al graficar los puntos en un plano xy es notorio que existen muchos valores de y para cada x y muchos valores de x para cada y . La zona en la que se agrupan los puntos es llamada *Zona de Confianza*, mostrada en la figura (2.4).

Tomando un valor de \vec{x} , por ejemplo x_1 , se debe hallar el mejor valor posible de \vec{y} para asignárselo. Se calcula la media de todos los y_1 posibles en esa x_1 :

$$y_1 = E[y|x = x_1]$$

Esta sería la estimación para x_1 .

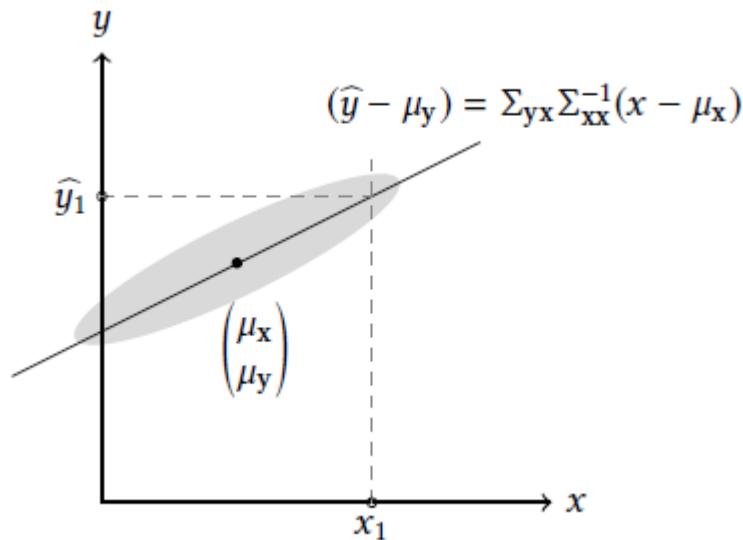


Figura 2.4: Gráfica de las observaciones x vs y [23].

En algunos problemas es posible asumir que existe una relación desconocida entre \mathbf{x} y \mathbf{y} y que la incertidumbre proviene del ruido.

Aunque hay muchos puntos (x_1, y) , los valores de y están encerrados alrededor de la línea mostrada en la figura. Entonces, el valor \hat{y}_1 es una estimación razonable para el valor de y correspondiente a x_1 . Esta línea se llama Mejor Estimación Lineal No Sesgada (BLUE).

¿Cómo calcular el BLUE?

Considerando el estimador lineal

$$\hat{\mathbf{y}}_{A,b} = A\mathbf{x} + \mathbf{b}$$

Se elige A y \mathbf{b} tales que minimizan el MSE.

Teorema 5.1

$$\text{Sea } \begin{pmatrix} \mathbf{x} \\ \mathbf{y} \end{pmatrix} \implies p \left(\begin{pmatrix} \boldsymbol{\mu}_x \\ \boldsymbol{\mu}_y \end{pmatrix}, \begin{pmatrix} \Sigma_{xx} & \Sigma_{xy} \\ \Sigma_{yx} & \Sigma_{yy} \end{pmatrix} \right)$$

El estimador $\hat{\mathbf{y}}_{A,b}$ para ser el óptimo debe considerar que:

$$\begin{aligned} \mathbf{b} &= \boldsymbol{\mu}_y - A\boldsymbol{\mu}_x \\ A &= \Sigma_{yx}\Sigma_{xx}^{-1} \end{aligned}$$

Para calcular \mathbf{b} solo se escribe $E[\hat{\mathbf{y}}] = E[\mathbf{y}]$, pues al tratarse de un estimador no sesgado, entonces el valor esperado es igual al valor real.

Se calcula A de forma que el error cuadrático medio sea mínimo, es decir, se calcula a través de hacer $\frac{d}{dA}MSE(\hat{\mathbf{y}}_A) = 0$

Por lo tanto, el mejor estimador lineal no sesgado es:

$$\boxed{\hat{\mathbf{y}} = \boldsymbol{\mu}_y + \Sigma_{yx}\Sigma_{xx}^{-1}(\mathbf{x} - \boldsymbol{\mu}_x)} \quad (2.23)$$

- No es necesario conocer alguna relación entre \mathbf{x} y \mathbf{y} . Solamente habrán de calcularse las medias y las matrices de correlación de la expresión 2.23.

2.4. Filtro Kalman para Sistemas lineales

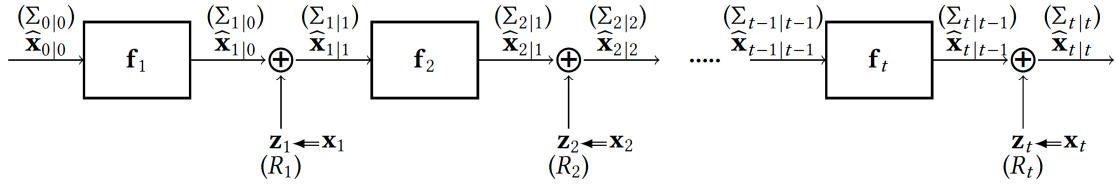


Figura 2.5: Estimación del estado usando Filtro Kalman [23].

- a) La Figura (2.6a) muestra cómo puede ser calculada la evolución del estado de cierto sistema en el tiempo con una condición inicial \mathbf{x}_0 , así, la dinámica ya está determinada.

El estado del sistema a cada tiempo es función del estado en el tiempo anterior y también función del control aplicado al sistema durante cierto intervalo de tiempo.

$$\mathbf{x}_t = f_t(\mathbf{x}_{t-1}, \mathbf{u}_t) = F_t \mathbf{x}_{t-1} + B_t \mathbf{u}_t$$

Aquí \mathbf{u} es el control y f_t generalmente no es lineal, por lo tanto puede cambiar en cada tiempo t . Así, es posible proponer un sistema lineal donde F_t y B_t sean matrices dependientes del tiempo que pueden ser determinadas por el sistema físico.

Si se conocen las condiciones iniciales \mathbf{x}_0 así como las matrices F_t y B_t (que en general no se conocen), se puede calcular de forma precisa la evolución del sistema.

- b) La Figura (2.6b) muestra el flujo del cálculo de los datos. Se asume que se tiene una estimación para el estado inicial $\hat{\mathbf{x}}_{0|0}$.
- *Estimación a priori*: es la predicción que hace uso de la información del tiempo anterior $t - 1$, denotada por $\hat{\mathbf{x}}_{t|t-1}$
 - *Estimación a posteriori*: es el resultado de fusionar dos componentes; la estimación a priori y la medición en el tiempo t ; \mathbf{z}_t . Este resultado se denota por $\hat{\mathbf{x}}_{t|t}$

Después, la estimación *a posteriori* se usa por el modelo para producir la estimación *a priori* para el siguiente paso en el tiempo $t + 1$.

2.4.1. Modelo de la Evolución del Estado y la predicción

La verdadera evolución del sistema está dada por la serie $\mathbf{x}_0, \mathbf{x}_1, \mathbf{x}_2, \dots$ a través de:

$$\mathbf{x}_t = F_t \mathbf{x}_{t-1} + B_t \mathbf{u}_t + \mathbf{w}_t \quad (2.24)$$

- \mathbf{u}_t : es la entrada de control, se asume como la entrada *determinista*.
- \mathbf{w}_t : es un termino de ruido con media cero el cual modela la incertidumbre del sistema. (los ruidos a diferentes tiempos no están correlacionados $E[\mathbf{w}_i, \mathbf{w}_j] = 0 \forall i \neq j$) ni tampoco con \mathbf{x}_0
- Q_t : es la matriz de covarianza de \mathbf{w}_t .
- La media y la matriz de covarianza de $\mathbf{x}_{i|j}$ se denotan por $\hat{\mathbf{x}}_{i|j}$ y $\Sigma_{i|j}$ respectivamente. Se asume que no existe sesgo, es decir; $E[\hat{\mathbf{x}}_{0|0}] = E[\mathbf{x}_0]$.

$$\hat{\mathbf{x}}_{t|t-1} = F_t \hat{\mathbf{x}}_{t-1|t-1} + B_t \mathbf{u}_t \quad (2.25)$$

$$\Sigma_{t|t-1} = F_t \Sigma_{t-1|t-1} F_t^T + Q_t \quad (2.26)$$

Estas expresiones predicen el valor de las variables del sistema en cada momento según cierto modelo matemático.

2.4.2. Fusión de Observaciones Completas

La medición es imprecisa y ruidosa, por lo tanto, se modela como:

$$\mathbf{z}_t = \mathbf{x}_t + \mathbf{v}_t \quad (2.27)$$

- \mathbf{v}_t : es ruido con media nula (no correlacionados entre sí $E[\mathbf{v}_i, \mathbf{v}_j] = 0 \forall i \neq j$), tampoco con \mathbf{x}_0 ni \mathbf{w}_k
- R_t : matriz de covarianza de \mathbf{v}_t
- \mathbf{x}_t en esta ecuación es el valor real del sistema al tiempo t .

Entonces, las estimaciones $\hat{\mathbf{x}}_{t|t-1}$ y \mathbf{z}_t son fusionadas usando las ecuaciones para 2 estimaciones vectoriales:

$$\begin{aligned} K_t &= \Sigma_{t|t-1} (\Sigma_{t|t-1} + R_t)^{-1} \\ \hat{\mathbf{x}}_{t|t}(\hat{\mathbf{x}}_{t|t-1}, \mathbf{z}_t) &= \hat{\mathbf{x}}_{t|t-1} + K_t (\mathbf{z}_t - \hat{\mathbf{x}}_{t|t-1}) \\ \Sigma_{t|t} &= (1 - K_t) \Sigma_{t|t-1} \end{aligned} \quad (2.28)$$

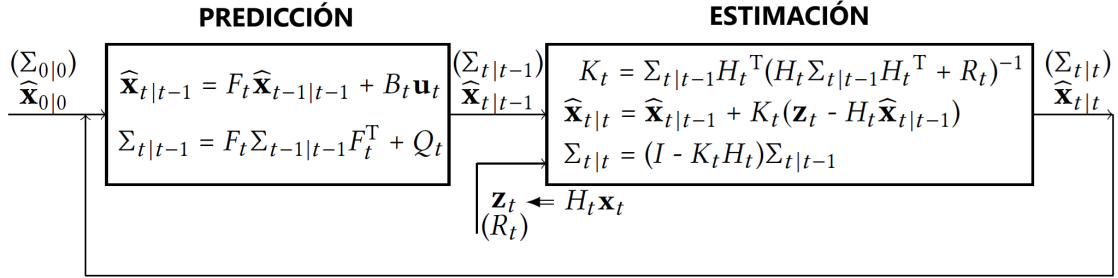


Figura 2.6: Estimación del estado usando Filtro Kalman [23].

2.4.3. Fusión de las Observaciones Parciales de un estado

En muchos problemas, solo es medible una parte del estado, esto hace que las variables no observadas deban ser estimadas de alguna manera.

- H_t : *matriz de observación*, en el estado es \mathbf{x} , lo que observamos es $H_t \mathbf{x}$. Ésta especifica la relación lineal entre el estado y la observación y también puede depender del tiempo.

$$\mathbf{z}_t = H_t \mathbf{x}_t + \mathbf{v}_t$$

- C_t : es la matriz que especifica la parte no observable.

Para fusionar los estados se siguen los siguientes pasos:

1. La estimación observable *a priori* $H_t \hat{\mathbf{x}}_{t|t-1}$ se relaciona con la medición \mathbf{z}_t usando la ecuación para la estimación BLUE (2.23) y las ecuaciones (2.22) para hallar el estado *a posteriori*.

$$\begin{aligned} H_t \hat{\mathbf{x}}_{t|t} &= H_t \hat{\mathbf{x}}_{t|t-1} + H_t \Sigma_{t|t-1} H_t^T (H_t \Sigma_{t|t-1} H_t^T + R_t)^{-1} (\mathbf{z}_t - H_t \mathbf{x}_{t|t-1}) \\ K_t &= \Sigma_{t|t-1} H_t^T (H_t \Sigma_{t|t-1} H_t^T + R_t)^{-1} \\ H_t \hat{\mathbf{x}}_{t|t} &= H_t \hat{\mathbf{x}}_{t|t-1} + H_t K_t (\mathbf{z}_t - H_t \mathbf{x}_{t|t-1}) \end{aligned} \tag{2.29}$$

2. La estimación *a priori* del estado no observable es $C_t \hat{\mathbf{x}}_{t|t-1}$.

La covarianza entre este y la parte observable $H_t \hat{\mathbf{x}}_{t|t-1}$ guiándonos del BLUE es $C_t \Sigma_{t|t-1} H_t^T$. La usamos para conocer la relación entre lo observable y lo no observable según las ecuaciones (2.23) de la siguiente manera.

$$\begin{aligned} C_t \hat{\mathbf{x}}_{t|t} &= C_t \hat{\mathbf{x}}_{t|t-1} + (C_t \Sigma_{t|t-1} H_t^T) (H_t \Sigma_{t|t-1} H_t^T)^{-1} H_t K_t (\mathbf{z}_t - H_t \mathbf{x}_{t|t-1}) \\ C_t \hat{\mathbf{x}}_{t|t} &= C_t \hat{\mathbf{x}}_{t|t-1} + C_t K_t (\mathbf{z}_t - H_t \mathbf{x}_{t|t-1}) \end{aligned} \quad (2.30)$$

3. Colocando finalmente ambas estimaciones en una sola ecuación;

$$\begin{pmatrix} H_t \\ C_t \end{pmatrix} \hat{\mathbf{x}}_{t|t} = \begin{pmatrix} H_t \\ C_t \end{pmatrix} \hat{\mathbf{x}}_{t|t-1} + \begin{pmatrix} H_t \\ C_t \end{pmatrix} K_t (\mathbf{z}_t - H_t \hat{\mathbf{x}}_{t|t-1}) \quad (2.31)$$

De aquí facilmente se encuentran las expresiones finales para la estimación *a posteriori*.

$$\boxed{\begin{aligned} \hat{\mathbf{x}}_{t|t} &= \hat{\mathbf{x}}_{t|t-1} + K_t (\mathbf{z}_t - H_t \hat{\mathbf{x}}_{t|t-1}) \\ \Sigma_{t|t} &= (I - K_t H_t) \Sigma_{t|t-1} \end{aligned}} \quad (2.32)$$

2.5. Filtro Kalman Extendido (EKF)

Cuando se tiene un sistema no lineal y se desea usar el filtro Kalman para estimar sus variables, se hace uso del Filtro Kalman Extendido.

Teniendo un sistema de ecuaciones diferenciales ordinarias no lineales y las mediciones con las que se desean realizar las estimaciones:

$$\begin{aligned}\frac{d\mathbf{x}}{dt} &= \mathbf{f}(\mathbf{x}) \\ \mathbf{z} &= \mathbf{h}(\mathbf{x}) + \mathbf{v}(t)\end{aligned}$$

Es necesario aproximar el sistema mediante algún método de resolución de sistemas de ecuaciones diferenciales ordinarias, por ejemplo: método de diferencias finitas de Euler.

$$\mathbf{x}_k \approx \mathbf{x}_{k-1} + \mathbf{f}(\mathbf{x}_{k-1}) * \Delta t$$

Aquí Δt es el tamaño del paso entre cada punto de la aproximación y se identifica:

$$\mathbf{F}(\mathbf{x}_{k-1}) = \mathbf{x}_{k-1} + \mathbf{f}(\mathbf{x}_{k-1}) * \Delta t \tag{2.33}$$

Es posible expandir $\mathbf{F}(\mathbf{x}_{k-1})$ y $\mathbf{h}(\mathbf{x}_k)$ en serie de Taylor alrededor de la estimación anterior y así aproximar la predicción y la próxima estimación de las variables \mathbf{x}_k .

Usando el modelo y las observaciones (las mediciones) se establecen las siguientes expresiones:

$$\begin{aligned}\mathbf{x}_k &= \mathbf{F}(\mathbf{x}_{k-1}) + \mathbf{w}_{k-1} \\ \mathbf{z}_k &= \mathbf{h}(\mathbf{x}_k) + \mathbf{v}_k\end{aligned}$$

La inicialización correspondería a:

$$\mathbf{x}_0^e = \mu_0$$

Se calcula la predicción del estado del sistema:

$$\begin{aligned}\mathbf{x}_k^p &\approx \mathbf{F}(\mathbf{x}_{k-1}^e) \\ \mathbf{S}_k^p &= \mathbf{J}_p(\mathbf{x}_{k-1}^e) \mathbf{S}_{k-1} \mathbf{J}_p^T(\mathbf{x}_{k-1}^e) + \mathbf{Q}_{k-1}\end{aligned}$$

Se calcula la estimación del sistema usando las mediciones:

$$\mathbf{K}_k = \mathbf{S}_k^e \mathbf{J}_h^T(\mathbf{x}_k^p) \left(\mathbf{J}_h(\mathbf{x}_k^p) \mathbf{S}_k^p \mathbf{J}_h^T(\mathbf{x}_k^p) + \mathbf{R}_k \right)^{-1}$$

$$\mathbf{x}_k^e \approx \mathbf{x}_k^p + \mathbf{K}_k (\mathbf{z}_k - \mathbf{h}(\mathbf{x}_k^p))$$

$$\mathbf{S}_k = (\mathbf{I} - \mathbf{K}_k \mathbf{J}_h(\mathbf{x}_k^p)) \mathbf{S}_k^p$$

dónde

$$\mathbf{J}_f = \begin{pmatrix} \frac{\partial f_1}{\partial x_1} & \frac{\partial f_1}{\partial x_2} & \cdots & \frac{\partial f_1}{\partial x_n} \\ \vdots & & \ddots & \vdots \\ \frac{\partial f_n}{\partial x_1} & \frac{\partial f_n}{\partial x_2} & \cdots & \frac{\partial f_n}{\partial x_n} \end{pmatrix} \quad \text{and} \quad \mathbf{J}_h = \begin{pmatrix} \frac{\partial h_1}{\partial x_1} & \frac{\partial h_1}{\partial x_2} & \cdots & \frac{\partial h_1}{\partial x_n} \\ \vdots & & \ddots & \vdots \\ \frac{\partial h_n}{\partial x_1} & \frac{\partial h_n}{\partial x_2} & \cdots & \frac{\partial h_n}{\partial x_n} \end{pmatrix}$$

Al realizar este procedimiento de forma iterada, se obtiene la estimación de todas las variables del sistema a cualquier tiempo t [24]

2.6. Filtro de Savitzky-Golay

El filtro Savitzky-Golay es ampliamente usado para el preprocesamiento de señales en múltiples áreas. Este es usado para reducir ruido de alta frecuencia debido a que *suaviza* la señal ya que la reduce a una con ruido de baja frecuencia [21].

Savitzky y Golay propusieron un método para suavizar una serie de datos basándose en una aproximación polinómica por mínimos cuadrados de forma local. Este filtro es atractivo por preservar la forma de los picos y ha tenido grandes aplicaciones como el procesamiento de electrocardiogramas; además, se ha aplicado al procesamiento de imágenes como las ecografías y al radar de apertura sintética [16].

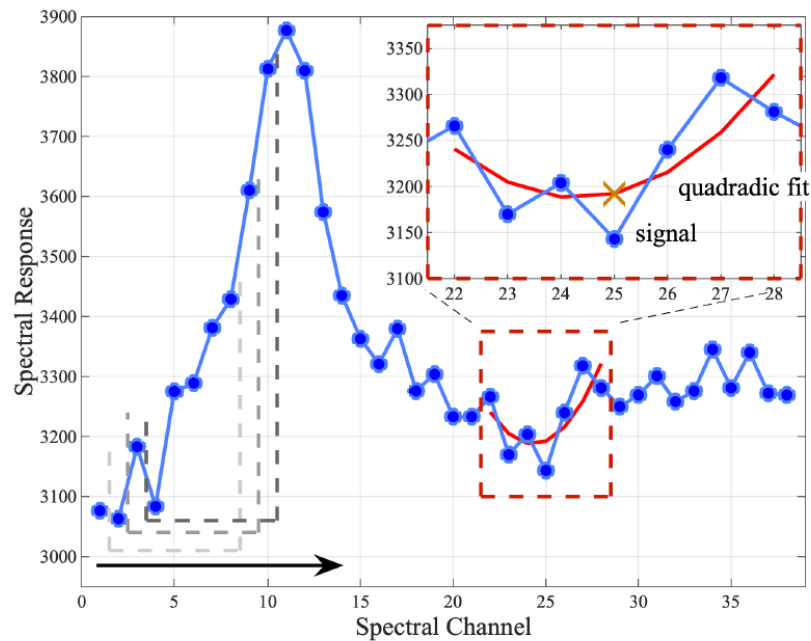


Figura 2.7: Señal discreta a la cual se le aplica un filtro S-G con una ventana $w = 7$ y ajuste cuadrático [21].

2.6.1. Implementación

Dada una señal con N puntos y una ventana de tamaño w , el filtro S-G ajusta un polinomio de orden O en cada ventana de puntos dada y el filtro se mueve a lo largo de toda la señal.

Este filtro estima el valor de la señal en el centro de la ventana evaluando el polinomio calculado en ese punto.

Un requisito importante es que el tamaño de la ventana tiene que ser mayor al del orden del polinomio que se ajustará [21].

$$O + 1 < w$$

Este filtro es sencillo de implementar en la actualidad mediante diferentes lenguajes de programación y *software*. Al aplicar este filtro a una señal, se debe proponer el tamaño de la ventana de puntos w y el orden del polinomio ajustado.

Capítulo 3

El Represilador y sus mediciones sintéticas

3.1. El represilador mediante diferencias finitas de Euler

Existen muchos métodos numéricos para resolver sistemas de ecuaciones diferenciales ordinarias no lineales; sin embargo, como primer acercamiento de este método para encontrar los parámetros mediante el uso indirecto del filtro Kalman, basta como ejercicio inicial con un método simple explícito para resolver el sistema del represilador.

Recordando las ecuaciones del represilador (2.2) :

$$\left\{ \begin{array}{l} \frac{dm_1}{dt} = -m_1 + \frac{\alpha}{(1+p_3^n)} + \alpha_0 \\ \frac{dm_2}{dt} = -m_2 + \frac{\alpha}{(1+p_1^n)} + \alpha_0 \\ \frac{dm_3}{dt} = -m_3 + \frac{\alpha}{(1+p_2^n)} + \alpha_0 \\ \frac{dp_1}{dt} = -\beta(p_1 - m_1) \\ \frac{dp_2}{dt} = -\beta(p_2 - m_2) \\ \frac{dp_3}{dt} = -\beta(p_3 - m_3) \end{array} \right. \quad (3.1)$$

En el método explícito de diferencias finitas de Euler se hace una aproximación de las ecuaciones diferenciales como:

$$\frac{dy}{dt} \approx \frac{y_{n+1} - y_n}{\Delta t} \implies \frac{y_{n+1} - y_n}{\Delta t} = f(y_n, t_n)$$

El Represilador y sus mediciones sintéticas
3.1 El represilador mediante diferencias finitas de Euler

Por lo tanto, las ecuaciones del represilador quedan de la siguiente manera:

$$\left\{ \begin{array}{l} \frac{m_{1,i} - m_{1,i-1}}{\Delta t} = -m_{1,i-1} + \alpha_0 + \frac{\alpha}{1 + p_{3,k-1}^n} \\ \frac{m_{2,i} - m_{2,i-1}}{\Delta t} = -m_{2,i-1} + \alpha_0 + \frac{\alpha}{1 + p_{1,k-1}^n} \\ \frac{m_{3,i} - m_{3,i-1}}{\Delta t} = -m_{3,i-1} + \alpha_0 + \frac{\alpha}{1 + p_{2,k-1}^n} \\ \frac{p_{1,i} - p_{1,i-1}}{\Delta t} = \beta(m_{1,i-1} - p_{1,i-1}) \\ \frac{p_{2,i} - p_{2,i-1}}{\Delta t} = \beta(m_{2,i-1} - p_{2,i-1}) \\ \frac{p_{3,i} - p_{3,i-1}}{\Delta t} = \beta(m_{3,i-1} - p_{3,i-1}) \end{array} \right. \quad (3.2)$$

Resolviendo el sistema usando los parámetros [1]:

- $n = 2$
- $\alpha_0 = 0,5$
- $\alpha = 500$
- $\beta = 5$

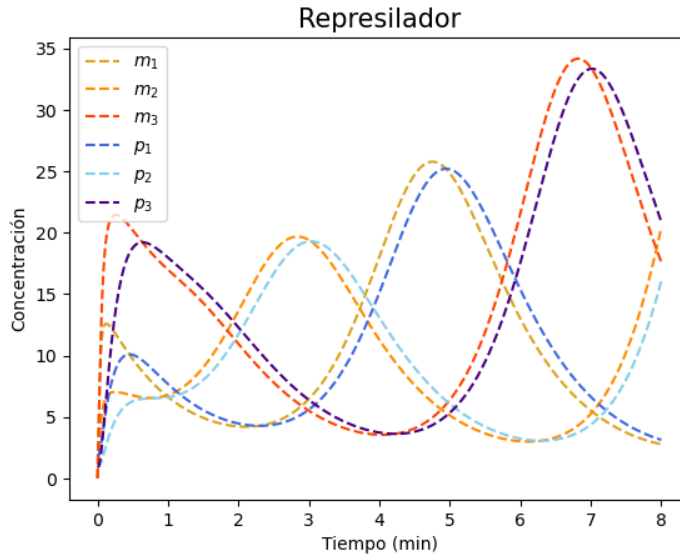


Figura 3.1: Gráficas de las moléculas de ARNm y de sus proteínas represoras durante los primeros 10 minutos de la dinámica con condiciones iniciales $m_1 = 0$, $m_2 = 0$, $m_3 = 0$, $p_1 = 2$, $p_2 = 1$ y $p_3 = 1$. Para esta simulación se usaron 1000 puntos.

3.2. Simulación de datos para el filtro Kalman

Se generaron datos sintéticos para evaluar el desempeño del filtro de Kalman. Para ello, se incorporó un término de ruido gaussiano aditivo con media cero y desviación estándar de 1.5. Dado que en muchos sistemas biológicos no todas las variables son directamente observables, el ruido se añadió únicamente a las concentraciones de proteínas supresoras. Estas proteínas, en el contexto experimental, suelen ser inferidas mediante técnicas de fluorescencia que emplean genes reporteros acoplados a promotores específicos [1].

Considerando la disponibilidad limitada de datos en escenarios experimentales reales, se simularon únicamente 100 observaciones. Estos datos corresponden a 100 puntos equidistantes extraídos de la solución numérica del sistema dinámico, obtenida mediante el método de diferencias finitas, y a cada uno se le añadió el término de ruido gaussiano previamente descrito.

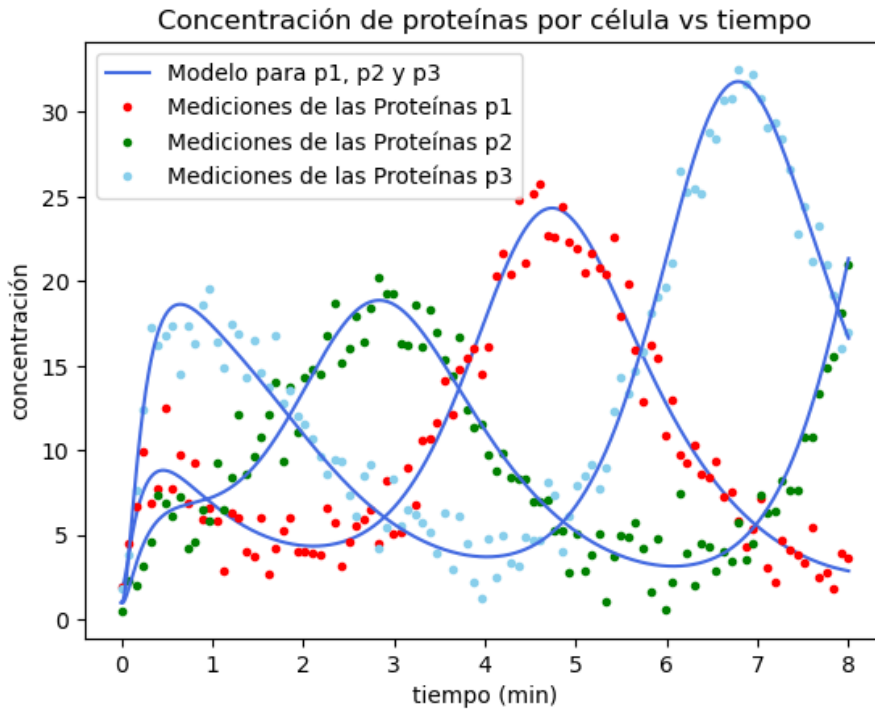


Figura 3.2: Modelo y Mediciones con ruido gaussiano ($\mu = 0$ y $\sigma = 1,5$)

Es importante aclarar que en el presente trabajo se considera que los datos experimentales en el estudio de estructuras biológicas, como lo son las redes de regulación genética, son bastante limitados y difíciles de obtener.

Capítulo 4

Estimación del represilador mediante Filtro Kalman Extendido

La estimación de las variables de la mayoría de los sistemas no se realiza en tiempo real, por lo tanto, tiene sentido usar todos los puntos de la estimación para realizar una mejor estimación usando los puntos *futuros*. Esto se logra introduciendo un filtro que suavice un intervalo fijo después de usar el filtro Kalman [17].

Con el objetivo de establecer el mejor resultado al encontrar los parámetros, se realizó la comparación entre el uso del filtro S-G antes o después del filtro Kalman. El filtro Kalman ha sido usado antes como estimador principal y después se aplica el refinamiento a través de S-G ya que elimina parte del ruido que Kalman no puede eliminar [25].

Métodos propuestos:

- **Estimación de las variables suavizando mediante el filtro Savitzky-Golay después del filtro Kalman:** Las mediciones se filtran mediante Kalman y se refina la estimación usando el filtro Savitzky-Golay.
- **Estimación de las variables suavizando mediante el filtro Savitzky-Golay antes del filtro Kalman:** Las mediciones se procesan primero por Savitzky-Golay y finalmente se estiman las variables mediante el filtro Kalman.

Es sabido que el filtro Kalman es deficiente si se tiene un número de datos muy limitado [17], y en el presente trabajo también se pretende evaluar si, aun en estas condiciones, es posible encontrar los parámetros que reproduzcan esa estimación; por esa razón, es fundamental usar la mejor estimación posible.

Se evaluarán ambos métodos calculando el MSE con respecto al modelo; así, posteriormente se decidirá qué estimación usar para calcular los parámetros. Para la realización de ambos, se deben obtener como primer paso las ecuaciones para la estimación de las variables del represilador mediante el filtro Kalman.

4.1. Ecuaciones para la estimación de las variables del represilador mediante el filtro Kalman Extendido

Para usar el Filtro Kalman Extendido en el represilador es necesario resolverlo.

Retomando el sistema discretizado 3.2 del represilador a partir del método de diferencias finitas de Euler:

$$\left\{ \begin{array}{l} m_{1,i} = m_{1,i-1} + \left(-m_{1,i-1} + \alpha_0 + \frac{\alpha}{1 + p_{3,k-1}^n} \right) \Delta t \\ m_{2,i} = m_{2,i-1} + \left(-m_{2,i-1} + \alpha_0 + \frac{\alpha}{1 + p_{1,k-1}^n} \right) \Delta t \\ m_{3,i} = m_{3,i-1} + \left(-m_{3,i-1} + \alpha_0 + \frac{\alpha}{1 + p_{2,k-1}^n} \right) \Delta t \\ p_{1,i} = p_{1,i-1} + \beta(m_{1,i-1} - p_{1,i-1})\Delta t \\ p_{2,i} = p_{2,i-1} + \beta(m_{2,i-1} - p_{2,i-1})\Delta t \\ p_{3,i} = p_{3,i-1} + \beta(m_{3,i-1} - p_{3,i-1})\Delta t \end{array} \right. \quad (4.1)$$

Como se comentó en la sección 2.5 y reacomodando algunos términos, el sistema se escribe de la siguiente forma:

$$\mathbf{x}_k \approx \mathbf{F}(\mathbf{x}_{k-1}) = \mathbf{x}_{k-1} + \mathbf{f}(\mathbf{x}_{k-1}) * \Delta t$$

$$\mathbf{F}(\mathbf{x}) = \begin{pmatrix} F_1(m_1, p_3) \\ F_2(m_2, p_1) \\ F_3(m_3, p_2) \\ F_4(m_1, p_1) \\ F_5(m_2, p_2) \\ F_6(m_3, p_3) \end{pmatrix} = \begin{pmatrix} m_1(1 - \Delta t) + \left(\alpha_0 + \frac{\alpha}{1 + p_3^n} \right) \Delta t \\ m_2(1 - \Delta t) + \left(\alpha_0 + \frac{\alpha}{1 + p_1^n} \right) \Delta t \\ m_3(1 - \Delta t) + \left(\alpha_0 + \frac{\alpha}{1 + p_2^n} \right) \Delta t \\ \beta(m_1 - p_1)\Delta t + p_1 \\ \beta(m_2 - p_2)\Delta t + p_2 \\ \beta(m_3 - p_3)\Delta t + p_3 \end{pmatrix}$$

Se expande en Serie de Taylor alrededor de la estimación \mathbf{x}_{k-1}^e para linealizar el sistema:

$$\mathbf{F}(\mathbf{x}_{k-1}) = \mathbf{F}(\mathbf{x}_{k-1}^e) + \mathbf{J}_{\mathbf{F}}(\mathbf{x}_{k-1}^e)(\mathbf{x}_{k-1} - \mathbf{x}_{k-1}^e) + \dots \quad (4.2)$$

Se calcula la matriz jacobiana de $\mathbf{F}(\mathbf{x}_{k-1})$.

$$\mathbf{J}_{\mathbf{F}} = \begin{pmatrix} 1 - \Delta t & 0 & 0 & 0 & 0 & -\frac{na_1 p_3^n \Delta t}{p_3(p_3^n + 1)^2} \\ 0 & 1 - \Delta t & 0 & -\frac{na_1 p_1^n \Delta t}{p_1(p_1^n + 1)^2} & 0 & 0 \\ 0 & 0 & 1 - \Delta t & 0 & -\frac{na_1 p_2^n \Delta t}{p_2(p_2^n + 1)^2} & 0 \\ \beta \Delta t & 0 & 0 & -\beta \Delta t + 1 & 0 & 0 \\ 0 & \beta \Delta t & 0 & 0 & -\beta \Delta t + 1 & 0 \\ 0 & 0 & \beta \Delta t & 0 & 0 & -\beta \Delta t + 1 \end{pmatrix} \quad (4.3)$$

Al aplicar el filtro Kalman se inicializa con $\mathbf{x}_0^e = \mu_0$, es decir, los valores en los que se espera que las variables se encuentren al inicio. Y la matriz de varianza \mathbf{S}_0 es una matriz diagonal en la cual las entradas s_{ii} toman valores asociados a la varianza de la condición inicial con el estado real del sistema (estos valores son ajustados por el usuario).

1. **Predicción:** Dada la expresión (4.2) se calcula la predicción del estado del sistema (también llamada estimación *a priori*) mediante el modelo matemático del represilador. La predicción y la estimación al tiempo k serán denotadas respectivamente por \mathbf{x}_k^p y \mathbf{x}_k^e .

$$\begin{aligned} \mathbf{x}_k^p &\approx \mathbf{F}(\mathbf{x}_{k-1}^e) \\ \mathbf{S}_k^p &= \mathbf{J}_{\mathbf{F}}(\mathbf{x}_{k-1}^e) \mathbf{S}_{k-1} \mathbf{J}_{\mathbf{F}}^T(\mathbf{x}_{k-1}^e) + \mathbf{Q}_{k-1} \end{aligned} \quad (4.4)$$

Dónde \mathbf{Q}_{k-1} es una matriz diagonal en la que cada elemento q_{ii} tiene el valor del ruido para cada variable asociado a la *confianza* que tenemos con el modelo (estos valores son ajustados por el usuario). Mientras que $\mathbf{J}_{\mathbf{F}}(\mathbf{x}_{k-1}^e)$ está dado por la expresión (4.3) y $\mathbf{F}(\mathbf{x}_{k-1}^e)$ está dada por:

$$\mathbf{F}(\mathbf{x}_{k-1}^e) = \begin{pmatrix} m_{1,k-1}^e (1 - \Delta t) + \left(\alpha_0 + \frac{\alpha}{1 + (p_{3,k-1}^e)^n} \right) \Delta t \\ m_{2,k-1}^e (1 - \Delta t) + \left(\alpha_0 + \frac{\alpha}{1 + (p_{1,k-1}^e)^n} \right) \Delta t \\ m_{3,k-1}^e (1 - \Delta t) + \left(\alpha_0 + \frac{\alpha}{1 + (p_{2,k-1}^e)^n} \right) \Delta t \\ \beta(m_{1,k-1}^e - p_{1,k-1}^e) \Delta t + p_{1,k-1}^e \\ \beta(m_{2,k-1}^e - p_{2,k-1}^e) \Delta t + p_{2,k-1}^e \\ \beta(m_{3,k-1}^e - p_{3,k-1}^e) \Delta t + p_{3,k-1}^e \end{pmatrix}$$

Estimación del represilador mediante Filtro Kalman Extendido

4.1 Ecuaciones para la estimación de las variables del represilador mediante el filtro Kalman Extendido

2. **Estimación:** Se calcula la ganancia Kalman usando el jacobiano de la matriz de observación que en este caso particular en el que se observan únicamente las proteínas traducidas tienen la forma siguiente:

$$\mathbf{h}(\mathbf{x}_k^p) = \begin{pmatrix} p_{1,k}^p \\ p_{2,k}^p \\ p_{3,k}^p \end{pmatrix}, \quad \mathbf{J}_h(\mathbf{x}_k^p) = \begin{pmatrix} 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 \end{pmatrix} \quad (4.5)$$

La matriz de ruido \mathbf{R}_k es una matriz diagonal en la cual sus entradas r_{ii} contienen valores relacionados al ruido de las mediciones (ajustados por el usuario) y \mathbf{z}_k es un vector columna en el que sus entradas contienen las mediciones de las proteínas.

$$\begin{aligned} \mathbf{K}_k &= \mathbf{S}_k^p \mathbf{J}_h^T(\mathbf{x}_k^p) \left(\mathbf{J}_h(\mathbf{x}_k^p) \mathbf{S}_k^p \mathbf{J}_h^T(\mathbf{x}_k^p) + \mathbf{R}_k \right)^{-1} \\ \mathbf{x}_k^e &\approx \mathbf{x}_k^p + \mathbf{K}_k (\mathbf{z}_k - \mathbf{h}(\mathbf{x}_k^p)) \\ \mathbf{S}_k &= (\mathbf{I} - \mathbf{K}_k \mathbf{J}_h(\mathbf{x}_k^p)) \mathbf{S}_k^p \end{aligned} \quad (4.6)$$

4.2. Estimación de las variables suavizando mediante el filtro Savitzky-Golay después del filtro Kalman

Se realizó el código en Python para encontrar la estimación mediante el Filtro Kalman usando los datos generados.

4.2.1. Estimación por filtro Kalman

Al implementar el filtro se deben colocar condiciones iniciales de los parámetros que estamos buscando, es por eso que se introdujeron valores aleatorios dentro del régimen de valores establecidos por el represilador para presentar oscilaciones [1];

$$\alpha_0/\alpha \approx 0,001 \quad \beta \approx 5 \quad n = 2 \quad (4.7)$$

Es importante mencionar que el resultado de la estimación realizada mediante el filtro Kalman depende en gran medida de los valores iniciales que se introduzcan, así como del número de datos (mediciones) que se tengan.

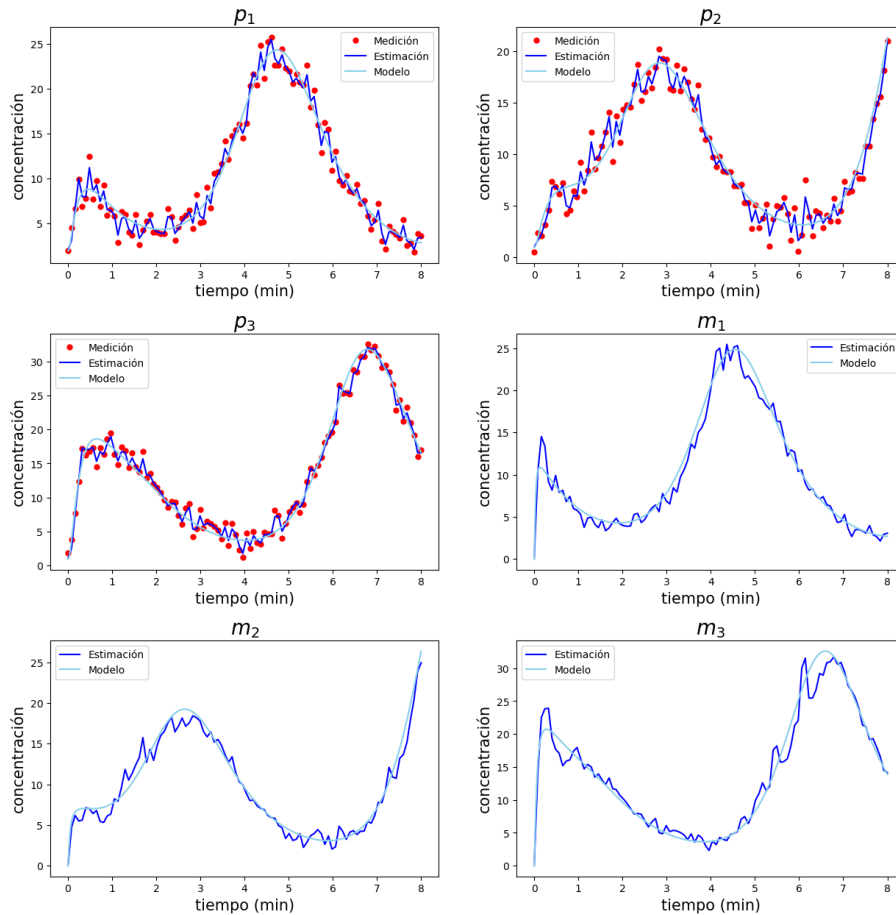


Figura 4.1: Gráficas de la Estimación realizada mediante el Filtro Kalman

Estimación del represilador mediante Filtro Kalman Extendido

4.2 Estimación de las variables suavizando mediante el filtro Savitzky-Golay después del filtro Kalman

Al obtener la medición, se calculó el error cuadrático medio de la estimación y las mediciones con el modelo para cada proteína.

Variable	MSE de las mediciones	MSE de la estimación Kalman
p_1	2.088	1.082
p_2	2.141	1.249
p_3	2.013	1.327

Tabla 4.1: Comparación de el error cuadrático medio entre las estimaciones y mediciones con el modelo del represilador.

Si las mediciones son ruidosas y se tiene la certeza de que el modelo describe correctamente la dinámica del fenómeno, entonces la estimación del filtro Kalman es mejor aproximación que las mediciones tomadas.

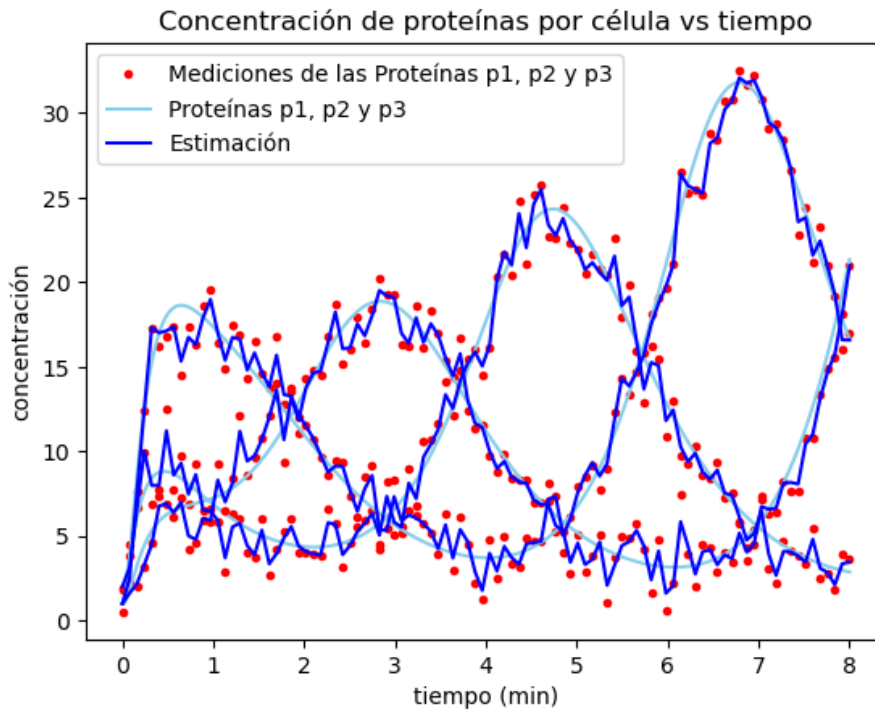


Figura 4.2: Estimación de las concentraciones de las proteínas del modelo. Para este filtro se usaron como parámetros iniciales $\alpha_0 = 0,7$, en consecuencia con (4.7) $\alpha_i \approx 700$ y $\beta = 7$.

Estimación del represilador mediante Filtro Kalman Extendido

4.2 Estimación de las variables suavizando mediante el filtro Savitzky-Golay después del filtro Kalman

4.2.2. Suavizado de la curva mediante filtro Savitzky-Golay

Aunque el filtro de Kalman proporciona una estimación óptima en presencia de ruido gaussiano bajo un modelo dinámico conocido, en escenarios con observaciones altamente ruidosas o con modelos parcialmente conocidos, las estimaciones pueden presentar oscilaciones de alta frecuencia. Para mitigar este efecto y obtener curvas más suaves que preserven las tendencias locales, se aplicó un filtro de Savitzky-Golay sobre las estimaciones. Este método de suavizado, basado en el ajuste local de polinomios mediante mínimos cuadrados, es particularmente útil para preservar la forma de la estimación.

La ventana móvil usada equivalente al 3% de los puntos de la estimación (en este caso particular, 3 puntos). La elección de la ventana es ajustada por el usuario y el mejor resultado fue logrado usando esta cantidad de puntos (comparando sus MSE), mientras que al ajustar un modelo lineal se evita el sobreajuste de los datos.

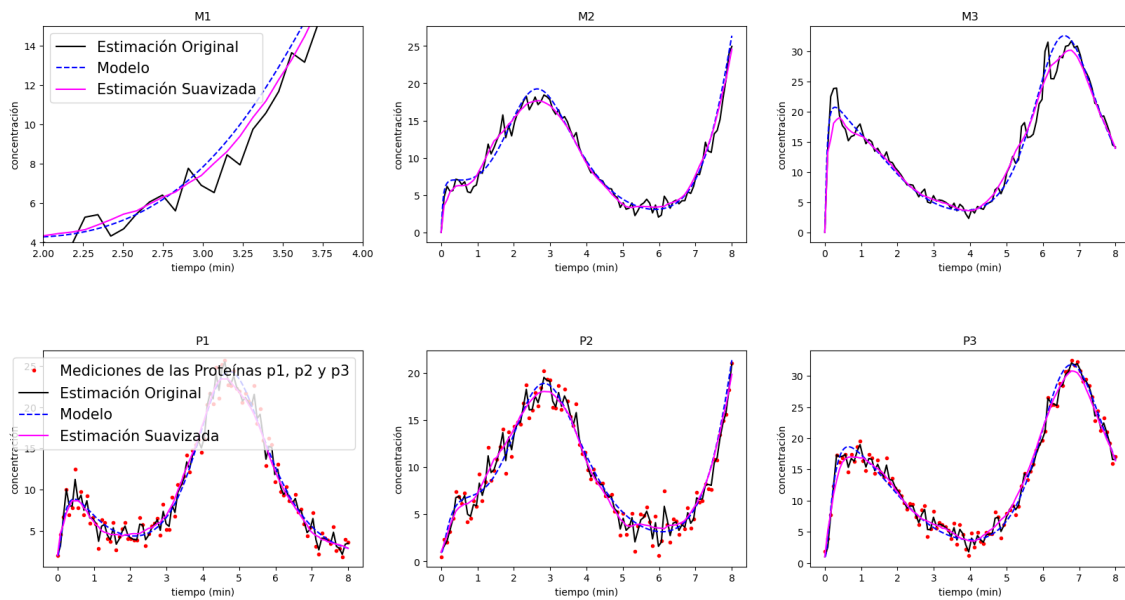


Figura 4.3: Gráficas del modelo con los parámetros reales, la estimación por kalman y la estimación filtrada por Savitzky-Golay.

Se calculó el error cuadrático medio de la estimación por kalman y de la estimación filtrada mediante Savitzky-Golay.

Variable	MSE de la estimación Kalman	MSE de la estimación suavizada
p_1	1.082	0.363
p_2	1.249	0.442
p_3	1.327	0.759

4.3. Estimación de las variables suavizando mediante el filtro Savitzky-Golay antes del filtro Kalman

4.3.1. Suavizado de las mediciones usando el filtro S-G

Como se estableció al principio del capítulo, ahora como segundo método, primero se suavizan las mediciones, notemos que en este caso, a diferencia del anterior, sólo podremos aplicar el suavizado a las proteínas p_1 , p_2 y p_3 pues de las moléculas de ARNm no tenemos mediciones. Al comparar el MSE de este resultado con el de la sección 4.2.2 se definirá a la estimación que se usará para encontrar los parámetros con el menor ruido posible.

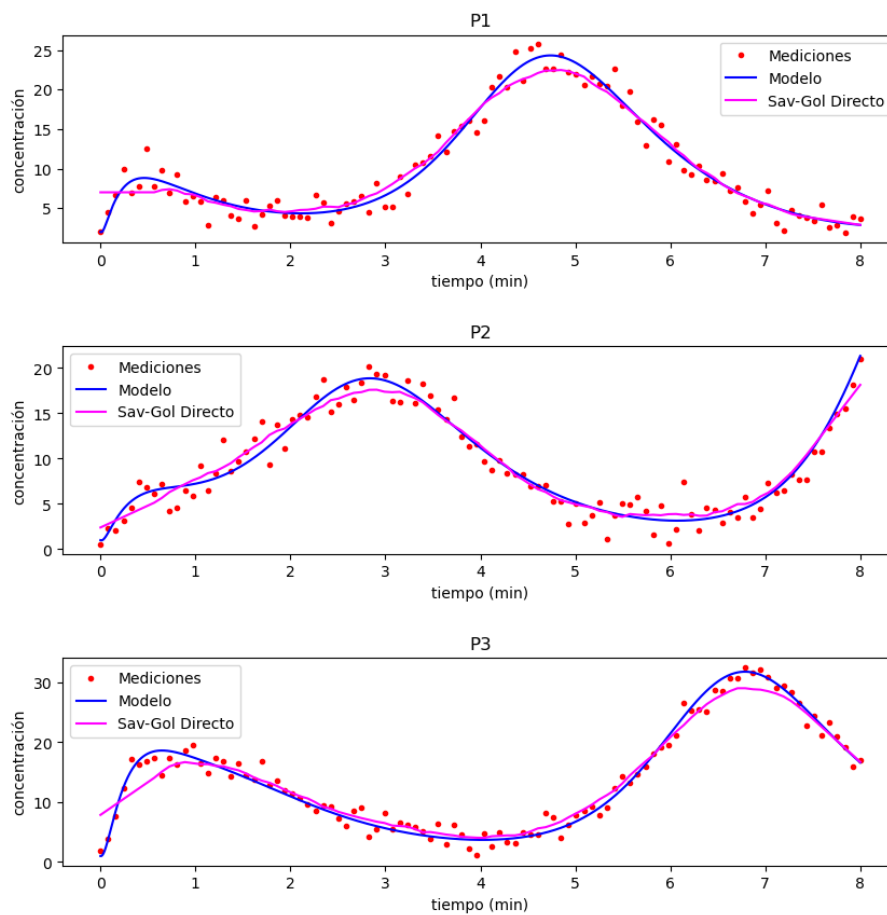


Figura 4.4: Gráficas de las mediciones de las proteínas, el modelo y las mediciones suavizadas.

4.3.2. Estimación por filtro Kalman usando mediciones suavizadas

Usando las mediciones suavizadas, se implementó el filtro Kalman.

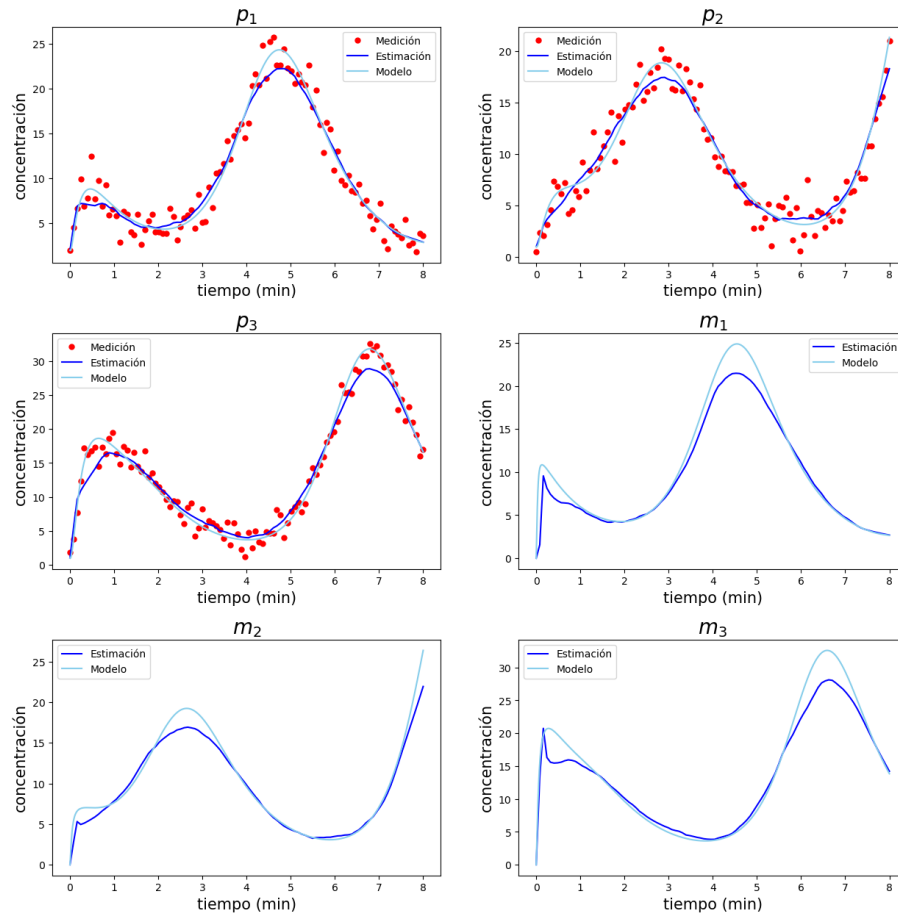


Figura 4.5: Gráficas de las estimaciones por Kalman usando mediciones suavizadas procesadas por S-G.

Se calcularon los errores cuadráticos medios de los datos suavizados por Savitzky-Golay y de la estimación por Kalman usando los datos suavizados.

Variable	MSE de los datos suavizados	MSE de la estimación con datos suavizados.
p_1	0.973	0.661
p_2	0.66	0.613
p_3	2.933	2.274

De estos resultados se concluye que la estimación que se usará para el cálculo de los parámetros corresponde a la realizada por Kalman y posteriormente suavizada por S-G, ya que el MSE fue menor.

Capítulo 5

Estimación de los parámetros del represilador

Al haber resuelto el sistema mediante diferencias finitas es posible escribir las expresiones de la siguiente manera;

$$\left\{ \begin{array}{l} m_{1,i} = m_{1,i-1}(1 - \Delta t) + \left(\alpha_0 + \frac{\alpha}{1 + p_{3,i-1}^n} \right) \Delta t \\ m_{2,i} = m_{2,i-1}(1 - \Delta t) + \left(\alpha_0 + \frac{\alpha}{1 + p_{1,i-1}^n} \right) \Delta t \\ m_{3,i} = m_{3,i-1}(1 - \Delta t) + \left(\alpha_0 + \frac{\alpha}{1 + p_{2,i-1}^n} \right) \Delta t \\ p_{1,i} = \beta(m_{1,i-1} - p_{1,i-1})\Delta t + p_{1,i-1} \\ p_{2,i} = \beta(m_{2,i-1} - p_{2,i-1})\Delta t + p_{2,i-1} \\ p_{3,i} = \beta(m_{3,i-1} - p_{3,i-1})\Delta t + p_{3,i-1} \end{array} \right. \quad (5.1)$$

Se desea calcular el valor de los parámetros variables en el tiempo $\alpha_0(t)$, $\alpha(t)$ y $\beta(t)$ de forma que el modelo reproduzca la estimación realizada por el filtro Kalman al sustituir estos parámetros.

$$\left\{ \begin{array}{l} m'_{1,i+1} = m_{1,i}(1 - \Delta t) + \left(\alpha'_{0,i+1} + \frac{\alpha'_{i+1}}{1 + p'_{3,i}} \right) \Delta t \\ m'_{2,i+1} = m_{2,i}(1 - \Delta t) + \left(\alpha'_{0,i+1} + \frac{\alpha'_{i+1}}{1 + p'_{1,i}} \right) \Delta t \\ m'_{3,i+1} = m_{3,i}(1 - \Delta t) + \left(\alpha'_{0,i+1} + \frac{\alpha'_{i+1}}{1 + p'_{2,i}} \right) \Delta t \\ p'_{1,i+1} = \beta'_{i+1}(m_{1,i} - p_{1,i})\Delta t + p_{1,i} \\ p'_{2,i+1} = \beta'_{i+1}(m_{2,i} - p_{2,i})\Delta t + p_{2,i} \\ p'_{3,i+1} = \beta'_{i+1}(m_{3,i} - p_{3,i})\Delta t + p_{3,i} \end{array} \right. \quad (5.2)$$

En las expresiones (5.2) las variables primadas ($m'_1, m'_2, m'_3, p'_1, p'_2$ y p'_3) toman el valor de la estimación $i + 1$ hallada por el filtro Kalman. Sin embargo, los parámetros primados θ'_i representan el valor que debemos hallar en cada tiempo i . Es importante aclarar que las variables no primadas son las variables del sistema i calculadas a partir del sistema (5.1) usando los parámetros θ'_i que se irán hallando recursivamente como se muestra en el diagrama de la figura (5.1).

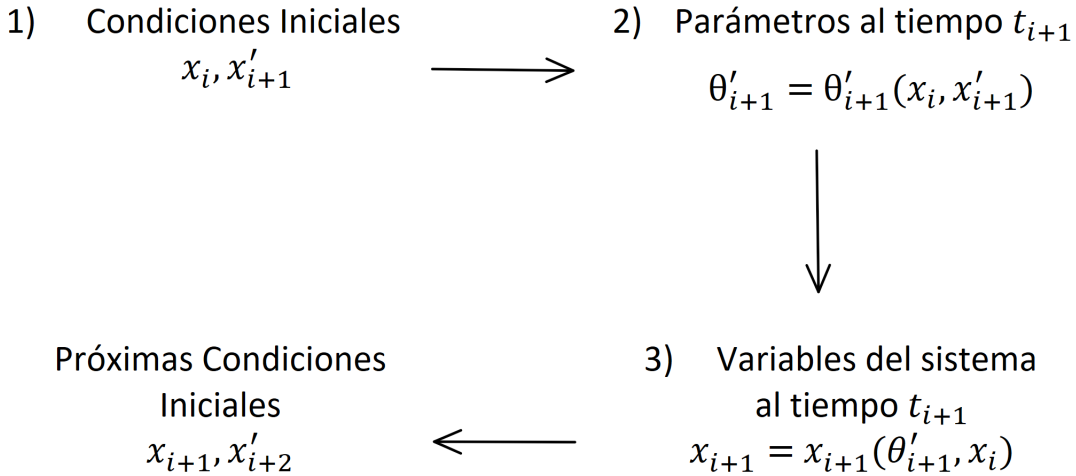


Figura 5.1: Esquema del algoritmo para calcular los parámetros en cada tiempo t_i . Las variables x'_i son las estimaciones por el filtro Kalman y las variables θ'_i corresponden a los parámetros calculados. 1) Se tienen las condiciones iniciales x_i y las estimaciones por Kalman x'_{i+1} . 2) Se usan los valores del punto anterior para calcular los parámetros que se necesitan para llegar a la mejor estimación. 3) Los parámetros hallados y el estado del sistema anterior son usados para encontrar el siguiente estado del sistema.

En este sistema de 3 variables con 6 ecuaciones (5.2) no es necesario usarlas todas para resolverlo. Usando únicamente la primera, la segunda y la cuarta ecuación, obtenemos cómo cambian los parámetros según el diagrama de la figura (5.1).

$$\left\{ \begin{array}{l} \alpha'_i = [m'_{1,i+1} - m'_{2,i+1} + (m_{1,i} - m_{2,i})(\Delta t - 1)] \frac{(1 + p_{3,i}^n)(1 + p_{1,i}^n)}{\Delta t(p_{1,i}^n - p_{3,i}^n)} \\ \alpha'_{0,i} = \frac{m'_{1,i+1} - m_{1,i}}{\Delta t} + m_{i,1} - \frac{\alpha'_i}{1 + p_{3,i}^n} \\ \beta'_i = \frac{p'_{1,i+1} - p_{1,i}}{\Delta t(m_{1,i} - p_{1,i})} \end{array} \right. \quad (5.3)$$

Es importante notar que las expresiones (5.3) se indefinen cuando se cumple:

$$\begin{aligned} p_{1,i}^n &= p_{3,i}^n \\ m_{1,i} &= p_{1,i} \end{aligned} \quad (5.4)$$

Con las expresiones (5.3) se encuentra la dinámica de los parámetros programando las ecuaciones en Python, no obstante, la presencia de los puntos (5.4) implica la existencia de singularidades en las ecuaciones, provocando que el cálculo de los parámetros sea indeterminado en múltiples instantes. En consecuencia, la estimación de los parámetros se vuelve fuertemente sensible a la elección del método numérico empleado para resolver las ecuaciones diferenciales, ya que la proximidad a las singularidades puede amplificar errores de redondeo y limitar la precisión alcanzable.

5.1. Interpolación

Se toma en cuenta que el tamaño de Δt depende del número de puntos que se usen. Del capítulo anterior obtuvimos las estimaciones por Kalman-S-G y S-G-Kalman. Sin embargo, únicamente tenemos 100 puntos disponibles, lo que hace que Δt sea muy grande. Por esa razón se realiza una interpolación de los datos. Esta interpolación busca al menos tener la misma cantidad de puntos que se usaron para resolver el sistema y así calcular el error cuadrático medio. En este caso particular 1000.

La interpolación es lineal mediante la función Python `frame.interpolate(method = 'linear')`. Entre cada uno de los puntos estimados se interpolan 9 puntos, dando un total de 1000.

5.1.1. Interpolación de la estimación Kalman

Realizando la interpolación de 100 a 1000 puntos, se obtuvo una nueva estimación para cada variable del sistema.

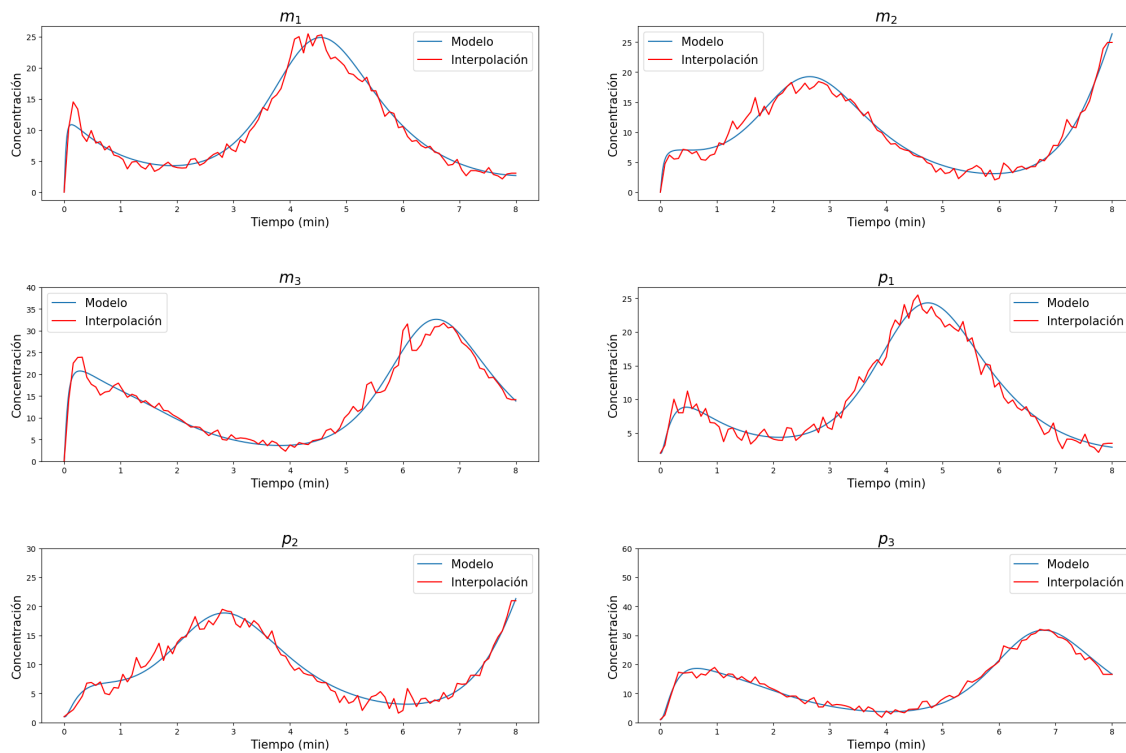


Figura 5.2: Gráfica de la interpolación de las estimaciones realizadas únicamente usando el filtro Kalman

5.1.2. Interpolación de la estimación Kalman-Savitzky-Golay (KSG)

Realizando la interpolación de 100 a 1000 puntos se obtuvo una nueva estimación para cada variable del sistema. En este caso se usó la estimación realizada por Kalman y posteriormente suavizada por S-G.

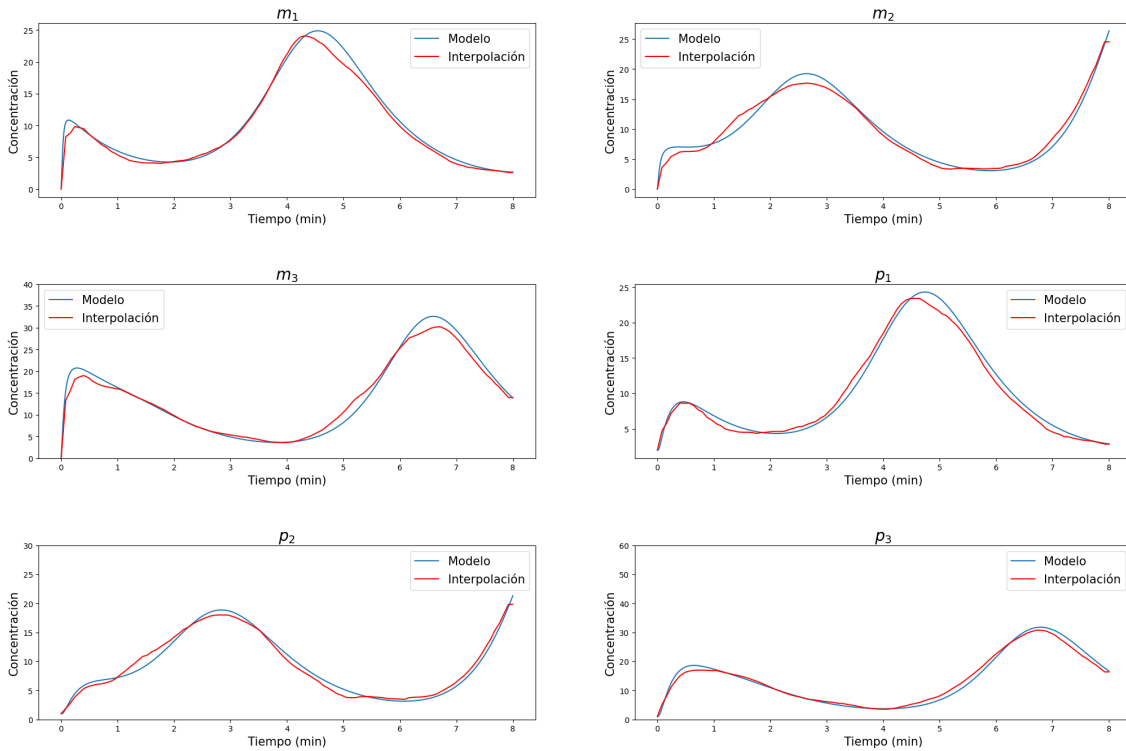


Figura 5.3: Gráfica de la interpolación de las estimaciones realizadas únicamente usando el filtro Kalman suavizada por S-G.

La calidad de las estimaciones interpoladas de ambas se compara en la tabla (5.1).

Variable	Kalman	KSG
m_1	1.109	0.833
m_2	0.936	0.928
m_3	2.5	1.952
p_1	0.959	0.722
p_2	0.938	0.614
p_3	1.228	1.13

Tabla 5.1: Tabla comparativa del error cuadrático medio de las interpolaciones de las estimaciones por Kalman y por KSG en cada una de las variables del sistema.

5.2. Cálculo de Parámetros

Se programaron en Python las ecuaciones 5.3 para encontrar los parámetros en cada punto usando ambas estimaciones (KSG y Kalman).

5.2.1. Parámetros usando la estimación Kalman

Al usar la estimación interpolada realizada con el filtro Kalman, obtenemos los parámetros de la figura (5.4) que, mediante el modelo, reproducen la dinámica observada en la figura (5.5)

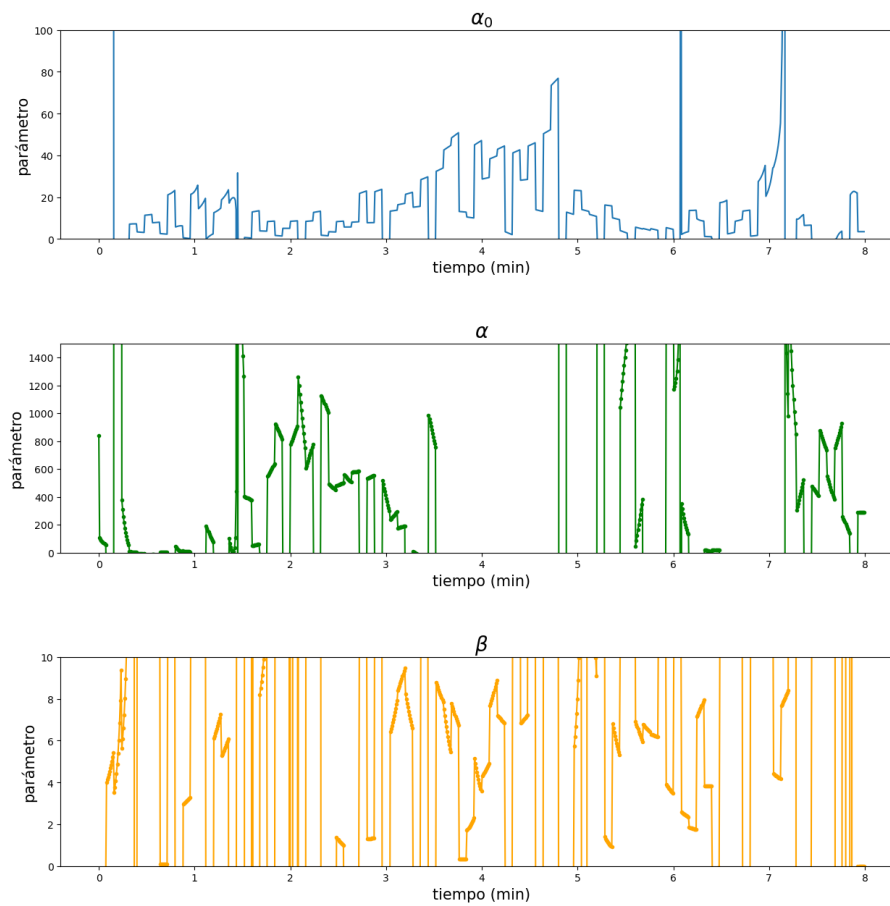


Figura 5.4: Parámetros usando directamente la estimación por filtro Kalman

Estimación de los parámetros del represilador

5.2 Cálculo de Parámetros

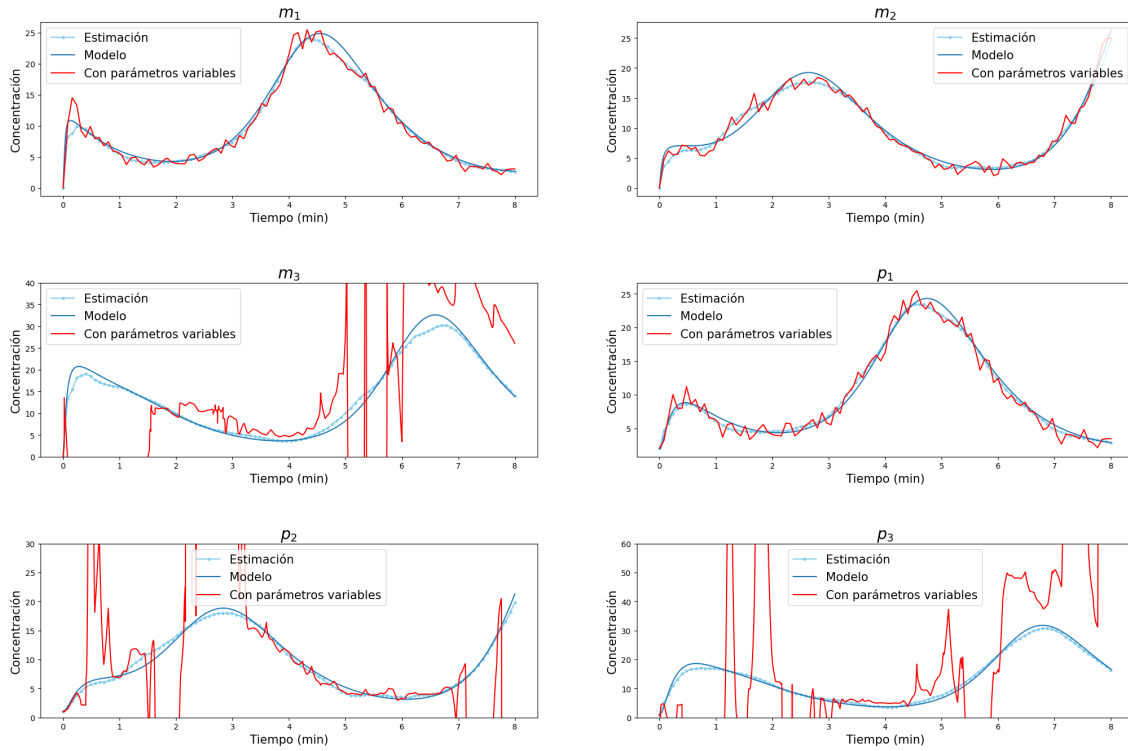


Figura 5.5: Dinámica de las variables del sistema usando los parámetros encontrados mediante la estimación directamente realizada por Kalman

Las gráficas p_1 , p_2 y p_3 en la figura 5.4 se acercan a la dinámica original dada por los parámetros constantes correctos. Sin embargo, es posible notar que las demás gráficas tienen cambios abruptos y, si bien no tienden hacia el infinito, son lo suficientemente grandes para no mostrar similitud con la dinámica original. Esto se debe en gran parte al método numérico empleado para resolver el sistema de ecuaciones diferenciales, como se comentó en la sección anterior. En este caso, el método de diferencias finitas de Euler no genera el resultado con la precisión deseada y es posible indagar en otros métodos para evaluar los resultados.

5.2.2. Parámetros usando la estimación KSG

Al usar la estimación interpolada realizada con el filtro Kalman, se calculan los parámetros de la figura (5.6) que, a su vez, mediante el modelo, reproducen la dinámica observada en la figura (5.7).

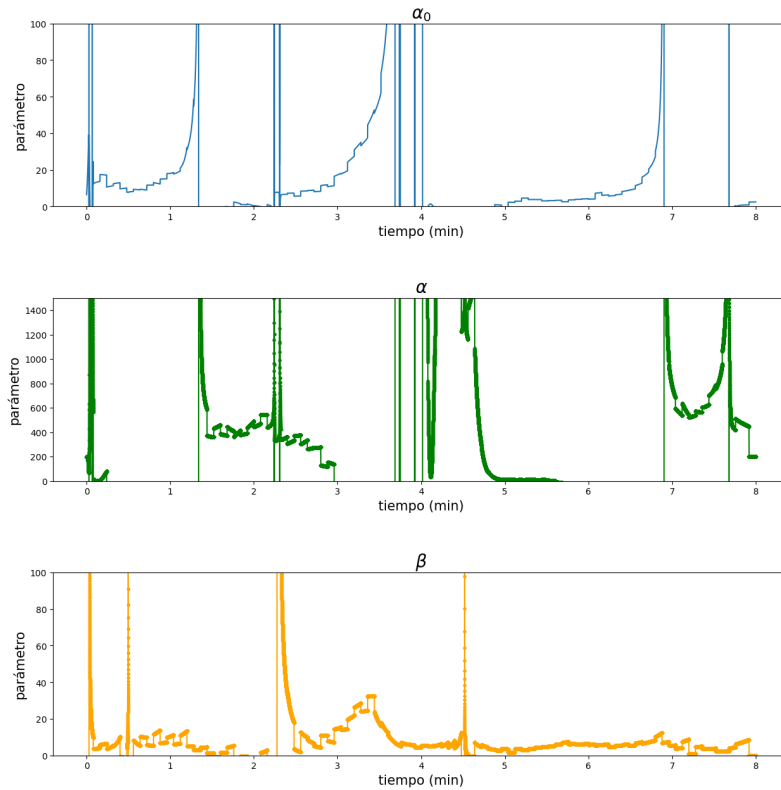


Figura 5.6: Gráficas de los parámetros α_0 , α y β

Es posible notar en la figura (5.6) que los parámetros tienen cambios muy pronunciados en múltiples puntos. Este efecto se debe a que la estimación por el filtro Kalman tiene cambios bruscos; por lo tanto, los parámetros cambian radicalmente para reproducir el resultado de esa estimación.

Estimación de los parámetros del represilador

5.2 Cálculo de Parámetros

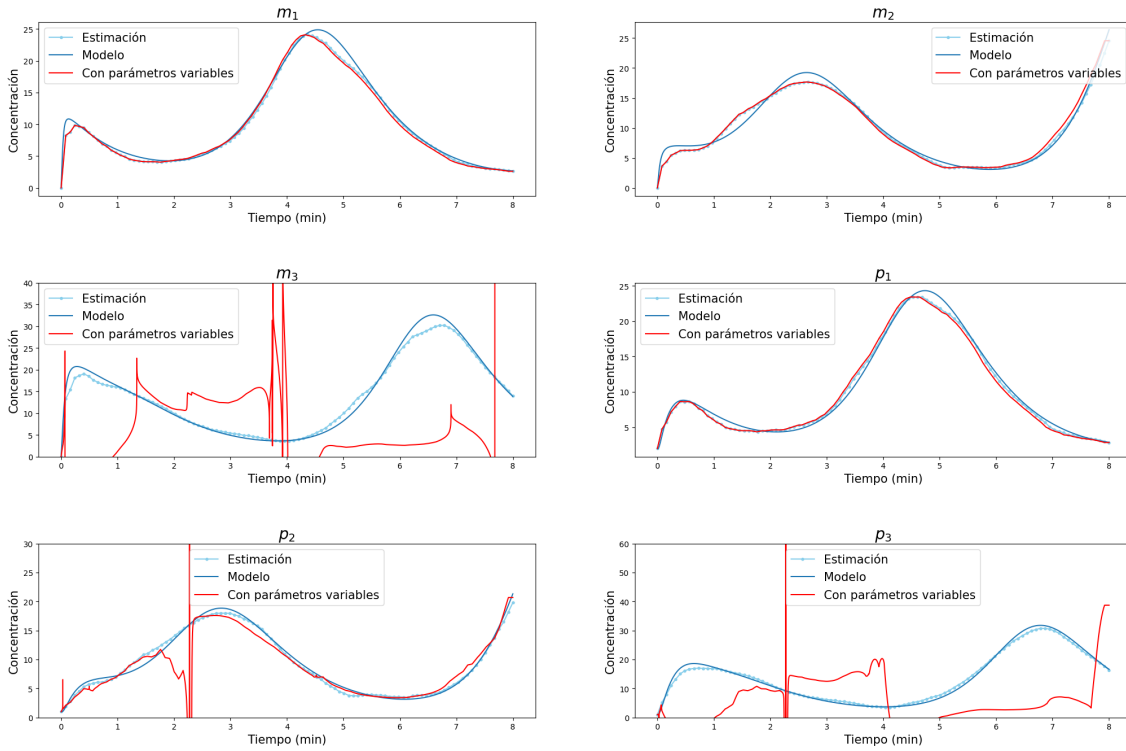


Figura 5.7: Dinámica de las concentraciones mediante el modelo usando los parámetros variables hallados.

En la figura (5.7) se puede observar que los parámetros hallados solo reproducen las estimaciones por el filtro Kalman de las variables m_1 , m_2 y p_3 y con mejor desempeño que usando la estimación directa por Kalman (figura 5.5), las cuales forman parte de las 3 ecuaciones del sistema que se usaron para encontrar los parámetros. Para la proteína p_3 se acerca a la dinámica real; sin embargo, varios puntos tienen esos cambios bruscos.

Es necesario considerar que para obtener los parámetros variables que reproduzcan las estimaciones (suavizadas) no todos los parámetros deben ser iguales, es decir, no existen únicamente 3 parámetros para el sistema, ya que no es necesariamente simétrico en la realidad [1].

Variable	Kalman	KSG
m_1	1.109	0.833
m_2	0.936	0.928
m_3	423.378	321.61
p_1	0.959	0.722
p_2	7.715×10^9	3.168×10^{27}
p_3	3.369×10^7	5.06×10^{26}

Tabla 5.2: Tabla comparativa del error cuadrático medio de las dinámicas producidas por el modelo usando los parámetros encontrados mediante las interpolaciones de las estimaciones por Kalman y por KSG en cada una de las variables del sistema.

5.3. Estimación de 6 parámetros diferentes para el represilador

Dado que los parámetros no reprodujeron todas las variables estimadas debido a que los parámetros que se encuentran únicamente toman en cuenta a las ecuaciones de m_1 , m_2 y p_3 y tomando en cuenta que el modelo del represilador abordado en el Capítulo 1 se simplificó tomando únicamente 3 parámetros, sin embargo se sabe que en la práctica no es así [1]. Se propone encontrar 6 parámetros variables de forma que reproduzcan las estimaciones realizadas.

$$\left\{ \begin{array}{l} m'_{1,i+1} = m_{1,i}(1 - \Delta t) + \left(\alpha'_{0,i+1} + \frac{\alpha'_{1,i+1}}{1 + p_{3,i}^n} \right) \Delta t \\ m'_{2,i+1} = m_{2,i}(1 - \Delta t) + \left(\alpha'_{0,i+1} + \frac{\alpha'_{2,i+1}}{1 + p_{1,i}^n} \right) \Delta t \\ m'_{3,i+1} = m_{3,i}(1 - \Delta t) + \left(\alpha'_{0,i+1} + \frac{\alpha'_{2,i+1}}{1 + p_{2,i}^n} \right) \Delta t \\ p'_{1,i+1} = \beta'_{1,i+1}(m_{1,i} - p_{1,i})\Delta t + p_{1,i} \\ p'_{2,i+1} = \beta'_{2,i+1}(m_{2,i} - p_{2,i})\Delta t + p_{2,i} \\ p'_{3,i+1} = \beta'_{3,i+1}(m_{3,i} - p_{3,i})\Delta t + p_{3,i} \end{array} \right. \quad (5.5)$$

Del sistema (5.5) se hallarán los 6 parámetros variables propuestos a través del método desarrollado en el presente trabajo.

$$\alpha'_{0,i}, \alpha'_{1,i}, \alpha'_{2,i}, \beta'_{1,i}, \beta'_{2,i}, \beta'_{3,i} \quad (5.6)$$

Por lo tanto, realizando la misma metodología que en la sección 5 encontramos las siguientes ecuaciones para cada uno de los parámetros:

Estimación de los parámetros del represilador
5.3 Estimación de 6 parámetros diferentes para el represilador

$$\left\{ \begin{array}{l} \alpha'_{2,i} = [m'_{3,i+1} - m'_{2,i+1} + (m_{3,i} - m_{2,i})(\Delta t - 1)] \frac{(1 + p_{2,i}^n)(1 + p_{1,i}^n)}{\Delta t(p_{1,i}^n - p_{2,i}^n)} \\ \alpha'_{0,i} = \frac{m'_{2,i+1} - m_{2,i}}{\Delta t} + m_{2,i} - \frac{\alpha'_{2,i}}{1 + p_{1,i}^n} \\ \alpha'_{1,i} = \left(\frac{m'_{1,i+1} - m_{1,i}}{\Delta t} + m_{1,i} - \alpha'_{0,i} \right) (1 + p_{3,i}^n) \\ \beta'_{l,i} = \frac{p'_{l,i+1} - p_{l,i}}{\Delta t(m_{l,i} - p_{l,i})}; \quad l = 1, 2, 3 \end{array} \right. \quad (5.7)$$

Para el cálculo de los seis parámetros (5.7) únicamente se usó la mejor estimación interpolada: KSG. Se realizó el algoritmo en Python y se obtuvo su valor en cada punto (representados en la figura 5.8). Es posible notar que α_1 y α_2 se acercan al valor real $\alpha = 500$ y en algunas partes $\beta_i, i = 1, 2, 3$ se acercan a $\beta = 5$.

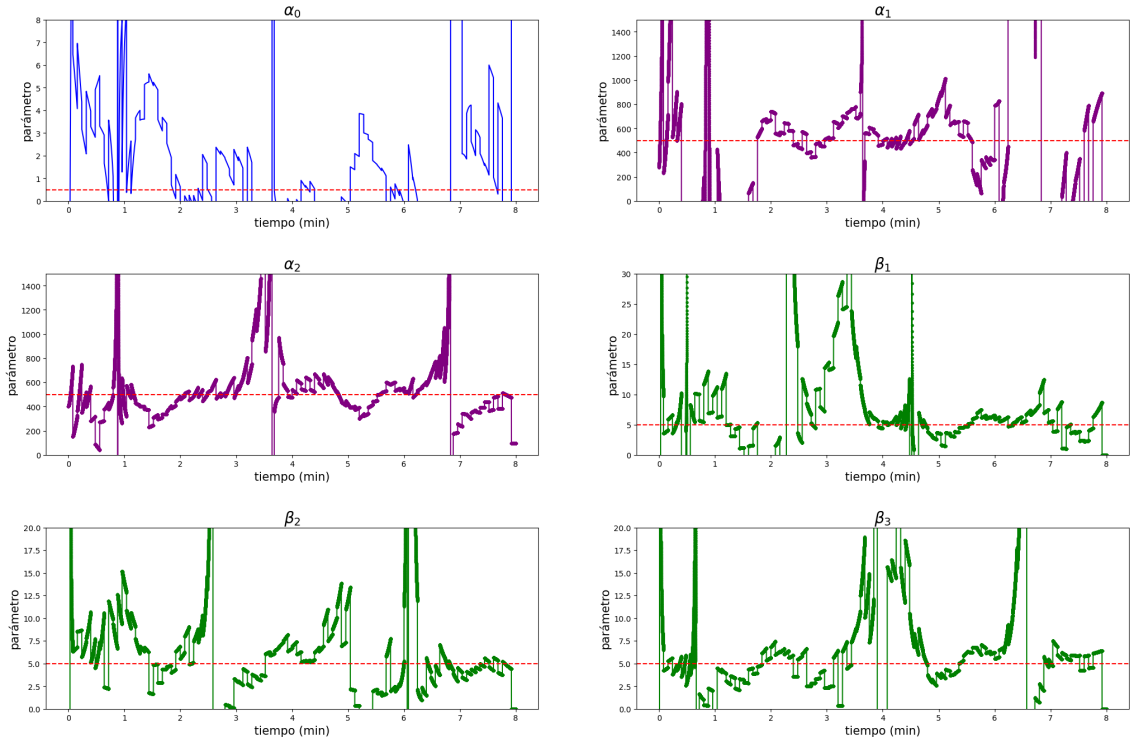


Figura 5.8: Gráfica de los seis parámetros variables propuestos

Estimación de los parámetros del represilador

5.3 Estimación de 6 parámetros diferentes para el represilador

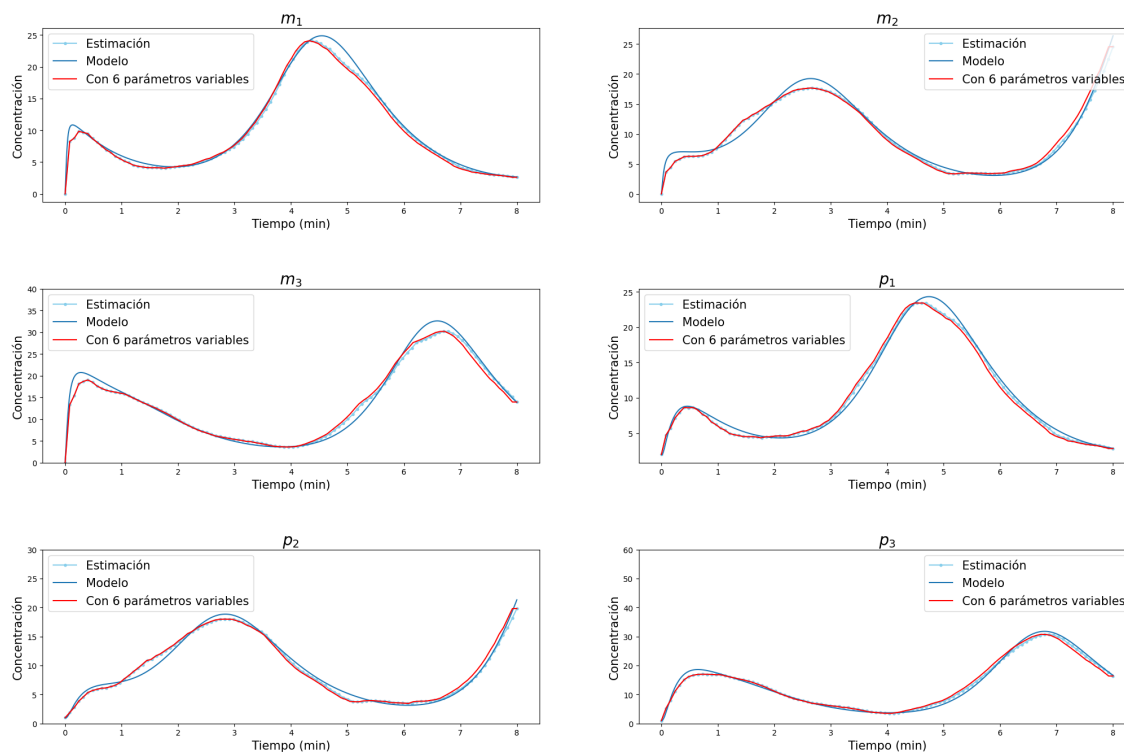


Figura 5.9: Gráficas de las concentraciones calculadas a partir de los seis parámetros hallados usando la estimación KSG.

En la figura 5.9 es posible notar que los seis parámetros calculados reproducen las estimaciones por KSG de todas las variables del sistema. El error cuadrático medio de este resultado se compara con el uso del error cuadrático medio como alternativa en lugar del filtro KSG (tabla 5.3).

Variable	Kalman	KSG	LSM	Parámetros Variables
m_1	1.215	0.516	-	0.833
m_2	1.206	0.733	-	0.928
m_3	2.901	1.501	-	1.952
p_1	1.082	0.363	0.973	0.722
p_2	1.249	0.442	0.660	0.614
p_3	1.327	0.759	2.933	1.306

Tabla 5.3: Tabla comparativa del error cuadrático medio de las dinámicas producidas por el modelo usando los parámetros encontrados mediante las estimaciones por Kalman y por KSG en cada una de las variables del sistema.

Capítulo 6

Conclusión

Este trabajo concluye que la estrategia más adecuada para estimar el estado de un sistema es aplicar primero el filtro de Kalman y posteriormente el filtro de Savitzky–Golay. Al invertir este orden, el filtro de Savitzky–Golay puede eliminar componentes de la señal que contienen información relevante para el proceso de estimación, lo que afecta negativamente el desempeño del filtro de Kalman. Además, se comprobó que esta combinación es más efectiva que un simple ajuste por mínimos cuadrados en una ventana móvil, ya que preserva mejor las características dinámicas del sistema.

Otro hallazgo importante es que, incluso en escenarios con una cantidad limitada de datos y con un sistema parcialmente observable, es posible aproximar parámetros del modelo que permiten reproducir una estimación confiable del estado. No obstante, se observó que los valores de dichos parámetros pueden fluctuar abruptamente debido al ruido presente en la estimación. Aunque los parámetros obtenidos se encuentran cerca de los valores reales, los resultados sugieren que aún es necesario aplicar técnicas adicionales de suavizado para estabilizar la estimación.

Finalmente, los resultados indican que este enfoque también muestra potencial como herramienta para la estimación de parámetros en sistemas no autónomos en tiempo real, lo que abre nuevas posibilidades para su aplicación en contextos experimentales y de control adaptativo.

Bibliografía

- [1] J. Garcia-Ojalvo, M. Elowitz, and S. Strogatz, *Proc. Natl. Acad. Sci. U.S.A.* 101, 10955 (2004).
- [2] Emmert-Streib, F., Dehmer, M., & Haibe-Kains, B. (2014). Gene regulatory networks and their applications: understanding biological and medical problems in terms of networks. En *Frontiers in Cell and Developmental Biology* (Vol. 2). Frontiers Media SA. <https://doi.org/10.3389/fcell.2014.00038>
- [3] Lajoie, M. J., Boyken, S. E., Salter, A. I., Bruffey, J., Rajan, A., Langan, R. A., Olshefsky, A., Muhunthan, V., Bick, M. J., Gewe, M., Quijano-Rubio, A., Johnson, J., Lenz, G., Nguyen, A., Pun, S., Correnti, C. E., Riddell, S. R., & Baker, D. (2020). Designed protein logic to target cells with precise combinations of surface antigens. *Science*, 369(6511), 1637-1643. <https://doi.org/10.1126/science.aba6527>
- [4] Li, M., Qian, H., Ichim, T. E., Ge, W., Popov, I. A., Rycerz, K., Neu, J., White, D., Zhong, R., & Min, W. (2004). Induction of RNA Interference in Dendritic Cells. *Immunologic Research*, 30(2), 215-230. <https://doi.org/10.1385/ir:30:2:215>
- [5] Yang, Q., Yang, F., Dai, W., Meng, X., Wei, W., Cheng, Y., Dong, J., Lu, H., & Dong, H. (2021). DNA Logic Circuits for Multiple Tumor Cells Identification Using Intracellular MicroRNA Molecular Bispecific Recognition. *Advanced Healthcare Materials*, 10(21). <https://doi.org/10.1002/adhm.202101130>
- [6] Gábor, A., & Banga, J. R. (2015). Robust and efficient parameter estimation in dynamic models of biological systems. *BMC Systems Biology*, 9(1). <https://doi.org/10.1186/s12918-015-0219-2>
- [7] Guido Sanguinetti & Vân Anh Huynh-Thu (eds.), *Gene Regulatory Networks: Methods and Protocols*, *Methods in Molecular Biology*, vol. 1883, https://doi.org/10.1007/978-1-4939-8882-2_1, Springer Science+Business Media, LLC, part of Springer Nature 2019
- [8] Diccionario de cáncer del NCI. (s.f.). *Cancer.gov*. <https://www.cancer.gov/espanol/publicaciones/diccionarios/diccionario-cancer/def/transcripcion>
- [9] Wu, S., Cui, T., Zhang, X., & Tian, T. (2020). A non-linear reverse-engineering method for inferring genetic regulatory networks. *PeerJ*, 8, e9065. <https://doi.org/10.7717/peerj.9065>
- [10] Saadatpour, A., & Albert, R. (2016). A comparative study of qualitative and quantitative dynamic models of biological regulatory networks. *EPJ Nonlinear Biomedical Physics*, 4(1). <https://doi.org/10.1140/epjnbp/s40366-016-0031-y>
- [11] Montero, S., Martin, R., Mansilla, R., Cocho, G., & Nieto-Villar, J. M. (2017). Parameters Estimation in Phase-Space Landscape Reconstruction of Cell Fate: A Systems Biology Approach. En *Methods in Molecular Biology* (pp. 125–170). Springer New York. https://doi.org/10.1007/978-1-4939-7456-6_8

- [12] Miramontes, P., Peralta, R., Garcés, H., & Cocho, G. (2015). Carcinogénesis y complejidad. *INTERdisciplina*, 3(6). <https://doi.org/10.22201/ceiich.24485705e.2015.6.49479>
- [13] Uri Alon. An introduction to systems biology: design principles of biological circuits. CRC Press, second edition, 2020 1.1.
- [14] Savitzky, A., & Golay, M. J. E. (1964). Smoothing and Differentiation of Data by Simplified Least Squares Procedures. *Analytical Chemistry*, 36(8), 1627-1639. <https://doi.org/10.1021/ac60214a047>
- [15] Schmid, M., Rath, D., & Diebold, U. (2022). Why and How Savitzky–Golay Filters Should Be Replaced. *ACS Measurement Science Au*, 2(2), 185-196. <https://doi.org/10.1021/acsmearsciau.1c00054>
- [16] Schafer, R. (2011). What Is a Savitzky-Golay Filter? [Lecture Notes]. *IEEE Signal Processing Magazine*, 28(4), 111-117. <https://doi.org/10.1109/msp.2011.941097>
- [17] Lillacci, G., & Khammash, M. (2010). Parameter identification of biological networks using extended Kalman filtering and χ^2 criteria. 2021 60th IEEE Conference On Decision And Control (CDC), 3367-3372. <https://doi.org/10.1109/cdc.2010.5717460>
- [18] Debomita Chakraborty, Raghunathan Rengaswamy, and Karthik Raman, *ACS Synthetic Biology* 2022 11 (4), 1377-1388 DOI: 10.1021/acssynbio.1c00557
- [19] Olguta Buse, Rodrigo Pérez, and Alexey Kuznetsov *Phys. Rev. E* 81, 066206 – Published 11 June 2010
- [20] Taboga, Marco (2021). "Gauss Markov theorem", *Lectures on probability theory and mathematical statistics*. Kindle Direct Publishing. Online appendix. <https://www.statlect.com/fundamentals-of-statistics/Gauss-Markov-theorem>.
- [21] Neal Gallagher (2020). Savitzky-Golay Smoothing and Differentiation Filter. *Eigenvector Research Incorporated*. DOI: 10.13140/RG.2.2.20339.50725
- [22] Brunton, S. L., & Kutz, J. N. (2021). *Data-Driven Science and Engineering*. Cambridge University Press. <https://doi.org/10.1017/9781108380690>
- [23] Pei, Y., Biswas, S., Fussell, D. S., & Pingali, K. (2019). An elementary introduction to Kalman filtering. *En Communications of the ACM* (Vol. 62, Issue 11, pp. 122–133). Association for Computing Machinery (ACM). <https://doi.org/10.1145/3363294>
- [24] Gabriel A. Terejanu, *Extended Kalman Filter Tutorial*, Department of Computer Science and Engineering University at Buffalo, Buffalo, NY 14260, terejanu@buffalo.edu.
- [25] Saleem, 2019, Weather Prediction Model using Savitzky–Golay and Kalman Filters, *Procedia Computer Science* (Vol. 163, pp. 397-406, 2019. Elsevier).
- [26] Julier, S., & Uhlmann, J. (2004). Unscented Filtering and Nonlinear Estimation. *Proceedings Of The IEEE*, 92(3), 401-422. <https://doi.org/10.1109/jproc.2003.823141>