

Aplicación del método del filtro de Kalman en la minería de datos en cursos de modelamiento estocástico

JOSÉ G. SOLORZANO MOVILLA, DIANA SUAREZ LÓPEZ,
AND LIZ KARINA BARRAGAN

Abstract. This article presents the successful application of the Kalman Filter method in two different contexts: the estimation of daily cases of COVID-19 in Colombia and the projection of points for two soccer teams in the Spanish Professional League.


In the first case, the complexity of predicting the evolution of the pandemic in a diverse country like Colombia is addressed. The Kalman Filter proved to be a robust tool to adjust and improve daily case estimates, providing more accurate results than conventional methods. The filter's ability to adapt to changes in epidemiological trends is revealed as a crucial factor in the effective management of the health crisis.

In the second case, the focus shifts to the sports field, specifically the Spanish Professional Football League. The application of the Kalman Filter to project the points obtained by two teams reveals its usefulness in the analysis of performance throughout the season. The results suggest that this method can be a valuable tool for coaches and analysts, allowing them to adjust strategies in real time and anticipate possible fluctuations in performance. In both cases, the versatility of the Kalman Filter emerges as a crucial factor, highlighting its ability to adapt to different types of data and contexts. These findings underline the potential of this technique not only in the management of health crises, but also in strategic decision-making in the field of sports, consolidating its position as a valuable analytical tool in various disciplines.

Date: Received: Day Month Year; Accepted: Day Month Year.

2010 *Mathematics Subject Classification.* Primary .

Key words and phrases. Kalman Filter, dynamical system, state spaces.

Author ONE  0000-0002-4176-0300.

Author TWO  0000-0001-9413-3434.

Author THREE  0009-0003-4786-8675.

Resumen. Este artículo presenta la aplicación exitosa del método del Filtro de Kalman en dos contextos distintos: la estimación de los casos diarios de COVID-19 en Colombia y la proyección de puntos para dos equipos de fútbol en la Liga Profesional Española.

En el primer caso, se aborda la complejidad de pronosticar la evolución de la pandemia en un país diverso como Colombia. El Filtro de Kalman demostró ser una herramienta robusta para ajustar y mejorar las estimaciones de casos diarios, proporcionando resultados más precisos que los métodos convencionales. La capacidad del filtro para adaptarse a cambios en las tendencias epidemiológicas se revela como un factor crucial en la gestión efectiva de la crisis sanitaria.

En el segundo caso, el enfoque se desplaza hacia el ámbito deportivo, específicamente la Liga Profesional Española de fútbol. La aplicación del Filtro de Kalman para proyectar los puntos obtenidos por dos equipos revela su utilidad en el análisis de rendimiento a lo largo de la temporada. Los resultados sugieren que este método puede ser una herramienta valiosa para entrenadores y analistas, permitiéndoles ajustar estrategias en tiempo real y anticiparse a posibles fluctuaciones en el desempeño.

En ambos casos, la versatilidad del Filtro de Kalman emerge como un factor crucial, destacando su capacidad para adaptarse a diferentes tipos de datos y contextos. Estos hallazgos subrayan el potencial de esta técnica no solo en la gestión de crisis sanitarias, sino también en la toma de decisiones estratégicas en el ámbito deportivo, consolidando su posición como una herramienta analítica valiosa en diversas disciplinas.



Esta publicación está bajo una
licencia Creative Commons
Reconocimiento-NoComercial 4.0

1. Introducción

El método del filtro de Kalman es uno de los usados para estimar los estados de un sistema dinámico estocástico, este se presenta por lo general para el filtrado de señales y de estimación de sistemas dinámicos expresados en espacios de estados. Sus aplicaciones son diversas y versátiles, dada su característica como el mejor estimador lineal insesgado, por lo cual su uso es explícito en áreas como la robótica, las ciencias del espacio, en especial sistemas dinámicos estocásticos.

En el siguiente artículo se presentan dos aplicaciones del filtro de Kalman en igual número de conjuntos de datos, como parte de las clases de un curso de modelos de ingeniería, en una institución de educación superior de la ciudad de Barranquilla, Colombia. El primero en la estimación de los casos de COVID 19, el segundo con dos equipos del fútbol español. El documento inicialmente se hace un análisis de los trabajos relacionados, luego se presenta la metodología, después los resultados y finalmente las conclusiones.

2. Trabajos relacionados

De acuerdo con [2], en 1960, Kalman presenta un conjunto de ecuaciones que toma como base el método de aproximación por mínimo cuadrados aportando una nueva forma de expresar los modelos en la que incluye las incertezas producto del ruido blanco, esto constituye la base de lo que se ha denominado el filtro de Kalman.

[5] por su parte describe que “El filtro de Kalman es un algoritmo recursivo desarrollado por Rudolf E. Kalman en 1960 que sirve para poder identificar el estado de un sistema dinámico lineal discreto, en presencia de ruido aleatorio”.

[4] desarrolla un “método para realizar capturas de movimiento mediante marcadores ópticos y placas de fuerza, de modo que se pueda reconstruir y visualizar el movimiento en tiempo real, así como estimar valores derivados del mismo, tales como esfuerzos motores y reacciones

Aplicación del método del filtro de Kalman en la minería de

articulares”.

[1] desarrolla un proyecto de ayuda a la navegación que consiste en “un navegador inercial de bajo costo hecho con filtros de Kalman, este se encuentra montado sobre un automóvil de juguete y el navegador se implementa por medio de Arduino, logrando así no solo tener un bajo costo de implementación sino una mejora significativa en la precisión de los sensores utilizados para hallar la posición del automóvil de juguete”.

Siguiendo la misma línea, [1] expone un Sistema para la ubicación de fallas eléctricas en líneas de transmisión basado en el filtro de Kalman, el cual se basa en el análisis de las señales de corriente y voltaje obtenidas del simulador EMPT en el cual se planteó varios escenarios con fallas de tipo simétricas y asimétricas en las líneas de transmisión del SEP. El problema está enfocado en ubicar fallas eléctricas en las líneas de transmisión del sistema IEEE de 9 barras.

En la tesis doctoral de [3] plantea que: “En el control de navegación de un vehículo autónomo, la posición y velocidad pueden ser determinadas con un sistema de posicionamiento global (“Global Positioning System” - GPS) y la velocidad angular y aceleración con un sistema de navegación inercial (“Inertial Navigation Systems”, INS). Sin embargo, los errores asociados a estas medidas no permiten que éste sea el único sistema de medida utilizado en un vehículo autónomo. Surge entonces la necesidad de implementar herramientas que permitan confiablemente determinar el estado del sistema (posición, velocidad e inclinación) en cualquier instante y realicen las acciones de control necesarias para cumplir con una tarea o trayectoria determinada en forma óptima y recursiva, teniendo en cuenta el modelo interno del sistema. La aplicación de un filtro de Kalman es clave en esta etapa, ya que permite la integración de la información del GPS”.

3. METODOLOGÍA

Para el desarrollo de las aplicaciones del FK que presentan en este trabajo, se implementaron las siguientes etapas:

- Etapa 1: Búsqueda de los datos en las páginas, ourworldindata.com y de la liga de fútbol de España.
- Etapa 2: Depuración de los datos
- Etapa 3: Implementación del algoritmo de Kalman, basados en programación en R y Python
- Etapa 4: Análisis de resultados

4. ANÁLISIS DE RESULTADOS

Para la implementación del FK se aplicó la forma unidimensional:

$$Est_t = Est_{t-1} + KG[MED. - Est_{t-1}]$$

Dónde:

Est_t , Representa el estado en el tiempo t. Est_{t-1} Considera los valores del estado en el tiempo anterior. $MED.$ Representa las medidas reales.

$KG = \frac{E_{Est}}{E_{Est} + E_{MED}}$ es el valor del vector de ganancia de Kalman, considerando E_{Est} error en la estimación, y E_{MED} el error en las medidas, para este caso se tomará como la varianza.

Luego de llevar a cabo la programación de estas ecuaciones en en R y Python, se pudieron graficar los resultados en las imagenes que presentan a continuación:

Para el caso de la estimación de contagios de COVID 19 para Colombia, se tomaron los datos del sitio web www.ourworldindata.com correspondientes a los casos diarios desde la detección de los primeros casos en el país hasta el día 81 (julio 15 de 2020).

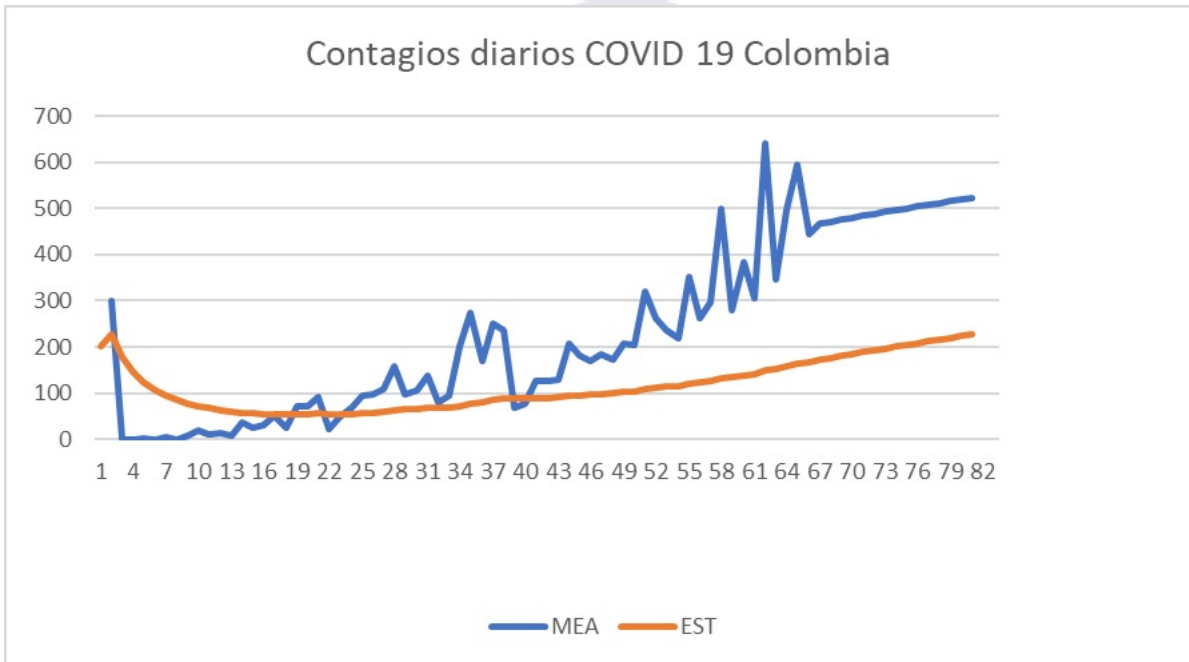


FIGURE 1. Proyección casos Covid

La línea azul en la gráfica representa los datos de contagios por día, y la línea roja la estimación de los casos a partir del FK, en este caso es posible notar que existe una diferencia significativa entre la estimación y los resultados reales, unicamente existe coincidencia entre los días 13 y 41 indicando en ese intervalo un alto valor de ruido blanco, posiblemente, no hubo una mejor estimación dados algunos factores como el reporte de falsos positivos.

Para el caso de los resultados de dos equipos de futbol de la liga española, Real Madrid, Barcelona, tomando los resultados históricos disponibles en la página web de la liga, se encontraron los siguientes resultados:

En el gráfico 2, se muestra el desempeño del Real Madrid Club de Fútbol, en el cual se evidencian pequeñas diferencias entre la estimación y los datos reales. A patir de esta proyección se determinó la cantidad de puntos que alcanzará el citado club deportivo estaria en el intervalo 80-87 para la campaña 2021-2022, comparando con los datos disponibles en el portal EuroSport, se encuentra que los puntos alcanzados fueron 86, temporada donde fue campeón.

Para la campaña, 2022-2023 se estimó que la cantidad máxima de puntos sería de 80, realizando el mismo ejercicio de comparación se encontro que los puntos fueron 78, en ese torneo el club finalizó segundo.

El gráfico 3 muestra la estimación del desempeño de Fútbol Club Barcelona, notando que los resultados de la implementación del FK son muy cercanos a los datos reales, teniendo en cuenta que para la campaña 2021-2022 se proyectó un máximo de 84 puntos, alcanzando un valor de 78 en los datos reales. Mientras que para la liga 2022-2023 se estimó un comportamiento mayor a 85 puntos y en el comparativo se pudo evidenciar que llegó a 88, siendo el campeón en esa campaña.

Aplicación del método del filtro de Kalman en la minería de

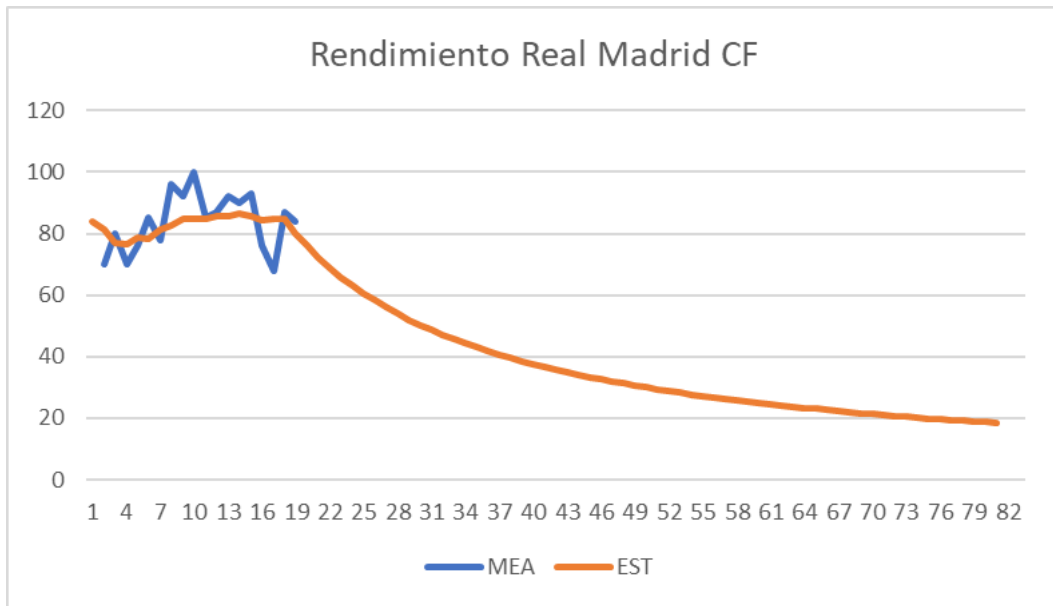


FIGURE 2. Proyección Rendimiento Real Madrid C.F.

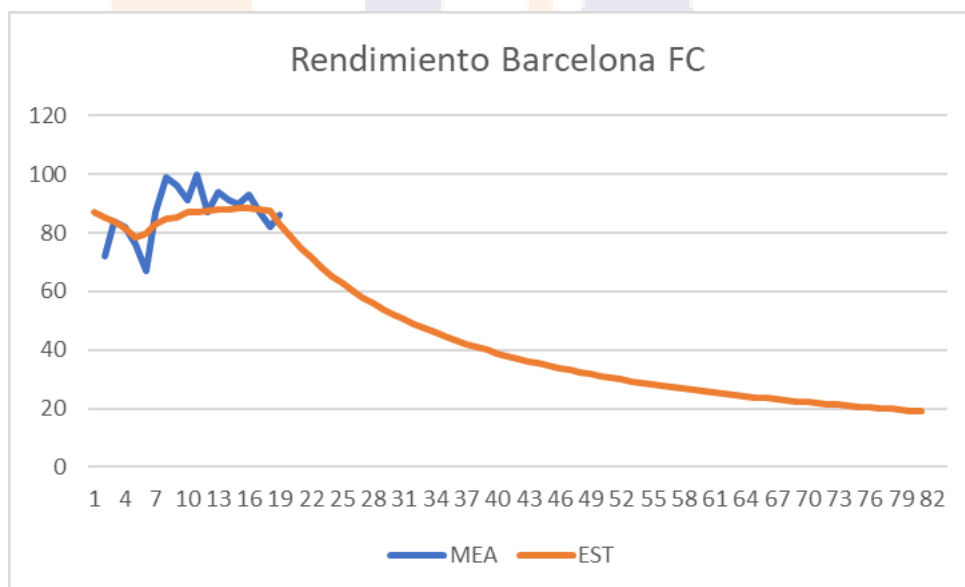


FIGURE 3. Proyección Rendimiento Barcelona F.C.

5. CONCLUSIONES

Los resultados de la implementación confirman la versatilidad del filtro de Kalman, en especial en lo referido a la estimación de fenómenos en los cuales no es explícita la relación entre las variables que inciden en la evolución del sistema dinámico.

Por otro lado, es relevante destacar una excepcionalidad al comparar los resultados entre los casos de COVID 19 y los puntajes de los equipos de fútbol en la liga española. Para el primero se tomaron datos consolidados de 80 días en Colombia, pero el método lo alcanzó a estimar con precisión, conllevando a que el modelo resultante no concordará con la realidad.

En el caso de los clubes de fútbol, la cantidad de datos fueron menores, al considerarse los últimos 22 años para ambos equipos y aunque como se evidencia en las gráficas 2 y 3, existen diferencias, entre la estimación y las medidas, la diferencia es menor en comparación con los resultados del análisis de los casos diarios de COVID 19.

Finalmente, este artículo presentó la forma unidimensional del FK, sin embargo, se plantea la necesidad de implementar las otras formas del FK en futuros trabajos usando un volumen de datos mayor.

REFERENCES

- [1] Bo Liu, Mohamed Mohandes, Huijian Li, Ali Al-Shaikhi, Xu Liu, and Ling Zhao. Likelihood-ratio-based recovery for seismic reflectivity series. *IEEE Transactions on Industrial Electronics*, 69(11):11622–11632, 2021. Cited on p. 3.
- [2] Jose Gregorio Solorzano, Martha Olaciregui Leemow, and Selene Coelho Loibel. Una aplicación del filtro de kalman en el movimiento parabólico. Cited on p. 2.
- [3] Máximo Giovanni Tandazo Espinoza. Aplicación del filtro de kalman difuso en una señal gps con ruido. No Publicado, Julio 2022. Cited on p. 3.
- [4] Helia Yousefnejad and Mohammadali Saniee Monfared. A control algorithm for a non-stationary batch service production system using kalman filter. *Expert Systems with Applications*, 207:117916, 2022. Cited on p. 2.
- [5] Ka-Veng Yuen, Yu-Song Liu, and Wang-Ji Yan. Estimation of time-varying noise parameters for unscented kalman filter. *Mechanical Systems and Signal Processing*, 180:109439, 2022. Cited on p. 2.

ESCUELA SUPERIOR DE ADMINISTRACIÓN PÚBLICA, BARRANQUILLA 127, COLOMBIA
Email address: jose.solorzanom@esap.edu.co

CORPORACIÓN UNIVERSITARIA AMERICANA, BARRANQUILLA 127, COLOMBIA
Email address: dsuarez@coruniamericana.edu.co

ESCUELA SUPERIOR DE ADMINISTRACIÓN PÚBLICA, BARRANQUILLA 127, COLOMBIA
Email address: lizbarr@esap.edu.co