

IMPLEMENTACIÓN DE FILTROS DE KALMAN COMO MÉTODO DE AJUSTE A LOS MODELOS DE PRONÓSTICO (GFS) DE TEMPERATURAS MÁXIMAS Y MÍNIMA PARA ALGUNAS CIUDADES DE COLOMBIA

Dorado, J; Ruíz, J. F.

Grupo de Modelamiento de Tiempo y Clima

Subdirección de Meteorología – IDEAM

RESUMEN

En este estudio se presentan los resultados obtenidos en la metodología aplicada del Filtro de Kalman como método de ajuste a las salidas del modelo GFS para el pronóstico de temperaturas máximas y mínimas en algunas ciudades de Colombia. Los datos correspondieron a descargas diarias del resultado de pronóstico del modelo GFS de 0.5° de las 12z y 18z (temperaturas mínimas y máximas), con horizonte a 4 días sobre los lugares donde la Oficina de Pronóstico y Alertas emite este tipo de información para los meses de Mayo hasta Octubre del 2013. Inicialmente se partió de una revisión bibliográfica del método y una descripción del análisis estadístico tipo multivariado para seleccionar la mejor variable explicativa de la temperaturas máximas y mínimas. Es importante aclarar que el método de Filtro de Kalman fue implementado en dos partes, una como método de ajuste y otra como modelo de pronóstico. El ejercicio realizado con Filtros de Kalman, consistió en corregir el pronóstico de una estación con base en el pronóstico dado por el modelo GFS en el tiempo actual (t), y los históricos GFS con los observados (t-1); por tanto, al suponer datos de GFS se generó el pronóstico corregido de la estación, siendo esta similar a lo observado. A partir de esto, se realizaron pruebas de aplicación y corrección del método en hoja electrónica como herramienta inicial. Las diferencias obtenidas entre lo modelado por el GFS y lo observado inicial, fueron grandes (correlaciones 0.6 en promedio) y la aplicación del modelo como método de ajuste perfeccionó los resultados, igualmente la proyección mantuvo un comportamiento según lo observado, mejorando significativamente las series (correlaciones 0.9 y errores cuadráticos medios de 0,32 en promedio general de las estaciones evaluadas). Una vez evaluado el método se procedió a generar el aplicativo ApplyKalmanFilter_Linux en R como herramienta a la oficina de Alertas y Pronóstico de IDEAM.

Palabras claves: Modelo GFS, Filtro Kalman, Pronóstico

ABSTRAT

In this study we present the results of the methodology applied to obtain the Kalman Filter as a method of adjustment to the GFS model output to forecast extreme temperatures in some cities in Colombia. Data were for daily downloads forecast result 0.5o GFS 12z and the 18z (minimum and maximum temperatures), with horizon 4 days about where Forecast Office and Alernas issues this type of information for months of May through October of 2013. Initially we started from a literature review of the method and a description of the type multivariate statistical analysis to select the best explanatory variable temperatures. It is important to note that the Kalman Filter method was implemented in two parts, one as the adjustment method and another as forecast model. Exercise on Kalman filters consisted of correcting the prognosis of a station based on the prognosis given by the GFS at the current time (t), and the observed historical GFS (t-1), therefore GFS assuming data generated corrected prognosis station, this being similar to that observed. From this, tests were performed and correction application of the method as an initial tool spreadsheet. The differences obtained between the GFS and modeled by the initial observed were large (0.6 on average correlations) and the application of the model as adjustment method perfected the results, the projection also maintained as observed behavior, significantly improving the series (Correlations of 0.9 and mean square errors of 0.32 in overall assessed stations). Having evaluated the method proceeded to generate the script ApplyKalmanFilter_Linux in R as a tool to office IDEAM Alerts and Forecast.

Key words: Model GFS, Kalman Filter, Forecast.

1. Introducción

El Filtro de Kalman es una herramienta poderosa desarrollada por Rudolf E. Kalman en 1960 empleada para controlar los sistemas ruidosos, específicamente para identificar el problema lineal cuadrático, cuando este es sometido a ruido blanco aditivo. Según Andrews (2001), el estimador resultante es estadísticamente óptimo con respecto a cualquier función cuadrática de estimación del error. La ventaja de este algoritmo respecto a otros, es que el filtro de Kalman es capaz de escoger la forma óptima cuando se conocen las varianzas de los ruidos que afectan al sistema, además es el mejor estimador lineal si sus vectores tienen estadísticas arbitrarias y si éstas son variables aleatorias normales (Deossa, 2009).

Cuando el sistema dinámico no es lineal (como en la mayoría de los casos), se aplica una extensión del filtro con un proceso de linealización, y se conoce como Filtro de Kalman Extendido (FKE), (Haykin, 1999). Con el desarrollo de este filtro se busca linealizar el modelo de estado NO LINEAL en cada instante de tiempo alrededor del último estado conocido, una vez hallado el modelo lineal se procede a aplicar el Filtro de Kalman.

En conclusión, el Filtro de Kalman, provee una solución al problema lineal de filtrado óptimo. Este se aplica en ambientes estacionario como no estacionario y permite actualizar el estado a partir del dato anterior (salida) y el nuevo dato (entrada).

Para la predicción del futuro, los filtros de Kalman son recursivos y se basa en el estado de la actual (posición, velocidad, aceleración, etc), así como una suposición acerca de lo que las partes controlables intentan hacer para influir en la situación, para conseguir una medición de la realidad, la comparación de los dos (presente y futuro), moderar esta diferencia y ajustar la estimación de este valor moderado.

Es importante denotar que entre más se entienda este modelo matemático más preciso será el filtro en sus resultados. Si su modelo es totalmente coherente con lo que está sucediendo realmente, la estimación del filtro de Kalman convergerá con la realidad.

En cuanto a los modelos meteorológicos de predicción, son afectados por errores sistemáticos que pueden ser derivados, por ejemplo, en parametrización físicas (color del terreno, elevación del lugar, vegetación, etc) y en los datos que inicializan al modelo. Por esta razón muchos centros meteorológicos utilizan métodos de post-procesamiento tratando de filtrar ruidos y errores sistemáticos implementando algún tipo de método adaptivo de regresión dinámica (Chacón 2009).

Respecto a las temperaturas extremas, el Filtro de Kalman, sirve como método de ajuste a los modelos de pronóstico de temperaturas máximas y mínimas, para este caso el modelo GFS, permitiendo evaluar si las salidas del modelo meteorológico corresponden a los datos observados de la oficina de Alertas y Pronósticos de IDEAM.

Chacón A. (2009), en su estudio "Evaluación de las salidas de los modelos meteorológicos GFS y MM5 para las zonas astronómicas de Paranal, Macón y Pachón", identificaron la aplicabilidad del filtro de Kalman como método de ajuste, por medio de la corrección de los errores sistemáticos de las salidas de modelos meteorológicos GFS y MM5, donde evaluaron estadísticamente si las variables meteorológicas extraídas de las salidas de estos modelos con los datos registrados por las estaciones meteorológicas estaban relacionadas entre sí con su valor y comportamiento, encontrando el Filtro de Kalman como método de ajuste para sus series.

2. Aspectos metodológicos

La metodología aplicada en este trabajo, se basó principalmente en 3 pasos:

1. Aplicabilidad del filtro de Kalman como método de ajuste (corrección) de las salidas de temperaturas máximas y mínimas del GFS bajo los observados, a partir de la optimización sucesiva del coeficiente de actualización.
2. Aplicabilidad del filtro de Kalman como modelo de pronóstico GFS en temperaturas máximas y mínimas.
3. Implementación de Filtros de Kalman para el pronóstico de temperaturas máximas y mínimas del Modelo GFS para 42 ciudades de Colombia.

3

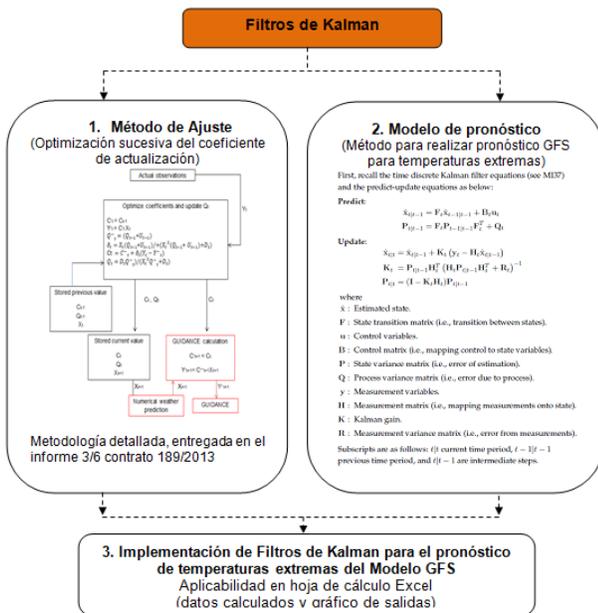


Figura 1. Diagrama de la aplicabilidad de Filtros de Kalman para temperaturas máximas y mínimas.

2.1 Análisis para la variable explicativa

Se partió por definir como variable explicativa, aquella que explica o influye sobre el cambio en la variable de respuesta (variable independiente). Para nuestro caso, se seleccionó las variables explicativas en relación con variables objetivas.

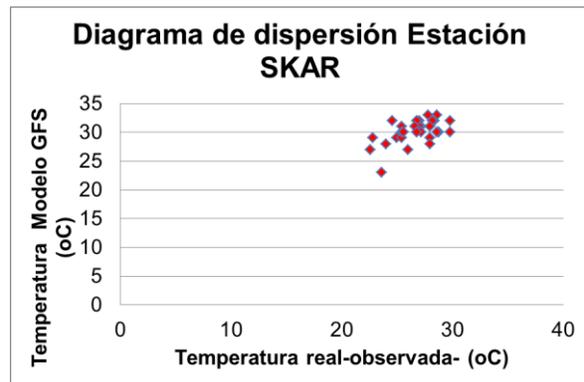
Por medio de un diagrama de dispersión (scatterplot), se indicó en el eje horizontal la variable explicativa (x), en nuestro caso las Temperaturas

reales, es decir las generadas por la oficina de pronóstico de IDEAM. En el eje vertical la variable de respuesta (y), las salidas del modelo GFS. Cada caso se representó por un punto (individuo)¹.

Diagrama de dispersión (scatterplot): Forma lineal

1) Asociación:

- Positiva (directa): valores por encima del promedio en una variable tienden a ir acompañados de valores por encima del promedio en la otra variable, y viceversa.
- Negativa (inversa): valores por encima del promedio en una variable tienden a ir acompañados de valores por debajo del promedio en la otra variable, y viceversa
- Fuerza de la relación: cuán cercanos están los puntos con respecto a la línea



Una anotación importante al seleccionar la variable explicativa es que cuando las observaciones están disponibles a mano, las últimas observaciones reales se pueden utilizar como variables explicativas (Ver referencia final de hoja), sobre todo, para hacer una predicción durante varias horas por venir, un buen uso de las observaciones reales determina si la predicción resulta correcta o incorrecta.

Relaciones entre dos variables de intervalo: Medida de asociación, correlación (r de Pearson).

Variables intervalo $-1 > r > 1$

-1: asociación perfecta (inversa)

0: no asociación

1: asociación perfecta

¹[http://metal.uniandes.edu.co/11%20-%20Relaciones%20entre%20dos%20variables%20\(metal\).pdf](http://metal.uniandes.edu.co/11%20-%20Relaciones%20entre%20dos%20variables%20(metal).pdf)

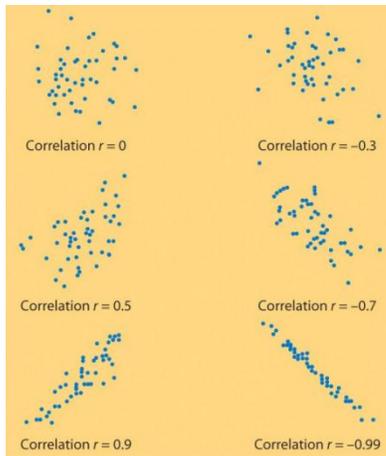


Figura 2. Comportamiento de correlación de Pearson

Importante al elegir una variable explicativa:

- Si la predicción no es exacta para un fenómeno que se produce en un patrón determinado, la clasificación puede emplearse. Como alternativa, también puede utilizar las variables explicativas adicionales para el patrón para mejorar la precisión.
- En la clasificación, los casos apropiados deben ser seleccionados con exactitud y objetividad.
- Es muy importante seleccionar con precisión los casos apropiados. Casos apropiados permiten obtener las variables explicativas más importantes y decir que la mitad de sus esfuerzos de desarrollo se ha finalizado correctamente.

2.2 Configuración de las variable explicativa en los parámetros requeridos por el filtro de Kalman²

D_t : Varianza del error de observación (Una gran varianza hace que sea difícil para los coeficientes que cambian.)

- Varianza del error de observación, que es relativa a la varianza del error de la predicción, se estima estadísticamente. Ejemplo: Cuando el RMSE (root mean square error o error cuadrático medio) de la temperatura máxima en determinado lugar es de 2 °C, el D_t , se considera aproximadamente 4 °C. En

² Yamagishi, Y. Introduction to the numerical weather prediction with emphasis on the application to the short term weather forecasting. Japan Meteorological Agency. Octubre 2003. Tokyo, Japan.

este caso, debido a que hay sólo pequeños cambios estacionales, que puede ser utilizado como un valor fijo (Yamagishi, Y, 2003).

- En la temporada de precipitación, por ejemplo, las varianzas de error son diferentes entre la temporada de menor precipitación y las temporadas de mayor precipitación (para el trópico). Por lo tanto, en el RMSE, los coeficientes se actualizan con el valor cuadrado de cada error y la media ponderada por el tiempo anterior.

Q_0 : Coeficiente de varianza del error inicial

- Varianza del error inicial de coeficientes óptimos a coeficientes reales. No hay ningún medio racional para estimarla.
- Asignar un valor determinado (un valor grande) Q_0 , probar el valor de los datos de los últimos meses (repetir la prueba, si es necesario), y obtener la Q cuando los cambios en Q se hacen más pequeños. Luego tome por Q_0 .
- En general, se incluyen dos o más variables explicativas. A continuación, Q es una matriz de covarianza.

$$Q = \begin{pmatrix} \sigma_{11} & \sigma_{12} & \cdots & \sigma_{1n} \\ \sigma_{21} & \sigma_{22} & \cdots & \sigma_{2n} \\ \cdots & \cdots & \cdots & \cdots \\ \sigma_{n1} & \sigma_{n2} & \cdots & \sigma_{nn} \end{pmatrix}$$

Cuando se utiliza este método para encontrar Q_0 , primero asignar un valor determinado a los componentes diagonales. A continuación, asigne 0 a otros componentes.

U_t : varianza de error del sistema (A gran diferencia hace que sea fácil para que los coeficientes cambien)

- Varianza de error utilizado para estimar los coeficientes para la próxima vez. No hay medios de agrupación racional para estimarla.
- Prácticamente, U puede ser utilizado sin ninguna consideración dada a tiempo.

• En general, se incluyen dos o más variables explicativas. Entonces U es también una matriz de covarianza.

• Por lo general, los errores del sistema para diferentes variables explicativas no están correlacionadas entre sí (covarianza = 0). Así que dar sólo componentes diagonales (U es diferente entre las distintas variables explicativas).

• U debe ser ajustado a un valor apropiado para la varianza del error inicial de cada coeficiente (Q₀) (Q₀/100, etc.) primero, y luego se ajusta mediante la evaluación de la precisión y la velocidad a la que se cambian los coeficientes (ajustado en un dígito -por dígitos base).

Notas importantes:

- 1) Cuando se cambian los coeficientes verdaderos (se cambia el mecanismo del desarrollo del fenómeno) en un ciclo corto: La configuración de U más grande tiene un efecto adverso sobre los cambios en los coeficientes, que resulta en errores más grandes. Ajustar un U más pequeño produce coeficientes medios. Para ello, debe utilizar la clasificación, agregar variables explicativas, o tomar otras medidas.
- 2) Cuando coeficientes verdaderos cambian mucho una vez y permanecen en ese estado por algún tiempo: Cuando U se fija más pequeño, el estado de cambio no puede ser atrapado con facilidad. U debe ser más grande.
- 3) Cuando los errores causados por los errores aleatorios de la predicción numérica del tiempo son significativos: Configuración del coeficiente de U más grandes genera causas para cambiar en gran medida cuando la predicción numérica del tiempo es "completamente equivocado". En este caso, U debe establecerse más pequeño.

Ejemplo, procedimiento para decidir Q₀ and U

$$U = f(x)Q_0$$

Donde f, actúa como el factor que configura el comportamiento del tamaño de U.

Por ejemplo:

(f= 1/10, 1/100, 1/1000, o 1/365, etc.)

Anotaciones:

- 1. Asignar un valor a Q₀ primero y ejecutar el filtro de Kalman con los datos de los últimos meses.
- 2. Tome Q obtenido finalmente por Q₀.
- 3. Repita los pasos 1 y 2 hasta que los cambios en Q₀ se hacen más pequeños.
- 4. Un cambio en f provoca un cambio en Q₀ (Configuración f causas más pequeñas U Para ser más pequeño, lo que resulta en una tasa más baja a la que se cambian los coeficientes). Repetir los pasos 1 a 3 utilizando diferentes valores para f, y decidir la F final mediante la adopción de la adecuación de los cambios en los coeficientes o RMSE de los valores previstos en consideración.

(Estas medidas tendrían una enorme cantidad de tiempo y energía sin necesidad de software de procesamiento automático. Estos pasos se pueden simplificar para las pruebas. Específicamente, puede seguir los pasos 1 a 3 con un determinado valor de f una vez y fijar Q₀. Para probar las tasas en que se modifican los coeficientes, se puede utilizar un valor fijo para Q₀ y cambiar f para cambiar U.)

Para este trabajo, consultar ApplyKalmanFilter_Linux, como herramienta ejecutable para realizar filtros de Kalman como método de ajuste a los modelos de pronóstico (GFS) de temperaturas máximas y mínimas³.

X_{0/0}: Los valores iniciales de coeficientes

• Para una guía de temperatura, donde los coeficientes cambian rápidamente y cada día, un valor de coeficiente inicial de 0 conducirán a coeficientes apropiados en unas dos semanas. Sin

³ Dorado, J; Ruíz, J. F. Manual de usuario ApplyKalmanFilter_Linux. Grupo de Modelamiento de Tiempo y Clima. Subdirección de Meteorología. IDEAM 2013.

embargo, para una guía precipitación, donde los coeficientes cambian lentamente y no todos los días, se debe repetir los cálculos con los datos de los últimos años para crear valores de los coeficientes iniciales poco precisos.

Verificación y monitoreo del coeficiente: Un sistema de filtro de Kalman obtiene observaciones reales asociados con cada predicción para actualizar los coeficientes. Estas observaciones reales permiten la verificación y el seguimiento de los coeficientes (modificación apropiada requiere esta verificación y supervisión).

- Seguimiento de predicción y observaciones reales, muestran los cambios en los valores pronosticados y observaciones reales en el tiempo (t) e identifica los errores.
- Seguimiento de los cambios en los coeficientes. Determine los coeficientes que causan errores en la predicción.
- Seguimiento de la contribución de las variables explicativas. Ver la contribución de las diferentes variables explicativas o valores de los coeficientes de las variables explicativas y determinar hasta qué punto afecta o contribuye a la predicción de los resultados.

2.3 Evaluación estadística

Se evaluaron los errores absolutos y error cuadrático medio, de los datos del modelo GFS y los datos observados, y los datos del GFS-K (GFS con filtro de Kalman) y los datos observados.

Para realizar la validación de las salidas de este modelo de predicción no existe en la actualidad un protocolo de validación. La mayoría de los estudios científicos realizan sus evaluaciones utilizando estadísticos estándar [White, et. al., 1999], los cuales son usados en este estudio. Los parámetros estadísticos son los siguientes:

1) Error Cuadrático Medio (RMSE), para obtener el error del pronóstico de las variables

meteorológicas simuladas por ambos modelos meteorológicos.

Para el cálculo de la precisión se utiliza el error cuadrático medio, definido como:

$$ECM = \sqrt{\frac{\sum_{i=1}^N (y_i - x_i)^2}{N}}$$

donde,

y_i : es el valor pronosticado de la variable meteorológica realizado por el modelo.

x_i : es el valor observado por la variable meteorológica registrada por la estación.

N: es el número total de datos analizados

2) Sesgo (BIAS), para observar la tendencia a subestimar o sobrestimar el pronóstico de las variables meteorológicas.

El sesgo (BIAS) nos proporciona información sobre la tendencia del modelo a sobreestimar o subestimar una variable, nos cuantifica el error sistemático de los modelos. Se define como:

$$BIAS = \sum_{i=1}^N \frac{(y_i - x_i)}{N}$$

donde,

y_i : es el valor pronosticado de la variable meteorológica realizado por el modelo.

x_i : es el valor observado por la variable meteorológica registrada por la estación.

N: es el número total de datos analizados

3) Correlación (Cor), para obtener la asociación lineal entre los datos simulados por los modelos meteorológicos y los datos medidos por las estaciones meteorológicas.

Calculamos la correlación entre los valores observados y valores pronosticados para encontrar si su relación es buena y determinar el desempeño de los modelos. La ecuación que define la correlación es:

$$Cor = \frac{\sum_{i=1}^N (x_i - \bar{x}) \cdot \sum_{i=1}^N (y_i - \bar{y})}{\sqrt{\sum_{i=1}^N (x_i - \bar{x})^2 \cdot \sum_{i=1}^N (y_i - \bar{y})^2}}$$

donde,

y_i es el valor pronosticado de la variable meteorológica realizado por el modelo.

x_i : es el valor observado por la variable meteorológica registrada por la estación.

\bar{y} es el promedio de la variable meteorológica pronosticada por el modelo.

\bar{X} : es el promedio de la variable meteorológica registrada por la estación meteorológica

N: es el número total de datos analizados

Aplicabilidad del Filtro de Kalman. Por lo mencionado anteriormente, se empleará una herramienta que permita disminuir errores correspondientes a las salidas del modelo GFS. Este filtro se podrá aplicar a las salidas directas de los puntos de grilla más cercano del modelo GFS donde no se encuentre el punto exacto de la toma del dato. Posterior a ello, los resultados obtenidos en la aplicabilidad del filtro en las series del modelo se compararán con los datos pronosticados.

2.4 Aplicabilidad del Filtro de Kalman como método de ajuste (corrección), a partir de la optimización sucesiva del coeficiente de actualización.

Metodología Filtros de Kalman: El sistema básico del Filtro Kalman

Ecuaciones básicas de la orientación por el filtro Kalman

$Y_t = C_t^{\wedge} X_t + v_t$ Relación entre los coeficientes y observaciones (1)

$C_t^{\wedge} = A C_{t-1}^{\wedge} + u_{t-1}$ Relación de coeficientes con coeficientes en el punto de tiempo anterior (2)

Y_t = Observaciones reales en el tiempo t (como la temperatura o la precipitación)

C_t^{\wedge} = coeficiente del vector verdadero en el tiempo t

X_t = Vector de variables explicativas derivadas de salida del modelo numérico

C_{t-1}^{\wedge} = Es cierto vector de coeficientes en el tiempo t-1

v_t = Errores debidos a errores de observación o el carácter incompleto de las ecuaciones de regresión, o ruido blanco.

u_t = Los errores del sistema (errores al estimar coeficientes para el siguiente punto de tiempo), o el vector de ruido blanco.

A= Matriz para estimar la cantidad del estado para el siguiente punto de tiempo

Aunque el valor del coeficiente cierto C_t^{\wedge} , es desconocido, el coeficiente de presumible en ese punto, C_t , se puede encontrar. El filtro de Kalman es un método para encontrarlo. El filtro de Kalman fue diseñado para utilizar las observaciones de series de tiempo (Y_0, Y_1, \dots, Y_t) para actualizar los coeficientes de C_t en diferentes puntos de tiempo a coeficientes más presumibles.

Flujo de sucesivas coeficiente de actualización (optimización sucesiva). En el tiempo t, el coeficiente presumible en el momento time t-1, C_{t-1} , es conocida y la variable explicativa en el momento t, X_t , se ha calculado a partir de la predicción numérica del tiempo en t-1. Entonces, las ecuaciones (1) y (2) pueden ser utilizados para obtener la primera estimación de C_t , C_t^- y la estimación de Y_t , Y_t^- , tal como se describe a continuación:

En primer lugar, calcula la primera estimación de C_t , C_t^- , de la ecuación (2). Debido u_t es desconocido en la ecuación, ignorar la cantidad que se muestra a continuación. Para obtener orientación, se puede asumir que los coeficientes no varían sistemáticamente con el tiempo y que $A = I$ (unidad de matriz).

$C_t^- = C_{t-1}$ (3)

Reemplace C_t en la ecuación (1) con la primera estimación de C_t , C_t^- , y calcular Y_t^- estimación como se muestra a continuación. Ignorar error v_t .

$Y_t^- = C_t^- X_t$ (4)

C_{t-1} : Coeficiente de presumible calculado a partir de observaciones en el momento t-1

C_t^- : Primera estimación del coeficiente en el tiempo t calcula a partir de coeficiente presumible C_{t-1}

Y_t^- : Estimación de Y_t en el tiempo t calculado desde la primera estimación del coeficiente C_t^- (valor predicho por la guía)

A continuación, corregir la primera estimación C_t^- de acuerdo a la diferencia entre la observación Y_t , para obtener un coeficiente más presumible C_t .

El filtro de Kalman está diseñado originalmente para actualizar un coeficiente sucesivamente a un coeficiente más presumible C_t .

C_t se actualiza como se muestra abajo. Esta es la ecuación de Kalman.

$$C_t = C_t^- + \partial_t(Y_t - Y_t^-) \quad (5)$$

∂_t ; Velocidad a la que se cambian los coeficientes (ganancia de Kalman)

Mediante la repetición de las ecuaciones (3), (4), y (5), C_t debe actualizarse sucesivamente.

El punto es como decir ∂_t .

(Generalmente, las ecuaciones de predicción contiene dos o más coeficientes. Sin embargo, para simplificar el proceso, la siguiente descripción utiliza las ecuaciones unidimensionales.

Aunque C_t es el coeficiente más presumible en ese punto, que tiene un cierto error con respecto al verdadero valor de C_t . Definir el error como e_t .

$$C_t^{\wedge} = C_t + e_t \quad : \text{Error } e_t \text{ es el ruido blanco.}$$

Error is expressed by the following equations:

$$e_t = C_t^{\wedge} - C_t \quad \text{Ecuación se supe en (5)}$$

$$e_t = C_t^{\wedge} - (C_t^- + \partial_t(Y_t - Y_t^-)) \quad \text{Sustituye (1) y (4)}$$

$$e_t = C_t^{\wedge} - (C_t^- + \partial_t(C_t^{\wedge} X_t + v_t - C_t^- X_t))$$

$$e_t = (1 - X_t \partial_t)(C_t^{\wedge} - C_t^-) - \partial_t v_t$$

$$e_t = (1 - X_t \partial_t)(C_t^{\wedge} - C_t^-) - \partial_t v_t$$

$$e_t = (1 - X_t \partial_t)(C_t^{\wedge} - C_t^-) - \partial_t v_t$$

$$e_t = (1 - X_t \partial_t)(u_{t-1} - e_{t-1}) - \partial_t v_t \quad (6)$$

Definir varianza Q_t en ecuación (6) de la siguiente manera:

$Q_t = \text{Var}(e_t)$; (Varianza de error del valor del coeficiente presumible)

La ecuación (6) puede ser utilizado para desarrollar Q_t de la siguiente manera:

$$Q_t = \frac{1}{n} \sum (e_t)^2$$

$$Q_t = (1 - X_t \partial_t)^2 (U_{t-1} + Q_{t-1}) + \partial_t^2 D_t$$

$$Q_t = \partial_t^2 (X_t^2 (U_{t-1} + Q_{t-1}) + D_t) - 2 \partial_t X_t (U_{t-1} + Q_{t-1}) + (U_{t-1} + Q_{t-1}) \quad (7)$$

Q_t es una ecuación bidimensional de δ_t . El δ_t que minimiza Q_t se puede obtener cuando el diferencial parcial debido a δ_t es 0.

$$2 \partial_t (X_t^2 (U_{t-1} + Q_{t-1}) + D_t) - 2 X_t (U_{t-1} + Q_{t-1}) = 0$$

$$\partial_t = X_t (U_{t-1} + Q_{t-1}) / (X_t^2 (U_{t-1} + Q_{t-1}) + D_t) \quad (8)$$

Suponiendo que Q_{t-1} es conocido, δ_t puede ser calculado a partir de la ecuación (8).

El Q_t se puede calcular mediante la sustitución de la ecuación (8) en la ecuación (7).

$$Q_t = (U_{t-1} + Q_{t-1}) - X_t^2 (U_{t-1} + Q_{t-1})^2 / (X_t^2 (U_{t-1} + Q_{t-1}) + D_t)$$

$$Q_t = D_t (U_{t-1} + Q_{t-1}) / (X_t^2 (U_{t-1} + Q_{t-1}) + D_t) \quad (9)$$

RESUMEN DEL SISTEMA DE FILTRO DE KALMAN

El sistema para actualizar los coeficientes requiere las siguientes ecuaciones:

$$C_t = C_{t-1} \quad (3)$$

$$Y_t = C_t X_t \quad (4)$$

$$\partial_t = X_t (U_{t-1} + Q_{t-1}) / (X_t^2 (U_{t-1} + Q_{t-1}) + D_t) \quad (8)$$

$$C_t = C_t^- + \partial_t (Y_t - Y_t^-) \quad (5)$$

$$Q_t = D_t (U_{t-1} + Q_{t-1}) / (X_t^2 (U_{t-1} + Q_{t-1}) + D_t) \quad (9)$$

1. En el tiempo t , C_{t-1} y Q_{t-1} optimizado en el tiempo $t-1$, and X_t obtenido a partir de la predicción numérica el tiempo en el momento $t-1$ están disponibles. Estos valores se pueden utilizar con las ecuaciones (3) y (4) para obtener coeficiente de estimación C_t y el cálculo de Y_t .

2. Una vez que se obtiene la observación real Y_t la ecuación (5) se puede utilizar con el obtenido a partir de la ecuación (8) para actualizar los coeficientes en el tiempo t (un coeficiente más presumible C_t se puede obtener).

- Al mismo tiempo, la ecuación (9) se utiliza para actualizar coeficiente de varianza del error de Q_t .
- El sistema de Filtro de Kalman repite este proceso para actualizar sucesivamente coeficiente C_t y el coeficiente de varianza de error Q_t que se utiliza para la optimización.
- Sin embargo, el valor inicial del coeficiente Q_0 , coeficiente inicial del error de la varianza de $Q=0$, error de la observación D_t , y el error del sistema U_t debe definirse por separado.

2.5. Aplicabilidad del Filtro de Kalman como modelo de pronóstico GFS en temperaturas máximas y mínimas.

Ecuaciones de Pronóstico y actualización: En primer lugar, recordar las ecuaciones de tiempo discretos de filtro de Kalman y las ecuaciones predicción-Actualización de la siguiente manera:

Pronosticado:

$$\hat{x}_{t|t-1} = F_t \hat{x}_{t-1|t-1} + B_t u_t$$

$$P_{t|t-1} = F_t P_{t-1|t-1} F_t^T + Q_t$$

Actualizaciones:

$$\hat{x}_{t|t} = \hat{x}_{t|t-1} + K_t (y_t - H_t \hat{x}_{t|t-1})$$

$$K_t = P_{t|t-1} H_t^T (H_t P_{t|t-1} H_t^T + R_t)^{-1}$$

$$P_{t|t} = (I - K_t H_t) P_{t|t-1}$$

donde,

x: estado estimado.

F: matriz de transición de estados (es decir, de transición entre dos estados).

u: Las variables de control.

B: Matriz de control (es decir, la asignación de controlar las variables de estado).

P: matriz de varianza del estado (es decir, el error de estimación).

Q: Proceso matriz de varianza (es decir, el error debido al proceso).

y: las variables de medición.

H: matriz de medición (es decir, la cartografía de las mediciones en el estado).

K: ganancia de Kalman.

R: medición de la varianza matriz (es decir, a partir de mediciones de error).

Subíndices son los siguientes: $t \setminus t$ período de tiempo actual, $t - s \setminus t - 1$ período de tiempo anterior, y $t | f \sim 1$ son pasos intermedios.

Se debe tener en cuenta que:

- El modelo es más suave que la medición de ruido, pero si hay un retraso en el valor actual, el sistema no está modelado correctamente
- La amplitud del filtro se hace más y más pequeño. Esto se debe a que el modelo está convergiendo lentamente al valor verdadero, en un nivel constante, todo lo cual es preciso con el tiempo.

3. Resultados

Las salidas del GFS⁴ y los datos del pronóstico⁵, correspondieron a 27 puntos distribuidos a nivel nacional, esta ubicación se presenta en la figura 3.

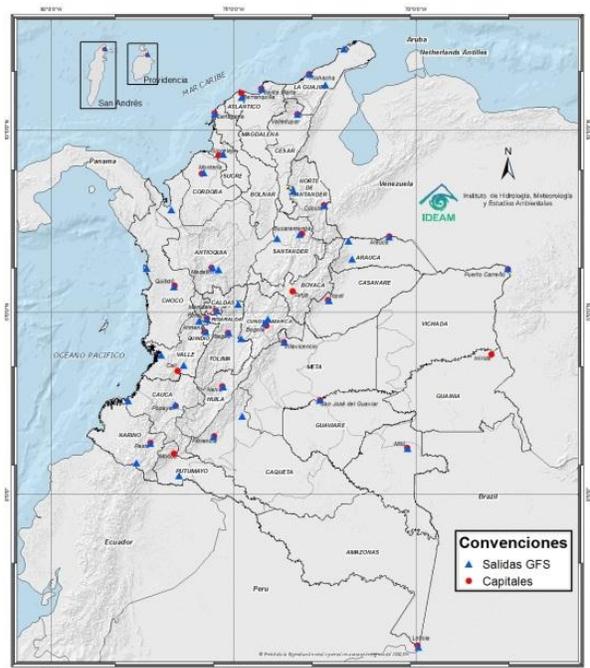


Figura 3. Distribución espacial de las salidas del modelo GFS y datos de pronóstico sobre Colombia

En la tabla 1 se presentan las ciudades

⁴ <http://www.ftp.ncep.noaa.gov/data/ncf/com/gfs/prod/>

⁵ Información Oficina de Alertas y Pronósticos. IDEAM

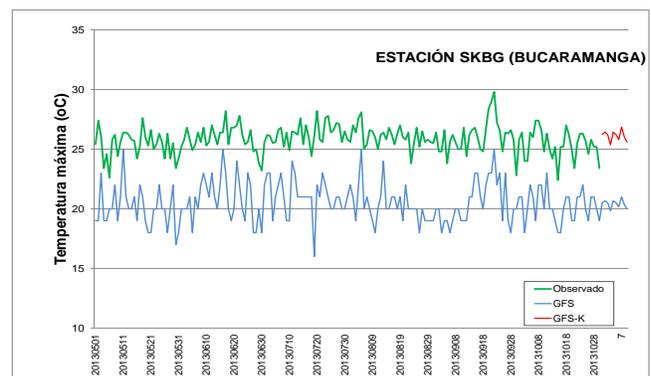
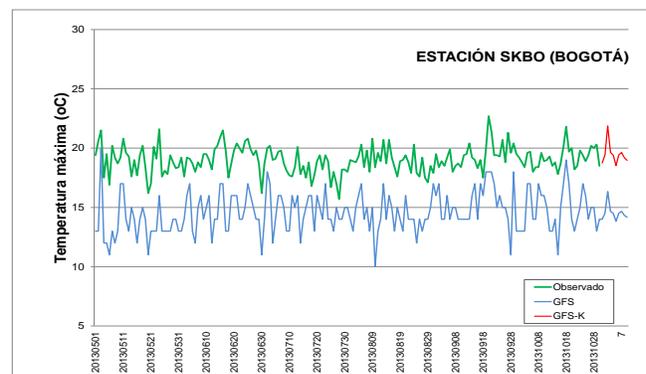
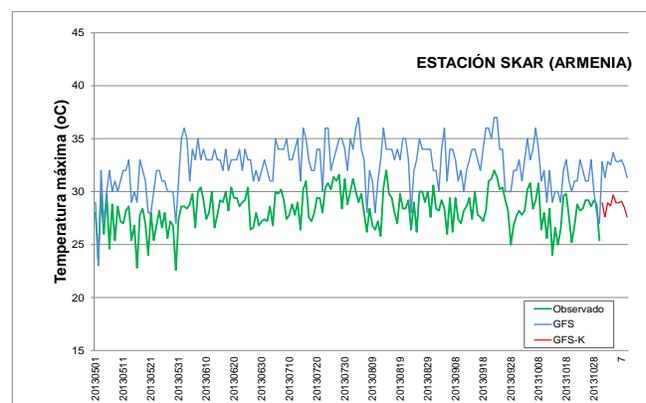
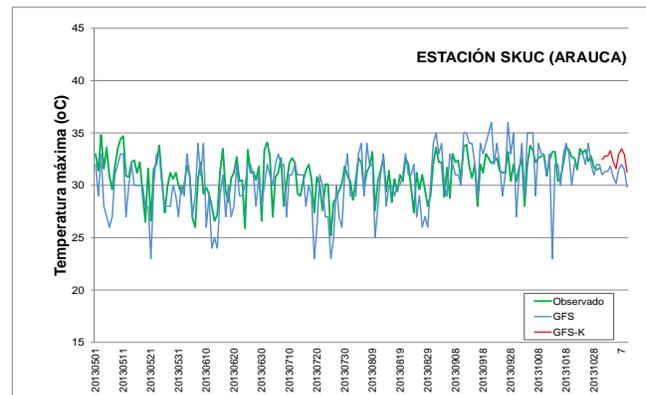
Tabla 1. Salidas pronóstico

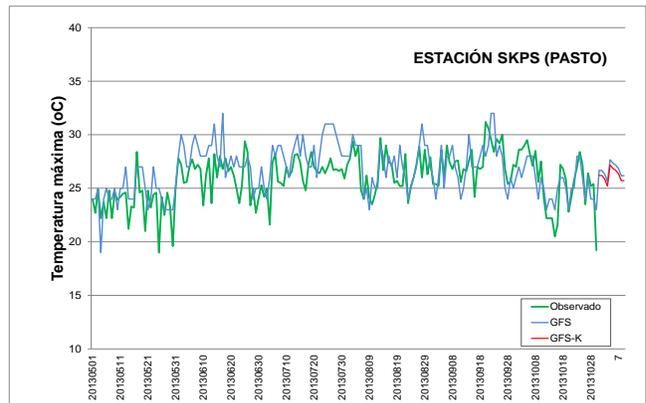
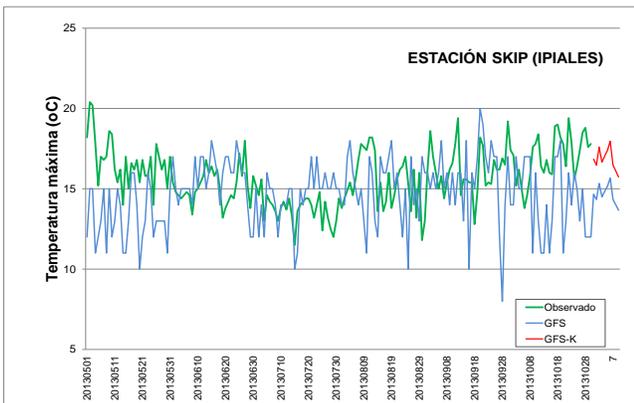
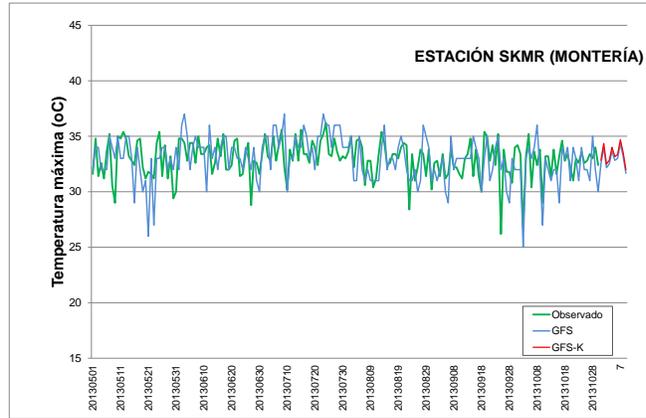
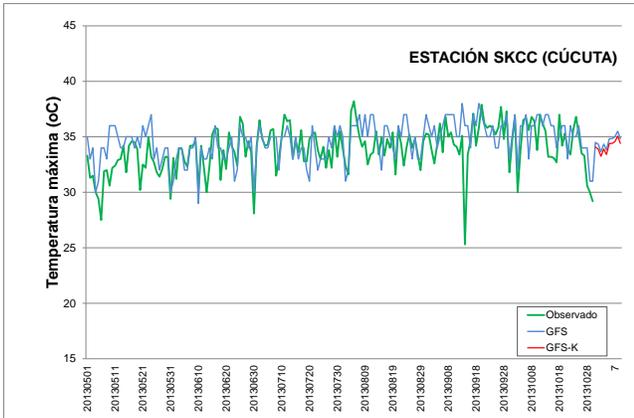
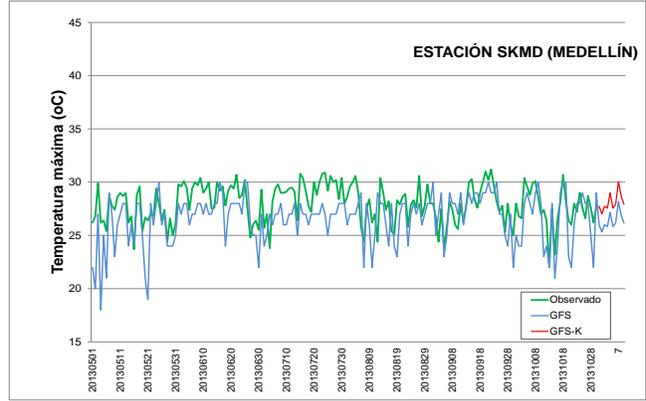
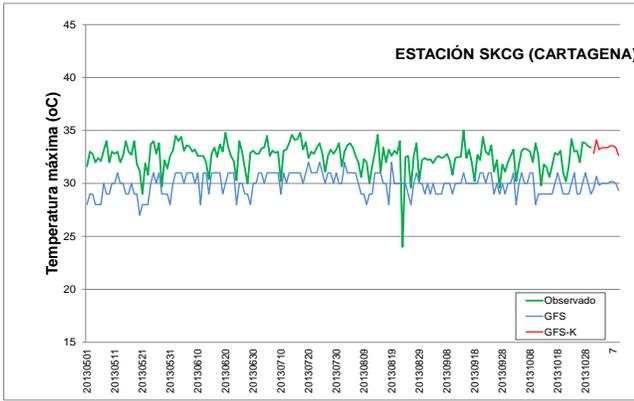
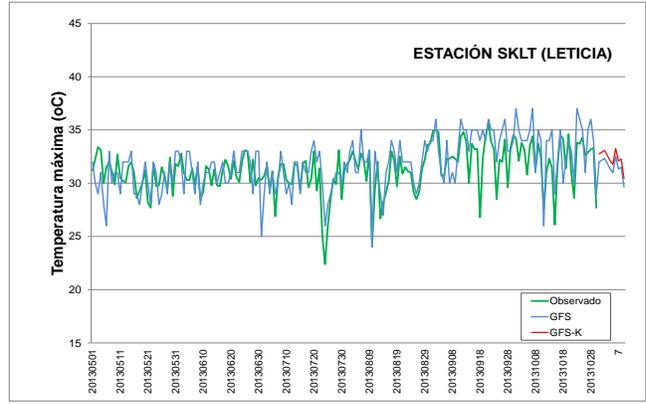
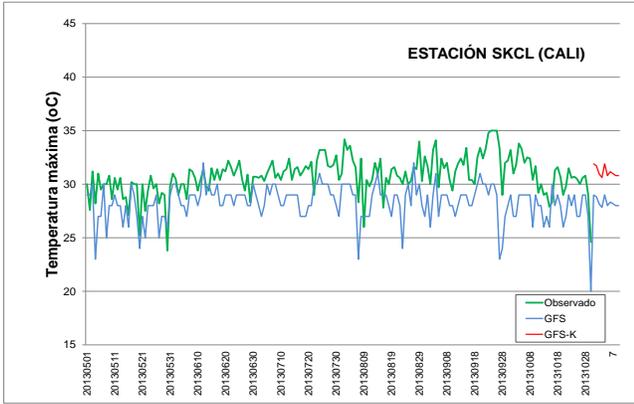
Ciudades donde se genera pronóstico		
Apartado	Cúcuta	Providencia
Arauca	Ibagué	Puerto Carreño
Armenia	Ipiales	Quibdó
Barrancabermeja	Leticia	Riohacha
Barranquilla	Medellín	Rionegro
Bogotá Eldorado	Montería	San Andrés
Bucaramanga	Neiva	Santa Marta
Cali	Pasto/Chachagüí	Valledupar
Cartagena	Pereira	Villavicencio

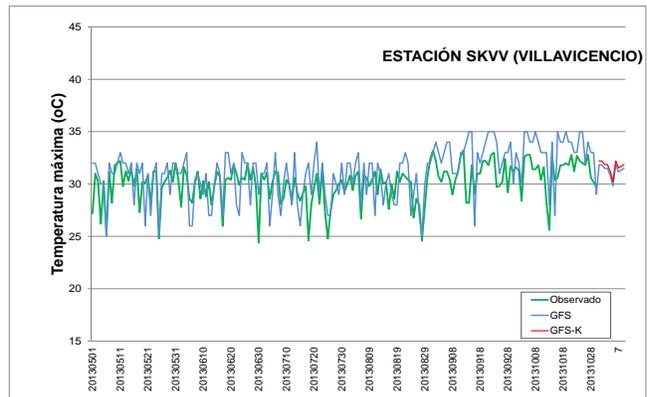
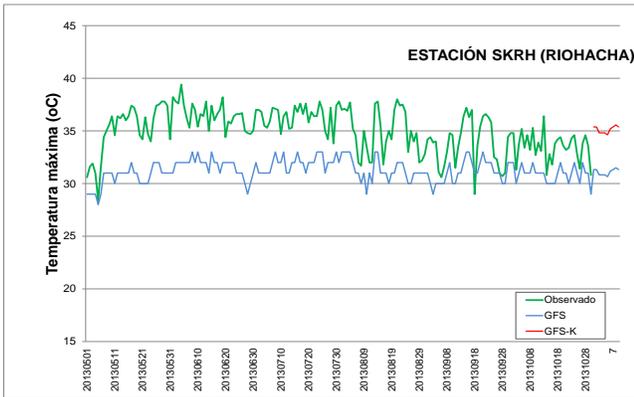
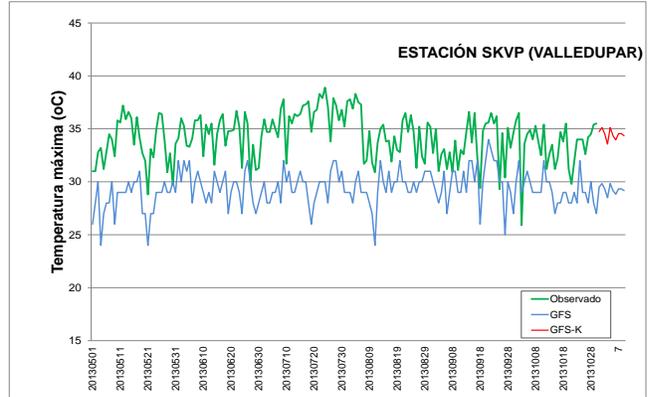
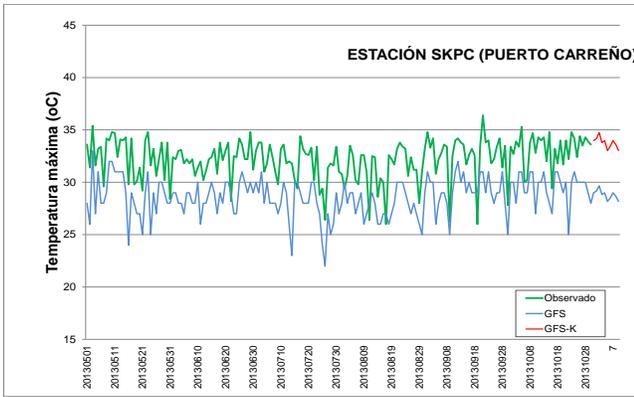
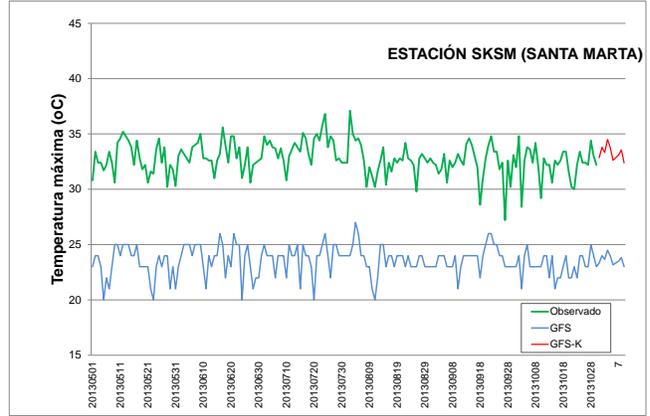
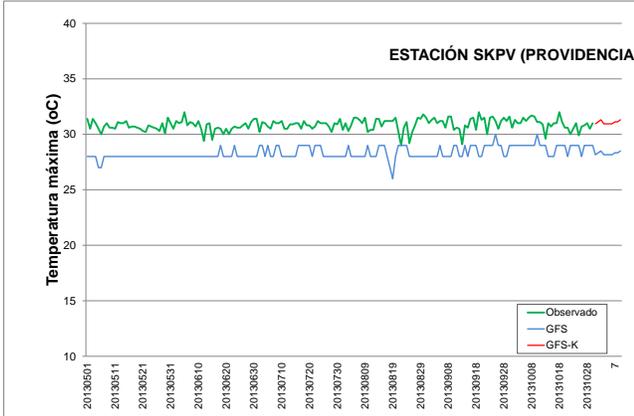
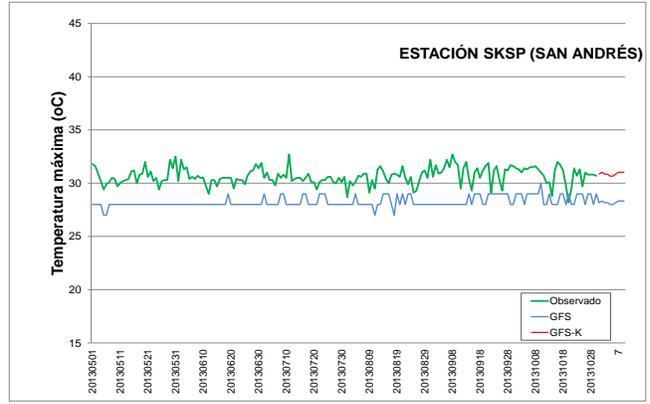
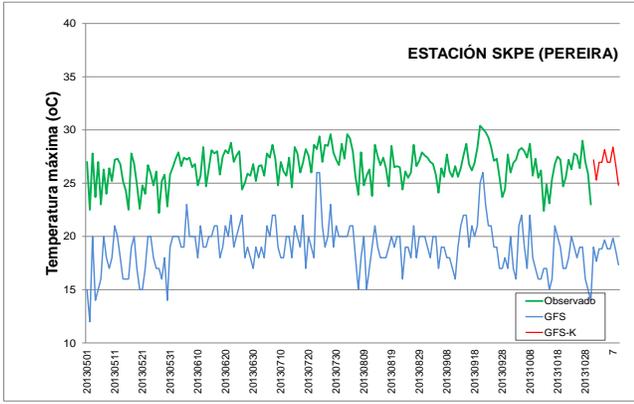
La implementación del método de Filtros de Kalman a los modelos de pronóstico GFS para temperaturas máximas y mínimas realizado en este trabajo, consistió en ajustar el pronóstico de las 27 estaciones (Información Oficina de Alertas y Pronóstico, IDEAM) con base en el pronóstico dado por el modelo GFS en el tiempo actual (t), y los históricos GFS y observados (t-1). Los resultados obtenidos al suponer datos de GFS, proyectaron el pronóstico corregido (de ahora en adelante GFS-K) para las 27 estaciones, ésta en forma similar a lo observado.

Las correlaciones tuvieron mejoras significativas, estas se calcularon entre los datos observados respecto al GFS y al nuevo GFS-K. En promedio pasaron de 0,65 a 0,97. A este procedimiento final se le agregó a la gráfica los valores de GFS-K y GFS (simulado) para ver el comportamiento del GFS pronosticado. Algunos de los resultados para temperatura máxima se presentan a continuación, donde verde son los datos observados (IDEAM), azul las salidas del modelo GFS y rojo las salidas del modelo GFS ajustado con Filtros de Kalman, de ahora en adelante GFS-K. Los datos simulados del GFS fueron en forma automática correspondiente al promedio de los primeros 10 días de la serie (x_1, x_2, \dots, x_{10}), esto para dar valores lógicos en el supuesto.

Temperaturas máximas:

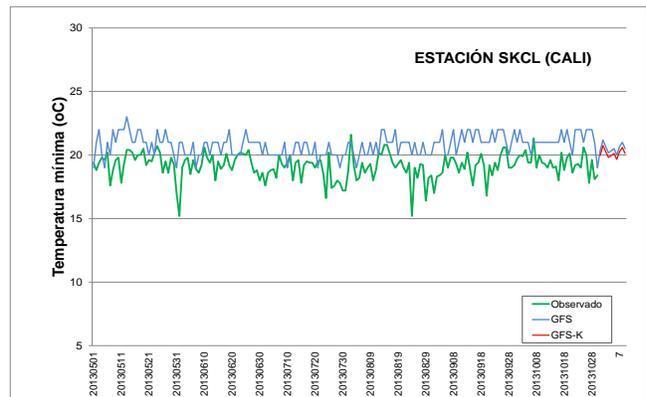
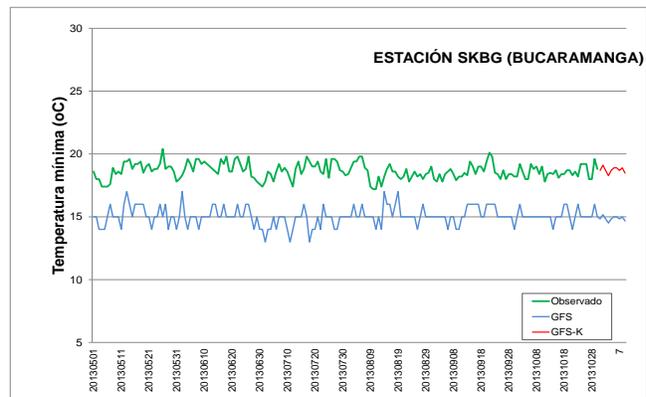
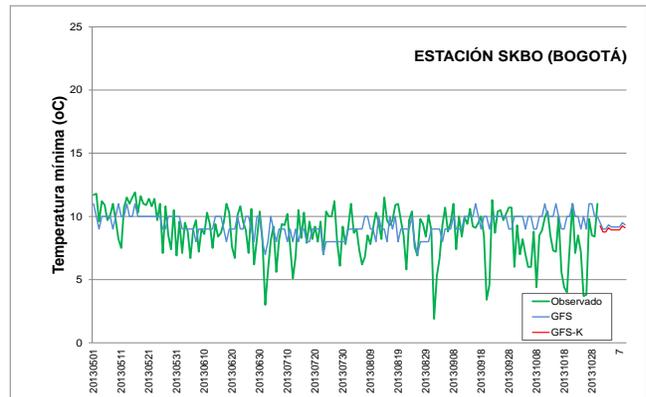
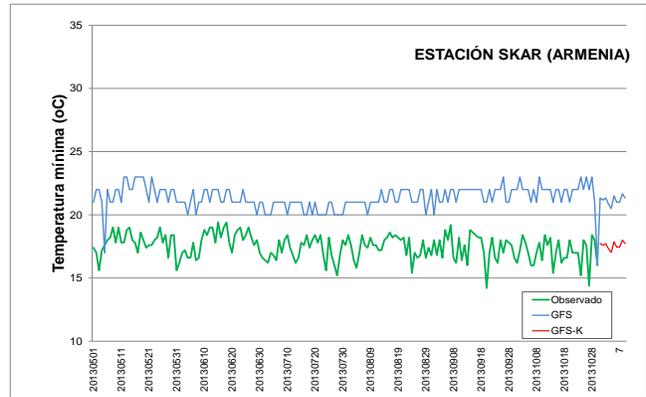




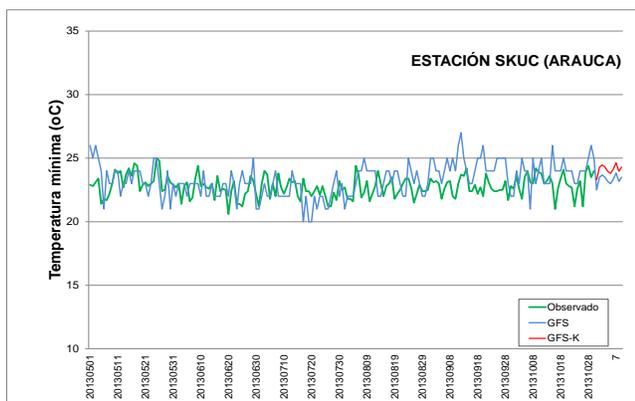


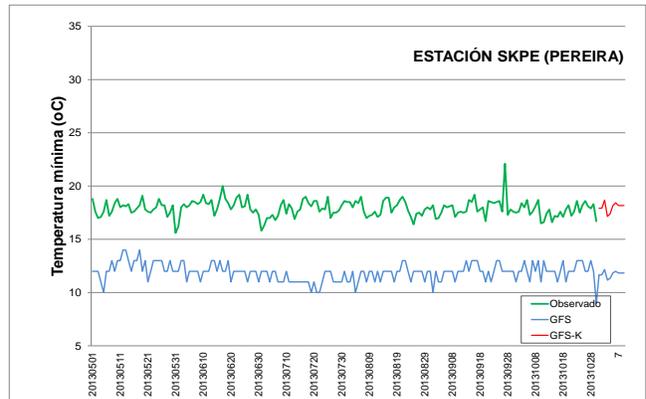
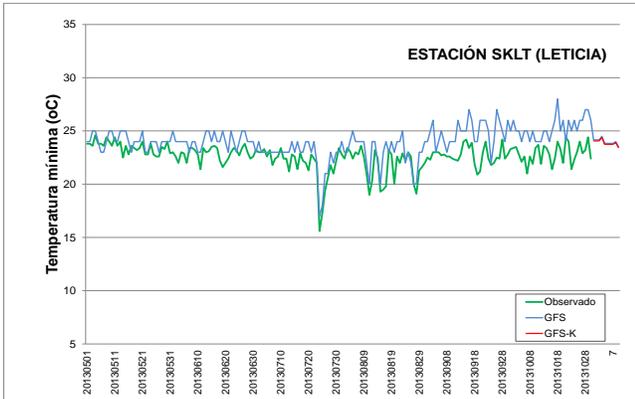
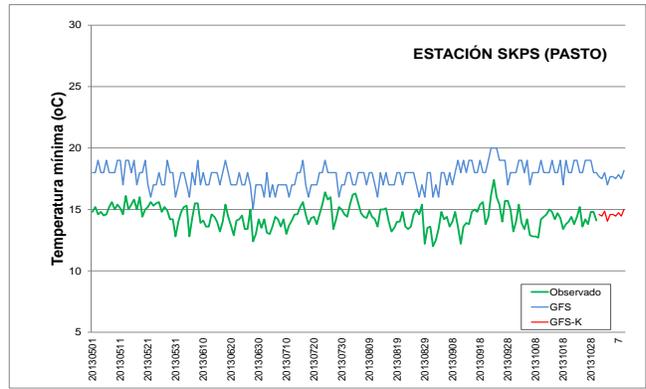
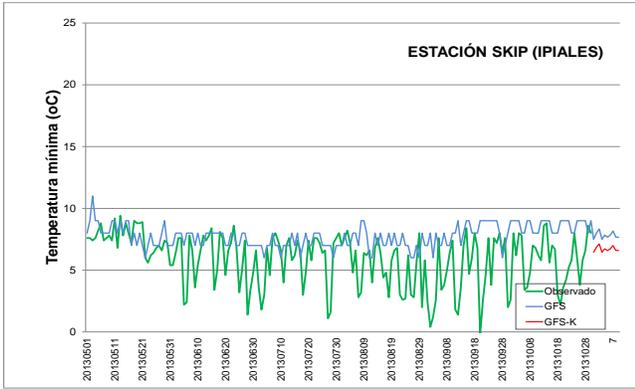
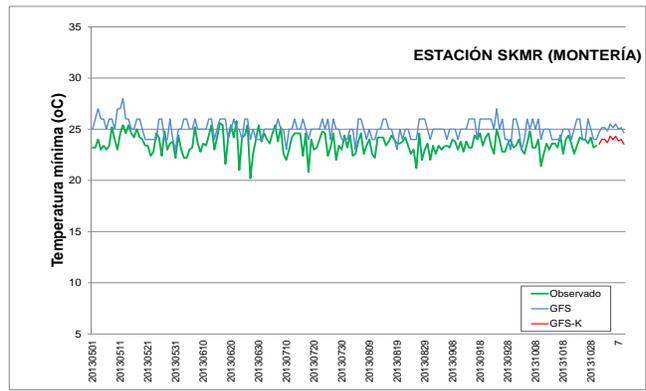
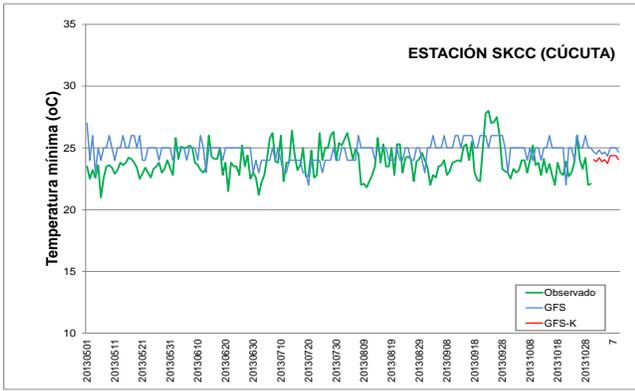
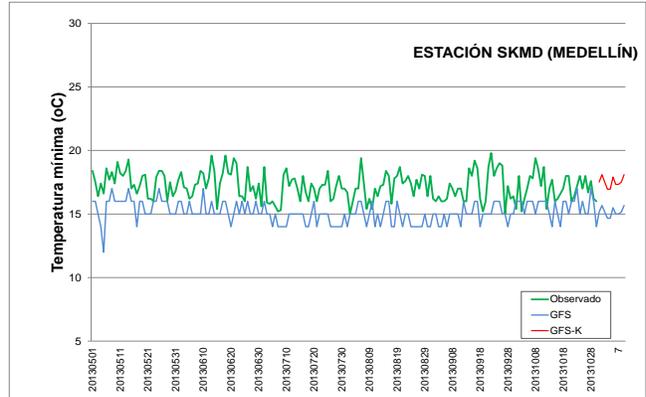
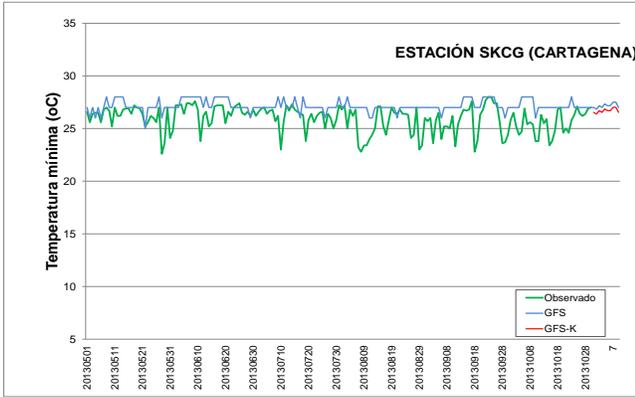
Correlaciones para tempeartura máxima			
Observado Arauca		Observado Montería	
GFS	0,60	GFS	0,48
GFS-K	0,99	GFS-K	0,98
Observado Armenia		Observado Pasto	
GFS	0,66	GFS	0,63
GFS-K	0,99	GFS-K	0,99
Observado Bogotá		Observado Pereira	
GFS	0,48	GFS	0,67
GFS-K	0,99	GFS-K	0,99
Observado Bucaramanga		Observado Providencia	
GFS	0,51	GFS	0,04
GFS-K	0,99	GFS-K	0,96
Observado Cali		Observado Pto Carreño	
GFS	0,55	GFS	0,59
GFS-K	0,99	GFS-K	0,98
Observado Cartagena		Observado Riohacha	
GFS	0,46	GFS	0,68
GFS-K	0,98	GFS-K	0,99
Observado Cúcuta		Observado San Andrés	
GFS	0,54	GFS	0,08
GFS-K	0,99	GFS-K	0,98
Observado Ipiales		Observado Santa Marta	
GFS	-0,05	GFS	0,47
GFS-K	0,99	GFS-K	0,99
Observado Leticia		Observado Valledupar	
GFS	0,70	GFS	0,64
GFS-K	0,99	GFS-K	0,99
Observado Medellín		Observado Villavicencio	
GFS	0,58	GFS	0,63
GFS-K	0,99	GFS-K	0,99

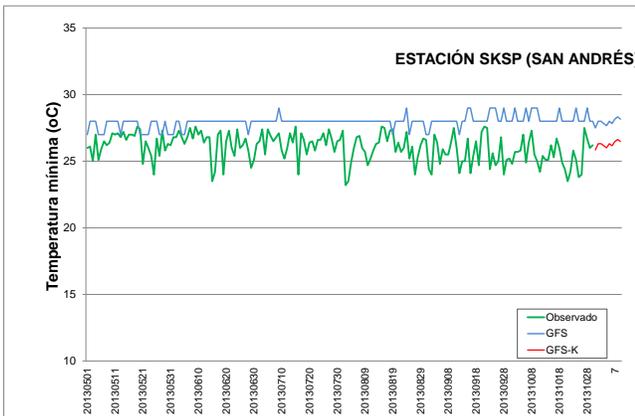
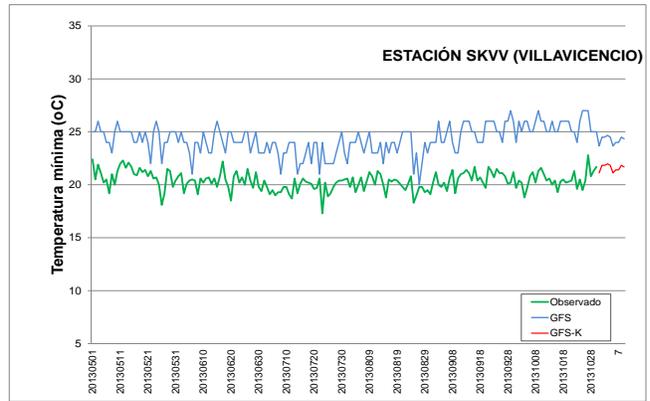
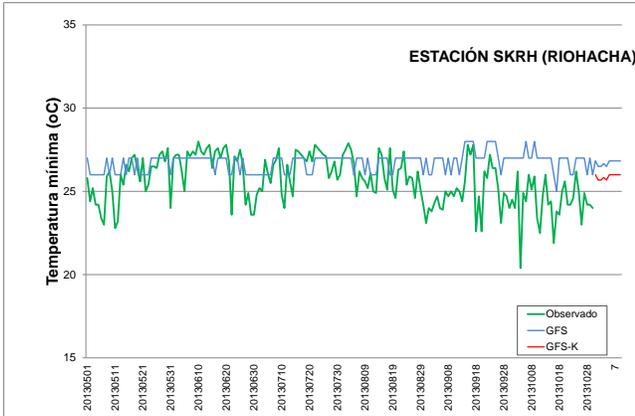
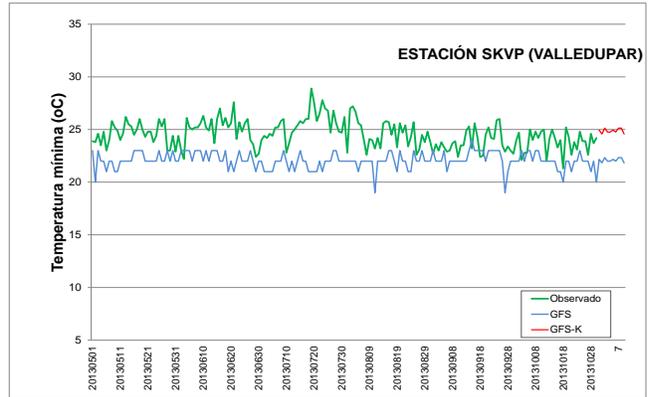
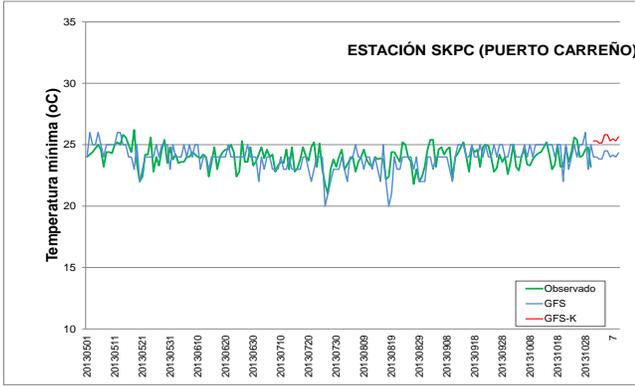
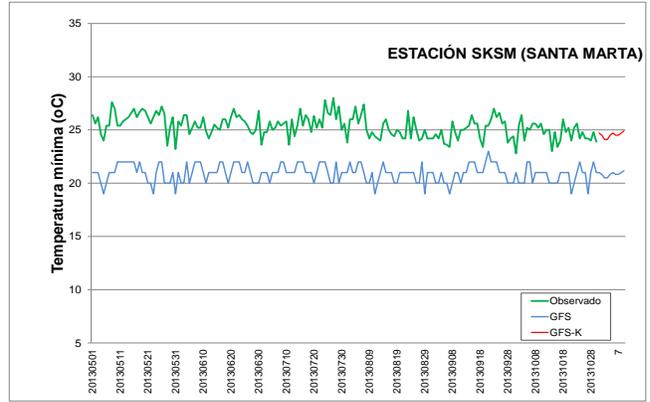
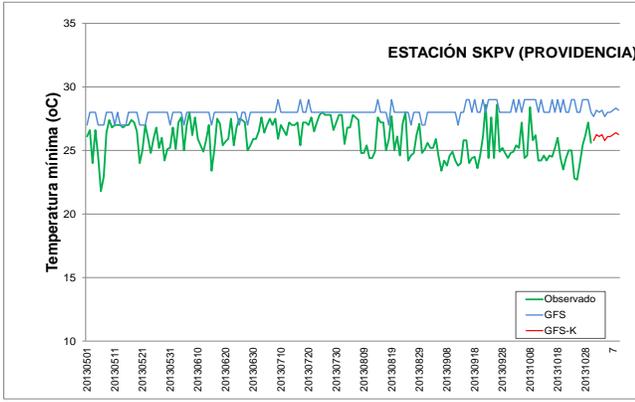
Figura 4. Correlaciones para temperaturas máximas



Temperaturas mínimas







Correlaciones para temperatura mínima			
Observado Arauca		Observado Montería	
GFS	0,24	GFS	0,30
GFS-K	0,98	GFS-K	0,98
Observado Armenia		Observado Pasto	
GFS	0,18	GFS	0,37
GFS-K	0,99	GFS-K	0,99
Observado Bogotá		Observado Pereira	
GFS	0,21	GFS	0,14
GFS-K	0,99	GFS-K	0,99
Observado Bucaramanga		Observado Providencia	
GFS	0,28	GFS	0,05
GFS-K	0,99	GFS-K	0,99
Observado Cali		Observado Pto Carreño	
GFS	0,30	GFS	0,42
GFS-K	0,99	GFS-K	0,99
Observado Cartagena		Observado Riohacha	
GFS	0,35	GFS	0,28
GFS-K	0,99	GFS-K	0,99
Observado Cúcuta		Observado San Andrés	
GFS	0,16	GFS	-0,08
GFS-K	0,99	GFS-K	0,99
Observado Ipiales		Observado Santa Marta	
GFS	0,13	GFS	0,34
GFS-K	0,99	GFS-K	0,99
Observado Leticia		Observado Valledupar	
GFS	0,61	GFS	0,10
GFS-K	0,99	GFS-K	0,99
Observado Medellín		Observado Villavicencio	
GFS	0,31	GFS	0,34
GFS-K	0,99	GFS-K	0,98

Figura 5. Correlaciones para temperatura mínima

El error cuadrático medio también mejoró al aplicar filtros de Kalman. A continuación se presentan los resultados del ECM para temperaturas mínimas y máximas.

Los valores de los errores en todas las estaciones está en promedio 0, lo que indicó con certeza la mejora en el ajuste.

Esto se presentó tanto para temperatura máxima como temperatura mínima.

En las figuras 6 y 7 se presentan los resultados de los errores cuadráticos medios.

Error cuadrático medio (ECM) para temperaturas máximas			
Arauca		Montería	
GFS vs OBS	2,30	GFS vs OBS	1,89
GFS-K vs OBS	0,40	GFS-K vs OBS	0,40
Armenia		Pasto	
GFS vs OBS	4,46	GFS vs OBS	2,07
GFS-K vs OBS	0,37	GFS-K vs OBS	0,37
Bogotá		Pereira	
GFS vs OBS	4,70	GFS vs OBS	7,80
GFS-K vs OBS	0,32	GFS-K vs OBS	0,37
Bucaramanga		Providencia	
GFS vs OBS	5,51	GFS vs OBS	2,63
GFS-K vs OBS	0,33	GFS-K vs OBS	0,30
Cali		Pto Carreño	
GFS vs OBS	2,99	GFS vs OBS	3,98
GFS-K vs OBS	0,36	GFS-K vs OBS	0,43
Cartagena		Riohacha	
GFS vs OBS	2,87	GFS vs OBS	4,11
GFS-K vs OBS	0,35	GFS-K vs OBS	0,35
Cúcuta		San Andrés	
GFS vs OBS	1,96	GFS vs OBS	2,58
GFS-K vs OBS	0,40	GFS-K vs OBS	0,30
Ipiales		Santa Marta	
GFS vs OBS	2,96	GFS vs OBS	9,35
GFS-K vs OBS	0,32	GFS-K vs OBS	0,36
Leticia		Valledupar	
GFS vs OBS	1,84	GFS vs OBS	5,36
GFS-K vs OBS	0,40	GFS-K vs OBS	0,41
Medellín		Villavicencio	
GFS vs OBS	2,37	GFS vs OBS	2,15
GFS-K vs OBS	0,36	GFS-K vs OBS	0,39

Figura 6. Error cuadrático medio para temperatura máximas

Error cuadrático medio (ECM) para temperaturas máximas			
Arauca		Montería	
GFS vs OBS	2,30	GFS vs OBS	1,89
GFS-K vs OBS	0,40	GFS-K vs OBS	0,40
Armenia		Pasto	
GFS vs OBS	4,46	GFS vs OBS	2,07
GFS-K vs OBS	0,37	GFS-K vs OBS	0,37
Bogotá		Pereira	
GFS vs OBS	4,70	GFS vs OBS	7,80
GFS-K vs OBS	0,32	GFS-K vs OBS	0,37
Bucaramanga		Providencia	
GFS vs OBS	5,51	GFS vs OBS	2,63
GFS-K vs OBS	0,33	GFS-K vs OBS	0,30
Cali		Pto Carreño	
GFS vs OBS	2,99	GFS vs OBS	3,98
GFS-K vs OBS	0,36	GFS-K vs OBS	0,43
Cartagena		Riohacha	
GFS vs OBS	2,87	GFS vs OBS	4,11
GFS-K vs OBS	0,35	GFS-K vs OBS	0,35
Cúcuta		San Andrés	
GFS vs OBS	1,96	GFS vs OBS	2,58
GFS-K vs OBS	0,40	GFS-K vs OBS	0,30
Ipiales		Santa Marta	
GFS vs OBS	2,96	GFS vs OBS	9,35
GFS-K vs OBS	0,32	GFS-K vs OBS	0,36
Letícia		Valledupar	
GFS vs OBS	1,84	GFS vs OBS	5,36
GFS-K vs OBS	0,40	GFS-K vs OBS	0,41
Medellín		Villavicencio	
GFS vs OBS	2,37	GFS vs OBS	2,15
GFS-K vs OBS	0,36	GFS-K vs OBS	0,39

Figura 7. Error cuadrático medio para temperatura mínimas

4. Conclusiones

Para temperaturas máximas y mínimas, el Filtro de Kalman se ajustó perfectamente, mejorando en un 90% las series.

Para temperaturas máximas, los coeficientes de correlación pasaron de 0.5 en promedio a un 0.99 y temperaturas mínimas de -0.08 a 0.99. Esto mostrando un comportamiento casi perfecto en el ajuste de las salidas del modelo GFS, llamado GFS-K, esto respecto a los observados.

En estaciones que se encuentran a nivel del mar, como por ejemplo, San Andrés y Providencia, Providencia, Santa Marta, Cartagena, el modelo no presentó buenos resultados, arrojando error en las series para pronóstico, sin embargo, al aplicarle el Filtro de Kalman (GFS-K), las series mejoraron significativamente.

Las series GFS-K, presentaron muy buena variabilidad respecto al observado, caso que no se logra solo con la salida del Modelo.

Los errores cuadráticos medios para temperaturas máximas y mínimas pasaron de 6,09 y 9,35 (serie sin ajustar) a 0,32 y 0,27 para la serie ajustada (GFS-K) respectivamente, resultados que confirman la efectividad del método.

Al estimar varios A (como función de estabilidad y generalidad del método) se denotó a 0 como función neutral, pero si la magnitud de diferencia entre lo observado y el modelo GFS son amplias, este valor es positivo (+) si el modelo es por encima o negativo (-) si el modelo está por debajo de la serie observada. La cantidad puede variar entre 1, 0, -1, respectivamente.

En cuanto al método de ajuste del filtro de Kalman empleado en este trabajo, es importante denotar su eficiencia y su bajo costo computacional, esto reflejado en la herramienta ApplyKalmanFilter_Linux, que permitió mantener el procedimiento en los datos con una mejor facilidad para el usuario que desee emplear esta aplicación en su información.

El filtro de Kalman es una herramienta capaz de tomar la información del error de entrenamiento y convertirla en un valor de corrección y pronóstico.

5. Recomendaciones

Como trabajo futuro se sugiere probar este filtro en temperaturas medias y realizar una metodología similar pero aplicada a precipitación.

Revisión en internet:

[http://metal.uniandes.edu.co/11%20-%20Relaciones%20entre%20dos%20variables%20\(metal\).pdf](http://metal.uniandes.edu.co/11%20-%20Relaciones%20entre%20dos%20variables%20(metal).pdf)

El método por su ganancia en resultados, vale la pena aplicarlo en un desarrollo computacional adsequible (fácil) para las personas que requieran hacer pronóstico a partir de los modelos GFS y las observaciones.

18

6. Referencias Bibliográficas

Andrews, Grewal. Kalman Filtering Theory And Practice Using Matlab. New York: Wiley, 2001

Application of Numerical Weather Prediction. Objective weather interpretation with direct use of numerical model output. Japan International Cooperation Agency Japan Meteorological Agency. 2003.

Chacón, A. C. Evaluación de las salidas de los modelos meteorológicos GFS y MM5 para las zonas astronómicas de Paranal, Macón y Pachón. Universidad de Valparaiso de Chile. 2009.

Deossa, P.A. Entrenamiento de Redes Neuronales con Filtros de Kalman. Universidad Nacional de Colombia, Sede Medellín. 2009.

Dorado, J; Ruíz, J. F. Manual de usuario ApplyKalmanFilter_Linux. Grupo de Modelamiento de Tiempo y Clima. Subdirección de Meteorología. IDEAM 2013.

Haykin, Simon. Neural Networks. A comprehensive Foundation. Pearson, India, 1999.

Yamagishi, Y., 2003. Introduction to the numerical weather prediction with emphasis on the application to the short term weather forecasting. Japan Meteorological Agency. Octubre 2003. Tokyo, Japan.