

Uso de métodos predictivos en estimaciones de capacidad en redes móviles

Use of predictive methods for capacity estimations in mobile networks
Use of predictive methods for capacity estimations in mobile networks

Ing. Ignacio Hernández Gutiérrez^{1*}, Ing. Lidia Hidalgo Rodríguez², Ing. Brian Jofre de Villegas Biosca³

Recibido: 02/2021 | Aceptado: 07/2021

Palabras clave

Métodos predictivos
Pronósticos
Capacidad de redes
Holt-Winters
Método ARIMA

Resumen

Las redes de telecomunicaciones tienen comportamientos complejos, debido a la gran heterogeneidad que existe en la actualidad. La penetración de dispositivos móviles ha generado un crecimiento exponencial de tráfico de datos que circula a diario en las diferentes redes de comunicaciones, lo que conlleva a que los operadores de servicio se mantengan en una constante revisión de la capacidad de sus enlaces. Tener la capacidad de evaluar las velocidades máximas y promedios alcanzadas, y a partir de ahí realizar un pronóstico de aquellos parámetros que tienen mayor impacto en los dimensionamientos necesarios de una red de telecomunicaciones resulta vital. Esto influye en gran medida en la toma de decisiones para los procesos de inversión y operación, y mantenimiento de la red, lo cual necesariamente tiene como resultado una correcta evaluación de los fondos destinados a tal efecto. Este artículo tiene como objetivo central, aplicar algunos de los métodos de pronósticos a diferentes parámetros que impactan en aquellas variables que determinan las necesidades de un crecimiento de una red a corto o mediano plazo.

Key words

Predictive methods
Forecasting
Network capacity
Holt-Winters
ARIMA method

Abstract

Telecommunication networks are complex regarding performance, due to the great heterogeneity that currently exists. The mobile device penetration has brought about an exponential growth of data traffic flowing daily in the different communication networks, which leads service operators to constantly review their link capacity. It is vital to be able to assess the maximum and average speeds reached, and then make a forecast of those parameters having the greatest impact on the necessary dimensioning of a telecommunications network. This greatly influences the decision making process for investment, operation and maintenance of the network, which necessarily results in a correct evaluation of the funds earmarked for this purpose. The

1 Vicepresidencia de Operaciones. Empresa de Telecomunicaciones de Cuba S.A., ETECSA. ignacio.hernandez@etecsa.cu

2 Vicepresidencia de Operaciones. Empresa de Telecomunicaciones de Cuba S.A., ETECSA. lidia.hidalgo@etecsa.cu

3 Vicepresidencia de Operaciones. Empresa de Telecomunicaciones de Cuba S.A., ETECSA.brian.jofre@etecsa.cu

main objective of this article is to apply some of the forecasting methods to different parameters which impact on those variables deciding the needs of a network growth in the short or medium term.

Introducción

El desarrollo de las tecnologías de comunicaciones en los últimos años por un lado, y la variedad y cantidad de servicios de valor agregado dentro de las redes móviles (MVAS), ha potenciado de tal forma el proceso de convergencia en el mundo digital hoy en día, que resulta impensable hablar de un dispositivo móvil que no tenga acceso a redes de datos o Internet propiamente dicha.

El crecimiento de usuarios a nivel global con acceso a dispositivos inteligentes es vertiginoso, especialmente en países que han llegado un poco más tarde a la revolución del acceso a Internet por tecnología móvil de Banda Ancha, especialmente la llamada tecnología 4G (LTE y sus variantes). Esto hace que para las empresas operadoras de servicios de telecomunicaciones resulte fundamental poder predecir cuánto crecerá el acceso a los servicios de datos o de voz dentro de sus redes.

La existencia de variados métodos de predicción dificulta el proceso de selección de aquellos que puedan ser utilizados en análisis para predecir los parámetros que, por sí solos, prácticamente definen las dimensiones de la red de telecomunicaciones. Estos parámetros se pueden circunscribir básicamente en dos elementos, las velocidades máximas alcanzadas y la cantidad de usuarios concurrentes dentro de una red.

La Empresa de Telecomunicaciones de Cuba S.A., opera una red heterogénea donde conviven las tecnologías móviles de segunda generación en adelante (GSM, WCDMA y LTE), con usuarios en cada una de las mismas, lo cual añade otro nivel de dificultad en el tema de su dimensionamiento. Resulta necesario en el proceso de Operación y Mantenimiento de esta, contar con un mecanismo que permita predecir estas variables de la manera más cercana a lo que sucederá en el futuro, para poder ajustar y equilibrar de forma correcta los gastos efectuados por concepto de crecimiento en la red.

Este artículo pretende evaluar distintos métodos predictivos con el objetivo de recomendar el uso de alguno de ellos para la predicción eficiente de los crecimientos de las variables velocidad máxima alcanzada en la red y cantidad de usuarios concurrentes, dentro de las especificidades del tráfico existente en Cuba.

Materiales y métodos

Se ha llevado a cabo un estudio de las series de tiempo relacionadas a la velocidad máxima de transmisión de datos y la cantidad de usuarios concurrentes en la red 4G de ETECSA, con el objetivo de extraer de la misma característica tales como Tendencia, Estacionalidad y Nivel, para evaluar distintos modelos de pronóstico.

El comportamiento de estas variables es fundamental para la toma de decisiones en cuanto a inversiones, gestión dinámica de tráfico o implementación de algún servicio en la red, poseer de antemano el mejor modelo para su previsión redundante en la disminución de tiempo en distintos análisis.

Los datos que componen estas series de tiempo, son obtenidos de los sistemas de gestión de las distintas tecnologías instaladas, que posteriormente son almacenados en diferentes bases de datos (RANKPI) del Departamento de Tráfico y Optimización de la Dirección de Operación y Mantenimiento de la Vicepresidencia de Operaciones.

En esta investigación se evaluaron los métodos de pronósticos de suavizado exponencial y de regresión (ARIMA) para las series descritas, los cuales son de tipo estadístico-matemático, los mismos están implementados en el software RStudio.

Resultados y discusión

Una serie de tiempo, es una secuencia de observaciones ordenadas en el tiempo, su estudio tiene como objetivo central desarrollar modelos estadísticos que expliquen el comportamiento de una variable aleatoria que varía con el tiempo, que permiten estimar pronósticos futuros de dicha variable aleatoria (Garzón y Wanumen, 2014). Por ello, el manejo de las Series de Tiempo es de vital importancia en los procesos de planeación y en aquellas áreas del conocimiento donde sea necesario evaluar el efecto de una política basada sobre una variable.

Pronosticar significa predecir valores futuros de una variable aleatoria basándose en el estudio de la estructura definida por las observaciones pasadas de variables que explican su variación, suponiendo que la estructura del pasado se conserva en el futuro (Garzón

y Wanumen, 2014). Es importante anotar, que existen factores o eventos que influyen en la calidad de un pronóstico. Estos factores pueden ser internos o controlables y externos o incontrolables. El éxito de la planeación depende de estos factores o eventos.

La elección de un método de pronóstico depende en gran medida de las características mostradas por la serie de tiempo, las cuales se enumeran a continuación:

Componente tendencia

Se puede definir como un cambio a largo plazo que se produce en la relación al nivel medio, o el cambio a largo plazo de la media. La tendencia se identifica con un movimiento suave de la serie a largo plazo.

Componente estacional

Muchas series temporales presentan cierta periodicidad o dicho de otro modo, variación de cierto periodo (semestral, mensual, etc.). Estos efectos son fáciles de entender y se pueden medir explícitamente, o incluso, se pueden eliminar de la serie de datos, a este proceso se le llama desestacionalización de la serie.

Componente aleatorio

Componente que no responde a ningún patrón de comportamiento, sino que es el resultado de factores fortuitos o aleatorios que inciden de forma aislada en una serie de tiempo.

Modelos de suavizado

El suavizado exponencial de series de tiempo tiene su fundamento sobre la base del método de promedios móviles ponderados, en este caso asigna ponderaciones decrecientes exponencialmente para las observaciones más nuevas a las más antiguas (Villarreal, 2016).

El suavizado exponencial se usa generalmente para hacer pronósticos a corto plazo, ya que los pronósticos a más largo plazo que utiliza esta técnica pueden ser poco confiables.

Existen varios tipos de suavizado. El suavizado exponencial simple usa un promedio móvil ponderado con pesos decrecientes exponencialmente. El suavizado exponencial doble con corrección de tendencias de *Holt* suele ser más confiable para manejar datos que muestran tendencias, en comparación con el procedimiento único. El suavizado exponencial triple (también llamado *Holt-Winters* multiplicativo o aditivo) suele ser más confiable para tendencias parabólicas o datos que muestran tendencias y estacionalidad.

Por las características de las series de estudio uno de los métodos a aplicar será el Suavizado exponencial triple, cuyas ecuaciones se muestran a continuación.

El modelo aditivo es:

$$\begin{aligned} L_t &= \alpha (Y_t - S_{t-p}) + (1 - \alpha) [L_{t-1} + T_{t-1}] \\ T_t &= \gamma [L_t - L_{t-1}] + (1 - \gamma) T_{t-1} \\ S_t &= \delta (Y_t - L_t) + (1 - \delta) S_{t-p} \\ P_t &= L_{t-1} + T_{t-1} + S_{t-p} \end{aligned}$$

Donde:

L_t Es el nivel en el tiempo t , α es la ponderación para el nivel

T_t Es la tendencia en el tiempo t

γ Es la ponderación para la tendencia

S_t Es el componente estacional en el tiempo t

δ Es la ponderación para el componente estacional

p Es periodo estacional

Y_t Es el valor de los datos en el tiempo t

Modelos de regresión

Se basa en la idea de que los valores presentes de una serie de tiempo pueden ser explicados como una función de p valores pasados, ponderados por un factor ϕ_j , conocido como el coeficiente autorregresivo, donde p determina el número de pasos necesitados en el pasado para predecir el valor actual; sumados a una función que representa las perturbaciones aleatorias del sistema (e_t) y que se modela como ruido blanco; este modelo se denota como AR(p) donde p es su orden (Brockwell y Davis, 2006), (Cryer y Chan, 2008), (González, 2009).

$$X_t = e_t + \Phi_1 X_{t-1} + \Phi_2 X_{t-2} + \dots + \Phi_p X_{t-p}$$

Donde:

X_t : Variable dependiente en el periodo t

Φ_1 : Coeficiente de orden 1 del modelo

X_{t-1} : Variable dependiente con un retardo

La elección de un modelo AR(p) se efectúa cuando se observa que los coeficientes de autocorrelación simple decrecen exponencialmente y además las correlaciones parciales reflejan picos grandes a lo largo de p rezagos.

En el modelo de medias móviles MA(q), una determinada observación está determinada por los impulsos aleatorios de las observaciones anteriores, el

proceso MA(q) tiene memoria corta, y absorbe rápidamente los impactos. La ecuación que la caracteriza es:

$$Y_t = a_t - \Theta_1 a_{t-1} - \Theta_2 a_{t-2} - \dots - \Theta_q a_{t-q}$$

Donde:

- Y_t : Variable dependiente para el periodo t
- Θ_1 : Coeficiente de orden 1 del modelo
- a_{t-1} : Residuo o error del periodo t-1
- a_t : Residuo o error del periodo t

Modelo autorregresivo integrado de medias móviles de orden p,d,q – ARIMA(p,d,q): es la combinación de los modelos autorregresivos y el de medias móviles, con la particularidad de incluir un proceso de restablecimiento (el cual se denomina integración) de inestabilidad original presente en una serie de tiempo. La ecuación que la caracteriza es la suma de las ecuaciones descritas anteriormente.

$$Y_t = e_t + \Phi_1 X_{t-1} + \Phi_2 X_{t-2} + \dots + \Phi_p X_{t-p} - \Theta_1 a_{t-1} - \Theta_2 a_{t-2} - \dots - \Theta_q a_{t-q}$$

La Unión Internacional de Telecomunicaciones en su recomendación E.507, define un grupo de modelos de pronósticos para series temporales que pueden emplearse para cálculos en entornos de redes de comunicaciones, específicamente para el tráfico, las variables de este estudio, tiene un comportamiento estadístico muy relacionado con el tráfico (UIT, Recomendación

E.507, en vigor). Los principales son los métodos de suavizado y autorregresivos.

Como se había planteado con anterioridad, los parámetros que se analizarán en este artículo son:

Cantidad de usuarios concurrentes en la red

Velocidad máxima alcanzada

Para esto se utilizarán los datos máximos de cada día correspondientes a los usuarios de la red LTE de ETECSA, pues representan el mayor peso en velocidades alcanzadas dentro de la red.

La muestra tomada corresponde con el período del 1ro de Noviembre de 2020, al 10 de febrero de 2021, para obtener el pronóstico correspondiente a los próximos 15 días dentro de la red.

La figura 1 muestra el comportamiento de la velocidad máxima alcanzada en la red en el periodo seleccionado.

La visualización de la serie no permite detallar características esenciales para determinar la posibilidad de uso de uno u otro modelo de predicción. Para esto es necesario la descomposición de las series pues posibilita determinar características como estacionalidad (*season*), tendencia (*trend*) y nivel (*level*). La figura 2 muestra la serie de velocidad máxima alcanzada descompuesta en dichos parámetros.

Como se observa en la figura 2, la serie posee los 3 componentes que resultan necesarios para una evaluación del método a utilizar, estacionalidad, tendencia y nivel.

La otra serie que se estudiará corresponde con la cantidad de usuarios 4G concurrentes en la red. La mis-

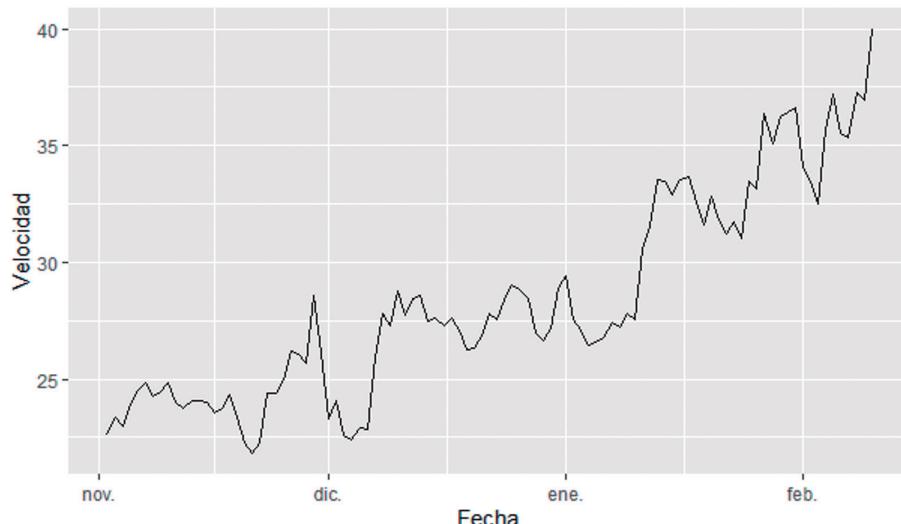


Figura 1. Comportamiento de la velocidad máxima diaria en la red 4G (en Gbps)

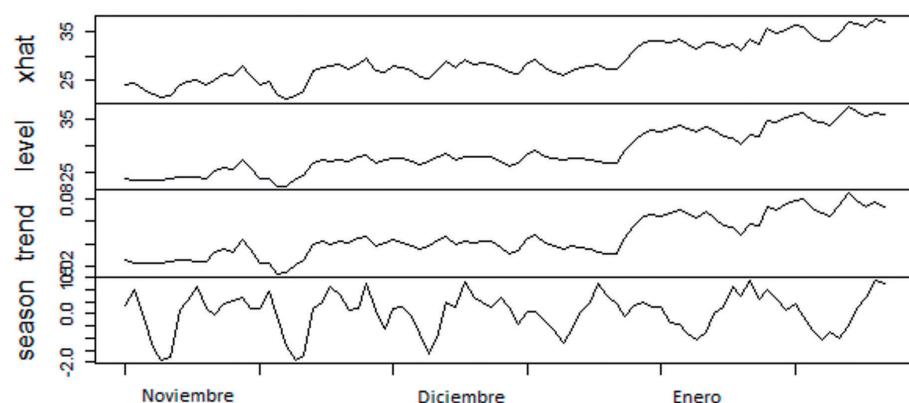


Figura 2. Descomposición de la serie velocidad máxima alcanzada

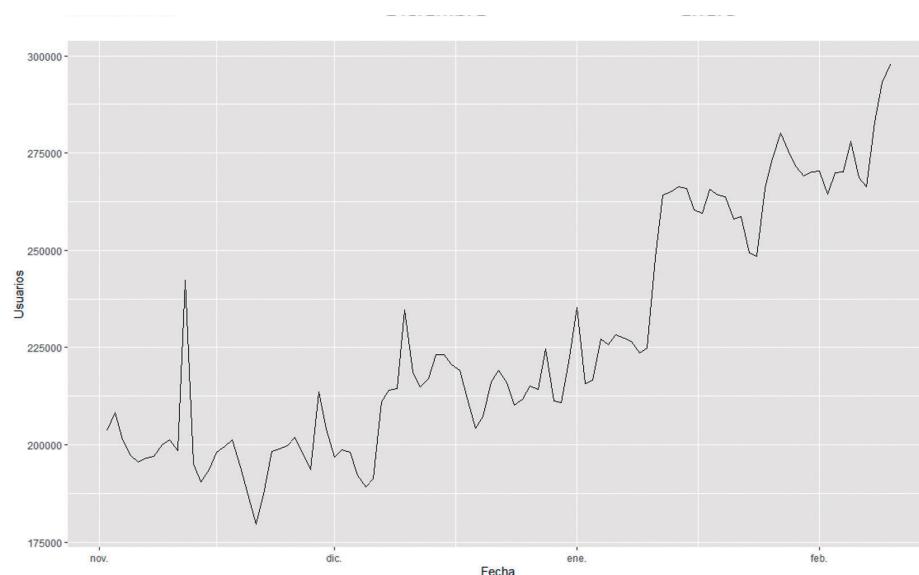


Figura 3. Cantidad de usuarios concurrentes en la red

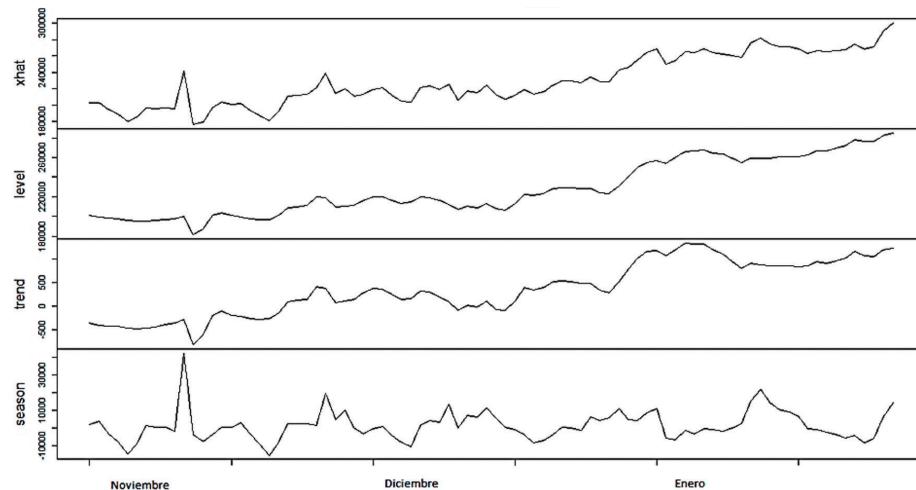


Figura 4. Descomposición de la serie usuarios concurrentes

ma es mostrada en la figura 3.

La descomposición de dicha serie es mostrada en la figura 4.

A partir de la descomposición de la serie en los parámetros correspondientes, se observa que no existe una estacionalidad regular, aunque sí posee los elementos de tendencia y nivel bien demarcados.

Para el caso de la primera serie (Velocidad máxima alcanzada) se justifica correctamente el uso del método de *Holt-Winters*, por la presencia de los 3 componentes y, a pesar de que la estacionalidad en la segunda serie (Usuarios concurrentes en la red) no sea tan marcada como en la anterior, es criterio de los autores emplear dicho método.

Se utilizó para realizar el pronóstico el software RStudio. Mediante la función *HoltWinter*, arroja los resultados con los valores óptimos de los parámetros para la realización del pronóstico. De aquí se nota la fuerte estacionalidad de la serie velocidad máxima (Valor de Gamma), y que poseen tendencias de suavizado bastante bajas (Valor de Beta).

Los resultados alcanzados de la aplicación de *Holt-Winters* en ambas series se muestran en las figuras 5 y 6.

Los componentes de atenuación resultados de la aplicación del método en ambas series son mostrados en la Tabla 1.

Otro de los métodos empleados fue ARIMA, como uno de los posibles métodos autoregresivos a utilizar, dado por la recomendación UIT E.507. Para este caso los pronósticos son obtenidos mediante las funciones `auto.arima` de la librería `forecast`, la cual proporciona una opción rápida para construir pronósticos con series temporales, debido a que evalúa entre todos los posibles modelos al mejor, considerando diversos criterios: estacionariedad, estacionalidad, diferencias, entre otras. Las figuras 7 y 8 muestran las predicciones utilizando este modelo.

Para el caso de la velocidad de transferencia se obtiene que el mejor modelo que brinda la función es ARIMA(0,1,0)(0,1,1), lo que significa que existe una correlación significativa de una componente de medias móviles de primer orden, con una diferenciación para llevar la serie a estacionaria.

El coeficiente es `sma 1: -0.7482`.

Para los usuarios concurrentes el mejor modelo que arroja la función es

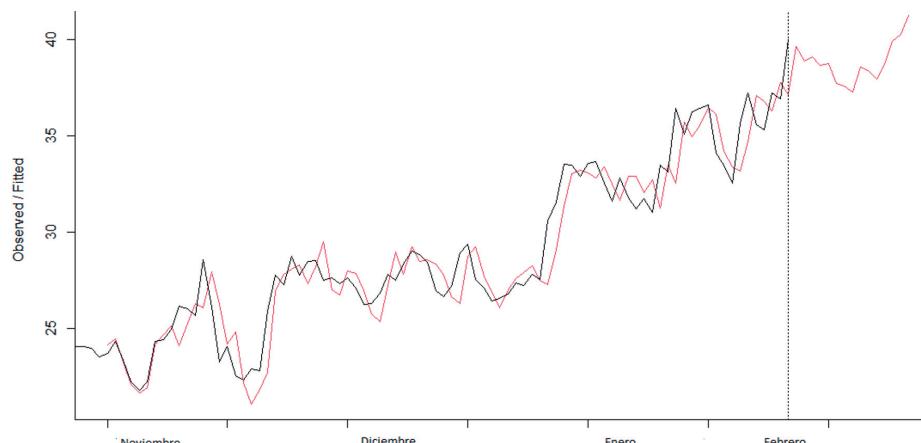


Figura 5. Predicción para 15 días de velocidad máxima alcanzada (en Gbps) mediante método *Holt -Winters*

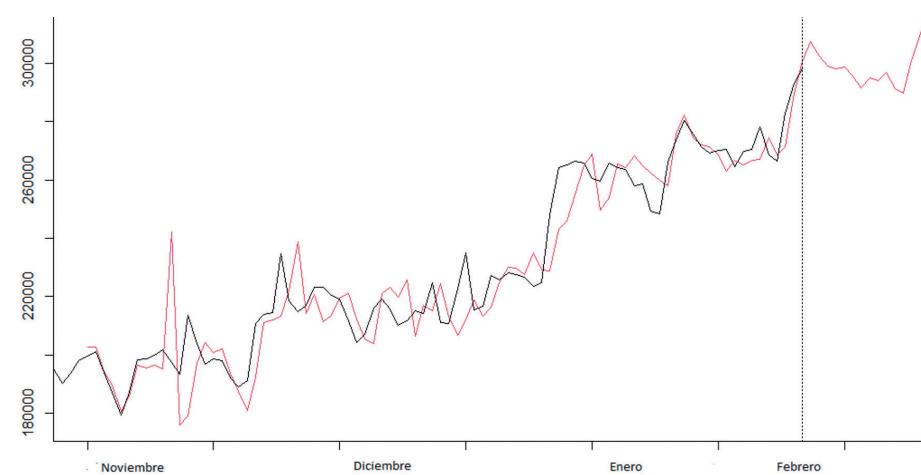


Figura 6. Predicción para 15 días de usuarios concurrentes en la red mediante método *Holt -Winters*

Serie	Alpha	Beta	Gamma
VELOCIDAD MÁXIMA	0.7040098	0.006076382	0.9351032
Usuarios concurrentes	0.4157313	0.03001612	0.8877335

Tabla 1. Componentes de suavizado del método *Holt -Winters*

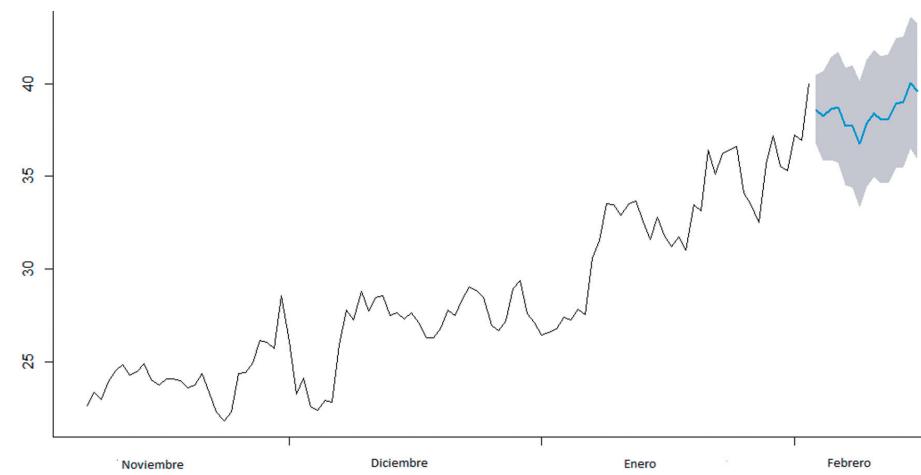


Figura 7. Predicción para 15 días de velocidad máxima alcanzada (en Gbps) mediante método ARIMA

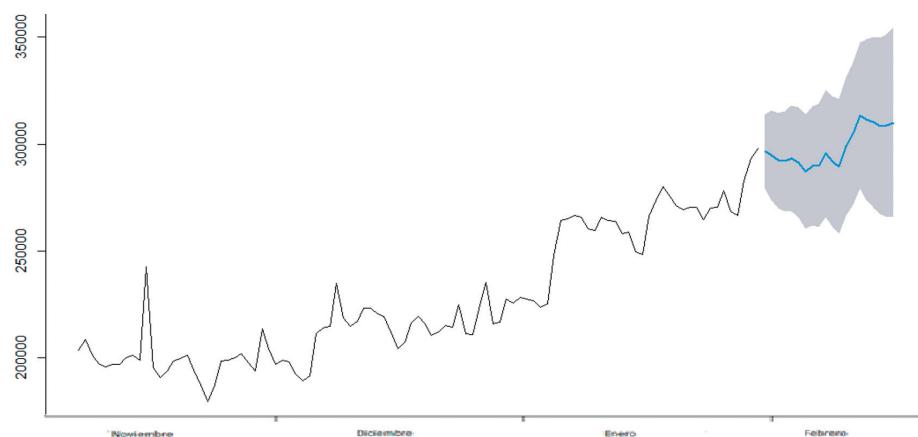


Figura 8. Predicción para 15 días de usuarios concurrentes mediante método ARIMA

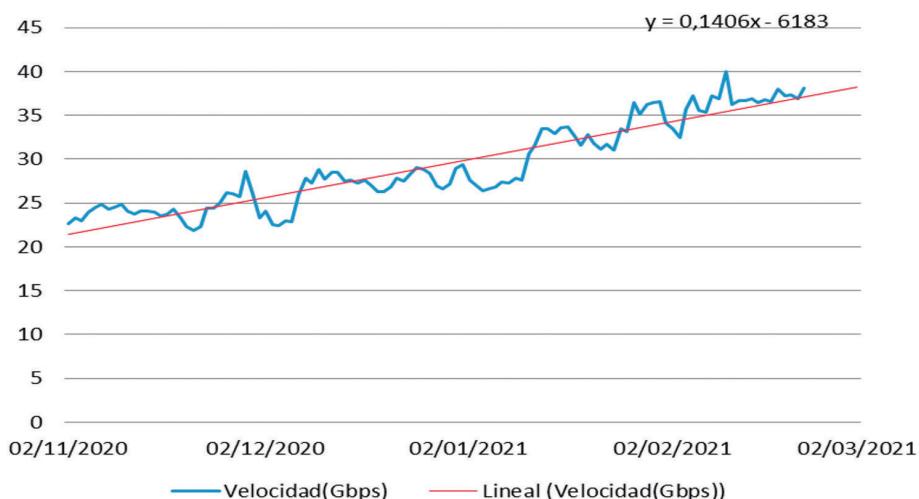


Figura 9. Predicción para 15 días de velocidad máxima alcanzada (en Gbps) mediante extrapolación lineal

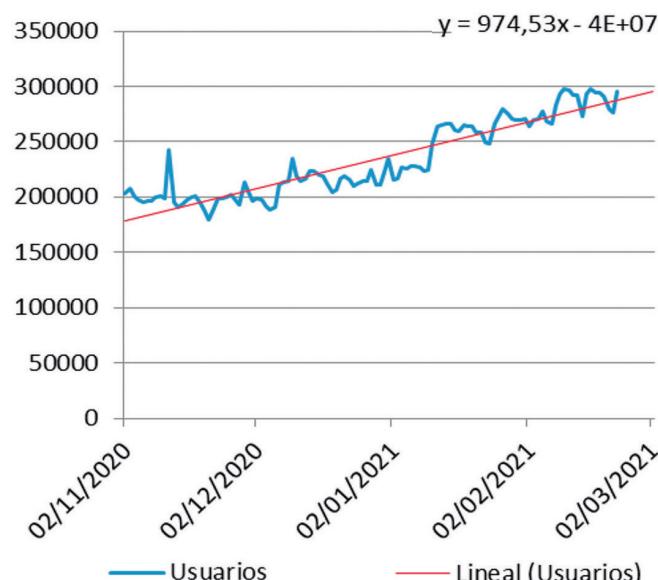


Figura 10. Predicción para 15 días de usuarios concurrentes en la red mediante extrapolación lineal

ARIMA(0,1,2)(1,1,1), los parámetros arrojados por esta función se muestran en la tabla 2.

Lo que significa que los rezagos con alta correlación son los de orden 1 y 2 de medias móviles, para la componente estacional tiene un componente autorregresivo de orden 1 y de medias móviles de orden 1. La serie para su análisis necesitó una transformación de diferenciación.

Se propone la utilización de una extrapolación lineal para ambas series, pues es una de las formas más simples y generalizadas de realizar algún tipo de pronóstico. Para este caso los resultados se muestran en las figuras 9 y 10.

Si se comparan los pronósticos realizados mediante las 3 variantes se obtienen los siguientes resultados, mostrados en la figura 11.

El comportamiento del error absoluto (MAD) en los métodos de predicción utilizados es observado en la figura 12.

Los errores relativos (MAPE) en ambas series pueden verse en la Tabla 3.

Los comportamientos de los errores tanto del MAD como del MAPE, se consideran como aceptables para los 3 pronósticos realizados.

Coeficientes del modelo			
ma1	ma2	sar1	sma1
-0.3095	-0.2579	0.2926	-0.8310

Tabla 2. Coeficientes del modelo para la velocidad de transmisión

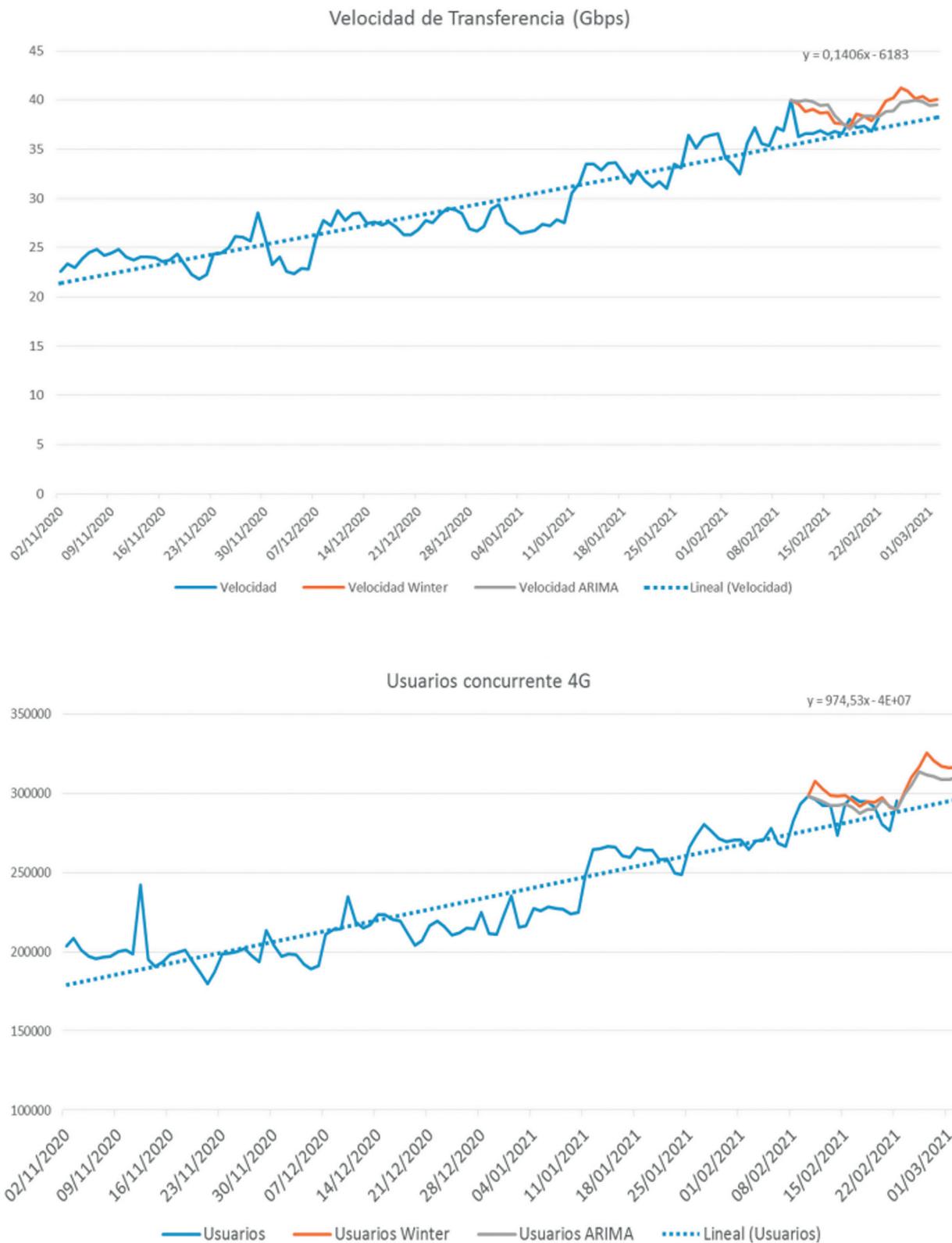


Figura 11. Comparación de los pronósticos obtenidos con los valores reales

Conclusiones

Debe tenerse en cuenta, que una buena práctica para la realización de predicciones consiste en la evaluación de varios métodos. Para las series de referencia utilizadas, los pronósticos realizados utilizando los métodos propuestos, no presentan errores relativos superiores al 10% de los valores reales.

En los análisis de errores (MAD y MAPE), los valores obtenidos son menores en una aproximación lineal, para el caso de Velocidad Máxima, aunque un análisis de extrapolación lineal no tendrá en cuenta las posibles estacionalidades en las series, que fueron observadas en las figuras 2 y 4. Por lo que para esta serie se propone para realizar un pronóstico: el modelo *Holt-Winter*.

Para el caso de usuarios concurrentes se recomienda como modelo a utilizar *Holt-Winters*, pues exponen valores inferiores de error, manteniendo la estacionalidad vista en dicha serie.

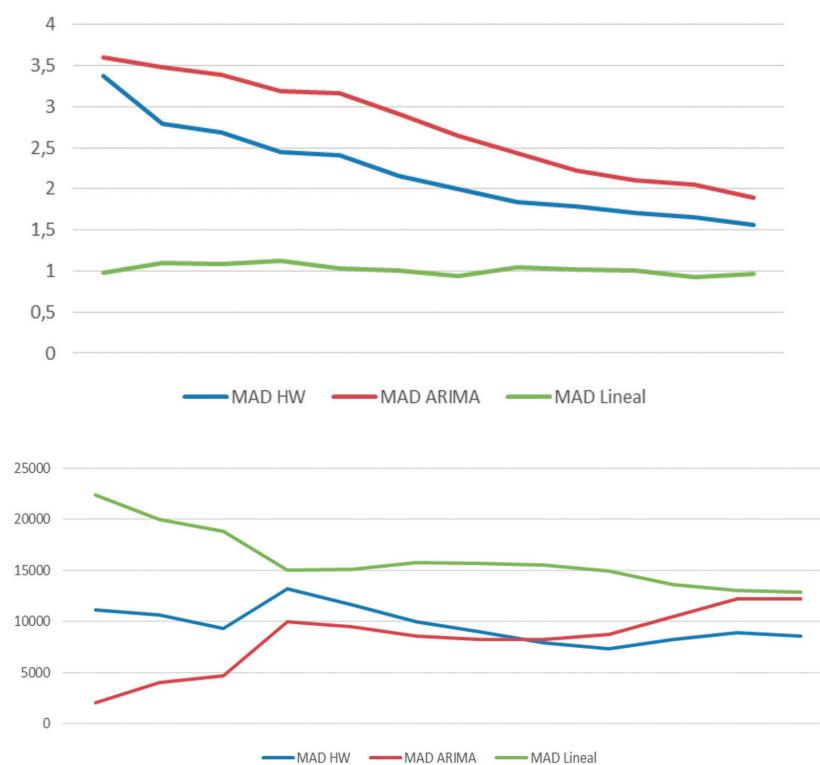


Figura 12. Comportamiento del MAD para las predicciones realizadas en ambas series

Método	Velocidad máxima	Usuarios concurrentes
Holt-winters	4.25%	3.03%
ARima	5.16%	4.30%
Regresión lineal	2.61%	4.39%

Tabla 3. Valores de los errores relativos (MAPE) en los pronósticos

Referencias bibliográficas

- Brockwell, P. y Davis, R. (2016). *Introduction to Time Series and Forecasting*. Springer.
- Cryer, J. y Chan, K. (2008). *Time Series Analysis With Applications in R*. Springer.
- Garzón Rodríguez, Yaqueline y Wanumen Silva, Luis Felipe (2014). *Tratamiento de los datos para el pronóstico de tráfico en redes Wi-Fi, mediante series de tiempo*. Universidad Distrital Francisco José de Caldas, Bogotá, Colombia.
- González Casimiro, Pilar. (2009). *Ánalisis de Series Temporales: Modelos ARIMA*. Departamento de Economía Aplicada III (Econometría y Estadística). Facultad de Ciencias Económicas y Empresariales. Universidad del País Vasco (UPV-EHU).
- UIT, Recomendación E.507 (En vigor) *Red telefónica y RDSI calidad de servicio, gestión de la red e ingeniería de tráfico modelos para la previsión del tráfico internacional (11/88)*.
- Villarreal, Fernanda (2016). *Introducción a los Modelos de Pronósticos*. Departamento de Matemática, Universidad Nacional del Sur.