

Análisis de expansión de redes de telefonía móvil empleando procesos Gaussianos

Analysis of the expansion of a GSM network using Gaussian Processes

Jhouben Janyk Cuesta Ramírez

Ingeniero Electricista, Joven Investigador
Laboratorio de Investigación en Automática.
Universidad Tecnológica de Pereira UTP, Colombia
jycuesta@utp.edu.co

Álvaro Ángel Orozco Gutiérrez

Ph.D. Bioingeniería, Docente Titular
Universidad Tecnológica de Pereira UTP, Colombia
aaog@utp.edu.co

Mauricio Alexander Álvarez López

Ph.D. Ciencias de la Computación, Docente Asociado
Universidad Tecnológica de Pereira UTP, Colombia
malvarez@utp.edu.co

Resumen— Durante la formulación de su plan de expansión, una red de telefonía móvil (GSM) requiere del análisis de aquellas variables que resultan clave para el buen desempeño de la red (KPIs). Los operadores de red poseen herramientas que analizan el comportamiento del KPI sólo para aquellos puntos de la red que contienen celdas. Este artículo propone una herramienta que ilustra de manera gráfica el comportamiento en el tiempo de un KPI, no sólo para aquellos puntos donde se encuentren las celdas, sino también en la totalidad de la zona geográfica donde la red de celdas se encuentra ubicada. Se aplica un Proceso Gaussiano de Regresión a mediciones obtenidas de las celdas pertenecientes a la red y se infiere una superficie que representa el comportamiento en el tiempo para toda la zona. Finalmente se observa cómo una región de la red con poca densidad de celdas sostiene valores elevados del KPI la mayoría del tiempo invitando al operador de red a tener en cuenta la solución del problema de dicha región en la formulación del plan de expansión.

Palabras clave— aprendizaje de máquina, KPIs, Procesos Gaussianos, red global de telefonía (GSM), regresión.

Abstract— the expansion plan of a Global System Mobile (GSM) network requires the analysis of some important variables known as key performance indicators (KPI) on the network. Network operators have tools for analyzing a KPI behavior on a particular network cell. This paper proposes a tool that illustrates graphically the behavior-in-time of a KPI in a whole geographical zone (including cell positions). A Gaussian process regressor is used over a real data set and time-space inference is performed. Finally we observe how a particular region presents high-KPI values most of the time. This alerts the network operator for including a solution in the formulation phase of the network expansion plan.

Keywords— Gaussian Processes, GSM networks, KPIs, machine learning, regression.

1. INTRODUCCIÓN

Los operadores de las redes globales de telefonía móvil (GSM) intentan siempre estar a la vanguardia en cuanto a calidad y tecnología del servicio prestado; evolucionan a la par de los sistemas y tecnologías para la comunicación, para entregar a sus usuarios un mejor servicio, para ello requieren de la formulación de un buen plan de expansión.

Un plan de expansión capaz de asegurar un óptimo manejo de los recursos de telefonía básicamente comprende de: el análisis del desempeño de la red actual; el análisis de la propagación del recurso en la zona; la identificación de celdas que requieran asignación de nuevos controladores de tráfico de la red (RNC) y centros de conmutación de datos móviles (MSC); la asignación de RNC y MSC a nuevas celdas, de acuerdo a las restricciones de la red; y la expansión de la conectividad de la red [1]. Los primeros 2 puntos mencionados requieren del monitoreo constante de indicadores de desempeño (PI) tales como: la tasa máxima a la cual la información puede ser enviada (ancho de banda), la tasa actual a la que se transfieren los datos (*throughput*) y la variación en el tiempo de llegada al receptor (*jitter*), entre otros. Aquellos PIs clave

para el buen desempeño de la empresa son denominados Indicadores Clave de Desempeño (KPIs) [2], [3] (en [4], se propone un listado de KPIs para redes 3G o superiores), por tanto, es necesario que el operador de red entienda de la mejor manera posible el comportamiento de los KPIs y, en consecuencia, pueda conocer las razones de los estados operacionales de la red [5], requeridos para realizar un correcto análisis del desempeño de su desempeño.

La cantidad de mediciones obtenidas de redes de telecomunicaciones se ha incrementado sustancialmente durante la última década [2]. Hoy en día es normal contar con grandes bases de datos correspondientes a mediciones como la tasa máxima a la cual la información puede ser enviada (tasa de bits), la tasa actual a la que se transfieren los datos (*throughput*) y la variación en el tiempo de llegada al receptor (*jitter*), entre otros. Aquellas que influyen en el desempeño y cumplimiento de los objetivos de una empresa son denominadas indicadores clave de desempeño (KPI). Por tanto, conocer el comportamiento de dichos KPIs resulta imprescindible en la etapa de análisis de la red, y muy especialmente, durante la formulación del plan de expansión.

El papel de análisis de los KPIs en telefonía ha sido estudiado con anterioridad en la literatura especializada. En [6] se propone una metodología de análisis basada en la capacidad y cobertura de las celdas, se aplica sectorización para obtener mediciones más precisas del KPI relacionado con la calidad del servicio (QoS) en las celdas. En [7] se emplea un análisis estadístico multivariado (correlación, covarianza, escalamiento dimensional y clasificación) basado en KPIs obtenidos de una red como base para inferir sobre el valor de la capacidad. En [8] se propone un método de *clasificación* basado en lógica difusa para el análisis y monitoreo automático de la calidad en las celdas pertenecientes a una red 3G, donde a cada celda se le asigna una clase relacionada al funcionamiento de acuerdo a la calidad monitoreada, con la cual el operador de red puede decidir acciones inmediatas o futuras para la solución de problemas así como tener en cuenta dichos comportamientos de la celda en la formulación del plan de expansión. Estos enfoques tienen como

objetivo principal servir de herramienta para conocer el comportamiento de los elementos y las variables en los diferentes puntos de la red, fundamentándose en analizar los KPIs en el punto donde las mediciones son obtenidas, sin proveer información alguna sobre puntos del espacio (e intervalos de tiempo) en los cuales no se hayan realizado mediciones.

Una forma general de obtener información en puntos geográficos desconocidos es aplicar un enfoque de *regresión*, para el cual el Proceso Gaussiano (GP) [9] provee un marco único, flexible y de buen desempeño. El Proceso Gaussiano, a pesar de ser una herramienta utilizada recientemente en la estadística, la ingeniería y otros campos [10] y poco utilizada en el área de telefonía [11]-[13]. Algunos de los enfoques existentes realizaban ya fuese la regresión de la posición de un usuario de la red con la intensidad de la señal [11], [12] o la regresión de la densidad de una señal WiFi en un entorno cerrado [14]. Aquí se propone una metodología de regresión basada en Procesos Gaussianos para interpolar el valor de un KPI en diferentes puntos del espacio y diferentes valores de tiempo en una zona geográfica específica donde se encuentra ubicada una red de telefonía 3.5G. El nuevo enfoque que se presenta pretende hacer uso de esta poderosa herramienta para explotar las relaciones entre las diferentes mediciones del KPI, lo cual dada la naturaleza móvil del usuario y el gran espacio geográfico que ocupa la red de celdas sirve (como se mostrará más adelante) como herramienta para capturar patrones que ayudan a entender tanto los valores que toma el KPI en diversos puntos, como el comportamiento de los usuarios de la red lo cual finalmente permite poder plantear mejores planes de expansión.

El artículo está compuesto por varias secciones. En la sección 2 se introducen los materiales y métodos utilizados en el experimento: inicialmente se realiza una descripción de la base de datos, luego se presenta el Proceso Gaussiano como herramienta de inferencia, el modelado de los datos seguido del procedimiento de validación utilizado para evaluar el desempeño del modelo y la metodología utilizada tanto para la elección del modelo como para la inferencia sobre la totalidad de la base de datos. La sección

3 contiene los resultados más importantes obtenidos durante la realización del experimento así como un análisis de los mismos. Finalmente, en la sección 4 se presentan las conclusiones y se proponen trabajos futuros al respecto.

2. MATERIALES Y MÉTODOS

2.1. Base de datos

Los datos utilizados en el experimento pertenecen a mediciones por estación base de un KPI perteneciente a una red GSM colombiana, ubicada en una zona geográfica determinada.

La Fig. 1 muestra la configuración espacial de 11 celdas pertenecientes a una red de telefonía. Se cuenta con mediciones diarias de un KPI relacionado con el manejo del recurso de datos durante 30 días consecutivos para cada una de las celdas mostradas; conformando así la base de datos utilizada durante el experimento. La zona geográfica fue determinada de acuerdo a los límites de la imagen tomando como origen de referencia el borde inferior izquierdo, los ejes X e Y se muestran en la figura, así como un valor representativo en kilómetros de la escala de distancias.

A manera de ejemplo, la Fig. 2 representa gráficamente las mediciones correspondientes a las celdas 1 y 5. El eje X muestra el tiempo en días y el eje Y el valor del KPI. Como medida comparativa entre los valores para cada celda y posteriormente entre los valores obtenidos como resultado del proceso de inferencia, se define un valor de KPI medio de 38, el cual corresponde al promedio de todas las mediciones.

Fig. 1. DISTRIBUCIÓN GEOGRÁFICA DE LAS CELDAS PERTENECIENTES A LA RED DE TELEFONÍA MÓVIL

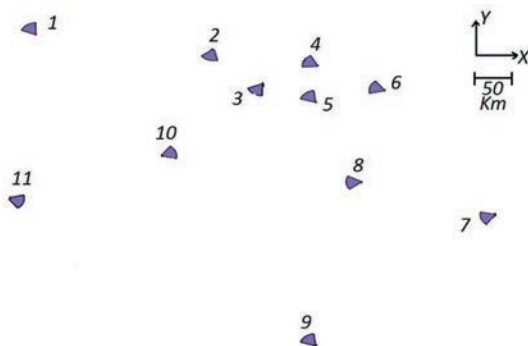
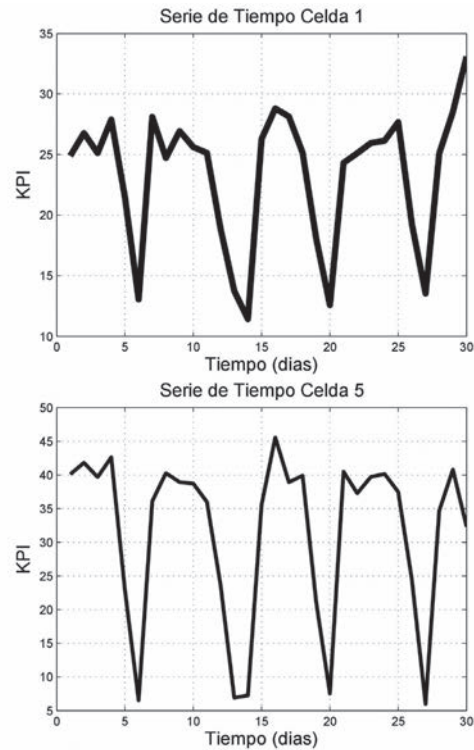


Fig. 2. SERIE DE TIEMPO CORRESPONDIENTE A LAS CELDAS 1 Y 5



2.2. Procesos Gaussianos

Un Proceso Gaussiano (GP) está compuesto por una colección posiblemente infinita de variables aleatorias escalares $\{f(x)\}_{x \in X}$ indexadas por un espacio de entradas $x: R^D \rightarrow R$ de tal forma que para cada combinación finita $X = \{x_1, \dots, x_n\}$, todas las funciones $f \triangleq [f(X_1) \dots f(X_n)]^T$ siguen una distribución Gaussiana multivariada [13]. Un GP queda formalmente definido con la existencia de una función media y una función de covarianza:

$$m(X) = E[f(X)], \quad (1)$$

$$k_f(X, X') = E[(f(X) - m(X))(f(X') - m(X'))^T] \quad (2)$$

Las ecuaciones (1) y (2) definen las propiedades del Proceso Gaussiano. La distribución de probabilidad sobre las funciones puede ser resumida mediante la notación: $f(x) \sim GP(m(x), k(x, x'))$.

En regresión, se parte de un conjunto de datos de entrenamiento $D = \{(x_i, y_i) | i = 1, \dots, n\}$ con los cuales se desea hacer predicciones f_* para nuevos valores x_* no pertenecientes al conjunto de entrenamiento.

Los Procesos Gaussianos cumplen con la propiedad de marginalización, lo cual implica que una distribución de probabilidad formada con

cualquier combinación de variables aleatorias dentro del proceso no es afectada por el resto de variables [9], [15].

Durante el modelado de los datos, es necesario definir una función de verosimilitud y una de covarianza.

Una *función de verosimilitud*, es aquella que representa la probabilidad condicional de las observaciones dado el modelo para así, con el *teorema de Bayes* calcular la distribución posterior (o inferencia) correspondiente. Es requerido también la elección o definición de una función de covarianza.

Una función de covarianza o *kernel*, es aquella que recibe a la entrada dos argumentos $x \in X$, $x' \in X$ y los transfiere al dominio R . En el Proceso Gaussiano, es la función que se encarga de definir el tipo de relación existente entre los valores pertenecientes al conjunto de entrenamiento. Los tipos de kernel más utilizados son del tipo exponencial, del tipo lineal, y de la clase *Mátern* [16].

En este experimento fueron utilizadas la función exponencial cuadrática (*Squared Exponential*) definida como:

$$k_{SE}(r) = \theta_f^2 \exp\left(-\frac{r^2}{2\theta_2^2}\right),$$

(El parámetro $\theta_f^2 = \sigma_f^2$ es la representación de la desviación estándar, el parámetro $\theta_2^2 = l$ representa la escala en el rango y la magnitud $r = (x-x')^T(x-x')$); al igual que las funciones de la clase *Mátern* del tipo 1/2, 3/2 y 5/2:

$$k_M(r) = \frac{2^{1-\nu}}{\Gamma(\nu)} \left(\frac{\sqrt{2\nu r}}{l}\right)^\nu K_\nu\left(\frac{\sqrt{2\nu r}}{l}\right),$$

(K_ν es una función de *Bessel* modificada, ν y l parámetros positivos).

Si se tiene una función de verosimilitud de la forma: $y = f(x) + \varepsilon$, donde el ruido $\varepsilon \sim N(0, \sigma_n^2)$ se supone independiente e idénticamente distribuido (i.i.d.), donde σ_n^2 corresponde a la varianza del ruido; la función de covarianza para dos salidas cualesquiera, toma la forma $cov(y_p, y_q) = k(x_p, x_q) + \sigma_n^2 \delta_{pq}$ (siendo δ_{pq} la función delta, la cual entrega un valor de la unidad únicamente cuando $p = q$ y cero en cualquier otro caso). Asumiendo función media cero en el prior, se obtiene la distribución de probabili-

dad conjunta de las salidas de entrenamiento y , y predicción f_* :

$$\begin{bmatrix} y \\ f_* \end{bmatrix} \sim N\left(0, \begin{bmatrix} K(X, X) + \sigma_n^2 I & K(X, X_*) \\ K(X_*, X) & K(X_*, X_*) \end{bmatrix}\right),$$

donde $K(X_*, X)$ corresponde a la matriz de covarianza resultado de evaluar la función de covarianza entre los valores de entrenamiento X y los valores de predicción X_* . De igual forma se obtienen $K(X, X)$, $K(X_*, X_*)$ y $K(X, X_*)$; y donde I corresponde a la matriz identidad.

Aplicado el teorema de Bayes y algunas propiedades de la distribución Gaussiana [15] se obtiene la distribución condicional posterior para las predicciones, donde $\bar{f}_* = E[f_* | X, y, X_*] = K(X_*, X)[K(X, X) + \sigma_n^2 I]^{-1} y$, $Y cov(f_*) = K(X_*, X_*) - K(X_*, X)[K(X, X) + \sigma_n^2 I]^{-1} K(X, X_*)$.

Finalmente, para obtener los valores de f_* , se requiere primero conocer los parámetros correspondientes al modelo del GP. Una forma de calcularlos, y la utilizada en este trabajo, consiste en maximizar la función de verosimilitud marginal [9].

Los cálculos de inferencia y optimización fueron realizados con ayuda del *toolbox* de Carl E. Rasmussen y Chris Williams en su versión para *MatLab*® [17].

2.3. Modelado de los datos

Matemáticamente la base de datos se expresó de la siguiente forma:

$$\begin{bmatrix} x_{x,1} & x_{y,1} & t_1 \\ x_{x,1} & x_{y,1} & t_2 \\ x_{x,1} & x_{y,1} & t_n \\ M & M & M \\ x_{x,p} & x_{y,p} & t_1 \\ x_{x,p} & x_{y,p} & t_{12} \\ x_{x,p} & x_{y,p} & t_n \end{bmatrix}, \begin{bmatrix} y_{1,1} \\ y_{1,2} \\ y_{1,n} \\ M \\ y_{p,1} \\ y_{p,2} \\ y_{p,n} \end{bmatrix} \quad (3)$$

donde los índices $x_{x,i}$ y $x_{y,i}$ indican la coordenada (x, y) de la celda i (con $i = \{1, L, p\}$; siendo $p = 11$, la cantidad total de celdas); y el valor t_k representa el día en el cual se efectuó la toma del dato (con $K = \{1, L, n\}$; $n = 30$, es el número total de días); y los valores de $y_{i,k}$ representan el valor promedio de las mediciones de los sectores pertenecientes a la celda i en el día k .

Para el modelado Gaussiano se asumió función media cero y función de verosimilitud con ruido Gaussiano independiente e idénticamente distribuido (iid); se utilizaron las funciones de covarianza mencionadas en la sección 2.2 con las cuales se realizó un procedimiento previo con la

finalidad de encontrar cual de las funciones era más apta para el proceso de inferencia (dicho procedimiento será explicado en detalle en las secciones 3.1.1 y 3.1.2).

Los parámetros iniciales escogidos para el modelo fueron ($I=I, \sigma_f=I, \sigma_n=0.1$) correspondientes al parámetro de escala en longitud, la desviación estándar en las funciones y la desviación estándar del ruido respectivamente, necesarios para hallar los parámetros finales del modelo con el criterio de máxima verosimilitud marginal.

2.4. Validación

Se utilizó el procedimiento de validación cruzada (*Cross Validation*) [15], el cual consta básicamente de subdividir el conjunto inicial de datos en k particiones de igual tamaño, realizar la etapa de entrenamiento con $k-1$ subconjuntos y validar con el restante, se repite este procedimiento para cada una de las k diferentes particiones. Esto con la finalidad particular de poder comparar el desempeño del modelo para “nuevos” valores. El desempeño del modelo fue analizado con las métricas de: Error Medio Cuadrático Normalizado o SMSE y la Pérdida Logarítmica Estándar Normalizada o MSL [18]. Cabe resaltar que un valor de SMSE bajo indica un menor error y un valor de MSL negativamente alto indica un mejor modelo [9].

2.5. Procedimiento de pre-selección para la elección del modelo

El procedimiento realizado como criterio de selección de la función de covarianza a ser utilizada en la inferencia del KPI para toda la zona geográfica, consistió en formar con las mediciones de cada celda 5 subconjuntos (de a 6 mediciones cada uno) y así proceder de la siguiente manera:

- 1) Se asume que cada celda de manera independiente y se calcula un GP de regresión para obtener la serie temporal asociada a cada celda. En este caso la entrada del GP corresponde al tiempo y la salida al valor del KPI en la ubicación geográfica específica de la celda. El proceso de entrenamiento del GP se realiza con 4 de los subconjuntos mientras que el subconjunto restante es utilizado en la validación, realizando finalmente 5 diferentes procesos de estimación para cada celda. Este

procedimiento fue nombrado por los autores como “procedimiento de inferencia Local”.

- 2) Se asume toda la configuración de celdas en la zona geográfica y se estima un único GP para obtener de igual manera que en el ítem anterior, la serie temporal asociada a cada una de las celdas. La entrada esta vez corresponde a la planteada en (3) y la salida el valor del KPI en la ubicación geográfica específica de la celda. Se diferencia del procedimiento descrito en el ítem anterior, debido a la inclusión en el conjunto de entrenamiento, del total de las mediciones de aquellas celdas a las que no corresponda la regresión. De igual forma se realizan 5 diferentes procesos de estimación para cada celda. Este procedimiento fue nombrado por los autores como: “procedimiento de inferencia Global”.

Finalmente, con los valores promedios del SMSE y el MSL se tuvo un criterio de selección para el modelo más apto.

2.6. Procedimiento para la inferencia espacio-temporal

El procedimiento final, consistió en calcular un GP espacio-temporal de la forma indicada en (3), utilizando como entrenamiento la totalidad de la base de datos y así inferir sobre un gran conjunto de puntos pertenecientes a la misma zona geográfica. Dichos puntos fueron tomados de una rejilla formada desde el origen de coordenadas hasta 600 km de distancia (3721 puntos entre 0-6) y tiempos de 1-30 (291 intervalos de tiempo en total) para un total de 1082811 puntos individuales a inferir.

Las distancias geográficas fueron divididas en 100, y los intervalos de tiempo fueron más cortos que el día, con la finalidad de obtener una función más suave y de paso tener mayor conocimiento del comportamiento del KPI.

3. RESULTADOS

3.1. Resultados procedimiento de pre-selección

Las Tablas I y II, relacionan los errores SMSE y MSL tanto locales como globales, correspondientes a los procesos de inferencia previamente mencionados en el apartado 3.1.1 y 3.1.2 respectivamente correspondientes a las celdas 09 y 07,

para cada una de las funciones de covarianza estudiadas. Ambas medidas de error se encuentran acompañadas de la respectiva desviación estándar. Se resalta también como el proceso de inferencia espacio-temporal presenta menor error que el proceso que involucra sólo una celda en el tiempo, así como el hecho que la función de covarianza Màtern de tipo 1, presenta parcialmente menores errores SMSE y mayores errores negativos MSLL (resaltados en negrita) para las dos celdas mencionadas.

Con los resultados obtenidos en la sección 3.3, se aplicó el procedimiento planteado en 3.2 a los datos, dando los siguientes resultados:

La Fig. 3 y la Fig. 4 correspondientes a la superficie de inferencia y al respectivo contorno, ilustran el comportamiento del KPI alrededor de toda la zona geográfica para un valor de tiempo entre los días 10 y 11 de mediciones. Los colores en las figuras corresponden a los niveles del KPI (en escala real), en la barra de la derecha de cada imagen (El color Azul

oscuro indica un valor mínimo del KPI de 0 y el color Rojo intenso un valor máximo de 160).

La Tabla III, resume los resultados de promedio de las 11 celdas analizadas durante los procedimientos descritos en la sección 3.1. De ésta se resalta la función de covarianza Màtern de tipo 1, la cual presentó menor error SMSE y a su vez mayor MSLL en el caso espacio temporal, por lo cual se escogió para realizar el procedimiento de inferencia descrito en la sección 3.2

3.2. Resultados procedimiento de inferencia espacio-temporal

La región superior compuesta por la mayor cantidad de celdas presenta valores promedios normales del KPI, lo cual se interpreta como resultado de la cercanía entre las celdas y que entre las mismas logran repartirse el recurso del KPI y por tanto, no se presentan valores altos del mismo en dicha región.

TABLA I
RESULTADOS DE LOS ERRORES SMSE Y MSLL CORRESPONDIENTES A LA CELDA 9, PARA LAS DIFERENTES FUNCIONES DE COVARIANZA

CELDA 9				
Cov.\ Error	SMSE Local	SMSE Global	MSLL Local	MSLL Global
Exp C.	0.8406±0.2051	0.3563±0.0622	0.4586±1.2612	-0.6824±0.0211
Màtern 5	0.7943±0.1516	0.3413±0.0608	-0.0527±0.1780	-0.6991±0.0263
Màtern 3	0.7731±0.1240	0.3313±0.0590	-0.1009±0.1135	-0.7096±0.0271
Màtern 1	0.7544±0.0658	0.3083±0.0485	-0.1383±0.0439	-0.7211±0.0244

Fuente: autores

Tabla II
RESULTADOS DE LOS ERRORES SMSE Y MSLL CORRESPONDIENTES A LA CELDA 7, PARA LAS DIFERENTES FUNCIONES DE COVARIANZA

CELDA 7				
Cov.\ Error	SMSE Local	SMSE Global	MSLL Local	MSLL Global
Exp C.	0.8392±0.1340	0.1772±0.0264	-0.0225±0.1757	-1.9536±0.3451
Màtern 5	0.7977±0.1022	0.1738±0.0286	-0.0744±0.1302	-1.8820±0.3899
Màtern 3	0.7818±0.0890	0.1702±0.0302	-0.0982±0.1033	-1.8896±0.3968
Màtern 1	0.7729±0.0612	0.1606±0.0353	-0.1230±0.0466	-2.0880±0.2841

Fuente: autores

Tabla III
RESULTADOS DE LOS ERRORES SMSE Y MSLL CORRESPONDIENTES A LA CELDA 7, PARA LAS DIFERENTES FUNCIONES DE COVARIANZA

RESULTADOS PROMEDIO EN TODAS LAS CELDAS PARA CADA FUNCIÓN DE COVARIANZA				
Cov.\ Error	SMSE Local	SMSE Global	MSLL Local	MSLL Global
Exp C.	0.7563±0.1525	0.4007±0.3396	-0.0329±0.2075	-0.7421±0.4594
Màtern 5	0.7429±0.1390	0.3796±0.3212	-0.1330±0.0819	-0.7534±0.4341
Màtern 3	0.7142±0.1320	0.3678±0.3111	-0.1473±0.0649	-0.7650±0.4324
Màtern 1	0.7558±0.0986	0.3491±0.2943	-0.1483±0.0379	-0.7907±0.4844

Fuente: autores

Un resultado diferente presenta la región cercana a la celda 7 ubicada en la parte inferior derecha de la Fig. 5 donde se presentan altos niveles en el KPI, esto sumado al hecho que se encuentra sin celdas vecinas, implicaría indirectamente la necesidad de ubicar una nueva celda cercana que ayude a reducir los altos niveles del KPI y así evitar saturaciones futuras.

Las Figuras 5 y 6 contienen varias de las superficies de inferencia calculadas y sus respectivos contornos. La barra de colores es igual a la de las Figuras 4 y 5. Cada contorno de la Fig. 6 corresponde al ubicado en la misma posición en la Fig. 5. Se puede observar parte de la dinámica del KPI en el tiempo. Para poder tener una perspectiva total del comportamiento en el tiempo del KPI y saber si estos comportamientos presentados en las Fig. 3 y 4 son generalizables, se realizó un video para la superficie y para el contorno utilizando las 291 muestras inferidas entre los 30 días.

Fig. 3 DÍA 10 DE LA INFERENCIA REALIZADA PARA TODA LA ZONA GEOGRÁFICA

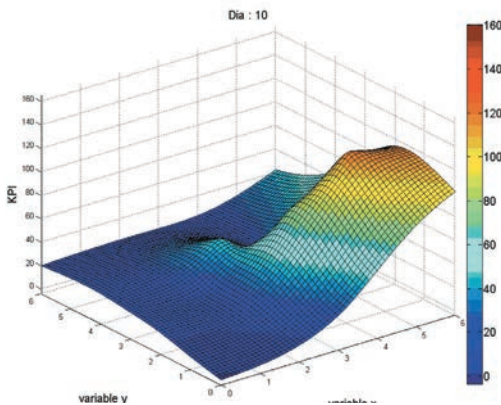


Fig. 4 CONTORNO CORRESPONDIENTE A LA INFERENCIA REALIZADA PARA TODA LA ZONA GEOGRÁFICA DURANTE EL DÍA 10

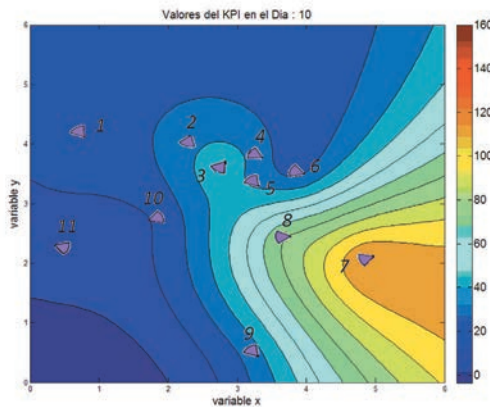


Fig. 5 ALGUNAS SUPERFICIES OBTENIDAS DURANTE EL PROCESO DE INFERENCIA ESPACIO-TEMPORAL

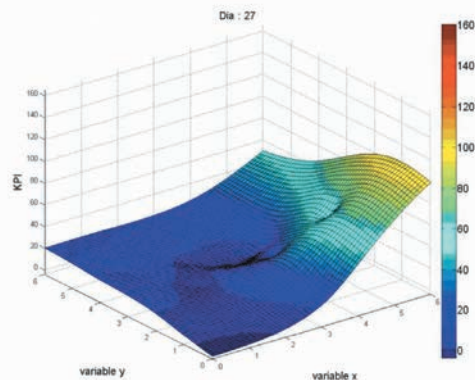
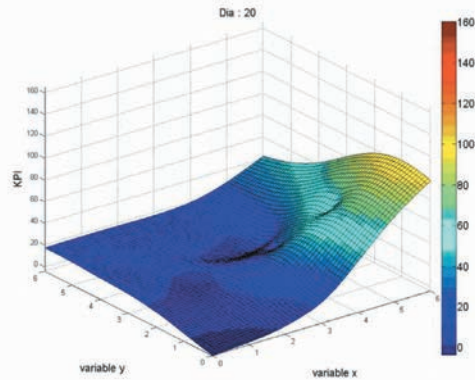
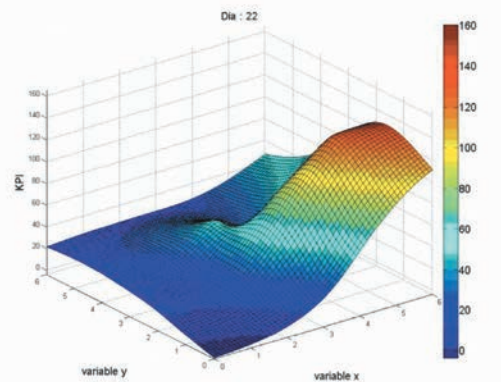
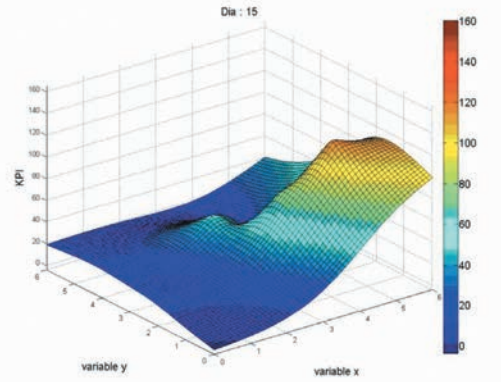
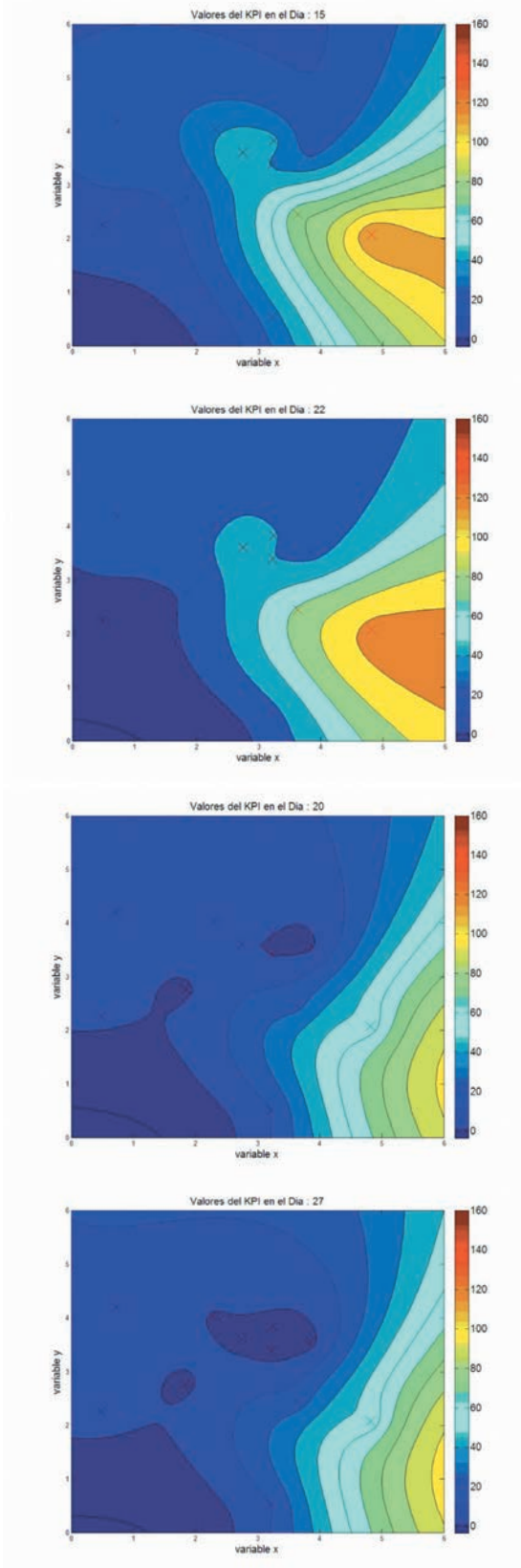


Fig. 6 ALGUNOS CONTORNOS CORRESPONDIENTES A SUPERFICIES OBTENIDAS DURANTE EL PROCESO DE INFERENCIA ESPACIO-TEMPORAL



Los videos pueden ser encontrados y descargados de manera gratuita en las siguientes direcciones: <http://www.mediafire.com/download/43aavdn58wv5r/GPVideoSuperficie.rary> <http://www.mediafire.com/download/dfmx-qlg22a2m3rk/GPVideoContorno.wmv>.

Luego de revisar la herramienta gráfica, se pudo comprobar y corroborar que el comportamiento de la celda 7, se mantenía constante en valores altos la mayor parte del tiempo, así como el hecho que la superficie tendía a comportarse de manera periódica, esto debido a que las series de tiempo asociadas (ver Fig. 2) también lo presentaron.

4. CONCLUSIONES

En este artículo se utilizó una metodología de regresión basada en Procesos Gaussianos para conocer de manera gráfica el comportamiento en el tiempo de un KPI perteneciente a una red de telefonía existente, lo cual al proveer información sobre los patrones tanto del KPI como el usuario, ayuda al operador de red durante el proceso de formulación del plan de expansión. Un ejemplo de esto es la celda 7, dado que en su región vecina presentó altos valores del KPI la mayor parte del tiempo.

El proceso de inferencia espacio-temporal mostró un comportamiento casi periódico que era de esperarse de variables que dependen de las situaciones diarias del usuario. Sugiere en un futuro, pensar en utilizar una función de covarianza del tipo periódico o cuasi-periódico y comparar los resultados.

Como resultado de algunos de los experimentos de regresión se obtuvieron valores pequeños negativos del KPI, lo cual en la realidad carece de algún tipo de significado físico. Para solucionar este problema, se puede plantear un trabajo futuro que incluya restricciones relacionadas con los valores máximos y mínimos que pueda tomar la salida del proceso Gaussiano relacionados a los valores físicos que pueda tomar el KPI.

AGRADECIMIENTOS

Este trabajo se desarrolló en el marco del proyecto de investigación "Análisis de Fallas en Redes de Telefonía Móvil usando reconocimien-

to estadístico de patrones” en el programa de jóvenes investigadores “Virginia Gutiérrez de Pineda”, fue financiado por medio de Colciencias y la Universidad Tecnológica de Pereira.

REFERENCIAS

- [1] C. Steven and P. Samuel, “*On the Expansion of Cellular Wireless Networks*”. École Polytechnique de Montréal, 2002.
- [2] A.R. Mishra, *Advanced Cellular Network Planning and Optimization: 2G/2.5G/3G... Evolution to 4G*, John Wiley & Sons, 2006.
- [3] J. Laiho, A. Wacker, and T. Novosad, *Radio Network Planning and Optimization for UMTS*, 2001 :Wiley
- [4] C. Roberto, I. Tiziano and V. Luca. “Network monitoring and performance evaluation in a 3.5G network”. *Computer Networks* Vol. 51. 2007.
- [5] Sukkhawatchani, P.; Usaha, W., “Performance evaluation of anomaly detection in cellular core networks using self-organizing map,” *Electrical Engineering/Electronics, Computer, Telecommunications and Information Technology*, 2008. ECTI-CON 2008. 5th International Conference, vol.1, no., pp.361, 364, 14-17 May 2008.
- [6] A.K.M. Fazlul, Mohamed Mir, K. Abu. “*Performance Analysis of UMTS cellular Network using Sectorization Based on Capacity and Coverage*” *IJACSA International Journal of Advanced Computer Science and Applications* Vol. 1, No.6, 2011.
- [7] Ye Ouyang and M. Hosein Fallah. “*A performance Analysis for UMTS Packet Switched Network Based on Multivariate KPIs*”. *IJNGN International Journal of Next Generation Network*, Vol.2, No.1, March 2010.
- [8] Kumpulainen P, Särkioja M. et al. “*Analysing 3G radio network performance with fuzzy methods*”. *Neurocomputing* Vol. 107. 2013.
- [9] C.E. Rasmussen and C. Williams, *Gaussian Processes for Machine Learning*. The MIT Press, 2006. 10 antes
- [10] J.Q.Shi, R. Murray-Smith and D.M. Titterington, “Hierarchical Gaussian Process Mixtures for Regression”. *Statistics and Computing*, Springer, 2005.
- [11] A. Schwaighofer, M. Grigoras, V. Tresp and C. Hoffmann, “GPPS: A Gaussian Process Positioning System for Cellular Networks” in book: *Advances in Neural Information Processing Systems 16*. The MIT press, 2004.
- [12] B. Ferris, D. Hähnel and D. Fox, “Gaussian Processes for Signal Strength-Based Location Estimation”. *Proc. of Robotics: Science and Systems*, 2006.
- [13] Liutkus A. and Badeau R. and Richard G., “Multi-dimensional signal separation with Gaussian processes,” *IEEE Statistical Signal Processing Workshop (SSP)*, 2011.
- [14] F. Duvallet and A.D. Tews, “WiFi Position Estimation in Industrial Environments Using Gaussian Processes”, *Intelligent Robots and Systems, 2008. IROS 2008. IEEE/RSJ International Conference on*, vol., no., pp.2216,2221, 22-26 Sept. 2008
- [15] C.M. Bishop, *Pattern Recognition and Machine Learning*. Springer, 2006.
- [16] K. P. Murphy, *Machine Learning a Probabilistic Perspective*. The MIT Press, 2012, p. 1067.
- [17] <http://www.gaussianprocess.org/#code>
- [18] K. Chalupka, “Empirical evaluation of Gaussian Process approximation algorithms”, Master’s thesis, School of Informatics, University of Edinburgh, 2011