

# ANÁLISIS DE LÍNEAS DE ESPERA A TRAVÉS DE TEORÍA DE COLAS Y SIMULACIÓN

## ANALYSIS OF WAITING LINES THROUGH QUEUE SYSTEMS AND SIMULATION

### RESUMEN

En este trabajo se presenta un contraste entre los modelos de la Teoría de Colas y la Simulación. El principal objetivo es evidenciar como estas dos áreas se complementan mutuamente. Lo anterior debido a que con el modelo matemático se puede validar el modelo de simulación y este último permite al investigador profundizar mucho más en el análisis del sistema de colas objeto de estudio. Para observar lo anterior, se presenta como caso de estudio un sistema de atención en una entidad bancaria, el cual está conformado por una línea de espera “fila Preferencial” y un servidor “El Cajero” encargado de atender los clientes respectivos.

**PALABRAS CLAVES:** Modelos, Procesamiento, Proceso Estocástico, Simulación, Sistema de Colas, Utilización.

### ABSTRACT

This paper talks about Queue systems and Simulation. The principal objective is to show how the Queue systems and Simulation are complement. This is possible because the mathematic model validates the results of simulation model. And with the use of simulation, the researcher can deepen in the analysis of the queue system. To observe the above-mentioned, it is showed a queue system in a bank entity, the one which this conformed by a waiting line “Preferential Line” and a location “The Cashier”. He assists the respective clients.

**KEYWORDS:** Models, Processing, Queue systems Stochastic Process, Simulation .

### LILIANA MARGARITA PORTILLA

Administrador Financiero, M. Sc.  
Profesor Asistente.  
Universidad Tecnológica de Pereira  
lilipor@utp.edu.co

### LEONEL ARIAS MONTOYA

Ingeniero Industrial, M. Sc.  
Profesor Asociado.  
Universidad Tecnológica de Pereira  
leoarias@utp.edu.co

### SERGIO A. FERNÁNDEZ HENAO

Ingeniero Industrial, M. Sc.  
Profesor Asistente.  
Universidad Tecnológica de Pereira  
sfernandez@utp.edu.co

## 1. INTRODUCCIÓN

En la mayoría de los procesos que se presentan en las empresas de manufactura y de servicio, aparecen las líneas de espera. Esto debido a que casi siempre, la capacidad de servicio (en algún momento) es menor que la capacidad demandada.

Este proceso de generación de líneas de espera, trae consigo diferentes tipos de inconvenientes que se reflejan a corto y mediano plazo. Por tal motivo, se cuenta con un conjunto de modelos matemáticos que se enmarcan en el área de “La Teoría de Colas”[1]. Estos modelos buscan encontrar el equilibrio entre el número de unidades que se encuentran en la línea de espera y la cantidad de servidores que satisfagan la demanda de servicio.

Existen ocasiones donde es pertinente que el investigador se apoye en la Simulación para analizar de una manera más flexible e integral el fenómeno de la línea de espera. Por tal razón, se muestra en este artículo las ventajas que se tienen al incorporar modelos de simulación en la Teoría de Colas.

Es así, como el artículo presenta una descripción general de lo concerniente a “Teoría de Colas” y “Simulación”, para finalmente presentar un caso modelo donde se evidencia la aplicabilidad de estas dos temáticas trabajadas en conjunto.

## 2. TEORÍA DE COLAS

### 2.1. Generalidades.

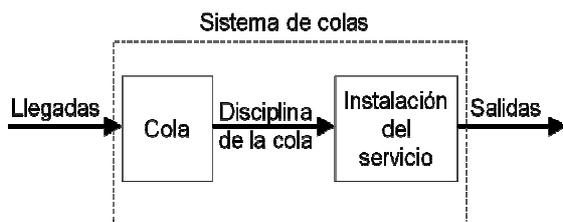
La teoría de colas es un conjunto de modelos matemáticos que describen sistemas de líneas de espera particulares. El objetivo principal es encontrar el estado estable del sistema y determinar una capacidad de servicio apropiada que garantice un equilibrio entre el factor cuantitativo (referente a costos del sistema) y el factor cualitativo (referente a la satisfacción del cliente por el servicio)[2].

Dado lo anterior, los agentes principales que participan en este procesos analítico, son los clientes y los servidores. Entendiéndose por cliente una persona, una orden de servicio, un automóvil, una maquina en espera de mantenimiento, entre otros y el servidor será aquella estación que este en facultad de realizar la respectiva actividad de servicio sobre el cliente, por ejemplo un cajero, una secretaria, una maquina, etc.

La figura 1, presenta un bosquejo de un sistema básico de líneas de espera para una sola cola y un servidor disponible, en donde es claro que cuando el cliente llega al sistema, si no hay nadie en la cola, pasa de una vez a recibir el servicio, de lo contrario, se une a la cola. Es importante señalar que la cola no incluye a quien está recibiendo el servicio.

Fecha Recepción: 9 de Septiembre de 2010

Fecha aceptación: 15 de Noviembre de 2010



**Figura 1.** Modelo básico de una Línea de Espera.

Con base a lo anterior en necesario tener en cuenta algunos componentes claves para ser analizados, los cuales son[3]:

1. Las llegadas de los clientes.
2. La capacidad de la cola.
3. La disciplina de la cola.
4. Los tiempos de servicio.
5. La cantidad de servidores.
6. Las etapas del sistema.

Es por eso que en la Teoría de Colas se utiliza una notación generalizada para indicar el tipo de sistema que se presenta. Esta notación tiene la siguiente forma:

$$A/B/c$$

En donde:

- A: Se refiere a la distribución de probabilidad que siguen las llegadas al sistema.
- B: Se refiere a la distribución de probabilidad que sigue el tiempo de servicio.
- C: Indica la cantidad de servidores con lo que cuenta el sistema.

La tabla 1 presenta la nomenclatura que se debe utilizar conforme al sistema analizado.

CARACTERÍSTICA	SÍMBOLO	EXPLICACIÓN
Distribución de tiempos de llegada (A)	M	Exponencial
	D	Determinista
	$E_k$	Erlang tipo-k ( $k=1,2,\dots$ )
Distribución de tiempos de servicio (B)	G	General

**Tabla 1.** Simbología utilizada en Líneas de Espera.

De esta manera un sistema de Líneas de Espera donde el patrón de llegadas siga una distribución exponencial, el patrón de servicio sea otra distribución exponencial y se cuente con un servidor, tendrá la siguiente notación:

$$M/M/1$$

Por otro lado, también es necesario identificar cuáles son las características operativas del sistema que se van a medir para evaluar la eficiencia respectiva. Entre las más utilizadas se tiene:

- ▶ Número promedio de unidades en la línea de espera.
- ▶ Número promedio de unidades en el sistema.
- ▶ Tiempo promedio de espera.
- ▶ Tiempo promedio de servicio.

Luego de haber citado algunas generalidades de lo concerniente a la Teoría de Colas, se continúa con algunas conceptualizaciones necesarias para el conocimiento integral de dicha temática.

### 2.2. Las Llegadas.

Este concepto hace referencia al análisis de cómo se alimenta el sistema de colas en donde se evalúa variables como el tiempo que transcurre entre dos llegadas sucesivas a dicho sistema.

Este valor es variable, por lo que se conoce como un proceso estocástico. Por lo tanto, es necesario analizar la distribución de probabilidad que presenta dicha variable.

Además de este tiempo entre llegadas, también se requiere analizar la cantidad de clientes que llegan al sistema, ya que puede ser de uno en uno o en lotes. De tal manera, es relevante analizar también la distribución probabilística asociada a la cantidad de clientes esperados que llegan por unidad de tiempo. Esta variable se conoce con el nombre de “Tasa Media de Llegadas” y su parámetro asociado es “ $\lambda$ ” Lambda.

### 2.3. La Capacidad de la Cola.

Es importante conocer de antemano cuál es la capacidad máxima de la cola, es decir, cuantos clientes pueden ubicarse en la línea de espera. Ya que se puede presentar casos en donde el sistema de colas presenta una línea de espera con capacidad limitada, otras donde es ilimitada y otras donde no hay líneas de espera (tal es el caso de un sistema de atención por vía telefónica en donde el usuario es bloqueado y rechazado si la línea telefónica se encuentra ocupada).

### 2.4. La Disciplina de la Cola.

Ésta hace referencia al modo como se acomodan las unidades o clientes en la cola antes de recibir el correspondiente servicio. Entre las formas más habituales se encuentran el sistema PEPS y el sistema UEPS. El primero se refiere a que la primera unidad que llega al sistema es la primera en ser atendida. El segundo indica que el último en ingresar a la cola es el primero en ser atendido. La aplicación de alguno de estos dos sistemas mencionados depende de la naturaleza de la unidad (Por ejemplo un producto no perecedero podrá ser trabajado con sistema UEPS, en cambio un producto perecedero deberá ser operado con un sistema PEPS).

Adicional a los sistemas mencionados, también se puede presentar sistemas de colas en donde la atención se da

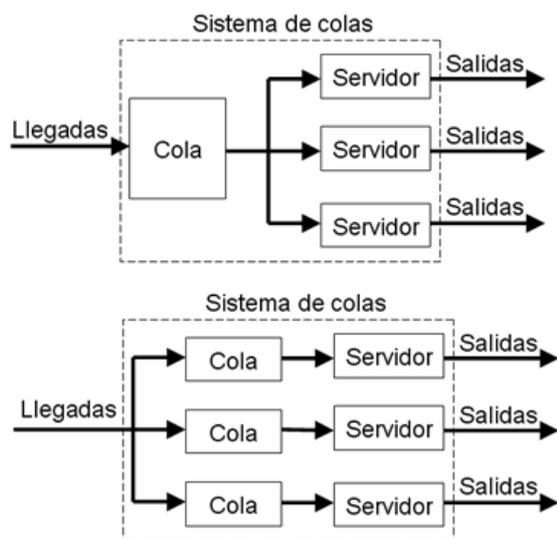
con base a los niveles de prioridad que lleven los clientes (un ejemplo típico de este caso es el sistema de urgencias médicas en un hospital).

### 2.5. Los Tiempos de Servicio.

El servicio puede ser brindado por un servidor o por servidores múltiples. Éste varía de cliente a cliente, por tal motivo es necesario analizar la distribución de probabilidad asociada a dicha variable. El tiempo esperado de servicio depende de la tasa media de servicio la cual es evaluada a través del parámetro ( $\mu$ ).

### 2.6. La Cantidad de Servidores.

En esta fase es importante conocer o identificar cuántos servidores están disponibles para atender los clientes que llegan al sistema. De esta manera se pueden presentar diferentes estructuras de sistemas de colas. La figura 2 presenta dos muy comunes, la primera representa un modelo con múltiples servidores alimentados por una sola cola y la segunda presenta un sistema en paralelo con una cola para cada servidor[4].



**Figura 2.** Estructuras de sistemas de colas.

Las características generales explicadas hasta el momento permiten identificar claramente el sistema de colas sobre el cual se realizarán los correspondientes análisis.

## 3. SIMULACIÓN.

Una técnica para ejecutar estudios piloto, con resultados rápidos y a un costo relativamente bajo, está basado en la modelación de escenarios a través de la simulación. El proceso de elaboración del modelo involucra un grado de abstracción y no necesariamente es una réplica de la realidad; consiste en una descripción que puede ser física, verbal o abstracta en forma, junto con las reglas de operación. Más aún debido a que el modelo es dinámico, su respuesta a diferentes entradas puede ser usada para

estudiar el comportamiento del sistema del cual fue desarrollado.

Por lo tanto al explicar ideas o conceptos complejos, los lenguajes verbales a menudo presentan ambigüedades e imprecisiones. Un modelo es la representación concisa de una situación; por eso representa un medio de comunicación más eficiente y efectivo[5].

Dado lo anterior, es importante tener claro las partes básicas que debe poseer un modelo, estas son[6]:

- Los componentes, son las partes constituyentes del sistema. También se les denomina elementos o subsistemas.
- Las variables, son aquellos valores que cambian dentro de la simulación y forman parte de funciones del modelo o de una función objetivo.
- Los parámetros, son cantidades a las cuales se les asignan valores, una vez establecidos los parámetros, son constantes y no varían dentro de la simulación.
- Las relaciones funcionales, muestran el comportamiento de las variables y parámetros dentro de un componente o entre componentes de un sistema.
- Las restricciones, son limitaciones impuestas a los valores de las variables o la manera en la cual los recursos pueden asignarse o consumirse.
- Las funciones de desempeño, se definen explícitamente los objetivos del sistema y cómo se evaluarán, es una medida de la eficiencia del sistema.

### 3.1. Componentes de la Simulación.

Todo modelo de simulación debe contener unos componentes básicos tal como se ilustra en la figura 3. El componente de "Entidad" se utiliza para referirse a todo lo que el sistema procesa (Una pieza, un producto, una orden, un recibo de pago, etc.).

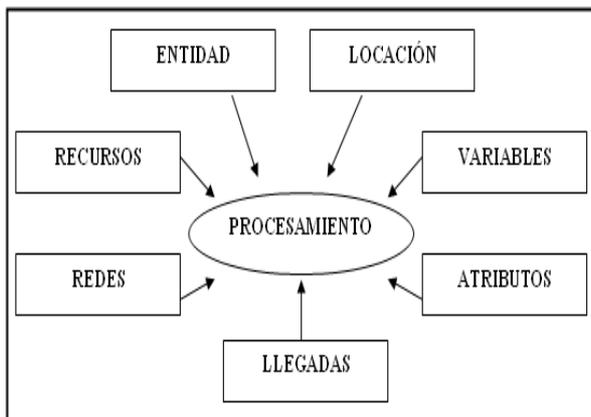
Las "Locaciones" representan los lugares fijos en el sistema a dónde se dirigen las entidades por procesar, el almacenamiento, o alguna otra actividad o fabricación (Una maquina, un área de trabajo, un área de espera, una cola, un ventanilla de pagos, etc.).

Las "Llegadas" indican cada cuanto y en que cantidad llegan nuevas entidades al sistema, esto con el fin de alimentar el sistema y activar su procesamiento.

Un "Recurso" es un operario, o una maquina que sirve para transportar, realizar operaciones puntuales, mantenimientos o asistencias complementarias para el procesamiento de entidades.

Una "Red de Rutas" se utiliza básicamente para construir caminos fijos por los cuales se mueven los recursos

(operarios, maquinas, etc.) para transportar entidades o dirigirse a otras estaciones.



**Figura 3.** Componentes básicos de un modelo de simulación.

Las “Variables” son útiles para capturar y guardar información numérica, de tipo real o entera, para ser utilizada en cálculos de ciertas estadísticas detalladas que puedan requerirse o para ciertos condicionamientos y/o restricciones del sistema analizado.

El “Atributo” es una condición inicial (como una marca), la cual puede ser asignada a entidades o a locaciones; entre ellos pueden contarse el peso de un material, su dureza, o cualquier otra característica ya sea física, química o de cualquier otro tipo que se quiera asignar a una entidad o locación. Este último, también puede utilizarse como medio para obtener información más detallada del sistema, por ejemplo tiempos de ciclo o niveles de eficiencia de laguna estación de trabajo.

El componente de “Proceso” define las rutas y las operaciones que se llevarán a cabo en las locaciones para las entidades en su viaje por el sistema. Generalmente se apoya en los diagramas de proceso u operación que se tienen para cada producto o servicio a simular. Por tal motivo es el último componente que se elabora, ya que necesita de los componentes ya mencionados para vincularlos en su construcción.

#### 4. DE LA TEORÍA A LA PRÁCTICA.

##### 4.1. Información General.

En pro de presentar las utilidades de aplicar los conceptos de la Teoría de Colas a través de la Simulación, se tomó como caso de estudio la cola preferencial de una sucursal bancaria y se analizó las medidas de desempeño de dicho sistema por medio del modelo matemático respectivo, para luego ser comparado con el modelo de Simulación.

Para la toma de la información se realizó en horario que no representaba “horas pico de servicio” y se utilizó

muestreo sistemático tanto para el tiempo de servicio del cajero como para las llegadas de los clientes a la cola[7],[8]. Para el tiempo de servicio se trabajó con una cota de error de 6 segundos y un nivel de confianza del 95%, lo que generó una muestra de 82 observaciones, tal como lo muestra la fórmula 1 (se aclara que la desviación estándar de 0.46 minutos para el tiempo de servicio fue obtenida mediante una prueba piloto).

$$n = \left( \frac{\sigma * Z}{B} \right)^2 = \left( \frac{0.46 * 1.96}{0.1} \right)^2 = 81.28 \cong 82 \quad (1)$$

Para el tiempo entre llegadas se trabajó con una cota de error de 6 segundos y un nivel de confianza del 95%, lo que generó una muestra de 78 observaciones, tal como lo muestra la fórmula 2 (nuevamente la desviación estándar de 0.45 minutos para el tiempo entre llegadas fue obtenida mediante una prueba piloto).

$$n = \left( \frac{\sigma * Z}{B} \right)^2 = \left( \frac{0.45 * 1.96}{0.1} \right)^2 = 77.79 \cong 78 \quad (2)$$

Es así, como se tomaron las muestras respectivas del tiempo entre llegadas de los clientes a la cola preferencial y el tiempo de servicio del cajero que los atendía.

##### 4.2. Aplicación del Modelo de Teoría de Colas.

Con la información tabulada se aplicó una prueba estadística de bondad de ajuste para determinar que distribución probabilística seguían los datos. Para ello, se utilizó el estadístico de prueba Chi Cuadrado en donde la hipótesis nula tanto para los tiempos de servicio como para los tiempos entre llegadas, era que los datos seguían una distribución “Exponencial”. El “P-Value” para los tiempos de servicio fue de 0.267 y para los tiempos entre llegadas fue de 0.348, con lo cual, se concluye en ambas pruebas que la distribución de cada conjunto de datos sigue un comportamiento Exponencial.

Con base a lo expuesto hasta el momento, el modelo de Teoría de Colas a utilizar es el “M/M/1” (ver figura 1). A continuación se presentan los parámetros de “Tasa Media de Servicio” y “Tasa Media de Llegadas” encontrados con la información tomada, con los cuales se obtienen las medidas de desempeño del sistema objeto de estudio a través de las ecuaciones propias del modelo “M/M/1”.

Tasa media de servicio ( $\mu$ ): **0.781** clientes/minuto.

Tasa media de llegadas ( $\lambda$ ): **0.508** clientes/minuto.

4.2.1. Medidas de desempeño. Aplicando los conceptos concernientes al modelo M/M/1, se presenta las siguientes medidas de desempeño con sus correspondientes fórmulas matemáticas:

- Porcentaje de utilización del sistema:

$$P = \frac{\lambda}{\mu} = \frac{0.508}{0.781} = 0.65 \quad (3)$$

- Cantidad promedio de clientes en la cola:

$$L_q = \frac{\lambda^2}{\mu(\mu - \lambda)} = \frac{0.508^2}{0.781(0.781 - 0.508)} = 1.21 \quad (4)$$

- Cantidad promedio de clientes en el sistema:

$$L = L_q + \frac{\lambda}{\mu} = 1.21 + 0.65 = 1.86 \quad (5)$$

- Tiempo promedio de un cliente en la cola:

$$W_q = \frac{L_q}{\lambda} = \frac{1.21}{0.508} = 2.38 \quad (6)$$

- Tiempo promedio de un cliente en el sistema:

$$W = W_q + \frac{1}{\mu} = 2.38 + 1.28 = 3.66 \quad (7)$$

Con los resultados obtenidos se puede observar que la utilización del sistema de colas para la línea preferencial es de 65%, luego hay un porcentaje de tiempo ocioso moderado del 35%. También se observa que el tiempo promedio que transcurre desde que el cliente entra en la cola hasta que sale del servicio es de 3.66 minutos. Adicionalmente, se encontró que la cantidad promedio de clientes haciendo fila en la cola preferencial es de 1.21.

#### 4.3. Aplicación del Modelo de Simulación.

En este numeral se procede a mostrar las características y resultados del modelo simulado a través del Software Promodel Student (Versión 7) en donde se aprecia el comportamiento del sistema objeto de estudio.

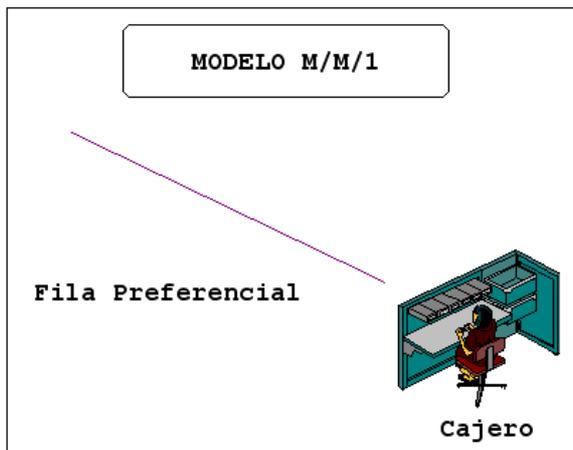


Figura 4. Layout del modelo simulado.

La figura 4 ilustra el diagrama del sistema, el cual se compone de dos locaciones. La primera hace referencia a la fila preferencial donde llegan los clientes denominados especiales o importantes del banco y la segunda estación se refiere al cajero encargado de atender a dichos clientes.

En la primera estación se asumió capacidad infinita (es decir que no hay un tope máximo de clientes que pueden llegar a hacer fila) y en la segunda estación se estableció capacidad de uno, es decir que el cajero solo puede atender un cliente a la vez.

Para alimentar este sistema se codifico el modulo de llegadas con base al valor inverso de la tasa media de llegadas. Es decir que el  $\lambda$  teórico de **0.508** clientes/minuto, equivale a decir que un cliente llega en promedio cada 1.97 minutos a la fila (Ver figura 5).

Entity...	Location...	Qty Each...	First Time...	Occurrences	Frequency
CLIENTE	Fila_Preferencial	1	0	INF	E(1.97)

Figura 5. Módulo de Llegadas.

Para alimentar el módulo del procesamiento se utilizó el valor inverso de la tasa media de servicio ( $\mu$ : **0.781** clientes/minuto), lo que equivale a decir que el cajero se demora atendiendo un cliente en promedio 1.28 minutos. Este proceso se ilustra en la figura 6.

Entity...	Location...	Operation...	Blk	Output...	Destination...
CLIENTE	Fila_Preferencial		1	CLIENTE	Cajero
CLIENTE	Cajero	WAIT E(1.28)			

Figura 6. Módulo de Procesamiento.

Con base a la codificación presentada del modelo a simular, se corrió la simulación durante 20 días en turnos de 8 horas diarios, lo que equivale a 160 horas. La figura 7 presenta un resumen de los resultados que permiten contrastar la veracidad del modelo simulado respecto al modelo teórico M/M/1.

Name	Scheduled Time (HR)	Avg Time Per Entry (MIN)	Avg Contents	% Utilization
Cajero	160,00	1,28	0,65	65,41
Fila Preferencial	160,00	2,65	1,36	0,00

Figura 7. Resumen de resultados del modelo simulado.

Es así, como al comparar las cifras de las medidas de desempeño del modelo teórico con las encontradas en el modelo simulado (ver tabla 2), se evidencia una muy buena representación de la realidad a través de la

simulación ejecutada[9], esto se corrobora gracias a que el margen de diferencia entre los resultados de los dos modelos (Teórico y Simulado) está alrededor del 7.2 %.

PARÁMETRO	MODELO		DIFERENCIA %
	TEÓRICO	SIMULADO	
P	0.65	0.6541	0.62
Lq	1.21	1.36	11.02
L	1.86	2.01	7.46
Wq	2.38	2.65	10.18
W	3.66	3.93	6.87

**Tabla 2.** Comparación de resultados.

Nuevamente, se concluye con los resultados obtenidos en el modelo simulado que este sistema está presentando un tiempo ocioso moderado. El cual, puede ser aprovechado a través de una actividad de refuerzo en donde este cajero puede atender clientes de otras filas mientras no tenga clientes preferenciales en espera de servicio.

## 5. CONCLUSIONES

Con la implementación del modelo de Teoría de Colas y el modelo de Simulación se evidenció como el sistema conformado por la fila de clientes preferenciales y el cajero asignado, está siendo subutilizado, esto debido a que el porcentaje de utilización es apenas del 65%, quedando un 35% libre, el cual puede ser utilizado en refuerzo de otras actividades.

Además de las medidas de desempeño que se encuentran tanto con el modelo teórico como con el modelo de simulación, también se puede analizar otros datos que el modelo de simulación entrega. Por ejemplo, al ver los reportes que la simulación entrega respecto al tiempo promedio que un cliente pasa en el sistema desde que llega a la fila hasta que es atendido por el cajero, se observa como el 42.7% de dicho tiempo es asignado a la atención y el porcentaje restante es asignado al tiempo de espera por el servicio. Esto indica claramente, que el tiempo ocioso del cajero no es por su rapidez en la atención, sino, por el tiempo entre llegadas que se presenta actualmente, es decir, que si este tiempo entre llegadas de clientes preferenciales se volviera más corto, se presentaría inmediatamente un crecimiento significativo en el número de personas en la fila causando un problema serio para la administración en lo que respecta a la satisfacción de sus clientes importantes.

Con lo anterior se puede corroborar una vez más la importancia de la implementación de modelos de simulación, ya que estos permiten profundizar mucho más en el comportamiento del sistema analizado.

Igualmente se resalta la importancia de apoyar la simulación con modelos teóricos, esto debido a que es una excelente forma de validar la representación del modelo simulado con respecto al modelo real, tal como se evidencio en la tabla 2. De esta manera, el investigador

podrá realizar cambios y ajustes al modelo con la tranquilidad de que los resultados obtenidos serán muy acordes con la realidad.

## 6. BIBLIOGRAFÍA

- [1] Giraldo. G. Norman. Procesos Estocásticos. Universidad Nacional de Colombia. Medellín. 2006.
- [2] Hillier, Liberman. Introducción a la investigación de operaciones, Mc. Graw. Hill, 1999.
- [3] Winston, W., "Investigación de Operaciones: Aplicaciones y Algoritmos" Cuarta Edición. Editorial Thomson, 2005.
- [4] Eppen, G. D.; Gould, F. J. y Schmidt, C. P. Investigación de Operaciones en ciencia Administrativa. Edit. Prentice Hall. 1999.
- [5] Carson, Y. Anu, M. "Simulation Optimization: Methods and Applications". Winter Simulation Conference. 1997.
- [6] Soto, J, "Fundamentos Teóricos de Simulación Discreta" Universidad Tecnológica de Pereira, Facultad de Ingeniería Industrial, 2007.
- [7] Freund, Jhon, Estadística Matemática con Aplicaciones, Editorial Pearson. 2006
- [8] Mendenhall, William, Estadística Matemática Aplicada. Editorial MG-Hill. 2007
- [9] Robert G. Sargent, "Verification and Validation of Simulation Models", Winter Simulation Conference, 1998.