

# Metodología de mejora de datos de conteo de pasajeros

**Sebastián Brie, Sebastián Guevara, Herman Moldován**

Universidad tecnológica Nacional, Facultad Regional Buenos Aires,  
Medrano 951, (C1179AAQ) Ciudad Autónoma de Buenos Aires, Argentina

*sbrie@frba.utn.edu.ar*

*Recibido el 15 de Noviembre de 2011, aprobado el 20 de Diciembre de 2011*

---

## Resumen

Se presenta una metodología para mejorar la calidad del conjunto de datos de conteo de pasajeros obtenidos de sensores instalados en las puertas de ingreso y egreso de la totalidad de los vehículos de un sistema de transporte automotor.

El planteo realizado se basa en dividir el problema en tres sub-problemas independientes, cada uno de ellos asociado a distintas estructuras físicas-conceptuales del problema: puerta-vehículo-recorrido. En cada uno de estos niveles se presenta un procedimiento simple para consistir los datos y asimilar información complementaria.

La metodología descrita en el trabajo ha sido desarrollada con el objeto de servir como base para una familia de algoritmos que tome los procedimientos y conceptos aquí enunciados y los adecue a las particularidades de cada sistema de transporte.

**PALABRAS CLAVE:** CONTEO DE PASAJEROS - APC - CONSISTENCIA DE DATOS - SISTEMA DE TRANSPORTE PÚBLICO

## Abstract

An APC (automatic passenger counters) data quality improvement methodology is presented here. The approach developed is based in dividing the problem in 3 (three) independent sub-problems, each one of them related to different physical/conceptual structures: door-vehicle-path. In each one of these levels a simple procedure is introduced for merging the data and assimilating the complementary information.

The methodology described in this work has been developed with aim to serve as a base for an entirely family of algorithms to deal with the particularities of each public transportation system.

**KEYWORDS:** AUTOMATIC PASSENGER COUNTERS - DATA CONSISTENCY -TRANSPORTATION SYSTEM DATA

## Introducción

El crecimiento demográfico de las grandes urbes, su capacidad de tracción de gente, y el constante aumento de la movilidad han puesto a los sistemas de transporte público de pasajeros en el centro de la discusión de políticas públicas de planeamiento urbano desde hace décadas. Para entender estos sistemas históricamente se ha caracterizado la demanda de los mismos realizando estudios de campo puntuales con una determinada periodicidad.

En los últimos años, gracias a los avances tecnológicos, en muchos sistemas de transporte se ha comenzado a monitorear en forma continua los valores de las variables claves para cada vehículo: (a) Posición-Tiempo y (b) Subidas-Bajadas. De esta forma, aquellos sistemas de transporte que posean estos datos medidos en forma continua y precisa cuentan con la información necesaria para dimensionar el sistema, adaptarse a los cambios y en forma general optimizar la operación del sistema (Furth 2006, Dueker 2004).

En este trabajo desarrollamos una metodología para mejorar la calidad del conjunto de datos de conteo de pasajeros obtenidos de sensores instalados en las puertas de ingreso y egreso de la totalidad de los vehículos de un sistema de transporte automotor.

La importancia de generar una metodología que permita corregir los datos generados por estos contadores automáticos reside en que todo sistema físico de medición sufre con el tiempo problemas de balance, calibración y normal deterioro. Sumado a este hecho, la alternativa de verificar en cada vehículo en forma física el correcto funcionamiento y calibración de los elementos es demasiado cara para ser tomada como alternativa logística corriente. Por ello, basándose en campañas de medición periódicas y aplicando la familia de algoritmos que surge de esta metodología, será mucho más económico mantener un sistema de medición dentro de los parámetros aceptables de certeza.

### Definiciones preliminares

A continuación se listan algunos de los términos con el significado que les daremos en el contexto de este trabajo:

**Pasajeros (*pax*):** Personas que ingresan a los autobuses para trasladarse de una parada a otra. Quedan excluidos de esta definición al chofer y/o personal de la empresa de transporte.

**Pasajeros Transportados (*paxT*):** Promedio de ingresos y egresos de pasajeros medidos a un determinado vehículo luego de una determinada cantidad de paradas.

**Vehículo:** Autobús de 2 (dos) o más puertas. El vehículo cuenta con contadores de pasajeros en sus puertas y un sistema de registro de transacciones de validación de boleto.

**APC:** Contador Automático de Pasajeros (CAP), acrónimo de *Automatic Passenger Counter*. Sensor de conteo automático de pasajeros instalado en cada puerta de cada vehículo.

**AVL:** Sistema de Rastreo Vehicular Automatizado (RVA) o AVL, acrónimo de *Automatic Vehicle Location*. Sistemas de localización en tiempo real, basado en el uso de un GPS y un sistema de transmisión.

**Parada:** La parada o estación corresponde al único lugar/ubicación donde los *pax* ingresan y/o egresan a cada vehículo.

**Viaje:** Es el trayecto realizado por cada *pax*. Queda especificado por sus paradas de ingreso/egreso y sus correspondientes horarios.

**Recorrido:** Trayectoria que realiza cada vehículo, diferenciando sentido. Cada recorrido tiene definida una lista de paradas y en particular una parada de inicio y otra parada de finalización. En caso de recorridos en loop se asigna una parada particular como inicial y final.

**Formación:** Es el recorrido que realiza un determinado vehículo siguiendo un horario previamente especificado. Es decir, una formación queda unívocamente determinada especificando: ID del vehículo, Recorrido, Horario.

**Línea:** Servicio de formaciones ofrecido para un determinado recorrido.

**Pax registrado (*paxR*):** Todo *pax* que cuente con boleto que sea validado por el sistema de validación de boletos será considerado como *pax* registrado.

Pax evasor ( $paxE$ ): Todo pax que ingresa a un vehículo sin validar un boleto se lo denominará pax evasor.

Pax medido ( $paxM$ ): Todo pax que ingresa al vehículo y es medido por el sistema de medición de pax del vehículo (APC).

### Sistema de medición

El sistema de medición está constituido por los distintos elementos que permiten obtener y registrar información sobre el funcionamiento del sistema de transporte. De esta forma, identificamos a los sensores de conteo (APC), a los equipos de localización (AVL) y a los dispositivos de registro y validación de boletos.

En este trabajo consideramos que la totalidad de los vehículos cuentan en cada una de sus puertas con un sensor que permite medir subidas y bajadas en forma automática. Cuando un vehículo deja una parada atrás, el sistema de medición de cada puerta consolida las mediciones de pasajeros entrantes y salientes, y envía las cuentas finales a la computadora de a bordo, desde la cual posteriormente se enviará al sistema central.

Se supone, además, que en cada vehículo se cuenta con un sistema de registro de transacciones en el cual los pasajeros pueden registrar/validar su viaje. Cada transacción, junto con la información del sistema AVL del vehículo permite determinar la cantidad de boletos validados en cada parada del recorrido.

### Metodología

Consideramos que el sistema de transporte está constituido por un conjunto de formaciones y tomamos a cada **formación** como unidad de estudio. Los motivos principales de esta decisión son que los datos medidos y sus errores dependen de:

**Sensores:** Cada vehículo se considera que tiene 2 o más sensores cuya calibración resulta obviamente independiente de la calibración del resto de los vehículos del sistema de transporte.

**Recorrido:** El comportamiento de los pax al ingresar/egresar al vehículo como así también sus características físicas depende de cada parada.

**Día/Hora:** Según el día y la hora se espera una problemática diferente.

Asimismo para cada **formación** se cuenta con los siguientes datos, que separamos según su tipo u origen:

1. Información estática:
  - ID del vehículo
  - Cantidad de asientos
  - Ocupación máxima
  - Parada de ocupación nula o mínima
  - Recorrido
  - Horario
2. Posicionamiento (AVL):
  - Para cada parada se conoce el tiempo de arribo ( $parada_i, T_i$ )
3. Contador de pasajeros (APC):
  - Ingresos medidos en cada parada:  $IN_i$
  - Egresos medidos de cada parada:  $OUT_i$
4. Sistema de validación de boletos:
  - Ingresos registrados en cada parada:  $IN_i^{reg}$
5. Otros:
  - Mapa de evasión
  - Conteos ingreso/egresos manuales
  - Curva de demanda por parada
  - Matriz Origen/Destino (OD), encuestas, etc.

Parte de esta información será empleada con el objeto de determinar con la mayor precisión posible los valores de ingreso y egreso de pasajeros en cada parada de una determinada formación. Por otra parte, la metodología plantea tratamientos diferentes para las variables ingreso y egreso. Los principales motivos para esto son:

. Si bien los sensores APC determinan ambas variables, la geometría de las puertas de egreso (preferencial de egreso) es diferente a la de las puertas de ingreso (preferencial ingreso).

. El proceso de ingreso es diferente al proceso de egreso (por ejemplo la velocidad de paso de los  $pax$ ) por lo cual es de esperar tener distintos valores y distribución del error en cada caso.

. La variable ingreso de *pax* es la variable fundamental del problema. Esto, dicho en el sentido de que la variable ingreso es la variable independiente, ya que el egreso en cada parada siempre estará condicionado (o relacionado) al valor de la carga de *pax* en la parada anterior.

. La variable ingreso está relacionada a la demanda de cada parada, siendo la demanda la variable del problema de menor variabilidad intrínseca.

. La determinación de *pax* registrados y la medición de la evasión en cada parada da una forma directa e independiente de estimar el valor de ingresos a cada vehículo. Para poder tener una estimación similar para los egresos en cada parada es necesario incorporar la información adicional contenida en la matriz Origen/Destino de la línea.

### Desarrollo

Considerando la información disponible se propone dividir el proceso de mejora de la calidad de los datos en tres fases o niveles. En cada una de ellas se busca mejorar la calidad de los datos desde el nivel más básico al más integral. Asimismo en cada nivel, se incorpora al análisis información de diferente tipo. En la tabla siguiente se pone en evidencia el concepto de nivel que mencionáramos y la relación con las estructuras físicas-conceptuales puerta-vehículo-formación.

Cada una de estas fases tiene un objetivo diferente como se detalla a continuación:

#### FASE 1: PUERTA - MEDICION

En esta fase se trabaja sobre cada sensor. En cada parada, cada puerta, cada APC genera un valor de ingreso (subidas) y egreso (bajadas) de *pax*. En esta fase se buscan dos objetivos:

- . Mejorar la calidad de cada dato medido
- . Determinar mejor el error de cada medición efectuada

Estudiando esta problemática se observa que para la tecnología actualmente disponible existe un conjunto de situaciones reales para las cuales las mediciones tienen un alto grado de certeza (ejemplo: subidas y/o bajadas individuales), mientras que hay otras donde se incrementa el error y otras donde resulta muy difícil obtener un dato cierto. Considerando este punto se propone que todo sensor APC genere en todos los casos 2 mediciones o componentes. La primera correspondiendo al valor determinado con mayor precisión y la segunda a los valores determinados con menor precisión. De esta forma, el APC de la puerta *p1* de la formación *m (Fm)* generará los siguientes datos de ingreso y egresos de *pax* en la parada *i-ésima*:

(1)

$$\begin{cases} INi_{Fmp1} = (\text{Ingresos\_confiables}, \text{Ingresos\_dudosos}) \\ OUTi_{Fmp1} = (\text{Egresos\_confiables}, \text{Egresos\_dudosos}) \end{cases}$$

En esta fase de la metodología se define el "peso" que tiene el valor confiable y el dudoso en la determinación del valor total. Asimismo, eventualmente en esta fase puede incorporarse la información de "uso" de puerta. Es decir, por ejemplo, considerar que una puerta es exclusiva de egreso y/o que otra puerta es de uso preferente de ingresos.

#### FASE 2: VEHÍCULO - CALIBRACIÓN

En esta fase se trabaja sobre el sistema de medición de cada vehículo.

En este nivel se trabaja con los valores de ingresos y egresos totales (sin diferenciar por puerta) en cada parada al vehículo.

Fases	Problemática	Procedimiento
Nivel 1	Puerta	MEJORA DE LA MEDICION
Nivel 2	Vehículo	CALIBRACION DEL SISTEMA
Nivel 3	Formación	AJUSTE O CONSISTENCIA DE DATOS

**TABLA 1. Relación entre los elementos de la problemática y los distintos niveles o fases del procedimiento de mejora del conjunto de datos.**

A partir de los datos generados por los APC en cada puerta (según se detalla en la Fase 1) y conteos manuales se determinan factores que permitan mejorar la calidad de los datos para las variables ingresos y egresos por parada para cada formación. Los factores de calibración (FCAL) se determinan por regresión lineal utilizando ambos conjuntos de datos: los generados por los sensores automáticos y los medidos manualmente.

Según la disponibilidad de datos o actualización de campañas de conteo manual se definen los factores de calibración a utilizar siguiendo los criterios que se listan a continuación:

1. Los conteos manuales se agrupan por tipo de día (hábil, fin de semana, feriado, otro) y por franja horaria.
  2. Idealmente se determinan los FCAL para cada formación/día/hora.
  3. En caso de formaciones con escasez de muestras manuales se determinará en primer lugar un FCAL general del recorrido y luego se intentará determinar factores diferenciando día/hora.
  4. En todos los casos se utilizarán los FCAL siempre y cuando se valide previamente la significancia estadística de los mismos.
  5. En esta fase se trabaja sobre los valores totales medidos con los APC, es decir ambas componentes que vienen dadas en cada medición. De esta forma los FCAL que se emplean van a modificar los valores de la componente de mayor error pero también los de menor aunque estos últimos en menor medida.
- Por ende, como resultado de esta fase se obtiene:

(2)

$$\begin{cases} IN_{iFm} = (\text{Ingresos\_confiables}^*, \text{Ingresos\_dudosos}^*) \\ OUT_{iFm} = (\text{Egresos\_confiables}^*, \text{Egresos\_dudosos}^*) \end{cases}$$

donde\* indica que los valores ya han sido modificados.

Es importante detenerse aquí y observar que un sistema de transporte real puede tener miles de formaciones, por lo que determinar individualmente los FCAL para cada una de ellas podría volverse impracticable. Considerando esto, los

algoritmos que se generen a partir de esta metodología deberán tener en cuenta este hecho. Por un lado incorporar el conjunto reducido de FCAL calculados siguiendo los criterios arriba listados. Por otro lado, la utilización de variables de control que permitan detectar si en una formación pudieran estar generándose valores inaceptables, para luego sí determinar los FCAL de esa formación particular.

### FASE 3: FORMACION - AJUSTE

En esta fase se trabaja sobre los datos del sistema para una formación.

En este nivel, el proceso de mejora se basa en un tratamiento diferencial de las variables de ingreso y las de egreso. Por una parte, en lo que respecta a la problemática, ambas variables son claramente diferentes, en especial debido a que los ingresos pueden considerarse independientes, en una primera aproximación, desacoplada de la definición del sistema de transporte<sup>1</sup>. Por otra parte, el siguiente análisis se basa en parte en que los datos de ingreso medidos presentan un error inferior a los correspondientes de egreso. Esta fase la dividiremos en los siguientes pasos:

#### Paso 1. Determinación de la cantidad de *pax* registrados en cada parada

Se supone que todo *pax* que ingresa al vehículo con boleto en la parada *i-ésima* lo valida (genera o se registra una transacción) en el trayecto entre las paradas *i* y la parada *i+1*. De esta forma los *pax* registrados de la parada *i-ésima* serán igual a la cantidad de transacciones registradas:  $T_i = paxR_i$

Este valor resulta ser de gran utilidad ya que representa una cota inferior a la variable ingresos en dicha parada, es decir:

$$\sum IN_{iFm_{pj}} \geq paxR$$

#### Paso 2. Estimación de la evasión – *paxE* por formación

La evasión en cada parada se define como la cantidad de ingresos al vehículo que no

<sup>1</sup> En realidad la variable demanda de pasajeros en cada parada, es la que se considera independiente y desacoplada al sistema de transporte. Los ingresos a cada formación obviamente dependen luego de la frecuencia entre formaciones. Si bien esta hipótesis no se verifica perfectamente, suele ser adecuada para cambios menores o leves ajustes en los horarios comunicados a los usuarios.

cuentan con una correspondiente validación de boleto.

Si bien los valores  $paxE$  de una formación no son idénticos a los correspondientes a la siguiente, es posible estimarlos por parada y franja horaria según la información de la que se disponga. Para cada formación, asumiendo que se cuenta con la evasión caracterizada y estimado su error y/o variabilidad intrínseca, los ingresos en cada parada pueden calcularse según la siguiente ecuación:

$$(3) IN_i^b = paxR_i + paxE_i$$

donde el superíndice "b" hace referencia explícita al uso de la información de boletos validados para este cálculo.

Esta determinación de los ingresos para cada parada tiene dos componentes. La primera que vamos a considerar exacta y la segunda con un error a determinar. De esta forma, recorriendo todas las paradas, se obtiene el siguiente resultado para la formación en cuestión:

$$(4) IN^b = \sum_{k=1}^N paxR_k + paxE$$

Aquí nuevamente, el primer término del miembro derecho se lo considera preciso, de hecho en este caso exacto, mientras que el error de  $paxE$  será un estimado cuyo valor y error normalmente son función de la cantidad de ingresos totales.

### Paso 3. Cálculo del factor de corrección de ingresos (FC)

Calculados los ingresos a la formación, a partir de información de transacciones, se lo puede utilizar para ajustar el valor de ingresos medidos en la formación con los APC calibrados siguiendo las fases 1 y 2 antes descriptas. De esta forma:

$$(5) FC_1 = \frac{\sum_{k=1}^N paxR_k + paxE}{\sum_{k=1}^N IN_{kmedido}}$$

donde el numerador representa los ingresos a la formación calculados a partir de datos transaccionales y el denominador corresponde al mismo valor determinado a partir de las mediciones de los contadores APC y su calibración y procesamiento posterior.

De igual forma, en caso de realizar campañas periódicas donde se obtienen conteos manuales (realizados de forma tal de poder considerarlos reales) puede calcularse el mismo factor de corrección ahora basados en los datos de la campaña:

$$(6) FC_2 = \frac{\sum_{k=1}^N IN_{kreal}}{\sum_{k=1}^N IN_{kmedido}}$$

El factor de corrección (cualquiera sea la forma en que haya sido calculado) corresponde a un valor definido para ajustar en forma global los ingresos totales de cada formación. Es decir que, más allá que pueda ser utilizado para ajustar el valor de ingresos en cada parada, tiene sentido real cuando se lo emplea a la sumatoria de los ingresos en cada una de las paradas de la formación.

Esto implica que hemos definido priorizar el valor de los ingresos/egresos totales de cada formación. Es decir la metodología que presentamos, si bien busca generar buenos datos de ingresos/egreso a cada parada, se enfoca en corregir los valores totales de la formación. Esto último también está implícito en el paso siguiente.

### Paso 4. Cálculo del factor de balanceo (FB)

Calculados los ingresos a la formación a partir del ajuste del paso anterior se plantea modificar los valores de los egresos incorporando la siguiente información elemental: todo pasajero que ingresa a la formación debe bajar. Los egresos son una variable dependiente acotada al valor de los ingresos previos a la formación. Bajo las consideraciones anteriores resulta razonable plantear:

$$(7) OUT'_{kmedido} = FB \times OUT_{kmedido}$$

donde:

$$(8) FB = \frac{\sum_{k=1}^N IN_{kmedido}^n}{\sum_{k=1}^N OUT_{kmedido}}$$

### Paso 5. Cálculo final de Ingresos y Egresos

En los pasos anteriores se presentaron distin-



No se hace mención respecto del tipo de la medición: ingreso/egreso, solamente se supone que son siempre del mismo tipo.

Dado que las muestras  $X_{ik}$  son mediciones de baja confianza, se puede plantear aplicar la calibración solamente sobre este conjunto. Se establece entonces la siguiente relación entre las variables en juego:

$$(9) \quad y_k = x_{ck} + \beta_i x_{ik} + e_k$$

con:

$$\begin{cases} \beta_i = \text{coeficiente de calibración de} \\ \text{las cuentas no certeras} \\ e_k = \text{error aleatorio } \sim N(0, \sigma^2) \end{cases}$$

Dada una cantidad de muestras  $N$ , el coeficiente de regresión se estima por cuadrados mínimos al minimizar el error cuadrático medio (Larson, 1982; Walpole, 2002; Kariya y Kurata, 2004):

$$(10) \quad E(\beta_i) = \sum_{k=1}^N (y_k - x_{ck} - \beta_i x_{ik})^2$$

$$\beta_i^* = \frac{\sum_{k=1}^N (y_k - x_{ck}) x_{ik}}{\sum_{k=1}^N (x_{ik})^2}$$

y la estimación de la varianza del error viene dada por la fórmula:

$$(11)$$

$$s_i^2 = \frac{1}{N-1} \sum_{k=1}^N (y_k - x_{ck} - \beta_i^* x_{ik})^2$$

y la estimación de la varianza del estimador del coeficiente de regresión es:

$$(12)$$

$$s_{\beta_i}^2 = \frac{s_i^2}{\sum_{k=1}^N (x_{ik})^2}$$

Queda implícito en este análisis que se está asumiendo, para simplificar, que las cuentas llamadas "certeras" no tienen asociadas un sesgo y por lo tanto, no necesitarían ser calibradas. Esto no será necesariamente así, por lo que un planteo más completo en el futuro, introduciría un coeficiente de calibración también para las cuentas certeras:

$$(13)$$

$$y_k = \beta_c x_{ck} + \beta_i x_{ik} + e_k$$

donde ahora:

$$\begin{cases} \beta_i = \text{coeficiente de calibración de} \\ \text{las cuentas no certeras} \\ \beta_c = \text{coeficiente de calibración de} \\ \text{las cuentas certeras} \\ e_k = \text{error aleatorio } \sim N(0, \sigma^2) \end{cases}$$

Se trata de una regresión múltiple cuya solución se simplifica al utilizar notación matricial y definir las siguientes matrices y vectores:

$$\begin{cases} \beta = \begin{bmatrix} \beta_i \\ \beta_c \end{bmatrix} \\ X' = \begin{bmatrix} x_{i1} & \dots & x_{iN} \\ x_{c1} & \dots & x_{cN} \end{bmatrix} \\ y' = [y_1 \quad \dots \quad y_N] \end{cases}$$

En consecuencia, el error cuadrático se escribe como:

$$(14)$$

$$E(\beta) = (y - X\beta)'(y - X\beta)$$

El estimador de  $\beta$  se encuentra minimizando el error como:

$$(15)$$

$$\beta = (X'X)^{-1}X'y$$

Este planteo no sería del todo correcto debido a que no considera el conocimiento "a priori" sobre las cuentas:

$$\left\{ \begin{array}{l} x_{ck} \text{ son cuentas certeras} \rightarrow \beta_c \text{ debería ser} \cong 1 \\ x_{ik} = \text{son cuentas menos certeras} \rightarrow \beta_i \text{ debería ser} > 1 \text{ (Se supone subconteo)} \end{array} \right.$$

Al no considerar esta información "extra", el realizar la regresión múltiple sin ningún control puede llevar a resultados erróneos. Se plantea en su lugar realizar la regresión en dos pasos según el siguiente procedimiento:

Se deben tomar dos conjuntos de muestras (o se divide una muestra en 2 subconjuntos) y se aplica el siguiente procedimiento de dos pasos:

1. Para el primer subconjunto de muestras se fija  $\beta_c = 1$  y se realiza una regresión lineal simple para encontrar  $\beta_i$  utilizando los valores de  $y_k - x_{ck}$  como variable dependiente y  $x_{ik}$  como variable independiente. Se calculan los siguientes parámetros:

$$\left\{ \begin{array}{l} \beta_i^* = \frac{\sum_{k=1}^N (y_k - x_{ck}) x_{ik}}{\sum_{k=1}^N (x_{ik})^2}, \\ \text{estimación del coeficiente de calibración de} \\ \text{las muestras no certeras} \\ s_i^2 = \frac{1}{N-1} \sum_{k=1}^N (y_k - x_{ck} - \beta_i^* x_{ck}), \\ \text{estimación del error cuadrático medio} \\ s_{\beta_i}^2 = \frac{s_i^2}{\sum_{k=1}^N (x_{ik})^2}, \\ \text{estimación de la varianza del estimador} \\ \beta_i^* \end{array} \right.$$

2. Utilizando el segundo subconjunto de muestras se calcula  $\beta_c$  utilizando ahora a  $y_k - \beta_i x_{ik}$  como variable dependiente y a  $x_{ck}$  como variable independiente. Se calculan los siguientes parámetros:

$$\left\{ \begin{array}{l} \beta_c^* = \frac{\sum_{k=1}^N (y_k - \beta_i^* x_{ik}) x_{ck}}{\sum_{k=1}^N (x_{ck})^2}, \\ \text{estimación del coeficiente de calibración} \\ \text{de las muestras certeras} \\ s_c^2 = \frac{1}{N-1} \sum_{k=1}^N (y_k - \beta_i^* x_{ik} - \beta_c^* x_{ck}), \\ \text{estimación del error cuadrático medio} \\ s_{\beta_c}^2 = \frac{s_c^2}{\sum_{k=1}^N (x_{ck})^2}, \\ \text{estimación de la varianza del estimador } \beta_c^* \end{array} \right.$$

De esta forma, el procedimiento para calibrar los datos de una formación en particular es:

1. Tomar una cantidad de muestras manuales y automáticas de la cantidad de subidas (IN) y bajadas (OUT). No se discriminan las mediciones por puerta o contador siendo todas mediciones de la formación en cuestión.
2. Aplicar el procedimiento para obtener los coeficientes de calibración de las mediciones de IN automáticas.
3. Aplicar el procedimiento para obtener los coeficientes de calibración de las mediciones de OUT automáticas.

Por último, un planteo análogo puede realizarse ahora para cada sensor o puerta y obtener como resultado el conjunto de factores FCAL de cada sensor.

En este último caso, en primer lugar pueden determinarse los FCAL genéricos para un sensor calibrado en laboratorio. Luego, eventualmente, determinar los FCAL de los sensores FCAL(s) que se detecte que así lo requieren.

## Caso de estudio

El estudio de datos correspondientes a una campaña de medición realizada en el sistema de transporte público automotor de la ciudad de Santiago de Chile, nos ha permitido validar parte de las hipótesis aquí propuestas. En particular, el procedimiento que hemos denominado "fase 3" o procedimiento de ajuste puede utilizarse independientemente de los anteriores y mostrar que su aplicación para los estudios de campo realizado disminuye sustancialmente el error definido según se define a continuación:

Para un determinado vehículo definimos el error de la siguiente forma:

(16)

$$Error\_PaxT_{1,2} = \frac{Pax\_T_1 + Pax\_T_2}{Pax\_T_2}$$

(17)

$$Pax\_T_{1,2} = \frac{\sum_{m=1}^{s,d} IN_{1,2} + \sum_{m=1}^{s,d} OUT_{1,2}}{2}$$

donde:

*s* = cantidad de paradas consideradas y *d* = cantidad de puertas del vehículo y los subíndices 1, 2 de ambas expresiones hacen referencia a medidas obtenidas con los sistemas APC de cada puerta y por observadores manuales respectivamente.

Asimismo puede demostrarse que el cociente entre el factor de corrección de ingresos *FC* y el factor de balanceo de egresos *FB* depende solamente de los valores totales medidos:

(18)

$$\frac{FB}{FC} = \frac{\sum_{k=1}^N IN_{kmedido}}{\sum_{k=1}^N OUT_{kmedido}}$$

Esta expresión aplicada a la definición de error realizada permite mostrar que para una determinada muestra de datos el error de toda sub-muestra (*Er*) es modificado siguiendo una transformación lineal:

$$(19) Er' = aEr + b$$

donde *a* y *b* resultan ser valores que dependen únicamente de los ingresos y egresos totales de la muestra. Esta transformación es tal que "ajusta" los errores de las sub-muestras para que su promedio tienda a cero<sup>2</sup>.

Sobre un conjunto de datos provistos se realizaron los cálculos descriptos previamente y se verificó la mejora sustancial de la calidad de datos. En este análisis se utilizaron datos correspondientes a distintas líneas donde se contaba por un lado el valor medido "crudo" por los contadores APC y por otro lado conteos manuales que hemos considerado para este estudio como sin error.

Para todos los casos estudiados se utilizó el tamaño de muestra más grande posible y se verificó que el error de las sub-muestras disminuye.

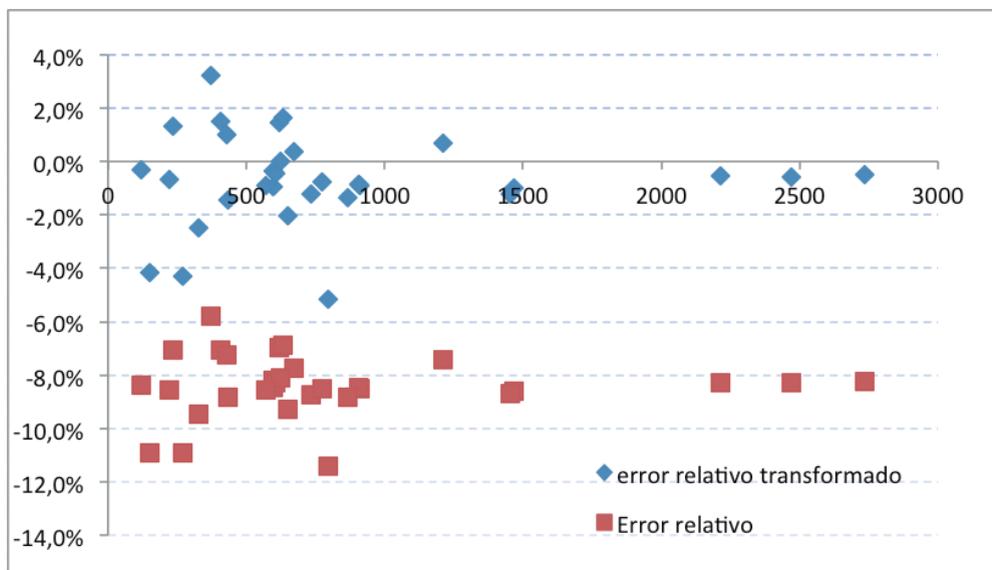
En la Fig. 1 se muestran los resultados obtenidos para el caso de una línea de ómnibus en particular. En este caso la muestra tiene una dimensión de aproximadamente 17.000 ingresos. En esta figura, el eje horizontal corresponde al tamaño de la sub-muestra (definido por la cantidad de ingresos considerados) y en el eje vertical se grafica el error relativo porcentual según fue definido más arriba, tanto para los datos originales como para los ajustados según se indicó previamente.

Asimismo, a modo de ejemplo, la Fig. 2 muestra, tal como indica la ecuación (19), la relación lineal entre el error y el error transformado para la línea de transporte analizada.

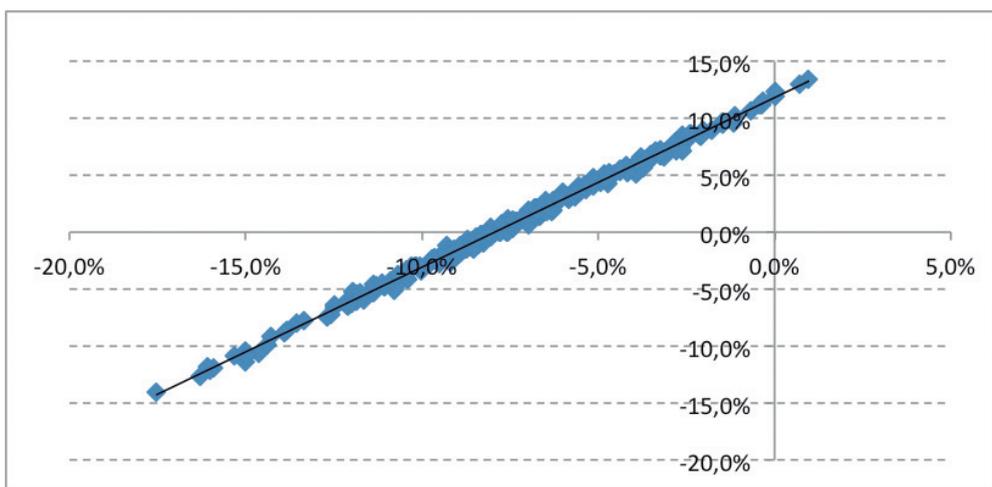
El procedimiento antes descrito fue repetido para otras líneas donde se contaba con compañías de medición manual. Los resultados obtenidos en todos los casos fueron semejantes.

En las líneas de transporte que fueron analizadas, para el conjunto de formaciones consideradas, el volumen de *pax* transportados fue cualitativamente distinto, lo cual nos permite intuir que los resultados hallados son independientes de la línea.

<sup>2</sup> Esto ocurre siempre y cuando los valores medidos manualmente y/o los estimados a partir de información de boletos se los pueda considerar exactos o "real"



**Fig. 1. Errores obtenidos para la línea 509 utilizando datos medidos manualmente y los generados por el APC**



**Fig. 2. Relación lineal entre el error y el error transformado**

## Discusión y conclusiones

Se ha descrito una metodología de mejora de calidad de datos medidos a partir de sistemas automáticos de conteo de pasajeros. La misma supone que se cuenta con información de posicionamiento y validación de boletos de cada vehículo del sistema de transporte. La metodología ha sido presentada con el objeto de servir como base de una familia de algoritmos que tome los procedimientos y conceptos aquí enunciados y los adecue a las particularidades de cada sistema de transporte.

La metodología presentada intenta ser un primer paso en el proceso de mejora, a partir de

la asimilación de datos complementarios como así también constituirse en un planteo genérico donde las particularidades y/o hipótesis adicionales sean realizadas especialmente al formular cada algoritmo de mejora particular. Una de las bases para lograr esta generalidad fue plantear como unidad de análisis cada recorrido realizado por un vehículo.

Por último, es importante resaltar que este trabajo ha sido motivado por un sistema de transporte automotor pero que en todo momento se ha realizado el esfuerzo de generalizar procedimientos y conceptos de forma tal que pueda ser empleado también para la mejora de datos de otros modos de transporte, como por ejemplo un sistema ferroviario.

## Referencias

- FURTH, P. G. (2006) Using Archived AVL-APC Data to Improve Transit Performance and Management, TCRP REPORT 113.
- DUEKER K. J., KIMPEL T. J. AND STRATHMAN J.G. (2004) Determinants of Bus Dwell Time, Journal of Public Transportation, Vol. 7, No. 1.
- LARSON, H. J. (1982) Introduction to Probability Theory and Statistical Inference. s.l. : Wiley Series in Probability & Mathematical Statistics.
- WALPOLE, R. E., et al. (2002) Probability and Statistics for Engineers and Scientists (7th Edition). s.l. : Prentice Hall.
- KARIYA, T., KURATA, H. (2004) Generalized Least Squares. s.l. : Wiley.