# Automatic Passenger Count (APC) Analysis Data Guide

November 22, 2013

# Table of Contents

# 1. Introduction

A key metric for the evaluation of the SF*park* pilot project is transit speed. More parking availability and less circling and double parking should help buses become faster and more reliable—especially on busy commercial corridors. To measure this, SF*park* staff utilized data from the Muni automatic passenger counter (APC) system. The SF*park* analysis involves extracting already-processed data from the existing APC system, loading it into the SF*park* data warehouse, and performing additional processing steps for analysis. Other divisions within the SFMTA have used APC data for various analyses and purposes; however, those analyses are not within the scope of the SF*park* evaluation. This document focuses on the processes and assumptions used for the evaluation of the SF*park* pilot project by the SF*park* team within the SFMTA.

## 1.1 Availability of Data

The APC dataset is available as an Excel workbook. A data dictionary defining key elements of the data is listed in section 6 of this document and is included as a worksheet in the Excel file. For all data requests and related inquiries, please contact info@sfpark.org and put "SFpark evaluation data request" in the subject line.

The APC file name, format and size are as follows:

- File name: SFpark_TransitData_20112013.xlsx
- File format: 2010 MS Office Excel workbook
- File size: 18.4 MB

# 2. Data collection

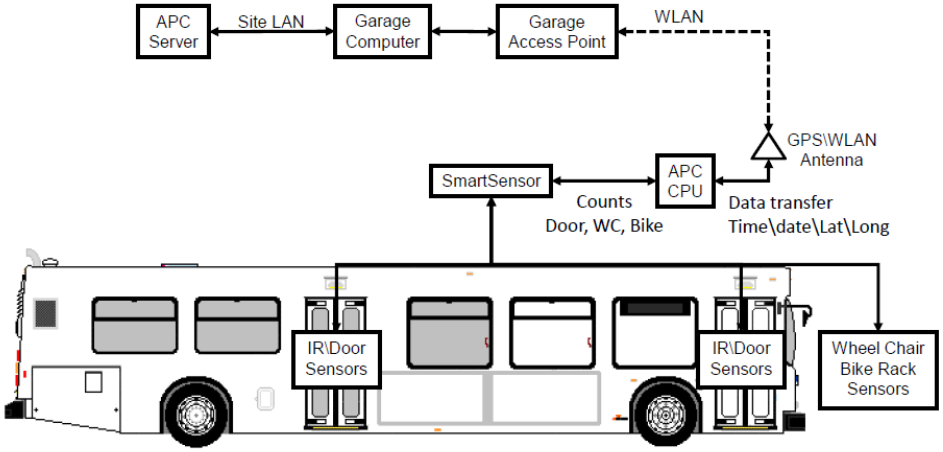
The SFMTA acquired an automatic passenger counter system in 2006 to support an analysis of the Muni system for the Transit Effectiveness Project (TEP), an ongoing program to improve transit travel speeds and reliability. The vendor, Urban Transportation Associates, provides on-board equipment and centralized servers for handling incoming data, and an algorithm for processing raw data.

## 2.1 Equipment

The APC system includes:

- A sensor that detects when doors open and close
- Infrared sensors installed at the front and rear doors that counts passengers boarding and alighting from vehicles
- Sensors that detect wheelchair and bike rack activity
- A GPS receiver that measures latitude and longitude

An on-board computer combines sensor data with a date/timestamp and location data. Once the bus enters the yard at the end of the day, the data is wirelessly transmitted to a division server. Data is then transmitted to the centralized APC server. The following diagram shows the configuration and flow of data:




## 2.2 Deployment

The SFMTA developed and implemented a sampling plan to ensure that vehicles with APC equipment were deployed to collect data from each line in a statistically significant way. The deployment of APC equipment currently covers 30 percent of the rubber-tire fleet.

The APC equipment has not been installed on any light rail vehicles, historic trolleys, or cable cars.

## 2.3 Data points

The centralized APC server processes the raw data points and associates them with predefined Muni stops and vehicles. This allows the data to be analyzed using meaningful dimensions (e.g., route, direction, trip). The centralized APC server stores data in a series of tables. The SF*park* analysis relies primarily on the "Stops" table, which includes (but is not limited to) the following:

- Stop-to-stop travel time by vehicle
- Stop-to-stop travel distance by vehicle
- Boardings and alightings by stop by vehicle
- Load by stop

# 3. SFpark processing for analysis

Two key metrics for the SF*park* evaluation are average transit speed and variability of speed. Additionally, these need to be measured only for those portions of selected bus routes through SF*park* pilot and control areas. This necessitated a new analytical unit – a segment – that was neither the entire length of a route nor stop to stop, but rather a portion of continuous stops along a route that falls within an SF*park* area. As an additional control to the SF*park* study areas, the remainder of the chosen routes was analyzed as well. This section outlines the additional steps undertaken by the SF*park* team for analysis.

## 3.1 Technical processes

Various technical/data warehousing steps were taken to further process APC data to support the SF*park* evaluation. The data was copied from the centralized APC server to the SF*park* data warehouse. Additional joins were created to properly associate stops with their respective routes and trips. Some aggregate tables were built to improve report performance.

## 3.2 Speed calculation

Speed is calculated as change in distance over change in time, and measured in miles per hour. Dwell time is excluded from this calculation for the SF*park* analysis.

The SF*park* team used the stop-to-stop travel time and stop-to-stop travel distance to calculate speed. Specifically, speed between Stop A and Stop B is calculated as follows:

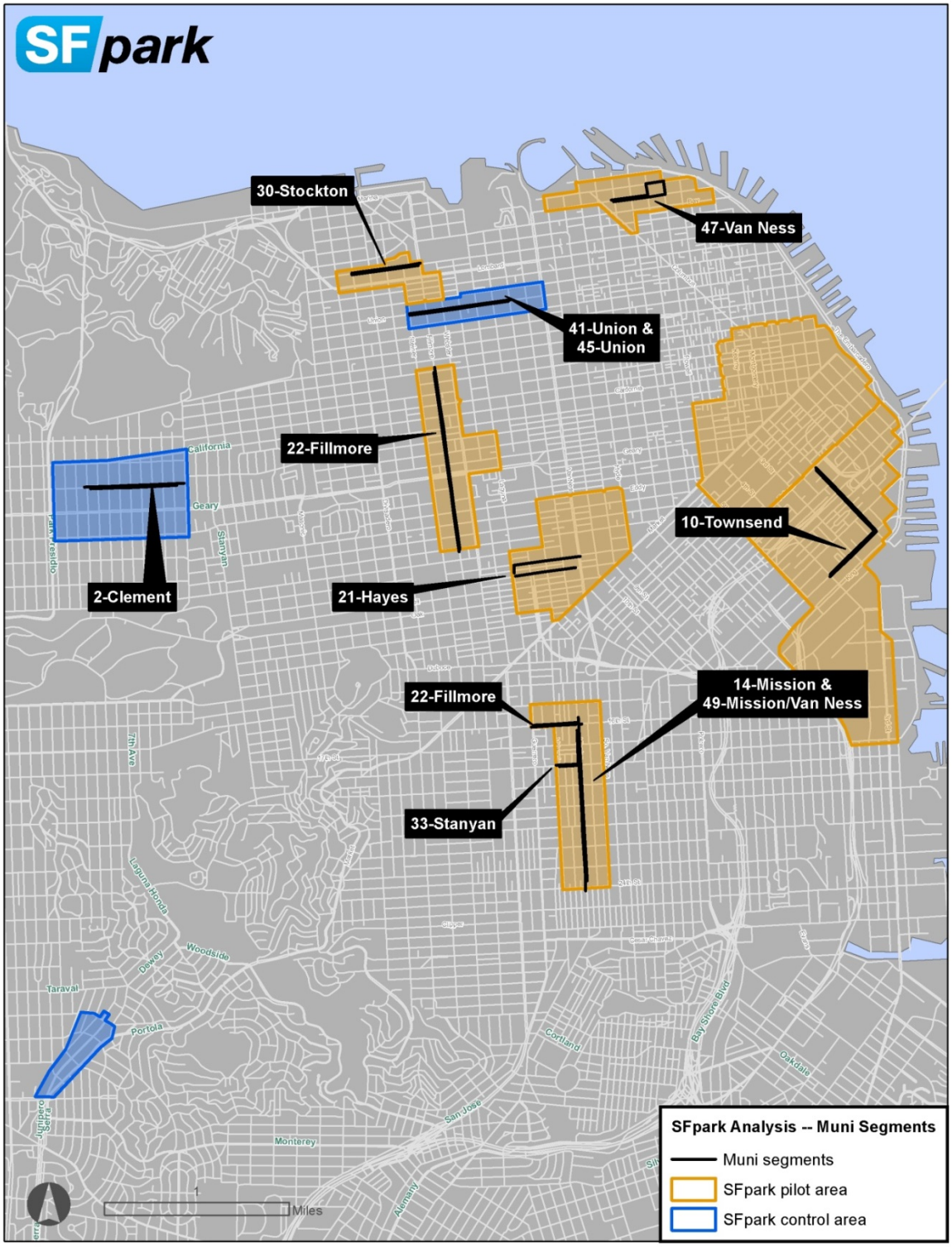
$$\frac{\text{Total distance traveled between Stop A and Stop B}}{\text{Total time elapsed between doors closing at Stop A and doors opening at Stop B}}$$

## 3.3 Segment definitions

The SF*park* team defined segments using the following criteria:

- Muni line that falls within a pilot or control area and has at least 3 continuous blocks that have parking sensors installed.
- Within pilot areas, blocks where SF*park* is conducting rate adjustments. This excludes many blocks/lines in Downtown.
- Where data exists (APC is a sample of bus and trolley bus lines only, no rail).
- No major service changes across or during evaluation periods (Spring 2011, 2012, 2013).

The following map shows the pilot and control segments:



### 3.4 Segment Areas

The following eleven bus routes were chosen because they travel through significant portions of control and pilot areas:

- 10 Townsend
- 14 Mission
- 2 Clement
- 21 Hayes
- 22 Fillmore
- 30 Stockton
- 33 Stanyan
- 41 Union
- 45 Union-Stockton
- 47 Van Ness
- 49 Van Ness-Mission

From the bus routes, SF*park* identified thirteen segments within the following areas:

- Pilot
  - Marina
    - 30-Stockton on Chestnut between Fillmore and Divisadero
  - Fillmore
    - 22-Fillmore on Fillmore between Jackson and McAllister
  - Civic Center
    - 21-Hayes on Hayes between Van Ness and Laguna
  - Mission
    - 14-Mission on Mission between 16th St and 24th St
    - 49-Van Ness-Mission on Mission between 16th St and 24th St
    - 33-Stanyan on 18th and Mission between 18th/Valencia and 16th/Mission
    - 22-Fillmore on 16th between 16th/Mission and 16th/Guerrero
  - South Embarcadero
    - 10-Townsend on 2nd St between Folsom and Townsend
    - 10-Townsend on Townsend between 2nd and 5th St
  - Fisherman's Wharf
    - 47-Van Ness on North Point between Jones and Powell
- Control
  - Richmond
    - 2-Clement on Clement between 11th Ave and Arguello
  - Union
    - 41-Union on Union between Steiner and Gough
    - 45-Union-Stockton on Union between Steiner and Gough

The remaining portions of the chosen bus routes that were not identified as pilot or control area segments were categorized as Non-SF*park* segments.

# 4. Data filters

## 4.1 Quality control columns

The vendor's processing of APC data includes the creation of quality control ("QC") variables that scores the consistency of the collected raw data and allows the inclusion or exclusion of trips for specific analytic purposes. For instance, APC for a particular vehicle's trip over the course of service day, the ridership figures may be skewed if the rear door sensors fail. However, the travel time information may be analytically valid. The SF*park* team worked with SFMTA's transit service planning staff to filter the collected the APC stop level transit data based on the same QC thresholds used for other SFMTA analyses:

- **Assignment Quality Code.** This variable measures the consistency of the expected and actual path a bus takes over the course of its schedule. The threshold is set to greater than or equal to 80 percent to account for GPS signal variation, but to exclude non-typical trips that may include a major detour, or road call that interrupts service.
- **GPS diagnostic.** This variable measures the consistency of vehicle collected GPS data with each stop's known position and it provides a way to identify APC devices with faulty GPS devices. For this analysis, the threshold is set to less than or equal to 10 to be consistent with other APC performance and ridership analyses.
- **Passenger counts.** This variable measures the percentage difference between a vehicle's expected and tabulated total boardings and alightings. A high difference can be an indicator of faulty equipment. For this analysis, the threshold is set to less than or equal to 10 to be consistent with other APC ridership analyses.

The SF*park* team performed a row count before and after filtering for these QC columns. SF*park* recommends removing data records where the speed is listed as 0 or null.

# 5. Considerations for analysis

The following lists potential considerations for analysis of the data:

- Due to less frequent service on weekends, samples for Saturdays and Sundays are considerably smaller relative to weekday data.
- For purposes of the internal SF*park* evaluation, the SF*park* team decided to exclude segments in the Mission pilot area due to construction between March and August 2012 that rerouted buses from Mission Street to South Van Ness. No APC data exists for this period. Additionally, sensors

have not been reinstalled along Mission Street, and there have been no demand-responsive price changes on this street since early 2012.

- Routes in the South Embarcadero pilot area may have been impacted by special events such as Giants baseball games, especially during the 2012 World Series that attracted more traffic to the area.

## 6. Data Dictionary

Each data point contains the following 19 attributes:

| FIELD NAME | DEFINITION | EXAMPLE |
|---|---|---|
| PERIOD | The period in which the data point was recorded, according to the season and year of collection. For each year, these periods include only the months of April, May and June.[1] SF*park* made this decision due to the lack of data for March 2013.[2] | Spring 2013 |
| PARKINGMANAGEMENT DISTRICT | This field specifies whether the segment ran in one of the six pilot areas, or one of the two control areas, or in a non-SF*park* area. | Fillmore |
| BUS_SEG_DESC | Lists the full name of the bus route, the direction of the bus, and the stops binding the segment. For portions of routes that do not fall within SF*park* pilot and control areas, the segment description is "Non-SF*park*." | 22-Fillmore, Outbound, Fillmore/Jackson - Fillmore/McAllister |
| DATE_DT | Exact date of data collection. | 6/23/2013 |
| TRIP_ID | The unique number associated with the trip a bus is scheduled to serve from one stop to another, every day, for a particular time band. The same trip ID is repeated for each day of week. | 805 |
| DAYOFWEEK | Either Weekdays, Saturday or Sunday. Weekdays are not specified. | Sunday |
| TIMEBAND | The twenty-four hour time range in which the data collection took place. | 0700-0900 |
| TRAVELDISTANCEMILES | The length of the trip distance, in miles. | 0.98 |
| TRAVELTIMEMINUTES | The duration of the trip, in minutes. | 6.8 |
| SPEED | The calculated bus speed for the trip, measured in miles per hour. Dwell times are excluded from this calculation. | 8.64705882352941 |

[1] The SFMTA recommends looking at three months' worth of APC data for the scope of this analysis.
[2] Due to an issue with the APC vendor.

| | | |
|---|---|---|
| LOAD | The net number of passengers on the bus. | 53 |
| BOARDINGS | Lists the number of passengers recorded as boarding the bus at the stop. | 8 |
| ALIGHTINGS | Lists the number of passengers recorded as alighting the bus at the stop. | 5 |
| QC_ASSIGNMENT | See page 6 for section on quality control columns. | 98 |
| QC_COUNTS | See page 6 for section on quality control columns. | 2 |
| QC_GPS | See page 6 for section on quality control columns. | 0 |
| BUS_LINE_DESC | The name of the bus route being analyzed. | 22 Fillmore |
| BUS_DIR_DESC | Lists the direction of the bus—either inbound or outbound. | Outbound |
| PMD_AREA_TYPE | Lists whether the data point was recorded in a pilot or control area. If the data point does not fall in a pilot or control area, it is listed as “Non-SF*park*.” | Pilot |