September 11th Memorial Scholarship Program
“Taxi Travel Estimation and Calibration Modeling Tool”

Stanislav Parfenov
New York City Department of Transportation, Division of Traffic & Planning, Modeling & Data Analysis Unit
Presented on September 18, 2013

Outline

1. Introduction
2. Simulation Modeling
3. Taxi GPS and Model Calibration
4. Findings
5. Data Visualization
6. Future Steps
7. Conclusions
8. Gratitude

I would like to honor the memory of three NYMTC employees who perished in the tragedy on September 11, 2001 in the World Trade Center:

Ignatius Adanga

Charles Esperance

See Wong Shum

Introduction

- Personal Info:
  - Polytechnic Institute of New York University (NYU Poly), Master of Science in Traffic Engineering and Planning (Graduated May 2013)
  - Graduate Advisor - Professor Elena Prassas, Ph.D.

- Topic Name:
- "Taxi Travel Estimation and Calibration Modeling Tool (TTEC MT)"

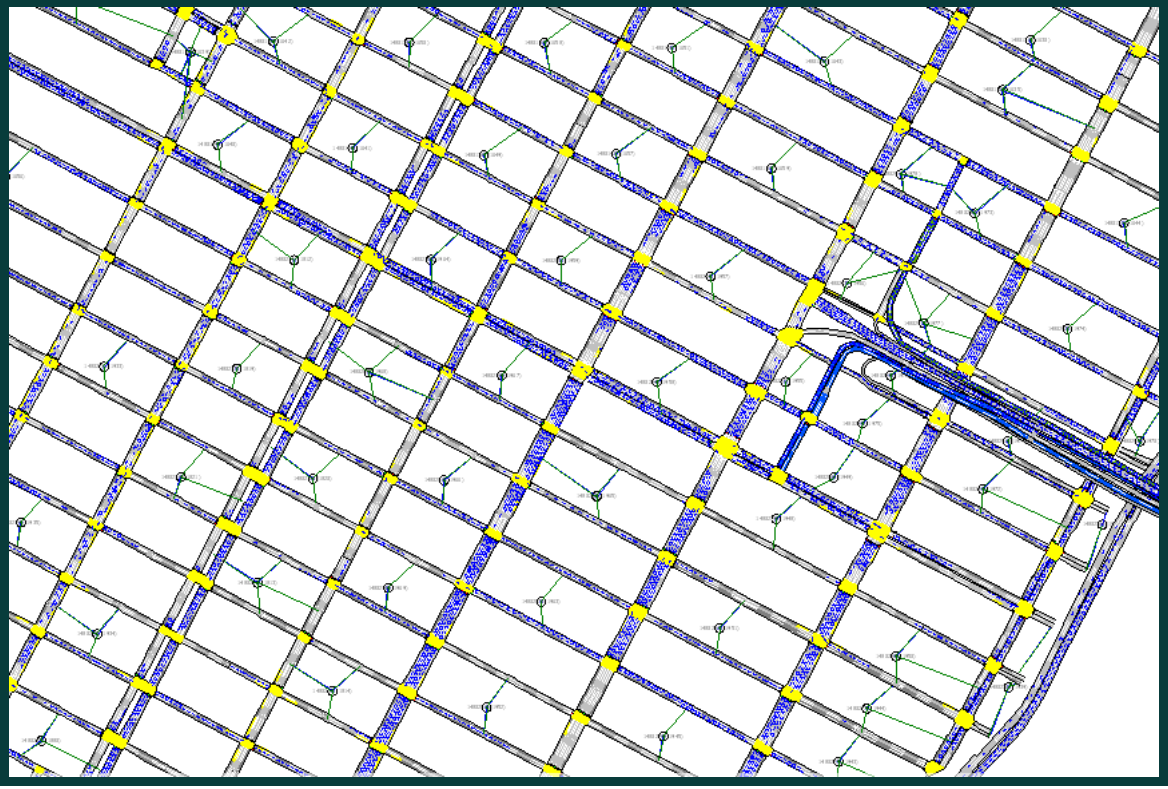
- Place of Internship:
  - New York City Department of Transportation (NYCDOT), Division of Traffic & Planning, Modeling & Data Analysis Unit
  - Supervisor – Michael Marsico, P.E.

Introduction

- Project's Main Points:
  - The project focuses on use of computer simulation models, innovative data calibration techniques and ways to increase predictability accuracy of traffic simulations.
  - Taxi GPS data is used for data validation purposes, as an innovative data source, capable of providing large data sample and reliable results.
  - Project focuses on the area of Manhattan Travel Model (MTM) project area, which is located in Manhattan - river to river from 14th St to 66th St.
  - Crosstown streets are specifically of interest as existing system prioritizes the avenues and north-south streets, while behavior of crosstown streets is less understood and there is an interest to see effects on those streets.

Simulation Modeling

- Aimsun 7 dynamic-simulation platform developed by TSS has been used to create the traffic simulation network known as Manhattan Travel Model (MTM).
- MTM model first developed in 2011 to assess crosstown bus travel, and went through recent updates.



Simulation Modeling

- The benefit of the simulation modeling is that it allows to see and predict the future network conditions and travel flows based on different scenarios.
- However, it requires feeding of new data into the Origin-Destination (OD) Matrix in order to keep the model up to date.
- Data could be updated from Turning Movement Counts (TMC), video traffic counts (Miovision), infrared readers, Radio Frequency Idenitification (RFID) chips, Automated Traffic Recorders (ATRs).
- However, most of these are rather costly and time consuming

- New innovative technique is to use Taxi GPS data to analyze traffic patterns and use it to validate simulation models

Taxi GPS

- Currently, NYC Taxi and Limousine Commission (TLC) collects travel data from all medallion taxi vehicles
- Data is spatially georeferenced to each individual trip
- Travel data that is collected for each trip includes, but not limited to: travel time and distance, average speed, fare and toll paid, and other variables.
- Data is processed for each month of the year and scrubbed for errors and outliers
- Each month has about 13 million records constituting a little under 2.5 GB in size

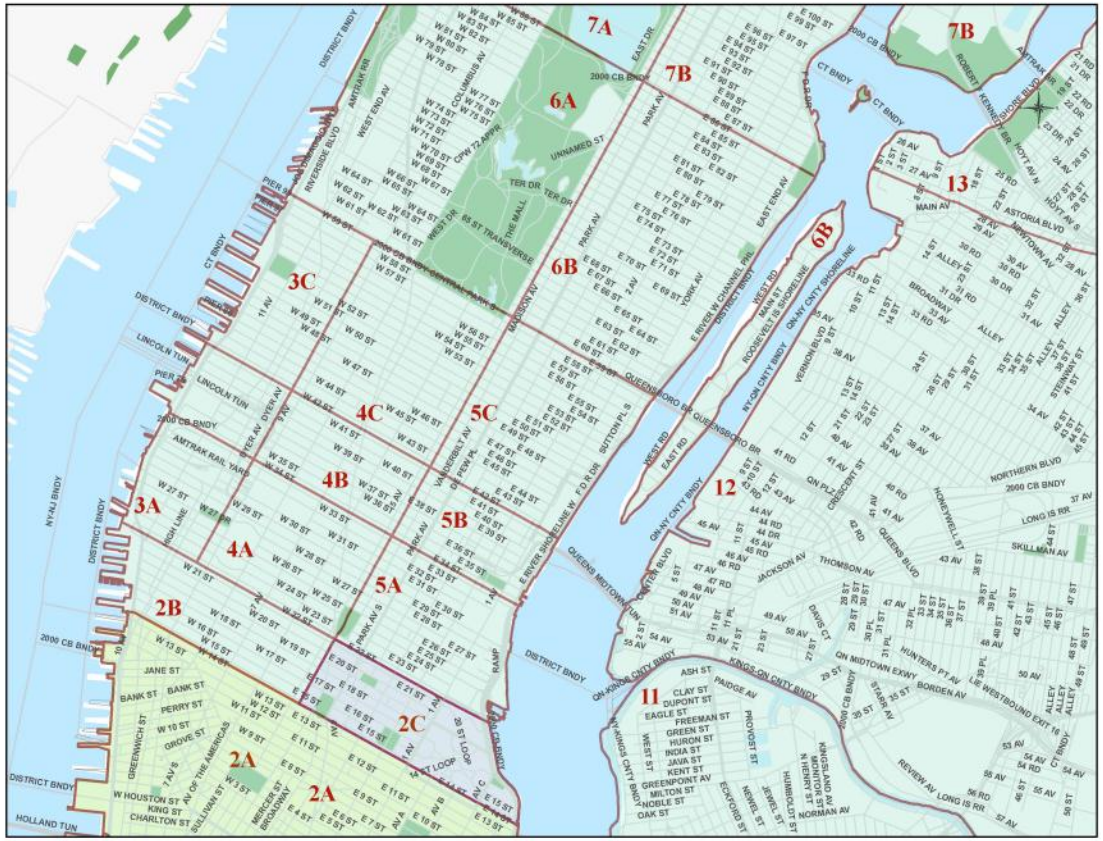
- For this project, data from 2012 and 2011 were used, which comprises about 16.4 GB of data generated by this project.

Taxi GPS

FareAmt	TipAmt	Surcharge	TotalAmt	TripTime	TripDistance	StartLon	StartLat	EndLon	EndLat	Speed	VendorName	TripNumber	PickUp
6.90	.00	.00	7.40	10.00	1.18	-73.980107	40.746412	-73.984002	40.754977	7.08	VTS	9830	5A
4.90	.60	.00	6.00	6.00	.89	-73.984600	40.780058	-73.987670	40.770133	8.90	VTS	31201	6A
6.90	2.00	.00	9.40	11.00	.88	-73.985150	40.756980	-73.975238	40.750480	4.80	VTS	26559	4C
5.70	.00	.00	6.20	8.00	.84	-73.982442	40.767328	-73.969525	40.760705	6.30	VTS	26768	4C
11.30	.00	.00	11.80	15.00	3.09	-73.982540	40.740647	-73.960962	40.777542	12.36	VTS	19813	5A
7.70	4.00	.00	12.20	12.00	1.28	-73.982422	40.768145	-73.974557	40.757210	6.40	VTS	21295	4C
6.50	3.00	.00	10.00	9.00	1.02	-73.990005	40.752143	-73.985945	40.758812	6.80	VTS	7431	4B
7.70	.00	.00	8.20	12.00	1.68	-73.974607	40.783158	-73.981818	40.766228	8.40	VTS	25993	6A
6.90	.00	.00	7.40	10.00	1.04	-73.991047	40.750822	-73.985628	40.740010	6.24	VTS	20090	4A
6.10	.00	.00	6.60	9.00	.58	-73.982290	40.768172	-73.973342	40.763710	3.87	VTS	26008	4C
7.30	.00	.00	7.80	10.00	1.54	-73.976700	40.788392	-73.988935	40.771930	9.24	VTS	7047	6A
6.10	.00	.00	6.60	8.00	1.26	-73.974993	40.760872	-73.974245	40.750442	9.45	VTS	14552	4C
10.10	.00	.00	10.60	19.00	1.27	-73.962283	40.779012	-73.968763	40.764188	4.01	VTS	13506	6A
11.30	1.00	.00	12.80	19.00	2.28	-73.979493	40.784570	-73.960507	40.762547	7.20	VTS	2572	6A
4.50	2.00	.00	7.00	5.00	.78	-73.984807	40.741782	-73.984102	40.735522	9.36	VTS	18073	5A
6.90	2.00	.00	9.40	10.00	1.27	-73.978567	40.762497	-73.992973	40.757815	7.62	VTS	24874	4C
5.30	1.00	.00	6.80	6.00	.55	-73.982068	40.770407	-73.978467	40.774490	5.50	VTS	7602	6A
8.10	.00	.00	8.60	12.00	1.57	-73.972312	40.765318	-73.973048	40.780147	7.85	VTS	2041	6A
7.70	.00	.00	8.20	11.00	1.74	-73.977447	40.753127	-73.982525	40.771555	9.49	VTS	26896	5C
12.50	.00	.00	13.00	25.00	.82	-73.963745	40.777147	-73.971753	40.766525	1.97	VTS	24845	6A
10.90	.00	.00	11.40	18.00	2.27	-73.998287	40.744540	-73.973320	40.756943	7.57	VTS	22364	2B

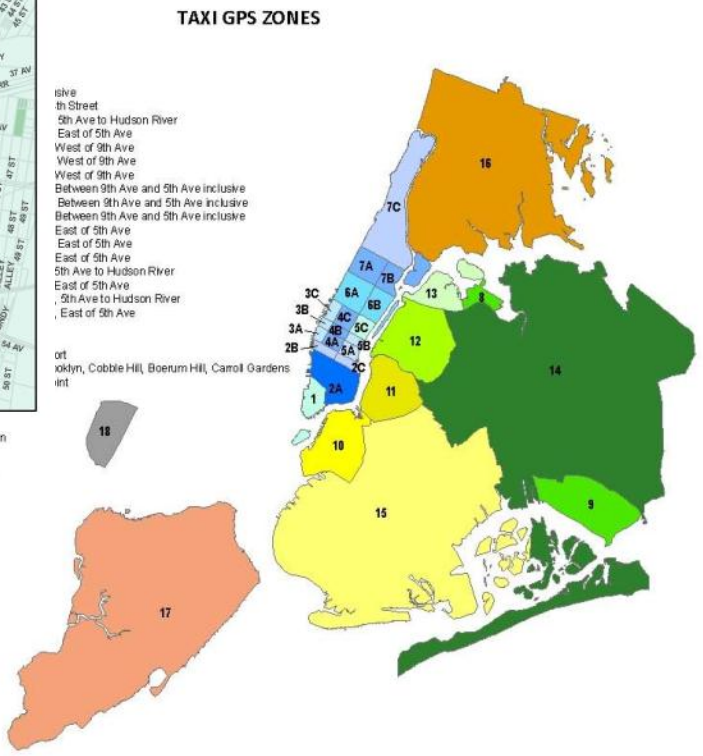
Sample Monthly Taxi GPS Data Table

Taxi GPS



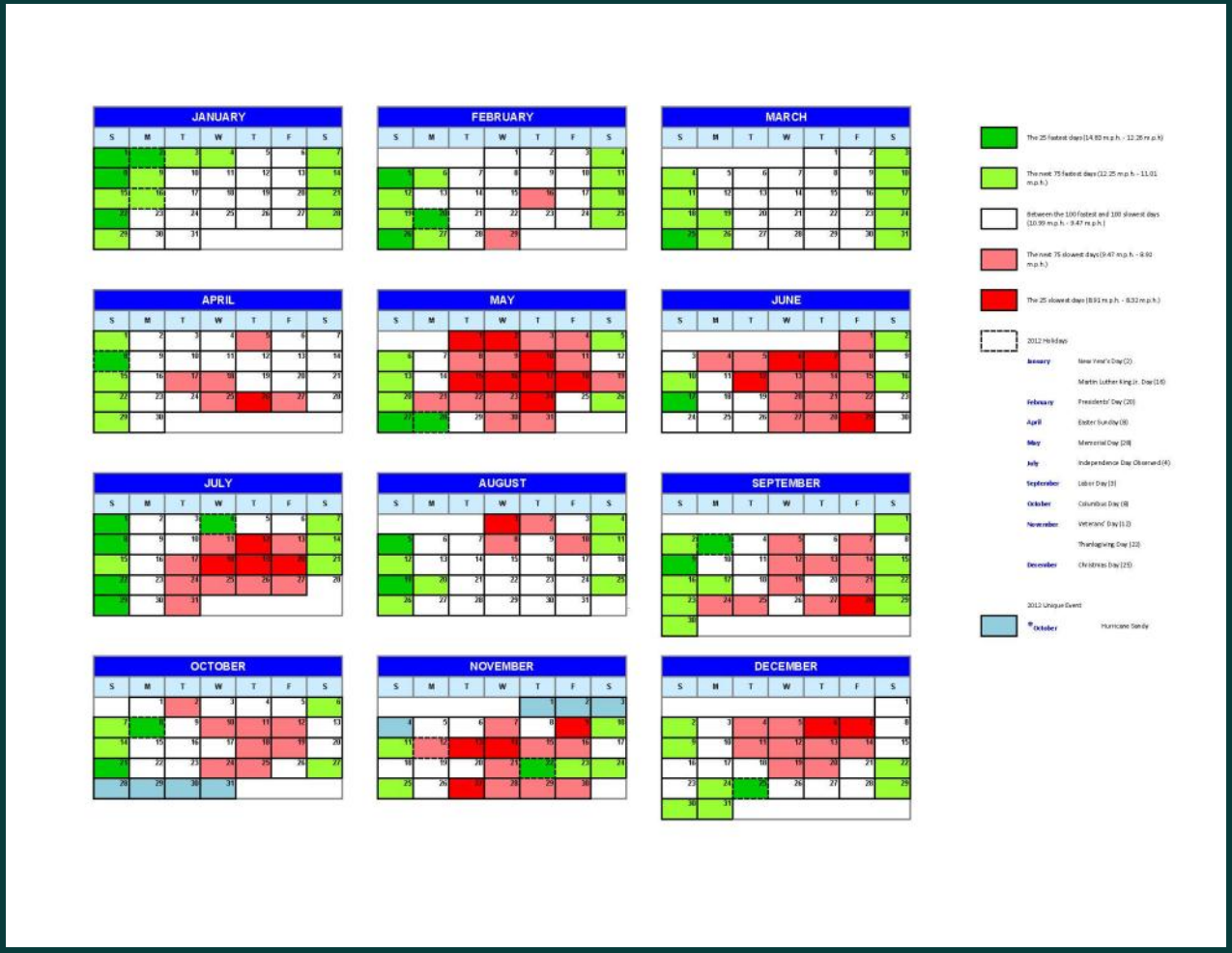
Taxi GPS Zones Manhattan

Taxi GPS Zones NYC



Taxi GPS

Taxi GPS could be used to analyze data for a specific day, week, month, season and/or year



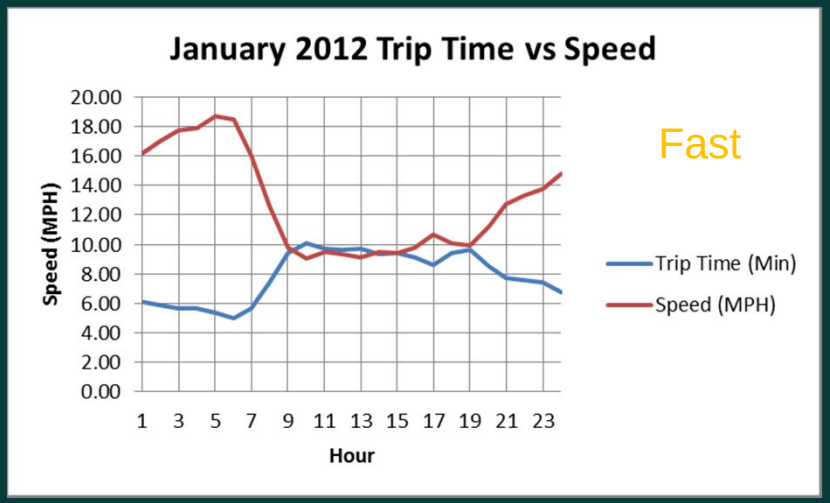
2012 Taxi GPS Calendar

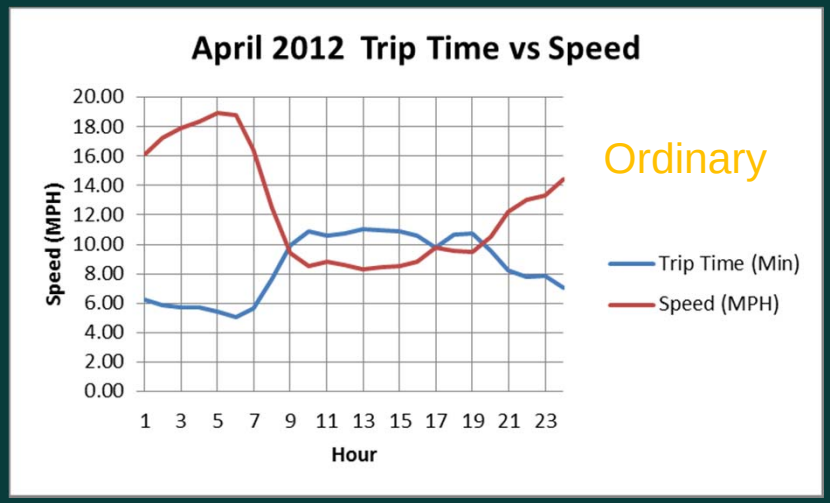
Findings

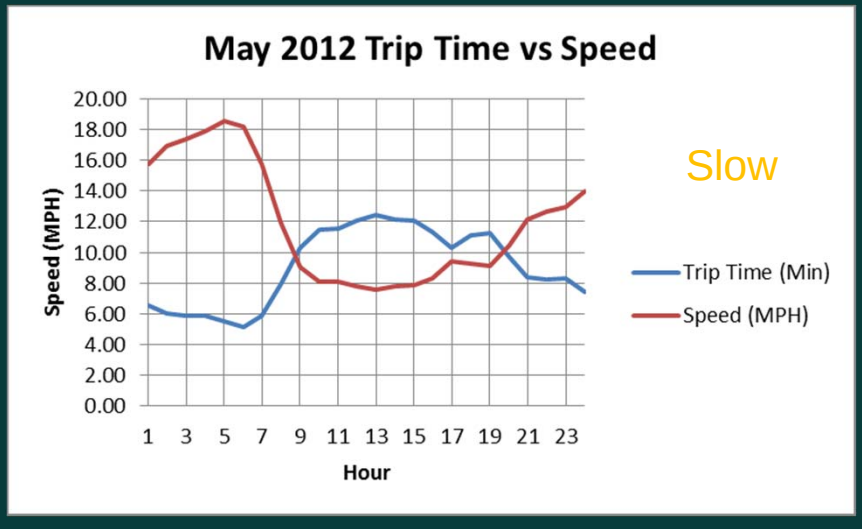
- Months of the year could be grouped into distributions based on travel speed as: Fast, Ordinary and Slow
  - Typical Fast Month is January
  - Typical Ordinary Month is April
  - Typical Slow Month is May
- These months could be thought of as different traffic pattern scenarios, where travel behavior differs based on month of the year
- Subtle differences provide us insights on how model will behave and this should be taken into account when models are created

Findings: Speed Patterns

Months of the year could be grouped into distributions based on travel speed as: Fast, Ordinary and Slow

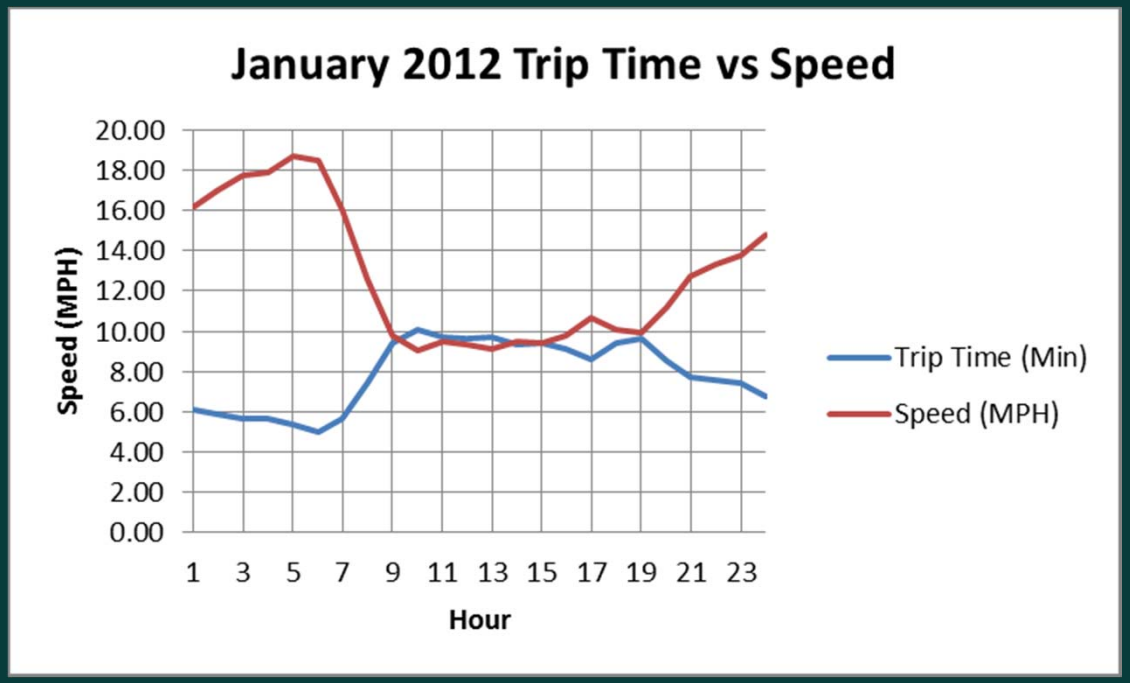






Findings: Speed Patterns

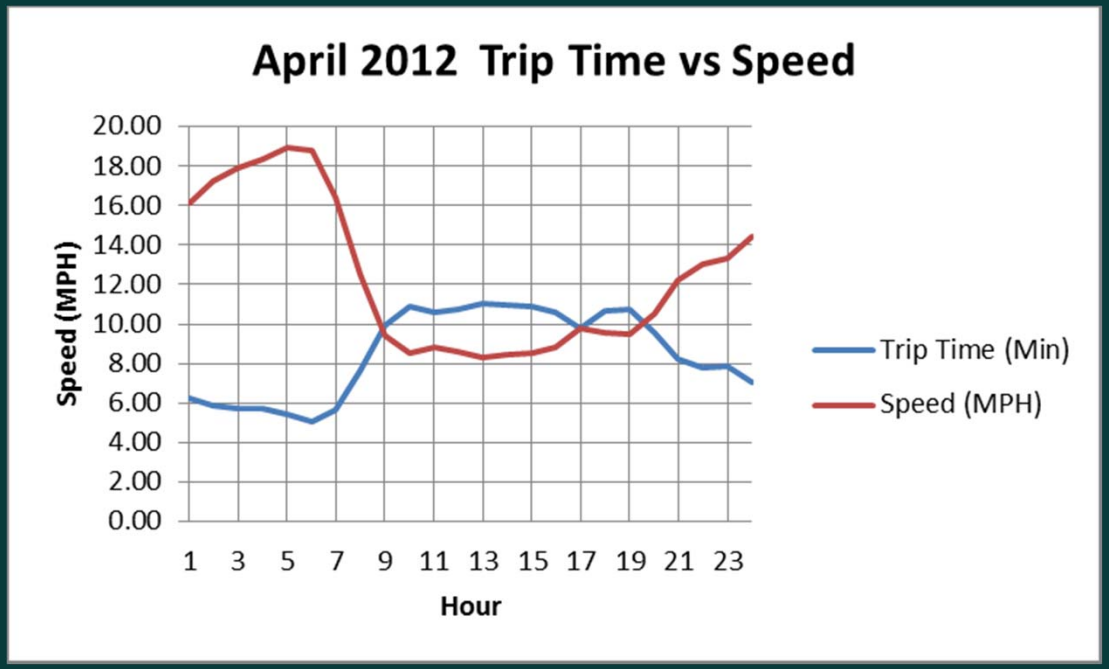
Fast Months – January, February, March



Average:	Total Trips	Trip Time (Min)	Trip Distance (Mile)	Speed (MPH)	Fare Amount ($)
January	118,160	7.86	1.48	12.78	6.74

Findings: Speed Patterns

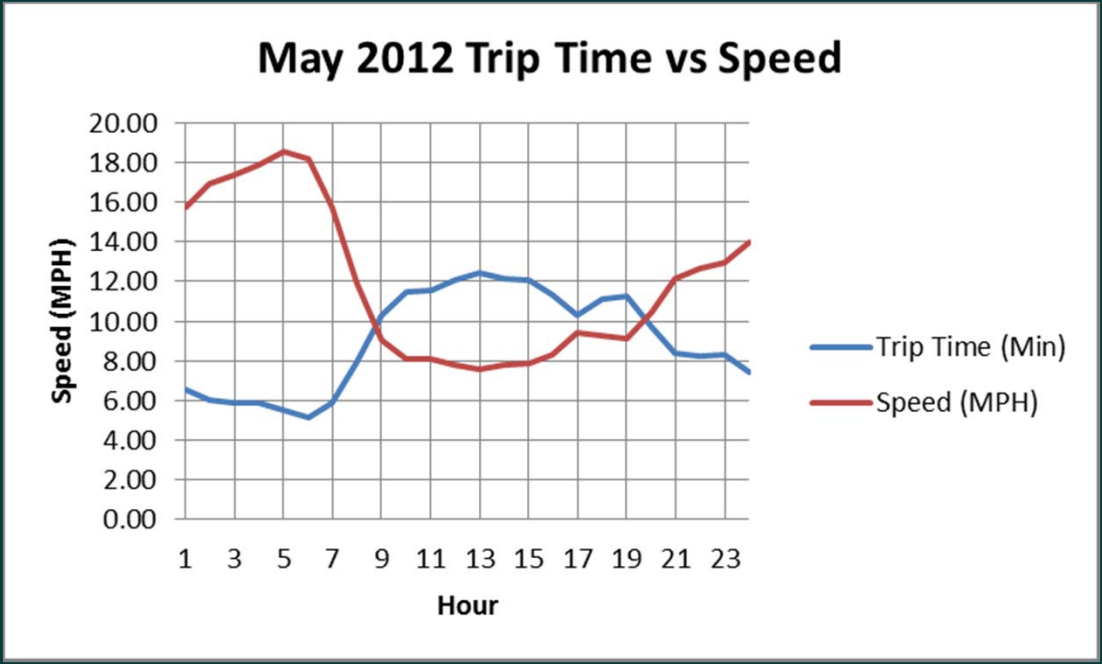
Ordinary Months - April, August, October, December



Average:	Total Trips	Trip Time (Min)	Trip Distance (Mile)	Speed (MPH)	Fare Amount ($)
April	111,227	8.53	1.50	12.41	7.01

Findings: Speed Patterns

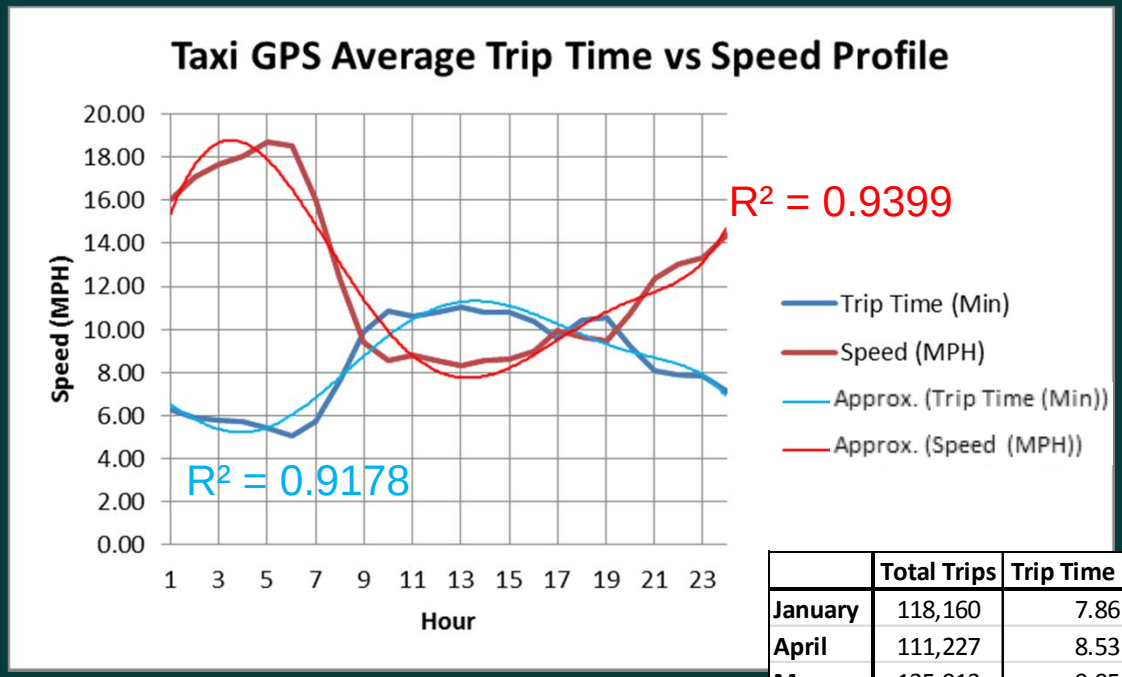
Slow Months – May, June, July, September, November



Average:	Total Trips	Trip Time (Min)	Trip Distance (Mile)	Speed (MPH)	Fare Amount ($)
May	135,913	9.05	1.50	11.96	7.19

Findings: Speed Patterns

- Overall, taxi behavior in Manhattan CBD could be described on average using the Trip Time vs Speed curves



	Total Trips	Trip Time (Min)	Trip Distance (Mile)	Speed (MPH)	Fare Amount ($)
January	118,160	7.86	1.48	12.78	6.74
April	111,227	8.53	1.50	12.41	7.01
May	135,913	9.05	1.50	11.96	7.19
Average	121,767	8.48	1.49	12.38	6.98

- Trip Speed and Trip Time could be described using 6th Degree Polynomial equations with $R^2 = 0.9399$ and $R^2 = 0.9178$ respectively, showing really high R^2, indicating a good fit.

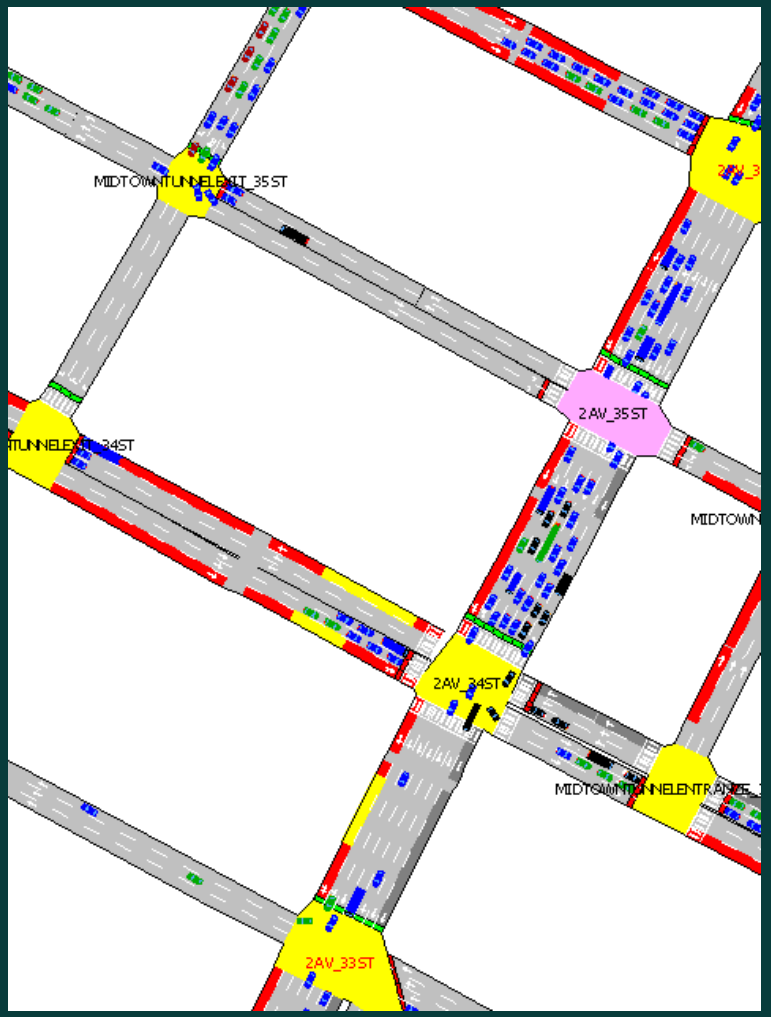
Findings: Crosstown Streets

- E/W streets are less studied than avenues, where detectors and progressions implemented.
- Taxi GPS data analysis can be used to gain understainding of speeds on E/W streets, calibrate the simulation model

- Case study: 34th St was used to analyze the pattern and compare it to the model's speeds.

Findings: Crosstown Streets

- Calibrated MTM 6 to 9 AM model was ran in Aimsun 7
- Several iterations were performed
- Special script was used to extract network specs and imported into ArcGIS.
- Taxi GPS data for January, April and May of 2012 was selected and examined for trips along 34th, 35th and 36th Streets only.



Findings: Crosstown Streets

Results:

- Comparing the results, Aimsun network crosstown speeds were within 10% of the taxi GPS data
- Network behaves closer to Fast month conditions based on Taxi GPS data
- Technique required use of Aimsun, SPSS and ArcGIS software

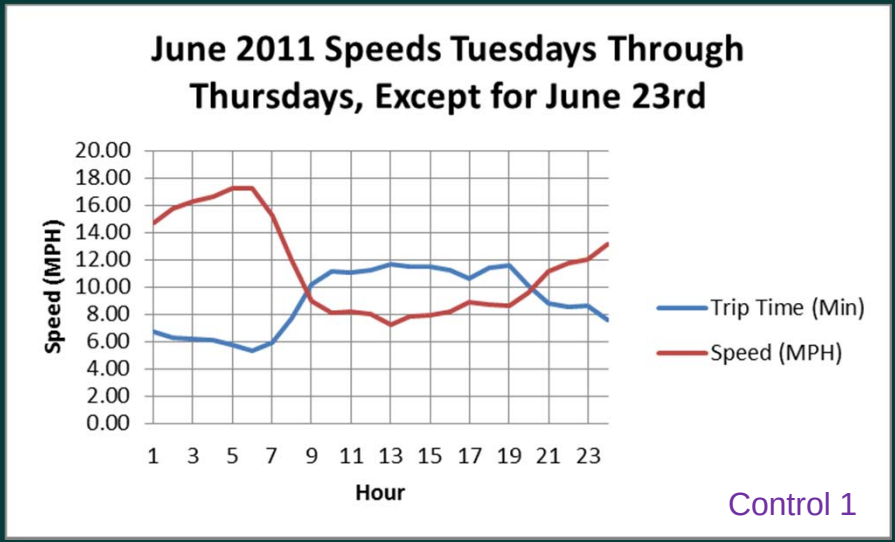
8-9 AM Peak Hour Scenario							
Corridor	**January Taxi Speeds**	**% Difference**	**April Taxi Speeds**	**% Difference**	**May Taxi Speeds**	**% Difference**	**Aimsun Speeds**
36th St	9.60	8%	9.12	12%	9.08	13%	10.42
35th St	8.34	7%	8.85	1%	8.49	5%	8.92
34th St	10.93	-8%	11.10	-10%	10.58	-5%	10.08

Findings: Road Closures

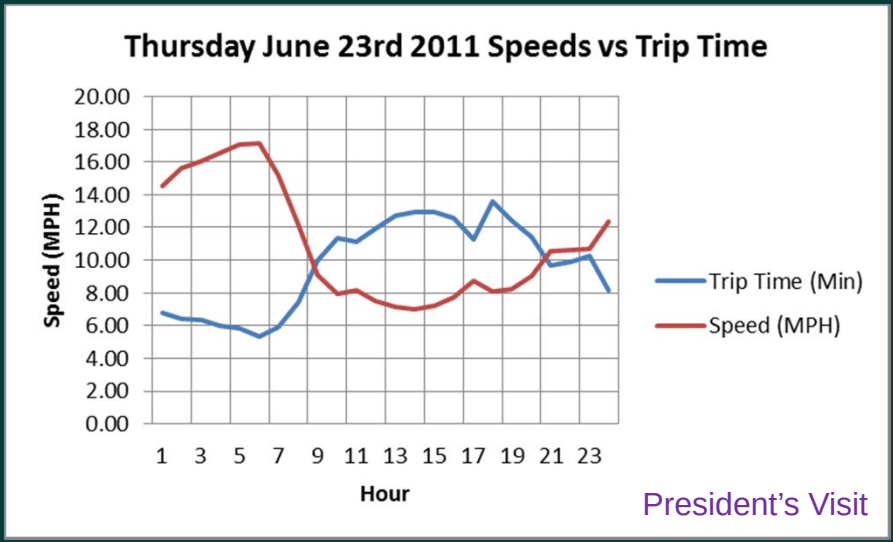
- Taxi GPS data can be effectively used in calibrating models that are used to simulate special events, which involve traffic re-routing and road closures.
- On June 23 2011 President Obama arrived to NYC for a fundraising event.
- Taxi GPS data was used to analyze the effects of special event. First, the base conditions were established for month of June 2011 for workdays Tuesday through Thursday and then compared to the conditions of Thursday, June 23.
  - Mr. President arrived at 5:10 PM at Lower Manhattan and then travelled to Midtown via car for events there, returning back and flying back from Lower Manhattan at 11:05 PM.
  - Taxi GPS data was used for the entire day to monitor the changing traffic conditions with a special focus from 5 to 11 PM.

Findings: Road Closures

Normal Profile vs Road Closures:



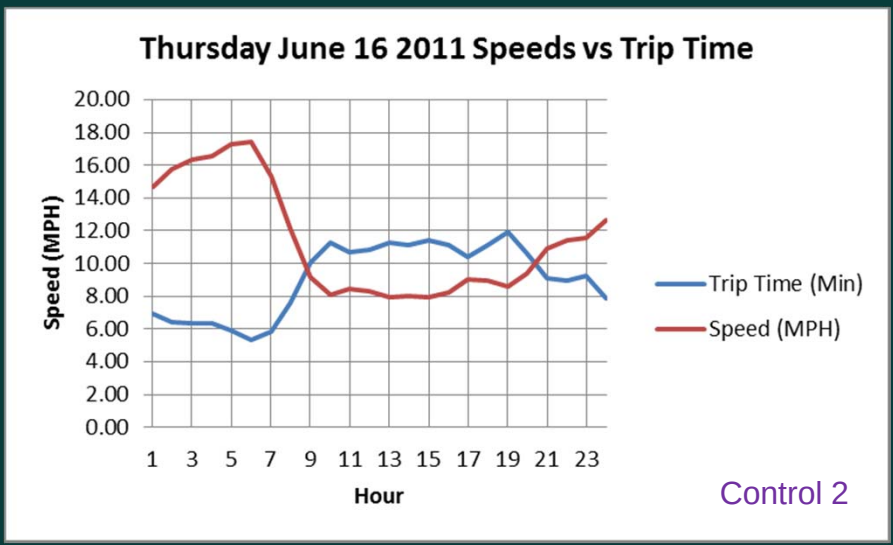
Control 1



President's Visit

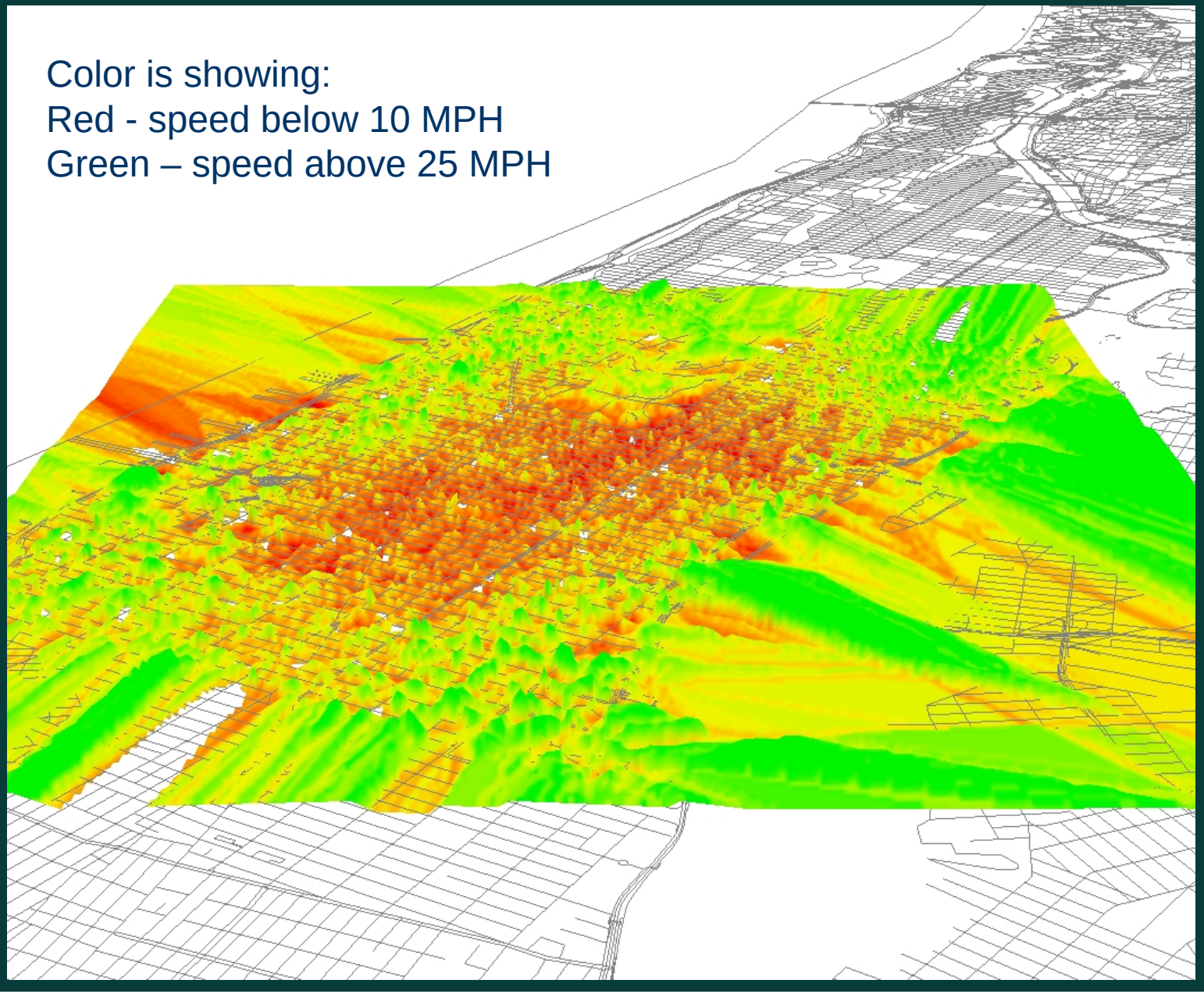
During President's Visit (5 – 11PM) compared to normal conditions, network experienced:

- 3,831 fewer taxi pick-ups
- Trip Time increased by 6.76 minutes
- Trip Distance decreased by 0.07 miles
- Speed decreased by 3.44 MPH
- Fare Amount paid increased by $2.58



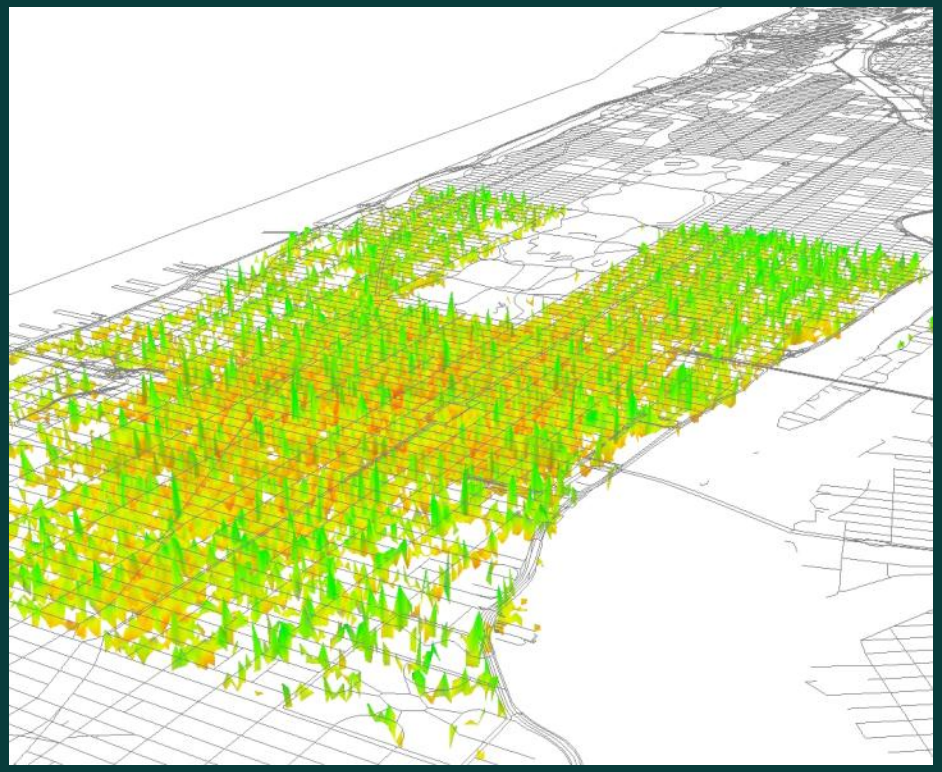
Control 2

Data Visualization

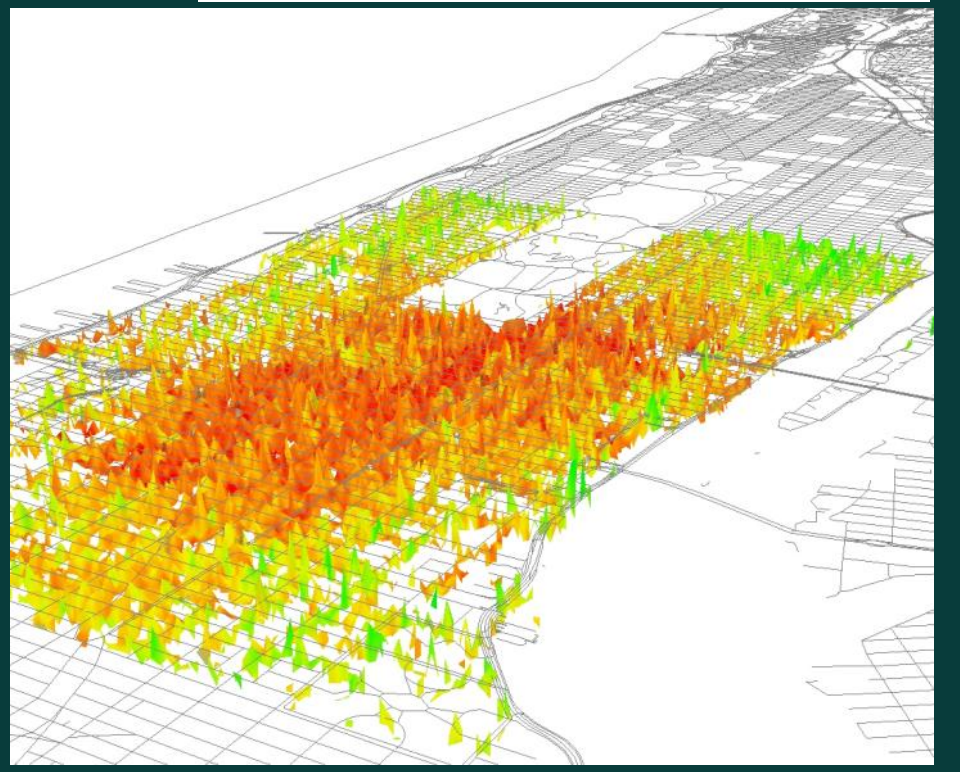


Data Visualization

Color is showing:
Red - speed below 10 MPH
Green – speed above 25 MPH

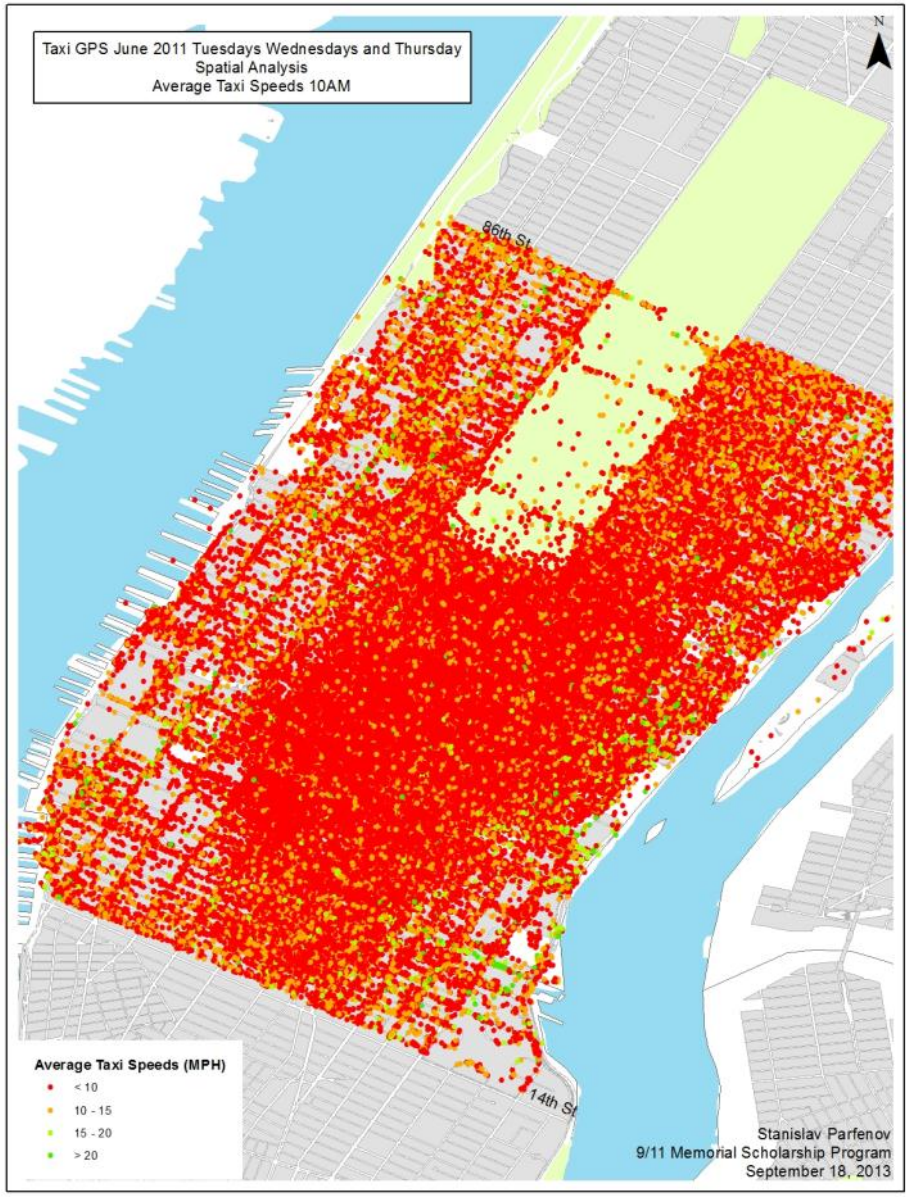


June 16 2011



June 23 2011

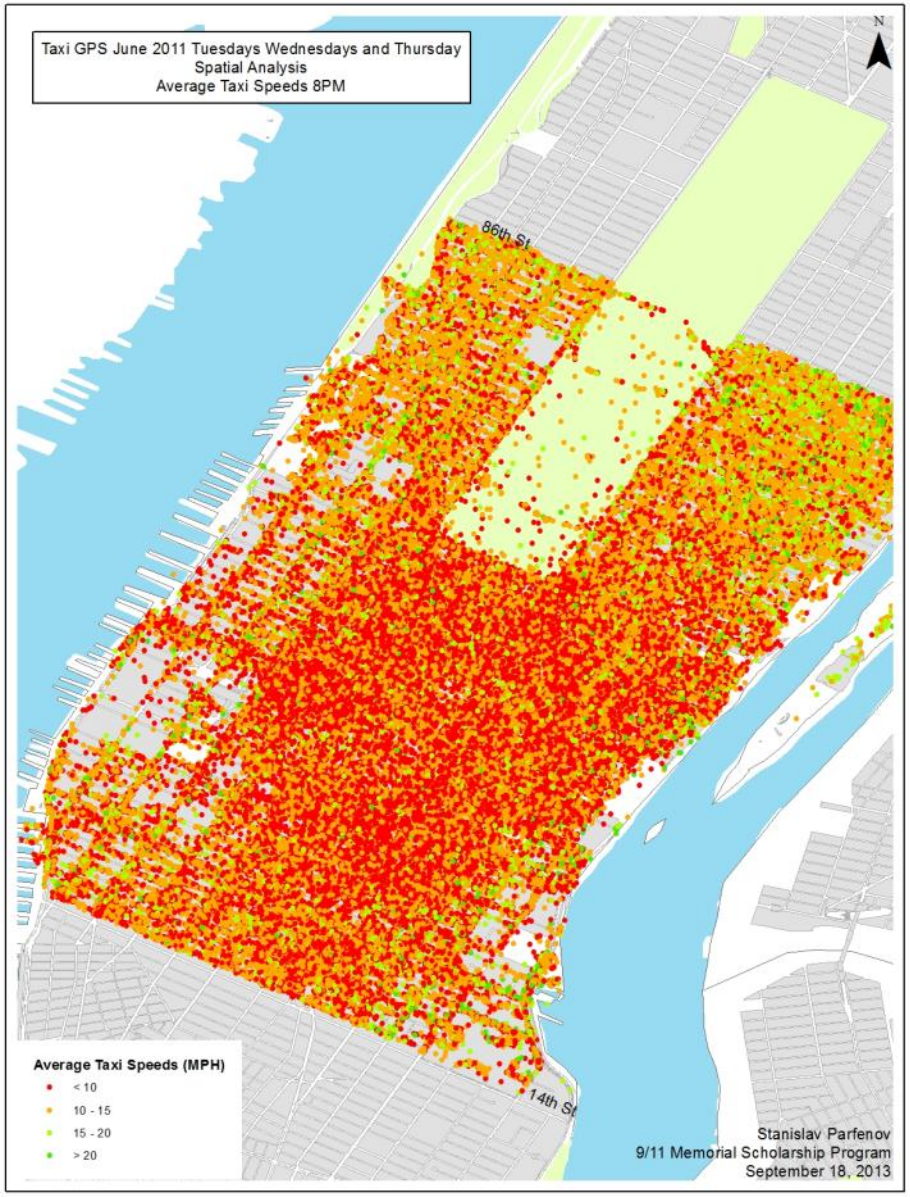
Data Visualization



Color is showing:
Red - speed below 10 MPH
Green – speed above 25 MPH

AM Peak Period
6 – 10 AM

Data Visualization



Color is showing:
Red - speed below 10 MPH
Green – speed above 25 MPH

PM Peak Period
3 – 8 PM

Future Steps

There are a number of future steps for both simulation and taxi GPS technologies which can work hand-in-hand:

- Use of GPS loggers that collect taxi XY coordinates and speed every second (currently a NYCDOT pilot project).
- NYCTLC is implementing new GPS data collection technology on board of their yellow taxis, where GPS XY location will be collected every 2 minutes
- Also, a about 1,500 “green cabs” will be distributed along the outer Boroughs to provide taxi services there (NYCTLC initiative)
- Software packages are moving online and soon Aimsun will be running its internet based cloud computing services, removing the need for a desktop only application.
- Traffic data will be fed into Aimsun’s online network, thus allowing to change traffic conditions on-the-fly with “look-ahead” prediction capabilities (FHTWA currently exploring this technology).

Conclusions

- Traffic simulation technology offers a great advantage to the planning process, especially in simulating/predicting traffic conditions during special events, road closures and construction projects.
- There is a need to move towards new sources of data, like taxi GPS and other on-board data collection devices, to collect both live and historical data, which would be then used to analyze and predict traffic conditions.
- Based on this project, it is evident that Taxi GPS data could be used as a reliable and innovative source of data, that can be used in the planning process and traffic analysis.

Gratitude

I would like to sincerely express my gratitude to a number of people who I had a pleasure to learn from, collaborate and get assistance:

- Mike Marsico, P.E. (NYCDOT)
- Elena Prassas, Ph.D. (NYU Poly)
- Penny Eickemeyer (University Transportation Research Center)
- Ophelia Ray-Fenner (formerly with NYCDOT)
- Alexander Parfenov, Ph.D.
- Andrew Weeks, Sophia Choi, Jannie Gao, Shane Zhang, Zamir Alam, Kim Kirby, Onyinye Akujuo and Catherine Matera (NYCDOT)
- Jeremy Safran (NYCDOT and fellow scholar)
- Mohamad Talas, Ph.D. (NYCDOT)
- Murat Aycin and Alex Gerodimos (TSS)
- NYCTLC for their continuing support with Taxi GPS Program
- Last but not least, NYMTC and UTRC staff, and all those who support and made this great program possible to honor those who are not with us today.

Questions?

Thank You