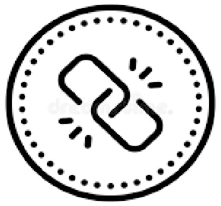
ACCELERATING EMERGENCY RESPONSE: OPTIMIZED AMBULANCE DISPATCH FOR MANHATTAN

15.C57 Optimization Methods

Ellie Yang, Naiqi Zhang, Sneha Sathish Kumar

MIT MANAGEMENT
SLOAN SCHOOL

WHAT IS THE PROBLEM?



Response time

Emergency response times in Manhattan have increased, with life-threatening emergencies now averaging 12 minutes and 26 seconds - a 29% rise from 2014

Resource constraints

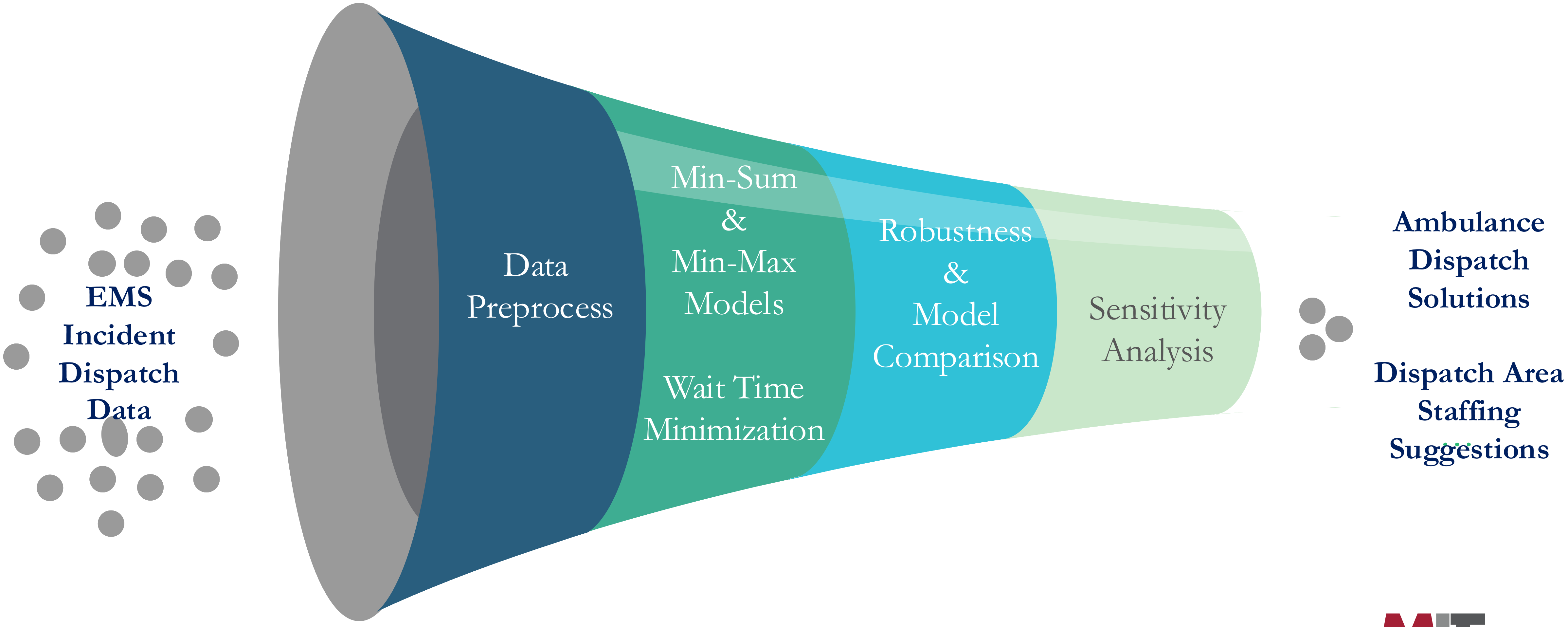
EMS resources are under pressure, handling approximately 5,000 daily emergency calls despite a capacity for 4,000. This is compounded by a 17.5% decrease in active certified EMS responders from 2019 to 2022

Traffic conditions

Traffic congestion continues to challenge ambulance efficiency, with average speeds of 4.8 mph in Midtown and 6.9 mph in Lower Manhattan.

MIT MANAGEMENT SLOAN SCHOOL

OUR METHODOLOGY



MIT MANAGEMENT SLOAN SCHOOL

DATA OVERVIEW

- EMS Incident Dispatch Data from New York City Fire Department (FDNY)
- October 2016 to October 2024: 27223682 entries of emergency calls
- Variables include datetime, response times(in seconds), dispatch and incident locations, etc.
- Coverage: New York City

Data Preprocess

- Manhattan: incident ZIP codes 10001–10282
- Response time:
  - maximum recorded response time between each pair of incident and dispatch area
  - large placeholder values for no-response cases
- Demand Estimation:
  - latest hourly data for incidents per incident and dispatch area
- Capacity Estimation:
  - peak hourly dispatch for each dispatch station (2023–2024)

BASIC MODELS AND THEIR ROBUST VERSION

Model 1: Min-Sum

- Objective: Minimizing the Overall Wait Time of all emergency incidents in Manhattan
- Optimize Overall system efficiency

Model Intuition

$$\min \sum_i \sum_j S_{ij} \cdot wait_{ij}$$

$$\sum_j S_{ij} \le C_i \quad \forall i \quad \text{(station capacity)}$$

$$\sum_i S_{ij} \ge d_j \quad \forall j \quad \text{(demand satisfaction)}$$

$$S_{ij} \ge 0 \quad \text{and integer} \quad \forall i, j$$

- Cons:
  - Doesn't guarantee fair service to all incidents
  - Some areas might wait longer than others

Trade Off !

Model 2: Min-Max

- Objective: Minimizing the longest wait time of one single incident in Manhattan
- Ensure nobody wait too long for ambulance, ensure equity

- Cons:
  - Average wait time might be longer
  - Uses more resources

Model Intuition

$$minmax z_{ij} wait_{ij}$$

$$\sum_j S_{ij} \le C_i, \quad \forall i \quad \text{(station capacity)}$$

$$\sum_i S_{ij} \ge d_j, \quad \forall j \quad \text{(demand satisfaction)}$$

$$S_{ij} \ge 0 \quad \text{and integer} \quad \forall i, j$$

$$z_{ij} \in \{0, 1\} \quad \forall i, j$$

These two models work well in different scenarios.

KEY FINDINGS

- Comparison of Min-Max / Min-Sum Models and their Robust Versions
  - Number of dispatches per station
  - Distribution of incident wait time

NUMBER OF DISPATCHES PER STATION

Basic Optimization Model

Station Code	Station Capacity	# of Ambulance Dispatched	
		Min-Sum Model	Min-Max Model
M1	5	5	5
M2	26	1	12
M3	28	2	15
M4	9	9	9
M5	16	6	0
M6	8	0	0
M7	20	0	2
M8	11	9	1
M9	16	5	1

Robust Optimization Model

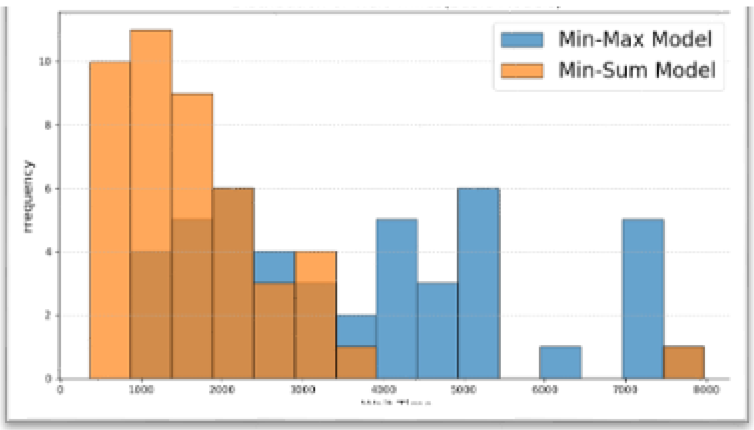
Station Code	Station Capacity	# of Ambulance Dispatched	
		Min-Sum Model	Min-Max Model
M1	5	4	4
M2	26	3	9
M3	28	8	8
M4	9	7	7
M5	16	12	10
M6	8	6	6
M7	20	16	14
M8	11	8	8
M9	16	12	10

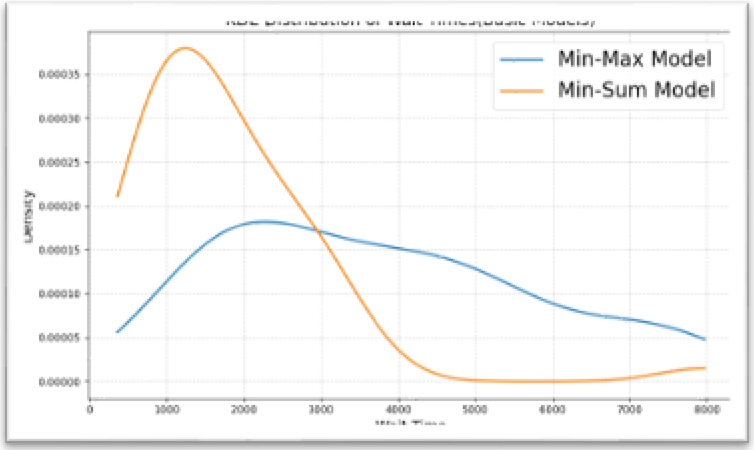
- Basic models exhibit **extreme allocation tendencies**:
  - M1 and M4 use full capacity; M5, M6, M7 use no capacity
- Robust versions have broader and more **balanced station utilization**

WAIT TIME DISTRIBUTION

Basic Model (Min-Sum & Min-Max)

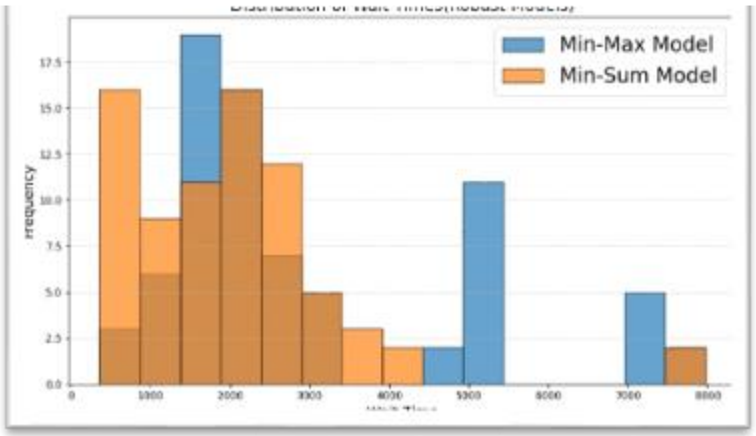
	Min-Sum Model	Min-Max Model
Total Wait Time(s)	80059	169691
Max Wait Time(s)	7968	7968
Min Wait Time(s)	362	1230

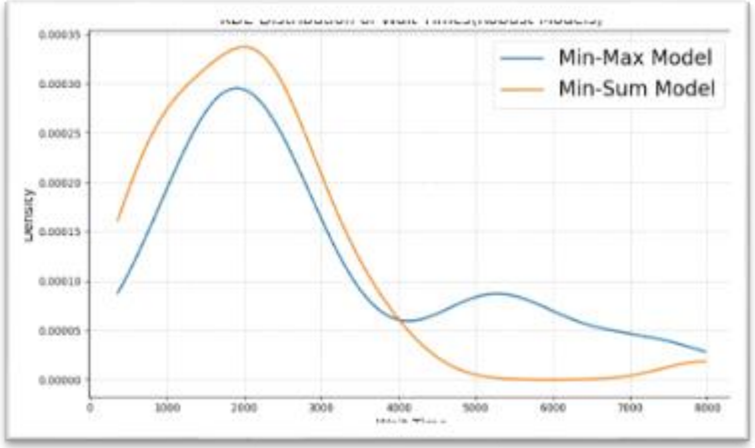




Robust Model (Min-Sum & Min-Max)

	Min-Sum Model	Min-Max Model
Total Wait Time(s)	155888	225587
Max Wait Time(s)	7968	7968
Min Wait Time(s)	478	362





- **Min-Sum vs. Min-Max**
  - Min-Sum has lower TOTAL wait time
  - Have the same MAX wait time (Due to zipcode 10013)

Min-Sum is better in our case!

- **Basic Model vs. Robust Model**
  - Robustness sacrifices immediate efficiency for reliability, increasing total wait time
  - Robustness has more balanced performance across demand scenarios

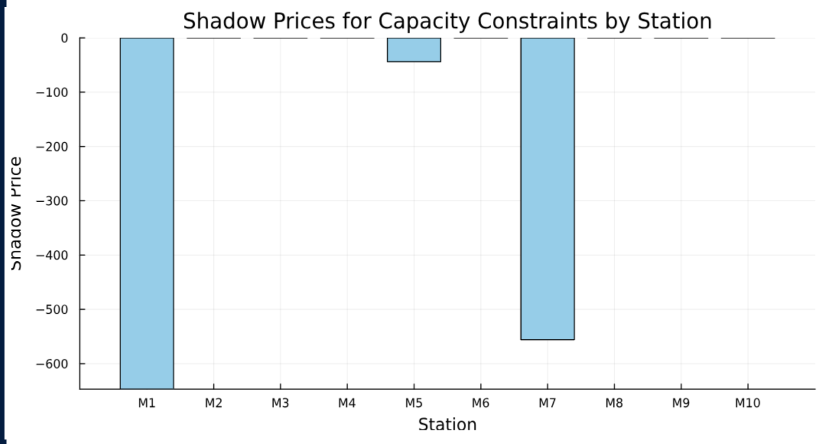
SENSITIVITY ANALYSIS

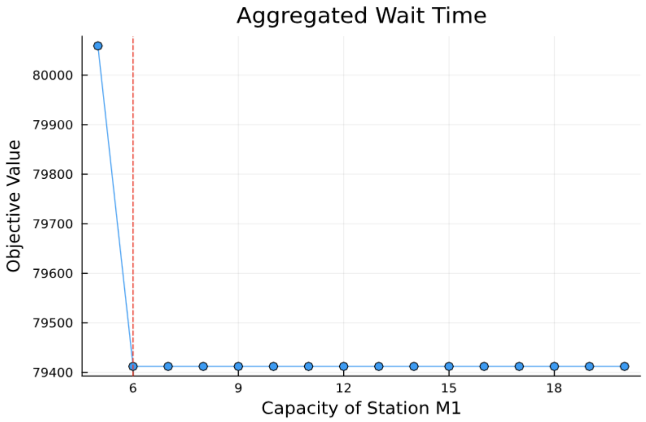
Critical Bottleneck:

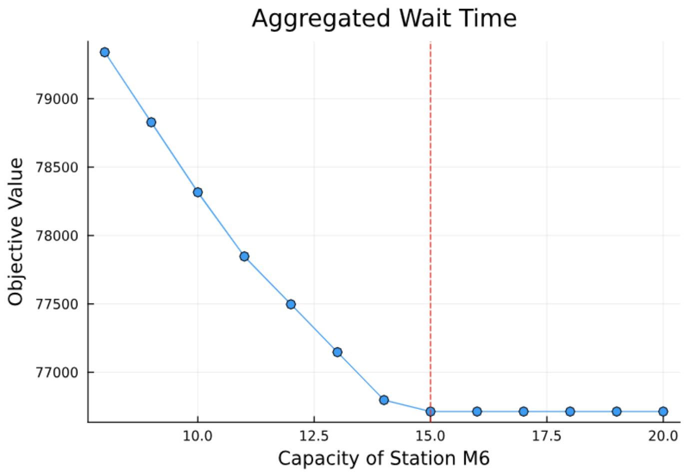
- **Min-Sum Model:**
  - Stations: M1, M4, M6
  - Shadow Prices: -647, -44, -551
- **Min-Max Model:**
  - Zero shadow prices across all stations

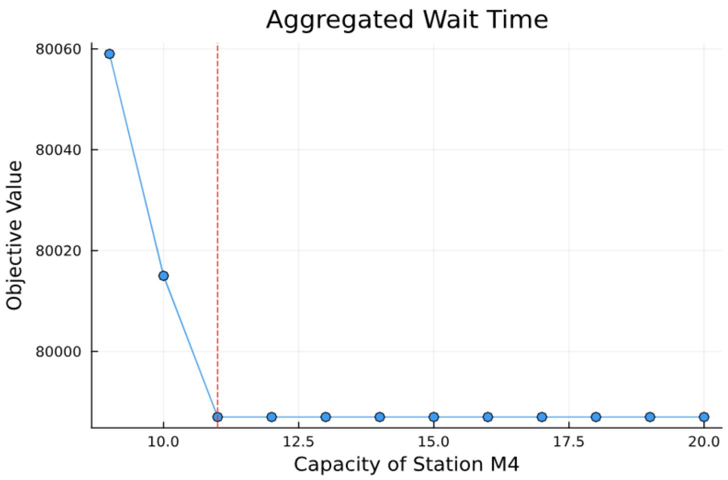
Knee Point:

- capacity 6 for Station M1
- capacity 11 for Station M4
- capacity 15 for Station M6









CONCLUSION & IMPACT

Efficiency and Equity

- ***Minimize Emergency Waiting Times:***
  - Min-Sum Model achieves lower total wait time (80,059 seconds); Min-Max Model addresses equity concerns
- ***Insights into underserved area:***
  - **ZIP Code 10013:** High wait times in Chinatown, SoHo, and Little Italy, a high-risk area needing additional resources

Robustness and Resiliency

- Add robustness to Min-Sum and Min-Max models to address uncertainties in demand and capacity
- Sacrifice immediate efficiency but provide better adaptability across demand scenarios

Staffing Suggestions

- Evaluate how capacity constraints impact system performance using shadow prices
- Target bottleneck station(M1,M4,M6) and make staffing suggestions to optimize system-wide efficiency and equity

MIT MANAGEMENT SLOAN SCHOOL

FUTURE DIRECTION

Improve Data Accuracy:	more precise and detailed data on ambulance and station capacities
Incorporate Incident Details:	accident types and severity levels for better prioritization
Multi-Stage Decision Making:	both initial and updated information (e.g., initial call and end-time data) for dynamic resource allocation.
Contextual Indicators:	Include factors like special events and localized incidents to enhance model adaptability

THANK YOU

MIT MANAGEMENT
SLOAN SCHOOL