Travel behavior, today and tomorrow: The promises and pitfalls of emerging data for transportation planning applications

Ekin Uğurel

University of Washington
Department of Civil and Environmental Engineering
URBDP 532: Current Topics in Transportation Planning and Policy

February 14, 2024

What is travel behavior?

(Goulias et al., 2020)

- "In this sense, travel behavior is the combination of doing things in different places at different times and how we move from one place to another. Travel behavior is also about feelings, emotions, perceptions, norms, beliefs, intentions, and attitudes. ... Moreover, travel behavior is how to go about deciding how to do things. Perhaps we form utilities for everything we do, or perhaps we use intuition, or perhaps we do both."
- "[W]e allocate time and other resources to activities and interactions with other people that evolve over time and space."

Dimensions of Travel Behavior

- Who: The trip maker
- What: Trip generation
- When: Departure choice, arrival time
- Where: Trip distribution, traffic assignment
- Why: Trip purpose
- How: Mode choice

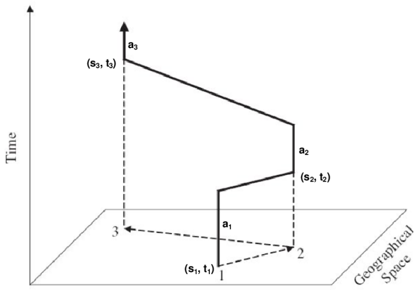
Influences on Travel Decisions

- Person and household-related attributes
  - Socioeconomics and demographics (McGuckin and Murakami, 1999; Nishii et al., 1988)
  - Attitudes and feelings (Bayarma et al., 2007)
- Built environment
  - Surrounding origin and destination
  - Density, diversity, and design (Cervero and Kockelman, 1997)
- Alternative-related attributes
  - E.g. for mode choice: What alternatives should one consider?

Space-Time Geography

"What about people in regional science?" (Hägerstrand, 1970)
Physical, temporal constraints to locations a person can go.

- Spatial: Origin/destination of trip, travel distance, path chosen, dispersion of trips
- Temporal: Departure time of trips, length of trip, length of tour, frequency of trips



Other important definitions

- **Anchor**: A primary trip destination (typically work, school, and home)
- **Trip**: Movement in time and space connecting one origin and destination (e.g., home to grocery store)
- **Tour**: Sequence of trips that start and end at the same location
- **Trip Chain**: A series of trips linked together during a single outing.
- **Accessibility**: The ease of reaching desired services, destinations, or activities.
- **Mode Split**: The distribution of travel made by various forms of transportation (e.g., the percentage of trips made by walking, cycling, public transit, and private automobile).

Wickedness of Planning Problems

One important aim of travel behavior analysis and modeling is transportation planning to solve problems such as congestion, accidents, waste of resources, pollution, and inequity. Most of the transportation planning problems are "wicked" problems (Rittel and Webber, 1973):

- they have unclear formulation of what the problem we need to solve is (vagueness);
- their solutions emerge when they are good enough, but never optimal (unknown optimum);
- progress occurs through a continuity of solutions that improve over time (incremental progression);
- not all intended and unintended consequences can be traced from the beginning (lack of complete observability);
- every solution to a problem leaves an unchangeable trace of the outcome(s) (path dependence and irreversibility);

Wickedness of Planning Problems (Cont.)

(Rittel and Webber, 1973)

- we cannot enumerate all possible solutions, consequences, and outcomes (indeterminacy);
- problems are unique in historical time and place with no repeatable paths to a solution (place-time uniqueness);
- a problem is a symptom of another problem from different domains of the life of people (nested hierarchy of problems);
- real-life planning work does not allow testing and experimentation using the scientific method (need for different methods)

Discussion

- Do you agree with Rittel and Webber's characterizations? Why or why not?
- Which dimensions of travel behavior do you think are most useful in understanding contemporary planning problems?
- Questions?

Travel Behavior → Travel Demand Forecasting

TDF: a process to predict changes in travel behavior for a specific time and place, based on changes in land use, demographics, preferences, technologies, and policy.

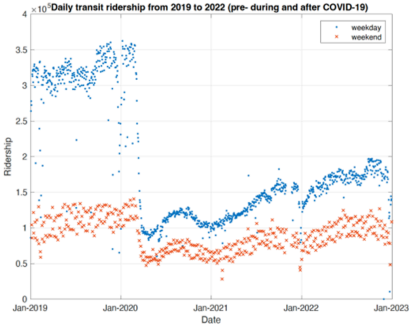
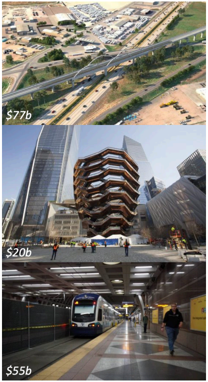


Figure: Bus ridership in King County between January 2019 and 2023.

Why is TDF important?

- Groundwork for transportation infrastructure investment decisions
- Critical for policy monitoring and evaluation
- Understand impacts of land use policies and development decisions on transportation
- Others?



Example Applications of TDF

Sound Transit light rail expansions

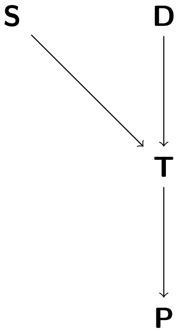
- To know where to build new lines, need to estimate future ridership and revenue, while accounting for construction and operation expenses

PSRC's VISION 2050

- Anticipated growth of 1.5 million people in next 30 years in Puget Sound region
- Focused growth in centers and near transit, reduce greenhouse gas emissions
- Other focuses: healthy environment, economic prosperity, social equity, affordable housing

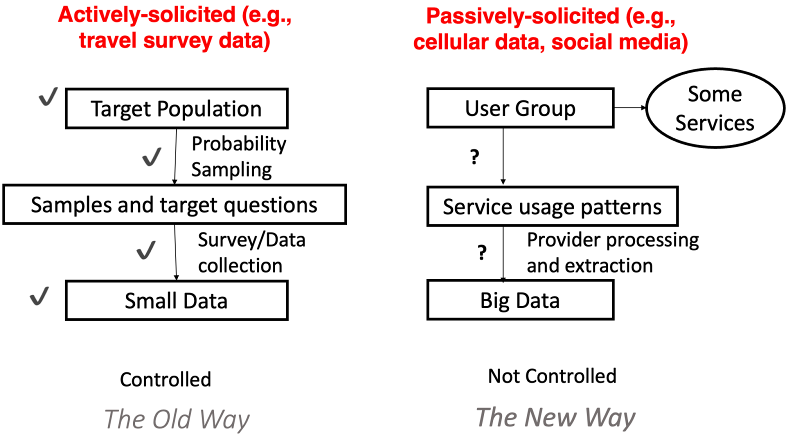
The TDF Process

- **S** (Supply): Characteristics of the general environment (transportation and land use)
- **D** (Demand): Numbers and characteristics of trip makers (households)
- **T** (Travel behavior): Trip-making in time and space
- **P** (Performance): Transportation system performance

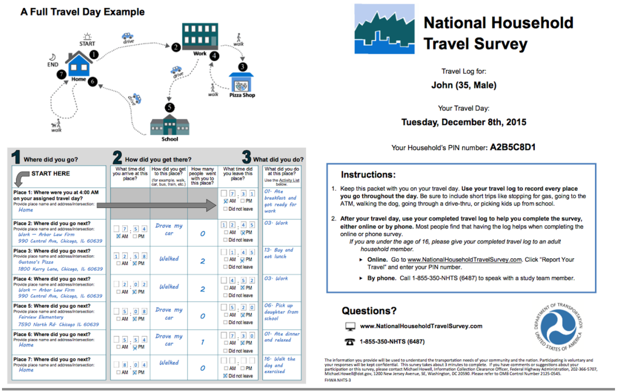


Emerging Data Sources

- Passively-generated mobile data (i.e., GPS traces)
- General Transit Feed Specification (GTFS) data
- Transit ridership data (i.e., from Automated Passenger Counters)
- Twitter/Yelp data
- Parking data (from third-party parking mgmt systems, sensors)
- Crowdsourced congestion and incident data (i.e., from Waze/Google Maps users)


Actively-solicited (e.g., travel survey data)
Target Population
Probability Sampling
Samples and target questions
Survey/Data collection
Small Data
Controlled
The Old Way
Passively-solicited (e.g., cellular data, social media)
User Group
Some Services
?
Service usage patterns
?
Provider processing and extraction
Big Data
Not Controlled
The New Way

Example of a Travel Survey



A Full Travel Day Example

1 Where did you go? START HERE	2 How did you get there? What time did you arrive at this place?	How did you get to this place? (for example, walk, car, bus, train, etc.)	How many people went with you to this place?	What time did you leave this place?	3 What did you do? What did you do at this place? Use the Activity List below.
Place 1: Where were you at 4:00 AM on your assigned travel day? Provide place name and address/intersection: Home				7:31 ☒ AM ☐ PM ☐ Did not leave	01- Ate breakfast and got ready for work
Place 2: Where did you go next? Provide place name and address/intersection: Work — Arbor Law Firm 990 Central Ave, Chicago, IL 60639	7:54 ☒ AM ☐ PM	Drove my car	0	12:45 ☐ AM ☒ PM ☐ Did not leave	03- Work
Place 3: Where did you go next? Provide place name and address/intersection: Gustoso's Pizza 1800 Kerry Lane, Chicago, IL 60639	12:58 ☐ AM ☒ PM	Walked	2	1:45 ☐ AM ☒ PM ☐ Did not leave	13- Buy and eat lunch
Place 4: Where did you go next? Provide place name and address/intersection: Work — Arbor Law Firm 990 Central Ave, Chicago, IL 60639	2:02 ☐ AM ☒ PM	Walked	2	4:52 ☐ AM ☒ PM ☐ Did not leave	03- Work
Place 5: Where did you go next? Provide place name and address/intersection: Fairview Elementary 7590 North Rd. Chicago IL 60639	5:08 ☐ AM ☒ PM	Drove my car	0	5:20 ☐ AM ☒ PM ☐ Did not leave	06- Pick up daughter from school
Place 6: Where did you go next? Provide place name and address/intersection: Home	5:54 ☐ AM ☒ PM	Drove my car	1	7:30 ☐ AM ☒ PM ☐ Did not leave	01- Ate dinner and relaxed
Place 7: Where did you go next? Provide place name and address/intersection: Home	8:04 ☐ AM ☒ PM	Walked	0	: ☐ AM ☐ PM ☒ Did not leave	16- Walk the dog and exercised

National Household Travel Survey

Travel Log for:

John (35, Male)

Your Travel Day:

Tuesday, December 8th, 2015

Your Household's PIN number: **A2B5C8D1**

Instructions:

1. Keep this packet with you on your travel day. **Use your travel log to record every place you go throughout the day.** Be sure to include short trips like stopping for gas, going to the ATM, walking the dog, going through a drive-thru, or picking kids up from school.
2. **After your travel day, use your completed travel log to help you complete the survey, either online or by phone.** Most people find that having the log helps when completing the online or phone survey.

   If you are under the age of 16, please give your completed travel log to an adult household member.

   - **Online.** Go to www.NationalHouseholdTravelSurvey.com. Click "Report Your Travel" and enter your PIN number.
   - **By phone.** Call 1-855-350-NHTS (6487) to speak with a study team member.

Questions?

www.NationalHouseholdTravelSurvey.com

1-855-350-NHTS (6487)

The information you provide will be used to understand the transportation needs of your community and the nation. Participating is voluntary and your responses will be kept confidential. This survey takes about 3 minutes to complete. If you have comments or suggestions about your participation or this survey, please contact Michael Howell, Information Collection Clearance Officer, Federal Highway Administration, 202-366-5707, Michael.Howell@dot.gov, 1200 New Jersey Avenue, SE, Washington, DC 20590. Please refer to OMB Control Number 2125-0545.

FHWA NHTS-3

Passively-generated mobile data

- Ubiquitous, and therefore massive sample sizes
- Self-selection bias
- Observation frequency varies greatly
- Missing data results in bias
- CityCast

Causes of sparsity in mobile data

- User-related causes
  - Phone on sleep mode (hibernation)
  - Restricted location data permissions
  - Restricted background app refresh
- Geographic/built-environment-related causes
  - Short gaps due to enclosed structures (e.g., tunnels)
  - The Urban Canyon Effect
- Stochastic/miscellaneous causes
  - Battery drain (device dead)
  - User leaves device at home
  - App shutdown/crash

Inferred travel behavior is a function of sparsity

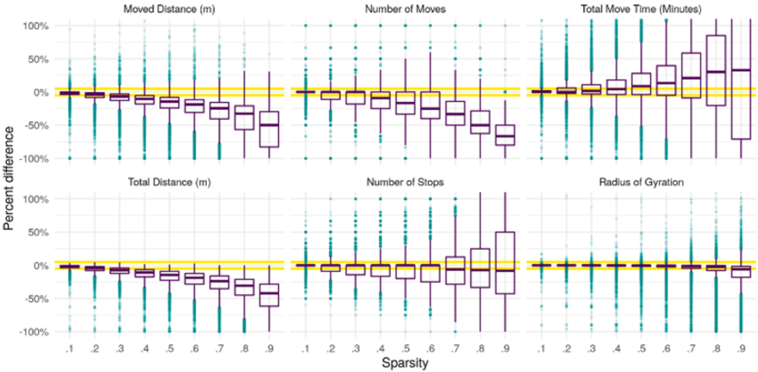
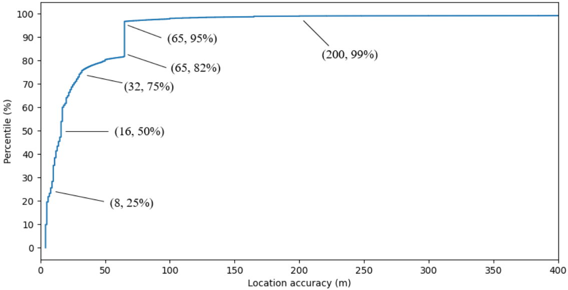
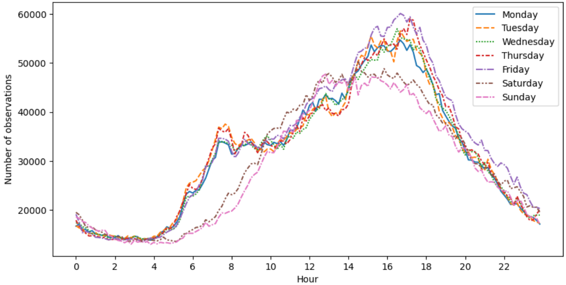


Figure: from McCool et al. (2022)

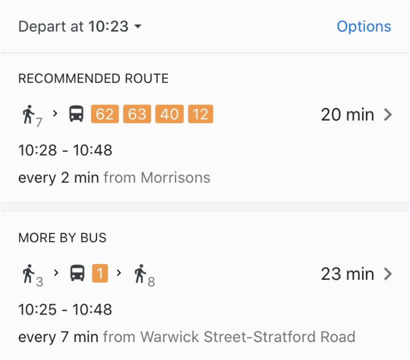
Privacy-protected mobile data from Spectus





Transit Network Data: GTFS

- Used to distribute relevant information about transit systems to riders
- As seen in OneBusAway, Google Maps
- Contains information about routes, schedules, fares, and geographic transit details, and it is presented in simple text files.



Ridership Data

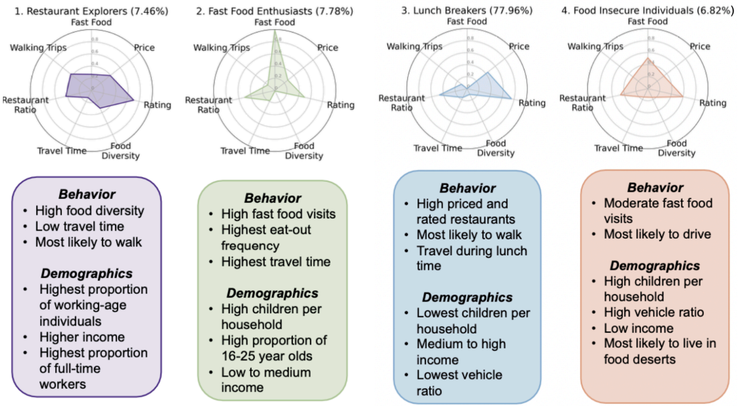
- Increasing availability due to ubiquity of automated passenger counters (APCs)
- Can accurately record boardings and alightings
- Helpful in providing real-time information on vehicle crowding to transit riders (was especially important during COVID!)

Twitter/Yelp Data

- Helpful for understanding attitudes towards points of interest (POIs)
- Can be scraped off the web and analyzed using natural language processing
- Liable to extremity bias

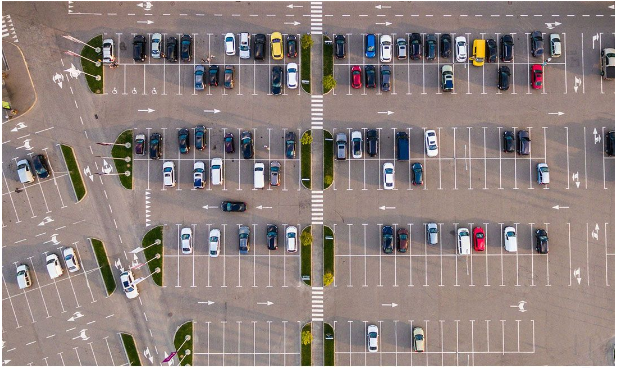
Applications of Yelp Data

Ng et al., 2024 (under review)



Parking Data

- Data on inventory, cost, and occupancy
  - PSRC Parking Inventory
  - Paid Parking Occupancy in Seattle
- Insights into demand
- Facilitates experimental designs with pricing models
  - Can help promote more sustainable modes of transport
- *even more novel:* Satellite imagery for this purpose

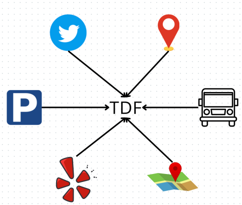


Takeaways

- Big data is useful for a variety of purposes, but more effort is required to process and derive meaning from big data compared to traditional data sources
- Results derived from big data *should be* validated by other independent data sources (i.e., traditional flow data, survey data, etc.)
- Data needs, collection procedures, and quality metrics should be defined/designed carefully for both big data and traditional data

Discussion

- How do the accuracy and reliability of emerging data sources like GPS traces, transit ridership data, and social media (e.g., Twitter, Yelp) compare to traditional data collection methods in transportation planning? What are the implications of these differences for travel demand forecasting?
- Do emerging data sources adequately represent the entire population and all modes of transportation? How might biases in these data sources impact travel demand forecasting and transportation policy decisions?
- Personal experiences?



My Research: Motivation

- Two pervasive issues:
  - As data collection practices become more transparent and user-centric, the sparsity issue only gets worse (DeGiulio et al., 2021)
  - Researchers are not able to share individual mobile data used in their studies due to privacy agreements with data providers (Gao et al., 2019; Rao et al., 2018; Liu and Onnela, 2021)
- The above motivates:
  - An imputation method to correct missing data in GPS traces at various levels (Ugurel et al., 2024)
  - A generative modeling framework for individual mobile data to create synthetic datasets replicating real travel behavior (Ugurel, E., Huang, S., Chen, C., under review)

Domain Challenges

- Travel behavior heterogeneity at the individual-level (Bayarma et al., 2007; Kitamura and Van Der Hoorn, 1987; McGuckin and Murakami, 1999; Nishii et al., 1988; Lee and McNally, 2006).
- Physical system complexities imposed by the built and natural environments

Research Questions

- Given time, how do we infer (predict) spatial locations?
- How do we infuse physics (i.e., constraints from velocity and bearing) into the inference problem from time to location, as stated above?

Research Questions

- Given time, how do we infer (predict) spatial locations?
- How do we infuse physics (i.e., constraints from velocity and bearing) into the inference problem from time to location, as stated above?

Research Questions[1]

- Given time, how do we infer (predict) spatial locations?
- How do we infuse physics (i.e., constraints from velocity and bearing) into the inference problem from time to location, as stated above?

[1] Papers:

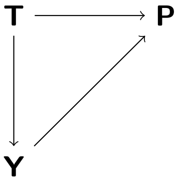
- Ugurel, E., Guan, X., Wang, Y., Huang, S., Wang, R., Chen, C., 2024. Correcting Missingness in Passively-generated Mobile Data using Multi-task Gaussian Processes. *To appear in latest issue of Transportation Research Part C: Emerging Technologies*.
- Ugurel, E., Huang, S., Chen, C., 2024. Uncovering physics-regularized data generation processes for individual human mobility: A multi-task Gaussian process approach based on multiple kernel learning. *Under review*.

Problem Definition

Let $\mathbf{T}$, $\mathbf{P}$, and $\mathbf{Y}$ be defined as follows

$$\mathbf{T} = \begin{bmatrix} t_{1,1} & \dots & t_{d,1} \\ \vdots & \ddots & \vdots \\ t_{1,n} & \dots & t_{d,n} \end{bmatrix} = \begin{bmatrix} \mathbf{t}_1 \\ \vdots \\ \mathbf{t}_n \end{bmatrix}, \mathbf{P} = \begin{bmatrix} v_1 & \beta_1 \\ \vdots & \vdots \\ v_n & \beta_n \end{bmatrix}, \mathbf{Y} = \begin{bmatrix} y_{\lambda,1} & y_{\phi,1} \\ \vdots & \vdots \\ y_{\lambda,n} & y_{\phi,n} \end{bmatrix}.$$

We assume the following causal structure between $\mathbf{T}$, $\mathbf{P}$, and $\mathbf{Y}$

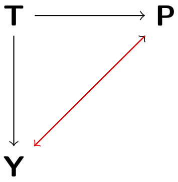


Problem Definition

Let **T**, **P**, and **Y** be defined as follows

$$\mathbf{T} = \begin{bmatrix} t_{1,1} & \dots & t_{d,1} \\ \vdots & \ddots & \vdots \\ t_{1,n} & \dots & t_{d,n} \end{bmatrix} = \begin{bmatrix} \mathbf{t}_1 \\ \vdots \\ \mathbf{t}_n \end{bmatrix}, \mathbf{P} = \begin{bmatrix} v_1 & \beta_1 \\ \vdots & \vdots \\ v_n & \beta_n \end{bmatrix}, \mathbf{Y} = \begin{bmatrix} y_{\lambda,1} & y_{\phi,1} \\ \vdots & \vdots \\ y_{\lambda,n} & y_{\phi,n} \end{bmatrix}.$$

We assume the following causal structure between **T**, **P**, and **Y**



Multi-task Gaussian Process

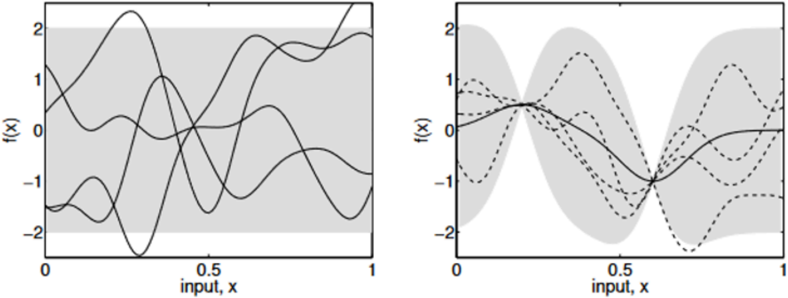
- First, let's focus on modeling the relationship $\mathbf{T} \rightarrow \mathbf{Y}$. Consider the task of learning a function $f_j : \mathbb{R}^d \rightarrow \mathbb{R}$ where j refers to either latitude ϕ or longitude λ. The basic form of our learning problem is

$$y_{ji} = f_j(\mathbf{t}_i) + \epsilon_{ji}, \tag{1}$$

  where f_j is a systematic function mapping inputs $\mathbf{t}_i$ to output y_{ji}, and $\epsilon_{ji} \sim \mathcal{N}(0, \delta_j^2)$ are independent random variables for noise associated with the j^{th} task.
- We place a GP prior on f_j such that $f_j \sim \mathcal{GP}(m(\cdot), k(\cdot, \cdot))$, where $m(\cdot) = \mathbb{E}[f_j(\cdot)]$ is the mean function, and $k(\cdot, \cdot)$ is the covariance (or kernel) function.

Intuition

GPs consider the space of all possible functions and return the most likely given your training data (+ your choice of kernel)



(a), prior (b), posterior

Panel (a) shows four samples drawn from the prior distribution. Panel (b) shows the situation after two datapoints have been observed. The mean prediction is shown as the solid line and four samples from the posterior are shown as dashed lines. Shaded region denotes twice the standard deviation at each input value x

Data preprocessing

- Oscillation Correction
  - Filter by maximum velocity (i.e., 200 km/h)
- Noise Filtering
  - Exclude observations with less than 300 meters in precision
- Input/output normalization
  - Mean of 0 and variance of 1

Defining Missingness

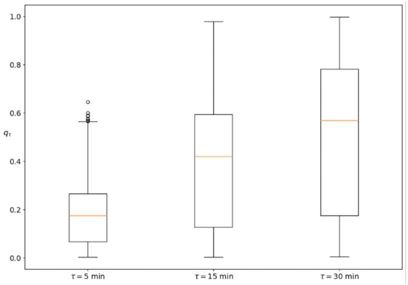
- Mobile data is irregularly sampled. Thus, we need a mathematical convention to denote varying levels of missingness
- Let $\mathcal{T}$ denote a the full length of a period, which we can discretize into P intervals of length τ. Let $\mathbf{I}_p$ be an indicator variable such that

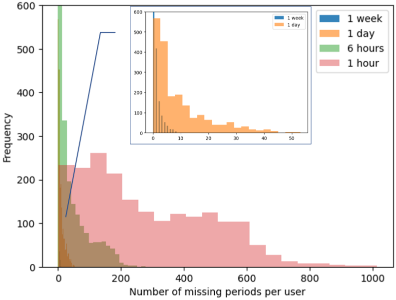
$$\mathbf{I}_p = \begin{cases} 1 & \text{if } p \text{ has at least one observation} \\ 0 & \text{otherwise} \end{cases}$$

- We can define temporal occupancy as

$$q_\tau = \frac{1}{P} \sum_{p=1}^{P} \mathbf{I}_p$$

Varying Levels of Missingness





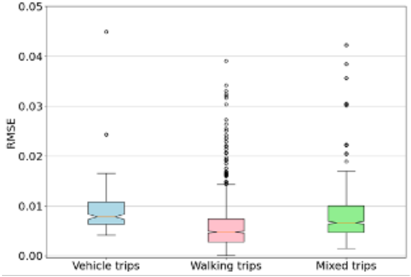
Experiment 1: Parameter Convergence

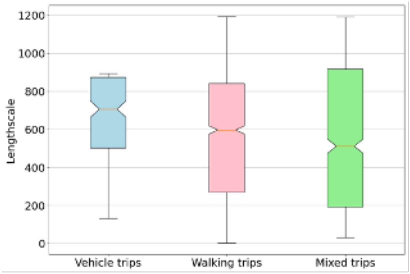
- K-means clustering to group together similar trips

Cluster	Mode	Avg. Vel. [m/s]	Distance [m]	Duration [s]	Heading Change Rate	Velocity Change Rate	Observations	Stop Rate
Slow, short trips	*Walk*	9.29	8,088	1,062	0.0019	0.0024	22.79	0.0007
Medium speed/ distance trips	*Mixed*	13.94	29,693	2,362	0.0007	0.0008	49.86	0.0002
Fast, distant trips	*Car*	17.86	59,299	3,449	0.0005	0.0006	141.8	0.0001

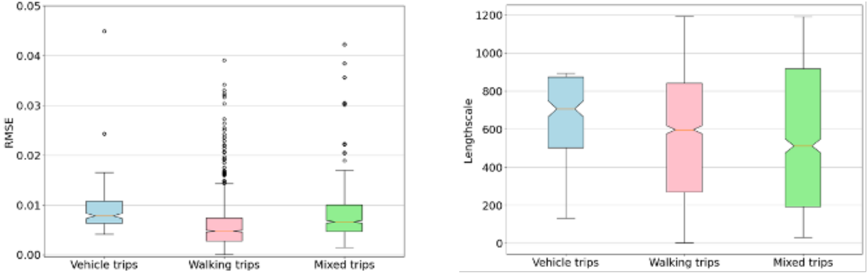
- Heading Change Rate: Ratio of consecutive points where a user changes direction with an angle exceeding a threshold (we use 0.33 rad)
- Velocity Change Rate: Ratio of consecutive points where the user exceeds a speed variation threshold (we use 26%)
- Stop Rate: Ratio of points with an inferred velocity lower than a threshold (we use 0.89 m/s)

Experiment 1: Parameter Convergence





Experiment 1: Parameter Convergence



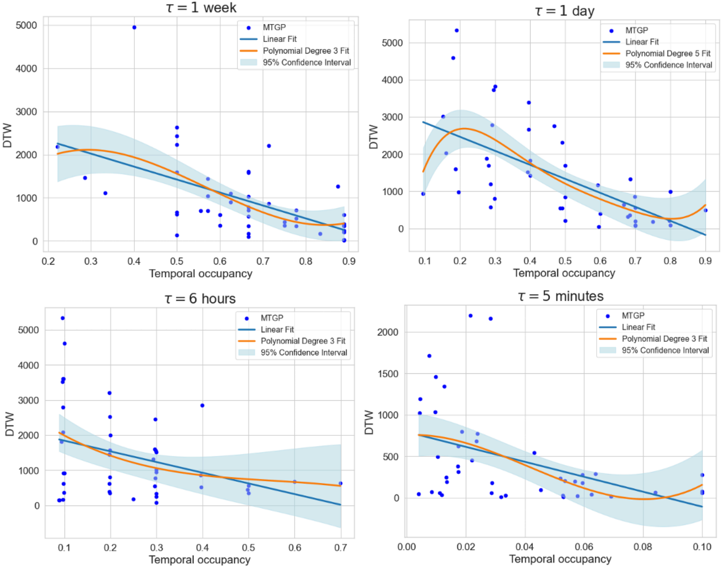
- The variability in lengthscales observed for walking trips may be attributed to the wide spectrum of walking behaviors.
- For mixed trips, the lower average lengthscale may be due to non-smooth transitions (or 'kinks') in the data introduced by mode changes

Experiment 2: Robustness Checks

- **Goal**: Assess model performance against other time-series imputation methods in a variety of missingness conditions
- **Method**: Simulate gaps by reserving a subset of data for testing, which we choose according to different temporal resolutions
- **Kernel**

$$k = \prod_{u=1}^{d} k_{RQ_u} \times k_{PER} + \prod_{u=1}^{d} k_{RQ_u} \times k_{PER}$$

Experiment 2: DTW



Experiment 2: Benchmarks

Time Gap	Method	Number of Locations	Radius of Gyration	Straight-Line Travel Distance	Random Entropy	Real Entropy	Uncorrelated Entropy
1 week	***MTGP***	**26**	**-0.07**	**205.029**	**0.045**	**0.278**	**0.153**
	RBF	-801	-0.835	-888.441	-9.647	-9.323	-9.527
	SES	-801	-0.835	-888.441	-9.574	-9.323	-9.527
	Holt	-801	-0.835	-888.441	-9.574	-9.323	-9.527
	ES	-778	-0.621	-643.909	-4.989	-7.726	-4.955
	ARIMA	-801	-0.835	-888.441	-9.276	-9.323	-9.527
	SARIMAX	-801	-0.835	-888.441	-9.647	-9.323	-9.527
1 day	***MTGP***	**33.5**	**-0.245**	**236.805**	**0.036**	**0.227**	**0.117**
	RBF	-1050	-0.909	-1303.06	-10.038	-9.612	-9.806
	SES	-1050	-0.871	-1303.06	-9.309	-9.612	-9.806
	Holt	-1050	-0.846	-1303.06	-9.223	-9.61	-9.803
	ES	-1027	-0.718	-768.678	-4.878	-7.907	-5.225
	ARIMA	-1050	-0.834	-1303.06	-8.506	-9.609	-9.803
	SARIMAX	-1050	-0.909	-1303.06	-10.038	-9.612	-9.806
6 hours	***MTGP***	**34**	**-0.187**	**-13.641**	**0.042**	**0.237**	**0.155**
	RBF	-956.5	-0.645	-1223.47	-9.809	-9.493	-9.751
	SES	-954	-0.645	-1177.23	-9.139	-9.493	-9.608
	Holt	-952	-0.645	-1171.79	-8.893	-9.493	-9.533
	ES	-929	-0.4	-768.56	-4.895	-7.869	-5.066
	ARIMA	-952	-0.645	-1178.65	-8.317	-9.493	-9.608
	SARIMAX	-956.5	-0.645	-1223.47	-9.901	-9.493	-9.751
1 hour	***MTGP***	**38.5**	**-0.074**	**389.308**	**0.053**	**0.29**	**0.157**
	RBF	-902	-0.94	-1319.47	-9.818	-9.548	-9.761
	SES	-901.5	-0.761	-1262.95	-7.989	-9.546	-9.642
	Holt	-901.5	-0.761	-1262.95	-7.041	-9.543	-9.642
	ES	-878.5	-0.627	-711.119	-4.8	-7.816	-5.174
	ARIMA	-898.5	-0.761	-1262.76	-6.801	-9.545	-9.613
	SARIMAX	-830.5	-0.644	-1161.14	-6.435	-9.455	-9.449
30 minutes	***MTGP***	**21**	**-0.435**	**123.606**	**0.048**	**0.314**	**0.166**
	RBF	-624.5	-1.36	-1043.86	-9.052	-8.954	-9.161
	SES	-614	-1.175	-1025.83	-7.405	-8.954	-9.006
	Holt	-614	-1.167	-1025.83	-6.872	-8.953	-8.952
	ES	-591	-1.145	-707.361	-4.099	-7.07	-4.445
	ARIMA	-614	-1.214	-1027.68	-6.282	-8.949	-8.952
	SARIMAX	-624.5	-1.36	-1043.86	-9.277	-8.954	-9.161
15 minutes	***MTGP***	**22**	**-0.299**	**-7.116**	**0.048**	**0.323**	**0.161**
	RBF	-670	-2.15	-1112.56	-8.925	-8.871	-9.125
	SES	-660	-1.71	-1162.11	-7.34	-8.871	-8.994
	Holt	-660	-2.099	-1162.07	-6.435	-8.871	-8.849
	ES	-637	-2.056	-513.035	-3.96	-6.771	-4.497
	ARIMA	-659	-1.931	-1168.42	-6.048	-8.871	-8.851
	SARIMAX	-670	-2.15	-1199.07	-9.39	-8.871	-9.146
5 minutes	**MTGP**	**21**	-0.824	**47.301**	**0.056**	**0.301**	**0.156**
	RBF	-896	-1.396	-1441.06	-9.791	-9.302	-9.571
	SES	-896	-1.274	-1391.03	-6.757	-9.302	-9.571
	Holt	-893	-1.012	-1391.03	-6.555	-9.302	-9.394
	ES	-872	**-0.744**	-666.535	-4.313	-6.643	-4.984
	ARIMA	-896	-1.339	-1391.03	-6.754	-9.302	-9.495
	SARIMAX	-896	-1.396	-1441.06	-9.809	-9.302	-9.571

Research Questions[2]

- Given time, how do we infer (predict) spatial locations?
- How do we infuse physics (i.e., constraints from velocity and bearing) into the inference problem from time to location, as stated above?
  - Note that this is different than the estimation problem $\mathbf{T} \rightarrow \mathbf{Y} \leftarrow \mathbf{P}$, when all variables are observed (albeit noisy)

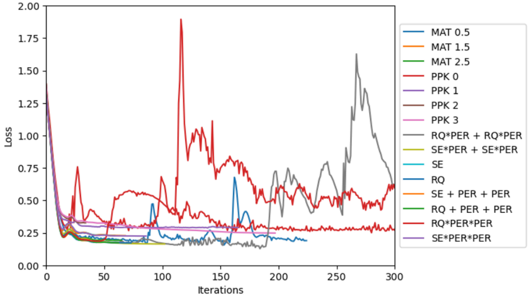
[2] Papers:

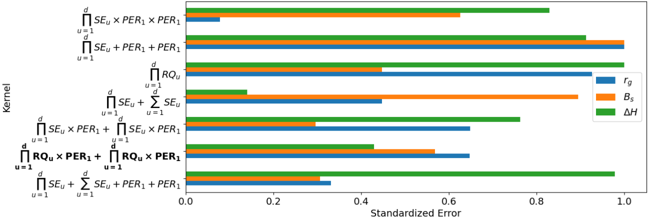
- Ugurel, E., Guan, X., Wang, Y., Huang, S., Wang, R., Chen, C., 2024. Correcting Missingness in Passively-generated Mobile Data using Multi-task Gaussian Processes. *Under review.*
- Ugurel, E., Huang, S., Chen, C., 2024. Uncovering physics-regularized data generation processes for individual human mobility: A multi-task Gaussian process approach based on multiple kernel learning. *Under review.*

Background

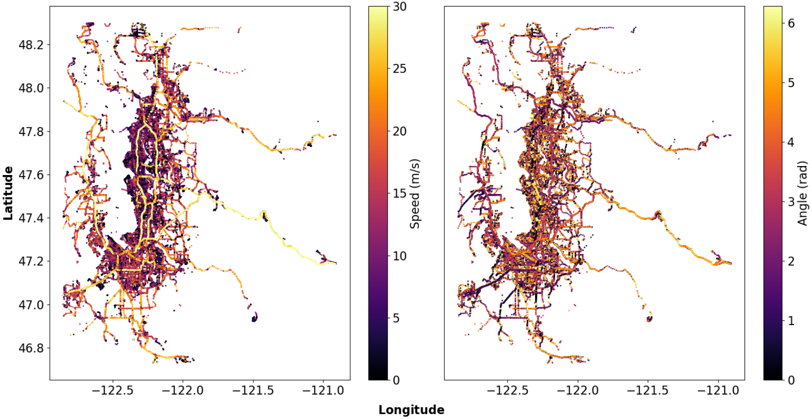
- Physical systems tend to have differential equations or other governing equations that describe the dynamics of the system.
- The Latent Force Model (Alvarez et al., 2013; Álvarez et al., 2009) has been successful in enforcing physical laws in a GP framework.
  - However, the LFM formulation is based on kernel convolution, and obtaining an analytical kernel after this process restricts usage to simple/smooth kernels (i.e., the Gaussian kernel).
  - This could hinder our ability to incorporate physical knowledge into kernels that are more intricate but extremely adaptable, such as those developed through our greedy learning algorithm.
- Inspired by Lasserre et al. (2006) and Wang et al. (2022), we propose a hybrid conditional-generative model that acts as a soft regularizer for the existing multi-task GP framework.
  - This model does not restrict the class of kernels that can be used, making it suitable for our approach.

Impact of Kernel Choice

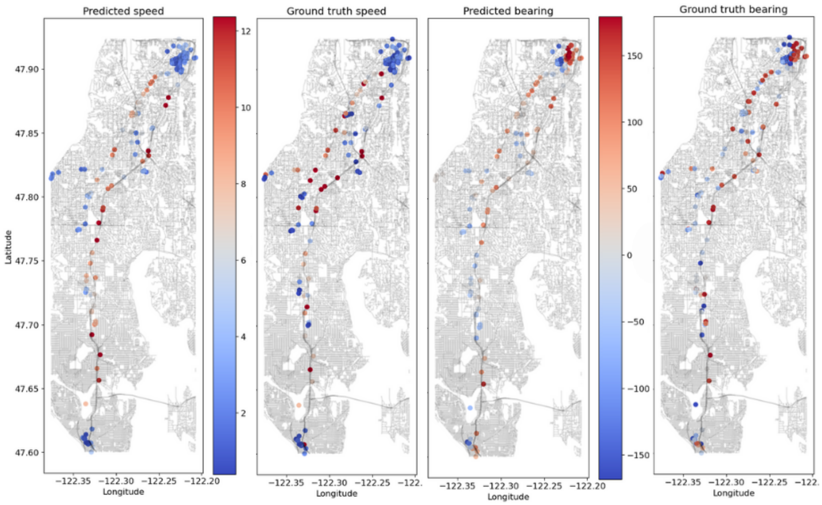




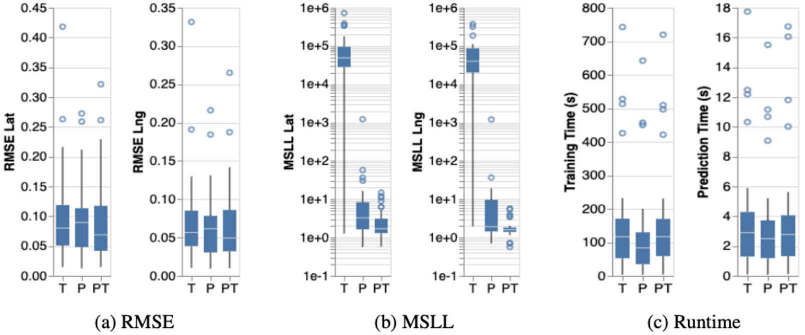
$\mathbf{P}_{gen}$ inference



P_{gen} inference



Performance



T, P, and PT denote the temporal-only, physical-only, and physics-regularized GP models, respectively. The MSLL plot is log-scaled in the y-axis.

Takeaways

- Different types of trips necessitate inherently different GP models
- GPs generalize better than traditional time-series extrapolation models
- The impact of kernel choice on mobility metrics derived from synthetic data is non-negligible
- Physics-regularization not only reduces model bias but also improves uncertainty estimates associated with the predicted locations.

Connect with me

- Email: `ekinokos2 [at] gmail [dot] com`
- Website: ekinugurel.github.io
- LinkedIn: linkedin.com/in/ekin-ugurel

References I

Alvarez, M. A., Luengo, D., and Lawrence, N. D. (2013). Linear latent force models using gaussian processes. *IEEE transactions on pattern analysis and machine intelligence*, 35(11):2693–2705.

Bayarma, A., Kitamura, R., and Susilo, Y. O. (2007). Recurrence of daily travel patterns: stochastic process approach to multiday travel behavior. *Transportation Research Record*, 2021(1):55–63.

Cervero, R. and Kockelman, K. (1997). Travel demand and the 3ds: Density, diversity, and design. *Transportation research part D: Transport and environment*, 2(3):199–219.

DeGiulio, A., Lee, H., and Birrell, E. (2021). "ask app not to track": The effect of opt-in tracking authorization on mobile privacy. In *International Workshop on Emerging Technologies for Authorization and Authentication*, pages 152–167. Springer.

References II

Gao, Y., Cheng, J., Meng, H., and Liu, Y. (2019). Measuring spatio-temporal autocorrelation in time series data of collective human mobility. *Geo-spatial Information Science*, 22(3):166–173.

Goulias, K. G., McBride, E. C., and Su, R. (2020). Life cycle stages, daily contacts, and activity-travel time allocation for the benefit of self and others. *Mobility and Travel Behaviour Across the Life Course*, pages 206–220.

Hägerstrand, T. (1970). What about people in regional science. *regional science association*, 24.

Kitamura, R. and Van Der Hoorn, T. (1987). Regularity and irreversibility of weekly travel behavior. *Transportation*, 14:227–251.

Lasserre, J. A., Bishop, C. M., and Minka, T. P. (2006). Principled hybrids of generative and discriminative models. In *2006 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'06)*, volume 1, pages 87–94. IEEE.

References III

Lee, M. and McNally, M. G. (2006). An empirical investigation on the dynamic processes of activity scheduling and trip chaining. *Transportation*, 33:553–565.

Liu, G. and Onnela, J.-P. (2021). Bidirectional imputation of spatial gps trajectories with missingness using sparse online gaussian process. *Journal of the American Medical Informatics Association*, 28(8):1777–1784.

McCool, D., Lugtig, P., and Schouten, B. (2022). Maximum interpolable gap length in missing smartphone-based gps mobility data. *Transportation*, pages 1–31.

McGuckin, N. and Murakami, E. (1999). Examining trip-chaining behavior: Comparison of travel by men and women. *Transportation Research Record*, 1693(1):79–85.

Nishii, K., Kondo, K., and Kitamura, R. (1988). An empirical analysis of trip chaining behavior.

References IV

Rao, W., Wu, Y.-J., Xia, J., Ou, J., and Kluger, R. (2018). Origin-destination pattern estimation based on trajectory reconstruction using automatic license plate recognition data. *Transportation Research Part C: Emerging Technologies*, 95:29–46.

Rittel, H. W. and Webber, M. M. (1973). Dilemmas in a general theory of planning. *Policy sciences*, 4(2):155–169.

Wang, Z., Xing, W., Kirby, R., and Zhe, S. (2022). Physics informed deep kernel learning. In *International Conference on Artificial Intelligence and Statistics*, pages 1206–1218. PMLR.

Álvarez, M., Luengo, D., and Lawrence, N. D. (2009). Latent force models. In van Dyk, D. and Welling, M., editors, *Proceedings of the Twelth International Conference on Artificial Intelligence and Statistics*, volume 5 of *Proceedings of Machine Learning Research*, pages 9–16, Hilton Clearwater Beach Resort, Clearwater Beach, Florida USA. PMLR.