Modeling Human Mobility from GPS Traces: Gaussian Processes, Physics-regularization, and Beyond

Ekin Uğurel

University of Washington
Department of Civil and Environmental Engineering
Presented at UW Urban Design & Planning/Interdisciplinary Research Seminar

February 13, 2024

Motivation

- The past: active solicitation (i.e., travel surveys)
  - Low sample sizes
  - Mixed reporting accuracy
  - Demographic info available
- The present (and future): passively-generated mobile data
  - Massive sample sizes
  - Found “in the wild”; data points are not generated due to any research-related processes
  - Prevalence of sparsity (large chunks of missing data)

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

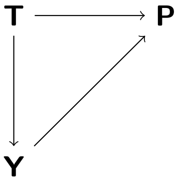
- Travel behavior heterogeneity at the individual-level (Bayarma et al., 2007; Kitamura and Van Der Hoorn, 1987; McGuckin and Murakami, 1999; Nishii et al., 1988; Lee and McNally, 2006).
- Physical system complexities imposed by the built and natural environments

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**

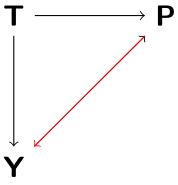


Problem Definition

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 $\mathbf{Y}$



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?
  - Note that this is different than the estimation problem $\mathbf{T} \rightarrow \mathbf{Y} \leftarrow \mathbf{P}$, when all variables are observed (albeit noisy)

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?
  - Note that this is different than the estimation problem $\mathbf{T} \rightarrow \mathbf{Y} \leftarrow \mathbf{P}$, when all variables are observed (albeit noisy)

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?
  - Note that this is different than the estimation problem $\mathbf{T} \rightarrow \mathbf{Y} \leftarrow \mathbf{P}$, when all variables are observed (albeit noisy)

[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.*
- 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.*

Multi-task Gaussian Process

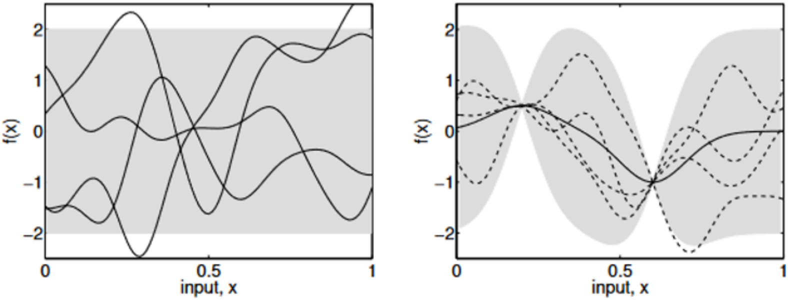
- First, let's focus on modeling the relationship $\mathbf{T} \rightarrow \mathbf{Y}$. Consider the task of learning a function $f_j : \mathbb{R}^d \rightarrow \mathbb{R}$ where j refers to either latitude ϕ or longitude λ. The basic form of our learning problem is

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

  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

Multi-task Gaussian Process

Assumption

The set $\mathbf{f}_j = [f_j(\mathbf{t}_1), \ldots, f_j(\mathbf{t}_n)]^\top$ *follows a multivariate normal distribution such that* $p(\mathbf{f}_j | \mathbf{T}) = \mathcal{N}(\mathbf{0}, \mathbf{K}_t)$, *where* $\mathbf{K}_t$ *is the covariance matrix such that* $[\mathbf{K}_t]_{i,g} = k_t(\mathbf{t}_i, \mathbf{t}_g)$.

We relate multiple tasks (each multivariate normal distributed) by leveraging the correlations between them. Thus, we specify the covariance matrix for all n observations and two tasks as

$$\mathbf{K}_t = \mathbf{K}^t(\mathbf{T}, \mathbf{T}) \otimes \mathbf{K}^f(\mathbf{y}_\lambda, \mathbf{y}_\phi), \tag{2}$$

where $\mathbf{K}^t$ is the $n \times n$ covariance matrix of the training times using any valid PSD kernel, $\otimes$ is the Kronecker product, and $\mathbf{K}^f$ is a PSD matrix containing the inter-task covariances (Bonilla et al., 2007). The dimension of $\mathbf{K}_t$ for two tasks is then $2n \times 2n$.

Model Inference

Assumption

The inferred location $f_j(\mathbf{t}_*)$ of a new input $\mathbf{t}_*$ conditioned on the training data is assumed to be distributed with the following form:

$$p(f_j(\mathbf{t}_*)|\mathbf{t}_*, \mathbf{T}, \mathbf{Y}, \delta_j^2) \sim \mathcal{N}(\boldsymbol{\mu}_*, \boldsymbol{\sigma}_*^2), \tag{3}$$

where

$$\begin{aligned} \boldsymbol{\mu}_* &= (\mathbf{k}_j^f \otimes \mathbf{k}_*)(\mathbf{K}^f \otimes \mathbf{K}^t + \mathbf{D} \otimes \mathbf{I})^{-1} vec(\mathbf{Y}) \\ \boldsymbol{\sigma}_*^2 &= (\mathbf{k}_j^f \otimes \mathbf{k}_{**}) - (\mathbf{k}_j^f \otimes \mathbf{k}_*)(\mathbf{K}^f \otimes \mathbf{K}^t + \mathbf{D} \otimes \mathbf{I})^{-1}(\mathbf{k}_j^f \otimes \mathbf{k}_*). \end{aligned} \tag{4}$$

Here, $\mathbf{k}_j^f$ selects the j^{th} column of $\mathbf{K}^f$, $\mathbf{k}_* = k(\mathbf{t}_*, \mathbf{T})$ is the vector of covariances between the test point and the training set, $\mathbf{D}$ is a 2×2 diagonal matrix with the variances of the noise processes for latitude and longitude δ_j^2, and $\mathbf{k}_{**} = k(\mathbf{t}_*, \mathbf{t}_*)$.

Model Training

We minimize the negative marginal log-likelihood (MLL) of the output vectors with respect to the training data in determining the optimal hyperparameters.

$$-\log(p(\mathbf{Y}|\mathbf{T},\Theta)) = \frac{1}{2}[\text{vec}(\mathbf{Y})^\top \mathbf{\Sigma}^{-1}\text{vec}(\mathbf{Y}) + \log(\det(\mathbf{K}_t)) + |\Theta|\log(2\pi)], \tag{5}$$

where Θ is the set of model parameters, $|\cdot|$ denotes cardinality, $\mathbf{\Sigma} = \mathbf{K}^f \otimes \mathbf{K}^t + \mathbf{D} \otimes \mathbf{I}$, and $\det(\mathbf{K}_t)$ is the determinant of the $\mathbf{K}_t$ matrix.

Kernels for Human Mobility

We consider the squared exponential (SE), rational quadratic (RQ), and Matern (MAT) 5/2 kernels:

$$k_{SE} = \eta \exp\left(-\frac{|x_i - x_g|^2}{2l^2}\right), \tag{6}$$

$$k_{RQ} = \eta \left(1 + \frac{|x_i - x_g|^2}{2\alpha l^2}\right)^{-\alpha}, \tag{7}$$

$$k_{MAT_{5/2}} = \eta \left(1 + \frac{\sqrt{5}|x_i - x_g|}{l} + \frac{5|x_i - x_g|^2}{3l^2}\right) \exp\left(-\frac{\sqrt{5}|x_i - x_g|}{l}\right), \tag{8}$$

where $|x_i - x_g|$ represents the Euclidian distance between any pair of inputs x_i and x_g; η is a scale parameter; the lengthscale l determines the smoothness of the function; and the scale mixture α determines the relative weight of large- and small-scale variations in the data.

Kernels for Human Mobility

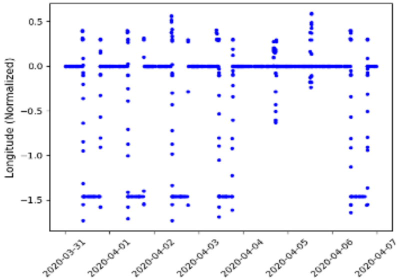
- We represent calendar-based structures like days, weeks, and months as binary variables using a one-of-k encoding.
  - For example, as the days of the week can take one of seven values $\{Mo, Tu, We, Th, Fr, Sa, Su\}$, a one-of-k encoding of We will correspond to $\{0, 0, 1, 0, 0, 0, 0\}$
- We embed categorical inputs in a GP framework by multiplying the same kernel across one-hot encodings

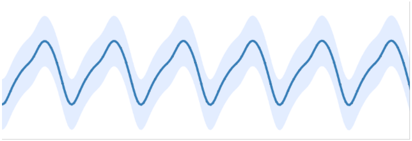
$$k_{RQ_{cat}} = \prod_{u=1}^{d} k_{RQ_u} \tag{9}$$

- We also use the periodic kernel, which allows GPs to model functions that repeat themselves

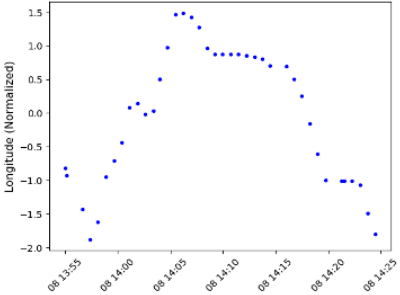
$$k_{PER} = \exp\left(-\frac{2\sin^2(\pi|x_i - x_g|/p)}{l^2}\right) \tag{10}$$

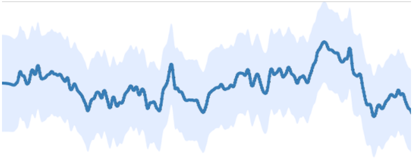
Kernels for Human Mobility





$$k = k_{PER} \times k_{SE}$$

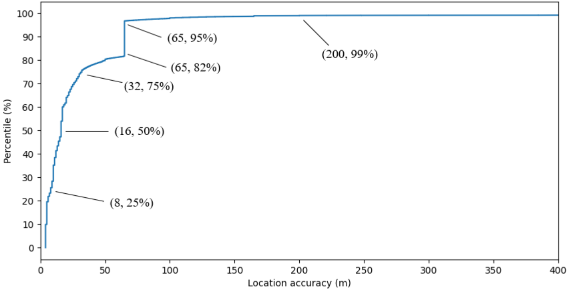


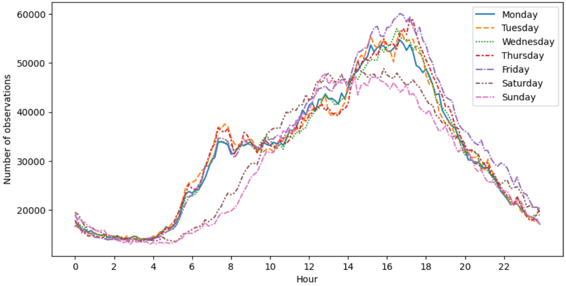


$$k = k_{RQ} \times k_{SE}$$

Dataset

- Privacy-protected mobile data from Spectus





Data Prepreprocessing

- 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

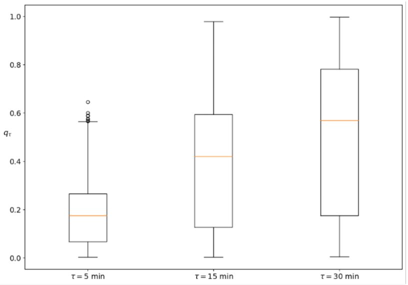
- Mobile data is irregularly sampled. Thus, we need a mathematical convention to denote varying levels of missingness
- Let $\mathcal{T}$ denote a the full length of a period, which we can discretize into P intervals of length τ. Let $\mathbf{I}_p$ be an indicator variable such that

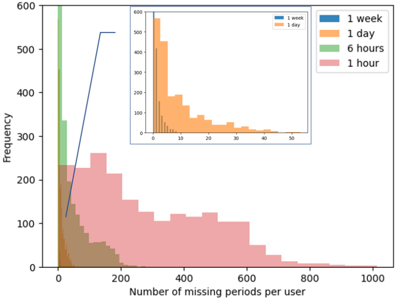
$$\mathbf{I}_p = \begin{cases} 1 & \text{if } p \text{ has at least one observation} \\ 0 & \text{otherwise} \end{cases}$$

- We can define temporal occupancy as

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

Varying Levels of Missingness





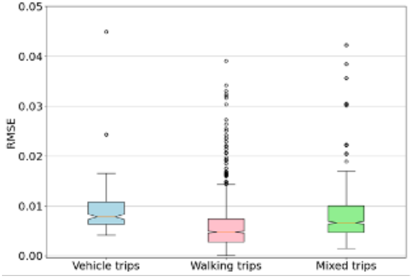
Experiment 1: Parameter Convergence

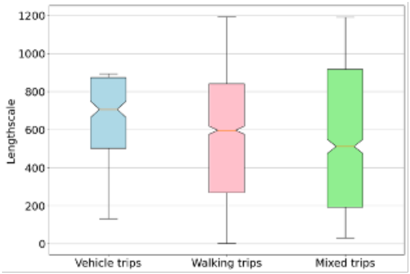
- K-means clustering to group together similar trips

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

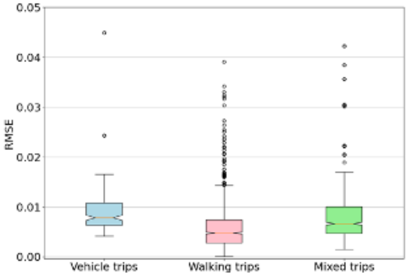
- Heading Change Rate: Ratio of consecutive points where a user changes direction with an angle exceeding a threshold (we use 0.33 rad)
- Velocity Change Rate: Ratio of consecutive points where the user exceeds a speed variation threshold (we use 26%)
- Stop Rate: Ratio of points with an inferred velocity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

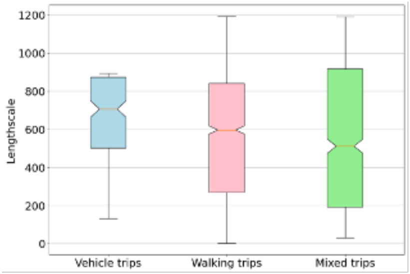
Experiment 1: Parameter Convergence





Experiment 1: Parameter Convergence





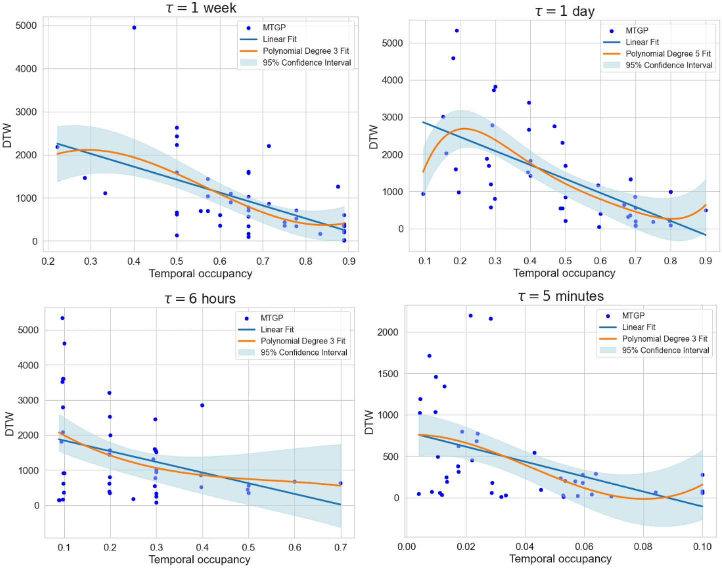
- The variability in lengthscales observed for walking trips may be attributed to the wide spectrum of walking behaviors.
- For mixed trips, the lower average lengthscale may be due to non-smooth transitions (or 'kinks') in the data introduced by mode changes

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.

Physics-regularized GP

Assume that the differential equation that describes the physical knowledge we want to embed in the GP takes the form

$$\Psi f(\mathbf{x}) = g(\mathbf{x}) \tag{11}$$

where Ψ is a differential operator and $g(\mathbf{x})$ is a latent function whose form we may not know. We propose the following system of equations that describe fine-grained individual human mobility

$$f_\lambda(t, v, \Theta) = \int_0^t v\cos(\Theta)dt, \quad f_\phi(t, v, \Theta) = \int_0^t v\sin(\Theta)dt \tag{12}$$

where taking the partial derivative with respect to time results in

$$\frac{\partial f_\lambda}{\partial t} = v\cos(\Theta) = v_\lambda, \quad \frac{\partial f_\phi}{\partial t} = v\sin(\Theta) = v_\phi \tag{13}$$

Physics-regularized GP

Estimating observations of segment velocity and bearings is then as easy as

$$v = \sqrt{v_\lambda^2 + v_\phi^2} \tag{14}$$

$$\beta = \arctan\left(\frac{v_\phi}{v_\lambda}\right) = \arctan\left(\frac{\partial f_\phi}{\partial f_\lambda}\right) \tag{15}$$

where β denotes the bearing. From here, we need to come up with a way to derive "virtual" observations of these variables at new locations $\mathbf{Y}_{gen}$ and times $\mathbf{T}_{gen}$. This would be the generative model for the latent function $g(\mathbf{x})$. The conditional component of the proposed model is that of the multi-task GP with predictions sampled from Equation 4.

Step 1: Generate a set of locations

We first generate set of locations (not necessarily in either training or testing data)

$$\mathbf{Y}_{gen} = [\mathbf{Y}_{gen,1}, \dots, \mathbf{Y}_{gen,m}]^\top = \begin{bmatrix} y_{gen,1\phi}, \cdots, y_{gen,m\phi} \\ y_{gen,1\lambda}, \cdots, y_{gen,m\lambda} \end{bmatrix}^\top$$

induced by a set of times $\mathbf{T}_{gen}$ using the conditional GP $f(\mathbf{T}_{cond}) \sim \mathcal{GP}(\mathbf{0}, \mathbf{K}_t(\mathbf{T}_{cond}, \mathbf{T}_{cond}))$ via Equations 3 and 4. The multivariate Gaussian projection of g on $\mathbf{Z} = [\mathbf{Y}_{gen}, \mathbf{T}_{gen}] = [\mathbf{z}_1, \dots, \mathbf{z}_m]$ can then be defined as

$$p(\mathbf{g}|\mathbf{Z}) = \mathcal{N}(\mathbf{0}, \mathbf{K}_\mathbf{Z}) = \mathcal{N}(\mathbf{g}, \mathbf{K}_\mathbf{Z}) \tag{16}$$

where $\mathbf{g} = [g(\mathbf{z}_1), \dots, g(\mathbf{z}_m)]^\top$, $[\mathbf{K}_\mathbf{Z}]_{ij} = k(\mathbf{z}_i, \mathbf{z}_j)$.

Step 2: Linking the conditional GP to the differential equation $g(\mathbf{x})$

The key idea here is that from the GP posterior distribution, we can construct a sample of the target function $f(\cdot) = \mu(\cdot) + \epsilon\sqrt{v(\cdot)}$ where $\epsilon \sim \mathcal{N}(0,1)$, $\mu(\cdot)$ is the posterior mean, and $\sqrt{v(\cdot)}$ is the posterior standard deviation functions. While ϵ makes f a random function, it still has a closed form and we can apply the differential operator Ψ to obtain the new function g

$$g(\cdot) = \Psi[\mu(\cdot) + \epsilon\sqrt{v(\cdot)}] \tag{17}$$

Thus, to sample the values of g on pairs of $\mathbf{y}_i$ and $\mathbf{t}_i$, we can sample from

$$p(\mathbf{g}|\epsilon, \mathbf{T}_{cond}, \mathbf{Y}_{cond}) = \prod_{j=1}^{m} \delta\left(\tilde{g}_j - \Psi\left[\mu(\mathbf{z}_j) + \epsilon\sqrt{v(\mathbf{z}_j)}\right]\right), \tag{18}$$

where $\tilde{g}_j = g(\mathbf{z}_j)$ and $\delta(\cdot)$ is the Dirac delta prior.

Step 3: The generative model

We can then sample virtual observations of the physical knowledge $\mathbf{P}_{gen}$ at the generated locations $\mathbf{Y}_{gen}$ and times $\mathbf{T}_{gen}$. This sampling process creates the generative model, with probability

$$\begin{aligned} p(\mathbf{P}_{gen}, \mathbf{Y}_{gen} | \mathbf{g}, \mathbf{T}_{gen}) &= p(\mathbf{Y}_{gen}) p(\mathbf{0} | \mathbf{Y}_{gen}, \mathbf{g}) \\ &= p(\mathbf{Y}_{gen}) \mathcal{N}(\mathbf{g}, \mathbf{K_Z}) \end{aligned} \tag{19}$$

where we marginalize the latent variable $p(\mathbf{T}_{gen})$ as it is modeler-specified and not dependent on any other variables.

Step 4: Combining the conditional and generative model

Using equations ...., we obtain the joint probability

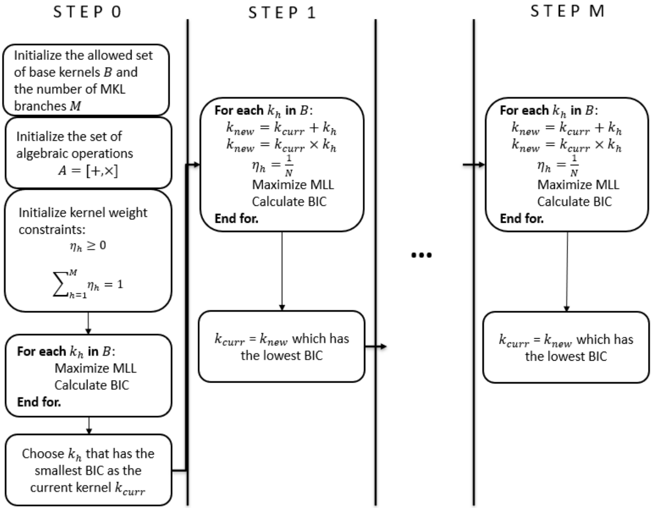
$$p(\mathbf{Y}_{cond}, \mathbf{P}_{gen}, \mathbf{Y}_{gen}, \epsilon, \mathbf{g} | \mathbf{T}_{cond}) \\ = p(\mathbf{Y}_{cond} | \mathbf{T}_{cond}) p(\mathbf{Y}_{gen}) p(\epsilon) p(\mathbf{g} | \epsilon, \mathbf{T}_{cond}, \mathbf{y}) p(\mathbf{P}_{gen} | \mathbf{g}, \mathbf{Y}_{gen}) \quad (20)$$

which simplifies to

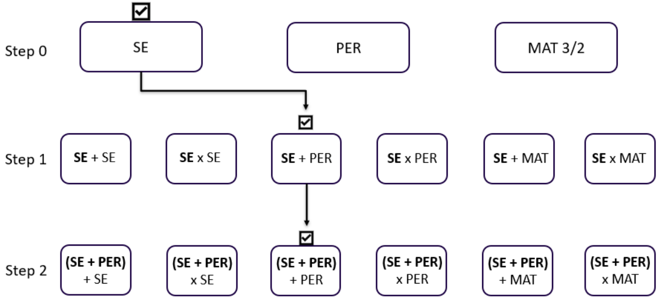
$$p(\mathbf{Y}_{cond}, \mathbf{P}_{gen}, \mathbf{Y}_{gen}, \epsilon, \mathbf{g} | \mathbf{T}_{cond}) \\ = \mathcal{N}(\mathbf{0}, \mathbf{K}_t) p(\mathbf{Y}_{gen}) \mathcal{N}(0, 1) \prod_{j=1}^{m} \left(\tilde{g}_j - \Psi \left[\mu(\mathbf{z}_j) + \epsilon \sqrt{v(\mathbf{z}_j)} \right] \right) \mathcal{N}(\mathbf{g}, \mathbf{K}_{\mathbf{Z}}) \quad (21)$$

Learning Multiple Kernel Structures

- To account for individual heterogeneity, we need a systematic method to find the optimal mix of kernels
- We propose a greedy multiple kernel learning algorithm

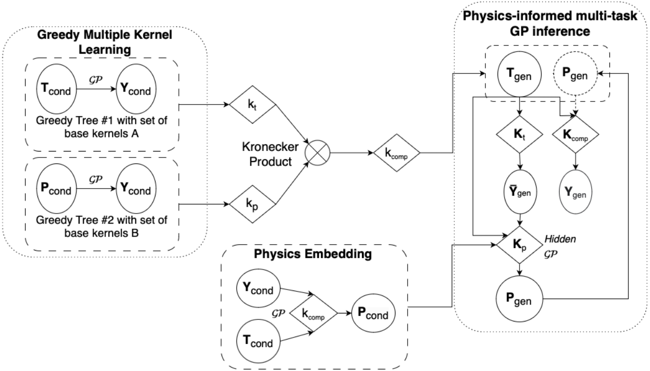


Learning Multiple Kernel Structures



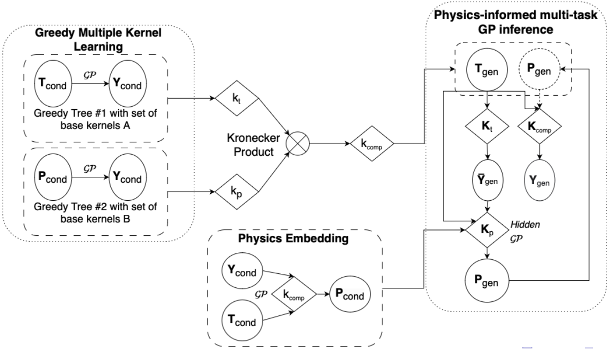
Model Inference

- Conditional-generative inference model (Lasserre et al., 2006)
- First, generate a set of noisy locations $\bar{\mathbf{Y}}_{gen} = [\bar{\mathbf{y}}_{gen,1}, \ldots, \bar{\mathbf{y}}_{gen,m}]^\top$ induced by a set of times $\mathbf{T}_{gen}$ using the conditional multi-task GP $f^y \sim \mathcal{GP}(\mathbf{0}, \mathbf{K}_t)$



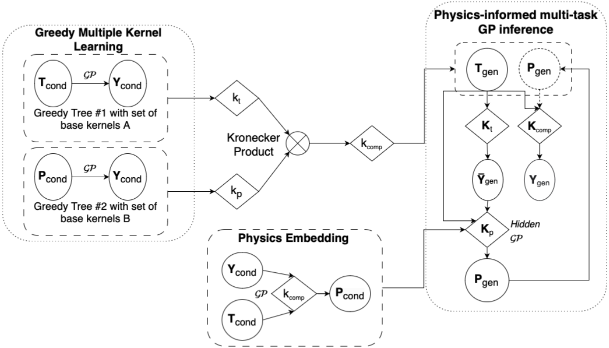
Model Inference

- To estimate $\mathbf{P}_{gen}$, we require a second posterior that takes in spatial and temporal observations $\mathbf{Z} = \begin{bmatrix} \mathbf{Y}_{cond} & \mathbf{T}_{cond} \end{bmatrix}$ and approximates a function $f^p : \mathbf{Z} \rightarrow \mathbf{P}_{cond}$.
- This is achieved by defining another multi-task GP for the physical variables $f^p \sim \mathcal{GP}(\mathbf{0}, \mathbf{K}_p)$ where $[\mathbf{K}_p]_{i,g} = k_{comp}(\mathbf{z}_i, \mathbf{z}_g)$. Sampling from the posterior distribution $\mathbf{P}_{gen} \sim p(\mathbf{T}_{gen})p(f^p|\bar{\mathbf{Y}}_{gen}, \mathbf{Y}_{cond}, \mathbf{T}_{cond}, \mathbf{P}_{cond}))$

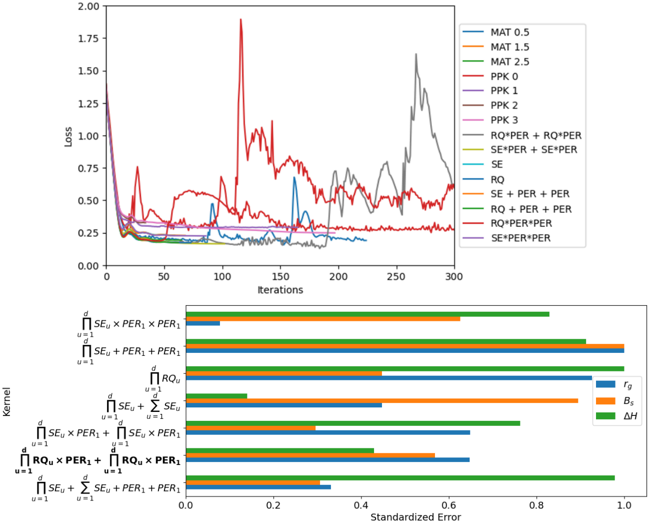


Model Inference

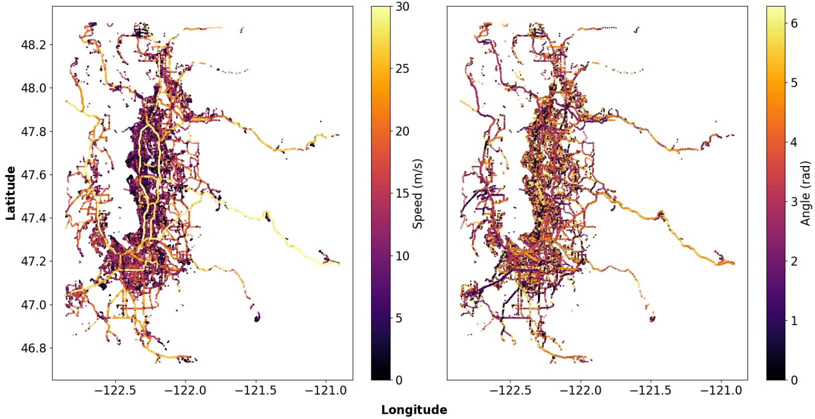
- The set of generated physical variables are then incorporated as inputs to the physics-regularized GP model.
- To generate physics-regularized synthetic data, we sample from the physics-regularized GP $f^y \sim \mathcal{GP}(\mathbf{0}, \mathbf{K}_{comp})$ where $[\mathbf{K}_{comp}]_{i,g} = k_{comp}(\mathbf{x}_i, \mathbf{x}_g)$.



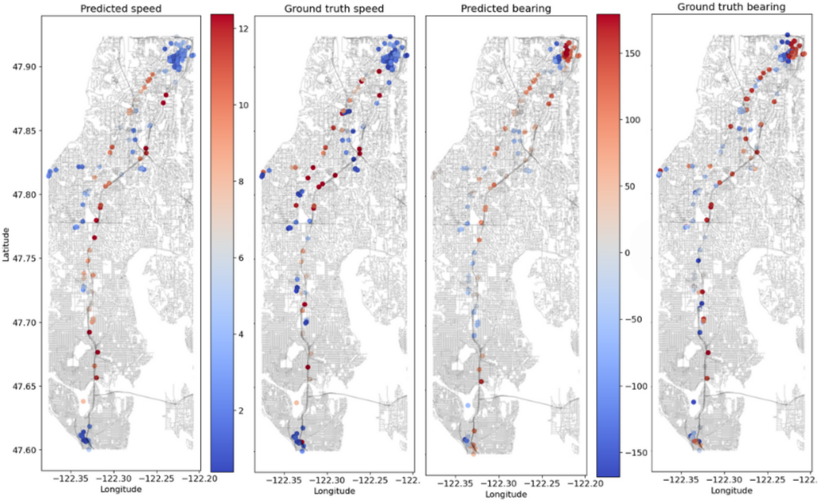
Impact of Kernel Choice



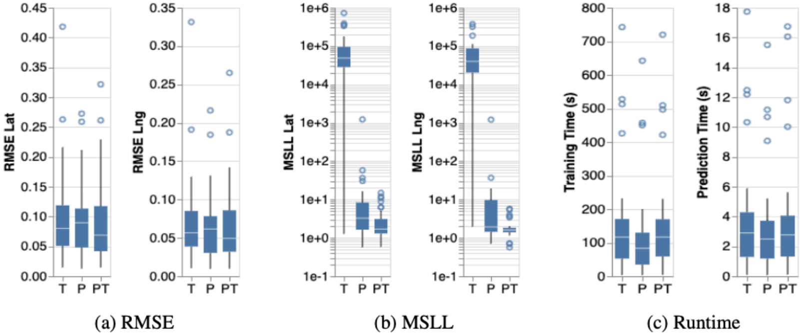
P_{gen} inference



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

Bonilla, E. V., Chai, K., and Williams, C. (2007). Multi-task gaussian process prediction. *Advances in neural information processing systems*, 20.

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.

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.

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

References III

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.

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.

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.

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.