FUTURE URBAN MOBILITY

Gaussian Process-Based Decentralized Data Fusion & Active Sensing Agents (UAI 2012, RSS 2013)

Towards Large-Scale Modeling & Prediction of Spatiotemporal Traffic Phenomena

J. Chen, **Bryan Kian Hsiang Low**, Ali Oran, & Patrick Jaillet

Multi-Agent Planning, Learning & Coordination Group (MapleCG)
Department of Computer Science, National University of Singapore
http://www.comp.nus.edu.sg/~lowkh/research.html
Singapore-MIT Alliance for Research & Technology (SMART)
Department of Electrical Engineering & Computer Science, MIT

J. Chen, K. H. Low, C. K.-Y. Tan, A. Oran, P. Jaillet, J. M. Dolan, & G. S. Sukhatme (2012). Decentralized Data Fusion & Active Sensing with Mobile Sensors for Modeling & Predicting Spatiotemporal Traffic Phenomena. In *Proceedings of the 28th Conference on Uncertainty in Artificial Intelligence (UAI)*, pages 163-173.
J. Chen & K. H. Low (2013). Gaussian Process-Based Decentralized Data Fusion and Active Sensing for Mobility-on-Demand System. In *Proceedings of the Robotics: Science and Systems (RSS)*.

This work is supported by SMART subaward agreements no. 33 & 41

SMART

FUTURE URBAN MOBILITY

Motivation

Objective. Modeling and predicting spatiotemporal traffic phenomena is important to the goal of achieving smooth-flowing, congestion-free traffic.

Practical applications. Route planning, detecting & forecasting congestion hotspots, road pricing, etc.

Sensor deployment is cost-constrained.

Drawbacks of using static sensors & passive mobile probes

- Sparse road network coverage
- Expensive installation & maintenance
- Abundant redundant data
- Inconsistent driving behaviors & privacy (mobile)
- No repositioning (static)

Deploy active mobile probes

- Since late 1920s
- Cover any segments of road network to **collect data that matters!**
- Consistent driving behavior

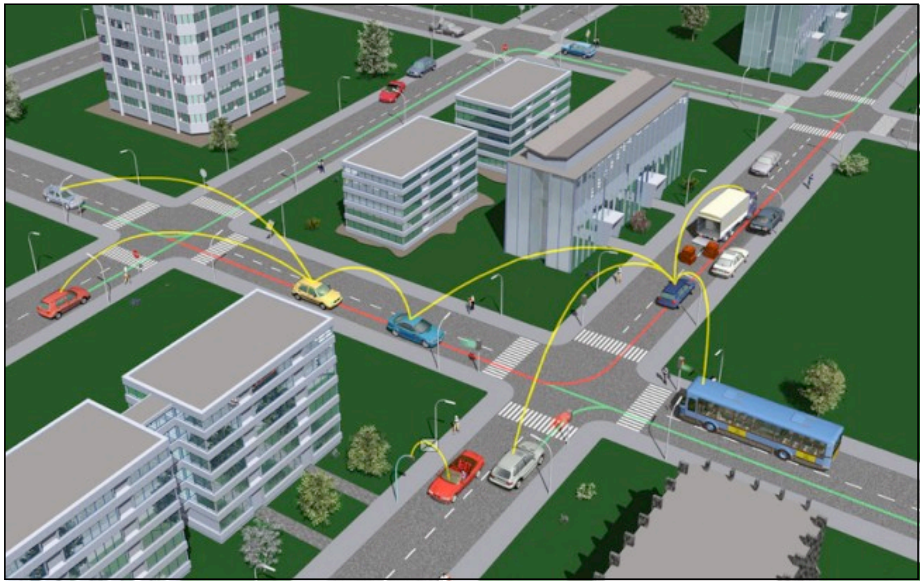
SMART

FUTURE URBAN MOBILITY

Research Problem & Challenges

Research problem

How do mobile probes actively explore the road network to gather & assimilate the most informative data for modeling & predicting a traffic phenomenon?



Challenges/Issues

Model for predicting traffic phenomena

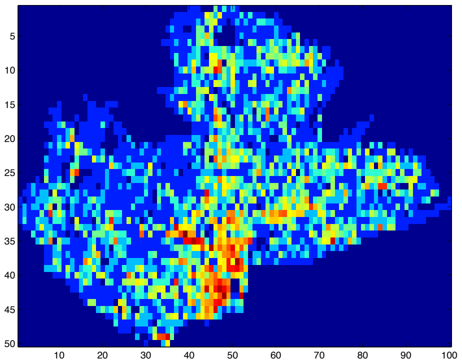
- Exploit spatiotemporal correlation structure to predict traffic phenomenon
- Exploit road segment features and network topology to model correlation

Data fusion

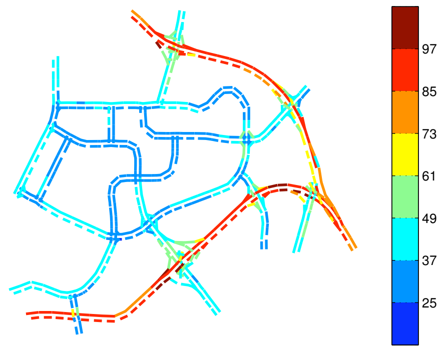
- Large data gathered "distributedly" by probes
- Real-time, efficient, scalable prediction and data fusion by decentralizing, parallelizing, and distributing in Google-like MapReduce paradigm

Active sensing

- Scale with large observations & probes



Mobility demand (pickups) pattern of CBD, SG



Traffic speeds (km/h) over urban road network of 775 segments in Tampines, SG during evening peak hours on April 20, 2011.

Gaussian Process (GP) Regression Model

Probabilistic regression model

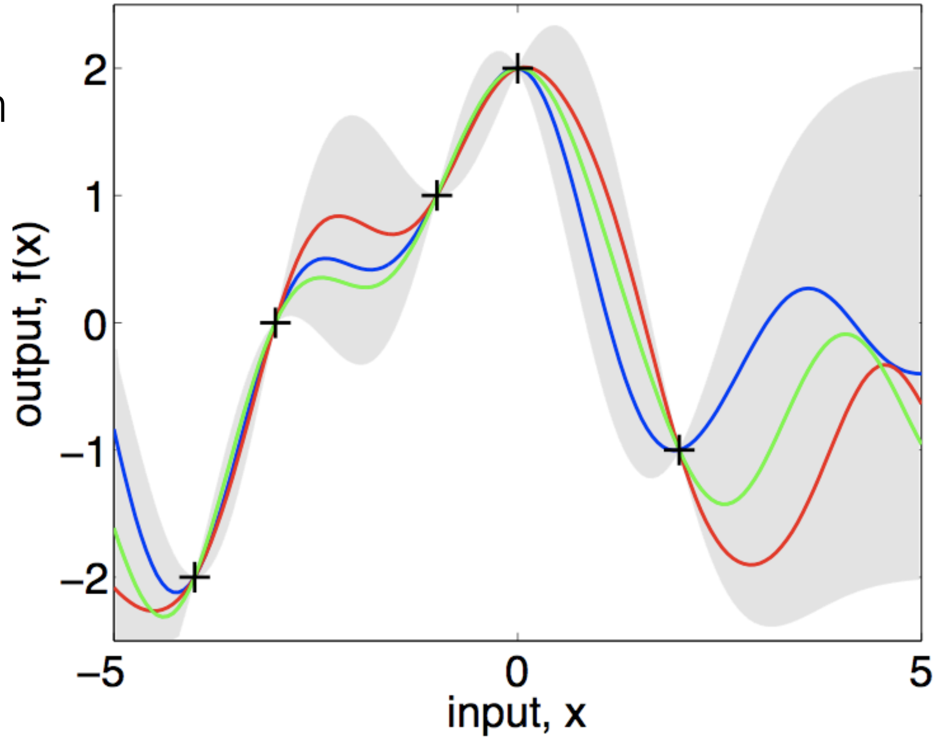
A Gaussian process (GP) is a set of random measurements, any finite subset of which has a joint Gaussian distribution.

Rich characterization of traffic phenomenon

Spatiotemporal correlation structure is exploited to predict traffic measurements (posterior mean & variance) of any unobserved road segment at any time using limited data.

Correlation of measurements between road segments

- Depend on road features (length, no. of lanes, speed limit) and network topology
- Graph-based kernel (UAI 2012)



Highway stretch: Define x as road segment or location & f(x) as corresponding speed measurement.

Formal measures of predictive uncertainty

Mobile probes can be directed to explore highly uncertain segments of road network.

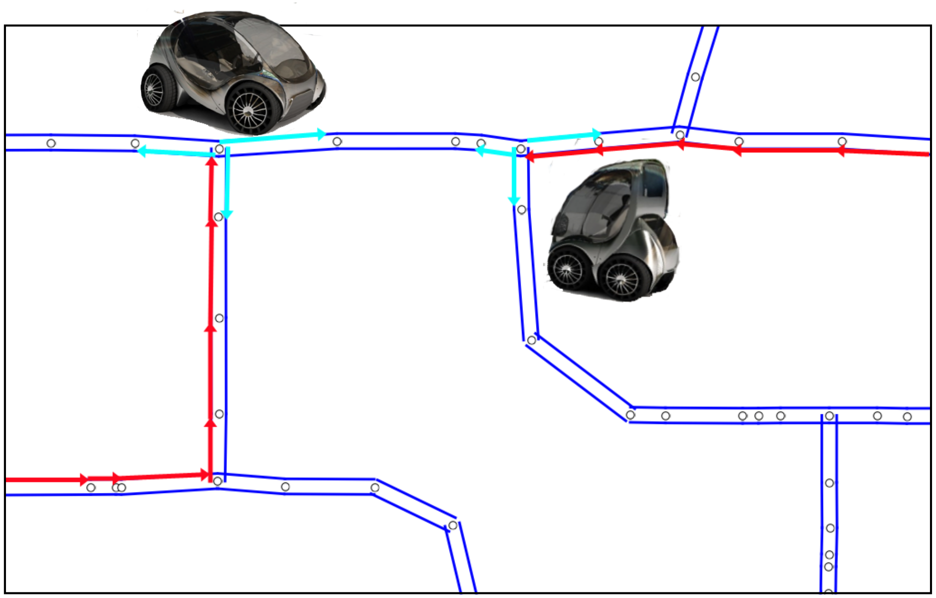
GP-Based Decentralized Data Fusion (GP-DDF)

Limitation of full GP

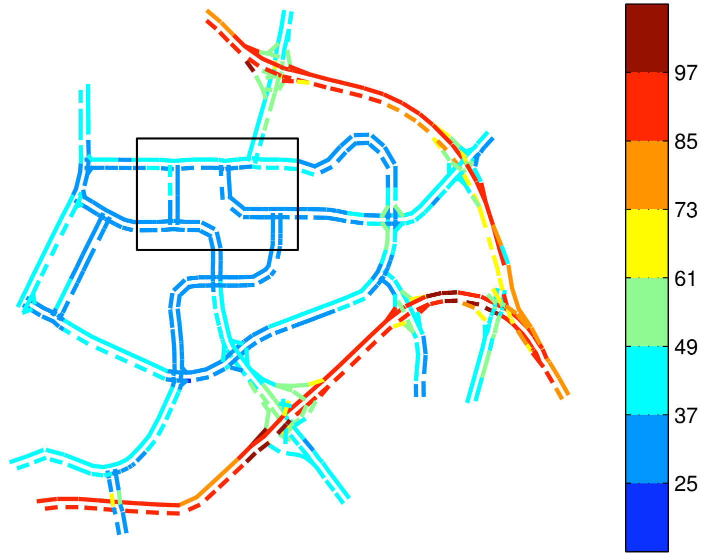
- Centralized: Cubic time in size of data
- Cannot perform prediction in real time

GP-DDF

- Distribute computational load among mobile agents to achieve efficient and scalable approximate GP prediction
- Idea. Each agent shares a **local summary** of its local data and assimilates them to form a **globally consistent summary** for prediction and active sensing



Traffic speeds (km/h) over urban road network of 775 segments in Tampines, SG during evening peak hours on April 20, 2011.



Theoretical Results

Performance guarantee. Predictive performance of GP-DDF is equivalent to that of a centralized sparse PITC approximation of full GP model (Quiñonero-Candela and Rasmussen, 2005).

Implication. Computation of centralized model can be distributed among all mobile agents, thereby achieving efficient and scalable prediction.

Time complexity. GP-DDF scales better than centralized GP models in size of data when number of agents is large.

Communication complexity. GP-DDF is more scalable since broadcasted local summary is independent of size of data.

FUTURE URBAN MOBILITY

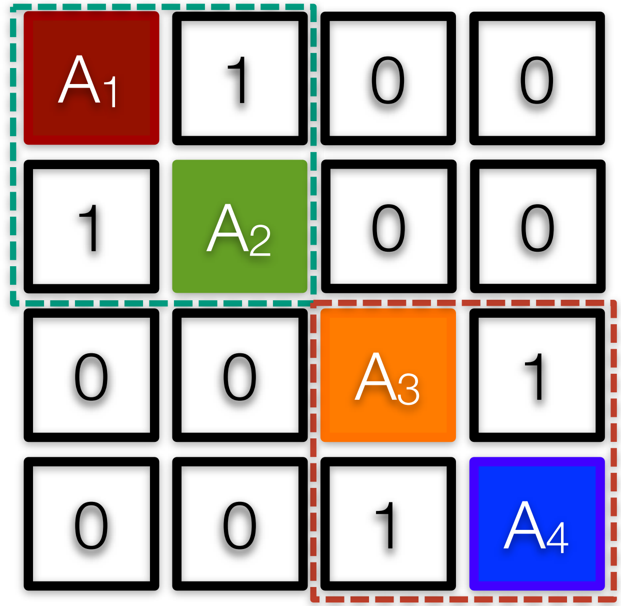
Decentralized Active Sensing (DAS)

Maximum entropy sampling

- Select max-entropy joint walk
- Minimize posterior joint entropy at remaining unobserved road segments
- Centralized: Exponential time in number of agents

Partially DAS

- Coordination graph: Two agents are adjacent if correlation between some pair of their possible walks is high enough
- Partition agents into disjoint subsets: Each corresponds to a connected component
- Select max-entropy joint walk for each subset



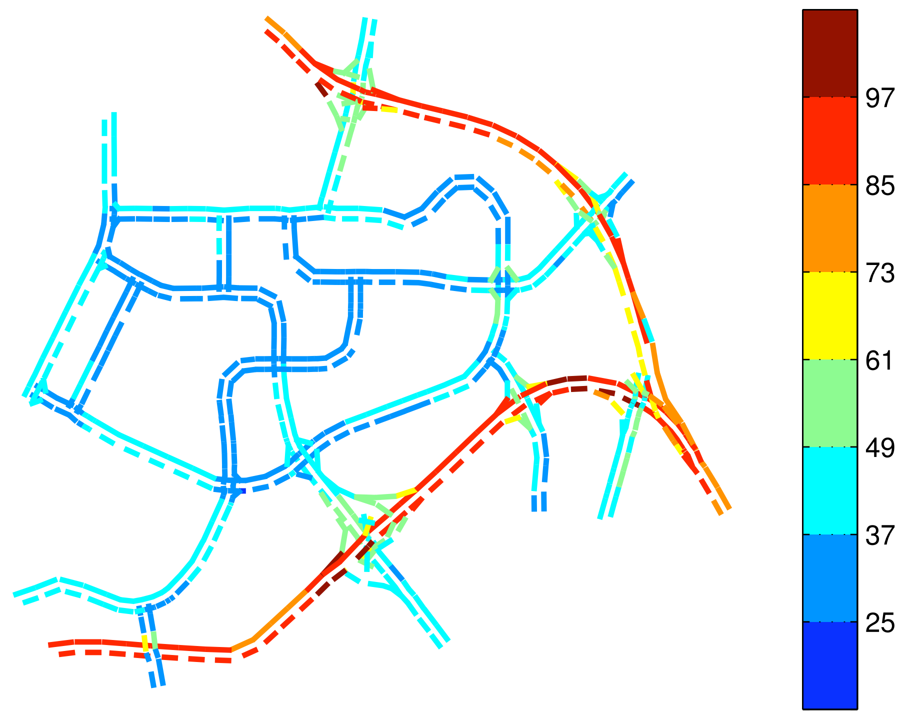
Adjacency matrix of coordination graph is used to determine its connected components.

Performance guarantee

Active sensing performance of DAS is near-optimal under various practical environmental conditions.

Experimental Setup

- A network of 4, 6, 8, 10, 20, 30 mobile agents is tasked to explore the road network to gather a total of up to 960 observations.
- Length of walk: 2
- Size of support set: 64
- Random initializations: 40



Traffic speeds (km/h) over urban road network of 775 segments in Tampines, SG during evening peak hours on April 20, 2011. Mean speed is 48.8 km/h and standard deviation is 20.5 km/h.

Comparing Data Fusion & Active Sensing Algorithms

Performance of D^2FAS (GP-DDF + DAS) is compared with that of two state-of-the-art centralized data fusion approaches to GP prediction coupled with centralized active sensing:

- Full GP (FGP): Full data is fused into GP model.
- Subset of Data (SoD) Approximation: Online greedy active subset selection of full data is performed, after which only the selected subset is assimilated into GP model.

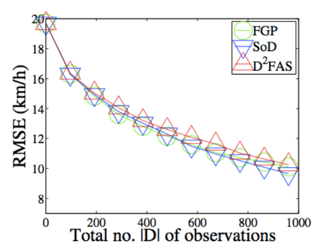
Experimental Results: Performance of D^{2}FAS

Time Efficiency

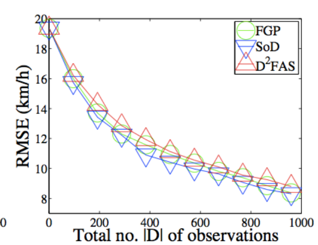
D^{2}FAS is significantly more time-efficient and scales better than centralized FGP and SoD with more data.

Predictive Accuracy

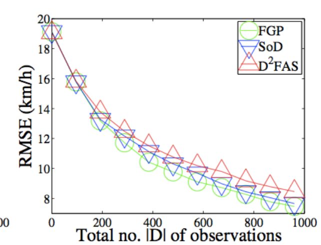
Predictive performance of D^{2}FAS is close to that of the centralized FGP and SoD.



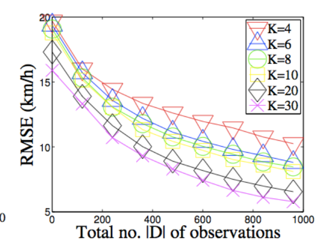
(a) $K = 4$



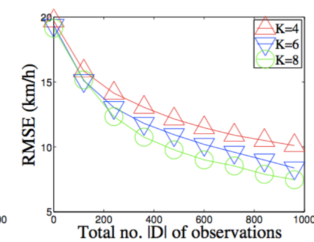
(b) $K = 6$



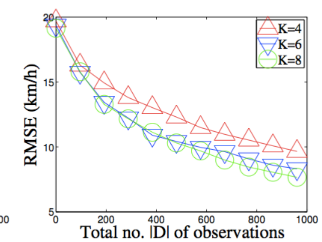
(c) $K = 8$



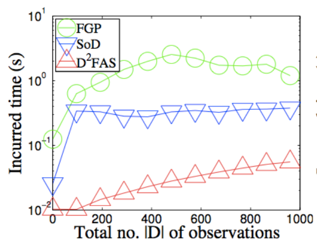
(a) D^{2}FAS



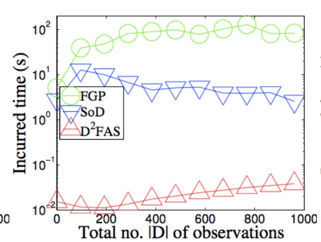
(b) FGP



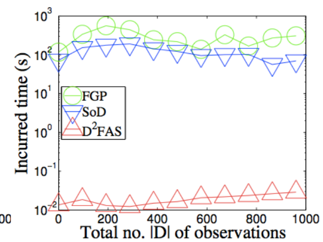
(c) SoD



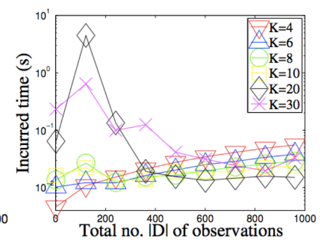
(d) $K = 4$



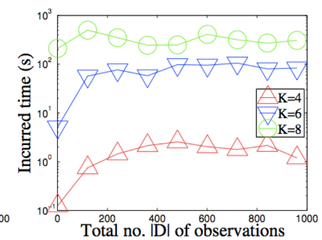
(e) $K = 6$



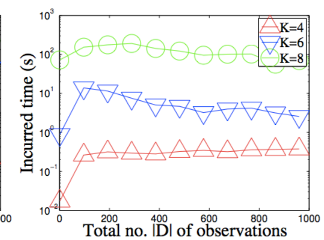
(f) $K = 8$



(d) D^{2}FAS



(e) FGP



(f) SoD

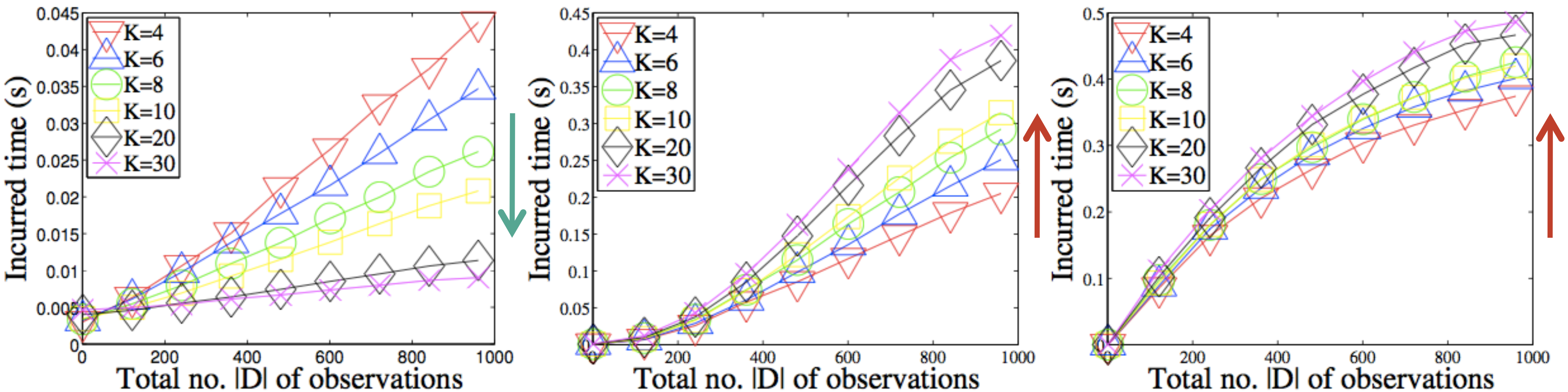
FUTURE URBAN MOBILITY

Experimental Results: Scalability of GP-DDF

Fully decentralized active sensing is used.

Time Efficiency

- With more agents, time incurred by D²FAS (FGP & SoD) decreases (increases)
- >=10 agents: D²FAS is at least 1 order of magnitude faster than FGP and SoD



(a) D^2FAS (b) FGP (c) SoD

SMART

Motivation

Private automobiles: Unsustainable personal urban mobility solutions

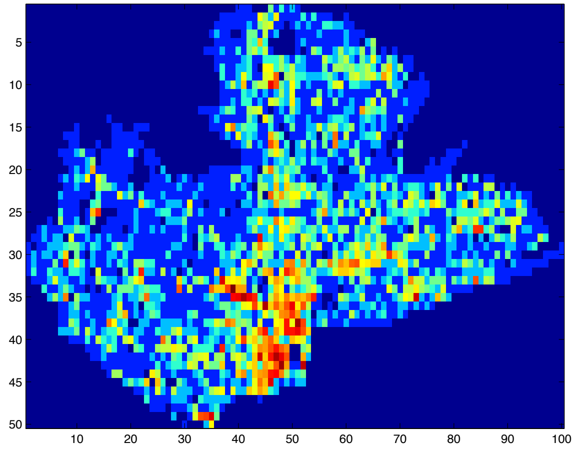
- 27.6% and 37% increase in private vehicles in HK & SG from 2003 to 2011
- Only 10% expansion of their road networks

Mobility-on-Demand (MoD) systems

- One-way vehicle sharing
- Racks and stacks of light electric vehicles
- Autonomous MoD vehicles cruise a road network to be hailed by users (like taxis)

Challenges (Mitchell, 2008)

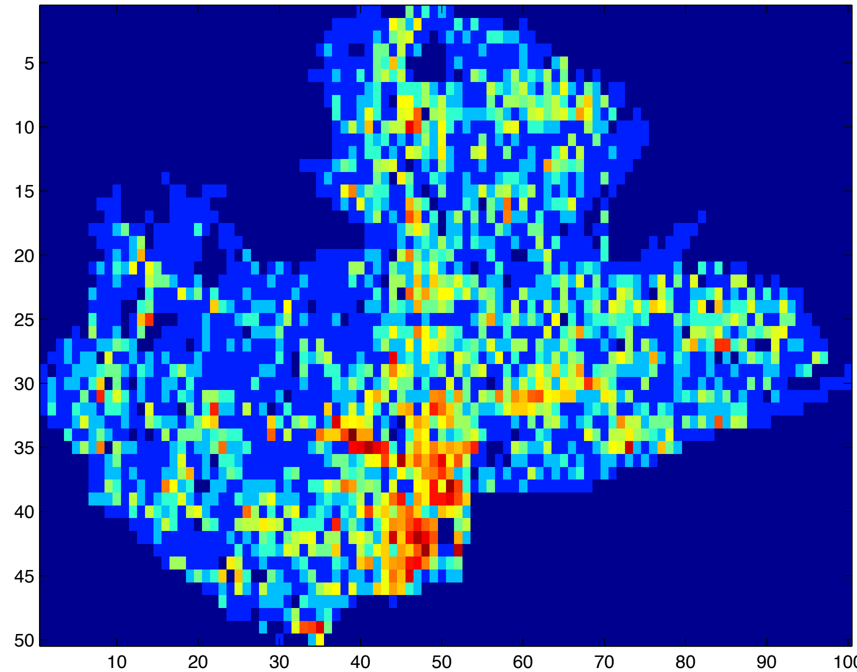
- Real-time, fine-grained mobility demand sensing, modeling, and prediction
- Real-time active fleet management to balance supply and demand



Mobility demand (pickups) pattern of CBD, SG

Research Problem

How do the vacant MoD vehicles actively cruise the road network to gather and assimilate the most informative demand data for predicting a mobility demand pattern?



Mobility demand (pickups) pattern of CBD, SG

Modeling & Predicting a Mobility Demand Pattern

- Gridding the service area into regions
- Pickup count (surrogate measure) of each region is tracked locally in a distributed manner (Jelasity et al., 2005)
- Stationary GP covariance structure violated by skewness & extremity → Problems: large hotspots predicted, small variances (Hohn, 1998)
- Model & predict log-demand using GP

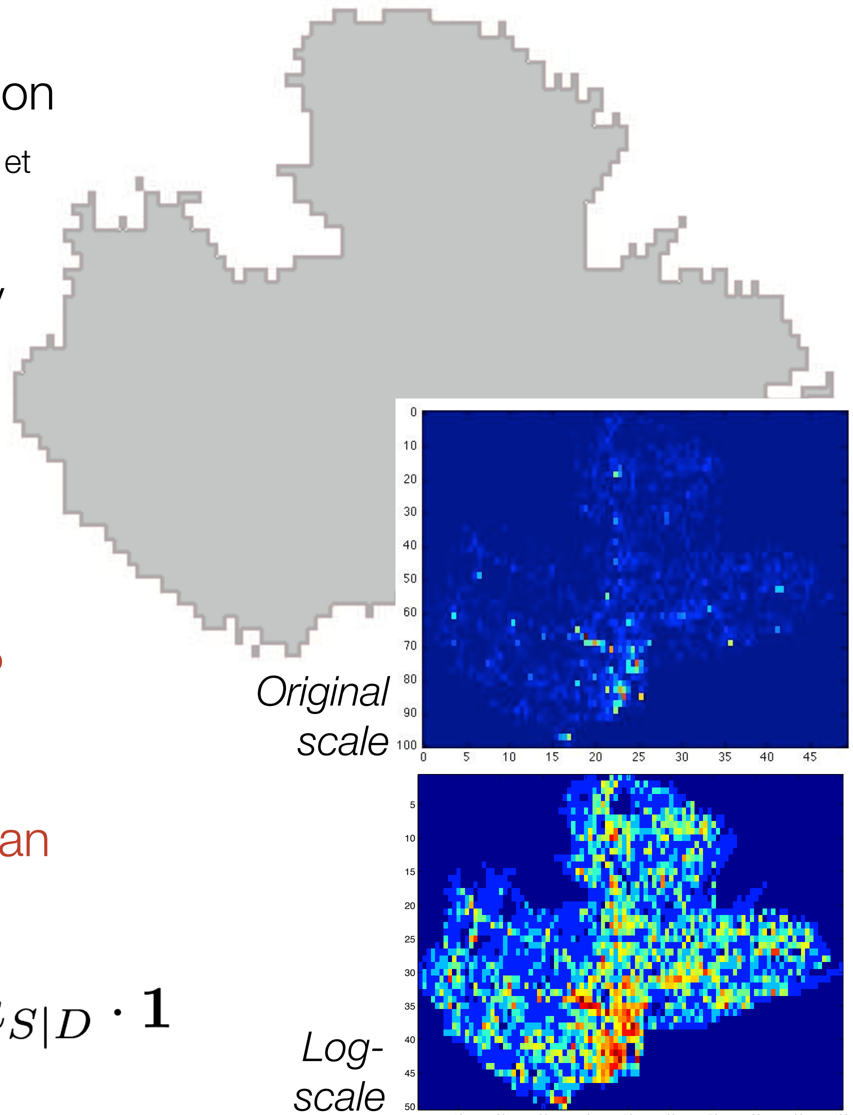
$$\mathcal{N}(\mu_{s|D}, \Sigma_{ss|D})$$

- Model & predict original demand using log-GP

$$\hat{\mu}_{s|D} \triangleq \exp(\mu_{s|D} + \Sigma_{ss|D}/2)$$

- Measure of predictive uncertainty: Log-Gaussian posterior entropy

$$\mathbb{H}[Y_S|Y_D] \triangleq \frac{1}{2}\log(2\pi e)^{|S|}\left|\Sigma_{SS|D}\right| + \mu_{S|D} \cdot \mathbf{1}$$



SMART

GP-Based Decentralized Data Fusion (GP-DDF$^+$)

GP-DDF

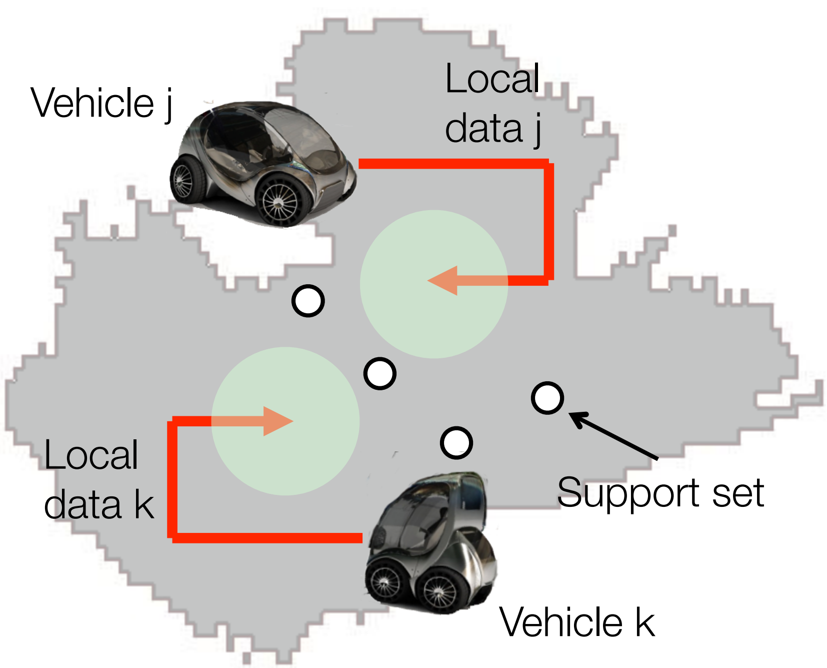
- Local data ➞ Local summary
- Local summaries ➞ Global summary
- Global summary ➞ Prediction & active sensing

Limitations of GP-DDF

- Information loss due to summarizing data
- Sparse coverage of hotspots by support set

GP-DDF$^+$

- Balances between predictive power of FGP and time efficiency of GP-DDF
- Idea. Exploit local data & summary information
- Global & local summaries + local data ➞ Prediction & active sensing



Performance guarantee

Predictive performance of GP-DDF$^+$ is equivalent to that of a centralized sparse PIC approximation of full GP model (Snelson, 2007).

Time complexity. Same as GP-DDF.

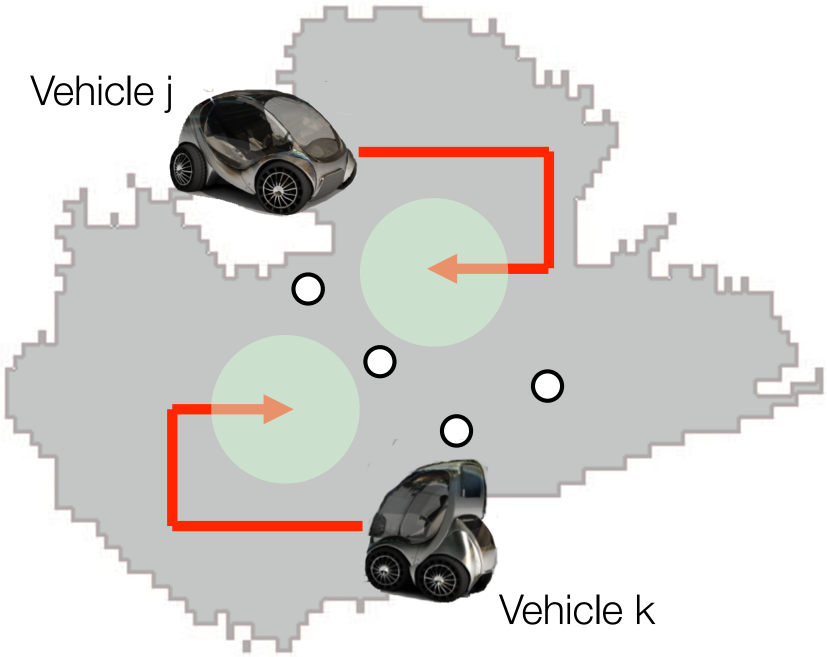
Decentralized Active Sensing

Maximum entropy sampling

- Select max-entropy walk

$$\mathbb{H}[Y_S|Y_D] \triangleq \frac{1}{2}\log(2\pi e)^{|S|}\boxed{\left|\Sigma_{SS|D}\right|} + \boxed{\mu_{S|D}\cdot\mathbf{1}}$$

- Cruising behavior: Explore hotspots and sparsely sampled areas
- Dual effect of fleet rebalancing to service mobility demands. Redistribute vacant MoD vehicles to these areas with high likelihood of pickups



FUTURE URBAN MOBILITY

Experimental Results: Performance of GP-DDF⁺

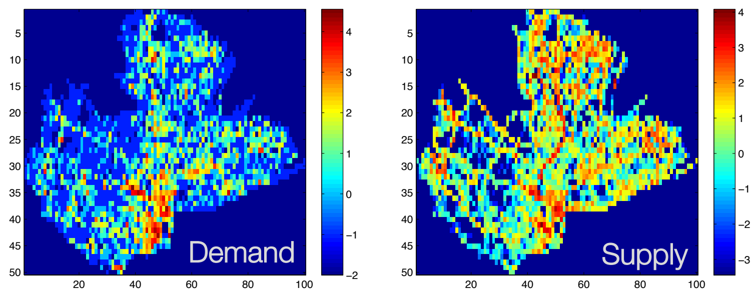
Experimental Setup

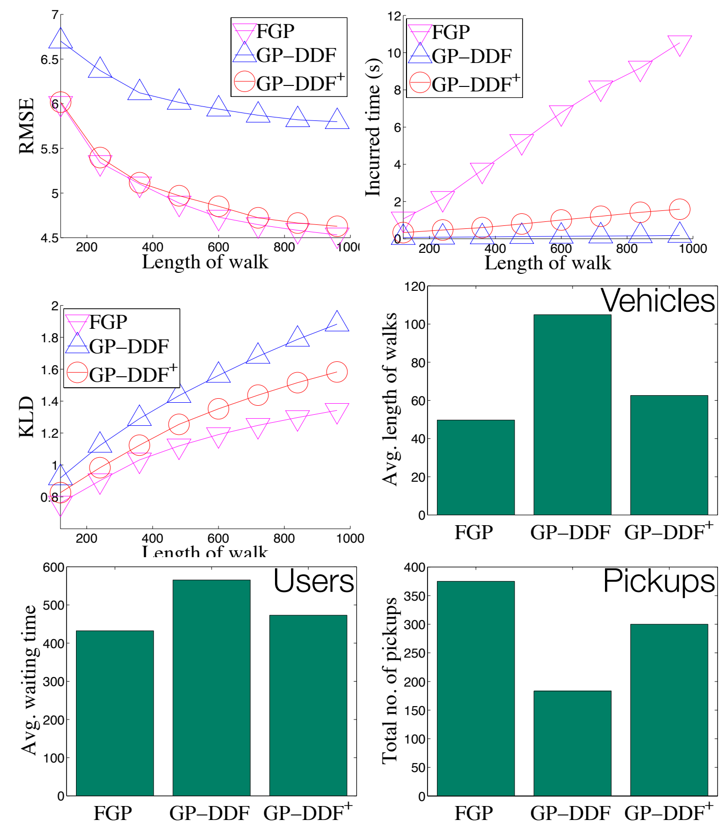
- 200 passengers, 20 MoD vehicles
- Results averaged over 30 instances

Observations

- GP-DDF⁺ achieves KL divergence (between vehicle and demand distributions) closer to centralized FGP than GP-DDF
- GP-DDF⁺ achieves RMSE lower than GP-DDF and comparable to FGP
- GP-DDF and GP-DDF⁺ are significantly more time-efficient than FGP
- GP-DDF⁺ achieves shorter cruising time, shorter waiting time, and more pickups than GP-DDF

Aug 2, 2010
9:30-10pm





SMART

FUTURE URBAN MOBILITY

Experimental Results: Scalability of GP-DDF$^+$

Experimental Setup

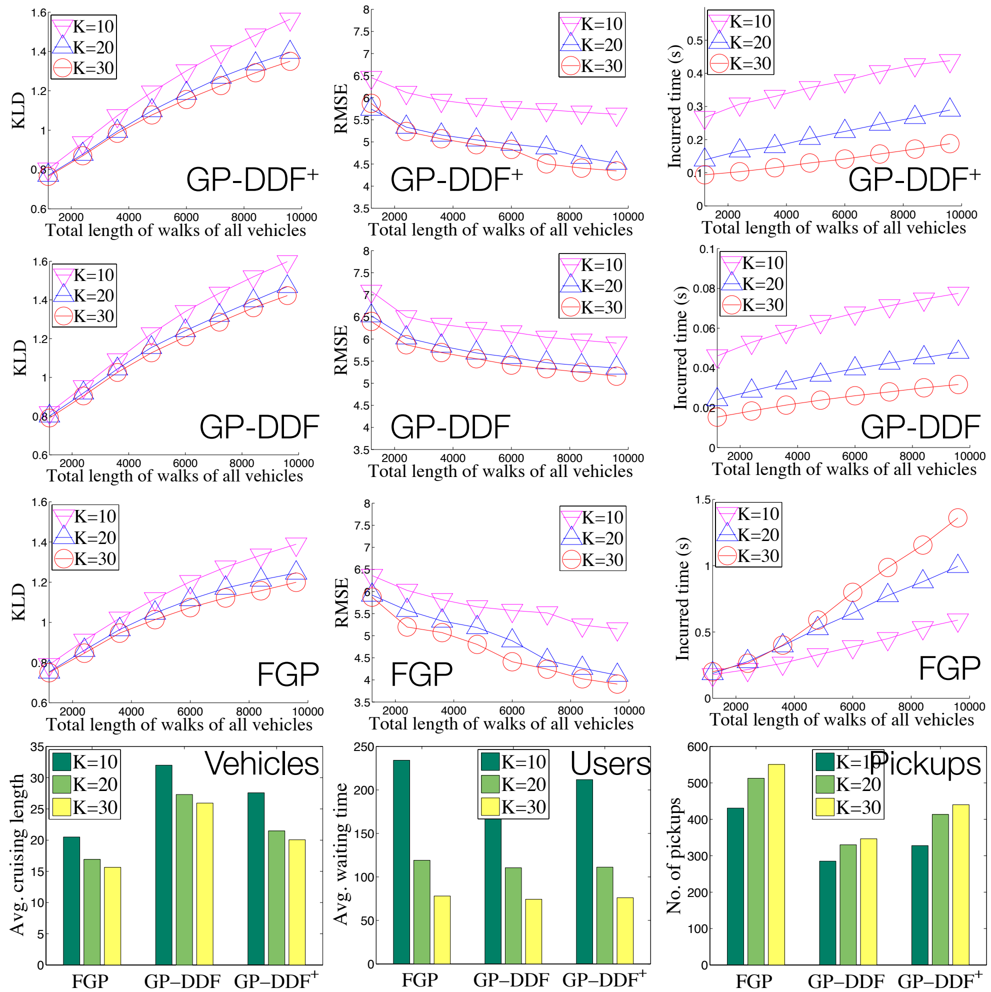
- 600 passengers
- Results averaged over 30 instances

Observations

By scaling to more agents (10→20→30), all three algorithms achieve

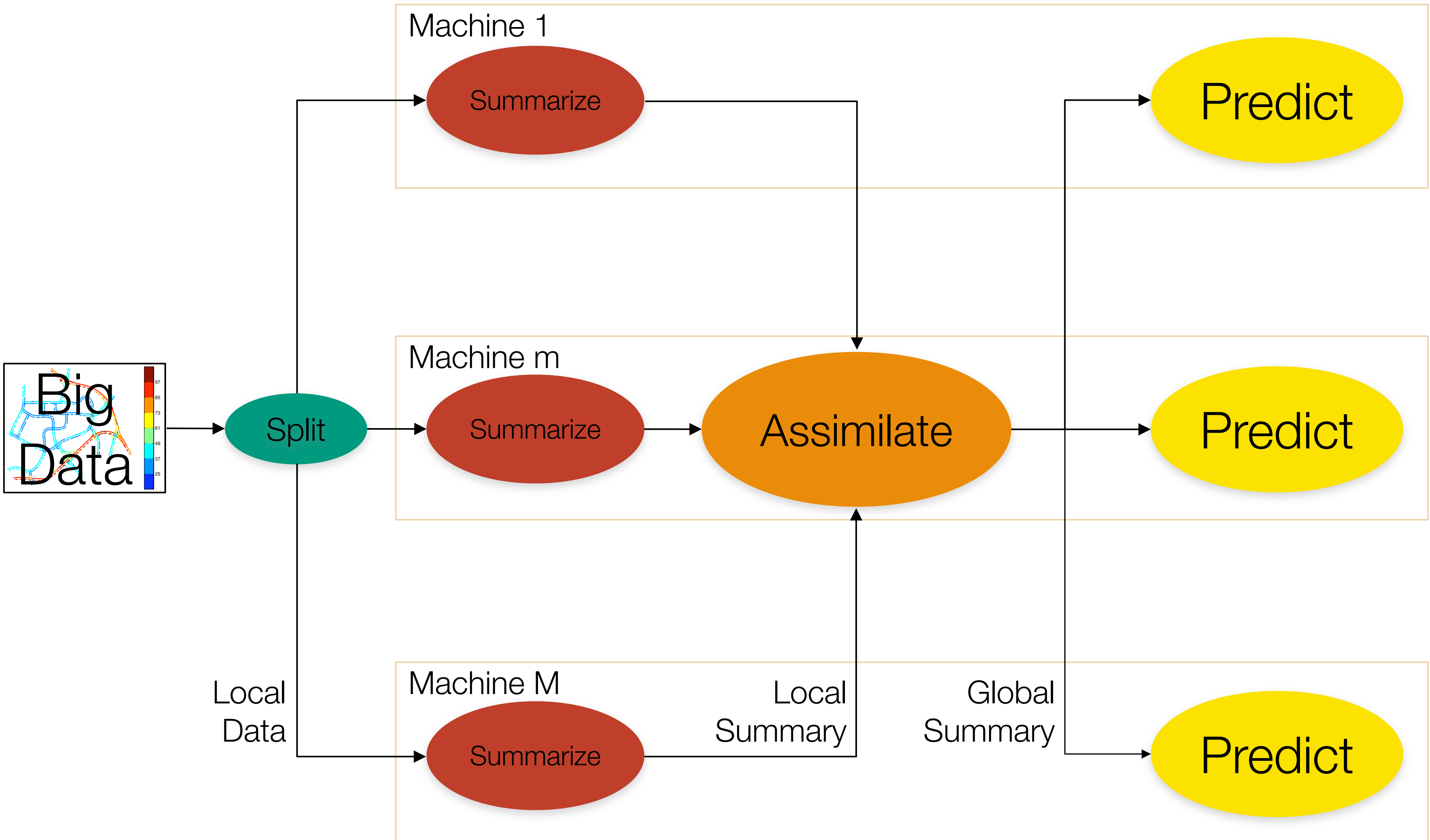
- Lower KL divergence;
- Lower RMSE;
- Shorter cruising time;
- Shorter waiting time; and
- More pickups.

But, GP-DDF and GP-DDF$^+$ become more time-efficient while FGP is less time-efficient.

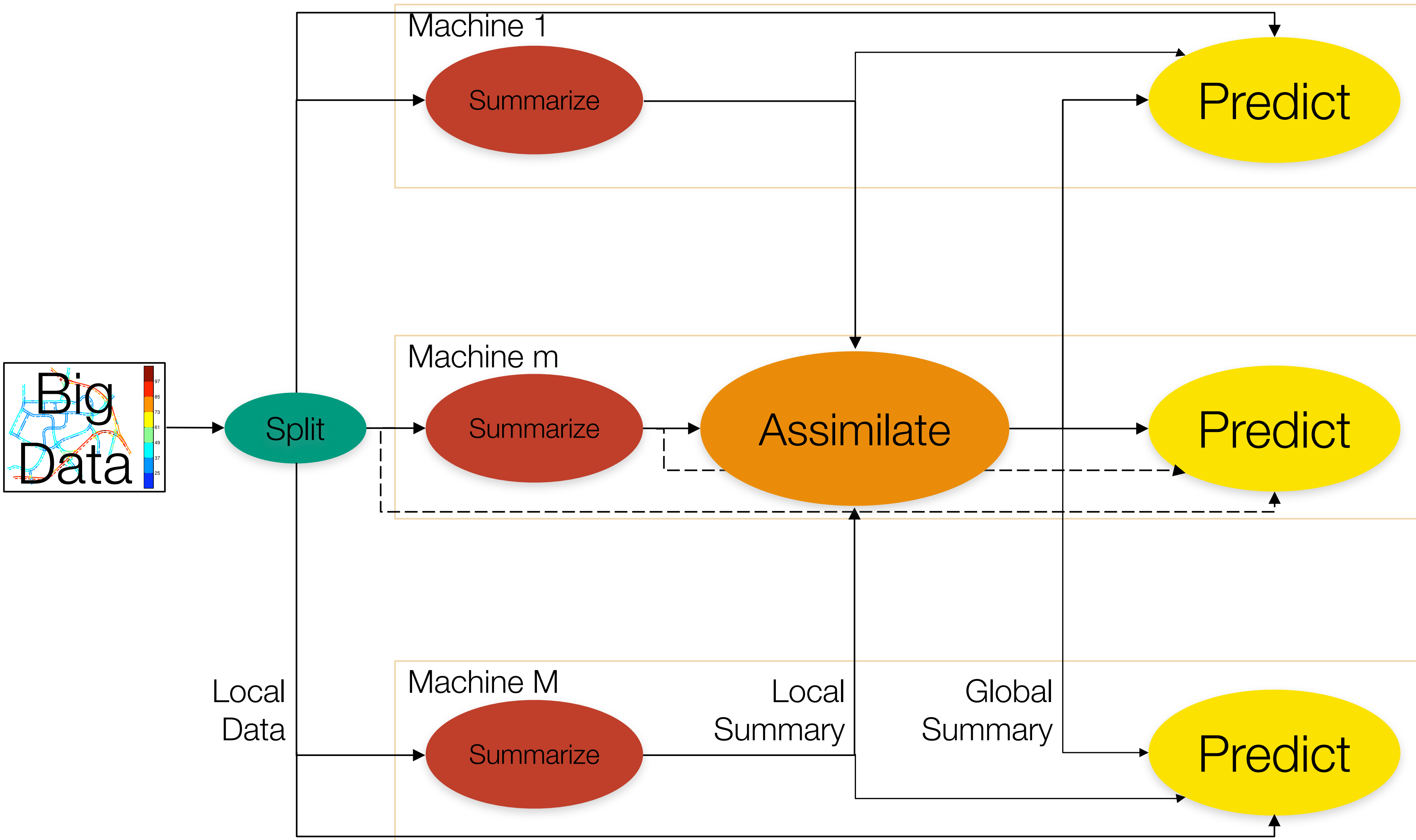


Parallel GP Regression: *p*PITC (MapReduce or MPI) (UAI 2013)

J. Chen, N. Cao, K. H. Low, R. Ouyang, C. K.-Y. Tan, & P. Jaillet (2013). Parallel Gaussian Process Regression with Low-Rank Covariance Matrix Approximations. In *Proceedings of the 29th Conference on Uncertainty in Artificial Intelligence (UAI)*.



Parallel GP Regression: *p*PIC (MapReduce or MPI)



FUTURE URBAN MOBILITY

Parallel GP Regression: Experimental Results

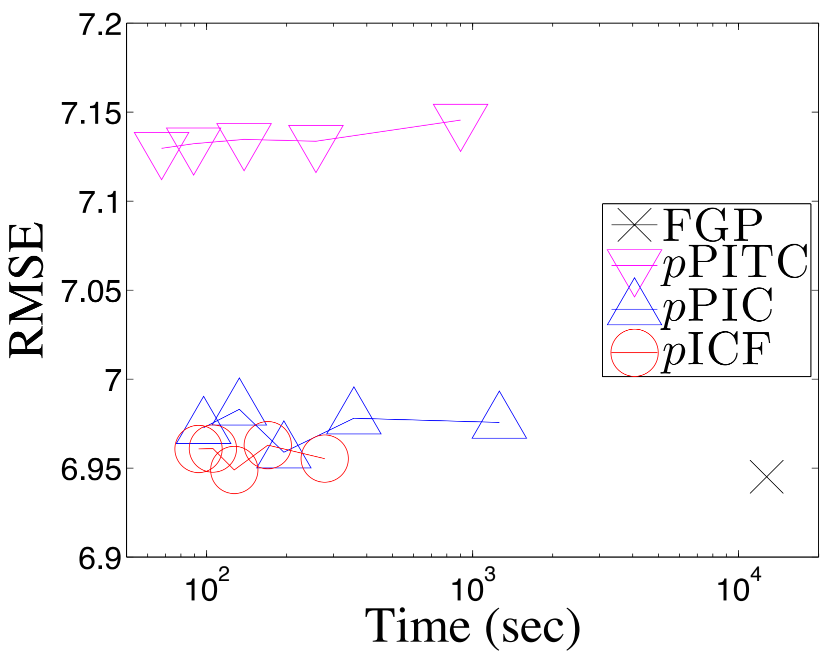
Time Efficiency

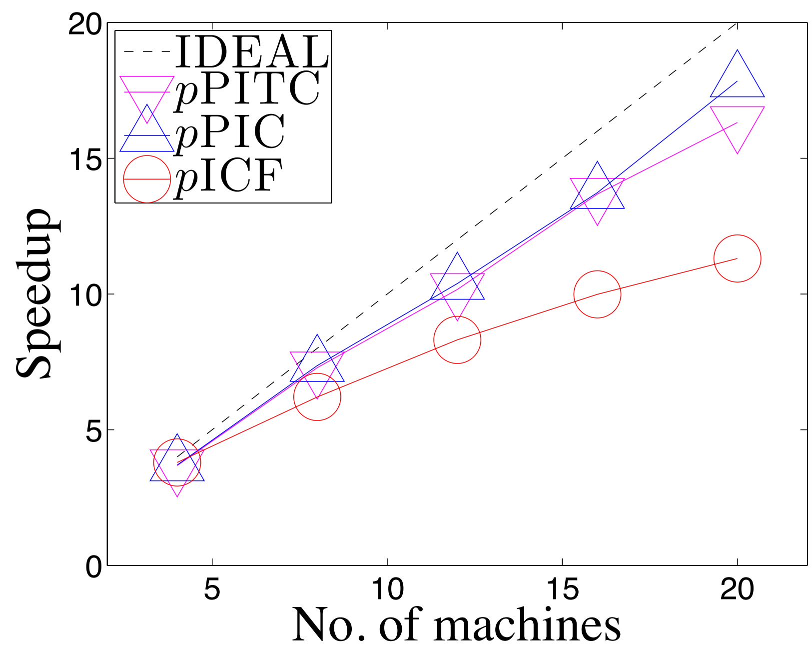
Parallelized GPs incur less time with larger number M of machines & 1-2 orders of magnitude less time than FGP.

For |D|=32000, parallelized GPs (M=20) incur 1-2 minutes while FGP incurs >3.5 hours.

Predictive Accuracy

Predictive performance of parallelized GPs is comparable to that of FGP (RMSE diff. <0.05 km/h).

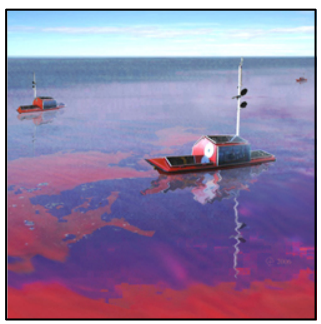




Active Sensing & Adaptive Sampling with GP & log-GP

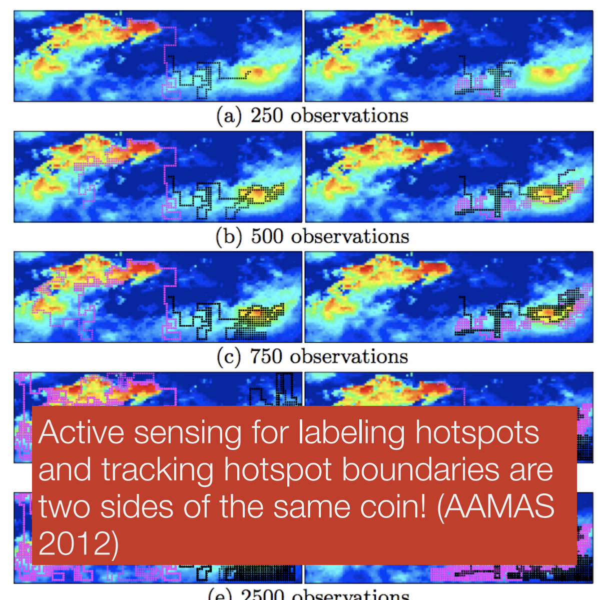
Joint work with John Dolan (CMU), Pradeep Khosla (CMU>UCSD), Steve Chien (JPL), David Thompson (JPL)

- Develop efficient GP-driven active sensing algorithms with performance guarantees
- Study formally how the structure and properties of spatiotemporally varying environmental phenomena affect the performance between different classes of active sensing algorithms
  - Degree of spatial patchiness of hotspots (AAMAS 2008, ICAPS 2009, AAMAS 2012), non-stationary, anisotropic (AAMAS 2011, AAMAS 2013) fields
  - Adaptive (AAMAS 2008, ICAPS 2009, AAMAS 2012) vs. non-adaptive (AAMAS 2011, AAMAS 2013), Markovian (AAMAS 2011, AAMAS 2013) vs. non-Markovian (AAMAS 2008, ICAPS 2009), greedy (AAMAS 2012) vs. non-myopic (AAMAS 2008, ICAPS 2009, AAMAS 2011, AAMAS 2013), centralized vs. distributed (AAMAS 2012, UAI 2012, RSS 2013)



Robotic adaptive path planning (AAMAS 2008, ICAPS 2009)

Robotic informative path planning (AAMAS 2011, AAMAS 2013)



Active sensing for labeling hotspots and tracking hotspot boundaries are two sides of the same coin! (AAMAS 2012)