

Creating Datasets to Study Urban Sustainability Problems in Developing Regions

Rijurekha Sen

Associate Professor, CSE, IITD


Road traffic dataset from Delhi and Noida intersections
(with Vehant and Aabmatica Technologies)

Road traffic dataset from Delhi and Noida intersections (with Vehant and Aabmatica Technologies)

7. List of Locations / Junctions – (Tentative)

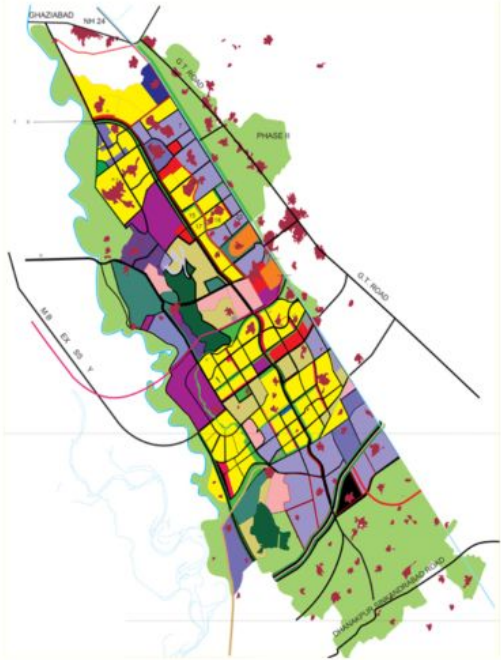
311 Locations / junctions as listed below are for Traffic management and Safe city. Other than these location / junctions 46 locations will be for vehicle speed detection devices. In total there would be $(311+ 46) = 357$ locations (ITMS, Safe City and Speed detection devices). Locations / junctions as per BoQ for VMDs, Location for PA System, Emergency call boxes (ECBs) and any other field component/s will also be made available to the successful bidder before starting the installation work. All the locations / junctions for field devices will be finalised in consultation with all the stakeholders and installation shall start only after prior approval by the authority. Implementation agency will take full responsibility to supply, install, test, commission, operate and maintain the overall solution as per RFP requirement. Variation in locations will be as per field feasibility and approval by Authority. GPS co-ordinates of all the locations will be finalised after detailed survey by the implementation agency (Successful bidder) and authority. Implementation agency shall prepare inception report and detailed project report (DPR) including all locations and junctions with its GPS coordinates along with all necessary details to make the system complete and submit the same to the Authority. Installation work shall start only after prior approval of Inception report and DPR submitted by IA.

S.no	Location / Junction	Police Station
1	P3 Circle SK Road	Beta-2
2	Grand Venice Mall	Beta-2
3	Kasna to Honda CL Chowk	Beta-2 /Kasna
4	Kasna Pulia	Beta-2/Kasna
5	Swarn Nagri Circle	Beta-2
6	Gram-Birauda Circle (Gail Society)	Beta-2
7	Wipro Circle	Beta-2
8	Dharna Sthal Near GNIDA office	Beta-2
9	Jaitpur Circle (Metro)	Beta-2
10	Delta Metro Gol Chakkar	Beta-2
11	Aicchar T Point	Beta-2
12	Achhar Market	Beta-2
13	Sector 37 Chauraha	Beta-2
14	Migsun Ultimo Gol Chakar	Beta-2
15	Aam ka Golchakar	Beta-2



Greater Noida Industrial Development Authority

* GREATER NOIDA CITY - PLANNED WITH AN OBSESSION *



Request for Proposal

E tender invitation no. – E&M-3/ 2024 / 1009 Dated: 07/08/2024

PM dataset collected with low cost platforms in Delhi and Kolkata

PM dataset collected with low cost platforms in Delhi and Kolkata

Permissions

GOVERNMENT OF NCT OF DELHI
TRANSPORT DEPARTMENT
(CLUSTER & DTC SECRETARIAT)
5/9, UNDER HILL ROAD, DELHI - 110 054

No. F.10/STA/Policy /Tpt/ 2011/333/40631

Date: 17/08/2020

To

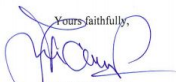
The CEO,
Delhi Integrated Multi Modal Transit System Ltd.,
8th Floor, Block-1, Delhi Technology Park,
Shastry Park, Delhi-110053.

Subject: Request for permission to install pollution sensing units in Cluster buses as a part of R&D Project by IIT, Delhi.

Sir,

Kindly refer to your letter no. DIMTS/Road Transport/2019/4398, dated 05.11.2019, on the abovementioned subject. DIMTS had requested for a formal approval to install pollution sensing units in 10 Cluster buses by CSE IIT, Delhi.

In this context, I am directed to convey the approval of Hon'ble Minister (Transport) for installing of pollution sensing units in 10 Cluster buses of the Kushak Nalah Depot by CSE, IIT Delhi.

Yours faithfully,

(Subodh Kumar)
Deputy Commissioner
(Cluster & DTC Sectt.)



Delhi Pollution Control Committee
5th Floor, I.S.B.T. Building Complex Kashmir Gate Delhi 110006
Visit us at :<http://dpcc.delhi.govt.in>

F. No. DPCC-1(12)(1)(260)Lab(R) 2020/2203

Date: 27/1/2020


To,

Dr. Rjurekha Sen
Department of Computer Science,
IIT Delhi, Hauz Khaz,
New Delhi-110016

Subject- Support Letter for Vehicle Mounted Low Cost PM Monitoring in Delhi

Madam,

With reference to your E-mail and telephonic discussion this organization is interested to know feasibility of Vehicle Mounted Low Cost PM Monitoring in Delhi and willing to share data generated by DPCC Ambient Air Quality Network to assess error percentage of Low Cost System.


(Dr. M. P. George)
Scientist



WEST BENGAL POLLUTION CONTROL BOARD
(Department of Environment, Government of West Bengal)
Paribesh Bhawan, I.A. Block - 2A, Sector III, Bidhannagar
Kolkata - 700 106, Ph: (033)23235-3913, Fax: (033) 2335-2813
website: www.wbpcb.gov.in, e-mail: net.wbpcb-wb@bangla.gov.in

Memo No. 2204/1K-5/2020(Pr-1)

Date: 25.11.2022

To,

Prof. Sagnik Dey
Institute Chair Professor
Centre for Atmospheric Sciences
Indian Institute of Technology (IIT), Delhi
Hauz Khan, New Delhi 110 016.

Sub : Support letter for the project 'Local Air Quality Management Plan'(LAMP)

Dear Sir,

I am directed to state that in response to your email about the support IIT-D is seeking for the deployment of portable sensors in Kolkata, West Bengal and executing the project LAMP in partnership with and support from Environmental Defence Fund (EDF).

The project objectives of carrying out hyperlocal measurement, in stationary and mobile mode and isolating local source characteristics from regional transport would provide critical inputs for research purposes to the Board's ongoing efforts in air quality management in Kolkata. This pilot project, if successful, shall be considered for scaling up for other non-attainment cities.

By keeping in view of the project objectives and deliverables in mind,

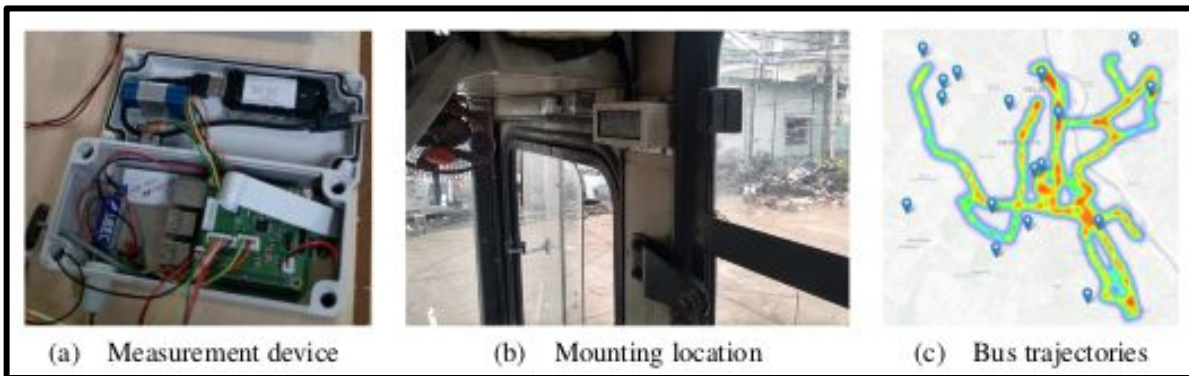
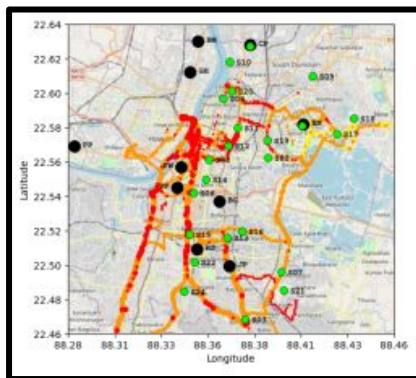
- 1) We allow to carry the proposed study related to air quality monitoring in and around monuments, in Kolkata to identify dispersed sources and track air quality at a hyperlocal scale.
- 2) It is also intimated that the CAAQMS maintained by WBPCB can be used for collocation of the portable sensors during the study to ensure data robustness from the portable sensors.
- 3) The bus routes for the deployment of mobile sensors may be finalized based on the discussion with the State Board
- 4) The study report shall be submitted to WBPCB after completion of the Pilot Project. The R & D section of the WBPCB shall examine the report for further necessary action from the end of the Board.
- 5) The data can be used by the Agency only after obtaining exclusive permission of the WBPCB in this regard.

Deployment



Particulate Matter dataset collected using low cost platforms mounted on public buses and schools/colleges in Delhi and Kolkata

(with respective Pollution Control Boards and Transport Ministries)

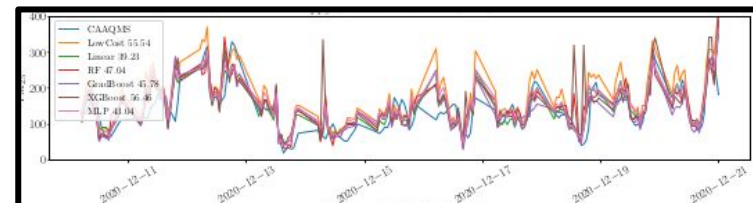
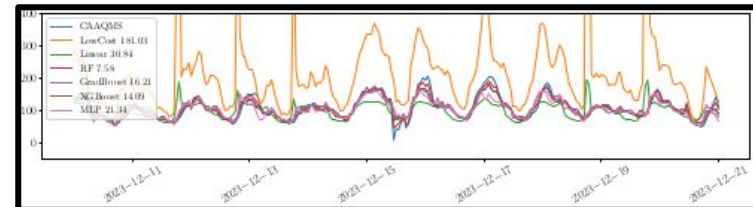


(a) Measurement device

(b) Mounting location

(c) Bus trajectories

Metric	Kolkata, India	Delhi-NCR, India
Total area	160 km ²	559 km ²
Total samples	104,447	12,542,183
PM _{2.5} Samples	104,447	12,542,183
Pollutants covered	PM _{2.5}	PM ₁ , PM _{2.5} , PM ₁₀
Meteorological	Temp, RH	Temp, RH, Pres, WS
Sensor source	College/Public bus	Public bus
Monitoring days	142	91
Calibration done	✓	× (Optionally)
PM _{2.5}	75.82 ± 35.49	207.92 ± 114.36
Missing Data	0	0



Pollution Exposure and Health Monitoring of Workers



[Home](#)

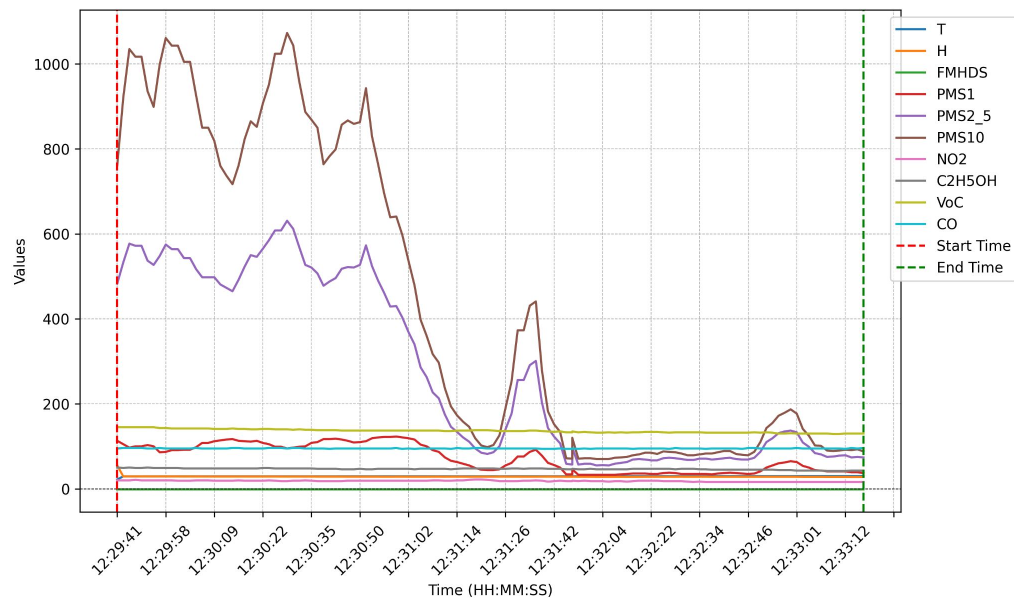
[About](#) ▾

[Our Work](#) ▾

CARE TO EACH OTHER

We Can Help Other People's Lives Sincerely

The Hazards Centre is a unit of the Sanchal Foundation that was set up by the Foundation in 1997 for the specific purpose of providing professional services to community and labour organisations. It consists of a multi-disciplinary team which has, for the past ten years, been assisting such organisations in identifying, understanding, and combating the “hazards” that beset them. The Centre is largely not pro-active, but responds to requests. At times, where several communities are affected by the same set of factors, the Centre may also undertake to do research on a common theme in order to disseminate the findings to a larger range of social groups.



Creating Datasets to Study Urban Sustainability Problems in Developing Regions

Sensible sensing

Privacy and security

Edge computing

Creating Datasets to Study Urban Sustainability Problems in Developing Regions

Sensible sensing

Where to sense?

When to sense?

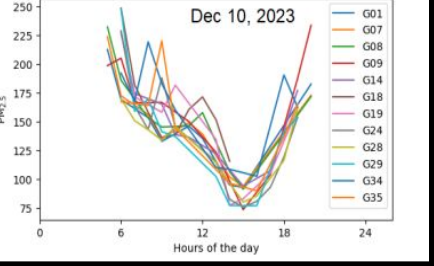
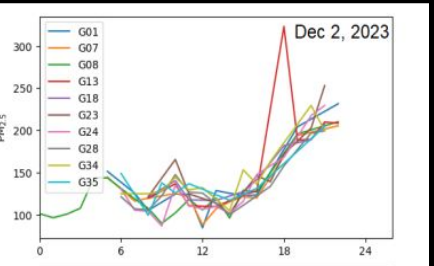
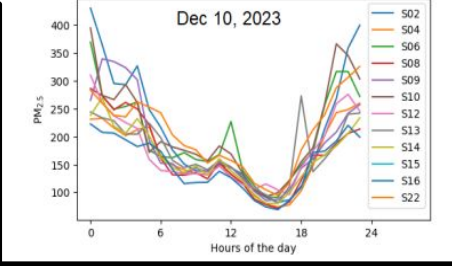
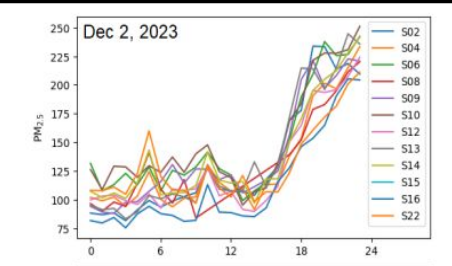
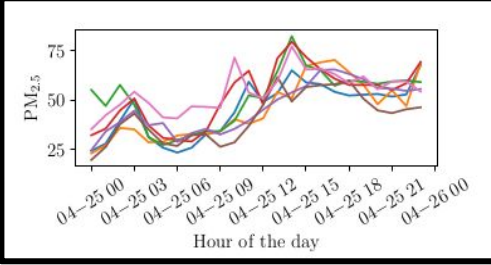
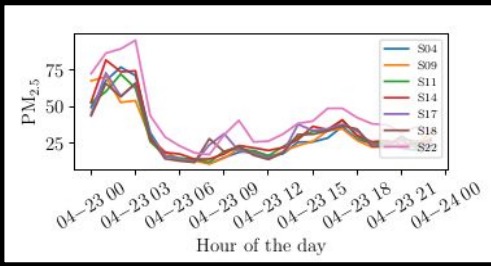
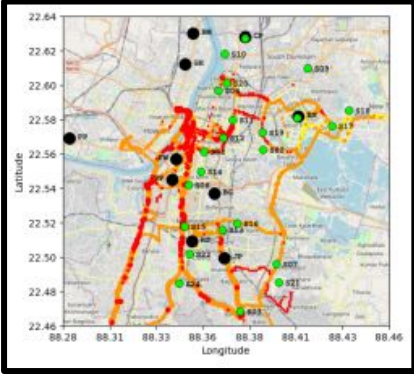
What to sense?

Privacy and security

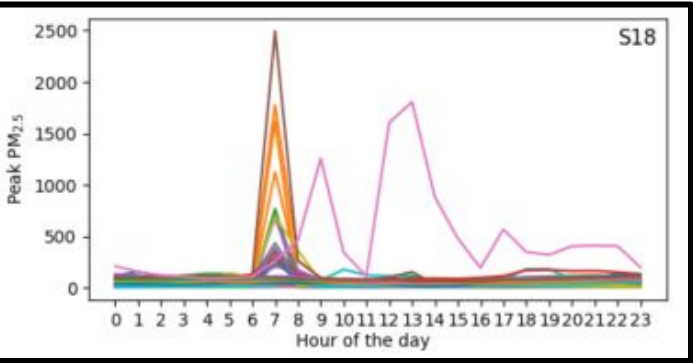
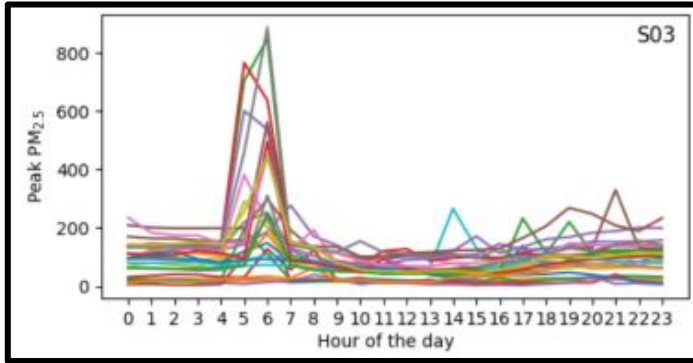
Edge computing

Where to sense? Motivation

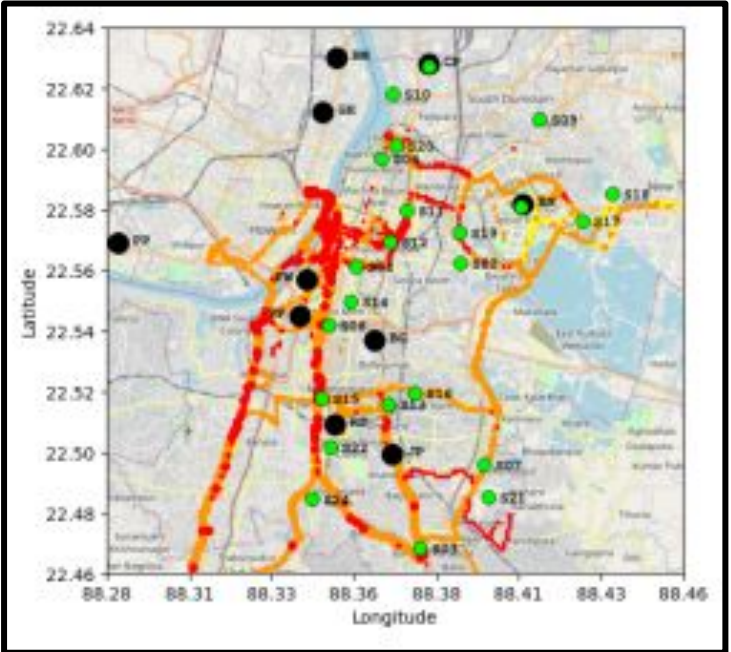
Regional patterns



Hyperlocal patterns



Where to sense? Motivation



(a) Spring

	BG	FW	RS	JP	BM	GH	PP
	13.36	15.37	19.46	8.95	17.03	16.03	16.02
BG	1.83	15.44	↑ 25.03	↑ 12.29	15.03	14.62	16.13
FW	↑ 17.04	1.03	13.86	10.58	↑ 21.38	↑ 20.38	16.04
RS	13.29	15.37	0.83	↑ 20.09	17.94	16.72	16.05
JP	11.79	15.34	19.97	1.95	17.95	16.96	15.95
BM	13.15	15.33	17.98	↑ 11.21	1.36	12.82	15.88
GH	11.74	15.41	20.45	9.22	15.40	1.67	16.05
PP	12.44	15.54	19.60	8.59	↑ 20.77	↑ 19.57	5.24

(b) Autumn

	BG	FW	RS	JP	BM	GH	PP
	9.81	17.27	9.86	7.91	12.46	25.56	12.23
BG	1.97	17.36	↑ 12.38	9.35	12.65	25.73	12.18
FW	↑ 12.35	1.47	↑ 14.73	↑ 11.66	12.84	26.08	15.41
RS	10.10	17.29	1.10	↑ 10.42	12.37	25.46	12.48
JP	9.05	17.34	9.91	1.36	12.58	25.62	12.47
BM	9.71	17.23	9.61	7.66	1.33	26.35	12.23
GH	9.13	17.27	9.65	7.61	12.59	2.40	12.18
PP	10.00	17.27	9.89	8.12	12.34	25.49	1.03

(c) Winter

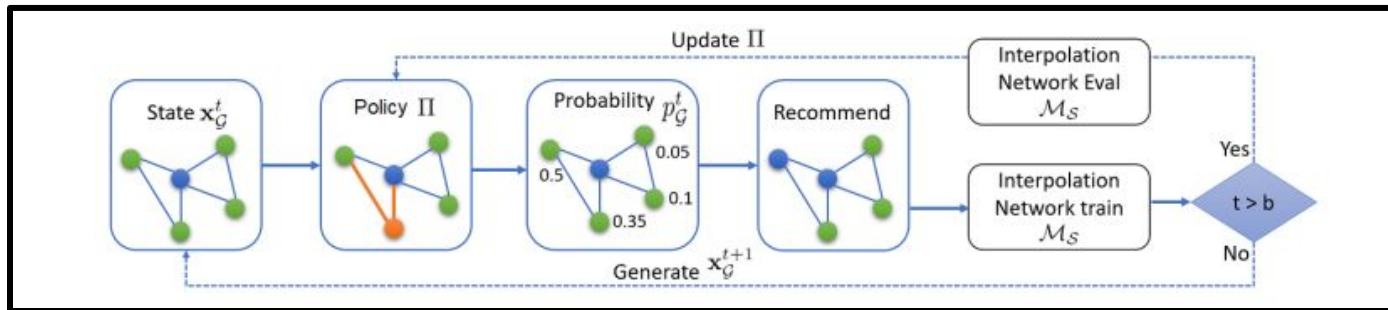
	BG	FW	RS	JP	BM	GH	PP
	19.59	33.21	22.15	19.16	16.68	62.05	18.86
BG	4.13	33.08	21.19	18.20	17.93	64.82	19.19
FW	21.34	2.83	23.54	21.92	20.08	65.45	↑ 26.33
RS	↑ 23.91	33.19	2.69	19.83	↑ 25.36	↑ 74.68	22.15
JP	14.86	33.23	17.96	2.68	16.59	59.91	18.44
BM	18.42	33.20	15.81	20.48	3.56	63.76	19.93
GH	↑ 31.50	33.22	↑ 45.62	↑ 23.61	↑ 57.74	6.64	22.07
PP	16.94	33.21	22.14	21.56	25.41	↑ 73.89	3.14

Where to sense? Location recommendation problem formulation

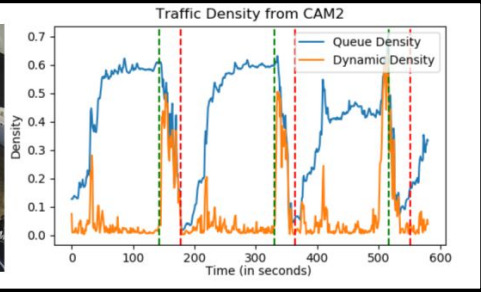
Problem 1 (Budget-constrained recommendation). *Training:* Given pollution data from static sensors \mathcal{S} and vehicle sensor data \mathcal{Q} , learn an inductive active-learning model Π such that the below condition holds in inference.

Inference: Given pollution data from a set of unseen sensors \mathcal{S}_2 , candidate locations $\mathcal{C} = \{\ell_1, \dots, \ell_n\}$ where PM monitoring is desired and therefore static sensors may be installed, and a budget b where $b \ll |\mathcal{C}|$, use Π to predict b locations from \mathcal{C} for installation of new sensors such that the marginal decrease in expected loss is maximized. Formally,

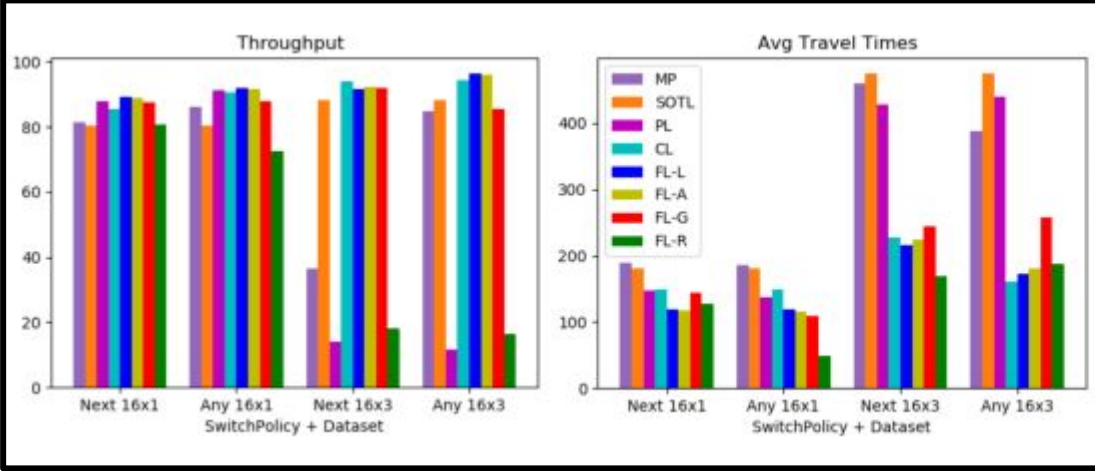
$$\begin{aligned}\Pi(\mathcal{M}, \mathcal{S}, \mathcal{C}, \mathcal{Q}, b) &= \arg \max_{\mathcal{A} \subseteq \mathcal{C}, |\mathcal{A}|=b} \{ \mathcal{L}(\mathcal{M}_{\mathcal{S}_1}) - \mathcal{L}(\mathcal{M}_{\mathcal{S}_1 \cup \mathcal{A}}) \} \\ &= \arg \min_{\mathcal{A} \subseteq \mathcal{C}, |\mathcal{A}|=b} \{ \mathcal{L}(\mathcal{M}_{\mathcal{S}_1 \cup \mathcal{A}}) \}\end{aligned}$$



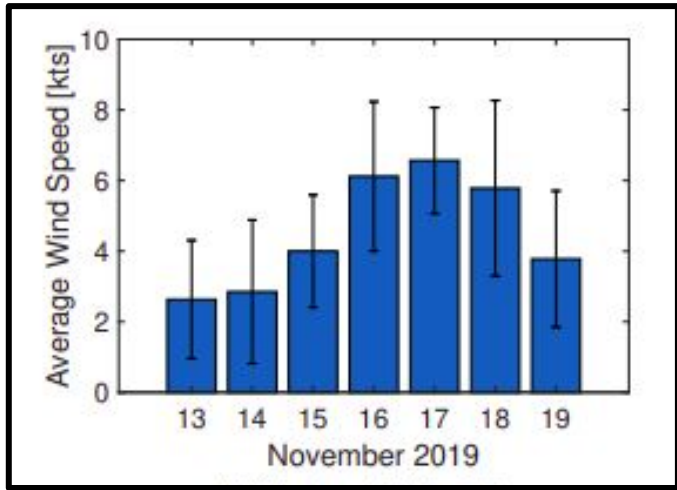
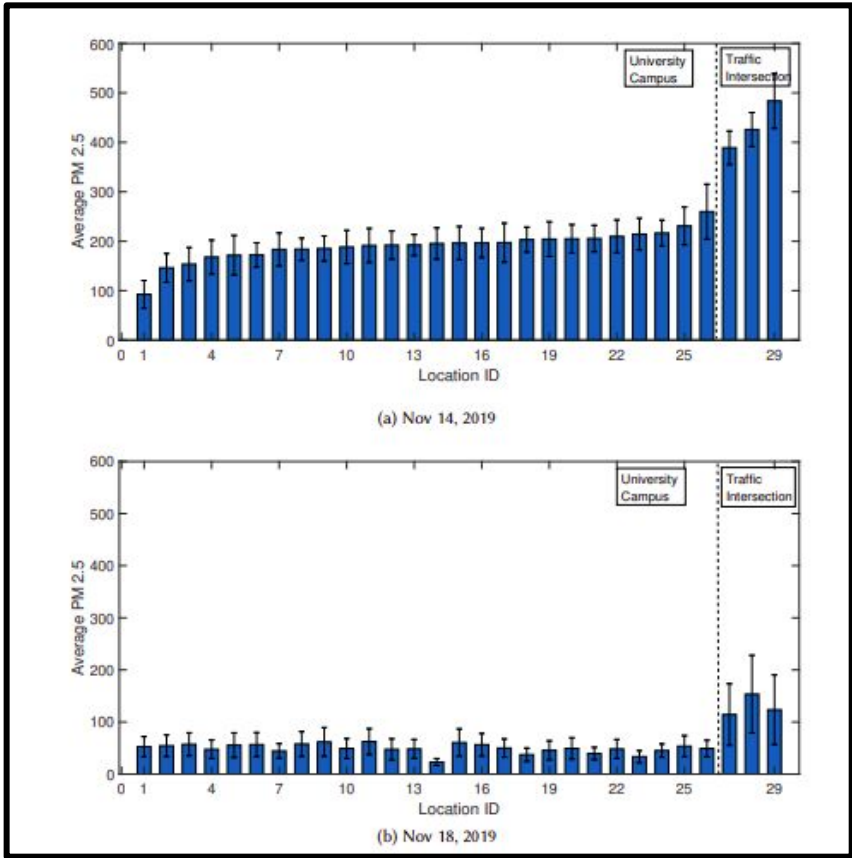
What to sense? When is less data useful?



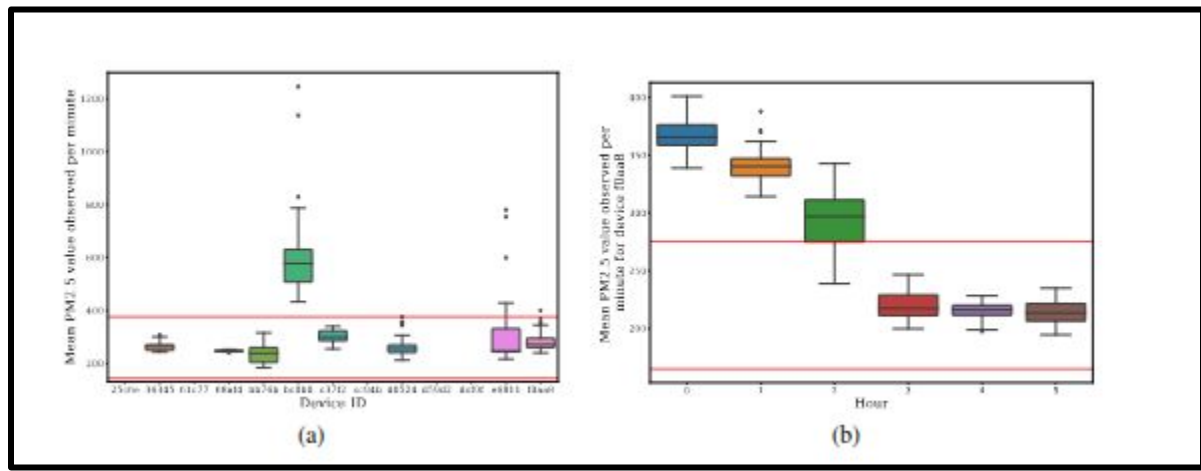
Model	RL State Size	DNN Arch	DNN Params
Presslight	80	-	2082-2124
Colight	1600-12480	-	6018-6084
FL-Lane	12	20x20	722-764
FL-Approach	4	15x15	347-379
FL-Group	2	10x10	162-184
FL-Relative	1	5x5	52-64



What to sense? When is more data useful?



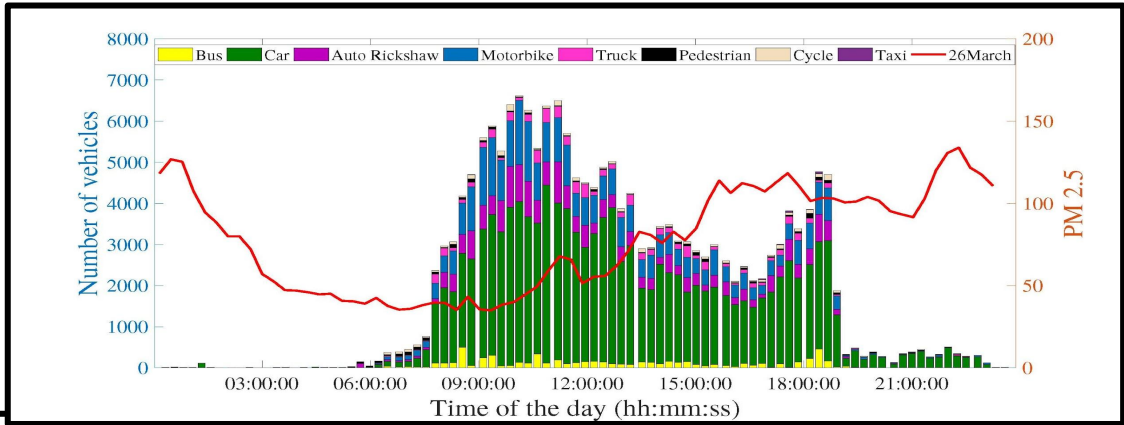
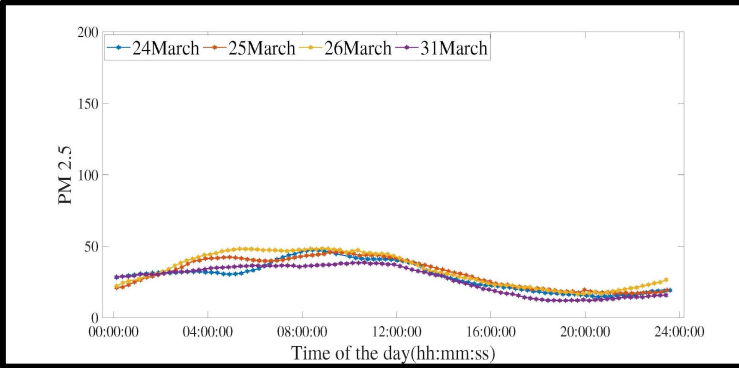
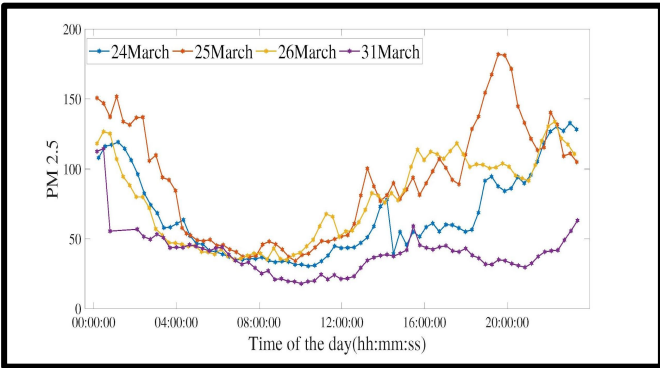
When to sense? Low cost sensor anomalies detected at bus depots.



When to sense? Delayed correlations of air pollution with road traffic level



Deployment of our instrument with road traffic camera in Noida



PM 2.5 does not directly peak at vehicle count peaks, but shows a delayed effect

Creating Datasets to Study Urban Sustainability Problems in Developing Regions

Sensible sensing

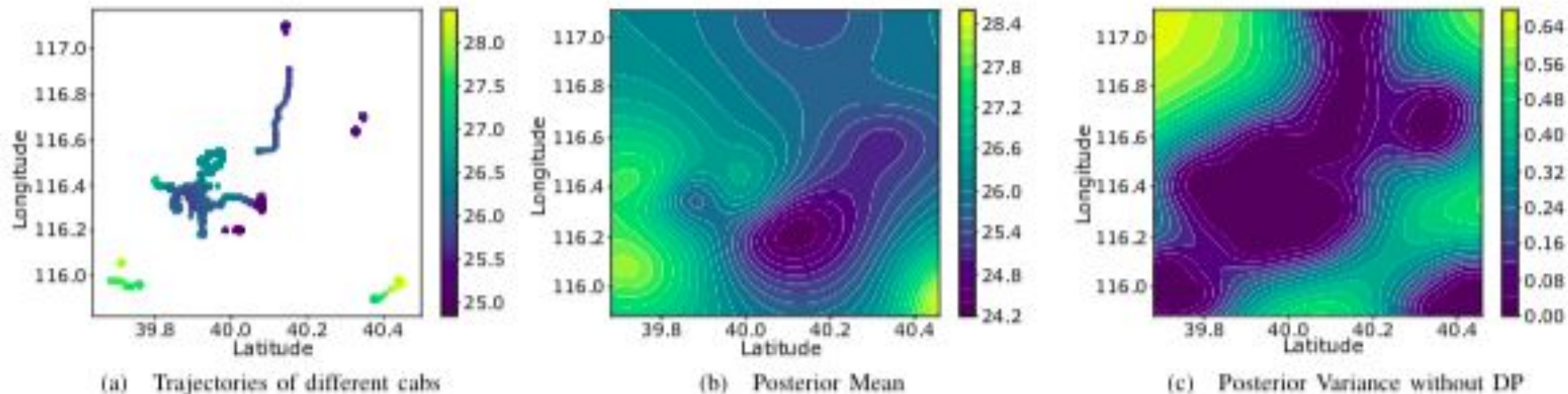
Privacy and security

Edge computing

Is there any privacy concern for the data? How to handle these concerns?

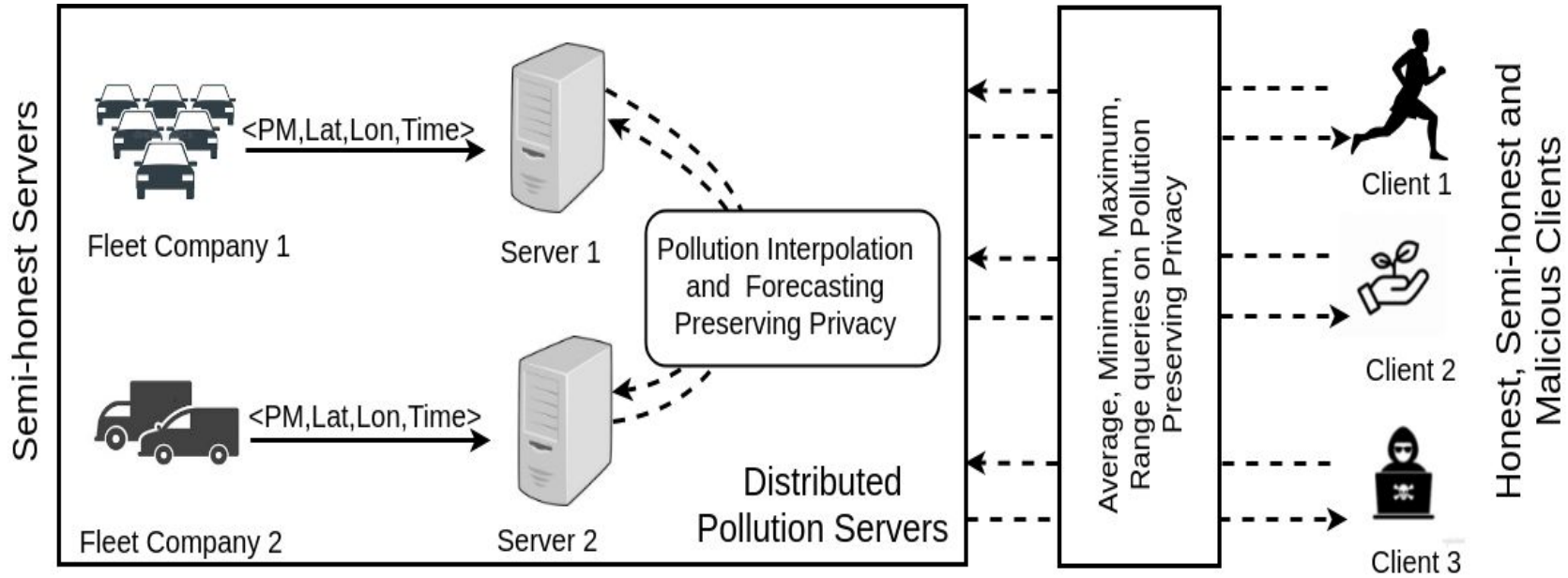
How to ensure integrity of the sensors and the edge computing platforms?

Organically arising privacy question: concerns of private fleet companies

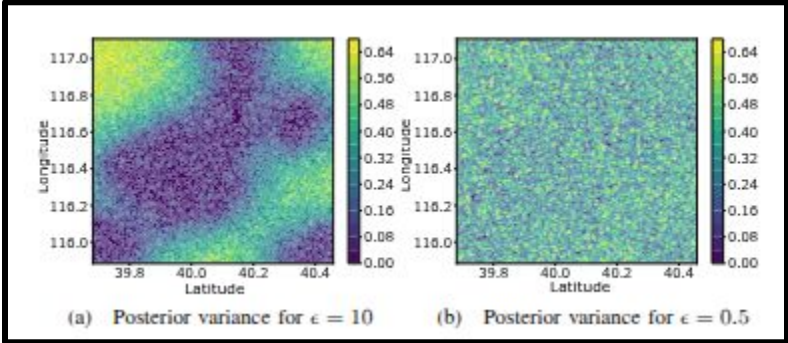


Anecdotal evidence after DIMTS deployment: OLA doesn't want Uber to know its location traces. How to aggregate data across fleet companies to train more accurate spatio-temporal air pollution models for interpolation and forecasting, preserving a private fleet's location privacy?

Privacy Preserving Machine Learning (PPML) optimized for our applications

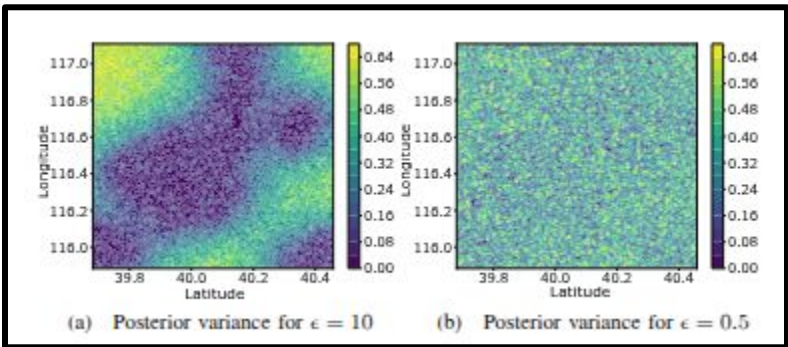


Examples of application specific research questions for PPML



Can differential privacy balance the utility-vs-privacy trade-offs?

Examples of application specific research questions for PPML



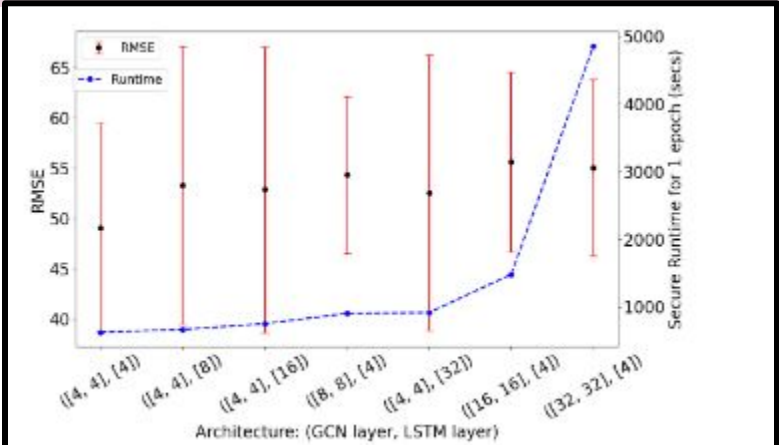
Can differential privacy balance the utility-vs-privacy trade-offs?

Query Function (f)	Function Formula	Propagated Error
Average	$\frac{\sum_{i=1}^n y_i}{\sum_{i=1}^n v_i}$	$\frac{\sqrt{\sigma_1^2 + \sigma_2^2 \dots \sigma_n^2}}{\sum_{i=1}^n v_i}$
Max or Min	$y^* + \frac{1}{\rho} \log(\sum_{i=1}^n \exp(\rho(y_i - y^*)))$	$\sqrt{\sum_{j=1}^n \left(\frac{\exp(\rho(y_j - y^*))}{\sum_{i=1}^n \exp(\rho(y_i - y^*))} \sigma_j \right)^2}$ (Max $\rho \rightarrow \infty$, Min $\rho \rightarrow -\infty$)
Count	$\sum_{i=1}^n S(y_i, \theta)$	$\sqrt{\sum_{j=1}^n (S(y_j, \theta)(1 - S(y_j, \theta)) \sigma_j)^2}$
Range	$\sum_{i=1}^n (S(y_i, \theta_1) - S(y_i, \theta_2)), \theta_1 < \theta_2$	$\sqrt{\sum_{j=1}^n (S(y_j, \theta_1)(1 - S(y_j, \theta_1)) - S(y_j, \theta_2)(1 - S(y_j, \theta_2)))^2 \sigma_j^2}$

Query function and error propagation formulae, implemented in the Garbled Circuit where $S(y_i, \theta) = \frac{1}{1 + \exp(-(y_i - \theta))}$

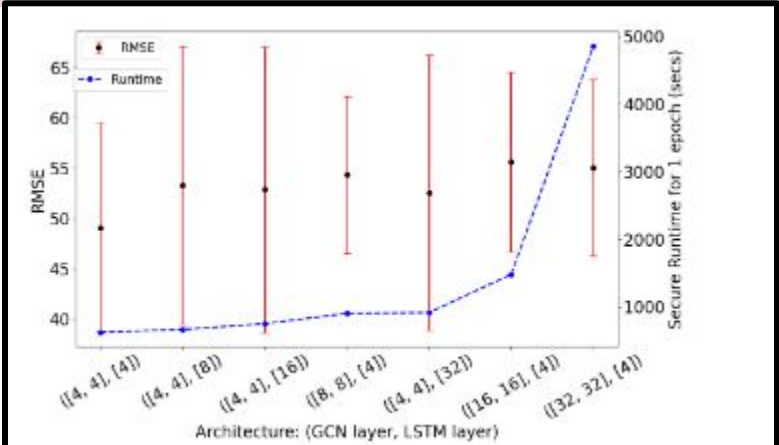
What mathematical functions need to be computed for our application?

Examples of application specific research questions for PPML

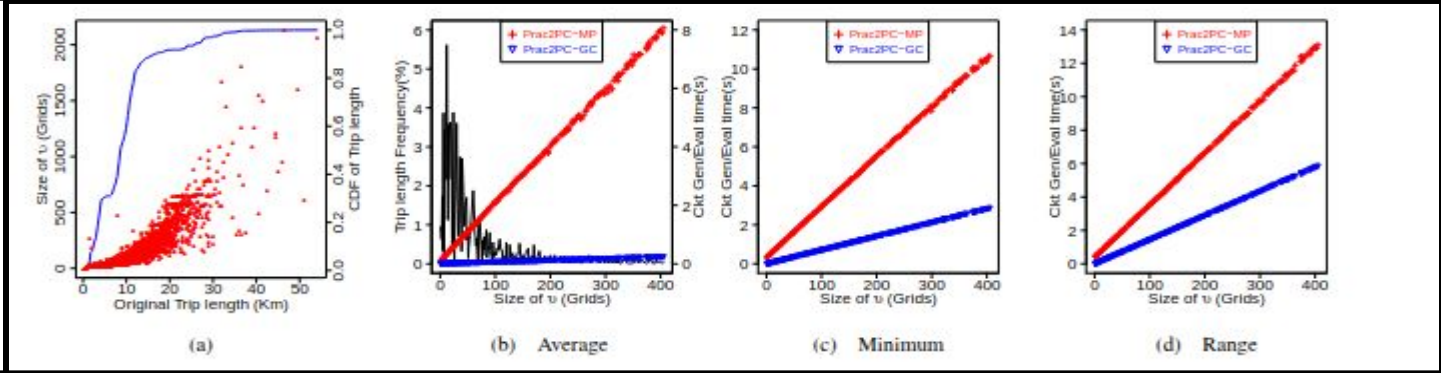


What neural network architectures are better performant for secure MPC?

Examples of application specific research questions for PPML



What neural network architectures are better performant for secure MPC?



What should be the end to end cryptographic pipeline? Which implementations perform better with same security guarantees?

Creating Datasets to Study Urban Sustainability Problems in Developing Regions

Sensible sensing

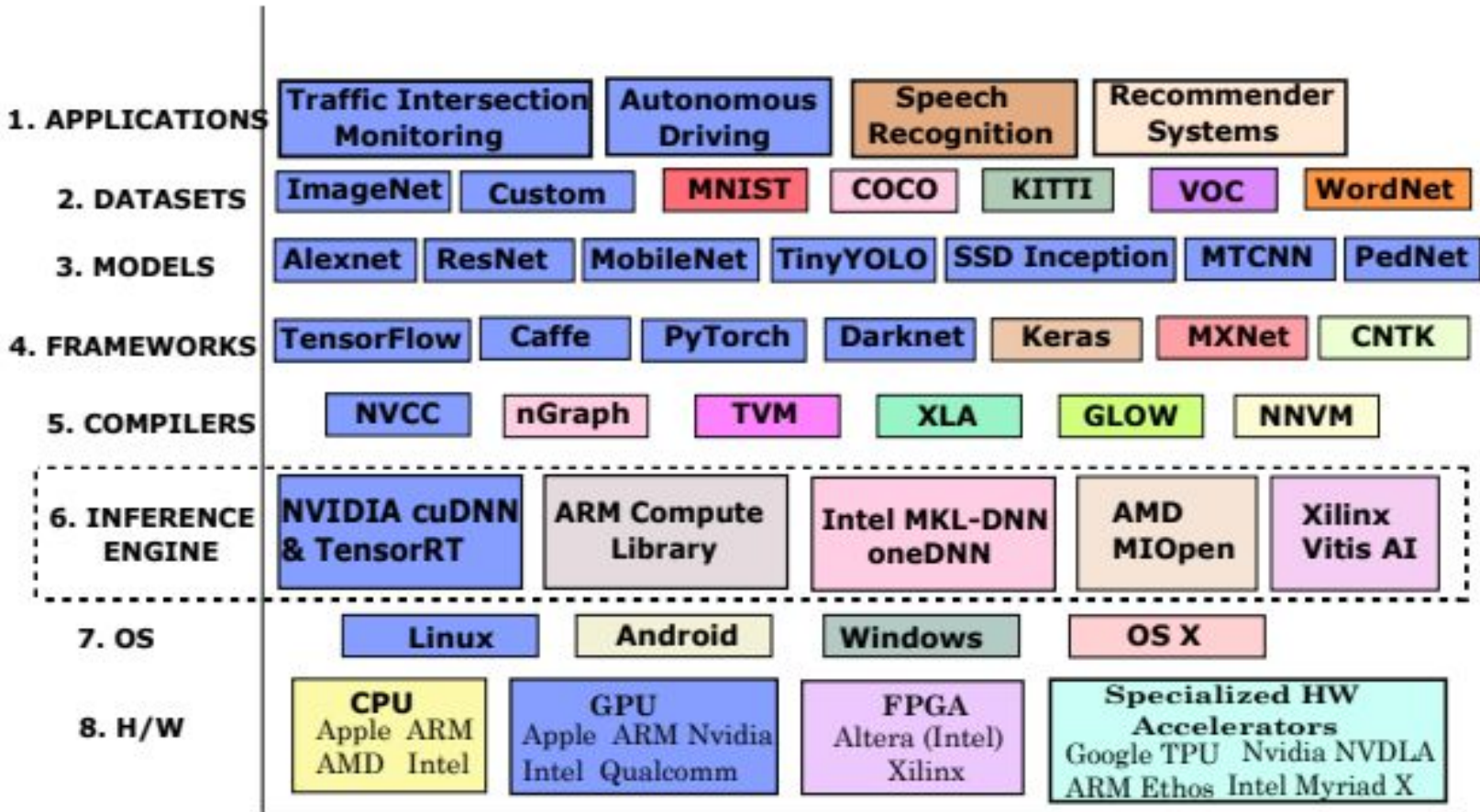
Privacy and security

Edge computing

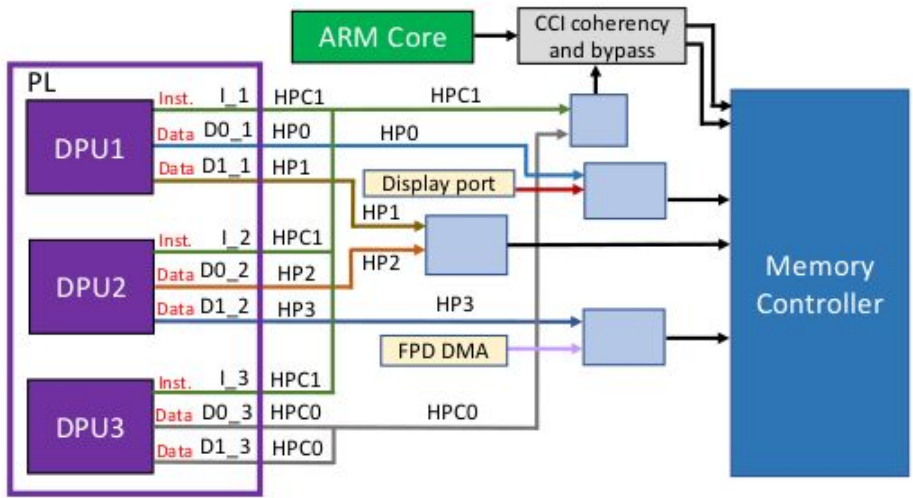
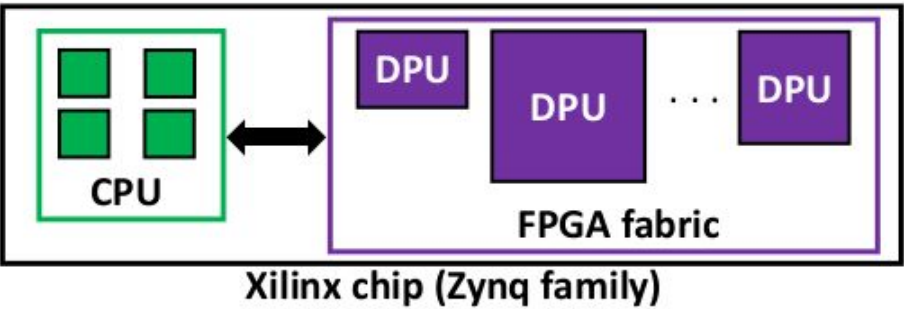
What to compute at edge?

How to balance performance vs. thermal safety?

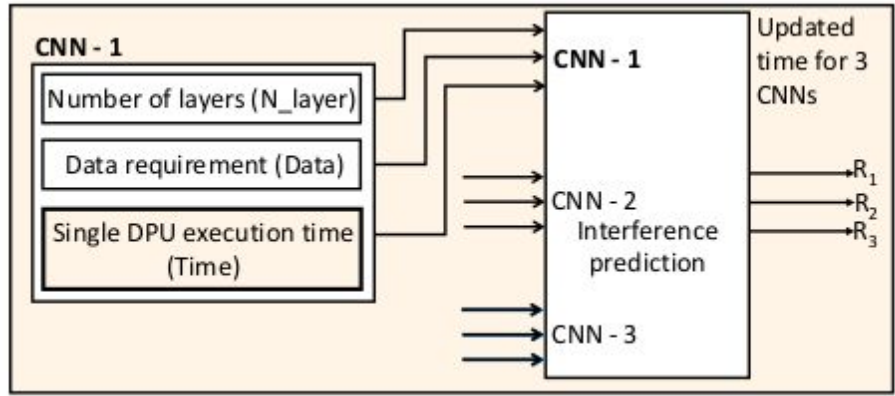
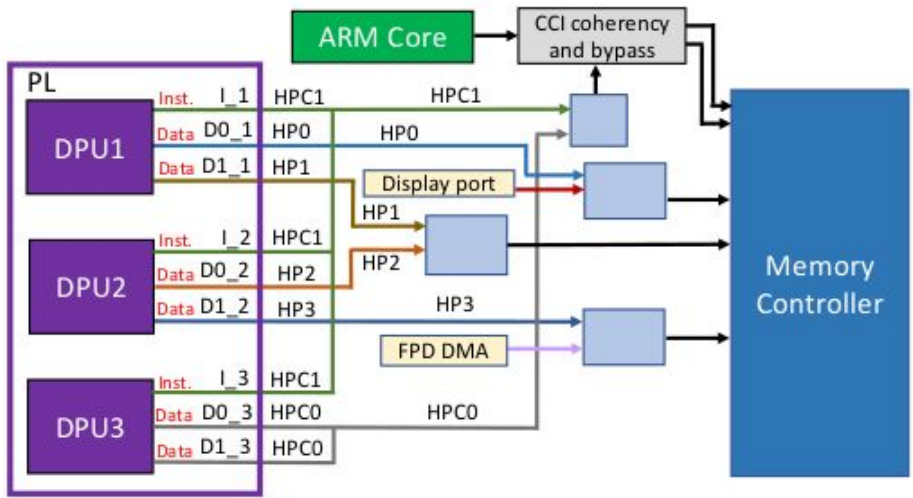
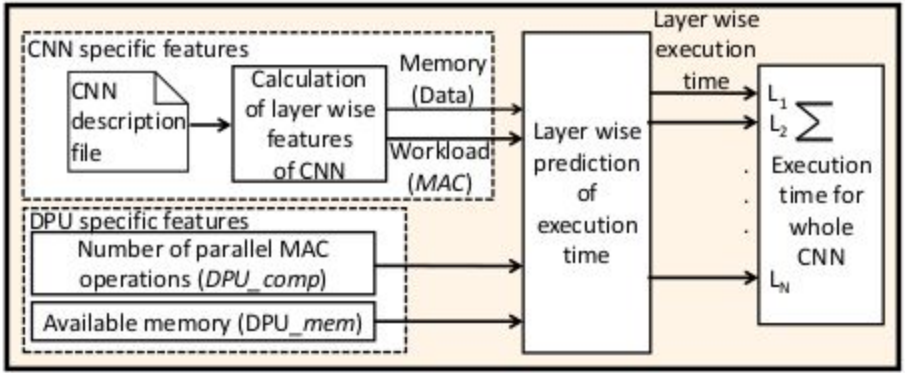
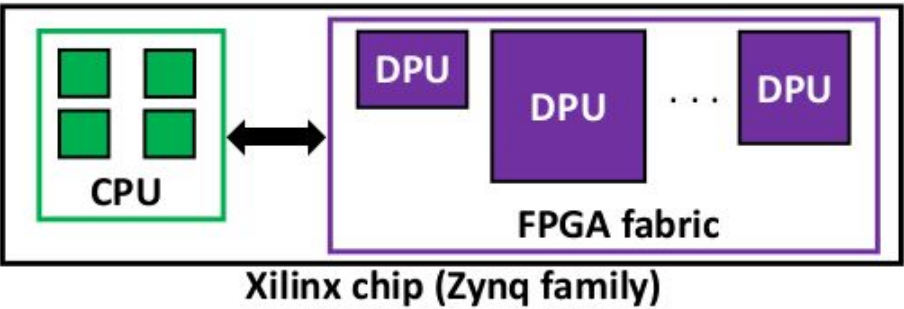
Edge computing hardware and software complexity with AI-ML boom



More transparency and control in XILINX embedded FPGA platforms



More transparency and control in XILINX embedded FPGA platforms



Understanding blackbox edge hardware and software for NN acceleration

Hardware Platform	CPU	GPU	NN-ACC	Memory	Storage	Cost
① Jetson Xavier AGX [35]	2.26GHz 8-core Carmel ARM v8.2	1.37GHz 512-core Volta	2x NVDLA v1	64GB 256-bit LPDDR4x	32GB eMMC	\$999
② Jetson Xavier NX [36]	1.4GHz 6-core Carmel ARM v8.2	1.1GHz 384-core Volta	2x NVDLA v1	16GB 128-bit LPDDR4x	64GB MicroSD	\$479
③ Jetson TX2 [37]	2GHz 2-core Denver, 2GHz 4-core Cortex A57	1.3GHz 256-core Pascal	-	8GB 128-bit LPDDR4	32GB eMMC	\$479
④ Khadas VIM 3 [38]	2.2GHz 4-core Cortex-A73, 1.8 GHz 2-core Cortex-A53	800MHz Mali-G52 MP4	5 TOPS Amlogic NPU	4GB LPDDR4	64GB MicroSD	\$159.90
⑤ Jetson Nano [39]	1.43GHz 4-core Cortex A57	950MHz 128-core Maxwell	-	4GB 64-bit LPDDR4	16GB eMMC	\$129
⑥ Odroid H2 [40]	2.3GHz 4-core Intel Celeron J4105	750MHz Intel UHD Graphics 600	-	16GB DDR4	64GB eMMC	\$119
⑦ Intel NCS2 [41]	-	-	Myriad X VPU	500MB LPDDR	-	\$110
⑧ Odroid M1 [42]	2GHz 4-core Cortex-A55	650MHz Mali-G52 MP2	0.8 TOPS Rockchip NPU	8GB LPDDR4	64GB MicroSD	\$95.50

Counter-intuitive observation 1: GPU with more compute resources is slower

	Xavier NX (GV10B)	Xavier AGX (GV10B)
CPU	6-core NVIDIA Carmel ARM@v8.2 64-bit CPU 6MB L2 + 4MB L3	8-core ARM® v8.2 64-bit CPU 8MB L2 + 4MB L3
# GPU cores	384 (64 per SM)	512 (64 per SM)
# SMs	6	8
# Tensor cores	48 (8 per SM)	64 (8 per SM)
L1 cache	128KB per SM	128KB per SM
L2 cache	512KB	512KB
Memory	8GB 128-bit LPDDR4x 51.2GB/s	32GB 256-bit LPDDR4x 137GB/s
GPU Clock	1.1 GHz	1.137 GHz
Technology	12nm	12nm

NN Model	cNX rNX	cNX rAGX	cAGX rAGX	cAGX rNX
Inception-v4	31.64 (0.33)	46.06(3.03)	42.60 (2.59)	34.56(1.89)
Pednet	29.75 (0.3)	34.55(0.9)	33.12 (0.37)	31.49(0.56)

Inference latencies in msec

Counter-intuitive observation 2: NN compilers are non-deterministic

NN Model	Engine1	Engine2	Engine3
Alexnet	44.47(0.62)	43.72(0.7)	44.86(0.34)
ResNet-18	13.94(1.73)	9.15(1.17)	9.02(0.39)
vgg-16	113.45(4.9)	129.5(5.09)	124.7(3.86)
inception-v4	62.83(4.65)	73.5(9.08)	68.15(2.53)
Googlenet	542.33(60.62)	541.14(57.9)	540.86(58.6)
ssd-inception-v2	41.18(1.90)	40.23(1.65)	42.2(2.2)
Detectnet-Coco-Dog	27.77(2.49)	25.56(2.74)	26.3(2.86)
pednet	37.96(3.69)	37.66(2.27)	37.78(4.17)
facenet	22.67(2.54)	22.54(2.45)	21.96(2.1)
Tiny-Yolov3	484.11(13.07)	484.29(12.96)	486.54(13.43)
Mobilenetv1	10.98(0.16)	13.25(0.35)	12.13(0.67)
MTCNN	861.2(41.2)	862.5(43.5)	861.67(41.9)
fcn-resnet18-cityscapes	29.87(1.36)	34.56(1.19)	35.23(0.85)

Inference latencies in msec

Platform	NN Model	Engines 1 - 2	Engines 2 - 3	Engines 1 - 3
NX	ResNet-18	105	105	0
AGX	vgg-16	269	0	269
AGX	inceptionv4	461	296	497
AGX	ResNet-18	243	224	183

Prediction labels different out of 1000 images

Counter-intuitive observation 3: Generic NN compilers outperform TensorRT

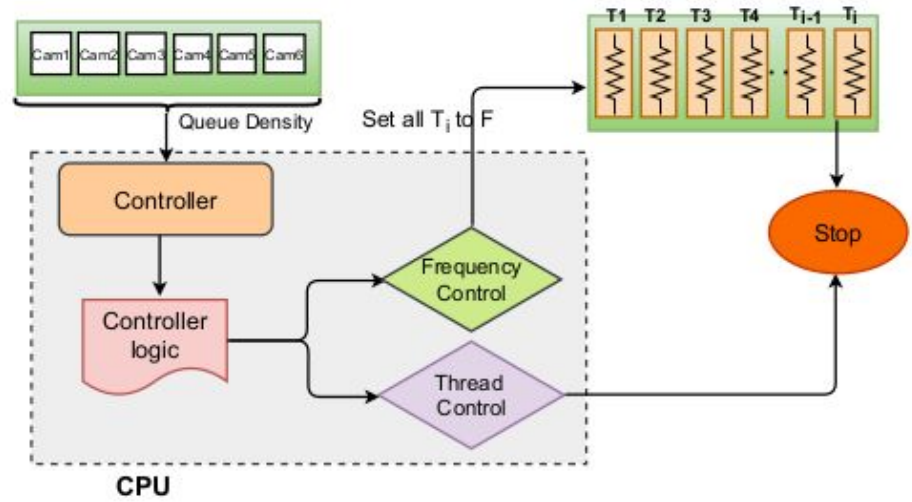
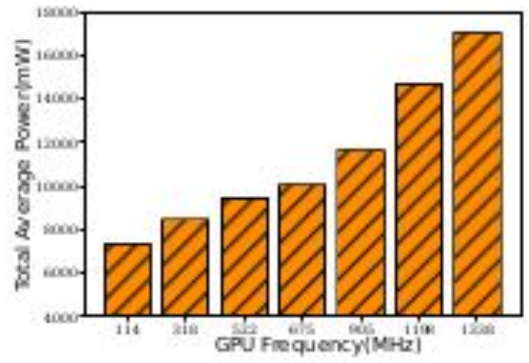
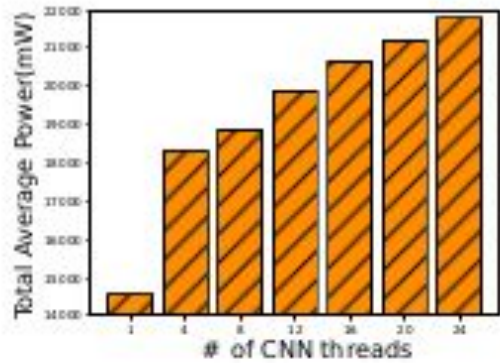
Listing 1 TVM correctness check at each optimization step

```
1 # Unoptimized output
2 out_np = np.maximum(conv_np + bias_np, 0.0)
3 # TVM output
4 out_tvm = tvm.nd.empty(out_np.shape, device=dev)
5 # Check correctness
6 np.testing.assert_allclose(out_np, out_tvm.numpy(), rtol=1e-3)
```

NN Model	TensorRT	AutoTVM	AutoScheduler
Resnet-18	3.93	0.49	0.36
Resnet-50	8.96	0.66	0.60
Squeezenet	4.03	0.56	0.45
Mobilenet_v2	5.65	0.70	0.5
Inception_v3	15.45	1.20	1.10
Vgg-16	8.82	0.49	0.228
Googlenet	16.97	1.18	0.80
Densenet-161	14.65	3.32	2.54
Yolov2	4.97	0.89	0.51
Retinanet	9.13	2.76	2.23

Inference latencies in msecs

Thermal safety on edge GPU platforms



Creating Datasets to Study Urban Sustainability Problems in Developing Regions

Sensible sensing

Where to sense?

When to sense?

What to sense?

Privacy and security

Is there any privacy concern for the data? How to handle these concerns?

How to ensure integrity of the sensors and the edge computing platforms?

Edge computing

What to compute at edge?

How to balance performance vs. thermal safety?

CSE PhD Research

Sensible sensing

Privacy and security

Edge computing

Graduated PhDs

1. Sachin Chauhan
Postdoc in Heidelberg
University, Geoscience Dept.,
funded by Red Cross

2. Ismi Abidi
Edgeless Systems, Confidential
Computing, Bochum, Germany
3. Mehreen Jabbeen
CPU software team lead Qualcomm

4. Shikha Goel
Performance analyst, NVIDIA
5. Omais Shafi
ARM Research, Cambridge

CSE PhD Research

Sensible sensing

Privacy and security

Edge computing

Graduated PhDs

1. Sachin Chauhan
Postdoc in Heidelberg
University, Geoscience Dept.,
funded by Red Cross

2. Ismi Abidi
Edgeless Systems, Confidential
Computing, Bochum, Germany
3. Mehreen Jabbeen
CPU software team lead Qualcomm

4. Shikha Goel
Performance analyst, NVIDIA
5. Omais Shafi
ARM Research, Cambridge

Ongoing research

System software for
edge and mobile

Exploit the reducing heterogeneity in the AI-ML compute stack with the growth of foundation models. Started trying these for sustainability applications - like air pollution forecasting with time-series LLMs. Will try time-series LLM based anomaly detection on personal devices, for improved data privacy.

Running foundation model inferences on edge and mobile under thermal and power constraints. Keep the system software stack as similar to cloud, as possible.

Ismi and Mehreen have developed cryptographic primitives for multi-party computations (privacy) and zero-knowledge proofs (integrity), and optimized these pipelines on edge CPUs. Currently, examining the acceleration of these cryptographic pipelines with edge accelerators like Nvidia edge GPUs and Xilinx edge FPGAs, for better hardware utilization and security.

Non-cryptographic secure and private systems at edge, to reduce computational overhead of cryptography, using ARM Confidential Compute Architecture (CCA). Comparison of security primitives with Intel TDY, AMD SEV-SNP etc. available for cloud systems.