Development and Validation of a Testbed for AI/ML QoS Prediction Algorithm Evaluation

Nick Turay
Maciej Muehleisen
Alex Palaios

Motivation

- Law by the Bundestag: “Act on Autonomous Driving”, *2021* [1]
  – Handles remote-supervision (incl. on-demand tele-operation)

- Caveat
  – Ensuring a communication link with sufficient Quality of Service (QoS)

- Predictive Quality of Service
  – Notification of upcoming changes to QoS
  – Prediction by (sim./emu. aided) Machine Learning (ML) models

[1] German Civil Code, “Amendment to the Road Traffic Act and the Obligation Insurance Act - Autonomous Driving Act," 2021-07

[2] 5GCroCo EU YouTube, https://www.youtube.com/watch?v=1v1i2vp8cEQ @ 10 s ,accessd 2022-08-08

Objective

- ***To what extent does uncertainty in input simulation/emulation models impact the Machine Learning prediction performance?***

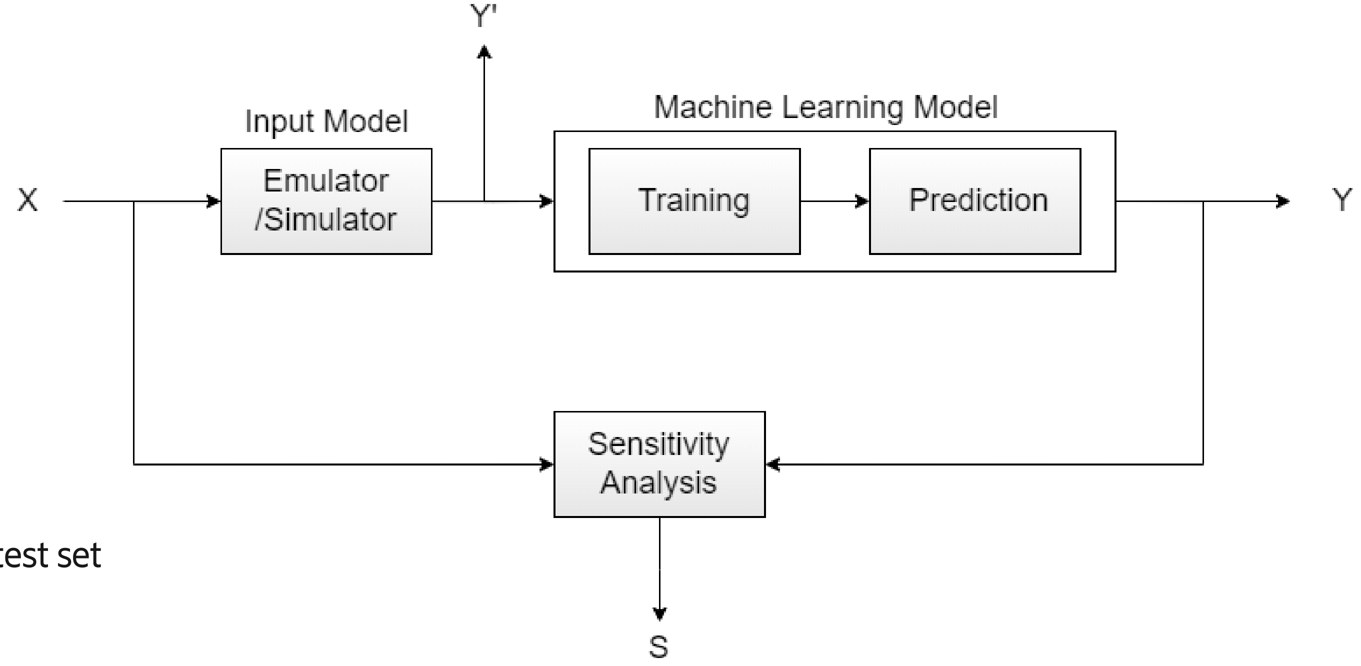
- Sensitivity analysis methodology in this work inspired by [3]:

1. *Definition of the system model*:
    - ➢ Input model and ML model
    - ➢ Input *X* – Emulator/ Simulator paramters
    - ➢ Output *Y*– Evaluated ML model predictions

2. *Generation of the input Y' for the ML model*:
    - ➢ Emulator/Simulator

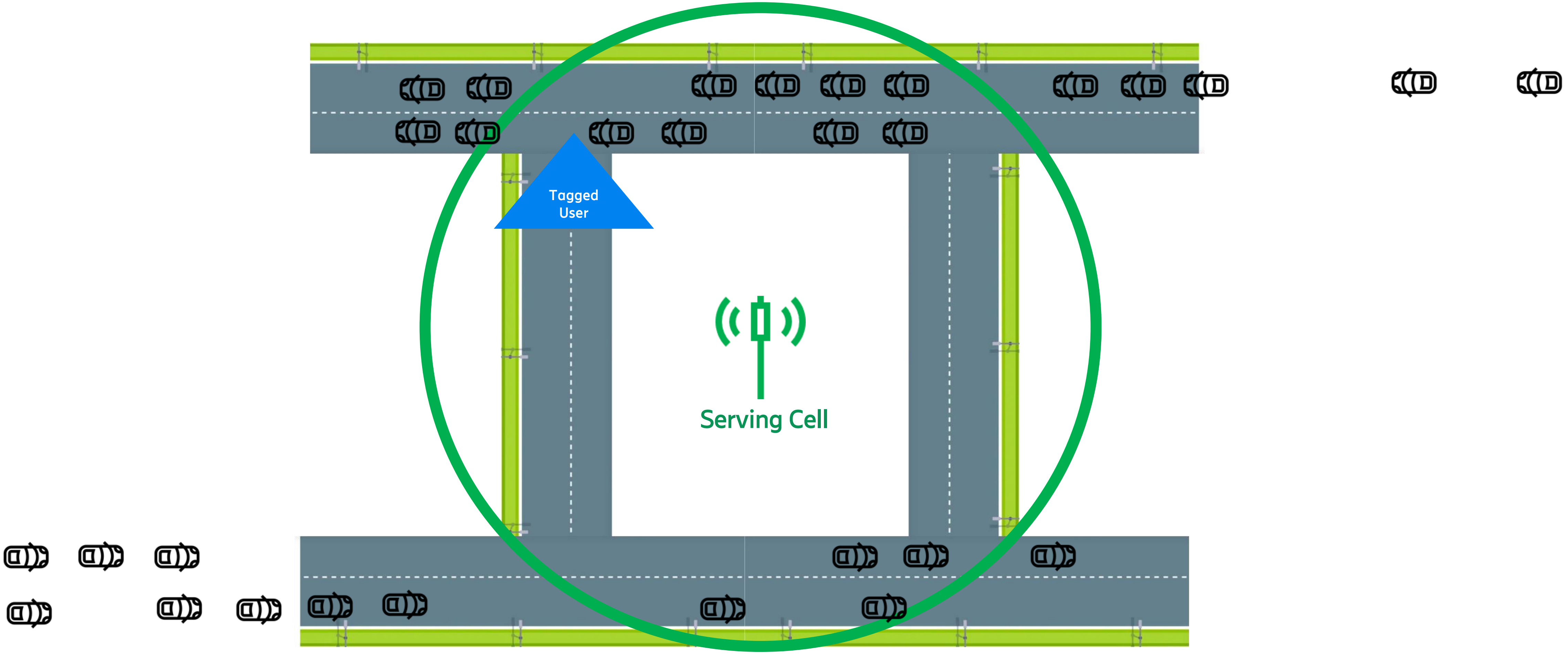
3. Model training on the training set and prediction on the test set

4. *Evaluation of the model sensitivity S, i.e., variability of Y due to change in X*
    - ➢ A variety of methods available, e.g., regression analysis [4]
    - ➢ In this work: The *Sensitivity Index* proposed by Hoffman and Gardner [5]

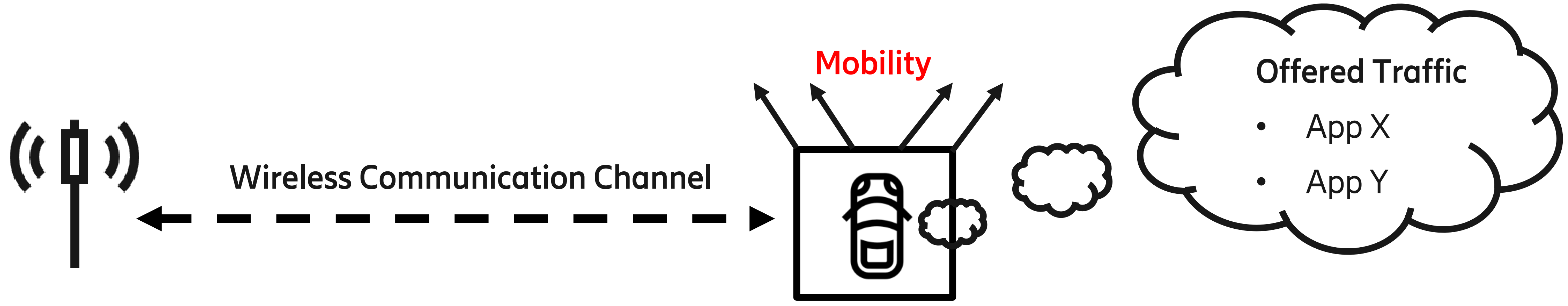


[3] R. L. Iman and J. C. Helton, "Comparison of uncertainty and sensitivity analysis techniques for computer models," 1985; [5] F. Hoffman and R. Gardner, "Evaluation of uncertainties in environmental radiological assessment models," 1983
[4] A. Saltelli, et al., "Sensitivity analysis in practice: a guide to assessing scientific models," 2004;

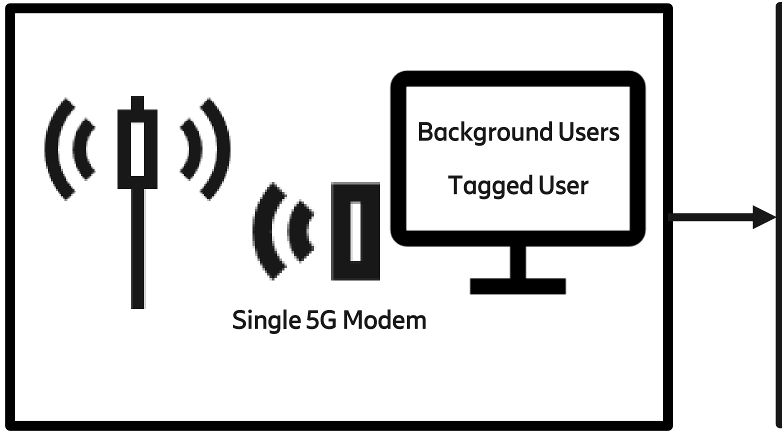
Story



Design Consideration



Small Indoor Lab Environment



Abstraction

Wireless Communication Channel: Max. MCS* always used

Offered Traffic: Fixed packet size

- Background Users: Poisson packet arrival
- Tagged Users: Fixed packet arrival period

Mobility: Number of users served by cell

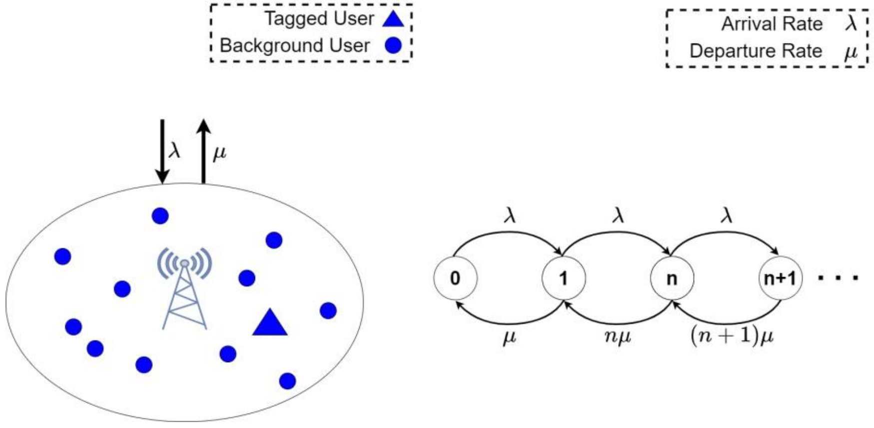
*Modulation Coding Scheme (MCS): Constant SNR of 30 dBm (no inter-cell or intra-cell interference)

Network Traffic - User Mobility

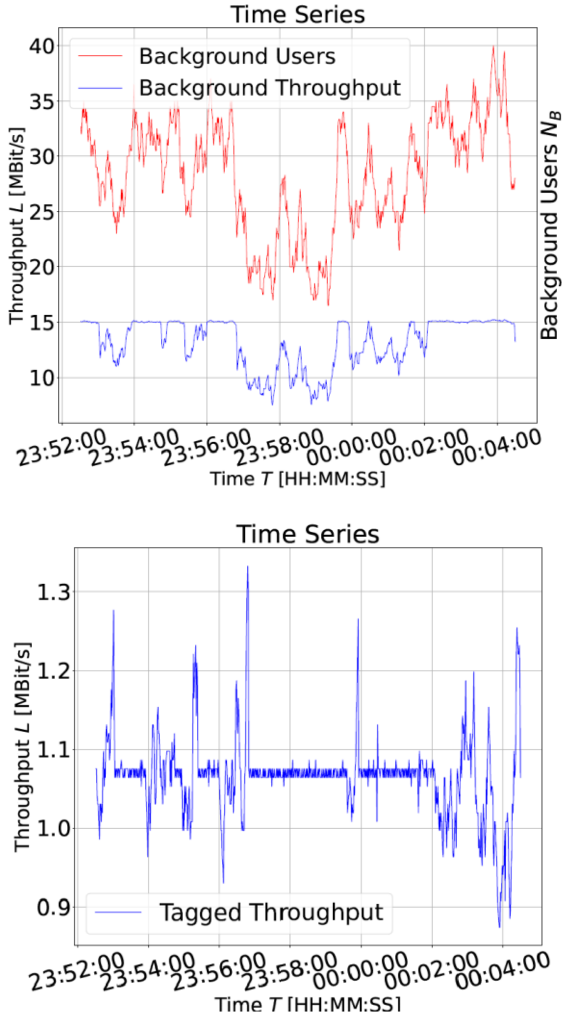
Stochastic Process: Birth-death Markov chain

- Experiment Exp $\triangleq \lambda, \mu$ modeled by $M/M/\infty$
- Sensitivity analysis: Variation of experiment Exp

Mobility Model: Number of users served by the cell



Model Input: User number and produced throughput



Machine Learning Pipeline

- Task: Binary classification *Will the tagged user throughput be above a certain (1 Mbit/s) threshold ?*

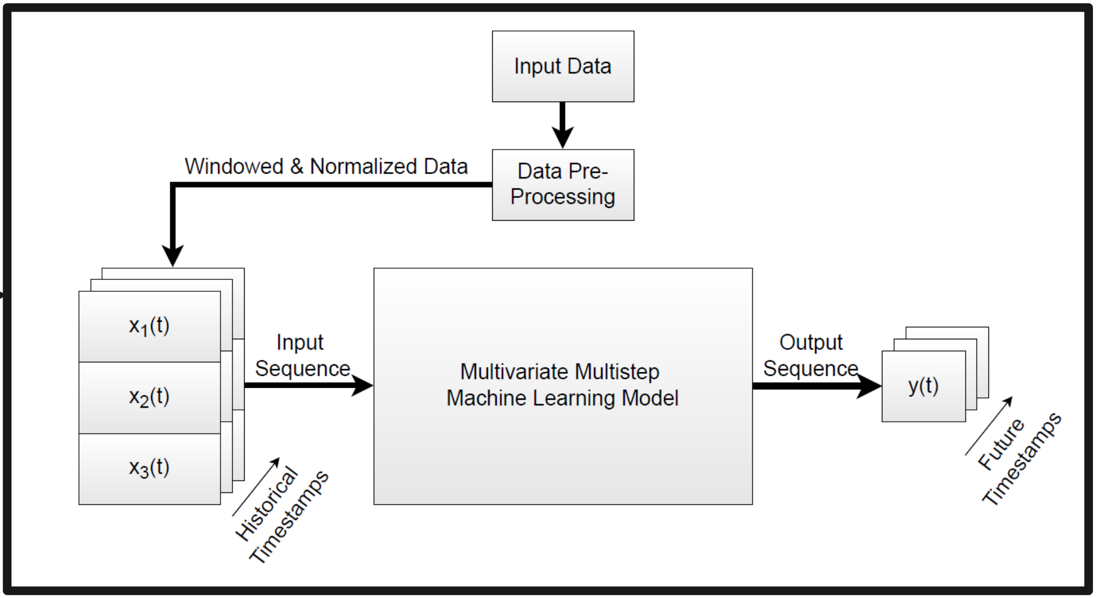
- Evaluation Metric*: $$F1 = \frac{precision \cdot recall}{precision + recall}$$

Input and Output

$$X = \begin{bmatrix} x^1_{taggedThroughput} & x^1_{backgroundThroughput} & x^1_{backgroundUsers} \\ x^p_{taggedThroughput} & x^p_{backgroundThroughput} & x^p_{backgroundUsers} \\ \vdots & \vdots & \vdots \\ x^P_{taggedThroughput} & x^P_{backgroundThroughput} & x^P_{backgroundUsers} \end{bmatrix}$$

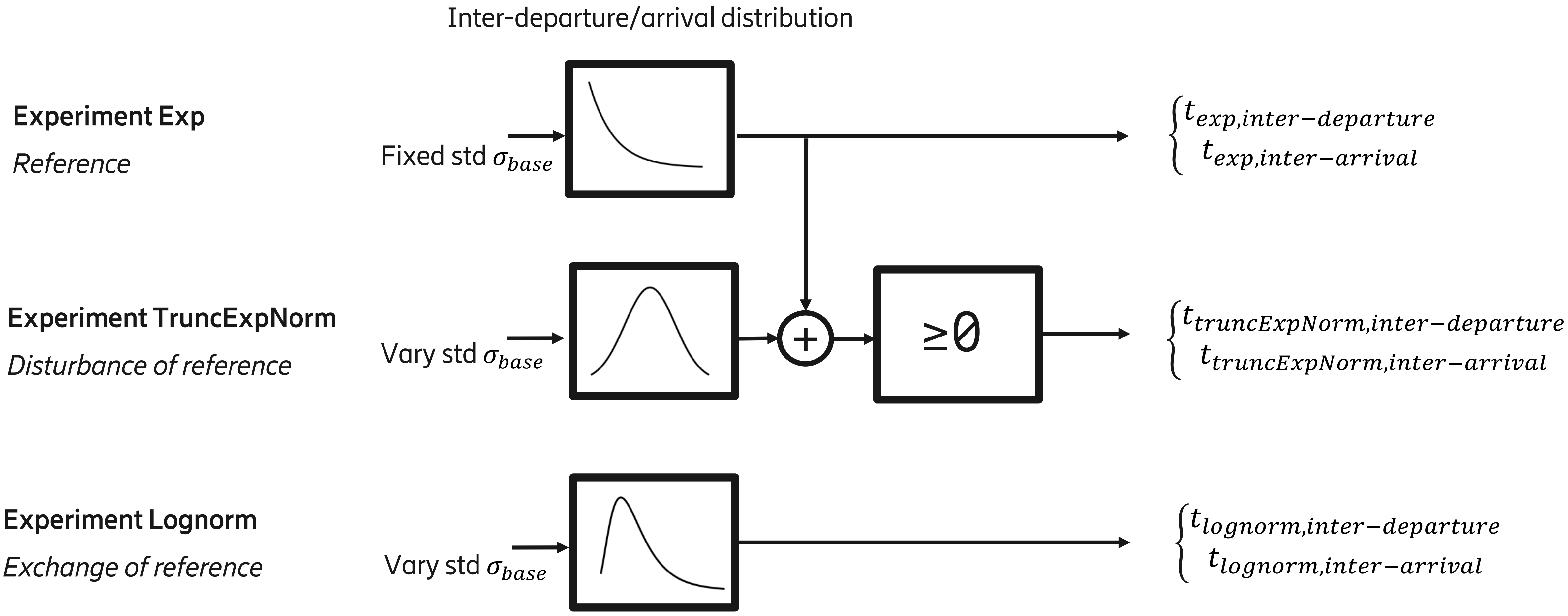
$$Y = \begin{bmatrix} y^{P+1}_{taggedThroughput} \\ y^f_{taggedThroughput} \\ \vdots \\ y^F_{taggedThroughput} \end{bmatrix}$$

Machine Learning Pipeline



* $precision = \frac{True\ Positive}{True\ Positive + False\ Positive}$, $recall = \frac{True\ Positive}{True\ Positive + False\ Negative}$

Trials



Introduction of a variation factor f_σ to summarize the applied change in standard deviation *: $f_\sigma = \frac{\sigma_{base}+\Delta\sigma}{\sigma_{base}}$

* σ_{base}: Baseline value that serves as reference, $\Delta\sigma$: Change to baseline

Trials Consideration

- Fixed configurations for all experimental trials

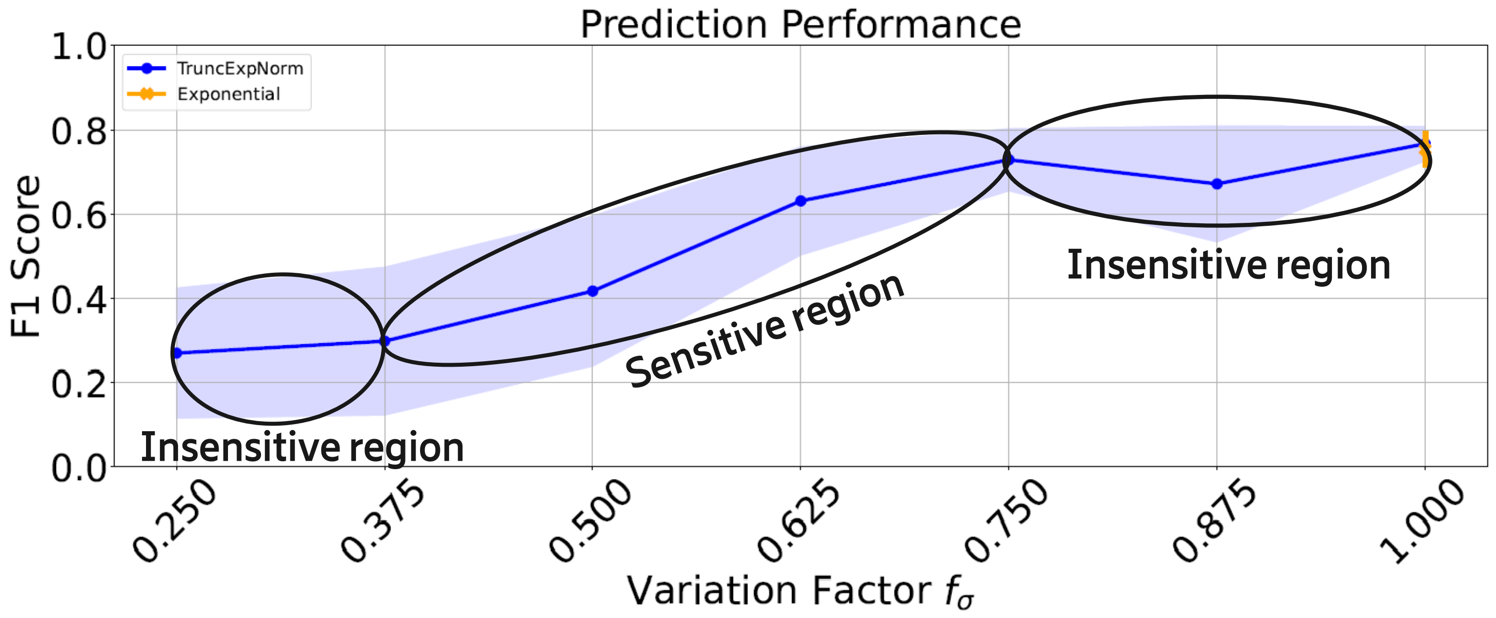
Parameter	Value
Future prediction horizion	4 s
Mean number of users served by the cell	30
Trial time	8 h
Number of samples (training: 70 %|testing: 30 %)	20160| 10 batches of 864 samples

- Variation of a baseline standard deviation used for the inter-arrival/departure mobility distribution

Parameter	Experiment Exp	Experiment TruncExpNorm	Experiment Lognorm
Standard deviation variation factor f_σ	1	[0.25, 1]	[1, 8]

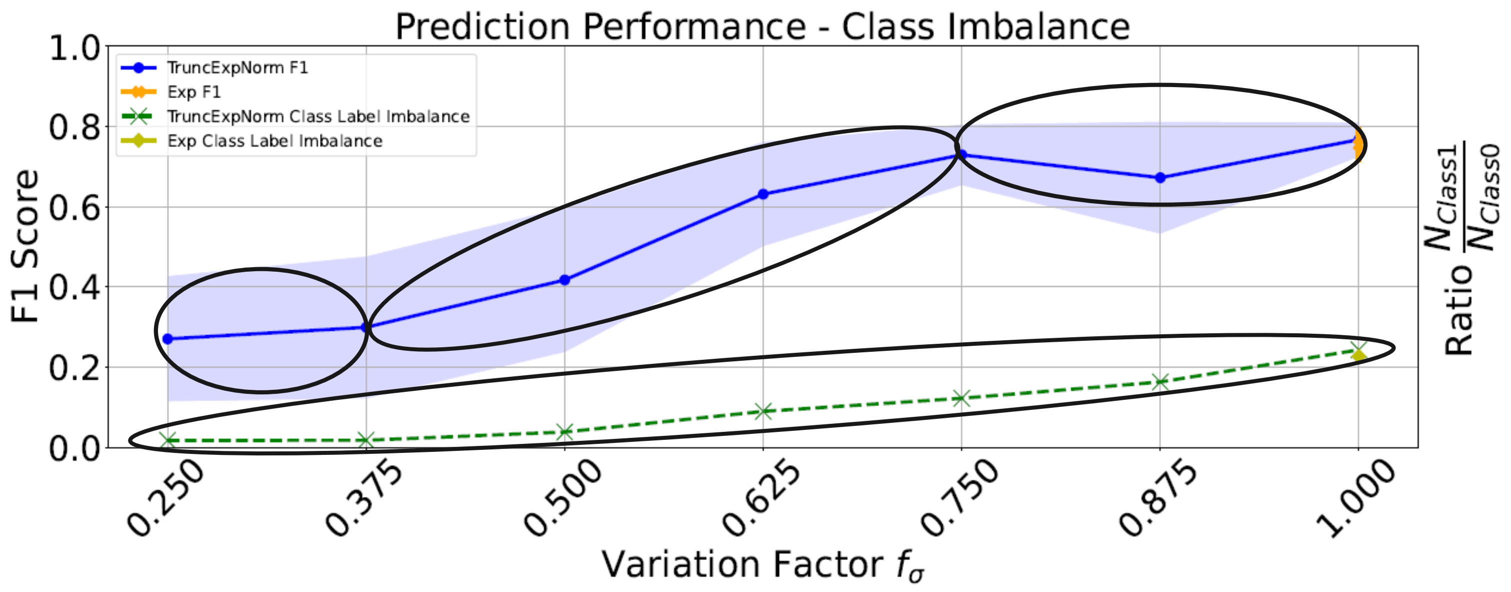
Results – Experiment TruncExpNorm

- ML Model behavior
  - Non-linearity of predictive performance in terms of F1 score
  - Different sensitivity regions



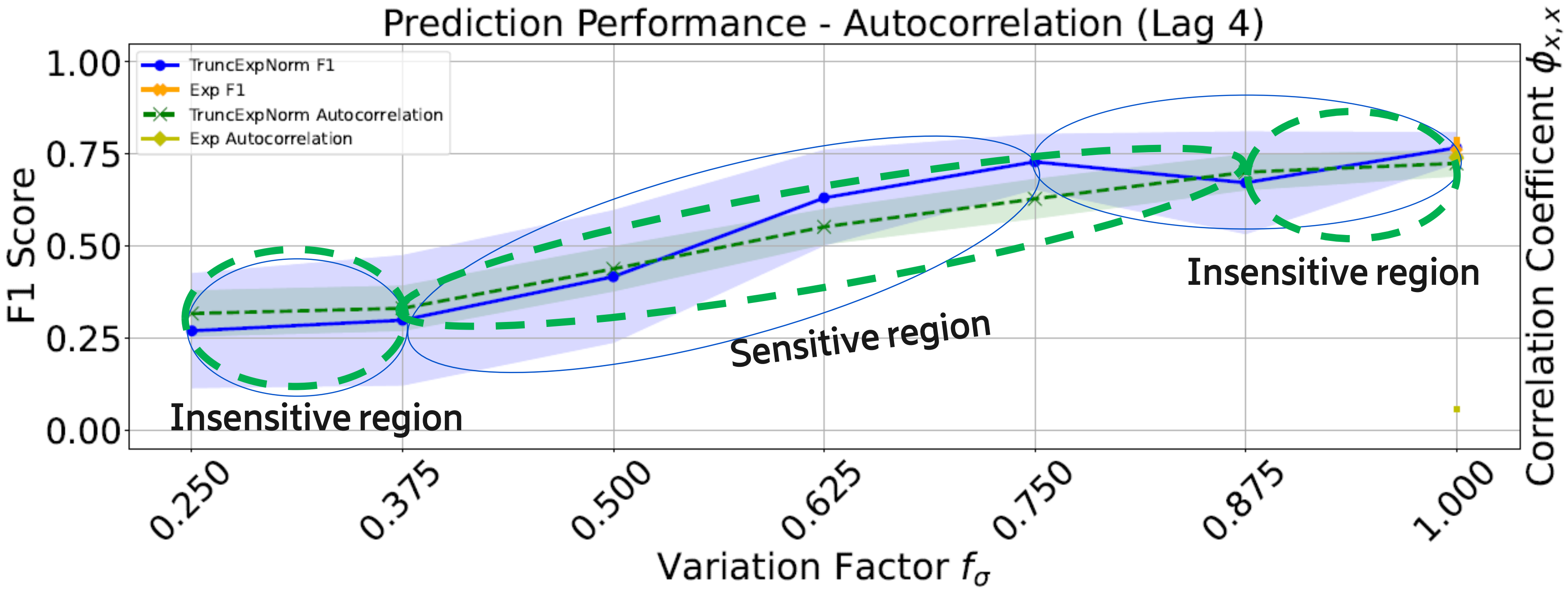
Results – Experiment TruncExpNorm: Class Ratio

- Class Ratio
  - Class 1: Tagged Throughput < 1 Mbit/s
  - Class imbalance: Rare cases of Class 1
  - Class 1 sample size grows linearly
- Comparision to F1 score
  - Unlike F1 score; linear curve progression
  - Can be roughly described by one sensitivity region



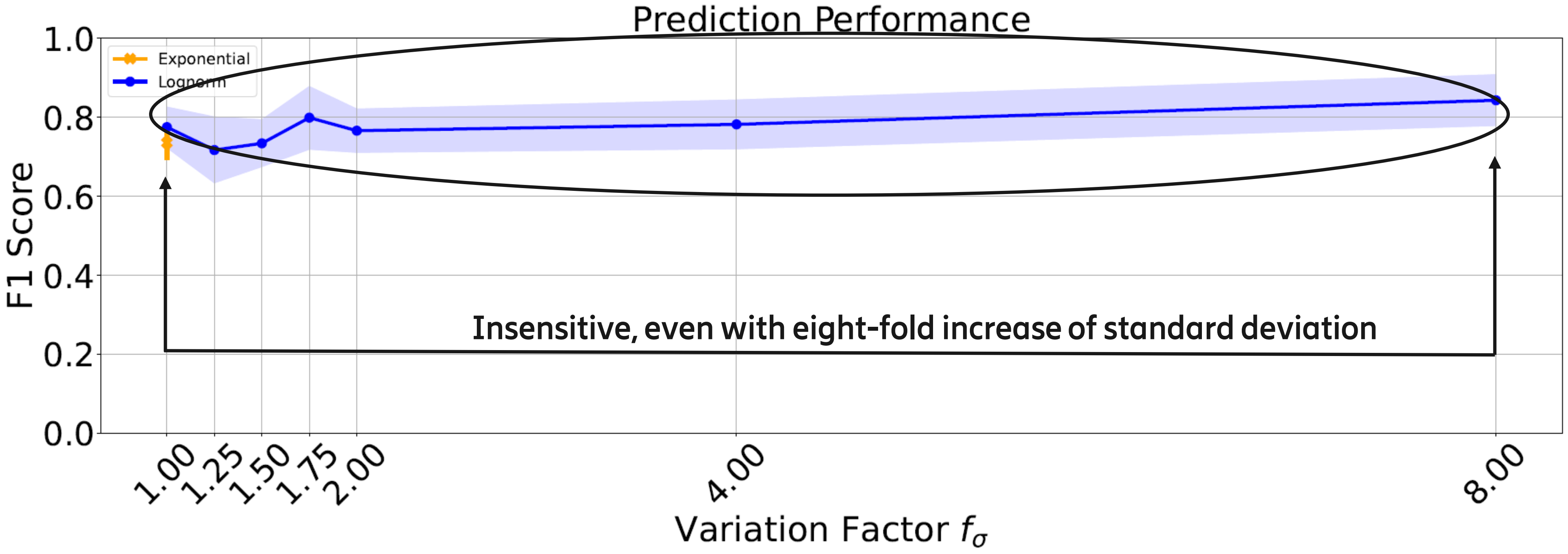
Results – Experiment TruncExpNorm : Autocorrelation

- Autocorrelation of tagged user throughput
  - Coefficient $\phi_{x,x}$ shows non-linear behavior
  - Different sensitivity regions
- Comparision to F1 score
  - Strong similarity between autocorrelation and F1 score
  - Unlike other aspects; similarity is consistent across variations



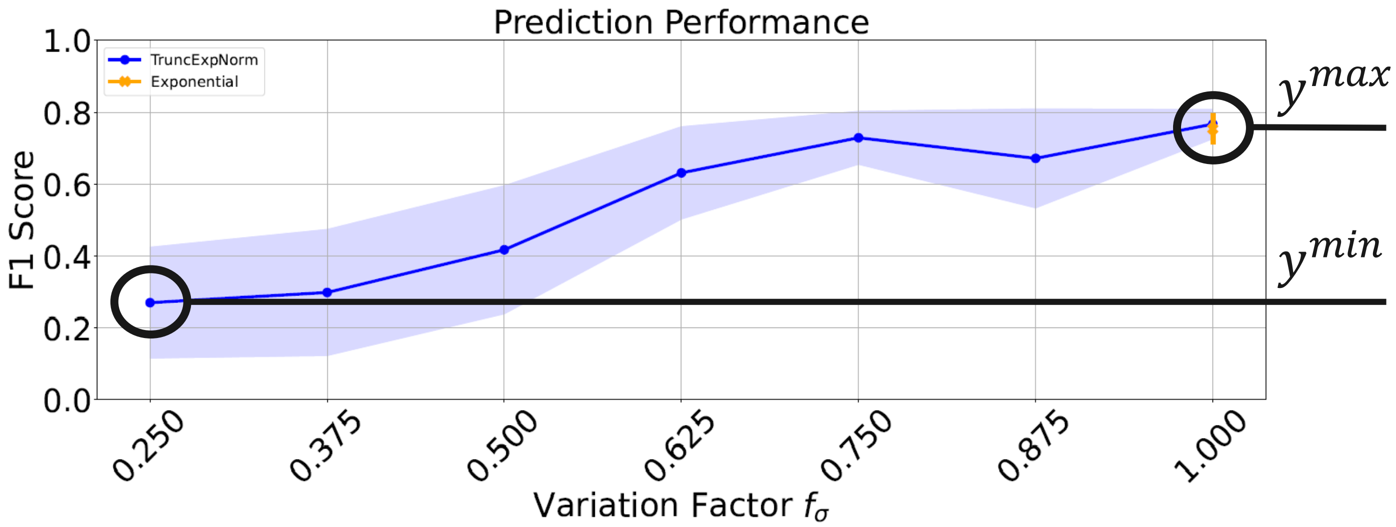
Results – Experiment Lognorm

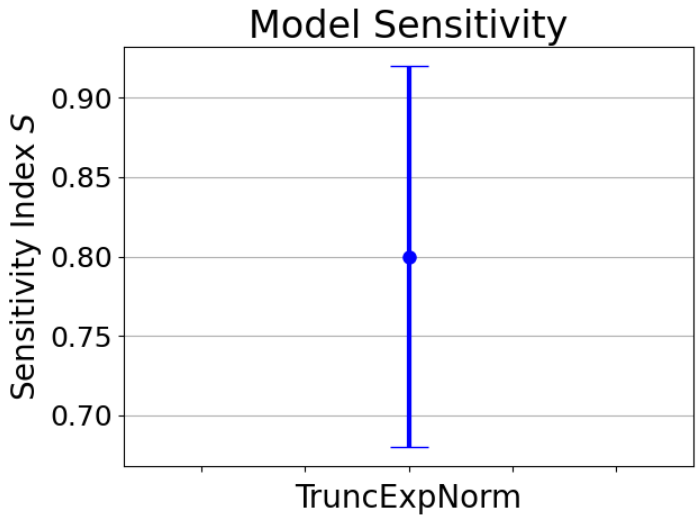
- Increase in standard deviation
  - Experiment TruncExpNorm only allows to reduce the standard deviation
- Comparision of mobility models
  - Similar performance
- ML Model behavior
  - Robust



Sensitivity Analysis

- Sensitivity index S proposed by Hoffmann and Gardner: $$S = \frac{y^{max} - y^{min}}{y^{max}} \text{ with } S \in [0, 1]$$





- The ML model shows a high sensitivity with a 95 % confidence of $S \approx [0.7, 0.9]$
  - Observation holds for the considered value range of f_σ: $\sigma_{arrival} = [0.3s, 1.2s]$, $\sigma_{departure} = [9s, 36s]$
- Non-linear behavior only visible in graph

Conclusion

- ***To what extent does uncertainty in input simulation/emulation models impact Machine Learning prediction performance?***

  – We identified and applied suitable sensitivity analysis methods to approach this question
  – We consider them applicable but **encourage others to also try them** to jointly come to a more tangible conclusion:
    - Identify where you have (stochastic) models in your AI/ML testbeds
    - Change the model (e.g. different distribution and/or temporal correlation) or adapt its parameters
    - Decide what should remain unchanged (e.g. the mean) to not change too much of the system

- Knowledge about sensitivity can be used for cautionary purposes, e.g.,
  – Building multiple models each for a specific range of values of a sensitive parameter
  – Focus research on a certain sensitive parameter to get a better understanding of the true probabilistic nature

- Outlook
  – Repeat with another model than the one for number of users in cell
  – Apply for more than one model/parameter at once, to determine relative significance/sensibility

Acknowledgment

This work was done in collaboration with the Technical University of Berlin

Special thanks to Prof. Dr. -Ing. Slawomir Stanczak

Thank You For Listining !

Configuration

- Fixed configuration for emulation/simulation, Machine Learning, and Radio Access Network

Model Parameter	Configruation
Packet Length M [Byte]	1500
Tagged User Throughput L [Mbit/s]	1.08
Mean Load ρ [%]	≈100
Mean inter-Departure Time [s]	36
Mean inter-Arrival Time [s]	1.2
Expectation Background User Number $E[N_B]$	30
Trial Time (Warm-Up Period) [h]([min])	8 (4)
Past Histroic Window P [s]	10
Future Prediction Horizon F [s]	4
Train:Test Datasize Ratio [%]	70:30
Number of testset batches N_{batch}	10 á 864

RAN Parameter	Configruation
Transmission Direction	Uplink
Transport Protocol	UDP
Max. Spatial Layers v_L	1
Max. Modulation Order Q_m	6
Max. Code Rate R	948/1024
Subcarrier Spacing Δf [kHz]	30
Channel Bandwidth B [MHz]	20
Max. Number of Resource Blocks N_{RB}	49 out of 51
Overhead Factor (Frequency Range 1)	0.08
Time Division Duplex UL-DL	2 out of 10 slots
Estimated Capacity $\tilde{C}$ [Mbit/s]	15.564
Measured Capacity C [Mbit/s]	≈15.633

M/M/inf

Birth-death Markov process motivated by literature

- Various models for vehicle arrivals and departure in literature such as Poisson or Lognormal
- Not asking which model is the best, but what impact does model inaccuracy have on prediction accuracy (sensitivity analysis)

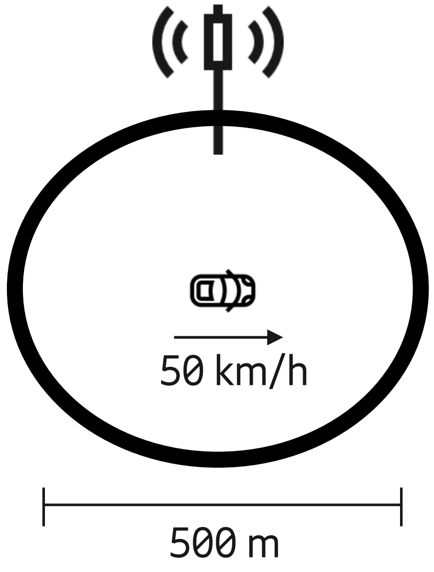
- Offered traffic Background User: Why Poisson traffic?
  - Palm's theorem: For many independent arrival processes, the overall process can be seen as a Poisson process
  - Assumption: Large number of mobile communication application services run by a user

- Offered traffic Tagged User: Why fixed?
  - Extremely simplified approach to model a video stream
  - Scheduler's point of view: The video stream is usually present in the form of large packets with constant size
  - Assumption: Each packet size correspond to the maximum transmission unit

Mean Number of Users

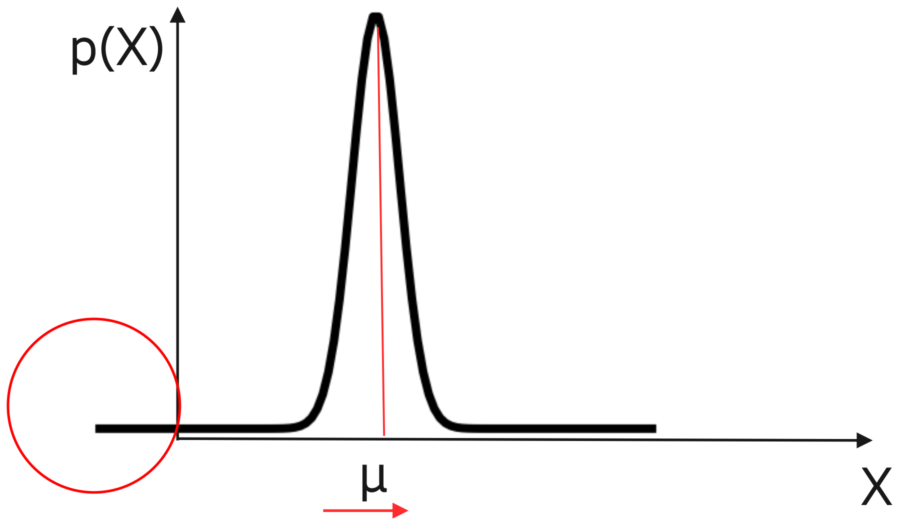
- Mean inter-departure time of a vehicle: 36 s
  - Linear vehicle trajectory with constant velocity and cell diameter
- Mean number of users: 30
  - Selection of relatively few users
  - Reduced bandwidth to 20 MHz: Allows to saturate the system with a lower CPU load
- Mean inter-arrival time of a vehicle entering cell: 1.2 s
  - Selection based on mean inter-departure and mean user number
  - Little’s Law for queuing systems:

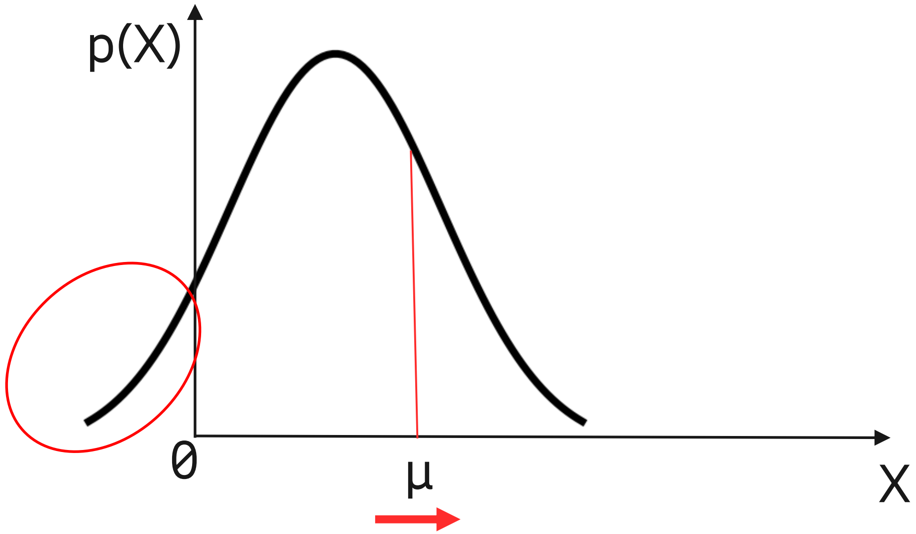
$$L = \boldsymbol{\lambda} W$$
$$L - Average\ number\ of\ items\ within\ the\ system$$
$$\boldsymbol{\lambda} - Average\ arrival\ rate\ of\ items\ into\ the\ sytem$$
$$W - Average\ time\ an\ item\ spends\ in\ the\ system$$



TruncExpNorm

- Sum of two random variables, i.e., convolution of an exponential PDF and normal PDF
- Truncation to prevent negative inter-time samples
  - The mean becomes a function of the standard deviation
  - The higher the standard deviation, the more the mean is shifted in the direction of ∞+

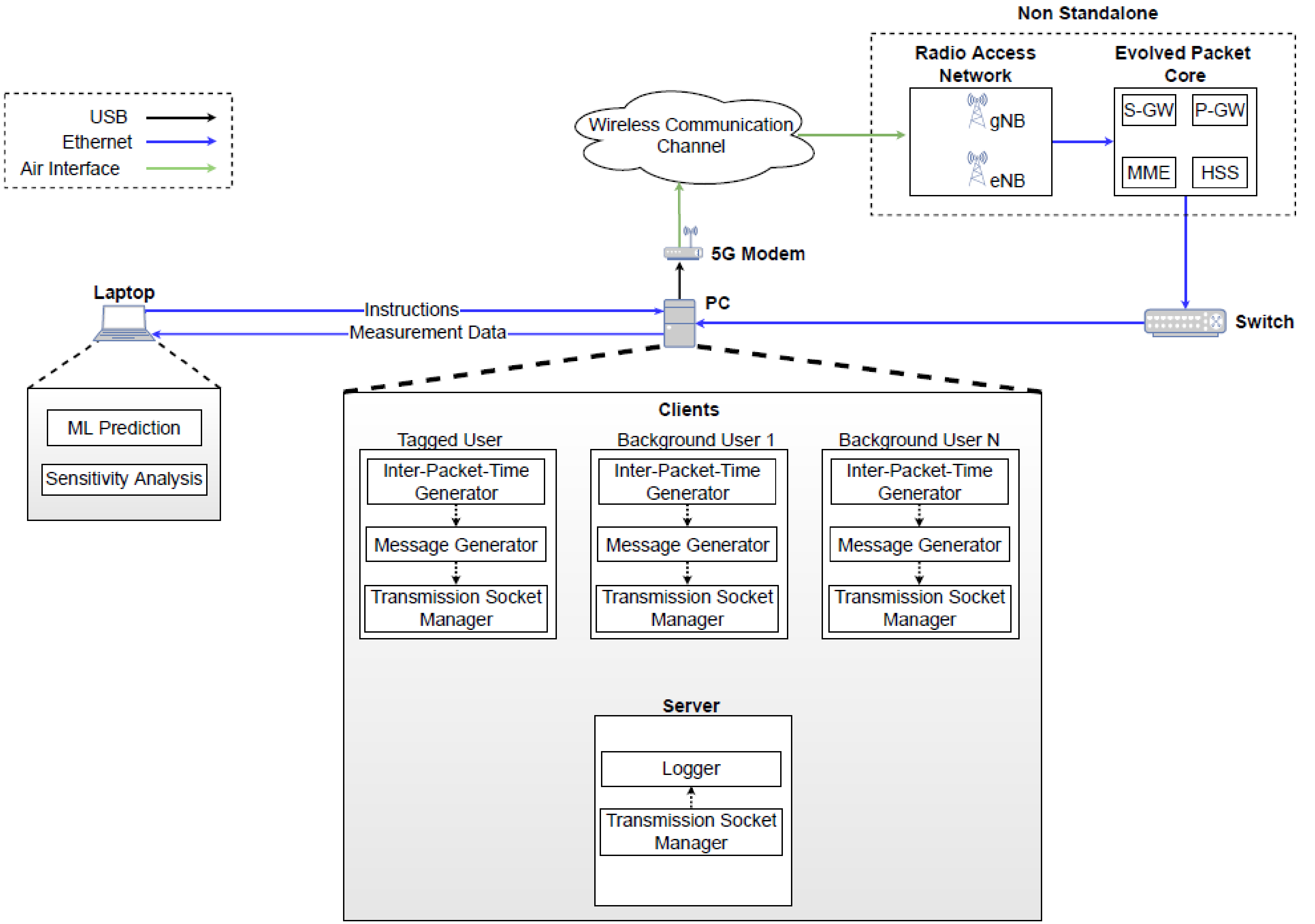




Sensitivity Analysis

1. Run prediction for each variation and collect max. and min. F1
2. Insert local extreme values into the formula for the sensitivity index and collect the sensitivity value
3. Repeat step 1 and 2 x times and collect x sensitivity indexes
4. Subdivide sensitivity values into k batches and calculate the mean of each individual batch
5. Estimate true mean using indivdual means
6. Use T-distribution (true standard deviation not known, samples size < 30) to get the confidence interval

Toolchain



USB
Ethernet
Air Interface
Non Standalone
Radio Access Network
Evolved Packet Core
gNB
eNB
S-GW
P-GW
MME
HSS
Wireless Communication Channel
5G Modem
PC
Switch
Laptop
Instructions
Measurement Data
ML Prediction
Sensitivity Analysis
Clients
Tagged User
Background User 1
Background User N
Inter-Packet-Time Generator
Message Generator
Transmission Socket Manager
Inter-Packet-Time Generator
Message Generator
Transmission Socket Manager
Inter-Packet-Time Generator
Message Generator
Transmission Socket Manager
Server
Logger
Transmission Socket Manager

Tele-operated Driving and Predictive Quality of Service