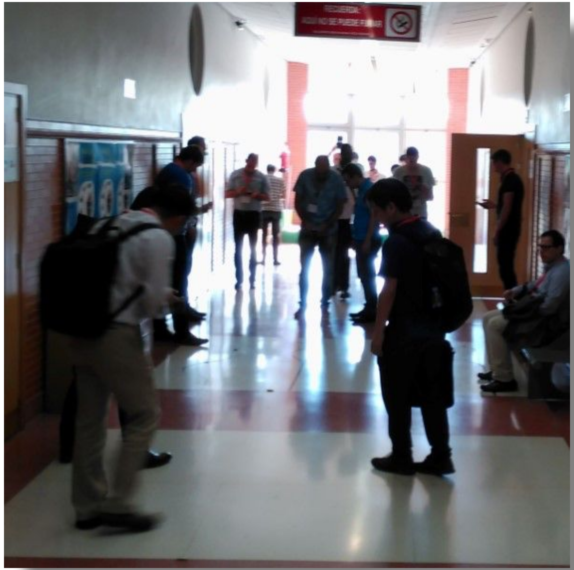
Course in **Machine Learning 2** -- A.A. 2021/2022
Robotics Engineering



UJI Indoor Localization

Compare WLAN fingerprinting indoor localization algorithms

Francesco Ganci - 4143910
Federico Civetta - 4194543

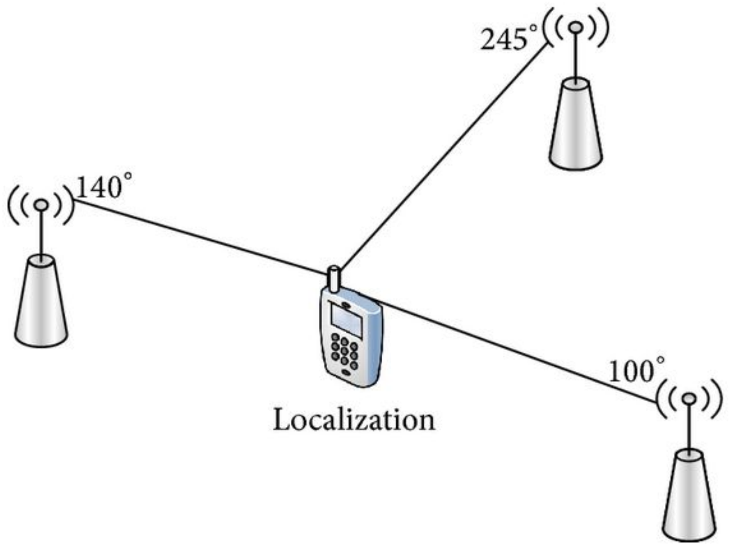
F

Intro - the Problem

Main points of this work:

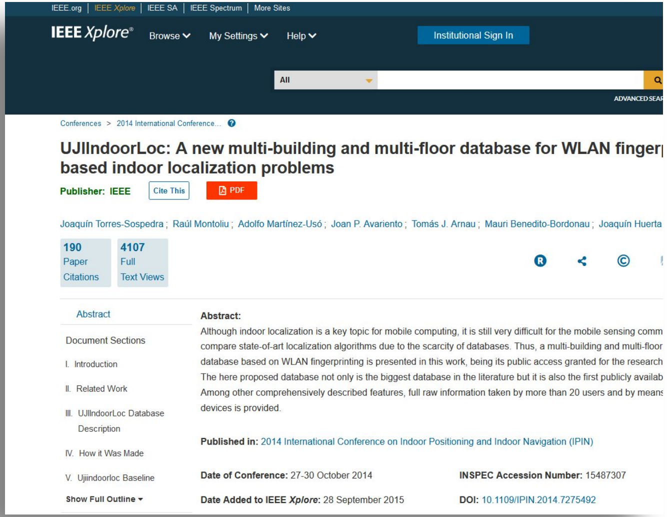
- a problem about **infrastructure-less 2D indoor localization** in a large scale environment
- *a problem with a paper*

- *Support Vector Regression* using the SciKitLearn framework



Ff

Intro - the Paper



IEEE.org | IEEE *Xplore* | IEEE SA | IEEE Spectrum | More Sites

IEEE *Xplore*® Browse My Settings Help Institutional Sign In

All

ADVANCED SEAR

Conferences > 2014 International Conference...

UJIIndoorLoc: A new multi-building and multi-floor database for WLAN finger based indoor localization problems

Publisher: IEEE Cite This PDF

Joaquín Torres-Sospedra ; Raúl Montoliu ; Adolfo Martínez-Usó ; Joan P. Avariento ; Tomás J. Arnau ; Mauri Benedito-Bordonau ; Joaquín Huerta

190 Paper Citations **4107** Full Text Views

Abstract

Document Sections

I. Introduction

II. Related Work

III. UJIIndoorLoc Database Description

IV. How it Was Made

V. Ujiindoorloc Baseline

Show Full Outline

Abstract:
Although indoor localization is a key topic for mobile computing, it is still very difficult for the mobile sensing comm compare state-of-art localization algorithms due to the scarcity of databases. Thus, a multi-building and multi-floor database based on WLAN fingerprinting is presented in this work, being its public access granted for the research The here proposed database not only is the biggest database in the literature but it is also the first publicly availab Among other comprehensively described features, full raw information taken by more than 20 users and by means devices is provided.

Published in: 2014 International Conference on Indoor Positioning and Indoor Navigation (IPIN)

Date of Conference: 27-30 October 2014

INSPEC Accession Number: 15487307

Date Added to IEEE *Xplore*: 28 September 2015

DOI: 10.1109/IPIN.2014.7275492

- Purpose: *making an **extensive free dataset** to allow other experts to test and compare different **indoor localization** algorithms using **WAP fingerprints** on a **large scale environment***
- A well-known dataset (see [IPIN2015 competition])

- inside: short paper review, dataset characteristics, examples of data visualization, methodology to obtain data

- **Paper available on IEEE** → [UJI Indoor Loc] (a paper published on 2014)
- Samples collected in three buildings of the *Jaume I University (in Spain)* → [link] [Maps]

f

About the Dataset

- already split into a training set and a test set (*calibration and operation*)
- *520 Wireless Access Point*, using RSSI measure for each WAP
- WLAN fingerprints are collected using different models of smartphone
- main features: *longitude*, *latitude*, *floor, timestamp, phone model, …*

- training set : *19,938 samples*
- test set : *1,111 samples*

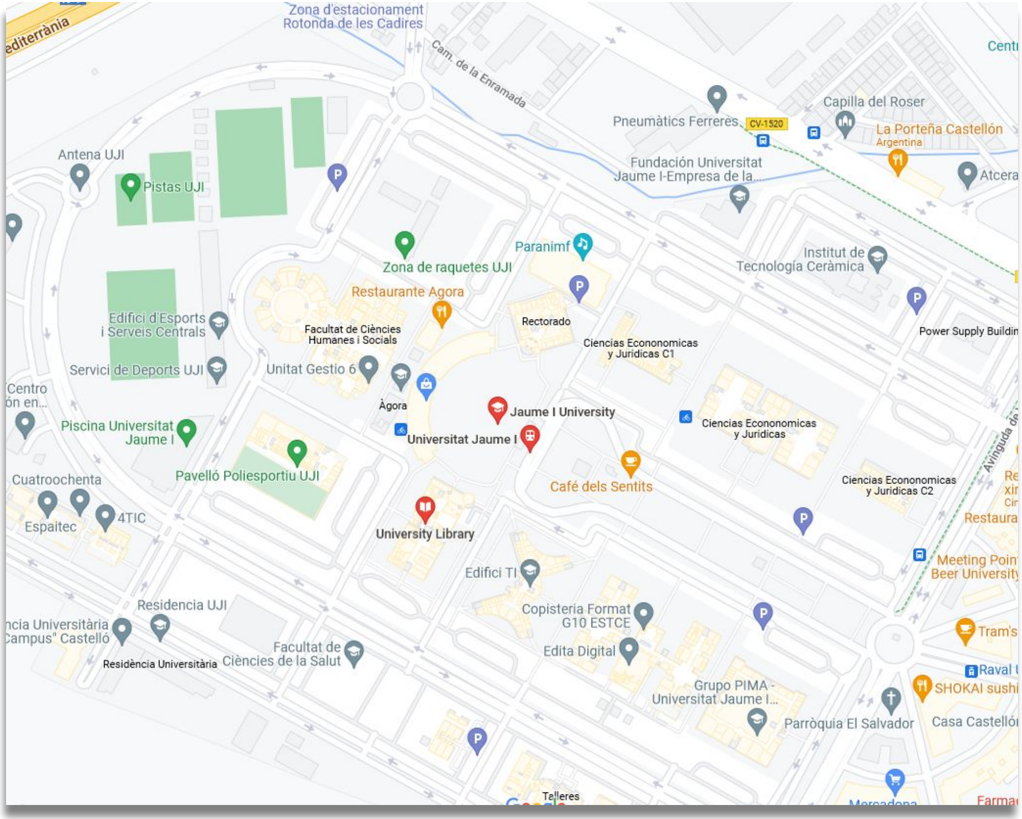
kaggle

UJI on Kaggle

UCI Machine Learning Repository

UJI on UCI ML repository

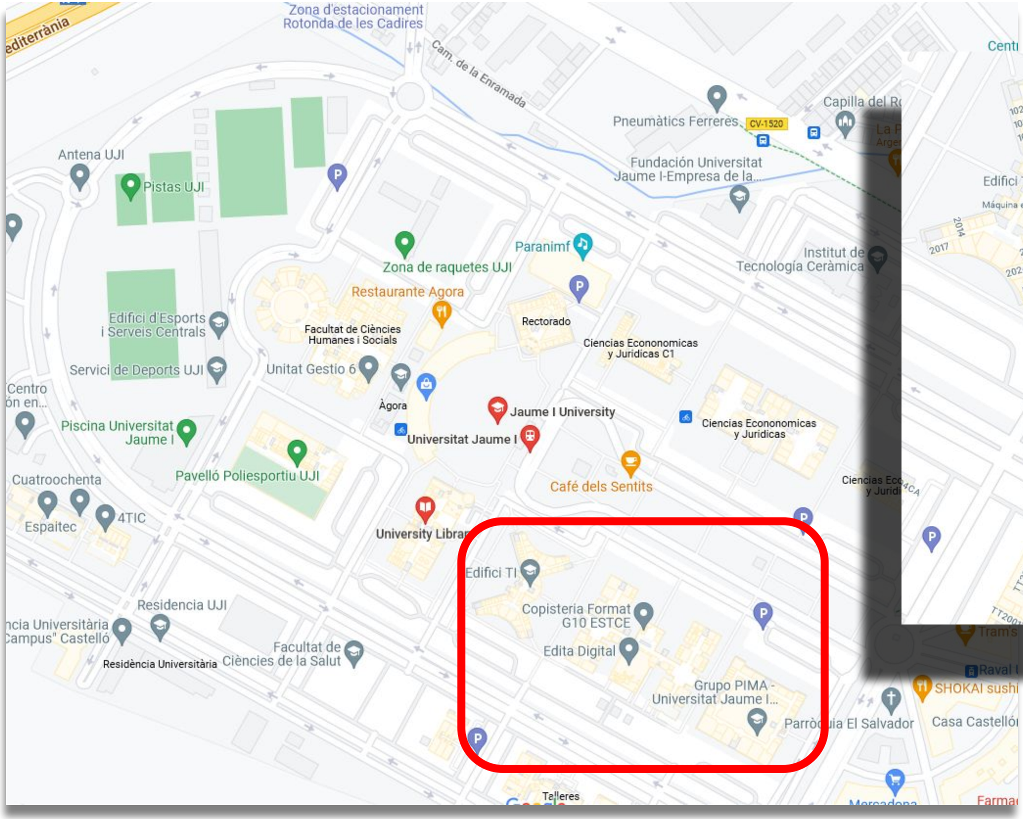
f

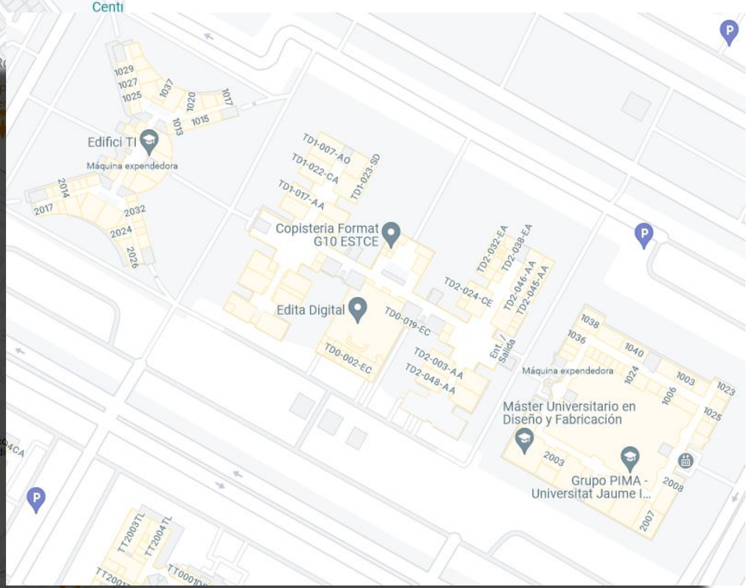


Infrastructure Less Approach

the localization system exploits WAP devices already available in the space, no specific localization system.

fF



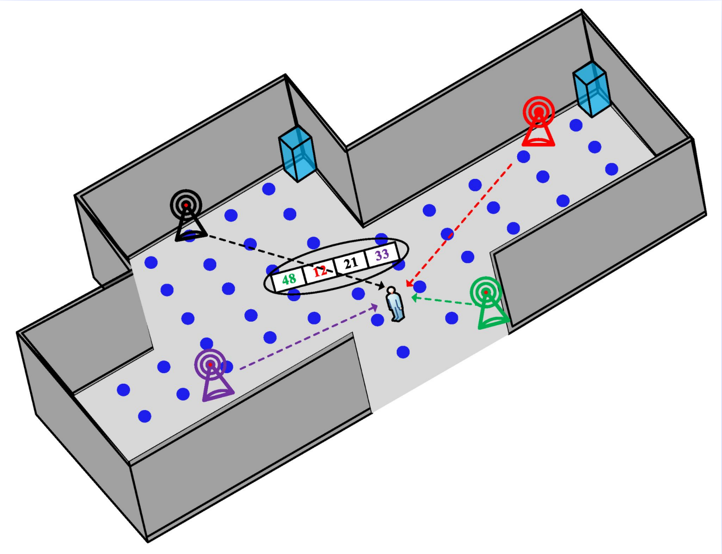


3 buildings over a surface of ***~109Km²***.

520 Wireless Access Points across the area.

F

About our Problem



Problem Statement:

→ *Absolute Indoor Infrastructure-less 2D geometrical Localization*

- Given the 520 RSSI values of the WAP fingerprint at a unknown location, …
- *… prediction of <u>longitude and latitude</u> of the user*

- **Regression Problem** : 520 to 2

Our solution Strategy

A solution with **Support Vector Regression (SVR)**:

- robust and flexible machine learning algorithm
- able to apply *data compression* “learning by mistakes”
- *so many implementations already available*

Our strategy:

- **two different learning machines**, one for each coordinate to predict, same input space for each learner

Ff

Data Preparation and Visualization

- samples taken from the sets using a uniform distribution
  - Training set: 70% (13,956 samples) out of 19,938 samples
  - Validation set: 30% (278 samples) out of 1,111 samples

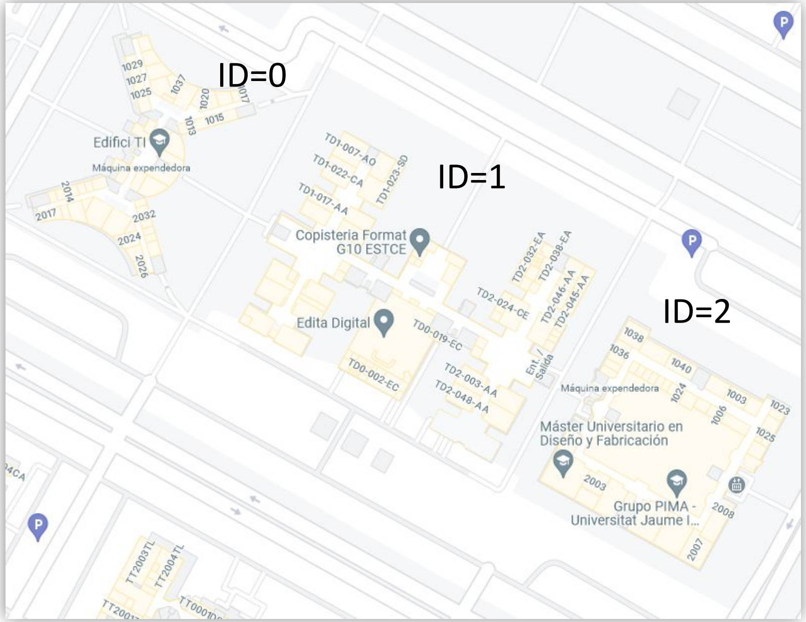
- data visualizations, according to the paper (see later)

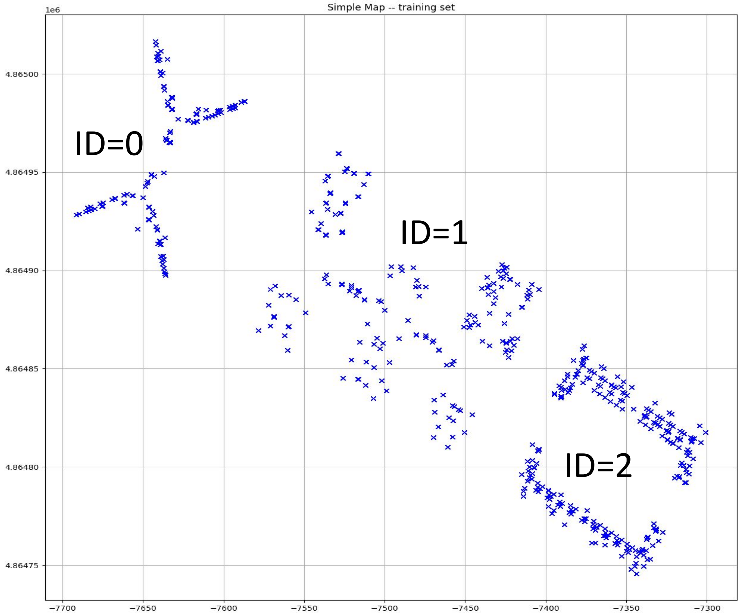
- *no data selection and handling*: data are simply taken *by chance* from the sets using a uniform distribution

- Standard normalization in [0, 1] of the inputs (using *MinMaxScaler* in sciKitLearn)

f

The real environment, as it appears on Google Maps:





The environment, obtained *from data*, over a small number of entries (**13,956** training samples out of 19,938)

f

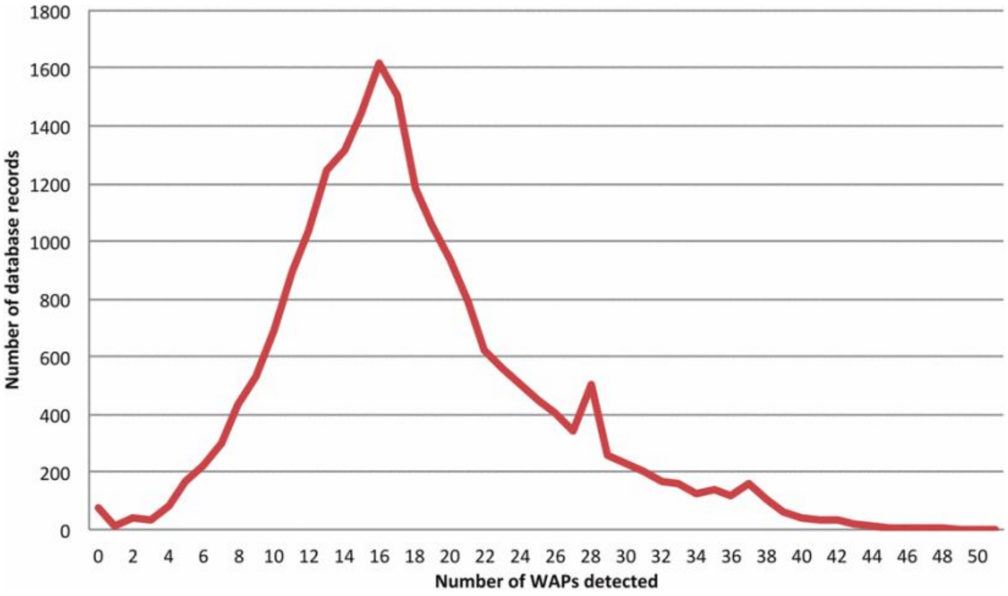
Visuals - 1 - num of WAP detected

(from the paper) About the graph:

- x axis : number of WAPs detected in a sample
- y axis: the number of samples detecting that number of WAPs

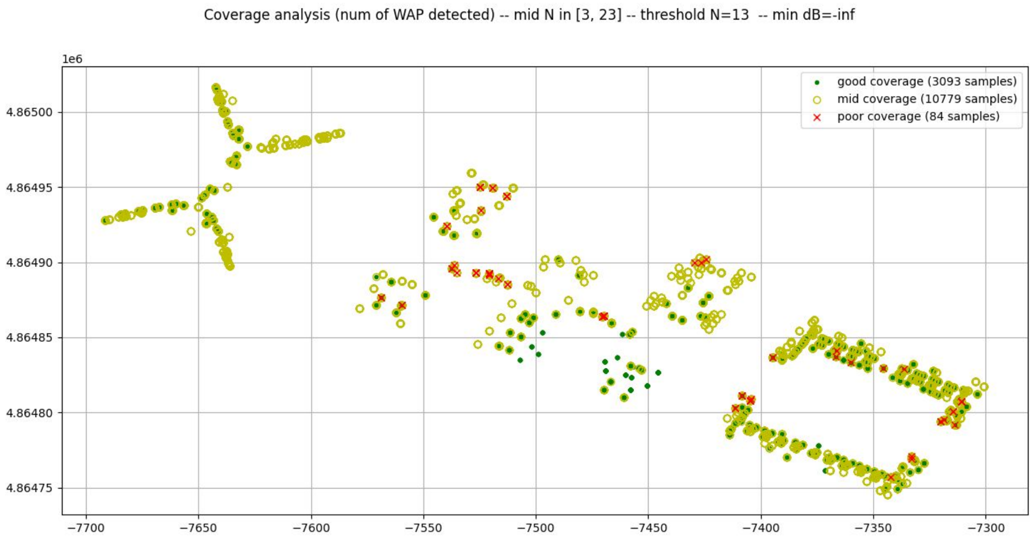
Observations:

- hard to detect more than 23 WAPS
- few samples detect less than 3 WAPs
- *working on a high number of WAP detections*



fF

Visuals - 2 - distribution map



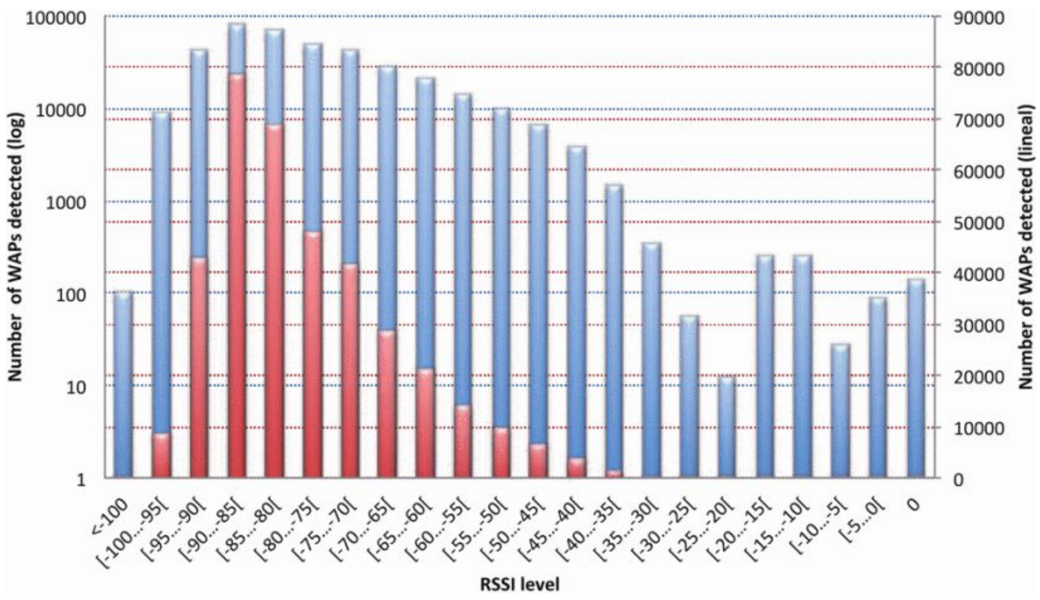
The map, obtained from the limited training set, shows:

- (red crosses) points detecting less than 3 WAPs
- (orange circles) from 3 to 23 WAPs
- (green points) more than 23 WAPs

Observations:

- no emerging coverage regions across the map
- as observed before, most of points detect from 3 to 23 WAPs

Visuals - 3 - RSSI values



(From the paper) Distribution of values from the successfully detected RSSI values.

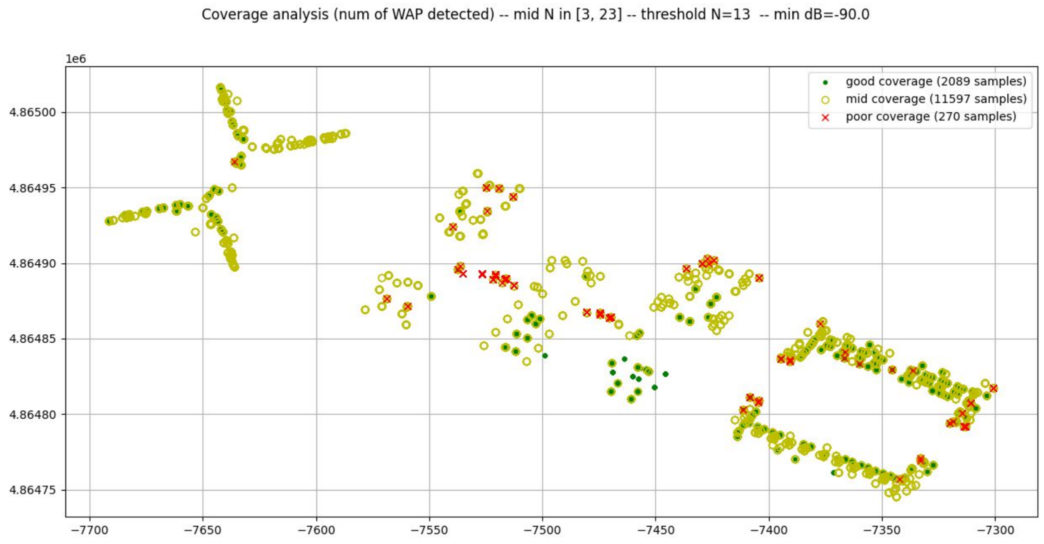
- red bars : num of samples in linear scale
- blue bars : num of samples in log scale

Observations:

- most of RSSI values are in a range from -90dB to -65dB
- <u>*working on small RSSI values from the WAPs*</u>

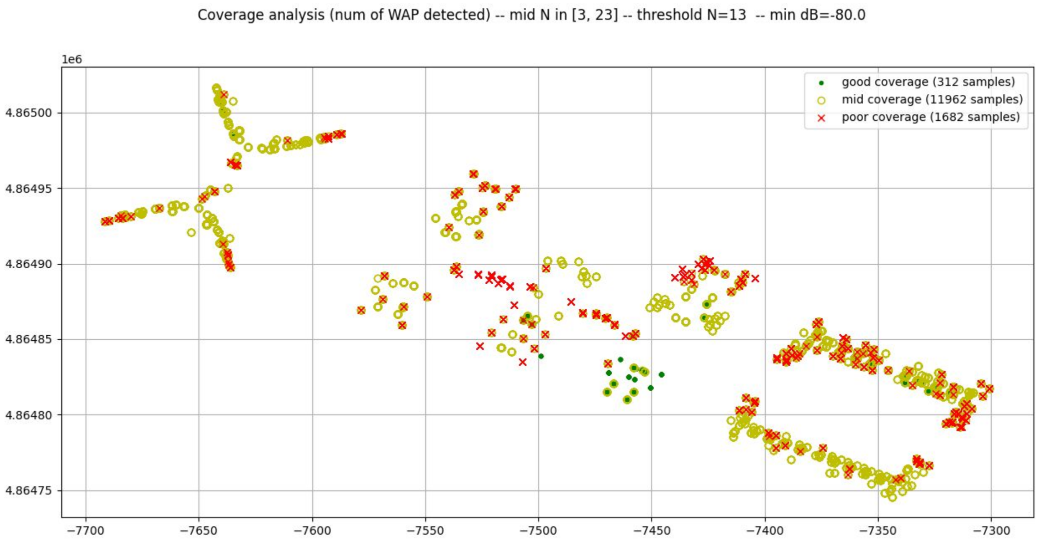
F

Visuals - 4a - distr. map with min RSSI



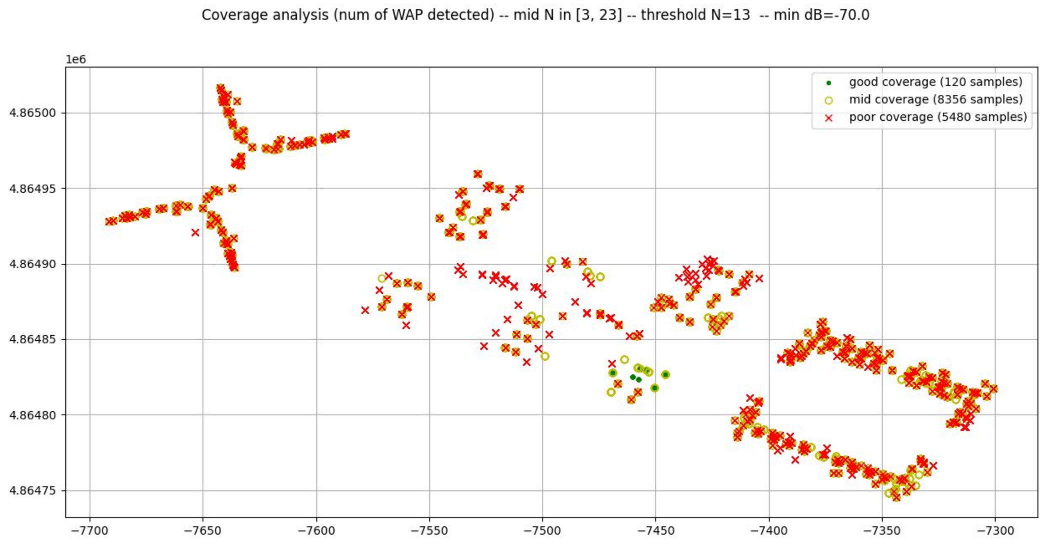
F

Visuals - 4b - distr. map with min RSSI



Ff

Visuals - 4b - distr. map with min RSSI



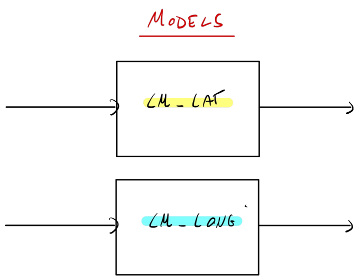
f

Model selection and Training

- First (experimental) Implementation with Colab
  - … quite unefficient (almost 1h 30min for only the 28% of the training set)
  - Links: the very first notebook, learning longitude, learning latitude, data visuals

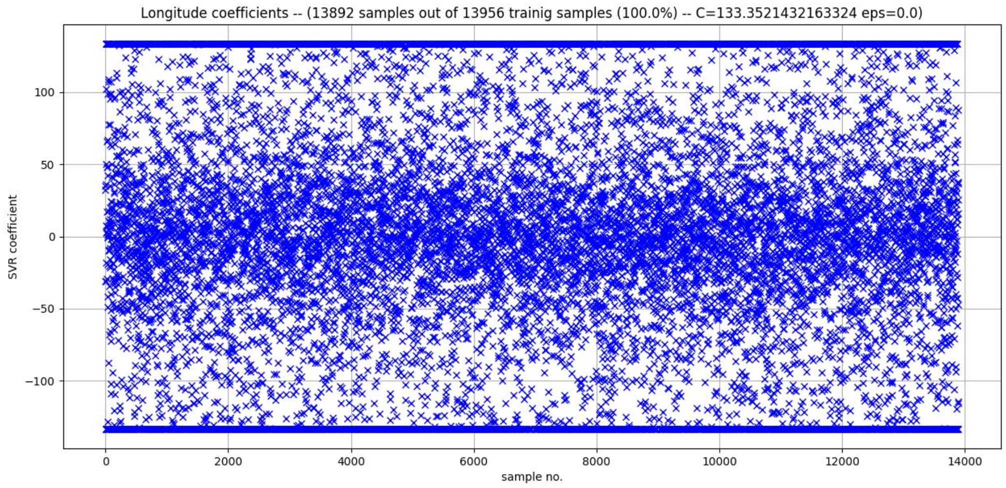
- Second implementation spreading the job on many Py threads:
  - 1h 30min for the 70% of the training set, see the training log
  - see the project On GitHub

- Research range for the hyperparameters (1250 combinations with cross validation 2):
  - C : logspace(-4, 3, 25)
  - gamma: logspace(-4, 3, 25)
  - epsilon: [0, 0.01]



f

Model selection and Training



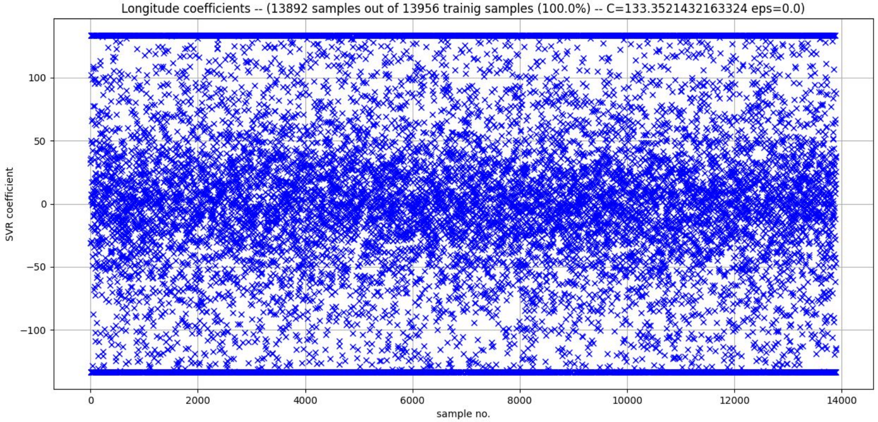
Functional form for SVR (from the SciKit Learn documentation):

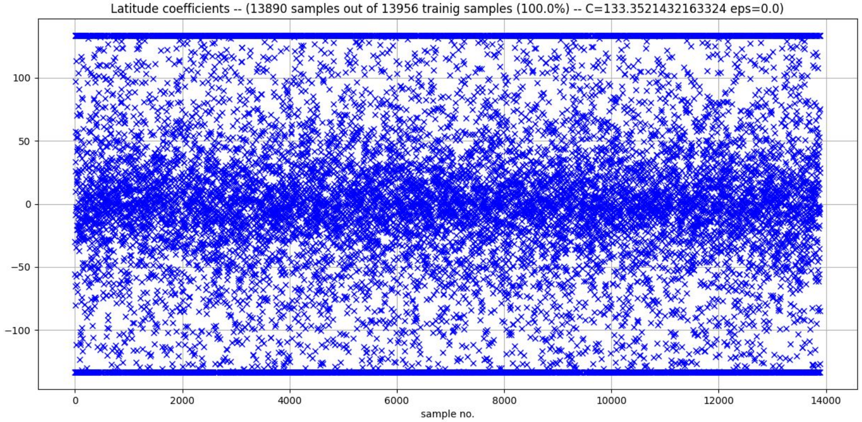
$$\sum_{i \in SV} (\alpha_i - \alpha_i^*) K(x_i, x) + b$$

- almost dense solution, because of the small epsilon
- coefficients follow the expected behaviour: the difference between coefficients is between -C and C.

f

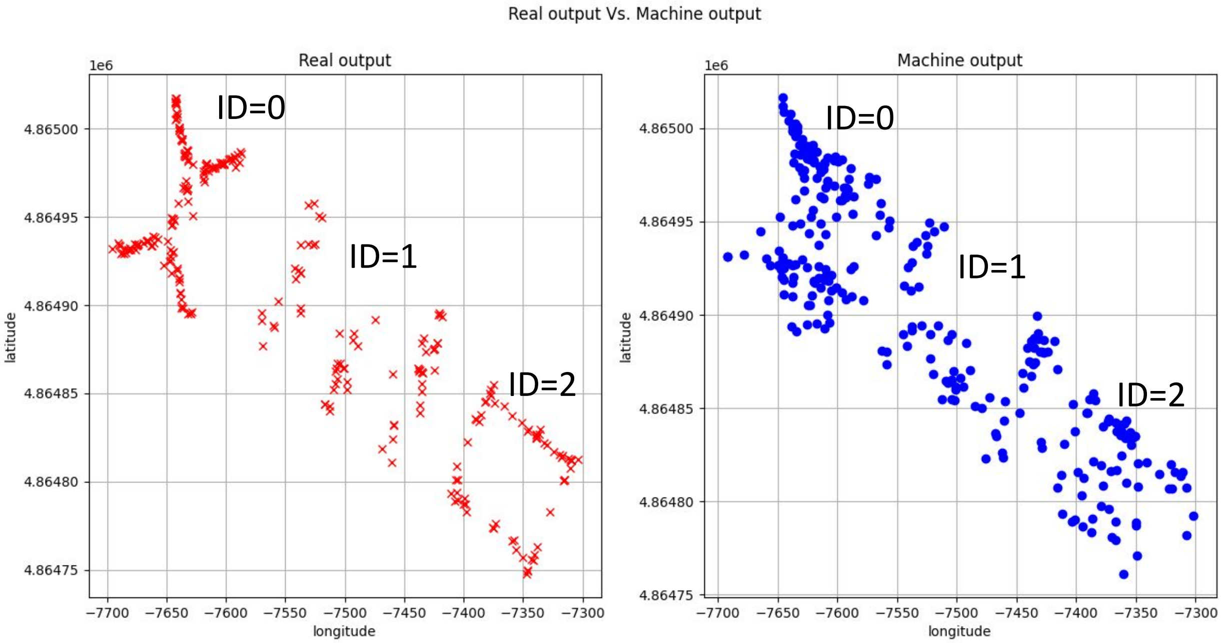
Model selection and Training





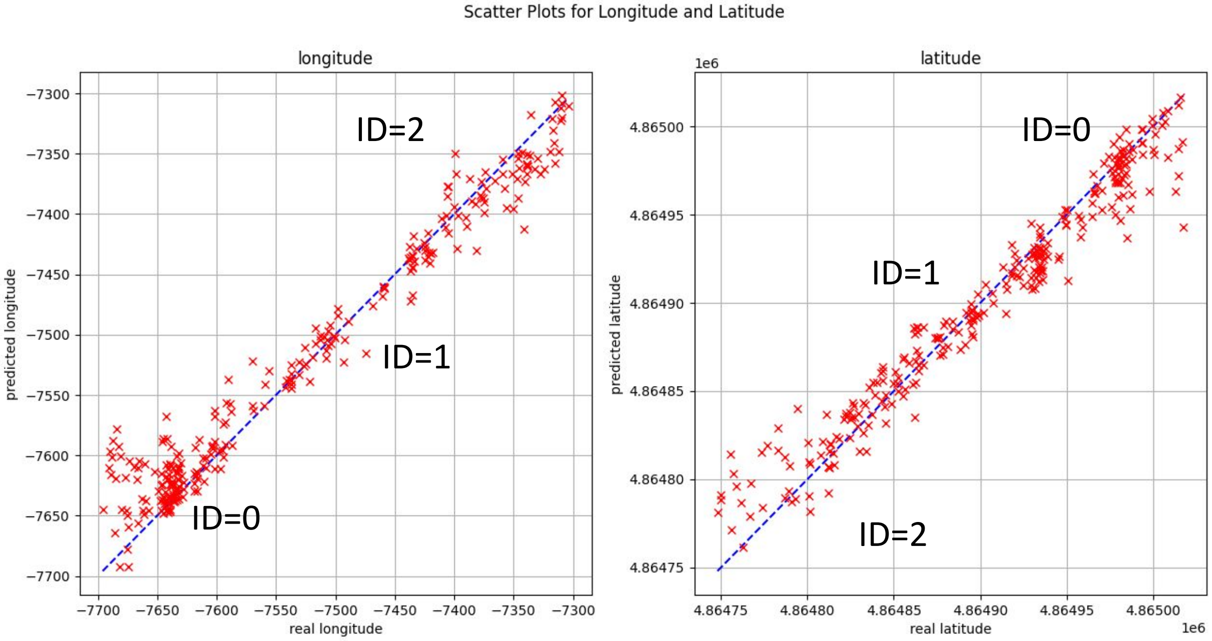
f

Results - 1 - first map comparison



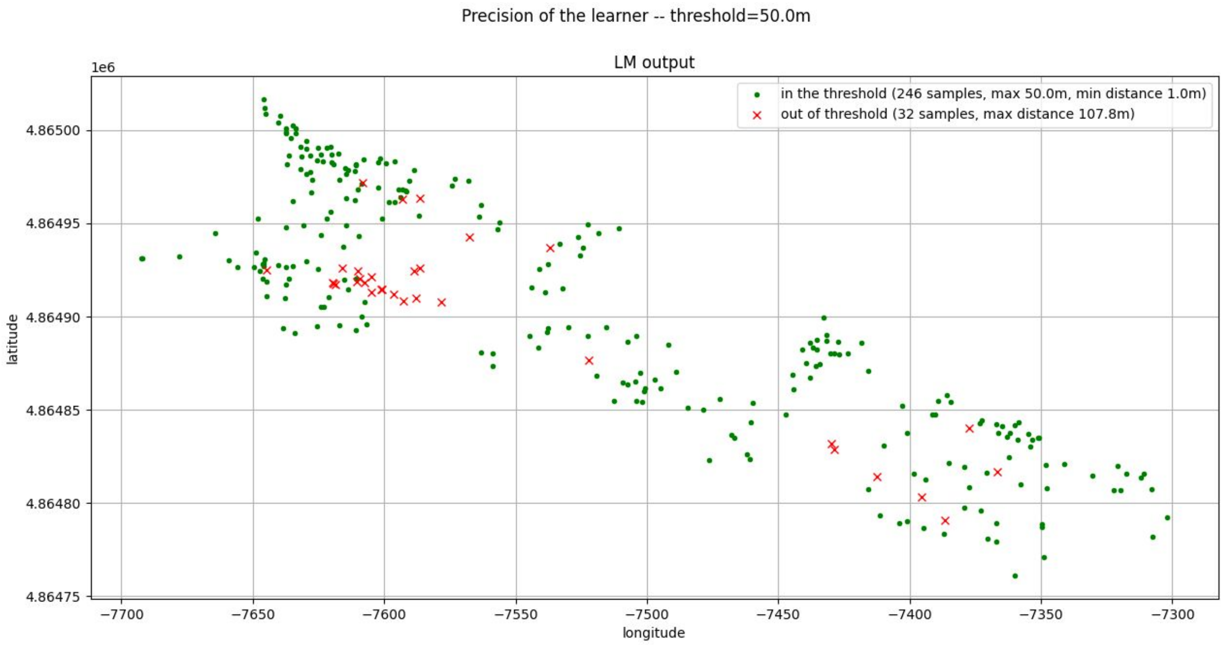
fF

Results - 2 - scatter plot



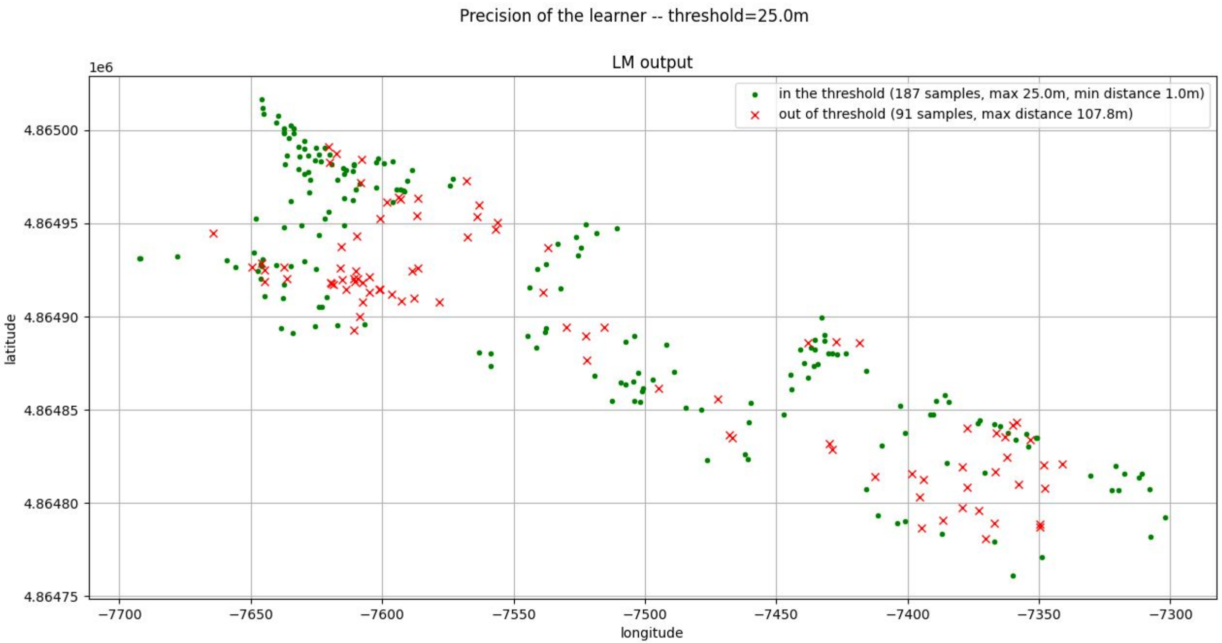
F

Results - 3a - precision in 50m



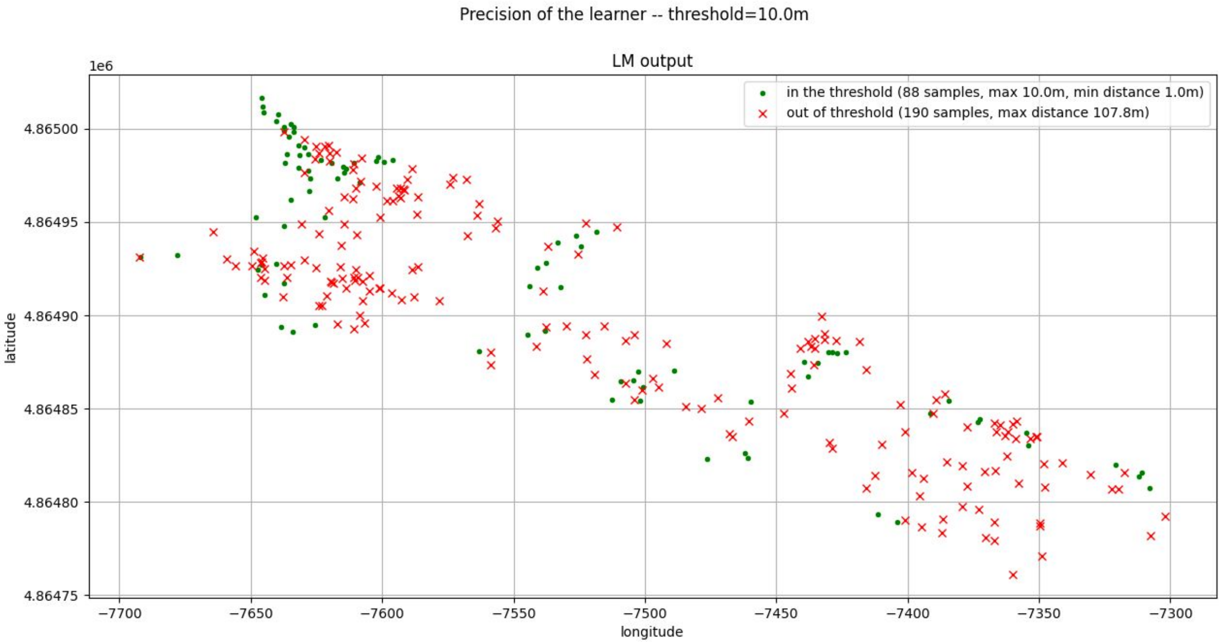
F

Results - 3b - precision in 25m



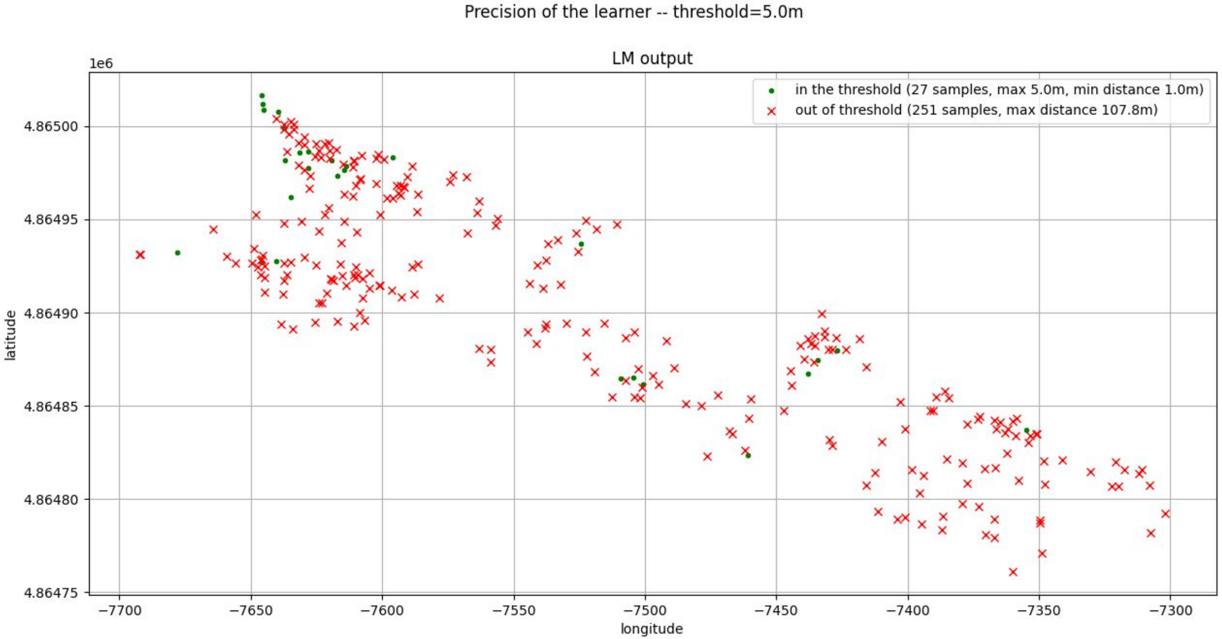
F

Results - 3c - precision in 10m



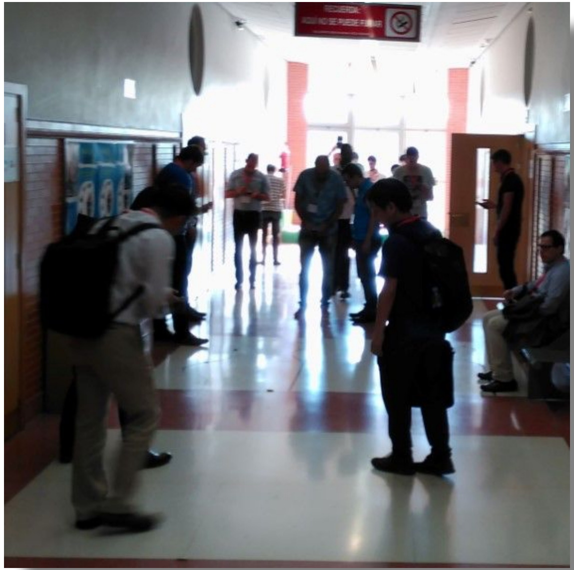
F

Results - 3d - precision in 5m



Links and Resources

- The dataset :
  - on Kaggle
  - on UCI
- The paper : on IEEE xplore
- ***The final version of the project***
  - ***on GitHub***
- Colab Notebooks (Experimental):
  - very first version
  - training Longitude, training Latitude
  - data analysis



Course in **Machine Learning 2** -- A.A. 2021/2022

UJI Indoor Localization

Compare WiFi fingerprinting indoor localization algorithms

Francesco Ganci - 4143910
Federico Civetta - 4194543