# Open research data: Sensor Data for ML based Indoor Positioning

Marcus Venzke (venzke@tuhh.de), Ahmad Nadeem Saigol, and
Nisal Hemadasa Manikku Badu
Institute of Telematics, Hamburg University of Technology
6.5.2024

**Abstract**

This data package contains open research data intended for machine learning from the WinOSens project. It is sensor data that was recorded by a sensor module attached to a sack truck pushed on the same path many times. It was meant to train ML models for recognizing the current position on the path. Sensor values are available for acceleration, gyration, magnetic flux density, air pressure, temperature, humidity, and the brightness of different colors. The current location of the sack truck is available as labels annotated by humans. The data is provided for 3 paths that cover indoor and outdoor areas.

## 1 Introduction

The project WinOSens[1] has investigated how to enable indoor positioning for transport boxes in factory environments using sensors and machine learning (ML) without an infrastructure for positioning. Each transport box is assumed to have a battery powered sensor module measuring physical values such as acceleration, gyration, magnetic flux density, and air pressure. ML models of different types were trained to recognize the current location on the path. The results achieved for supervised learning in the project were published in [Hemadasa 23] and [Hemadasa 24] also describing details on the models used. The WinOSens project was sponsored by the German Federal Ministry of Education and Research.

This data package contains sensor data that was recorded for the WinOSens project. It was recorded with a sensor module attached to a sack truck that was pushed repeatedly on the same paths on the campus of the Hamburg University of Technology. During these runs, log files were written containing acceleration, gyration, magnetic flux density, air pressure, temperature, humidity, brightness of different colors, and the current location on the path. The current location was manually entered by a human during the run. Sensor module and sack truck are shown in the YouTube video [Venzke 24] from timestamp 15:49 with comments in German language. Note that this data package only contains data from runs for which the WinOSens project results do not suggest that they might be faulty.

[1] The acronym WinOSens stands for the German project name „**W**artungs- und **in**frastrukturarme **O**bjektlokalisierung zur Steigerung von Effizienz und Transparenz in industriellen Logistikprozessen mithilfe des maschinellen Lernens in eingebetteten **Sens**orsystemen”.

The current locations are given as numbers identifying a particular section of the path. For that purpose, each path was divided into sections that were numbered with integers from 0. An example is given in Figure 1 for path 2. Its 11 sections are shown as arrows numbered from 0 to 10. The next higher number means, that the run has ended. It is 11 for path 2. -1 means, that the run has not yet started. The goal of the ML models for positioning is to determine the number of the current section from the sensor data during the run.

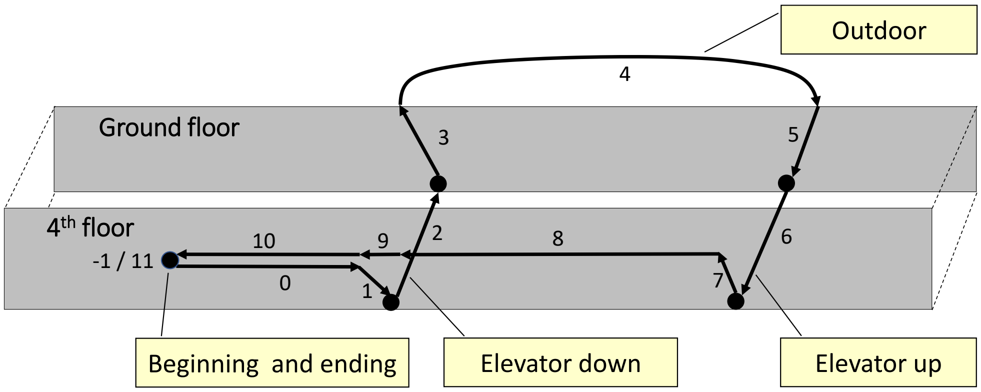


*Figure 1: Sketch of path 2*

## 2 Paths

The data package contains data for three paths named path 2, path 3, and path 4. Path 1 was omitted for quality reasons. Each path contains elevator rides and covers indoor and outdoor areas. In buildings, doors are passed through by turning the sack truck backwards, pathing through, and turning it forward again. Note that path 2 and 3 were recorded with a different sensor module than path 4.

A sketch of path 2 is shown in Figure 1. The path starts at the fourth floor of a building. A corridor leads to an elevator that is used to get to the ground floor. There the path leaves the building and continues to another entrance of the same building. Another elevator leads back to fourth floor. Here, the path follows the corridor back to the starting point. A run of path 2 is shown in the YouTube video [Venzke 24] from timestamp 17:30 including comments in German language.

As illustrated in Figure 2, a significant part of path 3 is outdoor. Path 3 starts in the same corridor as path 2 and follows the elevator down. After leaving the building it follows a ramp upwards, a cobblestone way, and a bridge, after which the sack truck is pushed down a staircase. The path leads back to the entrance, the elevator, and through the corridor to the starting point.

Path 4, illustrated in Figure 3, contains two breaks of a random time between 3 and 8 minutes. The first four sections are the same as for path 2 and 3. After leaving the building, there is the first break in front of another building. Afterwards it leads into the other building, through some corridors, and to an elevator leading the fourth floor. After crossing another corridor there is the second break. Afterwards the path leads back along the same route as on the way there, however, skipping the first break.

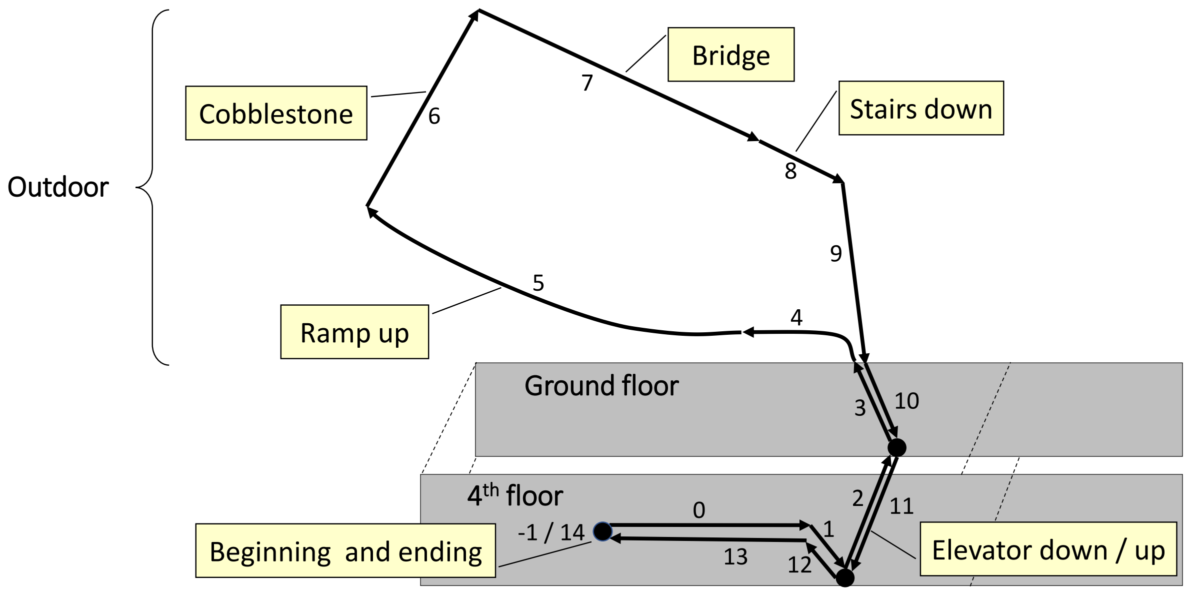


*Figure 2: Sketch of path 3*

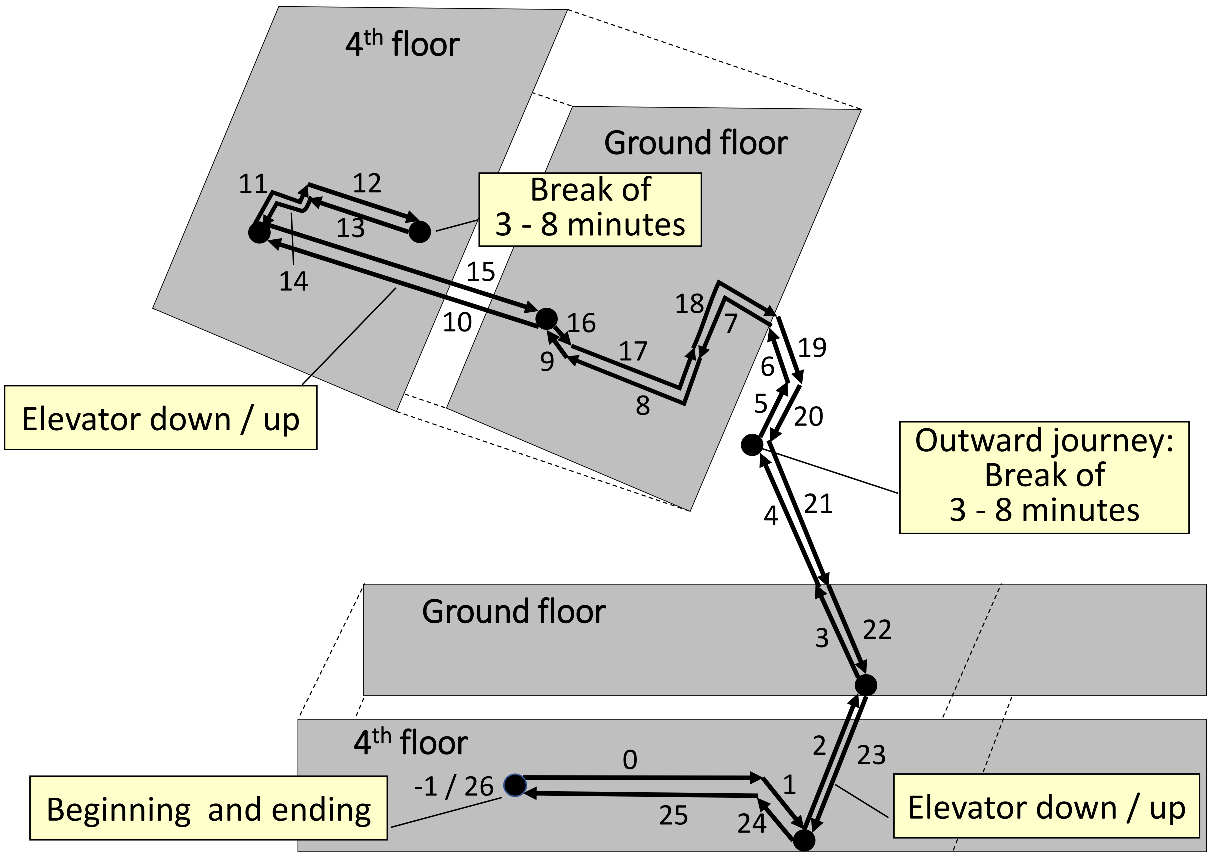


*Figure 3: Sketch of path 4*

# 3 Data package structure and log file format

The directory structure of the data package is shown in Figure 4. There is one directory for each path (`Path2`, `Path3`, and `Path4`). These have subdirectories named `Logs` containing one log file for each run. Log files are named with a number identifying the run (3 digits), the date of the run as year (4 digits), month (2 digits), and day (2 digits), and the time when the run started as hour, minute, and second (2 digits each). An example is `089_20220310_114318.csv`. For each run there is a notes file describing the weather and special occurrences during the run. It is named as the log file but ends with `_notes.txt`.

```
Data package
         ├──►Path2
         │      ├──►Logs
         │      └──►Logs_paths_with_anomalies
         ├──►Path3
         │      └──►Logs
         └──►Path4
                ├──►Logs
                └──►Logs_with_elevator_stopped_at_a_floor
```

*Figure 4: Directory structure of data package*

For path 2 and 4, the data package contains runs with anomalies in separate subdirectories. Each anomaly is described in the notes file of the run. For path 2 it is anomalies of different types that were performed intentionally during that run. The log files can be found in subdirectory `Logs_paths_with_anomalies`. For path 4 it consists of runs in which one of the elevators stopped on a floor between ground floor and fourth floor. These log files are stored in subdirectory `Logs_with_elevator_stopped_at_a_floor`.

The log file of a run is a comma separated file having the extension .csv. It contains records (i.e. lines of text) of three different types in temporal order. The record type can be determined from the first column that is a character. Possible values are:

- M: Record with data of MEMS sensors (accelerometer, gyroscope, magnetometer).
- E: Record with data of environmental sensors (all other sensors).
- L: Label with a location annotated by human.

## *3.1 M: Record with data of MEMS sensors*

MEMS data is measurements of acceleration, gyration and magnetic flux. It is 3D vectors provided as three components x, y, and z.

For path 2 and 3, MEMS data was recorded at a varying sampling rate of 24 Hz on average. It was acquired with the sensor TDK InvenSense ICM-20948 containing accelerometer, gyroscope, and magnetometer.

For path 4, MEMS data was recorded at a sampling rate of 25 Hz. The data was acquired with the two sensors STMicroelectronics LSM6DSOX (accelerometer, gyroscope) and STMicroelectronics LIS2MDL (magnetometer).

The columns in each row are:

0. Character "M"
1. Timestamp of measurement in seconds with 3 fraction digits.
2. Acceleration x in $m/s^2$
3. Acceleration y in $m/s^2$
4. Acceleration z in $m/s^2$
5. Gyration x in radian/s
6. Gyration y in radian/s
7. Gyration z in radian/s
8. Magnetic flux density x in micro tesla
9. Magnetic flux density y in micro tesla
10. Magnetic flux density z in micro tesla

Example record:

M,199.000,-0.107,0.390,9.973,0.003,-0.003,-0.003,57.150,-116.849,-58.950

### *3.2 E: Record with data of environmental sensors*

Environmental data is measurements of brightness, air pressure, temperature, and relative humidity.

For path 2 and 3, environmental data was recorded at a sampling rate of about 1/7 Hz (one sample about every 7 seconds). The data was acquired with the two sensors Bosch SensorTec BME688 (air pressure, temperature, and humidity), and AMS AS7341 (brightness of light colors). Note that the value for brightness of light (column 2) is always 0.0 as there was no brightness sensor in the data logger used for these paths. Brightness of clear light (column 14) could be used instead.

For path 4, environmental data was recorded at a sampling rate of 1 Hz. The data was acquired with the four sensors Texas Instruments OPT3004 (brightness of light), Bosch SensorTec BMP390 (air pressure), Sensirion SHT41 (temperature and humidity), and AMS AS7341 (brightness of light colors).

The columns in each row are:

0. Character "E"
1. Timestamp of measurement in seconds with 3 fraction digits.
2. Brightness of light [unit unknown] (Always 0.0 for path 2 und 3.)
3. Air pressure in hPa
4. Temperature in °C (in sensor module case)

5. Relative humidity in % (in sensor module case)
6. Brightness of color cyan [unit unknown]
7. Brightness of color red [unit unknown]
8. Brightness of color orange [unit unknown]
9. Brightness of color yellow [unit unknown]
10. Brightness of color green [unit unknown]
11. Brightness of color blue [unit unknown]
12. Brightness of color indigo [unit unknown]
13. Brightness of color violet [unit unknown]
14. Brightness of clear light [unit unknown]
15. Brightness of near infrared [unit unknown]

Example record:

E,199.000,141273.59,999.35,23.05,42.02,149,277,373,484,515,528,713,798,1000,1000

### *3.3 L: Label with a location annotated by human*

A label denotes a fixed location that is a section of the path. Labels were added, whenever a new location was reached. For that purpose, a human pressed a button on a smart phone (path 2 and 3) or tablet PC (path 4). Labels for locations range from 0 to 11 for path 2, from 0 to 14 for path 3 and from 0 to 26 for path 4.

The columns in each row are:

0. Character "L"
1. Label: An integer identifying the location

Example record:

L,6

The sensor data was acquired in the WinOSens project.

SPONSORED BY THE

Federal Ministry of Education and Research

## 4 License

The data package including this documentation is published under the Creative Commons license CC BY (version 4.0), as described at https://creativecommons.org/licenses/by/4.0/.

## References

[Hemadasa 23] Nisal Hemadasa Manikku Badu, Marcus Venzke, Volker Turau und Yanqiu Huang. Machine Learning-based Positioning using Multivariate Time Series Classification for Factory Environments. Technical Report Report arXiv:2308.11670, arXiv.org e-Print Archive - Computing Research Repository (CoRR), Cornell University, August 2023. https://doi.org/10.48550/arXiv.2308.11670

[Hemadasa 24] Nisal Hemadasa, Marcus Venzke, Volker Turau und Yanqiu Huang. Machine Learning-based Positioning using Multivariate Time Series Classification for Factory Environments. In Proceedings of OkIP International Conference on Automated and Intelligent Systems, CAIS 2023, OkIP Books, Oklahoma, USA, 2024. http://doi.org/10.55432/978-1-6692-0005-5_9

[Venzke 24] Marcus Venzke: Maschinelles Lernen für Mikrocontroller für die Indoor-Ortung in Produktionsprozessen im Projekt WinOSens. Video. Channel of TU-Hamburg, YouTube, February 2024. https://www.youtube.com/watch?v=AHIMf4QN7-I