Deliverable D4.5

“Vulnerability of sensors in the field”

J Jacob Wikner, Jan Ketil Rød, Radmil Popovic

1 Document purpose

The original short description of the deliverable as outlined in the proposal is:

> We will investigate the vulnerability of the sensors and eventual errors related to eventual data manipulation or tampering with sensors by doing random tests at the pilots where we contrast and compare data from secondary sources (texts and images) against sensor data.

Also here we will deviate slightly from the initial description in the work package. We will mainly present a technical description and focus on typical error sources in sensors given by the experiences we have obtained from operating the sensor stations.

The vulnerability of the data that the sensors capture is determined by issues like the following:

- Loss of power, theft, damage, maintenance.
  - Batteries are not properly charged
  - Power outlet plugs are pulled
- Manipulation of data
  - Deliberate or unintentional. For example, placing the device in shadows or moving them around.
- Incorrect calibration
  - Data is offset from the actual value and might even be considered an outlier and then discarded from the data set.
- Drift
  - Accuracy of the devices degrade throughout operating time and calibration intervals are too long.

In this delivery report we will list some of the experiences we have gathered from the work and point at the complexities and findings for mainly the Norrköping and Trondheim sites. The experiences can act as a best practices list for future projects and continued experiments.

An ERA4CS project funded by JPI Climate and co-funded by the European Union, FORMAS (Sweden),
RCN (Norway), NWO (The Netherlands), and FCT (Portugal)

3 Background

The setup and approach on how to deploy the network of sensors have been explained in detail in previous deliverables from this work package. In the project we deployed sensors in different ways: in (1) Rotterdam, we did not put any effort in installing new sensor stations, as a main focus here was on the app development. In (2) Porto, we relied on the already existing Porto Digital and Netatmo sensors. In (3) Trondheim, we did the same, used the already existing Netatmo and installed more advanced, professional equipment together with cheaper, commercially available products (for the home owners) and studied how they reported data and how well they correlated. In (4) Norrköping, we tried experimental sensor stations with different interaction schemes to test how we could build and interact with a sensor, but also installed commercial Netatmo modules with rain gauges to increase the density of precipitation measurement points. Sensor installations are further described in deliverable D4.5.

5 Results

As mentioned above, we will due to the circumstances present here the compiled findings from the installation of sensors in the different sites. For Porto, we have collected data from interviews with representatives of PortoDigital. For Rotterdam, we are not addressing the issue. For Norrköping and Trondheim, we compile the results below.

Assessment of vulnerability for sensors in the field in a network with LCS must be performed in the same way as it is done with any network or project where data collection takes place in real time. This means that the complete data chain must be observed, i.e. from the sensors itself to the communication elements (gateways, routers and nodes), internet, local server and database. To get data from sensors that we can trust, we need to consider the things listed below.

5.1 Data integrity per se

Even though data is stored in publicly available databases, the information cannot be treated as constant. There could be changes to the policy of the owners of the databases. For example, the wunderground weather network introduced a fee for access to data in their networks, and vice versa, some databases have opened access for everyone from previously having a limited access. The dataset thus changes and conclusions based on it might change.

Based on the findings in D4.4 and [Ref] that LCSs might be used to align with official, research-graded data we have decided to compose a scientific paper “Comparison of sources of sensor data for urban climate monitoring” by Röd, Popovic, and Wikner. In this paper, we study the dataset from Netatmo and other sources in a certain area and find the statistical distribution of reported numbers, how often data is reported, and how well they align with the nearest official measurement station.

- R_NetAtmo_1hr_m144.gdb - full dataset sliced into 144 hourly subset. In total 35 354 recordings. Minimum temperature value 14.8 degrees Celsius, maximum temperature value 50.2 degrees Celsius.
- R_NetAtmo3std_1hr.gdb - subsamples where outliers are removed.

Push out: We have evaluated some of the data from the data collected through the app. In some of the pilots during some of the time periods, there was less data collected than anticipated.

5.2 Physical security

Physical security applies to both sensors and all other elements in the data collection chain. If anyone uninvited can access the sensor or communication devices or server, then the entire network must be considered unreliable. Unauthorized access to the sensors can damage the sensors or change the operating environment of the sensor itself so that the sensor starts measuring incorrectly, irregularly or stops measuring at all. In one of the Norrköping sites, for example, one of the sensor stations was stolen. It was within a gated area, but unlocked. The value of the sensor is not high, and it is likely to assume it was damaged by plain sabotage.

Therefore, the sensors must be placed in areas that are out of reach of passersby, and usually at a height of more than 3 meters (on streetlight poles, on the roof of the building, etc.). Communication devices, such as the LoraWan gateway, must be mounted so that there is a minimum of obstructions between gateway and sensor unit.

Further on, an installation permit should be required from the owner of the building as well as those occupying the building. For that purpose, it should be recommended to set up a maintenance agreement, either that a person at the site can overview the installation regularly, and also that the project can visit the sites when needed. In case of Covid-19 this became very obvious and due to the restriction (as some of the campaigns in Norrköping directed themselves at caretakers and elderly nursery homes). We had no access to the premises due to that reason and it would have been beneficial if a local representative would have handled that. It would also increase the awareness of the sensor and the purpose it has.

5.3 Data security

To illustrate issues with data security, we highlight the installation in Trondheim, Norway and with comments related to the Norrköping trials. Figure 1 shows an overview of the realized and tested networks. There is a set of different approaches to data transfer from the sensors through the cloud to the time series database on the local server.

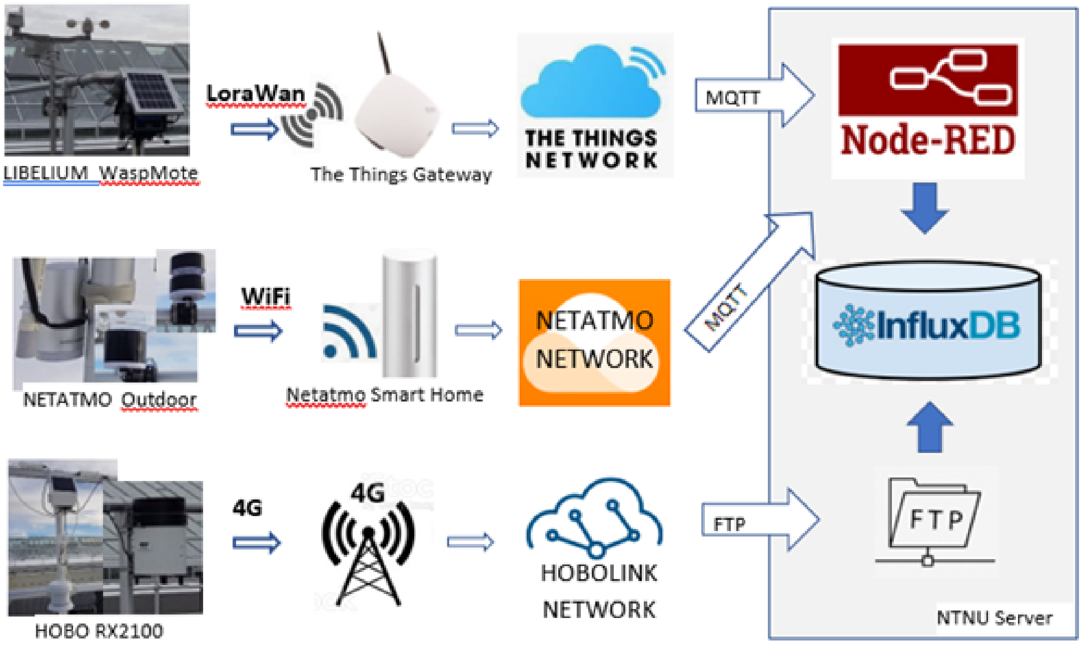

Figure 1: Overview of communication format and cloudstorage/data processing.

The different methods of communication with sensors also have different security challenges and levels of data protection. These are listed and discussed below:

Communication

- Data transfer over LoraWan[1] is mostly used in IoT projects in urban areas. It is a relatively new radio protocol and is therefore still not sufficiently standardized and is subject to frequent improvements in its security algorithms. LoraWan is the only low-power WAN protocol that allows us to easily build our own network with our own gateway. Developing our own network with the LoraWan protocol required an initial investment in equipment, but there are no major costs later during operation of the network.
- An alternative to using the LoraWan protocol is the use of the NB-IoT (narrow-band IoT) protocol,

[1] LoRa = Long Range, and Wan = WAN = wide-area network

An ERA4CS project funded by JPI Climate and co-funded by the European Union, FORMAS (Sweden),
RCN (Norway), NWO (The Netherlands), and FCT (Portugal)

which, with the introduction of 5G technology, can become the dominant communication method for data transmission in IoT networks. NB-IoT uses exclusively the mobile phone infrastructure. On the other hand, NB-IoT requires no initial investment (except for the hardware), but the subscription fee must be paid or paid according to the amount of data transferred.

- A mobile communication standard, 4G in our case, was used to transfer data from the Onset Weather Station RX2100 to HoboLink. HoboLink is Onset's own network-enabled platform for managing, controlling and visualizing sensor data from external sensors. Only owners of Onset sensors have access to the HoboLink network. Using Hobolink is free for data transfer every 30 minutes or slower. For higher communication rates, it is necessary to pay an annual subscription fee.
- WiFi is a technology used by the Netatmo sensor. (Internally, between the Netatmo main unit and the external units, a 433-MHz ISM band is used). Each owner of a Netatmo sensor has free access to the Netatmo WeatherMap network where the sensors can be registered and data from sensors visualized. The sensor owner has unlimited access to his own sensor data.
    - One problem with the wifi technology is the use of the network keys. In Norrköping, for example, not all wireless networks support the limited functionality that Netatmo offers. It is thus not possible to connect to, for example, eduroam or the city-wide wifi networks which requires a type of certificate that Netatmo is not supporting
- A combination of radio standards were tested in Norrköping and Linköping, Sweden. Some trials with Zigbee and Xbee were also performed but deemed too complicated/advanced for our purpose. Encryption is supported and mesh networks with could be used to increase reliability of a multitude of sensors installed in a certain area. This was not applicable to our case where sensor typically communicated pretty directly to the gateway/server. This is also in our recommendation as the best choice for highest reliability in terms of data communication.

Databases and data handling

- The Things Network (TTN) is a secure and open public network that supports true end-to-end encryption and supports 128-bit encryption for every single end device (sensor or actuator). Any user can create their user account in the TTN network and the use of TTN networks is free. Network access is protected by a user/password key. Each registered user can create an unlimited number of applications, with unlimited number of end devices (sensors or actuators) in each application.
    - Applications are identified by a unique Application ID. Each Application has one or more Access Keys (Device EUI, Application EUI, etc. to access application data and/or to manage

end devices

- Node-Red is a visual, flow-based development tool for connecting end devices in the IoT network, APIs and online services. Access to the data is protected by user ID and password key.
- InfluxDB is an open-source time series database. It used to quickly store and read large volumes of data in a short period of time applying time stamps. Access to the database is protected by a user ID and a password key.

In terms of access to the database and the sensor data, there are certain solutions, of various complexity, that can be used to enable access for the entire research team to the same data without sharing the access keys in a practical way. One also has to consider the license agreement, where sometimes a single user is not allowed to invite more users without them having to pay a fee too, or extended licensing schemes. It is suggested to have a plan to backup data, collect data from many sources (if licensing schemes allow such operations), and restrict the write access to data to avoid accidental flushing of data.

5.4 Energy and power supplies

The ability of the sensors' power supplies, in terms of harvesting of new energy and store that in batteries and/or access to main power outlets as well as data rate and the distance of the sensors to the communication node and to the gateway are elements that strongly affect the reliability of the data collected from the sensor.

The sensors that are installed in the field are typically forced to use batteries to operate. However, batteries have limited capacity and lifetime, and solar panels are commonly used to continuously charge the batteries.

In Trondheim we used comparatively small solar panels to power the Libelium and Onset Hobo sensors (Fig. 2). Although Trondheim is located at 63.42 degrees latitude, with very short days and little sunlight in the winter, charging batteries with small solar panels was sufficient for sensors to work reliably throughout the year.

An indicator of the amount of stored energy in a battery is its port voltage. Figure 3 shows the battery voltage level on the Hobo sensor depending on the number of sunny hours per day in Trondheim for the period with the shortest days of the year (winter).

An ERA4CS project funded by JPI Climate and co-funded by the European Union, FORMAS (Sweden), RCN (Norway), NWO (The Netherlands), and FCT (Portugal)

	The OnSet HOBO MicroRX2102 using an integrated solar panel that generates in average 1.7W for an area of 150x110 sqmm.		The Libelium WaspMote using a rigid solar panel that generates in average 3.5W for an area of 230x160 sqmm.

Figure 2: Examples of solar panels for the research-graded sensor stations (Hobo and Libelium).

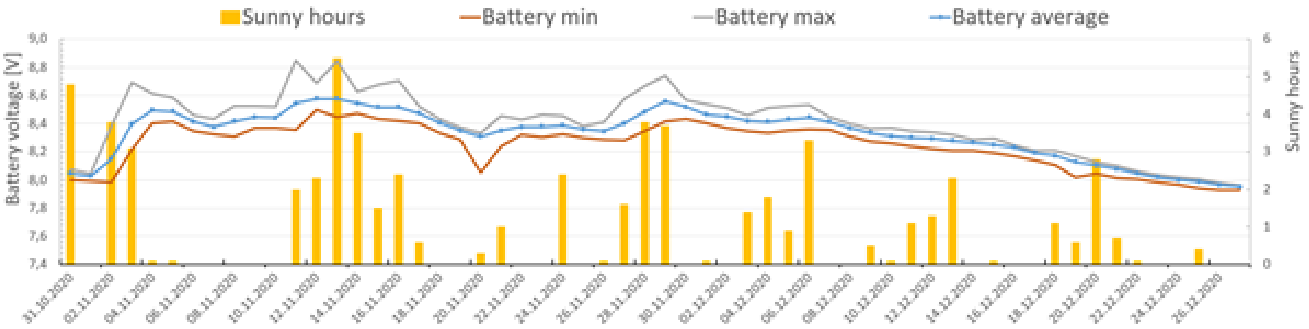

Figure 3: Example of the battery levels over time for Trondheim illustrating that even though the site is far north, the light is sufficient.

Sensors in Norrköping mainly had access to the main outlet. However, these components are sensitive towards power outage, and human errors in terms of pulling the plug. We experienced this in many cases for the Netatmo sensors in Norrköping. It suggested a robust way of controlling the access to the main unit. The issue, however, is that the main unit of the Netatmo is recording interesting data for the user and should thus not be set aside in a locked room (for example). Also here, some of the issues are overcome by having a local representative that can help us with inserting the plug again. The outdoor units of Netatmo drain their batteries comparatively fast and have to be replaced frequently (once a year). With a large set of these sensors, it becomes more cumbersome to maintain. Even though the task sounds easy, it is a component in the logistic chain that complicates maintaining a high quality power supply.

5.5 Communication range

The distance between the sensors and the gateway/node also affects the reliability of data transmission. This is determined by the physical distance per se, but also obstacles in the line-of-sight (assuming wireless

An ERA4CS project funded by JPI Climate and co-funded by the European Union, FORMAS (Sweden),
RCN (Norway), NWO (The Netherlands), and FCT (Portugal)

connection). We have experienced intermittent access to the sensors due to, for example, refurbishing of rooms, or similar, where we installed the gateways. (Typically, gateways are installed indoors to be better protected.)

- The Netatmo outdoor sensors must be installed very close to the Netatmo indoor smart home weather station. It is claimed that the distance is a maximum 100 meters without obstacles. In our tests, it was found that 100 meter was way too optimistic also for a clear line of sight. If the gateway was placed indoors, the sensor had to be within 10-20m to give a reliable signal. Thick concrete walls should be avoided.
- Hobo sensors are wired to the logger, which communicates over the mobile network with an integrated SIM card. Quality of service over the mobile network is good and high reliability is offered.
- Libelium sensors use LoraWan radio communications, and the distance from the sensor to the gateway can be up to 20 km. InTrondheim we installed our own gateway near our sensors, and thus experienced no range problems. At the same time, we tested our sensors with the LoraWan gateway, which was already installed in Trondheim city center. Although the distance to these sensors is relatively small, about 4 km, the communication was very unreliable and weather dependent. Figure 3 shows the altitude profile of terrain and distances between the citizen sensing sensor and Trondheim gateways.

It is suggested that during the installation of the sensors to be conservative with the range and consider different conditions. Weather conditions, like rain and snow, could affect the range, dependent on radio standard and tests should be done during several seasons.

Further on, it was also discovered during tests in Linköping that one of the LoRa gateways (public ones) closed and in that region we then did not receive any data. Therefore, it should be planned for your own gateway and/or overlap in the coverage of different LoRa gateways.

For some of the tests we used Bluetooth for communication between mobile phone and sensor station. Here it was found that the mobile phone (obviously) played a big role. We recommend that also here elaborate testing is required to cover many different use cases and scenarios with radio communication formats that should be “easy” to use.

An ERA4CS project funded by JPI Climate and co-funded by the European Union, FORMAS (Sweden),
RCN (Norway), NWO (The Netherlands), and FCT (Portugal)

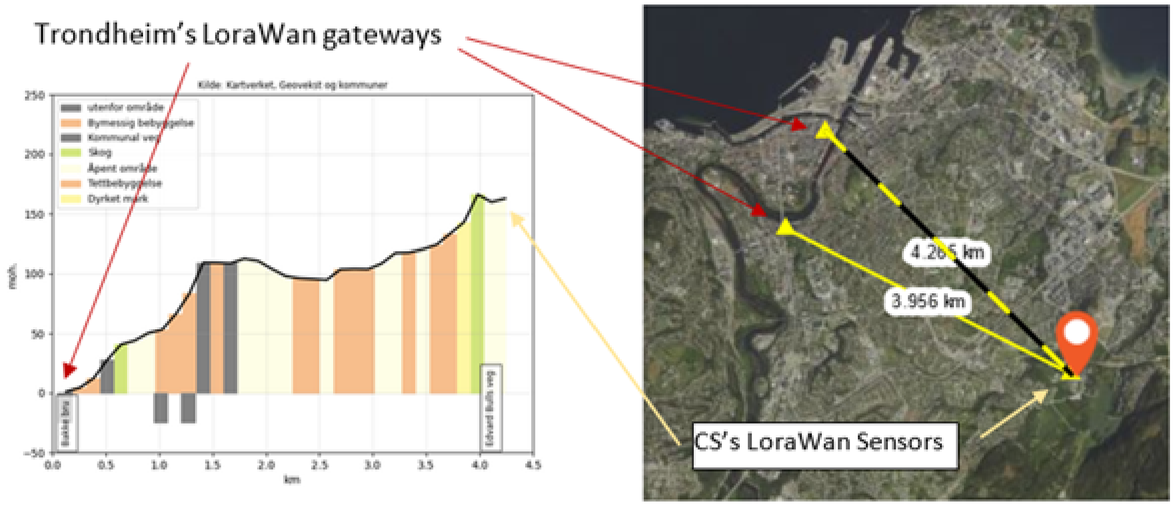

Figure 4: Distance profiles indicate that the terrain configuration prevents a good radio connection.

5.6 Locality

Mobile sensors can also be considered for the purpose that the citizen sensing project wants to fulfil. It could be sensors mounted on, for example, public transport vehicles, or bikes or similar. It could also be the user reporting data through the app that is developed within the project. The users of the app have had the opportunity to take pictures of events, such as drought, blue spots (local, flooded areas), phenomena of various kinds, or even just a snapshot of daily life illustrating for example clear, blue skies.

Regarding the user testing of the WayFinder functionality of the CitizenSensing app in Trondheim during 2020, a total of 128 (52%) out of 245 observations were made up to 20 meters away from the mapped blue spots, and a further 73 photos (30%) were taken even further away. Finally, 44 (18%) out of 245 observations were done within the mapped blue spots (e.g., Figure 5A).

Hence, most of the photos were taken at a certain distance from the objects in order to provide a view of their context, such as photo B in Figure 5. Most of the photos showed puddles of various sizes and shapes (Figure 5C), either on sidewalks or on streets, and large puddles' "traces," such as those in Figure 5A and B. Regarding observations registered under types other than "heavy rainfall," the participants observed debris or ice blocking water flows through drainage outlets. For instance, the observation in Figure 5D was registered as "snowfall," whereas the observation in Figure 5E was registered as "seasonal observation."

The three most common examples of locality issues include: (1) incorrect observation coordinates, (2)

An ERA4CS project funded by JPI Climate and co-funded by the European Union, FORMAS (Sweden),
RCN (Norway), NWO (The Netherlands), and FCT (Portugal)

overrepresentation of specific spots, and (3) “poor participation.”

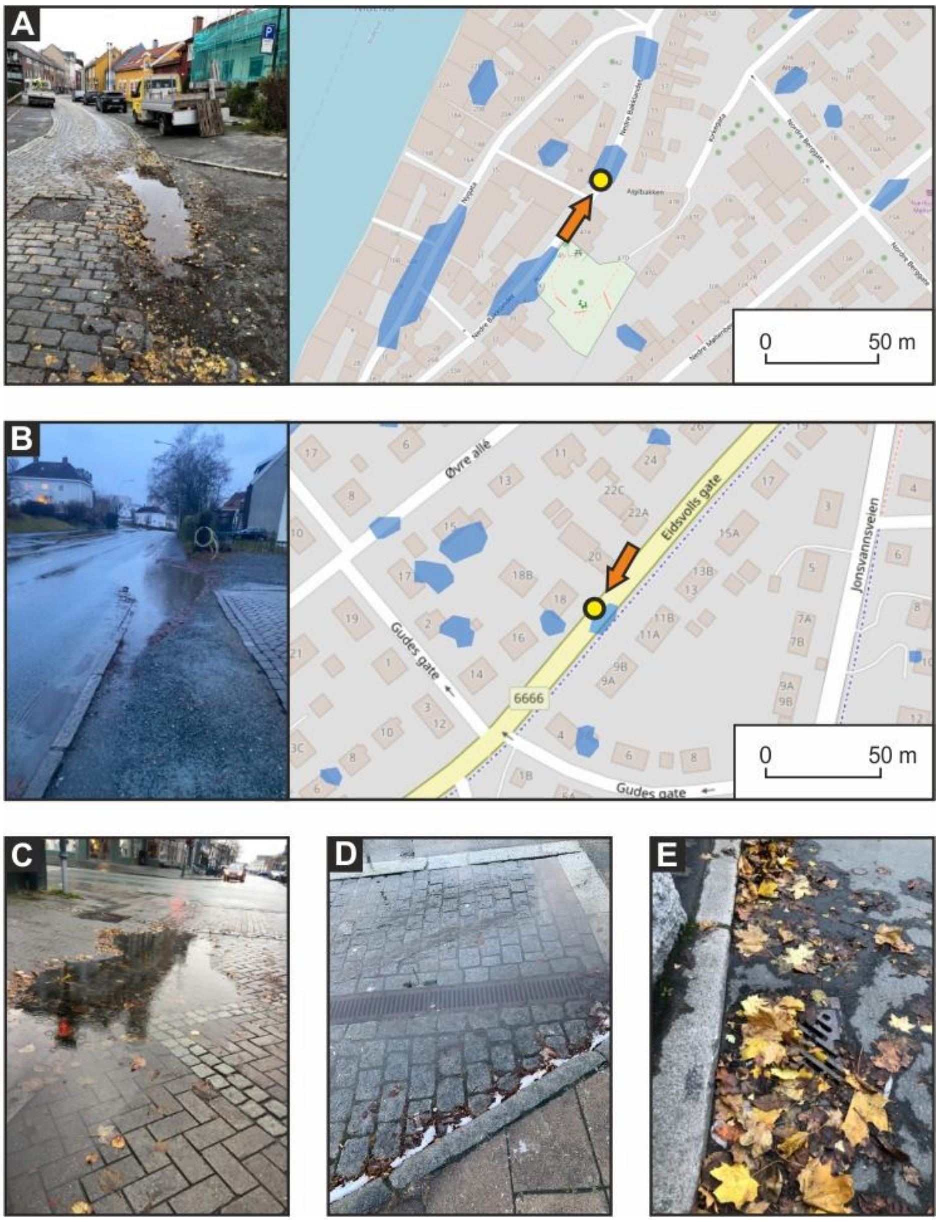

Figure 5. Examples of blue spot localities and picturesfrom Trondheim, Norway. The individual pictures (A through E) are further elaborated upon in the text.

Figure 5A exemplifies the issue of incorrect coordinates, in which observations were registered from inside of a shopping centre. After inspecting the observations’ attributes and photos in the portal, it turned out that they were registered by different users in different days and concerned the spots in front of the building.

An ERA4CS project funded by JPI Climate and co-funded by the European Union, FORMAS (Sweden), RCN (Norway), NWO (The Netherlands), and FCT (Portugal)

Thus, the incorrect geocoding probably occurred since the users registered observations after a while after taking the photos, for example when doing shopping, but forgot to ensure correct observation coordinates.

Incorrectly geocoded observations were also observed in other spots such as university buildings. The web portal has also facilitated revealing the overrepresentation issue. Since participants registered numerous reports near their housing, university buildings, when going to or back from their university classes, their observations were unequally distributed.

The optimal coverage of VGI would be a high number of observations distributed around the clock across the city, of course, depending on the events to be reported on. However, the number and spatiotemporal distribution of the data in our study was dependent on the set-up and context of the campaigns during which the data was collected, similar to the study by Hung et al. (2016). Thus, the participants privileged their neighbourhood.

Lastly, a number of participants were either too motivated, too lazy, or misunderstood the instructions. It resulted in "poor participation", in which there were too many observations registered without reason and numerous observations from the same place. For example, one participant of the 2020 campaign registered 282 observations (Figure 6B) of mostly wet sidewalks or streets with small puddles.

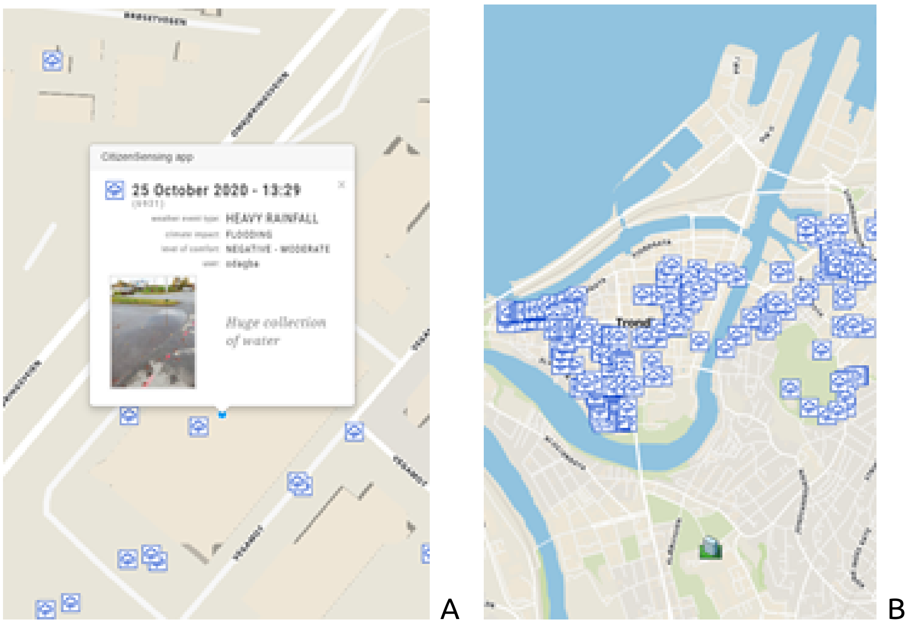

Figure 6. Common locality issues: (A) incorrect observation coordinates inside a building, and (B) "poor participation" by registering observations without reason.

An ERA4CS project funded by JPI Climate and co-funded by the European Union, FORMAS (Sweden), RCN (Norway), NWO (The Netherlands), and FCT (Portugal)

6 Conclusions and future work

In this report, we have not in detail commented on two factors that were mentioned in the original purpose of the deliverable:

- how the reported data through the app aligned with the data reported by sensors. This is a scientifically very interesting question, but also deemed too big for this deliverable and thus pushed forward to a scientific paper in which we can present a detailed study. It is highlighted as an observation to the reader of this report, and an action in the best-practices list.
- deliberate tampering of data through various means of affecting the results: stealing equipment, turning off power, reporting incorrect data in the app, actively affecting the sensor data by for example heating them or obstructing rain and wind gauges, etc. Also here, we deemed the task too big for the deliverable, as it opens up for a much larger discussion on what could be the rationale for tampering with data. In terms of security risks, there are models on how, for example, hardware equipment can be protected and levels of attacks (for example the FIPS 140 standard), countermeasures, counterattacks, etc., can be classified and handled. However, this is well out of the scope of the deliverable. In the deliverable, we have focused on the indirect tampering, or undeliberate, such as pulling plugs, reporting in the wrong location, or taking pictures of incorrect items.

7 References

[1] Hung et al. (2016)

[2] Maike & Jan

An ERA4CS project funded by JPI Climate and co-funded by the European Union, FORMAS (Sweden),
RCN (Norway), NWO (The Netherlands), and FCT (Portugal)