| Deliverable 3.4 | |
|---|---|
| **Pilots' use and performance evaluation** | |
| Lead Partner: | University of Parma - UNIPR |
| Authors: | Claudio Guerra, Dominic M. Kristaly, Niccolò Mora, Valentina Bianchi, Guido Matrella, Ilaria De Munari, Paolo Ciampolini |
| Contributors: | METEDA, VSRO, UNIPR |
| Date: | January 2017 |
| Revision: | V1.0 |
| Dissemination Level | PUBLIC |

*Project Acronym:* HELICOPTER
*Project full title:* HEalthy LIfe support through COmPrehensive Tracking of individual and Environmental Behaviors
*AAL project number:* AAL-2012-5-150

# TABLE OF CONTENT

# 1. Introduction

The main idea of the Helicopter project is the design and the test of an AAL-System oriented to the monitoring of users' behaviour and aimed at generating specific automatic triage functions. For these purposes, innovative algorithms and technologies have been devised and developed.

Experimental activity is the core of the whole Helicopter project and a great effort, in terms of production of the systems, installation and maintenance, was required to implement about 30 installation sites, split in two different counties: Sweden and the Netherlands.

The purpose of this document is to present the activities related to the installation of the pilot sites and an analysis of technical performances obtained by the Helicopter AAL-System. Such system is the result of the integration of several technologies, both hardware and software: the basic infrastructure consists in the home environmental sensors network (developed by UNIPR), and other components of the system are the clinical devices network (managed by METEDA), the "snowflake" poster (realized by CIID) and the APP for monitoring the state of the system (developed by VSRO).

In Section 2 is offered a technical description of the activities relating to the organization of the pilot, the architecture of the Helicopter AAL-System, the procedures activated for the installation (performed in two phases) and the maintenance, as well as the technical developments carried-on between the first and the second round of installations.

In Section 3 a brief presentation of the data produced by the pilot sites is showed. Some important parameters to evaluate the quality of the experimental activity are discussed, for example: a qualitative analysis of raw-data to show the feasibility of the adopted approach, and a more quantitative analysis to gather insights on users' behaviour.

A technical evaluation of the overall functioning of the pilots is the topic of section 4.

Eventually, some brief conclusions, to summarize the activities and take an overall evaluation on the experience of testing, close this document.

# 2. Technical description of pilots

In the next Chapter, the technical perspective of the pilot deployment will be presented. In particular, the recruit of the users, the overall architecture of the system, the installation phase and the development carried-on will be described further in details.

## 2.1 Pilot organization

Overall, 22 pilot sites were recruited in the Netherlands and 8 in Sweden, for a total of 30 pilot sites; among these, 6 in the Netherlands and 2 in Sweden had two users (husband and wife): in those pilots each person was given his own wearable sensor, while environmental sensors (except for the bed sensors) were common to both.

The users were healthy (none of them suffered from one of the eight diseases being monitored, see D4.1) elderly adults who were still able to live in their own home with no need of assistance by formal caregivers. The technical requirements for the recruitment were the house dimension (a small apartment\house with a maximum of two floors, but preferably on a single floor) and especially a broadband internet connection, with possibility to access the router from remote. Unfortunately, not always it has been possible to meet these requirements, hence the large house dimensions and the lack of a remote connection had an impact on the reliability of the wireless network, and on the collection of the data.

## 2.2 Architecture overview

The Helicopter AAL-System architecture was composed by the µserver, the environmental sensors, the wearable sensor (called MuSA [1]), the clinical sensors and the "snowflake" poster. Furthermore, a Lenovo 10" Tablet (running an APP that allows to know the state of the system sensors) completes the System.

Below, a brief description of the part of the system:

- **µserver**
  A Lenovo Thinkcentre M53, equipped with suitable radio transceivers (ZigBee and Bluetooth). The PC ran a supervision process, which took care of several functions, besides managing the actual sensor communication and being responsible for the communication with the cloud. The PC had to be accessible from remote for technical assistance purposes and for allowing quick data collection.
- **Wearable sensor**
  The wireless sensor platform MuSA was exploited, that allows to request assistance by pressing the help button and to automatically detect a fall. In the context of the Helicopter project, MuSA plays a key role because is the mobile device that enables the identification function in the pilot in which there is more than a person living in the home.
- **Environmental sensors**
  The use of sensors of many different kinds and functions was planned, in order to feed the behavioural model; both commercial devices and (whenever a more specific function is needed) purposely designed devices were used. In both cases, the IEEE 802.15.4/ZigBee wireless transmission protocol was adopted. Sensors have been selected based on user-related features, i.e., the need of actual user awareness in sensor management; installation requirements and intrusiveness were evaluated as well. Environmental sensors available for

exploitation in the HELICOPTER framework are described more in detail in Deliverable 3.3, and included:

- Door/drawer sensor, exploiting magnetic contact, useful for detecting behavioural patterns related to, for example, opening the food cabinet, indirectly relating to the food intake frequency and time
- Fridge sensor, providing information about fridge opening and internal temperature and humidity
- Bed/chair occupancy sensors, based on sensitive pads of different shapes and size, coupled to a wireless transceiver. Provided information about sleep patterns and other daily living habits.
- Toilet sensor, based on proximity sensors, monitoring the toilet usage frequency and time distribution.
- "Clinical push-button" sensor, used to identify which user is going to use any of the clinical devices listed below.

As previously mentioned, an innovative feature of the Helicopter project is the identification function: in multi-users pilot, it was necessary to identify the user who triggered the given environmental device (e.g. if the fridge sensor signalled, we had to understand which user opened or closed the fridge door). This was possible by exploiting the communication between the environmental and the wearable device, to find out which user was the closest to the environmental device when it signalled, hence which user was more likely to have triggered it. The identification feature is explained more in detail in Deliverable 3.1, and was extensively used in those devices accessible by all the users in the house, namely fridge, toilet, door\drawer, chair and clinical push-button sensors; the bed sensor was instead used without the identification feature, since usually both husband and wife sleep on "their own" side of the bed, that remains the same through every night.

- **Clinical sensors**

  Clinical sensors were exploited for the (self) assessment of physiological parameters: all involved sensors were easy to use, suitable for being used by the user himself (or by untrained relatives). The HELICOPTER system automatically managed data logging and transmission over the system infrastructure, i.e., no additional burden (with respect to customary device use) was placed on the user for dealing with system communication. A limited number of clinical sensors was used, hence they did not needed to be scattered over the whole home area, and thus did not need to cover large transmission paths; therefore the mainstream, commercial approach was followed, by using standard Bluetooth communication technology. Clinical devices are described more in detail in deliverable 3.3, and included:

  - Body Weight scale (A&D Medical UC-321PBT)
  - Blood Pressure Monitor (A&D Medical UA-767PBT)
  - Pulse Oximeter (SAT 300 BT fingertip device, Contec Medical Devices Co. Ltd)
  - Glucometer (FORA G31b, Fora Care Inc.)
  - Portable ECG (TD-2202B, TaiDoc Technology Corporation).

It is worth stressing that, since the system is compliant with the standard Bluetooth protocol and an open-approach has been used, it can communicate with any commercial device, and no specific assumption is made regarding brand or sensor type. Therefore, it would be very easy to add other clinical sensors to the system, because the only requirement would be to introduce a suitable descriptors in the database structure. Although the above device were included in the design and lab test procedures, only the body weight scale and the blood pressure monitor were generally included in actual pilots, with remaining ones possibly to be included upon specific conditions. This came from discussion with users and caregivers: for instance, people already acquainted with a specific model of portable glucometer wouldn't like to shift to a different model for such a short time. Due to different regulations in pilot countries, we were not allowed to store clinical data related to Swedish users: nevertheless, they were given the very same devices, to preserve the overall user experience, with pilot support teams possibly involved in managing related data, if needed.

- **Supervision infrastructure**
  To ensure a continuous collection of the sensors' data, for the local micro-system (the *HELICOPTER Home System*), software applications were created to ensure the wellbeing of the system:

  - *System self check*: responsible for the good functionality of the systems. It has to verify if the sensors work and transmit data to the desktop, verify if the batteries of the sensors are running low and transmit mails to specific technical persons with the sensors status.
  - *IP collector service*: transmits periodically to the Main Server the external IP addresses of each pilot site, so the micro-systems can be accessed remotely for periodical maintenance tasks or debugging.
  - *System initialization*: ensures that each time the micro-system is started/restarted it will have the same settings and the offered the same features.
  - *Automatic updates*: allows the remote or on-site update the software components of the micro-server; the update packages containing the instructions and software modules can be installed on the mini-PCs either by using an USB stick or uploading it in a pre-determined directory on the main server. The automatic update software component checks periodically for update packages and, if found, will copy them to the mini-PC's storage and, by restarting the machine, will install them.
  - *Data collecting and processing*: the data collected from the sensors are incrementally transferred to the Helicopter main server, where the „automatic triage" system can analyze it to infer the state of the monitored diagnostic suspicions.
  - *Logging service*: all system elements generate log information about their status, information that was transferred periodically to the main server. The information is related to the intrinsic functioning of the system, so it can be used when debugging problems.

The users were able to interact with the system using a tablet PC (Android-based *Lenovo Yoga 2* tablet). The most important information that the users received from the system, through the user interface, is about their wellbeing. The users were also be able to check the status of the sensors and sensors' batteries. The access to the user interface was granted using the username-password pair.

Two tablet applications were made available to the pilot site users:

- *HELICOPTER V1* tablet application: contained the features for selecting the colours of the sensor covers and defining the network of care (filling in the identity and contact information of the formal/informal caregivers); also, the users could get information about what type of sensors they were getting and what aspect of their lives they will monitor. Based on the feedback received from the partners and from pilot sites, several version of the application were released. The latest version is 1.8. From this application, the user had the possibility to upgrade to the second application (*V2*).
- *HELICOPTER V2* tablet application: contained all the features to get feedback from the HELICOPTER system. The application was design to be modular, so the features could be released incrementally (on one side not to overwhelm the inexperienced users, on the other to be able to speed up the release of improved version of the application). Several versions of the application were released, the latest being 2.9.

An automatic update mechanism was included in both application, but in a few pilot sites the security policy of the ISP (*Internet Service Provider*) interfered with the system and in some cases the different versions of Android made it possible/feasible only the manual update. However, links where made available to the pilot sites teams, and the users, to all the available versions of the applications.

A mobile application was made available to the caregivers, so they can track the evolution of the wellbeing of the pilot users. The application is compatible with Android-based terminals and the access is granted based on a token that is linked to a particular pilot user. Two major versions were released during the project's lifespan; no action was necessary from the users to update from a version to another.

All three mobile applications were available in 3 languages: English, Dutch and Swedish. A sample screenshot of the end-user app is given in Fig. 2.1 below.


**Figure 2.1: Sample screenshot of the end-user app**

The HELICOPTER main server was mainly in charge of collecting all the sensors data, managing the software updates for the home systems, offering web services for the user interface and running the „automatic triage" component (*HeliBrain*).

The main server also controlled the "snowflake" poster, by turning on and off the lights corresponding to each sensor. The lights were related to the messages sent to the users by the system; when the user read all the messages related to a particular sensor, the corresponding light would turn off. The control of the poster was done by communicating with the *particle.io* cloud services (REST calls).

By using probabilistic-based inference, built upon Bayesian Beliefs Networks and an anomaly detection algorithm developed by HIS, several health-related parameters (outliers) were computed on a daily basis for each participant in the pilots. Based on these parameters, a few diagnostic suspicions were evaluated and events triggered depending on the evaluation result. If the situation required, the user was prompted to fill in a short questionnaire that helped the "automatic triage" system to decide if the detection was a false positive. The data flow is as follows:

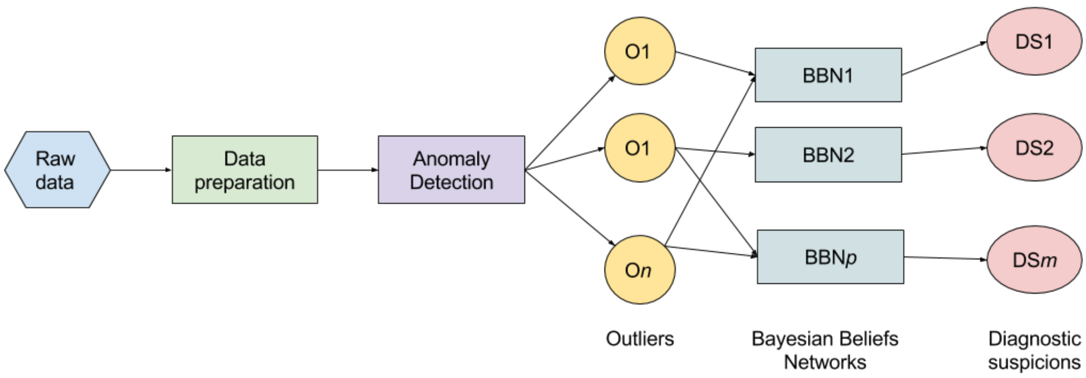



**Figure 2.2: Dataflow for the "automatic triage" (the HeliBrain component)**

The data gathered from the sensors is processed, so it matches the requirements of the anomaly detection algorithm, which generates truth values for a set of outliers. The Bayesian beliefs networks generate confidence intervals for each diagnostic suspicion using the computed outliers.

All raw and computed data was made available, through web-services (using SOAP messages), to the professional caregivers' application.

- **UI monitoring:** a further, innovative, component of the HELICOPTER vision consists of the introduction of specific tools for assessing quality and quantity of the user's interaction with system, by measuring it's interaction with the interface devices. This was initially thought as a way to quantify and monitoring interaction, as an additional indicator of the system usage. It was then considered in a much wider perspective: the idea is that such interaction could inherently carry behavioral information suitable to enter diagnostic suspicion models. More generally, some indicators could be devised (e.g., the number of mistaken operations, the time needed to perform basic interaction tasks, such as login, menu navigating, etc.) which could enter the behavioral profile, and provide valuable informations, with particular reference to cognitive decline issues. To this purpose, tools for detailed interaction of end-users with the HELICOPTER app were introduced, exploiting software environments originally conceived for e-learning applications. In particular, the "Experience API" (xAPI, [2]) software specifications were adopted. Basically, a number of specific interface actions were logged through xAPI references. A sample of such interaction set is shown in Fig. 2.3

-

| StatementView | | | StatementValidation | | | | TreeJsonViewer | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| | | | Statement | | | | | | | |
| | | User Interaction | I | Did | This | | | | stmJson | |
| | | | Actor (user) | Verb | Object | | result | context | | |
| O | L | | | | activity | activityType | | | | O |
| 1 | | user opens application | Unknown | opened | HelicopterWireframeV1 | application | yes | yes | {"actor":{"obje | 1 |
| 7 | | user close application | Unknown\|known | closed | HelicopterWireframeV1 | application | yes | yes | {"actor":{"obje | 7 |
| 2 | c | user starts creating an account | Unknown\|known | attempted | HelicopterRegistration | task | no | yes | {"actor":{"obje | 3 |
| 3 | c | user fills in the data about himself | Unknown\|known | edited | HelicopterRegistrationData | file | yes | yes | {"actor":{"obje | 4 |
| 4 | c | user saves a new account | Unknown\|known | registered | HelicopterRegistration | file | yes | yes | {"actor":{"obje | 5 |
| 5 | c | user is logged in | Unknown\|known | logged-in | HelicopterWireframeV1 | application | yes | yes | {"actor":{"obje | 2 |
| 6 | c | user is logged out | known | logged-out | HelicopterWireframeV1 | application | yes | yes | {"actor":{"obje | 6 |
| | c | User invites a caregiver: | known | invited | network of care group | na | no | yes | {"actor":{"obje | - |
| | c | user starts colors association | known | attempted | HelicopterColorAssociation | task | | | | |
| | c | user selects colors for sensors | known | assigned | HelicopterSensorClinical&Wea | organization | | | | |
| | c | user completes colors association | known | completed | HelicopterColorAssociation | task | | | | |
| | | user begins the Helicopter journey | known | started | Helicopter journey | task | | | | |
| | | user unlocks version 2 of the tablet application using the code on the sensor kit box | Unknown\|known | unlocked | HelicopterWireframeV2 | application | | | | |
| | | user switches between nigth and day activity | known | focused | HelicopterNigth&DayActivity | task | | | | |
| | c | User comments sensor data | known | edited | HelicopterSensorData | file | | | | |

Figure 2.3: set of xAPI-logged interactions

As the user interacts with the tablet app, interactions are logged on a cloud-based xAPI repository, from which they can be retrieved for evaluation and linked to data-analytics sections. Such an approach, which was not initially foreseen in the project work description, was considered as an interesting option and therefore experimentally implemented. Although in this case too the pilot size and duration is not sufficient for a thorough evaluation, the system proved to be fully functional, providing a promising insight and being suitable for extension to meny further different applications. In practice, whenever a tablet/smartphone/computer-based interaction comes into play, an expressive "behavioral" y sensor can be virtually obtained by such a tracking strategy.

- **Snowflake poster:** The snowflake poster, shown in fig 2.1, is a device developed by CID, designed to give a feedback to the user about the state of the environmental and wearable sensors. It features six LED lights of different colours, each of them representing a sensor deployed in the house: when a given light is turned on, the corresponding sensor needs assistance, e.g. the batteries must be changed or the device has gone offline. Also, the snowflake communicates in a soft and not alarming way that a sensor/s might have detected anomalies and that the user is invited to check the tablet app for more information. When the system is installed at the user's house, each environmental and wearable device is enclosed in a coloured box, and the correspondence between the LED and the device is obtained. Where two users were involved, a different poster was given to each of them. The snowflake must be connected to the wireless network of the house, because it must communicate with the server to obtain the information about the status of the devices.

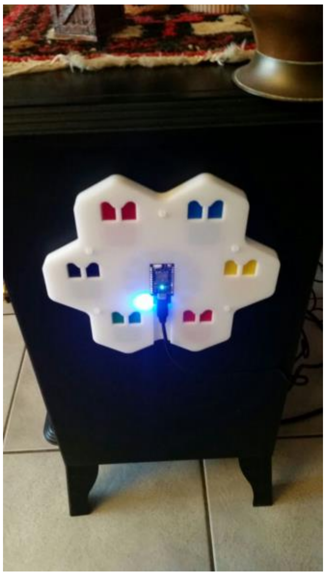


Figure 2.4: an example of snowflake poster installed in a user's home

## 2.3 Installation and maintenance activities

The pilot deployment started in November 2015. Both Swedish and Dutch pilot teams recruited a technician to handle the system installation. A friendly user interface, shown in Fig. 2.5, was developed by UNIPR, to guide the technicians during the installation of the ZigBee network (see D3.3). Using this interface it was possible to control in real time the status of every device in the network and to test its operation. Network configuration was carried out automatically, with each sensor registering to the ZigBee network when first turned on, and automatically entering the home ontological description.

A technician from UNIPR instructed the technicians on-site, and personally supervised the first couple of installations both in Sweden and in the Netherlands. Hereafter, during each installation remote assistance was provided, to help solving potential issues that could arise during the installations.

As explained in the paragraph below, a bug in the ZigBee stack provided by Texas Instruments caused a second trip to Sweden and The Netherlands by UNIPR technicians, to reprogram the faulty devices. In the following weeks, some improvements were carried-on in the MuSA firmware, to introduce new features and improve the quality of the data being collected (see next paragraph).

After installation, a web-based dashboard, shown in Fig. 2.6 and ran by VSRO, was made available to pilot support teams and technicians, allowing for continuous assessment of pilot status, and for recognizing needs of maintenance.

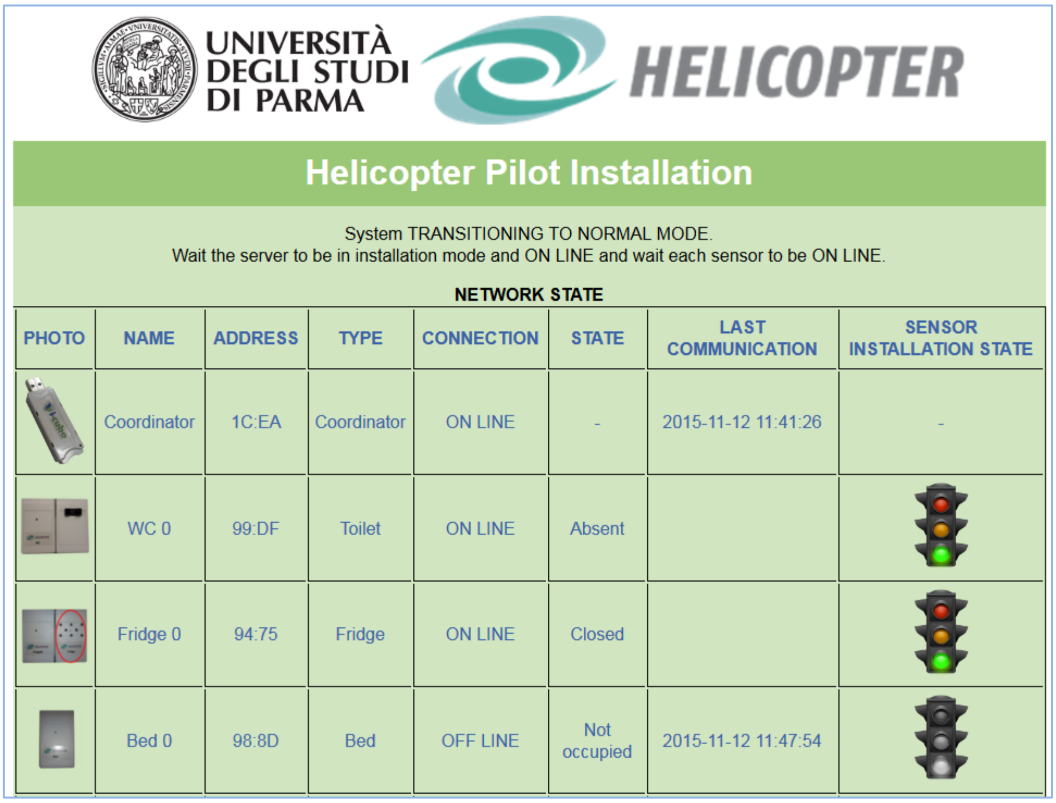



**Figure 2.5: sample screenshot of the wireless sensor network installation tool**

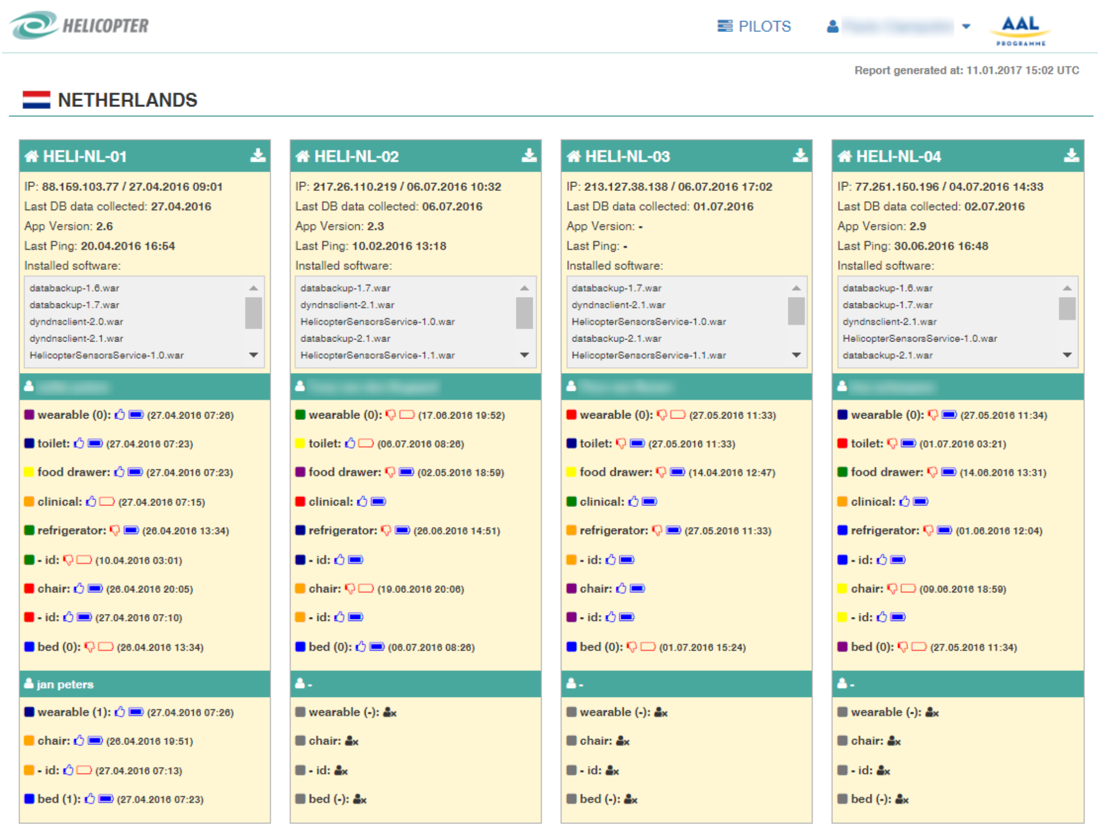
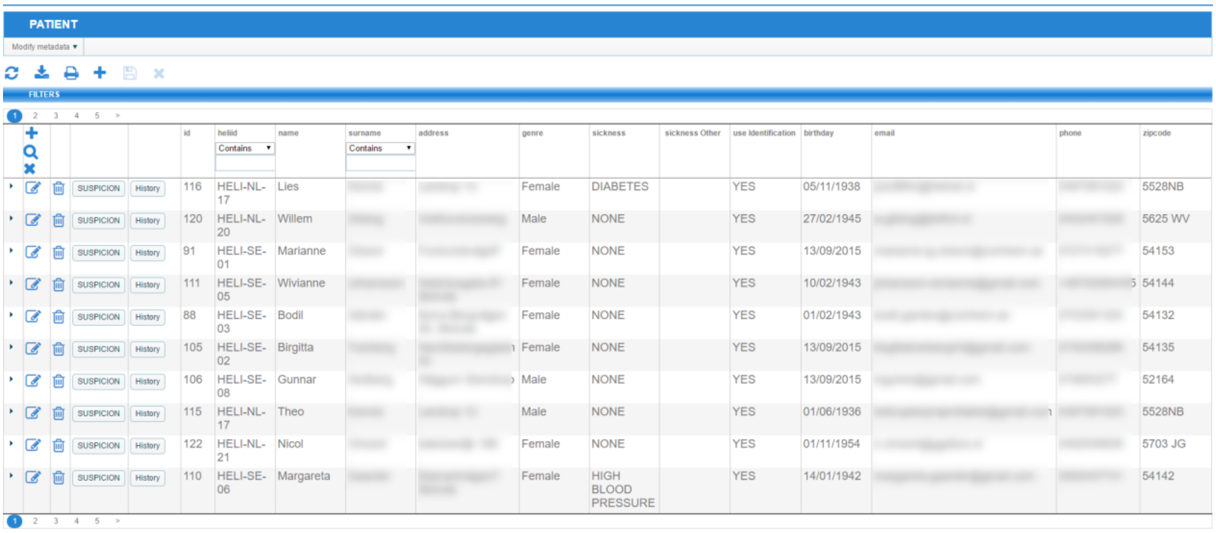



**Figure 2.6: sample screenshot of the web-based control dashboard (Helicopter Support Services, HSS)**

On the clinical side, a second dashboard was designed, aimed at health professional support. This was run by a different server, operated by METEDA. This web application can be used by the caregiver to supervise the automatic triage functions performed by the system. Caregiver can log into the application and access a specific list of patients (assigned by the administrator) that are under his care.

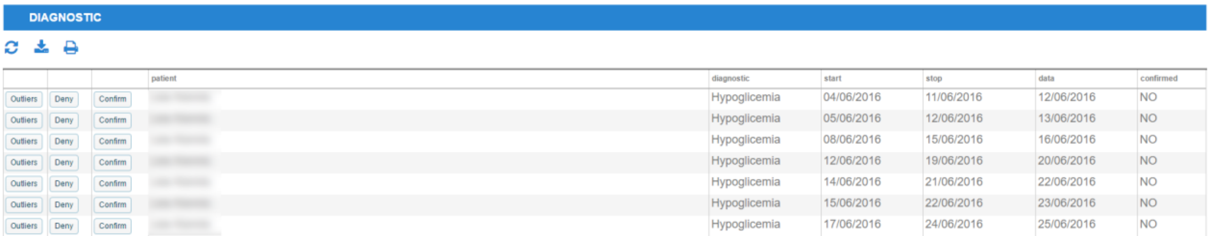


PATIENT

| id | heliid | name | surname | address | genre | sickness | sickness Other | use Identification | birthday | email | phone | zipcode |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 116 | HELI-NL-17 | Lies | [illegible] | [illegible] | Female | DIABETES | | YES | 05/11/1938 | [illegible] | [illegible] | 5528NB |
| 120 | HELI-NL-20 | Willem | [illegible] | [illegible] | Male | NONE | | YES | 27/02/1945 | [illegible] | [illegible] | 5625 WV |
| 91 | HELI-SE-01 | Marianne | [illegible] | [illegible] | Female | NONE | | YES | 13/09/2015 | [illegible] | [illegible] | 54153 |
| 111 | HELI-SE-05 | Wivianne | [illegible] | [illegible] | Female | NONE | | YES | 10/02/1943 | [illegible] | [illegible] | 54144 |
| 88 | HELI-SE-03 | Bodil | [illegible] | [illegible] | Female | NONE | | YES | 01/02/1943 | [illegible] | [illegible] | 54132 |
| 105 | HELI-SE-02 | Birgitta | [illegible] | [illegible] | Female | NONE | | YES | 13/09/2015 | [illegible] | [illegible] | 54135 |
| 106 | HELI-SE-08 | Gunnar | [illegible] | [illegible] | Male | NONE | | YES | 13/09/2015 | [illegible] | [illegible] | 52164 |
| 115 | HELI-NL-17 | Theo | [illegible] | [illegible] | Male | NONE | | YES | 01/06/1936 | [illegible] | [illegible] | 5528NB |
| 122 | HELI-NL-21 | Nicol | [illegible] | [illegible] | Female | NONE | | YES | 01/11/1954 | [illegible] | [illegible] | 5703 JG |
| 110 | HELI-SE-06 | Margareta | [illegible] | [illegible] | Female | HIGH BLOOD PRESSURE | | YES | 14/01/1942 | [illegible] | [illegible] | 54142 |

**Figure 2.7: sample screenshot of the web-based health professional dashboard**

For each patient, the caregiver can check if there's a diagnostic suspicion elicited by the system and get further details, verifying which events concurred to that specific warning.



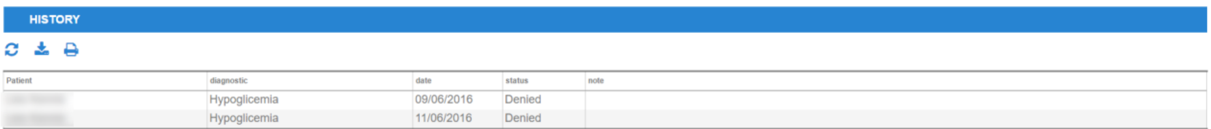
DIAGNOSTIC

| patient | diagnostic | start | stop | data | confirmed |
|---|---|---|---|---|---|
| [illegible] | Hypoglicemia | 04/06/2016 | 11/06/2016 | 12/06/2016 | NO |
| [illegible] | Hypoglicemia | 05/06/2016 | 12/06/2016 | 13/06/2016 | NO |
| [illegible] | Hypoglicemia | 08/06/2016 | 15/06/2016 | 16/06/2016 | NO |
| [illegible] | Hypoglicemia | 12/06/2016 | 19/06/2016 | 20/06/2016 | NO |
| [illegible] | Hypoglicemia | 14/06/2016 | 21/06/2016 | 22/06/2016 | NO |
| [illegible] | Hypoglicemia | 15/06/2016 | 22/06/2016 | 23/06/2016 | NO |
| [illegible] | Hypoglicemia | 17/06/2016 | 24/06/2016 | 25/06/2016 | NO |

**Figure 2.8: sample screenshot of the web-based health professional dashboard: active DS list**

He can check the outliers to find out what changed in the patient behavior, based on sensor data or user interaction with the tablet application. The mobile app, in fact, does not simply show the sensor status, but also asks targeted questions to the user based on his behavior (e.g., in case of "Severe decrease of walking velocity" the system asks if the patient skipped a meal or took a very little amount of food). Based on that information (or contacting directly the patient if needed), the caregiver decides to confirm or deny the suspicion and act accordingly in case of correct diagnosis.

The caregiver also may access the history of notifications during the monitoring period. He can check if the event was confirmed and review his notes to be able to identify a recurring behavior or an isolated event.



HISTORY

| Patient | diagnostic | date | status | note |
|---|---|---|---|---|
| [illegible] | Hypoglicemia | 09/06/2016 | Denied | |
| [illegible] | Hypoglicemia | 11/06/2016 | Denied | |

Figure 2.9: sample screenshot of the web-based health professional dashboard: patient DS log

Caregiver and system administrator may also  access a monitoring dashboard, showing statistics on the performance of the triage functions. Such a summary view includes, for each diagnostic suspicion managed by the system, the dashboard shows the number of occurrences and related confirmations, allowing to estimate false positive and false negative rates. Finally, for each diagnostic suspicion, relevance and reliability of each outlier contributing to the Bayesian Belief Network can be inspected.

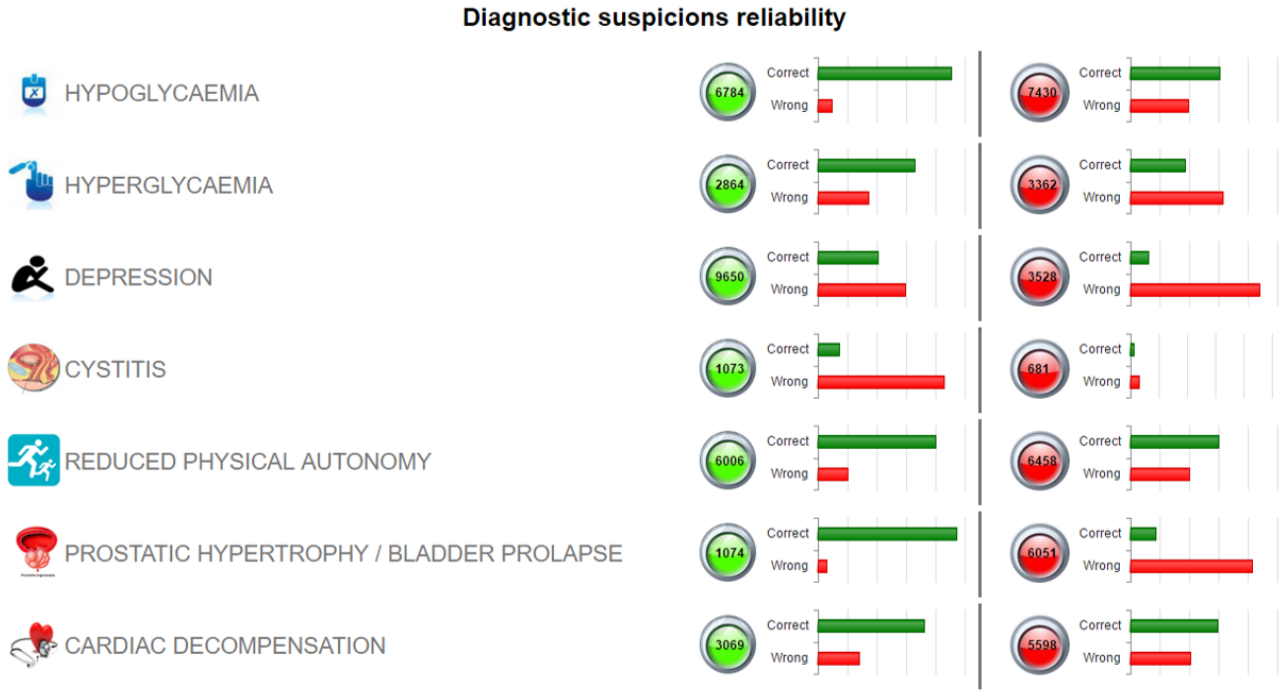



Figure 2.10: sample screenshot of the web-based health professional dashboard: DS statistics [data not meaningful]

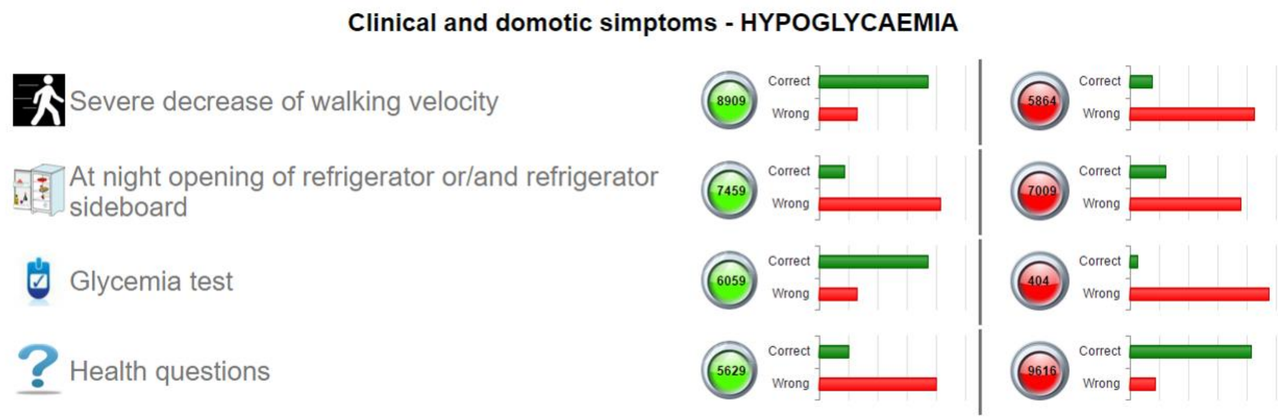



Figure 2.11: sample screenshot of the web-based health professional dashboard: DS-specific symptoms statistics [data not meaningful]

## 2.4 Development activities

The experience acquired as result of the first installations allowed to carry out some changes in the devices firmware to address some issues raised in the early stages of testing on the field.

A first revision activity concerned the toilet sensor: the sensitive element of this device is an infrared sensor, capable of measuring the distance at which an object or a body, placed in front of it, is located. The sensor is conceived to be installed with the infrared sensor pointing at the toilet: when a person is in front of it at a distance below a predetermined threshold, the sensor signals the presence and the data is interpreted as an access to the toilet. The definition of a unique value for the threshold turned out to be not so trivial, because the correct positioning of the sensor was influenced by external aspects, like, for example, the bathroom size. In some cases the rooms were very small and the pre-set value was too high, resulting in an over-determination of the accesses and increasing the false positives rate. Conversely, setting a low value may cause a greater difficulty in the detection. Because of these reasons, during the first installations it was not always possible to properly setup the toilet sensor. In order to remove this kind of issues, a simple threshold calibration procedure has been developed, which has to be executed by the technician during the sensor installation: this made possible to adapt the sensor to different scenarios. In Fig. 2.12, the control panel for range selection is shown.

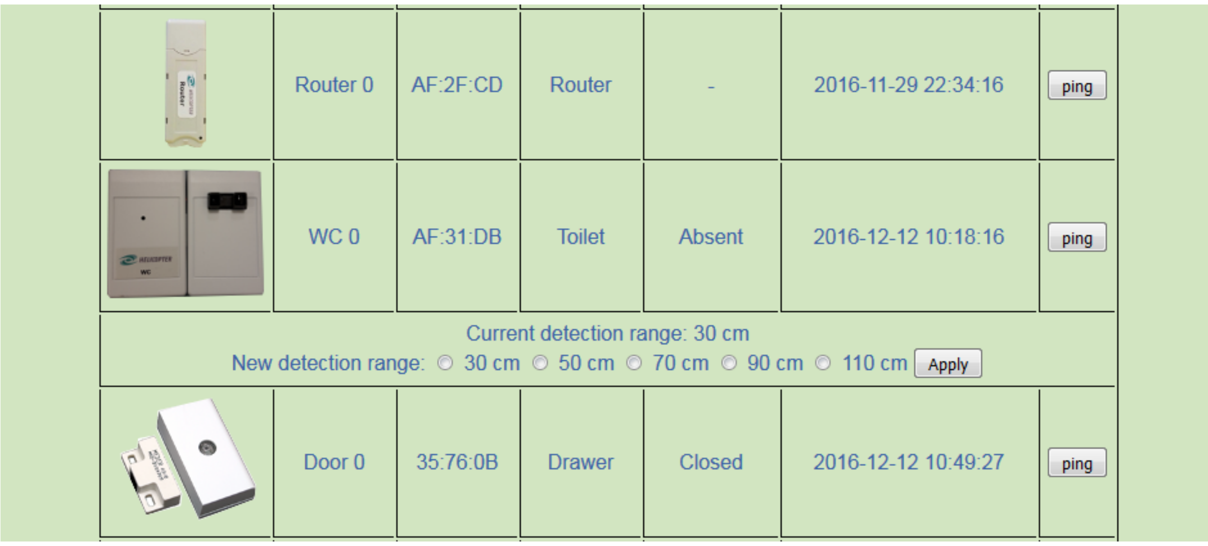



**Figure 2.12: toilet sensor - detection range control panel**

As stated before, some sensors are able to perform user identification. Some of these devices have this function integrated, some others exploit a separate "identifying" module for this purpose. To preserve battery lifetime, the identifying modules are put in a deep sleep mode (i.e. total absence of communication and elaboration) for most of the time and they are awoken only when the "main" device detects an activity; in order to communicate the result of the identification procedure they have to reconnect to the ZigBee network. In the first phase of pilot experience, when a particular network topology was built, they could not reconnect to the network after performing an identification, becoming orphan devices, hence unable to communicate their data. This issue turned out to be due to a bug in the ZigBee stack (which is a system component released by Texas Instrument) and it has been solved updating the firmware of all the devices with a new release of the stack. This operation was not trivial and required a lot of updates in the firmware of the devices, since the new stack presented some differences in some libraries and functions. Furthermore the reprogramming of the faulty devices in Sweden and in The Netherlands was needed, and it was carried-out in person by the UNIPR technicians.

The first tests have also allowed to re-evaluate the data provided by the wearable sensor. The analytic models need information about the user's activity in a quantitative sense. At first the intrinsic data derived from the user interaction with the sensors have been considered for this purpose: nevertheless, this data have proved to be not sufficient for a correct elaboration of the models. For this reason, the wearable sensor has been revised in order to produce more detailed information about the quantity of the user movement. The user speed is often indicated [3] as a good parameter for such an analysis: the velocity could be computed by the wearable sensor exploiting the on board Inertial Measurement Unit (composed by an accelerometer, a gyroscope and a magnetometer), through the integration of the acceleration signals generated by the body movements. However, given the incremental error due to integration over a large time window, an accurate speed calculation is not simple to be carried out on a platform with low computational capabilities. On the other hand, real-time transfer of large data streams could result in a very high shortening of the battery lifetime: to prevent this problem the radio-link has to be exploited to communicate only pre-elaborated and synthesized data.  A good compromise is represented by the Energy Expenditure parameter, namely the energy that a person needs to carry out a physical activity. This is somehow related to the velocity, but its calculation avoid integration over a long time base. According to [4], EE can be estimated by the equation:

$$EE = k_1 + k_2 I_{A,tot}$$

where k1 and k2 are empirical constants and $I_{A,tot}$ can be carried out from acceleration components (ax, ay, az) according to:

$$I_{A,tot} = \int_{T_W} a_x dt + \int_{T_W} a_y dt + \int_{T_W} a_z dt$$

Selecting an appropriate integration window (TW), the drift error is minimized. This algorithm does not require a large computational effort so it is indicated to be implemented in the wearable sensor firmware and thus enabling a continuous monitoring.

The acceleration components are sampled at a 60 Hz rate, filtered with a high-pass filter (Butterworth, 4th order), in order to eliminate frequency components at baseband related to gravity acceleration and steady-state movements, and eventually numerical integration is carried out.

Some tests have been performed: in fig. 2.13.a the estimated Energy Expenditure for several young healthy subjects, walking at different velocity on a treadmill, is depicted. This approach offers a good repeatability, with results substantially independent by the subject, and the ability of discriminating well among different walking velocity, proving that the Energy expenditure is a good estimator of human activity in the quantitative sense.

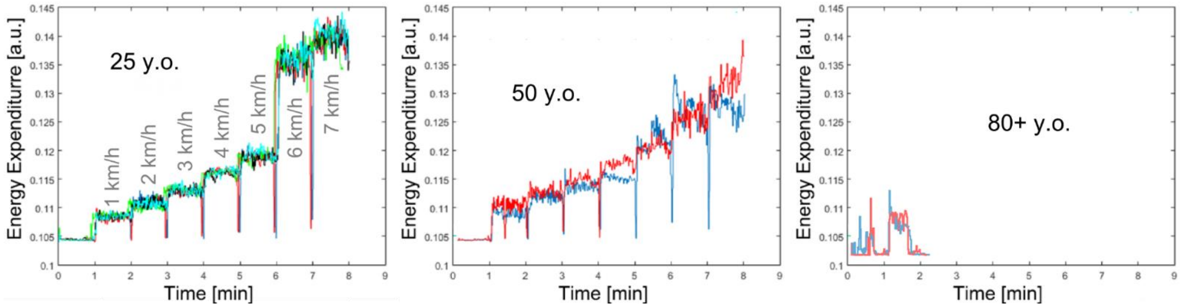



**Fig. 2.13 - Estimated energy expenditure for people of different ages (a. 25-30 years old, b. 50-60 years old, c. 80-90 years old), walking on a treadmill at different velocity.**

The other graphs (fig. 2.13.b and fig. 2.13.c) plot the results of the same test, but performed by subjects of different ages. The differences, easily recognizable comparing the graphs, can be linked to the relation existing between the difficulty in performing the exercise and the age of the subject: greater the age, greater the difficulty. This analysis confirms the ability of this parameter in discriminating among different levels of activity also in such cases.

# 3. Pilot data analysis

The analysis presented in this section involves 28 pilot sites:

- 20 in the Netherlands, named NL_01-22 (6 sites had two users at home). Pilots NL_09 and NL_10 withdrew from the project after a short period for personal reasons, and therefore they are not taken into account in the analysis.
- 8 in Sweden, named SE_01-08 (2 sites had two users at home).
- The pilot sites generated 46.9 million records during the pilot phase, averaging at 6652 records/day/pilot site and 1.23 million records/user.

The final setup and tuning of the systems deployment was carried-out between March and April, allowing data to be gathered continuously for about three months. The analysis consider the period from the beginning of April up to the beginning of July, gathering data from the pilots' stable period.

A preliminary analysis has been carried out to check the integrity of the data and to perform a first unrefined investigation of the pilot sites. Fig. 3.1 shows the data coming from NL_02 during the period from April 25$^{th}$ to May 9$^{th}$ (14 days). The blue or red ticks signal that the sensor detected the given action-of-interest; on the last row the EE coming from MuSA is displayed.

Even at a first glance, the data appear to be well organized and the devices seem to have worked correctly; during the night, the bed signals the presence continuously, while the other sensors are not activated; MuSA was worn correctly throughout the days, except for May 5$^{th}$ and 6$^{th}$.

A pattern is clearly visible for the chair sensor: it was placed in front of the television, being frequently occupied during the evening, right before going to bed. Furthermore, even if it is not so visible from the picture, the fridge activity is denser in the period around 12:30 pm and around 19:30 pm respect to the other parts of the day.

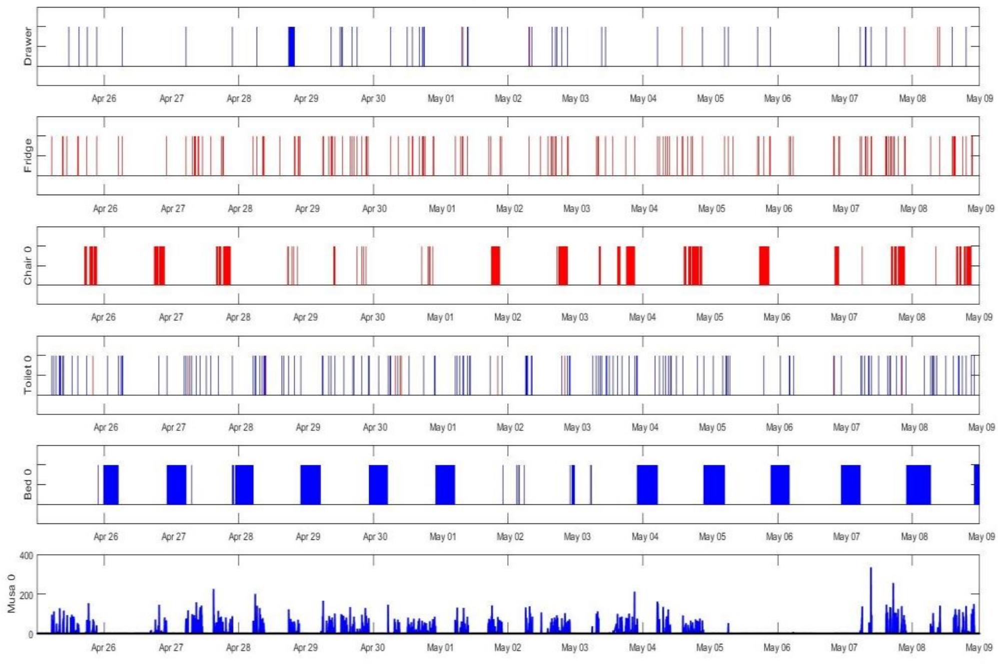



**Figure 3.1: some raw data from NL_02. The six plots represent, from top to bottom, Drawer, Fridge, Chair, Toilet, Bed and MuSA. A red or blue line signals the device activity in that moment**

Let's consider now a pilot with two users: Figure 3.2 shows the data coming from NL_01, on a two-days period from April 20th to April 22nd.

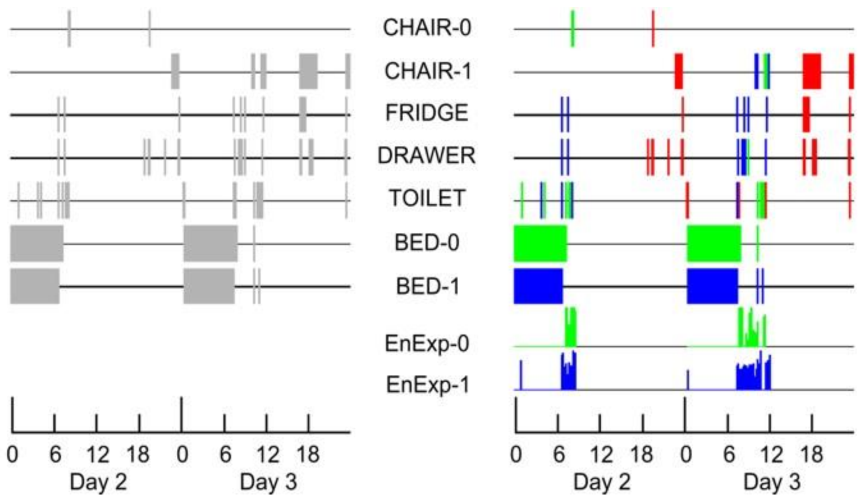



**Figure 3.2: raw data from NL_01, where two users are involved. On the right, the situation without identification; on the left, identification is introduced, allowing to distinguish the two users (green and blue).**

The diagram on the left refers to the picture coming from the environmental sensors only, while the diagram on the right shows improvements obtained by introducing MuSA features. In particular, personal energy expenditures appear on the two lowermost rows, and environmental sensors outputs were tagged. The color code is the following: green ticks refer to activity attributed to "user 0", while blue ticks stand for "user 1" actions. Red ticks instead, stand for actions not tagged (either because performed by a third person, or because the users didn't wear the MUSA device while acting).

The overall picture is quite clear and sound: in the experiment, users wore the device during the night and the morning, while they were not wearing it in the afternoon. Consistently, all activities carried out while carrying the wearable device were properly tagged, whereas afternoon activities remained untagged. As expected, the energy expenditure is low while resting in bed, and raises to higher values in the morning.

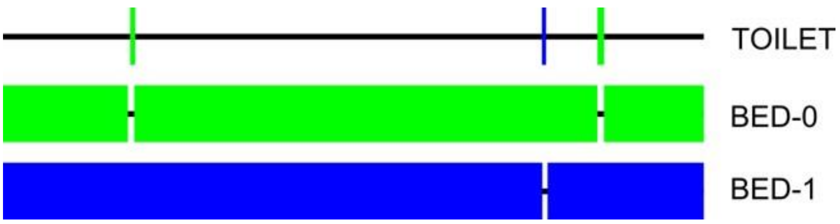



Figure 3.3: a detail from Fig 3.2, showing how the users that gets up from bed is correctly identified by the Toilet

Some consistency check can be easily done, by comparing different sensors outcomes. In Fig. 3.3, bed sensors and the shared toilet sensors are compared in a particular view, related to night time and showing consistent data: whenever the toilet recognizes a specific user, the related bed is shown to be empty.

In summary, about 220 sensors were deployed in the pilot environments. A total number of 24.36 M "events" were signaled by such sensors during the observation period. Average uptime for all sensors was 59,56% of the observation time. To this respect, it is to be noted that some sensors (most notably wearable sensors) inherently involve intermittent usage and that, also due to the prototypal nature of some of the sensors, some maintenance action (e.g., checking for proper connection, replacing exhausted batteries, etc.) was needed, either performed by the end-user himself or by the pilot support team. Of course, uptime statistics are quite dependent on the compliancy with such maintenance prescriptions. Such data are therefore more than acceptable in general, with lack of maintenance not jeopardizing the overall behavioral picture. If we limit ourselves to best-maintained pilots, uptime and usage statistics improves significantly, with a 30% subset of pilots exceeding 80% service time, and best ones approaching 99%, as shown in the table below, referring to best-performing pilots in either country:

| | NL [%] | SE [%] |
|---|---|---|
| **drw0** | 98.97 | 100.00 |
| **toilet0** | 98.97 | 100.00 |
| **fridge0** | 98.97 | 100.00 |
| **bed0** | 98.97 | 53.61 |
| **musa0** | 28.87 | 43.30 |
| **chair0** | 98.97 | 100.00 |
| **chair1** | 98.97 | 100.00 |
| **bed1** | 97.94 | 93.81 |

| | | |
|---|---|---|
| **med0** | 95.88 | n/a |
| **musa1** | 22.68 | 74.23 |
| **toilet1** | 98.97 | n/a |

**Figure 3.4: sensor uptime (% over total observation time) for best performing pilots in NL and SE.**

After pilot end, all sensors were inspected by UNIPR, and no major fault emerged. The only notable exception is related to the micro-USB connector in the MUSA device, which was exploited for daily recharging the battery. In some device, such connector was damaged: this indicates both the need of a more robust housing than that used in the pilot prototypes, and possibly the difficulty of farsighted elderly people in plugging such a connector (which however is compliant with EU recommendations for mobile USB chargers). This outcome, coming from long-term field test, results in valuable feedback for device re-design and engineering, and will be taken care of in subsequent version of the MUSA device.

Overall, testing of sensors designed for the HELICOPTER project was successfully: sensors provided data as expected and, after a few design iterations described above, they worked flawlessly during the pilot time. Lack of maintenance is an issue indeed, especially with reference to elderly people with limited technology skill: to this regard, however, maintenance needs will be greatly reduced in the final, market-ready devices. Maintenance consists of changing batteries (and being warned to do so) whenever necessary. This does not require any major skill, yet this bother can be significantly reduced: power management can now be effectively redesigned, avoiding much logging activity currently introduced in prototypes for testing and debugging purposes, and carefully tuning operating cycles on actual field-test outcomes. This will result in large improvement on battery lifetimes and less frequent maintenance needs. Moreover, final tuning of supporting apps will allow for simpler and more effective user notification strategies, allowing the user or the caregiver to promptly attend to maintenance tasks. This also apply to (desultory, indeed) network issues, possibly caused by accidental disconnection of a network router.

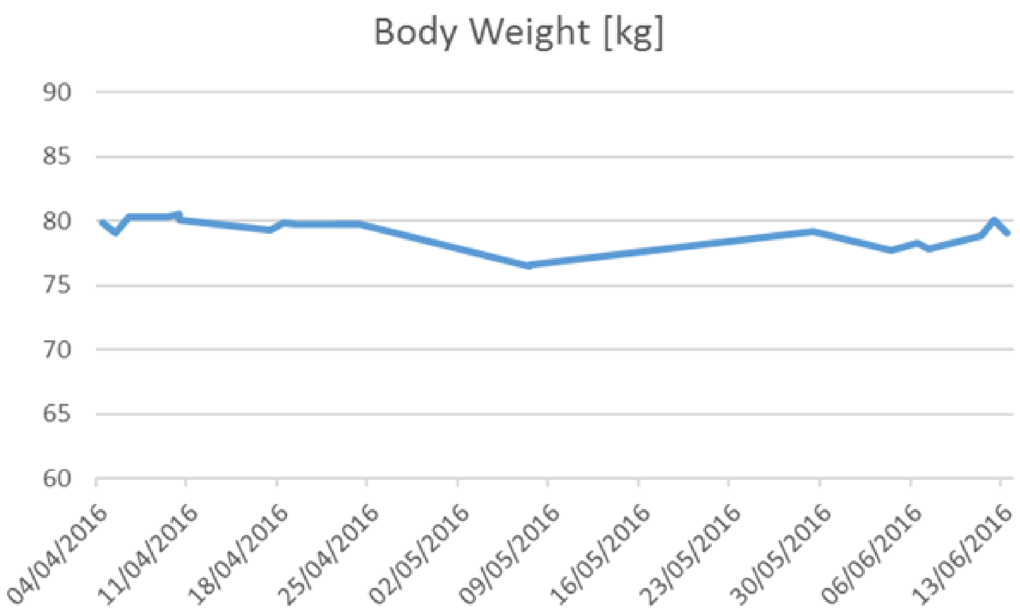



**Figure 3.5: sample plot of Bodyweight, as extracted from HELICOPTER database.**

Similar consideration apply to clinical sensors, embedded in the very same network: of course interaction with clinical sensor was less frequent and dependent on user habits. Also, due to national regulation issues, only data coming from Dutch pilots were logged. A total of 342 measurement were accumulated, distributed over the observation period. Sensor networking was operating as expected, with no service interruption. A sample of clinical data plot is shown in Fig. 3.5 and 3.6 below.

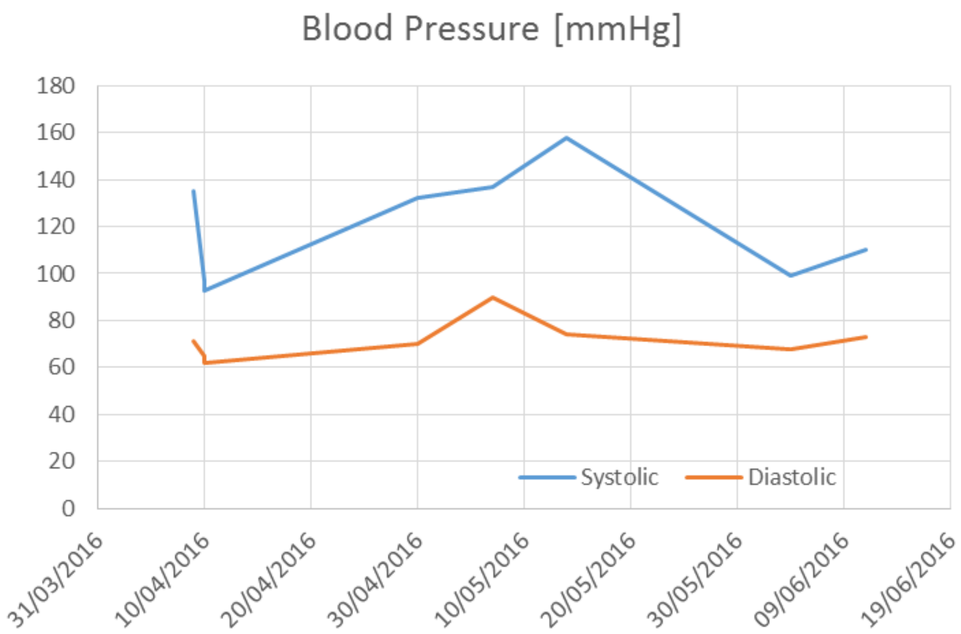



Figure 3.5: sample plot of Blood Pressure measurements, as extracted from HELICOPTER database.

The UI monitoring feature, although introduced only in late versions of the app, was tested as well. The system was fully functional and several thousands of interactions were actually logged.

| | Country | Interaction Number |
|---|---|---|
| Summary | NL | 2169 |
| | SE | 1229 |
| | Totale | 3398 |

Figure 3.6: Summary of xAPI logging activity.

In this case, only a proof-of-concept can be given, due to the limited test size. Nevertheless, we consider this as a quite promising option, to be investigated in more depth in project follow-up.

The home systems worked without noticeable interruptions during the whole pilot phase. The automatic update system put in place ensured the latest version of the software components were deployed on all home systems without the need of on-site human intervention.

Despite the satisfactory behavior of the overall infrastructure, however, an extensive validation of the "automatic triage" was not possible in the pilot timeframe. As mentioned elsewhere, this comes from the basic following aspects:

- The number of pilot users and the pilot duration was relatively limited. Although adequate to a technical validation, a clinical trial would have required a different scale, the effort and cost of which was beyond reach of the HELICOPTER project.
- To maximize the chances of eliciting "diagnostic suspicions", a more varied end-user population would have been requested, including persons more likely to suffer from medical conditions, as well as healthy persons to provide control reference. Actual pilot population was instead quite homogeneous with this respect. This was needed to provide data analytics techniques developed within the project with a reasonably sized statistical sample: given the inherent limitation in the pilot size, introducing a greater variability would have possibly resulted in too much statistical noise and would have jeopardized the development of data analysis modules.

Therefore, a complete validation of the "automatic triage" will require a larger scale trial, to be conducted with clinical supervision (e.g., with a randomized control trial approach). Nevertheless, we carried out further validation procedures, to deal with such an issue. Basically, neither "true positive" diagnostic suspicion events (DS) nor "false positive" DS was detected during the pilot run. We exploited therefore simulated approaches to induce positive events, by means of a twofold strategy. First, a simulated behavior was generated artificially, by means of the approach described in D4.4 and the ability of the model to recognize abnormal behavior, based on accumulated, user-specific knowledge was ascertained.

Then, a "perturbative" approach was exploited, by taking the usage data coming from an actual user (i.e., NL-14), originally yielding no DS. We faked a medical condition by tweaking some parameters, and we were able to trigger diagnostic suspicions:

```
COMPUTING FOR: User HeliBrain (HELI-NL-14) - BETWEEN 2016-06-04 -> 2016-06-11

- COMPUTING DIAGNOSTICS SUSPICIONS
-- RiskOfHeartFailure
--- Nodes: [age_above_sixtynine, decreased_movement_speed, diabetes,
hypertension, increased_body_weight, increased_diuresis_frequency,
increased_laying_down, increased_sitting_television, renal_failure, hypothesis]
--- Beliefs: [no]=0.4658706467661691; [yes]=0.5341293532338309
--- hypothesis: yes

-- RiskOfReducedAutonomy
--- Nodes: [age_above_sixtynine_and_below_eighty, age_above_or_is_eighty,
decreased_movement_speed, increased_diuresis_frequency, increased_laying_down,
increased_sitting_television, hypothesis]
--- Beliefs: [no]=0.43; [yes]=0.5700000000000001
--- hypothesis: yes
```

**Figure 3.5: fragment of HELICOPTER data analysis log, related to induced triggering of diagnostic suspicion**

Of course, this still does not validate the clinical vision at the model background, but allows to test the data chain as a whole and the system functionality: data coming from the field enters the behavioral analysis module, anomalies are actually discovered and the diagnostic suspicion model is properly triggered.

In order to check for "false negative" DS, we then referred to end user themselves, asking for significant changes in their health during the observation period. Our insight chances were actually limited by privacy concerns, so we just limited ourselves to ask if they were prescribed any changes in their therapies (assuming any relevant condition would result in some adjustment in therapies). All replies were negative, corroborating the system functional validation.

Furthermore, based on actual pilot data, a number of additional data analytics module were implemented and tested: i.e., besides the main HELICOPTER concepts, data coming from pilots enabled further offline studies related to behavioral analysis and its relevance in AAL activities. In particular, we focused on the need of extracting meaningful behavioral profiles even from noisy data, and to infer anomalies and trends in an unsupervised fashion, i.e., without the need of providing a-priori defined "normal" ranges and thresholds. Machine-learning techniques were exploited to this purpose, and the availability of data extracted from HELICOPTER pilots enabled testing and validation of such approaches.

For example it is possible to profile the expected user-sensor interaction throughout the day and, subsequently, it can be assessed whether there exist significant deviations through different periods (e.g. different sensor profiles during different months). In order to simplify the report, only some significant results are shown in the following figures, but this does not mean that similar conclusions cannot be retrieved from other pilots.

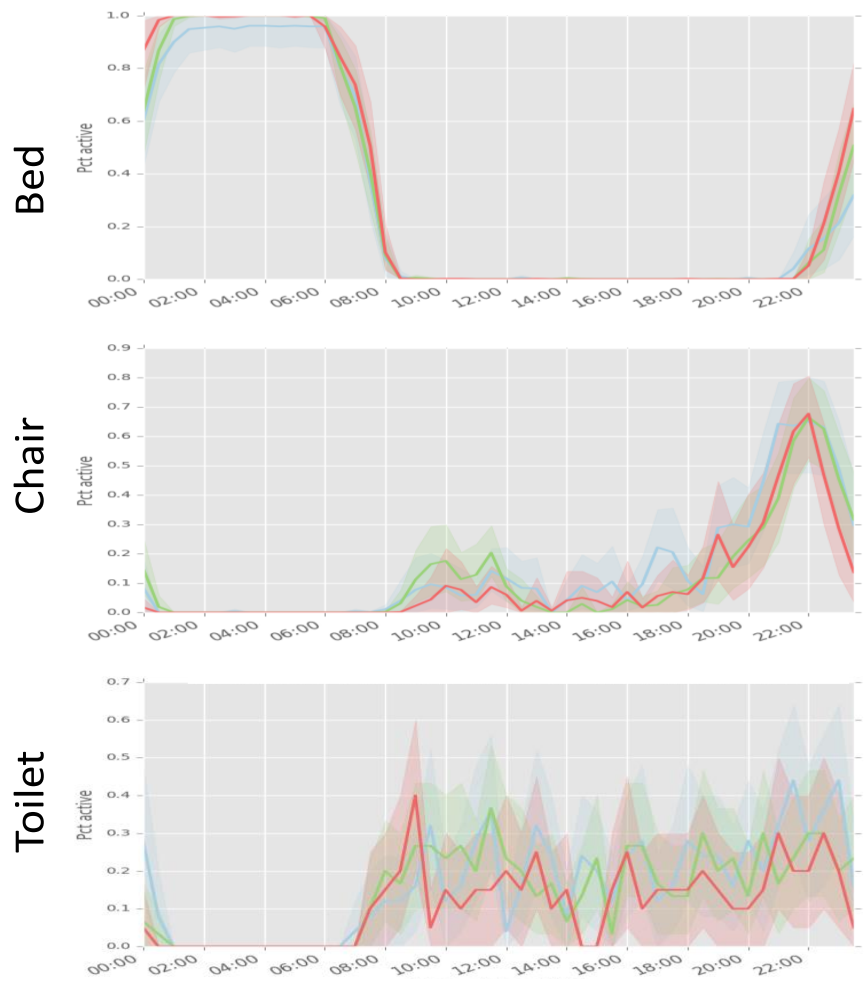



**Figure 3.4: daily usage of bed (top), chair (centre) and toilet (bottom) for NL_14. The x-axes represent the time of the day; the y-axes represent the probability of the sensor being activated in the given time slot; the color code is blue for April, Green for May and Red for June**

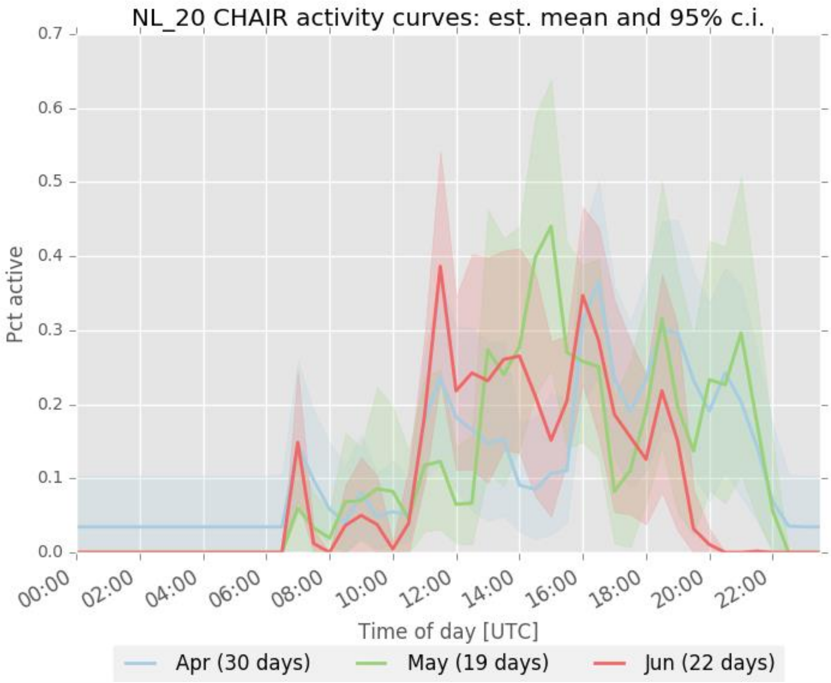



**Figure 3.5: daily usage of chair sensor for NL_20; the color code is blue for April, Green for May and Red for June**

The upper two plots of Fig. 3.4 show the daily usage of bed and chair in NL_14, divided by months (April-June-July): every day has been divided in 48 slots of 30 minutes each, and the curves represent the expected percent of time that the sensor will been active within those 30 minutes; the shaded areas, instead, represent 95% confidence interval of such prediction. A similar representation regarding the toilet is given in the lower plot of Fig. 3.4, but this time the curve represents the expected probability that, on average, the device is activated at least once within those 30 minutes.

As expected, it is highly probable for the bed to be active (i.e. occupied) during the night, whereas the toilet is used with regular frequency throughout the day; furthermore, the user spent time on the chair mostly during the evening, before going to bed. Interestingly, the toilet is never used during the night.

It can also be noticed that the user's behaviour is consistent throughout the observed months, i.e. the activity profiles match consistently (accounting the variability intervals) with very few deviations (significant from a statistical point of view, but apparently not so much from a behavioural perspective). Instead, Figure 3.5 shows a deviating trend during late evening: in June the user was less likely to spend time on the chair, on average.

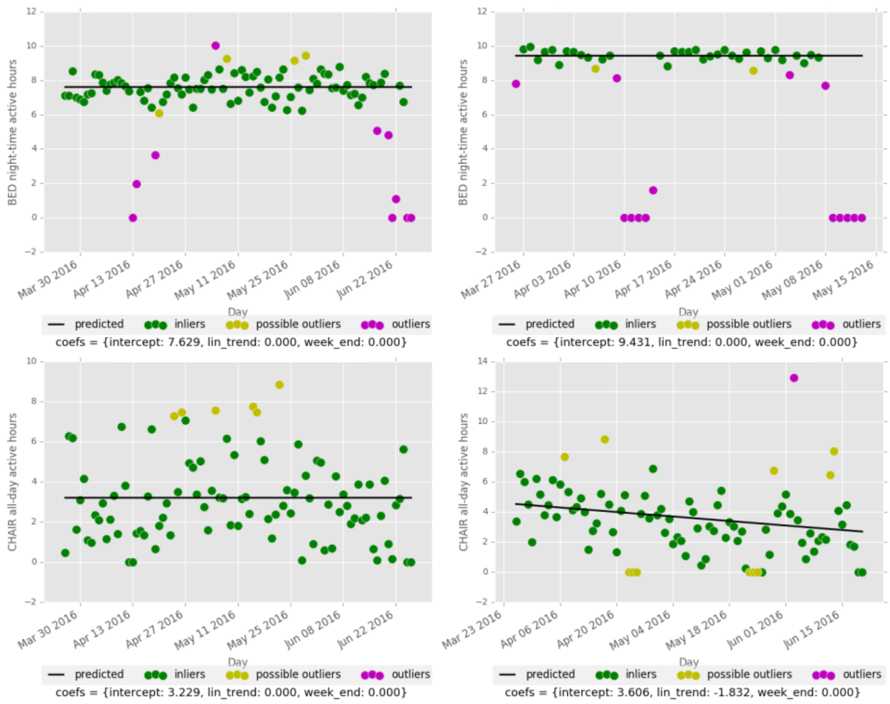



**Figure 3.6: regression plot of the bed (top) and chair (bottom) occupancy on NL_16 (left) and NL_18 (right). Green dots are the inliers, whereas yellow dots are possible outliers and purple dots are outliers. At the bottom of each graph are also reported the intercept and the weights given to linear trend and weekend.**

Starting from the amount of time spent in bed during each day, a predictive approach has been assessed, using the LASSO regression. The expected value of the parameter *y* can be determined with the formula

$$E\{y|data\} = N(E_0 + w_1 * f_1 + \cdots + w_n * f_n, \sigma^2)$$

Where $w_i$ is the weight given to the i-th feature $f_i$, and *N* is a Gaussian distribution. Starting from a list of features, with the LASSO regression it is possible to determine (within a given interval of confidence) which feature is non-significant, and assign to it a null weight.

Two features were considered, in addition to an intercept (average) term: week_end (=1 if the considered day was a weekday, =0 if it was a Sunday or Saturday) and lin_trend (to determine if there was a linear trend in the occupancy pattern).

To determine the outliers, a three-steps approach was followed:

1) Eliminate the points outside of the interquartiles , i.e. those points which values exceeded Q3+(Q3-Q1)*1.5 or were lower than Q1-(Q3-Q1)*1.5 (or equal to zero if this number resulted in something <0)

2) Apply the Student's t-distribution on the remaining samples to fit a model, and use that model to estimate a predictor on all the data (included the ones discarded after step 1)

3) Label the ones where 2,5<t-value<0,5 as possible outliers, and the ones where t-value<0,5 as outliers

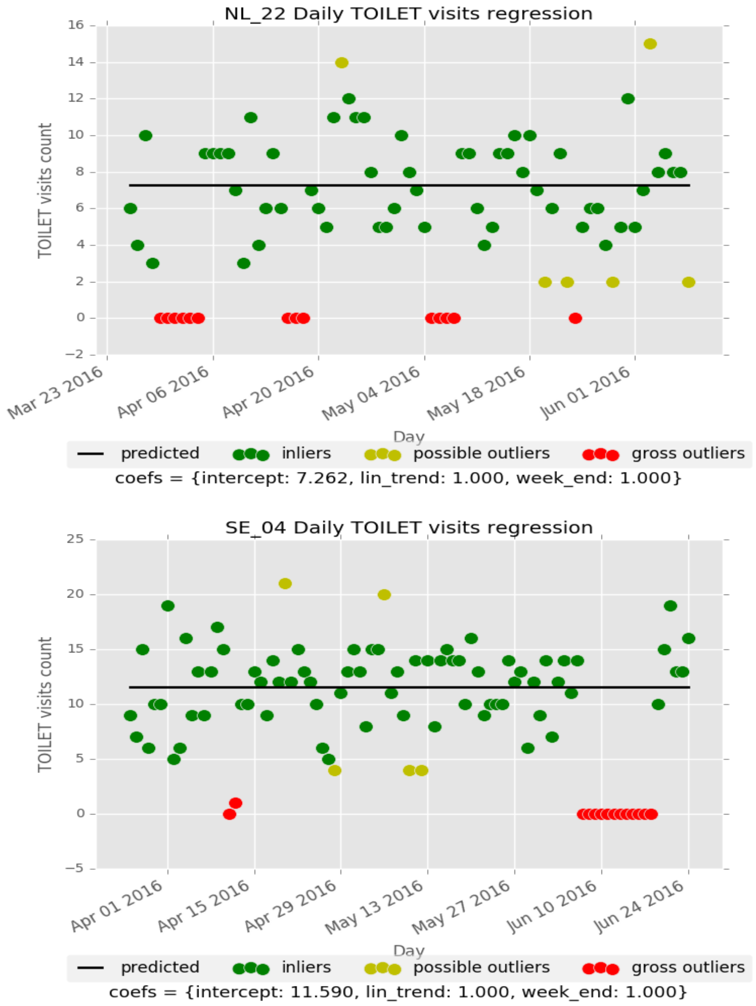



**Figure 3.7: regression plot of the Toilet for NL_22 (top) and SE_04 (bottom)**

Fig. 3.6 shows the regression plot of the bed and chair occupancy on NL_16 and NL_18: the x-axis represents the time (from April to June), while the colored dots represent the amount of time (in hours) that the device has been active during a given day, also showing inliers and outliers with different colors; The black solid line is the regression line; the Figure also reports the weight given to the two features and the intercept value, i.e. the average time spent in bed.

As can be seen, on NL_16 both bed and chair occupancies show no dependency on weekends nor linear trends (all the weights are equal to 0); however, on NL_18 the chair has a linear trend over time, expressed by the linear coefficient equal to -1.832.

To model the toilet usage a similar approach has been followed, this time using a Poisson regression, since the dataset was formed by integer numbers (counting the numbers of toilet activation during each days). The Poisson regression exploits the Generalized Linear Model, according to which

$$\ln(E\{y\}) = \beta X \rightarrow E\{y\} = e^{\beta X}$$

Where βX is the linear predictor, a linear combination of unknown parameters β.

In our case, considering the three parameters intercept, week_end and lin_trend ($\beta_0$, $\beta_1$, $\beta_2$) and their three weights ($X_0$, $X_1$, $X_2$), simple mathematical reasoning leads to

$$E\{y\} = e^{\beta_0} * e^{\beta_1 X_1} * e^{\beta_2 X_2}$$

Therefore, the parameters $\beta_0$ and $\beta_1$ have no influence if their weight is equal to 1, not to 0 as in the previous case ($X_o$ has been omitted from the above equation because its weight is always = 1, hence it has no influence on the multiplication with $\beta_0$).

Four different models have been created, taking into account four different combinations of the parameters: $\beta_0$, $\beta_0$-$\beta_1$, $\beta_0$-$\beta_2$, $\beta_0$-$\beta_1$-$\beta_2$. Model selection is performed, in order to select the most parsimonious model, according to the Bayesian Information Criterion.

The outliers have been determined in a similar way:

1) Create a Poisson distribution based on the predicted values
2) Label the values as possible outliers if the chance of falling outside of the distribution was between 97.5% and 99.5%, and as outliers if the chance was above 99.5%

Figure 3.7 shows the daily toilet visit for NL_22 and SE_04. Even if both of them show no dependencies on the weekend nor a linear trend, a marked difference can be noticed in the intercept value (7.2 against 11.59).

Despite some scarceness of integrity in the data and even if considering the devices on their own, it was yet possible to identify some significant patterns.

# 4. Pilot performance evaluation and conclusion

From the technical point of view, all major goals were satisfactorily reached. Most relevant achievement that were validated through the HELICOPTER pilot include:

- Some purposely designed sensors, particularly aimed at behavioral insights.
- The identification strategy, which allows to attribute activity detected by an environmental sensor by interacting with wearable sensors. This is a key feature in making the monitoring approach suitable for multi-user environment as well.
- The development of an heterogeneous, open sensor network, in which commercial sensors from different vendors (environmental, clinical) and custom-designed ones interoperate smoothly.
- The development of behavioral analysis models, based on machine-learning approaches and suitable for evaluating anomalies and trends in a reliable, user-aware fashion.
- The design and development of data fusion strategies for diagnostic suspicion elicitation, based on Bayesian Belief Networks.
- The implementation of a cloud-based pilot management system, including web services to manage software distribution and updates, pilot sites supervision and data storage.
- The implementation of different interfaces, aimed at end-user, caregiver and professional users, providing access to monitoring data in differentiated fashions.
- The experimental implementation of an UI interaction tracking mechanisms, suitable for feeding behavioral models with relevant insight about learning/cognitive performance.

Cooperation between all HELICOPTER system components was demonstrated and validated, and design of some components was refined during the pilot execution, based on pilot feedbacks. Further hints for future improvement and exploitation came from the pilots as well. In particular, it was found that wearable sensors were in principle accepted by end-users; nevertheless, reducing the need for active interaction turned out to be a key issue. E.g., daily charging of the wearable device, although being a common practice with smartphones and other mobile gadgets, was considered as a demanding procedure and efforts are being made to optimize such an issue.

More generally speaking, on the down side it should be mentioned again that the pilots were run in their final configuration (i.e., after technology tuning) for too a short period to obtain meaningful evidence of the clinical value of the monitoring approach. This also resulted in some lack of user motivation, emerging from users review. Nevertheless, the pilot provided full feasibility and proof-of-concept information, thus paving the way for further testing on a larger scale, which is now planned by business partners in the project consortium.

# Bibliography

[1] Bianchi V., Grossi F., De Munari I., Ciampolini P., 'MuSA: A multisensor wearable device for AAL', Federated Conference on Computer Science and Information Systems (FedCSIS), 2011, pp. 375-380

[2] https://www.adlnet.gov/adl-research/performance-tracking-analysis/experience-api/

[3] Studenski S., Perera S., et al '*Gait Speed and Survival in Older Adults*', Journal of the American Medical Association, 2011, 305, (1) pp. 50-58.

[4] Bouten C.V., Westerterp K.R., Verduin M., Janssen J.D., '*Assesment of energy expenditure for physical activity using a triaxial accelerometer*', Medicine & Science in Sports & Exercise, 1995, 26, (12), pp. 1516-1523.