D5.7 - Final DSF Predictive Models

Deliverable ID	**D5.7**
Deliverable Title	**Final DSF Predictive Models**
Work Package	**WP5**
Dissemination Level	**PUBLIC**
Version	**1.0**
Date	**2019/09/16**
Status	**Final**
Type	**Prototype**
Lead Editor	**LIBAL**
Main Contributors	**LIBAL, LINKS**

Published by the Storage4Grid Consortium

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 731155.

Document History

Version	Date	Author(s)	Description
0.1	2019-03-08	LIBAL	First ToC
0.2	2019-03-26	LIBAL	Input to SoH and RUL section
0.3	2019-04-05	LIBAL	Modification to SoH and RUL section
0.4	2019-04-11	LIBAL	Modification to figures and captions
0.6	2019-05-03	LINKS	Comments
0.7	2019-05-14	LIBAL	Comments adaption Modification to section 1 and 2
0.8	2019-08-01	LINKS	Contribution on PV, Load, ESS Predictive Models
0.9	2019-08-08	LIBAL	Text modification to section 1 and 2 Insert use cases overview table Insert Conclusion Comments
0.9.1	2019-08-26	LINKS	Input for section 1
0.9.2	2019-08-26	LIBAL	Adapt input from LINKS and ready for review
0.9.5	2019-09-13	LIBAL LINKS	Reply on reviewers' comments
1.0	2019-xx-xx	LIBAL	Adapted changes according to the review Ready for submission to the EC

Internal Review History

Review Date	Reviewer	Summary of Comments
2019-08-29	Baldini (EDYNA)	Approved: • Document ok • Minor corrections • Some observations
2019-08-28	Gustavo Aragón (FRAUNHOFER)	Approved:

Table of Contents

Executive Summary

D5.7 - Final DSF Predictive Models is a public prototype deliverable in Storage4Grid project. It will provide the final inputs of the predictive models' development based on D5.6 - Initial DSF Predictive Models.
These models are meant to empower S4G DSF and GESSCon with reliable forecasts and estimations, that in turn would optimize the ESS and RES exploitation.

GESSCon, based on Remaining Useful Life (RUL) and State-of-Health (SoH) estimation pattern would adapt the estimation result for each individual ESS to provide better operational instruction to improve efficiencies for those systems. These models are being designed, developed and implemented within LiBal's Battery Management System (BMS) product. In S4G project, since Bolzano commercial site and grid-side storage in Skive are using LiBal's energy storage, this feature will be available to the use cases HLUC2-PUC-2 and HLUC3-PUC-3 [D2.2].
DSF as the main component for storage analysis and planning requires an anticipated view over the state of the grid, this requirement essentially refers to the electricity consumption and generation. Distributed generation tied with the environmental aspects, load consumption bound directly to the human behaviour [1] and storage status data a combination of both. These models will be provided to assist the DSF analysis and scheduling.

1 Introduction

This deliverable document presents the final predictive models' development for the project, within T5.4 framework. We divide the predictive algorithms into two major categories which coincide with the division of activities:

1. Feature oriented: to develop estimation algorithms for electrochemical energy storage system (ESS) behaviour such as RUL (remaining useful life) and SoH (state of health), an effective feature has to be extracted from the measurement data. The term feature-oriented, refers to such features for SoH and RUL prediction.

2. Observation oriented: load consumption, RES production and ESS status forecasts are based on the human behaviour, climate and environment. These might be modelled either from the recorded/observed data in the past and can also be enriched with the reliable forecasts from 3rd parties' services which are obtained from numerous terrestrial and satellite measurements and observations.

Battery health predictive model will be operated in LiBal's BMS. Once these health-related parameters has been changed, it should be updated on GESSCon side as well for more accurate and optimized operation result [D4.3] [D4.5].

Observation-oriented predictive models determine some key inputs for GESSCon optimization framework. Thus, there is high dependency between GESSCon and DSF Predictive Models. In this deliverable the Storage4Grid (S4G) consortium provides the relevant inputs for the two prediction categories that mentioned above respectively.

In this deliverable section 1 contains the introduction of the prototypes and its relation to other deliverables and use cases within the project. Section 2 focuses on SoH and RUL algorithm and implementation explanation with lab test result. Section 3 aims at DSF predictive models and its implementation in the project. The predictive models are grouped in PV, load and storage status predictions and further detailed based on specifications. A conclusion will be stated in section 4.

1.1 Scope

The main goal for the T5.4 is to develop predictive algorithms for the DSF system given that a large number of entities in the DSF grid models are statistically predictable, and therefore can be described with sufficient confidence by means of simple probabilistic or even deterministic models. But also, some of the load, RES and storage systems are not known or predictable a priori [D3.2].
This task will integrate advanced lifetime estimation algorithms on Lithium Balance novel BMS platform. The advanced lifetime estimation includes State-of-health (SoH) and Remaining Useful Life (RUL) specific for stationary energy storage systems. These algorithms are not currently available for off-the-shelf battery management systems. These algorithms could predict End of life of the battery packs for easier maintenance and battery swap planning.
Alongside the ESS related estimation models, load consumption, RES production and ESS status predictive models are being developed within this task in order to supply DSF with reliable data.

1.2 Related documents

ID	Title	Reference	Version	Date
[D2.5]	Initial Lessons Learned and Requirements Report	D2.5	1.0	2017-05-30

[D3.1]	Initial S4G Components, Interfaces and Architecture Specification	D3.1	1.0	2017-09-15
[D3.2]	Updated S4G Components Interfaces and Architecture Specification	D3.2	1.5	2018-08-08
[D4.2]	Updated User-side ESS Control System	D4.2	1.0	2018-06-14
[D4.3]	Final User-side ESS Control System	D4.3	1.0	2019-06-13
[D4.5]	Final Grid-side ESS control System	D4.5	0.9	2019-09-13
[D4.7]	Final Cooperative EV charging station control algorithms	D4.7	1.0	2019-08-28
[D5.6]	Initial DSF Predictive Models	D5.6	1.0	2018-08-30

1.3 Use Cases Overview

Table 1 shows the use cases that has predictive models available in phase 3, according to the information available in D2.2 [S4G-D2.2].

Use-case	Predictive Models
HLUC-2-PUC-1: *Residential prosumer with storage and EV*	PV, Load, ESS status
HLUC-2-PUC-2: *Cooperative charging in the parking lot of a commercial test site*	PV, Load, SoH, RUL, ESS status
HLUC-2-PUC-3: *Simulation of high penetration of EV chargers and of prosumers with storage and residential EV charging*	PV, Load, ESS status
HLUC-3-PUC-1: *Support for analysing storage dimensioning and positioning in the low-voltage grid*	PV, Load
HLUC-3-PUC-3: *Voltage and flux control at grid side storage*	PV, Load, SoH, RUL
HLUC-3-PUC-4: *Coordinated Distributed storage in the grid*	PV, Load, ESS status

Table 1. Use cases overview

2 Battery Health Prediction

In this section, the final method that diagnose the battery health status will be presented. The State of Health (SoH) and Remaining Useful Life (RUL) implementation on BMS platform are documented, uncertainties and future work has also been considered.

2.1 Overview of Battery Health Prediction on LIBAL BMS

There is not a consensus concerning the definition of state-of-health (SoH) in industry or when looking into the literature. Generally, SoH can be seen as *a 'measure' that reflects the general condition of a battery and its ability to deliver the specified performance in comparison with a fresh battery*[1]

1 POP, V., BERGVELD, H. J., DANILOV, D., REGTIEN, P. P. L., AND

SoH can be defined and measured in various ways, concerning either:

- Capacity fade
- Internal resistance increase
- A measure of self-discharge over time
- Number of charge-discharge cycles
- ..

Or a combination of two or more in a given ratio.
Currently Lithium Balance c/n-BMS do only consider SoH as fade in capacity, estimated by combining the information gained by coulomb counting and SoC-OCV (Open Circuit Voltage) calibrations. The SoH is in principle only measured on the weakest cell but considered to be representative for the full package (as for the SoC-OCV estimation).
During this test series (section 2.2) the SoH is calculated as the capacity fade of a battery cell compared to the initial capacity. In other words, the full capacity at the present state compared to the initial full capacity of the battery cell, e.g.:

$$SoH = \frac{Q_{full}}{Q_{nominal}} \tag{1}$$

While $Q_{nominal}$ usually is provided by battery manufacturer (or measured by an initial labtest), Q_{full} at present conditions can be found by doing a full discharge in the lab. It is here important to test at conditions comparable to the nominal capacity that will often be specified by battery manufacturer at a specific temperature, discharge rate as well as upper and lower voltage limits after being charged by a specific charging scheme.
Q_{full} has the relationship shown in Equation (2).

$$Q_{full} = f(\Delta SoC_{OCV}, \Delta SoC_{CC}) \tag{2}$$

In an application a battery will most likely never be operated like in the lab and it is therefore required to have a method that can estimate SoH without doing a full charge or discharge as in Equation (1).
The SoH is during the following tests calculated by evaluating two SoC-OCV calibrations and an intermediate discharge sequence. By comparing the ΔSoC found from the two SoC-OCV calibrations (SoC_{OCV}) as in Euqation (3), with the ΔSoC found by coulomb counting (SoC_{CC}) as in Equation (4), the Q_{full} can be found. From the information gained during such operation it is possible to calculate a change in SoC found by OCV and a change in SoC found by coulomb counting:

$$\Delta SoC_{OCV} = SoC_{OCV\,1} - SoC_{OCV\,2} \tag{3}$$

$$\Delta SoC_{CC} = SoC_{cc\,1} - SoC_{cc\,2} \tag{4}$$

Initially, when there is no SoH estimation available in the BMS, Q_{full} is set equal to $Q_{nominal}$, the nominal capacity of the battery pack. Then, when the conditions above are met, Q_{full} is corrected from Equation (5)

$$Q_{full} = (1-k)Q_{full} + k \cdot f(\Delta SoC_{OCV}, \Delta SoC_{CC}) \tag{5}$$

k is a user-configured weight that determines how much trust to put in the most recent estimation of Q_{full}.
The new value of SoH is then updated as in Equation (1).

NOTTEN, P. H. L., 2008. BATTERY MANAGEMENT SYSTEMS: ACCURATE STATE-OF-CHARGE INDICATION FOR BATTERY-POWERED APPLICATIONS, 1ST ED. SPRINGER.

2.2 State of Health (SoH) system test

A brief introduction of the SoH estimation test profile is presented in the next subsections. Afterwards, the test result is shown and discussed.

2.2.1 Test profile

The tests are performed on the LIBAL's n-BMS battery setup. The tests are divided into two:

- Firstly 4 cells were dismounted from the battery package and a stepwise discharge was performed for getting the SoC-OCV characteristic and the full capacity. Following a capacity test was made on the 4 cells as explained in section 2.2.2.1.
- Secondly the four cells were mounted back in the battery package and the SoH tests were performed.

Recording SoC characteristics on a battery cell that is going to be used in a real application should be preferably made with a long resting time between each step and eventually with low discharge rate.
The actual SoH tests were performed by making partial discharges in different SoC-ranges. SoC-OCV calibration was performed before and after any partial discharge so that the ΔSoC_{OCV} and ΔSoC_{CC} could be found.

2.2.2 Test result

2.2.2.1 Capacity test

Capacity of the four cells were measured by doing full discharges of the cells. SoH was then calculated comparing the measured capacity to the design capacity as in Equation (1). Since $Q_{nominal}$ is 40 Ah,

CMU1 Cell2:

$$SoH = 84.1\%$$

CMU2 Cell5:

$$SoH = 92.7\%$$

CMU2 Cell6:

$$SoH = 90.8\%$$

CMU3 Cell11:

$$SoH = 91.4\%$$

The discharge sequence for the 4 cells are illustrated in Figure 1:

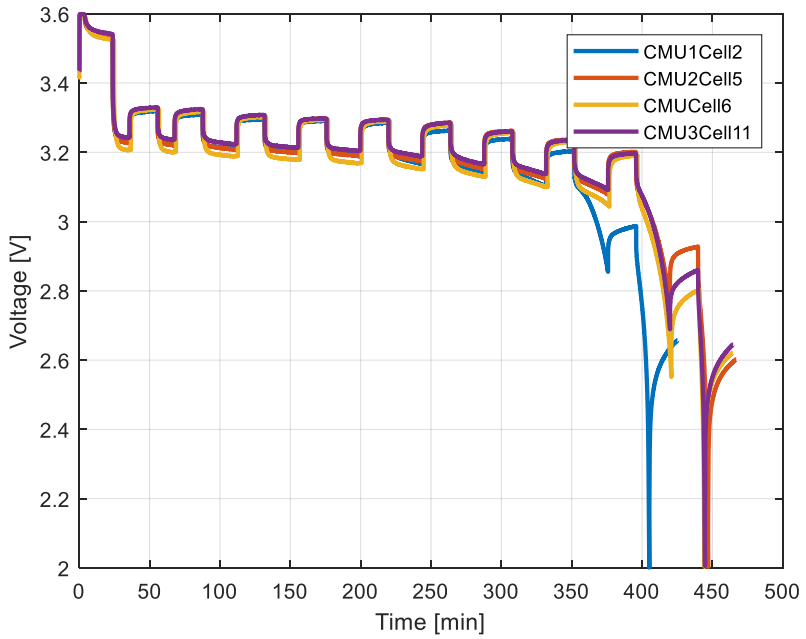

Figure 1. SoC-OCV characteristics according to designed discharge sequence

And the derived SoC-OCV relation is shown as in Figure 2:

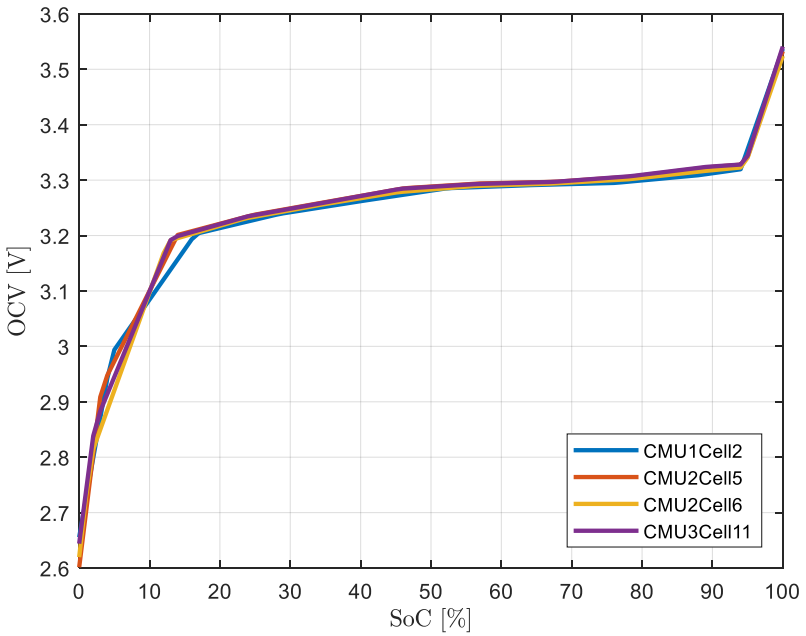

Figure 2. Derived SoC-OCV relation from the capacity test

2.2.2.2 SoH tests, discharge from 100% to 25% SoC

The battery pack is charged and fully balanced, then discharged to 25% SoC with 0.25C-rate. Figure 3 and Figure 4 show the SoH measured by the BMS and compares these with the SoH found by the battery cycler during the capacity test in section 2.2.2.1.

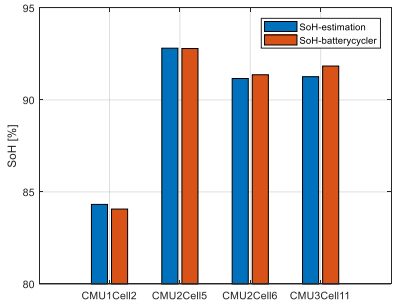

Figure 3. SoH-estimations compared to SoH measured by the battery cycler

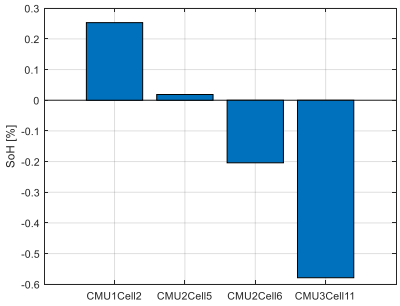

Figure 4. Deviation of the BMS SoH-estimate relative to the SoH measured by a full discharge in the battery cycler

2.2.2.2.1 Test improvement

The test was done again to verify the reproducibility of the SoH-estimates. At the same time the measured current (by shunt on BMS CMU board) accuracy was checked and found to be slightly wrong. The shunt resistor value was calibrated, which improved the accuracy of the SoH-estimates presented in Figure 5 and Figure 6.

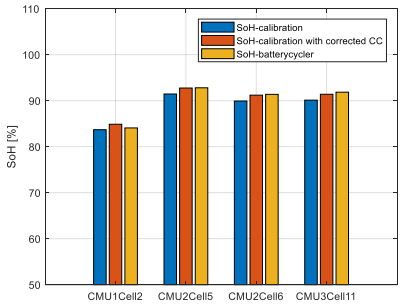

Figure 5. Reproduced SoH-estimations compared to SoH measured by the battery cycler

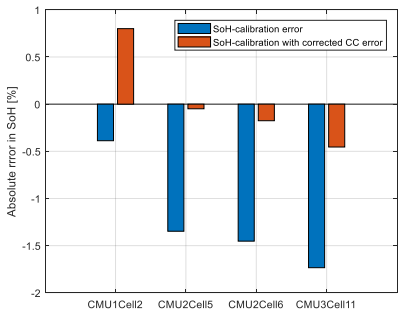

Figure 6. Reproduced Error in SoH-estimation

2.2.3 Uncertainties

Multiple measurements are needed to make a SoH-estimatation, and all have some uncertainties that adds up to the total uncertainty of the SoH-estimatation.
The main assumption by using this method for SoH-estimations is that the SoC-OCV relation is constant during aging of the battery and that there is no cell-to-cell variation. If that is not true it adds an uncertainty that can be evaluated as the variation between multiple SoC-OCV curves, e.g. the four measured here.
Secondly the SoC-OCV curve are very flat for LFP-cells, which gives quite a big uncertainty in SoC-estimation in some SoC-windows(taking the uncertainty on the cell voltage measurement into consideration).
Third there is the uncertainty on the current measurement. As SoH is only measured during a discharge sequence an offset on the current measurement will affect the SoH via the coulomb counting (the offset will not equal out as in other symmetrical charge-discharge type tests).

2.3 Remaining Useful Life (RUL)

The remianing useful life algorithm has been studied and review in the previous deliverable D5.6. However in order to implement such feature on BMS platform, currently only battery capacity has been considered for defining RUL, as in Equation (6)

$$RUL = f_1(Q_{full}) \tag{6}$$

RUL has direct correlation with the SoH result that has been documented in the previous section. In this case no additional lab test is needed for RUL. In the following subsection, the implementation of RUL will be presented.

2.3.1 Implementation

RUL is represented as an estimation date for battery end of life. A projection of the estimation date will be found by two Q_{full} measurements as shown in Equation (7)

$$RUL = T_{current} - \frac{(Q_{full_{current}} - Q_{full_{EoL}})(T_2 - T_1)}{Q_{full2} - Q_{full1}} \quad (7)$$

Where $T_{current}$ is the current date, $Q_{full\_current}$ and Q_{full_EoL} are battery capacity now and by the end of life. Q_{full1}, Q_{full2}, T_1 and T_2 are two battery capacities measured at two different time. Nomally Q_{full_EoI} is considered as 80% of the original capacity Q_{full}. But it is also relavent to what kind of application the ESS is doing, thus Q_{full_EoI} can be self-defined. For a chosen time range (T_1 and T_2), linear relation is assumed on battery capacity degradation and the lifetime. In this case RUL can be projected by the latest measurement. However, the chosen time range should be long enough in order to avoid error in the measurement. An initial assumption of the time range between T_1 and T_2 is having a difference for more than 2 weeks. This will be further verified and adjusted during operation.

$$RUL = (1 - k)RUL + k \cdot RUL \quad (8)$$

In Equation (8), k is a user-configured weight that determines how much trust to put in the most recent estimation of RUL.
In further implementation, more battery capacity measurement points can be used to derive the RUL instead of taking only two measurements.

3 DSF Predictive Models

DSF predictive models are all addressing time-series forecasts based on various features and sequences. In this chapter, several models have been introduced for each specific problem, complying with the nature of that data.

3.1 Photovoltaics production predictive model

In the context of S4G project Photovoltaics (PV) are the resource of renewable power. Photovoltaics is related to the direct conversion of the photons to electricity charge.
Forecast of this power needs the information about the maximum solar irradiation on a specific point of interest, declination and orientation of the pv panels, date and time in a first step. Then the weather condition that has a major impact on the sun's ray received by the panel must be taken into account. In a final step the physical characteristics of the PV system such as power converter and also PV modules come into effect. In the ongoing task, several data are being logged to further train and improve final results, these data are environment temperature, clouds in low elevation level range.

3.1.1 Solar Irradiation Calculation

To accurately estimate solar power, it is necessary to know the amount of pure solar radiation on a specific point based on the time and the point's exact position. There are various efficient models for calculating solar radiation in industrial sector and literatures, from which a model with a compromise between accuracy and simplicity has been selected for the present task, with some modifications and adaptations. The rationale behind this strategy is that even highly sophisticated models with powerful computational resources are subject to errors, therefore in a primary step the theoretical solar radiation is calculated from a promising method, then a machine learning routine completes the job. It is worth to note that the model apart from those

necessary adaptations contains many parameters that need to be computed and adjusted for each certain use-case.

3.1.1.1 Theoretical Solar Clear Day Radiation

Sun's relative position respect to the earth determines the intensity of radiation and thus power. The calculation of sun's position w.r.t a point on the earth surface is calculated with the following steps.
The year of forecast's horizon should be distinct to figure out if it is a leap year or not. Based on number of days per year, a Daylight Saving vector *DLS* is set to one for the interval between October 27th and March 31th and the rest equal to zero.
The Sun Declination Angle δ can be calculated for whole forecast horizon according to the formula 9:

$$\delta = -23.45 \cos\left(\frac{360}{t_d} \times (d + 10)\right) \tag{9}$$

Where *d* stands for the day of the year. The t_d assumes 365 in a normal and 366 in a leap year.
To compute local time correction, a constant *B* –only depends on the day *n* subject to forecast and its year- is computed from the Formula (9).

$$B = \frac{360\ (n - 81)}{t_d} \quad [deg] \tag{10}$$

Since a solar day is not exactly 24 hours due to irregularities of earth rotation, there is a need to take this fact into account through a correction term known as Equation of Time ϵ can be calculated following the Formula (10).

$$\epsilon = 0.165 \sin 2B - 0.126 \cos B - 0.025 \sin B \quad [hour] \tag{11}$$

Or from Formula (10):

$$\epsilon = 9.87 \sin 2B - 7.53 \cos B - 1.5 \sin B \quad [minutes] \tag{12}$$

The accuracy of the above approximated formulas is around half minute. Having ϵ at hand the Local Solar Time *LST* is computed from Formula 13:

$$LST = LT + \frac{\epsilon + L - 15 \times \Delta T_{UTC}}{15} \quad [hour] \tag{13}$$

With LT representing Local Time and L standing for Longitude. ΔT_{UTC} is the local time drift from UTC time.
Then from Formula (14) the Solar Hour Angle ω –that is zero at local noon- is obtained:

$$\omega = 15(LST - 12)\ [deg] \tag{14}$$

Now the angle of sunrise and sunset can also be calculated simply adding and subtracting the equation (15) to the local noon, respectively:

$$\omega_{0,\infty} = 12 \mp \frac{\cos^{-1}(-\tan\varphi \tan\delta) + L - LST}{15} + \epsilon \quad [hour] \tag{15}$$

After that, the sun's elevation and zenith angle should be calculated. Sun's elevation angle is the angular elevation of the sun from sun's ray projection on a specific point on the earth surface and is calculated as the Formula 16.

$$\alpha = \sin^{-1}(\cos\omega\cos\varphi\cos\delta + \sin\varphi\sin\delta)\ \ [deg] \tag{16}$$

With φ representing the latitude of the point of interest on the earth surface. Now it is possible to calculate sun's azimuth using formula (17):

$$\phi = \cos^{-1}\left(\frac{\sin\delta\cos\varphi - \cos\delta\cos\omega\sin\varphi}{\cos\alpha}\right)\ [deg] \tag{17}$$

It is necessary to pay attention to the sun's azimuth degree's sign, to have:

$$\phi = \begin{cases} \phi, & \omega \le 0 \\ 360 - \phi, & \omega > 0 \end{cases} \tag{18}$$

Respecting the sign conventions for solar azimuth angle ϕ, the angle between the sun and panel noted by γ can be computed simply by applying the Formula (19):

$$\gamma = |\phi - \Psi|\ [deg] \tag{19}$$

In which Ψ is the panel's orientation angle, and to take into account the panel tilt it is enough using Formula (20):

$$\theta = \cos^{-1}(\cos\alpha\cos\gamma\sin\xi + \sin\alpha\cos\xi)\ \ [deg] \tag{20}$$

Which calculates the incident angle θ based on the panel declination angle ξ.

There are some coefficients needed for calculations that are taken from experiments developed by the American Society of Heating, Refrigerating and Air-Conditioning Engineers (ASHRAE[1]), namely clear day solar flux model's table that is summarized in Table 2.

Month	B	C
Jan	0.142	0.058
Feb	0.144	0.06
Mar	0.156	0.071
Apr	0.18	0.097
May	0.196	0.121
Jun	0.205	0.134
Jul	0.207	0.136
Aug	0.201	0.122
Sep	0.177	0.092
Oct	0.16	0.073
Nov	0.149	0.063
Dec	0.142	0.057

Table 2. Extra-terrestrial Solar Irradiance and Related Data.

The solar power coming from sun's ray has different elements from which we focus on three major contributions; direct solar flux, diffuse flux on panel and reflected solar radiation. To calculate these metrics one needs to know pure direct insolation level that can be approximated by the Formula (21):

$$\mathrm{I}_D \approx 1367 \times 0.7^{\frac{1}{\cos \mathrm{Z}}}{}^{0678} \times e^{-B/\sin\alpha} \quad [\frac{W}{m^2}] \tag{21}$$

Z stands for sun's zenith angle in this context and is obtained simply using Formula (22) in the following:

$$Z = \cos^{-1}(\cos\omega\cos\varphi\cos\delta + \sin\varphi\sin\delta) = \sqrt{1-\sin^2\alpha} \quad [deg] \tag{22}$$

These coefficients are considered based on the month of the year and are obtained from several years of observations. Direct flux of the sun's rays on horizontal plane obtains from the Formula (23).

$$\mathrm{I}_{DH} = CI_D \quad [\frac{W}{m^2}] \tag{23}$$

So the diffuse flux on the declined panel is being calculated from Formula (24).

$$\mathrm{I}_{DD} = \mathrm{I}_{DH} \times \frac{1+\cos\xi}{2} \quad [\frac{W}{m^2}] \tag{24}$$

And reflected irradiation can be obtained by the Formula (25):

$$\mathrm{I}_R = \mu\, I_D \cos\theta \quad [\frac{W}{m^2}] \tag{25}$$

The μ is a coefficient that is determined according to the surrounding environment and materials. This can be transparent to the model in the beginning and it is one of the parameters to be found/estimated by the machine learning routine.
Finally, the total theoretical solar radiation is obtained by adding up these three sun rays' elements and following the Formula (26).

$$\mathrm{I}_R = (\mathrm{I}_D + \mathrm{I}_R + I_{DD}) \times \sin\alpha \quad [\frac{W}{m^2}] \tag{26}$$

The result of the presented model is logged in periodic intervals (1 hour for now) to be exploited by a further regressive model.

3.1.1.2 Presence of the Clouds and its Impact on Solar Radiation

In a second step, the impact of the clouds in low (up to 1500 m), medium (up to 3000m) and high (above 3000m) altitude levels come into effect. A uniform discrete distribution probability function takes into account the clouds density in different altitudes by a weighted sum scheme as in the Formula (27).

$$\mathrm{C}_D = \mathrm{w}_1\mathrm{C}_L + \mathrm{w}_2\mathrm{C}_M + \mathrm{w}_3\mathrm{C}_H\,; \qquad \sum_{j=1}^{3}\mathrm{w}_j = 1 \tag{27}$$

And the total cloud density C_D –as a normalized value- is applied to the clear day radiation through multiplying a random number X~U{0,1} by the

$$P(X) = \begin{cases} C_D, & R = 1 \\ 1 - C_D, & R = 0 \end{cases}; \quad I_R \leftarrow I_R \times R \tag{28}$$

The resulting vector will be very noisy, therefore a further step of discrete convolution is applied to smooth the profile, as reported in Formula (29):

$$I_R(n) \leftarrow \sum_{k=-K}^{K} I_R(k)\, \delta(n-k) \tag{29}$$

The impulse function δ is adapted by the optimal form described in the Formula (30):

$$D(k) = \frac{[u\left(k+\frac{K}{2}\right) - u\left(k-\frac{K}{2}\right)]}{K}; \qquad K = 8 \tag{30}$$

K which is the specification of impulse signal has been optimized for the solar radiation forecast, since it is a compromise between of the forecast resolution and its medium value accuracy.
Result of the described method is a vector with length and granularity equal to the forecast horizon and step-size set by the user.

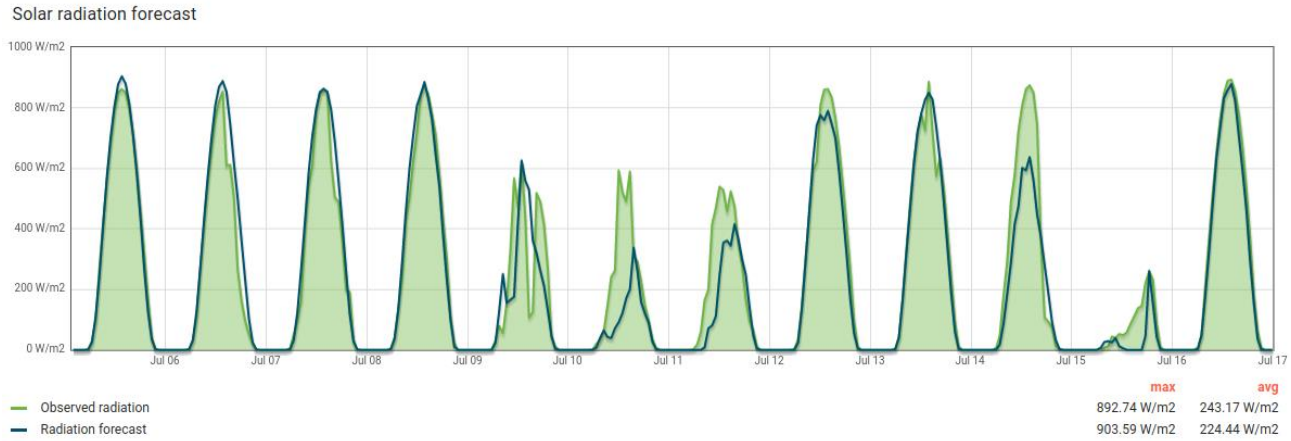

Figure 7. Solar radiation forecast and its closeness to the observed values.

3.1.1.3 Learning from the Field

There are many parameters and variables that are not related to sun and climate but to the photovoltaics complex system, its surrounding environment and materials. Hence it is necessary to pass calculated forecasted values so far to a regressive model as also the training data and the observed in the field as targets to find the unknown parameters, indirectly. Based on this reasoning, the machine learning process has been segmented in two subroutines:

- Clear day solar radiation: to figure out the right parameters from the field such as efficiencies of panel and converter, objects/buildings etc. that cause shadow and so forth.

- Cloud impact added: to find the correlations between the real power output and level of the clouds, temperature, etc.

To solve the regression problems, classic artificial neural network has been developed and implemented. Both regressive models are linear so a shallow network would be enough but one additional hidden layer is added to capture nonlinearities. The weighted some in output layer is linear and contains coefficients for above mentioned features. The last layer is set without any activation function, as so-called identity activation function.

For clear day radiation parameters, stored data should be filtered as a set of I_R^j for $\{ j \in J | \frac{dI_R^j}{dt} \leq \rho \}$, where ρ is determined by the experience but essentially means that smaller lower noise level is accepted.

Instead, to estimate right correlation of temperature, low, medium and high-altitude clouds over the quality of prediction a linear equation is adapted that sets the right coefficients for Formula (27) plus a weight for temperature and an intercept value.

The combination of theoretical and learning function results in reliable forecast that becomes more accurate by time as the dataset is getting enriched gradually. The Figure 8 depicts an example of forecast updated every hour versus monitored power output for the July 23, 2019. Although not all the predictions were as accurate as the brought example, nevertheless the model improves its performance progressively. Given that logging forecast and collected data from field form still a tiny dataset, and since clear sky days occurred rarely, no metric is presented in this report to evaluate the produced forecasts' errors.

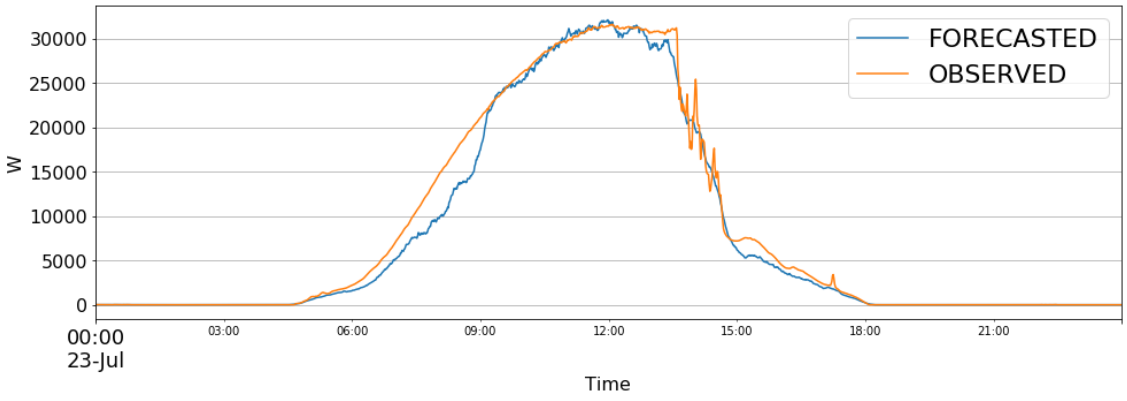

Figure 8. Example of PV generation forecast for Bolzano Commercial pilot.

3.2 Load predictive models

Load as a macro analysis subject can be sub-grouped into different categories since it might be a function of various motives from one case to the other. In S4G project case load predictive model is defined in three group:

- Substation (aggregated) load
- Electric vehicles load

- Household loads

Each gets treated therefore through a special method.

3.2.1 Substation (Aggregated) Load Prediction

Aggregated loads tend to have lower standard deviation as the number of its forming elements increases[2], so seasonality and repetitive patterns become more evident. Analysing aggregated loads in substation level unveils a strong weekly repetitive scheme. This fact reduces the uncertainty factors thus classic statistics' methods will produce high quality predictions even for a long horizon.

From Figure 9. Transformer current of supplying substation in Bolzano.Figure 9 one can observe low presence of random noise on aggregated load profile of a substation's transformer in Bolzano residential pilot's upstream.

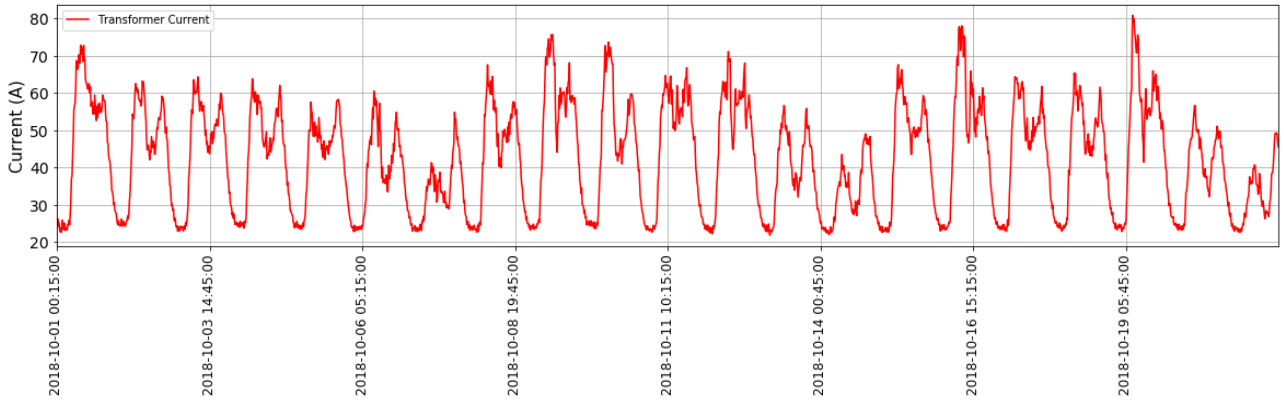

Figure 9. Transformer current of supplying substation in Bolzano.

The current representative for power as well- contains correlation between sampling values and observations with a certain lag in the past. Periodic patterns can be intuitively perceived from Partial Autocorrelation coefficients, as shown in the Figure 10.

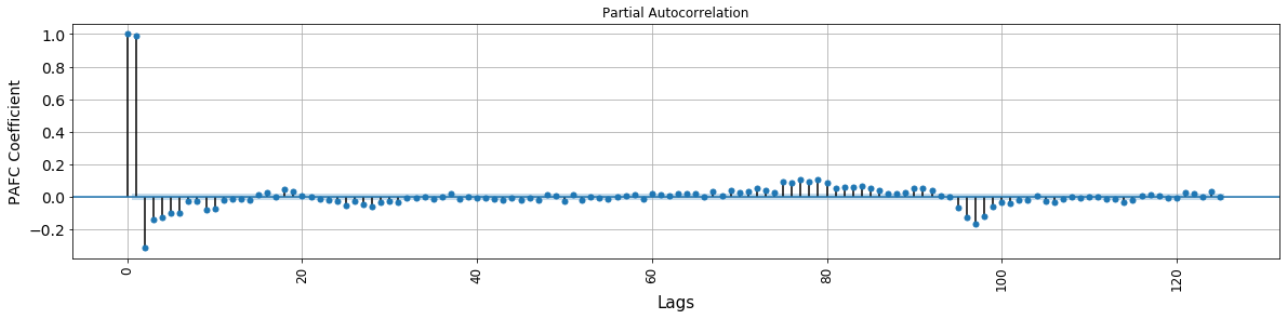

Figure 10. Correlation coefficients for transformer currents.

As intra-day correlations tend to become zero quickly, daily coefficients appear (lags of 15minutes in time). Therefore time series can be converted to stationary through differencing. Consequently, it can be concluded that an Autoregressive Integrated Moving Average (ARIMA) matches the requirements of such time series prediction.

A general convention of an ARIMA model can be formulated as the following:

$$\hat{Y}_t = \mu + \sum_{k=1}^{p} \emptyset_k \, Y_{t-k} \; - \sum_{j=1}^{q} \theta_j \, \varepsilon_{t-j} \tag{31}$$

Where $\hat{Y}_t$ is the prediction at time step t, μ is autoregressive intercept constant, p and q are the autoregressive and moving average orders respectively, $\emptyset_k$ is the coefficient of the regression weighted sum, for lag k, Y_{t-k} is the observation at k lag observation and θ_j is the error weight coefficient while error ε itself represent the moving average part of the complex formula.

Applying a differencing autoregressive of second order, or in other word ARIMA (2,1,0) model, the results are obtained as in the Figure 11.

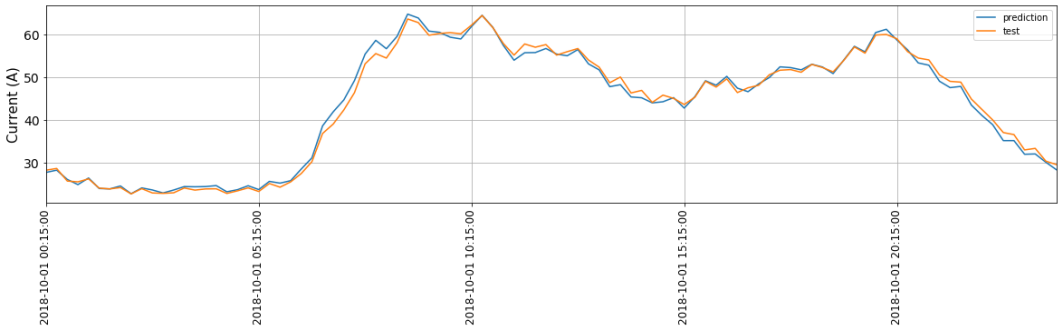

Figure 11. Transformer current prediction with a two lag second order ARIMA.

3.2.2 Electric Vehicles Load

Although the electric vehicles are nothing but active loads, however their prediction needs a different approach; the load modulation level is limited and known since there are a few standard active power modulations for the EV chargers, especially for the general-purpose ones installed in the public charging stations. Following the evolution of the S4G project there were two different kind of charging units subject to prediction:

- Public charging station in a commercial place.
- Electricity distribution operator's parking lot equipped and dedicated to only EV charging.

The analysis on the data gathered from the field showed different charging pattern for the two cases, and this needed to be treated separately. In general, for both cases a similar statistic model has been developed.

3.2.2.1 Public Charging Station

Charging pattern follows the citizen activities and habits. Also, charging behaviour (plugging time, parking duration, power and withdrawn energy) might be also different from one charging station to another, as shown in the Figure 12.

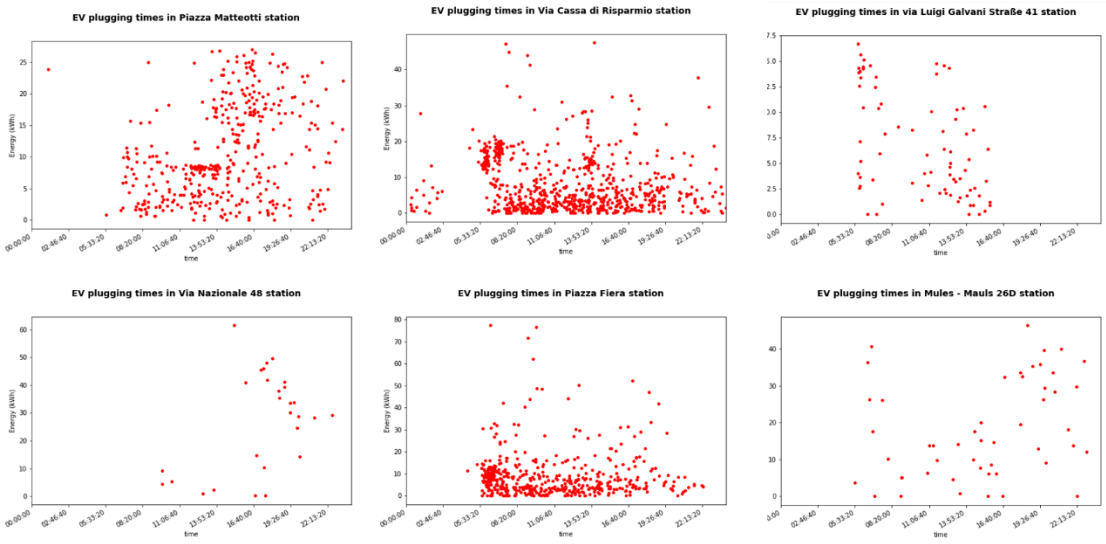

Figure 12. Charging events in various charging stations.

There is a significant correlation between plugging time, parking duration and received energy, however a hierarchical computation bottom-up tree is suggested to solve the problem as show in Figure 13.

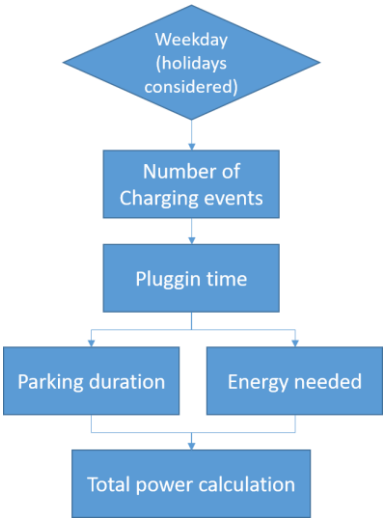

Figure 13. Prediction calculation hierarchical tree.

Number of events however is similar/proportional to the energy

As described briefly the number of EVs to charge in the public charging stations in Bolzano is determined from statistics and gets updated in determined interval of time. Figure 14 demonstrates the medium and standard deviation of charging events per weekdays.

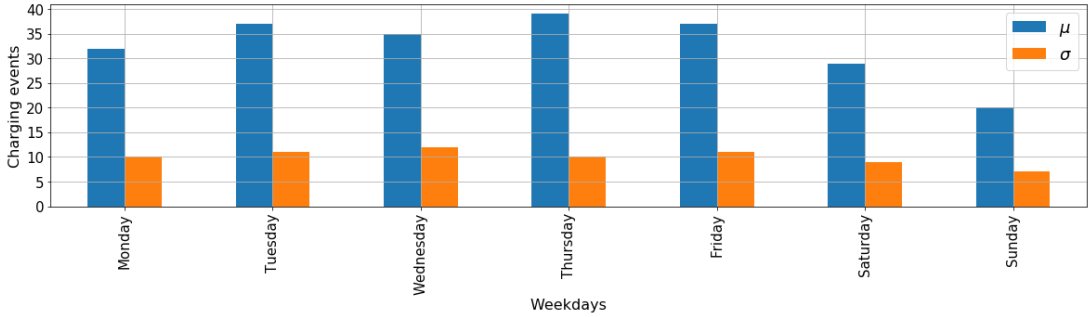

Figure 14. Weekday charging events in Bolzano public charging stations.

Next, the plugging time for those fleet to be charged is estimated. The Probability Distribution Function (PDF) of arrival time for a working day is as in the Figure 15.

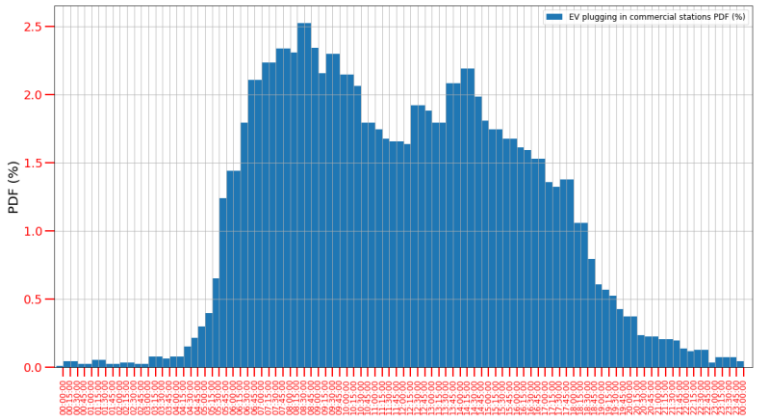

Figure 15. Probability Distribution Function of plugging time in a week's working day.

Then energy needed and parking duration are estimated according to the relevant statistical parameters.

The final result of prediction is a time-series for all the forecast horizon, like the Figure 16.

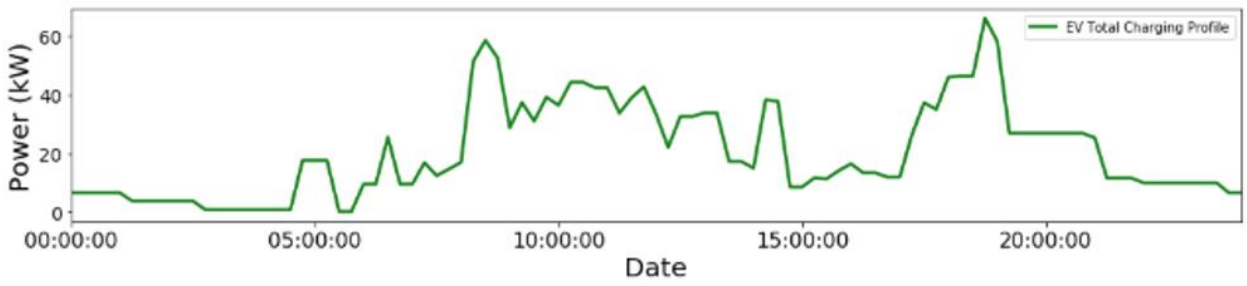

Figure 16. EV charging load prediction example.

3.2.2.2 EDYNA Parking lot

EDYNA has dedicated a parking lot to only electric vehicles, and there are 8 EV chargers installed. A similar approach like the public stations is applied to this case with some minor difference given that the data contents is different. However, charging events in the EDYNA parking lot has the Figure 17.

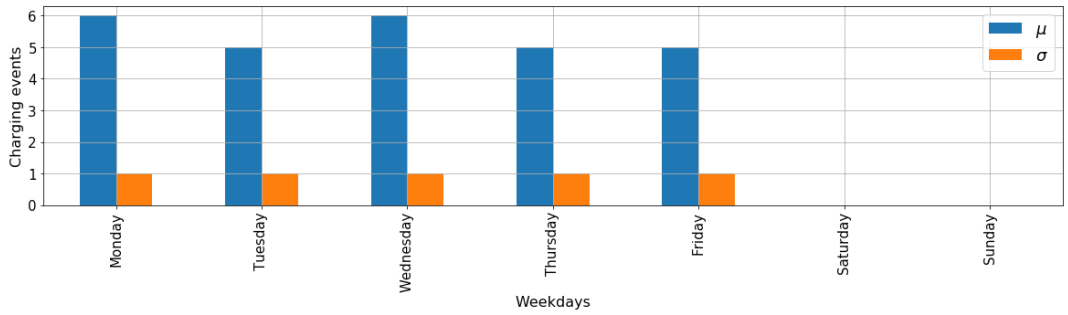

Figure 17. Charging events per day, EDYNA parking lot.

Once the number of EVs is known, the plugging time distribution determines the arrival time of the EVs. The plugging time distribution in EDYNA parking has the values that are reported in the Figure 18.

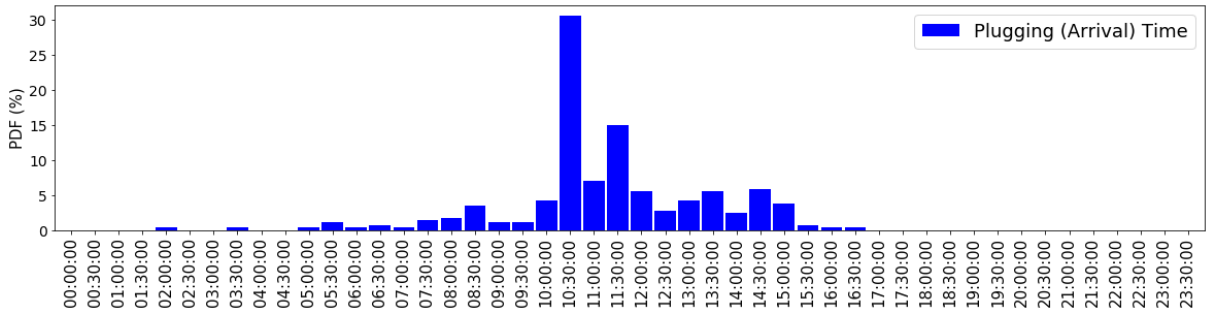

Figure 18. Plugging time in EDYNA parking stations.

Similar to the public stations, here also one can observe the charging duration is partially tied to the plugging time as is depicted in the Figure 19.

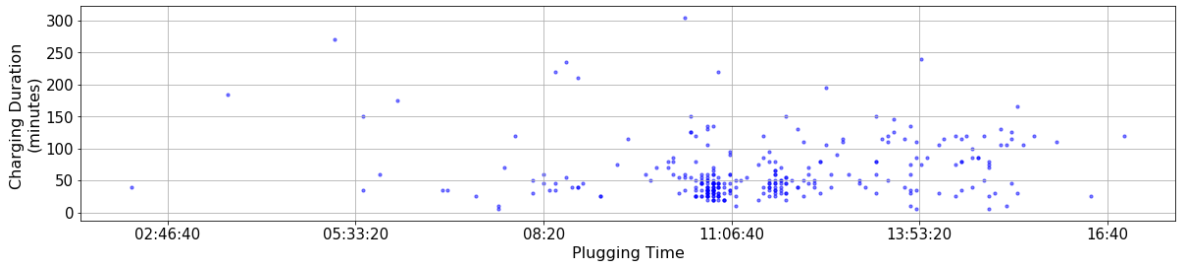

Figure 19. Charging time vs. plugging time.

This fact in other word means there are less deviations from mean values for certain hours. The mean value is depicted in the Figure 20.

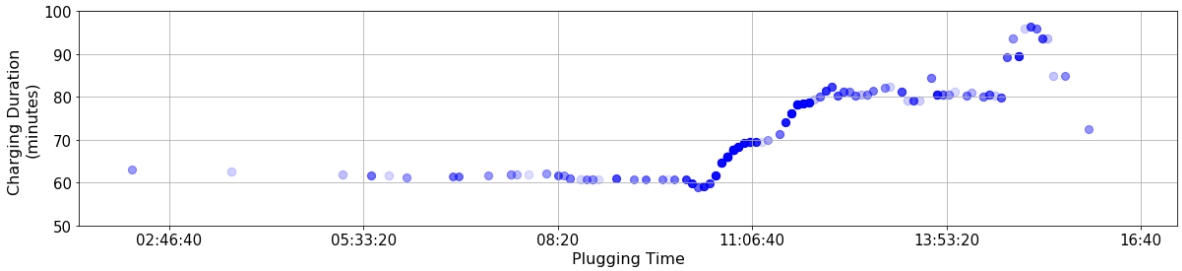

Figure 20. Average of charging duration versus plugging time.

Final result of the prediction is a time-series that is shifted in time each time is called. The Figure 21 depicts 2 examples of the model outputs for day-ahead EV charging load.

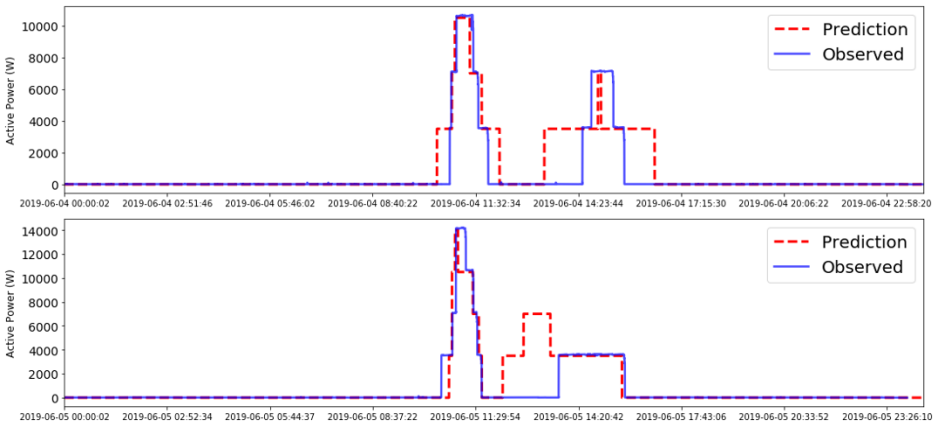

Figure 21. Examples of day-ahead forecast for two consecutive days.

The rolling horizon prediction mode allows to have a wider forecast window and taking into account the events up to the moment of query, as the sample is brought in the Figure 22.

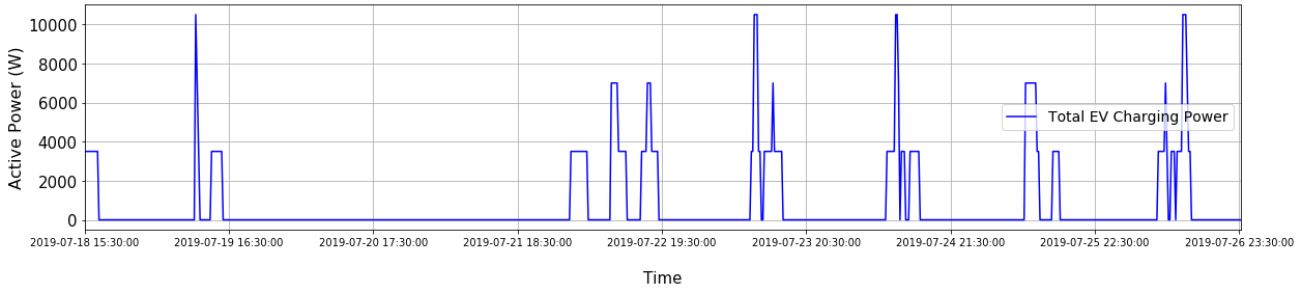

Figure 22. EV load rolling horizon forecast for desired time ahead.

3.2.3 Single House Load

Treating electricity consumption of a single user requires different measures than the previous two addressed cases due to presence of high random noise. The solutions in the literature are many e.g. classic "following pattern", predictive models or "energy" analytics rather than "power" [5].
In each of those many applied cases there are pros and cons, and specific S4G case entails application of methods with lower dependency on a huge recorded dataset. To this end, a method with both "classification" and "regression" is adopted to cope with the present task.

3.2.3.1 Classification

In a first stage, the model tries to classify the loads applying Non-Intrusive Load Monitoring (NILM) concept. The monitored loads in the test sites are not labelled in the S4G case, however the name of specific appliance is not of our interest. Therefore, dataset is being labelled in a preliminary process, to undergo later into a classification routine. In the labelling process, labeller class detects various appliances switching ON and OFF, and builds a dictionary with the following naming convention: as a prefix "UNK" and following a number that is in fact an index from a list of appliances ranked by the nominal power in decreasing order. Some features that come in the following are considered as signature of those loads:

- First derivative
- Second derivative
- Convolution of the first derivative with a variable impulse signal

The characteristics of the impulse signal is determined by the first derivatives magnitude. An optimizer tends to keep as minimum as possible the number of appliances in the dictionary, given that very sensitive model might label one device more than one time with a minor change in the features, this can happen especially for frequency and voltage dependent loads.

The resulting dictionary contains some appliances marked as UNK1, UNK2, …, UNKn, for each of those appliances in the dictionary, the above-mentioned signatures feature it uniquely, and also register the switching ON/OFF time stamp.

After labels are available, the dataset is fed to a bidirectional Long Short-Term Memory (LSTM) network to training model to detect loads' classes out of input time-series. Relying on the limited available data, the

models detects only loads with high power margins, this issue will be resolves gradually as the dataset gets enriched.

3.2.3.2 Regression

The output of the classification that is the probability of a set of appliances' functioning undergoes to a regressive model, which tries to correlate each appliance's working to a set of external/environmental drivers and exciter such as temperature, light, humidity and weekday. The number of drivers will raise in the future version, as various factor will be investigated.

The load prediction then starts by inserting known data for the entire forecast horizon. Weather data are provided by the 3rd party weather forecast, external light by the PV predictive model, and the weekday is known once forecast horizon is defined.

3.3 ESS Status Predictive model

The ESS status prediction is trivially correlated to the load consumption and PV production as the gathered data suggest. A linear regression problem that correlates simply the battery system's power to the PV power generation and load consumption may solve the problem. Starting to explore the right features, a regressive model is tried with inputs only as PV power generation and load consumption. Although this simple model returns accurate results during the day –where there is power produced from PV – the error raises when there is low PV power.
In the following, an example period and various attempts for battery power forecast are demonstrated. Figure 23 shows a monitored house PV power generation and loads during example period (more than 24 hours).

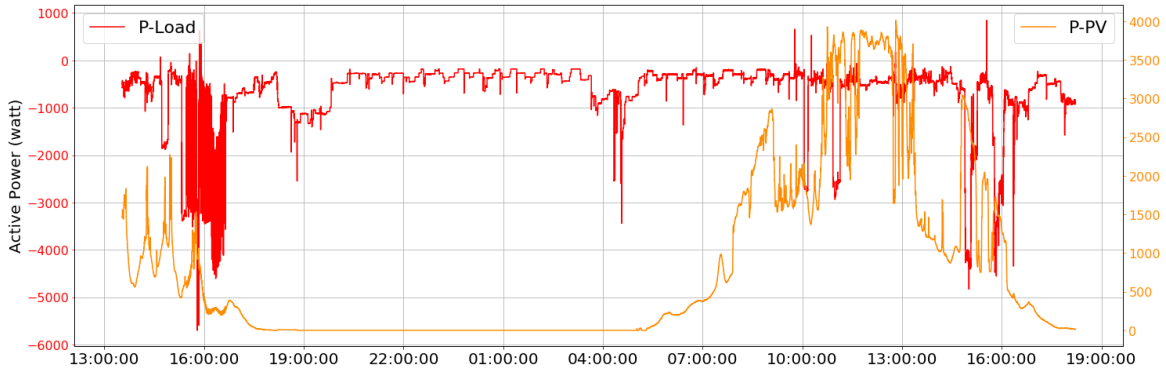

Figure 23. A generic time period of monitored PV generated power and consumed power in a residential building.

The prediction with only Load and PV power results with good quality of forecast during the day but wrong values during night.

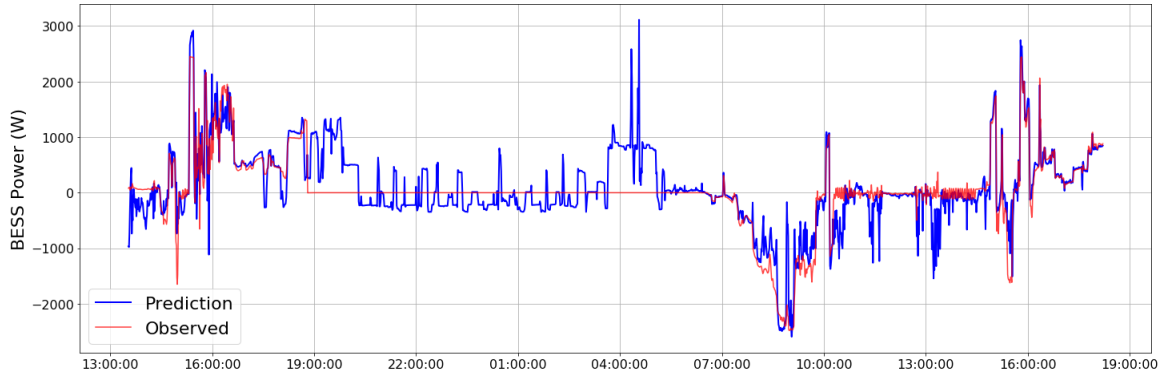

Figure 24. Prediction with PV and Load as features.

Adding State-of-the-Charge (SoC) to the features improves the model's accuracy. As a result, the error drops during no PV power available time and the Mean Square Error –as evaluation variable – is diminishing sharply from 135 (watt) to less than 25 for the considered period. However, this cannot be the desired model since the SoC itself is a state variable. Therefore, an additional measure must be taken into account, e.g. adding an additional vector that defines no-PV power time slots or a recurrent component could be added to give the evolution of SoC, but only from the latest sample i.e. a simple recurrent model with no-memory. The result gets better just by adding SoC as additional feature, this can be seen in Figure 25.

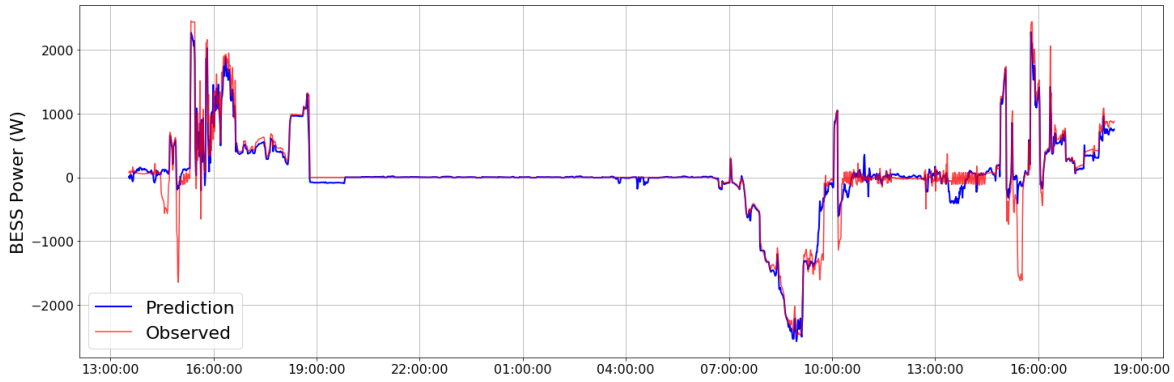

Figure 25. Prediction with PV, Load and SoC as features with recurrent component.

Once the battery status interpreted as the power, being in charging, discharging or idle mode is predicted with a relatively good accuracy, its integral, saturated between zero and one will return the battery's SoC during the prediction horizon.

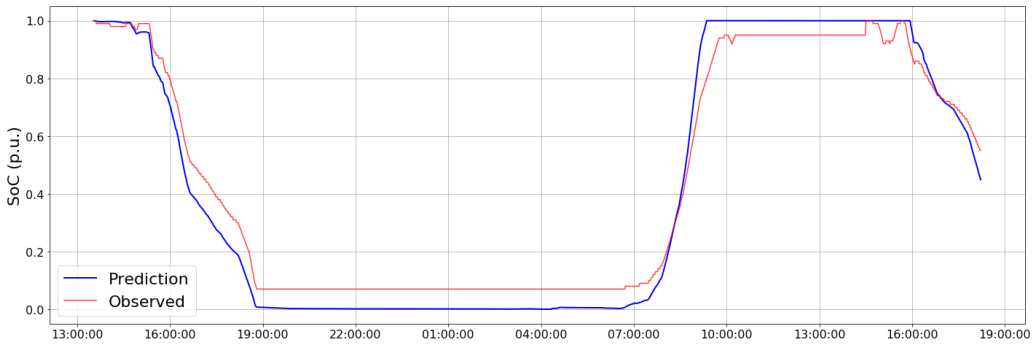

Figure 26. SoC prediction out of battery's power prediction.

4 Conclusions

This document "D5.7 – Final DSF Predictive Models" describes the final algorithm and test results of BMS real-time measurement-based SoH and RUL estimation, and large historical data-based PV, load and ESS status forecasts. SoH and RUL are related to use cases HLUC2-PUC-2 and HLUC3-PUC-3, where PV, load and ESS status predictive model are supporting all GESSCon related activities, which covers all HLUC-2 and HLUC-3 use cases.

A final evaluation "D6.12– Phase 3 Evaluation Report" will be presented in M39 which evaluates the quality of the predictive models' performance in Bolzano sites, Skive site and in the integration with DSF-SE simulation case. The current developed modules will be updated for further integration purpose within Storage4Grid project.

Acronyms

Acronym	Explanation
BESS	Battery Energy Storage System
BMS	Battery Management System
CMU	Cell Monitoring Units
DSF	Decision Support Framework
DSO	Distribution System Operator
DWH	Data Warehouse
ESS	Energy Storage System
GESSCon	Grid ESS Controller
OCV	Open Circuit Voltage
PROFESS	Professional Realtime Optimization Framework for Energy Storage Systems
PV	Photovoltaic
RES	Renewable Energy Sources
RUL	Remaining Useful Life
S4G	Storage4Grid
SEI	Solid Electrolyte Interphase
SMX	Smart Meter eXtension
SoC	State of Charge
SoH	State of Health
LFP	Lithium iron phosphate battery
LSTM	Long Short Term Memory

List of figures

References

[1] E. Delzendeh, S. Wu, A. Lee, Y. Zhou, "The impact of occupants' behaviours on building energy analysis: A research review". Renewable and Sustainable Energy Reviews, Vol. 80, 2017, p.p 1061-1071, doi: https://doi.org/10.1016/j.rser.2017.05.264

[2] American Society of Heating, Refrigerating and Air-Conditioning Engineers. (2009). 2009 ASHRAE handbook: Fundamentals. Atlanta, GA: American Society of Heating, Refrigeration and Air-Conditioning Engineers.

[3] Carpaneto, E & Chicco, Gianfranco. (2008). "Probabilistic characterisation of the aggregated residential load patterns". Generation, Transmission & Distribution, IET. 2. 373 - 382. 10.1049/iet-gtd:20070280.

[4] Panahi, Delshad & Deilami, Sara & A.S. Masoum, Mohammad. (2015). "Forecasting Plug-In Electric Vehicles Load Profile Using Artificial Neural Networks". 10.1109/AUPEC.2015.7324879.

[5] X. M. Zhang, K. Grolinger, Miriam A. M. Capretz. (2018). "Forecasting Residential Energy Consumption Using Support Vector Regressions", DOI: 10.1109/ICMLA.2018.00024.