Verification of services within EcoGrid 2.0

June 2019

EcoGrid 2.0 is a research and demonstration project funded by EUDP (Energiteknologisk Udviklings- og Demonstrationsprogram). The 9 partners in the project are:

Verification of services within EcoGrid 2.0

Daniel Esteban Morales Bondy, Emil M. Larsen,
Charalampos Ziras, Carsten Heinrich

June 2019

Contents

Chapter 1

Introduction

The following report is an expansion to the Deliverable 4.2.2 - Tool for Market interaction and service delivery verification. While the original deliverable outlined theoretically how the verification of services was to be carried out, this deliverable will present the application of this method to the services delivered during the second heating season of the project. It also presents adjustments to the original service verification definition, as well as alternative methods of calculating the baseline of the units; The baselines created for IBM and Insero aggregators are mathematically different, but the interpretation is broadly similar.

The purpose of this report is to answer the following questions:

- How much uncertainty did we observe for tests in HS2?
- How does the metering interval affect results?
- What are the pros and cons of different baseline methodologies?

Chapter 2

Background

Services delivered

During the second heating season a series of tests were carried out in order to evaluate the flexibility of available within the pool of houses controlled by two aggregators, Insero and IBM. Towards the end of the heating season, seven service tests for Insero were carried out and 78 for IBM, where a specific set of service parameters were bid, and the corresponding activation of service was attempted.

The offered services follow the definition of Load Reduction services presented in Chapter 6.2 of Deliverable 2.2, i.e. each service requires a response magnitude, response duration, rebound magnitude, rebound duration, and date/time of the service delivery. For all of the Insero tests and a majority of the IBM tests analysed here, a well-defined rebound magnitude and duration was not targeted. These scheduled activation tests without rebound are included anyway to create a big enough dataset to identify statistically significant differences in activation when analysed from a 5-minute, 15-minute and hourly resolution.

The service parameters are summarised in Tables 2.1-2.3 for Insero and IBM tests respectively. In the case of Insero, the bid response was to be held for one hour, and the corresponding rebound is assumed to be held for half an hour and half the magnitude of the response. In the case of IBM, the bid response varies between a half hour and a whole hour, and the rebound response magnitude was assumed to be one third of the initial response, lasting for the same duration as the initial response.[1]. The Insero table additionally presents the results of the control function, which estimates how many houses were needed to achieve the response[2] as well as the added safety margin due to uncertainty in communication with the houses. For IBM tests, this safety margin was not available.

Table 2.1: Overview of services delivered during the second heating season by Insero

ID	Date	Time (UTC)	Response	Rebound	Houses needed	Houses with 15% overhead
Test0	13/02/2018	18:00	70	35	140	165
Test1	14/02/2018	07:00	80	40	123	145
Test2	14/02/2018	19:00	60	30	133	156
Test3	15/02/2018	08:00	80	40	123	145
Test4	15/02/2018	20:00	50	25	152	179
Test5	16/02/2018	09:00	70	35	135	159
Test6	16/02/2018	21:00	50	25	167	196

For the IBM experiments, only the expected response has been registered, but a few of the experiments also included testing for ramping of the service delivery.

[1] In this set of experiments the rebound was uncontrolled, and the rebound magnitude and duration have been assumed in order to carry out the verification work

[2] This was calculated by dividing the response magnitude with an estimate of how much flexibility each house could deliver (measured in kW/house). This estimate was found through data analysis of the results of the tests carried out earlier in the heating season.

Table 2.2: Overview of services delivered during the second heating season by IBM

ID	Date	Time (UTC)	Response	Houses	Ramp	Min throttling period	Max throttling period
IBM 1	23/01/2018	13:00	180				
IBM 2	24/01/2018	03:00	150	135			
IBM 3	24/01/2018	14:00	100	100			
IBM 4	25/01/2018	08:30	130	125			
IBM 5	25/01/2018	21:00	205	179			
IBM 6	26/01/2018	04:00	205	179			
IBM 7	26/01/2018	15:00	205	179			
IBM 8	27/01/2018	08:00	205	179			
IBM 9	29/01/2018	10:00	205	179			
IBM 10	29/01/2018	17:00	30	41			
IBM 11	30/01/2018	01:00	70	62			
IBM 12	30/01/2018	11:00	180	170			
IBM 13	06/02/2018	09:00	100	66			
IBM 14	06/02/2018	17:00	100	178			
IBM 15	06/02/2018	00:00	200	131			
IBM 16	07/02/2018	08:00	150	100			
IBM 17	07/02/2018	20:00	290	185			
IBM 18	08/02/2018	04:00	260	170			
IBM 19	08/02/2018	17:00	100	195			
IBM 20	08/02/2018	21:00	100	67			
IBM 21	09/02/2018	11:00	100	67			
IBM 22	09/02/2018	13:00	80	195			
IBM 23	09/02/2018	21:00	120	195			
IBM 24	10/02/2018	02:00	250	170			
IBM 25	10/02/2018	07:00	140	195			
IBM 26	10/02/2018	12:00	200	143			
IBM 27	10/02/2018	14:00	120	195			
IBM 28	10/02/2018	22:00	100	195			
IBM 29	11/02/2018	03:00	150	114			
IBM 30	12/02/2018	18:00	150	89			
IBM 31	13/02/2018	04:00	200	117			
IBM 32	14/02/2018	14:00	240	144			
IBM 33	14/02/2018	22:00	280	167			
IBM 34	15/02/2018	08:00	240	145			
IBM 35	15/02/2018	20:00	230	149			
IBM 36	19/02/2018	17:00	95	149			
IBM 37	19/02/2018	20:00	200	145			
IBM 38	20/02/2018	06:00	200	135			
IBM 39	20/02/2018	08:00	100	149			
IBM 40	20/02/2018	18:00	200	143			
IBM 41	20/02/2018	20:00	100	149			
IBM 42	20/02/2018	00:00	100	151			
IBM 43	21/02/2018	03:00	200	121			
IBM 44	21/02/2018	10:00	100	151			
IBM 45	21/02/2018	15:00	200	138			

Table 2.3: Overview of services delivered during the second heating season by IBM (continued)

ID	Date	Time (UTC)	Response	Houses	Ramp	Min throttling period	Max throttling period
IBM 46	22/02/2018	21:00	115	152			
IBM 47	22/02/2018	23:00	240	149			
IBM 48	23/02/2018	04:00	115	152			
IBM 49	23/02/2018	10:00	200	123			
IBM 50	23/02/2018	15:00	100	151			
IBM 51	23/02/2018	22:00	220	150			
IBM 52	24/02/2018	04:00	120	152			
IBM 53	24/02/2018	06:00	260	152			
IBM 54	24/02/2018	20:00	258	151			
IBM 55	24/02/2018	21:00	124	152			
IBM 56	25/02/2018	05:10	130	152			
IBM 57	16/03/2018	16:00	50	149			
IBM 58	16/03/2018	23:00	80	149			
IBM 59	17/03/2018	09:00	70	149			
IBM 60	17/03/2018	16:00	50	149			
IBM 61	20/03/2018	16:00	50	152			
IBM 62	26/03/2018	10:00	175	147			
IBM 63	27/03/2018	09:00	170	145			
IBM 64	28/03/2018	11:00	175	141			
IBM 65	28/03/2018	20:00	175	141			
IBM 66	03/04/2018	20:00	148	147			
IBM 67	04/04/2018	06:00	148	147			
IBM 68	04/04/2018	15:00	135	145			
IBM 69	05/04/2018	09:00	135	145			
IBM 70	06/04/2018	09:00	145	138			
IBM 71	06/04/2018	20:00	145	138	10	60	105
IBM 72	07/04/2018	08:00	125	118			
IBM 73	07/04/2018	01:00	125	118	10	60	105
IBM 74	08/04/2018	09:00	90	112			
IBM 75	09/04/2018	03:00	90	112	10	60	105
IBM 76	15/04/2018	02:00	165	139	10	45	105
IBM 77	16/04/2018	02:00	165	140	10	45	105
IBM 78	18/04/2018	02:00	165	141	10	45	105

Available data

For each of the services the following data was collected:

- 5-minute energy-meter values of each household
- A list of which houses Insero was able to connect to
- A list of which houses IBM was able to connect to
- The expected response of each test
- In 6 of IBM's tests data for ramp limitiation, minimum throttling period and maximum throttling period were also registered

Chapter 3

Method

3.1 Verification procedure

In order to carry out the verification of the individual service delivery, the following steps are followed:

1. **Pooling of resources:** We need to know from the aggregator which units formed part of the pool that delivered the service. This is needed in order to derive appropriate baselines for the aggregator. It must be noted that it is the whole pool that was bid, not only those that responded or were controlled.

2. **Baseline calculation:** A baseline is calculated from the aggregator's resource pool (different proposals are shown in Sec. 3.2). An associated uncertainty to the baseline must also be estimated. These two, along the with the service contract, form the service model shown in Fig. 3.1 .

3. **Data analysis:** The smart-meter measurements for the aggregator pool are collected, summed up, and the resulting signal is compared to the service model time series.

4. **Performance evaluation:** The performance metrics of the service are calculated from the results of the data analysis.

Service Model and Performance Evaluation

In [1], the definition of the EcoGrid 2.0 services were refined in order to capture uncertainties in service delivery. An example of a load reduction service can be seen in Fig. 3.1, where it shows that the uncertainty in baseline[1] $\sigma_{up,down}$ must be added to the ideal response. Thus, an error function can be derived for the response, which takes the shape sketched in Fig. 3.2.

The performance measure used to evaluate the service delivery is the following has been reproduced here, with a slight alteration to notation:

$$\eta_{tot} = 1 - \left(\frac{1}{T} \sum_{t=1}^{n} f\left(P_{ideal,t} - P_t\right) \right) \tag{3.1}$$

where the error function $f(x)$ is sketched in Fig. 3.2, $P_{ideal,t}$ is the ideal service delivery defined by the contract, and P_t is the actual delivery by the aggregator. This definition means that the service performance ranges from $\eta_{tot} \approx 0$ (bad) to $\eta_{tot} \approx 1$ (good). The reason for using this metric is that we can capture the service quality into a single number, making it easy a verification entity to establish if the service was delivered or not. The minimum η_{tot} value that a service must comply with must be defined in the contract when the services are acquired.

Metrics for evaluating different metering sample rates

While the method defined above serves a single value to evaluate how well the service is provided within the acceptable error bounds (ϵ), we also want to analyse how much information about the service is lost when sampling at 15 minutes resolution, or 1 hour resolution, instead of the 5 minute measurements we currently have in the EcoGrid 2.0 setup.

[1] In the case where operational schedules are used instead of baselines, this uncertainty is defined as the uncertainty in the operational schedule.

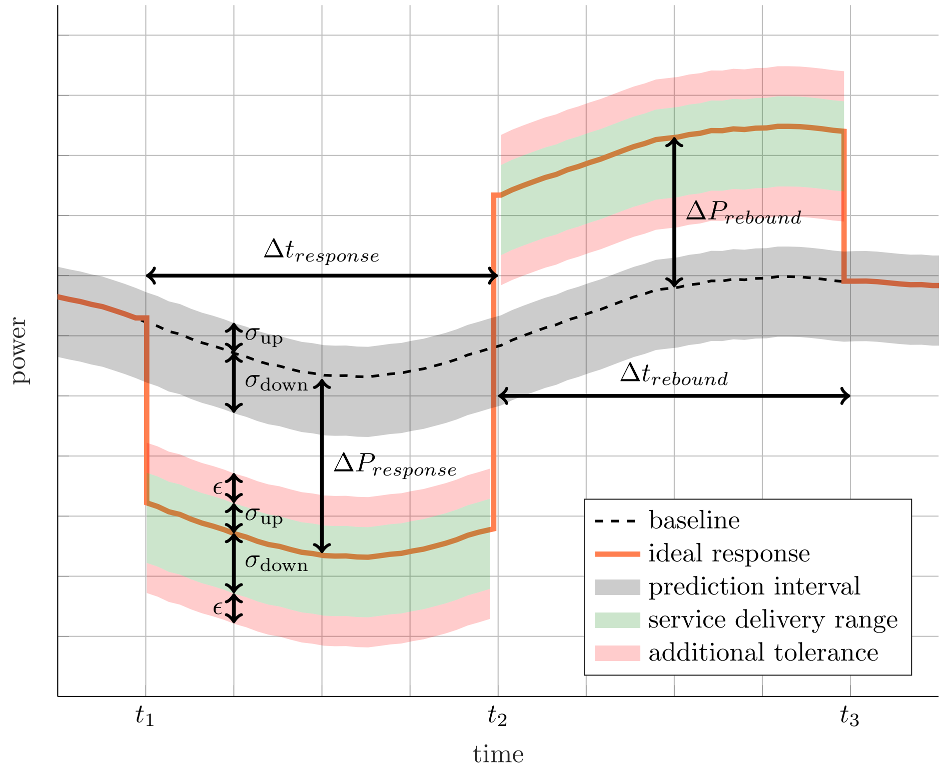

Figure 3.1: A time-series sketch of a load reduction with its associated rebound, uncertainty intervals, and acceptable service delivery.

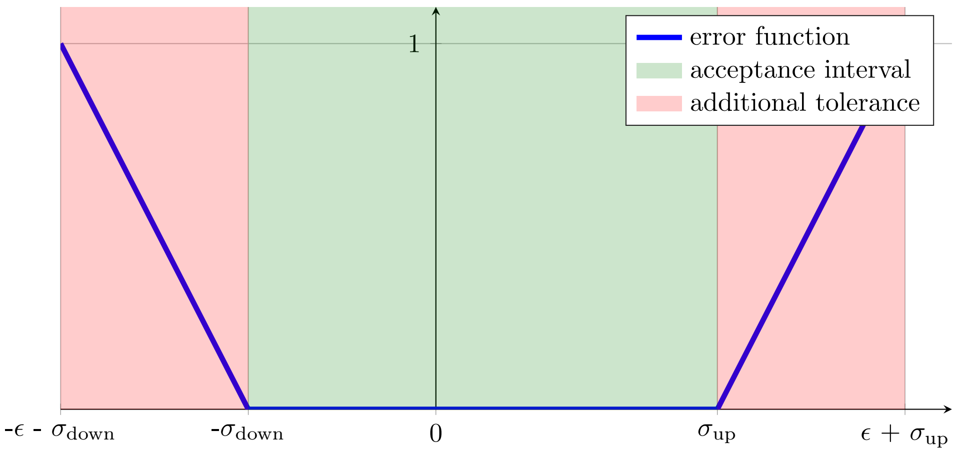

Figure 3.2: Error function used to evaluate the service performance.

In order to carry out this evaluation, we will look at the maximum and minimum of the aggregated power measurements of the portfolio. This can be defined as:

$$Peak = \max_{t\in[T]} P_t, \tag{3.2}$$

$$Dip = \min_{t\in[T]} P_t, \tag{3.3}$$

$$P_t = \sum_{h\in[H]} p_{h,t}, \tag{3.4}$$

where $p_{h,t}$ is the power measurement of a house h at time t, with $[T]$ being the set of time samples in the data set and $[H]$ the set of houses conforming that aggregation pool. Thus, P_t is the aggregation of all the household measurements at time t.

Furthermore, we use the Mean Absolute Error (MAE) in order to identify how much power is being distorted away with the resampling:

$$MAE = \frac{\sum_{t=1}^{T} |P_{ideal,t} - P_t|}{T}, \tag{3.5}$$

where $P_{ideal,t}$ is the optimal response (at time t) as defined by the service model (see e.g. the orange line in Fig. 3.1).

3.2 Definition of baseline methods

3.2.1 Insero baselines with the Hybrid Method

This baseline approach calculates profiles exclusively for the service delivery period. This time interval is the only period when the actual power consumption of the portfolio delivering the service is unknown. Baselines are calculated for each aggregation pool, and must therefore be recalculated when the composition of the aggregation pool changes. The method is thoroughly described in [2].

Table 3.1: Overview of the different baseline methods used for the hybrid method.

Method	Description
M1	Linear Interpolation
M2	Forward-Backward Auto regression
M3	Decomposition + Ignoring Residual
M4	Decomposition + FBA on Residual
M5	Decomposition + Linear Regression on Residual
M6	Hybrid

The procedure is as follows:

1. In the first step, a historic data set is created, representing the dynamics of the load, during times when the aggregator did not control the portfolio. To do this, all periods, when the aggregator was applying active control of the portfolio, are replaced. For this first step, a simple autoregressive forwards-backward (FBA) model is used.

2. The available historic data is split into training and test set (80% /20%). Five different baseline approaches are defined and optimised on the training set, using a ten-fold cross validation. The methods are summarised in table 3.1. For this purpose, random time intervals are chosen on the historic training data set, to represent service-periods. The baseline methods are applied to these time intervals and can be directly compared to the real historic data. Finally, the baseline performance is evaluated on random test periods on the test set.

   M1 The first baseline method is linear interpolation. Even though this method is very simple, it uses the information about the power consumption of the aggregation directly before and after the service period. It is in this regards the most simple autoregressive model, and outperforms most models, without any autoregressive component.

M2 The second method is the autoregressive forwards-backward model, which was used to create the initial historic data set. In this case, the FBA model is trained on 18 days before and after the service delivery. The autoregressive model is based on the previous two days - with 5 minute resolution, this corresponds to a model order of 576.

M3 This method relies on decomposition. The data set is split into three parts, a trend component, a seasonal component as well as a residual. To calculate the baseline, the residual is neglected, and the sum of trend component and seasonal component are used as the baseline.

M4 The forth method combines the benefits of decomposition and autoregression. The FBA model is applied to the residual of the decomposition, such that the residual is forecasted during service delivery. Then the sum of trend, seasonal component and forecasted residual are used as a baseline.

M5 The last method again combines decomposition with linear regression. The linear regression is carried out on the residual. Hour of the day, solar radiation and ambient temperature are used as regressors. Among the five methods, this one leads to the best results.

3. In order to improve the performance of the baseline calculation further, the five previously introduced baseline methods are linearly combined, to minimise the total error. The corresponding hybrid method **M6** combines forecasts of the 5 results linearly, the weighted sum is different for every time step, leading to the overall best results.

It is **M6** which is used for the rest of this work when referring to a Hybrid baseline.

3.2.2 IBM baselines with the Elastic Net

The Insero baseline approach requires perfect knowledge of consumption immediately before and after service delivery. The IBM baseline approach is a little simpler, not requiring knowledge of consumption before or after delivery. It has some similarity to steps **M3** and **M5** previously described and is comprehensively described in [3, 4].

The full model for demand can be expressed in linear model form as:

$$c_t = \tilde{z}_t^\intercal \tilde{\theta}_z + \tilde{\chi}_t^\intercal \tilde{\theta}_\chi + \epsilon_t = \boldsymbol{x}_t^\intercal \boldsymbol{\theta} + \epsilon_t \tag{3.6}$$

where c_t is the observed consumption of several smart meters at time t, z is an array of exogenous variables such as weather, holidays, and a Fourier series, and χ is an array of important interactions between exogenous variables. ϵ is the mismatch between the observed and modelled consumption. Assuming the model captures all physical aspects of the model, then ϵ should be normally distributed white noise.

To find the parameters of the model, θ, and ensure that they are not overfitted, an elastic net is employed. The objective of this is

$$\min \sum_{t=1}^{T} (c_t - \boldsymbol{x}_t^\intercal \boldsymbol{\theta})^2 + \lambda(\alpha \boldsymbol{\theta}^2 + (1-\alpha)|\boldsymbol{\theta}|) \tag{3.7}$$

The elastic net builds upon a conventional least squares regression with Lasso ($|\theta|$) and Ridge Regression (θ^2) parameter penalisation, where λ governs the overall strength of the penalisation and α dictates how much of each type of penalisation to use. λ and α are found using a ten-fold cross-validation, performed several times for a range α values between 0 and 1. The combination of λ and α that give the lowest mean square error on the validation fold is chosen before finding θ with the full dataset for training.

3.3 Definition of acceptable uncertainty as a function of portfolio size

The accuracy of the baseline estimate is dependent on the size of the aggregator portfolio. This is due to the baselines being an estimate of what the households would do on average, and following sampling statistics theory, the higher the sampling population, the closer the estimate will be to the true mean. At the same time, if the portfolio is larger, the magnitude of the absolute estimation error will be larger, while the proportional error will decrease. This is represented in Figure 3.3, where the error standard deviation and the per-household error standard deviation are plotted as functions of the portfolio size. The different curves represent the error for the baseline estimation methods described in Sec. 3.2.1.

We use these error estimates to define the acceptable uncertainty in the baseline $\sigma_{up,down}$ shown in Figure 3.2. Specifically, we define $\sigma_{up,down}$ as the band around the baseline estimate that contains 95% of the observations.

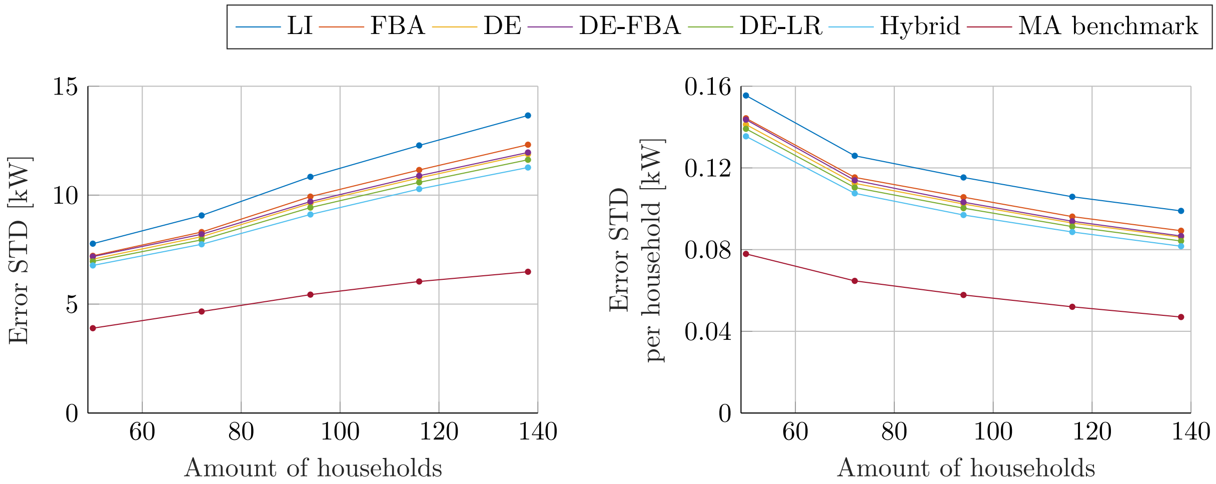

Figure 3.3: Baseline error as a function of portfolio size. Estimated on the Insero pool [2].

3.4 Re-sampling strategy

For all Insero baselines, the baseline model is trained with 5-minute resolution data, since this is also the highest available resolution for the smart meter measurements. In order to evaluate 15-minute sampling and 1-hour sampling, the baselines and data have been resampled by averaging over a 15-minute period or a 1-hour period.

For all IBM baselines, the baseline model is trained on 5-minute, 15-minute and 1-hour data respectively. In theory, the IBM approach should give more accurate 15-minute and 1-hour baselines than the Insero approach, 15-minute and 1-hour baselines are trained on less noisy data, which can result in less overfitting and fewer spurious correlations.

Chapter 4

Results

4.1 Insero verification

4.1.1 Activation overview

The verification method defined in Chapter 3 looks explicitly at the the overall response of the aggregation portfolio in comparison to service model. Before we show the results of this comparison in the next sections, we present an evaluation of the activation of the units in the portfolio. The presented numbers cover the Insero portfolio, since they explicitly distinguish between the optimal activation and the robustness margin, while IBM incorporates robustness directly into their model.

In Table 2.1 we presented the expected activations for each service provided by Insero. In Table 4.1 we compare the expected activated houses against the amount of houses Insero could communicate with.

Table 4.1: Overview of house availability compared to the expected activation for Insero portfolio

ID	Houses needed	Requested with 15% overhead	Houses available	Houses connected
Test 2018-02-15 21:00 Insero	140	165	165	139
Test 2018-02-14 08:00 Insero	123	145	145	114
Test 2018-02-16 10:00 Insero	133	156	156	126
Test 2018-02-13 19:00 Insero	123	145	145	113
Test 2018-02-14 20:00 Insero	152	179	168	140
Test 2018-02-16 22:00 Insero	135	159	149	122
Test 2018-02-15 09:00 Insero	167	196	214	176

From the data it can be seen that adding robustness margins to the activation is necessary since in all services, the number of houses Insero was able to connect to (and we therefore assume they responded to the control signal) was less than the amount of houses that were requested.

4.1.2 Results for 5-minute, 15-minute and 1-hour intervals

Figures 4.1-4.2 give a visual overview of the baseline time series and the measurements of the response for the first test. The figures present the service response as a deviation from the estimated baseline, along with the bounds defined in Sec. 3.1.

The results of the performance assessment are presented in Table 4.2. In the table, the following information is presented:

- The calculated η_{tot} for each service, for each time resolution,
- The improvement in the performance evaluation by going down in resolution, as a percentage of the 5 minute measurement performance (which gives the most accurate performance evaluation).

The following insights can be drawn from the results:

- The service model is also resampled (as can be seen in the figures) and therefore smoothed out. This leads to the improvement of the service delivery. If all three sampling rates were measured against a single service model sample, e.g. 5 minutes, the results would become worse as the sampling resolution is decreased.

- It can be seen that decreasing the time resolution of the measurements improves the performance of the service delivery.
- With a 15 minute resolution, the response is smoothed out, such that the ramp of the activation can be neglected when carrying out the verification.
- The mean improvement between 5 minute and 15 minute resolution is 2.35%, which means that for this specific metric, the change in resolution does not affect the results significantly.
- The mean improvement between 5 minute and 1 hour resolution is 10.52%, which means that there is a notable difference between the results.
- It seems that the performance difference between the time resolutions is dependent on the performance value itself. When the performance is low, the resampling improves the performance at a greater proportion than when the performance is high.

Table 4.2: η_{tot} results for Insero

Test ID	5 minute measurements	15 minute resampling/ (improvement)	1 hour resampling/ (improvement)
Test 2018-02-13 19:00 Insero	0.5632	0.5856/(4.0%)	0.6633/(17.8%)
Test 2018-02-14 08:00 Insero	0.8411	0.8493/(1.0%)	0.8775/(4.3%)
Test 2018-02-14 20:00 Insero	0.7167	0.7313/(2.0%)	0.7817/(9.1%)
Test 2018-02-15 09:00 Insero	0.8412	0.8494/(1.0%)	0.8776/(4.3%)
Test 2018-02-15 21:00 Insero	0.5778	0.5994/(3.7%)	0.6745/(16.7%)
Test 2018-02-16 10:00 Insero	0.8418	0.8499/(1.0%)	0.8781/(4.3%)
Test 2018-02-16 22:00 Insero	0.5728	0.5947/(3.8%)	0.6707/(17.1%)

*After these results were calculated, it was noticed that the 15-minute and 1-hour resampling results for Insero are incorrect. The mistake has resulted in performance improvements that are too small - i.e. the mean improvement between 5- and 15-minute resolution is greater than 2.35% and the mean improvement between 15-minute and 1-hour resolution is greater than 10.52%.

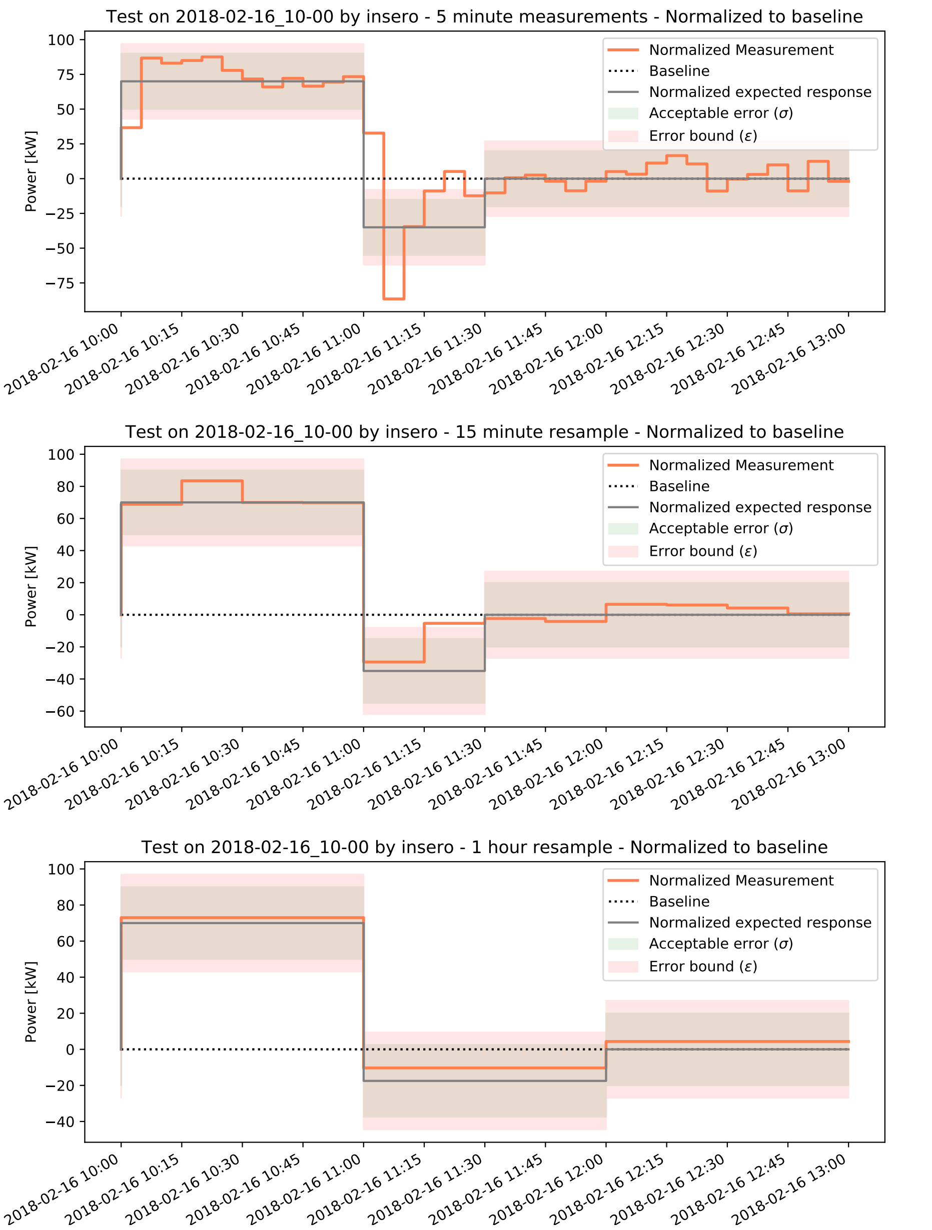

Figure 4.1: Results for the test with the best results, at different sampling rates

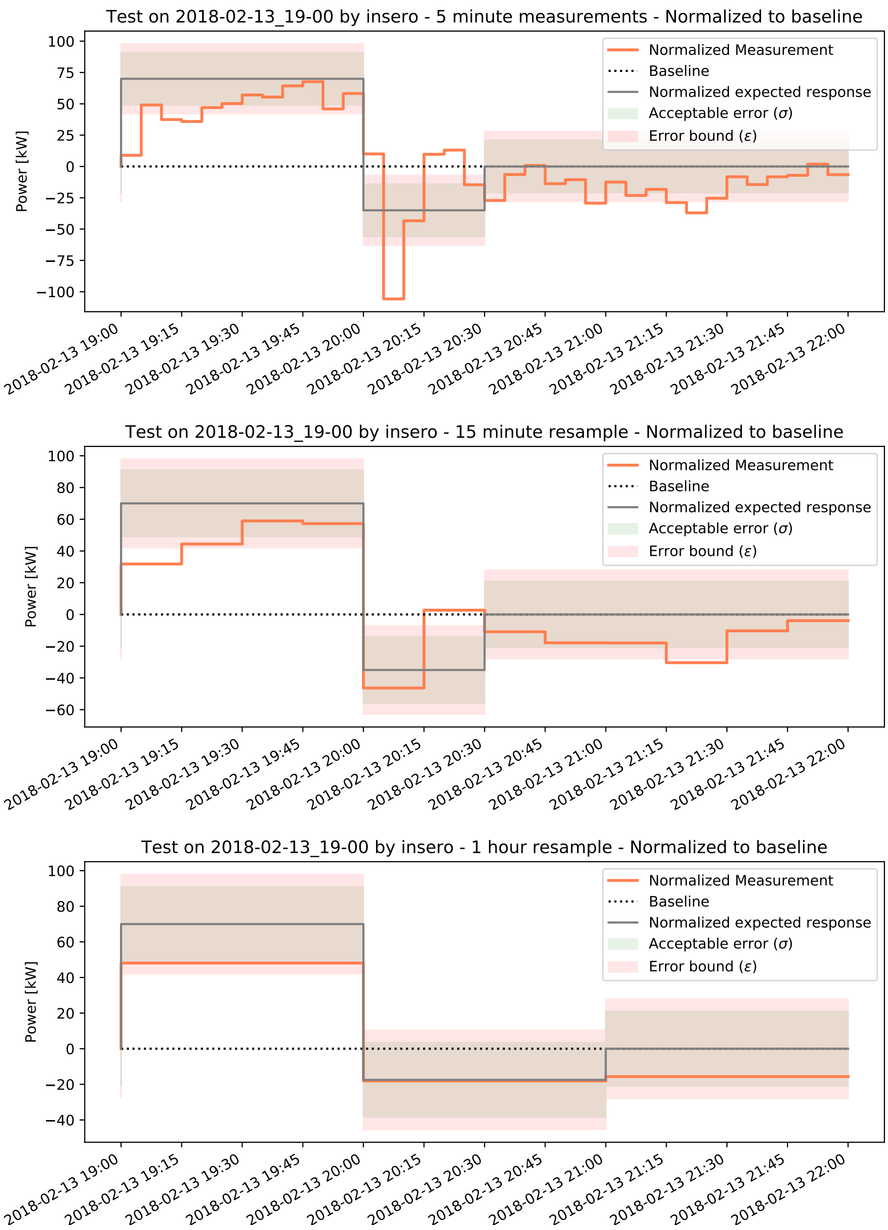

Figure 4.2: Results for the test with the worst results, at different sampling rates

4.2 IBM verification

4.2.1 Results for 5-minute, 15-minute and 1-hour intervals

Figures 4.3-4.4 give a visual overview of the baseline time series and the measurements of the response for two IBM tests. The figures present the service response as a deviation from the estimated baseline, along with the bounds defined in Sec. 3.1. In contrast to the Insero results, the hour before activation is shown, and activation is negative, as the IBM tests are interpretted as a load decrease test rather than a up-regulation test.

The results of the performance assessment are presented in Table 4.4, as with the Insero results. However, one subtle difference is that performance is only evaluated during the activation - not the rebound and before and after the test. On average, the performance metric η_{tot} is 0.78 for 5 minute sampling, 0.83 for 15 minute sampling and 0.91 for hourly sampling.

The following insights can be drawn from the results:

- For individual tests, 15-minute and hourly sampling can give worse results, as the baseline made for different time resolutions has different dynamics, contrary to the Insero approach.
- In general, decreasing the time resolution of the measurements improves the performance of the service delivery.
- IBM activations are less ”peaky” than Insero activations.
- The mean improvement for 15 minute and hourly resolutions is 5% and 14% respectively, which are both significant.
- IBM tests 45-minute and 30-minute activations, that cannot be evaluated with hourly resolution data, as is shown in 4.4.

Table 4.3: η_{tot} results for IBM

Test ID	5 minute measurements	15 minute resampling/ (improvement)	1 hour resampling/ (improvement)
Test 2018-01-23 13-00 IBM	1.0000	1.0000/(0.0%)	1.0000/(0.0%)
Test 2018-01-24 03-00 IBM	1.0000	1.0000/(0.0%)	1.0000/(0.0%)
Test 2018-01-24 14-00 IBM	1.0000	1.0000/(0.0%)	1.0000/(0.0%)
Test 2018-01-25 08-30 IBM	0.4817	0.4906/(1.9%)	nan/(nan%)
Test 2018-01-25 21-00 IBM	1.0000	1.0000/(0.0%)	1.0000/(0.0%)
Test 2018-01-26 15-00 IBM	0.3994	0.4098/(2.6%)	0.4395/(10.1%)
Test 2018-01-27 08-00 IBM	0.7302	0.7334/(0.4%)	0.7406/(1.4%)
Test 2018-01-29 10-00 IBM	0.4281	0.4403/(2.8%)	0.4540/(6.0%)
Test 2018-01-29 17-00 IBM	1.0000	1.0000/(0.0%)	0.1025/(-89.7%)
Test 2018-01-30 01-00 IBM	1.0000	1.0000/(0.0%)	1.0000/(0.0%)
Test 2018-01-30 11-00 IBM	1.0000	1.0000/(0.0%)	1.0000/(0.0%)
Test 2018-02-06 09-00 IBM	1.0000	1.0000/(0.0%)	1.0000/(0.0%)
Test 2018-02-06 17-00 IBM	0.0833	0.0900/(8.0%)	1.0000/(1100.3%)
Test 2018-02-07 00-00 IBM	0.6644	0.0342/(-94.8%)	1.0000/(50.5%)
Test 2018-02-07 08-00 IBM	1.0000	1.0000/(0.0%)	1.0000/(0.0%)
Test 2018-02-07 20-00 IBM	1.0000	1.0000/(0.0%)	1.0000/(0.0%)
Test 2018-02-08 04-00 IBM	1.0000	1.0000/(0.0%)	1.0000/(0.0%)
Test 2018-02-08 17-00 IBM	1.0000	1.0000/(0.0%)	1.0000/(0.0%)
Test 2018-02-08 21-00 IBM	0.6912	0.5435/(-21.4%)	1.0000/(44.7%)
Test 2018-02-09 11-00 IBM	1.0000	1.0000/(0.0%)	1.0000/(0.0%)
Test 2018-02-09 13-00 IBM	0.3211	0.2459/(-23.4%)	1.0000/(211.4%)
Test 2018-02-09 21-00 IBM	0.3441	0.0202/(-94.1%)	0.0147/(-95.7%)
Test 2018-02-10 02-00 IBM	0.0747	1.0000/(1238.8%)	1.0000/(1238.8%)
Test 2018-02-10 07-00 IBM	0.2002	0.1963/(-1.9%)	1.0000/(399.4%)
Test 2018-02-10 12-00 IBM	1.0000	1.0000/(0.0%)	1.0000/(0.0%)
Test 2018-02-10 14-00 IBM	1.0000	1.0000/(0.0%)	1.0000/(0.0%)
Test 2018-02-10 22-00 IBM	1.0000	1.0000/(0.0%)	1.0000/(0.0%)
Test 2018-02-11 03-00 IBM	0.0852	1.0000/(1074.1%)	1.0000/(1074.1%)
Test 2018-02-12 18-00 IBM	1.0000	1.0000/(0.0%)	1.0000/(0.0%)
Test 2018-02-13 04-00 IBM	1.0000	1.0000/(0.0%)	1.0000/(0.0%)
Test 2018-02-14 14-00 IBM	1.0000	1.0000/(0.0%)	1.0000/(0.0%)
Test 2018-02-14 22-00 IBM	1.0000	1.0000/(0.0%)	1.0000/(0.0%)
Test 2018-02-15 08-00 IBM	1.0000	1.0000/(0.0%)	1.0000/(0.0%)
Test 2018-02-15 20-00 IBM	1.0000	1.0000/(0.0%)	1.0000/(0.0%)
Test 2018-02-19 17-00 IBM	0.8364	0.7543/(-9.8%)	1.0000/(19.6%)
Test 2018-02-19 20-00 IBM	1.0000	1.0000/(0.0%)	1.0000/(0.0%)
Test 2018-02-20 06-00 IBM	0.3285	0.2445/(-25.6%)	1.0000/(204.4%)
Test 2018-02-20 08-00 IBM	0.0615	1.0000/(1525.8%)	1.0000/(1525.8%)
Test 2018-02-20 18-00 IBM	0.9180	1.0000/(8.9%)	1.0000/(8.9%)
Test 2018-02-20 20-00 IBM	0.5873	0.5051/(-14.0%)	0.0388/(-93.4%)
Test 2018-02-21 00-00 IBM	1.0000	1.0000/(0.0%)	1.0000/(0.0%)
Test 2018-02-21 03-00 IBM	1.0000	1.0000/(0.0%)	1.0000/(0.0%)
Test 2018-02-21 10-00 IBM	1.0000	1.0000/(0.0%)	1.0000/(0.0%)
Test 2018-02-21 15-00 IBM	1.0000	1.0000/(0.0%)	1.0000/(0.0%)
Test 2018-02-22 21-00 IBM	0.2598	0.0156/(-94.0%)	0.0279/(-89.3%)
Test 2018-02-22 23-00 IBM	1.0000	1.0000/(0.0%)	1.0000/(0.0%)
Test 2018-02-23 04-00 IBM	0.6802	0.7631/(12.2%)	1.0000/(47.0%)
Test 2018-02-23 10-00 IBM	1.0000	1.0000/(0.0%)	1.0000/(0.0%)
Test 2018-02-23 15-00 IBM	0.7554	0.7531/(-0.3%)	1.0000/(32.4%)
Test 2018-02-23 22-00 IBM	1.0000	1.0000/(0.0%)	1.0000/(0.0%)

Table 4.4: η_{tot} results for IBM (continued)

Test ID	5 minute measurements	15 minute resampling/ (improvement)	1 hour resampling/ (improvement)
Test 2018-02-24 04-00 IBM	0.0348	1.0000/(2777.2%)	1.0000/(2777.2%)
Test 2018-02-24 06-00 IBM	1.0000	1.0000/(0.0%)	1.0000/(0.0%)
Test 2018-02-24 20-00 IBM	0.9180	0.7599/(-17.2%)	1.0000/(8.9%)
Test 2018-02-24 21-00 IBM	0.3224	0.4784/(48.4%)	1.0000/(210.1%)
Test 2018-02-25 05-10 IBM	1.0000	0.7015/(-29.9%)	1.0000/(0.0%)
Test 2018-03-16 16-00 IBM	0.3205	1.0000/(212.0%)	1.0000/(212.0%)
Test 2018-03-16 23-00 IBM	0.5207	1.0000/(92.1%)	1.0000/(92.1%)
Test 2018-03-17 09-00 IBM	1.0000	1.0000/(0.0%)	1.0000/(0.0%)
Test 2018-03-17 16-00 IBM	1.0000	1.0000/(0.0%)	1.0000/(0.0%)
Test 2018-03-20 16-00 IBM	0.6621	0.4887/(-26.2%)	1.0000/(51.0%)
Test 2018-03-27 09-00 IBM	1.0000	0.5292/(-47.1%)	0.0263/(-97.4%)
Test 2018-03-28 11-00 IBM	1.0000	1.0000/(0.0%)	1.0000/(0.0%)
Test 2018-03-28 20-00 IBM	1.0000	1.0000/(0.0%)	1.0000/(0.0%)
Test 2018-04-04 06-00 IBM	1.0000	1.0000/(0.0%)	1.0000/(0.0%)
Test 2018-04-04 15-00 IBM	1.0000	1.0000/(0.0%)	1.0000/(0.0%)
Test 2018-04-05 09-00 IBM	1.0000	1.0000/(0.0%)	1.0000/(0.0%)
Test 2018-04-06 09-00 IBM	1.0000	1.0000/(0.0%)	1.0000/(0.0%)
Test 2018-04-06 20-00 IBM	1.0000	1.0000/(0.0%)	1.0000/(0.0%)
Test 2018-04-07 08-00 IBM	1.0000	1.0000/(0.0%)	1.0000/(0.0%)
Test 2018-04-08 01-00 IBM	1.0000	1.0000/(0.0%)	1.0000/(0.0%)
Test 2018-04-08 09-00 IBM	0.1363	0.0393/(-71.1%)	1.0000/(633.7%)
Test 2018-04-09 03-00 IBM	0.0113	1.0000/(8787.9%)	1.0000/(8787.9%)

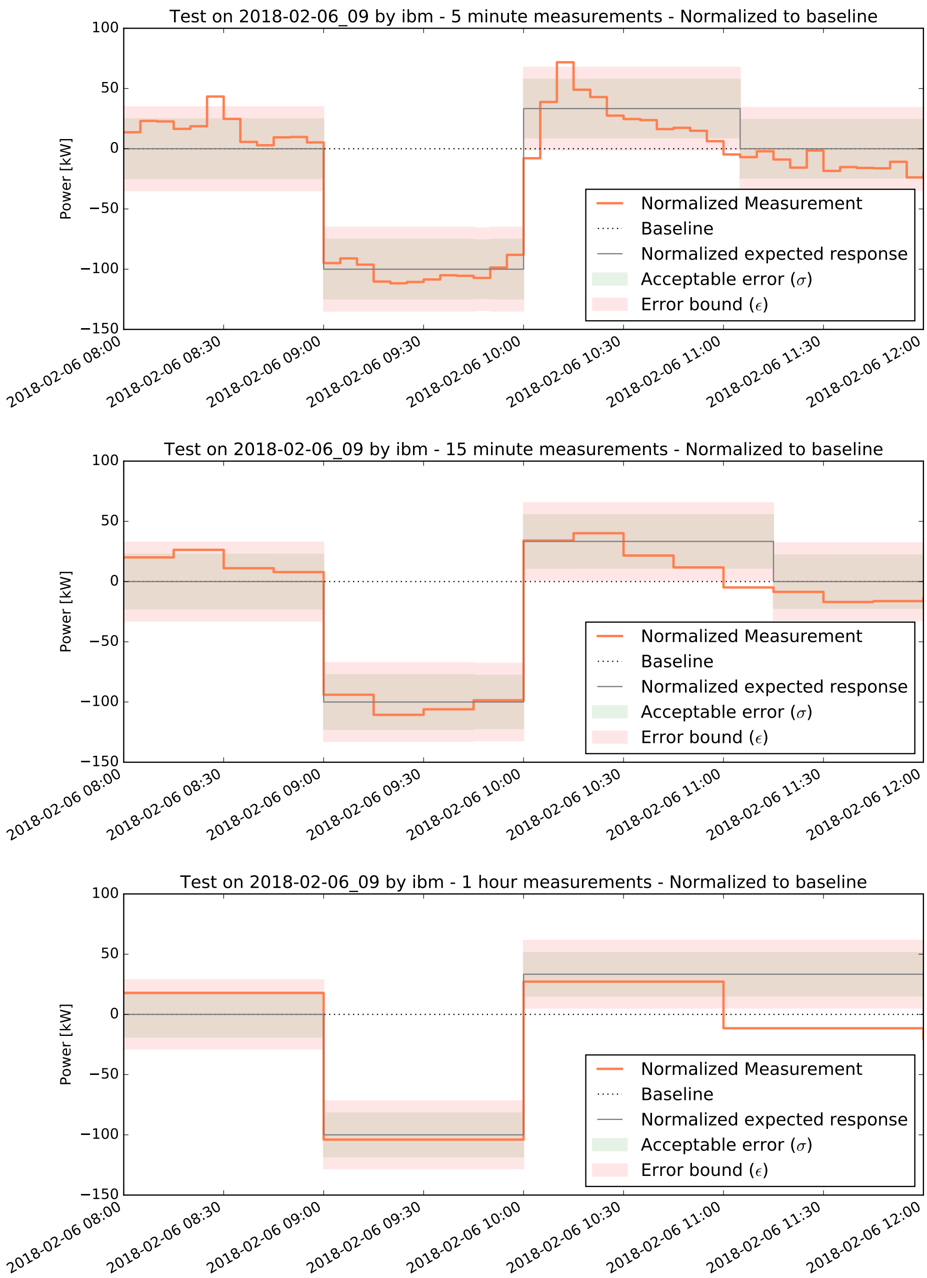

Figure 4.3: Results for the test with the best results, at different sampling rates

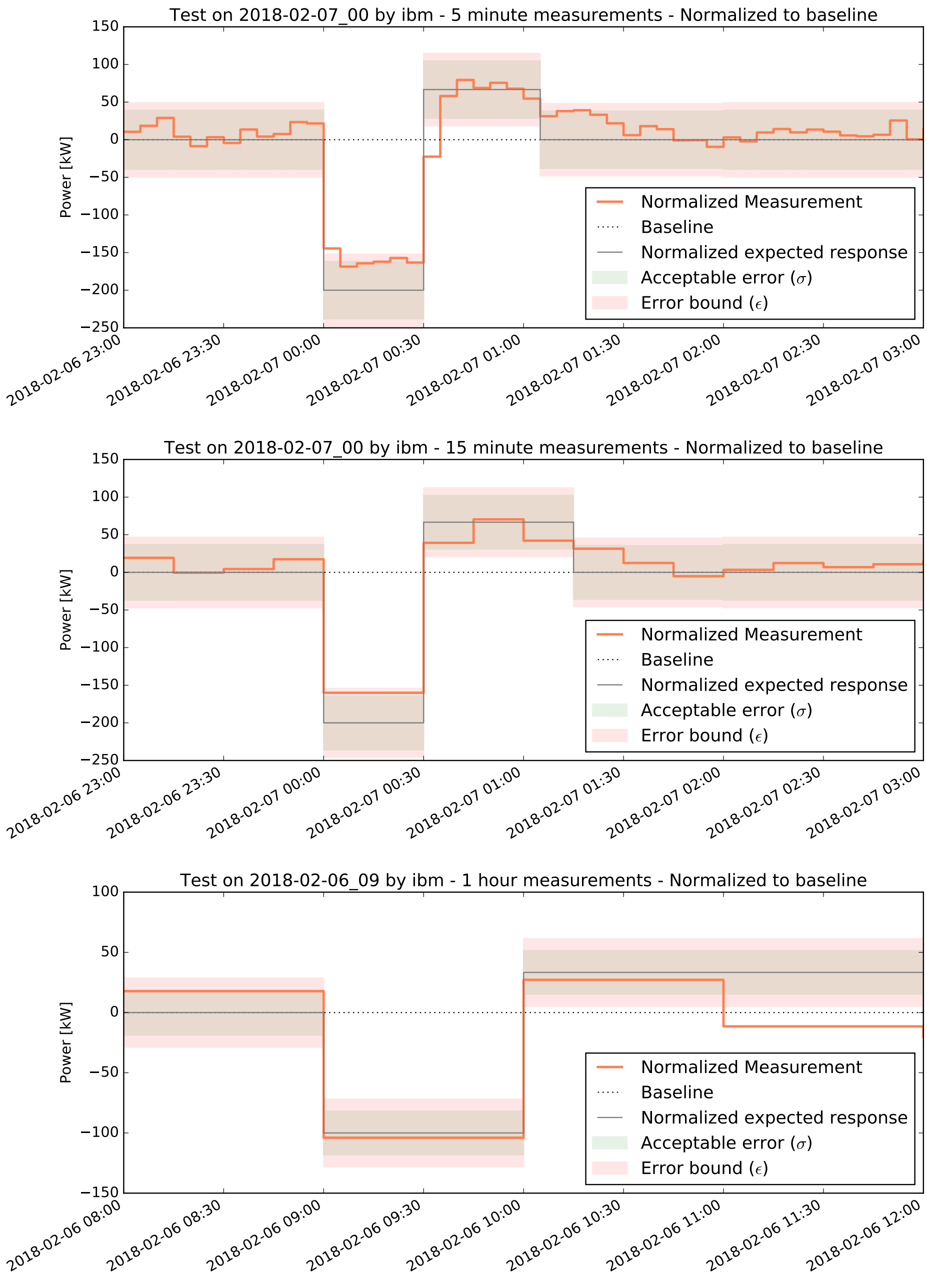

Figure 4.4: Results for an example where the service is less than the sampling rate, which means verification cannot be performed

4.3 Evaluation of sampling rates from a system perspective

In the previous two sections we compare the performance results at different sampling rates. This evaluates how much of the performance we distort by changing sampling rates. In this section we seek to understand how information is lost by the resampling from a system perspective, i.e. if the DSO contracts these services, will a 15 minute sampling rate provide the same information about the system than 5 minute or an hourly rate? This question is relevant since Denmark is in the process of updating its regulation with respect to metering obligations by the DSO. All DSOs must have roled out smart meters by 2020, and its currently under discussion if the metering should be based on hourly, 15 minute, or 5 minute basis.

Table 4.5: Peak and dip analysis of Insero services

Test ID	5 minute peak/dip	15 minute peak/dip (change)	1 hour peak/dip (change)
Test 2018-02-13 19:00 Insero	256.26/96.58	197.15/102.90 (23.06%/6.54%)	161.69/110.61 (36.90%/14.54%)
Test 2018-02-14 08:00 Insero	256.08/76.52	203.50/86.05 (20.53%/12.45%)	169.92/97.43 (33.65%/27.32%)
Test 2018-02-14 20:00 Insero	256.88/77.88	191.31/83.60 (25.53%/7.35%)	166.74/93.79 (35.09%/20.43%)
Test 2018-02-15 09:00 Insero	211.60/76.28	173.88/79.98 (17.82%/4.84%)	162.69/92.26 (23.11%/20.95%)
Test 2018-02-15 21:00 Insero	295.04/80.45	224.10/86.78 (24.05%/7.87%)	192.58/104.52 (34.73%/29.92%)
Test 2018-02-16 10:00 Insero	249.05/80.69	189.70/85.02 (23.83%/5.37%)	189.18/94.43 (24.04%/17.03%)
Test 2018-02-16 22:00 Insero	240.82/74.18	179.88/80.39 (25.30%/8.36%)	171.23/91.71 (28.89%/23.62%)

In Table 4.5 the peak and dips of the response signal over the different time samples is presented. In the worst case, Test 2018-02-14 20:00 Insero, 25% of the peak (256 kW) disappears when the measurements are resampled from 5 minute to 15 minute resolutions and 35% disappears when resampled from 5 minute to 1 hour resolution. These volumes can be significant for a DSO, in the case where a large part of the pool is under a single feeder, yet it may pose no problem if the pool is sufficiently spread.

In order to get a better idea of how the resampling affects the loss of information with respect to service provision, we take a look at the Mean Absolute Error between the baselines and the response. Since the sampling pool of the Insero services is small, we can not do a meaningful statistical analysis of these, so we focus on the 78 IBM tests. In Figure 4.5, it can be seen that the general trend is that the error is higher with 5 minute sampling resolution, and by increasing to 15 minute or 1 hour samples, the error is reduced. This is also what was expected, yet the trend is only apparent when the errors are large. In the lower end of the X-axis it is difficult to distinguish the trend of the three sampling resolutions.

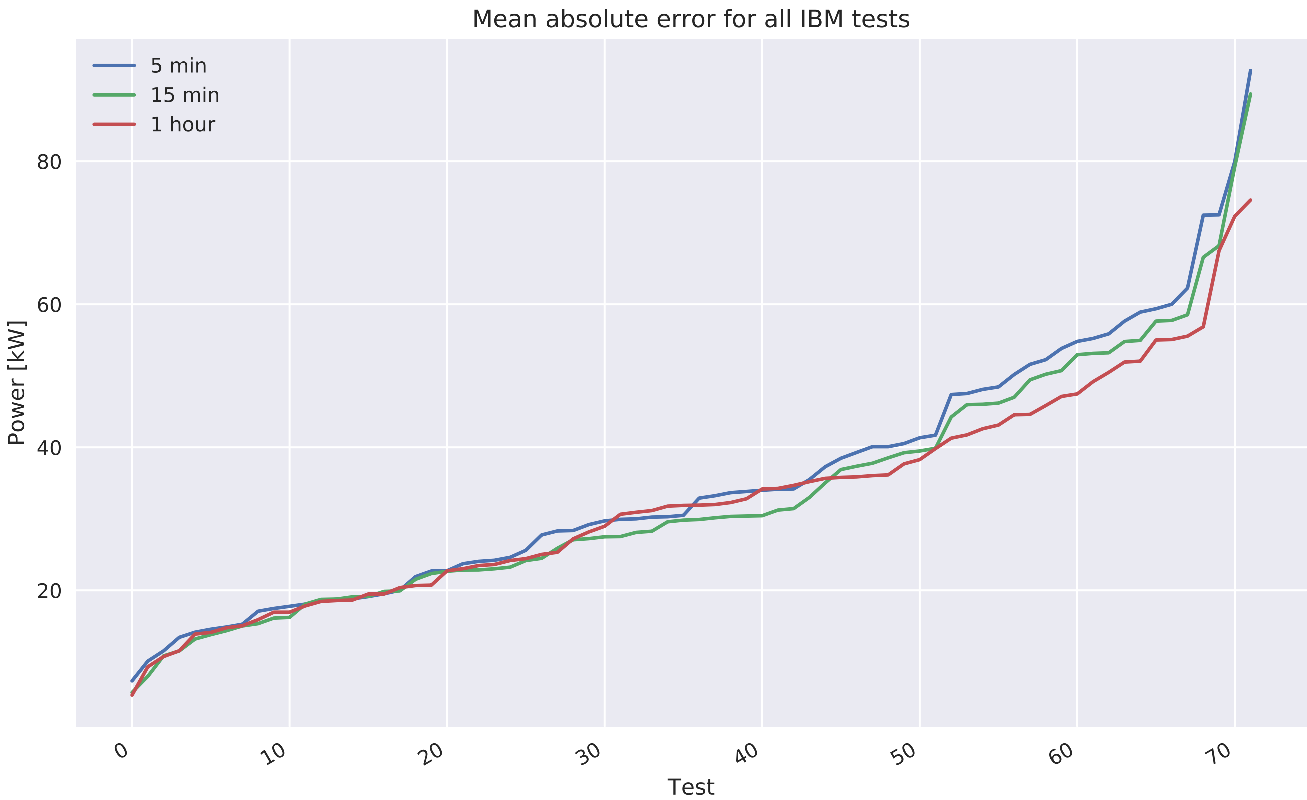

Figure 4.5: Mean average error as a function of metering resolution, estimated on the IBM pool

Chapter 5

Conclusions

In this report we have evaluated all the scheduled activation tests in EcoGrid 2.0's heating season 2. We have applied the methodology developed in WP2 in a systematic manner that would be compatible with a large-scale roll out of DR in Denmark.

5.1 Error discussion

The success of DR acuracy in EcoGrid 2.0 is subject to what a TSO or DSO would consider acceptable, which has not been investigated here. If the service delivery range $\sigma_{up,down}$ is a 95% prediction interval, then η_{tot} should be 0.95 for a 1 hour service with 1 hour resolution data. In this case, it is clear that the aggregators underperformed in heating season 2. Such a conclusion is, however, clearly unfair, as aggregators were continually improving their control algorithms, trying new cordination and communication approaches as is appropriate in an experimental setting.

15-minute and 1-hour resolution data reduces the burden on aggregators significantly, appearing to improve their performance by up to 14%. This is desirable for aggregators, as it reduces the complexity of aggregator coordination. The downside is that up to 25% of the peak disappears when 15-minute sampling is used and up to 35% of the peak disappears when 1-hour sampling is used. This means that short-term overloading may be a cause of concern for system operators.

5.2 Baseline discussion

The verification uncertainty is directly impacted by the accuracy of the baselines used and the period under investigation (before, activation, rebound and after). Increased uncertainty should, however, lead to a wider service delivery range $\sigma_{up,down}$. The baseline approach used for Insero has the benefit of being more accurate, as it draws upon information before and after activation. Its drawbacks are that it allows the aggregator to manipulate their verification results by, for example, pushing a rebound a little before or after the period of investigation, or by causing a congestion that the aggregator is subsequently employed to remedy. The baseline approached used for IBM is less accurate, as it does not rely on future information. It's main benefit is that an aggregator cannot manipulate verification to its advantage and it can investigate rebounds that continue for serveral hours after the initial activation.

5.3 Future research

Understanding what the TSO and DSO can live with of demand response uncertainty to maintain stable systems is an important future research topic. For example, it's perfectly possible that peaks lasting less than 15 minutes are not a problem at low- and medium-voltage levels. Another promising avenue for future research is to investigate the correlation between uncertainties for multiple aggregators delivering services at the same time, as it is possible that errors cancel out to a large extent, giving aggregators addition leeway (ϵ) in their coordination strategies.

Bibliography

[1] C. Heinrich, C. Ziras, D.E.M. Bondy,(2017), *D.4.2.2: Tool for Market Interaction and Service Delivery*, EcoGrid 2.0 Deliverable Report.

[2] C. Ziras, C. Heinrich, M. Pertl, H.W. Bindner, (2018), *Experimental Flexibility Modeling and Identification of Thermal Loads with Limited Information Requirements*, submitted to Applied Energy

[3] E. Larsen, P. Pinson, F. Leimgruber, F. Judex (2017), *Demand Response Evaluation and Forecasting — Methods and Results from the EcoGrid EU Experiment*, Sustainable Energy, Grids and Networks, vol. 10, pp. 75-83. https://doi.org/10.1016/j.segan.2017.03.001

[4] E. Larsen, A. Jónasdóttir, K. Rosenhørn, (2019), *Baselines for Evaluation Demand Response in the EcoGrid 2.0 Project*, 25th International Conference on Electricity Distribution, Madrid