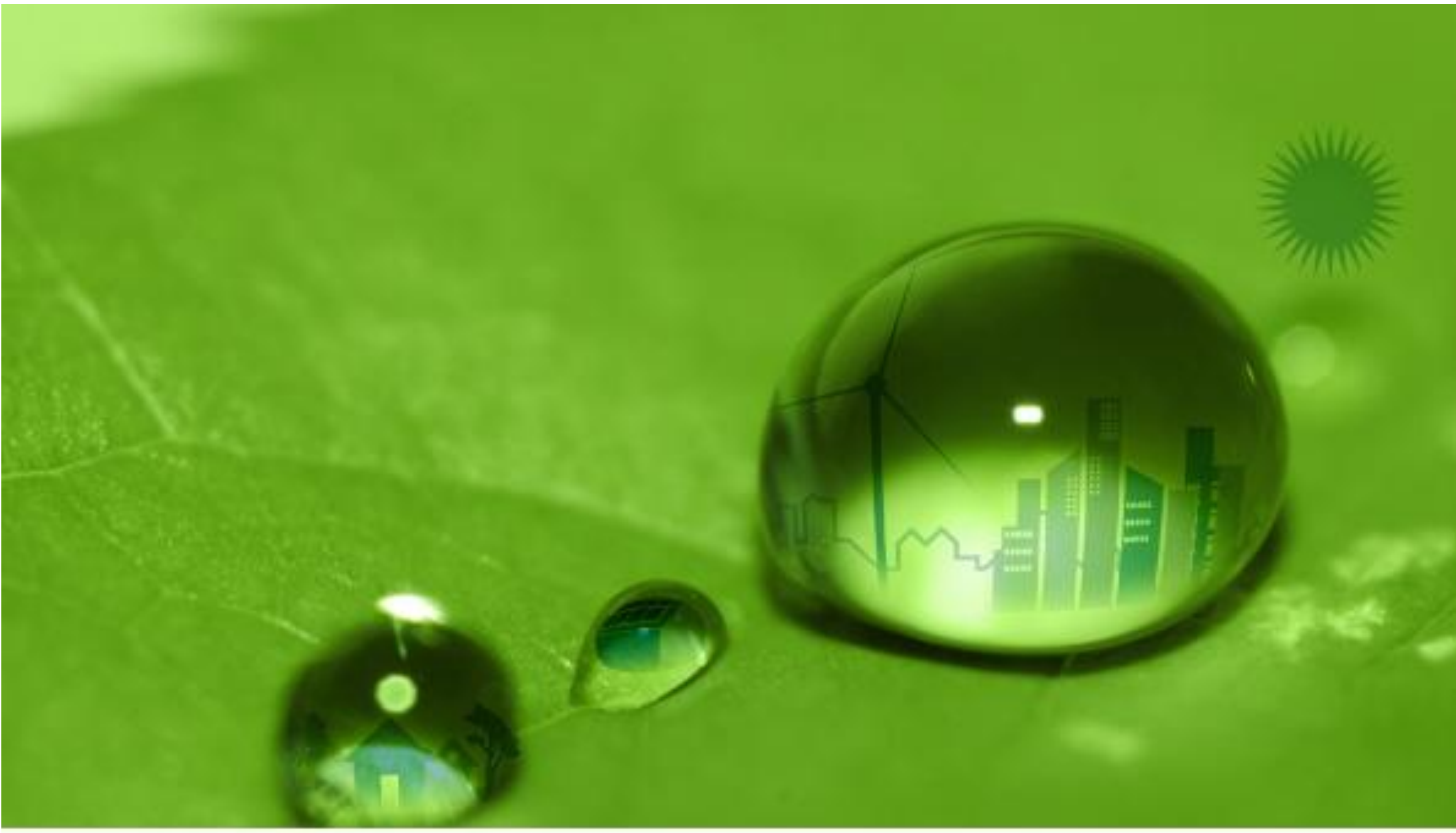

INNOVATION FOR ENERGY NETWORKS

Technical Spotlight DEMO6-S8

Spotlight on the load forecast in DEMO6

CO - FUNDED BY THE EUROPEAN UNION

This project has received funding from the European Union's

Table of content

1. Introduction

1.1 Scope

This document summarizes the work which has been carried out related to load forecasting in the frame of DEMO6. First the initial objectives and technical requirements are recalled, then a brief overview of the software developments and methodology for load forecasting are presented. Finally the main outputs and conclusions are given.

1.2 Tags & metadata (Technical Glossary)

Type of solution						
☐ Equipment / Hardware / Firmware		■ Information system		■ Process		
Manufacturer(s) implied (for equipment or hardware)						
EDF R&D, GE						
Work Stream considered						
■ Active Demand	■ DER integration	☐ Storage	☐ Islanding	☐ MV Innovation	■ LV Innovation	
Location / Topology (with regards to distribution grid)						
☐ HV/MV Substation	☐MV	■ MV/LV SS	■ LV	☐ DER	■ Meter	■ Downstream meter
■ Other Centralized system (calculations, information system)				☐ Other Decentralized system	☐ Other :	
Thematic(s)						
■ Grid Monitoring / state estimation		■ Active demand / DSM		☐ DER Integration / increased grid capacity		
☐ Islanding	☐ Anti Islanding protection		☐ Automatic Failure Detection		☐ Remote Grid Operations	
☐ Automatic Failure Management / Grid recovery		☐ Automatic Grid topology reconfiguration			☐ Other :	
Use Case(s)						
DEMO 6	☐ Islanding of a LV area	■ Peak demand reduction	■ Integrate massive PV production on LV network	■ Encourage resident to adopt smarter habits according to network state		
Key figures ▪ **620** clients below 36 kVA ▪ **5** large clients						

2. Objective and technical requirements

2.1 Context

Load forecasts are a key input to the **NEM** (Network Energy Manager) to determine the **current state of the grid**. They are compared with the **forecasts for the distributed renewable genera**tion and allow anticipating for mobilizing the flexibilities offered by the various aggregators accordingly.

2.2 Requirements

Every day, a processing sequence calculates the **half-hourly consumption forecasts** for that day and the following day. These forecasts concern load curves for around **620** "*tarif bleu*" customers (3-36 kVA), **five** major "*tarif jaune*"[1] customers (42-420 kVA) and six **secondary substations**.

This sequence automatically runs from start to finish every day, from acquiring real-time consumption data in XML format to generating forecasts in XML-**eCIM format** for the **NEM**. It runs every morning at 4:00 AM (local time) in the summer (5:00 AM in the winter) so that the results are ready for the **NEM** before 10:00 AM A Linux service (crontab) is used to run the sequence according to this schedule.

The sequence is comprised of:

- programs written in Java that:
    - convert received data from XML format to CSV format to make it easier to use, and convert forecasting results from CSV format to XML-**eCIM format** for the NEM.
    - generate temperature files (real and forecast temperature hour by hour) from data sent by OSGE (EDF's statistics office).
- a forecasting program written in R which also updates the consumption logs of Linky meters and SME (Small and Medium-sized Enterprise) meters, as well as the real temperature log.

3. Development and implementation

3.1 Main functions

Notes:

- In our terminology, Day D-1 corresponds to the current day (today). In theory, every day we were supposed to run the forecasts for Days D-1 and D with data from Day D-2 to provide the NEM with the results before 11:00 AM. In practice, data for Day D-2 is not available at 5:00 AM on Day D-1, which means this data is only processed two days later. As a result, the sequence is run using only the data for Day D-3. Forecasts for Days D-1 and D are therefore run using a horizon of two and three days.
- In addition, for each meter on the list we must provide the NEM with a half-hourly forecast for Days D-1 and D, **even if no data is received on Day D-2**.

Inputs

[1] **Tarif bleu** and *tarif jaune* are regulated tariff. Tarif jaune now longer exists since Januray 1st 2016.

Historical data logs

Historical log data is made up of R objects (RDATA files) stored locally. These logs are managed by our application and **must be updated every day before running forecasts**. If data is not received for certain timestamps, "For the consistency of historical data, particularly to avoid having holes in data, NA"s are generated for those timestamps. If data at those timestamps is received several days later, it is used to update the logs. For load curves in particular, since no D-2 data is available when the processing sequence is launched, the last 48 points are always set to "NA". If in the future this data arrived on time, it would be taken into account in the historical data logs. For temperatures, we work with real D-2 data.

Real-time meter data and weather data

In theory, the real-time data received on a daily basis is D-2 load curves and D-2 temperature data, but in reality, this data comes from D-3 because of the time it takes to collect and transmit Linky data.

- **Outputs: forecasts for all meters on the official list**
  - forecasts for Days D-1 and D in CSV format
  - forecasts for Days D-1 and D in XML-eCIM format

3.2 Architecture

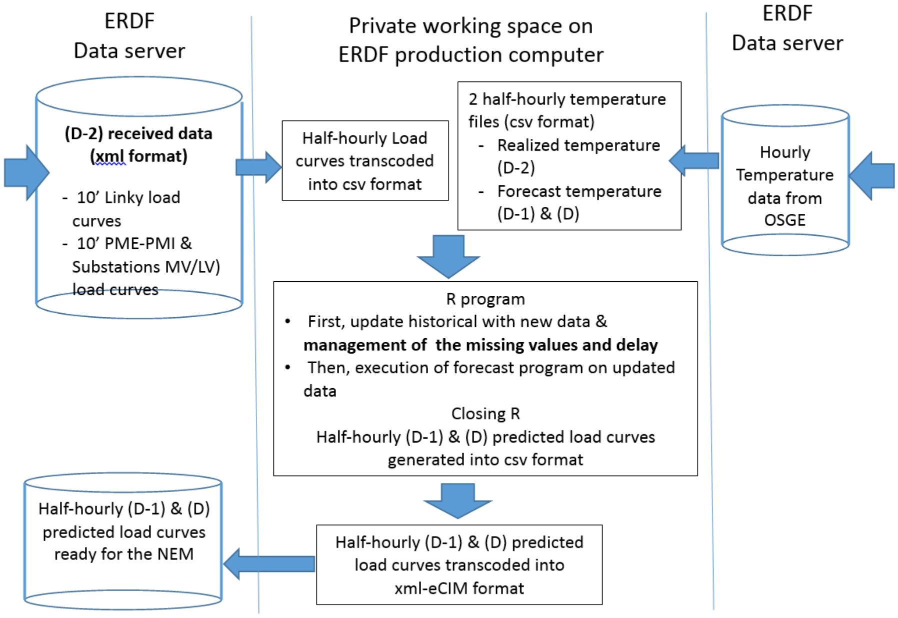

Figure 1 – Global architecture

3.3 Deployment

Programs for processing the data are all executed on ERDF production computer. Resulting forecast load curves are saved in a specific folder on ERDF data server before being used as an input for the calculations of the NEM.

3.4 Tools performance

The comparison between half-hourly forecast and realized load curves received from Linky and PME-PMI meters can be done over one day or over a period of consecutive days.

For a forecast day, we use for the accuracy indicator, the « Mean Average Percentage Error » (**MAPE**), slightly modified[2] :

$$MAPE_{Consumption} = \frac{\sum_{t=1}^{48} abs(P_{Forecast,t} - P_{Measured,t})}{\sum_{t=1}^{48} P_{Measured,t}} * 100$$

To fit with GRID4EU demonstrator, we use also the « normalized Root Mean Square Error » (**nRMSE**) for forecasts from 09:00 to 17:00 :

$$nRMSE_{Consumption} = \sqrt{\frac{\sum_{t=1}^{18} (P_{Forecast,t} - P_{Measured,t})^2}{\sum_{t=1}^{18} P_{Measured,t}^2}} * 100$$

Moreover, we should calculate the forecasts on many consecutive days (for instance, over all days of August 2015). The global indicator is then the mean of daily **nRMSE** over all these days.

4. Technical results

4.1 Individual meters

For individual meters, we used the **MAPE** as a **daily quality indicator.** Running a comparison over a week – from Monday, September 14 to Sunday, September 20, 2015 – provided insight into the quality of forecasts.

We began by comparing several **Linky meters**. The curves below represent the difference observed between forecast and real consumption for two **meters** with a good **MAPE** and two **meters** with a bad **MAPE** for the week.

- **Linky meters with relatively good forecasts (with a good MAPE)**

Linky meters 25454848017060 and 25451808945536 had mean weekly **MAPE**s of around 13.2% and 16.5%, respectively, although visually it may seem that the first **meter**'s load curve was less well forecast. This is due to a scaling effect. The differences are confined to a relatively limited range.

[2] The theoretical MAPE, for each point, divides the absolute value of difference between measured and forecast values by the measured value. As for individual load curves, some values can be very low even null, this ratio could be very high. So we decide to divide by the daily mean of realized day.

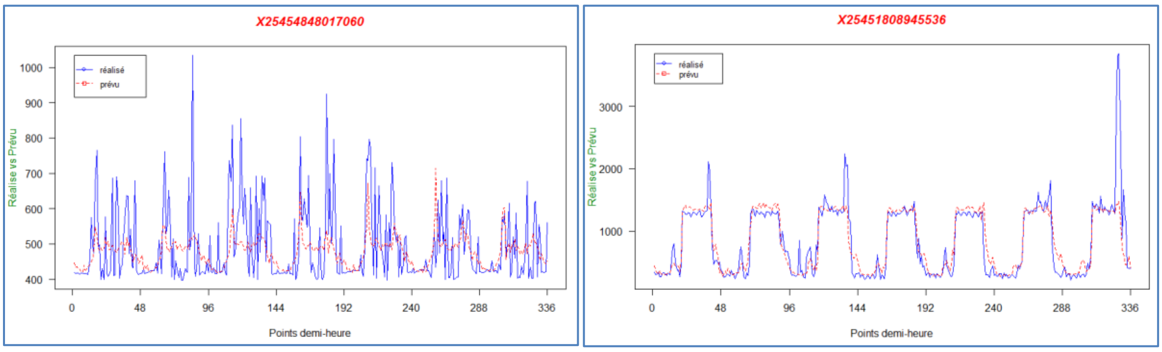

Figure 2 - Curves forecast and generated for two Linky meters: relatively accurate forecast

- **Linky meters: relatively poor forecast (with a bad MAPE)**

Linky meters 25486251781956 and 25458321157788 had mean weekly **MAPE**s of around 131.8% and 182.8%. Power usage varied widely, reached very high levels and was relatively erratic over time.

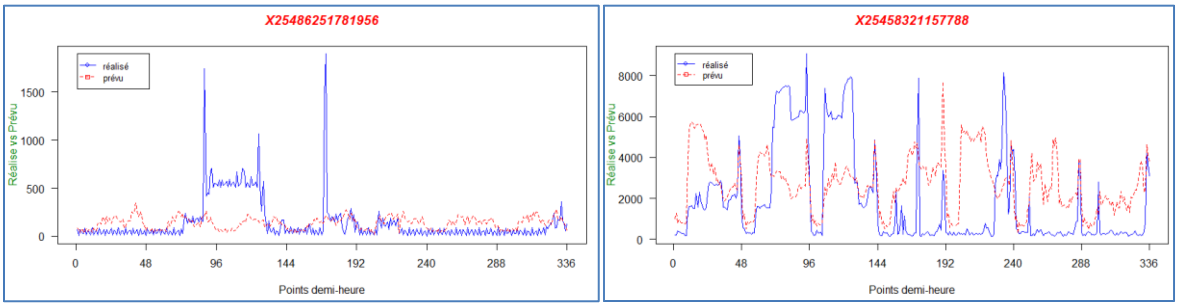

Figure 3 - Curves forecast and generated for two Linky meters: relatively poor forecast

4.2 At an aggregate level

As a **KPI** for "**Load Forecast Reliability Assessment"**[3], we used the **NRMSE** (see previous explanations above in §3.4) as a daily quality indicator and ran calculations on the period of the month of August 2015 (August 1-31 inclusive) for the 18 half-hour points (from 9:00 AM to 5:30 PM). We only took into consideration the **meters** for which we had load curves for EVERY day in August. 488 **Linky meters** (out of the 537 on the official list, which contains **meter** IDs and the substation each one is connected to) met this requirement. For these **meters**, we ran forecasts for Days D-1 and D.

We compared the total of the real load curves of these meters to the total of their daily forecasts. We also calculated the **NRMSE** for this period, which is a mean of the 31 days of daily **NRMSEs.**

The two diagrams below show the two totals – real and forecast – for this period. The number of points corresponds to the number of half-hour segments during the period, and the daily **NRMSEs** are calculated on a daily basis, i.e. per 48-point section.

[3] KPI # 10

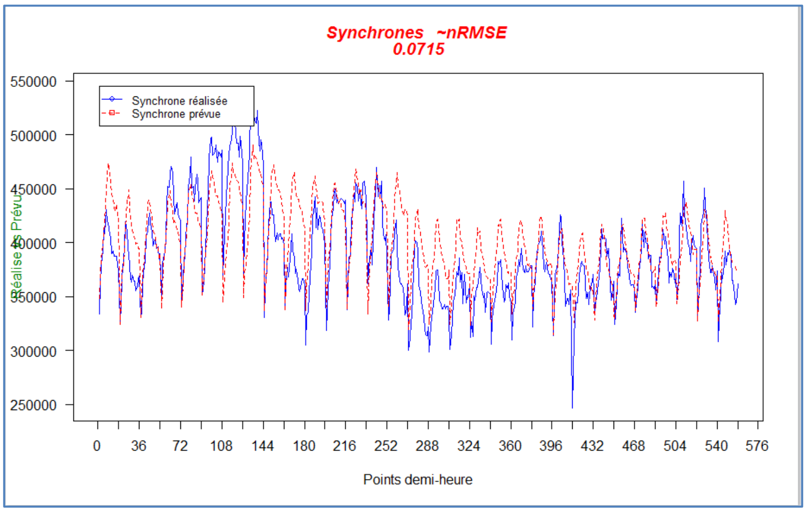

Figure 4 - Total real and forecast D-1 load curves calculated 2 days prior (x-axis = 31 days, y-axis = watts)

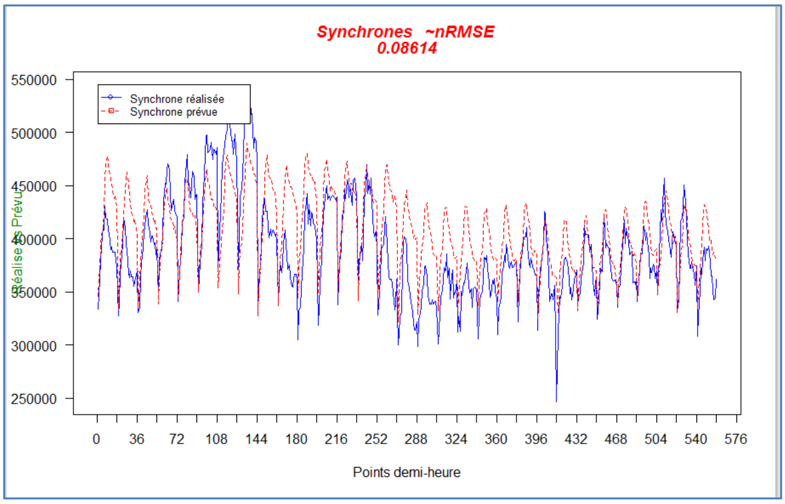

Figure 5 - Total real and forecast D load curves calculated 3 days prior (x-axis = 31 days, y-axis = watts)

The results were not bad: a mean of 7.2% for the month of August for D-1 forecasts and 8.6% for D forecasts, which is lower than the target (Consumption **NRMSE** <= 10%). Results are clearly less accurate when the forecast horizon is longer: D-1 forecasts are run for a two-day horizon and D forecasts for a three-day horizon.

4.3 Conclusion

In Figure 6, we provide the values of the mean individual **NRMSEs** in ascending order. The x-axis shows the **meters** (index of 488) and the y-axis shows the mean **NRMSEs** for the month of August 2015 in percent.

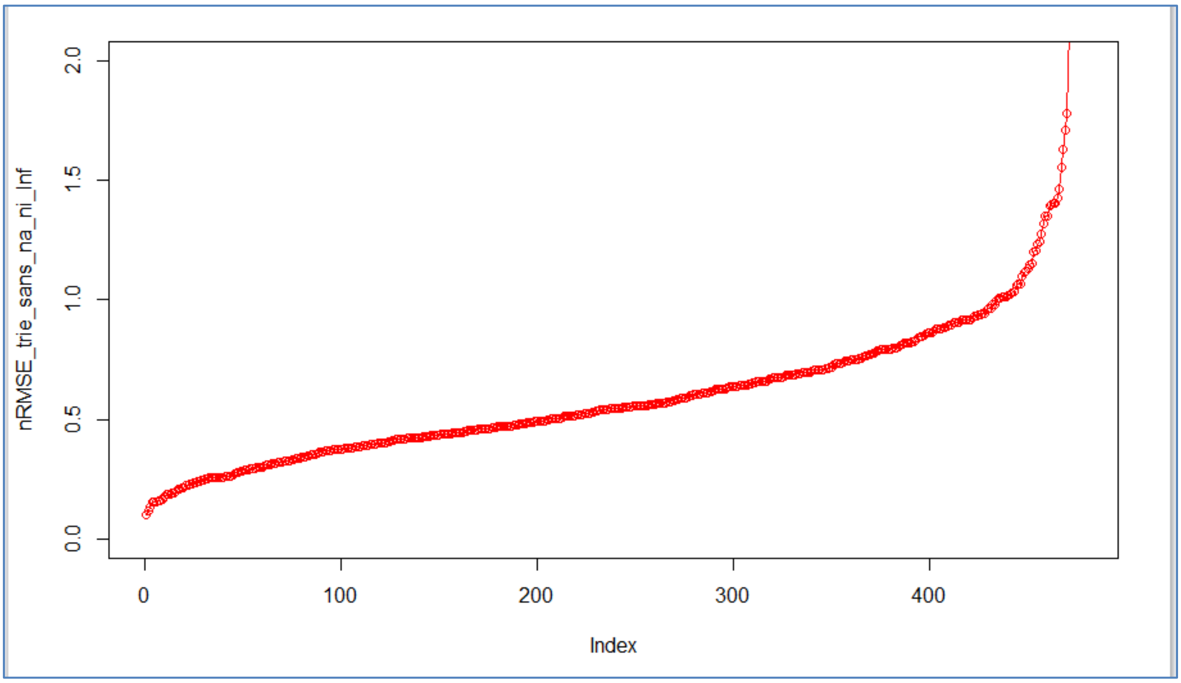

Figure 6 - Mean individual NRMSEs per meter and ordered by increasing value

The **meters** on the right-hand side of the curve were **very poorly forecasted**. Beyond around 400 on the x-axis, the indicator value is above 100%. For these cases, there is no real point in forecasting load curves because no repetitive (and therefore predictable) behavior is observed in the curves. To the left, however, the **meters** were fairly well forecast, although the results are still considerably off compared to the power consumption total for France available on the RTE website.

Note: We also observed this shape of curve for daily forecasts, regardless of the day. These results actually depend on customer consumption behavior. Customers with very repetitive consumption patterns from one day to the next have a better ***MAPE*** *or* ***NRMSE*** *than those who change their consumption every day.*

5. Conclusion and key messages

The **quality of individual forecasts** is **mediocre**, which is unsurprising given that the behavior of individuals in each household is hard to predict. At the individual level, **particular events** – absences, receiving guests, etc. – are always visible and greatly influence forecasting.

With KPI#10 (*Load Forecast Reliability*), when we only consider **meters** before rank 400, **the mean NRMSE is 50%** (here being the mean of the mean **NRMSE**s over the month of August for all **meters**); it is **64%** when we take into account all the **meters** with a valid mean **NRMSE**.

However, while the accuracy obtained for forecasts of individual curves is not very good overall (mean individual **NRMSE**s well above 10%), these forecasts could be used to predict totals **at a higher aggregation level**, such as for an **secondary substation** supposing that all the **meters** connected to it are load curve **meters** (in which case the aggregate **NRMSE** over the period **is less than 10% for one month**). This forecast could be used for LV grid forecasting management without needing to equip the substation.

We currently keep logs for two years – in case we need to make any retrospective verifications – and the forecasting algorithm only uses the last 200 days. At the aggregate level, more in-depth logs make for better forecasts. At the individual level, however, using 200 days may not be relevant: will real values from seven months ago improve the forecast for tomorrow?

It should be noted that temperature readings do not influence individual forecasts very much either.

6. Appendix

6.1 To go further

Document	Topic
dD6.7	Technical results (§2.2.2)
dD6.9.2	Key Performance Indicators (KPI) (*§2.6*)
dD6.9.3	Conclusions (*§2.7*)
Technical Spotlight S9	PV generation forecast

6.2 Glossary

Notion	Definition
CIM Format	CIM or *Common Information Model* is a data exchange model allowing application software to exchange information about an electrical network.[
Linky meter	Linky is not just an electricity meter. As well as providing accurate meter readings, it can perform remote operations, such as measuring the consumption and production of electricity, or resolving accidental outages. Linky also helps to control electricity consumption
MAPE	Mean Average Percentage Error $$MAPE_{Consumption} = \frac{\sum_{t=1}^{48} abs(P_{Forecast,t} - P_{Measured,t})}{\sum_{t=1}^{48} P_{Measured,t}} * 100$$
Network Energy Manager (NEM)	Developed by GE and operated by ERDF; analyses the network conditions of Carros area on a daily basis to identify situations where corrective actions can contribute to reduce grids constraints
nRMSE	normalized Root Mean Square Error $$nRMSE_{Consumption} = \sqrt{\frac{\sum_{t=1}^{18} (P_{Forecast,t} - P_{Measured,t})^2}{\sum_{t=1}^{18} P_{Measured,t}^2}} * 100$$
Secondary substation	Transformer substation from medium voltage (20 kV) to low voltage (400V) supplying end customers feeders. Transformer size range from 50 to 1000 kVA whereas supplied customers is very variable (up to 300 residential customers).
SME meter	SME meter is the smart meter for customer with a subscribed power greater than 36 kVA