Advanced forecasting tools for buildings and the distribution grid (algorithms)

This project has received funding from the European Union's Horizon 2020 research and innovation programme under Grant Agreement No. 645963.

Deliverable number: D3.4
Due date: 30.06.2016
Nature¹: R
Dissemination Level¹: PU
Work Package: WP3
Lead Beneficiary: 1 – ARMINES
Contributing Beneficiaries: Siemens
Reviewer(s): Arno Dentel, THN

¹ Nature: R = Report, P = Prototype, D = Demonstrator, O = Other
Dissemination level: PU = Public
PP = Restricted to other programme participants (including the Commission Services)
RE = Restricted to a group specified by the consortium (including the Commission Services)
CO = Confidential, only for members of the consortium (including the Commission Services)
Restrain UE = Classified with the classification level “Restrained UE” according to Commission Decision 2001/844 and amendments
Confidential UE = Classified with the mention of the classification level “Confidential UE” according to Commission Decision 2001/844 and amendments
Secret UE = Classified with the mention of the classification level “Secret UE” according to Commission Decision 2001/844 and amendments
<table>
<thead>
<tr>
<th>Version</th>
<th>Date</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.1</td>
<td>24.05.2016</td>
<td>First version by ARMINES</td>
</tr>
<tr>
<td>0.2</td>
<td>07.06.2016</td>
<td>Review done by THN</td>
</tr>
<tr>
<td>0.3</td>
<td>14.06.2016</td>
<td>Implementation of recommended changes by the reviewers</td>
</tr>
<tr>
<td>0.4</td>
<td>24.06.2016</td>
<td>Review done by coordinator</td>
</tr>
<tr>
<td>0.5</td>
<td>29/07/2016</td>
<td>Final version by ARMINES</td>
</tr>
<tr>
<td>0.5</td>
<td>16/08/2016</td>
<td>Approved by General Assembly</td>
</tr>
<tr>
<td>0.6</td>
<td></td>
<td>Final Version</td>
</tr>
</tbody>
</table>
# Table of Content

1. **Introduction**
   - 1.1 Purpose and Scope of the Deliverable
   - 1.2 References
   - 1.2.1 Internal documents
   - 1.2.2 External documents
   - 1.3 Acronyms

2. **Photovoltaic production forecast**
   - 2.1 Overview
   - 2.2 State of the art
   - 2.3 Forecast model
     - 2.3.1 Day ahead PV forecasting model
     - 2.3.2 Intraday PV forecasting model
   - 2.4 PV forecasts evaluation
     - 2.4.1 Datasets
     - 2.4.2 NWP data
     - 2.4.3 PV Forecasts Evaluation Criteria
     - 2.4.4 Day ahead PV forecasts evaluation
       - 2.4.4.1 Deterministic Forecasts Evaluation
       - 2.4.4.2 Probabilistic Forecasts Evaluation
     - 2.4.5 Intraday PV forecasts evaluation
       - 2.4.5.1 Deterministic Forecasts Evaluation
       - 2.4.5.2 Probabilistic Forecasts Evaluation
     - 2.4.6 Outcomes of PV forecasts evaluation

3. **Electric demand forecast**
   - 3.1 Overview
   - 3.2 State of the art
     - 3.2.1 Electric demand forecast
     - 3.2.2 Application
     - 3.2.3 Performance evaluation
   - 3.3 Missing data
   - 3.4 Forecast model
     - 3.4.1 Introduction
     - 3.4.2 Statistical framework
     - 3.4.3 Backup models
   - 3.5 Electric demand forecasts evaluation
     - 3.5.1 Datasets
       - 3.5.1.1 Electricity consumption
<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>3.5.1.2 Temperature</td>
<td>35</td>
</tr>
<tr>
<td>3.5.2 Day ahead</td>
<td>36</td>
</tr>
<tr>
<td>3.5.3 Intraday</td>
<td>38</td>
</tr>
<tr>
<td>3.5.4 Outcomes of electric demand forecasts evaluation</td>
<td>40</td>
</tr>
<tr>
<td>4 Heat demand forecast</td>
<td>41</td>
</tr>
<tr>
<td>5 Conclusions</td>
<td>43</td>
</tr>
<tr>
<td>6 Reference</td>
<td>44</td>
</tr>
</tbody>
</table>
Executive Summary

Within task 3.2.1 forecasting algorithms have been developed for electrical demand, heat demand, photovoltaic and wind production. This deliverable describes in detail these algorithms.

Concerning production, an algorithmic framework has been set up, that can do both the forecasting for the photovoltaic and the wind power in-feed.

On the demand side, the algorithm for the forecasting of heat demand has been developed and tested based on user specific data simulated in the context of the Nuremberg demonstrator. The forecasting algorithm for the electric demand has been trained and tested at data related to the Évora demonstration site.

The forecasting of electric demand and photovoltaic production are performed at the individual level, for each single user or producer.

The forecasts tool will be used in most of the Use cases as shown in D1.3 “Use Cases and Requirements” and in the three demonstrators.

For each forecast model, it is given an introduction, a state of the art and a description of the models. Finally an evaluation of its performance is presented.
1 Introduction

1.1 Purpose and Scope of the Deliverable
Within task 3.2.1 forecasting algorithms have been developed for electrical demand, heat demand, photovoltaic and wind production. This deliverable describes in detail these algorithms.
Concerning production, an algorithmic framework has been set up, that can do both the forecasting for the photovoltaic and the wind power in-feed.
On the demand side, the algorithm for the forecasting of heat demand has been developed and tested based on user specific data simulated in the context of the Nuremberg demonstrator. The forecasting algorithm for the electric demand has been trained and tested at data related to the Évora demonstration site.
The forecasting of electric demand and photovoltaic production are performed at the individual level, for each single user or producer.
The forecasts tool will be used in most of the Use cases as shown in D1.3 "Use Cases and Requirements" and in the three demonstrators.
For each forecast model, it is given an introduction, a state of the art and a description of the models. Finally an evaluation of its performance is presented.

1.2 References

1.2.1 Internal documents
D1.3 “Use Cases and Requirements"

1.2.2 External documents
See chapter 6 References, at the end of the document.

1.3 Acronyms
CRPS Continuous Rank Probability Score
ECMWF European Center for Medium-range Weather Forecast
KDE Kernel Density Estimation
MAE Mean Absolute Error
MTLF Medium-Term Load Forecast
MV Medium Voltage
NMAE Normalized Mean Average Error
NRMSE Normalized Root-Mean-Square Error
NWP Numerical Weather Prediction
**INTRODUCTION**

<table>
<thead>
<tr>
<th>Abbreviation</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>PCA</td>
<td>Principal Component Analysis</td>
</tr>
<tr>
<td>PV</td>
<td>Photovoltaic</td>
</tr>
<tr>
<td>QR</td>
<td>Quantile Reliability</td>
</tr>
<tr>
<td>RMSE</td>
<td>Root-Mean-Square Error</td>
</tr>
<tr>
<td>SMAPE</td>
<td>Symmetric Mean Absolute Percentage Error</td>
</tr>
<tr>
<td>SSR</td>
<td>Surface Solar Radiation</td>
</tr>
<tr>
<td>STLF</td>
<td>Short-Term Load Forecast</td>
</tr>
<tr>
<td>UTC</td>
<td>Coordinated Universal Time</td>
</tr>
<tr>
<td>VSTLF</td>
<td>Very Short-Term Load Forecast</td>
</tr>
</tbody>
</table>
2 Photovoltaic production forecast

2.1 Overview
Making day-ahead forecasts of PV production basically comes down to converting weather forecasts (mainly of solar irradiation) and measurements into power forecasts. This is generally done by statistical models which combine weather forecasts and historical production data to do so (see the schematic representation of the forecasting process in Figure 1). In this part is described the proposed work in developing and evaluating various PV forecasting models. In the evaluations showed here, focus is on forecasts for 1 to 2 days ahead with a 30-minutes resolution.

![Figure 1 — Generic principle of short-term PV power forecasting.](image)

2.2 State of the art
Nowadays, solar power (PV) capacity is undergoing a fast growth. The development of network management systems facilitating its penetration in the distribution network may rely on individual forecasts for each solar plant connected to the grid. Recent research work has undertaken the development of dedicated short-term (from a few hours to a few days ahead) PV forecasting models based on Numerical Weather Predictions (NWP), basically solar irradiation forecasts. Detailed modelling of electrical power as a function of solar irradiance\(^2\) has been investigated over the years. It generally relies on a parametric modelling of PV modules' performance.

---

\(^2\) One has to distinguish solar irradiation which represents the solar radiation energy by surface unit, from solar irradiance which is the solar power by surface unit (i.e. the solar radiation energy by unit of surface and unit of time). The latter is generally expressed in W.m\(^{-2}\).
efficiency with the incidental irradiance level and ambient temperature [7,8]. In a short-term forecasting context however, such a refined modelling may be unnecessary when considering the overall accuracy of NWP forecasts. Thus, a linear relationship between irradiance and power may be assumed [1]. When forecasting PV production from only past production data and solar irradiation forecasts, a statistical linear model may turn in fact, as performant as more complex nonlinear ones [9]. Nonlinear statistical models must nevertheless be useful if deciding to incorporate additional information such as temperature, air humidity or wind speed forecasts as input [2,5,6]. More generally, they may be useful when considering any additional source of nonlinearity.

NWP solar irradiation forecasts are generally provided on horizontal plane. On the other hand, PV plants generally have tilted, potentially multiple, panels' orientation(s)3 involving complex shading conditions. The relationship between irradiance levels, respectively on horizontal plane and on the panels' surface, may be complex and highly nonlinear [10-12]. Thus, in some cases, a raw linear assumption between power and forecast irradiance may not be satisfying (even) for short-term forecasting purpose. Advanced non-linear statistical algorithms may perform NWPs' recalibration to PV plant's orientation automatically. An alternative may be to re-estimate solar irradiance forecasts from horizontal to PV plant's orientation, using dedicated models, before carrying the conversion into power forecasts [3,4]. In [13], the conversion from horizontal to PV plant's orientation is first applied to clear-sky irradiance estimates. Then, tilted clear-sky irradiance estimates are combined with NWP forecasts characterizing future sky-clarity conditions as input of a Neural-Network.

2.3 Forecast model

In this work, we propose a short-term probabilistic PV forecasting model incorporating PV plant's orientation data. As in [13], it relies on clear-sky irradiance estimates derived for the PV plant's orientation. Then, we propose to model the power's distribution with a multi-components mixture, differencing clear-sky from cloudy conditions, and whose shape is parameterized by the clear-sky irradiance level. This parameterization must allow capturing the changes in the power's distribution shape both with the hour of daytime and atmospheric conditions. Forecasting the power's distribution then relies on forecasting production's regimes associated to different sky-clarity conditions, from a day to another. To that purpose, we consider different data-driven regime-switching approaches.

2.3.1 Day ahead PV forecasting model

The considered forecasting approach is as follow: one considers a statistical model to forecast the PV production from Numerical Weather Predictions (NWP) of surface solar irradiance. Parameter’s values are estimated depending on the hour of the day so as to capture interactions between the sun’s course, the PV panels' orientation, potential

---

3 We consider here plants with no sun-tracking system of any kind, thus with fixed panels' orientation.
shadowing effects, etc. It also allows capturing other effects from diurnal variations of meteorological parameters, such as the influence of temperature on modules’ efficiency. Moreover the model’s parameters are estimated adaptively using the most recent data available, so as to capture (seasonal/climatic) variations not explicitly modelled or even represented in the training data set (e.g. variations due to NWP model’s updates, to ageing or dirt on PV panels, etc.).

The chosen statistical model is a non-parametric model which has been considered in estimating the whole power distribution at once. It is based on a kernel density estimator (KDE) and can be written as:

\[
\hat{f}_{t+h}(p|I) = \frac{C_{t+h}}{b_{t+h}(k)} \sum_{i=1}^{N} w_{i,t+h}(I,k) \left\{ K\left(\frac{p - p_{t+h}}{b_{1,h}(k)}\right) + K\left(\frac{p + p_{t+h}}{b_{1,h}(k)}\right) + K\left(\frac{p + p_{t+h} - 2P_n}{b_{1,h}(k)}\right) \right\},
\]

where \( P_n \) is the power plant’s nominal capacity and the weights \( w_{i,t+h}(I,k) \) are given by:

\[
w_{i,t+h}(I,k) = \lambda^{\frac{t - t_i}{12}} K\left(\frac{I - \hat{I}_{t+h}}{b_{2,h}(k)}\right), \quad \text{and} \quad c_{i,k,h} = 1/\sum_{i=1}^{N} w_{i,t+h}(I,k).
\]

where \( \hat{I}_{t+h} \) is the forecast surface solar irradiance on horizontal plan at time \( t \) for horizon \( h \) and \( \lambda \) is a fixed factor.

Note that every day and for every forecast horizon \( h \), the parameters of the model are re-estimated using data from the last 120 days and an exponential forgetting factor \( \lambda \) to weight past observations (this factor is selected through trials and errors and fixed at \( \lambda = 1 - (0.25 \times 10^{-4}) \), see [1] for more details).

The model proposed above is a conditional kernel density estimator of the power distribution estimated conditionally to the forecast irradiance level \( I \). The bandwidth selection procedure is a \( k \)-nearest neighbours’ procedure (see [14]), with the same value \( k \) assumed for both the variable \( p \) and the covariate \( I \) and selected through trials and errors. We introduced boundary corrections from a reflection method [15] to take the bounded aspect of the PV production process into account. Unfortunately, it is impossible to know about the theoretical maximum production value at some given time without additional information about the power plant (i.e. the panels’ orientation). Boundary correction using nominal capacity \( P_n \) as high bound is not always appropriate. It involves associating non null probability masses to unrealistic high power values and then to overestimate distribution forecasts’ quantiles. Better results are obtained when truncating the power distribution’s support once unit mass is reached rather than using a renormalization, but still unreliable forecasts were observed for the highest quantiles (e.g. the 9th decile and higher). Thus, for those quantiles, we systematically replaced adaptive selection of an optimal value through cross-validation has been first considered but left apart because of the computational burden.

Because of using a varying bandwidth in (1), the estimator \( \hat{I}_{t+h} \) does not necessarily integrate to 1 and usually requires a renormalization so as to be a density [16].
forecasts made from the KDE model with those made from a linear model (LIN) which can be written as follow:

\[ \hat{p}_{t+h} = a_{t,h} \hat{I}_{t+h} + b_{t,h}, \]  

(2)

where \( \hat{p}_{t+h} \) is the production level forecast at time \( t \) for horizon \( h \) and \( \hat{I}_{t+h} \) is the forecast surface solar irradiance on horizontal plan.

When combining different models (i.e. correcting model (1) through using model (2) for the highest quantiles’ estimation), one may face the unwanted behaviour of quantiles crossing. To palliate such a phenomenon, we used the following procedure: if \( \hat{q}_{i} \), \( i = 1, \ldots, l \) denote quantiles of some distribution forecast in increasing probability order, then to ensure they are in increasing order as their associated probabilities, we do:

\[ \hat{q}_1 = \min(\hat{q}_1, P_n), \quad \hat{q}_i = \min(\hat{q}_i, \hat{q}_{i+1}) , \quad i < l. \]

2.3.2 Intraday PV forecasting model

The same KDE model has been used for evaluating the intraday forecasts. Despite the model works fine and gives required PV forecasts, it has not been designed for that purpose. We therefore planned to work on a dedicated PV forecasting model for the intraday PV forecasts to incorporate the use of the very latest known measures. This dedicated PV forecasting model will be evaluated at the end of the project and its results will be delivered through the online evaluation deliverable.

2.4 PV forecasts evaluation

2.4.1 Datasets

The offline evaluation has been done for two PV producers from a village named Valverde in the surrounding area of Evora (Portugal).
We named those two PV producers Valverde_A, with an installed capacity of 3.5 kWc and Valverde_B, with an installed capacity of 3.0 kWc. We have a period of one year from the 20th of April 2014 to the 20th of April 2015 for which measured data are available for both PV producers. But looking in detail at Figure 3, it shows that for both PV producers we have a period of null values (in February 2015 for Valverde_A and August 2014 for Valverde_B). For that reason we decided to evaluate the results for a restraint period for both PV producers. But to have still enough data to evaluate, we decided to separate the evaluations of the 2 PV producers. Indeed, the PV plant Valverde_A will be evaluated for a period of 8 months from the 1st of May 2014 until the 1st of January 2015 and the PV plant Valverde_B will be evaluated for a period of 8 months from the 15th of August 2014 until the 15th of April 2015.
2.4.2 NWP data

Accurate short-term forecasts of PV production must rely on Numerical Weather Predictions (NWP) as input. The evaluation results of PV forecasting approaches we present in this document rely on Surface Solar Radiation (SSR) forecasts from the European Center for Medium Range Weather Forecasts (ECMWF). Those forecasts have been initially derived on a grid with a spatial resolution of 0.125 (13.5 km in the meridional direction), and a 3-hours resolution. We consider weather forecasts at the 4 grid points around the solar plants and perform a linear interpolation to give weather forecasts at the solar plants. The weather forecasts are also linearly interpolated to a 30-minutes resolution. The NWP model delivers new weather forecasts every 12 hours (i.e. at 00:00 UTC and at 12:00 UTC) which cover forecast horizons from 30-minutes to 48-hours ahead.

Further work will include evaluation results from: (i) forecasting approaches used with additional weather parameters as input (e.g. air temperature forecasts), (ii) derived from other NWP models with potentially higher spatial and temporal resolution (e.g. the AROME model from the French meteorological institute MeteoFrance). Results from this work will be reported through following deliverables D4.2, D4.3 and D4.4.

2.4.3 PV Forecasts Evaluation Criteria

Let us denote a point power forecast derived at time $t$ for horizon $h$ by $\hat{p}_{t+h|t}$, while the associated observation will be denoted by $p_{t+h}$ and the related forecasting error will be denoted by $e_{t+h|t} = p_{t+h} - \hat{p}_{t+h|t}$. Since we consider the mean power level as point forecast, it is natural to consider a quadratic loss function as evaluation statistic [3], namely the Root Mean Square Error (RMSE):

$$RMSE_h = \sqrt{\frac{1}{n} \sum_{t=1}^{n} e_{t+h|t}^2}$$

Forecasts’ quality may also be evaluated through other statistics, such as the Mean Absolute Error (MAE). The latter is known for penalizing large forecast errors less than does the RMSE:
\[ MAE_h = \frac{1}{n} \sum_{t=1}^{n} |e_{t+h|t}| \]

To be comparable, those two criteria have been normalized by the installed capacity of the PV producers in the evaluation results below.

Part of the evaluation exercise consists in comparing the proposed approaches’ performance to the one of some benchmark approach. A naive approach, often used as benchmark, consists in forecasting the production’s level from the last day’s power outputs. This approach is generally referred to as persistence [1,9]:

\[ \hat{p}_{t+h|t}^{\text{pers}} = p_t + h \% 24 - 24 \]

where the horizon \( h \) is expressed in hours and \% denotes the modulus of the Euclidean division.

When considering probabilistic forecasts, namely distribution forecasts, it is generally recognized the latter must be evaluated according to two complementary properties: reliability and sharpness [17]. Let us consider distribution forecasts characterized by their quantiles. At a given moment \( t \), the \( \alpha \)-quantile (\( \alpha \in [0,1] \)) of the power’s distribution is defined by:

\[ q_t^\alpha = \inf \{ x | \mathbb{P}[p_t \leq x] \geq \alpha \} \]

Now, reliability may refer to the adequacy between the nominal proportions (i.e., \( \alpha \)) and the real proportions associated to forecast quantiles (the latter being evaluated on a test data set). In other words, reliability here relates to the “bias” of quantile forecasts. If denoting \( \hat{q}_{t+h|t}^\alpha \) a quantile forecast at \( t \) for horizon \( h \), we evaluate reliability using the criterion:

\[ Rel_h^\alpha = \alpha - \frac{1}{n} \sum_{t=1}^{n} p_{t+h} \leq \hat{q}_{t+h|t}^\alpha, \]

the closest to 0 the better. The evaluation of distribution forecasts’ reliability requires evaluating quantile forecasts’ reliability for a range of nominal proportions \( \alpha \) (here from the 10 to the 90 percentile, with 10% increment).

Probabilistic forecasts’ reliability is a prerequisite for forecasts to be informative. However, it does not guarantee to dispose of satisfying forecasts. One can imagine an approach deriving unconditional, spread out, and yet reliable distribution forecasts with not much valuable information. Thus, one often evaluates the ability of an approach to provide forecasts that change with the forecast (conditioning) information, and vary from the climatology (namely the sharpness / resolution property [17]). A (not so direct) way of quantifying such a property can be to measure the average width of centred prediction intervals, i.e.

\[ Sharp_h^\alpha = \frac{1}{n} \sum_{t=1}^{n} (\hat{q}_{t+h|t}^{1-\alpha/2} - \hat{q}_{t+h|t}^{\alpha/2}) , \]
the lower the better. Such an evaluation statistic has been often used in particular in wind power forecasting [18].

2.4.4 Day ahead PV forecasts evaluation

Day-ahead forecasts’ output cover forecast horizons from 30-minutes to 48-hours ahead. We have new PV forecasts every 12 hours (at 00:00 UTC and 12:00 UTC), when new weather forecasts are available.

As PV forecasts remain null during the night, it does not make sense to evaluate the predictions during the night. Figure below is a scatter plot showing the distribution of measured power data for the evaluated period of both considered PV plants. After analysing this figure, we decided that the NRMSE and NMAE criteria will be calculated only for horizons included between 9:00 UTC and 16:00 UTC.

2.4.4.1 Deterministic Forecasts Evaluation

Figure 5 shows the evaluation results for the NMAE criterion for the KDE PV forecasting model in comparison with a persistence model and a climatology model (designated by global_average in the figure). This figure shows that for all forecast horizons the errors of prediction, for the KDE model, are always much better than for the other model in comparison.
Figure 5 — Normalized Mean Absolute Error criteria for the next 48 hours forecast horizons for the two considered PV producers (Valverde_A on the left hand side and Valverde_B on the right hand side) for the two output PV forecast runs (at 00:00 UTC at the top and at 12:00 UTC at the bottom).

Table 1 gives more in detail the results over the NMAE criterion for some forecast horizons during the day. We can notice that the closest the forecast horizons are to the execution run, the lower is the NMAE. Indeed, results from Day 1 run at 00:00 UTC are better than results from Next Day run at 12:00 UTC which are better than Day 2 run at 00:00 UTC.

We can also notice that the calculated NMAE values are much lower for forecast horizons at the end of the day than at the beginning. The last line labelled ‘09:00 – 16:00’ represents the calculated NMAE values for all errors occurred between 09:00 UTC and 16:00 UTC. These NMAE values are all between 10.0 and 11.4. Therefore, we can say that approximately, we have a NMAE of 10.6 for the KDE PV forecasting model which is acceptable in regards to the literature.

<table>
<thead>
<tr>
<th>Time of day</th>
<th>Valverde_A</th>
<th>Valverde_B</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Day 1 run at 00:00 UTC</td>
<td>Day 2 run at 00:00 UTC</td>
</tr>
<tr>
<td>09:00</td>
<td>10.2</td>
<td>10.6</td>
</tr>
<tr>
<td>10:00</td>
<td>10.7</td>
<td>11.6</td>
</tr>
<tr>
<td>11:00</td>
<td>9.3</td>
<td>10.9</td>
</tr>
<tr>
<td>12:00</td>
<td>10.5</td>
<td>11.1</td>
</tr>
<tr>
<td>13:00</td>
<td>13.5</td>
<td>14.3</td>
</tr>
<tr>
<td>14:00</td>
<td>11.4</td>
<td>12.8</td>
</tr>
<tr>
<td>15:00</td>
<td>8.8</td>
<td>9.0</td>
</tr>
<tr>
<td>16:00</td>
<td>5.2</td>
<td>5.3</td>
</tr>
<tr>
<td>09:00 – 16:00</td>
<td>10.0</td>
<td>10.7</td>
</tr>
</tbody>
</table>

Table 1 — NMAE values for daylight forecast horizons (i.e. between 09:00 UTC and 16:00 UTC) for the two considered PV producers: Valverde_A and Valverde_B
Figure 6 shows the evaluation results for the NRMSE criterion for the KDE PV forecasting model in comparison with a persistence model and a climatology model (designated by global_average in the figure). This figure shows that for all forecast horizons the errors of prediction, for the KDE model, are always much better than for the other model in comparison.

Table 2 gives more in detail the results over the NRMSE criterion for some forecast horizons during the day.
### Table 2 — NRMSE values for daylight forecast horizons (i.e. between 9:00 UTC and 16:00 UTC) for the two considered PV producers: Valverde_A and Valverde_B

We can notice that the closest the forecast horizons are to the execution run, the lower is the NRMSE. Indeed, results from Day 1 run at 00:00 UTC are better than results from Next Day run at 12:00 UTC which are better than Day 2 run at 00:00 UTC.

The last line labelled ‘09:00 – 16:00’ represents the calculated NRMSE values for all errors occurred between 09:00 UTC and 16:00 UTC. These NRMSE values are all between 14.0 and 15.1. Therefore, we can say that approximately, we have a NRMSE of 14.5 for the KDE PV forecasting model which is acceptable in regards to the literature.

2.4.4.2 Probabilistic Forecasts Evaluation

In Figure 7 are represented reliability diagrams evaluating the probabilistic forecasts (i.e. all quantiles from 10% to 90%) of the KDE PV forecasting model. For each forecast run, focus is given to middle day hours (10-16h UTC) for which the production and forecast errors are the highest.
For all diagrams, we have a reference line and 4 curves, which are for the 10:00 UTC, 12:00 UTC, 14:00 UTC and 16:00 UTC forecast horizons. The two diagrams at the top rely on the evaluation of the 1st day of output PV forecasts from the run at 00:00 UTC. The two diagrams at the middle rely on the evaluation of the 2nd day of output PV forecasts from the run at 00:00 UTC. And the two diagrams at the bottom rely on the evaluation of the next day of output PV forecasts from the run at 12:00 UTC.

Small deviations from nominal coverage are systematically observed. This deviation is generally positive for high quantiles (being thus underestimated), while being negative for low quantiles (i.e. overestimated). Some deviation from perfect reliability may be acceptable and a threshold value of say, 5%, may be assumed for a deviation to be considered as significant.

In our evaluation, we can see from the figure above, that the probabilistic PV forecasts of the KDE model are quite reliable although we notice that the curves from the middle of the day (i.e. 12:00 UTC and 14:00 UTC) are always much closer to the reference line which means that the KDE model is more reliable at that time of the day.
Figure 8 represents the sharpness results describing the mean interval width of centred confidence intervals. Results are given for the KDE PV forecasting model, for the 48 hours ahead forecasts. The sharpness is computed for each horizon, as the average distance between two given quantiles (quantile 20% and 80% in our evaluation), normalized by the installed capacity. An ideal prediction should have a very low sharpness (near 0) while a sharpness of 1 means that the quantiles provide almost no information for the forecast.

In the diagrams above, although we can see that the results of the sharpness are higher for forecast horizons during the day, the values are always below 0.25 which means that the KDE model makes forecasts with narrow prediction intervals. We can then conclude that its forecasts are accurate.

### 2.4.5 Intraday PV forecasts evaluation

Intraday forecasts’ output cover forecast horizons from 15-minutes to 6-hours ahead with a temporal resolution of 15 minutes. New PV forecasts are produced every hour.

For the same reason as for the day-ahead forecasts, we will concentrate only on day time evaluation and more specifically on the evaluation of the 09:00 UTC output forecast run which gives output forecast horizons from 9:15 UTC to 15:00 UTC.

#### 2.4.5.1 Deterministic Forecasts Evaluation

Figure 9 shows the evaluation results for the NMAE criterion for the KDE PV forecasting model in comparison with a persistence model and a climatology model (designated by global_average in the figure). This figure shows that for all forecast horizons the errors...
of prediction, for the KDE model, are always much better than for the other model in comparison.

Figure 9 — Normalized Mean Absolute Error criteria for the next 6 hours forecast horizons for the two considered PV producers (Valverde_A on the left hand side and Valverde_B on the right hand side).

Figure 10 shows the evaluation results for the NRMSE criterion for the KDE PV forecasting model in comparison with a persistence model and a climatology model (designated by global_average in the figure). This figure shows that for all forecast horizons the errors of prediction, for the KDE model, are always much better than for the other model in comparison.
Figure 10 — Normalized Root Mean Square Error criteria for the next 6 hours forecast horizons for the two considered PV producers (Valverde_A on the left hand side and Valverde_B on the right hand side).

Table 3 gives more in detail the results over the NMAE and NRMSE criteria for some forecast horizons during the day (i.e. 10-15h UTC) from the 9:00 UTC output forecast run. We can notice that both NMAE and NRMSE values are much lower for the last forecast horizon. As for day-ahead forecasts, errors are becoming lower starting from 15:00 UTC.

<table>
<thead>
<tr>
<th>Time of day (in UTC)</th>
<th>Valverde_A</th>
<th>Valverde_B</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>NMAE</td>
<td>NRMSE</td>
</tr>
<tr>
<td>10:00</td>
<td>10.2</td>
<td>13.6</td>
</tr>
<tr>
<td>11:00</td>
<td>9.2</td>
<td>13.2</td>
</tr>
<tr>
<td>12:00</td>
<td>10.8</td>
<td>15.4</td>
</tr>
<tr>
<td>13:00</td>
<td>13.5</td>
<td>18.0</td>
</tr>
<tr>
<td>14:00</td>
<td>11.9</td>
<td>16.3</td>
</tr>
<tr>
<td>15:00</td>
<td>7.6</td>
<td>11.6</td>
</tr>
<tr>
<td>10:00 – 15:00</td>
<td>10.6</td>
<td>14.9</td>
</tr>
</tbody>
</table>

Table 3 — NMAE and NRMSE values from 09:00 UTC output forecast run for the next 6 hours forecast horizons (i.e. from 10:00 to 15:00 UTC) for the two considered PV producers: Valverde_A and Valverde_B

The last line labelled ‘10:00 – 15:00’ represents the calculated NMAE and NRMSE values for all errors occurred between 10:00 UTC and 15:00 UTC. We get a NMAE value of 10.6 and a NRMSE value of 14.9 for the PV plant Valverde_A and a NMAE value of
11.3 and a NRMSE value of 15.2 for the PV plant Valverde_B. All these results are almost the same as what we have in the day-ahead forecasts evaluation. The reason is that, as we said in a previous paragraph, the same KDE model has been used for evaluating the intraday forecasts. To get an improvement of the use of intraday PV forecasts in addition to day-ahead PV forecasts, a dedicated PV forecasting model for the intraday PV forecasts must now be developed.

2.4.5.2 Probabilistic Forecasts Evaluation

In Figure 31 are represented reliability diagrams evaluating the probabilistic forecasts (i.e. all quantiles from 10% to 90%) of the KDE PV forecasting model for some forecast horizons during the day (i.e. 10-15h UTC) from the 9:00 UTC output forecast run.

In our evaluation, we can see from the figure above, that all forecast horizons of the intraday probabilistic PV forecasts of the KDE model have a very good reliability. The results are much better than for the day-ahead forecasts as all curves are very close to the reference line.

In Figure 42 are represented sharpness results describing the mean interval width of centred confidence intervals. Results are given for the KDE PV forecasting model, for the 6 hours ahead forecasts. The sharpness is computed for each horizon, as the average distance between two given quantiles (quantile 20% and 80% in our evaluation), normalized by the installed capacity. An ideal prediction should have a very low sharpness (near 0) while a sharpness of 1 means that the quantiles provide almost no information for the forecast.
Figure 42 — Sharpness diagrams for the two considered PV plants: Valverde_A and Valverde_B

In the diagrams above, the values are always below 0.25 which means that the KDE model makes intraday PV forecasts with narrow prediction intervals. We can then conclude that its forecasts are accurate.

2.4.6 Outcomes of PV forecasts evaluation

The day-ahead PV forecasts evaluation results to a NMAE of 10.6 and a NRMSE of 14.5. The evaluation of the probabilistic forecasts shows that the quantile forecasts are quite reliable and accurate. All these results are considered to be inside the expected results in regards with what we have in the literature. However we can hope for an improvement in a whole year evaluation of day-ahead PV forecasts by the use of a meteorological model with a higher spatial resolution. The use of input measurement data without period of null values would also improve the performance of the KDE PV forecasting model.

The intraday PV forecasts evaluation results to a NMAE of 10.6 and a NRMSE of 14.9 for the PV plant Valverde_A and a NMAE of 11.3 and a NRMSE of 15.2 for the PV plant Valverde_B. There is no improvement of the use of intraday PV forecasts in addition to day-ahead PV forecasts although it is expected to. The reason is that the same KDE model has been used for evaluating the intraday PV forecasts. Therefore, a dedicated PV forecasting model for the intraday PV forecasts must now be developed to incorporate the use of the very latest known measures.
3 Electric demand forecast

3.1 Overview

To forecast an event is to predict a future happening knowing current and past situations. For a long time, to forecast has been similar to model. If a model explains exactly how a process evolves along time, we simply use the model to forecast a value. However, it is often difficult — if not impossible — to find a model describing complex mechanisms. Statistical methods have then been developed. These methods do not try to explain underlying processes but they just try to find links between causes and effects. If we have always observed that event Y comes just after event X, then, the next time event X happens, it is highly probable that event Y will also happen.

Many kinds of statistical models have been created over the years using past observations. The main problem remains to find what are the exact facts leading to an event, to identify the irreducible and complete set of causes leading to the effects. If we rely on too many causes, the particular situation will never occur again. Conversely, if we identify too few causes, situation will always have different outcomes. For the electricity consumption, measurements are numerical so we try to predict a numerical value. Once the causes are identified, at instant $t$, we record them and compute with the statistical model the value $\hat{y}_{t+h}$ which is supposed to happen later (in $h$ hours). At instant $t+h$, we observe the real value $y_{t+h}$. A model is accurate when $\hat{y}_{t+h}$ is close to $y_{t+h}$. The closer it is, the better the model is.

When the model computes a single value, it is called a deterministic prediction. However, sometimes, user might not only be interested in the most probable value. To minimize his or her risk, user might be interested in how often the real value is below the predicted one. Therefore, one has to stay cautious whenever one says that a model is efficient.

A more refined approach is to give a probabilistic prediction. Instead of computing a single value, the statistical model returns a probability density $f_{t+h}$ (or a cumulative function $F_{t+h}$, a list of quantiles, a prediction interval and so on). Note that you can always reduce this predictive function into a single value, e.g. computing the expected value of the distribution, to obtain a deterministic prediction. In most numerical cases, it is indeed impossible to find the exact value. The uncertainty is inherent to most processes: error of measurements, random outcomes, and arbitrary decisions. It mixes the uncertainty inherent to the process itself and the uncertainty due to model's efficiency.

Due to its complexity, a probabilistic model is usually more time consuming than a deterministic model. Moreover, assessing its quality requires different tools such as the CRPS or the Quantile Reliability (see Section 3.2.3).
3.2 State of the art

3.2.1 Electric demand forecast

Many models have been used to forecast electricity load in the last decades. They are often divided according to their horizon: very short-term load forecasting (VSTLF) for horizon lower than 1 hour, short-term load forecasting (STLF) up to 1 week, medium-term load forecasting (MTLF) ranges from 1 week to 1 year, and long-term load forecasting for higher horizon (LTLF). We mainly focus here on the short-term horizon. Some of them try to be understandable to the user by using a few number of inputs. These inputs are usually historical consumption data, local NWP temperature forecasts and day type (weekday and weekend or other classification).

Taylor developed over the years a derivation of the Holt-Winters exponential smoothing to capture the multiple seasonal cycles of the time series of national consumptions [19-21]. Hong et al. explained a naïve multiple linear regression for household using temperature and day type [22]. Ghofrani et al. used individual smart meter measures and trained a Kalman filter to forecast at a very short-term horizon [23]. An artificial neural network was trained for Bangladeshi households by Ahmed et al. [24]. Rana and Koprinska developed a wavelet neural network to decompose consumption time series in different frequencies for the Australian and Spanish consumptions [25]. Cho et al. predicted the French national load in two steps: general trends with generalized additive model and the between-days dependence structure with a linear regression curve [26]. Goude et al. also used a generalized additive model to do STLF and MTLF at the MV level [27]. At the MV level in Italy, Bianchi et al. reduced a high-dimensional time series of the entire day with PCA decomposition to do STLF [28]. Some studies have been made on non-residential building exclusively such as the studies of Penya et al. in Basque Country [29,30].

Although Hendricks and Koenker used hierarchical spline models for households in Chicago over twenty years ago [31], the probabilistic forecasting has not been growing until recently. Arora and Taylor used a conditional kernel density estimation using smart meters data from 800 households and 200 small enterprises in Ireland [32]. Gaillard et al. developed a quantile version of the generalized additive model [33]. The Global Energy Forecasting Competition 2014 recently proposed a competition with probabilistic prediction [34] where a lot of participants proposed probabilistic models, such as Xie and Hong who created probabilistic predictions by using different temperature scenarios [35].

Other researchers have tried to use additional information about the households to predict more precisely their consumption. Asare-Bediako et al. used a neural network with different numbers of input data for day-ahead predictions at households in the Netherlands such as humidity and holiday flags [36]. Hsiao et al. used daily schedule and context information about households in Taiwan [37]. An identification of weather and calendar factors which impact hourly consumption have been carried out by Dordonnat et al. [38]. A set of 20 inputs have been tried by Son and Kim to forecast households consumptions in South Korea with a horizon up to 1 month [39]. Fan and
Hyndman used 12 variables as input of a probabilistic STLF model for power systems in Australia [40].

Recently, methods for clustering clients in order to make more accurate predictions have been developed. Beckel et al. analyzed thousands of households in Ireland to detect characteristics from consumption [41]. Quilumba et al. predicted intraday load by considering households with the same characteristics [42]. A clustering method to understand consumption usage has been made by Dent et al. [43]. Kwac et al. used k-means to classify households in the US [44]. Haben et al. tried to cluster households with a finite mixture model and intraday characteristics [45]. A spatio-temporal method has been proposed by Tascikaraoglu and Sanandaji which tries to identify relevant surrounding houses to refine one’s consumption [46].

3.2.2 Application

Short-term load forecasting plays a key role for power systems for several reasons. The operators always need to match supply and demand. Forecasting allows them to optimally schedule their equipment and reduce their costs to supply electricity. The unit commitment function needs forecasts to determine the minimal cost strategies to schedule reservoir releases for hydraulic systems, startup and shutdown of thermal units, resources planning and so on.

A second application of the forecasts is to assess the power system security. Predictions are required to the offline network analysis in order to detect when and where the power system might be vulnerable.

In the smart meter era, local operators (working at small scale, i.e. the building-scale) are using precise predictions in order to optimally manage their network: whenever the demand is supposed to be important, buy the necessary electricity from the market or use electricity stored in batteries.

3.2.3 Performance evaluation

There are various ways to assess models performance.

Among the most usual is the Root-Mean-Square Error (RMSE). The RMSE of a model can be estimated over a period $T$ of length $n$ is

$$\text{RMSE} = \sqrt{\frac{1}{n} \sum_{t \in T} (y_t - \hat{y}_t)^2}$$

where $y_t$ is the measure and $\hat{y}_t$ is the forecast. This score is negatively oriented and is always positive. This error is directly comparable to the time series, meaning that if the RMSE is equal to 100Wh, there is approximately a 100Wh difference between the measured and the predicted consumption. However, the score is often normalized to obtain a dimensionless value such as a 10% error. It is convenient in order to compare different applications of the model. If we try to evaluate the quality of the model on two households, household A which consumes 1000Wh in average, and household B which consumes 200Wh in average, it is highly probable that the non-normalized error will be
significantly higher for A than for B. Various normalizations have been proposed in the literature. Our definition of the normalized RMSE (NRMSE) uses the maximum measured value of the consumption, i.e.

\[
\text{NRMSE} = \frac{1}{\max_{t \in \mathcal{T}} y_t} \sqrt{\frac{1}{n} \sum_{t \in \mathcal{T}} (y_t - \hat{y}_t)^2}.
\]

The Mean Absolute Error (MAE) is also widely spread

\[
\text{MAE} = \frac{1}{n} \sum_{t \in \mathcal{T}} |y_t - \hat{y}_t|
\]

and uses the absolute norm instead of the quadratic one. It is negatively oriented and positive. It is less penalized if the error at one instant is very large. It is therefore more robust to detect outliers. Once again, various normalizations exist. For instance, there is the Symmetric Mean Absolute Percentage Error (SMAPE) which divides each value by the mean between true and predicted consumption (to avoid division by zero),

\[
\text{SMAPE} = \frac{1}{n} \sum_{t \in \mathcal{T}} \frac{|y_t - \hat{y}_t|}{y_t + \hat{y}_t}.
\]

Whenever we have a probabilistic prediction, a predictive distribution or a set of quantile values, we need some specific errors to assess the quality. The Continuous Rank Probability Score (CRPS) uses the whole forecast distribution function

\[
\text{CRPS} = \frac{1}{n} \sum_{t \in \mathcal{T}} \int \left( \tilde{F}_t(y) - \mathbb{I}(y > y_t) \right)^2 dy
\]

where \( \tilde{F}_t \) is the forecast cumulated distribution function at instant \( t \), and \( \mathbb{I}(y > y_t) \) is equal to 1 if \( y > y_t \) and 0 otherwise. This score is negatively oriented and positive.

Since this score is often tedious to compute — you have to calculate a lot of points in order to obtain precise empirical distribution —, we sometimes use the simpler Quantile Reliability (QR). A predicted \( \tau \in [0,1] \) quantile is said to be reliable when the ratio of observed values is lower than the quantile value. If \( \hat{y}_t^{\tau} \) is the predicted \( \tau \)-quantile, then

\[
\text{QR}(\tau) = \frac{1}{n} \sum_{t \in \mathcal{T}} \mathbb{I}(y_t \leq \hat{y}_t^{\tau}) - \tau.
\]

The closer to 0, the better the reliability is. If the reliability is negative, it means that the predicted quantile is too low; if the reliability is positive, it means that the predicted quantile is too large. A common issue is a positive reliability for low quantile and a negative reliability for high quantile, meaning that the extreme values, i.e. tails of the distribution, is underestimated: we underestimate rare events.

### 3.3 Missing data

For all predictive models, there might be numerous problems with historical data supplied. For the demand (electrical or heating), to correctly train models, it at least requires half a year of data, including winter and summer times, to have a large range of temperatures and situations.
There are often a few missing measurements inside the training datasets, sometimes the hourly data has not been recorded, wrongly transmitted or not even measured. If it happens every now and then, it is not a big issue. The missing points can be filled out with a simple spline interpolation of the nearest values. However, if the missing data last more than 5 consecutive hours, it is nonsensical to fill the gap and the corresponding period is discarded from the training set.

For some households, too much data is discarded from the training set. In these cases, similarly to households without historical data, it is impossible to train a model specific to this household.

In any case, models require having a live stream of data corresponding to the latest observations — supplied every day or so. If the live measurements are missing, it is not as easy to fill as for the training set. Most of the time, missing data, lasting less than 5 hours, can be filled out with splines. Occasionally, it should not be approximated because the missing points are “at the end” of the live stream, and if they are at the end, it induces a spline extrapolation which can be grossly wrong. Therefore, if a measurement is not received, they are approximated with spline if the period does not last more than 5 hours, and if we know data before and after the missing periods.

Temperature prediction comes from reliable institutes that provide values without doubt. Since, there still can be some problems with data transmission, backup models are available to replace the complete model whenever the temperature is missing.

### 3.4 Forecast model

#### 3.4.1 Introduction

The forecast model is designed to give probabilistic predictions of the electrical consumption of one household. The horizon is short: prediction for the next few hours, or for the next day. The word “probabilistic” refers to a range of predictions rather than a traditional single point prediction.
The positive time series \( (l_t) \), with \( t \) the number of hours since an arbitrary reference instant, denotes the hourly electrical load used by one household. Figure 53 shows the first week of March 2015 for one house in Évora. It is clear that the electrical load of a household exhibits distinct patterns: yearly pattern, weekly pattern, and daily pattern. These patterns originate from different reasons: meteorological seasons and weather condition, public and school holidays, work and sleep schedule.

The yearly pattern is of low interest for our model, since it is designed for short-term predictions. However, the local temperature is used inside the model to describe variations throughout the year. Consumption is higher in the winter than in the summer because it is colder outside. Houses thus often use electrical heating devices, increasing their overall consumptions. We note \( (T_t) \) this temperature series. Since the temperature is not known in advance, this series will, in the following, designate temperature predictions made by the European Centre for Medium-Range Weather Forecasts (ECMWF).

We then focus on the weekly and daily effect. These effects are depicted in Figure 64 where an average of the electrical consumption for each day of the week is plotted. Significant differences among days can be observed. For instance, we can notice that inhabitants of the house seem to use electricity later in the morning during the weekend. A finer approach has been investigated by classifying days in categories. For instance, a public holiday Monday is more similar to a Sunday than another Monday. Such a method has not been applied in our model, because it does not significantly improve predictive quality.
To reduce the weekly impact on the consumption, we use a training set of available data to compute the median consumption of every day of the week. Median, rather than mean, is selected for its robustness. For every day of the week, day ∈ {Monday, ..., Sunday}, and for every hour h ∈ {0, ..., 23}, we note $m_{h}^{\text{day}}$ the median consumption of the household. These median consumptions are then subtracted from the electrical loads. By defining the day indicator function $\mathbb{I}_{\text{day}}(t)$ — equal to 1 if instant $t$ occurs during day, 0 otherwise —, and taking $h = t \mod 24$, we define a new series $(y_t)$ as

$$y_t = l_t - \sum_{\text{day}=\text{Monday}}^{\text{Sunday}} \mathbb{I}_{\text{day}}(t) m_{h}^{\text{day}}$$

for every $t$. Let us note that this newly defined time series has positive and negative values. These median shifts, which are different for every day of the week, capture a part of the daily and weekly pattern of the series.

The resulting $(y_t)$ series can be approximated by a sum of three functions depending on the hour of the day. For all $t$, and its associated hour of the day $h$,

$$y_t = a_h(y_{t-24}) + b_h(\bar{y}_t) + c_h(T_t) + \epsilon_t,$$

where $\bar{y}_t$ is the median of the previous week, i.e.

$$\bar{y}_t = \text{median}(y_{t-24}, y_{t-48}, y_{t-72}, y_{t-96}, y_{t-120}, y_{t-144}, y_{t-168}),$$

and $\epsilon_t$ is the noisy part. This model will later be referred to as model (A). It is therefore assumed that the consumption at instant $t$ is the sum of three decorrelated effects and a noise: consumption from the day before with function $a$, consumption from the previous week with function $b$ and the temperature — or rather the predicted temperature — at the current instant with function $c$. This framework assumes that there is no correlation...
between the three functions. A three dimensional function, i.e. a function \( f(y_{t-24}, y_t, T_t) \), is too complex to be estimated precisely.

### 3.4.2 Statistical framework

By considering that the noise is a series of realizations of centred random variables, i.e. \( \mathbb{E}[\varepsilon] = 0 \), we can estimate the three functions with known data, for each hour. We use the function \( \text{rqss} \) (for regression quantile smoothing splines) developed by Koenker [47] in the R-package \textit{quantreg}. It is similar to the well-known \textit{gam} (generalized additive model) but with a pinball loss function. We therefore obtain, for every \( \tau \in [0,1] \), using exclusively previous consumption values and the predicted temperature, as in the Model A below.

\[
\hat{y}_t^\tau = a_h^\tau (y_{t-24}) + b_h^\tau (\hat{y}_t) + c_h^\tau (T_t)
\]

To obtain the real electricity consumption (which is a positive value), you still have to shift back with median consumption. For the hour of the day \( h \) corresponding to instant \( t \),

\[
\hat{y}_t^\tau = \hat{y}_t^\tau + \sum_{\text{day} = \text{Monday}}^{\text{Sunday}} \mathbb{I}_{\text{day}}(t) m_h^\text{day}.
\]

Sometimes, predictions for the next day are computed at noon. In these cases, temperature predictions are computed at noon for the next day so, to predict next day afternoon, you have to use afternoon of the day before. Instead of taking the consumption of the previous day, we thus take the consumption of two days before prediction as explained in Figure 75.

![Figure 75 — Illustration of the inputs used to predict consumption on day \( n \). The prediction is computed at noon on day \( n - 1 \) (black circle). At this instant, we require past consumptions during the whole previous week, and the temperature predictions for next day. Since the afternoon consumption of day \( n - 1 \) is not yet available, we instead use the afternoon of day \( n - 2 \).](image)

In practice, we estimate the 72 functions (24 times each one of the three functions) for every decile (\( \tau = 0.1, \ldots, 0.9 \)) to have a precise empirical predictive distribution.

### 3.4.3 Backup models

For diverse reasons the previous \( \text{rqss} \) model can be inapplicable. The \( \text{rqss} \) model in the R-package \textit{quantreg} is inefficient for extrapolation. So, if the training set is not widely
Electric demand forecast

Spread, it is better not to use the rqss model. For instance, if the training set covers just the summer time, it cannot be used for winter, because winter temperatures have never been observed before. This is why the training set must last through winter and summer. Even with an extremely wide spread set, it can always happen that the newly observed inputs are extrapolations from the set. In this case, we use a simpler model, denoted as model (B), where the quantile $\hat{y}_{t}^{\tau}$ is calculated using the observations of the previous week,

$$\hat{y}_{t}^{\tau} = \text{quantile}_{\tau}(y_{t-24}, y_{t-48}, y_{t-72}, y_{t-96}, y_{t-120}, y_{t-144}, y_{t-168}).$$

Sometimes, due to the low quality of data, long periods of measures are missing. When these periods are short (just a couple of hours), it is possible to fill measurements with close available values. However, when these periods extends, model (A) can also be inapplicable. We therefore use model (B). When the quality of the previous week of measurements is too poor, we use the quantile observations of the training set and not these of the previous week: it is model (C).

In certain households, there are no data available at all. In this case, given that we do not have any additional information (number of people in the household, surface area, work schedule), we simply simulate an average household based on all households in the vicinity and create a quantile model using local temperature and hour of the day such as in the Model D below

$$\hat{y}_{t}^{\tau} = f_{h}^{\tau}(T_{t}).$$

Finally, if the temperature prediction is not available, or if there is an unexpected problem, we use quantile observations of the average household independently of the temperature: model (E).
3.4 Electric demand forecasts evaluation

3.5.1 Datasets

3.5.1.1 Electricity consumption

The dataset used to evaluate our model’s quality gather electrical consumption of 226 households in the village of Valverde, in the neighbourhood of Évora, Portugal. They will be referred to with their number (household 1, household 2…, household 226). During the year 2015, measurements of individual loads were recorded every hour, resulting into 8760 measures from the 1st of January 00:00, to the 31st of December 23:00.

<table>
<thead>
<tr>
<th>Available measures</th>
<th>less than 1 month</th>
<th>between 1 and 11 months</th>
<th>more than 11 months</th>
</tr>
</thead>
<tbody>
<tr>
<td>Number of households</td>
<td>28</td>
<td>34</td>
<td>164</td>
</tr>
</tbody>
</table>

Table 4 — Number of households with few or no data (less than 1 month), with partial data (between 1 and 11 months), and almost complete data (more than 11 months).

Table 4 details the overall quality of the data set. Most households, 164 out of 226, have very good measurements, with only a few missing values here and there. 28 households
have very little data, less than 1 month. It is therefore difficult to evaluate whether the forecast model perform well or not on them. The rest of the households have often extended periods where for any reason data are not measured. Prediction of these 34 households are challenging because they have regime ranges between different measured periods. For instance, household 194 has an average consumption which goes from 500Wh in April to 600Wh in June. However, since there are no data between these two months, it is difficult to assess whether the gap is normal (e.g. habits changing in the summer) or not. It is possible that this gap comes from a sudden change (e.g. new highly consuming electrical device or new residents) which requires a new training of the model.

Having this dataset for one year, we need to divide it into two in order to have a training set and a testing set. The training set is used to feed the statistical models (in our case, it is a necessary step for model (A), (C), (D) and (E); unnecessary for model (B) which only uses previous week data, see section 3.4.3) while the testing set is required to evaluate models’ performance. We have chosen to divide in a usual 80/20 way. Therefore, the training set goes from the 1\textsuperscript{st} of January 00:00 to the 11\textsuperscript{th} of October 23:00 (6816 hours) and the testing set from the 12\textsuperscript{th} of October 00:00 to the 31\textsuperscript{st} of December 23:00 (1944 hours). Although the training period extends from winter to summer, and covers a wide range of temperature, the testing period suffers from the fact that it only lasts during a cold period. It would have been better to select 80% of random days during the year for training and the remaining 20% for testing but we opt to keep continuous period to mimic the real behaviour of future applications.

3.5.1.2 Temperature

Two out of our five models (models (A) and (D), see Figure 86) require local temperature predictions. European Centre for Medium-Range Weather Forecasts (ECMWF) which uses national meteorological services provides predictions computed every 12 hours (at midnight and at noon) for the next few days by step of 3 hours.

To use our Évora consumption datasets, temperature predictions in the city were requested. For the day ahead forecasts, we use prediction for day \( n \) computed at noon on day \( n - 1 \), so forecast horizons are 12, 15, 18, ..., 36 hours. For the intraday forecasts, we requested the closest temperature predictions realized. Predictions made at midnight give temperatures for midnight-to-noon period, while predictions made at noon give temperatures for noon-to-midnight period.

Since temperature predictions are given with a time step of 3 hours, whereas measured electricity consumptions are given with a time step of one hour, a linear interpolation is applied on the temperature predictions to obtain values with the same time steps.
3.5.2 Day ahead

We compute hourly functions of model (A) (described in section 3.4.3) for every one of the 226 households. Functions are estimated on the training set. Should there be no missing data for one household, 284 points are available for each hourly function. For 8 households, with very poor data quality, functions are impossible to evaluate at all. Therefore, the average models ((D) and (E) ones), which do not take these into account, are used.

Figure 97 represents estimated functions for one particular household. Top row shows functions at 03:00 and bottom row shows functions at 14:00. Solid lines show the median function found for $\tau = 0.5$ and the dashed lines show extreme functions for $\tau = 0.1$ and $\tau = 0.9$. These graphs show that the influence of previous consumptions is almost linear, and that consumption seems to decrease when the temperature increases. These trends may not be as strong for other households especially the one concerning temperature since households in Portugal are hardly sensible to temperature changes.

It should also be noted that, for some households at some instants, it is possible to have two quantiles $\tau_1 < \tau_2$ but their corresponding predicted quantiles $\hat{y}_t^{\tau_1} > \hat{y}_t^{\tau_2}$. It may be caused by insufficient training points, or convergence problem of the rqss models. Currently, if two quantiles predictions give an inverted consumption predictions, these two are simply interchanged.
Figure 108 — First three days of November for household 6. In orange, the true consumption is plotted. Ticks represent instants when predictions for next day are computed. In solid line is plotted the 0.5 quantile prediction, the lightly shadowed area is the prediction interval between [0.1, 0.9], the darker area the prediction interval [0.3, 0.7].

On Figure 108, we represent three days of prediction for household 6. Every day at noon (on the 31st of October, the 1st and the 2nd of November), we compute quantile predictions using model (A) (for respectively the 1st, the 2nd and the 3rd of November). The true consumption is plotted in orange, the 50% quantile is plotted in black, light shadow interval is the 10%–90% prediction interval and the dark shadow interval is the 30%–70% interval. The predicted and the real consumptions roughly follow the same trends; however it is interesting to note the discrepancy on the evening of the 1st of November. The real consumption peaks after the predicted peak. Explaining the difference is not obvious: it might be that on Sunday, the household came home later for dinner. This kind of discrepancy might be reduced with a presence detector: should the family not be home on Saturday, it is then probable that the family is not here on Sunday either.

It is interesting to check prediction performance over all households. Since we predict for the whole day at noon of the previous day, the prediction horizons are not always the same. Indeed, to predict 01:00 consumption, it is 13 hours ahead; to predict 15:00 consumption, it is 27 hours ahead. In Figure 119, we plot errors for every horizon: from 12 hours ahead to 36 hours ahead. The top graph shows the NRMSE and the SMAPE scores. The slow deterioration of the scores is visible: the further in time it is, the harder it is to predict. Even though we use normalized scores, the consumption peak around 18:00 is still visible on the errors: it is at that time that the model is the least efficient according to these scores. In black is also plotted the persistent model. We just took the previous day consumption for the next day prediction. Our model outperforms it consistently (except for the curious 18:00 value). The bottom graphic shows the difference between the true quantiles and the quantiles predicted. So, if there are 15% of the real values that are beneath the predicted quantiles, there is a difference of +5%.
The graphic shows that the model underestimates small and large values. This behavior is common because extreme values are less likely to happen and require a large training set. Degradation of the quantiles over horizon is not as visible as for NRMSE and SMAPE.

**Figure 119** — Model performance over different horizons (from 12 hours to 36 hours ahead). The top graphic depicts NRMSE (orange solid line) and SMAPE (blue dashed line). The black lines show performance of the persistence model. The bottom graphic shows the quantiles reliability: differences between the true quantile and the predicted quantile.

### 3.5.3 Intraday

The intraday predictions are computed at midnight and at noon (when temperature predictions are supplied) for the following 12 hours. Currently, it is the same model as for the day ahead with only the temperatures which are adjusted. Therefore, since weather predictions are only updated every 12 hours, electricity demand predictions are the same during the gap between two weather predictions. We are working on a new model taking into account the latest consumption measured (to make use of the 9 o’clock consumption in order to predict the 11 o’clock consumption more precisely).

Nevertheless, in Figure 20, we plot the prediction over the same period as before, first three days of November. The prediction interval spans are similar to those of day ahead.
and there does not seem to be much difference between the intraday predictions and the day-ahead ones.

Figure 121 shows the scores over all the households. Scores are smoother over horizon and they globally are lower than those of day-ahead predictions.

![Household 6 graph](image)

*Figure 20 — First three days of November for household 6. In orange, the true consumption is plotted. Ticks represent instants when predictions are computed. In solid line is plotted the 0.5 quantile prediction, the lightly shadowed area is the prediction interval between [0.1, 0.9], the darker area the prediction interval [0.3, 0.7].*
3.5.4 Outcomes of electric demand forecasts evaluation

The model introduced for the day-ahead electric load building forecasting shows similar performance to those developed in the literature: the quadratic error (NRMSE) for one building is 22% in average. The probabilistic evaluation shows that the quantile prediction has a reliability error of approximately 5%. This reliability is poorer for extreme quantiles where the error can reach 10%. Such phenomenon is a well-known issue when dealing with limited training dataset (less than 1 year). The intraday model works with the same framework (inputs and model structure) as the day-ahead model but with updated prediction for the temperature. Performance is thus only slightly enhanced and the quadratic error is 21.4% in average.
4 Heat demand forecast

The heat demand time series exhibits characteristics similar to those of the electric demand series. Heat demand is non-smooth, irregular and highly volatile. It is also strongly correlated to activities in the household (occupancy, time of the day, day of the week) and to the local weather (temperature, humidity).

We propose to use a similar forecast framework as the one developed for the electric demand. Heat demand time series faces the same issues in data quality as the electric demand time series (incomplete or absurd measures), hence we use different methods for the different cases. The best model uses a quantile generalized model taking into account heat data measured during the previous week and the outside temperature. It is important to note that, currently, the temperature series is measured (and not forecasted). When heat data or temperature data are unavailable, like for electric demand, simpler models are used: such as average measures during a training set or average forecast of the surrounding households.

We have tested the model for just one household during one year. We divided the year in two: the training set up to 11th October and the testing set for the end of the year. The resulting NRMSE (normalized by the nominal power) is equal to 15.5%. The SMAPE is equal to 34.9 %. On Figure 13 is showed three days of the prediction. At noon on day n is computed prediction for day n+1 (day-ahead prediction). As of now, the model was solely tested with measured temperature; therefore the intraday prediction gives the same results as the day-ahead prediction. When the model is tested with predicted temperatures, it is expected to see a small global degradation due to temperature prediction uncertainty. With predicted temperatures, the intra-day performance is expected to outperform day-ahead performance.

![Figure 13 — Heat demand measured (in orange). In black, the mean prediction. The light shadow area is the 30%-70% prediction interval and the dark shadow area is the 10%-90% prediction interval.](image)

Figure 13 — Heat demand measured (in orange). In black, the mean prediction. The light shadow area is the 30%-70% prediction interval and the dark shadow area is the 10%-90% prediction interval.
It looks like the heat demand time series is null during the summer. One can consider finding a yearly period during which heat demand is null. But there is currently not enough measured data to validate this plan.
5 Conclusions

The evaluation of the day-ahead PV forecasts shows similar performance to those developed in the literature. It results to an average NMAE value of 10.6 % and an average NRMSE value of 14.5 %. The evaluation of the probabilistic forecasts is also encouraging as the quantile forecasts are quite reliable and accurate. The evaluation of the intra-day PV forecasts does not show any improvement of the performance model in terms of forecast errors. Indeed, it results to an average NMAE value of 11.3 % and an average NRMSE value of 15.2 %. A dedicated PV forecasting model for the intra-day PV forecasts should be used to expect an improvement of the performances. The evaluation of the probabilistic forecasts shows really good results with quantile forecasts being reliable and accurate.

The evaluation of the day-ahead electric load building forecasting also shows similar performance to those developed in the literature. The average value of the quadratic error (NRMSE) for one building is 22 %. The evaluation of the probabilistic forecasts shows that the quantile prediction has a reliability error of approximately 5 % which is also a good value. The evaluation of the intra-day electric demand forecasts does only show a very little improvement with an average value of NRMSE of 21.4 %. As for the PV forecasts, a dedicated electric demand forecasting model for the intra-day forecasts should be used to expect an improvement of the performances.

The Table 5 gives a summary of the NRMSE values for day-ahead and intra-day forecasts for both PV production and electric demand forecasts. It is to be noted that if NRMSE criteria is computed for the whole 24-hours day horizons for the electric demand forecasts, it is computed only for a restricted range of horizons for PV production forecasts (i.e. between 9:00 UTC and 16:00 UTC for day-ahead forecasts and between 10:00 UTC and 15:00 UTC for the intra-day forecasts). PV production is null during the night which does not make sense to compute its errors of prediction. The range of considered horizons to compute the NRMSE is shorter for intra-day forecasts because the model only outputs forecasts for the next 6 hours. We have seen that NRMSE is higher at midday than at the beginning and at the end of the day. This therefore explains why NRMSE is higher for intra-day than for day-ahead PV production forecasts.

<table>
<thead>
<tr>
<th></th>
<th>NRMSE of Day-ahead forecasts (%)</th>
<th>NRMSE of Intra-day forecasts (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>PV production</td>
<td>14.5 %</td>
<td>15.2 %</td>
</tr>
<tr>
<td>Electric demand</td>
<td>22 %</td>
<td>21.4 %</td>
</tr>
<tr>
<td>Heat demand</td>
<td>15.5 %</td>
<td>15.5 %</td>
</tr>
</tbody>
</table>

Table 5 — Day-ahead and Intra-day forecasts’ NRMSE for both PV production and Electric Demand
6 Reference


27. Goude Y, Nedellec R, Kong N: \textit{Local Short and Middle Term Electricity Load Forecasting With Semi-Parametric Additive Models}. \textit{Ieee Transactions on Smart Grid} 2014, \textbf{5}:440-446.


