Schweizerische Eidgenossenschaft
Confédération suisse
Confederazione Svizzera
Confederaziun svizra

Federal Department of the Environment, Transport, Energy and Communications DETEC

Swiss Federal Office of Energy SFOE
Energy Research and Cleantech Division

REEL Demo – Romande Energie ELectric network in local balance Demonstrator

Deliverable: 3d2 Improvement of numerical weather prediction at local scale using aggregated low-quality sensor data

Demo site: Chapelle

Developed by

Lorenzo Nespoli, ISAAC, SUPSI

Vasco Medici, ISAAC, SUPSI

Roman Rudel, ISAAC, SUPSI

[Canobbio, 17.06.2019]

1 Description of deliverable and goal

Distributed dispatchable loads can be operated to optimally counteract abrupt changes in the power produced by renewable energy sources installed in the distribution grid. Thermo-electrical loads such as heat pumps (HP), HVAC and electric boilers are among the most energivorous dispatchable loads and thus the most likely to be operated in this sense. This deliverable focuses on the development of a technique to increase the accuracy of temperature forecasts up to 36 hours into the future, exploiting the information coming from low-quality distributed thermal sensors. Typically, spatial resolution of numerical weather prediction (NWP) for temperature predictions is of few kilometers; in this range temperature variation can be significant, especially in mountain regions. Such spatially coarse prediction can be corrected using locally installed sensors, reconciling the observed values and the NWP forecasts. The final scope is double fold: in the first place, this would result in better scheduling of these devices whose performance are temperature dependent, like HPs and HVACs, for which the coefficient of performance (COP) is influenced by the external air temperature (Ta). A smart controller of these devices would thus benefit in knowing Ta in advance, e.g. waiting for periods in which Ta is higher (and so the COP) to start heating an household, in order to use less energy. On the second hand, a better forecast of Ta could increase the accuracy of power forecasts for single households, as well as for aggregated groups of consumers.

1.1 Executive summary

Temperature time series from distributed sensors have been used in order to increase the accuracy of meteorological forecasts. The signals come from thermal sensors of Netatmo stations [9]. Despite the reasonable accuracy of these devices, the resulting available data presents two main issues:

- The sensors are installed by the users. This could result in an improper installation (e.g. on the southern facade of a building, under direct sun irradiance).
- The data is incomplete (see Fig. 1). Temporal characteristics of unavailable data spans from short and regular periods, which are likely to be the result of connection issues, to very long periods in which the signals are completely absent.

The first issue has been mitigated by identifying and removing outliers, and those signals having a high correlation between the measured temperature and the irradiance for different positions of the sun. The second issue has been addressed by firstly discarding those signals presenting large quantity of missing data. The missing data of the remaining sensors has then been forecasted using proper data imputation techniques. The complete forecasted signals has then been used to correct the forecasts coming from a NWP service. The technique has been tested with annual data from 59 netatmo stations located nearby Bellinzona (TI), in the district supplied by the Azienda Multiservizi Bellinzona (AMB). The results show that the method consistently increase the accuracy of the forecasted signals for all the locations.

1.2 Research question

The research questions addressed in this deliverable are two:

1. Is it possible to increase the forecast accuracy for the ambient temperature, with respect to the one provided by NWP services, using low-quality sensor networks? In section 1.4.2 we investigate this possibility.
2. Is it possible to use the corrected temperature forecast in order to increase the accuracy of the power forecasts of the aggregated power consumption of given region? In section 1.4.3, we try do so using data from AMB clients. We try to achieve this by obtaining a more representative forecast for the temperature influencing the supplied region.

1.3 Novelty of the proposed solutions compared to the state-of-art

1.3.1 Increasing temperature forecast accuracy

As previously stated, the agreement between the NWP prediction, $\hat{T}_{NWP,t}$ and the temperature observations up to time t, $T_{obs,t}$,can be increased using a data reconciliation technique. In particular, we want to obtain an increase of accuracy of the forecasted temperature $\hat{T}$, given $\hat{T}_{NWP,t}$ and the observed NWP error; in other words, we want to obtain the conditional probability distribution $p\left[\hat{T}_{t+k}|\hat{T}_{NWP}, T_{obs}\right]$. This can be done using a Kalman filter, as suggested in [1, 2, 3]. However, this requires to define a model for the evolution of the error $\epsilon_t = T_{obs,t} - \hat{T}_{NWP,t}$. Since no assumptions can be made in general over the time evolution of ϵ, a simple autoregressive model subject to Gaussian noise is usually adopted, so that the final probability for the corrected temperature prediction can be obtained through:

$$
\begin{aligned}
\epsilon_{t+1} &= A\epsilon_t + w \\
y_t &= C\epsilon_t + v \\
\mathbb{E}\left[\hat{T}_{t+k}|\hat{T}_{NWP}\right] &= \epsilon^*_{t+k} + \hat{T}_{NWP,t+k} \quad k \in \mathbb{N}_{[1,H]}
\end{aligned}
\tag{1}
$$

where $w\ \mathcal{N}(0, Q)$ and $v\ \mathcal{N}(0, R)$ and ϵ^*_{t+1} is the updated error on the prediction, obtained through a Kalman filter update. In fact, under the assumption of Gaussian model and observation noise, the probability distribution of error ϵ_t can be updated up to time t, and then simulated using the first equation of 1, up to time $t + k$. Three main issues can be identified using this approach:

1. When $T_{obs,t}$ is missing, is not possible to update the Kalman filter estimation for the forecast error. This would result in less accurate corrections.

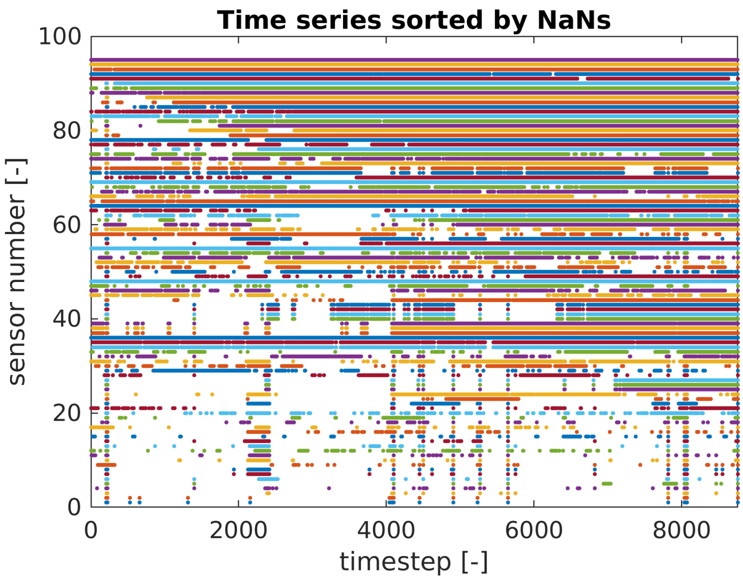

Figure 1: Missing data for the senors located in the sorroundings of Bellinzona, sorted by quantity of available data, for 2018. The dots represent the presence of the signal.

2. At time t, we only possess the error relative to the NWP forecasts up to time $t-1$. This means that the Kalman filter must run for the whole prediction horizon, which would drastically increase the covariance associated with the state estimation.

3. The autoregressive model for the error in 1 is not realistic in general. In fact, NWP services usually provides forecasts for the next 24 or 36 hours only a few times in a day (e.g. at midnight and at noon). This will result in systematic discontinuities in the error distribution, at the update time (if the forecasts are updated at 12, we expect that the error at $t = 13$ will be lower than the error at $t = 11$).

Instead of introducing all these modeling assumptions, we propose to use a regressor-based approach, in which the corrected probability distribution of the forecasted temperature for the next prediction horizon of H timesteps is obtained through H general-purpose interpolator:

$$\mathbb{E}\left[\hat{T}_{t+k}|\hat{T}_{NWP}\right] = f(\hat{T}_{NWP}, T_{obs}, X|t) \quad \forall \quad k \in \mathbb{N}_{[1,H]} \tag{2}$$

where $(z|t)$ refers to the history of variable z up to time t and X is a set of exogenous variables. Comparing 2 with 1, is possible to highlight the differences of the two approaches. The model in 1 uses a recursive strategy to get the updated forecast at time $t+k$. This is done by modeling the probability distribution of the error ϵ as a dynamic linear system. On the other hand, the set of regressors in 2 directly provides the probability distributions of the corrected forecast for all the timesteps in the prediction horizon.

1.3.2 Increasing consumption accuracy

Once the forecasts for the temperatures in a given geographical region have been corrected using historical observations, they can be used to forecast the composite power flow of heterogeneous electrical consumers/producers. We propose to reweight the temperature based on the consumption of the district in which the sensors are located, in order to have a more representative indicator for the power prediction task. The steps we propose in order to exploit the information from the distributed sensor network, are summarized in the following:

1. The temperature forecast $\hat{T}_{NWP}$ for a given region is corrected using historical values of temperature measurements from a group of sensors $\mathcal{S} = \mathbb{N}_{[1,S]}$to obtain more accurate sensor-dependent forecasts $\hat{T}_s$.

2. $\hat{T}_s$ are geographically smoothed using a Gaussian Process (GP), under the assumption of time-invariant GP's hyperparameters.

3. The obtained smoothed forecasts $\hat{T}_s$ are used to obtain an average temperature for each district in the geographic region, $\hat{T}_d$, with $d \in \mathcal{D}, \mathcal{D} = \mathbb{N}_{[1,D]}$ and $D < S$.

4. The district temperature forecasts are weighted by the relative consumption of their district (compared to the region) r_d, obtaining the final forecast for the temperature of the region, $\hat{T}_r = \sum_{d \in \mathcal{D}} r_d \hat{T}_d$

The geographical smoothing is introduced in order to obtain a more realistic representation of the temperature distribution over the region of interests. In fact, the GP finds a generative consistent probabilistic description of the data, assuming that the observations are linked by a given covariance structure. This probabilistic link will ensure that extreme values observed during the period of interest are smoothed out, and will actually de-noise the observations. This technique is also known under the name of kriging in the geographical information systems (GIS) literature [4, 5]. Usually, geographical smoothing does not take into account time-varying process. Two notable exceptions are represented by time-forward kriging, in which an autoregressive time structure is modeled by mean of augmented covariance matrices [6] and by Gaussian Markov random field, which parameters can be learned via integrated nested Laplace approximations [7]. Both these techniques requires a high computational burden. In our case we don't need to consider a temporal structure for the GP, and to apply these techniques will represent an overkill. The authors in [8] propose to fit different sets of hyperparameters of the GP for each time step, in order to predict the spatial distribution of wind pressure. We want however, to identify a common set of hyperparameters for the GP fitted at different timesteps. In fact, the set of hyperparamters defines the covariance structure and the level of noise in the data, which are reasonably time invariant. For this reason we propose to learn the hyperparameters by minimizing the sum of the standardized log loss (SLL) over all the timesteps.

1.4 Description

1.4.1 Dataset description

The used dataset consists in a set of temperature measurements coming from a distributed sensor network, a set of meteorological forecasts coming from a NWP service, and the aggregated power profile of the region in which the temperature sensors are located. All the signals refer to a one year period starting 1st January 2018, and to the region served by AMB. A detailed description of the data is in the following:

1. Temperature measurements coming from Netatmo stations [9]. All the sensors are labeled with their geographical locations in terms of longitude and latitude. The region of interest present 95 sensors, part of which unusable to our purposes, due to the high number of missing values, as shown in Fig. 1. We have discarded all sensors for which the ratio of missing data is greater than 40%. The resulting dataset consists of 65 sensors. The data sampling time is 1 hour.

2. NWP meteorological forecasts coming from meteoblue History+ service [10]. These forecasts are actually obtained by backward simulations of a meteorological model, and are thus de-facto more accurate than actual NPW forecasts. This means that the increase in accuracy of the temperature sensors we will obtain will be conservative w.r.t. what could be obtained using actual NWP forecasts. The data sampling time is 1 hour.

3. Aggregated power profile for the region of AMB, serving 13 commons and around 33500 single clients. The data sampling time is 15 minutes.

For the forecasting of the aggregate profile, we consider a prediction horizon of 36 hours, and assume that the forecasts must be communicated every day at noon for the following day.

1.4.2 Increasing temperature forecast accuracy

We have compared two different general purpose regressors in order to obtain the predicting function f in 2:

- Support vector regression (SVR). This regressor, under the assumption of Gaussian kernels, can be efficiently computed. Unfortunately, there is no out of the shelf method that allows for the use of datasets with missing data. The method will in fact discards (and won't predict) instances with missing regressors. This means that we can restrict f to be a function of the data that we are confident will be always available. This means that the interpolator cannot be function of the observed error ϵ, since this is a function of T_{obs}, which presence is not always guaranteed in this context, coming from an unreliable sensor. For this reason, in this case the regressor dataset is the following:

$$\hat{T}_{s,t+k|t} = f(\hat{T}_{NWP,t:t+k}, \theta_{az,t:t+k}, \theta_{el,t:t+k}, h_t) \quad \forall \quad k \in \mathbb{N}_{[1,H]} \tag{3}$$

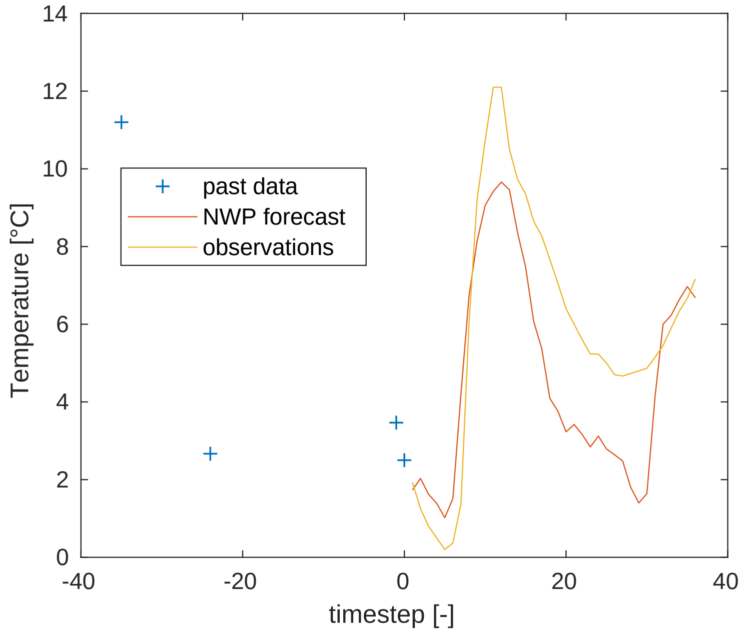

Figure 2: Crosses: example of historical values of the temperature, $T_{obs,t\in\mathcal{T}_{hist}}$, used as regressors with the QRF. Red line: regressor NWP forecasts for the region of interest. Yellow line: observations for the target temperature.

where $\hat{T}_{NWP,t:t+k}$ is a vector of length $k+1$, containing the temperature predicted by the NWP from time t up to time $t+k$, θ_{az}, θ_{el} are the sun azimuth and elevation, and h_t is the hour of the day at time t. The latter is passed to the SVR since, as stated previously, the NWP accuracy could be in general dependent on the time in which the forecast is made available.

- Quantile regression forest (QRF). This interpolator, well known for its performance on a broad variety of regression datasets, can automatically handle missing data, through surrogate splits. In this case, we can exploit also the information coming from the unreliable temperature sensors T_{obs} in order to correct the forecasts. In this case the regressors set can be augmented as follows:

$$\hat{T}_{s,t+k|t} = f(T_{obs,t\in\mathcal{T}_{hist|t}}, \hat{T}_{NWP,t:t+k}, \theta_{az,t:t+k}, \theta_{el,t:t+k}, h_t) \quad \forall \quad k \in \mathbb{N}_{[1,H]} \tag{4}$$

where $\mathcal{T}_{hist,t}$ is a set of past observations at time t, for the temperature sensors whose measurement we are trying to predict. For the numerical test we have chosen $\mathcal{T}_{hist} = -[36, 24, 1, 0]$, where each negative number denotes an hour in the past, and 0 denotes the value for the current hour. An example of the historical data for $T_{obs,t\in\mathcal{T}_{hist}}$ is shown in figure 2.

For each of the two methods, we have obtained the corrected forecasts $\hat{T}_s$ for the whole length of the dataset using 10 folds cross validation. In particular, since we are dealing with time series, and since the training dataset must be obtained

through Hankel matrices, we have built training and testing indices in such a way that training and testing periods are always separated by at least one day of data. In this way, we are sure that none of the data we have use in the test, has been seen by the interpolator during the training phase. In Fig. 3, an example of training and test indexes for the first 20 days, fir the first fold, is shown. Each 10 days, 7 contiguous days are used for training, while 1 is used for testing. The other folds are obtained by shifting the training and test indexes by one day; in this way we can obtain realistic predictions for the whole year.
In Fig. 4 the results in terms of RMSE and MAPE ratios with respect of the NWP forecasts is shown. Formally, the plotted quantities are:

$$\begin{aligned} RMSE_{r,s,k} &= \frac{RMSE_{corr,s,k}}{RMSE_{NWP,s,k}} \\ MAPE_{r,s,k} &= \frac{MAPE_{corr,s,k}}{MAPE_{NWP,s,k}} \end{aligned} \tag{5}$$

where $corr$ stands for the KPI obtained through the interpolators, while the s and k suffixes refers to the s_{th} sensor and k_{th} step ahead. The definition of $RMSE$ and $MAPE$ are the following:

$$\begin{aligned} RMSE &= \sqrt{\frac{1}{n_t}\sum_{t=1}^{n_t}(x_t - \hat{x}_t)^2} \\ MAPE &= \frac{1}{n_t}\sum_{t=1}^{n_t}\left|\frac{x_t - \hat{x}_t}{x_t}\right| \end{aligned} \tag{6}$$

Fig. 4 reports the $RMSE$ and $MAPE$ ratios for all the considered sensors, as a function of the prediction step (in hours). As can be seen, both the RF and SVR formulations consistently provide better KPIs (KPIs ratios below 1) for all the sensors and for each prediction step. In particular, the first row of the plots refers to the SVR interpolator, which doesn't consider historical values from $T_{obs,s}$ for the correction. This results in a constant increase in performance through the whole prediction horizon. As expected, the RF interpolator shows better performances in the first

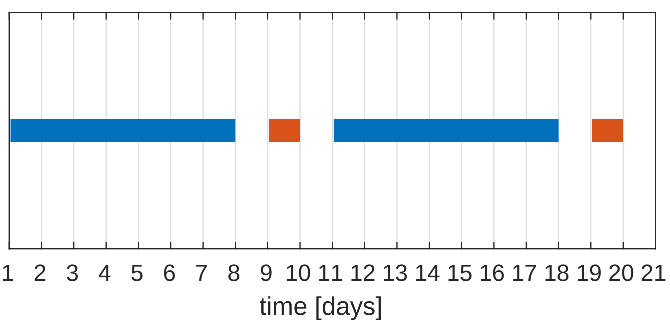

Figure 3: Example of training and test split for the first 20 days. Blue: training days. Red: test days.

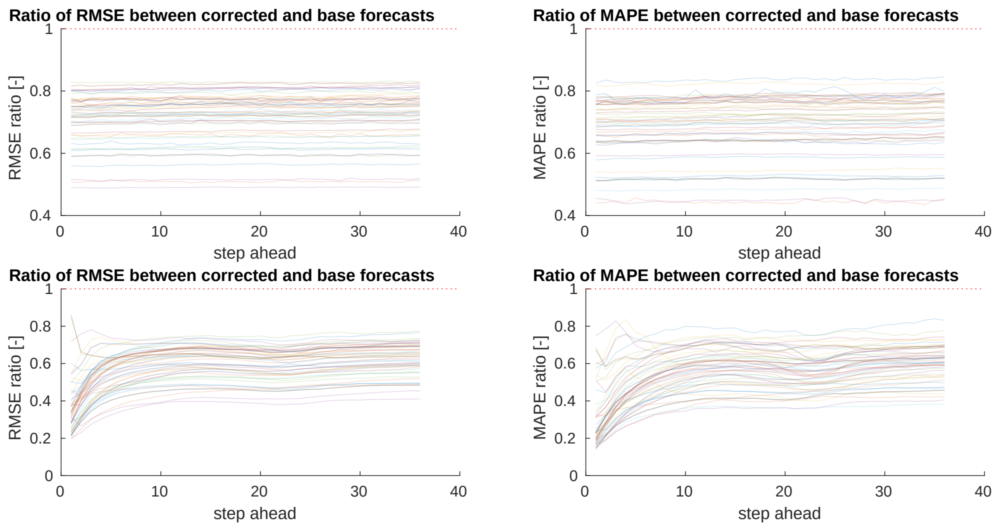

Figure 4: Ratio between the RMSE and MAE of the NWP forecast and of the corrected forecasts, as a function of step ahead. First row: SVR. Second row: RF. Each line indicates the ratios for one of the 59 considered sensors. The horizontal red line highlights the unit value; values below 1 indicates that the forecast correction has increased the accuracy of the prediction.

step ahead, and an increase of performance approaching the one of the SVR with increasing step ahead. We stress out that this increase of performance w.r.t. the SVR is possible due to the ability of the RF to automatically handle missing data in the regressors' dataset.

1.4.3 From temperature to power prediction

As anticipated, the corrected forecasts for the temperature can be used to improve the prediction of a group of prosumers, obtaining an average temperature for the region of interest. In order to be more correlated with the overall power signal, we propose to smooth the forecasts, and weight the single districts temperature with a relative consumption factor, as previously explained in 1.3.2. In order to smooth the forecasts, we use a GP, whose hyperparameters are learned minimizing the marginal log-likelihood of the observed data. Informally, a GP describes the joint probability distribution of the targets, $y \in \mathbb{R}^n$, as a multivariate Gaussian distribution, whose covariance matrix $\Sigma \in \mathbb{R}^{n \times n}$ is defined in terms of similarity between points in the design (or input, or feature) matrix $X_{tr} \in \mathbb{R}^{m \times n}$. We have chosen to adopt a scaled squared exponential kernel as the covariance function:

$$k(x_i, x_j) = e^{-\frac{1}{2}\left(\frac{x_i - x_j}{l}\right)^2} + b \tag{7}$$

where b and l are hyperparameters of the GP. The GP joint distribution for new predictions $\hat{y}$ is then completely defined in terms of the kernel function:

$$\begin{bmatrix} y \\ \hat{y} \end{bmatrix} \sim \mathcal{N}\left(0, \begin{bmatrix} k(X_{tr}, X_{tr}) + \sigma^2 I & k(X_{tr}, X_{te}) \\ k(X_{te}, X_{tr}) & k(X_{te}, X_{te}) \end{bmatrix}\right) \tag{8}$$

where X_{te} is the test input matrix relative to the new predictions $\hat{y}$. Through (8), one can retrieve the conditional expected value for $\hat{y}$:

$$\mathbb{E}\left[\hat{y}|X_{tr}, X_{te}, y_{tr}\right] = k(X_{te}, X_{tr})\left[k(X_{tr}, X_{tr}) + \sigma^2 I\right]^{-1} y_{tr} \tag{9}$$

Furthermore, under the GP assumption of Gaussian prior and likelihood for the predictions, $\hat{y}|X \sim \mathcal{N}(0, k(X, X))$, $y|\hat{y} \sim \mathcal{N}\left(\hat{y}, \sigma^2 I\right)$, one can compute the marginal likelihood (see [11] for details on this integration) and its logarithm w.r.t. the training set:

$$\begin{gathered} p(y|\theta) = \int \mathcal{N}(0, k(X, X))\mathcal{N}\left(\hat{y}, \sigma^2 I\right) \mathrm{d}\hat{y} = \mathcal{N}(0, \sigma^2 I + k(X_{tr}, X_{tr})) \\ L_t(\theta) = log(p(y_{tr,t}|\theta)) = -\frac{1}{2} y_{tr,t}^t \left(\sigma^2 I + k(X_{tr}, X_{tr})\right)\right)^{-1} y_{tr,t} - \frac{1}{2} log(|\sigma^2 I + k(X_{tr}, X_{tr}))|) \end{gathered} \tag{10}$$

where θ is the set of hyperparameters, which in this case is the tuple $\left(\sigma^2, l, b\right)$. We can now optimize the hyperparameters maximizing the log-likelihood on the training set. To do so, we make the aforementioned assumption that the hypeparameters set is time invariant. In our case, the feature matrix is unchanged at each time step, since the location of sensors are also time-invariant. This means that we can precompute $k(X_{tr}, X_{tr})$, and minimize the overall log-likelihood as:

$$\begin{aligned} \theta^* &= \underset{\theta}{\text{argmin}} \; -\sum_{t=1}^{T} L_t(\theta) \\ &= \underset{\theta}{\text{argmin}} \; \frac{1}{2} log(|\aleph(\theta)|) + \sum_{t=1}^{T} \frac{1}{2} y_{tr,t}^t \aleph(\theta)^{-1} y_{tr,t} \end{aligned} \tag{11}$$

where $\aleph(\theta) = \sigma^2 I + k(X_{tr}, X_{tr}))$, which is time invariant. An example of the resulting smoothed temperature field obtained using θ^* is shown in Fig. 5. As in the case of temperature forecast correction, the forecasted aggregated power for the region of AMB can be described in terms of a deterministic function acting on the input space:

$$\hat{P}_{t+k|t} = f(X_k, \hat{T}_{x,t:t+k}, \theta_{az,t:t+k}, \theta_{el,t:t+k}, h_t) \quad \forall \quad k \in \mathbb{N}_{[1,H]} \tag{12}$$

where $\hat{T}_x$ is a forecast for the temperature of the region of interest, obtained with the NWP service or with the method described before in section 1.3.2, and X_k is an additional regressors matrix.

Since for the given test case we are also in possess of the production of a subset of PV panels installed in the region, we include the forecasted PV production as a

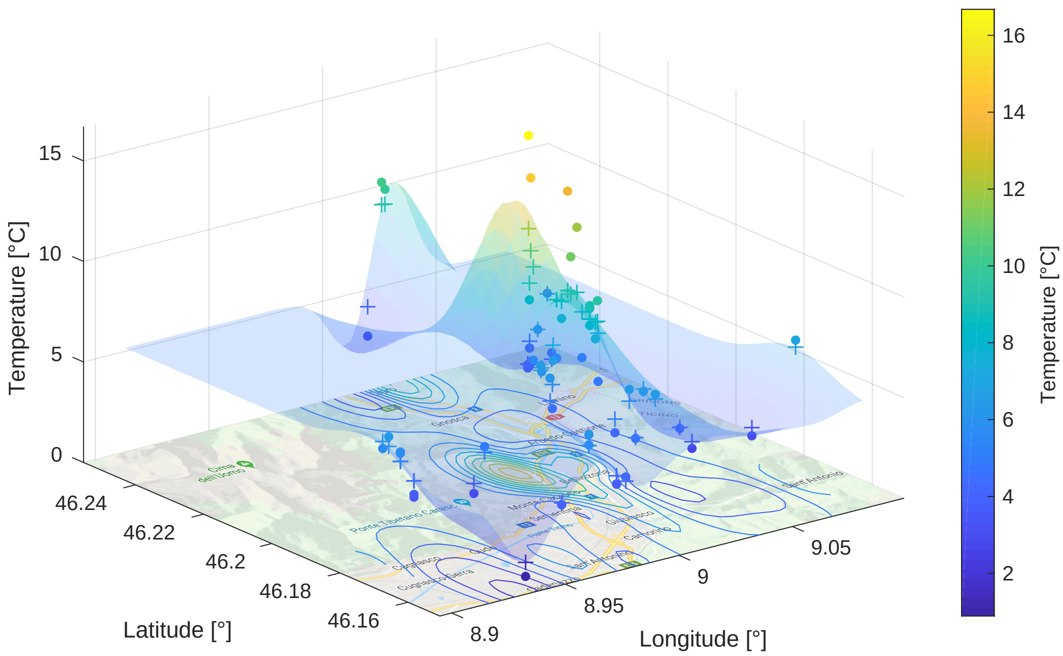

Figure 5: Mean value of the fitted GP. Points: original measures. Crosses: estimated values (projection onto the GP mean value).

regressor, as suggested in [12]. The additional regression matrix is thus defined as:

$$X_k = \hat{P}_{pv,t:t+k} \tag{13}$$

where $\hat{P}_{pv}$ is also obtained through an interpolator, starting from the forecasts for the global horizontal irradiance (GHI) and temperature:

$$\hat{P}_{pv,t+k|t} = f_{pv}(\hat{T}_{NWP,t:t+k}, G\hat{H}I_{NWP,t:t+k}, \theta_{az,t:t+k}, \theta_{el,t:t+k}, h_t) \quad \forall \quad k \in \mathbb{N}_{[1,H]} \tag{14}$$

The results of using the NWP temperature rather than $\hat{T}_r$ in terms of RMSE are shown in Fig. 6. Additionally, the results obtained when the temperature signals are not smoothed out is also shown. First of all, we can see that the RMSE for all the cases has a peak at noon. This can be explained by plotting the daily pattern of the target time series, after a weekly detrending, as shown in Fig. 7. As expected, the variance of the power profiles is maximal during daytime, while shrinking at night. From Fig. 7 it is also possible to see the peak of electric boilers' consumption in the evening, the UTC shift, and the three different clusters of profiles based on the day of the week. Further considerations can be done concerning Fig. 6. The method proposed in section 1.3.2 seems to effectively decreasing the RMSE, especially during night hours, as we would expect due to the influence of the temperature on heat pumps. In table 1 the mean (over the prediction horizon) nRMSE is reported for the base case, the corrected temperature, and the GP smoothed temperature. We can see that, when the corrected temperature is used instead of the NWP one, the nRMSE decreases from $5.84e-2$ to $5.54e-2$, which corresponds to a decrease of 5.14%. Secondly, the GP smoothing doesn't provide an additional increase in the accuracy, as is also shown in table 1. This could mean that the aggregation by district is enough to retrieve a representative temperature for the region of interest, and is robust enough to mitigate the influence of outliers.

Table 1: Mean normalized RMSE and MAPE over the prediction horizon, for different methods of retrieving the forecasted temperature of the region. NWP: the temperature is obtained through numerical weather prediction service. corr: the NWP temperature is corrected using observations from the distributed sensors. GP: the corrected temperature is additionally smoothed using a Gaussian Process.

method	nRMSE	MAPE
NWP	5.84e-02	4.20
corr	5.54e-02	4.08
GP	5.53e-02	4.08

2 Achievement of deliverable

2.1 Date

June 2019.

2.2 Demonstration of the deliverable

This deliverable consists of data analysis work. The comparison of the performance of the various forecasting techniques is presented in the previous sections.

3 Impact

This work explores the possibility of using distributed temperature sensors to increase the accuracy of NWP forecasts at their location. As shown, this approach can also increase the accuracy of the forecast of an aggregated power profile. As a result, the output of this deliverable can be used in the WP, in which the power forecasts for the next day ahead are required.

Acknowledgments

We would like to thank AMB, who kindly provided us their consumption and production data and allowed us to use them for this research work.

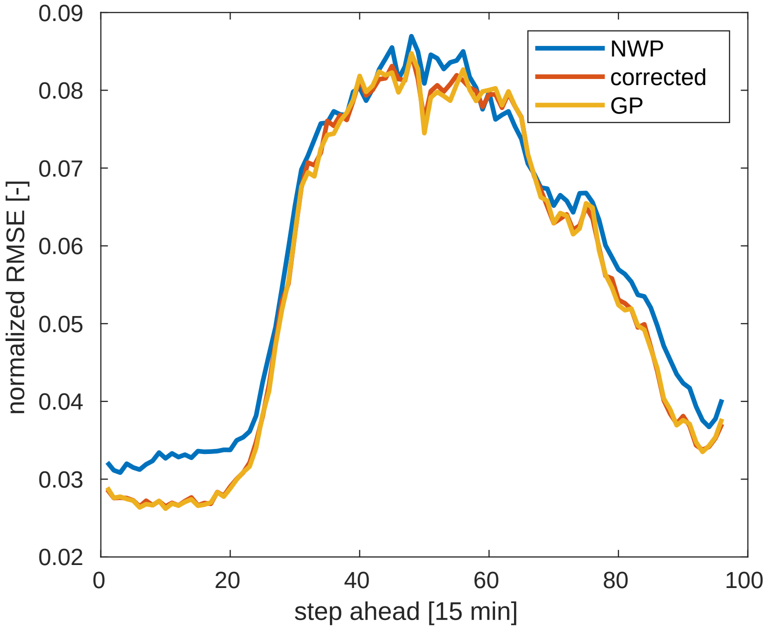

Figure 6: RMSE, normalized with the yearly mean, as a function of step ahead, using NWP forecasts, the corrected temperature, or the same value when the GP is applied.

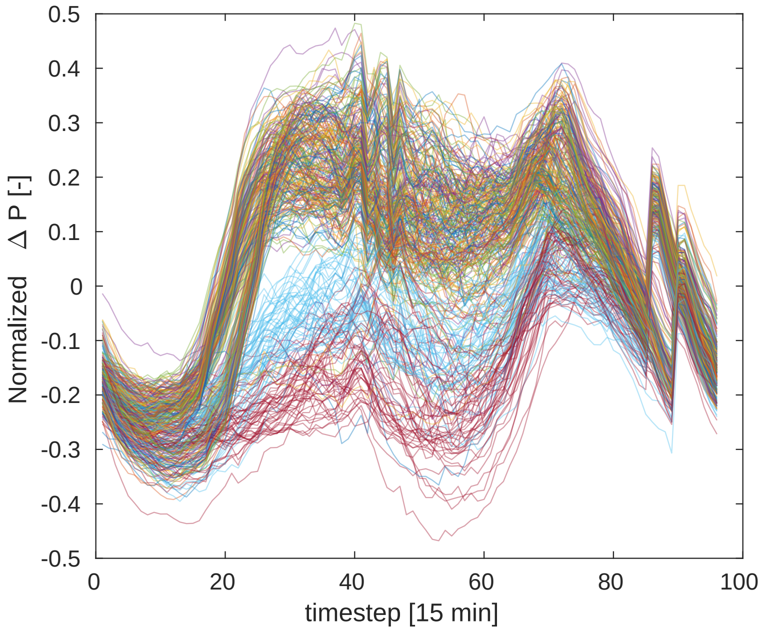

Figure 7: Daily profiles of the aggregated power profile, normalized with the yearly mean.

References

[1] F. Oldewurtel, "Stochastic Model Predictive Control for Energy Efficient Building Climate Control," no. 19908, 2011.

[2] J. R. Porto de Carvalho, E. D. Assad, and H. S. Pinto, "Kalman filter and correction of the temperatures estimated by PRECIS model," *Atmospheric Research*, 2011.

[3] R. Libonati, I. Trigo, and C. C. DaCamara, "Correction of 2m-temperature forecasts using Kalman Filtering technique," *Atmospheric Research*, 2008.

[4] J. P. Kleijnen, "Kriging metamodeling in simulation: A review," 2009.

[5] M. A. Oliver and R. Webster, "Kriging: A method of interpolation for geographical information systems," *International Journal of Geographical Information Systems*, 1990.

[6] G. Chaojun, D. Yang, P. Jirutitijaroen, W. M. Walsh, and T. Reindl, "Spatial load forecasting with communication failure using time-forward kriging," *IEEE Transactions on Power Systems*, vol. 29, no. 6, pp. 2875–2882, 2014.

[7] H. Rue, S. Martino, and N. Chopin, "Approximate Bayesian inference for latent Gaussian models by using integrated nested Laplace approximations," *Journal of the Royal Statistical Society. Series B: Statistical Methodology*, vol. 71, no. 2, pp. 319–392, 2009.

[8] X. Ma, F. Xu, and B. Chen, "Interpolation of wind pressures using Gaussian process regression," *Journal of Wind Engineering and Industrial Aerodynamics*, vol. 188, no. February, pp. 30–42, 2019.

[9] "www.netatmo.com."

[10] "https://www.meteoblue.com/en/historyplus."

[11] C. M. Bishop, *Pattern Recognition and Machine Learning*, vol. 53. 2013.

[12] Y. Wang, N. Zhang, Q. Chen, D. S. Kirschen, P. Li, and Q. Xia, "Data-Driven Probabilistic Net Load Forecasting With High Penetration of Behind-the-Meter PV," *IEEE Transactions on Power Systems*, vol. 33, no. 3, pp. 3255–3264, 2018.