



# ON THE PERSISTENCE FORECAST OF SINGLE HOME ELECTRICITY POWER CONSUMPTION

Naser Farag Abed<sup>1</sup>,  
Milan Milosavljević<sup>2</sup>

<sup>1</sup>Singidunum University  
32 Danijelova Street, Belgrade, Serbia

<sup>2</sup>University of Belgrade,  
Faculty of Electrical Engineering,  
Belgrade, Serbia

## Abstract:

This work analyzes an electricity power consumption forecast for a single home using various types of generalized persistence models. The predictor composes of parallel banks of simple averaging models (SAM) for each hour within a day. Fitting SAM is performed separately for each day of the week using the appropriate training sets of past electricity power consumption. The performance of the predictor was evaluated using real data which represents power consumption per minute measured over almost 4 years for a single home near Paris, France (approximately 2 million data points). Experiments show that the proposed system for predicting power consumption one day ahead gave mean absolute value of relative percentage error (MARPE) lower by more than 18% compared to ordinary persistent prediction, without significant computational cost.

## Key words:

energy consumption prediction; mean absolute value of relative percentage error, persistent prediction, growing window, sliding window.

## 1. INTRODUCTION

Electricity load forecasting has gained substantial importance nowadays in the modern electrical power management systems with elements of smart grid technology. A reliable forecast of electrical power consumption represents the starting point in policy development and improvement of energy production and distribution. At the level of individual households, the ability to accurately predict consumption of electricity power significantly reduces prices by appropriate systems for energy storage. Therefore, the energy efficient power networks of the future will require entirely new ways of forecasting demand on the scale of individual households [4], [5].

In this paper we present our electricity power consumption forecast results for single home data, which we call PARIS data. In order to compare our results with previously published, we choose data set which is comparable with TUM and REDD data sets, for which the prediction accuracy are publicly available [1]. In this way, we have three different data sets with possibility to control same granularity, window of training and forecast horizon. It enables us to crosscheck results and despite different complexity of this three data sets, render fair comparison.

## Correspondence:

Milan Milosavljević

## e-mail:

mmilan@etf.rs



## 2. GENERAL PROPERTIES OF PARIS DATA SET

The data set which we named PARIS data was obtained from the UCI data set repository [2]. It represents electricity power consumption per minute measured between December 16th 2006 and November 26th 2010 (47 months) for a single home near Paris, France, containing over 2 million data points. The data set contains nearly 1.25% of missing values. All experiments were performed in the MATLAB programming environment using Machine Learning Toolbox, [7]. We replace missing values by simple linear interpolation filtering.

All data can be analyzed at different level of time granularity, *i.e.*, sampling frequencies, from one minute to one day. In order to compare our results with the related work in the area, we chose time granularity of one hour. That means that one day of electricity power consumption is represented by 24 data points. Thus, the whole data set is represented by 34440 sampling points. In Figure 1 electricity power consumption is shown in the granularity of one minute (left) and one hour (right) during one day. Change in granularity is implemented by the appropriate moving average filtering. In Figure 2 shows three dimensional representation of whole PARIS data set at hourly granularity, where x axis denotes days (1435 days) while y axis denotes hours (1 -24).

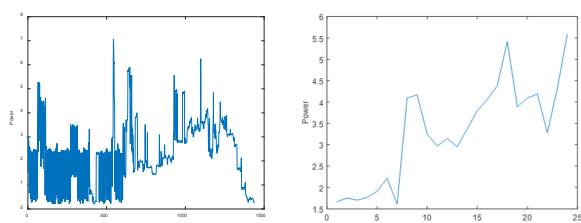


Figure 1. Electricity power consumption in the granularity of minute (left) and hour (right). In the granularity of one minute and one hour, one day is represented by 1440 data points, and 24 data points, respectively.

In Figure 3 shows average of original PARIS data taken over particular days of a week in the form of three dimensional representation (minutes, day, power), while Figure 4 shows long term average over all days. All these findings are the evidence of the fact that the main cause of electricity power consumption behavior is working dynamics of inhabitants during a week. This illustration gives clear evidence as to how hard it is to predict behavior of one day consumption based on long term averages. In this particular example MARPE=98.88%.

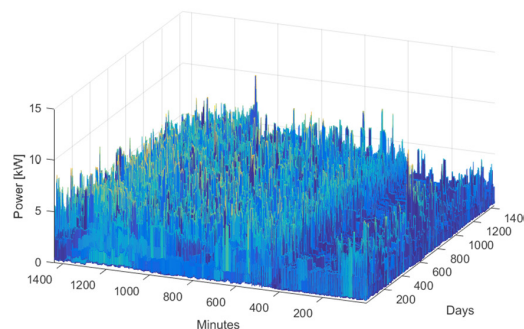


Figure 2. Three dimensional (Days, Minutes, Power) representation of original PARIS data.

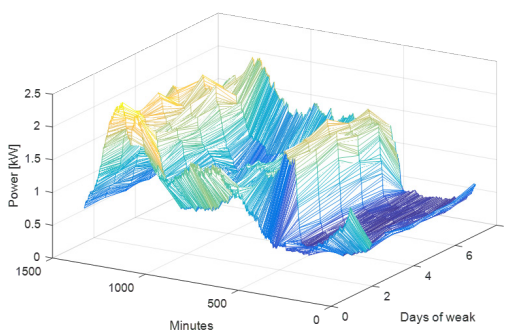


Figure 3. Average of original PARIS data taken over particular days of a week

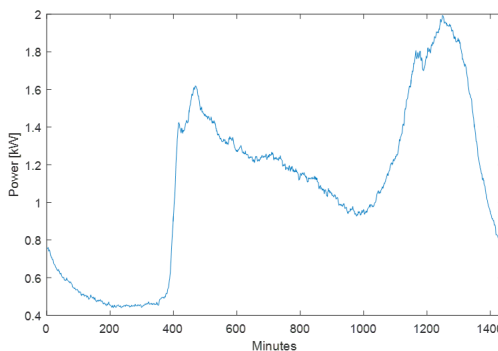


Figure 4. Average of original PARIS data taken over all days. If we use long term averages as prediction for actual data, we get MARPE=98.88%, which is pretty high error.

It is obvious that granularity of data at level of 1 minute produces very large quantity of data, beyond the practical interest for such high resolution. More practical scenario is granularity at level of one hour. Therefore we transform original PARIS data at hourly granularity. In this case, each day is represented by 24 measurements obtained by averaging 60 data for each hour.



### 3. PERSISTENCE FORECAST

The persistence method assumes that the conditions at the time of the forecast will not change, *i.e.*, all forecasts are equal to the last value before starting forecasting horizon. We will refer to this method as PERSIST. Corresponding MARPE has a form

$$MARPE_{per} = \frac{1}{n-1} \sum_{t=2}^{t=n} \left| \frac{x_t - x_{t-1}}{x_t} \right| \cdot 100 [\%] \quad (1)$$

We extend the notion of persistence forecast to the broader class of simple predictors, which gives forecast according to formula

$$\hat{x}_{t+1} = \frac{1}{N} \sum_{n=t-N+1}^t x_n, \quad (2)$$

*i.e.*, by simple averaging past observations within window of specified length  $N$ . We refer to this method as MA\_PERSIST (moving average persistence). Corresponding MARPE has a form

$$MARPE_{MAper} = \frac{1}{n} \sum_{t=1}^{t=n} \left| \frac{x_t - \hat{x}_t}{x_t} \right| \cdot 100 [\%]. \quad (3)$$

#### *Persistence forecast based on long term average*

We start our analysis by simplest persistence forecast model, where predicted value is equal to long term average over whole data. Mean value across whole data is  $96.8 \pm 79.03$  %. Figure 5 shows MARPE obtained by persistence forecast based on the long term average across all days of PARIS data in function of granularity, precisely for granularity at level of 1, 5, 10, 15 30 and 60 minutes. Notice that MARPE decreases with higher granularity, which is the expected effect. Namely, for higher granularity data exhibits lower variability, rendering persistence forecast more accurate.

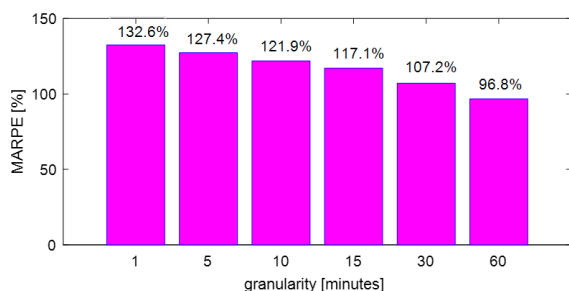


Figure 5. MARPE obtained by persistence forecast based on long term average across all days of PARIS data in function of granularity, ranging from 1 minute to 60 minutes.

In the remaining part of the work, we finally accept granularity of 60 minutes, which is in some sense the compromise between precision of modelling and practical usefulness of obtained results.

#### *Ordinary persistence forecast*

We will name persistence forecast based on formula (1) as the ordinary persistence forecast. Notice that the mean value of MARPE is 81.68% compared to 96.8% mean value of MARPE for long term average, provided the same granularity on the level of 60 minutes. Decrease in the forecasting error can be explained by the fact that ordinary persistence forecast better exploits local behavior of data.

#### *Persistence forecast based on moving window average*

Another class of simple persistence forecasting is forecast based on simple moving average within the sliding window of fixed length  $N$ , see formula (2) and (3).

First question we posted is what optimal length  $N$  is. In Figure 6 we present our experimental relationship between MARPE\_MAPER and length  $N$  of averaging sliding window. From Figure 6 we can see two interesting points, first minima at window size 7, and second minima at window size 1020. This findings suggest that all information needed for optimal forecast are strong local, just 7 days in the past. Therefore, optimal short term moving average persistence forecast system has accuracy 74.55%, and is shown in Figure 7. On the other hand, more information we can get only after very long period of time (minima at 1020, which is equal to the period of more than 2.5 years of observation), with negligible gain in accuracy of just 1.5% MARPE. These findings are important for our main thesis that PARIS data exhibits strong local dynamics.

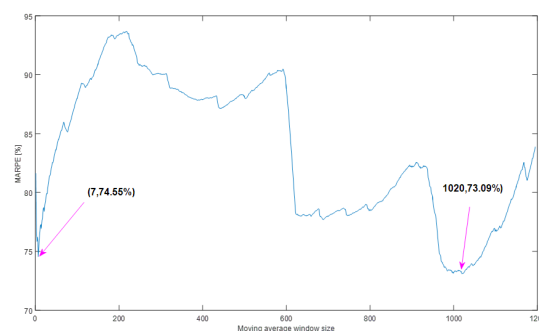


Figure 6. Relationship between moving average window size and MARPE\_MAPER for PARIS data at granularity level of 60 minutes.

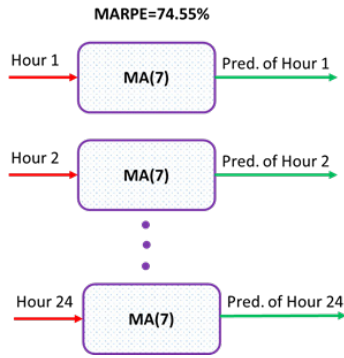


Figure 7. Optimal short term moving average persistence forecast system for PARIS data. Mean accuracy in terms of MARPE = 74.55%.

*Persistence forecast based on growing window average*

Let allow in our moving average persistence forecast system that each hour has its own predictor and that for each of them we find optimal window size. The results are presented in Table 1 and Figure 8 and Figure 9.

It is interesting to note from Table 1 that there two groups of hours: night, early morning and early non hours (1, 2, 3, 4, 5, 6, 7, 12, 13, 14, 15, 16, 17, 18), with extremely large window sizes, and other hours which are characterized by very short optimal window length (8, 9, 10, 11, 19, 20, 21, 22, 23). If we permit that window size of first type of hours grow unlimited, then we obtain very interesting forecasting system based on combined moving average, and growing window persistence forecasting, which means that the value of MARPE for whole PARIS data is 66.53%, see Figure 8 and Figure 9.

In Figure 10 is shown general architecture of this combined forecasting system. It is interesting to note that hours which cover growing window subsystem are locally unpredictable, and need as much information as possible. This hours are very close to subset of night and working hours, see Table 1.

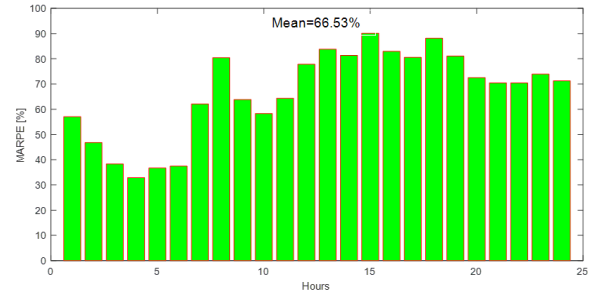


Figure 8. Distribution across the hours of MARPE obtained with combined moving average and growing average persistence forecast optimally designed for each hour of a day.

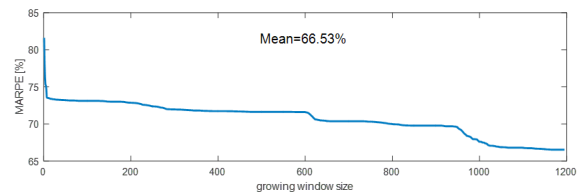


Figure 9. MARPE of combined moving average and growing window average persistence forecasting system in terms of length of growing window.

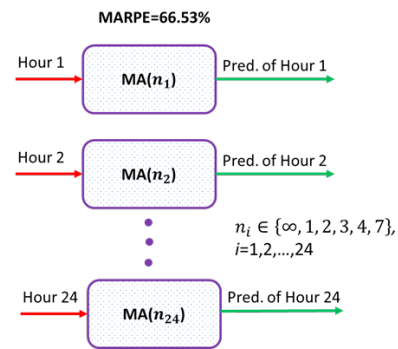


Figure 10. Combined forecasting system based on moving and growing size window averaging. On the PARIS data system gives MARPE=66.53%.

Hour	1	2	3	4	5	6	7	8	9	10	11	12
N	∞	∞	∞	∞	∞	∞	∞	1	1	4	2	∞
Hour	13	14	15	16	17	18	19	20	21	22	23	24
N	∞	∞	∞	∞	∞	∞	7	3	3	3	7	∞

Table 1. Optimal window length N for hourly persistence forecasting based on growing window average. The symbol ∞ means, that window is permanently growing. Table Styles



### Machine learning based forecasting

In order to get an insight into accuracy of SAM models, we just quote results of our broader investigation encompassing all relevant machine learning methods [6]. Especially, we include three classes of most promising approaches: Multilayer Neural Networks, Support Vector Machine (SVM) and K Nearest Neighbor (KNN) Regression, see Table 2.

## 4. SUMMARY OF OBTAINED RESULTS

Table 2 shows accuracy of all applied forecasting models on PARIS data, along with their mutual ranking. The most accurate method is KNN regression with so called combined forecasting function and number of neighbors K=12 [6]. The same findings are shown in Fig.11, along with so called modeling gain.

METHOD	MARPE	Rank
Persistence – global average	98.81%	7
Persistence - ordinary	81.68%	6
Persistence - optimal moving average	74.55%	5
Persistence – optimal growing + moving average	66.53%	4
Multilayer neural networks	61.16%	3
Linear kernel SVM	58.79%	2
KNN regression – combined method (K=12)	47.83%	1

Table 2. Summary accuracy of all applied forecasting models to PARIS data, along with their mutual ranking.

The most accurate method is KNN regression with combined forecasting function [6].

If we consider persistence forecast with global average as our starting point and KNN regression with combined forecasting function as our end point, then equivalent modeling gain is over 50% (98.81% against 47.83%). In the class of SAM models, the best model is optimal combination of moving and growing window model, which gain is 32.67% (98.81% against 66.53%) when the baseline model is Persistence - global average. If the baseline model is Persistence – ordinary, the gain of best SAM model is 18.5% (81.68% against 66.53%).

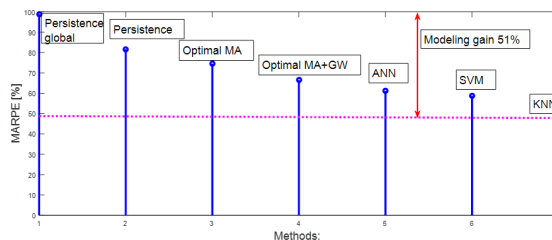


Figure 11. Final comparison of all method applied on PARIS data.

If we consider persistence forecast with global average as our starting point and KNN regression with combined forecasting function, equivalent modeling gain is over 50% (98.81% against 47.83%). The gain of the best SAM model is 32.67% for the same baseline model – Persistence – global average.

What is the main reason for such large modeling gain? Our explanation is the following:

1. PARIS data set possess large variability over days, weeks and seasons during a year. This fact supports high MARPE of persistence forecast (81.68%).
2. Moving average window forecasting method, shows that optimal value for window length is just 7 days. This fact supports strong local dynamics of PARIS data.
3. Good forecasting systems must exploit this two findings in an efficient way.

## 5. CONCLUSIONS

In order to build efficient forecasting systems, we must encompass all relevant information about past consumption data in very short window located as close as possible to the present time. Our feature engineering, did this job generating 36 features, which satisfy locality conditions [6]. This explains our unexpected good results for both SAM and Machine learning based forecasting methods. It is opposite to the statement from [1], where the authors claimed the following: “For short forecasting horizons and high granularities of consumption data, persistence forecasts are known as hard to beat by the other methods”. In this respect, we believe that the combination of local modeling of features and choice of the appropriate machine learning algorithm will be a good guide for building efficient electricity power consumption forecasting systems at the single home level.



## REFERENCES

- [1] Veit, A. *et al.* "Household Electricity Demand Forecasting – Benchmarking State-of-the-Art Methods." EprintarXiv:1040.0200 (2014).
- [2] <http://archive.ics.uci.edu/ml/datasets/Individual+household+electric+power+consumption>, source: Georges Hébrail (georges.hebrail@edf.fr), Senior Researcher, EDF R&D, Clamart, France, Alice Bérard, TELECOM ParisTech Master of Engineering Internship at EDF R&D, Clamart, France.
- [3] Naser Farag Abed, Milan M. Milosavljevic, "Single Home Electricity Power Consumption Forecast Using Neural Networks Model", *IJISSET - International Journal of Innovative Science, Engineering & Technology*, Vol. 3 Issue 1, pp.100 – 106, January 2016.
- [4] S. Chan, K. M. Tsui, H. C. Wu, Y. Hou, Y.-C. Wu, and F. Wu, "Load/price forecasting and managing demand response for smart grids: Methodologies and challenges," *Signal Processing Magazine, IEEE*, vol. 29, no. 5, pp. 68–85, 2012.
- [5] S. D. Ramchurn, P. Vytelingum, A. Rogers, and N. R. Jennings, "Putting the 'Smarts' Into the Smart Grid: A Grand Challenge for Artificial Intelligence," *Communications of the ACM*, vol. 55, no. 4, 2012.
- [6] Naser Farag Abed, *One class of electricity power consumption prediction based on machine learning techniques*, PhD thesis, Singidunum University, 2016.
- [7] <http://www.mathworks.com/products/matlab/>