Generalized Additive Model for Electricity Load Prediction in R

Sovjetka Krstonijević
Mihajlo Pupin Institute, Belgrade, Serbia
sovjetka.krstonijevic@pupin.rs

Abstract: The electricity load forecasting is an important activity in power grid which becomes a critical one in condition of growing demands, the presence of renewable energy sources and electricity market. This entails a question of an adequate electricity load model for forecasts generation. In this work the attention is on the generalized additive model, GAM, the configurable modelling framework, available in R environment, which is suitable for developing various tools for analysis, simulation and prediction. The recent results confirmed its potential and applicability for actual load prediction problems.

Keywords: generalized additive model (GAM); prediction model; electricity load prediction

I. Introduction

The power grid is the most complex man-made system which serves more than five billion people around the globe [1]. All in order to timely deliver the electricity to consumers. In the absence of the electricity storage on a large scale, the energy has to be generated and delivered as it is consumed. Thus, the most of the effort in managing power grid is about balancing the supply and demand as discrepancy between those two could cause the grid instability and unreliability. Consequently, the prediction of the electricity load is one of the most important activities in the power system. The prediction of electricity load is about recognizing and identifying the factors that drive the electricity, deduced from the past load behavior. Then, identified patterns are used to forecast the load expectations. This is the task for load prediction model.

The history of load prediction is over a century long. When lighting was the only electricity consumer, it was enough to estimate the rise of consumption in the evening, by counting the number of the installed bulbs. With the involvement of various electric appliances, the overall consumption exhibited the characteristic weekly and daily patterns, reflecting the consumers' habits of appliance usage. The prediction was based on the identified load patterns. With the further involvement of the heating/cooling systems into the grid, the electricity consumption became sensitive to meteorological factors. So the weather variables, such as temperature, wind, humidity, etc. were included into the model. In today's complex power system, present on electricity market, some economic and socio-demographic indices are, also incorporated into the model.

In the pre-PC era, all load prediction was made manually, using charts and tables. The first computationally supported load models are based on traditional statistical approaches like classical multiple linear regression or various time-series models [2]. Due to their simplicity, they are, still, widely used in practice, for load forecasting functionality, typically on a system or regional level. The new approaches are based on Machine Learning (ML) and Artificial Intelligence (AI) techniques, [3] [4] [5]. Their general advantage over the models of the previous generation is in increased ability to handle the nonlinearities in load, providing better accuracy.

In Smart Grid Era, the efficient grid management is based on ability to predict the load on more local level of aggregation, e.g. cities, neighborhoods, homes, etc. The renewable energy sources and changes in a way how the energy is consumed set new challenges in keeping the system reliable. Such trend changes the standard specification of load model. Not only the accuracy is critical, but, also, its fastness, robustness, adaptability. Most of the previously established models are developed under stable load conditions and customized to a particular part of the system. Although simple for implementation or successful in capturing the nonlinearities, they have some limitations in adapting to data, either because they need human expertise as in ARIMA models, or are computationally expensive as e.g. deep neural networks [6].

As a compromise, the generalized additive model, GAM, offers flexible and configurable framework for developing the load prediction model.It is useful extension of linear model, able to capture complex non-linear relationships, still retaining the simple estimation procedure.Because of its favorable properties it is suitable, both, for simulation and forecasting.

II. GAM in R environment

The generalized additive model, GAM, has been introduced by Hastie and Tibshirani in [7] [8], followed by several important publications concerned with the supporting mathematical foundations and its computational improvements, e.g. as in [9] [10]. However, for practitioners the recommended reading is a book by Simon Wood, *Generalized Additive Models: An Introduction with R* that offers some useful hints for GAM implementation. The GAM is a statistical framework that, generally, covers a wide range of model structures with some common properties:

1) Additivity, as the response variable that can be from the broad range of distribution is obtained as a summation of individual effects, represented with one or more terms.

DOI: 10.5281/zenodo.4067078

Table 1: Overview of the most important R packages for GAM model development

Package	Description	Reference
`mgcv`	Package for GAM model fitting, applicable for family of response distribution. The fitting is based on quadratically penalized likelihood type approach, (PIRLS).Several procedures for regulation of the model terms smoothness are available (GCV/UBRE/AIC/REML).The methods for variable selection, embedded into the fitting procedure are available	[20]
`gam`	Package for GAM model fitting, based on *back-fitting algorithm* that combines different smoothing and fitting methods. (Original GAM fitting version)	[7]
`GAMboost`	`Package for` fitting the GAM model by likelihood based boosting. It is recommended for models with a large number of terms. The procedure implements the mechanism for variable selection.	[11]
`Gamlss`	Package for GAM model development with the extension for Location, Scale and Shape (GAMLSS).It enables modeling all parameters of the response distribution via additive functions of the covariates.	[12]
`SpAM`	*Sparse Additive Model* for sparse estimator that simultaneously enforces smoothness of each term and sparsity across components.	[13]
`COSSO`	*Component Selection Shrinkage Operator* is a package for GAM variable selection. It imposes the lasso-type penalty term for constraining the sum of terms' norms and excludes them from the model through the optimization procedure.	[14]
`CASA`	*Component Automatic Selection in Additive models* variable selection method	[15] [16]
`gamsel`	*Generalized Additive Model Selection* uses penalized likelihood approach for fitting sparse generalized additive models in high dimension. The variable selection is embedded into the estimation procedure.	[17]
`gratia`	Improved *ggp*lot-based graphics and functions for visualization, adapted to the GAMs fitted using the `mgcv` package.	[18]
`mgcViz`	Advanced GAM visualization tool that enables interactive use, GAM additive structure exploitation, scale to large data sets that can be used in conjunction with a wide range of response distributions.	[19]

2) The model terms are smooth functions of a single or more variables and can be, either, linear or nonlinear, mostly represented with the splines. (The splines are typical GAM model building blocks. However, the GAM inherited, also, all types of terms from linear model.)

3) All model terms could be estimated simultaneously and in prediction, their individual contributions are simply summed up. In this regard, some recent developments of the GAM fitting methods, represented in [9] [10] enables its use for exploring very large data sets.

All tools for splines and variable selection, model fitting and validation as well as visualization needed for developing GAM model are available in R. Taxonomy of mostly used R packages for GAM, with the short description and references is given in Table 1. The list is by no means complete, as there are many other GAM-specific and common tools that support GAM development in R.

III. GAM Electricity Load Prediction

The GAM applicability to electrical load forecasting problem at the different network aggregation levels is confirmed in several publications. The good forecasting accuracy (about 1% MAPE error) for French load at the national level is reported in [21]. In addition, the French energy company *Electricite de France*, EDF, has implemented a GAM-based day-ahead load forecasting functionality, on a system level [9]. The improvement for on-line forecasting based on GAM is documented in [22].On the regional level, GAM performances were evaluated on data from National Electricity Market of Australia [23] and US utility company [24]. In [25] its application to the load forecasting at the substation level in France is shown. The flexibility of GAM model applicability to the load prediction problem is illustrated in the paper [26]. A platform for massive-scale load simulation in Smart Grids, starting from the individual households, over low- to medium-voltage network, up to the national level, is presented in this work. Three different scenarios covering the cases of network dynamic reconfigurations, abrupt and gradual changes in consumers' behaviour, and increasing capacity of distributed renewable energy sources is performed using GAM. Regarding the model, it has been shown that GAM is capable to simulate all aggregated loads for over 70 individual households.

A. GAM Load Prediction Model: A Toy Example

A simple GAM structure that implements basic load effects is used to demonstrate the GAM handily implementation in R. The load could be decomposed into the components on different temporal scales, related to the annual, weekly and daily cycles. Then one could suppose that in each instance in day, the load is a summation of the weekly and yearly effects, described with the spline `s(week)` and `s(year)`, respectively. The variable `week` is defined for each day in week, while `year` is for day in the year. Also, the weekly and annual cycles interacts, as weekly pattern changes during the year, that could be described with the multivariate spline `ti(week, year)`, representing only variation in weekly pattern due to year changes. The temperature impact on the load level is incorporated into the model and this nonlinearity is modelled with the spline `s(temperature)`. Based on these assumptions, the load GAM model structure is defined with the sum of identified load components.

For testing the above model, a data set with the hourly loads and daily temperatures for two years, collected from the site [27] was used. The separate model for each hour in day was fitted by using `gam` function from the `mgcv` package for one year sample of data. We used the penalized regression splines, with the control knots, 7 for week (daily knots), 12 for year (monthly knots) and 10 for temperature. For fitting and smoothing methods, the PIRLS/REML was chosen, (see `mgcv`).

Each package for GAM fitting has the standard statistical tools for model validation, residual checking, partial residual checking, etc. (In `mgcv`, typicaly, function `summary`, `residuals`, `gam.check`, `qq.gam` are used for preliminary model checking). The specificity of GAM is that it, also, provides the results for each individual effect that could be observed separately. The `draw` function from the `gratia` package was used to visualize the results of fitting. The weekly, yearly and temperature-based smooths, for one non-working and one working hour, learned from the data, are shown in Figures 1 and 2, respectively.

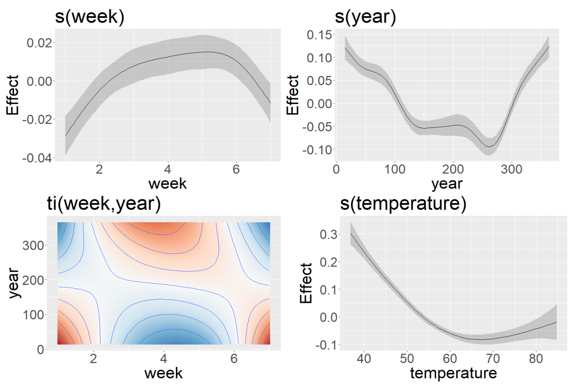

Figure 1: The load effects for the 1.h of the day

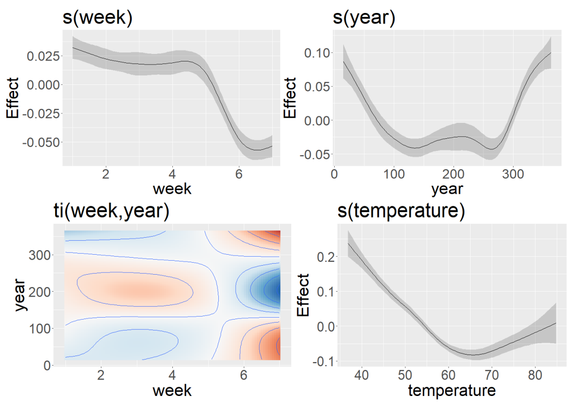

Figure 2: The load effects for the 14.h of the day

Some useful conclusions related to the differences in the weekly and yearly cycles, as well as response to a temperature, for different hours in the day could be drawn!

Based on the previously trained model and new values for temperature, the `mgcv` function `predict` produces the load predictions. The result for each hourly GAM model is combined for obtaining the whole day load prediction. For randomly chosen day, not included in model training, the results of prediction are represented in the Figure 3. The blue line is for predicted values and red for real one. The MAPE error for this day is 2.2%.

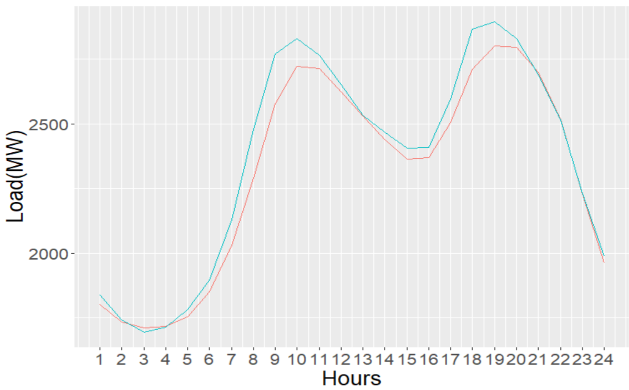

Figure 3: The results of prediction for one day

A stepping forward is in adding more assumption on the load model, which would improve the basic one. The terms for the load from the previous days, weeks, the temperature from the previous days, the other meteorological variables (humidity, wind, cloud cover), as well as economic parameters, could be incorporated into the model. To prevent the model overfitting, some of the variable selection procedure, listed in a Table 1, is available.

IV. Conclusion

The generalized additive model, GAM, and supporting resources for its development in R environment have been briefly overviewed. We are primary motivated by its potential to resolve the important actual problem of electricity load prediction. However, being flexible and easy to implement and interpret, providing high accuracy in prediction and free-software-based it is recommendable for wider application.

References

[1] T. Hong, Load Forecasting: White Paper and Case Study-NARUC, pubs.naruc.org, 2015

[2] J. R. Cancelo, A. Espasa, R. Grafe, "Forecasting the electricity load from one day to one week ahead for the Spanish system operator," *International Journal of Forecasting* 24(4), pp. 588–602,2008

[3] R. Wang, J. Wang, Y. Xu, "A Novel Combined Model Based on Hybrid Optimization Algorithm for Electrical Load Forecasting," *Applied Soft Computing Journal*, vol. 82, September 2019.

[4] J. G. Jetcheva, M. Majidpour, W. P. Chen, "Neural Network Model Ensembles for Building Level Electricity Load Forecasts," *Energy and Buildings*, vol. 84, pp. 214–223, 2014

[5] E. Ceperic, V. Ceperic, A. Baric, "A Strategy for Short-Term Load Forecasting by Support Vector Regression Machines," *IEEE Transactions on Power System*, Vol.28, Issue: 4 , Nov. 2013

[6] X. Liu, Z. Zhang, "A comparative study of the data-driven day-ahead hourly provincial load forecasting methods: From classical data mining to deep learning, "*Renewable and Sustainable Energy Reviews*, Vol. 119 , March 2020

[7] T. J. Hastie and R. J. Tibshirani, *Generalized additive models*, London: Chapman & Hall, 1990

[8] T. J. Hastie, R. J. Tibshirani, and J. H. Friedman, *The elements of statistical learning: data mining, inference, and prediction*, Springer series in statistics, Springer, New York, January 2017

[9] S. Wood, Y. Goude, S. Shaw, "Generalized additive models for large data sets,*" Journal of the Royal Statistical Society: Applied Statistics*, Vol.64 (1), pp. 139–155, 2015

[10] S. N. Wood, Z. Li, G. Shaddick, N. H. Augustin, "Generalized Additive Models for Gigadata: Modeling the U.K. Black Smoke Network Daily Data," *Journal of the American Statistical Association*, Vol.0, No.0,1-12,2017

[11] G. Tutz, H. Binder, "Generalized additive modeling with implicit variable selection by likelihood-based boosting," *Biometrics*, Vol. 62, pp. 961-971, December 2006

[12] R. A. Rigby, D. M. Stasinopoulos, "Generalized Additive Models for Location, Scale and Shape," *Journal of the Royal Statistical Society: Applied Statistics*, Vol. 54(3), pp. 507–554, 2005

[13] P. Ravikumar, J. Lafferty, H. Liu, L. Wasserman, "Sparse Additive Model," *Journal of the Royal Statistical Society: Statistical Methodology*, Vol. 71(5), November 2009

[14] Y. Lin, H. H. Zhang, “Component Selection and Smoothing in Multivariate Nonparametric Regression,“ *The Annals of Statistics*, Vol. 34(5), pp. 2272-2297, October 2006

[15] A. Antoniadis, X. Brossat, Y. Goude, J.M. Poggi, V. Thouvenot, *R package CASA: Component Automatic Selection in Additive models*, UseR! `a Aalborg, 2015

[16] V. Thouvenot, *Estimation et sélection pour les modèles additifs et application à la prévision de la con sommation électrique*, Ph.D. dissertation, Université Paris-Saclay, Français, 2015

[17] A. Chouldechova, T. Hastie, "Generalized Additive Model Selection*", Annals of Applied Statistics*, Stanford University, 2015

[18] G. Simpson, *Introducing gratia*, Oct. 23, 2014. [Online]. Available: https://fromthebottomoftheheap.net/2018/10/23/introducing-gratia/

[19] M. Fasiolo, R. Nedellec, Y. Goude, and S. Wood, "Scalable visualisation methods for modern Generalized Additive Models", *Journal of Computational and Graphical Statistics*, Vol.29(1), July 2019

[20] Wood, *Generalized Additive Models: An Introduction with R*, Second Edition, Chapman & Hall/CRC Texts in Statistical Science, 2006

[21] A. Pierrot, Y. Goude, "Short-term electricity load forecasting with generalized additive models", in *Proc. of ISAP power*, pp.593-600, 2011

[22] A. Ba, M. Sinn, Y. Goude, P. Pompey, "Adaptive learning of smoothing functions: Application to electricity load forecasting," *Advances in Neural Information Processing Systems*, 25, pp. 2519–2527, 2012

[23] S. Fan. R. J. Hyndman, "Short-term load forecasting based on a semi-parametric additive model", *IEEE Transactions on Power Systems*, 27(1): 134–141, 2012

[24] R. Nedellec, J. Cugliari, Y. Goude, "GEFCom2012: Electricity load forecasting and back casting with semi-parametric models", *International Journal of Forecasting*, 2013

[25] Y. Goude, R. Nedellec, N. Kong. "Local short and middle term electricity load forecasting with semi-parametric additive models," *IEEE Transactions on Smart Grid*, 2013

[26] P. Pompey, A. Bondu, Y. Goude, M. Sinn "Massive-Scale Simulation of Electrical Load in Smart Grids Using Generalized Additive Models," in *Modeling and Stochastic Learning for Forecasting in High Dimensions*, Vol 217. pp. 193-212, Springer, 2015

[27] Available: http://fkeynia.googlepages.com/loaddata [Accessed 08.09.2020]