

RESEARCH

Open Access



Towards smart energy systems: application of kernel machine regression for medium term electricity load forecasting

Miltiadis Alamaniotis^{1*}, Dimitrios Bargiotas² and Lefteri H. Tsoukalas¹

*Correspondence:

malamani@ecn.purdue.edu

¹ Applied Intelligent Systems Laboratory, School of Nuclear Engineering, Purdue University, 400 Central Dr., West Lafayette, IN 47907, USA
Full list of author information is available at the end of the article

Abstract

Integration of energy systems with information technologies has facilitated the realization of smart energy systems that utilize information to optimize system operation. To that end, crucial in optimizing energy system operation is the accurate, ahead-of-time forecasting of load demand. In particular, load forecasting allows planning of system expansion, and decision making for enhancing system safety and reliability. In this paper, the application of two types of kernel machines for medium term load forecasting (MTLF) is presented and their performance is recorded based on a set of historical electricity load demand data. The two kernel machine models and more specifically Gaussian process regression (GPR) and relevance vector regression (RVR) are utilized for making predictions over future load demand. Both models, i.e., GPR and RVR, are equipped with a Gaussian kernel and are tested on daily predictions for a 30-day-ahead horizon taken from the New England Area. Furthermore, their performance is compared to the ARMA(2,2) model with respect to mean average percentage error and squared correlation coefficient. Results demonstrate the superiority of RVR over the other forecasting models in performing MTLF.

Keywords: Relevance vector regression, Gaussian process regression, Medium term load forecasting, Smart energy systems

Introduction

Limitations in current power infrastructure together with world-wide concerns, like climate change and economic stability are the driving factors to ongoing research efforts for developing a new generation of smart energy systems (Fainti et al. 2014). Realization of smart energy systems is greatly accommodated by coupling information technologies with power systems. In particular, the advent of internet and advancements in communication technologies inspired the notion of an Energy Internet (Alamaniotis et al. 2011a, b), in which information networks interact with power generation, transmission, and distribution systems aiming at optimizing power system operation.

Smart energy systems utilize information to overcome the significant constraints of the current power grid infrastructure (Tsoukalas and Gao 2008). The limited delivery capacity and the lack of large scale energy storage may lead to grid destabilization causing distribution failures with high financial impact to grid participants. For instance, (i)

load demand beyond delivery capacity results in financially expensive system failures and blackouts (Alamaniotis et al. 2014b), and (ii) the amount of excess generated energy that cannot be stored is wasted since the generation does not closely follow the demand (Gao et al. 2003).

Electricity load forecasting has been recognized as a key issue in implementing smart energy systems (Alamaniotis et al. 2014a, b). Load forecasting may be used by all smart grid participants aiming at reaching their goals. For example, consumers utilize load forecasting for consumption planning and scheduling while grid operators for safe and secure electricity delivery. Depending on the forecasting time horizon, load forecasting may be identified as very short term (VSTLF) ranging from some minutes to an hour (Alamaniotis et al. 2012), short term (STLF) (Alamaniotis et al. 2011a, b) ranging from an hour to a week, medium term (MTLF) ranging from a week to a year (Ghiassi et al. 2006), and long term load forecasting (LTLF) for longer than a year ahead of time predictions (Kandil et al. 2002).

The current manuscript focuses on medium term load forecasting. MTLF is an efficient tool for implementing smart energy systems since it promotes optimal expansion planning by considering climate changes, maintenance scheduling, fuel purchase negotiating (for instance for nuclear power plants), component replacing or repairing, and maximizing utilization of renewable resources such as wind power. Furthermore, it is expected to play a crucial role in developing price directed energy markets in which entities will participate via intelligent meters (Gatsis and Giannakis 2012) and require forecasting tools to develop their electricity purchase strategies.

Though the number of proposed approaches for performing MTLF is limited, there are ongoing efforts for developing more sophisticated and advanced tools that satisfy the demands imposed by the advent of the “big data” era. The proposed approaches make use of tools coming from statistics and artificial intelligence fields. A dynamic artificial neural network is proposed in (Ghiassi et al. 2006), and a radial basis function neural network in (Xia et al. 2010), while combination of neural networks with expert systems in (Kim et al. 1995). Other methods employed adaptive neural networks (Tsekouras et al. 2006), particle swarm optimization (Rengcun et al. 2008), and singular value decomposition (Abu-Shikhah and Elkarmi 2011). Nonlinear multivariable regression for MTLF is presented in (Tsekouras et al. 2007), while a combination of linear and non-linear regression for MTLF is introduced in (Abu-Shikhah et al. 2011), and Gaussian processes for a year ahead monthly load forecasting in (Alamaniotis et al. 2014a). Furthermore, a support vector machine based approach for MTLF is discussed in (Bozic and Stojanovic 2011), while a hybrid methodology comprised of autoregressive integrated moving average (ARIMA) and artificial neural network is introduced and tested in (El Desouky and Elkateb 2000). The above methodologies, though effective, come at a cost of high prediction uncertainty. In addition they lack the necessary flexibility to update their predictions since they are unable to capture nonlinear load dynamics.

In this paper intelligent regression models for MTLF are examined. The proposed models make use of machine learning tools and more specifically of kernel machines (Scholkopf and Smola 2001). In particular, relevance vector regression (Tipping 2001) and Gaussian process regression (Rasmussen 2006) are utilized for making predictions for longer than a week ahead of time horizon. Generally speaking, kernel machines are

nonlinear methods that inherently make use of semi-positive definite matrices in order to make predictions (Hoffman et al. 2008). They are able of detecting the kind of dependencies that dominate the load properties by formulating the feature space in terms of kernels. Formulation of feature space by kernels is the advantage of kernel machines as opposed to the rest load forecasting methods mentioned earlier; it allows the modeler to control the forecasting process by selecting the kernel form, and promotes model flexibility by offering a high variety of kernels (Alamaniotis et al. 2015). For instance, kernel regression facilitates selection of a kernel that models particular data properties, for example stationarity, in contrast to artificial neural networks that require not only selection of neuron activation functions but also network architecture (Tsoukalas and Uhrig 1997). Assessment of the forecasting performance is done using the mean average percentage error (MAPE) and squared correlation coefficient (R^2), while the testing datasets are comprised of the daily demand for a 30-day-ahead horizon.

The roadmap of the paper is as follows: in the next two sections a brief presentation on kernel machines is provided and the proposed methodology is presented. Medium term load forecasting results are given in the “Results” section, while the last section concludes and summarizes the main points of the paper.

Background

Kernel machines

Analytical models that can be expressed as a function of a kernel are known as kernel machines (Bishop 2006). A kernel is any valid mathematical function that can be written with respect to the dual representation. The general form of the dual representation is given by:

$$k(x_1, x_2) = \varphi(x_1)^T \varphi(x_2) \quad (1)$$

with $\varphi(x)$ being any analytical function known as *basis function*, and $k(x, x)$ representing a kernel function. In general, formulating a function using Eq. (1) is known as the kernel trick. A few examples of common kernel functions are the linear and the polynomial kernels whose analytical formulas are given respectively by (Bishop 2006):

$$k(x_1, x_2) = x_1^T x_2 \quad (2)$$

$$k(x_1, x_2) = \left(x_1^T x_2 \right)^2. \quad (3)$$

Beyond the widely known kernels, new valid kernels may be created by composition of two, or more, valid kernels by applying the operations of addition and/or multiplication (Rasmussen 2006). The selection of an appropriate kernel function is a main design choice that must generally be made by the designer according to the specifications of the problem at hand.

Gaussian process regression

The set of random variables that has a joint Gaussian distribution is defined as a Gaussian process. A Gaussian process is fully determined by its mean $m(x)$ and covariance function $C(x, x')$, and therefore, the Gaussian process takes the form:

$$GP(m(x), C(x, x')) \tag{4}$$

where it is common to assume for convenience that $m(x) = 0$.

Gaussian processes are applied in regression problems where they deal with problems of predicting continuous parameters. Derivation of Gaussian process regression (GPR) has as a point of start the simple linear regression:

$$y(\mathbf{x}, \mathbf{w}) = \sum_{i=1}^N w_i \phi_i \tag{5}$$

where w_i are the regression weights and ϕ_i are the basis functions. Equation (5) may be written in vector form as given below:

$$y_n = \Phi \mathbf{w}. \tag{6}$$

Next, a prior normal distribution over the model weights is adopted:

$$P(\mathbf{w}) = N(\mathbf{0}, \sigma_w^2 \mathbf{I}) \tag{7}$$

where $\mathbf{0}$ represents the mean vector, σ_w^2 is the variance equal for all individual weights, and \mathbf{I} is the identity matrix. Therefore, the distribution over the vector output \mathbf{y} is also normal:

$$P(\mathbf{y}) = N(\mathbf{0}, \mathbf{C}_y). \tag{8}$$

Regression problems require taking into account noisy observed target values. If ε_n denotes the additive noise with zero mean and variance σ_n^2 , then the target values become:

$$t_n = y(\mathbf{x}^{(n)}) + \varepsilon_n. \tag{9}$$

Hence, the distribution over the target variables is also normal

$$P(\mathbf{t}) = N(\mathbf{0}, \mathbf{C}) = N(\mathbf{0}, \mathbf{C}_y + \sigma_n^2 \mathbf{I}). \tag{10}$$

In Gaussian process regression the Bayesian formalism is applied in order to infer a predictive distribution, i.e. a mean value and the associated variance. The prediction over the target t_{N+1} for an unknown input $\mathbf{x}^{(N+1)}$ is based on the previous observed targets \mathbf{t}_N and the respective inputs \mathbf{x}_N and thus the predictive distribution becomes

$$P(t_{N+1} | \mathbf{t}_N) \propto \exp \left\{ -\frac{1}{2} \begin{bmatrix} \mathbf{t}_N & t_{N+1} \end{bmatrix} \mathbf{C}_{N+1}^{-1} \begin{bmatrix} \mathbf{t}_N^T \\ t_{N+1} \end{bmatrix} \right\} \tag{11}$$

where it is apparent that the predictive distribution depends on the inverse of the covariance matrix \mathbf{C}_{N+1} . In order to ease computation of the predictive distribution parameters, the covariance matrix \mathbf{C}_{N+1} is subdivided into four submatrices (Williams 2002)

$$\mathbf{C}_{N+1} = \begin{bmatrix} [\mathbf{C}_N] & [\mathbf{k}] \\ [\mathbf{k}^T] & [k] \end{bmatrix} \tag{12}$$

with \mathbf{C}_N being the covariance matrix of the N observations, \mathbf{k} being a vector of length N encompassing the covariances between the $N + 1$ and each of the rest N points, and k

being the scalar value of the variance of the point $N + 1$. Thus, it can be shown (Mackay 1998) that the parameters of the normal predictive distribution, i.e. the mean and the covariance over $N + 1$, are given by the following formulas respectively:

$$m(\mathbf{x}^{(N+1)}) = \mathbf{k}^T \mathbf{C}_N^{-1} \mathbf{t}_N \tag{13}$$

$$\sigma^2(\mathbf{x}^{(N+1)}) = k - \mathbf{k}^T \mathbf{C}_N^{-1} \mathbf{k} \tag{14}$$

where it is noted that both equations depend on covariance matrix \mathbf{C}_N instead of \mathbf{C}_{N+1} .

Relevance vector regression

In the current manuscript we consider the regression form of relevance vector machines, which is known as relevance vector regression (RVR). In deriving RVR, initially, we assume that the target variable t given an input \mathbf{x} follows a normal distribution:

$$p\left(t|\mathbf{x}, \mathbf{w}, \frac{1}{\sigma^2}\right) = N(t|y(\mathbf{x}), \sigma^2) \tag{15}$$

where σ^2 is the variance of the data noise while the mean value $y(\mathbf{x})$ is given by:

$$y(\mathbf{x}) = \sum_{n=1}^M w_n \phi_n(\mathbf{x}) = \mathbf{w}^T \boldsymbol{\phi}(\mathbf{x}) \tag{16}$$

where $\phi()$ is a valid function called the *basis function*, M is the population of basis functions and \mathbf{w} is the weight vector. By using Eq. (16) and kernel functions, RVR is modeled as below:

$$y(\mathbf{x}) = \sum_{n=1}^N w_n k(\mathbf{x}, \mathbf{x}_n) + b \tag{17}$$

with b is the bias term and N is the population number of known observations (i.e., size of training dataset). Next, we consolidate the N input observations into a single matrix \mathbf{X} , and the respective N outputs into a vector \mathbf{t} . Thus, we get a likelihood function:

$$p\left(\mathbf{t}|\mathbf{X}, \mathbf{w}, \frac{1}{\sigma^2}\right) = \prod_{n=1}^N p\left(t_n|\mathbf{x}_n, \mathbf{w}, \sigma^2\right) \tag{18}$$

and a prior distribution over the weight vector \mathbf{w} :

$$p(\mathbf{w}|\boldsymbol{\alpha}) = \prod_{n=1}^M N(w_n|0, \alpha_n) \tag{19}$$

with α_n being the variance of weight w_n and M equal to $N + 1$. At this point we plug into the Bayes formula both Eq. (18) and (19) and hence we get the posterior distribution over \mathbf{w} :

$$p(\mathbf{w}|\mathbf{t}, \mathbf{X}, \boldsymbol{\alpha}, \sigma^2) = N(\mathbf{w}|\mathbf{m}, \boldsymbol{\Sigma}) \tag{20}$$

where mean is taken by:

$$\mathbf{m} = \frac{1}{\sigma^2} \Sigma \Phi^T \mathbf{t} \tag{21}$$

and respective variance by

$$\Sigma = \left(\mathbf{A} + \frac{1}{\sigma^2} \Phi^T \Phi \right)^{-1} \tag{22}$$

with $\mathbf{A} = \text{diag}(\alpha_i)$ and $\Phi = \mathbf{K}$; \mathbf{K} is a $(N + 1) \times (N + 1)$ dimensional matrix with elements given by the kernel function $k(x_n, x_m)$.

At this point it should be said that the unknown parameters α_i and σ^2 are evaluated by maximizing the logarithmic marginal likelihood:

$$\ln p\left(\mathbf{t} | \mathbf{X}, \boldsymbol{\alpha}, \frac{1}{\sigma^2}\right) = \ln N(\mathbf{t} | 0, \mathbf{C}) = -\frac{1}{2} \left\{ N \ln 2\pi + \ln |\mathbf{C}| + \mathbf{t}^T \mathbf{C}^{-1} \mathbf{t} \right\} \tag{23}$$

where $\mathbf{t} = (t_1, \dots, t_N)^T$ and \mathbf{C} is a $N \times N$ dimension matrix given by:

$$\mathbf{C} = \frac{1}{\sigma^2} \mathbf{I} + \Phi \mathbf{A}^{-1} \Phi \tag{24}$$

where \mathbf{I} is respectively the identity matrix.

Maximization of the marginal likelihood in Eq. (23) with an appropriate iterative method allows evaluation of its parameters. Therefore, the computed optimal values for $\boldsymbol{\alpha}$ and σ^2 are equal to $\boldsymbol{\alpha}^*$ and $(\sigma^2)^*$ respectively. Some of the elements of the vector $\boldsymbol{\alpha}^*$ are driven to infinity and thus the posterior distribution of their weights is normal with both mean and variance being equal to zero. As a result, the corresponding kernel functions have no contribution in prediction making driving the output to depend exclusively on the *non-zero weighted kernels*. The inputs associated with non-zero weighted kernels are called *relevance vectors*.

Therefore, RVR provides a predictive distribution over the target value t of a new input \mathbf{x} :

$$p\left(t | \mathbf{x}, \mathbf{X}, \mathbf{t}, a^*, \frac{1}{(\sigma^2)^*}\right) = N\left(t | \mathbf{m}^T \phi(\mathbf{x}), \sigma^2(\mathbf{x})\right) \tag{25}$$

with mean to be obtained by

$$\mathbf{m}^T \phi(\mathbf{x}) = \left(\frac{1}{(\sigma^2)^*}\right) \Sigma \Phi^T \mathbf{t} \phi(\mathbf{x}) \tag{26}$$

and variance by:

$$\sigma^2(\mathbf{x}) = \left(\frac{1}{(\sigma^2)^*}\right)^{-1} + \phi(\mathbf{x})^T \Sigma \phi(\mathbf{x}) \tag{27}$$

where $\phi()$ is vector of basis functions with non-zero elements for relevance vectors and zeros for the rest.

Medium-term-load-forecasting using kernel machine regression

Electricity load demand is a highly volatile signal and depends upon various factors such as: climate, day of the week, season. Capturing the dynamics of all those factors requires the use of appropriate datasets for training the kernel machines. In the present work the training datasets are consisted of historical electric load data of 1 month, 1 year, 2 years and 3 years earlier than the “target 30-day interval”. For convenience, Fig. 1 depicts the way training datasets are composed.

In the current manuscript we focus on applying kernel machine regression for medium term load forecasting. In particular, we adopt two kernel based methods for MTLF:

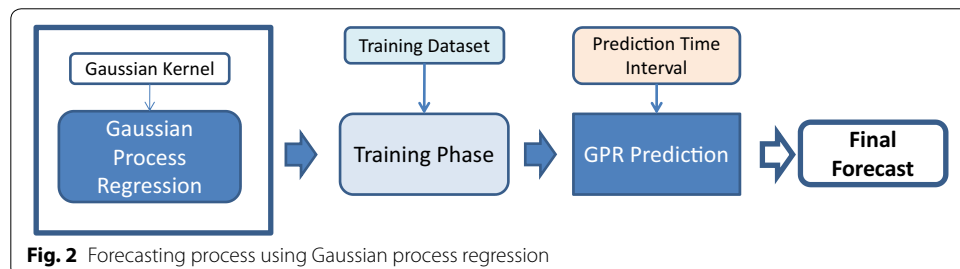
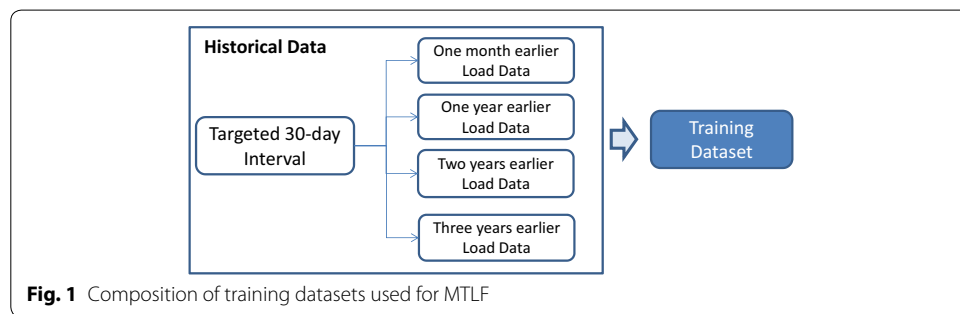
- (i) Gaussian process regression model equipped with a Gaussian kernel, and
- (ii) Relevance vector regression model equipped with a Gaussian kernel.

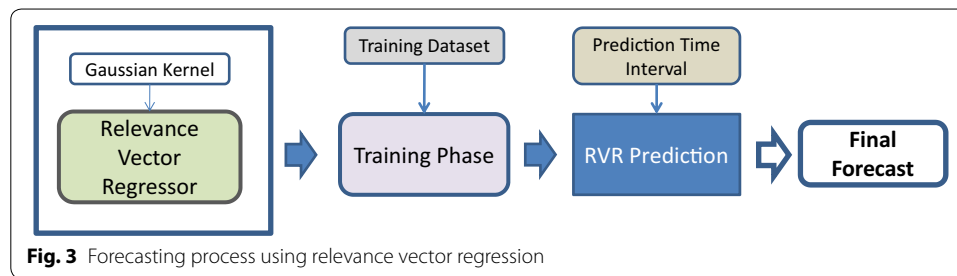
It should be noted that both forecasting models are kernel machines that are modeled using the Gaussian kernel whose analytical formula is given below (Bishop 2006):

$$k(x_1, x_2) = \exp \left(-\|x_1 - x_2\|^2 / 2\sigma^2 \right) \tag{28}$$

with σ^2 denoting a kernel parameter evaluated using the training data.

The block diagrams of applying GPR and RVR models in MTLF are presented in Figs. 2 and 3 respectively. We observe the process of forecasting being the same for both kernel machines; the difference lies in the model, i.e., GPR against RVR. Initially, the kernel machine is trained using the training data aiming at evaluating its kernel parameters. Once training ends, the model is suitable for prediction making. To that end, the trained kernel machine provides the final forecasts on the electricity load demand with respect





to a predetermined ahead-of-time horizon. The above process is repeated for both kernel machines in every targeted time interval.

In our study, we aim at making daily predictions for a 30-day-ahead horizon. Thus, the goal is to predict the load demand for every day in the next 30 days (overall 30 predicted values). To that end, we have our forecasters making predictions on a monthly basis (January–December) and therefore our study falls within the purpose of MTLF.

MTLF results

Problem statement

We apply the presented forecasters to medium term load forecasting for electricity demand load data obtained from the New England ISO (last accessed in 2015) for the period January 2004–August 2011. In particular, we analyze historical load datasets that represent the daily load demand in one of the hubs of the New England ISO Area. Taking into consideration the historical data at our disposal, the forecasters are applied to forecasting demand from January 2007 to August 2011.

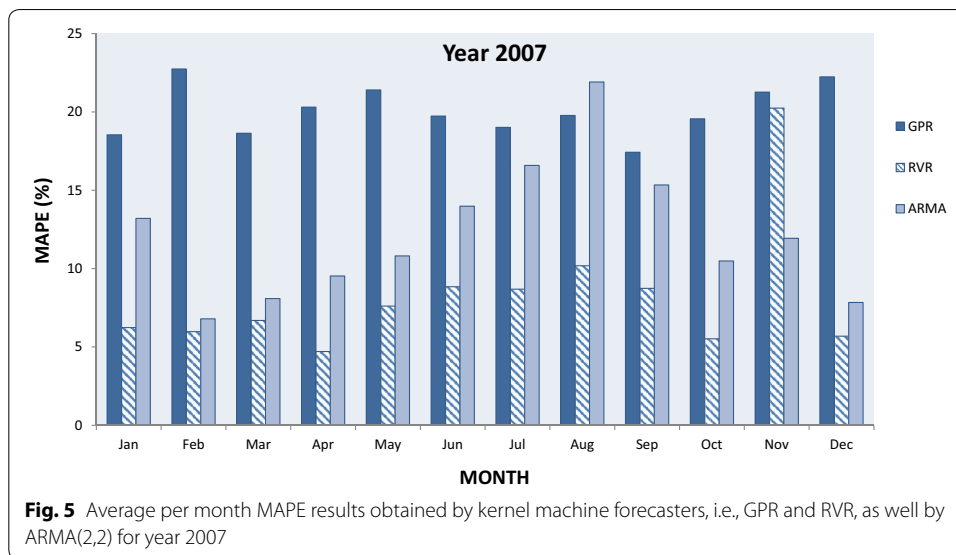
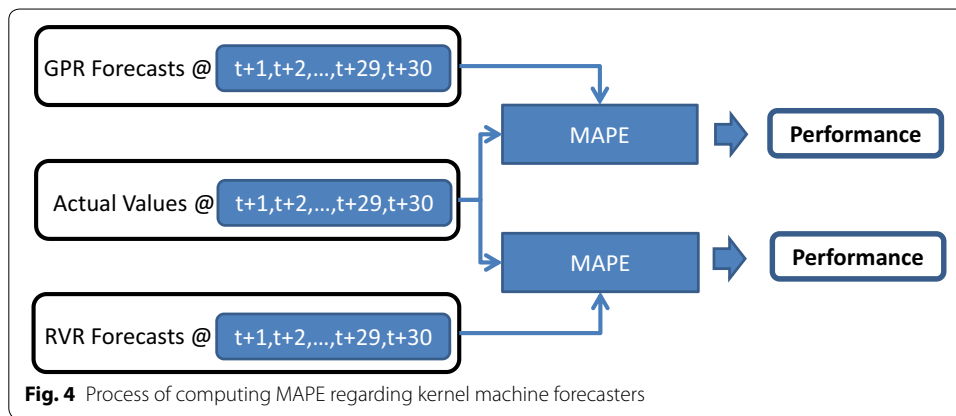
The kernel machine regression models have been applied to medium term load forecasting; the results are recorded and compared with respect to mean average percentage error (MAPE):

$$MAPE = \frac{100}{N} \sum_{t=1}^N \left| \frac{R_t - P_t}{R_t} \right| \quad (29)$$

with R_t, P_t being the real and predicted value at step t , and N is the number of timepoints considered in the prediction interval. In the current work, we obtain $N = 30$ as also indicated in Fig. 4, where the forecasting assessment procedure is depicted. Furthermore, the obtained results are compared to those taken with the statistical model of the autoregressive moving average(2,2), i.e., ARMA(2,2) (Huang and Shih 2003) with the ARMA coefficients to be determined by the Akaike Information Criterion (AIC) (Alamaniotis et al. 2012).

Test results

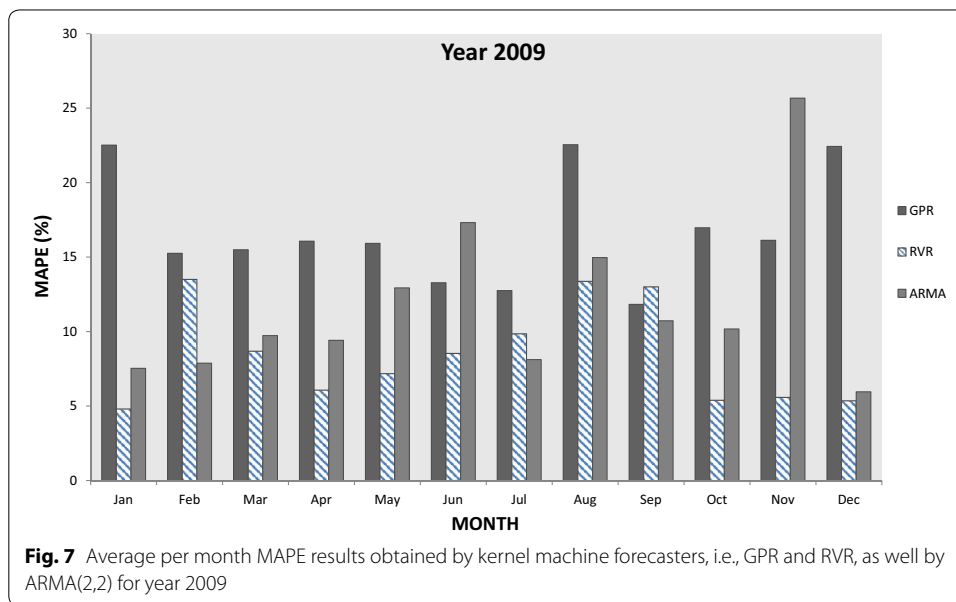
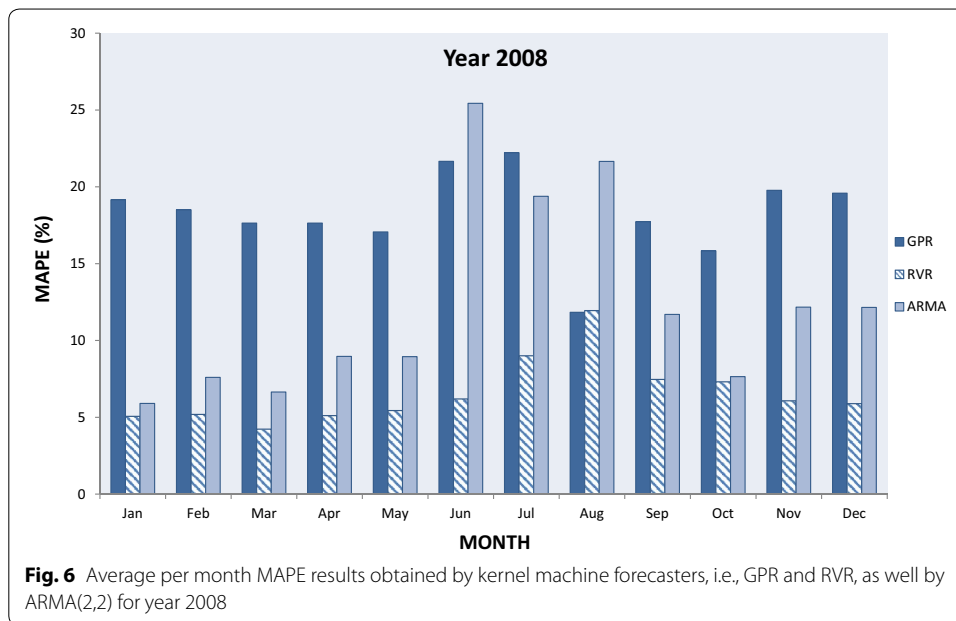
In this section, GPR and RVR have been applied in medium-term load forecasting and the results obtained for the designated time interval are plotted and compared to each other as well to results obtained with ARMA(2,2). In particular Figs. 5, 6, 7, 8 and 9 present the computed MAPE during the tested (almost) 5 year period for GPR, RVR and ARMA(2,2) forecasters. Results are depicted in terms of monthly intervals, giving 12



results for years 2007–2010 and 8 results for year 2011 (it was mentioned above that available tested data are from January 2007 to August 2011).

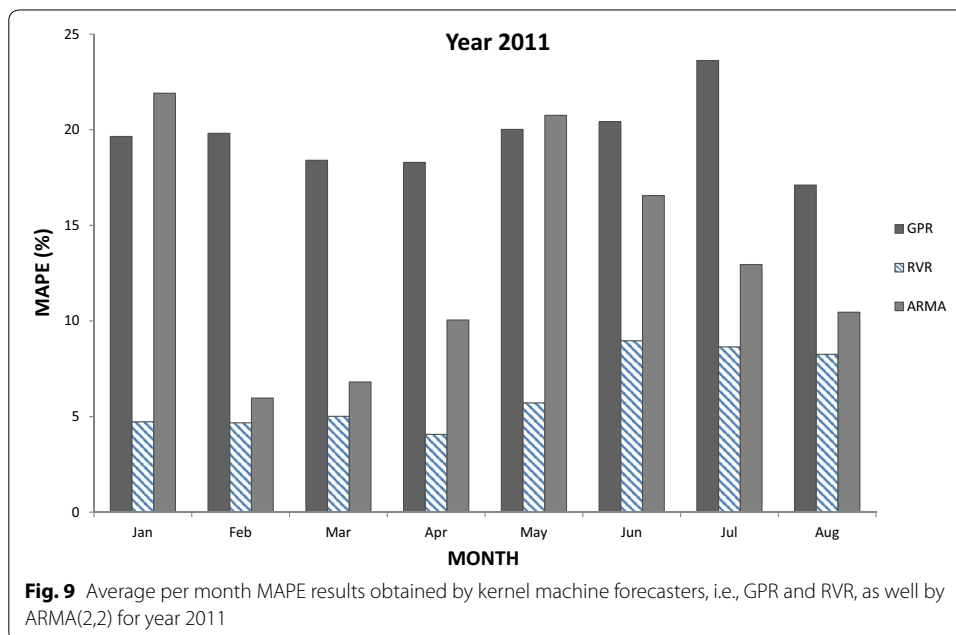
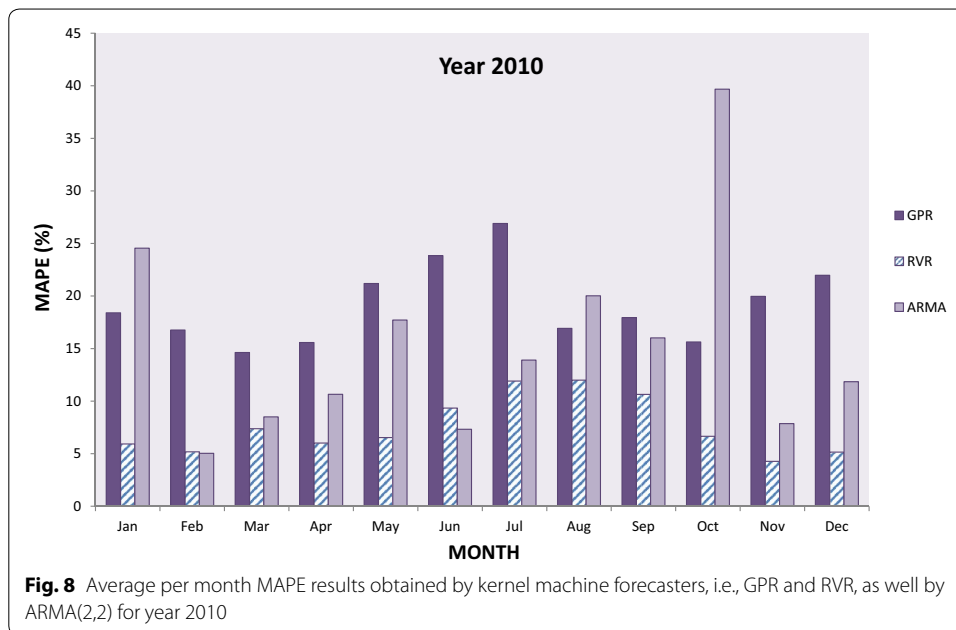
Figure 5 exhibits that RVR forecaster provides more accurate daily predictions for a month-ahead-horizon (i.e., 30-day ahead horizon) with respect to MAPE. In particular RVR gives the best performance for all months but November, where ARMA is the best forecaster. GPR gives the worst performance for all months in 2007 except for August. In 2008 data, Fig. 6 exhibits RVR as the best performing forecaster in all tested months except for August, where it is slightly outperformed by GPR. ARMA(2,2) performance is better than GPR and worse than RVR in the majority of the cases, with the exception of June and August 2008; for the latter months the ARMA forecasts are the least accurate among all forecasters.

In Fig. 7, we observe that RVR once more provides the best performance in the majority of the cases for year 2009—with the exception of February, July and September. For the same time interval (i.e., 2009), GPR provides the worst performance among three forecasters with a few exceptions. Furthermore, results for year 2010 presented in Fig. 8 drive to similar conclusions as earlier: RVR is the best forecaster in the majority of the



cases (in 10 out of 12), GPR the worst in most of them, while ARMA is the worst in two cases (January and October) and the best in other two (February and June). Additionally, in Fig. 9 provides the MAPE results for the first 8 months of year 2011: RVR clearly outperforms the other two forecasters in all cases, GPR provides the least accurate predictions in February, March, April, June, July and August, and ARMA is the least accurate for January and May.

In addition to monthly results, we present in Table 1 a yearly summary of the MAPE results obtained by each of the three forecasters. In particular, the average MAPE per tested year with respect to GPR, RVR and ARMA are given in the columns of Table 1.



Yearly averages exhibit that RVR is by far the most accurate forecaster for all tested years. The second most accurate is the ARMA model, with the GPR kernel machine to be the least accurate. For demonstration purposes, the forecasted demand by GPR and RVR are plotted against the actual demand for years 2007 and 2008 in Figs. 10 and 11 respectively. In both Figures we clearly observe that the RVR forecaster follows the actual demand closer than GPR.

In addition to MAPE criterion, we have also computed the squared correlation coefficient (R^2) between the predicted and the actual load values, despite the fact that R^2 is

Table 1 Average per year MAPE obtained by GPR, RVR and ARMA forecasters

Forecaster	MAPE (%)				
	Year 2007	Year 2008	Year 2009	Year 2010	Year 2011
GPR	20.0549	18.2197	16.7651	19.1449	19.6651
RVR	<i>8.2596</i>	<i>6.5793</i>	<i>8.4424</i>	<i>7.5811</i>	<i>6.2573</i>
ARMA(2,2)	12.2093	12.3496	11.6999	15.2576	13.1817

The lowest values are italicized

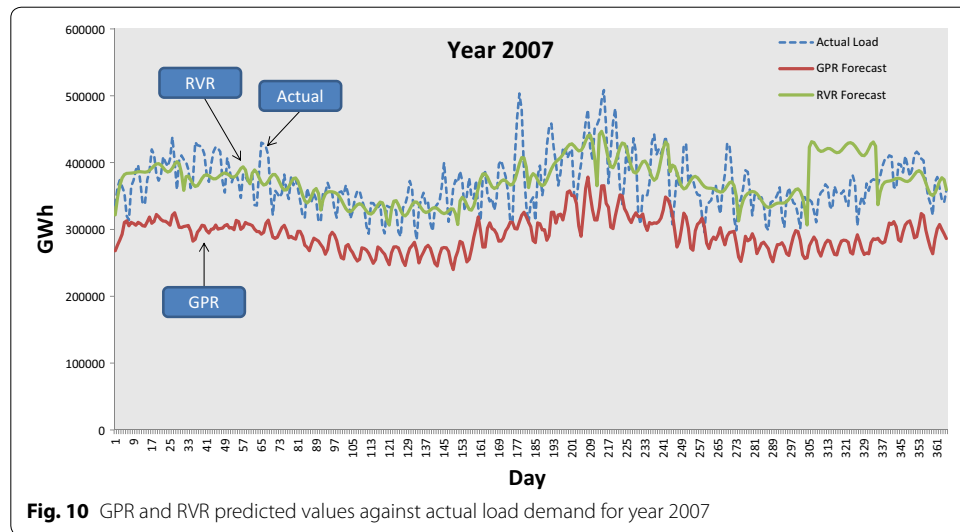


Fig. 10 GPR and RVR predicted values against actual load demand for year 2007

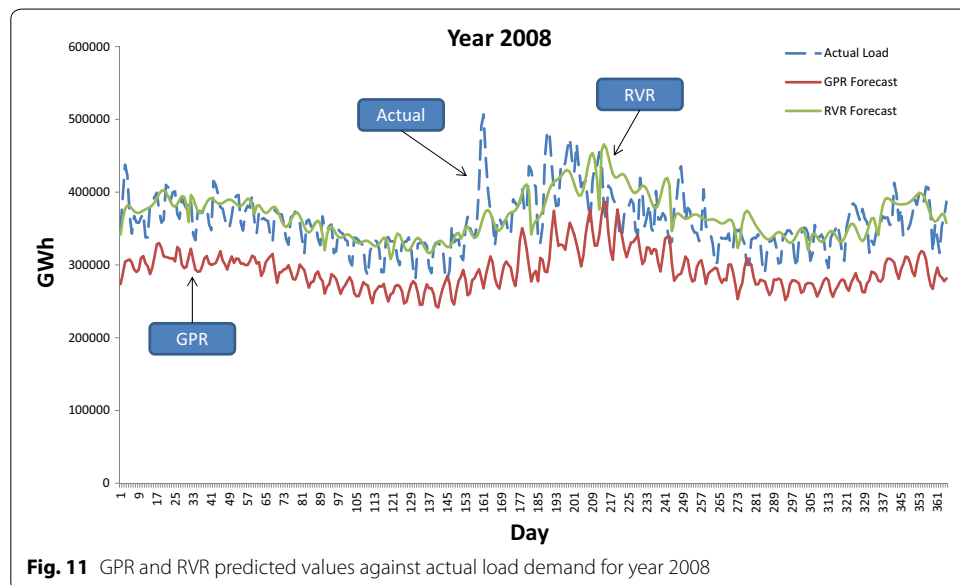


Fig. 11 GPR and RVR predicted values against actual load demand for year 2008

not very common criterion in load forecasting. This criterion does not express directly the performance of the forecaster but it shows how good a forecaster might be constructed from the predicted values. The obtained average per year R^2 for each of the

three forecasters is given in Table 2 where we observe that the ARMA provides the highest value for years 2007, 2008 and 2010, while RVR for 2009 and 2011. Overall combining observations from MAPE and R^2 from Tables 1 and 2, we may conclude that the ARMA captures the general trend of the load signal adequately in more cases than RVR but it is less accurate than RVR. In addition, it is slower in execution than both the kernel machines, with the GP to be the fastest. Figure 12 shows the average execution time of the models tested in this paper; models were run on an Intel i5 core laptop computer.

Therefore, we observe that depending on the selected model kernel machine may provide high accurate MLTF, as taken by RVR, or may provide low accuracy, as is the case with GPR.

Conclusion

The application of two types of kernel machines for medium-term load forecasting has been presented in this paper. The kernel machines studied are GPR and RVR whose performance is tested on actual historic data collected at the New England Area on a daily basis up to a month, with the tested time period being from January 2007 to August 2011. In addition, both forecasters are also compared to the ARMA(2,2) statistical tool that has been widely used in time series forecasting.

Obtained results show the superiority of RVR over the other two tested methods with respect to MAPE and R^2 . On a monthly comparison RVR provided the best accuracy in the majority of the cases while it is by far the best forecaster on a yearly based comparison. However, it should be emphasized that the kernel machines are equipped with

Table 2 Average per year squared correlation coefficient (R^2) obtained by GPR, RVR and ARMA forecasters

Forecaster	Squared correlation coefficient (R^2)				
	Year 2007	Year 2008	Year 2009	Year 2010	Year 2011
GPR	0.282	0.227	0.201	0.255	0.381
RVR	0.192	0.374	<i>0.286</i>	0.295	<i>0.438</i>
ARMA(2,2)	<i>0.444</i>	<i>0.533</i>	0.185	<i>0.320</i>	0.120

The best coefficients are italicized

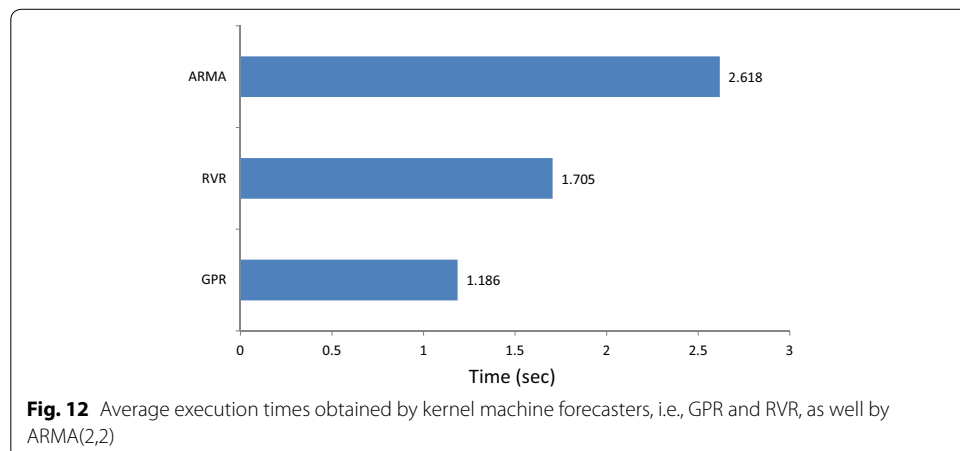


Fig. 12 Average execution times obtained by kernel machine forecasters, i.e., GPR and RVR, as well by ARMA(2,2)

a Gaussian kernel, which is the only kernel being tested in the current work; testing of other kernel functions is left for future work.

In addition, the promising method of core vector regression (Li and Liu 2010) will also be examined either as an independent forecaster or in combination with RVR and GP. Combination of kernel machines exhibits high potency for providing highly accurate medium term load predictions.

Authors' contributions

MA designed the study, developed the codes for the machine learning algorithms in Matlab, analyzed and interpreted the results and drafted the manuscript. DB designed and created the training and testing datasets, developed the statistical ARMA code in Matlab, and was involved in revising the manuscript for technical and intellectual content. LHT conceived the study and participated in its coordination and helped to draft the manuscript. All authors read and approved the final manuscript.

Author details

¹ Applied Intelligent Systems Laboratory, School of Nuclear Engineering, Purdue University, 400 Central Dr., West Lafayette, IN 47907, USA. ² Department of Electrical Engineering, Technological Institute of Stereas Elladas, 34400 Dimos Dirfion-Messapion, Psachna, Evia, Greece.

Acknowledgements

This work has been supported in part by the US National Science Foundation under Grant No. 1462393 and through the project "Hephaestus" under the auspices of "ARISTEIA" sponsored by the Hellenic General Secretariat for Research and Technology under the Action of Operational Program Education and Lifelong Learning co-funded by the European Social Fund and National Resources.

Competing interests

The authors declare that they have no competing interests.

Received: 1 April 2015 Accepted: 4 January 2016

Published online: 20 January 2016

References

- Abu-Shikhah N, Elkarmi F (2011) Medium-term electric load forecasting using singular value decomposition. *Energy* 36(7):4259–4271
- Abu-Shikhah N, Elkarmi F, Aloquili OM (2011) Medium-term load forecasting using multivariable linear and non-linear regression. *Smart Grid Renew Energy* 2:126–135
- Alamaniotis M, Gao R, Tsoukalas LH (2011) Towards an energy internet: a game-theoretic approach to price-directed energy utilization. In: *Energy-efficient computing and networking*, pp 3–11
- Alamaniotis M, Ikonomopoulos A, Tsoukalas LH (2011) A Pareto optimization approach of a Gaussian process ensemble for short-term load forecasting. In: *Proceedings of the 16th IEEE international conference on intelligent system application to power systems (ISAP)*, pp 1–6
- Alamaniotis M, Ikonomopoulos A, Tsoukalas LH (2012) Evolutionary multiobjective optimization of kernel-based very short-term load forecasting. *IEEE Trans Power Syst* 27(3):1477–1484
- Alamaniotis M, Chatzidakis S, Tsoukalas LH (2014) Monthly load forecasting using gaussian process regression. In: *Proceedings of the 9th Mediterranean conference on power generation, transmission, distribution, and energy conversion: MEDPOWER 2014*, pp 1–7
- Alamaniotis M, Tsoukalas LH, Bourbakis N (2014) Virtual cost approach: electricity consumption scheduling for smart grids/cities in price-directed electricity markets. In: *Proceedings of the 5th international conference on information, intelligence, systems and applications, IISA 2014*, pp 38–43
- Alamaniotis M, Bargiotas D, Bourbakis N, Tsoukalas LH (2015) Genetic optimal regression of relevance vector machines for electricity price forecasting in smart grids. *IEEE Trans Smart Grid* 6(6):2997–3005
- Bishop CM (2006) *Pattern recognition and machine learning*. Springer, New York
- Bozic M, Stojanovic M (2011) Application of SVM methods for mid-term load forecasting. *Serbian J Electr Eng* 8(1):73–83
- El Desouky AA, Elkateb MM (2000) Hybrid adaptive techniques for electric-load forecast using ANN and ARIMA. *IET Proc Gener Transm Distrib* 147(4):213–217
- Fainti R, Nasiakou A, Tsoukalas E, Vavalis M (2014) Design and early simulations of next generation intelligent energy systems. *Int J Monit Surveill Technol Res* 2(2):58–82
- Gao R, Wang X, Bougaev A, Schooley DC, Tsoukalas LH (2003) Short-term elasticities via intelligent tools for modern power systems. In: *Proceedings of 2003 medpower conference*, pp 1–6
- Gatsis N, Giannakis GB (2012) Residential load control: distributed scheduling and convergence with lost AMI messages. *IEEE Trans Smart Grid* 3(2):770–786
- Ghiassi MD, Zimbira DK, Saidane H (2006) Medium term system load forecasting with a dynamic artificial neural network model. *Electr Power Syst Res* 76(5):302–316
- Hoffman T, Scholkopf B, Smola AJ (2008) Kernel methods in machine learning. *Ann Stat* 33(3):1171–1220
- Huang SJ, Shih KR (2003) Short-term load forecasting via ARMA model identification including non-Gaussian process considerations. *IEEE Trans Power Syst* 18(2):673–679

- ISO New England Historical Data web page. <http://www.iso-ne.com>. Accessed in March 2015
- Kandil MS, El-Debeiky SM, Hasanien NE (2002) Long-term load forecasting for fast developing utility using a knowledge-based expert system. *IEEE Trans Power Syst* 17(2):491–496
- Kim KH, Park JK, Hwang KJ, Kim SH (1995) Implementation of hybrid short-term load forecasting system using artificial neural networks and fuzzy expert systems. *IEEE Trans Power Syst* 10(3):1534–1539
- Li Y, Liu, K (2010) Core vector regression with particle swarm optimization algorithm in short term load forecasting. In: *Proceedings of the IEEE second international conference on computer modeling and simulation (ICCMS'10)*, pp 325–329
- Mackay DJC (1998) Introduction to Gaussian processes. In: Bishop CM (ed) *Neural networks and machine learning*. Springer, Berlin
- Rasmussen CE (2006) *Gaussian processes for machine learning*. MIT Press, Cambridge
- Rengcun F, Jianzhong Z, Yongchuan Z, Qing-qing LI, Li LIU (2008) Application of particle swarm optimization based nonlinear grey Bernoulli model in medium-and long-term load forecasting. *Power Syst Technol (Beijing)* 32(12):60
- Scholkopf B, Smola AJ (2001) *Learning with kernels: support vector machines, regularization, optimization, and beyond*. MIT Press, Cambridge
- Tipping ME (2001) Sparse Bayesian learning and the relevance vector machine. *J Mach Learn Res* 1:211–244
- Tsekouras GJ, Hatzigiorgiou ND, Dialynas EN (2006) An optimized adaptive neural network for annual midterm energy forecasting. *IEEE Trans Power Syst* 21(1):385–391
- Tsekouras GJ, Dialynas EN, Hatzigiorgiou ND, Kavatza S (2007) A non-linear multivariable regression model for midterm energy forecasting of power systems. *Electr Power Syst Res* 77(12):1560–1568
- Tsoukalas LH, Gao R (2008) From smart grids to an energy internet: assumptions, architectures and requirements. In: *Proceedings of the third international conference on electric utility deregulation and restructuring and power technologies*, pp 94–98
- Tsoukalas LH, Uhrig R (1997) *Fuzzy and neural approaches in engineering*. Wiley Interscience, New York
- Williams CKI (2002) Gaussian processes. In: Arbib MA (ed) *Handbook of brain theory and neural networks*, 2nd edn. MIT Press, Cambridge, pp 466–470
- Xia C, Wang J, McMenemy K (2010) Short, medium and long term load forecasting model and virtual load forecaster based on radial basis function neural networks. *Int J Electr Power Energy Syst* 32(7):743–750

Submit your manuscript to a SpringerOpen[®] journal and benefit from:

- Convenient online submission
- Rigorous peer review
- Immediate publication on acceptance
- Open access: articles freely available online
- High visibility within the field
- Retaining the copyright to your article

Submit your next manuscript at ► springeropen.com
