

# The Effect of Driver Variables on the Estimation of Bivariate Probability Density of Peak Loads in Long-Term Horizon

Zohreh Kaheh, Morteza Shabanzadeh

Department of Power System Operation and Planning, Niroo Research Institute (NRI), Tehran, Iran

Correspondence: [mshabanzadeh@nri.ac.ir](mailto:mshabanzadeh@nri.ac.ir) ; [zohreh.kaheh@modares.ac.ir](mailto:zohreh.kaheh@modares.ac.ir)

## Abstract

It is evident that developing more accurate forecasting methods is the pillar of building robust multi-energy systems (MES). In this context, long-term forecasting is also indispensable to have a robust expansion planning program for modern power systems. While very short-term and short-term forecasting are usually represented with point estimation, this approach is highly unreliable in medium-term and long-term forecasting due to inherent uncertainty in predictors like weather variable in long terms. Accordingly, long-term forecasting is usually represented by probabilistic forecasting values which are based on probabilistic functions. In this paper, a self-organizing mixture network (SOMN) is developed to estimate the probability density function (PDF) of peak load in long-term horizons considering the most important drivers of seasonal similarity, population, gross domestic product (GDP), and electricity price. The proposed methodology is applied to forecast the PDF of annual and seasonal peak load in Queensland Australia.

**Keywords:** Long-term forecasting, Robust Multi-Energy Systems, Annual and seasonal peak load, Self-organizing mixture network, Probability density function, Driver variables.

## NOMENCLATURE

$i$	The index for pattern components
$c$	The index for winning node
$E$	Expected value
Var	Variance
$\mu$	Mean vector
$\sigma$	Standard deviation
$n$	Time step
$K$	The number of pattern components
$\rho$	Correlation coefficient
$\Sigma$	Covariance matrices
$p_i(x \theta_i)$	$i^{\text{th}}$ component-conditional density
$\theta_i$	Parameter vector for the $i^{\text{th}}$ conditional density for $i=1, 2, \dots, K$
$P_i$	Prior probability of the $i^{\text{th}}$ component
$\eta_c$	The neighborhood of the winner node in SOMN algorithm
$\alpha(n)$	Learning rate at the time step $n$
$[\cdot]^T$	Transpose function

design robust multi-energy systems, forecasting is of paramount importance; therefore, it is of significance to conduct novel and accurate forecasting methods in the multi-energy systems to arrange the operation mode of integrated energy system efficiently and economically [3].

Load forecasting as a dominant field of study in designing the multi-energy systems draws a lot of interest [2-5]. Conventional load forecasting approaches mainly concerned with only one type of loads, such as power loads, cooling loads, or heating loads. However, multi-energy load forecasting, as an ensemble forecasting approach considers the aggregated load, which has the intrinsic characteristics of the single load type, as well as the relevance among the series [4].

Long-term load forecasting is an indispensable tool for an effective planning of power systems. In long-term forecasting, inaccurate forecasts result in excessive investment, not fully utilized generating facilities, or insufficient generation and unfulfilled demand [6, 7]. Nevertheless, only few researchers have ever proposed new methods for long-term load forecasting in comparison with short-term forecasting [7].

## 1. Introduction

A new paradigm in the energy sector is MES, which captures the interactions among various energy carriers, e.g. electricity, heating, and cooling to improve the performance of the system [1, 2]. To

The current load forecasting literatures have mainly focused on point forecasting, in which the expected value of the future load is forecasted through different techniques. These forecasting techniques can be categorized as 1) statistical techniques, such as regression models, and time series models, 2) artificial intelligence techniques, such as neural networks and support vector machines, or 3) hybrid methods which are the combination of both statistical and artificial intelligence techniques. The point forecasting is mainly applied for very short-term and short-term forecasting, however, in medium-term and long-term forecasting, point forecast is not reliable since the inputs of forecasting models, which are mainly weather data, suffer from high uncertainty in long terms. Instead, probabilistic forecasting is applied for long-term forecasting where the possibility of having a demand is presented by a probabilistic value [7].

In spite of the importance of medium-term and long-term forecasting in operation and planning of power system, most of studies have focused on point forecasting in the short-term horizon, and few studies have been only conducted on probabilistic forecasting. Nonetheless, among these few studies on probabilistic load forecasting, most of them have focused on short-term forecasting. In [8], a review on probabilistic load forecasting is presented. Table 1 provides an overview of studies carried out in the literature of forecasting, taking into account inherent uncertainties in different contexts.

Table 1. Taxonomy of recent studies in probabilistic forecasting

Ref.	Forecasting methods	Context
[9-11]	Probabilistic time series	Medium-term load forecasting
[12]	Regression	Aesthetic quality assessments
[13]	Fuzzy regression	Short-term load forecasting
[14, 15]	Quantile regression	Probabilistic forecasting of solar power generation
[16]	Probabilistic SVR	General applications
[17]	Artificial intelligent methods	Very short-term load forecasting
[18]	Ensemble BMA package in R	Weather forecasting
[19, 20]	PDF estimation using machine learning techniques	Wind power ramp forecasting
[21]	Group method of data handling (GMDH)	Day-ahead electricity peak load interval forecasting

Fuzzy intervals are defined based on the covariance of data in different operating points, which are

characterized by linear regression models. In this context, a fuzzy regression method is presented in [12] to predict the aesthetic quality of a new product or service considering all uncertain objective drivers. In this method, a genetic programming is used to develop nonlinear structures of the models while model coefficients are determined by optimizing the fuzzy criteria. In short-term and medium-term load forecasting context, a fuzzy interaction regression is applied by [13] to forecast electric load in the short-term horizon with the help of fuzzy intervals. Moreover, a prediction interval construction model based on linear programming is presented in [14] to quantify the variability and uncertainty of the output of photovoltaic generating units for very short-term forecasting purposes (i.e., 5-min). This model is based on extreme learning machine and quantile regression. Apart from the considerations concerning the methods and applications of probabilistic forecasting, provided in Table 1, a fuzzy interval model, which is suitable for forecasting of electric demand and the output power of weather-dependent renewable energy sources that have limited dispatchability, is also presented in [22].

The authors of [23] present a practical methodology for probabilistic load forecasting based on a set of predictions, called *sister point forecasts*, generated from the same family of models. This approach performs the quantile regression on the average of sister point forecasts and generates prediction intervals of future electric loads. Reference [24] also presents a data-driven framework for probabilistic peak demand estimation using smart meter data of the consumers. This approach proposes four main steps including load modeling, customer grouping, maximum diversified demand estimation, and peak load estimation, to finally address both challenges of the unknown data of future loads and the influence of demand diversity among different customers. References [9, 10], among others, introduce a comprehensive class of time series models to precisely forecast the electric demand of industrial corporations. A simple procedure is also proposed to classify load profiles and present a probabilistic medium-term load forecasting tool for special types of industrial loads.

Among recent research works, the authors of [21] have presented a day-ahead electricity peak load interval forecasting that can easily convert an interval forecasting problem into a classification forecasting problem. The authors have applied a semi-supervised feature selection algorithm called group method of

data handling (GMDH) to address an electricity load classification forecasting issue. From a computational point of view, [17] has proposed a hybrid method for probabilistic load forecasting, including a generalized learning machine to train an improved wavelet neural network, and wavelet preprocessing as well as bootstrapping. This hybrid method provides a load forecasting with high reliability, accuracy and speed so that it would be more profitable for practical applications in the electricity market. However, as far as authors' knowledge concerned, this method has not been used for long-term forecasting purposes.

In a long-term context, [7] has presented a practical methodology for density forecast of the long-term peak electricity demand instead of common point-forecast approaches. The solution proposed by this methodology can hedge the financial risk accrued by uncertain demand. At first stage, the authors have used semi-parametric additive models to estimate the relationships between demand and the most influential driver variables such as temperature, calendar effects, and some economic variables. Then, they have forecasted the probability distribution of annual and weekly peak electricity demand up to ten years ahead by using a mixture of temperature simulation, future economic scenarios, and residual bootstrapping. This methodology captures the complex nonlinear effect of temperature and also other possible drivers such as calendar effects, price changes, and economic growth.

Another method which is recently proposed in this area is the probabilistic wind power ramp forecasting, which is presented in [19]. The authors have applied an ensemble machine learning technique to generate wind power scenarios and calculate the historical forecasting errors. Then, they used Gaussian mixture model to fit the probability distribution function (PDF) of forecasting errors. This method has not been used for demand forecasting purposes, although it is able to predict with a high level of accuracy.

In this paper, we develop the method proposed by [25] for long-term peak load forecasting considering different driver variables. In fact, we estimate the PDF of peak load in long-term horizons taking into account the most important drivers, like peak load in similar seasons in past years, peak load in the last season, population, and GDP. We apply a SOMN to estimate the PDF for the reason explained in [25] that much more accurate estimations with a rapid convergence could be obtained through this method. The results

show good forecasting capability of the proposed methodology at predicting the forecast PDF.

The paper is organized as follows. Section 2 presents the model and its concepts as well as the SOMN for estimating the PDF. The application and high performance of the proposed approach for a real case study are demonstrated in Section 3. Finally, the conclusions are drawn in Section 4.

## 2. Proposed method

### 2.1. The Concept of the bivariate distribution

Let the random variable  $Y$  denote the randomly selected peak load in a period of time, in MW. Then, suppose we are interested in determining the probability that  $Y$  would be between 9000 and 10000 MW, i.e.,  $P(9000 < Y < 10000)$ . It is clear that the peak load increases as the population or GDP increases. So, for the purpose of calculating the probability that  $Y$  is between 9000 and 10000 MW, we will find it more informative to first take into account a population or GDP value, say  $X$ . That is, we may want to find  $P(9000 < Y < 10000 | X = x)$ . To calculate such a conditional probability, we need to find the conditional distribution of  $Y$  given  $X = x$ . Based on three assumptions, we can easily find the conditional distribution of electric peak load ( $Y$ ) given electricity price, population, GDP or other drivers ( $x$ ). Required assumptions are stated below [26]:

- Peak load ( $Y$ ) follows a normal distribution (or easily transform to normal distribution [27]).
- $E(Y | x)$ , the conditional mean of  $Y$  given  $x$  is linear with respect to  $x$ .
- $Var(Y | x)$ , the conditional variance of  $Y$  given  $x$  is constant.

The first assumption is considered to facilitate using the proposed method, and is easily achievable through transforming from unknown distribution to normal distribution (Figure 4). The associated expected value and conditional variance for second and third assumptions are as follows, respectively [26].

$$\begin{aligned} E[Y|X = x] &= \mu_Y + \rho_{YX} \frac{x - \mu_X}{\sigma_X} \\ Var[Y|X = x] &= (1 - \rho^2)\sigma_Y^2 \end{aligned} \quad (1)$$

It should be mentioned that in machine-learning approaches (e.g., Bayesian method), it is common to select a prior distribution. Then, after observing data  $X_1, \dots, X_n$ , we can update our beliefs and calculate the posterior distribution  $f(\theta|X_1, \dots, X_n)$  [28].

In the next section, the multivariate PDF and conditional density for SOMN will be discussed.

### 2.1. Multivariate PDF and conditional density

In the case of a univariate normal distribution, the probability distribution or density function of variable  $y$  is represented as (1):

$$\varphi(y) = \frac{1}{\sqrt{2\pi}\sigma} \exp\left\{-\frac{1}{2\sigma^2}(y-\mu)^2\right\} \quad (2)$$

where  $y$  is the peak load (a random variable),  $\mu$  is the mean, and  $\sigma$  is the standard deviation.

In pattern estimation applications, each sample observation is assigned to a pattern component which has a prior probability. These situations are modeled by mixture distributions. The assumptions indicate that the conditional distribution of  $Y$  given  $X = x$  is:

$$Y | X = x \sim N\left(\mu_y + \rho(\sigma_y/\sigma_x)(X - \mu_x), \sigma_y^2(1 - \rho^2)\right) \quad (3)$$

where  $\rho$  is the correlation coefficient of  $X$  and  $Y$ .

Based on last three stated assumptions, we found the conditional distribution of  $Y$  given  $X = x$ . In this vein, the fourth assumption would be added;  $X$  follows a normal distribution for  $-\infty < x < \infty$ .

Based on the four stated assumptions, the joint probability density function of  $X$  and  $Y$  is defined as (4).

$$\begin{aligned} \varphi(x, y) &= \varphi(x) \times h(y|x) = \frac{1}{2\pi\sigma_x\sigma_y\sqrt{1-\rho^2}} \times \\ &\exp\left\{-\frac{1}{2(1-\rho^2)}\left[\frac{(x-\mu_x)^2}{\sigma_x^2} + \frac{(y-\mu_y)^2}{\sigma_y^2} - \frac{2\rho(x-\mu_x)(y-\mu_y)}{\sigma_x\sigma_y}\right]\right\} \end{aligned} \quad (4)$$

This joint PDF is called the *bivariate normal distribution*. In fact, the bivariate distribution represents the joint distribution of two random variables [29]. The two random variables  $X$  and  $Y$  are related to each other in the sense that they are not independent on each other. This dependency is

reflected by the correlation  $\rho$  between the two variables  $X$  and  $Y$ .

### 2.2. Self-organizing mixture network

According to [25], self-organizing mixture network (SOMN) is a powerful unsupervised learning method. This network contains two layers of nodes, including an input layer and an output layer. In the input layer, there is a weight vector and a position related to each node. The objective of SOMN is to maximize the degree of similarity of patterns within a cluster, as well as to minimize the similarity of patterns belonging to different clusters. In addition, SOMN transforms high dimensional input patterns into the responses of two-dimensional arrays of neurons, thus it can facilitate the detection of the innate structure and the interrelationship of data [30-33].

The learning process of SOMN is summarized as follows:

**Step 1.** Initialize random values for the weights associated with the input pattern.

**Step 2.** Find the winning node as one whose weights are very similar to the input vector considering the minimum distance Euclidean criterion.

**Step 3.** Update the weights of the winner and its neighborhood neurons in such a way that by strengthening them, this area would be more likely to fire up when a similar input pattern is presented next time. The significance of the strengthening decreases with the distance from the winner.

**Step 4.** The process of weight updating will be performed for a specified number of iterations. If the map is not unfolded, the algorithm must restart the training process with a different set of initial weights.

### 2.3. SOMN Structure for PDF estimation

The SOM structure for PDF estimation problem is illustrated in Fig. 1, where  $\{\mu_i, \Sigma_i\}$  are the mean vector and covariance matrix of the  $i^{th}$  value of assumed normal density function, respectively. Also,  $\eta_c$  is a neighborhood of the winner whose weight must be updated. According to previous section, given  $\theta_i = \{\theta_{i1}, \theta_{i2}\} = \{\mu_i, \Sigma_i\}$ , the conditional probability density of data sample is derived by (5), where  $p_i(y|\theta_i)$  is the  $i^{th}$  component-conditional density and  $P_i$  is the prior probability of the  $i^{th}$  component.

$$p(y|\theta) = \sum_{i=1}^k p_i(y|\theta_i) \cdot P_i \quad (5)$$

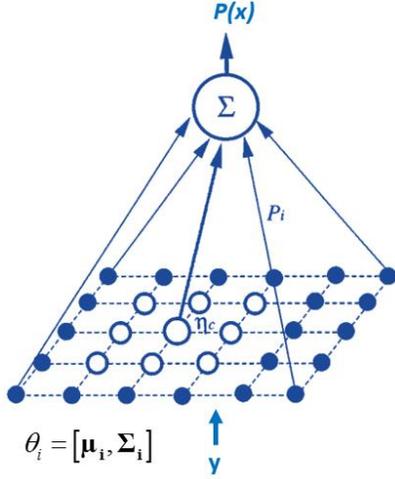


Fig. 1. The structure of SOMN [25]

Considering a limited number of conditions, the SOM network places  $M$  nodes in the input space. The parameters vector  $\theta_i$  includes mean vectors and covariance matrices related to the assumed bivariate normal density function which are considered as learning weights. At each iteration, a sample point is randomly taken from input space i.e., a finite data set. A winner is chosen according to its output multiplied by its estimated posterior probability [25].

$$\hat{p}(\hat{\theta}_i | y) = \frac{\hat{p}_i(y | \hat{\theta}_i) \times \hat{P}_i}{\hat{p}(y | \hat{\theta})} \quad (6)$$

The number of nodes should be equal to or greater than the number of conditions to avoid the under-represented problem [25]. The *Kullback–Leibler* information metric (7), also called *relative entropy* [34], measures the divergence between  $p(x)$  and  $\hat{p}(x)$ . In (7), the density function of the actual data and the estimated one are indicated by  $p(x)$  and  $\hat{p}(x)$ , respectively.

$$I = - \int \log \frac{\hat{p}(x)}{p(x)} \times p(x) dx \quad (7)$$

The optimal estimate of parameters in mixture distribution could be calculated by minimizing their partial differentials in respect to each model parameter by *Lagrangian method* considering the constraint  $\sum_{i=1}^k \hat{p}_i = 1$ . Also, according to [25], if the actual distribution function is unknown, the *Robbins–Monro* stochastic approximation method can be used instead of direct Lagrangian method. The parameters updating can be limited to a small neighborhood of the winning

node, which has the largest posterior probability. Therefore, the density can be approximated by a mixture of a small number of nodes at one time:

$$\hat{p}(x | \theta) \approx \sum_{i \in \eta_c} \hat{p}_i(x | \hat{\theta}_i) \hat{P}_i \quad (8)$$

The learning rules for updating the mean vector and covariance matrix in the SOM algorithms are as follow:

$$\Delta \hat{\mu}_i = \alpha(n) \hat{P}(i | x) [x(n) - \hat{\mu}_i(n)] \quad (9)$$

$$\Delta \hat{\Sigma}_i = \alpha(n) \hat{P}(i | x) \times \left\{ [x(n) - \hat{\mu}_i(n)] [x(n) - \hat{\mu}_i(n)]^T - \hat{\Sigma}_i(n) \right\} \quad (10)$$

A large neighborhood at the beginning of learning process means a large variance of the Gaussians as well as a high mobility for the neurons. This would be helpful to find the global optimum, or at least to result in a better local optimum, especially at the beginning of the learning. In contrast, small neighborhood sizes mean small variances for the Gaussians as well as a low mobility. As the learning progresses, the neighborhood during the process is shrink to provide an adjustment to the variance of the Gaussians [25].

### 3. Results

In this section, the proposed approach is applied on a real data in Queensland, Australia to derive long-term probabilistic forecasting. Half-hourly demand data during 2001 to 2016 were obtained from Australian Energy Market Operator (AEMO) [35]. The studied case is examined from two points of view including annual peak load and seasonal peak load. The yearly peak load data between 2001 and 2016 is illustrated in Fig. 2. The seasonal peak load data between 2007 and 2016 is also depicted in Fig. 3. It should be noted that a part of seasonal peak load data is ignored to avoid unnecessary historical data. The rest of seasonal data are presented in Table 2.

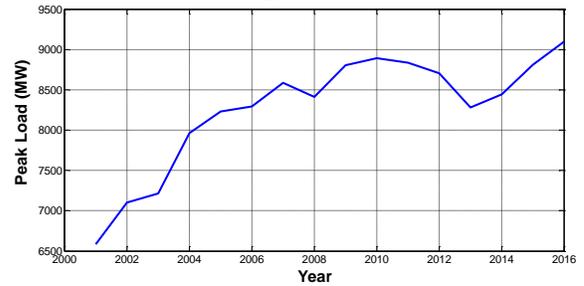


Fig. 2. Annual peak load for Queensland during 2001 to 2016

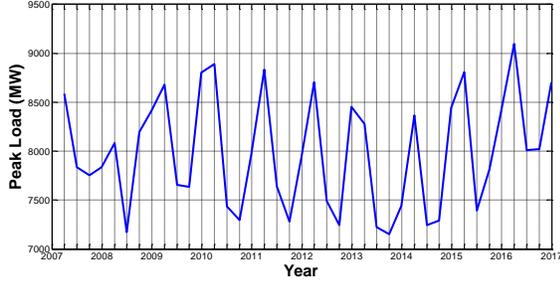


Fig. 3. Quarterly peak load for Queensland during 2007 to 2016

It is worth mentioning that the SOMN algorithm is implemented in MATLAB<sup>®</sup> and executed on a Windows-based PC with a Core™ i5 processor clocking at 3.2 GHz and 4 GB of RAM. In addition, all simulations for comparison and statistical tests are implemented in R-Studio<sup>®</sup> 3.4.2.

To derive long-term probabilistic forecasting, the univariate density estimation for two cases are studied separately for both training and learning purposes. Although the initial data do not show a normal distribution, they can easily transform to the normal distribution. The histogram graph of the initial and normalized annual and seasonal peak loads along with normal density function curve is illustrated in Figs. 6.

The seasonal and annual peak loads are subjected to different social, economic, and calendar drivers, such as population growth, changing technology, changing the economic condition, and so on [7].

The values of the *Pearson* correlation between the seasonal peak load and some influential drivers are provided in Table 3. As seen, the highest correlation is between the peak load and that of the similar season in the last year.

Table 3. Correlations between seasonal peak load and influential drivers

Variables	Correlation
Peak load & Peak load in similar season in the last year	0.716689
Peak load & Peak load in the last season	0.088361
Peak load & GDP	0.070420
Peak load & Price	0.1734
Peak load & Population	0.085395

Besides, we conduct a *Principal Components Analysis* (PCA) for different driver variables considering data provided in [36]. PCA aims to maximize the variance of a linear combination of the variables, and forms new variables which are linear composites of the original variables, and the new variables are uncorrelated among themselves [37, 38]. The results for data provided in [36] are illustrated in Figs 4 and 5. Fig 4 illustrates the coefficients of each variable in principal components. In Fig 5, the first two

eigenvalues form a steep curve as a bend at the beginning and then a straight-line trend with shallow slope. Accordingly, we need to keep those eigenvalues in the steep curve before the first one on the straight line, hereupon, two components can be retained as follows:

$$(11)$$

$$PC_1 : 0.3963789 \times GDP + 0.3976349 \times POP + 0.3951176 \times EP + 0.3898217 \times (GDP / POP) + 0.3933803 \times System\_Losses + 0.2821188 \times Load\_Factor + 0 \times Energy\_Cost$$

$$(12)$$

$$PC_2 : -0.073442070 \times GDP - 0.008094510 \times POP - 0.003872796 \times EP - 0.230324773 \times (GDP / POP) - 0.160632877 \times System\_Losses + 0.934934040 \times Load\_Factor + 0 \times Energy\_Cost$$

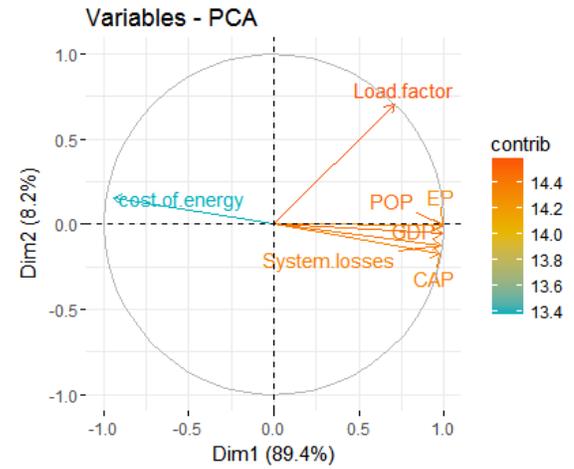


Fig 4. The share of initial driver variables in principal components for data provided in [36]

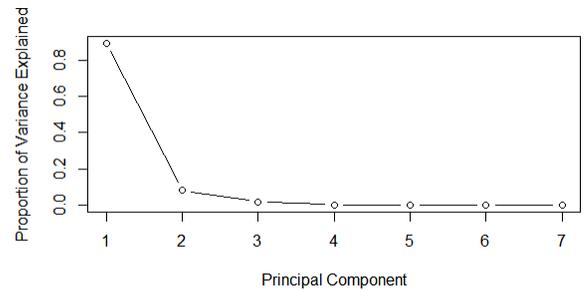
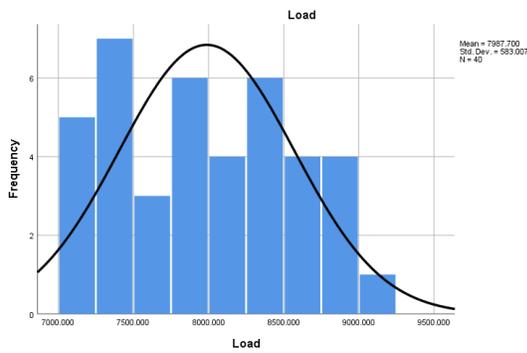


Fig 5. Scree graph for eigenvalues of data provided in [36]

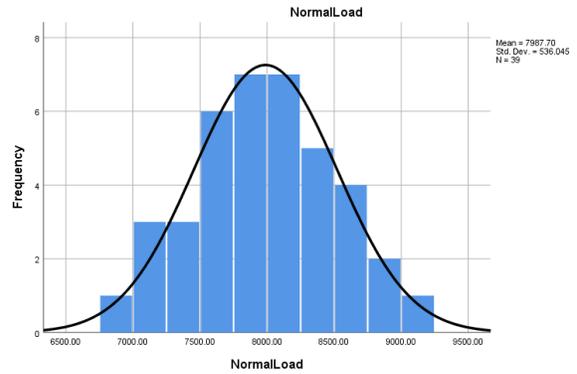
However, due to the lack of data for a wide array of variables in case of Australia, we inevitably avoided conducting PCA and only relied on available data.

**Table 2.** Seasonal data

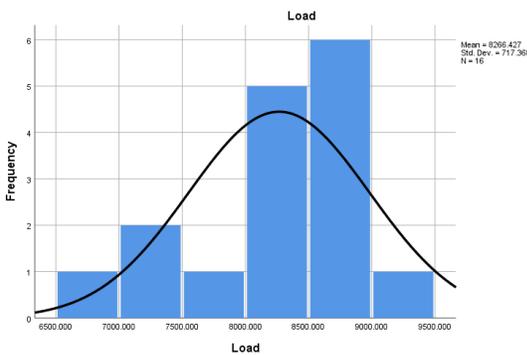
Year	Peak load	The same season in pre-year	Pre-season	GDP (M\$)	POP* 1000	Year	Peak load	The same season in pre-year	Pre-season	GDP (M\$)	POP* 1000
2008	8081.59	8588.69	7834.56	298005	21244	2012	7244.49	7281.96	7490.08	375212	22723
2008	7173.59	7837.02	8081.59	305846	21336	2012	8453.11	7943.88	7244.49	378399	22809
2008	8197.39	7753.15	7173.59	316532	21431	2013	8278.4	8706.76	8453.11	382165	22895
2008	8412.93	7834.56	8197.39	317375	21530	2013	7221.22	7490.08	8278.4	386665	22981
2009	8677.16	8081.59	8412.93	316712	21628	2013	7149.94	7244.49	7221.22	388898	23067
2009	7655.07	7173.59	8677.16	310452	21727	2013	7442.05	8453.11	7149.94	390076	23273
2009	7637.58	8197.39	7655.07	312363	21812	2014	8364.63	8278.4	7442.05	393896	23369
2009	8804.08	8412.93	7637.58	320164	21883	2014	7243.07	7221.22	8364.63	397788	23466
2010	8890.66	8677.16	8804.08	327262	21955	2014	7288.08	7149.94	7243.07	401675	23563
2010	7432.57	7655.07	8890.66	338507	22027	2014	8445.31	7442.05	7288.08	406538	23661
2010	7293.49	7637.58	7432.57	343252	22095	2015	8808.7	8364.63	8445.31	411571	23759
2010	7957.31	8804.08	7293.49	349162	22159	2015	7392.29	7243.07	8808.7	416604	23856
2011	8836.41	8890.66	7957.31	352805	22224	2015	7815.98	7288.08	7392.29	422172	23954
2011	7640.42	7432.57	8836.41	361382	22289	2015	8402.56	8445.31	7815.98	428304	24051
2011	7281.96	7293.49	7640.42	368345	22366	2016	9097.04	8808.7	8402.56	434799	24148
2011	7943.88	7957.31	7281.96	369812	22456	2016	8011.49	7392.29	9097.04	441460	24245
2012	8706.76	8836.41	7943.88	371867	22546	2016	8020.1	7815.98	8011.49	448206	24341
2012	7490.08	7640.42	8706.76	375808	22635	2016	8703.17	8402.56	8020.1	454916	24437



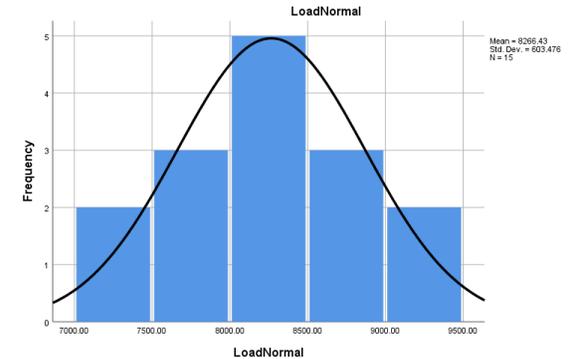
(a) Initial seasonal peak load



(b) Normalized seasonal peak load



(c) Initial annual peak load



(d) Normalized annual peak load

**Figure 6.** Initial and normalized peak load

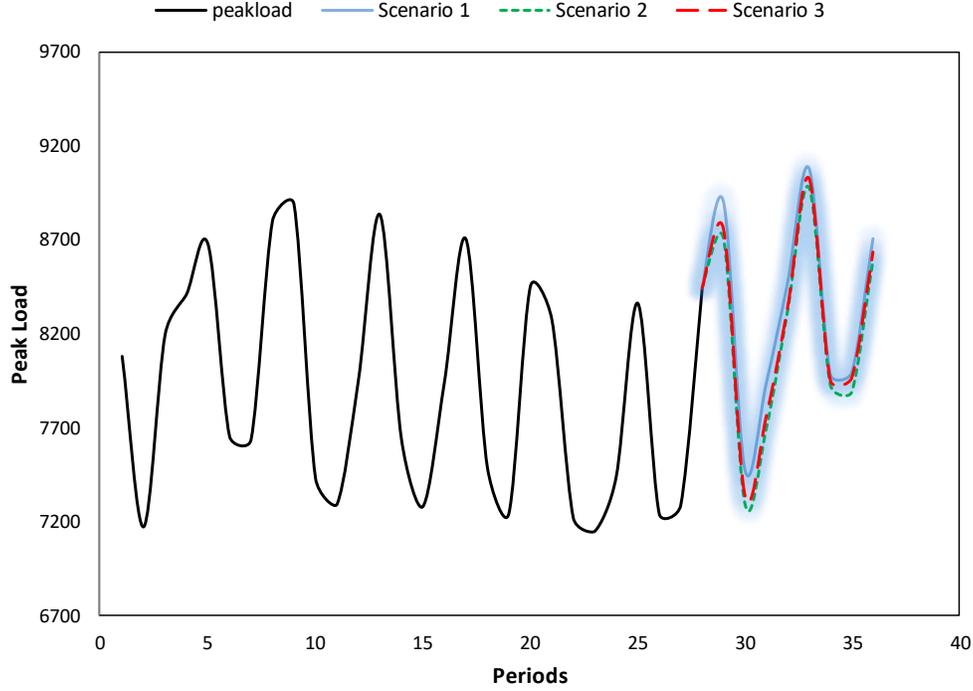


Figure 7. Three possible scenarios for seasonal peak load forecasting

Parameters of the component densities including Mean vectors and variance-covariance matrices, as well as prior probabilities are the learning weights. Hence, the initial mean vectors of a four-node Gaussian SOMN are set to small random vectors around the mean of standard normal distribution [0,0]. Besides, the initial variance-covariance matrices are defined as matrices equal to the initial sample variance plus a random coefficient (a random value between 1 and 30) of the initial sample variance, and the initial probabilities are equally set to 1/3.

At each iteration, one data point was randomly taken from the 120-point training set. The learning rates for means and for variances and mixing priors were decreasing from 0.5 and 0.05, respectively.

Three possible scenarios for seasonal peak load forecasting in the most probable range of values are illustrated in Fig. 7. To analyze the fitting performance, the error metric of root-mean-square error (RMSE) in (13) is applied.

$$\text{RMSE} = \sqrt{\frac{\sum_{i=1}^n (\hat{y}_i - y_i)^2}{n}} \quad (13)$$

RMSE is more preferable in comparison with other measures like Mean Absolute Error (MAE) and Mean Absolute Percentage Error (MAPE). For example, MAPE is a poor-accuracy indicator although it is a quite well-known measure among business managers. With reference to its mathematical formulation, MAPE divides each error individually by the demand, so it is clearly skewed. It means that high errors during low-demand periods will have a significant impact on MAPE. For this reason, optimizing MAPE will result in a strange forecast that will most likely some undershoots may be seen in the demand profile poor-accuracy indicator [39]. On the other hand, compared to MAE and MAPE, the indicator RMSE is more accurate and it does not treat each error as the same. It gives more importance (i.e., weight) on the most significant errors which means that one big error is enough to get a very bad RMSE. Thus, taking the square root of the average squared errors might have some interesting implications for RMSE because the errors are squared before they are averaged and thus, the RMSE gives a relatively high weight to large errors. This proves that the RMSE could be more useful when large errors are particularly undesirable [40].

To evaluate the average of the obtained RMSEs, the proposed algorithm is carried out ten times. The averages of RMSEs for these ten times considering each of driver variables as the second variable for seasonal and annual peak load are illustrated in Tables 4 and 5, respectively. The driver variables for seasonal peak load are population, GDP, peak load in last season, and peak load in similar season in the last year. However, the driver variables for annual peak load are population, GDP. Furthermore, to evaluate the fitted PDFs, the RMSEs of our proposed PDF estimation method are compared with those of the non-central multivariate ‘t’ distribution in the “*mvtnorm*” package of R.

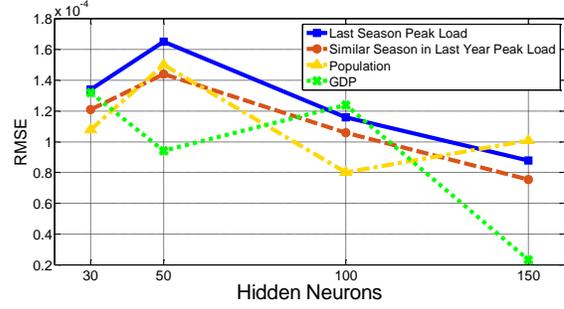


Fig. 8. A comparison between the calculated RMSEs for different numbers of hidden neurons considering different drivers for seasonal peak load

Table 4. The RMSEs of our proposed approach and those of the non-central multivariate ‘t’ distribution for seasonal data

Seasonal peak load	RMSE				
	SOMN				non-central multivariate ‘t’ distribution
	30	50	100	150	
Hidden neurons					
Last season peak load	0.000134	0.000164	0.000116	0.000088	0.000193
Peak load in similar season in the last year	0.000121	0.000144	0.000106	0.000075	0.000177
Population	0.000108	0.000149	0.000081	0.000101	0.000211
GDP	0.000131	0.000094	0.000123	0.000023	0.000189

Table 5. RMSEs of our proposed approach and those of the non-central multivariate ‘t’ distribution for annual data

Annual peak load	RMSE				
	SOMN				non-central multivariate ‘t’ distribution
	30	50	100	150	
Hidden neurons					
Electricity price	0.000116	0.000105	0.000103	0.000107	0.000213
Population	0.000108	0.000109	0.000029	0.000029	0.000173
GDP	0.000129	0.000108	0.000171	0.000019	0.000182

According to Table 4 for PDF estimation of the seasonal peak load, considering the driver variables lead to a lower RMSE. Accordingly, the lowest RMSE is obtained considering GDP as driver variable in the case with 150 hidden neurons. However, for some number of the hidden neurons, other driver variables lead to a lower RMSE. For instance, average RMSE associated with the PDF estimation of the seasonal peak load for 30 and 100 hidden neurons are decreased considering the population as driver variable. Table 4 for PDF estimation of the annual peak load provides similar results.

The key feature of our proposed methodology is providing a full density forecast for the peak demand with quantifiable probabilistic uncertainty, which captures the complex nonlinear effect of possible drivers. The RMSE results illustrate that the proposed method performs well on the historical data.

In light of the results presented in Tables 4 and 5 and the RMSE values in Fig. 8, we can conclude that the proposed method for PDF estimation is more

effective than the commonly used method in “*mvtnorm*” package. In addition, if we consider GDP as the second variable (see Fig. 8 and the last row in Tables 4 and 5), the result will lead to the lowest RMSE.

Nevertheless, according to the overall pattern of RMSEs, there is no robust evidence to define a relationship between “correlation between the dependent variable and each of driver variables” and “RMSE”; this constitutes the ground for future research work.

### 3. Discussion

In this paper, as a preliminary study, we aimed to find an appropriate forecasting approach for multi-energy systems. In this research, we focused on the importance of our proposed method on multi-energy systems. According to [1], the aim of multi-energy systems is considering the interaction among electricity, heat, cooling, fuels, transport at various

levels to improve technical, economical, and environmental performance at the operational and planning stage in comparison with classical energy systems whose sectors are treated separately [1].

Multi-energy systems have two aspects; first, these systems as integrated energy systems are known as robust systems due to their ability to stand various types of disturbances by increasing the system responsibility and decreasing the system volatility through providing various alternatives.

On the other hand, the main issues in integrated energy systems are the uncertainty and scalability [41]. To practically implement multi-energy systems, their uncertainty parameters and their uncertainty sets should be first defined. For instance, according to the robust design methodology proposed by [41, 42], the noise factors beyond the control of the designer should be considered in the multi-energy system design.

Therefore, it stands to reason that considering different driver variables via a comprehensive forecasting method, which deals with uncertainties in multi-energy systems, is of paramount importance. However, due to the inherent randomness of the underlying energy resource (e.g., wind speed, solar radiation) alongside economic and social impacts, there will be definitely a high uncertainty associated with the load forecasts especially over the long term.

In addition, to cope with the bottleneck of performance improvement, a practical methodology for density forecast of the long-term peak electricity demand instead of common point-forecast approaches is highly needed. Applying such approaches can hedge the financial risk imposed by uncertain demands. Such approaches also capture the complex nonlinear effect of different possible drivers. Besides, the applied method of PDF estimation is also necessary. For example, here, we have applied a SOMN algorithm to estimate the PDF, which produces accurate estimations with rapid convergence.

#### 4. Conclusion and further research

In this paper, an unsupervised learning method called SOMN was proposed for estimating the bivariate density functions of the annual and seasonal peak load. The major contribution of this paper is presenting a novel systematic methodology for forecasting the density of long-term peak electricity demand in multi-energy system. Using the measure RMSE, the performance of the proposed method was compared with the non-central multivariate  $t$  distribution. The simulation results demonstrated that

the proposed method outperforms the non-central multivariate  $t$  distribution.

According to the values of RMSE, it can be inferred that a high correlation between two variables does not necessarily lead to a low RMSE. In other words, there is no robust evidence to define a relationship between these concepts.

The results show that making a relationship between the “correlation between dependent variable and driver variables” and “RMSE” in bivariate probability density function is still needs further research. Furthermore, the method proposed in this paper would be developed from several aspects. The most important one is the improvement of the proposed algorithm through introducing ensemble method by combining several artificial intelligent algorithms.

Multi-energy systems (MES)  
Self-Organizing Mixture Network (SOMN)  
Probability Density Function (PDF)  
Gross Domestic Product (GDP)  
Group Method of Data Handling (GMDH)  
Root-Mean-Square Error (RMSE)  
Principal Components Analysis (PCA)

#### Declarations

- Availability of data and materials

The datasets analyzed during the current study are available from the corresponding author on request.

- Competing interests (Not applicable)
- Funding

This work was supported by *Niroo Research Institute* (NRI) under Contract No. PONPN06.

- Authors' contributions
  - **Zohreh Kaheh:**
    - Developed the theory
    - Collected the data
    - Performed the numerical simulations
    - Discussed the results
    - Wrote the manuscript
  - **Morteza Shabanzadeh:**
    - Supervised the project
    - Developed the theory
    - Collected the data
    - Performed the numerical simulations
    - Discussed the results and commented on the manuscript
    - Contributed to the final version of the manuscript

## References

- [1] P. Mancarella, "MES (multi-energy systems): An overview of concepts and evaluation models," *Energy*, vol. 65, pp. 1-17, 2014.
- [2] P. Gabrielli, M. Gazzani, E. Martelli, and M. Mazzotti, "Optimal design of multi-energy systems with seasonal storage," *Applied Energy*, vol. 219, pp. 408-424, 2018.
- [3] Z. Tieyan, L. Hening, H. Qian, K. Xuan, G. Shengyu, Y. Xiaochen and H. Huan (2019, April). Integrated Load Forecasting Model of Multi-Energy System Based on Markov Chain Improved Neural Network. In *2019 11th International Conference on Measuring Technology and Mechatronics Automation (ICMTMA)* (pp. 454-457). IEEE.
- [4] S. Wang, S. Wang, H. Chen, and Q. Gu, "Multi-energy load forecasting for regional integrated energy systems considering temporal dynamic and coupling characteristics," *Energy*, vol. 195, p. 116964, 2020.
- [5] P. Gabrielli, F. Furer, P. Murray, K. Orehounig, J. Carmeliet, M. Gazzani, and M. Mazzotti (2018). A Time-series-based approach for robust design of multi-energy systems with energy storage. In *Computer Aided Chemical Engineering* (Vol. 43, pp. 525-530). Elsevier.
- [6] S. Avdakovic, E. Becirovic, N. Hasanspahic, M. Music, A. Merzic, A. Tuhcc, and A. K. Loncареvic (2015). Long-term forecasting of energy, electricity and active power demand–Bosnia and Herzegovina case study. *Balkan Journal of Electrical and Computer Engineering*, 3(1), 11-16.
- [7] R. J. Hyndman and S. Fan, "Density forecasting for long-term peak electricity demand," *IEEE Transactions on Power Systems*, vol. 25, no. 2, pp. 1142-1153, 2010.
- [8] T. Hong and S. Fan, "Probabilistic electric load forecasting: A tutorial review," *International Journal of Forecasting*, vol. 32, no. 3, pp. 914-938, 2016.
- [9] K. Berk, "Probabilistic Forecasting of Electricity Load for Industrial Enterprises," Siegen, 2016.
- [10] K. Berk and A. Muller, "Probabilistic forecasting of medium-term electricity demand: a comparison of time series models," 2016.
- [11] H. Sangrody, N. Zhou, and X. Qiao, "Probabilistic models for daily peak loads at distribution feeders," in *2017 IEEE Power & Energy Society General Meeting*, 2017, pp. 1-5: IEEE.
- [12] K. Y. Chan, H.-K. Lam, C. K. F. Yiu, and T. S. Dillon, "A Flexible Fuzzy Regression Method for Addressing Nonlinear Uncertainty on Aesthetic Quality Assessments," *IEEE Transactions on Systems, Man, and Cybernetics: Systems*, vol. 47, no. 8, pp. 2363-2377, 2017.
- [13] T. Hong and P. Wang, "Fuzzy interaction regression for short term load forecasting," *Fuzzy optimization and decision making*, vol. 13, no. 1, pp. 91-103, 2014.
- [14] C. Wan, J. Lin, Y. Song, Z. Xu, and G. Yang, "Probabilistic forecasting of photovoltaic generation: An efficient statistical approach," *IEEE Transactions on Power Systems*, vol. 32, no. 3, pp. 2471-2472, 2017.
- [15] H. Sangrody and N. Zhou, "An initial study on load forecasting considering economic factors," in *2016 IEEE Power and Energy Society General Meeting (PESGM)*, 2016, pp. 1-5: IEEE.
- [16] C.-J. Lin and R.-C. Weng, "Simple probabilistic predictions for support vector regression," 2004.
- [17] T. N. Mehdi Rafiei, Jamshid Aghaei, Miadreza Shafie-khah, Joao P. S. Catalao, "Probabilistic Load Forecasting using an Improved Wavelet Neural Network Trained by Generalized Extreme Learning Machine," in *IEEE Transactions on Smart Grid*, ed, 2018.
- [18] C. Fraley, A. Raftery, T. Gneiting, M. Sloughter, and V. Berrocal, "Probabilistic Weather Forecasting in R," *R Journal*, vol. 3, no. 1, 2011.
- [19] M. Cui, C. Feng, Z. Wang, J. Zhang, Q. Wang, A. Florita, V. Krishnan and B.M. Hodge 2017, July. Probabilistic wind power ramp forecasting based on a scenario generation method. In *2017 IEEE Power & Energy Society General Meeting* (pp. 1-1). IEEE.
- [20] B. Khorramdel, H. Khorramdel, A. Zare, N. Safari, H. Sangrody, and C. Chung, "A nonparametric probability distribution model for short-term wind power prediction error," in *2018 IEEE Canadian Conference on Electrical & Computer Engineering (CCECE)*, 2018, pp. 1-5: IEEE.
- [21] L. Yang, H. Yang, H. Yang, and H. Liu, "GMDH-Based Semi-Supervised Feature Selection for Electricity Load Classification Forecasting," *Sustainability*, vol. 10, no. 1, p. 217, 2018.
- [22] D. Saez, F. Avila, D. Olivares, C. Canizares, and L. Marın, "Fuzzy prediction interval models for forecasting renewable resources and

- loads in microgrids," *IEEE Transactions on Smart Grid*, vol. 6, no. 2, pp. 548-556, 2015.
- [23] B. Liu, J. Nowotarski, T. Hong, and R. Weron, "Probabilistic load forecasting via quantile regression averaging on sister forecasts," *IEEE Transactions on Smart Grid*, vol. 8, no. 2, pp. 730-737, 2017.
- [24] M. Sun, Y. Wang, G. Strbac, and C. Kang, "Probabilistic Peak Load Estimation in Smart Cities Using Smart Meter Data," Institute of Electrical and Electronics Engineers.
- [25] H. Yin and N. M. Allinson, "Self-organizing mixture networks for probability density estimation," *IEEE Transactions on Neural Networks*, vol. 12, no. 2, pp. 405-411, 2001.
- [26] Y. L. Tong, *The multivariate normal distribution*. Springer Science & Business Media, 2012.
- [27] J. Osborne, "Improving your data transformations: Applying the Box-Cox transformation," *Practical Assessment, Research, and Evaluation*, vol. 15, no. 1, p. 12, 2010.
- [28] L. Wasserman, *All of statistics: a concise course in statistical inference*. Springer Science & Business Media, 2013.
- [29] D. P. Bertsekas and J. N. Tsitsiklis, *Introduction to probability*. Athena Scientific Belmont, MA, 2002.
- [30] S.-H. Hsu, J. P.-A. Hsieh, T.-C. Chih, and K.-C. Hsu, "A two-stage architecture for stock price forecasting by integrating self-organizing map and support vector regression," *Expert Systems with Applications*, vol. 36, no. 4, pp. 7947-7951, 2009.
- [31] F.-J. Chang, L.-C. Chang, H.-S. Kao, and G.-R. Wu, "Assessing the effort of meteorological variables for evaporation estimation by self-organizing map neural network," *Journal of Hydrology*, vol. 384, no. 1-2, pp. 118-129, 2010.
- [32] J. J. Verbeek, N. Vlassis, and B. J. Kröse, "Self-organizing mixture models," *Neurocomputing*, vol. 63, pp. 99-123, 2005.
- [33] G. F. Lin and L. H. Chen, "Time series forecasting by combining the radial basis function network and the self-organizing map," *Hydrological Processes*, vol. 19, no. 10, pp. 1925-1937, 2005.
- [34] S. Kullback and R. A. Leibler, "On information and sufficiency," *The annals of mathematical statistics*, vol. 22, no. 1, pp. 79-86, 1951.
- [35] <http://www.aemo.com.au/Electricity/National-Electricity-Market-NEM/Data-dashboard#price-demand>.
- [36] S. A.-h. Soliman and A. M. Al-Kandari, *Electrical load forecasting: modeling and model construction*. Elsevier, 2010.
- [37] A. C. Rencher, *Methods of multivariate analysis*. John Wiley & Sons, 2003.
- [38] S. Sharma and S. Sharma, "Applied multivariate techniques," 1996.
- [39] <https://towardsdatascience.com/forecast-kpi-rmse-mae-mape-bias-cdc5703d242d>.
- [40] <https://medium.com/human-in-a-machine-world/mae-and-rmse-which-metric-is-better-e60ac3bde13d>.
- [41] A. Arteconi, "An Overview about Criticalities in the Modelling of Multi-Sector and Multi-Energy Systems," *Environments*, vol. 5, no. 12, p. 130, 2018.
- [42] N. S. Jaddi, S. Abdullah, and A. R. Hamdan, "Taguchi-based parameter designing of genetic algorithm for artificial neural network training," in *2013 International Conference on Informatics and Creative Multimedia*, 2013, pp. 278-281: IEEE.