An Ensemble of Single Multiplicative Neuron Models for Probabilistic Prediction

Ufuk Yolcu
Department of Statistics
Ankara University
Ankara, Turkey
uyolcu@ankara.edu.tr

Yaochu Jin
Department of Computer Science
University of Surrey
Surrey, United Kingdom
yaochu.jin@surrey.ac.uk

Erol Egrioglu
Department of Statistics
Giresun University
Giresun, Turkey
erol.egrioglu@giresun.edu.tr

***Abstract*—Inference systems basically aim to provide and present the knowledge (outputs) that decision-makers can take advantage of in their decision-making process. Nowadays one of the most commonly used inference systems for time series prediction is the computational inference system based on artificial neural networks. Although they have the ability of handling uncertainties and are capable of solving real life problems, neural networks have interpretability issues with regard to their outputs. For example, the outputs of neural networks that are difficult to interpret compared to statistical inference systems' outputs that involve a confidence interval and probabilities about possible values of predictions on top of the point estimations. In this study, an ensemble of single multiplicative neuron models based on bootstrap technique has been proposed to get probabilistic predictions. The main difference of the proposed ensemble model compared to conventional neural network models is that it is capable of getting a bootstrap confidence interval and probabilities of predictions. The performance of the proposed model is demonstrated on different time series. The obtained results show that the proposed ensemble model has a superior prediction performance in addition to having outputs that are more interpretable and applicable to probabilistic evaluations than conventional neural networks.**

Keywords— probabilistic prediction; single multiplicative neuron model; ensemble; bootstrap technique; time series prediction

I. INTRODUCTION

Time series analysis has been one of the most widely investigated data mining research topic in the last decades. The main purpose of time series analysis as a prediction tool is to come as close as possible to an accurate picture of the future. Time series prediction problems, from this point of view, can be seen as an inference problem. It can be mentioned from three kinds of inference system in time series prediction; the first one is statistical inference systems based on probability, the second one is fuzzy inference systems based on fuzzy sets and fuzzy arithmetic, and the last one is computational inference systems based on artificial neural networks. Statistical inference systems, from the point of their outputs that include probabilistic predictions, produce substantial knowledge to decision-makers. However, they are not able to predict complex real-world time series since they need some strict assumptions to be satisfied such as model assumptions, normal distribution and the number of observation. To get rid of these presumptive problems of statistical inference systems, in recent years, computational inference systems based on artificial neural networks (ANNs) have been widely utilized as an efficient prediction tool.

In the literature, although a great number of neuron models have been introduced [1]-[5], the most commonly used ANN model in the time series prediction literature is the multilayer perceptron (MLP) built based upon the McCuloch & Pitts neuron model [6]. The studies which make use of MLP to predict time series were reviewed in [7], [8]. Crone and Kourentzes [9] and Crone *et al.* [10] investigated the prediction performance of ANNs. Various comparative studies evaluating the results of different kinds of ANNs and some other time series prediction tools were presented in [3], [11]-[16]. The hybrid ANN models based on various approaches were proposed in [17]-[23]. Furthermore, in [24] and [25], the dynamic ANN models were introduced. While some researchers proposed new ANN models to predict various time series as in [26]-[30], many successful applications have also been found [31]-[33]. In applying MLPs to time series prediction, determination of the NN architecture is a challenging issue since the number of unit of hidden layer plays an important role in the prediction performance of MLPs. To address this problem, Egrioglu *et al.* put forward an approach to model selection strategy [34]. A single multiplicative neuron model (SMNM) that does not suffer from the architecture selection problem was introduced in [35]. Because SMNM has just one neuron in the hidden layer, it uses fewer parameters than MLPs [36]. In contrast to MLPs that use an additive function, SMNM uses a multiplicative function in its neuron as the aggregation function.

In the time series prediction literature, various approaches have also been proposed. One most widely used approach is the ensemble approach. Some of ensemble approaches are based on simple statistical parameters such as the simple mean, trimmed mean, winsorized mean, and median as in [37]-[39]. An outperformance method that determines the combining weights from the number of times the corresponding models performed best in past in-sample prediction trials was proposed in [40]. In addition, Winkler and Makridakis proposed a method based on Differential Weighting Scheme that

adaptively estimates the combining weights from the past prediction records of the constituent models [41]. Some analysts used Ordinary Least Squares (OLS) method in the determination of the weights of component methods to minimize the combined prediction SSE [4], [37]-[42]. Another combination strategy in ensemble approaches is based on the inverse proportion of in-sample prediction error such as the sum of squared errors (SSE). In the ensemble approach suggested in [43], [44], the weights for the component models are determined in such a way that a model with a larger error receives a smaller weight and vice versa. Smith and Jin [19] investigated different selection methods to produce an ensemble recurrent neural network (R-NN) trained by a hybrid multi-objective evolutionary algorithm for time series prediction. Moreover, for regression and time series prediction, an ensemble of deep learning belief networks (DBN) was proposed in [45] and a detailed reviewing study was produced about traditional as well as state-of-the-art ensemble methods in [46].

Although, in general, the above-mentioned prediction tools have outstanding prediction performance for time series analysis, their outputs are difficult to interpret. From a machine learning perspective, probabilistic predictions of time series can be considered as a probabilistic regression problem [47]. An efficient prediction tool for this kind of problems is the Gaussian process (GP) [48]. In addition to point estimation, GP also provides an error bound. Kersting et al. [49] proposed a heteroscedastic Gaussian process model based on the maximum a posteriori prediction (MHGP). This MHGP model is an extension of the standard GP. A variational heteroscedastic Gaussian process model (VHGP) was proposed in [50]. A new hybrid model for the probabilistic prediction of electricity price (A-VHGP) was proposed by combining the idea of VHGP and active learning [47].

In a prediction problem, the decision-makers want to take little risk in their decision making process, which can be achieved by having as much knowledge as possible. In fact, probabilistic predictions including interval estimation and probabilities of predictions present richer knowledge than point estimations to the decision-maker. The decision-makers also want to have tools that can be utilized in complex real life problems including most time series prediction problems. From this point of view, producing a prediction tool that will have outstanding features of both computational and statistical inference systems is of essential importance. For this purpose, we proposed in this work an ensemble of SMNM based on bootstrapping technique for time series prediction.

In the proposed model, an experimental distribution of predictions (outputs) is obtained by means of different bootstrap samples. By way of experimental distribution of bootstrap predictions, the probabilities of all possible prediction values and bootstrap confidence interval are obtained. In the proposed ensemble model, sub-samples are determined by using bootstrap techniques and the determination of weights and biases of model, in other word, the training of model, is carried out by using a modified particle swarm optimization (MPSO). The performance of proposed model is demonstrated on several benchmark problems in comparison with some other probabilistic and non-probabilistic models that are widely used in the time series prediction literature.

In the rest of the paper, section II introduces the proposed model in detail. In section III, the performances of different models are compared on a number of benchmark problems. Finally, a discussion and summary are given in the last section.

II. An Ensembel of Single Multiplicative Neuron Models

Nowadays, SMNM has become one commonly used computational inference system for time series analysis. Although SMNM produces useful and effective predictions for time series, it involves point estimations of time series just like all of neural network models (including conventional feedforward neural network models) that have been used for time series prediction. This restriction makes the model difficult to be interpreted and limits its usability. Since better interpretability means less risk, it is always desirable if the outputs of an inference system involve a confidence interval (interval estimation) and probabilities about possible values of predictions on top of point estimation. The proposed ensemble model (E-SMNM) based on bootstrap samples aims to provide the decision-makers with not only the ability of handling uncertainty but also as much knowledge as possible by generating probabilistic outputs.

The "probabilistic nature" of the proposed model's outputs (predictions) is in the bootstrap confidence interval. The interval estimation indicates that the predictions will remain within the lower and upper bounds of this bootstrap confidence interval at a certain probability. This can be fulfilled by using an empirical cumulative distribution function (E-CDF) of predictions that are obtained from each of the bootstrap samples. The obtaining of the E-CDF provides probabilities of all possible prediction values. Supposing that # bootstrap replication (# bootstrap sample) and bootstrap prediction for i^{th} ($i=1,2,...,NR_{BST}$) bootstrap sample are represented by NR_{BST} and f_i^{BST}, respectively. Let $\dot{F}$ be the E-CDF of f_*^{BST}. The 1-2α bootstrap confidence interval is defined by the α and 1- α percentiles of $\dot{F}$:

$$[f_{lb}^{BST}, f_{ub}^{BST}] = [\dot{F}^{-1}(\alpha), \dot{F}^{-1}(1-\alpha)] \quad (1)$$

At the same time, above bootstrap confidence interval represents an estimation of probabilities for bootstrap predictions based on the empirical distribution, as follows:

$$\Pi\,(f_{lb}^{BST} < f \leq f_{ub}^{BST}) = 1-2\alpha \quad (2)$$

$$\Pi\,(f \leq f_{ub}^{BST}) = 1-\alpha \quad (3)$$

$$\Pi\,(f < f_{lb}^{BST}) = \alpha \quad (4)$$

In the analysis process, for each bootstrap sample, the training of E-SMNM is realized by using MPSO. The flow diagram of the proposed E-SMNM can be given as in Fig. 1.

Now, let us try to explain that the process of the forming of sub-samples by using bootstrap technique. N_{TS}, N_{Test} and N_{Train} (N_{TS}- N_{Test}) are the number of observation of time series, test set and training set, respectively. The number of observation of sub-sample N_{SS} is determined from *Uniform (a,b)*. Here *a* and *b*;

$$a = integer\,[(NP_{TS} - NP_{Test}) \times 0.50] \tag{5}$$

$$b = integer\,[(NP_{TS} - NP_{Test}) \times 0.75] \tag{6}$$

$$N_{SS} = (integer\,[Uniform\,(a - b)]) + 1 \tag{7}$$

The starting point of sub-sample (t_{strt}) is determined from *Uniform (c, d)*. Here *c* and *d*;

$$c=1 \tag{8}$$

$$d=N_{TS} - N_{Test} - N_{SS} \tag{9}$$

For time series y_t, sub-sample time series;

$$y_t^{SS} = y_{t+tstrt-1};\ t=1,2,...,N_{SS} \tag{10}$$

To obtain the outputs of E-SMNM, computation algorithm can be given step by step.

Step 1. The parameters of process are determined.

T bootstrap repetition, that is, the value of NR_{BST} represents that how many times repetition will be applied. # input of E-SMNM, that is, model order or values of $NR_{INP}=q$ represents the number of lagged variable for sub-sample time series.

Step 2. The i^{th} ($i=1,2,...,NR_{BST}$) sub-sample is determined by bootstrap.

Step 3. Reset the cycle counter *c*.

Step 4. Increase *c* by 1.

$$c=c+1 \tag{11}$$

Step 5. The calculations for each learning sample of i^{th} sub-sample are performed.

The calculations for l^{th} ($l=1,2,...,NR_{SS\text{-}q}$) learning sample of i^{th} sub-sample are performed. Since the proposed probabilistic model is basically a SMNM, it has same structure with SMNM and it can be demonstrated in Fig. 2.

In Fig. 1, $y_{t\text{-}1}^{SS(i)}$, $y_{t\text{-}2}^{SS(i)}$,..., $y_{t\text{-}q}^{SS(i)}$ lagged variables of sub-sample time series are inputs of E-SMNM, $\hat{y}_{t\text{-}1}^{SS(i)}$ is the prediction for t^{th} observation of i^{th} sub-sample time series. The target value for t^{th} observation of i^{th} sub-sample time series will also be $y_t^{SS(i)}$. Moreover, the function Ω composes of multiplication weighted inputs and *f* represents activation function. Activation value and output value of the model for l^{th} learning sample of i^{th} sub-sample;

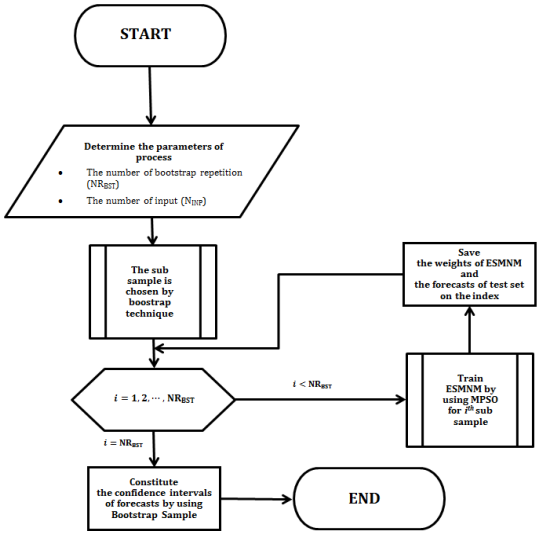

Fig. 1. The diagram of the proposed E-SMNM

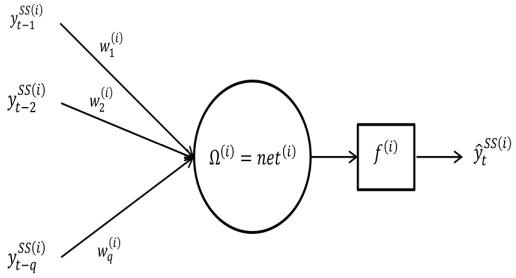

Fig. 2. The architecture of the E-SMNM

$$net_t^{(i)} = \prod_{n=1}^{q}\left(w_n^{(i)} \times y_{t-n}^{SS(i)} + b_n^{(i)}\right)\ ;t = 1,2,\cdots,N_{SS}^{(i)} \tag{12}$$

$$\hat{y}_t^{SS(i)} = f^{(i)}\left(net_t^{(i)}\right) = \left(1 - \exp\left(net_t^{(i)}\right)\right)^{-1} \tag{13}$$

Step 6. Check the stopped criteria.

If the cycles counter *c* equal NR_{BST} then stop the process, or else go to Steps 2 to determine new sub-sample by bootstrap technique.

For each bootstrap sample, while the proposed E-SMNM is producing the outputs for each learning sample, it uses weights and biases that are obtained by making use of MPSO in an optimization process. Particle swarm optimization (PSO) is a swarm optimization technique proposed by Kennedy and Eberhart [51]. The particle swarm concept originated as a simulation of simplified social system [52]. Distinguishing feature of this heuristic algorithm is that it simultaneously examines different points in different regions of the solution space to obtain the global optimum solution. The risk of getting trapped in a local optimum can be reduced because of the population based search method [36]. The MPSO algorithm

has time varying inertia weight as suggested in [53]. In a similar way, this algorithm also has varying changing acceleration coefficient as in [54].

In this optimization process, the weights and the biases, $w_n^{(i)}$ and $b_n^{(i)}$ $(n=1,2,\ldots,q)$, of the E-SMNM will be trained. Thus, each particle of the swarm has $2\times q$ positions. The structure of a particle in a swarm is illustrated in Fig. 3.

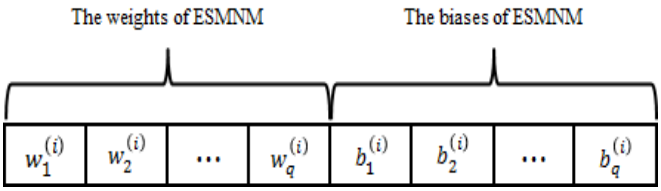

Fig. 3. The structure of a particle in a swarm

When the maximum iteration number is reached the optimum values of weights and biases are specified. And then the training of E-SMNM is completed. For the test set of time series, bootstrap predictions can be obtained as follow based on an ensemble of each individual prediction;

$$\hat{y}_t^{BST} = E\left(\hat{y}_t^{SS}\right) = \frac{1}{NR_{BST}} \sum_{i=1}^{NR_{BST}} \hat{y}_t^{SS(i)} \tag{14}$$

III. The Implementations

To examine the performance of proposed E-SMNM, various experiments are performed.

A. The Evaluation of the Point Estimation Performance

In the proposed model, predictions are obtained as a combination of the predictions of each bootstrap samples. Therefore we firstly intend to evaluate the prediction performance of the proposed method together with an ensemble method. For this purpose, two time series data sets, the Mackey-Glass and Sunspot time series are predicted, and the obtained results are interpreted with an ensemble time series prediction results given in [19].

The Mackey-Glass time series [5] is a benchmark problem that has been widely used in the literature due to its chaotic behaviour. In the implementation, for the Mackey-Glass time series with 1000 observations, the last 500 data points and for Sunspot time series data with 2000 data points from November 1834 to June 2001, the last 1000 data points are used as a test set as in [19] and the model parameters are chosen as $pn=30$,$vm=2$, $(c_{1i}, c_{1f}) =(2, 1)$, $(c_{2i}, c_{2f}) =(1, 2)$, $(w_1, w_2)=(0.4, 0.9)$, $itr_{max}=200$. The number of inputs is chosen between 2 and 10. The obtained results are evaluated in Table I in terms of root mean square error (RMSE) criteria calculated for test sets.

$$RMSE = \sqrt{\frac{1}{N}\sum_{i=1}^{N}\left(y_t - \hat{y}_t\right)^2} \tag{15}$$

TABLE I. RMSE Comparison of Point Estimation

	The Mackey-Glass		Sunspot	
Model	***RMSE***	***Rank***	***RMSE***	***Rank***
RNN (CC-SL) [55]	6.33E-03	2	1.66E 02	5
RNN (GD) [56]	**3.72E-05**	**1**	**1.19E 02**	**1**
ERNN (GA) [57]	---	---	1.29E 02	3
RNN (GA) [58]	1.22E-02	5	---	---
RNN (H-MOEA) [19]	7.53E-03	4	1.52E 02	4
E-SMNM	6.62E-03	3	1.25E 02	2

The results represented in this table except the proposed methods are taken from [19]. From Table I, although the proposed probabilistic prediction model does not offer the best point estimations in terms of RMSE, we can say that it is a competitive prediction tool.

B. The Evaluation of the Probabilistic Prediction Performance

Secondly, we intend to evaluate the probabilistic prediction ability of our method as well as ability of point estimation. To this end, the electricity price dataset in year 2013 in, New South Wales (NSW) taken from the Australian National Electricity Market website [59] is analyzed. Obtained results are compared with some other probabilistic models such as the standard GP, A-VHGP, generalized autoregressive conditional heteroscedasticity (GARCH) and splines quantile regression (SQR), and non-probabilistic models such as feed forward neural networks (FFNN), extreme learning machine (ELM), support vector regression (SVR) and adaptive neuro fuzzy inference system (ANFIS). In comparison, in order to assess the probabilistic prediction performance, two performance index; the reliability evaluation (RE) criterion and the sharpness evaluation (SE) criterion are employed. Moreover, mean absolute percentage error (MAPE) criterion is utilized to evaluate the point estimation performance of the methods. These criteria can be given as in (16)-(18).

$$RE = \left(\frac{\xi^{(1-\alpha)}}{N} - (1-\alpha)\right) \times 100\% \tag{16}$$

$$SE = \frac{1}{N}\sum_{t=1}^{N}\left(Ub_{1-\alpha}(\hat{y}_t) - Lb_{1-\alpha}(\hat{y}_t)\right) \tag{17}$$

$$MAPE = \frac{1}{N}\sum_{t=1}^{N}\left|\frac{y_t - \hat{y}_t}{y_t}\right| \tag{18}$$

Where N is # testing points, $\xi^{(1-\alpha)}$ is # times that actual target values do indeed lie within the α-level prediction intervals. $Lb_{1-\alpha}(\hat{y}_t)$ and $Ub_{1-\alpha}(\hat{y}_t)$ are the lower and upper bound of the α-level prediction interval. y_t and $\hat{y}_t$ are the target value and the output value at t, respectively. The observed proportions $\xi^{(1-\alpha)}/N$ should as close as possible to the pre-assigned probability $(1-\alpha)$, the difference between them is the bias of the probabilistic prediction method. This bias is measured by the RE criterion that is, the smaller the $|RE|$, the better. The SE criterion is used to measure the mean width of the prediction intervals, thus assessing the sharpness of the

predictive distribution. This criterion expresses the ability of the model to concentrate the uncertainty information of the probabilistic predictions. What is desired is the smaller SE value.

As in [47], to assess the capability of overcoming the different price patterns of different seasons, the whole dataset is divided into four subsets as the first, second, third, and fourth quarter of year. For each subset, the data of the last month is used as the testing set. In implementations of E-SMNM, the parameters were taken as in previous implementations. The number of inputs is chosen as 12, 18 and 24. Moreover, the bootstrap predictions are obtained by expected value of the predictions that are produced in each of 200 bootstrap repetitions.

The probabilistic prediction results (α=0.10) for the last 50 datum points of first quarter subset are plotted in Fig. 4, together with the point estimation and the actual price observations. From Fig. 4, it can be seen that while 1488th prediction is less than the real observation, 1489th prediction is bigger than the related real observation. In other words, in this boxplots, the points with white box indicate that the prediction is less than the related real observation and the points with black box are vice versa. At the same time, the boxes having lines their both side show that the confidence interval involves the related real observation.

To compare the probabilistic prediction of the methods, obtained SE and $|RE|$ criteria for 90%-level prediction intervals (PI) are shown in Table II and III. Apart from the proposed model, the criteria values are taken from [47].

From Table II, it is seen that the proposed E-SMNM has the smallest SE value for 1st and 3rd quarters**,** which means that E-SMNM has the smallest prediction intervals. On the one hand, the SE given by the proposed probabilistic neural network comes up with result better than the standard GP for 2nd quarter and also E-SMNM produces the largest prediction intervals for 4th quarter. From Table III it is clearly seen that the proposed probabilistic prediction model has the best performance in terms of this criterion for all quarter data. The absolute values of $|RE|$ obtained from E-SMNM are smaller than 4%. It means that the prediction intervals given by E-SMNM perfectly contain the actual observations especially for 1st quarter with 0.591%.

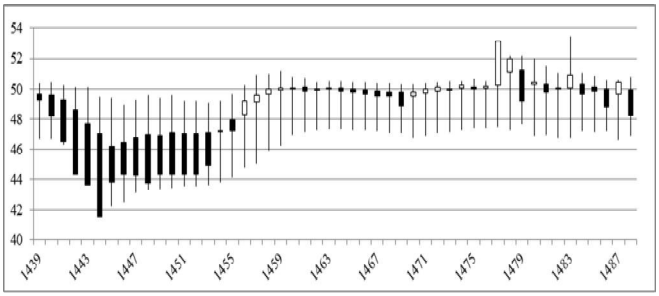

Fig. 4. 90%-level prediction intervals given by E-SMNM

TABLE II. SE CRITERION (%) OF THE 90%-LEVEL PI, FOR THE NSW

Dataset	1st Quarter		2nd Quarter		3rd Quarter		4th Quarter	
Model	*SE*	*Rank*	*SE*	*Rank*	*SE*	*Rank*	*SE*	*Rank*
SQR	5.548	3	8.051	3	5.907	3	3.221	2
GARCH	5.115	2	**7.441**	**1**	5.865	2	**3.089**	**1**
GP	7.489	5	9.507	5	7.175	5	3.427	3
A-VHGP	5.623	4	7.994	2	6.275	4	3.497	4
E-SMNM	**2.691**	**1**	8.869	4	**4.582**	**1**	5.019	5

TABLE III. RE CRITERION (%) OF THE 90%-LEVEL PI, FOR THE NSW

Dataset	1st Quarter		2nd Quarter		3rd Quarter		4th Quarter	
Model	*|RE|*	*Rank*	*|RE|*	*Rank*	*|RE|*	*Rank*	*|RE|*	*Rank*
SQR	5.623	5	5.095	3	5.964	3	4.789	4
GARCH	5.482	4	5.299	5	6.168	5	4.773	3
GP	4.561	3	5.247	4	6.119	4	4.839	5
A-HGP	3.553	2	3.253	2	3.809	2	2.860	2
E-SMNM	**0.591**	**1**	**2.639**	**1**	**3.542**	**1**	**2.769**	**1**

Finally, prediction performance of the proposed model is evaluated by using the MAPE criterion. In this comparative evaluation, the results of various prediction models, SVR, FFNN, ANFIS, ELM, standard GP, GARCH, and SQR are taken from [47]. The obtained findings are summarized in Table IV. From this table, it can be obviously inferred that the proposed E-SMNM has the best prediction performance for all quarter data sets except 2nd quarter. Even for 2nd quarter data, in terms of MAPE, the proposed model has better prediction results than whole models except A-HGP and ELM.

To support that the proposed model has a satisfactory prediction performance, New England ISO data in year 2011 [60] are also analyzed and the obtained MAPE criterion values are represented in Table V. The whole dataset, as in previous experiment, is divided into four subsets as the quarter of year and the data of the last month is used as the testing set for each subset. In implementations of E-SMNM, the parameters of the model are taken as in implementation of NSW data. From Table V, the proposed E-SMNM produces the best point estimation with respect to MAPE criterion for 1st and 3rd quarter data. Although the proposed prediction model does not have the best performance for 2nd quarter, it gives satisfactory prediction results with 4.611%. For 4th quarter, even if the proposed model is not one of the top four models in terms of point estimation success, because all models reveal around 7% MAPE values, the proposed neural network can be seen as competitive prediction tool with 7.203%.

TABLE IV. MAPE (%) COMPARISON OF POINT ESTIMATION FOR THE NSW

Dataset	1st Quarter		2nd Quarter		3rd Quarter		4th Quarter	
Model	*MAPE*	*Rank*	*MAPE*	*Rank*	*MAPE*	*Rank*	*MAPE*	*Rank*
ELM	4.288	2	6.117	2	4.658	3	3.881	7
ANFIS	5.002	8	7.128	8	4.876	6	3.915	8
FFNN	4.305	4	6.228	5	4.709	4	3.721	4
SVR	4.293	3	6.215	4	4.344	2	3.812	5
SQR	4.912	7	6.323	6	4.922	7	3.880	6
GARCH	5.203	9	8.990	9	6.784	9	5.033	9
GP	4.908	6	6.513	7	5.337	8	3.616	2
A-HGP	4.442	5	**6.110**	**1**	4.815	5	3.709	3
E-SMNM	**3.873**	**1**	6.132	3	**3.541**	**1**	**3.544**	**1**

TABLE V. MAPE (%) COMPARISON OF POINT ESTIMATION FOR THE ISO

Dataset	1st Quarter		2nd Quarter		3rd Quarter		4th Quarter	
Model	*MAPE*	*Rank*	*MAPE*	*Rank*	*MAPE*	*Rank*	*MAPE*	*Rank*
ELM	5.632	4	4.775	5	5.193	3	**6.966**	**1**
ANFIS	5.689	5	4.853	8	5.879	9	7.977	8
FFNN	5.553	3	4.562	2	5.255	5	7.171	4
SVR	5.238	2	4.780	6	5.254	4	7.032	2
SQR	5.911	8	4.813	7	5.842	7	7.213	6
GARCH	5.769	7	6.031	9	5.017	2	9.873	9
GP	6.223	9	**4.522**	**1**	5.859	8	7.347	7
A-HGP	5.739	6	4.595	3	5.279	6	7.040	3
E-SMNM	**5.022**	**1**	4.611	4	**3.535**	**1**	7.203	5

The interval estimation indicates that the predictions will remain within the lower and upper bounds of this bootstrap confidence interval at a certain probability. This can be fulfilled by means of an empirical distribution of predictions that are obtained from each of the bootstrap samples. Moreover, the obtaining of the empirical distribution provides to get the probabilities of all possible prediction values. In this respect, another predominant characteristic of the proposed E-SMNM is that it enables the probabilistic evaluation of the outputs containing additional information than point estimations.

In this study, as another indication of the probabilistic nature of the proposed method in addition to interval estimation accentuated before, we presented some probabilities of first prediction obtained for 4th quarter of NSW data. The graphs of E-CDF and some possibilities for the first prediction are given in Fig. 5 and in Table VI, respectively. In a similar way, E-CDF and probabilities can be obtained for all predictions.

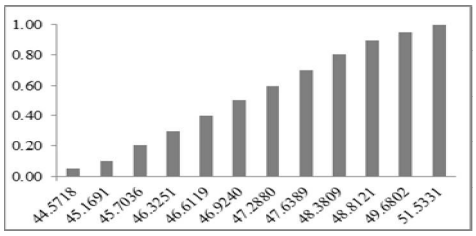

Fig. 5. The E-CDF for first prediction/NSW-4th quarter

TABLE VI. SOME PROBABILITIES OF FIRST PREDICTION FOR THE NSW / 4TH QUARTER

First Prediction	
$P(44.5718<f_1\leq 49.6802)=0.90$	$P(f_1>45.1691)=0.10$
$P(f_1\leq 48.8121)=0.90$	$P(f_1\leq 46.9240)=0.50$

C. The Evaluation of the Reliability

Moreover, to evaluate the reliability and the behaviour of E-SMNM in case of the variability that may emerge from different random initializations, 10 different real-world time series (daily observed data) are analyzed by performing 30 times with different random initializations. These time series are composed of logarithmic daily basis stock exchange of Dow Jones Futures (DJF) and National Association of Securities Dealers Automated Quotations (NSDQ). These time series and their some features used in the implementations are given in Table VII. In these implementations of E-SMNM and the other models based on MPSO, the parameters were taken as in previous implementations. Moreover, the bootstrap predictions are obtained by expected value of the predictions that are produced in each of 200 bootstrap repetitions. The mean values of the error criteria obtained from 30 performances are evaluated.

In the evaluation and comparison of results, 6 methods are used; SMNM-P (single multiplicative neuron model trained by PSO), SMNMRB-P (radial basis single multiplicative neuron model ANN trained by PSO), MLP-LM (multi-layer perceptron feed-forward NN trained by Levenberg-Marquardt), ELMAN-LM (multi-layer perceptron feed-back NN trained by Levenberg-Marquardt), SMNM-R-P (single multiplicative recurrent neuron model trained by PSO). Obtained results were evaluated together in RMSE criteria calculated for test sets. For five DJF time series, the results of methods obtained from 30 different random initializations are summarized in Table VIII. In consideration of Table VIII created for DJF time series, it is clearly seen that the proposed E-SMNM has the best performance.

To investigate whether there is a significant difference among the performance of the prediction models, all the results in terms of RMSE criteria values are statistically analyzed. In this analysis process, firstly, it is confirmed that the results obtained from each models are not normal-distributed and so in the statistical analysis, Kruskal–Wallis H test which is a non-parametric test are utilized to statistically compare the performance of the models in alpha level of 0.05. The analysis results demonstrate that there are significant statistical differences among the models' prediction performances for all DJF data sets ($p<0.001$). In this situation, it is necessary to determine which models have these statistical significant differences. For this purpose, Mann-Whitney U test which is also a non-parametric test is used to make a paired comparison the proposed E-SMNM with other prediction models. From the results of this test, it is said that the proposed prediction model, for DJF2010, DJF2013 and DJF2014 data sets, has the best performance (with smaller Mean Rank and $p<0.001$). Moreover, while the E-SMNM comes off better than all models except MLP-LM for DJF2011 (*Mean Rank of E-SMNM=**35.87**>Mean Rank of MLP-LM=**25.13*** and $p=0.017$), it shares the best performance with MLP-LM for DJF2011 ($p=0.214$).

Finally, we analyzed NSDQ data and also presented the evaluation values of the error criteria obtained from 30 different random initializations in Table IX. This table contains the result showing that the proposed model has the best performance except for the data belonging to 2013. Similarly, the results of Kruskal–Wallis H test are the evidence that there are significant statistical differences between the prediction performance of the models for NSDQ data sets ($p<0.001$). When it comes to the results of Mann-Whitney U test, the proposed model has the superior forecasting performance for NSDQ2011, NSDQ2012 and NSDQ2014 data sets (with smaller Mean Rank and $p<0.001$). Moreover, the E-SMNM shares the best performance with MLP-ELMAN-LM for NSDQ2010 ($p=0.605$).

TABLE VII. TIME SERIES USED IN IMPLEMENTATIONS

	DJF		NASDAQ	
Time Series	*# obs.*	*# test set*	*# obs.*	*# test set*
2010	252	22	252	22
2011	251	21	251	21
2012	250	20	250	20
2013	252	21	252	21
2014	252	22	252	22

TABLE VIII. THE SUMMERY RESULTS OF IMPLEMENTATIONS FOR DJF

	Data				
	DJF 2010	***DJF 2011***	***DJF 2012***	***DJF 2013***	***DJF 2014***
E-SMNM	**0.0178**	**0.0184**	**0.0080**	**0.0301**	**0.0191**
SMNM-P	0.0242	0.0246	0.0089	0.0333	0.0219
SMNMRB-P	0.0285	0.0240	0.0105	0.0362	0.0238
MLP-PSO	0.0485	0.0214	0.0092	0.0490	0.0339
MLP-LM	0.0250	0.0190	0.0085	0.1729	0.0981
ELMAN-LM	0.0248	0.0190	0.0087	0.1657	0.1062
SMNM-R-P	0.0431	0.0424	0.0103	0.0400	0.0331

TABLE IX. THE SUMMERY RESULTS OF IMPLEMENTATIONS FOR DJF

	Data				
	NSDQ 2010	***NSDQ 2011***	***NSDQ 2012***	***NSDQ 2013***	***NSDQ 2014***
E-SMNM	**0.0210**	**0.0130**	**0.0095**	0.0837	**0.0142**
SMNM-P	0.0344	0.0145	0.0118	0.0406	0.0196
SMNMRB-P	0.0435	0.0182	0.0147	0.0588	0.0373
MLP-PSO	0.0971	0.0176	0.0131	0.1115	0.0739
MLP-LM	0.0278	0.0160	0.0107	**0.0167**	0.0218
ELMAN-LM	0.0249	0.0156	0.0104	0.0310	0.0217
SMNM-R-P	0.0951	0.0137	0.0113	0.0764	0.0611

IV. CONCLUSION AND DISCUSSION

Particularly over the past decade, as a computational inference system, SMNMs that do not have the architecture selection problem have popular for time series prediction. Although SMNM produces impressive results in time series analysis, its outputs fail to provide interpretable information to decision-makers in comparison with outputs of statistical inference systems that make a stochastic evaluation available for them by including the confidence interval and probabilities about possible values of predictions in addition to point estimations. To overcome this drawback, we suggested an ensemble of SMNMs based on bootstrap technique. The suggested model produces an empirical distribution function of predictions that are obtained from each of the bootstrap samples so that the probabilities of all possible prediction values and bootstrap confidence intervals can be acquired in the form of probabilistic predictions. Besides having the distinguishing feature from the point of its outputs, the superior prediction performance of proposed model is demonstrated on various benchmark problems.

In the future, different kinds of probabilistic ANN structures will be tried based on other bootstrap techniques to further improve the prediction ability of the proposed approach.

ACKNOWLEDGMENT

This study is supported by "The Scientific and Technological Research Council of Turkey (TUBITAK)" as part of "2219 - International Postdoctoral Research Scholarship Programme".

REFERENCES

[1] M. Basu and T. K. Ho, "Learning behavior of single neuron classifiers on linearly separable or non-separable inputs," In: IEEE LICNN'99 /1, 1999.

[2] P. R. A. Firmino, P. S. G. De Mattos-Neto and T. A. E. Ferreira, "Error modeling approach to improve time series forecasters," Neurocomputing, vol. 153, pp. 242–254, 2015

[3] Y. Gong, Y. Zhang, S. Lan and H. Wang, "A Comparative Study of Artificial Neural Networks, Support Vector Machines and Adaptive Neuro Fuzzy Inference System for Forecasting Groundwater Levels near Lake Okeechobee Florida," Water Resources Management, vol. 30, pp. 375–391, 2016.

[4] C. W. J. Granger and R. Ramanathan, "Improved Methods of combining forecasts," International Journal of Forecasting, vol. 3, pp. 197-204, 1984.

[5] M. C. Mackey and L. Glass, "Oscillation and chaos in physiological control systems," Science, vol. 4300, pp. 287–289, 1977.

[6] W.S. McCulloch and W. Pitts, "A logical calculus of the ideas immanent in nervous activity," Bull Math Biophys vol. 5 pp. 115–133, 1943.

[7] G. Zhang, "Time series forecasting using a hybrid ARIMA and neural network model," Neurocomputting, vol. 50, pp. 159–175, 2003.

[8] G. Zhang, B.E. Patuwo, and Y.M. Hu, "Forecasting with artificial neural networks: the state of the art," International journal of Forecasting, vol. 14 pp. 35–62, 1998.

[9] S. Crone and N. Kourentzes, "Naive support vector regression and multilayer perceptron benchmarks for the 2010 neural network grand competition (NNGC) on time series prediction," In: Proceedings of the international joint Conference on neural networks, IJCNN'10, Barcelona, IEEE, New York, 2010.

[10] S. Crone, K. Nikolopoulos and M Hibon, "New evidence on the accuracy of computational intelligence for monthly time series prediction—results of the NN3 forecasting competition," International journal of Forecasting, vol. 27(3), pp. 635–660, 2011.

[11] Z. He, Y Zhang, Q. Guo and X. Zhao, "Comparative Study of Artificial Neural Networks and Wavelet Artificial Neural Networks for Groundwater Depth Data Forecasting with Various Curve Fractal Dimensions," Water Resources Management vol. 28, pp. 5297–5317, 2014.

[12] M. G. De-Giorgi, S. Campilongo, S. Ficarella and P. M. Congedo, "Comparison between wind power prediction models based on wavelet decomposition with least-squares support vector machine (LS-SVM) and artificial neural network (ANN)," Energies vol. 7, pp. 5251-5272, 2014.

[13] X. Y. Chen, K. W. Chau and A.O. Busari, "A comparative study of population-based optimization algorithms for downstream river flow forecasting by a hybrid neural network model," Engineering Applications of Artificial Intelligence vol. 46, pp. 258–268, 2015.

[14] N.H.A. Rahman, M.H. Lee, Suhartono and M. T. Latif, "Artificial neural networks and fuzzy time series forecasting: an application to air quality," Quality & Quantity, vol. 49, pp. 2633–2647, 2015.

[15] A. P. Piotrowski, M. J. Napiorkowski, J. J. Napiorkowski and M. Osuch, "Comparing various artificial neural network types for water temperature prediction in rivers," Journal of Hydrology vol. 529, pp. 302–315, 2015.

[16] O. Claveria, E. Monte and S. Torra, "Tourism demand forecasting with neural network models: different ways of treating information," International Journal of Tourism Research, Int. J. Tourism Res., vol. 17, vol. 492–500, 2015.

[17] S. Wei, H. Yang, J. Song, K. Abbaspour and Z. Xu, "A wavelet-neural networkhybrid modelling approach for estimating and predicting river monthly flows," Hydrological Sciences Journal, vol. 58, pp. 374–389, 2013.

[18] N. Babu and B. E. Reddy, "A moving-average filter based hybrid ARIMA–ANN model for forecasting time series data," Applied Soft Computing vol. 23, pp. 27–38, 2014.

[19] C. Smith, Y. Jin, "Evolutionary multi-objective generation of recurrent neural network ensembles for time series prediction," Neurocomputing vol. 143, pp. 302–311, 2014.

[20] J. J. Ruiz-Aguilar, I. J. Turias and M. J. Jiménez-Come, "Hybrid approaches based on SARIMA and artificial neural networks for inspection time series forecasting," Transportation Research Part E vol. 67, pp. 1–13, 2014.

[21] R. Adhikari and R. K. Agrawal, "A combination of artificial neural network and random walk models for financial time series forecasting," Neural Computing and Application vol. 24, pp. 1441–1449, 2014.

[22] A. K. Fard and M. R. Akbari-Zadeh, "A hybrid method based on wavelet, ANN and ARIMA model for short-term load forecasting," Journal of Experimental & Theoretical Artificial Intelligence, vol. 26(2), pp. 167–182, 2014.

[23] P. S. G. De Mattos-Neto, F. Madeiro, T. A. E. Ferreira and G. D. C. Cavalcanti, "Hybrid intelligent system for air quality forecasting using phase adjustment," Engineering Applications of Artificial Intelligence vol. 32, pp. 185–191, 2014.

[24] F. Almonacid, P. J. Pérez-Higueras, E. Fernández and L. Hontoria, "A methodology based on dynamic artificial neural network for short-term forecasting of the power output of a PV generator," Energy Conversion and Management vol. 85, pp. 389–398, 2014.

[25] A. A. Godarzi, R. M. Amiri, A. Talaei and T. Jamasb, "Predicting oil price movements: A dynamic Artificial Neural Network approach," Energy Policy, vol. 68, pp. 371–382, 2014.

[26] C. Voyant, P. Randimbivololona, M. L. Nivet, C. Paolic and M. Musellic, "Twenty four hours ahead global irradiation forecasting using multi-layer perceptron," Meteorological Applications, vol. 21, pp. 644–655, 2014.

[27] L. A. Laboissiere, R. A. S. Fernandes and G. G. Lage, "Maximum and minimum stock price forecasting of Brazilian power distribution companies based on artificial neural networks," Applied Soft Computing, vol. 35, pp. 66–74, 2015.

[28] C. T. Cheng, W. J. Niu, Z. K. Feng, J. J. Shen and K. W. Chau, "Daily reservoir runoff forecasting method using artificial neural network based on quantum-behaved particle swarm optimization," Water vol. 7, pp. 4232-4246, 2015.

[29] M.K. Kim, "T new approach to short-term price forecast strategy with an artificial neural network approach: application to the nord pool," J Electr Eng Technol., vol. 10(4), pp. 1480-1491, 2015.

[30] W. C. Wang, K. W. Chau, L. Qiu and Y. B. Chen, "Improving forecasting accuracy of medium and long-term runoff using artificial neural network based on EEMD decomposition," Environmental Research vol. 139, pp. 46–54, 2015.

[31] R. C. Deo and M. Sahin, "Application of the Artificial Neural Network model for prediction of monthly Standardized Precipitation and Evapotranspiration Index using hydro meteorological parameters and climate indices in eastern Australia," Atmospheric Research, vol. 161–162, pp. 65–81, 2015.

[32] Y. Falamarzi, N. Palizdan, Y. F. Huang and T.S. Lee, "Estimating evapotranspiration from temperature and wind speed data using artificial and wavelet neural networks (WNNs)," Agricultural Water Management, vol. 140, pp. 26–36, 2014.

[33] R. Ata, "Artificial neural networks applications in wind energy systems: a review," Renewable and Sustainable Energy Reviews, vol. 49, pp. 534–562, 2015.

[34] E. Egrioglu, C. H. Aladag and S. Gunay, A new model selection strategy in artificial neural network," Applied Soft Computing, vol. 195, pp. 591–597, 2008.

[35] R. N. Yadav, P. K. Kalra and J. John, "Time series prediction with single multiplicative neuron model," Applied Soft Computing, vol. 7, pp. 1157–1163, 2007.

[36] C. H. Aladag, E. Egrioglu, U. Yolcu, "Robust multilayer neural network based on median neuron model," Neural Computing and Applications, vol. 24, pp. 945–956, 2014.

[37] L. M. De Menezes, D. W. Bunn and J. W. Taylor, "Review of guidelines for the use of combined forecasts," European Journal of Operational Research, vol 120(1), pp. 190-204, 2000.

[38] J. G. Gooijer and R. J. Hyndman, "25 years of time series forecasting." International Journal of Forecasting, vol. 22, pp. 443-473, 2006.

[39] V. R. R. Jose and R. L. Winkler, "Simple robust averages of forecasts: some empirical results," International Journal of Forecasting, vol. 24(1), pp. 163-169, 2008.

[40] D. Bunn, "A Bayesian approach to the linear combination of forecasts," Operational Research Quarterly, vol. 26(2), pp. 325-329, 1975.

[41] R. L. Winkler and S. Makridakis, "The combination of forecasts," Journal of the Royal Statistical Society A, vol. 146(2), pp. 150-157, 1983.

[42] P. S. Frietas and A. J. Rodrigues, "Model combination in neural-based forecasting," European Journal of Operational Research vol. 173, pp. 801-814, 2006.

[43] C. Lemke and B. Gabrys, "Meta-learning for time series forecasting and forecast combination," Neurocomputing, vol. 73, pp. 2006–2016, 2010.

[44] R. Adhikari and R. K. Agrawal, "Performance evaluation of weight selection schemes for linear combination of multiple forecasts," Artificial Intelligence Review, vol. 42(4), pp. 529-548. 2014.

[45] X. Qiu, L. Zhang, Y. Ren, P. N. Suganthan and G. Amaratunga, "Ensemble Deep Learning for Regression and Time Series Forecasting," In: Computational Intelligence in Ensemble Learning (CIEL), IEEE Symposium, 2014.

[46] Y. Ren, L. Zhang and P. N. Suganthan, "Ensemble Classification and Regression – Recent Developments, Applications and Future Directions," IEEE Computational Intelligence Magazine, vol. 11 (1), pp.41-53, 2016.

[47] P. Kou, D. Liang, L. Gao and J Lou, "Probabilistic electricity price forecasting with variational heteroscedastic Gaussian process and active learning," Energy Conversion and Management, vol. 89, pp. 298–308, 2015.

[48] C. Rasmussen and C. Williams, Gaussian processes for machine learning, Cambridge: Massachusetts, London: The MIT Press, 2006.

[49] K. Kersting, C. Plagemann, P. Pfaff and B. Wolfram, "Most likely heteroscedastic Gaussian process regression," In: Proc of international conference on machine learning, Corvallis, OR, USA, 2007.

[50] M. Lazaro-Gredilla and M. Titsias, "Variational heteroscedastic Gaussian process regression," In: Proc of international conference on machine learning, Bellevue, WA, USA, 2011.

[51] J. Kennedy and R. Eberhart, "Particle swarm optimization,". In: Proceedings of IEEE international conference on neural networks. Piscataway, NJ, USA. IEEE Press, pp. 1942–1948, 1995.

[52] W. Shuhua, Z. Xueliang, L. Hainan, L. Shuyang and W. Jiaying, "A Modifled Particle Swarm Optimization Algorithm," In: Neural Networks and Brain, ICNN&B '05. International Conference, pp 318–321, 2005.

[53] Y. Shi and R. C. Eberhart, "Empirical study of particle swarm optimization," Proc IEEE Int Congr Evol Comput, vol. 3, pp. 101–106, 1999.

[54] Y. Ma, C. Jiang, Z. Hou and C. Wang, "The formulation of the optimal strategies for the electricity producers based on the particle swarm optimization algorithm," IEEE Transactions on Power Systems, vol. 21(4), pp. 1663–1671, 2006.

[55] R. Chandra and M. Zhang, "Cooperative coevolution of Elman recurrent neural networks for chaotic time series prediction," Neurocomputing, vol. 86, pp. 116–123, 2012.

[56] M. Ardalani-Farsa and S. Zolfaghari, "Chaotic time series prediction with residual analysis method using hybrid elman-narx neural networks," Neurocomputing vol. 73(13–15), pp. 2540–2553, 2010.

[57] Q. Ma, Q. Zheng, H. Peng, T. Zhong and L. Xu, "Chaotic time series prediction based on evolving recurrent neural networks," In: International Conference on Machine Learning and Cybernetics, vol. 6, pp. 3496–3500, 2007.

[58] A. A. Ferreira, T. B. Ludermir and R. R. DeAquino, "Comparing recurrent networks for time-series forecasting," In: The 2012 International Joint Conference on Neural Networks (IJCNN), pp. 1–8, 2012.

[59] National Electricity Market Management Company Limited. <http://www.nemmco.com.au>.

[60] ISO New England. < http://www.iso-ne.com/>