2014 International Joint Conference on Neural Networks (IJCNN)
July 6-11, 2014, Beijing, China

Aggregation of PI-based Forecast to Enhance Prediction Accuracy

Mohammad Anwar Hosen, Abbas Khosravi, Saeid Nahavandi and Douglas Creighton

***Abstract*—In contrast to point forecast, prediction interval-based neural network offers itself as an effective tool to quantify the uncertainty and disturbances that associated with process data. However, single best neural network (NN) does not always guarantee to predict better quality of forecast for different data sets or a whole range of data set. Literature reported that ensemble of NNs using forecast combination produces stable and consistence forecast than single best NN. In this work, a NNs ensemble procedure is introduced to construct better quality of PIs. Weighted averaging forecasts combination mechanism is employed to combine the PI-based forecast. As the key contribution of this paper, a new PI-based cost function is proposed to optimize the individual weights for NN in combination process. An optimization algorithm, named simulated annealing (SA) is used to minimize the PI-based cost function. Finally, the proposed method is examined in two different case studies and compared the results with the individual best NNs and available simple averaging PIs aggregating method. Simulation results demonstrated that the proposed method improved the quality of PIs than individual best NNs and simple averaging ensemble method.**

I. INTRODUCTION

NEURAL NETWORK (NN) is a popular tool to model nonlinear system as it does not require any predefined mathematical formulation for relationship between system inputs and targets [1], [2]. However, NN performance significantly drops in the presence of process disturbances and uncertainties [3], [4]. In a recent study, Khosravi et al. [5] reported that PI-based modelling technique is more reliable and effective than point forecasting to quantify the disturbances as well as uncertainty. Moreover, in contrast to the traditional point forecast-based NN, PI-based NN carries extra information such as the prediction accuracy. PI-based NN can be developed with a prescribed probability or predefined confidence level (CL) where CL relates to the prediction accuracy. The basic idea of NN-based PIs is shown in Fig. 1. Unlike tradition point-based NN, PI-based NN generates an interval (upper bound and lower bound) with a predefined CL. It is assumed that the target value should lie between the upper bound and lower bound.

There are several existing methods in literature to construct PIs. These include bootstrap [6], Bayesian [7], mean-variance estimation [8], and delta methods [9]. The main principle of these four methods to construct PIs is the same as they use traditional error-based cost function to train NNs. Moreover, the main strategy of all these methods is to minimize the prediction error, instead of trying to improve the PI quality (such as PI coverage probability and width)

Mohammad Anwar Hosen, Abbas Khosravi, Saeid Nahavandi and Douglas Creighton are with the Centre of Intelligent Systems Research (CISR), Deakin University, Australia (email: {ahosen, abbas.khosravi, saeid.nahavandi, douglas.creighton}@deakin.edu.au)

This research was fully supported by the Centre for Intelligent Systems Research (CISR) at Deakin University, Australia.

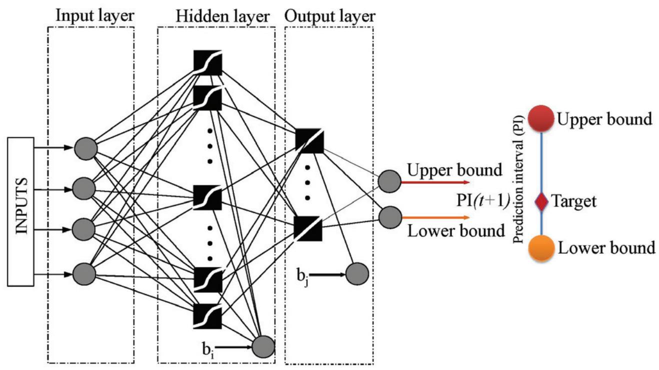

Fig. 1. Basic idea of PI-based NN

through changing parameters of NNs. Khosravi et al. [5] reported that the quality of PIs constructed in this way is questionable.

Recently another method, named lower upper bound estimation (LUBE) has been appeared in literature to construct PIs [5]. This method constructs quality PIs through optimizing NN structure and utilizing a PI-based cost function for a particular CL. The PI-based cost function includes two quality indexes of PIs, PI coverage probability ($PICP$) and width of prediction intervals. This method produces high quality PIs in terms of $PICP$ and PI width rather than other four methods that mentioned earlier.

It is argued that NN performance fluctuates from one replicate of training to other one, even when retraining with the same condition and the same data set. This is because the NN performance highly depends on its initial training parameters as well as perturbation of NN parameters. Literature reported that best trained NN is not always best for whole set or different data sets [10]. Recently, ensemble of NN has appeared as an additive tool to improve the prediction accuracy of NNs [11], [12]. An ensemble of NNs can greatly improve the overall representation accuracy, generalization, and robustness of NN predictions [13], [14]. The effects of a poor prediction from one NN in combined networks is simply minimized by effects of good predictions obtained from the other NNs [15]. In this technique, the forecast from several individual NNs are combined in a systematic way to get a united forecast.

The most popular forecast combination techniques are simple averaging and weighted averaging. In simple averaging mechanism, mean or median values of every sample instant are used. The main limitation of this method is that all NN ensemble members contribution is the same though they are not the same in terms of accuracy. In weighted averaging method, the weights (w) are assigned to each ensemble member based on their accuracy where their summation equal to one. In this case, the contribution of a better NN is more than a poor one. The application of weighted averaging mechanism in forecast combination method can be found in [16], [17].

Though there are vast application of ensemble technique for point forecast-based NNs, ensemble of PI-based NN is still limited. In recent years, Khosravi et al. [12] proposed a NN-based PI ensemble method using simple averaging of PIs generated from individual PI-NN models. It is the only existing method in literature to combine PI-based forecast. Firstly, they developed a couple of NN models and filter them based on their prediction performance. PIs are then constructed from those filtered NN models. Finally, constructed PIs are combined using simple averaging of PIs for every sample instance. They have shown that this simple averaging method can improve the quality of PIs in terms of PI width and $PICP$. It is also shown that the consistency of ensemble NN performance is better than individual ones. As described before, the main limitation of the simple averaging method is that it treats all individual NN ensemble members equally by ignoring their own prediction accuracy. The present work is the extended version of this paper where weighted averaging method is chosen to combine the NN forecasts for further improvement of PI quality.

II. NN-BASED PIs

LUBE method is an effective, reliable and fast method to generate quality PIs than other traditional methods [5]. In the present work, it is used to develop feedforward NN to construct quality PIs. NN is trained with a pre-defined confidence level. As there are no target values of PIs but the point values to train the NN, the quality of PIs is measured by the following indexes: (i) $PICP$, and (ii) width of PIs, known as PI normalized average width $(PINAW)$. A high $PICP$ with a low $PINAW$ means that the quality of PIs is good.

In contrast to traditional error-based cost functions, LUBE method uses a PI-based cost function to optimize the NN structure. PI-based cost function, known as the coverage width criterion (CWC) consists two quality indexes of PIs, $PICP$ and $PINAW$, and it is defined as:

$$CWC = PINAW\left(1+\gamma(PICP)e^{-\eta(PICP-\varphi)}\right) \quad (1)$$

where

$$\gamma = \begin{cases} 0, & PICP \geq \varphi \\ 1, & PICP \leq \varphi \end{cases}$$

Here η is a hyperparameter that scales up the PIs coverage error ($PICP$-CL) assigning more penalties undercoverage. φ corresponds to the nominal CL $(1-\alpha)$. The two objective parameters for CWC, $PICP$ and $PINAW$ are define as,

$$PICP = \frac{1}{n}\sum_{j=1}^{n} c_j \quad (2)$$

where

$$c_j = \begin{cases} 1, & t_j \in [L_j, U_j] \\ 0, & t_j \notin [L_j, U_j] \end{cases}$$

and

$$PINAW = \frac{1}{R}\left(\frac{1}{n}\sum_{j=1}^{n}(U_j - L_j)\right) \quad (3)$$

Here t_j, L_j and U_j are the target value, lower bound and upper bound for j^{th} sample, respectively. R is the range of the underlying target values.

As seen in (1), CWC not only increases the $PICP$ but also decreases the width of PIs leading to generating quality PIs. A simulated annealing optimization algorithm is used to optimize the NN structure through minimizing the CWC value. The detailed procedure for the LUBE method can be found in [5].

III. COMBINATION OF NN FORECASTS

The key contribution of this paper is to improve the LUBE PI quality through aggregating the PIs generated from the individual LUBE NN models. Many literatures works reported that the NN prediction accuracy can be improved by combining the forecasts of several NNs, even using a simple averaging method. In ensemble method, several networks $(1 \dots N)$ are trained for the same task (based on NN inputs and outputs). The structure of NN can be varied from one replicate to another one. Finally, the predicted output of each of these networks is combined to produce a single forecast. The proposed methodology of PI-based forecast combination follows three steps. This includes (i) development of NN models, (ii) choosing the ensemble members, and (iii) combining forecasts with an appropriate mechanism. The brief description of these three steps are given below.

A. Development of PI-based NN models

The very first step of the proposed ensemble method is randomly split the available sample data into the training (D_{train}), validation (D_{vald}) and test (D_{test}) data sets. LUBE method as described in Section II, is used to developed N_{total} NN-based PI models with a predefined CL utilizing PI-based cost function. Only the training data set is used in the NN training process. The rest of the data sets are reserved for examining the prediction performance of individual NNs and ensemble of NNs. Random initialization of NN parameters and different NN structures (by changing the number of neurons in the hidden layer) are chosen in the training process to diversify the NN models. SA optimization algorithm is employed to optimize the NN parameters through minimizing the cost function, CWC.

B. Selection of NN Ensemble Members

It is known that the prediction accuracy of all developed NNs with the same data set and the same procedure is not the same, as some of them produce very poor predictions [12]. Therefore N_{best} NNs are selected from the N_{total} NNs developed in the previous step. This is done by examining the prediction accuracy of NN models using D_{vald} data set. Total N_{total} PIs (PI$_{Ntotal}$) sets are constructed by using N_{total} LUBE NN models. N_{total} $PICP$ and $PINAW$ values are then determined by using PI$_{Ntotal}$ and calculate the N_{total} CWC (CWC_{Ntotal}). N_{total} NNs are then sorted based on CWC values. CWC is used here as a performance criterion of NN models. A NN with low CWC means that the prediction accuracy of NN model is better than a NN with large CWC. The NN with the minimum CWC is placed as rank one in terms of prediction accuracy. The NN models are sorted in ascending order of $CWC_{Ntotal,i}$, where $i = 1, 2, \ldots N_{total}$. Finally the first N_{best} NN (NN$_{Nbest}$) models are selected as ensemble members for the forecast combination process. The $PICP$ and CWC values for these N_{best} NNs are referred to as $PICP_{Nbest}$ and CWC_{Nbest} respectively.

C. Weighted Averaging of PI-based Forecasts

The PIs are generated for forecast combination from the N_{best} ensemble members. Total N_{best} sets of PIs (PI$_{test,em}$, where $em = 1, 2, \ldots N_{best}$) are constructed using D_{test} data set. This D_{test} data set is not used in the previous two steps. The $CWC_{test,em}$ are then calculated using (1-3) to check the prediction performance of individual NN ensemble members. The constructed PI$_{test,em}$ are then combined by using a weighted averaging mechanism. The general equation of weighted averaging combined PIs (PI$_{comb}$) can be defined as:

$$\text{PI}_{comb} = \sum_{em=1}^{N_{best}} w_{em} \text{PI}_{test,em} \tag{4}$$

where $em(em = 1, 2, \ldots N_{best})$ is the NN ensemble member with a ranking position and w_{em} is its corresponding weight. w_{em} is referred here to as ensemble parameters for easy understanding.

Traditionally, error-based cost functions, such as SSE, MSE and MAPE are used to optimize the weight in point-forecasting problems. However, there is no available literature works to optimize w_{em} for PI-based forecasting problems. As the key contribution of the present work, we use a PI-based cost function to obtain optimal w_{em}. The cost function CWC is utilized in the PI combination process. As mentioned before CWC covers both quality indexes of PIs, $PICP$ and width of PIs. It is expected that minimization of CWC through adjusting the ensemble parameters provides quality PIs (high coverage with low width).

CWC is nonlinear, nondifferentiable and complex as seen in (1). SA optimization algorithm is used to solve this cost function as this technique is suitable for this type of complex cost function [5]. Two constrains are defined in optimization process that allow positive contribution of all ensemble members. These include:

$$0 \leq w_{em} \leq 1 \text{ and } \sum_{r=1}^{N_{best}} w_{em} = 1$$

The initial ensemble parameters ($w_{initial,em}$) for optimization process are determined by using CWC values as CWC indicates the accuracy of NN model. CWC_{Nbest} values (using D_{vald} data set) obtained in the ensemble member selecting stage are used to calculate $w_{initial,em}$ as below:

$$w_{initial,em} = \frac{\frac{1}{CWC_{Nbest,em}}}{\sum_{em=1}^{N_{best}} \frac{1}{CWC_{Nbest,em}}} \tag{5}$$

In optimization, $w_{initial,em}$ is firstly set as the optimal ensemble parameters ($w_{opt,em}$). PI$_{comb}$ is calculated using (4). CWC_{opt} is then calculated with the help of (1-3). For every iteration, a new set of ensemble parameters ($w_{new,em}$) is generated through random perturbation of one of the current ensemble parameters. CWC_{new} is then determined. If $CWC_{new} \leq CWC_{opt}$, then $w_{opt,em}$ and CWC_{opt} are replaced with $w_{new,em}$ and CWC_{new} respectively; otherwise $w_{opt,em}$ remains unchanged. After completing the optimization process, the latest ensemble parameters ($w_{opt,em}$) are used to generate PI$_{comb}$ and, calculate and record $PICP_{comb}$, $PINAW_{comb}$ and CWC_{comb} to check the quality of PIs.

IV. Simulation Results

A. Case Studies

The performance of the proposed ensemble method is examined for two different case studies. A nonlinear model (time delay Mackey-Glass differential equation) is used as the first case study [18]. The Mackey-Glass equation can be defined as:

$$\frac{dx(t)}{dt} = \frac{0.2x(t-\tau)}{1 + x^{10}(t-\tau)} - \gamma x(t) \tag{6}$$

where τ is positive constant.

The parameter τ is set to 17. $x(0)$ is also set to 1.2 [19]. Total 1000 data samples are generated. Four lagged values are used as inputs to train the PI-NN model to predict $x(t)$. The inputs-output vector for NN training data is $[x(t-4), x(t-3), x(t-2), x(t-1); x(t)]$. The second case study is a nonlinear plant, where its output nonlinearly depends on both its past output values and the input values [20]. The plant difference model is given by,

$$y(t+1) = \frac{y(t)}{1 + y^2(t)} + u^3(t) \tag{7}$$

where

$$u(t) = \sin\left(\frac{2\Pi t}{100}\right) \tag{8}$$

$y(1)$ is set to 0.05. 500 samples are generated. Both $y(t)$ and $u(t)$ are used to predict $y(t+1)$. The input-output vector for NN training data is $[u(t), y(t), y(t+1)]$.

B. Data and parameters for proposed method

The NN models are developed using the LUBE method (with a 90% CL) for proposed forecasts aggregating method. The first step of the LUBE method is to prepare the training data. The collected data (case 1 and 2) for NN training are prepared by random sampling from the original data set. The prepared data are then split into training (50%), validation (30%) and testing (20%) data sets [21]. Only training data set

TABLE I. PARAMETERS USED IN LUBE METHOD TO TRAIN PI-NN MODEL

Parameters	values
Training data set (X_{train})	50% of total data
η	50
φ	0.90
T_{SA0}	5
Number of hidden layer	1
Number of neurons	6, 8, 10, 12, 14
Geometric cooling factor	0.9

is used to train LUBE NN models. Table I lists the parameters used in LUBE method. Five different NN structures are chosen in the training process to diversify the NN models. This is done by changing the number of neurons (Nu) in the hidden layer. Nu is set to 6, 8, 10, 12, and 14. This leads to $\frac{N_{total}}{5}$ NNs with a similar structure in each class.

SA optimization algorithm is employed to optimize the NN parameters by minimizing the cost function, CWC. The parameters for SA are chosen from the previous work of Khosravi et al. [5] as seen in Table I. A geometric cooling schedule with a cooling factor 0.9 is used for SA.

N_{total} and N_{best} are set to 150 and 10 respectively. It means that ten best NNs are selected as NN ensemble members out of 150 NNs. After selecting ensemble members, 10 sets of PIs are constructed by using D_{test} data set for all those ensemble members. These sets of PIs are then used for constructing combined PIs using the weighted averaging method as proposed in Section III.

C. Results and discussion

This section examine the quality of PI$_{comb}$, and compare the results with the PIs from individual ensemble members. The results of proposed weighted averaging forecast combination method is also compared with simple averaging forecast combination method proposed in [12]. Comparison is made in terms of the cost function, CWC, and coverage probability ($PICP$).

Firstly, the performance of the proposed ensemble method is compared with individual best NNs and simple averaging method in terms of the CWC. CWC indirectly declares the quality of PIs as it covers both quality indexes of PIs ($PICP$ and width of PIs). The smaller the CWC value, the better the PI quality. Fig. 2 shows the CWC values for the best ten individual NN ensemble members and two combined forecasts (simple averaging and weighted averaging) methods. For ease of reference, CWC values for simple averaging and weighted averaging with SA are referred to as $CWC_{comb,simple}$ and $CWC_{comb,SA}$ respectively. As can be seen in Fig. 2 (a and b), CWC values for the forecast aggregation methods are lower than CWC values for the individual NNs for both case studies. The order of CWC values is $CWC_{comb,SA} < CWC_{comb,simple} < CWC_{Nbest}$. This means that proposed weighted averaging forecast combination method produces high quality PIs than individual NNs and simple averaging ensemble method. The $CWC_{comb,SA}$ values for the first and second case studies are 14.68 and 12.78 respectively that are significantly lower than their corresponding mean values of $CWC_{Nbest,i}$ (20.97 and 303.80 for the first and the second case studies). The main interesting phenomena is that $CWC_{comb,SA}$ is lower than any $CWC_{Nbest,i}$ ($i = 1, 2, \ldots N_{best} = 10$) for both case studies. This indicates that the proposed method significantly improves the quality of PIs compared to individual NNs as well as simple averaging method.

Now, we examine the quality of combined PIs in terms of their coverage probability, $PICP$. In training process, the NN models are developed using LUBE method with a CL 90%. This means that PIs should cover at least 90% of target values ($PICP \geq 90\%$). Fig. 3 demonstrated the $PICP$ values for PI$_{Nbest,i}$, PI$_{comb,simple}$ and PI$_{comb,SA}$. It can be seen that $PICP_{comb,SA}$ value is 90% for both case studies which is satisfied with the predefined CL where the $PICP$ values for individual NNs fluctuate from 89.5-93% and 81-89% for the first and second cases, respectively. However, simple averaging ensemble method covers a higher percentage of the target value ($PICP_{comb,simple} = 94$ and 92% in the first and second cases, respectively) than individual NNs and proposed method. Though $PICP_{comb,SA}$ is less than $PICP_{comb,simple}$, weighted averaging forecast aggregation method constructs high quality PIs in terms of CWC as $CWC_{comb,SA}$ is lowers (see Fig. 2). A high $PICP$ with a large CWC value indicates that PIs are too wide and accordingly less informative. Note that the predefined CL is 90%. It is not practically desirable to have too wide intervals as they do not carry much information about the target values and their fluctuations.

It is noticeable that all $PICP_{Nbest,i}$ values are lower than nominal CL, 90% for the second case study. This is highly likely attributable to dissimilar patterns in training, validation and testing data sets, and small data size. However, forecast combination methods eliminate this unfortunate circumstance as PIs for these methods cover at least 90% of target values. This means that forecast combination method produces high quality PIs, even when it is a simple averaging method.

The improvement of the PI quality of individual NNs through the proposed aggregation method is presented in Fig. 4. The improvements ($I_{m,em}$) of individual NNs (in terms of the CWC) are calculated by the following equation

$$I_{m,em} = \frac{CWC_{N_{best,em}} - CWC_{comb,SA}}{CWC_{N_{best,em}}} \tag{9}$$

Fig. 4 depicts that the proposed ensemble method improves the PI quality on an average of 18% and 76% for the first and second case studies, respectively. Significant improvement can be observed for the second case as PIs is improved at least 70% in 80% (8 out of 10) of ensemble members. Fig. 4 also demonstrates that the proposed weighted averaging method improves the PI quality by 3% and 4% for the first and second case studies compared to simple averaging method proposed in [12].

In summary, weighted averaging aggregation method through minimizing PI-based cost function produces quality PIs better than individual NNs. This method also improves the PI quality at least 3% compared to the simple averaging ensemble method. Therefore, it is reasonable to conclude that proposed method constructs higher quality PIs than individual LUBE NN models and any other existing PI-based forecast ensemble methods.

V. CONCLUSIONS

In this paper, a new NN ensemble method is proposed to improve the quality of PIs that generated using the LUBE method. LUBE method is a reliable, fast and effective method to develop NN model for constructing quality PIs. Weighted averaging forecast aggregation mechanism is employed in the NN ensemble method. In contrast to traditional error-based cost function, a PI-based cost function, CWC is introduced to optimize the ensemble parameters (weights). A global optimization algorithm, named SA, is used to minimize the PI-based cost function. Finally, the proposed ensemble method is examined for two case studies.

Simulation results demonstrate that proposed method improves the LUBE PI quality on an average of 25% and 95% for the first and second case studies, respectively. Improvement of PI quality is measured in terms of CWC as this covers both quality indexes of PIs (width and coverage probability). The proposed aggregated PIs also improves the simple averaging PI combination method by at least 3% for both case studies. Further improvement of the proposed method can be achieved by determining the optimal structure of NN models.

REFERENCES

[1] M. A. Hosen, M. A. Hussain, and F. S. Mjalli, "Hybrid modelling and kinetic estimation for polystyrene batch reactor using artificial neutral network (ann) approach," *Asia-Pacific Journal of Chemical Engineering*, vol. 6, no. 2, pp. 274–287, 2011.

[2] M. A. Hosen, M. A. Hussain, F. S. Mjalli, A. Khosravi, D. Creighton, and S. Nahavandi, "Performance analysis of three advanced controllers for polymerization batch reactor: An experimental investigation," *Chemical Engineering Research and Design*, no. 0, p. Article in Press, 2013. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S026387621300333X

[3] A. Khosravi, S. Nahavandi, D. Creighton, and A. F. Atiya, "Comprehensive review of neural network-based prediction intervals and new advances," *Neural Networks, IEEE Transactions on*, vol. 22, no. 9, pp. 1341–1356, 2011.

[4] M. A. Hosen, A. Khosravi, S. Nahavandi, and D. Creighton, "Prediction interval-based neural network modelling of polystyrene polymerization reactor a new perspective of data-based modelling," *Chemical Engineering Research and Design*, no. 0, p. Article in Press, 2014. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S0263876214001063

[5] A. Khosravi, S. Nahavandi, D. Creighton, and A. Atiya, "Lower upper bound estimation method for construction of neural network-based prediction intervals," *Neural Networks, IEEE Transactions on*, vol. 22, no. 3, pp. 337–346, 2011.

[6] T. Heskes, "Practical confidence and prediction intervals," *Advances in Neural Information Processing Systems*, pp. 176–182, 1997.

[7] C. M. Bishop, *Neural Networks for Pattern Recognition*. Oxford university press, 1995.

[8] D. A. Nix and A. S. Weigend, "Estimating the mean and variance of the target probability distribution," in *Neural Networks, 1994. IEEE World Congress on Computational Intelligence., 1994 IEEE International Conference on*, vol. 1. IEEE, 1994, pp. 55–60.

[9] J. G. Hwang and A. A. Ding, "Prediction intervals for artificial neural networks," *Journal of the American Statistical Association*, vol. 92, no. 438, pp. 748–757, 1997.

[10] R. Furtuna, S. Curteanu, and F. Leon, "Multi-objective optimization of a stacked neural network using an evolutionary hyper-heuristic," *Applied Soft Computing*, vol. 12, no. 1, pp. 133–144, 2012.

[11] D. West, S. Dellana, and J. Qian, "Neural network ensemble strategies for financial decision applications," *Computers & Operations Research*, vol. 32, no. 10, pp. 2543 – 2559, 2005.

[12] A. Khosravi and S. Nahavandi, "Combined nonparametric prediction intervals for wind power generation," *Sustainable Energy, IEEE Transactions on*, vol. 4, no. 4, pp. 849–856, 2013.

[13] Q. ZHU, N. ZHAO, and Y. XU, "A new selective neural network ensemble method based on error vectorization and its application in high-density polyethylene (hdpe) cascade reaction process," *Chinese Journal of Chemical Engineering*, vol. 20, no. 6, pp. 1142–1147, 2012.

[14] J. Zhang, "Inferential estimation of polymer quality using bootstrap aggregated neural networks," *Neural Networks*, vol. 12, no. 6, pp. 927–938, 1999.

[15] Y. Tian, J. Zhang, and J. Morris, "Modeling and optimal control of a batch polymerization reactor using a hybrid stacked recurrent neural network model," *Industrial & engineering chemistry research*, vol. 40, no. 21, pp. 4525–4535, 2001.

[16] H. Song, S. F. Witt, K. F. Wong, and D. C. Wu, "An empirical study of forecast combination in tourism," *Journal of Hospitality & Tourism Research*, vol. 33, no. 1, pp. 3–29, 2009.

[17] B. E. Hansen, "Least-squares forecast averaging," *Journal of Econometrics*, vol. 146, no. 2, pp. 342 – 350, 2008.

[18] C.-F. Juang, R.-B. Huang, and W.-Y. Cheng, "An interval type-2 fuzzy-neural network with support-vector regression for noisy regression problems," *Fuzzy Systems, IEEE Transactions on*, vol. 18, no. 4, pp. 686–699, 2010.

[19] A. Khosravi, S. Nahavandi, and R. Khosravi, "Evaluation and comparison of type reduction algorithms from a forecast accuracy perspective," in *Fuzzy Systems (FUZZ), 2013 IEEE International Conference on*, 2013, pp. 1–7.

[20] C.-F. Juang and Y.-W. Tsao, "A self-evolving interval type-2 fuzzy neural network with online structure and parameter learning," *Fuzzy Systems, IEEE Transactions on*, vol. 16, no. 6, pp. 1411–1424, 2008.

[21] M. A. Hosen, M. A. Hussain, and F. S. Mjalli, "Control of polystyrene batch reactors using neural network based model predictive control (nnmpc): An experimental investigation," *Control Engineering Practice*, vol. 19, no. 5, pp. 454–467, 2011.

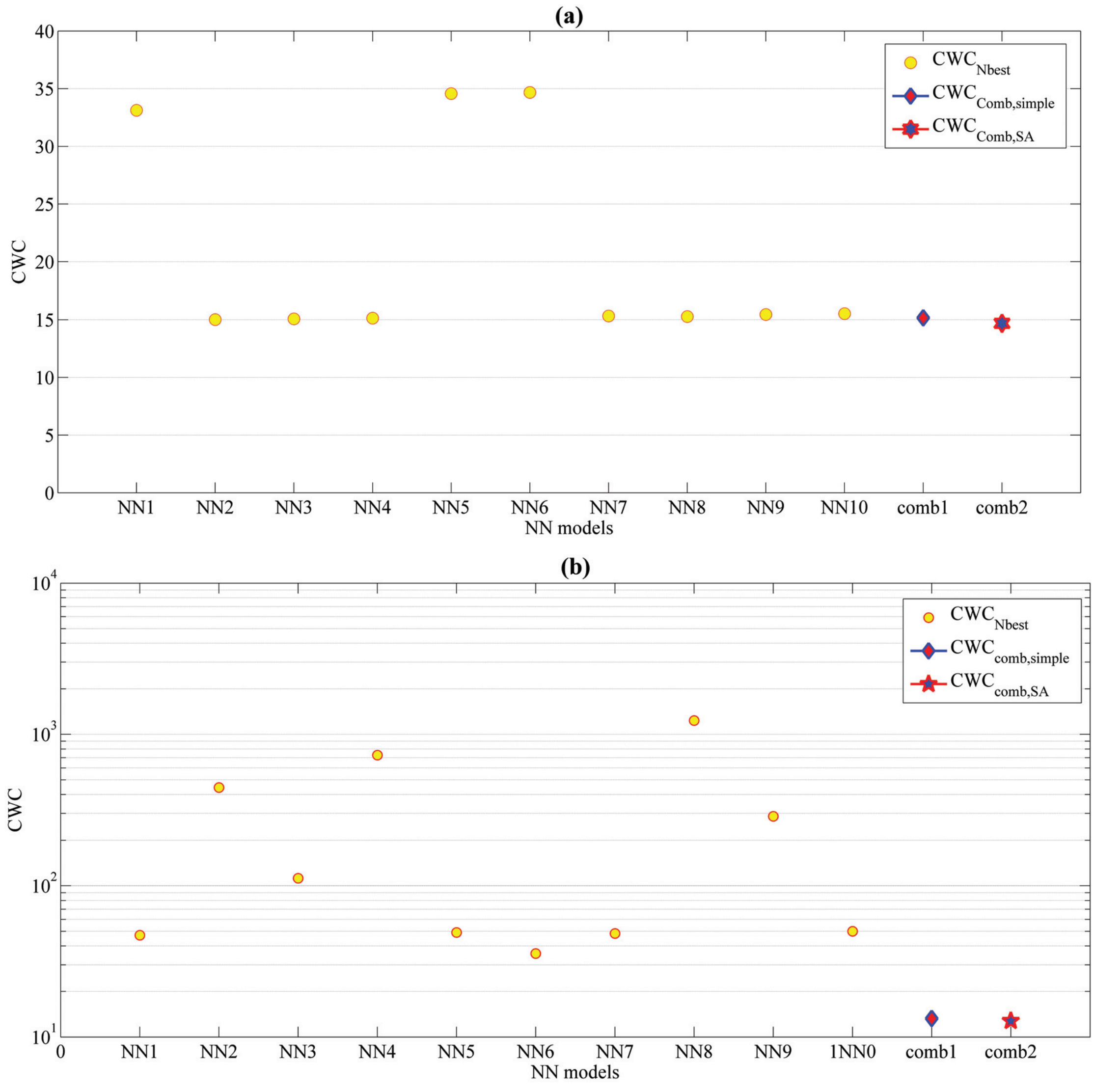

Fig. 2. CWC for best ensemble members and combined PIs, (a) for the first case study, and (b) for the second case study

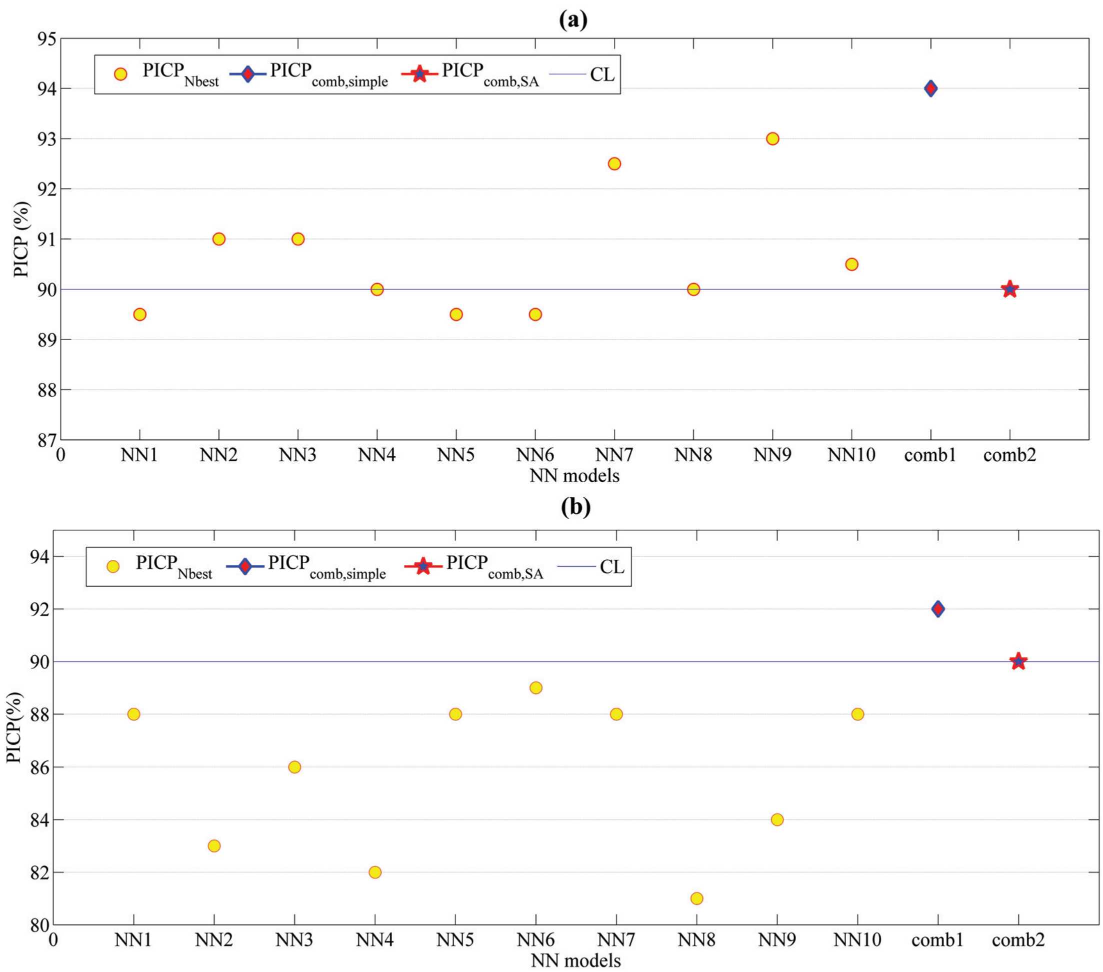

Fig. 3. PICP for best ensemble members and combined PIs, (a) for the first case study, and (b) for the second case study

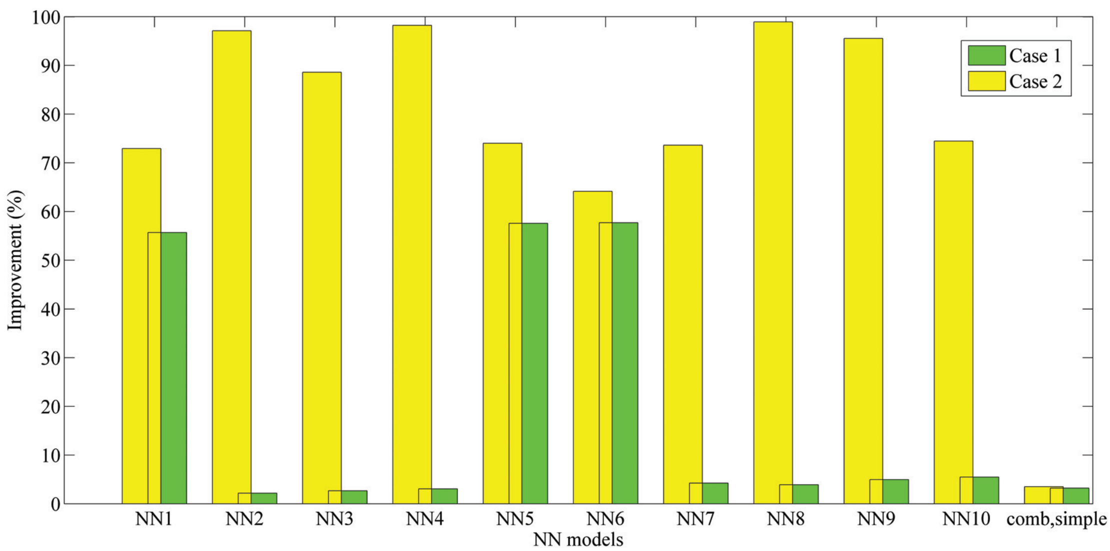

Fig. 4. Improvement of individual NN ensemble members and simple averaging aggregation method for both case studies