

Uncertainty Quantification for Wind Farm Power Generation

Abbas Khosravi, Saeid Nahavandi, Doug Creighton, Reihaneh Naghavi-zadeh

Centre for Intelligent Systems Research

Deakin University, Geelong, VIC, 3217, Australia

Email: {abbas.khosravi,saeid.nahavandi,douglas.creighton@deakin.edu.au} and naghavizade_r@yahoo.com

Abstract—Accurate forecasting of wind farm power generation is essential for successful operation and management of wind farms and to minimize risks associated with their integration into energy systems. However, due to the inherent wind intermittency, wind power forecasts are highly prone to error and often far from being perfect. **The purpose of this paper is to develop statistical methods for quantifying uncertainties associated with wind power generation forecasts. Prediction intervals (PIs) with a prescribed confidence level are constructed using the delta and bootstrap methods for neural network forecasts.** The moving block bootstrap method is applied to preserve the correlation structure in wind power observations. The effectiveness and efficiency of these two methods for uncertainty quantification is examined using two month datasets taken from a wind farm in Australia. It is demonstrated that while all constructed PIs are theoretically valid, bootstrap PIs are more informative than delta PIs, and are therefore more useful for decision-making.

Index Terms—wind energy, uncertainty, prediction intervals, neural networks.

I. INTRODUCTION

As the fastest growing source of renewable energy, wind energy is changing its role from being a marginal energy source to a major primary energy source. In fact, it is becoming a mainstream source of electricity supply in many developed countries worldwide, in particular in Europe. However, widespread penetration of wind power into energy system has some negative consequences that must be mitigated. Wind energy is stochastic in its nature. In contrast to the traditional electricity industry that relies on storable primary energies, wind power is non-storable and intermittent. The intermittency of wind energy and its large variability imparts a high level of uncertainties into the electricity generation, transmission, and distribution. These impose several difficulties and challenges on an independent system operator charged with retaining the balance and security of the network.

Accurate prediction and forecast of wind power generation is critical for safe and smooth operation of energy systems with a high penetration of wind energy. Wind power forecasting methods can be broadly divided into two groups: physical methods and data-driven (also called statistical) methods [1]. Meteorological data and physical laws governing atmospheric behavior are often used in the first group to develop a model for estimating the wind speed and direction. Numerical weather prediction and mesoscale models are examples of physical models [2] [3].

Statistical methods use historical wind power datasets and some explanatory variables to build models. A variety of techniques has been proposed in literature for linear and nonlinear model development. These include, but not limited to, ARMA [4], neural networks (NNs) [5], recurrent NNs [6], neuro-fuzzy systems [7], and support vector machines [8]. Hybrid approaches aggregating forecasts generated by methods from both groups have also been proposed in recent years [9], [10]. Compared with physical methods, data-driven models are usually simpler, less expert-expensive, and more suitable for small farms.

Research on wind power forecasting has been mainly driven by a best guess strategy. **The best forecast is often generated by an individual model or an ensemble and then used for decision-making.** Despite all the work carried out in literature, it is already known that wind power generation forecasts are always prone to error [11]. The fact is that forecasting errors always exist regardless of the employed forecasting method and the implemented procedures. This is mainly due to a high level of uncertainty associated with operation of wind farms. According to this, it is critical to consider and quantify to relevant uncertainties and wisely use them in the process of decision-making related to wind farms [12] [13].

Uncertainties associated with power generation forecasts can be appropriately quantified through construction of prediction intervals (PIs). PIs take into account both the uncertainty in model structure and noise in input data. They allow the users of the produced point predictions to perceive the accuracy of the predictions, and thus to make informed decisions to keep or reject the result. Instead of computing different quantiles and pairing them to construct forecasts intervals [14], this paper takes a more direct approach towards construction of PIs for wind power forecasts generated by NN models. **This paper focus on construction of PIs, rather than confidence intervals (CIs),** makes it distinct from existing wind forecasting literature. Feedforward NN models are developed and used for forecasting aggregated power generation in wind farms. Then, the delta and moving block bootstrap methods are applied to quantify uncertainties associated with these forecasts. Two confidence levels are considered for construction of PIs. Quantitative indices and measures are used in experiments for comprehensive and unbiased evaluation of the quality of PIs. Experiments are conducted with summer and winter datasets from two wind farms in Australia. Each experiment

is repeated several times and averaged results are used for drawing conclusions. Constructed PIs and obtained results are discussed from different perspectives such as the method efficiency, effects of the confidence level, and PI validity and informativeness.

The rest of this paper is organized as follows. Delta and bootstrap methods for construction of NN-based PIs are briefly introduced in Section II. Section III describes performance measures for assessing the quality of PIs. Simulation results are demonstrated in Section IV. Section V concludes the paper with some remarks and guidelines for future work.

II. PI CONSTRUCTION METHODS

By definition, a PI with an associated confidence level of $(1 - \alpha)\%$ is a random interval developed based on past observations, $X = (x_1, x_2, \dots, x_n)$ for future observations, $PI = [L(X), U(X)]$, such that, $Pr(L(X) < x_{n+1} < U(X)) = 1 - \alpha$. Before proceeding, it is important to emphasize the difference between a CI and a PI. While a CI describes the uncertainty in the prediction of an unknown but fixed value, a PI deals with the uncertainty in the prediction of a future realization of a random variable [15]. A PI accounts for more sources of uncertainty and is wider than the corresponding CI [16]. Many authors wrongly interchange these terminologies [17]. Similarly, only CIs are constructed in the work reported in [11], where indeed PIs are required for decision-making in real world applications.

A. Delta Method

The underpinning principles of the delta method have their roots in theories of nonlinear regression [18]. It is based on representing and interpreting a NN as a nonlinear regression model,

$$y = f(X, w^*) + \epsilon \quad (1)$$

where X and y are respectively the set of inputs (m independent variables) and the corresponding target (dependent variable). $f(\cdot)$ with the parameter set w^* is the nonlinear function representing the true regression function. ϵ is the noise with a zero expectation. The estimated NN parameters are represented by \hat{w} . It is obtained through minimizing the sum of squared error (SSE) cost function,

$$SSE = (y - \hat{y})^T (y - \hat{y}). \quad (2)$$

The NN parameter estimation problem can be viewed as a nonlinear regression problem, using the cost function defined in (2). Therefore, the standard methods for constructing the asymptotic PIs for nonlinear regression models are applicable. In the neighborhood of the true set of NN parameters, w^* , $f(X, w)$ can be approximated by a first-order Taylor series,

$$\hat{y} = f(X, w^*) + J(\hat{w} - w^*) \quad (3)$$

where the i, j -th element in the matrix J is $\frac{\partial f_i}{\partial w_j}$. These derivatives are calculated for the true set of parameters, w^* .

For the sample x_0 , the linear model is given by,

$$\hat{y}_0 = f(x_0, w^*) + g_0^T (\hat{w} - w^*) \quad (4)$$

where g_0^T is a vector whose j th entry is $\frac{\partial f(x_0)}{\partial w_j}$. With the assumption that ϵ in (1) are independently and normally distributed ($N(0, \sigma^2)$), the $(1 - \alpha)\%$ PI for \hat{y}_i is,

$$\hat{y}_0 \pm t_{df}^{1-\frac{\alpha}{2}} s \sqrt{1 + g_0^T (J^T J)^{-1} g_0}. \quad (5)$$

$t_{df}^{1-\frac{\alpha}{2}}$ is the $\frac{\alpha}{2}$ quantile of a cumulative t-distribution function with df degrees of freedom. df here is the difference between number of training samples (n) and number of NN parameters (p). J is also the Jacobian matrix of the NN model with respect to its parameters. s is an unbiased estimation of σ given below,

$$s^2 = \frac{SSE}{n - p}. \quad (6)$$

In some cases, matrix $J^T J$ may become singular. Consequently, constructed PIs using (5) are unreliable. To remedy this problem, one alternative is to include some regularizing terms in (2). Based on the weight decay concept [19], De Veaux et al. [20] added the sum of squares of adaptive parameters to the cost function in (2),

$$SSE + \lambda \hat{w}^T \hat{w} \quad (7)$$

where λ is the regularizing factor [19]. The generalization power of NN models is improved through using (7) as the training cost function. Rebuilding PIs based on (7) will yield the PIs shown in equation (8).

The procedure for calculation of s in (8) has small differences with the one used in (5). These differences arise from the corresponding cost function considered in each method. An acceptable estimate of s in (8) is given below,

$$s = \frac{\sqrt{SSE}}{n - \text{trace}(2\Gamma - \Gamma^2)} \quad (9)$$

where $\Gamma = J(J^T J + \lambda I)^{-1} J^T$. Inclusion of λ in (8) fixes the singularity problem resulting in better quality PIs. More discussion on these issues can be found in [20].

B. The Moving Block Bootstrap Method

The moving block bootstrap (MBB) method extends the standard bootstrap method [21] to the case that observations are correlated. To keep the temporal dependence in data, sequential observations are randomly resampled in blocks, rather than single observations. Mathematically, let $X_{l,i} = (x_i, \dots, x_{i+l-1})$, $i = 1, \dots, n - l + 1$, be the $n - l + 1$ overlapping blocks. The MBB method randomly with replacement selects r overlapping blocks $X_{l,i}^*$, $i = 1, \dots, r$. Then sampled blocks, $X_{l,i}^*$, are put together to form a time series of length n . Unless $\frac{n}{l}$ is an integer, one or more of the blocks have to be truncated to form a sample of length n . The whole procedure is repeated B times to form the B datasets, $\{D\}_{b=1}^B$, required for training of B models.

$$\hat{y}_0 \pm t_d^{1-\frac{\alpha}{2}} s \sqrt{1 + g_0^T (J^T J + \lambda I)^{-1} (J^T J) (J^T J + \lambda I)^{-1} g_0} \quad (8)$$

The bootstrap method estimates the variance due to model misspecification, $\sigma_{\hat{y}}^2$, by building B NN models. According to this assumption, the true regression is estimated by averaging the point forecasts of B models, $\hat{y}_i = \frac{1}{B} \sum_{b=1}^B \hat{y}_i^b$, where \hat{y}_i^b is the prediction of the i -th sample generated by the b -th bootstrap model. Assuming that NN models are unbiased, the model misspecification variance can be estimated using the variance of B model outcomes,

$$\sigma_{\hat{y}_i}^2 = \frac{1}{B-1} \sum_{b=1}^B (\hat{y}_i^b - \hat{y}_i)^2. \quad (10)$$

For the construction of PIs, the variance of errors, $\sigma_{\hat{\epsilon}_i}^2$, should be also estimated. $\sigma_{\hat{\epsilon}_i}^2$ can be calculated as follows,

$$\sigma_{\hat{\epsilon}_i}^2 \simeq E\{(t - \hat{y})^2\} - \sigma_{\hat{y}}^2. \quad (11)$$

According to (11), a set of variance squared residuals is developed,

$$r_i^2 = \max((t_i - \hat{y}_i)^2 - \sigma_{\hat{y}_i}^2, 0) \quad (12)$$

where \hat{y}_i and $\sigma_{\hat{y}_i}^2$ both are available from B trained NN models and (10). These residuals are linked by the set of corresponding inputs to form a new dataset,

$$D_{r^2} = \{(x_i, r_i^2)\}_{i=1}^n. \quad (13)$$

A new NN model, called $NN_{\sigma_{\hat{\epsilon}}}$, can be trained using the maximum likelihood estimation method to estimate the unknown values of $\sigma_{\hat{\epsilon}_i}^2$, so as to maximize the probability of observing the samples in D_{r^2} [16].

In total, $B + 1$ NN models are required in the bootstrap method for PI construction. Simplicity is the main advantage of using the bootstrap method for PI construction.

III. PERFORMANCE INDEX FOR PREDICTION INTERVALS

Three indices here are used for quantitative evaluation of the quality of PIs. The first index is PI coverage probability (PICP), which measures what portion of targets is covered by constructed intervals,

$$PICP = \frac{1}{n} \sum_{t=1}^n c_t \quad (14)$$

where, $c_t = 1$ if $y_t \in [L_t, U_t]$, otherwise $c_t = 0$. L_t and U_t are the lower and upper bounds of the t -th PI respectively.

The second measure is PI normalized averaged width (PINAW), that measures how wide PIs are on average,

$$PINAW = \frac{1}{nR} \sum_{t=1}^n (U_t - L_t) \quad (15)$$

where R is the range of the underlying target.

A combinational coverage width-based criterion (CWC) is used for simultaneous evaluation of the quality of constructed PIs from both informativeness (measured by PINAW) and validity (measured by PICP),

$$CWC = PINAW \left(1 + \gamma(PICP) e^{-\eta(PICP - \mu)}\right) \quad (16)$$

$\gamma(PICP)$ is given by the following step function,

$$\gamma = \begin{cases} 0 & PICP \geq \mu \\ 1 & PICP < \mu \end{cases} \quad (17)$$

η and μ in (16) are two hyperparameters controlling the location and amount of CWC jump. These measures can be easily determined based on the level of confidence associated with PIs. μ corresponds to the nominal confidence level associated with PIs and can be set to $1 - \alpha$. Further information about these measures and hyperparameters η and μ can be found in [22], [23], [24], [26], [28] [29].

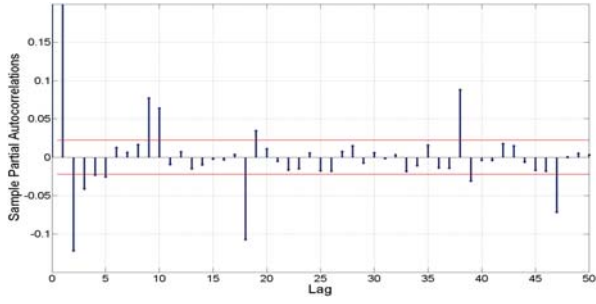
IV. CASE STUDIES AND RESULTS

A. Data and Experiment Procedure

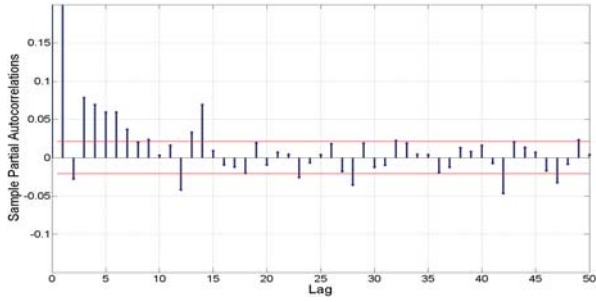
February and August 2010 datasets from Woolnorth wind farm are considered in this study for construction of PIs. These two months respectively correspond to summer and winter seasons in Australia in southern hemisphere. Located in northwest of Tasmania, Australia, the site is an ideal location for wind farm operation because of the steady and consistent winds. As a short term forecasting problem, we consider a five minute look-ahead time for forecasting wind power generation in this wind farm.

Autocorrelation and partial correlation functions are applied for analysis and selection of appropriate lagged values for model development. Up to fifty lagged wind power generation values are considered in correlation studies, and the peak values are selected as the set of inputs for each wind farm model in each month.

Single layer NNs with ten neurons in the hidden layer are used in this study. $NN_{\sigma_{\hat{\epsilon}}}$ in the bootstrap method has also a similar structure. One hundred datasets are formed using the moving block bootstrap methods. The block length (l) is set to fifty, the same as the number of lagged values used in correlation analysis. Generated datasets are then used for training of NN models in the bootstrap method ($B = 100$). To avoid any subjective judgment, experiments are repeated ten times and PIs are examined each time using PICP, PINAW, and CWC. PIs are constructed for three different levels of confidence: 70%, 80%, and 90%. These correspond to setting α to 0.3, 0.2, and 0.1, respectively. μ is set to $(1 - \alpha)\%$, which is the confidence level. Also, η is selected to be 50 in order to highly penalize PIs with a coverage probability lower than the



(a) February 2010



(b) August 2010

Fig. 1. Partial autocorrelation with 50 lags for two month datasets.

corresponding confidence level. Also, all datasets are preprocessed to have zero mean and unit variance. Hereafter and for notational simplicity, we use the following abbreviations and acronyms as subscripts: 'Feb' as February, 'Aug' as August, 'D' as delta, and 'MBB' as moving block bootstrap.

B. Results and Discussion

Fig. 1 shows the partial autocorrelations for fifty lags of the wind farm output for February and August 2010. The lagged outputs corresponding to the peak values of partial autocorrelation coefficients are considered as the vector of inputs to NN models. This results in 11 and 14 inputs to NNs for these two months respectively. It is obvious that the most recent lagged outputs form the major part of the input vector. After the tenth lagged output, the partial autocorrelation values significantly drop indicating a weak dependency.

Table I and II report values of PICP, PINAW, and CWC for ten replicates of each experiment. These results and the quality of PIs can be studied from different perspectives with different purposes: validity, effects of confidence level, seasons, and performance of methods.

PIs are theoretically valid when their PICP is greater than or equal to the corresponding confidence level. Violation of this requirement results in invalidity of PIs and ruins their credibility. Checking PICPs in Table I and II indicates that constructed PIs using both the delta and bootstrap methods satisfy this condition, and therefore are valid. In all cases, there is a positive margin between PICP and the associated confidence level highlighting the capability of both methods for producing reliable PIs, even in the presence of a high level of uncertainty in data and operation of wind farms. There

TABLE I
PICP, PINAW, AND CWC FOR PIS CONSTRUCTED WITH THREE
CONFIDENCE LEVELS FOR FEBRUARY 2010 DATASET

Confidence Level	Moving Block Bootstrap			Delta		
	PICP	PINAW	CWC	PICP	PINAW	CWC
70%	79.04	19.94	19.94	86.71	25.75	25.75
	77.60	19.42	19.42	83.41	24.41	24.41
	79.60	19.70	19.70	86.96	25.64	25.64
	77.48	19.18	19.18	87.84	25.76	25.76
	78.92	19.98	19.98	85.59	25.30	25.30
	78.10	19.53	19.53	86.28	25.51	25.51
	78.10	19.20	19.20	85.78	25.60	25.60
	77.92	19.25	19.25	87.09	25.95	25.95
	78.79	19.44	19.44	87.21	25.67	25.67
	77.36	19.27	19.27	87.40	25.77	25.77
80%	87.40	24.90	24.90	91.20	31.24	31.24
	86.84	24.41	24.41	91.52	31.39	31.39
	87.34	24.08	24.08	92.27	31.36	31.36
	86.78	24.48	24.48	92.08	31.65	31.65
	86.40	24.61	24.61	91.89	31.37	31.37
	86.65	24.25	24.25	92.02	31.48	31.48
	87.46	25.32	25.32	90.46	31.43	31.43
	85.40	23.72	23.72	91.70	31.63	31.63
	86.15	23.79	23.79	89.83	31.12	31.12
	86.15	23.89	23.89	92.08	31.64	31.64
90%	93.14	30.39	30.39	95.88	40.05	40.05
	93.45	31.30	31.30	95.63	40.60	40.60
	92.70	31.13	31.13	95.57	40.88	40.88
	92.02	30.83	30.83	96.20	41.14	41.14
	93.08	31.10	31.10	95.01	40.81	40.81
	92.45	30.50	30.50	95.20	40.46	40.46
	93.26	31.34	31.34	95.76	40.79	40.79
	92.20	30.93	30.93	94.57	39.90	39.90
	92.14	30.39	30.39	96.51	41.58	41.58
	92.45	30.16	30.16	96.38	40.92	40.92

TABLE II
PICP, PINAW, AND CWC FOR PIS CONSTRUCTED WITH THREE
CONFIDENCE LEVELS FOR AUGUST 2010 DATASET

Confidence Level	Moving Block Bootstrap			Delta		
	PICP	PINAW	CWC	PICP	PINAW	CWC
70%	85.98	10.97	10.97	82.04	10.07	10.07
	84.46	10.35	10.35	82.60	9.88	9.88
	85.53	10.49	10.49	81.53	10.07	10.07
	85.92	10.50	10.50	82.26	9.93	9.93
	85.87	11.15	11.15	82.94	9.91	9.91
	85.47	10.89	10.89	83.00	9.72	9.72
	86.49	10.89	10.89	81.03	9.65	9.65
	84.57	10.27	10.27	82.43	9.75	9.75
	85.98	10.74	10.74	82.83	9.91	9.91
	84.29	10.30	10.30	81.08	9.97	9.97
80%	90.20	13.60	13.60	86.71	12.00	12.00
	89.98	12.89	12.89	86.71	12.27	12.27
	91.33	14.30	14.30	87.16	12.70	12.70
	91.10	14.33	14.33	86.60	11.97	11.97
	91.55	14.92	14.92	85.81	12.43	12.43
	89.13	12.82	12.82	87.44	12.11	12.11
	88.68	12.74	12.74	86.37	12.34	12.34
	89.19	12.69	12.69	86.77	12.42	12.42
	89.58	12.43	12.43	88.01	12.78	12.78
	89.36	12.65	12.65	85.14	11.57	11.57
90%	95.38	20.88	20.88	90.54	15.69	15.69
	92.74	16.01	16.01	91.44	15.41	15.41
	93.53	16.88	16.88	91.95	16.17	16.17
	93.41	16.90	16.90	90.32	15.91	15.91
	93.30	16.02	16.02	92.17	16.40	16.40
	94.71	17.28	17.28	90.37	15.45	15.45
	93.69	16.87	16.87	89.47	15.85	15.85
	92.17	15.16	15.16	90.99	15.70	15.70
	93.13	16.45	16.45	89.13	15.81	15.81
	94.09	17.48	17.48	86.21	15.13	15.13

are only three cases in $PI_{D,Aug,90\%}$ where PICP is slightly lower than 90%. However, this is not a big issue, because (i) the gap is small, and (ii) the averaged value of PICPs in ten experiments is 90.38%, which is greater than the associated confidence level.

When the confidence level is 70%, the minimum value for $PICP_D$ and $PICP_{MBB}$ is 77.36%. However, this is 85.14% for the case of an 80% confidence level, and 90.32% for the case of a 90% confidence level (only checking PICPs greater than 90%). The positive margin decrement in these cases signifies that achieving a theoretically satisfactory coverage probability becomes more problematic as the confidence level goes up.

Both delta and bootstrap methods respond similarly to the requirement of a greater confidence level. To achieve the validity requirement, they widen PIs to cover more uncertainties in data. According to the results in Table I and II, for each dataset and each method, we have $PINAW_{90\%} \geq PINAW_{80\%} \geq PINAW_{70\%}$. This simply means that the greater the confidence level, the wider the PIs.

To graphically demonstrate this behavior, we plot $PI_{D,Feb,70\%}$, $PI_{D,Feb,80\%}$, and $PI_{D,Feb,90\%}$. Fig. 2 displays PI_D for test samples 400-1200 with three confidence levels. In this specific experiment, $PICP_{D,70\%} = 85.78\%$, $PICP_{D,80\%} = 90.46\%$, and $PICP_{D,90\%} = 95.76\%$. Therefore, PIs are valid and reliable. However, $PI_{D,70\%}$ and $PICP_{D,80\%}$ are covered by $PI_{D,90\%}$ as they are narrower for the majority of samples. This signifies that an extra cost must be paid in the form of wider PIs to satisfy the condition of a greater confidence level.

Comparing CWCs in Table I and II reveals that seasons have a significant impact on the quality of PIs. In all conducted experiments with three confidence levels and using both methods, $CWC_{Feb} > CWC_{Aug}$. This means that the quality of PIs constructed for August datasets is better than the quality of PIs constructed for corresponding February datasets. This is only due to wider PIs constructed for wind farm datasets for February 2010. The large values of CWC in February (summer time in Australia) imply that there are more uncertainties associated with operation of wind farms in summer than winter.

It is also important to compare performance of the delta and MBB methods for construction of PIs. The averaged values of CWCs for all experiments are shown in Table III. According to these results, CWC_{MBB} is smaller than CWC_D is 4 out of 6 case studies, and in 2 out of 6 cases, it is bigger. The issue is that when $CWC_{MBB} < CWC_D$, the gap is big and varies between 6-11. In contrast, when $CWC_{MBB} > CWC_D$, the gap is less 1. These figures indicate that PI_{MBB} have a better quality than PI_D .

As both PI_{MBB} and PI_D are theoretically valid, the superiority of PI_{MBB} stems from their informativeness aspect. From a practical point of view, narrow PIs carry more information than wide PIs (under the validity assumption). Therefore, PI_{MBB} are more useful and informative for decision-making than PI_D .

Obtained results in this paper can be used in both theo-

TABLE III
AVERAGED CWC VALUES FOR BOTH DATASETS AND THREE CONFIDENCE LEVELS

Dataset	Confidence Level	CWC_{MBB}	CWC_D
February	70%	19.49	25.54
	80%	24.34	31.43
	90%	30.81	40.71
August	70%	10.66	9.88
	80%	13.34	12.26
	90%	16.99	30.35

retical and practical studies of wind power generation. From a scientific point of view, other statistical models, such as support vector machines or neuro-fuzzy systems, can be hired for producing quality prediction intervals and uncertainty quantification. Also, studies can be conducted to enhance the quality of prediction intervals in terms of their informativeness (narrowing process) without compromising their validity ($PICP \geq (1 - \alpha)\%$). From the practical point of view, constructed intervals can be efficiently used by end users of wind power forecast. As a complementary source of information, they can be used along with point predictions for decision-making. As uncertainties associated with each forecast are quantified and reflected in the width of intervals, decision-makers will have a signal of the risk they bear when using point forecasts. Some preliminary studies in this field have started to appear in literature [12].

The quality of PIs and results in Table I and II can be improved in several ways. One approach is to diversify the set of NN inputs and feed NN models with various sorts of valuable information. Further research can also be done in this direction by applying more advanced methods for feature selection. Results can be further improved for both bootstrap and delta methods by applying their optimized versions proposed in [30] and [27].

V. CONCLUSIONS

The problem of uncertainty quantification through construction of PIs for the case of wind farm power generation was investigated in this paper. The delta and moving block bootstrap methods were applied to build PIs for wind farm power generation forecasts obtained by neural network models. The quality of PIs was quantitatively examined based on their width and coverage probability. Comprehensive experiments with real datasets revealed that the quality of power generation PIs greatly varies due to effects of the confidence level, seasons, and performance of methods. It was also found that while all constructed PIs are valid, the moving block bootstrap PIs are more informative than the delta PIs, and are therefore more useful for decision-making.

ACKNOWLEDGMENT

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

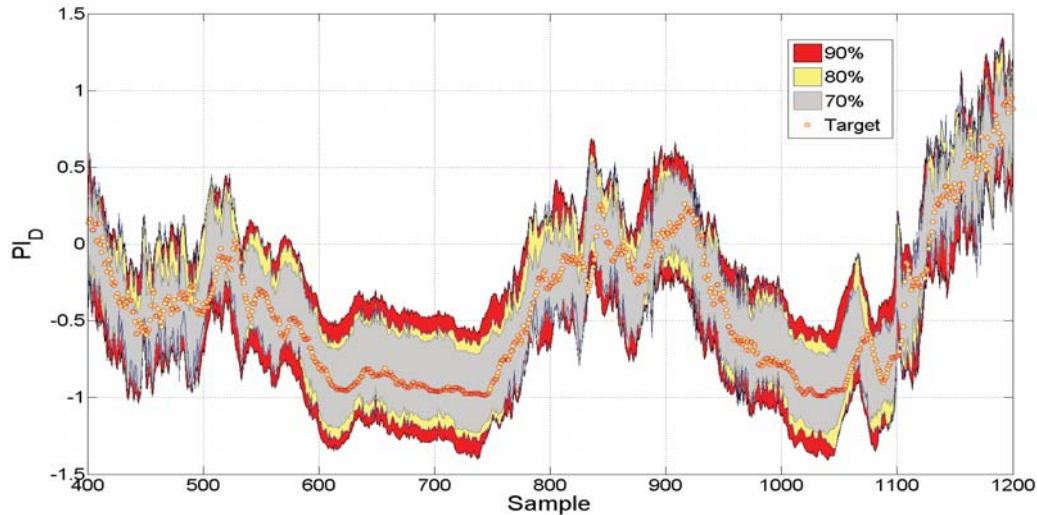


Fig. 2. $PI_{D, Feb}$ with three confidence levels: 70%, 80%, and 90%.

REFERENCES

- [1] A. M. Foley, P. G. Leahy, A. Marvuglia, and E. J. McKeogh, "Current methods and advances in forecasting of wind power generation," *Renewable Energy*, vol. 37, no. 1, pp. 1–8, Jan. 2012.
- [2] L. Landberg, G. Giebel, H. A. Nielsen, T. Nielsen, and H. Madsen, "Short-term predictionan overview," *Wind Energy*, vol. 6, no. 3, pp. 273–280, 2003.
- [3] M. Lange and U. Focken, *Physical Approach to Short-term Wind Power Prediction*. Springer Verlag, Heidelberg, 2005.
- [4] J. Torres, A. Garca, M. De Blas, and A. De Francisco, "Forecast of hourly average wind speed with arma models in navarre (Spain)," *Solar Energy*, vol. 79, no. 1, pp. 65–77, Jul. 2005.
- [5] G. Kariniotakis, G. Stavrakakis, and E. Nogaret, "Wind power forecasting using advanced neural networks models," *IEEE Transactions on Energy Conversion*, vol. 11, no. 4, pp. 762–767, 1996.
- [6] T. Barbounis, J. Theocharis, M. Alexiadis, and P. Dokopoulos, "Long-term wind speed and power forecasting using local recurrent neural network models," *IEEE Transactions on Energy Conversion*, vol. 21, no. 1, pp. 273–284, 2006.
- [7] T. Barbounis and J. Theocharis, "A locally recurrent fuzzy neural network with application to the wind speed prediction using spatial correlation," *Neurocomputing*, vol. 70, no. 7-9, pp. 1525–1542, 2007.
- [8] E. G. Ortiz-Garcia, S. Salcedo-Sanz, A. M. Perez-Bellido, J. Gascon-Moreno, J. A. Portilla-Figueras, and L. Prieto, "Short-term wind speed prediction in wind farms based on banks of support vector machines," *Wind Energy*, vol. 14, no. 2, pp. 193–207, 2011.
- [9] F. Molteni, R. Buizza, T. N. Palmer, and T. Petroligis, "The ecmwf ensemble prediction system: methodology and validation," *Quarterly Journal of the Royal Meteorological Society*, vol. 122, pp. 73–119, 1996.
- [10] P. Pinson, H. Nielsen, H. Madsen, and G. Kariniotakis, "Skill forecasting from ensemble predictions of wind power," *Applied Energy*, vol. 86, no. 7-8, pp. 1326–1334, 2009.
- [11] P. Pinson and G. Kariniotakis, "Conditional prediction intervals of wind power generation," *IEEE Transactions on Power Systems*, vol. 25, no. 4, pp. 1845–1856, 2010.
- [12] K. Methaprayoon, C. Yingvivatanapong, W.-J. Lee, and J. Liao, "An integration of ann wind power estimation into unit commitment considering the forecasting uncertainty," *IEEE Transactions on Industry Applications*, vol. 43, no. 6, pp. 1441–1448, 2007.
- [13] M. Shaheen, "Integration of wind resources into utility systems," *National Wind Coordinating Committee, Wind Energy Issue Brief*, 1997.
- [14] J. B. Bremnes, "Probabilistic wind power forecasts using local quantile regression," *Wind Energy*, vol. 7, no. 1, pp. 47–54, 2004.
- [15] N. Meade and T. Islam, "Prediction intervals for growth curve forecasts," *Journal of Forecasting*, vol. 14, pp. 413–430, 1995.
- [16] T. Heskes, "Practical confidence and prediction intervals," in *Neural Information Processing Systems*, T. P. M. Mozer, M. Jordan, Ed., vol. 9. MIT Press, 1997, pp. 176–182.
- [17] J. G. De Gooijer and R. J. Hyndman, "25 years of time series forecasting," *International Journal of Forecasting*, vol. 22, no. 3, pp. 443–473, 2006.
- [18] W. C. J. Seber, G. A. F., *Nonlinear Regression*. New York: Wiley, 1989.
- [19] C. M. Bishop, *Neural Networks for Pattern Recognition*. Oxford University, Press, Oxford, 1995.
- [20] R. D. d. Veaux, J. Schumi, J. Schweinsberg, and L. H. Ungar, "Prediction intervals for neural networks via nonlinear regression," *Technometrics*, vol. 40, no. 4, pp. 273–282, 1998.
- [21] B. Efron and R. J. Tibshirani, *An Introduction to the Bootstrap*. New York: Chapman and Hall, 1993.
- [22] A. Khosravi, S. Nahavandi, D. Creighton, and A. F. Atiya, "A lower upper bound estimation method for construction of neural network-based prediction intervals," *IEEE Transactions on Neural Networks*, vol. 22, no. 3, pp. 337–346, 2011.
- [23] A. Khosravi, E. Mazlumi, S. Nahavandi, D. Creighton, and J. W. C. van Lint, "Prediction intervals to account for uncertainties in travel time prediction," *IEEE Transactions on Intelligent Transportation Systems*, vol. 12, no. 2, pp. 537–547, 2011.
- [24] A. Khosravi, E. Mazlumi, S. Nahavandi, D. Creighton, and J. Van Lint, "A genetic algorithm-based method for improving quality of travel time prediction intervals," *Transportation Research Part C: Emerging Technologies*, vol. 19, no. 6, pp. 1364–1376, Dec. 2011.
- [25] A. Khosravi, S. Nahavandi, and D. Creighton, "A prediction interval-based approach to determine optimal structures of neural network metamodells," *Expert Systems with Applications*, vol. 37, pp. 2377–2387, 2010.
- [26] —, "Load forecasting and neural networks: A prediction interval-based perspective," *B.K. Panigrahi et al. (Eds.): Computational Intelligence in Power Engineering, SCI 302*, pp. 131–150, 2010.
- [27] —, "Construction of optimal prediction intervals for load forecasting problem," *IEEE Transactions on Power Systems*, vol. 25, pp. 1496–1503, 2010.
- [28] —, "Prediction interval construction and optimization for adaptive neurofuzzy inference systems," *IEEE Transactions on Fuzzy Systems*, vol. 19, no. 5, pp. 983–988, 2011.
- [29] A. Khosravi, S. Nahavandi, D. Creighton, and A. Atiya, "Comprehensive review of neural network-based prediction intervals and new advances," *IEEE Transactions on Neural Networks*, vol. 22, no. 9, pp. 1341–1356, 2011.
- [30] A. Khosravi, S. Nahavandi, D. Creighton, and D. Srinivasan, "Optimizing the quality of bootstrap-based prediction intervals," in *The 2011 International Joint Conference on Neural Networks (IJCNN)*, 2011, pp. 3072–3078.