

A linear hybrid methodology for improving accuracy of time series forecasting

Ratnadip Adhikari · R. K. Agrawal

Received: 11 May 2013 / Accepted: 22 August 2013 / Published online: 11 September 2013
© Springer-Verlag London 2013

Abstract Modeling and forecasting of time series data are integral parts of many scientific and engineering applications. Increasing precision of the performed forecasts is highly desirable but a difficult task, facing a number of mathematical as well as decision-making challenges. This paper presents a novel approach for linearly combining multiple models in order to improve time series forecasting accuracy. Our approach is based on the assumption that each future observation of a time series is a linear combination of the arithmetic mean and median of the forecasts from all participated models together with a random noise. The proposed ensemble is constructed with five different forecasting models and is tested on six real-world time series. Obtained results demonstrate that the forecasting accuracies are significantly improved through our combination mechanism. A nonparametric statistical analysis is also carried out to show the superior forecasting performances of the proposed ensemble scheme over the individual models as well as a number of other forecast combination techniques.

Keywords Time series · Forecast combination · Box-Jenkins models · Artificial neural networks · Elman networks · Support vector machines

1 Introduction

Many natural as well as synthetic phenomena can be expressed in terms of time series which are sequential

collections of observations measured over successive times. Analysis and forecasting of time series data have fundamental importance in various scientific and engineering applications. As such, improving forecasting accuracies is a matter of constant attentions of researchers [1]. Various linear and nonlinear forecasting models are the outcomes of extensive works in this area during the past three decades [1–3]. However, due to the stochastic nature of a time series, it is evident that no single model alone can capture all the intrinsic details of the associated data-generating process [4]. Hence, it is quite risky and inappropriate to rely upon only one of the available models for forecasting future data. On the other hand, combining forecasts from conceptually different models is a reliable approach of decreasing the model selection risk while at the same time improving overall forecasting precisions. Moreover, the accuracies obtained through combining forecasts are often better than any individual model in isolation [4–6].

Intrigued by their strengths and benefits, many forecast combination algorithms have been developed during the last two decades. At present, selection of the most promising combination scheme for a particular application is itself a nontrivial task [5]. Here, it is worth mentioning that a combination of multiple forecasts attempts to enhance the forecasting precisions at the expense of increased computational complexity. Obviously, a balanced trade-off between the accuracy improvement and associated computational cost is highly desirable from an ideal combination scheme. An extensive body of the literature has shown that simple methods for combining often provide much better accuracies than more complicated and sophisticated techniques [6, 7]. Thus, for combining forecasts, one should use straightforward tools and avoid intricacies.

The simple average in which all component forecasts are weighted equally is so far the most elementary, yet

R. Adhikari (✉) · R. K. Agrawal
School of Computer and Systems Sciences,
Jawaharlal Nehru University, New Delhi 110067, India
e-mail: adhikari.ratan@gmail.com

R. K. Agrawal
e-mail: rkajnu@gmail.com

widely popular combination method. Together with being easy to understand, implement, and interpret, it is also computationally inexpensive. A well-known fact is that unequal or judgmental weighting often suffers from miscalculations, biasness, and estimation errors [8]. From this viewpoint, a simple average is quite robust as it does not require any estimation of combining weights or other parameters. Moreover, several research evidences in the literature have shown that the naïve simple average in many occasions produced remarkably better forecasting results than various other intricate combination tools [1, 6, 7]. Due to these salient features, simple averages have been extensively used for combining time series forecasts. However, they are highly sensitive to extreme values, and so sometimes, there are considerable extent of variation and instability among the obtained forecasts [7, 8]. As a remedial measure, numerous studies have suggested the use of the median which is far less sensitive to extreme values than the simple average. But there are varied results regarding forecasting superiority of the simple average and median. The former produced better accuracies in the work of Stock and Watson [9], worse in the works of Larreche and Moinpour [10] and Agnew [11], and about the same in the work of McNees [12]. From these studies, it is not possible to rationally differentiate between the performances of these two statistical averages. But evidently, both simple average and median can achieve significantly improved forecasting accuracies that too at a very few computational costs.

In this paper, we propose a novel linear ensemble method that attempts to take advantage of the strengths of both simple average and median for combining time series forecasts. Our proposed method is based on the assumption that each future observation of a time series is a linear combination of the arithmetic mean and median of the individual forecasts, together with a random noise. Five different forecasting models are combined through the proposed mechanism that is then tested on six real-world time series datasets. A nonparametric statistical analysis is also carried out to compare the accuracies of our ensemble scheme with those of the individual models as well as other benchmark forecast combination techniques.

The rest of the paper is organized as follows. Section 2 describes the related works on combining time series forecasts. The details of our proposed combination scheme are described in Sect. 3. Section 4 presents a concise discussion on the five individual forecasting models, which are used to build up the proposed ensemble. The empirical results are reported in Sect. 5, and finally, Sect. 6 concludes this paper.

2 Related works

Combining multiple methods in scientific applications has a long history that dates back to early eighteenth century [6]. The notable use of models combining for time series forecasting started in late 1960s with important contributions from Crane and Crotty [13], Zarnowitz [14], and Reid [15]. However, the seminal work of Bates and Granger [16] was the first to introduce a general analytical model for effectively combining multiple forecasts. Till then, several forecast combination mechanisms have been developed in the literature.

The constrained *Ordinary Least Square (OLS)* method is one of the earlier tools for linear combination of forecasts. It determines the combining weights through solving a *Quadratic Programming Problem (QPP)* that minimizes the *Sum of Squared Error (SSE)* error between the original and forecasted datasets with the restriction that the weights are nonnegative and unbiased [17–19]. An alternative is the *Least Square Regression (LSR)* method [1, 4, 5, 19] that does not impose any restriction on the combining weights and often provides better forecasting accuracies than the constrained OLS method.

A familiar fact is that the weights of a linear combination of forecasts can be optimized with the knowledge of the covariance matrix of one-step-ahead forecast errors. As the covariance matrix is unknown in advance, Newbold and Granger [20] suggested five procedures for estimating the weights from the known data and these are commonly known as the *differential weighting schemes*. Winkler and Makridakis [21] performed an extensive empirical analysis of these five methods and found that two of them provided better forecasting results than the others.

In order to cope with the dynamic nature of a time series, a forecast combination algorithm should be able to recursively update the weights with additions of new data values. As such, a number of recursive combination schemes have also been developed in the literature, which include the *Recursive Least Square (RLS)* technique and its variants, such as the *Dynamic RLS*, *Covariance Addition RLS (RLS-CA)*, *Kalman filter (KF)*, etc. [22]. These algorithms are often reported to be more efficient than the fixed weighting methods [1, 6, 22].

The *Outperformance* method by Bunn [23] is another effective linear combination technique that is based on the Bayesian probabilistic approach. It determines how likely a component model outperforms the others. This method assigns subjective weight to a component model on the basis of the number of times it performed best in the past [5, 23].

Implicit combinations of two or more forecasting models are also developed by time series researchers.

One such benchmark technique, due to Zhang [24], adopts a hybridization of autoregressive integrated moving average (ARIMA) [2, 3, 24] and artificial neural network (ANN) models [2, 4, 24]. In this hybrid mechanism, the linear correlation structure of the time series is modeled through ARIMA, and then, the remaining residuals, which contain only nonlinear part, are modeled through ANN. With three real-world datasets, Zhang [24] showed that his hybrid scheme provided reasonably improved accuracies and also outperformed each component model. Recently, Khashei and Bijari [25, 26] further explored this approach, thereby suggesting a similar but slightly modified and more robust combination method.

Forecast combinations through nonlinear techniques are analyzed as well in the time series literature, but to a limited extent. This is mainly due to the lack of recognized studies which document success of such schemes [5]. Adhikari and Agrawal [27] recently developed a nonlinear-weighted-ensemble technique that considers both the individual forecasts as well as the correlation among pairs of forecasts. Their scheme was able to provide reasonably enhanced forecasting accuracies for three popular time series datasets.

In spite of several improved techniques, the robustness and efficiency of elementary statistical combination methods are always appreciated in the literature [1]. Many empirical evidences show that the naïve simple average notably outperformed more complex ensemble schemes [7, 28]. A robust alternative to the simple average is the trimmed mean in which forecasts are averaged by excluding an equal percentage of highest and lowest forecasts [5, 7, 8]. In a recent comprehensive study, Jose and Winkler [7] have found that trimmed means were able to provide slightly more accurate results than the simple averages and reduced the risk of high errors. But hitherto, there is no rigorous method for selecting the exact amount of trimming. The median (i.e., the ultimate trimming) has also been studied as an alternative to the simple average with varied results [7–12]. Thus, it seems to be advantageous to adequately combine both simple average and median.

3 The proposed combination method

3.1 Formulation of the proposed combination method

Let, $\mathbf{Y} = [y_1, y_2, \dots, y_N]^T \in \mathbb{R}^N$ be the actual testing dataset of a time series, which is forecasted using n different models, and $\hat{\mathbf{Y}}^{(i)} = [\hat{y}_1^{(i)}, \hat{y}_2^{(i)}, \dots, \hat{y}_N^{(i)}]^T$ be the i th forecast of \mathbf{Y} ($i = 1, 2, \dots, n$). Let, u_j and v_j , respectively, be the

mean and median of $\{\hat{y}_j^{(1)}, \hat{y}_j^{(2)}, \dots, \hat{y}_j^{(n)}\}$, $j = 1, 2, \dots, N$.

Then, the proposed combined forecast of \mathbf{Y} is defined as $\hat{\mathbf{Y}} = [\hat{y}_1, \hat{y}_2, \dots, \hat{y}_N]^T$, where

$$\hat{y}_j = \begin{cases} v_j, & \text{for } \alpha = 0 \\ u_j, & \text{for } \alpha = 1 \\ \alpha u_j + (1 - \alpha)v_j + \varepsilon_j, & \text{for } 0 < \alpha < 1 \end{cases} \quad (1)$$

$$\varepsilon_j \sim N(0, \sigma^2) \quad \forall j = 1, 2, \dots, N.$$

In Eq. 1, $\{\varepsilon_j | j = 1, 2, \dots, N\}$ is assumed to be a white noise process, i.e., a sequence of independent, identically distributed (i.i.d.) random variables, which follow the typical normal distribution with zero mean and a constant variance σ^2 . These white noise terms are introduced as the trade-offs between the accuracy improvement and proper combination of the two averages. The parameter α manages the contributions of the two averages in the final combined forecast. The median is more dominating for $0 \leq \alpha < 0.5$, whereas the simple average is more dominating for $0.5 < \alpha \leq 1$. The proposed ensemble scheme can be written in the vector form as follows:

$$\hat{\mathbf{Y}} = \alpha \mathbf{U} + (1 - \alpha)\mathbf{V} + \mathbf{E} \quad (2)$$

$$0 \leq \alpha \leq 1$$

where $\mathbf{U} = [u_1, u_2, \dots, u_N]^T$, $\mathbf{V} = [v_1, v_2, \dots, v_N]^T$ are, respectively, the vectors of means and medians, and $\mathbf{E} = [\varepsilon_1, \varepsilon_2, \dots, \varepsilon_N]^T$ is the vector of the white noise terms.

3.2 Selections of the tuning parameters

Our proposed scheme combines the simple average and median of individual forecasts in an unbiased manner. The success of the scheme solely depends on the suitable selection of the parameters α and σ . Here, we suggest effective techniques for selecting these parameters.

For selecting α , first, we consider one of the two ranges: $0 \leq \alpha < 0.5$ (*median dominating*) and $0.5 < \alpha \leq 1$ (*mean dominating*). In either of these ranges, we vary α in an arithmetic progression with a certain step size (i.e., common difference) s . The desired value of α is then taken to be the mean of its values in this particular range and is denoted as α^* . The precise mathematical formulation of α^* is given as follows:

$$\alpha^* = \frac{1}{N_s} \sum_{i=1}^{N_s} [l + (i - 1)s] \quad (3)$$

where $N_s = \lceil \frac{0.5}{s} \rceil$ and $l = 0, 0.5$ in the ranges $0 \leq \alpha < 0.5$ and $0.5 < \alpha \leq 1$, respectively.

It is obvious that smaller values of the step size s ensure better estimation of the final combined forecast. The value $s = 0.01$ is used for all empirical works in this paper.

The nature of the white noise terms has a crucial impact on the success of our combination scheme, and so the noise variance σ^2 must have to be chosen with utmost care. The value of σ^2 is closely related to the deviation between the simple average and median of the forecasts. Keeping this fact in mind, we suggest choosing σ^2 as the variance of the difference between the mean and median vectors, i.e.,

$$\sigma^2 = \text{var}(\mathbf{U} - \mathbf{V}) \quad (4)$$

Equation 4 provides a rational as well as robust method for selecting the noise variance in our combination scheme. After choosing the two tuning parameters α and σ through Eqs. 3 and 4, the combined forecast vector is given by

$$\hat{\mathbf{Y}} = \alpha^* \mathbf{U} + (1 - \alpha^*) \mathbf{V} + \mathbf{E} \quad (5)$$

The requisite steps of our combination method are concisely summarized in Algorithm 1.

Algorithm 1 The proposed linear combination of multiple forecasts

Inputs: The individual forecasts $\hat{\mathbf{Y}}^{(i)}$ ($i = 1, 2, \dots, n$) of \mathbf{Y}

Output: The combined forecast $\hat{\mathbf{Y}}$

1. Compute the mean vector \mathbf{U} and median vector \mathbf{V} from the individual forecasts
2. Select a certain step size s and a range for α from either $0 \leq \alpha < 0.5$ or $0.5 < \alpha \leq 1$
3. Compute the values of the parameters α^* and σ^2 through Eqs. 3 and 4, respectively
4. Generate N i.i.d. random variables $\varepsilon_j \sim \mathcal{N}(0, \sigma^2)$, ($j = 1, 2, \dots, N$) and define the vector $\mathbf{E} = [\varepsilon_1, \varepsilon_2, \dots, \varepsilon_N]^T$
5. Finally, calculate the proposed combined forecast vector through Eq. 5

A schematic depiction of our proposed combination method is presented in Fig. 1.

3.3 Salient features of our proposed combination mechanism

1. The primary advantage of the proposed scheme is that it reduces the overall forecasting error in more precise manner than various other ensemble mechanisms. In our method, a large extent of error is already reduced through simple average and median of the individual forecasts. Then, combining these averages results in further reduction in the forecasting error. Hence, the proposed methodology is evidently better than directly combining the forecasts from the individual models.
2. The proposed method benefits from the forecasting skills of diverse constituent models, unlike some others, which combines only a few particular ones. For example, Zhang [24] suggested a hybridization of two models, viz. ARIMA and ANN. Similarly, Tseng

et al. [29] suggested the ensemble scheme SARIMABP which combines seasonal ARIMA (SARIMA) and backpropagation ANN (BP-ANN) models for seasonal time series forecasting. However, in many situations, our linear ensemble method can be apparently more accurate as it combines a large number of competing forecasting models.

3. Considering a wide pool of available forecast combination schemes, nowadays, a major challenge faced by the time series research community is to select the most appropriate method of combining forecasts. Our proposed mechanism improves the forecasting accuracy as well as diminishes the model selection risk to a great extent. The formulation of our method suggests that it apparently performs much better than both simple average and median. As such, the proposed scheme provides a potentially good choice in the domain of time series forecast combination.
4. Our proposed scheme is notably simple and much more computationally efficient than various existing combination methods. This is due to the fact that many sophisticated methods, e.g., RLS, dynamic RLS, NRLS, outperformance, etc., require repeated in-sample applications of the constituent forecasting models, thus entailing large amount of computational times. On the contrary, our proposed method applies the component models only once and, hence, saves a lot of associated computations.

4 The component forecasting models

The effectiveness of a forecast combination mechanism depends a lot on the constituent models. Several studies document that for a good combination scheme, the component forecasting models are essentially to be as diverse and competent as possible [4, 8, 21]. Armstrong [8] in his extensive study on combining forecasts further emphasized on using four to five constituent models for achieving maximum combined accuracy. Based on these studies and recommendations, here, we use the following five diverse forecasting models to build up our proposed ensemble:

- *The Autoregressive Integrated Moving Average (AR-IMA)* [2, 3].
- *The Support Vector Machine (SVM)* [30].
- *The iterated Feedforward Artificial Neural Network (FANN)* [2, 25, 32, 33].
- *The iterated Elman ANN (EANN)* [5, 34].
- *The direct EANN* [5, 34].

In the forthcoming subsections, we concisely describe these five forecasting techniques.

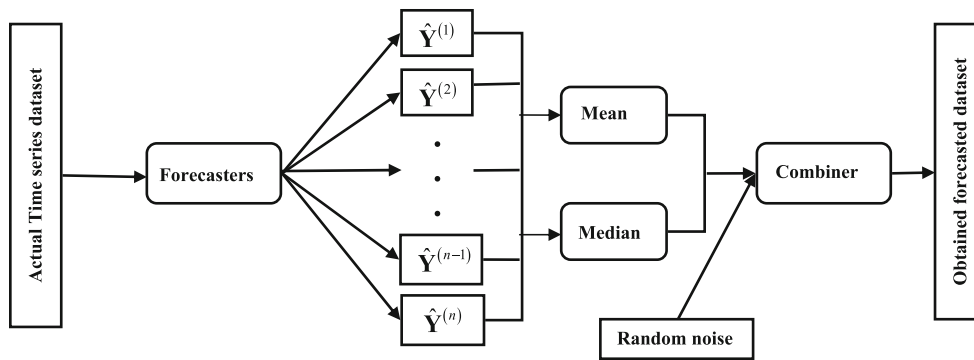


Fig. 1 The proposed forecast combination mechanism

4.1 The ARIMA model

ARIMA models are the most extensively used statistical techniques for time series forecasting. These are developed in the benchmark work of Box and Jenkins [3] and are also commonly known as the Box-Jenkins models. The underlying hypothesis of these models is that the associated time series is generated from a linear combination of predefined numbers of past observations and random error terms. A typical ARIMA model is mathematically given by

$$\phi(L)(1 - L)^d y_t = \theta(L)\varepsilon_t \tag{6}$$

where,

$$\phi(L) = 1 - \sum_{i=1}^p \phi_i L^i, \quad \theta(L) = 1 + \sum_{j=1}^q \theta_j L^j, \quad L y_t = y_{t-1}.$$

The parameters $p, d,$ and q are, respectively, the number of *autoregressive, degree of differencing,* and *moving average* terms; y_t is the actual time series observation, and ε_t is a white noise term. The white noise terms are basically i.i.d. normal variables with zero mean and a constant variance. It is customary to refer the model, represented through Eq. 6 as the ARIMA (p, d, q) model. This model effectively converts a nonstationary time series to a stationary one through a series of ordinary differencing processes. A single differencing is enough for most applications. The appropriate ARIMA model parameters are usually determined through the well-known Box-Jenkins three-step iterative model building methodology [3, 24].

The ARIMA (0, 1, 0), i.e., $y_t - y_{t-1} = \varepsilon_t$ is in particular known as the random walk (RW) model and is commonly used for modeling nonstationary data [24]. Box and Jenkins [3] further generalized the basic ARIMA model to forecast seasonal time series, as well, and this extended model is referred as the seasonal ARIMA (SARIMA). A SARIMA (p, d, q) \times (P, D, Q)^s model adopts an additional *seasonal differencing* process to remove the effect of seasonality from the dataset. Like ARIMA (p, d, q), the parameters

(p, P), (q, Q), and (d, D) of a SARIMA model represent the autoregressive, moving average, and differencing terms, respectively, and s denote the period of seasonality.

4.2 The SVM model

SVM is a relatively recent statistical learning theory, originally developed by Vapnik [30]. It is based on the principle of *Structural Risk Minimization (SRM)* whose objective is to find a decision rule with good generalization ability through selecting some special-training data points, viz. the *support vectors* [30, 31]. Time series forecasting is a branch of support vector regression (SVR) problems in which an optimal separating hyperplane is constructed to correctly classify real-valued outputs. But the explicit knowledge of this mapping is avoided through the use of a kernel function that satisfies the Mercer’s condition [31].

Given a training dataset of N points $\{\mathbf{x}_i, y_i\}_{i=1}^N$ with $\mathbf{x}_i \in \mathbb{R}^n, y_i \in \mathbb{R}$, the goal of SVM is to approximate the unknown data-generating function in the following form:

$$f(\mathbf{x}, \mathbf{w}) = \mathbf{w} \cdot \varphi(\mathbf{x}) + b \tag{7}$$

where \mathbf{w} is the weight vector, \mathbf{x} is the input vector, φ is the nonlinear mapping to a higher dimensional feature space, and b is the bias term.

Using Vapnik’s ε -insensitive loss function, given in Eq. 8, the SVM regression is converted to a quadratic programming problem (QPP) to minimize the empirical risk, as defined in Eq. 9.

$$L_\varepsilon(y, f(\mathbf{x}, \mathbf{w})) = \begin{cases} 0, & \text{if } |y - f(\mathbf{x}, \mathbf{w})| \leq \varepsilon \\ |y - f(\mathbf{x}, \mathbf{w})| - \varepsilon, & \text{otherwise} \end{cases} \tag{8}$$

$$J(\mathbf{w}, \xi_i, \zeta_i^*) = \frac{1}{2} \|\mathbf{w}\|^2 + C \sum_{i=1}^N (\xi_i + \zeta_i^*). \tag{9}$$

In Eq. 9, C is the positive regularization constant that assigns a penalty to misfit, and ξ_i, ζ_i^* are the nonnegative slack variables.

After solving the associated QPP, the optimal decision hyperplane is given by

$$y(\mathbf{x}) = \sum_{i=1}^{N_s} (\alpha_i - \alpha_i^*) K(\mathbf{x}, \mathbf{x}_i) + b_{\text{opt}}. \quad (10)$$

where N_s is the number of support vectors, α_i and α_i^* ($i = 1, 2, \dots, N_s$) are the Lagrange multipliers, b_{opt} is the optimal bias, and $K(\mathbf{x}, \mathbf{x}_i)$ is the kernel function.

From several choices for the SVM kernel function, the *Radial Basis Function (RBF)* kernel [31], defined as $K(\mathbf{x}, \mathbf{y}) = \exp(-\|\mathbf{x} - \mathbf{y}\|^2 / 2\sigma^2)$, σ being a tuning parameter is used in this paper. While fitting the SVM models, the associated hyperparameters C and σ are determined through precisely following the grid-search technique, as recommended and utilized by Chapelle [35].

4.3 The FANN model

Initially inspired from the biological structure of human brain, ANNs gradually achieved great success and recognition in versatile domains, including time series forecasting. The major advantage of ANNs is their nonlinear, flexible, and model-free nature [4, 25, 33]. ANNs have the remarkable ability of adaptively recognizing relationship in input data, learning from experience and then utilizing the gained previous knowledge to predict unseen future patterns. Unlike other nonlinear statistical models, ANNs do not require any information about the intrinsic data-generating process. Moreover, an ANN can always be designed that can approximate any nonlinear continuous function as closely as desired. Due to this reason, ANNs are referred to as the universal function approximators [24, 33].

The most common ANN architecture, used in time series forecasting, is the multilayer perceptron (MLP). An MLP is a feedforward architecture of an input, one or more hidden and an output layer in such a way that each layer

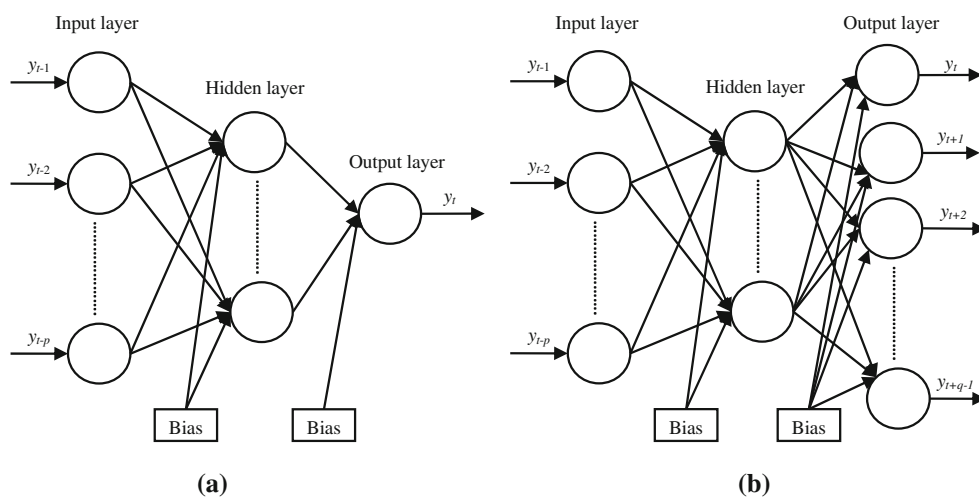
consists of several interconnecting nodes, which transmits processed information to the next layer. It is also known as an FANN model. An FANN with a single hidden layer is often sufficient for practical time series forecasting applications, and so FANNs with single hidden layers are considered in this paper. There are two extensively popular approaches for multi-periodic forecasts through FANNs, viz. *iterative* and *direct* [5, 32]. An iterative approach consists of one neuron in the output layer, and the value of the next period is forecasted using the current predicted value as one of the inputs. On the contrary, the number of output neurons in a direct approach is precisely equal to the forecasting horizon, i.e., the number of future observations to be forecasted. In short-term forecasting, the direct method is usually more accurate than its iterative counterpart, but there is no firm conclusion in this regard [32]. The network structures for iterative and direct FANN forecasting methods are shown in Fig. 2a, b respectively.

4.4 The EANN model

Relatively recently, EANNs attracted notable attention of time series forecasting community. An EANN has a recurrent network structure that differs from a common feedforward ANN through inclusion of an extra *context layer* and *feedback connections* [34]. The context layer is continuously fed back by the outputs from the hidden layer, and as such it acts as a reservoir of past information. This recurrence imparts robustness and dynamism to the network so that it can perform temporal nonlinear mappings. The architecture of an EANN model is shown in Fig. 3.

EANNs generally provide better forecasting accuracies than FANNs due to the introduction of additional memory units. But they require more network weights, especially the hidden nodes in order to properly model the associated temporal relationship. However, there is no rigorous

Fig. 2 The FANN architectures for the following: **a** iterative forecasting, **b** direct forecasting



guideline in the literature for selecting the optimal structure of an EANN model [5]. In this paper, we set the number of hidden nodes as 25 and the training algorithm as *traingdx* [36] for all EANN models.

5 Empirical results and discussions

Six real-world time series from different domains are used in order to empirically examine the performances of our proposed ensemble scheme. These are collected from the Time Series Data Library (TSDL) [37], a publicly available online repository of a wide variety of time series datasets. Table 1 presents the descriptions of these six time series,

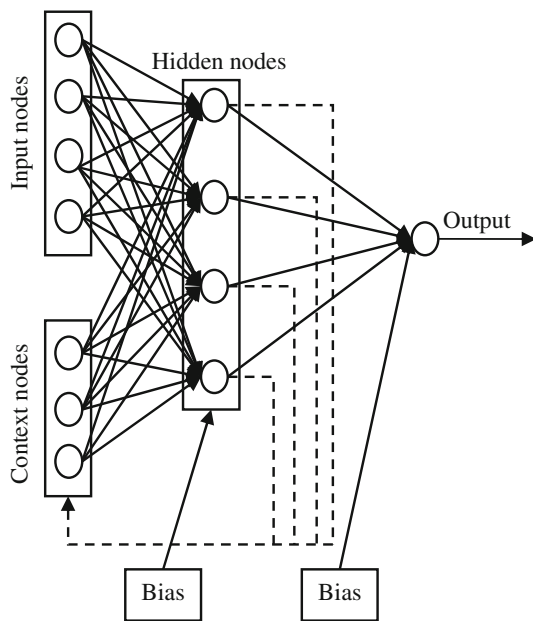


Fig. 3 Architecture of an EANN model

Table 1 Descriptions of the time series datasets

Time series	Description	Type	Total size	Testing size
Canadian Lynx (LYNX)	Number of lynx trapped per year in the Mackenzie River district of Northern Canada (1821–1934)	Stationary, nonseasonal	114	14
Sunspots (SNSPOT)	The annual number of observed sunspots (1700–1987)	Stationary, nonseasonal	288	67
USA Real GNP (RGNP)	Real GNP of USA in billions of dollars (1890–1974)	Nonstationary, nonseasonal	85	15
Child Births (BIRTHS)	Births per 10,000 of 23 year old women in USA (1917–1975)	Nonstationary, nonseasonal	59	10
Airline Passengers (AP)	Monthly number of international airline passengers (in thousands) (January 1949–December 1960)	Monthly, seasonal	144	12
USA expenditure (UE)	Quarterly new plant/equipment expenditures in USA (1964–1976)	Quarterly, seasonal	52	8

and Fig. 4 depicts their corresponding time plots. The horizontal and vertical axis of each time plot, respectively, represents the indices and actual values of successive observations. Here, we are considering short-term forecasting, and so the size of the testing dataset for each time series is kept reasonably small.

All experiments are performed on MATLAB. The default neural network toolbox [36] is used for the FANN and EANN models. The forecasting accuracies are evaluated through the mean-squared error (MSE) and the symmetric mean absolute percentage error (SMAPE), which are defined as follows:

$$MSE = \frac{1}{N} \sum_{t=1}^N (y_t - \hat{y}_t)^2 \tag{11}$$

$$SMAPE = \frac{1}{N} \sum_{t=1}^N \frac{|y_t - \hat{y}_t|}{(y_t + \hat{y}_t)/2} \times 100 \tag{12}$$

where y_t and \hat{y}_t are, respectively, the actual and forecasted values, and N is size of the testing set.

MSE and SMAPE are relative error measures and both provide a reasonably good idea about the forecasting ability of a fitted model. For better forecasting performance, the values of both these error statistics are desired to be as small as possible. The information about the determined optimal forecasting models for all the six datasets is presented in Table 2.

Five other linear combination schemes are considered for comparing with our proposed method. The obtained forecasting results of the individual models and linear combination methods for all six time series are, respectively, presented in Tables 3 and 4. The best forecasting accuracies, i.e., the least error measures in each of these tables, are shown in bold. Following previous works [24], the logarithms to the base 10 of the LYNX data are used in the present analysis. Also, the MSE values for the AP

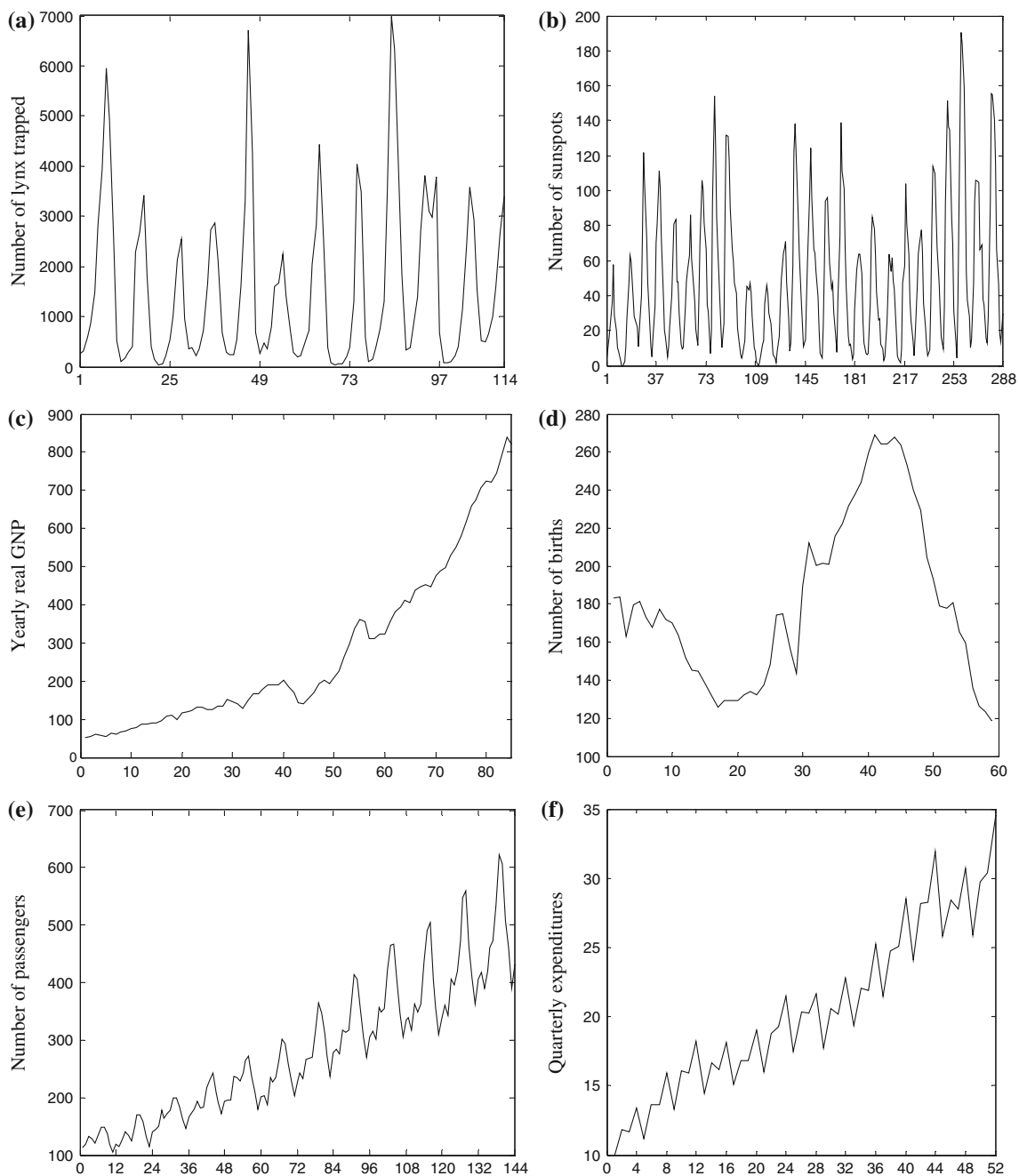


Fig. 4 Time plots of the following: **a** LYNX, **b** SNSPOT, **c** RGNP, **d** BIRTHS, **e** AP, **f** UE

Table 2 The appropriate forecasting models for the six time series datasets

Time series	ARIMA	Iterated FANN	Iterated EANN	Direct EANN
LYNX	ARIMA (12, 0, 0)	$7 \times 5 \times 1$	$7 \times 25 \times 14$	$7 \times 25 \times 14$
SNSPOT	ARIMA (9, 0, 0)	$11 \times 9 \times 1$	$11 \times 25 \times 1$	$11 \times 25 \times 67$
RGNP	Random walk	$4 \times 10 \times 1$	$4 \times 25 \times 1$	$12 \times 25 \times 15$
BIRTHS	Random walk	$6 \times 6 \times 1$	$6 \times 25 \times 1$	$8 \times 25 \times 10$
AP	SARIMA (0, 1, 1) \times (0, 1, 1) ¹²	$12 \times 3 \times 1$	$12 \times 25 \times 1$	$12 \times 25 \times 12$
UE	SARIMA (0, 1, 1) \times (0, 1, 1) ⁴	$4 \times 4 \times 1$	$4 \times 25 \times 1$	$4 \times 25 \times 8$

Table 3 Forecasting results of the individual models

Models	LYNX		SNSPOT		RGNP		BIRTHS		AP		UE	
	MSE	SMAPE	MSE	SMAPE	MSE	SMAPE	MSE	SMAPE	MSE	SMAPE	MSE	SMAPE
ARIMA	0.0153	3.40	803.34	44.44	1,906.05	4.92	98.00	4.71	0.0291	2.48	1.88	4.01
SVM	0.0527	6.12	792.96	33.375	1,194.95	4.00	61.87	4.53	0.0177	2.35	1.61	3.51
Iterated FANN	0.0319	5.25	669.03	40.472	813.38	2.75	93.37	5.48	0.0378	3.39	2.76	4.71
Iterated EANN	0.0365	5.48	899.89	30.435	322.17	2.27	75.50	5.41	0.0333	3.32	2.58	3.94
Direct EANN	0.0189	3.80	1,479.64	53.616	1,772.28	5.12	124.8	5.66	0.0967	5.64	1.94	3.80

Table 4 Forecasting results of the combination methods

Models	LYNX		SNSPOT		RGNP		BIRTHS		AP		UE	
	MSE	SMAPE	MSE	SMAPE	MSE	SMAPE	MSE	SMAPE	MSE	SMAPE	MSE	SMAPE
Simple average	0.0172	3.75	693.11	30.848	301.78	2.05	45.52	3.62	0.0189	2.23	0.91	2.28
Trimmed mean (40 % trimming)	0.0222	4.21	743.97	31.633	301.22	1.96	39.74	3.33	0.0196	2.33	1.11	2.63
Median	0.0232	4.29	795.67	31.007	296.84	2.00	39.35	3.42	0.0176	2.25	1.28	2.67
LSR	0.0254	4.10	881.16	113.62	263.48	1.90	81.06	4.31	0.0188	2.27	1.70	3.84
Outperformance	0.0152	3.60	686.79	31.654	285.65	2.18	45.87	3.64	0.0143	2.16	1.19	2.72
Proposed-I ($0 \leq \alpha < 0.5$)	0.0146	3.33	664.21	29.802	174.64	1.56	26.51	2.81	0.0127	1.97	0.792	2.08
Proposed-II ($0.5 \leq \alpha < 1$)	0.0115	3.03	646.00	30.245	255.27	1.72	29.90	2.82	0.0139	2.03	0.725	2.36

dataset are given in transformed scale (original $MSE = MSE \times 10^4$).

The following important observations are evident from Tables 3 and 4:

1. The obtained forecasting accuracies notably vary among the individual models, and no single model alone could achieve the best forecasting results for all datasets.
2. In terms of MSE, the simple average and the median outperformed the best individual models for three and four datasets, respectively. In terms of SMAPE, both of them outperformed the best individual model for four datasets.
3. Our proposed schemes, viz. Proposed-I and Proposed-II outperformed all individual forecasting models as well as linear combination methods in terms of both MSE and SMAPE.
4. Among themselves, the Proposed-I scheme achieved least MSE and SMAPE values for three and five datasets, respectively, whereas the Proposed-II scheme achieved least MSE and SMAPE values for three and one datasets, respectively.

We present the two bar diagrams in Fig. 5 for visual depictions of the forecasting performances of different methods for all six time series.

We have transformed the error measures for some datasets in order to uniformly depict them in Fig. 5a, b. In

Fig. 5a, the MSE values for SNSPOT and RGNP are divided by 10,000, and those for BIRTHS and UE are divided by 1,000 and 100, respectively. Similarly, in Fig. 5b, the SMAPE values for the SNSPOT data are divided by 10. Figure 5a, b clearly show that the two forms of our proposed scheme, viz. Proposed-I and Proposed-II achieved least MSE and SMAPE values throughout.

We further show the percentage reductions in MSE and SMAPE of the best individual models through our proposed schemes in Fig. 6a, b, respectively. From these figures, it can be seen that except SNSPOT, for all other datasets, our proposed schemes reduced the forecasting errors of the best individual models to considerably large extents. Only for SNSPOT, the amounts of error reductions are small, which can be credited to the reasonably good performances of the corresponding best individual models for this dataset.

The diagrams of the actual observations and their forecasts through the proposed combination scheme for all six time series are shown in Fig. 7. The closeness among the actual and forecasted observations for each dataset is clearly visible in all the six plots of Fig. 7.

We have carried out the nonparametric Friedman test for a statistical analysis of the obtained forecasting results. This test evaluates the null hypothesis (H_0) that all the forecasting methods are equally effective in terms of MSE or SMAPE against the alternative hypothesis (H_1) that all

of them are not equally effective [38]. The obtained Friedman test results are as follows:

- For forecasting MSE, the Friedman’s χ^2 statistic is 46.92 and $p = 0.0000022$.
- For forecasting SMAPE, the Friedman’s χ^2 statistic is 46.69 and $p = 0.0000024$.

Here, p represents the probability that the null hypothesis is true. From the sufficiently small values of p , we reject the null hypotheses with 95 % confidence level and conclude that the forecasting methods differ significantly in terms of both MSE and SMAPE.

The Friedman test results are depicted in Fig. 8a, b. In these figures, the mean rank of a forecasting method is

Fig. 5 Bar diagrams showing the performances of all fitted models on the basis of the following: **a** MSE, **b** SMAPE

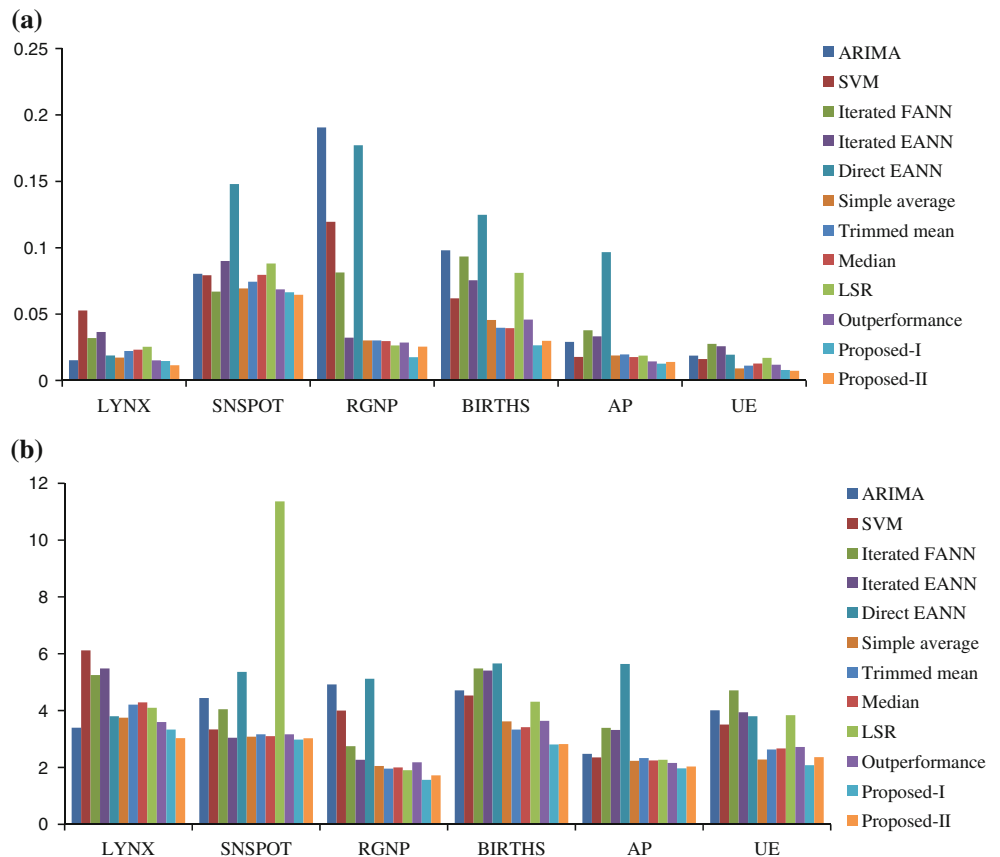
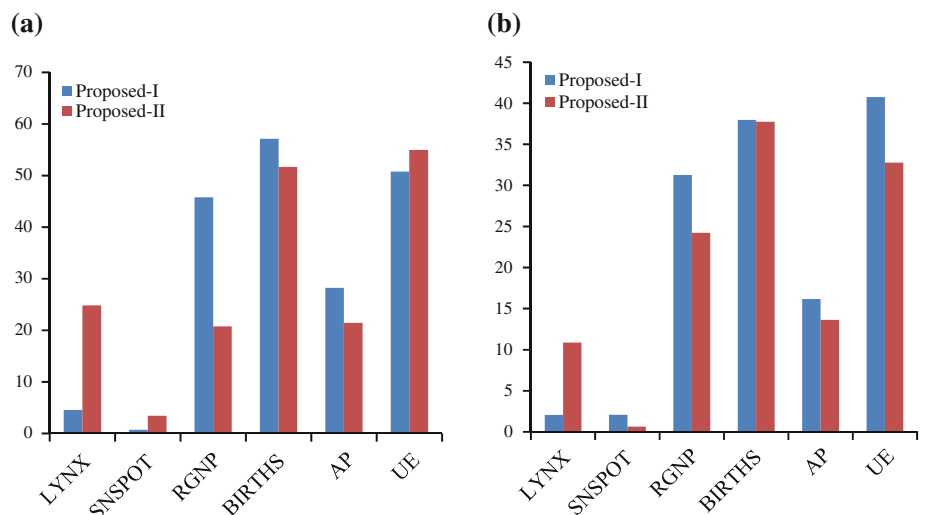


Fig. 6 Percentage improvements over the best individual model in terms of the following: **a** MSE, **b** SMAPE



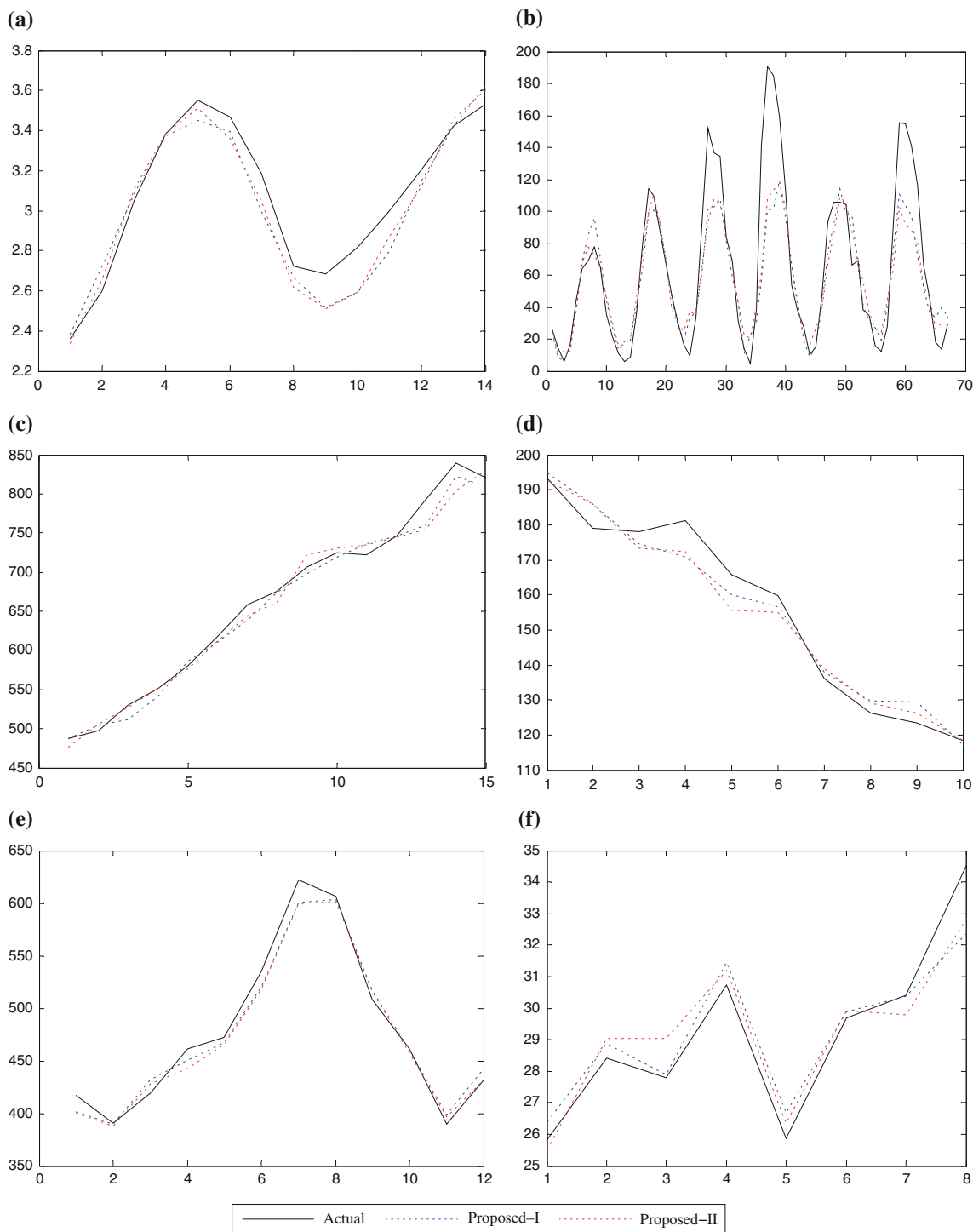
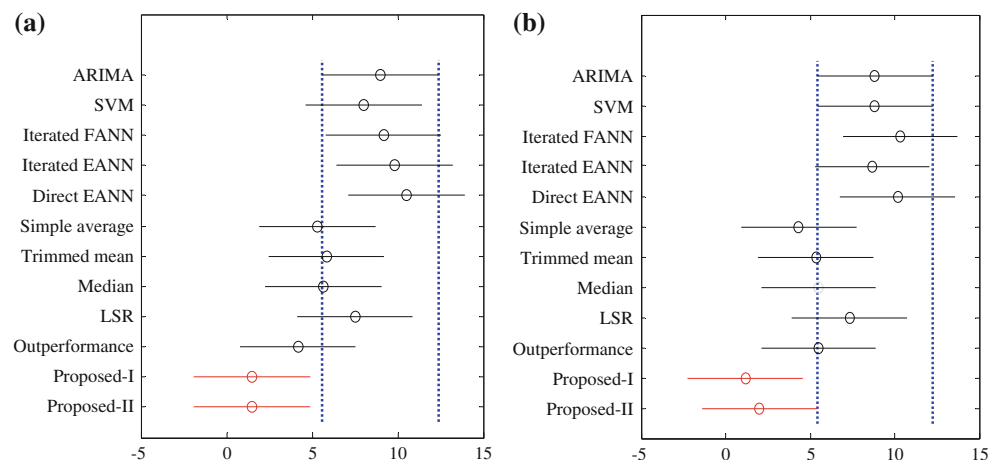


Fig. 7 Diagrams of actual and forecasted observations for the time series: **a** LYNX, **b** SNSPOT, **c** RGNP, **d** BIRTHS, **e** AP, **f** UE

pointed by a circle, and the horizontal bar across each circle is the critical difference. The performances of two methods differ significantly if their mean ranks differ by at least the critical difference, i.e., if their horizontal bars are nonoverlapping.

From Fig. 8a, b, it can be seen that the individual forecasting methods do not differ significantly among themselves, but many of them are significantly outperformed by our combination schemes in terms of both obtained MSE and SMAPE values.

Fig. 8 Friedman test result in terms of the following: **a** MSE, **b** SMAPE



6 Conclusions

Time series analysis and forecasting have major applications in various scientific and industrial domains. Improvement of forecasting accuracy has been constantly drawing the attentions of researchers during the last two decades. Extensive works in this area have shown that combining forecasts from multiple models substantially improves the overall accuracies. Moreover, in many occasions, the simple combinations performed considerably better than more complicated and sophisticated methods. In this paper, we propose a linear combination scheme that takes advantage of the strengths of both simple average and median for combining forecasts. The proposed method assumes that each future observation of a time series is a linear combination of arithmetic mean and median of the individual forecasts together with a random noise. Two approaches are suggested for estimating the tuning parameter α that manages the relative weights between simple average and median. Empirical analysis is conducted with six real-world time series datasets and five forecasting models. The obtained results clearly demonstrate that both forms of our proposed ensemble scheme significantly outperformed each of the five individual models and a number of other common linear forecast combination techniques. These findings are further justified through a nonparametric statistical test.

Acknowledgments The authors are thankful to the reviewers for their constructive suggestions which significantly facilitated the quality improvement of this paper. In addition, the first author likes to express his profound gratitude to the Council of Scientific and Industrial Research (CSIR), India, for the obtained financial support in performing this research work.

References

- De Gooijer JG, Hyndman RJ (2006) 25 years of time series forecasting. *J Forecast* 22(3):443–473
- Zhang GP (2007) A neural network ensemble method with jittered training data for time series forecasting. *Inform Sci* 177(23):5329–5346
- Box GEP, Jenkins GM (1970) *Time series analysis, forecasting and control*, 3rd edn. Holden-Day, California
- Andrawis RR, Atiya AF, El-Shishiny H (2011) Forecast combinations of computational intelligence and linear models for the NN5 time series forecasting competition. *Int J Forecast* 27(3): 672–688
- Lemke C, Gabrys B (2010) Meta-learning for time series forecasting and forecast combination. *Neurocomputing* 73: 2006–2016
- Clemen RT (1989) Combining forecasts: a review and annotated bibliography. *J Forecast* 5(4):559–583
- Jose VRR, Winkler RL (2008) Simple robust averages of forecasts: some empirical results. *Int J Forecast* 24(1):163–169
- Armstrong JS (2001) *Principles of forecasting: a handbook for researchers and practitioners*. Academic Publishers, Norwell, MA
- Stock JH, Watson MW (2004) Combination forecasts of output growth in a seven-country data set. *J Forecast* 23:405–430
- Larrece JC, Moïn pour R (1983) Managerial judgment in marketing: the concept of expertise. *J Market Res* 20:110–121
- Agnew C (1985) Bayesian consensus forecasts of macroeconomic variables. *J Forecast* 4:363–376
- McNees SK (1992) The uses and abuses of ‘consensus’ forecasts. *J Forecast* 11:703–711
- Crane DB, Crotty JR (1967) A two-stage forecasting model: exponential smoothing and multiple regression. *J Manage Sci* 13:B501–B507
- Zarnowitz V (1967) *An appraisal of short-term economic forecasts*. National Bureau of Econ Res, New York
- Reid DJ (1968) Combining three estimates of gross domestic product. *J Economica* 35:431–444
- Bates JM, Granger CWJ (1969) Combination of forecasts. *J Oper Res Q* 20:451–468
- Granger CWJ, Ramanathan R (1984) Improved methods of combining forecasts. *J Forecast* 3:197–204
- Granger CWJ, Newbold P (1986) *Forecasting economic time series*, 2nd edn. Academic Press, New York
- Aksu C, Gunter S (1992) An empirical analysis of the accuracy of SA, OLS, ERLS and NRLS combination forecasts. *Int J Forecast* 8:27–43
- Newbold P, Granger CWJ (1974) Experience with forecasting univariate time series and the combination of forecasts (with discussion). *J R Stat Soc A* 137(2):131–165
- Winkler RL, Makridakis S (1983) The combination of forecasts. *J R Stat Soc A* 146(2):150–157

22. Pollock DSG (1993) Recursive estimation in econometrics. *Comput Stat Data An* 44:37–75
23. Bunn D (1975) A Bayesian approach to the linear combination of forecasts. *J Oper Res Q* 26:325–329
24. Zhang GP (2003) Time series forecasting using a hybrid ARIMA and neural network model. *Neurocomputing* 50:159–175
25. Khashei M, Bijari M (2010) An artificial neural network (p, d, q) model for time series forecasting. *J Expert Syst Appl* 37(1): 479–489
26. Khashei M, Bijari M (2011) A novel hybridization of artificial neural networks and ARIMA models for time series forecasting. *J Appl Soft Comput* 11:2664–2675
27. Adhikari R, Agrawal RK (2012) A novel weighted ensemble technique for time series forecasting. In: *Proceedings of the 16th Pacific-Asia conference on knowledge discovery and data mining (PAKDD)*, Kuala Lumpur, Malaysia, pp 38–49
28. Winkler RL, Clemen RT (1992) Sensitivity of weights in combining forecasts. *Oper Res* 40(3):609–614
29. Tseng FM, Yu HC, Tzeng GH (2002) Combining neural network model with seasonal time series ARIMA model. *J Technol Forecast Soc Change* 69:71–87
30. Vapnik V (1995) *The nature of statistical learning theory*. Springer, New York
31. Suykens JAK, Vandewalle J (1999) Least squares support vector machines classifiers. *Neural Process Lett* 9(3):293–300
32. Hamzaçebi C, Akay D, Kutay F (2009) Comparison of direct and iterative artificial neural network forecast approaches in multi-periodic time series forecasting. *J Expert Syst Appl* 36(2): 3839–3844
33. Gheyas IA, Smith LS (2011) A novel neural network ensemble architecture for time series forecasting. *Neurocomputing* 74(18): 3855–3864
34. Lim CP, Goh WY (2005) The application of an ensemble of boosted Elman networks to time series prediction: a benchmark study. *J Comput Intel* 3(2):119–126
35. Chapelle O (2002) *Support vector machines: introduction principles, adaptive tuning and prior knowledge*. Ph.D. Thesis, University of Paris, France
36. Demuth H, Beale M, Hagan M (2010) *Neural network toolbox user's guide*. The MathWorks, Natic, MA
37. Hyndman RJ (2011) *Time series data library (TSDL)*, <http://robjhyndman.com/TSDL/>
38. Hollander M, Wolfe DA (1999) *Nonparametric statistical methods*. Wiley, Hoboken, NJ