Particle Swarm Optimization-based Empirical Mode Decomposition Predictive Technique for Nonstationary Data

Cry Kuranga*[1] · Njodzi Ranganai[2] · Tendai S. Muwani[2]

*Cry Kuranga: kurangacry@gmail.com
[1]Department of Computer Science, University of Pretoria, Lynnwood Road, Hillcrest, Pretoria 0002, South Africa
[2]Department of Computer Science and Information Systems, Manicaland State University of Applied Sciences, Stair Guthrie Rd., Fernhill, P. Bag 7001, Mutare, Zimbabwe

Abstract

Real-world nonstationary data are usually characterized by high nonlinearity and complex patterns due to the effects of different exogenous factors that make prediction a very challenging task. An ensemble strategically combines multiple techniques and tends to be robust and more precise compared to a single intelligent algorithmic model. In this work, a dynamic particle swarm optimization-based empirical mode decomposition ensemble is proposed for nonstationary data prediction. The proposed ensemble implements an environmental change detection technique to capture concept drift occurring and the intrinsic nonlinearity in time series, hence improving prediction accuracy. The proposed ensemble technique was experimentally evaluated on electric time series datasets. The obtained results show that the proposed technique improves prediction accuracy and it outperformed several stateof- the-art techniques in several cases. For future work direction, a detailed empirical analysis of the proposed technique can be considered such as the effect of the cost of prediction errors, and the technique's search capability.

Keywords: Empirical Mode Decomposition, Nonlinear Autoregressive, Ensemble Technique, Particle Swarm Optimization, Time Series Forecasting, Concept Drift, Nonstationary Data

1 Introduction

Real-world nonstationary data such as time series are characterized by high nonlinearity and complex patterns due to the effects of different exogenous factors such as weather condition, and economic fluctuation that make prediction a very challenging task [1]. Usually, statistical-based modeling techniques are used for time series forecasting due to their strong expansion ability and simple features [2]. However, these techniques usually fail to forecast nonstationary time series, especially when there are great fluctuations and classical chaos in the series [3]. To overcome the challenges of statistical-based modeling, machine learning techniques such as support vector regression (SVR) have been introduced into time series forecasting. Nevertheless, the accuracy of machine learning techniques like SVR mainly depends on hyperparameter tuning and is usually prone to overfitting [2].

The short-term electric load and pricing time series can be considered as a stream of incoming data, as such, incremental learning is ideal to model such time series. The presence of concept drifting in time series poses challenges especially in incremental learning due to the random drifting of statistical properties of the target variable [4]. Incremental learning can take a passive or active approach. A passive approach is ideal if the environment is continuously drifting in which the induced model parameters are continuously adapted to produce a model that effectively depicts the current distribution. Conversely, an active approach adapts the learning model only if an environmental change is detected [5].

The empirical mode decomposition (EMD), is an adaptive data processing technique tailored for nonlinear and nonstationary data [8]. The EMD decomposes data into intrinsic mode function (IMF) components and a residual. An ensemble technique is a combinatorial model that strategically combines multiple algorithms to yield a representational, computational, and statistically advantageous model [6]. The combinatorial model (hybrid) tends to be robust and more precise compared to a single intelligent algorithmic model [7]. Thus, EMD can be combined with a forecasting model in which each IMF component is modeled independently in anticipation of improved performance [9].

The dynamic particle swarm optimization-based nonlinear autoregressive with exogenous inputs (NARX-QPSO) presented in [10] is an adaptive incremental learning technique. The NARX-QPSO technique has strong generalization capability, copes with concept shifts in nonstationary data, and possesses good learning capabilities in which it takes into account the influence of exogenous factors. However, the predictive accuracy of NARX-QPSO was outperformed by the state-of-the-art techniques due to its inability to capture the intrinsic nonlinearity in the data [10].

In this work, a particle swarm optimization-based empirical mode decomposition predictive technique is proposed to improve on the predictive accuracy of NARX-QPSO. The proposed ensemble is build on NARX-QPSO in which it combines EMD to capture the intrinsic nonlinearity in time series, NARX-QPSO, and an environmental change detection technique to detect concept drift occurring. The EMD decomposes a time series into IMF components which are modeled independently by NARX-QPSO to construct an ensemble prediction technique. The proposed ensemble dynamically adapts when a concept shift is detected. Thus, this work experimentally evaluates the effectiveness of:

a) Hybridizing an EMD with NARX-QPSO, on improving forecasting performance;
b) Using the base regressors, namely the least-squares approximation and SVR in NARX-QPSO to induce a predictive model; and
c) Ensemble combining techniques, namely averaging and weighted average to the overall algorithm performance.

Also, the best performing technique is compared to EMD-based state-of-the-art ensemble techniques.

This work is organized as follows: Section 2 provides a discussion of related literature and Section 3 presents the proposed ensemble technique and experimental setup. Section 4 presents and discusses the obtained results. The conclusion and future work directions are discussed in Section 5.

2 Related Literature

2.1 Time Series Modeling

The purpose of time series modeling is to gather and analyze historical data to fit a model that describes the internal structure of a time series such as autocorrelation, trend, or seasonality. The created model is then used to forecast future values of the series in which the AR, ARX, and ARIMA models are among the most widely utilized models [11].

Three classes of time-series forecasting models are statistical, machine learning, and hybrids [12]. When there is a lack of knowledge and information about the data generating process and the factors that influence it, statistical models are used [11]. A statistical model fits a model to the data using calculated parameters, expressed as a mathematical function. Typically, statistical models are characterized by numerous computational possibilities which entail high complexity, especially for data patterns that exhibit nonlinearity.

Machine learning techniques use artificial intelligence (AI) algorithms to model time series data and its forecasting accuracy usually outperforms statistical models though exhibits several drawbacks such as overfitting and being trapped in local minima [15] [16] [5] [17]. As black-box models in AI are increasingly being implemented to make significant predictions in vital contexts, the demand for explainability and transparency is increasing in AI [13]. The challenge happens if the decision that are created are not justifiable, legitimate, or simply do not allow to obtain detailed explanations of their behavior [14] [20]. Conversely, it is usually easier to put everything into a black-box and optimize for the highest performance possible. Hybridization harnesses the strength of two or more techniques to improve on effectiveness and accuracy of the induced model. However, hybrid models are usually computationally expensive [18] [19].

A statistical learning model consists of measured variables inputs (X_k) as well as a target variable, outputs (Y_k) [11]. The assumption is that there is a connection between the X_k and the Y_k, and the statistical model tries to figure out how to mathematically describe it. Thus, creating a process model based on the information given. A commonly used statistical learning model for nonlinear data is a nonlinear autoregressive model with exogenous inputs (NARX).

The NARX model connects the current value of output to both prior values of the same output and current and previous values of externally influencing inputs [11]. The NARX model usually provides accurate predictions though it is one of the most computationally expensive statistical models and the least durable [21]. The NARX model can be described using a linear difference equation as follows [21]:

$$y(t) = a_1 y(t-1) + \cdots + a_k y(t-k) + b + b_0 x(t) + \cdots + b_l x(t-l) \quad (1)$$

where *y(t)* is the variable of interest, *x(t)* is exogenous input at time *t, a* and *b* are estimated parameters, and *(k, l)* is the maximum time lag of the two variables important to the model.

The autoregressive component of the model is represented by the dependency on lagged versions of *y*, while the exogenous or additional variable is represented by *x's*. Time is discretized, and at each sampling period, *t* varies by one unit [21].

2.2 Dynamic QPSO-based NARX Technique

Particle swarm optimization is a population-based metaheuristic algorithm that converges faster compared to several metaheuristic techniques [26]. However, PSO can easily be trapped into local optima [22]. A charged particle swarm optimization algorithm (PSO), referred to as quantum-inspired PSO (QPSO) is an optimization algorithm designed for dynamic environments [23]. The QPSO consists of both neutral and quantum particles [24] [25]. Neutral particles promote exploitation in the search space whereas quantum particles explore new optima, facilitating exploration. The literature suggests that QPSO can adapt in the presence of a concept drift occurring environment [27]. The PSO and its variants such as QPSO have been successfully implemented in hybrid with other techniques as an optimizer of either the inherent parameters of the other technique(s) or the induced model [50][51][52].

A *dynamic QPSO-based regression* technique (DynQPSO) was proposed in [10]. The DynQPSO is a data-driven predictive technique that can adapt the predictive model in an environment with concept shifts occurring to cope with uncertainty in the presence of change. The DynQPSO combines the NARX model and QPSO. Firstly, NARX estimates the model parameters and QPSO optimizes both the parameters and the structure of the induced model. The NARX uses the unconstrained least-squares technique which strives to minimize the prediction error. The fitness of each QPSO particle is measured using mean square error. A sliding window of analysis is used to model real-world streaming data. The sliding window of analysis is partitioned into training which consists of the first 80% of data points and testing, the remaining data points. Algorithm 1 summarizes the DynQPSO technique.

Algorithm 1 Dynamic QPSO-based Regression Algorithm

Set analysis_ window to w_s data points
BEGIN

    Estimate model coefficients using NARX

    DO

        Move analysis_window

        Run QPSO *n* iterations

        If environmental change is detected

            Update the coefficients of term-coefficient mappings of each particle

            Run QPSO *n* iterations

    REPEAT until no further data to analyze.

END

2.3 Empirical Mode Decomposition

Decomposing the original data stream into separate different components to deal with each component separately and combining the results usually improve predictive performance [7]**.** However, decomposing the data stream increases the prediction burden and brings multiple random mistakes. An example of a technique that can decompose a data stream is Discrete-wavelet transform (DWT) [7]. The DWT transforms nonlinear and nonstationary signals into separate different components. However, DWT fails to achieve fine resolutions for both time-domain and time-frequency which poses a significant challenge for short-term time series analysis [7].

Instead of decomposing signals based on wavelet basis and prior harmonic functions as in DWT and Fourier decomposition techniques respectively, a novel signal processing method, empirical mode decomposition (EMD) was proposed to decompose signals based on the time scale properties of the data [28]. The EMD approach, in contrast to wavelet decomposition, is intuitive, direct, posterior, and adaptive, allowing it to theoretically break down any form of signal to yield considerable benefits when working with non-linear and non-stationary data [29]. Thus, EMD extracts instantaneous frequency data from nonlinear and nonstationary data using an empirical approach to decompose a signal into multiple IMFs and a remnant that represents the trend. An IMF can be considered as a function with a single extreme between zero crossings and a mean value of zero [30].

The EMD technique employs multi-resolution to solve the problem of choosing a wavelet basis function in wavelet transformation. By drilling down non-linear or non-stationary time series into singular values independent IMFs and calculating the overall trend of the data series, the EMD approach can help determine its features. Also, EMD can efficiently decompose pixel value and avoid entrapment in a local optimum, resulting in improved model reliability and performance [31].

Decomposing a time series using EMD produces IMFs that reduce the number of variables in the predictive model to yield a high degree of fit. However, high volatility still exists for a lower-order IMF component which happens to have valuable information from the time series that makes it difficult to simply discard the component to reduce noise [2]. Also, the mode mixing problem is a significant disadvantage of EMD. One IMF may include signals from a wide range of frequencies, or multiple IMF signals in a comparable frequency band may be present. [32]. This problem is typically solved by the ensemble EMD (EEMD) [33]. The EEMD is based on repeatedly applying EMD to a time series signal with uncorrelated Gaussian noise and merging the results to remove the noise. The robustness of EEMD is realized by a noise-assisted technique that slightly perturbed signals from their initial position to perform several decompositions. The noise cancels each other to produce a pure decomposition when averaging the results of all IMFs [33].

The electric load is a nonstationary and nonlinear time series that consists of independent components. A variety of factors tend to influence electric load behavior such as random impacts, economic considerations, time, day, season, and weather. As a result, the EMD algorithm has the potential to be extremely useful for forecasting demand for an electric load.

2.4 EMD-based Ensemble Techniques

Ensemble learning approaches, often known as hybrid methods, aim to improve predicting performance by strategically combining various algorithms to realize representational, computational, and statistical effectiveness [6] [34]. An ensemble can either be parallel or sequential [35]. A parallel approach splits the dataset in which each sub-dataset is trained independently and the final model is the combination of each sub-dataset induced model. In a sequential approach, the results of a given technique become the input of the next technique [34]. Numerous parallel ensemble approaches have been proposed in the literature, including EMD and wavelet decomposition [36] [18] [37] [28].

A sequential ensemble learning strategy for a short-term electric load that combines EMD and random vector functional link network (RVFL) was proposed in the literature [38]. Five AEMO electric load datasets and six benchmark techniques were used in the experiments. The presented results suggest that EMD-based ensemble techniques, EMD-SVR, EMD-SLFN, and EMD-RVFL, were superior in performance compared to single structure models. Also, the proposed EMD-RVFL outperformed all other techniques under study.

An EMD-based incremental time series forecasting ensemble technique (DWT-EMD-RVFL) was proposed in the literature that combines three techniques, namely DWT, EMD, and RVFL [5]. The DWT deals with the frequency of the series, EMD decomposes the series, and the RVFL models each IMFs and a residue. The incremental RVFL creates the final ensemble predictive model. The DWT-EMD-RVFL technique was used to forecast AEMO electric load in which the proposed incremental forecasting technique outperformed several techniques under study.

Another EMD-based incremental ensemble, H-EMD-SVR-PSO was proposed in the literature to forecast electric load [17]. Instead of modeling each IMF, H-EMD-SVR-PSO defined four components: ‘A’ consists of the middle and random terms; ‘B’ consists of the residual and middle terms, and ‘C’ consists of the middle terms. The first three components are modeled independently using SVR-PSO and the last component, ‘D’ is computed as:

$$D = A + B - C \tag{2}$$

The AEMO electric load dataset was used in the experiments and the presented results suggest that the proposed EMD-based ensemble technique was capable to forecast time series data characterized with inherent nonlinearity and residual interactive effects to outperform several EMD-based ensemble techniques under study.

An ensemble kernel machine technique was proposed in the literature which consists of EMD, Kernel Ridge Regression (KRR), and SVR [39] The EMD decomposes the time series and a KRR models the extracted components, and SVR was used to combine the models. The performance of the suggested technique was evaluated using AEMO electricity pricing datasets. Furthermore, six benchmark techniques were used to perform a comparative study of the proposed model. The following conclusions were drawn from the presented results: hybrid techniques based on EMD, such as EMD-SVR, EMD-SLFN, and the EMD-KRR-SVR, outperform single structure models, KRR has the quickest computing time, and the proposed EMD-KRR-SVR yields the best prediction performance.

Several EMD-based ensemble approaches for various research domains have been presented in the literature. For example, for wind speed forecasting, EMD and upgraded variants were paired with SVR and ANN [2], a self-adapting EMD-based learning model for multi-step-ahead forecasting was proposed in [53]. EMD-based ensemble techniques to predict wind speed and to forecast tourism demand were proposed in the literature [40] [32]. An EMD-based forecasting model was proposed in [41] which hybridized least-squares SVMs and EMD. The presented results suggest that EMD-based techniques outperformed single classifier models. Readers who are interested in ensemble methods should read the survey paper [35].

3 Proposed Technique and Experimental Setup

The proposed technique combines EEMD, NARX-QPSO and concept drift detection technique. The EEMD decompose the series to yield IMFs which are modeled independently using NARX. The QPSO is tasked to optimize both the NARX parameters and the structure of the induced predictive model [10]. The induced models for each IMF are combined to create an ensemble. Algorithm 2 summarizes the proposed particle swarm optimization-based empirical mode decomposition technique (EMD-NARX-QPSO).

Algorithm 2 EMD-NARX-QPSO

Set sliding_window (data_window) to w_s data patterns
Ensemble = { }

BEGIN
Do
Slide a data_window
IF a change in the environment is detected (*gBest*_deterioration > *e*)
Decompose the data into its respective IMFs using EEMD
FOR each IMF
Induce a predictive model (p_model) using NARX-QPSO
Append *p_model* to *Ensemble*
Combine elements in *Ensemble* to produce *y_model*
Let *y_model* becomes the *current_model*
Perform forecasting using *current_model*
ELSE
Perform forecasting using the *current_model*
REPEAT until no further data to analyze.
END

A sliding window of analysis technique is adopted to model a nonstationary environment. A sliding window slides through the dataset to mimic a stream of incoming data. If the algorithm detects an environmental change in the current window, then the data is decomposed into its respective IMFs in which each IMF is modeled independently using a NARX-QPSO. The algorithm re-trains if and only if an environmental change is detected else it uses the current model to predict the incoming data.

The fitness function of each QPSO particle is computed using a root mean square error (RMSE). An environmental change technique used in NARX-QPSO is adapted in this work that makes use of a global best particle (*gBest*) in QPSO which is *re-evaluated before* being updated [10]. A significant fitness deterioration ($> e$) of *gBest implies* an environmental change (presence of concept drift in the data pattern). The parameter *e* is a user-defined parameter. For example, a value of $e > 10$ implies that a *gBest* fitness deterioration of more than 10% signifies concept drift occurring in the data.

Two base regressor techniques model are used independently in NARX: unconstraint least-squares approximation [10] and the popular nonlinear regression technique, SVR [42]. Three ensemble combining techniques are adopted in EMD-NARX-QPSO, namely averaging (A), weighted averaging (W), and an additive (H) proposed in [17]. Therefore, considering two base regressors, least-squares approximation (R) and SVR (S), and three ensemble combining techniques: A,W, and H, six variants of EMD-NARX-QPSO are derived. Table 1 presents the proposed variants.

Table 1: **E**MD-**N**ARX-**Q**PSO (ENQ) Variants

Base Technique	**Variants**		
NARX-QPSO	NARX_**R**	NARX_**S**	
EMD-**N**ARX-**Q**PSO (based on R)	ENQ _**R**_A	ENQ _**R**_W	ENQ _**R**_H
EMD-**N**ARX-**Q**PSO (based on S)	ENQ_**S**_A	ENQ _**S**_W	ENQ _**S**_H

The averaging technique computes the average model from each IMF to yield an overall model. The weighted averaging technique considers the inverse of the training accuracy (RMSE) of each model as the weight of that model. As such, the component with the highest precision contributes much to the overall model.

3.1 Performance Measure

The error metrics, root mean square error (RMSE) and mean absolute percentage error (MAPE), are used to ascertain the prediction accuracy of the induced model. The RMSE and MAPE are computed as:

$$RMSE = \sqrt{\frac{\sum_{i=1}^{n}(y_i' - y_i)^2}{n}} \tag{3}$$

$$MAPE = \sum_{i=1}^{n}\left|\frac{y_i' - y_i)}{y_i}\right| \times 100, \tag{4}$$

where n is the testing data patterns, y_i' is forecasted, and y_i is the actual values.

The RMSE indicates the degree of dispersion of error, as such, smaller values reflect favorable prediction stability [7]. Large errors are highly penalized in RMSE since the prediction errors are squared. The MAPE reflects the overall level of deviation between the actual value and the induced model prediction value [7]. As in RMSE, smaller values of MAPE entail a higher prediction accuracy. The RMSE and MAPE are used as performance measures that have the advantage of expressing the prediction error of the model in the same units as the predicted variable [45].

Statistical tests are performed using the non-parametric *Friedman test* based on the null hypothesis (H_0) – 'all techniques yield the same performance'. Friedman's test is ideal to find differences in performance across multiple algorithms. The algorithms (techniques) are ranked separately and for tie cases, average ranks are assigned. The *post-hoc Nemenyi test* is used to infer a statistically significant difference if the H_0 is rejected. A statistically significant difference exists if the difference of mean rank of the pair of techniques was greater than Nemenyi test critical distance (CD) [46].

3.2 Dataset

The proposed model implements a nonlinear autoregressive model with exogenous inputs. Therefore, the datasets with exogenous variables are selected to experimentally evaluate the performance of the proposed model. Sixteen electric datasets from the Australian Energy Market Operator (AEMO) that consist of fifteen electric load [5] and one electric pricing datasets are used in the experiments [47]. The electric pricing dataset was sampled at a half-hour interval for the period 7th May 1996 - 5th December 1998 [48]. The electric pricing dataset exhibits short-term irregular changes due to seasonal changes, and long-term regular changes due to weather fluctuations.

The fifteen electric load datasets from AEMO consist of five Australian states, namely New South Wales (NSW), Victoria (VIC), South Australia (SA), Tasmania (TAS), and Queensland (QLD) for three years (2013-2015) [47]. The data points were sampled at a half-hour interval to give 48 data points for a day and 17 520 data points for a year. Daily temperatures (minimum and maximum) for the following places are considered as *exogenous variables*: Melbourne for VIC, Adelaide for SA, Brisbane for QLD, Hobart and Launceston airport for TAS, and Sydney and Canberra for NSW [49]. The statistics of AEMO datasets are summarized in Table 2.

The dataset is split into training and testing sets. The last three months of each dataset constitute a testing set. The input features: the electric data's (load and pricing) value x_{t-48}, hour, day, month, and the temperature data are used to build a multiple regression model to forecast x_t for one-day horizon (48 steps). The predictive model is built using a training dataset whereas the forecasting performance of the built model is based on out-of-sample done using the testing dataset. Each dataset is linearly scaled to [0,1] computed as:

$$\bar{x}_i = \frac{x_u - x_t}{x_u - x_l} \tag{6}$$

where x_u and x_l are max and min values respectively, x_t is the data point and $\bar{x}_i$ is the scaled value.

Table 2: Summary of AEMO Datasets

Dataset	Year	Data points	Min	Max	Mean	Std.
TAS	2013	17520	659.5	1650.3	1129.3	142.3
	2014	17520	569.1	1630.1	1109.7	139.0
	2015	17520	479.4	1667.2	1138.2	145.3
QLD	2013	17520	4148.7	8278.4	5703.7	747.0
	2014	17520	4073.0	8445.3	5745.7	794.0
	2015	17520	4281.4	8808.7	6035.4	777.2
VIC	2013	17520	9587.5	3551.6	5511.8	895.9
	2014	17520	3272.9	10240.0	5324.4	921.4
	2015	17520	3369.1	8579.9	5194.6	864.7
NSW	2013	17520	5113.0	13788.0	7981.6	1190.9
	2014	17520	5138.1	11846.0	7917.8	1170.1
	2015	17520	5334.4	12602.0	7979.8	1232.7
SA	2013	17520	728.6	2991.3	1426.6	301.7
	2014	17520	682.5	3245.9	1403.3	312.8
	2015	17520	696.3	2870.4	1398.5	306.0

A sliding window (*data_window*) is set to 240 data points, the minimal possible *data_window* for the proposed model to induce a model. The data_window slides by 48 data points (one-day horizon). Thus, for each dataset, at least 90 slides are performed to forecast at least 4320 data points (testing set). When an environmental change is detected, the data_window is split: the first 80% to training and the remaining 20% to testing which is equivalent to one-day horizon. The order of the data points (x_t) in *data_window* are not altered since *t* had an inherent meaning to the data pattern.

3.3 Experimental Setup

The experimental setup described in [5] is adopted in this work. Experimental work is done in a Python environment using Scikit learn [43] on an 8 GB RAM Intel Core i7 processor (3.20 GHz) desktop. The predictive performance of the proposed model is benchmarked by NARX-QPSO [10] using either unconstrained least squares or SVR as a base regressors. Parameters for QPSO in NARX-QPSO and EMD-NARX-QPSO are optimized using a Scikit learn function, GridSearch for the selected range of values from the literature. A cross-domain parameter optimization approach is adopted [44]. The following obtained optimal values are used in the experiments: $e =$

10, QPSO parameters: $c_1 = c_2 = 0.496190$, $\omega = 0.729844$, $n = 100$, $swarm_{size} = 30$, and $quantum_{radius} = 2$.

4 Results and Discussion

This section evaluates experimentally the proposed EMD-NARX-QPSO discussed in Section 3 and presents the obtained generalization results on the evaluation of eight techniques presented in Table 1 using MAPE and RMSE. The statistical analysis was performed on the eight populations (techniques) at $\alpha = 0.05$ significance level.

Table 3 presents the obtained median (MD), mean absolute deviation (MAD), and the mean rank (MR) among all populations over the samples based on Friedman Test. A significant difference between populations exists if the mean rank is greater than CD.

Table 3: Friedman Test Results

Technique	MD		MAD		MR	
Load	**MAPE**	**RMSE**	**MAPE**	**RMSE**	**MAPE**	**RMSE**
ENQ_S_W	0.038±0.024	232.45±366.97	0.020	207.045	1.326	1.267
ENQ_S_H	0.037±0.024	231.67±364.54	0.020	214.799	2.315	2.378
ENQ_S_A	0.037±0.024	227.21±364.15	0.020	207.164	2.685	2.356
NARX_S	0.013±0.024	0.212±0.336	0.019	0.181	3.761	4.044
NARX_R	0.000±0.000	0.000±0.000	0.000	0.000	5.391	5.578
ENQ_R_W	0.000±0.000	0.000±0.000	0.000	0.000	6.087	5.578
ENQ_R_H	0.000±0.000	0.000±0.000	0.000	0.000	6.913	6.122
ENQ_R_A	0.000±0.000	0.000±0.000	0.000	0.000	7.522	7.289
Pricing						
NARX_R	0.075±0.096	0.000±0.005	0.022	0.000	1.000	3.500
ENQ_S_A	0.006±0.002	0.000±0.000	0.001	0.000	2.778	1.167
NARX_S	0.006±0.001	0.000±0.000	0.001	0.000	3.056	5.000
ENQ_S_H	0.006±0.002	0.000±0.000	0.001	0.000	4.000	2.556
ENQ_S_W	0.006±0.002	0.000±0.000	0.001	0.000	4.167	2.778
ENQ_R_W	0.000±0.000	0.000±0.000	0.000	0.000	6.583	6.056
ENQ_R_A	0.000±0.000	0.000±0.000	0.000	0.000	7.139	7.333
ENQ_R_H	0.000±0.000	0.000±0.000	0.000	0.000	7.278	7.611

The following results were obtained for Electric load datasets: $CD = 1.548$ for MAPE and $CD = 1.565$ for RMSE. A value of $p = 0.000$ was obtained for both MAPE and RMSE which implies that there was a significant difference between median values of the populations. The results obtained from post-hoc Nemenyi test indicates that there exists no significant difference among the following sets: {ENQ_S_W, ENQ_S_H, ENQ_S_A}; {ENQ_S_H, ENQ_S_A, NARX_S}; {NARX_R, ENQ_R_W, ENQ_R_H}; {ENQ_R_W, ENQ_R_H, ENQ_R_A} for MAPE and {ENQ_S_W, ENQ_S_A, ENQ_S_H}; {NARX_S, ENQ_R_W}; {ENQ_R_W, ENQ_R_H, ENQ_R_A}; {ENQ_R_H, ENQ_R_A, NARX_R} for RMSE.

The following results were obtained for Electric pricing datasets: $CD = 2.475$ for MAPE and $CD = 2.497$ for RMSE. A value of $p = 0.000$ was obtained for both MAPE and RMSE which implies that there was a significant difference between median values of the populations. The results obtained from post-hoc Nemenyi test indicates that there exists no significant difference among the following sets: { NARX_R, ENQ_S_A, NARX_S}; {ENQ_S_A, NARX_S, ENQ_S_H, ENQ_S_W}; {ENQ_S_W, ENQ_R_W}; {ENQ_R_W, ENQ_R_A, ENQ_R_H} for MAPE and {ENQ_S_A, ENQ_S_H, ENQ_S_W, NARX_R}; {ENQ_S_H, ENQ_S_W, NARX_R, NARX_S}; {NARX_S, ENQ_R_W, ENQ_R_A}; {ENQ_R_W, ENQ_R_A, ENQ_R_H} for RMSE.

Figure 1 and Figure 2 are graphical illustrations of the obtained post-hoc Nemenyi test based on RMSE (top) and MAPE (bottom) for Electric load and pricing datasets.

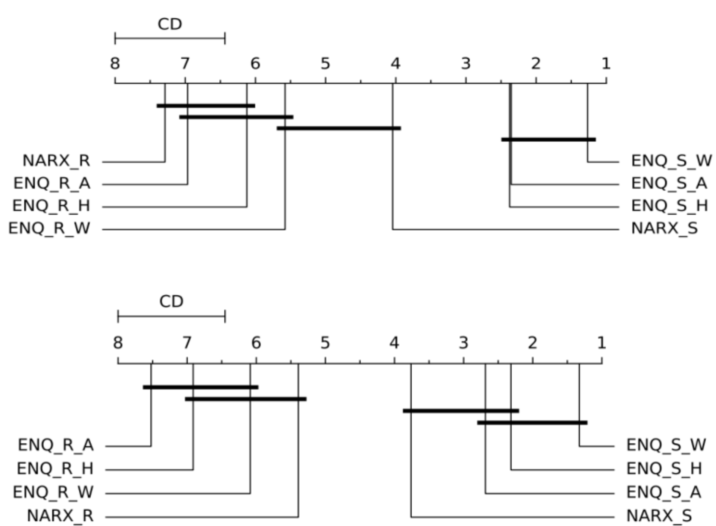

Figure 1: Electric Load Nemenyi Test for RMSE (top) and MAPE (bottom)

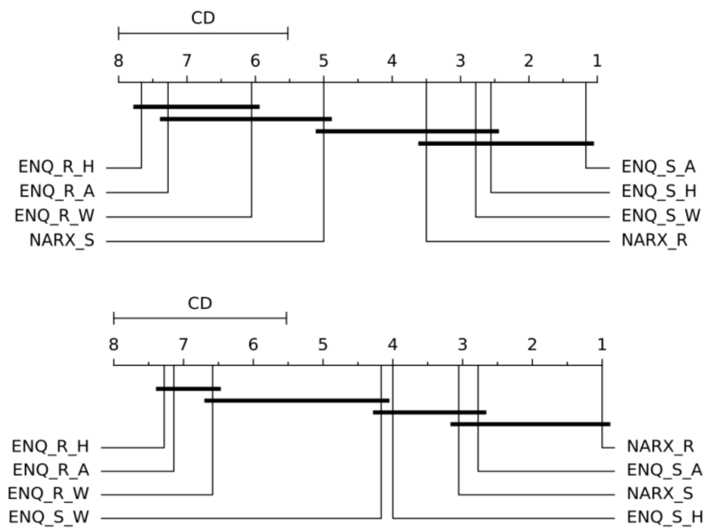

Figure 2: Electric Pricing Nemenyi Test for RMSE (top) and MAPE (bottom)

Table 4 presents the results for the eight techniques evaluated using Electric load and pricing datasets described in Section 3.2. The best value for each dataset is in **boldface.** A comparative study of the ensemble combining techniques was performed. As illustrated in Figure 1 and 2, there exists no statistically significant difference in performance among the ensemble combining techniques: averaging, weighted averaging, and additive. As such, the basic averaging technique may be ideal compared to the other two combining techniques.

A comparative study of the base regressors of the NARX model (SVR and unconstrained least squares) was performed. As presented in Table 4, the least square technique outperformed the SVR technique on both electric load and pricing datasets. The decomposed IMFs were easily modeled using least squares technique which entails that the data is from the same data generating process.

Given that each State was considered separately, for Vic, NARX_R obtained the best RMSE whereas the proposed ensemble ENQ _R_A obtained the best MAPE. For SA, the proposed ensemble technique ENQ_R_A yielded outstanding performance for all datasets. The NARX_R outperformed all other techniques for the NSW dataset on RMSE whereas ENQ_R_A yielded the best performance on MAPE. As observed for the NSW dataset, the same was also observed for QLD and TAS. The NARX_R obtained the best RMSE and ENQ _R_A the best MAPE for QLD and TAS datasets.

Table 4: Prediction results for Electric Load and Pricing datasets

Dataset	NARX_R	NARX_S	ENQ_R_W	ENQ_R_A	ENQ_R_H	ENQ_S_W	ENQ_S_A	ENQ_S_H
Electric Load								
VIC_2013_RMSE	**9,73e-23**	1,87e-01	9,94e-12	7,77e-12	8,39e-12	3,38e+02	3,23e+02	3,25e+02
VIC_2013_MAPE	1,39e-05	9,32e-05	1,22e-05	**1,79e-06**	2,09e-06	4,10e-02	4,00e-02	4,01e-02
VIC_2014_RMSE	2,64e-06	8,01e-01	4,59e-12	**3,91e-12**	4,00e-12	3,43e+02	3,40e+02	3,39e+02
VIC_2014_MAPE	1,62e-15	4,73e-05	**6,74e-16**	5,78e-16	5,88e-16	4,49e-02	4,37e-02	4,39e-02
VIC_2015_RMSE	**1,34e-22**	4,00e-04	7,82e-12	3,36e-12	3,93e-12	3,26e+02	3,14e+02	3,22e+02
VIC_2015_MAPE	7,44e-16	5,50e-05	1,20e-15	3,51e-15	**5,23e-16**	4,59e-02	4,51e-02	4,54e-02
SA_2013_RMSE	1,68e-05	8,47e-01	**2,06e-12**	**2,06e-12**	2,13e-12	1,53e+02	1,47e+02	1,48e+02
SA_2013_MAPE	6,31e-03	4,22e-02	**1,17e-15**	**1,17e-15**	1,21e-15	8,03e-02	8,01e-02	8,02e-02
SA_2014_RMSE	2,08e-01	2,03e-01	2,92e-12	**2,59e-12**	2,61e-12	1,27e+02	1,22e+02	1,22e+02
SA_2014_MAPE	1,96e-02	3,59e-02	1,64e-15	**1,48e-15**	1,49e-15	8,10e-02	8,09e-02	8,09e-02
SA_2015_RMSE	1,99e-01	5,49e-01	2,71e-12	**2,16e-12**	2,53e-12	1,38e+02	1,29e+02	1,32e+02
SA_2015_MAPE	2,01e-02	4,02e-02	1,53e-04	**4,89e-05**	5,06e-05	8,12e-02	8,08e-02	8,11e-02
NSW_2013_RMSE	**2,35e-22**	5,93e-01	7,50e-12	5,70e-12	5,77e-12	8,52e+02	8,46e+02	8,44e+02
NSW_2013_MAPE	9,33e-16	5,22e-02	7,23e-16	**5,47e-16**	5,67e-16	7,57e-02	7,45e-02	7,45e-02
NSW_2014_RMSE	**3,88e-23**	2,50e-01	5,80e-12	2,38e-12	3,83e-12	8,64e+02	8,45e+02	8,50e+02
NSW_2014_MAPE	8,51e-16	3,88e-02	6,82e-16	**5,28e-16**	5,83e-16	8,00e-02	7,79e-02	7,79e-02
NSW_2015_RMSE	**2,10e-22**	6,24e-01	1,30e-11	5,53e-12	5,77e-12	8,56e+02	8,50e+02	8,52e+02
NSW_2015_MAPE	1,21e-15	4,36e-02	1,21e-15	**5,33e-16**	5,50e-16	7,51e-02	7,52e-02	7,53e-02
QLD_2013_RMSE	**1,22e-22**	3,01e-01	5,91e-12	5,42e-12	5,46e-12	2,53e+02	2,50e+02	2,50e+02
QLD_2013_MAPE	1,14e-15	1,29e-06	8,29e-16	**6,72e-16**	7,64e-16	3,58e-02	3,48e-02	3,50e-02
QLD_2014_RMSE	**9,91e-23**	1,33e-01	6,57e-12	3,09e-12	3,23e-12	1,49e+02	1,37e+02	1,36e+02
QLD_2014_MAPE	1,15e-15	3,53e-05	9,14e-16	**4,34e-16**	4,50e-16	1,47e-02	1,43e-02	1,44e-02
QLD_2015_RMSE	**7,87e-23**	2,68e-01	7,33e-12	4,76e-12	4,69e-12	2,91e+02	2,72e+02	2,67e+02
QLD_2015_MAPE	8,20e-16	6,10e-03	2,94e-15	6,75e-16	**6,65e-16**	3,71e-02	3,66e-02	3,65e-02
TAS_2013_RMSE	**7,93e-24**	4,10e-01	1,30e-12	8,80e-13	8,67e-13	6,72e+01	6,41e+01	6,26e+01
TAS_2013_MAPE	7,10e-05	1,45e-02	9,17e-16	2,70e-13	**6,03e-16**	3,68e-02	3,59e-02	3,58e-02
TAS_2014_RMSE	**1,87e-24**	3,36e-01	1,23e-12	6,13e-13	6,47e-13	6,95e+01	6,84e+01	6,73e+01
TAS_2014_MAPE	7,97e-16	2,04e-02	8,50e-16	**4,27e-16**	4,43e-16	3,27e-02	3,24e-02	3,41e-02
TAS_2015_RMSE	**3,00e-24**	8,81e-02	1,27e-12	6,47e-13	6,73e-13	6,02e+01	6,01e+01	5,95e+01
TAS_2015_MAPE	8,97e-16	3,29e-02	8,70e-16	**4,47e-16**	4,60e-16	3,53e-02	3,46e-02	3,48e-02
Electric pricing								
NSW_RMSE	0.0001	5.52e-07	5.26e-07	3.95e-07	**3.88e-07**	0.0040	0.00379	0.0035
NSW_MAPE	3.2e+02	0.0049	0.0003	0.0003	**0.0003**	0.0329	0.0328	0.0324

Figure 3 is a graphical illustration of the prediction performance of the EMD-NARX-QPSO variants for the last 30 days for the NSW 2015 Electric load dataset and for the last 30 days for the NSW Electric pricing dataset. As illustrated in Figure 3, the prediction performance of all variants was very high, nearly resembling the actual values for both NSW Electric load and pricing datasets.

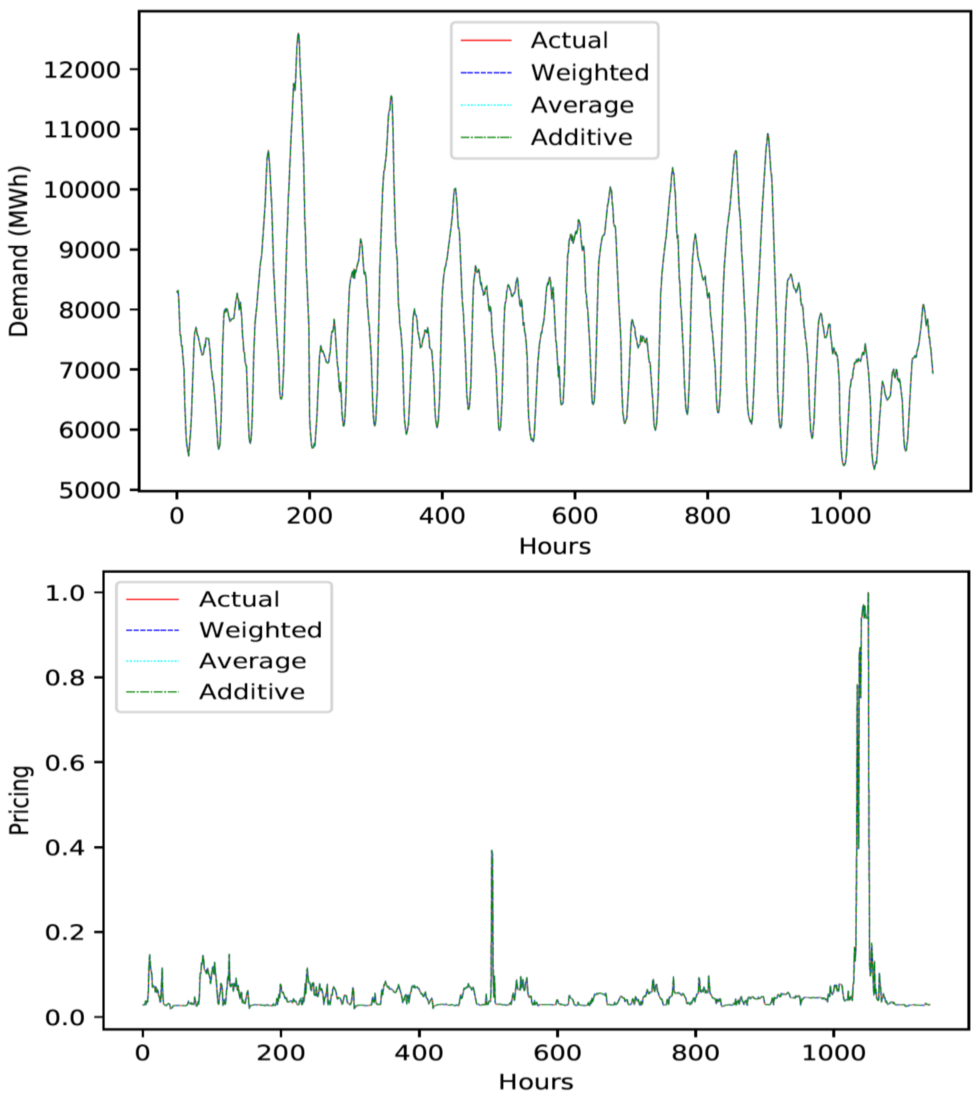

Figure 3: Prediction performance of the proposed technique on Electric Load and Pricing

Figure 4 illustrates the computational time of the base regressors and the EMD-NARX-QPSO variants for NSW and TAS load datasets. The simple regression-based technique, NARX_R obtained the least computational time (fastest). The computational performance of NARX_S was superior to all the ensembles. As expected, the ensembles were computational intensive due to the extra load introduced by EMD decomposition and independent IMF modeling. However, the NARX_R-based ensembles obtained favorable performance due to the computational speed of NARX_R.

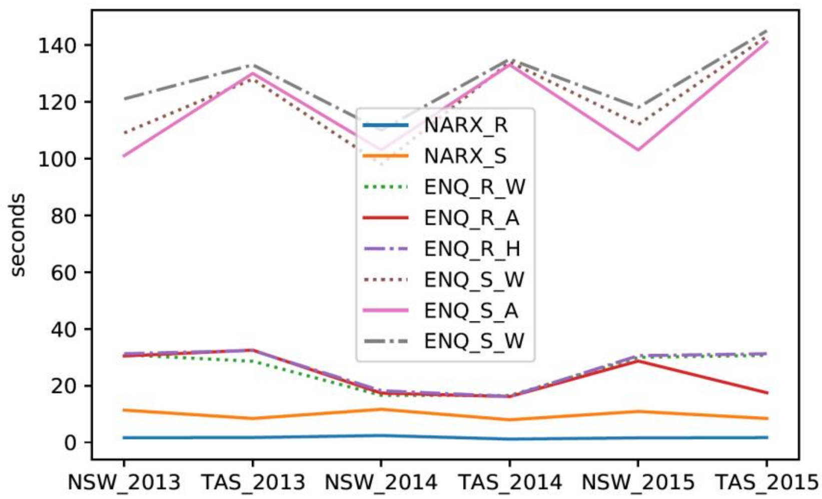

Figure 4: Computational Time for NSW and TAS Load Datasets

The best performing EMD_NARX variant, ENQ_R_W was compared to the benchmark forecasting techniques: general linear model-based load forecaster-benchmark (GLMLF-B) and Persistence [5]. Table 5 presents the obtained results of the experiments performed as described in [5]. As observed in the results presented in Table 5, ENQ_R_W, outperformed the benchmark techniques by a significant margin in all cases.

A comparative study was carried out to ascertain the performance of the proposed model with state-of-the-art techniques. Table 6 presents the results of the best performing variant, EMD-NARX-QPSO_R_A, and the EMD-based ensemble state-of-the-art techniques. The experiments were carried out as discussed in the literature [5].

The results presented in Table 6 show that the proposed ensemble technique, EMD-NARX-QPSO outperformed all state-of-the-art techniques on both RMSE and MAPE. Most of the state-of-the-art techniques were built based on the false stationarity assumption. As such, the induced predictive model usually performs sub-optimally at best or fails at worst if a drift is present in the data. However, the outstanding performance of EMD-NARX-QPSO is attributed to its ability to adapt the model whenever a drift is detected.

Thus, if a time series is decomposed into its equivalent IMFs, then simple regression modeling techniques such as the least square technique can be used to model each IMF to yield a forecasting model of very high precision.

Table 5: Prediction results for comparative study using Forecasting Benchmark Techniques

Dataset	NARX_R	NARX_S	ENQ_R_W	GLMLF-B	Persistence
Electric Load					
VIC_2013_RMSE	481.68	401.28	**398.18**	584.13	781.68
VIC_2013_MAPE	5.91	3.75	**2.84**	9.16	10.71
VIC_2014_RMSE	391.71	428.64	**361.25**	621.93	719.40
VIC_2014_MAPE	6.79	7.58	**5.86**	10.08	11.25
VIC_2015_RMSE	398.63	432.61	**325.37**	684.13	874.69
VIC_2015_MAPE	6.98	8.01	**4.88**	10.16	12.89
SA_2013_RMSE	153.26	201.93	**138.68**	167.37	206.10
SA_2013_MAPE	9.78	11.60	**9.53**	10.12	12.62
SA_2014_RMSE	141.93	198.74	**118.14**	179.00	242.36
SA_2014_MAPE	9.70	10.28	**7.98**	11.66	14.32
SA_2015_RMSE	179.68	418.52	**132.64**	244.25	365.14
SA_2015_MAPE	9.74	4.55	**5.91**	12.68	18.40
NSW_2013_RMSE	400.18	411.68	**357.85**	643.16	901.51
NSW_2013_MAPE	4.00	3.99	**2.87**	6.43	8.66
NSW_2014_RMSE	412.69	497.92	**328.50**	632.28	878.07
NSW_2014_MAPE	3.67	5.37	**2.39**	6.07	8.60
NSW_2015_RMSE	443.54	283.49	**385.31**	713.83	1055.41
NSW_2015_MAPE	4.80	3.07	**3.74**	6.67	9.69
QLD_2013_RMSE	221.96	343.16	**202.96**	355.50	492.59
QLD_2013_MAPE	2.85	4.27	**2.89**	4.33	6.35
QLD_2014_RMSE	275.00	328.19	**281.38**	399.91	588.71
QLD_2014_MAPE	3.34	4.47	**3.73**	5.07	7.14
QLD_2015_RMSE	286.70	326.38	**249.31**	369.41	553.08
QLD_2015_MAPE	3.40	3.53	**3.27**	4.75	6.63
TAS_2013_RMSE	57.09	67.90	**51.06**	84.10	97.86
TAS_2013_MAPE	4.48	5.37	**3.68**	6.04	6.87
TAS_2014_RMSE	57.81	72.68	**49.81**	82.19	86.86
TAS_2014_MAPE	4.61	5.27	**3.99**	6.32	6.52
TAS_2015_RMSE	58.61	68.93	**59.23**	75.81	83.99
TAS_2015_MAPE	4.46	5.34	**3.79**	4.30	5.45

Table 6: Prediction results for comparative study using NSW_2015 dataset

Dataset	DWT_EMD_RFVL	EMD_RFVL	EMD_RF	EMD_SLFN	EMD_NARX_QPSO
Oct_RMSE	193.80	403.27	428.39	379.87	**112.83**
Oct_MAPE	0.0186	0.0342	0.0333	0.0332	**0.0101**
Jul_RMSE	212.70	411.82	441.05	400.46	**123.65**
Jul_MAPE	0.0203	0.0386	0.0397	0.0403	**0.0137**
Apr_RMSE	296.74	423.29	402.97	440.43	**204.47**
Apr_MAPE	0.0296	0.0442	0.0414	0.0472	**0.0117**
Jan_RMSE	659.41	987.33	911.77	913.70	**417.05**
Jan_MAPE	0.0593	0.0704	0.0682	0.0722	**0.0338**

In some cases, the performance of EMD-NARX-QPSO was in the same error range with NARX-QPSO which implies that signifinance performance improvement was evident for complex time series.

The results from the previous studies shows that Electric load datasets exhibits nonlinearity and nonstationary behavior [5]. The comparative study of EMD-NARX-QPSO to the forecasting benchmark techniques presented in Table 6 indicates that EMD-NARX-QPSO had the ability to capture drifting concepts and adapts the forecasting model accordingly to yield an improved performance. The outstanding performance of EMD-NARX-QPSO could have been attributed to the ability to detect concepts drifts, dynamic adaptation of the model and independent modeling of each IMF. Also, the outstanding performance of EMD-NARX-QPSO compared to the state-of-the-art techniques could have been attributed to the optimal predictive model induced by NARX using least-squares as its base regressor and also, the adaptive incremental learning ability of the proposed model to environmental changes due to concept shifts.

Nevertheless, EMD formalism remains the one of the greatest challenges in the EMD-NARX-QPSO and also, decomposing the times series into its respective IMFs posed a load on resources. Thus, the improved performance was realized at the expense of interpretable, transparency computational complexity of the induced model. Alternatively, deep learning techniques could yield comparative results to EMD-NARX-QPSO. The scope of the study was limited to electricity load and pricing time series data.

5 Conclusion and Future Work

Electric load and pricing time series can be considered as a stream of incoming data; therefore, adaptive incremental learning becomes ideal. One of the most significant challenges in incremental learning is concept drift, which occurs when the statistical properties of the target variable drift in unexpected ways. This work proposed an ensemble technique that combines EMD and NARX-QPSO to capture changes due to concept drifting and inherent nonlinearity in time series. The experiments conducted in this work led to the following conclusions: hybridizing EMD with NARX-QPSO improves predictive accuracy. The performance measure used in the proposed technique promoted the induction of models with lower error dispersion and improved predictive accuracy. The extracted IMFs were modeled to a higher degree of accuracy using a simple base regressor, unconstrained least-squares approximation. The effects of ensemble combining techniques on the overall performance of the induced predictive model were insignificant and the performance of EMD-NARX-QPSO was competitive to the state-of-the-art techniques.

A detailed empirical analysis of EMD-NARX-QPSO can be considered for future work direction such as the effect of the combined performed measure, the cost of prediction errors, and the technique's search capability. Robust combining techniques such as bagging can be considered.

References

[1] Sousa, J. C., Jorge, H. M., & Neves, L. P. (2014). Short-term load forecasting based on support vector regression and load profiling. *International Journal of Energy Research, 38*, 350–362.

[2] Wang, X., & Wang, Y. (2016). A hybrid model of EMD and PSO-SVR for short-term load forecasting in residential quarters. *Mathematical Problems in Engineering*.

[3] Cuaresma, J. C., Hlouskova J., Kossmeier S., & Obersteiner, M. (2004). Forecasting electricity spot-prices using linear univariate time-series models. *Applied Energy, 77*, 87–106.

[4] Gepperth, A., & Hammer, B. (2016). Incremental learning algorithms and applications. *In European symposium on artificial neural networks (ESANN).*

[5] Qiu, X., Suganthan, P. N., & Amaratunga, G. A. (2018). Ensemble incremental learning random vector functional link network for short-term electric load forecasting. *Knowledge-Based Systems, 145*, 182-196.

[6] Dietterich, T. G. (2000). Ensemble methods in machine learning. *In International workshop on multiple classifier systems* (pp. 1-15). Berlin, Heidelberg: Springer.

[7] Zhu, G., Peng, S., Lao, Y., Su, Q., & Sun, Q. (2021). Short-Term Electricity Consumption Forecasting Based on the EMD-Fbprophet-LSTM Method. *Mathematical Problems in Engineering*.

[8] Zhang, X., Lai, K. K., & Wang, S. Y. (2008). A new approach for crude oil price analysis based on empirical mode decomposition. *Energy Economics*, *30*(3), 905-918.

[9] Wei, Y., & Chen, M. C. (2012). Forecasting the short-term metro passenger flow with empirical mode decomposition and neural networks. Transportation Research Part C: Emerging Technologies. *21*(1), 148-162.

[10] Kuranga, C., & Pillay, N. (2021). A Comparative Study of Nonlinear Regression and Autoregressive Techniques in Hybrid with Particle Swarm Optimization for Time-series Forecasting. *Expert Systems with Applications*, 116163.

[11] Fentis, A., Bahatti, L., Tabaa, M., & Mestari, M. (2019). Short-term nonlinear autoregressive photovoltaic power forecasting using statistical learning approaches and in-situ observations. *International Journal of Energy and Environmental Engineering, 10*(2), 189-206.

[12] Hammad, M. A., Jereb, B., Rosi, B., & Dragan, D. (2020). Methods and models for electric load forecasting: A comprehensive review. *Logist. Sustain. Transp*, *11*(1), 51-76.

[13] Preece, A., Harborne, D., Braines, D., Tomsett, R., & Chakraborty, S. (2018). Stakeholders in explainable AI. *arXiv preprint arXiv:1810.00184*.

[14] Gunning, D., & Aha, D. (2019). DARPA's explainable artificial intelligence (XAI) program. *AI magazine*, *40*(2), 44-58.

[15] Kandananond, K. (2012). A comparison of various forecasting methods for autocorrelated time series. *International Journal of Engineering Business Management*, *4*(4).

[16] Ristanoski, G., Liu, W., & Bailey, J. (2013, August). A time-dependent enhanced support vector machine for time series regression. *In Proceedings of the 19th ACM SIGKDD international conference on Knowledge discovery and data mining*, (pp. 946-954).

[17] Hong, W. C., & Fan, G. F. (2019). Hybrid empirical mode decomposition with support vector regression model for short-term load forecasting. *Energies*, *12*(6), 1093.

[18] Hooshmand, R. A., Amooshahi, H., & Parastegari, M. (2013). A hybrid intelligent algorithm-based short-term load forecasting approach. *International Journal of Electrical Power & Energy Systems*, *45*(1), 313-324.

[19] Zhao, J., & Liu, X. (2018). A hybrid method of dynamic cooling and heating load forecasting for office buildings based on artificial intelligence and regression analysis. *Energy and Buildings*, *174*, 293-308.

[20] Rudin, C. (2019). Stop explaining black box machine learning models for high stakes decisions and use interpretable models instead. *Nature Machine Intelligence*, *1*(5), 206-215.

[21] Munawar, M. A., & Ward, P. A. (2007, October). A comparative study of pairwise regression techniques for problem determination. *In Proceedings of the 2007 conference of the center for advanced studies on Collaborative research*, (pp. 152-166).

[22] Lin, L., Guo, F., Xie, X., & Luo, B. (2015). Novel adaptive hybrid rule network based on TS fuzzy rules using an improved quantum-behaved particle swarm optimization. *Neurocomputing*, *149*, 1003-1013.

[23] Blackwell, T. M., & Bentley, P. J. (2002, July). Dynamic search with a charged swarm. *In Proceedings of the 4th annual conference on genetic and evolutionary computation*, (pp. 19-26).

[24] Branke, J. (2012). *Evolutionary optimization in dynamic environments* (Vol. 3). Springer Science & Business Media.

[25] Blackwell, T., & Branke, J. (2004, April). Multi-swarm optimization in dynamic environments. *In Workshops on Applications of Evolutionary Computation* (pp. 489-500). Berlin, Heidelberg: Springer.

[26] Marouani, H., & Fouad, Y. (2019). Particle swarm optimization performance for fitting of Lévy noise data. *Physica A: Statistical Mechanics and Its Applications*, *514*, 708-714.

[27] Kuranga, C., & Pillay, N. (2020). In *Nonlinear regression in dynamic environments using particle swarm optimization* (pp. 133–144). Springer.

[28] Huang, N. E., Shen, Z., Long, S. R., Wu, M. C., Shih, H. H., Zheng, Q., ... & Liu, H. H. (1998). The empirical mode decomposition and the Hilbert spectrum for nonlinear and non-stationary time series analysis. *Proceedings of the Royal Society of London. Series A: mathematical, physical and engineering sciences*, *454(1971*, pp. 903-995.

[29] Lahmiri, S. (2017). Modeling and predicting historical volatility in exchange rate markets. *Physica A: Statistical Mechanics and its Applications*, *471*, 387-395.

[30] Chenxing, W., Kemao, Q., & Feipeng, D. (2016). Regenerated Phase-shifted Sinusoids Assisted Empirical Mode Decomposition.

[31] Zheng, H., Yuan, J., & Chen, L. (2017). Short-term load forecasting using EMD-LSTM neural networks with an XgBoost algorithm for feature importance evaluation. *Energies*, *10*(8), 1168.

[32] Chen, C. F., Lai, M. C., & Yeh, C. C. (2012). Forecasting tourism demand based on empirical mode decomposition and neural network. *Knowledge-Based Systems, 26*, 281-287.

[33] Wu, Z. & N. E. Huang, N. E. (2009). Ensemble empirical mode decomposition: A noise-assisted data analysis method. *Advances in Adaptive Data Analysis, 1*, 1-41.

[34] Qiu, X., Suganthan, P. N., & Amaratunga, G. A. (2017). Empirical mode decomposition based on ensemble deep learning for load demand time series forecasting. *Applied Soft Computing*, *54*, 246-255.

[35] Ren, Y., Zhang, L., & Suganthan, P. N. (2016). Ensemble classification and regression-recent developments, applications, and future directions. *IEEE Computational intelligence magazine*, *11*(1), 41-53.

[36] Guan, B. T. (2014). Ensemble empirical mode decomposition for analyzing phenological responses to warming. *Agricultural and Forest Meteorology*, *194*, 1-7.

[37] Abdoos, A., Hemmati, M., & Abdoos, A. A. (2015). Short-term load forecasting using a hybrid intelligent method. *Knowledge-Based Systems*, *76,* 139-147.

[38] Qiu, X., Suganthan, P. N., & Amaratunga, G. A. (2016, October). Electricity load demand time series forecasting with empirical mode decomposition based random vector functional link network. *In 2016 IEEE International Conference on Systems*, *Man, and Cyber*.

[39] Qiu, X., Suganthan, P. N., & Amaratunga, G. A. (2017). Short-term electricity price forecasting with empirical mode decomposition based ensemble kernel machines. *Procedia Computer Science*, *108*, 1308-1317.

[40] Wang, J., Zhang, W., Li, Y., Wang, J., & Dang, Z. (2014). Forecasting wind speed using empirical mode decomposition and Elman neural network. *Applied soft computing*, *23,* 452-459.

[41] Chevallier, J., Zhu, B., & Zhang, L. (2020). Forecasting Inflection points: Hybrid methods with multiscale machine learning algorithms. *Computational Economics*, 1-39.

[42] Awad, M., & Khanna, R. (2015). Support vector regression. *In Efficient learning machines* (pp. 67-80). Berkeley, CA: Apress.

[43] Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., ... & Duchesnay, E. (2011). Scikit-learn: Machine learning in Python. the Journal of Machine Learning research. *12*, 2825-2830.

[44] Gümüş, D. B., Ozcan, E., & Atkin, J. (2016). An investigation of tuning a memetic algorithm for cross-domain search. *In 2016 IEEE Congress on Evolutionary Computation (CEC)* (pp. 135-142). IEEE.

[45] Ramsami, P., & Oree, V. (2015). A hybrid method for forecasting the energy output of photovoltaic systems. *Energy Conversion and Management, 95*, 406-413.

[46] Demšar, J. (2006). Statistical comparisons of classifiers over multiple data sets. *The Journal of Machine Learning Research, 7*, 1-30.

[47] *Australian energy market operator* (2021). Retrieved from http://www.aemo.com.au/

[48] Harries, M., & Wales, N. S. (1999). *Splice-2 comparative evaluation: Electricity pricing.* Artificial Intelligence Group, School of Computer Science and Engineering. Sydney 2052, Australia: The University of New South Wales.

[49] *Australian bureau of meteorology* (2021). Retrieved from http://www.bom.gov.au/

[50] Zhang, Y., Ji, G., Yang, J., Wang, S., Dong, Z., Phillips, P., & Sun, P. (2016). Preliminary research on abnormal brain detection by wavelet-energy and quantum-behaved PSO. *Technology and Health Care*, *24*(s2), S641-S649.

[51] Larrea, M., Porto, A., Irigoyen, E., Barragán, A. J., & Andújar, J. M. (2021). Extreme learning machine ensemble model for time series forecasting boosted by PSO: application to an electric consumption problem. *Neurocomputing, 452*, 465-472.

[52] Li, W., Jiang, X., Sun, W., Wang, S. H., Liu, C., Zhang, X., ... & Miao, L. (2020). Gingivitis identification via multichannel gray-level co-occurrence matrix and particle swarm optimization neural network. *International Journal of Imaging Systems and Technology, 30*(2), 401-411.

[53] Ribeiro, M. H. D. M., Stefenon, S. F., de Lima, J. D., Nied, A., Mariani, V. C., & Coelho, L. D. S. (2020). Electricity price forecasting based on self-adaptive decomposition and heterogeneous ensemble learning. *Energies*, *13*(19), 5190.