Development of a Robust Identifier for NPPs Transients Combining ARIMA Model and EBP Algorithm

Khalil Moshkbar-Bakhshayesh and Mohammad B. Ghofrani

***Abstract*—This study introduces a novel identification method for recognition of nuclear power plants (NPPs) transients by combining the autoregressive integrated moving-average (ARIMA) model and the neural network with error back- propagation (EBP) learning algorithm. The proposed method consists of three steps. First, an EBP based identifier is adopted to distinguish the plant normal states from the faulty ones. In the second step, ARIMA models use integrated (I) process to convert non-stationary data of the selected variables into stationary ones. Subsequently, ARIMA processes, including autoregressive (AR), moving-average (MA), or autoregressive moving-average (ARMA) are used to forecast time series of the selected plant variables. In the third step, for identification the type of transients, the forecasted time series are fed to the modular identifier which has been developed using the latest advances of EBP learning algorithm. Bushehr nuclear power plant (BNPP) transients are probed to analyze the ability of the proposed identifier. Recognition of transient is based on similarity of its statistical properties to the reference one, rather than the values of input patterns. More robustness against noisy data and improvement balance between memorization and generalization are salient advantages of the proposed identifier. Reduction of false identification, sole dependency of identification on the sign of each output signal, selection of the plant variables for transients training independent of each other, and extendibility for identification of more transients without unfavorable effects are other merits of the proposed identifier.**

***Index Terms*—Auto regressive integrated moving-average (ARIMA), Bushehr nuclear power plant (BNPP), error back propagation (EBP), transient identification.**

I. Introduction

NUCLEAR POWER PLANTS (NPPs) are complex systems which are normally monitored by human. Transients which are generated by failures or faults of the plant systems cause difficulties to correctly interpret the trend of the interacting variables. Transient identification can thus be regarded as a useful support to the plant operators.

Many researchers have so far developed different types of transients identifiers using either model-based or model-free methods [1].

Manuscript received March 03, 2014; revised May 11, 2014; accepted June 02, 2014. Date of publication July 25, 2014; date of current version August 14, 2014.

The authors are with the Department of Energy Engineering, Sharif University of Technology, Tehran 145-65114, Iran (e-mail: moshkbar@energy.sharif.ir; ghofrani@sharif.edu).

Color versions of one or more of the figures in this paper are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/TNS.2014.2329055

In model-based methods, a mathematical model is used to illustrate behavior of the system. However, systems complexity causes difficulty in finding the accurate models, while robustness and uncertainty analysis have also to be considered [2], [3]. In spite of recent studies related to the model-based methods [4], their practical applications are still limited.

Model-free methods are more appropriate for transient identification in NPPs. These methods can be categorized into the following branches [1]:

1) Biological metaphors such as particle swarm optimization (PSO), genetic algorithm (GA), and biological metaphors in combination with the principles of quantum computation such as quantum ant colony optimization (QACO), quantum swarm evolutionary(QSE), and quantum inspired evolutionary algorithm (QEA).
2) Statistical methods such as hidden Markov model (HMM), support vector machine (SVM), and symbolic dynamic filtering (SDF)[5].
3) Fuzzy-based systems.
4) Artificial neural networks (ANNs) including feed-forward error back-propagation neural networks [6], [7], competitive networks [8], [9], localized networks [10], [11], and eventually methods concerned with time dependent data [12], [13].

While most of model-free techniques for transient identification are covered by the above classification, some particular methods may be addressed separately. For example, fault recognition of analogical channels of VVER 1000/440 NPPs was done based on analysis of the amplitude fluctuations of electrical signals at the output of channels [14].

Each method has its own capabilities and limitations [1]. Fuzzy systems usually do not cover adequate transients to provide an appropriate identification method for a real power plant [15]. Also, classification results of fuzzy based systems need to be interpreted [16]. Biological metaphors were applied to optimize NPPs transient identification problem. However, misidentification in presence of a noise and incorrect classification of "don't know" transients are limitations of these identifiers [17]. Difficult training and incorrect classification of "don't know" transients are major limitations of the HMM method [1]. Also, despite practical applications of SVM method, many problems such as the optimization of parameters and the option of kernel function are necessary to be solved [18]. Among the above mentioned methods, ANNs, especially multilayer perceptron (MLP) networks with error back- propagation (EBP) learning algorithm, are extensively used [1]. However, hard training of long temporal dependencies [19], difficulty in recognition of unlabeled transients, and weakness against noisy data are shortcomings of this method.

This study seeks to improve performance of EBP based identifiers. To achieve this, a novel identification method by combination of the autoregressive integrated moving-average (ARIMA) model and EBP algorithm is developed.

The proposed identification method includes three steps. First, an EBP based identifier is adopted to distinguish the plant normal states from the faulty ones. In the second step, time series of selected variables are forecasted by ARIMA models. In the third step, these forecasted time series are fed to the modular identifier which has been developed using the latest advances of EBP learning algorithm [20]. Recognition of transients is based on similarity of its statistical properties to the reference one, rather than the values of input patterns. To analyze the ability of the proposed method, the transients in Bushehr nuclear power plant (BNPP) are identified. BNPP is a Russian type PWR (i.e. WWER-1000).

The proceeding sections are organized as follows. In Section II, ARIMA model and its contribution for transient identification is presented. In Section III, the modular identifier based on EBP algorithm is explained. Section IV describes and illustrates the proposed new identifier. In Section V, BNPP plant variables and operation conditions, extracted from the plant final safety analysis report (FSAR) [21], are studied. Subsequently, in Section VI, the results of BNPP transients identification are presented and discussed. Section VII attends the conclusion.

II. Forecasting Time Series by ARIMA Model

ARIMA is general form of autoregressive moving-average (ARMA) model and is used for prediction of time series where data show evidence of non-stationary [23], [24]. This model is generally referred to ARIMA (p, d, q) where parameters p, d, and q are the order of the autoregressive (AR), Integrated (I), and moving-average (MA) processes.

In this section, short review of ARIMA processes is explained while detailed description is left to Appendix. Moreover, our innovative application of this method for transient identification in NPPs is presented.

A. Integrated Process

ARIMA model uses integrated process to convert non stationary time series to stationary ones where joint probability distribution does not change when shifted in time. Subsequently, forecasting can be done by AR, MA, or ARMA processes. This capability is appropriate to convert non-stationary time series of the plant variable to stationary ones.

For any time series, integrated process use differencing step of order d to remove non-stationary. First and second differences of X_t are given by (1).

$$\begin{cases} d = 1 \rightarrow \mathrm{Y}_t = \Delta X_t = X_t - X_{t-1} \\ d = 2 \rightarrow \mathrm{Y}_t = \Delta^2 X_t = \Delta(\Delta X_t) = X_t - 2X_{t-1} + X_{t-2} \end{cases} \quad (1)$$

Where, $\{Y_t\}$ is stationary time series. The dth difference of X_t can be resulted and is given by (2).

$$\mathrm{Y}_t = \Delta^d X_t = \sum_{i=0}^{d} (-1)^i \binom{d}{i} X_{t-i} \quad (2)$$

B. Autoregressive Process

The autoregressive process indicates that the output variable depends on its own previous values [25]. The autoregressive process of order p, AR (p), for stationary time series $\{Y_t\}$ is defined by (3).

$$Y_t = c + \varphi_1 Y_{t-1} + \varphi_2 Y_{t-2} + \ldots + \varphi_p Y_{t-p} + \varepsilon_t \quad (3)$$

Where, $\varphi_1, \ldots, \varphi_p$ are the parameters, c is constant, and ε_t is white noise defined by (4).

$$\mathrm{Covariance}(\varepsilon_t, \varepsilon_s) = 0 \, for \; all \; t \neq s \quad (4)$$

The standard deviation of ε_t is given by (5).

$$\sigma_\varepsilon^2 = \gamma_0 - \sum_{i=1}^{p} \gamma_i \varphi_i \quad (5)$$

Where, γ is auto-covariance of Y_t as a function of lag.

For example, the order of AR process for the core inlet flowrate in large break loss of coolant accident at the cold-leg (LBLOCA-C) of BNPP is estimated. Bipolar representation of this parameter is shown in Fig. 1 where values are calculated by (6)[20]. The partial autocorrelation function (PACF) as a function of lag accompanying with 95% confidence interval (CI) is presented in Fig. 2. According to this figure, the order of 4 is appropriate for AR process, shown in (6) at the bottom of the page.

C. Moving-Average Process

A moving-average process is conceptually a regression of the current value of the series against current and previous white noise error terms.

$$\frac{\text{Bipolar value} = (2 \times (\text{Real value}) - ((\text{Max value}) + (\text{Min value}))}{(\text{Max value}) - (\text{Min value})} \quad (6)$$

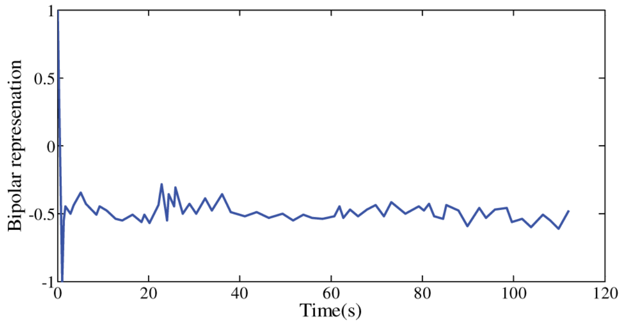

Fig. 1. Bipolar representation of the core inlet flow rate in LBLOCA-C.

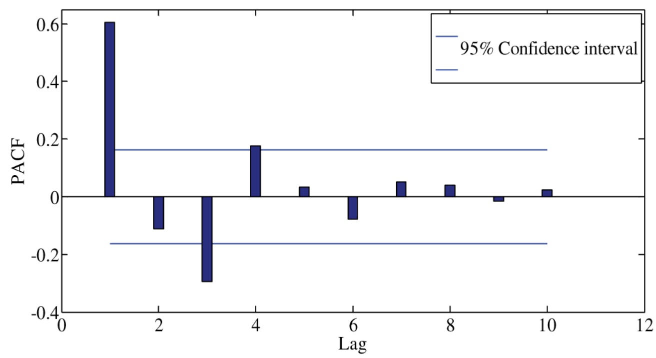

Fig. 2. PACF of the core inlet flow rate in LBLOCA-C as a function of lag.

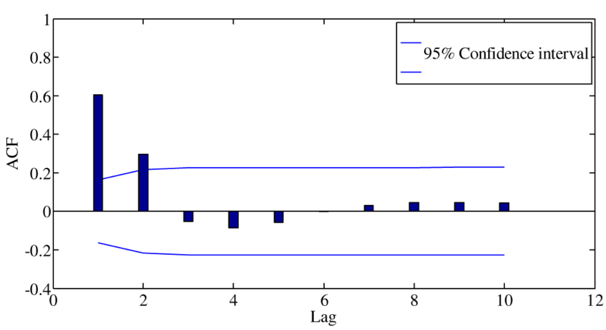

Fig. 3. ACF of the core inlet flow rate in LBLOCA-C as a function of lag.

The moving-average process of order q, MA (q), for stationary time series $\{Y_t\}$ is defined by (7).

$$Y_t = \mu + \varepsilon_t + \theta_1 \varepsilon_{t-1} + \theta_2 \varepsilon_{t-2} + \ldots + \theta_q \varepsilon_{t-q} \tag{7}$$

Where, $\theta_1, \ldots, \theta_q$ are the parameters, μ is the mean of the series.

For example, MA process for the core inlet flow rate in LBLOCA-C of BNPP is calculated. The autocorrelation function (ACF) as a function of lag is shown in Fig. 3.

D. *Autoregressive Moving-Average Process*

Combining AR (p) and MA (q) processes, we can define ARMA (p, q) process by (8).

$$Y_t = \varphi_1 Y_{t-1} \ldots + \varphi_p Y_{t-p} + \varepsilon_t + \theta_1 \varepsilon_{t-1} \ldots + \theta_q \varepsilon_{t-q} \tag{8}$$

After some manipulation, ARMA parameters are calculated.

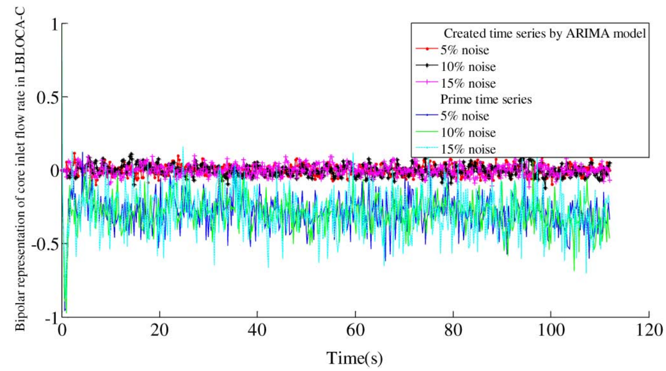

Fig. 4. Primary and created time series for bipolar representation of core inlet flowrate in LBLOCA-C with inserted noise up to 15%.

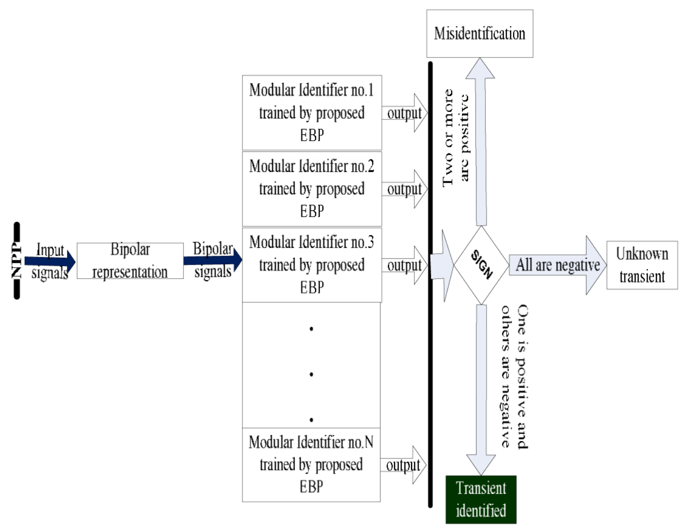

Fig. 5. The schematic view of the EBP based modular identifier.

The ARMA process is more suitable when the PACF or ACF do not have clear cutoff points. Akaike's Final Prediction Error (FPE) and Akaike's Information Criterion (AIC) [26], [27] are two methods which can be used to estimate the order of an ARMA (p, q).

E. *Use of ARIMA Model in Transient Identification*

ARIMA model is used for transient identification in NPPs via following steps:

1) Conversion non-stationary data into stationary ones using integrated process,
2) Developing AR, MA, or ARMA processes with unknown parameters,
3) Calculation of unknown parameters,
4) Estimation the order of model,
5) Selection of appropriate process (i.e. AR, MA, or ARMA),
6) Creation of new time series using selected process of step 5.

ARIMA model is furthermore derived as a case, for bipolar representation of the core inlet flow rate in LBLOCA-C transient. The performance of this application, illustrated in Fig. 4 is more robustness of the created time series against inserted noise. Therefore, feeding the created time series as input signals

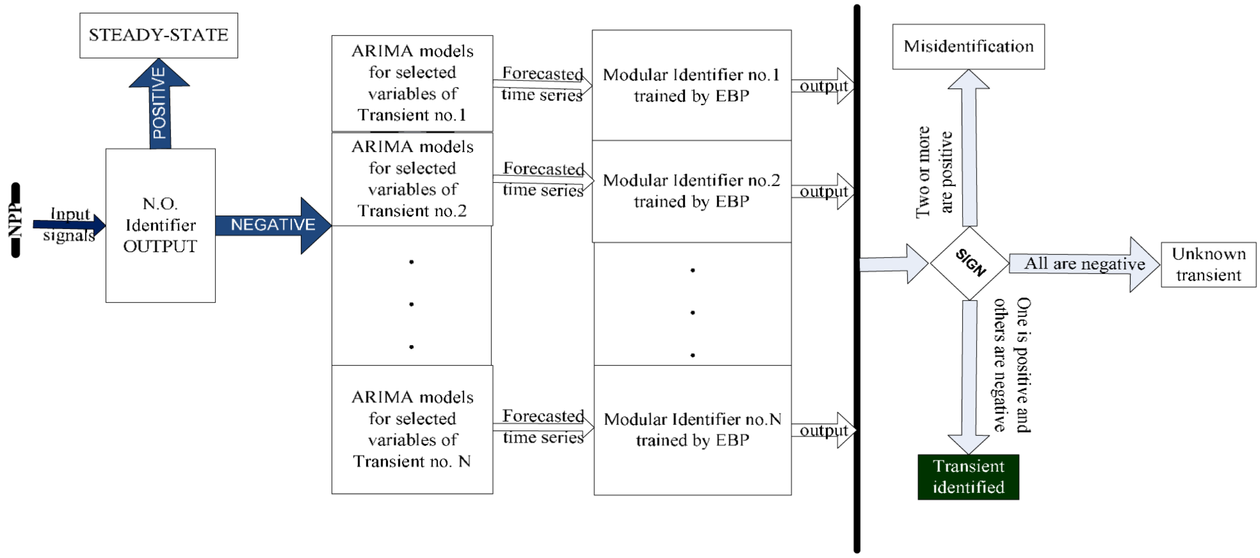

Fig. 6. Schematic view of the proposed new identifier.

to the modular EBP based identifier increases the percentage of correct identification in noisy environments.

The noise is manually superimposed on the input data using zero-mean normal distribution. The terminology of "N% noise" signifies that the variance (power) of normal distribution is equal to the variance (power) of uniform distribution with two boundaries $-\mathrm{N}/100$ and $+\mathrm{N}/100$. The standard deviation of applied zero-mean normal distribution is given by (9). Symbols a and b are the boundaries of uniform distribution.

$$\sigma_{normal} = \sqrt{\frac{(b-a)^2}{12}} \tag{9}$$

The equation for producing new time series in Fig. 4 is given by (10).

$$Y_t = PACF(1)*Y_{t-1} + PACF(2)*Y_{t-2} + PACF(3)*Y_{t-3} + PACF(4)*Y_{t-4} + N \tag{10}$$

where, PACFs are presented in Fig. 2. $\{Y_t\}$ is new created time series, and N is white noise with variance calculated by (5).

For better representation, the created time series is intentionally shifted upward.

III. The Modular EBP Based Identifier

An identifier for NPPs transients training and recognition based on the latest advances of EBP algorithm was previously developed and introduced [20].

Elements of EBP including input data, initial weights, learning rate, cost function, activation function, and weights updating procedure were investigated and an efficient neural network was developed.

Furthermore, a set of simple neural networks, one for each transient, was used as a pattern recognition system.

Using the modular EBP based identifier led to the following advantages:

1) Quick identification of transients occurred with few number of plant variables, while in previous EBP based identifiers, a larger number of variables (typically more than 15 variables) were needed [28].
2) Bipolar targets and non-saturated property of log function made possible to choose greater error in the training process. Therefore, the balance between memorization and generalization was improved.
3) Because of bipolar representation of targets, it was possible to train the network with a positive output for the specified transient and a negative one for all others. Therefore, identification only hung on the sign of each network output and qualitative consideration became unnecessary.
4) Choose of modular identifier made it possible to extend the number of transients without unfavorably affecting the identifier design.

Schematic view of the EBP based modular identifier is presented in Fig. 5.

IV. The Proposed New Identifier

Schematic view of the proposed new identifier is presented in Fig. 6 performing its function in three steps. First, an EBP based identifier is adopted to distinguish the normal operation (N.O.) from transients. In the second step, ARIMA models use integrated (I) process to convert non-stationary data of the selected variables into stationary ones. Subsequently, ARIMA processes, including AR, MA, or ARMA are used to create new time series for the selected variables of the target transients. In the third step, created time series are fed to the modular EBP based identifier, to identify the type of transients.

Bipolar representation of targets in N.O. identifier makes possible to train the network with positive number for the normal operation and negative number for all others. Therefore, recognition only depends on the sign of the network outputs.

ARIMA model is usually used for forecasting linear component of time series. In this research, ARIMA model is used as a

TABLE I
MAIN CHARACTERISTICS OF BNPP

Reactor nominal thermal power, MW	3000
Coolant flowrate m^3/h	84800
Coolant pressure at the core outlet, MPa	15.7±0.03
Coolant temperature at the reactor inlet, °C	291±3
Coolant temperature at the reactor outlet, °C	321±5
Coolant heating in the reactor, °C	30
Pressure differential in the reactor, MPa	0.381
Number of loops	4
Fuel height in the core in cold state, m	3.53
Flow area of the core, m^2	4.14
Number of FA in the core	163
Maximum linear heat rate, W/cm	448
Steam generator steam capacity, t/h	1470^{+103}
Steam pressure in the steam generator steam header, MPa	6.27±0.1
Temperature of generated steam, °C	278.5
Boric acid concentration, g H_3BO_3/kg H_2O	16…20

TABLE II
LIST OF THE TARGET TRANSIENTS

No.	Transient
1	Large break loss of coolant accident at cold-leg (LBLOCA-C)
2	Main steam line break (MSLB)
3	Instantaneous jamming of one reactor coolant pump set (IJRCP)
4	Steam generator feed- water line break (SGFWLB)
5	Trip of all four reactor coolant pump sets (TRCP)
6	Uncontrolled withdrawal of control rods (UWCR)

TABLE III
LIST OF THE SELECTED PLANT VARIABLES FOR THE TARGET TRANSIENTS

Transient	Selected variables
LBLOCA-C	Pressurizer pressure
	Core outlet pressure
	Coolant temperature at core inlet
	Core inlet flow rate
	Pressurizer water level
MSLB	Feed-water flow rate
	Steam-generator pressure
	Steam-generator steam flow
	Steam-generator water level
	Pressurizer water level
IJRCP	Core outlet pressure
	Core inlet flow rate
	Coolant temperature at core inlet
	Coolant temperature at core outlet
SGFWLB	Feed-water flow rate
	Steam-generator pressure
	Steam-generator steam flow
	Steam-generator water level
	Pressurizer water level
TRCP	Core outlet pressure
	Core inlet flow rate
	Coolant temperature at core inlet
	Coolant temperature at core outlet
UWCR	Relative thermal power
	Pressurizer water level
	Coolant temperature at core inlet
	Coolant temperature at core outlet

preprocessor, to create specific time series for the selected variables of the target transients. These time series are then fed to the modular EBP based identifier for transient identification. The main advantage of such combination is more robustness of the created time series against inserted noise to the primary time series, thus increasing percentage of correct identification in noisy environments.

The above design makes possible to have an efficient identifier with the most important characteristics (i.e. autocorrelation finding, cross-correlation detection and proximity measure). Autocorrelation of the plant variables is forecasted by ARIMA model, while EBP based identifier detects cross-correlation of input data.

Moreover, bipolar targets and non-saturated property of output function of modular EBP based identifier [20] make possible to choose greater error in the training process. Therefore, the balance between memorization and generalization which performs as proximity measure between new transient and the reference one is improved.

In the next section, the proposed identifier is applied for identification of BNPP transients.

V. CASE STUDY: TRANSIENTS OF BUSHEHR NUCLEAR POWER PLANT

In this section, BNPP transients are selected to show the ability and performance of the proposed identifier. *BNPP is a* water-moderated reactor type namely WWER-1000 (V-446). The main characteristics of this reactor are presented in Table I.

Furthermore, the plant variables, as well as input data needed for transients training are discussed.

A. Selection of Bushehr Nuclear Power Plant Transients

BNPP transients are selected based on the following criteria:

1) Coverage of the reactor core, primary and secondary loops transients.
2) Coverage of both types of anticipated operational occurrence (AOO) and design basis accident (DBA) transients.

Beyond design basis accidents (BDBA) are not considered since, for this class of accidents, mitigative countermeasures are implemented to reduce the core damage and radiological consequences.

List of the target transients is presented in Table II.

B. Selection of the Plant Variables for Transients Training

In recent years, several methods were developed for selection of the plant variables which are the most sensitive to the faults and malfunctions in transient classification [1], [29]. The main purpose of these methods is to decrease the number of common plant variables where, in NPPs, hundreds of parameters are monitored. Our innovative approach using modular identifier makes possible to identify the type of transients with few number of the plant variables and independent of each other. Moreover, selection of the plant variables is done according to expert judgment based on importance of each variable for identification of a specific transient. This would facilitate the recognition and increase performance of the identifier.

The selected BNPP variables are listed in Table III.

C. Input Data for Transients Training

Training is usually performed using simulator data. The main advantage of simulators is real-time analysis achieved by simple

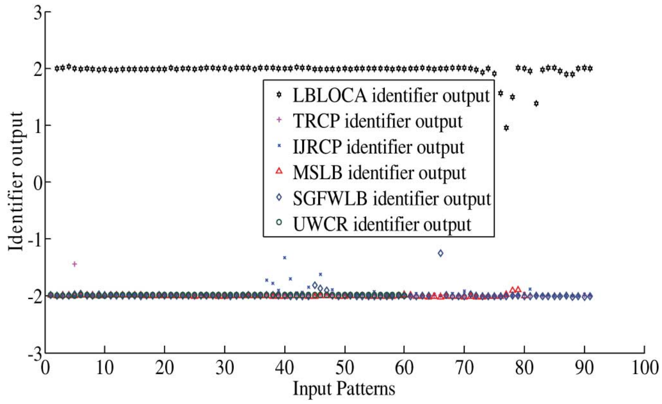

Fig. 7. Identification of large break loss of coolant accident at cold-leg.

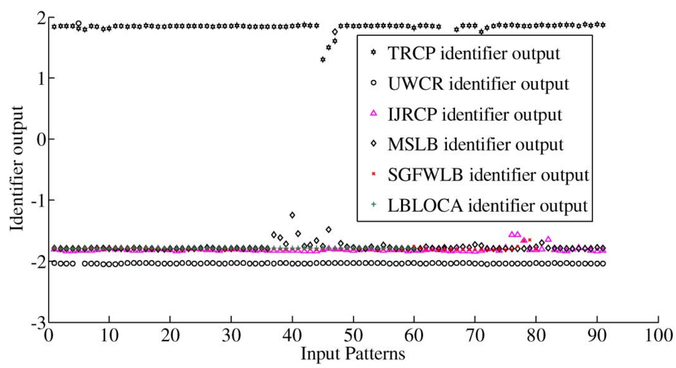

Fig. 8. Identification of trip of all four reactor coolant pump sets.

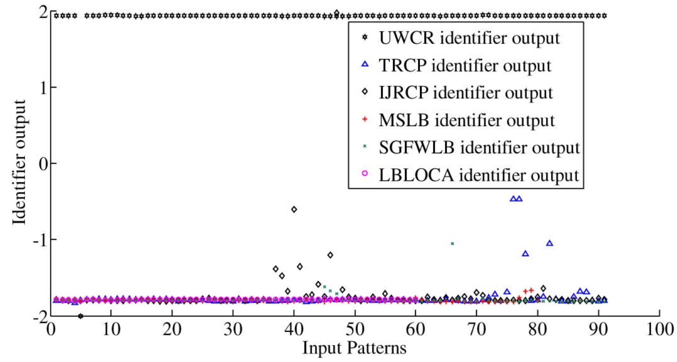

Fig. 9. Identification of uncontrolled withdrawal of control rods.

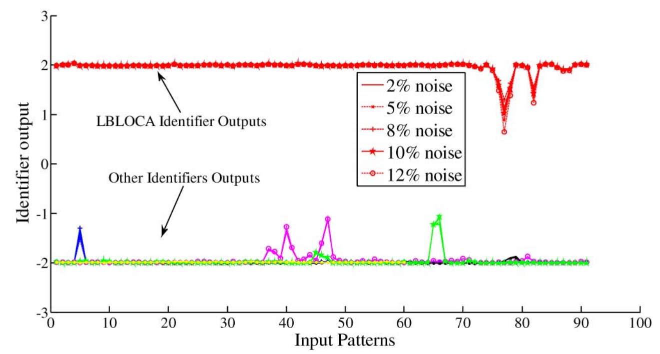

Fig. 10. Identification of large break loss of coolant accident at cold-leg with noisy input data by the proposed identifier.

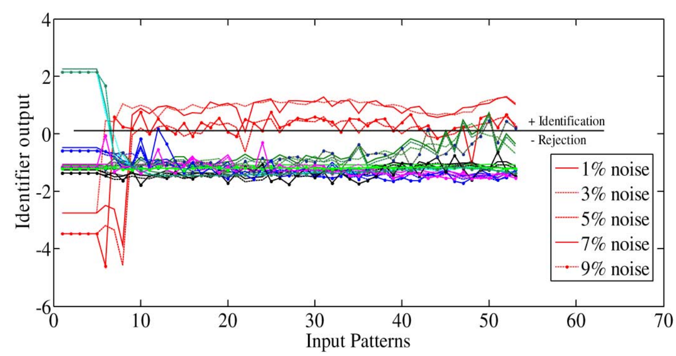

Fig. 11. Identification of large break loss of coolant accident at cold-leg without ARIMA model.

thermal-hydraulic and neutronic models. However, this simplicity would decrease the precision of analysis. Extracted data from FSAR of the BNPP is more reliable than simulator data according to the following reasons:

1) Results of the code package used for transient analysis in FSAR had been verified on the basis of experimental data arisen from preoperational work on WWER-1000 reactor type [21].
2) Interrnational benchmark problems and alternative system codes (such as RELAP, ATHLET) were used to verify results of the code package [21].

After creation of new time series by ARIMA model, approximately 500 input signals at intervals of 1/2 second are employed for transients training.

VI. Results and Discussion

In this section, the results of the proposed identifier for identification of BNPP transients are presented and discussed. The results for various plant transients, i.e. LBLOCA-C, trip of all four reactor coolant pump sets (TRCP) and uncontrolled withdrawal of control rods (UWCR) are demonstrated in Fig. 7, Fig. 8, and Fig. 9 respectively. The marker represents the output of each modular network for the input patterns. As seen from these figures, each trained identifier recognizes its related transient distinctively.

Noise superimposition resulted from the difference between real plant conditions and one assumed for training can affect the confidence level of any transient identifier. By the proposed identifier, recognition is performed using similarity of statistical properties of any new transient (i.e. statistical properties of variables) to the reference one. Robustness against noisy input data is thus increased. Performance of identification with and without ARIMA is presented by Fig. 10 and Fig. 11, respectively. This comparison clearly shows prominence of the proposed identifier in noisy environments.

Also, Balance between correct answer to the trained transients (i.e. memorization) and reasonable response to the new transients (i.e. generalization) as one the primary design criteria of any identifier can be improved. As an example, Fig. 12 presents this ability in recognition of large break loss of coolant accident at the hot-leg (LBLOCA-H).

Using the proposed identifier leads to the following advantages:

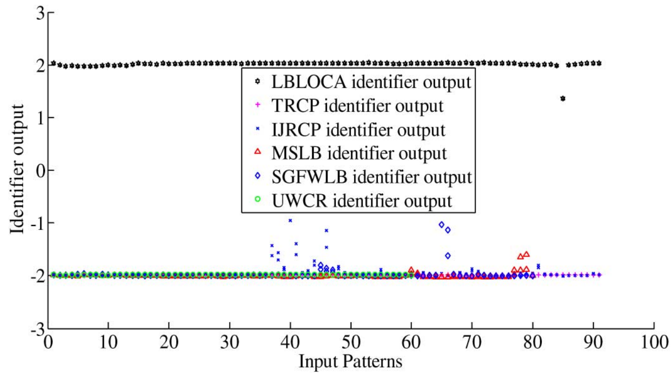

Fig. 12. Identification of large break loss of coolant accident at hot-leg.

TABLE IV
PERCENTAGE OF CORRECT IDENTIFICATION OF THE TARGET TRANSIENTS WITH NOISY DATA BY THE PROPOSED IDENTIFIER IN COMPARISON WITH OTHER TYPES OF EBP BASED IDENTIFIERS

Target transients	The proposed identifier with noisy data up to 12%	EBP based identifiers with noisy data up to 10%		
		Modular	Feed-forward	Recurrent
LBLOCA-C	100	66	40	56
MSLB	96	62	38	48
IJRCP	96	64	44	56
SGFWLB	92	78	56	60
TRCP	94	64	48	58
UWCR	100	100	68	68

1) Recognition of transients based on statistical properties of input patterns increases robustness against noisy data.
2) Using autocorrelation of plant variables makes possible to identify fast transients
3) Dependency to statistical properties rather than value of input patterns increases generalization
4) Plant variables for transients training can be selected independent of each other, while in the previous studies it was necessary to select common plant variables.
5) For transient identification, few number of the plant variables is enough, while in previous EBP based identifiers a larger number of variables (typically more than 15 variables) were needed [28].
6) Transient is recognized only by the sign of each classifier output.
7) Choose of modular identifier makes it possible to extend the number of transients without unfavorably affecting the identifier design.

Table IV, presents percentage of correct identification of transients with noisy data in comparison with other types of EBP based identifiers (i.e. modular, feed-forward, and recurrent networks). The results confirm more robustness of the proposed identifier in noisy environments.

Difficulty in modeling time dependent data is one of the major challenges of EBP based identifiers. Recurrent networks are usually used to overcome this challenge. However, recurrent networks are deficient in training of long temporal dependencies. In this research, we tried to tackle this problem by using ARIMA model, where there is no need to train the network with recurrent input data. Qualitative comparison of the proposed identifier with other types of EBP based ones is given in Table V. In this table, A, I, and T stand for "appropriate", "inappropriate", and "tolerable", respectively. A quick look to the abilities and weaknesses of different transient identifiers reveals noticeable performance of the proposed identifier.

TABLE V
QUALITATIVE COMPARISON OF THE PROPOSED IDENTIFIER WITH OTHER TYPES OF EBP BASED IDENTIFIERS

Transient identifier	The proposed identifier	EBP based identifiers		
		Modular	Feed-forward	Recurrent
Training of temporal data	A	I	I	A
Training of long temporal data	A	I	I	I
Identification of large number of transients	A	A	A	A
Identification of fast transients	A	I	I	A
Identification of untrained transients	T	T	T	T
Identification by few number of variables	A	A	T	T
Balance between memorization and generalization	A	T	T	T
Identification in noisy environments	A	T	T	T

Because of supervised approach of the proposed identifier, we will further investigate identification of transients out of collected knowledge (i.e. unlabeled transients).

VII. CONCLUSION

In this study, we developed a robust modular identifier for NPPs transients, by combination of ARIMA model and EBP algorithm. ARIMA model forecasts time series of the selected plant variables which provides input patterns for an EBP based identifier, used to recognize the type of transients. As a case study, the proposed identifier was applied successfully for recognition of BNPP transients. Noticeable advantages of the proposed identifier are robustness against noisy data and improvement balance between memorization and generalization. Reduction of false identification, sole dependency of identification on the sign of each output signal, selection of the plant variables independent of each other, and extendibility for identification of more transients without unfavorable effects, are other advantages of the proposed identifier.

Our studies show that an efficient classifier for transient identification in NPPs should have the ability to learn and find autocorrelation and cross-correlation of the plant variables. In particular, the identifier needs proximity measure between new transient and the trained one. Autocorrelation of the plant variables may be adequately forecasted by ARIMA models, while the choice of modular identifiers facilitates to detect cross-correlation of input data. Moreover, balance between memorization and generalization performs as proximity measure between new transient and the reference one. Accordingly, the proposed identifier could improve identification of the NPPs transients.

Finally, identification of unlabeled (i.e. untrained) transients will be further investigated in our ongoing research.

Appendix

Appendix: Detailed Description of ARIMA Processes

To calculate AR parameters, Yule-Walker equations defined by (A1) are used.

$$\underbrace{\begin{pmatrix} r_1 \\ r_2 \\ \vdots \\ r_{p-1} \\ r_p \end{pmatrix}}_{\boldsymbol{r}} = \underbrace{\begin{pmatrix} 1 & r_1 & \cdots & r_{p-1} \\ r_1 & 1 & \cdots & r_{p-2} \\ & \vdots & & \\ r_{p-2} & r_{p-3} & \cdots & r_1 \\ r_{p-1} & r_{p-2} & & 1 \end{pmatrix}}_{\boldsymbol{R}} \underbrace{\begin{pmatrix} \varphi_1 \\ \varphi_2 \\ \vdots \\ \varphi_{p-1} \\ \varphi_p \end{pmatrix}}_{\boldsymbol{\Phi}} \quad \text{(A1)}$$

Where, r is autocorrelation of Y_t as a function of lag.

Resultant equation for calculation of AR parameters is given by (A2).

$$\mathbf{\Phi} = \mathbf{R}^{-1}\mathbf{r} \quad \text{(A2)}$$

The PACF can be used to determine the order of AR process. PACF at lag k of stationary time series $\{Y_t\}$ is defined by (A3).

$$\begin{aligned} PACF(k) = corr(&Y_t - L_{\{Y_{t+1},\ldots,Y_{t+k-1}\}}(Y_t), \\ &Y_{t+k} - L_{\{Y_{t+1},\ldots,Y_{t+k-1}\}}(Y_{t+k})) \end{aligned} \quad \text{(A3)}$$

Where, L denotes the best linear predictor of Y_t and Y_{t+k} based on $Y_{t+1}, \ldots, Y_{t+k-1}$.

The idea is to use Yule-Walker equations to solve for successive values of p, which is the order of the AR process. Using Cramer's rule, the PACF can be calculated easily at any lag. The PACF up to lag 3 is given by (A4).

$$\begin{cases} PACF(1) = r_1 \\ PACF(2) = \frac{r_2 - {r_1}^2}{1 - {r_1}^2} \\ PACF(3) = \frac{-2r_1 r_2 - {r_1}^2 r_3 + r_3 + r_1 r_2^2 + r_1^3}{1 + 2r_1^2 r_2 - r_2^2 - 2{r_1}^2} \end{cases} \quad \text{(A4)}$$

If the true order of AR process is p, then φ_{p+1} is approximately normally distributed with mean zero. Therefore, confidence interval (CI) can be used for PACF to estimate the order of AR process.

To calculate MA parameters, ACF of Y_t is used. After some manipulation, MA parameters and standard deviation of white noise are given according to (A5).

$$\begin{cases} \gamma_0 = \sigma_\varepsilon^2(1 + \theta_1^2 + \theta_2^2 + \theta_3^2 \cdots + \theta_q^2) \\ r_1 = (\frac{\theta_1 + \theta_1\theta_2 + \theta_2\theta_3 + \cdots + \theta_{q-1}\theta_q}{1 + \theta_1^2 + \theta_2^2 + \theta_3^2 + \cdots + \theta_q^2}) \\ \qquad \vdots \\ r_k = (\frac{\theta_k + \theta_1\theta_{k+1} + \theta_2\theta_{k+2} + \cdots + \theta_{q-k}\theta_q}{1 + \theta_1^2 + \theta_2^2 + \theta_3^2 + \cdots + \theta_q^2}) \\ \qquad \vdots \\ r_q = (\frac{\theta_q}{1 + \theta_1^2 + \theta_2^2 + \theta_3^2 + \cdots + \theta_q^2}) \end{cases} \quad \text{(A5)}$$

The standard deviation of ACF is given by (A6). This is because the produced CIs are the so called "large-lag" standard errors of ACF [32].

$$\sigma_{ACF}(k) = \sqrt{\frac{1 + 2\sum_{i=1}^{k-1} r_i^2}{N}} \quad \text{(A6)}$$

Where, N is the length of $\{Y_t\}$.

To calculate ARMA parameters, we multiply both sides of (8) by Y_{t-l} and take expectance. Hence,

$$\begin{aligned} \langle Y_{t-l}Y_t \rangle = \langle Y_{t-l}(&\varphi_1 Y_{t-1} + \ldots + \varphi_p Y_{t-p} \\ &+ \varepsilon_t + \theta_1\varepsilon_{t-1} + \ldots + \theta_q\varepsilon_{t-q}) \rangle \end{aligned} \quad \text{(A7)}$$

After some manipulation, (A7) results (A8).

$$\begin{aligned} \langle Y_{t-l}Y_t \rangle = &\sum_{i=1}^{p} (\varphi_i \langle Y_{t-l}Y_{t-i} \rangle) + \langle Y_{t-l}\varepsilon_t \rangle \\ &+ \sum_{i=1}^{q} (\theta_i \langle Y_{t-l}\varepsilon_{t-i} \rangle) \end{aligned} \quad \text{(A8)}$$

Dividing (A8) by length of $\{Y_t\}$ gives (A9).

$$\gamma_l = \sum_{i=1}^{p} \varphi_i\gamma_{i-l} + \delta_{l0}\sigma_\varepsilon^2 + \sum_{i=1}^{q} \theta_i\sigma_\varepsilon^2\delta_{li} \quad \text{(A9)}$$

Where,

$$\delta_{li} = \begin{cases} 1 & l = i \\ 0 & l \neq i \end{cases}$$

l is an integer changing between 0 and p.

Similarly, by multiplying both sides of (8) by ε_{t-k} and take expectance

$$\begin{aligned} \langle \varepsilon_{t-k}Y_t \rangle = &\sum_{i=1}^{p} (\varphi_i \langle \varepsilon_{t-k}Y_{t-i} \rangle) + \langle \varepsilon_{t-k}\varepsilon_t \rangle \\ &+ \sum_{i=1}^{q} (\theta_i \langle \varepsilon_{t-k}\varepsilon_{t-i} \rangle) \end{aligned} \quad \text{(A10)}$$

Dividing (A10) by length of $\{Y_t\}$ gives (A11).

$$0 = \sum_{i=1}^{p} \varphi_i\sigma_\varepsilon^2\delta_{ki} + \sum_{i=1}^{q} \theta_i\sigma_\varepsilon^2\delta_{ki} \quad \text{(A11)}$$

Where, k is an integer changing between 1 and q.

Finally, ARMA parameters are presented by (A12).

$$\begin{cases} \gamma_0 = \sigma_\varepsilon^2 + \sum_{i=1}^{p} \varphi_i\gamma_i \\ \qquad \vdots \\ \gamma_l = \sum_{i=1}^{p} \varphi_i\gamma_{i-l} + \delta_{l0}\sigma_\varepsilon^2 + \sum_{i=1}^{q} \theta_i\sigma_\varepsilon^2\delta_{li} \\ l = \{n \in \mathbb{Z} : 0 \leq n \leq p\} \\ \qquad 0 = \varphi_1 + \theta_1 \\ \qquad \vdots \\ 0 = \sum_{i=1}^{p} \varphi_i\sigma_\varepsilon^2\delta_{ki} + \sum_{i=1}^{q} \theta_i\sigma_\varepsilon^2\delta_{ki} \\ k = \{m \in \mathbb{Z} : 1 \leq m \leq q\} \end{cases} \quad \text{(A12)}$$

References

[1] K. Moshkbar-Bakhshayesh and M. B. Ghofrani, "Transient identification in nuclear power plants: A review," *Progr. Nucl. Energy*, vol. 67, pp. 23–32, 2013.

[2] P. M. Frank and S. X. Ding, "Survey of robust residual generation and evaluation methods in observer-based fault detection systems," *J. Process Control*, vol. 7, pp. 403–424, 1997.

[3] R. Patton and J. Chen, "Observer-based fault detection and isolation: Robustness and applications," *Control Eng. Practice*, vol. 5, pp. 671–682, 1997.

[4] T. Y. Hsiao, C. Lin, and Y. Yuann, "Identification of initiating events for pressurized water reactor accidents," *Annals Nucl. Energy*, vol. 37, pp. 1502–1512, 2010.

[5] X. Jin, Y. Guo, S. Sarkar, A. Ray, and R. M. Edwards, "Anomaly detection in nuclear power plants via symbolic dynamic filtering," *IEEE Trans. Nucl. Sci.*, vol. 58, no. 1, pt. 2, pp. 277–288, 2011.

[6] A. Basu, "Nuclear power plant status diagnostics using a neural network with dynamic node architecture," Iowa State Univ. Sci. Technol., Ames, IA, USA, 1992.

[7] E. B. Bartlett and R. E. Uhrig, "Nuclear power plant status diagnostics using artificial neural networks," Dept. Nucl. Eng., Tennessee Univ., Knoxville, TN, United States, 1991.

[8] Y. Bartal, J. Lin, and R. E. Uhrig, "Nuclear power plants transient diagnostics using LVQ or some networks don't know that they don't know," in *Proc. IEEE Int. Conf. Computational Intell.*, 1994, pp. 3744–3749.

[9] H. Furukawa, T. Uedu, and M. Kitamura, "Use of self-organizing neural networks for rational definition of plant diagnostic symptoms," in *Proc. Topical Meet. Comput.-Based Human Support Syst.: Technol., Methods, Future*, 1995, pp. 441–448.

[10] Y. Bartal, J. Lin, and R. E. Uhrig, "Nuclear power plant transient diagnostics using artificial neural networks that allow "dont-know" classifications," *Nucl. Technol.*, vol. 110, pp. 436–449, 1995.

[11] M. G. Na, S. H. Shin, S. M. Lee, D. W. Jung, S. P. Kim, J. H. Jeong, and B. C. Lee, "Prediction of major transient scenarios for severe accidents of nuclear power plants," *IEEE Trans. Nucl. Sci.*, vol. 51, pp. 313–321, 2004.

[12] C. Lin and H. J. Chang, "Identification of pressurized water reactor transient using template matching," *Annals Nucl. Energy*, vol. 38, pp. 1662–1666, 2011.

[13] D. Roverso, "On-line early fault detection and diagnosis with the aladin transient classifier," in *Proc. PNPIC HMIT, 4th Amer. Nucl. Soc., Int. Topical Meet. Nucl. Plant Instrumentation Control Human-Machine Interface Technol.*, Columbus, OH, USA, Sep. 2004, pp. 19–22.

[14] M. Yazikov, G. Gola, Ø. Berg, J. Porsmyr, H. Valseth, D. Roverso, and M. Hoffmann, "On-line fault recognition system for the analogic channels of VVER 1000/400 nuclear reactors," *IEEE Trans. Nucl. Sci.*, vol. 59, no. 2, pp. 411–418, 2012.

[15] T. Iijima, "Application study of fuzzy logic methods for plant-state identification," 1995.

[16] M. Marseguerra, E. Zio, A. Oldrini, and E. Brega, "Fuzzy identification of transients in nuclear power plants," *Nucl. Eng. Design*, vol. 225, pp. 285–294, 2003.

[17] J. A. Carlos, C. Medeiros, and R. Schirru, "Identification of nuclear power plant transients using the particle swarm optimization algorithm," *Annals Nucl. Energy*, vol. 35, pp. 576–582, 2008.

[18] C. Gottlieb, V. Arzhanov, W. Gudowski, and N. Garis, "Feasibility study on transient identification in nuclear power plants using support vector machines," *Nucl. Technol.*, vol. 155, pp. 67–77, 2006.

[19] D. Roverso, "Plant diagnostics by transient classification: The aladdin approach," *Int. J. Intell. Syst.*, vol. 17, pp. 767–790, 2002.

[20] K. Moshkbar-Bakhshayesh and M. B. Ghofrani, "Development of an efficient identifier for nuclear power plant transients based on latest advances of error back-propagation learning algorithm," *IEEE Trans. Nucl. Sci.*, vol. 61, no. 1, pt. 3, pp. 602–610, 2014.

[21] BNPP, "Final safety analysis report (FSAR), Chapter 15, Rev.0," 2003.

[22] D. Dubois and H. Prade, "An introduction to bipolar representations of information and preference," *Int. J. Intell. Syst.*, vol. 23, pp. 866–877, 2008.

[23] J. J. Commandeur and S. J. Koopman, *An Introduction to State Space Time Series Analysis*. Oxford, U.K.: Oxford University Press, 2007.

[24] P. J. Brockwell and R. A. Davis, *Time series: theory and methods*. : Springer, 2009.

[25] G. E. Box, G. M. Jenkins, and G. C. Reinsel, *Time Series Analysis: Forecasting and Control*. New York, NY, USA: Wiley, 2013.

[26] I. C. Arndt, "Akaike's information criterion," in *Information Measures*. New York, NY, USA: Springer, 2001, pp. 347–361.

[27] H. Akaike, "Statistical predictor identification," *Annals Inst. Statistical Math.*, vol. 22, pp. 203–217, 1970.

[28] R. E. Uhrig and J. W. Hines, "Computational intelligence in nuclear engineering," *Nucl. Eng. Technol.*, vol. 37, pp. 127–138, 2005.

[29] E. Zio, P. Baraldi, and N. Pedroni, "Selecting features for nuclear transients classification by means of genetic algorithms," *IEEE Trans. Nucl. Sci.*, vol. 53, no. 3, pt. 3, pp. 1479–1493, 2006.

[30] M. J. Embrechts and S. Benedek, "Hybrid identification of nuclear power plant transients with artificial neural networks," *IEEE Trans. Industrial Electron.*, vol. 51, no. 3, pp. 686–693, 2004.

[31] K. C. Kwon, J. H. Kim, and P. H. Seong, "Hidden Markov model-based real-time transient identifications in nuclear power plants," *Int. J. Intell. Syst.*, vol. 17, pp. 791–811, 2002.

[32] O. D. Anderson, *Time Series Analysis and Forecasting: The Box-Jenkins Approach*. London, U.K.: Butterworths, 1976.