NIGERIA JOURNAL OF PURE AND APPLIED PHYSICS

Nigeria Journal of Pure & Applied Physics, Vol. 6, No. 1, pages 6 – 25, 2015

ARMAX, OE AND SSIF MODEL PREDICTORS FOR POWER TRANSMISSION AND DISTRIBUTION PREDICTIONS IN AKURE AND ITS ENVIRONS IN NIGERIA

Reginald O. A. Osakwe[1], Vincent A. Akpan[2†] and Sylvester A. Ekong[3]

[1]*Department of Physics, The Federal University of Petroleum Resources, P.M.B. 1221 Effurun, Delta State, Nigeria.*
[2]*Department of Physics Electronics, The Federal University of Technology, P.M.B. 704 Akure, Ondo State, Nigeria.*
[3]*Department of Physics & Energy Studies, Achievers University, P.M.B. 1030, Owo, Ondo State, Nigeria.*
† Corresponding Author, Email: avincent@eng.auth.gr

Abstract: Three mathematical model structures, namely: ARMAX, OE and a SSIF are first formulated followed by the formulation of their respective model predictors for the model identification and prediction of power transmission and distribution within Akure and its environs. A total of 51,350 data samples from the Power Holding Company of Nigeria were collected for thirteen different parameters that influences the evaluation and analysis in the case study area. The performances of these three model predictors are validated by one-step and five-step ahead prediction methods as well as the Akaike's final prediction error (AFPE) estimates. The results obtained from the application of these three model structures and their predictors for the modeling and prediction of power transmission and distribution as well as the validation results show that the OE model predictor outperforms the ARMAX and SSIF model predictors with much smaller prediction errors, good prediction and tracking capabilities and that the OE model structure and its predictor structure can be used for power transmission and distribution modeling and predictions in real scenarios.

Keywords: Auto-regressive moving average with exogenous input (ARMAX) model, output-error (OE) model, state-space innovations form (SSIF) model, mathematical modeling, model predictor, model structure, power distribution, power transmission.

1. INTRODUCTION

Electrical energy is produced at power stations. It is transmitted to long distance and distributed to loads using transmission lines and transformers. Power is the bulk transfer of electrical energy, from generating power plant to electrical substations located near demand centres. This is distinct from the local wiring between high-voltage substations and customers which are typically referred to as electric power distribution. Transmission lines, when interconnected with each other become transmission networks. The combined transmission and distribution network is known as the power grid.

Power distribution is the final stage in the delivery of electricity to end users. A distribution system's network carries electricity from the transmission system and delivers it to consumers. Typically, the network would include medium-voltage (2kV to 34.5kV) power lines, substations and pole-mounted transformers, low-voltage (less than 1 kV) distribution wiring and sometimes meters.

Due to huge structure of the area of power transmission and distribution systems, it is too difficult to observe behaviour of all the equipment used. Therefore, reliability and performances of the system become poor and to take an action for any fault can be very late. In this case, some fatal problems can occur. In the electricity supply industry, it is important to determine the future demand for power as far in advance as possible. If accurate estimates can be made for maximum and minimum load for each hour, day, month, season and year, utility companies can make significant economics in areas such as setting the operating reserve, maintenance scheduling and fuel inventory management [1, 2]. Thus, electric load prediction is performed over long, medium and short terms, to ensure that the resources are made available to match the demand. Long term prediction gives an insight into the future and helps in investment planning over years. Medium term prediction predicts months ahead and is needed to support energy procurement, energy marketing, tariff management, maintenance planning and network design functions.

Short-term prediction is the prediction of load demand, hours or days into future and is required to support energy trading and network control functions. The factors that affect short term prediction were well discussed by [3]. The prediction procedure depends on the manner in which historical data is analysed and on the type of information available at the time the forecast is prepared [4].

Various techniques have been applied to the problem of power transmission and distribution modeling and prediction. The statistical techniques that were widely used in short term electric load prediction are regression, time series analysis and general exponential smoothing. Papalexopoulos and Hesterberg [5] described a linear regression model for short term load prediction while Haida and Muto [6] presented a multivariate linear regression-based peak load prediction with a transformation technique. The transformation function is used to deal with non-linear relationship between temperature and load and the performance of the technique have been verified with simulations of actual load data from Tokyo Electric Power Company. Charytonuik and co-workers [7] proposed a novel approach to load forecasting by the application of non-parametric regression. The dual work of [8] and [9] developed novel approaches for short term load forecasting, which incorporates the time series modeling of the ARIMA (autoregressive integrated moving average) with the knowledge of experienced human operators. Although these statistical techniques are reliable, they fail to give accurate results when quick weather changes occur which form a non- linear relationship with daily load.

Nevertheless, various techniques have been applied to the problem of power transmission and distribution modeling and prediction. The authors in [6] presented a multi-variance linear regression-based peak load prediction with a transformation technique. The transformation reaction is used to deal with non-linear relationship between temperature and load. Performance of the technique is verified with simulations of actual load data of Tokyo Electricity Power Company.

2. FORMULATION OF THE RESEARCH PROBLEM AND EXPERIMENTAL DATA AQUISITION

2.1 Formulation of the Research Problem

A case study of Akure 132KV Transmission Station was chosen for this work. The transmission station is in general, nonlinear, time-varying multivariable system, subject to large disturbances where different physical and weather phenomena takes place. Many strategies have been proposed to analyze issues concerned with power transmission and distribution, but their evaluations and comparisons are difficult. This is partly due to the variability of the parameters, the complexity of the physical and weather phenomena, and the large range of time constants inherent in the work station. Additional difficulty in the evaluation is the lack of standard evaluation criteria to ascertain how being generated are transmitted and distributed from source to destination.

The aim of this study is on the development of mathematical models and model predictors for the prediction of power transmission and distribution within Akure and its environs using efficient modeling algorithms. This is achieved through: *1).* Data collection and data pre-processing to remove outliers; *2).* The development of three model predictors ARMAX, OE and SSIF model predictors for the modeling and prediction of present and future power transmission and distribution within Akure and its environs; and *3).* The validation of the efficiency of the ARMAX, OE and SSIF model predictors for the prediction of expected future power transmission and distribution in real scenarios to ascertain the best model structure and predictor for future use.

2.2 Experimental Data Acquisition

The data used in this study have been collected from Power Holding Company of Nigeria (PHCN), Akure which is now Benin Electricity Distribution Company (BEDC), from the Omotosho Work Centre, Akure, Ondo State, Nigeria. The measurement data of the transmitted and distributed power parameters are recorded on hourly basis as shown in Fig. 1 from PHCN. Although, the "*Hourly Reading Sheet*" as shown in Fig. 1 has 47 parameters across the 47 columns, only data 13 parameters are of interest in the present study based on the scope of the study and their impact on the intended power transmission and distribution analysis. Thus, a total of 51,350 six months hourly data from 1st October, 2012 to 31st March, 2013 was obtained for the 13 parameters each having 3,950 data for the present study. The thirteen parameters includes: *1).* Total voltage transmitted from Osogbo (TVFO); *2).* Total voltage received in Akure (TVRA); *3).* Total current transmitted from Osogbo (TCFO); *4).* Total current received in Akure (TCRA); *5).* Total power transmitted from Osogbo (TPFO); *6).* Total power received in Akure (TPRA); *7).* Total power distributed to Akure Area 1 in terms of current (T2B_Akure); *8).* Total power distributed to Akure Area 2 in terms of current (T2C_Akure); *9).* Total power distributed to Oba-Ile in terms of current (Oba_Ile); *10).* Total power distributed to Iju in terms of current (Iju); *11).*

Total power distributed to Owena in terms of current (Owena); *12).* Total power distributed to Owo in terms of current (Owo); and *13).* Total power distributed to Igbara-Oke in terms of current (Igbara_Oke).

Thus, the first step is to gather historical information or data about the total voltage (KV) transmitted from Osogbo (TVFO) and that received in Akure (TVRA); the total current (AMP) transmitted from Osogbo (TCFO) and the total current received in Akure (TCRA); and total power (MW) transmitted from Osogbo (TPRO) and the total power received in Akure (TPRA). The study also considers the current distribution to two sub-stations in Akure metropolis (T2B_Akure and T2C_Akure) and five sub-stations around Akure environs which includes Oba-Ile, Iju, Owena, Owo and Igbara-Oke as mentioned above.

3. PROBLEM FORMULATION FOR THE POWER TRANSMISSION AND DISTRIBUTION IN AKURE AND ITS ENVIRONS

3.1 Total Power Transmitted from Osogbo, Osun State and Received in Akure, Ondo State

From the historical data collected from Akure 132KV Line Transmission Station, the following results were collected after simulation using MATLAB. Fig. 2(a) shows the total voltage (in kilo-Volt, KV) transmitted from

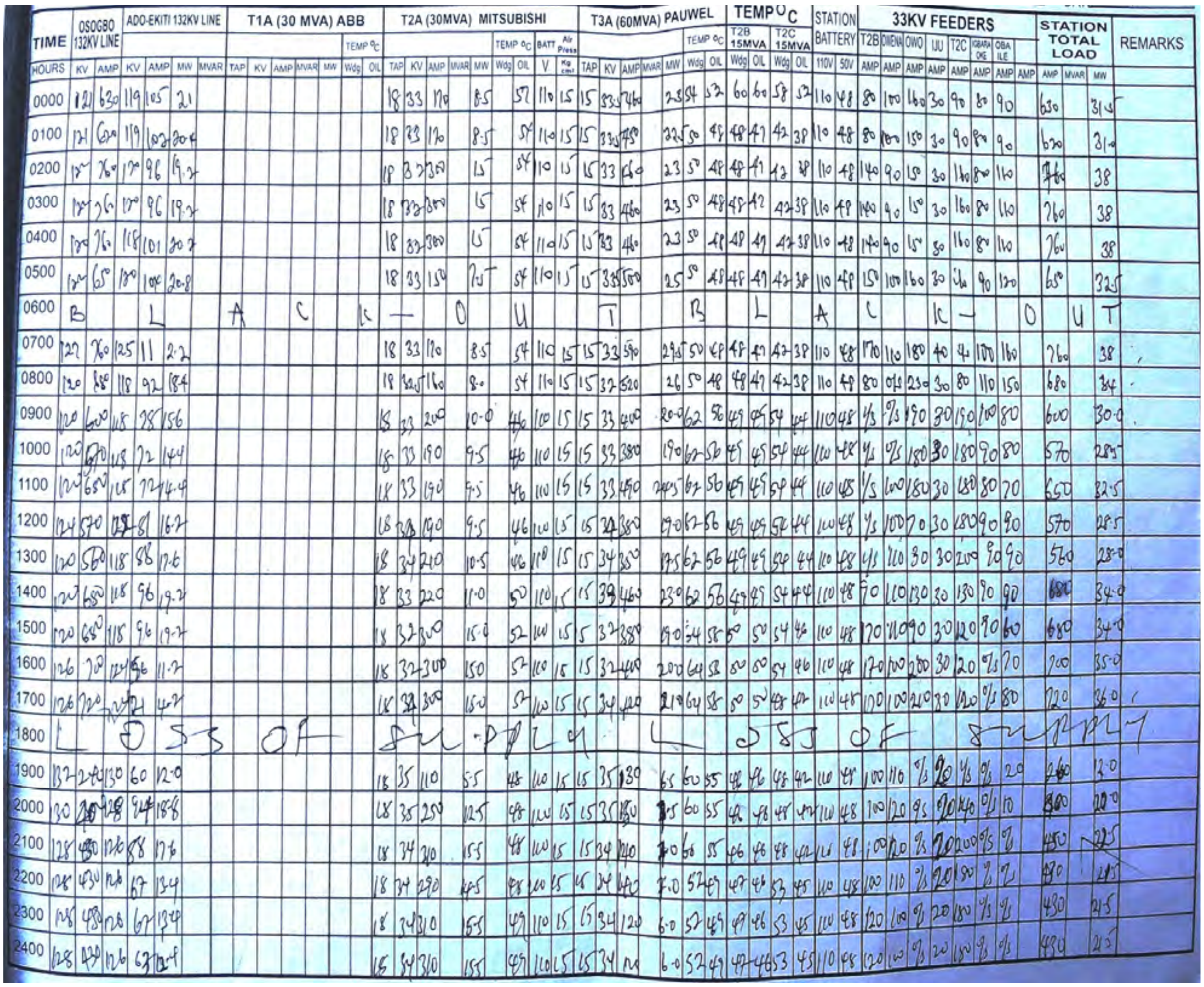

Fig. 1: A Sample of the hourly reading sheet for Sunday 1st October, 2012 showing detail information on transmission to Akure and distribution from Akure.

Osogbo (TVFO) while Fig. 2(b) show the total voltage received in Akure (TVRA). It is obvious from the graphs of Fig. 2(a) and (b) that the voltage received is almost half of the transmitted voltage from Osogbo. The total current (in amperes, AMP) as transmitted from Osogbo (TCFO) and the total current received in Akure (TCRA) are shown in Fig. 2(c) and (d) respectively. It can be seen that the total current from Osogbo and that received in Akure are both the same. This can easily be verified from Fig. 1 because the addition of all the currents distributed to Akure and its environs sums up to the total current transmitted from Osogbo which is equally received in Akure. Fig. 2(e) and (f) presents the total power (in mega-Watt, MW) transmitted from Osogbo (TPFO) and the total power received in Akure (TPRA) respectively. It can be observed from the graphs that the total power from Osogbo is higher than the total power received in Akure. The observed drop in power in Akure transmission station is specifically due to the voltage as well as some factors explained in the next two sub-sections.

3.1.1 Factors Affecting Transmission

1). *Generation cost:* The cost of fuel to generate electric power is a major factor influencing the flow of power into the grid. Under generation will cause the entire system to be slow, i.e. frequency drop and over generation will cause the system to behave in the other way round.
2). *Grid:* A big factor affecting the flow of power through the grid is the grid itself. As power grid age, many portions of it are at maximum transmission capacity. As demand for energy increases, some areas may experience energy demand that exceed the capacity of the existing transmission lines to deliver.

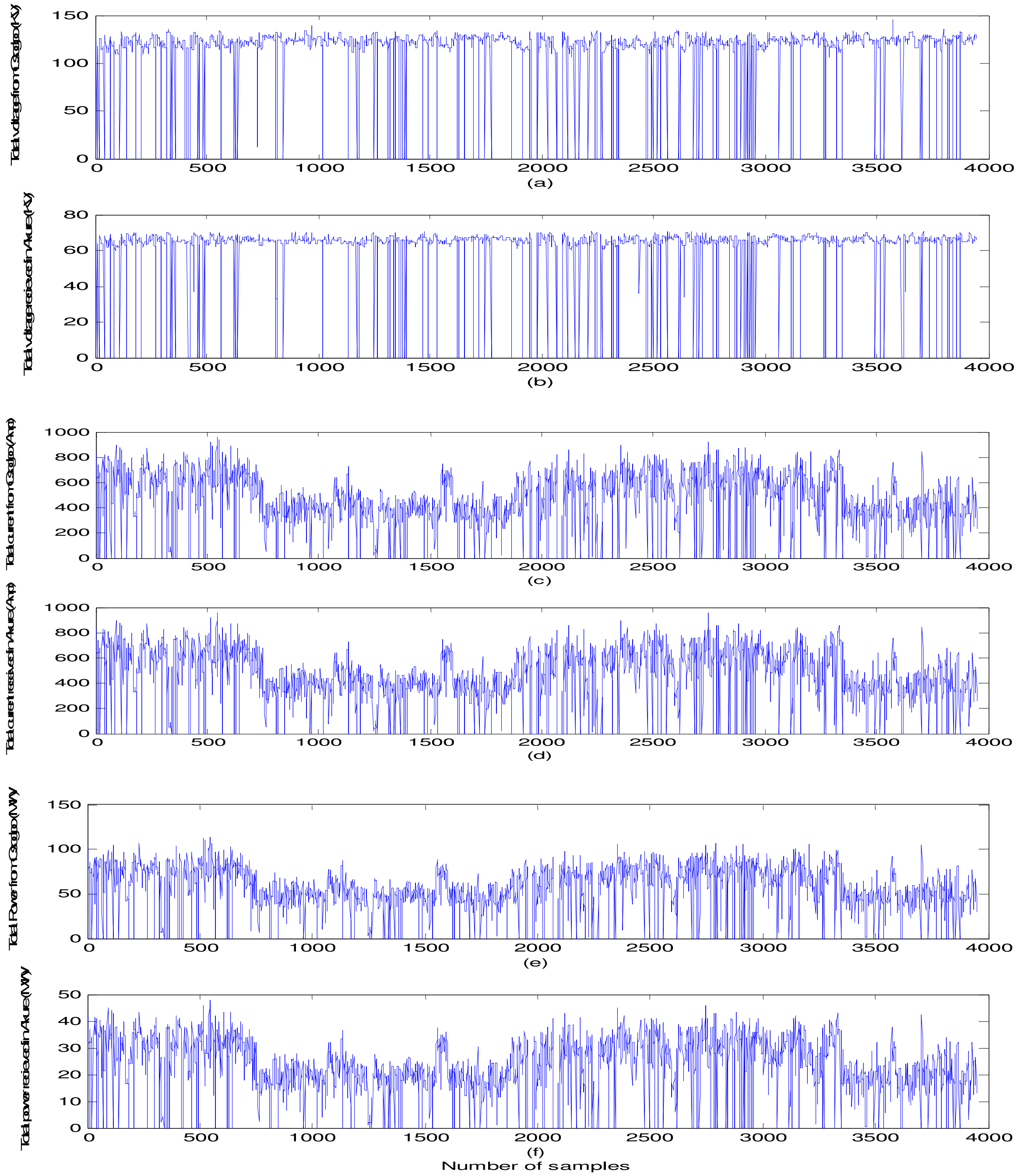

Fig. 2: Graph of total voltage (V), current (A) and power (MW) transmitted from Osogbo and received in Akure: (a) TVFO, (b) TVRA, (c) TCFO, (d) TCRA, (e) TPFO and (f) TPRA.

3). *Weather and Climate:* Weather conditions can have a serious effect on the flow of power in the electric grid. Summer heat waves increases the peak demand for energy to such high level that many businesses and factories must shut down during peak demand hour. Thunderstorms and winter weather can cause downed power lines and loss of power to whole communities.

In addition to the regular seasonal weather related energy problems, global climate change could also influence power flow in the grid. According to Dotto [10], global climate change can cause serious increase in demand for electric power as summer temperature increase.

4). Line Temperature: The principal limitation on the capacity of a line is its temperature. As a line gets warmer, it sags and in worst cases, it can touch trees or ground thereby causing earth fault.

5). Distance: Long distance transmission causes significant loses especially during hot weather. This follows the fact that the resistance of a metallic conductor increases proportionally with length.

6). Power Factor: This is the cosine of the phase difference between voltage and current. For purely resistive load, the voltage and current are in phase and power factor is 1. For a purely reactive load, the voltage and current are out of phase and power factor is 0. The power factor is leading if current leads voltage (capacitive load) and is lagging if the current lags voltage (inductive load). Most domestic loads (such as washing machines, air conditioners and refrigerators) and industrial loads (such as induction motors) are inductive and operate at a low lagging power factor. Loads with low power factors are costly to serve because they require large currents and invariably results in increased power loss. In view of this, power companies often encourage their customers to have power factors as close to unity as possible and penalized some customers who do not improve their load power factor. One of the ways for increasing power factor is the addition of a reactive elements usually static capacitor in parallel with the load in order to make the power factor closer to unity.

3.1.2 The Physics of Power Loss in the Transmission and Distribution Lines

Voltage drops in an electrical circuit normally occur when current is passed through the cable. Cables used to distribute power throughout the transmission and/or distribution lines have resistance associated with them. The longer the cables the larger the resistance and the larger the resistance the greater the voltage drop. Hence, voltage drop is inevitable if it has to carry current. There are many causes of voltage drop but the four fundamental causes of voltage are [11]: *1). Material:* copper is a better conductor than aluminium and will have less voltage drop than aluminium for a given length and cable size; *2). Cable Size:* Larger cable size will have less voltage than smaller cable size of the same length; *3). Cable Length:* shorter cables will have less voltage drop than longer cables of the same size (diameter); and *4). Current Being Carried:* Voltage drop increase with an increase in the current flowing through the cable.

In alternating current circuits, opposition to current flow does occur because of resistance (just as in direct current circuits). Alternating current circuits also present a second kind of opposition to current flow: reactance. This "*total*" opposition (i.e. the sum of resistance and reactance) is called impedance. The impedance in an alternating current circuit depends on the spacing and dimensions of the elements and conductors, the frequency of the alternating current, and the magnetic permeability of the elements, the conductors, and their surroundings. The voltage drop in an alternating current circuit is the product of the current and the impedance of the circuit. Electrical impedance, like resistance, is expressed in ohms. Electrical impedance is the vector sum of electrical resistance, capacitive reactance, and inductive reactance.

Thus, significant voltage drop is due to long transmission distance of approximately 186 km from Osogbo to Akure neglecting the cable size. Additional voltage drop occur again during distribution from the 132 KV transmission station (Omotosho Work Centre) in Akure to major load centers in Akure and its environs which results in frequent low voltages to consumers as can be seen from Fig. 1.

3.2 Individual Power Transmitted From Osogbo and Distributed to Akure and Its Environs.

Fig. 3(a) and (b) present the total current distributed to Akure area 1(T2B_Akure) and Akure area 2 (T2C_Akure). It can be seen that there is power outages as indicated by the blank spaces in Fig. 3(a) due to several factors such as those shown in Fig. 1 as well as several other hourly reading sheets used in this study though not shown here for space economy. The total power distributed to and consumed by Akure area1 in terms of current would have been approximately the same as that distributed to and consumed by Akure area 2 (T2C_Akure) except for the power outage periods as shown in Fig. 3(a). According to the Omotosho Work Centre in Akure, the areas that make up Akure area 1 (T2B_Akure) are Oyemekun and Olle-Eda while those that make up Akure area 2 (T2C_Akure) are Isinkan, Ondo road, Ajipowo and Ilesha road.

According to the PHCN *Hourly Reading Sheet*, the five towns that make up the so-called Akure environs are Oba-Ile, Iju, Owena, Owo and Igbara-Oke; and the total power distributed to these areas from the 132KV transmission station in Akure are shown respectively in Fig. 3(c) through Fig. 3(g).

Power distribution at the 132KV transmission state from the Omotosho Work Centre in Akure are as follows: Akure Area 1 (T2B_Akure) and Akure Area 2 (T2C_Akure) share a common 30MVA transformer called T2A and the total current distributed to Akure area 1 and 2 are as follows: *1).* the total current distributed to Akure area 1 (T2B_Akure) is 250 AMP; and *2).* the total current distributed to Akure area 2 (T2C_Akure) is 250 AMP. Thus, the total current distributed within Akure is from T2A is 500 AMP.

Owena and Igbara-Oke share a 30 MVA transformer called T1A and the total current distributed to these areas is as follows: the total current distributed to Owena is 150 AMP; the total current distributed to Igbara-Oke is 150 AMP; and the total current from T1A is 300 AMP. However, Owo, Oba-Ile, and Iju share a separate 60MVA transformer called T3A and the total current distributed to these areas is as follows: the total current distributed to Owo is 300 AMP; the total current (AMP) distributed to

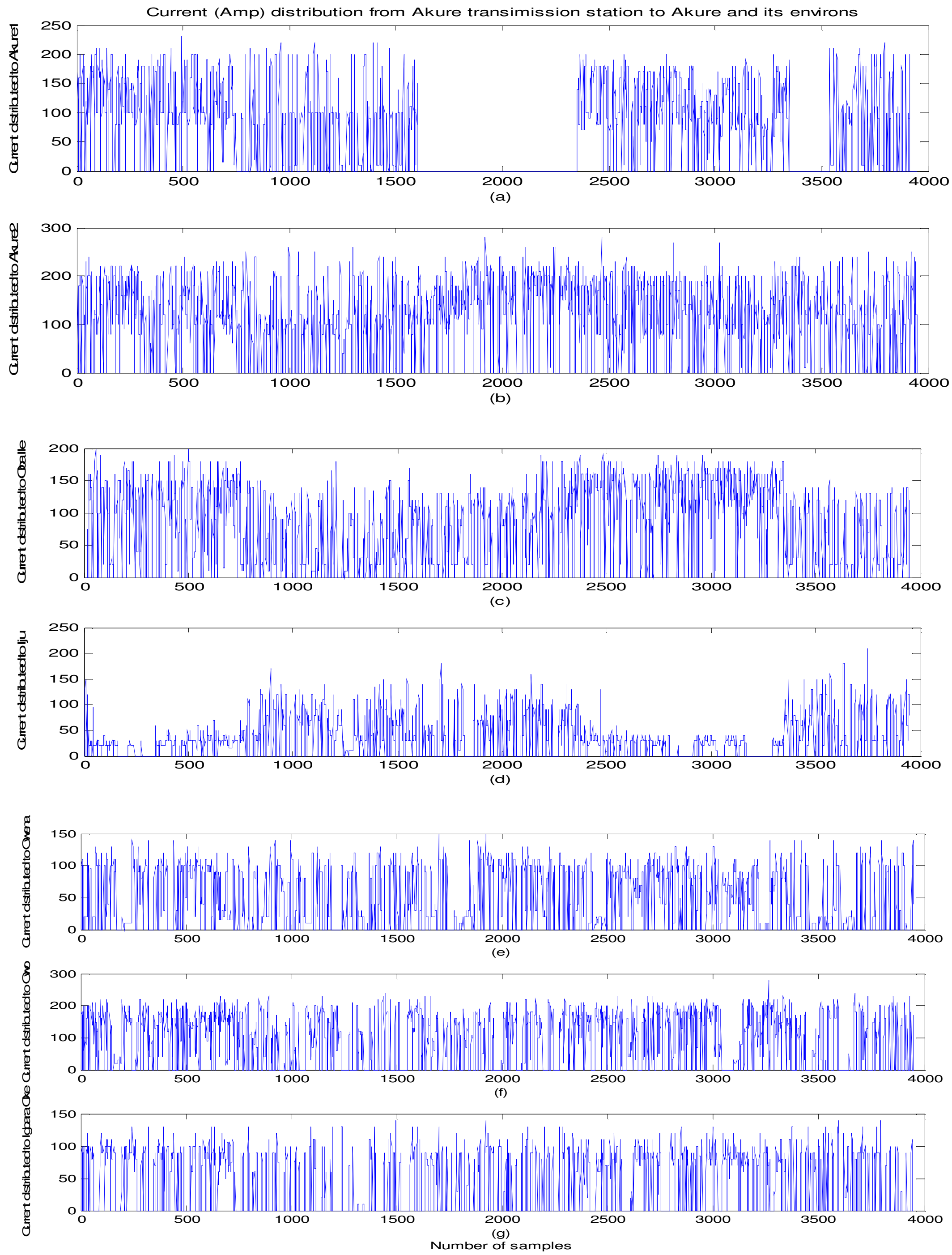

Fig. 3: Graph of total power in terms of current distributed from the 132kV Omotosho transmission station in Akure to: (a) T2B_Akure, (b) T2C_Akure, (c) Oba_Ile, (d) Iju, (e) Owena, (f) Owo and (g) Igbara_Oke.

Oba-Ile is about 200 AMP; the total current distributed to Iju is 200 AMP; and the total current from T3A is 700 AMP.

The total current rating of the three transformers for power distribution to Akure area 1 (T2B_Akure) and area 2 (T2C_Akure), Oba-Ile, Iju, Owena, Owo and Igbara-Oke is 1500 AMP. If we compare the total current distributed to Akure and its environs, it can be seen that it is equal to the total current received in Akure. Therefore, the total current received is equal to the total current distributed.

3.2.1 Factors Affecting Power Consumption

1). *Time Factor:* The time factor includes the time of the year, the day of the week, and the hour of the day. There are important differences in load between weekdays and weekends. The load on different weekdays can also behave differently. For example, Mondays and Fridays being adjacent to weekends, may have structurally different loads than Tuesdays through Thursdays.

2). *Seasonal Variation:* the power consumed during one week in winter cold due to increasing use of electric heaters differs from the power consumed during one week in summer warm, which also increases due to the use of air conditioning equipments.

3). *Economic Factors:* This estimates the relationship between energy consumption and factors influencing consumption. The economic approach classified consumers of utility industry into residential, commercial and industrial consumers. Electric utility industry is more responsive to commercial and industrial consumers than residential consumers.

4). *Population Size:* This has to do with the numbers of domestic consumers of electricity. The demand for electricity is a function of the number of domestic consumers in the market. Factors such as the sizes of the houses, the age of equipment, technology changes, consumers behaviour e.t.c. influence the demand for electricity significantly.

3.2.2 Interpretation and Discussion of the Power-Current Relationship

Considering the first row of Fig. 1 as an example, it can be seen in the first row that the expected power is approximately 76.23 MW while the total power transmitted and consumed within Akure and its environs is 31.5 MW, which is less than half of the expected power due to several factors as highlighted above. As can be seen in Fig. 2(a), there are periods of zero voltages which are caused by some factors such as the two shown in Fig. 1 and four others not shown here for space economy. Besides the two shown in Fig. 1, in the charts used for this study, we have the following: *1).* Blackout, Loss of Supply, and/or No Reading which are periods when there is no power supply from Osogbo Area Control station.; and *2).* Planned Outage and/or System Collapse (Unplanned Outage) which occur when Osogbo Area Control station is shutdown for transformer, equipment or routine maintenance during which period there is no power supply from Osogbo.

4. MATHEMATICAL MODELS OF DYNAMIC SYSTEMS AND THE DEVELOPMENT OF THEIR MODEL PREDICTORS

The method of representing the behaviour of dynamical systems by vector difference or differential mathematical relationships is well established in system and control theories [12–15]. These relationships constitute the so-called mathematical model of the system. One very common method of modeling the behaviour of a *p*-input *q*-output multivariable plant in the discrete time space is by the family of the following general mathematical relationship [12–15]:

$$A(z^{-1})Y(k) = z^{-d}\,\frac{B(z^{-1})}{F(z^{-1})}U(k) + \frac{C(z^{-1})}{D(z^{-1})}e(k) \qquad (1)$$

where $Y(k)$ is the vector of order n of the q outputs at the timing instant k responding to the vector input $U(k)$; $e(k)$ is the noise disturbance vector; and $A(z^{-1})$, $B(z^{-1})$, $C(z^{-1})$, $D(z^{-1})$ and $F(z^{-1})$ are polynomial matrices given by

$$\begin{aligned}
A(z^{-1}) &= I + A_1 z^{-1} + \cdots + A_{n_a} z^{-n_a} \\
B(z^{-1}) &= B_0 + B_1 z^{-1} + \cdots + B_{n_b} z^{-n_b} \\
C(z^{-1}) &= I + C_1 z^{-1} + \cdots + C_{n_c} z^{-n_c} \\
D(z^{-1}) &= I + D_1 z^{-1} + \cdots + D_{n_d} z^{-n_d} \\
F(z^{-1}) &= I + F_1 z^{-1} + \cdots + F_{n_f} z^{-n_f}
\end{aligned} \qquad (2)$$

d is the system delay, *A*, *C*, *D* and *F* are *monic* polynomial matrices in the backward shift operator z^{-1} Their dimensions are $n_a \times n_a$ $q \times n_a$, $q \times n_d$, $q \times n_d$ and $q \times n_f$ and their degree n, m, c, l, respectively; B is a $n_b \times p$ stable polynomial matrix (i.e. all its zeros are all inside the unit circle) of degrees of degree r. The term *monic* implies that the leading coefficients of *A*, *C*, *D* and *F* are identity matrices of appropriate dimension to avoid division by zeros and also because the magnitude of $e(k)$ can be adjusted to compensate for this if necessary. In this discussion, it is assumed that: *1)* the time delay d of the system is known, i.e. $d = 1$; *2)* the coefficients of the polynomials matrices $A(z^{-1})$, $B(z^{-1})$, $C(z^{-1})$, $D(z^{-1})$ and $F(z^{-1})$ are unknown; *3)* the polynomials matrices $A(z^{-1})$, $B(z^{-1})$, $C(z^{-1})$, $D(z^{-1})$ and $F(z^{-1})$ are relatively prime; and *4)* that the upper bound on the order or each polynomial matrix is known or can be specified exactly.

Since the noise term $e(k)$ enters the general model Equation (1) as a direct error term, the model of Equation (1) is often called an equation error model [12, 15]. Depending on how the five parameters A, B, C, D and F are combined, several model structures can be obtained from Equation (1)

The choice of the models that will represent the noise disturbances is as important as the choice of the system model. Depending on the different assumptions made about the spectral density of the noise, $e(k)$ and how the noise is assumed to enter the system given by Equation (1); 32 different model structures can be derived from Equation (1) based on the combination of the five parameters *A*, *B*, *C*, *D* and *F* [15]. However, the model structures considered in the present work is limited to the structures derived from the combination of the four parameters *A*, *B*, *C* and *F*, that is ignoring the *D* parameter in Equation (1). The reason for choosing these four parameters is because, as literature shows, they were adequate for the modeling needs of the model predictive control (MPC) for a wide range of dynamical systems [12, 16]. The combination of *A*, *B* and *C* gives an AutoRegressive Moving Average with eXogenous inputs (ARMAX) model while the combination of *B* and *F* results in an output error (OE) model. The OE model is a form of equation error model [12, 13, 15] and can also take the form based on *A*, *B*, *C* and *D* which is widely used in MPC literature [12, 16, 17]. Rather than using *A*, *B*, *C* and *D* to describe the OE model, the choice of using *B* and *F* is adopted in this work for the OE model [12].

Let $\theta(k)$ be a parameter vector which encapsulates the model parameters given in Equation (2) and defined as:

$$\left.\begin{aligned}\theta(k) = \big[-A_1,\ldots,-A_{n_a},\ B_1,\ldots,B_{n_b},\ C_1,\ldots,C_{n_c}, \\ -D_1,\ldots,-D_{n_d},\ -F_1,\ldots,-F_{n_f} \big]^T\end{aligned}\right\} \qquad (3)$$

Since the exact value of the parameter vector $\theta(k)$ in Equation (3) is unknown, a parameterized set of model structures Θ can be defined as a set of candidate models given as:

$$\Theta:\ \theta(k) \in \mathbb{Q}_\theta \subset \Re^\nu \to \hat{\theta}(k) \qquad (4)$$

where $\mathbb{Q}_\theta$ is some subset of $\Re^\nu$ inside which the search for a model is carried out; ν is the dimension of $\theta(k)$; $\hat{\theta}(k)$ is the desired model associated with the parameter vector $\theta(k)$ and contained in the set of models $\Theta = \{\theta(k)_1, \theta_2(k), \ldots, \theta_\tau(k)\}$; $\theta(k)_1, \theta_2(k), \ldots, \theta_\tau(k)$ } Each member of this set is a distinct value of $\theta(k)$; and $\tau = 1, 2, \ldots, \max_{iter}$ is the number of iterations required to determine the $\hat{\theta}(k)$ from Θ.

Thus, the minimum variance (one-step) ahead predictor of Equation (1) at time k based on the system information up to the time $k-1$ can be expressed as

$$\left.\begin{aligned}\hat{Y}(k \mid k-1, \theta(k-1)) = z^{-d}\frac{B(z^{-1})D(z^{-1})}{F(z^{-1})C(z^{-1})}U(k) \\ + \left[1 - A(z^{-1})\frac{D(z^{-1})}{C(z^{-1})}\right]Y(k)\end{aligned}\right\} \qquad (5)$$

Note the inclusion of $\theta(k)$ as an argument to indicate that the model structure represents a set of models. For notational convenience, the $k-1$ will be omitted henceforth. The prediction error $\varepsilon(k,\theta)$ can be computed directly from Equation (1) and Equation (5) as follows:

$$\left.\begin{aligned}\varepsilon(k,\theta(k)) &= Y(k) - \hat{Y}(k,\theta(k)) \\ &= \frac{D(z^{-1})}{C(z^{-1})}\left[A(z^{-1})Y(k) - z^{-d}\frac{B(z^{-1})}{F(z^{-1})}U(k)\right]\end{aligned}\right\} \qquad (6)$$

By introducing $\tilde{d}(k,\theta(k)) = z^{-d}\dfrac{B(z^{-1})}{F(z^{-1})}U(k)$ (7)

and $\tilde{v}(k,\theta(k)) = A(z^{-1})Y(k) - \tilde{d}(k,\theta(k))$ (8)

and using Equation (7) and Equation (8), Equation (6) can be expressed as

$$\varepsilon(k,\theta(k)) = Y(k) - \hat{Y}(k,\theta(k)) = \frac{D(z^{-1})}{C(z^{-1})}\tilde{v}(k,\theta(k)) \qquad (9)$$

Let the regression vector (the so-called state vector) derived from the difference equation form of Equation (1) be:

$$\left.\begin{aligned}\varphi(k,\theta(k)) = \big[&Y(k-1),\ldots,Y(k-n_a), \\ &U(k-d),\ldots,U(k-d-n_b), \\ &\varepsilon(k-1,\theta(k)),\ldots,\varepsilon(k-n_c,\theta(k)), \\ &\tilde{d}(k-1,\theta(k)),\ldots,\tilde{d}(k-n_f,\theta(k)), \\ &\tilde{v}(k-n_d,\theta(k)),\ldots,\tilde{v}(k-n_{d,\theta(k)})\big]\end{aligned}\right\} \qquad (10)$$

Using the parameter vector given in Equation (3) and the regression vector in Equation (10) above, equations Equation (7) and Equation (9) can be expressed respectively as:

$$\left.\begin{aligned}\tilde{d}(k,\theta(k)) = B_1U(k-d) + \cdots + B_{n_b}U(k-d-n_b) \\ + F_1\tilde{d}(k-1,\theta(k)) + \cdots + F_{n_f}\tilde{d}(k-n_f,\theta(k))\end{aligned}\right\} \qquad (11)$$

$$\left.\begin{aligned}\varepsilon(k,\theta(k)) = C_1\varepsilon(k-1,\theta(k)) + \cdots + C_{n_c}\varepsilon(k-n_c,\theta(k)) \\ + \tilde{v}(k,\theta) + D_1\tilde{v}(k-d) + \cdots + D_{n_d}\tilde{v}(k-d-n_d)\end{aligned}\right\} \qquad (12)$$

Inserting $\tilde{v}(k,\theta)$ from Equation (12) and substituting $\tilde{d}(k,\theta)$ from Equation (11) into Equation (6) gives

$$\varepsilon(k,\theta(k)) = Y(k) - \varphi(k,\theta(k))\theta(k) \qquad (13)$$

Thus, the one-step ahead predictor can then be expressed as:

$$\hat{Y}(k,\theta(k)) = \varphi(k,\theta(k))\theta(k) \qquad (14)$$

Remarks on the Disturbance Model

The disturbance model, i.e. the second term in Equation (1), plays significant role in modeling the overall system model behaviour. Let the disturbance model be defined as

$$D_M = \frac{C(z^{-1})}{D(z^{-1})}e(k) \qquad (15)$$

In MPC literature, the model of Equation (15) is usually called CARIMA (controlled auto-regressive and integrated moving average) model [12, 16, 17]. In practice, $e(k)$ cannot be measured but it can be estimated as deterministic or stochastic noise [12, 14, 15, 18].

4.1 Autoregressive Moving Average with Exogenous Input (ARMAX) Model

Considering the disturbance model in Equation (15), the stochastic case is somewhat more involved. Consider the case of modeling a stationary, zero-mean white noise process, namely $E\{e(k)^2\} = \sigma^2$, $E\{e(k)e(k-\lambda)\} = 0$ for all $\lambda \neq 0$, the probability distribution of $e(k)$ being the same for all (k), and each $e(k)$ being independent of $e(\lambda)$ if $\lambda \neq k$; where the term $E\{\bullet\}$ implies the expectation or mean value of its arguments. Then, if $C(z^{-1})/D(z^{-1})$ is an asymptotically stable transfer function, the (15) will be a stationary process with spectral density given by

$$\Phi_{SD}(\omega) = \sigma^2 \frac{\left|C(e^{-j\omega T})\right|^2}{\left|D(e^{-j\omega T})\right|^2} \tag{16}$$

where σ_2 is the spectral density. Note that since $\left|C(e^{-j\omega T})\right|^2 = C(e^{-j\omega T}) \cdot C(e^{-j\omega T})$, it is always possible to choose $C(z^{-1})$ such that all its roots lie inside the unit disc, i.e. without restricting the spectral densities which can be modeled in this way. Also for the same reason, the factors of $C(z^{-1})$ do not affect the spectral density. This property shows and guarantees a useful way of selecting $C(z^{-1})$ to lie inside the unit circle for models with moving average such as ARMAX and OE models including the SSIF models.

As in the previous sub-section 4.1, with the assumptions on Equation (15) and setting $D(z^{-1}) = F(z^{-1}) = 1$ in Equation (1), Equation (1) essentially reduces to an autoregressive moving average with exogenous input (ARMAX) model structure, which is usually unstable but finds applications in wide range of systems with coloured noise. From Equation (5), the ARMAX model predictor becomes:

$$\left.\begin{aligned}\hat{Y}(k \mid k-1, \theta(k-1)) = z^{-d} \frac{B(z^{-1})}{C(z^{-1})} U(k) \\ + \left[1 - A(z^{-1}) \frac{1}{C(z^{-1})}\right] Y(k)\end{aligned}\right\} \tag{17}$$

where the regression vector and the adjustable parameters of the ARMAX model predictor are given respectively from Equation (10) and Equation (3) as

$$\left.\begin{aligned}\varphi_{ARMAX}(k, \theta(k)) = [&\mathrm{Y}(k-1|\theta(k)), \ldots, \mathrm{Y}(k-r|\theta(\mathrm{k})), \\ &U(k-d), \ldots, U(k-d-m), \\ &\varepsilon(k-1, \theta(k)), \ldots, \varepsilon(k-n_c, \theta(k))] \\ \theta_{ARMAX} = [&-A_1, \ldots, -A_{n_a}, \mathrm{B}_0, \ldots, \mathrm{B}_{n_b}, C_0, \ldots, C_{n_c}]^T\end{aligned}\right\} \tag{18}$$

Note that the moving average filter $C(z^{-1})$ must be estimated at each time step and must equally lie within the left-hand plane of the unit circle for stability [12].

4.2 Output-Error (OE) Model

The OE is a special stochastic case which is achieved by setting $A(z^{-1}) = C(z^{-1}) = D(z^{-1}) = 1$ with the assumption that $e(k)$ is a zero-mean white noise with finite variance while its first few terms are made non-zero. Additional assumption on $e(k)$ is that it is independent of past inputs and that it can be characterized by some probability function [12, 15]. With these assumptions on Equation (15), Equation (1) essentially reduces to an output error (OE) model structure, which is sometimes unstable for wide range of operations due to the integrator term $F(\mathrm{z}^{-1})$ in Equation (1). The OE model or parallel model structure is used if the only noise affecting the system is white measurement noise. Thus, from Equation (5), the OE model predictor becomes:

$$\hat{Y}(k \mid k-1, \theta(k-1)) = z^{-d} \frac{B(z^{-1})}{F(z^{-1})} U(k) + e(k) \tag{19}$$

corresponding to the following choice of G and H given as:

$$\left.\begin{aligned}G(q^{-1}, \theta) &= z^{-d} \frac{B(z^{-1})}{F(z^{-1})} \\ H(q^{-1}, \theta) &= 1\end{aligned}\right\} \tag{20}$$

The OE model predictor for this system is given from Equation (5) and Equation (19) as:

$$\left.\begin{aligned}\hat{\mathrm{Y}}(k|\theta(\mathrm{k})) &= z^{-d} \frac{B(\mathrm{z}^{-1})}{F(\mathrm{z}^{-1})} U(t) \\ &= z^{-d} B(q^{-1}) \mathrm{U}(k) + [1 - F(\mathrm{z}^{-1})] \hat{\mathrm{Y}}(k|\theta(\mathrm{k})) \\ &= \varphi^T(k, \theta(k)) \theta(k)\end{aligned}\right\} \tag{21}$$

where the regression vector and the adjustable parameters of the OE model predictor are given respectively as

$$\left.\begin{aligned}\varphi_{OE}(k, \theta(k)) = [&\hat{\mathrm{Y}}(k-1|\theta(k)), \ldots, \hat{\mathrm{Y}}(k-r|\theta(\mathrm{k})), \\ &U(k-d), \ldots, U(k-d-m)] \\ \theta_{OE} = [&\mathrm{B}_0, \ldots, \mathrm{B}_{n_b}, -F_1, \ldots, -F_{n_f}]^T\end{aligned}\right\} \tag{22}$$

Therefore, for the predictions to be stable, the roots of F must be inside the unit circle. Note that the OE model predictor uses the system outputs $Y(k)$ and the model parameters $\theta(k)$ to predict future outputs $\hat{Y}(k)$.

4.3 State–Space Innovations Form (SSIF)

The state-space description is a widely used alternative to the input-output model structures such as the ARMAX and the OE model structures described in the last two subsections above. For a system that can be described by the following set of coupled first-order difference equation given by:

$$\left.\begin{aligned}x(k+1) &= A(\theta(k))x(k) + B(\theta(k))U(k) + w(k) \\ Y(k) &= C(\theta(k))x(k) + v(k)\end{aligned}\right\} \tag{23}$$

where *x(k)* and *x(k+1)* are the state variables at the current time step *(k)* and the next time step *(k+1)* respectively, *w(k)* and *v(k)* are white noise signals independent of the control signal, *U(k)* and the expectation value

$$\mathbf{E}\left\{\begin{bmatrix} w(k) \\ v(k) \end{bmatrix} \begin{bmatrix} w^T(k) v^T(k) \end{bmatrix}\right\} = \begin{bmatrix} R_w(\theta(k)) & R_{wv}(\theta(k)) \\ R_{wv}^T(\theta(k)) & R_v(\theta(k)) \end{bmatrix} \tag{24}$$

It has been shown in [19, 20] that the optimal one-step ahead model predictor for system Equation (23) takes the following form:

$$\left.\begin{aligned}\hat{x}(k+1, \theta(k)) &= A(\theta(k))\hat{x}(k, \theta(k)) + B(\theta(k))U(k) \\ &\quad + K(\theta(k))\varepsilon(k, \theta(k)) \\ \hat{\mathrm{Y}}(k|\theta(k)) &= C(\theta(k))\hat{x}(k, \theta(k))\end{aligned}\right\} \tag{25}$$

where the regression vector and the adjustable parameters of the SSIF model predictor are given respectively from Equation (10) and Equation (3) as

$$\left.\begin{aligned}\varphi_{SSIF}(k,\theta(k)) = [&\mathrm{Y}(k-1|\theta(k)),\ldots,\mathrm{Y}(k-r|\theta(k)),\\ &U(k-d),\ldots,U(k-d-m),\\ &\varepsilon(k-1,\theta(k)),\ldots,\varepsilon(k-n_c,\theta(k)),\\ &v(k-1,\theta(k)),\ldots,v(k-n_c,\theta(k)),\\ &w(k-1,\theta(k)),\ldots,w(k-n_c,\theta(k))]\\ \theta_{SSIF} = [-A_1,\ldots,&-A_{n_a}, \mathrm{B}_0,\ldots,\mathrm{B}_{n_b}, C_0,\ldots,C_{n_c}]^T\end{aligned}\right\} \quad (26)$$

Note that the moving average filter $C(z^{-1})$ must be estimated at each time step and must equally lie inside the left-hand plane of the unit circle for stability [12]. Omitting θ for notational convenience, $K(\theta)$ can be found from:

$$K = [APC^T + R_{wv}][CPC^T + R_v]^{-1} \quad (27)$$

where $P(\theta)$ represents the positive semi-definite solution to the stationary Ricatti equation given as:

$$\left.\begin{aligned}P &= APA^T +\\ R_w &- [APC^T + R_{wv}][CPC^T + R_v]^{-1}[APC^T + R_{wv}]^T\end{aligned}\right\} \quad (28)$$

The optimal predictor is also known as the *Kalman filter* and the matrix $K(\theta)$ is referred to as the Kalman gain. The form of $\hat{x}(t+1,\theta(k))$ in Equation (25) is called the *state space innovations form.*

A simple relationship between the state-space innovations form and the general input-output form exist from [15] as follows:

$$G(z^{-1},\theta(k)) = C(\theta(k))\left[zI - A(\theta(k))\right]^{-1} B(\theta(k)) \quad (29)$$

$$H(z^{-1},\theta(k)) = C(\theta(k))\left[zI - A(\theta(k))\right]^{-1} K(\theta(k)) + \mathrm{I} \quad (30)$$

By some matrix manipulations, it can be verified that the poles of the predictor are the eigenvalues of the matrix *A – KC* [19, 20]. The set D_m is thus given by:

$$D_m = \left\{\begin{matrix}\theta(k) \mid A(\theta(k)) - K(\theta(k))C(\theta(k)),\\ \textit{inside the unit circle}\end{matrix}\right\} \quad (31)$$

When estimating state-space models, the elements in $K(\theta(k))$ are typically estimated directly rather than the detour of estimating the covariance matrices and solving the Ricatti equation before computing $K(\theta(k))$.

The selection of a proper parameterization, i.e., the structure of *A, B, C* and *K* is a problem that disfavours the innovations form. It is far more involved than the input-output model structures. The problem is that a fully parameterized model structure, meaning that all elements in *A, B, C* and *K* must be estimated, is generally not identifiable from a set of input-output data. This is because such a structure contains more adjustable parameters than necessary: the same input-output relationship can be described by different choices of *A, B, C* and *K*. Sometimes the parameterization can be based on physical insight, which may remedy the problem. However, when taking a *black-box* approach some kind of "*generic*" identifiable parameterization is required. In the classical single-input single-output (SISO) case, a number of the so-called canonical forms exist and are frequently used for transforming a transfer function description to a state-space description [20]. In the modern multiple-input multiple-output (MIMO) case, it is more complicated. Ljung [15] proposed MIMO extensions called overlapping forms that are identifiable and thus suitable in a system identification context. According to Ljung's Appendix A [15] guidelines for selecting a parameterization are recapitulated in the following *Corollary 1* (where *n* specifies the model order and n_y specifies the number of past outputs):

Corollary 1:

> *"Let $A(\theta(k))$ initially be a matrix filled with zeros and with ones along the super-diagonal. Let the row numbers be $r_1, r_2, \ldots, r_{ny}$, where $r_{ny} = n$, be filled with parameters. Take $r_0 = 0$ and let $C(\theta(k))$ be filled with zeros, and then let row i have a one in column $r_{i-1}+1$. Let $B(\theta(k))$ and $K(\theta(k))$ be filled with parameters".*

Based on Ljung's [15] guidelines stated in *Corollary 1* above, the structural decisions to be made are thus the model order (*n*) and a set of row indices, $\{r_i\}_{i=1}^{n_y-1}$. Furthermore, note that the moving average filter $C(z^{-1})$ must be estimated at each time step and must equally lie within the left-hand plane of the unit circle for stability [12].

5. The ARMAX, OE and SSIF Model Validation Algorithms

5.1 Validation of the Output Predictions of the Model Predictors

Network validations are performed to assess to what extend the trained model predictors have approximated and captured the behaviour of the underlying dynamics of a system and as measure of how well the model being investigated will perform when deployed for the actual system modeling and future predictions [12, 15, 21].

The first test involves the comparison of one-step output predictions of the true training data using the three model predictors given in Equation (17), Equation (21) and Equation (25) and the evaluation of their respective corresponding prediction errors using Equation (13).

The second test involves the comparison of one-step output predictions of the true validation data that was not used during the model predictor development and the evaluation of their respective corresponding prediction errors using Equation (13).

5.2 K-Step Ahead Validation of the Output Predictions of the Model Predictors

The third method is the *K*-step ahead predictions [12, 15] where the outputs of the trained network are compared to the unscaled output training data. The K-step ahead predictor follows directly from Equation (8) and for

$\varphi(k) = \hat{\varphi}(k+K)$ and $\theta(k) = \hat{\theta}(k)$, takes the following form:

$$\hat{Y}((k+K)\,|\,k,\hat{\theta}) = \hat{J}(Z^N, \hat{\varphi}(k+K), \hat{\theta}(k)) \quad (32)$$

where $\hat{\varphi}(k+K) = [U((k+K-1)\,|\,\hat{\theta}),\ldots, U((k+K-m)\,|\,\hat{\theta}),$

$\hat{Y}((k+K-1)\,|\,\hat{\theta}),\ldots,\hat{Y}((k+K+1-\min(k,n))\,|\,\hat{\theta}),$

$Y((k+K-1)\,|\,\hat{\theta}),\ldots,Y((k+K-\max(n-k,0)\,|\,\hat{\theta})]^T$

The mean value of the K-step ahead prediction error (MVPE) between the predicted output and the actual training data set is computed as follows:

$$MVPE = mean\left(\sum_{k=m+K}^{N} \frac{Y(k) - \hat{Y}((k+K)\,|\,k,\hat{\theta})}{Y(k)}\right) \times 100\% \quad (33)$$

where $Y(k)$ corresponds to the actual output training data and $\hat{Y}((k+K)\,|\,k,\hat{\theta})$ the *K*-step ahead predictor output.

5.3 Akaike's Final Prediction Error (AFPE) Estimate

The fourth validation test is the Akaike's final prediction error (AFPE) estimate based on the weight decay parameter D [12, 13, 15]. A smaller value of the AFPE estimate indicates that the identified model approximately captures all the dynamics of the underlying system and can be presented with new data from the real process. Evaluating the $\varepsilon(k,\hat{\theta}(k))$ portion of Equation (13) using the trained network with $\theta(k) = \hat{\theta}(k)$ and taking the expectation $\mathbf{E}\{J(Z^N, \hat{\theta}(k))\}$ with respect to $\varphi(k)$ and $\tilde{d}(k)$ leads to the following AFPE estimate:

$$\hat{F}(Z^N \hat{\theta}(k)) \approx \frac{N + p_a}{N - p_b} J(Z^N \hat{\theta}(k)) + \gamma \quad (34)$$

where

$$p_a = \mathbf{tr}\left\{V(\hat{\theta}(k))\left[V(\hat{\theta}(k)) + D\right]^{-1} V(\hat{\theta}(k))\left[V(\hat{\theta}(k)) + D\right]^{-1}\right\}$$

and $\mathbf{tr}\{\cdot\}$ is the trace of its arguments and it is computed as the sum of the diagonal elements of its arguments, $p_b = tr\{V(\hat{\theta}^*)[V(\hat{\theta}^*) + (1/N)D]^{-1}\}$ and γ is a positive quantity that improves the accuracy of the estimate and can be computed according to the following expression:

$$\gamma = \frac{\hat{\theta}(k)^T D}{N^2}\left(R[\hat{\theta}(k)] + \frac{D}{N}\right)^{-1} R[\hat{\theta}(k)]\left(R[\hat{\theta}(k)] + \frac{D}{N}\right)^{-1} D\hat{\theta}(k)$$

6. Simulation Studies and Discussion of Results

The ARMAX, OE and SSIF mathematical models with their respective model predictors developed in Section 4 as well as the model validation algorithms discussed in Section 5 are applied for the modeling, prediction and validation of power transmission and distribution in Akure and its environs based on the available obtained data evaluated and discussed in Section 3.

6.1 Estimating the ARMAX, OE and SSIF Models

The input vector to the ARMAX, OE and SSIF model predictors are the past values of the inputs (n_b) and outputs (n_a) as well as the order of moving average filter (n_c) which constitute the regression vector each defined from Equation (10) for ARMAX, OE and SSIF as $\varphi_{ARMAX} = [\mathrm{n}_a, \mathrm{n}_b, \mathrm{n}_c]$ and $\varphi_{OE} = [\mathrm{n}_b, \mathrm{n}_f]$ and $\varphi_{SSIF} = [\mathrm{n}_a, \mathrm{n}_b, \mathrm{n}_c]$ respectively. Note that given A, B and C with initial random noise, $P(\theta)$ can be computed from Equation (28) and subsequently $K(\theta)$ from Equation (27). The inputs are the past values contained in the regression vector while the outputs are the predicted values of $\hat{Y}(k)$ given by Equation (17), Equation (21) and Equation (25) for the ARMAX, OE and SSIF model predictors respectively while the optimal value of the adjustable parameters of the models $\theta_{ARMAX}(k)$, $\theta_{OE}(k)$ and $\theta_{SSIF}(k)$ defined in Equation (18), Equation (22) and Equation (26) becomes $\hat{\theta}_{ARMAX}(k)$, $\hat{\theta}_{OE}(k)$ and $\hat{\theta}_{SSIF}(k)$.

For assessing the model prediction performances, the model predictors was trained for τ = 500 epochs (number of iterations) with the following selected parameters: $p = 13$, $q = 13$, $n_a = 4$, $n_b = 4$, $n_c = 4$, $\varphi_{ARMAX}(k,\theta(k)) = 156$ (ARMAX), $\varphi_{OE}(k,\theta(k)) = 105$ (OE) and $\varphi_{SSIF}(k,\theta(k)) = 156$ (SSIF). The details of these parameters are discussed in Section 4; where p and q are the number of inputs and outputs of the system, n_a, n_b and n_c are the orders of the regressors in terms of the past values, $\varphi(k,\theta(k))$ is the total number of regressors (that is, the total number of inputs to the model predictor).

The 3,950 data each collected for the thirteen parameters selected and used for the present case study is divided arbitrarily into two parts: 3160 (80%) to form the training data used for estimating the three models while the remaining 790 (20%) is reserved for the three estimated model validation.

6.2 Validation of the Estimated ARMAX, OE and SSIF Models

According to the discussion on model validation algorithms discussed in Section 5, an estimated model can be used to model a process once it is validated and accepted, that is, the model demonstrates its ability to predict correctly both the data that were used for its development and other data that were not used during the model development.

The results shown in Fig. 4 having (a) to (f) and Fig. 5 having (a) to (g) corresponds to the one-step ahead output predictions of the training data, Fig. 6 having (a) to (f) and Fig. 7 having (a) to (g) corresponds to the one-step ahead output predictions of the validation data, and 8 having (a) to (f) and Fig. 9 having (a) to (g) corresponds to the five-step ahead output predictions or the training data respectively for the thirteen parameters considered, namely: *1)* TVFO, *2)* TVRA, *3)* TCFO, *4)* TCRA, *5)* TPFO, *6)* TPRA, *7)* T2B_Akure, *8)* T2C_Akure, *9)* Oba_Ile, *10)* Iju, *11)* Owena, *12)* Owo, and *13)* Igbara_Oke. Their respective corresponding prediction errors are given in Table 1 for the one-step ahead training, one-step ahead validation and 5-step ahead prediction errors.

Table 1: One-step ahead training, one-step ahead validation and 5-step ahead prediction errors.

		Training Errors			Validation Errors			5-Step Errors			AFPE		
S/N	Parameters	ARMAX	OE	SSIF	ARMAX	OE	SSIF	ARMAX	OE	SSIF	ARMAX	OE	SSIF
1.	TVFO	87.6704	28.8931	-4.8525e13	89.4037	29.4414	2.7417e+03	113.9725	113.2539	274.5787	6.6244	0.5851	6.6195
2.	TVRA	46.2844	17.0458	18.0260	47.2540	17.3787	18.4159	61.3055	21.2727	61.2671	5.3343	0.3071	5.3434
3.	TCFO	120.2449	22.1596	4.7105e+236	51.3808	26.5078	5.0092e+98	475.9057	74.6507	2.1058e223	10.0144	2.9884	10.0150
4.	TCRA	2.0042e+20	20.2810	136.6336	1.4761e+08	21.7391	78.6518	4.9322e+06	75.9705	474.8910	10.0092	0.9838	10.0136
5.	TPFO	0.0433	3.5998e-02	11.4484	4.0586	5.6247e-03	12.2301	57.5389	37.5282	57.5694	5.7821	1.7610	5.7852
6.	TPRA	0.1706	1.3346	NaN	1.5345	0.2014	NaN	43.7927	22.7689	NaN	4.0029	0.9816	4.0101
7.	T2B_Akure	0.7435	2.5805e-02	5.5970	3.6121	0.1253	7.7175	67.0190	56.9740	67.0159	7.4276	2.4064	7.4694
8.	T2C_Akure	21.9985	17.0889	65.9152	20.4470	16.8652	60.9571	120.6421	82.6514	122.6159	8.0102	4.915e-3	8.0318
9.	Oba_Ile	3.2293	0.4425	4.0827e+14	7.5012	3.0375	8.5009e+04	72.8136	62.6882	1.7478e+03	7.4295	1.4078	7.4625
10.	Iju	3.6066	2.8008	4.6102	5.7658	2.9807	6.0487	34.9263	33.0915	34.9619	6.4340	0.4146	6.4520
11.	Owena	21.9652	5.2488	1.4228e+291	22.9434	5.6243	2.1604e+89	48.2391	29.2810	NaN	6.9458	0.9266	6.9559
12.	Owo	NaN	9.0823	209.1187	1.2432e+99	10.7283	229.6607	NaN	62.6867	92.7639	7.9972	0.9765	8.0265
13.	Igbara_Oke	40.1283	4.9606	18.3602	5.6050	0.8871	20.6645	39.3378	28.2818	39.3772	6.8272	0.8040	6.8425

6.3 Validation by the One-Step Ahead Predictions Simulations

In the one-step ahead output prediction method, the errors obtained from one-step ahead output predictions of the estimated model are assessed. In Fig. 4(a)–(f) and Fig. (a) – (g), the graphs for the one-step ahead predictions of the scaled training data (blue -) against the trained network output predictions (red --*) using the estimated ARX and ARMAX models are shown respectively for 500 epochs.

The mean value of the one-step ahead prediction errors are given in the 3rd, 4th and 5th columns of Table 1 for ARMAX, OE and SSIF respectively. It can be seen in these figures that the output predictions of the training data based on the OE model closely match the original training data while Fig. 4(a), (c), (d), (f) and Fig. 5(c), (e), (g) and (g) shows situations where the predictions based on the ARMAX and the SSIF models are completely out of phase in tracking the true training data with much larger errors when compared to the errors obtained based on the OE model as shown in the 3rd, 4th and 5th columns of Table 1. These small one-step ahead prediction errors are indications that the OE model captures and approximate the dynamics inherent in the data to an appreciable degree of accuracy and that the ARMAX and SSIF models predictor can be used for the future power transmission and distribution predictions.

Furthermore, the suitability of the ARMAX, OE and the SSIF model predictors for the model identification and prediction for use in real power transmission and distribution predictions is investigated by validating the estimated models with the 790 validation data that was not used during model predictor development. Graphs of the performances of the estimated ARMAX, OE and SSIF model predictor for one-step ahead output predictions (red --*) of the validation data with the actual validation data (blue -) using the ARMAX, OE and the SSIF models are shown in Fig. 6(a)–(f) and Fig. 7(a)–(g) for the thirteen parameters considered in the present study. Again, as one can observe in these figures that the output predictions of the validation data based on the OE model closely match the validation data while the predictions shown in Fig. (a), (c), (d), (f) and Fig. 7(c), (f) based on the ARMAX and the SSIF models are completely out of phase with the true validation data with much larger errors when compared to the errors obtained based on the OE model as shown in the 6th, 7th and 8th columns of Table 1. These small one-step ahead prediction errors of the validation data are indications that the ARMAX model can again be used for power transmission and distribution predictions in real scenarios.

6.4 *K*–Step Ahead Prediction Simulations

The results of the *K*-step ahead output predictions (red --*) using the *K*-step ahead prediction validation method discussed in Section 5 for 5-step ahead output predictions ($K = 5$) compared with the original training data (blue -) are shown in Fig. 8(a)–(f) and Fig. 9(a)–(g) for the estimated ARMAX, OE and the SSIF models. Again, the value $K = 5$ is chosen since it is a typical value used in most model predictive control (MPC) applications to investigate the capabilities of trained model for future distant predictions. The comparison of the 5-step ahead output predictions performance by the ARMAX, OE and the SSIF models indicate the superiority of the OE model predictor over the ARMAX and SSIF model predictors for distant predictions. This is further justified by the small output prediction errors produced by OE model when compared to relatively large and sometimes infinite (NaN) error produced by ARMAX and SSIF model as can be observed in the 9th, 10th and 11th columns of Table 1.

4.3.3 Akaike's Final Prediction Error (AFPE) Estimates

The implementation of the AFPE algorithm discussed in Section 5.3 for the regularized criterion for the ARMAX, OE and the SSIF model predictors and their respective AFPE estimates are given in the 12^{th}, 13^{th} and 14^{th} columns of Tables 1. These relatively small values of the AFPE estimate obtained from the OE model predictor indicates that the OE model captures the underlying dynamics of the thirteen parameters under investigation to appreciable degree of accuracy and that the model parameters are suitable without the OE model being over-estimated [15, 21]. This in turn implies that optimal OE model parameters have been selected including the weight decay parameters. Again, the results of the AFPE estimates computed for three model predictors based on the OE model are much smaller when compared to those obtained using ARMAX and the SSIF models.

7. CONCLUSION

This paper presents the formulation of an ARMAX, OE and SSIF model predictors for model identification and prediction of power transmission and distribution predictions in Akure and its environs. The 51,350 data used in the study has been obtained from the Power Holding Company of Nigeria, Akure. The results obtained from the application of these three model predictors for the modeling and prediction of power transmission and distribution predictions as well as the validation results show that the OE model outperforms the ARMAX and the SSIF models with much smaller predictions error and good prediction abilities with appreciable degree of accuracy and that the OE model predictor can be deployed for power transmission and distribution predictions in real scenarios.

Although all three models are unstable but the relatively poor performances of the ARMAX and the SSIF models could be attributed to the poor estimation of the moving average filter coupled with the nonlinear nature of the power transmission and distribution data. Even though, it is assumed that the output of the OE model is unknown but estimated based on given inputs, the appreciable excellent performance of the OE model could be attributed to the absence of the moving average filter but with the availability of numerator and denominator model parameters which readily estimated with appreciation degree of accuracy. Thus, the next aspect of the work is on the dynamic modeling and nonlinear model identification of the multivariable nonlinear systems using nonlinear neural network-based approaches for all three model predictors developed here for performance comparison which may give much better predictions and good tracking of the data for much more reliable power transmission and distribution predictions.

ACKNOWLEDGMENTS

The authors wish to express the gratitude to the staff and management of the PHCN, Omotosho Work Centre, 132KV transmission station, Akure, Ondo State, Nigeria for providing the data used in this study.

REFERENCES

[1] Seetha, H. (2006): "Data mining for short-term electric load prediction", *M. Ph. dissertation*, Vellore Institute of Technology (Deemed University), (2006).

[2] Seetha, H. and Saravanan, R. (2007): "Short Term Electric Load Prediction using Fuzzy BP", *Journal of Computing and Information Technology-CIT 15,* vol. 3, pp: 267-282.

[3] Gross, G. and Galiana, F. (1987): "Short-term Load Forecasting", *Proc. IEEE*, Vol. 75, No. 12, pp. 1558–1573.

[4] Lijesen, D. P. and Rosing, J. (1971): "Adaptive forecasting of hourly load based on load measurements and weather information", *IEEE Trans. Power App. Syst.*, Vol. PAS-90, No. 4, pp. 1757–1767.

[5] Pappalexopoulus, A. D. and Hesterberg, T. C. (1994): "A Regression-based approach to short-term load forecasting", *IEEE Transactions on Power Systems*, Vol. 5, No. 4, pp. 1535–1547.

[6] Haida T. and Muto S. (1994): "Regression-based peak load forecasting using a transformation technique", *IEEE Transactions on Power Systems*, Vol. 9, No. 4, pp. 1788–1794.

[7] Charytonuik, W., Chen, M. S. and Van Olinda, P. (1998): "Non-parametric Regression-based Short-term Load Forecasting", *IEEE Transactions on Power Systems*, Vol. 13, No. 3, pp. 725–730.

[8] Amjady, N. (2001): "Short term hourly load forecasting using time-series modeling with peak load estimation capability", *IEEE Transactions on Power Systems* Vol. 16, No. 3, pp. 498–505.

[9] Arafel, L. and Basbous, R. (2009): "Electric Power Load Short Term Forecasting", M. Sc. Thesis, Alquds University.

[10] Dotto, L. (1988): "Thinking the unthinkable: Civilization and Rapid Climate change", *Humanities Press*, U.S.A.

[11] Okunamiya, M. S., Udeozor, O. S. and Imade, O. J. (2009) Investigation and Evaluation of Voltage Drops: A Case Study of Guinness and Ikpoba Dam Injection Substations, *International Journal of Electrical and Power Engineering*, vol. 3, no. 2, pp. 105 – 111.

[12] Akpan, V. A. (Jul., 2011): "*Development of new model adaptive predictive control algorithms and their implementation on real-time embedded systems*", Department of Electrical and Computer Engineering, Faculty of Engineering, Aristotle University of Thessaloniki, Thessaloniki, Greece, Ph.D. Dissertation, 517 pages. [Online] Available: http://invenio.lib.auth.gr/record/127274/files/GRI-2011-7292.pdf.

[13] Akpan, V. A. and Hassapis, G. D. (2011) "Nonlinear model identification and adaptive model predictive control using neural networks", *ISA Transactions*, vol. 50, no. 2, pp. 177 – 194.

[14] Goodwin, G. C. and Sin, K. S. (1984): "Adaptive Filtering, Prediction and Control". *New Jersey, USA: Prentice-Hall.*

[15] Lung, L. (1999): "*System Identification: Theory for the User*". 2nd ed. Upper Saddle River, NJ: Prentice-Hall.

[16] Maciejowski, J. M. (2002): Predictive Control with Constraints, England: Pearson Education Limited.

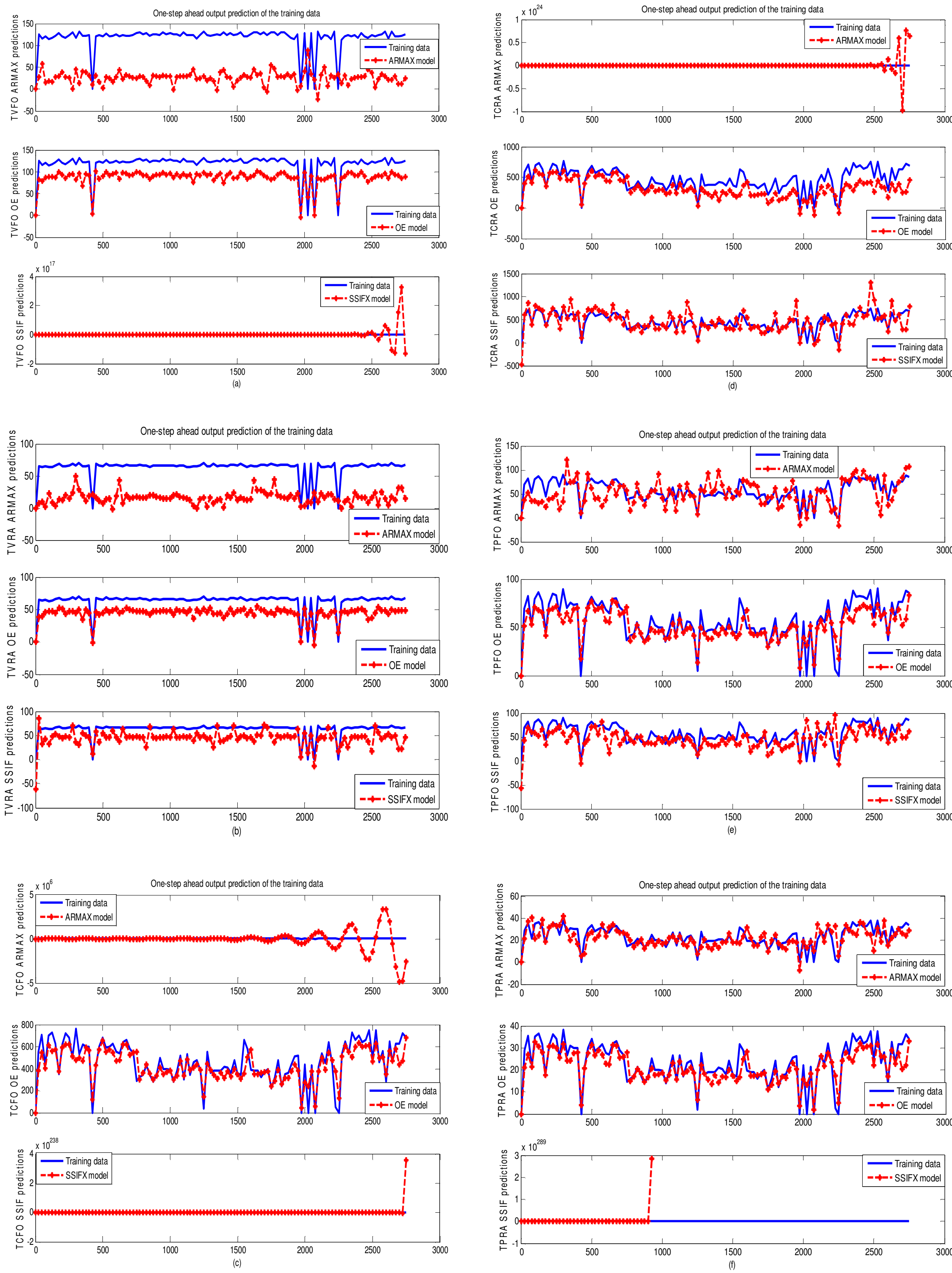

Fig. 4: Comparison of the one-step ahead output predictions of the training data by ARMAX, OE and SSIF for: (a) TVFO, (b) TVRA, (c) TCFO, (d) TCRA, (e) TPFO and (f) TPRA.

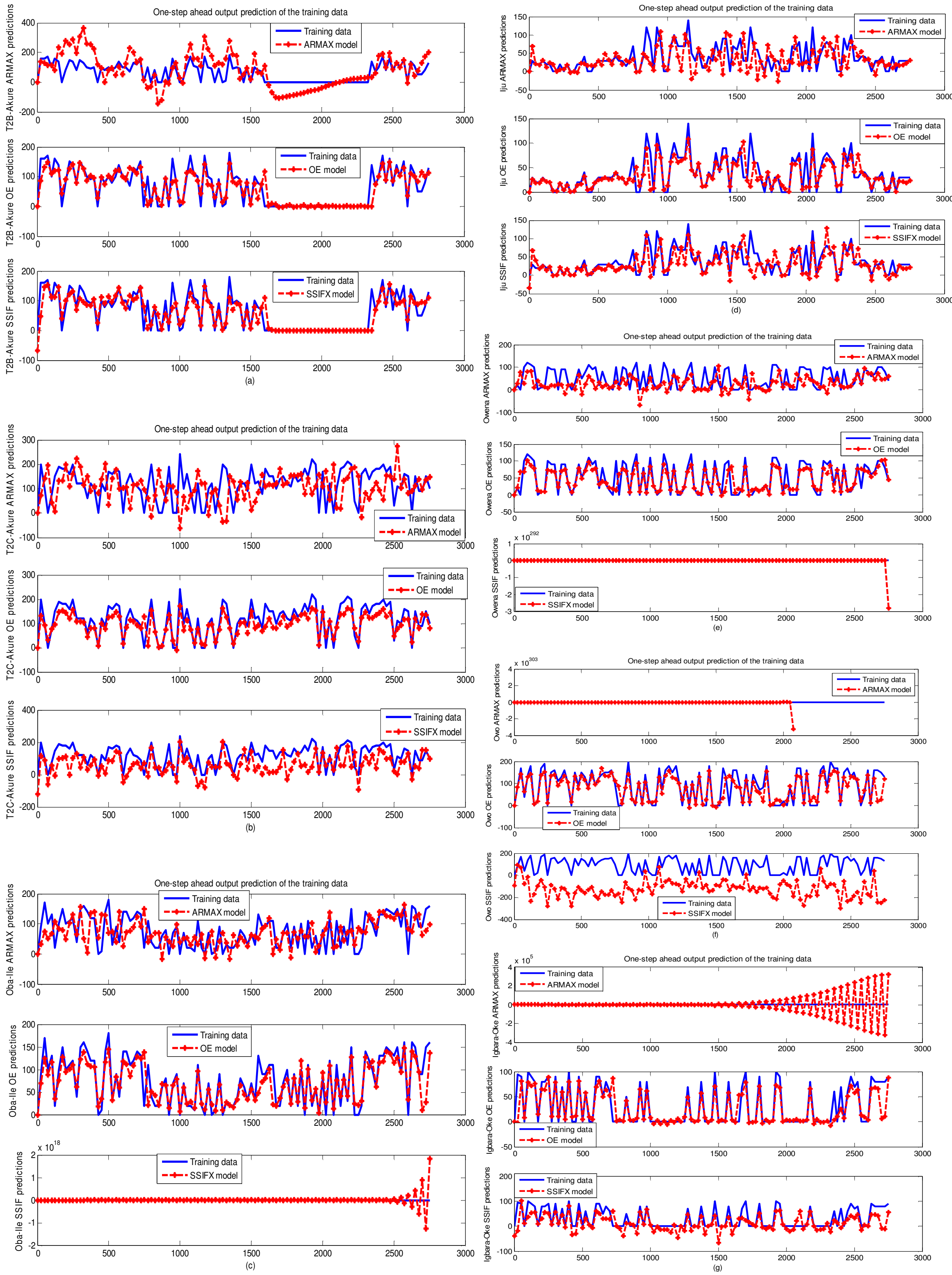

Fig. 5: Comparison of the one-step ahead output predictions of the training data by ARMAX, OE and SSIF for: (a) T2B_Akure, (b) T2C_Akure, (c) Oba_Ile, (d) Iju, (e) Owena, (f) Owo, and (g) Igbara_Oke.

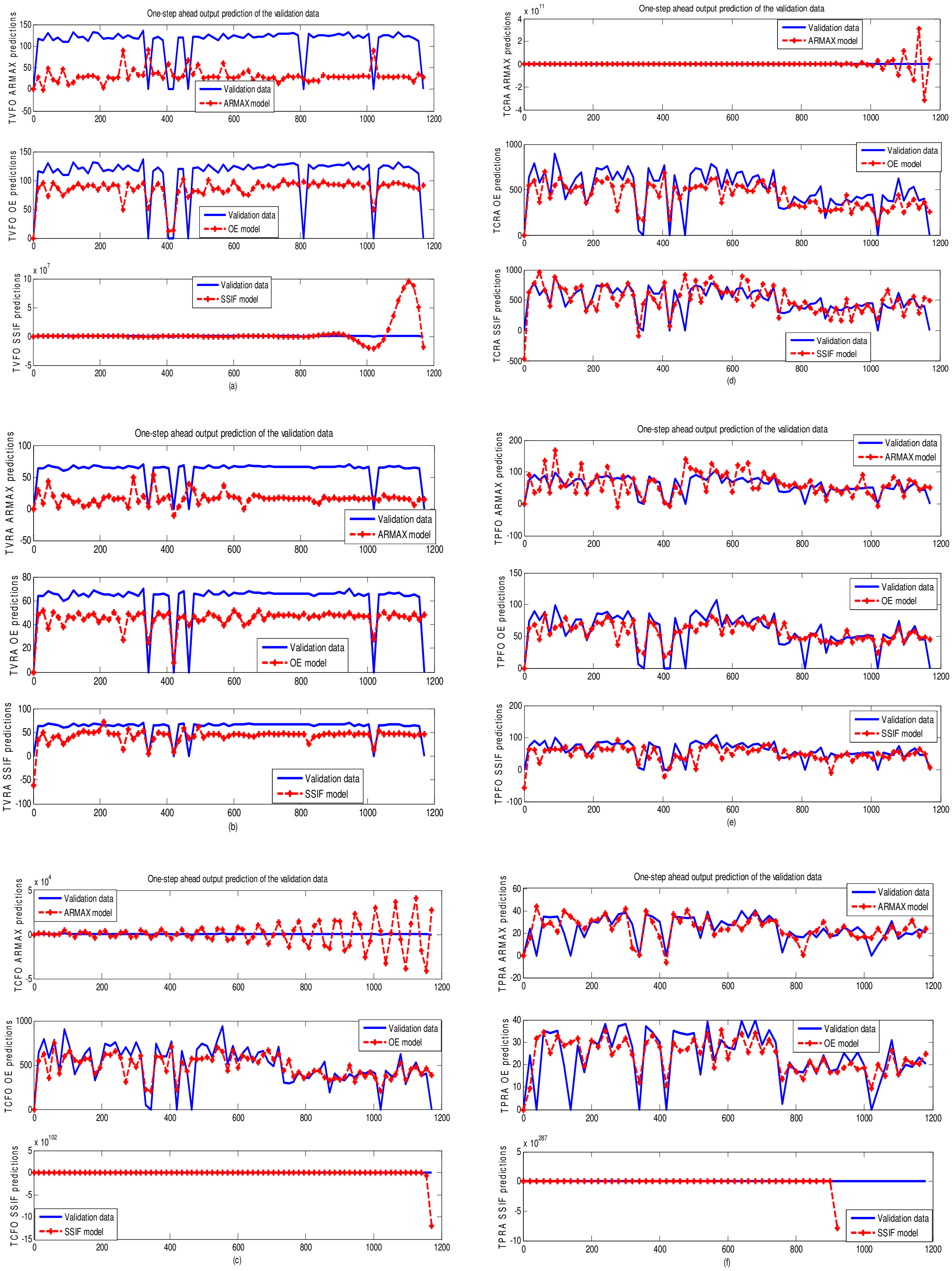

Fig. 6: Comparison of the one-step ahead output predictions of the validation data by ARMAX, OE and SSIF for: (a) TVFO, (b) TVRA, (c) TCFO, (d) TCRA, (e) TPFO and (f) TPRA.

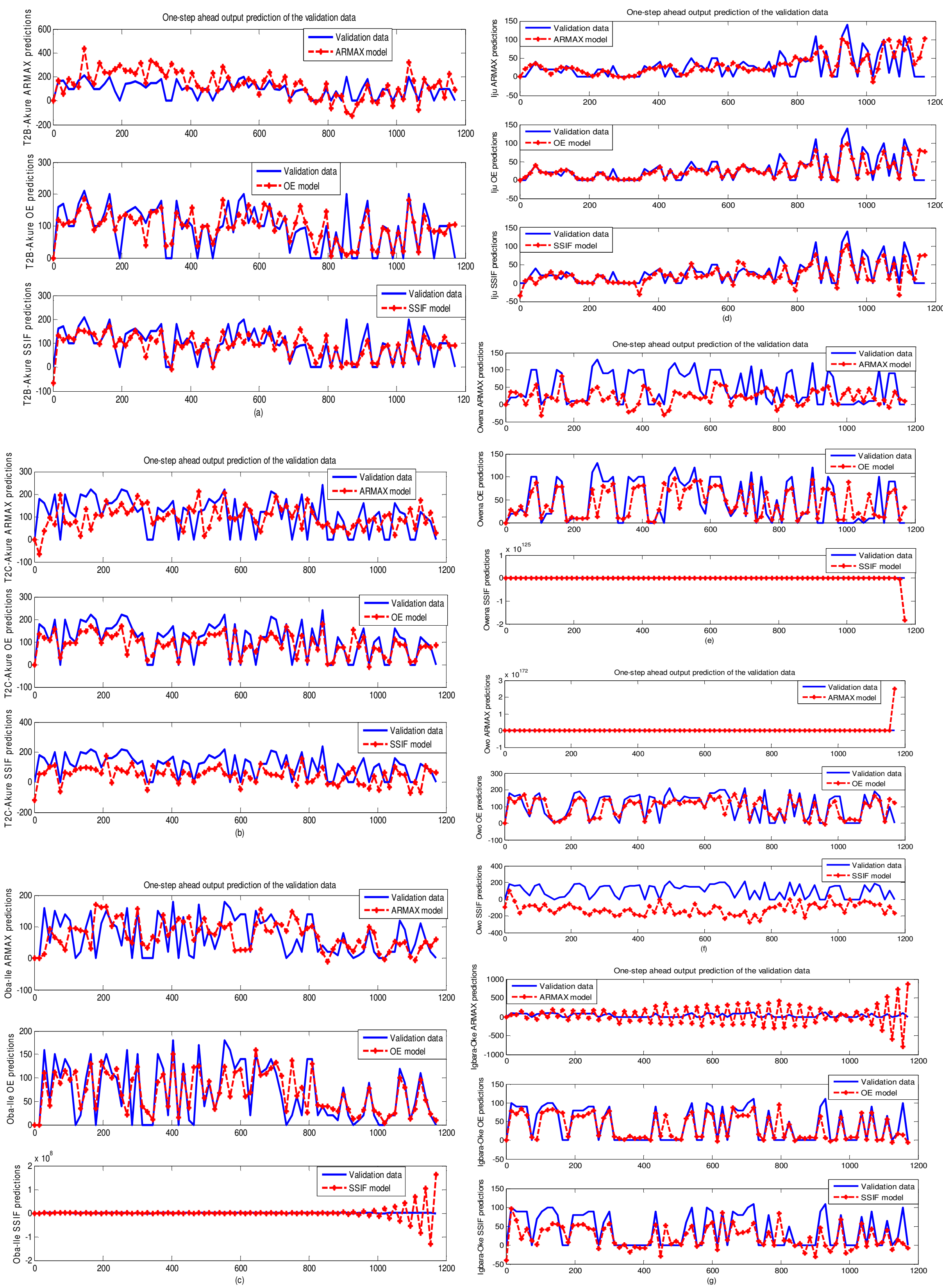

Fig. 7: Comparison of the one-step ahead output predictions of the validation data by ARMAX, OE and SSIF for: (a) T2B_Akure, (b) T2C_Akure, (c) Oba_Ile, (d) Iju, (e) Owena, (f) Owo, and (g) Igbara_Oke.

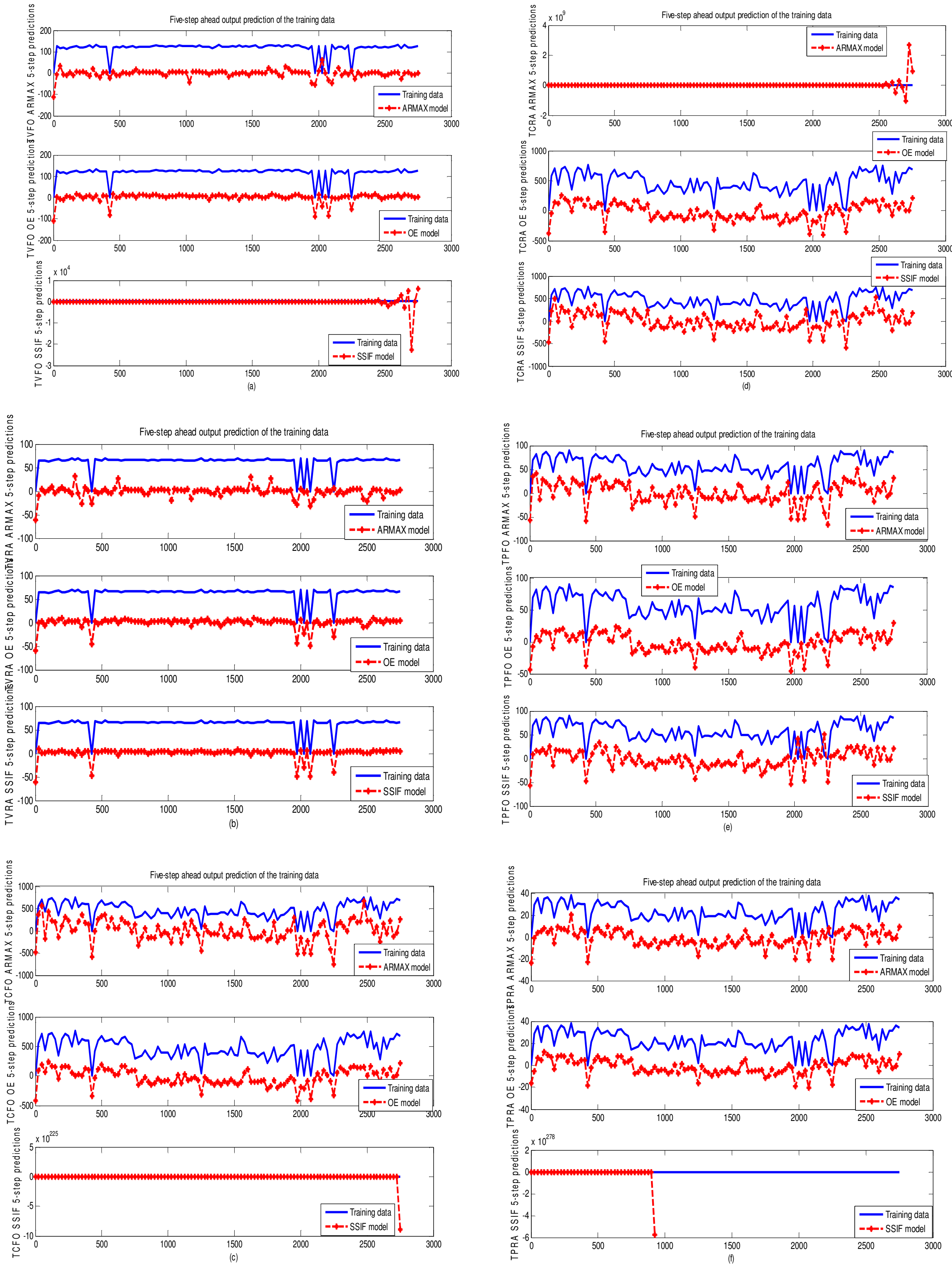

Fig. 8: Comparison of the five-step ahead output predictions of the training data by ARMAX, OE and SSIF for: (a) TVFO, (b) TVRA, (c) TCFO, (d) TCRA, (e) TPFO and (f) TPRA.

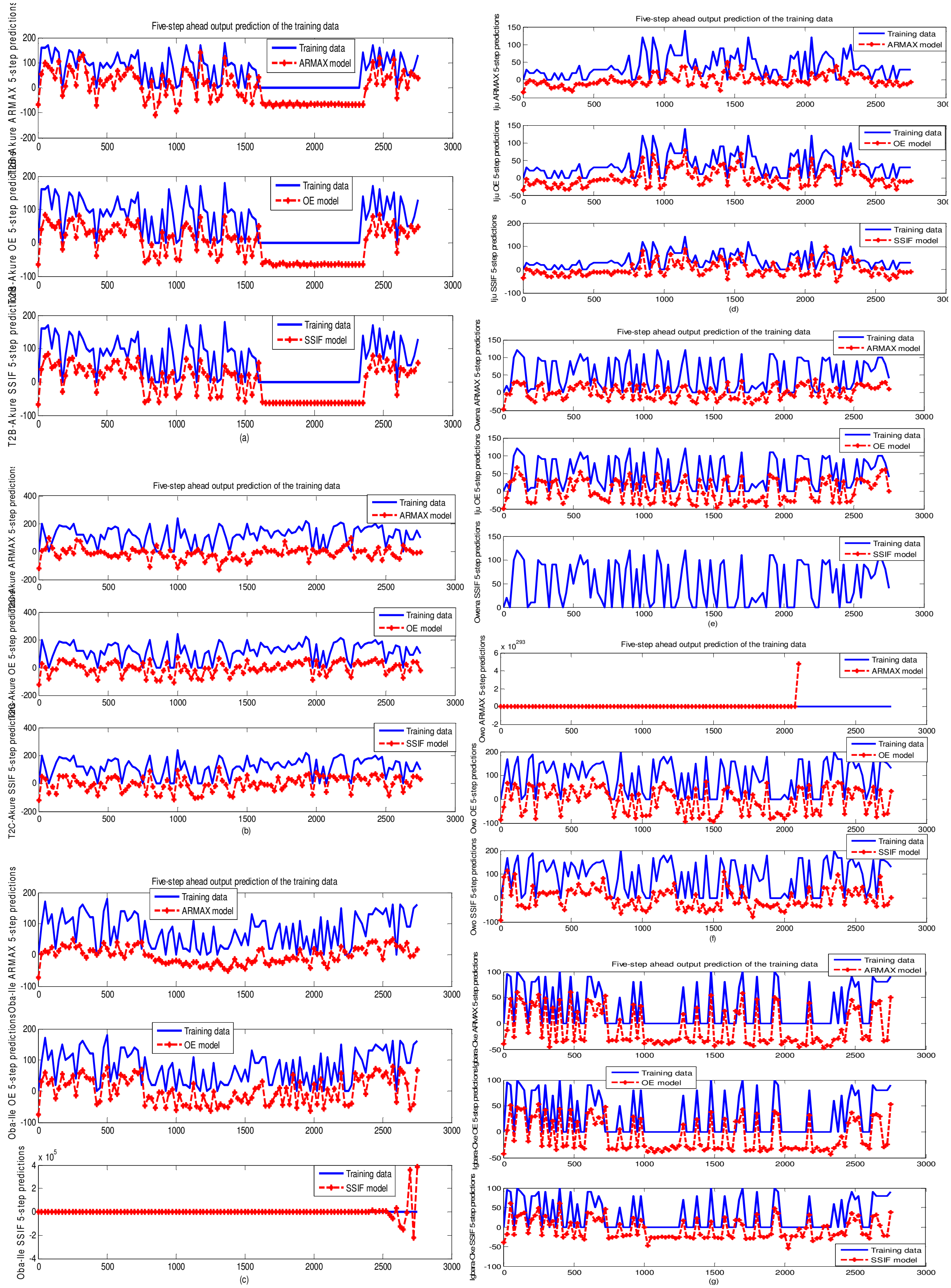

Fig. 9: Comparison of the five-step ahead output predictions of the training data by ARMAX, OE and SSIF for: (a) T2B_Akure, (b) T2C_Akure, (c) Oba_Ile, (d) Iju, (e) Owena, (f) Owo, and (g) Igbara_Oke.

[17] Normey-Rico, J. E. and Camacho, E. F. (2007): *Control of Dead-Time Process*, 2nd ed., London: Springer-Verlag.

[18] Narendra, K. S. and Annaswamy, A. M. (1989): "*Stable Adaptive Systems*". *Englewood Cliff, NJ: Prentice-Hall.*

[19] Söderström,T. and Stoica, P. (1989): "*System Identification*", Prentice Hall, London, U.K.

[20] Kwakernaak, H. and Sivan, H. (1972): "*Linear Optimal Control Systems*", John Wiley and Sons, New York.

[21] Sjöberg, J. and Ljung, L. (1995): Overtraining, regularization, and searching for minimum in neural networks, *International Journal of Control,* vol. 62: 1391-1408.