



HAL
open science

Modeling of IEC 61850 GOOSE Substation Communication Traffic Using ARMA Model

Ronak Feizimirkhani, Antoneta Iuliana Bratcu, Yvon Bésanger, Antoine Labonne, Thierry Braconnier

► **To cite this version:**

Ronak Feizimirkhani, Antoneta Iuliana Bratcu, Yvon Bésanger, Antoine Labonne, Thierry Braconnier. Modeling of IEC 61850 GOOSE Substation Communication Traffic Using ARMA Model. ISGT-Europe - 2019 IEEE PES Innovative Smart Grid Technologies Europe, Sep 2019, Bucharest, Romania. 10.1109/ISGTEurope.2019.8905634 . hal-02292918

HAL Id: hal-02292918

<https://hal.science/hal-02292918>

Submitted on 26 Nov 2019

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Modeling of IEC 61850 GOOSE Substation Communication Traffic Using ARMA Model

Ronak Feizimirkhani^{1,2}

¹Univ. Grenoble Alpes, CNRS, Grenoble INP*,
GIPSA-lab, 38000 Grenoble, France

²Univ. Grenoble Alpes, CNRS, Grenoble INP*,
G2Elab, 38000 Grenoble, France
ronak.feizimirkhani@grenoble-inp.fr

Antoneta Iuliana Bratcu

Univ. Grenoble Alpes, CNRS, Grenoble INP*,
GIPSA-lab, 38000 Grenoble, France
antoneta.bratcu@gipsa-lab.grenoble-inp.fr

Yvon Bésanger

Univ. Grenoble Alpes, CNRS, Grenoble INP*,
G2Elab, 38000 Grenoble, France
yvon.besanger@g2elab.grenoble-inp.fr

Antoine Labonne

Univ. Grenoble Alpes, CNRS, Grenoble INP*,
G2Elab, 38000 Grenoble, France
antoine.labonne@g2elab.grenoble-inp.fr

Thierry Braconnier

Univ. Grenoble Alpes, CNRS, Grenoble INP*,
G2Elab, 38000 Grenoble, France
thierry.braconnier@g2elab.grenoble-inp.fr

Abstract—Since smart grids raised much attention nowadays considering that there exist strong interactions between Information and Communication Technology (ICT) and electric power systems, it is paramount important to model the communication traffic behavior mathematically to be able to assess the cyber-physical interconnection quality, as their interrelated parameters may affect both electrical and communication systems. In this paper we propose a data-driven-based stochastic model of the IEC 61850 Generic Object Oriented Substation Events (GOOSE, a widely used protocol in smart grids) traffic over a real-time traffic generator. The network traffic is well estimated by the Auto Regressive Moving Average (ARMA) model using Box-Jenkins method. Signal processing and identification are performed in MATLAB[®]/Simulink[®] and the model is validated using an intelligent substation test bench. In further works, this model will be used for analysis and control of smart grids as cyber-physical systems.

I. INTRODUCTION

As the modern power grids are more and more automated and often remote controlled, different units are equipped with Intelligent Electronic Devices (IEDs) to connect different management centers and to collect the measured data over the electrical sensors distributed in a vast geographical area. A reliable and high-speed communication network is of main importance to ensure the stability and normal operation of the electrical grid. Since the maintenance, control and security of the network relates on the communication traffic characteristics, concerted research work has been carried out in the evaluating, modeling and characterizing the Substation

Communication Network (SCN) traffic behavior from different point of views [1]. As new smart technologies should be integrated into the existing power grids, the robustness and protective functionality of such a grid is considered by developing a framework to mathematically describe the smart grid as a stochastic hybrid dynamical system. In this method, encoding the dynamical system into symbol strings made it possible to analyze such a complex cyber-physical system [2]. A queueing model has been developed to describe a grouping network of Distributed Grid Intelligences (DGIs) concerning the variable communication delays, furthermore the model has been used to predict the grouping delays [3]. A recent research considers modeling and analyzing the SCN traffic attributes. Distinct from the Internet traffic in this work the self-similarity and Short Range Dependence (SRD) of the SCN traffic have been studied. A mathematical model is proposed for the SCN traffic using time-series analyzing methods and through this model a prediction of the data traffic is explored [4]. Smart grid infrastructure includes two main platforms: namely communication network and electrical power grid. While the communication part has a discrete concept and the electrical part has a continuous one, to understand the interaction between them and to be able to define a model describing both information and electric characteristics is an issue for the control designers.

In this work, we propose a mathematical model to describe the stochastic behavior of the communication traffic. The traffic is modeled as a time-series data, and the stochasticity of the traffic is mathematically described. The proposed model is validated on a substation test bench built within the SINARI collaborative project framework, funded by the French Na-

This work has been financially supported by Grenoble Institute of Technology in the framework of AGIR project CY-PHY-GRID.

* Institute of Engineering Univ. Grenoble Alpes

tional Research Agency (ANR).

This paper is organized as follows. Section II introduces different data traffic patterns of SCN network. Section III describes in detail time-series analyzing method and how the Auto Regressive Moving Average (ARMA) model has been applied to data. In Section IV, measured time series is preprocessed to be as stationary as possible in case of the ARMA model application. Section V presents a related model which is validated by modeling the experimental Generic Object Oriented Substation Events (GOOSE) traffic surveyed over a real-time smart grid traffic generator employing IEC 61850 protocol. Section VI concludes the key points of this work and the model implementation.

A. SCN communication traffic

According to the IEC 61850, SCN data flow over time could be divided into three groups which are periodic, burst and random data [5]. Considering the functionality of each message, it is distributed specifically in the time domain. Metering units are measuring and sending the values to the protection and control unit which includes the time critical information (current, voltage, power, etc.) as Sampled measured Value (SV) and to the server station level which includes the breaker status information as GOOSE messages. This type of messages has normally a fix length, being distributed equally over time. Other type of flow is event driven and triggered by abnormal situations like the trip messages as a short circuit fault or an unusual parameter deviation of the equipment parameters. They are randomly distributed with a probability of p . These types include the protection actions and changing status of the breakers transmitted by the GOOSE messages. When a fault occurs, the flow of GOOSE messages changes from cyclic to burst characteristics. Since in this work the power grid is assumed to be stable and without fault, the type of messages is GOOSE which has a periodic flow according to the definition of GOOSE protocol. The scenario considered here is sending of a predefined active power signal through the SCN while sampling the number of frames generated at the end point of one of the monitoring PC in the network as a receiver.

II. TIME-SERIES ANALYSIS

Time series is a discrete-time sequence of data taken at successive equally spaced points in time. Time-series analysis comprises methods for analyzing time-series data in order to extract meaningful statistics and other characteristics of the data [6]. These methods search for an existing linear relationship between the current and previous observations and also the exogenous events. Considering the SCN traffic, the cumulative number of data frames transmitted is sampled at a fixed period of time. There are different methods to analyze a time series, among which Box-Jenkins and exponential smoothing methods are most common [7]. In this paper, the Box-Jenkins model has been used to express the traffic behavior as a linear polynomial ARMA(p,q) model, where p and q are respectively the auto regressive and moving average polynomial order in terms of the one-sample-delay

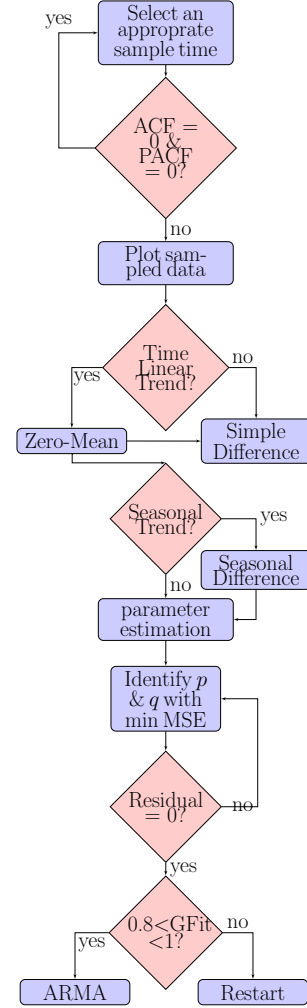


Fig. 1. Model implementation, estimation and the validation [9].

operator z^{-1} [8]. All the model implementation steps are illustrated in the flow chart of Fig. 1. The flowchart, its corresponding acronyms and parameters are explained in the following sections.

1) *Data pre-processing*: To use the ARMA model, the time series should be stationary and uncorrelated so it will be in a kind of statistical equilibrium (constant mean, constant variance). To do so, all the time-dependent characteristics should be smoothed using statistical tools. In order to find the nonstationarity through data, one way is to draw time series over time, then by looking to data, any linear trend will be observed.

Linear time trend: If a linear upward or downward trend exists over time, as mentioned before, it should be removed using a proper statistical tool. To remove a linear time trend a simple difference is used which subtracts each two consecutive values of the time series. If the time series is noted by y_t , then:

$$y'_t = z^{-1}y_t = y_t - y_{t-1}. \quad (1)$$

Series mean value: The mean value is removed from the series to set the data variation over zero:

$$\bar{y}'_t = y'_t - \text{mean}(y'_t), \quad (2)$$

where \bar{y}'_t is the zero-mean series.

Seasonal difference: Sampled data may show a repetitive behavior by a cyclic pattern called seasonal period and it may be recognized by some special statistical tools as Auto Correlation Function (ACF) and Partial Auto Correlation Function (PACF) or by drawing data over time and looking to its behavior [10]. The seasonal trend can be smoothed by a seasonal difference which subtracts each two values spaced by some detected cycle. If s is the seasonal period, w_t is the pre-processed series, and D is the number of seasonal difference required to apply, then the detrended time series is:

$$w_t = \Delta_s^D \bar{y}'_t = (1 - z^{-s})^D \bar{y}'_t \quad (3)$$

A. ARMA model

As shown in Fig. 2, ARMA(p,q) model is a linear filter with a white Gaussian noise as the stochastic input series and the data output which is modeled by a linear polynomial function. ARMA includes two components: Auto Regressive (AR) of order p which expresses the effect of the previous data points on the model, and Moving Average (MA) of order q which accounts for the influence of the white Gaussian noise effectiveness a_t :

$$w_t = \varphi_1 w_{t-1} + \varphi_2 w_{t-2} + \dots + \varphi_p w_{t-p} + a_t + \theta_1 a_{t-1} + \theta_2 a_{t-2} + \dots + \theta_q a_{t-q}, \quad (4)$$

where w_t is the prepared data as the result of the pre-processing steps on the periodically sampled data. φ_i and θ_j are respectively the process and noise coefficients which can be written as polynomial functions of z^{-1} , as below:

$$\begin{aligned} \varphi(z^{-1}) &= 1 + \varphi_1 z^{-1} + \varphi_2 z^{-2} + \dots + \varphi_p z^{-p} \\ \theta(z^{-1}) &= 1 - \theta_1 z^{-1} - \theta_2 z^{-2} - \dots - \theta_q z^{-q} \end{aligned} \quad (5)$$

where $z^{-k} w_t = w_{t-k}$, such that the transfer function of a discrete-time filter is obtained:

$$\psi(z^{-1}) = \frac{\theta(z^{-1})}{\varphi(z^{-1})} \quad (6)$$

B. Model validation

Model validation could be done by Goodness of Fit (GFit), the Minimum Squared Error (MSE), and the whiteness of the data residuals. Model identification refers to estimating the best orders of ARMA(p,q) which minimize the MSE. The estimation GFit is better as closer to 100%, acceptable values should be larger than 80%. Finally the residuals' whiteness will verify if the pre-processing has effectively removed the nonstationarity through data.

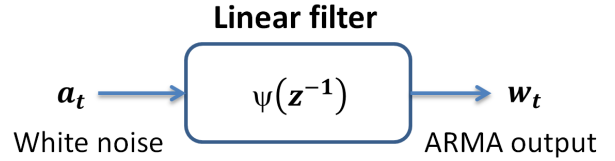


Fig. 2. ARMA(p,q) model could be considered as a linear filter with $\psi(z^{-1})$ as its transfer function.

III. MODEL APPLIED TO THE IEC 61850 GOOSE PROTOCOL

The model is applied to a series of data in which the number of frames has been measured. The frames distributed in the network are measured within stable conditions of the grid, so all the GOOSE messages are periodic.

A. Communication scenario and features

All experiments have been carried out on the test bench depicted in Fig. 3, where the employed devices may be identified. IEC 61850 GOOSE protocol is used in the experiments. It emulates a real MV/LV substation with real electrical devices, likes Remote Terminal Units (RTU) and protection relays. The test bench can be plugged on a real reduced-scale distribution grid, or on a current-injection box. The IEC 61850 protocol is implemented by CET850 simulator, the active power is generated by the ramping module of an Omicron signal generator box. A typical active power signal is broadcast over the network by a S80-Sepam PLC as the sender IED, while all the other units are listening. The power signal sent through the network is periodic of a total duration 144 s, the total profile is presented in Fig. 4. In this case, the power signal is just an example, in a real case, the power profile is not as periodic as this one and its variation through time may happen over a day or so. The power could vary more slowly, such as 24 h instead of 37 s. To generate the active power (P) by Omicron Ramping module, the power factor is selected to be 1, therefore the reactive power is zero. Each ramp (50 Hz three-phase electrical system) is constructed by the related current (I) and the phase-to-neutral voltage (V) values, so $P = 3VI\cos(\varphi)$. The ramping configuration is illustrated in Fig. 5. While the power is sending by the considered IED, the number of frames is registered by a monitoring PC via a network packet analyzer Wireshark which registers evolution of frames over time (including transmission times and packet size). One of the experiences is illustrated by the Wireshark analyzer window in Fig. 6.

B. Sampling and data pre-processing

As the data received are sorted by their transmission time and following the time-series definition, we need them to be sorted at a fixed sampling period, therefore we need to transform data into a sorted data over a sampling period. In order to have data at a constant sampling period, an interpolation by $T_s = 0.2$ ms is applied to data and results in 747 000 samples .

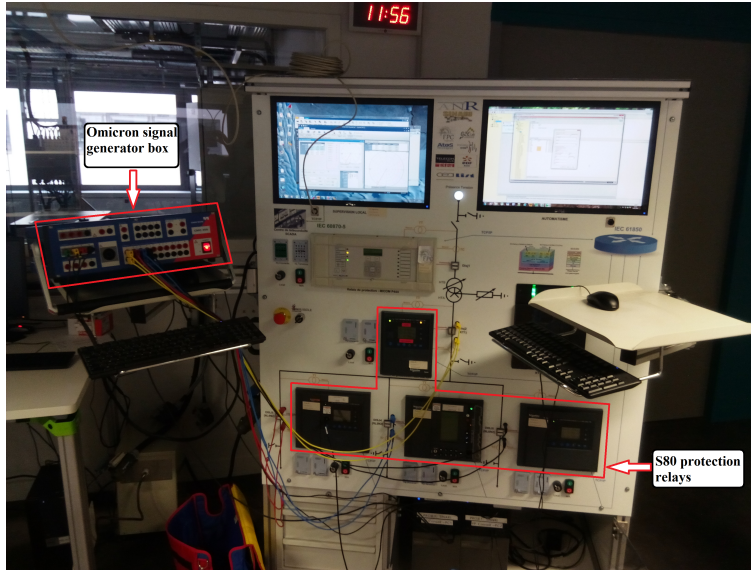


Fig. 3. Photos of the MV/LV substation emulation test bench used in the experiments.

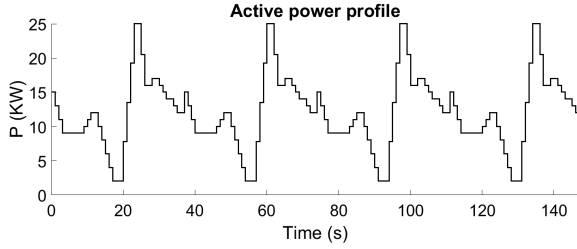


Fig. 4. The active power (kW) variation transmitted by GOOSE messages through the network.

Upward trend: Since the number of frames is cumulative, there is a clear upward trend over time. As previously explained, in case of stationarity, we apply a first-order simple difference to remove this trend. Fig. 7 shows that the upward trend is removed by the simple difference.

Zero mean: The mean value is subtracted to result in a mean value equal to zero.

Seasonality: At this level, by looking over the ACF/PACF or simply to the data drawn over time, four periodic components are recognized at each $s = 37.35 \text{ s} / T_s = 186750$ samples. Fig. 8 illustrates the periodic coefficients around T_s . So the first-order seasonal difference is applied to data and the seasonal trend is smoothed as much as possible. By this level, the remainder of data is mostly random part of the original one, so the ARMA identification can be applied. Fig. 9 shows time series after the seasonal difference application.

C. Model estimation and validation

The best orders leading to the minimum MSE are for AR ($p = 6$) and for MA ($q = 6$) which results in an almost zero MSE (4.73×10^{-8}). In addition, the GFit of this estimation is 82.48%. Finally, by looking through the residuals, we can

judge there still could be a solution to have a better fit since the residuals are not all in between the two confidence levels. Fig. 10 shows the estimation of the stochastic part of data.

Mathematical expression: Since the estimation validation is in an acceptable range, one can find the mathematical expression including all the deterministic components of data behavior. Therefore, one applies all the removed trends (seasonal difference, zero mean, simple difference) *via* linear filters back to an approximation of the original measured data and writes down the related mathematical equations. As the ARMA output is as follows,

$$w_t = \frac{\theta(z^{-1})}{\varphi(z^{-1})} a_t, \quad (7)$$

a sequence of white Gaussian noise a_t is passed through the ARMA(p, q) already identified to get w_t .

Seasonalizing: The seasonal trend is added to the output of ARMA by applying a linear filter as the reversed filter of the seasonal difference:

$$\bar{x}'_t = \left(\frac{\theta(z^{-1})}{\varphi(z^{-1})} a_t \right) \frac{1}{(1 - z^{-s})}, \quad (8)$$

where \bar{x}'_t denotes the series including the seasonal trend.

Adding the mean value:

$$x'_t = \left(\frac{\theta(z^{-1})}{\varphi(z^{-1})} a_t \right) \frac{1}{(1 - z^{-s})} + \text{mean}(y'_t), \quad (9)$$

where x'_t is the series after adding the mean value which was previously removed by the zero-mean process.

Simple difference filtered back: Finally, by applying the inverse of the simple difference – that is, discrete integration – one can find the original measured data with good precision. Fig. 11 compares the original sampled data to its estimation.

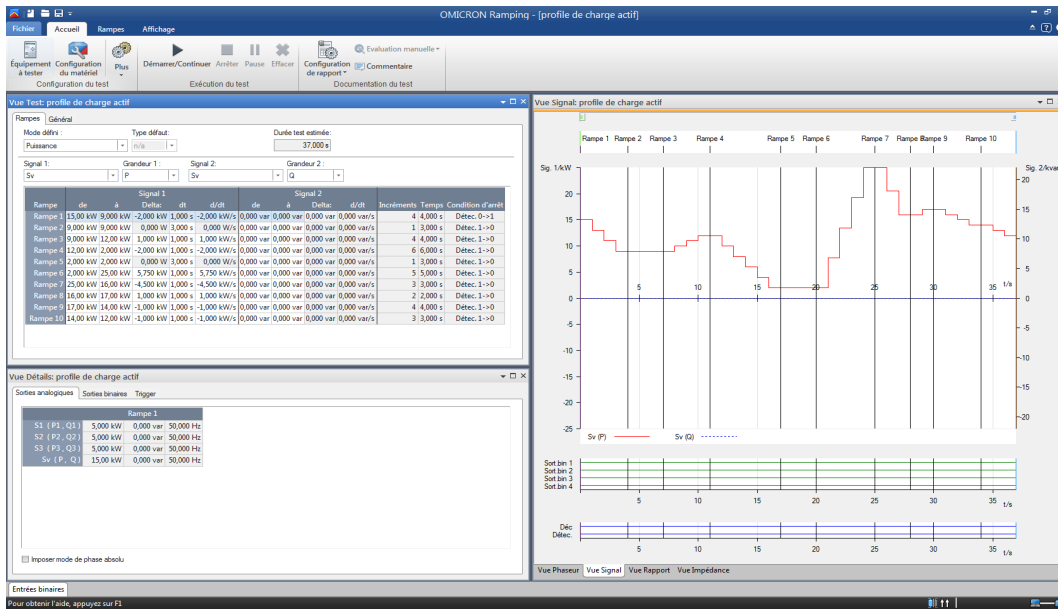


Fig. 5. The active power (kW) is entered in the Omicron Ramping module as piecewise-constant values by means of the corresponding current and voltage values.

No.	Time	Source	Destination	Protocol	Length	Info
1	0.000000000	fe80::94f3:ae4c:7eab:f96b	ff02::1:3	LLMNR	84	standard query 0x4607
2	0.000216000	192.168.66.151	224.0.0.252	LLMNR	64	standard query 0x4607
3	0.093935000	fe80::94f3:ae4c:7eab:f96b	ff02::1:3	LLMNR	84	standard query 0x4607
4	0.093986000	192.168.66.151	224.0.0.252	LLMNR	64	standard query 0x4607
5	0.296957000	192.168.66.151	192.168.66.255	NBNS	92	Name query NB WPAD-00>
6	0.983498000	Telemeca_d4:32:82	Iec-Tc57_01:00:00	GOOSE	151	
7	1.061138000	192.168.66.151	192.168.66.255	NBNS	92	Name query NB WPAD-00>
8	1.825523000	192.168.66.151	192.168.66.255	NBNS	92	Name query NB WPAD-00>
9	2.607974000	fe80::94f3:ae4c:7eab:f96b	ff02::1:3	LLMNR	84	standard query 0x3a80
10	2.608193000	192.168.66.151	224.0.0.252	LLMNR	64	standard query 0x3a80
11	2.714690000	fe80::94f3:ae4c:7eab:f96b	ff02::1:3	LLMNR	84	standard query 0x3a80
12	2.714740000	192.168.66.151	224.0.0.252	LLMNR	64	standard query 0x3a80
13	2.917767000	192.168.66.151	192.168.66.255	NBNS	92	Name query NB WPAD-00>
14	3.031449000	Telemeca_d4:32:82	Iec-Tc57_01:00:00	GOOSE	151	
15	3.681894000	192.168.66.151	192.168.66.255	NBNS	92	Name query NB WPAD-00>
16	4.446264000	192.168.66.151	192.168.66.255	NBNS	92	Name query NB WPAD-00>
17	5.079358000	Telemeca_d4:32:82	Iec-Tc57_01:00:00	GOOSE	151	
18	5.225967000	fe80::94f3:ae4c:7eab:f96b	ff02::1:3	LLMNR	84	standard query 0xb694
19	5.226257000	192.168.66.151	224.0.0.252	LLMNR	64	standard query 0xb694
20	5.335541000	fe80::94f3:ae4c:7eab:f96b	ff02::1:3	LLMNR	84	standard query 0xb694
21	5.335598000	192.168.66.151	224.0.0.252	LLMNR	64	standard query 0xb694
22	5.538547000	192.168.66.151	192.168.66.255	NBNS	92	Name query NB WPAD-00>
23	6.302626000	192.168.66.151	192.168.66.255	NBNS	92	Name query NB WPAD-00>
24	7.067022000	192.168.66.151	192.168.66.255	NBNS	92	Name query NB WPAD-00>
25	7.127271000	Telemeca_d4:32:82	Iec-Tc57_01:00:00	GOOSE	151	
26	7.848206000	fe80::94f3:ae4c:7eab:f96b	ff02::1:3	LLMNR	84	standard query 0xd2da
27	7.848425000	192.168.66.151	224.0.0.252	LLMNR	64	standard query 0xd2da
28	7.956216000	fe80::94f3:ae4c:7eab:f96b	ff02::1:3	LLMNR	84	standard query 0xd2da
29	7.956275000	192.168.66.151	224.0.0.252	LLMNR	64	standard query 0xd2da
30	8.159306000	192.168.66.151	192.168.66.255	NBNS	92	Name query NB WPAD-00>

Fig. 6. In the Wireshark packet analyzer the network transmission is monitored.

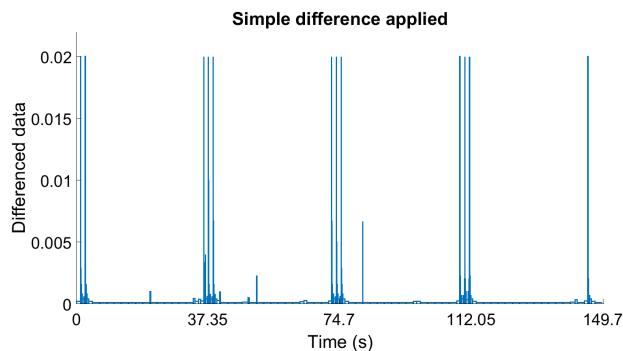


Fig. 7. Time series after simple difference application.

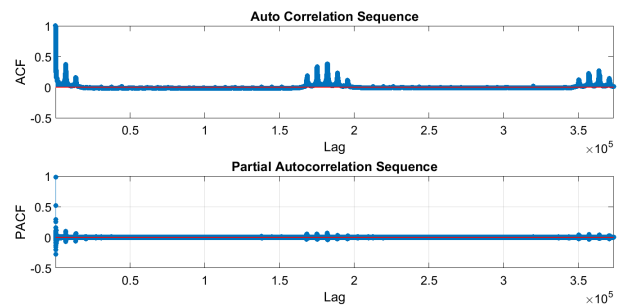


Fig. 8. Related ACF/PACF diagram clearly shows the cyclic behavior.

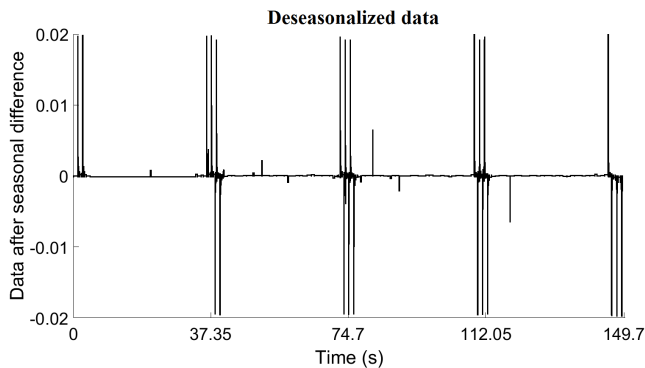


Fig. 9. Time series after seasonal difference application.

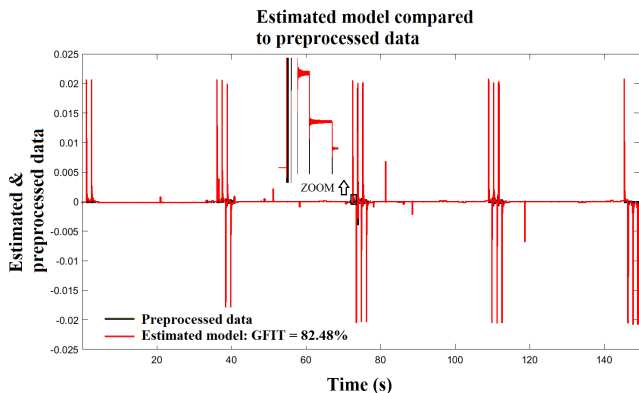


Fig. 10. The estimated data is compared to the sampled data.

The final model of time series estimation, including the deterministic and the stochastic parts, is written below:

$$x_t = \left[\left(\frac{\theta(z^{-1})}{\varphi(z^{-1})} a_t \right) \frac{1}{(1 - z^{-s})} + \text{mean}(y'_t) \right] \frac{1}{(1 - z^{-1})} \quad (10)$$

By estimation of the original data through applying back all the modeling process, the following recurrent equation is obtained:

$$\begin{aligned} w_t = & 0.7085w_{t-1} + 0.1694w_{t-2} + 0.1604w_{t-3} \\ & + 0.1745w_{t-4} + 0.6747w_{t-5} - 0.8896w_{t-6} + a_t \\ & - 0.0435a_{t-1} + 0.0652a_{t-2} - 0.1038a_{t-3} \\ & - 0.2110a_{t-4} - 0.8546a_{t-5} + 0.3268a_{t-6}, \end{aligned} \quad (11)$$

which represents the ARMA(6,6) model of the stochastic part of the traffic behavior.

IV. CONCLUSION

This work has dealt with a model that describes mathematically the IEC 61850 GOOSE traffic sampled over a SCN network where different IEDs exchange information in the context of decentralized control. In the considered scenario it is supposed that the load active power variation is known, then an IED sends out this information to all other IEDs in the network. This scenario has been implemented on a dedicated test bench offering the possibility of real-time data

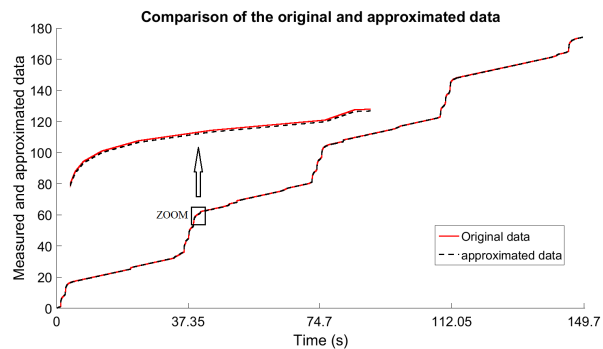


Fig. 11. The original signal is estimated through all the steps back.

packets analysis. After removing the deterministic components – such as trends and cyclic components – the traffic stochastic behavior was modeled by auto regressive moving-average (ARMA) models. From a control approach point of view, the proposed model could serve as a basic mathematical description of a single communication channel, viewed as an elementary brick of the communication network structure within a smart grid. This model will further be complexified in the sense of including modeling of the interaction between the electrical and communication grid, for analysis and control of smart grids as cyber-physical systems. A comparison of the proposed model with a model in [11], based on the queueing theory, which allows the message distribution delay being estimated, is in progress.

REFERENCES

- [1] M. Yu and N. Ansari, "Smart grid communications: Modeling and validation," *Journal of Network and Computer Applications*, vol. 59, pp. 247–249, 2016.
- [2] R. M. Kolacinski and K. A. Loparo, "A mathematic framework for analysis of complex cyber-physical power systems," *IEEE Power and Energy Society General Meeting*, no. July, pp. 1–8, 2012.
- [3] T. Yang, R. Zhao, W. Zhang, and Q. Yang, "On the Modeling and Analysis of Communication Traffic in Intelligent Electric Power Substations," *IEEE Transactions on Power Delivery*, vol. 8977, no. m, pp. 1–1, 2016. [Online]. Available: <http://ieeexplore.ieee.org/lpdocs/epic03/wrapper.htm?arnumber=7479543>
- [4] Z. Cai, M. Yu, M. Steurer, H. Li, and Y. Dong, "A network model for the real-time communications of a smart grid prototype," *Journal of Network and Computer Applications*, vol. 59, pp. 264–273, 2016. [Online]. Available: <http://dx.doi.org/10.1016/j.jnca.2015.05.024>
- [5] Z. Zhang, X. Huang, B. Keune, Y. Cao, S. Member, Y. Li, and S. Member, "8-Modeling and Simulation of Data Flow for VLAN-Based Communication in Substations," pp. 1–12, 2015.
- [6] A. Pankratz, *Forecasting with Univariate Box-Jenkins Models*, 1983.
- [7] Ni Ding, "Load models for operation and planning of electricity distribution networks with smart metering data," Ph.D. dissertation, 2017.
- [8] G. Box, George, Jenkins, Gwilym, Reinsel, Gregory, Ljung, *Time series analysis: forecasting and control*, 2016.
- [9] R. Feizimirkhani, A. I. Bratcu, and Y. Besanger, "Time-series modelling of iec 61850 goose communication traffic between ieds in smart grids—a parametric analysis," in *10th Symposium on Control of Power and Energy Systems (IFAC CPES2018)*, 2018.
- [10] P. J. Brockwell and R. A. Davis, *Introduction to Time Series and Forecasting*.
- [11] Y. Zhang, Z. Cai, X. Li, and R. He, "Analytical Modeling of Traffic Flow in the Substation Communication Network," *IEEE Transactions on Power Delivery*, vol. 30, no. 5, pp. 2119–2127, 2015.