energies

MDPI

Article

ARIMA Models in Electrical Load Forecasting and Their Robustness to Noise

Ewa Chodakowska, Joanicjusz Nazarko and Łukasz Nazarko *

Faculty of Engineering Management, Bialystok University of Technology, Wiejska 45A, 15-351 Bialystok, Poland; e.chodakowska@pb.edu.pl (E.C.); j.nazarko@pb.edu.pl (J.N.)
* Correspondence: l.nazarko@pb.edu.pl

Abstract: The paper addresses the problem of insufficient knowledge on the impact of noise on the auto-regressive integrated moving average (ARIMA) model identification. The work offers a simulation-based solution to the analysis of the tolerance to noise of ARIMA models in electrical load forecasting. In the study, an idealized ARIMA model obtained from real load data of the Polish power system was disturbed by noise of different levels. The model was then re-identified, its parameters were estimated, and new forecasts were calculated. The experiment allowed us to evaluate the robustness of ARIMA models to noise in their ability to predict electrical load time series. It could be concluded that the reaction of the ARIMA model to random disturbances of the modeled time series was relatively weak. The limiting noise level at which the forecasting ability of the model collapsed was determined. The results highlight the key role of the data preprocessing stage in data mining and learning. They contribute to more accurate decision making in an uncertain environment, help to shape energy policy, and have implications for the sustainability and reliability of power systems.

Keywords: ARIMA; electricity load; forecasting; model identification; tolerance to noise; robustness; simulation

check for updates

Citation: Chodakowska, E.; Nazarko, J.; Nazarko, Ł. ARIMA Models in Electrical Load Forecasting and Their Robustness to Noise. *Energies* **2021**, *14*, 7952. https://doi.org/10.3390/en14237952

Academic Editor: Jin-Li Hu

Received: 27 October 2021
Accepted: 22 November 2021
Published: 28 November 2021

1. Introduction

Electrical load forecasting plays a key role in the management and control of a power system. Electricity is a peculiar product—there is currently no practical possibility to store it on a large scale at a desired time. It is necessary to balance energy supply and demand in real time. Imbalance may cause problems with the stability of the power system. Breakdowns resulting from power system instability have serious implications for the sustainability of regional, national, and international energy systems. They may be a cause of many human systems failures and of serious environmental disasters. Precise analysis and forecasting of electric load are necessary to make rational decisions at all levels of energy sector control, management, and policy (technical, managerial, regulatory) as they are closely linked with countries' energy security, resources, and natural environment. Effective forecasting decreases uncertainty, thus allowing for more accurate decisions at the operational, strategic, and policy levels.

The importance of modelling and forecasting of electricity consumption is reflected in numerous studies [1]. There are many modelling approaches dedicated to the volume of demand/consumption/load. Depending on the time horizon, forecasts may be divided into: (i) very short-term load forecasting—VSTLF (up to one hour ahead), (ii) short-term load forecasting—STLF (from one hour to one month ahead), (iii) medium-term load forecasting—MTLF (from one week to one year ahead), and (iv) long-term load forecasting—LTLF (more than one year ahead) [2,3]. Diverse methods are applied depending on the time horizon and the aim of the forecasts. Three main classes of forecasting methods may be distinguished: (i) statistical methods, e.g., exponential smoothing models (ESM), multiple linear regression (MLR), and autoregressive and moving average (ARMA); (ii) artificial intelligence methods, e.g., artificial neural networks (ANN), fuzzy regression models

(FRM), and support vector machines (SVMs); and (iii) hybrid methods comprising of two or more methods from one or both classes [4,5]. A comprehensive review of models and techniques used in load forecasting is presented in [1–6].

Auto-regressive integrated moving average (ARIMA) models are among the most popular approaches in the statistical methods class successfully applied in electrical load forecasting, in VSTLF, STLF, and MTLF tasks [1–6]. ARIMA model identification means specifying the class: moving average—MA, auto-regressive—AR, or mixed —ARMA, and the order. The background for model identification may be found in the general guidelines on the pattern of autocorrelation (ACF) and partial autocorrelation functions (PACF); therefore, much depends on experts' knowledge and experience. Factors shaping the load process in the power system are to a large extent random in nature. Electric load is influenced by highly diverse and non-deterministic human activity, numerous technological processes and their random alterations, changing weather conditions, and other stimuli that are non-deterministic or difficult to include in the modeling. Random noise is an inherent component of all measurement data. It may come from various sources: measuring and transmission instruments as well as factors external to the process. Growth of the noise in the measured signal (observed load time series) significantly impacts the capabilities of forecasting models [7]. Taking into account the specificity of ARIMA models, the level of the random component in a time series (measured by its amplitude) significantly impacts the possibility of the correct model identification [8].

The above considerations have led to the formulation of the following research problem: insufficient knowledge on the impact of noise on the ARIMA model identification in electrical load forecasting. Authors' motivation for this research is two-fold: exploratory and pragmatic. First, the authors desire to fill the knowledge gap that exists in the research on noise laden electrical load time series forecasting with ARIMA models. Second, they wish to provide a robust methodology for the evaluation of the tolerance of ARIMA models to unavoidable noise occurring in the electric load time series. Additional motivation for this research is to provide practical guidelines for the use of ARIMA models in noise laden electrical load forecasting. Achieving those goals would imply more effective load forecasting, thus better-informed decisions in managing power systems.

The contribution of this paper to forecasting theory and practice is threefold:

- Development of studies on the impact of noise in time series on the forecasting model identifiability and their robustness,
- Assessment of the tolerance to noise of ARIMA models,
- Formulation of practical guidelines for the use of ARIMA models in noise laden electrical load forecasting.

The paper has the following structure. In this section (Section 1), the research problem is formulated and the motivation for the study is offered. Section 2 presents the essence of ARIMA models and their identification. Section 3 provides a review of the publications on ARIMA models in electrical load forecasting. Sections 2 and 3 are the basis for the formulation of the problem solution. Section 4 describes the adopted research methodology and the simulation experiment design. Section 5 presents the results of the experiment. The subsequent sections explore the ARIMA model identification for real load data and impact of random noise on ARIMA models identification and their ability to predict electrical load time series. Section 6 presents the discussion on the contribution of the results to the development of forecasting methodology and practice of load forecasting. The article ends with conclusions.

2. ARIMA Models and Their Identification

The ARIMA class of models, also referred to in the literature as the Box–Jenkins models due to the ground-breaking contribution of G.E.P. Box and G.M. Jenkins (1970) [9], integrate the autoregressive AR(p) and moving average MA(q) component so that:

$$Y_t = \phi_1 Y_{t-1} + \cdots + \phi_p Y_{t-p} + e_t - \theta_1 e_{t-1} - \cdots - \theta_q e_{t-q}, \quad (1)$$

where:

ϕ_i—coefficient at observation Y_{t-i},
e_t—error distributed as white noise,
p—the order expresses the earliest value included,
θ_i—coefficient at error e_{t-i},
q—the order of the series depends on the earliest previous error.

Using the backshift operator: $B^j y_t = y_{t-j}$, Equation (1) may be expressed by Equation (2):

$$\phi_p(B)Y_t = \theta_q(B)e_t \quad (2)$$

where:

$\phi_p(B) = 1 - \phi_1 B - \cdots - \phi_p B^p$—moving-average operator, represented as a polynomial in the backshift operator;
$\theta_q(B) = 1 - \theta_1 B - \cdots - \theta_q B^q$—autoregressive operator, represented as a polynomial in the backshift operator.

The ARMA assumes that the process is stationary. It means that time series has at least constant mean and variance, and its covariance function depends only on the time difference. When nonstationarity is observed, data transformations are needed. Considering nonstationarity in variance, logarithm transformation is the most popular solution, whereas nonstationarity in mean is commonly removed by differencing. Differenced processes are modelled by auto-regressive integrated moving average—ARIMA(*p*,*d*,*q*)—and the general form for the model is:

$$\phi_p(B)(1-B)^d Y_t = \theta_q(B)e_t \quad (3)$$

where:

d—order of integration (differencing).

The standard process of ARIMA modelling covers six consecutive phases: (i) preliminary analysis, (ii) transformation to stationary, (iii) identification of the components, (iv) parameters estimation, (v) testing, and (vi) application.

Identification of an adequate ARIMA model depends on the autocorrelation and partial autocorrelation pattern. The general idea is that ACF of the *p*-order AR process decays gently (exponentially), whereas the PACF cuts off after the *p*-th lag. In contrast, the ACF of the MA process of *q* order cuts off after the *q* lag, whereas the PACF gently decreases. If both ACF and PACF decay exponentially, this suggests a mixed ARMA process. A popular approach to the determination of the appropriate order of ARIMA is based on fitting [10]. A common approach is the automation of model identification. It relies on the algorithmic comparison of models with different parameters and results in the choice of a model which best fulfils the fit criteria [11–13]. Models that pass the Ljung-Box test are accepted as statistically adequate. To compare and determinate the fitting accuracy, several other criteria are used: simple statistical metrics, such as mean absolute error (MAE), mean absolute percentage error (MAPE), final prediction error (FPE), Akaike information criterium (AIC), or Bayesian information criteria (BIC) [14–16].

There is a large variety of ARIMA models. In the case of series with a seasonal component, the seasonal ARIMA (SARIMA) model may be used. The general notation of the SARIMA seasonal model is: ARIMA(*p*,*d*,*q*)(*P*,*D*,*Q*)*s*, where *s* is a number of seasons in the seasonal cycle:

$$\Phi_P(B^s)\phi_p(B)(1-B)^d(1-B^s)^D Y_t = \Theta_Q(B^s)\theta_q(B)e_t, \quad (4)$$

where:

$\Phi_P(B^s) = 1 - \Phi_1 B^s - \cdots - \Phi_P B^{Ps}$—seasonal autoregressive operator,
$\Theta_Q(B^s) = 1 - \Theta_1 B^s - \cdots - \Theta_Q B^{Qs}$—seasonal moving-average operator.

It is worth emphasizing that ARIMA models also consider multiple seasonality, which is especially useful in the case of energy forecasting:

$$\Omega_{P_2}(B^{s_2})\Phi_{P_1}(B^{s_1})\phi_p(B)(1-B)^d(1-B^{s_1})^{D_1}(1-B^{s_2})^{D_2}Y_t = \psi_{Q_2}(B^{s_2})\Theta_{Q_2}(B^{s_2})\theta_q(B)e_t, \quad (5)$$

where:

$\Phi_{P_1}(B^{s_1}) = 1 - \Phi_1 B^{s_1} - \cdots - \Phi_{P_1} B^{P_1 s_1}$—first seasonal autoregressive operator,
$\Theta_{Q_1}(B^{s_1}) = 1 - \Theta_1 B^{s_1} - \cdots - \Theta_{Q_1} B^{Q_1 s_1}$—first seasonal moving-average operator,
$\Omega_{P_2}(B^{s_2}) = 1 - \Omega_1 B^{s_2} - \cdots - \Omega_{P_2} B^{P_2 s_2}$—second seasonal autoregressive operator,
$\psi_{Q_2}(B^{s_2}) = 1 - \psi_1 B^{s_2} - \cdots - \psi_{Q_2} B^{Q_2 s_2}$—second seasonal moving-average operator.

Further extensions and variations of the classic ARIMA are models that include exogenous series as input variables, referred to as an ARIMAX:

$$Y_t = \mu + \sum_i \frac{\omega_i(B)}{\delta_i(B)} B^{k_i} x_{i,t} + \frac{\theta_q(B)}{\phi_p(B)} e_t \quad (6)$$

where:

$\omega_i(B)$—numerator polynomial of the transfer function for the ith input series,
$\delta_i(B)$—denominator polynomial of the transfer function for the ith input series,
k_i—the pure delay for effect of $x_{i,t}$. at time t.

Other variants include the multivariate approach, e.g., vector ARIMA (VARMA). When d is a fraction rather than an integer, the process is called fractionally integrated ARMA (ARFIMA or FARIMA). There are also models like ARCH/GARCH (generalized auto-regressive conditional heteroskedasticity) to deal with data with nonconstant auto-correlated variance. Commonly applied ARIMA based time series modeling approaches are hybrids derived from models of the same family, e.g., AR-GARCH—AR models with GARCH residuals or based on models with dissimilar assumptions, e.g., ARIMA and neural networks. In the case of energy load prediction, it is justified to use the Reg-ARIMA compound of regression and ARIMA time series errors, for example, hourly temperature data as a regressor [17].

Glossary of all abbreviations and acronyms used in this article can be found in Table 1.

Table 1. List of abbreviations and acronyms used in this article.

Abbreviation	Meaning
ACF	Autocorrelation Functions
AIC	Akaike Information Criterium
ANN	Artificial Neural Network
AR	Auto-Regressive
ARCH	Auto-Regressive Conditional Heteroskedasticity
ARFIMA	Fractionally Integrated ARMA
AR-GARCH	AR Models with GARCH Residuals
ARIMA	Auto-Regressive Integrated Moving Average
ARIMAX	Auto-Regressive Integrated Moving Average with Exogenous Variable
ARMA	Auto-Regressive and Moving Average
BIC	Bayesian Information Criteria
ESM	Exponential Smoothing Models
FARIMA	ARFIMA
FPE	Final Prediction Error
FRM	Fuzzy Regression Models
GARCH	Generalized Auto-Regressive Conditional Heteroskedasticity
LTLF	Long-Term Load Forecasting
MA	Moving Average
MAE	Mean Absolute Error

Table 1. *Cont.*

Abbreviation	Meaning
MAPE	Mean Absolute Percentage Error
MLR	Multiple Linear Regression
MTLF	Medium-Term Load Forecasting
MW	Megawatt
NSR	Noise to Signal Ratio
PACF	Partial Autocorrelation Functions
SARIMA	Seasonal Auto-Regressive Integrated Moving Average
STLF	Short-Term Load Forecasting
SVM	Support Vector Machines
VARMA	Vector ARIMA
VSTLF	Very Short-Term Load Forecasting

3. Background Literature

In the last dozen or so years, Clarivate WoS, the Scopus database, and IEEE Xplore have recorded several hundred publications each year, in which key words include the terms ARIMA and "electricity" or "energy" and "volume" or "demand", "consumption", "power", "load". The number of relevant publications from the last 10 years reported by different scientific databases is presented in Figure 1.

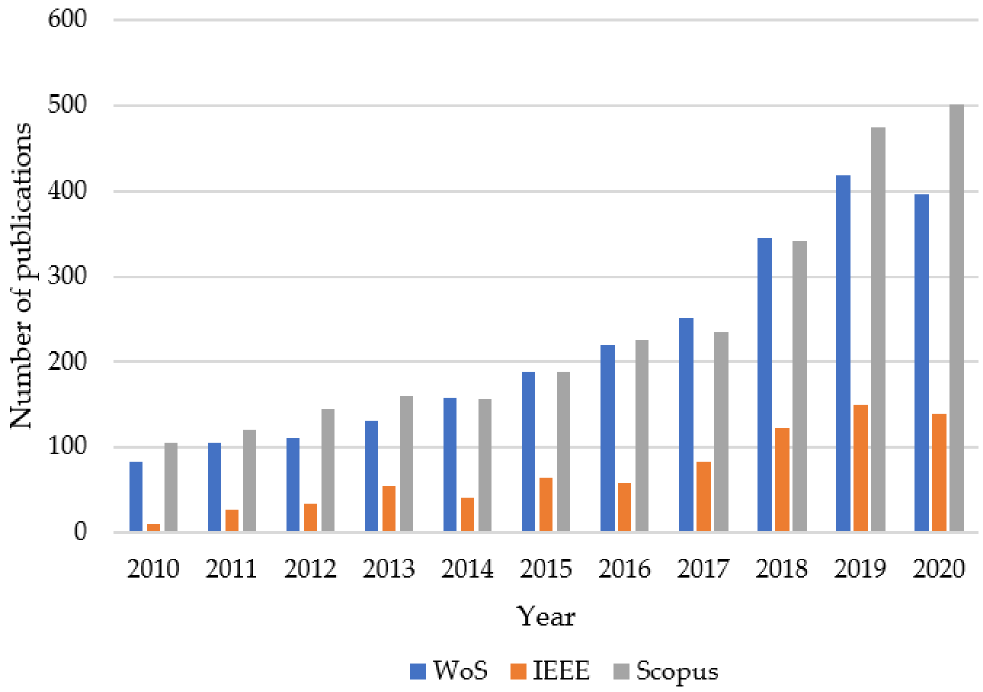


Figure 1. Number of publications with keywords "ARIMA" and "electricity" or "energy" and "volume" or "demand", "consumption", "power", "load" in Clarivate WoS, Scopus, and IEEE Xplore databases.

ARIMA models are among the most popular forecasting techniques in the energy sector alongside artificial neural network (ANN), support vector machines (SVM), and uncertainty solving approaches under discrete data such as grey or fuzzy [18] or rough [19] and are often employed in hybrid approaches [20]. A large group of articles focuses on comparing or combining approaches. The models used are, among others: noted earlier neural networks [21–24], linear regression with ARIMA [21,25], Holt-Winters or exponential smoothing [21,26], and seasonal and trend decomposition using loess [27], metabolic grey model [28], or data mining [29]. Often it is the hybrid approaches that are indicated as providing better accuracy. However, approaches to increase predictability also include, for example, innovative data filtering methods [30].

Considering applied areas, ARIMA is used to forecast energy issues at various levels of data aggregation, based on timeseries or panel data of a group of countries, e.g., the EU [31], individual countries [26], sectors of the economy [31–33], institutions [25,34], or production processes [22]. Among the articles that forecast the demand/consumption/load of energy in general, a popular subject is modelling the generation/consumption of energy from renewable sources. Forecasts address the development of renewable energy consumption in total [35] or from particular sources, i.e., wind [36–38], hydropower [39–41], solar [42], thermal [43], or biogas [44]. ARIMA models are also used for modelling and forecasting that are inextricably linked with energy CO_2 generation [45–47]. Another area is price forecasting [48] or a volatility index [49].

Considering the topic of this paper, the authors focus the literature review on the class of ARIMA models used in electric load forecasting. ARIMA models employed in load forecasting and modelling tasks identified in the literature are shown in Table 2. Data used to construct and verify a model, type of forecast, ARIMA model specification, and the source publication are presented.

Table 2. Identified ARIMA models used in load forecasting.

No	Year	Data/Forecast	ARIMA Model	Publications
1	2021	10-year monthly load, Electricité Du Cambodge, Cambodia Multi-step, monthly load one-year ahead	ARIMA(2,0,2)(4,0,2)12	Nop and Qin [50]
2	2020	47-month daily load of household, France Multi-step, daily load 7, 14, 28, and 31 days ahead	ARIMA(1,0,2)	Mpawenimana et al. [51]
3	2020	3-year daily 10 a.m. load, SLDC, Assam Multi-step, daily 10 a.m. load one-year ahead	(a) ARIMA(1,2,2) (b) ARIMA(0,1,1)(0,1,1)7	Goswami and Kandali [52]
4	2020	10-year daily load, Karnataka, India Multi-step, daily load month-wise one-year ahead	Twelve models, each for a certain month January ARIMA(8,1,1) February ARIMA(5,4,1) March ARIMA(9,1,1) April ARIMA(5,5,1) May ARIMA(1,6,1) June ARIMA(4,3,1) July ARIMA(3,3,1) August ARIMA(8,2,1) September ARIMA(3,3,1) October ARIMA(2,6,1) November ARIMA(9,1,1) December ARIMA(5,3,1)	Gupta and Kumar [53]
5	2020	10-year monthly load, Shaoxing, China Multi-step, monthly load one-year ahead	ARIMA(12,2,9)	Wang et al. [54]
6	2020	1-year hourly load, Mahakam East Kalimantan, Indonesia One-step, hourly load from one-week to one-month ahead Multi-step, hourly load from one-week to one-month ahead	ARIMA (0,1,1)(0,1,1)24(0,1,1)336	Dinata et al. [55]
7	2019	3-day 5 min interval load, Sichuan Province, China Multi-step, 5 min load 30 min ahead for one day	ARIMA(2,2)	Yang and Yang [56]
8	2019	126-weak daily load, Taiwan Power Company, Taiwan Daily load one-day ahead Multi-step, daily load one-week ahead	ARIMA(1,1,1)(1,1,1)7	Yu, Hsu and Yang [57]

Table 2. *Cont.*

No	Year	Data/Forecast	ARIMA Model	Publications
9	2019	3-year daily load, agriculture, PG&E, US Multi-step, monthly load one-year ahead	ARIMA(0,1,0)(1,1,1)12	Noureen et al. [32]
10	2019	14-year daily load, Toronto Canada Multi-step, daily load one-week ahead	General model (a) ARIMA(2,0,8) Seven models, each for a certain day of week (b): Mon ARIMA(0,1,10); Tue ARIMA(0,1,10); Wed ARIMA(0,1,12); Thu ARIMA(0,1,10); Fri ARIMA(0,1,9); Sat ARIMA(0,1,10); Sun ARIMA(0,1,9)	Tang, Yi and Peng [58]
11	2019	3-month 15 min load, Builders Temporary Supply Multi-step, 15 min load one day ahead	ARIMA(2,1,1)	Amin and Hoque [59]
12	2018	One-week 15 min load, Shiqu, Ganzi State, China Multi-step 15 min load 12 h ahead	ARIMA(13,1,15)	Zou et al. [60]
13	2017	N/A N/A	The most commonly used seasonal ARIMA is probably the ARIMA(0,1,1)(0,1,1)	Kuster, Rezgui, and Mourshed [6]
14	2017	2-year hourly peak load, TX, US Multi-step, hourly peak load two years ahead	Three seasonal periods are defined as 24, 168, and 8766 for daily, weekly, and annually effect respectively	Eljazzar and Hemayed [61]
15	2017	2-year hourly load, southern region, India Multi-step, hourly load one day ahead	ARIMA(4,0,1)24(2,1,2)168	Karthika, Margaret, and Balaraman [62]
16	2016	20-year monthly load, Regional Transmission Organization, US Multi-step, monthly load one year ahead	ARIMA(1,1,1)	Khuntia, Rueda, and van der Meijden [63]
17	2012	2-year half-hourly load, Java-Bali Indonesia Multi-step, half-hourly load two weeks ahead separately for each half-hour of a day	ARIMA(0,1,1)(0,1,1)7	Suhartono et al. [64]
18	2009	6-year monthly load data, China Multi-step, monthly load six months ahead	ARIMA(4,1,4)	Wei and Zhen-gang [65]
19	2009	One-week, 5 min load, substations located at Andradina, Ubatuba, and Votuporanga, Brazil Multi-step, 5 min load twelve steps ahead	ARIMA(3,2,2)(0,1,1)12 ARIMA(4,2,2)(0,1,1)12 ARIMA(3,2,2)(0,1,1)12	de Andrade and da Silva [66]
20	2006	33-year daily load, UK Multi-step, daily load four years ahead	ARIMIA(1,1,1)	Hor, Watson, and Majithia [67]
21	2006	1-year 15 min load, Hebei province, China Multi-step, 15 min load one day ahead	ARIMA(2,2,3)	He, Zhu, and Duan [68]
22	2005	2-month 15 min load of distribution substations, Bialystok, Poland Daily load profile modelling	ARIMA(0,1,1)(0,1,1)96	Nazarko Jurczuk, and Zalewski [69]
23	2004	1-year 15 min load, Hebei area, China Multi-step, 15 min load one day ahead	ARIMA(2,2,3)	Ran-chang Lu et al. [70]
24	1999	6-month 1 h load series, Red Electrica de Espana, Spain Multi-step, hourly load one day ahead	Logarithmic transformation ARIMA(1,1,0)(1,1,1)24(0,1,1)168	Juberias et al. [71]

Table 2. *Cont.*

No	Year	Data/Forecast	ARIMA Model	Publications
25	1995	1-year 5 min, Taipower, Taiwan Multi-step, hourly load one week ahead	Commercial load ARIMA(1,0,0)(2,1,1)24 Office load ARIMA(2,0,0)(1,0,0)168(0,1,1)24 Industrial load, logarithmic transformation ARIMA(1,0,2) Residential load, logarithmic transformation ARIMA([1,4,5],00)(0,1,1)24	Cho, Hwang, and Chen [72]

The review synthesized in Table 2 is a clear indication of the number and the variety of ARIMA models employed in load forecasting tasks. Single seasonal models (models 1,3b,8,9,13,17,19,22,25), double seasonal models (models 6,24,25), as well as models without a seasonal component (models 2,3a,4,5,7,10–12,16,18,20–23,25) are used. They may contain the AR and MA affixes or only the MA affix (models 3b,6,10b,13,17,22). They also may or may not include differencing (models 1,2,10a). Values determining the order of AR and MA components and the degree of differencing vary considerably. Closer analysis leads to the conclusion that the model type is closely tied with the length of the output time series and the load probing period, as well as the forecast step and horizon. ARIMA models are used to forecast the electric load with different time horizons: VSTLF (models 7,11,12,17,19,21,23), STLF (models 2,4,6,7,8,10,11,12,15,17,20,21,23,24,25), MTLF (1–5,9,16,18,20), and LTLF (14,20). Some models have been used to prepare mixed forecasts like VSTLF-STLF (models 7,11,12,17,21,23), STLF-MTLF (models 2,4,6,20), MTLF-LTLF (models 14,16,20), and STLF-MTLF-LTLF (model 20). ARIMA models are most frequently used in STLF, MTLF, and VSTLF forecasting tasks, less frequently in LTLF tasks. Typically, the model type has been selected on the basis of a preliminary analysis of the load curve in order to assess the occurrence of trend and seasonality and next on the basis of the ACF and PACF function plots. Sometimes the model selection was more mechanical; it was based on the comparison of many models to a chosen criterion function without deeper inspection of the time series structure (e.g., models 2,4,18,21).

Review of ARIMA models employed in electric load forecasting points at the importance of proper model identification. Unfortunately, none of the cited studies have studied the impact of noise (and its level) in the load time series on the adequacy of model identification and its predictive capacity. The authors have encountered only a few studies indirectly related to this problem [7,8,73,74]. In paper [7] a pattern recognition technique was used to examine the influence of noise on the one-step ahead time-series forecasting in the case of the exponential smoothing with non-linear neural networks methods. Results of studying different forecasting techniques (nearest neighbors, artificial neural networks, ARIMA, fuzzy neural networks, and nearest neighbors combined with differential evolution) from the perspective of their susceptibility to random fluctuation were presented in work [8]. Paper [73] provides the analysis of the possibility to reconstruct the attractor of a noise affected time series using a hybrid approach of nonparametric regression and optimal transformations. Two algorithms that estimate the noise level in a time series are exhibited in article [74]. In work [75], a data-filtering method for short-term load forecasting was proposed. It was demonstrated that statistical data-prefiltering improved the efficiency of STLF forecasting in the case of ARIMA models as well as the artificial neural networks.

In this paper the authors embrace the unexplored problematic of the robustness of ARIMA model identification in the case of time series affected by white noise of different noise to signal ratios (NSR) [8].

4. Research Methodology and Experiment Design

A dedicated research process was designed to study the tolerance to noise of ARIMA models in electrical load forecasting. Its logic and main stages are presented in Figure 2.

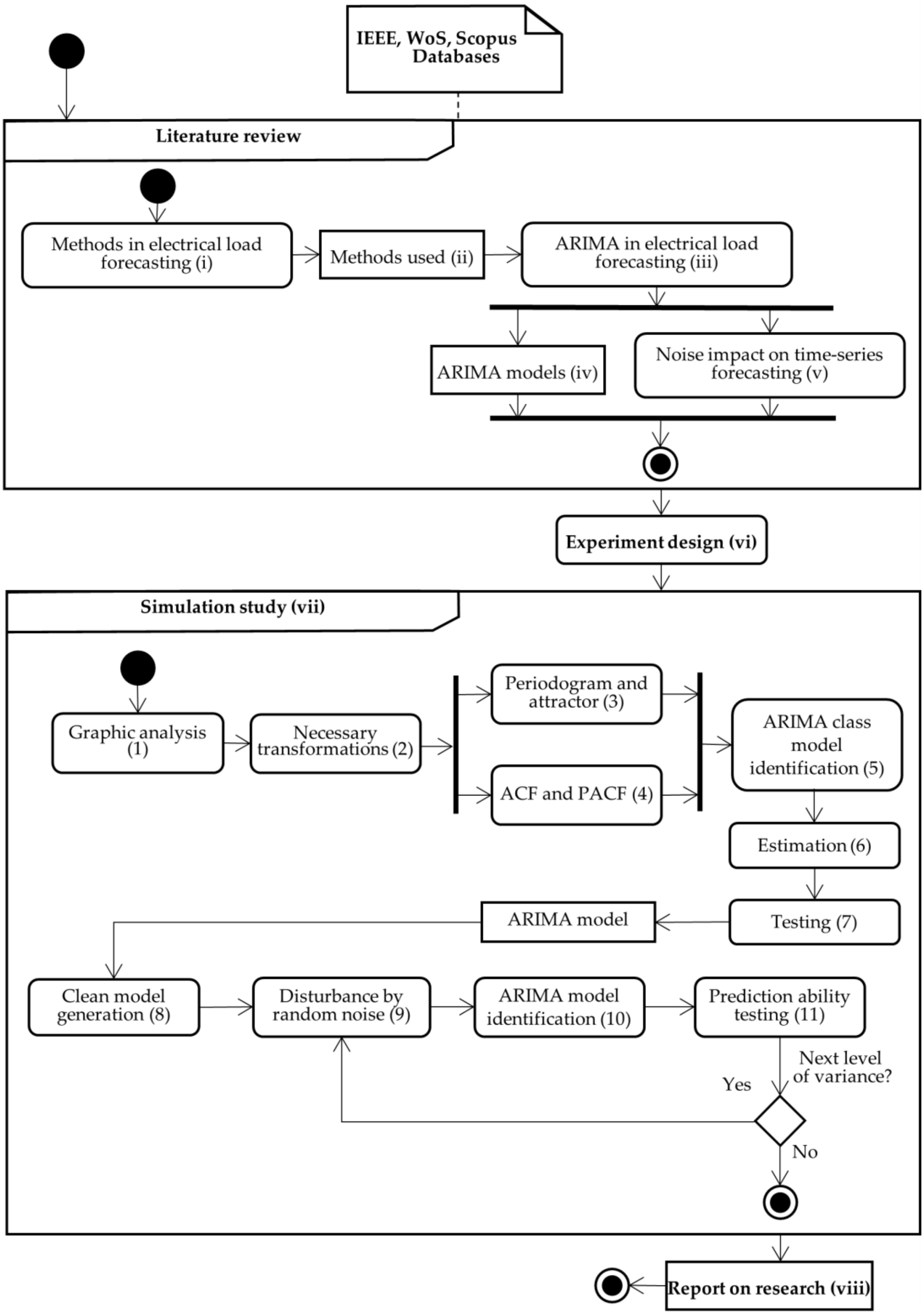


Figure 2. General research methodology.

The designed research process consists of the following stages: (i) review of the scientific literature related to the methods in electrical load forecasting, which resulted in (ii) the identification of methods used in electrical load forecasting; (iii) review of the scientific literature related to the applications of ARIMA method in load forecasting, which resulted in (iv) the specification of ARIMA models employed in electrical load forecasting; (v) review of the scientific literature related to the noise impact on time-series forecasting.

The literature review fed into the (vi) experiment design, which was the basis of the conducted (vii) simulation study that concluded with (viii) the final research report.

The simulation experiments play a key role in the study. The flow diagram of the experiment process is also presented in Figure 2. The starting point is (1) the graphic analysis of the electric load time series obtained from measurement data. It allows (2) to assess the time series with a view on the occurrence of trend and seasonal components, and consequently to decide on the needed transformations. The length of the seasonal periods is determined by (3) analyzing the periodogram and the time series attractors. The next step is (4) the determination of the ACF and PACF functions of the time series. This allows us to (5) specify the ARIMA model class and (6) estimate its parameters. The next step is (7) the analysis of model fit with the MSE and AIC criteria. The ARIMA model constructed in this process is used to (8) generate a clean time series model, which is treated as a reference in further study. In the following steps, (9) the reference time series is additively disturbed with noise of different levels measured by the ratio of the standard deviations of the signal and noise (NSR—noise to signal ratio). For each noise level, the identification of the ARIMA model of the disturbed time series is performed together with the assessment of its predictive capacity by setting the 95% confidence interval.

The designed experiment allows us to evaluate the stability of an identified ARIMA model class and to assess the changes in the model's predictive capacity in relation to the occurrence of different levels of white nose in the time series.

5. Simulation Results

5.1. ARIMA Model Identification

Energy load is a stochastic data series with values that depend on many factors: type of receivers; atmospheric conditions; time of the day, month, and year; sports and cultural events; and many other random events affecting the operation of receivers. In this paper, an hourly load time series registered in the Polish Power System (PPS) between 6 July 2020 and 27 September 2020 (12 weeks—2016 observations) is considered. The data were collected from Polish Power System Operation—Load of Polish Power System (https://www.pse.pl/ (accessed on 1 June 2021)). Basic characteristics of the data used in this study are presented in Table 3.

Table 3. Basic characteristics of data used in the study.

Characteristic	Value
Period of load data collection	6 July 2020—27 September 2020 (12 weeks)
Number of observations	2016
Mean load	18,134.96 MW
Minimum load value	11,903.98 MW
Maximum load value	23,222.25 MW
Load standard deviation	2810.35 MW

The plot of the studied time series is presented in Figure 3.

Two seasonal components (daily and weekly; lower load on weekends) as well as a slight linear trend are clearly visible in the time series. A periodogram was used to illustrate the harmonic structure of the data in more detail [76].

Two dominant periods, 24 h (1 day) and 168 h (1 week), are clearly visible in the periodogram (Figure 4), which indicates the daily and weekly seasonality of the load time series. This is quite a typical pattern in European countries [77,78].

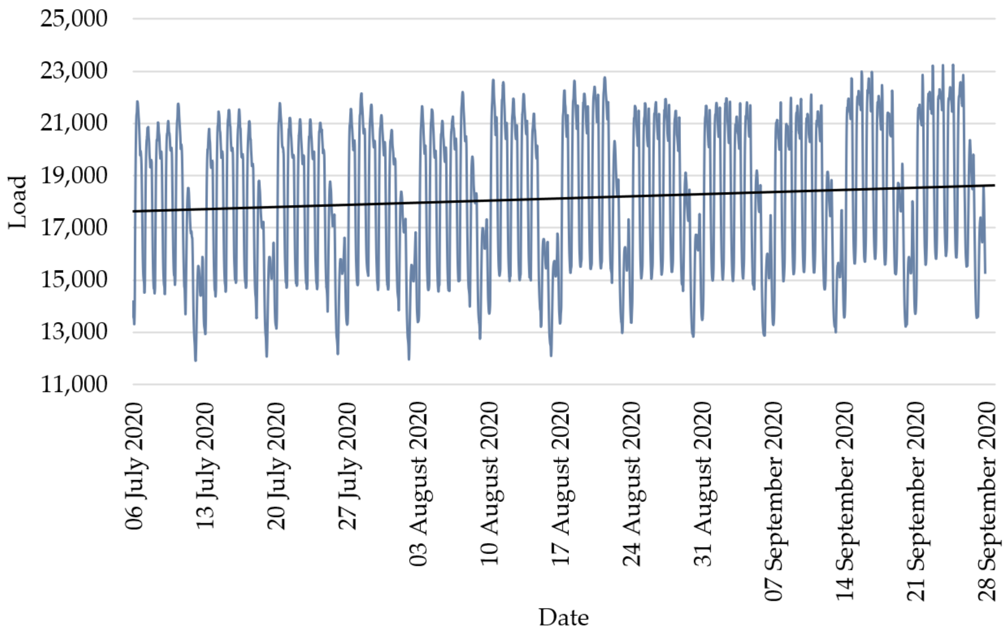


Figure 3. Hourly load data of the Polish Power System from 6 July 2020 to 27 September 2020 as a time series plot.

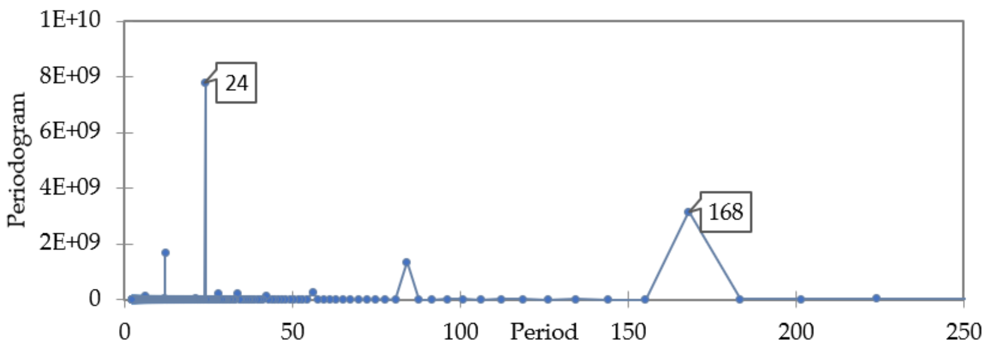


Figure 4. Periodogram of the load data time series.

Reconstructions of the studied load time series in two-dimensional phase-spaces (Y_t, Y_{t-1}), (Y_t, Y_{t-24}), and (Y_t, Y_{t-168}) are presented in Figure 5a–c, respectively. It may be noticed that the attractor is quite easy to distinguish in both cases. This implies good forecastability of the time series [79].

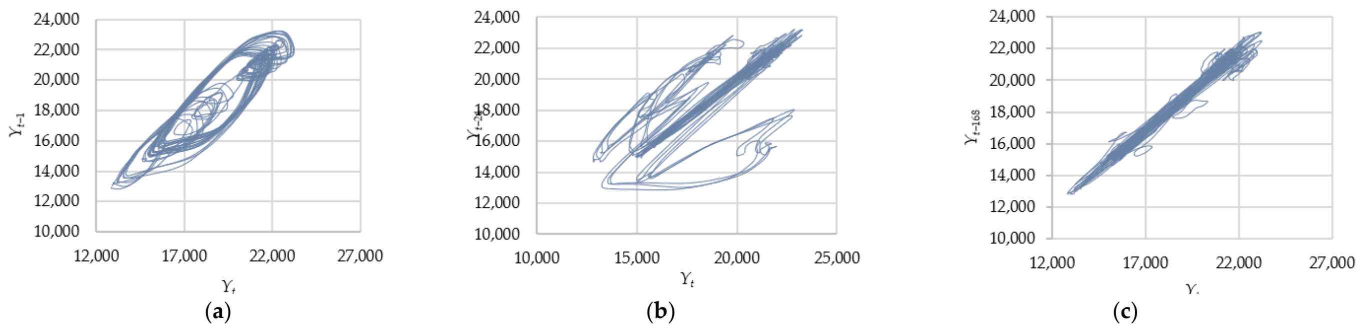


Figure 5. Attractors of the time series of load data in two-dimensional phase-spaces: (**a**) (Y_t, Y_{t-1}), (**b**) (Y_t, Y_{t-24}), and (**c**) (Y_t, Y_{t-168}).

Time series differencing is a standard procedure to remove the nonstationary components (trend and seasonality) from data. Trend is removed by single differencing (linear trend) or multiple differencing (equal to the degree of the polynomial describing the trend) with lag 1. Seasonal components are eliminated through seasonal differencing with the lag corresponding to the number of observations in the seasonal cycle [9]. In the case of the analyzed load time series, differencing with lag 1, 24, and 168 was carried out. ACF and PACF function plots for the differenced time series (d = 1, D_{24} = 1, D_{168} = 1) are presented in Figure 6.

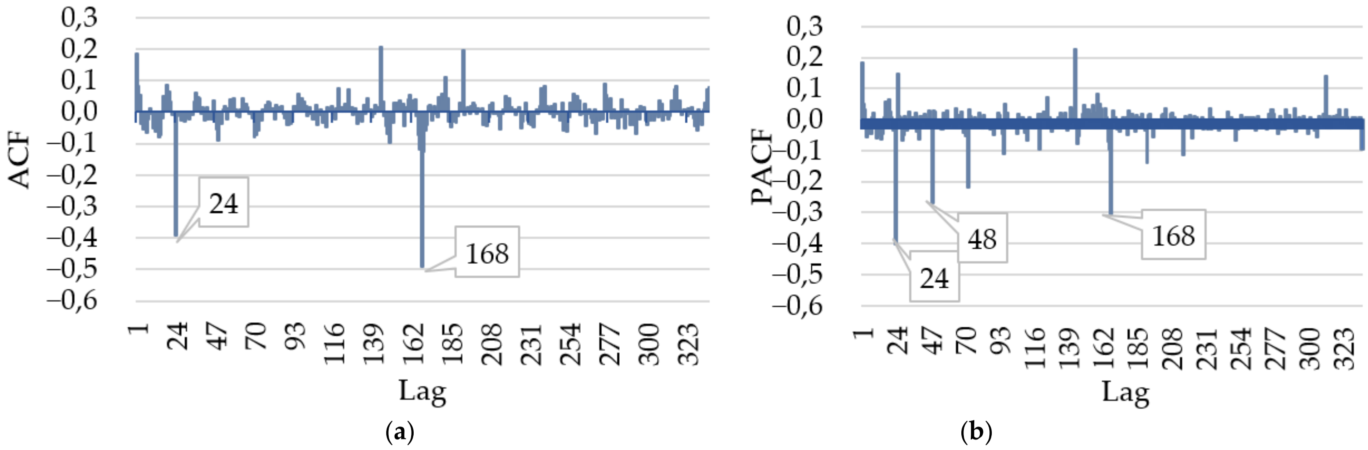


Figure 6. ACF and PACF functions of differentiated time series $(1 - B)(1 - B^{24})(1 - B^{168})Y_t$. (**a**) ACF, (**b**) PACF.

ACF function plot (significant values for delays 1, 24, and 168) and PACF function plot (combination of exponential decays starting from delays 1, 24, and 168) indicate the ARIMA(0,1,1)(0,1,1)$_{24}$(0,1,1)$_{168}$ model.

The ARIMA(0,1,1)(0,1,1)$_{24}$(0,1,1)$_{168}$ model is stationary and reversible. It may be expressed in the form of the backward shift notation as in Equation (7):

$$(1 - B)\left(1 - B^{24}\right)\left(1 - B^{168}\right)Y_t = (1 - \theta_1 B)\left(1 - \Theta_1 B^{24}\right)\left(1 - \vartheta_1 B^{168}\right)e_t \quad (7)$$

For the purpose of modelling and forecasting, Equation (7) may be transformed into Equation (8):

$$\begin{aligned} Y_t = Y_{t-1} + Y_{t-24} - Y_{t-25} + Y_{t-168} - Y_{t-169} - Y_{t-192} + Y_{t-193} - \theta_1 e_{t-1} - \Theta_1 e_{t-24} + \theta_1\Theta_1 e_{t-25} - \vartheta_1 e_{t-168} \\ +\theta_1\vartheta_1 e_{t-169} + \Theta_1\vartheta_1 e_{t-192} - \theta_1\Theta_1\vartheta_1 e_{t-193} + e_t \end{aligned} \quad (8)$$

Estimation of the model parameters θ_1, Θ_1, ϑ_1 was carried out with the maximum likelihood approach via nonlinear least squares using Marquardt's method, with computations performed in SAS Studio software. The following values were obtained:

$$\theta_1 = 0.20261, \text{ with std. error } 0.02220,$$

$$\Theta_1 = 0.72873, \text{ with std. error } 0.01864,$$

$$\vartheta_1 = 0.72605, \text{ with std. error } 0.02359.$$

The model described by Equation (9) was used to generate a reference (clean) load time series that was the basis for further simulations:

$$\begin{aligned} Y_t = Y_{t-1} + Y_{t-24} - Y_{t-25} + Y_{t-168} - Y_{t-169} - Y_{t-192} + Y_{t-193} - 0.20261 e_{t-1} - 0.72873 e_{t-24} \\ +0.20261 \times 0.72873 e_{t-25} - 0.72605 e_{t-168} + 0.20261 \times 0.72605 e_{t-169} \\ +0.72873 \times 0.72605 e_{t-192} - 0.20261 \times 0.72873 \times 0.72605 e_{t-193} + e_t \end{aligned} \quad (9)$$

The reference (clean) time series of hourly load values is presented in Figure 7.

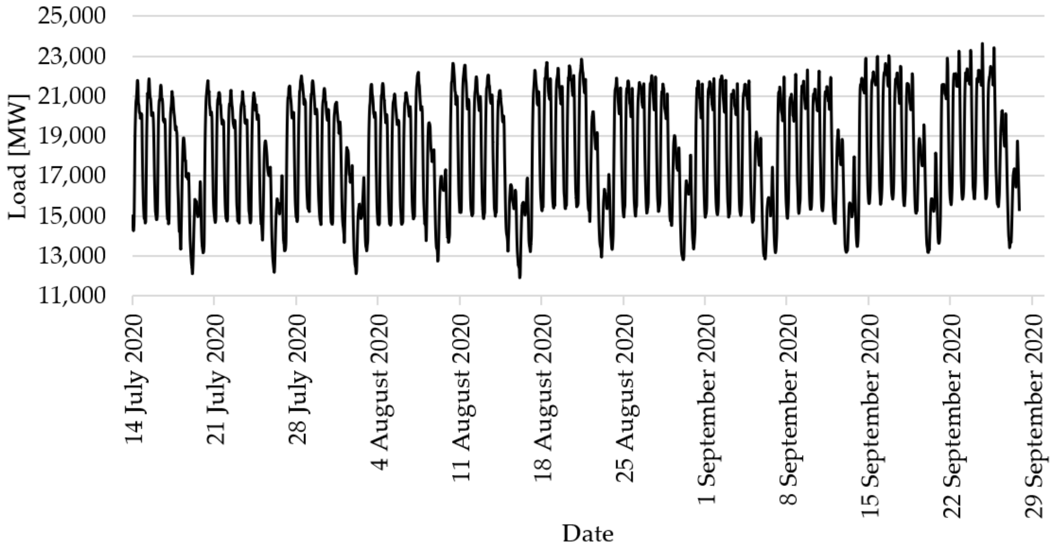


Figure 7. Reference (clean) load time series.

5.2. Simulation of the Impact of Random Noise

In the next step, the time series (signal) generated with use of Equation (9) was additively disturbed by white noise with zero mean ($\mu = 0$) and standard deviation equal to the product of NSR ratio multiplied by the signal standard deviation:

$$\sigma_{noise} = NSR\sigma_{signal} \tag{10}$$

where:

σ_{noise}—inference noise standard deviation,
σ_{signal}—signal standard deviation,
NSR—noise to signal ratio.

Signal standard deviation σ_{signal} was determined for the time series remainders (Equation (9)), i.e., after concluding the differencing operation ($d = 1$, $D_{24} = 1$, $D_{168} = 1$).

In Figure 8, the weekly load time series repeatedly disturbed with white noise with standard deviation (Equation (10)) determined for various NSR values from 10% to 500% is presented.

After the time series was additively disturbed, the ARIMA model parameters were re-estimated and the values of residual mean square (RMS) and the Akaike information criterion (AIC) were recalculated. Results of the calculations for NSR = 10%, 20%, 30%, 50%, 100%, and 200% are compiled in Table 4.

Simulation results indicate that, in the case of the analyzed time series, disturbances not exceeding NSR = 20% do not cause significant alterations in parameter estimation. The observed changes in parameters Θ_1 and ϑ_1 values do not exceed the 95% confidence interval of the reference model parameter estimation. Parameter θ_1 only slightly exceeds that interval. RMS and AIC values do not change significantly, either. Increasing the disturbance level above 30% causes more significant changes in the values of the estimated parameters and in the RMS and AIC values. Parameters were not estimated for noise levels higher than NSR = 200% because the ϑ_1 parameter value was above the irreversibility boundary of the model in that case.

The behavior of ACF and PACF functions is worth attention. Their plots for different noise levels are presented in Figure 9. As can be seen, the patterns still suggest MA seasonal process rather than AR. The obtained research results lead to the conclusion that the reaction of the load time series to random disturbances is relatively small. The functions of AC and PAC do not change significantly at all tested levels of disturbance, always indicating the original type of model. The changing values of estimated parameters indicate that the series is recognized as of the same type, but completely different due to the parameters' values.

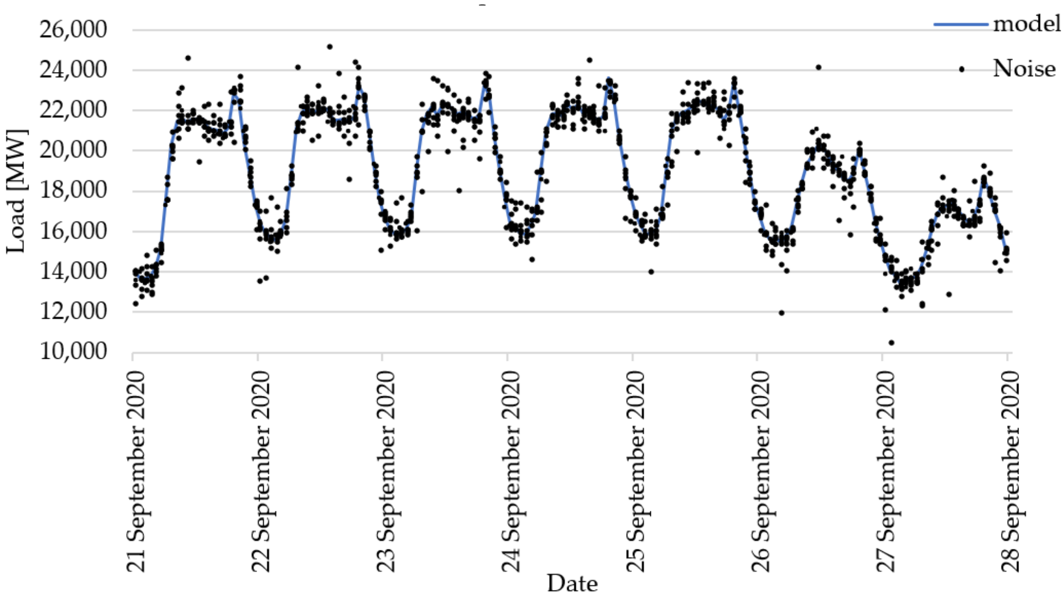
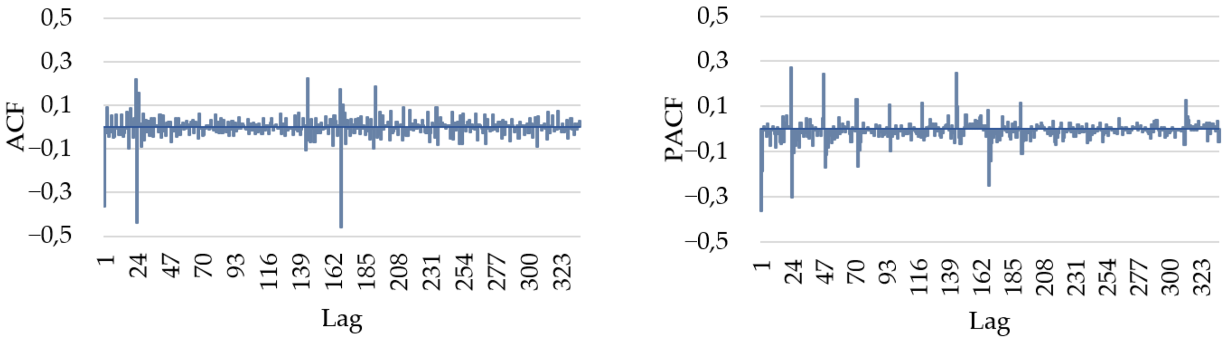


Figure 8. Noise-disturbed load time series.

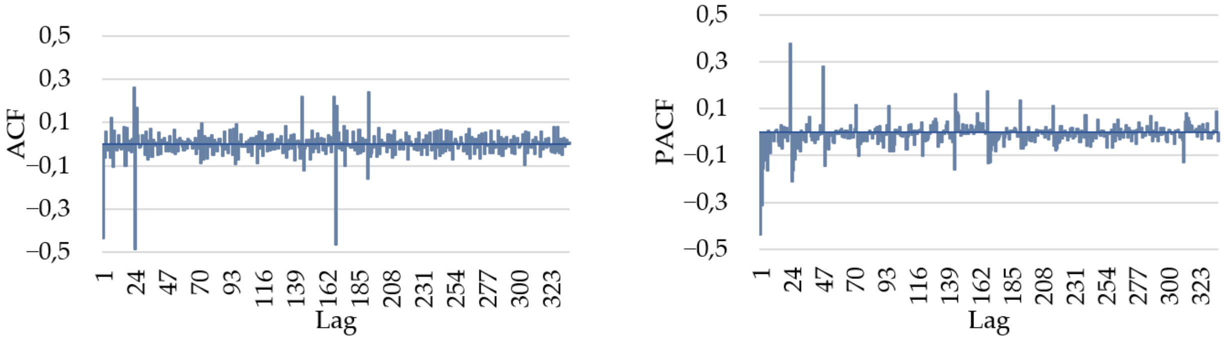
Table 4. Simulation experiment results.

Level of Disturbance (NSR)	Parameters	Std. Error	*p*-Value	95% Confidence Interval		RMS	AIC
Model σ_{signal} = 289.5 MW	θ_1 = 0.20261 Θ_1 = 0.72873 ϑ_1 = 0.72605	0.02382 0.01860 0.02588	<0.0001 <0.0001 <0.0001	0.15592 0.69227 0.67533	0.24930 0.76519 0.77677	179.9524	21,704.73
NSR = 10% Model + N(0; 29)	θ_1 =0.21643 Θ_1 = 0.72698 ϑ_1 = 0.74025	0.02368 0.01863 0.02647	<0.0001 <0.0001 <0.0001	0.17002 0.69047 0.68837	0.26284 0.76349 0.79213	184.8383	21,799.42
NSR = 20% Model + N(0; 59)	θ_1 = 0.28279 Θ_1 = 0.74010 ϑ_1 = 0.76365	0.02307 0.01829 0.02692	<0.0001 <0.0001 <0.0001	0.23757 0.70425 0.71089	0.32801 0.77595 0.81641	200.8707	22,085.67
NSR = 30% Model + N(0; 87)	θ_1 = 0.35262 Θ_1 = 0.76606 ϑ_1 = 0.77243	0.02239 0.01763 0.02739	<0.0001 <0.0001 <0.0001	0.30873 0.73151 0.71875	0.39650 0.80061 0.82611	221.5301	22,413.83
NSR = 50% Model + N(0; 145)	θ_1 = 0.46077 Θ_1 = 0.78613 ϑ_1 = 0.77771	0.02104 0.01738 0.02718	<0.0001 <0.0001 <0.0001	0.41953 0.75207 0.72444	0.50201 0.82019 0.83098	270.3925	23,070.28
NSR = 100% Model + N(0; 290)	θ_1 = 0.68715 Θ_1 = 0.83840 ϑ_1 = 0.94662	0.01615 0.01525 0.09050	<0.0001 <0.0001 <0.0001	0.65550 0.80851 0.76924	0.71880 0.86829 1.12400	395.6171	24,487.19
NSR = 200% Model + N(0; 580)	θ_1 = 0.84442 Θ_1 = 0.89227 ϑ_1 = 0.99989	0.01165 0.01441 45.70017	<0.0001 <0.0001 0.9825	0.82159 0.86403	0.86725 0.92051	661.1075	26,260.64

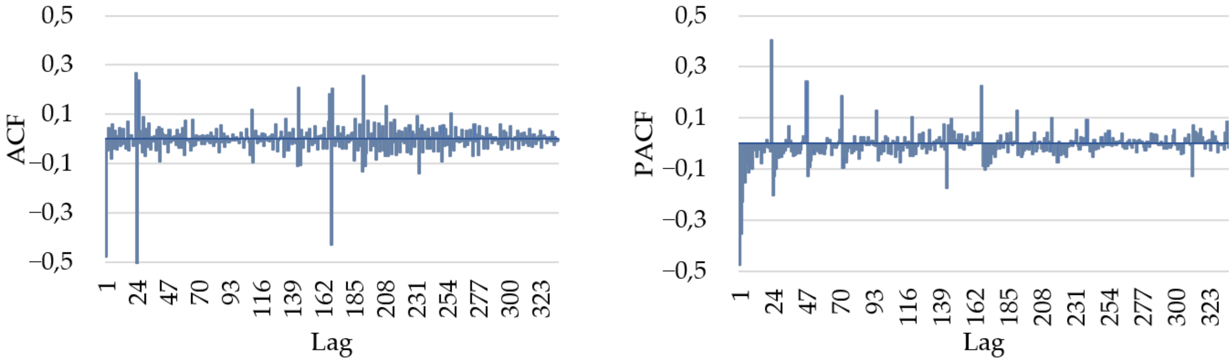
In the next step, the models developed for the reference model and the disturbed time series were used to calculate forecasts. Multi-step forecast 6 h ahead was prepared for each model. Obtained forecasts with the 95% confidence interval are compiled in Table 5 and illustrated in Figure 10.



(**a**)



(**b**)



(**c**)

Figure 9. ACF and PACF of differentiated noise-disturbed time series: (**a**) NSR = 30%, (**b**) NSR = 100%, (**c**) NSR = 200%.

Increasing the noise level enlarges the forecast confidence interval (lowers the accuracy), but obtained results are quite surprising. Forecasts for all noise levels are fairly consistent. Practically all forecasts made on the basis of the models derived from the disturbed time series fit within the 95% confidence interval of the forecast made on the basis of the reference model. It may be assumed that up to NSR = 30%, the model and its estimation is not very noise sensitive. Increasing the noise beyond this level significantly increases the width of the forecast confidence interval. Only increasing the noise level to NSR = 200% makes the model irreversible. This level of disturbance changes the possibilities of discovering the patterns of the energy load time series.

Table 5. ARIMA forecasts for the periods of 6 h (t = 2017, 2018, 2019, 2020, 2021, 2020) for 28 September 2020.

Level of Disturbance (NSR)	Hour of 28 September 2020	Forecast	95% Confidence Interval	
			From	To
NSR = 0 Model	1	14,569.2799	14,216.5795	14,921.9802
	2	14,231.1726	13,780.0701	14,682.2751
	3	14,074.4136	13,542.8222	14,606.0040
	4	14,116.8656	13,515.4628	14,718.2684
	5	14,552.4516	13,888.5380	15,216.3652
	6	15,601.6356	14,880.6105	16,322.6607
NSR = 10% Model + $N(0; 29)$	1	14,551.6886	14,189.4121	14,913.9651
	2	14,202.3811	13,742.1356	14,662.6266
	3	14,056.7153	13,515.9681	14,597.4626
	4	14,099.6095	13,488.8810	14,710.3380
	5	14,544.8341	13,871.3574	15,218.3109
	6	15,570.5080	14,839.6505	16,301.3655
NSR = 20% Model + $N(0; 58)$	1	14,647.8067	14,254.1075	15,041.5060
	2	14,239.7803	13,755.2924	14,724.2682
	3	14,078.4536	13,517.6881	14,639.2190
	4	14,183.4656	13,555.6223	14,811.3089
	5	14,556.4968	13,868.0807	15,244.9128
	6	15,640.8009	14,896.7270	16,384.8749
NSR = 30% Model + $N(0;87)$	1	14,589.3005	14,155.1095	15,023.4915
	2	14,265.3775	13,748.1437	14,782.6113
	3	14,156.0002	13,567.3240	14,744.6765
	4	14,189.9685	13,537.6275	14,842.3095
	5	14,627.1028	13,916.7806	15,337.4251
	6	15,659.0319	14,895.1165	16,422.9473
NSR = 50% Model + $N(0; 145)$	1	14,334.9439	13,804.9843	14,864.9036
	2	13,984.5530	13,382.4542	14,586.6518
	3	13,749.7201	13,083.2452	14,416.1949
	4	13,784.8641	13,059.7058	14,510.0223
	5	14,203.5181	13,424.0823	14,982.9539
	6	15,435.2713	14,605.0991	16,265.4435
NSR = 100% Model + $N(0; 290)$	1	14,146.3251	13,370.9299	14,921.7203
	2	14,264.9867	13,452.5307	15,077.4426
	3	14,028.2346	13,180.3362	14,876.1329
	4	14,113.0099	13,231.0924	14,994.9275
	5	14,410.1252	13,495.4529	15,324.7976
	6	15,543.8580	14,597.5639	16,490.1521
NSR = 200% Model + $N(0; 580)$	1	14,379.1325	13,083.3855	15,674.8794
	2	14,358.8916	13,047.5563	15,670.2269
	3	14,624.8744	13,298.1340	15,951.6149
	4	14,357.8940	13,015.9252	15,699.8628
	5	14,853.3425	13,496.3162	16,210.3688
	6	15,746.6589	14,374.7404	17,118.5774

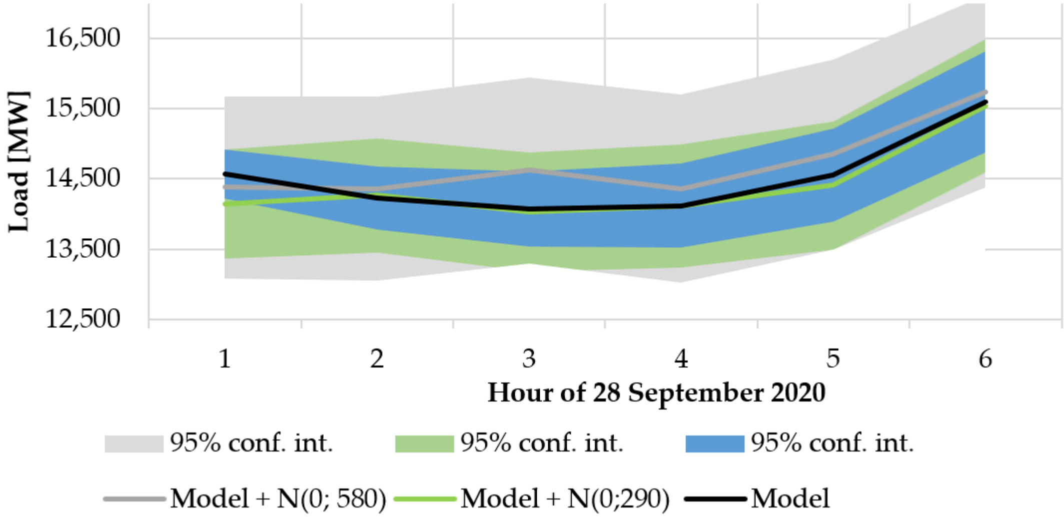


Figure 10. Model and disturbed time series forecasts with confidence interval.

6. Discussion

The main problem of time series modelling with ARIMA models is specifying the class (autoregressive and moving average) and the order of non-seasonal differencing, and the number of seasons and the order of seasonal differencing. Identification of ARIMA models is not strictly codified, and it depends to a large extent on the empirical knowledge and the intuition of a researcher and the quality of fit of the tested models. The basis for the identification is the analysis of ACF and PACF plots. In many cases, a given time series may be described by different ARIMA models. The autoregressive and moving average components may cancel each other's effect. There is also a relationship between the degree of differencing and the order of autoregression and moving average. The over-differencing of the series can be compensated for by considering the additional term of autoregression in the model, the under-differencing by the additional term of the moving average. The possibility of the correct identification and estimation always depends on the presence and variance of random noise. For this reason, it is important to define the disturbance level, which determines the possibility of applying specified models.

The designed and executed simulation experiment allowed us to evaluate the robustness of ARIMA models to noise in their ability to predict electrical load time series. This research activity follows an established research practice that consists of simulating various aspects of power system performance under changing noise intensity [80]. In the study, an idealized ARIMA model of electrical loads was disturbed by noise of different levels. The model parameters were then re-estimated and new forecasts were calculated. The experiment has provided many interesting observations. It may be concluded that the reaction of the ARIMA model to random disturbances of the modeled time series is relatively weak. ACF and PACF functions do not change significantly at all tested levels of disturbance, generally indicating the original type of model. However, changing values of the estimated parameters indicate that the series is recognized as of the same type, but with different parameter values. The correctness of the estimation stage of a given type of ARIMA model depends to a large extent on the level of random disturbances present in the series. The presence of disturbance over 30%, and strongly over 100% of standard deviation significantly influences the RMSE, AIC, and the width of the forecast confidence interval.

ARIMA models are frequently used in load forecasting. They are flexible and well interpretable. Obtained results constitute a valuable advice regarding the mode of conduct in practical applications of ARIMA in load modeling and forecasting. They reaffirm the key importance of data preprocessing stage in the ARIMA model implementation. It is also recommended to carry out a preliminary time series evaluation with regard to the noise presence and the possible noise filtering before the ARIMA model identification and estimation. The authors consider it reasonable to introduce two additional phases to the standard ARIMA model development process: noise level identification and signal filtering. Thus, the process of ARIMA modelling would cover eight consecutive phases: (i) preliminary analysis, (ii) noise level identification, (iii) signal filtering, (iv) transformation, (v) identification, (vi) estimation, (vii) testing, and (viii) application.

Too high of a noise-to-signal ratio may be a premise for the choice of other forecasting methods based on, e.g., machine learning or other artificial intelligence methods.

Certain limitations of the presented results must be also acknowledged. First of all, only one load time series describing the whole power system was analyzed. Consequently, such a time series was characterized by a large share of systematic components with well specified features and parameters. Second, simulations were carried out only for a single class of ARIMA model. Third, the considerations were limited to STLF forecasting. Identified limitations point at the possible directions of further research. They should concern the load of various elements (fragments) of the power system at different hierarchy levels. Different ARIMA model classes should be considered. Calculations of forecasts with various time horizons would also be valuable. It would be desirable to compare the results obtained in this study to other simulations based on data from different time periods, different forecast horizons, different power systems (and their sections), and different ARIMA model classes. In this paper, authors focus on the electric load processes, but the proposed methodology may as well be applied to study time series presenting observable data of other origins.

7. Conclusions

The obtained simulation results presented in this paper lead to the following conclusions:

- Noise loading of the signal significantly affects the identification of the time series ARIMA model type and the estimation of its parameters,
- The accuracy of the prediction of electrical loads strongly depends on the noise level in the observed signal,
- The observed time series of the electrical load should be carefully examined for the presence and the level of noise in the signal before the prediction is performed,
- Usefulness of extending the classic Box–Jenkins approach by the preliminary time series filtration is proven.

Despite the identified limitations, which partly result from the size constraints of this paper, it is justified to claim that the presented research contributes to the theory and practice of electric load forecasting, allowing for the preparation of more precise forecasts. Effectively, better forecasting decreases uncertainty and leads to better informed decisions at different hierarchical management levels of the power system, thus making the energy policy more robust to uncertainty, better aligned with the Goal 7 of the Sustainable Development Goals [81], and more environmentally viable.

Author Contributions: J.N. and E.C. were responsible for the study conception and research design; J.N. and Ł.N. developed the concept; E.C. performed computation and analysis; J.N. and E.C. were responsible for data interpretation; J.N. and Ł.N. discussed the results and contributed to the final manuscript; J.N., E.C., and Ł.N. wrote and edited the text. All authors have read and agreed to the published version of the manuscript.

Funding: The publication of the article for 11th International Conference on Engineering, Project, and Production Management - EPPM2021 was financed in the framework of the contract no. DNK/SN/465770/2020 by the Ministry of Science and Higher Education within the ”Excellent Science” programme.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: Publicly available datasets were analyzed in this study. This data can be found here: https://www.pse.pl/web/pse-eng/areas-of-activity/polish-power-system/system-load (accessed on 17 November 2021).

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Soliman, S.A.; Alkandari, A.M. *Electrical Load Forecasting: Modeling and Model Construction*; Butterworth-Heinemann: Burlington, MA, USA, 2010; ISBN 978-0-12-381543-9.
2. Nti, I.K.; Teimeh, M.; Nyarko-Boateng, O.; Adekoya, A.F. Electricity Load Forecasting: A Systematic Review. *J. Electr. Syst. Inf Technol* **2020**, *7*, 13. [CrossRef]
3. Hammad, M.A.; Jereb, B.; Rosi, B.; Dragan, D. Methods and Models for Electric Load Forecasting: A Comprehensive Review. *Logist. Sustain. Transp.* **2020**, *11*, 51–76. [CrossRef]
4. Hong, T.; Fan, S. Probabilistic Electric Load Forecasting: A Tutorial Review. *Int. J. Forecast.* **2016**, *32*, 914–938. [CrossRef]
5. Hahn, H.; Meyer-Nieberg, S.; Pickl, S. Electric Load Forecasting Methods: Tools for Decision Making. *Eur. J. Oper. Res.* **2009**, *199*, 902–907. [CrossRef]
6. Kuster, C.; Rezgui, Y.; Mourshed, M. Electrical Load Forecasting Models: A Critical Systematic Review. *Sustain. Cities Soc.* **2017**, *35*, 257–270. [CrossRef]
7. Singh, S. Noise Impact on Time-Series Forecasting Using an Intelligent Pattern Matching Technique. *Pattern Recognit.* **1999**, *32*, 1389–1398. [CrossRef]
8. Flores, J.J.; Calderon, F.; Cedeno Gonzalez, J.R.; Ortiz, J.; Farias, R.L. Comparison of Time Series Forecasting Techniques with Respect to Tolerance to Noise. In Proceedings of the 2016 IEEE International Autumn Meeting on Power, Electronics and Computing (ROPEC), Ixtapa, Mexico, 9–11 November 2016; IEEE: Piscataway, NJ, USA, 2016; pp. 1–6.
9. Box, G.E.P.; Jenkins, G.M. *Time Series Analysis: Forecasting and Control*; Holden-Day series in time series analysis and digital processing; Rev. ed.; Holden-Day: San Francisco, CA, USA, 1976; ISBN 978-0-8162-1104-3.
10. Rosadi, D. New Procedure for Determining Order of Subset Autoregressive Integrated Moving Average (ARIMA) Based on over-Fitting Concept. In Proceedings of the 2012 International Conference on Statistics in Science, Business and Engineering (ICSSBE), Langkawi, Kedah, Malaysia, 10–12 September 2012; IEEE: Piscataway, NJ, USA, 2012; pp. 1–5.
11. Awan, T.M.; Aslam, F. Prediction of Daily COVID-19 Cases in European Countries Using Automatic ARIMA Model. *J. Public Health Res.* **2020**, *9*, 1765. [CrossRef]
12. Villegas, M.A.; Pedregal, D.J. Automatic Selection of Unobserved Components Models for Supply Chain Forecasting. *Int. J. Forecast.* **2019**, *35*, 157–169. [CrossRef]
13. Awe, O.; Okeyinka, A.; Fatokun, J.O. An Alternative Algorithm for ARIMA Model Selection. In Proceedings of the 2020 International Conference in Mathematics, Computer Engineering and Computer Science (ICMCECS), Lagos, Nigeria, 18–21 March 2020; IEEE: Piscataway, NJ, USA, 2020; pp. 1–4.
14. Lénárt, B. Automatic Identification of ARIMA Models with Neural Network. *Per. Pol. Transp. Eng.* **2011**, *39*, 39. [CrossRef]
15. Włodarczyk, A. X-13-ARIMA-SEATS as a Tool Supporting Environmental Management Process in the Power Plants. *PJMS* **2017**, *16*, 280–291. [CrossRef]
16. Li, G.; Wang, Y. Automatic ARIMA Modeling-Based Data Aggregation Scheme in Wireless Sensor Networks. *J. Wirel. Com. Netw.* **2013**, *2013*, 85. [CrossRef]
17. Caro, E.; Juan, J.; Cara, J. Periodically Correlated Models for Short-Term Electricity Load Forecasting. *Appl. Math. Comput.* **2020**, *364*, 124642. [CrossRef]
18. Deb, C.; Zhang, F.; Yang, J.; Lee, S.E.; Shah, K.W. A Review on Time Series Forecasting Techniques for Building Energy Consumption. *Renew. Sustain. Energy Rev.* **2017**, *74*, 902–924. [CrossRef]
19. Chodakowska, E.; Nazarko, J. Assessing the Performance of Sustainable Development Goals of EU Countries: Hard and Soft Data Integration. *Energies* **2020**, *13*, 3439. [CrossRef]
20. Kao, Y.-S.; Nawata, K.; Huang, C.-Y. Predicting Primary Energy Consumption Using Hybrid ARIMA and GA-SVR Based on EEMD Decomposition. *Mathematics* **2020**, *8*, 1722. [CrossRef]
21. Son, H.; Kim, Y.; Kim, S. Time Series Clustering of Electricity Demand for Industrial Areas on Smart Grid. *Energies* **2020**, *13*, 2377. [CrossRef]
22. Trull, O.; García-Díaz, J.C.; Peiró-Signes, A. Forecasting Irregular Seasonal Power Consumption. An Application to a Hot-Dip Galvanizing Process. *Appl. Sci.* **2020**, *11*, 75. [CrossRef]
23. Parsai, S.; Mahajan, S. Anomaly Detection Using Long Short-Term Memory. In Proceedings of the 2020 International Conference on Electronics and Sustainable Communication Systems (ICESC), Coimbatore, India, 2–4 July 2020; IEEE: Piscataway, NJ, USA, 2020; pp. 333–337.
24. Lee, J.-W.; Kim, H.-J.; Kim, M.-K. Design of Short-Term Load Forecasting Based on ANN Using Bigdata. *KIEE* **2020**, *69*, 792–799. [CrossRef]

25. Sáenz-Peñafiel, J.-J.; Luzuriaga, J.E.; Lemus-Zuñiga, L.-G.; Solis-Cabrera, V. A Comprehensive Solution for Electrical Energy Demand Prediction Based on Auto-Regressive Models. In *Systems and Information Sciences*; Botto-Tobar, M., Zamora, W., Larrea Plúa, J., Bazurto Roldan, J., Santamaría Philco, A., Eds.; Advances in Intelligent Systems and Computing; Springer International Publishing: Cham, Switzerland, 2021; Volume 1273, pp. 443–454. ISBN 978-3-030-59193-9.
26. Nafil, A.; Bouzi, M.; Anoune, K.; Ettalabi, N. Comparative Study of Forecasting Methods for Energy Demand in Morocco. *Energy Rep.* **2020**, *6*, 523–536. [CrossRef]
27. Sun, T.; Zhang, T.; Teng, Y.; Chen, Z.; Fang, J. Monthly Electricity Consumption Forecasting Method Based on X12 and STL Decomposition Model in an Integrated Energy System. *Math. Probl. Eng.* **2019**, *2019*, 1–16. [CrossRef]
28. Han, X.; Li, R. Comparison of Forecasting Energy Consumption in East Africa Using the MGM, NMGM, MGM-ARIMA, and NMGM-ARIMA Model. *Energies* **2019**, *12*, 3278. [CrossRef]
29. Borthakur, P.; Goswami, B. Short Term Load Forecasting: A Hybrid Approach Using Data Mining Methods. In Proceedings of the 2020 International Conference on Emerging Frontiers in Electrical and Electronic Technologies (ICEFEET), Patna, India, 10–11 July 2020; IEEE: Piscataway, NJ, USA, 2020; pp. 1–6.
30. Bui, D.M.; Le, P.D.; Cao, M.T.; Pham, T.T.; Pham, D.A. Accuracy Improvement of Various Short-Term Load Forecasting Models by a Novel and Unified Statistical Data-Filtering Method. *Int. J. Green Energy* **2020**, *17*, 382–406. [CrossRef]
31. Jahanshahi, A.; Jahanianfard, D.; Mostafaie, A.; Kamali, M. An Auto Regressive Integrated Moving Average (ARIMA) Model for Prediction of Energy Consumption by Household Sector in Euro Area. *AIMS Energy* **2019**, *7*, 151–164. [CrossRef]
32. Noureen, S.; Atique, S.; Roy, V.; Bayne, S. Analysis and Application of Seasonal ARIMA Model in Energy Demand Forecasting: A Case Study of Small Scale Agricultural Load. In Proceedings of the 2019 IEEE 62nd International Midwest Symposium on Circuits and Systems (MWSCAS), Dallas, TX, USA, 4–7 August 2019; IEEE: Piscataway, NJ, USA, 2019; pp. 521–524.
33. Salam, M.A.; Yazdani, M.G.; Wen, F.; Rahman, Q.M.; Malik, O.A.; Hasan, S. Modeling and Forecasting of Energy Demands for Household Applications. *Glob. Chall.* **2020**, *4*, 1900065. [CrossRef] [PubMed]
34. Luan, C.; Pang, X.; Wang, Y.; Liu, L.; You, S. Comprehensive Forecasting Method of Monthly Electricity Consumption Based on Time Series Decomposition and Regression Analysis. In Proceedings of the 2020 2nd International Conference on Industrial Artificial Intelligence (IAI), Shenyang, China, 24–26 July 2020; IEEE: Piscataway, NJ, USA, 2020; pp. 1–6.
35. 3Mehedintu, A.; Sterpu, M.; Soava, G. Estimation and Forecasts for the Share of Renewable Energy Consumption in Final Energy Consumption by 2020 in the European Union. *Sustainability* **2018**, *10*, 1515. [CrossRef]
36. Méndez-Gordillo, A.R.; Cadenas, E. Wind Speed Forecasting by the Extraction of the Multifractal Patterns of Time Series through the Multiplicative Cascade Technique. *Chaos Solitons Fractals* **2021**, *143*, 110592. [CrossRef]
37. Marulanda, G.; Bello, A.; Cifuentes, J.; Reneses, J. Wind Power Long-Term Scenario Generation Considering Spatial-Temporal Dependencies in Coupled Electricity Markets. *Energies* **2020**, *13*, 3427. [CrossRef]
38. Bórawski, P.; Bełdycka-Bórawska, A.; Jankowski, K.J.; Dubis, B.; Dunn, J.W. Development of Wind Energy Market in the European Union. *Renew. Energy* **2020**, *161*, 691–700. [CrossRef]
39. Lei, Y.; Xue, P.; Li, Y. Comparison of Holt-Winters and ARIMA Models for Hydropower Forecasting in Guangxi. In Proceedings of the Proceedings of the 2020 3rd International Conference on Signal Processing and Machine Learning, Beijing, China, 22–24 October 2020; ACM: New York, NY, USA, 2020; pp. 63–67.
40. Jamil, R. Hydroelectricity Consumption Forecast for Pakistan Using ARIMA Modeling and Supply-Demand Analysis for the Year 2030. *Renew. Energy* **2020**, *154*, 1–10. [CrossRef]
41. Barzola-Monteses, J.; Mite-León, M.; Espinoza-Andaluz, M.; Gómez-Romero, J.; Fajardo, W. Time Series Analysis for Predicting Hydroelectric Power Production: The Ecuador Case. *Sustainability* **2019**, *11*, 6539. [CrossRef]
42. Belmahdi, B.; Louzazni, M.; Bouardi, A.E. One Month-Ahead Forecasting of Mean Daily Global Solar Radiation Using Time Series Models. *Optik* **2020**, *219*, 165207. [CrossRef]
43. 4Lu, Y.-S.; Lai, K.-Y. Deep-Learning-Based Power Generation Forecasting of Thermal Energy Conversion. *Entropy* **2020**, *22*, 1161. [CrossRef]
44. Dittmer, C.; Krümpel, J.; Lemmer, A. Power Demand Forecasting for Demand-Driven Energy Production with Biogas Plants. *Renew. Energy* **2021**, *163*, 1871–1877. [CrossRef]
45. Leerbeck, K.; Bacher, P.; Junker, R.G.; Goranović, G.; Corradi, O.; Ebrahimy, R.; Tveit, A.; Madsen, H. Short-Term Forecasting of CO2 Emission Intensity in Power Grids by Machine Learning. *Appl. Energy* **2020**, *277*, 115527. [CrossRef]
46. Wang, Q.; Li, S.; Pisarenko, Z. Modeling Carbon Emission Trajectory of China, US and India. *J. Clean. Prod.* **2020**, *258*, 120723. [CrossRef]
47. Malik, A.; Hussain, E.; Baig, S.; Khokhar, M.F. Forecasting CO_2 Emissions from Energy Consumption in Pakistan under Different Scenarios: The China–Pakistan Economic Corridor. *Greenh. Gas Sci. Technol.* **2020**, *10*, 380–389. [CrossRef]
48. Peshkov, A.V.; Alsova, O.K. Short-Term Forecasting of the Time Series of Electricity Prices with Ensemble Algorithms. *J. Phys. Conf. Ser.* **2020**, *1661*, 012070. [CrossRef]
49. Nametala, C.; de Faria, W.; Junior, B. A Volatility Index for Bilateral Electricity Contracting Auctions. *IEEE Lat. Am. Trans.* **2020**, *18*, 938–946. [CrossRef]
50. Nop, P.; Qin, Z. Cambodia Mid-Term Transmission System Load Forecasting with the Combination of Seasonal ARIMA and Gaussian Process Regression. In Proceedings of the 2021 3rd Asia Energy and Electrical Engineering Symposium (AEEES), Chengdu, China, 26–29 March 2021; IEEE: Piscataway, NJ, USA, 2021; pp. 700–707.

51. Mpawenimana, I.; Pegatoquet, A.; Roy, V.; Rodriguez, L.; Belleudy, C. A Comparative Study of LSTM and ARIMA for Energy Load Prediction with Enhanced Data Preprocessing. In Proceedings of the 2020 IEEE Sensors Applications Symposium (SAS), Kuala Lumpur, Malaysia, 28–31 March 2020; IEEE: Piscataway, NJ, USA, 2020; pp. 1–6.
52. Goswami, K.; Kandali, A.B. Electricity Demand Prediction Using Data Driven Forecasting Scheme: ARIMA and SARIMA for Real-Time Load Data of Assam. In Proceedings of the 2020 International Conference on Computational Performance Evaluation (ComPE), Shillong, India, 2–4 July 2020; IEEE: Piscataway, NJ, USA, 2020; pp. 570–574.
53. Gupta, A.; Kumar, A. Mid Term Daily Load Forecasting Using ARIMA, Wavelet-ARIMA and Machine Learning. In Proceedings of the 2020 IEEE International Conference on Environment and Electrical Engineering and 2020 IEEE Industrial and Commercial Power Systems Europe (EEEIC/I&CPS Europe), Madrid, Spain, 9–12 June 2020; IEEE: Piscataway, NJ, USA, 2020; pp. 1–5.
54. Wang, Y.; Zhang, W.; Li, X.; Wang, A.; Wu, T.; Bao, Y. Monthly Load Forecasting Based on an ARIMA-BP Model: A Case Study on Shaoxing City. In Proceedings of the 2020 12th IEEE PES Asia-Pacific Power and Energy Engineering Conference (APPEEC), Nanjing, China, 14–16 September 2020; IEEE: Piscataway, NJ, USA, 2020; pp. 1–6.
55. Dinata, S.A.W.; Azka, M.; Faisal, M.; Suhartono; Yendra, R.; Gamal, M.D.H. Short-Term Load Forecasting Double Seasonal ARIMA Methods: An Evaluation Based on Mahakam-East Kalimantan Data. In Proceedings of the AIP Conference, Pontianak, Indonesia, 23–25 September 2020; p. 020004. [CrossRef]
56. Yang, L.; Yang, H. A Combined ARIMA-PPR Model for Short-Term Load Forecasting. In Proceedings of the 2019 IEEE Innovative Smart Grid Technologies—Asia (ISGT) Asia, Chengdu, China, 21–24 May 2019; IEEE: Piscataway, NJ, USA, 2019; pp. 3363–3367.
57. Yu, K.W.; Hsu, C.H.; Yang, S.M. A Model Integrating ARIMA and ANN with Seasonal and Periodic Characteristics for Forecasting Electricity Load Dynamics in a State. In Proceedings of the 2019 IEEE 6th International Conference on Energy Smart Systems (ESS), Kyiv, Ukraine, 17–19 April 2019; IEEE: Piscataway, NJ, USA, 2019; pp. 19–24.
58. Tang, L.; Yi, Y.; Peng, Y. An Ensemble Deep Learning Model for Short-Term Load Forecasting Based on ARIMA and LSTM. In Proceedings of the 2019 IEEE International Conference on Communications, Control, and Computing Technologies for Smart Grids (SmartGridComm), Beijing, China, 21–23 October 2019; IEEE: Piscataway, NJ, USA, 2019; pp. 1–6.
59. Amin, M.A.A.; Hoque, M.A. Comparison of ARIMA and SVM for Short-Term Load Forecasting. In Proceedings of the 2019 9th Annual Information Technology, Electromechanical Engineering and Microelectronics Conference (IEMECON), Jaipur, India, 13–15 March 2019; IEEE: Piscataway, NJ, USA, 2019; pp. 1–6.
60. Zou, Z.; Wu, X.; Zhao, Z.; Wang, Q.; Bie, Y.; Zhou, M. Prediction of Short Term Electric Load Based on BP Neural Networks & ARIMA Combination. In Proceedings of the 2018 IEEE 4th Information Technology and Mechatronics Engineering Conference (ITOEC), Chongqing, China, 14–16 December 2018; IEEE: Piscataway, NJ, USA, 2018; pp. 1671–1674.
61. Eljazzar, M.M.; Hemayed, E.E. Enhancing Electric Load Forecasting of ARIMA and ANN Using Adaptive Fourier Series. In Proceedings of the 2017 IEEE 7th Annual Computing and Communication Workshop and Conference (CCWC), Las Vegas, NV, USA, 9–11January 2017; IEEE: Piscataway, NJ, USA, 2017; pp. 1–6.
62. Karthika, S.; Margaret, V.; Balaraman, K. Hybrid Short Term Load Forecasting Using ARIMA-SVM. In Proceedings of the 2017 Innovations in Power and Advanced Computing Technologies (i-PACT), Vellore, India, 21–22 April 2017; IEEE: Piscataway, NJ, USA, 2017; pp. 1–7.
63. Khuntia, S.R.; Rueda, J.L.; van der Meijden, M.A.M.M. Volatility in Electrical Load Forecasting for Long-Term Horizon—An ARIMA-GARCH Approach. In Proceedings of the 2016 International Conference on Probabilistic Methods Applied to Power Systems (PMAPS), Beijing, China, 16–20 October 2016; IEEE: Piscataway, NJ, USA, 2016; pp. 1–6.
64. Puspitasari, I.; Akbar, M.S.; Lee, M.H. Two-Level Seasonal Model Based on Hybrid ARIMA-ANFIS for Forecasting Short-Term Electricity Load in Indonesia. In Proceedings of the 2012 International Conference on Statistics in Science, Business and Engineering (ICSSBE), Langkawi, Kedah, Malaysia, 10–12 September 2012; IEEE: Piscataway, NJ, USA, 2012; pp. 1–5.
65. Wei, L.; Zhen-gang, Z. Based on Time Sequence of ARIMA Model in the Application of Short-Term Electricity Load Forecasting. In Proceedings of the 2009 International Conference on Research Challenges in Computer Science, Shanghai, China, 28–29 December 2009; IEEE: Piscataway, NJ, USA, 2009; pp. 11–14.
66. de Andrade, L.C.M.; da Silva, I.N. Very Short-Term Load Forecasting Based on ARIMA Model and Intelligent Systems. In Proceedings of the 2009 15th International Conference on Intelligent System Applications to Power Systems, Curitiba, Brazil, 8–12 November 2009; IEEE: Piscataway, NJ, USA, 2009; pp. 1–6.
67. Hor, C.-L.; Watson, S.J.; Majithia, S. Daily Load Forecasting and Maximum Demand Estimation Using ARIMA and GARCH. In Proceedings of the 2006 International Conference on Probabilistic Methods Applied to Power Systems, Stockholm, Sweden, 11–15 June 2006; IEEE: Piscataway, NJ, USA, 2006; pp. 1–6.
68. He, Y.; Zhu, Y.; Duan, D. Research on Hybrid ARIMA and Support Vector Machine Model in Short Term Load Forecasting. In Proceedings of the Sixth International Conference on Intelligent Systems Design and Applications, Jinan, China, 16–18 October 2006; IEEE: Jinan, China, 2006; Volume 1, pp. 804–809.
69. Nazarko, J.; Jurczuk, A.; Zalewski, W. ARIMA Models in Load Modelling with Clustering Approach. In Proceedings of the 2005 IEEE Russia Power Tech, St. Petersburg, Russia, 27–30 June 2005; IEEE: Piscataway, NJ, USA, 2005; pp. 1–6.
70. Lu, J.C.; Niu, D.X.; Jia, Z.Y. A Study of Short-Term Load Forecasting Based on ARIMA-ANN. In Proceedings of the Proceedings of 2004 International Conference on Machine Learning and Cybernetics (IEEE Cat. No.04EX826), Shanghai, China, 26–29 August 2004; IEEE: Piscataway, NJ, USA, 2004; pp. 3183–3187.

71. Juberias, G.; Yunta, R.; Garcia Moreno, J.; Mendivil, C. A New ARIMA Model for Hourly Load Forecasting. In Proceedings of the 1999 IEEE Transmission and Distribution Conference (Cat. No. 99CH36333), New Orleans, LA, USA, 11–16 April 1999; IEEE: Piscataway, NJ, USA, 1999; Volume 1, pp. 314–319.
72. Cho, M.Y.; Hwang, J.C.; Chen, C.S. Customer Short Term Load Forecasting by Using ARIMA Transfer Function Model. In Proceedings of the 1995 International Conference on Energy Management and Power Delivery EMPD '95, Singapore, 21–23 November 1995; IEEE: Piscataway, NJ, USA, 1995; Volume 1, pp. 317–322.
73. Voss, H.U.; Horbelt, W.; Timmer, J. Continuous Nonlinear Delayed-Feedback Dynamics from Noisy Observations. In Proceedings of the AIP Conference Proceedings, Albuquerque, New Mexico, 23–28 June 2002; AIP: Potsdam, Germany, 2002; Volume 622, pp. 163–173.
74. Cellucci, C.J.; Albano, A.M.; Rapp, P.E.; Pittenger, R.A.; Josiassen, R.C. Detecting Noise in a Time Series. *Chaos* **1997**, *7*, 414–422. [CrossRef] [PubMed]
75. Cao, M.T.; Pham, T.T.; Kuo, T.C.; Bui, D.M.; Nguyen, H.V.; Nguyen, T.H. Short-Term Load Forecasting Enhanced With Statistical Data-Filtering Method. In Proceedings of the 2020 IEEE International Conference on Power Electronics, Smart Grid and Renewable Energy (PESGRE2020), Cochin, India, 2–4 January 2020; IEEE: Piscataway, NJ, USA, 2020; pp. 1–8.
76. Akdi, Y.; Dickey, D.A. Pfriodograms of Unit Root Time Series: Distributions and Tests. *Commun. Stat.-Theory Methods* **1998**, *27*, 69–87. [CrossRef]
77. Pérez-Chacón, R.; Luna-Romera, J.; Troncoso, A.; Martínez-Álvarez, F.; Riquelme, J. Big Data Analytics for Discovering Electricity Consumption Patterns in Smart Cities. *Energies* **2018**, *11*, 683. [CrossRef]
78. Noussan, M.; Roberto, R.; Nastasi, B. Performance Indicators of Electricity Generation at Country Level—The Case of Italy. *Energies* **2018**, *11*, 650. [CrossRef]
79. Yao, W.; Essex, C.; Yu, P.; Davison, M. Measure of Predictability. *Phys. Rev. E* **2004**, *69*, 066121. [CrossRef] [PubMed]
80. Gambuzza, L.V.; Buscarino, A.; Fortuna, L.; Porfiri, M.; Frasca, M. Analysis of Dynamical Robustness to Noise in Power Grids. *IEEE J. Emerg. Sel. Top. Circuits Syst.* **2017**, *7*, 413–421. [CrossRef]
81. Cheba, K.; Bąk, I. Environmental Production Efficiency in the European Union Countries as a Tool for the Implementation of Goal 7 of the 2030 Agenda. *Energies* **2021**, *14*, 4593. [CrossRef]