Article

Usage of the Pareto Fronts as a Tool to Select Data in the Forecasting Process—A Short-Term Electric Energy Demand Forecasting Case

Michał Sabat and Dariusz Baczyński *

Electrical Power Engineering Institute, Warsaw University of Technology, Building of Mechanics, ul. Koszykowa 75, 00-662 Warszawa, Poland; michal.sabat.dokt@pw.edu.pl
* Correspondence: dariusz.baczynski@pw.edu.pl

Abstract: Transmission, distribution, and micro-grid system operators are struggling with the increasing number of renewables and the changing nature of energy demand. This necessitates the use of prognostic methods based on ever shorter time series. This study depicted an attempt to develop an appropriate method by introducing a novel forecasting model based on the idea to use the Pareto fronts as a tool to select data in the forecasting process. The proposed model was implemented to forecast short-term electric energy demand in Poland using historical hourly demand values from Polish TSO. The study rather intended on implementing the range of different approaches—scenarios of Pareto fronts usage than on a complex evaluation of the obtained results. However, performance of proposed models was compared with a few benchmark forecasting models, including naïve approach, SARIMAX, kNN, and regression. For two scenarios, it has outperformed all other models by minimum 7.7%.

Keywords: electric energy demand; Pareto fronts in forecasting; K nearest neighbors (kNN) algorithm; nondominated solutions; power engineering challenges

1. Introduction

After decades of a rather stable situation in the electric power engineering sector, almost every year brings new challenges. This sector faces dynamically growing intermittent sources of production of electric energy (wind, PV) [1], changes in energy storages [2], changes in electricity demand (e-mobility, standard of living, social influence) [3,4], changes in existing electricity production methods (pollution, climate) [5]. All of them cause the forecasting to be more complicated and more unreliable.

The obvious point is that accurate forecasting is an important part of any successful planning process. In the electric energy sector, transmission system operators (TSOs), distribution system operators (DSOs), commercial operators (COs), or commercial and technical operators (CTOs) perform numerous different forecasts to plan their activities in an optimal way.

The high accuracy forecasts enable, particularly in the short-term, the TSOs to provide more secure power system operation, i.e., to balance demand with production and minimalize its costs. The impact of forecasts' accuracy is multilevel: it influences not only the security, reliability of the power system, and cost of electrical energy production, but also procurement of both electrical energy and frequency reserves. Moreover, the accuracy influences the comfort of dispatchers' everyday work, which is due to operation on more reliable sets of data about the foreseen situations in the power system for the following, future time steps, used by them in the security management of power systems. In the long run, forecasts' accuracy will influence the possibility of green energy inclusion in energy mix and proper management of energy storage which can give benefits both for customer and energy provider [6].

As mentioned, accuracy is crucial in TSOs' planning and operating processes. In consequence, over the years many TSOs' projects have concentrated on inventing new forecasting models or adjusting the existing ones [7,8].

Future energy systems will create more complex connections and dependencies than the current hierarchical connection of TSOs and DSOs. Most countries change their economy to be environment friendly. That results in emergence of many new entities generating green energy and also many entities using energy responsibly. Responsibly means also usage of every possible energy source and form—not only electric energy. We should be able to manage and optimize multi energy systems [9,10], where harmoniously cooperating thermal, gas, electricity, and other sources will create new quality through synergy [11–13]. Undoubtedly, it will be important to use high-speed computer networks to effectively manage and control all entities [14]. Not only information about energy control signals must be transmitted but also predictions on different system levels must be elaborated and dispatched.

The aforementioned development of computer networks and saturation with measuring devices increases the amount of available data, which can be the basis for making forecasts. This influences the creation of an apparent picture that the time series of phenomena are sufficient to use any prognostic methods. But the mentioned dynamic changes in the energy sector result in variability of relationships between explanatory and explained variables. In other words, prediction models should be elaborated often. This entails necessity to develop and use forecasting methods that are capable of producing precise forecasts on the basis of small amount of data. This means we should have a method that uses short time series but is better than naïve approach or k nearest neighbors (kNN) algorithm. A current trend in energy forecasting is the use of increasingly complex and ensemble methods consisting of LSTM [15] and convolutional [16–19] neural networks. This ensures high quality forecasts, but it is in contradiction with the possibility of using short time series.

In this article, the authors, whose scope of expertise are power systems, propose a new forecasting model evaluated on the short-term electric energy demand forecasting problem (next hour demand). Presented model bases on the historic data sets, combining both Pareto fronts calculation to choose the data from the training set and one of the methodologies to compute the forecast value like the least-squares method. The similar model based on Pareto fronts was presented in [20], but they aimed to identify anomalies among data rather than to forecast values.

Pareto fronts have been used as parts of models forecasting different aspects of power systems [21–24], but never as proposed in this article.

The authors, in this article, focus rather on the range of different approaches to Pareto fronts usage in forecasting and its advantages and disadvantages than on complex evaluation of the obtained results. The process of evaluation will be in the scope of further publications.

The remainder of this article is organized as follows: Section 2 demonstrates the proposed idea of Pareto fronts application for forecasting; Section 3 presents results of experiments; Section 4 discusses the results; Section 5 concludes the article.

2. Materials and Methods

2.1. The Idea behind the Pareto Fronts Usage

The idea to use the Pareto fronts as a tool to select data in the forecasting process originated from the fact that similar, well-known and described in many articles [25–27] machine learning algorithm, k nearest neighbors (kNN), has been successfully applied to that task. This algorithm has been used and described in the literature both as a classification algorithm [28,29] and as a forecasting model. This algorithm is a powerful forecasting tool due to the combination of its simplicity and accuracy. It has been provided in this article, as the kNN algorithm had been chosen as one of the benchmark models. In the kNN models k represents the model's parameter, i.e., the number of data (facts)

selected from the training data set, which are the k closest ones in terms of the explanatory variables (e.g., calculated as the Euclidean distance) to the forecast one. The forecast value is calculated based on the selected k nearest training data values, e.g., as their arithmetic average (as in the Figure 1). Hence, in this approach, the number of similar data is the main factor determining the chosen data from the training data set (set items are used to make the forecast) [30].
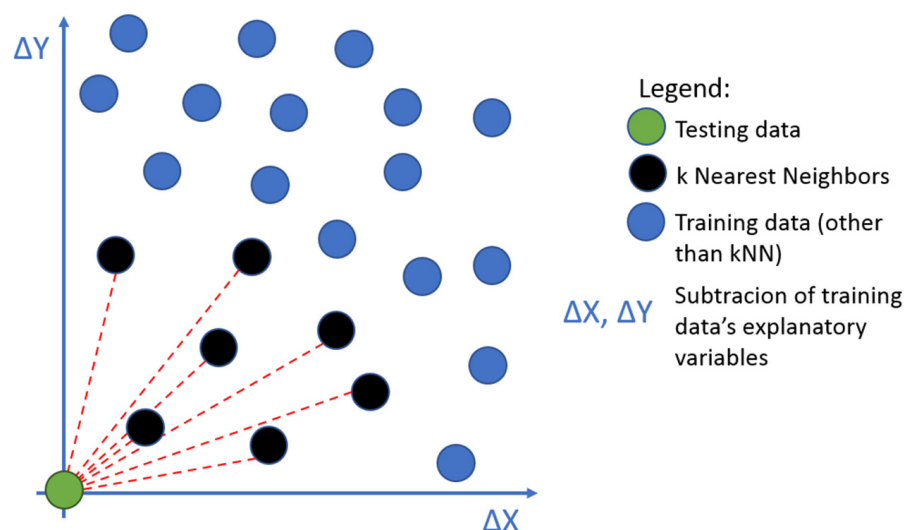

Figure 1. Example of choosing k nearest neighbors from the training data set for k = 7.

The authors' aim was to challenge kNN's approach and verify a similar concept. The new model proposed in this article assumed that the process of choosing similar data, from the training data set, was based on their belonging to the Pareto fronts rather than on the predefined number of nearest data points.

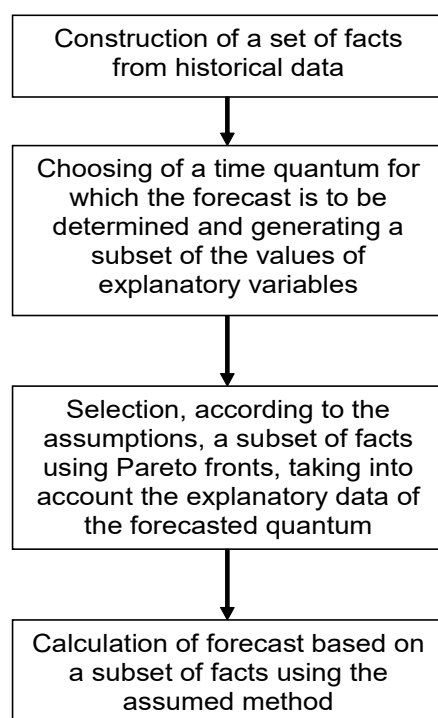The general idea of using Pareto fronts for forecasting is presented in Figure 2.

Figure 2. General idea of using Pareto fronts for forecasting.

The entire process of forecasting begins with the construction of a set of facts from historical data. A single fact (as it is in the case of neural networks) is understood as a set consisting of a subset of the values of the explanatory variables and the corresponding one value of the dependent variable. Facts are constructed from historical data. The method of including the explanatory data in the subset depends on the adopted data scenario on which the forecast is to be based.

Then the time quantum for which the forecast is to be determined should be chosen. For this quantum, the values of the explanatory variables should be determined (in the same way as a subset of the explanatory variables for individual fact).

The key step is to select, according to the assumed scenario, a subset of facts using the Pareto front. Central point location for determining of Pareto fronts is stated by coordinates—values of the explanatory variables of the predicted quantum.

The last step is to determine the appropriate forecast using the assumed method and using the subset of facts selected in the previous step. The methods/scenarios for selecting Pareto fronts and the forecast calculation are described in following sections.

2.2. Theory and Examples of Pareto Fronts

In optimization the Pareto front represents the set of nondominated solutions being chosen as optimal if no objective can be improved without sacrificing at least one other objective [31,32].

Nondominated solution is defined in [33] as "A point, $x* \in X$, is Pareto optimal if there does not exist another point, $x \in X$, such that $F(x) \leq F(x*)$, and $Fi(x) < Fi(x*)$ for at least one function." In the authors' approach to forecasting and choosing nearest neighbors, Pareto front represents training data points that are nondominated in sense of distance (in many dimensions) to testing point. Each explanatory variable creates one space dimension.

In other words, in case of problem with two explanatory variables (two dimensions $\Delta X, \Delta Y$) and when sought values for Pareto front are minimalized, as in the Figure 3, the points on the Pareto front (represented by the red dots in the Figure 3) are:

- Those elements whose values are the closest to the zero value in dimension ΔX or ΔY;
- Those elements which with one another have one value closer to zero in one dimension than other elements and one value further than other elements in the second dimension.

Mentioned above two-dimensional elements have their values calculated:

- For ΔX dimension as differences between testing data value representing value connected with first explanatory variable X and relevant corresponding training example' value of this variable;
- For ΔY dimension as differences between testing data value representing value connected with the second explanatory variable Y and relevant corresponding training example value of this variable.

As a result there is no unequivocal methodology to compare the points on Pareto front to one another in situations when parameters represented by the two dimensions ($\Delta X, \Delta Y$) are so different from each other (for example price and color). The testing data represents data for which the forecast value is calculated.
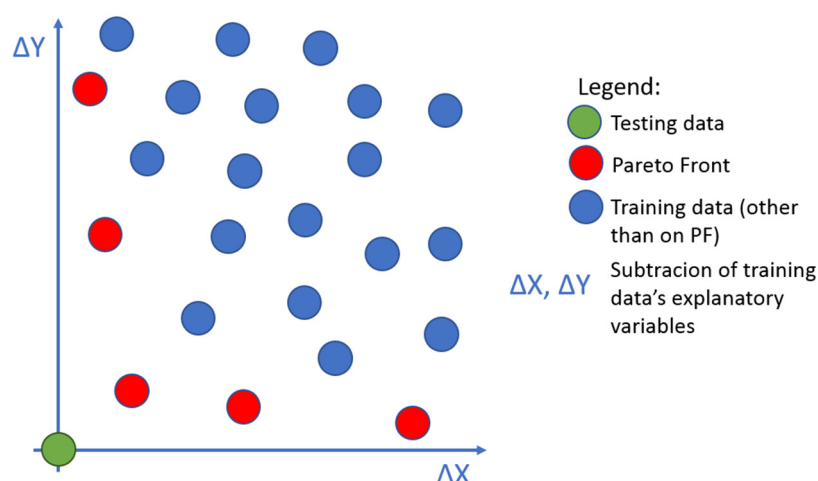
Figure 3. The example of training data belonging to the first Pareto front in one quadrant.

In the proposed model, as compared to kNN algorithm, there is no parameter directly defining the number of training data points belonging to the Pareto fronts. In consequence, the selection of the data for the forecast computation is, while still structured, randomized. This can be viewed as an advantage, as the users are not required to determine the optimal number of 'k' neighbors themselves, and a disadvantage considering the lack of control over the data selection process. However, the proposed model contains a new, different parameter defining the number of Pareto fronts, from which training data are used to set forecast values. An example of three Pareto fronts is presented in Figure 4.
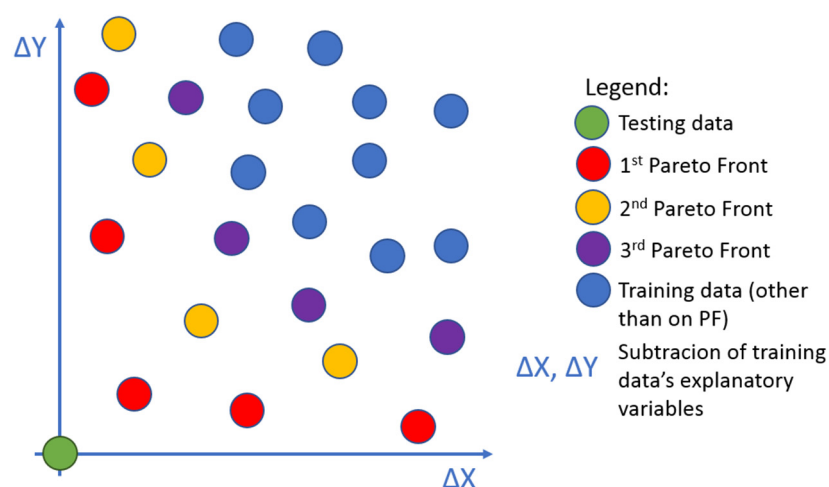
Figure 4. The example of testing data belonging to the first, the second, and the third Pareto front in one quadrant.

2.3. Implementation Options

There are many methods of applying proposed Pareto fronts model to select data from training data set and to use them further in the forecasting process. Those taken into account by authors, implemented and verified by them are described and presented below.

Scenario options and their combination on which Pareto fronts as model has been verified as described in Section 3: Application and Results.

Option 1: The set of training data contains all available training data (there are no limitations added, all historic data from the set are treated as the training data).

Option 2: The set of training data is limited exclusively to data representing historical hours equal to the forecast hour.

Option 3: The forecast value is calculated as the arithmetic average of testing data belonging to Pareto front/s.

Option 4: The forecast value is obtained as the result of the linear regression calculated on testing data belonging to Pareto front/s.

Option 5: Explanatory variables are:

(a) Historical hourly demand values for different historic hours (for example training data is characterized by two explanatory variables, i.e., on ΔX dimension by the demand in the previous hour (h-1) and on ΔY dimension by the demand on the same hour but on the previous day (h-24)).

(b) Days of the week (values are in range from -3 to 3, i.e., value equals 0 when the forecast hour comes from the same day of the week as the historic hour, for example, both are Wednesdays).

(c) Hours (values are in range from -11 to 12, i.e., value equals 0 when the forecast hour is the same as the historic hour from historic day, for example, both are 10 a.m.).

(d) Meteorological data (historical temperature demand values, for example, when training data is characterized by three explanatory variables and on ΔZ dimension is by the temperature in the previous hour (h-1)).

Option 6: There are differences in the size of set of training data.

All presented above scenario options have been verified and are presented in Section 3: Application and Results with necessary examples and results provided.

Additionally, to compare achieved results from Pareto front model the following benchmark approach, algorithm, and models have been implemented:

• Naïve approach (used as benchmark in [34,35],
• SARIMAX model (different variations of ARIMA model i.e., ARIMA, SARIMA and SARIMAX used to forecast demand has been described in [36–38],
• K nearest neighbors algorithm,
• Additional 'no data selection' model.

Above-mentioned approach, algorithm, and models have been characterized by the following sets of parameters:

(1) Naïve approach in three variants:

   (a) Forecast value for the hour equals previous hour's value (h-1);
   (b) Forecast value for the hour equals the value for the same hour but the previous day (h-24);
   (c) Forecast value for the hour equals the arithmetic average of values from points a and b (h-1, h-24).

(2) SARIMAX model with following values verified and giving the best results for the analyzed sets of data:

   (a) Trend parameters:

      • autoregression order (p) equals 2;
      • difference order (d) equals 1;
      • moving average order (q) equals 2.

   (b) Seasonal parameters:

      • autoregressive order (P) equals 1;
      • difference order (D) equals 1;
      • moving average order (Q) equals 0;
      • the number of time steps for a single seasonal period (m) equals 24.

(3) K nearest neighbors algorithm with the k parameter's value is from set (1, 7).

(4) 'No data selection' model in which forecast value is calculated equally to the corresponding Pareto front scenarios mentioned above, but the forecast calculation involves all available training data values without any kind of selection.

Scenario options for Pareto front model verification and benchmark approach, algorithm and models described above have been implemented and described both in two- and three-dimensional spaces (in Section 3: Application and Results; Sections 3.1 and 3.2, respectively). Each verification case contains set of results and detailed information about forecast period, range of available historic data, methodology of forecast value calculations.

For cases in two-dimensional spaces, Pareto fronts have been obtained in all quadrants individually.

The example of the first Pareto front visualization is presented in Figure 5 and of the first, the second, and the third Pareto fronts in Figure 6, respectively.
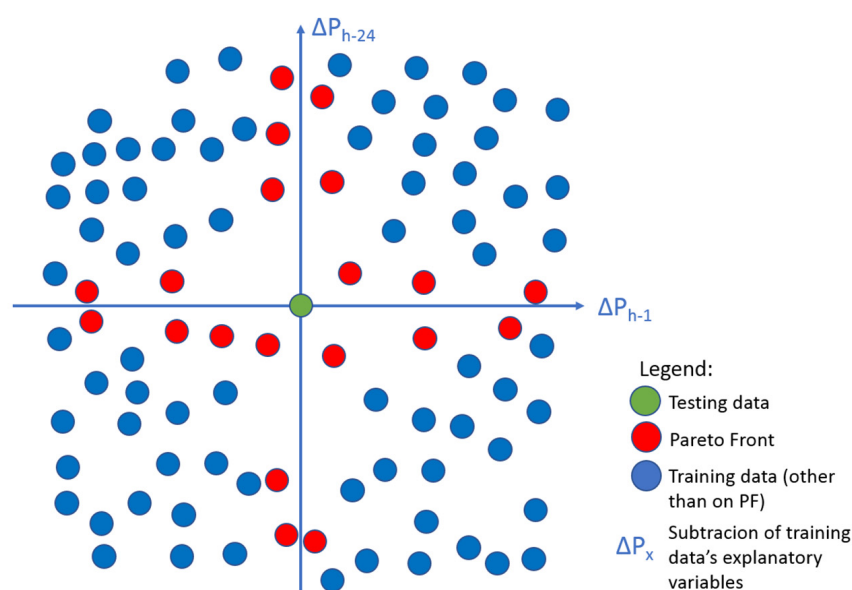

Figure 5. The example of testing data belonging to the first Pareto front in all four quadrants.
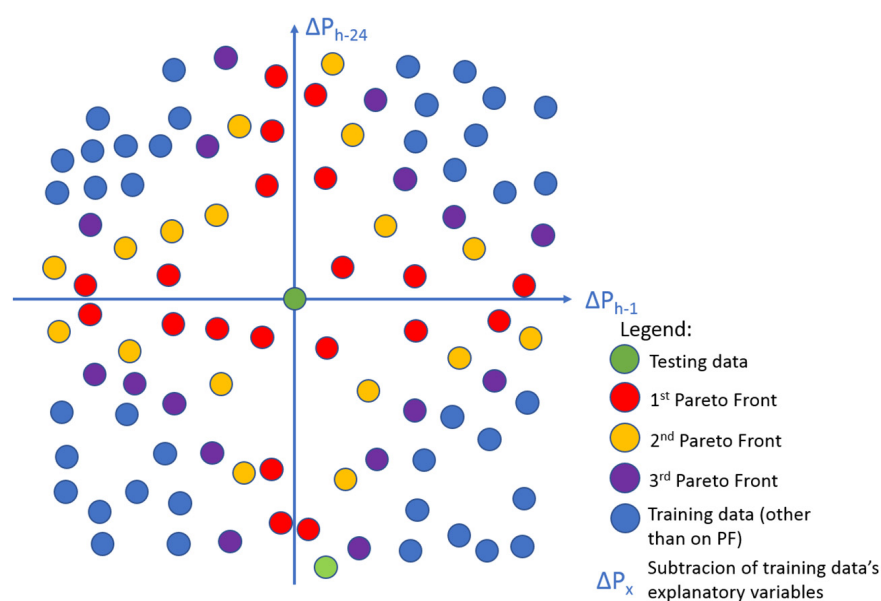

Figure 6. The example of testing data belonging to the first, the second, and the third Pareto front in all four quadrants.

The values of the elements in basic verification cases represents differences between the pairs of historical hourly demand values from h-1 and h-24 as explanatory variables (as described in this Section's scenario Option 5a). For example, when the demand for 6 p.m. on 1.06.2018 is forecast, the points in two-dimensional space are characterized by the two following values:

- The first one as the parameter representing ΔX dimension equals the difference between the value of demand for one hour before the forecast one which is for 5 p.m. on 1.06.2018 and the value of demand for one of the previous hours from the available data set like 4 p.m. on 1.06.2018.
- The second one as the parameter representing ΔY dimension equals the difference between the value of demand for the hour 24 hours before the forecast one which is for 6 p.m. on 31.05.2018 and the value of demand for the hour 24 hours before the hour chosen from the available data set in process of setting parameter ΔX which is 5 p.m. on 31.05.2018.

In that case if the element characterized by historical hourly demand values from 4 p.m. on 1.06.2018 and 5 p.m. on 31.05.2018 is situated on Pareto front, the historical demand value from 5 p.m. on 1.06.2018 will be used to calculate forecast value.

Based on information from this Section's scenario Options 3 and 4, forecast value might be calculated in different ways. Illustration of one of the approaches, which is when the forecast value is calculated based on linear regression modeled based on values located on the Pareto fronts, is presented in Figure 7. The presented orange rectangle represents the plane that most closely fits values (green dots) of the points on Pareto front (here the red dots) used to calculate forecast value. The green rhomb marks the point on the rectangle surface, when values of parameters ΔX and ΔY equal zero, representing the forecast value.
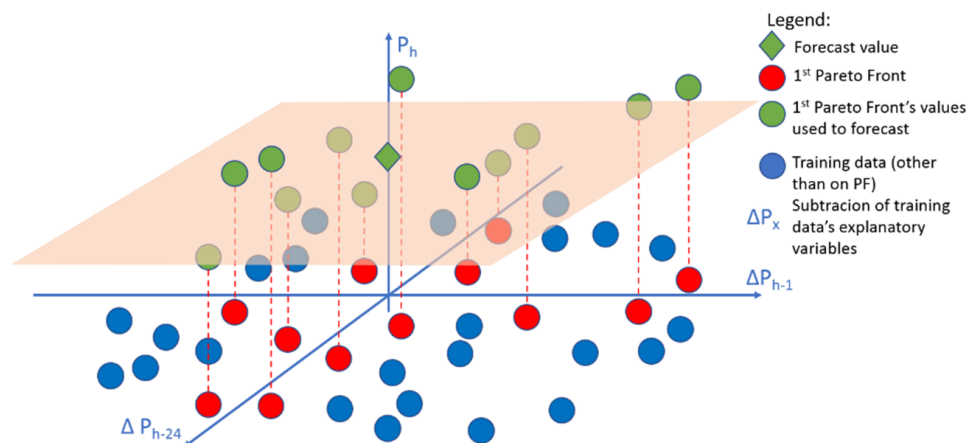

Figure 7. Example of choosing data belonging to the first, the second, and the third Pareto fronts in all four quadrants.

2.4. Data

The proposed model was implemented to forecast short-term electric energy demand in Poland using historical hourly demand values from Polish TSO. Polish TSO publishes historical hourly demand values for Poland on its website https://www.pse.pl/obszary-dzialalnosci/krajowy-system-elektroenergetyczny/zapotrzebowanie-kse (accessed on 5 September 2018). The exact forecasting task was to predict next hour demand having available data of hourly demand for previous hours and some additional explanatory variables—depending on case considered. For each case, the data was specially collected for all methods compared. The results depicted in this paper are only a small, more interesting part of numerous experiments.

The set of weather data was provided to authors directly by employees of the Interdisciplinary Centre for Mathematical and Computational Modelling at the University of Warsaw.

3. Results

In this Section, as mentioned in Section 1: Introduction, authors present the range of different approaches to Pareto fronts usage in forecasting rather than the complex evaluation of the obtained results. However, to present and compare different scenarios' results (verification cases) obtained from Pareto front model and benchmark approach, model, and algorithm within the same verification case and between different cases the mean absolute percentage error (MAPE) was calculated and used:

$$\text{MAPE} = \frac{1}{n} \sum_{t=1}^{n} \left| \frac{A_t - F_t}{A_t} \right| \cdot 100\% \tag{1}$$

where:

A_t—The actual hourly value

F_t—The forecast hourly value

n—The total number of the forecast values

The percentage of the error for every forecast horizon is averaged resulting in the MAPE value for the Pareto front and benchmark approach, model, and algorithm. The mean absolute percentage error is considered as standard measurement, for electric forecasting, in the process of verifying the accuracy of the models [36].

3.1. Results for the Two-Dimensional Spaces

Results for the first verification case are given in Tables 1 and 2. The first verification case assumptions (in line with Section 2 scenario Options 2, 3, and 5a):

(1) Forecast period: 1–30.06.2018;
(2) Available historic data from: 1.05.2018;
(3) The same historic hours as forecast;
(4) Pareto fronts model uses differences between the pairs of historical demand values from h-1 and h-24 as explanatory variables;
(5) Forecast value is calculated as arithmetic average of values on Pareto fronts.

Table 1. MAPE values for naïve approach, SARIMAX model, and kNN algorithm for the first verification case.

Naïve Approach [%]			SARIMAX [%]	kNN [%]						
h-1	h-24	h-1, h-24	(2,1,2) (1,1,0,24)	k = 1	k = 2	k = 3	k = 4	k = 5	k = 6	k = 7
3.181	7.598	4.470	0.918	1.213	1.223	1.293	1.407	1.446	1.579	1.676

Table 2. MAPE values for Pareto fronts and 'no data selection' models for the first verification case.

Pareto Fronts [%]			No Data Selection [%]
1 Front	2 Fronts	3 Fronts	-
3.511	4.540	5.130	13.660

Results for the second verification case are given in Tables 3 and 4. The second verification case assumptions (in line with Section 2 scenario Options 2, 4, and 5a, and no limitations in case of Option 6):

(1) Forecast period: 1–30.06.2018;
(2) Available historic data from: 1.05.2018;
(3) The same historic hours as forecast;

(4)   Pareto fronts model uses historical demand values from h-1 and h-24 as explanatory variables;

(5)   Forecast value is calculated based on linear regression modeled based on values located on the Pareto fronts/all training data.

Table 3. MAPE values for naïve approach, SARIMAX model, and kNN algorithm for the second verification case.

Naïve Approach [%]			SARIMAX [%]	kNN [%]						
h-1	h-24	h-1, h-24	(2,1,2) (1,1,0,24)	k = 1	k = 2	k = 3	k = 4	k = 5	k = 6	k = 7
3.181	7.598	4.470	0.918	1.213	1.223	1.293	1.407	1.446	1.579	1.676

Table 4. MAPE values for Pareto fronts and 'no data selection' models for the second verification case.

Pareto Fronts [%]			No Data Selection [%]
1 Front	2 Fronts	3 Fronts	-
1.100	0.687	0.705	0.718

Results for the third verification case are given in Tables 5 and 6. The third verification case assumptions (in line with Section 2 scenario Options 2, 4, 5a, and no limitations in case of Option 6), with longer forecast period than in the previous cases:

(1)   Forecast period 1.06–31.08.2018;

(2)   Available historic data from 1.05.2018;

(3)   The same historic hours as forecast;

(4)   Pareto fronts model uses historical demand values from h-1 and h-24 as explanatory variables;

(5)   Forecast value is calculated based on linear regression modeled based on values located on the Pareto fronts/all training data.

Table 5. MAPE values for naïve approach, SARIMAX model, and kNN algorithm for the third verification case.

Naïve Approach [%]			SARIMAX [%]	kNN [%]						
h-1	h-24	h-1, h-24	(2,1,2) (1,1,0,24)	k = 1	k = 2	k = 3	k = 4	k = 5	k = 6	k = 7
3.203	7.702	4.563	0.889	1.027	0.993	1.019	1.049	1.079	1.144	1.209

Table 6. MAPE values for Pareto fronts and 'no data selection' models for the third verification case.

Pareto Fronts [%]			No Data Selection [%]
1 Front	2 Fronts	3 Fronts	-
1.143	0.703	0.722	0.762

Results for the fourth verification case are given in Tables 7 and 8. The fourth assumptions (in line with Section 2 scenario Options 1, 4, 5a, and no limitations in case of Option 6):

(1)   Forecast period 1–30.06.2018;

(2)   Available historic data from 1.05.2018;

(3)   The all historic hours;

(4)   Pareto fronts methodology uses historical demand values from h-1 and h-24 as explanatory variables;

(5)   Forecast value is calculated based on linear regression modeled based on values located on the Pareto fronts/all training data.

Table 7. MAPE values for naïve approach, SARIMAX model, and kNN algorithm for the fourth verification case.

Naïve Approach [%]			SARIMAX [%]	kNN [%]						
h-1	h-24	h-1, h-24	(2,1,2) (1,1,0,24)	k = 1	k = 2	k = 3	k = 4	k = 5	k = 6	k = 7
3.181	7.598	4.470	0.918	2.360	2.236	2.266	2.249	2.219	2.259	2.267

Table 8. MAPE values for Pareto fronts and 'no data selection' models for the fourth verification case.

Pareto Fronts [%]			No Data Selection [%]
1 Front	2 Fronts	3 Front	-
2.489	2.607	2.705	3.126

Results for the first verification case are given in Tables 9 and 10. The fifth verification case assumptions (in line with Section 2 scenario Options 1, 4, combination of Options 5a and 5b, and no limitations in case of Option 6):

(1)   Forecast period 1–30.06.2018;
(2)   Available historic data from 1.05.2018;
(3)   The all historic hours;
(4)   Pareto fronts model uses historical demand value from h-1 and difference between days of the week as explanatory variables;
(5)   Forecast value is calculated based on linear regression modeled based on values located on the Pareto fronts/all training data.

Table 9. MAPE values for naïve approach, SARIMAX model, and kNN algorithm for the fifth verification case.

Naïve Approach [%]			SARIMAX [%]	kNN [%]						
h-1	h-24	h-1, h-24	(2,1,2) (1,1,0,24)	k = 1	k = 2	k = 3	k = 4	k = 5	k = 6	k = 7
3.181	7.598	4.470	0.918	3.962	3.505	3.362	3.298	3.273	3.237	3.226

Table 10. MAPE values for Pareto fronts and 'no data selection' models for the fifth verification case.

Pareto Fronts [%]			No Data Selection [%]
1 Front	2 Fronts	3 Fronts	-
3.461	3.232	3.251	3.220

Results for the sixth verification case are given in Tables 11 and 12. The sixth verification case assumptions (in line with Section 2 scenario Options 1, 4, combination of Options 5a and 5c, and no limitations in case of Option 6):

(1)   Forecast period 1–30.06.2018;
(2)   Available historic data from 1.05.2018;
(3)   The all historic hours;
(4)   Pareto fronts model uses historical demand value from h-1 and difference between hours as explanatory variables;
(5)   Forecast value is calculated based on linear regression modeled based on values located on the Pareto fronts/all training data.

Table 11. MAPE values for naïve approach, SARIMAX model, and kNN algorithm for the sixth verification case.

Naïve Approach [%]			SARIMAX [%]	kNN [%]						
h-1	h-24	h-1 h-24	(2,1,2) (1,1,0,24)	k = 1	k = 2	k = 3	k = 4	k = 5	k = 6	k = 7
3.181	7.598	4.470	0.918	1.002	0.924	0.937	0.934	0.979	1.012	1.038

Table 12. MAPE values for Pareto fronts and 'no data selection' models for the sixth verification case.

Pareto Fronts [%]			No Data Selection [%]
1 Front	2 Fronts	3 Fronts	-
2.536	2.500	2.581	3.207

Results for six verification cases described above are illustrated in Figures 8 and 9. In the case of proposed approach (Figure 8) best results were obtained for second and third cases. Other cases gave inferior forecast quality. For mentioned best cases it can be pointed out that MAPE is maximal for one Pareto front, having minimum for two Pareto fronts, and goes little up for three Pareto fronts. This is probably due to the fact that points from the third Pareto front are different from the forecasted point than those from first and second Pareto fronts. For kNN method (Figure 9) a similar phenomenon is visible. For the best three cases (first, third, and sixth), best results are obtained for two neighbors. Obviously, this can be called only a "similar phenomenon" because operation of these two methods is different.
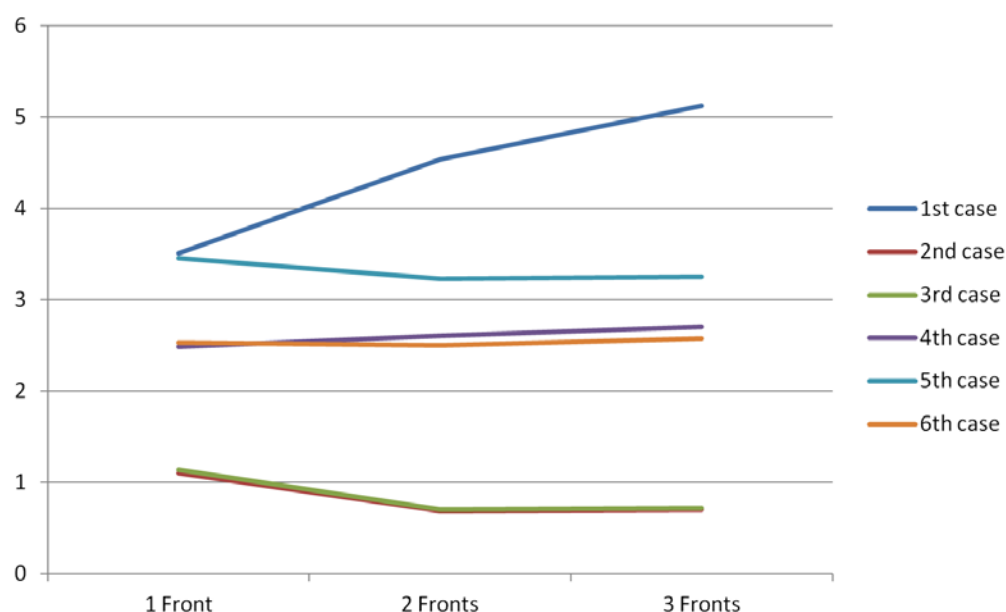
Figure 8. MAPE [%] for different cases of Pareto front utilization Table 2, Table 4, Table 8, Table 10, and Table 12).
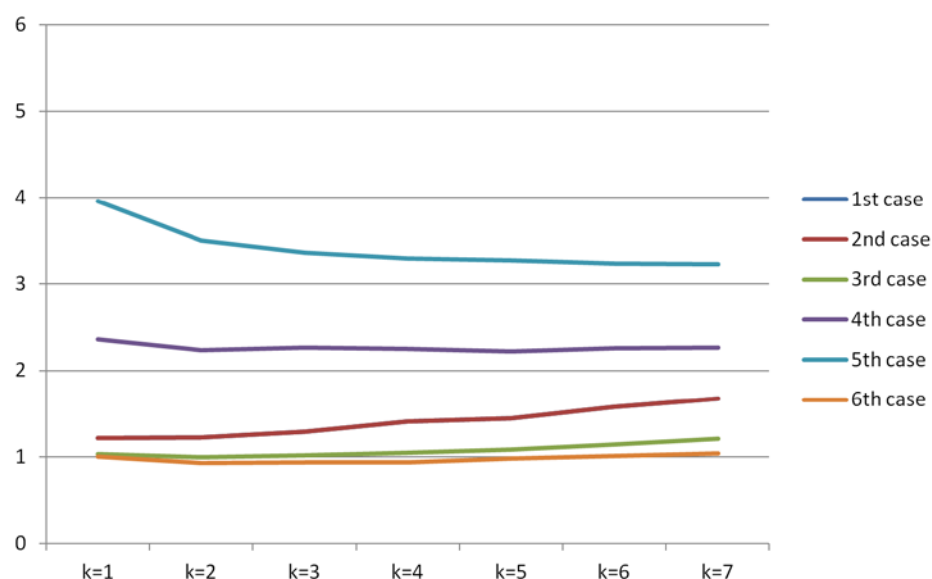
Figure 9. MAPE [%] for different cases of kNN method utilization (Table 1, Table 3, Table 7, Table 9, and Table 11).

Additionally, the impact of the data history length on the forecast accuracy has been under investigation with the results as in Table 13 (in line with Section 2 scenario Options 1, 4, 5a, and with limitations in case of Option 6):

(1)　Forecast period 1–30.06.2018;
(2)　Available historic data from 1.05.2018;
(3)　The different sets of historic hours;
(4)　Pareto fronts models uses historical demand values from h-1 and h-24 as explanatory variables;
(5)　Forecast value is calculated based on linear regression modeled based on values located on the Pareto fronts/all training data.

Table 13. MAPE values for Pareto fronts models with and without available historic data limitation.

	Number of Historic Day Available [%]					All Historic Data Available [%]
	10	**15**	**20**	**25**	**30**	
1 Front	2.484	2.522	2.569	2.519	2.686	2.489
2 Fronts	2.622	2.646	2.730	2.819	2.955	2.607
3 Fronts	2.714	2.789	2.893	2.921	3.022	2.705

In case of calculation involving three explanatory variables (presented in Section 3.2) the historical hourly demand values from 2015, not from 2018, were used due to having more available adequate meteorological data for that year. Therefore, one extra simplified simulation was run to obtain results for Pareto front model based on two explanatory variables from 2015 (Table 14) required to have a benchmark for the three-dimensional one.

The benchmark simplified 2015 verification case assumptions (in line with Section 2 scenario Options 1, 4, 5a and no limitations in case of Option 6.6):

(1)　Forecast period 1–30.06.2015;
(2)　Available historic data from 1.05.2015;
(3)　The all historic hours;
(4)　Pareto fronts model uses historical demand values from h-1 and h-24 as explanatory variables;
(5)　Forecast value is calculated based on linear regression modeled based on values located on the Pareto fronts/all training data.

Table 14. MAPE values for Pareto fronts and 'no data selection' models for the benchmark simplified 2015 verification case.

	Pareto Fronts [%]		No Data Selection [%]
1 Front	**2 Fronts**	**3 Fronts**	**-**
2.782	2.953	3.023	3.314

3.2. Results for the Three-Dimensional Spaces

Results for the three-dimensional verification case are given in Tables 15 and 16. The verification case assumptions (in line with Section 2 scenario Options 1, 4, combination of Options 5a and 5d, and no limitations in case of Option 6):

(1)   Forecast period 1–30.06.2015;
(2)   Available historic data from 1.05.2015;
(3)   The all historic hours;
(4)   Pareto fronts model uses historical demand values from h-1 and h-24 and temperature from h-1 as explanatory variables;
(5)   Forecast value is calculated based on linear regression modeled based on values located on the Pareto fronts/all training data.

Table 15. MAPE values for naïve approach, SARIMAX model, and kNN algorithm for the three-dimensional verification case.

Naïve Approach [%]			SARIMAX [%]	kNN [%]							
h-1	**h-24**	**h-1, h-24**	**(2,1,2) (1,1,0,24)**	**k = 1**	**k = 2**	**k = 3**	**k = 4**	**k = 5**	**k = 6**	**k = 7**	
3.349	8.057	4.742	0.972	4.180	4.174	4.475	4.591	4.776	5.060	5.285	

Table 16. MAPE values for Pareto fronts and 'no data selection' models for the three-dimensional verification case.

	Pareto Fronts [%]		No Data Selection [%]
1 Front	**2 Fronts**	**3 Fronts**	**All**
2.692	2.784	2.832	3.137

3.3. Two- and Three-Dimensional Spaces' Quadrants and Octants Analysis

All Pareto fronts results presented in Sections 3.1 and 3.2 were calculated based on all nondominated solutions situated in all four quadrants, for two explanatory variables in Section 3.1 and all eight octants in Section 3.2 for three explanatory variables. The focus of further analysis was on investigating results in individual quadrants and octants and their potential impact on the final forecast value. The following two options were considered and verified in terms of searching possible enhancement of Pareto front model:

(1)   Forecasts based on nondominated solutions situated in only one quadrant/octant (from four or eight available);
(2)   Usage of forecasts based on nondominated solutions situated in each of the quadrants/octants (four or eight) in order to calculate the final forecast demand value.

For both options individual quadrants and octants analyses were performed based on the assumptions like the one in the third scenario in the Section 3.1 and in scenario in the Section 3.2:

(1)   Forecast period 1–30.06.2015;
(2)   Available historic data from 1.05.2015;
(3)   The all historic hours;

(4) Pareto fronts model uses historical demand values from h-1 and h-24 (for four quadrants analysis) and additional temperature from h-1 as explanatory variables (for eight octants analysis);

(5) Forecast value is calculated based on linear regression modeled based on values located on the Pareto fronts.

A meticulously performed analysis, for both options described above, exhibit instability and unpredictability in terms of the number of historic data points appearing in each quadrant and octant. This instability and unpredictability resulted in the occurrence of hours for which the number of the data points on Pareto fronts in one or more quadrants/octants were not sufficient to use linear regression to calculate the forecast value; in an extreme example, that number equaled zero. When the number of data points was not sufficient to use linear regression but was still different from zero, that issue could be solved by using an alternative method to calculate the forecast for those hours, for instance the mean average. However, there were hours in which the number of training examples in a quadrant/octant equaled zero. In consequence there was no method allowing the forecast value calculation to apply, which could solve this issue. The simulations showed that the issue with the insufficient number of selected training examples was not connected with any particular quadrant nor octant. The insufficient number of selected data points occurred occasionally in each quadrant/octant depending on the data set. In consequence, the first option of enhancement proposed in this subsection cannot be treated as a potential possibility to achieve that, even though the analyzed cases proved that there was potential in this approach due to hours in which forecast values, calculated based on results from one of the spaces instead of from all of them, are more accurate.

As far as the second option of enhancement was concerned, the abovementioned instability and unpredictability foreclosed constructing an enhancing universal adjustment to the proposed Pareto front model. Moreover, the same obstacles, as described for the first option, apply to that one. For the second option, verified forecast calculation as mean average value from forecast values calculated for each quadrant/octant (for four and eight spaces) and tried to find equation coefficients for the following equation (for quadrants):

$$a * f_1 + b * f_2 + c * f_3 + d * f_4 = f_f \tag{2}$$

where:

   a—the first quadrant forecast value's coefficient
   b—the second quadrant forecast value' coefficient
   c—the third quadrant forecast value's coefficient
   d—the fourth quadrant forecast value' coefficient
   f_1—forecast value obtained based on the Pareto front in the first quadrant
   f_2—forecast value obtained based on the Pareto front in the second quadrant
   f_3—forecast value obtained based on the Pareto front in the third quadrant
   f_4—forecast value obtained based on the Pareto front in the fourth quadrant
   f_5—final forecast value

Both attempts to did not enhance the Pareto front model, therefore further work concerning them was not continued.

4. Discussion

The authors presented the new forecasting model based on the idea to use Pareto fronts as a tool to select data in the forecasting process with its various variants. The proposed model was implemented to forecast short-term electric energy demand in Poland.

The publication aimed to present the range of different approaches to Pareto fronts usage in forecasting, thus the descriptions of eight basic verification cases and two additional developed approaches were provided in the text. Naïve approach, SARIMAX model, and kNN algorithm, and 'no data selection' model were implemented as benchmarks to the proposed Pareto front model.

As far as two-dimensional spaces, i.e., with two explanatory variables concerned, the first verification case focused on using the set of training data that was limited exclusively to data representing historical hours equal to the forecast hour. Simple forecast value calculation as the arithmetic average of values on Pareto fronts did not deliver accurate results, comparing not only to SARIMAX but also with kNN and naïve approach's results. However, the change in forecast value calculation, like in the second and third cases, from obtaining arithmetic average to linear regression modelling based on values located on the Pareto fronts increased results accuracy. This additional training data preselection resulted, in the second verification case, in the smallest mean absolute percentage error not only among all benchmark models' results for that specific case but also as compared to results obtained in all other scenarios. Therefore, the second case's assumptions are the most promising for further investigation. On the other hand, the difference between MAPE value for Pareto front with two fronts and 'no data selection' models equaled 0.031% (MAPE), but it also means improvement by 4.3% in forecast quality. In this case, training data selection provided by Pareto fronts improved the accuracy of the obtained results, but not significantly.

The third case results confirmed those obtained in the second verification case presenting the Pareto front method advantage again. The difference between MAPE value for Pareto front with two fronts and 'no data selection' models equaled 0.059% (MAPE), but it also means stronger improvement than in previous case—by 7.7% in forecast quality. However, that case showed also that the set of training data containing data representing longer history than in the second case, may negatively influence the accuracy, e.g., by 0.016% in MAPE for two fronts.

Further investigation will focus both on the Pareto front model development and identifying this method's advantages over 'no data selection'.

Further verification cases, i.e., from number four to six, showed that lack of the additional training data preselection and, using the difference between days of the week or hours as explanatory variables, decreased the Pareto fronts models' forecast accuracy.

For the case aimed to verify results for three-dimensional spaces, i.e., with three explanatory variables both historical demand and temperature values, adding the third dimension improved the proposed method's forecast accuracy by 0.09% in MAPE for the best one front calculations. This slight improvement causes that three- and more-dimensional cases will be implemented and verified in further investigation.

5. Conclusions

In all analyzed cases so far, all Pareto fronts results were calculated based on all nondominated solutions situated in all four quadrants/eight octants.

Additionally, the authors proposed and examined the Pareto front model to obtain forecasts based on nondominated solutions situated in each quadrant/octant individually or on nondominated solutions situated in each of the quadrants/octants (four or eight) in order to calculate the final forecast demand value. Even though those approaches seemed intriguing, they both had a disadvantage of instability and unpredictability in terms of the number of historic data appearing in each quadrant and octant. The extreme examples when the number of data equaled zero were the reason why those ideas were not developed and are not planned to be in the future.

The new forecasting model based on the idea to use the Pareto fronts as a tool to select data in the forecasting process was presented and its various variants were analyzed. Two-dimensional cases assuming additional training data preselection and forecast obtained from linear regression modelling based on values located on the Pareto fronts provided promising results' accuracy. For two investigated cases (second and third) proposed model gave best results among all benchmark methods. The forecast quality was better by 4.3% and 7.7%, respectively, comparing to next best method.

As stated, the authors find the presented idea to be very promising. In this article they have presented more interesting results than they have obtained so far. These results give rise to many subsequent questions which the authors intend to analyze. As the most interesting, the authors indicate the following issues:

- Tests of the approach for other time series (e.g., wind energy forecast),
- Development of other methods of final forecast calculation from Pareto fronts sets,
- Tests of the approach in more than three-dimensional spaces,
- Development of a new approach for more than three-dimensional spaces,
- Development of hybrid and ensemble approaches.

References

1. Meshcheryakova, T.; Pigurin, A. RES development trends that determine the sustainable development of the energy system of the future. *E3S Web Conf.* **2019**, *135*, 04051. [CrossRef]
2. Maggio, G.; Nicita, A.; Squadrito, G. How the hydrogen production from RES could change energy and fuel markets: A review of recent literature. *Int. J. Hydrogen Energy* **2019**, *44*, 11371–11384. [CrossRef]
3. Kłos, M.; Marchel, P.; Paska, J.; Bielas, R.; Błędzińska, M.; Michalski, L.; Zagrajek, K. Forecast and impact of electromobility development on the Polish Electric Power System. *E3S Web Conf.* **2019**, *84*, 01005. [CrossRef]
4. Piotrowski, P.; Baczyński, D.; Robak, S.; Kopyt, M.; Piekarz, M.; Polewaczyk, M. Comprehensive forecast of electromobility mid-term development in Poland and its impacts on power system demand. *Bull. Pol. Acad. Sci. Tech. Sci.* **2020**, *68*, 697–709. [CrossRef]
5. Huang, J.; Gurney, K.R. Impact of climate change on U.S. building energy demand: Sensitivity to spatiotemporal scales, balance point temperature, and population distribution. *Clim. Chang.* **2016**, *137*, 171–185. [CrossRef]
6. Han, Y.; Chen, W.; Li, Q. Energy Management Strategy Based on Multiple Operating States for a Photovoltaic/Fuel Cell/Energy Storage DC Microgrid. *Energies* **2017**, *10*, 136. [CrossRef]
7. Chatillon, O.; Graeber, D. Efficient Management of Wind Energy In-feed at a Large German TSO. In Proceedings of the 2008 IEEE Power and Energy Society General Meeting-Conversion and Delivery of Electrical Energy in the 21st Century, Pittsburgh, PA, USA, 20–24 July 2008; pp. 1–5.
8. Miettinen, J.; Holttinen, H.; Ammala, J.; Piironen, M. Wind Power Forecasting at Transmission System Operator's Control Room. In Proceedings of the 2015 IEEE Power & Energy Society General Meeting, Denver, CO, USA, 26 July 2015; pp. 1–5.
9. Yushuai, L.; Gao, W.; Gao, W.; Zhang, H.; Zhou, J.A. Distributed Double-Newton Descent Algorithm for Cooperative Energy Management of Multiple Energy Bodies in Energy Internet. *IEEE Trans. Ind. Inform.* **2020**, 1–11. [CrossRef]
10. Li, Y.; Gao, D.W.; Gao, W.; Zhang, H.; Zhou, J. Double-Mode Energy Management for Multi-Energy System via Distributed Dynamic Event-Triggered Newton-Raphson Algorithm. *IEEE Trans. Smart Grid* **2020**, *11*, 5339–5356. [CrossRef]
11. Zhang, N.; Sun, Q.; Yang, L.; Li, Y. Event-Triggered Distributed Hybrid Control Scheme for the Integrated Energy System. *IEEE Trans. Ind. Inform.* **2021**, 1. [CrossRef]

12. Zhang, N.; Sun, Q.; Wang, J.; Yang, L. Distributed Adaptive Dual Control via Consensus Algorithm in the Energy Internet. *IEEE Trans. Ind. Inform.* **2021**, *17*, 4848–4860. [CrossRef]

13. Hafiz, F.; Awal, M.A.; de Queiroz, A.R.; Husain, I. Real-Time Stochastic Optimization of Energy Storage Management Using Deep Learning-Based Forecasts for Residential PV Applications. *IEEE Trans. Ind. Appl.* **2020**, *56*, 2216–2226. [CrossRef]

14. Tipantuña, C.; Hesselbach, X. NFV/SDN Enabled Architecture for Efficient Adaptive Management of Renewable and Non-Renewable Energy. *IEEE Open J. Commun. Society* **2020**, *1*, 357–380. [CrossRef]

15. Gong, G.; An, X.; Mahato, N.K.; Sun, S.; Chen, S.; Wen, Y. Research on Short-Term Load Prediction Based on Seq2seq Model. *Energies* **2019**, *12*, 3199. [CrossRef]

16. Tian, C.; Ma, J.; Zhang, C.; Zhan, P. A Deep Neural Network Model for Short-Term Load Forecast Based on Long Short-Term Memory Network and Convolutional Neural Network. *Energies* **2018**, *11*, 3493. [CrossRef]

17. Amorim, A.J.; Abreu, T.A.; Tonelli-Neto, M.S.; Minussi, C.R. A new formulation of multinodal short-term load forecasting based on adaptive resonance theory with reverse training. *Electr. Power Syst. Res.* **2020**, *179*, 106096. [CrossRef]

18. Chu, Y.; Xu, P.; Li, M.; Chen, Z.; Chen, Z.; Chen, Y.; Li, W. Short-term metropolitan-scale electric load forecasting based on load decomposition and ensemble algorithms. *Energy Build.* **2020**, *225*, 110343. [CrossRef]

19. Eskandari, H.; Imani, M.; Moghaddam, M.P. Convolutional and recurrent neural network based model for short-term load forecasting. *Electr. Power Syst. Res.* **2021**, *195*, 107173. [CrossRef]

20. Jan, L.G.; Hsiao, K.; Xu, K.S.; Calder, J.; Iii, A.O.H. Multi-criteria Anomaly Detection using Pareto Depth Analysis. *arXiv* **2011**, arXiv:1110.3741.

21. Alamaniotis, M.; Bourbakis, N.; Tsoukalas, L.H. Very-short term forecasting of electricity price signals using a Pareto composition of kernel machines in smart power systems. In Proceedings of the 2015 IEEE Global Conference on Signal and Information Processing (GlobalSIP), Orlando, FL, USA, 14–16 December 2015; pp. 780–784.

22. Alamaniotis, M.; Ikonomopoulos, A.; Tsoukalas, L.H. A Pareto optimization approach of a Gaussian process en-semble for short-term load forecasting. In Proceedings of the 2011 16th International Conference on Intelligent System Applications to Power Systems, Hersonissos, Greece, 25–28 September 2011; pp. 1–6.

23. Feng, L.; He, J.; Kong, Q.; Guo, L. Application of multi-objective algorithm based on particle swarm optimization in electrical short-term load forecasting. In Proceedings of the 2006 International Conference on Power System Technology, Chongqing, China, 22–26 October 2006; Volume POWERCON2006, pp. 1–5.

24. Wan, C.; Niu, M.; Song, Y.; Xu, Z. Pareto Optimal Prediction Intervals of Electricity Price. *IEEE Trans. Power Syst.* **2017**, *32*, 817–819. [CrossRef]

25. Valgaev, O.; Kupzog, F.; Schmeck, H. Building power demand forecasting using K-nearest neighbours model–practical application in Smart City Demo Aspern project. *CIRED-Open Access Proc. J.* **2017**, *2017*, 1601–1604. [CrossRef]

26. Zagar, A.; Grolinger, K.; Capretz, M.; Seewald, L. Energy Cost Forecasting for Event Venues. In Proceedings of the 2015 IEEE Electrical Power and Energy Conference (EPEC), London, ON, Canada, 26–28 October 2015; pp. 220–226.

27. Zhang, R.; Xu, Y.; Ieee, M.; Dong, Z.Y.; Member, S.; Kong, W.; Wong, K.P. A Composite k-Nearest Neighbor Model for Day-Ahead Load Forecasting with Limited Temperature Forecasts. In Proceedings of the 2016 IEEE Power and Energy Society General Meeting (PESGM), Boston, MA, USA, 17–21 July 2016; pp. 1–5.

28. Asadi Majd, A.; Samet, H.; Ghanbari, T. K-NN based fault detection and classification methods for power transmission systems. *Prot. Control. Mod. Power Syst.* **2017**, *2*, 32. [CrossRef]

29. Khoa, N.M. Classification of Power Quality Disturbances Using Wavelet Transform and K-Nearest Neighbor Classifier. In Proceedings of the 2013 IEEE International Symposium on Industrial Electronics, Taipei, Taiwan, 28–31 May 2013; pp. 1–4.

30. Zhang, Z. Introduction to machine learning: K-nearest neighbors. *Ann. Transl. Med.* **2016**, *4*, 1–7. [CrossRef] [PubMed]

31. Gambier, A. MPC and PID control based on multi-objective optimization. In Proceedings of the American Control Conference, Seattle, WA, USA, 11–13 June 2008; Volume 26, pp. 4727–4732.

32. Liu, X.; Reynolds, A.C. Gradient-based multi-objective optimization with applications to waterflooding optimization. *Comput. Geosci.* **2015**, *20*, 677–693. [CrossRef]

33. Marler, R.T.; Arora, J.S. Survey of multi-objective optimization methods for engineering. *Struct. Multidiscip. Optim.* **2004**, *26*, 369–395. [CrossRef]

34. Carvalho, L.M. Modeling Wind Power Uncertainty in the Long-Term Operational Reserve Adequacy Assessment: A Comparative Analysis between the Naive and the ARIMA Forecasting Models. In Proceedings of the 2016 International Conference on Probabilistic Methods Applied to Power Systems (PMAPS), Beijing, China, 16–20 October 2016; pp. 1–6.

35. Hassan, S.; Khosravi, A.; Jaafar, J. Neural network ensemble: Evaluation of aggregation algorithms in electricity demand forecasting. In Proceedings of the 2013 International Joint Conference on Neural Networks (IJCNN), Dallas, TX, USA, 4–9 August 2013; pp. 1–6.

36. He, H.; Liu, T.; Chen, R.; Xiao, Y.; Yang, J. High Frequency Short-term Demand Forecasting Model for Distribution Power Grid Based on ARIMA. In Proceedings of the 2012 IEEE International Conference on Computer Science and Automation Engineering (CSAE), Zhangjiajie, China, 25–27 May 2012; Volume 3, pp. 293–297.

37. Hutama, A.H.; Zuhri, M.; Candra, C. Medium Term Power Load Forecasting for Java and Bali Power System using Artificial Neural Network and SARIMAX. In Proceedings of the 2018 5th International Conference on Data and Software Engineering (ICoDSE), Mataram, Indonesia, 7–8 November 2018; pp. 1–6.
38. Mitkov, A.; Noorzad, N.; Gabrovska-evstatieva, K.; Mihailov, N. Forecasting the Energy Consumption in Afghani-stan with the ARIMA Model. In Proceedings of the 2019 16th Conference on Electrical Machines, Drives and Power Systems (ELMA), Varna, Bulgaria, 6–8 June 2019; pp. 1–4.