

doi:10.15199/48.2017.04.16

## Prognozowanie miesięcznego zapotrzebowania na energię elektryczną metodą $k$ najbliższych sąsiadów

**Streszczenie.** Przedstawiono model prognostyczny oparty na metodzie  $k$  najbliższych sąsiadów do prognozowania miesięcznego zapotrzebowania na energię elektryczną. Model wykorzystuje analogie pomiędzy fragmentami szeregów czasowych reprezentowanymi przez ich obrazy. Obrazy zapewniają ujednoczenie danych wejściowych i wyjściowych, odfiltrowanie trendu i uproszczenie modelowanej zależności. W części eksperymentalnej model przetestowano w prognozach dla wybranych państw europejskich.

**Abstract.** A forecasting model based on the  $k$  nearest neighbor method for forecasting monthly electricity demand is presented. The model uses analogies between fragments of time series represented by their patterns. Patterns ensure unification of input and output data, filtering out the trend and simplification of the modeled relationship. In the experimental part of the work the model was tested in forecasting for selected European countries. (*Forecasting monthly electricity demand using  $k$  nearest neighbor method*).

**Słowa kluczowe:** średnioterminowe prognozowanie zapotrzebowania na energię, prognozowanie oparte na podobieństwie obrazów, metoda  $k$  najbliższych sąsiadów, obrazy sekwencji szeregów czasowych.

**Keywords:** mid-term energy demand forecasting, pattern similarity-based forecasting,  $k$  nearest neighbor method, patterns of time series sequences.

### Wstęp

Prognozowanie miesięcznego zapotrzebowania na energię elektryczną odgrywa ważną rolę w planowaniu pracy systemu elektroenergetycznego. Potrzeba sporządzania prognoz średnioterminowych wynika z trzech przesłanek [1]:

- techniczno-eksploatacyjnych – ustalanie harmonogramów remontów, planowanie poziomu produkcji i zakupów paliwa, planowanie inwestycji sieciowych,
- ekonomicznych – negocjowanie kontraktów pomiędzy przedsiębiorstwami energetycznymi, zawieranie umów z odbiorcami,
- prawnych – obowiązek sporządzania prognoz narzucony przez ustawę Prawo Energetyczne.

W prognozowaniu średnioterminowym zapotrzebowania na energię elektryczną wykorzystuje się metody klasyczne, takie jak ARIMA, czy regresja liniowa [2] oraz metody inteligencji obliczeniowej, np. sieci neuronowe [3]. Proponowane w tej pracy metody oparte na podobieństwie (*similarity-based methods*) stanowią podstawę wielu metod uczenia maszynowego i rozpoznawania obrazów [4]. Podstawowe problemy rozwiązywane za pomocą tych metod to klasyfikacja i grupowanie danych, regresja, uzupełnianie brakujących wartości i tworzenie pamięci skojarzeniowych. U podstaw modeli opartych na podobieństwie leży założenie: nieznane cechy obiektu można przewidzieć na podstawie zbioru obiektów podobnych, dla których te cechy są znane. Obiekty widziane są poprzez swoje obrazy, tj. zbiory mierzalnych parametrów, atrybutów, własności. Definicja obrazów, czyli sposób reprezentacji danych, jest tutaj kluczowa. Innym kluczowym elementem jest definicja miary podobieństwa obrazów.

W dziedzinie prognozowania szeregów czasowych podejście oparte na podobieństwie wykorzystuje podobieństwa pomiędzy fragmentami szeregów czasowych (podobieństwa kształtu). Może to stanowić podstawę wyznaczania prognozy bez uciekania się do liczbowego opisu związków przyczynowo-skutkowych i jest szczególnie atrakcyjne, gdy zmienna przechodzi wiele faz rozwoju, które powtarzają się cyklicznie.

W odniesieniu do szeregów czasowych, pod pojęciem obrazu rozumie się wektor o składowych będących funkcjami rzeczywistych wyrazów szeregu. Definiuje się obrazy wejściowe  $\mathbf{x} = [x_1 \ x_2 \ \dots \ x_n]^T$  i obrazy prognoz (wyjściowe)  $\mathbf{y} = [y_1 \ y_2 \ \dots \ y_m]^T$ . Obrazy  $\mathbf{x}$  mogą być

definiowane inaczej niż obrazy  $\mathbf{y}$ . Obrazy łączy się w pary  $(\mathbf{x}_i, \mathbf{y}_i)$ , gdzie  $\mathbf{y}_i$  jest obrazem fragmentu szeregu czasowego następującego po fragmencie reprezentowanym przez obraz  $\mathbf{x}_i$ , a odległość w czasie pomiędzy tymi fragmentami równa jest horyzontowi prognozy  $\tau$ . Modele prognostyczne wykorzystujące podobieństwo obrazów opierają się na założeniu [5]: Jeśli obrazy wejściowe fragmentów szeregu czasowego  $\mathbf{x}_a$  i  $\mathbf{x}_b$  są do siebie podobne, to sparowane z nimi obrazy prognoz  $\mathbf{y}_a$  i  $\mathbf{y}_b$  również są do siebie podobne. Założenie to pozwala prognozować obraz  $\mathbf{y}_b$  na podstawie obrazów  $\mathbf{x}_a$ ,  $\mathbf{x}_b$  i  $\mathbf{y}_a$ , które wyznacza się z historii procesu.

### Obrazy fragmentów szeregów czasowych

Definicje funkcji przekształcających wyrazy szeregu czasowego w obrazy  $\mathbf{x}$  i  $\mathbf{y}$  są zależne od specyfiki szeregu (składowych periodycznych, trendu) oraz okresu i horyzontu prognozy. Funkcje te należy zdefiniować tak, aby obrazy przносиły jak najwięcej informacji o procesie, a jakość modelu prognostycznego była maksymalna. Ponadto funkcje definiujące obrazy  $\mathbf{y}$  powinny zapewniać możliwość dekodowania, czyli wyznaczania prognoz rzeczywistych wyrazów szeregu.

Rozważmy zadanie prognozy  $m$  wyrazów szeregu czasowego miesięcznego zapotrzebowania na energię elektryczną. Oznaczmy prognozowany fragment przez  $Y_i = \{E_{i+1} \ E_{i+2} \ \dots \ E_{i+m}\}$ . Fragment szeregu o długości  $n$  poprzedzający fragment prognozowany oznaczmy przez  $X_i = \{E_{i-n+1} \ E_{i-n+2} \ \dots \ E_i\}$ . Obraz wejściowy  $\mathbf{x}_i = [x_{i,1} \ x_{i,2} \ \dots \ x_{i,n}]^T$  reprezentuje fragment  $X_i$ . Składowe tego wektora są funkcjami elementów sekwencji  $X_i$ . Poniżej podano kilka definicji tych funkcji:

$$\begin{aligned} (1) \quad & x_{i,t} = E_{i-n+t} \\ (2) \quad & x_{i,t} = E_{i-n+t} - \bar{E}_i \\ (3) \quad & x_{i,t} = \frac{E_{i-n+t}}{\bar{E}_i} \\ (4) \quad & x_{i,t} = \frac{E_{i-n+t} - \bar{E}_i}{D_i} \end{aligned}$$

gdzie  $\bar{E}_i$  jest wartością średnią elementów sekwencji  $X_i$ ,

a  $D_i = \sqrt{\sum_{j=1}^n (E_{i-n+j} - \bar{E}_i)^2}$  jest miarą ich dyspersji.

Obrazy zdefiniowane za pomocą (1) kopiują elementy sekwencji  $X_i$  bez przetworzenia. Składowe obrazów (2) to różnice energii zapotrzebowanej w danym miesiącu i energii średniej w okresie objętym sekwencją  $X_i$ . Iloraz tych dwóch wielkości wyrażają składowe zdefiniowane wzorem (3). Obraz zdefiniowany za pomocą (4) jest unormowanym wektorem  $[E_{i-n+1} E_{i-n+2} \dots E_i]^T$ . Jego długość jest jednostkowa, wartość średnia jego składowych jest zerowa, a ich wariancja jednakowa dla wszystkich wektorów  $\mathbf{x}_i$ .

Obraz wyjściowy  $\mathbf{y}_i = [y_{i,1} y_{i,2} \dots y_{i,m}]^T$  reprezentuje fragment  $Y_i$ . Składowe tego wektora są funkcjami elementów sekwencji  $Y_i$ . Funkcje te zdefiniowane są podobnie jak dla obrazów  $\mathbf{x}$ :

$$(5) \quad y_{i,t} = E_{i+t}$$

$$(6) \quad y_{i,t} = E_{i+t} - \bar{E}_i$$

$$(7) \quad y_{i,t} = \frac{E_{i+t}}{\bar{E}_i}$$

$$(8) \quad y_{i,t} = \frac{E_{i+t} - \bar{E}_i}{D_i}$$

W powyższych wzorach  $\bar{E}_i$  i  $D_i$  to wielkości opisujące sekwencję  $X_i$ , a nie, jak można by się spodziewać,  $Y_i$ . Wielkości te w momencie sporządzania prognozy są znane i umożliwiają wyznaczenie prognozy zapotrzebowania na podstawie prognozy obrazu  $\mathbf{y}_i$ , zwróconej przez model prognostyczny. Używa się do tego przekształconych wzorów (5)-(8). Na przykład w przypadku definicji obrazu  $\mathbf{y}$  za pomocą (8), prognozę zapotrzebowania wyznaczamy ze wzoru:

$$(9) \quad \hat{E}_{i+t} = \hat{y}_{i,t} D_i + \bar{E}_i$$

Obrazy  $\mathbf{x}_i$  i  $\mathbf{y}_i$ , które reprezentują dwie sekwencje szeregu czasowego, poprzedzającą prognozę i prognozowaną, łączy się w parę  $(\mathbf{x}_i, \mathbf{y}_i)$ . Zbiór tych par dla historycznego szeregu czasowego wykorzystywany jest do tworzenia modelu prognostycznego (estymacji parametrów, uczenia).

### Model prognostyczny $k$ najbliższych sąsiadów

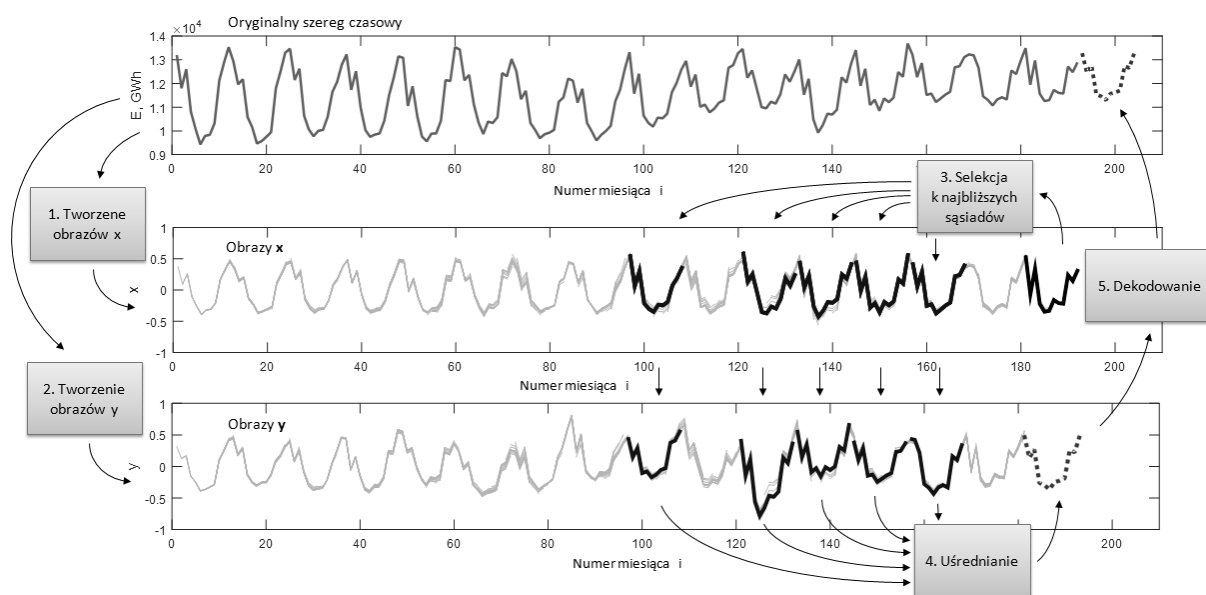
Metoda  $k$  najbliższych sąsiadów ( $k$ -NS) należy do jednej z najpopularniejszych metod rozpoznawania obrazów z nauczycielem. Pomimo swojej prostoty z powodzeniem konkuruje z bardziej wyrafinowanymi metodami klasyfikacji danych i jest nawet zaliczana do metod najskuteczniejszych [6]. Ograniczenie obszaru rozpoznania do otoczenia próbki wejściowej, które reprezentuje zbiór  $k$  najbliższych sąsiadów, wprowadza lokalny charakter modelu. Warto odnotować, że metoda  $k$ -NS nie wymaga przyjęcia żadnych założeń dotyczących dziedziny i reprezentacji obrazów poza tym, że jest dla nich określona pewna miara podobieństwa lub odległości [7]. Również żadne założenia o przeciwdziedzinie odwzorowania docelowego nie są potrzebne. Jeśli przeciwdziedziną jest zbiór liczb rzeczywistych, to algorytm rozwiązuje zadanie uczenia się aproksymacji. Gdy przeciwdziedziną jest pewnym zbiorem kategorii, wówczas  $k$ -NS służy do klasyfikowania przykładów.

Proponowany model prognostyczny należy do klasy nieparametrycznych modeli regresyjnych. Estymator funkcji regresji  $m(\cdot)$  typu  $k$ -NS zdefiniowany jest jako średnia zmiennych objaśnianych (obrazów  $\mathbf{y}$ ) w sąsiedztwie zmiennej objaśniającej  $\mathbf{x}$ . Sąsiedztwo zdeterminowane jest wartością parametru  $k$  i należy do niego  $k$  najbliższych (w sensie przyjętej metryki) sąsiadów zmiennej  $\mathbf{x}$ . Estymator funkcji regresji ma postać:

$$(10) \quad \hat{m}(\mathbf{x}) = \frac{1}{k} \sum_{i \in \Theta_k(\mathbf{x})} \mathbf{y}_i$$

gdzie  $\Theta_k(\mathbf{x})$  jest zbiorem indeksów  $k$  obrazów ze zbioru uczącego, które należą do najbliższych sąsiadów  $\mathbf{x}$ .

Zbiór uczący obejmuje pary  $(\mathbf{x}_i, \mathbf{y}_i)$  historyczne w stosunku do prognozowanej sekwencji szeregu czasowego, tzn. takie, dla których  $i = n, n+1, \dots, i^*-m$ , gdzie  $i^*$  jest indeksem ostatniego miesiąca przed prognozowaną sekwencją (zadanie prognostyczne polega na sporządzeniu prognoz dla kolejnych miesięcy  $i^*+1, i^*+2, \dots, i^*+m$ ).



Rys.1. Idea modelu prognostycznego  $k$  najbliższych sąsiadów

Zbiór najbliższych sąsiadów  $\Theta_k(x)$  aktualnego obrazu wejściowego  $x_{i^*}$  reprezentującego sekwencję  $\{E_{i^*-n+1} E_{i^*-n+2} \dots E_{i^*}\}$  znajdujemy przesuwając się po kolejnych obrazach  $x_i$  ze zbioru uczącego. Obrazy te reprezentują kolejne sekwencje  $\{E_{i-n+1} E_{i-n+2} \dots E_i\}$ ,  $i = n, n+1, \dots, i^*-m$ . Obrazy  $y_i$  sparowane z najbliższymi sąsiadami obrazu wejściowego są uśredniane (10) dając prognozę obrazu  $y$ , który reprezentuje prognozowaną sekwencję  $\{E_{i^*+1} E_{i^*+2} \dots E_{i^*+m}\}$ . Elementy tej sekwencji obliczamy z przekształconych wzorów (5)-(8), zależnie od przyjętej definicji obrazu  $y$ . Parametrami modelu  $k$ -NS są: liczba sąsiadów  $k$  oraz długość obrazu  $x - n$ . Parametry te dobiera się dla danego zadania prognostycznego.

Ideę działania modelu  $k$ -NS w prognozowaniu szeregów czasowych miesięcznego zapotrzebowania na energię elektryczną zobrazowano na rysunku 1. W pierwszej fazie szereg czasowy zostaje przekształcony w obrazy (kroki 1 i 2). Dla obrazu  $x_{i^*}$  znajduje się  $k$  najbliższych sąsiadów wśród obrazów historycznych (krok 3). Sparowane z nimi obrazy  $y$  są uśredniane, dając prognozę obrazu  $y_{i^*}$  (krok 4). Prognozowany obraz  $y_{i^*}$  jest dekodowany za pomocą jednego ze wzorów (5)-(8) po przekształceniu (krok 5). W wyniku otrzymujemy prognozę miesięcznego zapotrzebowania na energię elektryczną dla kolejnych miesięcy  $i^*+1, i^*+2, \dots, i^*+m$ .

### Badania symulacyjne

Proponowany model prognostyczny użyto do prognozowania miesięcznego zapotrzebowania na energię elektryczną dla czterech państw europejskich: Polski (PL), Niemiec (DE), Francji (FR) i Hiszpanii (ES). Dane pozyskano z repozytorium ENTSO-E ([www.entsoe.eu](http://www.entsoe.eu)). Dane dla PL obejmują lata 1998-2014, a dla pozostałych państw lata 1991-2014. Prognozy wykonano dla danych z roku 2014, wykorzystując dane z lat poprzednich do uczenia modelu. Prognozy sporządzano w dwóch wariantach:

A - dla wszystkich 12 miesięcy roku 2014 jednocześnie ( $i^*$  to indeks grudnia 2013,  $m = 12$ ),

B - indywidualnie dla kolejnych 12 miesięcy roku 2014 (utworzono 12 modeli dla  $i^*$  odpowiadającego indeksom: grudnia 2013, stycznia 2014, ..., listopada 2014,  $m = 1$ ).

Parametry modelu: liczbę najbliższych sąsiadów  $k$  oraz długość obrazu  $x - n$ , dobierano w procedurze przeglądu zupełnego, zmieniając  $k$  w granicach od 1 do 20, a  $n$  w granicach od 3 do 24. Minimalizowano błąd prognozy estymowany na zbiorze uczącym metodą minus jednego elementu (*leave-one-out cross-validation*). Wyniki: optymalne wartości parametrów, minimalny błąd walidacyjny  $MAPE_{wal}$  (na zbiorze uczącym) oraz błąd testowy  $MAPE_{test}$  (dla 2014 roku) zamieszczono w tabelach 1-8. We wszystkich przypadkach, z wyjątkiem prognoz dla Francji w wariantach A, najlepsze wyniki uzyskano dla obrazów (4)-(8). Błędy prognoz w rozbiu na poszczególne miesiące pokazano na rysunku 2. Jak widać, nie zawsze model B, w którym prognozuje się jeden krok do przodu, oferuje lepsze prognozy od modelu A, w którym horyzont prognozy wynosi 12 miesięcy.

Na rysunku 3 pokazano konstrukcję prognoz dla PL w wariantach A, przy różnych definicjach obrazów. Obrazy konstrukcyjne sparowane z najbliższymi sąsiadami obrazu wejściowego  $x_{i^*}$  po uśrednieniu formują prognozę obrazu  $y_{i^*}$  (linia przerywana). Jak widać, prognoza obciążona jest najmniejszym błędem, gdy obrazy definiuje się za pomocą wzorów (4) i (8).

Proponowany model prognostyczny jest konkurencyjny w stosunku do modeli opisanych w literaturze. Autorzy [8] do prognozowania miesięcznego zapotrzebowania na

energię elektryczną w Hiszpanii użyli sieci neuronowych. Zadanie zdekomponowano na prognozę trendu i sześciu składowych fluktuacji, które wyodrębniono za pomocą filtracji cyfrowej. Model złożony z siedmiu sieci neuronowych uzyskał błędy na poziomie 3,15%, a użyty dla porównania model ARIMA wykazywał jeszcze większe błędy. Model  $k$ -NS opisany w niniejszej pracy generował dla Hiszpanii błędy na poziomie 1,90%. W prognozowaniu z horyzontem jednomiesięcznym autorzy [3] uzyskali dla Hiszpanii błąd 1,89% używając modeli neuronowych i 4,19% używając modelu ARIMA. Również w tym przypadku model  $k$ -NS wykazuje mniejszy błąd – 1,66%.

Tabela 1. Wyniki dla PL wariant A

Obrazy	$n$	$k$	$MAPE_{wal}$ , %	$MAPE_{test}$ , %
(1)-(5)	17	5	3,11	3,05
(2)-(6)	18	4	3,02	2,59
(3)-(7)	21	3	3,07	3,32
(4)-(8)	9	8	2,90	1,65

Tabela 2. Wyniki dla DE wariant A

Obrazy	$n$	$k$	$MAPE_{wal}$ , %	$MAPE_{test}$ , %
(1)-(5)	8	2	3,29	3,63
(2)-(6)	10	12	3,17	4,08
(3)-(7)	11	11	3,16	3,68
(4)-(8)	9	13	3,23	2,03

Tabela 3. Wyniki dla FR wariant A

Obrazy	$n$	$k$	$MAPE_{wal}$ , %	$MAPE_{test}$ , %
(1)-(5)	21	6	3,27	4,96
(2)-(6)	24	4	3,43	5,22
(3)-(7)	12	5	3,33	5,65
(4)-(8)	13	17	3,35	7,71

Tabela 4. Wyniki dla ES wariant A

Obrazy	$n$	$k$	$MAPE_{wal}$ , %	$MAPE_{test}$ , %
(1)-(5)	24	6	2,82	2,13
(2)-(6)	9	8	2,88	2,45
(3)-(7)	22	7	2,85	3,37
(4)-(8)	22	5	2,89	1,90

Tabela 5. Wyniki dla PL wariant B

Obrazy	$n$	$k$	$MAPE_{wal}$ , %	$MAPE_{test}$ , %
(1)-(5)	10	4	2,41	2,31
(2)-(6)	10	6	2,27	1,89
(3)-(7)	24	3	2,25	2,05
(4)-(8)	10	8	2,03	1,48

Tabela 6. Wyniki dla DE wariant B

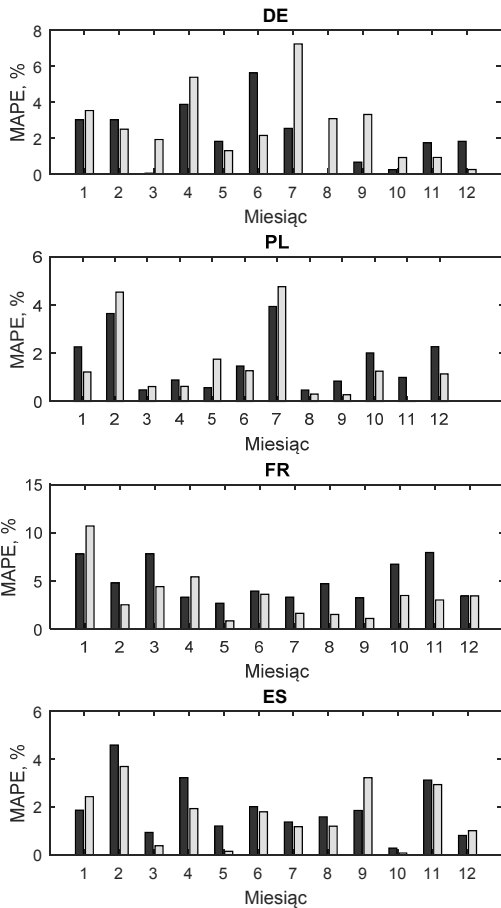
Obrazy	$n$	$k$	$MAPE_{wal}$ , %	$MAPE_{test}$ , %
(1)-(5)	8	3	2,54	5,89
(2)-(6)	10	7	2,51	4,18
(3)-(7)	11	10	2,53	3,30
(4)-(8)	9	6	2,48	2,70

Tabela 7. Wyniki dla FR wariant B

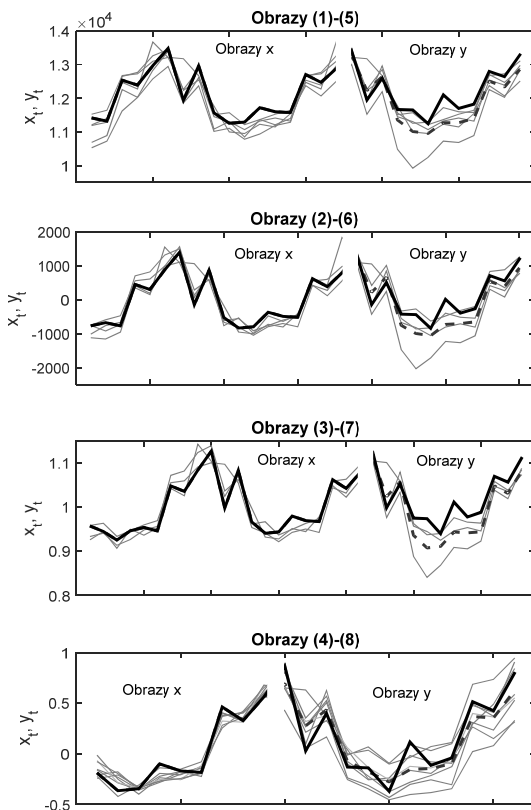
Obrazy	$n$	$k$	$MAPE_{wal}$ , %	$MAPE_{test}$ , %
(1)-(5)	22	6	3,13	4,22
(2)-(6)	14	7	3,02	3,95
(3)-(7)	14	10	2,98	4,03
(4)-(8)	14	6	2,85	3,46

Tabela 8. Wyniki dla ES wariant B

Obrazy	$n$	$k$	$MAPE_{wal}$ , %	$MAPE_{test}$ , %
(1)-(5)	24	7	2,79	2,20
(2)-(6)	13	10	2,40	2,07
(3)-(7)	14	9	2,33	2,64
(4)-(8)	18	6	2,38	1,66



Rys.2. Błędy prognoz dla poszczególnych miesięcy 2014 roku. Wariant A - ciemne słupki, wariant B - jasne słupki.



Rys.3. Konstrukcja prognoz dla PL w wariantcie A, przy różnych definicjach obrazów. Grubymi liniami narysowano obrazy  $x_t$  i  $y_t$ , cienkimi –  $k$  najbliższych sąsiadów  $x_{t^*}$  i sparowane z nimi obrazy konstrukcyjne  $y_{t^*}$ , linią przerywaną – prognozę obrazu  $y_t^*$

## Wnioski

W pracy przedstawiono model prognostyczny oparty na metodzie  $k$  najbliższych sąsiadów do prognozowania miesięcznego zapotrzebowania na energię elektryczną. Kluczowym elementem modelu jest sposób reprezentacji szeregów czasowych w postaci obrazów. Zależnie od definicji obrazu pozwalają zredukować problem niestacjonarności szeregu czasowego oraz odfiltrować trend. Zaletą modelu jest jego prosta i zrozumiała zasada działania oraz niewielka liczba parametrów do estymacji.

Lepsze wyniki działania modelu obserwuje się dla bardziej regularnych szeregów czasowych (o niskiej zawartości zakłóceń losowych) i przy stabilnej relacji wiążącej obrazu wejściowe z wyjściowymi. Stabilność tę osłabia nieliniowy trend i zmienna w czasie wariancja szeregu.

Wyniki badań symulacyjnych wskazują na dużą dokładność proponowanego modelu. Błędy na poziomie 2% należy uznać za dobry rezultat na tle doniesień literaturowych. W dalszej pracy planuje się wprowadzić wagi w modelu regresji uzależniające wpływ obrazów konstrukcyjnych od odległości pomiędzy obrazem wejściowym a jego najbliższymi sąsiadami.

**Autorzy:** dr hab. inż. Grzegorz Dudek, Politechnika Częstochowska, Instytut Informatyki, al. Armii Krajowej 17, 42-200 Częstochowa, E-mail: [Dudek@el.pcz.czyst.pl](mailto:Dudek@el.pcz.czyst.pl); Paweł Pełka, student kierunku Informatyka na Wydziale Elektrycznym Politechniki Częstochowskiej, E-mail: [Pavelle50@gmail.com](mailto:Pavelle50@gmail.com).

## LITERATURA

- [1] Piotrowski P., Prognozowanie w elektroenergetyce w różnych horyzontach czasowych. Oficyna Wydawnicza Politechniki Warszawskiej, Warszawa (2013)
- [2] Hor C.L., Watson S., Majithia S., Analyzing the Impact of Weather Variables on Monthly Electricity Demand. *IEEE Transactions On Power Systems*, 20 (2005), 2078-2085
- [3] González-Romera E., Jaramillo-Morán M., Carmona-Fernández D., Monthly Electric Energy Demand Forecasting Based on Trend Extraction. *IEEE Transactions On Power Systems*, 21(2006), 1946-1953
- [4] Duch W., Similarity-based Methods: A General Framework for Classification, Approximation and Association. *Control and Cybernetics*, 29 (2000), 937-968
- [5] Dudek G., Systemy uczące się oparte na podobieństwie obrazów do prognozowania szeregów czasowych obciążeń elektroenergetycznych. Akademicka Oficyna Wydawnicza EXIT, Warszawa (2012)
- [6] Michie D., Spiegelhalter D.J., Taylor C.C., Machine Learning, Neural and Statistical Classification. Ellis Horwood (1994)
- [7] Cichosz P., Systemy uczące się. WNT (2007)
- [8] Jaramillo-Morán M., González-Romera E., Carmona-Fernández D., Monthly electric demand forecasting with neural filters. *Electrical Power and Energy Systems*, 49 (2013), 253-263