

## Idea zastosowania zautomatyzowanych metod doboru zmiennych objaśniających dla celów optymalizacji kosztów zakupu danych wejściowych do prognozowania dobowego profilu obciążenia KSE

**Streszczenie.** Artykuł prezentuje możliwość skorzystania z metod statystycznych automatyzujących dobór zmiennych objaśniających na przykładzie dobowego obciążenia Krajowego Systemu Elektroenergetycznego. Automatyzacja pozwala na optymalizację kosztów zakupu prognoz wejściowych dzięki minimalizacji ich liczby, a uzyskane wyniki pozwalają dodatkowo na zmniejszenie nakładów pracy związanych z wyborem parametrów wejściowych (zmiennych objaśniających) na potrzeby późniejszego opracowywania prognoz dobowego obciążenia KSE.

**Abstract.** The paper presents the possibility of using statistical methods to automate the selection of explanatory variables to balance the daily load of the National Power System (NPS). With automation, the cost of input forecast purchase may be optimized by minimizing their number, and the results also allow for a reduction in the effort required to select input parameters (explanatory variables) for later forecasting of NPS daily loads. (The Concept of Explanatory Variables Automated Selection Methods Usage for Input Data Purchase Cost Optimization in the Scope of National Power System Daily Load Profile Forecasting).

**Słowa kluczowe:** zapotrzebowanie na moc, zmienne objaśniające, zmienne meteorologiczne, Data Mining.

**Keywords:** Power Demand, Explanatory Variables, Meteorological Variables, Data Mining.

### Wprowadzenie

Bezpieczne funkcjonowanie systemu elektroenergetycznego (SEE) wymaga przygotowania m.in. prognoz spodziewanego obciążenia tego systemu [1]. Operator sieci przesyłowej narażony jest m.in. na ryzyko znaczącego odchylenia prognozy od rzeczywistego obciążenia Krajowego Systemu Elektroenergetycznego (KSE) [2]. Optymalny wybór zmiennych objaśniających oraz skutecznej metody prognostycznej jest kluczowym etapem w procesie budowy modelu prognostycznego. Im większy stopień powiązania dobranych zmiennych objaśniających (ZO) ze zmienną objaśnianą (prognozowaną) tym większa dokładność opisu tej ostatniej [3]. Staranny dobór ZO jest kluczem do zbudowania skutecznego modelu prognostycznego. Dobór ten wymaga jednak często dużych nakładów pracy (czasochłonność), a pozyskiwanie pełnych i wiarygodnych danych historycznych, charakteryzujących się wysoką rozdzielczością, może być trudne i wiąże się ze sporymi nakładami finansowymi [4]. Idealną sytuacją, z punktu widzenia przygotowania prognozy dla operatora SEE, jest minimalizacja nakładów pracy oraz (bardzo często) minimalizacja trudności i kosztów pozyskiwania danych stanowiących zestaw ZO.

Minimalizacja nakładów pracy, obsługi danych w postaci ZO i kosztów ich pozyskania możliwa jest dzięki:

- zautomatyzowanym procesom typowania i rankingowania najkorzystniejszych zmiennych dla danego procesu (z zastosowaniem pakietów statystycznych);
- wyborowi najlepszych ZO z danej grupy pozyskiwanych danych – dzięki ich testowaniu przez różne metody statystyczne (tab. 1).

### Idea automatyzacji doboru zmiennych objaśniających vs koszty ich zakupu i obsługi

Wybór ZO w procesie budowy modelu prognostycznego w dużym stopniu determinuje poprawność uzyskiwanych wyników z tego modelu. Zwykle dobór ZO zależy od wiedzy i doświadczenia prognosty. Do ewentualnej zmiany zestawu ZO dochodzi zwłaszcza w sytuacji, gdy opracowany wcześniej model nie daje satysfakcjonujących wyników [5]. Jednak zwykle ponowny wybór ZO jest czasochłonny

i kosztowny, ponieważ wymaga przeprowadzenia wielu działań weryfikujących poprawność uzyskiwanych wyników dla zmienionego zestawu ZO. Stąd w niniejszej publikacji przedstawiono wyniki wybranych symulacji, które dotyczyły próby wykorzystania zautomatyzowanej metody doboru najlepszych ZO dla zadanego zadania prognostycznego. Jako najlepsze ZO postrzegane są zmienne, które jak najprecyzyjniej opisują zmienną objaśnianą. Najlepsze ZO typowano w powiązaniu z danymi historycznymi szczytowej wartości 15-minutowej mocy zapotrzebowania w KSE w ciągu doby, za pomocą wybranych metod statystycznych. W kolejnym kroku przeliczono potencjalne koszty ich całorocznego zakupu dla rozdzielczości godzinowej obciążenia dobowego. Pozwoliło to na ocenę możliwości ograniczenia liczby parametrów, dla których kupowane mogą być prognozy z wyprzedzeniem tygodniowym i z rozdzielczością godzinową. Idealnym stanem jest sytuacja, w której za pomocą jednej ZO możliwe jest opracowanie prognozy analizowanego parametru. Na obciążenie KSE wpływa jednakże więcej niż jeden parametr, dlatego też wydaje się, że każde ograniczenie nakładów finansowych i pracy na przygotowanie prognozy może być korzystne dla prognosty i podmiotu, na rzecz którego ten prognosta pracuje.

Zaprezentowane podejście do optymalizacji liczby rozpatrywanych ZO modelu prognostycznego może być przydatne dla trzech grup użytkowników:

- doświadczonych badaczy prognostów, których modele (metody) prognostyczne uodporniły się na obserwowane w przyrodzie zmiany (reagują z opóźnieniem na dynamiczne zmiany ich zachowań);
- początkujących badaczy prognostów, którzy posiadają podstawową wiedzę z zakresu przygotowywania danych na potrzeby prognoz i prognozowania, którzy działają przy ograniczonych zasobach ludzkich i czasowych (np. klastry energii);
- osób zarządzających kosztami pozyskiwania ZO do opracowywania prognoz.

Do symulacji jako zmienną objaśnianą (prognozowaną) przyjęto szczytową wartość 15-minutowej mocy zapotrzebowania w KSE w ciągu doby [6]. Model

prognostyczny zasilono na wejściu danymi historycznymi o kształtowaniu się poszczególnych parametrów. Po wytypowaniu liczby optymalnych ZO zasymulowano roczne koszty zakupu prognoz wytypowanych parametrów. Jako najkorzystniejsze rozwiązanie zadanego problemu założono minimalizację nakładów pracy i nakładów finansowych związanych z pozyskiwaniem zmiennych wejściowych do procesu opracowania prognoz. Testowanie skuteczności typowania ZO przez metody M1-M5 przeprowadzono przy wykorzystaniu metod przedstawionych w tabeli 1.

Tabela 1. Metody statystyczne wykorzystane do badań testowych

Metoda	Kod
Wielozmienna regresja adaptacyjna z użyciem funkcji sklepanych (MARSplines)	S1
Ogólne modele drzew klasyfikacyjnych i regresyjnych (C&RT wersja standardowa)	S2
Ogólne modele drzew klasyfikacyjnych i regresyjnych (C&RT wersja z układami)	S3
Automatyczny detektor interakcji za pomocą chi-kwadrat (CHAID wersja standardowa)	S4
Automatyczny detektor interakcji za pomocą chi-kwadrat (CHAID wersja z układami)	S5
Drzewa interakcyjne regresyjne (C&RT wersja interaktywna)	S6
Drzewa CHAID regresyjne (CHAID wersja interaktywna)	S7
Drzewa interakcyjne regresyjne (CHAID wersja wyczerpująca)	S8
Uogólnione modele addytywne z użyciem funkcji wiążącej identycznościowej	S9
Uogólnione modele addytywne z użyciem funkcji wiążącej logarytmicznej	S10
Regresja wieloraka	S11
Ogólne modele liniowe i nieliniowe	S12
Ogólne modele regresji	S13
Modele najmniejszych kwadratów (cząstkowa)	S14
Sztuczne sieci neuronowe	S15

### Zmienne objaśniające meteorologiczne

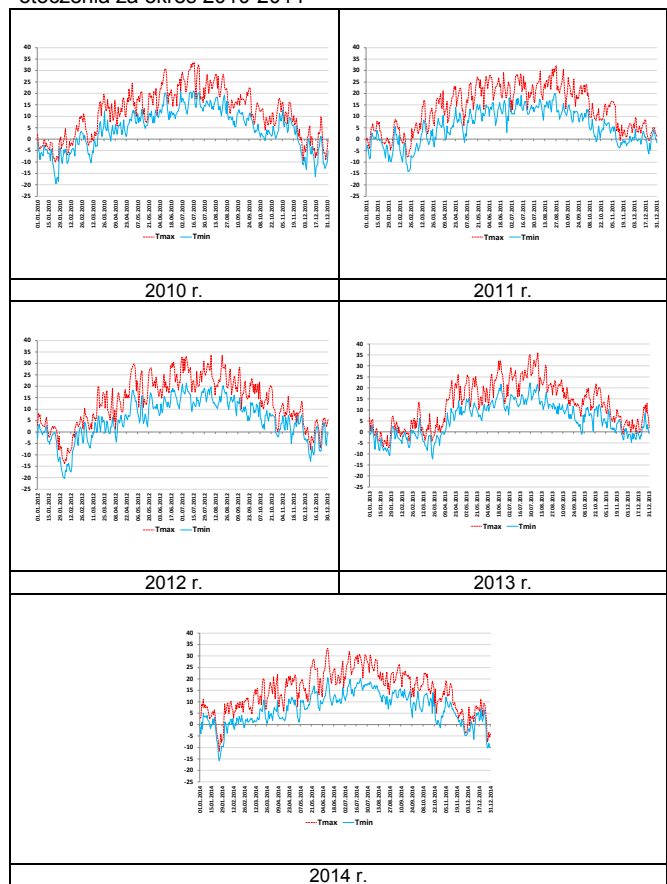
Jedne z najwyższych kosztów generuje pozyskanie ZO będących parametrami meteorologicznymi (dane historyczne oraz prognozy). Parametry meteorologiczne w sposób znaczący wpływają na obciążenie KSE (tab. 2) [7, 8]. Jako zbiór ZO meteorologicznych wybrano pomiary meteorologiczne z jednej ze stacji meteorologicznych na środkowym południu Polski. Wyniki porównań temperatur z kilkoma większymi miejscami pomiarów w kraju wskazują, że stację tę można przyjąć zgrubnie (ok.  $\pm 1^\circ\text{C}$ ) jako odzwierciedlającą średnią pomiarów z całego kraju.

Pierwszy zestaw zmiennych składał się z 14 następujących parametrów meteorologicznych, które mogą być przedmiotem zakupu od specjalistycznych podmiotów: temperatura otoczenia maksymalna ( $^\circ\text{C}$  - ZO6); temperatura otoczenia minimalna ( $^\circ\text{C}$  - ZO7); opady deszczu (mm - ZO8); prędkość wiatru średnia (km/h - ZO9); prędkość wiatru średnia ekspercka (km/h - ZO10); prędkość wiatru maksymalna (km/h - ZO11); ciśnienie atmosferyczne (hPa - ZO12); liczba stopniogrzejnych ( $^\circ\text{C}$  - ZO13); liczba stopniogrzejnych ( $^\circ\text{C}$  - ZO14); liczba godzin słonecznych (- - ZO15); energia słoneczna ( $\text{W/m}^2$  - ZO16); poziom promieniowania UV (- - ZO16); temperatura punktu rosy ( $^\circ\text{C}$  - ZO17); temperatura mokrego termometru ( $^\circ\text{C}$  - ZO18).

Do powyższego zestawu dołączono dwie ZO. Pierwsza z nich (ZO2) zawiera – w postaci zakodowanej – informację o dacie dokonanego pomiaru (rok/miesiąc/dzień) z rozróżnieniem kolejnych dni tygodnia oraz z uwzględnieniem podziału na dzień nieświęteczny i święteczny. Druga z dodatkowych zmiennych (ZO3) ma postać czasową niezakodowaną, w której zawarto informację o czasie (z rozdzielczością 15-minutową) wystąpienia szczytowej wartości mocy 15-minutowej

obciążenia KSE oraz o generacji wiatrowej w KSE [6] w każdej dobie analizowanego szeregu czasowego.

Tabela 2. Maksymalne i minimalne wartości dobowe temperatury otoczenia za okres 2010-2014



### Zmienne objaśniające pozostałe

Poza ZO meteorologicznymi, w procesie budowy modelu prognostycznego dodatkowo uwzględniono zmienne ZO20–ZO23 (w celu uwzględnienia wpływu warunków wiatrowych na obciążenie KSE), a także podjęto próbę oceny przydatności zakodowanej informacji o fazy księżyca (ZO5) do wyjaśniania zmiennej objaśnianej [6, 9]: udział dobowej szczytowej mocy 15-minutowej w szczycie tygodniowym (% - ZO4); maksymalna generacja w farmach wiatrowych (MW - ZO20); godzina wystąpienia maksymalnej generacji w farmach wiatrowych (- - ZO21); moc osiągalna w farmach wiatrowych (MW - ZO22); udział mocy generowanej w farmach wiatrowych do ich mocy zainstalowanej (% - ZO23).

Poszczególne ZO stanowią ciągi danych historycznych pobranych raz na dobę i obejmują okres pięciu lat, tj. 1 stycznia 2010 – 31 grudnia 2014.

### Metody typowania zmiennych objaśniających

Do metod zautomatyzowanych pozwalających na typowanie ZO zaliczono m.in. (w nawiasie wskazano kryterium doboru zmiennych):

- regresja wieloraka ( $\beta > \pm 0,04$ ) – M1 (metoda klasyczna);
- metoda MARSplines (ranking predyktorów) – M2 (metoda DM);
- współczynniki korelacji Pearsona ( $> 0,47$ ) – M4 (metoda klasyczna).

Do innych metod typowania ZO zaliczono:

- wybór ZO postrzeganych jako posiadające znaczący wpływ na obciążenie KSE (kryterium eksperckie) – M3;
- wybór wszystkich posiadanych ZO – M5.

Do typowania ZO dla poszczególnych wymienionych metod jako zmienną objaśnianą wykorzystano wspomnianą szczytową wartość 15-minutowej mocy zapotrzebowania w KSE w ciągu doby. Dla tak przyjętego podejścia uzyskano następujące zestawy ZO dla poszczególnych metod (tab.3).

Tabela 3. Wykaz zestawów ZO z podziałem na całkowitą ich liczbę i liczbę ZO płatnych

Metoda typująca	Liczba zmiennych	Liczba zmiennych płatnych	Zmienne objaśniające
M1	7	6	4, 6–7, 13, 16–17, 19
M2	8	5	3–4, 6–7, 16–17, 20, 22
M3	5	5	11, 13–15, 18
M4	6	5	4, 6–7, 13, 16–17
M5	22	14	1, 3–23

### Analiza potencjalnych kosztów pozyskania ZO dla prognoz dobowego obciążenia KSE wszystkich dób w roku z rozdzielczością godzinową

Zestawienie wstępnie wytypowanych zestawów ZO zaprezentowane w tabeli 3 wskazuje, że dla zmiennych płatnych metody zautomatyzowane wytypowały najmniej liczne zestawy ZO potrzebnych do opracowania jak najdokładniejszej prognozy szczytowego dobowego obciążenia KSE. Z wymienionej trójki najmniejszą liczbę ZO wytypowała metoda MARSplines oraz metoda wykorzystująca współczynnik Pearsona (5 ZO). Metoda regresji wielorakiej wytypowała 6 ZO. Dla opracowanych zestawów ZO dla każdej z metod zautomatyzowanego ich doboru przeprowadzono weryfikację skuteczności uzyskiwanych prognoz *ex post*, których wyniki zaprezentowano w tabeli 4. W tabeli 4 literą S oznaczono metodę statystyczną zastosowaną do weryfikacji skuteczności prognozowania wygasłego, literą M oznaczono metodę „zautomatyzowaną” doboru zmiennych objaśniających. W ostatnim wierszu zaprezentowano średni arytmetyczny błąd wynikający z zastosowania piętnastu metod statystycznych weryfikacyjnych.

Tabela 4. Uśrednione arytmetycznie wyniki mierników oceny dokładności prognoz *ex post* dla każdego roku z osobna

Metoda S/M	M1	M2	M3	M4	M5
S1	2,68	1,86	6,99	2,7	1,87
S2	3,10	2,78	7,18	3,1	2,56
S3	3,10	2,78	6,77	3,1	2,56
S4	5,55	4,06	7,33	5,55	5,36
S5	5,55	4,06	7,33	5,55	5,36
S6	3,34	3,16	7,03	3,34	9,40
S7	5,67	3,16	7,35	5,67	9,40
S8	4,05	3,69	7,73	4,05	9,40
S9	2,69	1,74	6,99	2,71	1,73
S10	2,66	1,71	6,97	2,68	1,68
S11	2,91	2,85	7,05	2,94	-
S12	2,86	2,78	7,07	2,89	-
S13	2,91	2,85	7,05	2,94	-
S14	2,91	2,85	7,05	2,94	2,54
S15	4,59	3,78	6,70	4,49	3,54
Śr.	3,64	2,94	7,11	3,64	4,62

W celu przeprowadzenia symulacji rocznych kosztów zakupu prognoz założono 4 warianty (W1–W4) kosztów pozyskania prognozy na tydzień w przód, z rozdzielczością godzinową dla pojedynczego parametru 100–1000 zł. Na potrzeby tej symulacji założono, że prognoza będzie dotyczyć obciążenia KSE na każdą godzinę doby w roku, bez określania, czy jest to wartość maksymalna, średnia czy minimalna. Zestawienie danych wykorzystanych do symulacji kosztowej przedstawiono w tabeli 5.

Dodatkowo w tabeli 4 (w ostatnim wierszu) zamieszczono uśrednione arytmetycznie wartości czterech mierników skuteczności prognoz: MPE, MAPE, RMSPE, Theila [10]. Uśrednienie dotyczyło błędów prognoz wyrażonych w % za okres pięciu lat wstecz przy

wykorzystaniu 15 metod prognostycznych wymienionych w tabeli 1.

Tabela 5. Symulacja rocznych kosztów zakupu prognoz ZO na potrzeby krótkoterminowej prognozy dobowego obciążenia KSE na tle uśrednionych mierników skuteczności prognoz *ex post*

Metoda	J.	M1	M2	M3	M4	M5
Liczba ZO płatnych	-	6	5	5	5	14
Liczba tygodni	-	52	52	52	52	52
Tygodniowy koszt zakupu prognozy zmiennych objaśnianej	ZO W1	1,00	1,00	1,00	1,00	1,00
	ZO W2	1,00	0,90	0,80	0,70	0,60
	ZO W3	0,10	0,20	0,30	0,40	0,50
	ZO W4	0,60	0,70	0,80	0,90	1,00
	ZO W5	0,80	0,90	1,00	0,90	0,80
	ZO W6	1,00	0,80	0,60	0,40	1,00
	ZO W7	1,00	0,50	1,00	0,50	1,00
	ZO W8	0,50	1,00	0,50	1,00	0,50
	ZO W9	1,00	1,00	1,00	0,50	0,50
	ZO W10	0,50	0,50	1,00	1,00	1,00
	ZO W11	1,00	1,00	1,00	1,00	0,50
	ZO W12	0,50	1,00	1,00	1,00	1,00
	ZO W13	0,50	0,50	0,50	1,00	1,00
	ZO W14	1,00	1,00	0,50	0,50	0,50
	ZO W15	1,00	1,00	1,00	0,50	0,50
	ZO W16	0,50	0,50	1,00	1,00	1,00
	ZO W17	1,00	1,00	1,00	1,00	0,50
	ZO W18	0,50	1,00	1,00	1,00	1,00
Roczny koszt zakupu prognozy zmiennych objaśnianej	W1	312,0	260,0	260,0	260,0	728,0
	W2	312,0	234,0	208,0	182,0	436,8
	W3	31,2	52,0	78,0	104,0	364,0
	W4	187,2	182,0	208,0	234,0	728,0
	W5	249,6	234,0	260,0	234,0	582,4
	W6	312,0	208,0	156,0	104,0	728,0
	W7	312,0	130,0	260,0	130,0	728,0
	W8	156,0	260,0	130,0	260,0	364,0
	W9	312,0	260,0	260,0	130,0	364,0
	W10	156,0	130,0	260,0	260,0	728,0
	W11	312,0	260,0	260,0	260,0	364,0
	W12	156,0	260,0	260,0	260,0	728,0
	W13	156,0	130,0	130,0	260,0	728,0
	W14	312,0	260,0	130,0	130,0	364,0
	W15	312,0	260,0	260,0	130,0	364,0
	W16	156,0	130,0	260,0	260,0	728,0
	W17	312,0	260,0	260,0	260,0	364,0
	W18	156,0	260,0	260,0	260,0	728,0
Roczny śr. koszt W1–W18	tys. zł	234,0	209,4	216,7	206,6	562,2
Błąd (%)	%	3,64	2,94	7,11	3,64	4,62

Legenda: J – jednostka; tyg. – tygodniowy; prog. – prognoza; ZO – zmienna objaśniająca.

Analiza danych z tabeli 5 wskazuje, że metody M4, M2 i M3 – dzięki wytypowaniu najmniejszej liczby zmiennych płatnych – pozwalają na uzyskanie najniższych uśrednionych kosztów zakupu prognoz tych zmiennych w skali roku (52 tygodni). Dodatkowo metoda M2 pozwala, oprócz minimalizacji kosztów zakupu prognoz ZO, na uzyskanie najdokładniejszych prognoz *ex post* za okres 2010–2014. Ponadto metoda M2 umożliwia minimalizację nakładów czasu poświęcanych na budowę modelu prognostycznego, dzięki umożliwieniu badaczowi wglądu w automatyczną ocenę ważności ZO z podanego zbioru zmiennych. Tym samym metoda M2 wydaje się jednocześnie najskuteczniejsza, najtańsza i najszybsza w warunkach przeprowadzonego eksperymentu symulacyjnego.

### Podsumowanie

Podstawowym celem prowadzonych badań, których wybrane wyniki zostały przedstawione w niniejszej publikacji, była analiza możliwości i skuteczności zautomatyzowania typowania optymalnego doboru zestawu ZO, z założeniem minimalizacji nakładów finansowych i nakładów pracy. Mając na uwadze powyższe, wnioskując się, że zalecany jest wybór metody M1 (regresja wieloraka), M2 (MARSplines) oraz M4 (współczynnik Pearsona) typowania zestawu zmiennych w procesie budowy modelu prognostycznego. Zauważalne jest także, że dla

zaprezentowanego zestawu zmiennych płatnych najmniejszą liczbę wymaganych danych wytypowano również w oparciu o badania literaturowe, doświadczenie i wiedzę autorską. Przedstawione podejście wskazuje na jego przydatność do wstępnego optymalizowania przewidywanych kosztów opracowywania prognoz. Założenie czterech różnych scenariuszy kształtowania się cen prognoz poszczególnych ZO wskazuje, że M2 (MARSplines), M3 (współczynniki korelacji Pearsona) oraz M4 (metoda ręcznego doboru) bez względu na rozkład kosztów zakupu prognoz poszczególnych ZO będą dawać najniższe średnioroczne koszty zakupu (ok. 200 tys. zł) dla założonych wariantów kosztów zakupu. Powyższe wynika z faktu, że dla każdej z metod wytypowana liczba zmiennych płatnych wynosiła 5. Tym samym wykazana została odporność tych trzech metod na potencjalnie różne koszty zakupu ZO. Wybór najkorzystniejszej metody spośród trzech najmniej kosztownych, która w przyszłości mogłaby okazać się najlepszą dla budowania skutecznych prognoz, jest możliwy dzięki porównaniu średniej arytmetycznej skuteczności mierników *ex post* wykonywanych prognoz z uśrednionymi kosztami rocznymi zakupu danych prognostycznych. Z wyróżnionej wcześniej trójki metod odrzucić należy metodę M3, która uzyskała najmniej korzystny wynik (7,11%) w ocenie skuteczności prognozowania za okres pięciu lat. Z pozostałej dwójki metod wyróżniającym się wynikiem skuteczności prognoz *ex post* charakteryzuje się metoda M2 (MARSplines), która jako jedyna pozwoliła na uzyskanie skuteczności prognozowania na poziomie poniżej 3% (2,94%). Metoda ta zapewnia tym samym zarówno minimalizację kosztów zakupu prognoz ZO, dzięki wyposażeniu badacza prognosty w możliwość zautomatyzowanego typowania ZO, jak i zapewnia najwyższą skuteczność prognozowania *ex post* spośród wytypowanych w tej publikacji metod. Tym samym należy uznać, że wykorzystanie metody MARSplines pozwala na minimalizację czasu potrzebnego na przygotowanie ZO do modelu prognostycznego, minimalizację kosztów pozyskiwania prognoz tych danych oraz zapewnia (w realiach zastosowanego podejścia) największą dokładność uzyskiwanych za jej pomocą prognoz. Konieczna jest weryfikacja skuteczności opracowanych zestawów danych (zmiennych) objaśniających dla prognozowania *ex ante* krótkoterminowego dobowego obciążenia KSE z rozdzielczością godzinową. Weryfikacja taka powinna obejmować zbiór metod klasycznych oraz zgłębiania danych DM. Należy zauważyć ciekawą zależność, że wykonane badania i analiza porównawcza wyników wskazują, że najkorzystniej będzie się skupić na metodzie MARSplines, która z grupy 8 ZO wytypowała jedynie 5 zmiennych wymagających zakupu [11, 12]. Metoda ta należy do grupy metod DM i oprócz analizy statystycznej oferuje szybką i zautomatyzowaną drogę do uzyskania najkorzystniejszego zestawu ZO. Warto podkreślić również, że uzyskane wyniki są najkorzystniejsze dla rozpatrywanego 5-letniego zbioru danych historycznych. Należy zaznaczyć jednakże na obecnym etapie, że metoda MARSplines na etapie uczenia (prognozowanie w trybie *ex post*) jest podatna na przeuczenie i w zderzeniu z realnym prognozowaniem może dawać mniej obiecujące wyniki. Cennym doświadczeniem mogłoby być uwzględnienie jako ZO danych historycznych o cenach energii elektrycznej na Rynku Bilansującym, Towarowej Giełdzie Energii lub przebiegów notowań ciągłych towarów giełdowych, np. gazu, ropy, węgla w ramach rynków

intraday i *day ahead* [12]. Pomimo znaczących zalet metody MARSplines i uzyskania drugiej pozycji w rankingu rocznych średnich kosztów (dla 18 wariantów) i pierwszej pozycji w skuteczności prognozowania *ex post* może ona – w porównaniu ze zbiorami zmiennych objaśniających typowanych w oparciu o macierz współczynników korelacji Pearsona – dawać mniej konkurencyjne wyniki w trybie *ex ante* i zniwelować różnicę 0,7% w porównaniu do metody MARSplines.

Przedstawione w publikacji proponowane podejście do typowania zmiennych objaśniających z zachowaniem jak największego automatyzmu pozwala nie tylko na skrócenie czasu ich doboru, ale także na minimalizację kosztów zakupu prognoz tych zmiennych. Zastosowane przez autorów podejście uwzględniające ocenę kosztową pozyskiwania różnych zestawów zmiennych objaśniających może stanowić narzędzie do optymalizacji ponoszonych kosztów opracowywania prognoz.

**Autorzy:** mgr inż. Rafał Czapał, PSE Innowacje Sp. z o.o., Centrum Kompetencji Badania i Rozwój, ul. Jordana 25, 40-056 Katowice, E-mail: [rafał.czapał@pse.pl](mailto:rafał.czapał@pse.pl); dr inż. Mateusz Szablicki, Politechnika Śląska, Wydział Elektryczny, Instytut Elektroenergetyki i Sterowania Układów, PSE Innowacje Sp. z o.o., E-mail: [mateusz.szablicki@pse.pl](mailto:mateusz.szablicki@pse.pl); dr inż. Piotr Rzepka, Politechnika Śląska, Wydział Elektryczny, Instytut Elektroenergetyki i Sterowania Układów, PSE Innowacje Sp. z o.o., E-mail: [piotr.rzepka@pse.pl](mailto:piotr.rzepka@pse.pl).

#### LITERATURA

- [1] Bartodziej G., Tomaszewski M., Polityka energetyczna i bezpieczeństwo energetyczne – wydanie II, Wydawnictwo „Nowa Energia”, Racibórz 2008.
- [2] Weron A., Weron R., Giełda Energii: strategie zarządzania ryzykiem, Wydawnictwo CIRE, Centrum Metod Stochastycznych im. Hugona Steinhausa, Politechnika Wroclawska, Wrocław 2000.
- [3] Klóska R., Hundert M., Czyżycy R., Wybrane zagadnienia z prognozowania, Wydawnictwo Economicus, Szczecin 2007.
- [4] Zeliaś A., Pawełek B., Wanat S., Prognozowanie ekonomiczne. Teoria, przykłady, zadania, Wydawnictwo Naukowe PWN, Warszawa 2003.
- [5] Kot S.M., Jakubowski J., Sokołowski A., Statystyka. Podręcznik dla studiów ekonomicznych, Centrum Doradztwa i Informacji, Warszawa 2007.
- [6] Witryna internetowa: [www.pse.pl](http://www.pse.pl).
- [7] Prognozowanie w elektroenergetyce. Zagadnienia wybrane, red. I. Dobrzyńska, Wydawnictwo Politechniki Śląskiej, Częstochowa 2012.
- [8] Kopecki K. i in., Analiza i prognoza obciążeń elektroenergetycznych, Wydawnictwa Naukowo-Techniczne, Warszawa 1971.
- [9] Pieczarko R., Sołtysik M., Analiza wpływu generacji źródeł wiatrowych na poziom kształtowania się cen energii elektrycznej na rynku SPOT, Konferencja Naukowa „Prognozowanie w elektroenergetyce”, Podlesice 2016.
- [10] Czapał R., Rzepka P., Szablicki M., Typowanie zmiennych objaśniających przy wykorzystaniu zautomatyzowanych metod statystycznych jako sposób optymalizacji wyboru metody estymacji szczytowego dobowego obciążenia KSE, Konferencja Naukowa „Prognozowanie w elektroenergetyce”, Podlesice 2016.
- [11] Czapał R., Optymalizacja kosztów zakupu danych wejściowych do prognoz dobowego obciążenia KSE przy wykorzystaniu zautomatyzowanych metod statystycznych, Konferencja Naukowa „APE'17”, Jastrzębia Góra 2017.
- [12] Czapał R., Optymalizacja kosztów zakupu danych wejściowych do prognoz dobowego obciążenia KSE przy wykorzystaniu zautomatyzowanych metod statystycznych, Acta Energetica 3/2017 (32), Gdańsk 2017.