

Typowanie zmiennych objaśniających przy wykorzystaniu zautomatyzowanych metod statystycznych jako sposób optymalizacji wyboru metody estymacji szczytowego dobowego obciążenia KSE

Streszczenie. Artykuł prezentuje możliwość skorzystania z metod statystycznych automatyzujących dobór zmiennych objaśniających na przykładzie szczytowego obciążenia dobowego KSE. Testy ex-post dotyczyły 10 zbiorów zmiennych objaśniających dla metod statystycznych klasycznych i typu Data Mining. Uzyskana macierz wyników pozwala wstępnie wybrać najkorzystniejszy zbiór zmiennych objaśniających i metodę statystyczną.

Abstract. The paper presents the possibility of using the statistical methods automating the selection of explicatory variables for the daily peak demand of the National Power Grid in Poland. Predictions in ex-post were conducted for 10 predictors' set and both for classical and Data Mining methods. The obtained matrix of ex-post predictions measures gives a chance for a primal selection of the best set and statistical method. (Selection of predictors using automated statistical methods as a way of choosing the estimation method of daily peak demand of the National Power Grid).

Słowa kluczowe: szczytowe 15-minutowe obciążenie KSE, zmienne objaśniające meteorologiczne, Data Mining, mierniki ex-post
Keywords: National Power System 15-minute Power Demand, explicatory variables, Data Mining, ex-post indicators.

Wstęp

Prognozowanie stanowi jeden z ważnych elementów składających się na bezpieczeństwo systemu elektroenergetycznego [1]. Operator sieci przesyłowej ponosi wiele ryzyk do których należy m.in. ryzyko znaczącego odchylenia prognozy od rzeczywistego obciążenia Krajowego Systemu Elektroenergetycznego (KSE) [2]. Dobór zmiennych objaśniających oraz metody statystycznej stanowi kluczowy etap w procesie budowy modelu prognostycznego. Im lepsze dopasowanie zmiennych objaśniających do zmiennej objaśnianej tym bardziej precyzyjny opis zmiennej objaśnianej przez zmienne objaśniające [3]. Wskutek powyższego możliwe jest uzyskanie dokładnego modelu prognostycznego dla założonej zmiennej objaśnianej. Staranny dobór zmiennych objaśniających jest kluczem do zbudowania jak najlepszego modelu prognostycznego. Dobór ten wymaga jednak często dużych nakładów pracy (proces ten jest bardzo czasochłonny), a pozyskiwanie pełnych i wiarygodnych danych historycznych charakteryzujących się wysoką rozdzielczością może być trudne i wiąże się niejednokrotnie ze sporymi nakładami finansowymi [4]. Idealnym stanem, z punktu widzenia przygotowania prognozy dla operatora systemu elektroenergetycznego, jest minimalizacja nakładów pracy oraz (bardzo często) minimalizacja trudności i kosztów pozyskiwania danych stanowiących zestaw zmiennych objaśniających. Minimalizacja ww. nakładów pracy, obsługi danych w postaci zmiennych objaśniających i kosztów ich pozyskania możliwa jest dzięki:

- zautomatyzowanym procesom typowania i rankingowania najkorzystniejszych zmiennych dla danego procesu - co możliwe jest dzięki zastosowaniu pakietów statystycznych;
- selekcji najlepszych zmiennych objaśniających z danej grupy pozyskiwanych danych - dzięki ich testowaniu przez różne metody statystyczne.

Jedne z najwyższych kosztów generuje pozyskanie wartości zmiennych objaśniających będących pomiarami parametrów meteorologicznych, które są wykorzystywane dla potrzeb gospodarczych/przemysłowych. Parametry meteorologiczne wynikające z położenia geograficznego [5], [6] w sposób znaczący wpływają na obciążenie KSE.

W zależności od długości analizowanego przedziału czasu (szeregu czasowego) w procesie budowy modelu prognostycznego oprócz doboru zmiennych objaśniających konieczny jest także optymalny wybór metody prognostycznej [7]. Wybór tych wielkości bardzo mocno rzutuje na uzyskane wyniki i w dużej mierze zależy od wiedzy i doświadczenia prognosty. W procesie budowy modelu prognostycznego ponowny wybór zmiennych objaśniających i/lub metody prognozowania w szczególności zachodzi w sytuacji, gdy opracowany wcześniej model nie daje satysfakcjonujących wyników [8].

Przeprowadzone badania, których wybrane wyniki prezentuje niniejsza publikacja, dotyczyły próby opracowania „zautomatyzowanej” metody doboru najlepszych zmiennych objaśniających. Przez najlepsze zmienne objaśniające rozumie się zmienne, które jak najbardziej precyzyjnie opisują zmienną objaśnianą. Najlepsze zmienne objaśniające typowano za pomocą wybranych metod statystycznych i następnie testowano je takimi metodami. Pozwoliło to na krzyżową weryfikację „jakości” grup zmiennych, w tym jakości dopasowywania do zmiennej objaśnianej. Dzięki takiemu podejściu uzyskuje się informację pod postacią macierzy uśrednionych wartości błędów ex-post dla każdego roku analizy z osobna, która pozwala ocenić jakie zmienne spośród zmiennych objaśniających będą najlepiej wyjaśniać zmienną objaśnianą (zmienną prognozowaną) oraz jaka metoda, dla analizowanego zakresu danych historycznych, umożliwi (potencjalnie) uzyskanie najdokładniejszej prognozy. Jako zmienną objaśnianą przyjęto maksymalną (szczytową) wartość mocy 15-minutowej obciążenia w ciągu doby KSE [9]. Jako najkorzystniejsze rozwiązanie zadanego problemu założono jednoczesną minimalizację obu funkcji celu (nakładów pracy i nakładów finansowych).

Zbiór zmiennych objaśniających - zmienne meteorologiczne

Jako zbiór zmiennych objaśniających wybrano pomiary meteorologiczne pozyskane z lokalizacji na środkowym południu Polski, które można uznać jako reprezentatywne dla całego KSE. Wymieniona lokalizacja nie stanowi „bieguna zimna” (Suwałki) ani nie stanowi „bieguna ciepła” (Wrocław), a wieloletnie obserwacje pozwalają oszacować,

że jest ona w dużym przybliżeniu zgrubną średnią (poniżej $\pm 1^{\circ}\text{C}$) z obu biegunów temperatury [10], [11]. Pierwszy zestaw zmiennych (najliczniejszy) składa się z 13 następujących parametrów meteorologicznych:

- (Z6) - temperatura otoczenia maksymalna ($^{\circ}\text{C}$),
- (Z7) - temperatura otoczenia minimalna ($^{\circ}\text{C}$),
- (Z8) - opady deszczu (mm),
- (Z9) - prędkość wiatru średnia (km/h),
- (Z10) - prędkość wiatru średnia (km/h) ekspercka,
- (Z11) - prędkość wiatru maksymalna (km/h),
- (Z12) - ciśnienie atmosferyczne (hPa),
- (Z13) - liczba stopniodni grzewczych ($^{\circ}\text{C}$ dzień),
- (Z14) - liczba stopniodni chłodniczych ($^{\circ}\text{C}$ dzień),
- (Z15) - liczba godzin słonecznych (-),
- (Z16) - energia słoneczna (W/m^2),
- (Z17) - poziom promieniowania UV (-),
- (Z18) - temperatura punktu rosy ($^{\circ}\text{C}$),
- (Z19) - temperatura mokrego termometru ($^{\circ}\text{C}$).

Do powyższego zestawu dołączono dwie zmienne objaśniające, które istotnie uzupełniają powyższy zbiór. Pierwsza z nich zawiera w postaci zakodowanej informacji o dacie dokonanego pomiaru (rok/miesiąc/dzień) z rozróżnieniem kolejnych dni tygodnia oraz z uwzględnieniem podziału na dzień nieświęteczny i święteczny. Druga ze wspomnianych zmiennych ma postać czasową niezakodowaną, w której zawarto informację o czasie (z rozdzielczością 15-minutową) wystąpienia szczytowej wartości mocy 15-minutowej obciążenia KSE - dane te pozyskano z publikacji PSE S.A. [9] w każdej dobie analizowanego szeregu czasowego. Zmienną objaśnianą była wspomniana szczytowa wartość mocy 15-minutowej dobowego obciążenia KSE.

Zbiór zmiennych objaśniających - pozostałe zmienne

Dodatkowo, w celu uwydatnienia wpływu warunków wiatrowych na obciążenie KSE, przetestowano zmienne Z20-23, a także podjęto próbę oceny przydatności zakodowanej informacji o cyklu księżycy do wyjaśniania zmiennej objaśnianej. Do zmiennych tych zaliczono poniższe pozycje:

- (Z4) - udział procentowy dobowej szczytowej mocy 15-minutowej w szczycie tygodniowym,
- (Z5) - zakodowana informacja o fazie księżycy,
- (Z20) - maksymalna generacja w farmach wiatrowych,
- (Z21) - godzina wystąpienia maksymalnej generacji w farmach wiatrowych,
- (Z22) - moc osiągalna w farmach wiatrowych,
- (Z23) - procentowy udział mocy generowanej w farmach wiatrowych do ich mocy zainstalowanej.

Wytypowany zbiór zmiennych objaśniających stanowi Główny Zbiór Zmiennych Objaśniających (GZZO). Poszczególne zmienne stanowią ciąg danych historycznych charakteryzujących się rozdzielczością dobową. Obejmują okres 5 lat, od 1 stycznia 2010 r. do 31 grudnia 2014 r.

Zbiór wstępnie wytypowanych metod statystycznych

W przeprowadzonych badaniach uwzględniano metody statystyczne klasyfikowane jako metody zgłębiania danych (ang. *Data Mining*) oraz klasyczne metody statystyczne.

Z metod *Data Mining* wybrano (zgodnie z klasyfikacją w pakiecie STATISTICA [12], [13], [14], [15], [16]):

- (S1) - Multivariate Adaptive Regression Splines (Wielozmienna regresja adaptacyjna z użyciem funkcji sklepanych),
- (S2) - Classification and Regression Trees /C&RT/ (Ogólne modele drzew klasyfikacyjnych i regresyjnych) - standardowy C&RT,

- (S3) - Classification and Regression Trees /C&RT/ (Ogólne modele drzew klasyfikacyjnych i regresyjnych) - C&RT z układami,
- (S4) - Chi-squared Automatic Interaction Detector (Automatyczny detektor interakcji za pomocą chi-kwadrat) - standardowy CHAID,
- (S5) - Chi-squared Automatic Interaction Detector (Automatyczny detektor interakcji za pomocą chi-kwadrat) - CHAID z układami,
- (S6) - Interactive Regression Trees (Drzewa interakcyjne regresyjne C&RT, CHAID),
- (S7) - Interactive Regression Trees (Drzewa interakcyjne regresyjne C&RT, CHAID) - CHAID,
- (S8) - Interactive Regression Trees (Drzewa interakcyjne regresyjne) - CHAID wyczerpujący,
- (S9) - Generalized Additive Models (Uogólnione modele addytywne) - funkcja wiążąca identycznościowa,
- (S10) - Generalized Additive Models (Uogólnione modele addytywne) - funkcja wiążąca logarymiczna.

Do grupy metod statystycznych klasycznych (zgodnie z klasyfikacją w pakiecie STATISTICA) wybrano metody tj.:

- (S11) - Multiple Regression (Regresja wieloraka),
- (S12) - Generalized Linear and Non-Linear Models (Ogólne modele liniowe i nieliniowe),
- (S13) - General Regression Models (Ogólne modele regresji),
- (S14) - Fragmentary Least Squares Models (Modele cząstkowych najmniejszych kwadratów),
- (S15) - Artificial Neural Networks (Sztuczne Sieci Neuronowe).

Powyższe metody były testowane w trybie *ex-post* (prognozy wygasłe na znanych wartościach historycznych) pod kątem jakości dopasowania zmiennych objaśniających do zmiennej objaśnianej. W tym celu przy pomocy metod określono 10 zestawów zmiennych objaśniających wytypowanych z GZZO automatycznie (przez niektóre z poniższych metod) i ręcznie metodą ekspercką. Do metod (M_n) zautomatyzowanych zaliczono (z odpowiednim kryterium):

- (M1) - regresję wieloraką ($\beta > \pm 0,04$),
- (M2) - metodę MARS (ranking predyktorów),
- (M4) - obliczanie współczynnika Pearsona ($> 0,47$),
- (M8) - metodę szybką C&RT (ranking predyktorów),
- (M10) - dobór i eliminację zmiennych (ranga zmiennej).

Do metod eksperckich zaliczono (M_n):

- (M3) - wybór zmiennych objaśniających postrzeganych jako posiadające znaczący wpływ na obciążenie KSE,
- (M5) - wybór wszystkich posiadanych zmiennych objaśniających (brak kryterium),
- (M6) - wybór wszystkich spośród najlepszych zmiennych objaśniających z metod M1-M4 (zgodnie z kryteriami dla metod od M1 do M4),
- (M7) - wybór wszystkich spośród najlepszych zmiennych objaśniających oraz ekspercki dobór dodatkowej/dodatkových zmiennych objaśniających,
- (M9) - regresję wieloraką przeprowadzaną iteracyjnie osobno dla każdej zmiennej objaśnianej ($\beta > \pm 0,1$).

Zastosowane zestawy zmiennych objaśniających

Po zastosowaniu opisanego powyżej podejścia uzyskano następujące zestawy (M) zmiennych objaśniających (Z):

- M1 - Z (4, 6-7, 13, 16-17, 19) - 7 zmiennych,
- M2 - Z (3-4, 6-7, 16-17, 20, 22) - 8 zmiennych,
- M3 - Z (11, 13-15, 18) - 5 zmiennych,

- M4 - Z (4, 6-7, 13, 16-17) - 6 zmiennych,
- M5 - Z (1, 3-23) - 22 zmienne,
- M6 - Z (3-4, 6-7, 11, 13-20, 22) - 14 zmiennych,
- M7 - Z (3-4, 6-7, 11-20, 22) /dodatkową zmienną jest ciśnienie - Z12/- 15 zmiennych,
- M8 - Z (4, 6-7, 12-20, 22) - 13 zmiennych,
- M9 - Z (4, 6-7, 13-14, 16-19, 22-23) - 11 zmiennych,
- M10 - Z (3-4, 6-7, 13-14, 16-19) - 10 zmiennych.

Zastosowane mierniki ex-post

Wytypowano 15 metod statystycznych z grupy metod *Data Mining* oraz klasycznych metod prognostycznych i następnie poddano testom *ex-post*. W testach wykorzystano 10 zbiorów zmiennych objaśniających wytypowanych ze zbioru GZZO). Rezultatem testów jest macierz wyników. Każdy z elementów tej macierzy stanowi średnią arytmetyczną czterech procentowych mierników *ex-post* jakości dopasowania danych do modelu, tj. [4], [7], [8], [12], [17]:

- MPE - Średni błąd procentowy (Mean Percentage Error),
- MAPE - Średni całkowity błąd procentowy (Mean Absolute Percentage Error),
- RMSPE - Pierwiastek Średniokwadratowego błędu procentowego (Root Mean Square Error),
- Theil - Współczynnik rozbieżności Theila.

Poziom wartości mierników *ex-post*, który uznano za kwalifikujący do dalszych rozważań, ustalono na poziomie nieprzekraczającym 5%. Założono również, że za najkorzystniejsze uznane zostaną zestawy zmiennych objaśniających gwarantujące jak najniższą średnią arytmetyczną ze wspomnianych wartości mierników *ex-post* dla każdego roku z osobna. Analogiczne założenie przyjęto dla wytypowania najkorzystniejszej metody/metod, które będą uznane za najlepiej rokujące dla przyszłych badań, w oparciu o zakres danych uwzględnionych w tej publikacji.

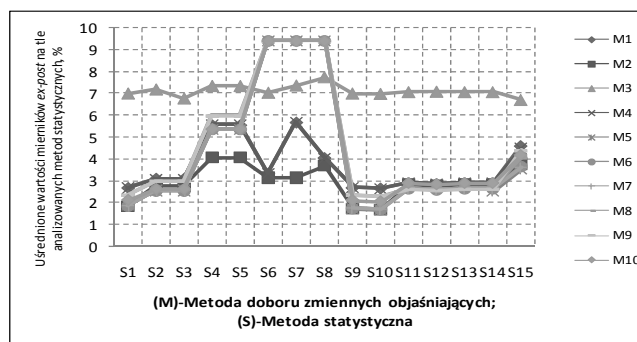
Analiza uzyskanych wyników uśrednionych

Macierz wyników testów *ex-post* wstępnie wytypowanych metod statystycznych i wstępnie wytypowanych zestawów zmiennych objaśniających zaprezentowano w tabeli 1. W wierszach zamieszczono wyniki uzyskiwane dla wybranych metod statystycznych oferowanych w pakiecie *STATISTICA*, a w kolumnach zamieszczono wyniki uzyskane dla poszczególnych zestawów zmiennych objaśniających. Odzwierciedleniem graficznym tabeli 1 jest rysunek 1.

Metodą doboru zmiennych objaśniających, która średnio najlepiej określiła zestaw takich zmiennych jest metoda MARS (M2). Uzyskana średnia arytmetyczna wszystkich uśrednionych mierników (dla każdego roku z osobna) jakości szacunków *ex-post* wyniosła 2,94%. Na drugim *ex-aequo* miejscu znalazły się regresja wieloraka (M1) oraz metoda wyznaczania wskaźnika Pearsona (M4) z wynikiem ok. 3.64%. Wyniki dla wszystkich 10 metod doboru zmiennych objaśniających prezentuje rysunek 2. Metodą statystyczną, która średnio najlepiej operowała na typowanych zestawach zmiennych objaśniających była uogólniona metoda modeli addytywnych (S10) dla funkcji logit (2,57%). Drugie miejsce zajęła ta sama metoda (S11) dla funkcji addytywnej (2,62%), a trzecie metoda (S1) MARS ze średnim błędem 2,67% (rys. 3). Najniższy wynik (1,68%) średniej arytmetycznej mierników dopasowania *ex-post* dla całej macierzy uzyskano dla zestawu zaproponowanych zmiennych objaśniających (M5) przy zastosowaniu metody uogólnionych modeli addytywnych dla funkcji *logit* (S10) - wynik ten zaznaczono pogrubioną czcionką w macierzy zawartej w tabeli 1.

Tabela 1. Macierz uśrednionych mierników procentowych *ex-post* z każdego roku z osobna - metoda statystyczna x metoda doboru zmiennych objaśniających (S x M)

S/M	M1	M2	M3	M4	M5	M6	M7	M8	M9	M10
S1	2,68	1,86	6,99	2,70	1,87	1,86	1,86	2,34	2,38	2,12
S2	3,10	2,78	7,18	3,10	2,56	2,57	2,57	2,95	3,00	2,61
S3	3,10	2,78	6,77	3,10	2,56	2,57	2,57	2,95	3,00	2,61
S4	5,55	4,06	7,33	5,55	5,36	5,36	5,36	5,97	5,97	5,36
S5	5,55	4,06	7,33	5,55	5,36	5,36	5,36	5,97	5,97	5,36
S6	3,34	3,16	7,03	3,34	9,40	9,40	9,40	9,40	9,40	9,40
S7	5,67	3,16	7,35	5,67	9,40	9,40	9,40	9,40	9,40	9,40
S8	4,05	3,69	7,73	4,05	9,40	9,40	9,40	9,40	9,40	9,40
S9	2,69	1,74	6,99	2,71	1,73	1,76	1,75	2,33	2,36	2,10
S10	2,66	1,71	6,97	2,68	1,68	1,71	1,70	2,28	2,32	2,04
S11	2,91	2,85	7,05	2,94	-	2,65	2,63	2,64	2,63	2,89
S12	2,86	2,78	7,07	2,89	-	2,60	2,59	2,59	2,59	2,83
S13	2,91	2,85	7,05	2,94	-	2,65	2,63	2,64	2,63	2,89
S14	2,91	2,85	7,05	2,94	2,54	2,65	2,63	2,64	2,63	2,89
S15	4,59	3,78	6,70	4,49	3,54	3,67	3,61	4,06	4,42	4,17
Śr.	3,64	2,94	7,11	3,64	4,62	4,24	4,23	4,50	4,54	4,40

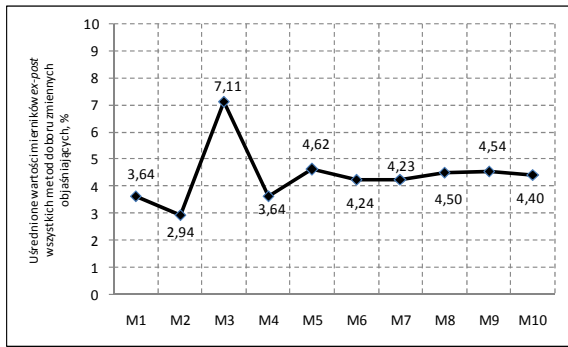


Rys.1. Uśrednione wartości mierników *ex-post* na tle analizowanych metod statystycznych

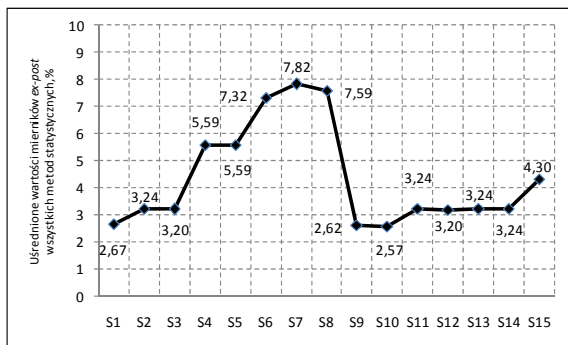
Osiągnięcie przez tę metodę tak dobrego wyniku należy uznać za bardzo interesujące i korzystne z punktu widzenia minimalizacji czasu poświęcanego na dobór zmiennych objaśniających (minimalizacja tylko jednej z dwóch założonych funkcji celu). Niestety taki wybór zestawu zmiennych objaśniających i metody statystycznej prowadzi do maksymalizacji nakładów czasowych i finansowych związanych z obróbką i pozyskiwaniem pomiarów meteorologicznych. Metoda ta dawała wyniki na podobnym poziomie (1,70%-1,71%) również dla zestawów zmiennych objaśniających (M2 - metoda MARS) oraz M6 i M7 - czyli wszystkich „najlepszych” zmiennych z metod M1-M4 wzbogaconych zmienną ciśnienie dla metody M7. Podkreśleniem w tej samej tabeli zaznaczono obiecujące, według autorów, wyniki.

Na podobnym, nieco „gorszym” poziomie, uzyskano wyniki dla tej samej metody, lecz dla funkcji addytywnej (S10). Podobnie jak w przypadku funkcji logitowej najkorzystniejsze zestawy zmiennych objaśniających były typowane przez metody M2, M5-M7 (poziom uśrednionych wartości błędów od 1,73% do 1,75% dla wymienionych metod doboru zmiennych objaśniających).

Trzecie miejsce dla omawianego kryterium, jednoczesnego wyboru najkorzystniejszej metody doboru zestawu zmiennych objaśniających i metody statystycznej, uzyskano dla metody MARS (S1) w połączeniu z zestawami (tak samo jak dla dwóch wcześniej kombinacji metod) M2 (MARS) oraz M5-M7 (wszystkie zmienne, wszystkie najlepsze zmienne oraz wszystkie najlepsze zmienne z autorskim doбором dodatkowej zmiennej, którą było ciśnienie).



Rys.2. Uśrednione arytmetycznie wartości wyników *ex-post* wszystkich analizowanych metod doboru zmiennych objaśniających



Rys.3. Uśrednione arytmetycznie wartości wyników *ex-post* wszystkich analizowanych metod statystycznych

Podsumowanie

Podstawowym celem prowadzonych badań, których wybrane wyniki zostały przedstawione w niniejszym artykule, było zautomatyzowanie i wytypowanie optymalnego doboru zestawu zmiennych objaśniających i optymalnej metody statystycznej z założeniem minimalizacji nakładów czasu i nakładów finansowych. Mając na uwadze powyższe zaleca się wybrać metodę M2 (MARS) typowania zestawu zmiennych oraz metodę statystyczną S1 (również MARS). W przypadku zakodowanej informacji o fazie księżyca nie odnotowano jej wpływu na osiągane wyniki przez żadną z analizowanych metod, co mogło wynikać z małej zmienności wartości tej zmiennej. Wyniki zostaną wykorzystane przyszłych badań, które będą prowadzone w trybie *ex-ante*.

Należy zauważyć ciekawą zależność, że wykonane badania i analiza porównawcza wyników przeprowadzona na podstawie macierzy zamieszczonej w tabeli 1 wskazują, że najkorzystniej (w sensie jednoczesnej minimalizacji funkcji celu polegającej na ograniczeniu nakładów czasu na dobór zmiennych i funkcji celu ograniczającej nakłady finansowe) jest skupić się na metodzie MARS. Metoda ta należy do grupy metod *Data Mining* i oprócz analizy statystycznej oferuje szybką i zautomatyzowaną drogę do uzyskania najkorzystniejszego zestawu zmiennych objaśniających. Warto podkreślić również, że uzyskane wyniki są najkorzystniejsze dla rozpatrywanego 5-letniego zbioru danych historycznych. Należy pamiętać jednakże na obecnym etapie, że metoda MARS na etapie uczenia (prognozowanie w trybie *ex-post*) jest podatna na przecuczenie i w zderzeniu z realnym prognozowaniem może dawać mniej obiecujące wyniki.

Pomimo faktu, że sztuczne sieci neuronowe nie uzyskały korzystnych, na tym etapie badań wyników, należy mieć na uwadze, że po poddaniu ich dalszemu doskonaleniu struktur i uczeniu mogą również w przyszłych badaniach okazać się skutecznym narzędziem prognostycznym. Przewiduje się, że przyszłe testy mogłyby obejmować kojarzenie metod statystycznych, przykładowo współpracę metody MARS z sieciami neuronowymi (po ich udoskonaleniu). Autorzy pragną podkreślić również, że kolejne badania prowadzone w tym zakresie mogą obejmować również dalsze poszukiwanie zmiennych objaśniających (nie tylko meteorologicznych), optymalizację długości ciągu czasowego umożliwiającego uzyskanie wyników na podobnym poziomie oraz testowanie wybranych metod w trybie *ex-ante*.

Autorzy: mgr. inż. Rafał Czapał-Atlas, PSE Innowacje sp. z o.o., ul. Jordana 25, 40-056 Katowice, E-mail: rafal.czapaj@pse.pl; dr inż. Piotr Rzepka, Politechnika Śląska, Wydział Elektryczny, Instytut Elektroenergetyki i Sterowania Układów, dr inż. Mateusz Szabliski, Politechnika Śląska, Wydział Elektryczny, Instytut Elektroenergetyki i Sterowania Układów

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., Czyżyci R., Wybrane zagadnienia z prognozowania, *Wydawnictwo Economicus*, n.5, Szczecin (2007)
- [4] Zeliaś A., Pawełek B., Wanat S., Prognozowanie ekonomiczne - Teoria, przykłady, zadania, *Wydawnictwo Naukowe PWN*, Warszawa (2003)
- [5] Dobrzańska I., (pod redakcją), Prognozowanie w elektroenergetyce - Zagadnienia wybrane, *Wydawnictwo Politechniki Śląskiej*, Częstochowa (2012)
- [6] Kopecki K., Dobrzańska I. i inni (praca zbiorowa.), Analiza i prognoza obciążeń elektroenergetycznych, *Wydawnictwa Naukowo-Techniczne*, Warszawa (1971)
- [7] Witkowska D., Podstawy ekonometrii i teorii prognozowania (Podręcznik z przykładami i zadaniami), *Oficyna Ekonomiczna*, Kraków (2005)
- [8] Kot S.M., Jakubowski J., Sokołowski A., Statystyka - Podręcznik dla studiów ekonomicznych, *Centrum Doradztwa i Informacji*, Warszawa (2007)
- [9] Witryna internetowa www.pse.pl
- [10] Witryna internetowa www.twojapogoda.pl
- [11] Witryna internetowa www.tvnmeteo.tvn24.pl
- [12] StatSoft Polska, STATISTICA - Przewodnik, *Wydawnictwo StatSoft Polska Sp. Z o.o.*, Kraków (2008)
- [13] Zastosowania statystyki i Data Mining, *Wydawnictwo StatSoft Polska Sp. Z o.o.*, Kraków (2006)
- [14] Praktyczne wykorzystanie analizy danych i Data Mining, *Wydawnictwo StatSoft Polska Sp. Z o.o.*, Kraków (2008)
- [15] Data Mining - Metody i przykłady - Seminarium, *Wydawnictwo StatSoft Polska Sp. Z o.o.*, Kraków-Warszawa (2002)
- [16] Statystyka i Data Mining w badaniach naukowych - Seminarium, *Wydawnictwo StatSoft Polska Sp. Z o.o.*, Kraków (2006)
- [17] Helt P., Parol M., Piotrowski P., Metody sztucznej inteligencji w elektroenergetyce, *Oficyna Wydawnicza Politechniki Warszawskiej*, Warszawa (2000)