

Analiza doboru zmiennych w zadaniu prognozowania ultrakrótkoterminowego produkcji energii elektrycznej w systemach fotowoltaicznych

Streszczenie. W tekście przedstawiono obszerną analizę różnych metod doboru zmiennych do modeli prognozowania ultrakrótkoterminowego produkcji energii przez system fotowoltaiczny, a w szczególności wpływu tych metod na jakość prognoz. Sformułowano wnioski końcowe z przeprowadzonych analiz i prognoz.

Abstract. The paper presents comprehensive analysis of input variables selection for chosen prediction models in ultra short-term electric energy production forecasting for photovoltaic systems, especially influence of input variables selection methods on quality of prediction models. The obtained results have been discussed and the final conclusions have been presented. (The analysis of input variables selection in ultra short-term electric energy production forecasting for photovoltaic systems)

Słowa kluczowe: prognozowanie ultrakrótkoterminowe produkcji energii elektrycznej, system fotowoltaiczny, analiza statystyczna danych, dobór zmiennych.

Keywords: ultra short-term forecasting of electric energy production, photovoltaic system, statistical analysis of data, variables selection.

doi:10.12915/pe.2014.04.02

Wstęp

Typowe cele prognoz ultrakrótkoterminowych w elektroenergetyce to procesy bieżącego sterowania oraz procesy planowania pracy prognozowanego systemu np. systemu smart grids, w którym energia elektryczna pozyskiwana jest z OZE.

Ogólny podział metody prognozowania mocy w systemach fotowoltaicznych to [1]:

- metody fizyczne,
- metody statystyczne,
- metody wykorzystujące sztuczną inteligencję.

Metod prognozowania mocy oraz produkcji energii elektrycznej w systemach fotowoltaicznych wykorzystujących modele sieci neuronowych jest dość dużo. Wybór jest uzależniony od horyzontu prognoz, dostępności różnych danych historycznych oraz prognoz otrzymanych z NWP. Wykorzystuje się różne typy sieci neuronowych, również w połączeniu z modulem logiki rozmytej. Istnieją ponadto metody wykorzystujące kilka modeli sieci neuronowych do generacji jednej prognozy. Stosowane są różne metody doboru danych wejściowych sieci neuronowej oraz ich wartości opóźnionych. Często implementowany jest w sieci neuronowej nieliniowy model statystyczny NARX (Nonlinear Autoregressive Exogenous Model) [2]. Wykorzystuje on zmienne opóźnione endogeniczne (szereg czasowy mocy/produkcji energii elektrycznej) oraz dodatkowe zmienne egzogeniczne opóźnione oraz ich wartości z okresu prognozy (zmienne egzogeniczne to najczęściej zmienne wybrane z danych mających wysoką korelację z mocą prognozowaną). Najczęściej wykorzystywaną siecią neuronową w prognozowaniu ultrakrótkoterminowym i krótkoterminowym produkcji energii oraz mocy w systemach fotowoltaicznych jest sieć neuronowa typu MLP (Multilayer Perceptron) [1,3,4]. Dość popularny jest algorytm uczenia Levenberga-Marquardta dla sieci neuronowych typu MLP.

Istnieją ponadto opisane w literaturze prognozy z wykorzystaniem sieci neuronowej typu SVM (Support Vector Machine) [5], sieci neuronowych rekurencyjnych RNNs (Recurrent Neural Networks) np. sieć Elmana [4], sieci neuronowej typu RBF (Radial Base Function) [4], sieci neuronowych dynamicznych typu FTDNN (Focused Time Delay NN) [6] oraz sieci neuronowych dynamicznych typu DTDNN [6]. Zdecydowana większość opracowań zwraca

uwagę, że największe znaczenie dla uzyskania wysokiej jakości prognozy mocy lub produkcji energii elektrycznej w systemie fotowoltaicznym ma jakość prognoz czynników meteorologicznych (prognozy NWP) stanowiących zmienne egzogeniczne w modelach, a w szczególności precyzyjne prognozy poziomu nasłonecznienia [1]. Niektóre opracowania zwracają uwagę na konieczność integracji kilku prognoz lokalnych w celu poprawy jakości prognoz. Ponadto sugeruje się, również korzystać z podziału zadania prognostycznego na osobne modele dla poszczególnych miesięcy w roku [1]. Istnieją również opracowania np. [5] sugerujące, że korzystnie jest budować osobne modele dla kilku kategorii pogody na dzień prognozy.

Porównanie modelu fizycznego oraz modelu sieci neuronowej typu MLP dla prognoz mocy o horyzoncie 15 minut przedstawiono w [1]. Model fizyczny wykorzystujący konwersję poziomu nasłonecznienia (prognoza z modelu NWP) do wartości mocy prognozowanej uzyskał błąd nRMSE od 11% do 17% w zależności od miesiąca. Im bardziej niebo było bezchmurne, tym jakość prognoz była wyższa. W przypadku wykorzystania sieci neuronowej typu MLP błędy nRMSE były niższe (około 10,5%). Bardzo ciekawe były wyniki testów wykorzystania różnych kombinacji danych meteorologicznych (prognoz na okres prognozy mocy) na wejściu sieci neuronowej. Wyniki przedstawiono w tabeli 1.

Tabela 1. Wpływ wyboru wejść na wyniki prognoz sieci neuronowej typu MLP (opracowano na podstawie [1])

T	C	R	H	Ps	Liczba neuronów w warstwie ukrytej	Błąd nRMSE [%]
+	+	+	+	+	11	10,5
+	+	+	+		12	11,53
	+	+	+		15	12,20
	+		+		10	12,40
+		+		+	11	12,36
+		+	+	+	13	11,47
+	+	+			13	11,89

gdzie: T – temperatura, C – współczynnik zachmurzenia, R – nasłonecznienie, H – wilgotność, Ps – pozycja słońca na niebie, „+” – wykorzystanie danej egzogenicznej na wejściu sieci neuronowej.

Najlepszy zestaw danych wejściowych zawierał komplet informacji z prognoz NWP. Zwraca uwagę dość duże znaczenie pozycji słońca na jakość prognoz. Ponadto autorzy uważają, że trudno jest zbudować efektywny model

całoroczny – korzystniej podzielić jest zadanie prognostyczne na kilka osobnych modeli np. dla każdej z pór roku.

Analiza doboru zmiennych dla wybranych modeli prognostycznych

Wykonane analizy doboru zmiennych do modeli prognostycznych oraz testowe prognozy dotyczyły systemu fotowoltaicznego o mocy 19,5 kW w Centrum Fotowoltaiki Wydziału EiTI Politechniki Warszawskiej. Szczegółowej analizie statystycznej poddane zostały dostępne dane z okresów: 15.01.2010 - 9.03.2010, 6.07.2010 - 16.08.2010 oraz 5.09.2010 - 31.12.2010 [7]. W wyniku tej analizy wytypowano wstępnie 13 potencjalnych zmiennych objaśniających do modeli prognostycznych (tab.2). Weryfikacja istotności korelacji z tabeli 2, wykazała we wszystkich przypadkach ich istotność z $p < 0,05$ (5% poziom istotności). Warto zwrócić uwagę, że współczynniki korelacji liniowej przedstawiają siłę zależności liniowych pomiędzy porównywanymi danymi. W praktyce zależności pomiędzy danymi (zmienną objaśnianą oraz zmienną objaśniającą) dość często są nieliniowe. Wartości współczynnika korelacji liniowej nie mogą być więc decydującym i jedynym kryterium wyboru zmiennych do modeli prognostycznych. Istotne jest również aby zmienne objaśniające miały wysokie wartości współczynnika korelacji ze zmienną objaśnianą ale równocześnie niskie z innymi zmiennymi objaśniającymi [8].

Tabela 2 Współczynniki korelacji liniowej Pearsona wybranych potencjalnych zmiennych objaśniających z godzinową produkcją energii elektrycznej przez system fotowoltaiczny.

nr	Kod	Potencjalne zmienne objaśniające	Współczynnik korelacji
Egzogeniczne			
1	GODZ	Godzina prognozy (liczba z zakresu od 1 do 24)	0,102
2	PORA	Kod pory dnia (1 - okres pomiędzy wschodem a zachodem słońca, 0 – okres od zmroku do świtu)	0,479
3	OPAD	Opady [mm/h]	-0,080
4	WIATR	Szybkość wiatru [m/s]	0,101
5	CISN	Ciśnienie atmosferyczne [hPa]	0,121
6	TEMP	Temperatura [stopnie Celcjusza]	0,412
7	ZACH	Zachmurzenie [skala 0-8, brak zachmurz. - 0]	-0,296
8	NASŁ	Nasłonecznienie [Wh/m ²]	0,913
9	DŁUG	Długość dnia w dniu prognozy (liczba z zakresu od 0 do 1 określająca jaką część całego dnia stanowi okres pomiędzy wschodem a zachodem słońca)	0,319
Endogeniczne			
10	PT01	Produkcja energii elektrycznej w okresie t-1	0,915
11	PT02	Produkcja energii elektrycznej w okresie t-2	0,766
12	PT03	Produkcja energii elektrycznej w okresie t-3	0,580
13	PT24	Produkcja energii elektrycznej w okresie t-24	0,720

W tabeli 3 przedstawiono macierz współczynników korelacji liniowej pomiędzy potencjalnymi zmiennymi objaśniającymi. Prostą pomocniczą metodą doboru właściwych zmiennych objaśniających jest metoda analizy grafów (metoda Bartosiewicza). Dla zadanego poziomu istotności (5% poziom istotności) oraz 13 stopni swobody obliczono wartość rozkładu t-Studenta, która wyniosła 2,20. Następnie wyznaczono wartość krytyczną współczynnika korelacji, która wyniosła 0,553. W tabeli 3, współczynniki korelacji powyżej wartości krytycznej pogrubiono, natomiast

współczynniki korelacji statystycznie nieistotne zostały przekreślone. Na podstawie macierzy korelacji zbudowano graf powiązań między kandydatami na zmienne objaśniające (wykorzystując jedynie zmienne o wartościach współczynników korelacji powyżej wartości krytycznej). Powstały grafy izolowane z węzłów 1, 3, 4, 5 oraz 7. Z tych węzłów wybrano zmienną 7 ponieważ tylko ten węzeł miał wysoki współczynnik korelacji ze zmienną objaśnianą. Ponadto powstały graf powiązań pomiędzy węzłami 6 i 9 (z tego grafu wybrano zmienną 6 z uwagi na wyższy współczynnik korelacji ze zmienną objaśnianą) oraz graf powiązań pomiędzy węzłami 2, 8, 10, 11, 12, 13 (z tego grafu wybrano zmienną 8 ponieważ miała najwięcej (5) powiązań z innymi zmiennymi w grafie). Wybór zmiennych objaśniających do modelu prognostycznego wg ścisłych zasad metody grafów to zestaw tylko trzech zmiennych objaśniających: 6 (temperatura), 7 (zachmurzenie) oraz 8 (nasłonecznienie). Wybór nieścisły wg metody grafów to zmienne 6, 7, 8 oraz dodatkowa zmienna objaśniająca 10 (produkcja energii elektrycznej w okresie t-1). Zmienna 10 miała 4 powiązania z innymi zmiennymi i spośród tych zmiennych miała najwyższy współczynnik korelacji ze zmienną objaśnianą.

Tabela 3 Macierz współczynników korelacji liniowej pomiędzy potencjalnymi zmiennymi objaśniającymi.

	1	2	3	4	5	6	7	8	9	10	11	12	13
1	1,00												
2	0,16	1,00											
3	0,02	0,00	1,00										
4	0,08	0,10	0,06	1,00									
5	0,00	0,00	-0,11	-0,04	1,00								
6	0,09	0,29	0,09	-0,11	-0,15	1,00							
7	0,00	-0,04	0,14	0,16	-0,39	-0,22	1,00						
8	0,12	0,55	-0,07	0,05	0,10	0,51	-0,26	1,00					
9	0,00	0,23	0,10	-0,14	0,09	0,85	-0,32	0,45	1,00				
10	0,17	0,48	-0,08	0,11	0,11	0,43	-0,29	0,87	0,32	1,00			
11	0,24	0,46	-0,08	0,10	0,11	0,44	-0,28	0,75	0,32	0,92	1,00		
12	0,31	0,41	-0,07	0,08	0,10	0,43	-0,28	0,59	0,32	0,77	0,92	1,00	
13	0,10	0,48	-0,07	0,11	0,09	0,41	-0,19	0,83	0,32	0,78	0,67	0,52	1,00

Kolejnym krokiem były wykonanie prognoz wybranymi metodami. Całość danych została podzielona na dane do estymacji parametrów modeli (treningowe) oraz dane do weryfikacji jakości prognoz (testowe). Ze 180 dni wybrano do weryfikacji jakości prognoz następujące okresy: 4.03.2010 - 9.03.2010, 11.08.2010 - 16.08.2010, 5.09.2010 - 10.09.2010 oraz 26.12.2010 - 31.12.2010 (łącznie 24 dni).

Z uwagi na wartości zerowe występujące w szeregu czasowym produkcji energii elektrycznej jako podstawową miarę błędu przyjęto znormalizowany pierwiastek średniokwadratowego błędu – nRMSE (wzór 1 [8]).

$$(1) \quad nRMSE = \frac{\sqrt{\sum_{i=1}^T (y_t^* - y_t)^2}}{E_n \cdot \sqrt{T}}$$

gdzie: y_t^* - prognoza produkcji energii, y_t - wartość rzeczywista produkcji energii, E_n - godzinowa produkcja energii przy pracy systemu fotowoltaicznego z mocą nominalną.

W metodzie odniesienia, którą była metoda naiwna, prognozę stanowiła wartość produkcji energii elektrycznej z ostatniej godziny sprzed godziny prognozy (przyjęto założenie, że jakość prognoz bardziej zaawansowanymi metodami powinna być wyższa niż jakość metody odniesienia).

Pierwszą metodą testowaną była metoda ekonometryczna (model jednorównaniowy liniowy

dynamiczny - regresja wieloraka). W pierwotnym modelu regresji wielorakiej wykorzystano wszystkie 13 zmienne objaśniające, zakładając w kolejnym etapie usunięcie tych, które będą statystycznie nieistotne. Weryfikacji poddano również jakość modelu z wykorzystaniem 2 wersji zestawów danych dobranych metodą analizy grafów (wybór ścisły oraz wybór nieściśły). Optymalizacja parametrów modeli wykonana była metodą Newtona (funkcja celu – błąd nRMSE) wykorzystując narzędzie Solver w arkuszu kalkulacyjnym Excel oraz klasyczną metodą najmniejszych kwadratów (funkcja celu – suma kwadratów odchyłeń wartości rzeczywistych i prognoz) w programie symulacyjnym Gretl. Wykonane prognozy miały dostarczyć z jednej strony informacji o statystycznej istotności zmiennych objaśniających ale również uszeregować ich ważność wg miar: współczynnika ważności (wzór 2 [8]) oraz wartości statystyk t-Studenta. W przypadku gdy istnieje pewność, że parametry modelu są statystycznie istotne czyli zmienne istotnie wpływają na zmienną objaśnianą można obliczyć, która ze zmiennych ma względnie najsilniejszy wpływ na kształtowanie się wartości zmiennej objaśnianej. Wykorzystać można do tego celu wspomniany współczynnik ważności zmiennej w modelu – relatywny wpływ zmiennej objaśniającej na zmienną objaśnianą. Im większa wartość tego współczynnika, tym relatywnie większy wpływ zmiennej objaśniającej na zmienną objaśnianą.

$$(2) \quad b_i = |a_i| \cdot \frac{\bar{x}_i}{\bar{y}}$$

gdzie: a_i - estymator (szacunek parametru), $\bar{x}_i$ - średnia wartość zmiennej objaśniającej przy której stoi parametr, $\bar{y}$ - średnia empiryczna wartość zmiennej objaśnianej.

W tabeli 4 przedstawiono wyniki prognoz metodą regresji wielorakiej oraz metodą naiwną wartości godzinowej produkcji energii elektrycznej. W nawiasach podano numery wykorzystanych zmiennych objaśniających zgodnie z tabelą 2. W metodzie naiwnej prognozę stanowiła wartość produkcji energii w okresie t-1.

Tabela 4. Wybrane wyniki prognoz metodą regresji wielorakiej godzinowej produkcji energii elektrycznej przez system fotowoltaiczny z horyzontem 1 godziny.

Metoda prognostyczna	Miara błędu	Zakres estymacji	Zakres weryfikacji
Metoda naiwna (10)	nRMSE [%]	6,930	7,633
	wsp.korelacji	0,915	0,928
Regresja wieloraka 1a (1,2,3,4,5,6,7,8,9,10,11,12,13)	nRMSE [%]	4,799	4,364
	wsp.korelacji	0,958	0,976
Regresja wieloraka 1b (3,4,5,6,7,8,9,10,11,12,13)	nRMSE [%]	4,799	4,362
	wsp.korelacji	0,958	0,976
Regresja wieloraka 1c (3,4,5,6,7,8,9,10,11,12,13)	nRMSE [%]	4,798	4,367
	wsp.korelacji	0,958	0,976
Regresja wieloraka 1d (6,7,8)	nRMSE [%]	6,742	5,934
	wsp.korelacji	0,916	0,956
Regresja wieloraka 1e (6,7,8,10)	nRMSE [%]	5,328	5,256
	wsp.korelacji	0,948	0,965
Regresja wieloraka 2a (1,2,3,4,5,6,7,8,9,10,11,12,13)	nRMSE [%]	4,732	4,405
	wsp.korelacji	0,960	0,976
Regresja wieloraka 2b (1,4,5,6,7,8,9,10,11,12,13)	nRMSE [%]	4,737	4,357
	wsp.korelacji	0,959	0,976

Gdzie: regresja wieloraka 1 – optymalizacja parametrów metodą Newtona, regresja wieloraka 2 – optymalizacja parametrów metodą KMNK.

Modele regresji wielorakiej 1a oraz 2a zawierają wszystkie 13 zmienne objaśniające. W przypadku modelu regresji wielorakiej 1b usunięto mechanicznie z modelu (bez ponownego ustalania jego parametrów) 3 najmniej

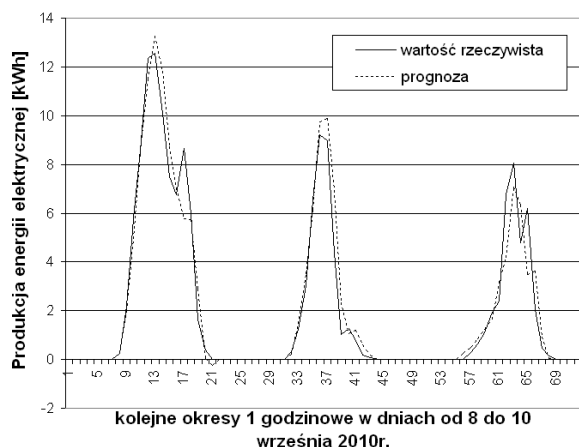
ważne zmienne objaśniające na wg rankingu ważności zmiennych (obliczone współczynniki ważności zmiennych). W przypadku modelu regresji wielorakiej 1c usunięto z modelu (ponownie ustalając jego właściwe parametry) 3 najmniej ważne zmienne objaśniające na wg rankingu ważności zmiennych. W przypadku modelu regresji wielorakiej 1d wykorzystano tylko 3 zmienne objaśniające (ściśle dobór metodą analizy grafów), natomiast w przypadku modelu regresji wielorakiej 1e wykorzystano 4 zmienne objaśniające (nieściśle dobór metodą analizy grafów). Model regresji wielorakiej 2b zawiera 11 zmiennych objaśniających (zmienne objaśniające 2 oraz 3 okazały się wg testów statystycznie nieistotne).

W tabeli 5 przedstawiono wyniki prognoz wykonanych siecią neuronową typu MLP z wykorzystaniem 3 różnych algorytmów uczących (algorytm Quasi-Newtona, algorytm Levenberga-Marquardta oraz algorytm szybkiej propagacji). W nawiasach podano numery wykorzystanych zmiennych objaśniających (wejścia sieci neuronowej). Wszystkie sieci neuronowe miały 1 warstwę ukrytą z sigmoidalną funkcją aktywacji. W warstwie wyjściowej wykorzystano liniową funkcję aktywacji. W przypadku wykorzystania na wejściu wszystkich 13 zmiennych objaśniających, sieć neuronowa zawierała 15 neuronów w warstwie ukrytej (struktura 13-15-1). Wykorzystując analizę wrażliwości wykonanej dla 3 sieci neuronowych z różnymi algorytmami uczącymi wykluczono 3 najmniej ważne zmienne objaśniające (były to najczęściej występujące zmienne na końcowych pozycjach w 3 rankingach analizy wrażliwości czyli zmienne objaśniające 3, 4 oraz 5). W przypadku wykorzystania na wejściu tylko 10 zmiennych objaśniających (3 usunięto), sieć neuronowa zawierała 13 neuronów w warstwie ukrytej (struktura 10-13-1). Liczba cykli uczących zależna była od algorytmu uczącego. Najmniej cykli uczących wymagał algorytm Levenberga-Marquardta (około 75), nieco więcej algorytm Quasi-Newtona (około 250), a najwięcej algorytm szybkiej propagacji (około 400).

Tabela 5. Wybrane wyniki prognoz sieci neuronowej typu MLP produkcji energii elektrycznej przez system fotowoltaiczny z horyzontem 1 godziny.

Metoda prognostyczna	Miara błędu	Zakres estymacji	Zakres weryfikacji
Sieć neuronowa 1a (alg uczący: Quasi-Newtona) (1,2,3,4,5,6,7,8,9,10,11,12,13)	nRMSE [%]	4,276	3,707
	wsp.korelacji	0,967	0,983
Sieć neuronowa 1b (alg uczący: Quasi-Newtona) (1,2,6,7,8,9,10,11,12,13)	nRMSE [%]	3,938	3,717
	wsp.korelacji	0,972	0,983
Sieć neuronowa 2a (alg uczący: Levenberga-Marquardta) (1,2,3,4,5,6,7,8,9,10,11,12,13)	nRMSE [%]	3,957	3,750
	wsp.korelacji	0,972	0,983
Sieć neuronowa 2b (alg uczący: Levenberga-Marquardta) (1,2,6,7,8,9,10,11,12,13)	nRMSE [%]	4,040	3,746
	wsp.korelacji	0,971	0,983
Sieć neuronowa 3a (alg uczący: szybka propagacja) (1,2,3,4,5,6,7,8,9,10,11,12,13)	nRMSE [%]	4,180	3,778
	wsp.korelacji	0,969	0,982
Sieć neuronowa 3b (alg uczący: szybka propagacja) (1,2,6,7,8,9,10,11,12,13)	nRMSE [%]	4,456	4,017
	wsp.korelacji	0,964	0,980

Na rysunku 1 przedstawiono Przykładowe prognozy produkcji energii elektrycznej z horyzontem 1 godziny przy użyciu sieci neuronowej typu MLP (sieć neuronowa 1a) w okresie od 8 do 10 września 2010r.



Rys. 1 Przykładowa prognoza produkcji energii elektrycznej z horyzontem 1 godziny przy użyciu sieci neuronowej typu MLP (sieć neuronowa 1a)

Analiza ważności zmiennych objaśniających dla wybranych modeli prognostycznych

W tabeli 6 przedstawiono rankingi ważności poszczególnych zmiennych objaśniających dla sieci neuronowych typu MLP na podstawie wykonanej analizy wrażliwości (zakres danych treningowych) oraz ranking ważności na podstawie wartości współczynników korelacji liniowej Pearsona pomiędzy zmienną objaśnianą a poszczególnymi potencjalnymi zmiennymi objaśniającymi (kolumna – współczynnik korelacji). W wyniku analizy wrażliwości dla sieci neuronowych 1a, 1b oraz 1c usunięto zmienne: opad, szybkość wiatru oraz ciśnienie i wykonano powtórnie analizę wrażliwości dla zestawu 10 zmiennych objaśniających.

W tabeli 7 przedstawiono rankingi ważności poszczególnych zmiennych objaśniających dla modeli regresji wielorakiej (współczynniki ważności). W przypadku metody - regresja 2a, zmienne kod pory dnia oraz opady były nieistotne statystycznie. W tabelach 6 oraz 7 pogrubiono w każdej kolumnie z kompletem 13 zmiennych 3 najlepsze zmienne wg rankingu ważności, natomiast pola 3 najgorszych zmiennych mają kolor szary.

Tabela 6. Rankingi ważności zmiennych objaśniających (analiza wrażliwości dla sieci neuronowych oraz współczynniki korelacji).

Kod zmiennej	Pozycja w rankingu ważności zmiennych						
	Współczynnik korelacji	Sieć 1a	Sieć 2a	Sieć 3a	Sieć 1b	Sieć 2b	Sieć 3b
GODZ	11	5	4	5	5	5	10
PORA	5	6	5	3	4	4	8
OPAD	13	12	13	13	-	-	-
WIATR	12	11	11	10	-	-	-
CISN	10	13	12	12	-	-	-
TEMP	6	9	9	11	8	10	9
ZACH	9	8	10	8	7	9	7
NASŁ	2	2	1	1	1	1	2
DŁUG	8	3	3	4	3	3	5
PT01	1	1	2	2	2	2	1
PT02	4	7	8	6	10	8	3
PT03	7	10	7	9	6	6	4
PT24	3	7	6	7	9	7	6

Tabela 7. Rankingi ważności zmiennych objaśniających dla modeli regresji wielorakiej (współczynniki ważności).

Kod zmiennej	Pozycja w rankingu ważności zmiennych					
	Regresja 1a	Regresja 1c	Regresja 1d	Regresja 1e	Regresja 2a	Regresja 2b
GODZ	11	-	-	-	11	11
PORA	12	-	-	-	13	-
OPAD	13	-	-	-	12	-
WIATR	9	9	-	-	9	8
CISN	3	3	-	-	10	4
TEMP	10	10	3	3	8	10
ZACH	4	5	2	2	4	5
NASŁ	2	2	1	4	2	3
DŁUG	7	7	-	-	5	2
PT01	1	1	-	1	1	1
PT02	6	6	-	-	7	6
PT03	5	4	-	-	3	7
PT24	8	8	-	-	6	9

Dobór zmiennych objaśniających został również wykonany przy użyciu algorytmu selekcji krokowej postępującej (Forward selection), algorytmu selekcji krokowej wstecznej redukującej (Backward selection) oraz algorytmu genetycznego w programie symulacyjnym Statistica neural networks. Wszystkie 3 algorytmy wskazały komplet 13 zmiennych jako właściwe (najmniejsze miary błędów). Algorytmy te nie wskazywały kolejności ważności zmiennych tylko listę wybranych zmiennych objaśniających.

Wnioski

Na podstawie przeprowadzonych analiz sformułować można szereg wniosków i sugestii odnośnie prognoz analizowanego szeregu czasowego oraz doboru zmiennych objaśniających i ich ważności:

- wnioski związane z wykorzystaniem modelu regresji wielorakiej:
 - szacowanie właściwych parametrów metodą Newtona oraz metodą KMNK dają niemal identyczne błędy prognoz,
 - wykorzystanie do doboru zmiennych metody analizy grafów w wersji ścisłej oraz nieścisłej zwiększa znacznie błędy prognoz w stosunku do modeli z większą liczbą parametrów (11, 12 lub 13 parametrów),
 - wykorzystanie do doboru zmiennych metody analizy grafów w wersji nieścisłej (4 parametry zamiast 3) zmniejsza nieco błędy prognoz w stosunku do wersji ścisłej,
 - wykorzystanie zbyt małej liczby zmiennych objaśniających w modelu jest niekorzystne (np. trzech zamiast trzynastu) i znacznie zwiększa błędy prognoz, ich wielkość jest zbliżona do błędów najprostszej metody naiwnej,
 - usunięcie w modelu „regresja wieloraka 2” dwóch zmiennych nieistotnych statystycznie (kod pory dnia, opady) nie zmieniło w praktyce błędów prognoz (minimalna poprawa a zakresie weryfikacji i minimalne pogorszenie w zakresie estymacji parametrów modelu),
 - w przypadku modelu „regresja wieloraka 1” usunięcie 3 zmiennych najmniej ważnych (godzina prognozy, kod pory dnia, opady) nie wpłynęło na jakość prognoz,
 - dla zdecydowanej większości przypadków w modelu regresji wielorakiej najważniejszą zmienną objaśniającą była produkcja energii w okresie t-1, natomiast do 3 najmniej ważnych zmiennych objaśniających należały: opad, kod pory dnia oraz godzina prognozy,

- rodzaj algorytmu optymalizacyjnego do doboru wartości parametrów modeli regresji nie miał wpływu na kolejność ważności zmiennych.
- b. wnioski związane z wykorzystaniem sieci neuronowych typu MLP:
- poszczególne algorytmy uczące dla metody wykorzystującej sieć neuronową typu MLP generowały dość zbliżone jakościowo prognozy (ich kolejność wg jakości prognoz „ex-post” to: algorytm uczący Quasi-Newtona, algorytm Levenberga-Marquardta, algorytm szybkiej propagacji),
 - usunięcie 3 zmiennych objaśniających (opady, wiatr, ciśnienie atmosferyczne) na podstawie analizy wrażliwości wykonanej dla sieci neuronowych nie wpłynęło w praktyce na jakość prognoz (dla algorytmów uczących Quasi-Newtona oraz szybkiej propagacji nastąpiło minimalne pogorszenie jakości prognoz, natomiast dla algorytmu Levenberga-Marquardta nastąpiła minimalna poprawa w zakresie prognoz „ex post”,
 - usunięcie 3 zmiennych objaśniających (opady, wiatr, ciśnienie atmosferyczne) na podstawie analizy wrażliwości nie wpłynęło znacząco na kolejność ważności zachowanych w modelu zmiennych objaśniających,
 - kolejność ważności zmiennych objaśniających nie była związana z algorytmem uczącym sieci neuronowych,
 - zmienna objaśniająca – godzina prognozy była istotną daną dla sieci neuronowych pomimo bardzo niskiej wartości współczynnika korelacji (0,1) ze zmienną objaśnianą, natomiast zmienne objaśniające opady oraz szybkość wiatru były na końcu rankingu zarówno dla analizy wrażliwości jak i dla rankingu opartego na współczynnikach korelacji,
 - zmienna objaśniająca nasłonecznienie oraz produkcja energii w okresie t-1 były najważniejszymi zmiennymi objaśniającymi w rankingu analizy wrażliwości oraz rankingu opartego na współczynnikach korelacji,
 - dla zdecydowanej większości sieci neuronowych najważniejszą zmienną objaśniającą była produkcja energii w okresie t-1 oraz nasłonecznienie, natomiast większość wyników analizy wrażliwości dla sieci neuronowych wykazała, że 3 zmienne (opady, wiatr, ciśnienie atmosferyczne) mają wartości ilorazów błędów uzyskanych przy uruchomieniu sieci dla zbioru danych bez jednej zmiennej i błędów uzyskanych z kompletem zmiennych nieco poniżej 1 czyli usunięcie tych zmiennych nie ma wpływu na jakość prognoz (ewentualnie mogłoby je poprawić, co jednak nie nastąpiło).
- c. wnioski o charakterze ogólnym:
- metoda naiwna uzyskała zdecydowanie największe błędy prognoz,
 - metoda wykorzystująca sztuczne sieci neuronowe typu MLP (nieliniowy estymator) okazała się znacznie korzystniejsza niż model liniowy regresji wielorakiej,
 - nadmierne usuwanie potencjalnych zmiennych objaśniających może silnie zwiększyć błędy prognoz (maleje potencjał informacyjny),
 - usuwanie wstępnie wytypowanych zmiennych objaśniających wymaga dużej ostrożności i analizy wpływu ich usunięcia na jakość prognoz [8,9],
 - metoda analizy grafów do doboru zmiennych objaśniających okazała się niekorzystna – błędy prognoz znacznie wzrosły w stosunku do modeli ze znacznie większą liczbą zmiennych objaśniających,
 - ważność poszczególnych zmiennych objaśniających dla modelu regresji wielorakiej oraz sieci neuronowych typu MLP wykazuje dość duże różnice, przykładowo dla regresji wielorakiej zmienna objaśniająca – godzina prognozy była mało ważna, a dla sieci neuronowej bardzo ważna, z kolei zmienna objaśniająca – ciśnienie atmosferyczne była ważna dla modelu regresji wielorakiej a mało ważna dla sieci neuronowych,
 - właściwe zmienne objaśniające i kolejność ich ważności mogą być różne w zależności od wykorzystywanej metody prognostycznej,
 - ranking ważności zmiennych objaśniających na podstawie współczynników korelacji dobrze wskazał najważniejsze zmienne objaśniające (duża zbieżność z analizą wrażliwości), natomiast gorsze były jego sugestie odnośnie zmiennych najmniej ważnych (np. zmienne godzina oraz długość dnia okazały się bardzo ważne w przypadku prognoz siecią neuronową typu MLP),
 - pomimo wielu badań zagranicznych z zakresu prognozowania krótkoterminowego systemów fotowoltaicznych niezbędne wydaje się wykonywanie analiz na danych lokalnych (krajowych) z uwagi na bardzo indywidualny charakter prognozowanych szeregów czasowych (różnice klimatyczne, odmienne zestawy danych objaśniających wykorzystane do prognoz).
- Autor dziękuje firmie Globema za udostępnienie danych w ramach projektu 4RES.*

LITERATURA

- [1] Yuehui H. et al.: „Comparative study of power forecasting methods for PV stations”, International Conference on Power System Technology (POWERCON), 2010,
- [2] Cococcioni M., et al.: „24-hour-ahead forecasting of energy production in solar PV systems”, 11th International Conference on Intelligent Systems Design and Applications (ISDA), 2011,
- [3] Baczyński D., Wasilewski J.: „Krótkoterminowe prognozowanie produkcji energii elektrycznej w systemach fotowoltaicznych”, Rynek Energii, 96 (2011), nr 5, 47-51,
- [4] Yona A., et al.: „Application of neural network to 24-hour-ahead generating power forecasting for PV system”, Power and Energy Society General Meeting - Conversion and Delivery of Electrical Energy in the 21st Century, 2008 IEEE,
- [5] Jie Shi, et al.: „Forecasting power output of photovoltaic system based on weather classification and support vector machine”, Industry Applications Society Annual Meeting (IAS), 2011 IEEE, Orlando 2011,
- [6] Al-Messabi N., et al.: „Forecasting of photovoltaic power yield using dynamic neural networks”, The 2012 International Joint Conference on Neural Networks (IJCNN),
- [7] Baczyński D., Helt P., Marzecki J., Piotrowski P., Wasilewski J.: „Opracowanie metod prognozowania produkcji energii elektrycznej dla źródeł odnawialnych”, praca na zlecenie firmy Globema w ramach umowy z Instytutem Elektroenergetyki Politechniki Warszawskiej,
- [8] Piotrowski P.: „Prognozowanie w elektroenergetyce w różnych horyzontach czasowych”, Monografia, Prace Naukowe Politechniki Warszawskiej „Elektryka”, z. 144, Oficyna Wydawnicza Politechniki Warszawskiej, Warszawa, 2013,
- [9] Piotrowski P.: „Prognozowanie krótkoterminowe godzinowych obciążeń w spółce dystrybucyjnej z wykorzystaniem sieci neuronowych – analiza wpływu doboru i przetworzenia danych na jakość prognoz”, Przegląd Elektrotechniczny (Electrical Review), (83) 2007, nr 7-8, 40-43.

Autor: dr inż. Paweł Piotrowski, Politechnika Warszawska, Instytut Elektroenergetyki, ul. Koszykowa 75, 00-662 Warszawa, E-mail: pawel.piotrowski@ien.pw.edu.pl