

Ekonometryczny model krótkoterminowego prognozowania zużycia gazu

Econometric model of short-term natural gas consumption forecasting

Tadeusz Kwilosz, Bogdan Filar

Instytut Nafty i Gazu – Państwowy Instytut Badawczy

STRESZCZENIE: W celu opracowania modelu krótkoterminowego zapotrzebowania na gaz konieczne jest dokonanie analizy najnowszych metod prognozowania matematycznego w celu wyboru i zaadaptowania właściwej z nich (spełniającej warunek efektywności i skuteczności). Niezbędne jest rozpoznanie i analiza czynników (głównie środowiskowych) wpływających na wynik krótkoterminowych prognoz oraz źródeł danych możliwych do wykorzystania. Efektem wykonanej pracy jest model numeryczny krótkoterminowego zapotrzebowania na gaz dla wybranej jednostki terytorialnej kraju. Opracowany model został skalibrowany i przetestowany na historycznych danych opisujących warunki środowiskowe i rzeczywiste zużycie gazu. Zaprojektowano i skalibrowano, na podstawie wybranego zestawu atrybutów (zmiennych objaśniających), niejednorodny liniowy model ekonometryczny. Dokonano statystycznej weryfikacji oszacowanych parametrów modelu. Warto zauważyć, że w krótkim terminie wykonania prognozy (7 dni) nie zachodzą znaczące zmiany w otoczeniu rynku gazowego (uruchomienie nowych inwestycji, podłączenie nowych użytkowników do systemu czy zmiany zapotrzebowania wynikające ze zmieniających się warunków makroekonomicznych). Inne czynniki techniczne, takie jak awarie linii produkcyjnych u odbiorców czy przestoje przemysłowe, są trudne do przewidzenia lub wiedza o nich rzadko jest dostępna. Z tego względu jedynymi czynnikami mogącymi mieć wpływ na zmiany zapotrzebowania gazu w krótkim terminie są czynniki pogodowe, które zostały wybrane jako zmienne objaśniające dla opracowanego modelu. Historyczne dane pogodowe zostały pobrane z usługi sieciowej (*web service*) *OpenWeatherMap History Bulk*. Jako zmiennej objaśnianej użyto dobowych wartości zużycia gazu dla jednego z województw Polski południowej. Dane zostały pobrane z systemu wymiany informacji operatora gazociągów przesyłowych. Dane dotyczą okresu trzyletniego, gdyż tylko takie dane zostały upublicznione. Zmienne objaśniające obejmują dobowe wartości danych pogodowych, takich jak: średnia temperatura, temperatura odczuwalna, temperatura minimalna, temperatura maksymalna, ciśnienie atmosferyczne, wilgotność względna, prędkość wiatru i kierunek wiatru.

Słowa kluczowe: model ekonometryczny, krótkoterminowe prognozowanie, zużycie gazu.

ABSTRACT: In order to develop a mathematical model of short-term gas demand, it is necessary to analyze the latest mathematical forecasting methods in order to select and adapt the right one (meeting the condition of efficiency and effectiveness). It is necessary to recognize and analyze factors (mainly environmental) affecting the result of short-term forecasts and sources of data that can be used. The result of the work is a numerical model of short-term gas demand for a selected territorial unit of the country. The developed model was calibrated and tested on historical data describing environmental conditions and real gas consumption. A heterogeneous linear econometric model was designed and calibrated on the basis of a selected set of attributes (explanatory variables). The estimated parameters of the model were statistically verified. It is worth noting that in the short term of the forecast (7 days) there are no significant changes in the gas market environment (launching new investments, connecting new users to the system, or changes in demand resulting from changing macroeconomic conditions). Other technical factors, such as production line failures at customers or industrial downtime, are difficult to predict, or knowledge about their occurrence is rarely available. For this reason, the only factors that may have an impact on changes in gas demand in the short term are weather factors, which were selected as explanatory variables for the developed model. Historical weather data was retrieved from the *OpenWeatherMapHistoryBulk* web service. Daily values of gas consumption for one of the voivodships of southern Poland were used as the response variable. The data was downloaded from the information exchange system of the transmission pipeline operator. The data covers a three-year period, as only such data has been made public. The explanatory variables include the daily values of weather data such as: average temperature, chilled temperature, minimum temperature, maximum temperature, atmospheric pressure, relative humidity, wind speed and wind direction.

Key words: econometric model, short-term forecasting, natural gas consumption.

Wstęp

W ostatnich latach obserwowany jest wzrost zużycia gazu w kraju (o około 1 mld m³ rocznie). Problem racjonalnego (pod względem rytmiczności i ciągłości dostaw) dostarczania gazu do odbiorców staje się ważny z przyczyn ekonomicznych oraz wymogów bezpieczeństwa energetycznego. W celu spełnienia tak postawionego zadania konieczna jest wiedza o bieżącym zapotrzebowaniu na gaz w rozbiciu na jednostki terytorialne kraju (Paliński i Łucki, 2006; Soldo, 2012; Rogowska, 2017, 2018; Siewierski et al., 2017; Tamba et al., 2018). Podstawowym sposobem uzyskania takiej wiedzy jest wykonywanie możliwie dokładnych krótkoterminowych prognoz (do 7 dni) zapotrzebowania na gaz. Skuteczną metodą wykonywania takich prognoz jest posługiwanie się opracowanym numerycznym modelem zużycia gazu. W celu opracowania modelu zapotrzebowania na gaz konieczne jest dokonanie analizy najnowszych metod prognozowania matematycznego w celu wyboru i zaadaptowania właściwej (spełniającej warunek efektywności i skuteczności).

Wybór modelu i podstawowe założenia

Z przeglądu literaturowego związanego z krótkoterminowym prognozowaniem zużycia energii wynika, że autorzy opracowań, publikacji i wdrożonych aplikacji stosowali różne rozwiązania badanego problemu:

- sieci neuronowe (Cieślak, 1996; Kelner, 2003; Wójcik, 2005; Maciejasz, 2006; Kogut, 2007; Jeżyk i Tomczewski, 2014; Szoplik, 2015; Cieślak i Kogut, 2016; Zheng et al., 2017);
- adaptacyjne metody regresyjne (Sokołowski i Pasztyła, 2004; Czapaj et al., 2019);
- algorytmy genetyczne (Yu i Xu, 2014);
- liniowe modele ekonometryczne (Łucki, 1988; Zieliński, 1995; Gawlik, 2008; Łucki et al., 2008; Bianco et al., 2014a, 2014b).

Większość opracowań oparta jest na algorytmach należących do szeroko pojmowanej grupy metod sztucznej inteligencji, w szczególności sieci neuronowych. Niejednokrotnie wybór metody zdeterminowany jest ograniczeniami związanymi z dostępem do wiarygodnych danych. Wybór modelu ekonometrycznego zastosowany w tym opracowaniu podyktowany był głównie ograniczonym dostępem do dobowych danych zużycia gazu dla danego obszaru kraju. Tematem artykułu jest próba opracowania ekonometrycznego modelu krótkoterminowego prognozowania zużycia gazu na podstawie bieżących prognoz pogody (Zeliaś et al., 2003).

Wybrano liniowy model ekonometryczny. W pierwszej kolejności zdefiniowano zmienną objaśnianą opisującą dobowe

zużycie gazu [tys. m³] oraz zmienne objaśniające opisujące czynniki pogodowe. Wzorując się na danych literaturowych, dodano do modelu zmienną, która wyraża bazowy dzień tygodnia wykonywania prognozy. Dokonano rangowania i roboczym dniom tygodnia przypisano wartość 1, sobotom – wartość 2, a niedzielom i świętom – wartość 3. Eksperymenty numeryczne wykonane na modelu z opisywaną zmienną nie wykazały żadnego istotnego wpływu tej zmiennej na wartość prognozy i dlatego została ona usunięta z modelu. Metoda wyboru zbioru danych historycznych, służących do kalibracji modelu, została opisana w dalszej części opracowania. W tym miejscu warto zaznaczyć, że zadbano o to, aby zalecane proporcje między liczbą parametrów modelu (k) a liczbą danych (n) były zachowane. Autorzy literatury przedmiotu wskazują, że liczba zmiennych modelu (szczególnie w przypadku modeli liniowych) nie może przekraczać liczebności serii danych (Wydymus, 1984; Szmigiel i Mercik, 2000). W rzeczywistości powinno być ich istotnie mniej, gdyż różnica pomiędzy liczbą danych (n) a liczbą zmiennych (k), zwana liczbą stopni swobody modelu, ma istotny wpływ na statystyczną weryfikację poprawności oszacowanych parametrów (Welfe, 1988; Osińska, 2006; Gładysz i Mercik, 2007; Goryl et al., 2009).

Niejednorodny liniowy model ekonometryczny z wieloma zmiennymi objaśniającymi można zapisać w postaci (Zeliaś et al., 2003):

$$y_t = \alpha_0 + \alpha_1 X_{t1} + \alpha_2 X_{t2} + \dots + \alpha_k X_{tk} + \varepsilon_t$$

gdzie:

y_t – zmienna objaśniana o wartościach empirycznych w okresie t ,

X_{tj} – zmienna objaśniająca o wartościach empirycznych w okresie t ,

α_j – nieznanne parametry modelu,

ε_t – składnik losowy.

Powyższy model można zapisać w postaci macierzowej:

$$\mathbf{y} = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix} \mathbf{X} = \begin{bmatrix} 1 & x_{11} & \dots & x_{1k} \\ 1 & x_{21} & \dots & x_{2k} \\ \vdots & \vdots & \ddots & \vdots \\ 1 & x_{n1} & \dots & x_{nk} \end{bmatrix} \boldsymbol{\alpha} = \begin{bmatrix} \alpha_0 \\ \alpha_1 \\ \vdots \\ \alpha_k \end{bmatrix} \boldsymbol{\varepsilon} = \begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \vdots \\ \varepsilon_n \end{bmatrix}$$

lub

$$y = X\alpha + \varepsilon$$

Po oszacowaniu parametrów strukturalnych model przyjmuje postać:

$$\hat{y}_t = a_0 + a_1 X_{t1} + a_2 X_{t2} + \dots + a_k X_{tk}$$

gdzie:

\hat{y}_t – teoretyczna wartość zmiennej objaśnianej w okresie t ,
 a_j – wyliczony estymator nieznanego parametru α_j .

Wyrażenie:

$$e_t = y_t - \hat{y}_t$$

nazywane jest resztą modelu.

Oszacowanie nieznanymi parametrów modelu przeprowadza się metodą najmniejszych kwadratów, polegającą na wyznaczeniu minimum sumy kwadratów reszt modelu:

$$\sum_{t=1}^n e_t^2 \rightarrow \min$$

Funkcja ta osiąga minimum w punkcie:

$$a = (X^T X)^{-1} X^T y$$

Po wstępnym wyborze zmiennych objaśniających modelu należy dokonać ich oceny i ewentualnej eliminacji. Czynnikiem oceny jakości zmiennych objaśnianych jest ich współczynnik zmienności:

$$V = \frac{s}{\bar{X}} 100\%$$

gdzie:

s – odchylenie standardowe z próby,

\bar{X} – średnia z próby.

Przyjmuje się arbitralnie, że wartość ta powinna przekraczać 10%.

Zasadniczym kryterium doboru zmiennych objaśniających modelu jest metoda pojemności informacyjnej Hellwiga. Polega ona na:

- wyznaczeniu wszystkich $k = 2^{n-1}$ podzbiorów zbioru zmiennych objaśniających,
- przypisaniu każdemu z tak wyznaczonych podzbiorów (k) integralnej pojemności nośników informacyjnych:

$$H_k = \sum_j h_{kj}$$

gdzie:

$h_{kj} = \sum_i \frac{r_{0j}^2}{|r_{ij}|}$ – indywidualna pojemność j -tej zmiennej w k -tej kombinacji (podzbiórze),

r_{0j} – współczynnik korelacji j -tej zmiennej objaśniającej ze zmienną objaśnianą,

r_{ij} – współczynnik korelacji j -tej zmiennej objaśniającej z i -tą zmienną występującą w tej kombinacji (zbiorze).

Parametr H_k przyjmuje wartość z przedziału $\langle 0, 1 \rangle$. Im jego wartość jest większa, tym wytypowana kombinacja zmiennych lepiej opisuje modelowane zjawisko. Na użytek tej pracy przyjęto, że wartość H_k nie może być mniejsza od 0,4.

Po wyznaczeniu parametrów strukturalnych modelu należy dokonać jego weryfikacji. Służy do tego celu szereg parametrów.

1. Miarą dopasowania oszacowanego modelu do danych empirycznych jest współczynnik determinacji:

$$R^2 = 1 - \frac{\sum_{t=1}^n e_t^2}{\sum_{t=1}^n (y_t - \bar{y})^2}$$

gdzie:

\bar{y} – średnia arytmetyczna z serii danych empirycznych zmiennej objaśnianej. Współczynnik ten przyjmuje wartości z przedziału $\langle 0, 1 \rangle$ i informuje, jaka część całkowitej zaobserwowanej zmienności zmiennej objaśniającej została wyjaśniona przez model.

2. Nieobciążonym estymatorem wariancji składnika losowego jest wariancja resztowa:

$$S_e^2 = \frac{\sum_{t=1}^n e_t^2}{n - k - 1}$$

Pierwiastek kwadratowy z wariancji resztowej S_e nazywany jest odchyleniem resztowym modelu. Informuje on, o ile średnio wartości teoretyczne modelu różnią się od wartości empirycznych zmiennej objaśniającej.

3. Współczynnik zmienności resztowej informuje, o ile średnio wartości zmiennej objaśnianej stanowią odchylenia losowe.

$$V_e = \frac{S_e}{\bar{y}} 100\%$$

4. Zakłada się, że estymatory a_j parametrów α_j mają rozkład normalny.

$$a_j : N(a_j, D(a_j))$$

Estymatorem nieznanego odchylenia standardowego $D(a_j)$ jest:

$$S(a_j) = S_e \sqrt{c_{jj}}$$

gdzie:

c_j – elementy głównej przekątnej macierzy $(X^T X)^{-1}$. Mając oszacowane średnie błędy estymatorów, można zapisać:

$$a_j = a_j + -S(a_j)$$

5. Miarą oceny trafności doboru poszczególnych zmiennych objaśniających jest test istotności pojedynczej zmiennej objaśniającej przeprowadzany w formie weryfikacji hipotezy statystycznej.

Zakładamy hipotezę zerową $H_0: \alpha_j = 0$ oznaczającą, że zmienna x_j nie ma wpływu na wartość zmiennej y , przeciw hipotezie alternatywnej $H_1: \alpha_j \neq 0$ oznaczającej, że zmienna x_j ma wpływ na wartość zmiennej y . Sprawdzianem hipotezy jest ocena statystyki

$$t(a_j) = \frac{a_j}{S(a_j)}$$

W przypadku gdy:

$$|t(a_j)| > t_{\alpha, Q}$$

odrzucaamy hipotezę zerową (H_0) na rzecz hipotezy alternatywnej (H_1), tzn. zmienna x_j ma istotny (statystycznie) wpływ na zmienną y . W przeciwnym przypadku nie ma przesłanek do odrzucenia hipotezy zerowej (Domański, 1990; Gajda, 2001).

Parametr $t_{\alpha, Q}$ jest wartością krytyczną rozkładu t -Studenta dla poziomu istotności α oraz dla $Q = n - k^{-1}$ stopni swobody. Dla celów tego opracowania przyjęto $\alpha = 0,05$.

Proponowany zbiór zmiennych objaśniających dla opracowanego modelu zawiera następujące atrybuty:

x_1 – średnia (dobowa) temperatura otoczenia [K];

x_2 – temperatura odczuwalna [K];

x_3 – temperatura (dobowa) minimalna [K];

x_4 – temperatura (dobowa) maksymalna [K];

x_5 – gradient średniej temperatury otoczenia [K];

x_6 – gradient temperatury odczuwalnej [K];

x_7 – gradient temperatury minimalnej [K];

x_8 – gradient temperatury maksymalnej [K];

x_9 – ciśnienie atmosferyczne [hPa];

x_{10} – wilgotność względna powietrza [%];

x_{11} – prędkość wiatru (uśredniona dobowo) [m/s];

x_{12} – kierunek wiatru [stopnie azymutu $N = 0^\circ$].

Gradient temperatury zdefiniowany jest w modelu jako różnica temperatury w dniu pomiaru i temperatury w dniu poprzednim.

Wybór modelu ekonometrycznego generuje pewne ograniczenia związane ze zjawiskiem sezonowości, które jest charakterystyczne zarówno dla zużycia gazu jak i zmian pogodowych. Zjawisko sezonowości występuje zarówno po stronie danych historycznych, jak i wyników prognozy. W przypadku prognoz krótkoterminowych (7 dni) zmienność warunków pogodowych nie jest jednak związana z sezonowością, którą (po stronie wyników) można zaniedbać. Jeśli chodzi o wyeliminowanie sezonowości po stronie danych historycznych, to opracowano model wyboru danych do kalibracji, który pozwala zminimalizować to zjawisko. Jest rzeczą oczywistą, że czynniki pogodowe są jednymi z wielu wpływających na zużycie gazu. Z tego względu zadaniem modelu jest, w zasadniczej mierze, wychwycenie jedynie zmienności zużycia gazu w krótkim okresie wynikającej ze zmiany warunków pogodowych, a nie „wyjaśnienie” dobowego zużycia gazu czynnikami pogodowymi.

Najbardziej istotną innowacją, odróżniającą opracowany model od innych rozwiązań prezentowanych w literaturze, jest algorytm wyboru danych historycznych wykorzystywanych do kalibracji parametrów. Przyjęto założenie, że warunki pogodowe

w krótkim okresie (tygodniowym) w odniesieniu do wartości średnich dobowych temperatur nie ulegają znaczącym zmianom. Parametry pogodowe w prognozowanym okresie powinny, z dużym prawdopodobieństwem, mieścić się w założonych przedziałach w odniesieniu do wartości zarejestrowanych w dniu wykonywania prognozy (tzw. dniu bazowym). Ten dzień w dalszej części opracowania jest nazywany dniem bazowym (oznaczonym przez T_0). Zakresy zmienności wszystkich parametrów pogodowych zostały empirycznie wyznaczone poprzez współczynniki zmienności wyliczone oddzielnie dla okresów jesienno-zimowych i wiosenno-letnich. Danymi wejściowymi do prognozy są:

- wartość odbioru gazu q_0 zarejestrowana w dniu wykonywania prognozy;
- wartości parametrów pogodowych x_{0j} ($j = 1 \dots 12$) zarejestrowanych w dniu wykonywania prognozy;
- wartości prognozowanych (w okresie siedmiodniowym) parametrów pogodowych x_{kj} ($k = 1 \dots 7, j = 1 \dots 12$);
- oznaczenie okresu (sezonu), w jakim wykonywana jest prognoza (1 – sezon jesienno-zimowy [15 października do 30 kwietnia], 2 – sezon wiosenno-letni [1 maja do 14 października]).

Algorytm wyboru danych historycznych, parametrów modelu i ich kalibracji odpowiada następującemu schematowi:

- 1) Wybierany jest podzbiór parametrów modelu (zmiennych objaśniających), dla których dokonana zostanie próba dopasowania do danych empirycznych. Zakłada się, że liczba wybranych parametrów będzie większa od dwóch oraz że przynajmniej jedna ze zmiennych z zakresu do x_1 do x_4 będzie reprezentowana. Na tej podstawie tworzony jest wektor wartości przypisanych wybranym atrybutom z dnia T_0 .
- 2) Dla tak wybranego wektora parametrów pogodowych ustalany jest górny i dolny zakres ich zmienności z uwzględnieniem sezonu. Następnie z całego zbioru danych historycznych wybierany jest podzbiór rekordów, dla których wartości wyznaczonych parametrów (w pkt 1) mieszczą się w wyznaczonych zakresach.
- 3) Dla tak wybranego podzbioru testowego wyliczane są nieznane parametry modelu a_j oraz wartości zmiennej objaśniającej. Następnie wyliczane są parametry dopasowania modelu, które muszą spełniać założone kryteria:
 - współczynnik integralnej pojemności nośników informacyjnych $H_k > 0,4$;
 - współczynnik determinacji $R^2 > 0,6$;
 wszystkie oszacowane parametry strukturalne modelu dla danej kombinacji zmiennych muszą spełniać kryterium istotności ($|t(a_j)| > t_{\alpha, Q}$).
- 4) W przypadku niespełnienia założonych kryteriów eliminowane są kolejne rekordy ze zbioru testowego z zastosowaniem zasady, że kolejno usuwane są rekordy, dla których różnice między wartościami empirycznymi i wyliczonymi (dla

zmiennej objaśniającej) są największe. W przypadku poprawnego dopasowania rejestrowany jest komplet wyników. Proces eliminacji rekordów trwa do uzyskania poprawnego dopasowania rozwiązania lub wyczerpania liczby rekordów możliwych do usunięcia. Graniczna liczba rekordów, poniżej której nie można już kontynuować procesu odrzucania, została ustalona jako trzykrotna liczba parametrów modelu.

- 5) W przypadku braku uzyskania rozwiązania w punkcie (4) następuje powrót do punktu (1), w którym losowany jest kolejny podzbiór atrybutów i procedura jest powtarzana.

Po zakończeniu opisanej wyżej procedury wyliczana jest teoretyczna wartość zużycia gazu P_0 w dniu wykonywania prognozy (dniu bazowym T_0) i w kolejnych dniach prognozy. Wielkość ta (w dniu T_0) powinna być odtworzona. Różnice pomiędzy wartością wyliczoną w tym dniu a wartością zarejestrowaną wynikają, poza oczywistymi czynnikami związanymi z błędem metody, z faktu, że dane historyczne dotyczące zużycia gazu rejestrowane były w innych warunkach związanych z otoczeniem rynku na tym terenie. Jak wcześniej zaznaczono, model ma opisywać, jak zmienia się, z dnia na dzień, zużycie gazu w reakcji na zmianę pogody.

W związku z tym wyliczone dane prognozy zostaną przeskalowane stałym czynnikiem multiplikatywnym, tak aby wyliczona wartość zużycia gazu odtwarzała wartość rzeczywistą z dnia T_0 . W ramach eksperymentów związanych z testowaniem modelu wymuszono działanie algorytmu dla kolejnych liczebności wybieranych zmiennych objaśniających, rozpoczynając od 3 zmiennych. Warto zauważyć, że w każdej realizacji opisywanego algorytmu budowany jest inny model (co do zakresu atrybutów modelu i wyboru podzbioru danych testowych).

Testowanie modelu

Podstawowym zagadnieniem prezentowanego opracowania była ocena danych historycznych. Rozdzielono je na dwa zbiory, reprezentujące sezon jesienno-zimowy i wiosenno-letni. Podział taki uzasadniony był tym, że w obydwu tych okresach następuje znacząca zmiana zużycia gazu związana z wykorzystaniem gazu do ogrzewania gospodarstw domowych i innych obiektów. Nie bez znaczenia pozostaje fakt, że zmienność parametrów pogodowych w obu tych okresach nieznacznie się różni. Dokonano oszacowania zmienności parametrów pogodowych przy wykorzystaniu współczynnika zmienności (tab. 1).

Ze względu na duże różnice w rzędach wielkości wartości poszczególnych zmiennych, aby uniknąć błędów wynikających

Tabela 1. Współczynniki zmienności zmiennych objaśniających modelu
Table 1. Coefficients of variation of the model's explanatory variables

Zmienna objaśniająca modelu	Współczynnik zmienności V [%]	
	sezon jesienno-zimowy	sezon wiosenno-letni
Średnia temperatura otoczenia	2,13	1,56
Temperatura odczuwalna	2,52	1,99
Temperatura minimalna	2,17	1,55
Temperatura maksymalna	2,11	1,60
Gradient średniej temperatury otoczenia	0,85	0,80
Gradient temperatury odczuwalnej	0,99	0,99
Gradient temperatury minimalnej	0,87	0,78
Gradient temperatury maksymalnej	0,87	0,84
Ciśnienie atmosferyczne	0,96	0,65
Wilgotność względna powietrza	15,88	15,84
Prędkość wiatru	46,23	37,07
Kierunek wiatru	39,01	39,54

z obliczeń numerycznych, wartości wszystkich zmiennych unormowano, przyjmując wartość 1 dla maksymalnej wartości z serii danych danej zmiennej. W celu korzystania z opracowanych modeli w wartościach realnych zachowano informacje o wartościach referencyjnych x_{ref} (odpowiadających unormowanej wartości 1) dla każdej zmiennej.

Uzyskane wyniki

Opracowany model przetestowano przy wykorzystaniu zbioru danych historycznych pochodzących z trzyletniego okresu. Dane dotyczące zużycia gazu pochodzą z publikowanego – w systemie wymiany informacji firmy przesyłowej – zbioru.

Zaobserwowano występowanie danych anomalnych, odbiegających od średnich z danego okresu, wynikających prawdopodobnie z przerw technicznych, awarii i innych tego typu zdarzeń (głównie po stronie odbiorców). Nie dokonano odrzucenia tych wartości na wstępnym etapie analizy danych, gdyż zastosowany algorytm sam je odrzuca w trakcie dopasowywania modelu.

Historyczne dane pogodowe dla badanego obszaru kraju zostały pobrane z usługi sieciowej (*web service*) *OpenWeatherMap History Bulk*.

Wybrano sześć zestawów danych wejściowych (tab. 2) reprezentowanych dla trzech sezonów jesienno-zimowych i trzech okresów wiosenno-letnich. Wybrane zestawy zostały zróżnicowane ze względu na warunki pogodowe charakterystyczne dla właściwych sezonów:

- zestaw 1 – reprezentuje temperatury dla mroźnej zimy z łagodnym wiatrem;

Tabela 2. Dane pogodowe ustalone w dniu wykonania prognozy (T_0)**Table 2.** Weather data determined on the day of the forecast (T_0)

	Zestaw 1	Zestaw 2	Zestaw 3	Zestaw 4	Zestaw 5	Zestaw 6
x_1	270,05	275,78	281,04	270,19	293,02	297,56
x_2	265,99	270,97	275,90	287,50	290,80	297,20
x_3	267,85	275,01	278,95	287,30	291,82	296,70
x_4	272,73	276,49	283,92	292,33	293,81	298,34
x_9	995,32	984,63	1014,38	1011,79	1016,83	1008,67
x_{10}	93,48	89,42	96,21	68,88	54,00	67,25
x_{11}	2,23	4,26	6,46	4,32	2,89	4,24
x_{12}	118,42	206,25	100,42	74,58	154,75	145,42

- zestaw 2 – reprezentuje warunki zimowe charakterystyczne dla odwilży (temperatury bliskie 0°C);
 - zestaw 3 – jest charakterystyczny dla chłodnej wiosny lub jesieni (średnia temperatura około 8°C);
 - zestaw 4 – cechują temperatury charakterystyczne dla późnej wiosny lub wczesnej jesieni;
 - zestaw 5 – jest charakterystyczny dla umiarkowanych temperatur letnich;
 - zestaw 6 – cechuje temperatury dla ciepłej letniej pogody.
- Uzyskane wyniki zaprezentowano w tabelach 3–14. Numerem „0” oznaczono parametr wolny modelu (model niejednorodny).

Tabela 3. Wyniki oszacowania parametrów modelu dla zestawu 1 z 4 zmiennymi objaśniającymi**Table 3.** The results of the model parameters estimation for set 1 with 4 explanatory variables

H	R^2	S_e	V_e
0,749	0,973	0,022	0,026
x_i	a_i	$t(a_i)$	$t_{a,Q}$
0	12,677	10,598	2,030
1	-27,562	-3,207	2,030
2	17,641	2,337	2,030
9	-2,250	-3,252	2,030
11	0,419	1,832	2,030

Tabela 4. Wyniki oszacowania parametrów modelu dla zestawu 1 z 3 zmiennymi objaśniającymi**Table 4.** The results of the model parameters estimation for set 1 with 3 explanatory variables

H	R^2	S_e	V_e
0,765	0,993	0,009	0,010
x_i	a_i	$t(a_i)$	$t_{a,Q}$
0	8,456	33,282	2,179
4	-4,084	-16,278	2,179
8	0,756	3,372	2,179
9	-4,401	-24,216	2,179

Tabela 5. Wyniki oszacowania parametrów modelu dla zestawu 2 z 4 zmiennymi objaśniającymi**Table 5.** The results of the model parameters estimation for set 2 with 4 explanatory variables

H	R^2	S_e	V_e
0,802	0,978	0,020	0,022
x_i	a_i	$t(a_i)$	$t_{a,Q}$
0	13,722	15,010	2,028
1	-12,887	-9,495	2,028
2	2,642	2,626	2,028
9	-2,961	-10,013	2,028
10	0,247	2,681	2,028

Tabela 6. Wyniki oszacowania parametrów modelu dla zestawu 2 z 3 zmiennymi objaśniającymi**Table 6.** The results of the model parameters estimation for set 2 with 3 explanatory variables

H	R^2	S_e	V_e
0,798	0,930	0,049	0,062
x_i	a_i	$t(a_i)$	$t_{a,Q}$
0	13,826	12,148	1,996
4	-16,801	-6,302	1,996
8	7,004	3,074	1,996
9	-3,394	-4,833	1,996

Tabela 7. Wyniki oszacowania parametrów modelu dla zestawu 3 z 4 zmiennymi objaśniającymi**Table 7.** The results of the model parameters estimation for Set 3 with 4 explanatory variables

H	R^2	S_e	V_e
0,722	0,988	0,021	0,025
x_i	a_i	$t(a_i)$	$t_{a,Q}$
0	14,427	12,593	2,026
2	-10,485	-24,696	2,026
9	-3,243	-3,481	2,026
11	-0,114	-3,931	2,026
12	-0,037	-2,056	2,026

Tabela 8. Wyniki oszacowania parametrów modelu dla zestawu 3 z 3 zmiennymi objaśniającymi

Table 8. The results of the model parameters estimation for set 3 with 3 explanatory variables

H	R^2	S_e	V_e
0,794	0,962	0,040	0,049
x_i	a_i	$t(a_i)$	$t_{a,Q}$
0	15,553	8,708	2,000
1	-8,125	-4,822	2,000
2	-3,590	-2,481	2,000
9	-3,235	-2,295	2,000

Tabela 9. Wyniki oszacowania parametrów modelu dla zestawu 4 z 4 zmiennymi objaśniającymi

Table 9. The results of the model parameters estimation for Set 4 with 4 explanatory variables

H	R^2	S_e	V_e
0,505	0,972	0,021	0,025
x_i	a_i	$t(a_i)$	$t_{a,Q}$
0	-4,861	-2,243	2,064
4	-14,783	-12,134	2,064
7	9,989	5,187	2,064
10	10,335	5,869	2,064
12	0,232	13,750	2,064

Tabela 10. Wyniki oszacowania parametrów modelu dla zestawu 4 z 3 zmiennymi objaśniającymi

Table 10. The results of the model parameters estimation for set 4 with 3 explanatory variables

H	R^2	S_e	V_e
0,513	0,892	0,057	0,071
x_i	a_i	$t(a_i)$	$t_{a,Q}$
0	-2,330	-2,062	2,018
7	3,437	3,025	2,018
10	-0,457	-5,516	2,018
12	0,178	5,108	2,018

Tabela 11. Wyniki oszacowania parametrów modelu dla zestawu 5 z 4 zmiennymi objaśniającymi

Table 11. The results of the model parameters estimation for set 5 with 4 explanatory variables

H	R^2	S_e	V_e
0,456	0,975	0,027	0,033
x_i	a_i	$t(a_i)$	$t_{a,Q}$
0	17,240	13,450	2,101
2	7,545	2,437	2,101
3	-14,417	-3,513	2,101
9	-9,926	-9,223	2,101
11	0,320	5,543	2,101

Tabela 12. Wyniki oszacowania parametrów modelu dla zestawu 5 z 3 zmiennymi objaśniającymi

Table 12. The results of the model parameters estimation for set 5 with 3 explanatory variables

H	R^2	S_e	V_e
0,491	0,950	0,034	0,040
x_i	a_i	$t(a_i)$	$t_{a,Q}$
0	2,752	4,130	2,012
2	-1,548	-2,280	2,012
10	-0,556	-11,100	2,012
12	0,131	5,448	2,012

Tabela 13. Wyniki oszacowania parametrów modelu dla zestawu 6 z 4 zmiennymi objaśniającymi

Table 13. The results of the model parameters estimation for set 6 with 4 explanatory variables

H	R^2	S_e	V_e
0,582	0,961	0,032	0,037
x_i	a_i	$t(a_i)$	$t_{a,Q}$
0	25,757	12,426	2,052
1	-40,281	-8,691	2,052
3	26,861	5,913	2,052
9	-11,744	-7,211	2,052
12	0,106	3,722	2,052

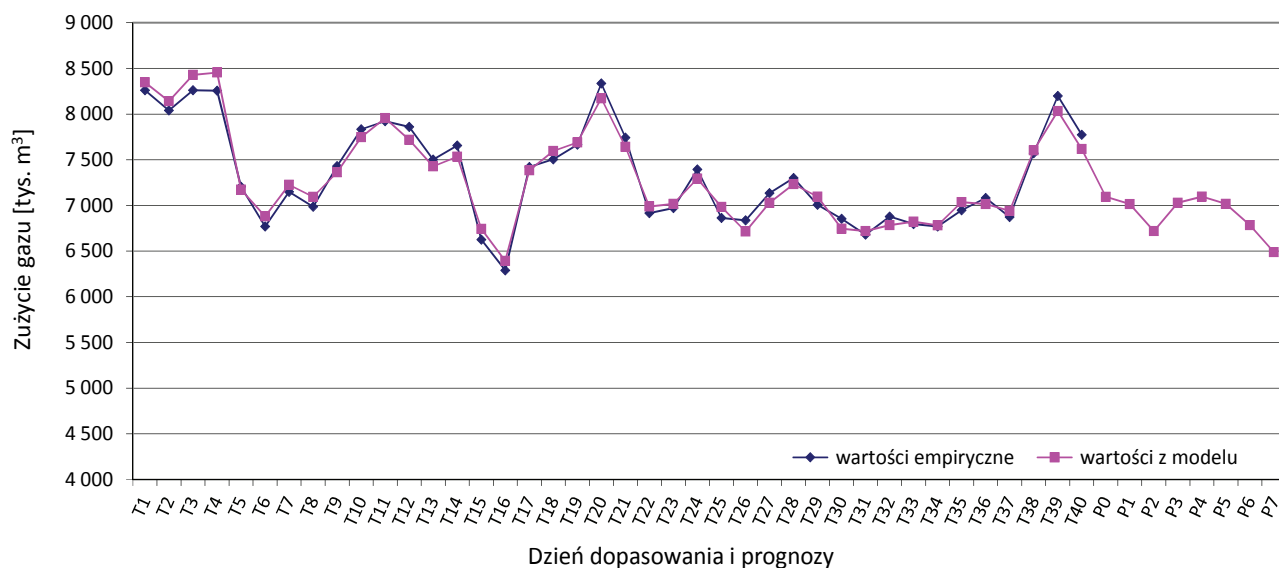
Tabela 14. Wyniki oszacowania parametrów modelu dla zestawu 6 z 3 zmiennymi objaśniającymi

Table 14. The results of the model parameters estimation for set 6 with 3 explanatory variables

H	R^2	S_e	V_e
0,550	0,977	0,014	0,015
x_i	a_i	$t(a_i)$	$t_{a,Q}$
0	16,663	19,019	2,160
2	-17,795	-6,248	2,160
10	5,586	2,230	2,160
12	-3,748	-5,364	2,160

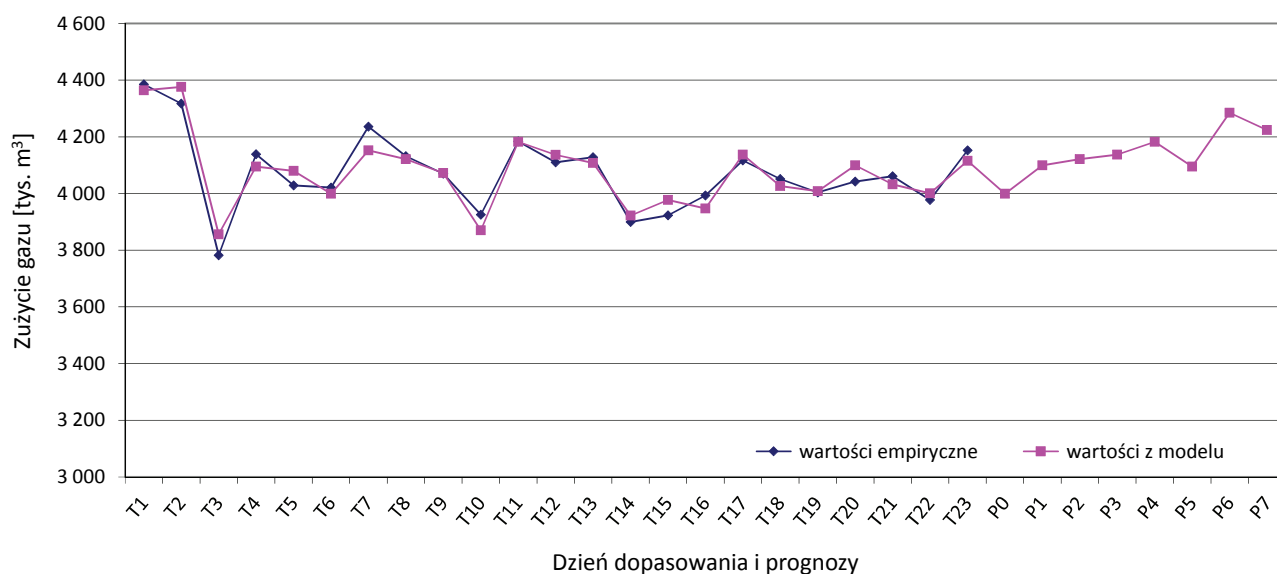
Na rysunkach 1 i 2 przedstawiono przykładowe wyniki prognoz (punkty oznaczone jako P1, P2, ..., P7 na wykresie) oraz wyniki dopasowania wartości empirycznych (punkty oznaczone jako T1, T2, ..., Tn na wykresie) do wartości wyliczonych z modelu w wariantach dla 4 zmiennych objaśniających. Analizując otrzymane wyniki, można zauważyć, że:

- uzyskanie akceptowalnego dopasowania dla więcej niż 4 zmiennych objaśniających napotyka na trudności i jest możliwe w nielicznych przypadkach;
- temperatury powietrza mają zasadniczy udział w wyliczonych wartościach zmiennej objaśnianej, co widać w wartościach współczynników modelu. Należy pamiętać, że obliczenia wykonywane są na wartościach unormowanych;



Rys. 1. Krótkoterminowa prognoza zużycia gazu – zestaw 1. Sezon jesienno-zimowy

Fig. 1. Short-term forecast of gas consumption – set 1. Fall-winter season



Rys. 2. Krótkoterminowa prognoza zużycia gazu – zestaw 5. Sezon wiosenno-letni

Fig. 2. Short-term gas consumption forecast – set 5. Summer-spring season

- dla sezonów jesienno-zimowych uzyskano lepsze wartości oszacowania niepewności wyznaczanych parametrów (zwłaszcza w przypadku współczynnika H) niż dla sezonów wiosenno-letnich.
2. Zbiór zaproponowanych zmiennych objaśnianych daje możliwości wyboru wielu kombinacji zmiennych, dla których spełnione jest kryterium dostatecznej integralnej pojemności informacyjnej i jednocześnie spełnione są kryteria istotności.
 3. Prognozowanie zużycia gazu w sezonie jesienno-zimowym daje lepsze rezultaty w porównaniu z prognozowaniem w sezonie wiosenno-letnim.
 4. Zastosowanie liniowych modeli ekonometrycznych w prognozowaniu krótkoterminowym zużycia gazu jest konkurencyjnym rozwiązaniem w porównaniu z zaawansowanymi metodami AI i big data w przypadku, gdy dysponujemy stosunkowo małym zbiorem danych empirycznych.

Wnioski

1. Zbudowany i zaprezentowany model ekonometryczny może być stosowany do krótkoterminowego prognozowania zapotrzebowania na gaz pod warunkiem rozszerzenia zbioru danych zużycia gazu o większą liczbę sezonów w porównaniu z danymi wykorzystanymi w opracowaniu.

Artykuł powstał na podstawie pracy statutowej pt. *Numeryczny model krótkoterminowego prognozowania zużycia gazu dla wybranych jednostek terytorialnych kraju* – praca INiG – PIB na zlecenie MNiSW; nr zlecenia: 0062/KP/2020, nr archiwalny: DK-4100-0050/2020.

Literatura

- Bianco V., Scarpa F., Tagliafico L.A., 2014a. Analysis and future outlook of natural gas consumption in the Italian residential sector. *Energy Conversion and Management*, 87: 754–764. DOI: 10.1016/j.enconman.2014.07.081.
- Bianco V., Scarpa F., Tagliafico L.A., 2014b. Scenario analysis of nonresidential natural gas consumption in Italy. *Applied Energy*, 113: 392–403. DOI: 10.1016/j.apenergy.2013.07.054.
- Cieślak M., 1996. Prognozowanie gospodarcze. *Wydawnictwo Akademii Ekonomicznej im. Oskara Langego, Wrocław*.
- Cieślak T., Kogut K., 2016. Prognozowanie pracy sieci gazowej za pomocą sztucznych sieci neuronowych. *Nafta-Gaz*, 6: 443–450. DOI: 10.18668/NG.2016.06.08.
- Czapaj R., Benalcazar P., Kamiński J., 2019. Prognozowanie krótkoterminowe zapotrzebowania na energię elektryczną w KSE z wykorzystaniem metody MARSplines. *Przegląd Elektrotechniczny*, 95(7): 133–136. DOI: 10.15199/48.2019.07.27.
- Domański C., 1990. Testy statystyczne. *PWE, Warszawa*.
- Gajda J., 2001. Prognozowanie i symulacje a decyzje gospodarcze. *Wydawnictwo C.H. Beck, Warszawa*.
- Gawlik L., 2008. Budowa i weryfikacja modelu ekonometrycznego dla określenia liniowej zależności pomiędzy kosztami pozyskania węgla a wielkością wydobycia. *Gospodarka Surowcami Mineralnymi*, 24(1/1): 27–44.
- Gładysz B., Mercik J., 2007. Modelowanie ekonometryczne. Studium przypadku. *Oficina Wydawnicza Politechniki Wrocławskiej, Wrocław*.
- Goryl A., Jędrzejczyk Z., Kukuła K., Osiewalski J., Walkosz A., 2009. Wprowadzenie do ekonometrii. *Wydawnictwo Naukowe PWN, Warszawa*.
- Jeżyk T., Tomczewski A., 2014. Krótkoterminowe prognozowanie zużycia energii elektrycznej z wykorzystaniem sztucznej sieci neuronowej. *Electrical Engineering*, 79: 121–130.
- Kelner M., 2003. Prognozowanie krótkoterminowe poborów gazu z sieci przemysłowej metodą sztucznych sieci neuronowych. *Gaz, Woda i Technika Sanitarna*, 6: 196–204.
- Kogut K., 2007. Analiza możliwości modelowania sieci przesyłowej gazu ziemnego. *Wydawnictwa AGH, Kraków*.
- Łucki Z., 1988. Modele ekonometryczne do zarządzania w górnictwie naftowym. *Zeszyty Naukowe AGH. Zagadnienia Techniczno-Ekonomiczne*, 41.
- Łucki Z. (red.), Byrska-Rapała A., Kozarkiewicz A., 2008: Statystyka i ekonometria. *AGH, Uczelniane Wydawnictwa Naukowo-Dydaktyczne, Kraków*.
- Maciejasz M., 2006. Zastosowanie sieci neuronowych do analizy pracy sieci przesyłowych. *Wydawnictwa AGH, Kraków*.
- Osińska M., 2006. Ekonometria finansowa. *PWE, Warszawa*.
- Paliński A., Łucki Z., 2006. Problemy zarządzania finansowego i zarządzania ryzykiem w sektorze handlu energią. *Polityka Energetyczna*, 9(2): 129–152.
- Rogowska D., 2017. Wykorzystanie OZE w energetyce a zrównoważony rozwój. *Nafta-Gaz*, 8: 616–623. DOI: 10.18668/NG.2017.08.09.
- Rogowska D., 2018. Produkcja biopaliw jako element gospodarki o obiegu zamkniętym. *Nafta-Gaz*, 2: 156–163. DOI: 10.18668/NG.2018.02.10.
- Siewierski J. i zespół, 2017. Raport Polski Rynek Paliw Płynnych 2016. *Polska Izba Paliw Płynnych, Warszawa*.
- Sokołowski A., Pasztyła A., 2004. Data Mining w prognozowaniu zapotrzebowania na nośniki energii. *StatSoft Polska Sp. z o.o.*: 91–102.
- Soldo B., 2012. Forecasting natural gas consumption. *Applied Energy*, 92(C): 26–37. DOI: 10.1016/j.apenergy.2011.11.003.
- Szmigiel C., Mercik J., 2000. Ekonometria. *Wydawnictwo Wyższej Szkoły Zarządzania i Finansów we Wrocławiu, Wrocław*.
- Szoplak J., 2015. Forecasting of natural gas consumption with artificial neural networks. *Energy*, 85: 208–220. DOI: 10.1016/j.energy.2015.03.084.
- Tamba J.G., Essiane S.N., Sapnken E.F., Koffi F.D., Nsouandélé J.L., Soldo B., Njomo D., 2018. Forecasting natural gas: A literature survey. *International Journal of Energy Economics and Policy, Econjournals*, 8(3): 216–249.
- Welfe A., 1988. Ekonometria. Metody i ich zastosowania. *PWE, Warszawa, 1998*.
- Wójcik M., 2005. Model sieci gazowniczej oparty o sztuczne sieci neuronowe. *Wydawnictwa AGH, Kraków*.
- Wydymus S., 1984. Zmienne syntetyczne w modelowaniu ekonometrycznym. *Ruch Prawniczy, Ekonomiczny i Socjologiczny*, 46(1): 187–200.
- Yu F., Xu X., 2014. A short-term load forecasting model of natural gas based on optimized genetic algorithm and improved BP neural network. *Applied Energy*, 134: 102–113. DOI: 10.1016/j.apenergy.2014.07.104.
- Zeliaś A., Pawełek B., Wanat S., 2003. Prognozowanie ekonomiczne. Teoria, przykłady, zadania. *PWN, Warszawa*.
- Zheng H., Yuan J., Chen L., 2017. Short-Term Load Forecasting Using EMD-LSTM Neural Networks with a Xgboost Algorithm for Feature Importance Evaluation. *Energies*, 10(8): 1168. DOI: 10.3390/en10081168.
- Zieliński Z., 1995. Liniowe modele zgodnie opisujące zależności symulacyjnych (zintegrowanych) procesów ekonomicznych. *Wydawnictwo Akademii Ekonomicznej, Kraków*: 77–87.



Dr Tadeusz KWILOSZ
Adiunkt w Zakładzie Podziemnego
Magazynowania Gazu
Instytut Nafty i Gazu – Państwowy Instytut Badawczy
ul. Lubicz 25 A
31-503 Kraków
E-mail: tadeusz.kwilosz@inig.pl



Mgr inż. Bogdan FILAR
Kierownik Zakładu Podziemnego
Magazynowania Gazu
Instytut Nafty i Gazu – Państwowy Instytut Badawczy
ul. Lubicz 25 A
31-503 Kraków
E-mail: bogdan.filar@inig.pl