

## APPLICATION OF PCA IN THE ANALYSIS OF PARAMETERS RELATED TO AGRICULTURAL GREENHOUSE GASES EMISSIONS IN EUROPE

### Summary

In the study, dimensionality reduction of data was carried out using the method of Principal Components Analysis (PCA). Two groups of variables were analyzed - size of population of livestock and crop areas in Europe. Two principal components explaining in 95% variability of the original data were obtained from dimension of the 13 variables.

**Key words:** farming; pollutants; carbon dioxide; methane; nitrous oxide; environmental pollution; main components; field experimentation; method; Principal Component Analysis

## WYKORZYSTANIE METODY PCA W ANALIZIE PARAMETRÓW POWIĄZANYCH Z ROLNICZYMI EMISJAMI GAZÓW CIEPLARNIANYCH W EUROPIE

### Streszczenie

W badaniach przeprowadzono redukcję wymiarowości danych z wykorzystaniem metody analizy składowych głównych. Analizie poddano dwie grupy zmiennych: wielkość pogłowia zwierząt gospodarskich oraz areal upraw w Europie. Z wymiaru 13 zmiennych otrzymano dwie główne składowe, które w 95% wyjaśniają zmienności pierwotnych danych.

**Słowa kluczowe:** rolnictwo; gazy cieplarniane; dwutlenek węgla; metan; podtlenek azotu; zanieczyszczenie środowiska; składowe główne; badania polowe; metoda; PCA

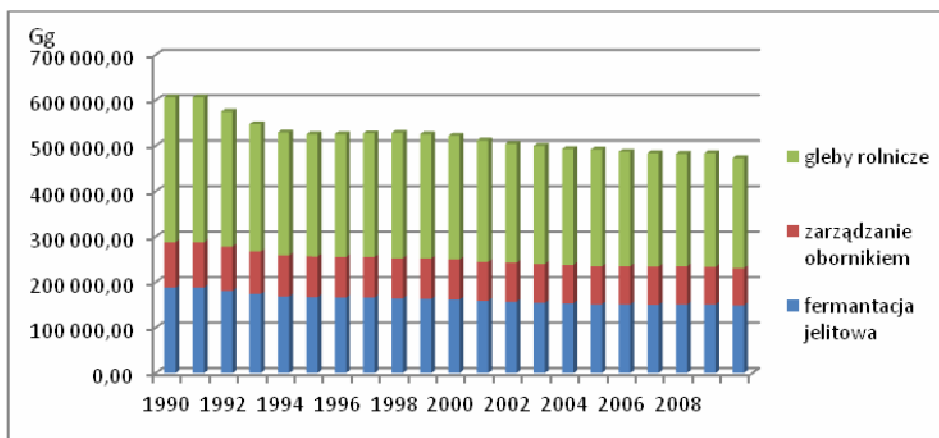
### 1. Wprowadzenie

Rolnictwo, po sektorze energetycznym, stanowi drugie źródło pod względem wielkości emisji głównych gazów cieplarnianych (GGC), a jego udział w ogólnej emisji stanowił 9,2% w 2009 roku [6]. Obecnie Europa uchodzi za lidera w zakresie skutecznego obniżania emisji powodowanych przez rolnictwo. W okresie minionych dekad w Europie rozpowszechniono szereg działań, m.in. redukcję pogłowia bydła, zmniejszenie wykorzystania nawozów, nowoczesne rozwiązania w zakresie uprawy gruntów oraz inne, które pozwoliły osiągnąć zamierzony cel (rys. 1). Łącznie z emisjami pochodzącymi ze zmian użytkowania gruntów sektor rolny jest odpowiedzialny za ok. 1/3 globalnych emisji GGC i stanowi najpoważniejsze źródło emisji gazów cieplarnianych innych niż dwutlenek węgla (CO<sub>2</sub>) emitując głównie metan (CH<sub>4</sub>) i podtlenek azotu (N<sub>2</sub>O). Emisje rol-

nicze mają swoje źródło głównie w intensywnej uprawie ziemi oraz hodowli zwierząt, przy czym struktura upraw oraz ich nawożenie, a także rodzaj hodowli mają wpływ na wielkość emisji gazów [3].

### 2. Metodyka badań

Celem badań była redukcja wymiaru danych powiązanych z emisjami z rolnictwa z wykorzystaniem analizy składowych głównych. Analiza składowych głównych (PCA – *Principal Component Analysis*) to technika pozwalająca na redukcję wymiarowości danych. Jest to metoda nieparametryczna, a więc nie wymagająca założeń o rozkładzie analizowanych zmiennych. Jej celem jest zastąpienie wejściowego zbioru skorelowanych cech przez niewielką ilość nieskorelowanych tzw. składowych głównych, które razem mogą wyjaśnić niemal całą zmienność danych.



Rys. 1. Emisja rolnicza GGC w krajach UE w ekwiwalencie CO<sub>2</sub>

Fig. 1. GHG agricultural emission in the EU countries in CO<sub>2</sub> equivalent

Znajduje zastosowanie w szerokim spektrum nauk [1, 4, 5]. Pierwsza składowa wyjaśnia najwięcej zmienności (składowe są kombinacjami liniowymi wejściowych zmiennych). Metoda dzięki redukcji pozwala na wskazanie niewidocznych przy dużej liczbie atrybutów związków między nimi. Składowymi głównymi są wektory własne macierzy korelacji.

Ze względu na dużą liczbę rozpatrywanych zmiennych (13 zmiennych), trudno dostrzec ewentualne korelacje. Dane do analiz pozyskano z baz danych Food and Agriculture Organization [2] oraz United Nations Framework Convention on Climate Change. Do badań przyjęto następujące dane: wielkość hodowli podstawowych grup zwierząt (bydło, drób, trzoda chlewna, konie, owce) oraz wiodące w Europie uprawy, głównie zboża (pszenica, jęczmień, pszenżyto, kukurydza, żyto, rzepak, owies, a także burak cukrowy). Badania przeprowadzono z wykorzystaniem pakietu Statistica v. 9.0.

### 3. Wyniki badań i interpretacja

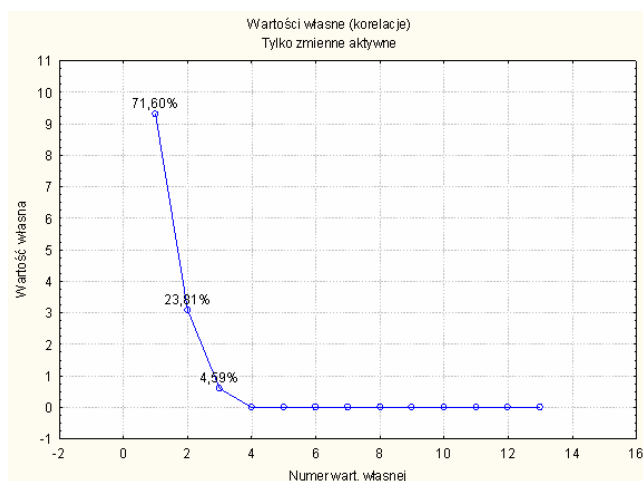
Odczytanie stopnia wzajemnych zależności pomiędzy poszczególnymi atrybutami możliwe jest dzięki macierzy korelacji (tab. 1). Wartość bezwzględna świadczy o sile korelacji między zmiennymi – im jest wyższa, tym związek jest silniejszy. Otrzymano silne korelacje zarówno dodatnie jak i ujemne. Daje się zauważyć wysoce dodatnio skorelowane zmienne, m.in. dla zmiennych B (drób) z D, F, G, H, I (odpowiednio: konie, pszenica, jęczmień, pszenżyto, kukurydza) lub zmienne skorelowane ujemnie, np. A (bydło) i E (owce). Dostrzec też można przypadki świadczące o braku korelacji pomiędzy zmiennymi – E (owce) z I (kukurydza).

Miarą zmienności pierwotnych danych przedstawionych

jako współrzędne składowe główne są wartości własne macierzy korelacji (tab. 2). Kryterium średniej wartości własnej wskazuje, że dwie pierwsze składowe odpowiadają za ponad 95% zmienności pierwotnych danych, co daje możliwość dalszej analizy w dwóch wymiarach.

Wykres osypiska przedstawiony na rys. 2 jest potwierdzeniem istotności dwóch pierwszych składowych, co tym samym pozwala na redukcję 13 wymiarowej przestrzeni do dwóch składowych.

W tab. 3 zestawiono wektory własne macierzy korelacji. Wartości i znaki elementów wektorów własnych mówią o sile i kierunku wpływu poszczególnych zmiennych na główne składowe.



Rys. 2. Wykres osypiska dla wartości własnych macierzy korelacji

Fig. 2. Scree plot for individual values of the matrix of correlation

Tab. 1. Macierz korelacji między zmiennymi

Table 1. Matrix of correlation between parameters

Zm.	Korelacje												
	A	B	C	D	E	F	G	H	I	J	K	L	M
A	1,000	0,687	0,738	0,709	-0,809	0,638	0,578	0,778	0,515	0,913	0,663	0,959	0,453
B	0,687	1,000	0,586	0,953	-0,184	0,963	0,946	0,948	0,973	0,919	0,956	0,776	-0,210
C	0,738	0,586	1,000	0,398	-0,733	0,373	0,304	0,460	0,533	0,680	0,371	0,572	0,606
D	0,709	0,953	0,398	1,000	-0,161	0,993	0,984	0,994	0,885	0,922	0,997	0,851	-0,293
E	-0,809	-0,184	-0,733	-0,161	1,000	-0,070	0,007	-0,262	-0,006	-0,516	-0,100	-0,627	-0,886
F	0,638	0,963	0,373	0,993	-0,070	1,000	0,996	0,978	0,920	0,889	0,998	0,790	-0,370
G	0,578	0,946	0,304	0,984	0,007	0,996	1,000	0,961	0,914	0,850	0,994	0,747	-0,442
H	0,778	0,948	0,460	0,994	-0,262	0,978	0,961	1,000	0,862	0,955	0,985	0,899	-0,195
I	0,515	0,973	0,533	0,885	-0,006	0,920	0,914	0,862	1,000	0,808	0,903	0,611	-0,332
J	0,913	0,919	0,680	0,922	-0,516	0,889	0,850	0,955	0,808	1,000	0,900	0,954	0,094
K	0,663	0,956	0,371	0,997	-0,100	0,998	0,994	0,985	0,903	0,900	1,000	0,814	-0,348
L	0,959	0,776	0,572	0,851	-0,627	0,790	0,747	0,899	0,611	0,954	0,814	1,000	0,197
M	0,453	-0,210	0,606	-0,293	-0,886	-0,370	-0,442	-0,195	-0,332	0,094	-0,348	0,197	1,000

A – bydło, B – drób, C – trzoda chlewna, D – konie, E – owce, F – pszenica, G – jęczmień, H – pszenżyto, I – kukurydza, J – burak cukrowy, K – żyto, L – rzepak, M – owies

Tab. 2. Wartości własne macierzy korelacji

Table 2. Individual values of correlation matrix

Nr wartości	Wartości własne (korelacje), pokrewne statystyki			
	Wartości własne	% ogółu wariacji	Skumulowane wartości własne	Skumulowane %
1	<b>9,308228</b>	71,60175	9,30823	71,6018
2	<b>3,094769</b>	23,80591	12,40300	<b>95,4077</b>
3	0,597003	4,59233	13,00000	100,0000

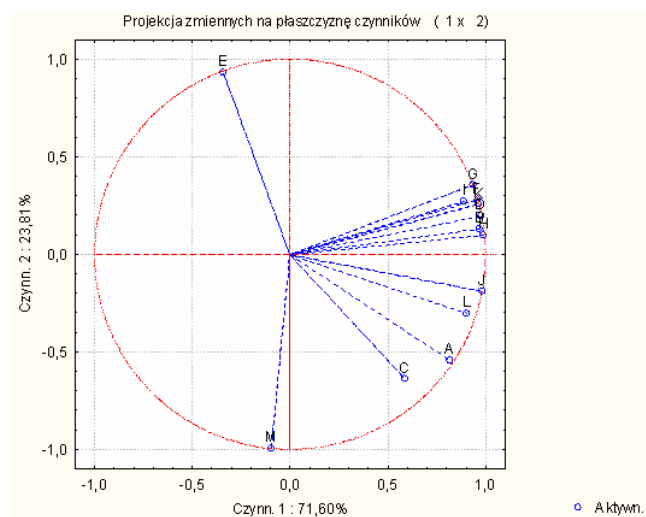
Tab. 3. Wektory własne macierzy korelacji  
 Table 3. Individual vectors of the correlation matrix

Zmienna	Wektory własne macierzy korelacji		
	Czynn. 1	Czynn. 2	Czynn. 3
A	0,268686	-0,309238	0,231771
B	0,317961	0,074880	-0,263958
C	0,192297	-0,362695	-0,645416
D	0,319770	0,112561	0,122751
E	-0,112176	0,529446	-0,160441
F	0,314550	0,159294	0,028898
G	0,306180	0,202036	0,042229
H	0,324046	0,055008	0,148788
I	0,291412	0,153758	-0,477940
J	0,321613	-0,107957	0,043663
K	0,316232	0,145210	0,080767
L	0,296464	-0,171348	0,390485
M	-0,030836	-0,565043	-0,071744

A – bydło, C – trzoda chlewna, E – owce, M – owies.

Pierwsza składowa główna składa się z wielu zmiennych, odnoszących się do zmiennych związanych z hodowlą drobiu, koni oraz rodzajem upraw tj. pszenicy, jęczmienia, pszenżyta, żyta i buraków cukrowych. Zmienne wnoszą podobny wkład w składową. Literatura przedmiotu podaje, iż dla zmiennych o korelacji < 0,3 nie powinno prowadzić się dalszej analizy [7]. Pierwsza składowa w ok. 72% wyjaśnia zmienności w danych. W skład drugiej składowej wchodzi przede wszystkim dwie zmienne o przeciwstawnych znakach, co sugeruje, iż wysokie wartości zmiennej pogłowie owiec odpowiadają niskim wartościom upraw owsa i odwrotnie. Druga składowa odpowiada za niemal 24% zmienności.

Wpływ poszczególnych zmiennych wejściowych na użyte składowe główne widoczny jest na wykresie na rys. 3 wyznaczonym przez dwie pierwsze składowe.



Rys. 3. Wykres konfiguracji punktów reprezentujących zmienne w układzie dwóch pierwszych osi czynnikowych (głównych składowych)

*Fig. 3. Chart of configuration of points representing the variables in the system of two first factorial axes (principal components)*

Większość z rozważanych atrybutów ma znamienne znaczenie, o czym świadczy długość ich wektorów. Im są one dłuższe, a zmienne położone są bliżej okręgu, tym większa część informacji zawartych w danej zmiennej wejściowej jest przenoszona przez główne składowe. Położenie wektorów w bliskim sąsiedztwie oznacza, że zmienne są wobec siebie silnie dodatnio skorelowane. W rozważanym przypadku możemy mówić o takich korelacjach pomiędzy grupą zmiennych np. jęczmień, kukurydza, pszenica, żyto, drób, konie, pszenżyto. Wektory zmiennych ujemnie skorelowanych położone są wobec siebie po przeciwnej stronie, np. trzoda chlewna – owce, natomiast wektory skierowane wobec siebie prostopadle, oznaczają brak korelacji między zmiennymi, np. owce – jęczmień.

#### 4. Podsumowanie

Obliczenie macierzy korelacji między zmiennymi wskazało niewidoczne wcześniej zależności. Zastosowanie metody analizy składowych głównych pozwoliło na redukcję pierwotnego zbioru 13 zmiennych do wymiaru dwóch składowych głównych. Składowa pierwsza, składa się z wielu zmiennych, których wielkość rzutuje na emisje GGC. Odnosi się przede wszystkim do struktury upraw, ale również wielkości hodowli drobiu oraz koni. W przypadku drugiej składowej, zmienne bydło, trzoda chlewna oraz owies są ujemnie skorelowane ze zmienną owce, co oznacza, iż wzrost tych pierwszych powoduje spadek wartości drugiego i odwrotnie.

#### 5. Bibliografia

- [1] Czernyszewicz E.: Zastosowanie analizy głównych składowych do opisu konsumenckiej struktury jakości jabłek, *Żywność. Nauka. Technologia. Jakość*, 2008, nr 2 (57): 119–127.
- [2] Food and Agriculture Organization of the United Nations, [www.faostat.fao.org](http://www.faostat.fao.org) [dostęp 4.12.2011].
- [3] Kaiser E. A., Ruser R.: Nitrous oxide emissions from arable soils in Germany – An evaluation of six long-term field experiments. *J. Plant Nutr. Soil Sci.*, 2000: 163, 249–260.
- [4] Kamińska A., Andrejko D.: Zastosowanie analizy składowych głównych w badaniu wpływu nawilżania nasion łubinu żółtego odmiany radames na ich wielkość, *Inżynieria Rolnicza*, 2006, nr 7 (82): 241–246.
- [5] Kuna-Broniowska I., Talik Z.: Badanie rozkładu temperatur w silosie betonowym podczas schładzania ziarna pszenicy urządzeniami GK-160 i KK-140, *Technica Agraria*, 2002, 1(2): 61–72.
- [6] Miatkowski Z., Turbiak J., Burczyk P., Myczko A., Karłowski J.: Prognozy zmian aktywności w sektorze rolnictwa, zawierające informacje niezbędne do wyliczenia szacunkowej wielkości emisji gazów cieplarnianych. Raport z realizacji umowy pomiędzy Ministerstwem Rolnictwa i Rozwoju Wsi a Instytutem Technologiczno-Przyrodniczym w Falentach, umowa o dzieło nr BDGzp-2120A-31/10 zawarta w dniu 30.06.2010, Bydgoszcz, Poznań, Instytut Technologiczno-Przyrodniczy w Falentach, 2010.
- [7] Sokołowski A., Sagan A.: Przykłady stosowania analizy danych w marketingu i badaniu opinii publicznej, [www.statsoft.pl/ czytelnia/marketing/adwmarketingu.html](http://www.statsoft.pl/czytelnia/marketing/adwmarketingu.html), [dostęp 6.12.2011].