

## ANALIZA SKŁADOWYCH GŁÓWNYCH W OCENIE ZALEŻNOŚCI SPOSOBU UŻYTKOWANIA GRUNTÓW Z EMISJĄ GAZÓW CIEPLARNIANYCH Z ROLNICTWA

*Alicja Kolasa-Więcek*

*Katedra Ekonomii i Badań Regionalnych, Politechnika Opolska*

**Streszczenie.** W badaniach wykorzystano metodę analizy składowych głównych w celu wyodrębnienia istotnie skorelowanych zmiennych powiązanych z emisją głównych gazów cieplarnianych z rolnictwa względem sposobu użytkowania gruntów. Metoda daje sposobność zastąpienia wejściowego zbioru skorelowanych cech poprzez niewielką liczbę nieskorelowanych składowych głównych, które stanowią liniowe kombinacje zmiennych. Minimalizacja liczby zmiennych potrzebnych do wyjaśnienia danej zmiennej upraszcza interpretację wyników. Badania przeprowadzono z wykorzystaniem programu Statistica 9.0. Wyodrębniono dwie składowe główne, które w 96% wyjaśniają zmienności danych wejściowych. Pierwsza główna składowa wyjaśnia 75% zmienności zawartych głównie w zmiennych grunty orne wraz ze stałą powierzchnią upraw i pastwiska oraz ujemnie skorelowanych zmiennych tj. inne grunty rolne, powierzchnia lasów a także ilość maszyn rolniczych. Druga składowa wyjaśnia 21% zmienności danych poprzez zmienną ilość zużytej w gospodarstwach rolnych energii.

**Słowa kluczowe:** analiza składowych głównych, PCA, główne gazy cieplarniane, emisje z rolnictwa

### Wprowadzenie

Wzrost populacji ludzkiej na świecie wymusza rosnącą potrzebę produkcji żywności. Znaczna intensyfikacja rolnictwa powoduje ciągły wzrost emisji zanieczyszczeń do środowiska, w tym gazów cieplarnianych. Dynamicznie zwiększa się ilość użytków rolnych wymagających ciągłego nawożenia, stosowania nowoczesnych maszyn rolniczych oraz powodującego wzrost zużycia energii.

Rolnictwo łącznie z emisjami pochodzącymi ze zmian użytkowania gruntów jest odpowiedzialne za niemal 1/3 globalnych emisji głównych gazów cieplarnianych (GGC) [WRI]. Gazy cieplarniane emitowane z rolnictwa to przede wszystkim  $\text{CH}_4$ ,  $\text{N}_2\text{O}$  oraz  $\text{CO}_2$ . Źródła

tych emisje obejmują: fermentację jelitową zwierząt gospodarskich ( $\text{CH}_4$ ), odchody zwierzęce ( $\text{CH}_4$  i  $\text{N}_2\text{O}$ ), gleby rolne ( $\text{N}_2\text{O}$ ) oraz spalanie resztek roślinnych ( $\text{CH}_4$  i  $\text{N}_2\text{O}$ ) oraz odrębny sektor przypisany użytkowaniu gruntów, zmianom użytkowania oraz leśnictwu [Bebkiewicz i inni, 2011].

Istotnym źródłem emisji  $\text{CO}_2$  jest uprawa roli oraz wylesianie. Ocenia się, iż sektor leśny odpowiada za 17,4% emisji gazów cieplarnianych [Nabuurs i in. 2007] a deforestacja dla potrzeb produkcji zwierzęcej stanowi 89,5% emisji  $\text{CO}_2$  uwalnianych z całej produkcji zwierzęcej [Steinfeld 2006]. Uważa się, że podjęcie działań zmierzających do stosowania w rolnictwie i leśnictwie dobrych praktyk może przyczynić się do zwiększenia zdolności zatrzymywania  $\text{CO}_2$  w glebie i lasach. Badanie wpływu produkcji rolnej na środowisko, a zwłaszcza emisji gazów cieplarnianych, stało się poważną problematyką podejmowaną zarówno przez naukowców jak i decydentów politycznych na całym świecie [np. Myczko i in. 2003; Pietrzak i in. 2002; Winnicki i in. 2010]. Wiele instytucji ukierunkowanych jest na badanie wpływu sposobów użytkowania gleb z emisją gazów cieplarnianych, np. podtlenku azotu, którego rolnictwo jest największym źródłem [Burezyk i in. 2011; Freney 1997; Oenema i in. 1999; Turbiak i in. 2011]. Coraz częściej w tym celu stosowane są nowoczesne narzędzia i techniki obliczeniowe pozwalające na interpretację zjawisk oraz zależności je łączących służące prognozowaniu możliwych w przyszłości wielkości emisji.

Wzrastające możliwości gromadzenia danych potrzebują metod, które pozwalają na odnalezienie cennych dopełniających wiedzę w danej dziedzinie informacji. Służą temu narzędzia algorytmiczne do eksploracji danych [Hand i in. 2005]. Poprzez nagromadzone dane eksperymentalne lub empiryczne uzyskane uprzednio dla potrzeb innych badań, można dotrzeć do zupełnie nowych, wartościowych naukowo informacji, których przesłanki nie zostały jeszcze wykorzystane [Tadeusiewicz 2006].

## Metodyka badań

W podjętych badaniach jako cel założono analizę zależności sposobu użytkowania gruntów i źródeł emisji towarzyszących uprawie, względem emisji gazów cieplarnianych. Do badań wykorzystano dane pochodzące z bazy danych Organization for Economic Cooperation and Development (OECD). Badania prowadzono dla Polski, na zbiorze danych z 15 lat (1990-2004). Dane pochodzą z międzynarodowej bazy danych OECD [OECD]. Do badań zastosowano analizę składowych głównych (Principal Component Analysis, PCA), która przy redukcji liczby zmiennych, pozwala opisać dany proces lub zjawisko z zachowaniem maksymalnego poziomu informacji. Metodę zastosowano celem redukcji wymiarowości danych. Metoda dzięki minimalizacji liczby zmiennych potrzebnych do wyjaśnienia danej zmiennej upraszcza interpretację wyników. W celu nazwania nowych zmiennych (głównych składowych) wyodrębnia się te o najwyższych ładunkach czynnikowych względem danych składowych. Zastosowanie metody jest możliwe jeśli między oryginalnymi zmiennymi zachodzą pewne zależności statystyczne. PCA polega na zastąpieniu wejściowego zbioru skorelowanych cech poprzez niewielką liczbę nieskorelowanych składowych głównych, które stanowią liniowe kombinacje zmiennych obserwowanych. Mogą one razem wyjaśnić prawie całą zmienność danych. Zakłada się, że kilka pierwszych składowych gromadzi w sobie większość zmienności pierwotnego zbioru danych [Morrison 1990].

Analizowane dane to: użytki rolne ogółem tj. grunty orne i stała powierzchnia upraw, pastwiska, inne kategorie gruntów rolnych, powierzchnia lasów, bezpośrednie zużycie energii w gospodarstwach, liczba maszyn rolniczych oraz całkowita emisja głównych gazów cieplarnianych z rolnictwa.

Poprzez metodę PCA wyodrębniono składowe opisujące zależność z emisją GGC z rolnictwa. W poddanych analizie 6 zmiennych trudno jest dostrzec związek pomiędzy nimi. Istotna redukcja zmiennych umożliwi jego wskazanie, co w pierwotnej wersji nie było możliwe. Analiza została przeprowadzona z wykorzystaniem pakietu Statistica 9.0. Wartości zmiennych wstępnie poddano standaryzacji.

## Wyniki badań i ich interpretacja

Wyniki analizy statystycznej analizowanych parametrów przedstawiono w tabeli 1.

Tabela 1. Charakterystyka statystyczna zmiennych  
Table 1. Statistical characteristics of variables

Wyszczególnienie	Jednostka	Średnia	Mediana	Odchylenie standardowe
Użytki rolne ogółem	[ha]	18 132 000,00	18 457 000,00	881513,0661
Grunty orne i stała powierzchnia upraw	[ha]	14 198 386,67	14 379 000,00	605430,1067
Pastwiska	[ha]	3 933 680,00	4 047 000,00	284309,7285
Inne kategorie gruntów rolnych	[ha]	12 837 560,00	12 811 400,00	1145890,005
Powierzchnia lasów	[ha]	9012533	9005600	90021,16
Bezpośrednie zużycie energii w gospodarstwach	[tony] ekwiwalentu ropy	4 453,27	4 627,70	566,0749
Maszyny rolnicze	[szt.]	1 388 304,13	1 415 957,00	88672,2466
Całkowita emisja z rolnictwa GGC	tony ekwiwalentu [CO <sub>2</sub> ]	25 354 066,67	24 982 000,00	3350625,048

*Źródło: obliczenia własne*

Zamieszczona w tabeli 2 macierz korelacji opisuje stopień zależności pomiędzy zmiennymi. Im większa wartość bezwzględna tym wyższa korelacja między nimi (np. pastwiska – grunty orne i stała powierzchnia upraw lub inne grunty rolne - powierzchnia lasów). Uznaje się w literaturze, że dla zmiennych o współczynniku korelacji < 0,3 nie powinno prowadzić się dalszej analizy [Sokołowski i Sagan 2011].

W kolejnym kroku badań wyznaczono wartości własne macierzy korelacji (tabela 3), które odzwierciedlają istotność głównych składowych w wyjaśnianiu zasobów informacyjnych zmiennych wejściowych (udział procentowy w zmienności zbioru danych). W celu określenia liczby głównych składowych skorzystano z metody wartości własnej większej od jedności tzw. kryterium Kaisera. Im współczynnik korelacji zmiennej ze składową wyższy, tym bardziej jest ta zmienna istotna dla danej składowej. W analizowanym przypadku dwie pierwsze składowe główne są decydujące i w niemal 96% opisują zmienności pierwotnych danych.

Tabela 2. Macierz współczynników korelacji pomiędzy zmiennymi  
Table 2. Matrix of coefficients of correlation between variables

Zmienna	Współczynnik korelacji					
	A	B	C	D	E	F
A	1,000000					
B	0,956778	1,000000				
C	-0,960936	-0,866693	1,000000			
D	-0,863956	-0,726083	0,935293	1,000000		
E	-0,165200	0,040224	0,351196	0,483020	1,000000	
F	-0,718107	-0,564334	0,842749	0,884114	0,678457	1,000000

gdzie: A – grunty orne i stała powierzchnia upraw, B – pastwiska, C – inne grunty rolne, D – powierzchnia lasów, E – zużycie energii w gospodarstwach rolnych, F – maszyny rolnicze.

where: A - arable lands and a constant surface area of crops, B - grazing lands, C - other arable lands, D - surface area of forests, E - energy consumption in agricultural farms, F - agricultural machinery.

*Źródło: obliczenia własne*

Tabela 3. Wartości własne macierzy korelacji  
Table 3. Eigenvalues of the correlation matrix

Nr wartości	Wartości własne (korelacje), statystyki pokrewne			
	Wartość własna	% ogółu wariancji	Skumulowana wartość własna	Skumulowana %
1	4,494351	74,90585	4,494351	74,9059
2	1,264492	21,07487	5,758843	95,9807
3	0,120515	2,00859	5,879358	97,9893
4	0,083581	1,39301	5,962939	99,3823
5	0,028494	0,47490	5,991433	99,8572
6	0,008567	0,14278	6,000000	100,0000

*Źródło: obliczenia własne*

W tabeli 4 zamieszczono wartości i znaki elementów wektorów własnych, z których można odczytać siłę i kierunek wpływu poszczególnych zmiennych na główne składowe.

Tabela 4. Wektory własne macierzy korelacji  
Table 4. Eigenvectors of the correlation matrix

Zmienna	Wektory własne macierzy korelacji (dane)					
	Czynn. 1	Czynn. 2	Czynn. 3	Czynn. 4	Czynn. 5	Czynn. 6
A	0,445459	0,272839	0,239063	0,107648	0,164868	0,794484
B	0,397372	0,447919	0,460935	-0,180520	-0,495887	-0,387958
C	-0,465190	-0,083771	0,019490	-0,090021	-0,750999	0,451773
D	-0,452961	0,087426	0,545671	-0,578679	0,388915	0,057455
E	-0,207959	0,777897	-0,555244	-0,194553	0,068240	0,028733
F	-0,422893	0,324258	0,352072	0,758288	0,083691	-0,100294

*Źródło: obliczenia własne*

Interpretacja uzyskanych składowych głównych dokonywana jest na podstawie wartości ich współczynników, które stanowią jednocześnie współczynniki korelacji liniowej pomiędzy zmiennymi wejściowymi i głównymi składowymi. Pierwsza główna składowa przenosi 75% informacji zawartych głównie w zmiennych grunty orne wraz ze stałą powierzchnią upraw i pastwiska oraz ujemnie skorelowanych zmiennych tj. inne grunty rolne, powierzchnia lasów a także ilość maszyn rolniczych. Druga zmienna wyjaśnia 21% zmienności danych poprzez zmienną ilość zużytej w gospodarstwach rolnych energii (tabela 5).

Tabela 5. Wartość współczynników głównych składowych  
Table 5. Value of principal components coefficients

Zmienna	Współrzędne czynnikowe zmiennych, na podstawie korelacji (danych)	
	Czynn. 1	Czynn. 2
A	0,944369	0,306806
B	0,842424	0,503683
C	-0,986197	-0,094200
D	-0,960273	0,098311
E	-0,440872	0,874743
F	-0,896529	0,364627

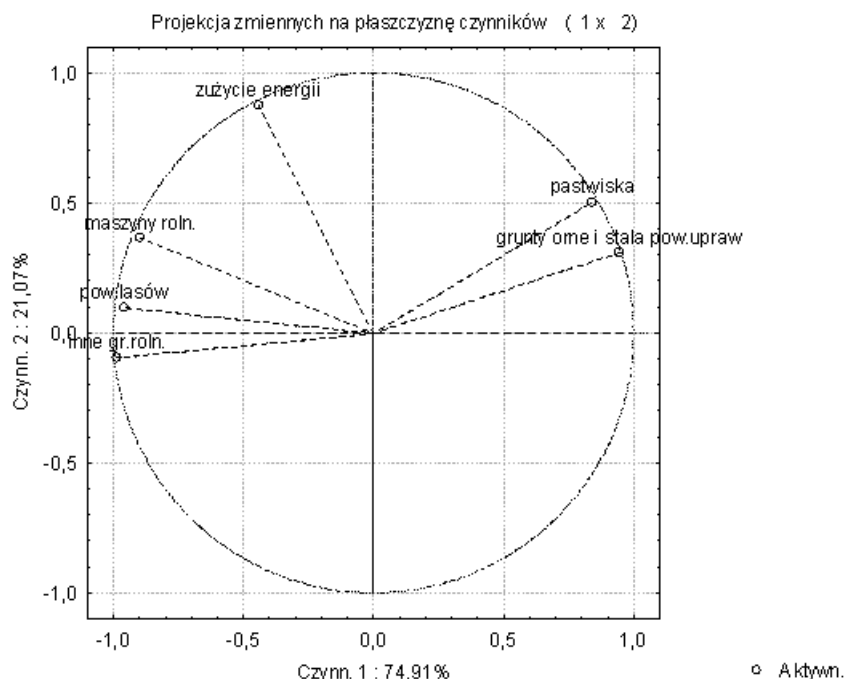
*Źródło: obliczenia własne*

Zależności pomiędzy zmiennymi pierwotnymi i uzyskanymi składowymi głównymi przedstawiono graficznie na rysunku 1. Każda zmienna jest reprezentowana przez wektor. Kierunek i długość wektorów określa w jakim stopniu poszczególne zmienne wpływają na składowe główne.

W analizowanym przykładzie niemal wszystkie zmienne wejściowe położone są w pobliżu okręgu co oznacza, iż większa część informacji zawartych w tych zmiennych jest przenoszona przez główne składowe.

Fakt położenia obok siebie dwóch zmiennych świadczy o silnej dodatniej korelacji (np. inne grunty rolne oraz powierzchnia lasów, lub grunty orne i stała powierzchnia upraw oraz pastwiska). Wektory prostopadle wobec siebie usytuowane świadczą o braku korelacji (np. pastwiska oraz zużycie energii w gospodarstwach rolnych). Zmienne położone po przeciwnej stronie są ujemnie skorelowane [Gramacki J. i Gramacki A. 2008] (np. grunty orne i stała powierzchnia upraw oraz powierzchnia lasów lub pastwiska względem innych gruntów rolnych).

Istotną rolę w odniesieniu do emisji odgrywa sposób użytkowania gruntów. Użytki zielone w ostatnich latach zwracają szczególną uwagę ze względu na wysoką emisję podtlenku azotu [Marcinkowski 2010]. Badania naukowców dowodzą, iż zmiana użytków zielonych na pola uprawne powoduje wzrost emisji GGC [Mosier i in. 1997]. Potwierdzeniem powyższych wniosków może być uzyskany w analizie wynik świadczący o silnej korelacji gruntów ornych i stałej powierzchni upraw ze zmienną pastwiska.



Źródło: obliczenia własne

Rys. 1. Wykres zmiennych. Położenie wektorów ładunków względem dwóch składowych głównych

Fig. 1. Plot of variables. Location of load vectors towards two principal components

## Wnioski

Wpływ konkretnych analizowanych czynników na emisję gazów cieplarnianych z rolnictwa można opisać za pomocą dwóch głównych składowych. Pierwsza główna składowa przenosi około 75% informacji o emisjach zawartych w zmiennych wejściowych. W jej skład wchodzi zmienne skorelowane dodatnio: grunty orne i stała powierzchnia upraw (A) oraz pastwiska (B) a także ujemnie: inne grunty rolne (C), powierzchnia lasów (D) oraz ilość maszyn rolniczych (F). Powyższe korelacje świadczą o powiązaniu wysokich wartości zmiennych A i B z niskimi wartościami C, D i F oraz odpowiednio ze wzrostem tych drugich maleją pierwsze. Druga składowa odpowiada przede wszystkim za zużycie energii w gospodarstwach rolnych. Można stwierdzić zatem przydatność metody analizy składowych głównych w redukcji wymiarowości zmiennych w opisie rozważanego zagadnienia.

## Bibliografia

- Bebkiewicz K., Cieślińska J., Dębski B., Kanafa M., Kargulewicz I., Olecka A., Olendrzyński K., Skośkiewicz J., Żaczek M.** (2011): Krajowy Raport inwentaryzacyjny 2011, Inwentaryzacja gazów cieplarnianych w Polsce dla lat 1988-2009, Raport wykonany na potrzeby Ramowej konwencji Narodów Zjednoczonych w sprawie zmian klimatu oraz Protokołu z Kioto. KOBIZE, Warszawa.
- Burczyk P., Miatkowski Z., Turbiak J.** (2011): Wstępne rozpoznanie emisji N<sub>2</sub>O w wybranych siedliskach łąkowych w różnych regionach Polski, *Woda-Środowisko-Obszary Wiejskie*, z. 4(36), 57-67
- Hand D., Mannila H., Smyth P.** (2005): Eksploracja danych. WNT, Warszawa, ISBN 83-204-3053-4.
- Freney R.** (1997): Emission of nitrous oxide from soils used for agriculture. *Nutrient Cycling in Agroecosystems*. 49, 1-6.
- Gramacki J., Gramacki A.** (2008): Statystyczne odkrywanie zależności w danych. *Przegląd telekomunikacyjny*. 6, 711-713.
- Marcinkowski T.** (2010): Emisja gazowych związków azotu z Rolnictwa. *Woda-Środowisko-Obszary Wiejskie*. 10 Z. 3(31), 175-189.
- Mosier A.R., Parton W.J., Valentine D.W., Ojima D.S., Schimel D.S., Heinemeyer O.** (1997): CH<sub>4</sub> and N<sub>2</sub>O fluxes in Colorado shortgrass steppe. 2. Long-term impact of land use change. *Global Biogeochem. Cycles*. 11, 29-42.
- Myczko A., Karłowski J., Szulc R.** (2003): Detailed Investigations of Methane and Dinitrogen Oxide Emissions from Enteric Fermentation and from Animal Excreta, in: *Elimination of Agricultural risks to health and environment*, edited by A. Myczko, Scientific Network Agrorisks. IB-MER, Poznań, ISBN 83-920061-0-0.
- Nabuurs G.J., Masera O., Andrasko K., Benitez-Ponce P., Boer R., Dutschke M., Elsidig E., Ford-Robertson J., Frumhoff P., Karjalainen T., Krankina O., Kurz W.A., Matsumoto M., Oyhantcabal W., Ravindranath N.H., Sanz Sanchez M.J., Zhang X.** (2007): "Forestry" In *Climate Change 2007: Mitigation. Contribution of Working Group III to the Fourth Assessment Report of the Intergovernmental Panel on Climate Change* [Metz B., Davidson O.R., Bosch P.R., Davey R., Meyer L.A. (eds)]. Cambridge University Press. Cambridge United Kingdom and New York USA.
- Pietrzak S., Sapek A., Oenema O.** (2002): Ocena emisji podtlenku azotu (N<sub>2</sub>O) ze źródeł rolniczych w Polsce. *Zeszyty Edukacyjne* 8. IMUZ, Falenty, 23-36.
- Oenema O., Pietrzak S., Sapek A.** (1999): Controlling nitrous oxide emissions from grassland farming systems in Poland; preliminary results. W: *Nitrogen cycle and balance in Polish agriculture*. Red. Sapek A. Falenty. IMUZ, 126-139.
- Sokołowski A., Sagan A.** Przykłady stosowania analizy danych w marketingu i badaniu opinii publicznej, [dostęp 21-11-2011]. Dostępny w Internecie: <http://www.statsoft.pl/czytelnia/marketing/adwmarketingu.html>
- Steinfeld H., Gerber P., Wassenaar T., Castel V., Rosales M., de Haan C.** (2006): *Livestock's long shadow: environmental issues and options*. Food and Agriculture Organisation, Rome. ISBN 978-92-5-105571-7.
- Tadeusiewicz R.** (2006): Data mining jako szansa na relatywnie tanie dokonywanie odkryć naukowych poprzez przekopywanie pozornie całkowicie wyeksploatowanych danych empirycznych, W: *Statystyka i data mining w badaniach naukowych*, red. J. Wątroba. StatSoft, Kraków, ISBN 10: 83-88724-31-2.
- Turbiak J., Miatkowski Z., Chrzanowski S., Gąsiewska A., Burczyk P.** (2011): Emisja podtlenku azotu z gleby torfowo-murszowej w dolinie Biebrzy w zależności od warunków wodnych. *Woda-Środowisko-Obszary Wiejskie*. 11 Z. 4(36), 239-245.

**Winnicki S., Jugowar J.L., Karłowski J.** (2010): Monitorowanie gospodarstw rolnych jako źródeł emisji rolniczych, Raport z realizacji umowy pomiędzy Ministerstwem Rolnictwa i Rozwoju Wsi a Instytutem Technologiczno-Przyrodniczym w Falentach, wersja 2 poprawiona, Poznań.  
Organization for Economic Co-operation and Development. Stats Extracts, [dostęp 21-11-2011]  
Dostępny w Internecie: <http://stats.oecd.org/index.aspx>  
World Resources Institute (WRI) Climate Analysis Indicators Toolkit (CAIT) [dostęp 16.03.2011].  
Dostępny w Internecie: <http://cait.wri.org/figures.php?page=/World-FlowChart>

## **PRINCIPAL COMPONENTS ANALYSIS IN EVALUATION OF RELATION BETWEEN LAND USE AND GREENHOUSE GASES EMISSION FROM AGRICULTURE**

**Abstract.** A method of principal components analysis was used in the research in order to separate significantly correlated variables related to emission of greenhouse gases from agriculture towards a method of using lands. The method gives opportunity to replace an input set of correlated features through a low number of non-correlated principal components, which constitute linear rearrangement of variables. Minimisation of the number of variables required to explain a particular variable simplifies interpretation of results. The research was carried out with the use of Statistica 9.0 programme. Two main components were separated, which in 96% explain variations of input data. The first main component explains 75% of variations included mainly in variables arable lands along with a constant surface area of crops and grazing lands as well as negatively correlated variables that is other arable lands, surface area of forests and number of agricultural machinery. The second component explains 21% of variations of data through a variable number of energy used in agricultural farms.

**Key words:** principal components analysis, PCA, main greenhouse gases, emission from agriculture

### **Adres do korespondencji:**

Alicja Kolasa-Więcek; e-mail: [a.kolasa-wiecek@po.opole.pl](mailto:a.kolasa-wiecek@po.opole.pl)  
Wydział Ekonomii i Zarządzania  
Politechnika Opolska  
ul. Waryńskiego 4  
45-047 Opole



*Dofinansowanie ze środków Wojewódzkiego Funduszu Ochrony Środowiska  
i Gospodarki Wodnej w Opolu*