

ANALIZA KORELACJI I REGRESJI POMIĘDZY ZAWARTOŚCIĄ  
METALI CIĘŻKICH W GLEBACH GÓRNOŚLĄSKIEGO OKRĘGU  
PRZEMYSŁOWEGO A ICH PODATNOŚCIĄ MAGNETYCZNĄ

JAROSŁAW ZAWADZKI<sup>1</sup>, TADEUSZ MAGIERA<sup>2</sup>, ZYGMUNT STRZYSZCZ<sup>2</sup>

<sup>1</sup> Politechnika Warszawska, Instytut Systemów Inżynierii Środowiska, Zakład Metod Matematycznych  
ul. Nowowiejska 20, 00-661 Warszawa

<sup>2</sup> Instytut Podstaw Inżynierii Środowiska Polskiej Akademii Nauk  
ul. M. Skłodowskiej-Curie 34, 41-819 Zabrze

Keywords: heavy metals, magnetic susceptibility, topsoil.

CORRELATION AND REGRESSION ANALYSIS OF HEAVY METAL CONTENT  
AND MAGNETIC SUSCEPTIBILITY IN SOILS FROM UPPER SILESIAN  
INDUSTRIAL REGION

The aim of this paper was to investigate the relationship between magnetic susceptibility of topsoil and content of heavy metal being the result of urban and industrial dust-fall. Tools for this study were some complementary statistic methods such as: correlation analysis using Pearson correlation coefficient, Spearman rank correlation coefficient, stepwise regression and „*chi-kwadrat*” test. The base for statistic analysis was dataset of ca. 600 topsoil samples (20 cm) from Upper Silesian Industrial Region, including content of As, Cd, Co, Cr, Cu, Fe, Mn, Ni and Pb as well as values of low-field specific magnetic susceptibility ( $\chi$ ) measured for the same samples.

The study clearly confirms a significant correlation between the level of inorganic contamination and the measured susceptibility value, although the correlations in soil are usually more sophisticated. The most often observed correlation coefficients between magnetic susceptibility and heavy metals content were on medium ( $r = 0.5-0.7$ ) and high ( $r = 0.7-0.9$ ) level. The statistic analysis of the studied parameters can not be based only on Pearson correlation coefficient. The use of some complementary statistic methods allows for more correct interpretation of existing relationships. The comparable values of Pearson linear correlation coefficient and Spearman rank the correlation coefficient, observed in studied dataset within the range of accuracy used, shows the existence of linear correlation. The similar conclusions have been drawn from the analysis of reverse stepwise regression. The observed model of linear multiple regression explains almost 80% of variability of the  $\chi$  value. Foregoing statistical analysis confirms some earlier observations that magnetometry based on topsoil magnetic susceptibility measurement could be a very interesting and alternative or complementary method for monitoring anthropogenic soil pollution and especially heavy metal contamination level.

Streszczenie

Celem niniejszej pracy było poznanie współzależności pomiędzy wartością podatności magnetycznej gleb a zawartością metali ciężkich pochodzących z opadu pyłów miejsko-przemysłowych przy pomocy kilku komplementarnych metod statystycznych: analizy korelacji przy użyciu współczynników korelacji Pearsona i współczynników korelacji rang Spearman'a, regresji krokowej, testu niezależności „*chi-kwadrat*”. Podstawą do obliczeń statystycznych było około 600 pomiarów zawartości As, Cd, Co, Cr, Cu, Fe, Mn, Ni i Pb, w górnej (20 cm) warstwie gleby z obszaru Górnośląskiego Okręgu Przemysłowego (GOP). Dane dotyczące zawartości metali ciężkich uzupełniono pomiarami niskopolowej specyficznej podatności magnetycznej ( $\chi$ ) dokonanymi w tych samych próbach.

Stwierdzono istnienie wyraźnej korelacji pomiędzy poziomem zanieczyszczeń antropogenicznych a wielkością mierzonej podatności, chociaż zależności te w glebie są na ogół skomplikowane. Najczęściej występują średnie na poziomie  $r = 0,5-0,7$  i wysokie, na poziomie  $r = 0,7-0,9$  korelacje pomiędzy wielkością podatności a zawartościami badanych metali. Stwierdzono, że badania zależności pomiędzy podatnością magnetyczną, a zawartością metali ciężkich nie powinny być oparte wyłącznie o wyznaczanie współczynnika korelacji Pearsona. Zastosowanie kilku komplementarnych metod statystycznych pozwala ocenić istniejące współzależności w sposób bardziej prawidłowy. Porównanie wartości współczynnika korelacji liniowej Pearsona oraz współczynnika korelacji rang Spermmana wskazuje, że w zakresie stosowanej dokładności związki korelacyjne są liniowe. Podobnych wniosków dostarczają wyniki regresji krokowej wstecznej. Otrzymany model liniowy regresji wielokrotnej wyjaśnia prawie 80% zmienności podatności magnetycznej. Powyższe analizy statystyczne potwierdzają wcześniejsze spostrzeżenia, że magnetometria bazująca na pomiarach podatności magnetycznej gleb stanowi bardzo interesującą alternatywę dla innych metod w zakresie monitoringu zanieczyszczeń antropogenicznych gleby, a w szczególności zanieczyszczeń metalami ciężkimi.

## WSTĘP

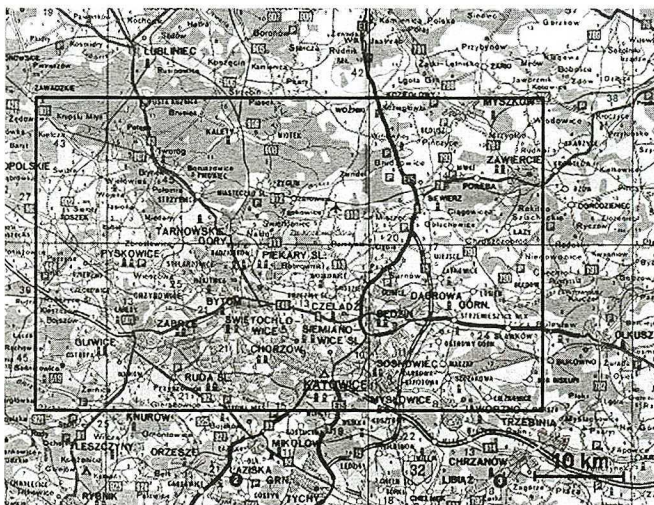
Istnieje wiele dowodów potwierdzających związek pomiędzy wzrostem specyficznej podatności magnetycznej ( $\chi$ ) w glebach a czynnikami o charakterze antropogenicznym [3, 10–12, 14]. Ważnym czynnikiem wskazującym na antropogeniczny charakter  $\chi$  jest zwiększona zawartość metali ciężkich oraz innych zanieczyszczeń w badanych glebach. Główną przyczyną tych zanieczyszczeń jest emisja pyłów przemysłowych, które zawierają związki ferromagnetyczne pochodzenia antropogenicznego. Związki te powstają w wyniku skomplikowanych procesów technologicznych w różnych gałęziach przemysłu. Istnienie różnorodnych związków magnetycznych w tym ferromagnetycznych, paramagnetycznych oraz diamagnetycznych pochodzenia naturalnego i antropogenicznego powoduje, że związek pomiędzy podatnością magnetyczną, a zawartością badanych zanieczyszczeń, mimo że bardzo wyraźny jest równocześnie bardzo złożony i niełatwy do wyjaśnienia. Stwierdzenie istnienia powyższych związków pozwala na wykorzystanie podatności magnetycznej w monitoringu środowiska. Metoda ta łączy prostotę pomiarów, możliwość szybkiego wykonywania pomiarów terenowych oraz znaczne obniżenie kosztów badań ze względu na zmniejszenie liczby analiz chemicznych koniecznych do wykonania [5–7].

Zrozumienie związku pomiędzy zawartością zanieczyszczeń antropogenicznych a podatnością magnetyczną wymaga zarówno badań fizykochemicznych jak również obliczeń o charakterze statystycznym, ze względu na różnorodność źródeł i rodzaju zanieczyszczenia gleb. Celem przeprowadzonych badań było zbadanie współzależności pomiędzy wartością  $\chi$  a zawartością zanieczyszczeń przy pomocy kilku komplementarnych metod statystycznych: analizy korelacji przy użyciu współczynników korelacji Pearsona i współczynników korelacji rang Spearman'a, regresji krokowej, testu niezależności „*chi-kwadrat*”. Zastosowanie kilku metod statystycznych pozwala lepiej zrozumieć badaną zależność pomiędzy  $\chi$  a zawartością zanieczyszczeń m.in., dlatego iż stosowanie pojedynczej metody nie jest najczęściej wiarygodne i wymaga porównania z wynikami innych metod. W szczególności interesujące było zagadnienie zasadności stosowania do wyżej opisywanych badań popularnego współczynnika korelacji Pearsona [13]. W celu uzyskania możliwie prostego modelu wiążącego podatność magnetyczną z zawartością kilku rodzajów zanieczyszczeń zastosowano analizę regresji krokowej. Metoda ta może okazać się perspektywiczna jako skuteczne i niezbyt skomplikowane narzędzie analizy wyników w wyżej opisywanych badaniach.

## METODYKA BADAŃ

Podstawą do obliczeń statystycznych było około 600 pomiarów zawartości As, Cd, Co, Cr, Cu, Fe, Mn, Ni i Pb w próbkach pobranych z obszaru Górnośląskiego Okręgu Przemysłowego (GOP) wykonanych przez Państwowy Instytut Geologiczny w ramach opracowywania „Atlasu geochemicznego Polski” [4]. Dane chemiczne zostały uzupełnione pomiarami niskopolowej specyficznej podatności magnetycznej ( $\chi$ ) mierzonej w laboratorium Instytutu Podstaw Inżynierii Środowiska PAN Zabrze. Pomiary  $\chi$  zostały wykonane na bazie tych samych prób, w których przeprowadzono oznaczenia chemiczne. Pomiary te zostały wykonane w ramach projektu naukowego „Geogeny lub antropogeny charakter glebowych anomalii magnetycznych na obszarze Polski – studium porównawcze zapisu petromagnetycznego i geochemicznego” finansowanego przez KBN (nr projektu 6 P04D 014 16). Analizą statystyczną objęto obszar o kształcie prostokąta, o wymiarach 72 km w kierunku wschód-zachód oraz 38 km w kierunku północ-południe (Rys. 1). Siatka punktów pomiarowych była nieregularna. Średnia odległość między punktami pomiarowymi wynosiła 2,3 km.

Próbki gleb pochodziły z zasobów archiwalnych Państwowego Instytutu Geologicznego. Poboru dokonywano w latach 1991–1992. Próbki o masie około 0,5 kg pobierano za pomocą ręcznej sondy o średnicy około 8 cm z głębokości 0,0–0,2 m. Na terenach leśnych przed poborem próby odrzucano ściółkę leśną, co miało zniwelować różnice między terenami rolnymi i leśnymi. Następnie próbki suszono w temperaturze pokojowej, przesiewano przez sito 1 mm i przez kwartowanie uzyskiwano próbkę analityczną o masie około 100 g. Oznaczenia zawartości metali ciężkich wykonano metodą spektrometrii emisyjnej ze wzbudzeniem plazmowym (ICP) spektrometrem firmy Philips 8060 [4]. Niskopolową podatność magnetyczną mierzone miernikiem podatności MS2B „Bartington” i przeliczano na podatność specyficzną (masową) uwzględniając gęstość próby.



Rys. 1. Położenie badanego obszaru na tle mapy GOP  
Localization of study area on the map of Upper Silesian Industrial Region (USIR)

## WYNIKI BADAŃ

*Statystyka opisowa*

Pierwszym etapem badań było wykonanie obliczeń podstawowych miar statystyki opisowej zawartości badanych pierwiastków oraz podatności w glebach GOP. Porównując dane z tabeli 1 z liczbami granicznymi podanymi przez Ministerstwo Środowiska [8] można wysnuć ogólny wniosek, że o ile wartości średnie zanieczyszczeń mieszczą się w granicach od zawartości podwyższonej do średniej, to wartości maksymalne leżą na ogół w zakresie bardzo silnych zanieczyszczeń. Ogólną cechą analizowanych danych są bardzo wysokie wartości odchyłeń standardowych, najczęściej większe od wartości średniej oraz wysoka asymetria prawostronna. Są to często spotykane cechy zbiorów danych dotyczących zanieczyszczenia środowiska.

Tabela 1. Statystyki opisowe zawartości badanych pierwiastków wraz z liczbami granicznymi oraz wartości  $\chi$  w glebach GOP

Descriptive statistic of heavy metal content, boundary values and  $\chi$  value in soil of USIR

	n	Średnia Mean mg/kg	Odchylenie stand. Standard deviation mg/kg	Min. mg/kg	Max. mg/kg	Skośność stand. Standard skewness	Kurtoza stand. Standard kurtosis	Wartości graniczne <sup>3)</sup> Boundary values <sup>3)</sup> [mg/kg]
As	593	8,78	14,15	2,5	238	91,50	625,88	20
Cd	593	4,90	10,27	0,25	159,5	89,51	544,32	4
Co	593	2,73	2,16	0,5	21	22,25	56,48	20
Cr	593	6,62	7,30	0,5	95	65,38	309,88	150
Cu	593	16,58	41,29	0,5	805	132,85	1160,45	150
Fe <sup>1)</sup>	593	0,76	0,54	0,05	3,57	14,41	14,37	no limits
Mn	593	380,47	447,92	5,0	7000	66,73	415,72	no limits
Ni	593	6,61	6,52	0,5	89	45,82	224,828	100
Pb	593	226,20	896,34	3,0	16972	139,05	1147,08	100
Zn	593	505,28	1041,23	15	11899	65,57	277,87	300
$\chi$ <sup>2)</sup>	593	121,31	186,11	1,25	1768,1	39,15	102,39	no limits

<sup>1)</sup> zawartość Fe jest wyrażona w % – Fe content in %,

<sup>2)</sup> wartości  $\chi$  wyrażone są w  $10^{-8} \text{m}^3/\text{kg}$  –  $\chi$  values in  $10^{-8} \text{m}^3/\text{kg}$ ,

<sup>3)</sup> na podstawie Rozporządzenie Ministra Środowiska [8] – based on Directive of the Polish Ministry of Environment [8]

W niniejszej pracy korzystając z bardzo dużego (około 600 próbek) zbioru danych można było obliczyć realistyczne wartości współczynnika korelacji Pearsona pomiędzy  $\chi$  a zawartością badanych pierwiastków.

Tabele 2 i 3 pokazują odpowiednio wartości współczynników korelacji Pearsona oraz rang Spearmana. Współczynniki korelacji obliczane były albo z całego zbioru danych (liczba przypadków 597 – Wariant 1) albo ze zbioru danych po odrzuceniu punktów odstających (liczba przypadków 526 Wariant 2). W tym drugim przypadku wykonywano regresję krokową na całym zbiorze danych, a następnie analizowano anomalne reszty. Jako wartości odstające przyjęto te, których moduł wartości studentyzowanej reszty był większy od 3. W rezultacie takiej procedury usunięto kilkadziesiąt punktów odstających.

Tabela 2. Tabela współczynników korelacji liniowej Pearsona (współczynniki istotne na poziomie ufności 95%). Wartości nad górze i na dole oznaczają odpowiednio współczynniki korelacji dla wszystkich danych oraz dla danych po usunięciu wartości odstających

Pearson linear correlation coefficients (coefficients are significant on confidence level of 95%). Values on the top of the cell are coefficients for the whole dataset. Values on the bottom of the cell are coefficients after the removal of extreme values

Elements	Cd	Co	Cr	Cu	Fe	Mn	Ni	Pb	Zn	$\chi$
As	0,6314 0,6789	0,2475 0,4728	0,4086 0,4923	0,4886 0,6204	0,5729 0,6408	0,4447 0,5194	0,4778 0,5411	0,6005 0,7106	0,7561 0,7060	0,6253 0,6303
Cd		0,1333 0,3754	0,1754 0,3926	0,3076 0,6130	0,4076 0,5000	0,3850 0,5212	0,3135 0,5183	0,6251 0,8711	0,8967 0,9273	0,3603 0,6209
Co			0,5433 0,8563	0,4133 0,6845	0,6989 0,8545	0,5762 0,7700	0,8189 0,8719	0,0374 0,3282	0,1349 0,4471	0,3153 0,4727
Cr				0,5352 0,7365	0,6741 0,8893	0,5663 0,7183	0,7469 0,8910	0,1661 0,3562	0,2246 0,4840	0,5454 0,5413
Cu					0,6355 0,7685	0,4324 0,6527	0,6641 0,8260	0,2283 0,6266	0,3582 0,7155	0,8391 0,8204
Fe						0,6412 0,7717	0,7803 0,8748	0,3789 0,4797	0,4497 0,5702	0,5834 0,6109
Mn							0,5962 0,7293	0,2448 0,4658	0,4049 0,5905	0,4497 0,5150
Ni								0,2218 0,4532	0,3709 0,6197	0,6198 0,6236
Pb									0,7923 0,8600	0,2378 0,6480
Zn										0,4228 0,7046

W celu dodatkowej weryfikacji wykonano również test niezależności „*chi-kwadrat*”, w celu potwierdzenia zależności (stochastycznych) pomiędzy  $\chi$  a zawartością metali ciężkich. Test ten wykazał brak niezależności pomiędzy  $\chi$  za zawartością każdego z pierwiastków na poziomie ufności 0,95.

#### *Wyniki regresji krokowej wstecznej*

Celem przeprowadzenia regresji krokowej było znalezienie wiarygodnego modelu wiążącego wartość  $\chi$  z zawartością badanych pierwiastków, zawierającego jedynie taką liczbę zmiennych, które istotnie wpływają na jakość modelu. Zmienną zależną  $\chi$  (wyjaśnianą) powiązано z zawartościami badanych pierwiastków. Regresję przeprowadzono w dwóch wariantach. W pierwszym z nich używano wszystkich danych bez analizy wartości odstających. W drugim przypadku przed wykonaniem regresji przeanalizowano wszystkie dane GOP pod kątem występowania w zbiorze danych wartości odstających. Jako wartości odstające przyjęto te, których moduł wartości studentyzowanej reszty był większy od 3. W obu przypadkach badano odpowiednie założenia modelu regresji liniowej jak na przykład *r* – homoscedastyczność modelu [9]. Dla porównania dokładności wyników uzyskanych metodą regresji krokowej wstecznej przedstawiono również model i wyniki otrzymane z klasycznej regresji wielokrotnej, w której jako zmienne niezależne przyjęto zawartości wszystkich mierzonych pierwiastków. Istotą regresji krokowej jest wyeliminowanie zmiennych niezależnych silnie ze sobą skorelowanych. Punktem wyjściowym do wykonania regresji krokowej wstecznej było 10 zmiennych, wliczając w to stałą modelu. Przyjęto domyślną wartość do usunięcia  $F = 4,0$ ,

informująca o tym jak mało istotna dla regresji musi być dana zmienna, aby ją usunąć z modelu. Wartość ta odpowiada poziomowi ufności równym około 95%.

Tabela 3. Tabela współczynników korelacji rang Spearmana (współczynniki są istotne na poziomie ufności 95%). Wartości nad górze i na dole oznaczają odpowiednio współczynniki korelacji rang dla wszystkich danych oraz dla danych po usunięciu wartości odstających  
Spearman rank correlation coefficients (coefficients are significant on confidence level of 95%). Values on the top of the cell are coefficients for the whole dataset. Values on the bottom of the cell are coefficients after the removal of extreme values

Elements	Cd	Co	Cr	Cu	Fe	Mn	Ni	Pb	Zn	$\chi$
Wariant 1 Alternative 1 Wariant 2 Alternative 2	Cd	Co	Cr	Cu	Fe	Mn	Ni	Pb	Zn	$\chi$
As	0,6199 0,7101	0,2221 0,5107	0,2646 0,5484	0,3727 0,6560	0,4911 0,6843	0,2813 0,5539	0,4132 0,5919	0,7626 0,7410	0,6581 0,7401	0,4618 0,6617
Cd		0,1731 0,4190	0,1593 0,4535	0,1865 0,6318	0,4214 0,5473	0,3600 0,5671	0,3181 0,5603	0,6378 0,8809	0,8516 0,9299	0,2989 0,6324
Co			0,5485 0,8645	0,2489 0,7162	0,7406 0,8601	0,5295 0,7626	0,7798 0,8817	0,0875 0,3806	0,2043 0,4886	0,4139 0,5271
Cr				0,3290 0,7784	0,6029 0,9033	0,3998 0,7324	0,7690 0,9090	0,1104 0,4268	0,2443 0,5414	0,4765 0,6060
Cu					0,3510 0,8025	0,1661 0,6695	0,3919 0,8501	0,2423 0,6505	0,3063 0,7352	0,4078 0,8324
Fe						0,5864 0,7789	0,7657 0,8936	0,3219 0,5355	0,4857 0,6173	0,6535 0,6663
Mn							0,5057 0,7442	0,1837 0,5175	0,3669 0,6288	0,3344 0,5548
Ni								0,2305 0,5082	0,3914 0,6562	0,6415 0,6719
Pb									0,6004 0,8761	0,2331 0,6714
Zn										0,3694 0,7214

### Wariant 1 – wszystkie dane

W trakcie procedury regresji krokowej wstecznej zostały usunięte z modelu: Cd i Mn. Końcowe równanie regresji krokowej wstecznej jest następujące:

$$\chi = -29,22 + 3,00724*AS\_PPM - 36,2393*CO\_PPM - 3,33961*CR\_PPM + 0,536258*CU\_PPM + 184,235*FE\_PROC + 17,0018*NI\_PPM - 0,0365503*PB\_PPM - 0,0152296*ZN\_PPM \quad (1a)$$

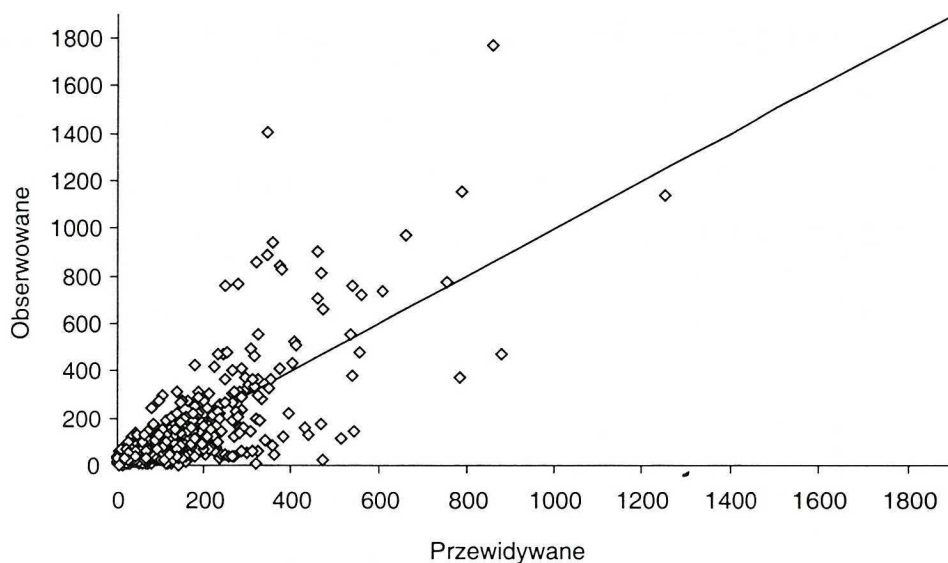
Poniżej dla porównania pokazano równanie regresji otrzymane w wyniku klasycznej regresji liniowej przy zastosowaniu wszystkich zmiennych niezależnych, tzn. wszystkich zawartości badanych pierwiastków):

$$\chi = -29,4564 + 2,99279*AS\_PPM^{1)} + 0,148529*CD\_PPM - 35,3166*CO\_PPM - 3,24926*CR\_PPM + 0,527505*CU\_PPM + 188,627*FE\_PROC^{2)} - 0,0167543*MN\_PPM + 16,943*NI\_PPM - 0,0371615*PB\_PPM - 0,0147815*ZN\_PPM \quad (1b)$$

<sup>1)</sup> wartość w ppm

<sup>2)</sup> wartość w %

Poniżej, na rysunku 2 przedstawiono wykres rozrzutu obserwowanych i przewidywanych wartości podatności magnetycznej  $\chi$  oraz prostą regresji, otrzymaną na podstawie modelu regresji krokowej wstecznej podanym wzorem (1a).



Rys. 2. Wykres rozrzutu obserwowanych i przewidywanych wartości podatności magnetycznej oraz prosta regresji otrzymana na podstawie modelu regresji krokowej wstecznej danej równaniem (1a)  
Scatter diagram of observed (vertical axis) and predicted (horizontal axis) values of magnetic susceptibility and regression line obtained on the base of reverse stepwise regression model described by formula (1a)

Dokładne parametry linii regresji pokazanej na rysunku 2 są następujące:

- współczynnik determinacji (informuje ile zmienności podatności magnetycznej jest wyjaśniana przez model) –  $R^2 = 57,4722\%$  (57,5696%);
- współczynnik determinacji (poprawiony ze względu zmianę ilości zmiennych niezależnych) –  $R_c^2 = 56,8896\%$  (56,8406%);
- błąd standardowy estymacji – 65,5708 (65,7689);
- błąd średni, bezwzględny (średnia wartość reszt) – 24,4649 (65,7689);
- statystyka Durbina-Watsona = 1,65888 (1,64159) (wysoka wartość statystyki Durbina-Watsona – powyżej 1,4 – informuje o braku istotnych statystycznie korelacji między resztami potwierdzając prawidłowość modelu).

Statystycznie istotny związek między zmiennymi został stwierdzony na poziomie ufności 0,99%, o czym informuje bliska zeru wartość parametru P analizy wariancji (Tab. 4).

Niewysokie wartości współczynników determinacji otrzymanego modelu zawierającego 8 zmiennych niezależnych zawartości As, Co, Cr, Cu, Fe, Ni, Pb i Zn oraz stałą wynikały z występowania wyraźnych punktów odstających, które usunięto w następnym etapie (warian nr 2).

Tabela 4. Wyniki analizy wariancji wykonanej na wszystkich danych  
Results of variance analysis on the base of whole dataset

Źródło zmienności Variation source	Suma kwadratów Sum square	Stopnie swobody Degree of freedom	Średnie kwadraty Quadratic mean	Statystyka F F statistic	Wartość-P P - value
Model Model	1,17848E7	8	1,47309E6	98,65	0,0000
Reszty Final segments of the set	8,72038E6	584	14932,2		
Zmienność całkowita Total variation	2,05051E7	592			

### Wariant 2 – dane po usunięciu punktów odstających

W trakcie procedury regresji krokowej usunięte zostały kolejno zmienne: Cd, Fe, Zn i Cr. Otrzymano następujący model:

$$\chi = 5,07404 + 3,34224*AS\_PPM - 11,6068*CO\_PPM + 4,37567*CU\_PPM + 0,0203893*MN\_PPM + 5,24575*NI\_PPM - 0,022788*PB\_PPM \quad (2a)$$

Poniżej dla porównania zaprezentowano model zawierający wszystkie zmienne.

$$\chi = 4,54632 + 3,44542*AS\_PPM - 0,00261307*CD\_PPM - 11,3547*CO\_PPM + 0,431644*CR\_PPM + 4,38287*CU\_PPM - 1,41298*FE\_PROC + 0,0199196*MN\_PPM + 4,9273*NI\_PPM - 0,0205905*PB\_PPM - 0,00207208*ZN\_PPM \quad (2b)$$

<sup>1)</sup> wartość w ppm

<sup>2)</sup> wartość w %

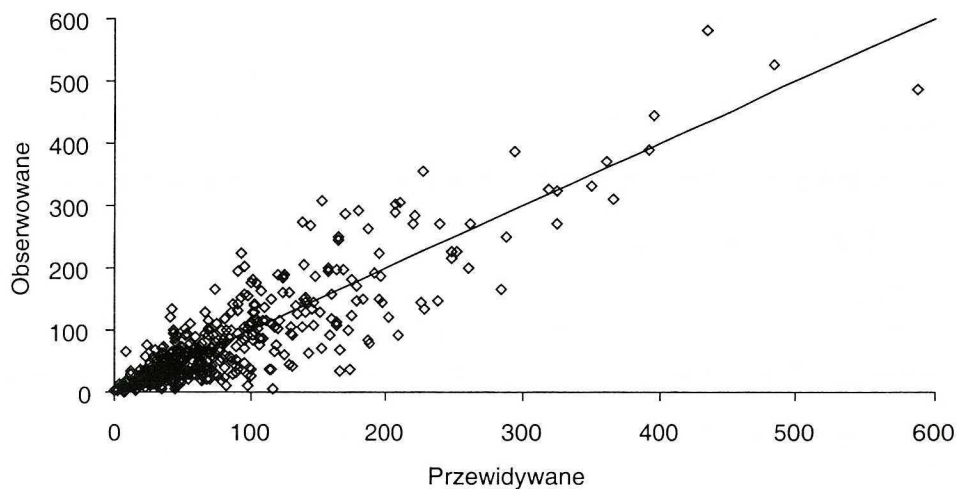
Poniżej na rysunku 3 przedstawiono wykres rozrzutu obserwowanych i przewidywanych wartości  $\chi$  oraz prostą regresji, otrzymaną na podstawie powyższego modelu regresji krokowej wstecznej.

Jak łatwo zauważyć na podstawie powyższego wykresu po wyeliminowaniu punktów odstających uzyskano znacznie lepszy stopień dopasowania. Potwierdziły to wartości otrzymanych parametrów linii regresji:

- współczynnik determinacji –  $R^2 = 80,0774\%$  (80,1297%);
- współczynnik determinacji (poprawiony ze względu na zmianę ilości zmiennych niezależnych) –  $R_c^2 = 79,8471\%$  (79,7439%);
- błąd standardowy estymacji 35,0813 (35,171);
- błąd średni, bezwzględny (średnia wartość reszt) 24,4649 (24,405);
- statystyka Durбина-Watsona = 1,7806 (1,79218).

W tabeli 5 przedstawiono tablicę analizy wariancji otrzymaną w wariancie drugim. Tym razem otrzymano wysoką wartość współczynników determinacji modelu zawierającego jedynie 6 zmiennych: As, Co, Cu, Mn, Ni i Pb, przy pominięciu zmiennych Cd, Fe, Zn, Cr.





Rys. 3. Wykres rozrzutu obserwowanych i przewidywanych wartości podatności magnetycznej oraz prosta regresji otrzymana na podstawie modelu regresji krokowej wstecznej danej równaniem (2a)  
Scatter diagram of observed (vertical axis) and predicted (horizontal axis) values of magnetic susceptibility and regression line obtained on the base of reverse stepwise regression model described by formula (2a)

Tabela 5. Wyniki analizy wariancji wykonanej na danych użytych do przeprowadzenia obliczeń regresji krokowej wstecznej

Results of variance analysis on the base of data after Cd, Fe, Zn and Cr removal

Źródło zmienności Variation source	Suma kwadratów Sum square	Stopnie swobody Degree of freedom	Średnie kwadraty Quadratic mean	Statystyka F F statistic	Wartość-P P - value
Model Model	2,56734E6	6	427890,0	347,68	0,0000
Reszty Final segments of the set	638734,0	519	1230,7		
Zmienność całkowita Total variation	3,20607E6	525			

## DYSKUSJA WYNIKÓW

W zagadnieniach dotyczących współzależności zmiennych losowych np. w badaniach środowiska najczęściej korzysta się ze współczynnika korelacji Pearsona. Podobnie było w dotychczasowych badaniach zależności  $\chi$  od zanieczyszczeń antropogenicznych. Dotychczasowe badania wykazały raczej silną zależność korelacyjną pomiędzy  $\chi$  a zanieczyszczeniem np. Pb, Zn, Cu, Cd [13]. Oczwistym jest, że wielkości współczynników korelacji zależą od wielu różnych czynników i mogą być różne np. w zależności od rodzaju występujących pyłów przemysłowych. Korzystając ze współczynnika korelacji liniowej Pearsona należy jednak pamiętać o jego ograniczeniach. Po pierwsze

powszechnie stosowany estymator współczynnika korelacji otrzymany metodą momentów  $r(x,y) = \text{cov}(x,y)/s_x s_y$  (gdzie,  $\text{cov}(x,y)$  jest kowariancją empiryczną z próby,  $s_x s_y$  jest iloczynem empirycznych zmiennych losowych  $X$  i  $Y$ ) ma rozkład, który jest bardzo słabo zbliżony do rozkładu normalnego  $N(\rho, 1-\rho/n^{1/2})$ . Zatem w celu uzyskania wiarygodnych wyników powinno korzystać się z bardzo licznych prób [2]. Po drugie, że współczynnik korelacji liniowej Pearsona zależy silnie od punktów odstających i jeden lub kilka nietypowych pomiarów może znacznie zmienić jego wartość. Po trzecie w przypadku rozkładów wielowymiarowych współczynnik korelacji nie jest zawsze miarą rzeczywistej współzależności, a jego wartość wynika raczej z zależności pozornych [1]. Po czwarte współczynnik korelacji liniowej Pearsona pozwala wyznaczyć jedynie korelacje liniowe.

Jak łatwo dostrzec w tabeli 2 korelacje liniowe określone współczynnikiem korelacji liniowej Pearsona zawartościami badanych pierwiastków są najczęściej przeciętne ( $0,3 \leq r < 0,5$ ), średnie ( $0,5 \leq r < 0,7$ ) lub wysokie ( $0,7 \leq r < 0,9$ ). Podobny wniosek można wysunąć odnośnie korelacji pomiędzy podatnością magnetyczną a zawartością badanych pierwiastków. Po usunięciu punktów odstających, w przypadku większości, nastąpił wyraźny wzrost współczynników korelacji. Większość obserwowanych korelacji była teraz średnia lub wysoka. Potwierdza to, że przypadkowe wartości odstające mogą poważnie wpłynąć na wyniki obliczeń, dlatego powinno się je raczej usuwać ze zbioru danych, jeśli chcemy scharakteryzować całą próbę. Podobnie wysokie wartości otrzymano obliczając, w analogiczny sposób, współczynniki rang Spearmana (Tab. 3). Są one nieco bardziej odporne na wartości odstające niż w przypadku współczynnika korelacji liniowej Pearsona oraz na brak liniowości współzależności korelacyjnej.

Jak widać wartości współczynnika korelacji rang są generalnie nieco wyższe niż współczynnika korelacji Pearsona. Świadczy to o liniowości badanych korelacji, co potwierdza celowość stosowania współczynników korelacji Spearmana do badania korelacji pomiędzy  $\chi$  a zawartością zanieczyszczeń antropogenicznych.

Analizując regresje krokową wsteczną pominięcie zawartości tak istotnych pierwiastków dla wyjaśnienia zmienności podatności, jak Fe i Zn, wymaga wyjaśnienia. Wiadomo, że podczas wykonywania analizy regresji ważne jest, aby zmienne niezależne były silnie skorelowane ze zmienną zależną, a jednocześnie były bardzo słabo skorelowane między sobą. Idealnie jest wówczas, gdy korelacje pomiędzy zmienną wyjaśnianą a zmiennymi objaśniającymi są znacznie większe niż korelacje pomiędzy zmiennymi objaśniającymi. Aby wykryć istnienie współliniowości pomiędzy zmiennymi przeanalizowano macierz korelacyjną pomiędzy współczynnikami modelu zawierającego wszystkie zmienne. Wysokie moduły wartości elementów tej macierzy informują o występowaniu silnych współliniowości. Na podstawie macierzy korelacyjnej widać, że bardzo duża współliniowość występuje pomiędzy: Cd a Zn (-0,80), Cd i Pb (0,34), Cr a Ni (-0,54), Cr a Co (0,36), Zn a Cd (-0,78), Zn a Pb (-0,63) i Zn a As (-0,42). Silne współliniowości znaleziono pomiędzy zawartością Fe i zawartościami prawie wszystkich metali, a w szczególności pomiędzy Fe a Co (-0,35). Podobne korelacje można dostrzec w tabelach odpowiednich współczynników korelacji (Tab. 2 i 3).

Weześniejsze badania również wykazywały stosunkowo niski stopień korelacji pomiędzy  $\chi$  a zawartością całkowitą Fe [15]. Wynika to z faktu, że tylko niewielka procentowo część żelaza jest ferrimagnetyczna. Pozostała część żelaza występująca w minerałach antyferromagnetycznych lub paramagnetycznych wykazuje znacznie mniejsze korelacje z  $\chi$ .

Przeprowadzona powyżej analiza regresji wielokrotnej potwierdza fakt, że podatność magnetyczna jest współzależna z zanieczyszczeniami antropogenicznymi i zależność ta ma charakter liniowy. Wyjaśnienie zmian tej podatności wymaga uwzględnienia nie jednego, lecz kilku badanych zanieczyszczeń. Ze względu na występowanie silnych korelacji pomiędzy zawartościami badanych pierwiastków w modelu nie należy wykorzystywać wszystkich zmiennych. Metoda regresji krokowej wstecznej może po bardziej szczegółowej adaptacji stanowić wygodną technikę analizy zależności  $\chi$  zanieczyszczeń zanieczyszczeniami antropogenicznymi.

## WNIOSKI

1. Istnieje wyraźna korelacja pomiędzy poziomem zanieczyszczeń antropogenicznych a wielkością mierzonej podatności, chociaż w glebie jest ona na ogół skomplikowana.
2. Badania zależności pomiędzy  $\chi$  a zawartością metali ciężkich nie powinny być oparte wyłącznie o wyznaczanie współczynnika korelacji Pearsona. Zastosowanie kilku komplementarnych metod statystycznych badania współzależności pozwala ocenić je w sposób bardziej prawidłowy.
3. Potwierdzono, że najczęściej występują średnie ( $0,5 \leq r < 0,7$ ) i wysokie ( $0,7 \leq r < 0,9$ ) korelacje pomiędzy wielkością  $\chi$  a zawartościami badanych metali.
4. Porównywalna wartości współczynnika korelacji liniowej Pearsona oraz współczynnika korelacji rang Spermmana wskazuje, że w zakresie stosowanej dokładności związki korelacyjne są liniowe. Podobnych wniosków dostarczają wyniki regresji krokowej wstecznej – otrzymano liniowy model regresji wielokrotnej wyjaśniający prawie 80% zmienności podatności magnetycznej.
5. Badania podatności magnetycznej stanowią bardzo interesującą alternatywę dla innych metod w zakresie monitoringu zanieczyszczeń antropogenicznych gleby, a w szczególności zanieczyszczeń metalami ciężkimi.

## LITERATURA

- [1] Dobosz M.: *Wspomagana komputerowo statystyczna analiza wyników badań*, Problemy współczesnej nauki. Teoria i zastosowania, Akademicka Oficyna Wydawnicza EXIT, 2001.
- [2] Greń J.: *Statystyka matematyczna. Modele i zadania*, PWN, Warszawa 1976.
- [3] Hay K.L., J.A. Dearing, S.M.J. Baban, P.J. Loveland: *A preliminary attempt to identify atmospherically-derived pollution particles in English topsoils from magnetic susceptibility measurement*, Phys. Chem. Earth, **22**, 207–210 (1997).
- [4] Lis J., A. Pasieczna: *Atlas geochemiczny Polski*, Państwowy Instytut Geologiczny, Warszawa 1995.
- [5] Magiera T., Z. Strzyszczyk: *Monitoring magnetyczny na terenach objętych imisjami przemysłowymi*, [w:] Monitoring Środowiska, II Forum Inżynierii Ekologicznej – Nałęczów 1998.
- [6] Magiera T., Z. Strzyszczyk: *Ferromagnetic minerals of anthropogenic origin in soils of some Polish national parks*, Water Air and Soil Poll., **124**, 37–48 (2000).
- [7] Magiera T., J. Lis, J. Nawrocki, Z. Strzyszczyk: *Podatność magnetyczna gleb Polski*, Państwowy Instytut Geologiczny, Warszawa 2002.
- [8] Rozporządzenie Ministra Środowiska z dnia 9 września 2002 w sprawie standardów jakości gleby oraz standardów jakości ziemi, Dz. U. 02. 165.1359 z dnia 4 października 2002.
- [9] *Statgraphics Plus. Podręcznik Użytkownika*, Manguistics, Inc. Kraków 1995.
- [10] Strzyszczyk Z.: *Ferromagnetic properties of forest soils being under influence of industrial pollution. Air pollution and forest decline*, [w:] Proceedings 14<sup>th</sup> International Meeting for Specialist in Air Pollution, Effects on Forest Ecosystems, IUFRO, Interlaken 1989, pp 201–207.
- [11] Strzyszczyk Z.: *Ferromagnetism of soil in some Polish National Parks*, Mitt. Deut. Boden. Ges., **66**,

- 1119–1122 (1991).
- [12] Strzyszczyński Z.: *Magnetic susceptibility of soils in the area influenced by industrial emissions*, [w:] Soil Monitoring, Monte Verita, Birkhäuser Verlag, Basel 1993, 255–269.
- [13] Strzyszczyński Z., T. Magiera: *Heavy metal contamination and magnetic susceptibility in soils of southern Poland*, Phys. Chem. Earth., **23**, 1127–1131 (1998).
- [14] Strzyszczyński Z., T. Magiera: *Podatność magnetyczna niektórych przemysłowych pyłów atmosferycznych i jej znaczenie w monitoringu ekologicznym powierzchni ziemi*, Prace i Studia Instytutu Podstaw Inżynierii Środowiska PAN, **54**, 205–215 (2000).
- [15] Strzyszczyński Z., T. Magiera: *Pomiary podatności magnetycznej torfowisk wysokich w monitoringu emisji przemysłowych*, Archiwum Ochrony Środowiska, **27** (2) 141–167 (2001).

Wpłynęło: 3 września 2003, zaakceptowano do druku: 20 stycznia 2004.