

Karolina Pirowska  
Instytut Nafty i Gazu, Kraków

## Zastosowanie algorytmu genetycznego do estymacji parametrów ośrodka geologicznego na podstawie pomiarów sejsmicznych

### Wstęp

Określenie pola prędkości jest kluczowym zadaniem w procesie rozpoznawania budowy ośrodka geologicznego. Model prędkości stanowi najważniejszą informację dla geofizyka i geologa, ponieważ w dalszym procesie przetwarzania danych geofizycznych jest on wyjściowy dla wyznaczenia własności geometrycznych, zmienności litologicznej oraz parametrów petrofizycznych skał. W efekcie możliwe jest pełniejsze poznanie ośrodka i bardziej dokładne zlokalizowanie złoża węglowodorów.

Analizą prędkości od lat zajmowało się i nadal zajmuje wielu autorów na całym świecie. Jednak optymalne rozwiązanie zadania wyznaczenia modelu prędkości wciąż pozostaje problemem otwartym. Ciągłe ulepszanie dotychczasowych metod czy tworzenie zupełnie nowych jest nieuniknione w związku z rozwojem technik pomiarowych i nowych urządzeń – zarówno w badaniach sejsmicznych, jak również w innych badaniach geofizycznych. Współcześnie znacznie wzrosła liczba, a także jakość wykonywanych pomiarów. Nowe możliwości otwierają się wraz ze wzrostem szybkości i mocy obliczeniowej komputerów, dzięki czemu realne staje się przetwarzanie dużej ilości danych. Tradycyjne rozwiązania okazują się być niewystarczające, gdyż wymagają wielu uproszczeń, a otrzymane za ich pomocą modele są mało dokładne. Coraz częściej stosuje się metody oparte na technikach ewolucyjnych, sieciach neuronowych czy logice rozmytej, które umożliwiają pełniejsze wykorzystanie dostępnych informacji w analizie i interpretacji danych [6].

Rosnącą rolę zastosowań tego typu metod uzmysławia między innymi artykuł O. Veleza-Langsa [10], który

dokonał przeglądu aplikacji algorytmów genetycznych (GA – *Genetic Algorithm*) zastosowanych do rozwiązania problemów w przemyśle naftowym. Algorytm genetyczny okazuje się efektywną i stosunkowo prostą w użyciu metodą, zwłaszcza gdy w danym problemie mamy do czynienia z wieloma parametrami i gdy istnieje wiele jednakowo dobrych rozwiązań. Praktyczne prace i zastosowania dotyczą tematyki związanej z charakterystyką skał zbiornikowych [7], podziemnym przechowywaniem gazu, zagadnieniami inwersji [1, 8, 9], wspomaganiami prac na polach roponośnych, przetwarzaniem ropy naftowej czy strategiami i monitorowaniem wydobywania.

Również w Polsce obserwuje się zainteresowanie technikami ewolucyjnymi w naukach geologicznych i geofizycznych. A. Leśniak i G. Pszczoła zastosowali algorytm genetyczny w dwóch zadaniach odwrotnych [5]. Pierwszym z nich było określenie położenia źródeł wstrząsów sejsmicznych, gdzie do nieznanych parametrów należały trzy współrzędne przestrzenne tego źródła oraz czas wystąpienia wstrząsu, drugie zadanie natomiast dotyczyło danych sondowania elektrooporowego i określenia parametrów ośrodka – miąższości i oporności poszczególnych warstw.

W niniejszym opracowaniu przedstawiono wyniki zastosowania algorytmu genetycznego do estymacji parametrów ośrodka geologicznego. Obliczenia przeprowadzono dla dwóch modeli syntetycznych o różnej geometrii. Aplikacja wymagała wielu testów wstępnych, koniecznych do rozpoznania sposobu działania algorytmu i dopasowania metody do konkretnego problemu.

## Sformułowanie problemu

Celem było zastosowanie algorytmu genetycznego do estymacji pola prędkości propagacji fali podłużnej oraz wyznaczenia geometrii ośrodka na podstawie danych sejsmiki refleksyjnej. Założono, że ośrodek geologiczny jest izotropowy, a prędkość jest zdefiniowana funkcją liniową. Zbiorem poszukiwanych parametrów były: prędkość początkowa, pionowy i poziomy gradient prędkości oraz głębokości poszczególnych granic odbi-

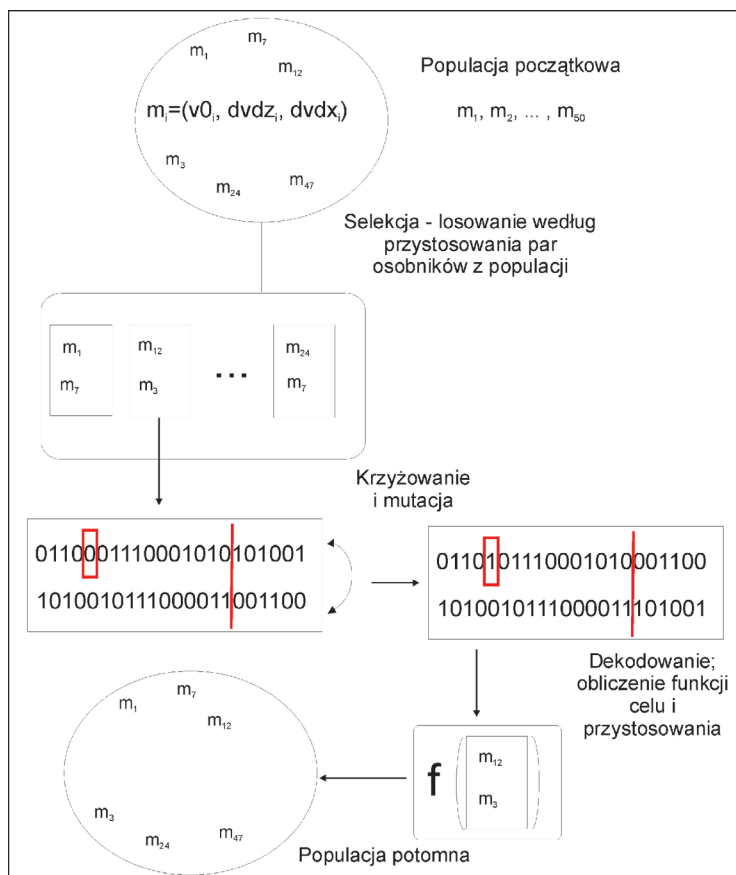
jających. Danymi były odpowiadające poszukiwanym granicom hodografy, czyli czasy przyścia fali z określonego punktu wzbudzenia do kolejnych odbiorników. Algorytm genetyczny został wykorzystany jako metoda optymalizacyjna w zadaniu wyznaczenia takich wartości poszczególnych parametrów, dla których wymodelowane hodografy syntetyczne najlepiej przybliżają hodografy wyznaczone z danych.

## Aplikacja algorytmu genetycznego

### Charakterystyka metody

Algorytm genetyczny (GA) próbuje znaleźć optymalne rozwiązanie w sposób, który naśladuje ewolucję biologiczną. Tak jak w naturze obserwować można ciągłą walkę gatunków o przetrwanie, a żyjące gatunki biologiczne reprezentują najsilniejsze, optymalne rozwiązanie problemu przetrwania we wrogim środowisku, tak w algorytmie genetycznym otrzymane pokolenie wyjściowe reprezentuje zbiór najlepszych rozwiązań danego problemu optymalizacyjnego. Działanie algorytmu genetycznego

przebiega w analogii do zaczerpniętych z medycyny praw genetyki. Po wystartowaniu z losowo przyjętej populacji początkowej, budowane jest pokolenie potomne. Populacja i kolejne pokolenia reprezentują zbiory rozwiązań danego problemu optymalizacyjnego. Do następnego pokolenia mogą przejść jedynie najlepiej przystosowane osobniki, czyli takie, dla których zdefiniowana dla problemu funkcja celu przyjmuje wartości optymalne. Rozwiązania w kolejnych iteracjach (to znaczy pokolenia potomne) wyznacza się za pomocą trzech operatorów: selekcji, krzyżowania i mutacji. Selekcja (reprodukcja) polega na losowym wyborze osobników z populacji, w procesie krzyżowania (rekombinacji) następuje zamiana części zakodowanych osobników w losowo wybranym punkcie, natomiast mutacja jest operatorem wprowadzającym różnicowanie w populacji. Na rysunku 1 schematycznie zobrazowano działanie opisanych operatorów. Teorię dotyczącą algorytmów genetycznych szczegółowo opisuje D. E. Goldberg [3].



Rys. 1. Działanie algorytmu genetycznego – generacja nowego pokolenia za pomocą operatorów genetycznych

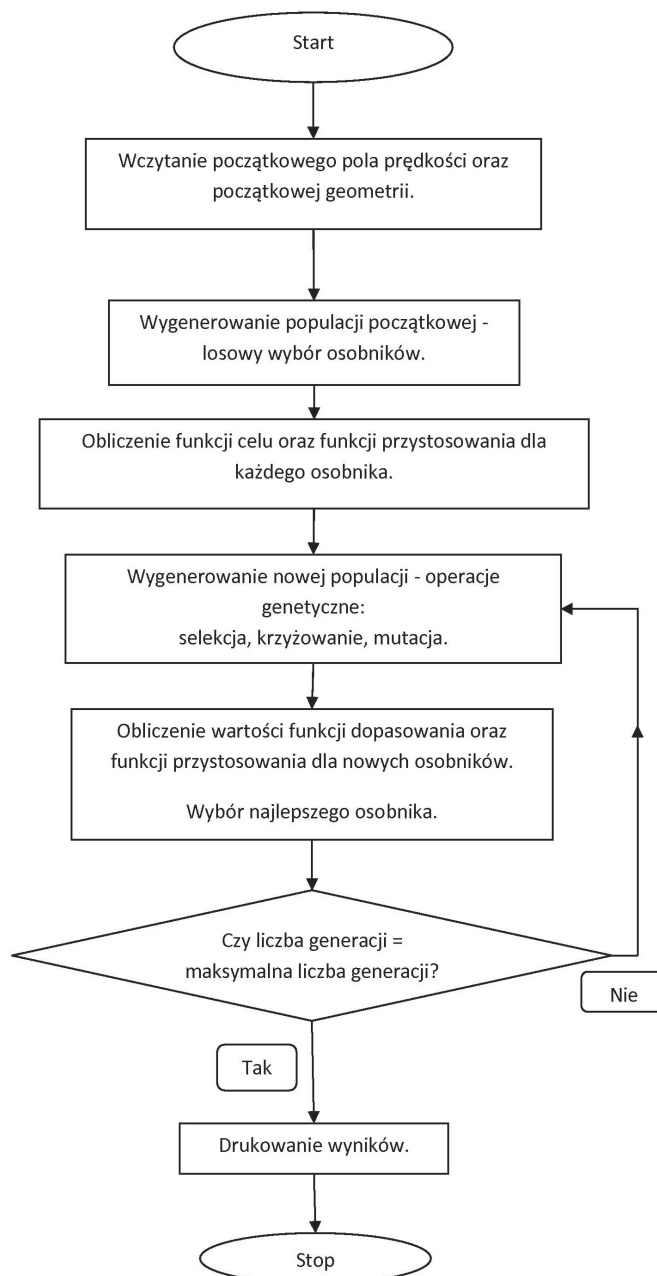
### Dopasowanie metody do konkretnego zadania

Algorytm genetyczny jest stochastyczną metodą optymalizacji globalnej, użyteczną przede wszystkim w skomplikowanych problemach nieliniowych oraz gdy poszukiwany model składa się z bardzo dużej liczby parametrów. Ogromną zaletą metody jest bowiem to, że nie wykorzystuje ona informacji o gradiencie funkcji celu oraz nie wymaga obliczeń na macierzach. Cechuje się również niezależnością od modelu początkowego – we wstępnej fazie działania algorytmu modele początkowe generowane są losowo. Jednak główną trudność stanowi dopasowanie algorytmu do konkretnego zadania. Przede wszystkim jest to wybór sposobu zakodowania modelu w ciąg binarny,

a także ustalenie sposobu oraz prawdopodobieństw wykonywania poszczególnych operacji modyfikujących model (prawdopodobieństwo selekcji, krzyżowania, mutacji). Aby algorytm genetyczny był optymalizacją użyteczną i efektywną, należy prawidłowo zdefiniować funkcję przystosowania, dobrać wielkość populacji, liczbę pokoleń oraz określić kryteria zatrzymania algorytmu. Dopasowanie parametrów determinujących działanie algorytmu genetycznego może być dosyć kłopotliwe i odbywa się praktycznie jedynie metodą prób i błędów. Konieczne staje się przeprowadzenie wielu wstępnych testów i analiz.

W przypadku postawionego zadania bazujący na algorytmie genetycznym program wyznacza wartości prędkości propagacji fali podłużnej oraz geometrię modelu (układ granic refleksyjnych), dla których wymodelowane hodografy syntetyczne najlepiej przybliżają hodografy pomierzone. Założono, że prędkość jest zadana funkcją liniową daną wzorem  $v(x, z) = v_0 + x \times dvdx + z \times dvdz$ , gdzie  $v_0$  oznacza prędkość początkową, natomiast  $dvdz$  i  $dvdz$  to odpowiednio pionowy i poziomy gradient prędkości. Zatem poszukiwane pole prędkości opisuje wektor złożony z trzech parametrów: prędkości początkowej oraz pionowego i poziomego gradientu prędkości (wektor  $m_i$  na rysunku 1). W pierwszej fazie działania programu następuje wczytanie danych. Docelowe dane to wypunktowane na rekordach połowych hodografy, jednak najpierw należało przeprowadzić obliczenia dla danych syntetycznych. Dla zaprezentowanych dalej przykładowych modeli użyto hodografów wymodelowanych za pomocą programu obliczającego seismogramy syntetyczne (opisanego w dalszej części artykułu). Danymi wejściowymi są również parametry początkowe modelu: prędkość początkowa, pionowy i poziomy gradient prędkości, a także głębokości poszczególnych granic. Następnie wygenerowana zostaje populacja początkowa, czyli zbiór osobników, wektorów ( $v_0$ ,  $dvdz$ ,  $dvdz$ ), reprezentujących pole prędkości. Dla każdego osobnika zostaje obliczona funkcja celu oraz funkcja przystosowania. Funkcja celu została zdefiniowana jako błąd średniokwadratowy pomiędzy hodografami pomierzonymi a hodografami wymodelowanymi dla konkretnego pola prędkości (osobnika). Funkcja przystosowania przekształca wartości funkcji celu w taki sposób, aby najlepiej przystosowanymi osobnikami były te, dla których obliczone hodografy najlepiej przybliżają hodografy pomierzone. Kolejnym etapem algorytmu jest wygenerowanie za pomocą operatorów genetycznych nowego pokolenia (rysunek 1), aktualizacja parametrów związanych z geometrią modelu, ponowne obliczenie funkcji celu i przystosowania oraz wybór najlepszego osobnika.

Algorytm zatrzymuje się, gdy spełnione zostaje kryterium zakończenia algorytmu, czyli osiągnięcie zdefiniowanej, maksymalnej liczby iteracji. Ogólny schemat działania programu optymalizacyjnego przedstawiono na rysunku 2.



Rys. 2. Schemat blokowy algorytmu obliczania parametrów ośrodka geologicznego

Posłużono się programem do modelowania *susynlv*, należącym do ogólnodostępnego oprogramowania geofizycznego Seismic Unix. Pakiet SU został stworzony i jest rozwijany w Center for Wave Phenomena w Department of Geophysical Engineering, w Colorado School of Mines [2]. Po zaakceptowaniu warunków licencji dotyczącej danej wersji pakietu można bez ograniczeń korzystać z kodów

programów dla celów przetwarzania danych lub tworzenia oprogramowania. Program *susynlv* służy do generowania sejsmogramów syntetycznych dla ośrodka izotropowego z prędkością zadaną funkcją liniową.

### Problemy aplikacyjne – ustalenie parametrów charakteryzujących algorytm genetyczny

Jak wspomniano wcześniej, aplikacja metody wymagała przetestowania działania algorytmu dla różnie przyjętych

parametrów odpowiedzialnych za efektywność i jakość otrzymywanych wyników. W kolejnych próbach zmieniano długość ciągu binarnego kodującego model, liczbę osobników w populacji, liczbę pokoleń, przyjmowano różne wartości prawdopodobieństwa krzyżowania i mutacji. Kolejną kwestią było dopasowanie funkcji przystosowania oraz obserwacja jakości wyników w stosunku do wielkości przedziałów wartości poszukiwanych w zadaniu parametrów.

### Przykłady i omówienie wyników

Program przetestowano dla dwóch modeli syntetycznych:

- 1) Model o trzech granicach: pierwsza granica była płaskorównoległa, o głębokości 200 m, druga i trzecia były nachylone odpowiednio pod kątem  $10^\circ$  i  $15^\circ$ , a ich początkowe głębokości dla modelu wynosiły 500 m i 800 m. Prędkość została zadana funkcją liniową o wartości początkowej równej 2200 m/s i gradientach: pionowym równym 0,2 i poziomym równym  $-0,1$ . Geometria modelu została przedstawiona na rysunku 3, natomiast pole prędkości na rysunku 4a,
- 2) Model o trzech granicach: pierwsza i druga granica miały kształt antykliny, trzecia granica była płaskorównoległa, o głębokości 900 m. Prędkość została zadana funkcją liniową o wartości początkowej równej 2200 m/s i gradientach: pionowym równym 0,32 i poziomym równym  $-0,1$ . Geometria modelu została przedstawiona na rysunku 6, a pole prędkości na rysunku 7a.

Za pomocą programu *susynlv* dla obu modeli obliczono hodografy dla kolejnych granic, dla 53 punktów wzbudzenia. Przyjęto rozstaw prawostronny z 45 odbiornikami rozstawionymi co 50 m. Wybrane wymodelowane hodografy potraktowano następnie jako dane pomiarowe, na podstawie których należało oszacować parametry ośrodka. W przypadku modelu 1 obliczenia prowadzono dla 5 rekordów, dla których punkty wzbudzenia umieszczone były w  $x = 0, 500, 1000, 1500, 2000$  m, natomiast dla modelu 2 punkty wzbudzenia umieszczone były w  $x = 600, 800, 1000, 1200, 1400$  m.

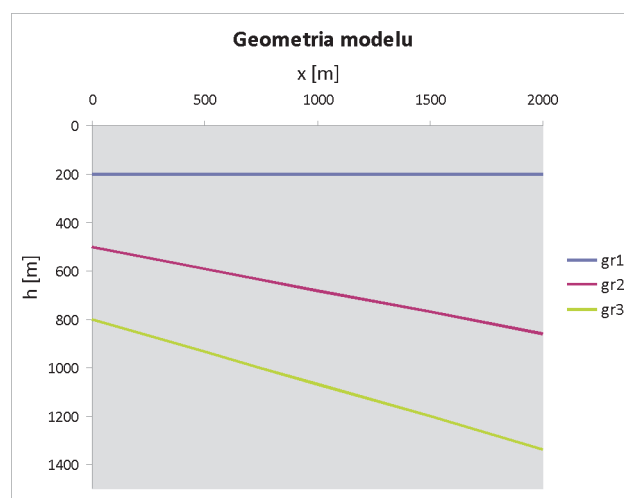
#### Model syntetyczny 1

Tablica 1. Parametry prędkości dla modelu syntetycznego 1

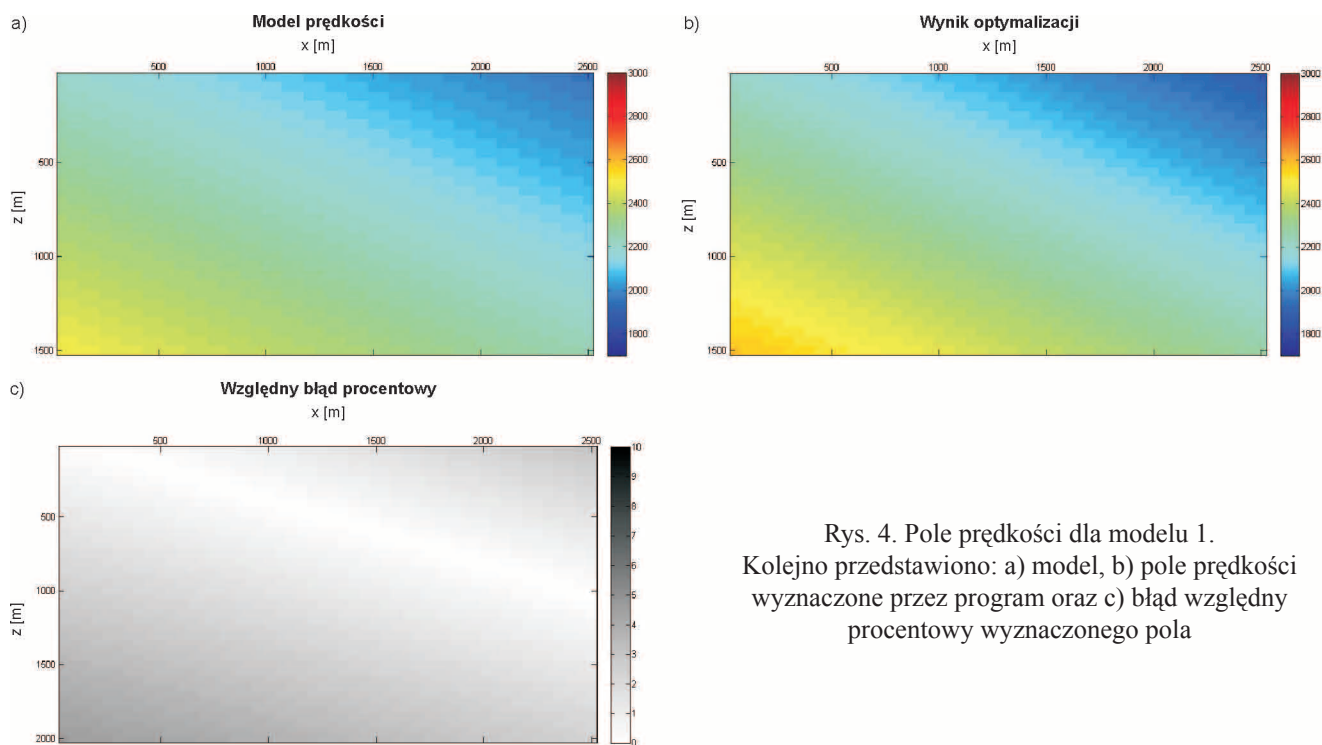
	$v_0$ [m/s]	$dvdz$	$dvdz$
Szukane	2200	0,2	$-0,1$
Startowe	2000	0	0
Optymalne	2204	0,2583	$-0,1252$

Tablica 2. Parametry głębokości dla modelu syntetycznego 1

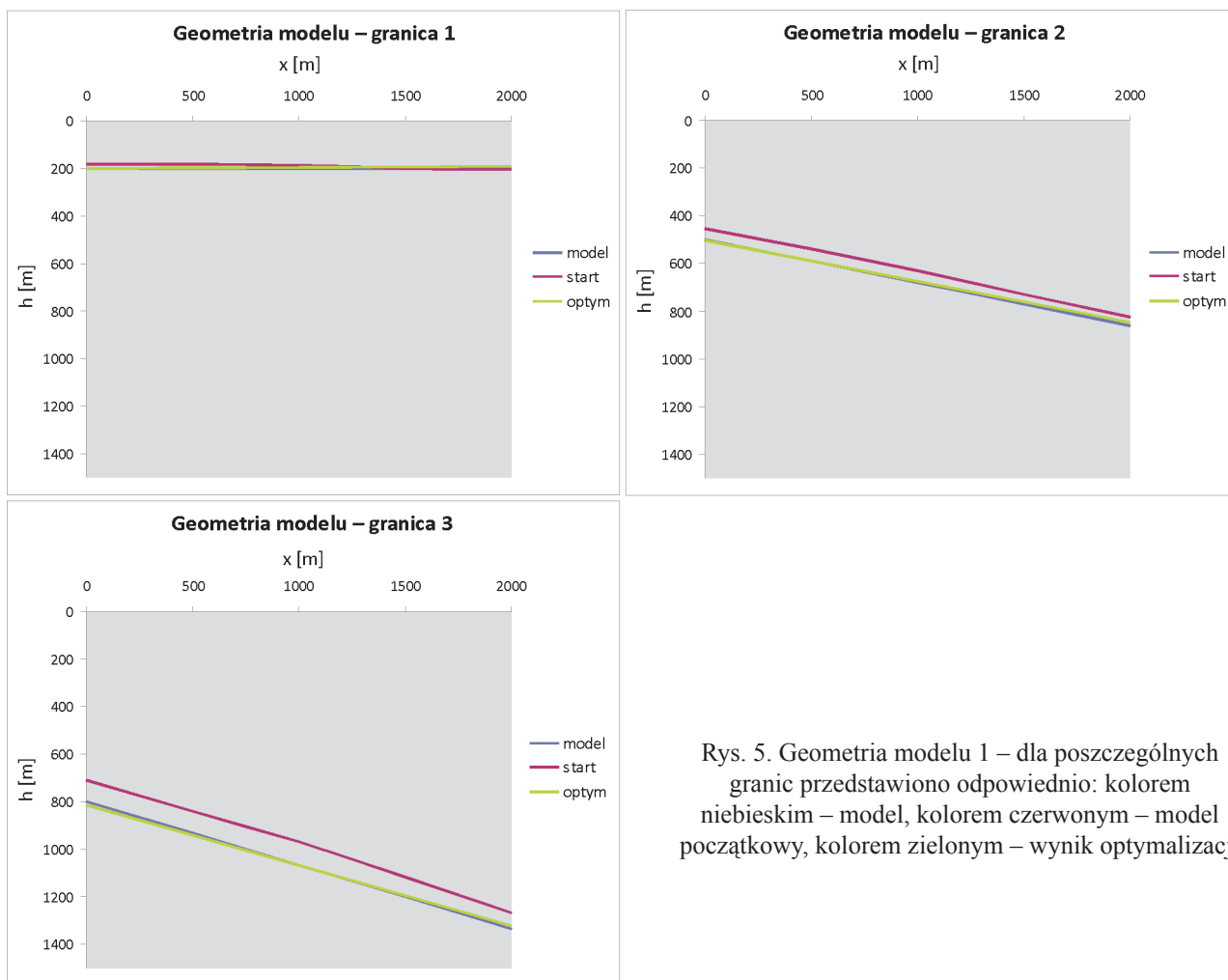
Granica	$x$	Szukane	Startowe	Optymalne
1	0	200	180	200
	500	200	180	200
	1000	200	186	198
	1500	200	200	195
	2000	200	205	192
2	0	500	454	502
	500	590	540	589
	1000	680	630	676
	1500	770	730	762
	2000	860	823	849
3	0	800	710	813
	500	934	840	941
	1000	1068	970	1069
	1500	1202	1117	1196
	2000	1336	1268	1324



Rys. 3. Geometria modelu 1



Rys. 4. Pole prędkości dla modelu 1. Kolejno przedstawiono: a) model, b) pole prędkości wyznaczone przez program oraz c) błąd względny procentowy wyznaczonego pola

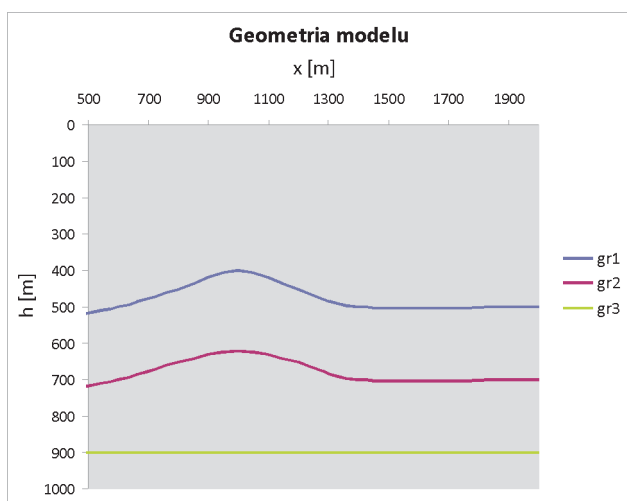


Rys. 5. Geometria modelu 1 – dla poszczególnych granic przedstawiono odpowiednio: kolorem niebieskim – model, kolorem czerwonym – model początkowy, kolorem zielonym – wynik optymalizacji

## Model syntetyczny 2

Tablica 3. Parametry prędkości dla modelu syntetycznego 2

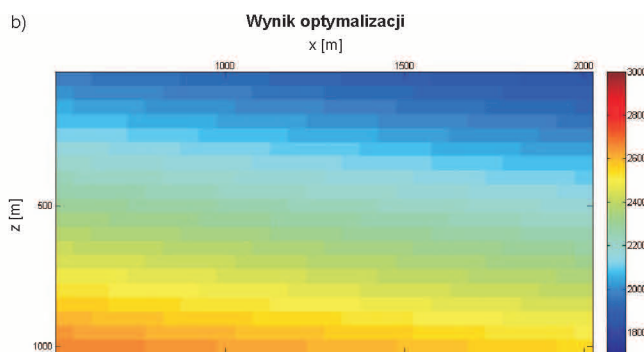
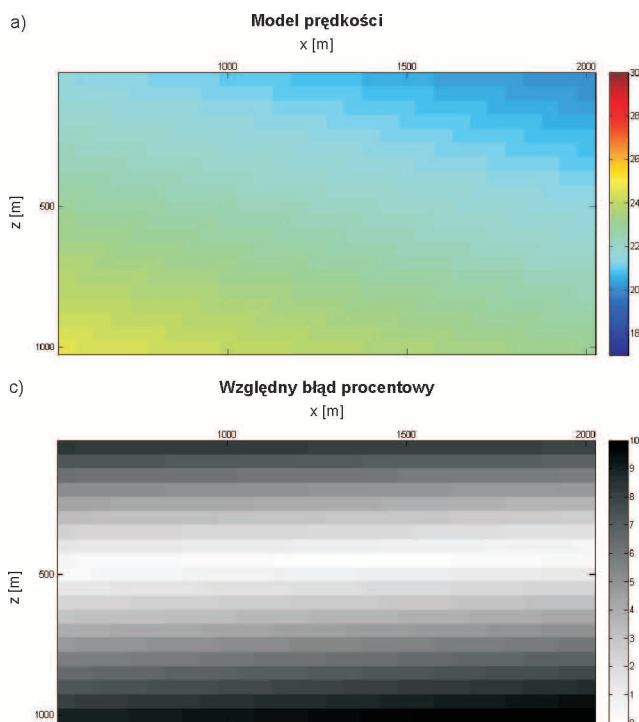
	$v_0$ [m/s]	$dv/dz$	$dv/dx$
Szukane	2200	0,32	-0,1
Startowe	2000	0	0
Optymalne	2015	0,7378	-0,0867



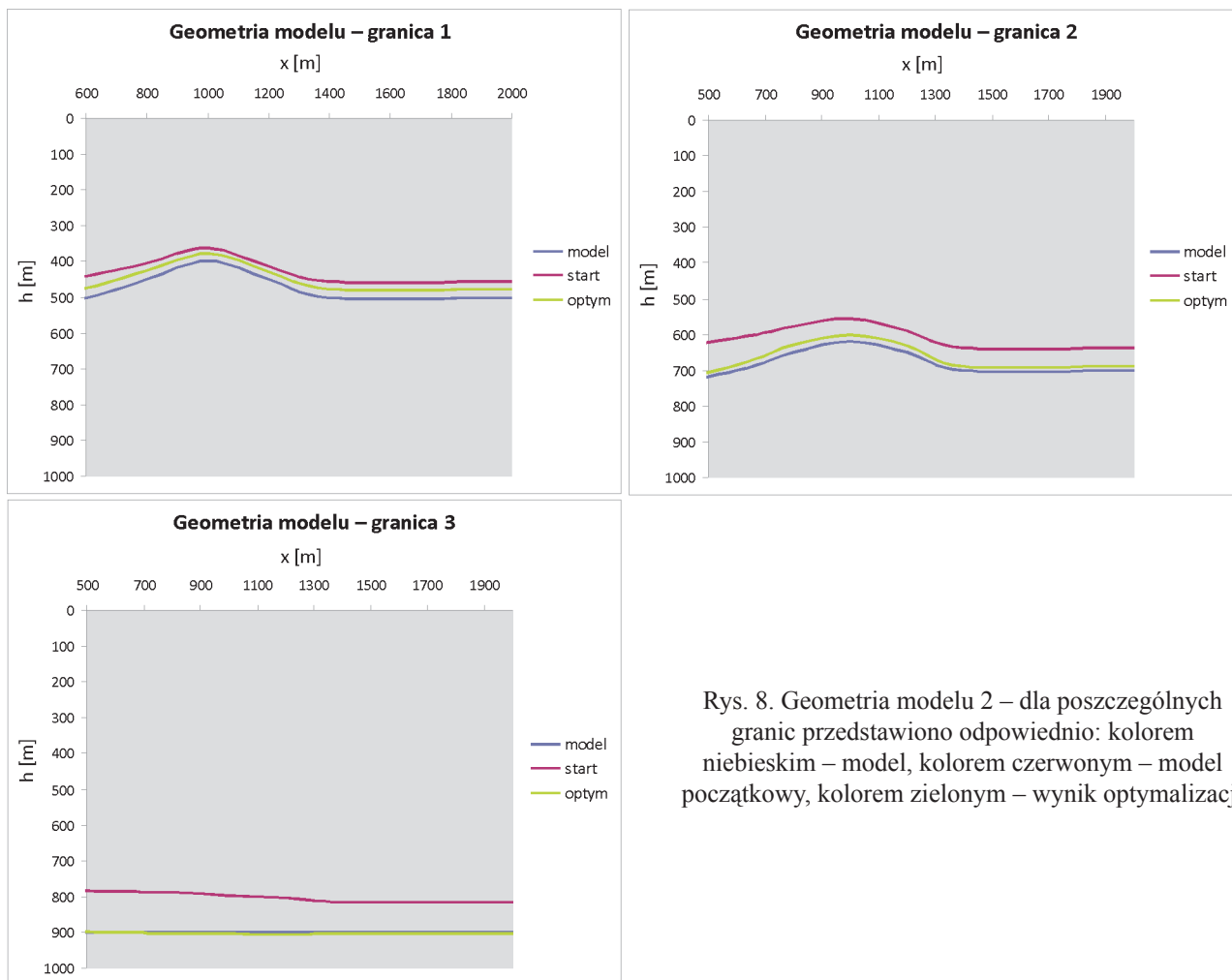
Rys. 6. Geometria modelu 2

Tablica 4. Parametry głębokości dla modelu syntetycznego 2

Granica	$x$	Szukane	Startowe	Optymalne
1	0	600	530	570
	600	500	442	480
	800	450	405	440
	1000	400	364	380
	1200	450	413	430
	1400	500	457	480
2	0	800	680	780
	600	700	610	680
	800	650	577	630
	1000	620	554	620
	1200	650	590	650
	1400	700	637	700
3	0	900	768	780
	600	900	785	900
	800	900	788	900
	1000	900	798	900
	1200	900	804	990
	1400	900	815	910



Rys. 7. Pole prędkości dla modelu 1. Kolejno przedstawiono: a) model, b) pole prędkości wyznaczone przez program oraz c) błąd względny procentowy wyznaczonego pola

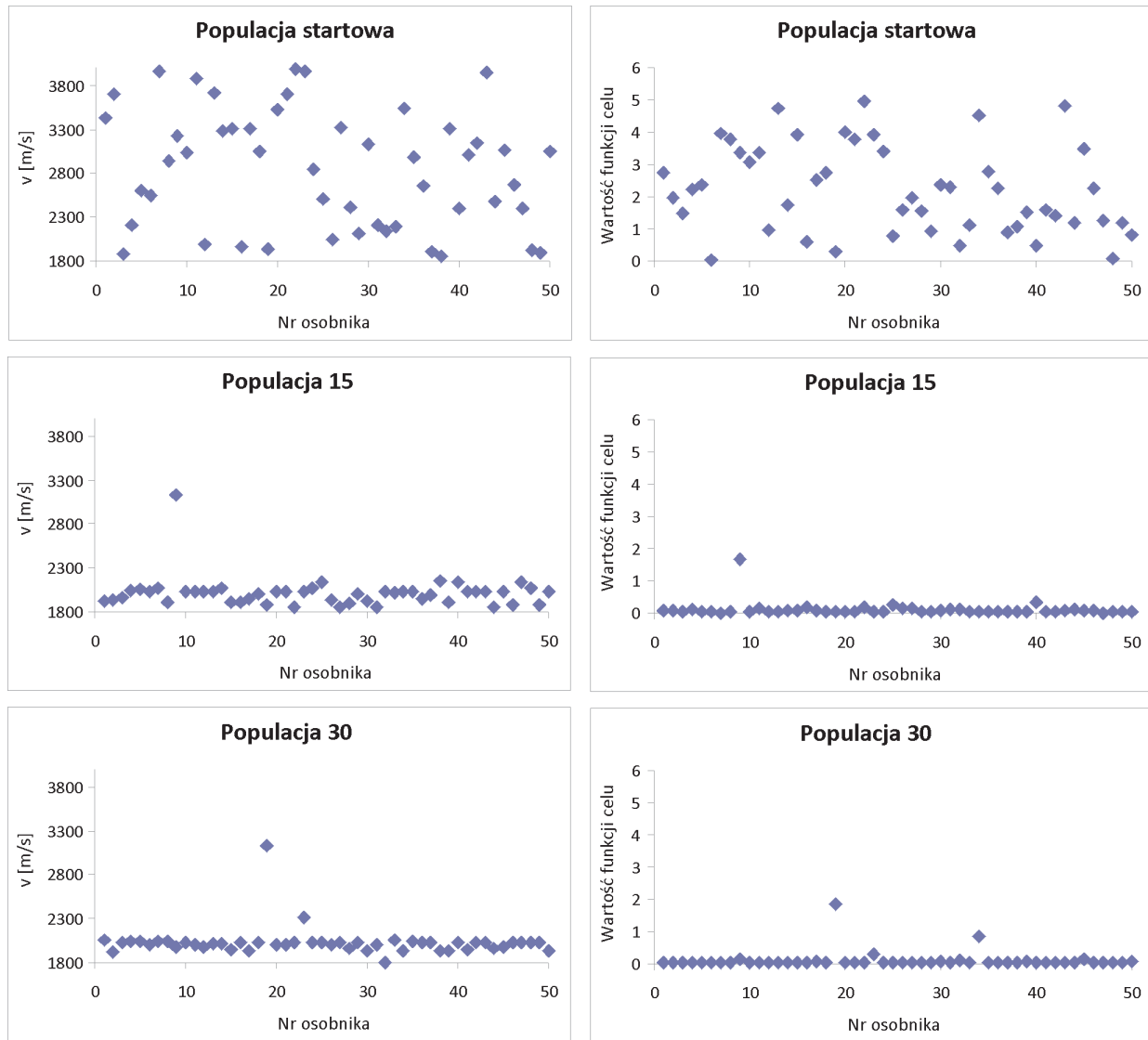


Rys. 8. Geometria modelu 2 – dla poszczególnych granic przedstawiono odpowiednio: kolorem niebieskim – model, kolorem czerwonym – model początkowy, kolorem zielonym – wynik optymalizacji

Analizując wyniki dla modelu pierwszego, można stwierdzić bardzo dobre przybliżenie zarówno pola prędkości, jak i głębokości poszczególnych granic refleksyjnych. Dokładne wartości parametrów dotyczących prędkości poszukiwanej, startowej i otrzymanej w wyniku działania programu umieszczono w tabelicy 1, natomiast pole prędkości optymalnej zobrazowano na rysunku 4b. Na rysunku 4c widać, że względny błąd procentowy jest największy w lewej dolnej części, co wynika z tego, że oszacowany przez program gradient pionowy jest wyższy od poszukiwanego. Rezultat taki może być związany z charakterem zalegania granic (ta część pola prędkości nie ma wpływu na kształt hodografu). Patrząc na tabelicę 2 i rysunek 5, można porównać parametry reprezentujące geometrię modelu. Oszacowane głębokości różnią się od poszukiwanych maksymalnie do 8 m dla granicy pierwszej i drugiej oraz do 13 m dla granicy trzeciej (na rysunku 5 można zaobserwować, że zaznaczone kolorem zielonym granice, będące wynikiem optymalizacji, pokrywają się z przedstawionymi na niebiesko granicami modelu).

Dla modelu drugiego przybliżenie parametrów dotyczą-

cych prędkości nie jest aż tak dobre jak we wcześniejszym modelu. Wartości parametrów dotyczących prędkości poszukiwanej, startowej i otrzymanej w wyniku działania programu umieszczono w tabelicy 3, natomiast pola prędkości optymalnej przedstawiono na rysunku 7b. Względny błąd procentowy oszacowania pola prędkości (rysunek 7c) jest najmniejszy w środkowej części modelu, a wzrasta w górnym i dolnym obszarze. Jest to związane z tym, że w wyniku działania programu wartość początkowa prędkości  $v_0$  została oszacowana z niedomiarem (2015 m/s zamiast 2200 m/s), natomiast otrzymany pionowy gradient prędkości został zawyżony (prawie 0,74 zamiast 0,32). W rezultacie w górnej części modelu prędkość jest za niska, w środkowej części osiąga wartości poszukiwane, a w dolnej części jest wyższa od poszukiwanej. Patrząc na tabelicę 4 i rysunek 8, można porównać parametry reprezentujące geometrię modelu. Oszacowanie głębokości jest nieco gorsze niż dla modelu pierwszego, ale zadowalające. Błąd bezwzględny oszacowania wynosi maksymalnie do 30 m dla granicy pierwszej, do 20 m dla granicy drugiej oraz do 120 m w przypadku jednego rekordu dla granicy trzeciej (na

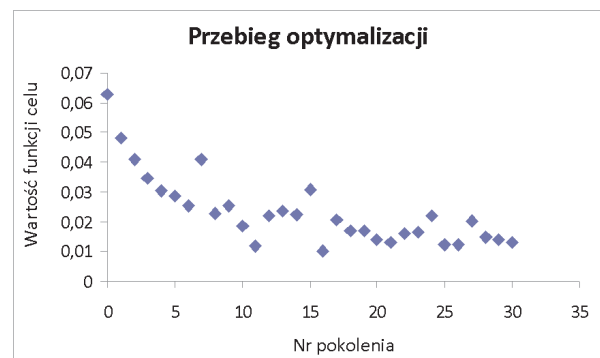


Rys. 9. Działanie algorytmu genetycznego – wybrane populacje

rysunku 8 zaznaczone kolorem zielonym granice, będące wynikiem optymalizacji, leżą blisko przedstawionych na niebiesko granic modelu).

Działanie algorytmu na przykładzie jednego z poszukiwanych parametrów dla modelu 1 przedstawia rysunek 9. W przypadku prędkości początkowej  $v_0$  populacja startowa składająca się z 50 osobników, czyli zbiór rozwiązań, jest losowana z przyjętego przedziału dopuszczalnych wartości [1800; 4000] m/s. W trakcie kolejnych iteracji, w wyniku działania operatorów genetycznych, zbiór rozwiązań skupia się wokół jednej (może się zdarzyć, że wokół kilku) wartości – jest to poszukiwana wartość 2200 m/s. Dla pokolenia początkowego funkcje celu dla poszczególnych osobników osiągają wartości do 5,1; dla pokolenia ostatniego – do 1,4. Można zaobserwować szybki spadek wartości funkcji celu w początkowej fazie działania algorytmu i stabilizację w przedziale [0,012; 0,016] dla

10 ostatnich pokoleń. Przebieg optymalizacji obrazuje rysunek 10. Na wykresie przedstawiono wartości funkcji celu dla najlepszych osobników z kolejnych 30 pokoleń (populacji potomnych).



Rys. 10. Wykres wartości funkcji celu dla najlepszego osobnika w kolejnych pokoleniach



### Obserwacje i wnioski

Przeprowadzone testy dla modeli syntetycznych są satysfakcjonujące. Otrzymano bardzo dobre wyniki dla modelu o granicach nachylonych (względny błąd procentowy oszacowania pola prędkości nie przekracza 5%) i nieco gorsze dla modelu z antykliną (względny błąd procentowy oszacowania pola prędkości nie przekracza 11%). Zauważono, że najgorzej szacowanym parametrem jest pionowy gradient prędkości  $dv/dz$ . Być może wynika to z dosyć małej głębokości reflektorów w stosunku do całego modelu oraz ograniczeń metody modelowania.

Przedstawione wyniki są obiecujące i stanowią zachętę do dalszego rozwijania metody oraz badania kryteriów jej stosowalności. Przykłady uwidoczniły jedną z podstawowych zalet zastosowania algorytmu genetycznego, jaką jest możliwość zadania dużych przedziałów wartości poszukiwanych parametrów i niezależność od modelu początkowego. Głównym zadaniem pozwalającym w peł-

ni ocenić użyteczność metody będzie przeprowadzenie w przyszłości testów dla danych rzeczywistych. Cenne byłoby również rozważenie zamiany programu do modelowania na taki, który pozwoliłby uniknąć założenia, że prędkość jest zdefiniowana funkcją liniową. W przypadku dostępu do programu modelującego sejsmogramy syntetyczne dla dowolnego pola prędkości, modyfikacja programu optymalizacyjnego tak, aby działał dla nowego modelowania, nie powinna stanowić problemu.

Ustalenie w pełni optymalnych parametrów charakteryzujących działanie algorytmu genetycznego jest trudne. Przeprowadzając różne testy i śledząc wyniki, również wyniki pośrednie, można jednak wskazać najbardziej efektywne dla danego problemu. Obecny etap badań pozwala uznać, że algorytm genetyczny jest metodą użyteczną w rozwiązaniu rozważanego zadania szacowania parametrów ośrodka geologicznego.

*Praca naukowa finansowana ze środków na naukę w latach 2010–2012 jako projekt badawczy własny, numer NN525 349038.*

### Literatura

- [1] Boschetti F., Dentithz M. C., List R. D.: *Inversion of seismic refraction data using genetic algorithms*. „Geophysics” 1996, vol. 61, s. 1715–1727.
- [2] Cohen J. K., Stockwell Jr. J. W.: *CWP/SU: Seismic Unix Release No. 38: a free postage for seismic search and processing*. Center for Wave Phenomena. Colorado School of Mines.
- [3] Goldberg D. E.: *Algorytmy genetyczne i ich zastosowania*. WNT, 1998.
- [4] Jędrzejowska-Tyczkowska H., Pieniążek K.: *Zastosowanie metod optymalizacji globalnej, a szczególnie symulowanego wyżarzania (simulated annealing), do rozwiązania odwrotnego zadania kinematycznego w metodzie sejsmicznej*. „Nafta-Gaz” 2007, nr 12.
- [5] Leśniak A., Pszczoła G.: *Algorytmy genetyczne w rozwiązywaniu zagadnień odwrotnych w geofizyce*. „Geoinformatica Polonica” 2006.
- [6] Nikravesh M., Aminzadeh F.: *Past, present and future intelligent reservoir characterization trends*. „Journal of Petroleum Science and Engineering” 2001, vol. 31 (2–4), s. 67–79.
- [7] Romero C. E., Carter J. N.: *Using genetic algorithm for reservoir characterization*. „Journal of Petroleum Science and Engineering” 2001, vol. 31, s. 113–123.
- [8] Sen M., Stoffa P. L.: *Global optimization methods in geophysical inversion*. Elsevier, 1995.
- [9] Stoffa P. L., Sen M. K.: *Nonlinear multiparameter optimization using genetic algorithms: Inversion of plane-wave seismograms*. „Geophysics” 1991, vol. 56, s. 1794–1810.
- [10] Velez-Langs O.: *Genetic algorithms in oil industry: An overview*. „Journal of Petroleum Science and Engineering” 2005, vol. 47, s. 15–22.



Mgr Karolina PIROWSKA – absolwentka kierunku Matematyka na Uniwersytecie Jagiellońskim, specjalizacja: Zastosowania Matematyki. Od lutego 2007 roku pracownik Zakładu Sejsmiki Instytutu Nafty i Gazu w Krakowie. Zajmuje się zastosowaniem metod optymalizacji stochastycznej w rozwiązaniu odwrotnego zadania kinematycznego w metodzie sejsmicznej.