

*Ryszard Snopkowski**

BŁĘDY I NIEDOKŁADNOŚCI W SYMULACJI STOCHASTYCZNEJ PROCESÓW – MOŻLIWOŚCI ICH WERYFIKACJI**

1. Wprowadzenie

Symulacja stochastyczna – jako metoda badawcza – stosowana jest od ponad pół wieku. Początki jej wykorzystywania to lata czterdzieste ubiegłego wieku. Istotą symulacji stochastycznej jest eksperyment przeprowadzany w komputerze, polegający na wielokrotnej symulacji działania analizowanego procesu oraz rejestracji szukanych charakterystyk wynikowych. Do lat osiemdziesiątych, symulacja stochastyczna była metodą z konieczności elitarną. Badania z jej użyciem wymagają bowiem komputera, a wówczas znajdowały się one jedynie w centrach obliczeniowych (np. ODRY, RIADY, IBM-y). W roku 1976 pojawił się pierwszy komputer osobisty i praktycznie od tego momentu symulacja stochastyczna jest metodą badawczą, której dostępność przestała być ograniczona. Współczesne komputery, tzw. osobiste (ten termin już właściwie jest coraz rzadziej stosowany), dysponują parametrami (szybkość obliczeń, pamięci), które dają użytkownikowi duże możliwości w zakresie wykorzystywania symulacji stochastycznej jako metody badawczej.

Jednocześnie jednak, stosowanie tej metody związane jest z możliwościami popełniania błędów lub niedokładności, na które autor zwraca uwagę w niniejszym artykule, podając także możliwości ich weryfikacji. Symulacja stochastyczna jest metodą badawczą, która jest wykorzystywana również w górnictwie, w analizie wybranych procesów produkcyjnych, o czym świadczą m.in. prace [2, 3].

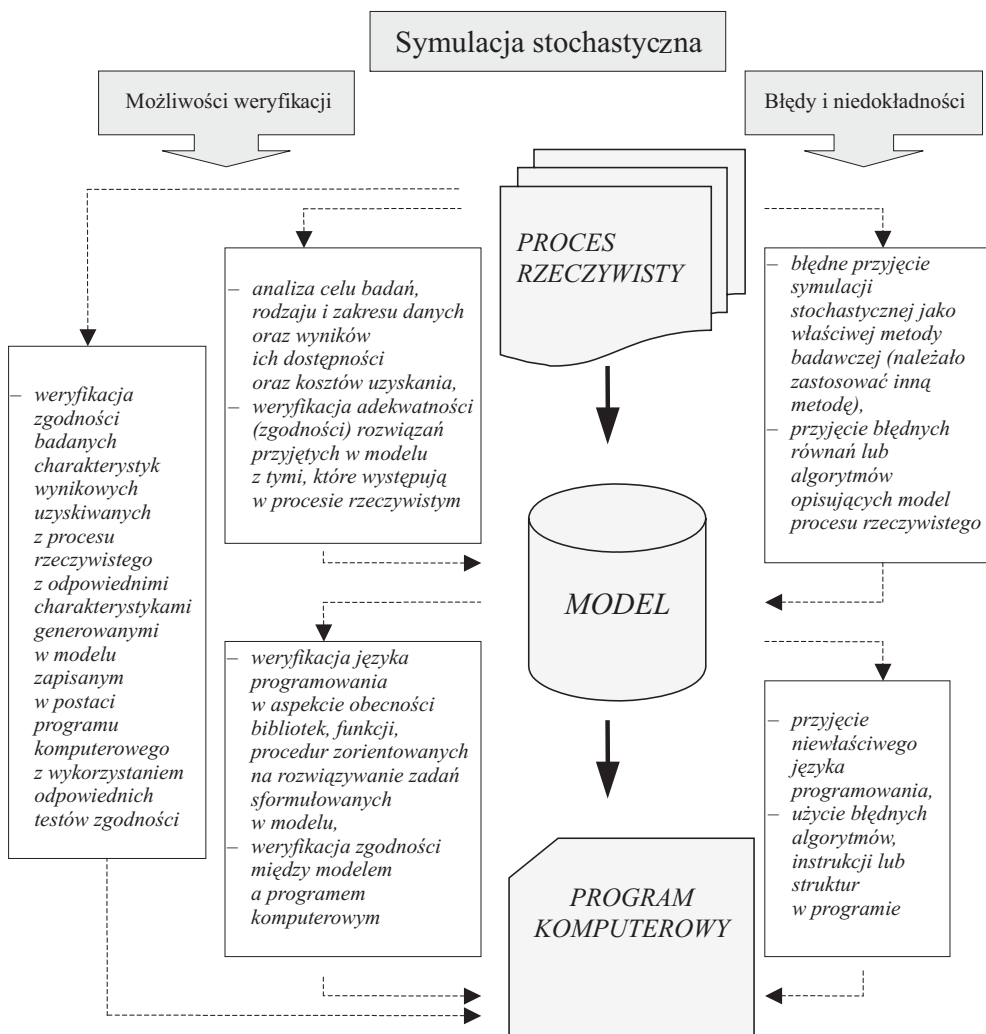
Zawarte w niniejszym artykule uwagi, mogą być – choć w części – przydatne w trakcie stosowania symulacji jako metody badawczej.

2. Błędy i niedokładności w symulacji stochastycznej

Wykorzystaniu symulacji stochastycznej – jako metody badawczej – może potencjalnie towarzyszyć popełnianie błędów lub niedokładności, które scharakteryzowano ogólnie na rysunku 1.

* Wydział Górnictwa i Geoinżynierii, Akademia Górniczo-Hutnicza, Kraków

** Praca została zrealizowana w ramach tematu badawczego 11.11.100.279



Rys. 1. Błędy i niedokładności oraz możliwości ich weryfikacji w symulacji stochastycznej

Poszczególne pojęcia oznaczają: *proces rzeczywisty* to proces, który podlega badaniu metodą symulacji; *modelem* jest zestaw równań, nierówności i (lub) algorytmów, które zostały przyjęte jako opis matematyczny procesu rzeczywistego; *programem komputerowym* jest zapis modelu w wybranym języku programowania.

Błędy i niedokładności, które potencjalnie mogą wystąpić, to:

- **Błędne przyjęcie symulacji stochastycznej jako właściwej metody badawczej.**

Symulacja stochastyczna jest metodą, która może być stosowana, jeżeli badany proces ma charakter stochastyczny, czyli taki, w którym występuje co najmniej jedna zmienna losowa. Praktycznie, jest to jedyny warunek, który musi być spełniony, by zastosować tę metodę.

Można zatem stwierdzić, iż symulacja stochastyczna nie wymaga spełnienia wielu warunków wstępnych, by ją wykorzystać w badaniu wybranego procesu rzeczywistego. To powoduje, iż należy dołożyć szczególnej staranności w trakcie podejmowania decyzji dotyczącej wyboru metody badawczej (może należy zastosować inną metodę?).

Jeśli przyjmuje się symulację stochastyczną jako metodę badań, należy także mieć świadomość jej wad. Metoda nie gwarantuje bowiem znalezienia rozwiązania optymalnego, choć może wskazywać kierunki zbliżania się do niego. Symulując np. przebieg pewnego procesu technologicznego, można uzyskać jego charakterystykę w długim okresie, zawierającą rozkłady niezawodnościowe układu technologicznego, rozkłady przerw w pracy, itd. W tym znaczeniu, wykorzystując symulację stochastyczną, można przyczynić się do optymalizacji tego procesu, wskazując jego słabe ogniwa.

— **Przyjęcie błędnych równań lub algorytmów opisujących model procesu rzeczywistego.**

Charakter procesu rzeczywistego (ciągły, nieciągły, statyczny, dynamiczny, itp.), wpływa na przyjęcie określonej metody jego zapisu w modelu. Warunkiem przyjęcia właściwych (adekwatnych) metod opisu procesu rzeczywistego jest jego dogłębna znajomość, a także wiedza z zakresu metod matematycznych, umożliwiających jego sformalizowany zapis.

— **Przyjęcie niewłaściwego języka programowania.**

Etap zapisu modelu w postaci programu komputerowego musi być poprzedzony analizą, której celem jest wybór właściwego języka programowania. Właściwego tzn. takiego, który jest zorientowany (w zakresie swojej struktury, bibliotek, itp.), na rozwiązywanie zadań sformułowanych w modelu.

Błąd w wyborze języka programowania może skutkować dłuższym programem (konieczność zapisu procedur, które w innych językach są już gotowe), wzrostem kosztów obliczeń, itp.

Przez program komputerowy można tu także rozumieć np. arkusz kalkulacyjny Excel. Wykorzystanie arkusza ma jednak miejsce zwykle w przypadkach niezbyt skomplikowanych modeli.

— **Użycie błędnych algorytmów, instrukcji lub struktur w programie.**

Kolejnym błędem, który może wystąpić, to brak zgodności między modelem a programem komputerowym. Dzieje się tak, jeśli programista popełni błędy w tworzeniu programu, które mogą polegać na użyciu błędnych algorytmów, instrukcji lub struktur w programie.

Błahy błąd programisty może skutkować całkowicie innym od zamierzonego działaniem programu. Świadczy o tym przykład, który nie jest wprawdzie związany z symulacją procesów produkcyjnych, lecz bardzo dobrze ilustruje konsekwencje drobnego błędu w tworzeniu programu. Przykład ten można znaleźć w literaturze z zakresu niezawodności oprogramowania.

Błąd programisty był – na pierwszy rzut oka – niewielki. W jednej z instrukcji programu, zamiast przecinka wstawił omyłkowo kropkę. Skutki tego były jednak kosztowne – w sumie kilka milionów dolarów. Program nie wykazywał błędu formalnego (niezgodności z zasadami języka), lecz jego działanie było całkowicie inne od zamierzonego, co przedstawia rysunek 2.

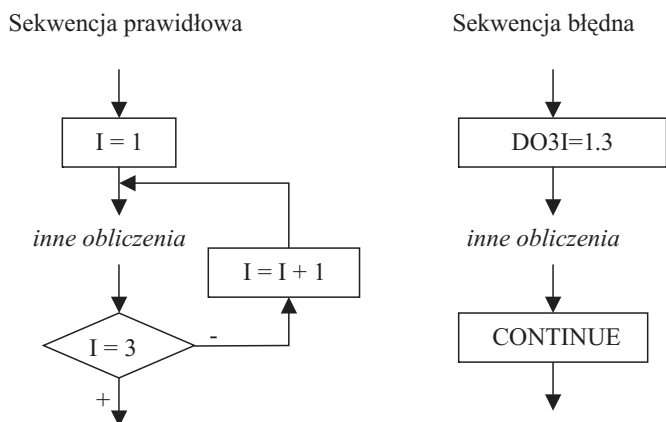
Poprawna sekwencja instrukcji powinna mieć postać następującą:
DO 3 I = 1, 3

.....
3 CONTINUE

Programista popełnił błąd, umieszczając zamiast przecinka, kropkę:
DO 3 I = 1. 3

.....
3 CONTINUE

Działanie obu sekwencji instrukcji zamieszczono na rysunku 2.

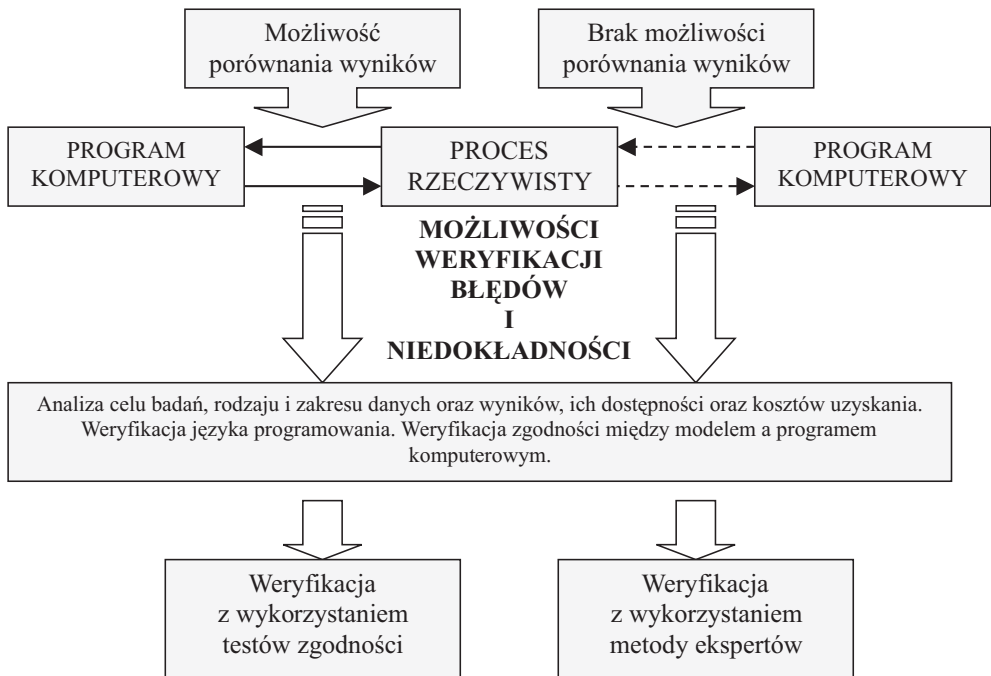


Rys. 2. Prawidłowa i błędna sekwencja instrukcji na schemacie blokowym

Obie postacie instrukcji, czyli „DO 3 I = 1,3” oraz „DO 3 I = 1.3” są poprawne z punktu widzenia reguł języka, lecz postać pierwsza to inicjacja pętli obliczeniowej (realizowanej wielokrotnie, do momentu uzyskania przez zmienną „I” wartości 3), postać „z kropką” nie inicjuje pętli (jest instrukcją podstawienia), stąd „inne obliczenia” były wykonywane tylko raz, co było przyczyną awarii. Program sterował działaniem jednej z pierwszych sond kosmicznych wysłanych w kierunku Wenus (napisano go w języku Fortran – obecnie już rzadziej stosowanym, choć w dalszym ciągu chętnie wykorzystywanym, np. przez fizyków).

3. Możliwości weryfikacji błędów i niedokładności w symulacji stochastycznej

Na rysunku 3 zamieszczono schemat możliwych działań w zakresie weryfikacji błędów i niedokładności w symulacji stochastycznej. Działania weryfikacyjne uzależnione są od możliwości porównania badanych charakterystyk wynikowych uzyskiwanych z procesu rzeczywistego z odpowiednimi charakterystykami generowanymi w modelu, zapisanym w postaci programu komputerowego.



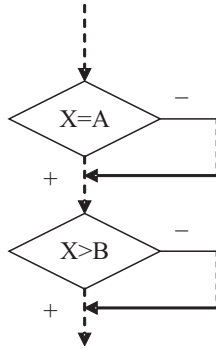
Rys. 3. Schemat możliwości weryfikacji błędów i niedokładności w symulacji stochastycznej

Możliwość porównania wyników ma miejsce w symulacji stochastycznej tych systemów, które można obserwować w rzeczywistości (mogą to być np. systemy produkcyjne). Brak możliwości porównania wyników jest związany z symulacją procesów, których rzeczywisty czas przebiegu jest długi (np. procesy geologiczne). Niezależnie od charakteru modelowanego procesu – w każdym przypadku – weryfikacja poprawności budowy i wykorzystania modelu symulacyjnego w badaniach powinna być poprzedzona szczegółową analizą celu badań, a także możliwości w zakresie uzyskania danych do budowanego modelu, w tym także oceny ich dostępności i kosztów uzyskania.

Weryfikacja poprawności programu komputerowego, będącego numerycznym modelem badanego procesu, jest zawsze trudna. Celem ilustracji tego zagadnienia – w dalszej części – zaprezentowano przykład, który pomimo szeregu założeń upraszczających może dać ocenę złożoności problemu weryfikacji poprawności programu komputerowego.

Jedną z metod weryfikacji (oczywiście nie jedyną) jest przeprowadzenie obliczeń testowych wszystkimi tzw. „ścieżkami” w programie i wnioskowanie o poprawności programu, jeśli obliczenia nie wykazują błędów. O tym, że nie zawsze jest to możliwe, świadczy poniższy przykład (opracowany przez autora i zamieszczony w pracy [1]).

Założmy, że w programie występuje tzw. „IF logiczne” w postaci zamieszczonej na rysunku 4. Znaki „-” oraz „+” oznaczają kierunek dalszych obliczeń w zależności od tego czy podane wyrażenie jest prawdą („+”) czy fałszem („-”). W miejscu przerywanej linii mogą występować inne instrukcje programu.



Rys. 4. Fragment zawierający dwie struktury typu „IF logiczne”

Jak łatwo zauważyć, by przetestować fragment zamieszczony na rysunku, należy wykonać obliczenia czterema ścieżkami: pierwszą, jeśli $X = A$ i $X > B$, drugą, jeśli $X = A$ i $X \leq B$, trzecią, gdy $X \neq A$ i $X > B$, czwartą, jeśli $X \neq A$ i $X \leq B$.

Założmy, że w programie występuje 100 wierszy (instrukcji), zawierających tzw. „IF logiczne” (takich jak na rysunku 4).

Liczba ścieżek może być zatem obliczona według wzoru

$$LS = 2^{LIF} \quad (1)$$

gdzie:

LS — liczba ścieżek w programie (obliczana w tym przypadku jedynie na podstawie instrukcji, tzw. „IF logiczne”),

LIF — liczba instrukcji „IF logiczne” w programie.

Dla stu instrukcji „IF logiczne” w programie („powiązanych” jak na rysunku) wzór ma postać

$$LS = 2^{100} \quad (2)$$

W przybliżeniu, można przyjąć, iż $2^{10} \approx 10^3$, czyli

$$LS = 10^{30} \quad (3)$$

Założmy, że czas obliczeń, przeprowadzanych jedną ścieżką trwa jedynie 0,001 sekundy, czyli czas obliczeń zrealizowanych wszystkimi ścieżkami (dla uproszczenia zakładamy, że czas ten dla każdej ścieżki jest taki sam) trwałby

$$CZAS\ OBLICZEŃ = 10^{30} \cdot 10^{-3} = 10^{27} \text{ [sekund]} \quad (4)$$

Jeden rok to około $3 \cdot 10^7$ sekund, czyli program należałoby testować przez czas równy

$$\text{CZAS OBLICZEŃ} = \frac{1}{3} \cdot 10^{20} \text{ [lat]} \quad (5)$$

Dla porównania, szacowany wiek Wszechświata to około 14 miliardów lat, czyli $1,4 \cdot 10^{10}$ lat.

Otrzymany czas obliczeń – pomimo swej absurdalności oraz wielu uproszczeń w przykładzie (np. programy dzieli się na mniejsze moduły) – daje pewne wyobrażenie o trudnościach w weryfikacji programów. Trudności te oczywiście rosną wraz ze wzrostem złożoności programów.

Stosowane techniki weryfikacyjne (wyszukujące błędy) powinny obejmować sposoby, strategie oraz plany testowania programów, sposoby oraz metody wykrywania i rozpoznawania błędów, zasady konserwacji oprogramowania, itp.

Nigdy jednak nie ma stuprocentowej pewności, że program jest pozbawiony błędów. Nawet po wielu latach użytkowania oprogramowania może się zdarzyć tzw. *execution error* (błąd wykonania), wynikający np. ze specyficznej konfiguracji danych, która do tej pory nie występowała. Przykładem jest system operacyjny każdego komputera (tu można wstawić dowolną nazwę), gdzie pomimo wielu lat użytkowania przez miliony osób, wciąż pojawiają się błędy (wprawdzie wraz z upływem czasu coraz rzadziej, ale jednak). Uwzględniając to wszystko, firmy tworzące programy nigdy nie biorą pełnej odpowiedzialności np. za szkody wynikłe z błędnego ich funkcjonowania.

3.1. Weryfikacja z wykorzystaniem testów zgodności

Zamieszczone poniżej metody weryfikacji wyników symulacji mogą być stosowane, jeżeli istnieje możliwość porównania zbioru $\{x_1, x_2, \dots, x_n\}$, ze zbiorem $\{x_1^{rz}, x_2^{rz}, \dots, x_{n_{rz}}^{rz}\}$.

Zbiór $\{x_1, x_2, \dots, x_n\}$ zawiera liczbową charakterystykę badanej cechy modelu, uzyskaną na drodze symulacji pewnego procesu rzeczywistego, natomiast $\{x_1^{rz}, x_2^{rz}, \dots, x_{n_{rz}}^{rz}\}$ jest zbiorem zawierającym liczbową charakterystykę tej samej cechy, uzyskaną z obserwacji tego procesu w warunkach rzeczywistych.

Można zatem postawić tezę, iż cały proces symulacji (budowa modelu, uruchamianie modelu) w sposób wiarygodny (na określonym poziomie ufności) oddaje istotę funkcjonowania procesu rzeczywistego wówczas, gdy testy statystyczne nie pozwalają na odrzucenie hipotezy o przynależności do tej samej populacji generalnej wyników uzyskanych na drodze symulacji oraz otrzymanych z procesu rzeczywistego.

Weryfikacja wyników symulacji polega zatem w tym przypadku na analizie dwóch prób statystycznych, pod kątem ich pochodzenia z tej samej populacji generalnej. W tym celu można skorzystać m.in. z następujących testów zgodności [1]:

- test równości wartości oczekiwanych dwóch zbiorowości generalnych,
- test równości wariancji dwóch zbiorowości generalnych,
- test zgodności rozkładów dwóch zbiorowości generalnych – test Smirnowa,
- test zgodności rozkładów dwóch zbiorowości generalnych – test Walda i Wolfowitza (test serii).

Wymienione powyżej testy umożliwiają weryfikację wyników symulacji jedynie wówczas, gdy istnieje możliwość porównania wyników uzyskanych z symulacji z procesem rzeczywistym.

Jeśli testy statystyczne odrzucają hipotezę o przynależności do tej samej populacji generalnej wyników uzyskanych na drodze symulacji oraz otrzymanych z procesu rzeczywistego, wówczas należy podjąć ponowną analizę poprawności procesu symulacji stochastycznej (analizę modelu, analizę programu komputerowego, w tym metod generowania liczb losowych, itp.).

W wielu przypadkach modeli symulacyjnych, dotyczących procesów górniczych, można z powodzeniem stosować metody weryfikacji z wykorzystaniem testów zgodności.

3.2. Weryfikacja z wykorzystaniem metody ekspertów

Symulacja stochastyczna jest metodą badawczą, którą stosuje się z powodzeniem w modelowaniu procesów trwających w rzeczywistości długi czas (dziesiątki, setki lub nawet tysiące lat). Mogą to być procesy dotyczące np. rozwoju i powstawania formacji geologicznych na jakimś obszarze. Do tych procesów można także zaliczyć procesy ewolucyjne, rozwoju gatunków, itp. Wszystkie łączy wspólna cecha – bardzo trudno jest, lub jest to wręcz niemożliwe, śledzić i porównywać wartości wynikowe z modelu symulacyjnego z wartościami uzyskiwanymi w procesach rzeczywistych.

Jedyną metodą weryfikacji modeli rozwiązywaną metodą symulacji stochastycznej jest wówczas szczegółowa analiza tych modeli dokonywana przez ekspertów. Analizowane są wszelkie – wykorzystywane w modelu – procedury symulacyjne. Szczegółnej analizie poddawany jest model w aspekcie wykorzystywanych w nim algorytmów. Badana jest ich zasadność (zgodność z procesem rzeczywistym).

W tej grupie weryfikowanych modeli stosunkowo rzadko występują modele procesów górniczych. Mogłyby do nich należeć np. modele obejmujące symulację funkcjonowania kopalni w okresie jej istnienia lub podobne, dotyczące długiego czasu funkcjonowania procesu rzeczywistego, podlegającego symulacji.

4. Wnioski końcowe

Symulacja stochastyczna jest metodą badawczą posiadającą wiele zalet. Do najważniejszych można zaliczyć możliwość „obserwacji” badanego procesu, przy wykorzystaniu do tego celu jego komputerowego modelu. Wielokrotne uruchamianie modelu (symulacja) umożliwia uzyskanie charakterystyk, które są osiągalne w trakcie obserwacji procesu rzeczywistego przez długi okres, co w wielu przypadkach nie jest możliwe.

Symulacja stochastyczna umożliwia odpowiedź na pytanie „co się stanie, jeśli...”. Co się stanie z systemem komunikacyjnym, jeśli wprowadzimy w jego części ruch jednokierunkowy, gdzie pojawiają się korki, jaki będzie przeciętny czas przejazdu, itp. Co się stanie z systemem produkcyjnym, jeśli wprowadzimy maszyny o innych charakterystykach niezawodnościowych, w jakim zakresie wpłynie to na jego funkcjonowanie jako systemu.

Symulacja stochastyczna jest metodą badawczą stosowaną z powodzeniem w analizie procesów górniczych. Metodę wykorzystywano w symulacji procesów produkcyjnych

w przodku ścianowym, w analizie transportu poziomego i pionowego w kopalni, w symulacji pracy zbiorników wyrównawczych, w modelowaniu robót przygotowawczych i udostępniających.

Stosowanie metody związane jest także z możliwością popełniania błędów lub niedokładności, rzutujących w efekcie na uzyskiwane wyniki. W niniejszym artykule zwrócono uwagę na różne rodzaje błędów, które mogą być popełniane na etapie tworzenia modelu oraz opracowywania programu komputerowego. Poruszono problematykę weryfikacji tych błędów, zamieszczając wiele przykładów oraz metod, które jednak nie wyczerpują tematyki tego złożonego zagadnienia.

LITERATURA

- [1] *Snopkowski R.*: Symulacja stochastyczna, Uczelniane Wydawnictwa Naukowo-Dydaktyczne AGH, Kraków 2007
- [2] *Snopkowski R.*: The use of the Stochastic Simulation for Identification of the Function of Output Probability Density, Archives of Mining Sciences, Polish Academy of Sciences, Vol. 50, No. 4, 2005, pp. 497–504
- [3] *Snopkowski R.*: Boundary Conditions for Elementary Functions of Probability Densities for the Production Process Realized in Longwalls, Archives of Mining Sciences, Polish Academy of Sciences, Volume 45, Issue 4, Wydawnictwo Naukowe PWN, Warszawa–Kraków 2000