ZACHODNIOPOMORSKI UNIWERSYTET TECHNOLOGICZNY
WYDZIAŁ INFORMATYKI

ROZPRAWA DOKTORSKA

Joanna Bobkowska

DETERMINISTYCZNA METODA ANALIZY DANYCH NA MAŁOLICZNYCH ZBIORACH PRÓBEK

Promotor rozprawy
prof. dr hab. inż. Walery Rogoza

Szczecin 2015

Składam serdeczne podziękowania mojemu Promotorowi, Panu prof. dr hab. inż. Waleremu Rogozie, za inspirację naukową do podjęcia tematu niniejszej pracy, poświęcony czas oraz cenne wsparcie merytoryczne, które przyczyniły się do finalnego kształtu pracy

Joanna Bobkowska

Spis treści

Wstęp

Jednym z zagadnień współczesnej informatyki jest opracowywanie nadchodzących ze źródeł zewnętrznych informacji o obiekcie w celu symulacji zachowania tego obiektu w warunkach jego współdziałania ze środowiskiem. Symulacja może mieć kilka celów: poznanie lub sprecyzowanie funkcji składników strukturalnych omawianego obiektu, w przypadku, gdy znana jest architektura obiektu (zagadnienie identyfikacji systemów w ujęciu wąskim), albo tworzenie matematycznych modeli całego obiektu, rozpatrując go jako czarną skrzynkę (zagadnienie identyfikacji w ujęciu szerokim) [1], określenie odpowiednich sposobów aktywnego oddziaływania na obiekt (problemy sterowania) [2] lub w przypadku, gdy nie dysponujemy takimi zasobami w pełnym zakresie, wyznaczenie z pewnym prawdopodobieństwem wiarygodności tych stanów obiektu, których możemy oczekiwać od danego obiektu w przyszłości (problemy prognozy) [3].

Pierwsze dwa zagadnienia rozwijane są w ramach klasycznej teorii systemów oraz mają dobrze opracowane, rozbudowane zasoby matematyczne. Natomiast trzecie, ze względu na to, że jego korzenie sięgają podstaw klasycznej cybernetyki [4], otrzymało w ostatnich latach jakościowo nowe zabarwienie w związku z udowodnieniem przez różnych naukowców tezy o zasadniczej możliwości i skuteczności zastosowania do rozwiązania problemów prognozy, metod ewolucyjnego opracowywania informacji, takich jak metody symulacji z wykorzystaniem sieci neuronowych, algorytmów genetycznych oraz symulacji ewolucyjnej. Za jedną z bardziej interesujących prac z tej dziedziny, wydanych w ostatnich latach można uznać, np. monografię [5], w której wykonano wszechstronne badania zastosowania metody grupowania argumentów w zadaniach zautomatyzowanej predykcji zachowania rynków finansowych.

Szczególnego znaczenia nabierają technologie dynamicznego opracowywania danych oraz ewolucyjnej syntezy matematycznych modeli zachowania się złożonych obiektów czy systemów. Według teorii systemów, obiekty i systemy złożone charakteryzują się ciągiem specyficznych właściwości [6], [7], z których najczęściej wyróżnia się następujące: 1) nieentropijność informacyjna, czyli zmniejszenie nieokreśloności (entropii[1] informacyjnej), reakcji systemu na zewnętrzne oddziaływania w trakcie działania systemu (obiektu) i odpowiednio zwiększenie w czasie stopnia przewidywalności przyszłych stanów systemu (uważa się, że nieentropijność złożonego

[1] Entropia jest miarą stopnia nieuporządkowania układu.

systemu powstaje w wyniku optymalizacji działań systemu, właściwej wszystkim adaptacyjnym i samoorganizującym się systemom), 2) heterogeniczność[2] składników, 3) unikalność zachowania się systemu oraz 4) emergentność[3], (ang., *emergence*).

Należy podkreślić, że istniejące tzw. „klasyczne" metody symulacji systemów, w większości przypadków okazują się nadzwyczaj nieefektywne, albo w ogóle nie są przydatne w przypadku, gdy mamy do czynienia z obiektami i systemami złożonymi, ponieważ metody te oparte są na konserwatywnych modelach obiektów, czyli modelach tworzonych apriorycznie, przed rozpoczęciem procesu symulacji (ich tworzenie jest możliwe w przypadku zagadnienia symulacji obiektów, których zachowanie można przewidzieć z wysokim stopniem prawdopodobieństwa w perspektywie oddalonej). Według omówionej powyżej cechy unikalności zachowania obiektów i systemów złożonych, badacz nie jest w stanie dokładnie przewidzieć, który z ewentualnych modeli zachowania będzie najbardziej przydatny do symulacji zmian stanów wspomnianego systemu (obiektu). W tych warunkach badacz może się spotkać z sytuacją, gdy albo (w najlepszym przypadku) apriorycznie wybrane losowo modele okazały się w pełni zadowalające (dzięki „doświadczeniu" i „intuicji" badacza), albo, w gorszym przypadku, wybrane modele mogą w ogóle nie odpowiadać rzeczywistemu zachowaniu się symulowanego obiektu. Jednakże, liczenie na łut szczęścia, nie jest rozsądnym ani profesjonalnym podejściem w badaniach naukowych. Dlatego od czasu, gdy w teorii systemów i symulacji komputerowej została ściśle określona kategoria złożonych systemów, jako obiektów badań o specyficznych właściwościach, studiowanie złożonych systemów nabyło status samodzielnego kierunku badań w informatyce oraz pokrewnych dziedzinach naukowych (cybernetyce, analizie systemowej, ekonomii itp.).

Dalej w tekście rozprawy przyjmujemy, że system może składać się z pojedynczego lub kilku obiektów. A jeżeli chodzi o system złożony, to wychodząc od cechy emergentności, zachowanie się złożonego systemu nie może zostać przedstawione, jako suma reakcji prostych obiektów, tzn. składniki złożonego systemu są również podsystemami złożonymi. Wskutek owej specyficzności złożonych systemów, wyniki dotyczące tych systemów w równej mierze mogą być stosowane do obiektów złożonych

[2] Heterogeniczność jest to niejednorodność (zróżnicowanie). Może tyczyć się każdej dziedziny życia. W środowisku informatycznym oznacza zróżnicowaną wielkość struktur danych, różną funkcjonalność, różne modele danych.

[3] Zjawisko (lub system) określa się jako emergentne, jeśli w jakimś sensie nie można go określić opisując jedynie jego części składowe. W ścisłym sensie naukowym oznacza to nieredukowalność danego zjawiska (systemu).

i odwrotnie, czyli w tym kontekście terminy „system”, „obiekt” lub „proces” będziemy uznawali za synonimy, bowiem otrzymane w pracy wyniki, w równej mierze dotyczą zarówno systemów złożonych, jak i obiektów czy też procesów złożonych.

Współczesne sposoby rozwiązywania opisanego zagadnienia symulacji systemów złożonych opierają się na idei tworzenia modeli na bieżąco, podczas dokonywania procesu symulacji. Takie podejście do symulacji niektórzy badacze określają jako samoorganizację modeli, lub symulację ewolucyjną (ze względu na to, że proces tworzenia modeli przypomina proces ewolucji naturalnej) [8], inni – jako adaptację modeli (ze względu na to, że modele są adaptowane do ciągle zmieniających się stanów obiektów podlegających symulacji) [9], [10]. Autor tej pracy podziela zdanie autorów książki [8], że adaptacja jest uogólniającą kategorią określającą zdolność modeli do przystosowania się do zmieniających się warunków, a skuteczną metodą realizującą adaptację jest samoorganizacja ewolucyjna modeli. Dalej, więc będziemy rozumieć termin „adaptacja” w ten sposób.

Proces adaptacyjnej budowy modeli obiektów złożonych realizowany jest poprzez proces obejmujący: 1) szczegółowe zbieranie danych o obiekcie z dostępnych źródeł; 2) analizę zebranych informacji w celu wyboru istotnych dla rozwiązania danego problemu wiadomości i faktów oraz wyrzucania nieistotnych; 3) formułowanie wiedzy o obiekcie badań; 4) tworzenie i dopracowywanie modelu zachowania się obiektu, który coraz bardziej precyzyjnie odzwierciedli rzeczywiste zachowanie obiektu; 5) predykcję zachowania się obiektu w przyszłości.

Mówiąc o prognozie, możemy podzielić ją na krótko- i długoterminową. Ponadto, zarówno pierwszą, jak i drugą można podzielić na jakościową (ocena możliwości lub niemożliwości wystąpienia pewnego zdarzenia w przyszłości) i ilościową (oceny możliwych liczbowych przedziałów wartości prognozowanych zmiennych). Problem ilościowej prognozy długoterminowej, przy odpowiedniej złożoności systemu jest wyjątkowo skomplikowany i jak dowodzą źródła naukowe, do końca nie został jeszcze rozwiązany [8]. Powodem jest to, że prognozy przeważnie oparte są na założeniach ekspertów, którzy są w stanie mniej lub bardziej dokładnie ocenić sytuację w najbliższej perspektywie, ale stają się bezradni, gdy chodzi o perspektywę bardziej odległą.

Jako przykład może służyć problem ocieplania klimatu na kuli ziemskiej. Obserwując tendencje zmiany średniej temperatury rocznej w różnych części Ziemi w ostatnich latach, niektórzy naukowcy doszli do wniosku, że mamy do czynienia z nieuniknionym ociepleniem klimatu. Jednak zima 2009-2010 roku poddaje

w wątpliwość tę hipotezę. Inny przykład – niestety eksperci nie są w stanie ściśle określić momentów kryzysowych, jakie od czasu do czasu wstrząsają światową ekonomią i mimo wysiłków wielu ekspertów, nie udaje się ich przewidzieć zawczasu.

Szeroko rozpowszechniona subiektywna analiza systemowa i modelowanie imitacyjne bazują na podejściu dedukcyjnym i wymagają dogłębnego badania symulowanych obiektów i gromadzenia dosyć obszernych o nich informacji. Takie podejście do symulacji klasyfikujemy, jako „klasyczne".

Klasyczne metody symulacji przewidują, że poziom wiedzy o obiekcie musi być na tyle wysoki, aby można było budować na tyle pełne matematyczne równania, na ile ścisłą ocenę liczbową chcemy odnaleźć dla wszystkich możliwych stanów obiektu. Niestety, w wielu dziedzinach, np. w biologii, socjologii, ekologii, klimatologii, hydrologii i innych naukach przyrodniczych i technicznych, nie udaje się nam osiągnąć takiego stanu wiedzy.

W przeciwieństwie do metod „klasycznych", wspólną własnością metod, określonych powyżej, jako metody adaptacyjne jest to, że oparte są one na podejściu indukcyjnym, dlatego w pewnym sensie wolne są od surowych wymagań właściwych metodom tworzonym na podstawie podejścia dedukcyjnego. Wybierając wersję metody opieramy się na pewnych kryteriach, są one formalizowane w taki sposób, że przerzucają olbrzymią część obliczeń na komputer, który pozbawiony jest możliwości subiektywnych ocen, tak charakterystycznych dla ekspertów-ludzi i zdolny jest do optymalizacji tworzonych modeli, niekiedy nawet na podstawie mało licznych próbek. Istotna możliwość tworzenia takich systemów symulujących i ich wysoka skuteczność działania potwierdzona jest szeregiem publikacji, oraz umocniona przykładami praktycznej implementacji tych systemów (np. [8], [9], [10], [11], [12]).

Poza tym, w pracach badaczy zajmujących się zagadnieniem tworzenia metod adaptacyjnych przeznaczonych do symulacji obiektów o złożonym charakterze zachowania się i opartych na podejściu indukcyjnym zaznaczona jest ich jedna wspólna, słaba strona, jak dotąd, dosyć charakterystyczna dla tych metod: czynnikiem decydującym, który zasadniczo wpływa na jakość tworzonych modeli i odpowiednio jakość rozwiązania problemów z wykorzystaniem tych modeli (dokładność rozwiązania problemu i szybkość znalezienia tego rozwiązania), jest należyte dobranie danych (czyli odpowiedniość i pełność zbioru dobieranych danych, które są w danym momencie wymagane dla poprawnego skonstruowania modelu symulowanego obiektu). Jeżeli dane zostały dobrane niewłaściwie (np. wybrane są informacje nieistotne), system formułuje

fałszywe przekonania o właściwościach obiektu i tworzy modele, które mogą być bardzo odległe od rzeczywistości. Sytuacja komplikuje się bardziej w momencie, gdy podczas symulacji złożonego obiektu, nadchodzące dane są niepełne, niekiedy sprzeczne ze sobą czy rozmyte. Jeżeli wrażliwość modelu na nadchodzące dane jest bardzo wysoka, system symulujący w odpowiednich, niesprzyjających warunkach może wygenerować modele niepoprawne, czyli modele nieodpowiadające rzeczywistemu stanowi rzeczy i ostatecznie straci on zdolność do poprawienia modelu nawet, jeżeli później będą napływały do systemu dane niezbędne. Pewną analogią do tej sytuacji w metodach matematyki obliczeniowej może być rozbieżność procesu obliczeń, dlatego zjawisko to będziemy nazywali *utratą zbieżności procesu syntezy modeli obiektów*.

Opisane sytuacje są przedmiotem badań analizy systemowej [13], w ramach której opracowane są specjalne metody zmniejszenia wrażliwości procesu identyfikacji systemów na nieścisłość nadchodzących danych – chodzi o tzw. *metody regularyzacji*. Tej kwestii poświęconych jest szereg fundamentalnych publikacji matematycznych moskiewskiej szkoły pod kierownictwem prof. A.N. Tichonowa (systemy symulacji, w których pokazana została silna zależność modeli od nadchodzących danych, otrzymały nazwę *systemów tichonowskich*). Tym nie mniej, kwestia zastosowania metod regularyzacji Tichonowa w celu zmniejszenia wrażliwości procesów symulacji na niedokładności danych wejściowych jest, jak na razie, otwarta w naukach stosowanych i oczekuje na swoje rozwiązanie.

Wstępna analiza stanu problematyki syntezy modeli zachowania się złożonych obiektów i systemów daje podstawy sformułować następujące cele i tezę danej rozprawy:

Celem pracy jest *opracowanie metody opartej na teorii identyfikacji systemów, wykorzystującej wielomian Kołmogorowa-Gabora (jak w metodzie GMDH* [5]*), która umożliwia analizę oraz rozwiązanie problemu prognozy procesów na małolicznych zbiorach próbek, o skuteczności większej niż uproszczona metoda podstawowa (GMDH), z wykorzystaniem narzędzia funkcji wrażliwości.*

Teza rozprawy jest następująca: *istnieje możliwość zwiększenia dokładności prognozy procesu na małolicznych zbiorach próbek w stosunku do istniejących metod w tym GMDH, przy wykorzystaniu deterministycznych metod identyfikacji systemu wykorzystujących funkcje wrażliwości i szereg tzw. modeli cząstkowych tworzonych na podstawie wielomianu Kołmogorowa-Gabora i podlegających adaptacji w trakcie rozwiązania problemu.*

Struktura pracy: praca składa się ze wstępu, czterech rozdziałów, wniosków, spisu ilustracji, spisu tabel oraz literatury.

W **rozdziale pierwszym** przedstawiono istotę zagadnienia badania, analizy, symulacji systemów złożonych oraz uzasadnienie, że jest to zagadnienie interdyscyplinarne, funkcjonujące na przecięciu wielu dziedzin takich jak teoria systemów, informatyka, ekonomia i inne, a ich badanie wymaga specjalnego podejścia. Pokazano, że w przypadku prób analizy zagadnienia, które wykazują wysoki stopień złożoności, zasadnym jest wykorzystanie spośród wielu dostępnych narzędzi i metod – koncepcji symulacji indukcyjnej w odniesieniu do badania i symulacji opisanych w rozprawie zagadnień.

Prognozę umiejscowiono wewnątrz zagadnienia ewolucyjnych metod analizy danych i procesów. Jest ona jedną z integralnych części systemów wspomagania decyzji, które są potężnym, szeroko stosowanym narzędziem współczesnej informatyki. Rozważa się rozmaite podejścia do rozwiązania problemów prognozowania przyszłych stanów systemów złożonych oraz uzasadnienie stosowności wykorzystania w tym celu metod klasyfikacji i regresji.

W następnej kolejności wskazuje się metodę grupowania elementów, jako tę, która wykorzystuje podstawowe zasady samoorganizacji modeli i może okazać się użyteczna podczas projektowania modeli predykcji złożonych systemów przedstawionych w postaci szeregów czasowych.

W części zawierającej wnioski określa się zadania, które podlegają rozwiązaniu w rozprawie.

W **rozdziale drugim** przedstawiono koncepcję metody opartej na zastosowaniu wielomianów Kołmogorowa-Gabora (WKG – wykorzystywanych również w algorytmie GMDH) zawierającej funkcje wrażliwości oraz jej matematyczne podstawy. Uzasadniono wykorzystanie indukcyjnej metody do analizy systemów złożonych przedstawionych w postaci szeregów czasowych oraz zaprezentowano przykłady ilustrujące skuteczność prognostyczną tej metody przy niewielkiej ilości danych uczących. Metoda generuje kilka modeli wyznaczających przypuszczalne prognozy wartości każdej ze zmiennych szeregu, spośród których wybieramy ex ante, jeden (najkorzystniejszy), oceniając jego wiarygodność na podstawie wartości średnic uzyskanych w każdym modelu rezultatów (rozrzutu uzyskanych prognoz) oraz biorąc pod uwagę dwie wybrane na potrzeby eksperymentów miary odległości (odległość Euklidesa oraz Canberra) i najmniejszą od średniej wszystkich rezultatów

poszczególnych modeli odległość. Prezentujemy tu sposób wykonywania obliczeń na przykładach.

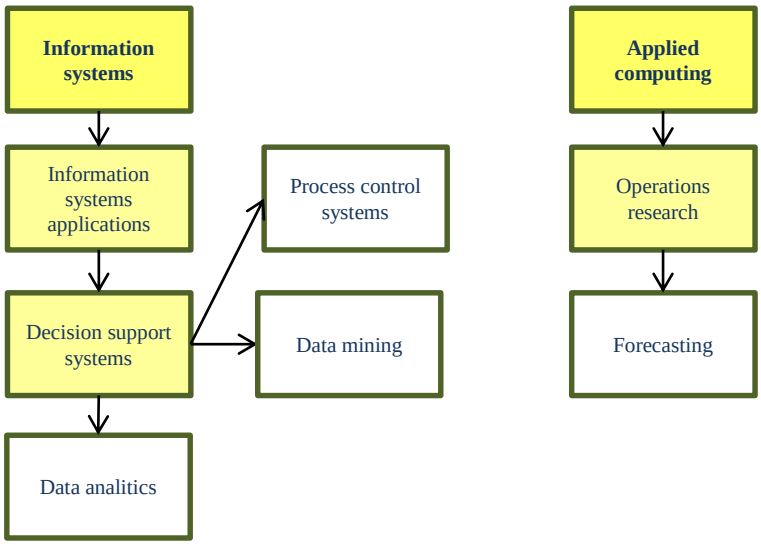
W tym rozdziale wyjaśniamy również, czym różni się nasza metoda, od metody GMDH, z której zapożyczamy koncepcję wykorzystania WKG. Zasadniczym elementem nowości zaproponowanej metody jest zastosowanie funkcji wrażliwości na bazie regresyjnych modeli cząstkowych oraz analiza szeregów czasowych przy niewielkiej ilości danych inicjujących. Niewielka ilość oznacza w naszym przypadku minimalnie siedmiu punktów czasowych (podczas, gdy w metodach statystycznych za niewielką ilość uważa się 50 próbek).

W **rozdziale trzecim** prezentujemy możliwość sterowania obiektami, systemami złożonymi i ich analizę oraz prognozę w przestrzeni stanów. Wskazujemy również możliwość zastosowania filtru Kalmana do prognozowania wartości przyrostów wyznaczonych przez naszą metodę analizy danych z funkcjami wrażliwości.

W **rozdziale czwartym** przedstawiono na przykładach sposób przeprowadzenia badań eksperymentalnych oraz uzyskane wyniki w ujęciu liczbowym i graficznym. Porównano wyniki prognozy uzyskane za pomocą naszej metody, uproszczonej metody GMDH oraz innych popularnych metod regresyjnych.

Ostatnią część stanowią wnioski podsumowujące wyniki badań przeprowadzonych w dysertacji.

Według klasyfikacji *The ACM 2012 Computing Classification System* [14], problem rozpatrywany w danej rozprawie można zakwalifikować następująco: (Rys. 0.1)



Rys. 0.1 Klasyfikacja problemu omawianego w rozprawie według The ACM 2012 Computing Classification System
Źródło: Opracowanie własne na podstawie [14]

1 Analiza istoty i stanu zagadnienia prognozy, jako dziedziny inteligentnej analizy danych i procesów

W tym rozdziale na podstawie analizy istoty i stanu zagadnienia prognozy zachowania się obiektów i procesów udowodnimy tezy, że: (1) osobną klasę wspomnianych obiektów i procesów tworzą tzw. złożone systemy, których badanie wymaga zastosowania specjalnych podejść; (2) analiza zachowania się złożonych systemów, w tym rozwiązanie problemów prognozy, jest jednym z elementów zagadnienia analizy danych i procesów; (3) w klasie metod analizy danych i procesów osobne miejsce zajmują metody bazujące na zasadach symulacji indukcyjnej z wykorzystaniem technik samoorganizacji ewolucyjnej modeli badanego systemu. Na podstawie przedstawionej analizy sformułowany będzie problem prognozy będący obiektem badań w rozprawie.

Na początku określimy cechy systemów złożonych, jako specyficznej kategorii systemów naturalnych i sztucznych. Dalej przeanalizujemy korzenie formowania się koncepcji symulacji indukcyjnej, jako narzędzia do badania zachowania systemów złożonych.

1.1 Badanie złożonych systemów jako zagadnienie interdyscyplinarne

Rozwój technologii informatycznych wiąże się, z jednej strony, z pojawianiem się nowych problemów, które w naturalny sposób wyłaniają się w miarę rozwoju społeczeństwa informacyjnego, z drugiej strony, z koniecznością posługiwania się nowymi metodami do rozwiązywania tzw. tradycyjnych problemów. Przy czym, nierzadko wokół tradycyjnych problemów pojawia się wiele dodatkowych, istotnych szczegółów, które nabierają głębszego naukowego znaczenia, co powoduje konieczność innowacyjnego ujęcia zgromadzonych informacji i inicjuje rozwój nowatorskich zasad oraz metod badań. Przykładem zagadnienia, które można uznać za nowe (gdyż wymaga nowoczesnych metod i technologii współczesnej informatyki do jego rozwiązania) oraz za klasyczne (gdyż ma wieloletnią historię) jest zagadnienie symulacji *złożonych obiektów i procesów* z wykorzystaniem zasobów informatyki.

Symulacja komputerowa jest koniecznym sposobem badania obiektów o złożonym charakterze zachowania i może być ukierunkowana na wyjawienie struktury

i funkcjonalności złożonych obiektów (problem identyfikacji), określenia skutecznych mechanizmów oddziaływania na obiekt (problem sterowania), albo przewidywania zachowania się badanego obiektu w przyszłości (problem prognozy). W tym ostatnim zagadnieniu ważne miejsce zajmują problemy tworzenia metod i algorytmów przewidywania zachowania się złożonych obiektów (systemów, procesów) w przyszłości na podstawie obserwacji ich współdziałania ze środowiskiem zewnętrznym w przeszłych chwilach czasu. Aktualność tych problemów rośnie z roku na rok, ponieważ wraz z rozwojem metod symulacji komputerowej dostępne stają się badania tak złożonych procesów, jak chociażby procesy ekologiczne i socjologiczne, makroekonomiczne czy z dziedziny planowania zużycia naturalnych zasobów Ziemi.

Badania specjalistów z dziedziny informatyki prowadzone wewnątrz tej problematyki skierowane są na zaspokojenie wymagań formułowanych przez ekspertów, analityków i inżynierów z różnych dziedzin nauki, w których badanie złożonych obiektów i procesów jest jednym z głównych przedmiotów ich działalności zawodowej. Mamy wszelkie podstawy, by uznać, że osiągnięcia w dziedzinie tworzenia metod i technik informatycznych przeznaczonych do badania złożonych obiektów i procesów nabierają nierzadko charakteru interdyscyplinarnego i stają się narzędziem do rozwiązania problemów współczesnej *teorii systemów sterowania* (łączącej w sobie teorię pewności, badania operacyjne, teorię gier, teorię automatów skończonych itd.), *chemii*, która rozwiązuje skomplikowane problemy symulacji procesów kinetyki chemicznej i syntezy organicznej, *cybernetyki* zajmującej się projektowaniem „inteligentnych" robotów przemysłowych, *ekonomii*, w której ważną rolę odgrywa badanie i prognozowanie rozwoju makroekonomicznych wskaźników regionów i całego kraju, w *socjologii* mającej do czynienia z badaniem procesów społecznych, w tym za pomocą metod informatycznych, czy też, bezpośrednio tej części *informatyki*, która jest związana z problemami tworzenia tzw. inteligentnych środowisk i obiektów (np. inteligentnych systemów rozproszonych), lub wreszcie *biologii***,** stosującej metody symulacji matematycznej do badań procesów zachodzących wewnątrz żywych organizmów, na poziomie komórkowym.

Ponadto, do listy tej całkiem słuszne byłoby dodać nauki rozwijające się integralnie, jako symbioza kilku dyscyplin. Można do nich zaliczyć np. *bioinformatykę* – dziedzinę nauki, która została określona, jako nauka XXI wieku, zgłębiająca procesy biochemiczne przyrody ożywionej za pomocą metod informatycznych, czy też *lingwistykę obliczeniową*, której celem jest opracowanie zdań języka naturalnego środkami

informatycznymi oraz niektóre inne dziedziny nowoczesnej nauki.

Nawet tak znacznie okrojona lista przykładów dziedzin nauki, w których *złożone systemy* są przedmiotem wytężonych studiów z wyraźnie zaznaczoną tendencją do wzajemnego przenikania się idei oraz metod, świadczy o aktualności zagadnienia badania fenomenu złożoności zasobami informatycznymi.

Fenomen złożoności w terminach kategorii analizy systemowej określimy jak w [15], [16]. Obiekt (system, proces) uważany jest za złożony, jeśli posiada następujące cechy:

1. Działanie systemu ma charakter stochastyczny.
2. Użyteczny system działa w taki sposób, że jego działania skierowane są na stopniowe zmniejszenie entropii informacyjnej (właściwość tzw. nieentropijności).
3. Właściwości systemu nie są pochodnymi właściwości jego podsystemów.
4. System jest emergentny (ang., *emergent*), czyli ma właściwości, których wstępnie nie da się przewidzieć
5. Cele użytecznego systemu polegają na znalezieniu zadowalających, ale niekoniecznie najlepszych rozwiązań spośród wszystkich hipotetycznie możliwych.

Pierwsza z tych właściwości oznacza, że stany i zachowanie złożonego systemu mogą się zmieniać w czasie nie tylko wskutek działania „wewnętrznych" mechanizmów współdziałania jego składników, lecz również z powodu zmiany stanów środowiska zewnętrznego.

Druga cecha, która jest charakterystyczna dla systemów naturalnych, zdolnych do przeżycia, a także dla systemów sztucznych zdolnych do optymalizacji swoich działań, oznacza umiejętność złożonego systemu do wzmacniania uporządkowania strukturalnego oraz racjonalizacji swojej funkcjonalności. Tak więc, omówione działanie systemu ukierunkowane jest na zmniejszanie nieokreśloności swego zachowania (czyli na zmniejszaniu entropii informacyjnej), gdyż w procesie funkcjonowania systemu rosną jego korzystne właściwości i maleją bezużyteczne a działanie systemu złożonego staje się w przyszłości coraz bardziej przewidywalne.

Trzecia właściwość oznacza, że cechy złożonego systemu nie są prostą sumą cech jego części składowych, ponieważ w przeciwnym razie dzieląc kolejno system na coraz mniejsze elementy, ostatecznie moglibyśmy dojść do tzw. składników „elementarnych" z przewidywalnym zachowaniem. Wówczas zachowanie całego systemu w przyszłości

byłoby problemem jednoznacznie rozwiązywalnym (przynajmniej teoretycznie).

Czwarta cecha oznacza, że dekompozycja złożonego systemu na składniki ma pewną granicę, poza którą składniki nie wykazują już cech systemów złożonych i wskutek tego, nie odzwierciedlają w ogóle rzeczywistych stanów badanego obiektu.

Ostatnia właściwość jest naturalnym rozszerzeniem informatycznego problemu obliczalności na klasie funkcji obliczalnych [17] (ta własność będzie przeanalizowana w następnym podrozdziale).

Uważa się, że różne rodzaje systemów złożonych nie muszą obowiązkowo posiadać wszystkich określonych wyżej cech, mimo to, je również należy zakwalifikować do kategorii złożonych [4], [18], [10]. Pomimo delikatnych nieścisłości definicji systemu złożonego, przyjmijmy ją, jako naszą wersję roboczą.

Aby udowodnić konieczność wykorzystania zasady symulacji indukcyjnej do badania złożonych systemów, przeanalizujemy historyczne korzenie problemu badania omówionych systemów, pochodzące z różnych dziedzin nauki.

1.2 Uzasadnienie konieczności rozwoju koncepcji symulacji indukcyjnej w naukach stosowanych

Korzeni koncepcji symulacji indukcyjnej należy szukać analizując istotę tych obiektów, do badania których przeznaczone są właśnie metody oparte o tę koncepcję, czyli istotę obiektów złożonych. Historycznie rzecz ujmując, badanie złożonych obiektów (systemów, procesów) stało się jednym z centralnych przedmiotów badań różnych nauk stosowanych jeszcze u zarania XX wieku, gdzie leżą podwaliny nauk takich, jak informatyka, teoria automatów, systemy sterowania oraz ekonomia.

Informatyka

W dziedzinie informatyki początek badań związanych z tą problematyką można powiązać z imieniem *Alana Turinga*, który od lat 30. ubiegłego wieku zajmował się badaniami koncepcji modelu komputera, tak zwanej abstrakcyjnej maszyny obliczeniowej (maszyny Turinga[4]). Celem tych badań było wyjaśnienie, jakie funkcje mogą zostać praktycznie obliczone.

Rozpatrując badania Turinga w retrospektywie historycznej, spójrzmy wstecz o kolejnych 30 lat, gdy w roku 1900 wybitny niemiecki matematyk *David Hilbert* na

[4] Stworzony przez Alana Turinga abstrakcyjny model komputera służącego do wykonywania algorytmów, składającego się z nieskończenie długiej taśmy podzielonej na pola.

Ogólnoświatowym Kongresie Matematyków w Paryżu sformułował swoje słynne dwadzieścia trzy zagadnienia, które weszły do annałów historii matematyki, jako „Problemy Hilberta". Ostatnie zagadnienie z tej listy (zagadnienie rozstrzygalności, czyli *Entscheidungsproblem*) brzmi następująco: *czy istnieje algorytm rozstrzygający prawdziwość bądź nieprawdziwość dowolnego zdania, zawierającego arytmetykę liczb naturalnych?* Ostatecznie, pytanie Hilberta sprowadza się do wyjaśnienia, czy istnieją takie właściwości obiektów matematycznych, które ograniczają moc efektywnych procedur udowadniania twierdzeń.

W roku 1930 austriacki matematyk *Kurt Gödel* wykazał, że istnieje efektywna procedura udowodnienia dowolnego prawdziwego zdania w logice pierwszego rzędu *G. Frege'a* i *S.J. Russell'a,* tym nie mniej jednak, logika pierwszego rzędu nie pozwala na wyrażenie zasady indukcji matematycznej. Ponadto, w roku 1931 K. Gödel opublikował uogólniające wyniki swoich badań, ustalające fakt istnienia rzeczywistych granic obliczalności [19]. Zostało udowodnione tzw. *twierdzenie o niezupełności*, w którym pokazano, że w dowolnym formalnym systemie logicznym istnieje zdanie, którego nie można ani udowodnić, ani sprostować w ramach aksjomatów tego systemu. Był to prawdopodobnie pierwszy ściśle udowodniony wynik zwracający uwagę uczonych na ograniczony charakter modeli typu dedukcyjnego.

Stosując ten rezultat do szczególnego problemu obliczalności funkcji, A. Turing doszedł do wniosku, że w rzeczywistości istnieją takie funkcje liczb całkowitych, które nie mogą zostać przedstawione za pomocą żadnego algorytmu, czyli nie mogą być obliczone przez maszynę Turinga (która, jak wiadomo [17], funkcjonuje na zasadach dedukcyjnych).

Napotykając, więc na granicę obliczalności przez maszynę Turinga, naukowcy doszli do wniosku, że istnieje pewna klasa systemów, których badanie za pomocą symulacji matematycznej wymaga rozwoju innych środków, niż te, które były oparte na zasadach dedukcji.

Kolejnym kierunkiem badań wyjaśniających istotę złożonych systemów było tworzenie i rozwój *teorii obliczalności*. Uważa się, że problem jest *nieobliczalny*, jeżeli czas potrzebny do rozwiązania tego problemu rośnie wykładniczo wraz ze wzrostem rozmiaru problemu. W badaniach *A. Cobhama* [20] i *J. Edmondsa* [21] pokazano, że należy rozróżniać wykładniczą i wielomianową złożoność obliczeniową. Tym nie mniej, obie klasy tych problemów zostały zakwalifikowane do kategorii trudnych lub NP-trudnych [17].

W klasycznej teorii obliczeń komputerowych uważa się, że problemy NP-trudne zasadniczo mogą zostać rozwiązane, lecz wymagają tyle czasu komputerowego, że komputer staje się bezużyteczny do poszukiwania rozwiązania niemalże każdego z takich problemów [17] co, jak pokażemy za chwilę, utrudnia rozwiązywanie problemów NP-trudnych z wykorzystaniem „zwykłych" modeli i kryteriów obliczalności opartych na zasadzie dedukcji. Badania zagadnienia obliczalności zostały rozwinięte w pracach *S.A. Cook'a* [22] i *R.M. Karpa*, [23], którzy ustalili fakt istnienia sporych klas problemów kanonicznych poszukiwania kombinatorycznego i formułowania rozważań, które są problemami NP-trudnymi.

W ostatnich latach pokazano sporo przykładów potwierdzających fakt znalezienia całkowicie dopuszczalnych dla celów praktycznych (chociaż, nie idealnych) rozwiązań problemów NP-trudnych za pomocą metod sztucznej inteligencji oraz nowoczesnych zasobów programowych (np., semantyczne sieci WWW i GRID). Klasycznym przykładem takiego NP-trudnego zagadnienia może być problem komiwojażera, który jest rozwiązywalny z dostateczną, dla celów praktycznych, dokładnością z pomocą metod sztucznych sieci neuronowych i metod programowania ewolucyjnego (patrz np. [24], [25]). Od lat 70. ubiegłego wieku trwają też teoretyczne prace nad powiązaniem teorii informacji i mechaniki kwantowej. W przyszłości ograniczenia czasowe obliczalności mogą zniknąć. Dzięki komputerom kwantowym budowanym na bazie tych teorii, znacznie skrócony zostanie czas skomplikowanych obliczeń. Na obliczenie trasy komiwojażera po Szwecji liczącej 24978 miejscowości współczesny komputer potrzebuje obecnie około 85 lat. Komputer kwantowy niezależnie od ilości danych rozwiązuje ten problem w ułamku sekundy. To jednak na razie wizja przyszłości [26].

Przykład problemów NP-trudnych wskazuje, więc na to, że klasa złożonych systemów wymaga innego sposobu symulacji zachowania się tych systemów niż metody oparte na zasadach dedukcji, co powoduje zmianę naszego nastawienia do „zasadniczej obliczalności" lub „precyzji rozwiązania" poszczególnych problemów.

Teoria automatów

Badania funkcji obliczalnych za pomocą metod matematycznych okazały się owocne i zapoczątkowały rozwój teorii automatów, która leży u podstaw współczesnej teorii komputerów. Epoka powstania tej teorii, czyli lata 40.-50. ubiegłego wieku, charakteryzowała się konstruowaniem najprostszych modeli komputerów, mających skończoną ilość ewentualnych stanów. Systemy te nazywamy obecnie *automatami skończonymi*.

Próby uogólnienia teorii automatów na inne dziedziny nauki w celu tworzenia teorii złożonych systemów zainicjowały pojawienie się w tym samym czasie cybernetyki. Jak wiadomo z pionierskich prac *Norberta Wienera* [4] i *Rossa Ashby'ego* [18], teoria automatów nabrała zupełnie innego zabarwienia i wywarła potężny wpływ na rozwój kilku nauk, w tym sztucznej inteligencji.

Naukowcy, idąc dalej w kierunku formalizacji teorii automatów, doszli do wniosku, że myślenie człowieka oparte jest na pewnych konstrukcjach językowych. Ponieważ dowolne stwierdzenia człowieka można wypowiedzieć w postaci fraz zawierających się w skończonym zbiorze reguł semantycznych, to ustalenie regularności budowy tych zbiorów bądź ich nieregularności jest drogą do zrozumienia algorytmów myślenia człowieka i odpowiednio drogą do zrozumienia algorytmów podejmowania optymalnych decyzji systemów inteligentnych.

Swego czasu do spektrum idei teorii złożonych systemów dodano zaskakujący wniosek ze strony nauki „nietechnicznej", a mianowicie lingwistyki matematycznej. Zwrotnym momentem w rozwoju tej dziedziny było seminarium w Massachusetts Institute of Technology w USA (wrzesień 1956r.), na którym lingwista *Noam Chomsky* przedstawił prezentację pt. „Three Models of Language". Zostały w niej opisane podstawy nowej w tym czasie teorii *gramatyk formalnych*, które nie będąc modelami maszyn obliczeniowych w dosłownym znaczeniu tego słowa, są ściśle powiązane z automatami abstrakcyjnymi, a później posłużyły do rozwoju niektórych najważniejszych komponentów oprogramowania, np. składników kompilatorów. Wspomniane seminarium wyróżniło się jeszcze dwoma referatami – *Allena Newell'a* i *Herberta Simon'a* „The Logic Theory Machine" [27] oraz *Georga Millera* „The Magic Number Seven Plus or Minus Two" [28]. Został w nich przeanalizowany problem tworzenia modeli złożonych systemów podobnych do mózgu człowieka i przeznaczonych do rozwiązywania problemów z dziedziny psychologii, zapamiętywania informacji rozmytej, opracowania zdań i pojęć języka naturalnego oraz myślenia logicznego w warunkach niepełnej i rozmytej informacji o obiekcie badań. Wśród współczesnych prac, w których rozwija się idee Chomskiego, Newella, Simona i Millera, można wspomnieć pracę *Franza Baadera* i współautorów [29] poświęconą rozwojowi *logik deskrypcyjnych*, jako podstaw wnioskowania logicznego również z wykorzystaniem ontologii[5].

[5] Ontologia w sensie informatycznym to formalna reprezentacja pewnej dziedziny wiedzy, na którą składa się zapis zbiorów pojęć (ang. *concept*) i relacji między nimi.

Z wykonanych badań wynikał istotny wniosek, że metody analizy dowolnych skomplikowanych systemów powinny przypominać program komputerowy. Działanie tego programu powinno być uporządkowane w postaci ciągu instrukcji i zawierać mechanizmy opracowania informacji pozwalające realizować *funkcję kognitywną,* na podstawie której może być realizowana zdolność systemu złożonego do uzyskiwania nowych informacji o obiekcie badań i *adaptacji* jego zachowania w stale zmieniających się warunkach. Do słownika naukowego wszedł termin „*adaptacja*", jako charakterystyka systemu zdolnego do samoorganizacji swoich działań.

Znamienną pracą, która wywarła decydujący wpływ na rozwój „inteligentnych" automatów zdolnych do adaptacji swoich działań, była publikacja *Jacoba Tsypkin'a* [9]. W późniejszym czasie idee Tsypkina zostały uogólnione i rozwinięte na klasy systemów złożonych dowolnej natury (tzn. systemy techniczne, ekonomiczne, socjologiczne) w pracach badaczy szkoły naukowej MIT *Johna Hollanda* [10] i *Johna Kozy* [30]. Po raz pierwszy zwrócono w nich uwagę na adaptacyjne właściwości dowolnych złożonych systemów, jako zasadniczą cechę tych systemów odróżniającą ją od „zwykłych" systemów, u podstaw których leżą modele alternatywne. Idee te były rozwijane w pracach poświęconych teorii systemów sterowania na podstawie zasady indukcyjnej.

Teoria systemów sterowania

Mówiąc o teorii systemów sterowania można wyodrębnić trzy najbardziej charakterystyczne okresy jej rozwoju [9]: okres determinizmu, stochastyczności oraz adaptacyjności.

W okresie determinizmu, zarówno równania opisujące stany obiektów, jak i zewnętrzne oddziaływania były określone jednoznacznie. W dziedzinie problemów liniowych pełna określoność pozwalała korzystać z klasycznego aparatu matematycznego opartego na teorii równań różniczkowych i całkowych, co pozwalało rozwiązywać szerokie spektrum aktualnych w tym czasie problemów. Rozwiązanie tych problemów znacznie ułatwiała zasada superpozycji[6].

W kwestii problemów nieliniowych otrzymano istotne wyniki dotyczące analizy i syntezy systemów automatycznych.

W teorii systemów stochastycznych uważa się, że wewnętrzne oddziaływania są ciągłe w czasie i nie mogą zostać wcześniej określone. Często przypuszczenia te dotyczyły również współczynników równań obiektów sterujących. Dlatego powstała

[6] Zasada superpozycji mówi, że pole (siła) pochodzące od kilku źródeł jest wektorową sumą pól (sił), jakie wytwarza każde z tych źródeł.

konieczność rozwoju i zastosowania nowego podejścia, uwzględniającego probabilistyczny charakter zewnętrznych sygnałów oraz równań. To podejście oparte jest na wykorzystaniu wiedzy o statystycznym charakterze funkcji przypadkowych.

Otrzymane w teorii sterowania automatycznego metody i wyniki były bezpośrednio stosowane do systemów sterowania automatycznego w warunkach dostatecznej informacji. Jednak obecnie, w dziedzinie systemów sterowania automatycznego priorytetowym zagadnieniem jest symulacja zachowania się systemów w rozmaitych warunkach, gdy równania obiektów podlegających sterowaniu i zewnętrzne sygnały (oddziaływania) nie są dokładnie określone, co więcej, niekiedy nie jesteśmy nawet w stanie określić ich w sposób eksperymentalny przed rozpoczęciem procesu sterowania.

Napotykamy, więc na początkową nieokreśloność. Wszystko to komplikuje sterowanie tymi obiektami, lecz nie powoduje, że sterownie to jest w praktyce niemożliwe. Świadczy to o nadejściu nowego okresu w teorii sterowania – okresu, w którym kluczowym zagadnieniem jest tworzenie metod projektowania i symulacji zachowania się systemów skomplikowanych zgodnie z zasadami *samokształcenia* i *adaptacji*.

Pierwszym etapem każdego z omówionych okresów, jego głównym zagadnieniem, było utworzenie skutecznych metod analizy systemów automatycznych i badanie ich właściwości. Naturalnie, niezbędnym wymogiem było utworzenie metod syntezy systemów w pewnym sensie optymalnych. W dziedzinie sterowania automatycznego zagadnienie optymalności rozwiązań stało się, więc jednym z najistotniejszych. Znaleziono efektywne, z teoretycznego punktu widzenia, sposoby rozwiązań, bazujące na *zasadzie maksimum Pontriagina* [31] i *metodzie programowania dynamicznego Bellmana* [32].

Omówione metody praktycznie wyczerpały deterministyczne problemy analizy i syntezy systemów oparte na modelach alternatywnych i utorowały drogę do badań systemów stochastycznych opartych na modelach imperatywnych będących też jednym z centralnych przedmiotów badań teorii ekonomii.

Teoria ekonomii

Teoria ekonomii ma również istotny wkład w rozwój teorii systemów złożonych. Jak wiadomo, ekonomia, jako nauka pojawiła się w roku 1776, kiedy to szkocki filozof *Adam Smith* opublikował swoją książkę „An Inquiry into the Nature and Causes of the Wealth of Nations" („Badania nad naturą i przyczynami bogactwa narodów"), w której określił stosunki ekonomiczne między ludźmi jako naukę, korzystając z idei, że dowolną

działalność można rozpatrywać jako aktywność osobnych agentów dążących do zmaksymalizowania swego stanu ekonomicznego.

Jako kryterium optymalizacji klasycznych problemów najczęściej wybierana jest maksymalizacja dochodów w ciągle zmieniających się warunkach konkurencji, czyli innymi słowy, ekonomia, jako nauka odpowiada na pytanie, w jaki sposób ludzie dokonują wyboru, który doprowadzi ich do najwłaściwszych dla nich rezultatów. Matematyczne potraktowanie pojęcia „najwłaściwsze rezultaty", czyli pojęcia użyteczności pochodzi od *Leona Walrasa* i zostało sprecyzowane przez *Franka Ramsey'ego* [33]. Później teoria ta była udoskonalona i uporządkowana przez *Johna von Neumana* i *Oskara Morgensterna* w znakomitej książce [34].

U podstaw ekonomii leży *teoria decyzji*, która łączy w sobie teorię prawdopodobieństwa oraz teorię użyteczności i tworzy pełną infrastrukturę potrzebną do podejmowania decyzji w warunkach nieokreśloności. Pod pojęciem nieokreśloności rozumiemy sytuację, gdy środowisko, w którym funkcjonuje pewna osoba podejmująca decyzję, najbardziej adekwatnie może być przedstawione jedynie za pomocą prawdopodobieństw. W dużych środowiskach ekonomicznych takie opisy przewidują, że nie każdy agent musi brać pod uwagę posunięcia innych agentów. A w niewielkich środowiskach ekonomicznych sytuacja bardzo przypomina grę, ponieważ działania jednego gracza mogą mieć istotny wpływ na działania innego. Z punktu widzenia użytkowników wychowanych na zasadach determinizmu, teoria gier Neumana i Morgensterna oferuje niekiedy dość niespodziewane rozwiązania. Np., w niektórych grach rozsądny agent musi działać w sposób przypadkowy, czy w sposób, który dla innych graczy wygląda na przypadkowy. Wówczas działania takiego gracza nie mieszczą się w ramach determinizmu, czyli w ramach modeli dedukcyjnych.

W ekonomii więc, dzięki wprowadzeniu teorii gier von Neumana i Morgensterna, podobnie jak w innych przeanalizowanych wyżej dziedzinach nauki, zamiast zasad determinizmu, silną pozycję zajęła koncepcja apriorycznej nieokreśloności przyszłych stanów systemów wykazujących zachowania o skomplikowanym charakterze.

Można tu zauważyć pewne podobieństwo pomiędzy omówionym rozwiązaniem z teorii gier a strategią, która została przyjęta w rozwiązaniu zagadnień za pomocą metod programowania ewolucyjnego. Główną rolę w tych strategiach odgrywa czynnik przypadkowego wyboru populacji rodzicielskich, punktów krzyżowania i punktów mutacji. Właśnie dzięki wykorzystaniu czynnika przypadkowości, poprzez wykorzystanie metod programowania ewolucyjnego bardzo często można znaleźć

rozwiązanie tych problemów, których nie udaje się rozwiązać za pomocą tradycyjnych metod.

Trzecia kwestia, która została przedstawiona powyżej, jako przedmiot badań teorii ekonomii, nabrała istotnego znaczenia w *badaniach operacyjnych*. W pracy *Richarda Bellmana* [32] została sformalizowana pewna klasa sekwencyjnych zagadnień podejmowania decyzji, które znane są jako *procesy podejmowania decyzji Markowa* (Markov Decision Processes – MDP), stosowanych w systemach stochastycznych. Prawdopodobnie po wprowadzeniu do teorii podejmowania decyzji łańcuchów Markowa ostatecznie umocniony został pogląd o systemach złożonych, jako kategorii systemów stochastycznych, czyli systemach, których zachowanie się zasadniczo nie może być symulowane z wykorzystaniem z góry określonych modeli (modeli opartych na zasadzie dedukcyjnej).

Teoria analizy systemowej

Charakterystyczne, że prace w dziedzinie badań operacyjnych wywarły istotny wpływ nie tylko na rozwój teorii ekonomii, lecz również na rozwój niektórych innych dziedzin nauki, w tym informatyki. Obecnie w informatyce mamy do czynienia z tzw. *agentami programowymi*, w szczególności z *agentami racjonalnymi* działającymi w imieniu użytkownika oraz zdolnymi do optymalizacji swoich działań na bieżąco w trakcie wykonywania zadania [24]. Jakiś czas temu wykorzystanie tradycyjnych metod do tworzenia agentów racjonalnych napotykało na przeszkody związane z konceptualną nieokreślonością inteligentnych właściwości agentów oraz zastosowaniem niedoskonałych metod optymalizacji procesów podejmowania decyzji w dużych parametrycznych przestrzeniach poszukiwania. Klasycznym przykładem tej niedoskonałości może być metoda programowania dynamicznego R. Bellmana, która została sklasyfikowana jako „przekleństwo dużego wymiaru".

Omówione okoliczności zmusiły do szukania innych metod rozwiązywania problemów podejmowania decyzji przez agentów i przede wszystkim innych sposobów oceny optymalności tych decyzji. Z czasem stało się jasne, że podejmowanie decyzji przez agentów programowych działających w imieniu użytkowników należy do zagadnienia, które obejmuje problemy klasyfikacji danych o symulowanym systemie, prognozie i optymalizacji działań w rozmytych parametrycznych przestrzeniach rozwiązań, przy nieściśle określonych warunkach [35]. Niektórzy naukowcy uważają, że problem tworzenia agentów racjonalnych zintensyfikował dociekania w dziedzinie sztucznej inteligencji [24], [25], w której mózg człowieka i mechanizmy myślenia stają

się idealnym obiektem badań. Tak więc, ze względu na to, że mózg człowieka posługuje się innymi strategiami oceny jakości rozwiązania problemów niż te, z których zwykle korzystamy w tradycyjnych systemach technicznych, w naturalny sposób powstała kwestia wyboru kryteriów rozwiązania problemów w sztucznych systemach inteligentnych.

Herbert Simon udowodnił [36], że najlepszy opis faktycznego zachowania się człowieka dają modele bazujące na zasadzie *satysfakcji*, czyli modele, na podstawie których system jest w stanie podejmować takie decyzje, które określa się jako „dopuszczalne", chociaż być może nie zawsze najlepsze, spośród wszystkich możliwych. Porównując więc ten wynik z przedstawionym w poprzednim podrozdziale, dotyczącym niedoskonałości definicji złożonych systemów na podstawie stopnia obliczalności lub precyzji rozwiązań, można dojść do wniosku, że pojęcie złożonego systemu musi określać inny cel działania owego systemu, niż liczbowa dokładność obliczeń.

Specyficzne, że przyjęcie takiego punktu widzenia zmieniło radykalnie stosunek badaczy do kwestii: jakiego typu zachowanie systemu złożonego należy uważać za racjonalne, a także, które z rozwiązań znalezionych przez agentów można uznać za optymalne. Za rezultaty otrzymane w tej dziedzinie Herbert Simon otrzymał w roku 1978 Nagrodę Nobla z ekonomii.

Przedstawiona wyżej analiza daje podstawy twierdzić, że we wszystkich naszych przykładach mamy do czynienia z systemami wykazującymi te czy inne cechy systemów złożonych. Ogólny wniosek jest następujący: *symulacja zachowania się tych systemów, w tym rozwiązanie problemów prognozowania ich stanów w przyszłości wymaga rozwoju i wykorzystania klasy imperatywnych modeli tworzonych na zasadach indukcji*.

Przystępując do symulacji i prognozy zachowania się obiektów złożonych, niezbędne jest ocenić, jakie właściwości mają modele i oprogramowanie przeznaczone do symulacji owych obiektów. Ogólną odpowiedź na to pytanie znaleźć można w tezie A.Turinga [37], która mówi, że system adekwatnie odzwierciedlający zachowanie się pewnego obiektu jest tak złożony, jak złożony jest obiekt podlegający symulacji za pomocą tego systemu. Zagadnienie symulacji zachowania się obiektów na podstawie analizy informacji o ich zachowaniu rozwiązują obecnie *systemy wspomagania decyzji* (ang., *Decision Support System, DSS*) [38]. Przechodzimy, więc do wniosku, że problem badany w danej rozprawie należy do klasy problemów ewolucyjnej analizy danych i procesów rozwiązywanych za pomocą współczesnych DSS.

1.3 Prognoza, jako zagadnienie ewolucyjnej analizy danych i procesów w systemach wspomagania decyzji

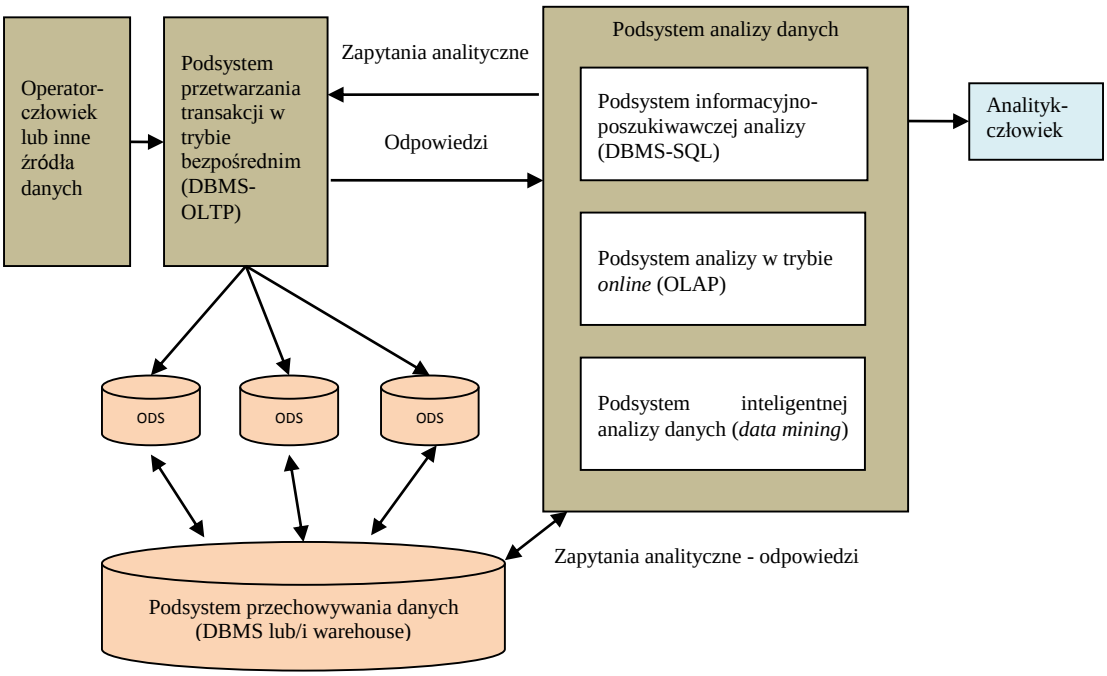
Jak wiadomo [39], u podstaw zagadnienia analizy danych i procesów leżą metody eksploracji danych, czyli zgłębiania danych (ang., *data mining*). Według definicji, przedstawionej w *Encyklopedii Komputerowej Microsoftu* [40], *data mining* określany jest jako proces identyfikacji komercyjnie użytecznych wzorców lub wzajemnych związków w bazie danych, bądź innych komputerowych składnicach informacji, dzięki użyciu zaawansowanych narzędzi statystycznych i niestatystycznych. Według *Wikipedii* natomiast, *zgłębianie danych* to proces analityczny, przeznaczony do badania dużych zasobów danych (zazwyczaj powiązanych z zagadnieniami gospodarczymi lub rynkowymi) w poszukiwaniu regularnych wzorców oraz systematycznych współzależności pomiędzy zmiennymi, a następnie do oceny wyników poprzez zastosowanie wykrytych wzorców do nowych podzbiorów danych. Finalnym celem *data mining* jest najczęściej przewidywanie stanów i działania pewnych obiektów będących przedmiotem badań (np. zachowań klientów, wielkości sprzedaży, prawdopodobieństwa utraty klientów, przewidywanie i usunięcie przyczyn awarii sprzętu itp.), dlatego rozwój metod *predykcyjnego data mining* jest aktualnym problemem tej części współczesnej informatyki, która jest powiązana z otrzymywaniem wiedzy o złożonych obiektach i procesach na podstawie analizy informacji przedstawionej w dużych zbiorach danych.

Jak wspomniano wyżej, algorytmy eksploracji danych są często realizowane jako część składowa systemu wspomagania decyzji DSS, którym jest zestaw programów i związanych z nimi danych przeznaczony do pomocy w analizie i podejmowaniu decyzji przez analityków-fachowców w danej dziedzinie nauki, inżynierii, ekonomii itd. Koncepcja DSS jest pewnym uogólnieniem koncepcji *systemu informacyjnego zarządzania* (ang., *Management Information System, MIS*) oraz koncepcji *systemu informacyjnego kierownictwa* (ang., *Executive Information System, EIS*). Poza tym, DSS dostarcza więcej wsparcia w formułowaniu decyzji niż MIS lub EIS, ponieważ w skład DSS wchodzi moduł hurtowni danych (której idea rozszerza ideę bazy danych), „język" używany do formułowania problemów i pytań oraz oprogramowanie symulujące zachowywanie obiektu badań i testujące wygenerowane modele.

Główne zagadnienie DSS polega na tym, że pozostawia on analitykom-ekspertom narzędzia do wykonania analizy danych i procesów. Przy czym nie gwarantuje najlepszego rozwiązania, a jedynie udostępnia analitykom dane w postaci odpowiedniej

do badania i analizy, bowiem DSS przeznaczone są do symulacji obiektów złożonych, których zachowanie nie jest ściśle przewidywalne (patrz podrozdział 1.1). Właśnie dlatego, systemy takie zapewniają funkcjonalność *wspomagania decyzji*, ale nie samodzielnego podejmowania decyzji i nie zastępują analityków-ludzi. Dla zwiększenia skuteczności wykorzystania DSS, badania ostatnich lat dotyczące ich tworzenia skierowane są głównie na nadanie tym systemom pewnych funkcji analityka-człowieka, czyli funkcji inteligentnych. Staje się to możliwe dzięki wprowadzeniu do tych systemów metod sztucznej inteligencji, teorii analizy systemowej, współczesnych technologii budowy hurtowni dużych zbiorów danych, nowoczesnych metod analitycznych oraz technologii komputerowego opracowania informacji.

Zaproponowano różne architektury systemów wspomagania decyzji [38], które w uogólnionej postaci można przedstawić tak, jak pokazano na Rys. 1.1.



Rys. 1.1 Uogólniona architektura systemu wspomagania decyzji (DSS)
(DBMS = data management system, system zarządzania danymi
OLTP = on-line transaction processing, przetwarzanie transakcyjne
ODS = online data sources, źródła danych dostępne w trybie bezpośrednim
Warehouse = hurtownia danych,
SQL = structured query language, strukturalny język zapytań
OLAP = on-line analytical processing, przetwarzanie analityczne
Źródło: Opracowanie własne

W zasadzie dane wejściowe mogą nadchodzić do systemu od operatorów-ludzi, czujników, sieci komputerowej, bądź bazy danych. Bazy danych obsługują duże zbiory danych i odpowiednio dużą liczbę transakcji składających się z zapytań i aktualizacji

zawartości danych. Tego rodzaju korzystanie z bazy danych nazywane jest *bezpośrednim przetwarzaniem transakcyjnym* (ang., *online transaction processing*, OLTP). Opisane procesy realizowane wewnątrz DSS, wykonywane są przez podsystem przetwarzania transakcji w trybie bezpośrednim (DBMS-OLTP). Podsystem odbiera nadchodzące na jego wejściu dane i aktualizuje pliki w systemie zarządzania bazą danych. Ta ostatnia jest podstawowym składnikiem podsystemu *hurtowni danych* (ang., *warehouse*) [41].

W proponowanych w literaturze realizacjach DSS przewidywane są różne sposoby dostępu podsystemu analizy danych do źródeł danych wejściowych – albo bezpośrednio do OLTP, albo za pośrednictwem hurtowni danych, gdzie te dane przeszły już wstępną filtrację i uporządkowanie [38], [39], [40]. Podsystem analizy danych może być tworzony na podstawie kilku składników, przeznaczonych do rozwiązywania następujących zagadnień:

- *poszukiwania informacji* (poszukiwanie niezbędnych danych, przy czym charakterystyczną cechą takiej analizy jest wykonanie wcześniej przygotowanych zapytań),
- *analizy danych* – rozwiązywane w czasie rzeczywistym (DSS wykonuje grupowanie i uogólnienie danych przedstawionych w postaci zrozumiałej dla analityka),
- *inteligentne* (poszukiwanie funkcjonalnych i logicznych regularności w przechowywanych zbiorach danych, tworzenie modeli i reguł, które tłumaczą znalezione regularności i/lub realizują predykcję rozwoju tych czy innych procesów z pewnym prawdopodobieństwem).

Końcowym składnikiem DSS jest analityk-człowiek, czyli osoba (lub kilka osób) generująca hipotezy na podstawie swojej wiedzy w danej dziedzinie przedmiotowej. Wiedzą, jednak, dysponuje nie tylko analityk, bowiem wiedza ukryta jest w danych akumulowanych w hurtowniach danych DSS.

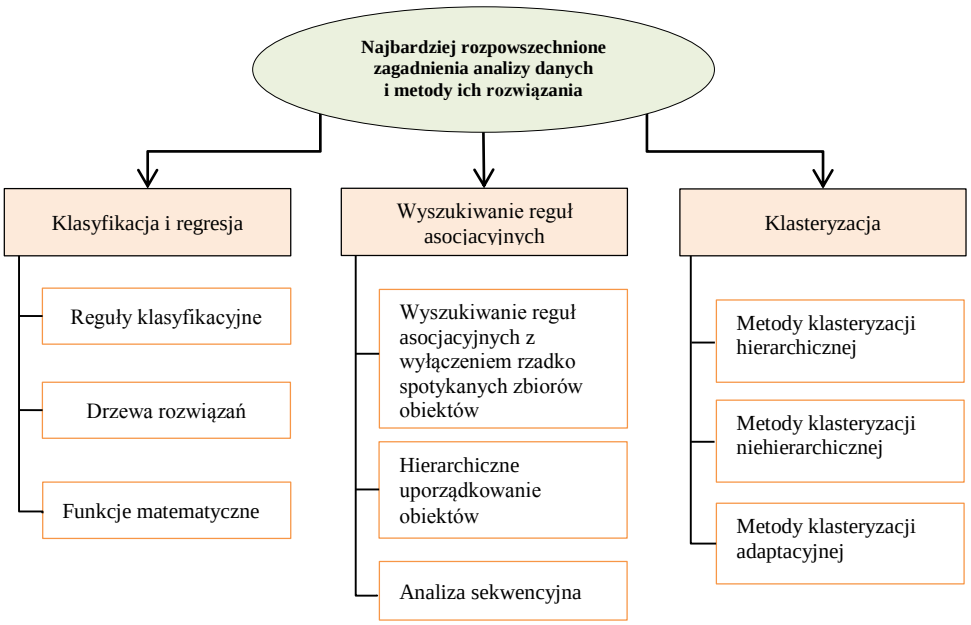
Ze względu na tera- i petabajty danych przechowywanych w hurtowniach danych DSS, eksploracja interesującej analityka wiedzy z przechowalni danych poprzez proste przeglądanie danych staje się zagadnieniem praktycznie niemożliwym, tym samym staje się praktycznie niemożliwym korzystanie z tej ukrytej wiedzy do wygenerowania hipotez stosowanych dla przyszłych stanów obiektu podlegającego badaniom. Istnieje, więc wysokie prawdopodobieństwo pomijania hipotez, które mogłyby przynieść wiele korzyści w rozwiązaniach konkretnych problemów.

Jest całkiem zrozumiałe, że dla wyjawienia „ukrytej” wiedzy o właściwościach

badanych obiektów, na podstawie której można byłoby tworzyć hipotezy o przyszłych stanach obiektów, należy korzystać z metod analitycznego badania danych w olbrzymich składnicach danych. Co więcej, analityk powinien mieć do swojej dyspozycji pewne *sformalizowane mechanizmy doboru najbardziej istotnych danych*, które mają największy wpływ na stany badanego obiektu.

W związku z tym, stosowne będzie przypomnieć jeszcze jedną definicję kierunku eksploracji danych, proponowaną przez jednego z twórców dziedziny eksploracji danych – Grzegorza Piateckiego-Szapiro [42]: „Data mining to badanie i odkrywanie przez „maszynę" (algorytmy, zasoby sztucznej inteligencji) ukrytej w surowych danych wiedzy, która wcześniej nie była znana, wiedzy nietrywialnej, praktycznie korzystnej i dostępnej do interpretacji przez człowieka."

Obecnie można wyodrębnić szereg zagadnień, które rozwiązywane są za pomocą metod eksploracji danych, wśród których zasadniczego znaczenia nabrały zagadnienia i metody przedstawione na Rys. 1.2 (rysunek tworzono na podstawie materiałów książek [38]- [42]).



Rys. 1.2 Najbardziej rozpowszechnione zagadnienia analizy danych i metody do ich rozwiązania
Źródło: Opracowanie własne na podstawie [38]-[42]

Zagadnienie *klasyfikacji* polega na podziale zbioru obiektów na klasy. Jest to zagadnienie określenia wartości jednego lub kilku parametrów analizowanego obiektu (tzw. zmiennych zależnych), na podstawie wartości innych parametrów (tzw. zmiennych niezależnych). Do tej grupy można zaliczyć np. problem oceny zdolności płatniczej klientów banku na podstawie analizy danych o klientach (średniego poziomu

wynagrodzenia, wieku klienta, posiadania nieruchomości, itd.). Jeśli wartościami zmiennych zależnych i niezależnych są liczby rzeczywiste, to problem nazywamy problemem regresji. Przykładem takiego problemu może być ocena dopuszczalnej kwoty kredytu, który może zostać udzielony przez bank pewnemu klientowi.

Istota zagadnienia *wyszukiwania reguł asocjacyjnych* polega na wskazaniu interesujących zależności obiektów w dużym zestawie takich zbiorów. W rzeczywistości, zagadnienie to jest przypadkiem szczególnym zagadnienia klasyfikacji.

Wreszcie, zagadnienie *klasteryzacji* (*analizy skupień*) polega na podziale badanego zbioru obiektów na grupy obiektów „podobnych", nazywanych klastrami.

W zależności od dziedziny zastosowań, omówione zagadnienia eksploracji danych dzielimy na dwie kategorie (którym odpowiadają dwie zasadnicze kategorie modeli):

- *zagadnienia opisowe* (*descriptive problems*) rozwiązywane z wykorzystaniem modeli opisowych,
- *zagadnienia prognostyczne* (*predictive problems*) rozwiązywane z wykorzystaniem modeli prognostycznych.

W zagadnieniach opisowych, główna uwaga skoncentrowana jest na ulepszeniu rozumienia analizowanych informacji (ich wzajemnej zależności, regularności zmian itd.). Kluczowe wymaganie dotyczące metod rozwiązujących te zagadnienia to łatwość ich zrozumienia i przejrzystość dla analityka. Do tej kategorii należą problemy klasteryzacji i wyszukiwania reguł asocjacyjnych.

W zagadnieniach prognostycznych należy ocenić przyszłe stany i/lub przyszłe zachowanie się obiektów na podstawie analizy ich stanów i funkcjonowania w przeszłości. Do tej kategorii zagadnień można zaliczyć klasyfikację i regresję, lub w niektórych przypadkach zagadnienie wyszukiwania reguł asocjacyjnych, jeśli te ostatnie używane są przez analityków do przewidywania pewnych zdarzeń w przyszłości.

Gdy weźmiemy pod uwagę sposób budowy modeli analizy danych, problemy te można podzielić na oparte na *uczeniu z nauczycielem* (ang., *supervised learning*) i/lub *uczeniu bez nauczyciela* (ang., *unsupervised learning*). W pierwszym przypadku (supervised learning), na początku tworzymy model analizowanych danych, innymi słowy tworzymy klasyfikator uczący, a na drugim etapie wykonujemy testowanie tego klasyfikatora. Do kategorii zagadnień *supervised learning* zaliczono zagadnienia klasyfikacji i regresji [39], [42].

Ważną rolę w doborze modeli rozwiązywania zagadnień klasyfikacji i regresji, a także oceny zdolności klasyfikatorów do poprawnego rozwiązania owych problemów

odgrywają *kryteria*, których ilość i sposób przedstawienia zależą od specyficznych cech konkretnego zadania.

Unsupervised learning łączy w sobie głównie zagadnienia, które oparte są na modelach opisowych. Rzeczywiście, jeśli zachodzą pewne regularności, to poprawnie tworzony model powinien umieć je przedstawić w sposób sformalizowany, dlatego niestosownie jest mówić o jego uczeniu. Zagadnienia tej kategorii mogą być rozwiązane bez wiedzy o danych podlegających analizie. Tym własnościom odpowiadają zagadnienia klasteryzacji i wyszukiwania reguł asocjacyjnych.

W przypadku zagadnień opartych na metodach *supervised learning*, takich jak problemy klasyfikacji i regresji, w tym problemów prognozy, faza uczenia klasyfikatora nabiera zasadniczego znaczenia, ponieważ jedynie na podstawie uczenia (czyli analizy zachowań obiektu w przeszłości) klasyfikator będzie w stanie ocenić ewentualne zachowanie się obiektu w przyszłości. Zwróćmy uwagę na pewne podobieństwo problemu zagadnienia prognozy do problemu ekstrapolacji funkcji.

Przedstawiona analiza problemów eksploracji danych pozwala wyjaśnić, jakie miejsce zajmuje w tej dziedzinie informatyki prognoza. Przedstawimy dalej ogólną klasyfikację metod przeznaczonych do analizy danych oraz procesów i na podstawie tej analizy spróbujemy określić pewne zasady metod rozwiązania problemów predykcji.

W zależności od teoretycznych podstaw, na których bazują metody inteligentnej analizy danych i procesów, metody te można podzielić na dwie obszerne klasy:

- podstawowe, oparte na technice przeglądania obiektów;
- budowane z wykorzystaniem zasobów sztucznej inteligencji (takich, jak sztuczne sieci neuronowe, algorytmy genetyczne oraz logika rozmyta) i/lub metod samoorganizacji heurystycznej.

Metody pierwszej grupy oparte są na rozmaitych realizacjach zasady przeglądania obiektów nadchodzących na wejściu systemu DSS. Prosty przegląd wszystkich analizowanych obiektów wymaga $O(2^N)$ operacji, gdzie N to ilość obiektów. Wraz ze zwiększeniem ilości obiektów, ilość obliczeń rośnie wykładniczo, co w przypadku ich dużej ilości, czyni rozwiązanie problemu praktycznie niemożliwym. Aby zmniejszyć złożoność obliczeniową, dla tych algorytmów używane są różnego rodzaju heurystyki. W niektórych przypadkach ilość obliczeń udaje się zmniejszyć do zależności prawie liniowej (od ilości obiektów). Równocześnie, udowodniono, że zależność złożoności obliczeniowej od ilości atrybutów, z reguły pozostaje wykładnicza [43]. Dlatego algorytmy wyszukiwania liniowego i ich modyfikacje, używane są przeważnie do

rozwiązywania stosunkowo prostych problemów.

Do metod podstawowych należą również te, które bazują na metodach statystycznych. Ich główna idea sprowadza się do analizy korelacyjnej, regresyjnej oraz innych typów analizy statystycznej. Główną wadą tego podejścia jednak jest to, że realizowane jest w nim uśrednienie wartości danych, co prowadzi do utraty informacyjnej wartości niektórych danych. To z kolei prowadzi do zmniejszenia wartości zdobytej wiedzy.

Ze względu na omówione okoliczności, dla celów rozwiązania problemów prognozy, bardziej skuteczne są metody drugiej grupy, na ogólnej charakterystyce których skupimy więcej uwagi.

Sposób tworzenia modeli w metodach opartych na *sztucznych sieciach neuronowych* odpowiada technikom dwuetapowego procesu – uczenia i testowania modeli, zarówno z wykorzystaniem nauczyciela, jak i bez [44]. Do istotnych zalet sztucznych sieci neuronowych należy zaliczyć to, że są one zdolne, po pierwsze, do wyodrębnienia najbardziej istotnych zmiennych i odrzucenia zmiennych drugorzędnych, a po drugie, do samouczenia (w tym sensie, że użytkownik nie musi samodzielnie dobierać wag gałęzi sieci neuronowej).

Należy jednak podkreślić, że podczas tworzenia takich modeli powstają trudności związane z doborem architektury sieci neuronowej i interpretacji wyników symulacji, ponieważ sieć neuronowa wygląda jak „czarna skrzynka" działająca w sposób niewidoczny dla użytkownika. Sieć neuronowa nastrojona na rozwiązanie pewnego problemu, wykazuje regularności istniejące w danych, ale nie odpowiada na pytanie, w jaki sposób te regularności zostały otrzymane. W pewnych przypadkach jest to zasadnicza przeszkoda dla wykorzystania sztucznych sieci neuronowych w systemach symulujących. Np., firmy, na których ciąży wysoka odpowiedzialność za generowane przez nich rekomendacje (firmy pracujące w obszarze badań strategicznych, makroekonomicznych itd.) nie akceptują wygenerowanych przez system wyników, jeśli sposób otrzymania tych wyników nie jest ściśle uzasadniony.

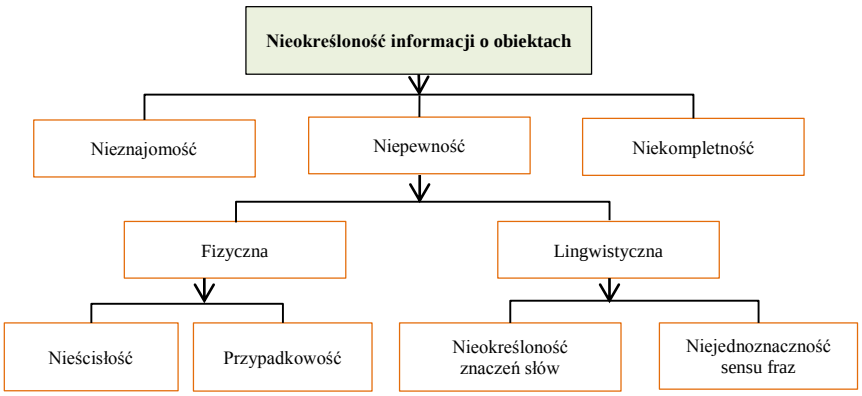
Metody oparte na *algorytmach genetycznych* (AG) należą do uniwersalnych metod optymalizacji pozwalających rozwiązywać problemy kombinatoryczne z ograniczeniami i bez ograniczeń o różnym stopniu złożoności. Jednak do wad algorytmów genetycznych można zaliczyć problemy nieokreśloności wyboru punktów krzyżowania i mutacji. W pracach badaczy poświęconych problemom AG można znaleźć raczej rekomendacje eksperckie niż konkretne algorytmy i kryteria do rozwiązania omówionych kwestii (patrz,

np. [45]).

W licznych pracach badawczych udowodniono, że skuteczność algorytmów genetycznych istotnie się zwiększa, jeśli są one integrowane z innymi metodami i technologiami [46]. Takie systemy symulujące nazywamy mieszanymi lub hybrydowymi. Wadą metod AG pozostaje problem doboru najbardziej znaczących parametrów podstawowych działań, a wady omówionych systemów hybrydowych łączących w sobie techniki AG i sieci neuronowych są takie, jak w przypadku sieci neuronowych.

Metody oparte na *logice rozmytej* bazują na przypuszczeniu, że, po pierwsze, informacja o badanych obiektach, ma pewną nieokreśloność a, po drugie, że mamy do czynienia z zagadnieniem na tyle skomplikowanym, że budowa adekwatnego modelu w terminach „ścisłych" modeli numerycznych staje się praktycznie niemożliwa.

Na Rys. 1.3 przedstawiona jest klasyfikacja głównych źródeł nieokreśloności, która rozjaśnia nieco zakres zastosowań metod logiki rozmytej (klasyfikacja przedstawiona jest na podstawie materiałów z artykułu [47]).



Rys. 1.3 Klasyfikacja głównych źródeł nieokreśloności informacji o badanych obiektach
Źródło: Opracowanie własne na podstawie [47]

Nieznajomość jest tym rodzajem nieokreśloności powstającej podczas badań obiektów i procesów, które są, jak na razie, nieznane przez badaczy albo wskutek niedostatecznej ich kwalifikacji w danej dziedzinie przedmiotowej, albo wskutek zasadniczej nowości przedmiotu badań.

Niekompletność powiązana jest z faktem, że informacje o obiekcie są niewystarczające, aby stworzyć model adekwatnie odzwierciedlający jego cechy charakterystyczne.

Zarówno w razie niejednoznaczności, jak i niekompletności nie ma ogólnych metod, które uzupełniłyby brakujące informacje. Niekiedy uda się uzupełnić dane na

podstawie przypuszczeń o najbardziej prawdopodobnym zachowaniu badanego obiektu wypowiedzianych przez ekspertów na podstawie ich doświadczenia.

Większej formalizacji podlegają metody przedstawiania i budowania modeli w razie nieokreśloności danych powiązanej z *niepewnością*. Niepewność może być *fizyczna* (gdy jej źródłem jest środowisko zewnętrzne) i *lingwistyczna* (powstaje wskutek słownego opisu, w ogólnym przypadku, nieskończonych sytuacji przez ograniczone zbiory słów i określony czas).

W razie niepewności fizycznej mamy do czynienia z nieścisłością określenia parametrów obiektów wskutek nieścisłości pomiarów oraz z nieścisłością wskutek przypadkowości tych zdarzeń, które są powiązane z zachowaniem obiektów. Do opracowania danych o niepewności fizycznej zwykle używane są metody teorii prawdopodobieństwa i klasycznej teorii mnogości.

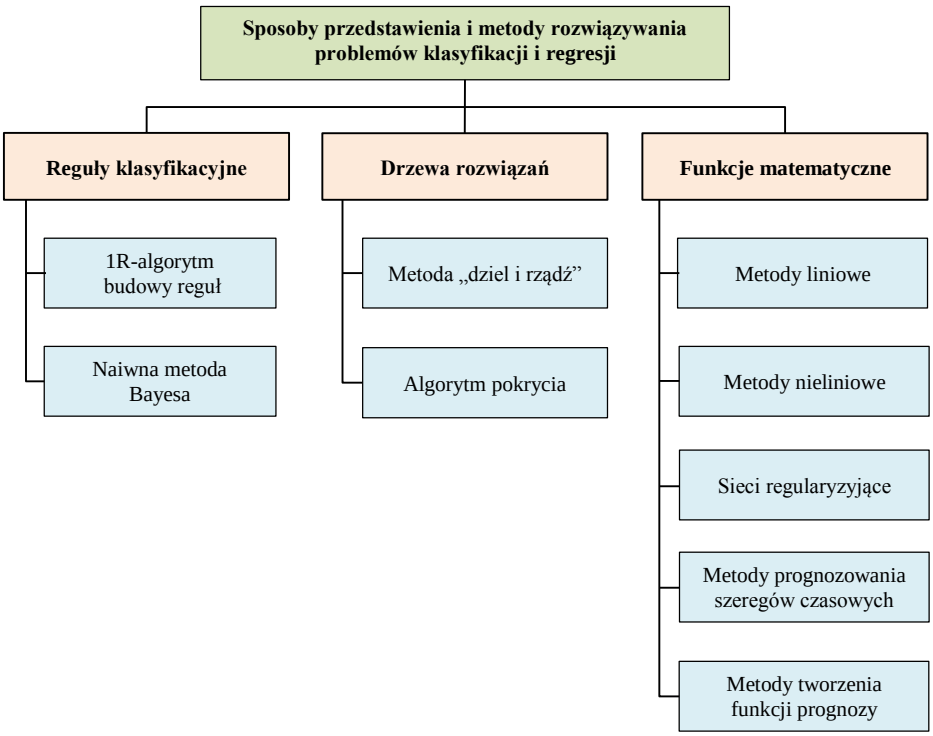
Wraz z rozwojem systemów, w których używane są metody teorii sztucznej inteligencji, coraz większą rolę odgrywają metody, w których kluczowego znaczenia nabierają opisy sytuacji i stanów obiektów za pomocą pojęć i relacji języka naturalnego, co niekiedy powoduje niepewność lingwistyczną. Idea wykorzystania lingwistycznych kategorii do symulacji obiektów ze słownymi opisami stanów i ich zachowań, stymulowała rozwój teorii logiki rozmytej L.A. Zadeha [48], [49].

Zwiększenie rozmiarów i złożoności systemów staje się barierą dla wykorzystania klasycznych metod symulacji do tworzenia adekwatnego modelu, opartych na metodach matematyki stosowanej. Chociaż użycie modelu słownego nie zapewnia dokładności liczbowej, jest jednak wygodnym sformalizowanym narzędziem opisu jakościowych właściwości systemów, co zbliża opis systemu do obrazowej realizacji systemu w mózgu człowieka. Metody symulacji złożonych systemów na podstawie logiki rozmytej i myślenia człowieka można, więc uznać w wielu wypadkach za heurystyczne. Jak pokazano w fundamentalnych pracach z dziedziny sztucznej inteligencji ([24], [25]) heurystyka, nie jest jednak cechą jedynie metod wnioskowania logicznego opartych na teorii zbiorów rozmytych, gdyż ta ostatnia uchyliła jedynie drzwi na świat nowych zasad symulacji systemów złożonych, w którym znalazły swoje godne miejsce również inne technologie przeznaczone do symulacji tak skomplikowanych procesów, których nie uda się badać za pomocą tradycyjnych metod „klasycznej" matematyki stosowanej.

Aby wszechstronnie ocenić zagadnienie prognozy, przeanalizujemy najpierw tę część omówionego zagadnienia, która dotyczy problemów eksploracji danych, czyli klasyfikacji i regresji.

1.4 Analiza metod klasyfikacji i regresji z punktu widzenia ich zdolności do rozwiązania problemów prognozy

Na podstawie przedstawionej analizy możemy, stwierdzić, że na ogół przewidywanie stanów i zachowania się złożonych obiektów i procesów należy do klasy zagadnień klasyfikacji i regresji. Na Rys. 1.4 pokazana została klasyfikacja podstawowych sposobów przedstawiania oraz podstawowe metody rozwiązywania problemów klasyfikacji i regresji (wykorzystane są materiały książek [24], [25], [39], [42], [50], [51]).



Rys. 1.4 Sposoby przedstawienia i metody rozwiązywania problemów klasyfikacji i regresji
Źródło: Opracowanie własne na podstawie [24], [25], [39], [42], [50], [51]

Idea *1R-algorytmu* jest bardzo prosta. Dla każdej wartości każdej zmiennej niezależnej (zmiennej wejściowej) budowana jest reguła klasyfikująca obiekty z próbki uczącej. W części wnioskującej każdej reguły wskazana jest najczęściej spotykana wśród obiektów podlegających testowaniu wartość zmiennej zależnej, która odpowiada wybranej wartości zmiennej niezależnej. Błąd reguły określany jest jako ilość obiektów mających te same wartości rozpatrywanej zmiennej, lecz nie zaliczonych do wybranej klasy. Dla każdej, więc zmiennej otrzymujemy zbiór reguł. Po ocenie błędów każdej próbki, wybrana będzie ta zmienna, dla której tworzono reguły z najmniejszym błędem.

Są dwie zasadnicze wady tego algorytmu. Po pierwsze, jeśli w próbce uczącej są

obiekty z pomijanymi wartościami zmiennych niezależnych, to 1R-algorytm zlicza te obiekty dla każdej ewentualnej wartości zmiennej, co powoduje tworzenie błędnych reguł. Po drugie, jeśli zmienna ma wartości zmiennoprzecinkowe (nie lingwistyczne), to ilość ewentualnych wartości może być nieskończona.

U podstaw *naiwnej metody Bayesa* leży wykorzystanie wzoru Bayesa do obliczenia prawdopodobieństwa. Idea metody polega na obliczeniu prawdopodobieństwa warunkowego przynależności obiektów do pewnych klas przy równości jego zmiennych niezależnych pewnym wartościom. Wada metody jest odzwierciedlona nawet w jej nazwie – „naiwna", pochodzącej od naiwnego przypuszczenia, że wszystkie zmienne wejściowe są względem siebie niezależne. W rzeczywistości przypuszczenie to nie zawsze jest prawdziwe.

Metodyka „dziel i rządzi" leży u podstaw większości metod budowy drzew decyzyjnych i oparta jest na rekurencyjnym podziale zbioru obiektów przedstawionych w próbce uczącej na podzbiory zawierające obiekty należące do tych samych klas. W taki sposób tworzy się tzw. drzewa rozwiązań. Wśród algorytmów, realizujących metodykę „dziel i rządzi" najbardziej powszechnie używane są algorytmy ID3 oraz C4.5. Procedura budowy drzewa leży u podstaw wielu tzw. „algorytmów zachłannych" [52]. Znaczy to, że jeśli raz wybrana została pewna zmienna i na jej podstawie dokonano podziału na podzbiory według różnych wartości tej zmiennej, to algorytm ten już się nie cofnie i nie będzie w stanie wybrać innej zmiennej, która doprowadziłaby do lepszego podziału. Dlatego nie możemy być pewni, że wybrana na początku zmienna doprowadzi do podziału optymalnego. Drugi problem występuje podczas budowy drzewa, a mianowicie problem przerwania procesu podziału. Aby pokonać wspomniane problemy, proponowane są obecnie różne reguły heurystyczne. Tym nie mniej, na razie nie istnieje jedyna zasada, która miałaby istotną wartość praktyczną [53]. Ponadto, w skomplikowanych przypadkach, gdy mamy do dyspozycji dużą ilość zmiennych, drzewa rozwiązań stają się zbyt złożone, aby być czytelne dla analityków.

W odróżnieniu od algorytmów budowania drzew rozwiązań „z góry na dół", *algorytm pokrycia* tworzy drzewo rozwiązań dla każdej klasy z osobna. Na każdym etapie testowany jest kolejny węzeł drzewa, który „pokrywa" kilka obiektów próbki uczącej jednocześnie. Po zakończeniu budowania drzewa dla jednej klasy, w taki sam sposób buduje się drzewo dla pozostałych klas. Algorytm nie gwarantuje, że uda się utworzyć reguły dla wszystkich obiektów bez wyjątku [25], niektóre obiekty mogą nie zostać w ogóle sklasyfikowane.

Jak można zauważyć na Rys. 1.4, klasa metod rozwiązywania problemów klasyfikacji i regresji z wykorzystaniem *funkcji matematycznych* obejmuje szerokie spektrum metod, co jest całkiem zrozumiałe, bowiem właśnie realizacja i rozwiązanie wskazanych problemów za pomocą zależności funkcjonalnych jest najbardziej naturalne i przybliżone do istoty zagadnień klasyfikacji i regresji na zbiorach wartości liczbowych.

Tworzenie funkcji klasyfikacji i regresji można najprościej przedstawić w sposób formalny jako zagadnienie wyboru funkcji z minimalnym błędem:

$$\min_{f \epsilon F} R(f) = \min_{f \epsilon F} \frac{1}{m} \sum_{i=1}^{m} c\big(y_i, f(x_i)\big) \tag{1.1}$$

gdzie F – to zbiór wszystkich ewentualnych funkcji; $c\big(y_i, f(x_i)\big)$ – funkcja strat (ang. *loss function*), w której $f(x_i)$ to wartość zmiennej zależnej, która jest wyznaczona poprzez obliczenie wartości funkcji f dla wektora zmiennych niezależnych $x_i \in T$ (T – dziedzina funkcji f), a y_i – dokładna wartość funkcji f.

W przypadku *metod liniowych*, funkcja ze zbioru F ma postać:

$$y = \omega_0 + \sum_{j=0}^{m} \omega_j x_j, \tag{1.2}$$

gdzie $\omega_0, \omega_1, \dots, \omega_m$ – to współczynniki (wagi) zmiennych niezależnych, w ilości m. Problem polega na tym, że należy znaleźć takie wartości współczynników ω, aby spełnić wymagania (1.1). Na przykład, podczas rozwiązania problemu regresji, współczynniki ω można obliczyć korzystając z kwadratowej funkcji strat:

$$c\big(x, y, f(x)\big) = (f(x) - y)^2, \tag{1.3}$$

oraz zbioru liniowych funkcji F:

$$F := \left\{ f \,\middle|\, f(x) = \sum_{i=1}^{n} \omega_i f_i(x), \omega_i \epsilon \Re \right\}, \tag{1.4}$$

gdzie funkcje f_i, w ilości n są pewnym podzbiorem funkcji przyjmujących wartości rzeczywiste $f_i: X \to \Re$.

W tym przypadku zagadnienie regresji najczęściej polega na wykorzystaniu metody najmniejszych kwadratów, która minimalizuje funkcjonał liniowy pewnej funkcji aproksymującej wartości zmiennej zależnej od zmiennych niezależnych [52]:

$$\min_{f \epsilon F} R(f) = \min_{f \epsilon \Re^n} \frac{1}{m} \sum_{i=1}^{m} \left(y_i - \sum_{j=1}^{n} \omega_j f_j(x_i) \right)^2 . \quad (1.5)$$

Obliczając pochodną $R(f)$ względem ω oraz wprowadzając oznaczenie $Y \coloneqq f_j(x_i)$, otrzymamy wniosek, że minimum jest osiągalne, gdy spełniony jest warunek:

$$Y^T y = Y^T Y \omega. \quad (1.6)$$

Rozwiązaniem tego równania jest:

$$\omega = (Y^T Y)^{-1} Y^T y. \quad (1.7)$$

Stąd wyznaczamy wartości współczynników ω.

Metody nieliniowe z reguły lepiej klasyfikują obiekty, jednak ich tworzenie i wykorzystanie jest bardziej skomplikowane. W tym przypadku problem, podobnie, polega na minimalizacji funkcjonału (1.5), z tym wyjątkiem, że zbiór *F* zawiera funkcje nieliniowe. Szeroko rozpowszechnione podejście do rozwiązania problemu klasyfikacji polega w tym przypadku na tworzeniu modeli liniowych. Dlatego początkowa przestrzeń obiektów X z wykorzystaniem pewnego operatora Φ jest przekształcona do nowej przestrzeni X':

$$\Phi: X \to X'. \quad (1.8)$$

W tej nowej przestrzeni tworzona jest funkcja liniowa, której w przestrzeni początkowej odpowiada pewna funkcja nieliniowa. Po rozwiązaniu problemu regresji, czyli po wykonaniu w przestrzeni X' procedury minimalizacji funkcjonału liniowego (1.5) wykonujemy zwrotne przekształcenie do przestrzeni początkowej.

Istotną wadą opisanego podejścia jest to, że procesy przekształceń przestrzeni funkcji są dość złożone z punktu widzenia ilości obliczeń i niejednoznaczności wyboru odpowiedniego operatora Φ. Dlatego wspomniane metody ograniczone są do rozwiązania problemów regresji o stosunkowo małych ilościach zmiennych. Ponadto w metodach nieliniowych nie ma uniwersalnej techniki przekształcania nieliniowej przestrzeni zmiennych do przestrzeni liniowej i na odwrót, co istotnie utrudnia ich zastosowanie.

Wszystkie opisane wyżej metody rozwiązania problemów klasyfikacji i regresji dążą do wykonania poprawnej klasyfikacji *wszystkich* bez wyjątku wektorów danych – niezależnie od stopnia złożoności funkcji i wkładu różnych zmiennych niezależnych w kształtowanie stanów obiektów (wartości zmiennych wyjściowych). Ponadto, opisane wyżej metody klasyfikacji i regresji prowadzą do mnogości rozwiązań, a w niektórych

przypadkach tworzą problemy niewłaściwej klasyfikacji i niestabilności obliczeniowej [42].

Aby pokonać te problemy, zaproponowano podejście regularyzacji operatorów typu A.N. Tichonowa [54]. W wyniku rozwoju tego podejścia otrzymano uogólnienie (1.5) o nazwie *sieci regularyzujące* (*Regularization Networks*). W rzeczywistości, badania sieci regularyzujących doprowadziły do uogólnienia tak szeroko rozpowszechnionych technik, jak *regresja grzbietowa* (ang., *Ridge Regression*) oraz *wygładzające krzywe sklejane* (*Smoothing Splines*). Teoretyczne uzasadnienie tej tezy zostało przedstawione głównie w pracach F. Girosi, M. Jones, T. Poggio ([42], [55]). W szczególności, w pracach wspomnianych autorów pokazano, że wiele współczesnych metod klasyfikacji i regresji może być interpretowanych jako metody rozwiązania następującego problemu regularyzacji:

$$\min_{f \epsilon F} R(f) = \min_{f \epsilon F} \frac{1}{m} \sum_{i=1}^{m} c\big(y_i, f(x_i)\big) + \lambda \varphi(f), \tag{1.9}$$

gdzie $\varphi(f)$ – to operator wyrównujący, $\varphi(f) = \|Pf\|^2$, tj. $Pf = \nabla f$, λ – parametr regularyzacji.

Pierwszy składnik w tym wyrażeniu $\frac{1}{m}\sum_{i=1}^{m} c(y_i, f(x_i))$ zapewnia dobrą adaptację funkcji klasyfikującej do danych uczących, podczas gdy drugi składnik $\varphi(f)$ „utrzymuje” funkcję klasyfikującą w postaci na tyle gładkiej, na ile jest to możliwe i zapewnia dobrą zdolność do uogólnienia. Pierwszy składnik, skłania się jakby ku „przeuczaniu” (*overfitting*), podczas gdy drugi – ku „niedouczaniu”. Dobierając wartość parametru regularyzacji λ, można osiągnąć kompromis między tymi dwoma składnikami.

Wiele metod aproksymacji, w tym tworzenie modeli addytywnych, funkcji radialnych oraz niektóre typy sztucznych sieci neuronowych może zostać wyprowadzonych poprzez wybór operatora regularyzacji.

Zasady leżące u podstaw tworzenia i wykorzystania sieci regularyzacyjnych stają się ważnym narzędziem do badania wielu metod klasyfikacyjnych i regresyjnych. W pracach wielu autorów [39], [41], [42], [43], [55] stwierdza się, że kierunek badań sieci regularyzacyjnych i ogólnych koncepcji leżących u ich podstaw jest bardzo obiecujący mimo, że jesteśmy dopiero na początkowym etapie rozwoju tych idei.

Przypadkiem szczególnym problemu klasyfikacji jest problem prognozowania *szeregów czasowych.* Szeregiem czasowym nazywamy ciąg zdarzeń uporządkowanych

w czasie ich obserwacji $E = \{e_1, e_2, \dots, e_j, \dots, e_n\}$, gdzie e_i to zdarzenie zachodzące w chwili czasu t_i, a $n \in N$.

W terminach teorii systemów, zdarzenia można rozpatrywać jako stany badanego obiektu (lub procesu) w chwilach czasu $t_1, t_2, \dots, t_j, \dots, t_n$. W ogólnym przypadku, każde zdarzenie może być opisane przez kilka atrybutów $\{x_1^i, x_2^i, \dots, x_j^i, \dots, x_m^i\}$, gdzie x_j^i – to *j*-y atrybut charakteryzujący jedną z właściwości badanego obiektu lub procesu w chwili t_i, przy czym w terminach teorii systemów wspomniane atrybuty można rozpatrywać jako *zmienne stanów* obiektu (procesu). Jeżeli jeden atrybut może być określony za pomocą wartości innych atrybutów w chwili bieżącej lub w poprzednich chwilach czasu, to taki atrybut jest zależny wówczas, gdy inne atrybuty są niezależne.

Zagadnienie analizy stanów obiektów (procesów) na podstawie szeregów czasowych brzmi następująco: Niech dany jest szereg czasowy $E = \{e_1, e_2, \dots, e_j, \dots, e_n\}$. Przy czym, możliwa jest jedna z dwóch postaci sformułowania problemu:

1. Należy określić stan e_{n-p} obiektu (procesu) w chwili czasu t_{n-p}, gdzie $p > 0$.
2. Należy określić stan e_{n+p} obiektu (procesu) w chwile czasu t_{n+p}, gdzie $p > 0$.

Pierwszy przypadek, który w pewnym sensie przypomina problem interpolacji funkcji, jest w rzeczywistości problemem opisu zachowania się obiektów na przedziale obserwacji, podczas gdy drugi, który przypomina problem ekstrapolacji funkcji – jest problemem prognozy zachowania się obiektów w przyszłych chwilach czasu, czyli poza przedziałem obserwacji.

Mówiąc o problemach prognozy, należy rozróżniać prognozę krótko- średnio- i długoterminową. Pojęcie terminowości prognozy może być dokładniej sprecyzowane w przypadkach, gdy mamy do czynienia z procesami stacjonarnymi.

Przypomnijmy, że proces stochastyczny nazywamy *stacjonarnym*, jeśli jego wartość oczekiwana jest stała (M_x = *const*), a funkcja korelacyjna jest funkcją przesunięcia między argumentami: $Kx(t_1, t_2) = Kx(\tau)$. Ponadto, jeśli stacjonarny proces przechodzi przez te same stany, niezależnie od tego, w jakiej chwili czasu ten proces się rozpoczął, to taki stacjonarny proces nazywamy *ergodycznym*. Prognoza procesu ergodycznego może być wykonana z wysoką dokładnością poprzez wykorzystanie modelu opisującego ten proces na wystarczająco długim odcinku czasu.

W przypadku, gdy stacjonarny proces podlegający prognozie, nie jest ergodyczny, sytuacja nieco się komplikuje. Dowolny problem prognostyczny jest zwykle rozwiązywany w pewnej bieżącej chwili czasu t_n obserwacji obiektu (nazwijmy ten

punkt czasowy punktem centralnym) z wykorzystaniem metodyki opartej na tworzeniu i wykorzystaniu klasyfikatora. Ogólnie, więc czasowy przedział symulacji obiektu (procesu) *T* składa się z czasowego przedziału tworzenia i uczenia klasyfikatora T_{ucz} oraz czasowego przedziału prognozy, na którym przewiduje się stany badanego obiektu w przyszłości T_{progn}, czyli:

$$T = \left[T_{ucz}, T_{progn}\right]. \tag{1.10}$$

Uczenie klasyfikatora realizowane jest na pewnym ciągu dyskretnych chwil czasu w przeszłości

$$T_{ucz} = \{t_{n-m}, t_{n-m+1}, t_{n-m+2}, \dots, t_n\}, \tag{1.11}$$

wówczas, problem prognozy rozwiązywany jest na ciągu dyskretnych chwil czasu w przyszłości

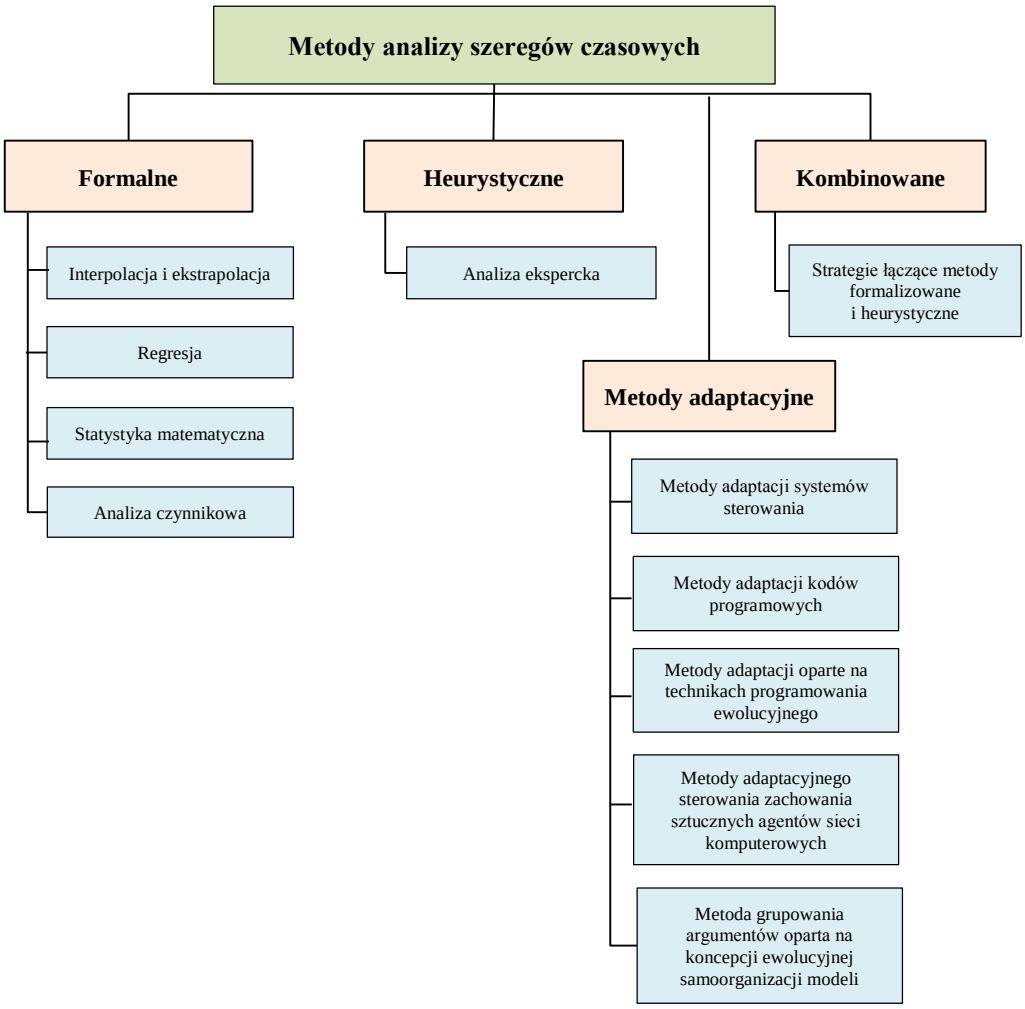
$$T_{progn} = \{t_{n+1}, t_{n+2}, t_{n+3}, \dots, t_{n+k}\}. \tag{1.12}$$

Ogólny, więc przedział czasu badania obiektu leży w przedziale

$$T = \left[T_{ucz}, T_{progn}\right] = [t_{n-m}, t_{n+k}] \tag{1.13}$$

Oczywiście, dla procesów stacjonarnych przedział czasu prognozy T_{progn} nie może być większy od czasu korelacji T_{kor}. Przy czym, zwykle przez czas korelacji stacjonarnego procesu stochastycznego y_t, z monotonicznie malejącą funkcją korelacyjną $R_y(t)$, rozumiemy czas, od którego zaczynając funkcja ta jest bliska zeru. Znaczy to, że procesów podobnych do szumu „białego" ($T_{kor} = 0$) nie można w zasadzie prognozować, podczas gdy procesy deterministyczne, np. procesy sinusoidalne ($T_{kor} = \infty$) można prognozować ze 100%-wą dokładnością na dowolnym odcinku czasu. Przyjmując, że zmiany stanów badanego obiektu to stochastyczne procesy stacjonarne z monotonicznie malejącej funkcją korelacji, będziemy uważać, że prognoza jest *krótkoterminowa*, jeśli przedział jej czasu T_{progn} nie przekracza 10% czasu korelacji T_{kor}, *średnioterminową*, jeśli T_{progn} nie przekracza 30% czasu korelacji T_{kor} i długoterminową, jeśli T_{progn} przekracza 30% czasu korelacji T_{kor}.

Ze wszystkich opisanych wyżej podejść do rozwiązania problemów klasyfikacji i regresji, istota problemu tworzenia szeregów czasowych najbardziej odpowiada istocie badanego w rozprawie problemu prognozy. Metody prognozy zachowania się obiektów (procesów) na podstawie szeregów czasowych (dalej krótko: *metody szeregów czasowych*) można podzielić na cztery grupy, jak pokazano na Rys. 1.5.



Rys. 1.5 Metody prognozy zachowania się obiektów (procesów) na podstawie szeregów czasowych
Źródło: Opracowanie własne

Wynikiem zastosowania *metod formalnych* są wskaźniki liczbowe opisujące ilościowe charakterystyki stanów obiektów i procesów. Do grupy tych metod można zaliczyć: metody bazujące na technikach interpolacji i ekstrapolacji funkcji [56], [57], metody statystyki matematycznej [58], metody analizy czynnikowej [59].

Metody heurystyczne oparte są na wykorzystaniu ocen eksperckich [60], [61]. Ekspert (lub grupa ekspertów), opierając się na swojej wiedzy w danej dziedzinie przedmiotowej oraz doświadczeniu, zdolny jest do przewidywania zmian zachowania się badanego obiektu lub procesu, zakładając, że proces ten jest stacjonarny. Metody te są szczególnie użyteczne, gdy należy przedstawić prognozę zmian jakościowych stanów obiektów w warunkach nieściśle określonych i w warunkach niepełnej wiedzy o wszystkich atrybutach opisujących stany obiektu. Charakterystyczne dla ocen eksperckich jest to, że zwykle mają postać nie liczbową, a jakościową.

Metody kombinowane łączą w sobie podejścia realizowane w metodach pierwszych dwóch grup, co w niektórych przypadkach prowadzi do dobrych wyników [62].

Podkreślmy, że metody wspomnianych wyżej trzech grup oparte są na formalnym przypuszczeniu, że badany *obiekt lub proces* posiada właściwość inercji[7] (co odpowiada definicji procesu stacjonarnego). Takie prawa zwykle formułuje się na podstawie badań „historii" stanów obiektów w przeszłości i nie zmieniają się one podczas ocen przyszłych stanów. Dlatego metody te przeważnie przeznaczone są do rozwiązania kategorii problemów interpolacji stanów obiektów w chwili czasu t_{n-p}, a także ekstrapolacji stanów obiektów w nieodległej chwili czasu w przyszłości t_{n+p}, tj. prognozy krótkoterminowej.

Osobne miejsce w naszej klasyfikacji zajmują *metody adaptacyjne*. Metody te tworzone są głównie do symulacji i sterowania obiektów o złożonym charakterze zachowania się. Do tej kategorii metod można zaliczyć podejścia mające różne zastosowania, lecz powiązane wspólną cechą, określoną słowami „zdolność do modyfikacji swoich działań w zmieniających się warunkach funkcjonowania": metody adaptacji systemów sterowania [9], metody adaptacji kodów programowych [30], adaptacyjne metody symulacji zachowania się złożonych obiektów tworzone z wykorzystaniem technik programowania ewolucyjnego [10], metody adaptacyjnego sterowania zachowania się sztucznych agentów sieci komputerowych [35], oraz różne wersje metody grupowania argumentów [5], [8], [50].

Natomiast idee *metod samoorganizacji modeli* zapoczątkowane były w kilku pracach i mają nieco odmienne interpretacje naukowe. Np. N. Wiener [4] utożsamiał samoorganizację z adaptacją systemów, czyli zdolnością systemów do przystosowania swojej struktury i parametrów do zmian zewnętrznych stanów środowiska. W.R. Ashby [18] porównywał samoorganizację sztucznych systemów ze zdolnością przystosowania się mózgu człowieka do rozwiązania nowych problemów. J. Tsypkin [9] i J. Holland [10] dopatrywali się w adaptacji zdolności do przeżycia obiektów przyrody ożywionej lub do zdolności systemów sztucznych do rozwiązania inteligentnych problemów ze zmieniającymi się warunkami. W pracy A.G. Ivakhnenki i współautorów [50] metody selekcji, ewolucji i adaptacji modeli połączone są pod wspólną nazwą – metody samoorganizacji heurystycznej. We współczesnych pracach badaczy metod sztucznej

[7] Założenie, że w przyszłości obiekt (proces) będzie rozwijał się zgodnie z prawami, według których działał on w przeszłości i istnieje obecnie.

inteligencji (patrz, np. [24], [25], [30]) adaptacja jest rozpatrywana w kontekście tworzenia specjalnych zasobów (algorytmów i oprogramowania) do realizacji zasad samoorganizacji ewolucyjnej w systemach sztucznych przeznaczonych do podejmowania decyzji na podstawie ekstrakcji wiedzy o badanym obiekcie z obszernych źródeł danych. Możemy, więc wnioskować, że tworzenie adaptacyjnych systemów prognozy jest zagadnieniem tej części informatyki, która zajmuje się analizą danych i procesów, czyli eksploracją danych, oraz strukturalną i algorytmiczną samoorganizacją informatycznych systemów, tak więc należy do kategorii problemów rozwiązywanych zasobami sztucznej inteligencji.

1.5 Sformułowanie problemu prognozy zachowania złożonych systemów

Prognoza zachowania złożonych systemów jest jednym z aktualnych problemów, wokół którego skupiają się badania systemów złożonych. Prognoza to naukowo uzasadnione oszacowanie przyszłych stanów obiektu, realizowane na podstawie analizy stanów poprzednich oraz dynamiki zachowania się obiektu. Powinna ona określać, kiedy i w jakiej kolejności będzie przebiegała zmiana stanów obiektu oraz jaki wpływ będzie miał stan obiektu na wykonanie postawionego zadania.

Prognoza jest praktycznie nieodłączną częścią każdej naukowej działalności. Aby przedstawić matematyczny opis obiektu, używa się informacji formalizujących zachowanie obiektu w postaci takich lub innych regularności podczas, gdy niektóre nieregularne wpływy środowiska zewnętrznego na obiekt rozpatrywane są jako przypadkowe oddziaływania. Jednocześnie zewnętrzne wpływy, które zmieniają przebieg procesów wewnątrz danego obiektu i mogą się zmieniać w czasie, nierzadko również są istotne i powinny zostać zdefiniowane, aby mieć możliwość utworzenia tzw. *prognozy normatywnej*, czyli prognozy opartej na ujawnieniu regularności zmian stanów obiektu i otoczenia zewnętrznego.

Ciąg prognoz uporządkowanych według reguły „jeżeli→to” nazywamy niekiedy *scenariuszami* przyszłości. Eksperci mogą być poproszeni o wybór scenariusza, który najlepiej (według ich opinii) odpowiada zadaniu rozwiązywanemu przez dany obiekt/system. Oprócz „przepracowania” omówionych scenariuszy możliwa jest też ślepa taktyka sterowania danym obiektem opierająca się na „przypadkowości”, co w większości przypadków rozwiązania problemów jest niedopuszczalne. Miejsce ekspertów podejmujących decyzje i wybierających potrzebny scenariusz mogą zajmować

eksperci-ludzie, systemy wyposażone w odpowiednie oprogramowanie lub systemy hybrydowe łączące w sobie ekspertów i oprogramowanie wykonujące funkcje doradcze.

Rozwój algorytmów heurystycznych prowadzi nas do bardziej ogólnego zagadnienia – budowy tzw. systemów samoorganizujących się. Wśród nich można wyszczególnić kategorię systemów, zasadniczą właściwością których staje się zdolność do samodzielnego nastrajania się, zapewniającego ich optymalne (w pewnym sensie) zachowanie się na podstawie odbierania, klasyfikacji i analizy informacji o obiekcie, bezpośrednio w trakcie działania symulowanego systemu. Jak już wspomniano wcześniej, takie systemy uznajemy za złożone. Bliższe zaznajomienie się ze źródłem rozwoju metod symulacji systemów złożonych daje podstawy stwierdzić, że granica między „klasycznymi" technologiami a nowoczesnymi metodami symulacji systemów jest dość płynna. Należy raczej mówić o pewnym zbiorze rozmaitych metod i algorytmów, których elastyczne wykorzystanie pozwala otrzymać jakościowo nowe zasady symulacji systemów i procesów o złożonym charakterze zachowania.

1.6 Analiza podstawowych zasad samoorganizacji modeli prognozy na przykładzie metody grupowania argumentów

Zrozumienie faktu, że prognoza jest jednym z zagadnień klasyfikacji i regresji z zakresu analizy danych i procesów, doprowadzi nas do kolejnego pytania: jakie modele matematyczne w postaci zależności funkcyjnych spośród tych, które zaproponowane są w teorii metod klasyfikacji i regresji, posiadają niezbędną zupełność do rozwiązania problemów prognozy?

Innymi słowy, tworząc model zachowania się systemu w postaci funkcji regresji, musimy mieć pewność, że model ten obejmuje wszystkie możliwe przyszłe stany, i nie ma „luk" niedostrzegalnych przez tworzone przez nas modele, czyli z reguły niewidocznych stanów systemu. Fenomen niewidoczności stanów obiektów został określony w teorii systemów jako *nieobserwowalność stanów* [63].

Odpowiedź na powyższe pytanie przedstawiona jest w twierdzeniu *Stone'a-Weierstrass'a* o przybliżaniu funkcji [64], [65]. Oznaczmy przez $\boldsymbol{C}([a,b])$ przestrzeń składającą się ze wszystkich funkcji ciągłych określonych na przedziale $[a,b]$. Na mocy twierdzenia Stone'a-Weierstrass'a, każdą funkcję ciągłą na przedziale domkniętym i ograniczonym $[a,b]$ możemy przybliżać w metryce przestrzeni $\boldsymbol{C}([a,b])$, a więc jednostajnie, za pomocą wielomianów. Twierdzenie to możemy sformułować

następująco: zbiór wielomianów określonych na przedziale $[a, b]$ jest gęsty w przestrzeni $\boldsymbol{C}([a, b])$.

Tak więc, wybierając strukturę tworzonych modeli, będziemy opierać się na ich postaci wielomianowej. Wśród modeli tej klasy osobne miejsce zajmuje *metoda grupowania argumentów* (ang., *Group Method of Data Handling, GMDH*) zaproponowana w pracach prof. *A.G. Ivakhnenki* [50], [8], która rozwija się obecnie w pracach jego uczniów oraz innych badaczy (np. [5]).

W danej rozprawie wspomniana metoda została wybrana jako podstawowa dla tworzenia modeli prognozy. Aby uzasadnić przewagę GMDH, jako podstawowej metody dla tworzenia modeli prognozy, przytoczymy jej główne cechy. W poniższej analizie wykorzystano głównie prace [5], [8], [50], [66].

Załóżmy, że mamy dany pewien system, którego stany poddają się opisowi za pomocą funkcji, w tym wielomianów. Ciągi zmiennych wejściowych (np. danych albo sygnałów) i danych wyjściowych (np. wartości funkcji opisujących przyszłe stany badanego systemu lub reakcje systemu na sygnały wejściowe) tworzą odpowiednio zbiory danych wejściowych i wyjściowych. Ponieważ w naszych badaniach będziemy mieli do czynienia z danymi w postaci liczb rzeczywistych, to wspomniane zbiory zawierające powiedzmy *n* danych wejściowych i *m* danych wyjściowych, można formalnie przedstawić odpowiednio jako *n*-wymiarowe wektory zmiennych wejściowych i *m*-wymiarowe wektory zmiennych wyjściowych: $x\epsilon\,\Re^n$, $y\epsilon\,\Re^m$, gdzie $\Re$ to euklidesowa przestrzeń rzeczywista.

Szukamy matematycznego opisu systemu, w postaci modelu „wejście-wyjście”

$$Y(t) = F\big(u(t)\big),\ t_0 \le t < T, \tag{1.14}$$

gdzie F to rozmaitość w przestrzeni zmiennych wejściowych i wyjściowych określonych w chwili czasu t, $Y(t)$ – wektor zmiennych wyjściowych określony przez zastosowanie jednej z funkcji wybranych do symulacji zachowania się systemu, $u(t)$ – wektor zmiennych wejściowych, $t\epsilon[t_0, T]$ – czasowy przedział obserwacji systemu. Ocenie jakości tworzonego modelu zachowania systemu służy kryterium oceny jakości modelu $Q\big(Y(t), Y_S(t)\big)$, w którym porównujemy dokładną wartość funkcji wyjścia systemu $Y(t)$ z wartością funkcji wyjścia $Y_S(t)$ otrzymaną w wyniku symulacji. To porównanie można wykonać mając do dyspozycji zbiory próbek testowych.

Stosując dyskretny odpowiednik szeregu Volterry dla systemu dynamicznego w postaci szeregu Kołmogorowa-Gabora, zapiszemy (1.14) w postaci:

$$y_S(t) = a_0 + \sum_{i=0}^{n_1} a_i x_{it} + \sum_{i=0}^{n_2}\sum_{j=0}^{n_3} x_{it} A_{ij} x_{jt} + \cdots, \qquad (1.15)$$

gdzie $x_{it} = u(t-i), a_0, \ldots, a_i, \ldots,$ oraz A_{ij} – współczynniki rzeczywiste.

Problem prognozy w klasycznej interpretacji przedstawionej w pracach twórców GMDH [8], [50], brzmi następująco. Na podstawie obserwacji zachowania systemu tworzymy pary „wejście-wyjście" (u_t, y_t), które dalej wykorzystujemy do syntezy modelu zachowania się systemu w przyszłości. Model ten przedstawiamy w postaci pewnej funkcji prognozy:

$$y_S(t+\tau) = f\big(y(t), y(t-1), \ldots, y(t-k), v(t), v(t-1), \ldots, v(t-k)\big), \qquad (1.16)$$

gdzie $y(t), y(t-1), \ldots, y(t-k)$ oraz $v(t), v(t-1), \ldots, v(t-k)$ to odpowiednio, wartości wektora wyjściowego oraz wartości parametrów wpływających na wartości zmiennych wyjściowych lub powiązanych z nimi zmiennych (w tym zmiennych wejściowych) w przeszłych chwilach czasu obserwacji $t, t-1, \ldots, t-k$.

Niestety najczęściej obiektem symulacji jest system nieergodyczny lub nawet niestacjonarny, dla którego ocena terminowości prognozy nie jest ściśle określona. W tych warunkach autorzy GMDH [8], [50] proponują ze względów praktycznych zakładać, że w przypadku gdy $\tau \leq 3 \ldots 10$, problem prognozy można zaliczyć do kategorii prognozy krótkoterminowej, a dla $\tau \gg 10$ – do kategorii problemów prognozy długoterminowej. Dokładne granice terminowości w klasycznym sformułowaniu problemu tworzenia GMDH są nieokreślone, ponieważ terminowość tworzonego modelu w sposób istotny zależy od cech badanego systemu (szybkości zmian stanów środowiska, w których działa dany system, szybkości reakcji systemu na te zmiany, obecności nieprzewidywanych wpływów, szumów itd.).

Złożoność modelu tworzonego na podstawie GMDH zależy od ilości punktów niezbędnych do określenia wartości współczynników modelu (1.15). W systemie komputerowym samoorganizacja modeli w klasycznej jej interpretacji [8], [50] jest realizowana poprzez wygenerowanie wystarczająco dużych ciągów alternatywnych modeli-pretendentów, z których na podstawie wykorzystania pewnych kryteriów odrzucane są zbędne lub nieistotne relacje. Uważa się, że wynikiem tego doboru będzie model *optymalnej złożoności*.

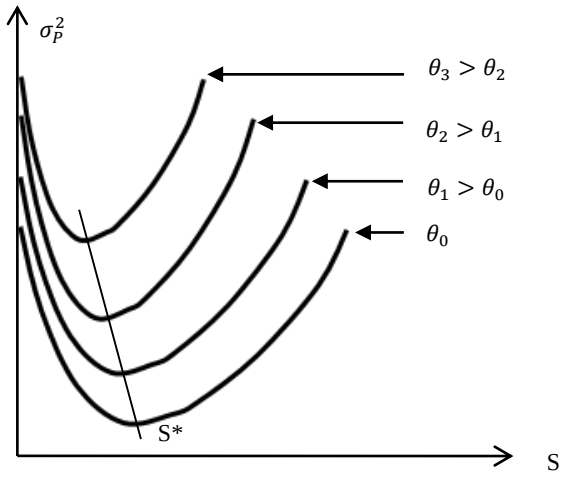
Zasady, na podstawie których tworzono algorytmy samoorganizacji modeli, a także główne cechy tych algorytmów można określić następująco: [5], [8], [50], [66], [67].

1. Ze względu na sposoby generowania modeli, algorytmy samoorganizacji różnią się *stopniem złożoności struktury modelu* (1.15) (jednowarstwowe, wielowarstwowe, z zerowaniem współczynników, wieloetapowe i inne) oraz *zbiorem kryteriów* (jedno- i wielokryterialne). Ponadto, generowanie modeli-pretendentów może być oparte na koncepcjach *indukcji zupełnej* albo *niezupełnej*.
2. *Zasadę samoorganizacji*, która jest jedną z fundamentalnych w GMDH, można sformułować następująco: po stopniowym zwiększeniu złożoności struktury modelu wartość kryteriów zewnętrznych najpierw maleje, a po osiągnięciu minimum pozostaje stała lub rośnie. Ponad to, jeśli kryterium doprowadzi do wieloznaczności wyboru modeli lub model jest zbyt czuły na małe zmiany danych wejściowych, to dodawanie do niego drugiego kryterium (*operatora regulującego*) z charakterystyką jednomodalną pozwala dokonać jednoznacznego wyboru modelu optymalnej złożoności. Udowodniono, że wykorzystanie regulujących operatorów jest mechanizmem istotnego zwiększenia skuteczności modeli prognozy [8].
3. *Zasada dopełnienia zewnętrznego* – jest stosowana do problemów mających wieloznaczne rozwiązanie. W szczególności, takim problemem jest problem prognozy. W celu rozwiązania tych problemów należy skorzystać z dopełnienia zewnętrznego, pod pojęciem którego rozumiemy kryterium zewnętrzne. To takie kryterium, które jest obliczane na podstawie danych niewchodzących w skład zbioru uczącego. Dla każdego dopełnienia zewnętrznego mogą istnieć różne rozwiązania problemu prognozy. Stosownie tej zasady do problemu samoorganizacji modeli oznacza, że korzystając z danych, które były już wykorzystane wcześniej do tworzenia modelu, bez dodatkowych zewnętrznych informacji, nie można ich użyć do znalezienia modelu optymalnej złożoności. W tych warunkach niezbędne jest kryterium, które ocenia błędy na nowych realizacjach, czyli na nowych danych.
4. *Podejście Gödel'a do samoorganizacji modeli* to sposób pokonywania wyników twierdzenia Gödel'a o niezupełności (podrozdział 1.2). W problemach samoorganizacji modeli idee Gödel'a można interpretować następująco: opierając się na minimum danego kryterium zewnętrznego można rozwiązać wszystkie problemy dotyczące wyboru funkcji nośnikowych, a także struktury i parametrów modelu, za wyjątkiem problemów wyboru algorytmów obliczeń i sposobów zastosowania kryteriów, czyli problemów drugiego poziomu hierarchii, a optymalizacja tych ostatnich wymaga wykorzystania modeli trzeciego poziomu hierarchii itd. Systemy obliczeniowe działające w ramach tylko pierwszego poziomu hierarchii, nazwijmy

systemami typu gödelewskiego, natomiast systemy działające w taki sposób, że są oparte na wykorzystaniu modeli i kryteriów wyższego poziomu hierarchii, nazwijmy *systemami typu niegödelewskiego*. Pod pojęciem podejścia Gödel'a do samoorganizacji modeli będziemy więc rozumieli strukturę algorytmu z rozwijającymi się w czasie modelami.

5. *Zewnętrzne kryteria selekcji modeli*. W problemach regresji dla różnych próbek tego samego procesu można otrzymać różne funkcje regresji. Które z tych modeli należy wybrać? Wybór odpowiedniego modelu może być dokonany na podstawie *kryterium minimalnej zmiany*. Kryterium to wymaga, aby modele tworzone z wykorzystaniem danych z pewnego zbioru *A* jak najmniej różniły się od modeli tworzonych z wykorzystaniem danych ze zbioru *B*. Inaczej mówiąc, chodzi, po pierwsze, o kryterium zgodności modeli tworzonych na podstawie różnych zestawów próbek dla tego samego obiektu, i po drugie, o konieczności podziału zestawu danych wejściowych na dwa lub nawet trzy zestawy: dane uczące *A*, dane testujące *B* oraz, ewentualnie, dane egzaminujące *C*.
6. *Zasada Gabora wsparcia wolności wyboru* mówi [67], że optymalność wyboru danych i modeli osiągana jest za pomocą przejścia do kolejnego etapu rozwiązań w taki sposób, aby do kolejnego etapu selekcji przekazane było nie jedno rozwiązanie, a kilka najlepszych rozwiązań otrzymanych w poprzednim etapie lub kilku poprzednich etapach. W swoim artykule [67] D. Gabor tłumaczy swoją zasadę następująco: w procesie kolejnego podejmowania decyzji, na każdym etapie algorytmu należy wybierać pewną ilość najlepszych rozwiązań *F*, w taki sposób, aby dalsze zwiększenie tej ilości nie doprowadziło do istotnego ulepszenia wyników selekcji, a także by w następnej chwili czasu, gdy będzie niezbędne kolejne rozwiązanie, była możliwa zagwarantowana swoboda wyboru rozwiązań.

Jak udowodniono w pracach dotyczących GMDH (patrz np. [5], [8], [50], [66]), zasada ta jest właściwa z matematycznego punktu widzenia, ponieważ udowodniono, że w miarę zwiększania złożoności modelu *S* (gdzie *S* to maksymalny rząd wielomianu aproksymującego), kryterium zewnętrzne (np., odchylenie średniokwadratowe na próbce testującej σ_P^2) przechodzi przez minimum; wspomniana regularność obserwowana jest nawet w warunkach tworzenia modeli z wykorzystaniem zaszumionych danych (Rys. 1.6).



Rys. 1.6 Wartość odchylenia średniokwadratowego na próbce testującej σ_P^2, jako funkcja złożoności modelu *S* i poziomu szumu θ (każdy wykres ma jawnie określone minimum)
Źródło: [66]

Zasada ta stosowana jest w metodach tworzenia wielowarstwowych algorytmów GMDH, w tym, w tworzeniu zbiorów kryteriów do wyboru najlepszych modeli symulacji. Więc, wielowarstwowość i zachowanie wolności wyboru można uznać za podstawowe cechy algorytmów samoorganizacji tworzonych na podstawie GMDH.

7. Na przykładach wielu praktycznych rozwiązań problemów prognozy z wykorzystaniem GMDH udowodniona została teza o istotności zasady *symulowania wielopoziomowego*, pod którą autorzy GMDH rozumieją jednoczesną symulację na różnych poziomach uogólnienia języka matematycznego opisu obiektu. Niestety zasada ta dotychczas nie została rozwinięta przez twórców GMDH do postaci metod lub algorytmów i jest przedmiotem dalszych badań.
8. *Zasada kombinatoryki jednoszeregowych algorytmów GMDH* - mówi, że algorytmy indukcji zupełnej (nazywane również algorytmami kombinatorycznymi) można otrzymać korzystając z wielomianu zupełnego odpowiedniego rzędu poprzez wykreślanie składników. Po wykreśleniu niektórych składników otrzymamy wielomiany niezupełne (częściowe). Np. równanie regresji dla dwóch argumentów x_1 i x_2 ma postać wielomianu zupełnego drugiego rzędu zawierającego 6 składników:

$$y = a_0 + a_1x_1 + a_2x_2 + a_3x_1^2 + a_4x_2^2 + a_5x_1x_2 \qquad (1.17)$$

Wykreślając z tego wielomianu różne składniki, można utworzyć ciąg wielomianów częściowych, np.:

$$y_1 = a_0 + a_1x_1;$$

$$y_2 = a_0 + a_1 x_2;$$
$$y_3 = a_0 + a_1 x_1 + a_2 x_2;$$
$$y_4 = a_0 + a_1 x_1 x_2;$$
itd.

Jeśli stały składnik a_0 jest obecny we wszystkich wielomianach częściowych, to przesortowaniu ulegają wszystkie pozostałe 5 składników ze wzoru (1.17), czyli ilość wielomianów częściowych będzie równa:

$C_5^1 + C_5^2 + \cdots + C_5^5 = 32$ wielomiany częściowe.

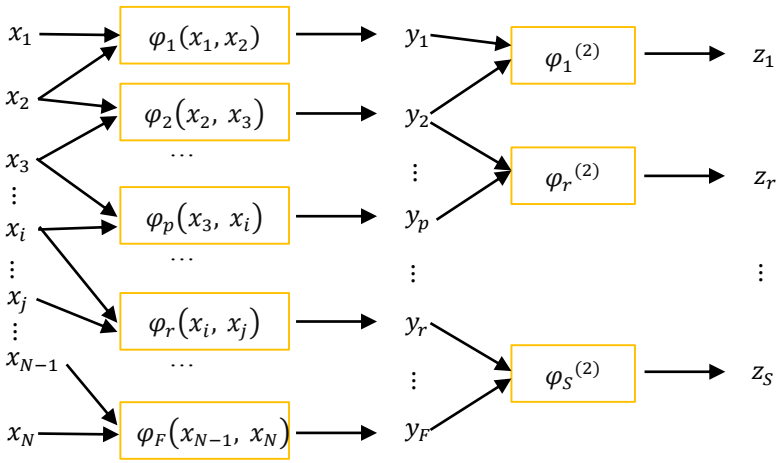
Dla wielomianu zupełnego trzeciego rzędu zawierającego 20 składników jest 2^{19} wielomianów częściowych. Dla wielomianu zupełnego czwartego rzędu zawierającego 70 składników można stworzyć 2^{69} wielomianów częściowych, itd. Elementarne obliczenia dają więc podstawę stwierdzić, że praktyczne wykorzystanie kombinatorycznego GMDH, bez stosowania pewnych sposobów redukcji tworzonych modeli jest ograniczone niskimi rzędami wielomianów zupełnych, co istotnie obniża wydajność wykorzystania GMDH w praktyce.

Ze względu na omówione ograniczenia, badacze zajmujący się rozwojem GMDH (np. patrz [5], [8]) zaproponowali różne sposoby redukcji kombinatorycznych algorytmów GMDH, w których przestrzeń poszukiwanych wielomianów wielokrotnie się zmniejsza.

Jeden z tych sposobów bazuje na eksperymentalnie ustalonym fakcie, że w większości praktycznych problemów niezależne zmienne tworzą pewne grupy (klastry) o bliskich sobie wartościach zmiennych wyjściowych podlegających prognozowaniu. W tych, więc warunkach, zamiast korzystać ze wszystkich zmiennych wchodzących w skład danego klastra, w zupełności wystarczy wybrać jedną zmienną, jako przedstawiciela danego klastra, a resztę wyłączyć z analizy. Nazwijmy tych przedstawicieli klastrów *zmiennymi głównymi*. Niestety, ich wybór jest nierzadko sztuką dla ekspertów rozwiązujących konkretny problem, ponieważ jak dotychczas, nie znaleziono jednej metody określenia zmiennych głównych dla rozwiązania problemów prognozy. Zauważmy, że w informatyce znana jest uogólniająca metoda nazywana *analizą głównych składowych* (ang., *Principal Component Analysis*, *PCA*), która jest jedną ze statystycznych metod analizy czynnikowej. Kwestia optymalizacji GMDH z wykorzystaniem technik klasteryzacji danych pozostaje, więc na razie otwarta.

Drugi sposób pokonywania „przekleństwa” zbyt dużej wymiarowości wielomianów aproksymacyjnych i ich ilości polega na ewolucyjnej strategii doboru

najlepszych przedstawicieli wielomianów aproksymacyjnych na każdym kolejnym kroku tworzenia modeli w algorytmach wielowarstwowych. Idea jest następująca. Na początku tworzymy wszystkie wielomiany zawierające wszystkie kombinacje dwóch zmiennych wejściowych. Spośród nich wybieramy te wielomiany, które w klasie próbek uczących rozwiązują problem prognozy w najlepszy sposób, a resztę wielomianów odrzucamy. Na następnym etapie tworzymy wszystkie kombinacje z wielomianów wybranych na poprzednim etapie i wybieramy spośród nich te, które w klasie próbek uczących rozwiązują problem prognozy w sposób najlepszy, a resztę odrzucamy itd. Krok po kroku zbiór wielomianów zawęża się i ostatecznie otrzymamy kilka lub jeden wielomian. Schemat tych działań przedstawia Rys. 1.7. Na tym rysunku przez $\varphi_1(x_1, x_2)$, $\varphi_2(x_2, x_3)$, …, $\varphi_F(x_{N-1}, x_N)$ oznaczone są wielomiany o różnych kombinacjach dwóch zmiennych wejściowych (x_1, x_2), (x_2, x_3), …, (x_{N-1}, x_N) (gdzie N – ilość tych zmiennych, F – ilość wielomianów), tworzonych na początkowym etapie syntezy modeli prognozy; $\varphi_1{}^{(2)}$, $\varphi_2{}^{(2)}$, …, $\varphi_S{}^{(2)}$ – wielomiany tworzone na drugim etapie syntezy modeli (gdzie S – ilość tych wielomianów), itd. Proces doboru ewolucyjnego trwa dopóki, według pewnego kryterium, (lub kilku kryteriów) nie zostanie ustalony fakt, że znaleziono wielomian odpowiadający wymaganiom precyzji i złożoności. Zwrócimy uwagę, że opisany proces w ogóle nie gwarantuje, że model z poszukiwanymi charakterystykami uda się znaleźć.



Rys. 1.7 Schemat doboru ewolucyjnego najlepszych składników modelu predykcji
Źródło: Opracowanie własne

Inny sposób ewolucyjnego doboru „najlepszych" (w określonym sensie) modeli polega na tym, że na początku wybierany jest (według pewnych kryteriów) najlepszy pojedynczy wielomian drugiego rzędu z dwoma argumentami spośród wszystkich tworzonych wielomianów z dwoma argumentami. Do niego, kolejno, dobieramy trzeci

argument i wielomian rozszerzamy zwiększając ilość wchodzących w jego skład argumentów, i ewentualnie, czwarty itd., dopóki nie zostanie wybrany ostatni argument. Oczywiście, ilość etapów jest ograniczona ilością zmiennych wejściowych.

Są również inne schematy ewolucyjnej rozbudowy modelu GMDH [8]. Wspólną cechą i jednocześnie wspólną wadą tych sposobów pokonywania problemu dużych wymiarów wielomianów i dużej ilości tych wielomianów jest przypuszczenie, że złe początkowe modele nie mogą doprowadzić do dobrego wyniku. Dlatego wybrana została strategia odrzucania na każdym kolejnym etapie wszystkich tych modeli, które nie odpowiadają wymaganiom określonej dokładności, czyli wielomianów, które uznano za zbyt niedokładne.

Zasada ta została przeniesiona z natury ożywionej, gdzie jest znana jako jedno z praw doboru naturalnego: „słabi rodzice nie mogą reprodukować silnego potomstwa". W systemach sztucznych, natomiast, to prawo nie działa. Zostało to dobrze udowodnione w teorii algorytmów genetycznych (AG). Jak wiadomo [68], w AG została zastosowana strategia doboru ewolucyjnego, zgodnie z którą w każdej kolejnej generacji wraz z najsilniejszymi przedstawicielami z poprzedniego pokolenia zostawiana jest pewna ilość słabych poprzedników, którzy przez kilka pokoleń niekiedy są w stanie doprowadzić do reprodukcji najbardziej efektywnych potomków. Problemu dokładności modeli w algorytmach ewolucyjnych należy więc szukać nie we właściwym doborze początkowych modeli w pierwszych krokach ewolucyjnej samoorganizacji, a raczej w doskonałym doborze zbiorów zmiennych wejściowych i modeli otrzymanych w poprzednich krokach ich syntezy.

1.7 Wnioski: zadania rozwiązywane w rozprawie

Przedstawiona w danym rozdziale analiza zagadnienia prognozy daje podstawy do wyciągania pewnych wniosków, na podstawie których następnie określony zostanie przedmiot badań tej rozprawy.

Wnioski podsumowujące naszą analizę problemu prognozy, są następujące:

1. Napotykając granicę obliczalności przez maszynę Turinga, naukowcy jeszcze na początku rozwoju nauk informatycznych doszli do wniosku, że istnieje pewna klasa systemów, których badanie za pomocą symulacji matematycznej wymaga rozwoju platformy innej niż ta, która była oparta na zasadach dedukcji (tzw. problemy NP-trudne). Fenomen złożoności systemów został ujawniony nie tylko w dziedzinach

informatycznych, ale w szeregu innych nauk stosowanych, w których symulacja matematyczna leży u podstaw analizy, projektowania i planowania systemów o skomplikowanym charakterze zachowania: teorii automatów, systemów sterowania, ekonomii, analizie systemowej, bioinformatyce, naukach o Ziemi, socjologii i innych. Należy więc uznać, że rozwój metod symulacji wspomnianych systemów za pomocą zasobów informatycznych ma charakter interdyscyplinarny w tym sensie, że wymaga zastosowania pewnych wspólnych metod bazujących na abstrakcyjnych kategoriach określonych w informatyce jako „analiza danych" czy „algorytmy". Wśród najbardziej znanych fundamentalnych prac, w których zostały sformułowane problemy badania systemów złożonych, jako zagadnienie interdyscyplinarne można wspomnieć prace Wienera, Ashby'ego, Gödel'a, Turinga, Chomski'ego, Newell'a, Simon'a, Miller'a, von Neuman'a, Morgenstern'a, Baader'a i innych.

2. Z wykonanych przez wspomnianych autorów badań wynika wniosek, że metody analizy dowolnych złożonych systemów powinny przypominać program komputerowy. Właśnie działanie takiego programu musi być uporządkowane podobnie do instrukcji kodu komputerowego zawierającego instrukcji warunkowe „jeżeli→to", czyli funkcje wyboru optymalnej drogi rozwiązania postawionego problemu z szeregu alternatywnych rozwiązań na podstawie oceny stanów systemu i środowiska, w którym ten system działa. Optymalizacja wyboru drogi może być oceniona na podstawie kryteriów wygenerowanych przez ekspertów lub przez system, jeżeli ten system posiada własności inteligentne, zdolne do podejmowania decyzji. Do naukowego słownika wszedł termin „adaptacja", jako charakterystyka systemu zdolnego do samoorganizacji działań na podstawie zasad indukcji.
3. Zgodnie z twierdzeniem A. Turinga, system dokładnie symulujący zachowanie się obiektu jest tak samo skomplikowany, jak wspomniany obiekt. Wychodząc z założenia, że zachowanie się obiektu należącego do klasy obiektów złożonych, z reguły nie może być w 100% przewidziane, H. Simon sformułował tezę, że system podejmowania decyzji, który symuluje zachowanie się złożonego obiektu, musi działać na zasadzie satysfakcji, czyli powinien być zdolny do podejmowania wystarczająco dobrych, lecz nie koniecznie najlepszych decyzji, które mogą być praktycznie nieosiągalne w skończonym czasie działania tego systemu.
4. Na podstawie analizy problemów rozwiązywanych za pomocą systemów wspomagania decyzji (DSS) przychodzimy do wniosku, że problem badany w danej

rozprawie należy do klasy problemów ewolucyjnej analizy danych i procesów rozwiązywanych za pomocą DSS. Jest to problem rozwiązywany metodami klasyfikacji, regresji.

5. Wśród metod klasyfikacji i regresji obiecującymi są metody bazujące na wykorzystaniu sieci regularyzacyjnych i metod szeregów czasowych. Metody z pierwszej grupy jednak, oparte są na formalnym przypuszczeniu, że badany obiekt lub proces posiada własność inercyjności. Jest to całkiem dopuszczalne dla rozwiązania problemów interpolacji oraz ekstrapolacji procesów stacjonarnych lub nawet niestacjonarnych w przypadku powoli zmieniających się stanów zewnętrznego środowiska i obiektu badań. Natomiast metody szeregów czasowych są przeznaczone do rozwiązania zarówno problemów interpolacji, jak i problemów ekstrapolacji stanów obiektów, są więc najbardziej przydatne do rozwiązania problemów prognozy. Poza tym, jak udowodniono w pracach niektórych badaczy (np. [8]), metody szeregów czasowych oraz inne metody klasyfikacji i regresji należy raczej rozpatrywać jako wzajemnie uzupełniające się, a nie jako alternatywne podejścia, jeżeli chodzi o analizę zachowania się złożonych obiektów.
6. W związku z powyższym, należy podkreślić, że osobne miejsce w naszej klasyfikacji metod zajmują metody adaptacyjne. Zasadniczą cechą różniącą te metody od innych wspomnianych grup jest to, że są one oparte na przypuszczeniu, że obiekt lub proces podlegający badaniu może nie posiadać właściwości inercyjności. W przypadku rozwoju metod prognozy opartych na tej zasadzie, można przypuszczać, że okażą się one przydatne do rozwiązania problemów prognozy krótko-, średnio- i długoterminowej.
7. Metodą spełniającą wymagania adaptacji modeli przeznaczonych do symulacji zachowania się złożonych obiektów jest metoda grupowania argumentów (GMDH).
8. Zalety GMDH są następujące:
   8.1. Zasadnicza przydatność GMDH do rozwiązania problemów krótko-, średnio- i długoterminowej prognozy zachowania się systemów.
   8.2. Dostateczność modeli prognozy tworzonych na podstawie GMDH dla przewidywania wszystkich możliwych stanów badanych systemów (wynika ona z faktu, że modele GMDH bazują na wykorzystaniu dyskretnych analogów szeregów Volterry w postaci szeregów Kołmogorowa-Gabora, czyli wielomianów, a te ostatnie, według twierdzenia Stone'a-Weierstrassa, pokrywają wszystkie ewentualne stany obiektów podlegających symulacji).

8.3. Zdolność modeli tworzonych na podstawie GMDH do symulacji złożonych systemów, gdyż bazują one na koncepcji syntezy indukcyjnej, czyli koncepcji samoorganizacji ewolucyjnej modeli matematycznych (niezbędnej, jak pokazano wyżej, do symulacji właśnie systemów złożonych).

8.4. Tworzenie modeli z wykorzystaniem wielokryterialnego wyboru alternatywnych wersji tych modeli (co też jest konieczne w przypadku systemów złożonych).

8.5. Szeroki zakres zastosowań modeli tworzonych na podstawie GMDH.

9. Natomiast za cechy GMDH wymagające dodatkowych badań, należy uznać następujące:

9.1. Niedokładnie określone są pojęcia krótko-, średnio- i długoterminowych prognoz dla nieergodycznych i niestacjonarnych procesów, wskutek czego ciężko ocenić jest granice przewidywalności tworzonych modeli prognozy.

9.2. Podczas tworzenia modeli prognozy używane są informacje o przeszłych stanach obiektu. Otrzymany w taki sposób model tradycyjnie podlega korekcie jedynie na zbiorach danych uczących i testujących. Nasuwa się pytanie, czy nie można rozszerzyć zakresu prognozowania, jeśli wprowadzić by możliwość modyfikacji modeli nie tylko na etapie ich tworzenia, ale też bezpośrednio na etapie rozwiązania bieżącego problemu prognozy.

9.3. Zasady dopełnienia zewnętrznego i swobody wyboru Gabora potwierdzają konieczność wykorzystania precyzujących danych dla rozwiązania problemu regularyzacji poszukiwanych modeli. Co więcej, ta sama zasada dopełnienia zewnętrznego wymaga, żeby struktura operatorów regularyzujących miała postać hierarchiczną. Dochodzimy, więc do wniosku, że aktualnym jest problem tworzenia modeli prognostycznych typu niegödelowskiego, które w większym stopniu odpowiadają istocie zagadnienia prognozy w warunkach niepełnej określoności typów tych zjawisk i procesów, które badamy.

9.4. Realizacja jednowarstwowych (kombinatorycznych) algorytmów GMDH dla rozwiązania problemów prognozy na dużych zbiorach danych powiązana jest z olbrzymią ilością obliczeń. Jak twierdzą autorzy metody GMDH [8], mimo, że zaproponowano szereg skutecznych sposobów, zmniejszających objętość obliczeń, szybko osiągamy praktyczne granice obliczalności i podatności na ocenę przez ludzi tak olbrzymiej ilości informacji. Z tego powodu algorytmy jednowarstwowe ograniczone są do wykorzystania jedynie wielomianów niskiego rzędu na stosunkowo niedużych zbiorach danych wejściowych.

9.5. W przypadku, gdy rzędy wielomianów przekraczają granice obliczalności, pomocne jest zastosowanie indukcji niezupełnej, w której dzięki rozsądnemu wyborowi nośnika funkcji często udaje się utrzymać ilość wersji w rozsądnych granicach. Dla zmniejszania objętości obliczeń używane są różne techniki redukcji modeli. Przyjęte w praktyce podejście do tej redukcji polega na odrzucaniu na każdym etapie tworzenia modeli ich nieudanych wersji. Zasada ta wychodzi z założenia, że „słabi rodzice nie mogą reprodukować silnego potomstwa", która, jak udowodniono w teorii algorytmów genetycznych, jest błędna. Dlatego też wielowarstwowe GMDH nie gwarantuje syntezy nawet „dopuszczalnych" (w simonowskim sensie) modeli prognozy.

9.6. Z omówionych w podpunktach 9.1. i 9.2 przyczyn, otwarte pozostaje pytanie o wybór głównych zmiennych wśród danych wejściowych, które są wystarczające do tworzenia „dobrych", lecz nie koniecznie „najlepszych" modeli. Pewną podpowiedzią, co do sposobu rozwiązania tego problemu mogłaby być idea realizowana w analizie głównych składowych, jednak idea ta stosowana do metody GMDH jak dotychczas pozostaje nieanalizowana i wymaga dalszych badań.

Podsumowując wyniki przedstawionej wyżej analizy, można wyciągnąć wniosek, że aktualnym w informatyce jest problem tworzenia metod syntezy modeli prognozy zachowania systemów (obiektów, procesów) bazujących na zasadach indukcji, które przewidują ewolucyjną samoorganizację modeli i adaptację tworzonych modeli do zmieniających się warunków działania wspomnianych systemów. Zagadnienie to leży na przecięciu kilku kierunków informatyki: eksploracji danych, symulacji zachowania się obiektów, a także syntezy modeli obiektów różnej natury na podstawie zasad interpolacji i ekstrapolacji funkcji.

Są podstawy, by twierdzić, że metoda grupowania argumentów GMDH jest w stanie sprostać głównym wymaganiom stawianym metodom wspomnianego typu.

Celem danej rozprawy jest przedstawienie adaptacyjnej metody syntezy modeli przewidywania zachowania się złożonych systemów z wykorzystaniem GMDH, o jakości metody podstawowej, która będzie funkcjonowała również na małolicznych próbkach danych. Badania w danej rozprawie są skierowane na tworzenie metody osłabiającej ograniczenia GMDH, sformułowane wyżej w podpunktach 9.1 – 9.6.

2 Tworzenie i uzasadnienie metody prognozy na podstawie symulacji indukcyjnej

2.1 Ujęcie zagadnienia analizy procesów przy niewielkiej ilości danych eksperymentalnych

Szeregi czasowe są jednym z mechanizmów, za pomocą których możemy opisać dynamikę pewnych procesów, mogą więc stanowić ich modele. Procesy natomiast mogą mieć różne pochodzenie – mogą to być procesy zachodzące w ekonomii, socjologii, naturze, technologii, lub procesy opisujące zmiany stanów pewnych złożonych obiektów i systemów. Z różnych też powodów procesy te są analizowane, badane. Po pierwsze, w celu wyjaśnienia przyczyn zmian stanów systemu, określenia stopnia zależności procesów od wpływających na nie czynników, a także prognozy przyszłych stanów tego procesu – zagadnienie analizy procesów. Po drugie, analiza procesów może towarzyszyć rozwiązaniu problemu sterowania procesem za pomocą skierowanych zmian tych czynników, które najistotniej wpływają na zachowanie się procesu w czasie – zagadnienie syntezy modeli sterowania procesami.

Najbardziej popularne metody data mining służące rozwiązaniu określonych zagadnień opierają się na metodach statystycznych. Tzn. badacz, mając do dyspozycji wystarczającą ilość danych eksperymentalnych analizuje statystyczne parametry procesów (wartość oczekiwaną parametrów procesu, ich dyspersję[8], prawo, według którego zmienne opisujące stany procesów są rozłożone w czasie itd.) (Rozdział 1.3 rozprawy). Ponadto wśród klasycznych podejść do analizy procesów przypadkowych należy wspomnieć też specjalną teorię o tej samej nazwie (teoria procesów przypadkowych), która, jak i wspomniane metody data mining, opiera się na metodach teorii prawdopodobieństwa i statystyki matematycznej [69]. Wszystkie te podejścia funkcjonują w warunkach, gdy badacz ma do dyspozycji wystarczającą ilość danych eksperymentalnych, czyli wystarczającą ilość próbek.

Jednak, w praktyce badań procesów badacz ma niekiedy bardzo ograniczoną ilość eksperymentów a musi w ten czy inny sposób rozwiązać wspomniane wyżej problemy. Metody statystyczne okazują się w takich warunkach nieprzydatne, bowiem badacz nie dysponuje wystarczającą ilością danych, aby te metody zastosować. Jeśli przy okazji nieliczne próbki eksperymentalne, którymi dysponuje badacz zawierają dane, które

[8] Zróżnicowanie jednostek zbiorowości statystycznej z uwagi na pewną cechę mierzalną

charakteryzują się istotną nieregularnością (np. wskutek zaszumienia danych, niskiej dokładności pomiarów lub nieokreśloności niektórych danych w eksperymentach), to uzupełnienie ich braku oraz/lub zwiększenie regularności istniejących danych nie jest na ogół możliwe, bowiem, jak wiadomo z teorii identyfikacji systemów [1], żadne, nawet najbardziej wyrafinowane narzędzia matematyczne nie są w stanie pokonać problemu braku informacji.

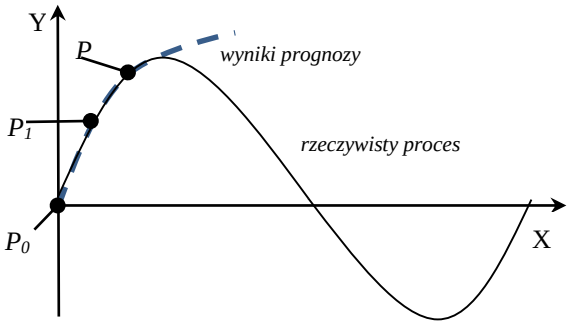
Natomiast, jeśli próbki mają ściśle określoną regularność, niski poziom błędów pomiarowych oraz praktycznie nie zawierają luk w danych, to te regularności w danych muszą pojawić się od razu zaczynając od pierwszych próbek aż do końca odcinka, na którym ten proces jest badany. Całkiem słuszne jest więc badanie zagadnienia tworzenia narzędzia matematycznego do analizy procesów w takich warunkach.

Celem naszych badań jest utworzenie takiego narzędzia matematycznego, które pozwala rozwiązać omówione wcześniej dwa zagadnienia na podstawie analizy niewielkiej ilości próbek (na krótkich zbiorach danych). U podstaw opracowywanej metody leżą głównie zasady deterministyczne (nie statystyczne).

Jednak nawet w przypadku, kiedy badany szereg jest szeregiem ze ściśle określoną regularnością, to nie mamy pewności, że ilość próbek będących do dyspozycji badacza jest wystarczająca do określenia wszystkich cech tego szeregu.

Przypuśćmy, że proces przedstawiony za pomocą pewnego szeregu, ma składnik cykliczny, który na małej ilości próbek będzie niewidoczny. Np. mamy periodyczny proces opisywany za pomocą zwykłej funkcji sinusoidalnej i trzy próbki, które dokładnie opisują wartości zmiennej *y* w trzech punktach P_0, P_1 oraz P_2, jak na rysunku Rys. 2.1.

Czy możliwe jest na podstawie jedynie tych trzech danych eksperymentalnych określić przyszły rozwój procesu, czyli ocenić czy proces ten jest periodyczny? Linia przerwana na Rys. 2.1. ilustruje, w jaki sposób prawdopodobnie przebiegałby ten proces w przyszłości, przewidywany za pomocą dowolnej metody prognozy.



Rys. 2.1 Przypuszczalny przebieg prognozy procesu periodycznego przy trzech danych wejściowych P_0, P_1, P_2
Źródło: Opracowanie własne

Podane punkty nie pozwalają praktycznie określić tej periodyczności. Konieczne jest posiadanie większej ilości punktów eksperymentalnych. Strategia badań w tym przypadku może być następująca. Na początku badań procesu, posługując się tymi danymi, którymi dysponujemy, tworzymy prognozę na jeden, dwa kroki czasowe wprzód. Dalej, otrzymując nowe dane eksperymentalne w kolejnych punktach, poprawiamy tworzony model i ponownie prognozujemy zachowanie się systemu na jeden, dwa kroki wprzód itd. Możemy napotkać następujące sytuacje: być może prognozowane stany systemu np. w punktach czasowych t_{i+1}oraz t_{i+2} będą na tyle bliskie wartościom rzeczywistym, że nie będzie potrzeby korekty utworzonego modelu na co najmniej kilku krokach wprzód. Może to oznaczać, że albo okres cyklicznych zmian systemu jest na tyle duży, że na krótkim odcinku czasu nie jest jeszcze widoczny, lub że okres ten już został uchwycony w naszym modelu. W obu przypadkach rozszerzamy zakres predykcji do ok. dwóch-trzech kroków czasowych i rozwiązujemy problem prognozy w tym zakresie korzystając ze starych modeli. Potem ponownie sprawdzamy dokładność modeli itd.

Jeśli natomiast po predykcji zachowania się procesu na kolejnych (jednym, dwóch) krokach t_{i+1}oraz t_{i+2} zauważamy istotne odchylenia wyników obliczeń od rzeczywistych wartości zmiennych procesu, oznacza to, że albo proces jest silnie zaszumiony i istotnie nieregularny, albo że nie jest zbyt zaszumiony i jest regularny, ale periodyczny o dużym okresie, który nie został jeszcze uchwycony w naszym modelu, bowiem zbiór próbek eksperymentalnych był zbyt mały. Wówczas modyfikujemy modele w tych nowych punktach czasowych o wysokich niedokładnościach i ponownie rozwiązujemy problem predykcji dla kolejnych punktów t_{i+3} oraz t_{i+4} itd. Ostatecznie, uzyskujemy modele, które zapewnią wymaganą dokładność.

Jak można zauważyć, ta strategia syntezy modeli predykcji jest strategią dynamiczną – tworzymy początkowy model, dalej sprawdzamy go i ponownie wykonujemy jeden, dwa kroki wprzód itd. Aby zrealizować taki proces, zbiór próbek musi zostać podzielony na dwie (niekoniecznie równe) części: podzbiór próbek inicjujących, na którym tworzymy początkowe modele, oraz podzbiór próbek testujących, na którym odbywa się korekta początkowych modeli:

$$Z_{próbek} = Z_{ini} + Z_{test}$$

Jak wspomniano powyżej, trudność praktycznej realizacji tej strategii polega na tym, że proces może być na tyle złożony, że do dyspozycji badacza będzie niewystarczająca ilość próbek (najczęściej niewystarczająca ilość próbek tego

pierwszego podzbioru Z_{ini}). W tej sytuacji oczywisty sposób rozwiązania powyższego problemu polega na tym, że jedną lub kilka próbek podzbioru Z_{test} przenosimy do podzbioru Z_{ini} i powtarzamy proces tworzenia modelu na rozszerzonym podzbiorze. W sprzyjającej sytuacji, gdy mamy do dyspozycji nowe próbki, ich kosztem uzupełniamy podzbiór Z_{test} i powtarzamy proces testowania. W mniej sprzyjającej sytuacji, gdy musimy tworzyć model na podstawie jedynie istniejących próbek, dokładność tworzonych modeli może być niska. Niestety jest to zasadnicza wada analizy procesów na krótkich zbiorach próbek, która uwarunkowana jest istotą obiektu badań, a nie niską efektywnością stosowanych metod analizy procesów. Problem niedoboru danych może być częściowo lub w pełni pokonywany poprzez uzupełnienie podzbioru Z_{ini} próbkami z podzbioru Z_{test}.

W odróżnieniu od statystycznych metod budowy modeli, opartych na analizie wystarczająco bogatych zbiorów danych, w warunkach małej ilości eksperymentów badacz tworzy model deterministyczny nie mając gwarancji, że dane, którymi ten badacz dysponuje, obejmują wszystkie ewentualne stany badanego obiektu, które mogą pojawić się w przyszłości. Ze względu na te okoliczności, najprawdopodobniej jedyna możliwa strategia ulepszenia dokładności tworzonych modeli polega na tym, że nowe (w tym – niewystarczające) dane pobieramy na bieżąco, w trakcie bezpośredniego rozwiązywania zagadnienia predykcji, tzn. badacz musi od czasu do czasu porównywać wyniki predykcji z rzeczywistymi wartościami zmiennych opisujących dany proces, uzyskanymi w trakcie rozwiązywania problemu. Ta właśnie strategia leży u podstaw metody, rozwijanej w danej pracy.

2.2 Zagadnienie prognozy zachowania procesów

Przypuśćmy, że mamy pewien obiekt, którego stany opisane są przez *n* zmiennych niezależnych $x_1(t), x_2(t), \dots, x_n(t)$ charakteryzujących stany tego procesu (obiektu, systemu) w różnych chwilach czasu *t.* Zmienne te nazywamy **zmiennymi stanów** procesu (obiektu, systemu). (Na razie przyjmujemy określenie zmiennych, jako „zmiennych stanów" bez precyzowania właściwości modeli stanów. Szczegóły modeli stanów opisane są w Rozdziale 3. tej rozprawy) Przypuśćmy, że jedyne informacje, którymi dysponujemy o tym obiekcie, to wartości zmiennych $x_1(t), x_2(t), \dots, x_n(t)$ w dyskretnych chwilach czasu $t_0, t_1, t_2, \dots, t_N$ uzyskane, np. z pomiarów lub jako uśrednione dane statystyczne. Jeśli ilość tych zmiennych wynosi *n*, to zmienne stanów opisują stan procesu (obiektu

itd.) w metrycznej przestrzeni *n*-wymiarowej wartości rzeczywistych $(x_1, x_2, \dots, x_n) \epsilon\, \Re^n$. Metrycznej, ponieważ odległość jednego punktu od drugiego w tej przestrzeni można określić za pomocą wprowadzonej w ten czy inny sposób metryki. Zagadnienie polega na tym, aby na podstawie zbiorów obserwowanych danych $x_1(t), x_2(t), \dots, x_n(t)$ określonych w minionych dyskretnych chwilach czasu $t_0, t_1, t_2, \dots, t_N$, utworzyć matematyczny model obiektu, który z dopuszczalnym błędem pozwala określić wartości poszczególnych zmiennych $x_1(t), x_2(t), \dots, x_n(t)$ w przyszłych dyskretnych chwilach czasu $t_{N+1}, t_{N+2}, \dots, t_{N+K}$, gdzie K nazywamy *przedziałem czasowego wyprzedzenia.*

Zasadniczym założeniem, które musi zostać w naszym zagadnieniu spełnione, jest to, że procesy opisywane przez zmienne $x_1(t), x_2(t), \dots, x_n(t)$ powinny podlegać pewnym regularnościom, które mogą zostać ujawnione poprzez eksperymenty i opisane za pomocą tworzonych modeli matematycznych. Oczywiście, do kategorii takich procesów można zaliczyć procesy ergodyczne, stacjonarne lub niestacjonarne, ale takie, w których obserwowana jest pewna powtarzalność stanów. Przykładem procesu, który w zasadzie nie może być prognozowany, jest tzw. „szum biały".

Zmienne $x_1(t), x_2(t), \dots, x_n(t)$ można interpretować jako składniki *n*-wymiarowego wektora

$$X = \begin{bmatrix} x_1(t) \\ x_2(t) \\ \dots \\ x_n(t) \end{bmatrix}. \qquad (2.1)$$

Wektor X określony w chwili czasu t_0 oznaczamy przez $X(0)$, w chwili czasu t_1 – przez $X(1)$, ..., w chwili czasu t_N – przez $X(N)$ itd. W przedstawionym zagadnieniu obserwowane dane $x_1(t), x_2(t), \dots, x_n(t)$ są grupowane w postaci *N+1* wektorów: $X(0)$, $X(1)$, $X(2)$,..., $X(N)$ w *n*-wymiarowej przestrzeni zmiennych $x_1(t), x_2(t), \dots, x_n(t)$. Kontynuując interpretację geometryczną, można uznać, że każdy wektor $X(0)$, $X(1)$, $X(2)$,..., $X(N)$ określa pewien punkt w *n*-wymiarowej przestrzeni zmiennych $x_1(t), x_2(t), \dots, x_n(t)$, a te ostatnie są współrzędnymi tychże punktów.

Uogólnione wyrażenie dla poszukiwanego modelu w postaci wektorowej może być przedstawione następująco:

$$X(N+k) = F\big(X(0), X(1), X(2), \dots, X(N)\big), \qquad k = 1, 2, \dots, K, \quad (2.2)$$

gdzie $X(N+1), X(N+2), \dots, X(N+K)$ to wektory X określone w przyszłych chwilach czasu, odpowiednio: $t_{N+1}, t_{N+2}, \dots, t_{N+K}$, F – pewne przekształcenie.

Jak można wnioskować z przeglądu problematyki eksploracji danych przedstawionego w Rozdziale 1. tej rozprawy, omówione zagadnienie przypomina problemy klasyfikacji i regresji, w tym problem prognozy szeregów czasowych. Nie mniej zwróćmy uwagę, że w przedstawionej wyżej postaci zagadnienie to zasadniczo różni się od klasycznego zagadnienia klasyfikacji i regresji, gdyż:

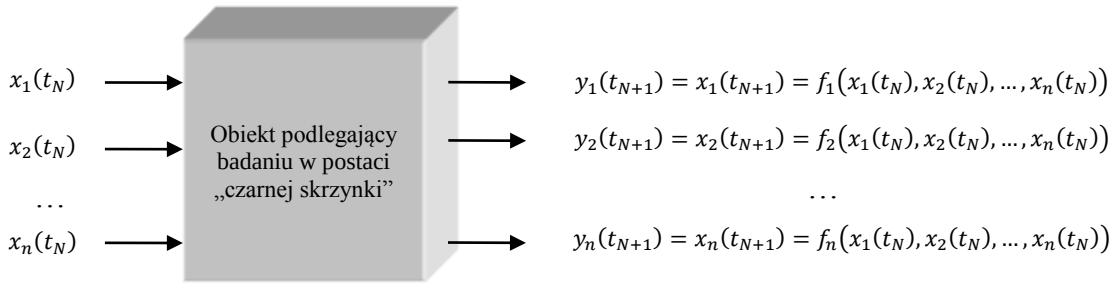
1. W odróżnieniu od „klasycznego" ujęcia problemu prognozy, będziemy zakładali, że mamy do czynienia ze stosunkowo niewielką ilością próbek danych, tzn. ilość punktów obserwacji *N+1* jest mała, np. $N < 10$.
2. Procesy podlegające badaniom mogą być zarówno stacjonarne, jak i niestacjonarne, ale takie, dla których w ogólności może być zbudowany matematyczny model, odzwierciedlający z wystarczającą dokładnością zmiany stanów na krótkim czasowym odcinku obserwacji.
3. Postać zależności funkcjonalnej między zmiennymi $x_1(t), x_2(t), \dots, x_n(t)$ jest nieznana, ale może zostać określona klasa modeli (np. wielomianów dowolnego rzędu i dowolnej nieliniowości), pozwalających z wymaganą dokładnością symulować stany badanego obiektu.

W tych warunkach stosowanie klasycznych metod analizy statystycznej, np. analizy regresji lub analizy dyspersji może prowadzić do błędnych wniosków. Jak udowodniono w pracach [1], [2], [3] i wspomniano w Rozdziale 1., podejście przydatne w tych warunkach, może być oparte na koncepcji indukcyjnej samoorganizacji modeli. Dlatego w tej rozprawie jako podstawowa została wybrana metoda GMDH [50], która jest reprezentantem takiego podejścia. Zakłada ona działanie na danych, złożonych procesach i przewidywanie ich przyszłych wartości bez wykorzystywania dodatkowych założeń i informacji o systemie z zewnątrz, oraz ich wewnętrznych związkach, korelacjach.

Wykorzystana przez nas do celów budowy modeli predykcji metoda GMDH to wymyślona 46 lat temu, przez ukraińskiego naukowca A. G. Ivakhnenkę, metoda maszyn uczących ze stopniowo komplikującymi się modelami aż do znalezienia optymalnej jego złożoności. Tworzy się w niej wielowarstwowe modele podobne do sieci neuronowych. To typowa metoda indukcyjnego modelowania, jedna z efektywniejszych metod identyfikacji strukturalno-parametrycznej obiektów i procesów (systemów) złożonych z danych w warunkach niekompletnej informacji. GMDH posiada wielką różnorodność możliwości w porównaniu z innymi metodami o podobnej konstrukcji. Dotyczy to przede wszystkim sposobów generowania modeli i kryteriów oceny ich jakości.

Do dalszych badań wygodniej będzie przedstawić obiekt podlegający analizie

w postaci „czarnej skrzynki”, czyli obiektu, którego architektura oraz funkcjonalne zależności między zmiennymi $x_1(t), x_2(t), \dots, x_n(t)$ nie są znane. Przy czym, zmienne po prawej stronie wyrażenia (2.2) można utożsamić z „sygnałami wejściowymi” czarnej skrzynki tworzącymi wektor sygnałów wejściowych $X(t)$, a zmienne po lewej stronie – z „sygnałami wyjściowymi” tworzącymi wektor sygnałów wyjściowych $Y(t)$ (Rys. 2.2).



Rys. 2.2 Obiekt podlegający badaniu w postaci „czarnej skrzynki”
Źródło: Opracowanie własne

Np. jeśli mając do dyspozycji zmienne $x_1(t_N), x_2(t_N), \dots, x_n(t_N)$ określone w dyskretnej chwili czasu t_N, szukamy wartości tych zmiennych w następnej dyskretnej chwili czasu t_{N+1}, to formalnie sytuację tę można przedstawić w taki sposób, że na wejściu „czarnej skrzynki” podane są sygnały tworzące wejściowy wektor $X(t_N) = \left(x_1(t_N), x_2(t_N), \dots, x_n(t_N)\right)^T$ (gdzie „*T*” - to znak transpozycji wektora), a na wyjściu „czarnej skrzynki” otrzymamy sygnały wyjściowe tworzące wyjściowy wektor $Y(t_{N+1}) = \left(y_1(t_{N+1}), y_2(t_{N+1}), \dots, y_n(t_{N+1})\right)^T$, gdzie:

$y_1(t_{N+1}) = x_1(t_{N+1}), y_2(t_{N+1}) = x_2(t_{N+1}), \dots, y_n(t_{N+1}) = x_n(t_{N+1})$

W nieco zmodyfikowanej postaci nasze zagadnienie można więc określić jako zagadnienie syntezy matematycznych modeli wejścia-wyjścia dla każdej zmiennej wyjściowej y_i, i = 1, …, n obiektu formalnie przedstawionego w postaci tzw. wielobiegunnika o *n* wejściach i *n* wyjściach (w ogólności ilość wejść nie musi być równa ilości wyjść).

Przy czym sygnały wejściowe określone są w punktach czasowych $t = t_0, t_1, t_2, \dots, t_N$ (gdzie *N* zależy od wymaganej dokładności modeli i osobliwości zmian zmiennych $x_1(t), x_2(t), \dots, x_n(t)$, od tego, czy są argumentami procesu stacjonarnego czy niestacjonarnego, liniowego czy nieliniowego, jak bardzo są zaszumione, na ile dużo niosą ze sobą informacji by znaleźć poszukiwane regularności itd.); sygnały wyjściowe określone są w punktach czasowych $t = t_{N+1}, \dots, t_{N+K}$. W takiej właśnie postaci tworzy się modele w metodzie GMDH.

Ważną okolicznością, która będzie miała zasadnicze znaczenie w dalszej realizacji modeli, jest to, że w modelu (2.2) wyrażenie po prawej stronie określone jest dla chwili czasu o jeden i więcej kroków czasowych wstecz w porównaniu ze składnikami wektora przedstawionego po stronie lewej. W szczególności, przedział prognozy $[t_N, t_{N+K}]$, na którym zasadniczo model może funkcjonować, zależy od wspomnianych wyżej osobliwości procesów podlegających analizie (stacjonarne bądź niestacjonarne, liniowe lub nieliniowe itd.).

W ogólnym przypadku składniki wektora *Y* należy uważać za funkcje wszystkich składników wektora *X:*

$$\begin{aligned} y_1 &= f_1(x_1, x_2, \dots, x_n); \\ y_2 &= f_2(x_1, x_2, \dots, x_n); \\ &\dots \\ y_n &= f_n(x_1, x_2, \dots, x_n); \end{aligned} \tag{2.3}$$

Omówiony wyżej problem, polega więc na znalezieniu matematycznych wzorów dla $f_1(), f_2(), \dots, f_n()$, jako funkcji argumentów $x_1(t), x_2(t), \dots, x_n(t)$.

2.3 Matematyczne zasady metody prognozy na podstawie symulacji indukcyjnej

Przystąpmy do opisu proponowanej metody przeznaczonej do rozwiązania omówionego problemu. Będziemy szukali wzorów matematycznych dla wspomnianych funkcji (2.3) w klasie modeli wielomianowych, w postaci wielomianów Kołmogorowa-Gabora będących dyskretnym analogiem szeregów Volterry [70], wykorzystywanych w klasycznej metodzie GMDH [50] (2.4).

$$Y = a_0 + \sum_{i=1}^{n} a_i x_i + \sum_{j=1}^{n} \sum_{i \le j} a_{ij} x_i x_j \sum_{i=1}^{n} \sum_{j \le i} \sum_{k \le j} a_{ijk} x_i x_j x_k + \cdots, \tag{2.4}$$

gdzie współczynniki $a_0, a_i, a_{ij}, a_{ijk}$ to wartości nieznane, podlegające obliczaniu, a *Y* to dowolna zmienna wyjściową $y_1(t), y_2(t), \dots, y_n(t)$.

Spośród wszystkich możliwych wersji wielomianów Kołmogorowa-Gabora (WKG), w których może zostać uwzględniona współzależność zmiennych $x_1(t), x_2(t), \dots, x_n(t)$, jednym z prostszym jest WKG drugiego rzędu o postaci:

$$Y = a_0 + a_1 x_1 + a_2 x_2 + a_3 x_1^2 + a_4 x_2^2 + a_5 x_1 x_2, \tag{2.5}$$

w którym dla uproszczenia korzystać będziemy z jednowymiarowej notacji

współczynników $a_0, a_1, \ldots$ itd. Ten typ wielomianu wybieramy też jako podstawowy model dla naszych dalszych badań.

Zauważmy, że we wzorach (2.4), (2.5) przewiduje się, że zmienne $x_i, x_j, x_k, \ldots$, przedstawione po prawej stronie tych równań, określone są w punkcie czasowym tylko o jeden krok wstecz w porównaniu ze zmienną *Y*. Ponadto, wyrażenie (2.5) jest funkcją tylko dwóch zmiennych, a nie wszystkich zmiennych, jak to powinno być w ogólnym przypadku (2.4). Dlatego model (2.5) nie odpowiada uogólniającemu modelowi (2.4), z tego też powodu będziemy nazywali wyrażenie (2.5) *cząstkowym modelem prognozy*. Aby uwzględnić zależności funkcji wyjściowych *Y* od wszystkich zmiennych wejściowych w różnych chwilach czasu, w metodzie GMDH przewiduje się, po pierwsze – wykorzystanie kombinacji różnych par zmiennych wejściowych $\{x_1,\ x_2\}, \{x_1,\ x_3\}, \ldots,$ $\{x_1,\ x_n\},\ \{x_2,\ x_3\}, \ldots, \{x_2,\ x_n\}, \ldots, \{x_{n-1},\ x_n\}$ (takich kombinacji, czyli cząstkowych modeli, dla *n* zmiennych wejściowych będzie C_n^2), i po drugie – stopniowe nawarstwianie modelu w procesie przejścia z jednej chwili czasowej do następnej z przechowaniem modeli otrzymanych w poprzednich krokach.

Kontynuując utożsamianie naszego problemu z analizą „czarnej skrzynki", wprowadzimy pojęcie *wrażliwości zmiennych* wyjściowych modelu na małe zmiany zmiennych wejściowych. Po raz pierwszy pojęcie funkcji wrażliwości zmiennych wyjściowych od zmiennych stanów obiektu były proponowane w fundamentalnej pracy *Bode* z teorii wzmacniaczy elektronicznych ze sprzężeniem zwrotnym [71]. Idee te były później rozwinięte w pracach z teorii układów elektronicznych i teorii systemów sterowania [72]. Funkcja wrażliwości zmiennej wyjściowej (np. sygnału wyjściowego) $f(x_1, \ldots, x_i, \ldots, x_n)$ od zmiennej wejściowej x_i może być opisana jednym z następujących wzorów:

- względna funkcja wrażliwości pierwszego rzędu

$$S_{x_i}^f = \frac{\frac{\partial f}{f}}{\frac{\partial x_i}{x_i}}, \tag{2.6}$$

czyli stosunek względnego przyrostu funkcji *f* do względnego przyrostu argumentu x_i, gdy ten ostatni dąży do zera. Zakłada się, że, po pierwsze, funkcja *f* jest różniczkowalna względem argumentu x_i w punkcie, w którym jest określona pochodna $\frac{\partial f}{\partial x_i}$, po drugie, ani wartość funkcji *f*, ani wartość argumentu x_i nie zeruje się w otoczeniu punktu określoności funkcji wrażliwości. Natomiast, jeżeli funkcja *f* zeruje się w tym otoczeniu,

to dla określenia funkcji wrażliwości może być stosowana następująca funkcja:
- półwzględna funkcja wrażliwości pierwszego rzędu

$$S_{x_i}^f = \frac{\frac{\partial f}{\partial x_i}}{x_i} = \frac{\partial f}{\partial x_i} \cdot x_i, \tag{2.7}$$

dla której zakłada się, że funkcja f jest różniczkowalna względem argumentu x_i w punkcie, w którym jest określona pochodna $\frac{\partial f}{\partial x_i}$, oraz wartość argumentu x_i nie zeruje się w otoczeniu punktu określoności funkcji wrażliwości. Natomiast, jeżeli ostatni warunek nie jest spełniony, wówczas można korzystać z następującej funkcji:
- różniczkowa funkcja wrażliwości pierwszego rzędu

$$S_{x_i}^f = \frac{\partial f}{\partial x_i}, \tag{2.8}$$

gdzie jedynym warunkiem stosowalności jest różniczkowalność funkcji f względem argumentu x_i.

Aby pokonać problemy związane ze wspomnianymi wyżej ewentualnymi wartościami zerowymi funkcji f i/lub zmiennej x_i można zastosować sztuczną metodę, czyli, dodać do funkcji f (argumentu x_i) pewną stałą wartość, a po obliczeniach odjąć tę wartość od otrzymanego wyniku. Nie jest to, więc zasadniczo ograniczeniem dla wykorzystania funkcji (2.6), (2.7) w praktycznych obliczeniach. Dlatego dalej będziemy korzystać z definicji (2.7), chociaż wszystkie otrzymane dalej rezultaty bez trudu można uzyskać korzystając z funkcji (2.6), (2.8).

Jak można zauważyć, wspólną częścią tych wzorów jest pierwsza pochodna funkcji f względem zmiennej x_i. Funkcje wrażliwości określone za pomocą pierwszych pochodnych należą do klasy *funkcji wrażliwości pierwszego rzędu.* Im większa jest bezwzględna wartość dowolnej z tych funkcji wrażliwości dla konkretnej zmiennej x_i, tym bardziej zmienia się wartość funkcji f podczas zmiany wartości argumentu x_i.

Korzystając z drugich pochodnych funkcji f względem zmiennych x_i i x_j, w podobny sposób można określić funkcje wrażliwości drugiego rzędu. Przedstawimy wzory dla tych funkcji na przykładzie półwzględnej funkcji wrażliwości drugiego rzędu:
- półwzględna funkcja wrażliwości drugiego rzędu dla zmiennej x_i,

$$S_{x_i^2}^f = \frac{\partial^2 f}{\partial x_i^2} \cdot x_i^2, \tag{2.9}$$

-mieszana półwzględna funkcja wrażliwości drugiego rzędu dla zmiennych x_i i x_j

$$S^f_{x_i x_j} = \frac{\partial^2 f}{\partial x_i \partial x_j} \cdot x_i \cdot x_j, \tag{2.10}$$

W przypadku wykorzystania modelu WKG w postaci (2.5), wzory dla funkcji (2.7), (2.9), (2.10) wyglądają następująco:

$$S^Y_{x_1} = \frac{\partial Y}{\partial x_1} \cdot x_1 = (a_1 + 2a_3 x_1 + a_5 x_2) \cdot x_1; \tag{2.11a}$$

$$S^Y_{x_2} = \frac{\partial Y}{\partial x_2} \cdot x_2 = (a_2 + 2a_4 x_2 + a_5 x_1) \cdot x_2; \tag{2.11b}$$

$$S^Y_{x_1^2} = \frac{\partial^2 Y}{\partial x_1^2} \cdot x_1^2 = (2a_3) \cdot x_1^2; \tag{2.11c}$$

$$S^Y_{x_2^2} = \frac{\partial^2 Y}{\partial x_2^2} \cdot x_2^2 = (2a_4) \cdot x_2^2; \tag{2.11d}$$

$$S^Y_{x_1 x_2} = \frac{\partial^2 Y}{\partial x_1 \partial x_2} \cdot x_1 \cdot x_2 = a_5 \cdot x_1 \cdot x_2; \tag{2.11e}$$

Pokażemy dalej, że funkcje wrażliwości mogą być stosowane w cząstkowych modelach prognozy.

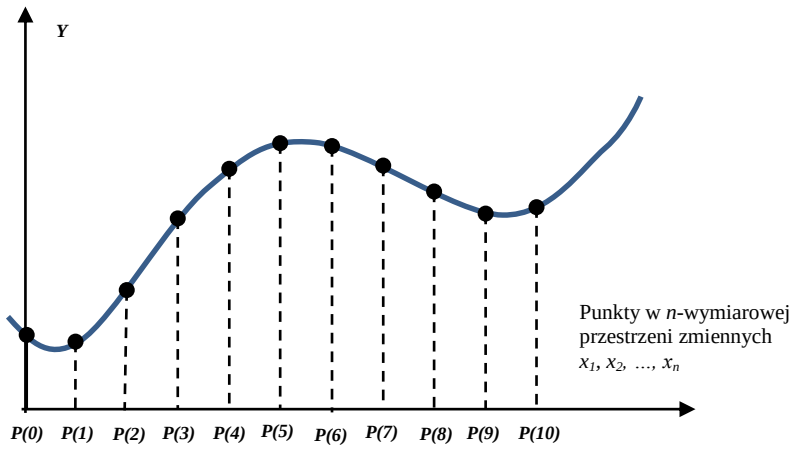
Niech ogólna próbka, z której wybieramy dane do tworzenia poszukiwanych modeli, składa się z szeregu pomiarów dokonywanych w chwilach czasu $t_0, t_1, t_2, \dots, t_N$. Oznaczmy te pomiary odpowiednio przez $P(0)$, $P(1)$, $P(2)$, ..., $P(N)$. Niech dla każdego z tych pomiarów, wartości wszystkich zmiennych $x_1(t), x_2(t), \dots, x_n(t)$ są znane. Aby zilustrować poglądowo ideę metody, przedstawimy nasze zagadnienie w następującej postaci. Przypuśćmy, że mamy pewną funkcję Y dwóch argumentów x_1 i x_2. Zarówno wartości tej funkcji, jak i wartości jej argumentów są znane (np. na podstawie dokonanych pomiarów). Każdy pomiar i symbolicznie oznaczamy jako pewien punkt $P(i)$ w przestrzeni (Y, x_1, x_2), czyli otrzymujemy punkty $P(0)$, $P(1)$, $P(2)$,... itd. (patrz Rys. 2.3).

W rozpatrywanym zagadnieniu pod pojęciem funkcji Y rozumiemy dowolną funkcję ze zbioru $\{y_1(t), y_2(t), \dots, y_n(t)\}$.

Dysponując wartościami zmiennych x_1, x_2 w punktach $P(0)$ i $P(1)$, w tym wartościami funkcji Y w tych punktach, czyli wartościami Y_0 i Y_1, spróbujemy utworzyć funkcję, która prognozuje przyrost wartości tej funkcji $\Delta Y = Y_2 - Y_1$, czyli spróbujemy ocenić, o ile zmienia się funkcja Y przy przejściu od punktu $P(1)$ do kolejnego punktu $P(2)$. Korzystając z rozwinięcia Taylora i ograniczając się pochodnymi drugiego rzędu, otrzymujemy:

$$Y_2 = Y_1 + \frac{\partial Y}{\partial x_1}\cdot(\Delta x_1)\Big|_{P(1)} + \frac{\partial Y}{\partial x_2}\cdot(\Delta x_2)\Big|_{P(1)} + \frac{1}{2}\cdot\frac{\partial^2 Y}{\partial x_1^2}\cdot(\Delta x_1)^2\Big|_{P(1)} + \\ + \frac{1}{2}\cdot\frac{\partial^2 Y}{\partial x_2^2}\cdot(\Delta x_2)^2\Big|_{P(1)} + \frac{\partial^2 Y}{\partial x_1 \partial x_2}\cdot(\Delta x_1)\cdot(\Delta x_2)\Big|_{P(1)} + R_2 \tag{2.12}$$

gdzie R_2 jest resztą rozwinięcia Taylora; $\Delta x_i|_{P(1)} = x_i^{(1)} - x_i^{(0)}$, i = 1, 2 to różnica wartości zmiennej x_i w punkcie $P(1)$ i wartości tej zmiennej w punkcie $P(0)$; wszystkie pochodne obliczane są w punkcie $P(1)$.



Rys. 2.3 Przedstawienie zależności funkcji *Y* od punktów, w których dokonywane są pomiary
Źródło: Opracowanie własne

Porównując składniki prawej strony równania (2.12) z definicjami różniczkowych funkcji wrażliwości (2.8), nietrudno dostrzec, że wyrażenie (2.12) można przedstawić w terminach funkcji wrażliwości następująco:

$$Y_2 \approx Y_1 + S_{x_1}^{Y}\Delta x_1|_{P(1)} + S_{x_2}^{Y}\cdot\Delta x_2|_{P(1)} + \frac{1}{2}\cdot S_{x_1^2}^{Y}\cdot(\Delta x_1)^2|_{P(1)} + \\ + \frac{1}{2}\cdot S_{x_2^2}^{Y}\cdot(\Delta x_2)^2|_{P(1)} + S_{x_1 x_2}^{Y}\cdot(\Delta x_1)\cdot(\Delta x_2)|_{P(1)} \tag{2.13}$$

W miejscu różniczkowych funkcji wrażliwości możemy użyć również względnych (2.6) czy półwzględnych (2.7) funkcji wrażliwości. Podstawiając zamiast funkcji wrażliwości wyrażenia (2.12), otrzymamy równanie:

$$Y_2 \approx Y_1 + (a_1 + 2a_3x_1 + a_5x_2)\cdot\Delta x_1|_{P(1)} + (a_2 + 2a_4x_2 + a_5x_1)\cdot\Delta x_2|_{P(1)} \\ + \frac{1}{2}\cdot(2a_3)\cdot(\Delta x_1)^2|_{P(1)} + \frac{1}{2}\cdot(2a_4)\cdot(\Delta x_2)^2|_{P(1)} + (a_5)\cdot(\Delta x_1)\cdot(\Delta x_2)|_{P(1)}, \tag{2.14}$$

które wyraża zależność wartości zmiennej *Y* od wartości przyrostów argumentów x_1 i x_2, przez poszukiwane współczynniki $a_0, a_1, \ldots, a_5$. Otrzymaliśmy, więc matematyczne wyrażenie przedstawiające zależność wartości funkcji uzyskanej w pewnym punkcie czasowym od wartości uzyskiwanych w poprzednim punkcie czasowym.

Model cząstkowy (2.5) zawiera sześć współczynników $a_0, a_1, \dots, a_5$, więc aby jednoznacznie wyznaczyć te współczynniki należy posiadać co najmniej sześć zgodnych ze sobą równań. W szczególności, tymi równaniami mogą być równania typu (2.5) sformułowane odpowiednio dla punktów czasowych $P(1)$, $P(2)$, $P(3)$, $P(4)$, $P(5)$, $P(6)$. Będziemy uznawali, że pewien model cząstkowy jest właściwy w matematycznym sensie, jeśli przy tych samych współczynnikach $a_0, a_1, \dots, a_5$ z dopuszczalnym błędem odpowiada wartościom oraz przyrostom funkcji Y i zmiennych x_1, x_2 we wszystkich wspomnianych punktach. W celu znalezienia współczynników $a_0, a_1, \dots, a_5$ korzystamy więc z następującego układu równań:

$$\begin{aligned}
&(1)\ Y_1 = a_0 + a_1x_1(0) + a_2x_2(0) + a_3x_1^2(0) + a_4x_2^2(0) + a_5x_1(0)x_2(0)\\
&(2)\ Y_2 = a_0 + a_1x_1(1) + a_2x_2(1) + a_3x_1^2(1) + a_4x_2^2(1) + a_5x_1(1)x_2(1)\\
&(3)\ Y_3 = a_0 + a_1x_1(2) + a_2x_2(2) + a_3x_1^2(2) + a_4x_2^2(2) + a_5x_1(2)x_2(2)\\
&(4)\ Y_4 = a_0 + a_1x_1(3) + a_2x_2(3) + a_3x_1^2(3) + a_4x_2^2(3) + a_5x_1(3)x_2(3)\\
&(5)\ Y_5 = a_0 + a_1x_1(4) + a_2x_2(4) + a_3x_1^2(4) + a_4x_2^2(4) + a_5x_1(4)x_2(4)\\
&(6)\ Y_6 = a_0 + a_1x_1(5) + a_2x_2(5) + a_3x_1^2(5) + a_4x_2^2(5) + a_5x_1(5)x_2(5).
\end{aligned} \tag{2.15}$$

A następnie uzyskane rezultaty stosujemy w modelach typu (2.14).

Będziemy, więc korzystać z wartości eksperymentów uzyskanych w siedmiu kolejnych punktach czasowych $t_0, \dots, t_6$ w celu obliczenia wartości w punkcie czasowym t_7, ewentualnie kolejnych.

W naszych badaniach korzystamy z następującej notacji:

x, y, z, ... – to zmienne stanów, zmienne procesu (zamiast x_1, x_2, ... – dla uproszczenia zapisu);

$X(x, y)$ – to wartość zmiennej *x* w kolejnej chwili czasu, przedstawiona jako funkcja cząstkowa tylko dwóch zmiennych (*x* oraz *y*), których wartości określone są na poprzedniej chwili czasu; formalnie przedstawiamy *X* jako funkcję zmiennych *x, y*;

$X(x, z)$ – analogicznie, ale argumentami są *x, z*;

$Y(y, x)$ – podobnie, ale obliczamy prognozowaną wartość zmiennej *y* jako funkcję argumentów *x, y;* itd.

W eksperymentach przedstawionych dalej korzystamy z półwzględnych funkcji wrażliwości. Natomiast główna przewaga względnych funkcji wrażliwości polega na tym, że posługujemy się tam wartościami przyrostów względnych (a nie bezwzględnych) zmiennych *x, y, z, ...* : $ddx = \frac{\Delta x}{x}$, $ddx = \frac{\Delta y}{y}$, $ddx = \frac{\Delta z}{z}$, ... itd.

Jest to wygodne, bowiem w rozwiązaniach praktycznych problemów, wartości różnych zmiennych mogą się różnić w zakresie kilku rzędów. Może to doprowadzić do ignorowania małych wartości, a więc, większych błędów w obliczeniach numerycznych. Przypuśćmy, że zmienna *x* zmienia się w zakresie 1000±100, a druga zmienna *y* – w zakresie 1±0.1, wtedy bezwzględne przyrosty pierwszej zmiennej *x* będą interpretowane jako dominujące, a zmiany drugiej zmiennej będą prawdopodobnie ignorowane. Aby wyrównać wagi tych zmian, korzysta się z wartości względnych. Wówczas przyrost ±100 pierwszej zmiennej (czyli 10% od jej początkowej wartości) będzie miał taką samą wagę, jak drugiej zmiennej (0.1 – czyli też 10%).

W naszej terminologii wspomniane funkcje $X(\,)$, $Y(\,)$, ... itd. są funkcjami cząstkowymi, bowiem nie zawierają wszystkich argumentów, a jedynie różne kombinacje par argumentów (dokładnie tak samo, jak w metodzie GMDH na pierwszym kroku budowy modelu).

$$S_x^{X(x,y)} = \frac{\partial X(x,y)}{\partial x} \cdot x \tag{2.16}$$

– funkcja wrażliwości pierwszego rzędu funkcji $X(x,y)$ względem argumentu *x* (uważamy, że wartość zmiennej *x* jest znana w tym wyrażeniu);

$$S_y^{X(x,y)} = \frac{\partial X(x,y)}{\partial y} \cdot y \tag{2.17}$$

– funkcja wrażliwości pierwszego rzędu funkcji $X(x,y)$ względem argumentu *y* (uważamy, że wartość zmiennej *y* jest znana w tym wyrażeniu);

$$S_{x^2}^{X(x,y)} = \frac{\partial^2 X(x,y)}{\partial x^2} \cdot x^2 \tag{2.18}$$

– funkcja wrażliwości drugiego rzędu funkcji $X(x,y)$ względem argumentu *x*;

$$S_{y^2}^{X(x,y)} = \frac{\partial^2 X(x,y)}{\partial y^2} \cdot y^2 \tag{2.19}$$

– funkcja wrażliwości drugiego rzędu funkcji $X(x,y)$ względem argumentu *y*;

$$S_{xy}^{X(x,y)} = \frac{\partial^2 X(x,y)}{\partial x \partial y} \cdot x \cdot y \tag{2.20}$$

– funkcja wrażliwości drugiego rzędu funkcji $X(x,y)$ względem argumentów *x* oraz *y*;
Dokładnie tak samo dla innych kombinacji funkcji i argumentów.

Jak wspomniano wcześniej, funkcje wrażliwości mogą być przedstawione z wykorzystaniem współczynników WKG w sposób następujący:

$$\left.\begin{aligned} S_x^{X(x,y)} &= \frac{\partial X(x,y)}{\partial y}\cdot x = (a_1 + 2a_3x + a_5y)\cdot x; \\ S_y^{X(x,y)} &= \frac{\partial X(x,y)}{\partial y}\cdot y = (a_2 + 2a_4y + a_5x)\cdot y; \\ S_{x^2}^{X(x,y)} &= \frac{\partial^2 X(x,y)}{\partial x^2}\cdot x^2 = 2a_3\cdot x^2; \\ S_{y^2}^{X(x,y)} &= \frac{\partial^2 X(x,y)}{\partial y^2}\cdot y^2 = 2a_4\cdot y^2; \\ S_{xy}^{X(x,y)} &= \frac{\partial^2 X(x,y)}{\partial x\partial y}\cdot x\cdot y = a_5\cdot x\cdot y; \end{aligned}\right\} \quad (2.21)$$

Dokładnie tak samo dla innych kombinacji funkcji i argumentów.

Podobnie jak w notacji proponowanej wcześniej, załóżmy, że mamy pewien ciąg próbek, określających punkty przestrzeni wielowymiarowej zmiennych *x, y, z,* .... Przedstawiamy funkcję *X* jako funkcję kombinacji par zmiennych *x, y, z,* ... (analogicznie dowolną inną funkcję $Y, Z,$ ...) w postaci symbolicznej zależności tej funkcji względem numeru próbki (Rys. 2.3).

Załóżmy teraz, że mamy kilka próbek procesu w punktach $P(0), P(1), P(2),$ .... Wartości każdej zmiennej *x, y, z,* ... odzwierciedlają wartości pewnych parametrów procesu rzeczywistego i odzwierciedlają wzajemne zależności między wszystkimi tymi zmiennymi. Niech proces, który badamy, jest regulowalny, czyli możemy zmieniać wartości każdej z tych zmiennych, próbując kierować zmianami tych zmiennych w przyszłości. W sytuacji tej powstaje pytanie: jeśli zmienimy wartości parametrów *x* oraz *y* w *i*-tym momencie czasowym (czyli w naszej notacji w punkcje $P(i)$), jakie będą oczekiwane wartości tych parametrów w kolejnym punkcje czasowym (czyli w punkcje $P(i+1)$), albo za kilka kroków w punktach $P(i+2), P(i+3)$ itd.? Nie możemy zmienić wzajemnych zależności między zmiennymi procesu, ponieważ to nie od nas zależy, zależności funkcjonalne między zmiennymi procesu określone są przez naturę procesu. Ale możemy kierować tym procesem zmieniając niektóre lub wszystkie zmienne w wymaganym przez nas kierunku.

Jeśli zmienne analizowanego przez nas procesu są wzajemnie niezależne, to proces ten nazywamy **prostym**. Sterowanie procesem prostym jest oczywiste. Powiedzmy, jeśli chcemy uzyskać wartość zmiennej $x_1(t_{i+1})$ zwiększoną o 10%, to dodajemy do wartości $x_1(t_i)$ przyrost $0.1\cdot x_1(t_i)$, i wiemy z pewnością, że wartość tej zmiennej w kolejnym momencie czasu będzie równa $x_1(t_{i+1}) = x_1(t_i) + 0.1\cdot x_1(t_i)$. Jest to słusznie jedynie

w przypadku, gdy zależność $x_1(t_{i+1})$ od $x_1(t_i)$ jest funkcją liniową. Problem w tym, że niekiedy budujemy matematyczne zależności, wychodząc z czysto matematycznych rozważań, np. dokładności aproksymacji, chociaż tworzone zależności mogą być sprzeczne ze "zdrowym rozsądkiem".

Np., dla modelu regresyjnego $Y = a_0 + a_1x_1 + a_2x_2 + a_3x_1^2 + a_4x_2^2 + a_5x_1x_2$, może zaistnieć taka sytuacja, gdy wszystkie współczynniki się zerują, za wyjątkiem współczynnika a_3. Wtedy wartość Y zależy od siebie jako funkcja $Y = a_3x_1^2$. Zmiana o 10% zmiennej x_1 spowoduje zmianę wartości zmiennej Y o wartość, inną niż 10%. Niemniej jednak, w przypadku procesu prostego, sterowanie jest uproszczone, ponieważ nie musimy szukać zmiennych, które wpływają najsilniej na konkretną składową tego procesu (np., korzystając z funkcji wrażliwości), bowiem dana składowa zależy jedynie od samej siebie.

Jeśli zmienne procesu są wzajemnie zależne, to proces ten nazywamy **złożonym.** Jako abstrakcję systemu, proces złożony można utożsamić z pewnym systemem, składającym się z ciągu podsystemów $S_1, S_2,\ldots,S_i,\ldots, S_n,$ z których każdy S_i, wygeneruje swój sygnał x_i sterujący danym procesem. Owa abstrakcja może być wygodna w przypadku analizy takich systemowych cech procesu, jak np. obserwowalność i sterowalność tych czy innych składników procesu. Wspomniane cechy mają duże znaczenie dla budowy algorytmów sterowania przebiegiem procesu (niektóre szczegóły tego zagadnienia powiązane z proponowaną w danej pracy metodą opisane są w Rozdziale 3.).

W terminach wspomnianej abstrakcji, poprzez złożoność systemu rozumiemy fakt, że składa się on z wielu elementów (podsystemów) oraz, że podsystemy wchodzące w skład całego systemu są ze sobą połączone i wzajemnie na siebie oddziałują, a ich rozdzielenie jest na ogół zbyt trudne lub niejednoznaczne. Mogą również wystąpić ograniczenia w dostępności do informacji o sygnałach dochodzących z zewnątrz do systemu. Dlatego też do systemów o strukturze złożonej nie da się w bezpośredni sposób zastosować algorytmów identyfikacji, stosowanych dla obiektu o pojedynczym sygnale wyjściowym.

Definicja systemu złożonego może mieć różne interpretacje dla różnych zastosowań. Często używana jest nieściśle, gdyż każdy badacz może subiektywnie oceniać złożoność. Za system złożony może być uważany taki, którego model matematyczny składa się z dużej liczby równań, lub jeżeli te równania są np. nieliniowe. Wówczas granica podziału systemów złożonych i prostych jest nieformalna i bardzo

płynna. Jeżeli chodzi o zagadnienie sterowania procesem złożonym, to istotnie się ono komplikuje w porównaniu do systemu prostego. Jak już było wspomniane powyżej, problem polega przede wszystkim na tym, że w ten czy inny sposób musimy ocenić to, które zmienne wejściowe mają największy wpływ na konkretne zmienne wyjściowe. W naszym podejściu do oceny siły wpływu zmiennych proponowana jest metoda obliczania funkcji wrażliwości cząstkowych modeli procesu i wyborze takich zmiennych do sterowania, których funkcje wrażliwości mają największe wartości.

Drugi zasadniczy problem polega na tym, że, mówiąc ogólnie, cząstkowe modele procesu mogą nie odzwierciedlać rzeczywistych zależności pomiędzy zmiennymi procesu, ponieważ dowolne matematyczne zależności, w tym nasze modele cząstkowe to jedynie formalne aproksymacje eksperymentalnych zależności. Naturalnym rozwiązaniem wspomnianego problemu jest zaznajomienie się przez badacza z architekturą badanego systemu albo zależnościami funkcjonalnymi występującymi pomiędzy zmiennymi stanów tego systemu. Np., jeżeli badany proces jest pewnym procesem chemicznym, to badacz najprawdopodobniej wie, które odczynniki chemiczne reagują ze sobą. Dlatego z pośród wszystkich ewentualnych modeli cząstkowych wybiera te, które odpowiadają procesowi rzeczywistemu. A ponadto, wśród ewentualnych zmiennych procesu może wybrać te zmienne sterujące, które mają największe wartości funkcji wrażliwości w tych modelach cząstkowych.

W gorszym przypadku, gdy architektura lub rzeczywiste zależności funkcjonalne pomiędzy zmiennymi pozostają dla badacza nieznane, jedynym sposobem sprawdzenia dopasowania modelu do procesu rzeczywistego jest testowanie tworzonych modeli cząstkowych na ciągu danych eksperymentalnych.

W dowolnym przypadku, dokładna ocena wpływu zmian wartości jednych zmiennych na wartości innych zmiennych za pomocą jedynie funkcji wrażliwości, obliczanych w modelach cząstkowych jest niemożliwa, ponieważ każdy model cząstkowy jest pewną idealizacją prawdziwych zależności między zmiennymi. Dlatego należy klasyfikować proponowaną w danej pracy metodę opartą na wykorzystaniu funkcji wrażliwości, raczej jako metodę jakościowej oceny wpływów sygnałów wejściowych na te czy inne sygnały wyjściowe. Jednak, jak świadczą liczne doświadczenia, wspomniane oceny, tak samo, jak i modele tworzone w metodzie GMDH i rozmaitych jej modyfikacjach, najczęściej dają wyniki liczbowe dosyć bliskie rzeczywistym wartościom zmian zmiennych stanów.

Rzeczywiście, wskutek wzajemnej zależności między zmiennymi, zmiana,

powiedzmy wartości zmiennej $x_1(t_i)$ może spowodować zmiany wartości niektórych pozostałych zmiennych stanów x_2, x_3,... itd., a te ostatnie z kolei mogą wpłynąć na wartość zmiennej $x_1(t_{i+1})$, wzmacniając lub zmniejszając jej wartość. Jest to efekt nieco podobny do efektu wpływu sprzężenia zwrotnego we wzmacniaczach elektronicznych – np. jeśli taki wzmacniacz ma dodatnie sprzężenie zwrotne, to niewielkie zwiększenie sygnału wejściowego może spowodować wielokrotne zwiększenie tego sygnału przez dodatnie sprzężenie zwrotne, czego efektem będzie sygnał wyjściowy wielokrotnie większy niż ten, który jest określony współczynnikiem wzmocnienia.

Problem komplikuje się tym bardziej, że najczęściej zmienia się nie pojedyncza zmienna (powiedzmy, $x_1(t_i)$), a od razu kilka lub wszystkie zmienne, które wpływają w sposób złożony na wartość $x_1(t_{i+1})$.

Istnieją dwa, znane z literatury, podejścia do rozwiązania tego zagadnienia. Pierwsze oparte jest na metodach teorii procesów przypadkowych. Drugie – na metodach teorii eksploracji danych (data mining). W tym drugim podejściu można wspomnieć metodę regresji i metodę sieci Bayesowskich. Niestety metoda regresji, mimo, że pozwala tworzyć modele deterministyczne, nie pozwala wyodrębnić z tych modeli składników, z których składa się proces złożony, czyli nie jest przydatna do rozwiązania sformułowanego problemu. Pozostałe ze wspomnianych podejść oparte są na teorii prawdopodobieństwa, a więc wymagają przeprowadzenia dużej ilości eksperymentów, czyli dużej ilości próbek, na podstawie których można określić pewne parametry statystyczne. Uważa się, że aby teoria prawdopodobieństwa działała skutecznie, ilość punktów czasowych, na których zbieramy dane eksperymentalne, powinna być o wiele większa (w praktycznych zastosowaniach około 10-krotnie), niż ilość tych, na których tworzone modele mogą być stosowane do sterowania procesem. Aktualnym jest, więc problem utworzenia metody, która pozwala przy niewielkiej ilości próbek testujących stworzyć model, którym dalej możemy się posługiwać w sterowaniu złożonych procesów i w prognozie.

Wskutek wzajemnej zależności między parametrami (zmiennymi) procesu *x, y, z, ...* , rzeczywiste wartości tych parametrów w momentach $P(i+1)$, $P(i+2)$ itd., będą się różnić się od ustalonych przez nas przyrostów wartości tych parametrów w momencie $P(i)$.

Problem wygląda, więc następująco: jakie będą rzeczywiste wartości zmiennych *x, y, z ...* w kolejnych punktach czasowych $P(i+1)$, $P(i+2)$, ... , oceniane na podstawie

modeli cząstkowych, jeśli każda z tych zmiennych będzie miała określony przez nas przyrost względny $ddx_i = \frac{\Delta x_i}{x_i}, ddy_i = \frac{\Delta y_i}{y_i}$ itd., bądź bezwzględny $ddx_i = \Delta x_i$, $ddy_i = \Delta y_i$ itd. w momencie $P(i)$, i jaki wpływ na zmiany tych wartości ma każda ze zmiennych procesu? Problem komplikuje się tym bardziej, że w rzeczywistości wzajemne zależności parametrów (zmiennych) *x, y, z, ...* są **nieliniowe**, natomiast, jak wiadomo z teorii data mining, wzajemna zależność między zmiennymi może zostać ustalona za pomocą statystycznej metody określenia współczynników korelacji jedynie dla zależności liniowych między zmiennymi. Co więcej, nawet w tych warunkach wyniki korelacji mogą prowadzić nas do błędnych wniosków.

Biorąc pod uwagę, że rzeczywiste zmiany zmiennych *x, y, z,...* określone w danych eksperymentalnych mogą istotnie różnić się od przyrostów $ddx_i = \frac{\Delta x_i}{x_i}, ddy_i = \frac{\Delta y_i}{y_i}$, ... określonych na modelach cząstkowych, pokażemy poniżej, jak można, z pewnym przybliżeniem, obliczyć wspomniane przyrosty.

Zagadnienie określone powyżej jest jednym z zagadnień sterowania procesami złożonymi. Można określić dwa problemy sterowania: **bezpośredni** i **odwrotny**.
Używając terminologii przyjętej w ogólnej teorii systemów sterowania można wytłumaczyć oba problemy następująco:
Niech pewien proces zostanie opisany matematycznie za pomocą układu równań stanu procesu, zawierającego zmienne stanów *x, y, z, ...*, który w postaci operatorowej zapiszemy tak:

$$(x_{i+1}, y_{i+1}, z_{i+1}, \dots) = F(x_i + ddx_i, y_i + ddy_i, z_i + ddz_i, \dots), \qquad (2.22)$$

gdzie *F* – to operator przejścia od jednego stanu procesu do innego w przestrzeni metrycznej *n*-wymiarowej zmiennych stanu $(x, y, z, \dots)$.

Operator ten określa, jaki stan będzie posiadał proces w kolejnym punkcie $P(i+1)$, jeśli w punkcie $P(i)$ przestrzeni stanów $(x, y, z, \dots)$ zmienne stanów będą zmodyfikowane o wartości przyrostów względnych $ddx_i = \frac{\Delta x_i}{x_i}, ddy_i = \frac{\Delta y_i}{y_i}$,…. Operator *F* jest ciągły względem procesów ciągłych w czasie.

Bezpośredni problem sterowania procesem złożonym polega na znalezieniu punktu $(x_{i+1}, y_{i+1}, z_{i+1}, \dots)$, jeśli operator *F* i wartości przyrostów względnych są określone, a **odwrotny problem** sterowania polega na określeniu rzeczywistych wartości przyrostów ddx_i, ddy_i, ddz_i, ... zmiennych procesu $(x_i, y_i, z_i, \dots)$, które doprowadzą proces do stanu $(x_{i+1}, y_{i+1}, z_{i+1}, \dots)$.

Problem ten na ogół nie może zostać rozwiązywany za pomocą metod statystycznych oraz metody GMDH. Przyczyna tkwi w tym, że metody te nie przewidują rozdzielenia procesu na części składowe, a działają na całych zbiorach zmiennych. Natomiast rozwiązanie zarówno bezpośredniego, jak i odwrotnego problemu sterowania procesami polega na pewnym wydzieleniu składowych części złożonego procesu i ocenie wkładu każdej z tych części na ogólny stan procesu, co można uzyskać np. za pomocą funkcji wrażliwości.

Przykład 1. Analiza szeregu czasowego o trzech zmiennych

Przeanalizujemy eksperyment na przykładzie procesu o trzech zmiennych stanów *x, y, z* (Tab. 2.1). Załóżmy, że zmieniają się wartości zmiennych *x* oraz *y*, a zmienna *z* pozostaje niezmienna. Na początku tworzymy modele cząstkowe (2.23) i (2.24) w postaci WKG drugiego stopnia.

Tab. 2.1 Zbiór danych eksperymentalnych o trzech zmiennych

Punkty czasowe	***x***	***y***	***z***
P(0)	0.2366	0.2068	0.2496
P(1)	0.2403	0.2462	0.2731
P(2)	0.3033	0.2406	0.2856
P(3)	0.2367	0.2688	0.2966
P(4)	0.2679	0.2607	0.2951
P(5)	0.3308	0.2484	0.2823
P(6)	0.241	0.3136	0.337
P(7)	0.3151	0.2736	0.3713
P(8)	0.4963	0.2788	0.3282
P(9)	0.2249	0.3085	0.3362
P(10)	0.2855	0.282	0.3291

…

Tworzymy model funkcji cząstkowej $X(x, y)$:

$$X(x, y) = a_0 + a_1 x + a_2 y + a_3 x^2 + a_4 y^2 + a_5 xy, \tag{2.23}$$

dla poszukiwania przyszłych wartości zmiennej *x*,

oraz:

$$Y(y, x) = b_0 + b_1 y + b_2 x + b_3 y^2 + b_4 x^2 + b_5 xy, \tag{2.24}$$

dla przyszłych wartości zmiennej *y*.

W tym celu tworzymy układ równań do określenia wartości współczynników $a_1, \dots, a_5$ postaci (2.15).

$$0.2403 = a_0 + a_1(0.2366) + a_2(0.2068) + a_3(0.2366)^2 + a_4(0.2068)^2 + a_5(0.2366) \cdot (0.2068)$$

$$0.3033 = a_0 + a_1(0.2403) + a_2(0.2462) + a_3(0.2403)^2 + a_4(0.2462)^2 + a_5(0.2403) \cdot (0.2462)$$

…

$0.241 = a_0 + a_1(0.3308) + a_2(0.2484) + a_3(0.3308)^2 + a_4(0.2484)^2 + a_5(0.3308) \cdot (0.2484)$

Po rozwiązaniu tego układu równań otrzymujemy:

$a_0 = 2.0455$, $a_1 = -16.2210$, $a_2 = 1.3224$, $a_3 = -14.3273$, $a_4 = -49.3550$, $a_5 = 95.4847$

Poszukiwany model przyjmuje, więc postać:

$$X(x, y) = 2.0455 - 16.221x + 1.3324y - 14.3273x^2 - 49.355y^2 + 95.4847xy$$

W podobny sposób tworzymy model $Y(x, y)$ w postaci WKG (2.24). A mianowicie, aby znaleźć wartości współczynników b_i, tworzymy układ równań:

$0.2462 = b_0 + b_1(0.2068) + b_2(0.2366) + b_3(0.2068)^2 + b_4(0.2366)^2 + b_5(0.2068) \cdot (0.2366)$

$0.2406 = b_0 + b_1(0.2462) + b_2(0.2403) + b_3(0.2462)^2 + b_4(0.2403)^2 + b_5(0.2462) \cdot (0.2403)$

…

$0.3136 = b_0 + b_1(0.2484) + b_2(0.3308) + b_3(0.2484)^2 + b_4(0.3308)^2 + b_5(0.2484) \cdot (0.3308)$

Rozwiązaniem tego układu równań jest:

$b_0 = 1.5550$, $b_1 = -5.6708$, $b_2 = -5.252$, $b_3 = 14.9988$, $b_4 = 12.8445$, $b_5 = -5.1894$.

Poszukiwany model ma, więc postać:

$$Y(y, x) = 1.5550 - 5.6708y - 5.252x + 14.9988y^2 + 12.8445x^2 - 5.1894yx$$

Dokładnie w taki sam sposób znajdujemy modele cząstkowe $X(x, z)$, $Y(y, z)$, $Z(z, x)$ oraz $Z(z, y)$, otrzymując:

$X(x, z) = 39.3362 - 250.489x - 54.623z + 96.1379x^2 - 197.77z^2 + 689.9113xz$,

$Y(y, z) = -25.29022 - 179.125x + 340.38z - 327.215x^2 - 1131.1z^2 + 1202.8xz$,

$Z(z, x) = 1.9487 + 2.0640z - 15.5134x - 11.7447z^2 + 18.0084x^2 + 20.3766zx$

$Z(z, y) = -20.3054 + 268.05z - 136.966y - 875.559z^2 - 243.37y^2 + 909.63zy$

Na początku badamy, jak działa **klasyczna (uproszczona)** metoda GMDH. W tym celu korzystamy z wyznaczonych modeli, aby wskazać kilka przyszłych stanów procesu. Podstawiamy do modelu $X(x, y)$ (2.23) oraz $Y(y, x)$ (2.24) wartości zmiennych $x_6 = 0.2410$ oraz $y_6 = 0.3136$ (czyli wartości zmiennych x oraz y w punkcie $P(6)$) w celu predykcji wartości tych zmiennych w momencie $P(7)$.

Oto rezultaty:

$x_{7,progn} = 0.0815$; $y_{7,progn} = 0.3398$.

Dokładne wartości (Tab. 2.1) wynoszą $x_{7,dokł} = 0.3151$; $y_{7,dokł} = 0.2736$. Jak można zauważyć, już w pierwszym kroku predykcji otrzymujemy dużą rozbieżność wyników dla wartości zmiennej x, $\delta_{x_7} = \frac{|x_{7,progn} - x_{7,dokł}|}{x_{7,dokł}} = 0.741$ (czyli około 74%) i znaczny błąd dla wartości zmiennej y, $\delta_{y_7} = \frac{|y_{7,progn} - y_{7,dokł}|}{y_{7,dokł}} = 0.242$ (czyli 24.2%).

Wyniki dla kolejnych punktów czasowych są bardziej satysfakcjonujące. Natomiast podstawiając do modeli $X(x,y)$ oraz $Y(y,x)$ dokładne wartości zmiennych x_7 oraz y_7, otrzymujemy prognozowane wartości tych zmiennych dla punktu $P(8)$:

$x_{8,progn} = 0.4136$ (dokładna wartość $x_{8,dokł} = 0.4963$, błąd $\delta_{x_8} = 0.167$, czyli 16%),

$y_{8,progn} = 0.299$ (dokładna wartość $y_{8,dokł} = 0.2788$, błąd $\delta_{y_8} = 0.072$, czyli 7%).

Natomiast podstawiając do modelu wartości obliczone modelem w punkcie $P(7)$, w celu obliczenia wartości w punkcie czasowym $P(8)$, otrzymujemy:

$x_{8,progn} = -1.9756$, czyli błąd $\delta_{x_8} = 4.98$ (498%) i $y_{8,progn} = 0.873$ (błąd $\delta_{y_8} = 2.131;\ 213\%$).

Dla pozostałych modeli otrzymujemy następujące wyniki: (Tab. 2.2)

Tab. 2.2 Rezultaty prognozy w punkcie *P(7)* i *P(8)* dla wszystkich modeli

Modele	$X(x,y)$	$Y(y,x)$	$Y(y,z)$	$X(x,z)$	$Z(z,x)$	$Z(z,y)$
$P(7)_{dokł}$	0.3151	0.2736	0.2736	0.3151	0.3713	0.3713
$P(7)_{progn}$	0.0815	0.3398	−0.2836	−0.2838	0.2725	−0.1623
$P(8)_{dokł}$	0.4963	0.2788	0.2788	0.4963	0.3282	0.3282
$P(8)_{progn}$	0.4136	0.299	−30.4869	35.232	5.915	−25.737

Proces obliczeniowy rozbiega się, więc dla wszystkich modeli cząstkowych. Oczywiście, w tym przykładzie metoda GMDH została uproszczona. W rzeczywistej metodzie GMDH należałoby wybrać najdokładniejszy model cząstkowy spośród modeli obliczających wartość każdej ze zmiennych według wstępnie obranego kryterium, a następnie przejść z wynikami do kolejnej warstwy obliczeń. Ale tak znaczna rozbieżność obliczeń niektórych zmiennych wskazuje na to, że metoda GMDH może spowodować dowolnie duże błędy od samego początku obliczeń, które mogą zostać wzmocnione na kolejnych krokach. Przyczyna tkwi w tym, że GMDH działa podobnie do krzywika, który usiłujemy połączyć z istniejącymi punktami w przestrzeni zmiennych procesu, a nie uwzględnia dynamiki zmian tych zmiennych. Zmiany te uwzględniamy w momencie wykorzystania omówionych wyżej funkcji wrażliwości, ponieważ funkcje te oparte są na obliczeniach pierwszej i drugiej pochodnej prognozowanych funkcji wyjściowych, względem zmiennych stanów. O prawidłowości tego stwierdzenia możemy się przekonać w licznych przykładach przedstawionych dalej w danej pracy.

W podobnych warunkach, metoda GMDH tym bardziej nie jest w stanie rozwiązać problemu predykcji danych na krótkich próbkach w sytuacjach, gdy dane te przedstawione są dodatkowo z wysokim poziomem szumu.

Skorzystamy wiec teraz z modelu zmodyfikowanego wykorzystującego funkcje wrażliwości na tych samych danych (Tab. 2.1). Obliczamy więc wartości funkcji

wrażliwości dla wszystkich potrzebnych w dalszych obliczeniach punktów od $P(0)$ do $P(5)$ (w Tab. 2.3 przedstawiamy wartości funkcji wrażliwości dla kliku punktów czasowych więcej) dla modelu $X(x, y)$, wykorzystując wzory (2.21).

Tab. 2.3 Wartości funkcji wrażliwości dla modelu $X(x, y)$

Punkt czasowy	$S_x^{X(x,y)}$	$S_y^{X(x,y)}$	$S_{x^2}^{X(x,y)}$	$S_{y^2}^{X(x,y)}$	$S_{xy}^{X(x,y)}$
P(0)	-0.77	0.724	-1.604	-4.2214	4.672
P(1)	0.0965	-0.0086	-1.655	-5.9832	5.649
P(2)	-0.588	1.572	-2.636	-5.714	6.968
P(3)	0.6303	-0.7015	-1.605	-7.132	6.0752
P(4)	0.2666	0.305	-2.0565	-6.709	6.669
P(5)	-0.6555	2.0839	-3.136	-6.09	7.846
P(6)	1.643	-2.076	-1.664	-9.708	7.2165
P(7)	0.2757	1.2045	-2.845	-7.389	8.2319

Następnie dla modelu $Y(y, x)$ (Tab. 2.4):

Tab. 2.4 Wartości funkcji wrażliwości dla modelu $Y(y, x)$

Punkt czasowy	$S_y^{Y(y,x)}$	$S_x^{Y(y,x)}$	$S_{y^2}^{Y(y,x)}$	$S_{x^2}^{Y(y,x)}$	$S_{yx}^{Y(y,x)}$
P(0)	-0.1438	-0.0585	1.283	1.438	-0.254
P(1)	0.1151	-0.0857	1.818	1.4834	-0.307
P(2)	-0.0066	0.3915	1.7365	2.363	-0.3787
P(3)	0.3129	-0.134	2.167	1.439	-0.3301
P(4)	0.1979	0.0743	2.039	1.844	-0.362
P(5)	0.0159	0.647	1.851	2.811	-0.4264
P(6)	0.7795	-0.1659	2.9501	1.492	-.0392
P(7)	0.2466	0.4483	2.2455	2.5506	-0.447

oraz $Y(y, z)$ (Tab. 2.5).

Tab. 2.5 Wartości funkcji wrażliwości dla modelu $Y(y, z)$

Punkt czasowy	$S_y^{Y(y,z)}$	$S_z^{Y(y,z)}$	$S_{y^2}^{Y(y,z)}$	$S_{z^2}^{Y(y,z)}$	$S_{yz}^{Y(y,z)}$
P(0)	-2.9461	6.1037	-27.9875	-140.939	62.0844
P(1)	-2.8967	5.1017	-39.668	-168.727	80.872
P(2)	1.6683	-4.6647	-37.8839	-184.526	82.65
P(3)	0.4595	-2.165	-47.2848	-199.014	95.893
P(4)	1.3573	1.3573	-44.478	-197.006	92.533
P(5)	-0.5315	0.1451	-40.38	-180.287	84.343
P(6)	6.5805	-15.101	-64.36	-256.922	127.114
P(7)	24.1907	-63.3128	-48.989	-311.88	122.19

Dla $X(x, z)$ (Tab. 2.6).

Tab. 2.6 Wartości funkcji wrażliwości dla modelu $X(x, z)$

Punkt czasowy	$S_x^{X(x,z)}$	$S_z^{X(x,z)}$	$S_{x^2}^{X(x,z)}$	$S_{z^2}^{X(x,z)}$	$S_{xz}^{X(x,z)}$
P(0)	-7.7592	2.4668	10.7635	-24.6422	40.743
P(1)	-3.8136	0.8577	11.1028	-29.501	45.276
P(2)	1.4761	11.898	17.688	-32.263	59.7618
P(3)	-0.08276	-2.5621	10.773	-34.796	48.435
P(4)	1.2362	3.978	13.8	-34.445	54.543
P(5)	2.606	17.485	21.04	-31.522	64.427
P(6)	6.8322	-7.2965	11.168	-44.921	56.033
P(7)	20.879	5.905	19.09	-54.531	80.717

oraz $Z(z,x)$ (Tab. 2.7)

Tab. 2.7 Wartości funkcji wrażliwości dla modelu $Z(z,x)$

Punkt czasowy	$S_z^{Z(z,x)}$	$S_x^{Z(z,x)}$	$S_{z^2}^{Z(z,x)}$	$S_{x^2}^{Z(z,x)}$	$S_{zx}^{Z(z,x)}$
P(0)	0.2551	-0.451	-1.463	2.01620	1.2033
P(1)	0.149	-0.311	-1.752	2.0798	1.3372
P(2)	0.4386	0.3731	-1.916	3.3132	1.7651
P(3)	-0.0237	-0.2236	-2.0664	2.0179	1.4305
P(4)	0.1744	0.03981	-2.0456	2.5849	1.6109
P(5)	0.6136	0.7123	-1.872	3.941	1.9029
P(6)	-0.3172	0.0081	-2.668	2.0919	1.655
P(7)	-0.088	1.0717	-3.238	3.576	2.384

i ostatecznie dla modelu $Z(z,y)$ (Tab. 2.8).

Tab. 2.8 Wartości funkcji wrażliwości dla modelu $Z(z,y)$

Punkt czasowy	$S_z^{Z(z,y)}$	$S_y^{Z(z,y)}$	$S_{z^2}^{Z(z,y)}$	$S_{y^2}^{Z(z,y)}$	$S_{zy}^{Z(z,y)}$
P(0)	4.7637	-2.1882	-109.095	-20.816	46.953
P(1)	3.7615	-2.0638	-130.60	-29.5037	61.1609
P(2)	-3.773	1.3746	-142.83	-28.177	62.505
P(3)	-2.0229	0.5358	-154.05	-35.169	72.521
P(4)	-3.412	1.1917	-152.49	-33.081	69.980
P(5)	-0.0947	-0.269	-139.55	-30.033	63.786
P(6)	-12.406	5.3112	-198.87	-47.869	96.133
P(7)	-49.480	18.497	-241.42	-36.436	92.407

Jak można zauważyć – wkłady różnych zmiennych w wartości zmiennych *x, y, z* na różnych odcinkach przedziału obserwacji są istotnie różne. A więc, funkcje wrażliwości mogą zawierać konieczne informacje, którymi możemy posługiwać się dla oceny wpływów zmiennych procesu w różnych punktach czasowych, w tym do analizy bieżących i określenia przyszłych stanów procesu.

Dalej wygodnie jest opisać procesy obliczeniowe, rozpoczynając od rozwiązania wspomnianego wyżej problemu odwrotnego. Przeanalizujemy przykładowo zmiany zachodzące w wartościach zmiennej *x* i opiszemy te zmiany za pomocą funkcji cząstkowej $X(x,y)$. W tym celu, przypuśćmy, że w procesie zawierającym trzy zmienne *x, y, z* zmieniają się dwie zmienne – *x* oraz *y*, a zmienna *z* jest niezmienna. Więc, proces ten może być opisany za pomocą dwóch funkcji cząstkowych $X(x,y)$ oraz $Y(y,x)$.

W tych eksperymentach tworzymy funkcję odpowiadającą operatorowi *F* (patrz (2.22)). W naszej metodzie (Metoda Analizy danych na podstawie Modeli Cząstkowych – MAMC) podstawą dla tworzenia tej funkcji są funkcje wrażliwości pierwszego i drugiego rzędu (patrz (2.21)). Wykonujemy kolejno następujące czynności:

1. Na początku tworzymy funkcję uproszczoną ($MAMC_{simp}$), ograniczając się do funkcji wrażliwości jedynie pierwszego rzędu. Rozwiązując odwrotny problem

sterowania, znajdujemy wartości przyrostów *ddx* oraz *ddy* dla punktów $P(0)$, $P(1)$, $P(2)$, …, stosując wzory (2.25), (2.26).

2. Następnie tworzymy funkcję rozszerzoną (MAMC$_{ext}$), korzystając z funkcji wrażliwości pierwszego i drugiego rzędu i ponownie rozwiązujemy problem odwrotny, czyli szukamy wartości przyrostów *ddx* oraz *ddy* dla punktów $P(0)$, $P(1)$, $P(2)$, …, stosując wzory (2.30), (2.31).
3. Porównujemy ze sobą wyniki obliczeń, aby ocenić na ile konieczne jest wykorzystanie modelu rozszerzonego (zawierającego funkcje wrażliwości pierwszego i drugiego rzędu).

Trzy etapy przebiegają w sposób następujący:

Etap 1.

Uproszczony wzór do obliczania wartości zmiennej *x* w kolejnym punkcie czasowym $P(i+1)$, czyli funkcji $X_{i+1}(x,y)$, zawierający funkcje wrażliwości jedynie pierwszego rzędu, jest następujący:

$$X_{i+1}(x_i, y_i) = x_i + S_{x_i}^{X(x,y)} \cdot ddx_i + S_{y_i}^{X(x,y)} \cdot ddy_i, \tag{2.25}$$

gdzie wszystkie składniki po prawej obliczane są w punkcie $P(i)$.

Podobnie, uproszczony wzór do obliczania wartości zmiennej *y* w punkcie $P(i+1)$, czyli funkcji $Y_{i+1}(y,x)$, zawierający jedynie funkcje wrażliwości pierwszego rzędu, jest następujący:

$$Y_{i+1}(y_i, x_i) = y_i + S_{y_i}^{Y(y,x)} \cdot ddy_i + S_{x_i}^{Y(y,x)} \cdot ddx_i, \tag{2.26}$$

gdzie wszystkie składniki po prawej obliczmy również w punkcie $P(i)$.

Ponieważ w tym momencie wartości współczynników a_i oraz b_i WKG zostały już wcześniej określone, korzystamy z nich, aby przedstawić wzory (2.25), (2.26) w terminach tych współczynników. Otrzymujemy więc:

$$\begin{cases} X_{i+1}(x_i, y_i) = x_i + (a_1 + 2a_3x_i + a_5y_i) \cdot x_i \cdot ddx_i + \\ \qquad\qquad +(a_2 + 2a_4y_i + a_5x_i) \cdot y_i \cdot ddy_i \\ Y_{i+1}(y_i, x_i) = y_i + (b_2 + 2b_4y_i + b_5x_i) \cdot y_i \cdot ddy_i + \\ \qquad\qquad +(b_1 + 2b_3x_i + b_5y_i) \cdot x_i \cdot ddx_i \end{cases} \tag{2.27}$$

Jest to układ równań liniowych względem niewiadomych dwóch wartości przyrostów ddx_i oraz ddy_i.

Rozwiązujemy taki układ dla każdego punktu czasowego $P(i)$. Np. dla punktu $P(0)$, czyli $i = 0$, a $i+1 = 1$, otrzymujemy:

$$0.2403 = 0.2366 + (-16.221 + 2 \cdot (-14.3273) \cdot 0.2366 + 95.4847 \cdot 0.2068) \cdot 0.2366 \cdot ddx_0 + (1.3224 + 2 \cdot (-49.355) \cdot 0.2068 + +95.4847 \cdot 0.2366) \cdot 0.2068 \cdot ddy_0 \quad (2.28)$$

$$0.2462 = 0.2068 + (-5.252 + 2 \cdot 12.8445 \cdot 0.2068 + (-5.1894) \cdot 0.2366) \cdot 0.2068 \cdot ddy_0 + (-5.6708 + 2 \cdot 14.9988 \cdot 0.2366 + +(-5.1894) \cdot 0.2068) \cdot 0.2366 \cdot ddx_0 \quad (2.29)$$

Rozwiązaniem tego układu jest $ddx_0 = -0.1899$; $ddy_0 = -0.1968$.

Innymi słowy, jeżeli w miejsce przyrostów cząstkowych po prawej stronie równań (2.28) i (2.29) podstawimy znalezione wartości ddx_0 oraz ddy_0, to z pewnym błędem arytmetycznym otrzymamy wartości w kolejnym punkcie czasowym dla tych dwóch zmiennych, czyli $x_1 \approx 0.2403$ oraz $y_1 \approx 0.2462$.

W taki sam sposób obliczamy przyrosty cząstkowe dla punktów $P(1)$ ...$P(7)$. Oto wyniki obliczeń (Tab. 2.9).

Tab. 2.9 Przyrosty cząstkowe ddx_i oraz ddy_i znalezione dla modelu uproszczonego (2.25), (2.26)

Punkt czasowy	ddx_i	ddy_i
P(0)	-0.1899	-0.1968
P(1)	0.6946	0.4683
P(2)	0.0718	-0.0155
P(3)	0.0395	-0.0089
P(4)	0.5373	-0.2637
P(5)	0.1010	-0.0113
P(6)	-0.027	-0.057
P(7)	-0.0814	0.169

Porównując przyrosty cząstkowe z pełnymi przyrostami zmiennych *x*, *y* z Tab. 2.1, można zauważyć istotne różnice tych wartości. Np., w $P(0)$ rzeczywiste wartości zmiennych *x* oraz *y* wynoszą $x_0 = 0.2366$; $x_1 = 0.2403$; $y_0 = 0.2068$; $y_1 = 0.2462$. Czyli rzeczywiste przyrosty względne tych zmiennych wynoszą $\delta_{x_0} = \frac{x_1 - x_0}{x_0} = \frac{0.2403 - 0.2366}{0.2366} = 0.0156$, $\delta_{y_0} = \frac{y_1 - y_0}{y_0} = \frac{0.2462 - 0.2068}{0.2068} = 0.1905$, natomiast przyrosty cząstkowe tych zmiennych podczas przejścia z punktu $P(0)$ do punktu $P(1)$ wynoszą: $ddx_0 = -0.1899$; $ddy_0 = -0.1968$.

Dane dla pozostałych punktów czasowych przedstawione są w Tab. 2.10.

Sprawdza się, więc nasza hipoteza, że wskutek wzajemnego wpływu zmiennych procesu, cząstkowe przyrosty zmiennych (czyli przyrosty, które nadajemy odrębnym zmiennym) mogą istotnie różnić się od rzeczywistych wartości przyrostów tych

zmiennych. Choć jak pokażemy w kolejnych przykładach, wartości te mogą być do siebie bardzo zbliżone. Wszystko zależy od charakteru rozpatrywanego procesu. W przypadku wykrycia zbieżności możemy wykorzystać ten fakt w analizowaniu, prognozowaniu kolejnych wartości tego procesu.

Tab. 2.10 Wartości przyrostów cząstkowych i bezwzględnych dla modelu ze zmiennymi *x* i *y*

Punkt czasowy	ddx_i	ddy_i	δ_{x_i}	δ_{y_i}
P(0)	-0.1899	-0.1968	0.0156	0.1905
P(1)	0.6946	0.4683	0.2621	-0.0228
P(2)	0.0718	-0.0155	-0.2196	0.1172
P(3)	0.0395	-0.0089	0.1318	-0.0301
P(4)	0.5373	-0.2637	0.2348	-0.0472
P(5)	0.1010	-0.0113	-0.2715	0.2625
P(6)	-0.027	-0.057	0.3075	-0.1276
P(7)	-0.0814	0.169	0.5750	0.0190

Etap 2.

Znajdujemy teraz przyrosty cząstkowe modelu rozszerzonego o funkcje wrażliwości drugiego rzędu (zawierającego funkcje wrażliwości pierwszego i drugiego rzędu). Funkcje $X(x,y)$ oraz $Y(y,x)$ przedstawiamy w postaci:

$$X_{i+1}(x_i,y_i) = x_i + S_{x_i}^{X(x,y)} \cdot ddx_i + S_{y_i}^{X(x,y)} \cdot ddy_i + 0.5 \cdot S_{x_i^2}^{X(x,y)} \cdot (ddx_i)^2 + 0.5 \cdot S_{y_i^2}^{X(x,y)} \cdot (ddy_i)^2 + S_{x_i y_i}^{X(x,y)} \cdot ddx_i \cdot ddy_i. \tag{2.30}$$

$$Y_{i+1}(y_i,x_i) = y_i + S_{y_i}^{Y(y,x)} \cdot ddy_i + S_{x_i}^{Y(y,x)} \cdot ddx_i + 0.5 \cdot S_{y_i^2}^{Y(y,x)} \cdot (ddy_i)^2 + 0.5 \cdot S_{x_i^2}^{Y(y,x)} \cdot (ddx_i)^2 + S_{y_i x_i}^{Y(y,x)} \cdot ddy_i \cdot ddx_i. \tag{2.31}$$

Otrzymaliśmy układ dwóch równań nieliniowych względem niewiadomych wartości ddx_i oraz ddy_i. Dokładnie w taki sposób, jak wcześniej, podstawiamy wartości x_i, y_i oraz otrzymane wcześniej wartości funkcji wrażliwości (Tab. 2.3 i Tab. 2.4) i rozwiązujemy ten układ dla każdego z punktów czasowych $P(0)$, potem dla $P(1)$, $P(2)$ itd. i otrzymujemy wartości przyrostów ddx_i oraz ddy_i dla modelu rozszerzonego.

Trudność rozwiązania tego układu równań polega na tym, że składa się on z równań nieliniowych. Istnieją różne metody rozwiązania układów równań nieliniowych. Np. metoda Newtona-Ralphsona, czy metoda iteracji prostych. W naszych obliczeniach korzystamy z metody Newtona-Ralphsona oraz/lub funkcji *fsolve* z Matlaba. Z teorii rozwiązań równań nieliniowych wiadomo, że szybkość znalezienia rozwiązania, zbieżność procesu i dokładność rozwiązania zależą, niestety, od wyboru punktu

początkowego. Jako punkt początkowy, od którego startuje metoda numeryczna możemy wybrać wartości przyrostów ddx_i oraz ddy_i, znalezione wcześniej z układu równań uproszczonych (2.27). Tzn., jeśli szukamy ddx_0^{ext} i ddy_0^{ext}, to jako wartości początkowe, wybieramy obliczone wcześniej $ddx_0 = -0.1899$, $ddy_0 = -0.1968$; dla znalezienia ddx_1^{ext} i ddy_1^{ext} jako wartości początkowe wybieramy $ddx_1 = 0.6946$ i $ddy_1 = 0.4683$, itd. (z Tab. 2.9).

Dokładność rozwiązania równań często określana jest za pomocą tzw. **resztki**. Przenosząc lewą część równania (2.30) na prawą stronę i oznaczając otrzymane wyrażenie przez f_1 mamy:

$$f_1 = -X_{i+1}(x_i, y_i) + x_i + S_{x_i}^{X(x,y)} \cdot ddx_i + S_{y_i}^{X(x,y)} \cdot ddy_i + \\ +0.5 \cdot S_{x_i^2}^{X(x,y)} \cdot (ddx_i)^2 + 0.5 \cdot S_{y_i^2}^{X(x,y)} \cdot (ddy_i)^2 + S_{x_i y_i}^{X(x,y)} \cdot ddx_i \cdot ddy_i \tag{2.32}$$

W taki sam sposób tworzymy równanie dla $Y_{i+1}(y_i, x_i)$:

$$f_2 = -Y_{i+1}(y_i, x_i) + y_i + S_{y_i}^{Y(y,x)} \cdot ddy_i + S_{x_i}^{Y(y,x)} \cdot ddx_i + \\ +0.5 \cdot S_{y_i^2}^{Y(y,x)} \cdot (ddy_i)^2 + 0.5 \cdot S_{x_i^2}^{Y(y,x)} \cdot (ddx_i)^2 + S_{y_i x_i}^{Y(y,x)} \cdot ddy_i \cdot ddx_i \tag{2.33}$$

Oczywiście, jeśli znaleziono dokładne rozwiązanie każdego z tych równań względem zmiennych ddx_i oraz ddy_i, to $f_1 = 0$ oraz $f_2 = 0$. Natomiast, jeśli rozwiązanie pierwszego (drugiego) równania znaleziono z pewnym błędem, to $|f_1| > 0$ (odpowiednio $|f_2| > 0$). Wartość $|f_1|$ nazywana jest resztką pierwszego równania, a wartość $|f_2|$ – resztką drugiego równania. Resztką układu tych dwóch równań nazywamy wartość:

$$f = \sqrt{f_1^2 + f_2^2} \tag{2.34}$$

Im bliższą zeru wartość posiada resztka tym większą dokładność ma rozwiązanie układu równań.

W obliczeniach, których wyniki przedstawione są poniżej, podczas realizacji procesu rozwiązania układów równań nieliniowych skorzystano z wartości przyrostów z Tab. 2.9, jako wartości początkowych. Wyniki obliczeń przedstawiono w Tab. 2.11.

Staramy się zmniejszyć wartość resztki dobierając inne punkty początkowe. Gdy udaje się znaleźć inne, zadowalające rozwiązanie świadczy ono o tym, że powierzchnia przestrzeni (ddx_i, ddy_i) zawiera kilka minimów. To, które z nich zostanie znalezione, zależy od wyboru punktu początkowego w przestrzeni (ddx_i, ddy_i).

Tab. 2.11 Przyrosty cząstkowe ddx_i oraz ddy_i znalezione dla modelu rozszerzonego (2.30), (2.31) dla punktów czasowych *P(0) ... P(6)*

Punkty czasowe	Przyrost cząstkowy ddx_i	Przyrost cząstkowy ddy_i	Resztka rozwiązania f
P(0)	-0.1024	-0.129	$169 \cdot 10^{-15}$
P(1)	0.0799	0.0599	0.046
P(2)	0.0598	-0.0132	$144 \cdot 10^{-15}$
P(3)	0.0443	-0.0126	$429 \cdot 10^{-15}$
P(4)	0.05214	-0.01723	0.066
P(5)	0.0848	-0.0083	$654 \cdot 10^{-15}$
P(6)	-0.0383	-0.0684	$78 \cdot 10^{-18}$

Jak można zauważyć (Tab. 2.11), rozwiązania układu równań nieliniowych zostały znalezione z dużą dokładnością.

Etap 3.

Porównanie wyników, przedstawionych w Tab. 2.9 i Tab. 2.11, daje podstawy stwierdzić, że przyrosty cząstkowe ddx_i oraz ddy_i, znalezione przy wykorzystaniu modelu uproszczonego (2.25), (2.26), mogą istotnie różnić się od tych, otrzymanych przy wykorzystaniu modelu rozszerzonego (2.30), (2.31). Ten ostatni model można rekomendować do użytku we wszystkich przypadkach, gdzie dokładność rozwiązania problemów analizy procesów złożonych ma zasadnicze znaczenie (zakładając, że również wartości przyrostów znaleziono z dużą dokładnością).

Tab. 2.12 Porównanie przyrostów uzyskanych w modelu uproszczonym oraz rozszerzonym

Punkt czasowy	ddx_i w modelu uproszczonym	ddx_i w modelu rozszerzonym	ddy_i w modelu uproszczonym	ddy_i w modelu rozszerzonym
P(0)	-0.1899	-0.1024	-0.1968	-0.129
P(1)	0.6946	0.0799	0.4683	0.0599
P(2)	0.0718	0.0598	-0.0155	-0.0132
P(3)	0.0395	0.0443	-0.0089	-0.0126
P(4)	0.5373	0.05214	-0.2637	-0.01723
P(5)	0.1010	0.0848	-0.0113	-0.0083
P(6)	-0.027	-0.0383	-0.057	-0.0684

Poniżej (Tab. 2.13) przedstawiamy porównanie wcześniejszych wyników prognozy dla szeregu z Tab. 2.1 z rezultatami uzyskanymi w modelu uproszczonym $MAMC_{simp}$ i rozszerzonym $MAMC_{ext}$ z funkcjami wrażliwości na jeden krok wprzód. Dla każdej zmiennej wyniki prognozy uzyskiwane w modelach z funkcjami wrażliwości okazały się dokładniejsze. W przypadku zmiennej *z*, błąd względny prognozy wyniósł $\delta_z = 0.15\%$.

Wykorzystanie funkcji wrażliwości może, więc znacznie zwiększyć precyzję prognozy w stosunku do metody podstawowej GMDH. Pamiętać należy, że metoda podstawowa pracuje dobrze na dużej ilości danych uczących. Tutaj możemy zauważyć, że wprowadzenie narzędzia funkcji wrażliwości daje dobre rezultaty predykcji nawet na

niewielkiej próbce danych.

Tab. 2.13 Porównanie najlepszych wyników prognozy uzyskanej za pomocą uproszczonej metody GMDH, metody uproszczonej i rozszerzonej z funkcjami wrażliwości

Prognoza	Wartość rzeczywista	w uproszczonym modelu GMDH	w modelu $MAMC_{simp}$	w modelu $MAMC_{ext}$
x_7	0.3151	0.0815	0.4304	0.3861
y_7	0.2736	0.3398	0.288	0.299
z_7	0.3713	0.2725	0.344	0.3719

Przykład 2. Analiza procesu o większej ilości zmiennych

Procesy zachodzące w świecie rzeczywistym (fizyczne, chemiczne i inne) zwykle zależą od znacznej ilości powiązanych ze sobą zmiennych. W przypadku większej ilości zmiennych metoda z funkcjami wrażliwości również daje pozytywne rezultaty. Przeanalizujmy zachowanie szeregu o pięciu zmiennych Tab. 2.14 (skorzystano z danych dostępnych w zasobach internetowych z dziedziny elektroniki). Prognozę wartości każdego argumentu możemy wykonać korzystając z czterech funkcji cząstkowych (dla zmiennej *x* są to modele $X(x,y)$, $X(x,z)$, $X(x,q)$, $X(x,r)$).

Tab. 2.14 Szereg czasowy o pięciu zmiennych

Punkt czasowy	*x*	*y*	*z*	*q*	*r*
P(0)	-466	-202	-881	1255	-557
P(1)	-1362	-336	-945	1194	-710
P(2)	-1297	-448	-1005	181	-658
P(3)	-1035	-462	-988	-1735	-659
P(4)	-1146	-377	-1010	-2691	-540
P(5)	-927	-263	-1005	-3407	-259
P(6)	-369	-276	-924	-4000	-43
P(7)	31	-206	-945	-4000	413
P(8)	266	-40	-1030	-2369	1252

…

Tworzymy model funkcji cząstkowej $X(x,y)$ rozwiązując układ równań postaci (2.15). Otrzymujemy współczynniki WKG następujących wartości:

$a_0 = -311.02$, $a_1 = -10.414$, $a_2 = 31.976$, $a_3 = -0.0068$, $a_4 = 0.0196$, $a_5 = 0.0131$.

Następnie obliczamy wartości poszczególnych funkcji wrażliwości tego modelu według (2.21), Tab. 2.15.

Podobne obliczenia wykonujemy dla funkcji $Y(y,x)$, która potrzebna nam jest do obliczenia przyrostów ddx_i oraz ddy_i w modelach uproszczonym oraz rozszerzonym. Otrzymujemy następujące współczynniki WKG:

$b_0 = -1725.9$, $b_1 = -9.777$, $b_2 = -0.08595$, $b_3 = -0.0164$, $b_4 = -0.00061$, $b_5 = 0.0019$,

oraz wartości funkcji wrażliwości Tab. 2.16.

Tab. 2.15 Wartości funkcji wrażliwości w modelu $X(x, y)$

Punkt czasowy	$S_x^{X(x,y)}$	$S_y^{X(x,y)}$	$S_{x^2}^{X(x,y)}$	$S_{y^2}^{X(x,y)}$	$S_{xy}^{X(x,y)}$
P(0)	3127.101	-3622.65	-2959.89	1602.253	1234.251
P(1)	-5100.94	-310.413	-25284.7	4433.093	6000.434
P(2)	-1803.71	1174.56	-22928.9	7881.055	7618.759
P(3)	2446.749	-121.875	-14601.1	8381.317	6269.727
P(4)	-301.926	-809.07	-17900.8	5580.987	5664.9
P(5)	1137.246	-2496.93	-11712.9	2716.063	3196.699
P(6)	3322.082	-4498.8	-1855.91	2991.207	1335.37
P(7)	-419.653	-5004.45	-13.0987	1666.337	-83.7327

Tab. 2.16 Wartości funkcji wrażliwości w modelu $Y(y, x)$

Punkt czasowy	$S_y^{Y(y,x)}$	$S_x^{Y(y,x)}$	$S_{y^2}^{Y(y,x)}$	$S_{x^2}^{Y(y,x)}$	$S_{yx}^{Y(y,x)}$
P(0)	813.1166	-48.2637	-1338.03	-264.478	176.1599
P(1)	439.5049	-1285.8	-3702.03	-2259.29	856.4187
P(2)	-1113.84	-849.915	-6581.39	-2048.79	1087.396
P(3)	-1587.26	-320.848	-6999.16	-1304.67	894.8537
P(4)	-166.117	-692.479	-4660.63	-1599.51	808.5292
P(5)	759.4828	-510.661	-2268.16	-1046.59	456.2525
P(6)	391.1551	56.47642	-2497.93	-165.833	190.5922
P(7)	610.5986	-15.7859	-1391.54	-1.17042	-11.9508

Obliczamy teraz wartości przyrostów w modelu uproszczonym oraz rozszerzonym i porównujemy z rzeczywistymi wartościami przyrostów względnych. Wyniki przedstawione są w Tab. 2.17.

Tab. 2.17 Wartości przyrostów cząstkowych oraz bezwzględnych w modelu uproszczonym oraz rozszerzonym dla modelu $X(x, y)$ i $Y(y, x)$

Punkt czasowy	ddx_i^{simp}	ddx_i^{ext}	δ_{x_i}	ddy_i^{simp}	ddy_i^{ext}	δ_{y_i}
P(0)	-0.5127	-0.4163	1.922747	-0.19523	-0.1555	0.663366
P(1)	0.002347	-0.7438	-0.04772	-0.24797	-0.5457	0.333333
P(2)	-0.09157	-0.6385	-0.202	0.082442	-0.6179	0.03125
P(3)	-0.04755	-0.5001	0.107246	-0.04394	-0.5346	-0.18398
P(4)	-0.1095	-0.2256	-0.1911	-0.22982	-0.2215	-0.30239
P(5)	-0.9513	-0.0835	-0.60194	-0.65675	-0.2099	0.04943
P(6)	0.303425	0.3841	-1.08401	0.135148	0.2004	-0.25362
P(7)	-2.90608	-3.414	7.580645	0.196733	0.0248	-0.80583

Ponownie zauważamy różnicę w wartości przyrostów cząstkowych i bezwzględnych. Kontynuujemy, więc obliczenia z wykorzystaniem metody z funkcjami wrażliwości.

Analogiczne obliczenia wykonujemy dla pary funkcji $X(x, z)$ oraz $Z(z, x)$. Współczynniki WKG wynoszą odpowiednio:

$c_0 = 615.84, c_1 = -21.318, c_2 = 44.022, c_3 = 0.0077, c_4 = 0.0536, c_5 = -0.040,$

oraz

$d_0 = -17226, d_1 = -30.58, d_2 = -0.7593, d_3 = -0.0124, d_4 = 0.00143, d_5 = -0.0041.$

W Tab. 2.18 przedstawiono porównanie wartości przyrostów dla modelu $Z(z,x)$.

Tab. 2.18 Wartości przyrostów cząstkowych w modelu uproszczonym oraz rozszerzonym dla modelu $Z(z,x)$

Punkt czasowy	ddx_i^{simp}	ddx_i^{ext}	δ_{x_i}	ddz_i^{simp}	ddz_i^{ext}	δ_{z_i}
P(0)	-3.82991	0.248765	1.922747	-0.46362	-0.00761	0.072645
P(1)	0.043533	0.011481	-0.04772	-0.07451	-0.04001	0.063492
P(2)	0.024597	0.03496	-0.202	0.019157	0.017929	-0.01692
P(3)	0.096746	0.083129	0.107246	0.005906	0.001752	0.022267
P(4)	0.113787	0.034743	-0.1911	0.022081	0.014885	-0.00495
P(5)	-0.12015	-0.09848	-0.60194	0.001494	0.002263	-0.0806
P(6)	0.485381	0.357824	-1.08401	0.058788	0.045986	0.022727
P(7)	-3.32558	-2.98811	7.580645	0.035892	0.032631	0.089947

Dla pary funkcji $X(x,q)$ oraz $Q(q,x)$ współczynniki WKG wynoszą odpowiednio:
$e_0 = 15414$, $e_1 = 37.588$, $e_2 = -3.272$, $e_3 = 0.01931$, $e_4 = 0.000255$, $e_5 = -0.00181$,
oraz
$f_0 = -33028$, $f_1 = 7.395$, $f_2 = -73.115$, $f_3 = 0.00072$, $f_4 = -0.0381$, $f_5 = -0.00341$.
Oraz dla pary funkcji $X(x, r)$ oraz $R(r, x)$ współczynniki WKG wynoszą odpowiednio:
$g_0 = -21247.1$, $g_1 = -22.61$, $g_2 = -18.598$, $g_3 = 0.00942$, $g_4 = 0.0462$, $g_5 = -0.06705$,
oraz
$h_0 = 19764.1$, $h_1 = 20.555$, $h_2 = 19.297$, $h_3 = -0.05$, $h_4 = -0.0115$, $h_5 = 0.0692$.

Po wykonaniu tych obliczeń otrzymujemy ciąg wartości funkcji wrażliwości pierwszego i drugiego rzędu oraz przyrosty cząstkowe ddx, ddy, ddz, ddq, ddr w punktach czasowych $P(0)$, …, $P(7)$. Wszystkie te zmienne można uznać za funkcje kolejnych kroków czasowych $t = 0, 1, 2, 3, ...$ (czyli określone są w dyskretnych chwilach czasu). Można dokonać ekstrapolacji wartości każdej z tych funkcji, aby uzyskać ich hipotetyczne wartości w kolejnych chwilach czasu. W tym celu można skorzystać, np. z metod interpolacji i ekstrapolacji funkcji wielomianami Lagrange'a lub Newtona czy innych, np. Sheparda.

Jeśli uda się ekstrapolować te zmienne na przyszłe chwile czasu (choćby na kilka kroków wprzód), to na podstawie wzorów podobnych do (2.25) i (2.26) czy (2.30) i (2.31), można byłoby prognozować przyszłe wartości zmiennych *x*, *y*, *z*, *q*, *r* w punktach czasowych $P(7)$, $P(8)$, czy nawet kolejnych, podstawiając do tych wzorów ekstrapolowane wartości przyrostów cząstkowych ddx, ddy, ddz, ddq, ddr. Sugerujemy, więc w celu rozwiązania problemu predykcji skorzystać ze wzorów (2.25) i (2.26) czy (2.30) i (2.31) zamiast WKG (uproszczonej metody GMDH).

W celu ekstrapolacji wartości przyrostów cząstkowych (na więcej niż jeden krok wprzód) korzystamy z funkcji Matlaba: *interp1*. Dla porównania wykorzystuję też

algorytm Newtona.

Wyniki ekstrapolacji przyrostów cząstkowych nie są zadowalające. Rezultaty obliczeń są rozbieżne w stosunku do rzeczywistych wartości. Sprawdzamy wyniki predykcji, którą uzyskamy w modelu $X(x,y)$ wykorzystując wzór (2.25) oraz (2.30) i obliczone wcześniej wartości.

Rzeczywista wartość zmiennej *x* w punkcie czasowym $P(7)$ wynosi $x_7^{real} = 31$.

Wartość prognozowana zmiennej *x* obliczona za pomocą wzoru (2.25) wynosi:

$$X_7(x_6, y_6) = x_6 + S_{x_6}^{X(x,y)} \cdot ddx_6 + S_{y_6}^{X(x,y)} \cdot ddy_6 = -369 + 3322.08 \cdot (-0.622) + \\ +(-4498.8) \cdot (-0.473) = -309.68 .$$

A za pomocą modelu rozszerzonego (2.30):

$$X_7(x_6, y_6) = x_6 + S_{x_6}^{X(x,y)} \cdot ddx_6 + S_{y_6}^{X(x,y)} \cdot ddy_6 + 0.5 \cdot S_{x_6^2}^{X(x,y)} \cdot (ddx_6)^2 + \\ +0.5 \cdot S_{y_6^2}^{X(x,y)} \cdot (ddy_6)^2 + S_{x_6 y_6}^{X(x,y)} \cdot ddx_6 \cdot ddy_6 = -369 + 3322.08 \cdot (-0.17) + \\ +(-4498.8) \cdot (-0.24) + 0.5 \cdot (-1855.9) \cdot (-0.17)^2 + 0.5 \cdot (2991.2) \cdot (-0.24)^2 + \\ +1335.37 \cdot (-0.17) \cdot (-0.24) = 263.42 .$$

Rzeczywista wartość zmiennej *y* w punkcie czasowym $P(7)$ wynosi $y_7^{real} = -206$.

Wartość prognozowana zmiennej *y* obliczona za pomocą wzoru (2.26) wynosi:

$$Y_7(y_6, x_6) = y_6 + S_{y_6}^{Y(y,x)} \cdot ddy_6 + S_{x_6}^{Y(y,x)} \cdot ddx_6 = -276 + 391.1551 \cdot (-0.473) + \\ +56.476 \cdot (-0.622) = -496.21.$$

W modelu rozszerzonym (2.31):

$$Y_7(y_6, x_6) = y_6 + S_{y_6}^{Y(y,x)} \cdot ddy_6 + S_{x_6}^{Y(y,x)} \cdot ddx_6 + 0.5 \cdot S_{y_6^2}^{Y(y,x)} \cdot (ddy_6)^2 + \\ +0.5 \cdot S_{x_6^2}^{Y(y,x)} \cdot (ddx_6)^2 + S_{y_6 x_6}^{Y(y,x)} \cdot ddy_6 \cdot ddx_6 = -276 + 391.1551 \cdot (-0.24) + \\ +56.476 \cdot (-0.17) + 0.5 \cdot (-2497.93) \cdot (-0.24)^2 + 0.5 \cdot (-165.83) \cdot (-0.17)^2 + \\ +(190.59) \cdot (-0.24) \cdot (-0.17) = -446.45.$$

Natomiast przyszłe wartości zmiennej *x* uzyskane w pozostałych modelach uproszczonych są następujące (Tab. 2.19):

Tab. 2.19 Rezultaty predykcji zmiennej *x* we wszystkich modelach

		Modele uproszczone				**Modele rozszerzone**			
Modele	x_7^{real}	$X_7(x_6, y_6)$	$X_7(x_6, z_6)$	$X_7(x_6, q_6)$	$X_7(x_6, r_6)$	$X_7(x_6, y_6)$	$X_7(x_6, z_6)$	$X_7(x_6, q_6)$	$X_7(x_6, r_6)$
Rezultaty	31	−309.68	−490.7	−309	−3089.5	263.42	−22.77	−268.4	−2537.3

Dla wybranego zestawu danych, dla zmiennej *x*, nie uzyskaliśmy zadowalających rezultatów predykcji. Nie oznacza to jednak, że metoda jest nieprawidłowa, a jedynie, że nie działa satysfakcjonująco dla tak dobranych danych, gdyż prawdopodobnie są one nieregularne, bądź zawierają szum.

Inną przyczyną mogą być niedokładnie rozwiązane układy równań nieliniowych, czy niefortunnie dobrana metoda interpolacji stosowana do wyznaczenia przyszłych wartości przyrostów cząstkowych oraz funkcji wrażliwości.

Dla zmiennej *y* wyniki przedstawiają się następująco (Tab. 2.20):

Tab. 2.20 Rezultaty predykcji zmiennej *y* we wszystkich modelach

		Modele uproszczone				**Modele rozszerzone**			
Modele	y_7^{real}	$Y_7(y_6, x_6)$	$Y_7(y_6, z_6)$	$Y_7(y_6, q_6)$	$Y_7(y_6, r_6)$	$Y_7(y_6, x_6)$	$Y_7(y_6, z_6)$	$Y_7(y_6, q_6)$	$Y_7(y_6, r_6)$
rezultaty	-206	-496.2	-1297.4	-237	-2046.3	-446.5	-1175.8	-353.07	-1585.7

W tym przypadku udało się uzyskać rezultat zbliżony do wartości rzeczywistej w modelu cząstkowym uproszczonym $Y_7(y_6, q_6)$.

Podobne obliczenia wykonujemy dla pozostałych modeli. Przyrosty cząstkowe rozszerzonego modelu cząstkowego mogą być w niektórych punktach czasowych obarczone pewnym błędem w związku z problemem doboru właściwego punktu początkowego obliczeń numerycznych. We wszystkich przypadkach, wskutek wzajemnego wpływu zmiennych procesu, cząstkowe przyrosty zmiennych (czyli przyrosty, które nadajemy odrębnym zmiennym) istotnie różnią się od rzeczywistych wartości przyrostów tych zmiennych.

Dla zmiennej *z* otrzymujemy następujące wyniki predykcji na jeden krok wprzód (Tab. 2.21):

Tab. 2.21 Rezultaty predykcji zmiennej *z* we wszystkich modelach

		Modele uproszczone				**Modele rozszerzone**			
Modele	z_7^{real}	$Z_7(z_6, x_6)$	$Z_7(z_6, y_6)$	$Z_7(z_6, q_6)$	$Z_7(z_6, r_6)$	$Z_7(z_6, x_6)$	$Z_7(z_6, y_6)$	$Z_7(z_6, q_6)$	$Z_7(z_6, r_6)$
rezultaty	-945	-936.5	-715	-677.6	-886.5	-866.4	-733.8	-1664.6	-913.5

Dla zmiennej *q* otrzymujemy następujące wyniki predykcji na jeden krok wprzód (Tab. 2.22):

Tab. 2.22 Rezultaty predykcji zmiennej *q* we wszystkich modelach

		Modele uproszczone				**Modele rozszerzone**			
Modele	q_7^{real}	$Q_7(q_6, x_6)$	$Q_7(q_6, y_6)$	$Q_7(q_6, z_6)$	$Q_7(q_6, r_6)$	$Q_7(q_6, x_6)$	$Q_7(q_6, y_6)$	$Q_7(q_6, z_6)$	$Q_7(q_6, r_6)$
rezultaty	-4000	-3161	-4674.8	-5691.7	-4435.6	-3888.2	-4215.2	-1099.9	-4186.3

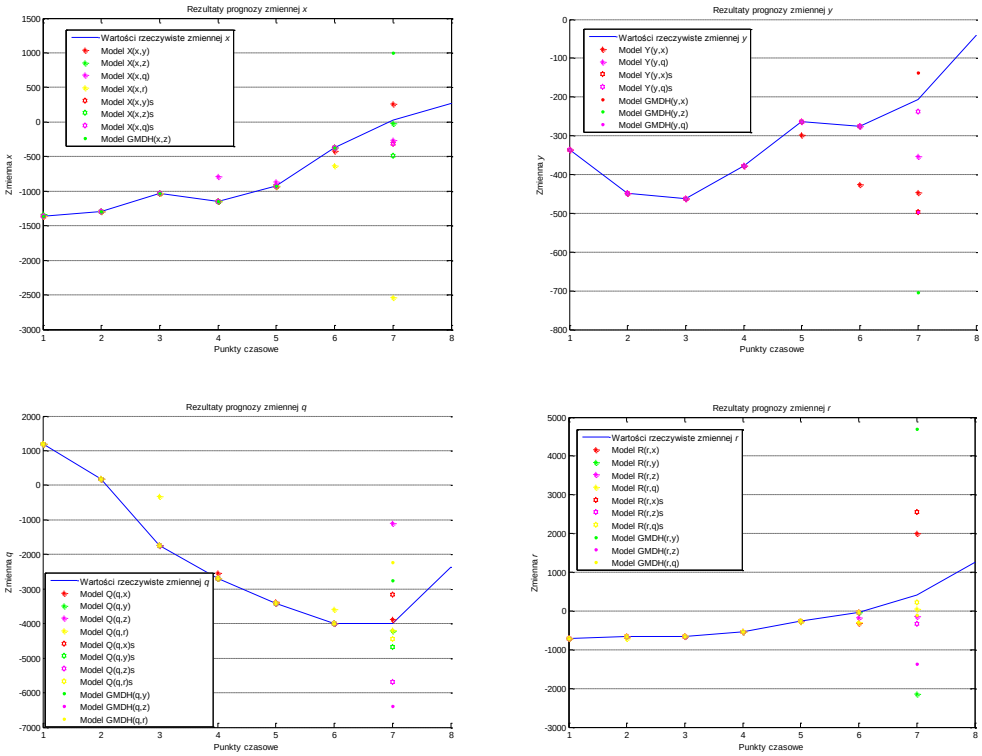
Dla zmiennej *r* otrzymujemy następujące wyniki predykcji na jeden krok wprzód (Tab. 2.23):

Tab. 2.23 Rezultaty predykcji zmiennej *r* we wszystkich modelach

		Modele uproszczone				**Modele rozszerzone**			
Modele	r_7^{real}	$R_7(r_6, x_6)$	$R_7(r_6, y_6)$	$R_7(r_6, z_6)$	$R_7(r_6, q_6)$	$R_7(r_6, x_6)$	$R_7(r_6, y_6)$	$R_7(r_6, z_6)$	$R_7(r_6, q_6)$
rezultaty	413	2554.8	-2873.8	-327.88	223.5	2006.3	-2148.5	-147.7	45.88

Trudno jest stosować znane metody statystyczne w celu badania wzajemnego wpływu zmiennych na siebie w przypadku tego procesu. Analizując rozpatrywany powyżej proces posiadamy jedynie niewielką ilość danych uczących, które niezbędne są do zbudowania poszczególnych modeli cząstkowych. Dodatkowo, z dużym

prawdopodobieństwem można przypuszczać, że zależności między poszczególnymi zmiennymi mają charakter nieliniowy, a narzędzia badające korelację między zmiennymi wskazują jedynie siłę zależności liniowej (bądź jej brak). Wskutek wzajemnego wpływu zmiennych na siebie, każdy z modeli cząstkowych (typu (2.25) czy (2.30)) daje odmienne rezultaty predykcji wartości poszczególnych zmiennych, nierzadko dosyć odległe od rzeczywistych, aczkolwiek w każdym przypadku, co najmniej jeden z modeli dostarczył zadowalające rezultaty (oznaczone kolorem zielonym). Porównując otrzymane wyniki z rezultatami uzyskanymi przez uproszczoną metodę GMDH ponownie zauważamy, że MAMC daje bardziej zbliżone do rzeczywistych wyniki. Przypomnijmy, że klasyczna metoda GMDH daje zadowalające wyniki przy znacznej liczbie danych inicjujących. Porównanie najbliższych właściwym, przykładowych rezultatów dla szeregu z Tab. 2.14 przedstawiono na Rys. 2.4.



Rys. 2.4 Rezultaty predykcji wartości zmiennych *x*, *y*, *q* oraz *r* dla danych z Tab. 2.14
Źródło: Opracowanie własne (Matlab)

Poza tym w przypadku metody GMDH brak jest jakiejkolwiek oceny dokładności prognoz uzyskanych przez poszczególne modele, więc nawet, gdy któryś z wyników jest bliski rzeczywistemu, to jesteśmy w stanie dokonać jedynie oceny rezultatu *ex post*. W przypadku naszego modelu proponujemy ocenę tej dokładności na podstawie porównania wyników prognozy otrzymanych za pomocą konkurujących modeli cząstkowych *ex ante*, wskazując w ten sposób model, którego wyniki przechodzą do

kolejnego kroku prognozy (wyznaczania przypuszczalnej wartości kolejnego punktu czasowego).

2.4 Ocena błędów prognozy

Powyższy przykład pokazuje, że w przypadku procesu złożonego może dojść do sytuacji, w której trudno jest ocenić stopień wiarygodności prognozy konkretnych zmiennych procesu, gdyż część z nich może być prognozowanych z wystarczającą dokładnością, a inne nie. W warunkach prognozy za pomocą powyższej metody nie korzystamy ze zbioru testowego, za pomocą którego moglibyśmy ocenić jej dokładność. W takich warunkach w zasadzie brak możliwości dokonania oceny bezpośredniej. Można natomiast zaproponować rozwiązanie pośrednie.

Podobne metody pośredniej oceny dokładności obliczeń wykorzystywane są w praktyce obliczeń numerycznych, w różnorodnych zagadnieniach matematyki stosowanej (rozwiązywanie układów równań różniczkowych), np. podczas oceny precyzji obliczeń kolejnego kroku całkowania w czasie. Stosujemy wtedy pewną metodę predykcji, która prognozuje, jaka będzie wartość poszukiwanej zmiennej na jeden krok wprzód (np. ekstrapolację Lagrange'a), a następnie inną metodę, która realizuje rolę korektora, w celu uściślenia wartości poszukiwanej zmiennej. Im bliższe sobie są wyniki obu metod, tym wiarygodność znalezionego rozwiązania jest większa, a błędy obliczeniowe mniejsze. Podobne postępowanie proponujemy w przypadku naszej metody.

Przypuśćmy, że mamy dany proces w przestrzeni zmiennych $\{x, y, z, q, r\}$. Oceńmy prognozę dla zmiennej *x*. W tym celu tworzymy modele cząstkowe $X(x, y)$, $X(x, z)$, $X(x, q)$, $X(x, r)$, za pomocą których uzyskujemy 8 różnych wartości prognozowanych (4 w modelach uproszczonych i 4 w rozszerzonych). Jeżeli te wartości są sobie bliskie, to możemy mieć pewność, że dokładność prognozy jest dobra, natomiast jeżeli te wartości są odległe od siebie (jak w przypadku szeregu z Tab. 2.14), może to świadczyć o tym, że dokładność prognozy jest niska.

Kryterium wiarygodności rozwiązań może być średnica zbioru wartości tych rozwiązań, rozumiana jako odległość pomiędzy najwyższą a najniższą wartością.

Dla modeli ze zmienną *x* (Tab. 2.14) średnica wynosi więc:

$263.42 - (-3089.5) = 3352.93.$

Poza tym, w ten sposób możemy również porównywać średnice różnych zbiorów,

tworzonych w celu prognozy różnych zmiennych. Dla modeli z pozostałymi zmiennymi otrzymujemy rezultaty przedstawione w Tab. 2.24.

Tab. 2.24 Porównanie średnic zbiorów prognoz w pierwszym kroku dla każdej ze zmiennych testowanego procesu z Tab. 2.14

Zmienna x	Zmienna y	Zmienna z	Zmienna q	Zmienna r
$\|263.42 - (-3089.5)\|$	$\|-237.02 - (-2046.4)\|$	$\|-1664.6 - (-677.6)\|$	$\|-1099.9 - (-5691.7)\|$	$\|2554.8 - (-2873.8)\|$
3352.93	1809.36	987.035	4591.8	5428.7

Oczywiście im większy rozrzut pomiędzy poszczególnymi rezultatami, tym mniejsze zaufanie do uzyskiwanej prognozy. Porównując średnice zbiorów wnioskujemy, że prawdopodobnie prognoza dla zmiennej *z* ma większą wiarygodność niż dla *y* czy *x*, a dla pozostałych zmiennych jest wyjątkowo niezadowalająca. Jednakże w przypadku zmiennych q oraz r, na tak niekorzystny wynik, wpływ mają rezultaty bardzo odległego od pozostałych pojedynczego modelu. Taki wynik można by wcześniej zidentyfikować i odrzucić przed ostatecznym dokonaniem wyboru modelu do kolejnego kroku prognozy.

Za pomocą odległości wyznaczanych pomiędzy poszczególnymi rezultatami prognozy możemy oceniać pośrednią dynamikę zmiany dokładności prognozy na kolejnych jej krokach. Mianowicie, po utworzeniu modeli w postaci WKG, które były budowane na próbkach z podzbioru Z_{ini}, wykonujemy jeden krok prognozy dla wszystkich zmiennych procesu i jej wyniki porównujemy z próbką z podzbioru Z_{test}. Prognoza zapewne będzie obarczona pewnym błędem. Jeżeli uznamy, że jest on dopuszczalny ze względu na ustalone kryterium, obliczamy średnicę uzyskanego zbioru rozwiązań dla każdej zmiennej uzyskując wartości δ_1^x, δ_1^y, δ_1^z, δ_1^q, δ_1^r. Następnie wykonujemy prognozę na kolejny krok wprzód, dla nowo uzyskanych wartości i ponownie oceniamy średnice δ_2^x, δ_2^y, $\delta_2^z, \delta_2^q, \delta_2^r$. Kontynuujemy ten proces dopóki po porównywaniu uzyskanych wartości z próbką testową otrzymujemy zadowalające rezultaty prognozy. Ostatecznie, gdy na pewnym kroku czasowym t_n, uzyskamy niezadowalające rezultaty δ_n^x, δ_n^y, $\delta_n^z, \delta_n^q, \delta_n^r$, wartości te przyjmujemy jako wskaźnik wyjścia prognozy poza zakres ufności. Czyli za każdym razem, gdy któraś ze średnic dla poszczególnych zmiennych przekroczy odpowiednio δ_n^x, δ_n^y, $\delta_n^z, \delta_n^q, \delta_n^r$ uznamy, że wyniki prognozy obarczone są zbyt dużym błędem.

Podczas prób prognozowania wartości szeregu najbardziej prawdopodobna jest sytuacja, w której wszystkie modele dają nieco różniące się rezultaty. Dlatego możemy też obliczyć różnice pomiędzy poszczególnymi prognozowanymi wartościami zmiennych w każdym z modeli $(\Delta_1(X(x,y)s, X(x,z)s)$, $\Delta_2(X(x,y)s, X(x,q)s)$,

$\Delta_3(X(x,y)s, X(x,q)s)$, …, Tab. 2.25), a następnie wykorzystując np. miarę odległości euklidesowej bądź/i jakiejś innej metryki (np. Canberra) w celu porównania rezultatów, sprawdzić, któremu modelowi pozostałe są najbliższe.

Tab. 2.25 Odległości pomiędzy wartościami uzyskanymi w każdym modelu ze zmienną *x*

Modele	$X(x,y)s$	$X(x,z)s$	$X(x,q)s$	$X(x,r)s$	$X(x,y)$	$X(x,z)$	$X(x,q)$	$X(x,r)$
$X(x,y)s$	0	181.0091	0.659016	2779.832	573.103	286.9107	41.28301	2227.66
$X(x,z)s$	181.0091	0	181.6681	2598.822	754.112	467.9197	222.2921	2046.651
$X(x,q)s$	0.659016	181.6681	0	2780.491	572.4439	286.2516	40.62399	2228.319
$X(x,r)s$	2779.832	2598.822	2780.491	0	3352.934	3066.742	2821.115	552.1716
$X(x,y)$	573.103	754.112	572.4439	3352.934	0	286.1923	531.8199	2800.763
$X(x,z)$	286.9107	467.9197	286.2516	3066.742	286.1923	0	245.6276	2514.571
$X(x,q)$	41.28301	222.2921	40.62399	2821.115	531.8199	245.6276	0	2268.943
$X(x,r)$	2227.66	2046.651	2228.319	552.1716	2800.763	2514.571	2268.943	0

Obliczamy wartości metryk dla każdego modelu (Tab. 2.26, Tab. 2.27).

Tab. 2.26 Porównanie wartości odległości euklidesowej oraz liczonej za pomocą metryki Canberra poszczególnych modeli dla prognozowanych wartości zmiennej *x* w każdym modelu

Modele	$X(x,y)s$	$X(x,z)s$	$X(x,q)s$	$X(x,r)s$	$X(x,y)$	$X(x,z)$	$X(x,q)$	$X(x,r)$
Odległość Euklidesa	3624.2	3441.7	3625	7150	4547.09	4031.6	3674.6	5808.4
Odległość Canberra	0.5203	0.3919	0.5206	0.1942	0.561	0.3519	0.4747	0.2054

Tab. 2.27 Porównanie wartości odległości euklidesowej oraz liczonej za pomocą metryki Canberra poszczególnych modeli dla prognozowanych wartości zmiennej *y* w każdym modelu

Modele	$Y(y,x)s$	$Y(y,z)s$	$Y(y,q)s$	$Y(y,r)s$	$Y(y,x)$	$Y(y,z)$	$Y(y,q)$	$Y(y,r)$
Odległość Euklidesa	2187.2	2010.2	2687.6	3554.6	2273.5	1868.9	2449.1	2507.4
Odległość Canberra	0.4926	0.5057	0.3328	0.3904	0.5122	0.5218	0.5044	0.4985

Z Tab. 2.26 i Tab. 2.27 wynika, że modele ze zmienną *x* są najbardziej skupione wokół modeli $X(x,y)s$, $X(x,q)s$, natomiast ze zmienną *y* wokół modeli $Y(y,z)s$ oraz $Y(y,z)$. Podobne czynności wykonujemy dla pozostałych zmiennych procesu (Tab. 2.28 – Tab. 2.30).

Tab. 2.28 Porównanie wartości odległości euklidesowej oraz liczonej za pomocą metryki Canberra poszczególnych modeli dla prognozowanych wartości zmiennej *z* w każdym modelu

Modele	$Z(z,x)s$	$Z(z,y)s$	$Z(z,q)s$	$Z(z,r)s$	$Z(z,x)$	$Z(z,y)$	$Z(z,q)$	$Z(z,r)$
Odległość Euklidesa	833.9	1022	1086.6	839.95	849.08	992.1	2253.8	833.691
Odległość Canberra	0.8115	0.7899	0.756	0.832	0.814	0.7663	0.492	0.8164

Tab. 2.29 Porównanie wartości odległości euklidesowej oraz liczonej za pomocą metryki Canberra poszczególnych modeli dla prognozowanych wartości zmiennej *q* w każdym modelu

Modele	$Q(q,x)s$	$Q(q,y)s$	$Q(q,z)s$	$Q(q,r)s$	$Q(q,x)$	$Q(q,y)$	$Q(q,z)$	$Q(q,r)$
Odległość Euklidesa	4154.4	4151	6148.1	3846.6	3559.5	3655.7	8731.8	3637.8
Odległość Canberra	0.659	0.759	0.644	0.7799	0.78	0.814	0.258	0.794

Tab. 2.30 Porównanie wartości odległości euklidesowej oraz liczonej za pomocą metryki Canberra poszczególnych modeli dla prognozowanych wartości zmiennej *r* w każdym modelu

Modele	$R(r,x)s$	$R(r,y)s$	$R(r,z)s$	$R(r,q)s$	$R(r,x)$	$R(r,y)$	$R(r,z)$	$R(r,q)$
Odległość Euklidesa	8901.5	9265.3	4902.2	4930.22	7647.9	7594.05	4856.59	4866.9
Odległość Canberra	0.401	0.3706	0.2615	0.276	0.34	0.3581	0.357	0.4034

Wartości wyznaczonych odległości mogą służyć za miarę błędu. Jeśli ich wartość nie przekracza zadanej apriorycznie granicy błędu, to uważamy, że wszystkie modele cząstkowe są dopuszczalne. Na bieżącym kroku prognozy wybieramy spośród nich ten model, do którego wszystkie inne są najbliższe, uznając go za najbardziej dokładny. Takie podejście ma tę przewagę nad klasyczną metodą GMDH, że po pierwsze, nie tworzymy złożonych modeli łącząc ze sobą poszczególne modele cząstkowe, jak to się robi w metodzie GMDH (komplikuje to istotnie proces obliczeniowy i jego sterowanie), po drugie, porównując modele cząstkowe ze sobą w pewien sposób oceniamy poziom błędu prognozy na każdym kolejnym kroku.

Jeśli wartość błędu przekracza dopuszczalną granicę, to przebudowujemy wszystkie modele cząstkowe, tworząc ponownie cząstkowe modele regresyjne na podstawie WKG (jak w metodzie GMDH).

Możemy również obliczyć zwykłą (arytmetyczną) średnią wszystkich uzyskanych dla każdej zmiennej rezultatów i wskazać modele dające wyniki najbardziej do tej średniej zbliżone (Tab. 2.31).

Tab. 2.31 Wartości średnich rezultatów prognozy dla każdej zmiennej rozpatrywanego procesu

Modele	Średnia arytmetyczna wartości modelu	Najbliższy średniej rezultat	Odpowiadający mu model
x	-845.5	354.8	$X_7(x_6, z_6)s$
y	-954.76	221.05	$Y_7(y_6, z_6)$
z	-924.25	10.7	$Z_7(z_6, r_6)$
q	-3919.1	30.88	$Q_7(q_6, x_6)$
r	-83.426	64.318	$R_7(r_6, z_6)$

Wyniki mogą być zaburzone poprzez wystąpienie pojedynczych, bardzo odległych od rzeczywistych wartości. Wpływają one na wielkość średnicy, tym samym już na samym początku określają stopień niepewności prognozy dla poszczególnych zmiennych. Biorąc pod uwagę wyznaczone miary odległości wskazujemy model, który według obranych kryteriów powinien być najkorzystniejszy, mając na uwadze, że wyniki prognoz są niepewne. (Tab. 2.32).

Do dalszych obliczeń wybieramy modele (wartości), które wskazane zostały przez powyższe obliczenia, jako najkorzystniejsze i powtarzamy procedurę.

Tab. 2.32 Wykaz modeli, które wybrano jako najkorzystniejsze dla każdej ze zmiennych, według obranych kryteriów

Zmienna	x	y	z	q	r
Wybór	$X(x,z)s$	$Y(y,z)$	$Z(z,r)$	$Q(q,y)$	$R(r,q)$
Rzeczywisty najlepszy model	$X(x,z)$, $X(x,y)$, $X(x,q)$	$Y(y,q)s$	$Z(z,x)s$, $Z(z,r)$	$Q(q,x)$, $Q(q,r)$, $Q(q,y)$	$R(r,q)s$, $R(r,q)$

Średnice dla poszczególnych zmiennych w kolejnym kroku przyjmują znacznie większe (gorsze) niż w poprzednim kroku wartości (Tab. 2.33).

Tab. 2.33 Porównanie średnic zbiorów prognoz w drugim kroku dla każdej ze zmiennych testowanego procesu z Tab 2.14

Zmienna x	Zmienna y	Zmienna z	Zmienna q	Zmienna r
$\lvert -18121.8 - (-153.5)\rvert$	$\lvert 13920.5 - (-29728.5)\rvert$	$\lvert -2005.1 - (-850.98)\rvert$	$\lvert 723.86 - (-34657.4)\rvert$	$\lvert -41993 - (10.831)\rvert$
17968.2	43649	1154.16	35381.3	52824.7

Możemy przypuszczać, że prognozy dla punktu czasowego $P(8)$ okażą się jeszcze mniej wiarygodne niż dla poprzedniego. Jednak nadal są dużo korzystniejsze niż w uproszczonej metodzie GMDH.

Tab. 2.34 Porównanie wartości odległości euklidesowej oraz liczonej za pomocą metryki Canberra poszczególnych modeli dla prognozowanych wartości zmiennej x w każdym modelu

Modele	$X(x,y)s$	$X(x,z)s$	$X(x,q)s$	$X(x,r)s$	$X(x,y)$	$X(x,z)$	$X(x,q)$	$X(x,r)$
Odległość Euklidesa	40755.1	20089	19532.8	16362.7	16308.4	20561.4	17689.3	15770.9
Odległość Canberra	0.1618	0.1985	0.2263	0.365	0.3997	0.5558	0.4029	0.3515

Tab. 2.35 Porównanie wartości odległości euklidesowej oraz liczonej za pomocą metryki Canberra poszczególnych modeli dla prognozowanych wartości zmiennej y w każdym modelu

Modele	$Y(y,x)s$	$Y(y,z)s$	$Y(y,q)s$	$Y(y,r)s$	$Y(y,x)$	$Y(y,z)$	$Y(y,q)$	$Y(y,r)$
Odległość Euklidesa	66074.7	38982.4	44330.3	80060.7	49023.8	38874	43887	63559
Odległość Canberra	0.3809	0.5019	0.3718	0.2824	0.511	0.5046	0.3268	0.184

Tab. 2.36 Porównanie wartości odległości euklidesowej oraz liczonej za pomocą metryki Canberra poszczególnych modeli dla prognozowanych wartości zmiennej z w każdym modelu

Modele	$Z(z,x)s$	$Z(z,y)s$	$Z(z,q)s$	$Z(z,r)s$	$Z(z,x)$	$Z(z,y)$	$Z(z,q)$	$Z(z,r)$
Odległość Euklidesa	1513.4	1195.2	1167.9	2469.7	1595.9	1283.6	2033	1515.4
Odległość Canberra	0.7507	0.7649	0.7471	0.5614	0.6802	0.7412	0.6178	0.7322

Tab. 2.37 Porównanie wartości odległości euklidesowej oraz liczonej za pomocą metryki Canberra poszczególnych modeli dla prognozowanych wartości zmiennej q w każdym modelu

Modele	$Q(q,x)s$	$Q(q,y)s$	$Q(q,z)s$	$Q(q,r)s$	$Q(q,x)$	$Q(q,y)$	$Q(q,z)$	$Q(q,r)$
Odległość Euklidesa	43239.7	79096.9	45074	44897.7	45891	77755	50381	43736.4
Odległość Canberra	0.449	0.206	0.5042	0.5077	0.6071	0.2478	0.3991	0.572

Miary odległości bardzo dobrze identyfikują wartości najbardziej odległe w każdym modelu. Natomiast, przy tak niskiej wiarygodności modeli odzwierciedlonej w zwiększonej wartości średnic każdego modelu, trudno spodziewać się precyzyjnego

wskazania modelu najlepszego (Tab. 2.34 – Tab. 2.38). Zbadajmy jeszcze najmniejszą odległość od średniej wyników dla każdego z modeli (Tab. 2.39).

Tab. 2.38 Porównanie wartości odległości euklidesowej oraz liczonej za pomocą metryki Canberra poszczególnych modeli dla prognozowanych wartości zmiennej *r* w każdym modelu

Modele	$R(r,x)s$	$R(r,y)s$	$R(r,z)s$	$R(r,q)s$	$R(r,x)$	$R(r,y)$	$R(r,z)$	$R(r,q)$
Odległość Euklidesa	61957.7	111701	71151.5	53508.6	59016.7	87125	54158	54075.5
Odległość Canberra	0.2437	0.1816	0.22656	0.1175	0.4466	0.2087	0.6055	0.6125

Możemy też zauważyć, że model, który w pierwszym kroku obliczeń nie był najlepszy, w kolejnym kroku daje korzystne rezultaty.

Tab. 2.39 Wartości średnich rezultatów prognozy dla każdej zmiennej rozpatrywanego procesu w drugim kroku prognozy

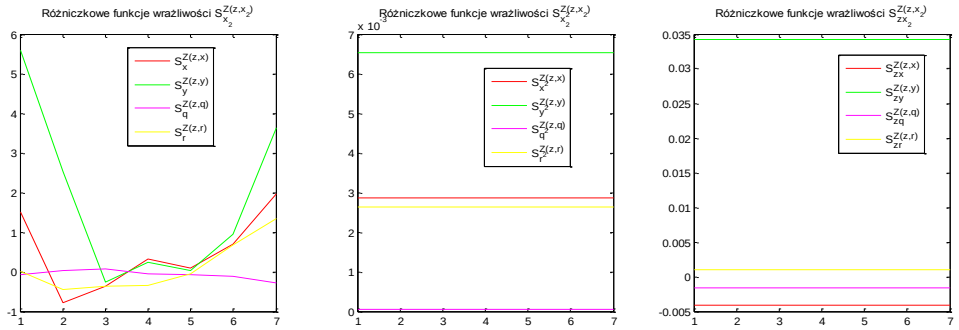
Modele	Średnia arytmetyczna wartości modelu	Najbliższy średniej rezultat	Odpowiadający mu model	Najlepszy rzeczywisty model
x	-4830.7	346.45	$X_8(x_7, r_7)$	$X(x,z)s, X(x,z)$
y	-4977.4	546.5	$Y_8(y_7, z_7)$	$Y(y,q), Y(y,q)s$
z	-1235.7	11.829	$Z_8(z_7, q_7)s$	$Z(z,y), Z(z,y)s$
q	-10398.9	6335.9	$Q_8(q_7, x_7)s$	$Q(q,z)s, Q(q,x)$
r	-6808.4	6019.5	$R_8(r_7, q_7)s$	$R(r,z), R(r,q)$

Niestety na drugim kroku prognozy nie możemy już liczyć na wskazania wybranych przez nas metryk. Możemy teraz przesunąć zakres badanych przez nas punktów czasowych i wykonać analogiczne obliczenia dla nowych wartości w celu próby uchwycenia ewentualnych regularności badanego szeregu czasowego.

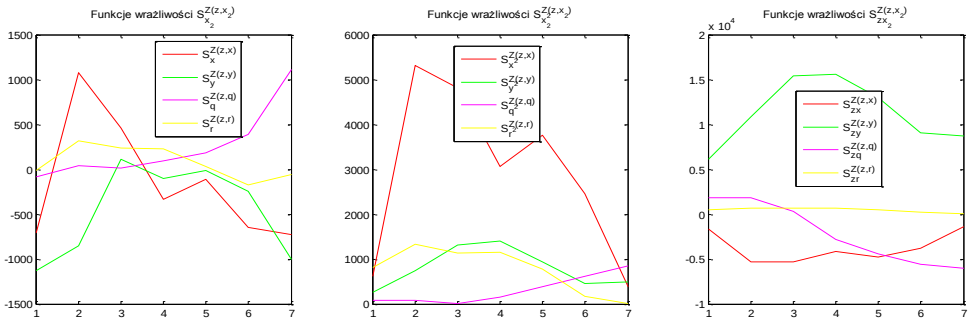
Dzięki zastosowaniu funkcji wrażliwości uzyskujemy też dodatkowe informacje dotyczące badanego szeregu, które badaczowi-ekspertowi dają wiedzę na temat wzajemnej zależności poszczególnych zmiennych szeregu pomagając w jego analizie. Jest ona tym bardziej ułatwiona, gdy badacz zna charakter (dziedzinę) badanego zbioru, czyli opiera się też na swojej wiedzy i doświadczeniu a nie tylko na narzędziu matematycznym.

Przykładowe wykresy funkcji wrażliwości różniczkowej, oraz półwzględnej (używanej w naszych obliczeniach) dla danych z Tab. 2.14 przedstawiają Rys. 2.5 – Rys. 2.8.

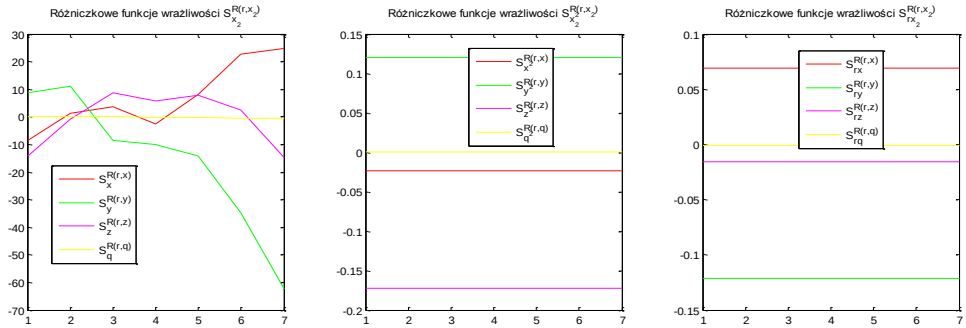
Im wyższa bezwzględna wartość funkcji wrażliwości, tym większy wpływ zmiennej, względem której jest liczona na wartość zmiennej głównej. Wartości bliskie zeru oznaczają brak wpływu. Bardzo wysokie wartości funkcji wrażliwości mogą też niestety dawać odległe wyniki predykcji, gdy wartości przyrostów cząstkowych nie zostaną znalezione z należytą dokładnością.



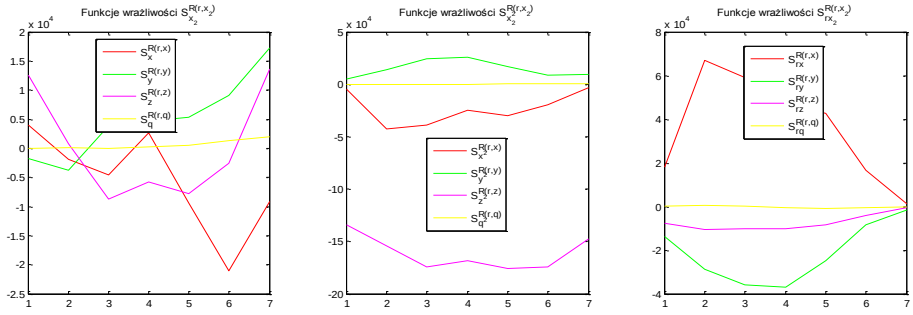
Rys. 2.5 Wykres różniczkowych funkcji wrażliwości I i II rzędu dla modelu ze zmienną *z*



Rys. 2.6 Wykres półwzględnych funkcji wrażliwości I i II rzędu dla modelu ze zmienną *z*



Rys. 2.7 Wykres różniczkowych funkcji wrażliwości I i II rzędu dla modelu ze zmienną *r*



Rys. 2.8 Wykres półwzględnych funkcji wrażliwości I i II rzędu dla modelu ze zmienną *r*

Najlepszy rzeczywisty model ze zmienną z to $Z(z,y)$, $Z(z,y)s$ (funkcje wrażliwości przyjmują wysokie wartości), a ze zmienną r to $R(r,z), R(r,q)$.

2.5 Wygładzanie analizowanych danych

Analiza danych pomiarowych otrzymanych w wyniku przeprowadzanych eksperymentów jest często uciążliwa, a czasami wręcz niemożliwa do wykonania również ze względu na szumy, które zawierają te dane. Mogą być one spowodowane licznymi czynnikami zewnętrznymi, ale również czułością urządzeń pomiarowych.
Łączny wpływ tych czynników sprawia, że wykres otrzymanych pomiarów jest obarczony dodatkowymi zakłóceniami. Prawidłowo przeprowadzona analiza wstępna (tzw. *pre-processing*) takich danych, oparta na algorytmach filtrowania, powinna umożliwić otrzymanie mniej lub bardziej uogólnionych trendów analizowanych danych. Natomiast dodatkowy szum, którym obarczone są pomiary, powinien zostać wyeliminowany. Oznacza to, że analizowanie pozbawionych szumu danych jest zdecydowanie łatwiejsze, a ostateczne wyniki bardziej wiarygodne. [73]

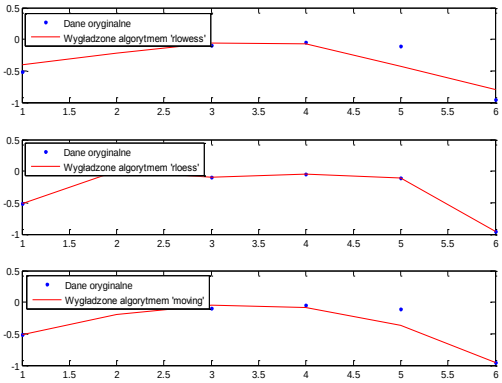
Obecnie mamy do dyspozycji wiele technik umożliwiających filtrowanie (wygładzanie) danych. Żadne z nich jednak nie mogą być traktowane w sposób uniwersalny, tzn., że wymagają dobierania odpowiednich parametrów w zależności od typu analizowanych danych i występujących w nich zakłóceń. Niejednokrotnie, nawet w przypadku takich samych pomiarów z różnymi parametrami wejściowymi, zachodzi konieczność doboru innych algorytmów filtrowania. Problem filtrowania danych jest w rzeczywistości problemem aproksymacji punktów pomiarowych. Część z nich umożliwia uzyskanie bardzo dokładnego dopasowania krzywej przefiltrowanej do przetwarzanego modelu danych. Jednakże dokładność otrzymywanych wyników okupiona jest wysokim kosztem obliczeń, a w przypadku dużej liczby punktów pomiarowych także bardzo długim czasem obliczeń. Inne cechuje duża szybkość działania, jednakże brak jest możliwości ustalenia właściwego warunku stopu algorytmu. Natomiast w wypadku mniej zakłóconych danych, przy zbyt dużej liczbie iteracji, metoda zwróci zbyt uogólnione wyniki, powodując, iż dalsza analiza tych danych jest mało wiarygodna. Na ostateczne wyniki filtrowania niemały wpływ ma więc też po prostu intuicja osoby filtrującej dane.

W metodzie MAMC filtrowanie (wygładzanie) danych można zastosować na etapie ekstrapolowania wartości wyznaczonych przyrostów cząstkowych. Na nielicznym zbiorze próbek danych inicjujących, jakie mamy do dyspozycji, często obserwujemy szum na pojedynczym punkcie czasowym. W przypadku wyznaczania przyrostów cząstkowych w modelu rozszerzonym szum może być spowodowany błędnymi

obliczeniami na etapie rozwiązywania układu równań nieliniowych (niemożności znalezienia dokładnego rozwiązania, źle dobranego punktu początkowego metody.)

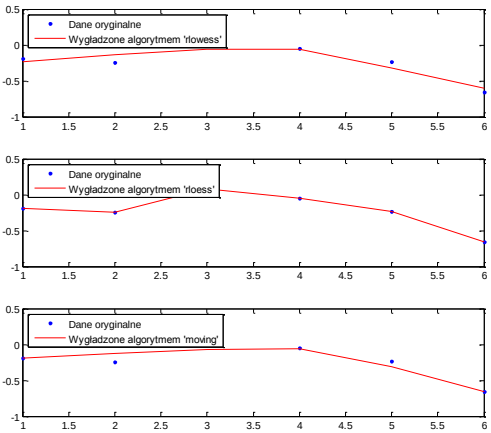
Stosujemy metody wygładzające na etapie ekstrapolacji wartości przyrostów cząstkowych do wszystkich zmiennych badanego powyżej szeregu czasowego (Tab. 2.14). Wybieramy kilka dostępnych w pakiecie Matlab metod z różnymi parametrami próbując uzyskać jak najdokładniejsze rezultaty. Metoda *rlowess* – metoda regresji lokalnej używająca liniowej metody najmniejszych kwadratów (pierwszego stopnia), *rloess* – metoda regresji lokalnej używająca liniową metodę najmniejszych kwadratów (drugiego stopnia), *moving* – metoda średnich kroczących.

Dla modelu $X(x,y)$ $(Y(y,x))$ rezultaty wygładzania wartości przyrostów cząstkowych ddx przedstawiają się następująco (Rys. 2.9):



Rys. 2.9 Porównanie wyników różnych sposobów wygładzenia wartości przyrostów cząstkowych ddx modelu $X(x,y)$

Dla przyrostów ddy uzyskujemy wyniki (Rys. 2.10):



Rys. 2.10 Porównanie wyników różnych sposobów wygładzenia wartości przyrostów cząstkowych ddy modelu $Y(y,x)$

Mimo, że efekty ekstrapolacji po wygładzeniu danych nadal odbiegają od rzeczywistych

wartości, to ostateczne wyniki predykcji uzyskane z ich zastosowaniem wykazały zwiększenie dokładności w porównaniu z wynikami bez zastosowania wygładzenia. Stąd też w naszych obliczeniach będziemy stosować metody wygładzania. Na podstawie badań eksperymentalnych wybrano metodę *rlowess* w celu wygładzenia otrzymywanych rezultatów przyrostów cząstkowych.

Funkcje wrażliwości możemy ekstrapolować w podobny sposób jak przyrosty cząstkowe, gdyż ich wartości uzyskane dla poszczególnych punktów czasowych tworzą pewien szereg czasowy. Jednakże stosowanie pewnych przybliżeń na etapie wygładzania tych danych, a następnie ich ekstrapolowania zwiększyłoby niepewność predykcji, dlatego też przyszłe wartości funkcji wrażliwości obliczamy ze wzorów (2.21), podstawiając w miejsce współczynników WKG $a_0, a_1, \dots, a_5$, początkowo wyznaczone wartości, a w miejsce zmiennych, ich predykowane wartości w kolejnym punkcie czasowym.

Funkcje wrażliwości można wykorzystać w celach predykcyjnych również w inny sposób, stosując je w równaniach stanów, za pomocą których można prognozować wartości samych przyrostów (cząstkowych czy pełnych) poszczególnych zmiennych. Tym samym prognozować te właśnie wartości zmiennych (Rozdział 3.).

3 Rozwiązanie problemów analizy zachowania się obiektów, sterowania nimi oraz ich prognozy w przestrzeni stanów

3.1 Ogólne cechy równań stanów obiektów

Jak już wspomniano w Rozdziale 2. rozprawy, badany obiekt można formalnie rozpatrywać jako „czarną skrzynkę", a zmienne charakteryzujące zachowanie się obiektu podzielić na dwie grupy (zmienne „wejściowe" i zmienne „wyjściowe"), które odzwierciedlają związki przyczynowo-skutkowe pomiędzy nimi. W teorii identyfikacji systemów [1], [63], taki sposób formalizacji opisu zachowania się obiektów jest nazywany *modelem wejście-wyjście*. Takie podejście do badanych procesów zostało przyjęte w Rozdziale 2. (patrz Rys. 2.2), w formie regresyjnych zależności zmiennych procesu $\{x_1(t_{N+1}), x_2(t_{N+1}), \dots, x_n(t_{N+1})\}$, określonych w punkcie czasowym t_{N+1} (zmienne te rozpatrywane są jako zmienne „wyjściowe" czarnej skrzynki), od zmiennych $\{x_1(t_N), x_2(t_N), \dots , x_n(t_N)\}$, określonych w poprzednim punkcie czasowym (zmienne „wejściowe"). W pewnych przypadkach podczas badania zachowania się obiektów wygodniej jest wprowadzić do modelu matematycznego dodatkowe zmienne, charakteryzujące „wewnętrzne" cechy badanych obiektów, takie jak np. częstotliwości własne obiektu czy inne. Równania zawierające zmienne wejściowe, wyjściowe oraz zmienne stanów, w teorii identyfikacji systemów nazywane są równaniami *wejście-stany-wyjście,* lub po prostu *równaniami (modelami) stanów obiektu* [1], [63].

Jak wiadomo [1], [63], w teorii identyfikacji systemów metody bazujące na narzędziu matematycznym równań stanów, tradycyjnie nazywane są metodami identyfikacji w przestrzeni stanów. Ważną przewagą tych metod jest ich uniwersalność, w tym sensie, że stosowane są do badań zarówno liniowych, jak i nieliniowych modeli obiektów. Korzyścią przedstawienia modeli matematycznych w przestrzeni stanów jest to, że pozwalają określić ciąg charakterystyk dynamicznych badanych obiektów, np. takich, jak sterowalność obiektów, ich obserwowalność, częstotliwości własne, stabilność i inne. Tak więc, ze względu na cechy modeli stanów, te ostatnie są ważnym narzędziem matematycznym do rozwiązania też problemów danej rozprawy.

W tym rozdziale proponujemy podejście do badania dynamicznych charakterystyk procesów, ich sterowania oraz prognozowania w przestrzeni stanów, rozwijane na podstawie modyfikacji cząstkowych modeli regresyjnych opisanych w poprzednim rozdziale danej rozprawy.

Jak wiadomo [63], pojęcie stanu jest pojęciem fundamentalnym, które nie może

zostać określone za pomocą bardziej „podstawowych" kategorii, podobnie, jak pojęcie „zbioru" w matematyce. Daje to podstawy do wyboru zmiennych stanów według uznania badacza, jednak pod warunkiem, że wybrane zmienne stanów i równania tworzone z ich wykorzystaniem spełniają pewne warunki. W teorii identyfikacji systemów warunki te zostały podzielone na dwie kategorie: warunki determinizmu zachowania się obiektów podlegających symulacji matematycznej oraz warunki dotyczące bezpośrednio równań stanów.

Niezależnie od tego, jaki sens fizyczny mają zmienne wejściowe, wyjściowe oraz zmienne stanów, w teorii identyfikacji systemów wspomniane zmienne formalnie rozpatrywane są jako „sygnały", co jednak nie wpływa na ogólny charakter przedstawionych definicji oraz wyników badań. Warunki determinizmu zachowania się obiektu przewidują, że zachowanie to w ogólności może zostać opisane matematycznie, chociaż niekoniecznie jedynie za pomocą równań stanów. A mianowicie, obiekt ten może zostać odniesiony do kategorii obiektów deterministycznych, jeśli jego zachowanie spełnia następujące warunki [63] (dalej literami pogrubionymi oznaczamy odpowiednie wektory):

1. Dla danego obiektu istnieje klasa funkcji czasowych $\boldsymbol{v}(t)$, nazywanych dopuszczalnymi funkcjami wejścia, które opisują sygnały podawane na wejściu obiektu (mogą to być nie tylko sygnały jako takie, ale również zewnętrzne wzbudzenia lub oddziaływania różnego pochodzenia, które powinny zostać wzięte pod uwagę podczas analizy obiektu).
2. Dla każdej chwili czasu *t* można skorzystać z pewnego zbioru zmiennych X_t, którego elementy $\boldsymbol{x}(t)$ opisują zmiany stanów badanego obiektu w czasie (zmienne te, jak już wspomniano powyżej, należą do kategorii zmiennych stanów obiektu).
3. Każdej parze $\boldsymbol{v}(t)$, $\boldsymbol{x}(t)$ odpowiada co najmniej jedna funkcja czasu, nazywana funkcją wyjścia $\boldsymbol{y}(t)$, oraz dla każdej chwili czasu $t' > t$ w $X_{t'}$, istnieje pojedynczy element $\boldsymbol{x}(t')$.

Aby można było uznać pewne równania zawierające zmienne wejściowe, wyjściowe oraz zmienne stanów za model stanów, powinny spełniać następujące wymagania [63]:

1. Jeśli $\boldsymbol{v_1}(t)$ i $\boldsymbol{v_2}(t)$ to dopuszczalne funkcje wejściowe, wtedy:
   $\boldsymbol{v_3}(t) = \boldsymbol{v_1}(t), t \leq t_0$;

$\boldsymbol{v_3}(t) = \boldsymbol{v_2}(t), t > t_0,$

powinna być również funkcją wejściową.

2. Przyszłe wartości zmiennych wyjściowych nie zależą od charakteru osiągania przez dany obiekt jego bieżącego stanu. Stan obiektu w chwili bieżącej oraz bieżące i przyszłe wartości jego wejść w sposób jednoznaczny określają bieżące i przyszłe wartości jego wyjść. Dlatego dla każdego $\boldsymbol{x}(t_0)$, należącego do X_{t_0}, oraz dla dopuszczalnych funkcji wejściowych $\boldsymbol{v_1}(t)$, $\boldsymbol{v_2}(t)$ (gdzie $\boldsymbol{v_1}(t) = \boldsymbol{v_2}(t)$ przy $t > t_0$) dowolna funkcja wyjściowa, odpowiadająca $\boldsymbol{x}(t_0)$, oraz $\boldsymbol{v_1}(t)$, jest identyczna do dowolnej funkcji wyjściowej, odpowiadającej $\boldsymbol{x}(t_0)$ i $\boldsymbol{v_2}(t)$ przy $t > t_0$.
3. Jeśli określone są początkowe stany obiektu i wejście $\boldsymbol{v}(t)$, $t \geq t_0$, to wyjście $\boldsymbol{y}(t)$, $t \geq t_0$ jest określone w sposób jednoznaczny.

Przedstawione wyżej trzy warunki można zapisać w postaci dwóch równań wektorowych, nazywanych równaniami stanów:

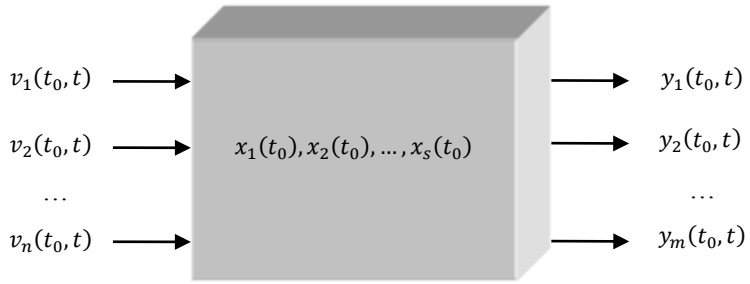
$$\dot{\boldsymbol{x}}(t) = \boldsymbol{f}[\boldsymbol{x}(t_0); \boldsymbol{v}(t_0, t)], \tag{3.1}$$

$$\boldsymbol{y}(t_0, t) = \boldsymbol{g}[\boldsymbol{x}(t_0); \boldsymbol{v}(t_0, t)], \tag{3.2}$$

gdzie $\boldsymbol{f}$ i $\boldsymbol{g}$ to jednoznaczne funkcje wektorowe. Z równania (3.2) wynika, że sygnał wyjściowy $\boldsymbol{y}$ na odcinku czasowym (t_0, t) jest jednoznaczną funkcją sygnału wejściowego $\boldsymbol{v}$ na tym odcinku oraz stanu na początku tego odcinka. A ponadto, z równania (3.1) wynika, że stan na końcu odcinka jest jednoznaczną funkcją stanu na początku odcinka.

Tak więc, wybór zmiennych wejściowych, wyjściowych oraz zmiennych stanów uwarunkowany jest właściwościami równań stanów oraz zmiennych wchodzących w skład tych równań. Ze względu na obecność zmiennych stanów, proces podlegający symulacji można formalnie przedstawić w formie obiektu, którego zachowanie się jest opisywane za pomocą trzech grup zmiennych wektorowych:
wektorem wejściowym $\boldsymbol{v}(t_0, t) = [v_1(t_0, t), \dots, v_n(t_0, t)]$, wektorem wyjściowym $\boldsymbol{y}(t_0, t) = [y_1(t_0, t), \dots, y_m(t_0, t)]$ oraz wektorem stanów $\boldsymbol{x}(t_0) = [x_1(t_0), \dots, x_s(t_0)]$, przy czym rozmiary tych wektorów zasadniczo mogą się różnić.

Korzystając, więc z terminologii teorii systemów, relacje (3.1) – (3.2) można formalnie rozpatrywać jako równania opisujące zachowanie się obiektu przedstawionego w postaci „czarnej skrzynki” (ignorując jej architekturę) z *n* sygnałami wejściowymi, *m* sygnałami wyjściowymi oraz *s* zmiennymi stanów (Rys. 3.1).



Rys. 3.1 Przedstawienie badanego procesu w formie obiektu, którego stany opisywane są za pomocą trzech grup zmiennych wektorowych: wektora sygnałów wejściowych $v(t_0,t)$, wektora sygnałów wyjściowych $y(t_0,t)$ oraz wektora stanów $x(t_0)$
Źródło: Opracowanie własne

Z przedstawionej wyżej analizy możemy wyciągnąć dwa wnioski. Po pierwsze, spełnienie warunków determinizmu jest wystarczającym warunkiem tego, aby dla danego obiektu można było wybrać zmienne stanów tak, aby były spełnione również warunki równań w przestrzeni stanów (np. w szczególnym przypadku zmienne wejściowe mogą być identyfikowane również jako zmienne stanów danego obiektu). Oraz, po drugie, w przypadku, gdy warunki determinizmu nie są spełnione, zachowanie się tego obiektu z reguły nie może zostać opisane przez jakikolwiek model matematyczny, bowiem warunki determinizmu odpowiadają warunkom immanentności modelu (patrz np. [1]). Dlatego, bez ograniczenia ogólności otrzymanych w danej pracy wyników, możemy uznać, że procesy podlegające badaniom, zawsze spełniają warunki determinizmu, a równania stanów, którymi będziemy się posługiwać dalej, spełniają wspomniane wyżej warunki modeli stanów.

Zmienne stanów obiektu można interpretować, jako minimalne informacje o obiekcie niezbędne do określenia przyszłych wartości jego zmiennych wyjściowych i zmiennych stanów przy znanych wartościach zmiennych wejściowych w chwili bieżącej. Okoliczności te wskazują na celowość posługiwania się równaniami stanów w rozwiązywaniu również problemów postawionych w danej pracy.

Modele stanów mogą być przedstawione w postaci schematów strukturalnych odzwierciedlających wzajemne związki między zmiennymi, wchodzącymi w skład tych modeli. W zależności od postaci równań stanów, mamy różne schematy strukturalne. Rozpatrzymy trzy schematy: 1) dla układu liniowych równań stanów; 2) dla układu nieliniowych równań stanów oraz 3) dla układu nieliniowych dyskretnych równań stanów w przyrostach.

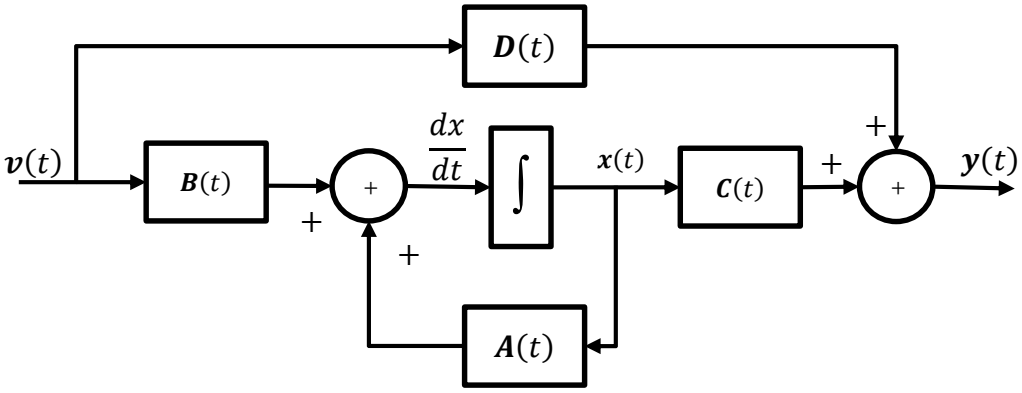
1. *Układ liniowych równań stanów*. Jeśli zachowanie się badanego obiektu pozwala na opis w postaci liniowych zależności między zmiennymi, to układ równań

(3.1) – (3.2) można przedstawić, jako układ liniowych równań różniczkowych i algebraicznych [63]:

$$\frac{d\boldsymbol{x}}{dt} = \boldsymbol{A}(t)\boldsymbol{x}(t) + \boldsymbol{B}(t)\boldsymbol{v}(t) \tag{3.3}$$

$$y(t) = \boldsymbol{C}(t)\boldsymbol{x}(t) + \boldsymbol{D}(t)\boldsymbol{v}(t) \tag{3.4}$$

gdzie $\boldsymbol{A}(t)$, $\boldsymbol{B}(t)$, $\boldsymbol{C}(t)$, $\boldsymbol{D}(t)$ to macierze ze stałymi lub zmieniającymi się w czasie elementami. Elementy macierzy $\boldsymbol{A}(t)$ są tymi parametrami, które określają wzajemne zależności między zmiennymi stanów (w teorii systemów [63] pokazano również, że właśnie za pomocą elementów tej macierzy można określić ciąg dynamicznych charakterystyk badanego obiektu, inwariantnych[9] w stosunku do sygnałów zewnętrznych, w tym częstotliwości własnych i innych). Elementy macierzy $\boldsymbol{B}(t)$ określają wpływ sygnałów wejściowych na wartości zmiennych stanów. Elementy macierzy $\boldsymbol{C}(t)$ określają wpływ zmiennych stanów na wartości zmiennych wyjściowych, a elementy macierzy $\boldsymbol{D}(t)$ odzwierciedlają bezpośredni wpływ sygnałów wejściowych na wartości sygnałów wyjściowych, jeśli obiekt ma bezpośrednie związki wejść i wyjść. Modelowi stanów (3.3) – (3.4) odpowiada schemat strukturalny pokazany na Rys. 3.2 [63]. Przebieg procesu dynamicznego w czasie zależy nie tylko od wartości wymuszeń w danej chwili, ale również od wartości tych wymuszeń w przeszłości. Można, więc powiedzieć, ze proces (układ) dynamiczny ma pamięć, w której są gromadzone skutki przeszłych oddziaływań.



Rys. 3.2 Schemat strukturalny odpowiadający modelowi stanów (3.3) – (3.4)
Źródło: [63]

2. *Układ nieliniowych równań stanów.* Równania stanów (3.1) – (3.2) w ogólnym przypadku mogą odzwierciedlać zarówno liniowe, jak i nieliniowe zależności

[9] W teorii sterowania układ, w którym sygnał na wyjściu nie zmienia się pod wpływem zakłócenia. W kontekście takiego układu mówi się też o sterowaniu inwariantnym

między wchodzącymi do nich zmiennymi. Ponadto zależności czasowe mogą być przedstawione zarówno jako funkcje ciągłej zmiennej czasowej, jak i dyskretnej zmiennej czasowej. W ostatnim przypadku uważa się, że zmienne te określone są w dyskretnej chwili czasu (np. podczas eksperymentalnych pomiarów tych zmiennych) $t_0, t_1, t_2, \dots, t_N, t_{N+1}, \dots$. Zbiory parametrów określone w dyskretnej chwili czasu nazywamy próbkami i przypisujemy im odpowiednio numery: 0, 1, 2, … , *N*, *N*+1, … . Modele zawierające zmienne określone w dyskretnej chwili czasu przyjmują postać *dyskretnych równań stanów* [1], [63]:

$$\boldsymbol{x}(NT + T) = \boldsymbol{f}[\boldsymbol{x}(NT); \boldsymbol{v}(NT)], \tag{3.5a}$$

$$\boldsymbol{y}(NT) = \boldsymbol{g}[\boldsymbol{x}(NT); \boldsymbol{v}(NT)], \tag{3.6a}$$

lub w uproszczonych oznaczeniach:

$$\boldsymbol{x}(N + 1) = \boldsymbol{f}[\boldsymbol{x}(N); \boldsymbol{v}(N)], \tag{3.5b}$$

$$\boldsymbol{y}(N) = \boldsymbol{g}[\boldsymbol{x}(N); \boldsymbol{v}(N)], \tag{3.6b}$$

przy czym przyjmujemy, że $\boldsymbol{f}$ i $\boldsymbol{g}$ to jednoznaczne, ciągłe, w ogólnym przypadku nieliniowe funkcje czasowe, a T – pewien dyskretny odczyt z chwili czasu (w przedstawionych równaniach przewiduje się, że odczyty czasowe realizowane są w równych odstępach czasowych *T*), *N* – to numer dyskretnej chwili czasu (czyli numer próbki). Zauważmy, jednak, że warunek jednakowych odstępów czasowych między kolejnymi próbkami nie jest obowiązkowy dla naszych badań, bowiem w używanych w danej pracy równaniach regresyjnych mamy zmienne określone jedynie na poprzedniej próbce, więc nie ma znaczenia, jakie odstępy czasu były wcześniej.

Niech zmienne stanów badanego procesu będą oceniane na próbkach N i $N + 1$, wówczas wektory stanów zawierają *n* zmiennych określonych na tych próbkach: $\boldsymbol{x}(N) = \{x_1(t_N), x_2(t_N), \dots, x_n(t_N)\}$ i $\boldsymbol{x}(N + 1) = \{x_1(t_{N+1}), x_2(t_{N+1}), \dots, x_n(t_{N+1})\}$. Ponadto, niech zmiennymi wyjściowymi $\boldsymbol{y}(N)$ są kombinacje liniowe zmiennych stanów, wówczas równania stanów będą miały postać [44]:

$$\boldsymbol{x}(N + 1) = \boldsymbol{f}[\boldsymbol{x}(N); \boldsymbol{v}(N)] \tag{3.7}$$

$$\boldsymbol{y}(N) = \boldsymbol{C}\boldsymbol{x}(N) \tag{3.8}$$

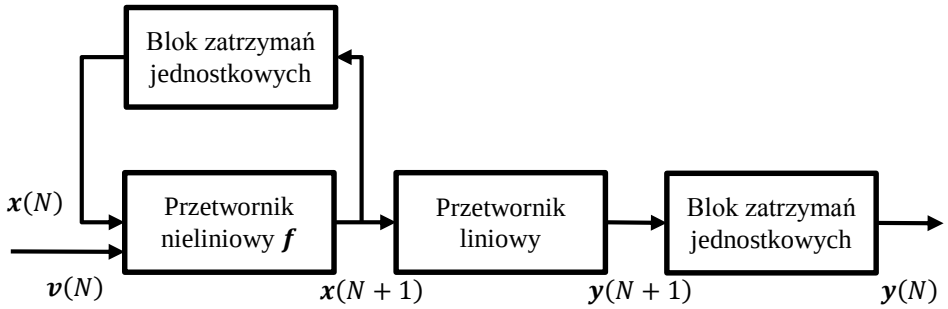
Danemu modelowi odpowiada schemat strukturalny na Rys. 3.3, [44].

3. *Układ nieliniowych dyskretnych równań stanów w przyrostach.* Jeśli po prawej stronie (3.7) funkcja wektorowa $\boldsymbol{f}$ jest funkcją różniczkowalną względem swoich argumentów w otoczeniu punktu t_N, to na podstawie nieliniowego układu równań

można otrzymać przybliżenie liniowe tego układu w otoczeniu punktu t_N stosując wzór Taylora; wynikiem tego przybliżenia będzie linearyzowany w otoczeniu punktu t_N układ równań stanów w przyrostach [1]:

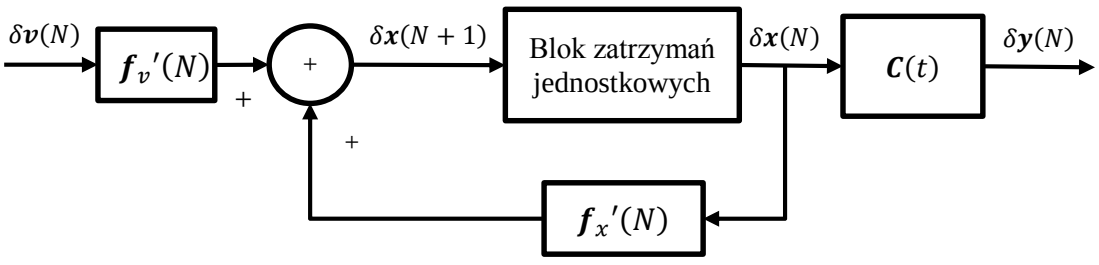
$$\delta x(N+1) = \boldsymbol{f}'_x(N)\delta\boldsymbol{x}(N) + \boldsymbol{f}_v{}'(N)\delta\boldsymbol{v}(N); \tag{3.9}$$

$$\delta\boldsymbol{y}(N) = \boldsymbol{C}\delta\boldsymbol{x}(N). \tag{3.10}$$



Rys. 3.3 Schemat strukturalny modelu stanów (3.7) – (3.8)
Źródło: [43]

Można przedstawić pewną odpowiedniość między układem równań stanów (3.3) – (3.4) oraz układem równań w przyrostach (3.9) – (3.10). Istotnie, wektorowi $\boldsymbol{x}(t)$ układu (3.3) – (3.4) odpowiada wektor $\delta\boldsymbol{x}(N)$ układu (3.9) – (3.10), wektorowi $\boldsymbol{v}(t)$ odpowiada wektor $\delta\boldsymbol{v}(N)$, wektorowi $d\boldsymbol{x}/dt$ odpowiada wektor $\delta\boldsymbol{x}(N+1)$, wektorowi $\boldsymbol{y}(t)$ odpowiada wektor $\delta\boldsymbol{y}(N)$, macierzy $\boldsymbol{A}(t)$ odpowiada macierz $\boldsymbol{f}'_x(N)$, oraz macierzy $\boldsymbol{B}(t)$ odpowiada macierz $\boldsymbol{f}_v{}'(N)$. Dalej skorzystamy z tych podobieństw przy budowie metod analizy badanych procesów tworzonych na podstawie równań w przyrostach. Tak więc, modelowi stanów (3.9) – (3.10) odpowiada schemat strukturalny pokazany na Rys. 3.4.



Rys. 3.4 Schemat strukturalny modelu stanów w przyrostach (3.9) – (3.10)
Źródło: Opracowanie własne

3.2 Związek równań stanów w przyrostach z cząstkowymi modelami regresyjnymi

Możemy sformułować następujący problem: wyrazić współczynniki równań stanów w przyrostach (3.9) – (3.10) poprzez współczynniki cząstkowych modeli regresyjnych rozpatrywanych w Rozdziale 2. W tym celu przeanalizujemy wybrane dwa cząstkowe modele regresyjne postaci (2.5), wprowadzając dla ułatwienia oznaczenia, z których korzystamy w bieżącym rozdziale:

$$X = a_0 + a_1 x + a_2 y + a_3 x^2 + a_4 y^2 + a_5 xy; \tag{3.11}$$

$$Y = b_0 + b_1 y + b_2 x + b_3 y^2 + b_4 x^2 + b_5 yx, \tag{3.12}$$

gdzie X i Y – to prognozowane wartości zmiennych x i y w kolejnym punkcie czasowym t_{N+1}, x i y – wartości zmiennych x i y w poprzednim punkcie czasowym t_N.

Zapiszmy równania (3.11) – (3.12) w przyrostach:

$$\delta X = a_0 + a_1(x+\delta x) + a_2(y+\delta y) + a_3(x+\delta x)^2 + a_4(y+\delta y)^2 + a_5(x+\delta x)(y+\delta y) - [a_0 + a_1 x + a_2 y + a_3 x^2 + a_4 y^2 + a_5 xy] = (a_1 + 2a_3 x + a_5 y)\cdot \delta x + (a_2 + 2a_4 y + a_5 x)\cdot \delta y + a_3(\delta x)^2 + a_4(\delta y)^2 + a_5 \delta x \delta y ; \tag{3.13}$$

$$\delta Y = (b_1 + 2b_3 y + b_5 x)\cdot \delta y + (b_2 + 2b_4 x + b_5 y)\cdot \delta x + b_3(\delta y)^2 + b_4(\delta x)^2 + + b_5 \delta x \delta y. \tag{3.14}$$

Teraz po porównaniu współczynników przy przyrostach δx oraz δy w (3.13) – (3.14) ze wzorami dla funkcji wrażliwości, które otrzymaliśmy w Rozdziale 2. (skorzystamy z definicji różniczkowej funkcji wrażliwości określonej wzorem (2.9), czyli $S_{x_i}^f = \frac{\partial f}{\partial x_i}$), mamy:

$$\delta X = S_x^X \delta x + S_y^X \delta y + \frac{1}{2} S_{x^2}^X (\delta x)^2 + \frac{1}{2} S_{y^2}^X (\delta y)^2 + S_{xy}^X \delta x \delta y; \tag{3.15}$$

$$\delta Y = S_y^Y \delta y + S_x^Y \delta x + \frac{1}{2} S_{y^2}^Y (\delta y)^2 + \frac{1}{2} S_{x^2}^Y (\delta x)^2 + S_{xy}^Y \delta x \delta y. \tag{3.16}$$

Lub w postaci macierzowej otrzymujemy:

$$\begin{bmatrix} \delta X \\ \delta Y \end{bmatrix} = \begin{bmatrix} S_x^X & S_y^X \\ S_x^Y & S_y^Y \end{bmatrix} \begin{bmatrix} \delta x \\ \delta y \end{bmatrix} + \begin{bmatrix} \frac{1}{2} S_{x^2}^X & \frac{1}{2} S_{y^2}^X & S_{xy}^X \\ \frac{1}{2} S_{x^2}^Y & \frac{1}{2} S_{y^2}^Y & S_{xy}^Y \end{bmatrix} \begin{bmatrix} (\delta x)^2 \\ (\delta y)^2 \\ (\delta x)\cdot(\delta y) \end{bmatrix}. \tag{3.17}$$

Można więc zauważyć, że między równaniami (3.9) i (3.17) jest pewna odpowiedniość, a mianowicie:

$$\delta\boldsymbol{x}(N+1)=\begin{bmatrix}\delta X\\ \delta Y\end{bmatrix};\qquad \delta\boldsymbol{x}(N)=\begin{bmatrix}\delta x\\ \delta y\end{bmatrix};\qquad \boldsymbol{f}'_x(N)=\begin{bmatrix}S_x^X & S_y^X\\ S_x^Y & S_y^Y\end{bmatrix};$$
$$\boldsymbol{f}'_y(N)=\begin{bmatrix}\frac{1}{2}S_{x^2}^X & \frac{1}{2}S_{y^2}^X & S_{xy}^X\\ \frac{1}{2}S_{x^2}^Y & \frac{1}{2}S_{y^2}^Y & S_{xy}^Y\end{bmatrix};\ \delta\boldsymbol{v}(N)=\begin{bmatrix}(\delta x)^2\\ (\delta y)^2\\ (\delta x)\cdot(\delta y)\end{bmatrix}. \tag{3.18}$$

Z równania (3.17) wynika ważny wniosek. Znając cząstkowe przyrosty δx i δy na poprzedniej próbce N, możemy obliczyć cząstkowe przyrosty tych zmiennych na kolejnej próbce, mamy więc do czynienia z prognozą przyrostów.

3.3 Prognoza za pomocą równań stanów w przyrostach

Na początku należy się zastanowić, co należy rozumieć przez δx oraz δy z (3.17), (3.18). Czy są to przyrosty cząstkowe ddx i ddy na poprzedniej próbce N, wtedy δX oraz δY oznaczaja cząstkowe przyrosty ddx i ddy na kolejnej próbce $(N+1)$. Jeżeli tak jest, to otrzymane wzory można zastosować do prognozowania tychże wartości. Jednakże przyrosty δx i δy oraz δX i δY mogą oznaczać zupełnie inne wielkości. Być może przyrosty te mają sens nie cząstkowych, a pełnych czy względnych przyrostów tychże zmiennych. W tym celu przeprowadzimy eksperymenty liczbowe. Każda możliwość jest prawdopodobna i użyteczna.

3.3.1 Prognoza stanów w przyrostach dla szeregu generowanego za pomocą wzorów matematycznych

Przeanalizujmy przypadek, w którym δX oraz δY oznaczają przyrosty cząstkowe obliczane za pomocą układu równań nieliniowych ((2.30) i (2.31)). Obliczenia wykonujemy na zbiorze danych Tab. 3.1.

Tab. 3.1 Szereg czasowy o trzech zmiennych, których dane wygenerowano za pomocą wzorów matematycznych

Punkty czasowe	*x*	*y*	*z*
P(0)	1.21	0.51076	1.424456
P(1)	1.331	0.577159	1.517032
P(2)	1.4641	0.652189	1.601241
P(3)	1.61051	0.736974	1.671262
P(4)	1.771561	0.832781	1.71972
P(5)	1.948717	0.941042	1.737606
P(6)	2.143589	1.063378	1.714393
P(7)	2.357948	1.201617	1.63849
P(8)	2.593742	1.357827	1.498261

…

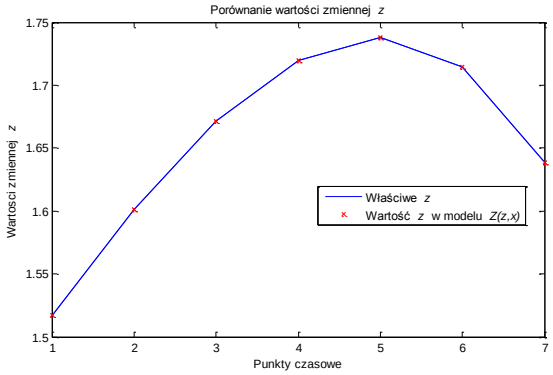
W celu zilustrowania zagadnienia wybieramy model $Z(z,x)$. Po rozwiązaniu

układu równań ze zmiennymi x i z otrzymujemy wartości przyrostów ddx i ddz (Tab. 3.2). Dla porównania obliczamy również wartości przyrostów względnych dla zadanego szeregu. Okazuje się, że nie odbiegają one znacznie wartościami od przyrostów cząstkowych. Tym samym dalsze wyniki obliczeń dla obu rodzajów przyrostów są podobne.

Tab. 3.2 Tabela zawierająca wartości przyrostów cząstkowych ddx oraz ddz w modelu $Z(z,x)$

Punkty czasowe	ddx	*Przyrosty względne*	ddz	*Przyrosty względne*
P(1)	0.090909	0.1	0.067894	0.06499
P(2)	0.090909	0.1	0.061137	0.05551
P(3)	0.090909	0.1	0.052615	0.043729
P(4)	0.090909	0.1	0.041763	0.028995
P(5)	0.090909	0.1	0.027772	0.010401
P(6)	0.090909	0.1	0.009423	-0.01336
P(7)	0.090909	0.1	-0.01721	-0.04427
P(8)	0.090909	0.1	-0.04935	-0.08558

Rozwiązania równań nieliniowych nierzadko bywają obarczone błędem, dlatego sprawdzamy, jak dokładne wyniki otrzymaliśmy. Podstawiamy w tym celu obliczone wartości przyrostów cząstkowych i wartości funkcji wrażliwości modelu $Z(z,x)$ ponownie do wzoru typu (2.30). Uzyskane wyniki przedstawione zostały na Rys. 3.5.



Rys. 3.5 Porównanie rzeczywistych wartości zmiennej z oraz uzyskanej w modelu $Z(z,x)$
Źródło: Opracowanie własne (Matlab)

Resztka rozwiązania układu równań liniowych jest bliska zeru. Za pomocą narzędzia Matlaba *fsolve* znaleziono bardzo dokładne rozwiązania układów równań dla poszczególnych punktów czasowych. Następnie wykorzystując uzyskane rezultaty stosujemy wzory (3.15) oraz (3.16) wstawiając w miejsce δX i δY wartości przyrostów cząstkowych ddx i ddz (oraz, dla porównania – wartości przyrostów względnych) modelu rozszerzonego oraz wartości funkcji wrażliwości dla modeli $X(x,z)$ i $Z(z,x)$ i rozwiązujemy ponownie układ równań nieliniowych. Po jego rozwiązaniu znajdujemy

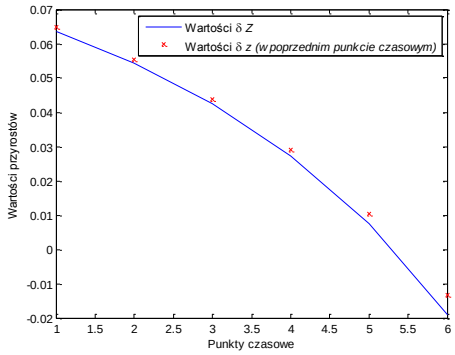
wartości przyrostów δx i δz, Tab. 3.3.

Podobnie, jak powyżej – sprawdzamy dokładność rozwiązań (resztkę rozwiązania). Bardzo podobne rezultaty uzyskujemy dla przyrostów względnych.

Tab. 3.3 Wartości przyrostów δx oraz δz dla modeli $X(x,z)$ i $Z(z,x)$

Punkty czasowe	δx	δz
P(1)	0.090909	0.070886
P(2)	0.090909	0.063561
P(3)	0.090909	0.054298
P(4)	0.090909	0.042494
P(5)	0.090909	0.02729
P(6)	0.090909	0.007419
P(7)	0.090909	-0.01909

Przystępujemy do sprawdzenia wzoru (3.15), który wskazuje, że przyrost δz oznacza przyrost cząstkowy na próbce N, podczas gdy δZ oznacza cząstkowy przyrost ddz na kolejnej próbce $(N+1)$. Wyniki porównania przedstawione są na Rys. 3.6.



Rys. 3.6 Porównanie wartości przyrostów δz w poprzednim punkcie czasowym z wartościami przyrostów δZ w bieżącym punkcie w modelu $Z(z,x)$
Źródło: Opracowanie własne (Matlab)

Dzięki uzyskanej dokładności możemy prognozować wartości przyrostów cząstkowych na przyszłą chwilę czasu, tym samym prognozując wartości szeregu.

Czyli, znając wartości przyrostów δX i δZ, znamy też wartość δx i δz w kolejnym punkcie czasowym.

$\delta X_6 = 0.0909, \delta x_7 = 0.0909, \delta Z_6 = -0.01721, \delta z_7 = -0.01721.$

Znając wartości przyrostów oraz wartości różniczkowych i półwzględnych funkcji wrażliwości obliczamy wartości δX_7 oraz δZ_7.

$$S'_{x_7,z_7} = \boldsymbol{f}'_x(N) = \begin{vmatrix} 1.1 & 33 \cdot 10^{-12} \\ 1.223 & -0.234 \end{vmatrix};$$

$$S''_{x_7,z_7} = \boldsymbol{f}'_y(N) = \begin{vmatrix} 29 \cdot 10^{-12} & -9 \cdot 10^{-12} & 795 \cdot 10^{-15} \\ 0.037 & 0.0733 & -0.072 \end{vmatrix}.$$

Otrzymujemy więc:

$\delta X_7 = 1.1 \cdot 0.0909 + 0 \cdot (-0.01721) + 0 \cdot (0.0909)^2 + 0 \cdot (-0.01721)^2 +$

$+0 \cdot 0.0909 \cdot (-0.01721) = 0.099.$

$\delta Z_7 = 1.223 \cdot (-0.01721) + (-0.234) \cdot 0.0909 + 0.037 \cdot (-0.01721)^2 +$

$+0.0733 \cdot (0.0909)^2 + (-0.072) \cdot 0.0909 \cdot (-0.01721) = -0.042.$

Rzeczywista wartość $\delta Z_7^{real} = -0.049.$

Obliczamy wartość zmiennej *x* podstawiając uzyskane wartości przyrostów cząstkowych oraz funkcji wrażliwości do wzoru typu (2.30).

$x_7 = 2.144,$

$S_z^{X(x,z)} = 2.358;\ S_x^{X(x,z)} \approx 0;\ S_{z^2}^{X(x,z)} \approx 0;\ S_{x^2}^{X(x,z)} \approx 0;\ S_{xz}^{X(x,z)} \approx 0.$

$x_8 = 2.114 + 2.358 \cdot 0.099 + 0 \cdot (-0.042) + 0 \cdot (0.099)^2 + 0 \cdot (-0.042)^2 +$

$+0 \cdot (-0.042) \cdot 0.099 = 2.34$

$x_8^{real} = 2.358$

Następnie obliczamy wartość zmiennej *z* podstawiając uzyskane wartości przyrostów cząstkowych oraz funkcji wrażliwości do wzoru typu (2.30).

$z_7 = 1.7144,$

$S_z^{Z(z,x)} = 2.0979;\ S_x^{Z(z,x)} = -0.504;\ S_{z^2}^{Z(z,x)} = 0.109;\ S_{x^2}^{Z(z,x)} = 0.337;\ S_{xz}^{Z(z,x)} = -0.264.$

$z_8 = 1.7144 + 2.0979 \cdot (-0.042) + (-0.504) \cdot 0.099 + 0.109 \cdot (-0.042)^2 +$

$+0.337 \cdot (0.099)^2 + (-0.264) \cdot (-0.042) \cdot 0.099 = 1.581$

Natomiast rzeczywista wartość $z_8^{real} = 1.638.$

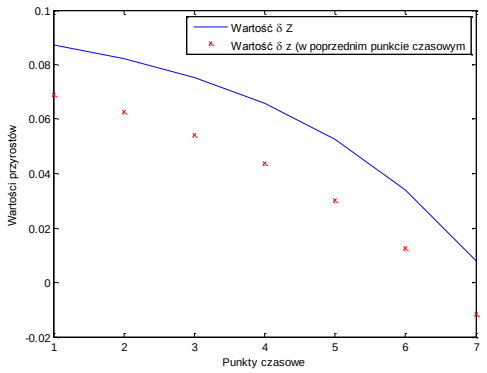
Błąd względny predykcji wynosi jedynie $\frac{|1.638-1.7144|}{1.638} = 0.0466.$

Dla zadanego powyżej szeregu, charakteryzującego się znacznymi regularnościami, można było również wykorzystać przyrosty względne (których wartości są bardzo bliskie przyrostom cząstkowym).

Jak można zauważyć na Rys. 3.6, wartości przyrostów δz niemalże pokrywają się z wartościami przyrostów δZ (wartości przyrostów δz w następnym punkcie czasowym). Otrzymując więc wartości przyrostów cząstkowych, możemy wyznaczyć prognozę wartości poszczególnych zmiennych szeregu na jeden krok wprzód. Musimy jedynie zidentyfikować charakter przyrostów δx i δz (czy są to przyrosty względne, bezwzględne, czy cząstkowe w modelu uproszczonym czy rozszerzonym na poprzednim kroku czasowym). Wtedy bez konieczności obliczania wartości prognoz w każdym z możliwych modeli i dokonywania wyboru najbardziej prawdopodobnego modelu poprzez analizę odległości (metryki) pomiędzy poszczególnymi rezultatami jesteśmy

w stanie prognozować wartości zmiennych szeregu. Charakter przyrostów δx i δz badamy porównując poszczególne wartości, jednocześnie w ten sposób wybieramy model, dla którego porównanie przyrostów okazuje się korzystniejsze.

Dla powyższego przykładu wartości δx i δz są odpowiednimi przyrostami cząstkowymi dla zmiennej x oraz z. W przypadku modelu ze zmienną y ($Z(z,y)$), porównanie przyrostów nie wypadło aż tak korzystnie (Rys. 3.7).



Rys. 3.7 Porównanie wartości przyrostów δz w poprzednim punkcie czasowym z wartościami przyrostów δZ w bieżącym punkcie w modelu Z(z,y)
Źródło: Opracowanie własne (Matlab)

3.3.2 Prognoza stanów w przyrostach dla szeregu odwzorowującego zależności fizyczne

Zastosujemy powyższy schemat obliczeń dla szeregu o znacznie bardziej skomplikowanych własnościach. Dane wykorzystane w obliczeniach pobrane zostały z ogólnodostępnych źródeł zasobów internetowych. W celu ilustracji wybieramy szereg o trzech zmiennych, Tab. 3.4.

Tab. 3.4 Szereg czasowy o trzech zmiennych.

Punkty czasowe	*x*	*y*	*z*
P(0)	1.6643	4.0362	6.9496
P(1)	1.7672	4.3238	7.2398
P(2)	1.8092	4.3153	7.5156
P(3)	1.8566	4.1618	7.8968
P(4)	1.9977	3.9617	8.6687
P(5)	2.1454	3.7198	8.9401
P(6)	2.0683	3.5165	8.4801
P(7)	2.2091	3.24	8.6287
P(8)	2.3371	3.2235	8.6201

…

Do badań wybieramy model $X(x,y)$. Po rozwiązaniu układu równań nieliniowych typu (2.30) i (2.31) ze zmiennymi x i y otrzymujemy wartości przyrostów Tab. 3.5. Dla

porównania korzystamy z metody Newtona-Ralphsona i metod dostępnych w pakiecie Matlab (funkcja *fsolve*).

Tab. 3.5 Wartości przyrostów cząstkowych ddx oraz ddy w modelach $X(x,y)$ i $Y(y,x)$

Punkty czasowe	ddx fsolve	ddx Newton-Ralphson	ddy fsolve	ddy Newton-Ralphson
P(1)	0.262927	0.017085	0.139656	0.02468
P(2)	0.161574	0.002567	0.053906	-0.00893
P(3)	0.054146	-0.013	0.020249	-0.03922
P(4)	0.065437	0.065437	-0.04018	-0.04018
P(5)	0.028778	0.159809	-0.02651	-0.22097
P(6)	0.007783	0.007783	-0.01117	-0.01117
P(7)	-0.03621	-0.00717	0.012574	0.018219

Sprawdzamy, jak dużym błędem obarczone zostały obliczenia. Resztka rozwiązania w przypadku modelu $X(x,y)$ jest różna od zera z uwagi na niedokładnie znalezione rozwiązania w punktach czasowych $P(5)$ oraz $P(7)$. Dla modelu $Y(y,x)$ wartość resztki jest bliska zeru. W Tab. 3.5 kolorem zielonym oznaczono dokładniejsze rozwiązania dla poszczególnych punktów czasowych.

Następnie, wykorzystując uzyskane wyniki stosujemy wzory (3.15) oraz (3.16) wstawiając w miejsce δX i δY wartości przyrostów cząstkowych modelu rozszerzonego oraz wartości funkcji wrażliwości dla modelu $X(x,y)$ i $Y(y,x)$ i rozwiązujemy ponownie układ równań nieliniowych. Po jego rozwiązaniu znajdujemy wartości przyrostów δx oraz δy (Tab. 3.6).

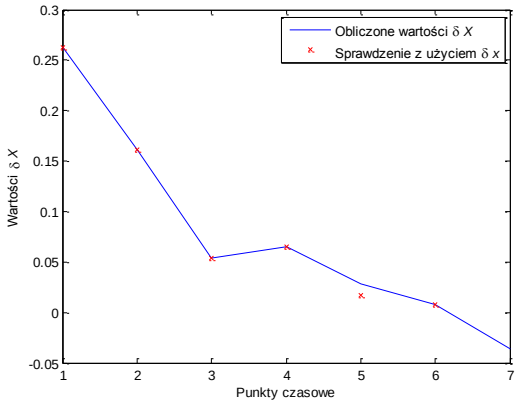
Obliczenia okazały się dokładne z wyjątkiem punktu $P(5)$, przystępujemy więc do sprawdzenia wzoru (3.15), który wskazuje, że przyrost δx oznacza przyrost cząstkowy na próbce N, gdy δX oznacza cząstkowy przyrost ddx na kolejnej próbce $(N+1)$. Wyniki porównania przedstawione są na Rys. 3.10 oraz Rys. 3.11.

Tab. 3.6 Wartości przyrostów δx oraz δy dla modeli $X(x,y)$ i $Y(y,x)$

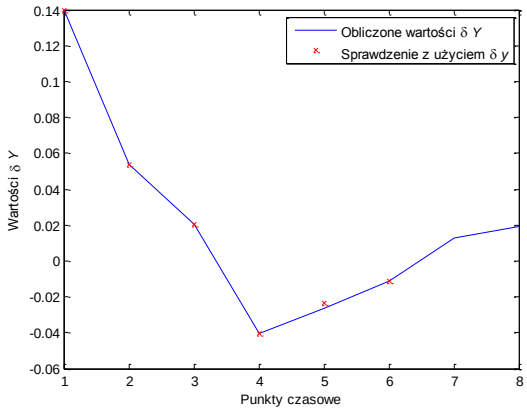
Punkty czasowe	δx	δy
P(1)	0.332786	0.391987
P(2)	0.235888	0.13796
P(3)	0.170313	0.145651
P(4)	0.042161	-0.04029
P(5)	0.062366	-0.09372
P(6)	-0.01087	0.016019
P(7)	0.042183	-0.03871

Na Rys. 3.8 oraz Rys. 3.9 przedstawiono wyniki sprawdzenia rezultatów obliczeń układu

równań.

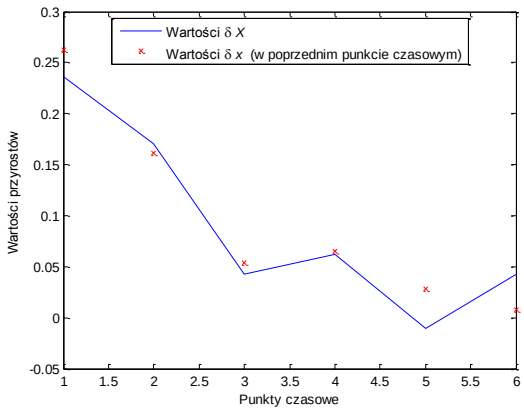


Rys. 3.8 Sprawdzenie dokładności obliczeń nieliniowego układu równań ze zmiennymi δx oraz δy
Źródło: Opracowanie własne (Matlab)

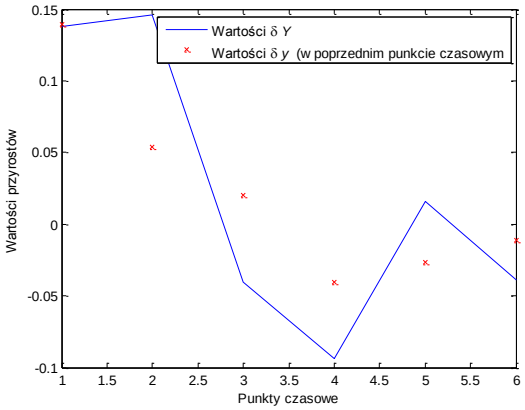


Rys. 3.9 Sprawdzenie dokładności obliczeń nieliniowego układu równań ze zmiennymi δx oraz δy
Źródło: Opracowanie własne (Matlab)

Dokonujemy prognozy wartości przyrostów cząstkowych na przyszłe chwile czasu, tym samym prognozując wartości szeregu.



Rys. 3.10 Porównanie wartości przyrostów δx w poprzednim punkcie czasowym z wartościami przyrostów δX w bieżącym punkcie
Źródło: Opracowanie własne (Matlab)



Rys. 3.11 Porównanie wartości przyrostów δy w poprzednim punkcie czasowym z wartościami przyrostów δY w bieżącym punkcie
Źródło: Opracowanie własne (Matlab)

Znając wartości przyrostów δX i δY, znamy też wartość δx i δy w kolejnym punkcie czasowym.

$\delta X_6 = 0.007783, \delta x_7 = 0.007783, \delta Y_6 = -0.01117, \delta y_7 = -0.01117.$

Znając wartości przyrostów oraz wartości różniczkowych i półwzględnych funkcji wrażliwości obliczamy wartości δX_7 oraz δY_7.

$$S'_{x_7,y_7} = \boldsymbol{f}'_x(N) = \begin{vmatrix} 3.819 & 5.168 \\ 17.571 & 15.4564 \end{vmatrix};$$

$$S''_{x_7,y_7} = \boldsymbol{f}'_y(N) = \begin{vmatrix} 8.885 & -8.0325 & -0.5365 \\ -23.828 & 60.63 & -2.844 \end{vmatrix}.$$

Otrzymujemy więc:

$\delta X_7 = 3.819 \cdot 0.007783 + 5.168 \cdot (-0.01117) + 8.885 \cdot (0.007783)^2 + (-8.0325) \cdot (-0.01117)^2 + (-0.5365) \cdot 0.007783 \cdot (-0.01117) = -0.0284.$

Rzeczywista wartość $\delta X_7^{real} = -0.036.$

$\delta Y_7 = 17.571 \cdot (-0.01117) + 15.4564 \cdot 0.007783 + (-23.828) \cdot (-0.01117)^2 + +60.63 \cdot (0.007783)^2 + (-2.844) \cdot (-0.01117) \cdot 0.007783 = -0.075.$

Rzeczywista wartość $\delta Y_7^{real} = 0.018219.$

$S_x^{X(x,y)} = 7.899$;$S_y^{X(x,y)} = 18.174$; $S_{x^2}^{X(x,y)} = 38.007$; $S_{y^2}^{X(x,y)} = -99.328$; $S_{xy}^{X(x,y)} = -3.9.$

$x_7 = 2.0683;$

$x_8 = 2.0683 + 7.899 \cdot (-0.0284) + 18.174 \cdot (-0.075) + 38.007 \cdot (-0.0284)^2 + +(-99.328) \cdot (-0.075)^2 + (-3.9) \cdot (-0.0284) \cdot (-0.075) = -0.05545.$

$x_8^{real} = 2.209.$

$S_y^{Y(y,x)} = 61.787$;$S_x^{Y(y,x)} = 31.968$; $S_{y^2}^{Y(y,x)} = -294.65$; $S_{x^2}^{Y(y,x)} = 259.37$;$S_{xy}^{Y(y,x)} = -20.684.$

$y_7 = 3.5165,$

$y_8 = 3.5165 + 61.787 \cdot (-0.075) + 31.968 \cdot (-0.0284) + (-294.65) \cdot (-0.075)^2 +$
$+259.37 \cdot (-0.0284)^2 + (-20.684) \cdot (-0.075) \cdot (-0.0284) = -3.52.$
$y_8^{real} = 3.24.$

Mimo, że wartości przyrostów δX pokrywają się z uzyskanymi przyrostami δx, to ze względu na wzajemny wpływ zmiennych *x* i *y,* uzyskane przez nas rezultaty predykcji wartości tych zmiennych nie są zadowalające. Przykład ten wskazuje wyraźnie na siłę związku poszczególnych zmiennych.

Nie oznacza to jednak, że idea prognozy w przyrostach jest błędna lecz, że działa dla szeregów wykazujących większe regularności. Na rezultaty naszych badań mogą wpływać też niedokładności w rozwiązywaniu układu równań nieliniowych.

Wykorzystamy teraz fakt, że wartości przyrostów możemy przedstawić za pomocą równań stanów postaci (3.17). Wtedy możliwe jest wykorzystanie dostępnych dla tej postaci równań filtrów prognozujących przyszłe wartości stanów tak przedstawionego procesu (przyrostów), odrzucających szumy procesu (np. filtr Kalmana).

3.4 Prognozowanie wartości przyrostów z użyciem filtru Kalmana

Filtr Kalmana jest ciekawym narzędziem, które w porównaniu do „zwykłych filtrów", przepuszczających przez siebie dane nic o nich nie wiedząc, oparty jest na modelu danego zjawiska. Aby możliwe było jego zastosowanie, nasz proces powinien dać się opisać za pomocą równań stanów postaci (3.17). To algorytm, który wykorzystuje szereg pomiarów dokonywanych w czasie, zawierających szum i niedokładności i tworzy oszacowania wartości zmiennych rządzących ewolucją procesu w czasie. Działa rekurencyjnie na strumieniach danych wejściowych w celu uzyskania statystycznie optymalnego oszacowania parametrów sterujących systemem.

Filtr ten można zastosować do prognozowania tychże wartości. Algorytm, zgodnie z którym działa filtr Kalmana (FK), dzielimy na dwie fazy (etapy). Pierwszy etap zwany jest predykcją a drugi korekcją. Podczas predykcji bazując jedynie na poprzedniej wartości stanu *x* wyznaczana jest nowa wartość stanu *x* oraz jej kowariancja. Wartości te wyznaczane są bez korzystania z informacji ze świata zewnętrznego, są one, więc niejako przewidywane na podstawie równań stanu, za pomocą których opisaliśmy nasz system.

Zasymulujmy, zatem działanie pewnego systemu wykorzystując FK na danych z Tab 3.1. Sprawdzamy przypadek, w którym δX oraz δY oznaczają przyrosty cząstkowe

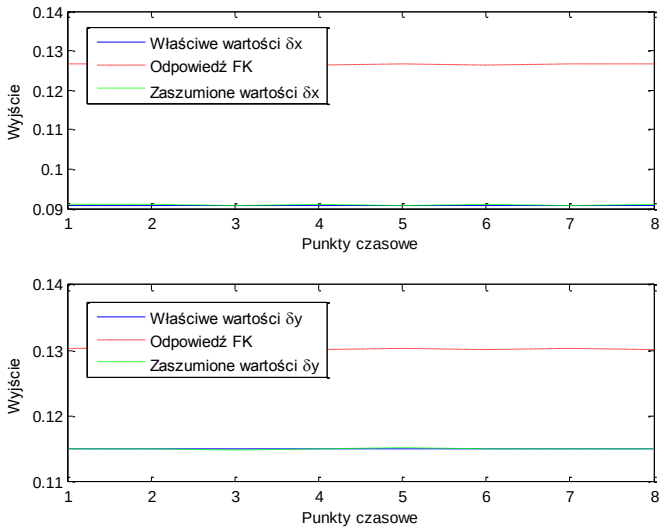
obliczane za pomocą układu równań nieliniowych ((2.30) i (2.31)). Sprawdzamy jak dokładną odpowiedź systemu na jeden krok wprzód możemy uzyskać (obliczając przyszłą wartość przyrostu). Czy można za pomocą FK znaleźć wartość przyrostu, który następnie wykorzystamy do znalezienia wartości zmiennej szeregu w przyszłym punkcie czasowym. Należy zauważyć, że mimo, iż pewne systemy jesteśmy w stanie przedstawić za pomocą równań stanów postaci (3.17), to nie zawsze da się je symulować za pomocą FK.

Dla danych z Tab. 3.1 rozszerzonych o wartości zmiennej q uzyskujemy następujące rezultaty.

$$A = \begin{vmatrix} a_1 + a_3 \cdot \delta x + a_5 \cdot \delta y & a_2 + a_4 \cdot \delta y + a_5 \cdot \delta x \\ b_1 + b_3 \cdot \delta y + b_5 \cdot \delta x & b_2 + b_4 \cdot \delta x + b_5 \cdot \delta y \end{vmatrix};\ B = \begin{vmatrix} a_3 & a_4 & a_5 \\ b_3 & b_4 & b_5 \end{vmatrix};\ C = \begin{vmatrix} 1 & 0 \\ 0 & 1 \end{vmatrix}$$

Gdzie $a_1, \dots, a_5,\ b_1, \dots, b_5,$ to współczynniki WKG używane do obliczenia wartości funkcji wrażliwości, δx i δy to wartości przyrostów w bieżącej chwili czasu.

Odpowiedzi systemu dla kilku punktów czasowych przedstawione są na Rys. 3.12. Na tym i następnych rysunkach przedstawiamy odpowiedzi FK wyznaczające wartości na jeden punkt czasowy wprzód na podstawie wartości z punktu poprzedniego.



Rys. 3.12 Odpowiedź FK na przyrosty cząstkowe w modelu $X(x, y)$ i $Y(y, x)$
Źrodło: Opracowanie własne (Matlab)

Stosując uzyskane rezultaty do wzoru (2.30) otrzymujemy prognozę dla punktu czasowego $P(7)$:

$$S_x^{X(x,y)} = 2.1436;\ S_y^{X(x,y)} \approx 0;\ S_{x^2}^{X(x,y)} \approx 0;\ S_{y^2}^{X(x,y)} \approx 0;\ S_{xy}^{X(x,y)} \approx 0.$$

$$\delta X = 0.1265;\ \delta Y = 0.13$$

$$x_7 = 2.1436 + 2.1436 \cdot 0.1265 + 0 \cdot (0.13) + 0 \cdot (0.1265)^2 + 0 \cdot (0.13)^2 +$$

$+0 \cdot (0.1265) \cdot 0.13 = 2.415.$

Rzeczywista wartość $x_7^{real} = 2.36.$

$S_y^{Y(y,x)} = 1.069; S_x^{Y(y,x)} = -0.00665; S_{y^2}^{Y(y,x)} = -0033; S_{x^2}^{Y(y,x)} = -0.01; S_{xy}^{Y(y,x)} = 0.0055.$

$y_7 = 1.063 + 1.069 \cdot 0.13 + (-0.00665) \cdot (0.1265) + (-0033) \cdot (0.13)^2 +$

$+(-0.01) \cdot (0.1265)^2 + 0.0055 \cdot (0.1265) \cdot 0.13 = 0.643.$

Rzeczywista wartość $y_7^{real} = 1.202.$

Dla modelu $Z(z, x)$ otrzymujemy następującą odpowiedź FK (Rys. 3.13):

$S_y^{Y(y,z)} = 1.063; S_z^{Y(y,z)} = -0.00019; S_{y^2}^{Y(y,z)} = -0.00033; S_{z^2}^{Y(y,z)} = 3 \cdot 10^{-5};$

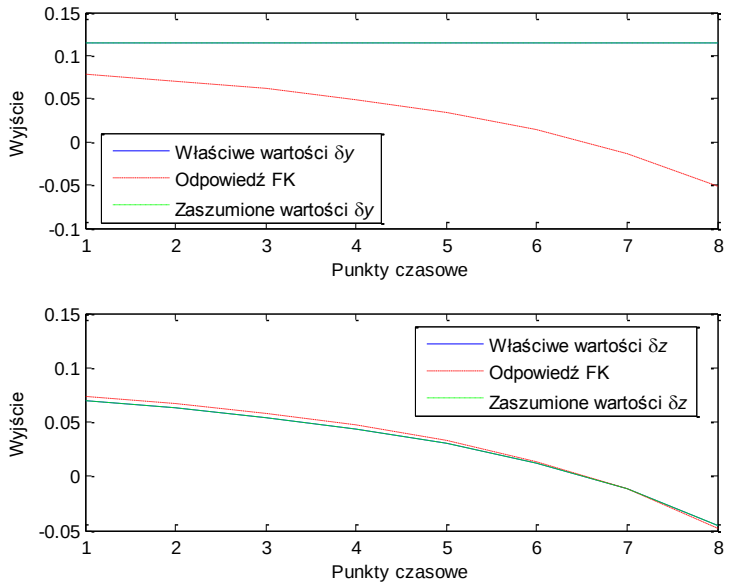
$S_{yz}^{Y(y,z)} = -4 \cdot 10^{-5}.$

$\delta Y = -0.013; \delta Z = -0.012.$

$y_7 = 1.063 + 1.063 \cdot (-0.013) + (-0.00019) \cdot (-0.012) + (-0.00033) \cdot (-0.013)^2 +$

$+(3 \cdot 10^{-5}) \cdot (-0.012)^2 + (-4 \cdot 10^{-5}) \cdot (-0.013) \cdot (-0.012) = 1.05.$

Rzeczywista wartość $y_7^{real} = 1.201.$



Rys. 3.13 Odpowiedź FK na przyrosty cząstkowe w modelu $Y(y, z)$ i $Z(z, y)$
Źródło: Opracowanie własne (Matlab)

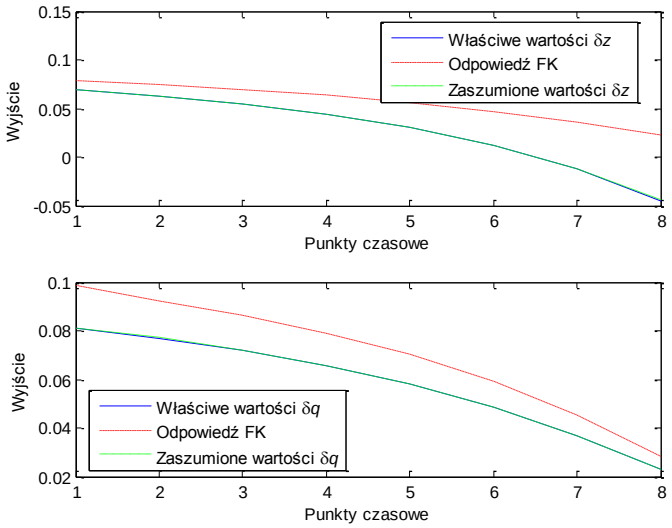
$S_z^{Z(z,y)} = 1.829; S_y^{Z(z,y)} = -0.382; S_{z^2}^{Z(z,y)} = 0.1668; S_{y^2}^{Z(z,y)} = -0.257; S_{yz}^{Z(z,y)} = -0.224.$

$z_7 = 1.714 + 1.829 \cdot (-0.012) + (-0.382) \cdot (-0.013) + (0.1668) \cdot (-0.012)^2 +$

$+(-0.257) \cdot (-0.013)^2 + (-0.224) \cdot (-0.012) \cdot (-0.013) = 1.697.$

Rzeczywista wartość $z_7^{real} = 1.63.$

Sytuację dla modelu $Z(z, q)$ $(Q(q, z))$ przedstawia Rys. 3.14.



Rys. 3.14 Odpowiedź FK na przyrosty cząstkowe w modelu $Z(z,q)$ i $Q(q,z)$
Źródło: Opracowanie własne (Matlab)

$S_z^{Z(z,q)} = 2.535;\ S_q^{Z(z,q)} = -1.089;\ S_{z^2}^{Z(z,q)} = 0.1176;\ S_{q^2}^{Z(z,q)} = -2.086;\ S_{zq}^{Z(z,q)} = 0.71.$

$\delta Z = 0.0352;\ \delta Q = 0.0445.$

$z_7 = 1.714 + 2.535 \cdot (0.0352) + (-1.089) \cdot (0.0445) + (0.1176) \cdot (0.0352)^2 +$

$+(-2.086) \cdot (0.0445)^2 + (0.71) \cdot (0.0352) \cdot (0.0445) = 1.752.$

$S_q^{Q(q,z)} = 1.94;\ S_z^{Q(q,z)} = 0.571;\ S_{q^2}^{Q(q,z)} = -2.088;\ S_{z^2}^{Q(q,z)} = 0.117;\ S_{qz}^{Q(q,z)} = 0.72.$

$q_7 = 2.778 + 1.94 \cdot (0.0445) + (0.571) \cdot (0.0352) + (-2.088) \cdot (0.0445)^2 +$

$+(0.117) \cdot (0.0352)^2 + (0.72) \cdot (0.0352) \cdot (0.0445) = 2.88.$

Rzeczywista wartość $q_7^{real} = 2.84.$

Obserwując odpowiedzi FK na wejścia w postaci przyrostów możemy ocenić, dla której zmiennej i w którym modelu uda nam się dokładniej przewidzieć przyszłą wartość zmiennej. Ze względu na wzajemny wpływ zmiennych na siebie, rezultaty prognoz nieco się od siebie różnią. Porównując wyniki prognozy zmiennej *z* w modelu ze zmienną *y* oraz *q* (Rys. 3.13, Rys. 3.14), zauważymy wyraźne różnice w dokładności odpowiedzi generowanych przez FK.

Powyższą symulację możemy przeprowadzić dla szeregów charakteryzujących się pewną regularnością. Otrzymujemy wartości przyrostów, które dają się symulować filtrem Kalmana, tym samym mamy możliwość wyznaczenia wartości takiego przyrostu w kolejnym punkcie czasowym, i również wartości samej zmiennej tego szeregu.

Dla szeregów modelujących pewne rzeczywiste systemy złożone możemy również uzyskać zadowalające rezultaty (Rozdział 4.).

4 Badania eksperymentalne

Celem przeprowadzenia badań eksperymentalnych było pokazanie skuteczności metody wykorzystującej funkcje wrażliwości, oraz wielomian Kołmogorowa-Gabora drugiego stopnia w prognozowaniu wartości danych przedstawionych w postaci wielowymiarowych szeregów czasowych różnego pochodzenia. W badaniach wykorzystano szeregi, których wartości poszczególnych zmiennych generowane były za pomocą wzorów matematycznych oraz ogólnodostępne zbiory danych w postaci szeregów czasowych o dużej ilości zmiennych, dostępne w zasobach internetowych.

Na podstawie wielu prób oceniamy, że przy znacznej nieregularności wartości poszczególnych zmiennych (gdy wartości na zmianę rosną i maleją w każdym kolejnym punkcie czasowym) uzyskanie właściwej wiarygodności prognozy (zarówno za pomocą MAMC, jak i wersji tej metody z równaniami stanów w przyrostach i FK), czy ostatecznie wskazanie korzystnego modelu, dającego najlepsze przybliżenie wartości prognozowanej zmiennej było niemożliwe. Każdy z modeli dawał mało satysfakcjonujące rezultaty. Nie oznacza to jednak, że w takich przypadkach nasza metoda jest całkowicie bezużyteczna. Może być wykorzystywana jako uzupełnienie innych metod predykcyjnych. Wskazujemy wiele przykładów, dla których prognoza daje zadowalające rezultaty.

4.1 Sposób przeprowadzania badań eksperymentalnych

Wszelkie obliczenia wykonywane były w środowisku Matlab. Każdy z prezentowanych szeregów reprezentuje inną charakterystykę danych.

Dla każdej zmiennej, każdego z szeregów obliczamy prognozowaną jej wartość na jeden krok wprzód używając modeli uproszczonych $MAMC_{simp}$ oraz rozszerzonych $MAMC_{ext}$ (dla szeregu o 5 zmiennych mamy po 4 modele każdego rodzaju, czyli 8 różnych rezultatów). Następnie oceniamy wiarygodność prognozy dla każdej zmiennej na podstawie wartości średnicy uzyskanych w każdym modelu rezultatów (rozrzutu uzyskanych prognoz) oraz biorąc pod uwagę dwie wybrane na potrzeby eksperymentów miary odległości (odległość Euklidesa oraz Canberra) i najmniejszą od średniej wszystkich rezultatów poszczególnych modeli odległość, wskazujemy ex ante, model, który dla obranych kryteriów wykazuje oczekiwane zachowanie.

Rezultaty wpasowania w poszczególne kryteria nie zawsze się pokrywają, wybieramy wiec model, który pasuje do największej ilości obranych przez nas kryteriów,

bądź uznajemy, że prognoza najprawdopodobniej będzie mało wiarygodna i nie uznajemy żadnego z uzyskanych wyników za właściwy w kolejnym kroku czasowym. Przesuwamy w takim przypadku zakres rozpatrywanych danych o jeden krok czasowy wprzód, wykorzystując kolejny zestaw danych i dla nich ponownie obliczamy współczynniki WKG drugiego stopnia, a następnie tworzymy wspomniane modele dla nowych wartości zmiennych i próbujemy ponownie prognozować wartości.

W przypadku uznania uzyskanych wyników za wystarczająco wiarygodne, dokonujemy prognozy na kolejny krok czasowy wprzód, korzystając z wybranych według kryteriów modeli oraz obliczając wartości przyrostów w kolejnym kroku.
Po każdym kroku dokonujemy oceny jakości wykonanej za pomocą wskazanego modelu prognozy (ex post) oraz wskazujemy dla porównania model, który w rzeczywistości był najlepszy.

Dla zilustrowania rozwiązania problemu predykcji szeregów za pomocą analizy równań w przyrostach wybrano zarówno szeregi czasowe generowane za pomocą wzorów matematycznych jak i takie, które zawierają pewne regularności i modelują przebieg pewnych rzeczywistych procesów złożonych.

4.2 Wyniki badań eksperymentalnych dla danych generowanych za pomocą wzorów matematycznych

Do badań wybieramy szereg czasowy o pięciu zmiennych, których wartości generowane są za pomocą wzorów matematycznych. Dla zmiennej *x*, począwszy od wartości równej 1.331, każda następna powstaje przez pomnożenie poprzedniej przez wartość 1.1. Dla zmiennej *y*, począwszy od wartości 0.9125, każda następna powstaje przez pomnożenie poprzedniej przez wartość 1.15. Zmienną *z* generujemy wzorem $z = x \cdot y \cdot e^{(-x^2-y^2)}$, zmienną *q* wzorem $q = x \cdot y \cdot z$, zmienną *r* wzorem $r = x^2 - z^2$. Dane są więc specyficzne, nieokresowe, a w każdym razie ilość branych pod uwagę próbek uczących nie jest w stanie objąć pełnego okresu danych. Dokonamy prognozy na kolejny punkt czasowy każdej zmiennej całego szeregu.

Wyznaczamy, więc wartości prognozowane zmiennej *x* w punkcie czasowym $P(7)$ w każdym z możliwych modeli. Następnie obliczamy odległości poszczególnych rezultatów od siebie (Tab. 4.1).

Tab. 4.1 Odległości wyników prognozy poszczególnych modeli od siebie dla zmiennej *x*

Modele	$X(x,y)s$	$X(x,z)s$	$X(x,q)s$	$X(x,r)s$	$X(x,y)$	$X(x,z)$	$X(x,q)$	$X(x,r)$
$X(x,y)s$	0	1.91E-11	1.91E-11	1.91E-11	1.57E-12	1.91E-11	1.92E-11	1.91E-11
$X(x,z)s$	1.91E-11	0	2.89E-14	1.82E-14	2.07E-11	1.15E-14	8.7E-14	3.86E-14
$X(x,q)s$	1.91E-11	2.89E-14	0	1.07E-14	2.07E-11	1.73E-14	1.16E-13	9.77E-15
$X(x,r)s$	1.91E-11	1.82E-14	1.07E-14	0	2.07E-11	6.66E-15	1.05E-13	2.04E-14
$X(x,y)$	1.57E-12	2.07E-11	2.07E-11	2.07E-11	0	2.07E-11	2.08E-11	2.07E-11
$X(x,z)$	1.91E-11	1.15E-14	1.73E-14	6.66E-15	2.07E-11	0	9.86E-14	2.71E-14
$X(x,q)$	1.92E-11	8.7E-14	1.16E-13	1.05E-13	2.08E-11	9.86E-14	0	1.26E-13
$X(x,r)$	1.91E-11	3.86E-14	9.77E-15	2.04E-14	2.07E-11	2.71E-14	1.26E-13	0

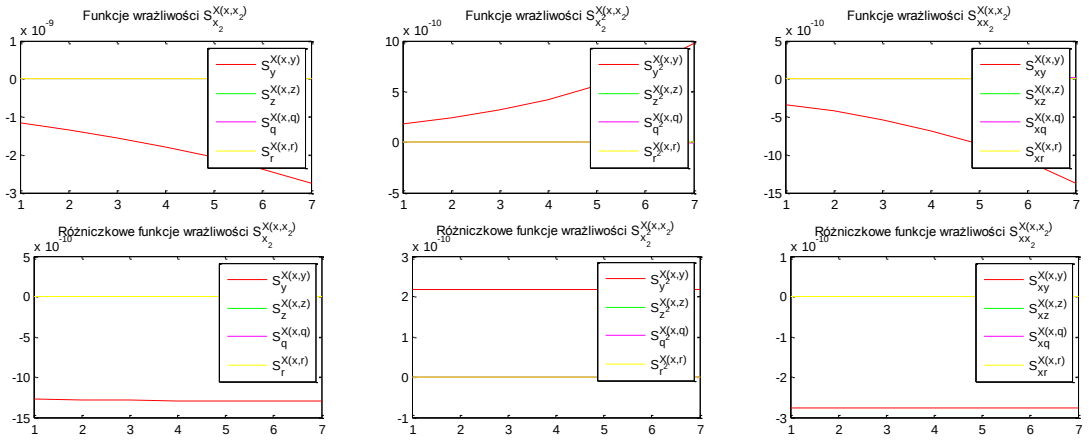
Z powyższej tabeli wynika, że wszystkie rezultaty są bardzo do siebie zbliżone (odległości są bliskie zeru). Tym samym odległości w metryce Euklidesa i Canberra dają podobne rezultaty, wskazują na nieznaczne różnice między poszczególnymi wynikami.

Tab. 4.2 Wartości odległości Euklidesa oraz Canberra prognoz uzyskiwanych przez poszczególne modele ze zmienną *x* oraz odległość od wartości średniej tych prognoz

Modele	$X(x,y)s$	$X(x,z)s$	$X(x,q)s$	$X(x,r)s$	$X(x,y)$	$X(x,z)$	$X(x,q)$	$X(x,r)$
Odległość Euklidesa	4.69E-11	2.82E-11	2.82E-11	2.82E-11	5.08E-11	2.82E-11	2.83E-11	2.82E-11
Odległość Canberra	1	1	1	1	1	1	1	1

Średnia	Rzeczywista wartość *x*	Średnica δ_x	Najmniejsza odległość od średniej	Najlepszy rzeczywisty	Dokonany wybór	Błąd względny prognozy
2.594	2.594	≈ 0	$X(x,r), X(x,q)s$	wszystkie	$X(x,r)$	0%

Średnica $\delta_x \approx 0$, co wskazuje na dużą wiarygodność uzyskanej prognozy i praktycznie daje dowolność w wyborze modelu, który przejdzie jako najdokładniejszy do kolejnego kroku. Ze względu jednak na ocenę modelu na podstawie najmniejszej odległości od średniej uzyskanej w każdym modelu wartości wybieramy model rozszerzony ze zmienną $r - X(x,r)$.



Rys. 4.1 Wykresy wartości półwzględnej oraz różniczkowej funkcji wrażliwości dla modeli ze zmienną *x*
Źródło: Opracowanie własne (Matlab)

Wartości funkcji wrażliwości zarówno różniczkowe, jak i półwzględne (Rys. 4.1) wskazują dużą regularność szeregu oraz na to, że zmiany wartości zmiennej

y w największym stopniu wpływają na zmianę wartości zmiennej *x*. Wyznaczamy wartości prognozowanej wartości zmiennej w punkcie czasowym $P(7)$ w każdym z rozpatrywanych modeli dla zmiennej *y* oraz ich odległości od siebie (Tab. 4.3).

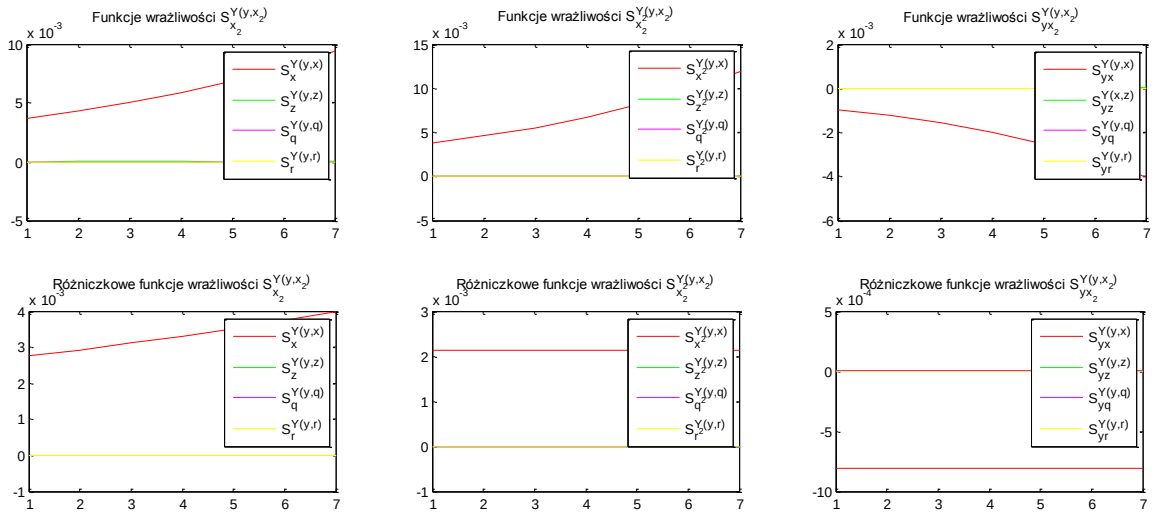
Tab. 4.3 Odległości wyników prognozy poszczególnych modeli od siebie dla zmiennej *y*

Modele	$Y(y,x)s$	$Y(y,z)s$	$Y(y,q)s$	$Y(y,r)s$	$Y(y,x)$	$Y(y,z)$	$Y(y,q)$	$Y(y,r)$
$Y(y,x)s$	0	3.51E-07	5.71E-07	3.35E-07	3.23E-07	3.24E-07	8.2E-07	3.95E-07
$Y(y,z)s$	3.51E-07	0	2.2E-07	1.61E-08	6.75E-07	2.76E-08	4.68E-07	4.35E-08
$Y(y,q)s$	5.71E-07	2.2E-07	0	2.36E-07	8.94E-07	2.47E-07	2.49E-07	1.76E-07
$Y(y,r)s$	3.35E-07	1.61E-08	2.36E-07	0	6.59E-07	1.15E-08	4.85E-07	5.96E-08
$Y(y,x)$	3.23E-07	6.75E-07	8.94E-07	6.59E-07	0	6.47E-07	1.14E-06	7.18E-07
$Y(y,z)$	3.24E-07	2.76E-08	2.47E-07	1.15E-08	6.47E-07	0	4.96E-07	7.11E-08
$Y(y,q)$	8.2E-07	4.68E-07	2.49E-07	4.85E-07	1.14E-06	4.96E-07	0	4.25E-07
$Y(y,r)$	3.95E-07	4.35E-08	1.76E-07	5.96E-08	7.18E-07	7.11E-08	4.25E-07	0

Tab. 4.4 Wartości odległości Euklidesa oraz Canberra prognoz uzyskiwanych przez poszczególne modele ze zmienną *y* oraz odległość od wartości średniej tych prognoz

Modele	$Y(y,x)s$	$Y(y,z)s$	$Y(y,q)s$	$Y(y,r)s$	$Y(y,x)$	$Y(y,z)$	$Y(y,q)$	$Y(y,r)$
Odległość Euklidesa	1.26E-06	9.21E-07	1.18E-06	9.17E-07	2.01E-06	9.15E-07	1.71E-06	9.45E-07
Odległość Canberra	1	1	1	1	1	1	1	1

Średnia	**Rzeczywista wartość *y***	**Średnica δ_y**	**Najmniejsza odległość od średniej**	**Najlepszy rzeczywisty model**	**Dokonany wybór**	**Błąd względny prognozy**
2.427	2.427	≈ 0	$Y(y,z), Y(y,r)s$	$Y(y,r)s, Y(y,z)s$	$Y(y,r)s$	0%



Rys. 4.2 Wykresy wartości półwzględnej oraz różniczkowej funkcji wrażliwości dla modeli ze zmienną *y*
Źródło: Opracowanie własne (Matlab)

Podobnie jak w przypadku zmiennej *x*, wartości funkcji wrażliwości zarówno różniczkowe, jak i półwzględne wskazują na dużą regularność szeregu oraz na to, że zmiany wartości zmiennej *x* w największym stopniu wpływają na zmianę wartości badanej zmiennej *y*.

Wyznaczamy teraz wartości prognozowanej wartości zmiennej w punkcie

czasowym $P(7)$ w każdym z modeli dla zmiennej z oraz ich odległości od siebie.

Tab. 4.5 Odległości wyników prognozy poszczególnych modeli od siebie dla zmiennej z

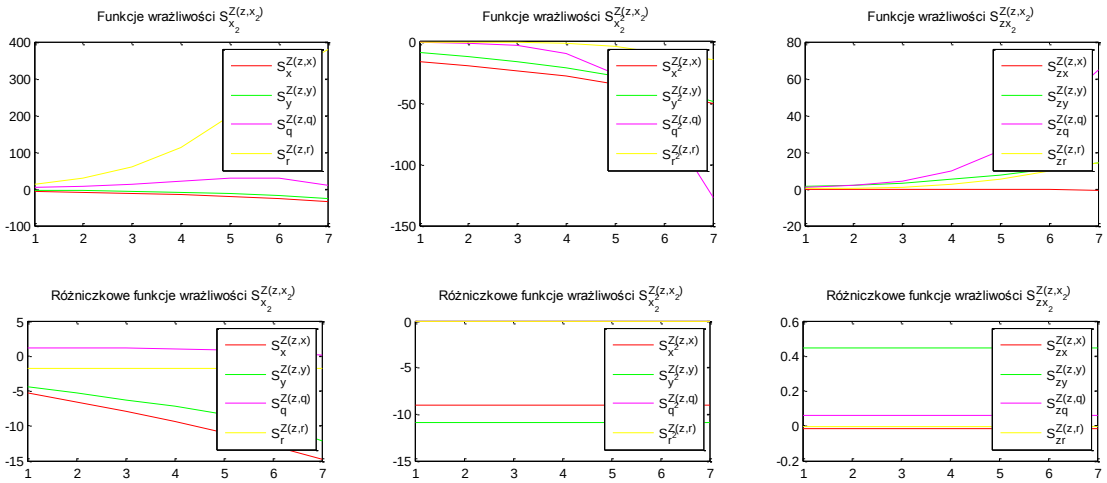
Modele	$Z(z,x)s$	$Z(z,y)s$	$Z(z,q)s$	$Z(z,r)s$	$Z(z,x)$	$Z(z,y)$	$Z(z,q)$	$Z(z,r)$
$Z(z,x)s$	0	0.11029	5.10724	2.3188	0.0786	0.1789	6.0694	1.8381
$Z(z,y)s$	0.11029	0	4.9969	2.2085	0.03166	0.0686	5.9591	1.72784
$Z(z,q)s$	5.1072	4.9969	0	2.7884	5.0286	4.9283	0.9622	3.269
$Z(z,r)s$	2.3188	2.2085	2.7884	0	2.2401	2.1399	3.7506	0.4807
$Z(z,x)$	0.0786	0.0317	5.0286	2.2402	0	0.1003	5.9908	1.7595
$Z(z,y)$	0.1789	0.0686	4.9283	2.1399	0.1003	0	5.8905	1.6592
$Z(z,q)$	6.0694	5.9591	0.9622	3.7506	5.9908	5.8905	0	4.2313
$Z(z,r)$	1.8381	1.7278	3.2691	0.48067	1.7595	1.659	4.2313	0

Tab. 4.6 Wartości odległości Euklidesa oraz Canberra prognoz uzyskiwanych przez poszczególne modele ze zmienną z oraz odległość od wartości średniej tych prognoz

Modele	$Z(z,x)s$	$Z(z,y)s$	$Z(z,q)s$	$Z(z,r)s$	$Z(z,x)$	$Z(z,y)$	$Z(z,q)$	$Z(z,r)$
Odległość Euklidesa	8.469211	8.268094	10.95522	6.474976	8.325129	8.146484	13.26019	6.405841
Odległość Canberra	0.861422	0.865797	0.748259	0.848796	0.871615	0.872994	0.688602	0.859864

Średnia	Rzeczywista wartość z	Średnica δ_z	Najmniejsza odległość od średniej	Najlepszy rzeczywisty model	Dokonany wybór	Błąd względny prognozy
14.224	14.52	6.0694	$Z(z,r)$	$Z(z,r), Z(z,r)s$	$Z(z,r)$	1%

Kolorem zielonym oznaczono najkorzystniejsze wartości poszczególnych modeli, czerwonym najmniej korzystne.



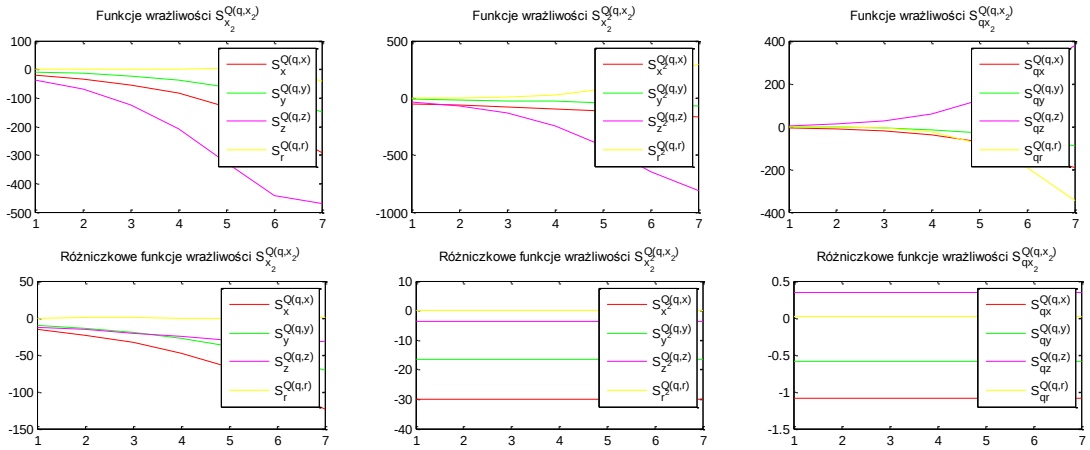
Rys. 4.3 Wykresy wartości półwzględnej oraz różniczkowej funkcji wrażliwości dla modeli ze zmienną z
Źródło: Opracowanie własne (Matlab)

Wyznaczamy wartości prognozowanej wartości zmiennej w punkcie czasowym $P(7)$ w każdym z możliwych modeli dla zmiennej q oraz odległości między nimi.

Tab. 4.7 Wartości odległości Euklidesa oraz Canberra prognoz uzyskiwanych przez poszczególne modele ze zmienną q oraz odległość od wartości średniej tych prognoz

Modele	$Q(q,x)s$	$Q(q,y)s$	$Q(q,z)s$	$Q(q,r)s$	$Q(q,x)$	$Q(q,y)$	$Q(q,z)$	$Q(q,r)$
Odległość Euklidesa	48.22342	45.82082	63.20014	41.54749	46.12079	42.50967	79.09915	37.20339
Odległość Canberra	0.873137	0.882872	0.760149	0.88631	0.887589	0.891496	0.696561	0.858508

Średnia	Rzeczywista wartość q	Średnica δ_q	Najmniejsza odległość od średniej	Najlepszy rzeczywisty model	Dokonany wybór	Błąd względny prognozy
88.09	91.41	35.75	$Q(q,r), Q(q,r)s$	$Q(q,r)s, Q(q,r)$	$Q(q,r)s$	3.8%



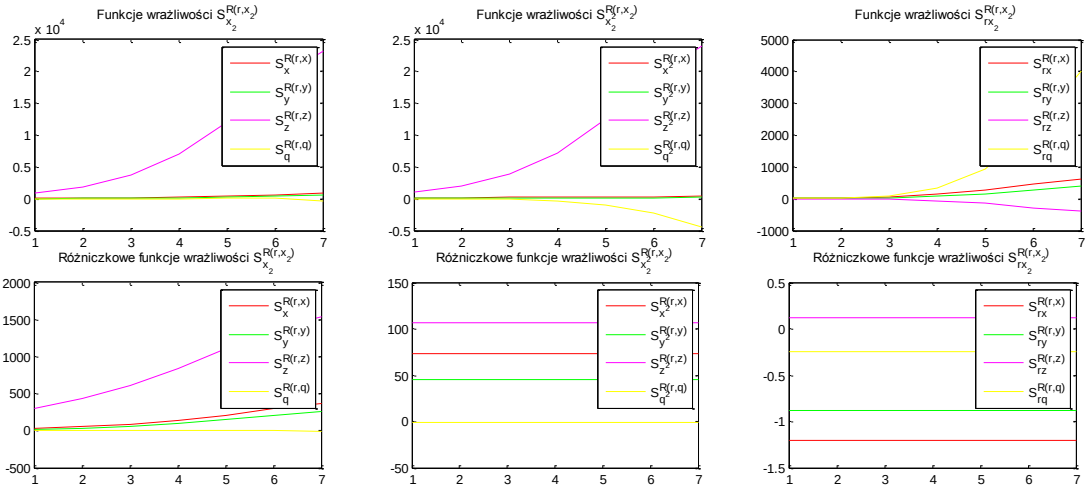
Rys. 4.4 Wykresy wartości półwzględnej oraz różniczkowej funkcji wrażliwości dla modeli ze zmienną q
Źródło: Opracowanie własne (Matlab)

Wyznaczamy wartości prognozowanej wartości zmiennej w punkcie czasowym $P(7)$ w każdym z możliwych modeli dla zmiennej r oraz odległości między nimi.

Tab. 4.8 Wartości odległości Euklidesa oraz Canberra prognoz uzyskiwanych przez poszczególne modele ze zmienną r oraz odległość od wartości średniej tych prognoz

Modele	$R(r,x)s$	$R(r,y)s$	$R(r,z)s$	$R(r,q)s$	$R(r,x)$	$R(r,y)$	$R(r,z)$	$R(r,q)$
Odległość Euklidesa	166.5669	160.3486	124.4724	126.6405	134.8792	126.4459	205.8927	222.5349
Odległość Canberra	0.794886	0.803504	0.779394	0.834491	0.838446	0.841908	0.664135	0.635873

Średnia	Rzeczywista wartość r	Średnica δ_r	Najmniejsza odległość od średniej	Najlepszy rzeczywisty model	Dokonany wybór	Błąd względny prognozy
-194.99	-204.067	110	$R(r,z)s, R(r,y)$	$R(r,q)s, R(r,y)$	$R(r,y)$	4.7%



Rys. 4.5 Wykresy wartości półwzględnej oraz różniczkowej funkcji wrażliwości dla modeli ze zmienną r
Źródło: Opracowanie własne (Matlab)

Ze względu na niewielką ilość danych inicjujących, podczas analizy danych nie korzystamy z innych niż średnia arytmetyczna narzędzi statystycznych.

Teraz dokonujemy ponownie obliczeń wykorzystując te same współczynniki WKG oraz wyznaczone w poprzednim kroku prognozowane wartości w celu wyznaczenia prognozowanych wartości poszczególnych zmiennych w punkcie czasowym $P(8)$.

Dla modeli ze zmienną *x* otrzymujemy:

Tab. 4.9 Wartości odległości Euklidesa oraz Canberra prognoz uzyskiwanych przez poszczególne modele ze zmienną *x* oraz odległość od wartości średniej tych prognoz dla punktu czasowego $P(8)$

Modele	$X(x,y)s$	$X(x,z)s$	$X(x,q)s$	$X(x,r)s$	$X(x,y)$	$X(x,z)$	$X(x,q)$	$X(x,r)$
Odległość Euklidesa	1.22E-10	7.52E-11	7.48E-11	7.5E-11	1.38E-10	7.51E-11	7.52E-11	7.5E-11
Odległość Canberra	1	1	1	1	1	1	1	1

Średnia	Rzeczywista wartość *x*	Średnica δ_x	Najmniejsza odległość od średniej	Najlepszy rzeczywisty	Dokonany wybór	Błąd względny prognozy
2.853	2.853	≈ 0	$X(x,q)s, X(x,r)s$	$X(x,r)s$	$X(x,q)s$	0%

Dla modeli ze zmienną *y* otrzymujemy:

Tab. 4.10 Wartości odległości Euklidesa oraz Canberra prognoz uzyskiwanych przez poszczególne modele ze zmienną *y* oraz odległość od wartości średniej tych prognoz dla punktu czasowego $P(8)$

Modele	$Y(y,x)s$	$Y(y,z)s$	$Y(y,q)s$	$Y(y,r)s$	$Y(y,x)$	$Y(y,z)$	$Y(y,q)$	$Y(y,r)$
Odległość Euklidesa	3.18E-06	2.65E-06	2.84E-06	2.35E-06	5.48E-06	2.54E-06	3.93E-06	2.35E-06
Odległość Canberra	1	1	1	1	1	1	1	1

Średnia	Rzeczywista wartość *y*	Średnica δ_y	Najmniejsza odległość od średniej	Najlepszy rzeczywisty model	Dokonany wybór	Błąd względny prognozy
2.791	2.791	≈ 0	$Y(y,r), Y(y,r)s$	$Y(y,z), Y(y,r)s$	$Y(y,r)$	0%

Dla modeli ze zmienną *z* otrzymujemy:

Tab. 4.11 Wartości odległości Euklidesa oraz Canberra prognoz uzyskiwanych przez poszczególne modele ze zmienną *z* oraz odległość od wartości średniej tych prognoz dla punktu czasowego $P(8)$

Modele	$Z(z,x)s$	$Z(z,y)s$	$Z(z,q)s$	$Z(z,r)s$	$Z(z,x)$	$Z(z,y)$	$Z(z,q)$	$Z(z,r)$
Odległość Euklidesa	46.11706	46.30504	65.41768	85.83869	46.16676	46.35189	81.02439	81.90038
Odległość Canberra	0.587212	0.588769	0.021405	0.3556	0.577705	0.579141	0.496556	0.368559

Średnia	Rzeczywista wartość *z*	Średnica δ_z	Najmniejsza odległość od średniej	Najlepszy rzeczywisty model	Dokonany wybór	Błąd względny prognozy
16.85	11.281	49.327	$Z(z,x)s, Z(z,x)$	$Z(z,y), Z(z,y)s$	$Z(z,x)s$	30%

Dla modeli ze zmienną *q* otrzymujemy:

Tab. 4.12 Wartości odległości Euklidesa oraz Canberra prognoz uzyskiwanych przez poszczególne modele ze zmienną *q* oraz odległość od wartości średniej tych prognoz dla punktu czasowego $P(8)$

Modele	$Q(q,x)s$	$Q(q,y)s$	$Q(q,z)s$	$Q(q,r)s$	$Q(q,x)$	$Q(q,y)$	$Q(q,z)$	$Q(q,r)$
Odległość Euklidesa	146.9358	142.2332	196.4447	166.8861	141.9137	135.0466	295.0846	134.4177
Odległość Canberra	0.719157	0.728865	0.290728	0.664236	0.760081	0.740594	0.192706	0.732735

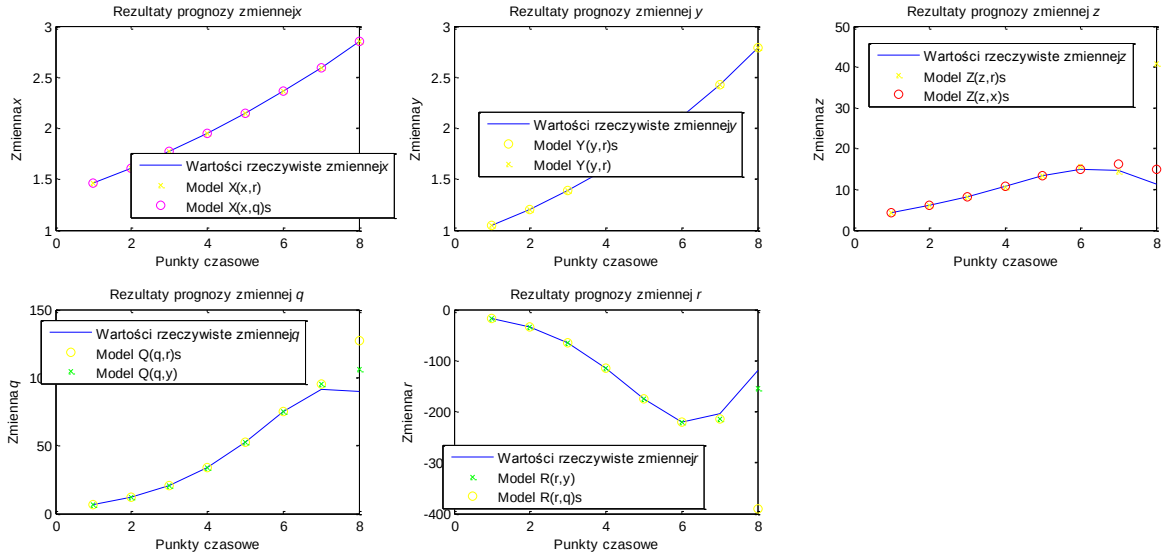
Średnia	Rzeczywista wartość q	Średnica δ_q	Najmniejsza odległość od średniej	Najlepszy rzeczywisty model	Dokonany wybór	Błąd względny prognozy
88.454	89.84643	133.346	$Q(q,r), Q(q,y)$	$Q(q,r)s, Q(q,y)$	$Q(q,y)$	18%

Dla modeli ze zmienną *r* otrzymujemy:

Tab. 4.13 Wartości odległości Euklidesa oraz Canberra prognoz uzyskiwanych przez poszczególne modele ze zmienną *r* oraz odległość od wartości średniej tych prognoz dla punktu czasowego $P(8)$

Modele	$R(r,x)s$	$R(r,y)s$	$R(r,z)s$	$R(r,q)s$	$R(r,x)$	$R(r,y)$	$R(r,z)$	$R(r,q)$
Odległość Euklidesa	1149.785	1159.077	2047.112	976.2505	1198.137	1216.65	1785.85	1132.328
Odległość Canberra	0.633494	0.632798	0.305614	0.445359	0.63324	0.612364	0.352243	0.599525

Średnia	Rzeczywista wartość r	Średnica δ_r	Najmniejsza odległość od średniej	Najlepszy rzeczywisty model	Dokonany wybór	Błąd względny prognozy
-411.07	-119.124	893.987	$R(r,q)s, R(r,q)$	$R(r,y), R(r,x)$	$R(r,q)s$	228%

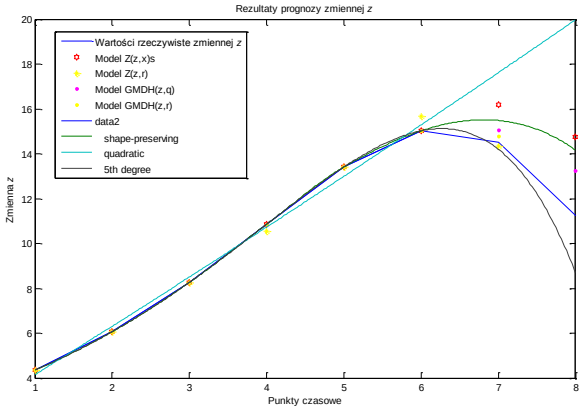


Rys. 4.6 Porównanie rezultatów prognozy najlepszych modeli każdej zmiennej
Źródło: Opracowanie własne (Matlab)

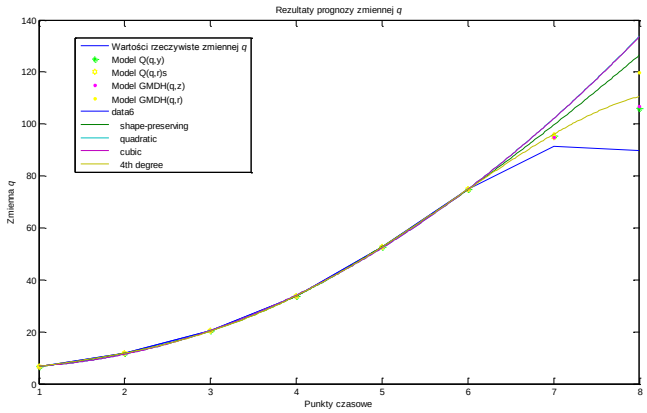
Jak można zauważyć, badanie odległości pozwala na wskazanie modelu, który na danym kroku w rzeczywistości okazuje się najbliższy wartości rzeczywistej. W trzecim kroku wykonywanym według ustalonej procedury, zmienne *x* oraz *y* nadal przewidywane są z dużą dokładnością, wartości pozostałych zmiennych wskazywane są z dokładnością mniej zadowalająca.

Porównanie wyników prognozy na jeden, dwa kroki wprzód z metodami interpolacyjnymi oraz uproszczoną wersją metody GMDH (wyznaczamy współczynniki WKG z układu sześciu równań z sześcioma niewiadomymi i wyznaczamy wartości w kolejnych punktach czasowych korzystając ze wzoru (2.5)) wskazuje skuteczność metody MAMC. W metodzie GMDH brak sposobu oceny dokładności prognozy ex ante,

więc nie możemy wskazać, który z modeli GMDH jest najlepszy, podobnie w przypadku metod interpolacyjnych, poznajemy ich skuteczność ex post). Rezultaty prognozy dla zmiennych *z*, *q* i *r* na rysunkach poniżej (Rys. 4.7, Rys. 4.8, Rys. 4.9).

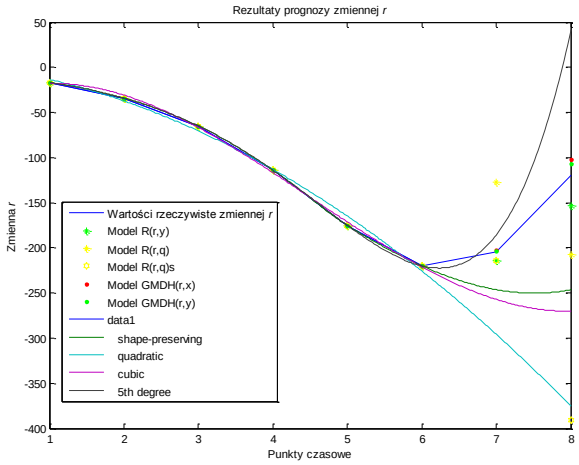


Rys. 4.7 Porównanie rezultatów prognozy najlepszych modeli z funkcjami wrażliwości, uproszczonej metody GMDH oraz kilku metod interpolacyjnych dla zmiennej *z*, na dwa kroki wprzód
Źródło: Opracowanie własne (Matlab)



Rys. 4.8 Porównanie rezultatów prognozy najlepszych modeli z funkcjami wrażliwości, uproszczonej metody GMDH oraz kilku metod interpolacyjnych dla zmiennej *q*, na dwa kroki wprzód
Źródło: Opracowanie własne (Matlab)

Na dokładność obliczeń, jak wspomniano w Rozdziale 2., wpływ ma dokładność rozwiązań nieliniowych układów równań, za pomocą których wyznaczamy wartości przyrostów, a następnie ich ekstrapolacje. Nawet w przypadku szeregów charakteryzujących się znaczną regularnością, wartości przyrostów są bardzo zróżnicowane. Poza tym, każda ze zmiennych w pewien sposób oddziałuje na pozostałe, co znajduje odzwierciedlenie w wartościach funkcji wrażliwości i również wpływa na rezultaty obliczeń. Jednakże, to cel determinuje miarę dokładności, jaką jesteśmy w stanie zaakceptować w przypadku naszej prognozy. Naszym celem jest uzyskanie zadowalających efektów dla jak największej ilości zmiennych szeregu czasowego jednocześnie, przy niewielkiej liczbie danych inicjujących.



Rys. 4.9 Porównanie rezultatów prognozy najlepszych modeli z funkcjami wrażliwości, uproszczonej metody GMDH oraz kilku metod interpolacyjnych dla zmiennej *r*, na dwa kroki wprzód
Źródło: Opracowanie własne (Matlab)

Po wykonaniu prognozy na jeden krok wprzód, uwzględniając najlepsze z otrzymanych rezultatów (modeli), dokonujemy przeliczenia współczynników WKG. Rezultaty dla punktów czasowych $P(8)$, $P(9)$ są następujące: (Tab. 4.14, Tab. 4.15, Tab. 4.16). Dla modeli ze zmienną *x* oraz *y* otrzymujemy błąd względny prognozy przyjmujący wartości bliskie 0%.

Dla modeli ze zmienną *z* otrzymujemy:

Tab. 4.14 Wartości odległości Euklidesa oraz Canberra prognoz uzyskiwanych przez poszczególne modele ze zmienną *z* oraz odległość od wartości średniej tych prognoz dla punktu czasowego $P(9)$

Modele	$Z(z,x)s$	$Z(z,y)s$	$Z(z,q)s$	$Z(z,r)s$	$Z(z,x)$	$Z(z,y)$	$Z(z,q)$	$Z(z,r)$
Odległość Euklidesa	22.36317	22.36839	51.48714	48.42312	22.36522	22.37513	22.35711	22.36998
Odległość Canberra	0.796925	0.797391	0.381638	0.173074	0.774732	0.772167	0.762154	0.769999

Średnia	Rzeczywista wartość z	Średnica δ_z	Najmniejsza odległość od średniej	Najlepszy rzeczywisty model	Dokonany wybór	Błąd względny prognozy
12.81	11.28116	31.584	$Z(z,q), Z(z,x)s$	$Z(z,y)$	$Z(z,x)s$	1%

Model $Z(z,q)$ wykazuje niezadowalające dopasowanie, do wartości rzeczywistych.

Dla modeli ze zmienną *q* otrzymujemy:

Tab. 4.15 Wartości odległości Euklidesa oraz Canberra prognoz uzyskiwanych przez poszczególne modele ze zmienną *q* oraz odległość od wartości średniej tych prognoz dla punktu czasowego $P(9)$

Modele	$Q(q,x)s$	$Q(q,y)s$	$Q(q,z)s$	$Q(q,r)s$	$Q(q,x)$	$Q(q,y)$	$Q(q,z)$	$Q(q,r)$
Odległość Euklidesa	203.2777	219.2539	508.9272	205.8581	200.4013	205.5962	185.9434	185.1831
Odległość Canberra	0.861491	0.804969	0.724318	0.857976	0.693629	0.695447	0.599409	0.641636

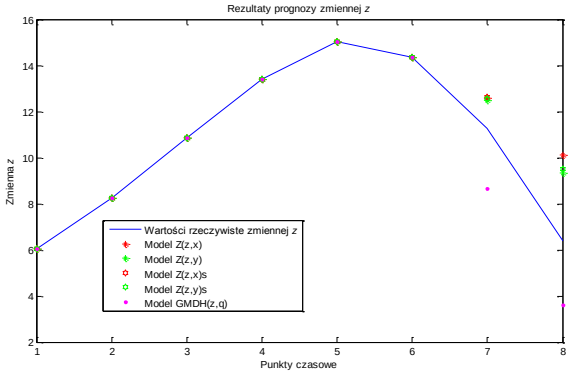
Średnia	Rzeczywista wartość q	Średnica δ_q	Najmniejsza odległość od średniej	Najlepszy rzeczywisty model	Dokonany wybór	Błąd względny prognozy
88.298	89.84643	209.872	$Q(q,r)$	$Q(q,r), Q(q,z)$	$Q(q,r)$	38%

Dla modeli ze zmienną *r* otrzymujemy:

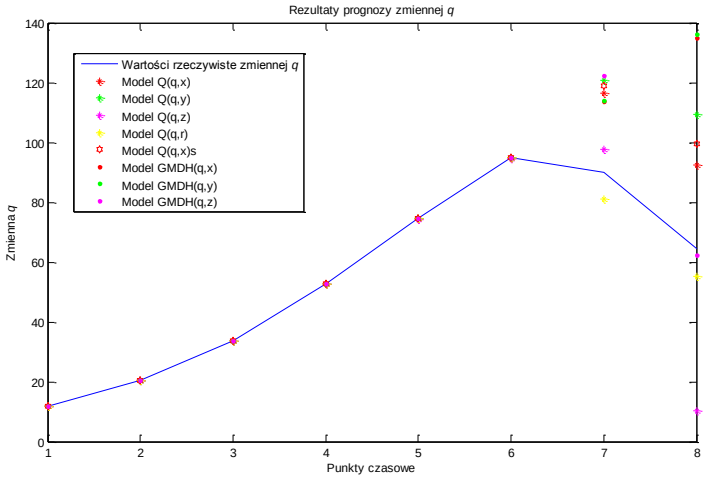
Tab. 4.16 Wartości odległości Euklidesa oraz Canberra prognoz uzyskiwanych przez poszczególne modele ze zmienną *r* oraz odległość od wartości średniej tych prognoz dla punktu czasowego $P(9)$

Modele	$R(r,x)s$	$R(r,y)s$	$R(r,z)s$	$R(r,q)s$	$R(r,x)$	$R(r,y)$	$R(r,z)$	$R(r,q)$
Odległość Euklidesa	249.7876	244.5763	566.0532	228.9206	258.7526	259.5066	242.4585	272.5565
Odległość Canberra	0.778512	0.780419	0.215812	0.748008	0.760283	0.758414	0.768482	0.448904

Średnia	Rzeczywista wartość *r*	Średnica δ_r	Najmniejsza odległość od średniej	Najlepszy rzeczywisty model	Dokonany wybór	Błąd względny prognozy
-149.2	-119.124	234.266	$R(r,q)s$	$R(r,q)s, R(r,q)$	$R(r,q)s$	11%

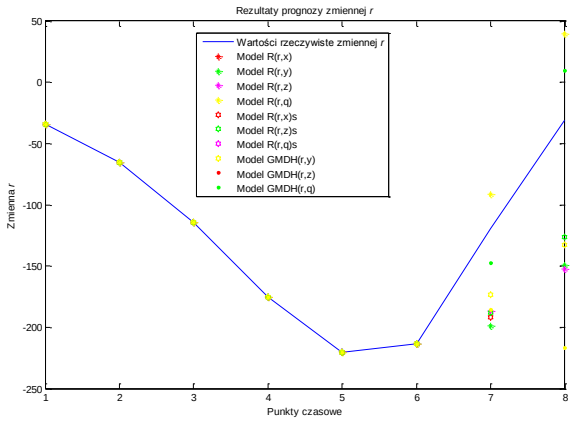


Rys. 4.10 Porównanie rezultatów prognozy najlepszych modeli z funkcjami wrażliwości, uproszczonej metody GMDH w punktach czasowych $P(8)$ i $P(9)$ dla zmiennej *z*
Źródło: Opracowanie własne (Matlab)



Rys. 4.11 Porównanie rezultatów prognozy najlepszych modeli z funkcjami wrażliwości, uproszczonej metody GMDH w punktach czasowych $P(8)$ i $P(9)$ dla zmiennej *q*
Źródło: Opracowanie własne (Matlab)

W wielu przypadkach średnica δ przyjmowała wysokie wartości, co wskazuje na niski poziom zaufania do wyników prognozy dla danej zmiennej, jednak na takie rezultaty zwykle wpływała jedna lub dwie znacznie odstające wartości uzyskiwane przez któryś z modeli. Więc nawet wysokie wartości δ nie musiały dyskwalifikować kroku prognozy dla określonej zmiennej.



Rys. 4.12 Porównanie rezultatów prognozy najlepszych modeli z funkcjami wrażliwości, uproszczonej metody GMDH w punktach czasowych $P(8)$ i $P(9)$ dla zmiennej r
Źródło: Opracowanie własne (Matlab)

Dla każdej zmiennej utworzone zostały dodatkowo 4 modele GMDH, dla których przyszłe wartości liczone były za pomocą WKG drugiego stopnia. Uzyskane w ten sposób rezultaty w pierwszym kroku (prognoza wartości punktu czasowego $P(7)$) dawały wyniki porównywalne z naszą metodą. W kroku kolejnym średnica modeli GMDH miała znacznie większą wartość niż średnica modeli MAMC. Na wykresach powyżej uwidocznione są rezultaty również najlepszych modeli GMDH.

Początkowy zestaw wartości $P(0), \ldots, P(6)$ dla zobrazowania funkcjonowania naszej metody analizy szeregów dobrano tak, aby zawierał minima bądź maksima lokalne (globalne). Tak, aby można było zauważyć, jak zachowają się modele przy zmianach monotoniczności wartości szeregu. Najlepsze z modeli wskazywały właściwy kierunek zmian wartości szeregu dla każdej jego zmiennej, wykazując większą dokładność od klasycznych metod ekstrapolacyjnych oraz uproszczonej metody GMDH.

4.3 Wyniki badań eksperymentalnych dla szeregu modelującego rzeczywisty proces fizyczny

Przykład wybrano z dostępnych w zasobach sieciowych zbiorów danych w postaci szeregu czasowego (Tab. 4.17). Przedstawia on zmiany w czasie rzeczywistych, wzajemnie powiązanych ze sobą czynników. Przykład obrazuje możliwość prognozowania szeregów czasowych, które nie są „sztucznie" uzyskiwane poprzez generowanie poszczególnych zestawów punktów czasowych $P(i)$ za pomocą wzorów matematycznych.

Tab. 4.17 Szereg czasowy o pięciu zmiennych modelujący zachowanie pewnego rzeczywistego procesu

Punkty czasowe	*x*	*y*	*z*	*q*	*r*
P(0)	13268	88942.94	18542	14661	11.648
P(1)	14343	93314.94	20438	16263	11.865
P(2)	15677	98455.94	21797	16952	12.074
P(3)	17076	103620.9	24658	19058.99	12.3
P(4)	18837	109770.9	27238	20817.99	12.553
P(5)	20837	114344.9	30027	23256.99	12.817
P(6)	23170	119030.9	33584	26067.99	13.067
P(7)	26001	125670.9	38306	30117	13.304
P(8)	30730	132515.9	46114	36300	13.505

Dla modeli ze zmienną *x* otrzymujemy:

Tab. 4.18 Wartości odległości Euklidesa oraz Canberra prognoz uzyskiwanych przez poszczególne modele ze zmienną *x* oraz odległość od wartości średniej tych prognoz dla punktu czasowego $P(7)$

Modele	$X(x,y)s$	$X(x,z)s$	$X(x,q)s$	$X(x,r)s$	$X(x,y)$	$X(x,z)$	$X(x,q)$	$X(x,r)$
Odległość Euklidesa	84549.69	55945.39	55714.96	56152.76	107576.3	54269.26	55706.34	55875.53
Odległość Canberra	0.358265	0.852463	0.853902	0.848525	0.560828	0.774416	0.804501	0.805384

Średnia	**Rzeczywista wartość *x***	**Średnica δ_x**	**Najmniejsza odległość od średniej**	**Najlepszy rzeczywisty**	**Dokonany wybór**	**Błąd względny prognozy**
15875.9	26001	44134.28	$X(x,z), X(x,q)$	$X(x,r)s, X(x,z)s$	$X(x,z)s$	0.5%

Dla modeli ze zmienną *y* otrzymujemy:

Tab. 4.19 Wartości odległości Euklidesa oraz Canberra prognoz uzyskiwanych przez poszczególne modele ze zmienną *y* oraz odległość od wartości średniej tych prognoz dla punktu czasowego $P(7)$

Modele	$Y(y,x)s$	$Y(y,z)s$	$Y(y,q)s$	$Y(y,r)s$	$Y(y,x)$	$Y(y,z)$	$Y(y,q)$	$Y(y,r)$
Odległość Euklidesa	453948.8	261111.1	318320.6	305183.1	562711.3	299811.6	304167.2	305947.8
Odległość Canberra	0.572243	0.686931	0.808822	0.852399	0.714023	0.77289	0.787478	0.786073

Średnia	**Rzeczywista wartość *y***	**Średnica δ_y**	**Najmniejsza odległość od średniej**	**Najlepszy rzeczywisty model**	**Dokonany wybór**	**Błąd względny prognozy**
66688	125671	243023	$Y(y,z)s, Y(y,z)$	$Y(y,r), Y(y,r)s$	$Y(y,z)$	598%

Dla modeli ze zmienną *z* otrzymujemy:

Tab. 4.20 Wartości odległości Euklidesa oraz Canberra prognoz uzyskiwanych przez poszczególne modele ze zmienną *z* oraz odległość od wartości średniej tych prognoz dla punktu czasowego $P(7)$

Modele	$Z(z,x)s$	$Z(z,y)s$	$Z(z,q)s$	$Z(z,r)s$	$Z(z,x)$	$Z(z,y)$	$Z(z,q)$	$Z(z,r)$
Odległość Euklidesa	49666.16	125223.9	54923.09	68498.46	49691.74	50485.72	49608.88	77667.98
Odległość Canberra	0.724691	0.012966	0.693254	0.64524	0.754346	0.721424	0.759205	0.621386

Średnia	**Rzeczywista wartość *z***	**Średnica δ_z**	**Najmniejsza odległość od średniej**	**Najlepszy rzeczywisty model**	**Dokonany wybór**	**Błąd względny prognozy**
40077.6	38306	61781.8	$Z(z,q)$	$Z(z,x), Z(z,x)s$	$Z(z,q)$	4%

Dla modeli ze zmienną *q* otrzymujemy:

Tab. 4.21 Wartości odległości Euklidesa oraz Canberra prognoz uzyskiwanych przez poszczególne modele ze zmienną q oraz odległość od wartości średniej tych prognoz dla punktu czasowego $P(7)$

Modele	$Q(q,x)s$	$Q(q,y)s$	$Q(q,z)s$	$Q(q,r)s$	$Q(q,x)$	$Q(q,y)$	$Q(q,z)$	$Q(q,r)$
Odległość Euklidesa	13585.31	14301.14	27369.44	30670.59	13600.25	13481.49	14049.64	13559.08
Odległość Canberra	0.889376	0.880043	0.750428	0.652888	0.890825	0.891657	0.885799	0.890214

Średnia	Rzeczywista wartość q	Średnica δ_q	Najmniejsza odległość od średniej	Najlepszy rzeczywisty model	Dokonany wybór	Błąd względny prognozy
30276.3	30117	18231	$Q(q,y), Q(q,r)$	$Q(q,x)s, Q(q,x), Q(q,y)$	$Q(q,y)$	3.2%

Dla modeli ze zmienną r otrzymujemy:

Tab. 4.22 Wartości odległości Euklidesa oraz Canberra prognoz uzyskiwanych przez poszczególne modele ze zmienną r oraz odległość od wartości średniej tych prognoz dla punktu czasowego $P(7)$

Modele	$R(r,x)s$	$R(r,y)s$	$R(r,z)s$	$R(r,q)s$	$R(r,x)$	$R(r,y)$	$R(r,z)$	$R(r,q)$
Odległość Euklidesa	0.526177	0.581431	0.5359	1.396433	0.515386	0.592326	0.534715	0.500886
Odległość Canberra	0.992454	0.990873	0.992433	0.96041	0.992247	0.990502	0.992554	0.990702

Średnia	Rzeczywista wartość r	Średnica δ_r	Najmniejsza odległość od średniej	Najlepszy rzeczywisty model	Dokonany wybór	Błąd względny prognozy
13.24	13.304	0.573	$R(r,q), R(r,x)$	$R(r,z), R(r,z)s, R(r,x)s$	$R(r,x)s$	6%

Ponownie dokonujemy obliczeń wykorzystując te same współczynniki WKG oraz wyznaczone w poprzednim kroku prognozowane wartości w celu wyznaczenia wartości poszczególnych zmiennych w punkcie czasowym $P(8)$.

Dla modeli ze zmienną x otrzymujemy:

Tab. 4.23 Wartości odległości Euklidesa oraz Canberra prognoz uzyskiwanych przez poszczególne modele ze zmienną x oraz odległość od wartości średniej tych prognoz dla punktu czasowego $P(8)$

Modele	$X(x,y)s$	$X(x,z)s$	$X(x,q)s$	$X(x,r)s$	$X(x,y)$	$X(x,z)$	$X(x,q)$	$X(x,r)$
Odległość Euklidesa	67426.3	69797.11	70388.9	72712.16	174649.1	69129.48	70420.47	73023.01
Odległość Canberra	0.339928	0.834785	0.840529	0.827349	0.662551	0.699991	0.729823	0.732442

Średnia	Rzeczywista wartość x	Średnica δ_x	Najmniejsza odległość od średniej	Najlepszy rzeczywisty model	Dokonany wybór	Błąd względny prognozy
17782	30730	70000	$X(x,y)s, X(x,z)$	$X(x,r), X(x,r)s$	$X(x,q)s$	8%

Dla modeli ze zmienną y otrzymujemy:

Tab. 4.24 Wartości odległości Euklidesa oraz Canberra prognoz uzyskiwanych przez poszczególne modele ze zmienną y oraz odległość od wartości średniej tych prognoz dla punktu czasowego $P(8)$

Modele	$Y(y,x)s$	$Y(y,z)s$	$Y(y,q)s$	$Y(y,r)s$	$Y(y,x)$	$Y(y,z)$	$Y(y,q)$	$Y(y,r)$
Odległość Euklidesa	408797	702976,8	565728,9	506077,2	867587	423534,3	482773,6	507766
Odległość Canberra	0.113231	0.602092	0.626418	0.650616	0.372195	0.390762	0.588496	0.61364

Średnia	Rzeczywista wartość y	Średnica δ_y	Najmniejsza odległość od średniej	Najlepszy rzeczywisty model	Dokonany wybór	Błąd względny prognozy
66688	132516	409388	$Y(y,x)s, Y(y,z), Y(y,r)s$	$Y(y,r), Y(y,r)s$	$Y(y,x)s$	50.9%

Ze względu na wysoką wartość średnicy δ_y, do dokładności prognozy ze zmienną y nie możemy mieć dużego zaufania.

Dla modeli ze zmienną z otrzymujemy:

Tab. 4.25 Wartości odległości Euklidesa oraz Canberra prognoz uzyskiwanych przez poszczególne modele ze zmienną z oraz odległość od wartości średniej tych prognoz dla punktu czasowego $P(8)$

Modele	$Z(z,x)s$	$Z(z,y)s$	$Z(z,q)s$	$Z(z,r)s$	$Z(z,x)$	$Z(z,y)$	$Z(z,q)$	$Z(z,r)$
Odległość Euklidesa	402202.7	915135.6	402036.3	612950.6	394632.8	427593.4	393726.5	668367.2
Odległość Canberra	0.277879	0.378885	0.364071	0.383933	0.48873	0.422682	0.542238	0.436715

Średnia	Rzeczywista wartość z	Średnica δ_z	Najmniejsza odległość od średniej	Najlepszy rzeczywisty model	Dokonany wybór	Błąd względny prognozy
47601	46114	483139	$Z(z,q)$	$Z(z,q), Z(z,x)$	$Z(z,q)$	14%

Najprawdopodobniej udało się wskazać model najdokładniejszy, wysoka wartość δ_z wynika z rzeczywistych wysokich wartości szeregu.

Dla modeli ze zmienną q otrzymujemy:

Tab. 4.26 Wartości odległości Euklidesa oraz Canberra prognoz uzyskiwanych przez poszczególne modele ze zmienną q oraz odległość od wartości średniej tych prognoz dla punktu czasowego $P(8)$

Modele	$Q(q,x)s$	$Q(q,y)s$	$Q(q,z)s$	$Q(q,r)s$	$Q(q,x)$	$Q(q,y)$	$Q(q,z)$	$Q(q,r)$
Odległość Euklidesa	72158.51	79633.44	108273.7	184488.8	72078.35	71102.94	75192.69	87330.6
Odległość Canberra	0.725691	0.732292	0.595784	0.623106	0.622152	0.692726	0.604442	0.595562

Średnia	Rzeczywista wartość q	Średnica δ_q	Najmniejsza odległość od średniej	Najlepszy rzeczywisty model	Dokonany wybór	Błąd względny prognozy
34867	36300	89067	$Q(q,y)$	$Q(q,y)$	$Q(q,y)$	2%

Dla modeli ze zmienną r otrzymujemy:

Tab. 4.27 Wartości odległości Euklidesa oraz Canberra prognoz uzyskiwanych przez poszczególne modele ze zmienną r oraz odległość od wartości średniej tych prognoz dla punktu czasowego $P(8)$

Modele	$R(r,x)s$	$R(r,y)s$	$R(r,z)s$	$R(r,q)s$	$R(r,x)$	$R(r,y)$	$R(r,z)$	$R(r,q)$
Odległość Euklidesa	3.590767	3.632424	4.05525	9.817939	3.590488	3.640307	4.152232	4.670678
Odległość Canberra	0.941166	0.943462	0.935812	0.734052	0.943515	0.946128	0.937111	0.920066

Średnia	Rzeczywista wartość r	Średnica δ_r	Najmniejsza odległość od średniej	Najlepszy rzeczywisty model	Dokonany wybór	Błąd względny prognozy
13.41	13.505	4.3	$R(r,x), R(r,x)s$	$R(r,x), R(r,x)s$	$R(r,x)$	0.25%

Gdy po dokonaniu prognozy na jeden krok wprzód, uwzględniając najlepsze z otrzymanych rezultatów, dokonamy przeliczenia współczynników WKG, wyniki

w kolejnych punktach czasowych $(P(8), P(9))$ są następujące: (Rys. 4.13 – Rys. 4.15, Tab. 2.28 – Tab. 2.37).

Dla modeli ze zmienną *x* otrzymujemy:

Tab. 4.28 Wartości odległości Euklidesa oraz Canberra prognoz uzyskiwanych przez poszczególne modele ze zmienną *x* oraz odległość od wartości średniej tych prognoz dla punktu czasowego $P(8)$

Modele	$X(x,y)s$	$X(x,z)s$	$X(x,q)s$	$X(x,r)s$	$X(x,y)$	$X(x,z)$	$X(x,q)$	$X(x,r)$
Odległość Euklidesa	44434.08	34320.9	33424	35441.23	91349.74	34166.9	33330.47	36630.31
Odległość Canberra	0.703213	0.818871	0.801741	0.811245	0.171564	0.791069	0.760647	0.782839

Średnia	Rzeczywista wartość *x*	Średnica δ_x	Najmniejsza odległość od średniej	Najlepszy rzeczywisty	Dokonany wybór	Błąd względny prognozy	Błąd względny prognozy bez przesunięcia
25089	30730	40603	$X(x,q)$, $X(x,q)s$	$X(x,r)$, $X(x,r)s$	$X(x,q)s$	12.7%	8%

Dla modeli ze zmienną *y* otrzymujemy:

Tab. 4.29 Wartości odległości Euklidesa oraz Canberra prognoz uzyskiwanych przez poszczególne modele ze zmienną *y* oraz odległość od wartości średniej tych prognoz dla punktu czasowego $P(8)$

Modele	$Y(y,x)s$	$Y(y,z)s$	$Y(y,q)s$	$Y(y,r)s$	$Y(y,x)$	$Y(y,z)$	$Y(y,q)$	$Y(y,r)$
Odległość Euklidesa	289000.5	214404.6	220185.1	231817.9	588210.9	215402.5	221828.3	232589.1
Odległość Canberra	0.69121	0.854171	0.883138	0.860142	0.841722	0.767598	0.735631	0.659156

Średnia	Rzeczywista wartość *y*	Średnica δ_y	Najmniejsza odległość od średniej	Najlepszy rzeczywisty model	Dokonany wybór	Błąd względny prognozy	Błąd względny prognozy bez przesunięcia
93218	132516	262793	$Y(y,z)s$, $Y(y,z)$	$Y(y,r)$, $Y(y,r)s$	$Y(y,z)s$	23%	50.9%

Dla modeli ze zmienną *z* otrzymujemy:

Tab. 4.30 Wartości odległości Euklidesa oraz Canberra prognoz uzyskiwanych przez poszczególne modele ze zmienną *z* oraz odległość od wartości średniej tych prognoz dla punktu czasowego $P(8)$

Modele	$Z(z,x)s$	$Z(z,y)s$	$Z(z,q)s$	$Z(z,r)s$	$Z(z,x)$	$Z(z,y)$	$Z(z,q)$	$Z(z,r)$
Odległość Euklidesa	36771.51	35051.12	33450.61	47289.22	35775.19	35594.06	33285.59	72650.68
Odległość Canberra	0.84327	0.854076	0.851087	0.643228	0.866178	0.86729	0.865393	0.460716

Średnia	Rzeczywista wartość *z*	Średnica δ_z	Najmniejsza odległość od średniej	Najlepszy rzeczywisty model	Dokonany wybór	Błąd względny prognozy	Błąd względny prognozy bez przesunięcia
45743	46114	30585	$Z(z,q)$, $Z(z,q)s$	$Z(z,q)$, $Z(z,q)s$	$Z(z,q)$	9%	14%

Dla modeli ze zmienną *q* otrzymujemy:

Tab. 4.31 Wartości odległości Euklidesa oraz Canberra prognoz uzyskiwanych przez poszczególne modele ze zmienną *q* oraz odległość od wartości średniej tych prognoz dla punktu czasowego $P(8)$

Modele	$Q(q,x)s$	$Q(q,y)s$	$Q(q,z)s$	$Q(q,r)s$	$Q(q,x)$	$Q(q,y)$	$Q(q,z)$	$Q(q,r)$
Odległość Euklidesa	242492	147111.3	142643.8	155302.2	268843.6	146531.6	142522.7	154535.3
Odległość Canberra	0.777953	0.85746	0.823851	0.850593	0.521194	0.69518	0.663839	0.73322

Średnia	Rzeczywista wartość q	Średnica δ_q	Najmniejsza odległość od średniej	Najlepszy rzeczywisty model	Dokonany wybór	Błąd względny prognozy	Błąd względny prognozy bez przesunięcia
16873	36300	114558	$Q(q,z)$, $Q(q,z)s$	$Q(q,z)$, $Q(q,z)s$	$Q(q,z)s$	7.4%	2%

Dla modeli ze zmienną *r* otrzymujemy:

Tab. 4.32 Wartości odległości Euklidesa oraz Canberra prognoz uzyskiwanych przez poszczególne modele ze zmienną *r* oraz odległość od wartości średniej tych prognoz dla punktu czasowego $P(8)$

Modele	$R(r,x)s$	$R(r,y)s$	$R(r,z)s$	$R(r,q)s$	$R(r,x)$	$R(r,y)$	$R(r,z)$	$R(r,q)$
Odległość Euklidesa	0.584027	0.591654	0.649317	0.617492	0.579074	0.576287	1.122171	0.640976
Odległość Canberra	0.989501	0.989402	0.983273	0.988756	0.984978	0.989417	0.970923	0.987845

Średnia	Rzeczywista wartość r	Średnica δ_r	Najmniejsza odległość od średniej	Najlepszy rzeczywisty model	Dokonany wybór	Błąd względny prognozy	Błąd względny prognozy bez przesunięcia
13.654	13.505	0.503	$R(r,y)$, $R(r,x)$	$R(r,q)$, $R(r,q)s$, $R(r,x)s$	$R(r,x)s$	0.26%	0.25%

Dokonujemy ponownie obliczeń wykorzystując te same współczynniki WKG oraz wyznaczone w poprzednim kroku prognozowane wartości w celu wyznaczenia wartości poszczególnych zmiennych w punkcie czasowym $P(9)$.

Dla modeli ze zmienną *x* otrzymujemy:

Tab. 4.33 Wartości odległości Euklidesa oraz Canberra prognoz uzyskiwanych przez poszczególne modele ze zmienną *x* oraz odległość od wartości średniej tych prognoz dla punktu czasowego $P(9)$

Modele	$X(x,y)s$	$X(x,z)s$	$X(x,q)s$	$X(x,r)s$	$X(x,y)$	$X(x,z)$	$X(x,q)$	$X(x,r)$
Odległość Euklidesa	231006.9	137926.4	133396.9	137387	329916.1	137849	134799	137416.9
Odległość Canberra	0.357025	0.570945	0.2275	0.570358	0.351333	0.566624	0.376754	0.56199

Średnia	Rzeczywista wartość x	Średnica δ_x	Najmniejsza odległość od średniej	Najlepszy rzeczywisty	Dokonany wybór	Błąd względny prognozy
15386	37446	174383	$X(x,q)s$, $X(x,q)$	$X(x,z)s$, $X(x,z)$	$X(x,q)$	89.5%

Dla modeli ze zmienną *y* otrzymujemy:

Tab. 4.34 Wartości odległości Euklidesa oraz Canberra prognoz uzyskiwanych przez poszczególne modele ze zmienną *y* oraz odległość od wartości średniej tych prognoz dla punktu czasowego $P(9)$

Modele	$Y(y,x)s$	$Y(y,z)s$	$Y(y,q)s$	$Y(y,r)s$	$Y(y,x)$	$Y(y,z)$	$Y(y,q)$	$Y(y,r)$
Odległość Euklidesa	1551050	996931.8	909639.2	1121333	2157632	970134.7	910528.8	1137011
Odległość Canberra	0.341272	0.409325	0.164302	0.447827	0.230419	0.27932	0.453578	0.473959

Średnia	Rzeczywista wartość y	Średnica δ_y	Najmniejsza odległość od średniej	Najlepszy rzeczywisty model	Dokonany wybór	Błąd względny prognozy
-1824.7	137193.9	1136426	$Y(y,q)s$, $Y(y,q)$	$Y(y,r)s$, $Y(y,r)$	$Y(y,q)$	85%

Dla modeli ze zmienną *z* otrzymujemy:

Tab. 4.35 Wartości odległości Euklidesa oraz Canberra prognoz uzyskiwanych przez poszczególne modele ze zmienną z oraz odległość od wartości średniej tych prognoz dla punktu czasowego $P(9)$

Modele	$Z(z,x)s$	$Z(z,y)s$	$Z(z,q)s$	$Z(z,r)s$	$Z(z,x)$	$Z(z,y)$	$Z(z,q)$	$Z(z,r)$
Odległość Euklidesa	844999.3	460031.9	482920.8	618912.8	866732.5	458246.7	461124.6	710056.4
Odległość Canberra	0.293519	0.623287	0.503415	0.514975	0.288857	0.577862	0.558115	0.583906

Średnia	Rzeczywista wartość z	Średnica δ_z	Najmniejsza odległość od średniej	Najlepszy rzeczywisty model	Dokonany wybór	Błąd względny prognozy
106299	55032	456791	$Z(z,y)$, $Z(z,y)s$	$Z(z,q)s$, $Z(z,q)$	$Z(z,y)$	33%

Dla modeli ze zmienną q otrzymujemy:

Tab. 4.36 Wartości odległości Euklidesa oraz Canberra prognoz uzyskiwanych przez poszczególne modele ze zmienną q oraz odległość od wartości średniej tych prognoz dla punktu czasowego $P(9)$

Modele	$Q(q,x)s$	$Q(q,y)s$	$Q(q,z)s$	$Q(q,r)s$	$Q(q,x)$	$Q(q,y)$	$Q(q,z)$	$Q(q,r)$
Odległość Euklidesa	2734408	1661303	1617945	1683569	2988819	1660905	1628085	1669306
Odległość Canberra	0.190445	0.602529	0.46117	0.548311	0.201255	0.72305	0.660081	0.703679

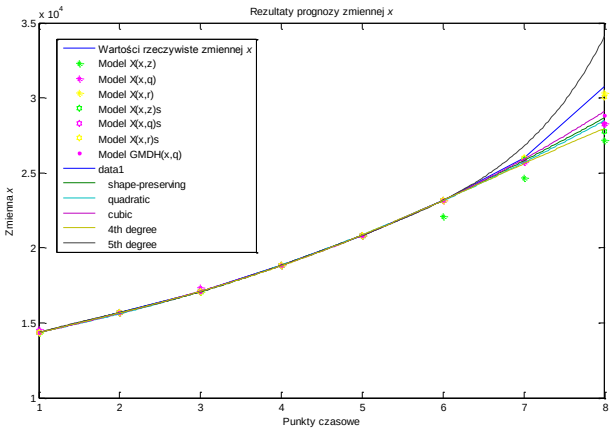
Średnia	Rzeczywista wartość q	Średnica δ_q	Najmniejsza odległość od średniej	Najlepszy rzeczywisty model	Dokonany wybór	Błąd względny prognozy
-218249	44277	1240397	$Q(q,z)s, Q(q,z)$	$Q(q,z)s, Q(q,z)$	$Q(q,z)$	12%

Dla modeli ze zmienną r otrzymujemy:

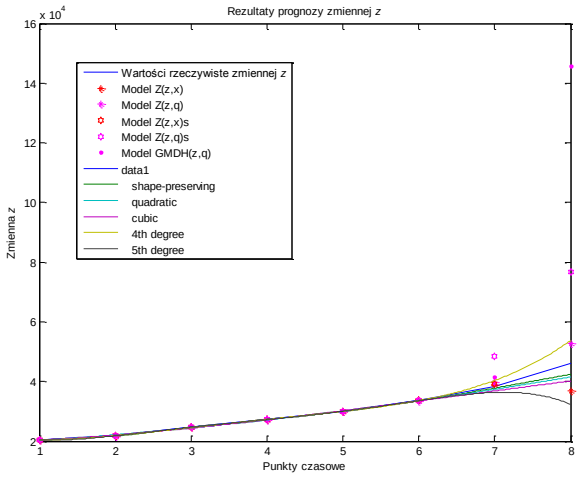
Tab. 4.37 Wartości odległości Euklidesa oraz Canberra prognoz uzyskiwanych przez poszczególne modele ze zmienną r oraz odległość od wartości średniej tych prognoz dla punktu czasowego $P(9)$

Modele	$R(r,x)s$	$R(r,y)s$	$R(r,z)s$	$R(r,q)s$	$R(r,x)$	$R(r,y)$	$R(r,z)$	$R(r,q)$
Odległość Euklidesa	4.670684	3.86638	4.289045	4.134662	5.046518	4.08446	6.140166	4.289486
Odległość Canberra	0.911695	0.919679	0.914025	0.920141	0.900121	0.918529	0.875107	0.917464

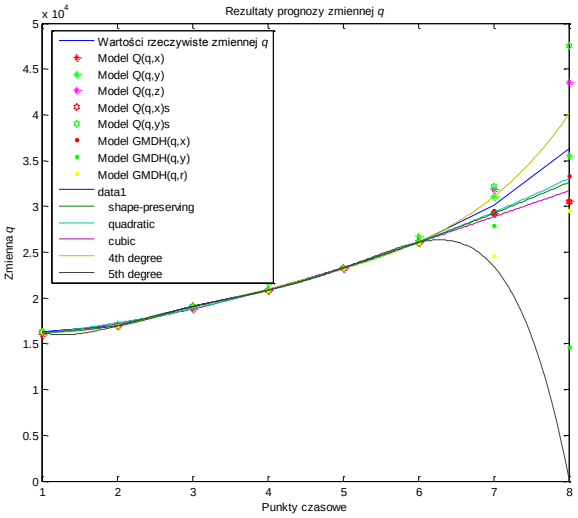
Średnia	Rzeczywista wartość r	Średnica δ_r	Najmniejsza odległość od średniej	Najlepszy rzeczywisty model	Dokonany wybór	Błąd względny prognozy
14.86	13.723	3.2	$R(r,y)s, R(r,y)$	$R(r,x)s, R(r,q)$	$R(r,y)s$	13.6%



Rys. 4.13 Porównanie rezultatów prognozy najlepszych modeli z funkcjami wrażliwości, uproszczonej metody GMDH oraz kilku metod interpolacyjnych dla zmiennej x, na dwa kroki wprzód
Źródło: Opracowanie własne (Matlab)



Rys. 4.14 Porównanie rezultatów prognozy najlepszych modeli z funkcjami wrażliwości, uproszczonej metody GMDH oraz kilku metod interpolacyjnych dla zmiennej *z*, na dwa kroki wprzód
Źródło: Opracowanie własne (Matlab)



Rys. 4.15 Porównanie rezultatów prognozy najlepszych modeli z funkcjami wrażliwości, uproszczonej metody GMDH oraz kilku metod interpolacyjnych dla zmiennej *q*, na dwa kroki wprzód
Źródło: Opracowanie własne (Matlab)

W przypadku szeregu modelującego zachowanie się złożonego procesu rzeczywistego, możliwe było wskazanie najlepszego modelu za pomocą porównania metryk Euklidesa i Canberra. Natomiast wysoka wartość średnicy poszczególnych modeli daje informacje o niskiej wiarygodności dokonywanych prognoz. W przypadku zmiennej *y*, wiarygodność ta była najniższa, co skutkowało wskazaniem niewłaściwego (nienajlepszego) modelu prognozy. Prognozy dokonywane za pomocą uproszczonych modeli GMDH dawały dużo gorsze rezultaty niż modele z funkcjami wrażliwości. Również popularne metody ekstrapolacji dawały gorsze rezultaty niż nasza metoda. Natomiast prognoza dla punktu czasowego $P(9)$ była niezadowalająca w przypadku większości modeli każdej ze zmiennych. Powyższy przykład wskazuje, że możliwe jest

uzyskanie poprawy skuteczności prognozy (na 1-2 kroków wprzód) nawet w przypadku szeregów, dla których modele MAMC nie są w stanie ujawnić zachodzących ewentualnych w nich regularności, gdyż wartości inicjujących było za mało.

4.4 Możliwość prognozowania wartości rzeczywistego procesu złożonego bez regularności

Przykład zaczerpnięto z dostępnych w zasobach sieciowych zbiorów danych w postaci szeregu czasowego. Przedstawia on zmiany w czasie rzeczywistych, wzajemnie powiązanych ze sobą czynników. Przykład obrazuje możliwość prognozowania szeregów czasowych, które nie są „sztucznie" uzyskiwane poprzez generowanie poszczególnych zestawów punktów czasowych $P(i)$ za pomocą wzorów matematycznych. Jednak jego zachowanie nie wykazuje zauważalnych regularności, (Tab. 4.38). Metoda oceny dokładności modeli jest skuteczna również w przypadku tego typu modeli. Jesteśmy, więc w stanie wskazać spośród dostępnych 8 modeli ten, który jest najbliżej wartości rzeczywistej, choć wskazywany przez niego rezultat nie zawsze jest zadowalający.

Tab. 4.38 Proces złożony przedstawiony za pomocą szeregu czasowego o pięciu zmiennych

Punkty czasowe	x	y	z	q	r
P(0)	62	-11	32	-255	508
P(1)	-52	11	46	-84	637
P(2)	73	8	16	128	807
P(3)	71	-39	95	235	807
P(4)	-29	-47	126	292	807
P(5)	53	-57	17	368	727
P(6)	44	-93	-160	376	659
P(7)	-85	-49	-319	274	663
P(8)	-190	-30	-319	154	628

…

Dla modeli ze zmienną *x* otrzymujemy:

Tab. 4.39 Wartości odległości Euklidesa oraz Canberra prognoz uzyskiwanych przez poszczególne modele ze zmienną *x* oraz odległość od wartości średniej tych prognoz dla punktu czasowego $P(7)$

Modele	$X(x,y)s$	$X(x,z)s$	$X(x,q)s$	$X(x,r)s$	$X(x,y)$	$X(x,z)$	$X(x,q)$	$X(x,r)$
Odległość Euklidesa	3057.6	1821.6	1546.9	1589.3	1655.3	3443.5	1547.997	1730.9
Odległość Canberra	0.308	0.4361	0.1359	0.4074	0.4779	0.307	0.5653	0.2792

Średnia	Rzeczywista wartość x	Średnica δ_x	Najmniejsza odległość od średniej	Najlepszy rzeczywisty	Dokonany wybór	Błąd względny prognozy
−86.36	−85.36	2025.69	$X(x,q)s, X(x,q)$	$X(x,q)$	$X(x,q)$	89%

Dla modeli ze zmienną *y* otrzymujemy:

Tab. 4.40 Wartości odległości Euklidesa oraz Canberra prognoz uzyskiwanych przez poszczególne modele ze zmienną *y* oraz odległość od wartości średniej tych prognoz dla punktu czasowego $P(7)$

Modele	$Y(y,x)s$	$Y(y,z)s$	$Y(y,q)s$	$Y(y,r)s$	$Y(y,x)$	$Y(y,z)$	$Y(y,q)$	$Y(y,r)$
Odległość Euklidesa	2126.4	2647.3	1737.5	1482.1	1520.13	2853.1	2157.11	1462.65
Odległość Canberra	0.4509	0.3911	0.3736	0.1851	0.383	0.2959	0.3702	0.5422

Średnia	Rzeczywista wartość *y*	Średnica δ_y	Najmniejsza odległość od średniej	Najlepszy rzeczywisty model	Dokonany wybór	Błąd względny prognozy
-52.23	-49	1429.38	$Y(y,r), Y(y,r)s$	$Y(y,r),$	$Y(y,r)$	90%

Dla modeli ze zmienną *z* otrzymujemy:

Tab. 4.41 Wartości odległości Euklidesa oraz Canberra prognoz uzyskiwanych przez poszczególne modele ze zmienną z oraz odległość od wartości średniej tych prognoz dla punktu czasowego $P(7)$

Modele	$Z(z,x)s$	$Z(z,y)s$	$Z(z,q)s$	$Z(z,r)s$	$Z(z,x)$	$Z(z,y)$	$Z(z,q)$	$Z(z,r)$
Odległość Euklidesa	27385.4	30717	47259.4	26659.4	26336.8	31858.1	39786.9	29822.4
Odległość Canberra	0.2411	0.4171	0.2373	0.2472	0.58387	0.5418	0.25599	0.5271

Średnia	Rzeczywista wartość *z*	Średnica δ_z	Najmniejsza odległość od średniej	Najlepszy rzeczywisty model	Dokonany wybór	Błąd względny prognozy
-5520.794	-319	22115	$Z(z,x), Z(z,r)s$	$Z(z,x)s$	$Z(z,x)$	300%

Dla modeli ze zmienną *q* otrzymujemy:

Tab. 4.42 Wartości odległości Euklidesa oraz Canberra prognoz uzyskiwanych przez poszczególne modele ze zmienną *q* oraz odległość od wartości średniej tych prognoz dla punktu czasowego $P(7)$

Modele	$Q(q,x)s$	$Q(q,y)s$	$Q(q,z)s$	$Q(q,r)s$	$Q(q,x)$	$Q(q,y)$	$Q(q,z)$	$Q(q,r)$
Odległość Euklidesa	11273.5	8941.5	18330.3	12202.2	11280.9	9125.7	15100	11254.7
Odległość Canberra	0.373	0.2614	0.2434	0.1344	0.531	0.277	0.284	0.5612

Średnia	Rzeczywista wartość *q*	Średnica δ_q	Najmniejsza odległość od średniej	Najlepszy rzeczywisty model	Dokonany wybór	Błąd względny prognozy
2854.8	274	8637.43	$Q(q,y)s, Q(q,y)$	$Q(q,x)$	$Q(q,y)$	1221%

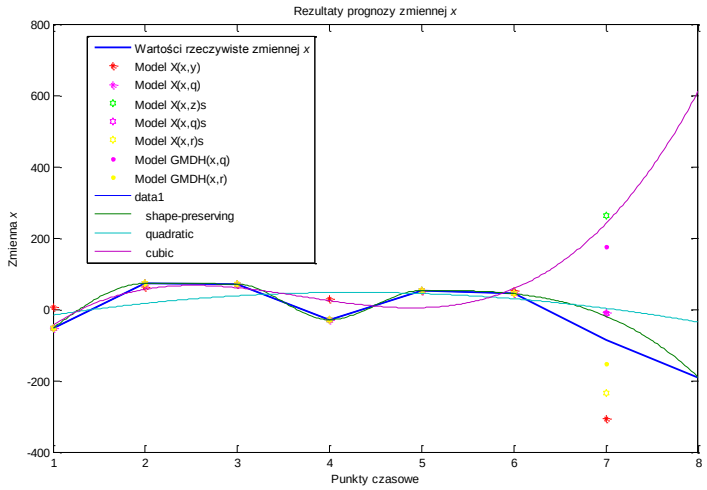
Dla modeli ze zmienną *r* otrzymujemy:

Tab. 4.43 Wartości odległości Euklidesa oraz Canberra prognoz uzyskiwanych przez poszczególne modele ze zmienną *r* oraz odległość od wartości średniej tych prognoz dla punktu czasowego $P(7)$

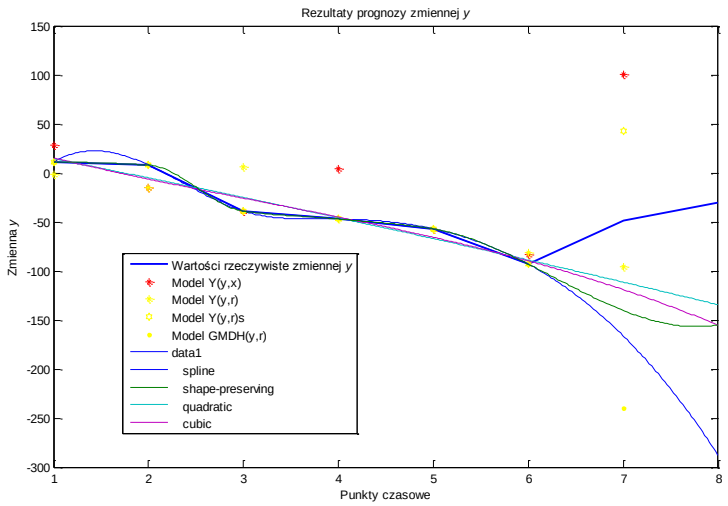
Modele	$R(r,x)s$	$R(r,y)s$	$R(r,z)s$	$R(r,q)s$	$R(r,x)$	$R(r,y)$	$R(r,z)$	$R(r,q)$
Odległość Euklidesa	7964.7	9021.9	21734.8	8608.2	7983.7	8005.76	8858.62	7984.85
Odległość Canberra	0.6063	0.5203	0.1443	0.5615	0.7355	0.7333	0.6636	0.637

Średnia	Rzeczywista wartość *r*	Średnica δ_r	Najmniejsza odległość od średniej	Najlepszy rzeczywisty model	Dokonany wybór	Błąd względny prognozy
9.025	663	8811.11	$R(r,x)s\ R(r,x),$ $R(r,q)$	$R(r,x)s$	$R(r,x)$	3%

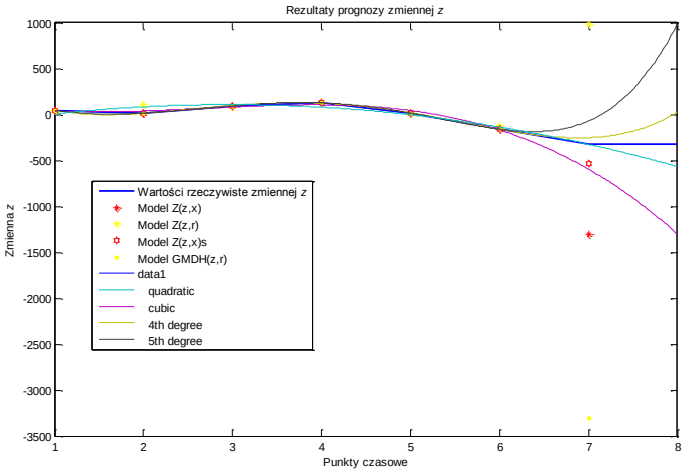
Najlepsze rezultaty uzyskane za pomocą uproszczonej metody GMDH były dla każdej zmiennej mniej korzystne niż w przypadku naszej metody. Podobnie dla popularnych metod ekstrapolacji (przykładowe rezultaty na Rys. 4.16 – Rys. 4.20). Prognozy dla kolejnego punktu czasowego ($P(8)$), dokonane zarówno przy pomocy najlepszych modeli oraz po przeliczeniu współczynników WKG znacznie odbiegają od wartości rzeczywistych. Jednak ze względu na precyzyjną identyfikację modeli najdokładniejszych oraz większą od najpopularniejszych metod ekstrapolacji skuteczność metody MAMC, można ją wskazać, jako wspomagającą dla innych metod prognostycznych.



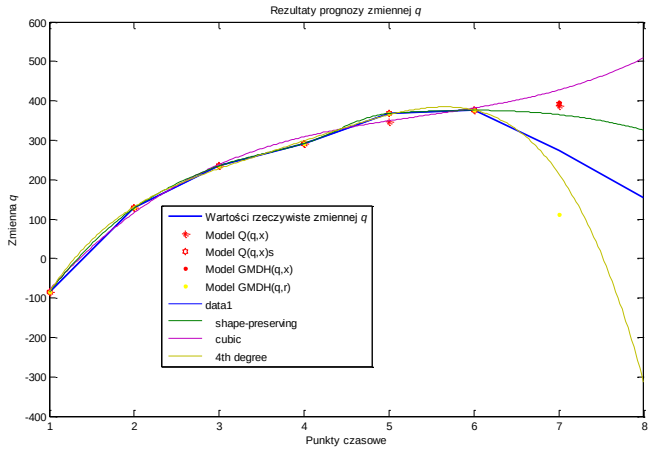
Rys. 4.16 Porównanie rezultatów prognozy najlepszych modeli z funkcjami wrażliwości, uproszczonej metody GMDH oraz kilku metod interpolacyjnych dla zmiennej *x*, na jeden krok wprzód
Źródło: Opracowanie własne (Matlab)



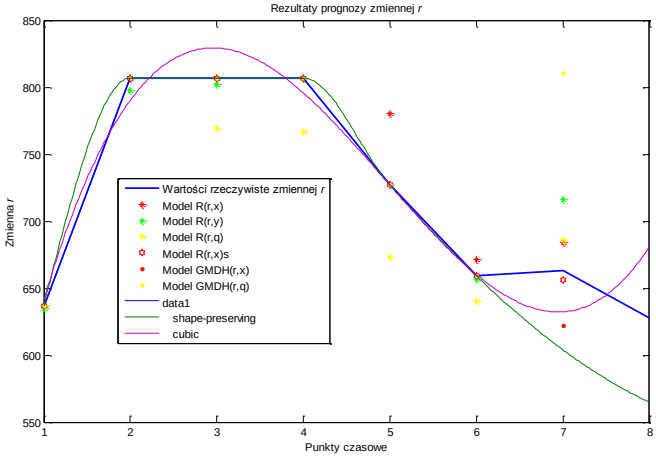
Rys. 4.17 Porównanie rezultatów prognozy najlepszych modeli z funkcjami wrażliwości, uproszczonej metody GMDH oraz kilku metod interpolacyjnych dla zmiennej *y*, na jeden krok wprzód
Źródło: Opracowanie własne (Matlab)



Rys. 4.18 Porównanie rezultatów prognozy najlepszych modeli z funkcjami wrażliwości, uproszczonej metody GMDH oraz kilku metod interpolacyjnych dla zmiennej *z*, na jeden krok wprzód
Źródło: Opracowanie własne (Matlab)



Rys. 4.19 Porównanie rezultatów prognozy najlepszych modeli z funkcjami wrażliwości, uproszczonej metody GMDH oraz kilku metod interpolacyjnych dla zmiennej *q*, na jeden krok wprzód
Źródło: Opracowanie własne (Matlab)



Rys. 4.20 Porównanie rezultatów prognozy najlepszych modeli z funkcjami wrażliwości, uproszczonej metody GMDH oraz kilku metod interpolacyjnych dla zmiennej *r*, na jeden krok wprzód
Źródło: Opracowanie własne (Matlab)

4.5 Prognoza w przyrostach dla szeregu generowanego za pomocą wzorów matematycznych o pięciu zmiennych

W Rozdziale 3. przedstawiono przykład szeregu czasowego o trzech zmiennych, dla którego dokonano prognozy w przyrostach. W przypadku szeregów o większej ilości zmiennych, wykazujących znaczne regularności udało się również wskazać zależności pomiędzy przyrostami $\delta x, \delta y$ oraz $\delta X, \delta Y$. Przykładowe rezultaty uzyskiwane dla danych z Tab. 4.44 przedstawiamy poniżej.

Tab. 4.44 Szereg czasowy o pięciu zmiennych

Punkty czasowe	x	y	z	q	r
P(0)	1.331	0.912525	3.105733	0.093449	1.482149
P(1)	1.4641	1.049404	4.357086	0.084663	1.775313
P(2)	1.61051	1.206814	6.061011	0.069621	2.172107
P(3)	1.771561	1.387836	8.264303	0.053982	2.69246
P(4)	1.948717	1.596012	10.85823	0.04003	3.362909
P(5)	2.143589	1.835414	13.40973	0.027662	4.224049
P(6)	2.357948	2.110726	15.02293	0.014139	5.347513
P(7)	2.593742	2.427335	14.51877	-0.01095	6.886451
P(8)	2.853117	2.791435	11.28116	-0.09852	9.250769

…

Po rozwiązaniu układu równań nieliniowych typu (2.30) i (2.31) ze zmiennymi y oraz r otrzymujemy wartości przyrostów cząstkowych Tab. 4.45.

Tab. 4.45 Wartości przyrostów cząstkowych ddy oraz ddr w modelu $Y(y,r)$

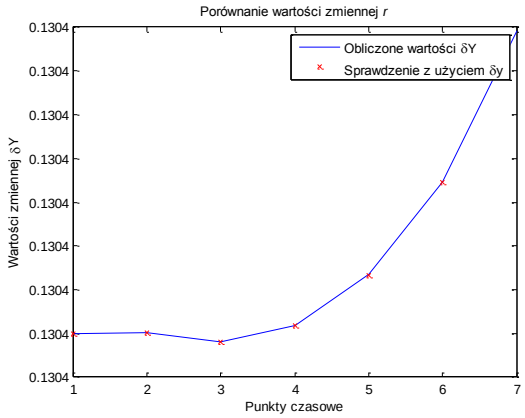
Punkty czasowe	ddy	ddr
P(1)	0.130435	0.145786
P(2)	0.130435	0.171447
P(3)	0.130435	0.188064
P(4)	0.130435	0.197205
P(5)	0.130435	0.201423
P(6)	0.130435	0.203492
P(7)	0.130435	0.206991
P(8)	0.130435	0.218104

Po rozwiązaniu układu równań nieliniowych z modelami $Y(y,r)$ i $R(r,y)$ sprawdzamy dokładność uzyskanych rozwiązań. Następnie, wykorzystując uzyskane wyniki przyrostów cząstkowych stosujemy wzory (3.15) oraz (3.16) wstawiając w miejsce δX i δY wartości przyrostów cząstkowych modelu rozszerzonego oraz wartości funkcji wrażliwości dla modelu $Y(y,r)$ i $R(r,y)$ i rozwiązujemy ponownie układ równań nieliniowych. Po jego rozwiązaniu uzyskujemy wartości przyrostów δy oraz δr (Tab. 4.46).

Tab. 4.46 Wartości przyrostów δy oraz δr dla modeli $Y(y,r)$ i $R(r,y)$

Punkty czasowe	δy	δr
P(1)	0.113422	0.10631
P(2)	0.113422	0.13689
P(3)	0.113422	0.16133
P(4)	0.113422	0.180649
P(5)	0.113422	0.196537
P(6)	0.113423	0.210551
P(7)	0.113423	0.224102

Podobnie, jak powyżej – sprawdzamy dokładność rozwiązań. Przykładowy rezultat przedstawia są na Rys. 4.21. Resztka rozwiązania jest bliska zeru, przystępujemy więc do sprawdzenia wzoru (3.15), który wskazuje, że przyrost δx w tym wzorze oznacza przyrost cząstkowy na próbce N, podczas gdy δX oznacza cząstkowy przyrost ddx na kolejnej próbce $(N+1)$. Wyniki porównania przedstawione są na Rys. 4.22, Rys. 4.23.



Rys. 4.21 Porównanie rzeczywistych wartości zmiennej *y* oraz uzyskanej w modelu $Y(y,r)$
Źródło: Opracowanie własne (Matlab)

Możemy teraz spróbować prognozować wartości przyrostów cząstkowych na przyszłą chwilę czasu, tym samym prognozując wartości szeregu.

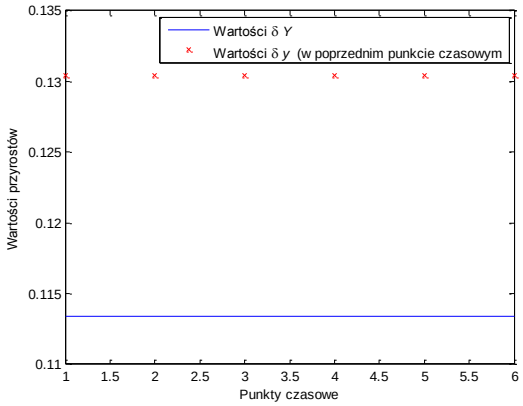
Czyli, znając wartości przyrostów δY i δR, znamy też wartość δy i δr w kolejnym punkcie czasowym.

$\delta Y_6 = 0.130435$, $\delta y_7 = 0.130435$,$\delta R_6 = 0.203492$, $\delta r_7 = 0.203492$.

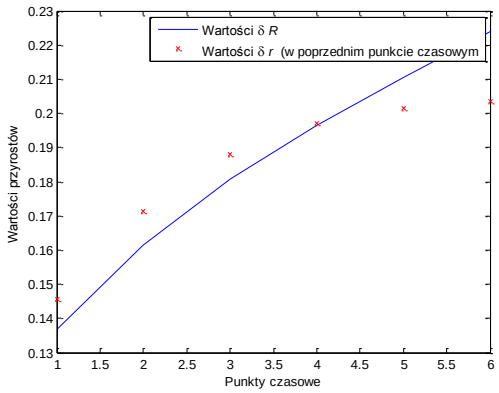
Znając wartości przyrostów oraz wartości różniczkowych i półwzględnych funkcji wrażliwości obliczamy wartości δY_7 oraz δR_7.

$$S'_{y_7,r_7} = \boldsymbol{f}'_x(N) = \begin{vmatrix} 1.149985 & 0.00000353 \\ 1.786 & -1.654 \end{vmatrix};$$

$$S''_{y_7,r_7} = \boldsymbol{f}'_y(N) = \begin{vmatrix} -0.000029597 & -0.000002 & 0.000007 \\ 0.19224 & -0.6419 & -0.24984 \end{vmatrix}.$$



Rys. 4.22 Porównanie wartości przyrostów δy w poprzednim punkcie czasowym z wartościami przyrostów δY w bieżącym punkcie
Źródło: Opracowanie własne (Matlab)



Rys. 4.23 Porównanie wartości przyrostów δr w poprzednim punkcie czasowym z wartościami przyrostów δR w bieżącym punkcie
Źródło: Opracowanie własne (Matlab)

Otrzymujemy więc:

$$\delta Y_7 = 1.149985 \cdot 0.130435 + 0.00000353 \cdot 0.203492 + (-0.000029597) \cdot (0.130435)^2 +$$
$$+(-0.000002) \cdot (0.203492)^2 + (0.000007) \cdot 0.130435 \cdot 0.203492 = 0.149999$$

Rzeczywista wartość $\delta Y_7^{real} = 0.13$.

Obliczamy wartość zmiennej y podstawiając uzyskane wartości przyrostów cząstkowych oraz funkcji wrażliwości do wzoru typu (2.30).

$Y_7 = 2.111,$

$S_y^{Y(y,r)} = 2.427; S_r^{Y(y,r)} = 0.0000189; S_{y^2}^{Y(y,r)} = (-0.0013); S_{r^2}^{Y(y,r)} = -0.000057;$

$S_{yr}^{Y(y,r)} = 0.000076.$

$$Y_8 = 2.111 + 2.427 \cdot 0.15 + 0.0000189 \cdot 0.138 + (-0.0013) \cdot (0.15)^2 +$$
$$+(-0.000057) \cdot (0.38)^2 + 0.000076 \cdot 0.15 \cdot 0.138 = 2.475.$$

Rzeczywista wartość zmiennej $Y_8^{real} = 2.42$

$$\delta R_7 = 1.786 \cdot 0.203492 + (-1.654) \cdot 0.130435 + 0.19224 \cdot (0.203492)^2 +$$

$+(-0.6419)\cdot(0.130435)^2+(-0.24984)\cdot 0.130435\cdot 0.203492=0.138.$

$R_7=5.347.$

$S_r^{R(r,y)}=9.552; S_y^{R(r,y)}=-3.492; S_{r^2}^{R(r,y)}=5.4973; S_{y^2}^{R(r,y)}=-2.85976; S_{ry}^{R(r,y)}=-2.82.$

$R_8=5.347+9.552\cdot 0.138+(-3.492)\cdot 0.15+5.4973\cdot(0.138)^2+$

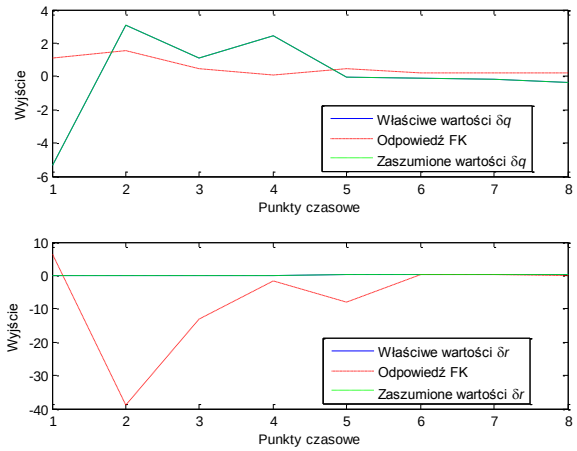
$+(-2.85976)\cdot(0.15)^2+(-2.82)\cdot 0.15\cdot 0.138=6.123$

$R_8^{real}=6.88.$

Tym razem wartość zmiennej r obliczona została z mniejszą dokładnością niż y, zachowany został jednak właściwy kierunek zmian wartości tej zmiennej.

Rezultaty uzyskane z wykorzystaniem FK dla tego samego szeregu czasowego przedstawiają również obiecujące rezultaty. Dla przykładowych modeli $Q(q,r)$ oraz $R(r,q)$ otrzymujemy: (Rys. 4.24).

Wyznaczmy wartości zmiennych q i r danego szeregu w punkcie czasowym $P(7)$ wykorzystując wyznaczone wcześniej wartości funkcji wrażliwości oraz wyznaczone za pomocą FK wartości przyrostów cząstkowych.



Rys. 4.24 Odpowiedź FK na przyrosty cząstkowe w modelu $Q(q,r)$ i $R(r,q)$
Źródło: Opracowanie własne (Matlab)

$q_7=0.053982+(-0.04268)\cdot(0.22)+(-0.12467)\cdot(0.139)+(-0.00195)\cdot(0.22)^2+$

$+(-0.1352)\cdot(0.139)^2+(-0.1069)\cdot(0.22)\cdot(0.139)=0.0213.$

$q_7^{real}=0.04.$

$r_7=2.6925+(3.643)\cdot(0.139)+(0.1528)\cdot(0.22)+(0.1281)\cdot(0.139)^2+$

$+(0.001135)\cdot(0.22)^2+(0.2265)\cdot(0.22)\cdot(0.139)=3.242.$

Rzeczywista wartość $r_7^{real}=3.363.$

Można zauważyć, że w przypadku szeregów regularnych, otrzymujemy wartości przyrostów, które dają się symulować filtrem Kalmana, tym samym mamy możliwość

wyznaczenia wartości takiego przyrostu w kolejnym punkcie czasowym i również wartości samej zmiennej tego szeregu. W momencie, gdy wartości szeregu naprzemiennie rosną i maleją FK nie dawał oczekiwanych rezultatów. Jest to zasadniczym ograniczeniem naszej metody.

4.6 Podsumowanie

W Rozdziale 4. przedstawiono sposób przeprowadzenia oraz szczegółowe wyniki badań eksperymentalnych, których celem było: wskazanie korzyści metody deterministycznego prognozowania szeregów czasowych z wykorzystaniem funkcji wrażliwości, porównanie rezultatów uzyskiwanych z wykorzystaniem naszej metody z innymi popularnymi metodami interpolacji (ekstrapolacji) oraz uproszczoną metodą GMDH.

W związku z tym, że metoda MAMC korzysta z niewielkiej ilości danych uczących, nie możemy stosować narzędzi statystycznych do badania zależności między poszczególnymi zmiennymi badanych procesów, jak to się odbywa w wielu popularnych metodach prognozujących wartości szeregów czasowych.

Ze względu na pewne niedogodności wynikające z małej liczności próbki danych, trudności obliczeń (rozwiązywanie układów równań nieliniowych czy nierzadko uzyskiwanie bliskich zeru wyznaczników macierzy podczas rozwiązywania układów równań liniowych dla modeli uproszczonych), metodę naszą można zarekomendować, jako dopełnienie innych metod prognozujących wartości szeregów czasowych.

Nie mniej jednak, na co wskazują opisane przykłady, nawet niewielka liczba danych okazuje się wystarczająca do precyzyjnej prognozy na 1-3 kroki wprzód, gdy mamy do czynienia z szeregami czasowymi, w których wartości poszczególnych zmiennych wykazują regularności na obserwowanym zakresie danych. Metoda nie może być stosowana do analizy procesów wyjawiających regularność, która może być wykazana jedynie na długim odcinku czasu, przekraczającym ten odcinek czasu, na którym badacz wybiera próbki.

Wartości funkcji wrażliwości stopnia pierwszego i drugiego służą za wskaźnik siły wpływu poszczególnych zmiennych na siebie w modelach odpowiednio uproszczonym i rozszerzonym. Informacje te mogą być przydatne w przypadku, gdy nie mamy dostępu do wiedzy eksperckiej na temat analizowanego procesu.

Wnioski

Cel niniejszej pracy to:
opracowanie metody opartej na teorii identyfikacji systemów o skuteczności większej niż uproszczona metoda podstawowa (GMDH), wykorzystującej wielomian Kołmogorowa-Gabora (podobnie jak w metodzie GMDH), która umożliwia analizę oraz rozwiązanie problemu prognozy procesów złożonych na małolicznych zbiorach próbek z wykorzystaniem narzędzia funkcji wrażliwości.

Aby zrealizować powyższy cel należało *wykazać, że istnieje możliwość zwiększenia dokładności prognozy procesu na małolicznych zbiorach próbek w stosunku do istniejących metod w tym GMDH, przy wykorzystaniu deterministycznych metod identyfikacji systemu wykorzystujących funkcje wrażliwości i szereg tzw. modeli cząstkowych tworzonych na podstawie wielomianu Kołmogorowa-Gabora i podlegających adaptacji w trakcie rozwiązania problemu.*

Aby zweryfikować prawdziwość przyjętej tezy, należało wskazać przykłady szeregów czasowych modelujących zachowanie się pewnych złożonych procesów, systemów, dla których za pomocą naszej metody dokonamy predykcji wartości poszczególnych zmiennych takiego szeregu z dokładnością porównywalną lub większą niż uproszczona metoda GMDH, czy popularne metody ekstrapolacyjne.

Metoda omawiana w danej rozprawie zakłada analizę niewielkiej ilości próbek danych uczących szeregu czasowego przedstawiającego pewien proces, system (minimum – siedem) oraz prognozę na ich podstawie kolejnych wartości tego szeregu. Konstrukcja metody wymusza taką ilość danych inicjujących, gdyż pierwszym jej krokiem jest znalezienie współczynników WKG, a w tym celu należy rozwiązać układ sześciu równań (2.15). Czy możliwa jest prognoza na tak niewielkiej próbce danych uczących? Minimalna ilość danych obserwowalnych potrzebnych do analizy szeregu czasowego wynosi 2. Krótko mówiąc, wystarczy połączyć 2 punkty szeregu na układzie współrzędnych a następnie przedłużyć prostą na taką odległość (dla kolejnych punktów czasowych), jaka jest nam potrzebna. Jednak intuicyjnie wiemy, że najprawdopodobniej proces nie będzie przebiegał liniowo, więc problemem jest raczej ustalenie, jaka jest minimalna ilość punktów eksperymentalnych, dla których można by przeprowadzić dobrą prognozę choćby na 1-3 kroki wprzód. Co znaczy – „dobrą”? Wystarczająco

dokładną. Box i Jenkins [74], rekomendowali np. minimum 50 obserwacji dla modeli ARIMA (*autoregressive integrated moving average*), co jest zwykle zapewniane dla uzyskania oczekiwanych efektów, zauważania sezonowości zmian. To konieczny limit, który nie zawsze jest w takim przypadku wystarczający. Przy takiej ilości danych uczących możemy stosować narzędzia statystyczne do oceny wpływu zmiennych wchodzących w skład badanego procesu na siebie oraz innych użytecznych zależności precyzujących zachowanie się analizowanego procesu. W takich warunkach proces można prognozować na 5-8 kroków wprzód (co też nie zawsze jest możliwe). Co daje nam stosunek $\frac{ilość\ probek\ uczących}{ilość\ udanych\ kroków\ prognozy} = \frac{>50}{od\ 5\ do\ 8}$ czyli mniej więcej 10:1. W przypadku MAMC, przy tak niewielkiej ilości próbek uczących, ten stosunek wynosi $\frac{od\ 7\ do\ 10}{od\ 1\ do\ 3}$ czyli mniej więcej 4:1. Oczywiście w przypadku, gdy ilość próbek uczących jest wystarczająca, aby ocenić na tak krótkim odcinku czasu ewentualną regularność badanego procesu. Taki rezultat jest zadowalający w przypadku, gdy mamy do czynienia z problemem, w którym musimy prognozować proces natychmiast po jego rozpoczęciu. W sytuacji, gdy regularności procesu nie udaje się uchwycić za pomocą niewielkiej ilości danych, nasza metoda okazuje się mało przydatna, gdyż żadna nawet najlepsza metoda matematyczna nie jest w stanie pokonać problemu braku informacji (danych).

Analiza danych procesu (systemu) jest tym trudniejsza im bardziej są one niepełne (niepewne). Jak np. w przypadku systemów czarnych czy szarych. System nazywamy szarym, gdy zawiera informacje zarówno znane i nieznane (niepełne) [75]. Teoria szarych systemów, rozwinięta oryginalnie przez Ju-Deng [76], zakłada analizę niepewności systemów na podstawie niewielkiej ilości danych eksperymentalnych i niekompletności danych. Może stanowić podstawę do prognozowania np. przyszłych ofert w przypadku wyjątkowo krótkich szeregów czasowych, w których liczba obserwacji wynosi $n \geq 4$ (w skrajnych przypadkach zakłada się, że długość szeregu czasowego wynosi $n \geq 2$). Pomija ona nieodłączne ułomności konwencjonalnych metod takich jak teoria prawdopodobieństwa czy praca z niepełnymi, niepewnymi danymi w celu oszacowania zachowania szeregów czasowych [77]. W świecie znana jest dopiero od 1990 roku z uwagi na fakt, iż pierwsze prace były publikowane w języku chińskim. Proces endogeniczny obserwowalny w rzeczywistości, wyjaśniany jest w czasie *N* zmiennymi niezależnymi (objaśniającymi) ($GM(I, N)$, gdzie *I*, to równanie różniczkowe *I*-tego rzędu opisujące system, w zastosowaniach praktycznych najczęściej przyjmowany jest model szary w postaci $GM(1,1)$. Nasza metoda, natomiast, zakłada modelowanie procesów

o większej niż 2 ilości zmiennych, co daje jej przewagę nad powyżej wspomnianą.

W opisywanej metodzie analizy ważne jest również, aby oszacować, w jakim stopniu pełne czy niepewne są posiadane przez nas dane. Nie zawsze też możemy ocenić, co jest źródłem ewentualnej niedokładności naszej metody prognozy. Jej przyczyny bywają różne i mogą wynikać zarówno z małej ilości próbek uczących, ale również z innego rodzaju okoliczności. Podsystem data mining, jak wspomniano w Rozdziale 1. rozprawy, może być częścią większego Systemu Wspomagania Decyzji (DSS). Wtedy takie czynniki zaburzające mogą (powinny) być wzięte pod uwagę przez badacza i pokonywane, np. poprzez sformułowanie dodatkowych zapytań kierowanych do DSS. Bywają czynniki, których nie uda się pokonać nawet z użyciem dodatkowej informacji o procesie, wskazanej przez badacza (analityka, eksperta), np. zbyt wysoka niedokładność pomiarów. Dlatego przedstawiona w pierwszym rozdziale klasyfikacja przyczyn nieokreśloności modeli ma zasadnicze znaczenie dla poprawnego wyboru strategii współdziałania badacza i Systemu Wspomagania Decyzji, w skład którego wchodzi podsystem prognozy.

W dysertacji wskazano przykłady szeregów czasowych, których wartości poszczególnych zmiennych w kolejnych punktach czasowych generowane były za pomocą wzorów matematycznych oraz takich, które przedstawiały rzeczywiste procesy złożone (z powszechnie dostępnych źródeł internetowych, z dziedziny elektroniki i fizyki). Dokonano prognozy wartości poszczególnych zmiennych tych szeregów stosując metodę, której podstawą jest WKG (podobnie jak w metodzie GMDH) drugiego stopnia, rozszerzony o wartości półwzględnych funkcji wrażliwości stopnia pierwszego i drugiego.

Za podstawę naszej metody uznajemy GMDH, należy jednak zwrócić uwagę, że wspólną częścią metody prognozy MAMC oraz GMDH jest wykorzystanie WKG, natomiast obie te metody różnią się w kilku aspektach:

a) W strategii budowy modeli, czyli sposobie, w jaki posługujemy się WKG. W metodzie GMDH na każdym kroku (w każdej warstwie) stopień wielomianu może rosnąć (znacznie komplikuje się wykorzystywany wielomian, tym samym obliczenia). Elementarne obliczenia dają podstawę twierdzić, że praktyczne wykorzystanie GMDH, bez stosowania pewnych sposobów redukcji tworzonych modeli jest ograniczone niskimi rzędami wielomianów zupełnych, co istotnie obniża wydajność tej metody w praktyce. MAMC, natomiast stosuje wielomiany jedynie drugiego stopnia, tworząc ciąg tzw. modeli cząstkowych,

które ze sobą **konkurują.** Na każdym kroku wykonujemy też takiego samego rodzaju ciąg obliczeń (nie komplikując modeli).

b) W metodzie GMDH zapamiętywane zostają obliczenia poszczególnych modeli cząstkowych, które wykorzystywane są w kolejnych warstwach, ponownie, wikłając obliczenia. Przyjęte w praktyce podejście do redukcji modeli GMDH polega na odrzucaniu na każdym etapie tworzenia modeli, ich nieudanych wersji (w myśl zasady, że „słabi rodzice nie mogą reprodukować silnego potomstwa", która, jak udowodniono w teorii AG, jest błędna). Dlatego wielowarstwowe GMDH nie gwarantuje syntezy nawet „dopuszczalnych" (w simonowskim sensie) modeli prognozy. Natomiast w naszej metodzie wybierany jest model najlepszy i z jego rezultatami przechodzimy do kolejnego kroku obliczeń (prognozy następnego punktu czasowego), jednocześnie modele, które zostały odrzucone w jednym kroku, mogą okazać się zadowalające w kroku kolejnym (nie dyskwalifikujemy modeli najsłabszych).

c) W metodzie GMDH brak sposobu oceny dokładności prognozy, z kolei my proponujemy ocenę tej dokładności na podstawie porównania wyników prognozy, odległości między uzyskiwanymi w poszczególnych konkurujących modelach cząstkowych rezultatami. Im bliższe sobie są wartości prognozowanej zmiennej, otrzymane przez wykorzystanie różnych modeli cząstkowych, tym mniejszy błąd prognozy dla danej zmiennej. Jest to pewnego rodzaju ocena dokładności (ufności) prognozy. W taki właśnie sposób oceniamy błędy dla wszystkich zmiennych. Następnie obliczamy odległości pomiędzy rezultatami każdego modelu, wykorzystując dwie metryki Euklidesa oraz Canberra.

d) Proponujemy zastosowanie funkcji wrażliwości, które dają w ręce badacza pewne narzędzie matematyczne służące ocenie, które zmienne mają największy wpływ na stany badanego obiektu, natomiast w GMDH brak jest takiego narzędzia matematycznego, jednakże w celu oceny wpływu zmiennych na siebie można w niej dodatkowo stosować pewne metody statystyczne, co nie jest możliwe w naszej metodzie ze względu na niewielką ilość próbek danych uczących.

e) MAMC jest zorientowana na nieliczną ilość próbek, zaś GMDH działa dobrze, gdy mamy do dyspozycji duże zbiory próbek.

f) Stosowanie metody GMDH może dawać wyniki obarczone dowolnie dużym błędem od samego początku obliczeń, które mogą zostać wzmocnione na kolejnych krokach. Przyczyna tkwi w tym, że GMDH działa podobnie do

krzywika, który usiłujemy połączyć z istniejącymi punktami w przestrzeni zmiennych procesu, a nie uwzględnia dynamiki zmian tych zmiennych. Metoda ta nie przewiduje rozdzielenia procesu na części składowe, a działa na całych zbiorach zmiennych. Rozwiązanie odwrotnego problemu sterowania procesami polega na pewnym wydzieleniu składowych części złożonego procesu i ocenie wkładu każdej z tych części na ogólny stan procesu, co można uzyskać np. za pomocą funkcji wrażliwości. MAMC, z racji wykorzystania funkcji wrażliwości, uwzględnia dynamikę zmian procesu. Każdy z modeli cząstkowych może dać inny rezultat prognozy.

Badania przeprowadzono na dużej ilości szeregów z różnych dziedzin. Na podstawie tych badań stwierdzamy, że za korzyści naszej metody można uznać to, że:

a) MAMC działa skutecznie w przypadku procesów wykazujących regularności, które udaje się uchwycić na niewielkiej ilości próbek danych,
b) Z reguły działa skuteczniej od metody GMDH, która daje nierzadko bardzo odległe od oczekiwanych rezultaty, a w przypadku, gdy wyniki są bliskie rzeczywistym, dla metody GMDH nie ma metody oceny dokładności rozwiązań poszczególnych modeli
c) Metoda posiada narzędzie oceny jakości modeli i dokładności obliczanych rezultatów prognozy ex ante. Za jego pomocą (średnica – odległość między skrajnymi wartościami uzyskiwanymi przez poszczególne modele) możemy wskazać zmienną, do której predykcji mamy największe zaufanie, a za pomocą odległości Euklidesa i Canberra wskazujemy model najbardziej zbliżony do rzeczywistej wartości. To kryterium wskazuje poprawnie najbliższy rzeczywistej wartości rezultat nawet w przypadku niezadowalającej dokładności prognozy (szeregi ze znacznymi nieregularnościami). Bardzo dokładnie wskazuje też model najbardziej odległy (np. najbardziej wrażliwy na zmiany wartości argumentów).
d) MAMC funkcjonuje (działa poprawnie) na niewielkiej liczbie danych inicjujących (minimum 7).
e) Metoda działa na szeregach czasowych modelujących systemy złożone o wielu zmiennych.
f) Metodę można wykorzystać przedstawiając badany problem w postaci równań stanów w przyrostach.

g) Metoda posiada korzystny współczynnik $\frac{ilość\ probek\ uczących}{ilość\ udanych\ kroków\ prognozy}$.

h) MAMC nie musi być stosowana samodzielnie, można ją stosować jako uzupełniającą w innych metodach predykcyjnych, bądź tworząc hybrydy z innymi metodami predykcyjnymi.

i) Naszą metodę można stosować w analizie, predykcji szeregów czasowych z różnych dziedzin nauki (fizyka, chemia).

Za ograniczenia metody można uznać to, że:

a) MAMC nie działa na dowolnych szeregach czasowych ze względu na konieczność rozwiązywania układu równań liniowych (złe uwarunkowanie macierzy), w takim przypadku przesuwamy zakres badanych danych o wymaganą liczbę kroków i próbujemy rozwiązać problem prognozy ponownie.

b) Istnieje konieczność rozwiązywania układów równań nieliniowych, dla precyzji rozwiązań których, ważne jest wskazanie właściwego punktu początkowego obliczeń (tym samych, w niektórych przypadkach obliczenia okazują się skomplikowane lub obarczone błędem).

c) Metoda nie działa zadowalająco na nieregularnych zbiorach danych, choć mimo to, daje często bardziej precyzyjne rozwiązania od uproszczonej metody GMDH i popularnych metod ekstrapolacyjnych.

d) Czynnikiem decydującym, który zasadniczo wpływa na jakość tworzonych modeli i odpowiednio jakość rozwiązania problemów z wykorzystaniem tych modeli (dokładność rozwiązania problemu i szybkość znalezienia tego rozwiązania), jest należyte dobranie danych (czyli odpowiedniość i pełność zbioru dobieranych danych, które są w danym momencie wymagane dla poprawnego skonstruowania modelu symulowanego obiektu).

Osiągnięcia naukowe rozprawy:

1. Opracowanie metody analizy, predykcji systemów złożonych, która działa na krótkich próbkach danych,
2. Wykorzystanie w metodzie predykcji funkcji wrażliwości, które dają w ręce badacza narzędzie oceny dynamiki zmian procesu, wpływu poszczególnych zmiennych na siebie w zastępstwie narzędzi statystycznych, które mogą dawać przekłamane wyniki przy tak niewielkiej ilości próbek.

3. Opracowanie metody predykcji, która jest skuteczniejsza od metody podstawowej, na której jest oparta (GMDH),
4. Opracowanie metody deterministycznej, w której opieramy się na pewnych kryteriach, które są formalizowane w taki sposób, że przerzucają olbrzymią część obliczeń na komputer, który pozbawiony jest możliwości subiektywnych ocen, tak charakterystycznych dla ekspertów-ludzi i zdolny jest do optymalizacji tworzonych modeli, nawet na podstawie mało licznych próbek.

Perspektywy dalszych badań nad rozważanym problemem są optymistyczne, powinny one iść w kierunku:

1. Bardziej szczegółowej oceny dziedziny szeregów czasowych, na których metoda jest pomocna w analizie.
2. Opracowania dalszych analiz i prac rozwojowych idących w kierunku usprawnienia wyboru właściwych danych inicjujących.
3. Wykorzystania metody jako uzupełnienia innych metod predykcyjnych (element metody hybrydowej).

Spis ilustracji

Spis tabel

Bibliografia

[1] D. Graupe, Identification of systems, 2 ed. ed., Malabar: Krieger publishing company, 1976.

[2] L. Schimansky-Geier, Analysis and control of complex nonlinear processes in physics, chemistry and biology, Singapore; Hackensack, NJ: World Scientific, 2007.

[3] P. Dittman and Z. Hellwiga, Teoria prognozy z zastosowaniami ekonomicznymi, Wrocław, 1977.

[4] N. Wiener, Cybernetics or control and communication in the animal and the machine, Cambridge: MIT Press, 1961.

[5] A. Wiliński, GMDH - metoda grupowania argumentów w zadaniach zautomatyzowanej predykcji zachowań rynków finansowych, Warszawa: Instytut Badań Systemowych PAN, 2009.

[6] M. Mesarowicz and Y. Takachara, Teoria systemów złożonych, London, NY, San Francisco: Academic Press, 1975.

[7] V. Rogoza and J. Bobkowska, "Analiza zagadnienia prognostyki, jako informatycznego problemu eksploracji danych," *Metody Informatyki Stosowanej,* p. 107–120, 2011.

[8] A. Ivakhnenko and J. Müller, Selbstorganisation von vorhersagemodellen, Berlin: Verlag Technik, 1984.

[9] Z. Y. Tsypkin, Adaptation and Learning in Automatic Systems, New York: Academic Press, 1971.

[10] H. Holland, Adaptation in Natural and Artificial Systems, London: A Bradford Book, The MIT Press Cambridge, Massachusetts, 1992.

[11] A. A. Freitas, Data mining and knowledge discovery with evolutionary algorithms, Berlin, New York: Springer, 2002.

[12] A. Taudes, Adaptive information systems and modelling in economics and management science, Wien; New York: Springer, 2005.

[13] N. Moiseew, Matematyczne problemy analizy systemowej, Moskwa: Nauka, 1981.

[14] "The 2012 ACM Computing Classification System http://www.acm.org/about/class/class/2012 [," [Online]. [Accessed dostęp: 2013].

[15] P. Fryer, "A brief description of Complex Adaptive Systems and Complexity Theory," [Online]. Available: http://www.trojanmice.com/articles/complexadaptivesystems.htm. [Accessed 24 1 2010].

[16] E. Ahmed, A. Elgazzar and A. Hegazi, "An overview of complex adaptive systems,"

[Online]. Available: http://arxiv.org/abs/nlin/0506059. [Accessed 15 6 2012].

[17] J. Hopcroft, R. Motwani and J. Ullman, Introduction to Automata Theory, Languages, and, Addison-Wesley Publ. Comp, 2001.

[18] W. R. Ashby, Design for a Brain. The Origin of Adaptive Behaviour, London: Chapman&Hall Ltd, 1960.

[19] K. Gödel, "Über formal unentscheidbare Sätze der Principia Mathematica und verwandter Systeme I.," *Monatshefte für Mathematik und Physik, 38,* p. 173 – 198., 1931.

[20] A. Cobham, "The intrinsic computational difficulty of functions.," in *Proceedings of the 1964 International Congress for Logic, Methodology, and Philosophy of Science*, Jerusalem, 1964.

[21] J. Edmonds, "Paths,trees and flowers," *Canadian Journal of Mathematics*, p. 449 – 467., 1965.

[22] S. Cook, "The complexity of theorem-proving procedures," *In Proceedings of the 3th Annual ACM Symposium on Theory of Computing*, p. 151 – 158, 1971.

[23] R. Karp, "Reducibility among combinatorial problems," *Complexity of Computer Computations*, p. 85 – 103, 1971.

[24] J. Russell and P. Norvig, Artificial Intelligence. A Modern Approach, Prentice Hall, 2003.

[25] F. Luger, Artificial Intelligence, Addison-Wesley Publ. Comp., 2002.

[26] M. Bieńkowski, "Obliczenia na kwantach," *Wiedza i życie*, no. 3, pp. 16-23, 2014.

[27] A. Newell and H. A. Simon, "The logic theory machine: A Complex Information Processing System," *Information Theory, IRE Transactions on*, vol. 2, no. 3, pp. 61 - 79, 1956.

[28] A. G. Miller, "The Magical Number Seven, Plus or Minus Two: Some Limits on Our Capacity for Processing Information," *The Psychological Review*, vol. 63, pp. 81-97, 1956.

[29] F. Baader, D. Calvanese, D. L. McGuinness, D. Nardi, P. F. Pater-Schneider and F. Peter, "The Description Logic Handbook. Theory, Implementation, and Application," p. 556, 2003.

[30] J. R. Koza, M. A. Kean and M. J. e. a. Streeter, Genetic Programming IV. Routine Human-Competitive Machine Intelligence, Springer, 2005.

[31] L. S. Pontriagin, W. G. Boltianski, R. W. Gamkrelidze and E. F. Miszczenko, Matematyczna teoria procesów optymalnych., Fizmatgiz, 1961.

[32] R. Bellman, "Dynamic programming, system identification and suboptimization," *Journal of the SIAM on Control*, vol. 4, no. 1, 1966.

[33] F. P. Ramsey, Truth and probability, H. B. Jovanovich, Ed., New York: The Foundations of Mathematics and Other Logical Essays, 1931.

[34] J. Neumann von and O. Morgenstern, The Theory of Games and Economic Behavior, Princeton, New Jersy: Princeton University Press, 1944.

[35] M. (. Klusch, Intelligent Information Agents. Agent-Based Information Discovery and Management on the Internet, Berlin Heidelberg: Springer-Verlag, 1999.

[36] H. A. Simon, Administrative behavior, New York: Macmillan, 1947.

[37] A. Turing, "Computing machinery and intelligence.," *Mind,* no. 59, p. 433–460, 1950.

[38] J. Zieliński, Inteligentne systemy w zarządzaniu. Teoria i praktyka, Warszawa: PWN, 2000.

[39] *Encyklopedia of Data Warehousing and Mining,* Idea Group Inc, 2000.

[40] *Encyklopedia komputerowa Microsoft,* Warszawa: MIKOM, 2002.

[41] W. H. Immon, Building the Data Warehouse, Wiley Publishing, Inc., 2005.

[42] A. A. Barsegian, M. S. Kuprijanow and et al., Analiza danych i procesów, Sankt-Petersburg: BHW-Petersburg, 2009.

[43] D. Larose, Metody i modele eksploracji danych, Warszawa: Wydawnictwo Naukowe PWN, 2008.

[44] S. Haykin, Neural Networks. A Comprehensive Foundation, Prentice Hall, 1998.

[45] T. D. Gwiazda, Algorytmy genetyczne. Kompendium, tom 1: Operator krzyżowania dla problemów numerycznych, tom 2: Operator mutacji dla problemów numerycznych, MIKOM, 2007.

[46] D. J. Montana, "Neural network weight selection using genetic algorithms," in *Intelligent Hybrid Systems*, 1995.

[47] V. Rogoza and A. Ishchenko, "Complicated systems: fuzziness and adaptation," in *Proceedings on the Advanced Computer Systems*, Międzyzdroje, 2002.

[48] L. A. Zadeh, "Fuzzy sets," in *Information and control,8,* 1965.

[49] L. A. Zadeh, Fuzzy sets as a basis for a theory of possibility, Berkeley: Electronics Research Laboratory, College of Engineering, University of California, 1977.

[50] A. Ivakhnenko, J. P. Zajczenko and W. D. Dimitrow, Podejmowanie decyzji na zasadzie samoorganizacji, Moskwa: Sov. Radio, 1976.

[51] D. Hand, H. Mannila and P. Smyth, Eksploracja danych, Warszawa: WNT, 2005.

[52] P. Wróblewski, Algorytmy, struktury danych i techniki programowania, Helion, 2010.

[53] H. A. Taha, Operations Research. An Introduction, Prentice Hall, 1997.

[54] A. N. Tichonow and W. J. Arsenin, Metody rozwiązywania niepoprawnych problemów, Moskwa: NAUKA, 1974.

[55] F. Girosi, M. Jones and T. Poggio, Priors, stabilizers and basis funcrions: from regularization to radial, tensor and additive splines, Artificial Intelligence Laboratory, MIT, 1993.

[56] L. Cao, S. Y. Philip, C. Zhang and H. Zhang, Data Mining for Business Application, Springer Science; Business Media, 2008.

[57] D. Pyle, Business Modeling and Data Mining, Morgan Kaufmann Publishers, 2003.

[58] P. Giudici, Applied Data Mining: Statistical Methods for Business and Industry, Wiley Publishing, Inc., 2003.

[59] M. Kantardzic, Data Mining: Concepts, Models, Methods, and Algorithms, Wiley Publishing, Inc., 2003.

[60] G. Fabrice and H. (. Hamilton, Quality Measures in Data Mining, Berlin,Heidelberg: Springer-Verlag, 2007.

[61] H. A. Abbas, R. A. Sarker and C. S. Newton, Data Mining: A Heuristic Approach, Australia: University of New South Wales, , Idea Group Publishing, 2002.

[62] M. Berthold and D. (. Hand , Intelligenbt Data Analysis, Berlin,Heidelberg: Springer-Verlag, 2007.

[63] P. M. Derusso, R. J. Roy and C. M. Clowe, State Variables for Engineers, John Wiley & Sons, Inc., 1978.

[64] V. Hutson and J. Pym, Applications of Functional Analysis and Operator Theory, Academic Press, 1980.

[65] I. Bronstein, K. Siemiendiajew, G. Musiol and Mühlig, Nowoczesne Kompendium Matematyki, Warszawa: PWN, 2007.

[66] A. Ivakhnenko, Prognostyka długoterminowa i sterowanie systemów złożonych, Kijów: Technika, 1975.

[67] D. Gabor, Perspektywy planowania, Automatyka, 1972.

[68] D. E. Goldberg, Algorytmy genetyczne i ich zastosowania, Warszawa: WNT, 2003.

[69] M. Fisz, Rachunek prawdopodobieństwa i statystyka matematyczna, Warszawa: PWN, 1969.

[70] V. Volterra, Theory of functionals and of integral and integro-differential equations, Dover, New York, 1959.

[71] H. W. Bode , Network Analysis and Feedback Amplifier Design, Prinston: D. Van Nostrand, 1945.

[72] K. J. Åström and N. A. Kheir, Control Sytems Engineering Education, Vols. Vol. 32, No. 2, Pergamon: Automatica, 1996, pp. 147-166.

[73] Ł. Rauch, J. Talar, T. Zak and J. Kusiak, Filtering of thermomagnetic data curve using artificial neural network and wavelet analysis, Proc. 7th ICAISC 2004, Conf. ed., Zakopane: Zakopane, 2004, p. 1093 – 1098.

[74] G. Box and G. Jenkins, Time series analysis: Forecasting and control, San Francisco: Holden-Day, 1970.

[75] E. Kayacan, B. Ulutas and O. Kaynak, *Grey system theory-based models in time series prediction. Expert Syst. Appl., 37,* pp. 1784-1789, 2010.

[76] D. Ju-Deng, "Control problems of grey system.," *Syst. Control Lett., 1,* pp. 288-294., 1982.

[77] G. Li and T. Wang, *A new model for information fusion based on grey theory. Inform. Technol. J., 10:,* pp. 189-194, 2011.

[78] C. Cortes and V. Vapnik, "Support-Vector Networks,Machine Learning," 1995. [Online]. Available: http://www.springerlink.com/content/k238jx04hm87j80g/.

Publikacje własne

[79] V. Rogoza, J. Bobkowska, „Metodologiczne aspekty przewidywania zachowania złożonych systemów z wykorzystaniem metod symulacji indukcyjnej", Metody Informatyki Sosowanej, nr 1, 2010 (22), s. 143-161

[80] J. Bobkowska, V. Rogoza, „Nowoczesne metody ewolucyjne w data mining oraz knowledge discovery", Zeszyty Naukowe Stargardinum

[81] V. Rogoza, J. Bobkowska, „Analiza zagadnienia prognostyki, jako informatycznego problemu eksploracji danych", Metody Informatyki Sosowanej, nr 1, 2011 (26) s. 107-119

[82] V. Rogoza, J. Bobkowska, "Using sensitivity functions to simulation of complex processes", Przegląd elektrotechniczny, nr 10b, 2012, s. 188-191

[83] J. Bobkowska, „The comparison of the results of time series prognosis obtained with the classical GMDH algorithm and the modified method containing sensitivity functions", PAK, nr 7, 2013, s. 688-691

[84] V. Rogoza, J. Bobkowska, "Analiza zachowania się obiektów, ich kontrola i prognoza w przestrzeni stanów", Przegląd elektrotechniczny, nr 3, 2014, s. 124-127

[85] J. Bobkowska, „Wady i zalety metody prognozowania zachowania złożonych procesów opartej na metodzie GMDH z funkcjami wrażliwości ", PAK, nr 12, 2014, s. 1136-1139