

Recenzja rozprawy doktorskiej
mgr. inż. Karola Nienałtowskiego

**Parametric and non-parametric methods
to address complexity of cellular signaling pathways**

złożonej do Rady Naukowej Instytutu Podstawowych Problemów Techniki
Polskiej Akademii Nauk.

1. Struktura i cele rozprawy

Rozprawa jest skonstruowana dosyć nietypowo. Zawiera ona pięć jakby zwykłych rozdziałów oraz trzy wieloautorskie artykuły:

- K. Nienałtowski, T. Jetka, M. Komorowski (2018) Sensitivity analysis. Rozdział w redagowanej książce: B. Munsky, L. S. Tsimring, W. S. Hlavacek (eds.) *Quantitative Biology: Theory, Computational Methods, and Models*, wydanej przez MIT Press, Cambridge, MA, 293-320 (wydawnictwo zaliczone do poziomu II, za 200 pkt. w wykazie MNiSW) – artykuł (1);
- K. Nienałtowski, M. Włodarczyk, T. Lipniacki, M. Komorowski (2015) Clustering reveals limits of parameter identifiability in multi-parameter models of biochemical dynamics. *BMC Systems Biology* 9:65 (czasopismo ocenione na 70 pkt. w wykazie MEiN) – artykuł (2);
- K. Nienałtowski, R.E. Rigby, J. Walczak, K.E. Zakrzewska, E. Głów, J. Rehwinkel, M. Komorowski (2021) Fractional response analysis reveals logarithmic cytokine responses in cellular populations. *Nature Communications*, 12:4175 (czasopismo ocenione na 200 pkt. w wykazie MEiN) – artykuł (3).

Prócz tego podano odnośniki do dwóch pakietów oprogramowania napisanego w języku R, opracowanych w celu wykonania obliczeń zamieszczonych w rozprawie.

W rozdziale 2, str. 37, rozprawy Doktorant stwierdza, że ogólnym celem rozprawy jest opracowanie ilościowych metod, które mogą sobie poradzić ze złożonością biochemicznych szlaków sygnałowych. Doktorant formułuje także trzy cele szczegółowe (przedstawione poniżej w moim tłumaczeniu):

- rozpoznanie możliwości zastosowania analizy wrażliwości parametrów w badaniach szlaków sygnałowych;
- wykorzystanie analizy identyfikowalności do badania wieloparametrowych modeli szlaków sygnałowych;
- rozwój metod nieparametrycznych dostosowanych do skomplikowanych procesów zachodzących w szlakach sygnałowych.

Odpowiadają one wprost trzem powyżej wymienionym publikacjom. Tematyka rozprawy bardzo dobrze mieści się w dyscyplinie *inżynieria biomedyczna*.

Zgodnie z art. 187, punkt 3 ustawy z dnia 20 lipca 2018 r. *Prawo o szkolnictwie wyższym i nauce* Dz. U. 2018 poz. 1668 rozprawę doktorską może stanowić praca pisemna, w tym monografia naukowa, zbiór opublikowanych i powiązanych tematycznie artykułów naukowych, praca projektowa, konstrukcyjna, technologiczna, wdrożeniowa lub artystyczna, a także samodzielna i wyodrębniona część pracy zbiorowej. Biorąc pod uwagę wcześniejszą analizę, przyjmuję, że rozprawę stanowi przede wszystkim zbiór opublikowanych trzech artykułów, chociaż Doktorant potraktował je w opracowanym opracowaniu nieco po macoszemu, drukując je w zminiaturowanym, trudnym do czytania formacie, bez dołączenia istotnych suplementów, do których są liczne powołania w zasadniczych tekstach. Pozostałą część rozprawy, z typowymi rozdziałami, traktuję jako omówienie szerszego tła omawianych prac oraz objaśnienia i uzupełnienia do zasadniczych artykułów.

Doktorant jest pierwszym autorem we wszystkich trzech artykułach. W rozprawie jest dodatek pod mylącym tytułem „Declarations of Author Contribution”, ale zawiera on listę współautorów w publikacjach Doktoranta, nie tylko trzech włączonych do rozprawy, ale także wymienionych pięciu innych, opublikowanych w znanych czasopismach (*Nature Communications*, *BMC Systems Biology*, *PLoS Computational Biology*), innym rozdziale wspomnianej książki wydanej przez MIT Press oraz w materiałach European Control Conference. Te dodatkowo wymienione publikacje są bliskie tematyce rozprawy i Doktorant figuruje w trzech z nich jako drugi autor. Jeżeli natomiast chodzi o wkład Doktoranta w publikacje włączone do rozprawy, to Doktorant pisze o tym w końcowej części wprowadzenia na str. 21, w rozdziale 2 na str. 37, a także we wnioskach na str. 70, w wierszach 16-19 od góry, czyli skrótowo 70¹⁶⁻¹⁹, którego to zapisu będę używał też w dalszym ciągu, przy czym strona z liczbą w indeksie dolnym będzie oznaczała wiersz od dołu. Z tych wymienionych opisów wynika, że Doktorant wniósł we wszystkich pracach decydujący wkład. Niemniej jednak spodziewam się, że w prezentacji wyników w trakcie obrony rozprawy Doktorant określi ten wkład bardziej precyzyjnie.

Typowe rozdziały rozprawy obejmują: wprowadzenie, cele rozprawy, metody, cele (objectives) – zawierający komentarze do artykułów i zakończenie.

2. Zakres rozprawy i uwagi o terminologii

Rozprawa jest związana z budową modeli biochemicznych szlaków sygnałowych. Modele te odzwierciedlają sieć komunikacyjną między komórkami. Sieć ta ma jednak specyficzną postać. Przenoszenie sygnałów odbywa się w niej przez uwalnianie cząsteczek sygnałowych do przestrzeni międzykomórkowej przez jedne komórki, a następnie wiązanie ich przez inne komórki. Ten sposób przenoszenia wiąże się z losowością procesów uwalniania, przepływu i wiązania cząsteczek. Opisowi działania tej sieci jest poświęcony punkt 1.1 wprowadzenia. Nie jestem specjalistą od biochemicznych szlaków sygnałowych oraz związanych z nimi modeli i moje uwagi koncentrują się bardziej na metodologii, a w szczególności odniesieniach do metodologii budowy modeli z użyciem danych empirycznych przyjętej w badaniach systemowych.

W rozprawie uwagi o metodologii budowy modeli koncentrują się głównie we wprowadzeniu. I tak, w punkcie 1.1 doktorant przedstawia trudności występujące przy modelowaniu sieci szlaków sygnałowych oraz wprowadza kilka modeli zaproponowanych wcześniej w literaturze. Omawiając modele parametryczne stwierdza (20₂₃₋₂₁), że „narzędzia (tools)” – które na

podstawie przedstawionych przykładów (identyfikowalność, analiza wrażliwości) nazwałbym raczej „pojęciami” – opracowane w ramach nauk inżynierskich oraz przez „statystykę teoretyczną” są „niezbyt przydatne w bardziej złożonych modelach biologii ilościowej”. I rzeczywiście, w spisie literatury zupełnie nie ma pozycji związanych z teorią estymacji statystycznej, czy z ogólną teorią i praktyką modelowania systemów rozwijanych nie tylko w ramach nauk inżynierskich, ale także i innych, jak na przykład w fizyce. W takiej literaturze są zawarte informacje i pojęcia o charakterze ogólnym, co według mnie jest przydatne, nawet jeżeli dyskutowane w rozprawie typy modeli mają szczególny charakter wyróżniający je wśród większości systemów rozpatrywanych dotąd w teorii systemów. Są też publikacje bliżej związane z tematyką rozprawy, jak na przykład książka Waltera i Contrerasa.¹ Natomiast w spisie literatury jest dużo pozycji, także książkowych, poświęconych budowie modeli biologicznych podobnego typu jak rozpatrywany w pracy.

To „zerwanie” z podstawowym nurtem zapewne spowodowało, że częściowo rozminęła się też terminologia w obu nurtach. Zanim przejdę do omówienia szczegółów, zacznę od podania kilku dosyć ogólnie przyjętych podstawowych wiadomości na temat modelu, aby wprowadzić nazewnictwo, którym się będę posługiwał. Mam też nadzieję, że znajomość tych różnic terminologicznych przyda się Doktorantowi w dalszych pracach,

Model matematyczny, bo o takich modelach mówimy, opisuje system przy użyciu pojęć matematycznych i matematycznego języka. Przedstawia on zależności między pewnymi wydzielonymi zmiennymi, ciągami lub funkcjami (ewentualnie zmiennymi losowymi lub procesami stochastycznymi), zwanymi sygnałami wyjściowymi czy zależnymi, które są reakcjami na sygnały wymuszające aktywność systemu, zwane sygnałami wejściowymi lub niezależnymi. Prócz tego na sygnały wyjściowe mogą wpływać (pośrednio lub bezpośrednio) nieznanne (nieobserwowane, niemierzone) sygnały zwane zakłóceniami.

Model jest zazwyczaj tworzony w pewnym zamierzonym celu. Wspomnę o kilku najbardziej typowych. A więc może to być wyznaczenie (predykcja) sygnałów wyjściowych rzeczywistego systemu, gdy są ustalone sygnały wejściowe. Do tego celu wystarczy opracowanie dostatecznie dokładnej (satisfakcjonującej budującego model) zależności między sygnałami wyjściowymi i wejściowymi, w postaci deterministycznej lub probabilistycznej. Matematyczny opis tej zależności nie musi odpowiadać faktycznemu działaniu systemu. Dobrym przykładem takich modeli są tak zwane „modele neuronowe”, które mają strukturę sieci powiązanych między sobą prostych elementów opisanych modelami wzorowanymi na działaniu neuronów. Podobne modele można stosować do sterowania, jednak wtedy musi występować zależność przyczynowa sygnałów wyjściowych od wejściowych, nie wystarczy w nich zależność korelacyjna, jak w przypadku modeli predykcyjnych. Modele do sterowania wymagają więc większej wiedzy o systemie. Powyższe modele są często zwane modelami wejściowo-wyjściowymi, gdyż nie wymagają znajomości struktury rzeczywistego systemu.

Większej wiedzy o strukturze systemu wymagają modele do wyznaczania nieobserwowalnych sygnałów wewnętrznych (stanów) systemu. Pełnej wiedzy o strukturze wymagają natomiast modele opisujące działanie fizyczne systemu zwane często właśnie modelami opisowymi, ale także modelami fenomenologicznymi lub dedukcyjnymi. Modele te mogą służyć na przykład do analizy działania systemu, roli wewnętrznych składowych systemu, czy symulacji jego działania przy przyjętych sygnałach wejściowych i ewentualnie warunkach początkowych, ale

¹ G.G. Walter, M. Contreras: *Compartmental Modelling with Networks*. Birkhäuser 1999.

także mogą służyć do wyznaczania nieznanymi parametrów lub niemierzonych sygnałów wewnętrznych systemu na podstawie obserwacji sygnałów wejściowych i wyjściowych. Modele parametryczne dyskutowane w rozdziale 1 należą do klasy modeli opisowych.

Modele opisowe buduje się na podstawie znajomości praw fizycznych, chemicznych itp. Z reguły występują w nich parametry, których wartość może być znana z wcześniejszych badań. Ale mogą być wśród nich parametry nieznanne. Wartości takich parametrów trzeba ustalić, typowo na podstawie pomiarów sygnałów wejściowych i wyjściowych systemu obserwowanych w trakcie eksperymentów, czynnych lub biernych. Wtedy do budowy modeli jest użyta wiedza empiryczna. Jednak takie modele, jak każde inne, powinny być dodatkowo sprawdzone, głównie pod kątem dokładności rekonstrukcji sygnałów wyjściowych dla niezależnych danych, czyli walidowane. Jeżeli walidacja nie jest zadowalająca, to model powinno się poprawić, na przykład przez zmodyfikowanie struktury modelu. Badania modelu mogą być także prowadzone w kontekście wyznaczenia obszaru stosowalności modelu.

Wracając teraz do dyskusji treści rozdziału 1, w drugim akapicie na stronie 22 modele są podzielone na opisowe i predykcyjne, przy czym te pierwsze „pozwalają na porównanie wiedzy apriorycznej z rzeczywistym zachowaniem się komórek”. Wydaje się, że w tym pierwszym przypadku chodzi o wcześniej omówioną klasę modeli fenomenologicznych (opisowych). Jednak co to są „modele predykcyjne”? Według zapisu w rozprawie „mogą one być użyte do określenia różnych, często wzajemnie rozłącznych hipotez i przebadania powstałych zachowań”. To wygląda właśnie na badania służące do analizy i wyznaczenia obszaru stosowalności modelu, o czym wspomniałem powyżej. Do takich badań stosuje się w zasadzie modele opisowe, gdyż w przypadku modeli opartych tylko na wiedzy empirycznej może to oznaczać ekstrapolację poza obszar, w którym model był dopasowany do obserwacji, co jest zawsze bardzo ryzykowne. Tak więc, chociaż w pewnym kontekście mówi się o wartościach wyjść modeli „predykcje”, a modele opisowe mogą być także używane do predykcji, to nazwa „modele predykcyjne” kojarzy się jednak ze specjalistycznymi prostszymi modelami, o których napisałem powyżej. Nie wydaje mi się, Doktorant miał na myśli takie uproszczone modele. Czym wobec tego różnią się te „modele predykcyjne” od modeli opisowych?

W trzecim akapicie są wymienione etapy tworzenia modelu: (i) konstrukcja, (ii) weryfikacja, przy czym „na etapie weryfikacji sprawdzamy, czy zachowanie modelu jest zgodne z wiedzą teoretyczną” (iii) kalibracja, (iv) walidacja. Nie rozumiem, na czym polega etap weryfikacji. Czy oznacza to, że przy konstrukcji modelu nie posługiwaliśmy się właśnie wiedzą teoretyczną? Czy też może chodzi o to, że pierwszy etap wykonuje ktoś słabiej kwalifikowany, na przykład student, i w drugim etapie musi to sprawdzić osoba lepiej znająca teorię? Mam też trochę zastrzeżeń do samego słowa „weryfikacja”, które oznacza, bardziej nawet w języku angielskim niż w polskim, rozstrzygnięcie prawdziwości. A tymczasem nie ma gwarancji, że nawet najlepsza obecna wiedza o modelowanym zjawisku czy procesie jest rzeczywiście prawdziwa, gdyż właściwie nasza wiedza jest tylko modelem mentalnym, a więc z założenia przybliżeniem rzeczywistości. Na podstawie takiej wiedzy nie da się rozstrzygnąć prawdziwości modelu, można najwyżej sprawdzić jego zgodność z obecną wiedzą.

Nieporozumienia terminologiczne mogą się pojawić w związku z wprowadzonymi w rozprawie pojęciami nieidentyfikowalności (non-identifiability lub unidentifiability) i obserwowalności (observability). Doktorant rozróżnia za [126] „nieidentyfikowalność strukturalną” i „nieidentyfikowalność praktyczną”. Nie jest to zbyt wygodny podział. Jest sporo definicji

identyfikowalności i funkcjonują różne nazwy, między innymi używa się nazwy „nieidentyfikowalność strukturalna”. W znacznej części literatury związanej z modelowaniem używa się też może zrzeczniejszego terminu „nieidentyfikowalność parametryczna”. Oba te terminy oznaczają to samo, a mianowicie że nie jest możliwe jednoznaczne wyznaczenie nieidentyfikowanych parametrów nawet przy najbardziej sprzyjających warunkach pomiarowych (nieograniczonych pomiarach, braku błędów pomiarowych, dowolnych sygnałach wejściowych). Jest to właściwość modelu, ale nieidentyfikowalność parametryczna odnosi się też do konkretnych parametrów w tym sensie, że niektóre parametry mogą być identyfikowalne, a inne nie. Podany w (26²²⁻²⁴) – przypadek (1) – przykład zależności funkcyjnej między parametrami jest prostym przykładem nieidentyfikowalności parametrycznej, jeżeli wyobrazimy sobie, że ta funkcja jest zbiorczym parametrem modelu, na przykład model zależy od iloczynu dwóch parametrów ab , bo wtedy zmiana jednego parametru może być skompensowana odpowiednią zmianą drugiego i nie da się jednoznacznie ustalić ich wartości. Współliniowość (collinearity) zmiennych niezależnych jest terminem używanym do określenia liniowej zależności między wektorami pomiarów tych zmiennymi. Wystąpienie takiego przypadku teoretycznie uniemożliwia estymację parametrów w statycznych modelach liniowych – regresji liniowej. Używam słowa „teoretycznie”, bo w przypadku zakłóceń pomiarowych ta liniowa zależność może nie być aż tak ewidentna. Nie spotkałem się z terminem „parameters collinearity” użytym w rozprawie (26²⁴) i niezbyt wiadomo, jak to rozumieć, bo parametry są skalarami. Dopiero po przeczytaniu dołączonego artykułu (2) w *BMC Systems Biology* można podejrzewać, że chodzi tu pewnie o liniową zależność kolumn macierzy informacyjnej Fishera. Może więc lepiej mówić o zależności parametrów (parameters dependence) lub powiązanych parametrach (correlated parameters), tak jak w artykule (2), chociaż to ostatnie może być mylone ze statystycznym znaczeniem słowa korelacja. Prócz tego, zależność funkcyjna jako taka, w formie dodatkowego ograniczenia równościowego w modelu, raczej poprawia identyfikowalność. Więc Doktorantowi chodzi raczej o zależność ukrytą, gdy jeden z parametrów modelu jest funkcją innych parametrów, o czym nie wiemy lub co przeoczyliśmy. W takiej sytuacji w przestrzeni parametrów minimum wskaźnika występuje na podprzestrzeni. Taki model jest oczywiście nieidentyfikowalny.

Brak obserwowalności modelu oznacza, że część modelu nie wpływa na sygnały wyjściowe. Parametrów występujących w części nieobserwowalnej nie da się oczywiście estymować na podstawie pomiarów sygnałów wyjściowych. Objawem takiej nieidentyfikowalności będzie właśnie wspomniany w (26²⁴⁻²⁶) punkt (2) mówiący o niewrażliwości sygnału wyjściowego na zmiany parametrów. Doktorant pisze (26¹⁷⁻¹⁶), że systemy biologiczne są często tylko „partially observed” nie tłumacząc, co to znaczy. Po wytłumaczeniu trzeba sięgnąć do [126], gdzie autorzy rozumieją przez to, że nie wszystkie stany są obserwowane. Nie musi to powodować nieidentyfikowalności, bo nawet jeżeli część stanów nie jest obserwowana, to system może być mimo to obserwowalny. Warto też wspomnieć, że przy modelach wejściowo-wyjściowych jest ważna jeszcze jedna właściwość modelu, zwana sterowalnością. W przybliżeniu, brak sterowalności oznacza, że nie jesteśmy w stanie zmieniać dowolnie wartości wszystkich stanów modelu przez jakikolwiek wybór sygnałów wejściowych. Taki model jest nieidentyfikowalny przy zerowych warunkach początkowych, ale może być identyfikowalny, gdy w części niesterowalnej warunki początkowe są niezerowe. W ogólności, przed badaniem identyfikowalności powinno się więc też sprawdzić, czy model jest sterowalny. O ile sprawdzanie obserwowalności i sterowalności modeli liniowych jest dosyć łatwe, to niestety dla modeli nieliniowych jest to znacznie trudniejsze.

„Strukturalna identyfikowalność” (termin użyty też na przykład w 26²¹⁻²²) może się kojarzyć z właściwością, która w literaturze związanej z modelowaniem jest zwana rozróżnialnością (distinguishability²) struktur i z grubsza mówiąc polega na stwierdzeniu, czy dla dwóch różnych struktur modelu nie znajdują się takie ich parametry (inne dla każdej struktury), aby dla dowolnych sygnałów wejściowych sygnały wyjściowe dla obu struktur były takie same. Jeżeli taka sytuacja nie występuje, to modele są rozróżnialne. W rozpatrywanych w rozprawie modelach takie różne modele nierozróżnialne mogłyby się na przykład różnić drogami przesyłania sygnałów w sieci połączeń, w których można by było tak dobrać współczynniki (rate constants) w obu modelach, aby wielkości wyjściowe modeli były zawsze takie same. Nie wiem, czy taki przypadek może być rozważany w modelu z rozprawy, bo wtedy nierozróżnialność modeli mogłaby być dużym zagrożeniem dla uzyskania poprawnego modelu.

Wszystkie omawiane wcześniej właściwości teoretycznie wykluczają prawidłowe wyznaczenie modeli na podstawie nawet najlepiej zaplanowanych pomiarów o nieskończonej długości (z wyjątkiem tak zwanej nieidentyfikowalności lokalnej, o której wspomnę dalej). Uzdrawienie nieidentyfikowalności parametrycznej wymaga właściwie zmiany modelu, włączając w to na przykład niezależne ustalenie części parametrów czy regularyzację, przez co rozumiem przyjęcie dodatkowych warunków w procesie estymacji parametrów. Ale w przypadku wystąpienia w systemie zakłóceń może się udać uzyskanie jakichś ocen parametrów, nawet jeżeli model jest nieidentyfikowalny. Ich wartość będzie wtedy jednak mocno przypadkowa, mocno zależna od zakłóceń, a do tego będą się one charakteryzowały dużymi szacowanymi błędami. Wyjątkiem będzie nierozróżnialność, w której oceny parametrów mogą być nieprawidłowe, a tymczasem ich szacowane dokładności mogą być małe. Taka „zakłócona” nieidentyfikowalność parametryczna może być bardzo trudna do rozróżnienia od innych przyczyn „praktycznej nieidentyfikowalności”, jak na przykład spowodowanej zbyt małą liczbą pomiarów. A ponieważ zakłócenia występują dosyć powszechnie w pomiarach, to w rezultacie wydaje mi się, że podział na nieidentyfikowalność strukturalną i praktyczną dopiero wtedy będzie naprawdę użyteczne w praktyce modelowania, gdy wypracuje się test rozróżniający te dwa przypadki.

Kłopoty z estymacją parametrów mogą też wyniknąć z nieprawidłowo przeprowadzonego eksperymentu pomiarowego. Powinien on pozwolić na określenie prawidłowej zależności wejściowo-wyjściowej modelu przy dostatecznie dużej liczbie pomiarów, co w latach 1980-tych Lennart Ljung nazywał „identyfikowalnością systemową”³. Jest to związane z wymaganiami prawidłowego przeprowadzenia eksperymentu, czyli prawidłowego zebrania obserwacji (pomiarów) użytych później do estymacji parametrów. O tego typu warunkach wspomniano w rozdziale dopiero dalej, ale nie chciałbym tego tematu rozwijać, bo wymagałoby to nieco dłuższego wywodu, a w rozpatrywanym w rozprawie przypadku nie wydaje się to być specjalnym powodem występowania kłopotów.

Natomiast chciałbym poruszyć wspomniane w wierszach 29³⁻⁶ planowanie eksperymentu. W pracach poświęconych budowie modeli z udziałem danych eksperymentalnych przez planowanie eksperymentu rozumie się taki dobór sygnałów wejściowych, aby uzyskać jak najbardziej dokładne oceny parametrów. O planowaniu eksperymentu wspomina się w [126], ale tam model zawiera sygnały wejściowe, a zgodnie ze sformułowaniem modelu w rozprawie

² Termin wprowadzony przez zespół Keitha Godfrey'a w University of Warwick.

³ L. Ljung: *System Identification. Theory for the User*. Prentice Hall, Hatfordshire, 1989.

tych sygnałów w nim nie ma. Można oczywiście planować jak najlepszy dobór warunków początkowych, jeżeli jest możliwość ich ustalania, ale wydaje mi się, że prawdopodobnie rozwiązaniem byłyby jak największe możliwe wartości tych warunków. Natomiast w artykule (2) na stronach 7 (prawa szpalta) i 8 (lewa szpalta) proponuje się niezależne wyznaczanie niektórych parametrów w wydzielonym podmodelu nazywając to planowaniem eksperymentu. Niezależne ustalenie wartości niektórych parametrów jest typową strategią stosowaną w modelach wieloparametrowych, gdyż pozwala to na bardziej efektywne wyznaczanie pozostałych parametrów. Nie jest też nowy pomysł, aby estymować część parametrów w cząstkowych podmodelach, gdy można je odpowiednio wydzielić, na przykład gdy są one słabo połączone. Jednak nazywanie takich podejść planowaniem eksperymentu jest mylące.

3. Uwagi o charakterze ogólnym

W punkcie 1.2.1 wprowadzono model deterministyczny w postaci układu równań różniczkowych zwyczajnych. Tu właśnie, w pierwszym akapicie, używa się pojęć zmiennych zależnych, czyli wyjść modelu, oraz zmiennych niezależnych, podając jako przykłady tych ostatnich czas i przestrzeń. Także w 23¹⁶ twierdzi się, że jedyną zmienną niezależną w modelu jest czas. Jest to dosyć dziwne, bo w dyskutowanych dalej w tym punkcie modelach występują sygnały wyjściowe zmienne w czasie i parametry, które wydają się być stałymi liczbami rzeczywistymi, a w modelach nie ma bezpośredniej zależności od czasu. Jeżeli zaś chodzi o przestrzeń (chyba chodzi o zmienne przestrzenne?), to nie tylko ich nie ma, ale wtedy równania musiałyby być cząstkowe, a nie zwyczajne. Niezbyt też rozumiem sens stwierdzenia „a set of primary conditions. i.e., kinetics parameters and initial values of variables”. Chyba nie chodzi to o warunki początkowe równania (initial conditions), a więc co oznaczają (pierwotne?) warunki?

Wbrew wcześniejszym zapowiedziom, model (4) ma dosyć klasyczną postać tak zwanych równań stanu z niezerowymi warunkami początkowymi, bez sygnałów wejściowych. Jego sygnały wyjściowe ewoluują z punktu zadanego warunkami początkowymi do pewnego rozwiązania ustalonego. W dalszej części tego punktu (24¹²⁻¹⁴) Doktorant interpretuje zapis (4) jako model kompartmentowy, a wobec tego warunki początkowe można też zinterpretować jako wynik wymuszenia w postaci bolusa, w zapisie matematycznym przedstawianego w formie delty Diraca. I te wymuszenia można ewentualnie traktować jako sygnały wejściowe modelu. Niezależnie od interpretacji, taki typ modeli jest znany od dawna. Trudnością może być skomplikowana forma połączeń i duża wymiarowość modelu, o czym tutaj Doktorant nie pisze, ale jest to wspomniane w innych miejscach. Dodatkowym utrudnieniem mogą być nieliniowości, o których Doktorant wspomina w wierszu 24¹⁵. Nie jest też dla mnie oczywiste pojęcie kompartmentu w tych modelach. Czy są to pule komórek tego samego typu, a liczba tych komórek decyduje o objętości kompartmentu?

Dalej (23²⁻¹) jest stwierdzenie, że rozwiązaniem jest przebieg przejściowy, który zbiega do wartości stałej (chyba powinno być „stałych”, bo jest wiele wyjść), która jest rozwiązaniem równania (6) z zerowymi pochodnymi. Bez dodatkowych założeń to stwierdzenie jest nieprawdziwe. Po pierwsze, rozwiązania mogą być rozbieżne, o czym się nie wspomina. Jednak nawet wykluczając taki przypadek, to rozwiązaniami równania (6) z zerowymi pochodnymi są punkty równowagi (ekwilibria), które mogą być stabilne lub niestabilne, w tym sensie, że po wytrąceniu systemu z punktu równowagi przebieg rozwiązania wraca z powrotem do tego punktu lub nie. W przypadku pierwszym mówimy o asymptotycznej stabilności rozwiązania

w tym punkcie. W drugim przypadku rozwiązanie może zbiegać do innego punktu równowagi lub oscylować wokół punktu wyjściowego. Sądząc po kilkukrotnych wcześniejszych wzmiankach (18¹⁹⁻²¹, 22²³⁻²², 23¹⁵⁻¹³) o możliwości wystąpienia pętli w grafie połączeń, to takie rozwiązania oscylacyjne mogą wystąpić, a trudno o nich mówić, że są to rozwiązania stałe.

W równaniu (9) wprowadza się macierz współczynników stechiometrycznych. Ponieważ rozprawa dotyczy szlaków sygnałowych, co się kojarzy z przesyłaniem sygnałów, to przydałoby się dokładniejsze objaśnienie, jaki charakter mają zmienne zapisane w wektorze \mathbf{x} lub x (w zależności od wyboru wcześniejszych definicji), gdyż sygnały w ogólności przenoszą informację, a nie masę lub energię, więc nie muszą podlegać bilansowaniu.

Punkt 1.2.2 zaczyna się znowu wątpliwymi sformułowaniami. Modele ilościowe nie muszą być modelami parametrycznymi. Z modeli nieparametrycznych też można uzyskiwać dane ilościowe. Wbrew zapisanemu dalej stwierdzeniu modele biologiczne mogą być dosyć proste. Na przykład modele metabolizmów lub modele rozptyłu leków stosowane w farmakologii bywają opisywane równaniami różniczkowymi drugiego, a nawet pierwszego rzędu. Dalej (25₁₆) jest użyty termin „funkcja chi-kwadrat” i nie do końca wiadomo, co Doktorant przez to rozumie. Są utarte pojęcia zmiennej o rozkładzie chi-kwadrat i testu chi-kwadrat. Wydaje się, że w tym sformułowaniu chodzi o statystykę w postaci sumy kwadratów, która przy odpowiednich założeniach ma rozkład chi-kwadrat. Dla modelu (8), gdy zmienne losowe $\varepsilon(t)$ są białym szumem, statystyka taka jest sumą kwadratów różnic pomiarów i wyjścia modelu, zwana często funkcją strat. Niezależnie od niejasnego znaczenia $\chi^2(\theta)$, to dla tego, aby zmienna wyrażona wzorem (9) miała rozkład chi-kwadrat, jest potrzebne założenie o niezależności błędów pomiarowych w czasie, o czym w rozprawie nie wspomniano.

Szacowanie dokładności ocen modelu jest poruszone w dalszym ciągu punktu. Ten fragment korzysta miejscami z materiału z pozycji [126]. Nie brakuje tu jednak nieścisłych sformułowań. Wzór (10) miał być chyba wzorem (7) z [126] sądząc z bardzo podobnego jego opisu, ale przez obcięcie części oryginalnego wzoru nie przedstawia przedziału ufności. We wzorze (11) widać bezpośrednio, jak kłopotliwe jest oznaczenie $\chi^2(\theta)$, niezależnie od wspomnianej wcześniej niejasnego jej określenia. To samo oznaczenie jest użyte i jako funkcja wektora parametrów, i jako kwantyl rozkładu. Dalej (27¹⁵) $\chi^2(\theta)$ jest nazwane funkcją wiarygodności, co nie jest prawdziwe, a w wierszu (27¹⁸⁻¹⁹) ponownie mówi się o współliniowości parametrów, przy czym środkiem na zaradzenie „strukturalnej nieidentyfikowalności” ma być uzyskanie „najmniejszej współliniowości”, co jest całkowicie niezrozumiałe, bo współliniowość może występować lub nie. Interpretacji tego niezręcznego sformułowania, w sensie intuicyjnym, bo nie ścisłym, trzeba chyba szukać w artykule (2), gdzie jednym ze wskaźników jest norma macierzy wrażliwości, która powinna mieć dostatecznie dużą wartość, ale bez zauważenia tego związku trudno zrozumieć, o co chodzi.

W trakcie estymacji parametrów modeli opisanych liniowymi równaniami różniczkowymi często występuje więcej niż jedno minimum globalne i liczne minima lokalne. Na przykład, układ dwóch liniowych równań różniczkowych zwyczajnych o stałych współczynnikach może mieć rozwiązanie $y(t) = a_1 e^{b_1 t} + a_2 e^{b_2 t}$. Ze względu na symetrię, czy inaczej mówiąc dowolność nazwania parametrów w składnikach powyższej sumy, w idealnych warunkach i bez błędów pomiarowych wskaźnik jakości estymacji będzie miał dwa optima globalne w przestrzeni parametrów, to znaczy (a_1, a_2, b_1, b_2) i (a_2, a_1, b_2, b_1) . Przy wzroście rzędu modelu liczba takich symetrii może wzrastać. Taki model nie jest globalnie identyfikowalny

parametrycznie, ale jest on identyfikowalny lokalnie, to znaczy istnieją obszary w przestrzeni parametrów, w którym jest tylko jedno optimum globalne. W powyższym przypadku może to być na przykład obszar ograniczony nierównościami $b_1 < b_2$. Wielomodalność wskaźnika jakości powoduje też pojawianie się licznych punktów stacjonarnych funkcji, a wśród nich minimów lokalnych. Przykłady mnożących się punktów stacjonarnych ze wzrostem liczby parametrów były wykazywane już na początku lat 1980-tych. Jest to spore utrudnienie dla metod numerycznych, szczególnie przy większych zakłóceniach pomiarowych. O tych trudnościach Doktorant nie wspomina i zajmuje się badaniem punktu wyznaczonego w trakcie estymacji numerycznej, zakładając chyba milcząco, że jest to właściwy minimalny punkt globalny.

4. Uwagi szczegółowe

W odróżnieniu od artykułów, które są napisane w sposób precyzyjny, rozdziały rozprawy są przygotowane niezbyt starannie. Część uwag na ten temat umieściłem w poprzednim punkcie. W tym punkcie podaję przykłady niestaranności we wprowadzaniu i objaśnianiu oznaczeń; które bardzo utrudniają zrozumienie przedstawianych konstrukcji matematycznych.

W czwartym akapicie punktu 1.2, str. 22-23, wprowadzono kilka oznaczeń. Wektor parametrów modelu jest oznaczony literą θ , a odpowiedzi modelu tworzą wektor $\mathbf{x} \in \mathbb{R}^N$, co zgodnie z wcześniejszym opisem należy chyba interpretować tak, że mamy do czynienia z systemem o N wyjściach. W następnym zdaniu stwierdza się, że θ można rozumieć jako wejścia modelu, przez co Doktorant chce chyba powiedzieć, że są one wymuszeniem, czyli źródłem zmienności występującej w modelu. Potwierdzałyby to wzór (1). Jednak dalej, w pierwszym paragrafie punktu 1.2.1, Doktorant wprowadza pojęcie zmiennych niezależnych i pisze, że jedyną zmienną niezależną modelu jest czas. Zamieszanie pogłębia stwierdzenie przed wzorem (3), że całkowanie względem θ następuje po „wszystkich warunkach środowiskowych” bez wyjaśnienia, jak jest tu rozumiane środowisko. Z pewnością przydałoby się więc precyzyjniej wyjaśnić, jak są rozumiane parametry modelu.

Podobnie niejasne jest oznaczenie wspomnianego powyżej wektora odpowiedzi modelu \mathbf{x} , gdyż we wzorze (3) na następnej stronie wektor \mathbf{x} jest argumentem funkcji wiarygodności, a tymczasem funkcja wiarygodności jest określona na zbiorze pomiarów, co w tym wielowymiarowym przypadku oznaczałoby albo argument w postaci macierzy, albo wektor nN -wymiarowy, gdzie n jest liczbą pomiarów. Dalej, przed równaniem (4) jest oznaczenie $\mathbf{x} = (x_1, \dots, x_N)$ jako zbioru N zmiennych, domyślnie wyjść modelu, a w (5) jest oznaczenie $\mathbf{x} = (x(t_1), \dots, x(t_n))^T$ sugerujące, że chodzi jednak może o zbiór n pomiarów wektorowych ułożonych jeden za drugim i tworzących właśnie wektor nN -wymiarowy, przy czym jest to wektor pionowy ($nN \times 1$), w odróżnieniu od poprzednio definiowanych wektorów poziomych. Ale jeszcze dalej (25₁₉) dane eksperymentalne są oznaczone przez \mathbf{y}^D . Te niejednoznaczne oznaczenia wprowadzają zamieszanie i utrudniają zrozumienie wprowadzanych wzorów.

W drugim akapicie niejasno jest wyjaśnione znaczenie wektora (5). W zdaniu poprzedzającym stwierdza się, że jest to wyjście modelu, przez co zazwyczaj rozumie się zbiór zmiennych zależnych. W (5) jest to zbiór zmiennych zależnych obserwowanych w n chwilach. Można to nazwać zbiorem obserwacji lub pomiarów sygnałów wyjściowych, ale nazwanie tego wektora wyjściem jest mylące.

Na rys. 3 na str. 28 oznaczenia w górnej części A, zapożyczonych wprost z [126], różnią się od oznaczeń w dolnej części B, a progi zaznaczone liniami przerywanymi nie są objaśnione.

W wierszu 31₂₁ napisano, że entropia jest wyrażona w bitach. Tak jest rzeczywiście, pod warunkiem że logarytm we wzorze (15) miałby podstawę 2. Ale ponieważ nie napisano przy nim podstawy, więc umownie oznacza to, że jest to logarytm naturalny o podstawie e , a wtedy entropia jest wyrażona w natach (nitach, nepitach).

W wierszu 38₄ powinno być chyba „nie większe niż $\min(m, n)$ ”, a w 38₃₋₂ „zmiennie kanoniczne”, a nie „korelacje kanoniczne”.

We wzorze (33) na str. 42 nie wiadomo, co oznacza $H[(t, t - 1)]$.

We wzorze (36) na tej samej stronie nie wiadomo jak rozumieć zapis $\log Q(p - 1)$ w górnym równaniu, jeżeli Q jest rozkładem prawdopodobieństwa.

Czy równość przed wyrażeniem $D(P||Q)$ w środkowym wierszu wzoru (36) oznacza może definicję tego wyrażenia? Bo wcześniej nie jest ono zdefiniowane. Jeżeli tak jest, to powinno się to zaznaczyć lub opisać.

Jak należy rozumieć definicję we wzorze (40a) na str. 43, jeżeli x jest dowolną wartością spełniającą warunek $p(x) > 0$, gdzie $p(x)$ jest gęstością prawdopodobieństwa? Z tego wzoru wynika, że $H_0(X)$ jest funkcją x , gdy tymczasem zdefiniowane w (40b) $H_\alpha(X)$ dla dowolnej wartości α od x nie zależy.

Co oznacza $P_{X|Y}(\cdot | Y)$ we wzorze (42a) na str. 44 i dalej?

Wykładnik we wzorze (42b) powinien być α^{-1} , a nie $-\alpha$.

Co oznacza p_{NML} w wierszu 47₂₀?

Co to jest C_{min}^* w wierszu 50⁷?

We wzorze (72) na str. 55 i dalej powinno chyba być $I(\hat{\theta}_A, \hat{\theta}_B)$.

We wzorze (73) na stronie 56 ostatni czynnik powinien być podniesiony do kwadratu.

Jak się ma (85) do (64)?

W podpisie do rysunku 9 na str. 69 jest powołanie na panel e rysunku, ale na rysunku takiego panelu nie ma.

5. Główne wyniki rozprawy

Po przedstawieniu uwag o rozdziałach rozprawy przejdę teraz do omówienia głównych wyników, zawartych przede wszystkim w artykułach dołączonych w załączniku 2, przy czym przedstawię to w kolejności chronologicznej: artykuły (2), (1), (3), nawiązując przy okazji do zapisów w rozdziałach rozprawy.

W artykule (2) Doktorant opisuje metodę będącą ulepszeniem metody zaproponowanej w pozycji [126] wykazu literatury, mającej na celu wykrywanie przypadków „strukturalnej nieidentyfikowalności” (czyli nieidentyfikowalności parametrycznej) i „praktycznej nieidentyfikowalności” na podstawie kształtu funkcji strat, czyli kwadratowego wskaźnika jakości estymacji w otoczeniu pewnego (globalnego?) minimum w przestrzeni parametrów. Metoda zaproponowana w [126] jest heurystyczna i sprawdza się w niej numerycznie, czy w otoczeniu punktu są linie, wzdłuż których wskaźnik jakości jest płaski (co odpowiada

strukturalnej nieidentyfikowalności) lub wzrasta nieznacznie (co oznacza praktyczną nieidentyfikowalność). Przy większych zakłóceniach pomiarowych rozróżnienie między strukturalną nieidentyfikowalnością i praktyczną nieidentyfikowalnością w rozważanym punkcie może być trudne, ale ten podział jest wygodny ze względów algorytmicznych.

Numeryczne badanie identyfikowalności nie pojawia się oczywiście po raz pierwszy w [126]. Piszą o tym Walter i Pronzato⁴ w książce, której oryginał francuski ukazał się w 1994 r., a powołują się przy tym na wcześniejszą pozycję z 1982 r. Warto też zwrócić uwagę, że do badania funkcji strat można skorzystać z punktów uzyskanych w trakcie minimalizacji wskaźnika jakości estymacji. Metody wspomniane w wierszach 26⁸⁻¹⁰ generują dużą liczbę praktycznie przypadkowych punktów wokół minimum i wartości wskaźnika jakości w tych punktach można użyć do oszacowania obszaru ufności ocen parametrów, co zaproponował w niedawno bronionej w Politechnice Warszawskiej pracy doktorskiej D.A. Pięta⁵. Punktów tych można by było także użyć do estymacji kształtu funkcji strat w pobliżu minimum, co mogłoby doprowadzić do metody konkurencyjnej w stosunku do metody opisanej w [126].

Znane są oczywiście metody analitycznego badania identyfikowalności. Analityczne metody dają pewność, że model jest identyfikowalny, jeżeli uda się to wykazać, oraz pozwala na wyznaczenie zależności między parametrami, które powodują nieidentyfikowalność modelu, jeżeli uda się wykazać jego nieidentyfikowalność. Ale praktycznie mogą być one stosowane raczej do prostszych modeli, szczególnie w przypadku, gdy są one nieliniowe. Przykład przedstawiony w [126] jest stosunkowo nieduży – składa się z 5 kompartmentów – i ma dosyć specyficzną strukturę ułatwiającą analizę, ale ma nieliniowości i chociaż jego badanie metodami analitycznymi mogłoby sprawiać trudności, można by było spróbować to zrobić, chociażby po to, aby porównać wyniki teoretyczne z numerycznymi. Metody numeryczne pozwalające na badanie identyfikowalności dużych, skomplikowanych modeli są z pewnością przydatne, chociaż przynoszą one mniej informacji i w sytuacjach wątpliwych interpretacja uzyskanych wyników może być kłopotliwa.

Metoda przedstawiona w artykule (2) polega na odejściu od numerycznego badania otoczenia punktu i zastąpienia go badaniem macierzy informacyjnej Fishera, decydującej o asymptotycznym kształcie badanej funkcji strat w otoczeniu minimum. Wprowadzono dwa wskaźniki. Jeden, to kąt między wektorami kolumn, który powinien być jak największy, i drugi, to miara wypukłości funkcji strat. Oprócz inspiracji zasadniczą ideą podejścia do badania nieidentyfikowalności z artykułu [126], zaprezentowana metoda jest właściwie zupełnie nowa. Jest ona oparta na znacznie bardziej zaawansowanych pojęciach i wyprowadzeniach matematycznych i jest niewątpliwie znaczącym krokiem ulepszającym pomysł przedstawiony w [126].

Jak wspomniałem wcześniej, jeżeli model jest parametrycznie nieidentyfikowalny, to potrzebne jest albo niezależne ustalenie części parametrów, albo użycie metod regularyzacji, albo modyfikacja struktury modelu. W modelach opisowych struktura jest właściwie wyznaczana na podstawie różnych zależności fizycznych, chemicznych i podobnych innych, a parametry

⁴ É. Walter, L. Pronzato: *Identification of Parametric Models from Experimental Data*. Springer, London, 1997, punkt 2.6.1.6.

⁵ D. A. Pięta: *Metoda oceny jakości wyników eksperymentów wzbudzeń kulombowskich z wykorzystaniem algorytmu genetycznego*. Rada Naukowa Dyscypliny Informatyka Techniczna i Telekomunikacja Politechniki Warszawskiej, 2021.

mają na ogół sens fizyczny. Wbrew pozorom, uproszczenie struktury modelu w takich przypadkach może być potrzebne. W systemach biologicznych występują duże różnice międzypersonalne, a także wynikające z różnorodnych schorzeń. To może powodować, że dane pomiarowe od różnych osób wymagają uproszczonego opisu, bo część pełnego modelu ma znikomy wpływ na sygnał wyjściowy, co zazwyczaj kończy się kłopotami przy estymacji parametrów. Wypracowane historycznie metody doboru struktury modelu bazują głównie na porównywaniu modeli o różnych strukturach, tak jak jest to na przykład opisane w monografii Burnhama i Andersona⁶, co jest jednak zbyt kłopotliwe do stosowania w dużych modelach. Badanie otoczenia punktu minimum, jak w rozprawie, w przypadku wykrycia nieidentyfikowalności może dać wskazówki co do ukierunkowania poszukiwania lepszej struktury. Wydaje mi się, że warto by było pomyśleć o powiązaniu obu tych podejść, czyli zaproponowanie nowej struktury w wyniku badania otoczenia minimum, a następnie sprawdzenie, czy ta struktura jest lepsza w sensie statystycznym za pomocą uznanych testów bądź kryteriów.

Metodą pomocną w wykrywaniu tych fragmentów modelu, które nie wpływają na sygnał wyjściowy, jest analiza wrażliwości parametrów. Temu zagadnieniu jest poświęcony punkt 1.2.3, punkt 4.1 oraz artykuł (1) będący rozdziałem redagowanej książki wydanej przez MIT Press. Punkt 1.2.3 jest krótkim wprowadzeniem do zagadnienia, chociaż zrozumienie przekształceń utrudnia brak objaśnienia oznaczeń podstawowych zmiennych Y i X wprowadzonych tu po raz pierwszy. Punkt 4.1 jest szerszym rozwinięciem i pewnym skrótem materiału zamieszczonego w artykule (1). Natomiast artykuł (1) jest uporządkowanym przeglądem metod badania wrażliwości obejmującym, oprócz powszechnie znanego pojęcia wrażliwości lokalnej dla modeli deterministycznych, także mało znane pojęcia wrażliwości globalnej dla takich modeli oraz wrażliwości lokalnej i globalnej dla modeli stochastycznych. Przedstawiono także przykłady zastosowań tych metod do analizy wrażliwości parametrów dwóch modeli. Rozdział ten oceniam bardzo dobrze. Mam tylko kilka uwag związanych z umieszczonym w nim materiałem. Na str. 94 wprowadzono podstawowy model deterministyczny (1.5) będący układem równań różniczkowych z warunkami początkowymi. W modelu tym nie ma sygnałów wejściowych, tak jak we wcześniej omawianych modelach w regularnych rozdziałach rozprawy. Co ciekawe, na samym początku punktu 1.2 stwierdzono nawet, że ogólnie model matematyczny systemu biologicznego to przekształcenie parametrów modelu na odpowiedzi modelu, wykluczając wpływ nie tylko wymuszeń przez sygnały wejściowe, ale też wpływ warunków początkowych. Tymczasem w modelach z przykładów ze str. 104 i 105 ewidentnie są sygnały wymuszające (wejściowe). W modelu (1.41) sygnałem wejściowym jest generacja (produkcja) molekuł mRNA. Ta generacja wymusza przepływy innych sygnałów. Ponieważ jest ona stała, to oznaczono ją jako parametr θ_1 . Nie znam tego procesu, ale wyobrażam sobie, że mogą być przypadki, kiedy taka generacja zmienia się w czasie. Taki przypadek wyraźnie unaoczniałby wymuszający charakter tego sygnału. Jeszcze bardziej wyraźny charakter sygnału wejściowego ma zmienna $u(t)$ w modelu SP, w dolnej części str. 105 (zewnętrzkomórkowe stężenie ligandu), które aktywizują kinazę i którego sam zapis sugeruje zmienność w czasie (dalej przyjęto, że ma on charakter skoku z 0 do 100 w chwili $t = 0$). Te przykłady wskazują, że założony model bez sygnałów wejściowych jest zbyt uproszczony i nie oddaje pełnej złożoności rozpatrywanych zagadnień.

⁶ K. P. Burnett, D. R. Anderson: *Model Selection and Multimodel Inference. A Practical Information-Theoretic Approach*. Springer, New York, 2002.

Z drobniejszych uwag, definicja elementarnego efektu EE, wzór (1.30) na stronie 99 jest niejasna. W wierszach 99₅₋₄ elementarnymi efektami nazywa się siatki (1.29), aby w następnym zdaniu zdefiniować to jako stosunki (1.30). W Box 1 na str. 101 niepoprawnie formalnie są zapisy przyrównujące parametry i rozkłady prawdopodobieństwa. W tabl. 1.1 na str. 114 podano wartości dla 12 parametrów modelu SP, gdy tymczasem w zapisie modelu na str. 105 jest tylko 11 parametrów.

Artykuł (3) przedstawia nową metodę badania komórkowych systemów biologicznych. Polega ona na analizie odpowiedzi populacji komórkowych na kilka poziomów wymuszeń. Metoda ma charakter dosyć ogólny i pozwala na charakteryzację różnych populacji komórkowych, lecz jest wprowadzana i objaśniana na przykładzie konkretnych systemów, w szczególności odpowiedzi komórek na zawartość interferonu typu 1 w mononuklearnych komórkach krwi. Pomysł polega na przykładaniu po kolei kilku poziomów dawek wymuszeń oraz analizie nakładających się rozkładów odpowiedzi komórek. Analiza przesunięć tych rozkładów, nazwana analizą odpowiedzi cząstkowych (fractional response analysis) pozwoliła na utworzenie wykresów charakteryzujących przyrosty odpowiedzi. W szczególności wprowadzono krzywą odpowiedzi cząstkowej (FRC – fractional response curve), której wykresy pozwoliły na ustalenie przybliżonej relacji typu logarytmicznego między dozą wymuszenia i odpowiedzią, a w konsekwencji wniosek, że przyrost reakcji jest proporcjonalny do względnego przyrostu dawki. W artykule podano także sposób oszacowania frakcji pobudzonych komórek za pomocą regresji logistycznej. Chociaż nie znam obszaru badań związanych z biochemicznymi szlakami sygnałowymi, to wynik ten oceniam jako bardzo istotny, charakteryzujący ilościowo odpowiedź na zadane dozy wymuszenia i wnoszący nową wiedzę o reakcji komórek sygnałowych na wymuszenia.

Ciekawym wynikiem jest też powiązanie wskaźnika FCR z min-informacją Rényi`ego. Te wyprowadzenia są przeniesione do załącznika z informacjami uzupełniającymi, natomiast w rozprawie znajdują się dwa dosyć obszerne podrozdziały przeglądowe wprowadzające pojęcia informacji w sensie Rényi`ego. Są one interesujące i po dopracowaniu może mogłyby być opublikowane jako materiał przeglądowy, podobny do artykułu (1).

6. Podsumowanie

Oceniam bardzo wysoko trzy artykuły wchodzące w zakres rozprawy i przedstawione w nich wyniki. Ukazują one dużą umiejętność dostrzeżenia możliwości uzyskania nowych rozwiązań oraz doboru odpowiednich środków w celu efektywnego uzyskania tych rozwiązań. W szczególności dotyczy to umiejętności zastosowania zaawansowanych narzędzi statystycznych i matematycznych. Dobrze to uwidoczniają artykuły (2) i (3) przedstawiające istotnie nowe wyniki w stosunku do wyników znanych wcześniej. Całkiem nowe podejście przedstawione w artykule (3) opublikowanym w bardzo prestiżowym czasopiśmie *Nature Communications* pozwoliło na wykazanie istotnej zależności między pobudzeniem komórek i ich reakcją. Nowa metoda numerycznej analizy identyfikowalności parametrów opisana w artykule (2) znacznie zaawansowała ten kierunek badania modelu. Dobrze też oceniam artykuł (1) z przeglądem mało znanych wyników i porównaniem ich na kilku przykładach.

Mniej przemyślane i niezbyt starannie napisane są regularne rozdziały rozprawy. Nie chodzi mi przy tym specjalnie o terminologię, gdyż celem mojej dyskusji na ten temat było głównie przedstawienie alternatywnego nazewnictwa oraz zwrócenia uwagi na pojęcia, które są w rozprawie używane w innym sensie niż dosyć powszechnie ustalone w literaturze. Bardziej

chodzi mi o to, że niezbyt rozumiem sens dodatkowych opisów materiału zawartego w artykułach. Jeżeli celem tego miało być bardziej przystępne i intuicyjne wprowadzenie do zawartości artykułów, to liczne nieprecyzyjności i braki objaśnień spowodowały, że właściwie musiałem najpierw przeczytać same artykuły, aby móc zrozumieć te materiały objaśniające w regularnych rozdziałach rozprawy. Na pewno można się było lepiej postarać o to, aby napisać te rozdziały bardziej przyjaźnie dla czytelnika.

7. Wniosek końcowy

Bardzo doceniam zaprezentowaną w artykułach intuicję badawczą i umiejętności rozwiązywania postawionych zadań, co stanowi najważniejszy element badań naukowych i charakteryzuje dobrego pracownika naukowego. Oprócz artykułów włączonych do rozprawy Doktorant ma jeszcze spory dodatkowy dorobek publikacyjny: pięć artykułów opublikowanych przez renomowane wydawnictwa, co jest wyróżniającym osiągnięciem na tle większości doktorantów broniących rozpraw. Zarówno postawiony w rozprawie cel ogólny jak i wszystkie przyjęte cele szczegółowe zostały osiągnięte. Nie mam wątpliwości, że rozprawa spełnia, i to z nadmiarem, warunki zapisane w obowiązujących aktach prawnych, a także zwyczajowe warunki stawiane rozprawom doktorskim. Wnoszę o dopuszczenie mgr. inż. Karola Nienałtowskiego do publicznej obrony rozprawy.

