

<i>I. Аутор</i>
Име и презиме: Милован Миливојевић
Датум и место рођења: 4.10.1964. Каменица, Ужице
Садашње запослење: предавач на Високој пословно-техничкој школи у Ужицу
<i>II. Докторска дисертација</i>
Наслов: Методe развоја и адаптације регресионих модела базиране на генетским алгоритмима
Број страница: 166
Број слика: 64
Број библиографских података: 139
Установа и место где је рад израђен: Природно-математички факултет Универзитета у Крагујевцу
Научна област (УДК): 004 (рачунарство)
Ментор: др Бобан Стојановић, ванредни професор Природно-математичког факултета Универзитета у Крагујевцу, ужа научна област: Програмирање
<i>III. Оцена и одбрана</i>
Датум пријаве теме: 9. јануар 2013. год.
Број одлуке и датум прихватања докторске дисертације:
Комисија за оцену подобности теме и кандидата: 1. др Драгић Банковић, редовни професор Државног универзитета у Новом Пазару 2. др Бобан Стојановић, ванредни професор Природно-математичког факултета Универзитета у Крагујевцу 3. др Милош Ивановић, доцент Природно-математичког факултета Универзитета у Крагујевцу
Комисија за оцену докторске дисертације: 1. др Бобан Стојановић, ванредни професор Природно-математичког факултета Универзитета у Крагујевцу 2. др Дејан Дивац, научни саветник у Институту за водопривреду „Јарослав Черни“, Београд 3. др Владимир Ранковић, ванредни професор Економског факултета у Крагујевцу 4. др Милош Ивановић, доцент Природно-математичког факултета Универзитета у Крагујевцу
Комисија за одбрану докторске дисертације: 1. др Бобан Стојановић, ванредни професор Природно-математичког факултета Универзитета у Крагујевцу 2. др Дејан Дивац, научни саветник у Институту за водопривреду „Јарослав Черни“, Београд 3. др Владимир Ранковић, ванредни професор Економског факултета у Крагујевцу 4. др Милош Ивановић, доцент Природно-математичког факултета Универзитета у Крагујевцу
Датум одбране дисертације:



**УНИВЕРЗИТЕТ У КРАГУЈЕВЦУ
ПРИРОДНО-МАТЕМАТИЧКИ ФАКУЛТЕТ**

Милован Миливојевић

**МЕТОДЕ РАЗВОЈА И АДАПТАЦИЈЕ
РЕГРЕСИОНИХ МОДЕЛА
БАЗИРАНЕ НА ГЕНЕТСКИМ АЛГОРИТМИМА**

докторска дисертација

Крагујевац, 2016.

Предговор

Почетком последње декаде XX века, на Машинском факултету, Универзитета у Београду (Катедра за производно машинство, флексибилне производне технологије и роботiku), сам се бавио проучавањем Теорије експеримената¹, математичким моделирањем и оптимизацијом објеката и процеса у домену производних технологија. Један од резултата тих напора је и магистарски рад који сам одбранио на Машинском факултету, Универзитета у Београду, 1996. год. под називом *Прилог математичком моделовању вишефакторних процеса и система на основу централних композиционих планова*. Теоријску основу овог рада су чинили истраживачки радови R. Fisher-а (R. Fisher, 1935) и радови Вох-а и Вилсон-а (G.E.P. Vox и K.B. Wilson, 1951). Током рада је развијен софтвер за аутоматизовану примену пуних факторних експерименталних планова² и математичко моделирање засновано на полиномијалним и *response surface* регресијама до нивоа четвртог реда. Методологија је верификована у три домена: оптимизација процеса каљења челика, математичко моделирање и оптимизација процеса инјекционог бризања пластичних маса са циљем постизања високе тачности одговорних делова, и генерисање аналитичких једначина површи намењених обради помоћу нумерички управљаних машина.

Поред класичних вишеструких линеарних регресија, генерализација идеје описивања процеса, објеката и система помоћу нелинеарних регресионих модела, у научној заједници је доживљавала реинкарнацију у више временских периода у форми вештачких неуронских мрежа.

Проучавања и истраживања у овој области, којима сам се бавио у периоду 2003-2008. год., резултирала су софтверским пакетом за математичко моделирање засновано на вештачким неуронским мрежама који је заокружен до 2008. год. Овај софтверски пакет је тестиран и унапређен на Институту за математику и информатику Природно-математичког факултета у Крагујевцу од стране тима који су чинили: др Бобан Стојановић, др Милош Ивановић, др Вишња Симић и мр Милован Миливојевић. Један од исхода рада у овом домену је и истраживање и математичко моделирање процеса екстракције никла за налазиште Руђинци, Врњачка Бања, које је реализовано заједно са истраживачима са IME Process Metallurgy and Metal Recycling, RWTH Aachen University³.

Као плод вишегодишње сарадње са RWTH Aachen University, у периоду 2012-2014. год. је реализован и DAAD пројекат под називом *Artificial Neural Network modelling of silver nanoparticle formation after thermal decomposition of an aerosol*⁴. Пројектом су са стране RWTH Aachen руководили Prof. Dr.-Ing. Bernd Friedrich и Dr.-Ing. Srecko Stopic, а са стране Природно-математичког факултета др Бобан Стојановић, ванредни професор. Експериментални рад и оптимизација у процесу добијања наночестица алуминијума је реализован на институту IME Process Metallurgy and Metal Recycling.

¹ DOE - Design of Experiments

² Full factorial Central Composite Design plan - CCD

³ Резултати су публиковани у часопису *International Journal of Minerals, Metallurgy and Materials*, Volume 19, Number 7, July 2012, 584-595. DOI: 10.1007/s12613-012-0599-x

⁴ Акроним пројекта је ANN-NanoAg (видети на www.daad.de)

Недостаци вештачких неуронских мрежа, као што су непостојање математичког модела моделираних процеса у аналитичкој форми, могућност достизања локалног уместо глобалног минимума функције грешака, потреба за великим бројем мерења и експеримената који су често веома скупи, променљивост скупова предикторских варијабли, комплексност све захтевнијих система за моделирање (вишефакторна анализа) итд. су инспирисали даља интересовања за домен напредних регресионих форми као што су: *hierarchical regression*, *stepwise multiple regression*, *robust regression*, *ridge regression*, *partial least squares regression*. Део резултата из сегмента везаног за *stepwise regression* и вештачке неуронске мреже је објављен у часопису Mettal⁵.

Посебно питање које је додатно заокупило моју пажњу, а које је и даље отворено, је питање софтверског генерисања поузданих регресионих модела са аспекта тачности и комплексности (броја и облика регресора).

У складу са претходним, у периоду од 2010. до краја 2015. год., под менторством проф. др Бобана Стојановића, сам радио у групи истраживача са Природно-математичког факултета у Крагујевцу на истраживању могућности примене регресионих модела и оптимизационих хеуристика као што су генетски алгоритми у области хидро-информатике.

Део резултата истраживања у овом пољу је и публиковани рад *Adaptive System for Dam Behavior Modeling Based on Linear Regression and Genetic Algorithms*, аутора: Стојановић Б., Миливојевић М., Ивановић М., Миливојевић Н., Дивац Д., 2013., *Advances in Engineering Software* (M21).

Ова докторска дисертација представља део резултата истраживања у наведеним доменима.

Фокус рада је на примени генетских алгоритама у креирању адекватних регресионих модела за одабране примере примењених истраживања као и на верификацији софтверских решења креираних за претходне намене.



Велику захвалност дугујем ментору, др Бобану Стојановићу, ванредном професору Природно-математичког факултета у Крагујевцу, на инспиративној подршци у хиперпростору научног истраживања, које је резултирало овом докторском дисертацијом. Његова помоћ је била вишедимензиона: научна строгост, трагање за есенцијом знања, инсистирање на резултатима, подршка у разradi софтвера, подршка у раду на пројектима и у објављивању радова из ове области, јасноћа и кристалност изведених закључака, тимски рад, несебични временски хоризонт у креирању, разradi и реализацији идеја, стварање услова за креативни успон на духовној вертикали...

Др Милошу Ивановићу се посебно захваљујем за инспиративне научне дијалоге у домену регресија, на учешћу у креирању научних чланака, студиозном приступу, помоћи у области паралелног рачунарства...

Такође сам веома захваљан и колегама др Николи Миливојевићу дипл. инг., шефу Одсека за хидроинформатику на Институту за воде, Јарослав Черни, из Београда као и др Дејану Дивцу дипл. инг., научном саветнику овог института, на заједничком раду везаном за моделирање структурног понашања бране Бочац на реци Врбас и бране Гранчарево на реци Требишњици у Републици Српској. Др Дејану Дивцу, др Владимиру Ранковићу, ванредном професору Економског факултета у Крагујевцу и др

⁵ Milivojevic M., Stopic S., Stojanovic B., Drndarevic D., Bernd F., *Forward stepwise regression in determining dimensions of forming and sizing tools for self-lubricated bearings*, METTAL Internationale Fachzeitschrift fur metallurgie, April, 2013, Vol. 67, pp. 147-153, ISSN 0026-0746

Милошу Ивановићу, доценту на Природно-математичком факултету у Крагујевцу, посебно хвала, што су прегледом дисертације и корисним сугестијама допринели њеном квалитету.

Имао сам срећу да овакав тим води, контролише и инспирише мој рад.

Такође, велики је број колега и пријатеља у научном, стручном и мотивационом смислу допринео да ова докторска дисертација буде што боља. Поменућу само неке, који су, непосредно или посредно, уткани у странице ове дисертације: проф. др Срећко Стопић, који је одговоран за реинкарнацију математичких модела из моје магистарске тезе и који је инспиратор мог учешћа на пројектима и примени раније стечених знања у домену теорије експеримената, математичког моделирања и експерименталних истраживања за област екстракције никла, нанопрахова и рафинације олова на RWTH Aachen; др Вишња Симић која је несебично пружила помоћ у домену C# програмирања и унапређењу софтвера за моделирање процеса и система помоћу вештачких неуронских мрежа; др Драгољуб Дрндаревић који је још далеке 1995. год. допринео да упознам домен вештачких неуронских мрежа, област која је тих година била у зениту научне и стручне јавности; Горан Којадиновић дипл. инг. ел. који је помогао кристалисању идеја у домену редукције димензија и генерално *Data Mining*-а; Иван Ђоковић дипл. оец. и Томислав Милетић дипл. инг. ел. , који су несебично учествовали у инспиративним разговорима и подстицали разраду и аналитику идеја у овом домену; Милоје Лазаревић који је стрпљиво и зналачки рачунарски припремао слике, шеме и дијаграме за научне чланке који су проистекли из ове докторске дисертације; младе колеге Дарко Антонијевић, Славиша Крстић, Срђан Николић, Ђорђе Форст, који су дали допринос развоју апликација, којима смо тестирали идеје и унапређивали технике моделирања и оптимизације базиране на принципима вештачке интелигенције.

Свима њима дугујем неизмерну захвалност.

Захваљујем се колегиницама и колегама на Институту за математику и информатику, Природно-математичког факултета Универзитета у Крагујевцу на пруженој подршци и стварању пријатног амбијента и оквира за рад на тези.

Захваљујем се колективу Високе пословно-техничке школе у Ужицу, који је одобрио део финансијске подршке у изради ове докторске дисертације. Велику захвалност дугујем и колегиницама и колегама са Високе пословно-техничке школе који су ме неуморно подстицали да у вишегодишњем раду истрајем.

Посебну, неизмерну захвалност дугујем Зори, Милану и мојој мајци Душанки, за разумевање и подршку.

Хвала и мојој широкој породици,

и свима које нисам поменуо, а који су допринели да на овом путовању кроз науку уживам.

Ужице, Крагујевац

јануар, 2016. год.

Милован Миливојевић

Сажетак

Већина постојећих регресионих метода моделирања претпоставља временску непроменљивост моделираних објеката и захтева сталан скуп улазних параметара. У реалним апликацијама, сталне промене објеката и откази мерне опреме могу довести до ситуација у којима усвојени модел постаје неупотребљив. Из тог разлога је неопходно развити методе и системе за аутоматско генерисање што адекватнијих модела за дату ситуацију. У оквиру ове дисертације су развијена два хибридна метода који нуде део решења за наведене проблеме.

MLR/GA хибрид омогућава генерисање линеарног регресионог модела (MLR) који је, за дате услове, оптимизован помоћу генетских алгоритама по критеријуму тачности и критеријуму комплексности. За разлику од постојећих метода, MLR/GA метод омогућава генерисање адаптивних модела који су отпорни на променљивост скупа улазних променљивих и променљивост скупа измерених вредности. Развијени MLR/GA метод је имплементиран у виду GenReg софтверског агента, чије перформансе су тестиране у поступку моделирања радијалног померања одабраних тачака бетонске бране Боцац, на реци Врбас у Републици Српској. Модели генерисани коришћењем MLR/GA метода су у случају отказа појединих сензора показали значајно бољу способност за предикцију у односу на MLR моделе који подразумевају сталан скуп улазних променљивих. Додатно, хибридни метод је показао способност да при генерисању модела одбацује предикторе који нису од значаја за описивање посматраног објекта.

ANN/GA је хибридни метод за развој и адаптацију регресионих модела заснованих на вештачким неуронским мрежама (ANN). Коришћењем генетских алгоритама ANN/GA метод оптимизује структуру и параметре неуронске мреже у складу са актуелним скуповима улазних и излазних променљивих, и мерених вредности. За разлику од сличних постојећих решења, ANN/GA метод оптимизује скоро све елементе неуронске мреже. Хибрид врши самоподешавање модела тако што оптимизује број скривених слојева, број неурона у тим слојевима, избор активационе функције, алгоритам учења, као и вредности параметара учења у складу са одабраним алгоритмом. Развијени ANN/GA метод је имплементиран у виду DEVONNA софтверског агента који је валидован кроз студију случаја бране Гранчарево, на реци Тебишњици у Републици Српској, а резултати су поређени са резултатима добијеним коришћењем еквивалентног MLR/GA хибрида. Реализовани тестови су показали да модели генерисани ANN/GA хибридом могу дати предикције структурног понашања бране са већом тачношћу од MLR модела. Међутим, за разлику од модела у облику MLR, који су отпорни на температурне фазне помаке присутне на различитим географским локацијама, модели у форми ANN показују нестабилно понашање под таквим околностима. Поред тога, генерисање ANN модела је временски знатно захтевније.

Компаративна анализа модела генерисаних на основу MLR/GA и ANN/GA метода са једне, и модела у форми постепених регресија, са друге стране, је показала да представљени методи у појединим аспектима превазилазе могућности постојећих метода за генерисање регресионих модела. Уз примену техника редукције димензија простора истраживања, предложени хибридни методи и развијени софтверски агенти представљају моћан алат за моделирање реалних објеката и система.

Abstract

Most of existing regression modeling methods presuppose the time immutability of the modeled objects and require a constant set of input parameters. In real applications, the constant changes of the objects and failures of measuring equipment can lead to situations in which the adopted model becomes unusable. For this reason it is necessary to develop methods and systems for automatic generation of the most adequate models for the given situation. In this dissertation two hybrid methods that offer part of the solution to the above problems have been developed.

MLR/GA hybrid is able to generate a linear regression model (MLR) which is, for the given conditions, optimized by using genetic algorithms according to the criterion of accuracy and complexity criterion. Unlike the existing methods, MLR/GA method is enable to generate the adaptive models that are resistant to the variability of the set of input variables and the growing set of measured values. The developed MLR/GA method is implemented in the form of GenReg software agent, whose performances have been tested in the process of modeling the radial displacement of the selected points of Bočac concrete dam on the Vrbas river, in the Republic of Srpska. In the case of failure of individual sensors, models generated by using MLR/GA method showed a significantly better prediction compared to the MLR models that implied a constant set of input variables. In addition, the hybrid method has shown the capability of rejecting predictors that have no influence on the modeled object.

ANN/GA is a hybrid method for the development and adaptation of regression models based on artificial neural networks (ANN). Using genetic algorithms ANN/GA method optimizes the structure and parameters of neural network in accordance with the current sets of input and output variables and measured values. Unlike similar existing solutions, ANN/GA method optimizes nearly all the elements of a neural network. The hybrid performs self-tuning of the model by optimizing the number of hidden layers, the number of neurons in these layers, the choice of activation function, learning algorithm, as well as the values of learning parameters of the selected algorithm. The developed ANN/GA method was implemented in the form of DEVONNA software agent that was validated through a case study Grancarevo, on the Tebisnjica river, in the Republic of Srpska, and the results were compared to the results obtained using the equivalent MLR/GA hybrid. Completed tests showed that the models generated by ANN/GA hybrid could give predictions of structural behavior of the dam with a higher accuracy than the MLR model. However, unlike the models in the form of MLR, which are resistant to temperature phase offsets present at different geographical locations, the models in the form of ANN exhibit unstable behavior under such circumstances. In addition, the generation of an ANN model has shown much higher computational demands.

The comparative analysis of the models generated by the MLR/GA and ANN/GA methods on the one hand, and the models in the form of stepwise regression, on the other hand, has shown that the presented methods in some aspects surpass the capabilities of existing methods for generating the regression models. With the application of the research space dimension reduction the proposed hybrid methods and the developed software agents represent a powerful tool for modeling real objects and systems.

Садржај

1. Увод	1
1.1. Платформа	1
1.2. Циљ и почетне хипотезе.....	3
1.2.1. Циљ и значај.....	3
1.2.2. Полазне хипотезе.....	4
1.3. Преглед садржаја дисертације	5
1.4. Преглед стања у подручју истраживања.....	6
2. Математички модели.....	9
2.1. Кибернетски принципи моделирања	9
2.2. Општи преглед и подела математичких модела	11
2.2.1. Појам математичког модела. Поделе.....	11
2.2.2. Методологије креирања математичких модела	12
2.3. Формализам математичких модела и рационални агенти.....	13
2.3.1. Надгледано обучавање. Простор хипотеза. Сложеност. Уопштавање	13
2.3.2. Преприлагођавање (overfitting). Бирање најбоље хипотезе	15
2.3.3. Губитак.....	17
2.3.4. Трошак. Регулација.....	18
2.4. Развојни оквир за избор хипотеза базиран на MLR и GA. Шематски приказ	19
3. Регресија.....	22
3.1. Основна формулација – регресиони модел	22
3.2. Проста линеарна регресија	22
3.3. Мере репрезентативности линеарне регресије	23
3.3.1. Коефицијент детерминације – релативна мера репрезентативности.....	24
3.4. Вишеструка линеарна регресија	25
3.5. Мере квалитета регресионих модела. Кориговани коефицијент детерминације ..	27
3.6. Врсте регресија.....	30
3.7. Регресори. Базичне форме. Припрема процеса моделирања	31
4. Вештачке неуронске мреже.....	32
4.1. Вештачке неуронске мреже са простирањем унапред	32
4.2. Активационе функције	33
4.3. Алгоритми учења.....	34
4.3.1. Алгоритам са опадајућим градијентом (пакетно опадање и стохастичко опадање) (BPGD).....	35
4.3.2. Алгоритам учења са еластичном пропагацијом (RPROP).....	35
4.4. Мрежна архитектура и правила учења	36

4. Мера квалитета ANN модела	37
5. Генетски алгоритми	38
5.1. Увод	38
5.2. Природно одабирање	38
5.3. Концептулани појмови GA хеуристике	39
5.3.1. Процедура уопштеног GA	39
5.3.2. Кодирање решења. Представљање јединки	41
5.3.3. Функција добротe (<i>fitness</i>)	42
5.4. Селекција	43
5.5. Репродукција	45
5.5.1. Укрштање (Crossover)	45
5.5.2. Мутација	45
5.6. Параметри GA	46
6. Адаптација линеарних регресионих модела помоћу генетских алгоритама.....	48
6.1. Стање у области	48
6.2. Опис GenReg система	49
6.3. Опис MLR/GA методе	51
6.3.1. Математичка формулација оптимизационог проблема	51
6.3.2. Опис примењеног генетског алгорита	53
6.3.2.1. Структура хромозома	53
6.3.2.3. Оператори генетског алгорита	53
6.3.2.4. Селекција и елитизам	53
6.3.3. Софтверска имплементација MLR/GA методе	53
6.4. Валидација MLR/GA методе	54
6.4.1. Моделирање структурног понашања бетонских брана	54
6.4.2. GenReg систем за бетонске бране	55
6.4.3. Математичко-програмерски модел за бетонске бране	56
6.4.4. Студија случаја: Примена MLR/GA методе за брану Бочац	57
6.4.5. Резултати и дискусија	58
7. Адаптација ANN модела помоћу генетских алгоритама.....	63
7.1. Стање у области	63
7.2. ANN/GA модел. Опис система	64
7.2.1. Кибернетски приказ ANN/GA модела	64
7.2.2. Математичка формулација оптимизационог проблема	65
7.2.3. Генетски алгоритми за оптимизацију вештачких неуронских мрежа	66
7.2.4. Структура ANN хромозома за ANN/GA хибрид	66
7.2.5. Оператори генетског алгорита за ANN/GA хибрид	67

7.2.6. Евалуација ANN јединки.....	68
7.2.7. Селекција	69
7.3. Софтверска имплементација.....	69
7.4. Валидација ANN/GA хибрида.....	70
7.4.1. Моделирање структурног понашања брана помоћу ANN/GA парадигме	70
7.4.2. Имплементација ANN/GA хибрида за моделирање понашања брана	71
7.5. Студија случаја: Брана Гранчарево.....	72
7.5.1. ANN/GA хибридни параметри и резултати	75
7.5.2. Параметри и резултати хибрида MLR/GA.....	76
7.5.3. Дискусија	78
8. Компарација MLR/GA и ANN/GA регресионих модела са моделима у форми постепених регресија	80
8.1. Постепене регресије. Теоријске основе.....	80
8.2. Етапе у креирању постепених регресија	81
8.3. Примена постепених регресија у хидро-информатици.....	82
8.4. Поређење модела генерисаних помоћу MLR/GA и ANN/GA хибрида са моделима у облику постепених регресија.....	83
8.4.1. Студија случаја: Брана Бочац	83
8.4.2. Резултати моделирања помоћу MLR/GA и ANN/GA хибрида	84
8.4.3. Резултати моделирања постепеним регресијама	87
8.4.4. Компаративна анализа модела које су генерисали MLR/GA и ANN/GA хибриди са моделима у облику постепених регресија.....	90
9. Редукција димензија простора истраживања код регресионих модела заснованих на MLR/GA и ANN/GA хибридама	92
9.1. Анализа главних компоненти (Principal Components Analysis - PCA).....	92
9.2. PCA и редукција димензија у функцији MLR/GA и ANN/GA хибрида	94
9.2.1. Разлози за примену PCA редукције димензија при моделирању структурног понашања бетонских брана.....	94
9.2.2. PCA технике – стање и примена у области хидрологије.....	96
9.3. PCA редукција променљивих бране Гранчарево.....	97
9.4. Регресиони модели структурног понашања брана на бази MLR/GA и ANN/GA хибрида у PCA редукованом простору	110
9.4.1. Подаци и мерења	110
9.4.2. Параметри и резултати добијени на основу MLR/GA хибридне методе.....	111
9.4.3. Дискусија резултата моделирања заснованог на хибриду MLR/GA.....	116
9.4.4. Параметри и резултати добијени на основу ANN/GA хибридне методе.....	117
9.4.5. Дискусија резултата моделирања заснованог на хибриду ANN/GA.....	123
9.5. Компаративна анализа MLR/GA и ANN/GA генерисаних хибридних модела....	125

10. Закључак и правци даљег истраживања	129
10.1. Закључак	129
10.2. Правци даљег развоја	132
Библиографија	133

Списак слика

- Слика 2.1. Црна кутија објекта истраживања, 9
- Слика 2.2. Процеси анализе, моделирања и оптимизације објекта истраживања, 10
- Слика 2.3. Математички модели: подела, 11
- Слика 2.4. Пример обучавања функције једне променљиве, 14
- Слика 2.5. Сложеност хипотеза: уопштавање и прилагођавање (а) и проширење простора хипотеза (б), 15
- Слика 2.6. Скупови за учење и тестирање модела, 16
- Слика 2.7. Утицај величине простора хипотеза на *overfitting*, 16
- Слика 2.8. Развојни оквир математичког моделирања објеката и система, 21
- Слика 3.1. Укупно, објашњено и необјашњено одступање, 23
- Слика 3.2. Пример MLR са две независне променљиве, 25
- Слика 4.1. Структура *feed-forward* неуронске мреже, 32
- Слика 4.2. Трансфер и активациона функција, 33
- Слика 4.3. Модел надгледаног учења, 35
- Слика 5.1. Шема генетског алгоритма, 40
- Слика 5.2. Хромозом $v(i)$ као једна од 2^n тачака равномерно распоређених на интервалу $[x_d, x_g]$, 41
- Слика 5.3. Број са покретним зарезом, једностуке и двоструке тачности, 42
- Слика 5.4. Кумулативна доброта q_i и вероватноћа селекције p_i , 44
- Слика 5.5. Оператори укрштања (а) и мутације (б), 45
- Слика 6.1. Адаптивни систем за моделирање објеката (GenReg) базиран на MLR/GA методи, 50
- Слика 6.2. Генетска структура хромозома у GenReg систему, 53
- Слика 6.3. Адаптивни систем за моделирање структурног понашања брана базиран на MLR /GA методи, 55
- Слика 6.4. Генетска структура хромозома, 56
- Слика 6.5. Поглед на брану Бочац, 57
- Слика 6.6. (а) Поглед на брану Бочац - узводно и (б, ц) попречни пресек Блока 5, 58
- Слика 6.7. Поређење оригиналног и адаптираног предикционог модела, 61
- Слика 6.8. Апсолутна грешка предикције оригиналног и адаптираног модела, 61
- Слика 7.1. Самоподешавајући систем за моделирање објеката базиран на прилагодљивим неуронским мрежама, 65
- Слика 7.2. Генетска структура ANN/GA хромозома, 66
- Слика 7.3. Примери генетске структуре ANN/GA хромозома, 67

- Слика 7.4.** Примери ANN јединки након операције укрштања и мутације, 68
- Слика 7.5.** Процес евалуације ANN јединки, 69
- Слика 7.6.** Самоподешавајући систем за моделирање структурног понашања брана базиран на прилагодљивим неуронским мрежама – DEVONNA, 72
- Слика 7.7.** Поглед на брану Гранчарево, 72
- Слика 7.8.** Радијално померање тачке P1 (а) и попречни пресек блока 17 (б), 73
- Слика 7.9.** Скуп података за учење за брану Гранчарево за период 1984-2011., 74
- Слика 7.10.** Пример хромозома који репрезентује ANN јединку, 75
- Слика 7.11.** Вредности $RMSE_{Test}$ грешака за различите скупове предиктора и LDS:TDS односа код ANN и MLR модела, 78
- Слика 7.12.** Просечне вредности $RMSE_{Test}$ за различите подскупе предиктора, 79
- Слика 8.1.** Улазно/излазни подаци бране Бочац за период 1.1.2004-31.12.2009. године, 83
- Слика 8.2.** Улазно/Излазни подаци бране Бочац за период 1.1.2004-31.12.2009.год., 84
- Слика 8.3.** Промена вредности R_{adj}^2 у моделирању постепене регресије за ρ_{II} , 88
- Слика 8.4.** Квалитет модела генерисаних техником постепене регресије, 89
- Слика 8.5.** Просечни квалитет модела добијен техником постепених регресија, као и помоћу MLR/GA и ANN/GA хибрида, 91
- Слика 9.1.** Време потребно за генерисање ANN/GA модела радијалног померања тачке P1 у моделирању структурног понашања бране Гранчарево, 95
- Слика 9.2.** Мерне станице за мерење температуре бетона на брани Гранчарево, 97
- Слика 9.3.** Термометри за мерење температуре бетона – шема уградње, 99
- Слика 9.4.** Вредности температуре бетона мерне станице бр.3 за период 30.8.2011-9.10.2013., 100
- Слика 9.5.** Критеријум дијаграма превоја (Scree test) (извор: SPSS), 102
- Слика 9.6.** Дијаграм три главне компоненте, 103
- Слика 9.7.** Дијаграм две главне компоненте, 106
- Слика 9.8.** Улазно/излазне променљиве ($T_a, H, f_1, f_2, f_3; y_{17}$) и подаци за период 2011-2013., 111
- Слика 9.9.** Просечне вредности $RMSE_{Test}$ за одабране *cross-validation* тестове и одабране контејнере базисних функција, 113
- Слика 9.10.** Мерене (плаво) и моделске (црвено) вредности радијалног померања тачке P1 за скуп података за обучавање LDS (579 записа), за однос LDS:TDS=75:25, који одговара регресионом моделу највеће тачности, 116
- Слика 9.11.** Предиктори за ANN јединке које одговарају контејнерима од ρ_I - ρ_{VII} , 118
- Слика 9.12.** Нестабилност квалитета ANN/GA модела за мали број генерација и јединки, 119

Слика 9.13. Приказ квалитета добијених ANN/GA модела, 122

Слика 9.14. Приказ нестабилности и неотпорности на температурни *offset* модела генерисаних помоћу ANN/GA хибрида (модел *d* и *f*), 123

Слика 9.15. Приказ нестабилности и неотпорности на температурни *offset* модела генерисаних помоћу ANN/GA хибрида (модел *e* и *g*), 124

Слика 9.16. Просечне вредности грешака $RMSE_{Test}$ за ANN/GA моделе *d*, *f*, *e* и *g*, 124

Слика 9.17. Најтачнији ANN/GA модели, 125

Слика 9.18. Квалитет модела за LDS:TDS=60:40, 126

Слика 9.19. Квалитет модела за LDS:TDS=70:30, 126

Слика 9.20. Квалитет модела за LDS:TDS=75:25, 127

Слика 9.21. Квалитет модела за LDS:TDS=80:20, 127

Списак табела

Табела 4.1. Примери активационих функција, 34

Табела 5.1. Оквирне вредности параметара GA, 46

Табела 6.1. Резултати теста за скуп података у периоду 1.1.2010 и 31.12.2010., 59

Табела 7.1. Структура бинарног ANN хромозома, 75

Табела 7.2. Квалитет ANN/GA модела изражен преко $RMSE_{Train}$ и $RMSE_{Test}$, 76

Табела 7.3. Регресори који су коришћени за моделирање радијалног померања тачке P1 (y_{17}), 77

Табела 7.4. Квалитет MLR модела изражен преко $RMSE_{Test}$, 77

Табела 8.1. Елементи дескриптивне статистике улазно/излазних променљивих бране Боцац за период 1. 1.2004-31.12.2009. год., 83

Табела 8.2. Дефинисани односи величина скупова података за учење и тестирање, 84

Табела 8.3. Утицај GA стохастике и провера односа $RMSE_{Learn}$ и $RMSE_{Test}$, 85

Табела 8.4. Утицај облика регресора на квалитет регресионих модела-MLR/GA хибрида, 86

Табела 8.5. Модели креирани MLR/GA хибридом за скуп регресора ρ_{II} , 86

Табела 8.6. Модели креирани ANN/GA хибридом за скуп регресора ρ_{II} , 86

Табела 8.7. Квалитет најбољих постепених регресија над ρ_{II} скупом регресора ($CrossVal4$), 88

Табела 8.8. Квалитет најбољих постепених регресија над ρ_{II} скупом регресора, 89

Табела 8.9. Квалитет најбољих постепених регресија над ρ_{III} скупом регресора, 90

Табела 8.10. Упоредни квалитети модела $RMSE_{Test}$, добијени помоћу: MLR/GA, ANN/GA хибрида и постепених регресија за скупове регресора ρ_{II} и ρ_{III} , 91

Табела 9.1. Позиције термометара на брани Гранчарево, 98

Табела 9.2. Елементи дескриптивне статистике за средњу дневну температуру (T_a) и температуре у телу бране Гранчарево на мерним станицама од ms11 до ms81, за период 30. 8.2011-9.10.2013. год., 100

Табела 9.3. Pearson-ови коефицијенти температуре ваздуха и температура у телу бране, 101

Табела 9.4. Укупни објашњени варијабилитет (извор: SPSS), 101

Табела 9.5. Поређење λ_i добијених са PCA и Horn-овом паралелном анализом, 102

Табела 9.6. Факторске тежине свих променљивих (за ротиране факторе): Pattern Matrix, 104

Табела 9.7. Матрица Φ , за две главне компоненте, 105

Табела 9.8. Збирни преглед различитих варијанти PCA анализе, 107

Табела 9.9. Матрица F . Решење PCA анализе за податке бране Гранчарево, 108

Табела 9.10. Матрица факторских тежина (A) и матрица корелација променљивих и фактора (S) за PCA , са *oblimin* ротацијом трофакторског решења за мерење температура у систему бране Гранчарево, 109

Табела 9.11. Опис контејнера базисних функцијаи, 112

Табела 9.12. Резултати моделирања радијалног померања тачке $P1$ бране Гранчарево (31. 8.2011-9.10.2013. год.), хибрид MLR/GA , 114

Табела 9.13. Резултати моделирања радијалног померања тачке $P1$ бране Гранчарево (31.8.2011-9.10.2013. год.), хибрид ANN/GA , 120

Списак коришћених акронима

MLR	- Multiple Linear Regression / Вишестука линеарна регресија
GA	- Genetic Algorithm / Генетски алгоритми
GP	- Genetic Programming / Генетско програмирање
ANN	- Artificial Neural Networks / Вештачке неуронске мреже
AI	- Artificial Intelligence / Вештачка интелигенација
PCA	- Principal Component Analysis / Анализа главних компоненти
SVM	- Support Vector Machines / Анализа вектора подршке
SMLR	- Stepwise Multiple Linear Regression / Постепене вишеструке регресије
CV	- Cross Validation / Укрштена провера
ACO	- Ant Colony Optimization / Оптимизација мрављом колонијом
PSO	- Particle Swarm Optimization / Оптимизација ројем честица
AIA	- Artificial Immune Algorithm / Вештачки имуни алгоритми
kNN	- k Nearest Neighbors / Техника k најближих суседа
kMean	- k кластеризација заснова на центроидима
ANFIS	- Адаптивни Neuro-Fuzzy Inference System
ARMA	- Autoregressive moving-average
MLR/GA	- Софтверски хибрид
ANN/GA	- Софтверски хибрид
TDS	- Test Data Set / Скуп података за тестирање
LDS	- Learning Data Set / Скуп података за учење
RMSE	- Root Mean Squared Error / Корен средње квадратне грешке
DOE	- Design of Experiments / Дизајн експеримената
CCD	- Full factorial Central Composite Design plan / Потпуни факторијални централни композициони план експеримената
VBA	- Visual Basic for Application / Дијалект програмског језика Visual Basic

*"Кад видиш свет у зрну песка,
И небо у дивљем цвету,
Држиш бескрајност на длану своје руке,
И вечност у једном сату"*

William Blake

1. Увод

1.1. Платформа

Идеје о свођењу целокупног света интелекта на неку врсту оперисања симболима фасцинирају мислиоце већ преко двадесет пет векова. Значајан напредак у њиховој техничкој реализацији је везан за другу половину XX века и савремено доба. Велики број научника и стручњака, велика средства и године рада ангажовани су у тражењу апликативних решења за њихову имплементацију [1].

Идеје су преточене у научне дисциплине, методе и технике које чине реторту знања, која су коришћена у оквиру ове дисертације. У те научне дисциплине спадају: Теорија система, Кибернетика, Теорија математичког моделирања, Теорија оптимизације и друге. Свака од ових теорија и дисциплина је развила и унапредила своје аксиоме, теореме и специфичну нотацију у циљу математичке формализације и могућности представљања, коришћења и унапређења знања.

Поменуте математичке формализације односно математичке дескрипције реалних објеката, процеса и система се могу посматрати у три домена:

- у креирању математичких модела
- у оптимизацији и
- у класификацији.

Сваки од ових домена је у последњим декадама доживео неслућене размере развоја.

У неким аспектима концепти, идеје и апстракције су преведене у практичне апликације до нивоа који можемо окарактерисати парадигмом *вештачка интелигенација (Artificial Intelligence - AI)*. Основне компоненте таквих апликација су *математички модели*. Они могу имати различите форме и облике и могу припадати различитим типовима и класама. Појмови који су везани за ове моделе су: стохастика, динамика, вишефакторни утицаји итд, а математички облици којима могу бити приказани су: регресиони полиномијални облици, матрице, тежински коефицијенти, трансцендентне функције, Насе-ови дијаграми [2], графови и други облици.

Од назначених модела, у практичној употреби су највише присутни статистички модели засновани на *вишеструким линеарним регресијама (Multiple Linear Regression - MLR)*. Актуелне су и унапређене форме ових регресија од којих се истичу: *hierarchical regression, stepwise multiple regression, robust regression, ridge regression, partial least squares regression* итд.

Предности регресионих модела се огледају у једноставности формулације, брзини процесирања и могућности корелационе анализе за произвољно посматране k -торке независних и зависних променљивих.

Недостаци регресионих модела су везани за проблеме који се односе на број регресора⁶, као и на проблеме везане за облике одабраних регресора. Оба ова елемента, и број и облик регресора, значајно утичу на тачност, комплексност и способност уопштавања. Додатне потешкоће се јављају у апликацијама које укључују динамички променљиве скупове улазних података или временски променљиве скупове предикторских променљивих. Такође, сложене су и ситуације када у реалном времену треба креирати нове математичке моделе као резултат повратне спреге у адаптивним системима регулације.

У циљу превазилажења назначених проблема, у последње време се регресиони модели за моделирање, класификацију и оптимизацију, унапређују разним облицима хеуристика из домена вештачке интелигенције. Неке од ових техника су *вештачке неуронске мреже* (*Artificial Neural Network* - ANN), *генетски алгоритми* (*Genetic Algorithm* - GA), *методе вектора подршке* (*Support Vector Machines* - SVM), *оптимизација мрављом колонијом* (*Ant Colony Optimization* - ACO), *оптимизација ројем честуца* (*Particle Swarm Optimization* - PSO), *вештачки имуни алгоритми* (*Artificial Immune Algorithm* - AIA) и друге.

Трагање за бројем и облицима регресора који довољно добро описују процес/систем, је веома комплексан нелинеарни оптимизациони проблем који не може бити решен класичним методама, јер у себи често садржи дискретне променљиве и/или високо нелинеарне релације. Претрага за оптималним решењем грубом силом⁷ је неподесна, јер са повећањем броја независних променљивих и/или њихових базичних форми, захтева експоненцијално повећање рачунарског процесорског времена.

За превазилажење наведених проблема је могуће, као оптимизациону методу, искористити GA. Генетски алгоритми, припадају класи хеуристика које су, пре свега, намењене за оптимизацију. Ова метода је базирана на концепту Дарвинове теорије природне селекције [3], где јаче јединке имају већу шансу да опстану и пренесу своје гене будућим генерацијама кроз процес репродукције. Уколико се могућа решења неког проблема третирају као јединке, онда се итеративна парадигма еволуције може применити на добијање оптимизованих решења. Математичка формализација GA оптимизације се реализује операторима укрштања, којима се бољи гени преносе са родитеља на потомке и операторима мутације, помоћу којих се покушавају избећи локални екстреми. Одабир јединки је базиран на процесу евалуације на основу дефинисане функције, односно критеријума добротe (ваљаности).

У складу са претходним, платформа за израду ове дисертације је базирана на знањима из домена:

- математичког моделирања;
- класичне корелационе, дисперзионе и регресионе анализе;
- унапређених линеарних и нелинеарних, параметарских и непараметарских регресионих математичких модела;
- кибернетике и теорије адаптивног управљања;
- оптимизације;

⁶ У употреби су и синтагме: *basis functions* – базичне функције, *basis states* – базична стања

⁷ *brute force*

- AI хеуристика, а посебно хеуристике генетских алгоритама и хеуристике вештачких неуронских мрежа;
- домена факторске анализе са фокусом на *анализу главних компоненти (Principal Component Analysis - PCA)*.

1.2. Циљ и почетне хипотезе

1.2.1. Циљ и значај

О значају регресионих модела реалних процеса и система у веома широком спектру фундаменталних и примењених истраживања говори велики број радова у скоро свим научним дисциплинама. У последњој деценији је приметан скок броја научних чланака, у најпрестижнијим часописима, који тангирају AI домен, а који се тичу покушаја унапређивања регресионе парадигме по различитим основама. Регресиони модели опстају упркос својој времешности.

Даља истраживања у овој области су неопходна јер се, поред раније наведених потешкоћа са аспеката комплексности, тачности, одзива у реалном времену, стабилности броја предиктора у функцији времена и слично, у реалним апликацијама примена регресионих модела сусреће са додатним изазовима. Неки од тих додатних изазова су: потреба за великим бројем мерења у моделима са великим бројем фактора, могућност конвергенције грешке модела ка локалном уместо глобалном минимуму, немогућност визуелног предвиђања типа и облика регресионих зависности када је реч о вишедимензионим проблемима, итд.

Трагање за оптималним регресионим формама које уважавају претходне проблеме, би било веома мукотрпно или скоро немогуће на бази класичног приступа, па је за моделирање сложених реалних објеката и система и њихову оптимизацију корисно употребити модерне технике, као што су генетски алгоритми и/или њима сродне технике.

На основу претходних чињеница се може истаћи да је примарни циљ дисертације: **развој и имплементација метода за генерисање и адаптацију регресионих математичких модела помоћу генетских алгоритама.**

Дефинисани циљ ће бити остварен кроз:

- развој MLR/GA хибрида, односно кроз развој метода за генерисање, оптимизацију и адаптацију модела у форми вишеструких линеарних регресија коришћењем GA концепта;
- креирање система за адаптацију регресионих модела у реалном времену;
- компаративну анализу модела добијених помоћу MLR/GA хибрида и модела у облику *постепених регресија*;
- развој ANN/GA хибрида, односно GA метода за генерисање, као и параметарску и структурну оптимизацију и адаптацију нелинеарних регресионих модела у форми вештачких неуронских мрежа;
- програмску имплементацију GA формализма и адаптивну оптимизацију регресионих модела по броју и типу регресора,
- унапређење MLR/GA и ANN/GA хибрида редукцијом димензија помоћу PCA технике, као и кроз
- тестирање регресионих парадигми у примењеним инжењерским областима, и

- генерализовање примењене методологије и креирање софтверског пакета, за употребу у ширем спектру примењених истраживања.

Значајно је да дисертација садржи читав опус стратегија које тангирају домен регресија, тако да може да послужи као својеврстан водич за примену различитих метода у сложенем послу математичког моделирања комплексних објеката и система.

1.2.2. Полазне хипотезе

Основна хипотеза овог рада је да је моделирање класичним вишеструким линеарним регресијама недовољно добро за: реалне објекте и системе који се одликују великим бројем улазно/излазних променљивих, као и великим обимом улазно/излазних података са значајним шумом; за објекте код којих је скуп улазних мерења или скуп предиктора динамички променљив, као и за објекте моделирања који захтевају адаптирање математичког модела у реалном времену.

Полазне хипотезе су следеће:

1. Генетски алгоритми могу оптимизовати број и тип регресора при моделирању објекта, процеса и система вишеструким линеарним регресијама.
2. Примена GA парадигме обезбеђује мању комплексност добијених регресионих модела (*регуларизација*⁸).
3. Хибридни оптимизовани математички модели базирани на парадигмама: вишеструке линеарне регресије-генетски алгоритми и вештачке неуронске мреже – генетски алгоритми (MLR-GA, ANN-GA), могу на адекватан начин описати одабране процесе и допринети бољој предикцији излазних ефеката ових процеса.
4. Примена MLR/GA и ANN/GA хибрида омогућава упоредиву или бољу тачност уз упоредиву комплексност у односу на постепене регресије.
5. Алгоритми предпроцесирања података засновани на PCA техникама обезбеђују креирање адекватних регресионих модела смањене комплексности.

⁸ Регуларизација – појам објашњен у одељку 2.3.4.

1.3. Преглед садржаја дисертације

У предговору је приказан развојни пут аутора дисертације у истраживању инспиративног домена теорије експеримента, математичког моделирања и оптимизације током претходне две декаде. Објашњена је мотивација за рад на дисертацији.

Дисертација се састоји од десет поглавља, чији је кратак садржај дат у наставку.

У **првом поглављу** је описана шира платформа простора истраживања. Дефинисан је циљ и значај дисертације и постављене су полазне хипотезе. У завршном делу поглавља је дат преглед стања у овој области истраживања уз навођење литературних извора.

Друго поглавље у уводном делу дефинише кибернетски приступ у поступку математичког моделирања и оптимизације реалних објеката, процеса и система. Дате су поделе математичких модела по разним основама и дефинисана методологија креирања математичких модела. Дефинисано је значење математичких модела у терминологији рационалних агената и парадигми вештачке интелигенције. Описани су појмови који ближе описују математичке моделе као што су: надгледано обучавање, простори хипотеза, преприлагођавање, губитак, трошак, регуларизација... Поглавље се завршава постављањем развојног оквира за избор најбољих математичких хипотеза које су засноване на линеарним и нелинеарним регресионим моделима и генетским алгоритмима.

Треће поглавље даје кључне теоријске поставке простих линеарних регресија и вишеструких линеарних регресија. Фокус је стављен на приказ мера квалитета регресионих модела, а посебно коригованог коефицијента вишеструке детерминације. Отворено је питање ефикасности и ефективности регресионих модела, како са аспекта тачности, тако и са аспекта сложености појединачних регресора и регресионих модела у целини.

У **четвртном поглављу** су описани кључни појмови и поставке вештачких неуронских мрежа са простирањем унапред. Дат је приказ и најчешће коришћених мера квалитета математичких модела у форми неуронских мрежа.

Пето поглавље даје теоријску подлогу и опис кључних одредница у домену оптимизационе методе генетских алгоритама. Дата је процедура уопштеног GA, приказане варијанте кодирања и представљања јединки, као и поступци селекције и репродукције. Описани су оператори укрштања и мутације и дефинисани остали базични параметри генетских алгоритама.

У оквиру **шестог поглавља** дисертације су приказани развијена MLR/GA метода као и GenReg софтверски агент, за генерисање адаптивних регресионих модела оптимизованих помоћу генетских алгоритама. Валидација развијене методе је реализована на студији случаја бране Бочац на реци Врбас у Републици Српској. У поглављу су дати резултати и дискусије којима се потврђују прве две хипотезе на којима се заснива спроведено истраживање.

Седмо поглавље садржи приказ развијене ANN/GA методе за моделирање објеката и система помоћу GA оптимизованих нелинеарних регресионих модела у облику неуронских мрежа. Валидација развијене методе је реализована на студији случаја бране Гранчарево на реци Требишњици у Републици Српској. У поглављу су дати резултати и дискусије којима се потврђује трећа хипотеза ове дисертације.

У **осмом поглављу** дисертације су дати приказ перформанси и упоредна анализа модела који су генерисани помоћу MLR/GA и ANN/GA метода и одговарајућих софтверских агената, са моделима у облику постепених регресија. Валидација је реализована на студији случаја бране Боцац на реци Врбас у Републици Српској. Резултати и дискусије су потврдили четврту хипотезу овог истраживања.

Девето поглавље даје анализу перформанси развијених MLR/GA и ANN/GA модела структурног понашања бетонске лучне бране Гранчарево у РСА редукованом простору. У поглављу су дати резултати и дискусије којима се потврђује пета хипотеза ове дисертације.

Након описаних поглавља, у **дестом поглављу** су дати сумарни, сажети прегледи добијених резултата, коначне упоредне анализе и закључци о развијеним софтверским агентима. У складу са изведеним закључцима, на крају овог поглавља су дефинисани и могући правци даљих истраживања.

1.4. Преглед стања у подручју истраживања

Модел базиран на вишеструким линеарним регресијама се користе за моделирање линеарних односа између зависних променљивих⁹ и независних, односно предикторских променљивих. У многим применама функционалне форме предиктора¹⁰ нису линеарне и могу представљати комбинације једноставнијих функционалних облика. На пример, базисна функција може имати облик $f_m : x_{im} \cdot x_{jm}^2 \cdot \sin(x_{jm})$ где је f_m , m -та базисна форма i -тог и j -тог предиктора. Облик и структура ових форми зависи од природе феномена који се моделира [4-6].

У последњим декадама је за решавања све сложенијих MLR моделирања (велики број значајних фактора, велики шум, динамички променљив скуп улазно/излазних података и променљивих ...) истраживачки процес све више усмерен на могућност употребе модификованих и унапређених регресионих техника као што су: *hierarchical regression*, *robust regression*, *ridge regression*, *partial least squares regression* итд.

Једна од метода, која је значајно истраживана, је и постепена регресија [7]. Постепене регресије су есецијално базиране на *t*-статистици и *парцијалним коефицијентима корелације*. Правилно коришћена, постепена регресија може поседовати квалитетнији информативни садржај него оригинална регресија. Ово посебно важи у случајевима када модел обухвата велики број независних променљивих и/или када се математички модел фино подешава, додавањем или одузимањем променљивих. Неправилно коришћена, ова метода може конвергирати ка лошем моделу уз пад поузданости и пад поверења у регресионе моделе генерално [8]. У научној заједници су подељена мишљења о квалитету и думетима ове методологије [9, 10].

Ипак, у класичним MLR или модификованим MLR, као што су постепене регресије, број могућих подмодела за евалуацију и испитивање тачности и комплексности као и потребно процесорско време, експоненцијално расте са порастом броја улазних фактора [11-14].

У научно-стручној литератури и научним чланцима, су приказане различите идеје о томе на који начин се потенцијално комплексни MLR модели могу учинити

⁹ Зависне променљиве - предиканти

¹⁰ Користе се синоними: функционалне форме предиктора, регресори (*regressors*), базисне функције (*basis functions*), стања (*states*)

једноставнијим. У суштини, већина техника и метода своди се на три стратегије или њихову комбинацију.

Прва стратегија је везана за појам *значајности* предиктора, односно на избор релативно малог броја значајних улазних фактора. На овај начин се и хиперпростор за избор могућих регресора смањује, а тиме и потенцијална сложеност модела. Реч је, дакле, о редукцији димензија. Ова стратегија углавном примењује добро познате и разрађене статистичке методе као што су: латински квадрати, латински кубови [15], Плекет-Берманови вишефакторни селекциони планови, Фишеров тест сигнификантности [15, 16, 1], PCA факторска анализа [17, 9, 10], Hasse-ови дијаграми [2], итд.

Друга стратегија је везана за концепт *непараметарских регресија*, који је пре свега фокусиран на скупове емпиријских података/мерења и експерименталних резултата, а тек посредно на број и облик регресора. Суштина је да се у процесу учења, односно генерисања и адаптирања математичких модела, не користи целокупан скуп, већ један део података. Овакав приступ најчешће доводи до једноставнијих модела. Сам редуковани скуп *примера за обучавање* регресионог модела може бити дефинисан на различите начине: концептом *блиских (сличних)* тачака (што представља есенцију kNN филозофије) [17-19], техникама базираним на хибридном kNN/GA *методологијама кластеризације* [18-20], методама проналажења *оптималних језгара* која, према одабраном закону, смањују утицај података (тачака) пропорционално њиховој удаљености од центра језгра [21], као и другим техникама или њиховим комбинацијама.

Трећа стратегија помоћу које се екстензивни модели могу упростити се односи на идеје регуларизације. Идеје регуларизације или кажњавања сложених *хипозета*¹¹ се могу реализовати на различите начине, било да је реч о класичним MLR или генерализованим облицима нелинеарних регресија као што су ANN. Једна од техника оптимизације која материјализује овакав приступ је GA хеуристика [22, 23].

Примена генетских алгоритама у креирању регресионих модела у последње време се значајно унапређује разним додатним хибридном техникама из домена вештачке интелигенције. За потврду наведене тврдње, као пример, се наводи кратак преглед истраживања и примене класичних, унапређених и хибридных регресионих модела у домену структурног понашања брана у хидро-информационим системима.

Веома дуго су статистички модели примењивани за надгледање сигурности брана и проналажења утицаја појединих величина на њено понашање [24]. Овакви модели, базирани на MLR и њеним унапређењима, као што су хијерархијске регресије, постепене регресије, вршне регресије и друге MLR модификације, су се показали више или мање успешним у опису структурног понашања брана [25]. Додатно су развијане и детерминистичке методе које су захтевале решавање диференцијалних једначина, за које је одређивање коначних решења, понекад тешко или немогуће [26]. Велику групу техника чине и методе базиране на нумеричким моделима, као што је метода коначних елемената [27]. Ипак, данас се многе методе за моделирање брана обогаћују разноврсним хеуристикама из домена вештачке интелигенције, креирајући хибридне моделе који комбинују њихове предности. Неке од тих техника су вештачке неуронске мреже [28], генетски алгоритми [29], методе вектора подршке [30], адаптивни неуро-фази системи (*Adaptive Neuro-Fuzzy Inference System* - ANFIS) [31, 32], *Monte Carlo* симулације [33], вештачки имуни алгоритми [34] и др.

¹¹ Хипотеза – математички модел. У овој дисертацији хипотеза представља могући регресиони модел

Тако је Mata у свом раду, [28] презентовао поређење између MLR и ANN модела за опис понашања лучне бране Alto Rabagao под дејством спољних оптерећења. SVR техника (*Support Vector Regression* - SVR) надограђена са генетским алгоритмима се често комбинује са постојећим ANN и ARMA техникама (*AutoRegressive Moving Average* - ARMA) у циљу унапређења предикције понашања брана [30]. Ранковић и др. [31] су презентовали студију чији је циљ био развој ANFIS система за предвиђање радијалног померања лучне бране Боцац на реци Врбас. ANFIS модел је био предложен као алтернативни приступ и у процени испаравања за Yuvacik Dam брану у раду [32]. Wen-Chun W. и коаутори у - студији [35] су истраживали могућности неколико техника вештачке интелигенције за моделирање временских серија речних протока, који укључују ANN приступ, ANFIS технике, GP модел и SVM метод, и поредили њихове перформансе са традиционалним техникама као што су ARMA модели. Хибридни алгоритми и њихове примене у предикцији и оптимизацији су такође презентоване у бројним чланцима. Gholizadeh и др. [36] користе хибридну методологију комбиновану од метахеуристика GA, PSO и неуронских мрежа за предлагање ефикасног софтверског модела за остварење оптималног облика лучних брана. Постоје и бројни радови који користе AIA алгоритме, који имитирају функције природних имуних система. На пример, Xi и др. [37] предлаже имуни статистички модел за решавање проблема анализе хоризонталног узводно-низводног померања брана.

Као што се из претходног текста види, могућности за истраживање у домену хибридног моделирања заснованог на регресијама и GA су неограничене, било да је реч о моделирању структурног понашања бетонских брана, било да је одабрана нека друга област као што је хидрометалургија, нанотехнологија, медицина или било која друга област научних или примењених истраживања.

2. Математички модели

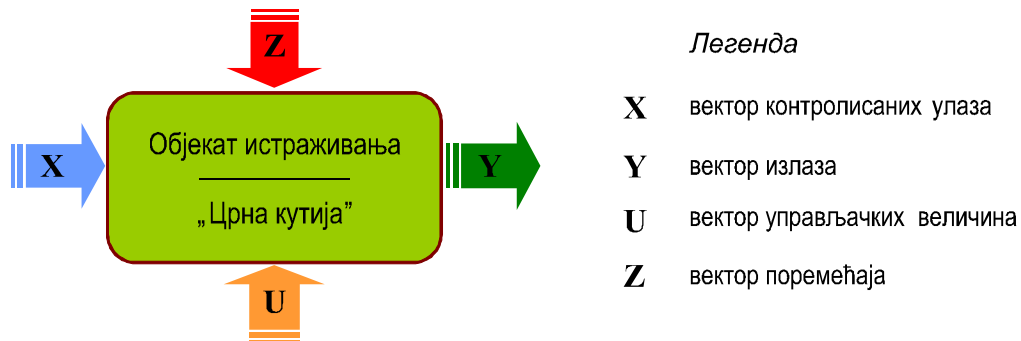
2.1. Кибернетски принципи моделирања

Процесу математичког моделирања, у оквиру ове дисертације се приступа са кибернетског становишта [38]. У складу са таквим приступом, се могу издвојити следеће кључне одреднице које чине окосницу за развој апликација: Wiener-ова *црна кутија* (*black box*) као апстрактни репрезент моделираног објекта, *математички модел* као срж ове апстракције, *електронски рачунар* као средство за реализацију идеја и *циклична природа процеса истраживања и моделирања* као основ повратне спреге у процесу адаптације.

У складу са Wiener-овим кибернетским поимањем, процес истраживања, моделирања, оптимизације и управљања одабраним објектом Θ , се може посматрати као уређени скуп вектора улаза X , излаза Y , вектора поремећаја Z и вектора управљачких величина U :

$$\Theta = \{X, Y, Z, U\} \quad (2.1)$$

Основни облик *black box* приступа истраживању је дат на слици 2.1.



Слика 2.1. Црна кутија објекта истраживања

Детаљнији садржај процеса који омогућавају да се објекат апроксимиран црном кутијом анализира, разуме, моделира и оптимизује је дат на слици 2.2. Кључне одреднице таквих процеса су:

- 1) дефинисање циља (циљева), прикупљање априорних информација, коришћење литературе и искуства експерата, анализа потребног хардвера и потребне софтверске подршке, процена допустивих губитака времена и средстава, анализа претходних резултата, разматрање структуре и параметара система, декомпоновање система, дефинисање услова управљања објектом, почетна поставка модела;
- 2) избор улазних фактора односно независних променљивих као и домена њиховог варирања (простора истраживања), са посебним освртом на *управљивост*, *стохастичност* и *мерљивост*. Избор вектора зависних променљивих. Реализација експерименталног плана (мерење, прикупљање, аквизиција и анализа података о објекту истраживања);

- 3) математичко моделирање одабраном класом математичких модела над скупом добијених улазно/излазних података намењених за обучавање (*Learning Data Set* - LDS) као и тестирање добијених модела над скупом података за тестирање (*Test Data Set* - TDS);
- 4) дефинисање и избор критеријума и метода оптимизације;
- 5) оптимизација;
- 6) корекција математичког модела (ако је потребно);
- 7) повратак на фазу 2 или неки њен део (условно);
- 8) интерпретација резултата (анализа добијених вредности параметара модела, процена поузданости модела...).



Слика 2.2. Процеси анализе, моделирања и оптимизације објекта истраживања

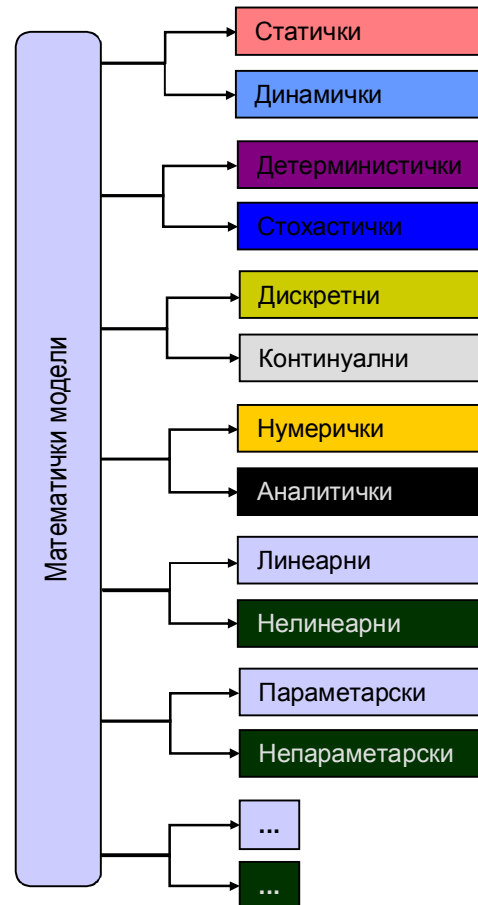
Са слике 2.2 се може видети да креирање математичког модела представља кључни део у процесу оптимизације и адаптивног управљања реалним објектима и системима са кибернетског аспекта. По том основу детаљнији описи, класификације и формалне нотације математичких модела су дате у посебном одељку.

За практичну имплементацију идеја математичког моделирања, коришћење рачунара се подразумева. Ово посебно важи у домену хеуристика које не поседују децидну аналитичку форму. Способност рачунара да оперише са апстрактним симболима 0 и 1, чије комбинације могу значити било шта, као и способност извршавања логичких операција, које могу поставити у однос све са свачим, представља делимично отелотворење *вештачког ума*, способног да се ухвати у коштац са тежином и лепотом математичке духовне вертикале.

2.2. Општи преглед и подела математичких модела

2.2.1. Појам математичког модела. Поделе

Под математичким моделом реалног система се подразумева скуп математичких форми и релација (формула, једначина, неједначина, логичких услова, оператора, тежинских коефицијената, графова итд.) које описују функционисање система, односно одређују карактеристике стања система у зависности од параметара система, улаза, почетних услова и времена [1, 15]. Због комплексности физичких појава и процеса у реалним условима, математички модели најчешће представљају апроксимације реалних објеката које обухватају њихове најзначајније особине. Класификација математичких модела се може вршити према различитим критеријумима (слика 2.3).



Слика 2.3. Математички модели: подела

Према променама у времену математички модели могу бити *статички* и *динамички*. У статичким моделима величине и релације нису у функцији времена. Динамички модели садрже величине које су функције времена.

Према критеријуму неодређености (неизвесности) модели се деле на *детерминистичке* и *стохастичке*. У детерминистичким моделима излаз може бити потпуно описан на основу улаза. У стохастичким моделима излазни ефекти су случајне величине. Отуда детерминистички математички модел експлицитно гласи:

$$\mathbf{Y} = \varphi(\mathbf{X}, \mathbf{U}) \quad (2.2)$$

а стохастички;

$$Y = \varphi(X, U, Z) \quad (2.3)$$

Значење вектора X , Y , U , и Z , је дато на слици 2.1. Детерминистички модел често представља приближан, поједностављен математички опис реалног објекта, док је природа већине реалних објеката стохастичка. Према томе, ако се жели адекватан математички опис неког реалног објекта, тада се мора користити стохастички модел. Са друге стране, када су стохастичка обележја у реалном објекту мањег степена, или када се жели приближан и поједностављен модел, може се користити детерминистички модел објекта.

Према континуалности модели могу бити: *дискретни* и *континуални*. У континуалном моделу величине имају континуалне вредности. Дискретни модели садрже дисконтинуалне промене.

На основу математичких поступака којима се анализирају, модели се обично сврставају у две групе: *нумеричке* и *аналитичке*. Нумерички модели се обрађују нумеричким поступцима (величине и функције могу да буду дате и табеларно). Аналитички модели садрже само функције у аналитичкој форми, а могу се решавати аналитички и/или нумерички.

У литератури се среће и подела математичких модела на *линеарне* и *нелинеарне*. Линеарни модел садржи само линеарне функције, а нелинеарни може садржати и нелинеарне функције.

Поред наведених, у последње време је актуелна и подела математичких модела на *параметарске* и *непараметарске* моделе. Ова подела је везана за обим коришћених расположивих података о понашању објекта или система који се моделира [21]. Код параметарских модела, читав скуп улазно/излазних података, се трансформише у параметре модела (на пример коефицијенте MLR или тежинске коефицијенте код ANN), независно од тога да ли је у питању 200, 100.000 или 2.000.000 података. Дакле, број параметара је фиксан, без обзира на обим примера за обучавање. Непараметарске методе карактерише коришћење дела улазно/излазних података који су основ за генерисање *локалних* математичких модела. Број параметара у моделу може бити променљив, јер је и број примера за обучавање, одабран по неком критеријуму, променљив. На пример, број примера за обучавање k , код неких kNN техника зависи од позиције тачке упита (x_j). Код непараметарских регресија које су базиране на *језгрима*, број тачака које покрива језгро зависи од позиције *центра језгра*. Сличне напомене важе и за *непараметарске класификаторе* (линеарни класификатор, логистички класификатор...) [21].

2.2.2. Методологије креирања математичких модела

За генерисање математичког модела неког реалног објекта, у научним методологијама постоје два концепта: *аналитички метод* и *експериментални метод*¹².

Аналитички метод базира на фундаменталним теоријским истраживањима унутрашњих интердејстава и механизма у процесима, објектима или системима и на примени познатих општих научних закона. Модели добијени овим методом су опште важећи за читаву фамилију сродних објеката, што је и њихова главна предност у односу на емпиријске моделе. Из математичког модела могу се, за било који објекат из класе, одредити излази (карактеристике) објекта за било које вредности улаза. Мана

¹² емпиријски, метод *црне кутије*

аналитички добијених модела је њихова применљивост на релативно једноставније или идеализоване објекте проучавања.

Експериментални метод се, због своје тачности, далеко више примењује. Ова значајна предност знатно се умањује узимањем у обзир и недостатака овог приступа као што су: релативно велики трошкови експерименталних испитивања, важење модела само за испитивани домен варијације улаза, итд.

Компромис између ова два полазишта у изучавању представља **комбиновани метод**. На основу аналитичког метода се поставља општи облик система једначина, неједначина, форми, математичких нотација и структура којима се формално описује дати реални објекат. Како су у оваквим моделима параметри најчешће непознати, њихово одређивање и провера адекватности управо се решавају применом експерименталног концепта [15].

2.3. Формализам математичких модела и рационални агенти

Класичну терминологију везану за математичко моделирање у новије време увелико замењује терминологија из AI домена. Тако се појам *генерисања математичког модела* може заменити појмом *обучавања рационалног агента* или *softbot-a*.

Рачунарски агент је рачунарски програм од кога се очекује да аутономно функционише, да опажа своје окружење, да се адаптира на промене и да креира и следи циљеве. Рационални агент је онај агент који делује тако да остварује *најбољи исход* или, када постоји неизвесност, *најбољи очекивани исход* [21].

Свака компонента агента може да се побољша обучавањем над подацима. Успех зависи од одговора на четири главна питања:

- Коју компоненту агента треба побољшати: извођење закључака, пресликавање услова тренутног стања у акције, анализу података о томе како се окружење развија/понаша или компоненту корисности?
- Које априорно знање агент већ има?
- Како је представљено знање (исказне и логичке реченице / за логичке агенте; Бајесовске мреже / за агенте у теорији одлучивања; факторисано представљање знања / код агената у домену машинског обучавања...)?
- Која се повратна спрега користи у обучавању (ненадгледано обучавање; обучавање појачавањем; надгледано обучавање на бази парова улаз/излаз, полунадгледано обучавање...)?

У оквиру ове докторске дисертације ће се користити методологија *надгледаног обучавања рационалног агента*, што је пандан креирању математичког модела објекта истраживања на бази парова улаз/излаз у класичној терминологији.

2.3.1. Надгледано обучавање. Простор хипотеза. Сложеност. Уопштавање

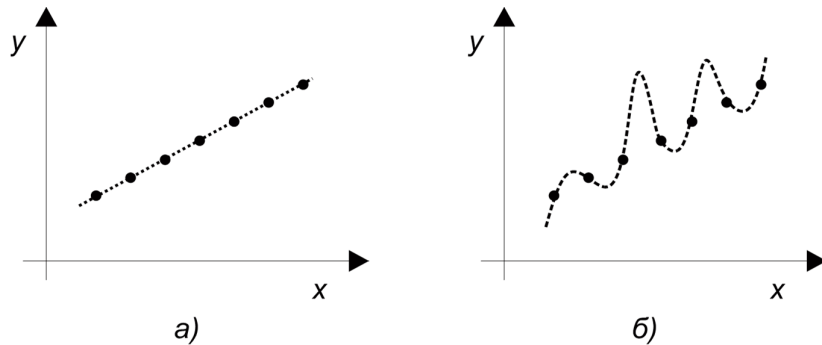
Задатак надгледаног обучавања је следећи: Ако је дат скуп обучавања од n примера парова улаз/излаз: $(x_1, y_1), (x_2, y_2), \dots, (x_j, y_j), \dots, (x_n, y_n)$, где је x_j улаз и y_j излаз, а свако y_j је генерисала једна непозната функција $y_j = f(x_j)$ - треба наћи функцију h која представља апроксимацију функције f .

Уопштено, функција f може представљати пресликавање *вектора улаза* на *вектор излаза* тј. $Y_j = f(X_j)$, где је X_j , j -ти пример улаза односно j -та каторка улазних фактора $(x_{j,1}, x_{j,2}, \dots, x_{j,i}, \dots, x_{j,m})$ (индекс i означава i -ти фактор j -те каторке, а m број фактора). Вектори X и Y могу бити било шта. Не морају бити бројеви.

Функција h се често назива *хипотезом*. *Обучавање* је трагање за погодном функцијом h у *простору хипотеза*. Хипотеза h је *погодна* ако даје добре перформансе и ван скупа обучавања. Прецизност хипотезе се мери помоћу *тест скупа примера*. Хипотеза h добро *уопштава* ако правилно предвиђа вредност Y за нове примере.

Обучавање хипотеза се може односити на класификацију и моделирање. Класификација се може реализовати у више подскупова из коначног скупа вредности, или на два подскупа ако је у питању *буловска* односно *бинарна* класификација. Код моделирања, хипотезе попримају разноврсне математичке форме. Неке од тих форми су регресиони модели.

Ако се посматра обучавање функције једне променљиве $y = f(x)$ у простору хипотеза χ , а *примери обучавања* су тачке (x, y) у равни xOy , онда се кроз дефинисане тачке могу поставити различите врсте конзистентних¹³ хипотеза, $h \in \chi$. Примери хипотеза су: хипотеза у облику линеарне функције $h: b_0 + b_1 \cdot x$ (слика 2.4 а) или хипотеза у облику полинома петог степена $h: b_0 + b_1 \cdot x + b_2 \cdot x^2 + b_3 \cdot x^3 + b_4 \cdot x^4 + b_5 \cdot x^5$, (слика 2.4 б).

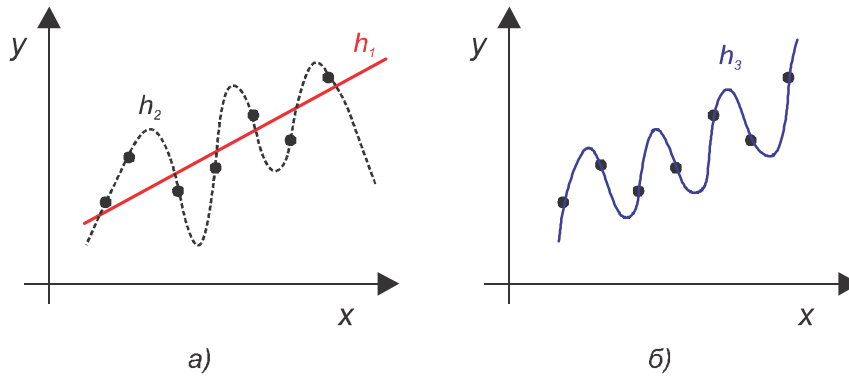


Слика 2.4. Пример обучавања функције једне променљиве

При избору хипотезе која репрезентује модел се јавља проблем: Како одабрати једну од више конзистентних хипотеза? Природни одговор би био примена *Ockham*¹⁴-ове оштрице (Слика 2.4 а): избор најједноставније хипотезе која је конзистентна подацима. На слици 2.5 а, је приказан други скуп података. За овај скуп података не постоји конзистентна права линија већ права линија која добро уопштава (*једноставна хипотеза* h_1). Ако се трага за хипотезом која је конзистентна, онда би то могао бити полином шестог степена јер је у питању 7 тачака (*сложена хипотеза* h_2). У општем случају се трага за компромисом између *једноставније* хипотезе која добро *уопштава* и *сложене* хипотезе која је добро *прилагођена* подацима.

¹³ Под *конзистентном* хипотезом се подразумева функција која се подудара са свим тачкама у скупу обучавања. Појам је аналоган са *интерполацијом*

¹⁴ William of Ockham (1280-1349) - енглески филозоф (*Pluralitas non sunt penenda sine necessitate*) [21]



Слика 2.5. Сложеност хипотеза: уопштавање и прилагођавање (а) и проширење простора хипотеза (б)

Питање једноставности и сложености хипотеза је директно везано за појам *изражајности* простора хипотеза. До промене изражајности простора хипотеза долази, на пример, када се породици хипотеза у облику полинома, дода класа трансцендентних функција. Могуће је, на пример, да краћа и једноставнија функција $y = x + \sin(x)$, (хипотеза h_3) из простора хипотеза чија је изражајност обogaћена тригонометријским функцијама, боље уопштава податке а да је по прилагођености слична полиному шестог степена. Трагање за квалитетним уравнотежењем једноставности, уопштавања и изражајности, је веома комплексан задатак.

2.3.2. Преприлагођавање (overfitting). Бирање најбоље хипотезе

Већина модерних техника из AI домена су по својој природи итеративне. На пример код ANN, при довољно великим бројем циклуса у току итеративне процедуре грешка учења ће конвергирати ка некој малој вредности. Познато је да се у поступку моделирања помоћу ANN јавља прекомерно учење¹⁵ и да поступак обучавања треба прекинути у зонама са минимумом грешке добијене на тестном скупу података. Ово значи да, иако би даљим учењем ANN модел могао бити тачнији, његова способност за уопштавање би се смањивала.

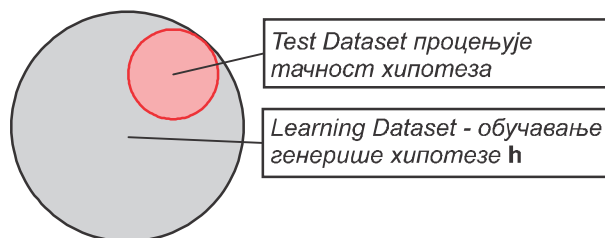
Пример функције која је консистентна, али лоше уопштава (јер је на располагању мали број тачака-примера) је хипотеза h_2 дата на слици 2.5 а. Овакви случајеви називају се *преприлагођавање*¹⁶. Реч је о ангажовању превише сложене функције за релативно једноставан скуп тачака. Такво стање проузрокује, да функције имају велико *таласање* између чворних тачака па лоше уопштавају на тестном скупу података. Типичан пример таквих хипотеза су Њутнови интерполациони полиноми.

Генерално *overfitting* постаје вероватнији са проширењем простора хипотеза и бројем улазних фактора, а мање вероватан са повећањем броја примера за обучавање. За одређивање најбоље прилагођене хипотезе будућим подацима, је потребно да расподела вероватноћа примера $\{(x_1, y_1); (x_2, y_2); \dots; (x_n, y_n)\}$ буде стационарна у времену.

Појам *најбоље прилагођености* се односи на тест скуп података. Питање прилагођености је веома комплексно и, поред осталог, је везано за однос величина скупова за учење и тестирање (LDS:TDS) (слика 2.6).

¹⁵ *overlearning*

¹⁶ *overfitting*



Слика 2.6. Скупови за учење и тестирање модела

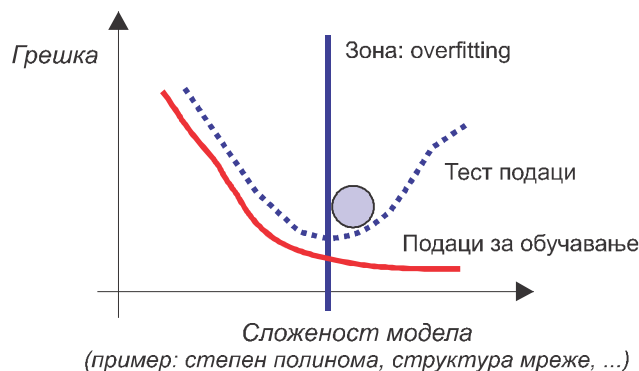
Ако је тај однос 50:50, могућ је избор лоше h , а ако је, на пример, 90:10 могућа је лоша процена стварне прецизности усвојене функције.

Неки од начина за превазилажење потенцијалних проблема по овом питању су: *k-тострука унакрсна провера*¹⁷ [39-41] или *унакрсна провера са изостављањем једног примера*¹⁸ [19].

На основу претходног, видимо да је избор најбоље h омеђен са: (а) дефинисањем простора хипотеза и (б) проналажењем *најбољих хипотеза* у том простору.

Разматрање природе простора хипотеза (степен полинома, фреквенције трансцендентних функција...) доводи до појма *величине* простора хипотеза. На пример, уколико би се за простор хипотеза одабрали полиноми, онда би линеарне функције имале величину 1, квадратне параболе величину 2, а кубне параболе величину 3. Истраживачи теже да пронађу вредност параметра *величина* који даје најбољу равнотежу подприлагођавања (*underfitting*) и преприлагођавања (*overfitting*), уз најбољу прецизност на тест скупу података. Поступак одређења праве величине простора је веома комплексан и није једнозначан.

Имплементација претходне идеје подразумева програмски оквир који за аргументе узима алгоритам обучавања (на пример MLR) и простор хипотеза одређене димензије. Оквир набраја моделе према параметру *величина*. За сваку величину се користи унакрсна оцена ваљаности и израчунава просечан износ грешке над скупом за обучавање и скупом за тестирање. Почиње се од најједноставнијих, најмањих модела (који су вероватно подприлагођени подацима) и реализује се итеративни поступак, разматрајући на сваком кораку све сложеније моделе, док модели не почну да преприлагођавају (слика 2.7).



Слика 2.7. Утицај величине простора хипотеза на *overfitting*

¹⁷ Cross-validation

¹⁸ Leave-One-Out Cross-Validation

2.3.3. Губитак

Код *надгледаног обучавања*, које се користи у оквиру ове дисертације, се тежи максимизацији очекиване *корисности*. Уобичајено је да се *корисност* изражава помоћу тзв. *функције губитка*.

Функција губитка $L(X, Y, \hat{Y})$ се дефинише као количина корисности изгубљене због предвиђања да је $\hat{y} = h(x)$ када је исправан одговор $y = f(x)$ (2.4).

$$\begin{aligned} L(X, Y, \hat{Y}) = & \text{Korisnost}(\text{rezultat koriscenja } Y \text{ ako je ulaz } X) \\ & - \text{Korisnost}(\text{rezultat koriscenja } \hat{Y} \text{ ako je ulaz } X) \end{aligned} \quad (2.4)$$

Често се користи упрошћена варијанта $L(Y, \hat{Y})$ која не зависи од X .

Као пример се наводи класификатор *spam*-а у *e-mail* преписци. Губитак корисности ако се *лоша* порука окарактерише као *добра* се може квантификовати са, на пример, $L(\text{Otpad}, \text{NijeOtpad}) = 1$, док би губитак корисности, ако се оно што није отпад прогласи отпадом могао бити 10 пута већи: $L(\text{NijeOtpad}, \text{Otpad}) = 10$.

Неке од функција које реализују логику губитка су: апсолутна вредност разлике (губитак L_1), квадрат разлике (губитак L_2) и губитак $L_{0/1}$. Функција губитка $L_{0/1}$ је намењена за дискретне дихотомне променљиве и има вредност 0 ако је одговор тачан. Ове функције губитка се рачунају према следећем:

- Апсолутна вредност губитка: $L_1(y, \hat{y}) = |y - \hat{y}|$
- Квадрат грешке губитка: $L_2(y, \hat{y}) = (y - \hat{y})^2$
- Губитак 0/1: $L_{0/1}(y, \hat{y}) = 0 \text{ ако је } y = \hat{y}, \text{ иначе } 1$

Из претходних једначина се види да функција губитка L_2 одговара класичној методи најмањих квадрата.

При математичком моделирању се може теоретски максимизовати *очекивана корисност* бирањем хипотезе која минимизује *очекивани губитак* на свим паровима улаз/излаз који се могу појавити. Нека је $\mathcal{E}$ скуп свих могућих примера улаз/излаз. Тада је *очекивани губитак уопштавања* за хипотезу h једнак:

$$\text{GenGub}_L(h) = \sum_{(x,y) \in \mathcal{E}} L(y, h(x)) \cdot P(x, y) \quad (2.5)$$

где је $P(x, y)$ априорна вероватноћа хипотезе, а најбоља хипотеза, h^* , је она са минимумом очекиваног губитка:

$$h^* = \arg \min_{h \in \mathcal{H}} \text{GenGub}_L(h) \quad (2.6)$$

Пошто вероватноћа $P(x, y)$ није позната, може се само проценити *губитак уопштавања* са *емпиријским губитком* на основу скупа примера из узорка E :

$$\text{EmpGub}_{L,E}(h) = \frac{1}{N} \sum_{(x,y) \in E} L(y, h(x)) \quad (2.7)$$

Процењена најбоља хипотеза, h^* , је тада она са минималним емпиријским губитком:

$$\hat{h}^* = \arg \min_{h \in \chi} EmpGub_{L,E}(h) \quad (2.8)$$

Разлози због којих $\hat{h}^*$ може да се разликује од стварне функције f су:

- *неостварљивост*
 - f може да не буде у χ .
 - f може да има малу вероватноћу.
- *варијанса*
 - алгоритам обучавања ће вратити различите h за различите *скупове примера*, што доводи до веће или мање варијансе у предвиђањима.
 - варијанса ће постојати увек јер је реч о случајним променљивим; биће мања ако је број примера већи.
- *шум*
 - увек постоји у стохастичким процесима, а већина процеса у реалности је таква.
- *рачунарска сложеност*
 - када је простор χ сложен, могло би да буде рачунарски несавладиво да се систематски претражи цео простор хипотеза.
 - могуће је локално претраживање, али се тиме може заобићи глобални оптимум¹⁹.

За простор хипотеза и обучавање је везан и појам *димензија*.

За *обучавање малих димензија* је карактеристично:

- везано је за традиционалне проблеме у статистици и прве године машинског обучавања (број примера за обучавање: 10 до неколико хиљада).
- у оваквим моделима на грешку уопштавања утиче грешка у избору h , а не да ли се права f налази у простору хипотеза.

Обучавања већих димензија су везана за последње декаде истраживања у домену AI и математичког моделирања, а одређена су поред осталог и са:

- милионима примера за обучавање
- овде на грешку уопштавања утичу ограничења рачунања
- података има довољно, а простор је довољно богат, да се може пронаћи h која је довољно блиска f , али је рачун за проналажење превише сложен, па се често компромис налази у субоптимизацији.

2.3.4. Трошак. Регуларизација

Као што је раније описано, један од начина избора најповољнијег математичког модела/хипотезе је базиран на унакрсној провери и трагању за оптималном величином модела. Другим речима, избор најповољнијег математичког модела је базиран на минимизацији емпиријског губитка. Алтернативни метод је минимизација *укупног трошка*. Под укупним трошком се подразумева пондерисани збир емпиријског губитка и сложености хипотезе:

¹⁹ Један од кључних разлог за примену GA хеуристике у овој дисертацији је трагање за глобалним оптимумом и стварање платформе за моделирање реалних проблема који имају елементе рачунарске несавладивости

$$Trosak(h) = EmpGub(h) + \lambda \cdot Slozenost(h) \quad (2.9)$$

Ово значи да се трага за хипотезом за коју важи да је:

$$\hat{h}^* = \arg \min_{h \in \mathcal{H}} Trosak(h) \quad (2.10)$$

где је λ број који служи као фактор конверзије и уједначава димензије губитка и сложености. Дакле, овде се поред величине хипотезе трага и за λ . Бира се λ које даје најбоље оцене у тестном скупу података.

Претходно описани поступак експлицитног кажњавања сложених хипотеза се зове *регуларизација* (тражи се функција која је регуларнија, односно мање сложена) [42].

Избор функције регуларизације зависи од простора хипотеза. На пример, један од облика регуларизације за полиноме је збир квадрата коефицијената. Ако се модел развија тако да тај збир буде што мањи, избегавају се таласасте полиноми, односно превелико одступање између чворних тачака.

Ова техника је разрађена и анализирана у посебној класи регресија које се називају *робустне регресије*²⁰ [17].

Поред наведеног на регуларизацију битно могу утицати елементи редукције димензија (факторска анализа: PCA) и програмирана редукција скупа примера за обучавање као што је kNN платформа, kMean кластеризација и друге технике.

2.4. Развојни оквир за избор хипотеза базиран на MLR и GA. Шематски приказ

У претходним деловима овог поглавља су дефинисани појмови неопходни за квалитетну реализацију процеса математичког моделирања. Дефинисан је *простор хипотеза* и описане одреднице као што су *величина*, *корисност*, *губитак*, *трошак*, *регуларизација*, *базисне функције*, *сложеност*, *рачунарска несавладаивост* (*неукротивост*)... Сви ови елементи су неопходни за синтезу знања о објекту истраживања.

Ипак, кибернетски приступ је неопходно преточити у модуле које је могуће практично развити и имплементирати. На слици 2.8 је приказан развојни оквир који омогућава решавање таквог задатка у креирању и развоју математичких модела објекта истраживања.

На првом нивоу, односно нивоу података, по мишљењу аутора, су могућа два приступа, опредељена количином информативне грађе о објекту истраживања. Уколико је реч о новим истраживањима препоручује се примена DOE²¹ приступа (слика 2.8-А). Концепт планираног дизајнирања експеримената обезбеђује добре перформансе са аспекта броја експеримената, времена и утрошених средстава, са једне стране, и количине добијених информација о објекту, са друге стране. Изузетно је погодан за добијање почетних модела заснованих на лабораторијским испитивањима. Везан је за генерисање довољно добрих хипотеза и *обућавање малих димензија*. Адаптирање и побољшање математичког модела се реализује током животног циклуса како модела

²⁰ Robust regression

²¹ Design Of Experiments – DOE; приказ теорије и софтвер за примену пуних факторних експерименталних планова (*Full factorial Central Composite Design plan / CCD*), аутор је развио у оквиру магистарске тезе [1]

тако и објекта моделирања. Ово адаптирање је најчешће везано за реалне услове експлоатације (индустријски услови, реални процеси...) са континуалном аквизицијом података о битним улазно/излазним карактеристикама самог објекта/процеса. У зрелом стадијуму модела, методе и технике се постепено трансформишу у концепт Б.

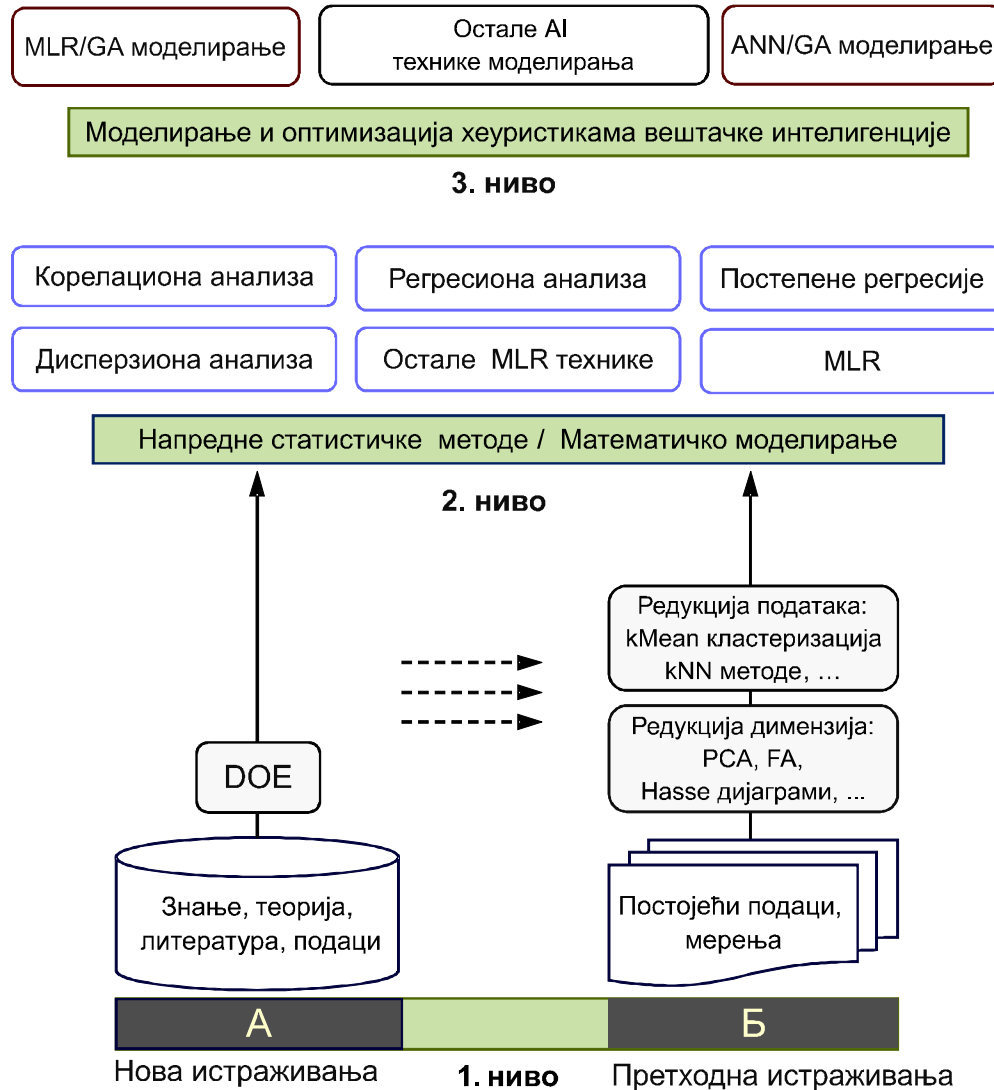
Концепт Б (слика 2.8-Б) се препоручује у случајевима када о објекту моделирања постоји значајна количина података, односно када улазно/излазни вектор примера за обучавање има значајне димензије (понекад је у питању обучавање великих димензија). Скуп података за обучавање се дели на два подскупа: скуп података за учење и скуп података за тестирање. Скуп података за тестирање може обезбедити спречавање *overfitting*-а и тиме дати добро уопштавање. У варијанти Б, у случају потенцијалних проблема као што су неостварљивост, варијанса, шум или рачунарска сложеност, аутор препоручује смањење димензионалности, редукцијом димензија и/или примену парадигме *сличних тачака* у циљу редукције скупа података за обучавање.

За редукцију димензија односно факторску анализу, истраживачи могу користити различите технике: *анализу главних фактора*, *факторизацију методом максималне веродостојности*, *алфа факторизацију*, *непондерисану методу најмањих квадрата* итд. У оквиру ове дисертације коришћена је РСА техника. Поред наведених, су могуће и редукције димензија применом идеја као што су симплификације базиране на Hasse-овим дијаграмима и друге технике. Редукција обима података и примера за обучавање се може реализовати различитим методама и техникама које у себе укључују идеју *блиских тачака*. Неке од ових метода су методе најближих суседа (kNN технике), класичне методе кластеризације (на пример *kMean clustering*), или њихова AI хибридна унапређења (kMean/GA).

На другом нивоу развојног оквира, се реализује додатно предпроцесирање података у смислу провере аутокорелација, уклањања нетипичних тачака²² или системских грешака које нису уклоњене на првом нивоу. Такође, на другом нивоу, за креирање и адаптацију математичких модела, се ангажују стандардне и напредне статистичке технике. Овде се под стандардним техникама подразумева примена корелационе, регресионе и дисперзионе анализе. Напредне технике подразумевају моделирање са MLR које су унапређене разноврсним техникама. Пример оваквих техника је моделирање помоћу постепених регресија.

²² outliers

Трећи ниво подразумева примену AI хибридних техника моделирања, базираних на GA оптимизованим вишеструким линеарним регресијама и вештачким неуронским мрежама које су оптимизоване помоћу GA хеуристике.



Слика 2.8. Развојни оквир математичког моделирања објеката и система

Код MLR/GA хибрида, се избор броја и типа регресора третира кроз оптимизационе GA алгоритме. На овај начин се тежи креирању регресионих модела у којима је примењена идеја кажњавања сложености односно смањења трошка. Другим речима, за одабрану класу проблема циљ је генерисање довољно једноставних хипотеза у облику MLR које добро уопштавају. ANN/GA хибриди обухватају примену GA оптимизационих техника за нелинеарне моделе. GA оптимизација ANN јединки је опсежна и обухвата како топологију и алгоритме учења, тако и активационе функције и правила учења.

Предложеним оквиром се могу остварити постављени циљеви ове дисертације. Конкретна имплементација модула и валидација перформанси софтверских хибрида развијених у оквиру тезе, је реализована применом Б-концепта развојног оквира за математичко моделирање објеката, процеса и система.

3. Регресија

3.1. Основна формулација – регресиони модел

Регресиона анализа је једна од најзначајнијих метода статистичке анализе са применом у пракси. Она омогућава предвиђање и оцењивање међусобних веза и утицаја две или више појава у квантитативном погледу. Појам *регресија* је увео енглески научник Francis Galton 1885. год. [12].

За разлику од интерполационих модела, регресиони модел покушава да на „најприроднији начин” опише одговарајући процес или систем. Овакав приступ је у пракси веома често присутан, јер даје боље резултате у односу на интерполационе методе за тачке које не припадају скупу примера за обучавање. Дакле, код регресионог модела, *хиперповрш* не пролази нужно кроз задате тачке, али је најчешће сума квадрата одступања појединачних тачака од регресионог модела минимална. Овај концепт се, као што је већ поменуто, назива методом најмањих квадрата (губитак L_2), а присутан је у науци скоро 200 година.

3.2. Проста линеарна регресија

За случај једнофакторних објеката и система (или оних код којих једна променљива доминантно утиче на излазну променљиву), *дијаграм расипања* указује да ли између независне променљиве (x) и зависне променљиве (y) постоји линеарна веза, односно да ли је реч о простом линеарном регресионом моделу.

Хипотеза линеарног модела се може записати у облику:

$$h_w(x) = w_0 + w_1 \cdot x \quad (3.1)$$

Кључни корак регресионе анализе је *учење* коефицијената w_0 и w_1 , који дефинишу вектор тежина $\mathbf{w} = [w_0 \ w_1]$, са циљем да се пронађе најприлагођеније h_w односно да се минимизира емпиријски губитак.

Гаус је доказао да, ако зависна променљива (y_j) има нормалну расподелу шума, онда се најприлагођеније $\mathbf{w}$ добија минимизовањем збира квадрата грешака¹. Ово значи да је уобичајено да се користи функција квадрата губитка, L_2 , сумирана по свим примерима обучавања

$$Gubitak(h_w) = \sum_{j=1}^N L_2(y_j, h_w(x_j)) = \sum_{j=1}^N (y_j - h_w(x_j))^2 = \sum_{j=1}^N (y_j - (w_1 \cdot x_j + w_0))^2 \quad (3.2)$$

где је N , број примера за обучавање.

Трага се за

¹ Least squered method – метод најмањих квадрата

$$\mathbf{w}^* = \arg \min_{\mathbf{w} \in \chi} \text{Gubitak}(h_{\mathbf{w}}) \quad (3.3)$$

што се добија парцијалним изводима:

$$\frac{\partial}{\partial w_0} \sum_{j=1}^n (y_j - (w_1 \cdot x_j + w_0))^2 = 0 \quad \text{и} \quad \frac{\partial}{\partial w_1} \sum_{j=1}^n (y_j - (w_1 \cdot x_j + w_0))^2 = 0.$$

Овај систем има јединствено решење:

$$w_1 = \frac{\bar{y} \cdot \sum x - \sum y \cdot \bar{x}}{\bar{x} \cdot \sum x - \sum x^2} \quad \text{и} \quad w_0 = \frac{\sum y - w_1 \cdot \sum x}{n} \quad (3.4)$$

где $\bar{x}$ и $\bar{y}$, представљају аритметичке средине независне и зависне променљиве, респективно.

Многе врсте обучавања подешавају тежине да би се минимизирао губитак. Детаљи који су везани за просте линеарне регресије опширније су дати у [12-13].

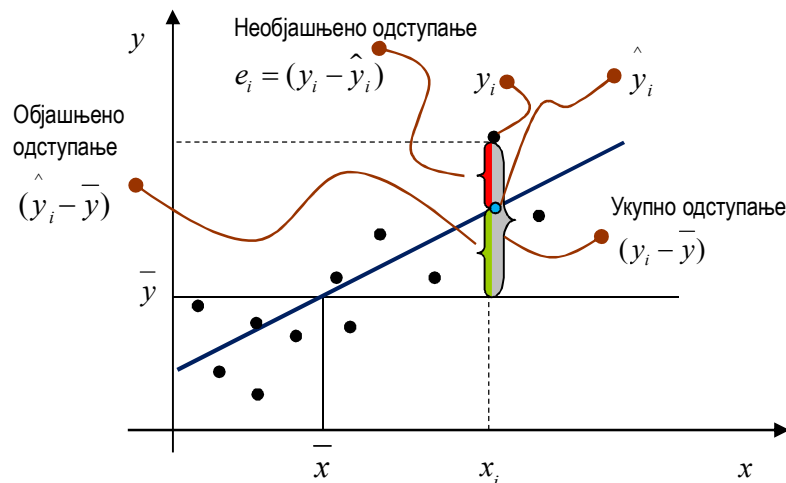
3.3. Мере репрезентативности линеарне регресије

Мере репрезентативности регресије дефинишу прилагођеност регресије емпиријским подацима. Издвајају се два показатеља од којих је један апсолутни, а други релативни по својој природи. Ти показатељи су [12-13]:

- *стандардна грешка регресије* - која представља апсолутну меру одступања емпиријских тачака од моделских вредности и
- *коефицијент детерминације* - који је релативни показатељ.

Ове две мере су последица варијабилитета у регресионом моделу (слика 3.1).

Анализа варијабилитета се заснива на чињеници да, ако сваки од парова (x_i, y_i) из узорка задовољава у одређеној мери једначину (3.1), онда то важи и за аритметичке средине података из узорка $(\bar{x}, \bar{y})$. Дакле, и аритметичке средине задовољавају утврђену линеарну регресију. Управо је одступање појединачних вредности y_i од аритметичке средине $\bar{y}$ основ анализе.



Слика 3.1. Укупно, објашњено и необјашњено одступање

То укупно одступање се састоји од два дела:

$$y_i - \bar{y} = \hat{y}_i - \bar{y} + y_i - \hat{y}_i \quad (3.5)$$

Укупно Објашњено Необјашњено
одступање одступање одступање (резидуал)

где је $\hat{y}_i$ представља вредност која одговара регресионом моделу (моделска вредност, предикција), а значења осталих чланова једначине су описана раније.

Претходне једначине важе и када се обе стране квадрирају и сумирају за све вредности у узорку [12-13].

$\sum (y_i - \bar{y})^2 = \sum (\hat{y}_i - \bar{y})^2 + \sum (y_i - \hat{y}_i)^2$		
SKU	SKO	SKN
Укупна сума квадрата	Објашњена сума квадрата	Необјашњена сума квадрата
Укупни варијабилитет	Објашњени варијабилитет	Необјашњени варијабилитет
	Регресиона сума квадрата	Резидуална сума квадрата –сума квадрата грешке

На основу претходне једначине се долази до мера репрезентативности и то до: стандардне грешке регресије - засноване на SKN односно на необјашњеном варијабилитету (резидуална сума квадрата) и коефицијенту детерминације - који је заснован на поређењу SKO и SKU (дакле односу: објашњеног и укупног варијабилитета).

3.3.1. Коефицијент детерминације – релативна мера репрезентативности

Коефицијент детерминације (R^2) описује проценат учешћа објашњеног варијабилитета у укупном варијабилитету: $\frac{SKO}{SKU} = 1 - \frac{SKN}{SKU}$ односно:

$$R^2 = 1 - \frac{\sum (y_i - \hat{y}_i)^2}{\sum (y_i - \bar{y})^2} = w_1^2 \cdot \frac{\sum x_i^2 - n \cdot \bar{x}^2}{\sum y_i^2 - n \cdot \bar{y}^2} \quad (3.6)$$

где је w_1 оцена нагиба регресионе праве.

Коефицијент детерминације представља релативну меру репрезентативности. Из формуле за израчунавање је јасно да се вредност овог коефицијента креће у границама $[0,1]$ односно да је $0 \leq R^2 \leq 1$ тако да су карактеристични следећи случајеви:

$R^2 \rightarrow 0$	све је мање објашњеног варијабилитета и линеарна регресија све лошије описује дати скуп
$R^2 = 0$	не постоји линеарна зависност
$R^2 = 1$	функционална зависност

3.4. Вишеструка линеарна регресија

Линеарни регресиони модели који садрже више од једне предикторске променљиве су вишеструки линеарни регресиони модели - MLR. Једначина (3.7).

$$\eta = \beta_0 + \beta_1 \cdot x_1 + \beta_2 \cdot x_2 \quad (3.7)$$

је пример вишеструке линеарне регресије *првог реда* са два предиктора, x_1 и x_2 , и функцијом одзива (површи реаговања – *response surface*) η . Линеарност се односи на коефицијенте β_1 и β_2 , а не на променљиве. Коефицијенти β_1 и β_2 се називају *парцијалним регресионим коефицијентима*. Приказана регресија је првог реда јер је максимални број са којим су степеноване променљиве 1.

Наравно, MLR могу имати и сложеније облике. Једначина (3.8) приказује MLR која садржи члан са међудејствима предикторских променљивих:

$$\eta = \beta_0 + \beta_1 \cdot x_1 + \beta_2 \cdot x_2 + \beta_{1,2} \cdot x_1 \cdot x_2 \quad (3.8)$$

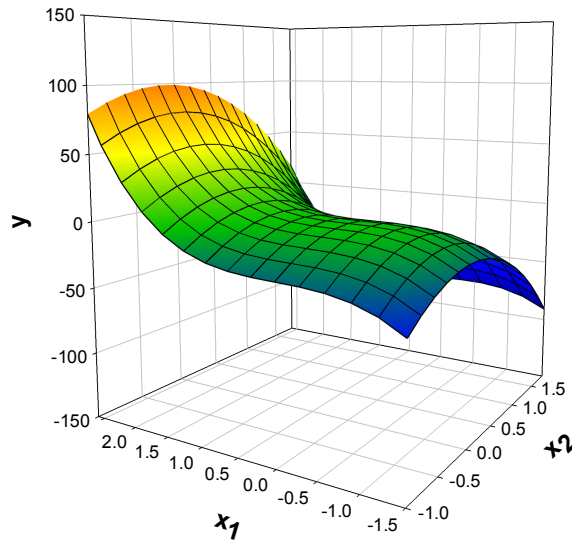
Функција одзива регресије трећег реда са два фактора, би имале форму:

$$\eta = \beta_0 + \sum_{i=1}^2 \beta_i \cdot x_i + \beta_{1,2} \cdot x_1 \cdot x_2 + \beta_{1,22} \cdot x_1 \cdot x_2^2 + \beta_{11,2} \cdot x_1^2 \cdot x_2 + \sum_{i=1}^2 \beta_{i,i} \cdot x_i^2 + \sum_{i=1}^2 \beta_{i,i} \cdot x_i^3 \quad (3.9)$$

а графички приказ једне од таквих регресија,

$$y = 0.12 \cdot x_1 - 23.4 \cdot x_2^2 - 1.45 \cdot x_1 \cdot x_2 + 5.56 \cdot x_1^3 \quad (3.10)$$

је дат на слици 3.2.



Слика 3.2. Пример MLR са две независне променљиве

Генерално, регресори (чланови регресије) могу имати сложеније облике као што је приказано у једначини (3.10)

$$\eta = \beta_0 + \beta_1 \cdot \sqrt{x_1} + \beta_2 \cdot \ln(x_2) + \beta_3 \cdot x_1 \cdot \sin(x_2) \cdot \sqrt{x_3} + \beta_4 \cdot x_2^2 + \dots + \beta_5 \cdot x_1 \cdot x_3^3 + \dots \quad (3.10)$$

Регресори, односно *базисне форме* f_i различитих облика се увек могу заменити новим линеарним променљивим, тако да модел задржава линеарност.

Број регресора d може бити већи од броја предикторских променљивих k .

Уз претпоставку да је $f_0 = x_0 = 1$, и да $\mathbf{X}^T = [x_1 \ x_2 \ \dots \ x_{k-1} \ x_k]$ означава транспоновану матрицу колоне независних променљивих x_i , $i = 1, k$, MLR се може записати у концизној форми:

$$\eta = \eta(\mathbf{X}, \boldsymbol{\beta}) = \eta(x_0, x_1, x_2, \dots, x_k, \beta_0, \beta_1, \dots, \beta_d) = \sum_{i=0}^d \beta_i \cdot f_i(\mathbf{X}^T) \quad (3.11)$$

Једначина представља теоријску MLR, а њени коефицијенти $\boldsymbol{\beta} = [\beta_1 \ \beta_2 \ \dots \ \beta_d]^T$, теоријске коефицијенте регресије. Непознате вредности коефицијената β_i , $i = 0, d$ процењују се на основу експерименталних резултата (комбиновани метод описан у 2.2.2.) вредностима b_i , $i = 0, d$ који опредељују емпиријски регресиони модел:

$$\hat{y} = \hat{y}(x_0, x_1, x_2, \dots, x_k, b_0, b_1, \dots, b_d) = \sum_{i=0}^d b_i \cdot f_i(\mathbf{X}^T) \quad (3.12)$$

Регресиони модел (3.12) важи, у одређеној мери, за сваки од n експерименталних резултата.

У развијеној форми емпиријски модел вишеструке линеарне регресије тако поприма облик:

$$\begin{aligned} y_1 &= b_0 \cdot f_{10} + b_1 \cdot f_{11} + b_2 \cdot f_{12} + \dots + b_d \cdot f_{1d} + \varepsilon_1 \\ y_2 &= b_0 \cdot f_{20} + b_1 \cdot f_{21} + b_2 \cdot f_{22} + \dots + b_d \cdot f_{2d} + \varepsilon_2 \\ &\dots \\ y_n &= b_0 \cdot f_{n0} + b_1 \cdot f_{n1} + b_2 \cdot f_{n2} + \dots + b_d \cdot f_{nd} + \varepsilon_n \end{aligned} \quad (3.13)$$

чија је матрична форма (3.14) далеко практичнија за употребу:

$$\mathbf{Y} = \mathbf{F} \cdot \mathbf{B} + \boldsymbol{\varepsilon} \quad (3.14)$$

У моделу (3.14) матрица $\mathbf{F}$ се често назива *дизајн матрицом* (*design matrix*), јер садржи информације о нивоима (вредностима) предикторских променљивих за које је добијена вредност зависне променљиве.

Ако се узме у обзир стохастичност изучаваног објекта или система, јасно је да је $\boldsymbol{\beta} = E(\mathbf{B})$ и $\boldsymbol{\eta} = E(\mathbf{Y})$, где је $\mathbf{B} = [b_0 \ b_1 \ b_2 \ \dots \ b_d]^T$, вектор колоне коефицијената емпиријске регресије, а $\mathbf{Y} = [y_1 \ y_2 \ y_3 \ \dots \ y_n]^T$, вектор колоне експерименталних резултата. Коефицијенти емпиријске регресије се одређују, као и код прости линеарне регресије, методом најмањих квадрата (губитак L_2), применом матричног рачуна. Овом оптималном методом за дати план експеримената $\mathbf{X}$, односно матрицу дизајна $\mathbf{F}$, резултате $\mathbf{Y}$ и матрицу колону моделских вредности зависне променљиве $\hat{\mathbf{Y}}$, се добија минимална дисперзија s_{bi}^2 коефицијената регресије b_i . Суштину методе најмањих квадрата, чини минимум *резидуалне суме квадрата*. Резидуална сума $\mathbf{S}_R$ (3.16), се одређује преко резидуалне матрице колоне (3.15):

$$\mathbf{R} = \mathbf{Y} - \hat{\mathbf{Y}} = \mathbf{Y} - \mathbf{F} \cdot \mathbf{B} \quad (3.15)$$

$$\mathbf{S}_R = \mathbf{R}^2 = (\mathbf{Y} - \mathbf{F} \cdot \mathbf{B})^T \times (\mathbf{Y} - \mathbf{F} \cdot \mathbf{B}) \quad (3.16)$$

Метода најмањих квадрата подразумева минималну вредност ове суме, те се непознати коефицијенти емпиријске регресије одређују њеним парцијалним диференцирањем по матрици $\mathbf{B}$ и изједначавањем добијеног израза са нула матрицом.

$$\frac{\partial \mathbf{S}_R}{\partial \mathbf{B}} = 0 \quad (3.17)$$

Након примене техника матричне алгебре, се добија коначна форма за одређивање вредности емпиријских регресионих коефицијената:

$$\mathbf{B} = (\mathbf{F}^T \cdot \mathbf{F})^{-1} \cdot \mathbf{F}^T \cdot \mathbf{Y} \quad (3.18)$$

Када је одређена матрица $\mathbf{B}$, на основу једначине (3.18), се може поставити коначни облик оцењеног модела MLR (*fitted model*):

$$\hat{\mathbf{Y}} = \mathbf{F} \cdot \mathbf{B} \quad (3.19)$$

На основу једначина (3.18) и (3.19) регресиони модел се може приказати у различитим формама:

$$\begin{aligned} \hat{\mathbf{Y}} &= \mathbf{F} \cdot \mathbf{B} \\ \hat{\mathbf{Y}} &= \mathbf{F} \cdot (\mathbf{F}^T \cdot \mathbf{F})^{-1} \cdot \mathbf{F}^T \cdot \mathbf{Y} \\ \hat{\mathbf{Y}} &= \mathbf{H} \cdot \mathbf{Y} \end{aligned} \quad (3.20)$$

Матрица $\mathbf{H}$ је такозвана *hat* матрица². Она трансформише матрицу колону мерених резултата ($\mathbf{Y}$) у матрицу колону моделских вредности ($\hat{\mathbf{Y}}$).

За ближе описивање коефицијената MLR модела користе се различите врсте тестова. Сваки од њих се може реализовати само ако је испуњена претпоставка нормалности случајне грешке (са стандардизованом средњом вредношћу једнаком нули и стандардном девијацијом $\sigma = 1$). Постоје три типа тестова:

- тестирање значајности целог регресионог модела;
- тестирање регресионих коефицијената појединачно - t тест (Student-ов тест) и
- тестирање значајности броја регресионих коефицијената, али и појединачних коефицијента (F тест).

Детаљнији приказ наведених тестова је дат у [12, 13, 16].

3.5. Мере квалитета регресионих модела. Кориговани коефицијент детерминације

Квалитет вишеструке линеарне регресије, којом се жели описати објекат или систем истраживања, се може мерити на различите начине. Једна од мера је слична мери репрезентативности просте линеарне регресије. Реч је о *коефицијенту вишеструке детерминације* који је дефинисан као однос објашњених и укупних одступања:

$$R^2 = \frac{SS_R}{SS_T} = 1 - \frac{SS_E}{SS_T} = 1 - \frac{\sum (y_i - \hat{y}_i)^2}{\sum (y_i - \bar{y})^2} = \frac{\sum_{i=1}^n (\hat{y}_i - \bar{y})^2}{\sum_{i=1}^n (y_i - \bar{y})^2} \quad (3.21)$$

У једначини (3.21), SS_R представља регресиону суму квадрата, SS_T суму квадрата укупних одступања, а SS_E суму квадрата резидуала, док y_i , $\hat{y}_i$ и $\bar{y}$, означавају i -то

² *hat matrix*

мерење, предикцију тог мерења и аритметичку средину свих мерења, респективно. Ова мера квалитета модела није најпогоднија јер се са повећањем броја регресора, R^2 повећава, чак и када се у модел додаје регресор који није значајан. Ово значи да повећање R^2 код MLR модела не гарантује да је модел и квалитетнији.

Мера квалитета која је погоднија је *кориговани коефицијент вишеструке линеарне детерминације*³:

$$R_{adj}^2 = 1 - \frac{MS_E}{MS_T} = 1 - \frac{\frac{SS_E}{(n-(d+1))}}{\frac{SS_T}{(n-1)}} = 1 - \frac{(n-1)}{(n-(d-1))} \cdot (1 - R^2) \quad (3.22)$$

где је MS_E средња грешка модела, MS_T средње укупно одступање, док n представља број примера за обучавање, а d број регресора у регресионом моделу.

R_{adj}^2 се повећава само ако се у регресиони модел дода значајан регресор. Додавање чланова који нису значајни може довести до пада вредности *коригованог коефицијента вишеструке линеарне детерминације*.

Поред наведених, за квалитет модела се користе и следеће мере: RMSE- корен средње квадратне грешке; s - стандардна грешка модела⁴; r^* - коефицијент вишеструке линеарне корелације⁵; *PRESS* - предикција суме квадрата грешака⁶; R_{pred}^2 - предикција коефицијента R^2 заснована на мери квалитета *PRESS*; и друге мере. Мере RMSE, R^2 , R_{adj}^2 , s и r^* се односе на квалитет модела, док се мере *PRESS* и R_{pred}^2 односе на квалитет предикција нових посматрања. Детаљни приказ наведених параметара којима се може оцењивати квалитет регресионих модела је дат у [16]. Свака од ових мера, понаособ или комбиновано, може допринети квалитетном генерисању корективних фактора у поступку регуларизације, који је описан у делу 2.3.4.

У оквиру ове докторске дисертације је најчешће, за оцену квалитета модела, коришћен корен средње квадратне грешке, рачунат над тест скупом података ($RMSE_{Test}$). Такође је коришћен и кориговани коефицијент вишеструке детерминације R_{adj}^2 , за одређивање *fitness*-а вишеструких линеарних регресија, при моделирању и оптимизацији процеса и система са MLR/GA хибридом.

Поред наведених величина које дефинишу експлицитну, квантитативну меру квалитета регресионих модела, на квалитет регресионих модела утиче и читав низ других параметара и претпоставки. У наредном тексту се наводе неки од њих.

Обим узорка

Вишеструка линеарна регресија спада у породицу веома осетљивих статистичких техника. Заснована је на више претпоставки о подацима и може бити непоуздана ако су ове претпоставке нарушене. Не може се употребљавати на малим узорцима, нити када је расподела резултата веома асиметрична⁷. Овде је реч способности модела да добро

³ *adjusted coefficient of multiple determination*

⁴ *standard error of the model*

⁵ *multiple correlation coefficient*

⁶ *PRrediction Error Sum of Squares*

⁷ Постоје различите математичке технике трансформације за побољшање симетричности расподеле резултата мерења (пресликавање у Нормални закон расподеле). О коришћењу ових метода постоје дијаметрално различита мишљења у научним круговима.

уопштава о чему је било речи у деловима 2.3.1. и 2.3.2. Разни аутори дају различите препоруке броја примера за обучавање при генерисању модела вишеструке линеарне регресије. Stevens [43], за истраживања у друштвеним наукама, предлага 15 мерења, односно примера по предиктору. Tabachnick и Fidell [10], дају формулу за израчунавање величине узорка: $n = 50 + 8 \cdot m$, где је m број независних променљивих. Када је зависна променљива асиметрична потребно је још више примера. За постепене регресије препоручује се 40 примера обучавања по независној променљивој [9, 10].

Квалитетнијем моделирању помоћу MLR, такође, могу допринети програмирана сажимања скупова примера за обучавање, која су конципирана према методологији *k-најближних суседа* (*kNN- k Nearest Neighbours*) као и концептима кластеризације (*kMean clustering*).

Корелација предиктора и зависне променљиве

Квалитетан регресиони модел подразумева да између зависне променљиве и независних променљивих постоји макар и слаба веза. Било би пожељно да *линеарна корелација* буде изнад 0.3 за сваки пар (независна променљива, зависна променљива) [9, 10].

Сингуларност

Вишеструки линеарни регресиони модели су осетљиви на сингуларност. Сингуларност настаје када је једна независна променљива заправо комбинација других независних променљивих тј. када уопште није независна [9, 10].

Мултиколинеарност

На квалитет регресионог модела значајно може утицати *мултиколинеарност*⁸. Мултиколинеарност значи да постоји јака зависност међу предикторским променљивим. У оваквим моделима обрачун коефицијента значајно зависи од тога које улазне променљиве су додатно укључене у модел, што у крајњем може довести и до промене знака неког од њих. Student-ов *t* тест и Fisher-ов парцијални *F* - тест, у регресијама са мултиколинеарношћу треба узимати са резервом. У исту анализу не би требало истовремено укључивати независне променљиве чија је линеарна корелација 0.7 и више [9, 10]. Овакве ситуације налажу да се једна од корелираних променљивих изостави, или да се од две променљиве формира изведена променљива, која ће ући у регресиони модел. Мултиколинеарност се не може у потпуности одредити Pearson-овом корелационом матрицом, па су уведене и додатне мере за проверу мултиколинеарности: *толеранција* (*Tolerance*), и *фактор повећања варијансе* (*VIF*)⁹. Детаљна анализа и препоруке за примену ових параметара су дате у [10, 9, 16].

Неке од метода које потенцијално могу решити проблем мултиколинеарности су: уклањање предикторских променљивих које нису значајне из модела, сажимање више зависних предикторских променљивих у једну променљиву, редизајн матрице **F**, односно репозиционирање мерних тачака у хиперпростору истраживања, и кодирање предикторских променљивих. Примери моделирање са кодираним предикторским променљивим су дати детаљно у [1]. Кодирањем се, потенцијално, могу избећи проблеми у одређивању дисперзионе матрице, $\mathbf{C} = (\mathbf{F}' \cdot \mathbf{F})^{-1}$, а тиме и проблеми у оцени регресионих коефицијената на бази *F* распореда.

Проблем мултиколинеарности је везан и за могућност редукције димензија. Другим речима, квалитетној регуларизацији и избору консистентних хипотеза (део

⁸ *multicollinearity*

⁹ *the Variance Inflation Factor*

2.3.2) у случајевима изражене мултиколинеарности може допринети и редукција димензија. Једна од техника које омогућавају редукцију димензија, техника анализе главних компоненти, је приказана у делу 9.1.

*Нетипичне вредности*¹⁰

За квалитет регресионог модела битно је да улазни подаци и резултати мерења буду у очекиваним границама. Нетипичне вредности (веома велике или веома мале вредности) могу значајно утицати на квалитет регресионог модела, па их због тога треба открити, уклонити или модификовати пре примене MLR моделирања. Ово треба урадити за све променљиве које ће се употребљавати у регресионој анализи, дакле и за зависне и за независне променљиве.

Нетипичне тачке се могу детектовати на разне начине: помоћу *трага hat* матрице односно њених дијагоналних елемената (h_{ii}) [16], помоћу дијаграма у облику правоугаоника (*boxplot*) [9, 10] и на друге начине.

Утицај нетипичних тачака и екстремних вредности улазних и излазних променљивих се може делимично оценити и помоћу мере 5% *тримоване средине* (5% *Trimmed Mean*), њеним поређењем са класичном аритметичком средином [9].

Постојање нетипичних тачака се може утврдити и методама резидуалне анализе. Резидуална анализа даје додатну графичку представу о фитовању регресионог модела у облику дијаграма резидуала. *Стандардизовани резидуали* d_i су погодни за откривање системских грешака у мерењима и генерално нетипичних резултата. Дијаграм стандардизованих резидуала (*Normal Probability Plot: Normal P-P*), као и дијаграм стандардизованих резидуала у облику *ScatterPlot*, које уобичајено генеришу већина статистичких пакета, омогућавају да се визуелно утврде *outliers* тачке. Стандардизовани резидуали, као мера квалитета регресионог модела, имају и недостатке јер могу да доведу до погрешне процене стварног увећања резидуала. Због тога се у детаљним анализама могу користити и друге мере као што су: *студентизовани резидуали*¹¹ (r_i) и *спољни студентизовани резидуали*¹² (t_i). Спољни студентизовани резидуал покушава додатно да утврди утицај нетипичних тачака на *фитовање* регресионог модела. Детаљи и препоруке за откривање нетипичних тачака помоћу стандардизованих резидуала, студентизованих резидуала и спољних студентизованих резидуала, су дати у [10, 9, 16].

Поред тога што је, за боље фитовање модела, значајно да се утврди постојање *outliers* тачака и које су то тачке, значајно је да се утврди и степен њиховог утицаја на регресиони модел. Једна од мера која то обезбеђује је Cook-ово растојање¹³. Детаље о препорукама и примени Cook-овог растојања у откривању нетипичних тачака видети у [16, 10, 9].

3.6. Врсте регресија

Моделирање објеката, процеса и система вишеструким регресијама обухвата породицу техника помоћу којих се може истражити веза једне, непрекидне, зависне

¹⁰ *outliers, unusual observation, outlying x observations*

¹¹ *studentized residuals*

¹² *external studentized residuals*

¹³ *Cook's distance*

променљиве и више независних променљивих односно предиктора. Неке од главних врста вишеструких регресија су:

- Стандардна вишеструка регресија (истовремена, симултана) - У овој врсти регресије, која се иначе најчешће користи, све предикторске променљиве се истовремено уносе у једначину. Сусрећу се различити дизајни ових регресија: *Multiple Regression, Factorial regression, Polynomial Regression, Response Surface Regression, All-possible-regressions*;
- Хијерархијска или секвенцијална регресија (*hierarchical regression*);
- Постепена вишеструка регресија (*stepwise regression*);
- Робустна (отпорна) регресија (*robust and resistant regression*);
- Вршна (рубна) регресија (*ridge regression*);
- Непараметарска регресија;
- Регресија са индикаторским променљивим;
- Логистичка регресија (*logistic regression*) и друге регресије.

3.7. Регресори. Базичне форме. Припрема процеса моделирања

Регресије су веома осетљиве математичке творевине, тако да квалитетно одабран полаз даје веће шансе за креирање адекватних модела. Од великог је значаја теоријско знање о процесима који се моделирају, претходно истраживачко искуство као и познавање природе веза међу предикторима и узрочно-последичних веза међу предикторима и излазним променљивим.

Додатни ефекат, позитиван или негативан, има избор броја, типа и облика предикторских променљивих и регресора који улазе у *контејнер потенцијалних базисних форми* у почетној фази процеса моделирања.

У оквиру ове дисертације, правилан избор регресора из контејнера регресора, је базиран на примени GA парадигме.

4. Вештачке неуронске мреже

Алгоритми и технике вештачких неуронских мрежа су, након своје велике реинкарнације, оличене у истраживањима Rumelhart-a и Mc Clelland-a и резултатима публикованим у њиховом научном чланку [44], у току последње три декаде доживеле неслућену примену у практично свим областима модерне науке и технологије. Фокус научних и стручних интересовања и истраживања последњих година се односи, не само на трагања за новим применама, већ и на оптимизацију великог броја параметара, алгоритама, ентитета и правила учења којима се померају ове апстрактне математичке творевине, које oponaшају функционисање човековог мозга.

Бројна унапређења из домена комбинованих AI хеуристика су усмерена на ANN концепте везане за типове ANN мрежа, као и на њихову топологију, архитектуру, активационе функције, трансфер функције, алгоритме учења и бројне параметре којима се њихове перформансе прилагођавају специјализованим наменама.

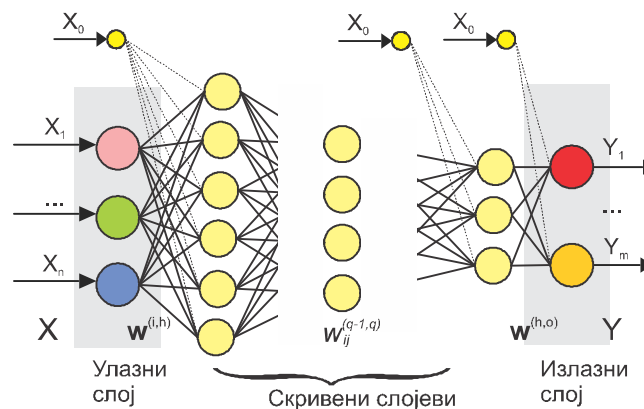
ANN представљају генерализацију нелинеарних регресионих модела.

4.1. Вештачке неуронске мреже са простирањем унапред

Вештачке неуронске мреже су математичке парадигме које апроксимирају структуру и рад природних неуронских мрежа. То су рачунарски системи сачињени од великог броја једноставних, *процесних елемената* или *неурона*¹ [45]. Процесни елементи су груписани у следеће слојеве: улазни слој, један или више скривених слојева и излазни слој. Неурони су спојени везама које су пондерисане *тежинским коефицијентима*.

Посебну врсту ANN мрежа чине мреже са простирањем унапред², које пропaгирају сигнале од улазног ка излазном слоју. Подаци који се процесирају се могу протирати кроз више слојева, од улаза ка излазу, али није дозвољена конекција у облику повратне спреге ка неуронима претходног слоја, нити са неуронима истог слоја. Мреже овог типа су предмет оптимизације у оквиру ове докторске дисертације.

Шематски приказ мреже са простирањем унапред дат је на слици 4.1.



Слика 4.1. Структура *feed-forward* неуронске мреже

¹ *neuron*-неурон; користи се и термин: *node*-чвор

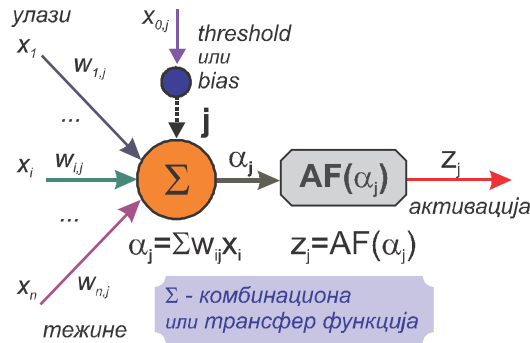
² *feed-forward networks*

где $\mathbf{X}$ означава вектор предикторских променљивих, $\mathbf{Y}$ је вектор зависних променљивих, $\mathbf{w}^{(i,h)}$ је матрица-колона тежинских коефицијената између неурона у улазном слоју и неурона првог скривеног слоја, а $\mathbf{w}^{(h,o)}$ је матрица-колона тежинских коефицијената између неурона последњег скривеног слоја и неурона излазног слоја. Члан $w_{ij}^{(q-1,q)}$ означава i тежински коефицијент између i -тог неурона $(q-1)$ -вог слоја и j -тог неурона (q) -тог слоја. Да би се унапредиле перформансе неуронске мреже, сваком скривеном слоју, као и излазном слоју неурона, је придружен по један посебан неурон, чија је улога слање константног сигнала x_0 (*bias*) ка сваком неурону у овим слојевима. За процесне елементе чија је улога линеарна трансформација улазних сигнала, *bias* члан је еквивалентан одсечку који линеарни регресиони модел креира на ординати [46].

У складу са *сигма правилу* укупни улаз $\alpha_j^{(q)}$ у процесни елемент j , који се налази у q -том слоју, је сума свих пондерисаних³ излаза $x_i^{(q-1)}$ из претходног слоја:

$$\alpha_j^{(q)} = \sum_{i=0}^{n^{(q-1)}} x_i^{(q-1)} \cdot w_{ij}^{(q-1,q)}, \quad j = 1, 2, \dots, n^{(q)}, \quad q = 1, 2, \dots, N_h \quad (4.1)$$

где је N_h број скривених слојева, док су $n^{(q-1)}$ и $n^{(q)}$, бројеви неурона у два суседна слоја (слика 4.2).



Слика 4.2. Трансфер и активациона функција

На исти начин, улазни сигнали за неуроне излазног слоја се рачунају као функције излаза из последњег скривеног слоја.

4.2. Активационе функције

Када улазни сигнал $\alpha_j^{(q)}$ пролази кроз неурон, он се процесира и трансформише у излазни сигнал $x_j^{(q)}$ коришћењем *активационе функције* (слика 4.2):

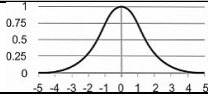
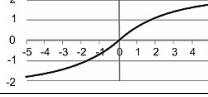
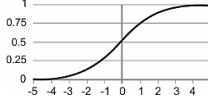
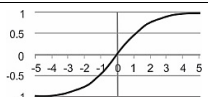
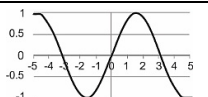
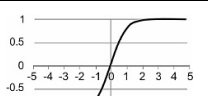
$$x_j^{(q)} = AF(\alpha_j^{(q)}) \quad (4.2)$$

Активациона функција (AF) је неопходна да трансформише суму свих пондерисаних сигнала који се предају неурону, тако да се може рећи да она

³ Под синтагмом *пондерисани сигнал* између неурона i у слоју $(q-1)$ и неурона j у слоју (q) подразумева се вредност која се добија када се излазни сигнал из неурона i помножи тежинским коефицијентом између ова два неурона $(x_i^{(q-1)} \cdot w_{ij}^{(q-1,q)})$.

детерминише трансформациону способност неурона [45]. Неке од најчешће коришћених активационих функција су дате у табели 4.1.

Табела 4.1. Примери активационих функција

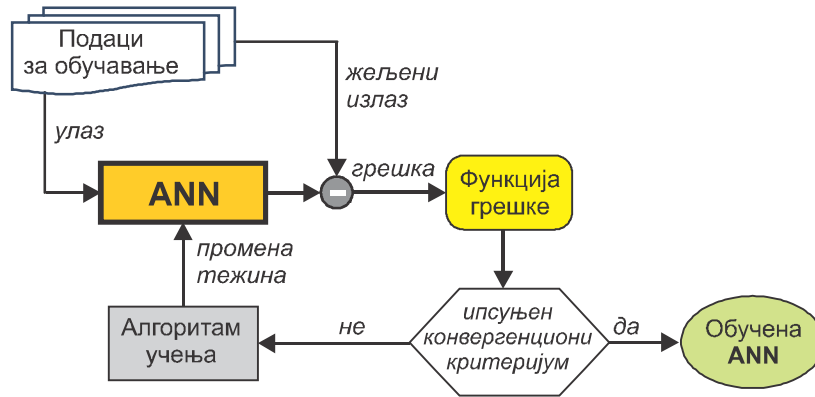
Назив	Математичка форма	Напомена	Изглед функције
<i>Gaussian</i>	$x_j^{(q)} = a \cdot e^{-\frac{(\alpha_j^{(q)} - b)^2}{2 \cdot c^2}}$	фамилија кривих у облику звона	
<i>log</i>	$x_j^{(q)} = \begin{cases} \log(1 + \alpha_j^{(q)}), & \alpha_j^{(q)} \geq 0 \\ -\log(1 - \alpha_j^{(q)}), & \alpha_j^{(q)} < 0 \end{cases}$	функција може бити корисна за избегавање засићења	
<i>sigmoid</i>	$x_j^{(q)} = \frac{1}{1 + e^{-\alpha_j^{(q)}}}$	такође се назива <i>log</i> , <i>log-sigmoid</i> , <i>uni-polar sigmoid</i> или <i>logistic function</i>	
<i>bipolar sigmoid</i>	$x_j^{(q)} = \frac{1 - e^{-\alpha_j^{(q)}}}{1 + e^{-\alpha_j^{(q)}}}$	слична функцији <i>sigmoid</i> AF, али обезбеђује излазе у интервалу [-1,1]	
<i>sine</i>	$x_j^{(q)} = \sin(\alpha_j^{(q)})$	корисна за податке који имају периодичну природу	
<i>hyperbolic tangent (TANH)</i>	$x_j^{(q)} = \frac{e^{2 \cdot \alpha_j^{(q)}} - 1}{e^{2 \cdot \alpha_j^{(q)}} + 1}$	слична функцији <i>log</i> AF, излазни опсег [-1,1], понекад води ка засићењу	

Због добрих својстава са аспекта диференцијабилности и способности да ради и са позитивним и са негативним бројевима *hyperbolic tangent* функција (TANH) је, вероватно, једна од најчешће коришћених активационих функција. Карактеристике различитих активационих функција (AFs) су детаљније дате у [47].

4.3. Алгоритми учења

За одабрану архитектуру мреже, иницијалне вредности тежинских коефицијената се бирају стохастички. Да би се обрачунале и одредиле вредности на излазним неуронима, вредности независних променљивих из подскупа података за учење се пропуштају кроз мрежу. На почетку процеса учења, израчунате вредности за дате улазе се значајно разликују од мерених. Разлика између жељених (измерених) и моделских (израчунатих) вредности се мери помоћу функције грешке (**E**). Циљ процеса учења је подешавање (прилагођавање) величина тежинских коефицијената на вредности које обезбеђују да функција грешке буде минимална.

Један од најпопуларнијих алгоритама учења, који минимизује функцију **E**, је алгоритам учења са пропагацијом грешке уназад (*backpropagation algorithm*). У складу са *backpropagation* алгоритмом, грешка **E** се током бројних итерација пропагира назад кроз мрежу, а тежински коефицијенти се прилагођавају од излазног ка улазном слоју (слика 4.3). Процедура се наставља док се не испуни критеријум конвергенције [45].



Слика 4.3. Модел надгледаног учења

Постоји велики број варијација *backpropagation* алгоритма учења. У овој докторској дисертацији су коришћени неки од најпознатијих: Алгоритам са простирањем грешке уназад и опадањем градијента (*Backpropagation gradient descent algorithm* - *BPGD*) и Алгоритам са еластичним пропагирањем грешке (*Resilient propagation algorithm* - *RPROP*).

4.3.1. Алгоритам са опадајућим градијентом (пакетно опадање и стохастичко опадање) (BPGD)

Базична идеја алгоритма учења са опадањем градијента (*BPGD*) је понављање процедура и правила уланчавања, која узимају у обзир утицај сваке тежине у мрежи на бази функције грешке E коју арбитарно бира истраживач [5]. Након што се израчуна парцијални извод функције грешке по сваком тежинском коефицијенту, минимизација грешке се остварује реализацијом методе падајућег градијента:

$$w_{ij}(t+1) = w_{ij}(t) - \eta \cdot \mu \cdot \frac{\partial E}{\partial w_{ij}}(t) + \mu \cdot \Delta w_{ij}(t-1) \quad (4.3)$$

где је η , *фактор брзине учења* а μ , *моментум фактор* који узима у обзир утицај претходног корака, при ажурирању тежинских коефицијената.

На брзину конвергенције се не може значајније утицати повећањем вредности фактора брзине учења η , зато што такав приступ може водити до *прескакања* минимума функције грешке и понекад до осцилација у процесу учења. Коефицијент μ (моментум фактор), чини процедуру учења стабилнијом и убрзава конвергенцију у области која је блиска оптимуму функције грешке.

4.3.2. Алгоритам учења са еластичном пропацијом (RPROP)

RPROP је скраћеница за *алгоритам учења са еластичном пропацијом* (*Resilient propagation algorithm*) који представља ефикасну шему учења за реализацију директне адаптације тежинских коефицијената на бази информација о локалном градијенту. Да би се ефикасност адаптације унапредила *Riedmiller* и *Braun* [48] су за сваки тежински коефицијент увели коригујућу вредност $\Delta_{i,j}$, која одређује величину односно интензитет промене.

Ова коригујућа вредност се адаптира и прилагођава током процеса учења на основу промене знака парцијалног извода функције грешке E , посматрано локално на нивоу ажурирања тежинског коефицијента w_{ij} , у складу са следећим правилима учења:

$$\Delta_{ij}^{(t)} = \begin{cases} \eta^+ \cdot \Delta_{ij}^{(t-1)}, & \text{if } \frac{\partial E^{(t-1)}}{\partial w_{ij}} \cdot \frac{\partial E^{(t)}}{\partial w_{ij}} > 0 \\ \eta^- \cdot \Delta_{ij}^{(t-1)}, & \text{if } \frac{\partial E^{(t-1)}}{\partial w_{ij}} \cdot \frac{\partial E^{(t)}}{\partial w_{ij}} < 0 \\ \Delta_{ij}^{(t-1)}, & \text{else} \end{cases} \quad (4.4)$$

где је $0 < \eta^- < 1 < \eta^+$. Исказано речима, правило адаптације ради према следећем: Сваки пут када парцијални извод функције грешке по одговарајућем коефицијенту тежине $w_{i,j}$ промени свој знак, што указује да је претходна промена била превише велика и да је алгоритам прескочио локални минимум, вредност коригујућег параметра $\Delta_{i,j}$ се умањује фактором η^- . Ако парцијални извод задржава знак (позитивни или негативни) коригујући параметар $\Delta_{i,j}$ се благо повећава, да би се убрзала конвергенција у подручју оптимума (минимума функције грешке E).

Након што се вредност коригујућег параметра $\Delta_{i,j}$ за сваку тежину адаптира, промена самих вредности тежинских коефицијената прати веома једноставно правило: ако је извод позитиван (повећавање грешке), тежина се умањује овим коригујућим параметром, а ако је извод негативан, вредност коригујућег параметра се додаје:

$$\Delta w_{ij}^{(t)} = \begin{cases} -\Delta_{ij}^{(t)}, & \text{if } \frac{\partial E^{(t)}}{\partial w_{ij}} > 0 \\ +\Delta_{ij}^{(t)}, & \text{if } \frac{\partial E^{(t)}}{\partial w_{ij}} < 0 \end{cases} \quad (4.5)$$

$$w_{ij}^{(t+1)} = w_{ij}^{(t)} + \Delta w_{ij}^{(t)} \quad (4.6)$$

Детаљна објашњења су дата у [49-51].

4. 4. Мрежна архитектура и правила учења

Дизајн ANN је дефинисан мрежном архитектуром (елементима као што су: број скривених слојева и број неурона, тип активационих функција, итд), као и правилима учења. И архитектура мреже и правила учења су веома важни, тако да њихово правилно компоновање и одабир значајно утиче на перформансе мреже [45].

Број слојева и неурона одређује комплексност мреже. Када је ова комплексност исувише велика долази до проблема преприлагођавања [5, 44, 52]. Трагање за оптималном комплексношћу ANN је задатак који је, генерално гледано, још увек нерешен, тако да многи истраживачи користе *метод покушаја и грешака*⁴ за одабир одговарајућег броја скривених слојева и неурона у конкретним апликацијама [53, 54].

У складу са претходним, постоје веома различите препоруке за ове намене.

Додатне идеје су везане за дефинисање обима, односно потребног броја података за обучавање мреже. Број примера за обучавање има утицај на квалитет модела који ће

⁴ *trial-and-error method*

се добити. Истраживачи предлажу различите величине скупова за обучавање у опсегу од 60% до 80% у односу на укупни расположиви скуп података [55-57].

Правила учења дефинишу иницијални скуп тежина, као и параметре η и μ у BPGD алгоритму, односно параметре η^+ и η^- у алгоритму RPROP. Ова правила опредељују како ће тежински коефицијенти бити адаптирани у процесу учења у циљу унапређења мрежних перформанси. Иницијализација тежинских коефицијената има знатан утицај на брзину конвергенције, тако да бројни истраживачи дају различите препоруке за опсеге из којих се вредности тежина у почетку бирају. Опсег иницијалих вредности тежинских коефицијената најчешће коришћен у литератури је $[-0.5, 0.5]$ [58]. Да би се умањио утицај стохастичке природе иницијалних тежина на квалитет ANN модела, Messer и Kittler предлажу да се процес обучавања понови најмање 5-10 пута са различитом иницијализацијом вредности тежинских коефицијената [45, 59].

Алгоритам BPGD је веома осетљив на η и μ параметре. Ако је вредност фактора учења η превише велика, алгоритам учења може ући у стање осцилација и нестабилности. Ако је фактор учења превише мали, алгоритму ће бити потребно пуно времена да постигне конвергенцију. Препоручени интервал за иницијалне вредности параметра η је $[0.2-0.95]$ [60]. Моментум фактор μ има стабилизујућу улогу у BPGD алгоритму [45]. Wythoff [61] препоручује да се моментум фактор бира из опсега од 0.4 до 0.9. За одговарајуће параметре η^+ и η^- за RPROP алгоритам, Riedmiller и Braun [48] препоручују вредности $\eta^+ = 1.2$ и $\eta^- = 0.5$.

4. 5. Мера квалитета ANN модела

Квалитет ANN модела може бити исказан преко Pearson-овог коефицијента линеарне корелације, корена средње квадратне грешке (RMSE), средње квадратне грешке (*Mean Squared Error* - MSE), стандардне девијације σ , итд. Свака од ових мера се израчунава на бази моделских и мерених вредности за податке подскупа за тестирање (TDS).

Најчешће коришћена мера, RMSE, се израчунава по формули:

$$RMSE_{Test} = \sqrt{\frac{1}{M \cdot P} \sum_{p=1}^P \sum_{i=1}^M (t_{pi} - y_{pi})^2} \quad (4.7)$$

где P означава број *pattern*-а⁵ у подскупу TDS, M означава број излазних неурона (чворова), а t_{pi} и y_{pi} су, респективно, циљани и актуелни моделски излази на i -том излазном неурону за p -ти *pattern* [46].

⁵ pattern- шема, односно к-торка улазно-излазних података

5. Генетски алгоритми

5.1. Увод

Теорија оптимизације проучава како да се опише и постигне оно што је најбоље, ако се зна како се мери и разликује шта је добро, а шта лоше. Она обухвата квантитативно проучавање оптимума и метода за одређивање оптимума. Њен је задатак да се изврши избор најбоље варијанте из низа могућих или повољних варијанти у смислу усвојеног критеријума. Таква, најбоља варијанта се назива оптимално решење оптимизационог задатка [1].

Математички гледано проблем оптимизације се своди на одређивање глобалног екстремума критеријумске функције. У ту сврху су, у оквиру теорије управљања, оптимизације и математичког моделирања и програмирања, развијене различите методе. Једна од оптимизационих метода, која је у последње три декаде доживела велику примену, је метода генетских алгоритама.

Генетски алгоритми припадају породици техника усмереног случајног претраживања простора решења¹ [62-78]. За разлику од методе класичне локалне претраге, као што је, на пример, градијентна метода, GA технике трагају за глобалним оптимумом, али са одређеном вероватноћом са аспекта одговора на питање: да ли је добијен глобални или неки од локалних оптимума.

Веома је широк спектар области у којима је ова техника нашла своје место, било да је употребљена самостално, било као компонента неке хибридне методологије. Подручје примене је веома широко: оптимизација неуронских мрежа, транспортни проблеми, оптимизација упита над базом података, итд. Резултати новијег датума, који дају приказ различитих имплементација GA парадигме, су приказани у радовима [79-90].

На жалост, резултати постигнути на теоријском подручју су двосмислени тако да генетски алгоритми остају и даље хеуристичке методе. Детаљнији приказ теоријских поставки генетских алгоритама је дат у [91].

5.2. Природно одабирање

Природна селекција – GA платформа

Генетски алгоритми своје исходиште налазе у Дарвиновој теорији еволуције [3]. Терминологија, концепти, оператори и идеје су изведене из парадигме борбе за преживљавање у којој опстају јединке које имају боља² својства/карактеристике. На тај

¹ Неке техника из ове породице су: *еволуционе стратегије (evolutionary strategies)* - развијене у Немачкој шездесетих година прошлог века, *симулирано каљење (simulated annealing)* - које као узор користи термодинамичко кретање материје ка стању минималне енергије и *генетско програмирање (genetic programming)*.

² Боља својства у смислу опстанка као на пример: јединке су брже, снажније, отпорније на болести, интелигентније...

начин боља својства имају веће шансе да се преносе из генерације у генерацију у континуално измењивом окружењу.

Својства/карактеристике јединки неке врсте живих бића записана су у генима, који граде ланчасте творевине, хромозоме.

Преношење генетског материјала са родитеља на потомке се остварује процесом укрштања. Сваки родитељ потомку даје половину хромозома. Тада долази до измене појединих сегмената хромозомских нити што представља један од кључних процеса, *процес укрштања*. Поред процеса укрштања у току репродукције, повремено, са мањом вероватноћом се јавља и *мутација*. Мутација је случајна промена гена, која код живих бића најчешће има крајње негативни ефекат, али може довести и до појаве повољних особина, чиме се обогаћује генетски материјал врсте.

5.3. Концептулани појмови ГА хеуристике

5.3.1. Процедура уопштеног ГА

Иако су основе и парадигме генетских алгоритама као метода оптимизације интензивно развијане током 60-тих година XX века (*еволуционе стратегије*), битан искорак у њиховом утемељењу везује се за рад John-а H. Holland-а из 1975. године [22].

Holland је у свом раду предложио идеју симулације природног процеса еволуције и дефинисао кључне поставке ове оптимизационе хеуристике.

Према Holland-у, генетски алгоритми су технике претраге инспирисане теоријом природне селекције. Јединке или хромозоми представљају решења у простору потенцијалних решења ($\mathbf{x} \in \mathbf{X}$). Хромозоми се, као и у генетици, састоје од дискретних елемената који се називају генима. Сваки од гена контролише једно својство хромозома [6]. Квалитет сваког решења $\mathbf{x}_i(t)$ се оцењује мером доброте односно мером квалитета тог решења (*fitness*). Јединке се приказују неком врстом апстракције кодиране на једнообразан начин.

Генетски алгоритми су стохастички алгоритми који популацију јединки, $P(t) = \mathbf{x}_1(t), \dots, \mathbf{x}_n(t)$, кроз генерације $(t, t+1, \dots)$ одржавају константном по обиму. Генетски алгоритми обухватају два процеса: *процес селекције* и *процес репродукције*.

У процесу селекције јаке јединке имају већу вероватноћу за опстанак и преношење својих гена на будуће генерације [22]. Селекција обухвата избор скупа хромозома (јединки) који ће преживети и пренети добар генетски материјал у следећу $(t+1)$ генерацију. У ГА постоје различите процедуре селекције прилагођене природи проблема који се оптимизује, као и природи *fitness*-а којим се описује квалитет јединки.

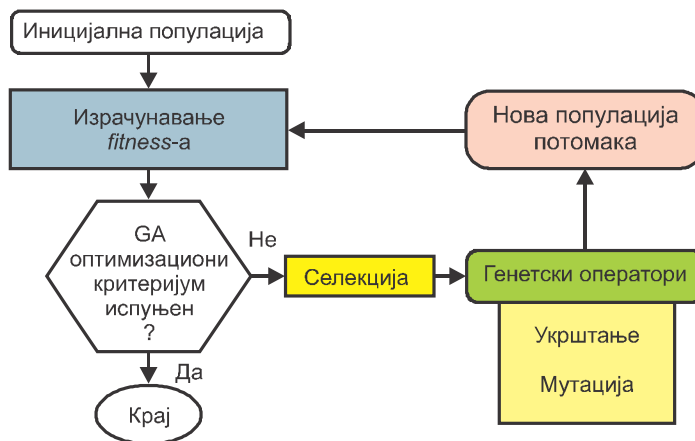
У процесу репродукције, ГА користи два оператора за генерисање нових решења на бази постојећих: бинарни оператор *укрштања*³ и унарни оператор *мутације*⁴. У укрштању се два хромозома, које називамо *родитељима*, комбинују формирајући нове хромозоме, које називамо *потомцима*. Итеративном применом селекције и оператора укрштања, гени добрих хромозома се јављају чешће у популацији. Ово обезбеђује конвергенцију у постизању оптималног решења [6]. Мутација обезбеђује свежу

³ *crossover operator*

⁴ *mutation operator*

генетску грађу и напуштање зоне локалних минимума и/или генетски уједначених хромозома у популацији.

Шематски приказ процедуре уопштеног GA [23] је дат на слици 5.1.



Слика 5.1. Шема генетског алгоритма

У већини имплементација GA величина популације (број хромозома) је константна у току процеса еволуције. Иницијална популација јединки је најчешће рандомизирана. Могућа је и варијанта да се у почетни скуп хипотеза укључи и неко оптимално решење добијено неком другом методом.

После евалуације и селекције, иницијална популација се подвргава операторима укрштања и мутације са циљем генерисања следеће генерације у процесу еволуције. Итеративна процедура се прекида када се добије довољно добро решење или када се реализује задати број итерација. Најбоља јединка из последње популације, би требало да буде близу глобалног оптимума, али за то не постоји гаранција, нити одговарајући математички доказ. Из тог разлога за GA са сигурношћу можемо да тврдимо само то да из генерације у генерацију неће дати лошија решења од постојећих.

Примена GA је могућа кроз две стратегије. Прва стратегија подразумева прилагођавање GA проблему који се решава. На овај начин се добијају специјализована ефикасна оптимизациона решења са стриктном наменом⁵. Друга стратегија подразумева прилагођавање оптимизационог проблема GA концепту [91].

У приступу у којем се реални оптимизациони проблем прилагођава GA концепту, посебна пажња је посвећена:

- апстрактној презентацији оптимизационог проблема односно пресликавању могућих решења у јединке, такозваном кодирању јединки и
- дефинисању *fitness-a*, односно мере квалитета тих јединки.

При свему овоме се јављају различите врсте проблема. На пример, након укрштања и/или мутације се могу јавити решења која не припадају области могућих решења. Даље, кодирање јединки може бити неодговарајући за посматрани проблем, итд [91].

Узимајући у обзир одреднице наведене у овом одељку, се може закључити да генетски алгоритми представљају скуп комплексних процедура често заснованих на

⁵ Генетски алгоритми одговарајуће специјализоване намене најчешће се називају *еволуционим програмима*.

стохастичи и емпирији, али уједно и скуп процедура које дају квалитетна решења у многим сложеним оптимизационим задацима.

5.3.2. Кодирање решења. Представљање јединки

Процес кодирања могућих решења из простора решења је од огромног значаја јер може битно утицати на ефикасност GA. Математички посматрано, хромозом може бити било каква структура података која описује својства јединке. Тако, хромозоми могу бити бинарни записи или реални бројеви са покретним зарезом за оптимизацију функција са једном или више променљивих. Хромозоми могу имати форму поља бројева која је погодна за кодирање проблема распореда. Хромозоми могу имати и облик матрица⁶, низова, стабла, графова итд. Након кодирања следи развој одговарајућих GA оператора за укрштање и мутирање. Развијени оператори не би смели да генеришу јединке које не припадају простору решења односно простору хипотеза. У тексту који следи наводе се примери кодова који се често сусрећу у практичним применама.

Хромозоми са бинарним кодом

У оваквом начину приказивања, хромозом представља бинарни вектор одређене димензије. Уколико је реч о једној димензији, бинарни вектор дужине n bit-ова, омогућава кодирање 2^n хромозома. Ови хромозоми, са одређеном тачношћу, представљају бинарни запис реалних бројева из интервала $x \in [x_d, x_g]$, где је x_d доња, а x_g горња граница интервала у којем се тражи оптимално решење.

Бинарни код се, дакле, може записати у облику:

$$b = \sum_{i=0}^{n-1} B_i \cdot 2^i, \quad B_i \in \{0,1\} \quad (5.1)$$

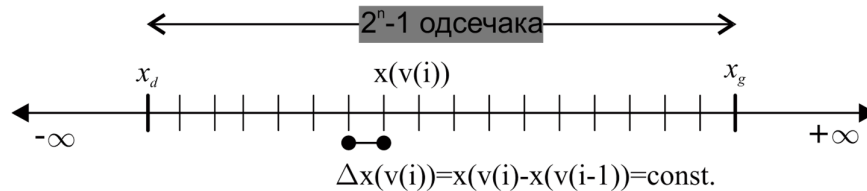
Ако се за задати интервал $[x_d, x_g]$ тражи оптимално решење са тачношћу p , онда дужина хромозома мора бити:

$$n \geq \frac{\log[(x_g - x_d) \cdot 10^p + 1]}{\log 2} \quad (5.2)$$

Хромозоми су еквидистантно распоређени на интервалу $[x_d, x_g]$ са међусобним растојањем:

$$\Delta x = x(v(i)) - x(v(i-1)) = \frac{1}{2^n - 1} \cdot (x_g - x_d), \quad \forall i \in [1, 2^n - 1] \quad (5.3)$$

где $v(i)$ представља бинарни вектор i -тог хромозома (слика 5.2) [91].



Слика 5.2. Хромозом $v(i)$ као једна од 2^n тачака равномерно распоређених на интервалу $[x_d, x_g]$

⁶ Пример: проблем 8 дама

У пракси се показало да бинарна представа даје најбоље резултате у већини примера где је могућа примена овог начина кодирања [91].

Хромозоми са Gray-евим кодом

Главни недостатак класичног бинарног кодирања је својство кода да Hamming-ова удаљеност суседних бројева може бити велика⁷. Ово значи да, ако је алгоритам достигао решење блиско оптимуму, онда је потребно променити Hamming-ов број *bit*-а, да би се постигао стварни оптимум. У Gray-овом коду⁸ суседни декадни бројеви се разликују само у једном *bit*-у (Hamming-ово растојање је 1) јер је овај код циклични.

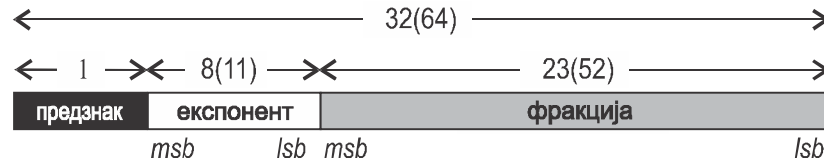
Једначине за трансформацију бинарног броја, $b = b_m b_{m-1} \dots b_2 b_1$, у Gray-ев код, $g = g_m g_{m-1} \dots g_2 g_1$, и обрнуто гласе:

$$g_m = b_m, \quad g_k = b_k \text{ XOR } b_{k-1}, \quad b_k = \sum_{j=k}^m g_j \pmod{2}, \quad j = 1, 2, \dots, m \quad (5.4)$$

Циклични кодови су у теорији GA погодни јер већина једнобитних мутација изазива мале промене (*hill climbing effect*). Експериментални резултати, ипак, не указују на неке значајне предности примене Gray кода [91].

Кодирање са реалним бројевима у покретном зарезу

За решавање задатка оптимизације, код функција са реалним аргументима, за кодирање гена се могу користити и реални бројеви, директно, без кодирања у бинари код. Приказ ових гена у облику предзнака, мантисе и експонента (слика. 5.3), одговара IEEE 754-1985 стандарду:



Слика 5.3. Број са покретним зарезом, једностуке и двоструке тачности

Презентација решења, директно, у реалним бројевима подразумева развој специфичних GA оператора. Један од таквих оператора је, на пример, динамички оператор мутације. Динамички оператор мутације обезбеђује фино подешавање решења у оптабилној области.

5.3.3. Функција добротe (*fitness*)

Један од највећих проблема у реалној примени GA парадигме је дефинисање мере квалитета јединки односно хромозома. Та мера квалитета, као што је већ речено, често се назива *добротом* односно *функцијом добротe* или *fitness*-ом. Понекад сама дефиниција није проблем, али је велики проблем реално мерење односно квантификовање функције добротe. Што је доброта јединки већа, то је већа вероватноћа да ће се јединка укрштати и да ће преживети процес репродукције.

Суштина процеса оптимизације кроз еволуцију је тежња да *просечна доброта* $\bar{D}$, на нивоу популације, расте из генерације у генерацију:

⁷ Иако се бројеви 255_{10} и 256_{10} у декадном бројном систему разликују за само једну јединицу, Hamming-ово растојање (број *bit*-ова у којима се бројеви разликују) је 9 у бинарном бројном систему: $255_{10} = 01111111_2$ и $256_{10} = 10000000_2$.

⁸ под Gray-евим кодом најчешће се подразумева бинарно рефлектиран Gray-ев код.

$$\bar{D} = \frac{\sum_{i=1}^{VelPop} fitness(v_i)}{VelPop} \quad (5.5)$$

где $VelPop$ представља величину популације, а v_i вектор јединке. У најједноставнијој варијанти $fitness$ је једнак самој вредности циљне функције:

$$fitness(v) = f(x) \quad (5.6)$$

5.4. Селекција

У процесу селекције се врши одабир бољих јединки са циљем чувања доброг генетског материјала. Добри гени се на тај начин преносе у следећу генерацију.

Уместо једноставног процеса сортирања и одбацивања лошијих јединки (што иначе доводи до преурађене конвергенције и губитка дела доброг генетског материјала које поседују лоше јединке), у пракси су развијени различити алгоритми селекције који балансирају потребу да добре јединке преживе, али и да лошије јединке добију одређену шансу да допринесу процесу еволуције. Генетски алгоритми, с обзиром на врсту селекције, се деле на:

- генерацијске (једноставна селекција, турнирска селекција...) и
- елиминацијске генетске алгоритме (*GA with steady-state reproduction*).

Генерацијске селекције подразумевају истовремено постојање две генерације у појединим тренуцима рада ГА. Елиминацијске селекције оперишу са једном популацијом која се ажурира. Кратак опис неких селекција из ових група је дат у тексту који следи.

Једноставна селекција (*Roulette wheel parent selection*)

Циљ једноставне селекције је одабир јединки чија је вероватноћа селекције пропорционална њиховој доброти. Један хромозом се може појавити више пута у следећој генерацији. Величина популације $P(t)$ је константна и иста као и у претходној генерацији $P(t-1)$.

Ако се као пример оптимизационог задатка одабере одређивање максимума функције једне променљиве, процедура једноставне селекције се може описати кроз следеће кораке [91]:

- Ако $\exists f(x_i) < 0, i = 1, 2, \dots, VelPop$ онда се додаје позитивна константа⁹
 $C = \min(f(x_1), \dots, f(x_i), \dots, f(x_{VelPop}))$, тако да је: $fitness(v_i) = f(x_i) + C$ и
 $fitness(v_i) \geq 0, i = 1, 2, \dots, VelPop$
- Израчунају се вредности доброте за сваку јединку: $fitness(v_i)$
- Одреди се укупна добротa: $D = \sum_{i=1}^{VelPop} fitness(v_i)$
- Израчунава се кумулативна добротa за сваки хромозом:

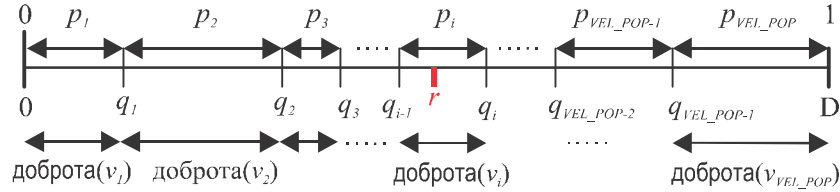
⁹ Описани поступак се назива **транслација** или помак функције циља (*windowing*). На овај начин $fitness$ јединки не може бити негативан, а избегавају се и приближно исте вероватноће избора јединки које би се добиле ако би се, у циљу избегавања негативних вредности, вредностима функције једноставно додао неки велики број M ($g(x)=f(x)+M$).

$$q_k = \sum_{i=1}^k \text{fitness}(v_i), \quad k = 1, 2, \dots, \text{VelPop}$$

- Израчунава се вероватноћа селекције за сваки хромозом:

$$p_k = \frac{\text{fitness}(v_k)}{D}, \quad k = 1, 2, \dots, \text{VelPop}$$

- Генерише се случајни број r у интервали $[0, D]$ и одреди i -ти хромозом за који важи да је $r \in (q_{i-1}, q_i)$ и управо овај хромозом преноси у следећу генерацију (слика 5.4)



Слика 5.4. Кумулативна доброта q_i и вероватноћа селекције p_i

Ако поједине јединке имају значајно већи *fitness* у односу на друге, понекад се као финални корак једноставне селекције, додатно реализује и *линеарна нормализација*.

Генерацијска турнирска селекција

Генерацијска турнирска селекција *VelPop* пута бира k јединки из популације, евалуира њихове *fitness*-е и најбољу јединку копира у базен за репродукцију [91].

Једноставна елиминацијска селекција

За разлику од једноставне генерацијске селекције, елиминацијска селекција уместо добрих, бира лоше хромозоме које одбацује мењајући их са јединкама које ГА оператори генеришу од јединки које су преживеле. Уместо *fitness*-а, се дефинише *казна*.

$$\text{kazna}(v_k) = \max_{i=1}^{\text{VelPop}} (\text{fitness}(v_i) - \text{fitness}(v_k)), \quad k = 1, 2, \dots, \text{VelPop} \quad (5.7)$$

Пошто је казна најбоље јединке једнака нули, произилази да код овог облика селекције нема потребе за посебном процедуром којом се чува најбоља јединка (*елитизам*). После елиминације сваке јединке, поново се мора обрачунати кумулативна доброта и укупна доброта.

Како се током процеса селекције могу јавити дупликати, у неким имплементацијама ГА су уграђени програмски механизми за спречавање дупликата, јер они могу довести до успоравања ГА.

Елиминацијска турнирска селекција

За разлику од генерацијске турнирске селекције, елиминацијска турнирска селекција *VelPop* пута бира k јединки из популације, евалуира њихове *fitness*-е и уместо најбоље, најлошију јединку из групе, елиминише и мења је са новом јединком коју су оператори ГА генерисали помоћу јединки из групе које су преживеле.

Елитизам

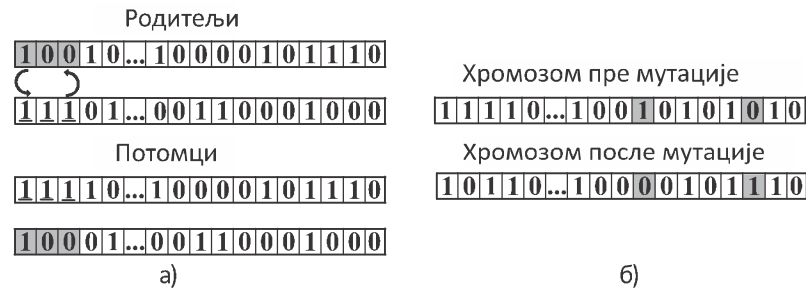
Елиминацијском селекцијом по аутоматизму је обезбеђено преношење најбољег хромозома у следећу генерацију. Код генерацијских селекција овај захтев се мора програмски подржати кроз посебне процедуре. Елитизам обезбеђује асимптотско кретање ка глобалном оптимуму.

5.5. Репродукција

Поред селекције, кључни процес у реализацији ГА парадигме је репродукција. Као што је раније наведено, репродукција се остварује преко оператора укрштања и мутације.

5.5.1. Укрштање (Crossover)

Укрштањем се обезбеђује да се наследне особине преносе са родитеља на децу. Број тачака прекида је арбитраран. На слици 5.5 а је приказано дејство оператора укрштања код бинарних хромозома у једној тачки прекида (*the single point crossover*). Код ове врсте укрштања вероватноћа да дете наследи особине оба родитеља је једнака. Након што се тачка прекида одабере на случајан начин, оба хромозома се деле у одабраној тачки прекида а потом се размењује генетски материјал [6].



Слика. 5.5. Оператори укрштања (а) и мутације (б)

Поред униформног разликујемо укрштање код ког је вероватноћа наслеђивања гена од једног родитеља p а од другог родитеља $(1-p)$, као и укрштање на основу маске код којег се вероватноћа наслеђивања од мајке или оца разликује од гена до гена понаособ. Код укрштања са више тачака прекида (*сегментно укрштање*) позиција прекида је случајна за свако поједино укрштање.

Поред описаних постоје и друге форме оператора укрштања као што је, на пример, *мешајуће укрштање* код кога се у првој фази мешају *bit*-ови код сваког родитеља, потом се хромозоми укрсте на класичан начин, а затим се код сваког родитеља измешани *bit*-ови врате на своја места.

Код хромозома у облику реалног броја са покретним зарезом, оператор укрштања даје као резултат случајан реалан број из интервала чије границе дефинишу родитељи:

$$x_{\text{дете}} = \text{rand у интервалу } (x_{\text{мајка}}, x_{\text{отац}}) \quad (5.8)$$

Претпоставља се да је управо оператор укрштања то што разликује ГА од других метода оптимизације¹⁰ [91].

5.5.2. Мутација

Мутација је механизам за избегавање локалних минимума. То је унарни оператор јер делује над једним хромозомом. Реч је о случајној промени једног или више гена. На Слици 5.5.б. је приказана, за бинарне хромозоме, мутација на нивоу једног гена (*bit-flip mutation operator*).

Параметар који одређује вероватноћу мутације је уједно и параметар ГА.

¹⁰ Оператор мутације се јавља и код симулираног кађења, као и код генетског програмирања

Улога мутације је, такође, и обнављање изгубљеног генетског материјала. Ако се догоди да све јединке имају исти ген на одређеном месту, укрштањем се тај ген никада не би могао променити. На тај начин би се изгубило чак пола простора претраживања [91].

Код оптимизационих проблема код којих се хромозоми репрезентују у форми реалних бројева са покретним зарезом, мутација може бити случајан број из читавог опсега варирања независне променљиве: $x' = rand$ у интервалу (x_{DG}, x_{GG}) . На тај начин се обезбеђује претрага целог простора потенцијалних решења у току процеса еволуције.

У циљу повећања ефикасности GA процедуре, у оптабилној области, се употребљава *динамички оператор мутације* који обезбеђује да се простор мутације смањује са повећањем броја итерација¹¹. Уместо вредности x' се бира коригована, динамичка, вредност x'_{din} :

$$x'_{din} = rand \text{ у интервалу } (DG_{din}, GG_{din}) \quad (5.9)$$

где је

$$DG_{din} = \max(DG, x' - (GG - DG) \cdot r(t)) \quad (5.10)$$

$$GG_{din} = \min(GG, x' + (GG - DG) \cdot r(t))$$

$$r(t) = 1 - s^{(1 - t/T)^b} \quad (5.11)$$

а значења параметара су: s је случајни број из интервала $[0,1]$, b је параметар који одређује степен зависности од броја итерација и одређује се арбитарно, t означава број текуће итерације, док T означава укупан број итерација [91].

Поред описаних генетских оператора, у литератури је приказан, и у практичним имплементацијама употребљен, велики број њихових варијација и унапређења са различитих аспеката (променљиве дужине хромозома, променљиви интервали деловања, итд).

5.6. Параметри GA

Независно од типа кодирања оптимизационог проблема, примењене селекције, врсте и специфичности оператора, кључни параметри GA су: величина популације, број генерација, вероватноћа укрштања, вероватноћа мутације и број јединки за елиминацију код елиминацијских GA. Оптимизовање ових параметара је веома сложен поступак и захтева извођење великог броја експеримената. Могуће препоруке за вредности параметара за *мале* и *велике* популације дате су у табели 5.1 [91].

Табела 5.1. Оквирне вредности параметара GA

Параметри	Ознака	Мала популација	Велика популација
Величина популације	$VelPop$	30	100
Вероватноћа укрштања	p_c	0.9	0.6
Вероватноћа мутације	p_m	0.01	0.001
Број јединки за елиминацију	M	$VelPop / 2$	$VelPop / 4$

¹¹ Идеја је позајмљена од хеуристике *симулираног каљења*, где са временом (с бројем итерација), температура, као параметар алгоритма, пада.

Као што је раније истакнуто, неки од ових параметара могу бити динамички, а избор опсега у којима се крећу њихове вредности умногоме зависи и од природе проблема који се решава (облика циљне функције). Циљ је налажење компромиса у конвергирању ка оптимуму са једне стране, и способности за истраживање нових подручја у трагању за глобалним оптимумом, са друге стране.

Неки аутори предлажу да се за избор вероватноћа p_c и p_m развију посебни генетски алгоритми који би се реализовали паралелно са главним GA (генетски алгоритам над генетским алгоритмом). Srinivais и Patnaik [92, 93], за генерацијске GA предлажу прилагодљиви генетски алгоритам (*AGA-Adaptive Genetic Algorithm*) у коме се p_c и p_m мењају у функцији тренутног стања популације са аспекта степена конвергенције према неком оптимуму. Као меру степена конвергенције ови аутори предлажу разлику доброте најбољег члана у популацији и просечне доброте текуће популације ($D_{\max} - \bar{D}$). Добри и лоши хромозоми немају исте вредности за p_c и p_m и разликују се за сваког члана текуће популације:

$$\begin{aligned} p_c &= k_1 \cdot (D_{\max} - D') / (D_{\max} - \bar{D}), \quad k_1 \leq 1.0 \\ p_m &= k_2 \cdot (D_{\max} - D) / (D_{\max} - \bar{D}), \quad k_2 \leq 1.0 \end{aligned} \quad (5.12)$$

У претходним једначинама D' , је *fitness* бољег хромозома од два који су одабрани за укрштање, а D је доброта хромозома одабраног за мутацију. Лако је закључити да су p_c и p_m најбољег хромозома једнаке нули (*елиитизам*). Додатна ограничења и препоруке за *AGA* су детаљније дате у [93].

Уопштено посматрано, генетски алгоритми дају лошије резултате при фином подешавању решења у подручју глобалног оптимума. Решавање ове проблематике се може усмерити у више праваца. Један од њих је, према Holland-у, употреба GA као предпроцесора у припреми глобалног оптимума, а потом примена неких стандардних аналитичких техника за одређивање оптимума са жељеном тачношћу. Остала решења, која смо већ поменули, су усмерена на смањење простора претраге путем развоја динамичких оператора мутације (за хромозоме кодиране са реалним бројевима са покретним зарезом) или динамичком променом дужине хромозома у процесу еволуције (код реалних бројева са покретним зарезом приказаних у форми *bit*-ова).

6. Адаптација линеарних регресионих модела помоћу генетских алгоритама

У оквиру овог поглавља су приказани развијена MLR/GA метода и одговарајући GenReg софтверски агент за адаптивно моделирање понашања објеката, у облику вишеструких линеарних регресија, које карактерише:

- аутоматско генерисање предиктивних регресионих модела који су оптимизовани са аспекта тачности и комплексности и
- робустност на могуће промене у скупу улазних променљивих.

Софтверски агент је назван GenReg, јер помоћу генетских алгоритама реализује аутоматизовану адаптацију регресионих модела.

MLR/GA метода омогућава генерисање MLR модела који је оптимизован за важећи скуп мерења, прикупљених од почетка посматраног периода до тренутка генерисања самог модела. Како се регресори могу јавити у различитим облицима, MLR/GA метода тежи проналажењу регресионог модела који ће се прилагодити тренутно расположивим подацима на најбољи могући начин¹. Оптимизација броја и облика регресора поверена је GA хеуристици. Јединке у популацији су регресиони модели. Оператори укрштања и мутације, селекција на основу коефицијената детерминације и примена идеје елитизма, су платформа помоћу које GenReg, на основу MLR/GA методе, омогућава генерисање *довољно доброг*² математичког модела који описује дати објекат или процес. Додатно, GenReg агент омогућава генерисање MLR модела који су, поред важећег скупа мерења, оптимизовани и за важећи скуп променљивих односно скуп тренутно активних сензора помоћу којих се реализује аквизиција података.

6.1. Стање у области

За опис, предикцију и оптимизацију понашања најразноврснијих објеката и система у широком дијапазону техничких, друштвених и природних феномена, у последњих неколико деценија су развијени бројни детерминистички, статистички и хибридни математички модели.

Већ је речено да се предности статистичких модела огледају у једноставности формулације, брзини израчунавања и успостављања било ког типа односа и корелације између независних и зависних променљивих. Унапређене и специјализоване статистичке технике, међу којима се посебно истичу постепене регресије, нашле су велику примену и постале практично стандард [25].

¹ *the best fitted model*

² Синтагму *довољно добар модел* не треба поистоветити са синтагмом *адекватан модел*. Појам адекватности детаљно је приказан у [15].

Као што је наведено у делу 1.4, у последње време статистичке методе су обogaћене великим бројем хеуристика из домена вештачке интелигенције, које комбинују предности класичног, статистичког приступа и AI парадигми [28-37, 94-97].

Ипак, без обзира на развој и унапређења модерних приступа, у широкој лепези бројних облика и врста статистичких математичких модела, модели у облику вишеструких линеарних регресија и даље представљају стожер и заузимају централно место. Због тога је од огромног значаја даље проучавање и решавање проблема који су за њих везани.

Примарни проблеми у примени MLR модела су везани за избор одговарајућих регресора, како по питању облика, односно форме, тако и по питању оптималног броја ових регресора.

Поред ових, кључних проблема, без обзира на историју унапређења и двовековну научну еволуцију у домену MLR тематике, велики број питања је и даље отворен.

Да ли користити параметарске или непараметарске регресије? Како одговорити изазову огромних база мерених података, које генеришу технолошки све савршенији сензори? Колика треба да буде сложеност MLR? Како генерисати хипотезе које су адекватне, али које добро уопштавају? Да ли и када применити редукцију димензија, а када редукцију података?

Додатно, за апликације које раде у реалном времену, а које се баве предикцијом понашања посматраних система на бази података који се добијају континуалном аквизицијом, проблем може бити повремени нерегуларан рад мерне опреме. Нерегуларан рад мерне опреме доводи до дисконтинуитета (несталности) у базама мерених вредности односно до временске променљивости структуре скупа улазних променљивих. Ово директно утиче на обим и структуру потенцијалних регресора.

На основу претходно изнетог следи, да је за адаптивно генерисање оптималних регресионих модела комплексних објеката и процеса, неопходно развити систем који је толерантан на несталност како скупа мерења тако и скупа улазних променљивих односно скупа активних сензора.

6. 2. Опис GenReg система

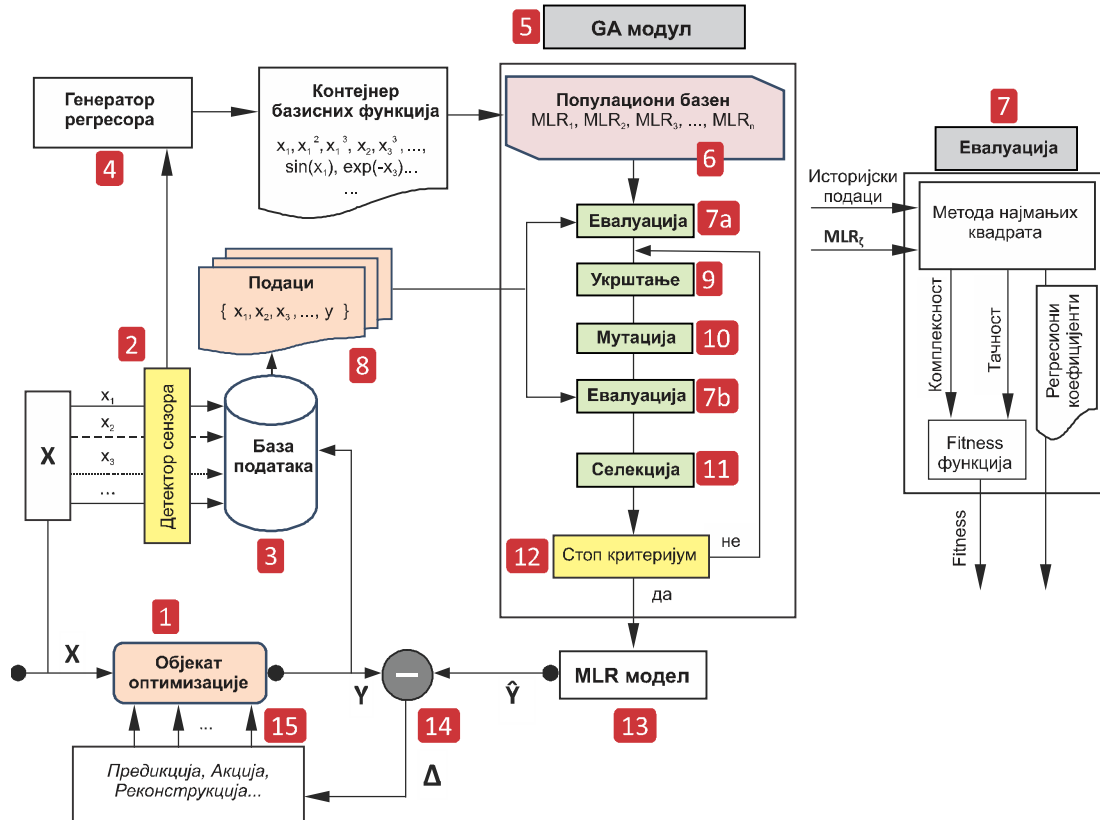
Понашање објекта оптимизације (1) (слика 6.1), се може посматрати као *црна кутија* која трансформише вектор улазних променљивих $\mathbf{X}$ у вектор излазних променљивих $\mathbf{Y}$. Овај систем је приближно описан математичким моделом (13) који је заснован на MLR. Математички модел омогућава предикцију понашања објекта моделирања ($\hat{\mathbf{Y}}$) за познате улазне променљиве и њихове вредности.

Као што је раније поменуто, класични регресиони модели нису толерантни на динамичку промену скупа актуелних мерења као и на промену сета улазних променљивих. Због тога је, у оквиру GenReg агента, на основу MLR/GA метода, развијен модул за адаптивно генерисање оптималног регресионог модела у складу са скупом тренутно активних сензора и скупом тренутно актуелних мерења.

MLR/GA метода је хибридна метода базирана на вишеструким линеарним регресијама и генетским алгоритмима. Језгро хибрида је део за предикцију понашања посматраног објекта коришћењем MLR. У складу са теоријом генетских алгоритама, одређивање оптималног броја и облика регресора у MLR моделу је разматрано као

оптимизациони проблем у коме сваки MLR модел представља јединку у популацији. Мера доброте (мера квалитета, *fitness*) сваке јединке (хромозома) је одређена као функција тачности и комплексности модела саме те јединке.

Слика 6.1, даје шематски приказ развијеног GenReg система у чијем средишту је новоразвијена MLR/GA метода.



Слика 6.1. Адаптивни систем за моделирање објеката (GenReg) базиран на MLR/GA методи

MLR/GA метода подразумева предпроцесирање података и припрему контејнера базисних функција. За наведене задатке задужени су поједини елементи GenReg система. За креирање скупова за учење и тестирање (8), који су подлога за генерисање MLR модела, вредности значајних улазних променљивих као и вредности излазних променљивих се складиште у базу података (3). Зависно од броја и типа активних сензора (2), генератор регресора (4) креира односно ажурира контејнер базисних функција (5). Дакле, ови регресори одговарају текућем расположивом скупу сензора. На пример, ако сензор који мери вредности предиктора x_3 , није активан или је дошло до нерегуларности у раду овог сензора, онда контејнер регресора односно базисних функција, не укључује базисне функције које би се односиле на улазну променљиву x_3 .

Коришћењем претходно креираних базисних функција, GA модул (5), који имплементира MLR/GA методу, генерише иницијалну популацију регресионих модела (6). Свака јединка (MLR модел) је формирана случајним избором регресора из скупа расположивих регресора. Мера квалитета сваке јединке у иницијалној популацији се одређује кроз евалуациони процес (7a) коришћењем два критеријума: *критеријума тачности* и *критеријума комплексности* регресионог модела. Критеријум тачности

мери степен подударанја резултата који генерише математички регресиони модел, са резултатима добијеним мерењем за исти скуп улазних података (8).

За оцену тачности MLR модела а на основу методе најмањих квадрата израчунавају се коефицијенти регресионог модела, уважавајући чињеницу да хромозоми репрезентују само број и тип регресора. Када се коефицијенти регресије израчунају, тачност модела се одређује помоћу стандардних статистичких показатеља, као што су корен средње квадратне грешке (RMSE) или коефицијент детерминације вишеструке линеарне регресије (R^2).

Критеријум комплексности обухвата оцену модела са аспекта рачунарског времена које је потребно за израчунавање резултата и одређен је бројем и типом регресора.

Нове јединке се добијају из постојећих у процесу репродукције, применом оператора укрштања (9) и мутације (10) а њихов *fitness* се израчунава као што је претходно описано (76). Јединке које ће потенцијално прећи у следећу генерацију пролазе процес селекције (11) на бази својих *fitness*-а.

Трагање за бољом комбинацијом регресора се наставља док се не испуни неки од критеријума за прекид алгорита као што су: постигнута задовољавајућа тачност модела, реализован задати број итерација (генерација), или протекло, унапред дефинисано време трајања оптимизационе процедуре (12). Када се критеријум престанка задовољи, добија се оптимизовани регресиони модел (13) који одговара скупу тренутно активних сензора. Овај модел може бити даље коришћен за предикцију понашања моделираног објекта, за додатне анализе, поређења са мерењима и могуће акције. Повратна спрега (15) омогућава поређење (14) мерених вредности и вредности добијених моделом, и на тај начин обезбеђује подлогу за праћење неуобичајеног понашања објекта.

6.3. Опис MLR/GA методе

MLR/GA метода је хибридна метода која је намењена за решавање комплексног оптимизационог задатка избора броја и облика регресора који граде регресиони модел објекта моделирања. Математичка формулација наведеног оптимизационог проблема, алгоритам за његово решавање и структура софтверске имплементације су описани у деловима који следе.

6.3.1. Математичка формулација оптимизационог проблема

Математички регресиони модел датог објекта или система, поред комплексности у смислу величине и обима, може имати регресоре различитих врста. На располагању су: фрагменти полиномијалних функција, степене функције, експоненцијалне функције, рационалне функције, тригонометријске функције, логаритамске функције итд. На пример, регресори могу имати облике: x_1 , x_2 , x_3 , x_1^2 , x_2^3 , $\sqrt{x_3}$, e^{-x_3} , $\ln(x_1)$, $\sin(x_2)$, и друге. Наравно, могуће су и комбинације сегмената појединих функција у најразноврснијим варијантама ($x_1 \cdot \sin(x_3)$, $\sqrt{x_1} \cdot \ln(x_4)$, ...), јер вишеструке линеарне регресије, као што је раније поменуто (Одељак 3.4), подразумевају линеарност коефицијената, а не регресора.

На основу природе оптимизационог проблема, теоријских знања и емпирије, истраживач опредељује структуру скупа потенцијалних регресора, односно даје оквир

за контејнер потенцијалних базисних функција. Број базисних функција регресионог модела у принципу се разликује од броја независних променљивих.

Уопштено, MLR модел зависне променљиве (*хипотеза*, у AI терминологији) може бити исказан у следећој форми:

$$y = \sum_{j=1}^m \sum_{k=1}^s C_{jk} \cdot \beta_{jk} \cdot f_{jk}(x_j) \quad (6.1)$$

где је m укупан број предиктора, s број облика у коме се може наћи сваки предиктор, f_{jk} је k -та базисна форма j -тог предиктора, β_{jk} је коефицијент за базисну форму f_{jk} , а x_j је вредност j -тог предиктора, односно компоненте вектора независних променљивих $\mathbf{X}$. Параметар C_{jk} је бинарна променљива која је једнака јединици ако је j -ти предиктор у k -тој форми (облику) део конкретног MLR модела, а у противном је једнака нули.

Ако $\mathbf{M}$ означава скуп свих могућих MLR модела који могу бити креирани на бази расположивог скупа базисних функција f_{jk} , онда се сваки модел ζ из $\mathbf{M}$ може написати као:

$$y^{(\zeta)} = \sum_{j=1}^m \sum_{k=1}^s C_{jk}^{(\zeta)} \cdot b_{jk}^{(\zeta)} \cdot f_{jk}^{(\zeta)}(x_j), \quad \zeta = 1, N \quad (6.2)$$

где је N укупан број модела.

Ако се тачност модела ζ мери коригованим коефицијентом вишеструке детерминације $R_{adj}^2(\zeta)$, (једначине 3.21 и 3.22), онда се оптимизациони проблем, који проналази најбољу комбинацију линеарних регресора, може дефинисати у следећој форми:

$$\text{Максимизовати } R_{adj}^2(\zeta), \quad \zeta = 1, N \quad (6.3)$$

Регресори f_{jk} , као што је већ наведено, по свом броју и облику су дефинисани у складу са теоријским и емпиријским знањима о објекту моделирања и оптимизације. Оцене $b_{jk}^{(\zeta)}$ коефицијената $\beta_{jk}^{(\zeta)}$ се добијају методом најмањих квадрата (део 3.4).

Из разлога што функције циља³ (6.3) садрже, како бинарне тако и нелинеарне променљиве, оптимизација таквих функција се може класификовати као високо комплексан, нелинеарни оптимизациони проблем, који не може бити решен класичним методама.

Додатни проблем се јавља са порастом броја независних променљивих и њихових регресионих облика. Повећање комплексности узрокује екстремно повећање обима претраге у простору решења. Уколико би се за решавање проблема обима претраге, користила нека од класичних метода (из породице *brute force* техника), онда би захтев за потребним рачунарским ресурсима растао експоненцијално. У системима који раде у реалном времену овакав приступ би био непримењив.

За превазилажење претходно наведених проблема, у овој дисертацији су искоришћени генетски алгоритми, на начин описан у следећим деловима.

³ *objective function*

6.3.2. Опис примењеног генетског алгорита

6.3.2.1. Структура хромозома

Полази се од претпоставке да се функција циља може описати линеарним регресионим моделом, чији регресори представљају базисне функције независних променљивих. Сваки ген у хромозому укључује одговарајући регресор у регресиони модел или га искључује. Дакле, сви хромозоми имају облик раније описане бинарне променљиве C_{jk} , чији се саставни елементи (гени) могу наћи у једном од два могућа стања: 0 (регресор није укључен у регресиони модел) и 1 (регресор је укључен у регресиони модел). Пример генетске структуре јединке у популационом базену је дат на слици 6.2.

	x_1 гени				x_2 гени				x_i гени				x_n гени				
MLR _c	1	0	1	...	0	1	...	...	0	1	..	1	...	1	0	0	...
	x_1	x_1^2	$e^{-x_1/2}$	...	x_2	$\sin(x_2)$	...	...	x_i	$x_i^{1/2}$			...	x_n	e^{-x_n}	$\ln(x_n)$	...

Слика 6.2. Генетска структура хромозома у GenReg систему

6.3.2.2. Евалуација

У складу са дефинисаним оптимизационим критеријумом, све јединке (регресионе моделе) би требало рангирати на бази R_{adj}^2 . Додатно, да би се избегло нагомилавање регресора, односно креирање регресионих модела са превеликим бројем регресора, функција добротe укључује и критеријум комплексности. Критеријуми комплексности могу бити различитог нивоa софистицираности (ред полиномијалних функција, процесорско време потребно за обрачун итд) а њихов облик дефинише истраживач на бази теоријских и емпиријских знања.

6.3.2.3. Оператори генетског алгорита

У оквиру MLR/GA методе, предложено је и примењено укрштање у једној тачки (*the single point crossover operator*). У циљу избегавања проблема локалних минимума, у дисертацији је предложена и примењена мутација у којој сваки бит може мутирати независно, са вероватноћом p (*the bit-flip mutation operator*).

6.3.2.4. Селекција и елитизам

За развијену MLR/GA методу је предвиђена *генерацијска турнирска селекција* која из популације бира по два хромозома и на основу дефинисаног *fitness*-a, победника копира у нову генерацију. Да би се обезбедило да најбоља јединка преживи и да се нађе у новој генерацији MLR/GA метода укључује и стратегију елитизма.

6.3.3. Софтверска имплементација MLR/GA методе

MLR/GA метода је имплементирана кроз развој GenReg система. GenReg систем је развијен у Microsoft.NET софтверском окружењу и састоји се од *сервиса за аквизицију података*, *MS SQL базе*, *модула за адаптивно генерисање регресионих модела оптимизованих помоћу генетских алгоритама* и *предикционог модула*. Специјална компонента система је интерфејс заснован на *R* софтверу [17]. Ова компонента је задужена за израчунавање регресионих коефицијената, као и коефицијента детерминације MLR модела у процесу евалуације.

6.4. Валидација MLR/GA методе

Валидација презентоване MLR/GA методе и GenReg система је приказана кроз моделирање структурног понашања бетонских брана. Као пример зависне променљиве, одабрана је променљива која представља радијално померање у хоризонталној равни посматране тачке на телу бране. Код сензора који прате стање предикторских променљивих се јавља повремено нерегуларан рад, чиме је креиран оквир за тестирање развијеног адаптивног система у пуном обиму.

6.4.1. Моделирање структурног понашања бетонских брана

Бране имају веома јаку интеракцију са окружењем, са хидрауличким и геомеханичким факторима, као што су температуре воде и ваздуха, ниво воде у акумулацији, притисак у пукотинама, деформабилност стена и околног тла и други фактори. Сваки од њих утиче на структурно понашање брана [98]. За математички опис структурног понашања брана у последњих неколико декада, као и генерално у другим областима, развијени су бројни детерминистички, статистички и хибридни системи.

У дужем временском периоду, статистички модели су коришћени за мониторинг и праћење сигурности брана као и за утврђивање појединачних утицаја екстерних оптерећења на померање карактеристичних тачака на бранама [24]. Бројни математички модели базирани су на MLR и њиховим напредним формама [24]. Предности и недостаци ових метода, као и у другим доменама, дошли су до изражаја. Ово важи и за детерминистичке и нумеричке методе.

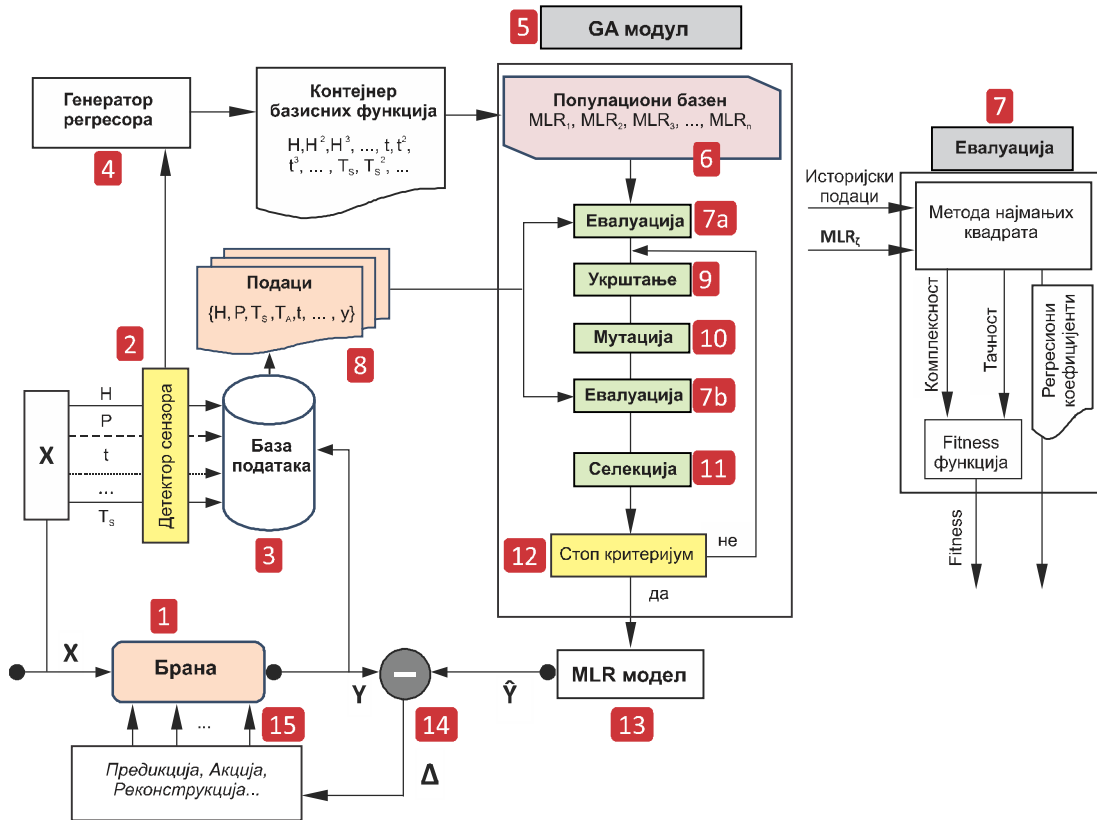
Моделирање структурног понашања брана се нашло и у фокусу модерних AI техника (ANN, GA, ANFIS, SVM, AIA, PCO ..) и хибридних приступа. Истраживачи у последње две декаде покушавају да комбинују предности сваке од ових техника и да осветле проблеме са различитих аспеката. Део референци [28, 31-34, 36,37] је, као пример, наведен у делу 1.4. *Преглед стања у подручју истраживања* а број радова у овој области се значајно повећава последњих година [94, 95, 97, 99].

Тако су, за унапређење предикције, SVR (*support vector regression*) технике обogaћене са GA комбиноване са постојећим ANN и ARMA техникама [94]. У студији [99], Wang и остали су истраживали могућности примене неколико техника из домена вештачке интелигенције за моделирање месечних временских серија које описују процесе варирања речног тока. Ова истраживања обухватила су ANN приступ, ANFIS технику, GP моделе и моделе вектора подршке (SVM), и поређење перформанси ових техника са традиционалним техникама за моделирање временских серија као што су ARMA модели (*autoregressive moving-average models*).

Претходно наведене методе, играју круцијалну улогу у моделирању структурног понашања брана. Ипак, све наведене методе захтевају континуалну расположивост свих улазних параметара, тј. сва мерења морају бити расположива у току читавог периода испитивања и мониторинга. Као што је раније поменуто, при дефинисању уопштеног модела, у реалној примени апликација ситуација је другачија, посебно код инструмената за аквизицију података старијег датума. Откази на оваквој опреми су присутни, што се може утврдити на основу анализе историјских података. Ово значи да се у појединим периодима мониторинга и оцене стања бране, не може рачунати на комплетан скуп података. Другим речима, актуелни модел понашања бране неће бити употребљив јер поједине улазне променљиве неће бити на располагању. Због тога је, као што је наведено у деловима 6.1 и 6.2, развијена хибридна метода која је у одређеној мери отпорна на неправилности у раду аквизиционе опреме.

6.4.2. GenReg систем за бетонске бране

Структура GenReg система дефинисана и приказана у делу 6.1.2 важи генерално, па тако и за моделирање структурног понашања бетонских брана (слика 6.3).



Слика 6.3. Адаптивни систем за моделирање структурног понашања брана базиран на MLR/GA методи

Све одреднице и концепти функционисања GenReg система, као и дефинисаног оптимизационог алгорита важе, уз напомену да се скуп улазно/излазних променљивих, контејнер регресора као и скуп правила и претпоставки за математички модел конкретизују у складу са природом проблема.

За креирање конкретног модела одабрани су следећи значајни фактори: H - горњи ниво воде у акумулацији, P - количина падавина, T_A - температура ваздуха, T_s - температура структуре (температура бетона) као и t - време протекло од почетка експлоатације бране.

Ови фактори (предиктори) су основ за креирање контејнера регресора односно базисних функција. Регресори се могу поделити у неколико група: групу која обухвата ефекте који настају као последица хидростатичког притиска, групу која обухвата ефекте који су проузроковани температуром и групу која обухвата ефекте који настају као последица трајања и експлоатације (старење, које је по својој природи иреверзибилно). Утицаји варирања хидростатичког притиска уобичајено се изражавају помоћу полиномијалних, степених, експоненцијалних и логаритамских форми које зависе од нивоа воде у акумулацији. Померања надгледаних тачака на брани, која настају као

последица температурних напрезања и варијација, могу бити моделирана на различите начине: коришћењем података о температурама ваздуха и/или коришћењем температура карактеристичних тачака у телу бране. Уобичајено је да се број година од почетка експлоатације конструкције узима за анализу, у намери да се математичким моделом обухвате процеси старења материјала.

6.4.3. Математичко-програмерски модел за бетонске бране

У складу са претходним, ако се усвоји претпоставка да се независне променљиве (H, P, T_A, T_S, t) могу представити различитим функционалним формама

$$(H, H^2, H^3, \dots, P, P^2, \sqrt{P}, \dots, T_A, T_A^2, T_A^3, T_A^4, T_A^5, e^{-T_A}, e^{-T_A/2}, \sqrt{T_A}, \ln(T_A), \dots, T_S, T_S^2, T_S^3, T_S^4, T_S^5, e^{-T_S}, \sqrt{T_S}, \ln(T_S), \dots, t, e^{-t}, \ln(t)),$$

онда оптимизациони проблем, који омогућава добијање најбоље комбинације линеарних регресора за моделирање померања тачака на структури бране, добија конкретизовану форму:

$$\text{Максимизовати } R_{adj}^2(\zeta), \quad \zeta = 1, N \quad (6.4)$$

уз следећа ограничења

$$f_{jk} \in \{\chi, \pi, \theta, \tau\} \quad (6.5)$$

$$\chi : \{H, H^2, H^3, \dots\} \quad (6.6)$$

$$\pi : \{P, P^2, \sqrt{P}, \dots\} \quad (6.7)$$

$$\theta : \{T_A, T_A^2, T_A^3, T_A^4, T_A^5, e^{-T_A}, e^{-\frac{T_A}{2}}, \sqrt{T_A}, \ln(T_A), \dots, T_S, T_S^2, T_S^3, T_S^4, T_S^5, e^{-T_S}, \sqrt{T_S}, \ln(T_S), \dots\} \quad (6.8)$$

$$\tau : \{t, e^{-t}, \ln(t), \dots\} \quad (6.9)$$

У претходним изразима χ је подскуп регресора који потичу од хидростатичких ефеката (6.6), π је подскуп регресора који потиче од падавина (6.7), θ је скуп регресора насталих као последица дејства промена у температурном домену (6.8), а τ је подскуп ефеката који су последица старења и експлоатације структуре (6.9). Као што је раније поменуто, једначина (6.5) репрезентује скуп функција које потенцијално могу бити коришћене као регресори у моделу, а конкретна форма ових функција је дефинисана теоретским и емпиријским предзнањима.

У складу са претходним, пример генетске структуре хромозома у популационом базену је приказан на слици 6.4.

	χ гени				π гени				θ гени				τ гени			
MLR	1	0	1	...	0	1	...	1	0	0	1	...	1	0	0	...
	H	H^2	H^3	...	P	P^2	...	T_A	T_A^2	T_S^3	$\ln(T_S)$	...	t	e^{-t}	$\ln(t)$	...

Слика 6.4. Генетска структура хромозома

Прва група гена укључује или искључује утицај регресора који потичу од хидростатичког притиска (χ гени). Другу групу чине гени који се односе на падавине (π гени). Потом следе гени који обухватају утицај температурних ефеката (θ гени) и гени коју узимају у обзир неповратне временске процесе (τ гени). MLR модел који је

репрезентован хромозомом на слици 6.4, је дат једначином (6.10).

$$y_{\zeta} = \beta_0 + \beta_1 \cdot H + \beta_2 \cdot H^3 + \beta_3 \cdot P^2 + \beta_4 \cdot T_A + \beta_5 \cdot \ln(T_S) + \beta_6 \cdot t \quad (6.10)$$

Ипак, додатно је потребно истаћи да прилагођени коефицијент вишеструке детерминације, R_{adj}^2 , узима у обзир број регресора, али је утицај броја регресора на R_{adj}^2 незнатан у случају веома обимног скупа примера за обучавање, односно скупа мерених вредности.

Због тога се, као мера комплексности, у складу са идејама регуларизације описаним у делу 2.3.4, у оквиру ове дисертације уводи додатни коефицијент λ . Ако λ означава минималну промену у вредности коефицијента детерминације због које вреди додати нови регресор, онда се функција добротe може изразити у коначном облику као:

$$\psi(\zeta) = R_{adj}^2(\zeta) - \lambda \cdot n^{(\zeta)}, \quad (6.11)$$

где $n^{(\zeta)}$ представља број регресора у моделу ζ .

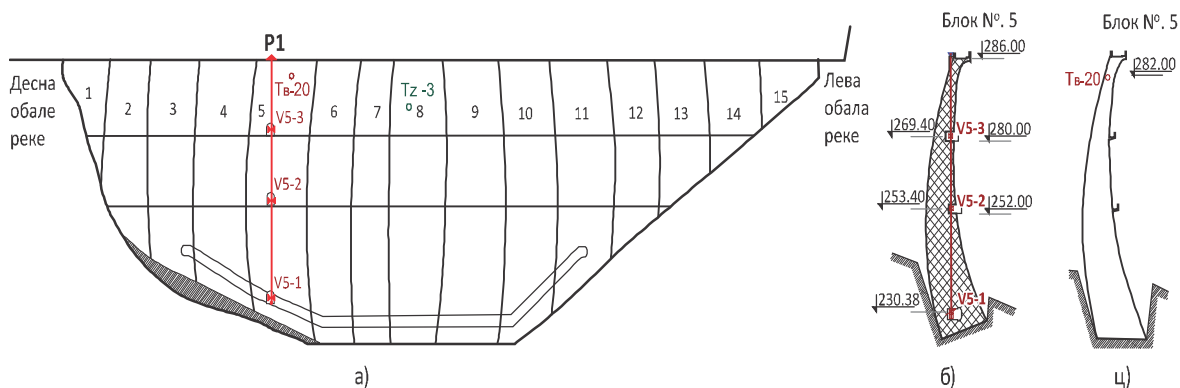
6.4.4. Студија случаја: Примена MLR/GA методе за брану Бочац

Брана Бочац је брана средње величине, која се налази у Републици Српској на реци Врбас, на удаљености приближно 25 km од Бања Луке (слика 6.5.). Ова брана је бетонска лучна брана са двоструким луком, висине 66 m, са дужином круне бране од 221.4 m. Брана је опремељена мониторинг системом који мери параметре као што су стање бетонске структуре, температуре воде и ваздуха, ниво воде у акумулацији, хоризонтална и вертикална померања, ротацију тела бране, померање спојева, напоне, притиске, силе узгона, померање темеља, процуривање... [31].



Слика 6.5. Поглед на брану Бочац

Три клатна (виска) су инсталирана за мерење радијалних и тангенцијалних деформација. За демонстрацију концепта и могућности развијеног GenReg система у моделирању и анализи структурног понашања бетонских брана, је одабрано радијално померање тачке P1 на блоку бр.5. Ово радијално померање је мерено координетром V5-1 (слика 6.6. а и б).



Слика 6.6. (а) Поглед на брану Бочац - узводно и (б, ц) попречни пресек Блока 5

База података је обухватала 2550 мерења (к-торки вектора улазних података односно примера за обучавање). Скуп примера за обучавање је променљив по обиму и садржи дневна мерења почев од 1. јануара 2004. год. до момента рекалкулације оптимизованог модела радијалног померања тачке P1. Поновни обрачун и креирање оптимизованог модела се реализује сваки пут када дође до промене у скупу сензора, односно када год неки од сензора промени своје стање активан/неактиван.

Ефекти хидростатичког притиска на померања у брани се узимају у обзир преко променљиве H , као што је приказано у једначини (5.5). Интервал у коме се ова променљива мења је од 272.28 до 282.24 m. Утицај температуре је узет преко температуре ваздуха T_A (-10.72 до 32.5 °C), чије вредности се мере помоћу сензора Tz-3 (слика 6.6 а), и температуре структуре T_s (1.63 до 20.52 °C), коју мери сензор Tв-20 (слика 6.6. а и ц).

Између температуре ваздуха и температуре бетона постоји линеарна корелација. Pearson-ов коефицијент, који мери ову корелацију, има вредност $r = 0.66$, дакле мању од 0.7, што омогућава да се променљиве T_A и T_s користе као независне улазне променљиве [9]. Падавине немају утицај на понашање бетонских брана, али је и ова променљива (P) третирана као улазна, са циљем да се испита способност развијеног адаптивног система, да одбаци предикторе који немају физички утицај на понашање бране. За посматрани тест период, 15-то дневне падавине су варирале од 0 до 367.6 mm. Временски ефекти укључују утицај иреверзибилних промена у структури материјала, које настају као последица старења материјала и експлоатације објекта (*lifetime*) на померања посматране тачке. Све улазне променљиве су нормализоване на интервал од 0 до 1. У овом примеру, радијално померање тачке P1 је разматрано као излазна променљива, односно зависна променљива. Интервал промене зависне променљиве је био од -4.1 до 15.8 mm.

6.4.5. Резултати и дискусија

Развијена MLR/GA метода и на њој базирани GenReg систем су тестирани на подацима који су мерени у периоду од 1.1.2010. до 31.12.2010. год. у коме су се дешавале периодичне неисправности и нерегуларности у функционисању мерне опреме (табела 6.1).

Табела 6.1. Резултати теста за скуп података у периоду 1.1.2010 и 31.12.2010.

N_0	Период предвиђања	H	P	T_A	T_S	R_{adj}^2	RMSE[mm] оригинални модел	RMSE [mm] адаптирани модел
I	1.1.- 20.1.2010	✓	✓	✓	✓	0.902	1.236	
II	21.1.-18.2.2010	✓	✓		✓	0.792	2.423	0.988
III	19.2.-25.6.2010	✓	✓	✓	✓	0.901	1.835	
IV	26.6.-17.8.2010			✓		0.699	3.718	1.743
V	18.8.-1.12.2010	✓	✓	✓	✓	0.878	2.137	

Параметри GA за сваки од периода из табеле 6.1 су били: број јединки (хромозома) у популационом базену 100, број генерација 70, док је параметар комплексности био $\lambda = 0.05$, што је вредност која се показала оптималном, а утврђена је емпијски.

На почетку разматраног периода, сви сензори су били активни; тако да је скуп од 30 потенцијалних регресора ρ_I обухватао све расположиве променљиве, као што је приказано у следећој једначини:

$$\rho_I : \{H, H^2, H^3, \sqrt{H}, P, P^2, P^3, T_A, T_A^2, T_A^3, T_A^4, T_A^5, e^{-T_A}, e^{-\frac{T_A}{2}}, \sqrt{T_A}, \ln(T_A), T_S, T_S^2, T_S^3, T_S^4, T_S^5, e^{-T_S}, \ln(T_S), \sqrt{T_S}, t, t^2, \sqrt{t}, e^{-t}, e^{-\frac{t}{2}}, e^{-\frac{t}{4}}\} \quad (6.12)$$

GenReg систем је, на бази MLR/GA методе, генерисао оптимизовани MLR модел користећи базу мерења за период од 1.1.2004. до 31.12.2009. год., обухватајући сва потребна мерења истовремено. Најбоље решење, дато једначином (6.13), је садржало 9 регресора, а прилагођени коефицијент вишеструке детерминације је био $R_{adj}^2 = 0.902$.

$$y = 204.95 + 0.89 \cdot H^3 - 109.16 \cdot T_A^3 - 135.46 \cdot e^{-T_A} - 65.72 \cdot \sqrt{T_A} - 62.47 \cdot T_S^4 - 64.09 \cdot e^{-T_S} - 31.34 \cdot t + 48.22 \cdot \sqrt{t} \quad (6.13)$$

У складу са MLR моделом, који је приказан једначином (6.13), се израчунава радијално померања тачке P1 током периода I, када су сви сензори били активни (Табела 1/I). Корен средње квадратне грешке је био $RMSE_{orgI} = 1.236$ mm. Дана 21.1.2010.год., се десио отказ на температурном сензору Tz-3. Ако би се предикција померања тачке P1 и даље заснивала на оргиналном моделу, који подразумева да су сви сензори активни, онда би се грешка предикције, због недостајућих података о температурама ваздуха, повећала на $RMSE_{orgII} = 2.423$ mm. Да би елеминисао нетачност, GenReg систем користи скуп базисних функција ρ_{II} , без температуре ваздуха (6.14) и генерише нови оптимизовани MLR модел који је приказан једначином (6.15).

$$\rho_{II} : \{H, H^2, H^3, \sqrt{H}, P, P^2, P^3, T_S, T_S^2, T_S^3, T_S^4, T_S^5, e^{-T_S}, \ln(T_S), \sqrt{T_S}, t, t^2, \sqrt{t}, e^{-t}, e^{-\frac{t}{2}}, e^{-\frac{t}{4}}\} \quad (6.14)$$

$$y = 73.20 - 62.40 \cdot \sqrt{T_S} + 9.73 \cdot \ln(T_S) - 9.48 \cdot t^2 - 21.66 \cdot e^{-t} \quad (6.15)$$

Прилагођени коефицијент вишеструке детерминације за период обучавања од 1.1.2004. до 20.01.2010. (само дани у којима су сви сензори осим Tz-3 били активни) је

био $R_{adj}^2 = 0.792$, док је грешка предикције адаптираног модела за период **II** била $RMSE_{adpII} = 0.988$ mm, што је значајно боље предвиђање од оног које нуди првобитни модел. Слика 6.7 даје приказ поређења измерених померања тачке P1 и померања која су предвиђена коришћењем оригиналног и адаптивног модела. Скуп података који је служио за обуку није приказан, због прегледности.

Почев од 19.2.2010. па до 25.6.2010. год. сви сензори су били поново активни. Дана 19.2.2010. год., GenReg софтверски агент је, према MLR/GA методологији, генерисао нови, оптимизовани модел, користећи скуп података за обучавање из периода 1.1.2004. до 18.02.2010. год (само за дане у којима су сви сензори били активни). Нови модел је, над скупом за учење, имао тачност од $R_{adj}^2 = 0.901$ и грешку предикције за период **III**, $RMSE_{orgIII} = 1.835$ mm.

Тачност предикције регресионог модела које је GenReg оптимизовао је била у фокусу анализе посебно током периода у којима значајан број сензора не ради правилно или није активан. На пример, током периода **IV**, једини активни сензор је сензор који мери температуру ваздуха (видети слику 6.7). Ако би се предикција наставила коришћењем оригиналног модела, који подразумева да су сви сензори активни, онда би тачност предвиђања рапидно пала на $RMSE_{orgIV} = 3.718$ mm. Уместо тога, GenReg систем, 26.6.2010. год., генерише нови оптимизовани модел (6.17) коришћењем расположивих базисних функција из контејнера ρ_{III} :

$$\rho_{III} : \{T_A, T_A^2, T_A^3, T_A^4, T_A^5, e^{-T_A}, e^{-\frac{T_A}{2}}, \sqrt{T_A}, \ln(T_A), t, t^2, \sqrt{t}, e^{-t}, e^{-\frac{t}{2}}, e^{-\frac{t}{4}}\} \quad (6.16)$$

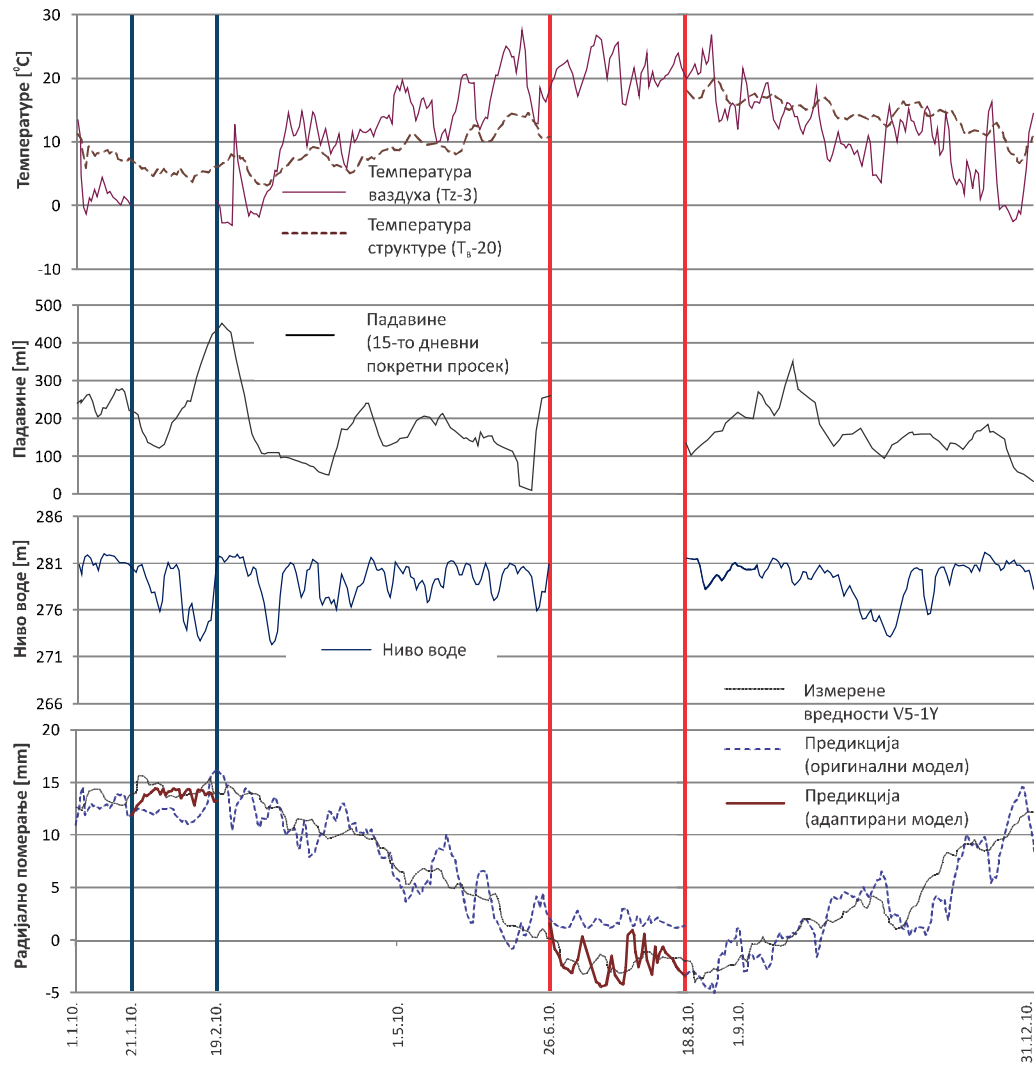
$$y = 32.20 - 58.82 \cdot T_A^3 + 45.68 \cdot T_A^5 - 8.24 \cdot t^2 - 20.85 \cdot e^{-\frac{t}{2}} \quad (6.17)$$

Иако у моделу приказаном једначином (6.17) недостаје један од важних параметара, висина нивоа воде у акумулацији (H), овај тип модела још увек може да генерише прихватљиве предикције, док оригинални модел губи функционалост у великом обиму ако неки од значајних улазних параметара није расположив.

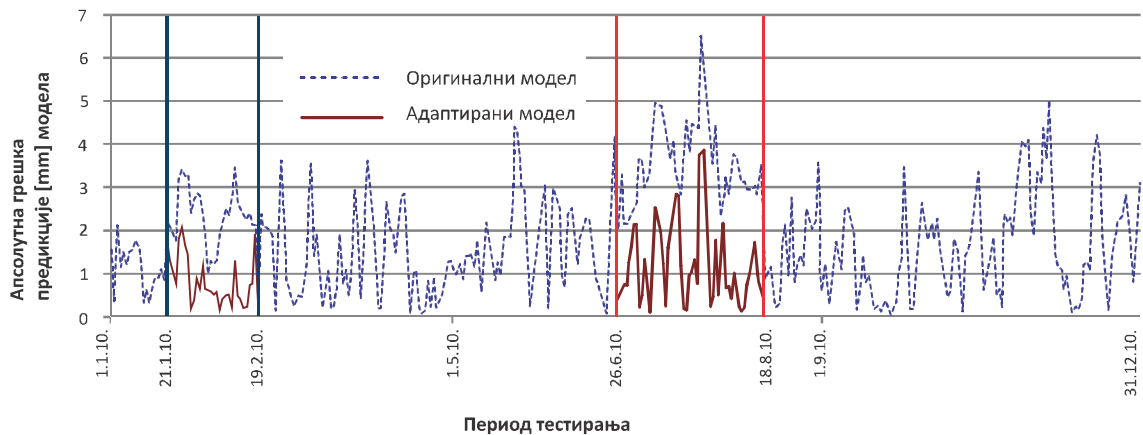
Тачност модела приказаног у једначини 6.17, који је креиран над скупом података за учење, за период 1.1.2004 – 25.6.2010. год. (за дане у којима је активан само сензор температуре) био је $R_{adj}^2 = 0.699$, док је грешка предикције за период **IV** била $RMSE_{adpIV} = 1.743$ mm, што је поново значајно боље у односу на оригинални модел.

У току периода од 18.8.2010. до 31.12.2010. год., сви сензори су били активни. У складу са описаном процедуром, GenReg систем је генерисао нови оптимизациони модел чија је тачност $R_{adj}^2 = 0.878$, а грешка предикције $RMSE_{orgV} = 2.137$ mm.

Компарација тачности између оригиналног модела и модела адаптираних на бази MLR/GA методе помоћу GenReg система је дата на слици 6.8.



Слика 6.7. Поређење оригиналног и адаптираног предикционог модела



Слика 6.8. Апсолутна грешка предикције оригиналног и адаптираног модела

Анализа прилагођеног коефицијента вишеструке детерминације у табели 6.1, посматрано са једне стране, води ка закључку да тачност модела умногоме зависи од

броја расположивих улазних променљивих; овај коефицијент варира од вредности 0.902 (када су све променљиве укључене) до 0.699 (када се разматрају само мерења температуре ваздуха).

Посматрано са друге стране, модели високе тачности, који су базирани на великом броју улазних параметара, показују рапидан пад у тачности предикције када, на пример, због отказа одговарајућег сензора неке од улазних променљивих више нису расположиве. Модели са мањим бројем улазних параметара имају уобичајено мању тачност. Ипак, они су корисни у ситуацијама у којима је број мерења редукован. Они производе боље предикције од предикција које се добијају помоћу модела који имају велики број улазних променљивих, али је део променљивих недоступан.

GenReg систем адаптира предикциони модел сваки пут када се скуп расположивих сензора промени. Грешка оригиналног предикционог модела у случају отказа сензора за мерење температуре би износила $RMSE=2.423$ mm. Коришћење модела који GenReg систем под тим условима генерише, ову грешку смањује са $RMSE=2.423$ mm на 0.988 mm. Слично, у периоду када је сензор за мерење температуре једини активни сензор, оригинална грешка $RMSE=3.718$ mm се редукује на 1.743 mm коришћењем адаптивног модела.

Из једначина (6.13, 6.15 и 6.17), се може приметити да се регресори са независном променљивом P , која означава падавине, не користе ни у једном моделу, иако су ови регресори расположиви у контејнеру базисних функција. Овакав избор је природан јер падавине немају значајан утицај на бетонске бране, које су биле предмет анализе у овој студији.

У циљу јасноће и прецизности, само једна излазна променљива (радијално померање тачке P1) је моделирана у овој студији. У пракси, мониторинг систем за надгледање брана, уобичајено омогућава мерење великог броја излазних променљивих, које могу бити коришћене за оцену структурног понашања. У складу са карактеристикама MLR метода, да моделира сваку од излазних променљивих независно, презентирани и тестирани принцип се може применити за моделирање више излазних компоненти истовремено и независно један од другог.

Генерални закључци о перформансама развијене MLR/GA методе и њеној способности да путем GenReg софтверског агента генерише адаптивне MLR моделе структурног понашања бетонских брана, за променљиве скупове података и променљиве скупове предикторских променљивих, су дати у поглављу 10.

7. Адаптација ANN модела помоћу генетских алгоритама

У оквиру ове дисертације GA су као оптимизационе хеуристичке технике, поред других примена, употребљене и као алат за свеобухватну оптимизацију математичких модела процеса, објеката и система у форми вештачких неуронских мрежа. ANN са својом архитектуром (бројем слојева и бројем неурона по слојевима), алгоритмом учења, активационим функцијама и правилима учења, представља јединку у GA популацији. Популација представља скуп мрежа, различитих топологија, различитих алгоритама учења, различитих активационих функција и правила учења. Из генерације у генерацију, у складу са GA парадигмом, кроз процесе селекције и репродукције, боље мреже преживљавају и просечни *fitness* популације расте. Концепт елитизма обезбеђује да се генетска грађа најквалитетније мреже чува током еволуције. Након испуњења критеријума заустављања генетског алгорита се очекује да ANN колонија, која одговара последњој генерацији, садржи и вештачку неуронску мрежу, која дати процес или систем описује на довољно добар начин.

7.1. Стање у области

Већ је истакнуто да је примена вештачких неуронских мрежа нашла своје место у скоро свим научним дисциплинама, али да избор и дизајн ANN, још увек представља проблем за чијим решењем се трага и чија оптимизација представља изазов. У наредном тексту се, у циљу увида у актуелне тенденције по питањима ANN оптимизације, даје преглед неколико радова објављених последњих година.

У раду [1], Fernando Gaxiola са коауторима, даје приказ перформанси и предности генетских алгоритама и PSO техника, при оптимизацији тежинских коефицијената неуронских мрежа, које се обучавају по *backpropagation* алгоритму. Компаративна студија је базирана на *benchmark* случају предикције у проблему Mackey-Glass временских серија. Hosseinia S.M. и Mahjouri N., у раду [101] примењују хибридную технику засновану на методама регресионих вектора подршке и вештачким неуронским мрежама за моделирање веома комплексних, нелинеарних прогноза дневних падавина. Капацитет предложеног SVR-ANN концепта је тестиран за симулацију падавина у три одабрана подручја Ирана. Способности предложеног хибрида, поређене су са класичним SVR, ANFIS моделима, као и хибридом вештачких неуронских мрежа, чији су тежински коефицијенти оптимизовани са генетским алгоритмима (ANN-GA). У већем броју случајева SVR-ANN је показао низ предности. У раду [102], Chanda S. са коауторима, даје приказ развијеног комбинованог ANN-GA модела за оптимално геометријско дизајнирање имплантата и утицаја облика на постоперативни поступак. Развијени хибрид је алтернатива класичној оптимизацији геометрије облика помоћу коначних елемената. Tolga Baklacioglu у својој студији [103], користи генетске алгоритме за оптимизацију топологије вештачке неуронске мреже која предвиђа потрошњу горива у авиотранспорту, на бази реалних податка о летовима. У раду [104], Hui Liu и коаутори, користе неколико AI техника, за моделирање брзине снажних

ветрова у циљу боље заштите од могућих последица. Коришћена су два хибрида [FEEMD¹-MEA²-MLP/FEEMD-GA-MLP] помоћу којих су сложене серије података дељене у више слојева (подскупова). Потом је реализовано прецизно предвиђање брзине снажних ветрова, помоћу ANN које су оптимизоване са MEA или GA техникама. У раду [105], Oleg E. Bukharov и Dmitry P. Bogolyubov, дају приказ развијеног система на бази генетских алгоритама и вештачких неуронских мрежа, за подршку одлучивању код комплексних система. Sánchez D. и коаутори су у студији [106] развили нови модел модуларних неуронских мрежа³ које су оптимизоване са хијерархијским генетским алгоритмима, за истраживања у области препознавања на бази информација и параметара о човековом лицу. У раду [107], Chandwani V. и коаутори примењују ANN мреже оптимизоване генетским алгоритмима, за моделирање слегања унапред припремљеног бетона⁴ у функцији концентрације саставних елемената као што су: цемент, пепео, песак, агрегати, додатне мешавине, као и количинског односа вода-везивно средство. Студија показује да ANN мрежа, хибридризована са GA брже конвергира и да има бољу тачност у односу на класични приступ.

Из претходних приказа се види да је модерна примена математичких модела у форми ANN и даље актуелна.

У складу са наведеним, у овој докторској тези се предлаже и развија само-подешавајући ANN/GA хибрид, односно рационални AI агент за моделирање понашања објеката и система који:

- генерише математичке моделе у форми вештачких неуронских мрежа, које су оптимизоване са аспекта топологије, алгоритама учења, активационих функција и правила учења и
- који је робустан на промене, како у скупу улазних променљивих, тако и на промене у скупу мерења, чији се обим стално повећава захваљујући развоју модерне аквизиционе и мерне технике.

7.2. ANN/GA модел. Опис система

7.2.1. Кибернетски приказ ANN/GA модела

Самоподешавајући хибрид ANN/GA је развијен на еволутивном систему вештачких неуронских мрежа које се оптимизују помоћу генетских алгоритама (слика 7.1).

Понашање објекта моделирања (1) се може посматрати као црна кутија која трансформише вектор улазних променљивих $\mathbf{X}$ (2) у вектор излазних променљивих $\mathbf{Y}$. Измерене вредности улазних променљивих, $x_1, x_2, \dots, x_i, \dots, x_n$, као и вредности излазних променљивих од интереса, се складиште у базу података (3).

Хибрид ANN/GA на случајан начин генерише иницијалну популацију јединки у форми ANN (4). Својства сваке од иницијално генерисаних мрежа, репрезентована су бинарним хромозомом. Да би се пронашла вештачка неуронска мрежа која најбоље описује понашање објекта моделирања, популација мрежа се оптимизује у еволуционом процесу (5, 6, 7, 8, 9), који је описан у деловима 5.3.1, 5.4, 5.5, 6.2 и 6.4.2.

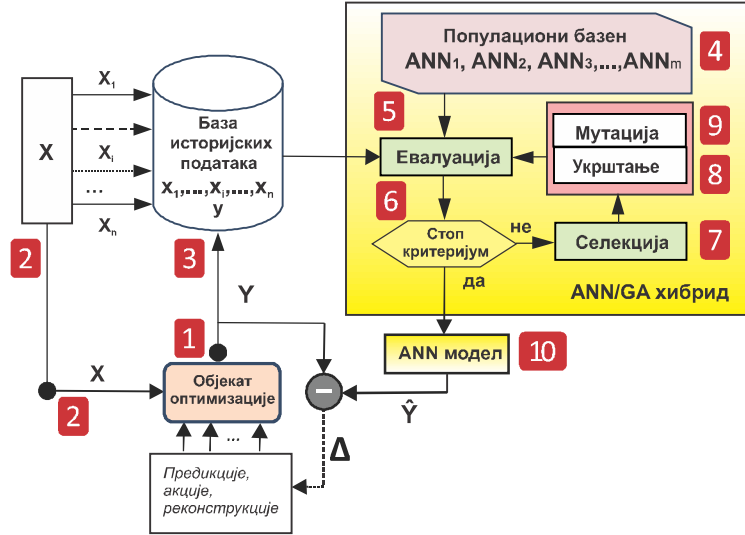
¹ Fast Ensemble Empirical Mode Decomposition

² Mind Evolutionary Algorithm

³ Modular Neural Network

⁴ Ready Mix Concrete (RMC)

Оптимизациона процедура се понавља неколико пута да би се редуковао ефекат инхерентне стохастичке природе GA.



Слика 7.1. Самоподешавајући систем за моделирање објеката базиран на прилагодљивим неуронским мрежама

Крајњи производ ANN/GA рационалног агента, добијен за познате вредности улазних променљивих, је оптимизовани математички модел (10) у форми ANN. Овај модел може предвидети понашање објекта моделирања и то понашање изразити нумерички кроз матрицу $\hat{Y}$. Тако добијени математички модел се даље може користити за предикцију понашања моделираног објекта, за анализе и поређења са мерењима, као и за предузимање могућих акција.

7.2.2. Математичка формулација оптимизационог проблема

Предложени систем одликује способност оптимизације и функционисања у реалном времену. Концепт подразумева оптимизацију свих елемената мреже са циљем генерисања модела који, на најбољи начин, описује понашање моделираног објекта под датим условима.

Елементи који су предмет оптимизације су број скривених слојева, број неурона по слојевима, активационе функције, алгоритми и правила учења.

Нека Ω означава скуп свих могућих ANN модела који садрже највише $N_{\max}$ скривених слојева са максимално $n_{\max}$ неурона по слоју. Нека сваки од неурона у оквиру јединке користи активациону функцију (AFs) која се бира на нивоу мреже и која се може разликовати од јединке до јединке. Додатно, нека сваки ANN модел користи један од понуђених алгоритама учења и нека се сваки од алгоритама учења додатно, фино, подешава параметрима, чије су вредности из препоручених интервала. Нека $RMSE_{Test}^{(w)}$ означава корен средње квадратне грешке, коју генерише обучена мрежа $ANN^{(w)}$ из Ω , за тест скуп података.

Онда, оптимизациони проблем који проналази најбољу ANN топологију и најбољи избор AF, алгоритама учења и правила учења, може бити дефинисан у следећој форми:

$$\text{Минимизовати } RMSE_{Test}^{(\psi)}, ANN^{(\psi)} \in \Omega \quad (7.1)$$

при чему важи

$$N^{(\psi)} \leq N_{\max} \quad (7.2)$$

$$n^{(\psi)} \leq n_{\max} \quad (7.3)$$

$$AF^{(\psi)} \in \{Gaussian, sine, TANH, \dots\} \quad (7.4)$$

$$LA^{(\psi)} \in \{BP Gradient Descent, RPROP, \dots\} \quad (7.5)$$

$$a \in [a_{\min}^{(\psi)}, a_{\max}^{(\psi)}] \quad (7.6)$$

где су $N^{(\psi)}$, $n^{(\psi)}$, $AF^{(\psi)}$, $LA^{(\psi)}$, број скривених слојева, број неурона у скривеним слојевима, активациона функција и алгоритам учења неуронске мреже $ANN^{(\psi)}$, респективно. У једначини 7.6, променљива a означава неки од параметара η , μ , η^+ или η^- .

Као што је већ поменуто, у циљу рачунарске ефикасности, математичка формулација оптимизационог проблема подразумева и ограничење да се иста, случајно изабрана активациона функција (AF), користи за све неуроне у скривеним слојевима, као и за неуроне у излазном слоју неуронске мреже $ANN^{(\psi)}$.

7.2.3. Генетски алгоритми за оптимизацију вештачких неуронских мрежа

У складу са комплексношћу оптимизационог проблема, у развијеном ANN/GA рационалном агенту, као оптимизациона техника је коришћена техника генетских алгоритама. Развијена оптимизациона методологија (ANN/GA хибрид) је базирана на итеративној стратегији приказаној на слици 7.1, где индивидуа представља вештачку неуронску мрежу из Ω . Свака неуронска мрежа $ANN^{(\psi)}$ из иницијалне популације се евалуира у циљу одређивања њеног *fitness*-а. Мреже са нижим $RMSE_{Test}^{(\psi)}$ се сматрају бољим, тако да је *fitness* рачунат по формули $1 / RMSE_{Test}^{(\psi)}$, јер је у дисертацији развијена софтверска имплементација GA парадигме, заснована на максимизовању *fitness*-а. У складу са одабраним критеријумом селекције и концептом елитизма, мреже са нижим вредностима *fitness*-а се одбацују. Када се дефинисани критеријум оствари (дефинисано време, захтевана тачност), добија се популација оптимизованих вештачких неуронских мрежа. Пошто постоји само један оптимизациони критеријум (RMSE), мрежа са најбољим *fitness*-ом у финалној популацији се бира као оптимална.

7.2.4. Структура ANN хромозома за ANN/GA хибрид

У складу са Holland-овим оригиналним приступом, јединке у популационом базену имају форму бинарних хромозома. Примењена генетска структура јединки у популационом базену је дата на слици 7.2.

ANN ^(ψ)																								
N ^(ψ) гени					n ^(ψ) гени					AF ^(ψ) гени			LA гени		LP1 ^(ψ) гени				LP2 ^(ψ) гени					
0	1	0	...	1	0	1	...	1	0	0	0	1	...	0	1	...	1	0	...	1	1	...	0	1

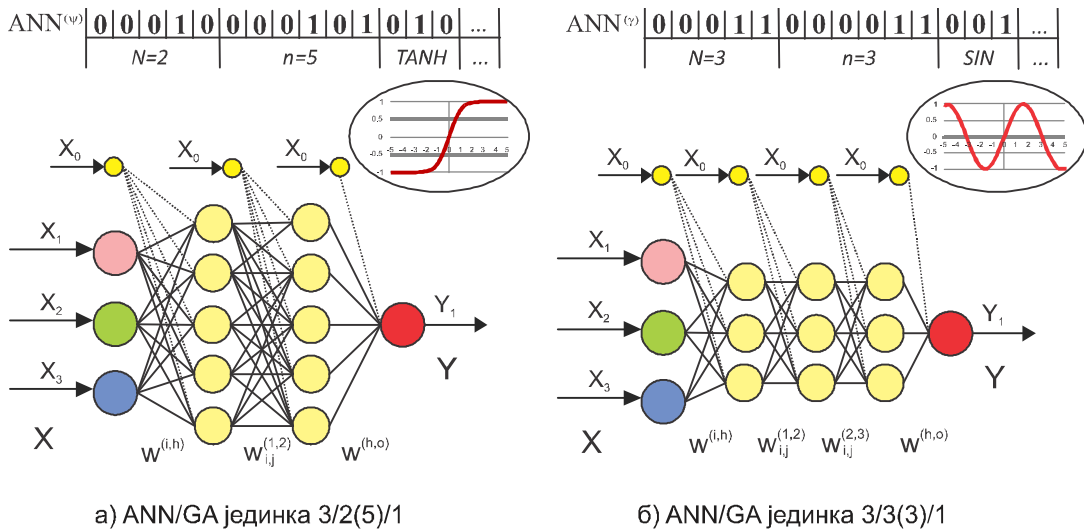
Слика 7.2. Генетска структура ANN/GA хромозома

Прве две групе гена репрезентују број скривених слојева $N^{(\psi)}$ и број неурона $n^{(\psi)}$ у овим слојевима. Може се уочити да је усвојено, да број неурона у свим скривеним слојевима буде исти $n_1^{(\psi)} = \dots = n_{N_{max}}^{(\psi)} = n^{(\psi)}$, да би се задржала иста дужина хромозома у читавој популацији.

Следећа група гена репрезентује тип активационе функције $AF^{(\psi)}$. Гени $LA^{(\psi)}$ у хромозому дефинишу врсту алгорита учења, док $LPI^{(\psi)}$ и $LP2^{(\psi)}$ гени укључују параметре одабраног алгорита (η и μ за BPGD, или η^+ и η^- за RPROP алгоритам).

Као пример, слика 7.3, приказује две ANN јединке, где обе мреже имају 3 предиктора и једну зависну променљиву. Бинарни хромозом на слици 7.3 а, репрезентује ANN са 2 скривена слоја, који имају по 5 неурона са TANH активационим функцијама. Због јасноће приказа, гени који репрезентују алгоритам учења и правила учења нису приказани на слици.

Хромозом $ANN^{(\psi)}$, који је приказан на слици 7.3 б, репрезентује кодирани облик ANN са 3 скривена слоја од по 3 неурона и синусном активационом функцијом. У означавању топологије мреже коришћена је нотациона форма $I / N(n) / O$, где су I , N , n и O : број предиктора, број скривених слојева, број неурона по слојевима и број излазних променљивих, респективно. На тај начин, топологија јединки $ANN^{(\psi)}$ и $ANN^{(\gamma)}$ може бити записана у форми $3/2(5)/1$ и $3/3(3)/1$.



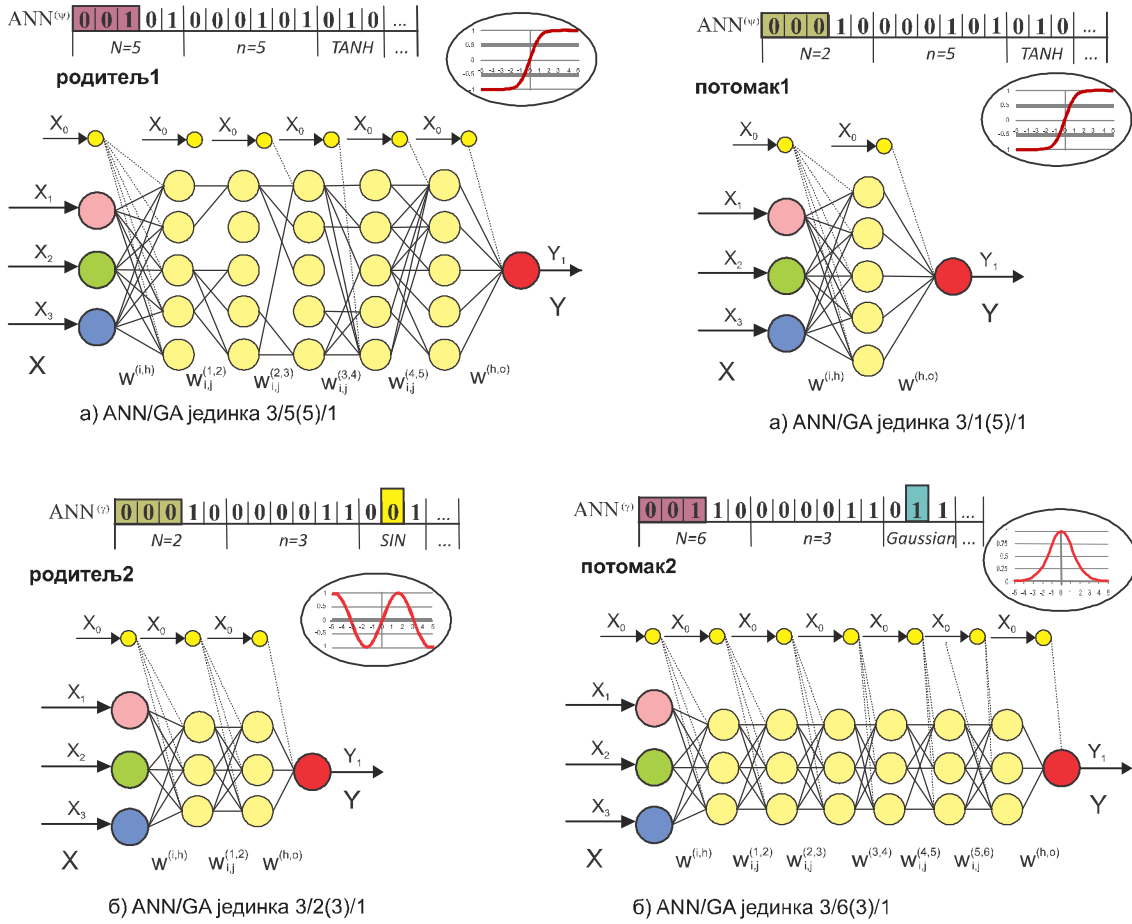
Слика 7.3. Примери генетске структуре ANN/GA хромозома

7.2.5. Оператори генетског алгорита за ANN/GA хибрид

За оператор укрштања ANN јединки у оквиру ANN/GA хибрида, је коришћен оператор укрштања у једној тачки (*single point crossover*). У циљу избегавања проблема локалног минимума, сваки од бита може мутирати независно са задатом вероватноћом мутације p , коришћењем оператора мутације: a bit-flip mutation operator [6].

Укрштањем и мутацијом хромозома настају мреже различитих архитектура, алгорита учења и разноврсних комбинација активационих функција и правила учења. Управо ове карактеристике обезбеђују усмерену глобалну претрагу простора могућих решења (различитих ANN) и проналажење оптималне мреже са великом вероватноћом.

На слици 7.4, је дат пример мрежа насталих у процесу репродукције.



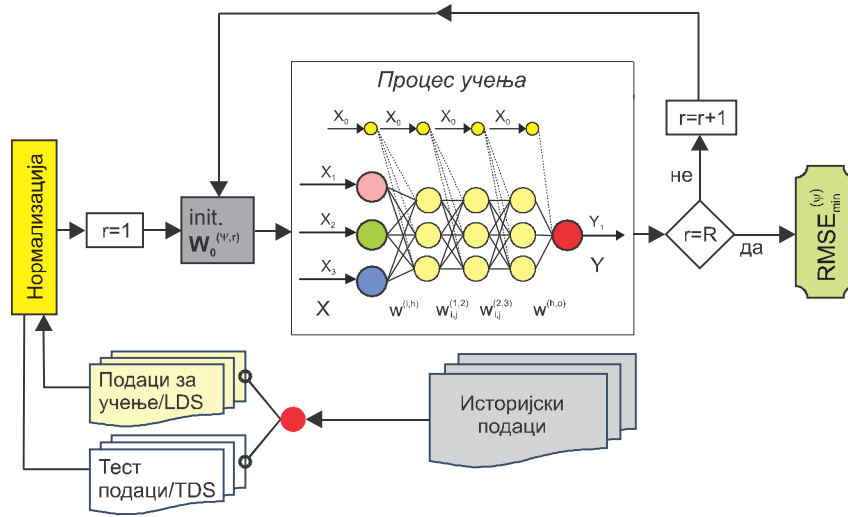
Слика 7.4. Примери ANN јединки након операције укрштања и мутације

Топологија јединки родитеља, $ANN^{(\psi)}$ са активационом функцијом TANH и $ANN^{(\gamma)}$ са синусном активационом функцијом (*sine*), може бити записана у форми 3/5(5)/1 и 3/2(3)/1. Једнопозицијско укрштање (*single point crossover*) се одиграло у тачки која одговара трећем гену са леве стране. Група гена 001 са леве стране хромозома **родитеља1** замењена је групом гена 000 **родитеља2**, и обрнуто. Као резултат креиране су и две нове јединке (мреже потомци) са топологијом 3/1(5)/1 и 3/6(3)/1. Како је реч о једнопозиционом укрштању, **потомак1** има различиту архитектуру у односу на своје родитеље а задржао је активациону функцију TANH. Такође, **потомак2** има значајно различиту топологију у односу на своје родитеље. Код ове јединке је дошло и до мутације на средишњем гену у групи гена који описују тип активационе функције. Синусну активациону функцију је заменила активациона функција у облику Гаусове криве (*Gaussian*).

7.2.6. Евалуација ANN јединки

Евалуација сваке $ANN^{(\psi)}$ јединке се реализује по концепту који је приказан на слици 7.5. Садржај хромозома дефинише активациону функцију, алгоритам учења и правила учења који се користе у неуронској мрежи која је репрезентована овим хромозомом. У складу са активационом функцијом, подаци у подскупу за обучавање (LDS) и подскупу за тестирање (TDS) су нормализовани. После иницијализације

тежинских коефицијената w_{ij} , стохастички одабраним вредностима из задатог интервала, реализује се алгоритам учења са дефинисаним правилима учења као што је раније описано.



Слика 7.5. Процес евалуације ANN јединки

Процес учења се прекида у зони *прекомерног учења*⁵ [52]. У циљу редукције утицаја стохастичког генерисања иницијалне матрице тежинских коефицијената ($\mathbf{W}_0^{(\psi,r)}$) на квалитет $ANN^{r(\psi)}$ модела, читава процедура се понавља неколико пута са различитим иницијалним вредностима. На крају евалуационог процеса, $fitness ANN^{r(\psi)}$ јединке се рачуна као $1/RMSE_{Test}^{(\psi)} = 1 / \min(RMSE_{Test}^{(\psi, \mathbf{W}_0^{(\psi,r)})})$, $r=1, R$, где је R број понављања, који арбитарно бира истраживач.

7.2.7. Селекција

Према оптимизационом критеријуму који је дефинисан у презентираном математичко-програмерском моделу, све јединке (моделу) се рангирају на бази $RMSE_{Test}^{(\psi)}$, $\psi \in \Omega$. За развијени ANN/GA хибрид у овој докторској дисертацији, коришћена је *бинарна турнирска* селекција, која реализује такмичење пара вештачких неуронских мрежа и бира мрежу победника коју преноси у следећу генерацију. У циљу обезбеђења да најбоља ANN мрежа увек опстане и пренесе се у следећу генерацију, такође се примењује и стратегија елитизма [6].

7.3. Софтверска имплементација

Презентирани ANN/GA агент је развијен у Microsoft .NET софтверском окружењу и садржи сервис за аквизицију података, MS SQL базу података и модул за оптимизацију ANN модела објекта који је предмет моделирања. Специјална компонента система је интерфејс ка софтверској библиотеци *Encog software library* [47], чија је улога везана за процес учења ANN.

⁵ *overlearning zone*

7.4. Валидација ANN/GA хибрида

Валидација презентираних концепта ANN/GA хибрида, као и код MLR/GA софтверског агента, је приказана кроз моделирање структурног понашања бетонских брана.

Као што је раније наведено, пуна адаптивност модела брана у реалним околностима, захтева способност модела да оперише са сталним повећањем скупа мерених података о систему као и променљивим скупом предикторских променљивих. Променљивост скупа предикторских променљивих узрокована је континуалним променама и отказима компоненти система за мониторинг и мерење. У намери да се валоризује ANN модел понашања бетонских брана, који је у потпуности прилагодљив променљивом скупу како мерених података тако и предиктора, у овом делу дисертације се приказује конкретизована GA методологија за оптимизацију модела структурног понашања брана у форми вештачких неуронских мрежа.

7.4.1. Моделирање структурног понашања брана помоћу ANN/GA парадигме

У великом броју радова [45, 108-111, 53] бројни истраживачи су уложили значајан напор за проналажење оптималних ANN структура, које су користили за моделирање различитих проблема, а који се директно или индиректно односе на моделирање понашања бетонских брана или тангирају ову проблематику. Majdi и остали [45] су комбиновали неуронске мреже са генетским алгоритмима за предикцију модула деформисања код стена. GA концепт је коришћен за проналажење оптималног броја неурона у скривеном слоју, фактора учења и моментума како у скривеном тако и у излазном слоју мреже. Коришћењем стандардног алгоритма са падајућим градијентом и простирањем грешке уназад, они су тестирали мреже са линеарним и сигмоидним активационим функцијама. Zhou и коаутори су у студији [109] презентирали комбиновану процедуру ортогоналног пројектовања, FE анализе, ANN и GA за инверзно моделирање *seepage/leakage* проблема. Одабрана неуронска мрежа користила је сигмоидну трансфер функцију и имала је фиксан број слојева: један улазни слој, два скривена слоја и један излазни слој. Број неурона у скривеним слојевима одређен је минимизацијом функције грешке за тест подкуп података коришћењем метода проба и грешака. За добијање краћег времена учења и високе тачности уопштавања, је коришћен LM алгоритам⁶ комбинован са Bayesian регуларизацијом. У студији [108], хибридна анализа коначних елемената/граничних елемената (FE-BE)⁷ укрштена са ANN процедурама је предложена за предикцију динамичких карактеристика постојеће бетонске гравитационе бране. Укрштени градијентни алгоритам⁸ (CGA) и LM алгоритам су имплементирани за брзо тренирање ANN. Аутори су тестирали неуронску мрежу са једним скривеним слојем где је број неурона био одређен методом проба и грешака. Hooshyaripour и други у чланку [111] су показали да је трослојни ANN модел прихватљив за рад са проблемом пуцања бране, који има два улаза: ниво и запремина воде у акумулацији у тренутку оштећења, и један излаз: пик пражњења акумулације. У њиховој студији је коришћен модел мреже са једним скривеним слојем и простирањем унапред. Применом Hecht-Nielsen критеријума [53], утврђено је да ANN са четири неурона у скривеном слоју има најбоље перформансе. Алгоритам LM је ангажован за

⁶ the Levenberg-Marquardt back-propagation algorithm

⁷ FE-BE – a hybrid finite element-boundary element analysis

⁸ The conjugate gradient algorithm

тренирање ANN мрежног модела, док су као активационе функције примењене тангенс-сигмоидна функција и линеарна функција у скривеном и излазном слоју, респективно.

7.4.2. Имплементација ANN/GA хибрида за моделирање понашања брана

У поглављу 6., приказан је развијени адаптивни систем за моделирање структурног понашања брана, базиран на линеарним регресионим моделима. Развијени систем је, за дате услове, оптимизован генетским алгоритмом. MLR/GA софтверски хибрид, који је развијен током истраживања, поседује способност да кроз еволуциони процес у приближно реалном времену прилагођава регресоре MLR модела у складу са тренутно активним сензорима.

У даљој разради идеја моделирања, у циљу уопштавања, компарације и унапређења квалитета математичких модела за структурно понашање брана, пројектован је систем базиран на вештачким неуронским мрежама, способан да се адаптира сталним променама скупа предиктора и базе мерених података. Да би се остварила пуна адаптивност ANN модела у реалним условима, као задатак је постављен циљ да ANN/GA хибрид буде у стању да оптимизује све значајне мрежне параметре у складу са расположивим улазним сензорима и базом историјских података.

Према сазнањима аутора дисертације и чланова тима који су учествовали у разради идеје, постојећа решења у овом домену оптимизују само ограничени скуп ANN параметара, док се остали параметри бирају арбитрарно, на бази искуства, литературе или методом пробе и грешке.

У деловима дисертације који следе, је презентиран систем за аутоматско генерисање ANN модела брана, који оптимизује све значајне елементе ANN архитектуре. Управљан променљивим скупом предиктора, систем континуално оптимизује мрежну топологију, избор активационих функција и алгоритама учења са намером да креира моделе који ће бити добро прилагођени растућој бази мерених података. Оптимизација параметера је реализована помоћу генетских алгоритама. Квалитет предложеног ANN/GA хибридног модела брана је тестиран помоћу студије случаја из реалног света и поређен са еквивалентним моделом бране заснованим на вишеструким линеарним регресијама и GA (MLR/GA).

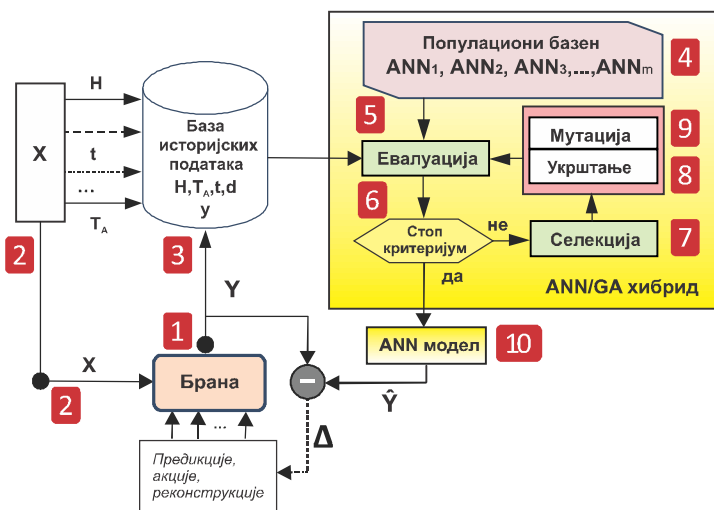
Развијени ANN/GA систем⁹ за моделирање брана у потпуности одговара кибернетској структури, оптимизационом проблему и теоријским одредницама описаним у делу 7.2.1. Једина разлика је што је скуп улазно/излазних променљивих конкретизован за потребе моделирања структурног понашања брана (слика 7.6).

Измерене вредности улазних променљивих, као што су висина нивоа воде у акумулацији H , температура ваздуха T_a , и време протекло од почетка експлоатације бране t (старење), затим вредности излазних променљивих од интереса, као што су померања и напони, се складиште у базу података (3). Ипак, ако су мерења температуре некомплетна или нерасположива, тригонометријске функције, као циклична (сезонска) температурна оптерећења могу бити коришћене за опис деформационе шеме [29]:

$$\begin{aligned}\sin(2 \cdot k \cdot \pi \cdot d / 365), k = 1, 2, 3 \dots \\ \cos(2 \cdot k \cdot \pi \cdot d / 365), k = 1, 2, 3 \dots\end{aligned}\tag{7.7}$$

где је d редни број дана у години.

⁹ A self-tuning system for Dam behavior modeling based on **EV**olving Neural Networks - DEVONNA



Слика 7.6. Самоподешавајући систем за моделирање структурног понашања брана базиран на прилагодљивим неуронским мрежама - DEVONNA

Као што је описано у делу 7.2.1, хибрид ANN/GA генерише, на случајан начин, иницијалну популацију јединки у форми ANN [26]. Да би се пронашла мрежа која најбоље представља историју понашања бране, популација мрежа се оптимизује у еволуционом процесу (5, 6, 7, 8, 9) који је описан у деловима 5.3.1, 5.4, 5.5, 6.2 и 6.4.2. У процесу евалуације сваке јединке (неуромреже) тежи се смањењу утицаја стохастичке иницијализације вектора тежинских коефицијената на квалитет добијеног математичког модела (слика 7.5). Крајњи производ DEVONNA система је оптимизовани математички модел $\hat{Y}$ структурног понашања бране (10) у форми ANN, расположив за различите намене.

У следећим деловима овог поглавља је приказана валидација развијеног DEVONNA система коришћењем студије случаја за моделирање померања бране Гранчарево.

7.5. Студија случаја: Брана Гранчарево

Брана Гранчарево представља први корак у хидроенергетском систему реке Требишњице у Републици Српској (слика 7.7).



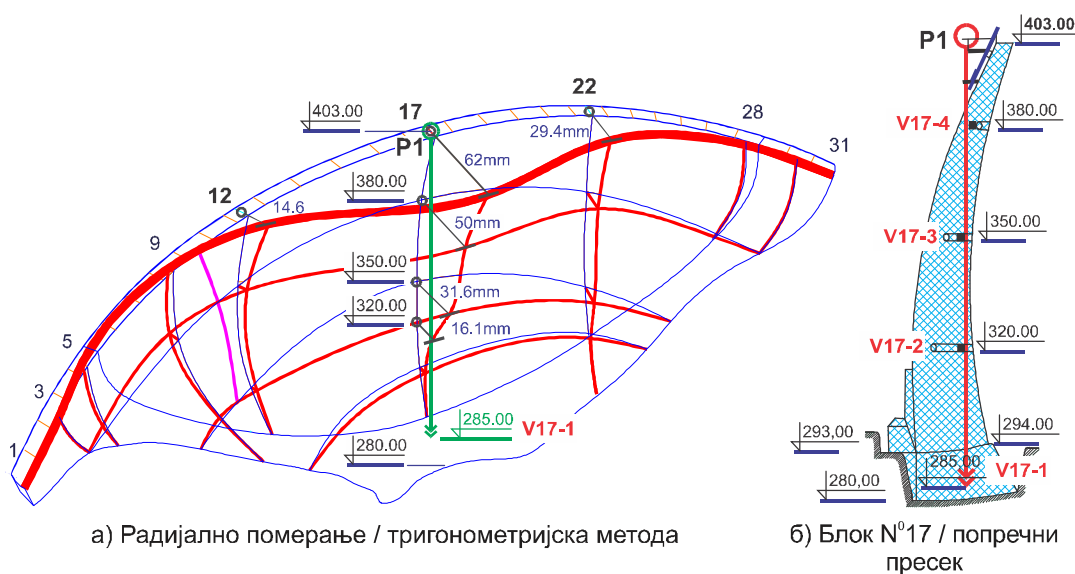
Слика 7.7. Поглед на брану Гранчарево

Брана је лоцирана 18 km од извора и 17 km узводно од града Требиња. Акумулација је запремине $1278 \cdot 10^6 \text{ m}^3$. Тело бране представља бетонска маса од 377000 m^3 у форми 31. инјектираног блока, који су нумерисани с лева на десно (слика 7.8). Ова брана је двострука лучна ($R=185.48 \text{ m}$) брана, висине 123 m, дужине круне од 439.3 m, са дебљином од 4.6 m у врху и 26.9 m у дну бране. Опремљена је мониторинг системом за мерење различитих параметара као што су стање бетона, температуре воде и ваздуха, нивоа воде у акумулацији, хоризонталних и вертикалних померања, ротације тела бране, кретања спојева, напона, притисака, сила узгона, померања темеља и других величина, у укупно 465 тачака.

Три клатна (виска) су инсталирана за мерење радијалних и тангенцијалних деформација. За валидацију и оцену квалитета DEVONNA система одабрано је моделирање радијалног померања тачке P1 на круни бране (надморска висина 403 m) (слика 7.8 а), на блоку бр. 17 (слика 7.8 б). Померање тачке је мерено коришћењем термометра V17-1.

Уобичајено је да мониторинг систем брана, омогућава праћење већег броја излазних величина, које могу бити коришћене за оцену структурног понашања. Иако вештачке неуронске мреже имају способност да истовремено моделирају више излазних величина, због јасноће је, у дисертацији презентован принцип коришћењем само једне излазне променљиве.

Брана је изграђена 1967. год. и до периода у коме је вршена анализа и моделирање је прикупљено око 14500 парова (к-торки) улазно-излазних података о померању тачке P1. Ипак, због проблема велике количине недостајућих података у интервалу од 1967 до 1984. год., за моделирање је одабран период од јануара 1984. до краја августа 2011. год. У циљу побољшања рачунарске ефикасности (процесорског времена потребног за генерисање оптималних модела) је коришћена свака друга k -торка улазно-излазних података (примера за обучавање), тако да је укупан обим скупа података био 5042 узорка. У складу са *cross-validation* стратегијом, ови подаци су дељени у LDS и TDS подскупове у задатим односима [112, 113, 114].



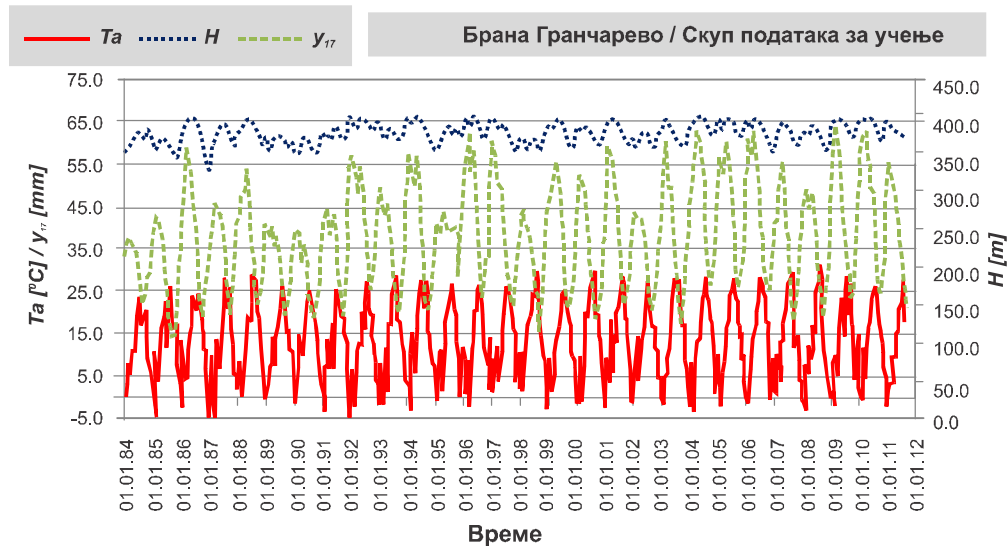
Слика 7.8. Радијално померање тачке P1 (а) и попречни пресек блока 17 (б)

Ефекти хидростатичког притиска на померање бране су експлицитно узети у обзир кроз променљиву H , која изражава ниво воде у акумулацији. Интервал варирања

улазне променљиве H је био од 331.05 до 401.28 m. Термички ефекти су били обухваћени променљивом која представља средњу дневну температуру ваздуха T_a (-7.10 до 32.10 °C). Термички ефекти су, такође, били изражени и тригонометријским функцијама, као циклична температурна оптерећења (оптерећења која имају сезонску природу), које могу описати слику деформација преко улазне променљиве d . Променљива d представља време, мерено у данима, протекло од почетка текуће године до тренутка мерења. Ово значи да је интервал варирања променљиве d био од 1 до 365 дана.

Међутим, на различитим географским локацијама, температурне осцилације могу имати фазни померај у односу на почетак календарске године. У циљу тестирања стабилности модела на поменути температурни померај, коришћене су изведене улазне променљиве $d20$ и $d50$, које означавају фазни померај од двадесет ($d20 = d + 20$) или педесет ($d50 = d + 50$) дана, респективно. Ефекат старења, у функцији времена t , је укључивао деградацију својстава материјала током животног века и експлоатације моделиране структуре. Као што је раније поменуто, радијално померање тачке P1, исказано променљивом y_{17} , је одабрано за излазну променљиву, односно функцију циља која се моделира. Интервал померања у анализираном периоду је био у опсегу од +13.26 до +64.4 mm, при чему је усвојено да се радијално померање у низводном правцу сматра позитивним.

Пример скупа података који су служили за учење (LDS), који репрезентује 60% података од укупно расположивих мерења, је дат на слици 7.9. Као последица хидростатичког притиска, радијална померања посматране тачке се повећавају са повећањем нивоа воде у акумулацији H , и обрнуто. За разлику од тога, радијална померања се смањују са порастом температуре ваздуха T_a , и обрнуто.



Слика 7.9. Скуп података за учење за брану Гранчарево за период 1984-2011.

Као што је већ речено, валидација предложене ANN/GA методологије и развијеног DEVONNA система је реализована на скупу података који су измерени и прикупљени у периоду 1.1.1984 до 29.8.2011. год. Према *cross-validation* стратегији, реализоване су четири серије тестова са различитим односом LDS:TDS (60:40, 70:30, 75: 25 и 80:20), при чему је подскуп примера за учење биран на случајан начин. Добијени резултати су

поређени са измереним вредностима, али и са резултатима који су добијени коришћењем еквивалентног MLR/GA модела презентованог у [6].

Параметри GA за оба хибрида су били: број јединки у популационом базену (број ANN или MLR модела) 70, а број генерација 50. За селекцију је одабрана бинарна турнирска селекција, а за репродукцију укрштање у једној тачки и униформна мутација са вероватноћама од 0.9 и 0.125, респективно.

7.5.1. ANN/GA хибридни параметри и резултати

Структура хромозома који је коришћен за моделирање радијалног померања тачке P1 је формирана у складу са табелом 7.1.

Табела 7.1. Структура бинарног ANN хромозома

Гени	Ознака	Напомена	Опсег	Број бита	Пример кодирања
Број скривених слојева	$N^{(\psi)}$		[1-32]	5	00010 2 скривена слоја
Број неурона у скривеном слоју	$n^{(\psi)}$		[1-64]	6	000111 7 неурона по скривеном слоју
Активациона функција	$AF^{(\psi)}$		[0-7]	3	001- <i>Sin AF</i> 010- <i>Hyperbolic tangent AF</i> ...
Алгоритам учења	$LA^{(\psi)}$			1	0-BPGD алгоритам 1-RPROP алгоритам
Правила учења	$LPI^{(\psi)}$	за RPROP: η^+	[1.0-1.60]	4	0001=1.04 ;... ; 0101=1.20; 1001=1.36;...
	$LPI^{(\psi)}$	за BPGD: η	[0.2-0.95]	4	0001=0.25;... ; 1110=0.90; 1111=0.95
	$LP2^{(\psi)}$	за RPROP: η^-	[0-1]	4	0001=0.15;... ; 1001=0.50;...;1110=0.90;....
	$LP2^{(\psi)}$	за BPGD: μ	[0.4-0.9]	4	..., 1101=0.83; 1110=0.86; 1111=0.90;...

Слика 7.10 приказује пример хромозома који репрезентује ANN јединку са 11 скривених слојева, од којих сваки има по 20 неурона. Мрежа моделира радијално померање тачке P1 помоћу RPROP алгоритма учења и TANH активационом функцијом у неуронима. Параметри LP1 и LP2 имају вредности $\eta^+ = 1.2$ и $\eta^- = 0.5$.

ANN ^(ψ)													
$N^{(\psi)}$ гени				$n^{(\psi)}$ гени				$AF^{(\psi)}$ гени		$LA^{(\psi)}$ гени	$LPI^{(\psi)}$ гени		$LP2^{(\psi)}$ гени
0	1	0	1	1	0	1	0	1	0	0	0	1	0
N=11				n=20				TANH		RPROP	$\eta^+=1.2$		$\eta^-=0.5$

Слика 7.10. Пример хромозома који репрезентује ANN јединку

Иницијалне вредности тежинских коефицијената $\mathbf{W}_0$ бирани су на случајан начин из интервала $[-0.5, 0.5]$. Да би се умањио утицај стохастичког избора матрице $\mathbf{W}_0$, процес учења је за сваку ANN јединку понављан пет пута ($R = 5$) (видети слику 7.5).

Моделирање радијалног померања тачке P1 је реализовано за сваки LDS-TDS пар и различите скупове предикторских променљивих, а резултати анализе су дати у табели 7.2.

Табела 7.2. Квалитет ANN/GA модела изражен преко $RMSE_{Train}$ и $RMSE_{Test}$

ANN/GA	$RMSE_{Train}$ [mm]				$RMSE_{Test}$ [mm]			
<i>cross-valid.</i>	предиктори				предиктори			
LDS:TDS однос	H, T_a t	H, T_a t, d	H, T_a $t, d20$	H, T_a $t, d50$	H, T_a t	H, T_a t, d	H, T_a $t, d20$	H, T_a $t, d50$
60-40 (3031:2021)	4.3701	2.1384	1.8815	1.5252	4.6598	2.1044	1.8854	1.5845
70-30 (3536:1516)	4.3745	2.1492	2.3184	3.2118	4.3759	2.1044	2.2654	3.2673
75-25 (3789:1263)	4.2113	2.0638	1.8562	5.8963	4.5877	2.0655	1.8647	5.8434
80-20 (4042:1010)	4.5001	1.8538	1.7574	1.6811	4.8433	2.3182	1.9145	1.6739

Због комплексности вештачких неуронских мрежа просечно трајање процеса генерисања модела варијало је од 1.5 до 5 h. Модели са улазним променљивим H , T_a и t показују стабилно понашање за све унапред дефинисане односе LDS:TDS. Вредност $RMSE_{Test}$ за ове случајеве је била приближно 4.5 mm (9% од опсега померања тачке P1).

Након укључивања променљиве d (дани у години), квалитет добијеног модела значајно је унапређен, тако да $RMSE_{Test}$ за скуп предикторских променљивих H , T_a , t и d има вредност од приближно 2 mm што представља износ од око 4% опсега радијалног померања тачке P1. Међутим, у тестовима са *dummy* променљивим $d20$ и $d50$, је уочено да се јавља одређена нестабилност у квалитету модела за различите LDS:TDS односе.

У више од 90% тестова, неуронске мреже са најбољим карактеристикама су имале TANH активациону функцију. Модел вештачке неуронске мреже највеће тачности ($RMSE_{Test} = 2.0655$ mm) је добијен за предикторе H, T_a, t и d са 75:25 LDS:TDS односом. Оптимизовани мрежни модел има 5 скривених слојева, са по 30 неурона у сваком слоју. Активациона функција је била TANH, док је за алгоритам учења за ову мрежу, систем определио RPROP алгоритам са параметрима $\eta^+ = 1.15$ и $\eta^- = 0.65$. Оптимизациона процедура је трајала око 135 минута.

7.5.2. Параметри и резултати хибрида MLR/GA

Да би се модели брана ANN/GA и MLR/GA учинили упоредивим, параметар комплексности модела вишеструке линеарне регресије у MLR/GA хибриду [6] је подешен на $\lambda = 0$. Креирана су два базена потенцијалних регресора (ρ_I и ρ_{II}) који описују радијално померање тачке P1, проузроковано хидростатичким притиском (H), температурним варијацијама (T_a), и старењем (t). Слично као и код ANN модела, променљива d је уведена у процес испитивања да би се моделирали термички ефекти, односно сезонска термичка оптерећења (табела 7.3).

Табела 7.3. Регресори који су коришћени за моделирање радијалног померања тачке P1 (y_{17})

ρ^*	Регресори		
	Ниво воде	Температуре	Старење
ρ_I	H, H^2, H^3	$T_a, T_a^2, T_a^3, T_a^4, T_a^5$	$t, t^2, \sqrt{t}$
	$\sqrt{H}$	$e^{-T_a}, e^{-T_a/2}$	$e^{-t}, e^{-t/2}, e^{-t/4}$
ρ_{II}	H, H^2, H^3	$T_a, T_a^2, T_a^3, T_a^4, T_a^5$	$\sin(2k \cdot \pi \cdot d / 365), k = 1, 2, 4$
	$\sqrt{H}$	$e^{-T_a}, e^{-T_a/2}$	$\cos(2k \cdot \pi \cdot d / 365), k = 1, 2, 4$
*Базен регресора			

У базену ρ_I је разматран скуп од 17 потенцијалних регресора. Регресори су обухватали све расположиве променљиве у различитим формама осим тригонометријских облика. Популациони базен ρ_{II} је имао 25 регресора, који обухватају све променљиве у одабраним формама, укључујући и тригонометријске форме променљиве d . У циљу тестирања отпорности на температурни померај, су реализоване додатне анализе, где је улазна променљива d мењана са *dumty* променљивим $d20$ и $d50$.

Резултати добијени коришћењем MLR/GA хибрида за *cross-validation* тестове су дати у табели 7.4.

Табела 7.4. Квалитет MLR модела изражен преко $RMSE_{Test}$

MLR/GA	$RMSE_{Train}$ [mm]				$RMSE_{Test}$ [mm]			
cross-validation	предиктори				предиктори			
LDS:TDS однос величина скупова	H, T_a	H, T_a	H, T_a	H, T_a	H, T_a	H, T_a	H, T_a	H, T_a
	t	t, d	$t, d20$	$t, d50$	t	t, d	$t, d20$	$t, d50$
60-40 (3031:2021)	5.0362	2.0147	2.0125	2.0138	5.0891	2.0144	2.0120	2.0134
70-30 (3536:1516)	5.1048	2.1378	2.1375	2.1374	4.9386	2.0750	2.0735	2.0748
75-25 (3789:1263)	5.0420	2.1504	2.1524	2.1539	5.0947	2.0170	2.0193	2.0237
80-20 (4042:1010)	5.0556	2.0514	2.0520	2.0513	5.1230	2.4311	2.4339	2.4370

У случајевима када у популационом базену није било регресора у тригонометријском облику (ρ_I скуп регресора), грешке модела су биле приближно око 5 mm (10% од опсега радијалног померања тачке P1). Након укључивања променљиве d и тригонометријских регресора, износ грешке $RMSE_{Test}$ је био редукован на приближно 2 mm, што представља износ од око 4% померања тачке P1. Коришћење *dumty* променљивих $d20$ и $d50$ уместо променљиве d није имало значајан утицај на износ грешке.

Вредност грешке $RMSE_{Test}$ за све *cross-validation* тестове је била приближно око 2 mm, осим за однос LDS:TDS=80:20, где је $RMSE_{Test}$ грешка била нешто већа, односно око 2.43 mm.

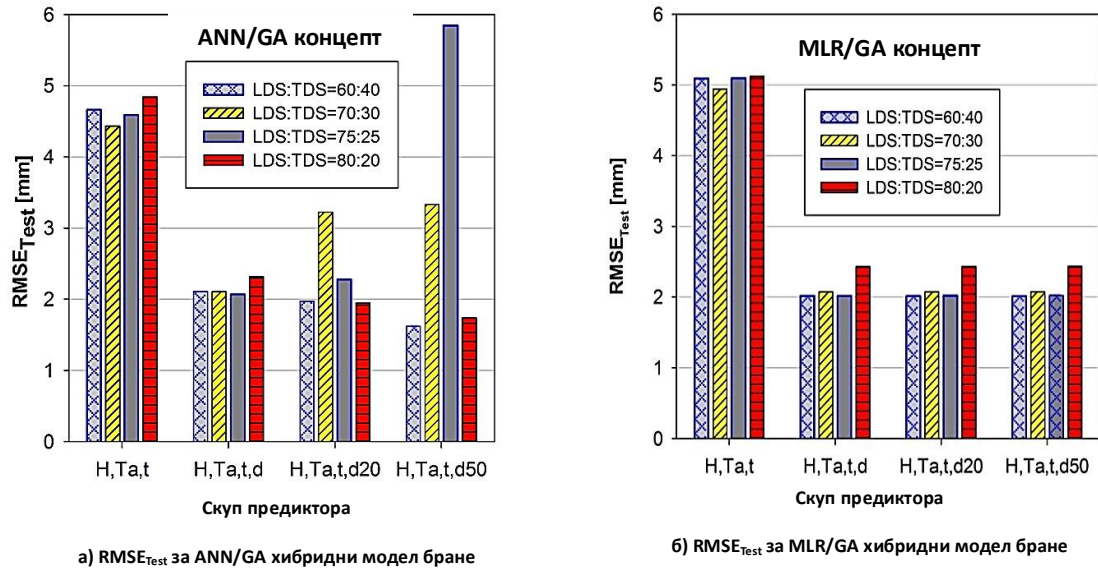
Најбољи од свих оптимизованих модела је био модел са грешком $RMSE_{Test} = 2.0144$ mm, који је добијен за однос LDS:TDS=60:40, скуп предиктора $\{H, T_a, t, d\}$ и 22 регресора из популационог базена ρ_{II} (7.8):

$$\begin{aligned}
y_{17} = & (7.906e+05) + (5.974e+03) \cdot H - (5.482e+00) \cdot H^2 + (0.020e-03) \cdot H^3 \\
& - (1.215e+05) \cdot \sqrt{H} - (1.591e-01) \cdot T_a - (1.682e-02) \cdot T_a^2 + (2.623e-03) \cdot T_a^3 \\
& - (1.360e-04) \cdot T_a^4 + (2.148e-06) \cdot T_a^5 + (3.656e-03) \cdot (e^{-T_a}) \\
& + (8.787e+00) \cdot \sin(2 \cdot \pi \cdot d / 365) + (8.795e-02) \cdot \sin(4 \cdot \pi \cdot d / 365) \\
& + (4.055e+00) \cdot \cos(2 \cdot \pi \cdot d / 365) + (8.806e-01) \cdot \cos(4 \cdot \pi \cdot d / 365) \\
& + (8.543e-02) \cdot \cos(6 \cdot \pi \cdot d / 365) + (1.186e-01) \cdot \cos(8 \cdot \pi \cdot d / 365) \\
& + (6.263e+03) \cdot t - (5.263e+01) \cdot t^2 - (3.819e+04) \cdot \sqrt{t} + \\
& - (5.435e+04) \cdot e^{-t} + (1.334e+04) \cdot e^{-t/2} - (4.239e+04) \cdot e^{-t/4}
\end{aligned} \tag{7.8}$$

Оптимизација MLR модела је трајала у просеку 5 до 10 минута.

7.5.3. Дискусија

Слика 7.11 приказује збирне резултате реализованих тестова помоћу ANN/GA и MLR/GA хибрида, где је квалитет модела изражен са $RMSE_{Test}$.



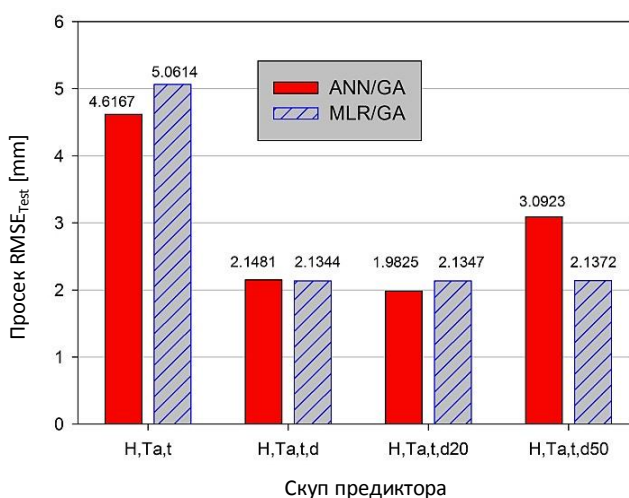
Слика 7.11. Вредности $RMSE_{Test}$ грешака за различите скупове предиктора и LDS:TDS односа код ANN и MLR модела

Са слике се може видети да, у анализама у којима су променљиве H , T_a и t коришћене као улазне променљиве, оба хибрида показују стабилно понашање за све унакрсне тестове. Вредност $RMSE_{Test}$ грешке за ANN моделе је била приближно 4.5 mm, што је око 10% мање у поређењу са величином грешке $RMSE_{Test}$ коју генерише MLR/GA хибрид, а која је износила око 5 mm.

Квалитет оба хибрида је значајно побољшан када је у систем уведена и улазна променљива d , тако да је вредност за $RMSE_{Test}$ грешку редукована на приближно 2 mm. Унапређење квалитета ANN и MLR модела који користе дане у години као улазну променљиву се може објаснити способношћу променљиве d да боље репрезентује циклична температурна оптерећења.

Тестови у којима је променљива d замењена *dummy* променљивим $d20$ и $d50$ показују да су ANN модели нестабилни када је присутан температурни померај у односу на почетак године. Нестабилност ANN модела је делимично уочљива у ситуацијама када скуп предиктора садржи променљиву $d50$ (слика 7.11 а), где $RMSE_{Test}$ вредности варирају од 1.6 mm до 5.84 mm. За разлику од ANN модела, MLR модели су се показали веома стабилним у тестовима са *dummy* променљивим $d20$ и $d50$ (слика 7.11 б). Стабилност MLR модела се може објаснити способношћу тригонометријских регресора да моделирају фазни померај једноставним суперпонирањем синусних и косинусних функција.

Слика 7.12 приказује просечне вредности грешке $RMSE_{Test}$ за ANN/GA и MLR/GA модел и то за сваки од предикторских подскупова. Са слике се, такође, може видети да је просечна грешка у тестовима са *dummy* променљивом $d50$ значајно лошија када је за моделирање коришћен ANN/GA хибрид.



Слика 7.12. Просечне вредности $RMSE_{Test}$ за различите подскупове предиктора

Са аспекта потребних рачунарских ресурса, просечно време потребно за генерисање оптимизованог ANN модела је драстично веће него време потребно за генерисање еквивалентног MLR модела (1.5 до 5h за ANN/GA у поређењу са 5-10 минута код MLR/GA хибрида).

Закључци о ANN/GA хибриду и његовој способности генерисања вештачких неуронских мрежа оптимизоване топологије, активационих функција, алгоритама и правила учења, као и закључци који следе као резултат компарације са перформансама MLR/GA хибрида односно GenReg система, су дати и у поглављу 10.

8. Компарација MLR/GA и ANN/GA регресионих модела са моделима у форми постепених регресија

У поступцима креирања регресионих модела објеката и процеса, истраживачи теже креирању једноставнијих модела, не само због филозофских (регуларизација, Оскат-ова оштрица), већ и из практичних разлога. Једноставније моделе је лакше поново тестирати у понављањима и *cross-validation* студијама. Једноставнији модели за контролу и предикцију су економичнији у пракси, лакши за разумевање и оцену, и чак поседују лепоту коју њихови компликованији пандани немају [8]. Код вишеструких линеарних регресионих модела, број могућих подмодела, са повећањем броја предикторских променљивих, рапидно расте. Једна од метода којом се екстензивни MLR модели могу упростити, је примена концепта постепених регресија¹. Постепене регресије представљају уобичајену стратегију у креирању статистичких модела различитих намена.

У овом делу дисертације су дате основе постепених регресија и приказано је поређење регресионих математичких модела структурног понашања бетонских брана, креираних помоћу MLR/GA и ANN/GA софтверских хибрида, са моделима заснованим на постепеним регресијама. Компаративна анализа је реализована на студији случаја бране Бочац, на реци Врбас у Републици Српској.

8.1. Постепене регресије. Теоријске основе

Модели у форми постепених регресија настају процесом додавања или уклањања променљивих и регресора, на бази t статистика њихових процењених коефицијената. Правилно коришћене, ове регресионе математичке форме, дају истраживачу додатну снагу за ефикасно описивање сложених објеката и система у односу на класичне MLR. Оне су посебно користне за пажљиви избор предиктора из обимног скупа потенцијалних улазних променљивих и/или за фино подешавање модела, укључивањем или искључивањем променљивих из регресионих једначина. Неправилно коришћене, оне могу водити ка лошим моделима и смањењу поверења у моћ регресионих парадигми, генерално [8].

Постепене регресије су есенцијално базиране на *семипарцијалној корелацији*². Семипарцијална корелација се изражава преко *коефицијента семипарцијалне корелације* (sr) и квадрата овог коефицијента (sr^2). Она обезбеђује додатно тумачење релативне важности независних променљивих у одређивању зависне променљиве.

¹ Постепене регресије - *stepwise regression*

² Уместо синтагме *semipartial correlation*, у статистичком пакету SPSS, као и у различитим литературним изворима, се користи синтагма *part correlation*, што може стварати конфузију

Семипарцијална корелација показује колико свака променљива уникатно доприноси коефицијенту детерминације R^2 . Другим речима, квадрат коефицијента семипарцијалне корелације неке променљиве, указује колико ће се смањити вредност R^2 уколико се та променљива уклони из регресионе једначине.

Нека је

χ - скуп свих независних променљивих $\mathbf{X}$, а

ψ_k - скуп свих независних променљивих $\mathbf{X}$ изузев x_k ,

онда је квадрат коефицијента семипарцијалне корелације:

$$sr_k^2 = R_\chi^2 - R_{\psi_k}^2 \quad (8.1)$$

Дакле, да би се добио уникатни допринос предиктора x_k коефицијенту детерминације регресионог модела, R^2 , прво је потребно регресирати зависну променљиву y на бази свих независних променљивих (χ), а потом регресирати зависну променљиву y на бази свих променљивих изузев x_k (ψ_k). Разлика добијених вредности за R^2 , представља квадрат коефицијента семипарцијалне корелације [7].

Форма исказивања коефицијента семипарцијалне корелације може бити различита, а у оквиру ове дисертације је коришћен облик:

$$sr_k = \frac{t_k \cdot \sqrt{1 - R_\chi^2}}{\sqrt{residualDF}} \quad (8.2)$$

где је t_k , *Student*-ова t статистика за k -ти регресор у вишеструком линеарном моделу, $residualDF = N - K - 1$, је број степени слободе резидуалне суме, N је број мерења, а K број регресора.

Када се нека променљива додаје или уклања из регресионе једначине, све вредности семипарцијалних коефицијената се мењају. Дакле, семипарцијална корелација говори о промени R^2 , за једну променљиву у једној трансформацији регресионог модела.

Детаљнија анализа вишеструке корелације је дата у [7].

8.2. Етапе у креирању постепених регресија

Техника креирања постепених регресионих модела обухвата следеће етапе:

1. Идентификација иницијалног модела

Иницијални модел је свеобухватан модел који укључује све доступне, мерљиве утицаје феномена који се истражује. Иницијални модели могу имати различите форме као што су: *Multiple Regression*, *Factorial Regression*, *Polynomial Regression*, *Response Surface Regression* или неки други облик. Као пример наводи се *Quadratic response surface* регресија која је дизајнирана као хибрид који садржи све ефекте, односно регресоре као и полиномијална регресија реда 2, са додатим регресорима који обухватају међудејства другог реда.

2. Итеративни кораци постепене процедуре

Ово су кораци у којим се алтернативно регресији додају променљиве, или се из регресионе једначине уклањају променљиве у складу са критеријумом додавања или

уклањања регресора. Током изградње модела истраживачи често користе неку од следећих процедура: *The Forward Entry Method*, *The Backward Removal Method*, *The Forward Stepwise Method*, *The Backward Stepwise Method*.

У *forward stepwise* регресији, променљива која обезбеђује највеће увећање за R^2 , односно променљива која би требало да има највећи семипарцијални коефицијент, се додаје у регресиону једначину, обезбеђујући тако статистичку значајност.

Код *backward stepwise* регресије, променљива која производи најмањи пад код R^2 , односно променљива са најмањим семипарцијалним коефицијентом, се уклања из регресије јер је статистички најмање значајна.

3. Прекид итеративне процедуре

Итеративна процедура се прекида када је испуњен дефинисани критеријум или је остварен задати максимални број итеративних корака. Постоје различити критеријуми за рангирање добијених регресионих подмодела са аспекта добротe односно квалитета³. Најчешћи критеријуми су: коефицијент вишеструке детерминације - R^2 , кориговани коефицијент вишеструке детерминације - R^2_{adj} и Mallows's C_p статистика.

8.3. Примена постепених регресија у хидро-информатици

Без обзира што се концепт моделирања помоћу постепених регресија може сматрати класичним, његова примена у математичком моделирању процеса везаних за структурно понашање бетонских брана и процеса који имају додирне тачке са доменом хидроинформатике, и даље је актуелна. Бројни су радови у наведеној области од којих се, као пример, наводи неколико. Fang Nu-Fang и коаутори у раду [115] користе постепене регресије за моделирање процеса таложења наноса у хидролошкој анализи поплава и воденог тока у сливу реке *Wangjiaqiao* (Кина) у области *Three Gorges Area*. У раду [116], Tsung-Ren Peng и Chung-Ho Wang користе постепене регресије за анализу извора и узрока процуривања земљане акумулације *Xin-Shan* на рукавцу реке *Keelung* на северу Тајвана. Због ограничења, која се огледају у недовољној стабилности у односу на скуп предикторских променљивих, у последњој декади често се са постепеним регресијама комбинују и користе и хибридне AI технике. У раду [117], Xi Guang-yong и коаутори дају компаративну анализу математичких модела узводно-низводног померања тачака бетонских брана. Хибридни регресиони модели су оптимизовани помоћу AI хеуристике вештачких имуних алгоритама (AIA) и поређени са моделима у форми постепених регресија. Квалитет добијених модела изражаван је преко коефицијената корелације и стандардне грешке регресије (S). Показало се да су обе стратегије моделирања генерисале моделе упоредиве тачности, уз благу предност AIA модела по питању тачности, али и уз мању ефективност са аспекта времена потребног за генерисање модела.

³ Као и код техника описаних у претходним деловима дисертације, за меру квалитета и код постепених регресија користе се термини *fitness* или *goodness*.

8.4. Поређење модела генерисаних помоћу MLR/GA и ANN/GA хибрида са моделима у облику постепених регресија

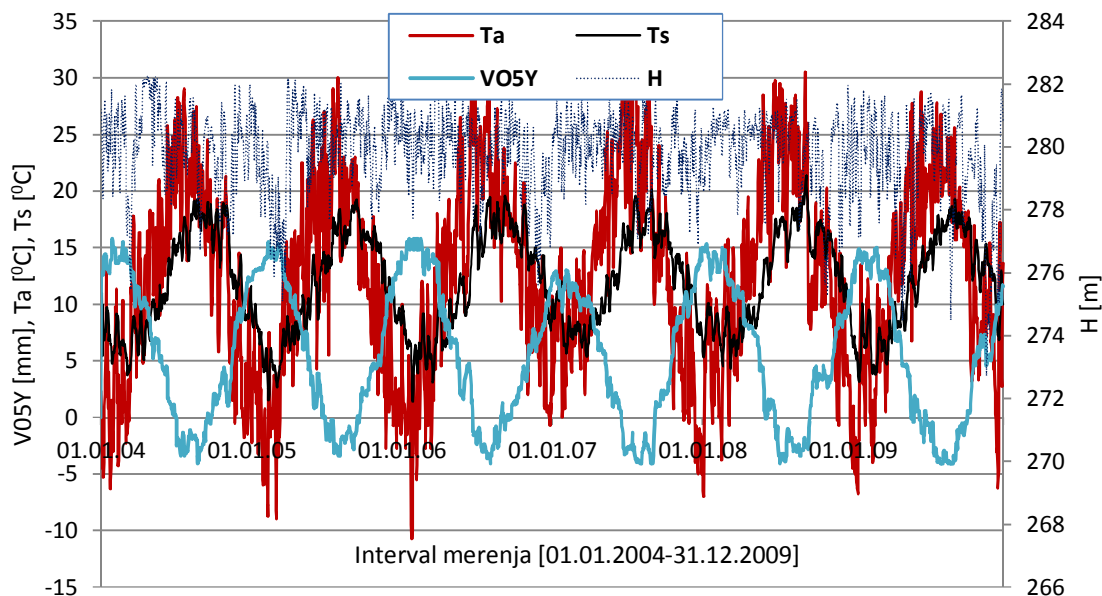
8.4.1. Студија случаја: Брана Бочац

За поређење модела креираних помоћу GenReg и DEVONNA система, односно MLR/GA и ANN/GA софтверских хибрида са моделима заснованим на постепеним регресијама, у оквиру студије случаја бране Бочац на реци Врбас у Републици Српској, је реализовано моделирање радијалног померања тачке P1 на блоку бр.5., (поглавље 6, слика 6.5 и слика 6.6 а, б и ц).

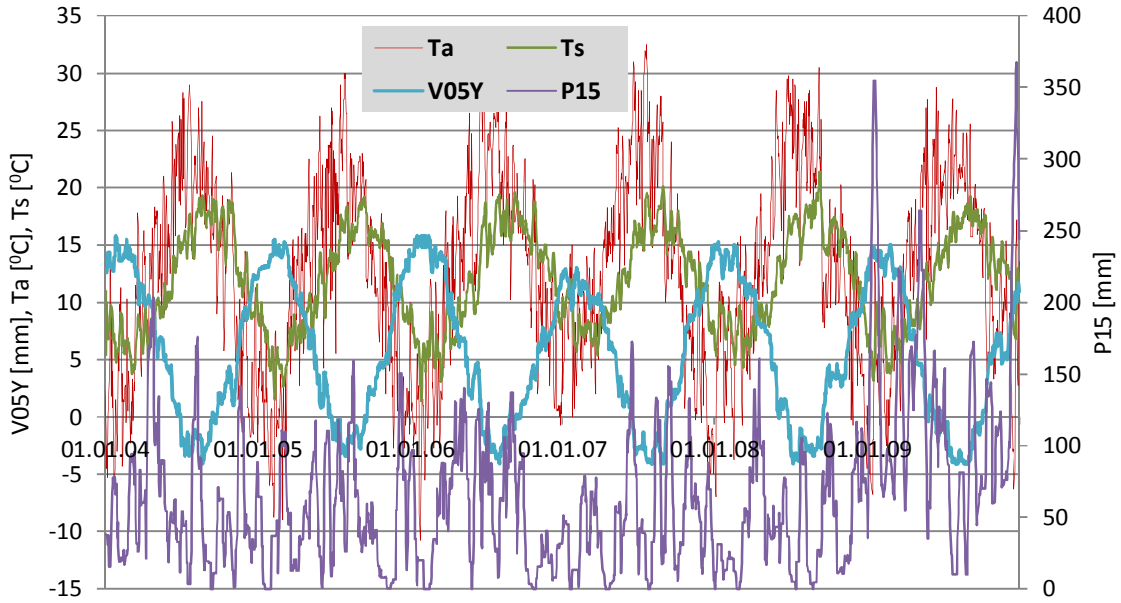
Део базе података је обухватао интервал мерења од 1. јануара 2004. год. до 31. децембра 2009. год. Скуп података је обухватао 2191 к-торку вектора улазних података. Кота горње воде, H , је варијала у интервалу [272.71, 282.24 m]. Средња температура ваздуха, T_a , и температура бетона у телу бране, T_s , су се кретале у интервалима [-10.75, 32.5 °C] и [1.21, 21.11 °C], респективно. Вредности променљиве P , која представља 15-то дневне падавине, су варијале од 0 до 367.6 mm. Ефекат старења материјала је презентован променљивом t , која мери време од почетка експлоатације бране, изражено у данима [1461, 3651]. Све улазне променљиве су нормализоване на интервал од 0 до 1. Интервал промене радијалног померања тачке P1, односно зависне променљиве је био од -4.1 до 15.8 mm. Елементи дескриптивне статистике података из базе су дати у табели 8.1, а приказ улазно излазних променљивих на сликама 8.1 и 8.2.

Табела 8.1. Елементи дескриптивне статистике улазно/излазних променљивих бране Бочац за период 1. 1.2004-31.12.2009. год.

	H [m]	T_a [°C]	T_s [°C]	P [mm]	$V05.1Y$ [mm]
мин.	272.71	-10.75	1.22	0	-4.1
макс.	282.24	32.50	21.12	367.6	15.8
просек	279.83	12.59	11.89	64.3	5.73
станд.дев.	1.358	8.61	4.29	55.7	6.11



Слика 8.1. Улазно/излазни подаци бране Бочац за период 1.1.2004-31.12.2009. године



Слика 8.2. Улазно/Излазни подаци бране Бочац за период 1.1.2004-31.12.2009.год.

У складу са *cross-validation* стратегијом [112-114], подаци из базе су дељени у LDS и TDS подскупове у односима који су приказани у табели 8.2. Избор података у скупове за учење и тестирање је реализован стохастички.

Табела 8.2. Дефинисани односи величина скупова података за учење и тестирање

Тест	LDS:TDS	LDS:TDS број n -торки
1. CrossVal 1	50: 50	1096: 195
2. CrossVal 2	60: 40	1315: 876
3. CrossVal 3	70: 30	1534: 657
4. CrossVal 4	75: 25	1644: 547
5. CrossVal 5	80: 20	1743: 438

8.4.2. Резултати моделирања помоћу MLR/GA и ANN/GA хибрида

На основу претходног, за моделирање помоћу MLR/GA хибрида, су креирана три скупа регресора: ρ_I , ρ_{II} и ρ_{III} . Скуп регресора ρ_I (8.3), не садржи температуру бетона, већ средњу температуру ваздуха T_A .

$$\rho_I : \{H, H^2, H^3, \sqrt{H}, P, P^2, P^3, T_A, T_A^2, T_A^3, T_A^4, T_A^5, e^{-T_A}, e^{-\frac{T_A}{2}}, \sqrt{T_A}, t, t^2, \sqrt{t}, e^{-t}, e^{-\frac{t}{2}}, e^{-\frac{t}{4}}\} \quad (8.3)$$

Скуп регресора ρ_{II} (8.4) обухвата и температуру бетона T_S

$$\rho_{II} : \{H, H^2, H^3, \sqrt{H}, P, P^2, P^3, T_A, T_A^2, T_A^3, T_A^4, T_A^5, e^{-T_A}, e^{-T_A/2}, \sqrt{T_A}, T_S, T_S^2, T_S^3, T_S^4, T_S^5, e^{-T_S}, \sqrt{T_S}, t, t^2, \sqrt{t}, e^{-t}, e^{-\frac{t}{2}}, e^{-\frac{t}{4}}\} \quad (8.4)$$

док је скуп регресора ρ_{III} (8.5), контејнер у којем су уместо температуре структуре T_s , за узимање утицаја термичких оптерећења, коришћени дани у години d у облику тригонометријских форми:

$$\begin{aligned} \rho_{III} : \{ & H, H^2, H^3, \sqrt{H}, P, P^2, P^3, T_A, T_A^2, T_A^3, T_A^4, T_A^5, e^{-T_A}, e^{-T_A/2}, \sqrt{T_A}, \\ & \sin(2 \cdot \pi \cdot d), \cos(2 \cdot \pi \cdot d), \sin(4 \cdot \pi \cdot d), \cos(4 \cdot \pi \cdot d), \sin(6 \cdot \pi \cdot d), \\ & \cos(6 \cdot \pi \cdot d), \sin(8 \cdot \pi \cdot d), \cos(8 \cdot \pi \cdot d), t, t^2, \sqrt{t}, e^{-t}, e^{-t/2}, e^{-t/4} \} \end{aligned} \quad (8.5)$$

У првој фази моделирања радијалног померања тачке P1 помоћу MLR/GA хибрида, су генерисани регресиони модели за скуп регресора ρ_I , који са температурног аспекта, садржи регресоре креиране над променљивом T_A . Како T_A променљива има изражену варијансу, квалитет модела је био релативно мали за све пробе ($R_{adj}^2 < 0.7$).

Даљи поступак моделирања је реализован над скуповима регресора ρ_{II} и ρ_{III} . У одређеном броју случајева корен средње квадратне грешке над скупом за тестирање је био мањи него над скупом за учење. Овакав однос величина поменутих грешака није очекиван. Да би се разјасниле недоумице око величине и односа грешака $RMSE_{Learn}$ и $RMSE_{Test}$, урађено је огледно моделирање за однос величина скупова LDS:TDS=50:50 и LDS:TDS=60:40, и то за регресоре из контејнера ρ_{II} . Сви прорачуни су поновљени три пута, да би се додатно анализирао и ефекат стохастичне природе GA, како на коначни избор регресора, тако и на величину грешака. Резултати су дати у табели 8.3.

Табела 8.3. Утицај GA стохастике и провера односа $RMSE_{Learn}$ и $RMSE_{Test}$

Однос LDS:TDS	Проба	Регресори добитијених модела	$RMSE_{Learn}$ [mm]	$RMSE_{Test}$ [mm]	R^2	R_{adj}^2
50:50	а	$\{H, H^2, H^3, P, P^2, T_A^4, T_A^5, e^{-T_A}, e^{-T_A/2}, \sqrt{T_A},$ $T_S, T_S^2, T_S^3, T_S^4, T_S^5, t^2, \sqrt{t}, e^{-t/4}\}$	1.9380	1.7793	0.8976	0.8958
	б	$\{H, H^2, H^3, P, P^2, T_A, T_A^2, T_A^5, e^{-T_A}, e^{-T_A/2},$ $T_S^2, T_S^3, T_S^4, T_S^5, e^{-T_S}, \sqrt{T_S}, \sqrt{t}, e^{-t/2}, e^{-t/4}\}$	1.9351	1.7838	0.8979	0.8961
	ц	$\{H, H^2, H^3, P, P^2, P^3, T_A, T_A^4, e^{-T_A}, \sqrt{T_A},$ $T_S^2, T_S^3, T_S^4, e^{-T_S}, \sqrt{T_S}, t^2, \sqrt{t}, e^{-t/4}\}$	1.9375	1.7894	0.8977	0.8959
60:40	а	$\{H^2, H^3, P, P^2, P^3, T_A^2, T_A^3, T_A^4, e^{-T_A}, \sqrt{T_A},$ $T_S, T_S^2, T_S^3, T_S^4, T_S^5, e^{-T_S}, t, t^2, \sqrt{t}, e^{-t}, e^{-t/2}\}$	1.7735	1.9893	0.9158	0.9140
	б	$\{H, H^2, P, P^2, T_A^2, T_A^3, T_A^5, e^{-T_A/2}, \sqrt{T_A},$ $T_S, T_S^3, T_S^4, T_S^5, e^{-T_S}, \sqrt{T_S}, t, \sqrt{t}, e^{-t}, e^{-t/2}, e^{-t/4}\}$	1.7748	1.9868	0.9152	0.9139
	ц	$\{H^2, P, P^2, T_A, T_A^2, T_A^5, e^{-T_A}, e^{-T_A/2}, T_S, T_S^3,$ $T_S^4, t, \sqrt{t}, e^{-t}, e^{-t/2}, e^{-t/4}\}$	1.7796	1.9814	0.9147	0.9137

Добијени резултати указују да је за однос LDS:TDS=50:50, $RMSE_{Learn} > RMSE_{Test}$, док је код односа LDS:TDS=60:40, $RMSE_{Learn} < RMSE_{Test}$, те се може закључити да величина грешке $RMSE_{Learn}$ и $RMSE_{Test}$, зависи од обима и од стохастике података у скупу за учење и скупу за тестирање. Другим речима, за овакве природне феномене и велики обим података, не мора се јављати уобичајени и очекивани однос по коме је $RMSE_{Learn} < RMSE_{Test}$. Поновљени обрачуни (а, б, ц), за константан однос величина скупова за учење и тестирање дају приближно исте резултате, односно параметре модела, без обзира што се структура, број и облик регресора незнатно мења. Може се закључити да унутрашњи механизми GA стохастике не доводе до значајних варирања квалитета модела.

За моделе генерисане помоћу MLR/GA хибрида, анализиран је и утицај облика регресора на тачност и квалитет добијених модела. За контејнер регресора ρ_{III} и однос LDS:TDS=75:25, реализовани су тестови чији су резултати приказани у табели 8.4.

Табела 8.4. Утицај облика регресора на квалитет
регресионих модела-MLR/GA хибрида

Однос LDS:TDS	Проба	Регресори добијених модела	RMSE _{Learn} [mm]	RMSE _{Test} [mm]	R ²	R ² _{adj}
75:25	а	$\{H^2, H^3, P, P^2, T_a^2, T_a^3, T_a^4, \sin(2 \cdot \pi \cdot d), \cos(2 \cdot \pi \cdot d), \sin(4 \cdot \pi \cdot d), \cos(4 \cdot \pi \cdot d), \sin(6 \cdot \pi \cdot d), \cos(6 \cdot \pi \cdot d), t^2, e^{-t/2}, e^{-t/4}\}$	1.1753	1.1002	0.9629	0.9626
	б	$\{H^2, \sqrt{H}, P, P^2, T_a^2, T_a^3, T_a^4, \sin(2 \cdot \pi \cdot d), \cos(2 \cdot \pi \cdot d), \sin(4 \cdot \pi \cdot d), \cos(4 \cdot \pi \cdot d), \sin(6 \cdot \pi \cdot d), \cos(6 \cdot \pi \cdot d), \sin(8 \cdot \pi \cdot d), t^2, e^{-t/2}, e^{-t/4}\}$	1.1759	1.0997	0.9629	0.9625
	ц	$\{H^2, \sqrt{H}, P, P^2, T_a^2, T_a^3, e^{-T_a/2}, \sin(2 \cdot \pi \cdot d), \cos(2 \cdot \pi \cdot d), \sin(4 \cdot \pi \cdot d), \cos(4 \cdot \pi \cdot d), \sin(6 \cdot \pi \cdot d), t^2, e^{-t}, e^{-t/2}, e^{-t/4}\}$	1.1747	1.0995	0.9630	0.9626

Увођење тригонометријских регресора у контејнер базисних функција је довело до значајног побољшања квалитета модела. Средња вредност корена средње квадратне грешке се побољшала са $\overline{RMSE}_{Test} = 1.885 \text{ mm}$ на $\overline{RMSE}_{Test} = 1.1065 \text{ mm}$, што представља 5.56% од распона варирања радијалног померања тачке P1. Кориговани коефицијент вишеструке детерминације се повећао оквирно са $R_{adj}^2 = 0.902$ на $R_{adj}^2 = 0.9626$.

У циљу добијања што потпуније слике о перформансама развијених MLR/GA и ANN/GA софтверских агената, за скуп регресора ρ_{II} су реализоване комплетне процедуре моделирања радијалног померања тачке P1, за све дефинисане односе LDS:TDS из табеле 8.2. Резултати моделирања су дати у табели 8.5 и табели 8.6.

Табела 8.5. Модели креирани MLR/GA хибридом за скуп регресора ρ_{II}

Скуп података	RMSE _{Test} [mm]	RMSE _{Train} [mm]	Train R2
CrossVal 1 50:50	1.7800	1.9301	0.9101
CrossVal 2 60:40	1.8802	1.7767	0.9213
CrossVal 3 70:30	1.9234	1.8502	0.9120
CrossVal 4 75:25	1.7905	1.8712	0.9127
CrossVal 5 80:20	1.8219	1.8624	0.9123

Табела 8.6. Модели креирани ANN/GA хибридом за скуп регресора ρ_{II}

Скуп података	Број скривених слојева	Број неурона по слоју	RMSE _{Test} [mm]	RMSE _{Train} [mm]	Train R2
CrossVal 1 50:50	4	28	1.7619	1.3978	0.9512
CrossVal 2 60:40	3	23	1.6956	1.4234	0.9502
CrossVal 3 70:30	6	28	1.7173	1.4854	0.9412
CrossVal 4 75:25	6	29	1.6665	1.6545	0.9263
CrossVal 5 80:20	3	25	1.6500	1.6482	0.9266

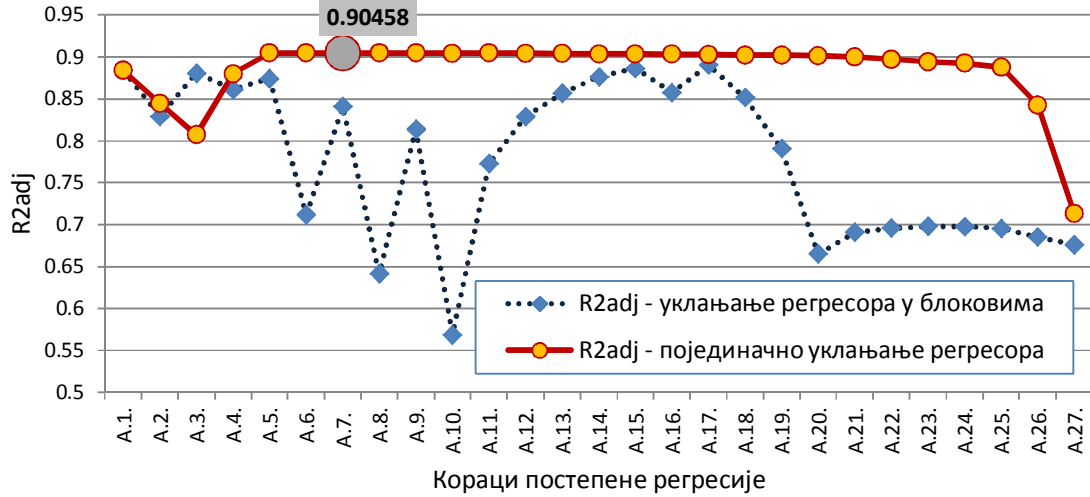
Коефицијент вишеструке детерминације је служио као мера за квалитет и поређење у процедурама генерисања регресионих модела хибрида MLR/GA и ANN/GA. За коначну оцену квалитета добијених модела је коришћен корен средње квадратне грешке за податке из тестног скупа, $RMSE_{Test}$. За скуп регресора ρ_{II} , просечан квалитет генерисаних модела у облику вишеструких регресија је био $\overline{RMSE}_{Test} = 1.8392$ mm, што представља 9.24% распона узводно-низводног померања тачке P1. Код модела у облику оптимизованих неуронских мрежа, просечна вредност корена средње квадратне грешке је износила $\overline{RMSE}_{Test} = 1.6983$ mm, што чини 8.53% распона радијалног померања посматране тачке.

Као што се види, по питању тачности исказане преко $RMSE_{Test}$, ANN генерисани модели су у благој предности, за око 8%, у односу на моделе у облику вишеструких линеарних регресија. Са друге стране, за генерисање MLR модела је потребно око десет пута мање времена (5-7 min) у односу на време (50-80 min) за које ANN/GA хибрид реализује оптимизацију мрежне топологије, алгоритама учења, активационих функција и правила учења.

8.4.3. Резултати моделирања постепеним регресијама

Сви тестови и моделирања радијалног померања тачке P1, бране Бочац коришћењем постепених регресија су реализована коришћењем софтверског пакета SPSS ver.16. Квалитет генерисаних модела у форми постепених регресија за сваки од тестова је изражаван коригованим коефицијентом вишеструке детерминације, рачунатим над скупом података за учење и/или кореном средње квадратне грешке, која је рачуната за тест скуп података.

Као и код моделирања на бази MLR/GA хибрида, за креирање квалитетнијих регресионих модела структурног понашања бетонских брана помоћу постепених регресија, је било неопходно у моделе уградити термичке ефекте у облику температуре структуре или дана у години. Из наведених разлога, даља моделирања су реализована за скуп регресора ρ_{II} . Због обима и броја извештаја, као пример, за *CrossVal4* (LDS:TDS=75:25=1644:547) се даје само дијаграм вредности коригованих коефицијената вишеструке детерминације. Вредности се односе на одабране режиме групног и појединачног уклањања регресора (слика 8.3). Може се уочити да је моделирање режимом у коме је реализовано фино појединачно уклањање регресора са најмањим семипарцијалним коефицијентом, стабилније и тачније у односу на режим у коме се регресори уклањају у блоковима (регресор са најмањим sr^2 , два регресора са два најмања sr^2 , три регресора са три најмања sr^2 итд).



Слика 8.3. Промена вредности R_{adj}^2 у моделирању постепене регресије за ρ_{II}

За претходни поступак моделирања, најтачнији модел је модел тачности $R_{adj-A7}^2 = 0.90458$, добијен у седмом кораку итеративне процедуре у режиму појединачног уклањања регресора из иницијалног регресионог модела. Једноставнији по форми, а високог нивоа тачности су и модели чији су кориговани коефицијенти вишеструке регресије: $R_{adj-A17}^2 = 0.902765$ и $R_{adj-A18}^2 = 0.902144$. Ови модели су добијени у 17., односно 18. кораку итеративне процедуре, респективно. Најједноставнији модел је генерисан у 18. кораку. Његов кодирани облик је дат једначином (8.6).

$$\begin{aligned}
 y_{V05.1} = & 0.8125234 + 0.0560041 \cdot H^2 - 0.2950477 \cdot P + 0.37147419 \cdot P^2 \\
 & + 1.1568288 \cdot T_a - 3.2599364 \cdot T_a^2 + 1.8898029 \cdot T_a^3 - 3.7599455 \cdot T_s^3 + \\
 & 5.1670149 \cdot T_s^4 - 1.8880244 \cdot T_s^5 - 4.1921665 E - 03 \cdot t
 \end{aligned} \quad (8.6)$$

Најтачнији модел, генерисан у режиму блоковског уклањања регресора је модел генерисан у 17. кораку итеративног поступка, са $R_{adj-A17}^2 = 0.89031$.

За моделе у облику постепених регресија, поред R_{adj}^2 , у циљу потпуне упоредивости са моделима које генеришу MLR/GA и ANN/GA, су одређивани и коефицијенти вишеструке детерминације R^2 . Вредности наведених коефицијената и мера квалитета за претходно генерисане моделе су дате у табели 8.7.

Табела 8.7. Квалитет најбољих постепених регресија
над ρ_{II} скупом регресора (CrossVal4)

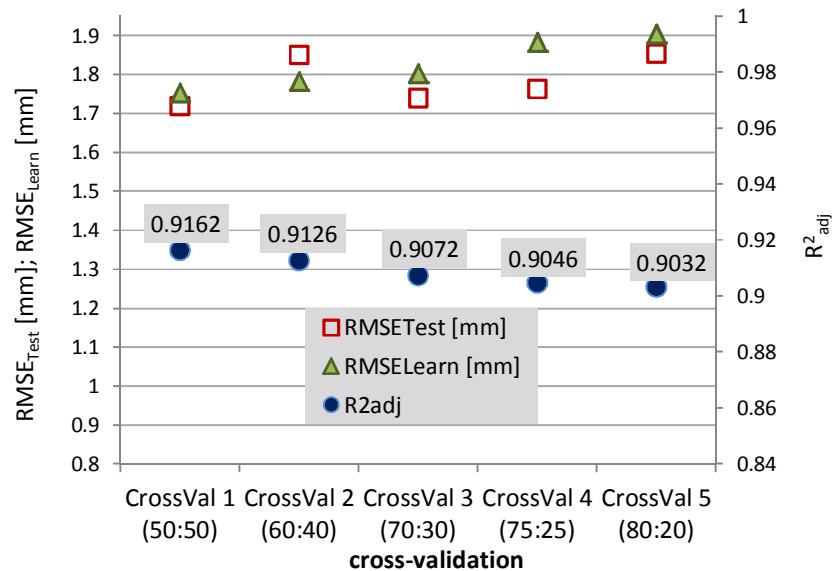
R_{adj}^2	R^2	RMSE _{Learn} [mm]	RMSE _{Test} [mm]	Режим/корак/број регресора
0.9046	0.9059	1.8825	1.7627	Појединачно уклањање регресора / A.7. / 21 регресор $\rho_{II} : \{H, H^2, H^3, P, P^2, T_A, T_A^2, T_A^3, T_A^5, e^{-\frac{T_A}{2}},$ $T_S, T_S^2, T_S^3, T_S^4, T_S^5, \sqrt{T_S}, t, \sqrt{t}, e^{-t}, e^{-\frac{t}{2}}, e^{-\frac{t}{4}}\}$
0.9022	0.9028	1.9133	1.7930	Појединачно уклањање регресора / A.17./ 11 регресора $\rho_{II} : \{H^2, P, P^2, T_A, T_A^2, T_A^3, T_S^2, T_S^3, T_S^5, t, e^{-t}\}$

На основу претходног су дефинисане следеће одреднице: а) На објектима моделирања код којих је присутна интензивна стохастика али и доминантан утицај једног или два фактора са мањом варијансом података (T_s , H ,...) се могу постепеним регресијама добити модели задовољавајуће тачности ($R_{adj}^2 \geq 0.9$) и задовољавајуће једноставности; б) Метода постепеног, појединачног уклањања регресора из регресионе једначине показује боље ефекте у односу на режим у коме се регресија креира блоковским елиминисањем регресора; в) Примена постепених регресија је временски ефикасна, јер је број циклуса односно корака израчунавања релативно мали.

Водећи рачуна о тачности и сложености, са једне стране, и стабилности процеса генерисања што тачнијих модела, са друге стране, у даљем току истраживања су сви тестови реализовани у режиму постепеног појединачног уклањања регресора из иницијалног регресионог модела. У циљу реализације комплетне *cross-validation* стратегије, моделирања су реализована за све преостале односе скупова за учење и тестирање, дефинисане табелом. 8.2. Добијени резултати су дати збирно у табели 8.8, а дијаграмски приказ квалитета генерисаних модела је дат на слици 8.4.

Табела 8.8. Квалитет најбољих постепених регресија над ρ_{II} скупом регресора

LDS:TDS	R_{adj}^2	R^2	RMSE _{Learn} [mm]	RMSE _{Test} [mm]	Регресори модела
CrossVal 1 50:50	0.9162	0.9174	1.7025	1.7886	$\{H^2, P, P^2, T_A, T_A^2, T_A^5, T_S, T_S^5, e^{-T_S}, t^2\}$
CrossVal 2 60:40	0.9126	0.9133	1.7325	1.7001	$\{H, P, P^2, T_A^2, T_A^3, T_A^5, T_S, T_S^5, e^{-T_S}\}$
CrossVal 3 70:30	0.9072	0.9077	1.8025	1.7388	$\{H, P, P^2, T_A^2, T_A^3, T_A^5, T_S, T_S^5, e^{-T_S}\}$
CrossVal 4 75:25	0.9046	0.9059	1.8825	1.7627	$\{H, H^2, H^3, P, P^2, T_A, T_A^2, T_A^3, T_A^5, e^{-T_A/2}, T_S, T_S^2, T_S^3, T_S^4, T_S^5, \sqrt{T_S}, t, \sqrt{t}, e^{-t}, e^{-t/2}, e^{-t/4}\}$
CrossVal 5 80:20	0.9032	0.9036	1.9043	1.8529	$\{H, P, P^2, T_A, T_A^2, T_A^5, T_S, \sqrt{T_S}, T_S^5, e^{-T_S}\}$



Слика 8.4. Квалитет модела генерисаних техником постепене регресије

На основу добијених резултата се може закључити следеће: а) Моделирање помоћу постепених регресија одликује стабилност за читав домен *cross-validation* тестирања; б) Просечна грешка модела добијених постепених регресијама у којима се као термичка променљива користи температура структуре T_s , је $\overline{RMSE}_{Test} = 1.7686$ mm, што представља 8.89% од распона варирања радијалног померања тачке P1; в) Грешке модела обрачунате над скупом података за учење, могу бити веће од грешака модела које су израчунате над тест скупом података; г) Модели имају просечну величину од 10 регресора; д) Регресори са ефектима старења материјала углавном су елиминисани из модела, што указује на њихов незнатан утицај на деформације у телу бране Бочац; ђ) Регресори који описују падавине су присутни у генерисаним моделима, иако за бетонске бране њихов утицај није значајан. Кодирани модели регресија нису приказани због обима, али се на основу кодиране једначине модела (8.6), за *CrossVal 4*, може закључити да је утицај регресора са падавинама, релативно мали.

Као и код моделирања помоћу MLR/GA и ANN/GA хибрида, за анализу утицаја облика регресора на квалитет добијених модела су тестиране и постепене регресије. За термичка оптерећења, уместо температуре структуре T_s , су коришћени дани у години d (контејнер регресора ρ_{III}). Моделирање је реализовано за скупове података са односом LDS:TDS=75:25. Све улазне променљиве, укључујући и променљиву d , су нормализоване на интервал [0,1]. Резултати моделирања за режим блоковског уклањања регресора и режим појединачног уклањања регресора, су дати у табели 8.9.

Табела 8.9. Квалитет најбољих постепених регресија над ρ_{III} скупом регресора

Режим	R2	R2adj	RMSE _{Train} [mm]	RMSE _{Test} [mm]	Регресори у финалном моделу
блоковско уклањање регресора	0.96385	0.96331	1.1002	1.1027	$T_a, \sqrt{H}, t, t^2, P, P^2, \sin(2 \cdot \pi \cdot d), \cos(2 \cdot \pi \cdot d), \sin(4 \cdot \pi \cdot d), \cos(4 \cdot \pi \cdot d), \sin(6 \cdot \pi \cdot d)$
појединачно уклањање регресора	0.96366	0.96226	1.1087	1.1103	$T_a, \sqrt{H}, t^2, P, P^2, \sin(2 \cdot \pi \cdot d), \cos(2 \cdot \pi \cdot d), \cos(4 \cdot \pi \cdot d),$

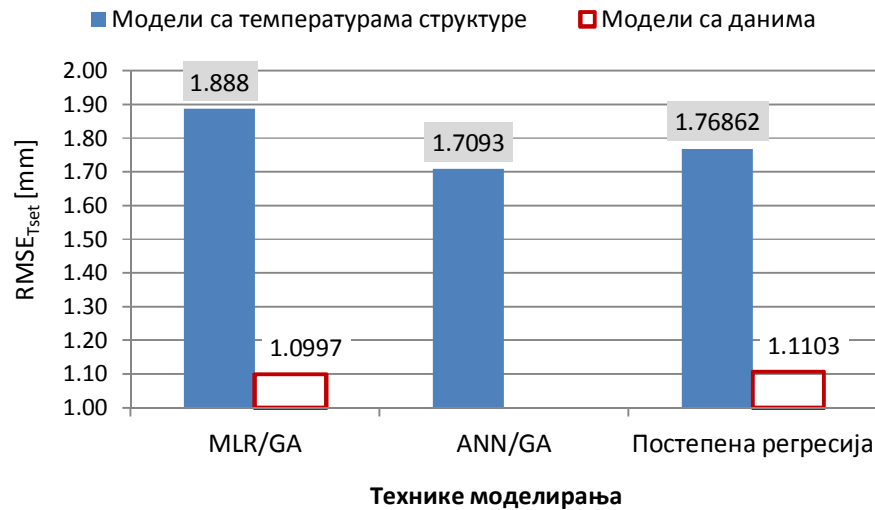
На основу добијених резултата се може закључити: а) Квалитет модела битно зависи од облика регресора у контејнеру из кога се генерише иницијални регресиони модел. Другим речима, природа проблема и предзнање о процесу и избору регресора од стране истраживача битно утичу на квалитет модела; б) Периодичност процеса је доминантна, тако да је увођење променљиве d уместо променљиве која обухвата температуру структуре, довело да значајног побољшања квалитета модела. Средња вредност корена средње квадратне грешке односно квалитет регресионог модела са вредности, $\overline{RMSE}_{Test} = 1.7686$ mm се побољшао на $\overline{RMSE}_{Test} = 1.1065$ mm, што представља 5.56% од распона варирања радијалног померања тачке P1.

8.4.4. Компаративна анализа модела које су генерисали MLR/GA и ANN/GA хибриди са моделима у облику постепених регресија

У циљу компарације, у табели 8.10 и на слици 8.5, су дати упоредни резултати за сва три поступка моделирања: MLR/GA хибрид и ANN/GA хибрид и постепене регресије.

Табела 8.10. Упоредни квалитети модела $RMSE_{Test}$, добијени помоћу: MLR/GA, ANN/GA хибрида и постепених регресија за скупове регресора ρ_{II} и ρ_{III}

Скуп података	MLR/GA	ANN/GA	Постепена регресија	MLR/GA	ANN/GA	Постепена регресија
	ρ_{II} : $RMSE_{Test}$ [mm]			ρ_{III} : $RMSE_{Test}$ [mm]		
CrossVal 1	1.78	1.7800	1.7886			
CrossVal 2	1.98	1.7900	1.7001			
CrossVal 3	1.92	1.7173	1.7388			
CrossVal 4	1.94	1.6092	1.7627			
CrossVal 5	1.82	1.6500	1.8529	1.0997		1.1103
просек	1.888	1.7093	1.7686	1.0997		1.1103



Слика 8.5. Просечни квалитет модела добијен техником постепених регресија, као и помоћу MLR/GA и ANN/GA хибрида

На основу резултата у оквиру овог дела дисертације и табеле 8.10 се могу извести следећи закључци: а) Величина грешке $RMSE_{Learn}$ и $RMSE_{Test}$, зависи од стохастике и обима података у скупу за учење и скупу за тестирање. При моделирању података великог обима не мора се јављати уобичајени однос $RMSE_{Learn} < RMSE_{Test}$; б) За константан однос величина скупова за учење и тестирање, унутрашњи механизми GA стохастике не доводе до значајних варирања квалитета модела; в) Као и при моделирању са MLR/GA хибридом, увођење тригонометријских регресора у контејнер базисних функција за постепене регресије доводи до значајног побољшања квалитета модела; г) Када се термички ефекти исказују преко температуре структуре, ANN/GA хибриди генеришу моделе највишег квалитета. Постепене регресије генеришу моделе који су нешто вишег нивоа квалитета у односу на моделе креиране помоћу MLR/GA хибрида; д) Када се термички ефекти исказују кроз дане у години, модели генерисани са MLR/GA хибридом, су нешто веће тачности у односу на моделе добијене са постепеним регресијама; е) Квалитет модела је упоредив; е) Са аспекта временске ефикасности техника постепених регресија је у предности у односу на хеуристичке вештачке интелигенције.

9. Редукција димензија простора истраживања код регресионих модела заснованих на MLR/GA и ANN/GA хибридика

Број предикторских променљивих и потенцијалних регресора утиче на сложеност регресионог модела као и на потребно процесорско време за његово генерисање. По том основу, упутно је, у поступку предпроцесирања размотрити методе за смањење потенцијалне сложености модела помоћу техника *факторске анализе*¹. Оне омогућавају да се мноштво повезаних (потенцијално корелираних) предикторских променљивих редукује на број који је погодан за моделирање помоћу других техника, као што су технике засноване на вишеструким линеарним регресијама [10]. Основне технике факторске анализе, PCA и Анализа заједничких фактора (*Common Factor Analysis - CFA*), генерално доводе до смањења броја улаза а тиме и обима контејнера базисних функција.

У оквиру ове дисертације се, за редукцију димензија простора истраживања, користи PCA техника. Њена валидација је реализована у оквиру студије случаја моделирања структурног понашања бране Гранчарево. PCA доприноси регуларизацији регресионих модела које генеришу развијени MLR/GA и ANN/GA хибриди и смањењу потребног процесорског времена за њихово генерисање.

9.1. Анализа главних компоненти (Principal Components Analysis - PCA)

PCA² је техника и метода мултиваријационе анализе која покушава да скуп потенцијално корелираних променљивих претвори у скуп променљивих и компоненти (фактора) који нису у корелацији [118]. Теоријске основе и примери примене PCA парадигме, су описани у великом броју литературних референци ([119], [10] и друге). Због обима, теоријска основа није приказана у оквиру ове дисертације, већ су, у основним цртама, описане само главне етапе и препоруке за њену примену.

PCA се реализује у оквиру три етапе, уз поштовање препорука које су често емпиријског карактера.

Оцена прикладности података за факторску анализу

У оквиру ове етапе се разматрају два главна питања: *величина узорка и јачина веза између предикторских променљивих*.

- Препоруке за величину узорка се веома разликују, али се генерално може уопштити да је за квалитетну факторску анализу потребно најмање 300 примера за обучавање, а да је 150 опсервација доња граница [10]. Nunnally

¹ У домену *факторске анализе* под појмом *фактор* односно *компонента* подразумева се група повезаних променљивих, тако да овај термин нема исто значење као код *регресионе* или *дисперзионе анализе* где се термин *фактор* често употребљава као синоним за *предикторску* променљиву, односно *независну променљиву*.

² Ову технику је први пут описао Karl Pearson 1901. године. Детаљнију разраду 1933. год., дао је Hotelling. Ипак, због потребе за компликованим и заморним израчунавањима за матрице са већим димензијама, техника је доживела своју пуну употребу тек са развојем и применом рачунара.

препоручује минимално 10 случајева (опсервација) за сваку ставку коју треба факторски анализирати [120].

- Јачина веза (корелација) између предикторских променљивих се може проценити уопштено на бази корелационе матрице. Према [10], скуп података чија корелациона матрица садржи већи број Pearson-вих коефицијената већих од 0.3, је добар кандидат за факторску анализу. Ако је ових вредности мало, ти подаци су неприкладни за факторску анализу. Детаљнија анализа адекватности се може реализовати преко *Бартлетовог теста сферичности*³ [121], чији показатељ треба да буде значајан ($p < 0.005$) да би подаци били погодни за факторску анализу. Друга техника је Кајзер-Мејер-Олкинов⁴ [122-123] показатељ (КМО) чија вредност треба да буде једнака или изнад 0.6 као предуслов за квалитетну факторску анализу.

Издајање фактора

У оквиру ове етапе, на бази помоћних техника, истраживач треба да одреди број компоненти (број димензија), које најбоље описију односе међу предикторским променљивим. Најчешће се, у циљу што квалитетније процене, примењују следеће технике:

- **Кајзеров критеријум** који је базиран на *карактеристичним вредностима фактора* које представљају укупне варијансе свих променљивих које су тим фактором објашњене. Фактори чије су карактеристичне вредности веће од 1.0 су погодни за даље анализе.
- **Критеријум дијаграма превоја (Scree test)** [124] – Овај дијаграм представља дијаграм карактеристичних вредности⁵ свих фактора. Потребно је на овом дијаграму уочити место у коме дијаграм мења облик и прелази у хоризонталну линију. Кател препоручује да се оставе за даљу анализу сви фактори изнад превоја, јер највише доприносе објашњавању варијансе у скупу података.
- **Паралелна Ногн-ова анализа** [125-127] - која пореди износе карактеристичних вредности фактора са онима који су добијени на једнако великом скупу случајно генерисаних података⁶ и задржава оне факторе чије *eigenvalues* су веће.

Ротација и тумачење фактора

Ово је етапа у којој је потребно да фактори добију тумачење и значење. Ротација открива групе сродних предикторских променљивих које потенцијално могу чинити компоненту, али именовање и тумачење компоненти је посао истраживача. Често се ради о *ортогонална* и *коса ротација фактора*, а користе резултати оних ротација које се лакше могу протумачити и које су физикално јасније. Идеални случај којем се истраживач може надати је *једноставна структура* у којој је свака предикторска променљива јако колерирана са само једном компонентом, односно свака компонента је састављена од више променљивих које су са њом у јакој корелацији [128]. У ортогоналне технике спадају *Varimax*, *Quairtimax* и *Equamax*, а у косе *Direct Oblimin*, *Promax* и друге. Поређење различитих врста ротација је дато у књизи [10].

³ *Bartlett's Test of Sphericity*

⁴ *Kaiser-Meyer-Olkin Measure of Sampling Adequacy*

⁵ *eigenvalues*

⁶ За поступке PCA паралелне анализе се могу, поред осталих, користити софтверски пакети SPSS, R и MonteCarlo PCA for Parallel Analysis (Marley Watkins, 2000). У оквиру статистичког софтвера R, PCA техника је имплементирана кроз функцију *prcomp* у оквиру пакета *stats* [118].

Поред наведених тестова и показатеља, за квалитетну PCA анализу је неопходно да су предикторске променљиве у линеарној вези и потребно је елиминисати нетипичне тачке (*outliers*) у току почетног пречишћавања података.

Техника PCA је веома суптилна статистичка метода. Она од истраживача очекује доношење већег броја одлука, заснованих како на теоријском знању, тако и на емпиријској подлози. У основи је реч о итеративној процедури чији параметри морају бити прилагођени специфичној, конкретној области истраживања.

У литератури се могу наћи различити приступи у тумачењу компаративних предности и недостатака ове методе [43, 10, 129].

9.2. PCA и редукција димензија у функцији MLR/GA и ANN/GA хибрида

У оквиру ове дисертације, у поглављу бр. 6 и бр. 7, су приказани развијени GenReg и DEVONNA системи односно MLR/GA и ANN/GA хибриди за генерисање и адаптацију линеарних и нелинеарних регресионих модела на бази генетских алгоритама. Развијени софтверски агенти су намењени за моделирање комплексних објеката па је повећање њихове ефикасности, са аспекта времена потребног за генерисање модела, од великог значаја. Додатно се очекује да креирани математички модели имају елегантну форму, односно да су засновани на идејама регуларизације.

Временску ефикасност и регуларизацију је могуће значајно унапредити код објеката које карактерише висок степен мултиколинеарности. Мултиколинеарност је сценарио у коме више променљивих мери исту или сличну величину која се налази у подлози истраживаног процеса или објекта. У многим системима, има само неколико оваквих снажних, суштинских компоненти. У том контексту се PCA техника може сматрати претходником у процесима математичког моделирања оваквих процеса, објеката и система.

Валидација MLR/GA и ANN/GA хибрида у поглављима 6 и 7, је реализована за бране Бочац и Гранчарево. Како је реч о објектима моделирања које карактерише висока мултиколинеарност улазних променљивих (бројни сензори температуре у различитим тачкама бране), бране су, због редудансе података [45], одабране и за редукцију димензија простора истраживања, помоћу PCA технике.

Најпре се у делу 9.2.1, наводе додатни разлози за примену ове факторске технике за бетонске бране, а потом даје кратак преглед научне литературе у домену хидрологије у PCA контексту (део 9.2.2). У делу 9.3, се даје приказ резултата PCA за брану Гранчарево. За редуковани простор истраживања, су генерисани модели радијалног померања одабране тачке помоћу MLR/GA и ANN/GA хибрида и у делу 9.4 приказана компаративна анализа квалитета ових модела.

9.2.1. Разлози за примену PCA редукције димензија при моделирању структурног понашања бетонских брана

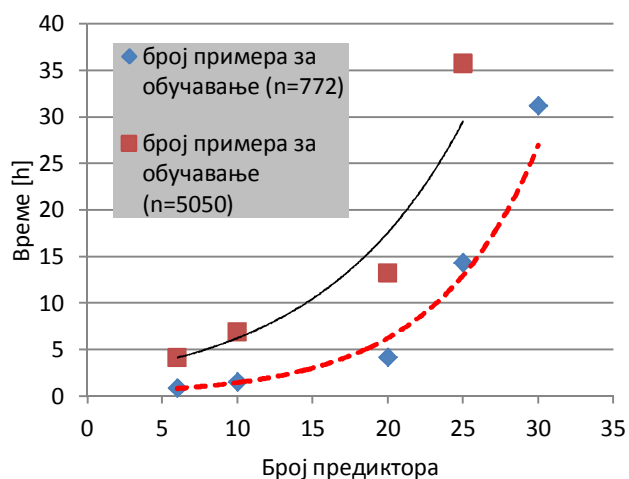
Савремени мониторинг бетонских лучних брана на акумулационим језерима обухвата неретко од 50 до 500 сензора, па је оперисање са оваквим бројем улазних променљивих у креирању регресионих модела оваквих објеката, велики изазов.

Моделирање радијалног померања карактеристичних тачака у структури бране помоћу MLR/GA хибрида карактеришу модели који имају линеарну регресиону форму. Сложеност израза којима се MLR приказују, зависи од обима и природе података али и

од обима контејнера базисних функција. Ако је број улазних променљивих у интервалу од 50 до 500, што је реална претпоставка, и ако свака променљива генерише бар 10 регресора различитог облика, онда контејнер базисних функција може имати од 500 до 5000 регресора. MLR модели који настају над овако обимним контејнером базисних функција, често су преприлагођени и имају превише сложен и неелгенатан аналитички облик. Процес моделирања траје у просеку од 5-10 минута за контејнере мањег обима и мањи број примера за обучавање, до 0.5 h, за контејнере већег обима, веће сложености регресора и скупове података већег обима. Иако у великом броју апликација, време генерисања модела није критичан фактор, критичан фактор је регуларизација, односно потенцијална сложеност генерисаних MLR модела. По овом основу примена PCA факторске анализе са аспекта MLR/GA хибрида је потпуно оправдана.

Иако вештачким неуронским мрежама, због њихове природе, одговарају обимни скупови података за обучавање и иако оне могу радити са великим бројем улазних променљивих, адаптација и оптимизација архитектуре, алгоритама и правила учења код ANN/GA хибрида (поглавље 7) отвара значајне проблеме са аспекта времена које је потребно за генерисање модела. Природа GA платформе подразумева оперисање са 70 до 100 генерација величине до 120 јединки, односно ANN модела по генерацији (поглавље 4). Даље, свака од ових јединки (са потенцијалних 50 до 500 улаза, може имати од 1 до преко 30 слојева са различитим бројем неурона, са различитим AF и параметрима учења. Ако се претходном придружи и *cross-validation* концепт и обим скупова од 1000 до 5000 примера за обучавање, онда се може закључити да ANN/GA хибрид у процесу моделирања, адаптације и оптимизације треба да оперише са преко 20000 комплексних мрежа и да проблем моделирања може постати рачунарски несавладив.

На слици 9.1, је дат приказ времена потребног за генерисање ANN/GA модела у функцији броја предиктора и обима података. Резултати су добијени при тестирањима за брану Гранчарево. За предикторе су биране променљиве: H , d , t , T_a (чије значење је дато у делу 7.4), као и T_{s1} до T_{s45} (45 променљивих које мере температуре бетона у мерним тачкама у телу бране). Радијално померање тачке P1 (слика 7.8), је представљало функцију циља (слика 7.8).



Слика 9.1. Време потребно за генерисање ANN/GA модела радијалног померања тачке P1 у моделирању структурног понашања бране Гранчарево

Код обимних скупова података и скупова предиктора, време за које ANN/GA хибрид генерише модел експоненцијално расте, без обзира на интензиван раст процесорске снаге савремених рачунара, на процесоре са више језгара, на паралелно рачунарство... Из неведених разлога, примена редукције димензија је у потпуности оправдана и при генерисању нелинеарних регресионих модела помоћу ANN/GA хибрида.

9.2.2. PCA технике – стање и примена у области хидрологије

Редукција димензија помоћу PCA техника се примењује за различите намене које тангирају математичке моделе у домену хидрологије.

У раду [130], M. Varol и коаутори, приказују примену техника мултиваријационе анализе. Примењене су технике PCA и дискриминационе анализе за оцену квалитета воде на бранама *Kralkiz*, *Dicle* и *Batman* у базену реке Тигар, за период од 2008. до 2009. године. Ове методе су коришћене и за оцену временско/просторних варијација података о изворима загађења. Сличан приступ, заснован на PCA/FA за евалуацију временских и просторних варијација квалитета воде, је дат у [131-135].

Westra и коаутори у свом раду [136], пореде перформансе PCA и ICA (*Independent Component Analysis*) техника у моделирању мултиваријабилних хидролошких серија. У научном чланку [137], Wu-jiao Dai и остали аутори указују на недостатке PCR (*Principal Components Regression*) и постепених регресија односно PLSR (*Partial Least-Squares Regression*) техника у моделирању деформација бетонских брана, које се уобичајено користе у случају постојања вишег нивоа мултикорелације међу предикторима. Главни проблем у њиховој примени је, по наведеним ауторима, то што ови статистички модели веома мало могу допринети физичкој интерпретацији процеса деформисања структура бетонских брана. По том основу они предлажу примену технике ICR (*Independent Component Regression*).

У раду [45], Abbas Majdi и Morteza Beiki користе ANN за предикцију модула деформисања стена у подлози брана на локацији четири акумулације и хидроелектране. Помоћу PCA технике је број предиктора сведен са седам предиктора на једну главну компоненту, која обухвата 63% варијансе почетног скупа података. На основу матрице коефицијената оптерећења, је закључено да прва главна компонента представља линеарну комбинацију три предиктора: UCS (*the Uniaxial Compressive Strength of the intact rock*), GSI (*the Geological Strength Index*) и RQD (*Rock mass Quality Designation*), и да ће се она користити за развој модела за предикцију модула деформисања стенских подлога.

У раду [138], D. Santillan и коаутори су дефинисали једнодимензионални модел термичког поља у бетонској структури бране, увођењем еквивалентне амбијенталне температуре θ_{eq} . Методологија је примењена на двоструку лучну бетонску брану *La Baells*, која се налази на реци *Llobregat* у провинцији *Barcelona*, Шпанија. У бетонску структуру је уграђен 21 термометар у средњем пресеку бране и 5 термометара за мерење температуре воде на узводном лицу бране. Мерење је и ниво воде. Такође, дневно су мерење и вредности неколико климатских променљивих помоћу метеоролошке станице инсталиране на локацији бране: брзина ветра, глобално сунчево зрачење, средња температура ваздуха, и ваздушни притисак. Еквивалентна амбијентална температура, θ_{eq} , је апроксимирана збиром,

$$\theta_{eq} = \theta_a + \theta_{nc} + \theta_s + \theta_{ec} \quad (9.1)$$

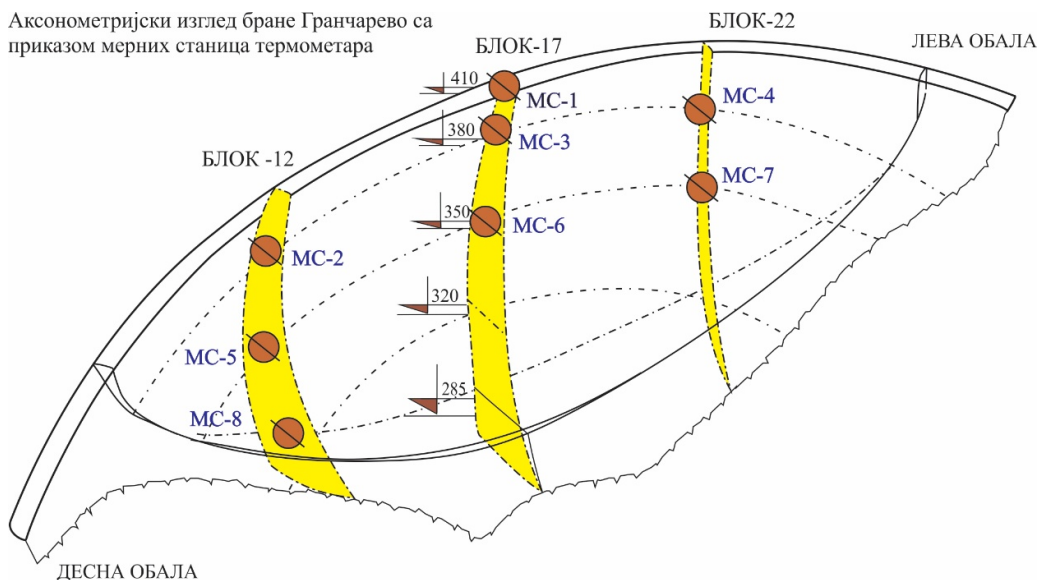
где су θ_a , θ_{nc} , θ_s и θ_{ec} , амбијентална температура, температурна девијација услед ноћног хлађења, температурна девијација која настаје као последица сунчеве радијације, и температурна девијација која настаје као последица испаравања воде, респективно [138]. Променљиве θ_a , θ_{nc} , θ_s и θ_{ec} , имају улогу главних компоненти, али нису одређене РСА техником, већ помоћу теоријских знања и аналитичких израза из домена простирања топлоте.

Према сазнањима аутора ове дисертације, РСА техника за редукцију димензија која се односи на температуру тела бетонских брана на хидро-акумулацијама, у радовима који су били доступни, није била предмет анализе. По том основу је, у оквиру израде докторске тезе, реализовано истраживање и примена РСА, за редукцију броја променљивих које репрезентују температуру ваздуха и температуре у 44 тачке, у блоковима, односно бетонској структури бране Гранчарево на реци Требишњици (Република Српска). Резултати редукције су приказани у делу који следи.

9.3. РСА редукција променљивих бране Гранчарево

За студију случаја у примени РСА технике у оквиру ове дисертације, је одабрана бетонска брана Гранчарево приказана и описана у делу 7.5.

Систем за праћење и предикцију структурног понашања је од августа 2011. године модернизован. Поред више од 400 сензора, за мерење различитих физичких величина, је опремљен и са 44 термометра за мерење температуре структуре (бетона). Термометри су распоређени у 8 мерних станица (слика 9.2).

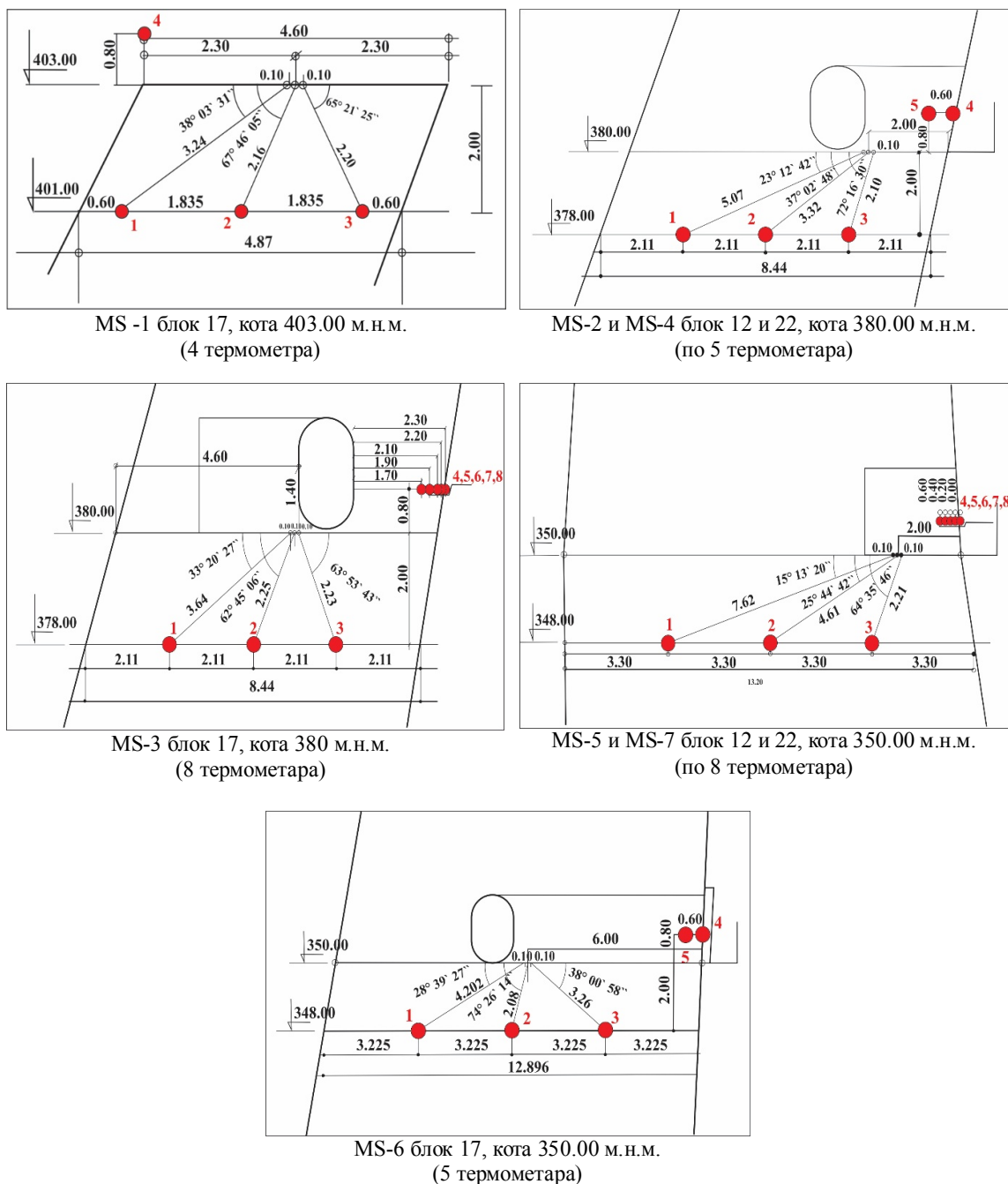


Слика 9.2. Мерне станице за мерење температуре бетона на брани Гранчарево

Позиције термометара дате су у табели 9.1, а шематски приказ локације уградње на слици 9.3.

Табела 9.1. Позиције термометара на брани Гранчарево

Термометар	Кога уградње изнад нивоа мора [м]	Блок	Дубина уградње
MS-11	401,00	17	0,60 м од узводног лица бране
MS-12	401,00	17	2,43 м од оба лица (средина)
MS-13	401,00	17	0,60 м од низводног лица бране
MS-14	403,80	17	на површини
MS-21	378,00	12	2,11 м од узводног лица бране
MS-22	378,00	12	4,22 м од оба лица (средина)
MS-23	378,00	12	2,11 м од низводног лица бране
MS-24	380,80	12	на површини
MS-25	380,80	12	0,60 м од низводног лица бране
MS-31	378,00	17	2,11 м од узводног лица бране
MS-32	378,00	17	4,22 м од оба лица (средина)
MS-33	378,00	17	2,11 м од низводног лица бране
MS-34	380,80	17	0,60 м од низводног лица бране
MS-35	380,80	17	0,40 м од низводног лица бране
MS-36	380,80	17	0,20 м од низводног лица бране
MS-37	380,80	17	0,10 м од низводног лица бране
MS-38	380,80	17	на површини
MS-41	378,00	22	2,11 м од узводног лица бране
MS-42	378,00	22	4,22 м од оба лица (средина)
MS-43	378,00	22	2,11 м од низводног лица бране
MS-44	380,80	22	на површини
MS-45	380,80	22	0,60 м од низводног лица бране
MS-51	348,00	12	3,30 м од узводног лица бране
MS-52	348,00	12	6,60 м од оба лица (средина)
MS-53	348,00	12	3,30 м од низводног лица бране
MS-54	350,80	12	0,60 м од низводног лица бране
MS-55	350,80	12	0,40 м од низводног лица бране
MS-56	350,80	12	0,20 м од низводног лица бране
MS-57	350,80	12	0,10 м од низводног лица бране
MS-58	350,80	12	на површини
MS-61	348,00	17	3,22 м од узводног лица бране
MS-62	348,00	17	6,45 м од оба лица (средина)
MS-63	348,00	17	3,22 м од низводног лица бране
MS-64	350,80	17	на површини
MS-65	350,80	17	0,60 м од низводног лица бране
MS-71	348,00	22	3,30 м од узводног лица бране
MS-72	348,00	22	6,60 м од оба лица (средина)
MS-73	348,00	22	3,30 м од низводног лица бране
MS-74	350,80	22	0,60 м од низводног лица бране
MS-75	350,80	22	0,40 м од низводног лица бране
MS-76	350,80	22	0,20 м од низводног лица бране
MS-77	350,80	22	0,10 м од низводног лица бране
MS-78	350,80	22	на површини
MS-81	330,00	12	0,60 м од низводног лица бране

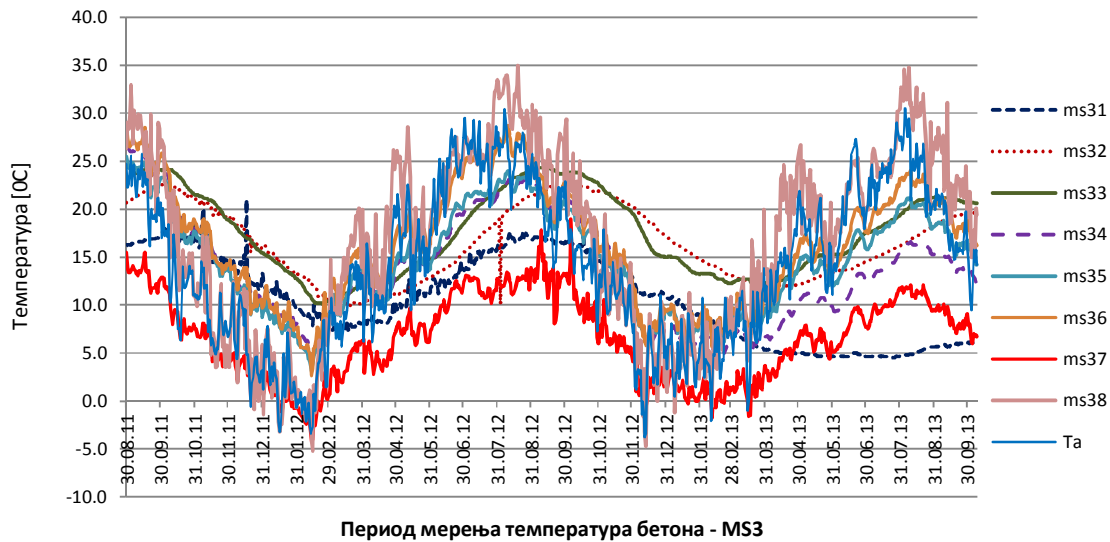


Слика 9.3. Термометри за мерење температуре бетона – шема уградње

База прикупљених података, поред осталих података, садржи 772 записа са вредностима средње температуре ваздуха (T_a) и температурама бетона са 44 термометра MS- ij , за период 30.8.2011-9.10.2013. год. У променљивим ms_{ij} , које репрезентују очитане вредности са ових термометара, i означава број мерне станице, а j број термометра у оквиру i -те станице.

На слици 9.4. је, као пример, приказан дијаграм температура бетона које мере термометри мерне станице бр.3. Уочава се да поједине променљиве (ms_{37} , ms_{38})

имају веома велики шум. То је и природно јер су термометри, који одговарају назначеним променљивим, уграђени на површини низводног лица бране, које је у већем делу дана изложено сунцу и променама температуре ваздуха (T_a). За разлику од наведених термометара, термометри које репрезентују променљиве $ms32$ и $ms33$, мере промену температуре у дубини тела бране, па је промена вредности циклична, али без великог шума.



Слика 9.4. Вредности температуре бетона мерне станице бр.3 за период 30.8.2011-9.10.2013.

Уочава се значајан степен корелације који постоји међу вредностима које појединачни термометри мере. Сличне закономерности и промене температуре бетонске структуре бране важе и за остале мерне станице, а очекивана је висока корелација и између вредности температуре које мере термометри у различитим мерним станицама.

Елементи дескриптивне статистике за вредности температура у бази података за посматрани период 30.8.2011-9.10.2013. год., су дати у табели 9.2.

Табела 9.2. Елементи дескриптивне статистике за средњу дневну температуру (T_a) и температуре у телу бране Гранчарево на мерним станицама од $ms11$ до $ms81$, за период 30.8.2011-9.10.2013. год.

	Ta	ms11	ms12	ms13	ms14	ms21	ms22	ms23	ms24	ms25	ms31	ms32	ms33	ms34	ms35	ms36
		1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
Мин	-3.75	1.5	3.5	-2.9	-8.1	7.7	7.9	11.5	-1.8	6.9	4.5	9.7	10.1	4.6	3.1	2.7
Просек	14.15	13.75	16.18	9.55	17.71	14.17	13.83	17.63	20.73	18.39	10.64	16.41	17.52	13.76	14.61	16.36
Станд. дев.	8.18	6.41	7.33	6.51	11.51	3.95	3.5	4.14	9.48	5.89	4.48	3.91	4.33	5.55	5.29	6.1
Макс.	30.50	25.6	27.8	21	39.4	20.6	19.1	24.6	37.1	29.2	21	22.7	24.4	26.5	25.5	28.9

	ms37	ms38	ms41	ms42	ms43	ms44	ms45	ms51	ms52	ms53	ms54	ms55	ms56	ms57	ms58
	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30
Мин	-4.2	-5.2	7.8	10.7	10.3	-3.5	7.2	10.2	7.8	11.9	5.8	5.2	3.1	1.1	-5
Просек	6.61	16.73	15.65	16.29	15.88	15.58	17.82	12.41	10.72	16.61	18.25	18.36	16.02	16.78	17.09
Станд. дев.	4.47	9.35	4.27	3.77	3.95	7.8	5.51	1.59	2.26	2.94	6.7	7.04	6.81	7.64	9.15
Макс.	19	35	25.8	29.5	23.4	38.8	27.8	15.8	16.1	21.5	30.4	33.6	32.9	33.2	34.1

	ms61	ms62	ms63	ms64	ms65	ms71	ms72	ms73	ms74	ms75	ms76	ms77	ms78	ms81
	31	32	33	34	35	36	37	38	39	40	41	42	43	44
Мин	4.8	4.8	4.8	-4.4	7.1	10.8	3.3	13.4	6.3	5	2.9	1.4	-3.4	3.2
Просек	10.99	9.9	11.01	16.61	17.56	12.35	7.36	17.89	18.88	17.99	17.46	16.92	17.98	13.11
Станд. дев.	4.03	3.37	4.13	8.68	5.06	1.31	3.36	3.08	6.36	6.47	6.79	7.2	7.53	5.58
Макс.	17	17.9	17.3	31.3	26.4	15.7	15.7	27.9	30.2	30.1	32.2	32.9	42.1	24.7

За 45 предикторских променљивих (средња температура ваздуха и температура бетона у 44 позиције у телу бране) је реализована *редукција димензија* помоћу PCA.

Анализа података је показала да је температура ваздуха у јакој корелацији са температурама у телу бране. Из табеле 9.3, се види (засенчена поља), да је број Pearson-ових коефицијената између променљивих које представљају температуру ваздуха и променљивих које представљају температуре у телу бране, са вредностима $|r| > 0.3$, преко 70%.

Табела 9.3. Pearson-ови коефицијенти температуре ваздуха и температура у телу бране

Варијабле за температуре бетона на мерним местима																						
	ms11	ms12	ms13	ms14	ms21	ms22	ms23	ms24	ms25	ms31	ms32	ms33	ms34	ms35	ms36	ms37	ms38	ms41	ms42	ms43	ms44	ms45
Та	0.880	0.825	0.897	0.959	0.182	0.090	0.326	0.903	0.862	0.083	0.218	0.514	0.698	0.858	0.890	0.885	0.944	0.365	-0.095	0.359	0.952	0.846
Варијабле за температуре бетона на мерним местима																						
	ms51	ms52	ms53	ms54	ms55	ms56	ms57	ms58	ms61	ms62	ms63	ms64	ms65	ms71	ms72	ms73	ms74	ms75	ms76	ms77	ms78	ms81
Та	-0.369	0.016	0.527	0.900	0.879	0.907	0.948	0.955	-0.213	-0.353	-0.031	0.938	0.895	-0.526	-0.201	0.248	0.915	0.921	0.945	0.959	0.802	0.876

У складу са претходним, PCA процедура је реализована *са* и *без променљиве* T_a . Циљ је био да, по могућству, *главне компоненте* интегришу T_a променљиву, како би математички модели предикције структурног понашања бране Гранчарево имали што једноставнију форму.

У тексту који следи, најпре се даје детаљан приказ корака редукције димензија са укљученом T_a променљивом. Редукција је реализована на основу корелационе матрице, **R**.

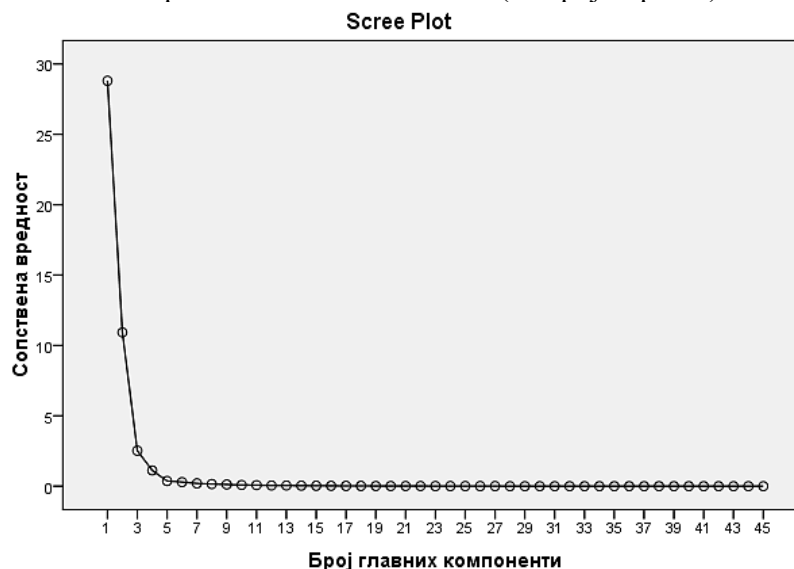
Укупно 45 димензија (T_a и 44 променљиве које представљају температуре бетона у различитим тачкама бетонске структуре) је подвргнуто анализи главних компоненти. Пре спровођења PCA је оцењена прикладност података за факторску анализу. Обим података говори да је реч о узорку великог обима (772). Прегледом корелационе матрице отрквено је много коефицијената вредности веће од 0.3 (73.23%). Вредност Кајзер-Мејер-Олкиновог показатеља била је 0.956, што значајно премашује препоручену граничну вредност [122, 123]. И Бартлетов тест сферичности [121] је достигао статистичку значајност, што указује на факторабилност корелационе матрице. PCA анализа је открила присуство 4 компоненте са сопственим вредностима преко 1 (Кајзеров критеријум) које објашњавају 64.016%, 24.243%, 5.619, и 2.484% варијансе, што укупно чини 96.363% варијабилитета (табела 9.4).

Табела 9.4. Укупни објашњени варијабилитет (извор: SPSS)

Component	Initial Eigenvalues			Extraction Sums of Squared Loadings			Rotation Sums of Squared Loadings ^a
	Total	% of Variance	Cumulative %	Total	% of Variance	Cumulative %	Total
1	28.807	64.016	64.016	28.807	64.016	64.016	27.579
2	10.909	24.243	88.259	10.909	24.243	88.259	7.945
3	2.529	5.619	93.879	2.529	5.619	93.879	16.315
4	1.118	2.484	96.363	1.118	2.484	96.363	6.877
5	.370	.823	97.185				
6	.295	.656	97.841				

Прегледом дијаграма превоја (*Scree Plot*,) је утврђено постојање прелома иза треће

(благи лом), односно четврте тачке или компоненте (оштрији прелаз).



Слика 9.5. Критеријум дијаграма превоја (Scree test) (извор: SPSS)

Дакле, на основу Кателовог критеријума [124], је утврђено да се могу разматрати опције са 2 или 3 главне компоненте. То су подржали и резултати паралелне анализе [125] са 3 компоненте чије сопствене (карактеристичне) вредности премашују одговарајуће вредности прага, добијене помоћу једнако велике матрице случајних бројева (45 променљивих x 772 мерења) (табела 9.5). Списак карактеристичних вредности из табеле 9.4. је поређен са сопственим вредностима добијеним на скупу случајно генерисаних бројева. Генерисање узорака случајних бројева је реализовано помоћу програма *Monte Carlo PCA for Parallel Analysis* (Marley Watkins, 2000). Програм је генерисао 100 скупова случајних бројева који су исте величине као датотека са стварним подацима. Средње вредности сопствених вредности од тих 100 узорака случајних бројева, су упоређиване са првом, другом, трећом, итд. сопственом вредношћу, које су добијене помоћу SPSS-а. Задржаване су само компоненте чије SPSS вредности су веће.

Табела 9.5. Поређење λ_i добијених са PCA и Horn-овом паралелном анализом

Редни број компоненте	Стварна сопствена вредност из PCA	Вредност добијена паралелном анализом	Одлука
1.	28.807	1.4925	прихватити
2.	10.909	1.4458	прихватити
3.	2.529	1.4049	прихватити
4.	1.118	1.3432	одбацити
5.	.370	1.3146	одбацити

На основу претходних критеријума је закључено да треба урадити анализу са избором три, као и са избором две главне компоненте.

Трокомпонентно решење објашњава укупно 93.879% варијансе, при чему је допринос по *Компоненти-1*, *Компоненти-2* и *Компоненти-3*, 64.016%, 24.243% и 5.619%, респективно (табела 9.4).

Главне компоненте након ротације

The 3D scatter plot displays the first three principal components (Komponenta 1, 2, 3) after rotation. The axes are labeled: Komponenta 1 (vertical, 0.0 to 1.5), Komponenta 2 (bottom-left, -0.2 to 1.2), and Komponenta 3 (bottom-right, 0.0 to -1.2). Data points are colored by group: orange (top-left), black (middle), and red (bottom-right). The orange group (ms57, ms58, ms76, ms56, ms64, ms77, ms75, ms81, ms36, ms11, ms55, ms38, ms44, ms74, ms54, ms35, ms34, ms13, ms65, ms12, ms25, ms37, ms45, Ta) is clustered in the upper-left region. The black group (ms53, ms33, ms41, ms43, ms23, ms73, ms31, ms63, ms32, ms21, ms22, ms52, ms61, ms42, ms51, ms62, ms71) is clustered in the middle region. The red group (ms57, ms58, ms76, ms56, ms64, ms77, ms75, ms81, ms36, ms11, ms55, ms38, ms44, ms74, ms54, ms35, ms34, ms13, ms65, ms12, ms25, ms37, ms45, Ta) is clustered in the lower-right region.

Иако на основу дијаграма делује да ротација одлично раздваја предикторе, ипак, компоненте не дају једноставну структуру [128], већ су неке од њих корелиране, што следи из *матрице корелације међу главним компонентама*, Φ (*Factor/Component Correlation Matrix*), која због обима није приказана. Вредности из ове матрице указују да:

- Са друге стране, ротирано решење (приказано преко *Pattern Matrix*⁷: **A**, табела 9.6), је олакшало физикално тумачење сваке од компоненти на основу позиције мерних места у блоковима бране (слика 9.3). Тако су компоненте именоване са **F1=TK1_{NL}** (*Температура Компонента 1 - Низводно Лице*); **F2=TK2_{US}** (*Температура Компонента 2 - Узводно Средиња*); **F3=TK3_{USS}** (*Температура Компонента 3 - Узводно Средиња Средиња*). Тумачење значења компоненти и додела имена је реализована на основу променљивих које имају највеће тежине за односне компоненте. Као пример се наводи тумачење значења и додела назива *Компоненти 1*. Променљиве, односно термометри који за *Компоненту 1* имају велико факторско оптерећење, су разврстани по мерним

⁷ Матрица ротираних факторских оптерећења - **A**

станицама (Колона I). Уочава се да термометри који припадају овој главној компоненти имају бројеве од 4 до 8 у оквиру мерних станица којима припадају. На основу њихових распореда (слика 9.3), се може закључити да се ови термометри налазе на низводном лицу бране или близу површине низводног лица бране.

Табела 9.6. Факторске тежине свих променљивих (за ротиране факторе): *Pattern Matrix*

Компоненте				Колона I	Колона II	Колона III
A	1	2	3	Компонента 1: TK1 _{NL}	Компонента 2: TK2 _{US}	Компонента 3: TK3 _{USS}
ms56	1.018			Са мерне станице бр.1. (MS1)	Са мерне станице бр.3. (MS3)	Са мерне станице бр.1. (MS1)
ms14	1.003			ms14- термометар бр. 4	ms31- термометар бр. 1	ms12- термометар бр. 2
ms58	1.002			ms11- термометар бр. 1	ms34- термометар бр. 4	Са мерне станице бр.2. (MS2)
ms76	.999			ms13- термометар бр. 3	Са мерне станице бр.4. (MS4)	ms22- термометар бр. 2
ms57	.995			ms12- термометар бр. 2	ms42- термометар бр. 2	ms21- термометар бр. 1
ms75	.991			Са мерне станице бр.2. (MS2)	Са мерне станице бр.5. (MS5)	ms23- термометар бр. 3
Ta	.988			ms24- термометар бр. 4	ms51- термометар бр. 1	Са мерне станице бр.3. (MS3)
ms81	.982			ms25- термометар бр. 5	ms52- термометар бр. 2	ms31- термометар бр. 1
ms77	.981			Са мерне станице бр.3. (MS3)	Са мерне станице бр.6. (MS6)	ms32- термометар бр. 2
ms38	.975			ms38- термометар бр. 8	ms61- термометар бр. 1	ms33- термометар бр. 3
ms44	.969			ms36- термометар бр. 6	ms63- термометар бр. 3	Са мерне станице бр.4. (MS4)
ms55	.968			ms37- термометар бр. 7	ms62- термометар бр. 2	ms45- термометар бр. 5
ms64	.964			ms35- термометар бр. 5	Са мерне станице бр.7. (MS7)	ms41- термометар бр. 1
ms24	.962			ms34- термометар бр. 4	ms71- термометар бр. 1	ms42- термометар бр. 2
ms36	.960			ms33- термометар бр. 3	ms73- термометар бр. 3	ms43- термометар бр. 3
ms54	.949			Са мерне станице бр.4. (MS4)	ms72- термометар бр. 2	Са мерне станице бр.5. (MS5)
ms11	.947			ms44- термометар бр. 4		ms52 - термометар бр. 2
ms74	.935			ms45- термометар бр. 5		ms53 - термометар бр. 3
ms37	.916			ms43- термометар бр. 3		Са мерне станице бр.7. (MS7)
ms35	.902			Са мерне станице бр.5. (MS5)		ms78- термометар бр. 8
ms65	.896			ms56- термометар бр. 6		ms73- термометар бр. 3
ms78	.874	.301		ms58- термометар бр. 8		ms72- термометар бр. 2
ms25	.856			ms57- термометар бр. 7		
ms45	.829	-.328		ms55- термометар бр. 5		
ms13	.827			ms54- термометар бр. 4		
ms34	.812	.368		ms53- термометар бр. 3		
ms12	.809	-.334		Са мерне станице бр.6. (MS6)		
ms61	1.026			ms64- термометар бр. 4		
ms51	.998			ms65- термометар бр. 5		
ms63	.933			Са мерне станице бр.7. (MS7)		
ms62	.912			ms76- термометар бр. 6		
ms71	-.426	.726		ms75- термометар бр. 5		
ms31	.677	-.328		ms77- термометар бр. 7		
ms52	.576	-.445		ms74- термометар бр. 4		
ms22		-1.067		ms78- термометар бр. 8		
ms32		-1.012		ms71- термометар бр. 1		
ms21		-1.001		Са мерне станице бр.8. (MS8)		
ms41		-.876		ms81 - термометар бр. 1		
ms42		.300	-.826	+Ta		
ms23		-.819				
ms33	.381	-.803				
ms53	.393	-.792				
ms43	.312	-.718				
ms73		.316	-.693			
ms72		.379	-.490			

Показује се да математичка подлога верно одсликава природне ефекте осунчаности и утицаја окружења на температурно поље у овој зони (зрачење, конвекцију ...). Другим речима, *Компонента 1* мери збирни термички утицај на низводној страни бране. Овој компоненти припадају и променљиве које мере температуре у оквиру *мерне станице бр. 1*, која се налази на врху блока бр. 17, где је утицај осунчавања, такође доминантан. *Компонента 1 Низводно Лице* је примарна, јер има пуно факторских тежина. Тумачење и додела имена за остале две главне компоненте реализовано је на платформи сличних разматрања.

SPSS пакет генерише и матрицу *комуналитета*⁸, чије вредности представљају варијабилитет сваке променљиве, објашњен главном компонентом којој односна променљива припада. Мали бројеви (<0.3) указују да се променљива вероватно не уклапа добро у одређену компоненту, односно да се не уклапа добро са осталим променљивим, што у анализи података за брану Гранчарево није случај. Због обима, подаци и *матрица комуналитета* биће приказани само у резимеу PCA анализе (табела 9.10). Поред претходних анализа и *резидуална матрица* ($\mathbf{R}_{\text{RES}}$) је показала да је реч о квалитетном факторисању јер је број резидуала који је већи од 0.05 само 3%. Резидуална матрица није приказана због обима.

SPSS програмски пакет генерише и матрицу $\mathbf{B}$ (*Component Score Coefficient Matrix*), помоћу које се израчунава и матрица нормализованих *факторских скорова*, $\mathbf{F} = \mathbf{Z} \cdot \mathbf{B}$. Матрица $\mathbf{F}$ (*Factor Scores Matrix*) и матрица $\mathbf{Z}$ (*нормализована матрица оригиналних података X према z score*) су погодне за даље анализе и примене.

Двокомпонентно решење, објашњава још увек велики проценат варијансе: 88.259%, при чему је допринос по *Компоненти-1* и *Компоненти-2*, 64.016%, 24.243%, респективно (табела 9.4).

Као и код анализе са три компоненте, у циљу лакшег тумачења значења компоненти и доделе имена, је спроведена *oblimin* ротација (слика 9.8). Променљива T_a , која представља средњу температуру ваздуха, је на дијаграму означена кругом црвене боје (доле десно). Ова променљива припада првој главној компоненти.

Из матрице корелације главних компоненти, Φ (*Factor/Component Correlation Matrix*) (табела 9.7) се види да је корелација између главних компоненти релативно мала (15.3%), дакле да ове компоненте нису узајамно линерано зависне.

Табела 9.7. Матрица Φ , за две главне компоненте

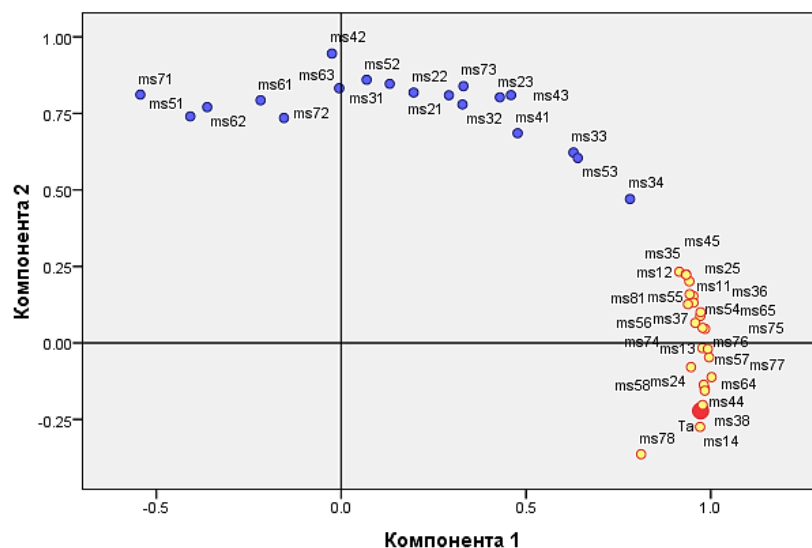
Матрица корелације главних компоненти		
Компонента	1	2
1	1.000	.153
2	.153	1.000

Ово се види и из дијаграма (слика 9.7), који приказује јасно груписање променљивих које представљају термометре са различитих мерних станица. Термометри који су уграђени на узводном лицу бране (бројеви 1 и 2) или средини тела бране (број 3), примарно су придружени *Компоненти 2*, док је група термометара (променљивих) који су уграђени ближе низводном лицу бране или на површини низводног лица бране (термометри са бројевима 4-8), придружена *Компоненти 1*.

⁸ communalities

Ротирано решење (*Pattern Matrix: A*, није приказана због обима) је, као и у случајну три компоненте, олакшало физикално тумачење сваке од компоненти на основу позиције мерних места у блоковима бране Гранчарево. Тако су компоненте именоване са **F1=TK1_{NL}** (*Температура Компонента 1 - Низводно Лице*) и **F2=TK2_{US}** (*Температура Компонента 2 - Узводно Средина*). Тумачење значења компоненти и додела имена је реализована на основу променљивих које имају највеће тежине за односне компоненте као и код трокомпонентног решења.

Дијаграм главних компоненти након ротације



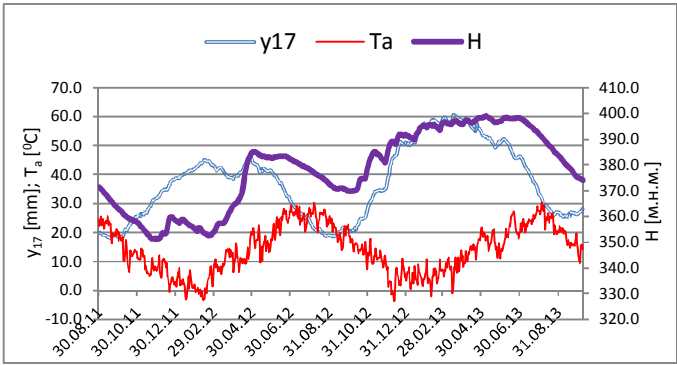
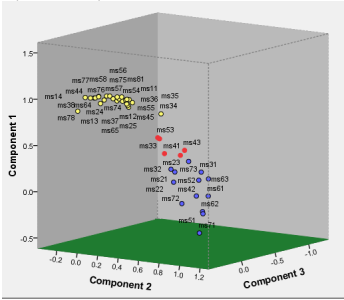
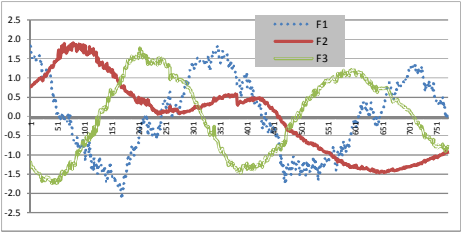
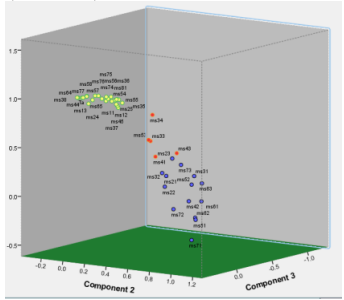
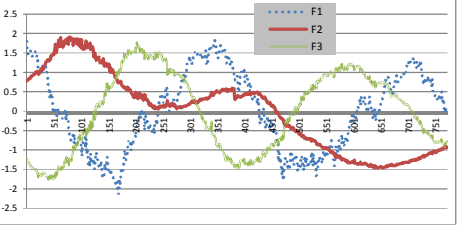
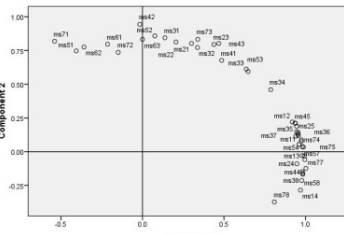
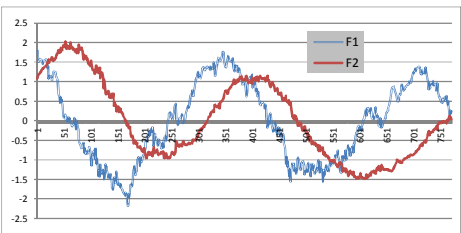
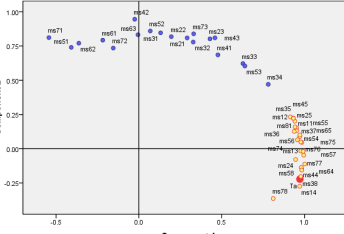
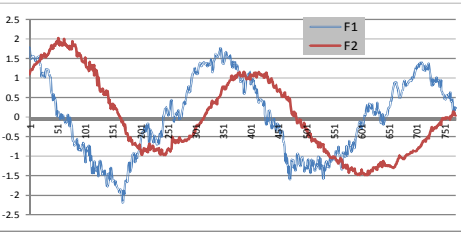
Слика 9.7. Дијаграм две главне компоненте

Као што је раније наведено, комплетна PCA анализа је реализована за скуп од 45 променљивих који је, због јаке корелације и потребе за симплификацијом регресионих модела, укључивао и средњу температуру ваздуха (T_a), и то за трокомпонентно и двокомпонентно решење. Да би се анализирао ниво утицаја променљиве T_a , коју карактерише велики шум, комплетна PCA процедура поновљена је и реализована и за скуп од 44 променљиве, који је садржао само скуп података о температурама структуре бране у задатим тачкама, без температуре ваздуха. Ове процедуре су такође спроведене како за трокомпонентно, тако и за двокомпонентно решење.

Због бројности и величина генерисаних матрица, у табели 9.8 се даје само збирни преглед укупне варијансе које поједине PCA анализе обухватају, дијаграми груписања оригиналних променљивих у главне компоненте, као и графици главних компоненти по решењима.

Иако двокомпонентна решења, која обухватају око 88% укупног варијабилитета свих података, могу водити ка једноставнијим облицима регресионих модела, аутор дисертације се одлучио за трокомпонентно решење које обухвата близу 94% укупне варијансе података.

Табела 9.8. Збирни преглед различитих варијанти PCA анализе

	Дијаграм: Кота воде (H), Температура ваздуха (T_a), Радијални помераји (y_{17}) 
ТРОКОМПОНЕНТНО РЕШЕЊЕ	без T_a А: укупно објашњава 93.782% варијансе 63.791%, 24.284%, 5.708%   Дијаграм главних компоненти
	са T_a Ц: укупно објашњава 93.879% варијансе 64.016%, 24.243%, 5.619%   Дијаграм главних компоненти
ДВОКОМПОНЕНТНО РЕШЕЊЕ	Б: укупно објашњава 88.074% варијансе 63.791%, 24.284% Component Plot in Rotated Space   Дијаграм главних компоненти
	Д: укупно објашњава 88.259% варијансе 64.016%, 24.243% Component Plot in Rotated Space   Дијаграм главних компоненти

Другим речима, у циљу обухвата што већег процента варијабилитета за даље анализе и моделирања је одабрана варијанта (Ц), односно решење са три главне компоненте, које објашњава укупно 93.879% варијансе свих података, при чему свака главна компонента обухвата 64.016%, 24.243% и 5.619% варијансе, редом. Ово решење укључује променљиву T_a , јер је разлика трокомпонентног решења са и без T_a незнатна са аспекта обухваћене варијансе (0.097%).

У табели 9.8, такође, је дат и дијаграм са графицима који приказују: коте нивоа воде (H), средње температуре ваздуха (T_a) и радијалне помераје (y_{17}), за посматрани период. Може се уочити да *Компонента 1-Низводно Лице* ($F1=TK1_{NL}$), која обухвата температуре са термометара који су близу низводне површине бране, одговара промени средње температуре ваздуха, што је и природно, јер компонента $TK1_{NL}$ репрезентује термичку зону која је током већег дела дана слободно изложена сунчевом зрачењу. Компоненте $F2=TK2_{US}$ (*Температура Компонента 2-Узводно Средина*) и $F3=TK3_{USS}$ (*Температура Компонента 3-Узводно Средина Средина*) су фазно померене због термичке инерције зона које су у средини бетонске структуре или ближе површи која је у додиру са водом у акумулацији.

За одабрано трофакторско решење, матрица стандардизованих факторских скорова F , има вредности приказане у табели 9.9.

**Табела 9.9. Матрица F . Решење РСА анализе
за податке бране Гранчарево**

Factor Scores Matrix											
	F1	F2	F3		F1	F2	F3		F1	F2	F3
1	1.82	0.76	-1.19	26	1.10	1.03	-1.64	51	0.04	1.45	-1.68
2	1.62	0.78	-1.27	27	1.10	1.05	-1.63	52	0.11	1.58	-1.63
3	1.51	0.79	-1.32	28	0.99	1.10	-1.67	53	0.10	1.73	-1.70
4	1.52	0.81	-1.32	29	1.01	1.05	-1.64	54	-0.03	1.64	-1.59
5	1.58	0.81	-1.31	30	1.17	1.05	-1.57	55	-0.03	1.50	-1.56
6	1.59	0.83	-1.33	31	1.24	1.07	-1.56	56	-0.11	1.56	-1.65
7	1.56	0.85	-1.38	32	1.23	1.12	-1.60	57	-0.01	1.69	-1.57
8	1.58	0.85	-1.38	33	1.25	1.13	-1.58	58	0.07	1.71	-1.48
9	1.51	0.89	-1.42	34	1.22	1.17	-1.60	59	0.00	1.84	-1.42
10	1.51	0.89	-1.43	35	1.23	1.19	-1.58	60	-0.06	1.66	-1.43
11	1.44	0.90	-1.46	36	1.08	1.25	-1.66	61	0.05	1.75	-1.46
12	1.53	0.90	-1.44	37	1.10	1.26	-1.65	62	-0.03	1.67	-1.46
13	1.53	0.93	-1.47	38	1.04	1.29	-1.67	63	-0.05	1.90	-1.52
14	1.52	0.93	-1.48	39	0.99	1.29	-1.68	64	-0.08	1.69	-1.45
15	1.53	0.94	-1.49	40	0.68	1.32	-1.75	65	-0.15	1.67	-1.50
16	1.56	0.96	-1.49	41	0.60	1.27	-1.70	66	-0.17	1.78	-1.41
17	1.58	0.98	-1.49	42	0.58	1.29	-1.66	67	-0.14	1.64	-1.44
18	1.57	1.00	-1.49	43	0.38	1.30	-1.72	68	-0.13	1.69	-1.45
19	1.48	0.97	-1.52	44	0.56	1.31	-1.64	69	-0.12	1.71	-1.37
20	1.33	1.01	-1.60	45	0.42	1.43	-1.72	70	0.07	1.78	-1.30
21	1.37	0.99	-1.59	46	0.50	1.40	-1.66	71	0.01	1.83	-1.28
22	1.01	0.99	-1.69	47	0.45	1.41	-1.65	72	-0.08	1.74	-1.25
23	1.10	0.92	-1.59	48	0.35	1.43	-1.70	73	-0.25	1.77	-1.34
24	1.10	0.99	-1.61	49	0.06	1.45	-1.74	74	0.03	1.80	-1.17
25	1.09	1.02	-1.63	50	0.00	1.49	-1.73	75	-0.07	1.86	-1.24
								...	...	...	...
								...	...	...	...

Напомена: приказано је, због обима, само 75 од укупно 772 записа

Главне компоненте, чији су стандардизовани скорови дати у матрици F , су подлога за развој регресионих модела радијалног померања посматраних тачака бетонске структуре бране Гранчарево. Ови модели су описани у наредном поглављу.

На основу претходне детаљне РСА анализе може се генерисати резиме дат у табели 9.10, са кључним матрицама A (*Pattern Matrix*) и S (*Structure Matrix*).

Табела 9.10. Матрица факторских тежина (А) и матрица корелација променљивих и фактора (S) за PCA, са oblimin ротацијом трофакторског решења за мерење температура у систему бране Гранчарево

	Pattern Matrix: A			Structure Matrix: S			
	Компоненте			Компоненте			Communalities
ГР11-13	1	2	3	1	2	3	део варијансе објашњен заједничким факторима
ms56	1.000	0.164	0.061	0.981	0.045	-0.384	0.982129
ms14	1.000	-0.105	0.194	0.942	-0.281	-0.127	0.949217
ms58	1.000	-0.037	0.098	0.969	-0.171	-0.252	0.951944
ms76	0.999	0.013	0.009	0.994	-0.081	-0.363	0.988575
ms57	0.995	-0.016	0.011	0.993	-0.111	-0.346	0.986156
ms75	0.991	0.069	-0.016	0.991	-0.013	-0.410	0.98781
Ta	0.988	-0.100	0.136	0.948	-0.250	-0.182	0.936877
ms81	0.982	0.174	-0.002	0.967	0.085	-0.437	0.964699
ms77	0.981	-0.095	0.008	0.987	-0.188	-0.309	0.983863
ms38	0.975	-0.122	0.090	0.953	-0.250	-0.213	0.94023
ms44	0.969	-0.108	0.040	0.964	-0.213	-0.267	0.946863
ms55	0.968	0.160	-0.054	0.973	0.096	-0.479	0.982997
ms64	0.964	-0.121	0.035	0.963	-0.224	-0.265	0.94644
ms24	0.962	-0.010	0.057	0.942	-0.122	-0.290	0.890192
ms36	0.960	0.125	-0.083	0.979	0.075	-0.490	0.989995
ms54	0.949	0.029	-0.103	0.984	-0.011	-0.464	0.98142
ms11	0.947	0.098	-0.097	0.974	0.055	-0.487	0.974833
ms74	0.935	-0.046	-0.130	0.987	-0.074	-0.452	0.985565
ms37	0.916	0.031	-0.153	0.969	0.015	-0.502	0.965363
ms35	0.902	0.102	-0.191	0.963	0.104	-0.566	0.987226
ms65	0.896	-0.061	-0.210	0.978	-0.050	-0.511	0.98625
ms78	0.874	-0.109	0.301	0.774	-0.321	0.029	0.719668
ms25	0.856	-0.013	-0.281	0.960	0.033	-0.589	0.987166
ms45	0.829	-0.034	-0.328	0.952	0.035	-0.616	0.989414
ms13	0.827	-0.270	-0.283	0.955	-0.220	-0.467	0.980826
ms34	0.812	0.368	-0.210	0.855	0.387	-0.669	0.977813
ms12	0.809	-0.029	-0.334	0.934	0.045	-0.618	0.960708
ms61	0.069	1.026	0.102	-0.061	0.975	-0.374	0.958218
ms51	-0.116	0.998	0.142	-0.259	0.946	-0.254	0.93879
ms63	0.215	0.933	-0.046	0.147	0.934	-0.535	0.927442
ms62	-0.129	0.912	0.016	-0.218	0.916	-0.337	0.857876
ms71	-0.426	0.726	-0.217	-0.412	0.860	-0.380	0.882055
ms31	0.204	0.677	-0.328	0.263	0.803	-0.700	0.82662
ms52	0.081	0.576	-0.445	0.192	0.764	-0.728	0.77977
ms22	-0.140	-0.080	-1.067	0.257	0.402	-0.980	0.977887
ms32	0.011	-0.068	-1.012	0.388	0.376	-0.986	0.976966
ms21	-0.011	-0.023	-1.001	0.358	0.418	-0.987	0.974618
ms41	0.208	-0.041	-0.876	0.532	0.325	-0.934	0.915723
ms42	-0.195	0.300	-0.826	0.080	0.681	-0.887	0.920983
ms23	0.225	0.152	-0.819	0.511	0.491	-0.968	0.982036
ms33	0.381	-0.039	-0.803	0.678	0.279	-0.925	0.990513
ms53	0.393	-0.049	-0.792	0.688	0.263	-0.915	0.982186
ms43	0.312	0.260	-0.718	0.551	0.547	-0.947	0.994042
ms73	0.205	0.316	-0.693	0.430	0.602	-0.907	0.905921
ms72	-0.205	0.379	-0.490	-0.060	0.613	-0.582	0.529454

Напомена: За сваку од главних компоненти, факторске тежине веће од 0.3 су исписане појачано

Скуп од 45 променљивих (44 $msXY + Ta$), које представљају температуре бетона и средњу температуру ваздуха на брани Гранчарево, је био подвргнут анализи главних компоненти (PCA) у 16. верзији SPSS-а. Пре спровођења PCA, је била оцењена прикладност података за факторску анализу. Обим података говори да је реч о узорку великог обима (772). Прегледом корелационе матрице је откривено много коефицијената вредности веће од 0.3 (73.23%). Вредност Кајзер - Мејер-Олкиновог показатеља је била 0.956, што значајно премашује препоручену граничну вредност 0.6 [122, 123]. И Бартлетов тест сферичности [121] достигао је статистичку значајност, што указује на факторабилност корелационе матрице.

PCA анализа је открила присуство 4 компоненте са карактеристичним вредностима преко 1 (Кајзеров критеријум), које објашњавају 64.016%, 24.243%, 5.619, и 2.484% варијансе, што укупно чини **96.363%** варијабилитета.

Прегледом дијаграма превоја је утврђено постојање прелома иза треће (благи лом) односно четврте тачке (оштрији прелаз). Дакле, на основу Кателовог критеријума се могу разматрати случајеви са 2 или 3 компоненте. То су подржали и резултати паралелне анализе [125] са три компоненте чије карактеристичне вредности премашују одговарајуће вредности прага, добијене помоћу једнако велике матрице случајних бројева (45 променљивих x 772 мерења) (Monte Carlo PCA for Parallel Analysis). Аутор дисертације је детаљно анализирао оба решења: са две и са три главне компоненте и одлучио се за решење са три компоненте.

ТРОКОМПОНЕНТНО решење је објаснило укупно **93.879%** варијансе, при чему је допринос по компонентама: 64.016%, 24.243%, 5.619, редом.

Да би се лакше протумачиле те ТРИ компоненте, спроведена је *oblimin* ротација. Ротирано решење је олакшало физикално тумачење сваке од компоненти, па су фактори (компоненте) именоване са: **F1-TK1_{NL}** (Температуре Компонента 1- Низводно Лице); **F2-TK2_{US}** (Температуре Компонента 2- Узводно Средина); **F3-TK3_{USS}** (Температуре Компонента 3- Узводно Средина Средина). Прва компонента је доминантна и има много факторских тежина. Компоненте (фактори) не дају једноставну структуру [128], већ су неке од њих корелиране у одређеној мери:

- **F1:F2**, мала корелација (-0.091), дакле ове две компоненте нису узајамно зависне
- **F1:F3**, корелација није мала (-0.366), ове две компоненте су ипак узајамно зависне (слаба негативна корелација)
- **F2:F3**, корелација није мала (-0.444), дакле ове две компоненте су значајније узајамно зависне

9.4. Регресиони модели структурног понашања брана на бази MLR/GA и ANN/GA хибрида у PCA редукованом простору

У оквиру овог поглавља су описани развијени линеарни и нелинеарни регресиони модели структурног понашања бетонске лучне бране Гранчарево у PCA редукованом простору (решење са три главне компоненте). Као и у моделима који су описани у делу 7.5, моделирано је радијално померање тачке P1 на круни бране (слика 7.8 а), на блоку бр.17 (слика 7.8 б).

9.4.1. Подаци и мерења

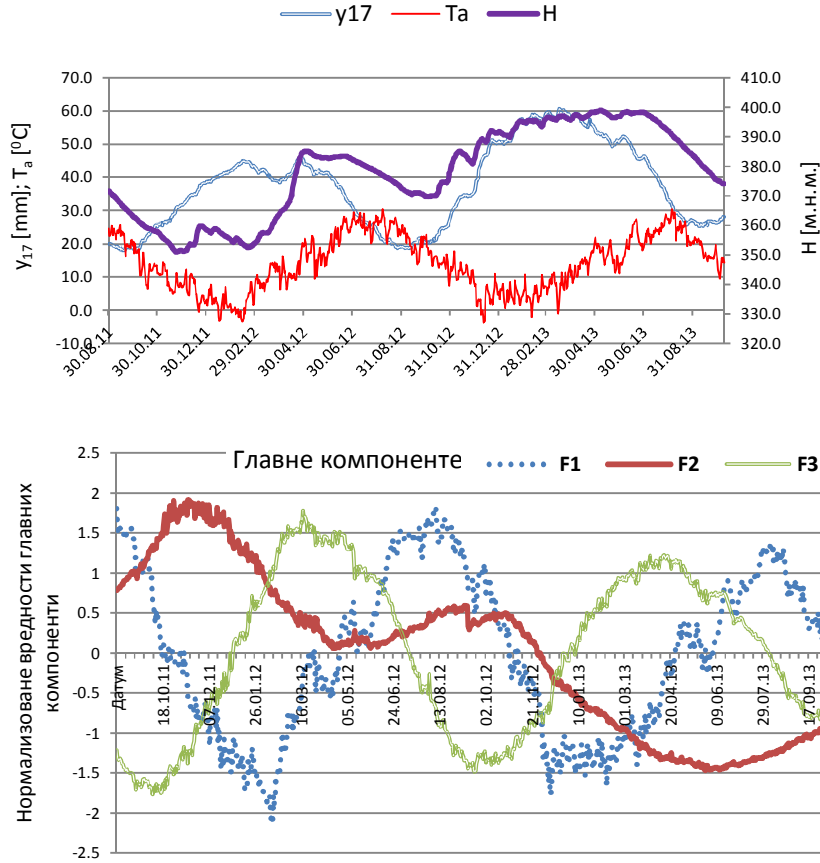
Мерења су вршена у периоду 31.8.2011-9.10.2013. год., тако да је база мерења располагала са 772 записа, са средњом температуром ваздуха, температурама бетонске структуре бране на 44 локације и висином нивоа воде у акумулацији. У складу са *cross-validation* стратегијом [112-114], ови подаци из базе су дељени у LDS и TDS подскупове у задатим односима.

Интервал варирања улазне променљиве H за посматрани период је био од 351.01 до 398.95 m. Термички ефекти су били обухваћени променљивом која представља средњу дневну температуру ваздуха T_a (од -3.75 до 30.50 $^{\circ}\text{C}$), као и променљивим које представљају температуре бетонске структуре на 44 позиције. Интервали варирања температуре бетона су дати у табели 9.2. Након примене PCA методологије, сви термички предиктори (њих 45) су редуковани на три главне компоненте, које су дате у нормализованом облику (табела 9.9). Главним компонентама, односно факторима (F_1, F_2, F_3), су придружене променљиве f_1, f_2 , и f_3 , респективно.

Термички ефекти су такође, као и раније у делу 7.5, били изражени и тригонометријским функцијама, преко улазне променљиве d [1-365], која представља време, мерено у данима, протекло од почетка текуће године до тренутка мерења. Термички ефекти су, у циљу тестирања стабилности модела на температурни померај, испитивани и преко улазних променљивих $d20$ и $d50$, које означавају фазни померај од двадесет ($d20 = d + 20$) или педесет ($d50 = d + 50$) дана, респективно. Ефекат старења у функција времена t је укључивао деградацију својстава материјала током животног века и експлоатације моделиране структуре. Радијално померање тачке P1, исказано променљивом y_{17} , је одабрано за излазну променљиву. Интервал померања променљиве y_{17} у анализираном периоду је био у опсегу од $+17.90$ до $+60.70$ mm, при чему је усвојено да се радијално померање у низводном правцу сматра позитивним. Померање тачке P1 је мерено коришћењем координометра V17-1 (слика 7.8 б).

Дијаграмски приказ дела података и променљивих из базе прикупљених података, намењених за креирање регресионих модела су дати на слици 9.8.

Са слике 9.8, се види да се као последица хидростатичког притиска, радијална померања посматране тачке повећавају са повећањем нивоа воде у акумулацији H , и обрнуто. За разлику од тога, радијална померања се смањују са порастом температуре ваздуха T_a , и обрнуто.



Слика 9.8. Улазно/излазне варијабле ($T_a, H, f_1, f_2, f_3; y_{17}$) и подаци за период 2011-2013.

9.4.2. Параметри и резултати добијени на основу MLR/GA хибридне методе

Вишеструки линеарни регресиони модели су генерисани помоћу GenReg агента на основу MLR/GA методе, која је детаљно описан у поглављу бр. 6 и раду [23].

У складу са *cross-validation* стратегијом су реализоване четири серије тестова са различитим односом LDS:TDS (60:40, 70:30, 75:25 и 80:20), при чему је подскуп примера за учење биран на случајан начин. Параметри генетског алгоритма за MLR/GA хибрид су били: број јединки у популационом базену 100, а број генерација 70. За селекцију је одабрана бинарна турнирска селекција, а за репродукцију укрштање у једној тачки и униформна мутација са вероватноћама од 1.0 и 0.18, респективно. У циљу упоредивости модела брана добијених помоћу MLR/GA и ANN/GA хибрида, параметар комплексности модела вишеструке линеарне регресије у MLR/GA хибриду је подешен на $\lambda = 0$ (видети део 6.4.3).

Креирано је више базена потенцијалних регресора (ρ_I до ρ_{VIII}) који описују радијално померање тачке P1, на бази предиктора који се директно могу мерити (H , T_a , t , и d), као и главних компоненти f_1 , f_2 и f_3 , које су генерисане као резултат PCA редукције димензија. Концептуални опис контејнера је дат у табели 9.11, а форме регресора за поједине контејнере су дате у табели 9.12.

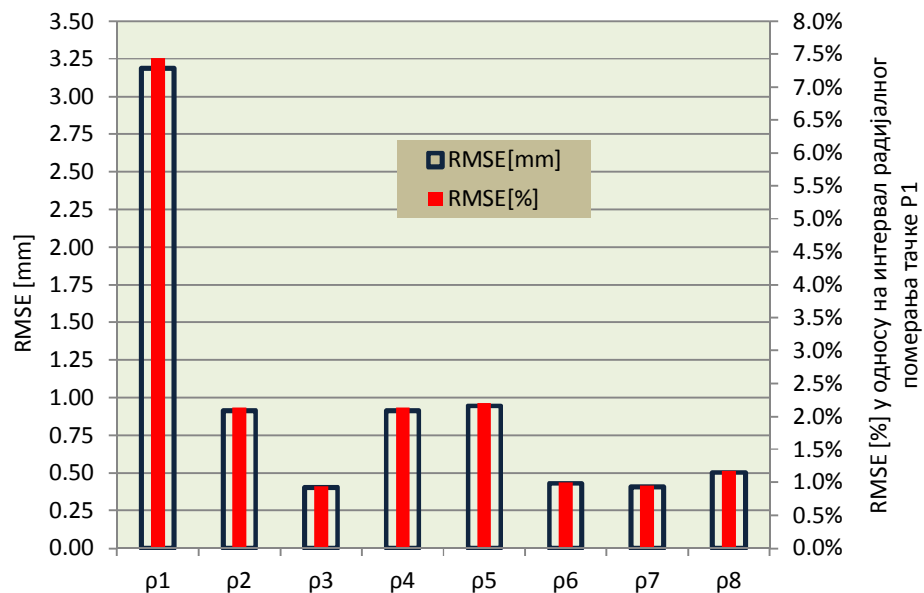
Табела 9.11. Опис контејнера базисних функција

Ознака контејнера	Опис контејнера	Број регресора
ρ_I	Обухвата утицај коте воде (H), средњу температуру ваздуха (T_a) и ефекте старења бране (t)	17
ρ_{II}	Обухвата контејнер ρ_I као и термичке ефекте, односно сезонска термичка оптерећења, помоћу тригонометријских функција променљиве d	25
ρ_{III}	Обухвата утицај коте воде (H), ефекте старења бране (t), сезонска термичка оптерећења, помоћу тригонометријских функција променљиве d и термичке ефекте изражене кроз главне компоненте (f_1, f_2, f_3) које обухватају 44 температуре бетона и средњу температуру ваздуха (T_a)	39
ρ_{IV}	Као ρ_{II} али је d замењен са фазним померајем d_{20}	25
ρ_V	Као ρ_{II} али је d замењен са фазним померајем d_{50}	25
ρ_{VI}	Као ρ_{III} али је d замењен са фазним померајем d_{20}	39
ρ_{VII}	Као ρ_{III} али је d замењен са фазним померајем d_{50}	39
ρ_{VIII}	Као ρ_{III} али без d	31

У базену ρ_I је разматран скуп од 17 потенцијалних регресора. Регресори су обухватили све расположиве променљиве у различитим формама, осим тригонометријских облика. Контејнер ρ_{II} је имао 25 регресора, који обухватају све променљиве у одабраним формама, укључујући и тригонометријске форме променљиве d . Контејнер базисних функција ρ_{III} је имао 39 регресора који су обухватили хидростатичке ефекте (H), старење (t), тригонометријске форме променљиве d и полиномијалне форме главних компоненти (f_1, f_2 и f_3). У циљу тестирања отпорности на температурни померај су реализоване додатне анализе, где је улазна променљива d мењана са *dumty* променљивим d_{20} и d_{50} (ρ_{IV} , ρ_V , ρ_{VI} и ρ_{VII}). Да би се испитала значајност променљиве d у односу на главне компоненте (f_1, f_2 и f_3), креиран је и контејнер ρ_{VIII} са 31. регресором, тако што су из базена ρ_{III} уклоњени тригонометријски регресори предиктора d .

У случајевима када у популационом базену није било регресора у тригонометријском облику (ρ_I) грешке модела су биле приближно око 3.2 mm (7.5% од опсега радијалног померања тачке P1) (слика 9.9). Након укључивања променљиве d и тригонометријских регресора (ρ_{II}), износ грешке $RMSE_{Test}$ је био редукован на приближно 0.92 mm, што представља износ од око 2.1% померања тачке P1. Коришћење *dumty* променљивих d_{20} и d_{50} уместо променљиве d није имало значајан утицај на износ грешке (ρ_{IV} и ρ_V у односу на ρ_{II}). Значајно побољшање тачности модела настаје увођењем главних компоненти у модел (ρ_{III}), тако да је $RMSE_{Test}$ приближно 0.4 mm, што представља 0.95% интервала радијалног померања тачке P1. Заменом предиктора d у моделима који одговарају контејнеру ρ_{III} , предикторима d_{20} и d_{50} (ρ_{VI} и ρ_{VII}), такође, није дошло до значајне промене квалитета модела ($RMSE_{Test} \approx 0.4$ mm). Када је из

контејнера ρ_{III} уклоњена променљива d , квалитет модела базиран на контејнеру ρ_{VIII} је пао за око 0.1 mm ($RMSE_{Test} \approx 0.5$ mm).



Слика 9.9. Просечне вредности $RMSE_{Test}$ за одабране *cross-validation* тестове и одабране контејнере базисних функција

Детаљни резултати, добијени коришћењем MLR/GA хибрида за одабране *cross-validation* тестове, су дати у табели 9.12.

Табела 9.12. Резултати моделирања радијалног померања тачке P1 бране Гранчарево (31. 8.2011-9.10.2013. год.), хибрид MLR/GA

Модели	Предик.	Почетни скуп регресора	LDS: TDS	Регресори одбачени после MLR-GA оптимизације	$R^2_{GA/adj}$	RMSE _{Train} [mm]	RMSE _{Test} [mm]
cross-validation LDS: TDS 60:40 (463:309) 70:30 (540:232) 75:25 (579:193) 80:20 (618:154)							
<i>a1</i>	H, T_a, t	$\rho_I:$ $\{H, H^2, H^3, \sqrt{H}, T_a, T_a^2, T_a^3, T_a^4 T_a^5, e^{-T_a}, e^{-T_a/2}, t, t^2, \sqrt{t}, e^{-t}, e^{-t/2}, e^{-t/4}\}$	60:40	$H^3, T_a^2, e^{-T_a}, e^{-T_a/2}, \sqrt{t}$	0.94345	2.96043	3.22595
до			70:30	$H^3, T_a, T_a^4 T_a^5, e^{-T_a}, e^{-T_a/2}, \sqrt{t}$	0.93925	3.16673	3.04076
<i>c3</i>			75:25	H^3, T_a, T_a^2, t, t^2	0.93925	3.01445	3.25765
			80:20	$H^3, T_a^2, e^{-T_a}, e^{-T_a/2}$	0.94553	3.02229	3.22028
<i>d1</i>	H, T_a, t, d	$\rho_{II}:$ $\{H, H^2, H^3, \sqrt{H}, T_a, T_a^2, T_a^3, T_a^4 T_a^5, e^{-T_a}, e^{-T_a/2}, t, t^2, \sqrt{t}, e^{-t}, e^{-t/2}, e^{-t/4}, \sin(2k \cdot \pi \cdot d / 365), \cos(2k \cdot \pi \cdot d / 365), k = 1, 4\}$	60:40	$H^2, H^3, \sqrt{t}, e^{-t}$	0.99539	0.81537	0.93630
до			70:30	$H^2, H^3, e^{-T_a}, e^{-T_a/2}, t^2, e^{-t/4}$	0.99474	0.87770	0.89456
<i>d2</i>			75:25	$H^2, H^3, e^{-T_a}, e^{-T_a/2}, t, \cos(6 \cdot \pi \cdot d / 365)$	0.99546	0.84074	0.95040
			80:20	$H^3, e^{-T_a/2}, t, e^{-t/2}$	0.99521	0.86382	0.87297
<i>e1</i>	H, f_1, f_2, f_3, t, d	$\rho_{III}:$ $\{H, H^2, H^3, \sqrt{H}, f_1, f_1^2, f_1^3, f_1^4, f_1^5, e^{-f_1}, e^{-f_1/2}, f_2, f_2^2, f_2^3, f_2^4, f_2^5, e^{-f_2}, e^{-f_2/2}, f_3, f_3^2, f_3^3, f_3^4, f_3^5, e^{-f_3}, e^{-f_3/2}, t, t^2, \sqrt{t}, e^{-t}, e^{-t/2}, e^{-t/4}, \sin(2k \cdot \pi \cdot d / 365), \cos(2k \cdot \pi \cdot d / 365), k = 1, 4\}$	60:40	$f_1^2, f_1^5, e^{-f_1}, f_3^2, f_3^4, \sqrt{t}, \sin(2 \cdot \pi \cdot d / 365)$	0.99903	0.38421	0.40736
до			70:30	$H^2, f_1^2, e^{-f_1}, f_2, f_2^4, e^{-f_2}, e^{-t/2}, \sin(8 \cdot \pi \cdot d / 365), \cos(4 \cdot \pi \cdot d / 365)$	0.99896	0.38027	0.39390
<i>e3</i>			75:25	$\sqrt{H}, f_2^3, f_2^5, e^{-f_2/2}, \sqrt{t}, e^{-t/2}, \cos(6 \cdot \pi \cdot d / 365)$	0.99896	0.37916	0.38869
			80:20	$H^2, f_1, f_2^5, e^{-f_2/2}, f_3, f_3^4, t, t^2, e^{-t/2}, \sin(8 \cdot \pi \cdot d / 365), \cos(4 \cdot \pi \cdot d / 365), \cos(6 \cdot \pi \cdot d / 365)$	0.99908	0.41560	0.41673
<i>f1</i>	H, T_a, t, d_{20}	$\rho_{IV}:$ $\{H, H^2, H^3, \sqrt{H}, T_a, T_a^2, T_a^3, T_a^4 T_a^5, e^{-T_a}, e^{-T_a/2}, t, t^2, \sqrt{t}, e^{-t}, e^{-t/2}, e^{-t/4}, \sin(2k \cdot \pi \cdot d_{20} / 365), \cos(2k \cdot \pi \cdot d_{20} / 365), k = 1, 4\}$	60:40	$H^3, \sqrt{H}, e^{-T_a}, \sqrt{t}, e^{-t}$	0.99530	0.81887	0.93518
до			70:30	$H, H^3, e^{-T_a}, t^2, \sqrt{t}$	0.99465	0.86460	0.88006
<i>f2</i>			75:25	$H^3, \sqrt{H}, t^2, \sqrt{t}$	0.99540	0.83735	0.95325
			80:20	$H^2, \sqrt{t}, e^{-t}$	0.99518	0.86046	0.88160

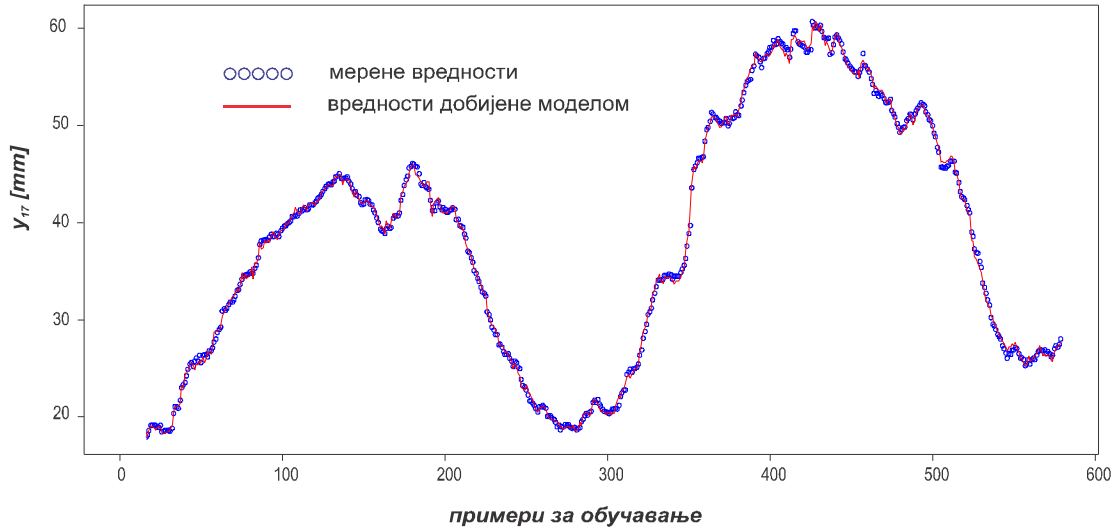
Табела 9.12. Резултати моделирања радијалног померања тачке P1 бране Гранчарево (31. 8.2011-9.10.2013. год.), хибрид MLR/GA (наставак)

Модели	Предик.	Почетни скуп регресора	LDS: TDS	Регресори одбачени после MLR-GA оптимизације	$R^2_{GA/adj}$ RegDesign	RMSE _{Train} [mm]	RMSE _{Test} [mm]
$f1l$ до $f2l$	$H, T_a,$ t, d_{50}	$\rho_V:$ $\{H, H^2, H^3, \sqrt{H}, T_a, T_a^2, T_a^3, T_a^4 T_a^5,$ $e^{-T_a}, e^{-T_a/2}, t, t^2, \sqrt{t}, e^{-t}, e^{-t/2}, e^{-t/4},$ $\sin(2k \cdot \pi \cdot d_{50} / 365), \cos(2k \cdot \pi \cdot d_{50} / 365), k = 1, 4\}$	60:40	$H, t^2, e^{-t/4}, \cos(4 \cdot \pi \cdot d_{50} / 365)$	0.99538	0.84866	0.97376
			70:30	$e^{-T_a}, t^2, e^{-t/2}, \cos(4 \cdot \pi \cdot d_{50} / 365)$	0.99471	0.89558	0.92893
			75:25	$H^2, \sqrt{H}, \sqrt{t}, e^{-t}$	0.99544	0.83833	0.95479
			80:20	$H^2, H^3, t, e^{-t/4}, \cos(4 \cdot \pi \cdot d_{50} / 365)$	0.99521	0.90235	0.92426
$g1$ до $g2$	$H, f_1,$ $f_2, f_3,$ t, d_{20}	$\rho_{VI}:$ $\{H, H^2, H^3, \sqrt{H}, f_1, f_1^2, f_1^3, f_1^4, f_1^5, e^{-f_1}, e^{-f_1/2},$ $f_2, f_2^2, f_2^3, f_2^4, f_2^5, e^{-f_2}, e^{-f_2/2}, f_3, f_3^2, f_3^3, f_3^4, f_3^5,$ $e^{-f_3}, e^{-f_3/2}, t, t^2, \sqrt{t}, e^{-t}, e^{-t/2}, e^{-t/4},$ $\sin(2k \cdot \pi \cdot d_{20} / 365), \cos(2k \cdot \pi \cdot d_{20} / 365), k = 1, 4\}$	60:40	$H^2, f_1, f_1^3, e^{-f_1}, f_2, e^{-f_2/2}, f_3^3, f_3^4,$ $\sin(2 \cdot \pi \cdot d_{20} / 365)$	0.99901	0.39025	0.40551
			70:30	$f_1^2, f_1^4, f_1^5, f_2, f_2^4, e^{-f_2}, f_3^2, t$	0.99927	0.39025	0.39462
			75:25	$H^3, f_1^2, f_1^3, f_1^5, t^2, e^{-t/2}, \sin(6 \cdot \pi \cdot d_{20} / 365)$	0.99904	0.44222	0.49489
			80:20	$H, f_1, f_1^4, f_1^5, e^{-f_1}, e^{-f_2}, e^{-t/2}$ $\sin(4 \cdot \pi \cdot d_{20} / 365), \sin(6 \cdot \pi \cdot d_{20} / 365)$	0.99907	0.38168	0.42426
$g1l$ до $g2l$	$H, f_1,$ $f_2, f_3,$ t, d_{50}	$\rho_{VII}:$ $\{H, H^2, H^3, \sqrt{H}, f_1, f_1^2, f_1^3, f_1^4, f_1^5, e^{-f_1}, e^{-f_1/2},$ $f_2, f_2^2, f_2^3, f_2^4, f_2^5, e^{-f_2}, e^{-f_2/2}, f_3, f_3^2, f_3^3, f_3^4, f_3^5,$ $e^{-f_3}, e^{-f_3/2}, t, t^2, \sqrt{t}, e^{-t}, e^{-t/2}, e^{-t/4},$ $\sin(2k \cdot \pi \cdot d_{50} / 365), \cos(2k \cdot \pi \cdot d_{50} / 365), k = 1, 4\}$	60:40	$H^3, f_1^2, f_1^3, f_1^5, e^{-f_1}, f_2, f_2^2, f_2^4, \sqrt{t},$ $\sin(2 \cdot \pi \cdot d_{50} / 365)$	0.99902	0.38272	0.39391
			70:30	$H^3, f_1^4, f_2, f_2^3, f_2^4, e^{-f_3}, e^{-f_3/2}, t, e^{-t/2},$ $\sin(4 \cdot \pi \cdot d_{50} / 365)$	0.99895	0.38868	0.39028
			75:25	$f_1^5, f_2, f_2^4, e^{-f_2}, t, e^{-t/2}$	0.99907	0.41227	0.43221
			80:20	$e^{-f_1/2}, f_2^3, e^{-f_3/2}, t^2, e^{-t/2},$ $\sin(4 \cdot \pi \cdot d_{50} / 365), \cos(4 \cdot \pi \cdot d_{50} / 365)$	0.99904	0.37834	0.40694
$h1$	$H, f_1,$ $f_2, f_3,$ t	$\rho_{VIII}:$ $\{H, H^2, H^3, \sqrt{H}, f_1, f_1^2, f_1^3, f_1^4, f_1^5, e^{-f_1}, e^{-f_1/2},$ $f_2, f_2^2, f_2^3, f_2^4, f_2^5, e^{-f_2}, e^{-f_2/2}, f_3, f_3^2, f_3^3, f_3^4, f_3^5,$ $e^{-f_3}, e^{-f_3/2}, t, t^2, \sqrt{t}, e^{-t}, e^{-t/2}, e^{-t/4}\}$	60:40	$H^2, f_2^2, f_2^4, f_2^5, e^{-f_2/2}, f_3^5, \sqrt{t}$	0.99866	0.48053	0.48883
			70:30	$f_1^3, f_2, f_2^4, e^{-f_2}, f_3^2, \sqrt{t}$	0.99869	0.48264	0.48704
			75:25	$H^2, f_2, e^{-f_2}, f_3, f_3^2$	0.99875	0.49132	0.49643
			80:20	$H^3, f_1^4, f_2^3, f_3^2, f_3^5, e^{-f_3}, \sqrt{t}$	0.99891	0.46292	0.53906

Најбољи од свих оптимизованих модела је био модел са грешком $RMSE_{Test} = 0.38869$ mm, који је добијен за однос $LDS:TDS=75:25$, скуп предиктора $\{H, f_1, f_2, f_3, t, d\}$ и 32 регресора из популационог базена ρ_{III} (9.2):

$$\begin{aligned}
 y_{17} = & (2.2070E+07) + (5.3650E+01) \cdot H - (1.5540E-01) \cdot H^2 + (1.5000E-04) \cdot H^3 \\
 & + (4.5730E+03) \cdot f_1 - (1.1000E+03) \cdot f_1^2 + (1.6840E+02) \cdot f_1^3 - (1.6640E+01) \cdot f_1^4 \\
 & + (7.1670E-01) \cdot f_1^5 - (1.8410E+02) \cdot (e^{-f_1}) + (9.5250E+03) \cdot (e^{-f_1/2}) \\
 & - (5.4870E+00) \cdot f_2 + (4.1540E+00) \cdot f_2^2 - (4.1790E-01) \cdot f_2^4 - (7.8910E-01) \cdot (e^{-f_2}) \\
 & + (5.7340E+04) \cdot f_3 - (1.3920E+04) \cdot f_3^2 + (2.1830E+03) \cdot f_3^3 - (2.3920E+02) \cdot f_3^4 \\
 & + (1.5780E+01) \cdot f_3^5 - (1.6630E+03) \cdot (e^{-f_3}) + (1.1800E+05) \cdot (e^{-f_3/2}) \\
 & + (1.0560E-01) \cdot \sin(2 \cdot \pi \cdot d / 365) + (3.3160E-01) \cdot \sin(4 \cdot \pi \cdot d / 365) \\
 & + (4.2630E-01) \cdot \sin(6 \cdot \pi \cdot d / 365) - (7.2720E-02) \cdot \sin(8 \cdot \pi \cdot d / 365) \\
 & + (3.3070E+00) \cdot \cos(2 \cdot \pi \cdot d / 365) + (3.3790E-01) \cdot \cos(4 \cdot \pi \cdot d / 365) \\
 & + (4.7120E-01) \cdot \cos(8 \cdot \pi \cdot d / 365) - (2.0040E+06) \cdot t + (4.6900E+04) \cdot t^2 \\
 & + (3.9360E+11) \cdot e^{-t} - (1.1950E+08) \cdot e^{-t/4}
 \end{aligned} \tag{9.2}$$

На слици 9.10, су приказане мерене и моделиране вредности за LDS који се састојао од 579 примера за обучавање.



Слика 9.10. Мерене (плаво) и моделске (црвено) вредности радијалног померања тачке P1 за скуп података за обучавање LDS (579 записа), за однос $LDS:TDS=75:25$, који одговара регресионом моделу највеће тачности

Оптимизације MLR модела трајале су у просеку 5 до 10 минута.

У табели 9.12, у првој колони са леве стране, су уведене ознаке добијених MLR регресионих модела, због упоредивости са моделима који су описани у наредном делу, где су приказани модели генерисани помоћу ANN/GA хибрида.

9.4.3. Дискусија резултата моделирања заснованог на хибриду MLR/GA

Резултати и модели који су базирани на контејнеру ρ_I , и $RMSE_{Test}$ грешка од око 7.5% (3.2 mm), указују да средња дневна температура, коју карактерише велики шум и расипање вредности, има значајан утицај на релативно лошији квалитет добијеног модела. Увођењем променљиве d посредно се узима у обзир цикличност у промени

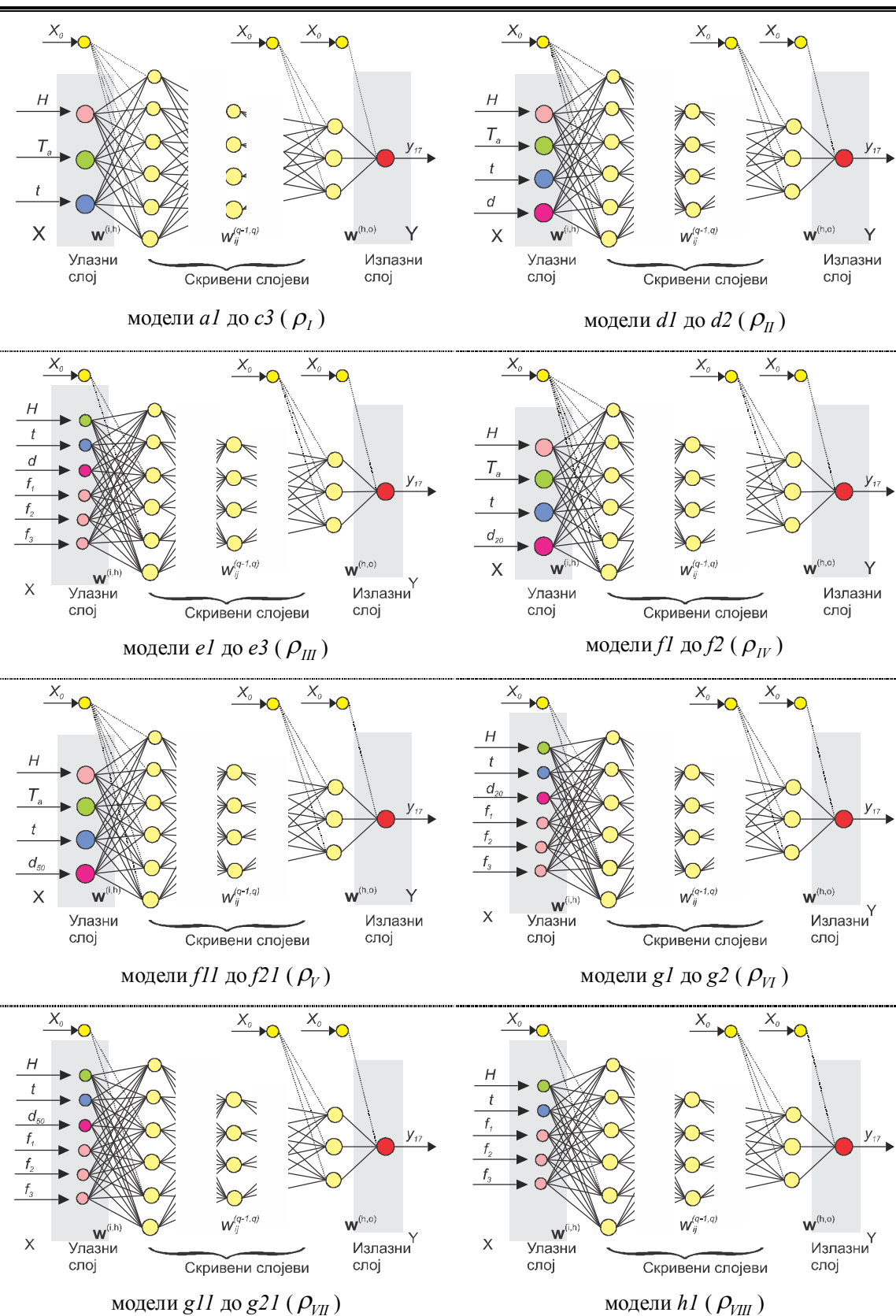
температурног поља у телу бране током календарске године. Квалитет регресионих модела се унапређује до 3.5 пута ($RMSE_{Test} \approx 0.92$ mm, 2.1% интервала радијалног померања тачке P1).

Као што је показано у поглављу 7, MLR регресиони модели су отпорни на фазни померај d_{20} и d_{50} . Такав закључак се потврђује и у оквиру овог истраживања, јер је $RMSE_{Test}$ за моделе базиране на регресорима из контејнера ρ_{II} , ρ_{IV} и ρ_V , сличан и износи између 0.9 и 9.95 mm. За разлику од променљиве d , која посредно одсликава утицај температурних ефеката, главне компоненте (f_1 , f_2 и f_3) представљају линеарну комбинацију директних мерења температура бетонске структуре. Отуда се добија и максимално побољшање квалитета добијених регресионих модела креираних над скупом ρ_{III} ($RMSE_{Test} \approx 0.4$ mm, што представља 0.95% интервала радијалног померања тачке P1). Отпорност на температурни *offset* је потврђена и за моделе креиране над контејнерима ρ_{VI} и ρ_{VII} . Модели креирани над контејнером ρ_{III} обухватају директне (44 позиције за температуре бетонске структуре бране и T_a) и индиректне температурне утицаје (d). Другим речима, променљива d обухвата, поред осталих, и термичке утицаје које не обухватају уграђени сензори за мерење температуре бетонске структуре бране. Због тога је, након уклањања тригонометријских регресора базираних на предиктору d , квалитет добијених регресионих модела нешто лошији (око 0.5 mm уместо 0.4 mm).

9.4.4. Параметри и резултати добијени на основу ANN/GA хибридне методе

Валидација ANN/GA методологије за брану Гранчарево је реализована помоћу DEVONNA рационалног агента, чије су карактеристике и параметри детаљно описани у поглављу 7. Због упоредивости је *cross-validation* концепт реализован над истим скупом података и над истим скуповима за учење и тестирање као и при примени MLR/GA хибрида. За све моделе, осим моделе са ознакама a и b , параметри GA за ANN/GA хибрид су били: број јединки у популационом базену 50, а број генерација 40. За селекцију је одабрана бинарна турнирска селекција, а за репродукцију укрштање у једној тачки и униформна мутација са вероватноћама од 0.9 и 0.125, респективно.

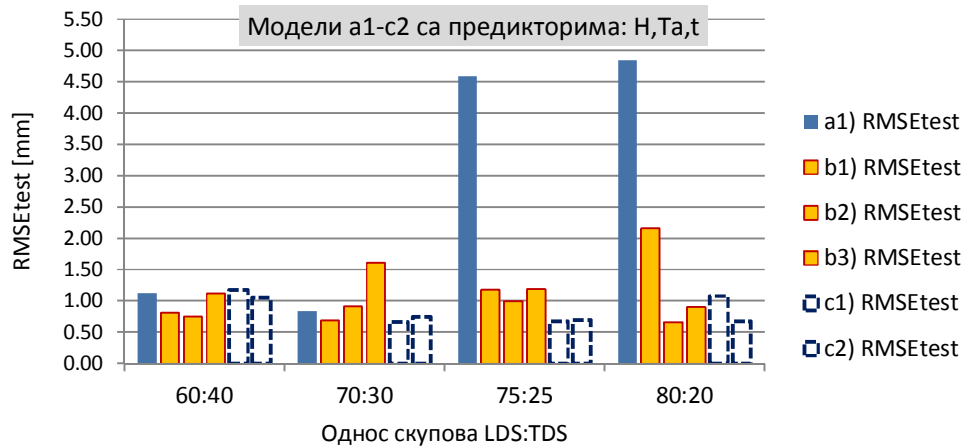
Код MLR/GA хибрида различити скупови предикторских променљивих су генерисали различите контејнере базисних функција. Код ANN/GA хибрида различити скупови предикторских променљивих представљају променљиве на улазном слоју мреже (слика 9.11). Структура хромозома која је коришћена за дефинисање архитектуре мрежа, избор активационих функција, алгоритама учења и параметара учења, формирана је у складу са табелом 7.1 (део 7.5.1).

Слика 9.11. Предиктори за ANN јединке које одговарају контејнерима од ρ_I - ρ_{VIII}

Иницијалне вредности тежинских коефицијената $\mathbf{W}_0$ су биране на случајан начин из интервала $[-0.5, 0.5]$. Да би се умањио утицај стохастичког избора матрице $\mathbf{W}_0$, процес учења је за сваку ANN јединку понављан пет пута ($R = 5$) (видети слику 7.5).

Моделирање радијалног померања тачке P1 је реализовано за сваки LDS-TDS пар и различите скупове предикторских променљивих, а резултати су дати у табели 9.13. Просечно трајање процеса генерисања модела варијало је од 45 до 85 мин.

Пошто су модели са ознакама a (10 генерација и 15 јединки) и b (15 генерација и 20 јединки), показали релативну нестабилност по питању квалитета (слика 9.12), већина модела (модел са ознакама c, d, g, f и h) је креирана кроз GA итеративну процедуру која је подразумевала 40 генерација и 50 јединки у популационом базену.



Слика 9.12. Нестабилност квалитета ANN/GA модела за мали број генерација и јединки

Табела 9.13. Резултати моделирања радијалног померања тачке P1 бране Гранчарево (31.8.2011-9.10.2013. год.),
хибрид ANN/GA

ANN/GA	RMSE [mm]			RMSE [mm]			RMSE [mm]			RMSE [mm]			RMSE [mm]		
Cross-validation	предиктори		модел	предиктори		модел	предиктори		модел	RMSE [mm]		модел	RMSE [mm]		
LDS:TDS однос	Train	Test		Train	Test		Train	Test		Train	Test		Train	Test	
	H, T_a, t			H, T_a, t, d			H, d, t, f_1, f_2, f_3			H, T_a, t, d_{20}			H, T_a, t, d_{50}		
60-40 (463:309)	a1	0.9544	1.1281	d1	0.5096	0.5729	e1	1.8674	1.9230	f1	0.6059	0.8923	f11	0.8159	0.9406
	b1	0.7349	0.8155	d1	0.4756	0.5589	e2	1.0728	1.1124	f2	0.6650	0.9091	f21	0.4912	0.5490
	b2	0.6488	0.7450				e3	1.2572	1.3319						
	b3	0.9761	1.1228												
	c1	1.2940	1.1713												
	c2	0.9301	1.0439												
	c3	0.5658	0.7052												
70-30 (540:232)	a1	1.7321	1.8189	d1	0.8358	0.9233	e1	1.8355	1.9216	f1	0.9944	1.0025	f11	0.6106	0.6923
	b1	0.6064	0.6877	d2	1.6273	1.7300	e2	0.3387	0.4023	f2	0.8716	1.7283	f21	0.4889	0.5401
	b2	0.8944	0.9090	d3	0.6617	0.8023	e3	0.9305	0.9999						
	b3	1.5947	1.6118												
	c1	0.5881	0.6662												
	c2	0.6562	0.7366												
75-25 (579:193)	a1	4.2113	4.5877	d1	0.66174	0.80232	e1	0.37725	0.40339	f1	0.8238	1.0172	f11	0.7271	0.8343
	b1	1.1787	2.4672	d2	0.70603	1.20915	e2	0.37836	0.43435	f2	1.0398	1.1843	f21	0.6752	0.7519
	b2	0.9918	1.0977	d3	1.85720	1.95907	e3	0.56524	1.47061						
	b3	1.1860	1.3928												
	c1	0.6623	0.6737												
	c2	0.5817	0.6900												
80-20 (618:154)	a1	4.5001	4.8433	d1	0.4987	0.5348	e1	0.3983	0.3939	f1	0.5205	0.5810	f11	0.4956	0.5238
	b1	2.1629	2.3496	d2	0.7947	1.0072	e2	0.5074	0.5745	f2	0.5649	1.3728	f21	1.0926	1.0370
	b2	0.6560	0.7317				e3	0.4757	0.5002						
	b3	0.8956	1.2072												
	c1	0.9441	1.0743												
	c2	0.6891	0.6734												

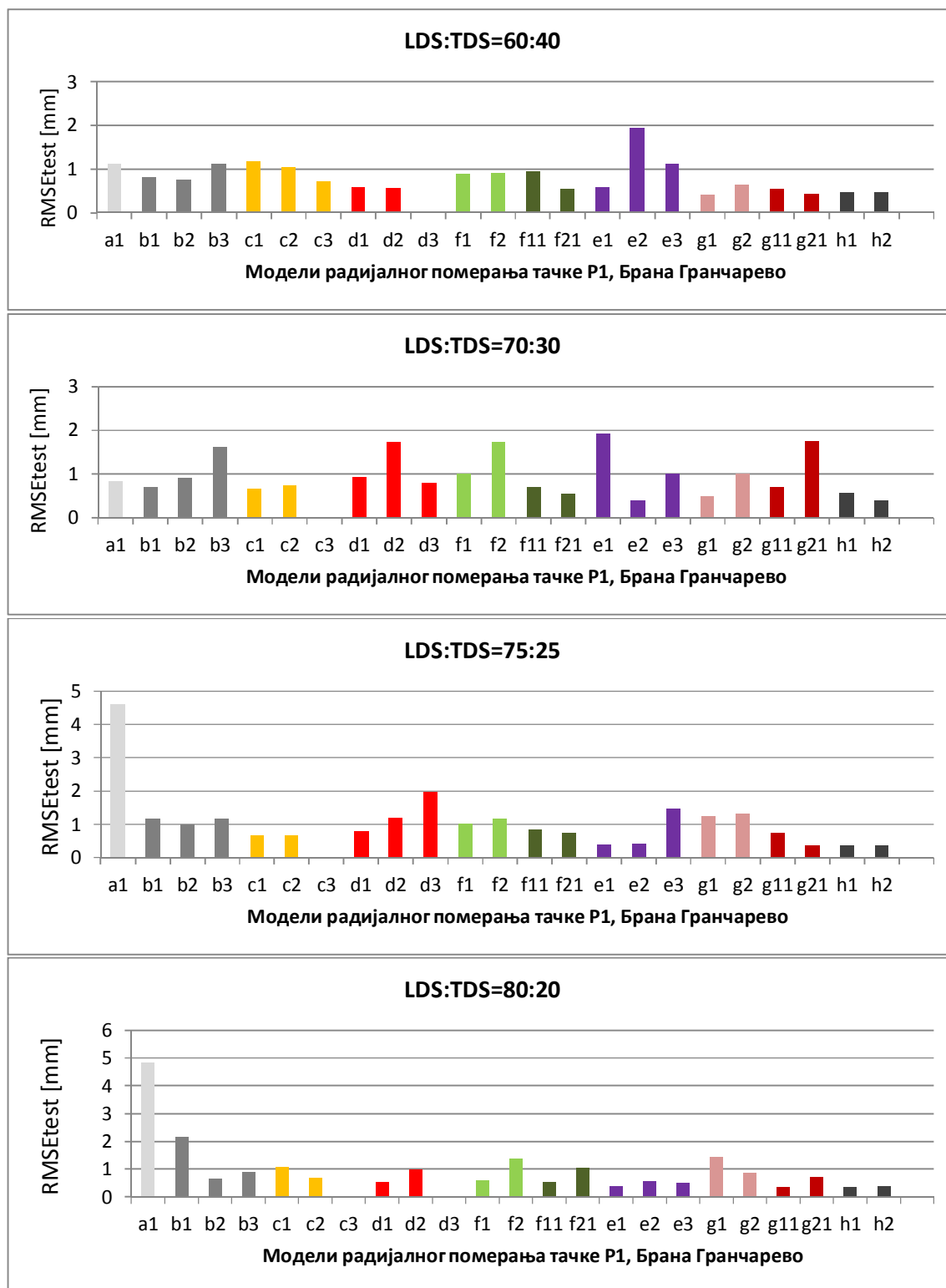
Table 9.13. Резултати моделирања радијалног померања тачке P1 бране
Гранчарево (31.8.2011-9.10.2013. год.), хибрид ANN/GA (наставак)

ANN/GA	модел	RMSE [mm]		модел	RMSE [mm]		модел	RMSE [mm]	
Cross-validation		Предиктори			Предиктори				
		Train	Test		Train	Test		Train	Test
LDS:TDS однос		$H, d_{20}, t, f_1, f_2, f_3$	$H, d_{50}, t, f_1, f_2, f_3$		H, t, f_1, f_2, f_3				
60-40 (463:309)	g1	0.3620	0.4024	g11	0.4458	0.5524	h1	0.3526	0.4637
	g2	0.4671	0.6290	g21	0.3947	0.4176	h2	0.3629	0.4591
70-30 (540:232)	g1	0.3812	0.4870	g11	0.5575	0.6979	h1	0.4020	0.5545
	g2	0.5851	0.9949	g21	1.6815	1.7523	h2	0.3630	0.3900
75-25 (579:193)	g1	0.9893	1.2543	g11	0.5248	0.7409	h1	0.3378	0.3787
	g2	1.2586	1.3208	g21	0.3634	0.3722	h2	0.3864	0.3683
80-20 (618:154)	g1	1.2321	1.4295	g11	0.3541	0.3748	h1	0.3754	0.3723
	g2	0.6386	0.8623	g21	0.5679	0.7390	h2	0.3542	0.3818

Вредност $RMSE_{Test}$ за моделе (c) са улазним променљивим H , T_a и t је била приближно 0.85 mm (око 2% од опсега померања тачке P1) (табела 9.13 и слика 9.13).

Након укључивања променљиве d (дани у години), квалитет добијених модела (модел d) је варирао од оних који су имали већу тачност ($RMSE_{Test} \approx 0.57$ mm) ($d1, d2$ за LDS:TDS=60:40), до оних са падом тачности ($RMSE_{Test} \approx 1.43$ mm) ($d2$ за LDS:TDS=70:30 и $d3$ за LDS:TDS=75:25, респективно). Увођењем у моделе променљивих које одсликавају главне компоненте (модел e), генерисани су још тачнији модели ($e1, e2$ за LDS:TDS=75:25), али се јављала и велика нестабилност ($e2$ за LDS:TDS=60:40 као и $e1$ за LDS:TDS=70:30, респективно). У тестовима са *dummy* променљивим $d20$ и $d50$ потврђено је да је ANN/GA хибрид осетљив на температурни *offset* (модел f у односу на моделе са ознакама d , као и модели са ознакама g у односу на моделе са ознакама e). ANN/GA модели у којима су поред предиктора H и t биле и главне компоненте (f_1, f_2, f_3), али без предиктора d , су имали приближну тачност $RMSE_{Test} \approx 0.42$ mm (око 1% од опсега померања тачке P1).

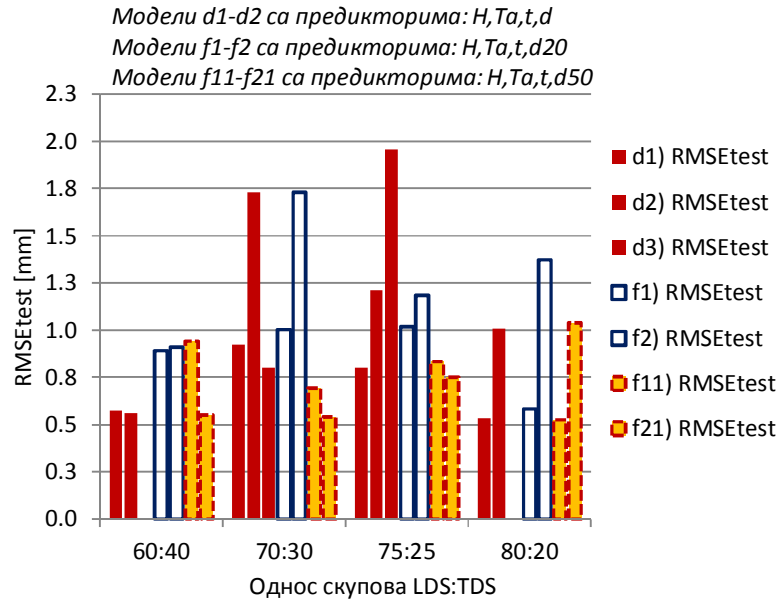
У више од 95% тестова, неуронске мреже са најбољим карактеристикама су имале *hyperbolic tangent* активациону функцију (TANH). Модел вештачке неуронске мреже највеће тачности ($RMSE_{Test} = 0.3683$ mm, 0.85% од опсега померања тачке P1) је добијен за предикторе H , t , f_1 , f_2 , и f_3 са 75:25, LDS:TDS односом. Оптимизовани мрежни модел има 4 скривена слоја, са по 30 неурона у сваком слоју. Активациона функција је била TANH, док је за алгоритам учења за ову мрежу систем определио RPROP алгоритам са параметрима $\eta^+ = 1.07$ и $\eta^- = 0.72$. Оптимизациона процедура је трајала око 65 минута.



Слика 9.13. Приказ квалитета добијених ANN/GA модела

9.4.5. Дискусија резултата моделирања заснованог на хибриду ANN/GA

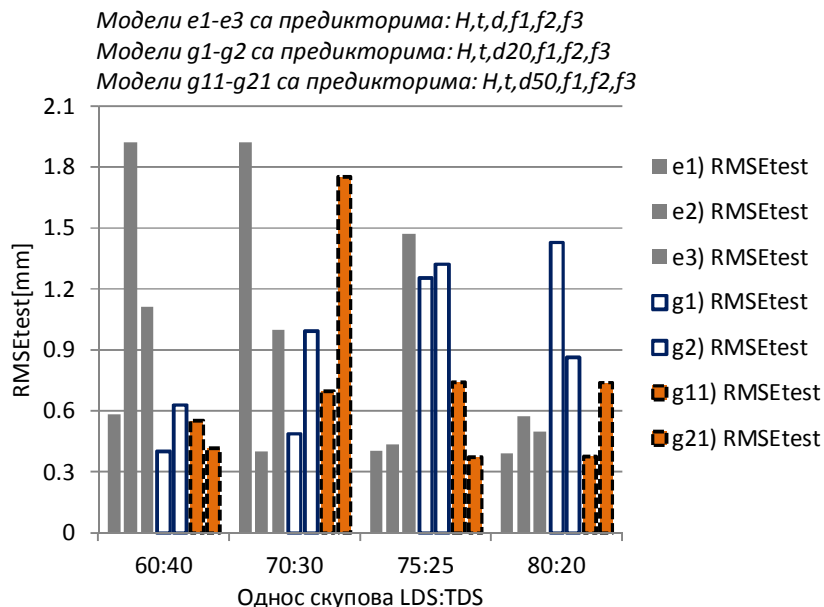
Просечна $RMSE_{Test}$ грешка за скуп предиктора H , T_a , и t од око 2% (0.85 mm), указује да средњом дневном температуром (T_a), коју карактерише велики шум и расипање вредности, ANN модели могу да овладају, због своје природе и унутрашње структуре. На бази модела у којима су поступно укључиване променљива d (модел d), променљиве f_1 , f_2 и f_3 (модел e), и *dumt*у променљиве $d20$ и $d50$ (модел f и g), се могу извести закључци о неспособности ANN/GA хибрида да генерише стабилне моделе као и закључци о његовој неотпорности на температурни *offset* (слика 9.14, слика 9.15).



Слика 9.14. Приказ нестабилности и неотпорности на температурни *offset* модела генерисаних помоћу ANN/GA хибрида (модел d и f)

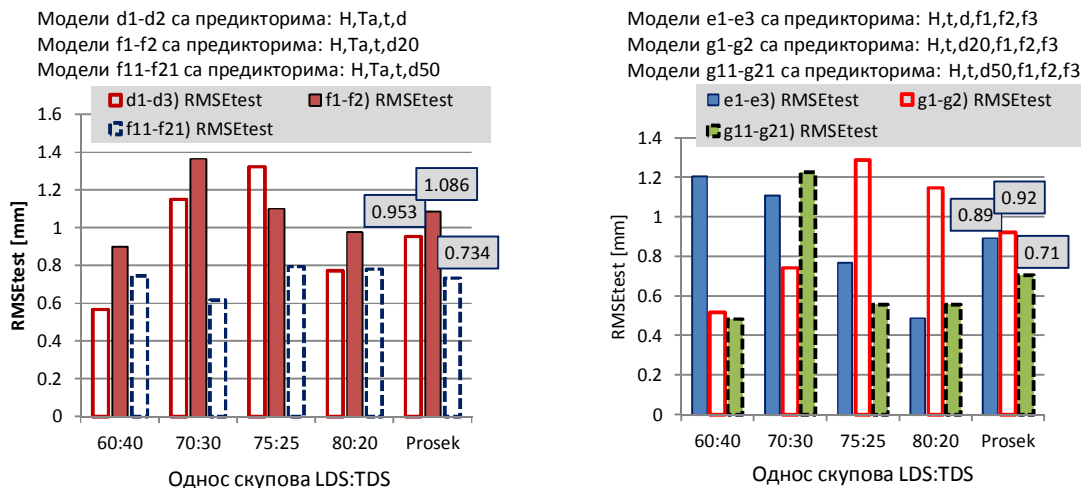
Са слике 9.15 и слике 9.16 се уочава изражени недостатак ANN/GA хибрида да, за различите односе LDS:TDS, генерише моделе стабилног квалитета (модел d , f_1 , f_2 , e , g). Такође се уочава и јасна промена у квалитетима модела без главних компоненти (модел f) у које је укључен фазни померај температуре од 20 дана (модел f_1 , f_2) и фазни померај од 50 дана (модел f_{11} , f_{21}) (слика 9.15). Слично се уочава и јасна промена у квалитетима модела са главним компонентама (модел g) за $d20$ (модел g_1 , g_2) и $d50$ (модел g_{11} , g_{21}) (слика 9.16).

Поред уграђених механизма, помоћу којих се жели елиминисати утицај стохастичности при иницијализацији тежинских коефицијената и примене *cross-validation* стратегије са циљем смањења стохастике у процесу креирања хромозома, неопходно је даље радити на питањима стабилности квалитета генерисаних ANN/GA модела. Повећање броја генерација и броја јединки у GA је један од могућих праваца, али је овај правац ограничен генерисањем оптималног модела у разумном времену.



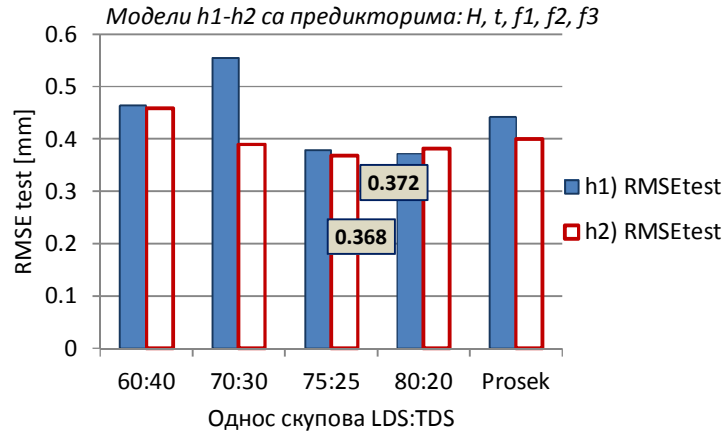
Слика 9.15. Приказ нестабилности и неотпорности на температурни *offset* модела генерисаних помоћу ANN/GA хибрида (модели *e* и *g*)

Разлика у квалитету ANN/GA модела, са аспекта температурног *offset*-а, се види и на дијаграмима просечних вредности грешака $RMSE_{Test}$ за ANN/GA моделе *d* и *f* (0.953 mm; 1.086 mm; 0.734 mm), као и моделе *e* и *g* (0.89 mm; 0.92 mm; 0.71 mm) (слика 9.16).



Слика 9.16. Просечне вредности грешака $RMSE_{Test}$ за ANN/GA моделе *d*, *f*, *e* и *g*

Ипак, ANN/GA хибриди могу генерисати и моделе високе тачности (слика 9.17), као на пример: $RMSE_{Test} \approx 0.368$ mm, за однос LDS:TDS=75:25, или $RMSE_{Test} \approx 0.372$ mm, за однос LDS:TDS=80:20, што представља 0.85%, односно 0.87% од опсега померања тачке P1, респективно. Наведени модели високе тачности су добијени за предикторе H, t, f_1, f_2 и f_3 (модел *h*).

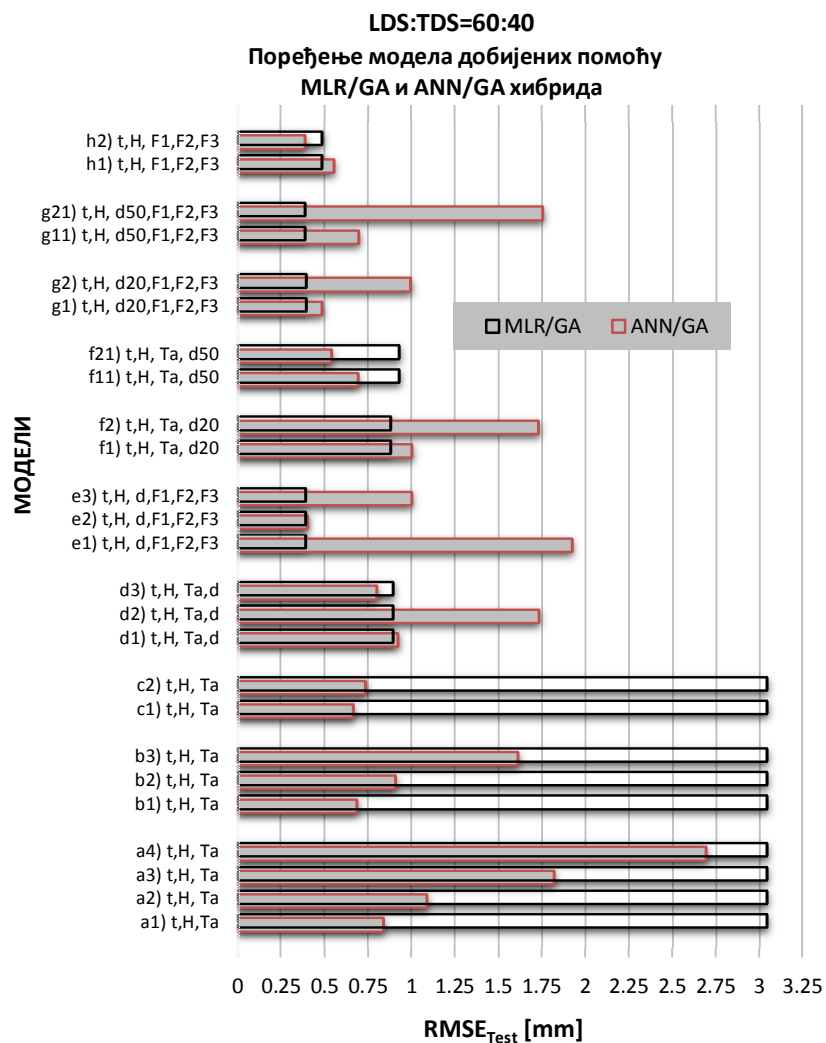


Слика 9.17. Најтачнији ANN/GA модели

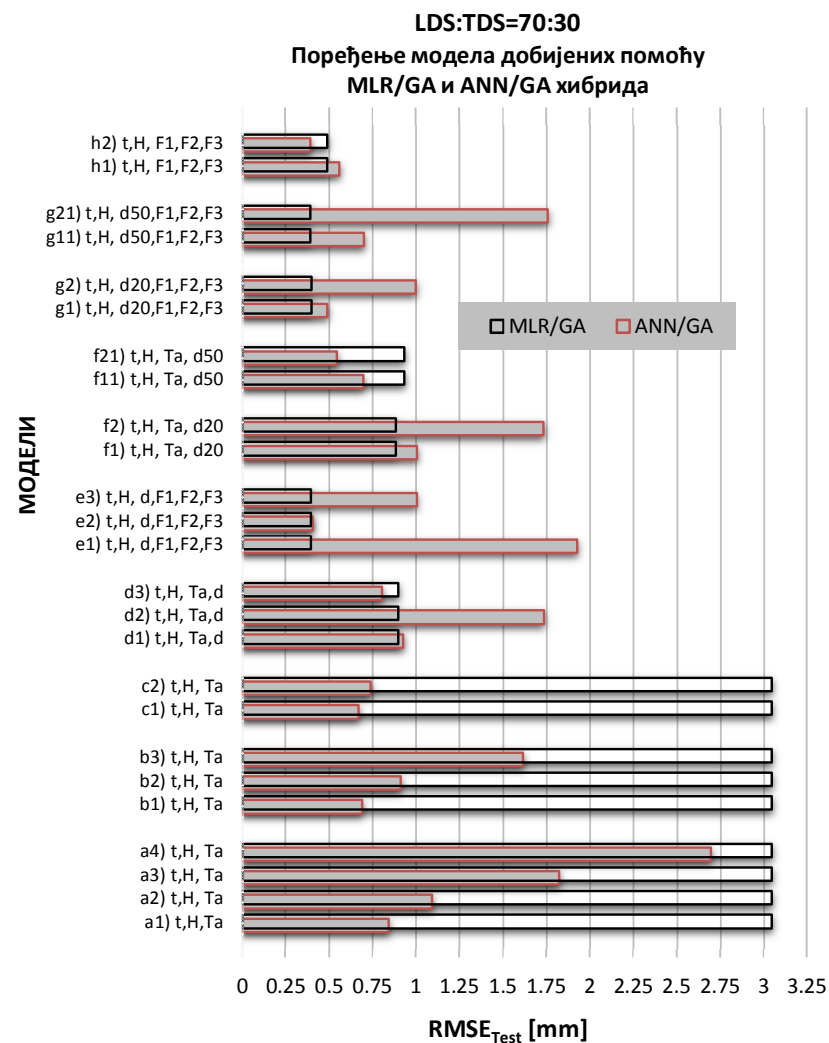
9.5. Компаративна анализа MLR/GA и ANN/GA генерисаних хибридних модела

У оквиру овог дела се даје упоредни приказ MLR/GA и ANN/GA генерисаних хибридних модела (слика 9.18, слика 9.19, слика 9.20, слика 9.21).

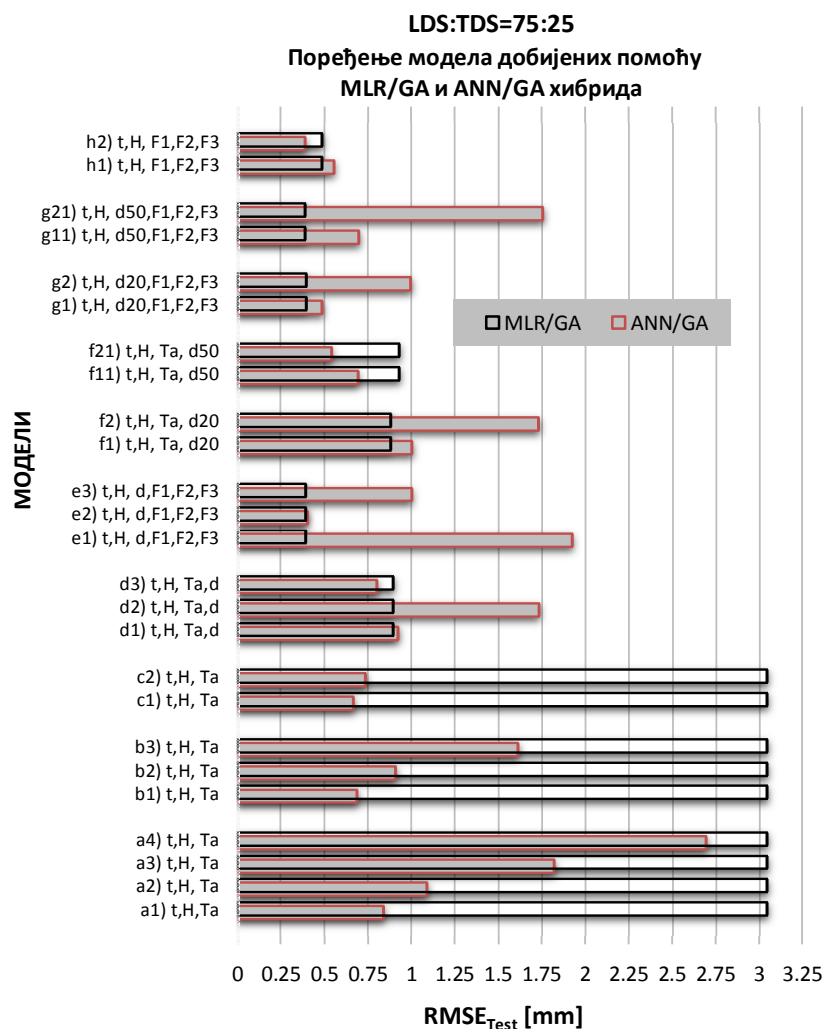
У моделима означеним од a до c , који су базирани на регресорима креираним над контејнером ρ_l за MLR/GA, односно предикторима H, T_a и t за ANN/GA, иако нестабилног квалитета, модели генерисани са ANN/GA хибридом показали су се квалитетнијим од модела генерисаних са хибридом MLR/GA. Једино одступање је код ANN/GA модела a (10 генерација и 15 јединки), што се може објаснити малим бројем генерација и хромозома у GA процедури. Крећући се ка моделима c , повећањем броја генерација и броја хромозома (40 генерација и 50 јединки), добијени ANN модели постају значајно квалитетнији, а вредности грешака $RMSE_{Test}$ се стабилизују на интервалу 0.5-0.75 mm. У овим случајевима долази до изражаја унутрашња природа и робустност ANN у моделирању процеса са значајним шумом. Регресиони MLR/GA модели из ове групе имају релативно малу тачност, јер су MLR осетљиве на изражену стохастичку података, тако да је $RMSE_{Test} \approx 3.2$ mm (7.5% од опсега радијалног померања тачке P1).



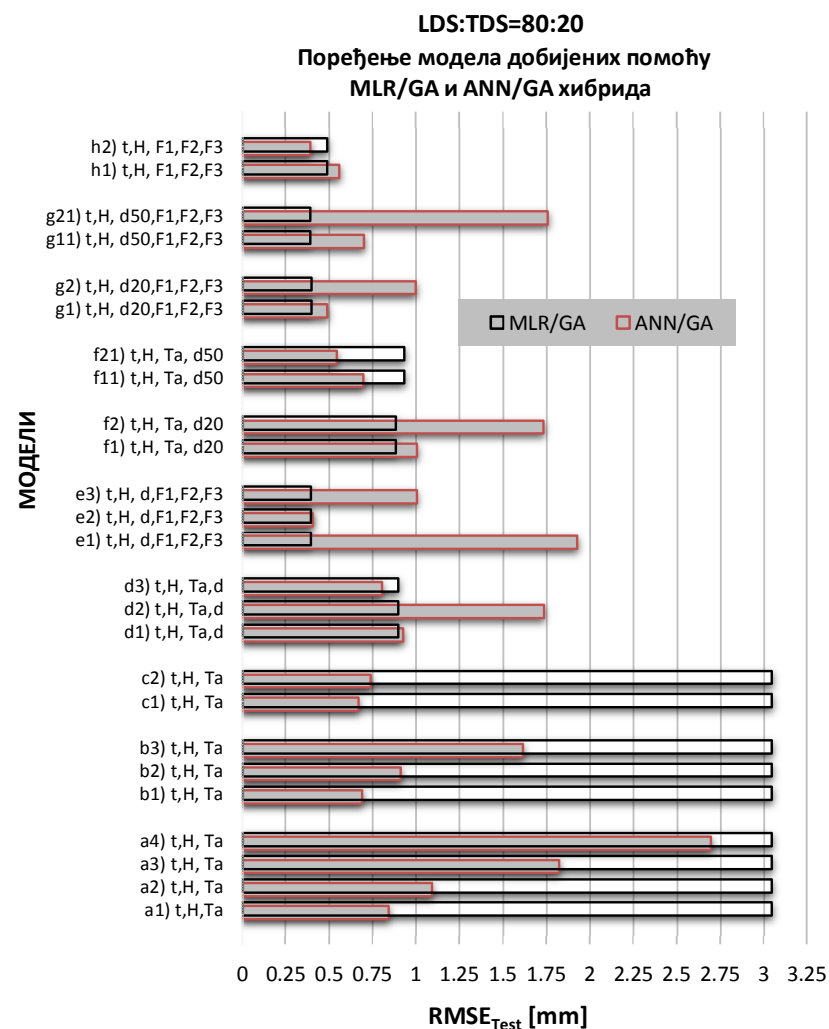
Слика 9.18. Квалитет модела за LDS:TDS=60:40



Слика 9.19. Квалитет модела за LDS:TDS=70:30



Слика 9.20. Квалитет модела за LDS:TDS=75:25



Слика 9.21. Квалитет модела за LDS:TDS=80:20

Код модела d , променљива која репрезентује дане у години (d), значајно обухвата цикличну природу феномена температурног поља бетонске бране, тако да се грешка за MLR моделе генерисане са MLR/GA стабилизује на вредности од око $RMSE_{Test} \approx 0.9$ mm. Ово је последица присуства тригонометријских регресора у контејнерима ρ_{II} , ρ_{IV} и ρ_V . Сличан квалитет MLR модела (f) се добија и након температурног *offset*-а (d_{20}, d_{50}), што је последица способности тригонометријских регресора да моделирају фазни померај суперпонирањем синусних и косинусних функција. ANN моделе d и f са предикторима H, T_a, t, d односно H, T_a, t, d_{20} или H, T_a, t, d_{50} , карактерише нестабилан ниво квалитета, односно променљиве вредности $RMSE_{Test}$ у интервалу од 0.6-1.5 mm. Дакле, ANN/GA хибрид може генерисати моделе бољег квалитета, али такође и моделе који су значајно лошији од MLR модела. Такође, ови модели не поседују отпорност на температурни *offset*.

Иста запажања важе и за моделе код којих су у разматрање уведене главне e компоненте (f_1, f_2 и f_3) (модели e , контејнер ρ_{III}) и додатно *dumty* променљиве d_{20} и d_{50} (модели g , контејнери ρ_{VI} и ρ_{VII}). Овде се може уочити да је квалитет модела унапређен са $RMSE_{Test} \approx 0.9$ mm на $RMSE_{Test} \approx 0.4$ mm. Без обзира на висок ниво тачности у неким тестовима, код ANN модела је поново присутна нестабилност.

Већ је објашњено да се уклањањем регресора који потичу од предиктора d , а задржавањем главних компоненти f_1, f_2 и f_3 (модели h), квалитет MLR модела незнатно умањује (са $RMSE_{Test} \approx 0.4$ mm на $RMSE_{Test} \approx 0.5$ mm), јер се тиме елиминише део цикличних термичких ефеката, који нису обухваћени вредностима средње температуре ваздуха и вредностима мереним са 44 термометра у телу бране.

Максимални квалитет једне и друге врсте модела је упоредив: $RMSE_{Test} = 0.38869$ mm, (0.87% од опсега померања тачке P1) за MLR моделе и $RMSE_{Test} = 0.3683$ mm, (0.85% од опсега померања тачке P1) за ANN моделе. Са друге стране, време које је потребно за генерисање ових модела је неупоредиво (5-10 минута за MLR моделе, насупротив времену од 45-85 min., које је потребно за генерисање ANN модела).

10. Закључак и правци даљег истраживања

10.1. Закључак

Већина постојећих регресионих метода моделирања претпоставља временску непроменљивост моделираних објеката и захтева сталан скуп улазних параметара. У реалним апликацијама, сталне промене објеката и откази мерне опреме могу довести до ситуација у којима усвојени модел постаје неупотребљив. Из тог разлога је неопходно развити методе и системе за аутоматско генерисање што адекватнијих модела за дату ситуацију. У оквиру ове дисертације су развијене две хибридне методе које нуде део решења за наведене проблеме. Обе методе развоја и адаптације регресионих модела су засноване на генетским алгоритмима.

Након дефинисања циља, претпоставки и анализе стања у области истраживања, у дисертацији је приказана теоријска подлога о математичким моделима, регресијама, вештачким неуронским мрежама и генетским алгоритмима. У поглављима бр. 6 и бр. 7 су разрађене и примењене MLR/GA и ANN/GA хибридне методе и развијени одговарајући софтверски агенти. Перформансе софтверских агената су валидоване кроз студије случаја брана Бочац и Гранчарево, које се налазе у Републици Српској.

У поглављу 6 је развијена и приказана MLR/GA хибридна метода, која за разлику од постојећих метода, омогућава генерисање адаптивних модела отпорних на променљивост скупа улазних варијабли и променљивост скупа измерених вредности. Коришћењем генетских алгоритама, метода оптимизује модел вишеструке линеарне регресије у складу са тренутно активним сензорима и актуелним подацима. MLR модел је, за дате услове, оптимизован по два критеријума: критеријуму тачности и критеријуму комплексности. За мерење тачности модела је коришћен прилагођени коефицијент вишеструке детерминације, док је укупни број регресора коришћен као мера комплексности модела. Развијена MLR/GA метода је имплементирана у виду GenReg софтверског агента, чије су перформансе тестиране у поступку моделирања радијалног померања одабраних тачака бетонске бране Бочац, на реци Врбас у Републици Српској. GenReg агент је генерисао адаптивни регресиони модел радијалног померања одабране тачке бране у функцији нивоа воде у акумулацији, падавина, температуре и времена које је протекло од почетка експлоатације бране. Према развијеном систему, сваки пут када неки сензор постане активан или неактиван, GenReg агент, на основу MLR/GA методе, прилагођава регресиони модел и усклађује га са скупом активних сензора, односно података које ови сензори прикупљају. Модели генерисани коришћењем MLR/GA методе су у случају отказа појединих сензора показали значајно бољу способност за предикцију у односу на MLR моделе који подразумевају сталан скуп улазних променљивих. Додатно, хибридна метода је показала способност да, при генерисању модела, одбацује предикторе који нису од значаја за описивање посматраног објекта. На основу приказаних перформанси у студији случаја бране Бочац, се може закључити да MLR/GA метода и развијени GenReg софтверски агент чине моћни и робустни статистичко-софтверски оквир за математичко моделирање, мониторинг, предикцију и управљање објектима, процесима и системима. Развијена MLR/GA хибридна метода је потврдила прве две претпоставке, које су дефинисане на почетку истраживања у оквиру ове дисертације.

У поглављу 7 је развијена ANN/GA хибридна метода и DEVONNA софтверски агент за моделирање објеката и система у форми самоподешавајућих еволуционих вештачких неуронских мрежа (ANN). Коришћењем генетских алгоритама ANN/GA метода оптимизује структуру и параметре неуронске мреже у складу са актуелним скуповима улазних и излазних променљивих, и мерених вредности. За разлику од сличних постојећих решења, ANN/GA метода оптимизује скоро све елементе неуронске мреже. Хибрид врши самоподешавање модела тако што оптимизује број скривених слојева, број неурона у тим слојевима, избор активационе функције, алгоритам учења, као и вредности параметара учења у складу са одабраним алгоритмом. Мрежа се адаптира, оптимизује и обучава на основу актуелних историјских података. Развијена ANN/GA метода је валидована кроз студију случаја бране Гранчарево која се налази на реци Тебишњици у Републици Српској. У студији је помоћу DEVONNA софтверског агента моделирано радијално померање тачке која се налази у структури бране у функцији нивоа воде у акумулацији, температуре и времена од почетка експлоатације бране. Да би се тестирале перформансе ANN/GA хибрида и DEVONNA агента, генерисани модели су поређени са моделима који су добијени коришћењем еквивалентног MLR/GA хибрида. Резултати су показали да се могу генерисати ANN модели чија је тачност предикције структурног понашања бране приближно 10% већа у поређењу са моделима који су базирани на MLR/GA методи. Међутим, за разлику од MLR модела, који су отпорни на температурне фазне помаке присутне на различитим географским локацијама, ANN модели показују нестабилно понашање под таквим околностима. Ова предност MLR модела се тумачи способношћу линеарних регресија да суперпонирањем тригонометријских регресора постигну жељени фазни помак. Поред тога, генерисање ANN модела је временски знатно захтевније. Генерисање оптимизованих ANN модела брана може захтевати неколико сати, што је значајно дуже од 5 до 10 минута потребних за генерисање еквивалентног MLR модела. Овај недостатак ANN/GA хибрида може бити превазиђен коришћењем паралелних софтверских оквира за оптимизацију еволуционим алгоритмима, као што је WoBinGO [139], који обезбеђује скоро линеарно убрзање. Коришћењем WoBinGO оквира на просечној Grid инфраструктури, која се састоји од неколико стотина чворова, може скратити време генерисања оптимизованог ANN модела на неколико минута.

Резултати моделирања структурног понашања бетонских брана који су приказани у поглављима 6 и 7, добијени на основу MLR/GA и ANN/GA методе, доказују трећу претпоставку истраживања спроведених у оквиру ове дисертације.

У оквиру 8. поглавља дисертације су регресиони модели генерисани коришћењем MLR/GA и ANN/GA методе поређени са моделима у форми постепених регресија, у циљу оцене перформанси новоразвијених хибридних метода у односу на постојеће. Структурно понашање бетонске лучне бране Боцац, које је служило као платформа за поређење је, као и у претходним поглављима, исказано преко радијалног померања посматраних тачака на круни бране. На основу добијених резултата могу се извести следећи закључци: а) Квалитет добијених модела је упоредив; б) Када се термички ефекти исказују преко температуре структуре, ANN/GA хибриди генеришу моделе највишег квалитета. Постепене регресије генеришу моделе који су нешто вишег нивоа квалитета у односу на моделе креиране коришћењем MLR/GA хибрида; в) Када се термички ефекти исказују кроз дане у години, модели генерисани на бази MLR/GA хибрида су нешто веће тачности у односу на моделе добијене постепеним регресијама; г) Са аспекта временске ефикасности, техника постепених регресија је у предности у односу на хеуристике вештачке интелигенције. Компаративна анализа модела

генерисаних на основу MLR/GA и ANN/GA метода, са једне, и модела у форми постепених регресија, са друге стране, је показала да представљене методе у појединим аспектима превазилазе могућности постојећих метода за генерисање регресионих модела. Тиме је потврђена четврта претпоставка истраживања у оквиру ове дисертације.

Развој модерних сензора и мерне и аквизиционе технике води ка рапидном повећању комплексности објеката моделирања, како са аспекта броја улазних променљивих, тако и са аспекта обима мерења. Ово би нужно водило до вишеструких линеарних регресија неприхватљиве дужине и сложености или до веома комплексних модела у облику вештачких неуронских мрежа. Проблеми преприлагођавања, енормног повећања времена потребног за генерисање оваквих модела и други проблеми би у таквим случајевима посебно дошли до изражаја. Из тих разлога се, као императив, намеће потреба да се могући проблеми предупреду и да се побољшају перформансе предложених хибридних метода и у таквим околностима. Из тог разлога је у завршном делу дисертације примењена PCA техника редукције димензија простора истраживања, у циљу регуларизације регресионих модела који су засновани на MLR/GA и ANN/GA методама. За студију случаја је одабран скуп података о измереним температурама, везаним за мониторинг бране Гранчарево. PCA метода факторске анализе је примењена за редукцију 45 променљивих: 1 променљиве која репрезентује средњу температуру ваздуха и 44 променљиве, које репрезентују температуре бетонске структуре на различитим позицијама. На основу детаљне PCA анализе, за две и за три компоненте, усвојено је трокомпонентно решење са *oblmin* ротацијом, које објашњава укупно 96.363% варијабилитета, односно по главним компонентама: 64.016%, 24.243%, 5.619% укупне варијансе, редом. У простору истраживања који је у погледу термичких ефеката редукован са 45 на 3 компоненте генерисани су различити регресиони модели и тестиране перформансе модела заснованих на MLR/GA и ANN/GA хибридама. Резултати и детаљне, нумеричке и графичке, појединачне и компаративне анализе квалитета добијених модела су показале да обе методе након редукције димензија могу генерисати моделе високе тачности и прихватљиве сложености, чак и у околностима када је почетни, факторабилни простор истраживања карактерисао велики број улазних променљивих.

PCA и редукција димензија простора истраживања омогућили су креирање квалитетних, а једноставнијих модела заснованих на развијеним MLR/GA и ANN/GA методама и тиме потврдили пету претпоставку реализованих истраживања у оквиру ове дисертације.

Независно од примењене технике моделирања, добијени резултати указују на следеће: а) Ако се у регресијама за приказ термичких оптерећења користе само регресори који одсликавају варирање средње температуре ваздуха, квалитет генерисаних модела ће бити незадовољавајући због великих осцилација у односу на температуру конструкције. За креирање квалитетнијих регресионих модела структурног понашања бетонских брана, неопходно је у моделе уградити регресоре који боље репрезентују термичке ефекте на конструкцију, као што су температуре структуре (бетона) или дани у години који прате периодичност у променама температурног поља; б) Величина грешке $RMSE_{Learn}$ и $RMSE_{Test}$ зависи од стохастике и обима података у скупу за учење и скупу за тестирање. При моделирању података великог обима не мора се јављати уобичајени однос $RMSE_{Learn} < RMSE_{Test}$; в) За константан однос величина скупова за учење и тестирање, стохастика GA не доводи до значајних варирања квалитета модела; д) Увођење тригонометријских регресора у контејнер базисних функција, доводи до значајног побољшања квалитета модела.

Као коначан закључак ове дисертације се може констатовати да предложене хибридне методе и развијени софтверски агенти, уз примену техника редукције димензија простора истраживања, представљају моћан алат за моделирање реалних објеката и система.

10.2. Правци даљег развоја

За даља истраживања у овој области се могу дефинисати следећи правци:

- а) Паралелизација развијених метода коришћењем WoBinGO софтверског оквира за дистрибуирану евалуацију у еволуционим алгоритмима;
- б) Унапређење критеријума комплексности, коришћењем статистичке сигнификантности сваког регресора у моделу;
- в) Развој еволуционих хеуристика које би требале да оснаже генетске модификације хромозома у складу са статистичком значајношћу њихових регресора;
- ђ) Креирање stepwise regressions/GA хибрида, помоћу кога би се, са постепеним регресијама, генерисао довољно тачан модел, а потом помоћу GA трагало за тачнијим и једноставнијим моделом.

Рад на даљем унапређењу регресионих модела и даље је изазов. Верујем да ће могући правци истраживања, бити инспирација за нове докторске дисертације и да ће кондензована знања у овој области, у будућности, довести до нових математичких дисциплина.

мр Милован Миливојевић, дипл. инг.

Библиографија

- [1] М. Миливојевић. *Математичко моделирање реалних процеса и система на бази централних композиционих планова*, магистарски рад, Београд: Машински факултет, Београд, 1996.
- [2] P.R. Duchowicz, E.A. Castro. *The Order Theory in QSPR - QSAR Studies, Mathematical Chemistry Monographs*, I. Gutman [ed], Kragujevac: University of Kragujevac, Faculty of Science, Kragujevac, 2008., pp. 15-20.
- [3] Ч. Дарвин. *Постанак врста путем природног одабирања или очување повлашћених раса у борби за живот* (превод оригинала: Charles Darwin, *Origin of Species (6th edition)*), ICON Classics, San Diego, CA, USA, 2008.), Нови Сад: Академска књига, 2009.
- [4] H. Hasheminia, S. T. A. Niaki. A genetic algorithm approach to find the best regression/econometric model among the candidates, *Applied Mathematics and Computation*, no. 183, 2006., pp. 337-9.
- [5] M. Milivojevic, S. Stopic, B. Friedrich, B. Stojanovic, D. Drndarevic. Computer modeling of high-pressure leaching process of nickel laterite by design of experiments and neural networks, *International Journal of Minerals, Metallurgy and Materials*, vol. 19, no. 7, 2012., pp. 584-94.
- [6] B. Stojanovic, M. Milivojevic, M. Ivanovic, N. Milivojevic, D. Divac. Adaptive System for Dam Behavior Modeling Based on Linear Regression and Genetic Algorithms, *Advances in Engineering Software*, vol. 65, 2013., pp. 182-90.
- [7] J. Cohen, P. Cohen, S. G. West, L. S. Aiken. *Applied Multiple Regression/Correlation Analysis for the Behavioral Sciences*, vol. 3rd Edition, Taylor & Francis, 2002.
- [8] www, www.duke.edu/~rnau/regstep.htm, January 2012. [Online]. [Accessed January 2012].
- [9] J. Pallant. *A Step by Step Guide to Data Analysis Using SPSS for Windows (Version 15). 3rd edition*, Berkshire: Open University Press, McGraw-Hill Education, 2007.
- [10] B.G. Tabachnick, S. Fidell. *Using Multivariate Statistics (5th edition)*, Boston: Pearson Education, 2007.
- [11] www, www.uta.edu/faculty/sawasthi/Statistics/stgsr.html, January 2012. [Online]. [Accessed January 2012].
- [12] М. Жижић, М. Ловрић, Д. Павличић, *Методи статистичке анализе*, Београд: Економски факултет, Београд, 2006.
- [13] P.S. Mann, *Uvod u statistiku* (autorizovani prevod originala: *Introductory Statistics (6th edition)*), Prem S. Mann, Eastern Connecticut State University, John Wiley & Sons Inc.), Београд: Економски факултет, Београд, 2009.
- [14] P. Blattner, L. Ulrich, K. Cook, T. Dyck, *Specilano izdanje: Vodič kroz Microsoft Excel 2000* (Autorizovani prevod: *Special Edition Using Microsoft Excel 2000*, 1999, QUE Corporation), Београд: CET, 2000.
- [15] Ј. Станић, *Метод инжењерских мерења*, Београд: Машински факултет, Београд, 1990.
- [16] www, www.weibull.com, June 2013. [Online]. [Accessed June 2013].
- [17] R Development Core Team (2012). *R: A language and environment for statistical computing*, Vienna, Austria: R Foundation for Statistical Computing, 2012. ISBN 3-

- 900051-07-0, URL <http://www.R-project.org>
- [18] M.J. Berry, G.S. Linoff. *Nearest Neighbor Approaches: Memory-Based Reasoning and Collaborative Filtering in Data Mining Techniques, Second Edition*, Indianapolis, Indiana, Wiley Publishing, Inc., 2004.
- [19] I. H. Witten, E. Frank, M.A. Hall. *Credibility: Evaluating What's Been Learned in Data Mining Practical Machine Learning Tools and Techniques, Third Edition*, Burlington, MA 01803, USA, Morgan Kaufmann Publishers is an imprint of Elsevier Copyright © 2011 Elsevier Inc., 2011.
- [20] N. Suguna, K. Thanushkodi. An Improved k-Nearest Neighbor Classification Using Genetic Algorithm, *International Journal of Computer Science Issues*, vol. 7: (4 / No 2), 2010., pp. 18-20.
- [21] S. Russel, P. Norvig. *Artificial Intelligence: A modern Approach, 3rd edition*, New York, Pearson Education, 2010.
- [22] J.H. Holland. *Adaptation in natural and artificial systems*, Michigan: Ann Arbor: University of Michigan Press, 1975.
- [23] D.E. Goldberg. *Genetic algorithms in search, optimization, and machine learning*, Boston: Addison-Wesley Longman Publishing Co., 1989.
- [24] R. Arrdito, G. Cocchetti. Statistical approach to damage diagnostic of concrete dam by radar monitoring: formulation and pseudo-experimental test, *Engineering Structures*, vol. 28, 2006., pp. 2036-45.
- [25] D. Hu, Z. Zhou, Y. Li, X. Wu. Dam Safety Analysis Based on on Stepwise Regression Model, *Advanced Materials Research*, Vol. 204(210), 2011., pp. 2158-61.
- [26] A. Szostak-Chrzanowski, A. Chrzanowski, M. Massiera, Use of deformation monitoring results in solving geomechanical problems, case studies, *Engineering Geology*, Vols. 79(1-2), 2005., pp. 3-12.
- [27] C. Yifeng, H. Ran, L. Wenbo, L. Dianqing, Z. Chuangbing. Modeling coupled processes of non-steady seepage flow and non-linear deformation for a concrete-faced rockfill dam, *Computers & Structures*, vol. 89 (13–14), 2011., pp. 1333-51.
- [28] J. Mata, Interpretation of concrete dam behavior with artificial neural network and multiple linear regression models, *Engineering Structures*, vol. 33, 2011., pp. 903-10.
- [29] X. Chang, Y. Dongjie, D. Chengfa. Hybrid GA/SIMPLS as alternative regression model in dam deformation analysis, *Engineering Applications of Artificial Intelligence*, vol. 25, 2012., pp. 468-75.
- [30] C. Kuan-Yu. Forecasting systems reliability based on support vector regression with genetic algorithms, *Reliability Engineering and System Safety*, vol. 92, 2007., pp. 423-32.
- [31] V. Rankovic, N. Grujovic, D. Divac, N. Milivojevic, A. Novakovic. Modeling of dam behavior based on neuro fuzzy identification, *Engineering Structures*, vol. 35, 2012., pp. 107-13.
- [32] E. Dogan, M. Gumrukcuoglu, M. Sandalci, M. Opan. Modeling of evaporation from the reservoir of Yuvacik dam using adaptive neuro-fuzzy inference systems, *Engineering Applications of Artificial Intelligence*, vol. 23, 2010., pp. 961-7.
- [33] M. Rohaninejad, Z. Mahdi. Combining Monte Carlo and finite difference methods for effective simulation of dam behavior, *Advances in Engineering Software*, vol. 45, 2012., pp. 197–202.
- [34] R.A. Yildiz. A new design optimization framework based on immune algorithm and

- Taguchi's method, *Computer in Industry*, vol. 60, 2009., pp. 613-20.
- [35] W. Wen-Chun, C. Kwok-Wing, C. Chun-Tian, Q. Lin. A comparison of performance of several artificial intelligence methods for forecasting monthly discharge time series, *Journal of Hydrology*, vol. 374, 2009., pp. 294-306.
- [36] S. Gholizadeh, M.S. Seyedpoor. Shape optimization of arch dam by metaheuristics and neural networks for frequency constraints, *Scientia Iranica A: Civil Engineering*, vol. 18(5), 2011., pp. 1020-7.
- [37] Y.G. Xi, P.J. Yue, B.X. Zhou, P. Tang. Application of an artificial immune algorithm on statistical model of dam displacement, *Computer and Mathematics with Applications*, vol. 62, 2011., pp. 3980-6.
- [38] N. Wiener. *Cybernetics*, Wiley, 1948.
- [39] S.C. Larson. The shrink age of the coefficient of multiple correlation, *Journal Educational Psychology*, vol. 22, pp. 45-55, 1931.
- [40] M. Stone, Cross-validatory choice and assessment of statistical predictions, *J. Royal Statistical Society*, vol. 36, 1974., pp. 113-33.
- [41] G. Golub, M. Heath, G. Wahba, Generalized cross-validation as a method for choosing a good ridge parameter, *Technometrics*, vol. 21, no. 2, 1979.
- [42] A.N. Tikhonov, *Solution of incorrectly formulated problems and regularization method*, Soviet Math Dok., vol. 5, 1963.
- [43] J. Stevens. *Applied multivariate statistics for the social sciences (3rd ed)*, Mahwah: NJ: Lawrence Erlbaum, 1996.
- [44] D.E. Rumelhart, J.L. McClelland. *Parallel Distributed Processing: Explorations in the Microstructure of Cognition*, vol. Volume 1: Foundations, Cambridge MA: MIT Press, 1986.
- [45] M. Majdi, M. Beiki. Evolving neural network using a genetic algorithm for predicting the deformation modulus of rock masses, *International Journal of Rock Mechanics & Mining Sciences*, vol. 47, 2010., pp. 246-53.
- [46] Z. Yu, *Feed-forward Neural Networks and Their Applications in Forecasting*, University of Houston, 2000.
- [47] J. Heaton. *Programming Neural Networks with Encog3 in C#*, Heaton Research Inc., <http://goo.gl/Zo8k70> (accessed 3.6.2015), 2015.
- [48] M. Riedmiller, H. Braun. A direct adaptive method for faster backpropagation learning: the RPROP algorithm, *Neural Networks, IEEE International Conference on: San Francisco, CA*, vol.1, 1993., pp. 586-91.
- [49] M. Riedmiller. Advanced supervised learning in multi-layer perceptrons - From backpropagation to adaptive learning algorithms, *Computer Standards & Interfaces*, vol. 16, 1994., pp. 265-78.
- [50] A.D. Anastasiadis, G.D. Magoulas, M.N. Vrahatis. New globally convergent training scheme based on the resilient propagation algorithm, *Neurocomputing*, vol. 64, 2005., pp. 253-70.
- [51] F. Günther, S. Fritsch. Neuralnet: Training of Neural Networks, *The R Journal*, vol. 2(1), 2010., pp. 30-8.
- [52] B.H.V. Topping, J. Sziveri, A. Bahreinejad, J.P.B. Leite, B. Cheng. Parallel processing, neural networks and genetic algorithms, *Advances in Engineering Software*, vol. 29(10), 1998., pp. 763-86.
- [53] R. Hecht-Nielsen. Kolmogorov's mapping neural network existence theorem, *In:*

- Proceedings of first IEEE international conference on neural networks*, vol. 3, 1987., pp. 11–4.
- [54] M.M. Poulton. Neural networks as an intelligence amplification tool: a review of applications, *J Geophys*, vol. 67(3), 2002., pp. 979–93.
- [55] K. Swingler. *Applying neural networks: a practical guide*, New York, Academic Press, 1996.
- [56] C.G. Looney. Advances in feed-forward neural networks: demystifying knowledge acquiring black boxes, *IEEE Trans Knowl Data Eng.*, vol. 8(2), 1996., p.p. 211–26.
- [57] M. Nelson, W.T. Illingworth. *A practical guide to neural nets*, Reading. Mass, Addison-Wesley, 1990.
- [58] C.G. Looney, *Pattern recognition using neural networks, theory and algorithms for engineers and scientists*, Oxford University Press, 1997.
- [59] K. Messer, J. Kittler. Choosing an optimal neural network size to aid search through a large image database, *In: Proceedings of ninth British machine vision conference (BMVC98)*, 1998., pp. 235–44.
- [60] M.H Beale, M.T. Hagan, H.B. Demuth. *Neural Network ToolboxTM, User's guide, R2015a*, 2015.
- [61] B.J. Wythoff. *Backpropagation neural networks: a tutorial*, Chemometr Intell Lab Syst., vol. 18, pp. 115–55.
- [62] L. He, X. Yu, Z. Huang, A. Mahgoub, A. Talab. The two-dimensional double-entropy threshold based on the parallel genetic simulated annealing algorithms. *Optik*, vol. 127, 2016., pp. 96–101.
- [63] Y.X. Jie, Y. Liu. Simulated annealing based algorithm for node generation in seepage analysis with meshless method, *Mechanics Research Communications*, vol. 43., 2012., pp. 96–100.
- [64] X. Zhang, X. Yuan, Y. Yuan. Improved hybrid simulated annealing algorithm for navigation scheduling for the two dams of the Three Gorges Project, *Computers and Mathematics with Applications*, vol. 56, 2008., pp. 151–9.
- [65] Y. Xiao, A. K. Abdullah. A simulating annealing algorithm to solve the green vehicle routing & scheduling problem with hierarchical objectives and weighted tardiness, *Applied Soft Computing*, vol. 34, 2015., pp. 372–88.
- [66] M.A. Franco, J. Bacardit, Large-scale experimental evaluation of GPU strategies for evolutionary machine learning, *Information Sciences*, vol. 330, 2016., pp. 385–402.
- [67] H. Hajabdollahi, A. Ganjehkaviri, M. Nazri, M. Jaafar. Assessment of new operational strategy in optimization of CCHP plant for different climates using evolutionary algorithms, *Applied Thermal Engineering*, vol. 75, 2015., pp. 468–80.
- [68] D. Shaw, J. Miles, A. Gray. *Genetic Programming Within Civil Engineering*, Engineers House, Clifton, Bristol UK, 2004.
- [69] D. J. Montana, Strongly typed genetic programming, *Evolutionary computation*, vol. 3(2), 1995., pp. 199–230.
- [70] Y.S. Hong, R. Bhamidimarri. Evolutionary self-organising modelling of a municipal wastewater treatment plant, *Water Research*, vol. 37, 2003., pp. 331–8.
- [71] P.A. Whigham, P.F. Crapper. Modelling rainfall-rainoff using genetic programming, vol. 33, 2001., pp. 707–21.
- [72] Y. Yang, S.K. Soh. Automated optimum design of structures using genetic programming, *Computers and Structures*, vol. 80, 2002., pp. 1537–46.

- [73] O.M. Rezapour, L.T. Shui, A.A. Deghhani. Review of genetic programming in water resource engineering, *Australian Journal of Basic and Applied Science*, vol. 4(11), 2010., pp. 5663-7.
- [74] T. Khu, S. Liong, W. Babovic, H.Madsen, N. Muttill. Genetic programming and its application in real-time runoff forecasting, *Journal of American water resources association*, vol. 37(2), 2001., pp. 439-51.
- [75] M. Makkeasorn, N.B. Chang, X. Zhou. Short-term streamflow forecasting with global climate change implications - a comparative study between genetic programming and neural network models, *Journal of Hidrology*, vol. 352, 2008., pp. 336-54.
- [76] J.R. Koza. *Genetic Programming: On the Programming of Computers by means of Natural selection*, Cambridge, MA: The MIT Press, 1992.
- [77] E.L. Harris, V. Babovic, R.A. Falconer. Velocity prediction in compound channels with vegetated floodplains using genetic programming, *Int.J. River Basin Manag.*, vol. 1(2), 2003., pp. 117-23.
- [78] A. Ayttek, O. Kishi. A genetic programming approach to suspended sediment modelling, *Journal of Hidrology*, vol. 351, 2008., pp. 288-98.
- [79] R. Leardi. Genetic algorithm in chemistry, *Journal of Chromatography A*, vol. 1158, 2007., pp. 226-33.
- [80] J. Lee, D-W Kim. An effective initialization method for genetic algorithm-based robot path planning using a directed acyclic graph, *Information Sciences*, vol. 332, 2016., pp. 1-18.
- [81] M. Shahnewaz, R. Machial, M.S. Alam, A. Rteil. Optimized shear design equation for slender concrete beams reinforced with FRP bars and stirrups using Genetic Algorithm and reliability analysis, *Engineering Structures*, vol. 107, 2016., pp. 151-65.
- [82] M. Hosseini, S. Alireza, M.Naeini, A.A. Dehghani, Y. Khaledian. Estimation of soil mechanical resistance parameter by using particle swarm optimization, genetic algorithm and multiple regression methods, *Soil & Tillage Research*, vol. 157, 2016., pp. 32-42.
- [83] X. Fu, S. Ricci, C. Bisagni. Minimum-weight design for three dimensional woven composite stiffened panels using neural networks and genetic algorithms, *Composite Structures*, vol. 134, 2015., pp. 708-15.
- [84] V.D. Tsoukalas, N.G. Fragiadakis. Prediction of occupational risk in the shipbuilding industry using multivariable linear regression and genetic algorithm analysis, *Safety Science*, vol. 83, 2016., pp. 12-22.
- [85] M. Marzouk, A. Abubakr. Decision support for tower crane selection with building information models and genetic algorithms, *Automation in Construction*, vol. 61, 2016., pp. 1-15.
- [86] R. Dou, C. Zong, G.Nan. Multi-stage interactive genetic algorithm for collaborative product customization, *Knowledge-Based Systems*, vol. 92, 2016., pp. 43-54.
- [87] X. Xiaosong Zhao, C-Y Hsu, P-C Chang, L. Li. A genetic algorithm for the multi-objective optimization of mixedmodel assembly line based on the mental workload, *Engineering Applications of Artificial Intelligence*, vol. 47, 2016., pp. 140-6.
- [88] S.A.F. Oliveira, R. Alajmar, R.Neto, F.N. Bezzer. A novel Genetic Algorithms and SURF-Based approach for image retargeting, *Expert Systems With Applications*, vol. 44, 2016., pp. 332-43.
- [89] S. Chiappone, O. Giuffre, A. Grana, R. Mauro, A.Sferlazza. Traffic simulation models

- calibration using speed–density relationship: An automated procedure based on genetic algorithm, *Expert Systems With Applications*, vol. 44, 2016., pp. 147–55.
- [90] G. Yang, S. Wu, Q. Jin, J. Xu. A hybrid approach based on stochastic competitive Hopfield neural network and efficient genetic algorithm for frequency assignment problem, *Applied Soft Computing*, vol. 39, 2016., pp. 104–16.
- [91] M. Golub, *Genteski algoritmi*, Zagreb: Sveuciliste Zagreb, 2007.
- [92] M. Srinivas, L.M. Patnaik. Genetich Algorithms: A Survey, *Computer*, June 1994., pp. 17-26.
- [93] M. Srinivas, L.M. Patnaik. Adaptive Probabilities of Crossover and Mutation in Genetic Algorithms, *Man and Cybernetics*, April 1994.
- [94] Y. Chen, R. Hu, W. Lu, D. Li, C. Yhou. Modeling coupled processes of non-steady seepage flow and non-linear deformation for a concrete-faced rockfill dam, *Comput. Structur.*, Vols. 89(13-14), 2011., pp. 1333-51.
- [95] C. Xu, D. Yue, C. Deng, Hybrid GA/SIMPLS as alternative regression model in dam deformation analysis, *Eng Appl Artif Intell*, vol. 25, 2012., pp. 468-75.
- [96] Y. K. Cen. Forecasting systems reliability based on support vector regression with genetic algorithms, *Reliab Eng Sys Saf*, vol. 92, 2007., pp. 423-32.
- [97] S. Li, L. Ding, L. Zhao, W. Zhou. Optimization design of arch dam shape with modified complex metod, *Adv Eng Softw*, vol. 40, 2009., pp. 804-8.
- [98] A. De Sortis, P. Paoliani. Statistical analysis and structural identification in concrete dam monitoring. *Engineering Structures*, vol. 29, 2007., pp.110-20.
- [99] W.C. Wang, K.W. Chau, C.T. Cheng, L. Qui. A comparison of performance of several artificial intelligence methods for forecasting monthly discharge time series, *Journal of Hidrology*, vol. 374, 2009., pp. 294-306.
- [100] F. Gaxiola, P. Melin, F. Valdez, J.R.Castro, O. Castillo. Optimization of type-2 fuzzy weights in backpropagation learning for neural networks using GAs and PSO, *Applied Soft Computing*, vol. 38, 2016., pp. 860–71.
- [101] S.M. Hosseini, N. Mahjouri. Integrating Support Vector Regression and a geomorphologic Artificial Neural Network for daily rainfall-runoff modeling, *Applied Soft Computing*, vol. 38, 2016., pp. 329–45.
- [102] S. Chanda, S. Gupta, D.K.Pratihar. A combined neural network and genetic algorithm based approach for optimally designed femoral implant having improved primary stability, *Applied Soft Computing*, vol. 38, 2016., pp. 296–307.
- [103] T. Baklacioglu. Modeling the fuel flow-rate of transport aircraft during flight phases using genetic algorithm-optimized neural networks, *Aerospace Science and Technology*, vol. 49, 2016., pp. 52–62.
- [104] H. Liu, H. Tian, X. Liang, Y. Li. New wind speed forecasting approaches using fast ensemble empirical model decomposition, genetic algorithm, Mind Evolutionary Algorithm and Artificial Neural Networks, *Renewable Energy*, vol. 83, 2015., p.p. 1066-75.
- [105] O.E. Bukharov, D.P. Bogolybov, Development of a decision support system based on neural networks and a genetic algorithm, *Expert Systems with Applications*, vol. 42, 2015., pp. 6177–83.
- [106] D. Sánchez, P. Melin, O. Castillo. Optimization of modular granular neural networks using a hierarchical genetic algorithm based on the database complexity applied to human recognition, *Information Sciences*, vol. 309, 2015., pp. 73–101.

- [107] V. Vinay Chandwani, V. Agrawal, R. Nagar. Modeling slump of ready mix concrete using genetic algorithms assisted training of Artificial Neural Networks, *Expert Systems with Applications*, vol. 42, 2015., pp. 885–93.
- [108] I. Karimi, N. Khaji, M. T. Ahmadi, M. Mirzayee. Evolving neural network using a genetic algorithm for predicting the deformation modulus of rock masses, *Engineering Structures*, vol. 32, 2010., pp. 3583–91.
- [109] C-B Zhou, W. Liua, Y-F Chen, R. Hua, K. Wei. Inverse modeling of leakage through a rockfill dam foundation during its construction stage using transient flow model, neural network and genetic algorithm, *Engineering Geology*, vol. 187, 2015., pp. 183–95.
- [110] R. Fedele, G. Maier, B. Miller. Evolving neural network using a genetic algorithm for predicting the deformation modulus of rock masses, *International Journal of Fracture*, vol. 137, 2006. pp. 151–72.
- [111] F. Hooshyaripor, A. Tahershamsi, S. Golian. Application of copula method and neural networks for predicting peak outflow from breached embankments, *Journal of Hydro-environment Research*, vol. 8, 2014., pp. 292–303.
- [112] K. Haddad, A. Rahman, M. A. Zaman, S. Shrestha. Applicability of Monte Carlo cross validation technique for model development and validation using generalised least squares regression, *Journal of Hydrology*, vol. 482, 2013., pp. 119–28.
- [113] Y. Zhang, Y. Yin, D. Guo, X. Yu, L. Xiao. Cross-validation based weights and structure determination of Chebyshev-polynomial neural networks for pattern classification, *Pattern Recognition*, vol. 47, 2014., pp. 3414–28.
- [114] C. Bergmeir, J. M. Benitez, On the use of cross-validation for time series predictor evaluation, *Information Sciences*, vol. 191, 2012., pp. 192–213.
- [115] F. Nu-Fang, S. Zhi-Hua, L. Lu, J. Cheng. Rainfall, runoff, and suspended sediment delivery relationships in a small agricultural watershed of the Three Gorges area, China, *Geomorphology*, vol. 135, 2011., pp. 158–66.
- [116] T. R. Peng, C. H. Wang, Identification of sources and causes of leakage on a zoned earth dam in northern Taiwan: Hydrological and isotopic evidence, *Applied Geochemistry*, vol. 23, 2008., pp. 2438–51.
- [117] X. Guang-yong, Y. Jian-ping, Z. Bao-xing, T. Pu. Application of an artificial immune algorithm on a statistical model of dam displacement, *Computers and Mathematics with Applications*, vol. 62, 2011., pp. 3980–6.
- [118] A. Adler. *R in a Nutshell*, 2010.
- [119] V. Rancic. *Metode za smanjenje dimenzionalnosti podataka i njihova primena u prirodnim naukama; master rad*, Univerzitet u Novom Sadu, Departman za matematiku i informatiku, 2013.
- [120] J. O. Nunnally. *Psychometric theory*, New York: McGraw-Hill, 1978.
- [121] M. S. Bartlett, A note on the multiplying factors for various chi square approximations, *Journal of the Royal Statistical Society*, vol. 16 (Series B), 1954., pp. 296–8.
- [122] H. Kaiser, A second generation Little Jiffy, *Psychometrika*, vol. 35, pp. 401–15.
- [123] H. Kaiser, An index of factorial simplicity, *Psychometrika*, vol. 39, pp. 31–6.
- [124] R. B. Catell, The scree test for the number of factors, *Multivariate Behavioral Research*, vol. 1, 1966., pp. 245–76.
- [125] J. L. Horn, A rationale and test for the number of factors in factor analysis, *Psychometrika*, vol. 30, 1965., pp. 179–85.
- [126] N. Choi, D. Fuqua, N. B. W. Griffin. Exploratory analysis of the structure of scores from

- the multidimensional scales of perceived self efficacy, *Educational and Psychological Measurements*, vol. 61, 2001., pp. 475-89.
- [127] J. Strober. The Frost multidimensional perfectionism scale revisited: more perfect with four (instead of six) dimensions, *Personality and Individual Differences*, vol. 24, 1998., pp. 481-91.
- [128] L. L. Thurstone, *Multiple factor analysis*, University of Chicago Press, Chichago, 1947.
- [129] M.A. Pett, N.R. Lackey, J.J. Sullivan. *Making Sense of factor analysis: the use of factor analysis for instrument development in health care research*, Thousand Oaks: Sage, 2003.
- [130] M. Varol, B. Gokot, A. Bekleyen, B.Sen. Spatial and temporal variations in surface water quality of the dam reservoirs in the Tigris River basin, Turkey, *Catena*, vol. 92, 2012., pp. 11–21.
- [131] N. Lambrakis, A. Antonakos, G. Panagopoulos. The use of multicomponent statistical analysis in hydrological environmental research, *Water Research*, vol. 38, 2004., pp. 1862–72.
- [132] C. Mendiguchia, C. Moreno, D.M. Galindo-Riano, M. Garcia-Vargas. Using chemometric tools to assess anthropogenic effects in river water. A case study: Guadalquivir River (Spain), *Analytica Chimica Acta*, vol. 515, pp. 143–149.
- [133] K.P. Singh, A. Malik, D. Mohan, S. Sinha. Multivariate statistical techniques for the evaluation of spatial and temporal variations in water quality of Gomti River (India): a case study, *Water Research*, vol. 38, pp. 3980–92.
- [134] K.P. Singh, A. Malik, S. Sinha. Water quality assessment and apportionment of pollution sources of Gomti River (India) using multivariate statistical techniques: a case study, *Analytica Chimica Acta*, vol. 538, pp. 355–74.
- [135] D.A. Wunderlin, M.P. Diaz, M.V. Ame, S.F. Pesce, A.C. Hued, M.A. Bistoni. Pattern recognition techniques for the evaluation of spatial and temporal variations in water quality. A case study: Suquia River basin (Cordoba, Argentina), *Water Research*, vol. 35, pp. 2881–94.
- [136] S. Westra, C. Brown, U. Lall, A. Sharma. Modeling multivariable hydrological series: Principal component analysis or independent component analysis?, *Water resources research*, vol. 43, 2007.
- [137] W-J. Dai, B. Liu, X-L Ding, D-W Huang. Modeling dam deformation using independent component regression method, *Trans. Nonferrous Met. Soc. China*, vol. 23, 2013., pp. 2194–200.
- [138] D. Santill, E. Salet, M.A. Toledo. A new 1D analytical model for computing the thermal field of concrete dams due to the environmental actions, *Applied Thermal Engineering*, vol. 85, 2015., pp. 160-71.
- [139] M. Ivanovic, V. Simic., B. Stojanovic. A. Kaplarevic-Malisic, B. Marovic. Elastic grid resource provisioning with WoBinGO: A parallel framework for genetic algorithm based optimization, *Future Generation Computer Systems.*, vol. 42, 2015., pp. 44–54.

Биографија

Милован Миливојевић је рођен 1964. године у Каменици код Ужица. Основну школу завршио је у Каменици (Луново Село) као носилац дипломе *Вук Караџић* а гимназију у Ужицу (математичко-технички смер, програмер) као носилац дипломе *Вук Караџић* и дипломе *Михаило Петровић - Алас*, за математику. Дипломирао је на Машинском факултету, Универзитета у Београду, 08.05.1990. год. са оценом 10 на тему *Развој CIQ модела за технолошке фабрике*, код проф. др Јока Станића. Током студија остварио је просечну оцену 9.63 и два пута био носилац *Универзитетске награде* (1985. и 1986. год.). Последипломске-магистарске студије одслушао и све предмете положио током периода 1991-1992. год., на *Катедри за производно машинство* Машинског факултета, Универзитета у Београду, смер *Флексибилни производни системи и роботика*. У периоду 1991-1995. год. радио на истој Катедри (смер *Флексибилни производни системи и роботика*) као асистент на различитим предметима почев од *Компјутерске графике* до *Пројектовања технолошких процеса*. Такође, на истој Катедри магистрирао 1996. год. на тему *Прилог математичком моделовању вишефакторних процеса и система на основу централних, композиционих експерименталних планова* (ментор: проф. др Јоко Станић).

У периоду од 1996. до 2000. год., водио рачунски центар *Ваљаонице алуминијума у Севојну, Ужице* и руководио пројектом *CIIS Seval (Computer Intergrated Information System)*, Ваљаонице алуминијума Севојно. Током периода 1997-2012. у оквиру *Школе рачунара Олимпија, Ужице* радио хонорарно као руководилац, инструктор и предавач на рачунарским курсевима и семинарима за преко 6500 полазника (просечан обим курсева 150 до 200 часова).

Од 1996. год., радио као спољни сарадник на Високој пословно-техничкој школи у Ужицу а од 2000. год. је стално запослен у овој школи. Један је од утемељивача *Информатичког одсека* на ВПТШ Ужице а током 2004/2005 био је в.д. шефа Информатичког одсека. У периоду 2010-2014. год. био је помоћник директора за наставу ВПТШ, Ужице. Предавач је на предметима: *Апликативни софтвер*, *Базе података*, *Увод у објектно програмирање*, *Алгоритми и структуре података*, *Техничко цртање са нацртном геометријом* - на основним студијама и *Компјутером интегрисана производња/СИМ*, *Рачунаром интегрисани ГИС*, *Веиштачка интелигенција*, као и *Информатичка методологија истраживачког рада* на специјалистичким струковним студијама.

У периоду од 1996-2013. год. похађао је велики број семинара у домену информатике и информационо-комуникационих технологија. Председник је *Одбора за Информатику, Регионалне привредне коморе, Ужице* (2006-2015.).

Један је од коаутора књиге, *РС рачунари и њихова примена* (ВПТШ, Ужице, 1996.). Објавио је око 60 радова у међународним и домаћим часописима и на међународним и домаћим конференцијама.

Учествовао у реализацији 4 ТЕМПУС пројекта и на DAAD пројекту / Germany: *Artificial Neural Network modelling of silver nanoparticle formation after thermal decomposition of an aerosol* / ANN-NanoAg, RWTH Aachen University, IME Process Metallurgy and Metal Recycling (руководиоци пројекта Prof. Bernd Friedrich / Dr.-Ing. Srecko Stopic) & Природно математички факултет, Универзитет Крагујевац, Институт за математику и информатику (руководилац пројекта др Бобан Стојановић) / www.daad.de. Такође је учесник у припреми немачког DFG пројекта, чији је предлог поднет у октобру 2015.

Шира област интересовања мр Милована Миливојевића: Флексибилни производни системи, CAD-CAM-CIQ, Рачунарски интегрисани ГИС, Информациони системи, Softcomputing, Објектно орјентисано програмирање, Релационе базе података, Моделовање и оптимизација реалних процеса и система, Теорија експеримента (*Design of Experiments – DOE*), Математичко моделовање, Регресиона анализа и примењена статистика, Факторска анализа (Principal Component Anlisis-PCA), Вештачка интелигенција (*Artificial Intelligence-AI*) са фокусом на: вештачке неуронске мреже (*Artificial Neural Networks-ANN*), генетске алгоритме (*Genetic Algorithms-GA*), оптимизацију ројем честица (Particle Swarm Optimization-PSO), Data Mining (kMean Clustering)...

Списак објављених радова, публикација, пројеката и техничких решења

Магистарска теза:

Mathematical Modeling of Real Objects, Processes and Systems by Means of CCF
Design of Experiments,
Faculty of Mechanical Engineering, University of Belgrade, Serbia, 1996

Књиге:

Milivojevic M., Petrovic S., Dikovic Lj., Raonic V., PCs and their applications,
Technical and Business College, Uzice, 1996.

А) Радови у часописима на SCI листи

- a.1. Milivojevic M., Stopic S., Friedrich B., Stojanovic B., Drndarevic D. (2012): *Computer Modeling of High Pressure Leaching Process of Nickel Laterite by Design of Experiments and Neural Networks*, International Journal of Minerals, Metallurgy and Materials, Volume 19, Number 7, July 2012, 584-595
- a.2. Stojanovic B., Milivojevic M., Ivanovic M., Milivojevic N., Divac D., Adaptive system for dam behavior modeling based on linear regression and genetic algorithms, *Advances in engineering software*, (2013), vol. 65, pp. 182-190
- a.3. Стојановић Б., Миливојевић М., Миливојевић Н., Антонијевић Д., Self-tuning system for dam behavior modeling based on evolving artificial neural networks, *Advances in engineering software*, (рад предат на рецензију у августу 2015.).

Б) Радови у научним часописима

- б.1. Majstorovic V., Milivojevic M., Stanic J., *CIQ development model for technological factory*, Tehnika, no 3-4, Belgrade, 1991
- б.2. Stanic J., Milivojevic M., Drulovic M., *Analytical methods of search optimization*, Tehnika, no. 10-11, Belgrade, 1994
- б.3. Milivojevic M., Stopic S., Stojanovic B., Drndarevic D., Bernd F., *Forward stepwise regression in determining dimensions of forming and sizing tools for self-lubricated bearings*, METTAL Internationale Fachzeitschrift fur metallurgie, April, 2013, Vol. 67, pp. 147-153, ISSN 0026-0746
- б.4. Milivojevic M., Stopic S., Friedrich B., Stojanovic B., Drndarevic D., *Computer Modeling of High Pressure Leaching Process of Nickel Laterite by Design of Experiments and Neural Networks*, International Journal of Minerals, Metallurgy and Materials, Volume 19, Number 7, July 2012, 584-595 (M22)
- б.5. Stojanovic B., Milivojevic M., Ivanovic M., Milivojevic N., Divac D., *Adaptive system for dam behavior modeling based on linear regression and genetic algorithms*, ADVANCES IN ENGINEERING SOFTWARE, (2013), vol. 65, pp. 182-190 IF2014=1.402 (M21)
- б.6. Drndarevic D., Milivojevic M., *Modelling with neural networks*, Journal of Mechatronics, Vol. 2., No. 2., June 2014., (143-146)

В) Радови на међународним научним конференцијама

- в.1. Milivojevic M., Drulovic M., A contribution to torsional vibration transmission systems, 19. JUPITER conference, 6. Yugoslav symposium CAD/CAM, Prohor Pčinjski, 1993.
- в.2. Stanic J., Drulovic M., Milivojevic M., Development of software for technical optimization of real objects by search methods, International Conference – Heavy Machinery - TM'93, Krusevac-Vrnjacka Banja, Serbia, 1993.
- в.3. Stanic J., Milivojevic M., Ivkovic N., Modeling and optimization of real objects using analytical search methods, 25. Yugoslav Conference on Production Engineering, Beograd, Serbia, 1994.
- в.4. Ivkovic N., Milivojevic M., Contribution to the modeling 3D surfaces, 20. JUPITER Conference 7. Yugoslav symposium CAD/CAM, Beograd, Serbia, 1994.
- в.5. Maneski T., Milivojevic M., Development of software for 3D surface modeling using interpolation methods, 25. Yugoslav Conference on Production Engineering, Beograd, Serbia, 1994.
- в.6. Milivojevic M., Stanic J., Modeling the surface using central composition plans, 21. JUPITER Conference, Beograd, Serbia, 1995.
- в.7. Milivojevic M., Djuricic M., Software Quality, Conference Quality in Services, Regional Chamber of Commerce, Uzice, Serbia, 1995.
- в.8. Djuricic M., Kaljevic M., Milivojevic M., Acimovic Z., Contribution to the methodology of assessment of quality systems in service industry, Conference Quality in Services, Regional Chamber of Commerce, Uzice, Serbia, 1995.
- в.9. Milivojevic M., Stanic J., Modeling and optimization of dimensional tolerances of the Hostaform gears, 22. JUPITER Conference, Beograd, Serbia, 1996.
- в.10. Panic S., Milivojevic M., Drndarevic D., Modeling of friction forces at the outbreak of the pallet from the mold in the process of molding metal powders, 24 JUPITER Conference, Zlatibor, Serbia, 1998.
- в.11. Drndarevic D., Milivojevic M., Radojicic N., Sokic D., System identification operation factors in Aluminum mill factory, 30 JUPITER Conference, Beograd, Serbia, 2004. (1.19-1.22)
- в.12. Drndarevic D., Milivojevic M., Sokic D., Application of the system characteristics in labeling the factors of business, 30. HIPNEF symposium, V.Banja, Serbia, 2006. (351-356)
- в.13. Drndarevic D., Milivojevic M., Djuricic M., Laboratory for process control I, Scientific conference, Technical College, Uzice, Serbia, 2006.
- в.14. Drndarevic D., Milivojevic M., Panic S., Modelling of sintering process, IV Balkan Conf. on Metallurgy, Zlatibor, Serbia, 2006.
- в.15. Drndarevic D., Milivojevic M., Petrovic S., Panic S., Behavior of neural networks in a number of cycles, 34 JUPITER Conference., Beograd, Serbia, 2008. (4.24-4.27)
- в.16. Drndarevic D., Milivojevic M., Panic S., Laboratory for processes control II, 31 Conference HIPNEF 2008, V. Banja, Serbia, 2008. (475-480)
- в.17. Drndarevic D., Petrovic S., Milivojevic M., Panic S., Modeling of powder metallurgy process using neural networks, 33 Serbian Symposium for Production Engineering, Belgrade, Serbia, 2009. (211-214)
- в.18. Drndarevic D., Milivojevic M., Panić S., The control system for presses for metal powders, 32 kongres HIPNEF 2009, V. Banja, Serbia, okt. 2009. (391-396)
- в.19. Milivojevic M., Drndarevic D., Stopic S., Simic V., Stojanovic B., Modeling steel annealing process based on BP Neural Network, 3. Int. Conf. SED 2010, Uzice, Serbia, 2010.
- в.20. Drndarevic D., Milivojevic M., Petrovic S., Panić S., Application of entity characteristics in the classification of items in manufacturing systems, 36 Jupiter konf., Beograd, Serbia, May, 2010. (4.24-4.28)
- в.21. Milivojevic M., Stojanovic B., Drndarevic D., Petrovic N., Dikovic Lj., Computer modeling of steel annealing based on the theory of experiment, 3. Int. Conf. SED 2010, Uzice, Serbia, 2010.
- в.22. Dikovic Lj., Ivkovic N., Milivojevic M., Visualizing and understanding limits via Internet technologies, 3.Int. Conf. SED 2010, Uzice, Serbia, 2010.
- в.23. Drndarevic D., Milivojevic M., Petrovic S., Djuricic M., Error of neural network model in dependence on inputs, 3. Int. Conf. SED 2010, Uzice, Serbia, 2010. (2.7-2.11)
- в.24. Forst Dj., Milivojevic M., Dikovic Lj., Mijovic M., Development of integrated data management system based on GIS for cropland and agricultural machinery, International student scientific conference-2011., Nitra, Slovakia, 2011.
- в.25. Milivojevic M., Dikovic Lj., Nedovic J., Development of application software for the analysis of the flow function, numerical integration and derivation, 1'st Mathematical Conf. of Serbian Republic, Department of Mathematics and Computing, Faculty of Philosophy, University of East Sarajevo, 2011.
- в.26. Dikovic Lj., Ivkovic N., Milivojevic M., Interactive Java applets in the function of integration, 1'st Mathematical Conf. of Serbian Republic, Department of Mathematics and Computing, Faculty of Philosophy, University of East Sarajevo, 2011.

- в.27. Forst Dj., Milivojevic M., Mijovic M., Development of classes in VBA programming environment, 4. Int. Conf. SED 2011, Uzice, 2011. (2.91-2.96)
- в.28. Petkovic M., Milivojevic M., Drndarevic D., Database on the results of scientific and research works, 4. Int. Conf. SED 2011, Uzice, Serbia, 2011. (2.97-2.101)
- в.29. Marinkovic D., Milivojevic M., Drndarevic D., Gordic V., Relational model of database of orthopedic department of medical center in Uzice, 4. Int. Conf. SED 2011, Uzice, Serbia, 2011. (2.102-2.106)
- в.30. Drndarevic D., Milivojevic M., Modelling with Neural Networks, 4. Int. Conf. SED 2011, Uzice, Serbia, 2011. (3.12-3.17)
- в.31. Sokic M., Drndarevic D., Milivojevic M., Making assembly using AMINSERT to create assembly constraint between three parts with circular edges in the AutoDesk Mechanical Desktop 6 program, 4. Int. Conf. SED 2011, Uzice, Serbia, 2011. (3.1-3.4)
- в.32. Milivojevic M., Drndarevic D., Sokic M., Nikolic A., Geometrical tolerances and CAQ support, 4. Int. Conf. SED 2011, Uzice, Serbia, 2011. (3.82-3.85)
- в.33. Milivojevic M., Stojanovic B., Simic V., The Simulation of probability distribution function in queuing theory, 4. Int. Conf. SED 2011, Uzice, 2011. (2.43-2.46)
- в.34. Stevanovic B., Milivojevic M., Drndarevic D., Implementation of modular architectures in software application form multiusers, 4. Int. Conf. SED 2011 (Student section), Uzice, Serbia, Oct., 2011. (1.71-1.76)
- в.35. Drndarevic D., Milivojevic M., Modelling with Backpropagation Algorithm, Proceedings of the 1st Regional Conference Mechatronics in Practice and Education (MECH-CONF 2011), December, Subotica, Serbia, 2011. (134-138)
- в.36. Forst Dj., Milivojevic M., Dikovic Lj., Drndarevic D., Rational Agricultural Agent for Sustainable Crop Protection based on Artificial Neural Networks, International student scientific conference-2012., Nitra, Slovakia, 2012.
- в.37. Drndarevic D., Milivojevic M., Panic S., Computer antifire systems, 7th. International scientific conference: *Engineering of risk and security assessment*, Kopaonik, Serbia, febr., 2012. (97-102)
- в.38. Drndarevic D., Milivojevic M., Modeling and control using backpropagation neural network, 33 međunarodni kongres HIPNEM 2012, Beograd, Serbia, okt. 2012. (8-12)
- в.40. Drndarevic D., Milivojevic M., Petrović S., Panić S., Modelling of the spring-back of PM parts by neural networks, 38 Jupiter konf., Beograd, Mayj, 2012. (3.47-3.51)
- в.41. Drndarevic D., Milivojevic M., An approach at business articles classification, 5. Int. Conf. SED 2012, Uzice, Serbia 2012. (182-184), ISBN 978-86-83573-28-8, COBIS.SR-ID 197071372
- в.42. Dikovic Lj. Milivojevic M., The various use of mathematical technology, 5. Int. Conf. SED 2012, Uzice, Serbia, 2012. (188-190), ISBN 978-86-83573-28-8, COBIS.SR-ID 197071372
- в.43. Stopic S., Bogovic J., Schwinger A., Friedrich B., Dikovic Lj., Milivojevic M., Drndarevic D., Stojanovic B., Computer modeling of metallurgical processes by design of experiments and neural networks, 5. Int. Conf. SED 2012, Uzice, Serbia, 2012. (216-219), ISBN 978-86-83573-28-8, COBIS.SR-ID 197071372
- в.44. Drndarevic D., Milivojevic M., Modelling of dimensional changes of pm parts after sizing, 6. Int. Conf. SED 2013, Uzice, Serbia 2013. (1:47-51), ISBN 978-86-83573-39-4, COBIS.SR-ID 201568780
- в.45. Drndarevic D., Milivojevic M., Djuricic M., Laboratory for automation, 6. Int. Conf. SED 2013, Uzice, Serbia, 2013. (1:52-55), ISBN 978-86-83573-39-4, COBIS.SR-ID 201568780
- в.46. Nikolić A., Milivojević M., Puzović R., Forst Dj., Parametric modeling of tool holders for external turning in the Solid-works software, 6. Int. Conf. SED 2013, Uzice, 2013. (1:65-68), ISBN 978-86-83573-39-4, COBIS.SR-ID 201568780
- в.47. Milivojevic M., Drndarevic D., Dikovic Lj., Forst Dj., Arnold A., Stopic S., Development of software for mathematical modeling using artificial neural network, 6. Int. Conf. SED 2013, Uzice, Serbia, 2013. (2:12-21), ISBN 978-86-83573-39-4, COBIS.SR-ID 201568780
- в.48. Dikovic Lj., Milivojevic M., Bogovic J., Friedrich B., Stopic S., Stojanovic B., Jankovic B., Normal vs. Lognormal distributions in aerosol synthesis, 6. Int. Conf. SED 2013, Uzice, Serbia, 2013. (2:28-34), ISBN 978-86-83573-39-4, COBIS.SR-ID 201568780
- в.49. Milivojevic M., Dikovic Lj., Forst Dj., Stopic S., Drndarevic D., Software for determining k Nearest Neighbours using Genetic Algorithm, 6. Int. Conf. SED 2013, Uzice, Serbia, 2013. (2:35-41), ISBN 978-86-83573-39-4, COBIS.SR-ID 201568780
- в.50. Drndarevic D., Milivojevic M., Modelling of the spring-back of PM parts after compaction using backpropagation learning algorithm, 7. Int. Conf. SED 2014, Uzice, Serbia, Oct., 2014. (1.15 – 1.19), ISBN 978-86-83573-42-24, COBIS.SR-ID 209983756
- в.51. Milivojevic M., Forst D., Stopic S., Drndarevic D., Stevanovic M., Development of software for k-means klustering, 7. Int. Conf. SED 2014, Uzice, Serbia, Oct., 2014. (2.10 – 2.17), ISBN 978-86-83573-42-24, COBIS.SR-ID 209983756

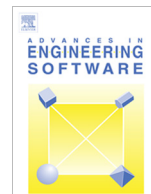
- в.52. Drndarević D., Đuričić R., Milivojević M., Application of entity characteristics in the classification of items in business, 39 Jupiter konf., Beograd, Serbia, okt., 2014. (2.1-2.6)
- в.53. Milivojević M., Diković Lj., Popović P., Forst Dj., Stevanović M., Logistic regression of students success rates in the exams, 7. Int. Conf. SED 2014, Uzice, Serbia, Oct., 2014. (2.10 – 2.17), ISBN 978-86-83573-42-24, COBIS.SR-ID 209983756
- в. 54. Drndarević D., Milivojević M., Modelling of dimensional changes of powder metallurgy parts after pressing, 2 Int. Conf. of New Technology NT-2015, Mostar, BiH, April, 2015. (170-177)
- в. 55. Milivojević M., Forst Dj., Pot M., Diković Lj., Demo software for simple genetic algorithms, 8. Int. Conf. SED 2015, Uzice, Serbia, 2-3 Oct., 2015. (2.50 – 2.57), ISBN 978-86-83573-61-5, COBIS.SR-ID 220039436
- в. 56. Pot M., Milivojević M., Tomographic reconstruction of binary matrices using two orthogonal projections, 8. Int. Conf. SED 2015, Uzice, Serbia, 2-3 Oct., 2015. (2.29 – 2.34), ISBN 978-86-83573-61-5, COBIS.SR-ID 220039436
- в. 57. Diković Lj., Stevanović S., Ivković N., Milivojević M., Teaching and investigation statistical correlation process, 8. Int. Conf. SED 2015, Uzice, Serbia, 2-3 Oct., 2015. (2.35 – 2.40), ISBN 978-86-83573-61-5, COBIS.SR-ID 220039436
- в. 58. Drndarević D., Milivojević M., Two models of a powder metallurgy process using neural networks, 8. Int. Conf. SED 2015, Uzice, Serbia, 2-3 Oct., 2015. (1.32 – 1.35), ISBN 978-86-83573-61-5, COBIS.SR-ID 220039436

Г) Пројекти

- г1. Project “CIIS (Computer Integrated Information System) of the Aluminum Mill Sevojno” in the period 1996-2000, Project leader
- г2. TEMPUS project 511044: N° 511044-1-TEMPUS-2010-1-UK-JPCR: Modernisation of Post-Graduate Studies in Chemistry and Chemistry Related Programmes (MCHEM), Registration number: 511044-2010 TEMPUS project, Project number – 5 17153-TEMPUS-1 -201 1-1-DE-TEMPUS-JPGR, CONGRAD – Conducting graduate surveys and improving alumni services for enhanced strategic management and quality improvement, Registration number: 517153 // Institutional Coordinator
- г3. TEMPUS project, No: 517200- TEMPUS-1-2011-1-BE TEMPUS-SMGR, Establishing and capacity building of the Southern Serbian Academy and National Conference for Vocational Higher Education in Serbia ECBAC, Registration number: 517200
- г4. **DAAD project: Artificial Neural Network modelling of silver nanoparticle formation after thermal decomposition of an aerosol / ANN-NanoAg**, RWTH Aachen University, IME Process Metallurgy and Metal Recycling (Project leaders Prof. Bernd Friedrich / PhD Eng. Srecko Stopic) & Institute of Mathematics and Informatics, Faculty of Science, University of Kragujevac (Project leader, PhD Boban Stojanovic)

Д) Техничка решења

- д.1. Drndarević D., Milivojević M., Radojčić N., Sokić D., Milivojević P., Identification Factors of Operations in Production System - The Project for Implementation of Integrated Production Management System of "Aluminum Mill", Sevojno, 2000.
- д.2. Milivojević M., Mijović M., Stevanović V., Computer Intergrated GIS, **Water Supply Company**, “Elan”, Kosjerić, 2007, practical implementation
- д.3. Milivojević M., Mijović M., Stevanović V., Computer Intergrated GIS, Heating Supply Company, “Gradska Toplana”, Uzice, 2009, practical implementation
- д.4. Milivojević M., Mijović M., Stevanović V., Client –Server GIS system, Heating Supply Company, “Gradska Toplana”, Uzice, 2011, practical implementation



Adaptive system for dam behavior modeling based on linear regression and genetic algorithms



B. Stojanovic^{a,*}, M. Milivojevic^b, M. Ivanovic^a, N. Milivojevic^c, D. Divac^c

^a Faculty of Science, University of Kragujevac, Radoja Domanovica 12, 34000 Kragujevac, Serbia

^b Technical and Business College, Trg Sv. Save 34, 31000 Uzice, Serbia

^c "Jaroslav Cerni" Institute for the Development of Water Resources, 11000 Belgrade, Serbia

ARTICLE INFO

Article history:

Received 25 February 2013

Received in revised form 15 May 2013

Accepted 26 June 2013

Available online 20 July 2013

Keywords:

Dam
Modeling
Linear regression
Genetic algorithms
Adaptive system
Monitoring

ABSTRACT

Most of the existing methods for dam behavior modeling require a persistent set of input parameters. In real-world applications, failures of the measuring equipment can lead to a situation in which a selected model becomes unusable because of the volatility of the independent variables set. This paper presents an adaptive system for dam behavior modeling that is based on a multiple linear regression (MLR) model and is optimized for given conditions using genetic algorithms (GA). Throughout an evolutionary process, the system performs real-time adjustment of regressors in the MLR model according to currently active sensors. The performance of the proposed system has been evaluated in a case study of modeling the Bocac dam (at the Vrbas River located in the Republic of Srpska), whereby an MLR model of the dam displacements has been optimized for periods when the sensors were malfunctioning. Results of the analysis have shown that, under real-world circumstances, the proposed methodology outperforms traditional regression approaches.

© 2013 Elsevier Ltd. All rights reserved.

1. Introduction

Dams have strong interactions with environmental, hydraulic and geomechanical factors, such as air and water temperature, water level, pore pressure, and rock deformability, each of which influences the structural behavior of the dam [1]. To describe and predict the structural behavior of dams, over the past decades, a number of deterministic, statistical and hybrid mathematical models have been developed.

For a long time, statistical models have been applied to dam safety monitoring to find out the contribution of external loads to dam displacements [2]. A number of statistical models based on multiple linear regression (MLR) and their advanced forms such as hierarchical regression, stepwise multiple regression, robust regression, ridge regression, and partial least squares regression have been shown to be more or less successful in dam modeling [3]. The advantages of the statistical models are the simplicity of formulation, the speed of execution and the availability of any type of correlation between independent and responses variables. In contrast, deterministic models require solving differential equations, for which closed form solutions could be difficult or

impossible to obtain [4]. Therefore, many models that are based on numerical methods, such as the finite element method, have been developed as well [5]. Recently, numerical and statistical methods have been enriched with various heuristics from the artificial intelligence (AI) domain, creating hybrid models that combine their advantages. Some of these artificial intelligence techniques and heuristic algorithms are artificial neural networks (ANN) [6], genetic algorithms (GA) [7], support vector machines (SVM) [8], adaptive neuro-fuzzy inference systems (ANFIS) [9,10], Monte Carlo simulations [11], the modified complex method [12] and the artificial immune algorithm (AIA) [13].

Recently, Rankovic et al. [9] presented a study in which the objective was to develop an adaptive neuro-fuzzy inference system (ANFIS) to predict the radial displacements of Bocac arch dam. ANFIS models have been proposed as an alternative approach for an evaporation estimation of the Yuvacik Dam [10]. In his paper, Mata [6] presented a comparison between MLR and ANN models for the characterization of dam behavior under environmental loads for the Alto Rabagao arch dam. In their study, Wang et al. [14] investigated several Artificial Intelligence techniques for modeling monthly river flow discharge time series, which included an ANN approach, an ANFIS technique, GP models and support vector machines (SVM), and compared their performance with traditional time series modeling techniques, such as autoregressive moving-average (ARMA) models. To improve prediction, support vector regression (SVR) upgraded with GAs is often combined with existing ANN and ARMA techniques [8]. Hybrid algorithms and their

* Corresponding author. Tel.: +381 69 11 543 75.

E-mail addresses: bobi@kg.ac.rs (B. Stojanovic), milovan.milivojevic@vpts.edu.rs (M. Milivojevic), mivanovic@kg.ac.rs (M. Ivanovic), nikola.milivojevic@gmail.com (N. Milivojevic), ddivac@eunet.rs (D. Divac).

ОБРАЗАЦ 1.

Изјава о ауторству

Потписани-а Милован Миливојевић
број уписа _____

Изјављујем

да је докторска дисертација под насловом

МЕТОДЕ РАЗВОЈА И АДАПТАЦИЈЕ РЕГРЕСИОНИХ МОДЕЛА
БАЗИРАНЕ НА ГЕНЕТСКИМ АЛГОРИТМИМА

- резултат сопственог истраживачког рада,
- да предложена дисертација у целини ни у деловима није била предложена за добијање било које дипломе према студијским програмима других високошколских установа,
- да су резултати коректно наведени и
- да нисам кршио/ла ауторска права и користио интелектуалну својину других лица.

У Крагујевцу, 26.01.2016.

Потпис аутора

М. Миливојевић

ОБРАЗАЦ 2.

Изјава о истовестности штампане и електронске верзије докторског рада

Име и презиме аутора Милован Миливојевић
Број уписа _____
Студијски програм СТУДИЈЕ ПО СТАРОМ ПРОГРАМУ
Наслов рада МЕТОДЕ РАЗВОДА И АДАПТАЦИЈЕ РЕГРЕСИОНИХ МОДЕЛА БАЗИРАНЕ НА ГЕНЕТСКИМ АЛГОРИТМИМА
Ментор Др БОБАН СТОЈАНОВИЋ

Потписани Милован Миливојевић

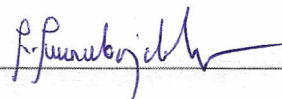
изјављујем да је штампана верзија мог докторског рада истоветна електронској верзији коју сам предао/ла за објављивање на порталу **Дигиталног репозиторијума Универзитета у Крагујевцу.**

Дозвољавам да се објаве моји лични подаци везани за добијање академског звања доктора наука, као што су име и презиме, година и место рођења и датум одбране рада.

Ови лични подаци могу се објавити на мрежним страницама дигиталне библиотеке, у електронском каталогу и у публикацијама Универзитета у Крагујевцу.

Потпис аутора

У Крагујевцу, 26.01.2016.



ОБРАЗАЦ 3.

Изјава о коришћењу

Овлашћујем Универзитетску библиотеку да у Дигитални репозиторијум Универзитета у Крагујевцу унесе моју докторску дисертацију под насловом:

МЕТОДЕ РАЗВОЈА И АДАПТАЦИЈЕ РЕГРЕСИОНИХ МОДЕЛА

БАЗИРАНЕ НА ГЕНЕТСКИМ АЛГОРИТМИМА

која је моје ауторско дело.

Дисертацију са свим прилозима предао/ла сам у електронском формату погодном за трајно архивирање.

Моју докторску дисертацију похрањену у Дигитални репозиторијум Универзитета у Крагујевцу могу да користе сви који поштују одредбе садржане у одабраном типу лиценце Креативне заједнице (Creative Commons) за коју сам се одлучио/ла.

1. Ауторство
2. Ауторство - некомерцијално
- ☒ 3. Ауторство - некомерцијално - без прераде
4. Ауторство - некомерцијално - делити под истим условима
5. Ауторство - без прераде
6. Ауторство - делити под истим условима

(Молимо да заокружите само једну од шест понуђених лиценци, чији је кратак опис дат је на обрасцу број 4.).

Потпис аутора

У Крагујевцу, 26.01.2016.

