Modeliranje voznih ciklusa uključujući vremenski promjenjive značajke nagiba ceste, mase vozila i zagušenja prometa

Topić, Jakov

Doctoral thesis / Disertacija

2022

Degree Grantor / Ustanova koja je dodijelila akademski / stručni stupanj: **University of Zagreb, Faculty of Mechanical Engineering and Naval Architecture / Sveučilište u Zagrebu, Fakultet strojarstva i brodogradnje**

Permanent link / Trajna poveznica: https://urn.nsk.hr/urn:nbn:hr:235:063660

Rights / Prava: In copyright / Zaštićeno autorskim pravom.

Download date / Datum preuzimanja: **2024-03-29**

Repository / Repozitorij:

Repository of Faculty of Mechanical Engineering and Naval Architecture University of Zagreb







Sveučilište u Zagrebu

Fakultet strojarstva i brodogradnje

Jakov Topić

MODELIRANJE VOZNIH CIKLUSA UKLJUČUJUĆI VREMENSKI PROMJENJIVE ZNAČAJKE NAGIBA CESTE, MASE VOZILA I ZAGUŠENJA PROMETA

DOKTORSKI RAD

Zagreb, 2022.



Sveučilište u Zagrebu

Fakultet strojarstva i brodogradnje

Jakov Topić

MODELIRANJE VOZNIH CIKLUSA UKLJUČUJUĆI VREMENSKI PROMJENJIVE ZNAČAJKE NAGIBA CESTE, MASE VOZILA I ZAGUŠENJA PROMETA

DOKTORSKI RAD

Mentor: prof. dr. sc. Joško Deur

Zagreb, 2022.



Sveučilište u Zagrebu

Faculty of Mechanical Engineering and Naval Architecture

Jakov Topić

MODELLING OF DRIVING CYCLES INCLUDING TIME-VARYING FEATURES OF ROAD SLOPE, VEHICLE MASS AND TRAFFIC CONGESTION

DOCTORAL THESIS

Supervisor: Joško Deur, PhD, Full Professor

Zagreb, 2022.

Podaci za bibliografsku karticu

UDK:	621.313:629.3
Ključne riječi:	vozni ciklusi, sinteza, validacija, Markovljevi lanci, regresijska analiza, odabir značajki, strojno učenje, predviđanje, neuronska mreža
Znanstveno područje:	Tehničke znanosti
Znanstveno polje:	Strojarstvo
Institucija:	Sveučilište u Zagrebu, Fakultet strojarstva i brodogradnje
Mentor:	prof. dr. sc. Joško Deur
Broj stranica:	174
Broj slika:	87
Broj tablica:	18
Broj korištenih bibliografskih jedinica:	125
Datum obrane:	28.1.2022.
Povjerenstvo:	dr. sc. Josip Kasać, red. prof., predsjednik (Fakultet strojarstva i brodogradnje, Zagreb) dr. sc. Branimir Škugor, doc., član (Fakultet strojarstva i brodogradnje, Zagreb) dr. sc. Tonči Carić, red. prof., član (Fakultet prometnih znanosti, Zagreb)
Institucija u kojoj je rad pohranjen:	Sveučilište u Zagrebu, Fakultet strojarstva i brodogradnje, Nacionalna i sveučilišna knjižnica u Zagrebu

Podaci o mentoru

Joško Deur rođen je 1966. godine u Zadru. Diplomirao je 1989. godine na Elektrotehničkom fakultetu Sveučilišta u Zagrebu, smjer Automatika. Na istom je Fakultetu magistrirao 1993. i doktorirao 1999. godine.

Od svibnja 1990. godine uposlen je na Zavodu za robotiku i automatizaciju proizvodnih sustava, Fakulteta strojarstva i brodogradnje Sveučilišta u Zagrebu. U zvanje redovitog profesora u trajnom zvanju izabran je 2018. godine. Nakon jednogodišnjeg znanstvenog usavršavanja u istraživačkom centru tvrtke Ford Motor Company u Dearbornu, SAD (2000), uspostavio je široku suradnju s tvrtkama Ford Motor i Jaguar Cars kroz više od 14 znanstveno-istraživačkih projekata na području modeliranja i regulacije automobilskih sustava. Osim toga, od 2001. godine vodio je desetak znanstveno-istraživačkih projekata poduprtih od strane Ministarstva znanosti i obrazovanja Republike Hrvatske, Hrvatske zaklade za znanosti, te Europske komisije (FP7, COST, Interreg, Horizon2020). Bio je voditelj nekoliko istraživačko-razvojnih projekata poduprtih od domaće industrije.

Znanstveno zanimanje prof. Deura obuhvaća područja automobilske mehatronike, električnih servopogona, električnih vozila i njihove integracije u energetske sustave, autonomnih vozila te sustava upravljanja općenito. Dobitnik je Državne nagrade za znanost za 2006. godinu u kategoriji godišnje nagrade za značajno znanstveno dostignuće u području tehničkih znanosti, srebrne plakete "Josip Lončar" za izrazito uspješnu doktorsku disertaciju, te pet nagrada za najbolje članke na međunarodnim konferencijama. Autor je preko 74 članaka objavljenih u međunarodnim časopisima (od toga 48 objavljena u CC/SCI-indeksiranim časopisima), preko 127 radova objavljenih na međunarodnim konferencijama, četiri patenata, šest poglavlja u međunarodnim enciklopedijama i knjigama te jedne znanstvene knjige. Bio je organizator više specijalnih sekcija i mini-simpozija na međunarodnim konferencijama, te jedne međunarodne ljetne škole.

Prof. Deur je član društava KoREMA, IEEE, SAE i IFAC, te IFAC “Technical Committee 7.1 Automotive Control“ i “IEEE Technical Committee on Automotive Control“. Član je uređivačkog odbora i/ili pridruženi urednik triju časopisa (IEEE Transactions on Vehicular Technology, Vehicle System Dynamics, International Journal of Powertrains i World Electric Vehicle Journal).

Zahvale

Prije svega zahvaljujem se mentoru prof. dr. sc. Jošku Deuru na iznimnom strpljenju i mnogim sugestijama tijekom vođenja ovog rada, te na motivaciji koja mi je pomogla provesti ovo istraživanje. Ovom prilikom zahvalio bih se i na strpljivom mentorstvu tijekom izrade diplomskog rada, kojim me je potakao na istraživački način razmišljanja i ulaganje u vlastito obrazovanje, te mi omogućio sudjelovanje na brojnim zanimljivim projektima.

Zahvaljujem se svih kolegama iz istraživačke grupe, a posebice doc. dr. sc. Branimiru Škugoru na bliskoj suradnji, izvrsnim diskusijama te mnoštvu korisnih sugestija pruženih tijekom izrade ovog rada.

Zahvaljujem se tvrtki Libertas Dubrovnik d.o.o. za omogućenje snimanja i prikupljanja podataka praćenja flote gradskih autobusa u Dubrovniku, bez kojih se ovaj rad ne bi mogao realizirati u svojoj punini.

Također se zahvaljujem Hrvatskoj zakladi za znanost koja je podržala moj istraživački rad kroz projekt "Adaptivno i prediktivno upravljanje utičnim hibridnim električnim vozilima (ACHIEVE)", te Znanstvenom centru izvrsnosti „Napredne metode i tehnologije u znanosti o podatcima i kooperativnim sustavima (DATACROSS)" u čijem radu sam sudjelovao.

Posebna hvala mojim roditeljima što su mi omogućili školovanje, te braći, sestrama, prijateljima i bližoj rodbini na svakodnevnoj moralnoj podršci.

Najveću zahvalnost želim izraziti svojoj supruzi Karmeli na riječima podrške, strpljenju i razumijevanju kroz sve godine rada, te na rođenju naše dvije prekrasne kćeri Elene i Korine koje nas uveseljavaju i upotpunjuju iz dana u dan.

Zagreb, siječanj 2022. Jakov Topić

„Poznavanje drugih je inteligencija;
poznavanje sebe je prava mudrost.
Ovladavanje drugima je snaga;
ovladavanje sobom istinska je moć.“
Lao Tze

Sažetak

Utična hibridna električna vozila (*engl. Plug-in Hybrid Electric Vehicle*, PHEV) predstavljaju jednu od ključnih tranzicijskih tehnologija prema energetski učinkovitom, čistom, tihom i održivom transportu u budućnosti. Budući da su vozni ciklusi vezani uz uvjete vožnje i ponašanje vozača, oni su od ključnog značaja za procjenu potrošnje energije i emisija štetnih plinova, te se naširoko koriste kod projektiranja i parametriranja strukture PHEV pogona i njegove upravljačke strategije. Većina studija vezanih uz PHEV kao i druge tipove vozila temelje se na certifikacijskim voznim ciklusima, koji ne uzimaju u obzir realne uvjete vožnje uključujući nagib ceste, već se temelje na "umjetno" generiranom profilu brzine vozila u ovisnosti o vremenu. Međutim, za potrebe projektiranja i ispitivanja sustava upravljanja vozilom trebali bi se koristiti realistični vozni ciklusi, koji odražavaju stvarno ponašanje u vožnji i uvjete na cesti. U tom smislu, glavni cilj ovog rada usmjeren je na sintezu i validaciju realističnih voznih ciklusa predstavljenih vremenskim profilima brzine i ubrzanja vozila te nagiba ceste, koji su izvedeni iz snimljenih GPS podataka praćenja vozila.

U radu se prvo uspostavlja stohastički model voznih ciklusa temeljen na Markovljevim lancima, koji uključuje kombinacije diskretnih vrijednosti brzine vozila, akceleracije vozila, nagiba ceste i vremenske promjene nagiba ceste kao Markovljeva stanja. U tu svrhu prikupljen je bogat skup voznih ciklusa gradskih autobusa u gradu Dubrovniku. Zatim se slučajnim uzorkovanjem višedimenzionalne matrice prijelaznih vjerojatnosti (*engl. Transition Probability Matrix*, TPM) generira veliki skup sintetičkih voznih ciklusa. Kako bi se smanjio zahtjev za memorijom, a istovremeno poboljšala računalna učinkovitost, predlaže se programska implementacija TPM u obliku rijetke matrice temeljene na rječniku ključeva. K tome, razmatra se postupak uključivanja mase putnika u sintezu voznih ciklusa temeljen na događajima zaustavljanja vozila. Zagušenje prometa uzima se u obzir kroz tvorbu zasebnih TPM za svaku kategoriju zagušenosti prometa, koja se provodi klasteriranjem snimljenih voznih ciklusa s obzirom na uspostavljene kriterije zagušenja prometa.

Za potrebe validacije sintetičkih voznih ciklusa predlaže se neuronska mreža koja za dani vozni ciklus predviđa potrošnju goriva. Kao podrška validaciji sintetičkih voznih ciklusa, uzimaju se u obzir brojne statističke značajke iz vremenske i frekvencijske domene, uključujući jedinstvene pokazatelje unakrsne korelacije brzine i ubrzanja vozila te nagiba ceste. Značaj svake nominirane statističke značajke i njezin utjecaj na potrošnju goriva raščlanjuje se pomoću linearne regresijske analize i LASSO (*engl. Least Absolute Shrinkage and Selection Operator*)

metode odabira značajki. Za jednoznačnu, jednokriterijsku validaciju sintetičkih voznih ciklusa u odnosu na skup snimljenih voznih ciklusa predlaže se regresijski model koji uključuje samo nekoliko najrelevantnijih statističkih značajki kao ulaze i potrošnju goriva kao izlaz. Također se kao potencijalno daljnje unaprjeđenje ispituje metoda validacije voznih ciklusa temeljem višestrukih kriterija. Predloženi pristupi validacije temelje se na minimiziranju odstupanja statističkih značajki sintetičkih voznih ciklusa u odnosu na očekivane vrijednosti snimljenih voznih ciklusa.

Konačno, u radu se predlažu deterministički i stohastički modeli za predviđanje brzine vozila na horizontu u budućnosti, a s tim povezanim i značajki voznih ciklusa relevantnih za potrebe modelskog prediktivnog upravljanja PHEV pogonom. Prediktivni modeli temelje se na višeslojnoj unaprijednoj neuronskoj mreži, koja se parametrira korištenjem podatka o trenutnoj poziciji vozila, trenutnoj brzini ili povijesti brzina vozila, dobu dana i dana u tjednu, a kao izlaz daje sekvencu brzina vozila na pomičnom vremenskom horizontu. Točnost predviđanja brzine vozila demonstrira se na izdvojenom skupu snimljenih voznih ciklusa.

Ključne riječi: vozni ciklusi, sinteza, validacija, Markovljevi lanci, regresijska analiza, odabir značajki, strojno učenje, predviđanje, neuronska mreža

Extended summary

Plug-in Hybrid Electric Vehicles (PHEV) represent a key enabling technology to make the transport system more efficient, cleaner, quieter, and less dependent on oil reserves. Since driving cycles are related to driving conditions and driver behaviour, they are crucial for estimating energy consumption and pollutant emissions, and they are widely used in designing and parameterising the structure of PHEV powertrain and its control strategy. Nowadays, various studies related to PHEVs, as well as other vehicle types, are usually based on certification driving cycles, which do not take into account realistic driving conditions including road slope, but are rather based on "artificially" generated vehicle velocity-time profiles. However, for the purposes of designing and testing a vehicle control system, realistic driving cycles should be used, which reflect the actual driving behavior and road conditions. In this sense, the main aim of this thesis is focused on the synthesis and validation of realistic driving cycles represented by vehicle velocity, acceleration and road slope time profiles, which are derived from recorded GPS vehicle tracking data.

The thesis first establishes a stochastic model of driving cycles based on Markov chains, which includes combinations of discrete values of vehicle velocity, vehicle acceleration, road slope and road slope time derivative as Markov states. For this purpose, a rich set of driving cycles of city buses operating in the city of Dubrovnik has been collected. By random sampling of a multidimensional transition probability matrix (TPM), a large set of synthetic driving cycles is generated. To reduce the memory requirement while improving computational efficiency, the implementation of TPM in the form of a sparse matrix based on a dictionary of keys (DOK) is proposed. In addition, including of passengers' mass into driving cycle synthesis process is considered based on vehicle stopping events. Traffic congestion is taken into account through the formation of separate TPMs for each traffic congestion category, which is carried out through clustering of recorded driving cycles based on established traffic congestion criteria.

For the purpose of synthetic driving cycles validation, a dedicated neural network (NN) is proposed, which predicts the vehicle fuel consumption for a given driving cycle. The NN approach is based on driving cycle time series input data arranged in a form of fixed-dimension histogram of counted discrete values of vehicle velocity, acceleration, and road slope. In support of validation of synthetic driving cycles, a rich set of statistical features from the time and frequency domains are considered, including unique indicators of cross-correlation of vehicle velocity, vehicle acceleration, and road slope. The significance of each nominated statistical

feature and its impact on fuel consumption is analyzed by using a linear regression analysis and Least Absolute Shrinkage and Selection Operator (LASSO) feature selection method. For unambiguous, single-criteria validation of synthetic driving cycles with respect to recorded driving cycles, a reduced-order LASSO regression model is proposed that takes only a few most significant statistical features as inputs and gives fuel consumption as output. The driving cycle validation method based on multiple criteria is also examined as a potential further improvement. The proposed validation approaches are based on minimizing the deviation of statistical features of synthetic driving cycles from the mean/expected values of recorded driving cycles.

Finally, the thesis proposes deterministic and stochastic models for predicting vehicle velocity on receding time horizon, and accordingly driving cycle features relevant for model predictive control of PHEV. Predictive models are based on a feedforward multilayer neural networks, which are trained by using the data on current vehicle position, current vehicle velocity or velocity history, time of day, and weekday, and as an output give sequence of vehicle velocities on the receding time horizon. The accuracy of vehicle velocity prediction is demonstrated on a separate set of recorded driving cycles.

The main aim of the thesis is to propose a systematic procedure of multi-dimensional driving cycle synthesis, which utilizes driving data recorded using a vehicle telemetry tracking system and at the end results in a minimum number of representative synthetic driving cycles for a wide range of driving conditions.

The thesis is organized in eight chapters, whose content is summarized in what follows.

Chapter 1, „Introduction“ outlines the motivation for the conducted research, presents the literature overview, and provides the main hypothesis and an overview of the thesis.

Chapter 2, „Recording, pre-processing and analysis of driving cycles“ describes the procedure of driving data recording by utilizing GPS/GPRS vehicle tracking devices for the case of a city bus fleet. Pre-processing of recorded driving data is carried out in terms of extracting driving cycles defined by recorded data between end stations of a certain route, filtering of driving cycles with respect to various criteria, reconstruction of precise road slope profiles using regression method based on Gaussian processes, and categorization of driving cycles in accordance to a traffic congestion index. A comprehensive analysis of interdependence of vehicle velocity, acceleration, road slope, and passengers mass is performed to determine whether the synthesis of driving cycles should be carried out jointly with respect to velocity,

acceleration, road slope and passengers mass' states or the road slope and passengers' mass can be modeled independently of velocity and acceleration.

Chapter 3, „Synthesis of multidimensional driving cycles“ proposes a method for synthesis of multidimensional driving cycles based on Markov chains. Different ways of realizing the transition probability matrix (TPM) are considered along with related implementation aspects, including computational efficiency analysis. It is proposed to adopt a stochastic driving cycle model represented by an 8D TPM using vehicle velocity, acceleration, road slope, and road slope time derivative states, and based on this model a rich set of synthetic driving cycles is generated. An independent, vehicle stopping event-based synthesis method of passengers’ mass profiles is also proposed. Traffic congestion is considered through the formation of separate TPMs for three levels of congestion (light, medium, and heavy). The chapter concludes with a preliminary analysis of validity of the generated synthetic driving cycles, based on a comparison of distributions of characteristic statistical features between synthetic and recorded driving cycles.

Chapter 4, „Regression model for predicting fuel consumption“ establishes a regression model for predicting fuel consumption based on appropriately formed input features of synthetic driving cycles and corresponding neural network (NN). The emphasis is on including a road slope feature when designing the model inputs to improve prediction accuracy. The proposed regression models are trained, validated, and tested on an augmented dataset of combined driving cycles, which are derived through random addition of recorded microcycles whose length corresponds to resolution of fuel consumption measurement. A comparative analysis of fuel consumption prediction accuracy is performed between NN and simpler, linear-in-parameter polynomial regression models, for the case with included and excluded road slope information. In addition, several methods for fine-tuning of NN models are conducted and described in detail, which are related to dimensionality reduction, finding the most appropriate architecture, determining the optimal learning rate, and regularizing the NN. Although the proposed regression model for predicting fuel consumption can be applied to various transport-related studies, in this thesis it is used for determining the fuel consumption related to a set of synthetic driving cycles generated in third chapter for the purpose of their validation.

Chapter 5, „Multi-criteria validation of driving cycles“ deals with validation of multidimensional synthetic driving cycles generated in the third chapter. A rich set of driving cycle-related statistical features are nominated to describe different driving patterns, which in addition to general statistical indices includes frequency domain ones and cross-correlation of

vehicle speed, vehicle acceleration and road slope features. A comparative analysis of the characteristic statistical features calculated for each recorded and synthetic driving cycle is performed to verify that the distributions of the statistical features for the synthetic driving cycles resembles well those obtained for the recorded driving cycles. For the needs of unambiguous validation of driving cycles, several lumped indicators of driving cycle representativeness are derived, which are based on consolidation of individual statistical features or similarity index between two vectors/matrices used to model driving cycles. The selection of the most appropriate representativeness indicator for the final validation of driving cycles is carried out based on correlation analysis in relation to fuel consumption deviation from the mean value of concatenated recorded driving cycles. Finally, a multi-criteria validation procedure is proposed, which highlights and analyzes several most representative driving cycles from the Pareto frontier of lumped representativeness indicators.

Chapter 6, „Feature selection-supported validation of driving cycles“ deals with feature selection techniques based on linear regression analysis and least absolute shrinkage and selection operator (LASSO), which serves to determine the relevance of each nominated statistical feature. The competitiveness of the fuel consumption prediction accuracy of linear and LASSO regression models of reduced order is examined in relation to baseline neural network model established in the fourth chapter. The procedure of extracting representative synthetic driving cycles based on LASSO-predicted fuel consumption is presented and compared with the Euclidean distance lumped indicator-based approach. To further improve the process of unambiguous validation of synthetic driving cycles, a dual-criterion validation method based on both Euclidean distance and LASSO-predicted fuel consumption is examined, as well.

Chapter 7, „Prediction of driving cycle features“ first describes the process of data preparation for learning, validation, and testing of a vehicle velocity prediction models. Next, static stochastic and deterministic deep feedforward neural network prediction models are proposed, where the former predicts the vehicle velocity distribution along the route, while the latter predicts the vehicle velocity profile on receding time horizon. A comprehensive analysis of the prediction accuracy of the developed predictive models is performed based on test dataset and considering different forecast quality indicators. Deterministic model is additionally examined in terms of influence of each input variable candidates (current vehicle position, current velocity or history of vehicle velocities, time of day, and weekday), velocity history length, and prediction horizon length on the accuracy of vehicle velocity profile prediction. Finally, a

dynamic stochastic model is proposed, which combines the properties of both previously developed predictive models, in terms of predicting expectations and standard deviations of velocity patterns on the receding time horizon.

Chapter 8, „Conclusion“ outlines the main results, guidelines for future research, and the following major contributions of the doctoral thesis: 1) Driving cycle model based on Markov chains, which in addition to vehicle velocity and acceleration as the Markov states accounts for the road slope, vehicle mass, and traffic congestion features; 2) Validation of driving cycle model based on recorded fuel consumption data and fuel consumption predictions obtained by using a neural network parameterized based on the driving cycle model features; 3) Stochastic model for predicting driving cycle statistical features meant to be applied in stochastic model predictive control of PHEV.

Keywords: driving cycles, synthesis, validation, Markov chains, regression analysis, feature selection, machine learning, prediction, neural network

Sadržaj

Sadržaj I
Popis slika IV
Popis tablica XII
Popis oznaka i kratica XIV
Popis oznaka XIV
Popis kratica XVIII
1. Uvod 1
1.1. Motivacija 1
1.2. Dosadašnje spoznaje 3
1.2.1. Sinteza voznih ciklusa 3
1.2.2. Validacija modela voznih ciklusa 4
1.2.3. Predviđanje značajki voznih ciklusa 6
1.3. Cilj i hipoteze istraživanja 8
1.4. Pregled rada 9
2. Prikupljanje, predobrada i analiza voznih ciklusa 12
2.1. Prikupljanje voznih ciklusa 12
2.2. Predobrada snimljenih voznih ciklusa 14
2.2.1. Izdvajanje voznih ciklusa 14
2.2.2. Filtriranje voznih ciklusa 17
2.2.3. Rekonstruiranje profila nagiba ceste 20
2.2.4. Kategorizacija voznih ciklusa 25
2.3. Analiza snimljenih voznih ciklusa 27
2.3.1. Glavne karakteristike 28
2.3.2. Međuovisnosti brzine, akceleracije i nagiba ceste 30
2.4. Akvizicija broja putnika u autobusu i pripadna analiza 33
3. Sinteza višedimenzionalnih voznih ciklusa 37
3.1. Koncept sinteze voznih ciklusa 37
3.1.1. Teorijska podloga Markovljevih lanaca 37
3.1.2. Konceptualni prikaz cjelokupnog postupka sinteze voznih ciklusa 39
3.2. Tvorba matrice prijelaznih vjerojatnosti 41
3.2.1. Koncepti realizacije 41
3.2.2. Implementacijski aspekti 45
3.2.3. Usporedna analiza računalnih performansi 48
3.2.4. Usvajanje i tvorba konačnog koncepta 50

3.3. Generiranje sintetičkih voznih ciklusa uzimajući u obzir promjenjivu masu putnika i stupanj zagušenja prometa 52
3.3.1. Koncept nezavisne sinteze profila mase putnika 52
3.3.2. Rezultati generiranja sintetičkih profila mase putnika 56
3.3.3. Rezultati generiranja sintetičkih voznih ciklusa s i bez uzimanja u obzir stupnja zagušenja prometa 58
4. Regresijski model za predviđanje potrošnje goriva 63
4.1. Motivacija i konceptualni prikaz metode 63
4.2. Priprema voznih ciklusa za parametriranje regresijskog modela 65
4.2.1. Obogaćivanje ulaznog skupa podataka 65
4.2.2. Transformacija voznih ciklusa u histograme fiksnih dimenzija 68
4.3. Modeliranje potrošnje goriva 70
4.3.1. Polinomski modeli 70
4.3.2. Teorijska podloga neuronskih mreža 71
4.3.3. Modeli temeljeni na unaprijednoj neuronskoj mreži 75
4.4. Usporedna analiza točnosti predviđanja potrošnje goriva 78
4.5. Fino-podešavanje neuronske mreže 81
4.5.1. Reduciranje dimenzionalnosti ulaza 81
4.5.2. Određivanje prikladne arhitekture 84
4.5.3. Određivanje optimalne stope učenja 86
4.5.4. Regularizacija neuronske mreže 88
5. Višekriterijska validacija voznih ciklusa 91
5.1. Statističke značajke voznih ciklusa 91
5.1.1. Nominiranje početnog skupa statističkih značajki 91
5.1.2. Usporedna analiza razdioba statističkih značajki 92
5.2. Definiranje pokazatelja reprezentativnosti voznog ciklusa 95
5.2.1. Matematička formulacija pokazatelja reprezentativnosti voznog ciklusa 96
5.2.2. Izbor najprikladnijih pokazatelja reprezentativnosti za validaciju voznih ciklusa 98
5.3. Izdvajanje reprezentativnih voznih ciklusa 100
5.3.1. Postupak višekriterijske validacije voznih ciklusa 100
5.3.2. Analiza reprezentativnih voznih ciklusa 104
6. Validacija voznih ciklusa podržana odabirom značajki 107
6.1. Izdvajanje podskupa signifikantnih statističkih značajki 108
6.1.1. Formulacija linearnog regresijskog modela 108
6.1.2. LASSO regresijska analiza 108
6.1.3. Učenje i testiranje linearnih regresijskih modela 109
6.1.4. Rezultati regresijske analize 110
6.2. Izdvajanje reprezentativnih voznih ciklusa 113

6.2.1. Definiranje naprednog pokazatelja reprezentativnosti temeljenog na LASSO regresijskom modelu 113
6.2.2. Postupak validacije voznih ciklusa temeljem regresijskog indeksa (RI) 114
6.2.3. Analiza reprezentativnih voznih ciklusa 116
6.2.4. Usporedna analiza rezultata jednoznačne i višekriterijske validacije voznih ciklusa 118
7. Predviđanje značajki voznih ciklusa 121
7.1. Statički stohastički model predviđanja brzine vozila 121
7.1.1. Modeliranje unaprijedne višeslojne neuronske mreže sa statičkim stohastičkim izlazom 121
7.1.2. Priprema podataka za učenje, validaciju i testiranje prediktivnog modela 123
7.1.3. Rezultati učenja, validacije i testiranja prediktivnog modela 126
7.2. Deterministički model predviđanja brzine vozila 131
7.2.1. Modeliranje unaprijedne višeslojne neuronske mreže s deterministčkim izlazom 131
7.2.2. Priprema podataka za učenje, validaciju i testiranje prediktivnog modela 132
7.2.3. Definiranje pokazatelja kvalitete predviđanja 133
7.2.4. Određivanje konačnog skupa ulaznih značajki 134
7.2.5. Određivanje optimalne duljine intervala povijesti brzine vozila 137
7.2.6. Analiza točnosti predviđanja u ovisnosti o duljini horizonta predviđanja 138
7.3. Dinamički stohastički model predviđanja brzine vozila 142
7.3.1. Modeliranje, učenje i validacija unaprijedne višeslojne neuronske mreže s dinamičkim stohastičkim izlazom 142
7.3.2. Analiza točnosti predviđanja prediktivnog modela 144
8. Zaključak 148
Literatura 154
Prilozi 164
Prilog A – Kandidirane značajke voznih ciklusa 164
Životopis 172
Curriculum vitae 174

Popis slika

Slika 2-1: Prikaz ruta gradskih autobusa u Dubrovniku na kojima su se provodila snimanja, zajedno s lokacijama krajnjih stanica i depoa. *14*
Slika 2-2: Primjer GPS-rekonstruiranog Di *i temeljem odometra-korigiranog* $Di*$ *profila prijeđenog puta između dva susjedna mjerenja odometra.* *17*
Slika 2-3: Rasipanje snimljenih geografskih koordinata na odabranoj ruti za smjer vožnje a) Babin kuk-Pile i b) Pile-Babin kuk. *19*
Slika 2-4: Primjeri raznih valjanih snimljenih voznih ciklusa za smjer a) Babin kuk-Pile i b) Pile-Babin kuk. *20*
Slika 2-5: Blokovski dijagram metode rekonstrukcije profila nagiba ceste duž rute temeljem snimljenih GPS podataka vožnje. *21*
Slika 2-6: a) Snimljene geografske koordinate za rutu Babin kuk-Pile zajedno sa pripadnim lokacijama polazišta i odredišta, te b) rezultati određivanja pripadnih referentnih geografskih koordinata (za broj klastera $K = 500$ *i broj referentnih točaka* $R = 5000$*).* *22*
Slika 2-7: a) Konačni profil nagiba ceste rekonstruiran za odabranu rutu uporabom regresijske metode temeljene na GP-u, zajedno s individualnim profilima nagiba ceste za smjer b) Babin kuk-Pile i c) Pile-Babin kuk. *25*
Slika 2-8: Krivulja ovisnosti ukupne inercije $E_{\sum}$ *u odnosu na broj klastera, uključujući odabrani optimalni broj klastera* $K*$ *prema Elbow metodi.* *27*
Slika 2-9: Rezultati klasteriranja voznih ciklusa u odnosu na srednju brzinu i broj zaustavljanja vozila po kilometru. *27*
Slika 2-10: Razdiobe a) prijeđenih udaljenosti, b) trajanja vožnje i c) prosječne brzine po putovanju za oba smjera vožnje na ruti Babin kuk – Pile, uključujući prosječne vrijednosti. *28*
Slika 2-11: Toplinske mape rasipanja snimljenih brzina vozila duž rute za oba smjera vožnje u domeni a), c) prijeđene udaljenosti i b), d) vremena, uključujući uprosječene vrijednosti brzina (crvena linija). *29*
Slika 2-12: Histogrami snimljenih uzoraka a) brzine vozila, b) akceleracije vozila i c) nagiba ceste za objedinjene podatke oba smjera vožnje na ruti Babin kuk-Pile. *30*
Slika 2-13: Toplinske mape a) pozitivnog i b) negativnog ubrzanja vozila s obzirom na brzinu vozila za objedinjenu rutu BK-P-BK, uključujući po brzini uprosječene vrijednosti ubrzanja vozila. *31*

Slika 2-14: Toplinske mape brzine vozila s obzirom na a) negativne i b) pozitivne vrijednosti nagiba ceste za objedinjenu rutu BK-P-BK, uključujući po nagibu uprosječene vrijednosti brzine vozila. *32*
Slika 2-15: Razdioba pozitivnog i negativnog ubrzanja vozila u odnosu na brzinu vozila za a), c) pozitivne i b), d) negativne vrijednosti nagiba ceste za objedinjenu rutu BK-P-BK. *33*
Slika 2-16: Objedinjena razdioba pozitivnog ubrzanja vozila u odnosu na brzinu vozila za pozitivne vrijednosti nagiba ceste. *33*
Slika 2-17: a) Vremenski profil broja putnika u autobusu na odabranoj ruti dobiven prebrojavanjem, te izvedeni histogrami b) broja putnika u autobusu i c) razlike broja putnika između dviju stanica. *34*
Slika 2-18: Matrica prijelaza broja putnika između susjednih stanica. *35*
Slika 2-19: Razdioba snimljenih akceleracija vozila u odnosu na brzinu vozila i masu putnika. *36*
Slika 3-1: Blokovski dijagram metode sinteze voznih ciklusa temeljene na Markovljevim lancima, uključujući osnove validacije voznih ciklusa. *40*
Slika 3-2: Prikaz različitih koncepata (modela) realizacije Markovljevog lanca u smislu dimenzionalnosti matrice prijelaznih vjerojatnosti Π *43*
Slika 3-3: Koncept modela temeljenog na hijerarhijskoj sintezi voznih ciklusa (ML-HIER) koji se sastoji od nezavisne $\Pi N4D$ *i zavisne* $\Pi Z5D$ *matrice prijelaznih vjerojatnosti.* *45*
Slika 3-4: Ilustracija a) 2D , b) 4D i c) 6D TPM-a implementiranog u obliku niza. *46*
Slika 3-5: Ilustracija a) Markovljevog lanca i b) pripadne 6D TPM implementirane u obliku rijetke matrice zasnovane na rječniku ključeva (DOK). *47*
Slika 3-6: Dijagram toka koncepta nezavisne sinteze profila mase putnika temeljene na događaju zaustavljanja vozila. *54*
Slika 3-7: Histogrami i pripadni kutijasti dijagrami snimljenih a) vremena stajanja autobusa, te b) prijeđene udaljenosti između dva zaustavljanja autobusa, uključujući postavljene vrijednosti pragova Dst,min *i* $Tst,min.$ *55*
Slika 3-8: Primjeri sintetičkih profila brzine vozila i mase putnika za slučajeve a) niske, b) srednje i c) visoke popunjenosti autobusa putnicima. *57*
Slika 3-9: Histogrami zabilježenih i sintetiziranih a) vrijednosti mase putnika i b) razlika u masi putnika između dvaju zaustavljanja vozila. *57*
Slika 3-10: Histogrami a) prosječnih vrijednosti i b) standardnih devijacija mase putnika, izračunatih za svaki individualni sintetički vozni ciklus ms,i*, uključujući vrijednosti izračunate za kombinirani sintetički te snimljeni profil mase putnika.* *58*

Slika 3-11: Razdiobe značajki zagušenja prometa sintetičkih i snimljenih voznih ciklusa uključujući pripadne srednje vrijednosti značajki, za slučajeve a) sveukupnog skupa podataka i b) uzimajući u obzir informaciju o stupnju zagušenosti prometa. 60
Slika 3-12: Reprezentativni primjeri vremenskih profila brzine vozila za a), c), e) snimljene i b), d), f) sintetičke vozne cikluse dobivene za klastere lagane (zelena boja), srednje (plava boja) i teške zagušenosti prometa (crvena boja). Oznaka BZPK odnosi se na broj zaustavljanja vozila po kilometru..... 61
Slika 3-13: Nasumično odabrani primjeri sintetičkih voznih ciklusa generiranih ne uzimajući u obzir stupanj zagušenja prometa: a), c), e) profili brzine u vremenu, te b), d), f) pripadni profili nagiba ceste u ovisnosti o prijeđenoj udaljenosti. 62
Slika 3-14: Histogrami snimljenih i sintetičkih uzoraka a) brzine vozila, b) ubrzanja vozila i c) nagiba ceste..... 62
Slika 4-1: Konceptualni prikaz metode predviđanja potrošnje goriva temeljem različitih familija regresijskih modela..... 65
Slika 4-2: Ilustracija postupka razdvajanja snimljenog voznog ciklusa na skup mikrociklusa ekvivalentne potrošnje goriva od 0.5 L za slučaj a) fine i b) grube razlučivosti snimljene potrošnje goriva. 67
Slika 4-3: Ovisnost a) trajanja vožnje i b) potrošnje goriva za snimljene vozne cikluse, izdvojene mikrocikluse i kombinirane vozne cikluse (sintetizirane iz mikrociklusa). 67
Slika 4-4: Razdioba a) prosječne brzine i b) srednje pozitivne i negativne akceleracije vozila za snimljene vozne cikluse, izdvojene mikrocikluse i kombinirane vozne cikluse. 68
Slika 4-5: Potrošnja goriva u odnosu na a) prijeđenu udaljenost i b) prosječni nagib ceste za kombinirane vozne cikluse (one koji sadrže snimljene mikrocikluse)..... 68
Slika 4-6: Primjer voznog ciklusa transformiranog u H2D predstavljenog u a) 3D i b) 2D obliku..... 70
Slika 4-7: Ilustracije polinomskih modela za predviđanje potrošnje goriva: a) model $Poly_{1D}$ i b) model $Poly_{2D}$, zajedno sa snimljenim podacima..... 71
Slika 4-8: Model statičkog (umjetnog) neurona..... 72
Slika 4-9: Ilustracija unaprijedne višeslojne neuronske mreže. 73
Slika 4-10: Predložene arhitekture neuronskih mreža za predviđanje potrošnje goriva, na temelju a) 3D histograma – H3D i b) 2D histograma – H2D vezanog za vozni ciklus. 76
Slika 4-11: Krivulje pogreške učenja i validacije kroz epohe učenja za NN-H2D (puna linija) i NN-H3D (isprekidana linija). 78

Slika 4-12: Predviđene u odnosu na stvarne potrošnje goriva dobivene za različite regresijske modele na temelju testnog skupa podataka. 79
Slika 4-13: Razdiobe a) apsolutnih i b) relativnih pogrešaka predviđanja za različite regresijske modele i slučaj testnog skupa podataka. 79
Slika 4-14: Rijetkost dva različita tipa ulaza u NN, a) H2D koji sadrži podatke o brzini i ubrzanju vozila, i b) H3D koji proširuje H2D informacijom o nagibu ceste. 82
Slika 4-15: Profil sačuvane varijance u ovisnosti o broju glavnih komponenti. 83
Slika 4-16: Krivulje pogreški učenja i validacije kroz epohe učenja za NN-PCA (puna linija) i NN-H3D (isprekidana linija). 84
Slika 4-17: Arhitekture NN s varijabilnim brojem skrivenih slojeva i pripadnih neurona koje se razmatraju prilikom pronalaska najprikladnijeg modela potrošnje goriva. 85
Slika 4-18: a) Ovisnost RMSE potrošnje goriva o arhitekturi NN definirane brojem skrivenih slojeva i pripadnih neurona, te b) pripadna toplinska mapa (brojke unutar zagrada odnose se na ukupni broj neurona u modelu). Cijan bojom označena je najprikladnija arhitektura NN modela (NN-OPT). 86
Slika 4-19: Ilustracija postupka određivanja optimalne stope učenja primjenom metode LRRT u obliku dijagrama toka. 87
Slika 4-20: Krivulja ovisnosti funkcije gubitka o stopi učenja dobivena primjenom LRRT metode. 88
Slika 4-21: Rezultati regularizacije NN-H3D modela: a) krivulje pogreške učenja i validacije i b) razlika pogreške učenja i validacije u ovisnosti o vrijednosti faktora regularizacije λ. Obojana skala (ΔJV) označava porast validacijske pogreške u odnosu na vrijednost dobivenu bez regularizacije. 90
Slika 5-1: Razdiobe vjerojatnosti za odabrani skup statističkih značajki $\mathcal{FR}$ izračunatih za snimljene i sintetičke vozne cikluse, zajedno s očekivanim vrijednostima odgovarajućih kombiniranih voznih ciklusa. 94
Slika 5-2: Histogram relativnih odstupanja očekivanih vrijednosti svih 100 razmatranih statističkih značajki ($\Delta frel, m$) za kombinirani sintetički u odnosu na snimljeni vozni ciklus, zajedno s pripadnom prosječnom vrijednosti ($\Delta frel$) i standardne devijacije ($\Delta frel\sigma$). 95
Slika 5-3: Primjer aproksimacijske krivulje definirane normalnom razdiobom $\mathcal{N}i$ čiji se parametri $\mu i, \sigma i$ uče na originalnoj razdiobi reziduala statističkih značajki $\mathcal{N}i$. 98
Slika 5-4: Rezultirajuća korelacijska matrica prikazana u obliku toplinske mape za sve razmatrane pokazatelje reprezentativnosti voznog ciklusa (plavom bojom označen je redak korelacije svakog pokazatelja u odnosu na FCD). 99

Slika 5-5: Ovisnosti pokazatelja FCD o a) ED, b) MAE, c) STD, d) SI-TPM, te e) SI-H3D pokazateljima (svaka točka na grafu odnosi se na jedan sintetički vozni ciklus). *100*
Slika 5-6: Blokovski dijagram postupka višekriterijske validacije voznih ciklusa..... *101*
Slika 5-7: Ovisnost broja validnih sintetičkih voznih ciklusa Nr *u podskupovima* $V1$, $V2$ *te* $V3$ *o graničnim vrijednostima pokazatelja reprezentativnosti definiranim p-tim percentilom*... *103*
Slika 5-8: Konačni podskup $V2$ *od* $Nr = 15$ *reprezentativnih sintetičkih voznih ciklusa (obojani simboli) za* $p = 5$ *te uži skup* $Nr = 10$ *za slučaj* $V3$ *(simboli s magenta obrubom), dobiveni primjenom višekriterijske metode validacije voznih ciklusa s obzirom na FCD, ED, MAE te STD pokazatelje reprezentativnosti*..... *103*
Slika 5-9: a), c), e) Profili brzine vozila u vremenu te b), d), e) nagiba ceste u ovisnosti o prijeđenoj udaljenosti za tri višekriterijskom validacijom-izdvojena reprezentativna sintetička vozna ciklusa označena kao #2037, #603 i #1696 na slici 5-8, uključujući snimljene profile nagiba ceste za oba smjera vožnje (označene isprekidanim crtama). *104*
Slika 6-1: Blokovski dijagram predloženog postupka validacije voznih ciklusa temeljenog na odabiru značajki. *107*
Slika 6-2: Krivulje kvalitete predviđanja LASSO modela opisane R^2 *vrijednostima u ovisnosti o vrijednosti faktora regularizacije* λ*, s obzirom na koje je određena optimalna vrijednost* $\lambda*$*.* *110*
Slika 6-3: Usporedba a) predviđene u odnosu na stvarnu potrošnju goriva i b) razdiobe reziduala predviđanja potrošnje goriva za LR, LASSO i NN-H3D regresijske modele i testni skup podataka...... *111*
Slika 6-4: Vrijednosti koeficijenata regresije βj *za slučaj a) LR i b) LASSO modela punog i reduciranog reda; te postignute* $R2$ *vrijednosti c)LR i d) LASSO modela u ovisnosti o broju ulaznih značajki tj. parametara.*..... *112*
Slika 6-5: Ovisnosti pokazatelja odstupanja potrošnje goriva (FCD) o pokazateljima reprezentativnosti voznih ciklusa a) ED i b) RI za slučaj 3000 sintetičkih voznih ciklusa, uključujući pripadne vrijednosti indeksa korelacije ($\rho ED, FCD$, $\rho RI, FCD$*).* *114*
Slika 6-6: Pozicije reprezentativnih sintetičkih voznih ciklusa dobivenih temeljem a) jednoznačne validacije prema kriteriju minimalne vrijednosti pokazatelja RI i ED ($\psi RI*$, $\psi ED*$*) te b) minimalne udaljenosti od ishodišta RI – ED koordinatnog sustava (*$\psi*$*).* *115*
Slika 6-7: a), c) e) Profili brzine vozila u vremenu te b), d), e) nagiba ceste u ovisnosti o prijeđenoj udaljenosti za tri reprezentativna sintetička vozna ciklusa izdvojena odabirom značajki podržanom validacijom te označena kao $\psi RI*$ *(#2022),* $\psi ED*$ *(#1347), te* $\psi*$

(#1241), uključujući snimljene profile nagiba ceste za oba smjera vožnje (označene isprekidanim crtama). *116*
Slika 6-8: Pozicije svih šest reprezentativnih sintetičkih voznih ciklusa izdvojenih temeljem postupka jednoznačne ili višekriterijske validacije u ED-FCD koordinatnom sustavu. *120*
Slika 6-9: Razdiobe relativnih odstupanja statističkih značajki ($\Delta frel,m$) za svih šest reprezentativnih sintetičkih voznih ciklusa izdvojenih temeljem postupka jednoznačne ili višekriterijske validacije, zajedno s pripadnim prosječnim vrijednostima ($\Delta frel$) i standandardnim devijacijama ($\Delta frel\sigma$), izračunate za cjelokupan skup nominiranih statističkih značajki $\mathcal{F}$ *120*
Slika 7-1: Predložena arhitektura statičke stohastičke unaprijedne višeslojne neuronske mreže za predviđanje rasipanja brzine vozila duž rute, vj, na temelju pozicije vozila dk (NN-STC). *122*
Slika 7-2: Ilustracija postupka kodiranja trenutne brzine vozila vk u vektor $Ck \in \mathbb{R}1\ x\ J$ koji predstavlja oznaku tj. ciljanu vrijednost pri učenju NN-STC modela. *123*
Slika 7-3: Prikaz izdvojenih slučajnih uzoraka brzine vozila vn na primjeru jednog snimljenog voznog ciklusa, odnosno pripadnom profilu brzine vozila u domeni a) vremena i b) pozicije. *125*
Slika 7-4: Histogram ukupnog broja izdvojenih uzoraka po voznom ciklusu, izračunat za skup svih validnih snimljenih voznih ciklusa. *125*
Slika 7-5: Toplinske mape rasipanja snimljenih brzina vozila vk duž rute u domeni pozicije vozila, sadržanih u skupu podataka za a) učenje, b) validaciju i c) testiranje NN-STC modela, uključujući po poziciji uprosječene vrijednosti brzine vozila (crvena linija). *126*
Slika 7-6: Krivulje pogreške učenja (crna linija) i validacije (zelena linija) kroz epohe učenja za NN-STC model. *127*
Slika 7-7: a) Predviđene srednje vrijednosti brzine vozila μk dobivene primjenom NN-STC modela na temelju testnog skupa podataka u odnosu na snimljene vrijednosti brzine vozila vk, te b) pripadna toplinska mapa. *128*
Slika 7-8: Primjeri usporedbe predviđenih i snimljenih profila brzine vozila za tri vozna ciklusa prikazani u domeni pozicije (lijevi stupac) i vremena (desni stupac). *128*
Slika 7-9: Predviđeni i stvarni profili srednje vrijednosti brzine vozila, zajedno s pripadnim intervalima pouzdanosti i rasipanjem snimljenih vrijednosti brzine vozila, sve dobiveno za testni skup podataka. *129*

Slika 7-10: Ovisnost predviđenih očekivanih o stvarnim srednjim vrijednostima brzine vozila (gornji red) kao i predviđenih u odnosu na stvarne standardne devijacije (donji red). *130*
Slika 7-11: Histogrami odstupanja predviđenih u odnosu na stvarne a) srednje brzine i b) standardne devijacije brzine. *130*
Slika 7-12: Predložena arhitektura determinističke unaprijedne višeslojne neuronske mreže (NN-DET) za predviđanje profila brzine vozila na horizontu u budućnosti vp,k *na temelju trenutne pozicije vozila* dk*, trenutne brzine* vk *(ili povijesti brzina* vh,k*), doba dana* tk *i dana u tjednu* wk..... *131*
Slika 7-13: Ilustrativni prikaz postupka generiranja ulazno-izlaznih podataka (tj. skupova $\mathcal{X}$ *i* $\mathcal{Y}$*) za učenje, validaciju i testiranje NN-DET modela.* *133*
Slika 7-14: Rezultirajuće vrijednosti funkcije gubitka $RMSEv$ *dobivene za svaku kombinaciju ulaznih značajki NN-DET-*I_n *modela te za skupove podataka za učenje (crna krivulja) i validaciju (plava krivulja).* *136*
Slika 7-15: Predviđeni profili brzine vozila (crvena isprekidana linija) u odnosu na snimljene vrijednosti (crna puna linija), prikazani na primjeru jednog testnog voznog ciklusa i za svaki NN-DET-I_n *model zasebno.* *136*
Slika 7-16: Vrijednosti funkcije gubitka $RMSEv$ *dobivene za različite duljine ulaznog vektora povijesti brzina (*Hh*) NN-DET-*H_h *modela te za skupove podataka za učenje (crna krivulja) i validaciju (crvena krivulja).* *138*
Slika 7-17: Vrijednosti pogreške testiranja $RMSEv$ *za različite duljine horizonta predviđanja* Hp *i različite prediktivne modele.* *139*
Slika 7-18: Razdiobe vrijednosti pokazatelja $RMSEk$ *dobivenih za sva individualna predviđanja profila brzine vozila i danih za različite duljine horizonta predviđanja* Hp*.* *140*
Slika 7-19: Karakteristične krivulje $R2$ *vrijednosti za svaki* j*-ti diskretni korak predviđanja duž horizonta duljine* Hp *i testni skup podataka.* *141*
Slika 7-20: Predložena arhitektura dinamičke stohastičke unaprijedne višeslojne neuronske mreže (NN-DSTC) za predviđanje razdioba brzine vozila na horizontu u budućnosti na temelju trenutne pozicije vozila dk*, povijesti brzina* vh,k*, doba dana* tk *i dana u tjednu* wk..... *142*
Slika 7-21: Krivulje pogreške učenja i validacije kroz epohe učenja za NN-DSTC model. .. *144*
Slika 7-22: Predviđene očekivane vrijednosti μk *u odnosu na stvarne vrijednosti* vk *brzine vozila za slučaj a) NN-DSTC i b) NN-DET modela. Predviđanja za svaki* j*-ti vremenski korak duž horizonta predviđanja označena su zasebnom bojom.* *145*

Slika 7-23: a) Predviđanja značajki brzine vozila ($\mu k, j$, $\sigma k, j$) za slučaj jednog testnog voznog ciklusa, uključujući b) uvećani segment profila brzine vozila. *146*
Slika 7-24: Ovisnost točnosti predviđanja NN-DSTC modela o diskretnom koraku duž horizonta *predviđanja $j = 1,2, \dots, Hp$, izražena kroz a) kutijasti dijagram standardne devijacije σj, b) prosječni $RMSEj$, te c) $R2$ vrijednost, uključujući međuovisnosti pokazatelja d) $RMSE$ i e) $R2$ o σj; gdje ρ predstavlja odgovarajuće indekse korelacije.*..... *147*

Popis tablica

Tablica 2-1: Podjela izdvojenih voznih ciklusa prema smjeru vožnje i razlučivošću snimljene potrošnje goriva. *16*
Tablica 2-2: Podjela filtriranih voznih ciklusa prema smjeru vožnje i razlučivosti snimljene potrošnje goriva. *19*
Tablica 2-3: Udjeli odbačenih voznih ciklusa prema individualnim kriterijima filtriranja. ... *19*
Tablica 2-4: Indeksi korelacije između brzine vozila, pozitivnog i negativnog ubrzanja vozila te pozitivnog i negativnog nagiba ceste (polja koja sadrže zamjetne korelacije označena su zelenom bojom). *31*
Tablica 3-1: Memorijski zahtjevi za različite dimenzije i pristupe implementacije TPM-a te grube i fine razlučivosti stanja (za potonje su rezultati dani u zagradama). *49*
Tablica 3-2: Prosječna vremena potrebna za generiranje jednog sintetičkog voznog ciklusa za različite dimenzije i pristupe implementacije TPM-a te grube i fine razlučivosti stanja (za potonje su rezultati dani u zagradama). *49*
Tablica 3-3: Broj neuspjelih pokušaja generiranja voznih ciklusa za različite dimenzije i pristupe implementacije TPM-a te grube i fine razlučivosti stanja (za potonje su rezultati dani u zagradama). *50*
Tablica 4-1: Pokazatelji kvalitete predviđanja potrošnje goriva za različite modele izračunati na testnom skupu podataka. *81*
Tablica 4-2: Pokazatelji kvalitete predviđanja potrošnje goriva za NN-H3D i NN-PCA modele izračunati na testnom skupu podataka. *84*
Tablica 5-1: Parametri Volvo 7900 konvencionalnog autobusa. *92*
Tablica 5-2: Očekivane vrijednosti statističkih značajki iz skupa osam odabranih značajki $\mathcal{FR}$, izračunate za kombinirani snimljeni fz,m i sintetički fs,m vozni ciklus, zajedno s pripadnim relativnim odstupanjima značajki $\Delta frel,m$. *95*
Tablica 5-3: Vrijednosti odabranih statističkih značajki iz limitiranog skupa $\mathcal{FR}$ za kombinirani snimljeni vozni ciklus ψz i odabrane reprezentativne sintetičke vozne cikluse označene kao #2037, #603 i #1696. ..... *105*
Tablica 6-1: Pokazatelji kvalitete predviđanja potrošnje goriva za LR i LASSO regresijske modele te NN-H3D model izračunati na testnom skupu podataka. *111*

Tablica 6-2: Konačno usvojen podskup najsignifikantnijih značajki LASSO regresijskog modela, uključujući pripadne vrijednosti koeficijenta regresije βj i minimalne te maksimalne vrijednosti odgovarajućih značajki (izračunate prije min-max normalizacije). *112*

Tablica 6-3: Vrijednosti odabranih statističkih značajki iz limitiranog skupa $\mathcal{FL}$ za kombinirani snimljeni vozni ciklus ψz i odabrane reprezentativne sintetičke vozne cikluse označene kao $\psi RI*(\#2022)$, $\psi ED*(\#1347)$, te $\psi*(\#1241)$. *118*

Tablica 7-1: Statistika odstupanja predviđenih u odnosu na stvarne srednje vrijednosti brzine i standardne devijacije brzine. *131*

Tablica 7-2: Rezultati učenja, validacije i testiranja NN-DET-H_h modela, uključujući broj podesivih parametara modela. *138*

Tablica 7-3: Statistika razdioba pokazatelja kvalitete individualnih predviđanja $RMSEk$ za različite duljine horizonta predviđanja Hp, za slučaj NN-DET-$H_{p,a}$ modela (crna boja) i NN-DET-$H_{p,b}$ modela (crvena boja). *141*

Popis oznaka i kratica

Popis oznaka

Oznaka	Jedinica	Opis
a	m/s^2	Ubrzanje vozila
A_f	m^2	Čeona površina vozila
BZPK	km^{-1}	Broj zaustavljanja vozila po prijeđenom kilometru
C_d	-	Koeficijent aerodinamičkog otpora
$\boldsymbol{C}_k$	-	Binarni vektor koji označava pripadnost određenoj klasi brzine vozila
d	-	Prijeđena udaljenost
$D_{st,min}$	m	Minimalna zahtijevana prijeđena udaljenost između dviju zaustavljanja vozila
$\bar{E}_{vc}^{+}$	MJ/km	Srednja specifična energija vožnje po prijeđenom kilometru
ED_i	-	Euklidska udaljenost između statističkih značajki
$\boldsymbol{f}_s$	-	Vektor izračunatih vrijednosti statističkih značajki za sintetički vozni ciklus
$\boldsymbol{f}_z$	-	Vektor izračunatih vrijednosti statističkih značajki za snimljeni vozni ciklus
$\mathcal{F}$	-	Skup nominiranih statističkih značajki voznih ciklusa
F_w	N	Zahtijevana sila na kotaču vozila
FCD_i	L/100 km	Apsolutno odstupanje potrošnje goriva sintetičkog voznog ciklusa od očekivane vrijednosti snimljenih podataka
g	m/s^2	Gravitacijska konstanta
H_h	s	Vremenski interval povijesti brzina vozila
H_p	s	Diskretni vremenski horizont predviđanja
K	-	Broj klastera korišten u algoritmu K-sredina

J	-	Funkcija gubitka
$\tilde{J}$	-	Regularizirana funkcija gubitka
m	kg	Masa vozila
$\bar{M}_s$	-	Skup sintetičkih profila mase putnika
$\bar{M}_z$	-	Snimljeni profil mase putnika
MAE_i	-	Srednja apsolutna pogreška statističkih značajki
MBP	-	Maksimalan deklariran broj putnika u autobusu
N_{DC}	-	Ukupni broj snimljenih voznih ciklusa
N_{DC}^{V}		Ukupni broj validnih snimljenih voznih ciklusa
$N_{DC,BK-P}$	-	Ukupni broj snimljenih voznih ciklusa za smjer Babin kuk - Pile
$N_{DC,P-BK}$	-	Ukupni broj snimljenih voznih ciklusa za smjer Pile - Babin kuk
$N_{DC,BK-P}^{V}$	-	Ukupni broj validnih snimljenih voznih ciklusa za smjer Babin kuk - Pile
$N_{DC,F}^{V}$	-	Ukupan broj snimljenih voznih ciklusa s finom razlučivošću potrošnje goriva
$N_{DC,G}^{V}$	-	Ukupan broj snimljenih voznih ciklusa s grubom razlučivošću potrošnje goriva
$N_{DC,P-BK}^{V}$	-	Ukupni broj validnih snimljenih voznih ciklusa za smjer Pile – Babin kuk
N_s^K	-	Ukupni broj generiranih sintetičkih voznih ciklusa koji uključuju kategorizaciju prema stupnju zagušenja prometa
N_s^M	-	Ukupni broj generiranih sintetičkih profila mase putnika
N_s^R	-	Ukupni broj generiranih sintetičkih voznih ciklusa koji ne uključuju kategorizaciju prema stupnju zagušenja prometa
N_s^U	-	Ukupni broj generiranih sintetičkih voznih ciklusa

NLL	-	Funkcija cilja određena negativnom log-vjerojatnosti koja maksimizira vjerojatnost ispravno predviđenih kategorija brzine vozila
$p_{i,j}$	-	Prijelazne vjerojatnosti (između Markovljevog stanja *i* u Markovljevo stanje *j*)
$\widehat{\boldsymbol{p}}_k$	-	Vjerojatnosna razdioba brzine vozila predviđena od strane stohastičkog (NN-STC) modela
$\mathbb{P}$	-	Uvjetna vjerojatnost
P_m	W/kg	Specifična snaga vozila
P_w	W	Zahtijevana snaga na kotaču vozila
PMP	kg	Prosječna masa putnika
R_0	-	Koeficijent otpora kotrljanja
R^2	-	Koeficijent determinacije
R_c^2	-	Korigirani koeficijent determinacije
RI_i	-	Regresijski indeks izveden iz regresijskih koeficijenata dobivenih za slučaj LASSO regresijskog modela reduciranog reda
SI_i	-	Kosinusova sličnost između dviju ulaznih vektora/matrica
STD_i	-	Standardna devijacija aproksimacijske krivulje reziduala statističkih značajki *i*-tog sintetičkog voznog ciklusa
t	s	Diskretni vremenski trenutak
$\bar{T}_p$	s	Prosječno vrijeme izvođenja (predviđanja) modela
$T_{st,min}$	s	Minimalno zahtijevano vrijeme stajanja vozila prilikom zaustavljanja
v	km/h	Brzina vozila
$\bar{v}$	km/h	Srednja brzina vozila
v_h	km/h	Horizontalna komponenta brzine vozila
v_v	km/h	Vertikalna komponenta brzine vozila

V	-	Konačni podskup validnih (reprezentativnih) sintetičkih voznih ciklusa
V_f	L	Apsolutna potrošnja goriva
$\bar{V}_f$	L/100 km	Prosječna potrošnja goriva
$\boldsymbol{W}$	-	Podesivi težinski koeficijenti tj. parametri neuronske mreže
X_{GPS}, Y_{GPS}	°	Snimljene geografske koordinate vozila (geografska duljina, geografska širina)
λ	-	Faktor regularizacije
β_j	-	Koeficijenti linearnog i LASSO regresijskog modela
$\hat{\beta}_j$	-	Estimirane vrijednosti koeficijenata linearnog i LASSO regresijskog modela
μ	-	Srednja/očekivana vrijednost uzoraka
σ	-	Standardna devijacija uzoraka
η	-	Stopa učenja
$\bar{\boldsymbol{\psi}}_s$	-	Kombinirani sintetički vozni ciklus
$\bar{\boldsymbol{\psi}}_z$	-	Kombinirani snimljeni vozni ciklus
Ψ_z	-	Skup snimljenih voznih ciklusa
Ψ_s		Skup sintetičkih voznih ciklusa
Ψ_{LZ}	-	Skup snimljenih voznih ciklusa koji pripadaju kategoriji laganog zagušenja prometa
Ψ_{SZ}	-	Skup snimljenih voznih ciklusa koji pripadaju kategoriji srednjeg zagušenja prometa
Ψ_{TZ}	-	Skup snimljenih voznih ciklusa koji pripadaju kategoriji teškog zagušenja prometa
Ψ_{LZ}^{s}	-	Skup sintetičkih voznih ciklusa koji pripadaju kategoriji laganog zagušenja prometa
Ψ_{SZ}^{s}	-	Skup sintetičkih voznih ciklusa koji pripadaju kategoriji srednjeg zagušenja prometa

Ψ_{TZ}^{S}	-	Skup sintetičkih voznih ciklusa koji pripadaju kategoriji teškog zagušenja prometa
Ψ_{R}^{S}	-	Skup sintetičkih voznih ciklusa s isključenom kategorizacijom prema stupnju zagušenja prometa
θ	°	Nagib ceste
$\dot{\theta}$	°/s	Vremenska derivacija nagiba ceste
ρ	-	Koeficijent korelacije
ρ_z	kg/m^3	Gustoća zraka
$\mathbf{\Pi}$	-	Matrica prijelaznih vjerojatnosti

Popis kratica

ADAM	Algoritam optimiranja neuronske mreže koji se temelji na adaptivnoj procjeni momenta (*engl. Adaptive Moment Estimation*)
BEV	Baterijsko električno vozilo (*engl. Battery Electric Vehicle*)
CSV	Vrijednosti razdvojene zarezom (*engl. Comma Separated Value*)
DOK	Zapis rijetke matrice u obliku rječnika ključeva (*engl. Dictionary of Keys*)
ED	Euklidska udaljenost (*engl. Euclidean Distance*)
FCD	Odstupanje potrošnje goriva (engl. *Fuel Consumption Deviation*)
FS	Tehnike odabira značajki (*engl. Feature Selection*)
GP	Gaussov proces (*engl. Gaussian Process*)
GPRS	Paketna, bežična podatkovna komunikacijska usluga (*engl. General Packet Radio Service*)
GPS	Globalni položajni/satelitski radionavigacijski sustav (*engl. Global Positioning System*)
LASSO	Operater najmanjeg apsolutnog skupljanja i odabira (engl. Least Absolute Shrinkage and Selection Operator)
LR	Linearni regresijski model

LRTT	Test raspona stope učenja (*engl. Learning Rate Range Test*)
LSTM	Neuronska mreža s dugom kratkoročnom memorijom (*engl. Long Short-Term Memory*)
MAE	Srednja apsolutna pogreška (*engl. Mean Absolute Error*)
ML	Strojno učenje (*engl. Machine Learning*)
MLE	Metoda procjene maksimalne vjerojatnosti (*engl. Maximum Likelihood Estimation*)
MLP	Unaprijedna višeslojna perceptron neuronska mreža (*engl. Multilayer Perceptron*)
MPC	Modelsko prediktivno upravljanje (*engl. Model Predictive Control*)
MSE	Srednja kvadratna pogreška (*engl. Mean Squared Error*)
NLL	Negativna log-vjerojatnost (*engl. Negative Log-Likelihood*)
NLML	Negativna log-marginalna-vjerojatnost (*engl. Negative Log Marginal Likelihood*)
NN	Neuronska mreža
PCA	Metoda za reduciranje dimenzionalnosti temeljena na analizi glavnih komponenti (*engl. Principal Component Analysis*)
PHEV	Utično hibridno električno vozilo (*engl. Plug-in Hybrid Electric Vehicle*)
ReLU	Vrsta aktivacijske funkcije kod neuronskih mreža (*engl. Rectified Linear Unit*)
RI	Regresijski indeks korišten kao pokazatelj reprezentativnosti prilikom validacije sintetičkih voznih ciklusa
RMSE	Korijen srednje kvadratne pogreške (*engl. Root Mean Square Error*)
SI	Indeks sličnosti korišten kao pokazatelj reprezentativnosti prilikom validacije voznih ciklusa
STD	Pokazatelj reprezentativnosti temeljen na aproksimaciji originalne razdiobe reziduala statističkih značajki s normalnom (Gaussovom) razdiobom
TPM	Matrica prijelaznih vjerojatnosti (*engl. Transition Probability Matrix*)

1. Uvod

1.1. Motivacija

Tijekom posljednja dva desetljeća povećanje ekološke osviještenosti predstavlja ključni pokretač uvođenja strožih propisa o smanjenju štetnih emisija vozila. Primjerice, Europska unija (EU) obvezala se smanjiti emisije stakleničkih plinova iz cestovnog prometa za 60% do 2050. godine u usporedbi s razinama iz 1990. godine [1]. Potrošnja energije vozila, a time i emisije štetnih plinova ovise o individualnom profilu vožnje (tzv. voznom ciklusu), kao i o karakteristikama vozila, vremenskim uvjetima, vrsti goriva te stupnju opterećenja pomoćnih uređaja [2-4]. Vozni ciklus uobičajeno se definira kao profil brzine vozila u funkciji vremena [5, 6], te kao takav predstavlja karakteristični uzorak vožnje koji odražava ponašanje vozača (npr. agresivnost vožnje) i stanje okoline (vrsta ceste, nagib ceste, prometni uvjeti, gustoća naseljenosti i sl.) [7, 8]. Počevši od ranih 60-ih godina, vozni ciklusi u mnogim zemljama svijeta služe kao standardizirana osnova za mjerenje emisija štetnih i stakleničkih plinova te potrošnje goriva tijekom postupka certifikacije [4, 5]. Postoje dva tipa voznih ciklusa [9]: (i) modalni, koji uključuju dugotrajna razdoblja pri konstantnim brzinama, te (ii) tranzijentni, koji uključuju mnoge promjene brzine što je tipično primjerice za gradsku vožnju. Modalni vozni ciklusi izvode se teorijski, dok su tranzijentni rezultat izravnog mjerenja uzorka vožnje koji se smatra reprezentativnim, gdje je reprezentativnost definirana kao stupanj do kojeg razmatrani vozni ciklus odražava stvarnu vožnju [10]. Za potrebe certifikacijskih ispitivanja kao i raznih analiza vezanih uz projektiranje strukture i upravljačke strategije pogona vozila, vozni ciklusi bi u čim većoj mjeri trebali predstavljati stvarna ponašanja vozača i uvjete na cesti [11].

Iako su propisani standardi uspješni u kontinuiranom smanjenju potrošnje goriva i emisija stakleničkih plinova, uočen je značajni jaz između certifikacijskih, tj. laboratorijskih (dinamometarskih) ispitivanja i potrošnje goriva u realnim uvjetima vožnje, tj. na cesti [12, 13]. Na temelju izvješća objavljenog 2013. godine od strane Međunarodnog vijeća za čisti prijevoz (*engl. International Council on Clean Transportation*, ICCT) [14], koje je prvo razmatralo prekomjernu emisiju NO_x plinova i zatražilo istragu, službene emisije CO_2 proizvođača automobila od 2001. do 2015. smanjile su se za približno 30% (sa 170 g/km na 120 g/km). Međutim, stvarne emisije CO_2 na cesti smanjene su samo za približno 10%, sa 183 g/km na 167 g/km. Ova razlika objašnjava se većim dijelom razlikama između certifikacijskih i stvarnih voznih ciklusa [12].

Kako bi se točnije procijenila potrošnja goriva i emisije ispušnih plinova razvijaju se novi postupci ispitivanja, koji uobičajeno uključuju primjenu novih standardiziranih voznih ciklusa s većim obuhvatom utjecajnih faktora (npr. inicijalna temperatura, opterećenje pomoćnih uređaja i sl.) [15]. Na primjer, Europska komisija odlučila je od rujna 2017. uvesti globalno usklađeni testni postupak za laka vozila (*engl. Worldwide Harmonised Light Vehicle Test Procedure*, WLTP), zamjenjujući prethodni postupak temeljen na NEDC (*engl. New European Driving Cycle*) modalnom voznom ciklusu [16, 17]. WLTP predstavlja certifikacijski tranzijentni vozni ciklus koji odražava 756.000 kilometara podataka stvarne vožnje prikupljenih u raznim dijelovima svijeta (SAD, Europa, Indija, Japan, itd.) [18]. Nadalje, uveden je test stvarne emisije vožnje (*engl. Real Driving Emissions*, RDE) [19], kojim se mjere emisije ispušnih plinova u stvarnim uvjetima vožnje, kako bi se regulirale emisije NO_x i štetnih čestica.

Alternativni pristupi unaprjeđenju procjene potrošnje goriva i emisija štetnih plinova, u koje se ulažu sve veći napori, temelje se na zamjeni standardiziranih voznih ciklusa sa statistički reprezentativnim voznim ciklusima izvedenim iz snimljenih, GPS (*engl. Global Positioning System*) podataka praćenja vozila [20, 21]. Sinteza takvih voznih ciklusa uglavnom se izvodi integriranjem snimljenih podataka višestrukih putovanja u jednu cjelinu kako bi se obuhvatili različiti uzorci vožnje. U tu se svrhu obično koriste stohastičke metode temeljene na spajanju mikro-ciklusa [22, 23] ili Markovljevim lancima [24-26]. Budući da je primjenom ovih postupaka moguće generirati neograničeni skup sintetičkih voznih ciklusa različitih karakteristika, ostaje pitanje kako procijeniti njihovu reprezentativnost, tj. koju validacijsku metodu i s tim povezane statističke značajke voznih ciklusa treba koristiti za izdvajanje jednog ili nekoliko najreprezentativnijih voznih ciklusa [27, 28].

S druge strane, kako bi proizvođači vozila ispunili sve strože zakonske regulative vezane uz potrošnju goriva i emisije štetnih plinova, posebna pažnja posvećuje se razvijanju učinkovitijih pogonskih sklopova temeljenih na alternativnim oblicima energije (npr. električna energija) [29]. Iako potpuno električna vozila poznata i kao baterijska električna vozila (*engl. Battery Electric Vehicle*, BEV) predstavljaju krajnji cilj za ostvarenje budućeg čišćeg i energetski učinkovitog transportnog sustava [30], utična hibridna električna vozila (*engl. Plug-in Hybrid Electric Vehicle*, PHEV) prepoznata su kao dobro prijelazno rješenje [31], koje zaobilazi trenutne probleme vezane uz primjenu BEV (visoka cijena baterije, dugo vrijeme punjenja te ograničen domet vozila) [32]. S druge strane, PHEV imaju složeniju strukturu pogona temeljenu na više izvora snage, zahtijevaju ći tako složenije upravljačke strategije za različite načine rada pogona i režime vožnje [33].

Projektiranje pogonske konfiguracije i upravljačke strategije PHEV pogona može biti učinkovitije ako su ispitni vozni ciklusi realističniji te ako je vozna ruta poznata unaprijed i/ili ako su statističke značajke voznih ciklusa poznate iz prethodno provedenih analiza transportnog sustava [34, 35]. Unaprijedno znanje voznih ruta je često dostupno (barem ugrubo) u transportnim sustavima dostavnog i javnog gradskog prijevoza zbog poznatih rasporeda vožnje [36]. Ovo vozila tih sustava čini pogodnim kandidatima za primjenu adaptivnih ili čak prediktivnih upravljačkih strategija koje se temelje na poznavanju ili čak predviđanju statističkih značajki voznih ciklusa na pomičnom vremenskom horizontu [37]. Budući da učinkovitost i praktičnost prediktivnih upravljačkih strategija upravljanja uvelike ovise o kvaliteti predviđanja budućih profila brzine vozila, nameće se potreba za razvojem preciznih, robusnih i računalno učinkovitih prediktivnih modela voznih ciklusa [38].

1.2. Dosadašnje spoznaje

1.2.1. Sinteza voznih ciklusa

Vozni ciklusi bilježe ponašanje vozača, lokalne prometne parametre i uvjete na cesti u obliku skupa podatkovnih točaka koje predstavljaju brzinu vozila u funkciji vremena [4], te se naširoko koriste iz perspektive procjene potrošnje goriva i emisija štetnih plinova za konvencionalna i hibridna cestovna vozila, tj. potrošnje električne energije za potpuno električna vozila [2, 3, 5]. Pored ovih primjena, vozni ciklusi mogu se koristiti za odabir optimalnih konfiguracija pogonskog sklopa vozila i razvoj optimalnih strategija upravljanja vozilom [39, 40], kao i za određivanje dometa vozila, očekivanog vijeka trajanja baterije i optimalne strategije upravljanja punjenjem u slučaju električnih vozila [41-44]. U te se svrhe uobičajeno koriste umjetno generirani, certifikacijski vozni ciklusi [45, 46]. Najčešće korišteni certifikacijski vozni ciklusi, dani redom od dinamičnijih prema stacionarnijim jesu: (i) Urban Dynamometer Driving Schedule (UDDS), (ii) Federal Highway Driving Schedule (FHDS), (iii) Aggressive Supplemental Federal Test Procedure (US06), (iv) New European Driving Cycle (NEDC), te (v) Air Conditioning Supplemental Federal Test Procedure (SC03) [40]. Budući da certifikacijski vozni ciklusi mogu rezultirati pristranim procjenama i analizama (vidi diskusiju u poglavlju 1.1), nedavni istraživački napori usmjereni su na razvoj statistički reprezentativnih sintetičkih voznih ciklusa izvedenih iz snimljenih podataka GPS praćenja vozila [20, 22, 24, 25].

Glavni cilj sinteze voznih ciklusa je zamijeniti široki skup snimljenih voznih ciklusa s nekoliko sintetičkih ciklusa koji su reprezentativni u statističkom smislu, tj. odražavaju prosječne karakteristike polaznog skupa. Najučestaliji pristup sintezi zasnovan je na metodi Markovljevih lanaca [27, 28, 47, 48], čija se glavna prednost odnosi na fleksibilnost u generiranju neograničenog broja sintetičkih ciklusa vožnje u smislu vremena trajanja ili ukupne prijeđene udaljenosti [20, 21]. Pored toga, ova metoda posjeduje inherentno svojstvo rješavanja problema nedostajućih podataka (npr. zbog povremenih gubitaka GPS signala), koje se svodi na izostavljanje neregularnih prijelaza između stanja pri određivanju modela voznih ciklusa u formi matrice prijelaznih vjerojatnosti (*engl. Transition Probability Matrix*, TPM). Većina studija koje se bave sintezom voznih ciklusa temeljem metode Markovljevih lanaca koriste brzinu i akceleraciju vozila za definiranje Markovljevog stanja [20, 24], zanemarujući pritom nagib ceste. Time se zanemaruje utjecaj energije potrebne za svladavanje nagiba ceste, te energije koja se regenerira tijekom kočenja ili vožnje nizbrdo u slučaju električnih vozila [25, 40]. Dostupne studije koje uključuju nagib ceste kao Markovljevo stanje dijele nedostatak implementacije TPM-a u obliku N-dimenzionalnog niza (tj. tenzora) [28, 47, 48], što zbog visoke dimenzionalnosti TPM-a rezultira prekomjernim zahtjevom za računalnom memorijom, kao i niskom računalnom učinkovitošću. Potencijalno dodatno stanje Markovljevog lanca je masa vozila, koja je naročito relevantna za dostavna vozila te za vozila javnog gradskog prijevoza (npr. za kamione i autobuse) zbog značajne varijabilnosti mase tereta i putnika te posljedičnog utjecaja na potrošnju goriva [49].

1.2.2. Validacija modela voznih ciklusa

Budući da se postupak sinteze voznog ciklusa temelji na stohastičkom modeliranju i generiranju, koristeći TPM i generator slučajnih brojeva, moguće je izgenerirati neograničen broj sintetičkih voznih ciklusa [50]. Stoga su potrebni određeni validacijski kriteriji i postupci kako bi se iz čitavog skupa sintetičkih voznih ciklusa izdvojili oni najreprezentativniji u statističkom smislu [51, 52].

Pristup validaciji sintetičkih voznih ciklusa obično se temelji na korištenju niza signifikantnih statistički značajki neovisnih o vozilu te tipično povezanih s brzinom vozila, akceleracijom vozila, brojem zaustavljanja i sl. [21, 27, 48, 50]. Pritom se dozvoljeni prag odstupanja svake statističke značajke od referentne vrijednosti uglavnom postavlja na opseg od ±10-15% [21, 27, 47]. Glavni nedostatak ovog pristupa je što s porastom broja razmatranih statističkih značajki

raste broj potrebnih sintetičkih voznih ciklusa koje je potrebno generirati kako bi se pronašao vozni ciklus koji zadovoljava dani uvjet za svaku statističku značajku [52].

Kako bi se izbjegla višekriterijska provjera valjanosti, često se iz mnoštva statističkih značajki izvodi jedinstveni konsolidirani pokazatelj reprezentativnosti voznog ciklusa, npr. zbroj apsolutnih odstupanja vrijednosti statističkih značajki od ciljanih vrijednosti [28]. Međutim, ovdje treba imati na umu da sve statističke značajke nemaju jednaku važnost iz perspektive reprezentativnosti voznog ciklusa. Stoga ostaje pitanje kako odrediti stupanj relevantnosti svake značajke u mnoštvu mogućih kandidata. Kako bi se odgovorilo na ovo pitanje, posebnu pozornost treba posvetiti metodama strojnog učenja ili, točnije, tehnikama odabira značajki [53]. Odabir značajki postupak je odabira podskupa relevantnih značajki (prediktorske varijable ili ulazi u model) iz početnog skupa značajki prilikom razvijanja modela predviđanja zavisne/mjerene varijable (izlaza iz modela) [54]. Glavni razlozi korištenja tehnike odabira značajki sa stajališta modeliranja su tipično [55]: (i) pojednostavljivanje modela radi poboljšanja interpretabilnosti putem uklanjanja značajki koje su ili nebitne ili suvišne tj. redundantne, uz istovremeno ublažavanje prenaučenosti modela; (ii) smanjenje računalne složenosti modela; te (iii) smanjenje dimenzionalnosti problema.

Postoji nekoliko vrsta metoda koje se obično koriste za odabir značajki, a kategoriziraju se kao metode filtra, metode omotača i ugrađene metode [56]. Dodatna kategorizacija odnosi se na to jesu li značajke odabrane na temelju zavisne varijable (nadzirane metode; *engl. supervised learning*) ili skupa nezavisnih varijabli (nenadzirane metode; *engl. unsupervised learning*). Metode filtra koriste statističke tehnike za procjenu odnosa između svake nezavisne/ulazne varijable i zavisne varijable, gdje se dotične ocjene/rangovi koriste kao osnova za odabir (filtriranje) nezavisnih varijabli koje će se koristiti u modelu. Metode omotača najprije izdvajaju različite podskupove ulaznih značajki i na svakom podskupu uče novi model. Potom se kao najrelevantniji podskup značajki odabire onaj koji rezultira najtočnijim modelom prema određenom pokazatelju kvalitete predviđanja. Ugrađene metode automatski izvršavaju odabir značajki kao dio procesa učenja modela (npr. modeli temeljeni na stablu ili penalizirani regresijski modeli). Primjer ovog pristupa je LASSO metoda (*engl. Least Absolute Shrinkage and Selection Operator*) [57], koja vrši odabir značajki skupljanjem, tj. L_1 regularizacijom koja funkciji najmanjeg kvadratnog gubitka koju treba minimizirati dodaje kaznu jednaku zbroju apsolutnih vrijednosti koeficijenata regresije. Alternativni pristup odabiru značajki temelji se na linearnoj regresijskoj analizi [27], koja se oslanja na odbacivanje nepotrebnih značajki prema kriteriju najnižih vrijednosti koeficijenata regresije i statističkom T-testu.

Ostali pristupi validaciji voznih ciklusa oslanjaju se na jedinstveni pokazatelj reprezentativnosti, koji se temelji na specifičnoj potrošnji energije vozila [58] ili na podudaranju razdioba brzine i ubrzanja vozila (*engl. Speed Acceleration Frequency Distribution*, SAFD) za sintetičke i snimljene vozne cikluse [4, 6]. Iako jednostavni u svojoj prirodi, glavni nedostatak ovih pristupa je u tome što ne uzimaju u obzir sveobuhvatan prostor značajki i stoga ne jamče primjeren stupanj reprezentativnosti voznog ciklusa s obzirom na sve statističke značajke. Na primjer, zanemarivanjem značajki frekvencijskog odziva, rezultirajući profili brzine i nagiba ceste mogu odstupati od snimljenih uzoraka u smislu broja i trajanja zaustavljanja vozila i učestalosti promjena nagiba ceste, a što sve može imati bitan utjecaj na potrošnju goriva [59]. Također, ukoliko se provodi združena sinteza profila brzine, akceleracije i nagiba ceste, za opsežniju validaciju potrebno je uzeti u obzir i značajke unakrsne korelacije koje kvantificiraju međuovisnost između brzine, ubrzanja i nagiba ceste [60].

1.2.3. Predviđanje značajki voznih ciklusa

Projektiranje sustava optimalnog upravljanja HEV/PHEV pogonom s ciljem minimiziranja potrošnje goriva u širokom rasponu voznih ciklusa predstavlja značajan izazov [61, 62]. Na primjer, globalna optimizacija potrošnje energije HEV/PHEV, koja se uobičajeno provodi primjenom algoritma dinamičkog programiranja, zahtijeva poznavanje značajki čitavog voznog ciklusa unaprijed, dakle na početku putovanja [63]. U praksi se ovo reducira na optimiranje na pomičnom vremenskom horizontu, tipično kroz modelsko prediktivno upravljanje (*engl. Model Predictive Control*, MPC). Kako je već spomenuto, glavne značajke voznih ciklusa izvode se iz profila brzine vozila u funkciji vremena [64], kojeg je stoga potrebno predviđati na pomičnom horizontu. U predviđanju brzine vozila postoji značajna inherentna nesigurnost te složenost, zbog nepredvidivih čimbenika voznog okruženja poput zagušenja prometa, utjecaja semafora, prometnih znakova, pješaka i sl. [65-67].

Nedavna istraživanja naprednih upravljačkih strategija HEV/PHEV pogonima usmjerena su k adaptivnim i prediktivnim strategijama, pri čemu se prvotna metoda temelji na adaptaciji upravljačkih parametara s obzirom na estimaciju voznih značajki povijesnog uzorka vožnje, a potonja na kratkoročnim predviđanjima profila brzine vozila u budućnosti [68]. Za potrebe adaptacije upravljačke strategije PHEV, analize stila vožnje i dr., u literaturi se razmatra kategorizacija voznih ciklusa (tzv. prepoznavanje uzorka vožnje), koja se uobičajeno provodi s obzirom na srednju brzinu i akceleraciju vozila kako bi se odredio stil vožnje (npr. mekan, regularan ili agresivan) i tip ceste (npr. vožnja gradskim cestama ili autocestama [69, 70]).

Osnovno načelo prepoznavanja uzorka vožnje je uzorkovanje i izdvajanje statističkih značajki iz povijesti vožnje uz pretpostavku da će ostati nepromijenjeni u neposrednoj budućnosti, nakon čega se fino-podešeni klasifikatori koriste za prepoznavanje trenutnih uzorka vožnje [71-73]. Osim adaptacije s obzirom na značajke voznog ciklusa, upravljačka strategija može također biti adaptirana s obzirom na predviđanja tipa ceste i zagušenja prometa [74].

Drugi relevantan pristup odnosi se na MPC, koji se i inače široko koristi u automobilskoj mehatronici [75]. MPC algoritam provodi optimiranje upravljačkih trajektorija na pomičnom vremenskom horizontu, te kao takav pretpostavlja poznavanje budućeg profila brzine vozila i korištenje pripadnog modela pogona vozila [66, 76]. Zbog neizvjesnosti i složenosti realnih voznih ciklusa, predviđanje brzine vozila teško je precizno provesti primjenom tradicionalnih pristupa predviđanja vremenskih nizova zasnovanih na linearnim regresijskim modelima poput ARIMA modela (*engl. AutoRegressive Integrated Moving Average*) [68]. Stoga se je potrebno orijentirati k podacima-vođenim prediktivnim modelima temeljenim na strojnom učenju, koje karakterizira automatsko učenje značajki iz dostupnih podataka [77]. Najčešće korišteni modeli odnose se na umjetne neuronske mreže (NN), kao univerzalne aproksimatore sposobne preslikati nelinearne karakteristike složenog sustava s pomoću nelinearne aktivacijske funkcije [78]. Povrh toga, pristupi temeljeni na strojnom učenju nemaju ograničenje na broj ulaznih varijabli za razliku od alternativnih modela temeljenih na Markovljevim lancima [68], gdje veličina TPM-a raste eksponencijalno s porastom broja ulaznih varijabli. Glavni izazov se pritom odnosi na određivanje arhitekture NN, za što ne postoje egzaktne metode [79].

U literaturi se za predviđanje profila brzine vozila koriste različiti modeli, počevši od jednostavnih eksponencijalnih modela preko Markovljevih lanaca pa do neuronskih mreža [76]. Na primjer, u [67] autori koriste geografske koordinate, brzinu vozila, datum i vrijeme na početku putovanja te povijesne podatke vožnje iz prometne baze podataka kao ulaze u duboku neuronsku mrežu koja predviđa brzinu na određenom segmentu ceste. Međutim, unatoč bogatom skupu ulaznih varijabli, dobivene su grube točnosti predviđanja zbog nedostatne količine podataka za učenje i testiranje prediktivnog modela. Kako bi poboljšali točnost predviđanja brzina, autori u [80] koriste sekvence povijesnih brzina, ubrzanja i prijeđenih udaljenosti kao ulaz u prediktor brzine zasnovan na LSTM (*engl. Long Short-Term Memory*) NN. Nešto drugačiji pristup korišten je u [81], gdje se predlaže napredni E-LLM alat (*engl. Evolutionary Least Learning Machine*) za predviđanje sekvence brzine na danom vremenskom horizontu i izvještava o iznimnoj točnosti predviđanja, ali za vrlo ograničeni skup od pet voznih ciklusa. U [82], autori sustavno uspoređuju performanse različitih prediktivnih modela, koje

dijele u dvije kategorije: (i) determinističke modele, koji predviđaju očekivane vrijednosti brzine te (ii) stohastičke modele, koji predviđaju čitavu razdiobu brzina. Za parametriranje prediktivnih modela koriste se brojni ulazni podaci, uključujući interne podatke vozila (podaci s CAN sabirnice) i eksterne podatke poput pozicija semafora i znakova zaustavljanja te nadolazećih zavoja (radar i prometna baza podataka). U [83] sustavno se uspoređuju različite izvedbe multivarijantnih Markovljevih lanaca te se predlaže pristup temeljen na višestrukim TPM-ovima, gdje svaka TPM^t predviđa sljedeće stanje udaljeno *t* koraka u budućnosti.

Nadalje, u nekim se studijama predviđanje brzine vozila izvodi u domeni pozicije [84, 85], budući da su nesigurnosti vezane uz geometriju ceste (zavoji), semafore, stajališta i sl. izravno povezane s domenom pozicije, a ne vremena. Pritom se tipično kombiniraju različiti prediktivni modeli. Na primjer, u [84] se Markovljevi lanci kombiniraju s NN, te se navodi da u usporedbi s pojedinačnim modelima predviđanja, kombinirani model u prosjeku poboljšava točnost predviđanja za približno 25%. Zanimljiv pristup predložen je u [85], gdje se profili brzina generiraju korištenjem stohastičkog modela temeljenog na Markovljevom lancu, uz ograničenja na brzinu vozila određenih primjenom statističkih metoda na snimljene podatke vožnje duž razmatrane rute. Nedostatak postupaka formuliranih u domeni pozicije jest da se za primjenu u MPC sustavima zahtijevaju vremenski profili predviđanja brzine, pri čemu postupak transformacije profila iz domene puta u vrijeme nije jednoznačan, posebno u točkama singulariteta vezanim uz trenutke zaustavljanja vozila.

Ispitivanje prediktivnih modela tipično se provodi u smislu točnosti predviđanja, vremena izvođenja i kvalitete prediktivnog upravljanja [68, 76]. Na primjer, u [86] se poboljšanja u smanjenju potrošnje goriva PHEV uspoređuju za sljedeća tri slučaja: (i) poznavanje punog voznog ciklusa, (ii) savršena predviđanja brzine na pomičnom horizontu od 10 sekundi, te (iii) stvarna predviđanja brzine na pomičnom horizontu od 10 sekundi dobivena primjenom LSTM NN. Kao pokazatelji kvalitete predviđanja uglavnom se koriste metrike poput srednje kvadratne pogreške (*engl. Root Mean Square Error*; RMSE), srednje apsolutne pogreške (*engl. Mean Absolute Error*, MAE) i R^2 vrijednosti između stvarnog i predviđenog profila brzine [68].

1.3. Cilj i hipoteze istraživanja

Cilj istraživanja provedenog u okviru ovog rada je razrada postupaka koji će temeljem voznih podataka snimljenih pomoću telemetrijskog sustava praćenja vozila rezultirati minimalnim brojem reprezentativnih sintetičkih voznih ciklusa za širok raspon uvjeta vožnje.

Temeljna hipoteza rada je da se proširenjem matrice prijelaznih vjerojatnosti mogu pored učinka brzine vozila modelirati i učinci promjenjivog nagiba ceste i mase vozila te značajke zagušenja prometa, te da se temeljem tog modela može izvesti stohastički model predviđanja statističkih značajki voznih ciklusa potrebnih za prediktivne upravljačke strategije šire skupine hibridnih električnih vozila.

1.4. Pregled rada

Rad je organiziran u osam poglavlja, koja se pored uvoda i zaključka odnose na sustavnu predobradu i analizu snimljenih voznih ciklusa, sintezu bogatog skupa višedimenzionalnih voznih ciklusa, modeliranje potrošnje goriva primjenom neuronske mreže, odabir signifikantnih značajki voznih ciklusa, validaciju sintetičkih voznih ciklusa te predviđanje značajki voznih ciklusa. U nastavku je dan opis pojedinih poglavlja.

Drugo poglavlje: *Prikupljanje, predobrada i analiza voznih ciklusa***.** Opisuje se postupak snimanja podataka vožnje uz primjenu GPS/GPRS uređaja za praćenje flote gradskih autobusa. Predobrada snimljenih podataka vožnje provodi se u smislu izdvajanja voznih ciklusa između krajnjih autobusnih stanica određene rute, filtriranja voznih ciklusa s obzirom na razne kriterije, rekonstrukcije pripadnog profila nagiba ceste primjenom regresijske metode temeljene na Gaussovim procesima, te kategorizacije voznih ciklusa s obzirom na stupanj zagušenja prometa. Provodi se sveobuhvatna analiza međuovisnosti brzine, akceleracije, nagiba ceste i mase putnika kako bi se ustanovilo treba li sintezu voznih ciklusa provoditi združeno s obzirom na stanja brzine, ubrzanja, nagiba ceste i mase putnika ili se nagib ceste i masa putnika mogu modelirati neovisno o brzini i akceleraciji.

Treće poglavlje: *Sinteza višedimenzionalnih voznih ciklusa***.** Predlaže se sinteza višedimenzionalnih voznih ciklusa primjenom stohastičke metode Markovljevih lanaca. Razmatraju se različiti načini realizacije matrice prijelaznih vjerojatnosti i pripadni implementacijski aspekti uključujući analizu računalne učinkovitosti. Predlaže se usvajanje stohastičkog modela voznih ciklusa koji je reprezentiran 8D matricom prijelaznih vjerojatnosti za stanja brzine i akceleracije vozila te nagiba i derivacije nagiba ceste, te se temeljem tog modela generira bogat skup sintetičkih voznih ciklusa. Također se razmatra uključivanje mase putnika u postupak sinteze voznih ciklusa. Zagušenje prometa uzima se u obzir kroz tvorbu zasebnih matrica prijelaznih vjerojatnosti za tri stupnja prigušenja (lagano, srednje i teško). Poglavlje se zaključuje s preliminarnom analizom validnosti generiranih sintetičkih voznih

ciklusa, temeljenoj na usporedbi razdioba i očekivanja uobičajenih statističkih značajki sintetičkih i snimljenih voznih ciklusa.

Četvrto poglavlje: *Regresijski model za predviđanje potrošnje goriva***.** Predlaže se regresijski model za predviđanje potrošnje goriva temeljem prikladno formiranih ulaznih značajki sintetičkih voznih ciklusa te odgovarajuće neuronske mreže. Naglasak je na uključivanju značajke nagiba ceste pri oblikovanju ulaza u model radi poboljšanja točnosti predviđanja u odnosu na model koji se zasniva samo na značajkama brzine i akceleracije vozila. Predloženi regresijski modeli parametriraju se, validiraju i testiraju na obogaćenom skupu kombiniranih voznih ciklusa, izvedenih slučajnim nadovezivanjem snimljenih mikro-ciklusa čija duljina odgovara razlučivosti mjerenja potrošnje goriva. Provodi se usporedna analiza točnosti predviđanja potrošnje goriva primjenom neuronske mreže u odnosu na jednostavnije linearne regresijske modele, s uključenom i isključenom informacijom o nagibu ceste. Opisuju se i primjenjuju metode za reduciranje dimenzionalnosti ulaza, određivanje najprikladnije arhitekture, određivanje optimalne stope učenja, te regularizaciju neuronske mreže. Iako se predloženi model za predviđanje potrošnje goriva može primijeniti za razne transportne studije, u ovom se radu koristi za potrebe određivanja potrošnje goriva za skup sintetičkih voznih ciklusa generiranih u trećem poglavlju, a u svrhu njihove validacije.

Peto poglavlje: *Višekriterijska validacija voznih ciklusa***.** Predlaže se postupak validacije višedimenzionalnih sintetičkih voznih ciklusa generiranih u trećem poglavlju. Kako bi se obuhvatili različiti obrasci vožnje, nominira se skup od 100 statističkih značajki voznih ciklusa, uključujući značajke vremenske i frekvencijske domene te unakrsne korelacije brzine vozila, ubrzanja vozila i nagiba ceste. Provodi se komparativna statistička analiza karakterističnih statističkih značajki izračunatih za svaki snimljeni i sintetički vozni ciklus pojedinačno, kako bi se provjerilo nalikuju li distribucije statističkih značajki za sintetičke vozne cikluse onima dobivenima za snimljene vozne cikluse. Za potrebe jednoznačne validacije voznih ciklusa izvodi se nekoliko združenih pokazatelja reprezentativnosti voznih ciklusa, koji se temelje na kombiniranju individualnih statističkih značajki ili na indeksu sličnosti dviju matrica kojima se modeliraju vozni ciklusi. Korelacijskom analizom združenih pokazatelja reprezentativnosti u odnosu na točnost predviđanja potrošnje goriva izdvajaju se najprikladniji pokazatelji za finalnu validaciju sintetičkih voznih ciklusa. Konačno, predlaže se postupak višekriterijske validacije kojim se na Pareto fronti združenih pokazatelja reprezentativnosti izdvaja i analizira nekoliko najprikladnijih voznih ciklusa.

Šesto poglavlje: *Validacija voznih ciklusa podržana odabirom značajki***.** Određuje se relevantnost kandidiranih statističkih značajki na temelju linearne regresijske analize, kao i tehnike odabira značajki temeljene na operateru najmanjeg apsolutnog skupljanja i odabira (LASSO). Ispituje se konkurentnost točnosti predviđanja potrošnje goriva linearnog i LASSO regresijskog modela reduciranog reda u odnosu na referentni model temeljen na neuronskoj mreži, uspostavljen u četvrtom poglavlju. Predstavlja se postupak izdvajanja reprezentativnih sintetičkih voznih ciklusa na temelju predviđene potrošnje goriva primjenom LASSO modela te se uspoređuje s pristupom temeljenim na Euklidskoj udaljenosti odstupanja statističkih značajki od ciljanih vrijednosti dobivenih iz skupa snimljenih voznih ciklusa. S ciljem daljnjeg unaprjeđenja postupka jednoznačne validacije sintetičkih voznih ciklusa dodatno se ispituje dvo-kriterijska metoda validacije temeljena na Euklidskoj udaljenosti i LASSO kriteriju predviđanja potrošnje goriva.

Sedmo poglavlje: *Predviđanje značajki voznih ciklusa***.** Opisuje se postupak pripreme podataka za učenje, validaciju i testiranje modela predviđanja brzine vozila. Modelira se statička stohastička i deterministička duboka unaprijedna višeslojna neuronska mreža za predviđanje razdiobe brzine vozila duž rute te profila brzine vozila na pomičnom vremenskom horizontu u budućnosti. Provodi se sveobuhvatna analiza točnosti predviđanja razvijenih prediktivnih modela temeljem raznih pokazatelja kvalitete predviđanja. Deterministički model detaljno se analizira u pogledu utjecajnosti svake kandidirane ulazne značajke (trenutna pozicija vozila, trenutna brzina ili povijest brzina vozila, doba dana i dan u tjednu), duljine vremenskog prozora povijesti brzina te duljine horizonta predviđanja na točnost predviđanja profila brzine vozila. Konačno, predlaže se dinamički stohastički model koji objedinjuje svojstva dva prethodno razvijena prediktivna modela, u smislu predviđanja očekivanja i standardnih devijacija uzoraka brzine na pomičnom horizontu.

Osmo poglavlje: *Zaključak*. Na kraju rada iznose se zaključna razmatranja, smjernice za buduća istraživanja, kao i znanstveni doprinosi doktorskog rada.

2. Prikupljanje, predobrada i analiza voznih ciklusa

U ovom poglavlju opisuje se postupak snimanja voznih ciklusa i njihove predobrade za potrebe daljnje uporabe, te se analiziraju njihove glavne karakteristike. U prvom potpoglavlju opisuje se postupak prikupljanja podataka dobivenih praćenjem reprezentativnog uzorka flote gradskih autobusa. U drugom potpoglavlju prikazuje se postupak predobrade snimljenih podataka u smislu izdvajanja i filtriranja voznih ciklusa, rekonstrukcije pripadnog profila nagiba ceste primjenom regresijske metode temeljene na Gaussovim procesima, te kategorizacije voznih ciklusa u odnosu na zagušenje prometa. U trećem potpoglavlju analiziraju se glavne karakteristike izdvojenih voznih ciklusa te provodi detaljna analiza međuovisnosti brzine, akceleracije i nagiba ceste. U četvrtom potpoglavlju opisuje se postupak akvizicije broja putnika u autobusu za potrebe uključivanja promjenjive mase putnika u postupak sinteze voznih ciklusa predložen u poglavlju 3.

2.1. Prikupljanje voznih ciklusa

Snimljeni podaci prikupljeni su praćenjem flote od deset gradskih autobusa tipa MAN Lion City koji prometuju u gradu Dubrovniku, te koji pokrivaju većinu glavnih gradskih autobusnih ruta. Snimanje je izvedeno korištenjem komercijalnog GPS/GPRS uređaja za praćenje vozila (STM-Eagle, proizvođača Artronic d.o.o, Hrvatska) instaliranog na odabranim autobusima. Snimljeni podaci pored GPS podataka uključuje i dostupne te relevantne podatke pogona vozila dobivene putem CAN sabirnice. Snimanje je započelo dana 1. 10. 2018., te se provodilo kontinuirano, 24 sata dnevno, tijekom razdoblja od šest mjeseci. Vrijeme uzorkovanja postavljeno je na jednu sekundu kako bi se osigurala dovoljna gustoća podataka s obzirom na dinamiku promjene brzine vozila. Svi, u „oblaku“ spremljeni podaci zaprimljeni su putem automatske, svakodnevne e-pošte.

Dodatno je korišten precizni GPS prijamnik (Novatel ProPak G2) s jednom aktivnom antenom (Novatel, Model 521, za L1 nosač valova 1575.42 MHz) [87] i to na ograničenom skupu voznih ciklusa duž rute Babin kuk-Pile, kako bi se prikupili podaci o horizontalnoj i vertikalnoj brzini vozila, koji su potrebni za preciznu rekonstrukciju profila nagiba ceste.

Relevantni podaci snimljeni s pomoću GPS/GPRS uređaja za praćenje vozila uključuju:

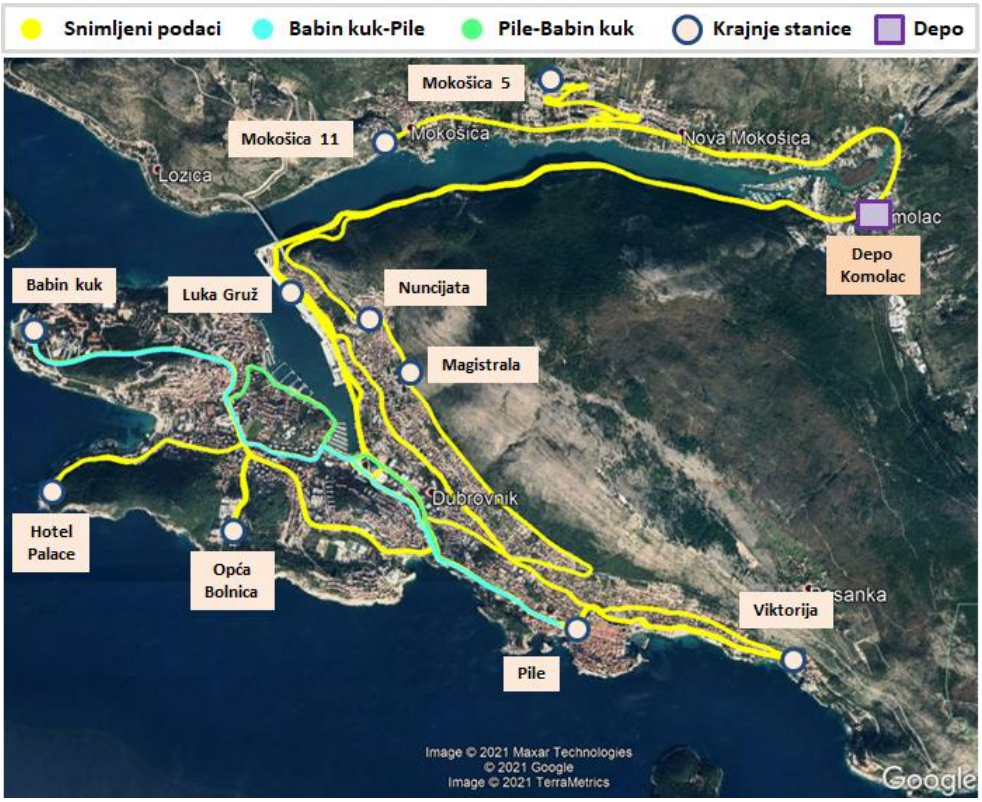
- garažni broj autobusa,

- datum i vrijeme slanja podataka u formatu “godina-mjesec-dan sat:minuta:sekunda“ uz razlučivost od 1 s,
- geografske koordinate u stupnjevima (zemljopisna širina i dužina; razlučivost 1e-7°),
- nadmorska visina (razlučivost 0.1 m),
- uzdužna brzina vozila (razlučivost 0.1 km/h),
- prijeđena udaljenost (od brojača kilometara-odometra; razlučivost 1 km),
- kumulativna potrošnja goriva (razlučivost 0.001 L ili 0.5 L, ovisno o autobusu).

Dodatni podaci snimljeni s pomoću Novatel uređaja za praćenje vozila uključuju:

- datum i vrijeme slanja podataka u formatu “godina-mjesec-dan sat:minuta:sekunda“ uz razlučivost od 1 s,
- geografske koordinate u stupnjevima (zemljopisna širina i dužina),
- brzina vozila (horizontalna i vertikalna komponenta).

Slika 2-1 prikazuje zemljopisne koordinate ruta na kojima se provodilo snimanje, zajedno s krajnjim stanicama pojedinih autobusnih ruta, kao i depoa gdje su autobusi tipično parkirani tijekom noći. Kružna ruta Babin kuk-Pile-Babin kuk odabrana je kao osnova za demonstraciju metoda predstavljenih u sklopu ovog rada, zbog relativno velike duljine (≈5.1 km), protezanja kroz različite dijelove grada (od otvorenih cesta do napučenog gradskog središta) i značajno varirajućeg nagiba ceste.



Slika 2-1: Prikaz ruta gradskih autobusa u Dubrovniku na kojima su se provodila snimanja, zajedno s lokacijama krajnjih stanica i depoa.

2.2. Predobrada snimljenih voznih ciklusa

Početni korak je pohrana snimljenih podataka u bazu podataka, čime se sirovi podaci praćenja, dobiveni u obliku skupa CSV (*engl. Comma Separated Values*) datoteka, objedinjuju u jednu cjelinu te se omogućava njihovo efektivno grupiranje, filtriranje i pretraživanje. U tu svrhu korištena je SQLite relacijska baza podataka zajedno s Python objektno-orijentiranim programskim jezikom (verzija 3.8) [88]. Zatim slijedi izdvajanje voznih ciklusa za odabranu rutu i smjer vožnje temeljem algoritma opisanog u odjeljku 2.2.1. Nadalje, provodi se filtriranje izdvojenih voznih ciklusa kako bi se uklonile određene nepravilnosti. Povrh toga, temeljem povijesnih podataka svih vožnji duž dotične rute, odnosno snimljenih podataka o brzini vozila i prijeđenom putu, rekonstruira se profil nagiba ceste te se pridružuje svakom voznom ciklusu, kako je navedeno u odjeljku 2.2.3. Naposljetku se provodi kategorizacija voznih ciklusa u odnosu na zagušenje prometa (odjeljak 2.2.4).

2.2.1. Izdvajanje voznih ciklusa

Snimljeni podaci vožnje najprije su segmentirani u vozne cikluse, gdje je svaki vozni ciklus definiran vremenskim nizom podataka praćenja (brzina vozila, prijeđeni put, geografske

koordinate, nadmorska visina, kumulativna potrošnja goriva) između dvije krajnje stanice Babin kuk i Pile. Prilikom određivanja voznog ciklusa mora biti zadovoljen sljedeći kriterij: odgovarajuće vozilo mora biti parkirano određeno vrijeme na odgovarajućoj početnoj ili krajnjoj stanici. Jednom kada je navedeni kriterij ispunjen, svim pripadnim voznim podacima snimljenima nakon polaska s početne stanice i dolaska na krajnju stanicu dodjeljuje se jedinstveni identifikator voznog ciklusa. Detaljan opis algoritma izdvajanja voznih ciklusa iz snimljenih podataka vožnje za odabranu rutu opisan je u nastavku.

Ulazni podaci u algoritam izdvajanja snimljenih voznih podataka prema smjeru vožnje za razmatranu rutu uključuju geografske koordinate (geografska dužina i širina) početne i krajnje stanice rute ($X_{GPS,BK}$, $Y_{GPS,BK}$ za stanicu Babin kuk i $X_{GPS,P}, Y_{GPS,P}$ za Pile), te niz apsolutnih vremena $\boldsymbol{t}_v$ i snimljenih geografskih koordinata svakog vozila $\boldsymbol{X}_{GPS,v}, \boldsymbol{Y}_{GPS,v}$; $v = 1,2, \dots, V$, gdje je $V = 10$ određen brojem razmatranih autobusa opremljenih GPS uređajem za praćenje. Za potrebe detekcije ulaska i izlaska vozila u/iz zone početne ili krajnje stanice izvršena je pretvorba geografskih koordinata (X_{GPS}, Y_{GPS}) iz stupnjeva u udaljenost od referentne/nulte točke u metrima X, Y, primjenom jednadžbe koja se obično koristi za projiciranje geografskih koordinata na referentni 2D koordinatni sustav [21]

$$X(k) \cong \frac{R_z \pi}{180} \cdot X_{GPS}(k), \tag{2-1}$$

$$Y(k) \cong \frac{R_z \pi}{180} \cdot Y_{GPS}(k), \tag{2-2}$$

gdje $R_z = 6371$ km označava srednji radijus Zemlje.

Algoritam radi na način da prolazeći kroz vremenski sortirane snimljene podatke vozila v za svaki diskretni vremenski korak k, temeljem geografskih koordinata vozila $X_v(k)$ i $Y_v(k)$, prvo provjerava nalazi li se vozilo u zoni početne/krajnje stanice

$$(X_v(k) - X_s)^2 + (X_v(k) - Y_s)^2 \leq R_s^2 ; \quad s \in \{BK, P\} \tag{2-3}$$

gdje $R_s = 20$ m predstavlja razmatrani radijus stanice.

Zatim se određuje sljedeće:

- Ako se vozilo u *k*-tom vremenskom koraku nalazi unutar R_s, a u prethodnom $k - 1$ vremenskom koraku je bilo izvan R_s, te se zadržava unutar R_s barem sljedećih 30 s, registrira se dolazak vozila na stanicu.

- Na sličan način, ako se vozilo u trenutnom *k*-tom vremenskom koraku nalazi izvan R_s, a u prethodnom $k-1$ vremenskom koraku bilo je unutar R_s, te ako se nalazi izvan R_s barem sljedećih 30 s, registrira se odlazak vozila sa stanice.

Time se dobivaju sekvence posjećenih krajnjih stanica za svako razmatrano vozilo, čijim se daljnjim rastavljanjem na parove (polazište/odredište) određuje vozni ciklus i pripadni smjer vožnje. Primjenom opisane metode izdvojeno je ukupno $N_{DC} = 5031$ voznih ciklusa, od kojih je $N_{DC,BK-P} = 2524$ (50.2%) snimljene na relaciji Babin kuk–Pile, a $N_{DC,BK-P} = 2507$ (49.8%) na relaciji Pile–Babin kuk. Povrh toga, vozni ciklusi su dodatno grupirani ovisno o razlučivosti snimljene potrošnje goriva (fina razlučivost od 0.001 L ili gruba razlučivost od 0.5 L), kako je dano u tablici 2-1. Ukupan broj voznih ciklusa s finom razlučivošću snimljene potrošnje goriva iznosi 2374 (47.2%), a s grubom razlučivošću 2657 (52.8%).

Tablica 2-1: Podjela izdvojenih voznih ciklusa prema smjeru vožnje i razlučivošću snimljene potrošnje goriva.

	Smjer vožnje		∑
	Babin kuk-Pile	Pile-Babin kuk	
Broj voznih ciklusa za dotični smjer vožnje	**2524**/5031 (50.2%)	**2507**/5031 (49.8%)	**5031**
Broj voznih ciklusa s finom razlučivošću snimljene potrošnje goriva	**1193**/2524 (47.3%)	**1181**/2507 (47.1%)	**2374**/5031 (47.2%)
Broj voznih ciklusa s grubom razlučivošću snimljene potrošnje goriva	**1331**/2524 (52.7%)	**1326**/2507 (52.9%)	**2657**/5031 (52.8%)

Nadalje, kako bi se rekonstruirao precizan profil prijeđenog puta između svakog uzorka zabilježenog odometrom (s razlučivosti od 1 km) za svaki *i*-ti vozni ciklus ($i = 1,2,\dots,N_{DC}$), najprije su izračunati pomaci $\Delta X_i, \Delta Y_i$ u referentnom 2D koordinatnom sustavu između svih susjednih vremenskih trenutaka

$$\Delta X_i(k+1) = X_i(k+1) - X_i(k), \tag{2-4}$$

$$\Delta Y_i(k+1) = Y_i(k+1) - Y_i(k). \tag{2-5}$$

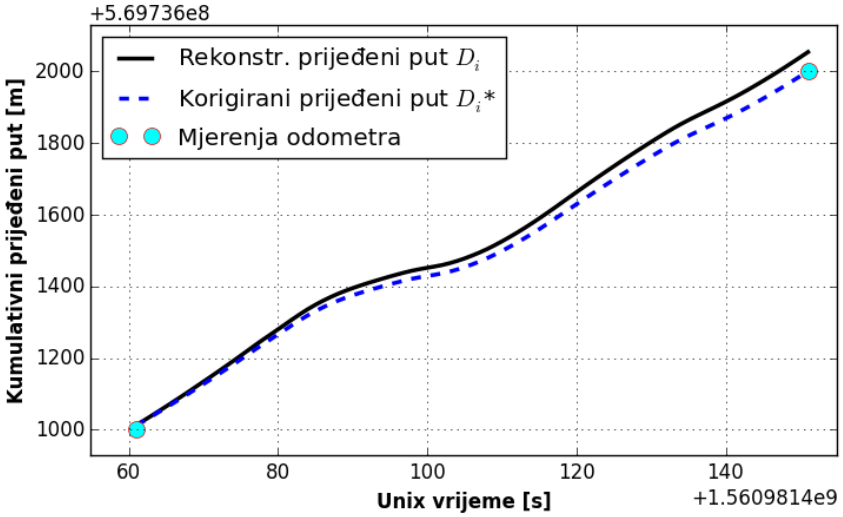
Potom je za svaki *i*-ti vozni ciklus korištenjem jednadžbi (2-4) i (2-5) izračunat kumulativni prijeđeni put D_i prema

$$D_i(k+1) = D_i(k) + \sqrt{\Delta X_i(k+1)^2 + \Delta Y_i(k+1)^2}. \tag{2-6}$$

Konačno, rekonstruirane vrijednosti prijeđene udaljenosti D_i korigiraju se temeljem snimljenih podataka odometra, koji se smatraju točnijim. Na primjer, neka $t_{ODO}(j)$ i $t_{ODO}(\mathrm{j}+1)$ predstavljaju uzastopne korake uzorkovanja u kojima je odometar izmjerio skok u prijeđenoj udaljenosti s $d_{ODO}(j)$ na $d_{ODO}(j+1)$. Budući da za danu razlučivost zabilježenog signala odometra vrijedi $d_{ODO}(j+1) - d_{ODO}(j) = 1$ km, onda se za svaki $k \in [t_{ODO}(j), t_{ODO}(j+1)]$ izvodi korekcija vrijednosti $D_i(k)$ prema principu min-max normalizacije

$$D_i^*(k) = d_{ODO}(j) + \frac{D_i(k) - D_i(t_{ODO}(j))}{D_i(t_{ODO}(j+1)) - D_i(t_{ODO}(j))} \cdot (d_{ODO}(j+1) - d_{ODO}(j)). \tag{2-7}$$

Time se uklanjaju akumulirane pogreške prisutne u rekonstruiranim vrijednostima prijeđene udaljenosti D_i, koje su rezultat projiciranja geografskih koordinata sfere na ravninu, zanemarivanja nadmorske visine, te linearne interpolacije udaljenosti između snimljenih točaka. Ilustracija rekonstruiranog D_i i korigiranog D_i^* profila prijeđenog puta između dvije mjerene vrijednosti s odometra prikazana je na slici 2-2.



Slika 2-2: Primjer GPS-rekonstruiranog D_i i temeljem odometra-korigiranog D_i^* profila prijeđenog puta između dva susjedna mjerenja odometra.

2.2.2. Filtriranje voznih ciklusa

Vozni ciklusi filtriraju se primarno kako bi se izdvojili oni koji ne uključuju gubitke GPS signala ili vrijednosti određenih varijabli koje povremeno značajno odstupaju od glavnine zapažanja (lažni signali), a koje su tipično posljedica pogreške mjerenja. Popis svih kriterija koje vozni ciklus mora zadovoljiti kako bi bio proglašen valjanim uključuje:

I. vremenski razmak između svakog snimljenog uzorka duž voznog ciklusa mora biti jednak vremenu uzorkovanja koje iznosi 1 s,
II. broj pojavaka numerički neodređenih vrijednosti (*engl. Not a Number*, NaN) mora biti jednak 0,
III. početna brzina voznog ciklusa mora biti jednaka 0 km/h,
IV. konačna brzina voznog ciklusa mora biti jednaka 0 km/h,
V. akceleracija vozila mora biti unutar intervala [-3, 3] m/s^2,
VI. udio vremena stajanja autobusa (brzina vozila jednaka 0 km/h) mora biti manja od 75% ukupnog vremena trajanja voznog ciklusa,
VII. vozni ciklus mora prolaziti referentnom trajektorijom rute (ne smije uključivati skretanja s referentne rute i vožnju sporednim cestama, npr. zbog radova na cesti, odlazak u depo ili prelazak na neku drugu rutu).

Svaki vozni ciklus koji ne zadovoljava jedan od gore navedenih uvjeta se odbacuje. Ovakvo relativno restriktivno filtriranje voznih ciklusa moguće je zbog sveukupno velikog broja snimljenih voznih ciklusa.

Kako je dano u tablici 2-2, postupak filtriranja rezultirao je ukupnim brojem valjanih voznih ciklusa jednakim $N_{DC}^{V} = 3216$, od kojih se $N_{DC,BK-P}^{V} = 2313$ (71.9%) odnosi na smjer Babin kuk – Pile, a $N_{DC,P-BK}^{V} = 903$ (28.1%) na smjer Pile – Babin kuk. Nadalje, ukupan broj voznih ciklusa s finom razlučivošću snimljene potrošnje goriva iznosi $N_{DC,F}^{V} = 1527$ (47.5%), a s grubom razlučivošću $N_{DC,G}^{V} = 1689$ (52.5%).

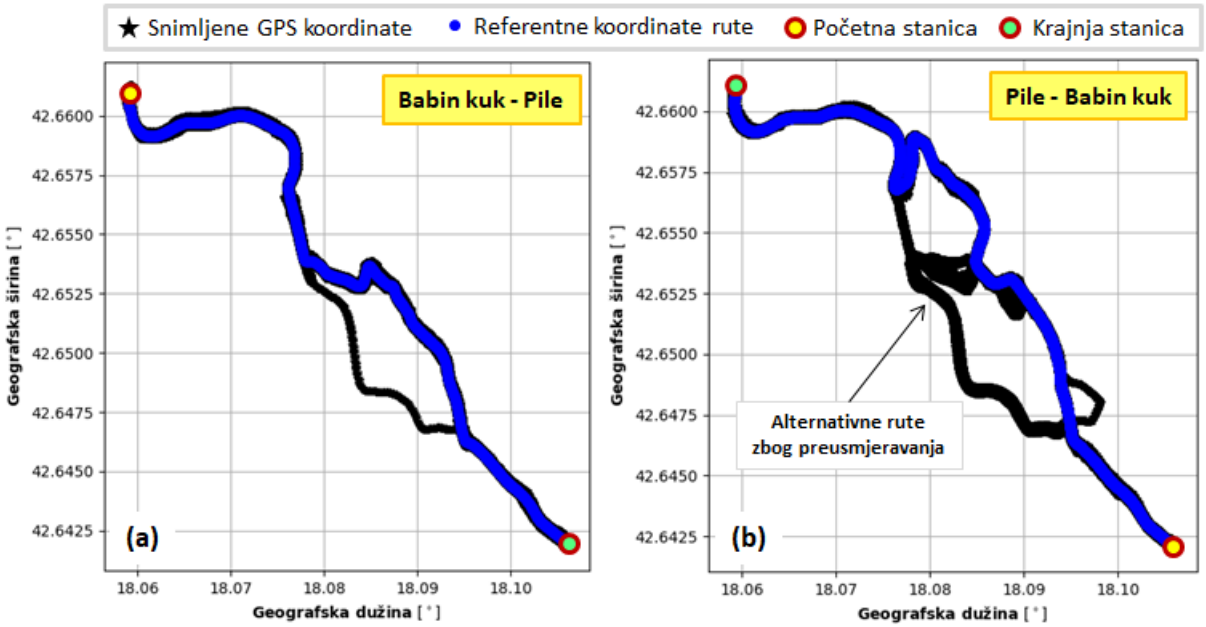
Detaljna statistika rezultata filtriranja voznih ciklusa prema svakom od razmatranih kriterija dana je u tablici 2-3. Rezultati u tablici 2-3 pokazuju kako najveći utjecaj na količinu odbačenih voznih ciklusa generalno imaju kriteriji I, II i V, koji su redom vezani za vremenski razmak između snimljenih točaka, pojavak numerički neodređenih vrijednosti te ograničenje akceleracije vozila. Nadalje, razlog odbacivanja značajne količine voznih ciklusa za smjer Pile – Babin kuk (ukupno 1604 odbačenih voznih ciklusa; 64% od ukupnog broja) rezultat je primjene kriterija VII, koji zahtijeva da vozilo prolazi referentnom rutom. Na slici 2-3 vidljivo je kako to nije bio slučaj za relaciju Pile – Babin kuk, jer se mnoštvo voznih ciklusa odnosilo na alternativne segmente ceste, što je vjerojatno posljedica izvođenja radova ili neke druge izvanredne okolnosti u danom periodu snimanja. Primjeri valjanih voznih ciklusa za različite datume i vremena polazaka prikazani su na slici 2-4, gdje su vidljive različite varijacije brzine vozila koje su rezultat različitih stanja prometa.

Tablica 2-2: Podjela filtriranih voznih ciklusa prema smjeru vožnje i razlučivosti snimljene potrošnje goriva.

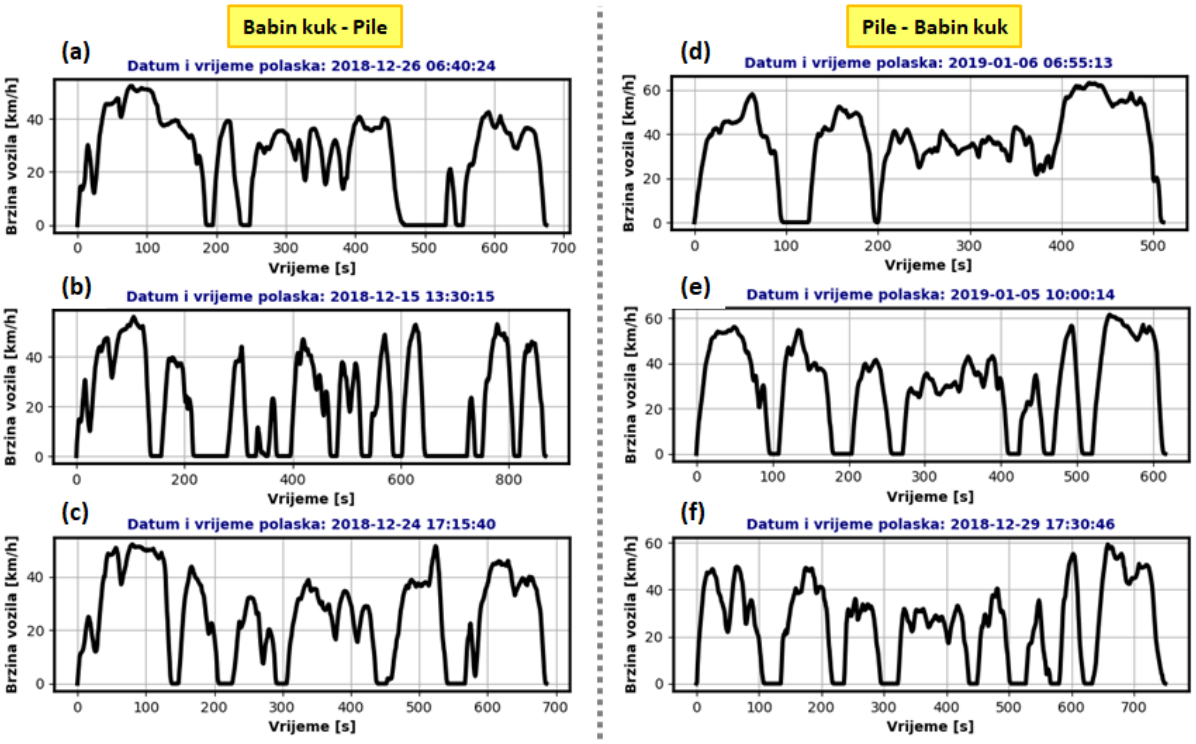
	Smjer vožnje		Σ
	Babin kuk-Pile	Pile- Babin kuk	
Omjer valjanih voznih ciklusa	**2313**/2524 (91.6%)	**903**/2507 (36.0%)	**3216**/5031 (63.9%)
Omjer valjanih voznih ciklusa s finom razlučivošću snimljene potrošnje goriva	**1102**/1193 (92.4%)	**425**/1181 (36.0%)	**1527**/3216 (47.5%)
Omjer valjanih voznih ciklusa s grubom razlučivošću snimljene potrošnje goriva	**1211**/1331 (91.0%)	**478**/1326 (36.0%)	**1689**/3216 (52.5%)

Tablica 2-3: Udjeli odbačenih voznih ciklusa prema individualnim kriterijima filtriranja.

Smjer vožnje	Kriterij filtriranja						
	I	II	III	IV	V	VI	VII
Babin kuk – Pile → / 2524	104 (≈4.1%)	41 (≈1.6%)	0 (0%)	0 (0%)	46 (≈1.8%)	1 (≈0%)	1 (≈0%)
Pile – Babin kuk → / 2507	74 (≈3.0%)	42 (≈1.7%)	0 (0%)	1 (0%)	160 (≈6.4%)	0 (0%)	1541 (≈61.5%)



Slika 2-3: Rasipanje snimljenih geografskih koordinata na odabranoj ruti za smjer vožnje a) Babin kuk-Pile i b) Pile-Babin kuk.



Slika 2-4: Primjeri raznih valjanih snimljenih voznih ciklusa za smjer a) Babin kuk-Pile i b) Pile-Babin kuk.

2.2.3. Rekonstruiranje profila nagiba ceste

Poznavanje profila nagiba ceste od izrazite je važnosti jer ima značajan utjecaj na potrošnju goriva te se uključuje u sintezu voznih ciklusa. Nagib ceste rekonstruira se na temelju snimljenih geografskih koordinata, horizontalne i vertikalne brzine vozila, te podataka o prijeđenom putu. Cjelokupni postupak rekonstrukcije prikazan je blokovskim dijagramom na slici 2-5. Metoda rekonstrukcije nagiba ceste demonstrirana je za smjer Babin kuk – Pile, dok se konačni rezultati daju za oba smjera vožnje.

Geografske koordinate svih snimljenih voznih ciklusa za smjer Babin kuk – Pile prikazane su na slici 2-6a. Kako bi se temeljem snimljenog skupa koordinata odredio manji broj referentnih geografskih koordinata koje reprezentiraju trajektoriju rute, provedeno je klasteriranje podataka primjenom algoritma *K*-sredina (*engl. K-means*) [89], koji pripada polju nenadziranog strojnog učenja. Time se provodi klasifikacija neobilježenih podataka grupiranjem po određenim značajkama. Cilj algoritma *K*-sredina je podijeliti ulazne podatke u *K* različitih klastera te odrediti položaj težišta mase/centroida za svaki klaster. Prototip funkcije algoritma *K*-sredina u Python programskom okruženju (*scikit-learn* modul) glasi

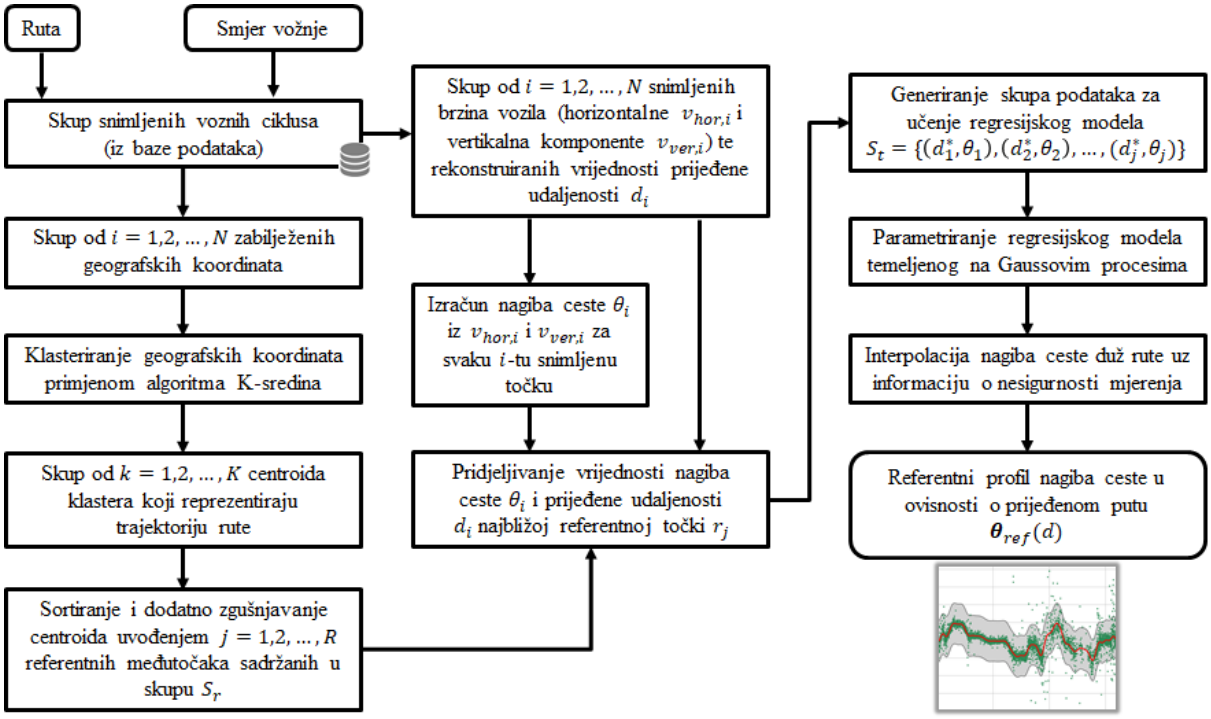
$$S_C, S_I, E_\sum = \text{k_means}(\mathbf{X}, K),$$

gdje **X** označava matricu dimenzija *N* x *M*, pri čemu broj redaka *N* označava broj uzoraka (ovdje broj snimljenih točaka), a *M* broj značajki po kojima se provodi klasteriranje (ovdje geografske koordinate). Ulazni parametar *K* označava zadani broj klastera. Funkcija kao izlaz daje (i) skup težišta centroida u razmatranom prostoru značajki $S_c \in R^{K \times M}$ (ovdje geografske koordinate), (ii) skup indeksa centroida pridruženih svakom pojedinom uzorku $S_I \in R^{N \times 1}$, te (iii) zbroj kvadratnih udaljenosti svakog uzorka do najbližeg centroida E_Σ.

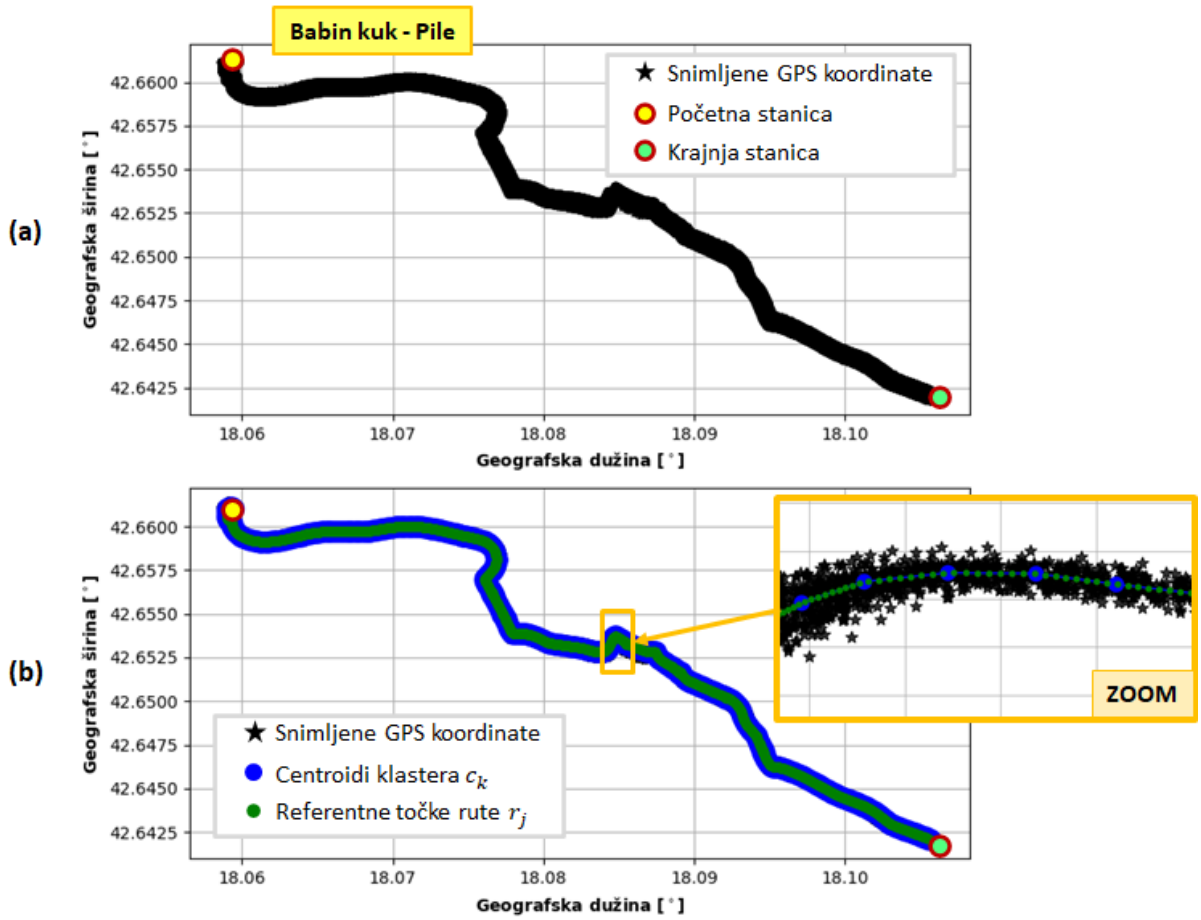
Algoritam *K*-sredina temelji se na iterativnom postupku koji se sastoji od sljedećih koraka:

1) u *M*-dimenzionalni Euklidski koordinatni sustav slučajno se postavlja *K* točaka koje predstavljaju centroide klastera,
2) izračunavaju se udaljenosti *M*-dimenzionalnih podataka do svakog centroida, te se svaki podatak pridodjeljuje najbližem klasteru prema Euklidskoj udaljenosti,
3) Računaju se nova težišta centroida svakog klastera usrednjavanjem svih podataka u klasteru,
4) Ponavljaju se koraci 2-4 sve dok pozicije novoizračunatih centroida ne postanu približno jednake starim pozicijama.

Cilj algoritma je minimiziranje sume udaljenosti svakog od podataka do pripadnog centroida klastera kojem je taj podatak dodijeljen (tzv. maksimiziranje očekivanja).



Slika 2-5: Blokovski dijagram metode rekonstrukcije profila nagiba ceste duž rute temeljem snimljenih GPS podataka vožnje.



Slika 2-6: a) Snimljene geografske koordinate za rutu Babin kuk-Pile zajedno sa pripadnim lokacijama polazišta i odredišta, te b) rezultati određivanja pripadnih referentnih geografskih koordinata (za broj klastera $K = 500$ i broj referentnih točaka $R = 5000$).

Prilikom klasteriranja geografskih koordinata, broj klastera K postavljen je metodom pokušaja i pogreške na 500, budući da bi traženje optimalnog broja klastera za količinu od $N = 1{,}908.209$ snimljenih točaka (703.654 u slučaju relacije Pile-Babin kuk) bio veoma dugotrajan proces. Time je dobiven skup centroida $S_c = \{c_1, c_2, \dots, c_k\}; k = 1, 2, 3, \dots, K$, gdje svaki centroid c_k predstavlja određeni segment ceste (vidi sliku 2-6b; plave točke). Budući da je skup od K centroida nedovoljan za preciznu rekonstrukciju profila nagiba ceste zbog velikih razlika u prijeđenim udaljenostima između susjednih centroida (≈10 m u prosjeku), provodi se dodatno zgušnjavanje centroida uvođenjem velikog broja referentnih međutočaka predstavljenih skupom $S_r = \{r_1, r_2, \dots, r_j\}; j = 1, 2, \dots, R$. Budući da duljina rute iznosi ≈5.1 km, broj referentnih točaka R postavljen je na 5000 kako bi prosječna udaljenost između svake susjedne referentne točke bila ≈1 m. Svaka referentna točka ekvidistantno je razmaknuta jedna od druge, a dobivena je linearnom interpolacijom geografskih koordinata centroida u ovisnosti o njihovoj međusobnoj Euklidskoj udaljenosti (vidi sliku 2-6b; zelene točke). Na ovaj se način

veliki broj snimljenih koordinata N zamjenjuje znatno manjim brojem referentnih koordinata ruta R koje se ne preklapaju, a koje precizno opisuju trajektoriju razmatrane rute.

Sljedeći korak je pridruživanje podataka o poziciji i nagibu ceste svakoj referentnoj točki rute r_j. Podatak o poziciji referentnih točaka moguće je odrediti preko prijeđenih udaljenosti vozila u odnosu na početnu stanicu na Babinom kuku rekonstruiranih iz snimljenih podataka odometra (vidi odjeljak 2.2.2), dok se nagib ceste izračunava korištenjem snimljenih podataka o vertikalnoj (v_{ver}) i horizontalnoj (v_{hor}) brzini vozila na sljedeći način:

$$\theta_i = \text{atan}\left(\frac{v_{ver,i}}{v_{hor,i}}\right). \tag{2-8}$$

Pridruživanje izračunatih vrijednosti prijeđene udaljenosti d_i i nagiba ceste θ_i svake *i*-te snimljene točke (gdje je $i = 1,2,\dots,N$) najbližoj referentnoj točki r_j provodi se po principu najmanje Euklidske udaljenosti. Time se za svaku *j*-tu referentnu točku dobiva skup pridijeljenih prijeđenih udaljenosti $D_j = \{d_1, d_2, \dots, d_n\}$ i nagiba ceste $\Theta_j = \{\theta_1, \theta_2, \dots, \theta_n\}$, gdje n predstavlja ukupan broj pridodjeljenih točaka. Konačno, aproksimativna pozicija *j*-te referentne točke $\tilde{d}_j$ duž rute računa se kao medijan prijeđenih udaljenosti svih snimljenih točaka pridruženih toj referentnoj točki (sadržanih u skupu D_j)

$$\tilde{d}_j = median\big(D_j = \{d_1, d_2, \dots, d_n\}\big), \tag{2-9}$$

te predstavlja kriterij prema kojemu se provodi sortiranje referentnih točaka kako bi se dobio njihov pravilan redoslijed duž rute.

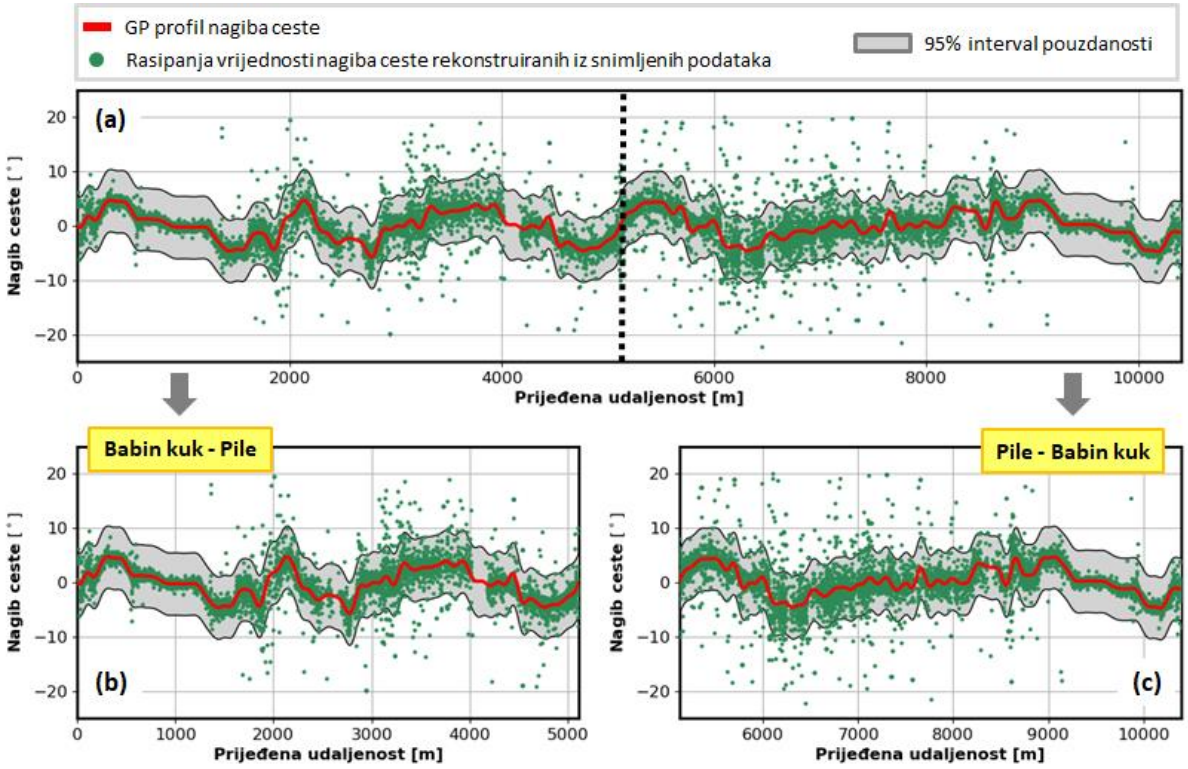
Iz jednadžbe (2-8) može se vidjeti da izračunati nagib ceste θ_i može postati vrlo osjetljiv u uvjetima male horizontalne brzine v_{hor}. U tom slučaju, male varijacije u komponentama vertikalne v_{ver} i horizontalne brzine v_{hor} uzrokovane šumom mjerenja rezultiraju značajnom promjenom izračunatog nagiba ceste te smanjuju točnost rekonstruiranog nagiba ceste. Kako bi se smanjila osjetljivost i dobili pouzdani profili nagiba ceste, korišteni su podaci višestrukih prolazaka dotičnom rutom, kao što prikazuje slika 2-7. Međutim, i dalje se mogu primijetiti određena rasipanja vrijednosti nagiba ceste (zelene točke). Stoga je za dobivanje pouzdanog profila nagiba ceste u ovisnosti o prijeđenoj udaljenosti za razmatranu rutu korištena regresijska metoda temeljena na Gaussovim procesima (*engl. Gaussian Process*, GP) [90], koja učinkovito uključuje više podataka i suzbija šum. U postupku provođenja regresije, podaci nagiba ceste θ_i svake *j*-te referentne točke rute smatrani su izlaznim vektorom $\boldsymbol{y}$, dok su pripadne vrijednosti prijeđene udaljenosti $\tilde{d}_j$ smatrane ulaznim vektorom $\boldsymbol{x}$. Metoda je implementirana korištenjem klase *GaussianProcessRegressor* iz *scikit-learn* modula za Python programski jezik [91].

GP metoda zahtijeva odabir odgovarajuće kovarijacijske funkcije cov i njezinih hiperparametara. U tu svrhu, koristi se kvadratna eksponencijalna funkcija $cov(x, y)$ koja je definirana kao

$$cov(x, y) = \sigma_f^2 \cdot e^{\left(-\frac{(x-y)^2}{2l^2}\right)} + \sigma_n^2 \delta_{x,y}, \tag{2-10}$$

gdje je l karakteristično mjerilo duljine procesa koje određuje glatkost konačnog GP profila, σ_f^2 je varijanca signala koja određuje varijacije vrijednosti GP profila od njihove srednje/očekivane vrijednosti, σ_n^2 je varijanca fluktuacije šuma koja određuje koliko se očekuje šuma prisutnih u podacima, a δ je Kroneckerova delta funkcija. Hiperparametri su optimizirani tijekom učenja modela s ciljem minimiziranja negativne log marginalne vjerojatnosti (*engl. negative log marginal likelihood*; NLML). Rezultirajući optimalni parametri GP algoritma iznose redom $l = 67.2, \sigma_n = 2.91, \sigma_f = 2.26$.

Konačni profil nagiba ceste koji predstavlja očekivanje GP-a prikazan je na slici 2-7 crvenom linijom, zajedno s odgovarajućim intervalima pouzdanosti od 95%, izvedenima iz varijance GP-a. Budući da postoje određena preklapanja rute za različite smjerove vožnje Babin kuk-Pile i Pile-Babin kuk, na slici 2-7a se može zamijetiti određena simetričnost profila nagiba ceste s obzirom na horizontalnu os, primjerice na početku i kraju rute, gdje su predznaci nagiba obrnuti zbog različitih smjerova vožnje. Individualni profili nagiba ceste za oba smjera vožnje prikazani su na slici 2-7b i 2-7c, te su kao takvi linearnom interpolacijom po prijeđenoj udaljenosti pridijeljeni svakom valjanom snimljenom voznom ciklusu prema tablici 2-2. Rutu Babin kuk-Pile u većoj mjeri karakterizira vožnja nizbrdo uz prosječnu vrijednost i standardno odstupanje nagiba ceste od -0.24° i 2.69°, a rutu Pile-Babin kuk vožnja uzbrdo uz prosječnu vrijednost nagiba ceste od +0.21° i nešto niže standardno odstupanje od 2.41°.



Slika 2-7: a) Konačni profil nagiba ceste rekonstruiran za odabranu rutu uporabom regresijske metode temeljene na GP-u, zajedno s individualnim profilima nagiba ceste za smjer b) Babin kuk-Pile i c) Pile-Babin kuk.

2.2.4. Kategorizacija voznih ciklusa

Budući da realna vožnja ovisi o duljini putovanja, odredištu i prometnoj situaciji, vozni ciklusi se tipično kategoriziraju na temelju prijeđene udaljenosti (kratki, srednji i dugi), stanju prometa (vožnja gradskim ulicama, glavnim prometnicama, brzim cestama ili autocestama), te stilu vožnje (mekan, regularan, agresivan) [24]. Razlog tome je što različite vožnje proizvode različite uzorke voznih ciklusa. Kategorizacija s obzirom na prijeđenu udaljenost preporučuje se kada se žele dobiti generalizirani obrasci vožnje koji sadrže atipičnu mješavinu vožnje lokalnim cestama prije spajanja s glavnom cestom/autocestom [20], dok se kategorizacija u odnosu na srednju brzinu preporuča za određivanje stanja prometa [70]. Za kategorizaciju voznih ciklusa prema stilu vožnje se zbog kompleksnosti samog problema koriste razne statističke značajke vezane za brzinu i akceleraciju vozila [69, 71].

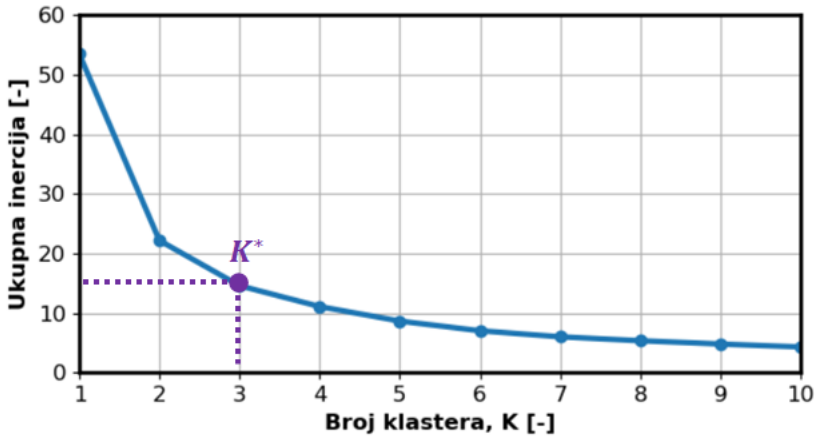
Budući da se u ovom radu podrazumijeva skup voznih ciklusa snimljenih na identičnoj autobusnoj ruti koji ne posjeduju značajne varijacije u prijeđenim udaljenostima (vidi potpoglavlje 2.1), te također ne sadrže izraženi utjecaj stila vožnje zbog karakteristika samog vozila (velika masa koja rezultira tromom dinamikom vožnje), usvaja se kategorizacija voznih ciklusa prema stupnju zagušenja prometa. Pritom se kao značajke relevantne za određivanje

zagušenosti prometa usvajaju: (i) srednja brzina vozila i (ii) broj zaustavljanja vozila po prijeđenom kilometru, koje se izračunavaju za svaki vozni ciklus zasebno (vidi sliku 2-4) [60]. Izračunate vrijednosti značajki skalirane su na raspon od 0 do 1 prema principu min-max normalizacije kako bi svaka značajka imala jednak utjecaj prilikom klasteriranja. Optimalni broj klastera određuje se temeljem Elbow metode [92], koja se sastoji od sljedeća četiri koraka:

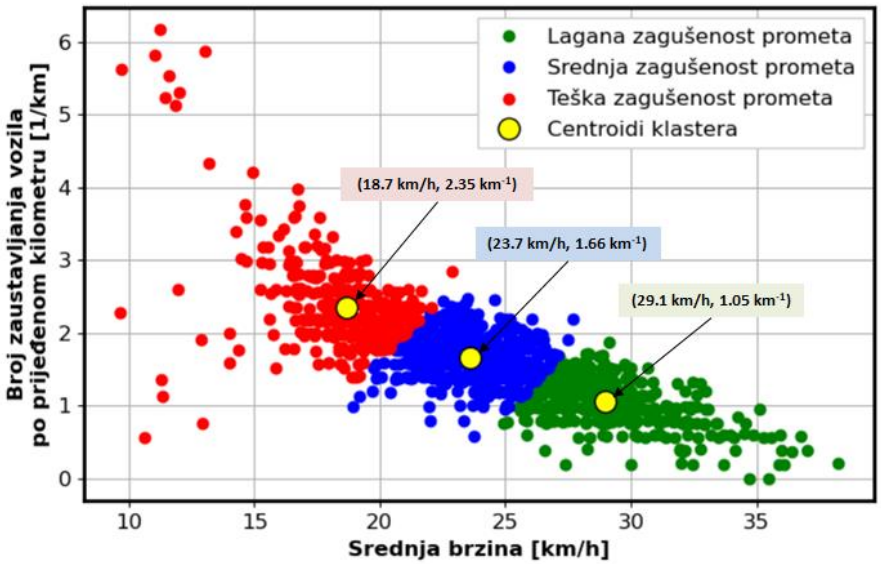
1) Provođenje klasteriranja za različiti broj klastera $K = 1,2,\ldots,N$.
2) Izračunavanje ukupne inercije $E_{\sum}$ definirane kao zbroj kvadratnih udaljenosti uzoraka do najbližeg središta/centroida klastera za svaki K (vidi potpoglavlje 2.2.3 za više detalja).
3) Iscrtavanje krivulje ukupne inercije $E_{\sum}$ kao funkcije broja klastera K.
4) Odabir vrijednosti K^* koji se nalazi na "laktu" krivulje (točka nakon koje se ukupna inercija $E_{\sum}$ počinje linearno smanjivati) kao optimalnog broja klastera.

Rezultirajuća krivulja ukupne inercije $E_{\sum}$ u odnosu na broj klastera $K = 1,2,\ldots,10$ dobivena primjenom algoritma *K*-sredina prikazana je na slici 2-8, te ističe kako je optimalni broj klastera jednak $K^* = 3$. Sukladno tome, zagušenost prometa kategorizira se kao lagana, srednja i teška. Slika 2-9 prikazuje razdiobu odabranih značajki zagušenosti prometa za sve vozne cikluse snimljene na odabranoj ruti, gdje se može uočiti kako laganu zagušenost prometa karakterizira niski broj zaustavljanja po kilometru i visoka srednja brzina, dok za tešku zagušenost vrijedi obratno. Prosječne vrijednosti srednjih brzina i broja zaustavljanja vozila po prijeđenom kilometru za svaki pripadni klaster određene su centroidima (označenima žutim kružićima na slici 2-9), a iznose redom 29.1 km/h i 1.05 km^{-1} za laganu, 23.7 km/h i 1.66 km^{-1} za srednju, te 18.7 km/h i 2.35 km^{-1} za tešku zagušenost prometa. Udio voznih ciklusa u kategoriji lagane zagušenosti prometa iznosi 41.8%, srednje 29.9%, a teške 28.3%.

Tako kategorizirani vozni ciklusi dalje se koriste za generiranje tri nezavisne matrice prijelaznih vjerojatnosti, iz kojih se potom generiraju sintetički vozni ciklusi za svaku od navedenih kategorija (za detaljne informacije vidi poglavlje 3). Time se u postupak modeliranja voznih ciklusa uključuju vremenski promjenjive značajke zagušenja prometa.



Slika 2-8: Krivulja ovisnosti ukupne inercije $E_{\sum}$ *u odnosu na broj klastera, uključujući odabrani optimalni broj klastera* K^* *prema Elbow metodi.*



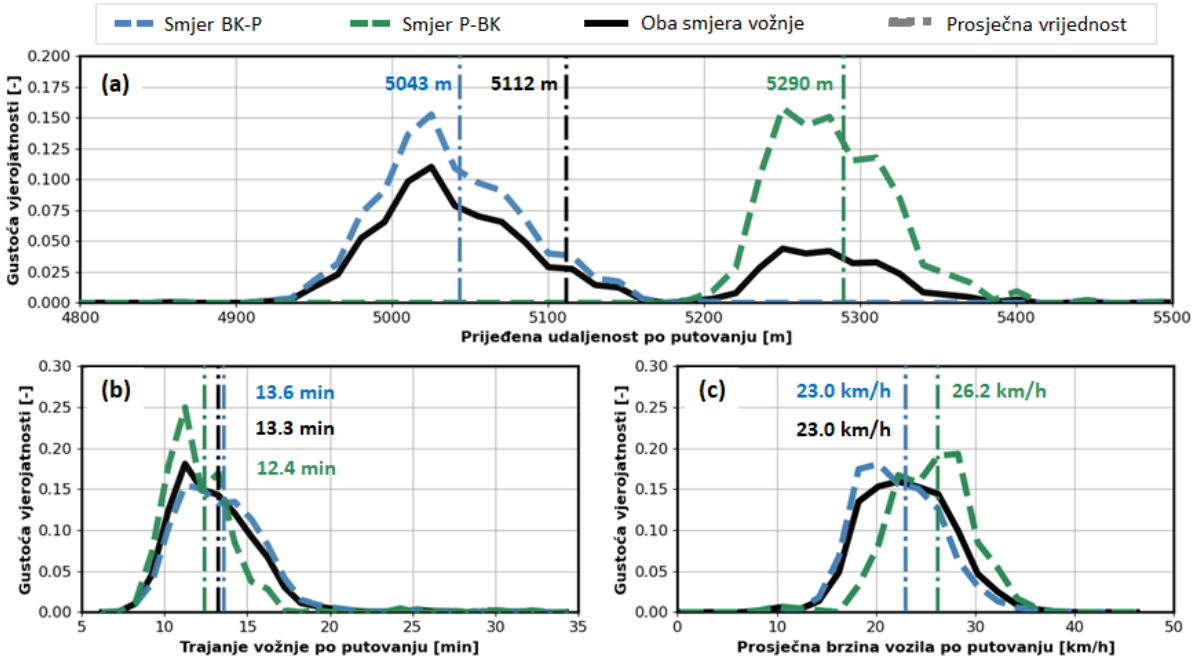
Slika 2-9: Rezultati klasteriranja voznih ciklusa u odnosu na srednju brzinu i broj zaustavljanja vozila po kilometru.

2.3. Analiza snimljenih voznih ciklusa

U ovom potpoglavlju opisuju se glavne karakteristike snimljenih voznih ciklusa na ruti Babin kuk–Pile za oba smjera vožnje, te se utvrđuju rasponi snimljenih brzina, akceleracija i nagiba ceste za potrebe parametriranja stohastičkog modela za generiranje sintetičkih voznih ciklusa opisanog u poglavlju 3 (odjeljak 2.3.1). Pored toga, provodi se analiza međuovisnosti brzine, akceleracije i nagiba ceste, kako je opisano u odjeljku 2.3.2.

2.3.1. Glavne karakteristike

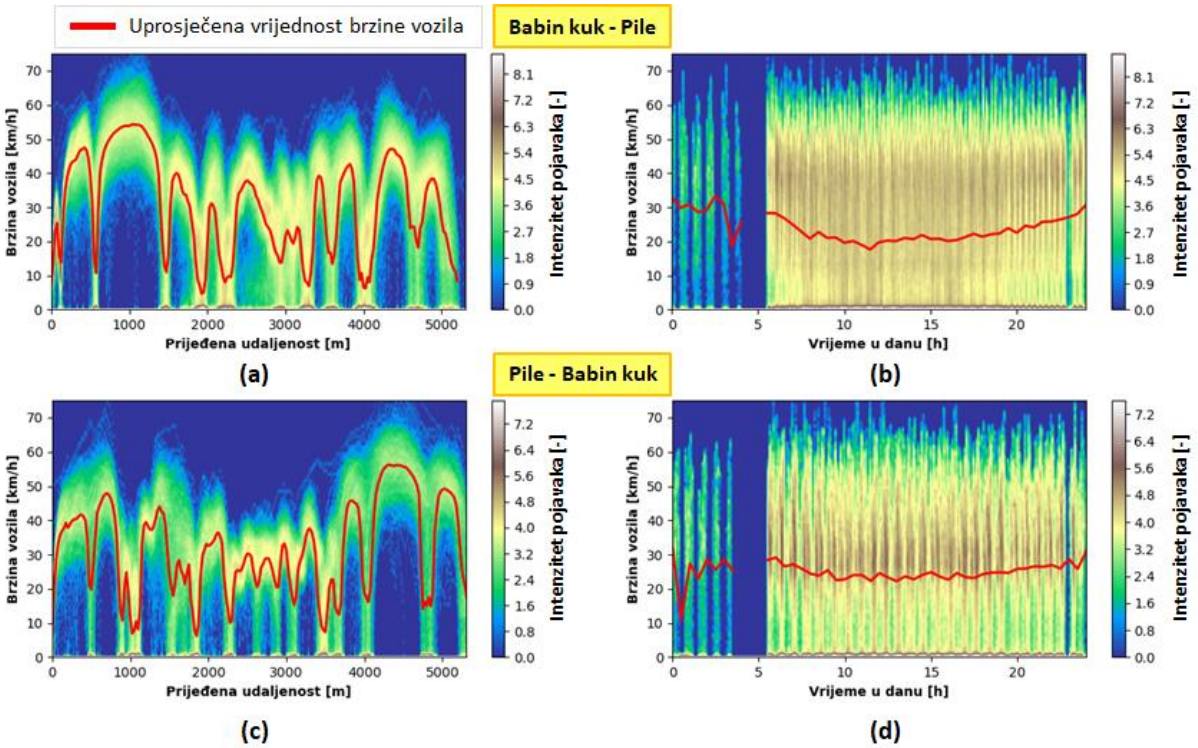
Slika 2-10 prikazuje razdiobe prijeđenih udaljenosti, trajanja vožnje te prosječne brzine vozila po putovanju za oba smjera vožnje, dok su vrijednosti osnovnih statističkih pokazatelja dane u tablici 2-3 (srednja vrijednost, standardna devijacija, te 25-ti, 50-ti (medijan) i 75-ti percentil). Prosječne prijeđene udaljenosti za relacije Babin kuk-Pile (BK-P) i Pile–Babin kuk (P-BK) iznose redom 5043 m i 5290 m, a za objedinjene podatke oba smjera vožnje 5112 m (slika 2-10a). Vidljive su određene razlike u prijeđenim udaljenostima vozila između individualnih putovanja (približno ±100 m, tj. ±2%), koje su prvenstveno rezultat rekonstrukcije vremenskog profila prijeđene udaljenosti iz snimljenih GPS koordinata zbog grube razlučivosti odometra (vidi odjeljak 2.2.1), uključujući klizanje pogonskih guma, učestalost promjena traka tijekom vožnje te nezaustavljanje na svim međustanicama. Prosječno trajanje vožnje po putovanju za objedinjene podatke oba smjera vožnje iznosi 13.3 min, te se ponešto razlikuje između smjerova vožnje (13.6 min za smjer BK-P u odnosu na 12.4 min za smjer P-BK; slika 2-10b), što je rezultat nešto više/niže prosječne brzine vozila duž rute (slika 2-10c). Značajna rasipanja u trajanju vožnje (do ≈40% oko prosječne vrijednosti, slika 2-10b) te slično i srednje brzine (slika 2-10c) mogu se pripisati različitim prometnim uvjetima duž dana.



Slika 2-10: Razdiobe a) prijeđenih udaljenosti, b) trajanja vožnje i c) prosječne brzine po putovanju za oba smjera vožnje na ruti Babin kuk – Pile, uključujući prosječne vrijednosti.

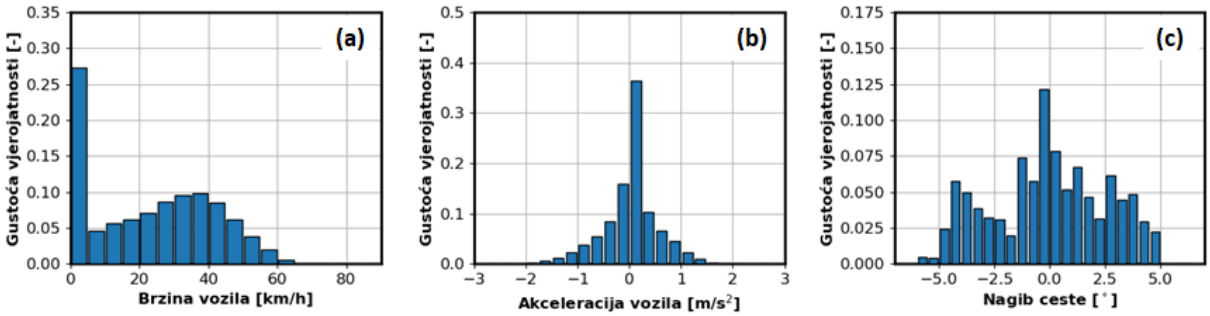
Slika 2-11 prikazuje rasipanja snimljenih brzina vozila duž rute za oba smjera vožnje u domeni vremena (slika 2-11a i 2-11c) i prijeđene udaljenosti/pozicije (slika 2-11b i 2-11d), uključujući uprosječene vrijednosti brzina (crvena linija). Iz slika 2-11b i 2-11d je vidljivo kako je

zagušenost prometa najviša u vremenskom intervalu između 8-20 sati (silazni trend prosječne brzine), te da značajno varira između snimljenih uzoraka budući da ponajviše ovisi o karakteristikama ceste i lokalnom prometu. Razlog nedostatka podataka u ranim jutarnjim satima je u tome što autobusi u tom vremenskom periodu miruju u depou. S druge strane, slike 2-11a i 2-11c pokazuju kako se s porastom količine prikupljenih podataka na fiksnoj ruti kroz vrijeme počinju odražavati karakteristike te rute (npr. lokacije raskrižja/semafora, autobusnih stajališta, pješačkih prijelaza i sl.), što je vidljivo iz pozicija s najučestalijim zaustavljanjima vozila te vožnje na otvorenoj cesti (s visokom brzinom).



Slika 2-11: Toplinske mape rasipanja snimljenih brzina vozila duž rute za oba smjera vožnje u domeni a), c) prijeđene udaljenosti i b), d) vremena, uključujući uprosječene vrijednosti brzina (crvena linija).

Slika 2-12a prikazuje histogram svih snimljenih uzoraka brzina vozila iz kojeg se vidi da se brzine vozila uglavnom kreću u rasponu od 0 do 60 km/h. Maksimalna izmjerena brzina vozila iznosi 81.2 km/h. Raspon akceleracija uglavnom se kreće od -1.3 m/s^2 do 1.2 m/s^2 (slika 2-12b), dok minimalna i maksimalna vrijednosti iznose redom -2.91 i 2.86 m/s^2. U slučaju nagiba ceste (slika 2-12c), raspon je definiran minimalnom i maksimalnom vrijednosti profila rekonstruiranog u odjeljku 2.2.3 (vidi sliku 2-7), koje iznose redom -5.73° i 4.71° za smjer BK-P, te -4.74° i 4.68° za smjer P-BK.

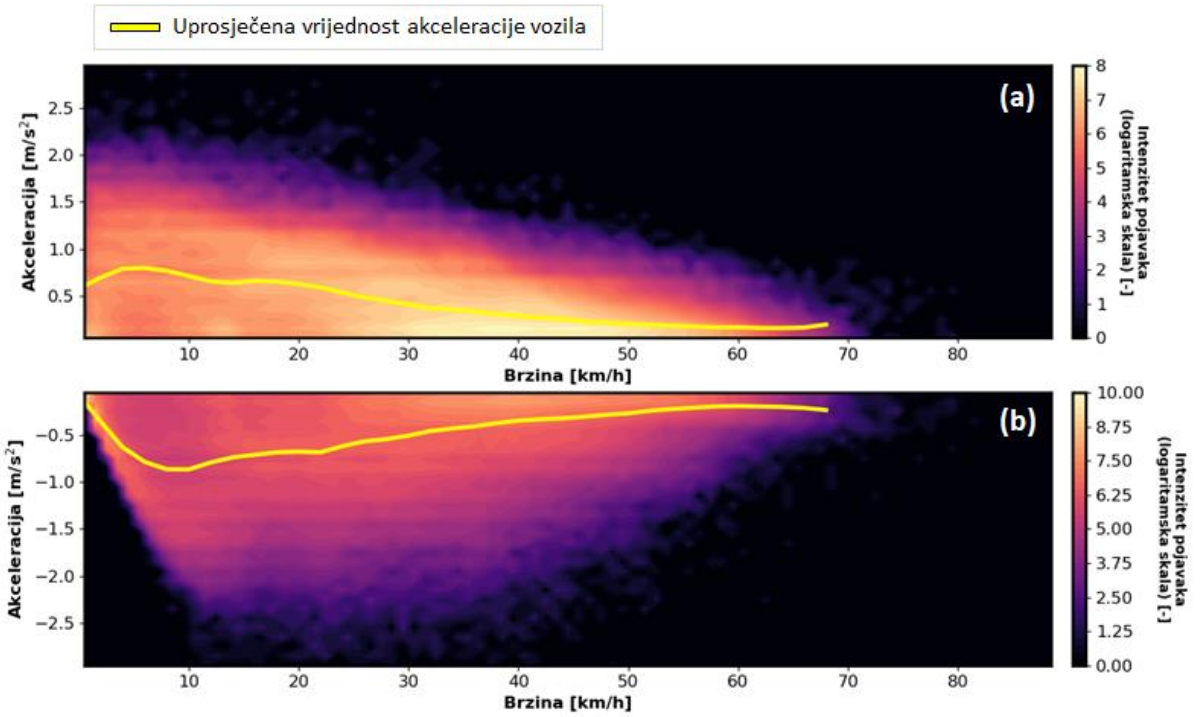


Slika 2-12: Histogrami snimljenih uzoraka a) brzine vozila, b) akceleracije vozila i c) nagiba ceste za objedinjene podatke oba smjera vožnje na ruti Babin kuk-Pile.

2.3.2. Međuovisnosti brzine, akceleracije i nagiba ceste

Cilj ove analize je utvrditi postoji li značajna međuovisnost brzine i akceleracije vozila te nagiba ceste, što se smatra važnim sa stajališta sinteze 3D voznih ciklusa (u združenom ili razdvojenom obliku). Analiza je provedena nezavisno za slučaj voznih ciklusa prikupljenih na relacijama Babin kuk-Pile (BK-P) i Pile-Babin kuk (P-BK), te u konačnici za slučaj objedinjenog skupa snimljenih voznih ciklusa za oba smjera vožnje (Babin kuk-Pile-Babin kuk; BK-P-BK).

Razdioba akceleracije vozila u ovisnosti o brzini vozila za slučaj podataka BK–P–BK prikazana je na slici 2-13, te jasno pokazuje kako se akceleracija vozila generalno sužava s povećanjem brzine vozila, kako za pozitivne (slika 2-13a) tako i za negativne (slika 2-13b) vrijednosti akceleracije. Ova tvrdnja je dodatno potkrijepljena izračunatim vrijednostima koeficijenta korelacije između brzine vozila i pozitivnih te negativnih akceleracija vozila koje iznose -51.4% te +20.4% (vidi tablicu 2-4; vrijednosti koeficijenata korelacija izračunate su primjenom Pythonove funkcije *corrcoef(.)* iz *numpy* modula [93]). U tablici 2-4 dane su vrijednosti indeksa korelacije izračunate nezavisno za slučaj voznih ciklusa snimljenih na relacijama BK-P te P-BK, gdje su vidljive određene razlike u odnosu na vrijednosti dobivene za objedinjene podatke (slučaj BK-P-BK). Konkretnije, za slučaj BK-P izraženije su korelacije između pozitivnih akceleracija i pozitivnih nagiba ceste te negativnih akceleracija i negativnih nagiba ceste, dok su za slučaj P-BK izraženije korelacije između brzine i pozitivnih te negativnih nagiba ceste te pozitivnih akceleracija i negativnih nagiba ceste. Međutim, budući da je broj snimljenih (i validnih) voznih ciklusa na relaciji BK-P znatno veći od onih snimljenih na P-BK (2313 naspram 903; vidi tablicu 2-2 u odjeljku 2.2.2), rezultati korelacijske analize dobiveni za slučaj BK-P-BK bliži su onima dobivenima za slučaj BK-P.



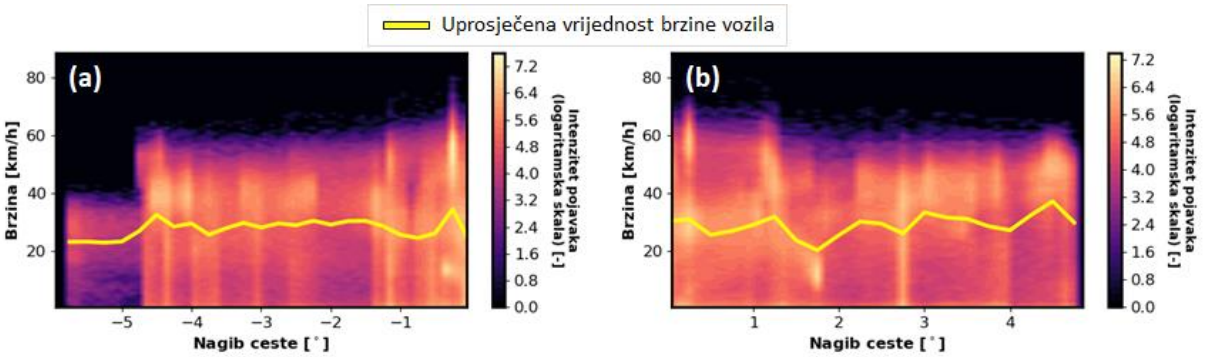
Slika 2-13: Toplinske mape a) pozitivnog i b) negativnog ubrzanja vozila s obzirom na brzinu vozila za objedinjenu rutu BK-P-BK, uključujući po brzini uprosječene vrijednosti ubrzanja vozila.

Tablica 2-4: Indeksi korelacije između brzine vozila, pozitivnog i negativnog ubrzanja vozila te pozitivnog i negativnog nagiba ceste (polja koja sadrže zamjetne korelacije označena su zelenom bojom).

Uspoređene varijable	**Indeksi korelacije [%]**		
	BK-P	P-BK	BK-P-BK
Brzina i pozitivna akceleracija	**-50.8**	**-53.5**	**-51.4**
Brzina i negativna akceleracija	**+18.8**	**+24.5**	**+20.4**
Brzina i pozitivni nagib ceste	+6.4	**+12.8**	+7.2
Brzina i negativni nagib ceste	+3.8	**-12.2**	+0.5
Pozitivna akceleracija i pozitivni nagib ceste	**-11.2**	-5.2	-9.4
Pozitivna akceleracija i negativni nagib ceste	+0.7	**-11.0**	-1.7
Negativna akceleracija i pozitivni nagib ceste	-6.9	-7.9	-8.3
Negativna akceleracija i negativni nagib ceste	**+10.6**	-2.2	+6.5

Ovisnost brzine vozila o nagibu ceste, prikazana na slici 2-14, pokazuje da se utjecaj nagiba ceste reflektira kroz nešto smanjene maksimalne vrijednosti brzine vozila s povećanjem apsolutne vrijednosti nagiba ceste. Karakteristika prosječne brzine vozila u ovisnosti o nagibu ceste (označena žutom linijom na slici 2-14) ne otkriva gotovo nikakav utjecaj u slučaju negativnih vrijednosti nagiba ceste (nepostojeća korelacija), dok za pozitivne vrijednosti nagiba

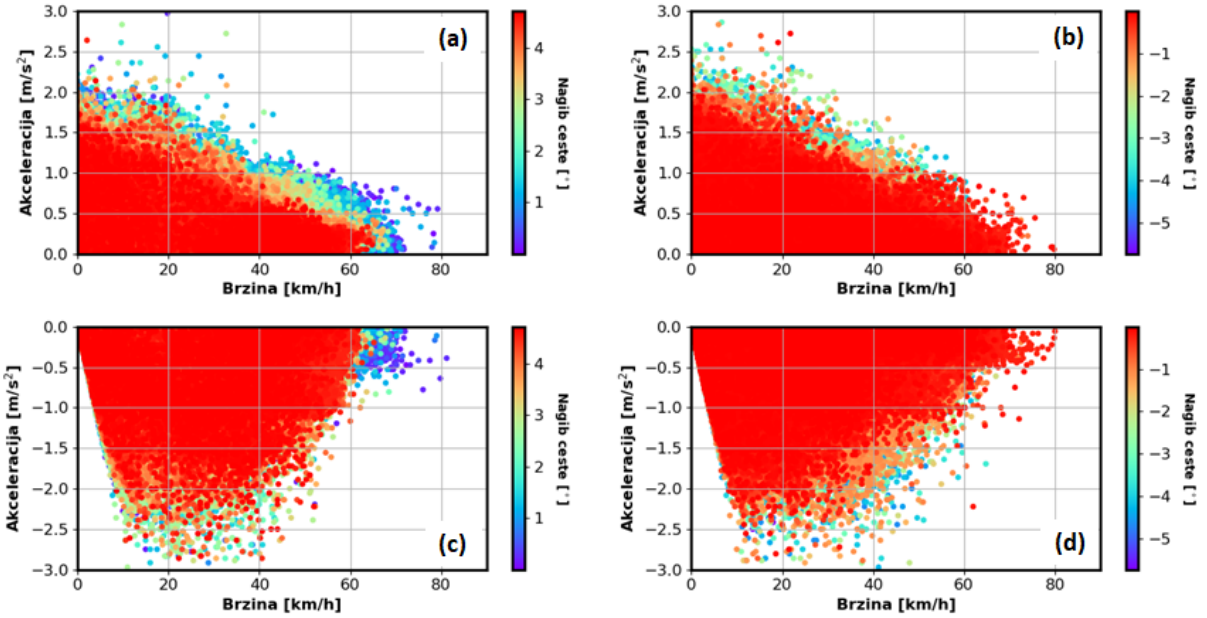
ceste pokazuje blago rastući trend, što se odražava u pozitivnoj korelaciji (vidi tablicu 2-4). Ovaj se trend može smatrati kontraintuitivnim, jer je za očekivati da se brzina vozila smanjuje s povećanjem nagiba ceste, što se može primijetiti u već opisanim vrijednostima maksimalne brzine vozila. Međutim, uočeni trend prosječne brzine potencijalno je rezultat karakteristika razmatrane rute koja sadrži otvorene dionice ceste pod nagibom na kojima je brzina vozila tipično veća pa stoga korelira s većim vrijednostima nagiba ceste.



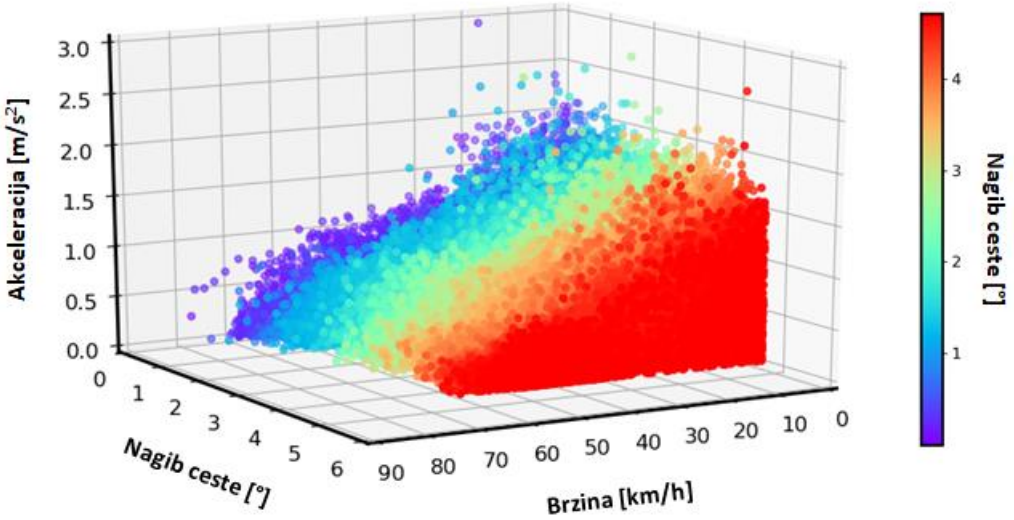
Slika 2-14: Toplinske mape brzine vozila s obzirom na a) negativne i b) pozitivne vrijednosti nagiba ceste za objedinjenu rutu BK-P-BK, uključujući po nagibu uprosječene vrijednosti brzine vozila.

Utjecaj nagiba ceste može se dodatno analizirati dodavanjem osi nagiba ceste u razdiobu akceleracije i brzine vozila (slika 2-15). Slika 2-15a i njezin 3D prikaz na slici 2-16 pokazuju kako se razdioba pozitivne akceleracije u odnosu na brzinu vozila značajno sužava uz povećanje nagiba ceste za slučaj vožnje uzbrdo, tj. pozitivnih nagiba ceste. Nešto manji, ali još uvijek postojeći utjecaj može se primijetiti u slučaju negativnih ubrzanja i pozitivnih vrijednosti nagiba ceste (slika 2-15c). U slučaju negativnih vrijednosti nagiba ceste taj je utjecaj gotovo zanemariv (slike 2-15b i 2-15d). Navedena zapažanja su potkrijepljena indeksima korelacije danima u tablici 2-4.

Gore navedena zapažanja sugeriraju da bi profil nagiba ceste trebao biti sintetiziran zajedno s brzinom i ubrzanjem vozila, kako bi se opisale njihove međuovisnosti koje su evidentne u stvarnim/snimljenim voznim ciklusima.



Slika 2-15: Razdioba pozitivnog i negativnog ubrzanja vozila u odnosu na brzinu vozila za a), c) pozitivne i b), d) negativne vrijednosti nagiba ceste za objedinjenu rutu BK-P-BK.



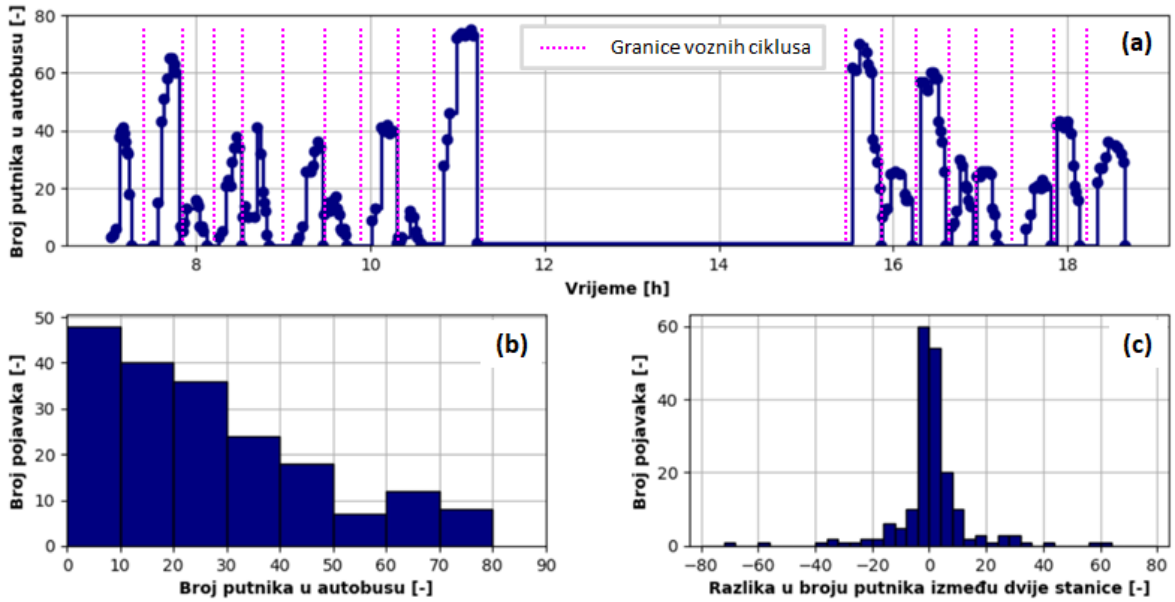
Slika 2-16: Objedinjena razdioba pozitivnog ubrzanja vozila u odnosu na brzinu vozila za pozitivne vrijednosti nagiba ceste.

2.4. Akvizicija broja putnika u autobusu i pripadna analiza

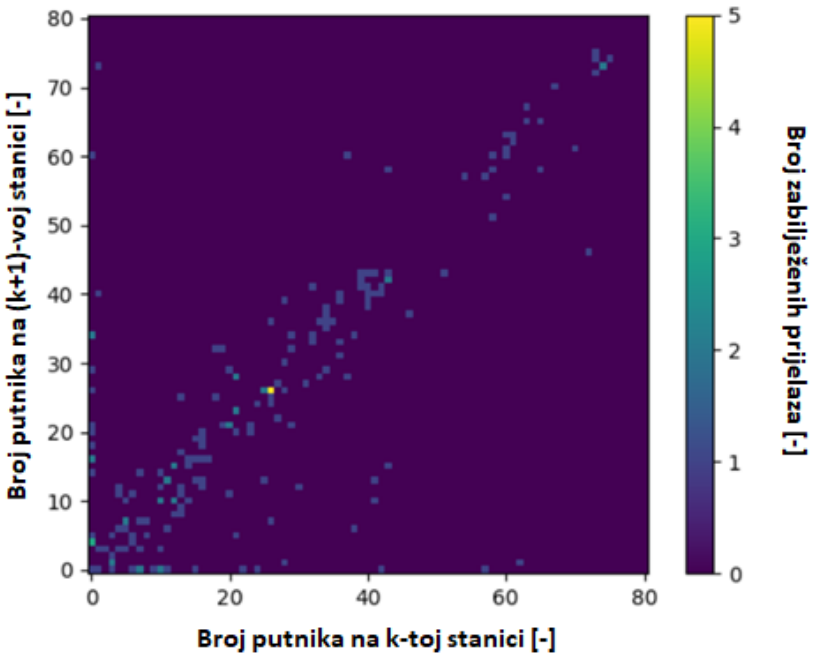
U svrhu uključivanja vremenski-promjenjive mase putnika u postupak sinteze voznih ciklusa, provedena je akvizicija broja putnika i to prebrojavanjem putnika u autobusu na razmatranoj ruti BK-P-BK. Prebrojavanje se izvodilo dvokratno, u jutarnjem i popodnevnom terminu, kako bi se obuhvatio čim raznolikiji profil broja putnika. Na taj način zabilježeno je ukupno 193 uzoraka broja putnika sadržanih u 18 voznih ciklusa. Prebrojavanje je provedeno istovremeno

od strane tri osobe raspoređene uz svako od troje vrata gradskog autobusa. Točnost prebrojavanja je potvrđena provjerom da je kumulativni zbroj putnika na kraju voznog ciklusa bio jednak nuli (autobus kreće prazan u novi vozni ciklus).

Na slici 2-17a prikazan je zabilježeni vremenski profil broja putnika koji varira od 0 do 75, s tim da je maksimalan deklariran broj putnika za dani tip autobusa jednak 80. Pritom su isprekidanim magenta linijama označene granice između individualnih voznih ciklusa, koje odgovaraju dolascima na krajnje stanice. Dulji vremenski razmaci odgovaraju čekanju na krajnjim stanicama. Slika 2-17b prikazuje razdiobu zabilježenog broja putnika, koja je unatoč relativno niskom broju uzoraka prilično solidno pokrivena od minimalnog do maksimalnog deklariranog broja putnika. Slika 2-17c prikazuje razdiobu razlike broja putnika između dviju stanica, a koja je na slici 2-18 prikazana u formi matrice prijelaza gdje svaka ćelija tj. element matrice sadrži broj prijelaza u zabilježenim podacima. Srednja vrijednost razlike broja putnika između dviju stanica iznosi 0, iz čega se može zaključiti kako je tipično broj ukrcanih putnika približno jednak broju iskrcanih putnika na stanici (slika 2-17c). Veći broj ukrcavanja i iskrcavanja putnika događa se relativno rijetko, osim inicijalno na krajnjim stanicama, što potkrjepljuje matrica prijelaza sa slike 2-18 na kojoj je vidljiva najveća koncentracija zabilježenih prijelaza broja putnika oko dijagonale matrice (relativno male promjene putnika na međustanicama), te na horizontalnoj i vertikalnoj osi (ukrcaji i iskrcaji putnika na krajnjim stanicama).

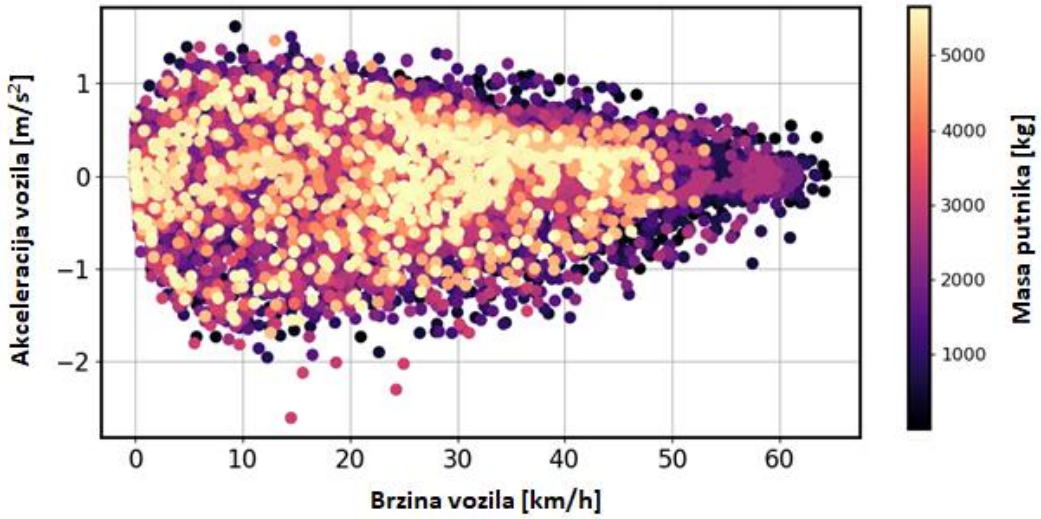


Slika 2-17: a) Vremenski profil broja putnika u autobusu na odabranoj ruti dobiven prebrojavanjem, te izvedeni histogrami b) broja putnika u autobusu i c) razlike broja putnika između dviju stanica.



Slika 2-18: Matrica prijelaza broja putnika između susjednih stanica.

Jasno je da masa, uz brzinu i akceleraciju vozila te nagib ceste, utječe na potrošnju goriva (vidi izraz (3-6)). No, unaprijed nije jasno postoji li korelacija između mase putnika i brzine te akceleracije vozila. Na slici 2-19 prikazana je ovisnost snimljenih akceleracija o brzini i masi putnika, gdje se može vidjeti kako zaliha akceleracije i deceleracije značajno opada s povećanjem brzine vozila (značajna međukorelacija; tablica 2-4). Povećanjem mase vozila ova razdioba se dodatno sužava (točke označene svjetlijim bojama). Međutim, ispitivanje međuovisnosti mase putnika i brzine te akceleracije vozila pokazuju kako između brzine vozila i mase putnika postoji slaba korelacija (indeks korelacije iznosi 16%), dok između akceleracije vozila i mase putnika ne postoji korelacija (indeks korelacije približno jednak 0). S druge strane, indeksi korelacije između nagiba ceste i brzine te akceleracije vozila su bili konzistentno udaljeni od nule (tablica 2-4). Stoga se pretpostavlja raspregnutost mase putnika s brzinom i akceleracijom vozila te nagibom ceste, te se u poglavlju 3 razmatra metoda nezavisne sinteze mase putnika. Za određenije zaključke potreban je puno širi skup snimljenih podataka o masi putnika, kojeg je teško dobiti bez ugrađenih senzora mase šasije, što je rijetko dostupno.



Slika 2-19: Razdioba snimljenih akceleracija vozila u odnosu na brzinu vozila i masu putnika.

3. Sinteza višedimenzionalnih voznih ciklusa

U ovom se poglavlju razmatra sinteza višedimenzionalnih voznih ciklusa primjenom stohastičke metode Markovljevih lanaca. U prvom potpoglavlju daje se teorijska podloga Markovljevih lanaca i konceptualni prikaz cjelokupnog postupka sinteze voznih ciklusa, uključujući osnove njihove validacije. U drugom potpoglavlju izvodi se osam različitih modela Markovljevog lanca s obzirom na različite izbore stanja i pripadne dimenzije matrice prijelaznih vjerojatnosti (TPM). Kako bi se smanjio zahtjev na memoriju i poboljšala računalna učinkovitost, predlaže se implementacija TPM u obliku rijetke matrice temeljene na rječniku ključeva. Za potrebe usvajanja konačnog modela voznih ciklusa provodi se usporedna analiza računalnih performansi svakog od predloženih modela. U trećem potpoglavlju predstavlja se koncept nezavisne sinteze vremenskog niza mase putnika i način njegove integracije u proces sinteze voznog ciklusa, te se demonstriraju rezultati primjene predložene metode sinteze voznih ciklusa.

3.1. Koncept sinteze voznih ciklusa

3.1.1. Teorijska podloga Markovljevih lanaca

Markovljev lanac je stohastički model koji opisuje niz mogućih događaja, tj. stanja u kojima vjerojatnost svakog sljedećeg događaja $\mathbb{P}(Z_{n+1})$ ovisi samo o stanju Z_n iz prethodnog koraka n, tj. ne ovisi o prošlim stanjima $Z_k;\ k = 0, 1, \dots, n-1\ (n \geq 1)$. Ovo svojstvo naziva se Markovljevo svojstvo [94, 95], čija matematička formulacija glasi

$$\mathbb{P}(Z_{n+1} = j \mid Z_n = i_n, Z_{n-1} = i_{n-1}, \dots, Z_0 = i_0) = \mathbb{P}(Z_{n+1} = j \mid Z_n = i_n), \tag{3-1}$$

gdje $i_0, i_1, \dots, i_n, j \in \mathbb{S}$ označavaju realizacije stanja sustava u određenim diskretnim vremenskim trenutcima. Svaki proces koji zadovoljava Markovljevo svojstvo naziva se Markovljevim procesom, dok je Markovljev lanac vrsta Markovljevog procesa koja razmatra diskretni (prebrojivi) prostor stanja $\mathbb{S}$. Markovljevi lanci naširoko se koriste kao statistički modeli raznih procesa [96].

U svakom diskretnom koraku sustav može prijeći u neko novo stanje ili može ostati u istom stanju. Koraci sustava najčešće se odnose na diskretne vremenske trenutke, ali se mogu odnositi i na fizičku udaljenost ili bilo koje drugo diskretno mjerenje. Promjene stanja sustava nazivaju se tranzicije, dok se uvjetne vjerojatnosti $p_{i.j} \coloneqq \mathbb{P}(Z_{n+1} = j \mid Z_n = i)$ povezane s tranzicijama

između različitih stanja nazivaju prijelazne vjerojatnosti. Razdiobe prijelaznih vjerojatnosti između stanja definiraju se kroz matricu prijelaznih vjerojatnosti **Π,** koja se definira kao

$$\mathbf{\Pi} = \left[p_{i,j}\right]_{i,j\in\mathbb{S}} = \left[\mathbb{P}(Z_{n+1} = j \mid Z_n = i)\right]_{i,j\in\mathbb{S}}, \tag{3-2}$$

gdje početno stanje i odgovara broju retka u matrici **Π**, dok konačno stanje j odgovara broju stupca. Za slučaj konačnog prostora stanja $\mathbb{S} = \{i_0, i_1, \dots, i_N\}$ matrica **Π** poprima dimenzije $\mathrm{N} \times \mathrm{N}$

$$\mathbf{\Pi} = \begin{bmatrix} p_{1,1} & \cdots & p_{1,N} \\ \vdots & \ddots & \vdots \\ p_{N,1} & \cdots & p_{N,N} \end{bmatrix}. \tag{3-3}$$

Prema zakonu ukupne vjerojatnosti [94] slijedi da zbroj svih vjerojatnosti prijelaza između stanja mora zadovoljiti uvjet

$$\sum_j p_{i.j} = \sum_j \mathbb{P}(Z_{n+1} = j \mid Z_n = i) = 1, \tag{3-4}$$

za svaki indeks retka $i \in \mathbb{S}$. Za stanje $k \in \mathbb{S}$ se kaže da je apsorbirajuće ako je $p_{k.k} = 1$, odnosno, ako se jednom uđe u to stanje, nemoguće je iz njega izaći.

Markovljev lanac u potpunosti je opisan prostorom stanja $\mathbb{S}$, pripadnom matricom prijelaznih vjerojatnosti **Π**, te početnim stanjem sustava Z_0 (ili početnom razdiobom vjerojatnosti opisanom vektorom retkom $p_0 = [\mathbb{P}(Z_0 = i_o), \mathbb{P}(Z_0 = i_1), \dots, \mathbb{P}(Z_0 = i_N)] \in \mathbb{R}^{1\times\mathrm{N}}$). Prema konvenciji se pretpostavlja da su sva moguća stanja i tranzicije uključeni u definiciju procesa, tako da uvijek postoji sljedeće stanje koje proces može poprimiti. Budući da se stanje sustava zbog svoje stohastičke prirode nasumično mijenja, generalno je nemoguće sa sigurnošću predvidjeti stanje Markovljevog lanca u određenoj točki u budućnosti. Međutim, prema Markovljevom svojstvu, moguće je predvidjeti statistička svojstva budućnosti sustava, tj. razdiobu vjerojatnosti *m* vremenskih koraka u budućnosti

$$\begin{gathered} p_1 = p_0\mathbf{\Pi} \\ p_2 = p_1\mathbf{\Pi} = (p_0\mathbf{\Pi})\mathbf{\Pi} = p_0\mathbf{\Pi}^2 \\ \vdots \\ p_m = p_{m-1}\mathbf{\Pi} = p_0\mathbf{\Pi}^m \end{gathered} \tag{3-5}$$

Ovo pravilo vrijedi samo ako je Markovljev lanac vremenski homogen [94], tj. ako matrica prijelaznih vjerojatnosti **Π** ostaje ista nakon svakog vremenskog koraka.

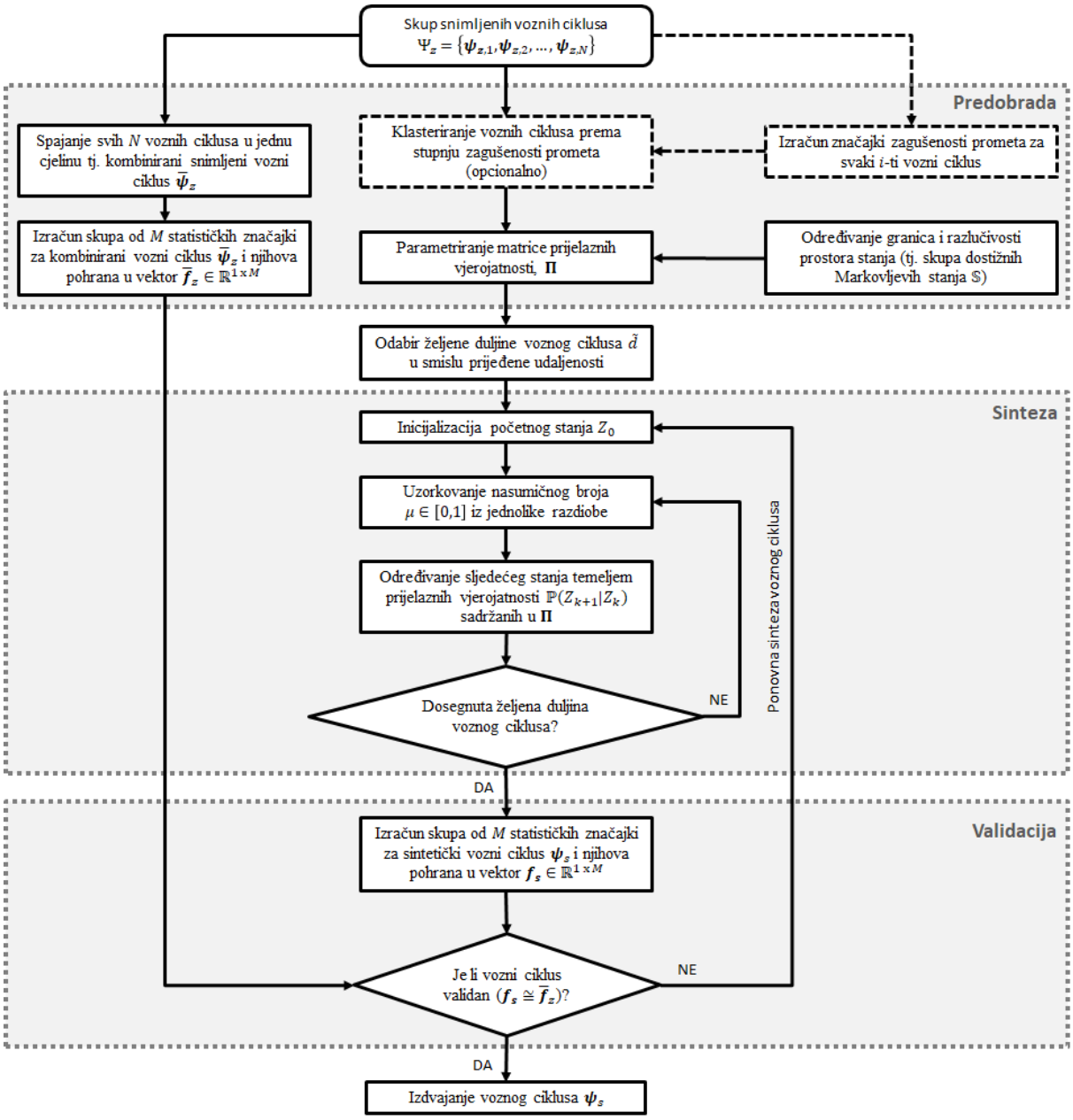
U ovom se radu smatra da je $(Z_n)_{\,n\in\mathbb{N}}$ stohastički proces koji je diskretan u vremenu te poprima vrijednosti u diskretnom prostoru stanja $\mathbb{S}$. Pritom se snimljeni vozni ciklusi predstavljeni u

obliku vremenskih nizova brzine, akceleracije i nagiba ceste koriste kao sljedovi prijelaza između različitih stanja Markovljevog lanca.

3.1.2. Konceptualni prikaz cjelokupnog postupka sinteze voznih ciklusa

Sinteza voznih ciklusa uobičajeno se odvija u četiri glavna koraka [21, 27]: (i) grupiranje/klasteriranje snimljenih voznih ciklusa u karakteristične skupine prema određenim statističkim značajkama, (ii) parametriranje modela voznih ciklusa tj. matrice prijelaznih vjerojatnosti **Π** na povijesnim podacima vožnje, (iii) generiranje bogatog skupa sintetičkih voznih ciklusa, te (iv) validacija generiranih sintetičkih ciklusa temeljem specificiranih statističkih značajki u frekvencijskoj ili vremenskoj domeni, te izdvajanje nekolicine reprezentativnih. Cjelokupni postupak sinteze voznih ciklusa temeljen na metodi Markovljevih lanaca prikazan je blokovskim dijagramom na slici 3-1. Ulazni podaci uključuju skup od $i = 1,2,\dots,N_{DC.F}^{V}$ snimljenih voznih ciklusa s finom rezolucijom potrošnje goriva $\Psi_z = \{\boldsymbol{\psi}_{z,1}, \boldsymbol{\psi}_{z,2}, \dots, \boldsymbol{\psi}_{z,N}\}$ (vidi potpoglavlje 2.2), gdje je svaki i-ti snimljeni vozni ciklus opisan diskretnim vremenskim profilima brzine $\boldsymbol{v}$ i akceleracije $\boldsymbol{a}$ vozila, te nagiba ceste $\boldsymbol{\theta}$ i njegove vremenske derivacije $\dot{\boldsymbol{\theta}}$; $\boldsymbol{\psi}_{z,i} = [\boldsymbol{v}^T\ \boldsymbol{a}^T\ \boldsymbol{\theta}^T\ \dot{\boldsymbol{\theta}}^T]$ (gdje su $\boldsymbol{v}$, $\boldsymbol{a}$, $\boldsymbol{\theta}$ i $\dot{\boldsymbol{\theta}}$ vektor-retci).

U fazi predobrade snimljeni vozni ciklusi sadržani u skupu Ψ_z najprije se dijele u tri karakteristične grupe prema stupnju zagušenosti prometa na način kako je opisano u poglavlju 2 i dano na slici 2-9, rezultirajući tako skupom voznih ciklusa označenim kao Ψ_{LZ} za laganu, Ψ_{SZ} za srednju te Ψ_{TZ} za tešku zagušenost prometa. Općenito gledano, klasteriranje voznih ciklusa može se smatrati opcionalnim korakom ukoliko se želi dobiti model voznih ciklusa koji objedinjuje karakteristike svih snimljenih obrazaca vožnje, ali pod uvjetom da ne postoji značajna razlika između tih obrazaca (npr. miješanja gradske vožnje i vožnje autocestom). Razlog tome je što takvo miješanje podataka može dovesti do pojave da tako generirani modeli tj. matrica prijelaznih vjerojatnosti **Π,** a time i generirani sintetički vozni ciklusi Ψ_s, ne uspiju precizno predstaviti karakteristike niti jedne od razmatranih kategorija [28]. Jednom kada su određeni ulazni skupovi podataka, provodi se popunjavanje matrice prijelaznih vjerojatnosti **Π** prebrojavanjem tranzicija između Markovljevih stanja opisanih snimljenim podatkom o brzini i akceleraciji vozila te nagiba ceste, $\mathbb{P}(Z_{n+1} | Z_n)$. Prethodno je potrebno definirati raspone (granice) i razlučivosti Markovljevih stanja kako bi se u potpunosti definirao obuhvaćeni prostor stanja $\mathbb{S}$.



Slika 3-1: Blokovski dijagram metode sinteze voznih ciklusa temeljene na Markovljevim lancima, uključujući osnove validacije voznih ciklusa.

Faza sinteze voznih ciklusa inicijalno zahtijeva odabir željene duljine sintetičkih voznih ciklusa u smislu prijeđenog puta $\bar{d}$, koji predstavlja primarni kriterij prema kojem se prekida postupak generiranja voznog ciklusa. Zatim se definiraju početne vrijednosti Markovljevog stanja Z_0, koje se tipično postavljaju na nulu. Sinteza voznog ciklusa $\boldsymbol{\psi}_s$ provodi se uzorkovanjem iz prethodno popunjene matrice prijelaznih vjerojatnosti $\mathbf{\Pi}$, polazeći od početne vrijednosti Markovljevog stanja Z_0. Uzorkovanje $\mathbf{\Pi}$ ostvaruje se pomoću generatora slučajnih brojeva $\mu \in [0, 1]$ (uzorkovanih iz jednolike razdiobe). Konkretnije, za svaki diskretni vremenski korak $k = 0, 1, \ldots, K$ sljedeće stanje Z_{k+1} određuje se uzorkovanjem iz pripadajuće razdiobe prijelaznih

vjerojatnosti $\mathbb{P}(Z_{k+1}|Z_k)$ za trenutno stanje Z_k. Iterativnim postupkom uzorkovanja generira se sintetički vozni ciklus $\boldsymbol{\psi}_s$ propisane duljine $\bar{d}$, koja je u ovom radu postavljena na 5.1 km kako bi odgovarala srednjoj prijeđenoj udaljenosti dobivenoj za slučaj objedinjenih podataka oba smjera vožnje (vidi sliku 2-10).

Kad je jednom generiran, sintetički vozni ciklus $\boldsymbol{\psi}_s$ podvrgava se postupku validacije kako bi se provjerilo u kojoj mjeri odražava prosječne značajke svih snimljenih voznih ciklusa. U tu svrhu uvodi se kombinirani snimljeni vozni ciklus $\bar{\boldsymbol{\psi}}_z$, koji se dobiva spajanjem svih individualnih snimljenih voznih ciklusa u jednu cjelinu. Postupak validacije temelji se na usporedbi vrijednosti ograničenog skupa od M statističkih značajki, koje se u fazi predobrade izračunavaju za kombinirani snimljeni vozni ciklus $\bar{\boldsymbol{\psi}}_z$ ($\bar{\boldsymbol{f}}_z \in \mathbb{R}^{1\,\mathrm{x}\,M}$), te u fazi sinteze zasebno za svaki sintetički vozni ciklus $\boldsymbol{\psi}_s$ ($\boldsymbol{f_s} \in \mathbb{R}^{1\,\mathrm{x}\,M}$). Sintetički vozni ciklusi čije su statističke značajke $\boldsymbol{f_s}$ bliže srednjim vrijednostima kombiniranog voznog ciklusa $\bar{\boldsymbol{f}}_z$ (odnosno, za koje vrijedi da je $\boldsymbol{f_s} \cong \bar{\boldsymbol{f}}_z$), smatraju se reprezentativnijima. Detaljni opis postupka validacije sintetičkih voznih ciklusa dan je u poglavlju 5.

3.2. Tvorba matrice prijelaznih vjerojatnosti

3.2.1. Koncepti realizacije

Inicijalni korak nužan za modeliranje Markovljevog lanca je odabir broja stanja. Potencijalne kandidate stanja Markovljevog lanca moguće je izdvojiti analizirajući jednadžbu uzdužnog gibanja vozila [64]:

$$F_w(k) = \underbrace{m(k)\cdot a(k)}_{\text{Inercijska sila}} + \underbrace{m(k)\cdot g\cdot sin(\theta(k))}_{\text{Sila svladavanja uspona}} + \\ + \underbrace{m(k)\cdot R_o\cdot cos(\theta(k))}_{\text{Otpor kotrljanja}} + \underbrace{0.5\cdot \rho_z\cdot C_d\cdot A_f\cdot v(k)^2}_{\text{Aerodinamički otpor}}, \tag{3-6}$$

gdje je F_w sila prenesena na podlogu od strane pogonskih kotača, m je masa vozila (zajedno s putnicima), a je ubrzanje vozila, v je brzina vozila, θ je nagib ceste, g je gravitacijska konstanta, R_0 je koeficijent otpora kotrljanja, ρ_z je gustoća zraka, C_d je koeficijent aerodinamičkog otpora, a A_f je čeona površina vozila.

Prema (3-6) slijedi da u skup vremenski-promjenjivih varijabli relevantnih s aspekta gibanja vozila i potrošnje energije spadaju: 1) brzina vozila v, 2) ubrzanje vozila a, 3) nagib ceste θ, te 4) masa vozila m koja se u konkretnom slučaju mijenja prilikom dolaska autobusa na svaku stanicu. Prilikom odabira stanja Markovljevog lanca u svrhu sinteze voznih ciklusa, treba imati

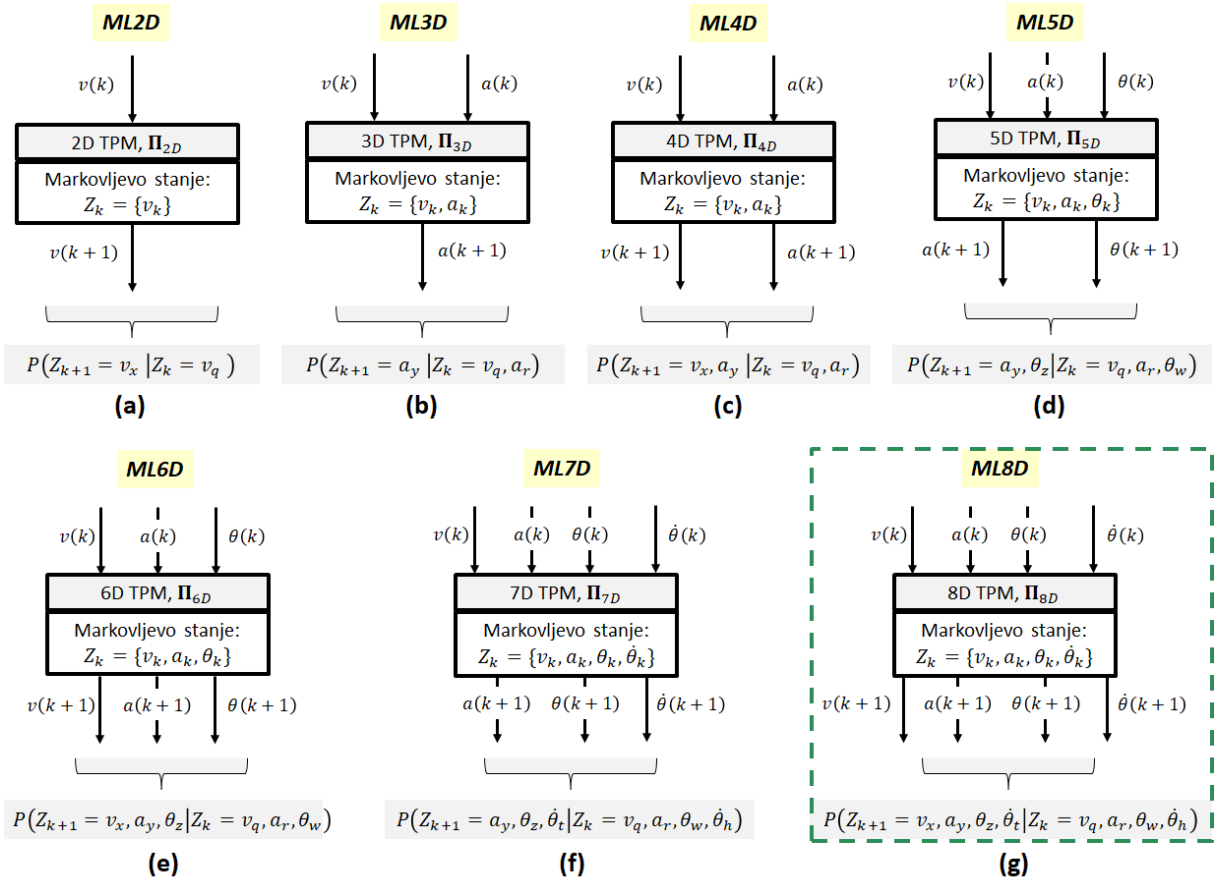
na umu sljedeće. Analize iz [21, 59] pokazale su kako sintetički profili brzine dobiveni uporabom isključivo brzine vozila kao stanja Markovljevog lanca uključuju nerealne visokofrekvencijske oscilacije, kao i nerealan broj i oblik zaustavljanja vozila. Uključivanjem ubrzanja vozila u stanje Markovljevog lanca dobivaju se realističniji profili brzine i ubrzanja vozila, čije razdiobe statističkih značajki statistički vjerodostojno rekonstruiraju razdiobe snimljenih voznih ciklusa [21, 59]. Razlog tome je što ubrzanje vozila pruža informacije o trendu brzine tj. namjeri vozača, čime se osigurava da faze usporavanja, krstarenja ili ubrzanja vozila traju dovoljno dugo, tj. da se ne događaju nasumično tijekom prijelaza iz jednog stanja u drugo. Uključivanjem i nagiba ceste u stanje Markovljevog lanca pruža se mogućnost generiranja profila nagiba ceste kao dodatnog izlaza [47, 48, 59], koji je izuzetno važan sa stajališta potrošnje goriva. Kao i u slučaju ubrzanja vozila (tj. derivacije brzine vozila), uključivanjem vremenske derivacija nagiba ceste u stanje Markovljevog lanca dobivaju se vjerniji profili nagiba ceste (glatkiji, s boljim podudaranjem frekvencija) [52]. Korelacijskom analizom opisanom u odjeljku 2.3.2 pokazano je kako je sintezu voznih ciklusa preporučljivo izvoditi združeno s obzirom na brzinu i ubrzanje vozila te stanje nagiba ceste, kako bi se osigurale dosljedne značajke međusobne korelacije pojedinih sintetiziranih profila [47, 60]. Budući da s povećanjem varijabli stanja Markovljevog lanca značajno rastu memorijski zahtjevi TPM, valja razmotriti neizravna rješenja realizacije TPM koja sljedeće stanje brzine vozila izračunavaju temeljem trenutne vrijednosti brzine i ubrzanja vozila.

Slijedeći gore navedene smjernice te polazeći od opće definicije Markovljevog lanca (3-1) i odabira različitih skupova varijabli za opisivanje stanja sustava Z, moguće je realizirati matricu prijelaznih vjerojatnosti **Π** (engl. skrać. TPM) različitih dimenzionalnosti, kako je prikazano na slici 3-2. Stoga se dalje u tekstu za različite izvedbe **Π**, koje predstavljaju pripadne modele Markovljevih lanaca, uvode oznake *MLxD*, gdje *x* poprima vrijednost u rasponu od 2 do 8 i označava dimenzionalnost **Π**.

Model *ML2D* (slika 3-2a) razmatra samo diskretne vrijednosti brzine vozila v kao stanja Markovljevog lanca, gdje je razdioba prijelaznih vjerojatnosti između diskretnih stanja brzine predstavljena dvodimenzionalnom (2D) TPM kao

$$p_{q,x} \coloneqq P\left(Z_{k+1} = v_x \middle| Z_k = v_q\right); \quad \mathbf{\Pi}_{2D} \coloneqq \left[p_{q,x}\right], \tag{3-7}$$

pri čemu element matrice $p_{q,x}$ označava vjerojatnost prijelaza iz stanja brzine v_q u stanje v_x.



Slika 3-2: Prikaz različitih koncepata (modela) realizacije Markovljevog lanca u smislu dimenzionalnosti matrice prijelaznih vjerojatnosti $\boldsymbol{\Pi}$.

Model *ML3D* (slika 3-2b) razmatra kombinaciju diskretnih vrijednosti brzine v i ubrzanja a vozila kao stanja Markovljevog lanca. Razdioba vjerojatnosti prijelaza između diskretnih stanja brzine i ubrzanja u ovom slučaju definirana je kroz 3D TPM kao

$$p_{qr,y} \coloneqq P\left(Z_{k+1} = a_y \middle| Z_k = v_q, a_r\right); \quad \boldsymbol{\Pi}_{3D} \coloneqq \left[p_{qr,y}\right], \tag{3-8}$$

gdje vektorski element $p_{qr,y}$ označava vjerojatnost prijelaza iz trenutnog stanja brzine i ubrzanja (v_q, a_r) u k-tom diskretnom vremenskom koraku u sljedeće stanje ubrzanja a_y u $(k+1)$-vom diskretnom vremenskom koraku. Sljedeće stanje brzine $v_{k+1} \equiv v_x$ dobiva se temeljem trenutnih vrijednosti brzine $v_k \equiv v_q$ i ubrzanja vozila $a_k \equiv a_r$ prema

$$v_{k+1} = v_k + a_k \cdot \Delta T \equiv v_x = v_q + a_r \cdot \Delta T. \tag{3-9}$$

Neizravnim tj. naknadnim izračunavanjem sljedećeg stanja brzine vozila (bez njegova uključivanja u **Π)** smanjuje se dimenzionalnost **Π**, čime se povećava računalna učinkovitost cjelokupnog postupka sinteze voznih ciklusa te smanjuju zahtjevi za memorijom.

Model *ML4D* (slika 3-2c) razlikuje se od *ML3D* u tome što se sljedeće stanje brzine v_x dobiva izravno iz razdiobe prijelaznih vjerojatnosti definiranim kroz 4D TPM kao

$$p_{qr,xy} \coloneqq P\left(Z_{k+1} = v_x, a_y \middle| Z_k = v_q, a_r\right); \quad \mathbf{\Pi}_{4D} \coloneqq \left[p_{qr,xy}\right], \tag{3-10}$$

gdje element matrice $p_{qr,xy}$ označava vjerojatnost prijelaza iz trenutnih stanja ubrzanja i brzine (v_q, a_r) u sljedeća stanja ubrzanja i brzine (v_x, a_y).

Model *ML5D* (slika 3-2d) uz brzinu v i akceleraciju a vozila koristi i diskretne vrijednosti nagiba ceste θ za određivanje stanja Markovljevog lanca. U ovom slučaju razdioba prijelaznih vjerojatnosti između diskretnih stanja brzina, ubrzanja i nagiba ceste definirana je kroz 5D TPM kao

$$p_{qrw,yz} \coloneqq P\left(Z_{k+1} = a_y, \theta_z \middle| Z_k = v_q, a_r, \theta_w\right); \quad \mathbf{\Pi}_{5D} \coloneqq \left[p_{qrw,yz}\right], \tag{3-11}$$

gdje element matrice $p_{qrw,yz}$ označava vjerojatnost prijelaza iz trenutnog stanja brzine, ubrzanja i nagiba ceste (v_q, a_r, θ_w) u sljedeće stanje ubrzanja i nagiba ceste (a_y, θ_z), iz kojeg se potom dobiva sljedeće stanje brzine v_x prema (3-9). Motivacija reduciranja dimenzionalnosti $\mathbf{\Pi}$ je ista kao u slučaju modela *ML3D*.

Model *ML6D* (slika 3-2e) nadograđuje model *ML5D* samo kroz izračunavanje sljedećeg stanja brzine v_x izravno iz razdiobe prijelaznih vjerojatnosti definiranim kroz 6D TPM kao

$$p_{qrw,xyz} \coloneqq P\left(Z_{k+1} = v_x, a_y, \theta_z \middle| Z_k = v_q, a_r, \theta_w\right); \quad \mathbf{\Pi}_{6D} \coloneqq \left[p_{qrw,xyz}\right]. \tag{3-12}$$

Model *ML7D* (slika 3-2f) se razlikuje od modela *ML5D* samo po tome što uključuje dodatnu informaciju o vremenskoj derivaciji nagiba ceste $\dot{\theta}$ kod određivanje stanja Markovljevog lanca. Time se razdiobe prijelaznih vjerojatnosti između stanja definiraju kroz 7D TPM kao

$$p_{qrwh,yzt} \coloneqq P\left(Z_{k+1} = a_y, \theta_z, \dot{\theta}_t \middle| Z_k = v_q, a_r, \theta_w, \dot{\theta}_h\right); \quad \mathbf{\Pi}_{7D} \coloneqq \left[p_{qrwh,yzt}\right], \tag{3-13}$$

dok motivacija za reduciranje dimenzionalnosti $\mathbf{\Pi}$ ostaje ista kao u slučaju modela *ML3D* i modela *ML5D*.

Model *ML8D* (slika 3-2g) nadograđuje model *ML7D* samo kroz izračunavanje sljedećeg stanja brzine v_x izravno iz razdiobe prijelaznih vjerojatnosti definiranim kroz 8D TPM kao

$$p_{qrwh,xyzt} \coloneqq P\left(Z_{k+1} = v_x, a_y, \theta_z, \dot{\theta}_t \middle| Z_k = v_q, a_r, \theta_w, \dot{\theta}_h\right); \ \mathbf{\Pi}_{8D} \coloneqq \left[p_{qrwh,xyzt}\right]. \tag{3-14}$$

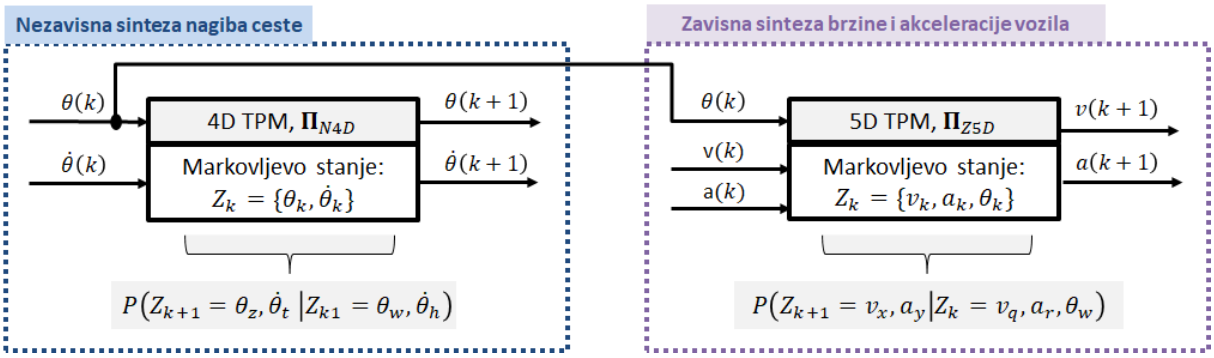
Alternativni koncept realizacije modela voznih ciklusa koji razmatra hijerarhijsku sintezu voznih ciklusa prikazan je na slici 3-3 (dalje označen kao *ML-HIER*). Glavna motivacija ovog pristupa je činjenica da sljedeće stanje brzine i akceleracije vozila ovisi o nagibu ceste, dok

suprotno ne vrijedi**.** Stoga metoda podrazumijeva nezavisnu sintezu profila nagiba ceste primjenom $\mathbf{\Pi}_{N4D}$, koji se potom koristi kao ulaz za zavisnu sintezu profila brzine i akceleracije primjenom $\mathbf{\Pi}_{Z5D}$. Time se model raščlanjuje na dvije TPM ($\mathbf{\Pi}_{N4D}$ i $\mathbf{\Pi}_{Z5D}$), čije su razdiobe prijelaznih vjerojatnosti redom definirane kao

$$p_{wh,zt} \coloneqq P\left(Z_{k+1} = \theta_z, \dot{\theta}_t \middle| Z_k = \theta_w, \dot{\theta}_h\right); \quad \mathbf{\Pi}_{N4D} \coloneqq \left[p_{wh,zt}\right], \tag{3-15}$$

$$p_{qrw,xy} \coloneqq P\left(Z_{k+1} = v_x, a_y \middle| Z_k = v_q, a_r, \theta_w\right); \quad \mathbf{\Pi}_{Z5D} \coloneqq \left[p_{qrw,xy}\right]. \tag{3-16}$$

Primjenom $\mathbf{\Pi}_{N4D}$ na trenutno stanje nagiba ceste $\theta(k) \equiv \theta_w$ i $\dot{\theta}(k) \equiv \dot{\theta}_h$ dobivaju se sljedeća stanja nagiba ceste $\theta(k+1) \equiv \theta_z$ i $\dot{\theta}(k+1) \equiv \dot{\theta}_t$ dok se $\theta(k)$ dalje koristi zajedno s trenutnom brzinom $v(k) \equiv v_q$ i akceleracijom $a(k) \equiv a_r$ vozila kao ulaz u $\mathbf{\Pi}_{Z5D}$ kako bi se dobila sljedeća stanja brzine $v(k+1) \equiv v_x$ i akceleracije $a(k+1) \equiv a_y$. Tako izračunata nova stanja θ_z, $\dot{\theta}_t$, v_x, a_y postaju ulazi modela u sljedećem koraku. Pritom se sinteza izvodi za $k = 1, 2, \ldots, K$ diskretnih vremenskih koraka, uz rubne uvjete $\theta(0) = \theta_0$ i $\dot{\theta}(0) = \dot{\theta}_0$, te $v(0) = v_0$ i $a(0) = a_0$.

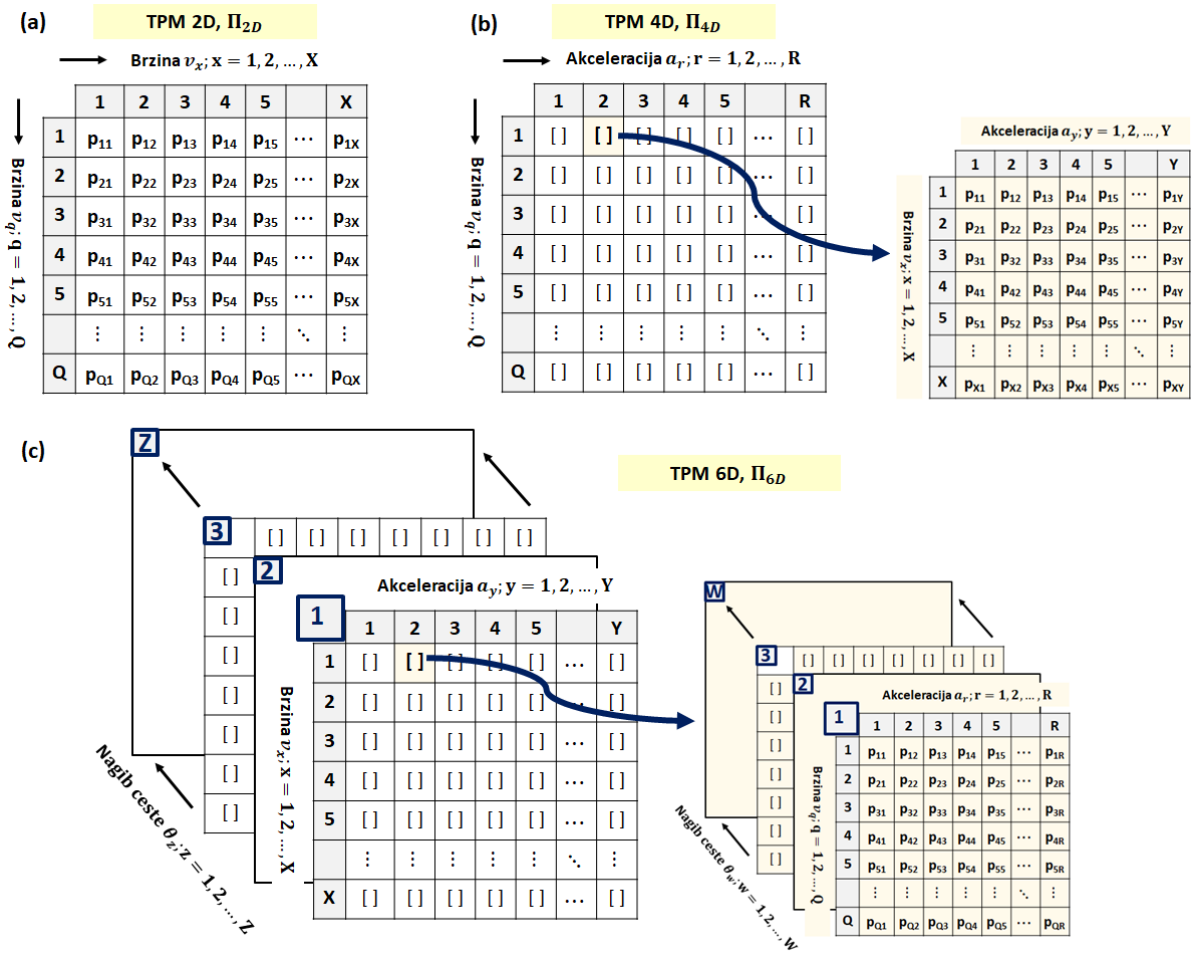


Slika 3-3: Koncept modela temeljenog na hijerarhijskoj sintezi voznih ciklusa (ML-HIER) koji se sastoji od nezavisne $\mathbf{\Pi}_{N4D}$ i zavisne $\mathbf{\Pi}_{Z5D}$ matrice prijelaznih vjerojatnosti.

3.2.2. Implementacijski aspekti

TPM dimenzije *N* tradicionalno se realizira u obliku 2*N*-dimenzionalnog niza, gdje *N* ovisi o broju razmatranih varijabli stanja Markovljevog lanca [27]. Na primjer, TPM za Markovljev lanac koji razmatra samo jednu varijablu stanja (model *ML2D* opisan s $\mathbf{\Pi}_{2D}$) određen je 2D-nizom, dok je za slučaj dvije varijable stanja (model *ML4D* opisan s $\mathbf{\Pi}_{4D}$) dan 4D nizom, itd. (za grafičku ilustraciju vidi sliku 3-4). Glavni nedostatak ovog pristupa implementacije je u tome što rezultirajući TPM sadrži mnogo elemenata (ćelija) jednakih nuli, tj. mnoge prijelazne vjerojatnosti iznose nula (tzv. svojstvo „rijetkosti“, *engl. sparsity*), što dovodi do prekomjernih memorijskih zahtjeva [97]. Pored toga, ovaj pristup karakterizira niža brzina izvođenja, budući da se pri formuliranju razdioba prijelaznih vjerojatnosti razmatra puni prostor stanja $\mathbb{S}$, odnosno

sve prijelazne vjerojatnosti, uključujući one jednake nuli. Sukladno tome, ovaj pristup može dovesti do problema s memorijom u slučaju TPM-a velike dimenzije (npr. *ML8D*), čime se nameću ograničenja na dimenzionalnost TPM-a, tj. na razlučivost stanja Markovljevog lanca, što posljedično smanjuje točnost sinteze voznog ciklusa.

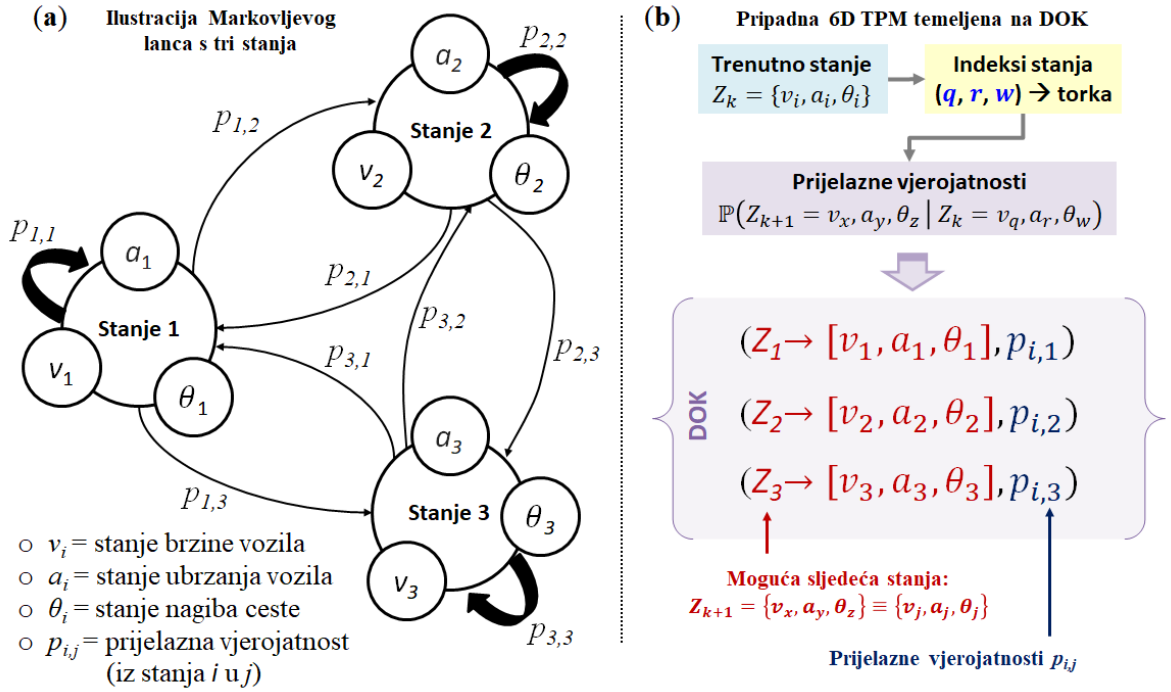


Slika 3-4: Ilustracija a) 2D , b) 4D i c) 6D TPM-a implementiranog u obliku niza.

Kako bi se riješio problem velikog zahtjeva za memorijom i niske računalne učinkovitosti, u ovom radu predlaže se implementacija visoko-dimenzionalnih TPM-ova u obliku „rijetke“ (*engl. sparse*) matrice temeljene na rječniku ključeva (*engl. Dictionary of Keys*, DOK [98]). Rječnik predstavlja podatkovnu strukturu koja je općenitije poznata kao asocijativni niz, koji se tipično izvodi u obliku hash-tablice (*engl. hash table*), a sastoji se od zbirke parova ključ-vrijednost, gdje svaki par preslikava ključ na njegovu pridruženu vrijednost. Hash tablica je podatkovna struktura koja koristi hash funkciju za učinkovito preslikavanje određenih ključeva u njima pridružene vrijednosti. Pritom se hash funkcija koristi za transformiranje ključa u indeks (tzv. hash kôd), koji određuje mjesto u nizu na kojemu je potrebno tražiti odgovarajuću vrijednost. Temeljna razlika između niza i rječnika je u tome što se elementima niza pristupa

putem indeksa koji mora biti cijeli broj, dok se elementima rječnika pristupa putem ključeva koji mogu biti i složeniji tipovi podataka (npr. riječ, *engl. string* i sl.). Zbog potrebe za transformacijom ključa u indeks temeljem hash funkcije, dohvaćanje vrijednosti pohranjene pod odgovarajućih ključem iz rječnika ima nešto nižu računalnu učinkovitost u odnosu na indeksirani niz.

Na opisani se način uređene n-torke (*engl. tuples*), koje predstavljaju stanje Markovljevog lanca, koriste kao ključevi rječnika kojima se dohvaćaju prijelazne vjerojatnosti $\mathbb{P}(Z_{k+1} \mid Z_k)$ za određeno stanje Z_k. Na primjer, u slučaju $\mathbf{\Pi}_{6D}$ dan izrazom (3-12) svaka n-torka definirana je indeksima (q, r, w) diskretnih vrijednosti brzine vozila v_q, ubrzanja vozila a_r, te nagiba ceste θ_w. Značajno povećanje računalnih performansi postiže se izostavljanjem pohranjivanja prijelaznih vjerojatnosti jednakih nuli u TPM (tj. prijelaza koji nisu prisutni u snimljenim podacima). Konkretnije, u slučaju rječnika memorijski zahtjevi određeni su ukupnim brojem pohranjenih parova ključ-vrijednost, dok inicijalizacija niza zahtjeva alokaciju memorije za čitavi prostor stanja unaprijed (koji je znatno veći od broja prijelaza između pojedinih stanja). Ilustracija Markovljevog lanca s tri varijable stanja, zajedno s pripadnom $\mathbf{\Pi}_{6D}$ implementiranom u obliku rijetke matrice zasnovane na DOK, prikazana je na slici 3-5.



Slika 3-5: Ilustracija a) Markovljevog lanca i b) pripadne 6D TPM implementirane u obliku rijetke matrice zasnovane na rječniku ključeva (DOK).

3.2.3. Usporedna analiza računalnih performansi

Performanse TPM-ova zasnovanih na N-dimenzionalnom nizu ispituju se za različite dimenzije TPM-a prema slikama 3-2 i 3-3, te se uspoređuju s onima dobivenima za TPM-ove zasnovane na rječniku ključeva (DOK). Testiranje se provodi u smislu zahtjeva za memorijom TPM-a i prosječnog vremena potrebnog za generiranje jednog voznog ciklusa duljine 5.1 km koristeći dotični TPM (za detalje generiranja TPM, vidi odjeljak 3.1.2). Uprosječivanje vremena generiranja voznog ciklusa se provodi na 100 uzoraka. Izračuni se izvode na radnoj stanici HP Z440, koja ima 16 GB RAM-a i Intel® Xeon® procesor E5-1620 v3 @ 3.50 GHz. U analizi se razmatraju dva skupa Markovljevih stanja dobivenih za slučajeve:

i. grubih razlučivosti brzine, ubrzanja, nagiba ceste i derivacije nagiba ceste postavljenih na $\Delta v = 2$ km/h, Δa =0.5 m/s^2, $\Delta\theta = 1°$, i $\Delta\dot{\theta} = 0.25°$, te
ii. finih razlučivosti brzine, ubrzanja, nagiba ceste i derivacije nagiba ceste postavljenih na $\Delta v = 0.5$ km/h, Δa =0.25 m/s^2, $\Delta\theta = 0.5°$, i $\Delta\dot{\theta} = 0.15°$.

Razlog tome je što realizacija 6D, 7D i 8D TPM-a u obliku nizova nije moguća za Markovljeva stanja fine razlučivosti zbog prekomjernih zahtjeva za memorijom (> 16 GB). Rezultati ispitivanja dani su u tablicama 3-1 i 3-2, gdje vrijednosti u zagradama predstavljaju rezultate dobivene za Markovljeva stanja fine razlučivosti, dok se vrijednosti izvan zagrada odnose na slučaj grube razlučivosti.

Rezultati predstavljeni u tablici 3-1 pokazuju da u slučaju implementacije TPM-a u DOK obliku ne postoje restriktivna ograničenja za dimenzionalnost TPM-a, budući da je zahtjev za memorijom drastično smanjen (<60 MB), u usporedbi s konvencionalnim pristupom gdje određeni slučajevi TPM-a nisu izvedivi za Markovljeva stanja fine razlučivosti (6D, 7D i 8D TPM). Povrh toga, TPM-ovi implementirani u DOK obliku također imaju puno bolju računalnu učinkovitost u smislu vremena generiranja sintetičkih voznih ciklusa u usporedbi s konvencionalnim pristupom, posebno za više dimenzije TPM-a (tablica 3-2). Nešto sporije vrijeme izvođenja može se uočiti za 3D, 5D i 7D TPM-ove kod DOK pristupa. Razlog tome je što ova tri slučaja uključuju dodatno računanje sljedećeg stanja brzine u svakom diskretnom vremenskom trenutku (vidi izraz (3-9)).

Tablica 3-1: Memorijski zahtjevi za različite dimenzije i pristupe implementacije TPM-a te grube i fine razlučivosti stanja (za potonje su rezultati dani u zagradama).

Model	Memorijski zahtjevi	
	N-dimenzionalni niz	Rječnik ključeva (DOK)
ML2D (2D-TPM; $\mathbf{\Pi_{2D}}$)	0.017 MB (0.257 MB)	0.014 MB (0.146 MB)
ML3D (3D-TPM; $\mathbf{\Pi_{3D}}$)	0.030 MB (0.409 MB)	0.069 MB (0.609 MB)
ML4D (4D-TPM; $\mathbf{\Pi_{4D}}$)	1.34 MB (73.97 MB)	0.21 MB (2.07 MB)
ML5D (5D-TPM; $\mathbf{\Pi_{5D}}$)	4.92 MB (255.41 MB)	1.42 MB (13.80 MB)
ML6D (6D-TPM; $\mathbf{\Pi_{6D}}$)	226.3 MB (47.3 GB)	3.29 MB (27.11 MB)
ML7D (7D-TPM; $\mathbf{\Pi_{7D}}$)	241.0 MB (31.6 GB)	5.10 MB (37.56 MB)
ML8D (8D-TPM; $\mathbf{\Pi_{8D}}$)	11.09 GB (5.72 TB)	9.91 MB (57.91 MB)
ML-HIER (4D-TPM + 5D-TPM; $\mathbf{\Pi_{N4D}}$ i $\mathbf{\Pi_{Z5D}}$)	17.4 MB (1.85 GB)	1.55 MB (15.33 MB)

Tablica 3-2: Prosječna vremena potrebna za generiranje jednog sintetičkog voznog ciklusa za različite dimenzije i pristupe implementacije TPM-a te grube i fine razlučivosti stanja (za potonje su rezultati dani u zagradama).

Model	Prosječno vrijeme generiranja voznog ciklusa	
	N-dimenzionalni niz	Rječnik ključeva (DOK)
ML2D (2D-TPM; $\mathbf{\Pi_{2D}}$)	12.4 ms (22.9 ms)	3.2 ms (5.1 ms)
ML3D (3D-TPM; $\mathbf{\Pi_{3D}}$)	13.3 ms (16.6 ms)	7.3 ms (9.5 ms)
ML4D (4D-TPM; $\mathbf{\Pi_{4D}}$)	55.2 ms (298.1 ms)	3.4 ms (4.6 ms)
ML5D (5D-TPM; $\mathbf{\Pi_{5D}}$)	33.8 ms (97.1 ms)	8.1 ms (10.7 ms)
ML6D (6D-TPM; $\mathbf{\Pi_{6D}}$)	538.2 ms (nepoznato)	4.4 ms (6.0 ms)
ML7D (7D-TPM; $\mathbf{\Pi_{7D}}$)	170.4 ms (nepoznato)	8.8 ms (11.5 ms)
ML8D (8D-TPM; $\mathbf{\Pi_{8D}}$)	3.98 s (nepoznato)	5.1 ms (5.5 ms)
ML-HIER (4D-TPM + 5D-TPM; $\mathbf{\Pi_{N4D}}$ i $\mathbf{\Pi_{Z5D}}$)	77.1 ms (357.7 ms)	5.5 ms (6.41 ms)

Nadalje, dodatno se provodi analiza broja potrebnih ponavljanja postupka sinteze prilikom generiranja skupa od 100 voznih ciklusa. Drugim riječima, analizira se broj neuspjelih pokušaja generiranja voznih ciklusa, do kojih dolazi zbog prijelaza u određena stanja Markovljevog lanca koja ne postoje u snimljenim podacima temeljem kojih je definiran TPM. Takva pojava zaglavljivanja u neviđenim stanjima ima za posljedicu prekidanje postupka generiranja voznog ciklusa te ponavljanja istog sve dok se uspješno ne provede, čime se povećava ukupno vrijeme izvođenja. Rezultati analize, dani u tablici 3-3, pokazuju kako se ovaj problem pojavljuje samo kod modela koji podrazumijevaju neizravno izračunavanje sljedećeg stanja brzine vozila (3D,

5D i 7D TPM) ili zavisnu sintezu sljedećih stanja brzine i akceleracije vozila (hijerarhijska sinteza; *ML-HIER*). U ovim je slučajevima vidljiv trend porasta broja neuspjelih generiranja voznih ciklusa s povećanjem dimenzije TPM-a te prelaskom na finije razlučivosti stanja. Zbog ovog nedostatka tj. njegove implikacije na produljeno vrijeme generiranja sintetičkih ciklusa, navedeni modeli (*ML3D*, *ML5D*, *ML7D* i *ML-HIER*) nisu pogodni za izravne primjene te stoga nisu nadalje razmatrani u ovom radu.

Tablica 3-3: Broj neuspjelih pokušaja generiranja voznih ciklusa za različite dimenzije i pristupe implementacije TPM-a te grube i fine razlučivosti stanja (za potonje su rezultati dani u zagradama).

Model	Broj neuspjelih pokušaja generiranja voznih ciklusa	
	N-dimenzionalni niz	Rječnik ključeva (DOK)
ML2D (2D-TPM; $\mathbf{\Pi_{2D}}$)	0 (0)	0 (0)
ML3D (3D-TPM; $\mathbf{\Pi_{3D}}$)	0 (37)	0 (29)
ML4D (4D-TPM; $\mathbf{\Pi_{4D}}$)	0 (0)	0 (0)
ML5D (5D-TPM; $\mathbf{\Pi_{5D}}$)	21 (220)	11 (232)
ML6D (6D-TPM; $\mathbf{\Pi_{6D}}$)	0 (nepoznato)	0 (0)
ML7D (7D-TPM; $\mathbf{\Pi_{7D}}$)	46 (nepoznato)	63 (8578)
ML8D (8D-TPM; $\mathbf{\Pi_{8D}}$)	0 (nepoznato)	0 (0)
ML-HIER (4D-TPM + 5D-TPM; $\mathbf{\Pi_{N4D}}$ i $\mathbf{\Pi_{Z5D}}$)	26 (377)	28 (364)

3.2.4. Usvajanje i tvorba konačnog koncepta

Prilikom usvajanja konačnog koncepta realizacije matrice prijelaznih vjerojatnosti $\mathbf{\Pi}$ u svrhu sinteze voznih ciklusa potrebno je imati na umu sljedeće prednosti te nedostatke razmatranih modela:

1) Modeli koji podrazumijevaju neizravno izračunavanje sljedećeg stanja brzine vozila ($\mathbf{\Pi}_{3D}$, $\mathbf{\Pi}_{5D}$ i $\mathbf{\Pi}_{7D}$) pokazali su izraženu sklonost zaglavljivanja u stanjima neviđenima u snimljenim podacima kojima je definiran TPM, što ih čini nepraktičnima za daljnje primjene (odjeljak 3.2.3). Istu sklonost pokazao je i model hijerarhijske sinteze voznih ciklusa *ML-HIER* temeljen na $\mathbf{\Pi}_{N4D}$ i $\mathbf{\Pi}_{Z5D}$.
2) U slučaju implementacije TPM u DOK obliku, i najsloženiji model *ML8D* ima niski zahtjev za memorijom (58 MB; tablica 3-1) i razumno vrijeme izvođenja (6 ms; tablica 3-2), čak i uz finu rezoluciju stanja.

Stoga se kao finalni model za sintezu voznih ciklusa u ovom radu usvaja ultimativni model *ML8D* izveden u obliku $\mathbf{\Pi}_{8D}$ (slika 3-2g), a koji koristi kombinaciju diskretnih vrijednosti

brzine vozila v, ubrzanja vozila $a = \dot{v}$, nagiba ceste θ i vremenske derivacije nagiba ceste $\dot{\theta}$ kao stanja Markovljevog lanca. Pripadni rasponi i razlučivosti varijabli stanja (v, a, θ i $\dot{\theta}$) postavljeni su kako slijedi:

- Raspon brzine vozila $\boldsymbol{v}_s = [v_{\min}, v_{\min} + \Delta v, \dots, v_{\max}]$ postavljen je na interval od 0 km/h do 90 km/h, pri čemu $v_{\max}$ = 90 km/h odgovara maksimalnoj propisanoj brzini razmatranih autobusa, a maksimalna izmjerena brzina iznosila je bliskih 81.2 km/h (slika 2-12a). Razlučivost brzine Δv postavljena je na 0.1 km/h kako bi bila jednaka snimljenoj, što rezultira ukupnim brojem od 901 diskretnih stanja brzina ($\boldsymbol{v}_s \in R^{1 \times 901}$).
- Ubrzanje vozila a rekonstruirano je numeričkim deriviranjem izmjerenih vrijednosti brzine vozila, s vremenom uzorkovanja ΔT od jedne sekunde. Raspon ubrzanja $\boldsymbol{a}_s = [a_{\min}, a_{\min} + \Delta a, \dots, a_{\max}]$ postavljen je na interval od -2 m/s^2 do 2 m/s^2, jer se većina ubrzanja sadržanih u snimljenim voznim ciklusima nalazi unutar tog raspona (vidi sliku 2-12b). Razlučivost ubrzanja Δa postavljena je na 0.15 m/s^2, što rezultira ukupnim brojem od 28 diskretnih stanja ubrzanja ($\boldsymbol{a}_s \in R^{1 \times 28}$).
- Nagib ceste θ je diskretiziran s razlučivošću $\Delta\theta$ od 0.1° u punom rasponu $\boldsymbol{\theta}_s = [\theta_{\min}, \theta_{\min} + \Delta\theta, \dots, \theta_{\max}]$ od $-6°$ do 6° (vidi sliku 2-12c), što rezultira ukupnim brojem od 121 diskretnih stanja nagiba ceste ($\boldsymbol{\theta}_s \in R^{1 \times 121}$).
- Raspon vremenske derivacije nagiba ceste $\dot{\boldsymbol{\theta}}_s = [\dot{\theta}_{\min}, \dot{\theta}_{\min} + \Delta\dot{\theta}, \dots, \dot{\theta}_{\max}]$ postavljen je na interval od -0.75 °/s do 0.75 °/s s razlučivošću $\Delta\dot{\theta}$ od 0.25 °/s, što rezultira ukupnim brojem od 7 diskretnih stanja derivacije nagiba ceste ($\dot{\boldsymbol{\theta}}_s \in R^{1 \times 7}$).

Tvorba $\boldsymbol{\Pi}_{8D}$ realizirana je u obliku rječnika ključeva (DOK). Nešto finije razlučivosti stanja od onih korištenih u prethodnoj analizi (odjeljak 3.2.3) uvedene su s ciljem poboljšanja točnosti sinteze voznih ciklusa, zadržavajući pritom razuman memorijski zahtjev $\boldsymbol{\Pi}_{8D}$ od 213.8 MB te čak kraće prosječno vrijeme generiranja voznog ciklusa od 4 ms (zbog manjeg broja zabilježenih prijelaznih vjerojatnosti za svako stanje). Punjenje $\boldsymbol{\Pi}_{8D}$ provodi se sekvencijalnim procesiranjem svih $N_{DC,F}^{V} = 1527$ valjanih snimljenih voznih ciklusa s finom razlučivošću potrošnje goriva, odnosno svakog uzorka opisanog vrijednostima brzine vozila $v(k)$, ubrzanja vozila $a(k)$, nagiba ceste $\theta(k)$ i pripadne derivacije $\dot{\theta}(k)$ zabilježenih u k-tom diskretnom vremenskom trenutku za određeni autobus (vidi poglavlje 2). Nužan uvjet je da svaki vozni ciklus ima pravilno raspoređene uzorke u vremenu. Sekvencijalno procesiranje podrazumijeva brojanje prijelaza između diskretnih Markovljevih stanja u susjednim diskretnim vremenskim koracima, $Z_k = (v_k, a_k, \theta_k, \dot{\theta}_k)$ i $Z_{k+1} = (v_{k+1}, a_{k+1}, \theta_{k+1}, \dot{\theta}_{k+1})$, te njihovo pohranjivanje u

odgovarajuće ćelije $\mathbf{\Pi}_{8D}$. Tijekom ovog postupka stvarne tj. snimljene vrijednosti brzine, ubrzanja, nagiba i derivacije nagiba zaokružuju se na najbliže diskretne vrijednosti Markovljevog stanja. Završna faza uključuje skaliranje $\mathbf{\Pi}_{8D}$, tako da zbroj prijelaznih vjerojatnosti iz svakog Markovljevog stanja u isto ili bilo koje drugo stanje bude jednako 1 (vidi izraz (3-4)). Ukupni broj stanja Markovljevog lanca iznosi 21,368.116, dok broj zabilježenih prijelaza između stanja tj. parova ključ-vrijednost sadržanih u $\mathbf{\Pi}_{8D}$ iznosi 528.634 (što odgovara rijetkosti matrice od 97.5%).

Time je dobivena TPM koja predstavlja stohastički model voznih ciklusa, koji se dalje u radu koristi za sintezu 3D voznih ciklusa koji uključuju međusobno korelirane vremenske profile brzine i ubrzanja vozila, te nagiba ceste.

3.3. Generiranje sintetičkih voznih ciklusa uzimajući u obzir promjenjivu masu putnika i stupanj zagušenja prometa

U ovom potpoglavlju najprije se razmatra uključenje promjenjive mase putnika u cjelokupni postupak sinteze voznih ciklusa, uključujući usklađivanje sintetičkih ciklusa mase putnika s ciklusima brzine vozila. Budući da se masa putnika ne mijenja kontinuirano kroz vrijeme već tek prilikom zaustavljanja autobusa na stanici i tek je u manjoj mjeri korelirana s brzinom i akceleracijom vozila (prvenstveno iz razloga što je prikupljen vrlo ograničen skup zabilježenih podataka; poglavlje 2), predlaže se nezavisan pristup sinteze profila mase putnika temeljen na događajima (odjeljak 3.3.1). U odjeljku 3.3.2 provodi se usporedna analiza sintetiziranih ciklusa mase putnika u odnosu na zabilježene. U konačnici se demonstriraju rezultati primjene predložene metode sinteze voznih ciklusa za slučaj s i bez uzimanja u obzir informacije o stupnju zagušenja prometa (odjeljak 3.3.3).

3.3.1. Koncept nezavisne sinteze profila mase putnika

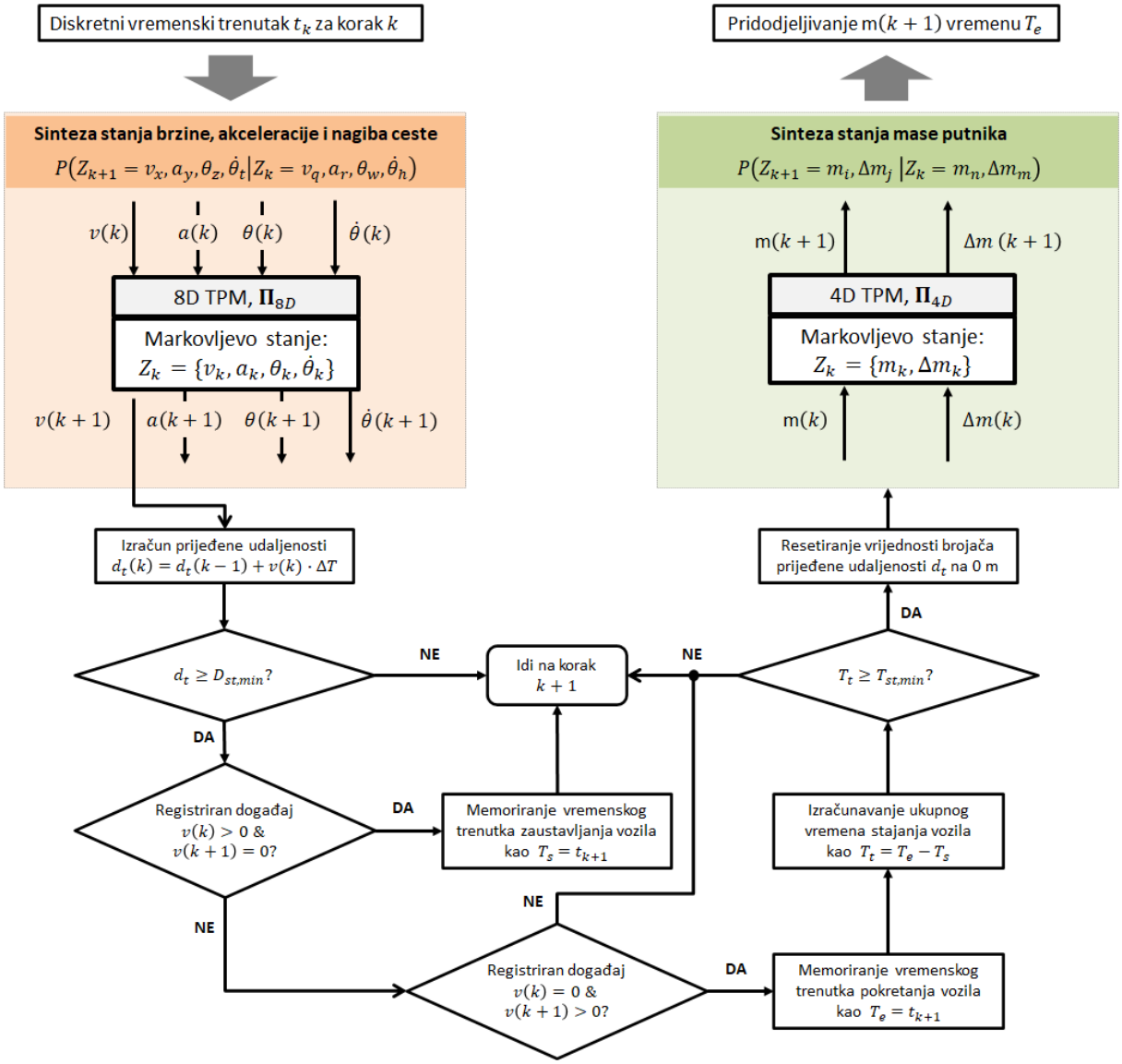
Sintezu profila mase putnika moguće je provesti na više načina. Na primjer, podatak o masi putnika može se uključiti u matricu prijelaznih vjerojatnosti (TPM), čime se za opći slučaj osigurava spregnutost mase putnika s brzinom i ubrzanjem vozila te nagibom ceste. Međutim, nedostatak ovog pristupa je u porastu dimenzionalnosti TPM s 8D na 10D, ili čak 12D ukoliko se pri definiranju stanja uz masu putnika razmatra i promjena mase putnika između susjednih stanica. Ovaj pristup također zahtijeva veliku količinu snimljenih voznih ciklusa s uključenim podatkom o masi putnika, što ovdje nije slučaj, a prema preliminarnoj analizi u poglavlju 2 nije neophodan zbog male koreliranosti mase putnika s brzinom i akceleracijom vozila.

Stoga, alternativni pristup sinteze podrazumijeva rastavljanje razmatrane rute na segmente ceste između dviju uzastopnih autobusnih stajališta, te potom mapiranje svih snimljenih voznih ciklusa i podataka mase putnika po segmentima. Zatim se provodi sinteza brzine vozila duž svakog segmenta uz forsiranje zaustavljanja vozila na stajalištima (ukoliko je temeljem mapiranih vjerojatnosti zaustavljanja tako određeno) [99]. U konačnici se provodi pridjeljivanje mapiranih vrijednosti mase putnika svakom segmentu. Iako ovaj pristup predstavlja egzaktno rješenje, njegov glavni nedostatak je zahtjev za poznavanjem svih podataka o ruti (npr. lokacije autobusnih stajališta) i prostorno-vremenskih mapa srednjih brzina, vjerojatnosti zaustavljanja vozila i vremena stajanja na stajalištu [99].

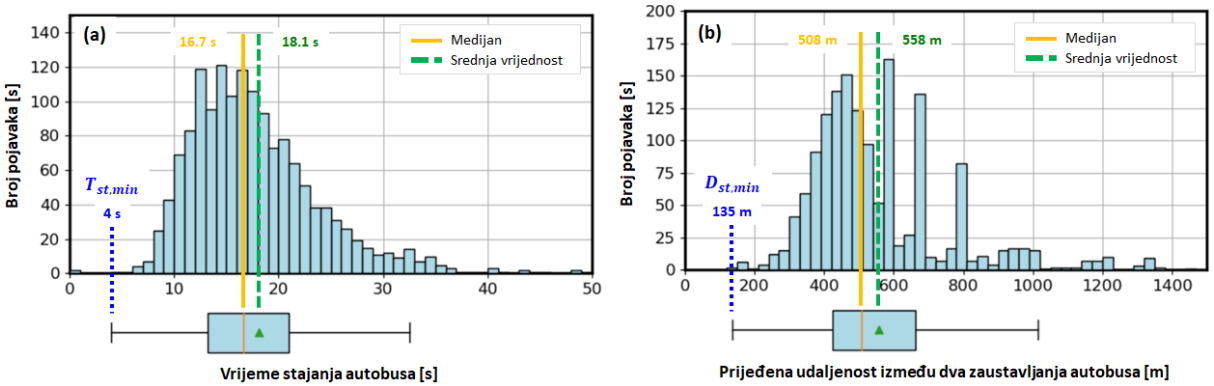
Vodeći se gore navedenim nedostatcima razmatranih pristupa sintezi vremenskog profila mase putnika, u ovom radu predlaže se postupak nezavisne sinteze profila mase putnika temeljen na događajima zaustavljanja autobusa, koji pretpostavlja raspregnutost mase putnika s ostalim stanjima Markovljevog lanca. U nedostatku informacija o stajalištima a bez gubitka općenitosti, sintetički ciklusi mase putnika usklađeni su s ciklusima brzine vozila, na način da promjena broja putnika može nastupiti tek nakon što se vozilo zaustavilo, odnosno kada se brzina vozila spustila na nulu. Pritom se zbog nemogućnosti određivanja razloga zaustavljanja (pored stajališta to može biti zbog npr. zagušenja prometa, semafora i sl.) postavlja uvjet da promjena mase putnika može nastupiti tek nakon što je vozilo prevalilo određeni put $D_{st,min}$, uz vrijeme stajanja veće od $T_{st,min}$. Vrijednosti $D_{st,min}$ i $T_{st,min}$ određuju se temeljem statističke analize zaustavljanja vozila provedene na snimljenim voznim ciklusima kako je opisano u nastavku.

Dijagram toka cjelokupnog postupka nezavisne sinteze profila mase putnika prikazan je na slici 3-6. Ideja nezavisne sinteze profila mase putnika je da se prilikom generiranja profila brzine vozila na temelju 8D-TPM tj. $\mathbf{\Pi}_{8D}$ registrira svaki događaj zaustavljanja autobusa (brzina vozila $v = 0$ km/h) nakon što je prijeđena određena udaljenost $d_t \geq D_{st,min}$; te ako je stajanje autobusa trajalo duže od određenog vremena, $T_t \geq T_{st,min}$, sintetizira se sljedeće stanje mase putnika $m(k+1)$ primjenom 4D-TPM ($\mathbf{\Pi}_{M4D}$). Pritom $D_{st,min}$ označava minimalnu zahtjevanu prijeđenu udaljenost između dva zaustavljanja vozila, a $T_{st,min}$ minimalno zahtjevano vrijeme stajanja autobusa. Vrijednosti ovih parametara određene su statističkom obradom snimljenih podataka o prijeđenim udaljenostima između dva zaustavljanja autobusa i pripadnim vremenima stajanja autobusa, koji su na slici 3-7 prikazani u obliku histograma zajedno s pripadnim kutijastim dijagramima (*engl. box plot*). Za potrebe sinteze profila mase putnika usvojeni su pragovi $D_{st,min} = 135$ m i $T_{st,min} = 4$ s, koji odgovaraju donjim granicama pripadnih interkvartilnih raspona odgovarajućih varijabli. Povećanjem ovih pragova smanjuje

se učestalost promjena mase putnika kod sintetičkih ciklusa, dok u suprotnom slučaju vrijedi obratno.



Slika 3-6: Dijagram toka koncepta nezavisne sinteze profila mase putnika temeljene na događaju zaustavljanja vozila.



Slika 3-7: Histogrami i pripadni kutijasti dijagrami snimljenih a) vremena stajanja autobusa, te b) prijeđene udaljenosti između dva zaustavljanja autobusa, uključujući postavljene vrijednosti pragova $D_{st,min}$ i $T_{st,min}$.

Tvorba $\mathbf{\Pi}_{M4D}$ podrazumijeva isti postupak kao i u slučaju $\mathbf{\Pi}_{4D}$ (opisan u potpoglavlju 3.2), ali uz stanja Markovljevog lanca definirana kao mase putnika m i njene promjene između dviju uzastopnih autobusnih stajališta Δm. Time su razdiobe prijelaznih vjerojatnosti predstavljene kroz 4D TPM kao

$$p_{nm,ij} \coloneqq P\left(Z_{k+1} = m_i, \Delta m_j \middle| Z_k = m_n, \Delta m_m\right); \quad \mathbf{\Pi}_{M4D} \coloneqq \left[p_{nm,ih}\right], \tag{3-17}$$

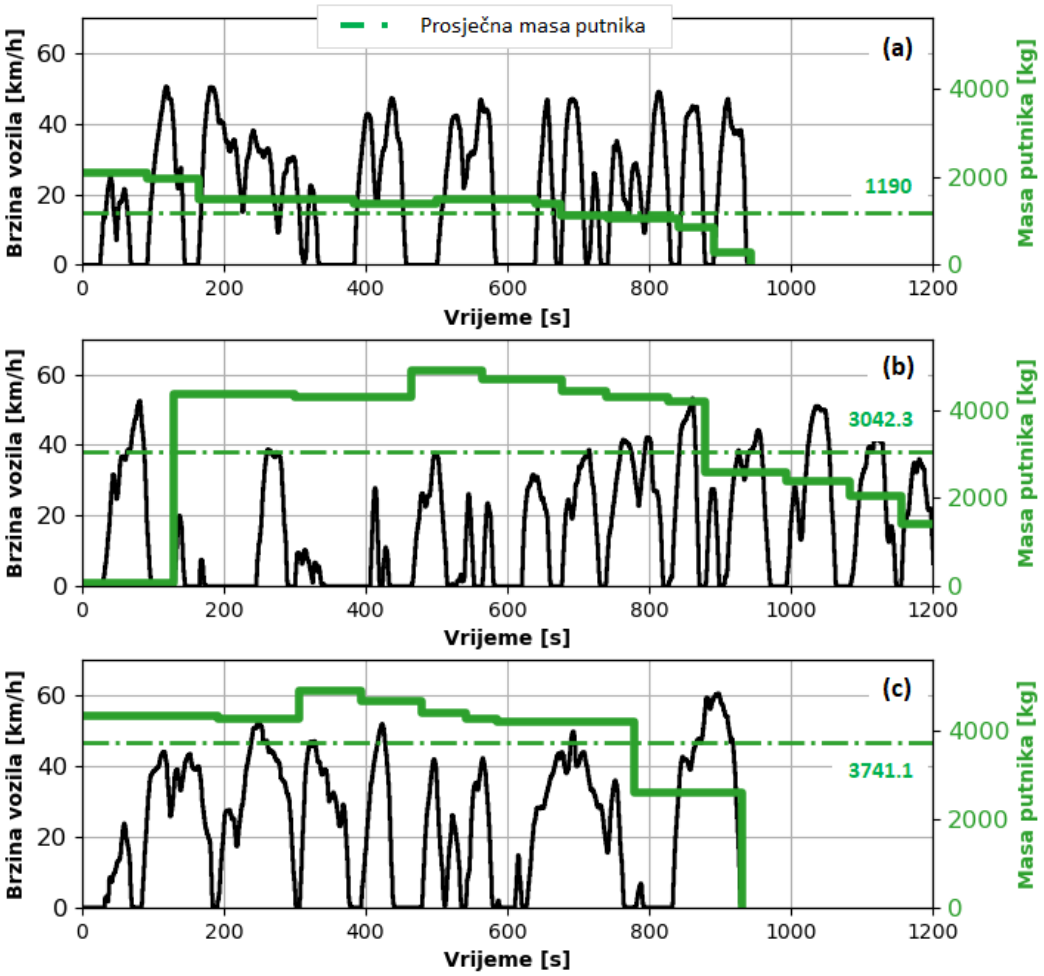
gdje element matrice $p_{nm,ij}$ označava vjerojatnost prijelaza iz trenutnog stanja mase putnika $(m_n, \Delta m_m)$ u sljedeće stanje $(m_i, \Delta m_j)$.

Za potrebe sinteze profila mase putnika, odnosno tvorbe $\mathbf{\Pi}_{M4D}$ pretpostavlja se prosječna masa jednog putnika (PMP) od 70 kg, koja se množi sa snimljenim brojem putnika kako bi se u konačnici dobio profil mase putnika. Raspon mase putnika $\boldsymbol{m}_s = [m_{\min}, m_{\min} + \Delta m, \ldots, m_{\max}]$ postavljen je na interval od 0 do 5600 kg, što odgovara umnošku prosječne mase putnika (PMP) i maksimalnog deklariranog broja putnika za dani tip autobusa (MBP jednak 80). Razlučivost mase putnika postavljena je na PMP, što rezultira ukupnim brojem od 81 diskretnih stanja mase putnika ($\boldsymbol{m}_s \in R^{1\,\mathrm{x}\,81}$). Budući da na svakoj stanici mogu ući ili izaći svi putnici (maksimalne zabilježene vrijednosti su -72 i +61; vidi sliku 2-17c), raspon promjene mase putnika Δm postavljen je na interval od -5600 kg do 5600 kg, dok je razlučivost postavljena na $10 \cdot PMP$, što rezultira ukupnim brojem od 17 diskretnih stanja brzina ($\Delta\boldsymbol{m}_s \in R^{1\,\mathrm{x}\,17}$). Prema tome, vrijedi da je $\mathbf{\Pi}_{M4D} \in R^{81\,\mathrm{x}\,17\,\mathrm{x}\,81\,\mathrm{x}\,17}$, odnosno da ukupni broj stanja Markovljevog lanca koji se koristi za sintezu profila mase putnika iznosi 1377.

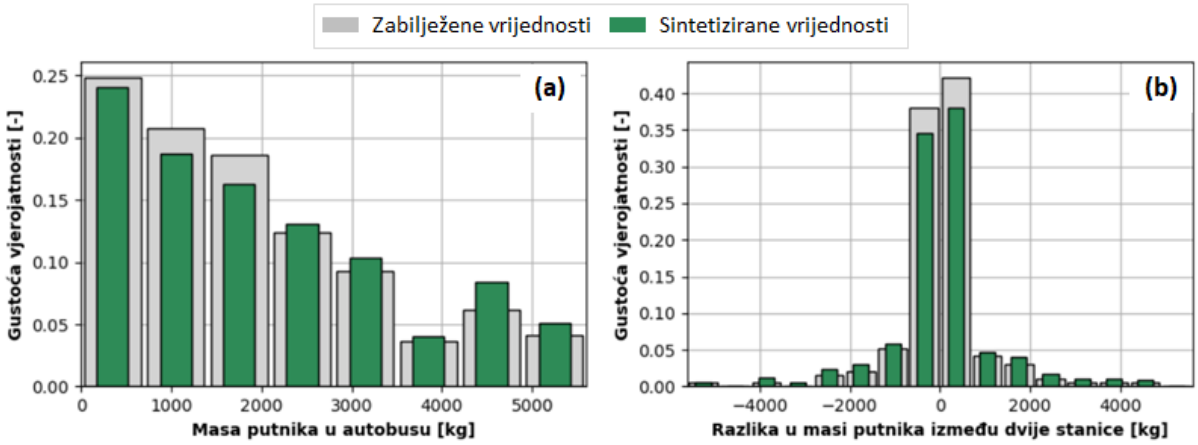
3.3.2. Rezultati generiranja sintetičkih profila mase putnika

Primjenom metode nezavisne sinteze voznih ciklusa i profila mase putnika temeljene redom na sljedećim TPM-ovima $\mathbf{\Pi}_{8D}$ i $\mathbf{\Pi}_{M4D}$ (slika 3-6) generirano je ukupno $N_s^M = 1000$ sintetičkih voznih ciklusa s uključenim profilom mase putnika, čija prijeđena udaljenost iznosi ≈5.1 km (što odgovara prosječnoj prijeđenoj udaljenosti snimljenih voznih ciklusa; vidi odjeljak 2.3.1 i sliku 2-10a). Pritom se podrazumijeva da je za generiranje sintetičkih voznih ciklusa korišten TPM $\mathbf{\Pi}_{8D}$ definiran čitavim skupom snimljenih voznih ciklusa Ψ_z (neovisno o tome kojem klasteru zagušenosti prometa pripadaju). Početna stanja mase putnika m_0 i Δm_0 određena su nasumičnim odabirom zabilježenih stanja sadržanih u $\mathbf{\Pi}_{M4D}$ (kojih je ukupno 107; rijetkost od 92.2%). Razlog tome je što masa putnika na početku voznog ciklusa, tj. početnoj autobusnoj stanici može poprimiti bilo koju vrijednost između 0 i $m_{\max}$, za razliku od brzine v_0 i akceleracije a_0 vozila čije početne vrijednosti uvijek iznose 0 km/h i 0 m/s^2, budući da autobus kreće s vožnjom iz stanja mirovanja. Na kraju sinteze svakog voznog ciklusa masa putnika se postavlja na nulu, uz realnu pretpostavku da su svi putnici izašli iz autobusa. Na slici 3-8 prikazani su izdvojeni primjeri sintetičkih profila mase putnika s različitim razinama popunjenosti autobusa putnicima (niska, srednja i visoka). Usporedno su prikazani pripadni profili brzine vozila koji pokazuju da promjena mase putnika nastupa tek prilikom zaustavljanja vozila. Razlog relativno niskih varijacija izdvojenih sintetičkih profila mase putnika oko pripadnih prosječnih vrijednosti proizlazi iz sličnih izvornih karakteristika zabilježenih profila mase putnika (slika 2-17). Opsežnija usporedba cjelokupnog skupa generiranih sintetičkih profila mase putnika sa zabilježenim skupom dana je u nastavku ovog odjeljka.

Kako bi se pokazala valjanost predložene metode sinteze profila mase putnika, provodi se usporedba histograma sintetiziranih vrijednosti mase putnika i promjena mase putnika između dviju susjednih stanica u odnosu na zabilježene vrijednosti. Pritom se $\bar{M}_s$ odnosi na skup svih individualnih sintetičkih profila mase putnika $\bar{M}_s = \{m_{s,1}, m_{s,2}, \dots, m_{s,i}\}; i = 1,2,3, \dots 1000$, spojenih u jednu cjelinu, a $\bar{M}_z$ na zabilježeni profil mase putnika (slika 2-17a). Iz rezultata prikazanih na slici 3-9 vidljivo je kako histogrami sintetičkih masa putnika $\bar{M}_s$ u velikoj mjeri nalikuju onima dobivenima za zabilježene vrijednosti $\bar{M}_z$. Međutim, postoje određena odstupanja koja se odražavaju kroz blago smanjeni broj masa putnika u području do < 2000 kg, uz blago povećani broj onih iznad te granice. Razlog ove pojave očito je u nasumičnom odabiru početnih vrijednosti mase putnika m_0, koje nisu u tolikoj mjeri učestale u zabilježenim podacima, te relativno malom broju od 193 zabilježenih uzoraka $\bar{M}_s \in \mathbb{R}^{1\,\mathrm{x}\,193}$.



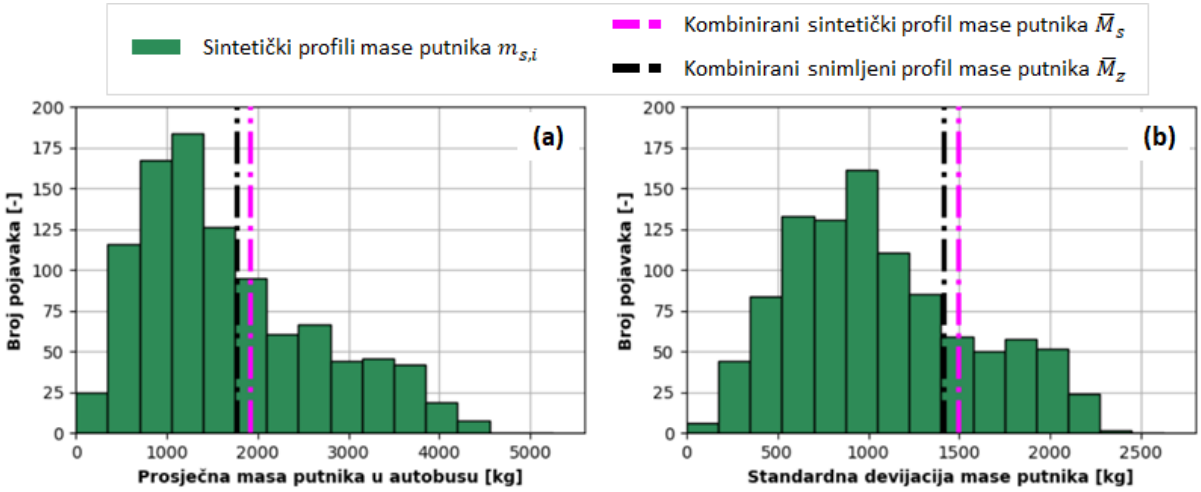
Slika 3-8: Primjeri sintetičkih profila brzine vozila i mase putnika za slučajeve a) niske, b) srednje i c) visoke popunjenosti autobusa putnicima.



Slika 3-9: Histogrami zabilježenih i sintetiziranih a) vrijednosti mase putnika i b) razlika u masi putnika između dvaju zaustavljanja vozila.

Također je provedena usporedna analiza skupova $\bar{M}_s$ i $\bar{M}_z$ u smislu prosječnih vrijednosti i standardne devijacije mase putnika. Slika 3-10 prikazuje histograme ovih metrika izračunatih

za svaki individualni sintetički vozni ciklus $m_{s,i}$ zasebno, te također one dobivene za čitave skupove $\bar{M}_z$ i $\bar{M}_s$. Prosječna masa putnika za zabilježene podatke $\bar{M}_z$ iznosi 1772 kg, dok za sintetičke $\bar{M}_s$ iznosi 1916 kg, što rezultira relativnim odstupanjem od +8.1% (slika 3-10a). Standardna devijacija mase putnika također je veća za $\bar{M}_s$ u odnosu na $\bar{M}_z$ (1495 kg naspram 1415 kg; +5.7%; slika 3-10b). Rasipanja prosječne mase putnika individualnih sintetičkih profila mase putnika $m_{s,i}$ kreću se u rasponu od 150 kg do 4500 kg, dok se pripadne vrijednosti standardne devijacije kreću u rasponu od 120 kg do 2400 kg. Međutim, mnoštvo prosječnih vrijednosti mase putnika koncentrira se u intervalu od 500 kg do 2000 kg (slika 3-10a), a standardnih devijacija u intervalu od 400 kg do 1400 kg (slika 3-10b). Iako postoje određena odstupanja sintetiziranih profila mase putnika od onih zabilježenih, ta odstupanja nisu značajna (<10%), pa se stoga predložena metoda sinteze profila mase putnika može smatrati valjanom za daljnju uporabu. Očekuje se da bi točnost sinteze porasla s brojem zabilježenih uzoraka broja putnika u odnosu na relativno uzak skup raspoloživ za ovu studiju. U tom slučaju moglo bi se provesti i klasteriranje mase vozila s obzirom na doba dana, kako bi se osigurala međusobna korelacija (od praznog autobusa noću do punog autobusa u vršnim satima), što bi u konačnici rezultiralo unaprjeđenjem predložene metode sinteze profila mase putnika.



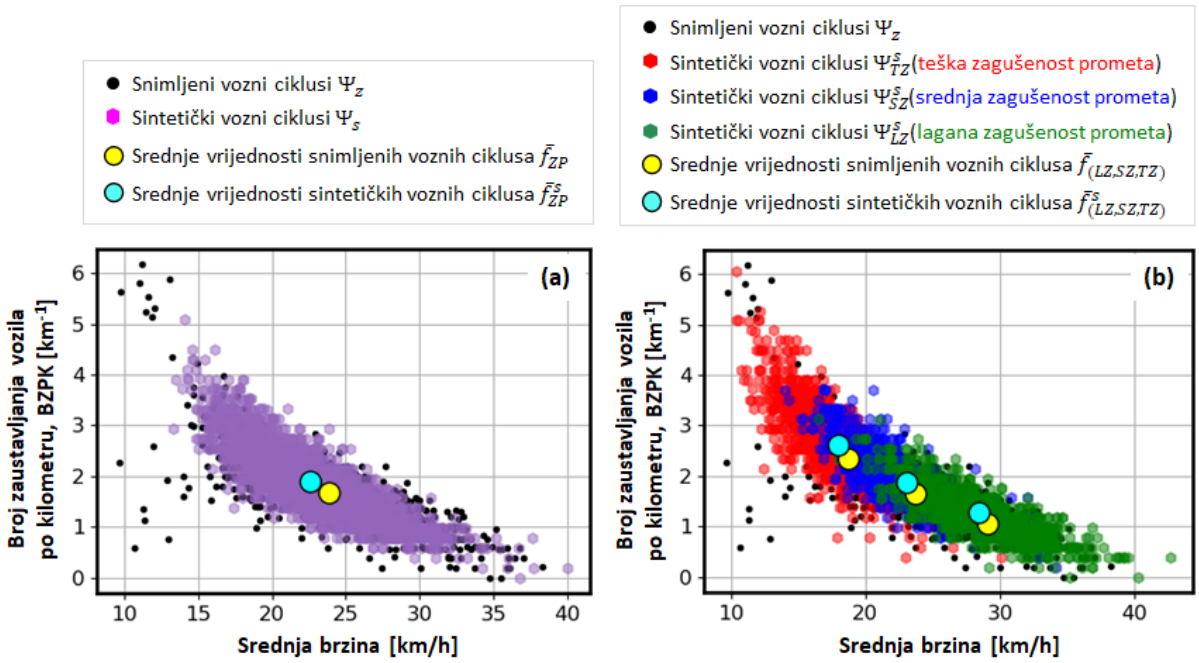
Slika 3-10: Histogrami a) prosječnih vrijednosti i b) standardnih devijacija mase putnika, izračunatih za svaki individualni sintetički vozni ciklus $m_{s,i}$, uključujući vrijednosti izračunate za kombinirani sintetički te snimljeni profil mase putnika.

3.3.3. Rezultati generiranja sintetičkih voznih ciklusa s i bez uzimanja u obzir stupnja zagušenja prometa

Primjenom usvojenog modela voznih ciklusa temeljenog na Markovljevom lancu predstavljenim s 8D-TPM tj. $\mathbf{\Pi}_{8D}$ i postupka sinteze opisanog prethodnim potpoglavljima generirano je ukupno $N_s^U = 6000$ sintetičkih voznih ciklusa duljine ≈5.1 km, od kojih $N_s^R =$

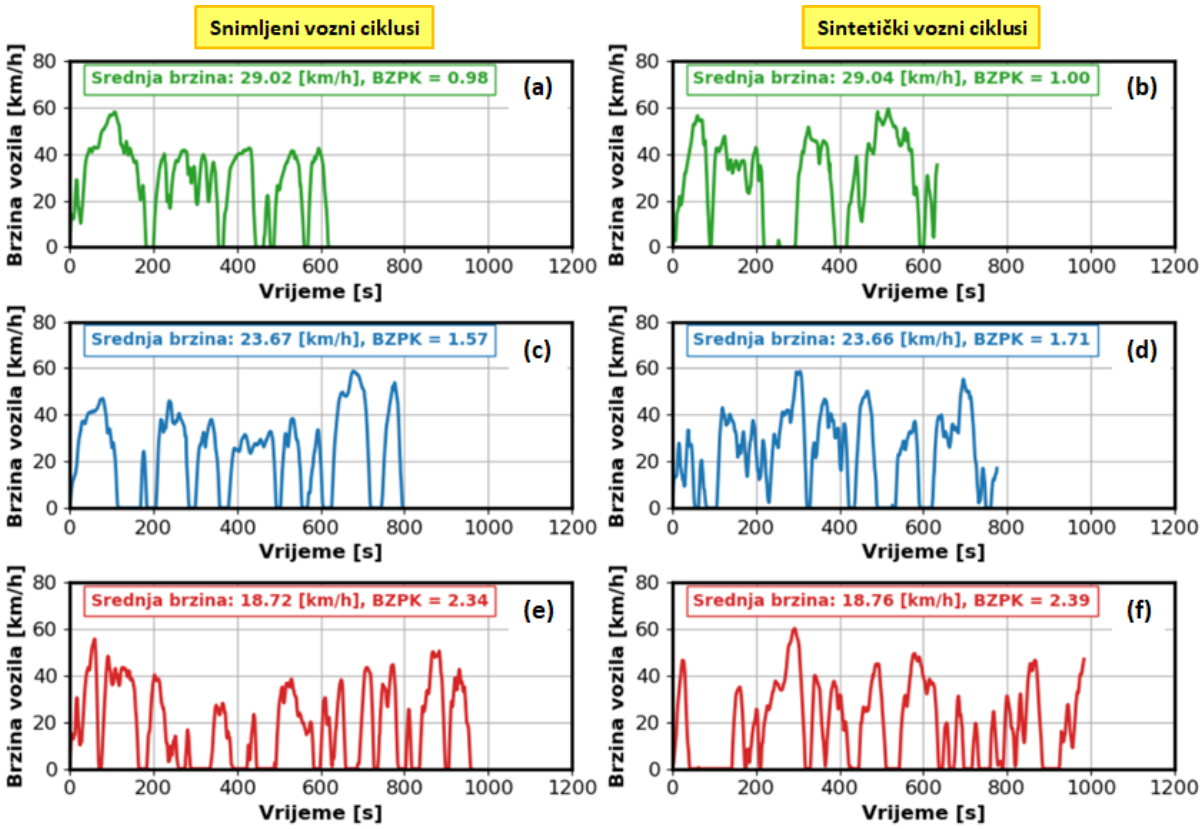
3000 sadržanih u skupu Ψ_R^S ne uključuje kategorizaciju prema stupnju zagušenja prometa (vidi odjeljak 2.2.4). Preostalih $N_S^K = N_S^U - N_S^R = 3000$ sintetičkih voznih ciklusa dalje se dijeli u podskupove Ψ_{LZ}^S za laganu, Ψ_{SZ}^S za srednju te Ψ_{TZ}^S za tešku zagušenost prometa, gdje svaki podskup sadrži 1000 voznih ciklusa. Pritom se podrazumijeva da je za generiranje Ψ_R^S korišten TPM $\mathbf{\Pi}_{8D}^R \equiv \mathbf{\Pi}_{8D}$ definiran čitavim skupom snimljenih voznih ciklusa Ψ_z (neovisno o tome kojem klasteru zagušenosti prometa pripadaju), dok su za generiranje Ψ_{LZ}^S, Ψ_{SZ}^S i Ψ_{TZ}^S korištene zasebne TPM ($\mathbf{\Pi}_{8D}^{LZ}$, $\mathbf{\Pi}_{8D}^{SZ}$ i $\mathbf{\Pi}_{8D}^{TZ}$) definirane na podskupovima snimljenih voznih ciklusa Ψ_{LZ}, Ψ_{SZ} i Ψ_{TZ} određenih pripadnim klasterima zagušenosti prometa prema slici 2-9.

Slika 3-11 prikazuje razdiobe značajki zagušenja prometa za slučaj čitavog skupa snimljenih Ψ_z i generiranih sintetičkih voznih ciklusa (Ψ_R^S na slici 3-11a te Ψ_{LZ}^S, Ψ_{SZ}^S i Ψ_{TZ}^S na slici 3-11b), uključujući pripadne srednje/očekivane vrijednosti $\bar{\boldsymbol{f}}_{ZP}$ i $\bar{\boldsymbol{f}}_{ZP}^S$ (označene žutim i cijan kružićima). Rezultati na slici 3-11 pokazuju kako razdiobe značajki izračunate za sintetičke vozne cikluse Ψ_R^S statistički vjerodostojno rekonstruiraju razdiobe značajki snimljenih voznih ciklusa Ψ_z, što se dodatno odražava kroz bliskost pripadnih srednjih vrijednosti $\bar{\boldsymbol{f}}_{ZP}^S \sim \bar{\boldsymbol{f}}_{ZP}$ (relativna odstupanja značajki iznose ≈5% u prosjeku). U slučaju kategoriziranih sintetičkih voznih ciklusa Ψ_{LZ}^S, Ψ_{SZ}^S i Ψ_{TZ}^S postoje određena preklapanja značajki između klastera zagušenosti prometa, što se može pripisati stohastičkoj prirodi postupka sinteze temeljenog na Markovljevom lancu. Neovisno o tome, može se primijetiti kako se dotične značajke distribuiraju prema očekivanjima (usp. sliku 2-9), npr. sintetički vozni ciklusi s lakim zagušenjem prometa Ψ_{LZ}^S imaju veliku srednju brzinu i mali broj zaustavljanja vozila po kilometru, dok za Ψ_{TZ}^S vrijedi obratno.

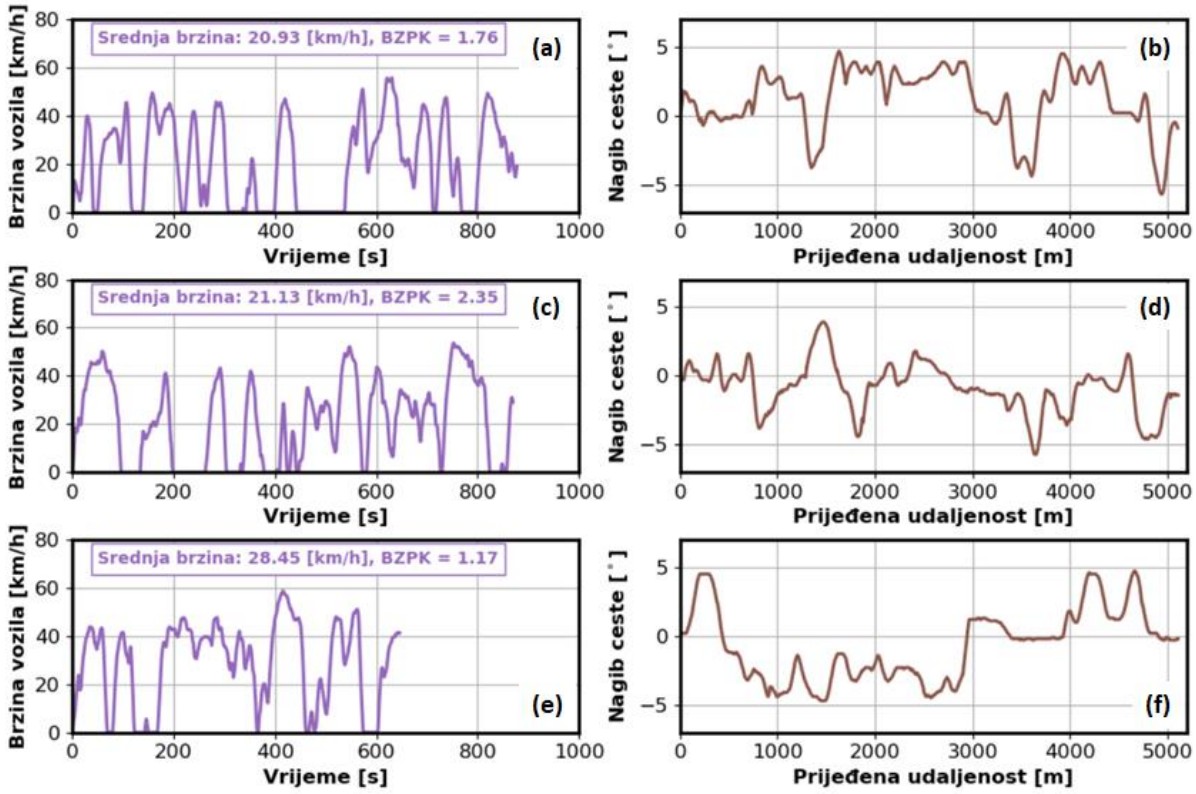


Slika 3-11: Razdiobe značajki zagušenja prometa sintetičkih i snimljenih voznih ciklusa uključujući pripadne srednje vrijednosti značajki, za slučajeve a) sveukupnog skupa podataka i b) uzimajući u obzir informaciju o stupnju zagušenosti prometa.

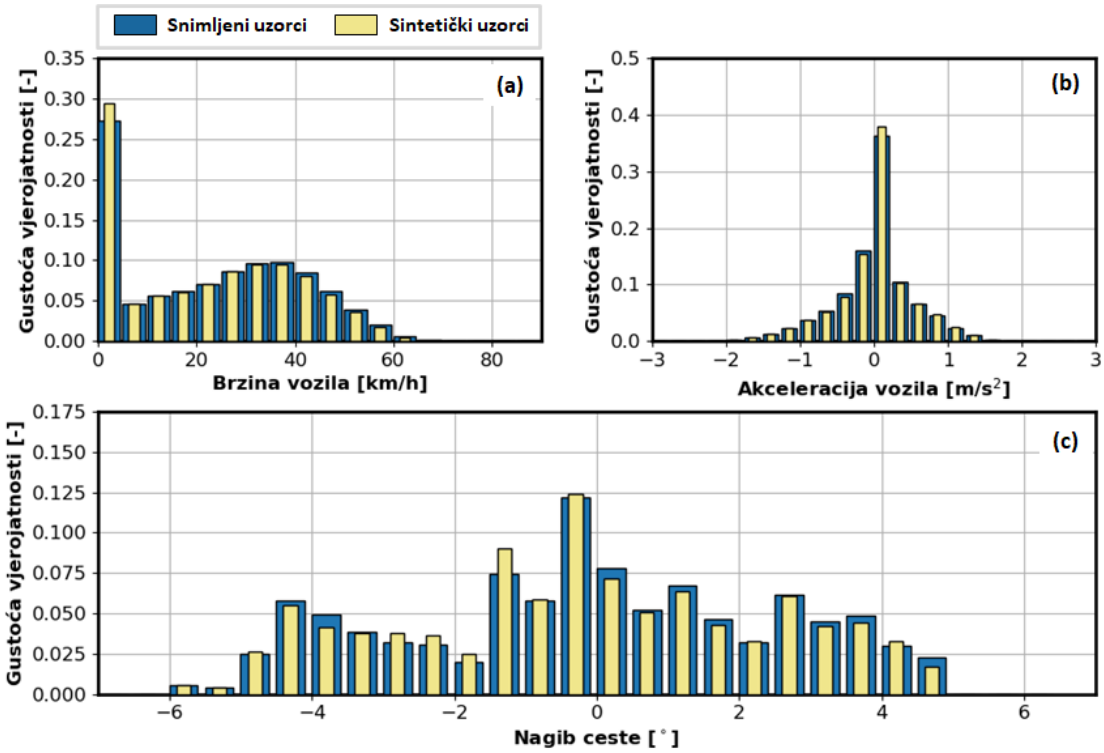
Reprezentativni primjeri snimljenih i sintetičkih voznih ciklusa za svaki klaster zagušenosti prometa (oni najbliži srednjim vrijednostima dviju značajki $\bar{\boldsymbol{f}}_{ZP}$ snimljenih voznih ciklusa) prikazani su na slici 3-12, zajedno s pripadnim vrijednostima srednjih brzina i broja zaustavljanja po vozila kilometru (BZPK). Pritom valja imati na umu kako uvjet na konačne vrijednosti brzine i ubrzanja vozila jednake nuli nije razmatran u okviru ovog rada, no detaljno je obrađen u popratnoj studiji [99]. Vizualnom inspekcijom lijevog i desnog stupca na slici 3-12 može se zaključiti kako izdvojeni reprezentativni sintetički vozni ciklusi u velikoj mjeri nalikuju snimljenima. Nadalje, nasumičnim odabirom izvučena su tri različita primjera generiranih sintetičkih voznih ciklusa iz skupa Ψ_R^S, čiji su profili brzine u vremenu te nagiba ceste u ovisnosti o prijeđenom putu prikazani na slici 3-13. Varijacije između sintetičkih profila nagiba ceste u odnosu na snimljene profile (usp. desni stupac slike 3-13 sa slikom 2-7) proizlaze iz stohastičke prirode uzorkovanja 8D-TPM ($\boldsymbol{\Pi}_{8D}$) definirane na objedinjenim podacima oba smjera vožnje. Kako bi se dokazala valjanost predložene metode sinteze voznih ciklusa, provodi se usporedba histograma svih sintetičkih uzoraka brzine i ubrzanja vozila te nagiba ceste iz skupa Ψ_R^S u odnosu na snimljeni skup Ψ_z. Rezultati su prikazani na slici 3-14 te pokazuju kako se histogrami sintetičkih i snimljenih voznih ciklusa podudaraju u značajnoj mjeri. Potpuna validacija generiranog skupa sintetičkih voznih ciklusa Ψ_R^S po znatno širem skupu značajki obrađuje se u poglavljima 5 i 6.



Slika 3-12: Reprezentativni primjeri vremenskih profila brzine vozila za a), c), e) snimljene i b), d), f) sintetičke vozne cikluse dobivene za klastere lagane (zelena boja), srednje (plava boja) i teške zagušenosti prometa (crvena boja). Oznaka BZPK odnosi se na broj zaustavljanja vozila po kilometru.



Slika 3-13: Nasumično odabrani primjeri sintetičkih voznih ciklusa generiranih ne uzimajući u obzir stupanj zagušenja prometa: a), c), e) profili brzine u vremenu, te b), d), f) pripadni profili nagiba ceste u ovisnosti o prijeđenoj udaljenosti.



Slika 3-14: Histogrami snimljenih i sintetičkih uzoraka a) brzine vozila, b) ubrzanja vozila i c) nagiba ceste.

4. Regresijski model za predviđanje potrošnje goriva

Ovo poglavlje bavi se tvorbom regresijskog modela za predviđanje potrošnje goriva temeljem značajki voznih ciklusa. Naglasak je na uključivanju značajki nagiba ceste pri oblikovanju ulaza u model radi poboljšanja točnosti predviđanja u odnosu na model koji se zasniva samo na značajkama brzine i akceleracije vozila. U prvom potpoglavlju daje se motivacija i konceptualni prikaz metode predviđanja potrošnje goriva temeljem različitih familija regresijskih modela. U drugom potpoglavlju opisuje se postupak pripreme voznih ciklusa za parametriranje regresijskih modela, koji uključuje nadopunjavanje ulaznog skupa voznih ciklusa za potrebe bolje uvjetovanog učenja modela, te transformaciju voznih ciklusa u oblik pogodan za korištenje kao ulaz u modele temeljene na neuronskoj mreži (NN). U trećem potpoglavlju predstavljeni su i matematički opisani razmatrani regresijski modeli, uključujući polinomske modele i one temeljene na NN. U četvrtom potpoglavlju provodi se detaljna usporedna analiza postavljenih regresijskih modela u smislu točnosti predviđanja potrošnje goriva. U petom potpoglavlju opisuju se metode koje su korištene za redukciju dimenzionalnosti ulaznih podataka, te fino-podešavanje arhitekture NN modela i vrijednosti pripadnih hiperparametara.

4.1. Motivacija i konceptualni prikaz metode

Predviđanje potrošnje goriva temelji se na modelu koji bi trebao uzeti u obzir ponašanje u vožnji (stil vožnje) i vremenski-promjenjive prometne uvjete (zagušenje prometa i s tim povezane značajke poput srednje brzine, broj zaustavljanja i sl.) uključujući promjenjivi broj putnika. Budući da se većina varijabli koje utječu na potrošnju goriva vozila tipično ne mjere (npr. masa vozila i prometni uvjeti), ulažu se značajni napori u modeliranje potrošnje goriva samo s podskupom ključnih, standardno dostupnih varijabli [100]. Metode korištene za modeliranje potrošnje goriva mogu se generalno podijeliti u dvije glavne kategorije: (i) modeli temeljeni na fizikalnim zakonima koji opisuju dinamiku vozila u svakom vremenskom koraku koristeći detaljne matematičke jednadžbe za svaku komponentu vozila [101]; te (ii) podacima-vođeni modeli temeljeni na strojnom učenju (*engl. Machine Learning*; ML) koji predstavljaju apstraktno preslikavanje skupa ulaznih tj. nezavisnih varijabli u izlazni prostor koji predstavlja izlaznu tj. zavisnu varijablu [102].

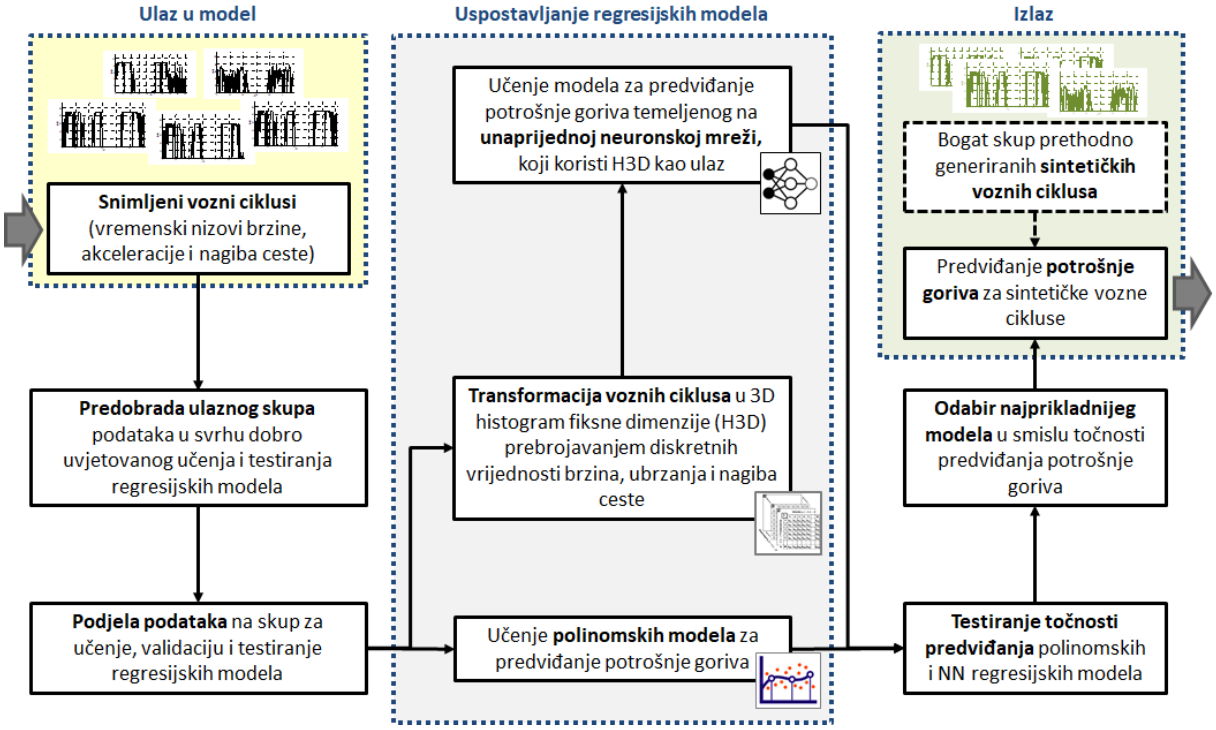
Pristupi temeljeni na preciznim analitičkim modelima pogonskih sklopova vozila mogu pružiti visoku točnost predviđanja [103, 104], ali po cijenu sporijeg vremena izračuna, što nije

prikladno za određene primjene koje zahtijevaju izvođenje velikog broja izračuna za razne hipotetske scenarije (npr. optimiranje rutiranja vozila). Drugi nedostatak fizikalnih modela jest to što zahtijevaju poznavanje svih mapa i parametara pogonskog sklopa vozila, a koje su u velikoj većini primjena nedostupni jer ih proizvođači vozila obično ne objavljuju i nerado ustupaju. Stoga postoji značajna potreba za razvojem brzog, makroskopskog modela koji predviđa potrošnju goriva za cijeli vozni ciklus odjednom, umjesto oslanjanja na mikroskopski fizički model koji se tipično izvodi na sekundnoj bazi. U tu svrhu potrebno je koristiti neki od suvremenih pristupa modeliranja temeljenih na ML [105-107], koje karakterizira svojstvo automatskog učenja značajki iz dostupnih podataka. Najčešće korišteni modeli u te svrhe su umjetne neuronske mreže (NN), jer su univerzalni aproksimatori i sposobni su preslikati nelinearne karakteristike složenog sustava pomoću nelinearnih aktivacijskih funkcija [108, 109]. Pored toga, NN-ove je lakše ponovno parametrirati za različita vozila i eventualne dodatne ulazne varijable, što nije slučaj s fizikalnim modelima. Glavni izazov kod modeliranja primjenom NN predstavlja određivanje arhitekture NN uključujući strukturiranje ulaza u model. Arhitektura NN obično se određuje postupkom "pokušaja i pogreške" (*engl. trial and error*) ili iscrpnog pretraživanja po unaprijed definiranom skupu mogućih arhitektura ("pretraživanja polja"; *engl. grid search*) [78], dok definiranje ulaza u model ovisi o problemu koji se razmatra i često predstavlja glavni dizajnerski doprinos.

Za potrebe predviđanja potrošnje goriva vozila parametriranje modela tipično se ostvaruje na podskupu empirijski odabranih statističkih značajki voznih ciklusa, izvučenih iz povijesnih podataka o vožnji [110]. Alternativni pristupi razmatraju dodatnu primjenu određene tehnike odabira značajki (*engl. Feature Selection*, FS), kako bi se pronašao podskup najvažnijih značajki za dani zadatak predviđanja [106, 111]. Glavni nedostatak navedenih pristupa je potreba za nominiranjem skupa ulaznih značajki, umjesto korištenja potpunih informacija o voznim ciklusu i oslanjanja na svojstvo automatskog izdvajanja značajki karakterističnog za NN.

U ovom radu predlaže se NN model koji se parametrira na temelju snimljenih voznih ciklusa gradskih autobusa (vidi poglavlje 2). Slijedeći osnovni koncept iz povezanog rada [112], vozni ciklusi različite duljine predstavljaju se histogramom fiksne dimenzije, čije osi odgovaraju stanju vozila i ceste (brzina vozila, ubrzanje vozila, i nagib ceste), kao što je prikazano na slici 4-1. Vremenski-promjenjiva masa putnika tj. vozila ne razmatra se pri modeliranju NN jer je uobičajeno nepoznata i jer se može pokazati da NN model može implicitno izvući znanje o masi vozila iz spomenutih temeljnih značajki voznih ciklusa [113]. U svrhu dobro uvjetovanog

učenja NN, koje uzima u obzir ograničenu razlučivost mjerene potrošnje goriva, snimljeni vozni ciklusi se predobrađuju u bogati skup voznih ciklusa temeljenih na mikro-ciklusima. Performanse predviđanja potrošnje goriva ispituju se u usporedbi s polinomskim modelima, kao i s NN koja ne uključuje podatak o nagibu ceste.



Slika 4-1: Konceptualni prikaz metode predviđanja potrošnje goriva temeljem različitih familija regresijskih modela.

Iako predloženi model za predviđanje potrošnje goriva ima potencijalno širok raspon primjene (npr. kod problema rutiranja vozila (*engl. Vehicle Routing Problem*, VRP) [116], gdje se dodatno minimizira potrošnja goriva), u ovom radu koristi se isključivo za potrebe određivanja potrošnje goriva generiranog skupa sintetičkih voznih ciklusa (poglavlje 3), a u svrhu njihove validacije (poglavlje 5 i 6).

4.2. Priprema voznih ciklusa za parametriranje regresijskog modela

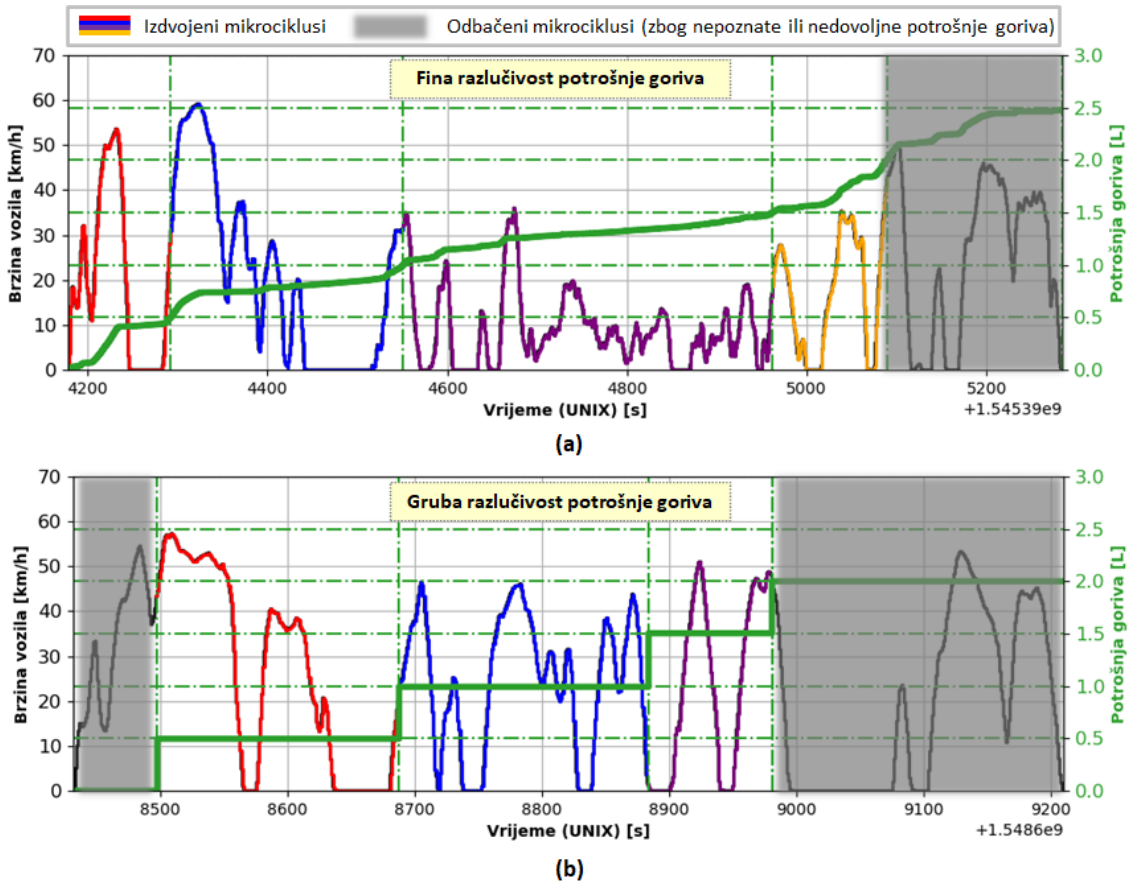
4.2.1. Obogaćivanje ulaznog skupa podataka

Glavna motivacija vezana za obogaćivanje ulaznog skupa podataka proizlazi iz toga što skup snimljenih voznih ciklusa uključuje samo sitne varijacije snimljenih potrošnji goriva, koje su posljedica relativno fiksnih prijeđenih udaljenosti zbog razmatranja samo jedne autobusne ruta (Babin kuk-Pile poglavlje 2). Time se pokriva vrlo ograničen prostor stanja definiran značajkama razmatrane rute, što nije pogodno sa stanovišta modeliranja potrošnje goriva,

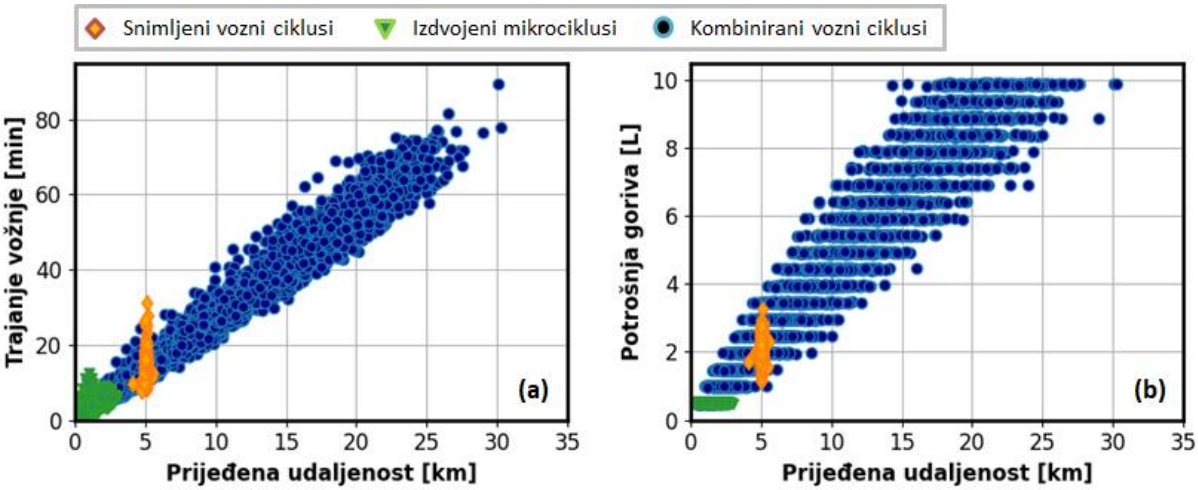
odnosno njegove primjene u šire svrhe. K tome, dodatan izvor pogreške predstavlja limitirana razlučivost mjerenja potrošnje goriva.

Kako bi se izbjegli navedeni problemi, snimljeni vozni ciklusi najprije se razlažu na mikrocikluse, na način da svaki microciklus odgovara potrošnji goriva jednakoj (gruboj) razlučivosti mjerenja od 0.5 L (vidi ilustraciju na slici 4-2). Zatim se kombiniranjem različitog broja nasumično odabranih mikrociklusa sastavlja bogat skup kombiniranih voznih ciklusa. Primjenom dotične metode generirano je ukupno 20 skupina od 200 kombiniranih voznih ciklusa, gdje svaka skupina uključuje različit broj mikrociklusa sadržanih u kombiniranom ciklusu, počevši od jednog za 1. skupinu do dvadeset za 20. skupinu. Konkretnije, generirano je ukupno 4000 kombiniranih voznih ciklusa, s rasponom prijeđenih udaljenosti do 30 km i potrošnje goriva od 0.5 L do 10 L. Na taj je način pripremljen vrlo širok skup podataka u smislu prijeđene udaljenosti i potrošnje goriva, pri čemu se uklanjaju određene pogreške povezane s (grubom) razlučivošću snimljenih potrošnji goriva. Ovo se ilustrira slikom 4-3, koja prikazuje ovisnost potrošnje goriva i vremena trajanja putovanja o prijeđenoj udaljenosti za snimljene vozne cikluse (narančasti dijamanti), mikrocikluse (zeleni trokuti), te kombinirane vozne cikluse (plavi kružići). Valjanost kombiniranih voznih ciklusa potvrđena je na slici 4-4, koja pokazuje kako razdiobe prosječnih brzina i ubrzanja kombiniranih voznih ciklusa statistički vjerodostojno odražavaju razdiobe dobivene za slučaj izvorno snimljenih voznih ciklusa.

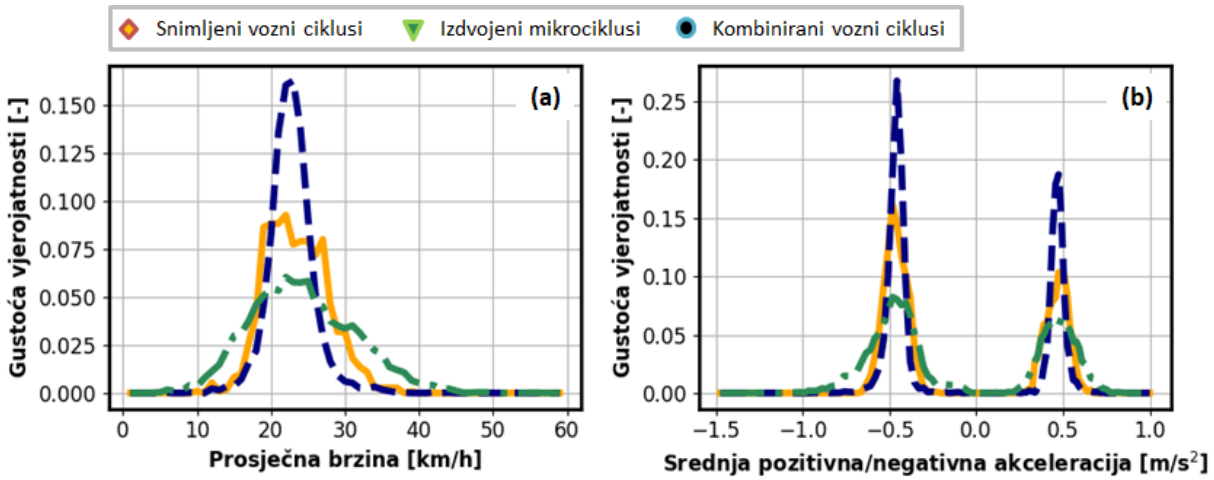
Nadalje, slika 4-5 prikazuje ovisnost potrošnje goriva o prijeđenoj udaljenosti i prosječnom nagibu ceste u slučaju kombiniranih voznih ciklusa. Evidentno je kako i prijeđena udaljenost i prosječni nagib ceste imaju snažan utjecaj na potrošnju goriva, s vrlo visokim vrijednostima koeficijentima korelacije od redom $\rho = 0.96$ i $\rho = 0.8$ (izračunate primjenom Pythonove funkcije *corrcoef(.)* iz *numpy* modula [93]). Stoga je u model predviđanja potrošnje goriva potrebno uključiti obje značajke kao ulaz, kako bi se osigurala visoka točnost predviđanja. Za potrebe razvoja modela potrošnje goriva, kombinirani vozni ciklusi su nasumično podijeljeni u tri glavna skupa: skup za učenje, validaciju i testiranje modela, s udjelima redom od 70%, 15% i 15% u ukupnim podacima. Skup podataka za učenje koristi se za učenje tj. parametriranje modela, skup podataka za validaciju koristi se za podešavanje hiperparametara modela i sprječavanje prenaučenosti (*engl. overfitting*) modela, dok je testni skup usmjeren na nepristranu evaluaciju točnosti modela.



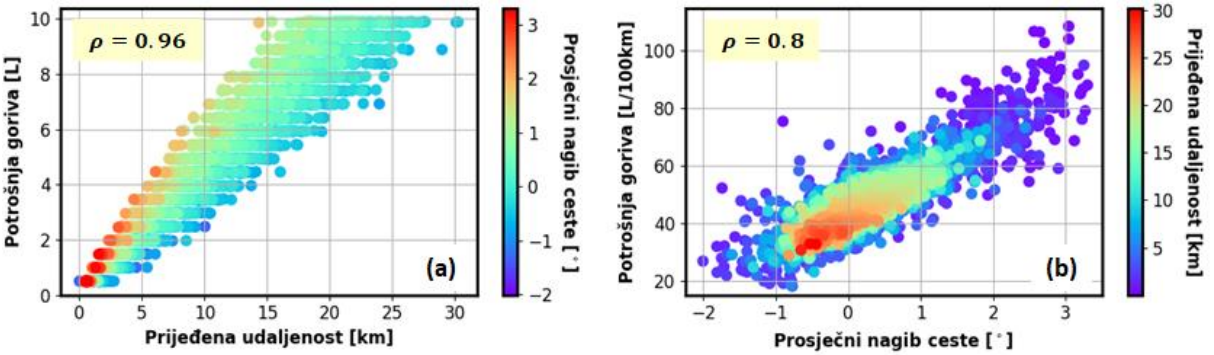
Slika 4-2: Ilustracija postupka razdvajanja snimljenog voznog ciklusa na skup mikrociklusa ekvivalentne potrošnje goriva od 0.5 L za slučaj a) fine i b) grube razlučivosti snimljene potrošnje goriva.



Slika 4-3: Ovisnost a) trajanja vožnje i b) potrošnje goriva za snimljene vozne cikluse, izdvojene mikrocikluse i kombinirane vozne cikluse (sintetizirane iz mikrociklusa).



Slika 4-4: Razdioba a) prosječne brzine i b) srednje pozitivne i negativne akceleracije vozila za snimljene vozne cikluse, izdvojene mikrocikluse i kombinirane vozne cikluse.



Slika 4-5: Potrošnja goriva u odnosu na a) prijeđenu udaljenost i b) prosječni nagib ceste za kombinirane vozne cikluse (one koji sadrže snimljene mikrocikluse).

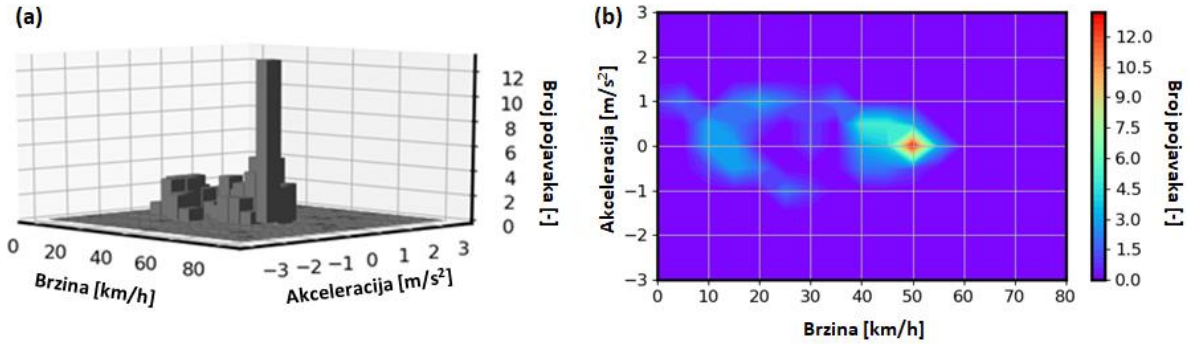
4.2.2. Transformacija voznih ciklusa u histograme fiksnih dimenzija

Transformacija voznih ciklusa različitog broja snimljenih uzoraka u histograme fiksnih dimenzija provodi se za potrebe modeliranja unaprijedne NN, koja zahtjeva statički ulaz. Ovaj oblik ulaza posebno je pogodan za napredne NN, poput konvolucijskih NN (*engl. Convolutional Neural Network*, CNN), koje su posebno dizajnirane i optimirane za zadatke klasifikacije slika [114, 115], jer se višedimenzionalni podatkovni niz različite duljine pretvara u sliku opisanu matricom fiksne dimenzije. Relevantna primjena CNN pokazana je u nedavnoj povezanoj studiji [112], gdje je predstavljeno modeliranje energetske potražnje električnog vozila proširenog dometa temeljem različitih formata ulaza, usmjereno ka predviđanju potrošnje goriva i stanja napunjenosti baterije te električnog dometa vozila na odredištu. Slijedeći pristup iz [111], za potrebe transformacije voznih ciklusa razmatra se 2D histogram (dalje označen kao *H2D*) definiran osima koje predstavljaju brzinu i ubrzanje vozila, te se predlaže i 3D histogram (dalje označen kao *H3D*) koji uključuje dodatnu os vezanu za nagib

ceste. Svaki element matrice kojom je opisan histogram označava broj uočenih stanja koja su se pojavila u odgovarajućem voznom ciklusu, gdje je svako stanje opisano kombinacijom diskretnih vrijednosti brzine i ubrzanja vozila u slučaju *H2D*, te dodatno nagibom ceste u slučaju *H3D*. Ovaj postupak inspiriran je tvorbom TPM-a, uz bitnu razliku da se kod TPM prebrojavaju prijelazi između diskretnih stanja za potrebe stohastičkog modeliranja voznog ciklusa (poglavlje 3).

Tvorba *H2D* i *H3D* zahtjeva određivanje raspona i razlučivosti razmatranih varijabli, pri čemu treba voditi računa da se ukupna dimenzija histograma kreće u razumnim granicama. U suprotnom je upitna izvedivost parametriranja modela zbog značajnog porasta njegove kompleksnosti. Prema tome, brzina vozila je u ovom slučaju diskretizirana s razlučivošću od 2 km/h u rasponu od 0 do 80 km/h (ukupno 41 diskretnih stanja brzine), dok je ubrzanje vozila diskretizirano s razlučivošću od 0.5 m/s^2 u rasponu od −3 do 3 m/s^2 (ukupno 13 diskretnih stanja ubrzanja), što rezultira ukupnim brojem od 533 moguća stanja u slučaju *H2D* (iste, grube razlučivosti odgovarajućih varijabli stanja razmatrane su u poglavlju 3). Nadalje, informacije o nagibu ceste uključene su proširenjem prikazane 2D ulazne matrice s dodatnom dimenzijom. Nagib ceste je diskretiziran s razlučivošću od 0.5° u rasponu od −6° do 6° (ukupno 25 diskretnih stanja nagiba ceste), rezultirajući ukupnim brojem od 13.325 moguća stanja u slučaju *H3D*. Neovisno o dimenzionalnosti histograma (*H2D* ili *H3D*), prebrojavanjem diskretnih stanja za zadano fiksno vrijeme uzorkovanja sve značajke voznog ciklusa koje utječu na potrošnju goriva (uključujući informaciju o prijeđenoj udaljenosti) implicitno su sadržane unutar matrice, koja zadržava istu (jedinstvenu) strukturu za vozne cikluse različite duljine [112].

Na slici 4-6 dana je ilustracija *H2D* u 3D i 2D obliku za primjer jednog kombiniranog voznog ciklusa. U slučaju *H2D* predstavljenog 2D oblikom (slika 4-6b), broj pojavaka odgovarajućih stanja brzine i ubrzanja opisan je ljestvicom boja. Na taj je način vozni ciklus transformiran u svojevrsnu sliku (opisanu 2D matricom u slučaju *H2D*), koja je, kao što je prethodno navedeno, izrazito pogodna za učenje CNN. Zapravo, zapis voznog ciklusa u obliku slike u slučaju *H2D* predstavlja ovisnost akceleracije o brzini vozila (tzv. SAFD; usp. sliku 4-6b sa slikom 2-15). Tako je na primjer iz slike 4-6b vidljivo kako se mnoštvo snimljenih stanja brzine i akceleracije gomila oko 50 km/h i 0 m/s^2, što ukazuje na visoki udio krstarenja tijekom vožnje. Vozni ciklus koji sadrži veći broj različitih obrazaca vožnje rezultira boljom popunjenosti matrice/histograma i obratno. Daljnjim svođenjem svakog *H2D* i *H3D* na 1D vektore fiksne dimenzije ($\mathbb{R}^{1\,x\,533}$ za *H2D* te $\mathbb{R}^{1\,x\,13.325}$ za *H3D*) dobivaju se ulazi pogodni za unaprijedne višeslojne NN. Ovaj postupak se u nastavku rada navodi kao „poravnavanje“ ulaza.



Slika 4-6: Primjer voznog ciklusa transformiranog u H2D predstavljenog u a) 3D i b) 2D obliku.

4.3. Modeliranje potrošnje goriva

4.3.1. Polinomski modeli

Najprije se razmatra jednostavan linearni model s jednim ulazom, koji se odnosi na prijeđenu udaljenost (d [m]). Model se formulira sljedećom jednadžbom te se nadalje označava kao $Poly_{1D}$

$$V_f = p_0 + p_1 d, \tag{4-1}$$

gdje je V_f [L] potrošnja goriva, a p_0 i p_1 su parametri modela.

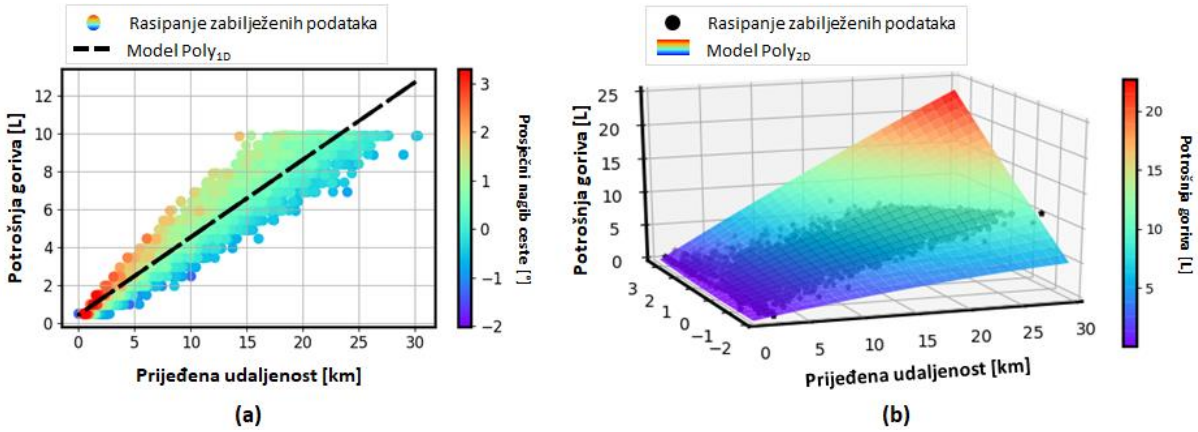
Zatim se uspostavlja polinomski (preciznije kvadratni) regresijski model linearan u parametrima koji kao ulaze uključuje prijeđenu udaljenost (d) i prosječni nagib ceste ($\bar{\theta}$). Model se formulira kao

$$V_f = p_{00} + p_{10} d + p_{01} \bar{\theta} + p_{20} d^2 + p_{11} d\bar{\theta} + p_{02} \bar{\theta}^2, \tag{4-2}$$

gdje $p_{ij}; i,j = 0,1,2$ predstavljaju parametre modela, te se nadalje označava kao $Poly_{2D}$.

Parametri oba regresijska modela ($Poly_{1D}$ i $Poly_{2D}$) identificirani su primjenom metode najmanjih kvadrata (*engl. Least Square Method*) na skupu podataka za učenje modela (odjeljak 4.2.1).

Identificirani modeli $Poly_{1D}$ i $Poly_{2D}$ vizualizirani su redom na slici 4-7a i 4-7b, zajedno sa snimljenim podacima. Iz slike 4-7 vidljivo je kako $Poly_{1D}$ bilježi očekivanje podataka, ali ne i njihove varijacije koje su uglavnom uzrokovane promjenjivim nagibom ceste, što nije slučaj kod $Poly_{2D}$ budući da uključuje prosječni nagib ceste kao dodatni ulaz u model.



Slika 4-7: Ilustracije polinomskih modela za predviđanje potrošnje goriva: a) model $Poly_{1D}$ i b) model $Poly_{2D}$, zajedno sa snimljenim podacima.

4.3.2. Teorijska podloga neuronskih mreža

Neuronska mreža nastala je iz vrlo popularnog algoritma strojnog učenja pod nazivom perceptron. Perceptron je 1950-ih i 1960-ih razvio znanstvenik Frank Rosenblatt, inspiriran ranijim radovima Warrena McCullocha i Waltera Pittsa. Perceptron uzima vrijednosti realnih brojeva na ulazu, te koristi aktivacijsku funkciju oblika praga osjetljivosti koja kao izlaz daje binarnu vrijednosti 0 ili 1. Iz tog razloga je primjenjiv samo na probleme koji su linearno odvojivi. Kombiniranjem više slojeva perceptrona dobivaju se tzv. višeslojni perceptroni (*engl. Multilayer Perceptron*; MLP). Ukoliko se umjesto perceptrona razmatraju višestruki statički neuroni s aktivacijskom funkcijom drugačijeg oblika, dobivaju se unaprijedne (*engl. feedforward*) NN, koje su u mogućnosti opisati složene odnose između ulaza i izlaza, pa su stoga pogodne za rješavanje nelinearnih problema [117]. Uspješnost primjene NN u području regresijskog predviđanja i klasifikacije osjetno je povećana u nedavno vrijeme razvojem i primjenom naprednijih tipova NN, poput CNN, koje koriste operator konvolucije nad ulazom u formi slike umjesto općeg množenja matrica [114]. Međutim, u nedavnoj povezanoj studiji [112] pokazano je kako je primjenom CNN moguće ostvariti tek neznatno poboljšanje točnosti predviđanja potrošnje energije električnog vozila proširenog dometa u odnosu na unaprijednu NN. Stoga se pod pretpostavkom istog tj. zanemarivog doprinosa točnosti predviđanja CNN za slučaj predviđanja potrošnje goriva u ovom radu nadalje primjenjuje i analizira unaprijedna NN zbog manje složenosti i bržeg učenja modela.

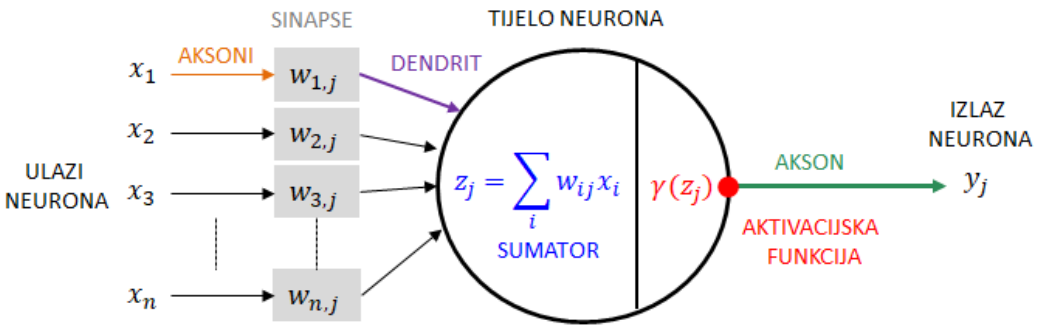
Opći model statičkog (umjetnog) neurona prikazan je na slici 4-8, a sastoji se od dvije temeljne podfunkcije [118]: (i) ***funkcija sume*** $\sum$, te (ii) ***aktivacijske funkcije*** γ. Pritom sumator predstavlja tijelo biološkog neurona, a ulazi u sumator dendrite. Izlaz sumatora odgovara aksonu biološkog neurona, dok se prag osjetljivosti preslikava preko aktivacijske funkcije (koja

može biti linearna ili nelinearna). Težinski faktori w_{ij} odgovaraju sinaptičkim vezama biološkog neurona s njegovom okolinom, a uglavnom su predstavljeni pozitivnim ili negativnim brojem (ukoliko je težina jednaka nuli, neuron je neaktivan). Osnovna značajka svakog neurona je posjedovanje više ulaza i samo jednog izlaza. Matematički model statičkog neurona opisan je ulaznim vrijednostima $x_i; i = 1,2,..n$ koji su realni brojevi te ***težinskim koeficijentima*** w_{ij}. Svaka ulazna vrijednost x_i najprije se množi s pripadnim težinskim koeficijentom w_{ij} i zatim zbraja u z_j:

$$z_j = \sum_i w_{ij} x_i. \tag{4-3}$$

Zatim se z_j obrađuje s pomoću aktivacijske funkcije γ_j te se dobiva izlazna vrijednost y_j:

$$y_j = \gamma(z_j). \tag{4-4}$$



Slika 4-8: Model statičkog (umjetnog) neurona.

Struktura unaprijedne mreže dobiva se organizacijom neurona u potpuno umrežene slojeve, odnosno na način da je svaki neuron promatranog sloja vezan sa svakim neuronom prethodnog sloja. Pritom je svaka veza opterećena pripadnim težinskim koeficijentom (w_{ij}).

Cilj unaprijedne mreže je aproksimirati neku funkciju f^* temeljem poznatih tj. definiranih parova ulaza $\boldsymbol{x}$ i izlaza $\boldsymbol{y}$ (nadzirano učenje). Na primjer, u slučaju klasifikacije $y = f^*(\boldsymbol{x})$ preslikava se ulaz $\boldsymbol{x}$ u kategoriju y, a u slučaju regresije u očekivanu vrijednost $y = E[\boldsymbol{x}]$. Konkretnije, unaprijedna mreža definira preslikavanje $\hat{y} = f(\boldsymbol{x}; \boldsymbol{W})$ i uči vrijednosti parametara (težinskih koeficijenata neurona) $\boldsymbol{W}$ koji rezultiraju najboljom aproksimacijom funkcije [78]. Mreža se naziva ***unaprijednom*** jer informacije teku samo u jednom smjeru (unaprijed), polazeći od ulaznog sloja koji prima $\boldsymbol{x}$, preko skrivenog sloja (jednog ili više) koji definira f, do izlaznog sloja koji vraća $\hat{y}$ (vidi sliku 4-9). Model unaprijedne NN može se prikazati usmjerenim acikličkim grafom koji opisuje na koji način su aktivacijske funkcije $(\gamma_s, \gamma_{s-1}, \dots, \gamma_1)$ povezane

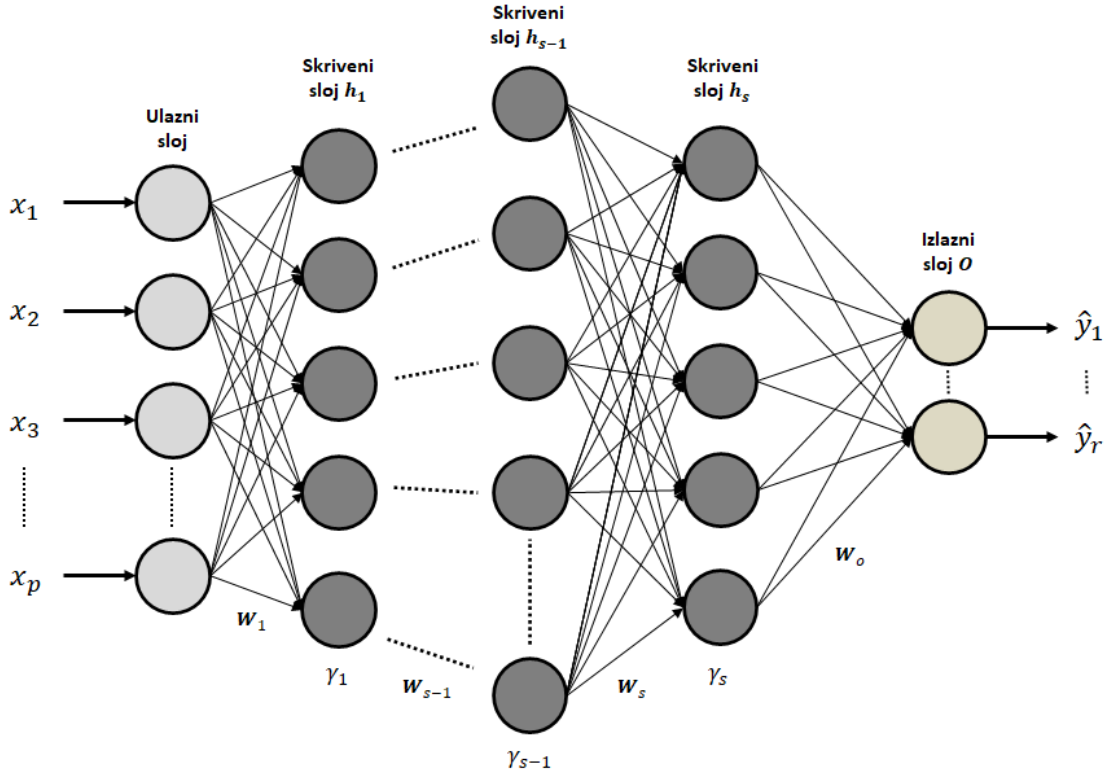
$$\hat{y} = f(\boldsymbol{x}; \boldsymbol{W}) = O(\gamma_s(\gamma_{s-1}(\dots(\gamma_1(\boldsymbol{x}, \boldsymbol{W}_1)\dots), \boldsymbol{W}_{s-1}), \boldsymbol{W}_s), \boldsymbol{W}_o), \tag{4-5}$$

gdje O označava aktivacijsku funkciju izlaznog sloja, a $\boldsymbol{W}_o$ pripadne težinske koeficijente.

Izraz (4-5) se može napisati i preko pomoćnih varijabli $\boldsymbol{h}_s, \boldsymbol{h}_{s-1}, \dots, \boldsymbol{h}_1$ koje se nazivaju ***značajke skrivenih slojeva*** kao

$$\begin{aligned} \boldsymbol{h}_1 &= \gamma_1(\boldsymbol{x}, \boldsymbol{W}_1), \\ &\vdots \\ \boldsymbol{h}_{s-1} &= \gamma_{s-1}(\boldsymbol{h}_{s-2}, \boldsymbol{W}_{s-1}), \\ \boldsymbol{h}_s &= \gamma_s(\boldsymbol{h}_{s-1}, \boldsymbol{W}_s), \\ f(\boldsymbol{x}; \boldsymbol{W}) &= O(\boldsymbol{h}_s, \boldsymbol{W}_o). \end{aligned} \tag{4-6}$$

Ukupni broj skrivenih slojeva s određuje ***dubinu*** mreže, dok dimenzionalnost skrivenih slojeva tj. broj neurona određuje ***širinu*** mreže. Dakle, kod unaprijednih NN nema povratnih veza u kojima se izlazi modela vraćaju natrag u model. Ukoliko postoji barem jedna povratna veza dobiva se tzv. povratna mreža (*engl. Recurrent Neural Network*, RNN) [78] koja je posebno dizajnirana za rad sa sekvencijalnim podacima jer posjeduje unutarnju memoriju. RNN i pripadne izvedenice poput LSTM-NN prvenstveno se koriste u područjima obrade prirodnog jezika, računalnog vida i analize vremenskih nizova (*engl. time series*), te nisu razmatrane u okviru ovog poglavlja [119].



Slika 4-9: Ilustracija unaprijedne višeslojne neuronske mreže.

Tijekom učenja unaprijedne NN podešava se funkcijska ovisnost $f(\boldsymbol{x};\boldsymbol{W})$ kako bi odgovarala ovisnosti $f^*(\boldsymbol{x})$ koja je određena ulazno-izlaznim skupom podataka za učenje. Svaki ulazni uzorak $\boldsymbol{x}$ popraćen je numeričkom vrijednošću $y = f^*(\boldsymbol{x})$, koja predstavlja ***ciljnu vrijednost*** koju izlazni sloj mora proizvesti za dani ulaz. Ponašanje ostalih slojeva nije izravno specificirano podacima za učenje, odnosno, algoritam učenja sam odlučuje kako će koristiti te slojeve za dobivanje željenog izlaza. Unaprijedne NN obično se uče s pomoću iterativnih optimizacijskih metoda temeljenih na gradijentnim algoritmima, kojima je cilj svesti ***funkciju gubitka*** $J(\boldsymbol{x}, y; \boldsymbol{W})$ na što nižu vrijednost (manju od nekog praga ε). Funkcija gubitka za slučaj klasičnog regresijskog zadatka jest srednja kvadratna pogreška (*engl. Mean Squared Error*, MSE) [117] definirana kao

$$J(\boldsymbol{x}, y; \boldsymbol{W}) = \frac{1}{N}\sum_{i=i}^{N}(f(\boldsymbol{x};\boldsymbol{W}) - y_i)^2 = \frac{1}{N}\sum_{i=i}^{N}(\hat{y}_i - y_i)^2, \tag{4-7}$$

gdje je y_i ciljna vrijednost za ulazno-izlazni par $(\boldsymbol{x}_i, y_i)$, $\hat{y}_i$ je izračunati izlaz mreže za ulaz $\boldsymbol{x}_i$, a N ukupni broj uzoraka u skupu podataka za učenje.

Postoje razne varijante učenja NN od kojih je najpopularnija metoda povratnog propagiranja pogreške (*engl. backpropagation*; BP) pogreške [118]. Pritom se BP odnosi samo na metodu izračunavanja gradijenta funkcije gubitka s obzirom na težinske koeficijente

$$\boldsymbol{\nabla_W J}(\boldsymbol{x}, y; \boldsymbol{W}) = \frac{\partial J}{\partial \boldsymbol{W}}, \tag{4-8}$$

dok se optimizacijski algoritam, npr. tipično stohastički gradijentni spust (*engl. Stochastic Gradient Descent*; SGD) [78], koristi za provođenje učenja NN, odnosno, podešavanje težina $\boldsymbol{W}$ s pomoću izračunatih gradijenta $\nabla_{\boldsymbol{W}}\boldsymbol{J}$ prema

$$\boldsymbol{W}(k+1) = \boldsymbol{W}(k) - \underbrace{\eta \cdot \nabla_{\boldsymbol{W}}\boldsymbol{J}(k)}_{\Delta\boldsymbol{W}(k)}, \tag{4-9}$$

gdje je k trenutni korak učenja (iteracija), a η koeficijent brzine učenja koji se tipično naziva ***stopa učenja*** (*engl. learning rate*) [78]. Vrijednost hiperparametra η potrebno je postaviti vlastoručno prije početka učenja NN. Izrazi (4-8) i (4-9) osiguravaju promjenu težinskih koeficijenata (parametara učenja) $\boldsymbol{W}$ u smjeru najstrmijeg pada ukupne pogreške $\nabla_{\boldsymbol{W}}\boldsymbol{J}$. Međutim, kao glavni nedostatak BP algoritma pokazao se u praksi zahtjev za velikim brojem iteracija učenja, pa stoga postoje brojne modifikacije algoritma koje tipično uključuju ***momentum*** (količinu gibanja), odnosno zamah $\alpha \cdot \Delta\boldsymbol{W}(k-1)$ koji se dodaje u izraz (4-9) kao [118]

$$\boldsymbol{W}(k+1) = \boldsymbol{W}(k) \underbrace{-\eta \cdot \nabla_{\boldsymbol{W}} \boldsymbol{J}(k) + \alpha \cdot \Delta \boldsymbol{W}(k-1)}_{\Delta \boldsymbol{W}(k)}. \tag{4-10}$$

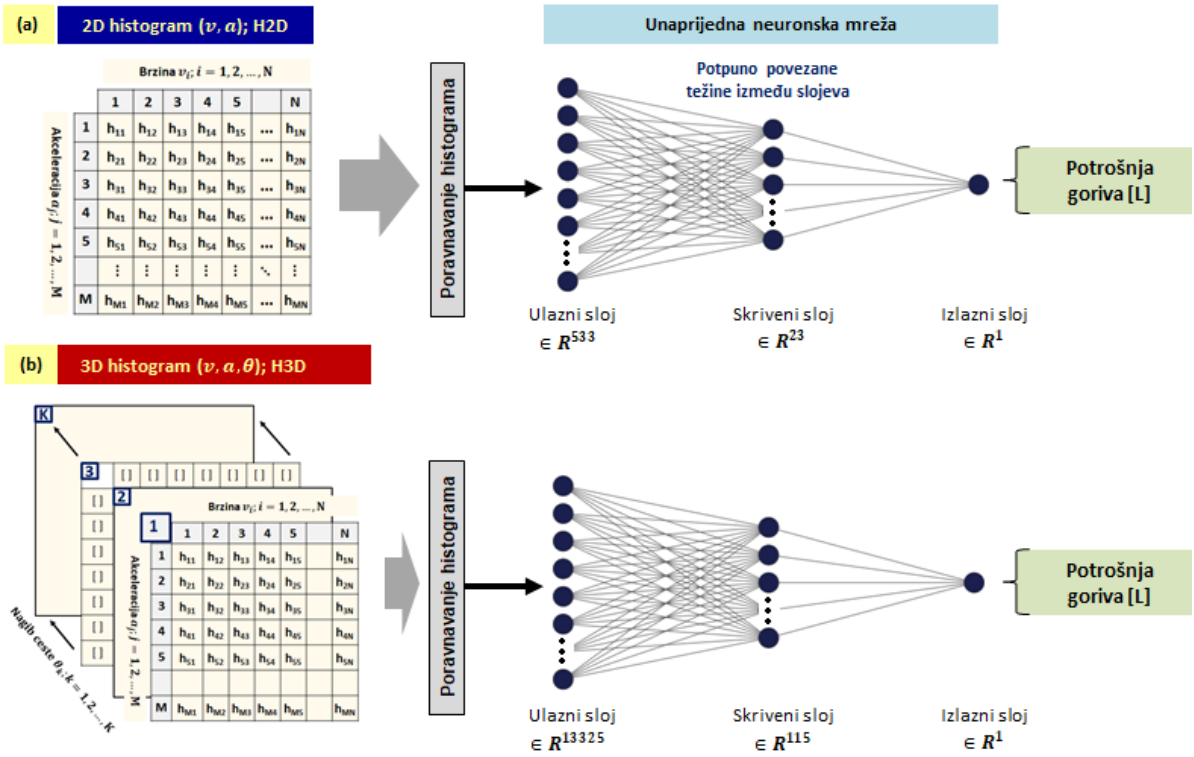
Pritom se vrijednost hiperparametra α tipično postavlja na neku od vrijednosti sadržanih u intervalu [0.2, 0.9]. Ovakvim proširenjem izraza (4-9) se brzina BP algoritma povećava i do 10 puta [118]. Nešto naprednije modifikacije BP algoritma uključuju dodatno adaptiranje stope učenja η kroz iteracije učenja NN modela te korekciju momentuma (npr. Nesterov ubrzani gradijent; *engl. Nesterov Accelerated Gradient,* NAG) [78, 120].

4.3.3. Modeli temeljeni na unaprijednoj neuronskoj mreži

Kako je navedeno u potpoglavlju 4.1, korištenje NN za modeliranje potrošnje goriva motivirano je univerzalnim aproksimacijskim teoremom, koji navodi kako unaprijedna NN (vidi odjeljak 4.3.2) s linearnim izlaznim slojem i najmanje jednim skrivenim slojem koji sadrži konačan (dovoljan) broj neurona, može aproksimirati bilo koju kontinuiranu funkciju konačno dimenzionalnog prostora uz određeni iznos pogreške ε [78].

Za potrebe modeliranja potrošnje goriva razmatraju se dva napredna NN modela, pri čemu je prvi (dalje označen kao *NN-H2D*) predstavljen arhitekturom prikazanom na slici 4-10a i koristi poravnati *H2D* definiran osima brzine i ubrzanja vozila, dok drugi (dalje označen kao *NN-H3D*) posjeduje istu arhitekturu ali koristi poravnati *H3D* kao ulaz, koji proširuje *H2D* dodatnom dimenzijom vezanom za nagib ceste (slika 4-10b). Izlaz iz NN modela predstavljen je apsolutnom potrošnjom goriva za čitav ulazni vozni ciklus, izraženom u litrama.

Budući da ne postoji egzaktna metoda za odabir optimalnog broja skrivenih slojeva i pripadnih neurona, razmatrane NN arhitekture odabrane su prema pravilu „geometrijske piramide“ predloženom u [79]. Ovo „pravilo palca“ (*engl. rule of thumb*) navodi kako bi za troslojnu mrežu s *N* ulaznih i *M* (ovdje 1) izlaznih neurona skriveni sloj trebao imati približno $\sqrt{N \cdot M}$ neurona, što rezultira kompetentnom inicijalnom arhitekturom NN. Prema tome, broj skrivenih neurona odgovarajućeg (jednog) skrivenog sloja postavljen je na 23 za slučaj *NN-H2D* te 115 za slučaj *NN-H3D* (*N* je jednak redom 533 i 13.325; odjeljak 4.2.2). Naknadnim provođenjem metode pretraživanja polja temeljem kriterija generalizacijske pogreške [78] (odjeljak 4.5.2), utvrđeno je kako je daljnjim podešavanjem arhitekture NN modela moguće ostvariti samo zanemarivo poboljšanje u točnosti predviđanja potrošnje goriva (RMSE manji za ≈0.01 L). Stoga su u konačnici zadržane inicijalno određene NN arhitekture s jednim skrivenim slojem i relativno niskim brojem skrivenih neurona.



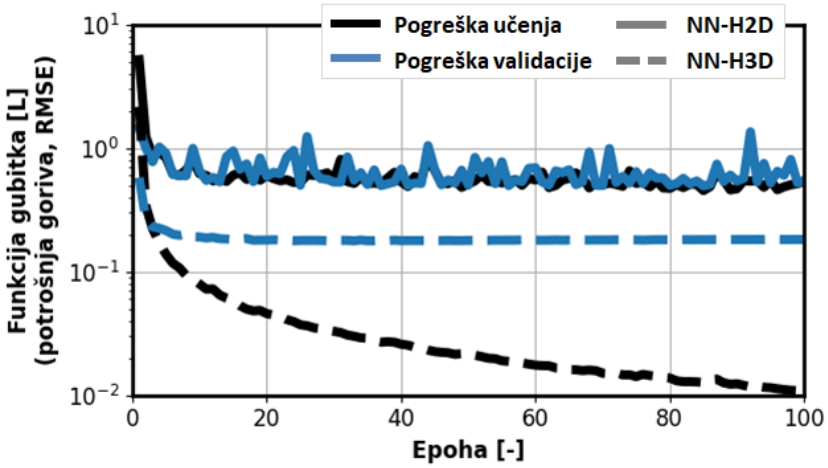
Slika 4-10: Predložene arhitekture neuronskih mreža za predviđanje potrošnje goriva, na temelju a) 3D histograma – H3D i b) 2D histograma – H2D vezanog za vozni ciklus.

Realizacija NN modela sa slike 4-10 izvodi se unutar *Python* programskog okruženja korištenjem *Keras* modula [121] s *Tensorflow* modulom kao pozadinom [122]. Za učenje NN modela koristi se ugrađeni optimizacijski algoritam ADAM (*engl. Adaptive Moment Estimation*) [78], koji je posebno pogodan za učenje dubokih neuronskih mreža te predstavlja proširenje SGD-a danog izrazom (4-9) odnosno (4-10). Razlika je u tome što SGD održava fiksnu stopu učenja za sva ažuriranja težinskih koeficijenata (parametara) mreže, dok se u slučaju ADAM algoritma stopa učenja izračunava za svaku individualnu težinu mreže zasebno te se adaptira kontinuirano tijekom odvijanja procesa učenja. Nadalje, provođenjem ispitivanja raspona stope učenja (*engl. Learning Rate Range Test*, LRTT) [123] pokazalo se kako se optimalna početna vrijednost stope učenja η za ovaj zadatak predviđanja nalazi negdje između 10^{-3} i 10^{-2} (vidi odjeljak 4.5.3 za detaljno pojašnjenje), te je konačno postavljena na vrijednost $\eta = 3 \cdot 10^{-3}$. Količina uzoraka iz koje se određuje gradijent (*engl. batch size*) postavljena je na 16, dok je broj epoha učenja postavljen na 100. U skrivenom sloju koristi se ReLU (*engl. Rectified Linear Unit*) [78], a u izlaznom sloju linearna aktivacijska funkcija. Razlog korištenja ReLU je u tome što izbjegava negativne vrijednosti preslikavajući ih u nulu, rezultirajući učinkovitijim učenjem NN. Kao funkcija gubitka koju treba minimizirati odabrana je MSE (usp. izraz (4-7)) potrošnje goriva

$$J_{MSE} = \frac{1}{N} \sum_{i=1}^{N} \left(V_{f,model,i} - V_{f,rec,i} \right)^2, \tag{4-11}$$

gdje N označava ukupan broj kombiniranih voznih ciklusa korišten za učenje NN modela, dok $V_{f,model,i}$ i $V_{f,rec,i}$ redom predstavljaju predviđenu i snimljenu potrošnju goriva za i-ti kombinirani vozni ciklus. Zbog bolje interpretabilnosti rezultata, u daljnjem tekstu analizira se korijen srednje kvadratne pogreške tj. funkcije gubitka $J_{RMSE} = \sqrt{J_{MSE}}$ [L].

Slika 4-11 prikazuje napredak funkcije gubitka J_{MSE} kroz epohe učenja oba NN modela, koja je dana za pogreške učenja i validacije. Pritom se težinski koeficijenti modela za koje se tijekom procesa učenja NN modela dobivaju minimalne validacijske pogreške usvajaju za parametriranje konačnih NN modela. U slučaju *NN-H2D*, zasićenje validacijske pogreške nastupa već nakon 10. epohe, dok se minimum od 0.492 L doseže u 84. epohi. Ovaj oblik napretka funkcije gubitka ukazuje na kvalitetu učenja i odsutnost prenaučenosti NN modela (dobra generalizacija modela). U slučaju *NN-H3D*, pogreške učenja i validacije značajno su smanjene u usporedbi s *NN-H2D*, čime se preliminarno ukazuje na važnost uključivanja informacije o nagibu ceste kao ulaz u NN model. Međutim, postoji određeni jaz između te dvije krivulje, koji se povećava kako napreduje učenje, što znači da se prilagođavanje NN modela sve bolje podudara sa skupom podataka za učenje, no isto ne vrijedi (barem ne u istoj mjeri) za skup podataka za validaciju. Unatoč ovoj pojavi, očuvana je monotonost krivulje validacijske pogreške, odnosno, nije došlo do prenaučenosti NN modela (krivulja validacijske pogreške ne počinje se povećavati nakon određenog broja epoha). Ukoliko bi se dotični problem želio eliminirati potrebno je primijeniti neku od metoda regularizacije modela (vidi odjeljak 4.5.4 za više informacija). Kao i u slučaju *NN-H2D*, zasićenje validacijske pogreške kod *NN-H3D* nastupa već nakon 10. epohe, dok se minimum od 0.176 L doseže u 29. epohi.

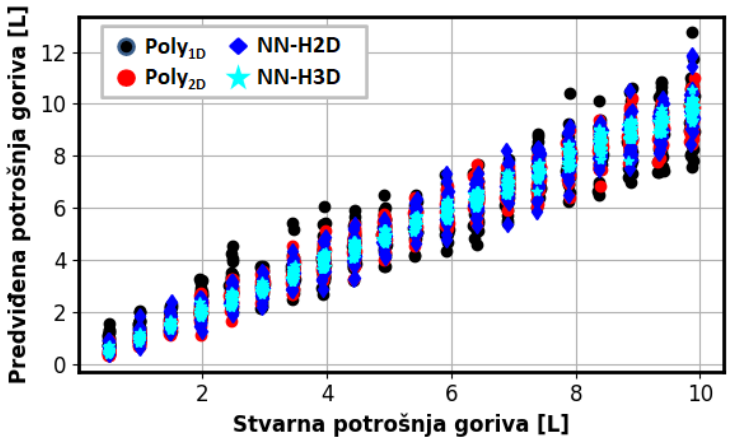


Slika 4-11: Krivulje pogreške učenja i validacije kroz epohe učenja za NN-H2D (puna linija) i NN-H3D (isprekidana linija).

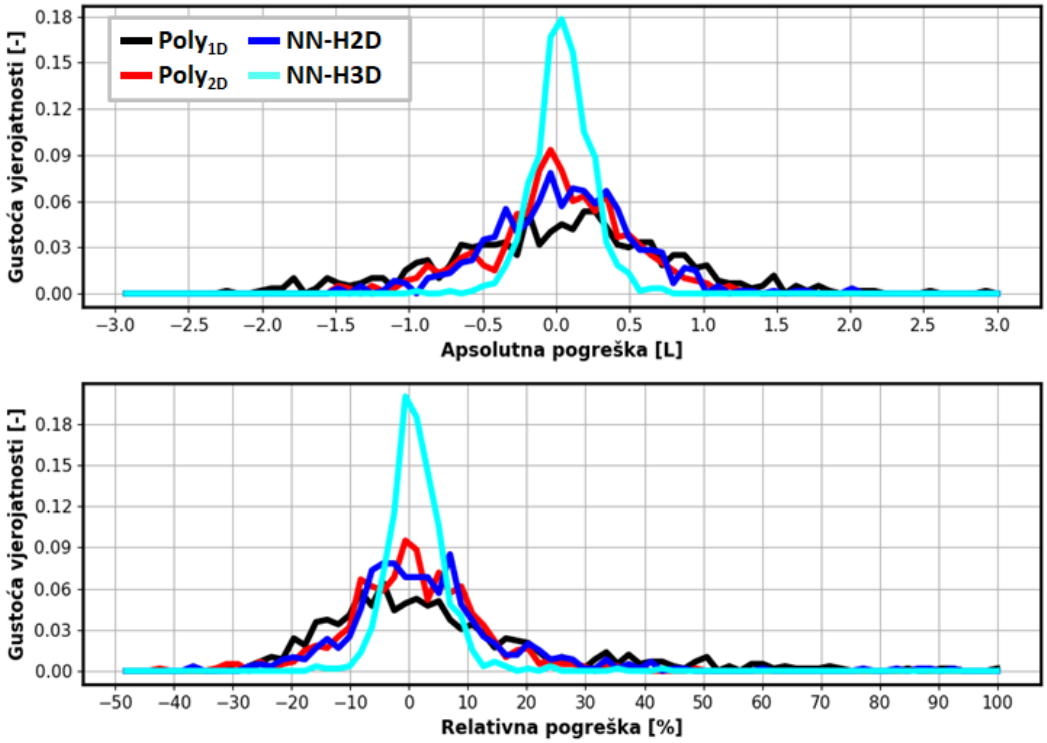
4.4. Usporedna analiza točnosti predviđanja potrošnje goriva

Učinkovitost predviđanja potrošnje goriva na temelju četiri različita modela definirana u potpoglavlju 4.3 ispitana je na skupu podataka za testiranje. Slika 4-12 prikazuje ovisnost predviđenih vrijednosti u odnosu na stvarne/snimljene vrijednosti potrošnje goriva. Evidentno je kako najjednostavniji linearni model $Poly_{1D}$ ima najveće pogreške u predviđanju potrošnje goriva (tj. odstupanja u odnosu na idealnu 1:1 ravnu liniju), dok NN-H3D model daje najtočnija predviđanja.

Odgovarajuće razdiobe apsolutnih i relativnih pogrešaka (reziduala) predviđanja prikazane su na slici 4-13. Zanimljivo je da složeniji linearni regresijski model $Poly_{2D}$ ima usporedivu razdiobu reziduala s jednostavnijim *NN-H2D* modelom, iz čega proizlazi da dodavanje informacije o nagibu ceste kao dodatnog ulaza može učiniti linearni regresijski model konkurentnim u odnosu na *NN-H2D* model (koji ne razmatra nagib ceste). Nadalje, vidljivo je kako *NN-H3D* model značajno nadmašuje ostale modele, s razdiobom pogreške predviđanja koja se uglavnom nalazi unutar ±0.5 L (tj. unutar ±10% relativne pogreške). Povrh toga, predviđanja modela su dobro uravnotežena, sa srednjom/očekivanom pogreškom blizu nule. Razlog daljnje prisutnosti određenih pogrešaka predviđanja kod *NN-H3D* modela je u tome što ne uzima u obzir utjecaj pomoćnih uređaja (npr. klimatizacije), koji nemaju izravne veze s voznim ciklusima, te utjecaj promjenjive mase vozila, čiju je informaciju moguće implicitno obuhvatiti kroz vozni ciklus samo djelomično [113].



Slika 4-12: Predviđene u odnosu na stvarne potrošnje goriva dobivene za različite regresijske modele na temelju testnog skupa podataka.



Slika 4-13: Razdiobe a) apsolutnih i b) relativnih pogrešaka predviđanja za različite regresijske modele i slučaj testnog skupa podataka.

Kvaliteta predviđanja modela dodatno se kvantificira R^2 vrijednošću (koeficijent determinacije), koja se određuje na testnom skupu podataka kao:

$$R^2 = 1 - \frac{\sum_{i=1}^{n}\left(V_{f,rec,i} - V_{f,model,i}\right)^2}{\sum_{i=1}^{n}\left(V_{f,rec,i} - \bar{V}_{f,rec}\right)^2}. \tag{4-12}$$

Koeficijent determinacije R^2 često se tumači kao relativni iznos varijance u snimljenim podacima, koju je moguće objasniti razmatranim modelom. Koeficijent poprima vrijednosti između 0 i 1, gdje $R^2 = 1$ odgovara idealnom predviđanju, dok $R^2 = 0$ odgovara najlošijem slučaju kada je izlaz modela konstantan i jednak srednjoj vrijednosti snimljene potrošnje goriva $\bar{V}_{f,rec}$. Nadalje, uvodi se korekcija R^2 na način da se predviđanja potrošnje goriva *Poly$_{1D}$* modela $V_{f,Poly1D,i}$ koristi umjesto $\bar{V}_{f,rec}$ u izrazu (4-12). Ovaj pokazatelj kvalitete predviđanja dalje je označen kao R_c^2, a može se tumačiti kao dodatna varijanca u snimljenim podacima koju je moguće objasniti razmatranim modelom, ali ne i *Poly$_{1D}$* modelom.

U tablici 4-1 dana su standardna odstupanja reziduala predviđanja σ_r, odgovarajuće vrijednosti pokazatelja R^2 i R_c^2, te prosječna vremena izvođenja $\bar{T}_p$ za svaki model (sve izračunato na skupu podataka za testiranje). Standardno odstupanje reziduala predviđanja σ_r iznosi 0.19 L za model *NN-H3D*, čineći ga približno 2.5 do 4 puta manjim u usporedbi s *NN-H2D, Poly$_{2D}$* i *Poly$_{1D}$* modelima. Interval pouzdanosti koji sadrži 95% predviđanja modela može se približno izračunati kao $\pm 2 \cdot \sigma_r$, a u slučaju *NN-H3D* modela iznosi 0.35 L. Svi modeli imaju visoke vrijednosti R^2 (> 0.9), dok se *NN-H3D* model približava gotovo idealnoj vrijednosti 1. Relativni odnosi različitih modela bolje su opisani (s višom osjetljivošću) u slučaju pokazatelja R_c^2, koji pokazuje kako *Poly$_{2D}$* i *NN-H2D* objašnjavaju oko 60% varijance u snimljenim podacima više od *Poly$_{1D}$* modela, dok u slučaju *NN-H3D* taj udio prelazi 90% (tj. opet se približava idealnoj vrijednosti 1). Prosječna vremena izvođenja modela, $\bar{T}_p$, nalaze se u rasponu od 1.5 do 4 μs za slučaj polinomskih modela te 920 do 1420 μs za slučaj NN modela. Značajno povećanje vremena izvođenja za NN modele izravno je povezano sa složenošću modela (tj. brojem parametara za učenje; BPZU), što se odražava kroz 35% niže vrijeme izvođenje za NN-H2D u odnosu na NN-H3D (ulazni vektor s 13.325 spram 533 elemenata).

Tablica 4-1 također sadrži rezultate dobivene za skup snimljenih podataka s grubom razlučivošću potrošnje goriva (jednakom 0.5 L; poglavlje 2). Ovi su rezultati dani unutar zagrada te potvrđuju da se slična razina performansi može postići bez obzira na to jesu li dostupni podaci o potrošnji goriva snimani s niskom ili visokom razlučivošću.

Tablica 4-1: Pokazatelji kvalitete predviđanja potrošnje goriva za različite modele izračunati na testnom skupu podataka.

Model	σ_r[L]	R^2[-]	R_c^2[-]	$\bar{T}_p$ [2] [µs]	BPZU [-]
Poly1D	0.757 (0.697)[1]	0.931 (0.943)	0.0 (0.0)	1.47	2
Poly2D	0.472 (0.398)	0.973 (0.981)	0.611 (0.674)	3.75	5
NN-H2D	0.484 (0.470)	0.972 (0.973)	0.593 (0.534)	920	12.306
NN-H3D	0.194 (0.133)	0.995 (0.998)	0.934 (0.964)	1420	1,532.606

[1] Brojevi u zagradama odgovaraju rezultatima dobivena u slučaju korištenja podataka potrošnje goriva niske (grube) razlučivosti.

[2] Prosječna vremena izvođenja mjerena na radnoj stanici sa 16 GB RAM-a i Intel® Xeon® Processor E5-1620 v3 @ 3.50GHz

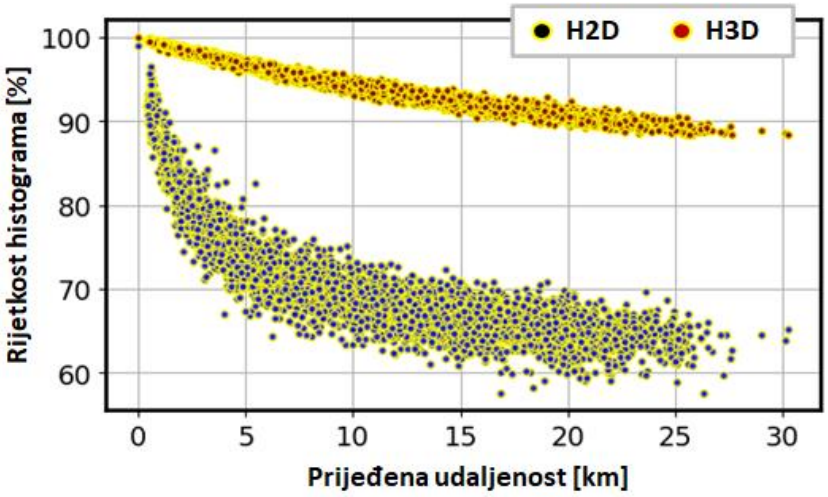
4.5. Fino-podešavanje neuronske mreže

4.5.1. Reduciranje dimenzionalnosti ulaza

Budući da su i *H2D* i *H3D* NN ulazi prilično dimenzionalni (posebno *H3D*), skloni su rijetkoj popunjenosti pripadnih matrica (tj. posjeduju velik broj elemenata matrice jednak nuli; slika 4-6). Rijetkost *H2D* i *H3D*, definirana kao postotak elemenata jednakih nula, analizirana je na slici 4-14 za svaki individualni kombinirani vozni ciklus. Rezultati sa slike 4-14 pokazuju kako se rijetkost *H2D* kreće u rasponu od 65% do 95% (crni kružići), a u slučaju *H3D* od 90% do 100% (crveni kružići). To ukazuje kako je dimenzija *H3D* (a također i *H2D* u određenoj mjeri) nepotrebno velika, te bi se mogla smanjiti reorganizacijom odgovarajućeg ulaza u NN, čime bi se smanjili memorijski zahtjevi povezani sa složenošću modela i povećala brzina učenja.

U tu svrhu potrebno je koristiti neku od tehnika reduciranja dimenzionalnosti koje provode transformaciju podataka iz prostora visoke dimenzije u prostor niske dimenzije $X_V \in \mathbb{R}^q \rightarrow X_N \in \mathbb{R}^w; w \ll q$ na način da nisko-dimenzionalna reprezentacija X_N zadržava sva značajna svojstva izvornih podataka X_V. Konkretnije, cilj je značajno reducirati dimenzionalnost podataka, uz očuvanje što je moguće više informacija. Jedna od tipično korištenih tehnika za reduciranje dimenzionalnosti je analiza glavnih komponenti (*engl. Principal Component Analysis*; PCA) [124], koja provodi linearno preslikavanje visoko-dimenzionalnih podataka u nisko-dimenzonalni prostor, uz maksimiziranje varijance. Princip rada PCA metode može se podijeliti u tri glavna koraka:

1) Standardizacija ulaznih varijabli $\boldsymbol{X} \in \mathbb{R}^{n \times m}$, gdje je n broj uzoraka (ovdje 4000 kombiniranih voznih ciklusa), a m broj značajki (ovdje poravnati *H3D* s 13.325 elemenata), kako bi svaka podjednako doprinosila analizi te kako bi pripadne vrijednosti bile centrirane oko ishodišta.
2) Izračun matrice kovarijanci $\sum \in R^{m \times m}$ kako bi se identificirale korelacije između značajki.
3) Izračun svojstvenih vektora $\boldsymbol{\Lambda}_i; i = 1,2, \dots, m$ i svojstvenih vrijednosti λ_i dekompozicijom matrice kovarijanci $\sum$ na singularne vrijednosti. Pritom vrijedi da $\boldsymbol{\Lambda}_i$ predstavljaju smjerove osi u kojima postoji najveća varijacija (glavne komponente), dok pripadne λ_i određuju količinu varijance svake glavne komponente.
4) Određivanje glavnih komponenti temeljem $\boldsymbol{\Lambda}_i$ i λ_i i odbacivanje onih od manjeg značaja (nižih vrijednosti λ_i), čime se dobiva vektor značajki $\boldsymbol{F} \in R^{m \times p}$ (matrica koja kao stupce ima vlastite vektore $\boldsymbol{\Lambda}_j; j = 1,2, \dots, p$ zadržanih komponenti).
5) Transformacija inicijalnih podataka u nove značajke projiciranjem na osi glavnih komponenti $\boldsymbol{X}^* = \boldsymbol{XF}$.

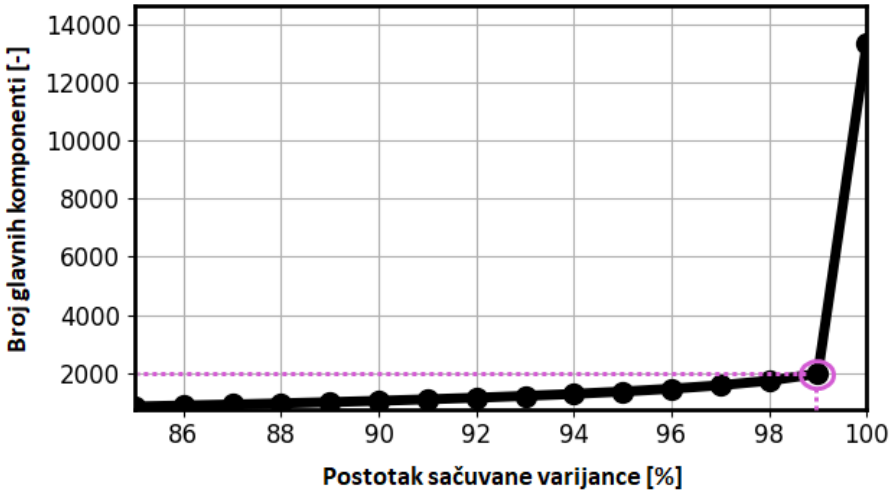


Slika 4-14: Rijetkost dva različita tipa ulaza u NN, a) H2D koji sadrži podatke o brzini i ubrzanju vozila, i b) H3D koji proširuje H2D informacijom o nagibu ceste.

Primjenom PCA metode dobivaju se glavne komponente koje predstavljaju nove, nekorelirane značajke koje su konstruirane linearnim kombiniranjem (ili miješanjem) početnih značajki. Geometrijski gledano, glavne komponente predstavljaju smjerove podataka koji objašnjavaju maksimalnu količinu varijance. Kako bi se odredila važnost svake glavne komponente, one se sortiraju od većih prema nižim vrijednostima pripadnih λ_i. Na taj način prva glavna komponenta (PC_1) sadrži najveću količinu varijance ulaznih podataka λ_1, druga glavna

komponenta (PC_2) nešto manju λ_2, itd. Postotak varijance (informacije) sadržane u svakoj komponenti σ_i izračunava se dijeljenjem svojstvene vrijednosti svake komponente λ_i sa zbrojem svih svojstvenih vrijednosti $\sigma_i = \frac{\lambda_i}{\sum_i \lambda_i}$. Konačni broj glavnih komponenti određuje se željenim postotkom zadržane varijance.

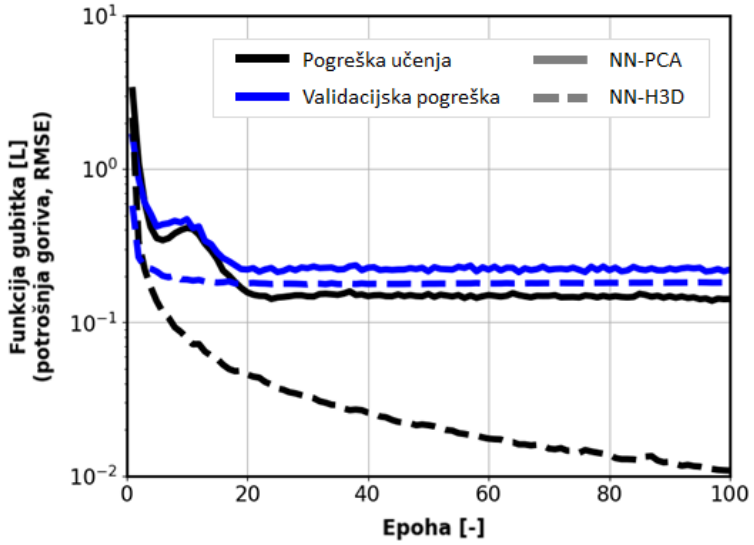
PCA metoda primijenjena je na ulazni skup podataka $X_{H3D} \in R^{4000\,\mathrm{x}\,13.325}$, sačinjen od poravnatih *H3D* (odjeljak 4.2.2) generiranih za svaki individualni kombinirani vozni ciklus. Rezultati prikazani na slici 4-15 pokazuju kako je moguće sačuvati 99% varijance izvornog skupa podataka $\boldsymbol{X}_{H3D}$, uz redukciju dimenzionalnosti s 13.325 značajki na 1976 glavnih komponenti (redukcija od 85%).



Slika 4-15: Profil sačuvane varijance u ovisnosti o broju glavnih komponenti.

Nadalje, ponovljeno je učenje *NN-H3D* modela s reduciranim skupom podataka $\boldsymbol{X}^*_{H3D} \in R^{4000\,\mathrm{x}\,1976}$ (dalje označavanog kao *NN-PCA*) kako bi se ispitalo u kojoj mjeri redukcija dimenzionalnosti utječe na kvalitetu točnosti predviđanja modela. Na slici 4-16 je prikazan napredak funkcije gubitka J_{RMSE} kroz epohe/iteracije učenja *NN-PCA* modela, gdje je vidljivo kako redukcija dimenzionalnosti *H3D* ulaza također ima regularizacijski učinak (značajno smanjenje generalizacijske pogreške; vidi odzive pogreške učenja na slici 4-16). Ovaj učinak pogodan je u slučajevima predviđanja izvan poznatog raspona značajki (tzv. ekstrapolacijski problem), gdje se zahtijevaju dobra generalizacijska svojstva modela. Usporedna statistika točnosti predviđanja za slučaj testnog skupa podataka, memorijskih zahtjeva (odraženih kroz BPZU), te vremena učenja T_u i prosječnog vremena izvođenja jednog predviđanja $\bar{T}_p$ za *NN-H3D* i *NN-PCA* modele dana je u tablici 4-2. Rezultati pokazuju kako je reduciranjem dimenzionalnosti *H3D* ulaza moguće smanjiti vrijeme učenja za ≈85% te vrijeme izvođenja za 30% (razina bliska *NN-H2D*), uz neznatno lošiju kvalitetu točnosti predviđanja potrošnje goriva

(R_c^2 od 0.908 za *NN-PCA* spram 0.934 za *NN-H3D*, što rezultira razlikom standardnih devijacija reziduala predviđanja $\Delta\sigma_r$ od +0.04 L; tablica 4-2). Dodatna prednost vezana je uz značajno smanjenje memorijskih zahtjeva NN-PCA modela od ≈95% s obzirom na NN-H3D, koje može biti od iznimne važnosti u realnim primjenama gdje su ograničeni kapaciteti memorije mikroprocesorske jedinice.



Slika 4-16: Krivulje pogreški učenja i validacije kroz epohe učenja za NN-PCA (puna linija) i NN-H3D (isprekidana linija).

Tablica 4-2: Pokazatelji kvalitete predviđanja potrošnje goriva za NN-H3D i NN-PCA modele izračunati na testnom skupu podataka.

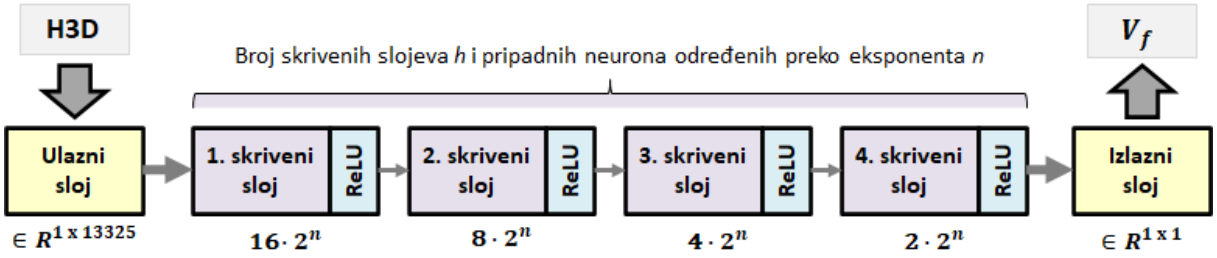
Model	σ_r [L]	R^2 [-]	R_c^2 [-]	$\bar{T}_p$[1] [µs]	T_u[1] [s]	BPZU [-]
NN-PCA	0.230	0.994	0.908	983	27.5	87.033
NN-H3D	0.194	0.995	0.934	1420	170.6	1,532.606

[1] Vremena izvođenja i učenja mjerena su na radnoj stanici sa 16 GB RAM-a i Intel® Xeon® Processor E5-1620 v3 @ 3.50GHz.

4.5.2. Određivanje prikladne arhitekture

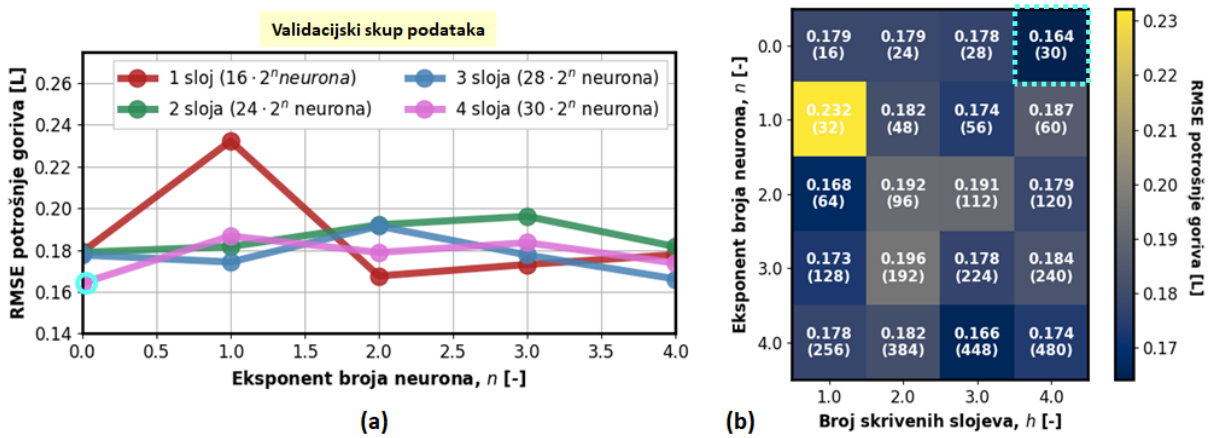
Kao što je opisano u potpoglavlju 4.3, za potrebe modeliranja potrošnje goriva razmatra se unaprijedna NN koja kao ulaz prima vozni ciklus transformiran u 3D histogram (*H3D*). Pritom je arhitektura NN modela inicijalno određena pravilom geometrijske piramide [79], koje ne garantira optimalnost u smislu definiranog kriterija (4-11). Stoga se za potrebe pronalaska najprikladnije arhitekture NN modela provodi metoda pretraživanja polja, pri čemu se

razmatraju različiti kandidati u smislu kombinacija broja skrivenih slojeva $h = 1, 2, 3, 4$ i pripadnih neurona $16 \cdot 2^n$, $8 \cdot 2^n$, $4 \cdot 2^n$ i $2 \cdot 2^n$ (vidi sliku 4-17), gdje je n eksponent čije se vrijednosti variraju u rasponu od $n = 1, 2, 3, 4$. Pritom je broj neurona u svakom skrivenom sloju postavljen tako da slijedi geometrijsku progresiju [79]. Najprikladnija arhitektura modela odabire se temeljem kriterija generalizacijske pogreške dobivene na skupu podataka za validaciju, koja predstavlja mjeru kojom se opisuje točnost predviđanja modela na neviđenim podacima [78].



Slika 4-17: Arhitekture NN s varijabilnim brojem skrivenih slojeva i pripadnih neurona koje se razmatraju prilikom pronalaska najprikladnijeg modela potrošnje goriva.

Rezultati ispitivanja različitih NN modela prikazani su na slici 4-18, gdje je cijan bojom označena arhitektura NN koja je rezultirala najmanjom validacijskom pogreškom (ovaj model dalje se označava kao *NN-OPT*). Analiza pokazuje kako se najbolje performanse NN modela postižu za arhitekturu koja se sastoji od $h = 4$ skrivena sloja, gdje svaki skriveni sloj (počevši od prvog) posjeduje redom 16, 8, 4, 2 neurona (eksponent $n = 0$; vidi cijan kružić na slici 4-18a). Glavni razlog ovog poboljšanja performansi NN modela u odnosu na NN-H3D model s jednim skrivenih slojem može se pripisati poboljšanoj sposobnosti izdvajanja skrivenih značajki koje je karakteristično za duboke (višeslojne) NN. Nadalje, rezultati na slici 4-18b pokazuju kako se točnost predviđanja NN modela, izražena kroz RMSE apsolutne potrošnje goriva, uglavnom kreće u rasponu od 0.16 L do 0.18 L, uz izuzetak modela s jednim skrivenim slojem s 32 neurona (RMSE jednak 0.23 L). Za najprikladniji, približno-optimalni model (*NN-OPT*) RMSE se smanjuje za 0.012 L (-6.8%) u odnosu na inicijalno razmatrani model *NN-H3D*. Time se potvrđuje kako je primjenom znatno jednostavnijeg pravila piramide [79] moguće postići performanse gotovo identične NN modelu čija je arhitektura određena znatno složenijim i vremenski zahtjevnijim postupkom pretraživanja polja predstavljenom u ovom odjeljku.



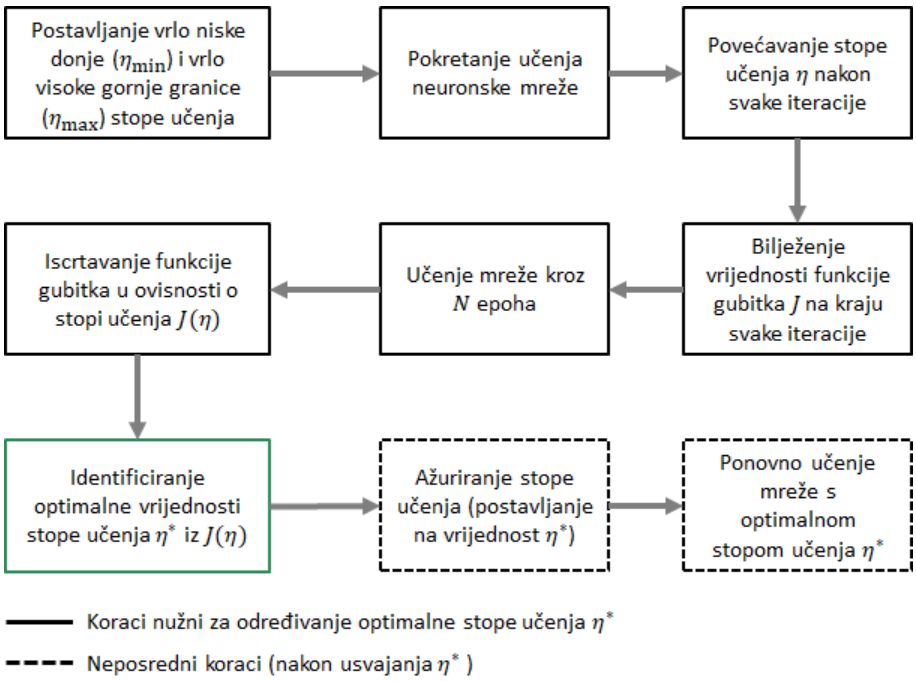
Slika 4-18: a) Ovisnost RMSE potrošnje goriva o arhitekturi NN definirane brojem skrivenih slojeva i pripadnih neurona, te b) pripadna toplinska mapa (brojke unutar zagrada odnose se na ukupni broj neurona u modelu). Cijan bojom označena je najprikladnija arhitektura NN modela (NN-OPT).

4.5.3. Određivanje optimalne stope učenja

Stopa učenja η predstavlja najvažniji hiperparametar u procesu učenja NN, jer određuje koliki udio gradijenta će se povratno propagirati kroz mrežu tijekom učenja (vidi izraz (4-9) u odjeljku 4.3.2), što u konačnici određuje brzinu konvergencije prema minimumu funkcije gubitka J_{MSE} [78]. Niske vrijednosti stope učenja rezultiraju sporom konvergencijom funkcije gubitka prema minimumu, dok visoke vrijednosti stope učenja dovode do divergencije. Stoga je stopu učenja potrebno postaviti na korektnu vrijednost koja je inicijalno nepoznata. Iako je za potrebe učenja *NN-H3D* korišten ADAM optimizacijski algoritam, koji adaptira stopu učenja kroz iteracije učenja, u slučajevima relativno malog broja epoha učenja (kao ovdje) potrebno je odrediti njezinu početnu vrijednost.

Manualni pronalazak najprikladnije stope učenja može biti vremenski zahtjevan zadatak ukoliko se radi o složenom NN modelu (koji se sporo uči), te također ne garantira optimalnost. Zato se u novijoj literaturi preporučuje provođenje testa raspona stope učenja (LRTT) [123], koji se pokazao kao učinkovita metoda za procjenu optimalne vrijednosti stope učenja η^*. Ilustracija cjelokupnog postupka određivanja optimalne stope učenja primjenom LRRT metode prikazana je na slici 4-19. Ideja LRRT-a je započeti učenje NN modela s nekom vrlo niskom vrijednošću stope učenja $\eta_{\min}$, te zatim povećavati vrijednost stope učenja nakon svake iteracije učenja sve dok funkcija gubitka ne krene rapidno rasti (eksplodirati), ili dok se ne izvrši određeni broj epoha učenja N (tada η poprima maksimalnu vrijednost $\eta_{\max}$). Zatim je potrebno iscrtati zabilježene vrijednosti funkcije gubitka (dobivene na skupu podataka za validaciju) u

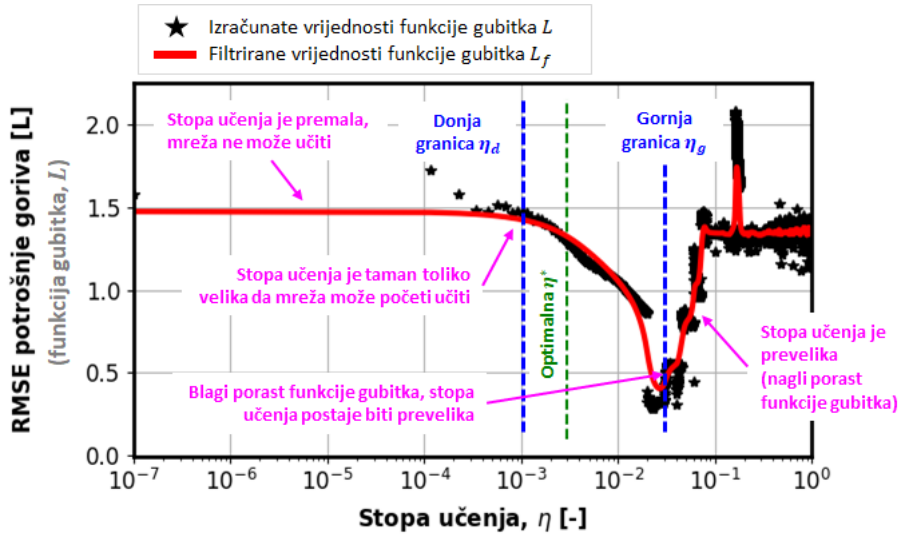
ovisnosti o stopi učenja $L(\eta)$. Vrijednost η za koju se dobiva najniža vrijednost funkcije gubitka L predstavlja maksimalnu vrijednosti stope učenja do koje valja ići (gornja granica η_g), dok je minimalna vrijednost stope učenja (donja granica $\eta_d < \eta_g$) određena vrijednošću η za koju funkcija gubitka L kreće strmije padati (početak učenja NN). Ukoliko se razmatra fiksna stopa učenja, optimalnom stopom učenja η^* smatra se bilo koja vrijednost u intervalu $[\eta_d, \eta_g]$. Ova metoda izrazito je učinkovita ukoliko se kod učenja NN modela umjesto adaptivnih optimizacijskih algoritama razmatra ciklička stopa učenja, koja, umjesto monotonog smanjenja stope učenja dopušta da se stopa učenja ciklički mijenja između graničnih vrijednosti η_d i η_g [123].



Slika 4-19: Ilustracija postupka određivanja optimalne stope učenja primjenom metode LRRT u obliku dijagrama toka.

Gore opisana LRRT metoda primijenjena je na *NN-H3D* model definiran u odjeljku 4.3.3. Pritom je vrijednostima $\eta_{\min} = 10^{-7}$ i $\eta_{\max} = 10^{0}$ određen raspon stope učenja $\eta \in [\eta_{\min}, \eta_{\max}]$. Broj epoha učenja postavljen je na $N = 50$, što rezultira ukupnim brojem od 8750 iteracija učenja za slučaj ulaznog skupa podataka za učenje koji se sastoji od 2800 kombiniranih voznih ciklusa (vidi potpoglavlje 4.2), uzimajući pritom u obzir da je potrebno 16 uzoraka za izračunavanje gradijenta. Rezultati primjene LRRT metode prikazani su na slici 4-20, a predstavljeni su krivuljom ovisnosti funkcije gubitka J_{RMSE} (RMSE potrošnje goriva) o vrijednosti stope učenja η. Budući da su izvorno dobivene vrijednosti funkcije gubitka jako

zašumljene (crne zvjezdice), dodatno su zaglađene primjenom dvostranog Butterworth filtra 2. reda (crvena linija). Vidljivo je kako NN model nije u mogućnosti učiti (funkcija gubitka J_{RMSE} se ne smanjuje) sve dok stopa učenja ne postigne vrijednost $\eta \approx 10^{-3}$ (gornja granica η_g), nakon čega J_{RMSE} počinje strmije padati do minimuma za $\eta \approx 10^{-2}$ (donja granica η_d). Daljnjim porastom stope učenja $\eta > 10^{-2}$ funkcija gubitka J_{RMSE} kreće naglo rasti, što označava kraj testa. Prema LRRT metodi vrijedi da se optimalna vrijednost stope učenja nalazi negdje u intervalu $\eta^* \in [10^{-3}, 10^{-2}]$, pa se stoga u konačnici odabire vrijednost $\eta^* = 3 \cdot 10^{-3}$ (označena isprekidanom zelenom linijom na slici 4-20). Odabrana vrijednost η^* korištena je kao početna vrijednost ADAM optimizacijskog algoritma prilikom učenja *NN-H2D* i *NN-H3D* modela definiranim u potpoglavlju 4.3.



Slika 4-20: *Krivulja ovisnosti funkcije gubitka o stopi učenja dobivena primjenom LRRT metode.*

4.5.4. Regularizacija neuronske mreže

Središnji problem strojnog učenja je kako učiniti da razmatrani model posjeduje željene performanse ne samo na podacima za učenje, već i na neviđenim ulaznim podacima (tj. da dobro generalizira). Pritom općenito vrijedi pravilo da modeli jednostavnije strukture posjeduju bolja generalizacijska svojstva (tj. manje su sklona prenaučenosti). Strategije kojima je cilj smanjenje generalizacijske pogreške, ali ne i pogreške učenja, nazivaju se regularizacijske tehnike. Postoje mnoge tehnike regularizacije, od kojih se najčešće koristi penaliziranje L^2 norme parametara modela (poznata kao hrbatna regresija - *engl. ridge regression* ili Tikhonova regularizacija) [78]. Ova tehnika regularizacije dovodi parametre/težinske koeficijente NN modela bliže ishodištu (tj. prema nuli) dodavanjem regularizacijskog člana $\Omega(\boldsymbol{W}, \lambda) = \lambda \|\boldsymbol{W}\|_2^2$

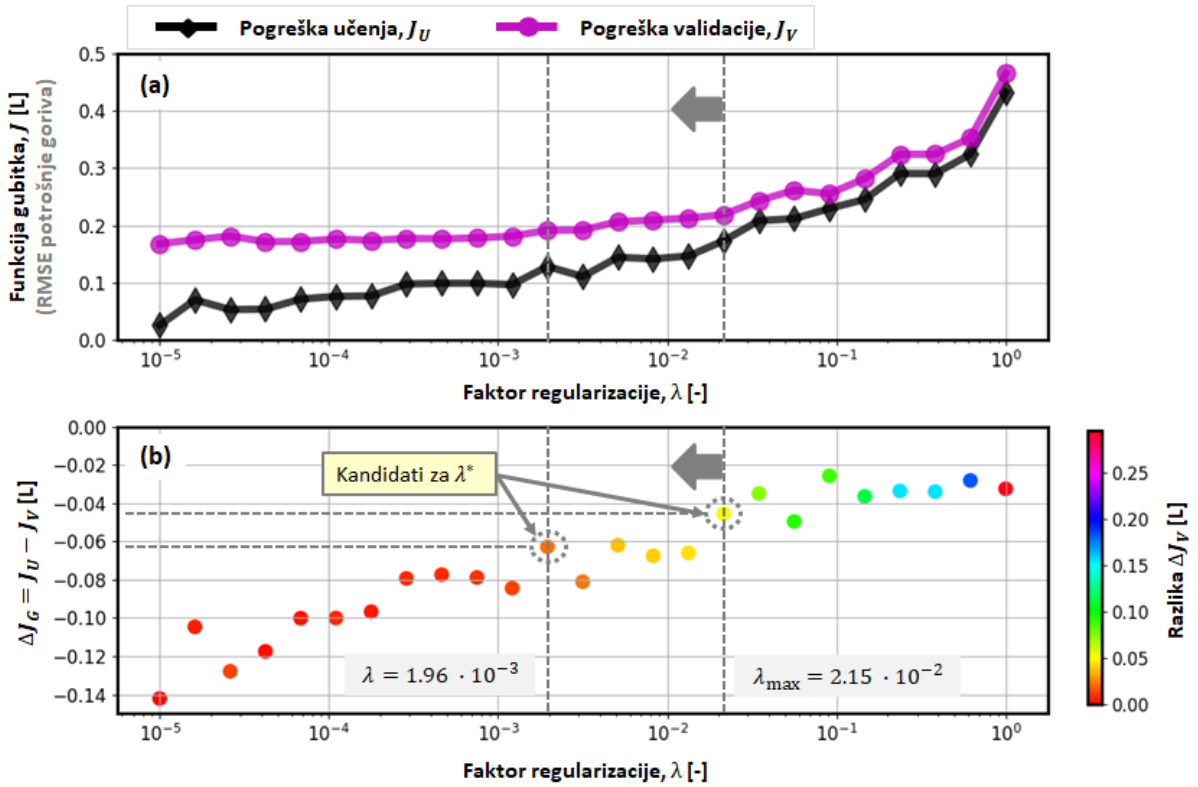
jednakim kvadratnoj magnitudi težina NN modela $\boldsymbol{W}$ funkciji gubitka $J(\boldsymbol{x}, y; \boldsymbol{W})$ (usp. izraz (4-7))

$$\tilde{J}(\boldsymbol{x}, y; \boldsymbol{W}) = J(\boldsymbol{x}, y; \boldsymbol{W}) + \Omega(\boldsymbol{W}, \lambda), \tag{4-13}$$

gdje je λ faktor regularizacije, a $\tilde{J}(\boldsymbol{x}, y; \boldsymbol{W})$ regulizirana funkcija gubitka.

Povećanjem vrijednosti λ jače se kažnjavaju veće vrijednosti težina $\boldsymbol{W}$, pa stoga optimizacijski algoritam tijekom učenja modela nastoji smanjiti njihove vrijednosti, što posljedično dovodi do jednostavnijeg modela koji je manje sklon prenaučenosti. S druge strane, vrijednost λ ne smije biti prevelika jer bi mogla dovesti do problema nedovoljnog prilagođavanja (*engl. underfitting*). Stoga je potrebno pronaći optimalnu vrijednost λ koja će rezultirati minimalnom generalizacijskom pogreškom. U tu svrhu, provedena je regularizacija *NN-H3D* modela definiranog u odjeljku 4.3.3 za različite vrijednosti λ, koje se kreću u rasponu od 10^{-5} do 1. Pritom je prostor pretraživanja diskretiziran s fiksnim razmakom u logaritamskom mjerilu na 25 vrijednosti λ.

Rezultati prikazani na slici 4-21a ukazuju kako s porastom vrijednosti λ raste vrijednost pogreške učenja J_U i pogreške validacije J_V (izražene kroz RMSE potrošnje goriva), dok se istovremeno smanjuje njihova razlika $L_U - L_V$ (slika 4-21b). Međutim, kako porast pogreške validacije J_V postaje intenzivniji nakon određene granice $\lambda > \lambda_{\max}$, potrebno je odabrati λ koji u dovoljnoj mjeri smanjuje jaz između pogreške validacije i učenja ($\Delta J_G = J_U - J_V$), a istovremeno zadržava vrijednost J_V relativno blisku onoj dobivenoj bez regularizacije NN modela, J_V^0, odnosno za koju razlika $\Delta J_V = J_V - J_V^0$ nije značajna. Kandidati koji predstavljaju optimalnu vrijednost λ^* označeni su na slici 4-21b sivim kružićima, te pokazuju kako je moguće smanjiti $|\Delta J_G|$ s 0.14 L na < 0.07 L, zadržavajući pritom porast validacijske pogreške ΔJ_V manjim od 0.05 L ($\lambda \in \{1.96 \cdot 10^{-3}, 2.15 \cdot 10^{-2}\}$). Ukoliko je cilj maksimalno regularizirati NN model, optimalnu vrijednost faktora regularizacije potrebno je postaviti na $\lambda^* = \lambda_{max} = 2.15 \cdot 10^{-2}$, dok je u suprotnom slučaju najprikladnija vrijednost $\lambda^* = 1.96 \cdot 10^{-3}$. Time se NN modelu daje određena robusnost s obzirom na neopćenite/neviđene podatke koji nisu sadržani u skupu podataka za učenje. Međutim, budući da to nije slučaj kod obogaćenog skupa snimljenih voznih ciklusa na kojem je izvorno parametriran *NN-H3D* model (potpoglavlje 4.3), u nastavku rada se za potrebe predviđanja potrošnje goriva sintetičkih voznih ciklusa koristi neregularizirani model.



Slika 4-21: Rezultati regularizacije NN-H3D modela: a) krivulje pogreške učenja i validacije i b) razlika pogreške učenja i validacije u ovisnosti o vrijednosti faktora regularizacije λ. Obojana skala (ΔJ_V) označava porast validacijske pogreške u odnosu na vrijednost dobivenu bez regularizacije.

5. Višekriterijska validacija voznih ciklusa

Ovo poglavlje bavi se validacijom višedimenzionalnih sintetičkih voznih ciklusa, generiranih u poglavlju 3 primjenom stohastičke metode sinteze temeljene na Markovljevim lancima. Najprije se u potpoglavlju 5.1 nominira te za svaki snimljeni i sintetički vozni ciklus izračunava bogati skup statističkih značajki, koje uz opće statističke indekse uključuje značajke frekvencijske domene i unakrsne korelacije brzine vozila, ubrzanja vozila i nagiba ceste. Također se s ciljem dodatne potvrde valjanosti predložene metode sinteze voznih ciklusa provodi usporedna analiza razdioba nominiranih statističkih značajki sintetičkih i snimljenih voznih ciklusa. Zatim se u potpoglavlju 5.2 za potrebe jednoznačne validacije voznih ciklusa izvode različiti združeni pokazatelji reprezentativnosti, koji se temelje na kombiniranju individualnih statističkih značajki ili na indeksu sličnosti dviju matrica, konkretnije, TPM-ova i histograma H3D za sintetičke i snimljene vozne cikluse. Nadalje, provodi se korelacijska analiza izračunatih vrijednosti razmatranih pokazatelja reprezentativnosti voznih ciklusa s obzirom na odstupanje potrošnje goriva predviđene NN-H3D neuronskom mrežnom iz poglavlja 4, kako bi se otkrilo koji pokazatelji reprezentativnosti su najprikladniji za validaciju sintetičkih voznih ciklusa. Konačno, u potpoglavlju 5.3 predstavljen je i ispitan postupak izdvajanja reprezentativnih sintetičkih voznih ciklusa, koji se temelji na višestrukim kriterijima.

5.1. Statističke značajke voznih ciklusa

5.1.1. Nominiranje početnog skupa statističkih značajki

Za potrebe opisivanja različitih obrazaca vožnje nominiran je bogat skup od ukupno 100 statističkih značajki voznih ciklusa (dalje označavan kao $\mathcal{F}$), razvrstanih u sedam glavnih skupina koje se odnose na brzinu vozila, ubrzanje vozila, nagib ceste, karakteristike rute, dinamiku vožnje, unakrsnu korelaciju brzine vozila, ubrzanja vozila i nagiba ceste, te značajke frekvencijskog odziva (vidi Prilog A). Za izračun odgovarajućih statističkih značajki, korišteni su sljedeći diskretni vremenski profili snimljenih podataka (gdje operator $\circ$ označava Hadamadrov produkt, odnosno množenje po elementima vektora):

- brzina vozila $\boldsymbol{v}$,
- ubrzanje vozila $\boldsymbol{a}$,
- nagib ceste $\boldsymbol{\theta}$,
- horizontalna komponenta brzine vozila $\boldsymbol{v}_h = \boldsymbol{v} \circ cos(\boldsymbol{\theta})$,
- vertikalna komponenta brzine vozila $\boldsymbol{v}_v = \boldsymbol{v} \circ sin(\boldsymbol{\theta})$,

- zahtijevana snaga na kotaču vozila $\boldsymbol{P}_w = \boldsymbol{F}_w \circ \boldsymbol{v}$,
- specifična snaga vozila $\boldsymbol{P}_m \cong \boldsymbol{v} \circ \boldsymbol{a}$.

Pritom se ukupna sila na kotaču vozila $\boldsymbol{F}_w$ izračunava pomoću općeg modela uzdužne dinamike vozila danog u izrazu (3-6), i ovisna je o tipu (parametrima) vozila. Budući da podaci za MAN Lion City autobus koji prometuje na razmatranoj ruti u Dubrovniku nisu dostupni, za računanje $\boldsymbol{F}_w$ koriste se parametri sličnog Volvo 7900 konvencionalnog (Diesel) autobusa dani u tablici 5-1.

Tablica 5-1: Parametri Volvo 7900 konvencionalnog autobusa.

Parametar	Vrijednost
Masa praznog vozila, m_v	12.635 kg
Koeficijent otpora kotrljanja, R_0	0.012
Koeficijent aerodinamičkog otpora, C_d	0.7
Čeona površina vozila, A_f	7.52 m^2

Statističke značajke $\mathcal{F}$ dalje se dijele prema tome jesu li izvedene iz vremenskih profila ili frekvencijskih odziva razmatranih varijabli. Cjelovit popis razmatranih statističkih značajki voznih ciklusa iz vremenske domene (ukupno 77 značajki) i frekvencijske domene (ukupno 23 značajki) s pripadnim objašnjenjima dan je u Prilogu A.

5.1.2. Usporedna analiza razdioba statističkih značajki

Važan predkorak validacije voznih ciklusa uključuje usporednu analizu razdioba nominiranih statističkih značajki definiranih u odjeljku 5.1.1 (i sadržanih u skupu $\mathcal{F}$) za slučajeve sintetičkih voznih ciklusa sadržanih u skupu $\Psi_R^S \rightarrow \boldsymbol{f}_S \in R^{N_S^R \text{ x } M}$ i snimljenih voznih ciklusa s finom razlučivosti potrošnje goriva sadržanih u skupu $\Psi_Z \rightarrow \boldsymbol{f}_Z \in R^{N_{DC.F}^V \text{ x } M}$, koji isključuju kategorizaciju po stupnju zagušenosti prometa (za definiciju voznih ciklusa vidi poglavlje 3) [52]. U tu svrhu, iz svake od pripadnih kategorija statističkih značajki odabrana je po jedna reprezentativna značajka, čime se dobiva skup od osam statističkih značajki $\mathcal{F}_R$:

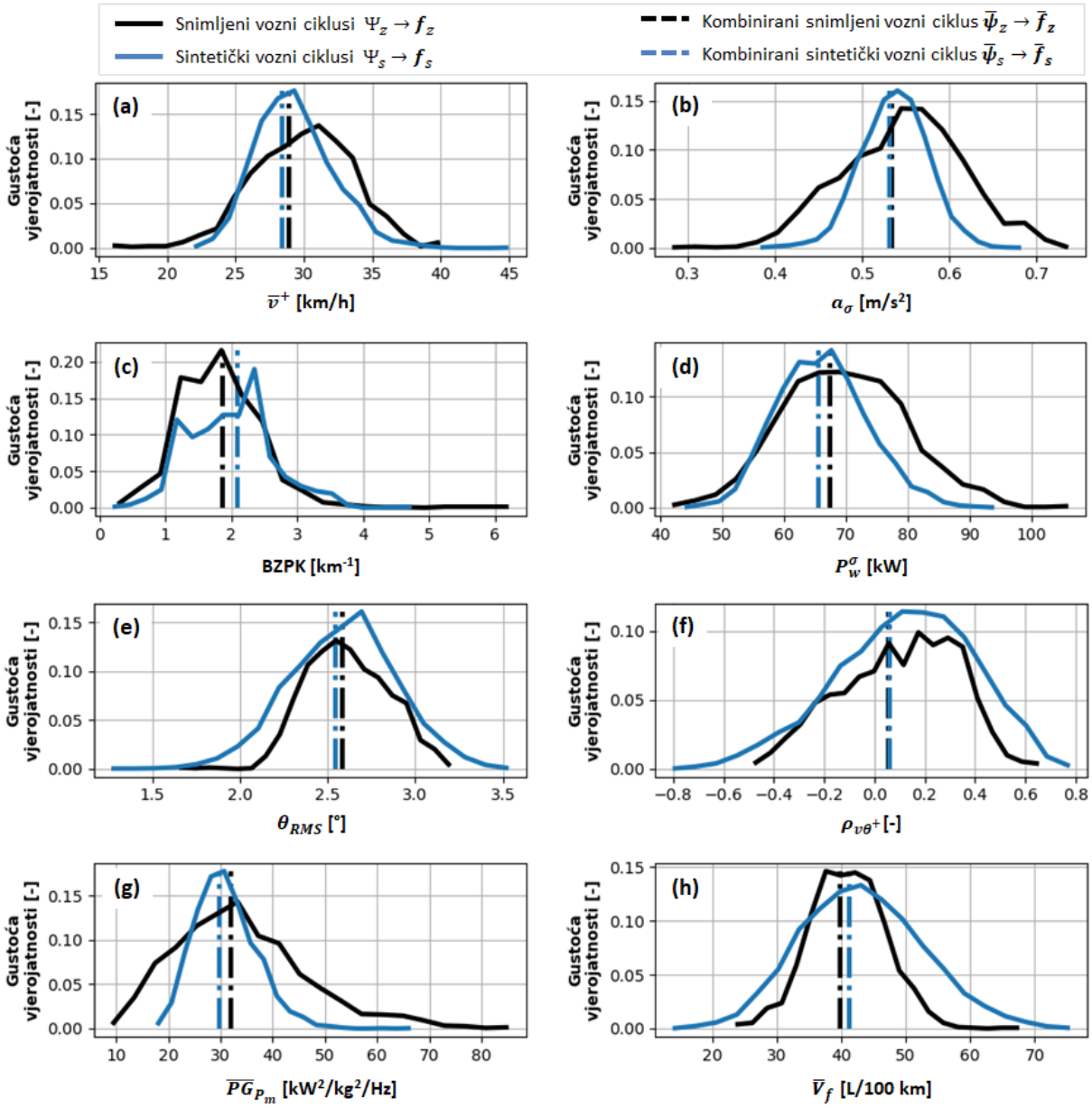
1) srednja pozitivna brzina vozila $\bar{v}^+$ (km/h),
2) standardna devijacija ubrzanja vozila a_σ (m/s^2),
3) broj zaustavljanja vozila po prijeđenom kilometru BZPK (km^{-1}),
4) standardna devijacija zahtijevane snage na kotaču vozila P_w^σ (kW),
5) RMS nagiba ceste θ_{RMS} (°),
6) Korelacija brzine vozila i pozitivnog nagiba ceste $\rho_{v\theta^+}$ (-),

7) Srednja vrijednost frekvencijskog odziva (periodograma) specifične snage vozila $\overline{PG}_{P_m}$ (kW^2/kg^2/Hz),
8) Prosječna potrošnja goriva $\bar{V}_f$ (L/100 km), koja se za sintetičke vozne cikluse predviđa temeljem NN-H3D.

Slika 5-1 prikazuje razdiobe vjerojatnosti statističkih značajki $\mathcal{F}_R$ dobivenih za slučaj snimljenih voznih ciklusa $\boldsymbol{f}_z \in R^{N_{DC.F}^V \text{ x } 8}$ (crna linija) i generiranih sintetičkih voznih ciklusa $\boldsymbol{f}_s \in R^{N_s^R \text{ x } 8}$ (plava linija), uključujući očekivane vrijednosti odgovarajućih statističkih značajki izračunatih za kombinirane vozne cikluse $\bar{\boldsymbol{\psi}} \rightarrow \bar{\boldsymbol{f}}$ (isprekidana linija) dobivene spajanjem svih pojedinačnih snimljenih i sintetičkih voznih ciklusa (njih redom 1527 i 3000). Rezultati na slici 5-1 pokazuju kako se razdiobe i očekivanja odabranih statističkih značajki $\mathcal{F}_R$ sintetičkih i snimljenih voznih ciklusa dobro podudaraju. Nadalje, za potrebe daljnje usporedbe izračunavaju se relativna odstupanja očekivanih vrijednosti razmatranih statističkih značajki ($m = 1,2, \ldots ,8$) prema

$$\Delta f_{rel,m} = 100 \cdot \frac{\bar{f}_{s,m} - \bar{f}_{z,m}}{\bar{f}_{z,m}} \text{ [\%]}, \tag{5-2}$$

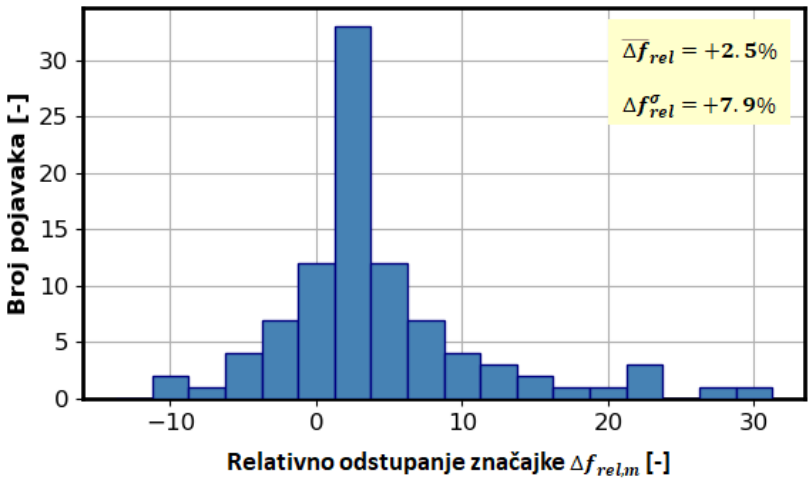
te se koriste za određivanje prosječnog relativnog odstupanja statističkih značajki $\overline{\Delta f_{rel}}$ i pripadne standardne devijacije Δf_{rel}^{σ}. Tablica 5-2 pokazuje da se sve $\Delta f_{rel,m}$ izračunate za skup značajki $\mathcal{F}_R$ nalaze unutar približno ±10% odstupanja, dok $\overline{\Delta f_{rel}}$ i Δf_{rel}^{σ} iznose redom -0.76% i +5.41%. Ukoliko se razmatra čitav skup nominiranih statističkih značajki $\mathcal{F}$ (ukupno $M = 100$ značajki), dobiva se razdioba dana na slici 5-2, pri čemu $\overline{\Delta f_{rel}}$ poprima vrijednost +2.5%, dok Δf_{rel}^{σ} iznosi +7.9%. Ovi rezultati potvrđuju valjanost metode sinteze voznih ciklusa predložene u poglavlju 3.



Slika 5-1: Razdiobe vjerojatnosti za odabrani skup statističkih značajki $\mathcal{F}_R$ izračunatih za snimljene i sintetičke vozne cikluse, zajedno s očekivanim vrijednostima odgovarajućih kombiniranih voznih ciklusa.

Tablica 5-2: Očekivane vrijednosti statističkih značajki iz skupa osam odabranih značajki $\mathcal{F}_R$, izračunate za kombinirani snimljeni $\bar{f}_{z,m}$ i sintetički $\bar{f}_{s,m}$ vozni ciklus, zajedno s pripadnim relativnim odstupanjima značajki $\Delta f_{rel,m}$.

Redni broj, *m*	**Statistička značajka**	$\bar{f}_{z,m}$	$\bar{f}_{s,m}$	$\Delta f_{rel,m}$
1	Srednja pozitivna brzina vozila $\bar{v}^+$ (km/h)	28.89	28.44	-1.6%
2	Standardna devijacija ubrzanja vozila a_σ (m/s^2)	0.535	0.531	-0.7%
3	Broj zaustavljanja vozila po prijeđenom kilometru BZPK (km^{-1})	1.87	2.09	+11.9%
4	Standardna devijacija zahtijevane snage na kotaču vozila P_W^σ (kW)	67.46	65.67	-2.7%
5	RMS nagiba ceste θ_{RMS} (°)	2.59	2.54	-1.6%
6	Korelacija brzine vozila i pozitivnog nagiba ceste $\rho_{v\theta^+}$ (-)	0.0055	0.059	+7%
7	Srednja vrijednost frekvencijskog odziva (periodograma) specifične snage vozila $\overline{PG}_{P_m}$ (kW^2/kg^2/Hz)	32.01	29.72	-7.2%
8	Prosječna potrošnja goriva $\bar{V}_f$ (L/100 km)	39.84	41.23	+3.5%
	$\overline{\Delta f_{rel}}$ (Δf_{rel}^σ)	-0.76% (+5.41%)		



Slika 5-2: Histogram relativnih odstupanja očekivanih vrijednosti svih 100 razmatranih statističkih značajki ($\Delta f_{rel,m}$) za kombinirani sintetički u odnosu na snimljeni vozni ciklus, zajedno s pripadnom prosječnom vrijednosti ($\overline{\Delta f_{rel}}$) i standardne devijacije (Δf_{rel}^σ).

5.2. Definiranje pokazatelja reprezentativnosti voznog ciklusa

U svrhu jednoznačne validacije voznih ciklusa izvodi se nekoliko združenih pokazatelja reprezentativnosti voznog ciklusa iz širokog skupa statističkih značajki razmatranog u odjeljku

5.1.1. Cilj je sublimirati nominirane statističke značajke sadržane u skupu $\mathcal{F}$ u jedan pokazatelj reprezentativnosti tj. metriku koja određuje koliko blisko određeni sintetički vozni ciklus slijedi srednje/očekivano ponašanje snimljenog skupa voznih ciklusa. Također, izvode se jedinstveni pokazatelji reprezentativnosti koji se temelje na indeksu sličnosti dviju matrica. Definirani pokazatelji reprezentativnosti uspoređuju se s pokazateljem vezanim uz potrošnju goriva predviđenu korištenjem neuronske mreže iz poglavlja 4, kako bi se otkrilo koji je pokazatelj najprikladniji za validaciju voznih ciklusa.

5.2.1. Matematička formulacija pokazatelja reprezentativnosti voznog ciklusa

Za potrebe validacije sintetičkih voznih ciklusa razmatraju se sljedeća pet pokazatelja reprezentativnosti:

1) Odstupanje potrošnje goriva (*engl. Fuel Consumption Deviation*; FCD), koje se definira kao apsolutna vrijednost odstupanja potrošnje goriva sintetičkog voznog ciklusa $V_{f,s,i}$, predviđene korištenjem *NN-H3D* modela opisanog u poglavlju 4, u odnosu na potrošnju goriva izračunatu za kombinirani snimljeni vozni ciklus $\bar{\boldsymbol{\psi}}_z \rightarrow \bar{V}_{f,z}$

$$\mathrm{FCD}_i = \left|V_{f,s,i} - \bar{V}_{f,z}\right|; \quad i = 1,2,\dots,N, \tag{5-3}$$

gdje i predstavlja *i*-ti sintetički vozni ciklus (od ukupno $N \equiv N_s^R = 3000$ generiranih).

2) Srednja apsolutna pogreška (*engl. Mean Absolute Error*, MAE)

$$\mathrm{MAE}_i = \frac{1}{M} \sum_{m=1}^{M} \left|f_{s,i,m} - \bar{f}_{z,m}\right| \tag{5-4}$$

gdje $\bar{f}_{z,m}$ predstavlja vrijednost *m*-te statističke značajke za kombinirani snimljeni vozni ciklus $\bar{\boldsymbol{\psi}}_z$, dok $f_{s,i,m}$ predstavlja vrijednost iste statističke značajke za *i*-ti sintetički vozni ciklus. Pritom se razmatra cjelokupni skup $\mathcal{F}$ od M = 100 statističkih značajki, čije su vrijednosti prethodno skalirane na raspon od 0 do 1 prema principu min-max normalizacije kako bi se poništio utjecaj različitih raspona vrijednosti raznih značajki.

3) Euklidska udaljenost (*engl. Euclidean Distance*, ED)

$$\mathrm{ED}_i = \sqrt{\sum_{m=1}^{M} (f_{s,i,m} - \bar{f}_{z,m})^2}, \tag{5-5}$$

koja snažnije naglašava veća odstupanja statističkih značajki u odnosu na MAE.

4) Indeks sličnosti (*engl. Similarity Index*, SI) između dva vektora (ili poravnate matrice), koji je definiran kao kosinusova sličnost [125] prema

$$\mathrm{SI}_i(\mathbf{A}_i, \mathbf{B}) = \frac{\sum_{j=1}^{K} A_{i,j} B_j}{\sqrt{\sum_{j=1}^{K} A_{i,j}^2} \sqrt{\sum_{j=1}^{K} B_j^2}}, \tag{5-6}$$

gdje su $A_{i,j}$ i B_j komponente ulaznih vektora $\mathbf{A}_i$ i $\mathbf{B}$, a K je ukupan broj elemenata u pripadnim vektorima ($\mathbf{A}_i, \mathbf{B} \in R^{1\,\mathrm{x}\,K}$). Vrijednost SI kreće se u rasponu od −1 do 1, gdje -1 označava potpunu suprotnost, 0 označava ortogonalnost (nema korelacije), a 1 označava potpunu sličnost datih vektora. Pritom se vektori tj. izvorno matrice $\mathbf{A}_i$ i $\mathbf{B}$ izračunavaju za svaki i-ti sintetički vozni ciklus ($\mathbf{A}_i$) zasebno te kombinirani snimljeni vozni ciklus $\bar{\boldsymbol{\psi}}_z$ ($\mathbf{B}$), razmatrajući sljedeća dva tipa transformacije voznog ciklusa:

i. 6D TPM (*ML6D*) definiran izrazom (3-12), u kojem su prebrojani prijelazi između diskretnih stanja brzine i ubrzanja vozila te nagiba ceste,
ii. 3D histogram (*H3D*) prebrojanih diskretnih stanja brzine i ubrzanja vozila te nagiba ceste prikazan slikom 4-6.

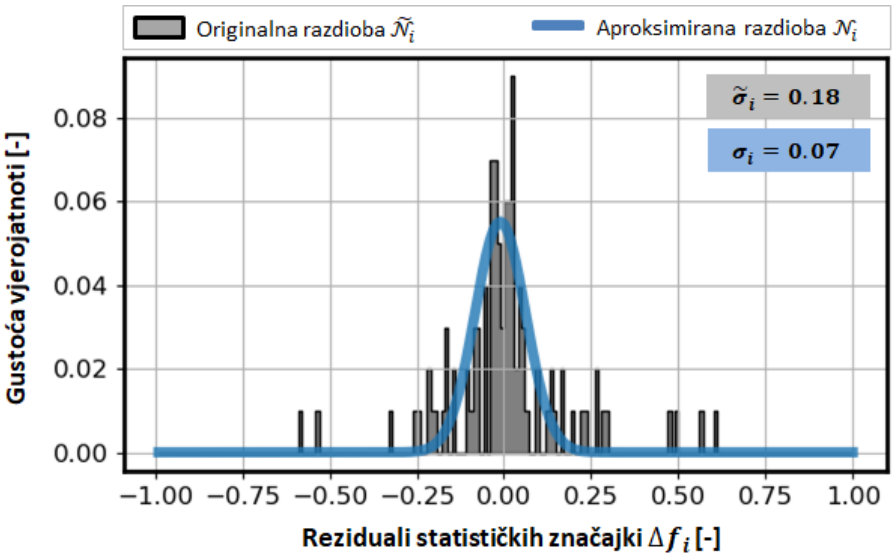
Pritom se za 6D TPM podrazumijevaju jednaki rasponi i razlučivosti diskretnih stanja brzine vozila, ubrzanja vozila i nagiba ceste, kako su definirani za *H3D*. Sličnost TPM-a dalje se označava kao SI-TPM, a *H3D* kao SI-H3D.

5) Konačni pokazatelj reprezentativnosti temelji se na aproksimaciji originalne razdiobe reziduala statističkih značajki $\widetilde{\mathcal{N}}_i(\Delta f_i(m))$ s normalnom (Gaussovom) razdiobom, čija je funkcija gustoće vjerojatnosti $\mathcal{N}_i(\mu_i, \sigma_i)$ formulirana kao

$$\mathcal{N}_i(\Delta f_i(m);\, \mu_i, \sigma_i) = \frac{1}{\sigma_i \sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{\Delta f_i(m) - \mu_i}{\sigma_i}\right)^2}; \quad m = 1,2,\ldots,M, \tag{5-7}$$

gdje μ_i i σ_i predstavljaju očekivanu vrijednost i standardnu devijaciju razdiobe $\mathcal{N}_i$, dok se reziduali statističkih značajki izračunavaju prema $\Delta f_i(m) = f_{s,i}(m) - \bar{f}_z(m)$. Parametri razdiobe $\mathcal{N}_i$ (μ_i, σ_i) uče se s pomoću nelinearne metode najmanjih kvadrata implementirane kroz funkciju *curve_fit* iz *scipy* modula u Pythonu [98]). Pritom se standardna devijacija σ_i aproksimacijske krivulje $\mathcal{N}_i$ usvaja kao relevantni pokazatelj te se dalje označava kao STD. Primjer rezultirajuće aproksimacijske krivulje prikazan je na slici 5-3, gdje je jasno vidljiv njezin učinak na filtriranje velikih odstupanja određenih statističkih značajki na vrijednost standardne devijacije, tj. standardna devijacija aproksimacijske krivulje (σ_i), značajno je niža od one izračunate za razmatrani uzorak ($\tilde{\sigma}_i$).

Pokazatelji reprezentativnosti voznog ciklusa definirani izrazima (5-3)-(5-7) izračunati su za svaki pojedinačni (i-ti) sintetički vozni ciklus te se sustavno analiziraju u sljedećem odjeljku.

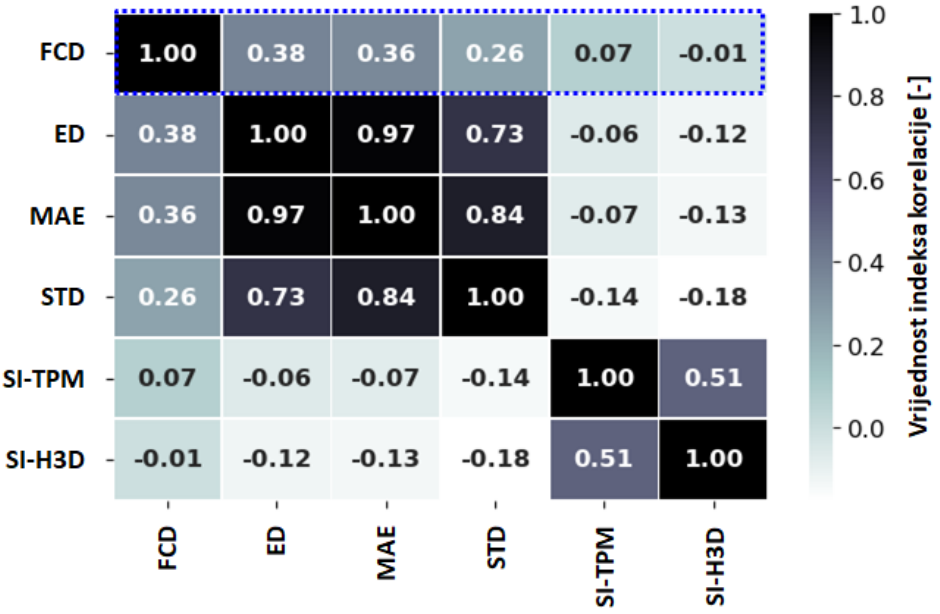


Slika 5-3: Primjer aproksimacijske krivulje definirane normalnom razdiobom $\mathcal{N}_i$ čiji se parametri μ_i, σ_i uče na originalnoj razdiobi reziduala statističkih značajki $\widetilde{\mathcal{N}}_i$.

5.2.2. Izbor najprikladnijih pokazatelja reprezentativnosti za validaciju voznih ciklusa

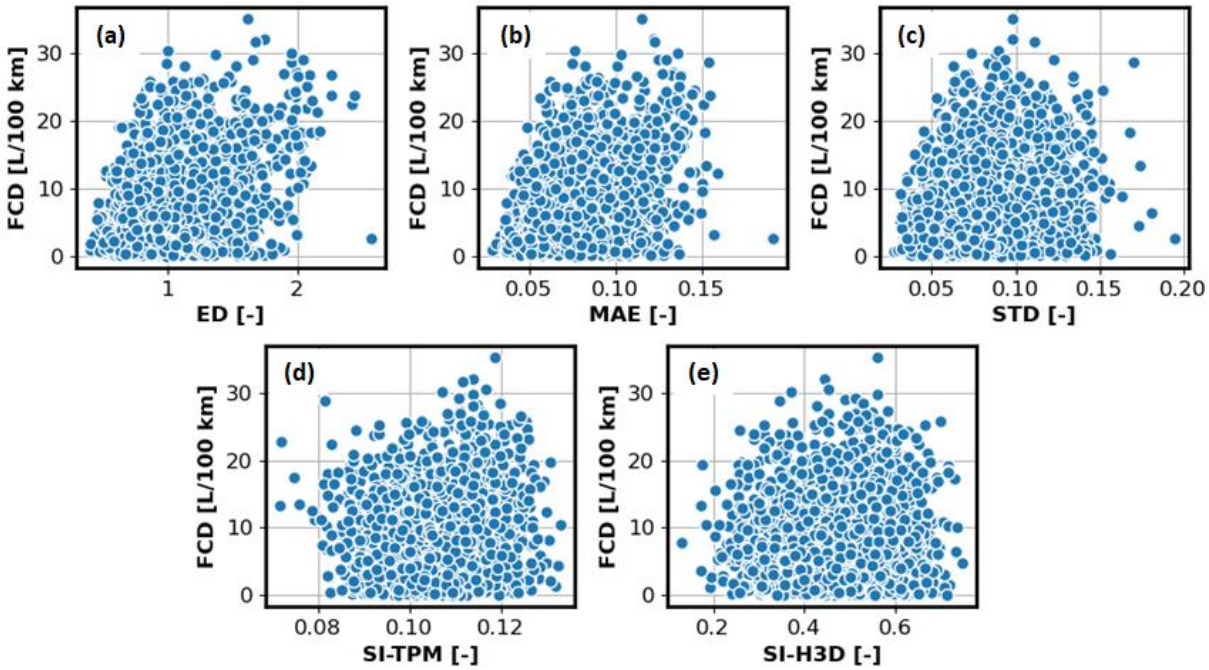
Odabir najprikladnijih pokazatelja reprezentativnosti voznog ciklusa za validaciju sintetičkih voznih ciklusa provodi se korelacijskom analizom svakog pokazatelja (ED, MAE, SI-TPM, SI-H3D, i STD) s obzirom na odstupanje potrošnje goriva FCD. Izračun indeksa korelacije ostvaren je s pomoću funkcije *corrcoef* iz *numpy* modula u Pythonu [93] i za punu korelaciju ima iznos 1.

Slika 5-4 prikazuje rezultirajuću toplinsku mapu korelacijske matrice za sve razmatrane pokazatelje. Vrijednosti indeksa korelacije u padajućem redoslijedu iznose redom 0.38, 0.36, 0.26, 0.07 i -0.01 za pokazatelje ED, MAE, STD, SI-TPM i SI-H3D s obzirom na FCD (prvi redak korelacijske matrice). Stoga se ED, MAE i STD (zajedno s FCD-om) usvajaju kao konačni pokazatelji na temelju kojih se provodi validacija voznih ciklusa. Izrazito visoka vrijednost indeksa korelacije od 0.97 između ED i MAE ukazuje na značajnu sličnost ovih dvaju pokazatelja. Pokazatelji ED i MAE također posjeduju izražene ali nešto različite korelacije s obzirom na STD pokazatelj (indeksi korelacije iznose redom 0.73 i 0.84). U slučaju pokazatelja SI-TPM i SI-H3D vidljiva je samo njihova međusobna korelacija (indeks korelacije jednak 0.51), dok je za ostale pokazatelje prisutna u vrlo niskoj razini (indeks korelacije manji od 0.18), što ukazuje na neprikladnost njihovog korištenja za potrebe validacije voznih ciklusa.



Slika 5-4: Rezultirajuća korelacijska matrica prikazana u obliku toplinske mape za sve razmatrane pokazatelje reprezentativnosti voznog ciklusa (plavom bojom označen je redak korelacije svakog pokazatelja u odnosu na FCD).

Nadalje, kako bi se dobio bolji uvid u međuovisnosti sa slike 5-4, na slici 5-5 prikazana je ovisnost odstupanja potrošnje goriva FCD u ovisnosti o svakom pokazatelju reprezentativnosti za cijeli skup sintetičkih voznih ciklusa. Vizualnom inspekcijom slike 5-5 može se ustanoviti kako se vrijednosti svih pokazatelja reprezentativnosti značajno rasipaju čak i za niske vrijednosti FCD, što objašnjava relativno niske vrijednosti indeksa korelacije na slici 5-4. Iz toga proizlazi da jednoznačna validacija voznih ciklusa prema kriteriju minimalne vrijednosti FCD ne garantira ujedno i reprezentativne vrijednosti statističkih značajki voznog ciklusa (tj. niski ED, MAE, STD ili visoki SI-TPM i SI-H3D). Isto vrijedi i za slučaj jednoznačne validacije voznih ciklusa temeljem kriterija minimalne vrijednosti ED, MAE, STD, SI-TPM i SI-H3D, koja pak ne garantira minimalnu vrijednost FCD. Pritom treba imati u vidu da je primjenom stohastičkog modela voznih ciklusa moguće generirati široki skup sintetičkih voznih ciklusa, pa je moguća „slučajna“ podudarnost minimalne vrijednosti FCD s nekim od pokazatelja reprezentativnosti. No, općenito se može zaključiti kako jednoznačnom validacijom po bilo kojem združenom pokazatelju reprezentativnosti nije moguće izdvojiti najreprezentativniji vozni ciklus. Stoga je potrebno razviti naprednije postupke validacije višedimenzionalnih voznih ciklusa, kako je izloženo u narednom potpoglavlju i poglavlju 6.



Slika 5-5: Ovisnosti pokazatelja FCD o a) ED, b) MAE, c) STD, d) SI-TPM, te e) SI-H3D pokazateljima (svaka točka na grafu odnosi se na jedan sintetički vozni ciklus).

5.3. Izdvajanje reprezentativnih voznih ciklusa

5.3.1. Postupak višekriterijske validacije voznih ciklusa

Višekriterijska metoda validacije voznih ciklusa temelji se na FCD, ED, MAE i STD pokazateljima reprezentativnosti voznog ciklusa, koji su se pokazali kao najbolji kandidati prema korelacijskoj analizi provedenoj u potpoglavlju 5.2. Blokovski dijagram cjelokupne metode višekriterijske validacije voznih ciklusa prikazan je na slici 5-6.

Metoda najprije uključuje izolaciju reprezentativnih podskupova sintetičkih voznih ciklusa *F*, *E*, *D* i *H* s obzirom na svaki razmatrani pokazatelj reprezentativnosti:

$$F = \{\boldsymbol{\psi}_i \mid \mathrm{FCD}_i \leq \varepsilon_{FCD}\} \quad \text{gdje je} \quad F \subseteq \Psi_R^S, \tag{5-8}$$

$$E = \{\boldsymbol{\psi}_i \mid \mathrm{ED}_i \leq \varepsilon_{ED}\} \quad \text{gdje je} \quad E \subseteq \Psi_R^S, \tag{5-9}$$

$$D = \{\boldsymbol{\psi}_i \mid \mathrm{MAE}_i \leq \varepsilon_{MAE}\} \quad \text{gdje je} \quad D \subseteq \Psi_R^S, \tag{5-10}$$

$$H = \{\boldsymbol{\psi}_i \mid \mathrm{STD}_i \leq \varepsilon_{STD}\} \quad \text{gdje je} \quad H \subseteq \Psi_R^S, \tag{5-11}$$

gdje $\boldsymbol{\psi}_i$ predstavlja i-ti kandidirani sintetički vozni ciklus sadržan u skupu svih 3000 generiranih sintetičkih voznih ciklusa $\Psi_R^S = \{\boldsymbol{\psi}_1, \boldsymbol{\psi}_2, \dots, \boldsymbol{\psi}_i\}$, dok su FCD_i, ED_i, i MAE_i pripadne vrijednosti razmatranih pokazatelja reprezentativnosti. Pragovi ε_{FCD}, ε_{ED}, ε_{MAE} i ε_{STD} predstavljaju gornje granične vrijednosti spomenutih pokazatelja reprezentativnosti, koje se određuju pripadnim vrijednostima p-tog percentila. Pritom p-ti percentil označava vrijednost

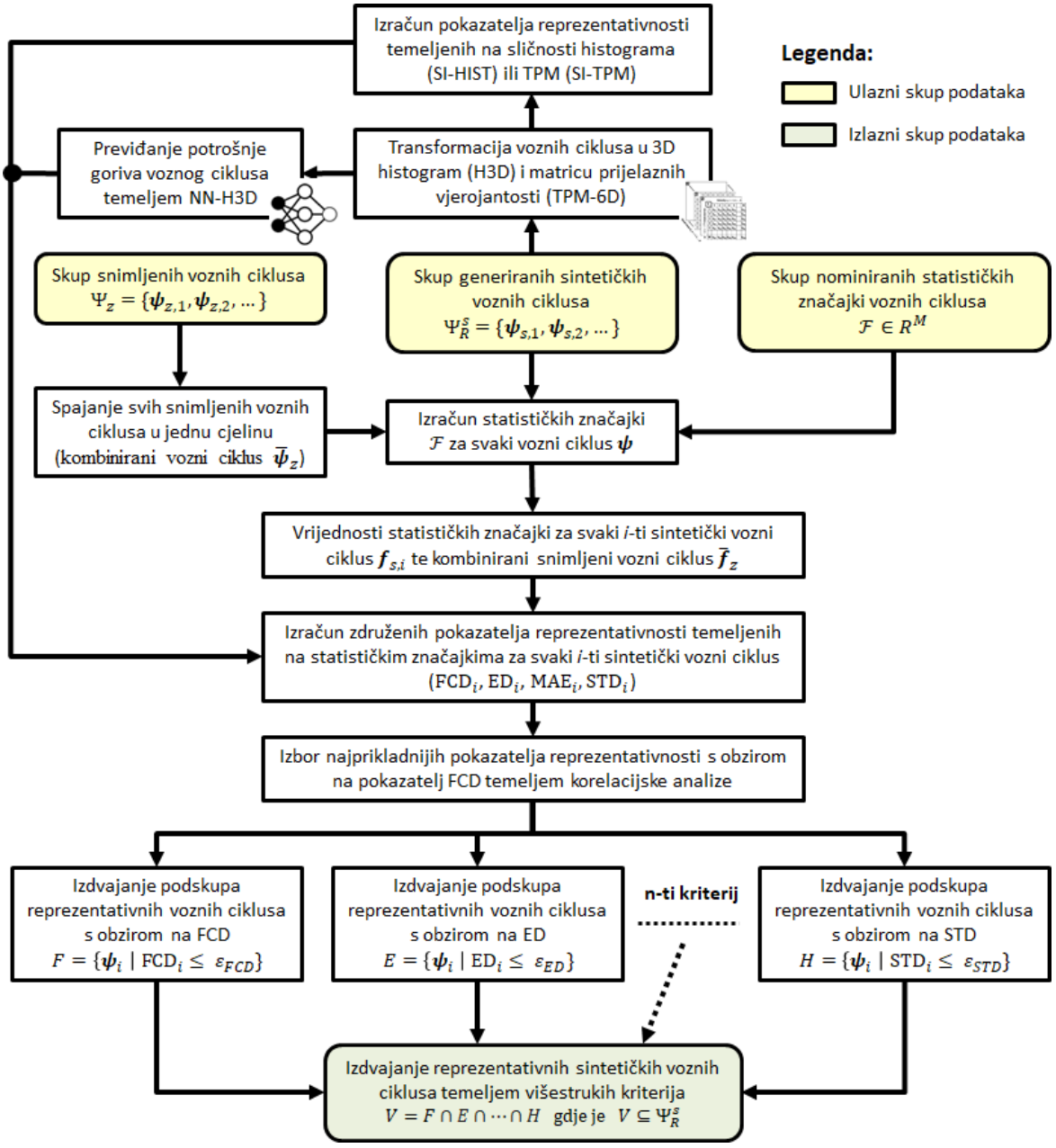
ispod koje pada zadani postotak opažanja u skupini opažanja. Na primjer, 25. percentil je vrijednost ispod koje se može naći 25% opažanja, itd.

Konačni podskup validnih (reprezentativnih) sintetičkih voznih ciklusa V moguće je odrediti kao presjek podskupova F, E, D i H u sljedećim kombinacijama, koje proizlaze iz dodavanja pokazatelja reprezentativnosti u redoslijedu određenim vrijednostima indeksa korelacija s obzirom na FCD

$$V_1 = F \cap E \quad \text{gdje je } V_1 \subseteq \Psi_R^S. \tag{5-12}$$

$$V_2 = F \cap E \cap D \quad \text{gdje je } V_2 \subseteq \Psi_R^S. \tag{5-13}$$

$$V_3 = F \cap E \cap D \cap H \quad \text{gdje je } V_3 \subseteq \Psi_R^S. \tag{5-14}$$



Slika 5-6: Blokovski dijagram postupka višekriterijske validacije voznih ciklusa.

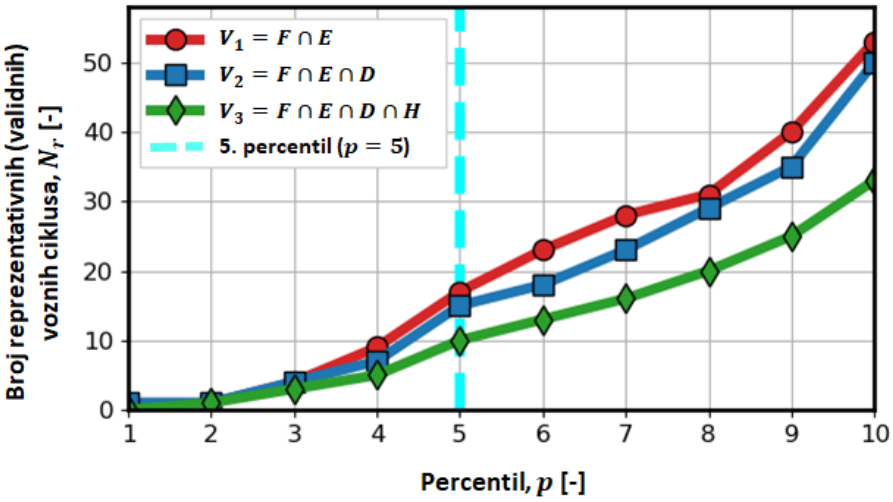
Nadalje, slika 5-7 prikazuje ovisnost broja validnih sintetičkih voznih ciklusa N_r (onih iz skupova V_1, V_2 i V_3) o broju percentila $p \in [0, 10]$ (kojim se definiraju vrijednosti ε_{FCD}, ε_{ED}, ε_{MAE} i ε_{STD}) za različiti broj razmatranih pokazatelja reprezentativnosti. Vidljivo je kako primjena $p \leq 2$ rezultira samo s jednim validnim (najreprezentativnijim) sintetičkim voznim ciklusom ($N_r = 1$), dok za $p > 2$ krivulja $N_r(p)$ počinje poprimati blago progresivni oblik, te se dobiva primjerice $N_r \in [10, 17]$ za $p = 5$ i $N_r \in [33, 53]$ za $p = 10$. Najreprezentativniji sintetički vozni ciklus $|\psi^*|$ za slučaj $V_3 = F \cap E \cap D \cap H$ (vidi izraz (5-14)) moguće je izdvojiti prema kriteriju minimalne udaljenosti FCD_i i ED_i od ishodišta FCD-ED koordinatnog sustava kao

$$|\psi^*| = \arg\min_i \left(\sqrt{FCD_i^2 + ED_i^2} \right). \tag{5-15}$$

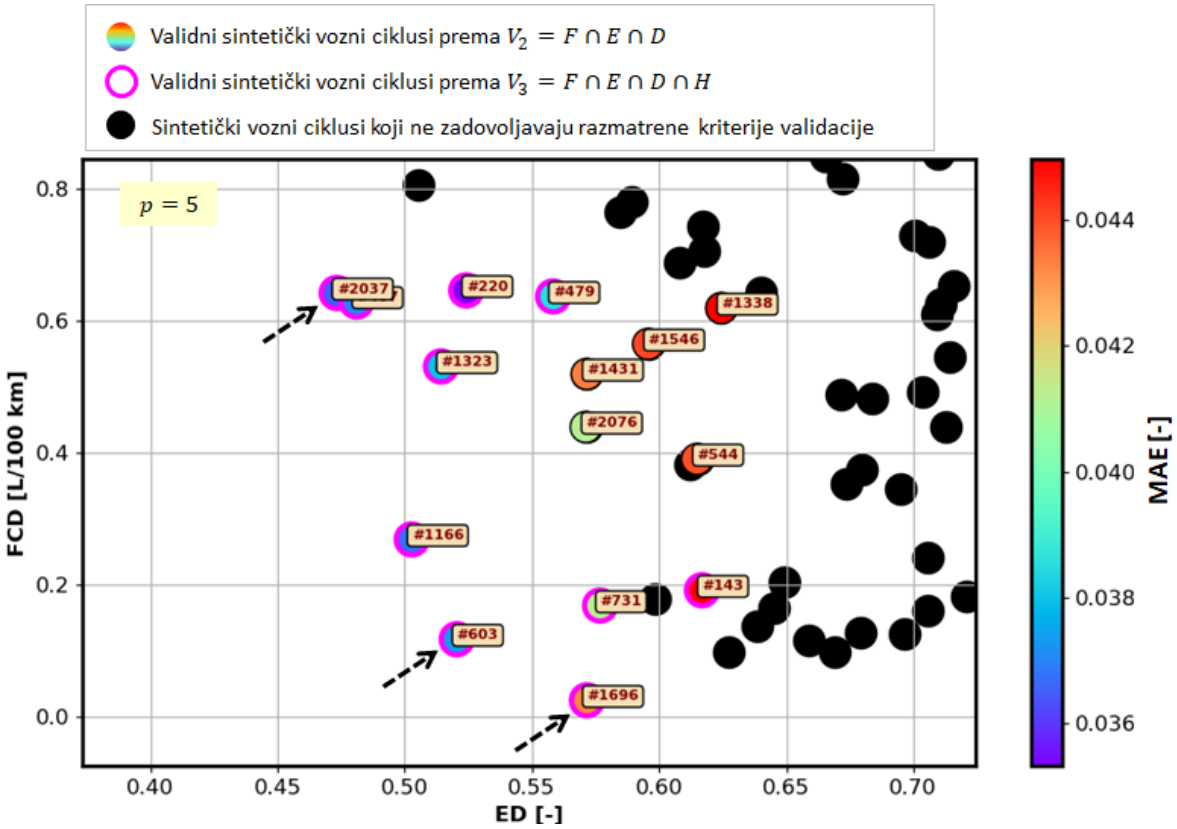
Na sličan način se mogu u korijenski zbog u (5-15) dodati ostali pokazatelji (MAE_i i STD_i).

Konačno usvojene gornje granične vrijednosti za FCD, ED, MAE i STD određene su 5. percentilom ($p = 5$; cijan linija na slici 5-7), a iznose redom $\varepsilon_{FCD} = 0.683$ L/100 km, $\varepsilon_{ED} = 0.627$, $\varepsilon_{MAE} = 0.047$ i $\varepsilon_{STD} = 0.046$. Time je izolirano ukupno $N_r = 15$ reprezentativnih sintetičkih voznih ciklusa za $V_2 = F \cap E \cap D$, čija je ovisnost ED u odnosu na FCD i MAE prikazana na slici 5-8 kao svojevrsna Pareto fronta. Ukoliko se i pokazatelj STD uključi kao dodatni kriterij validacije ($V_3 = F \cap E \cap D \cap H$), dobiva se rezultirajući skup od $N_r = 10$ reprezentativnih sintetičkih voznih ciklusa označenih magenta okvirima na slici 5-8. Crni kružići označavaju sintetičke vozne cikluse koji ne zadovoljavaju razmatrane kriterije validacije, a upadaju u prikazani koordinatni sustav.

U svrhu daljnje analize, s prednje fronte na slici 5-8 izdvojena su tri karakteristična sintetička vozna ciklusa, označena labelama #2037, #603 i #1696 te strelicama. Vozni ciklus #2037 ima najniži ED (0.473), ali najveći FCD (0.643 L/100 km), dok za vozni ciklus #1696 vrijedi suprotno ($\mathrm{ED} = 0.572$ i $\mathrm{FCD} = 0.025$ L/100 km). Vozni ciklus #603 pozicioniran je po sredini fronte, s ED od 0.521 i FCD od 0.118 L/100 km, te predstavlja najreprezentativniji vozni ciklus prema kriteriju (5-15). Pritom se MAE pokazatelj izostavlja jer gotovo u potpunosti korelira s ED (koeficijent korelacije iznosi 0.97; vidi sliku 5-4 te također sliku 5-8). Povrh toga, razmatrani vozni ciklusi #2037, #603 i #1696 također zadovoljavaju kriterij prema STD pokazatelju (spadaju u kružiće obrubljene magenta linijom). Sustavna analiza izdvojenog skupa od tri reprezentativna sintetička vozna ciklusa provodi se u narednom odjeljku.



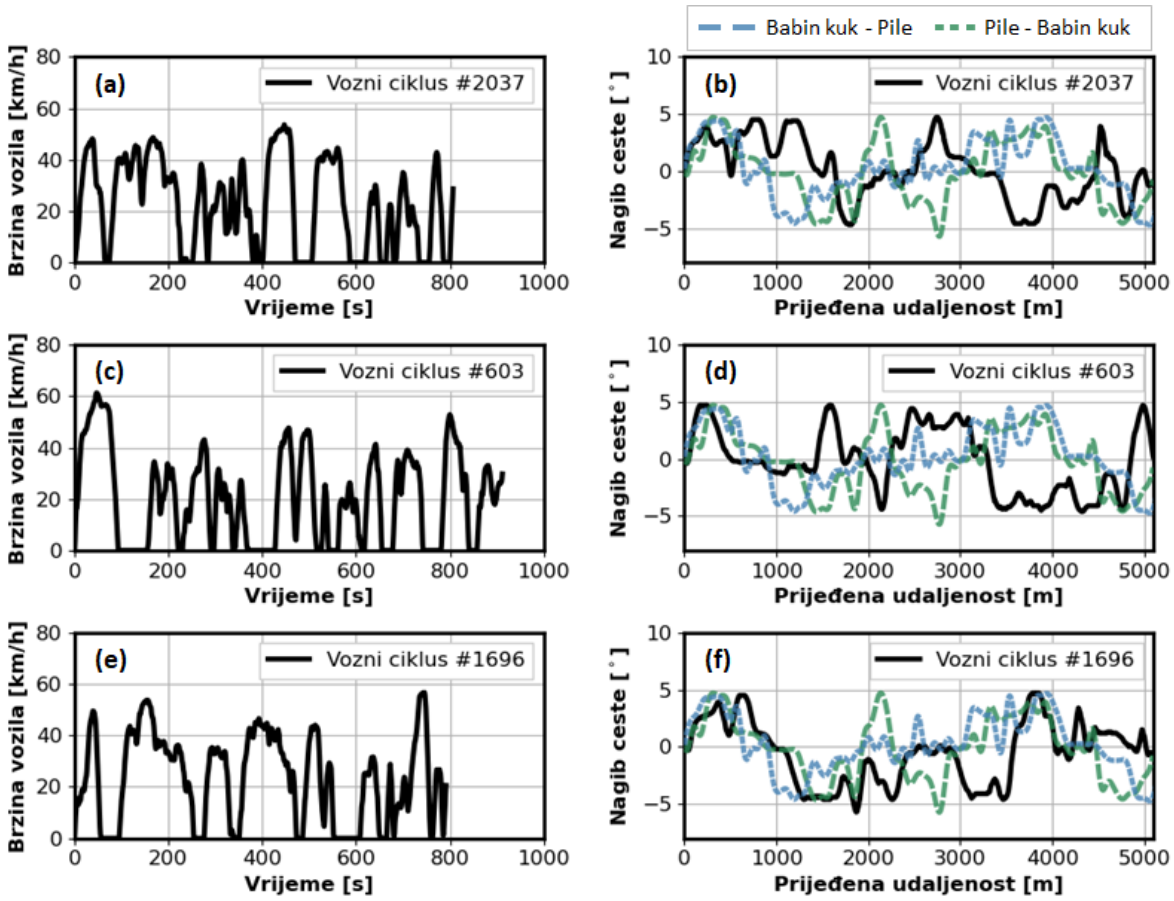
Slika 5-7: Ovisnost broja validnih sintetičkih voznih ciklusa N_r u podskupovima V_1, V_2 te V_3 o graničnim vrijednostima pokazatelja reprezentativnosti definiranim p-tim percentilom.



Slika 5-8: Konačni podskup V_2 od $N_r = 15$ reprezentativnih sintetičkih voznih ciklusa (obojani simboli) za $p = 5$ te uži skup $N_r = 10$ za slučaj V_3 (simboli s magenta obrubom), dobiveni primjenom višekriterijske metode validacije voznih ciklusa s obzirom na FCD, ED, MAE te STD pokazatelje reprezentativnosti.

5.3.2. Analiza reprezentativnih voznih ciklusa

Rezultirajući vremenski profili brzine vozila za reprezentativne vozne cikluse #2037, #603 i #1696 izdvojene višekriterijskom validacijom u prethodnom odjeljku prikazani su na slikama 5-9a, 5-9c i 5-9e, dok su odgovarajući profili nagiba ceste u ovisnosti o prijeđenoj udaljenosti prikazani na slikama 5-9b, 5-9d i 5-9f (zajedno sa snimljenim profilima nagiba ceste; usp. sliku 2-7). Ovdje valja imati na umu kako postupak sinteze rezultira različitim profilima nagiba ceste za svaki sintetički vozni ciklus zbog stohastičke prirode metode njihova generiranja. Neovisno o tome, svaki sintetički profil nagiba ceste vjerno zadržava statistička svojstva snimljenih profila nagiba ceste, uključujući unakrsnu korelaciju s brzinom i ubrzanjem vozila.



Slika 5-9: a), c), e) Profili brzine vozila u vremenu te b), d), e) nagiba ceste u ovisnosti o prijeđenoj udaljenosti za tri višekriterijskom validacijom-izdvojena reprezentativna sintetička vozna ciklusa označena kao #2037, #603 i #1696 na slici 5-8, uključujući snimljene profile nagiba ceste za oba smjera vožnje (označene isprekidanim crtama).

Odstupanja vrijednosti odabranih statističkih značajki razmatranih reprezentativnih voznih ciklusa od odgovarajućih srednjih vrijednosti snimljenog kombiniranog voznog ciklusa $\bar{\boldsymbol{\psi}}_z$ dane su u tablici 5-3, pri čemu su relativna odstupanja dana u zagradama. Budući da je

nominirani skup statističkih značajki $\mathcal{F}$ voznih ciklusa vrlo bogat (tj. sadrži ukupno 100 značajki), usporedba se provodi samo u odnosu na limitirani skup $\mathcal{F}_R$ od osam odabranih značajki, koji je definiran u odjeljku 5.1.2, a sadrži po jednu reprezentativnu statističku značajku iz svake od pripadnih kategorija. Detaljnija usporedba dana je u poglavlju 6.

Tablica 5-3: Vrijednosti odabranih statističkih značajki iz limitiranog skupa $\mathcal{F}_R$ za kombinirani snimljeni vozni ciklus $\bar{\boldsymbol{\psi}}_z$ i odabrane reprezentativne sintetičke vozne cikluse označene kao #2037, #603 i #1696.

Statistička značajka	$\bar{\boldsymbol{\psi}}_z$	**Reprezentativni sintetički vozni ciklus**		
		#2037	**#603**	**#1696**
Srednja pozitivna brzina vozila $\bar{v}^+$ (km/h)	28.89	27.84 (-3.65%)*	27.41 (-5.15%)	28.70 (-0.68%)
Standardna devijacija ubrzanja vozila a_σ (m/s^2)	0.53	0.54 (+1.03%)	0.54 (+0.03%)	0.52 (-3.20%)
Broj zaustavljanja vozila po prijeđenom kilometru BZPK (km^{-1})	1.87	1.96 (+4.79%)	1.95 (+4.71%)	1.37 (-26.61%)
Standardna devijacija zahtijevane snage na kotaču vozila P_w^σ (kW)	67.46	65.79 (-2.48%)	63.02 (-6.59%)	66.66 (-1.19%)
RMS nagiba ceste θ_{RMS} (°)	2.59	2.63 (+1.69%)	2.66 (+2.67%)	2.53 (-2.33%)
Korelacija brzine vozila i pozitivnog nagiba ceste $\rho_{v\theta^+}$ (-)	0.06	-0.03	0.18	-0.15
Srednja vrijednost frekvencijskog odziva (periodograma) specifične snage vozila $\overline{PG}_{P_m}$ (kW^2/kg^2/Hz)	32.01	28.00 (-12.55%)	25.66 (-19.84%)	27.82 (-13.09%)
Prosječna potrošnja goriva $\bar{V}_f$ (L/100 km)	39.84	39.20 (-1.61%)	39.96 (+0.30%)	39.82 (-0.06%)

* Vrijednosti u zagradama predstavljaju relativna odstupanja u odnosu na snimljene vrijednosti $\bar{\boldsymbol{\psi}}_z$, čije su minimalne vrijednosti označene zelenom bojom.

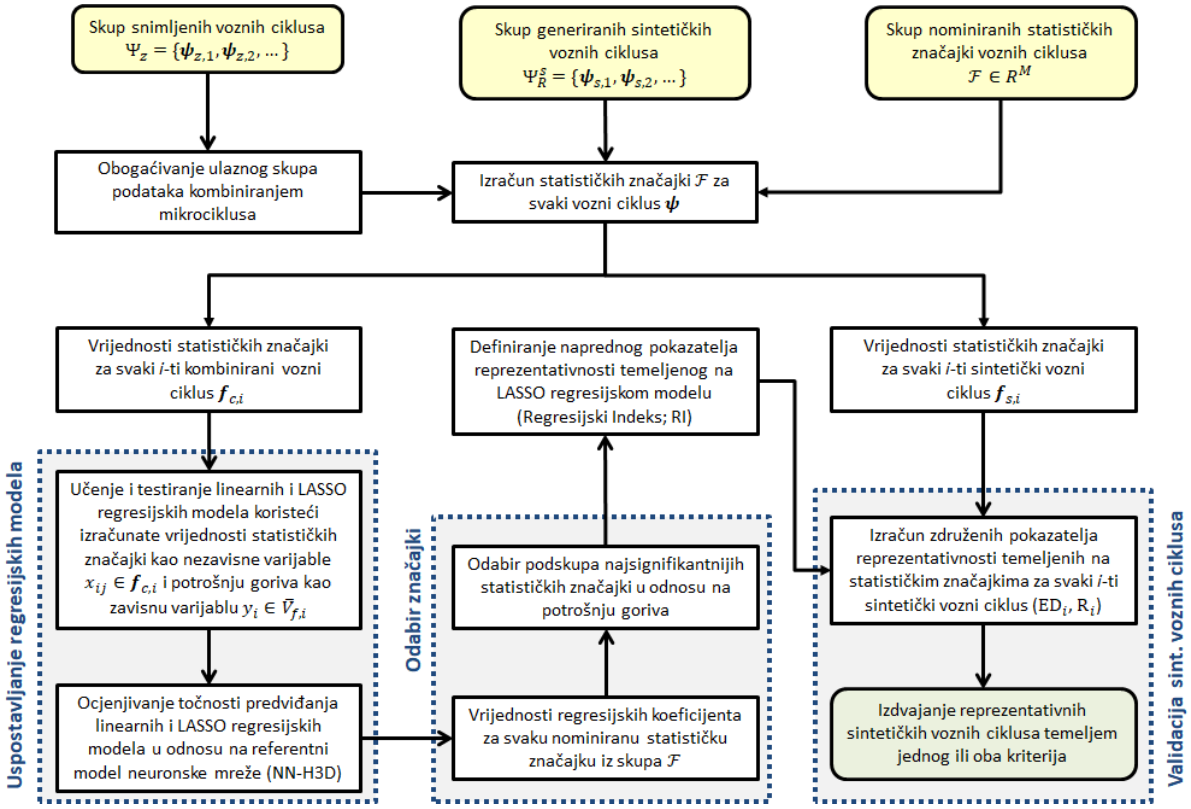
Vizualnom inspekcijom slika 5-9a, 5-9c i 5-9e može se ustanoviti kako razmatrani reprezentativni vozni ciklusi imaju međusobno slične vremenske profile brzine, koje se odražavaju kroz bliske vrijednosti srednje brzine vozila $\bar{v}^+$, te broja zaustavljanja vozila po prijeđenom kilometru BZPK (tablica 5-3). Maksimalna brzina vozila doseže najveću vrijednost od 61.3 km/h za vozni ciklus #603, te tome bliske vrijednosti za vozne cikluse #2037 i #1696 (redom 53.6 km/h i 56.7 km/h). Prosječna vremena mirovanja prilikom zaustavljanja vozila iznose redom 18.2 s, 26.4 s te 19.2 za vozne cikluse #2037, #603 i #1696, dok srednja vrijednost za $\bar{\boldsymbol{\psi}}_z$ iznosi 20.3 s. Nadalje, rezultati na slikama 5-8b, 5-8d i 5-8f pokazuju kako razmatrani sintetizirani profili nagiba ceste u velikoj mjeri nalikuju na snimljene profile, što je potkrijepljeno niskim relativnim odstupanjima RMS nagiba ceste θ_{RMS} (<3%; tablica 5-3). Generalno gledano, odstupanja/reziduali vrijednosti svih razmatranih statističkih značajki za

kandidirane vozne cikluse vrlo su bliske srednjim vrijednostima snimljenog voznog ciklusa $\bar{\boldsymbol{\psi}}_z$ (tj. većina relativnih odstupanja statističkih značajki $\Delta f_{rel,m}; m = 1,2, \dots 8$ nalazi se unutar ±10% tolerancije; tablica 5-3). Maksimalno odstupanje od -26.6% registrirano je za sintetički vozni ciklus #1696, kojeg karakterizira osjetno manji broj zaustavljanja vozila po prijeđenom kilometru (BZPK od 1.37 km^{-1} naspram 1.87 km^{-1} za $\bar{\boldsymbol{\psi}}_z$; tablica 5-3).

Iako vozni ciklus #603, pozicioniran po sredini fronte na slici 5-8, predstavlja najreprezentativniji vozni ciklus prema kriteriju (5-15), i svi preostali, brojčanom labelom označeni reprezentativni vozni ciklusi mogu se smatrati jednako vrijednima, ovisno kojem kriteriju tj. pokazatelju (FCD, ED, MAE ili STD) se želi dati veći značaj.

6. Validacija voznih ciklusa podržana odabirom značajki

Iz perspektive validacije voznih ciklusa, važno je saznati koje su statističke značajke iz nominiranog skupa $\mathcal{F}$ s ukupno 100 značajki (poglavlje 5) najutjecajnije u odnosu na zavisnu/mjerenu varijablu, a to je potrošnja goriva V_f. Stoga se u ovom poglavlju najprije određuje relevantnost svake statističke značajke na temelju linearne regresijske analize, odnosno tehnike odabira značajki temeljene na operateru najmanjeg apsolutnog skupljanja i odabira (LASSO). Povrh toga, predlaže se metoda validacije sintetičkih voznih ciklusa podržana odabirom značajki, koja se temelji na jedinstvenom pokazatelju reprezentativnosti voznog ciklusa određenim s odabranim podskupom najsignifikantnijih statističkih značajki (umjesto sa čitavim skupom od 100 nominiranih statističkih značajki). Metoda je opisana blokovskim dijagramom prikazanim na slici 6-1.



Slika 6-1: Blokovski dijagram predloženog postupka validacije voznih ciklusa temeljenog na odabiru značajki.

U potpoglavlju 6.1 prvo se predlažu linearni i LASSO regresijski modeli reduciranog reda koji koriste samo nekoliko najsignifikantnijih statističkih značajki kao ulaz i potrošnju goriva kao izlaz, te čija se točnost predviđanja potrošnje goriva uspoređuje s referentnim NN-H3D modelom definiranim u poglavlju 4. Zatim se u potpoglavlju 6.2. predstavlja postupak

izdvajanja reprezentativnih sintetičkih voznih ciklusa na temelju LASSO predviđene potrošnje goriva, te se uspoređuje s pristupom temeljenim na Euklidskoj udaljenosti odstupanja statističkih značajki od ciljanih vrijednosti iz poglavlja 5. S ciljem daljnjeg unaprjeđenja postupka jednoznačne validacije sintetičkih voznih ciklusa dodatno se ispituje dvo-kriterijska metoda validacije temeljena na dva spomenuta pristupa/kriterija.

6.1. Izdvajanje podskupa signifikantnih statističkih značajki

6.1.1. Formulacija linearnog regresijskog modela

Za potrebe izdvajanje podskupa signifikantnih statističkih značajki najprije se razmatra jednostavan linearni regresijski (*LR*) model, koji pretpostavlja linearni odnos između zavisne varijable y_i i nezavisnih varijabli (prediktora) x_{ij} [54]

$$y_i = \beta_0 + \beta_1 x_{i1} + \cdots + \beta_j x_{ij} = \sum_j \beta_j x_{ij}\,; \quad i = 1,2,\dots,N; \quad j = 1,2,\dots,M, \tag{6-1}$$

$$y_i \sim N(\mu_i, \sigma^2); \quad i = 1,2,\dots,N; \tag{6-2}$$

gdje je β_0 posmak, $\beta_1, \beta_2, \dots, \beta_j$ parametri regresijskog modela, μ_i i σ^2 su redom prosječna vrijednost i varijanca razdiobe zavisne varijable y_i, *N* je ukupni broj opažanja u snimljenim podacima, a *M* broj nominiranih statističkih značajki voznog ciklusa iz skupa $\mathcal{F}$.

Model (6-1) može se pogodno zapisati u matričnom obliku

$$\mathbf{Y} = \mathbf{X}\boldsymbol{\beta}, \tag{6-3}$$

gdje se $\mathbf{Y}_{N\times 1}$ naziva vektor odziva, $\mathbf{X}_{N\times M}$ matrica dizajna, a $\boldsymbol{\beta}_{M\times 1}$ vektor koeficijenta. Optimalni regresijski koeficijenti *LR* modela $\widehat{\boldsymbol{\beta}}$, koji minimiziraju funkciju gubitka definiranu u smislu najmanje kvadratne pogreške (MSE), pronalaze se rješavanjem kvadratičnog minimizacijskog problema [54]

$$\widehat{\boldsymbol{\beta}} = \arg\min_{\beta}\left(\frac{\|\mathbf{Y} - \mathbf{X}\boldsymbol{\beta}\|_2^2}{N}\right) = (\mathbf{X}^{\mathbf{T}}\mathbf{X})^{-1}\mathbf{X}^{\mathbf{T}}\mathbf{y}. \tag{6-4}$$

6.1.2. LASSO regresijska analiza

Nadalje, uvodi se LASSO metoda regresije, koja provodi odabir značajki „skupljanjem“ (L1 regularizacija) kako bi se poboljšala interpretabilnost i izbjegla prenaučenost *LR* modela (6-1) [55, 57]. L1 regularizacija dodaje kaznu jednaku zbroju apsolutnih vrijednosti regresijskih koeficijenata u obliku ograničenja tipa nejednakosti $\|\boldsymbol{\beta}\|_1 \leq t$ na kvadratnu funkciju gubitka danu u (6-4). Prema tome, LASSO estimat dobiva se kao rješenje linearnog kvadratnog optimizacijskog problema s L1 ograničenjem [57]

$$\min\left(\frac{\|\mathbf{Y}-\mathbf{X}\boldsymbol{\beta}\|_2^2}{N}\right) \text{ uz ograničenje } \|\boldsymbol{\beta}\|_1 = \sum_{j=1}^{M}|\beta_j| \leq t, \tag{6-5}$$

gdje je t gornja granica zbroja apsolutnih vrijednosti parametara modela, tj. stupanj regularizacije. Optimizacijski problem (6-5) ekvivalentan je procjeni parametara $\widehat{\boldsymbol{\beta}}$ koji se u Lagrangeovom obliku zapisuje kao [57]

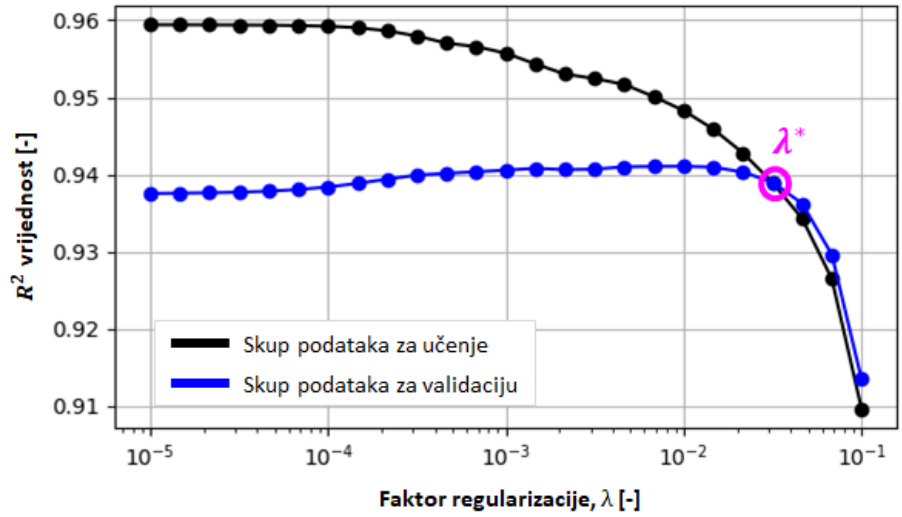
$$\widehat{\boldsymbol{\beta}}(\lambda) = \arg\min_{\beta}\left(\frac{\|\mathbf{Y}-\mathbf{X}\boldsymbol{\beta}\|_2^2}{N} + \lambda\|\boldsymbol{\beta}\|_1\right). \tag{6-6}$$

gdje λ predstavlja hiperparametar koji kontrolira snagu penalizacije, tj. kako se λ povećava, više se regresijskih koeficijenata β smanjuje u nulu i na taj način eliminira. U slučaju $\lambda = 0$ ne eliminira se niti jedan regresijski koeficijent, a optimizacijski problem se svodi na oblik najmanjih kvadrata (6-4). Statističke značajke koje nakon procesa „skupljanja“ imaju regresijski koeficijent različit od nule ($\beta_j(\lambda) \neq 0$), proglašavaju se relevantnim. Stupanj utjecajnosti svake statističke značajke u odnosu na zavisnu varijablu određen je rezultirajućom vrijednošću odgovarajućeg koeficijenta regresije $\hat{\beta}_j(\lambda)$. U daljnjim analizama se za *LR* model s LASSO kriterijskom funkcijom (6-6) usvaja notacija *LASSO* model.

6.1.3. Učenje i testiranje linearnih regresijskih modela

Učenje i testiranje *LR* i *LASSO* modela provedeno je na odvojenim skupovima podataka u omjeru 70% za učenje i 15% za testiranje (vidi poglavlje 4 za detalje). Kod LASSO modela se u svrhu određivanja optimalne vrijednosti hiperparametra λ dodatno koristi skup podataka za validaciju (15% od ukupnog skupa podataka). Pritom se vrijednost svake statističke značajke prethodno skalira na zajednički raspon od 0 do 1 prema principu min-max normalizacije kako bi se poništio utjecaj različitih raspona vrijednosti raznih značajki. Izračunate vrijednosti nominiranih statističkih značajki se potom koriste kao ulazne varijable tj. prediktori x_{ij}, dok se snimljena potrošnja goriva izražena u L/100 km koristi kao izlazna tj. zavisna varijabla y_i. Sukladno tome, ukupan broj opservacija *N* iznosi 4000, što odgovara ukupnom broju kombiniranih mikrociklusa (poglavlje 4), dok broj statističkih značajki *M* iznosi 100 (odnosno, jedna opservacija statističkih značajki iz skupa $\mathcal{F}$ po voznom ciklusu).

Optimalna vrijednost hiperparametra λ dobivena je ispitivanjem točnosti predviđanja *LASSO* modela za različite vrijednosti faktora regularizacije λ u rasponu $[10^{-5}, 10^{-1}]$, vodeći pritom evidenciju o generalizacijskoj pogrešci diskutiranoj u poglavlju 4. Postupak je ilustriran slikom 6-2 te se kao optimalan hiperparametar dobiva $\lambda^* = 3.16 \cdot 10^{-2}$.



Slika 6-2: Krivulje kvalitete predviđanja LASSO modela opisane R^2 vrijednostima u ovisnosti o vrijednosti faktora regularizacije λ, s obzirom na koje je određena optimalna vrijednost λ^*.

6.1.4. Rezultati regresijske analize

Točnost predviđanja potrošnje goriva za *LR*, *LASSO* i *NN-H3D* modela (definiranih redom u odjeljcima 6.1.1, 6.1.2 te poglavlju 4) ispitana je na testnom skupu kombiniranih mikrociklusa. Pokazatelji kvalitete predviđanja koji se uzimaju u obzir pri ocjeni točnosti predviđanja razmatranih modela uključuju koeficijent determinacije R^2 definiran izrazom (4-12), te srednju vrijednost μ_r i standardnu devijaciju σ_r reziduala predviđanja modela. Povrh toga, složenost regresijskih modela definira se ukupnim brojem parametara za učenje (BPZU) te brojem parametara čije vrijednosti nakon učenja modela iznose nula (BP0).

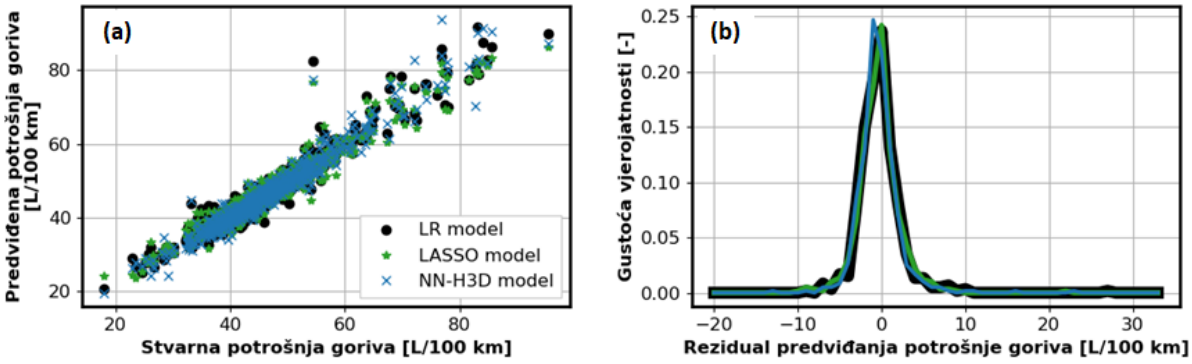
Izračunate vrijednosti spomenutih pokazatelja kvalitete predviđanja dane su u tablici 6-1, dok je na slici 6-3 prikazana ovisnost predviđene u odnosu na stvarnu (snimljenu) potrošnju goriva, zajedno s pripadnim razdiobama reziduala predviđanja. Rezultati ispitivanja triju regresijskih modela pokazuju kako oni imaju slične kvalitete predviđanja i razdiobe reziduala predviđanja potrošnje goriva. Konkretnije, *NN-H3D* model karakterizira najveći R^2 od 0.942 i najniži σ_r od 2.47 L/100 km, te kao takav predstavlja ultimativnu točnost predviđanja. Iako su *LR* i *LASSO* modeli znatno jednostavniji (BPZU jednak 100 u odnosu na 1,532.606 u slučaju *NN-H3D*), oni vrlo blisko slijede *NN-H3D* model, odnosno, njihove R^2 vrijednosti iznose redom 0.938 i 0.939, dok je σ_r veći za samo oko 5%. Također su rezultirajuće vrijednosti srednje pogreške predviđanja μ_r bliske nuli, što ukazuje na dobru uravnoteženost predviđanja svih triju modela. Nadalje, rezultati učenja modela pokazuju kako je *LASSO* sposoban smanjiti početni/ulazni skup $\mathcal{F}$ od 100 nominiranih statističkih značajki na njih 33, tj. ukupno 67 regresijskih koeficijenata $\hat{\beta}$ smanjeno je na nulu (BZ0 $= 67$). Ova tvrdnja potkrijepljena je grafom crne boje

na slici 6-4b, koja prikazuje rezultirajuće vrijednosti regresijskih koeficijenata $\hat{\beta}_j$ za svaku nominiranu statističku značajku iz skupa $\mathcal{F}$.

Tablica 6-1: Pokazatelji kvalitete predviđanja potrošnje goriva za LR i LASSO regresijske modele te NN-H3D model izračunati na testnom skupu podataka.

Model	μ_r[L]	σ_r[L]	R^2[-]	BPZU [-]	BP0
*LR**	0.22	2.54	0.938	101	0
*LASSO**	0.17	2.53	0.939	101	67
NN	0.12	2.47	0.942	1,532.606	0

* Rezultati se odnose na *LR* i *LASSO* modele punog reda (vidi tekst u nastavku za detaljno pojašnjenje).

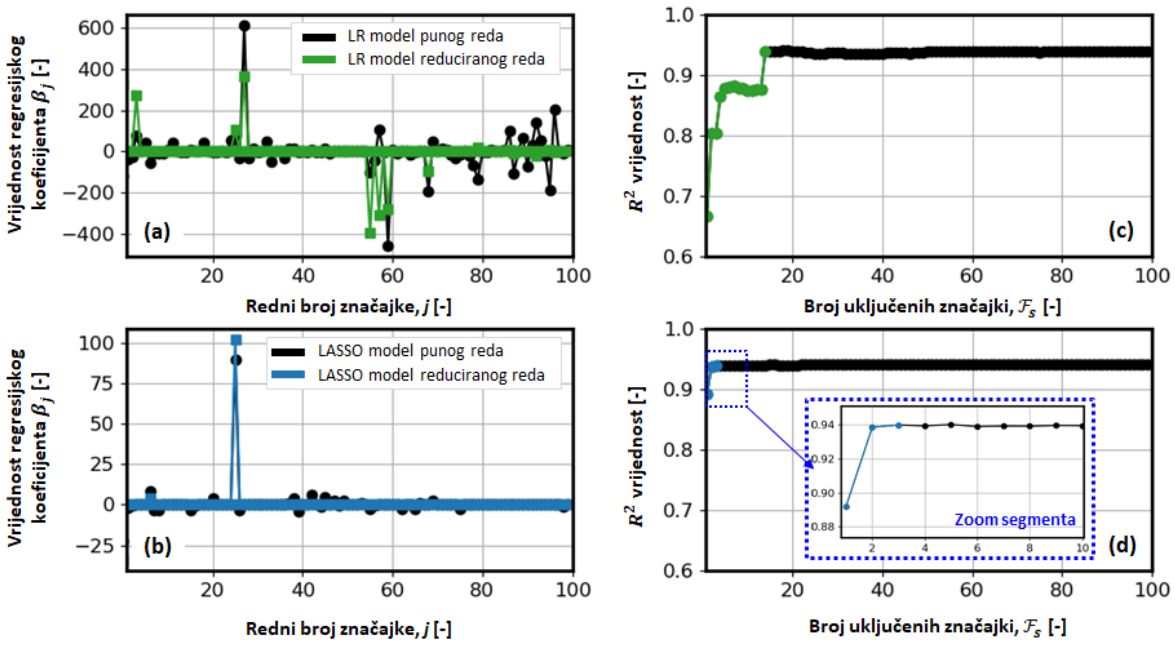


Slika 6-3: Usporedba a) predviđene u odnosu na stvarnu potrošnju goriva i b) razdiobe reziduala predviđanja potrošnje goriva za LR, LASSO i NN-H3D regresijske modele i testni skup podataka.

Kako bi se istražila daljnja mogućnost smanjenja broja relevantnih statističkih značajki, provodi se analiza osjetljivosti točnosti predviđanja *LR* i *LASSO* modela u ovisnosti o broju ulaznih značajki. Prvo se statističke značajke sortiraju prema relevantnosti (tj. vrijednosti $|\hat{\beta}_j|$) u opadajućem redoslijedu. Zatim se za svaki podskup $\mathcal{F}_s$ koji uključuje $s = 1,2,\dots S$ sortiranih značajki (gdje je $S = 100$) provodi učenje novih/zasebnih modela te se ispituje njihova točnost predviđanja temeljem R^2 vrijednosti dobivene na testnom skupu podataka. Konačno, provjerava se koji od modela rezultira s točnošću predviđanja bliskoj izvornom modelu punog reda danom u tablici 6-1.

Rezultati prikazani na slici 6-4 pokazuju kako je u slučaju *LR* modela dovoljno prvih 14 značajki za zadani zadatak predviđanja (vidi sliku 5-4c i zelenu krivulju na slici 5-4a), dok su samo prve tri značajke dovoljne kod *LASSO* modela (vidi sliku 5-4d i plavu krivulju na slici 5-4a). Budući da *LASSO* model rezultira boljim performansama i manjim skupom ulaznih značajki u odnosu na *LR* model (R^2 od 0.939 naspram 0.938, te 3 naspram 14 ulaznih značajki),

isti se usvaja kao referentni model za daljnje analize. S druge strane, glavni razlog za odabir *LASSO* modela umjesto *NN-H3D* modela je taj što, pored niže složenosti, kao ulaz koristi limitirani skup statističkih značajki (koje mogu biti poznate iz prometnih analiza/planiranja) umjesto punih informacija o voznom ciklusu, te također provodi automatski odabir značajki putem L1 regularizacije (vidi odjeljak 6.1.2). Stoga je za daljnju validaciju sintetičkih voznih ciklusa dovoljno poznavati samo vrijednosti koeficijenata regresije $\hat{\beta}_j$ dobivene za odgovarajući podskup od tri najsignifikantnije statističke značajke, koje su dane u tablici 6-2.



Slika 6-4: Vrijednosti koeficijenata regresije $\hat{\beta}_j$ za slučaj a) LR i b) LASSO modela punog i reduciranog reda; te postignute R^2 vrijednosti c)LR i d) LASSO modela u ovisnosti o broju ulaznih značajki tj. parametara.

Tablica 6-2: Konačno usvojen podskup najsignifikantnijih značajki LASSO regresijskog modela, uključujući pripadne vrijednosti koeficijenta regresije $\hat{\beta}_j$ i minimalne te maksimalne vrijednosti odgovarajućih značajki (izračunate prije min-max normalizacije).

Br.	**Statistička značajka**	**Min/max vrijednost**	**Mjerna jedinica**	**Vrijednost regresijskog koeficijenta***
1.	Srednja specifična energija vožnje, $\bar{E}_{vc}^{+}$	1.23/14.31	MJ/km	+101.61
2.	Srednja brzina vozila, $\bar{v}$	0.05/43.57	km/h	−34.11
3.	Srednje pozitivno ubrzanje vozila, $\bar{a}^{+}$	0.22/1.12	m/s^2	+3.89
4.	Posmak		L/100 km	33.98

* Vrijednosti koeficijenta regresije dobivene za slučaj koji uključuje min-max normalizaciju ulaznih značajki.

6.2. Izdvajanje reprezentativnih voznih ciklusa

6.2.1. Definiranje naprednog pokazatelja reprezentativnosti temeljenog na LASSO regresijskom modelu

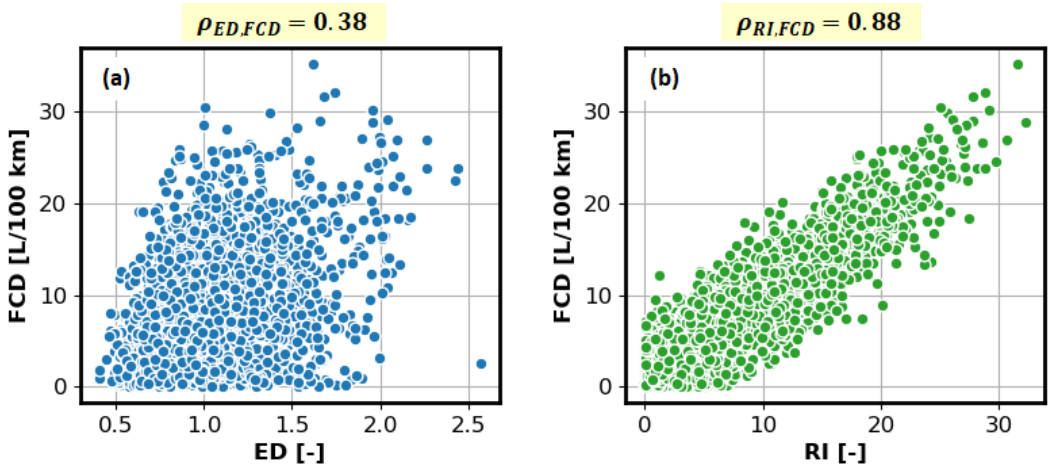
U poglavlju 5 izvedeno je nekoliko pokazatelja reprezentativnosti voznog ciklusa vezanih uz statističke značajke voznih ciklusa, te su sustavno valorizirani s obzirom na pokazatelj odstupanja potrošnje goriva u odnosu na mjerenu vrijednost (FCD; vidi izraz (5-3)), kako bi se ustanovilo koji su pokazatelji najprikladniji za validaciju sintetičkih voznih ciklusa. Rezultati korelacijske analize (slika 5-4) pokazali su kako je ED najprikladniji pokazatelj za validaciju voznih ciklusa, s najvećom vrijednosti indeksa korelacije $\rho_{ED,FCD} = 0.38$. Međutim, niti jedan od razmatranih pokazatelja nije se pokazao isključivo dostatnim za validaciju voznih ciklusa s obzirom na FCD, te je provedena višekriterijska validacija u Pareto optimalnom smislu. Ovdje se s ciljem približavanja FCD-u predlaže novi pokazatelj nazvan Regresijski Indeks (nadalje označen skraćenicom RI), koji se izvodi iz regresijskih koeficijenata $\hat{\beta}_j$ dobivenih za slučaj LASSO regresijskog modela reduciranog reda (tablica 6-2) na sljedeći način

$$\mathrm{RI}_i = \left|V_{f,s,i}^{LASSO} - \bar{V}_{f,z}^{LASSO}\right| \equiv \left|\sum_{j=1}^{J} \hat{\beta}_j \left(f_{s,i,j} - \bar{f}_{z,j}\right)\right|, \tag{6-7}$$

gdje $\bar{f}_{z,j}$ i $f_{s,i,j}$ predstavljaju redom vrijednosti *j*-te statističke značajke za kombinirani snimljeni vozni ciklus $\overline{\boldsymbol{\psi}}_z$ i *i*-ti sintetički vozni ciklus. Pritom J označava ukupni broj najsignifikantnijih statističkih značajki (ovdje, $J = 3$; vidi tablicu 6-2).

Od interesa je prvo utvrditi je li novopredloženi i računski jednostavniji pokazatelj RI točniji tj. reprezentativniji od pokazatelja ED. Odgovor na ovo pitanje traži se provođenjem korelacijske analize RI s obzirom na FCD. Na slici 6-5 prikazana je ovisnost odstupanja potrošnje goriva opisane pokazateljem FCD o pokazateljima ED i RI za 3000 sintetičkih voznih ciklusa. Rezultati korelacijske analize pokazuju kako ovisnost FCD o RI poprima gotovo linearni oblik (slika 6-5b), za razliku od ED u slučaju kojeg se javlja značajno rasipanje rezultata (slika 6-5a). Ovaj nalaz se potkrepljuje vrijednostima odgovarajućih indeksa korelacije $\rho_{RI,FCD} = 0.88$ i $\rho_{ED,FCD} = 0.38$, što jasno ukazuje na dominaciju RI nad ED u pogledu praćenja FCD pokazatelja. Povrh toga, budući da RI indirektno odražava odstupanje potrošnje goriva kao i FCD, njegova je glavna prednost u tome što, nakon što su odabrane signifikantne značajke voznih ciklusa (tablica 6-2), nema potrebe za daljnjim korištenjem podataka o potrošnji goriva (kako snimljenih tako i onih predviđenih temeljem složenog *NN-H3D* modela definiranog u poglavlju 4) za validaciju sintetičkih voznih ciklusa. Stoga se RI usvaja kao svojevrsni

ultimativni pokazatelj za daljnju validaciju sintetičkih voznih ciklusa prikazanu u narednom odjeljku.



Slika 6-5: Ovisnosti pokazatelja odstupanja potrošnje goriva (FCD) o pokazateljima reprezentativnosti voznih ciklusa a) ED i b) RI za slučaj 3000 sintetičkih voznih ciklusa, uključujući pripadne vrijednosti indeksa korelacije ($\rho_{ED,FCD}$, $\rho_{RI,FCD}$).

6.2.2. Postupak validacije voznih ciklusa temeljem regresijskog indeksa (RI)

Neka je $\Psi_R^S = \{\boldsymbol{\psi_1}, \boldsymbol{\psi_2}, \dots, \boldsymbol{\psi_i}, \dots, \boldsymbol{\psi_N}\}$ generirani skup od $N \equiv N_s^R = 3000$ sintetičkih voznih ciklusa, $\boldsymbol{\psi_i} = [\boldsymbol{v}_i^T\ \boldsymbol{a}_i^T\ \boldsymbol{\theta}_i^T\ \dot{\boldsymbol{\theta}}_i^T]; i = 1, 2, \dots, N$ je individualni (i-ti) sintetički vozni ciklus, dok su FCD_i, ED_i, i RI_i pripadne vrijednosti pokazatelja reprezentativnosti voznog ciklusa. Reprezentativni sintetički vozni ciklus $\boldsymbol{\psi}_r^*; r \in \{\mathrm{RI}, \mathrm{ED}\}$ moguće je izdvojiti prema kriteriju minimalnog RI (ili ED) kao

$$\boldsymbol{\psi}_{\boldsymbol{RI}}^* = \arg\min_i \left(\mathrm{RI}_i(\boldsymbol{\psi}_i)\right) \quad \text{ili} \quad \boldsymbol{\psi}_{\boldsymbol{ED}}^* = \arg\min_i \left(\mathrm{ED}_i(\boldsymbol{\psi}_i)\right). \tag{6-8}$$

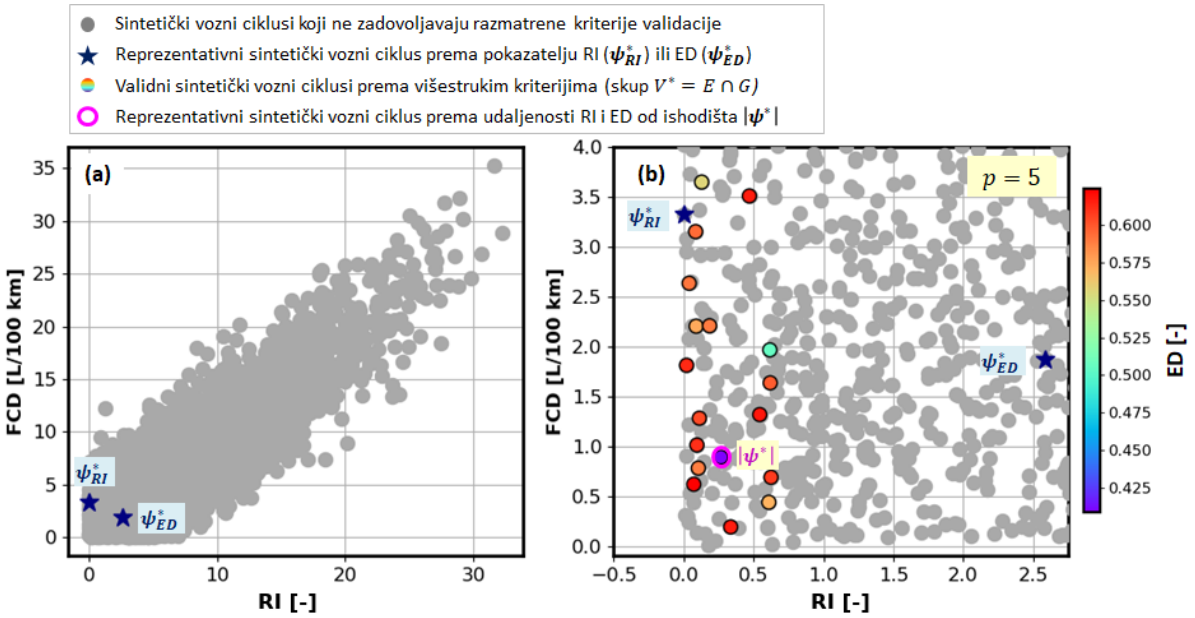
Međutim, kako je navedeno u poglavlju 5, postupak jednoznačne validacije voznih ciklusa definiran izrazom (6-8) ne garantira nužno i minimalno odstupanje potrošnje goriva definirano pokazateljem FCD. Ova tvrdnja potkrijepljena je slikom 6-6a, koja prikazuje pozicije rezultirajućih reprezentativnih voznih ciklusa $\boldsymbol{\psi}_{\boldsymbol{RI}}^*$ i $\boldsymbol{\psi}_{\boldsymbol{ED}}^*$ u koordinatnom sustavu pokazatelja RI i FCD, a iz kojih je vidljivo kako je za slučaj razmatranog skupa Ψ_R^S vrijednost FCD za $\boldsymbol{\psi}_{\boldsymbol{RI}}^*$ (3.32 L/100 km) viša od one dobivene za $\boldsymbol{\psi}_{\boldsymbol{ED}}^*$ (1.87 L/100 km), te se nalazi daleko od minimalne vrijednosti. Ukoliko se iz skupa Ψ_R^S izdvoji skup validnih sintetičkih voznih ciklusa V^* prema kriteriju $\mathrm{ED}_i \leq \varepsilon_{ED}$ (podskup E definiran izrazom (5-9)) i kriteriju $\mathrm{RI}_i \leq \varepsilon_{RI}$ (definiranog podskupom G) prema

$$G = \{\boldsymbol{\psi}_i \mid \mathrm{RI}_i \leq \varepsilon_{RI}\} \subseteq \Psi_R^S, \tag{6-9}$$

$$V^* = E \cap G, \tag{6-10}$$

gdje je granična vrijednost ε_{RI} (te također i ε_{ED}) određena vrijednošću 5. percentila ($p = 5$), dobivaju se rezultati prikazani na slici 6-6b, koji poboljšavaju uvid u daljnju mogućnost unaprjeđenja validacijske metode (6-8).

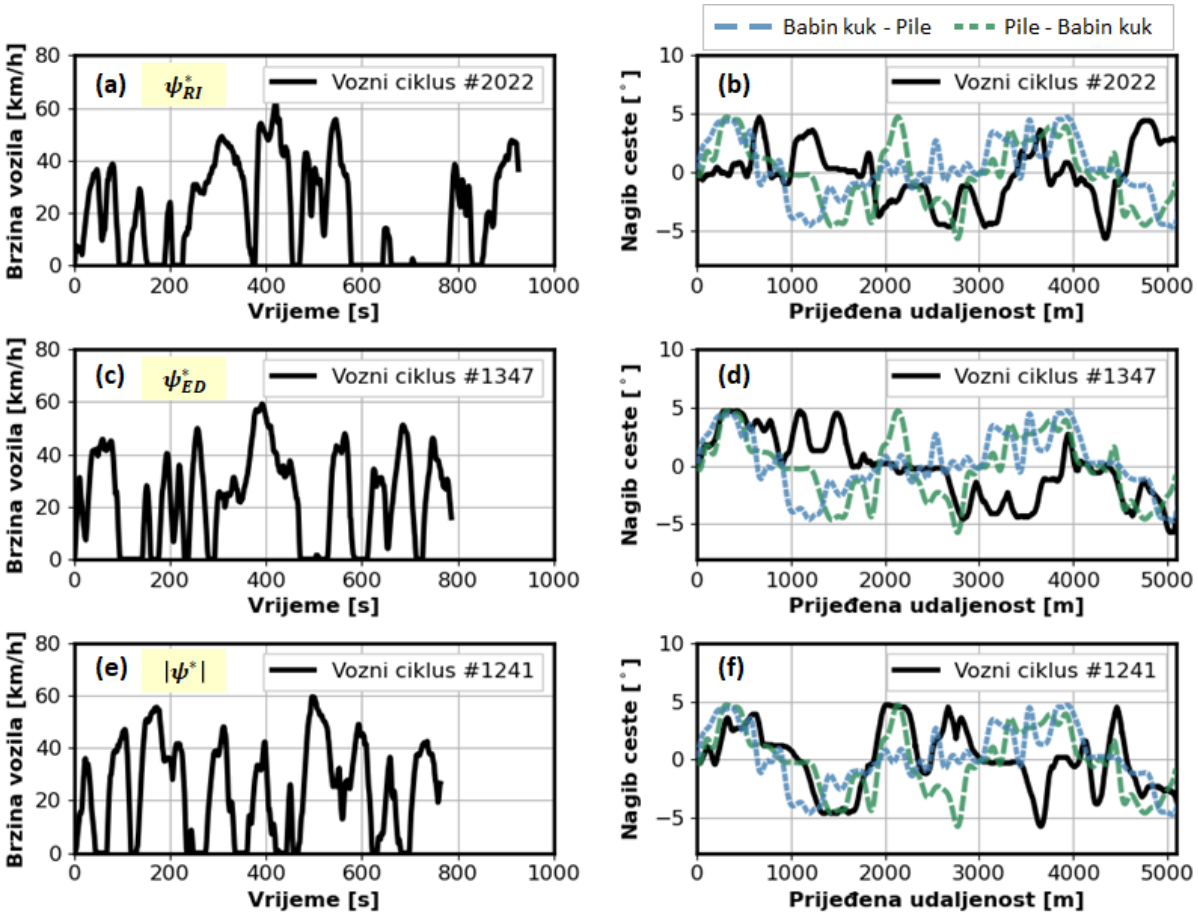
Vizualnom inspekcijom slike 6-6b može se ustanoviti kako postoje sintetički vozni ciklusi $\boldsymbol{\psi_i}$ koji za pripadne vrijednosti RI_i, koje nisu minimalne, istodobno posjeduju znatno niže vrijednosti FCD_i i ED_i u odnosu na $\boldsymbol{\psi^*_{RI}}$. Odgovarajući, među njima najreprezentativniji vozni ciklus označen je na slici 6-6b kao $|\boldsymbol{\psi}^*|$ (kružić s magenta okvirom), a moguće ga je izdvojiti iz skupa Ψ_R^S prema kriteriju minimalne udaljenosti RI_i i ED_i od ishodišta RI - ED koordinatnog sustava shodno izrazu (5-13). Vrijednosti pokazatelja RI i ED za rezultirajući vozni ciklus $|\boldsymbol{\psi}^*|$ iznose redom 0.89 L/100 km (-73.2% naspram RI od 3.32 L/100 km za $\boldsymbol{\psi^*_{RI}}$), te 0.41 (-54.9% naspram ED od 0.91 za $\boldsymbol{\psi^*_{RI}}$). Time se potvrđuje mogućnost značajnog unaprjeđenja metode jednoznačne validacije, koje se ostvaruje dodatnim izračunavanjem ED_i za svaki i-ti sintetički vozni ciklus. Izdvojeni reprezentativni sintetički vozni ciklusi $\boldsymbol{\psi^*_{RI}}$, $\boldsymbol{\psi^*_{ED}}$ i $|\boldsymbol{\psi}^*|$ detaljno se analiziraju u narednom odjeljku.



Slika 6-6: Pozicije reprezentativnih sintetičkih voznih ciklusa dobivenih temeljem a) jednoznačne validacije prema kriteriju minimalne vrijednosti pokazatelja RI i ED ($\boldsymbol{\psi^*_{RI}}$, $\boldsymbol{\psi^*_{ED}}$) te b) minimalne udaljenosti od ishodišta RI – ED koordinatnog sustava ($|\boldsymbol{\psi}^*|$).

6.2.3. Analiza reprezentativnih voznih ciklusa

Analiza reprezentativnih sintetičkih voznih ciklusa $\boldsymbol{\psi}^*_{\boldsymbol{RI}}$, $\boldsymbol{\psi}^*_{\boldsymbol{ED}}$ i $|\boldsymbol{\psi}^*|$, odabranih prema različitim kriterijima u prethodnom odjeljku, provodi se u smislu pripadnih statističkih značajki. Pritom je reprezentativni vozni ciklus $\boldsymbol{\psi}^*_{\boldsymbol{RI}}$ dalje označen kao #2022, $\boldsymbol{\psi}^*_{\boldsymbol{ED}}$ kao #1347, a $|\boldsymbol{\psi}^*|$ kao #1241, gdje # označava redni broj sintetičkog voznog ciklusa od ukupno 3000 voznih ciklusa. Rezultirajući vremenski profili brzine razmatranih reprezentativnih voznih ciklusa #2022, #1347 i #1241 prikazani su na slici 6-7, zajedno s odgovarajućim profilima nagiba ceste u ovisnosti o prijeđenoj udaljenosti.



Slika 6-7: a), c) e) Profili brzine vozila u vremenu te b), d), e) nagiba ceste u ovisnosti o prijeđenoj udaljenosti za tri reprezentativna sintetička vozna ciklusa izdvojena odabirom značajki podržanom validacijom te označena kao $\boldsymbol{\psi}^*_{\boldsymbol{RI}}$ (#2022), $\boldsymbol{\psi}^*_{\boldsymbol{ED}}$ (#1347), te $|\boldsymbol{\psi}^*|$ (#1241), uključujući snimljene profile nagiba ceste za oba smjera vožnje (označene isprekidanim crtama).

Nadalje, provedena je usporedna analiza statističkih značajki izdvojenih reprezentativnih voznih ciklusa. Izračunate vrijednosti statističkih značajki dane su u tablici 6-3, zajedno s odgovarajućim srednjim vrijednostima snimljenog kombiniranog voznog ciklusa $\bar{\boldsymbol{\psi}}_z$ i

pripadnim relativnim odstupanjima statističkih značajki $\Delta f_{rel,m}$ danim u zagradama. Zbog velike količine nominiranih statističkih značajki ($\mathcal{F} \in \mathbb{R}^{1\,x\,100}$), usporedba se provodi za limitirani skup $\mathcal{F}_L$ od 10 odabranih značajki. Ovaj skup uključuje osam bazičnih validacijskih značajki iz skupa $\mathcal{F}_R$ iz tablice 5-3 i dodatne tri signifikantne značajke iz tablice 6-2 izdvojene temeljem LASSO regresijske analize.

Vizualnom inspekcijom slika 6-7a, 6-7c i 6-7e može se ustanoviti kako razmatrani reprezentativni vozni ciklusi $\boldsymbol{\psi}^*_{\boldsymbol{RI}}$ (#2022), $\boldsymbol{\psi}^*_{\boldsymbol{ED}}$ (#1347) te $|\boldsymbol{\psi}^*|$ (#1241) posjeduju donekle različite vremenske profile brzine u smislu ukupnog vremena putovanja (i mirovanja pri zaustavljanju), koje se direktno odražavaju na različite vrijednosti srednje brzine vozila $\bar{v}$ (redom 19.85 km/h, 23.33 km/h te 24.05 km/h; tablica 6-3). Razlog tome su uglavnom različita vremena mirovanja vozila pri zaustavljanju, koja su u prosjeku najveća za $\boldsymbol{\psi}^*_{\boldsymbol{RI}}$ (29.4 s), srednja za $\boldsymbol{\psi}^*_{\boldsymbol{ED}}$ (19 s), te najmanja za $|\boldsymbol{\psi}^*|$ (13.2 s). Maksimalna brzina vozila doseže najveću vrijednost od 63 km/h za vozni ciklus $\boldsymbol{\psi}^*_{\boldsymbol{RI}}$, te gotovo identične vrijednosti za vozne cikluse $\boldsymbol{\psi}^*_{\boldsymbol{ED}}$ i $|\boldsymbol{\psi}^*|$ (redom 59.1 km/h i 59.5 km/h).

Rezultati na slikama 6-7b i 6-7d pokazuju kako sintetizirani profili nagiba ceste za $\boldsymbol{\psi}^*_{\boldsymbol{RI}}$ **i** $\boldsymbol{\psi}^*_{\boldsymbol{ED}}$ imaju nešto izraženije srednjo-frekvencijske karakteristike u usporedbi sa snimljenim profilima, dok to nije slučaj za$|\boldsymbol{\psi}^*|$ (slika 6-7f). Razlog tome može se pripisati utjecaju profila nagiba ceste za smjer Pile - Babin kuk (pri tvorbi TPM; vidi poglavlje 3), koji je oscilatorniji od onog rekonstruiranog za suprotni smjer. Neovisno o tome, razmatrani sintetizirani profili nagiba ceste i dalje u velikoj mjeri nalikuju na snimljene profile, što je potkrijepljeno niskim relativnim odstupanjima RMS nagiba ceste θ_{RMS} (< 4%; tablica 6-3), izuzev za vozni ciklus $\boldsymbol{\psi}^*_{\boldsymbol{RI}}$ (-18.09%).

Općenito gledano, rezultati prikazani u tablici 6-3 ukazuju na primjetnu različitost razmatranih statističkih značajki za reprezentativne vozne cikluse $\boldsymbol{\psi}^*_{\boldsymbol{RI}}$ (#2022), $\boldsymbol{\psi}^*_{\boldsymbol{ED}}$ (#1347), te $|\boldsymbol{\psi}^*|$ (#1241), gdje vidljivu dominaciju posjeduje $|\boldsymbol{\psi}^*|$ s najvećim brojem minimalnih relativnih odstupanja statističkih značajki $\Delta f_{rel,m}$ (polja označena zelenom bojom). Najlošije karakteristike pokazuje vozni ciklus $\boldsymbol{\psi}^*_{\boldsymbol{RI}}$ koji posjeduje najveća odstupanja statističkih značajki $\Delta f_{rel,m}$, uključujući odstupanja prosječne potrošnje goriva $\bar{V}_f$ (-8.34%) i srednje specifične energije vožnje po prijeđenom kilometru $\bar{E}^+_{vc}$ (-7.44%). Pritom $\bar{E}^+_{vc}$ predstavlja najznačajniju nezavisnu varijablu tj. prediktor odabran LASSO metodom (vidi potpoglavlje 6.1), koja je usko povezana s potrošnjom goriva V_f. Vozni ciklus $\boldsymbol{\psi}^*_{\boldsymbol{ED}}$ pokazuje solidne karakteristike, izuzev nešto viših odstupanja BZPK (-16.07%) te $\overline{PG}_{P_m}$ (-17.45%), koje su i u slučaju višekriterijske

validacije bile veće (usp. tablicu 5-2). Povrh toga, odstupanje $\bar{V}_f$ je manje od 5%, pa se stoga vozni ciklus $\boldsymbol{\psi}^*_{\boldsymbol{ED}}$, pored $|\boldsymbol{\psi}^*|$, može također smatrati prihvatljivim odabirom, ovisno o primjeni.

Tablica 6-3: *Vrijednosti odabranih statističkih značajki iz limitiranog skupa $\mathcal{F}_L$ za kombinirani snimljeni vozni ciklus $\bar{\boldsymbol{\psi}}_z$ i odabrane reprezentativne sintetičke vozne cikluse označene kao $\boldsymbol{\psi}^*_{\boldsymbol{RI}}$ (#2022), $\boldsymbol{\psi}^*_{\boldsymbol{ED}}$ (#1347), te $|\boldsymbol{\psi}^*|$ (#1241).*

Statistička značajka	$\bar{\boldsymbol{\psi}}_z$	**Reprezentativni sintetički vozni ciklus**		
		$\boldsymbol{\psi}^*_{\boldsymbol{RI}}$ **(#2022)**	$\boldsymbol{\psi}^*_{\boldsymbol{ED}}$ **(#1347)**	$\|\boldsymbol{\psi}^*\|$ **(#1241)**
Srednja pozitivna brzina vozila, $\bar{v}^+$ (km/h)	28.89	28.52 (-1.30%)*	29.46 (+1.95%)	29.20 (+1.05%)
Standardna devijacija ubrzanja vozila, a_σ (m/s^2)	0.53	0.48 (-9.67%)	0.51 (-5.16%)	0.52 (-3.26%)
Broj zaustavljanja vozila po prijeđenom kilometru, BZPK (km^{-1})	1.87	1.76 (-5.79%)	1.57 (-16.07%)	1.76 (-5.65%)
Standardna devijacija zahtijevane snage na kotaču vozila, P_W^σ (kW)	67.46	61.70 (-8.54%)	63.42 (-5.99%)	69.52 (+3.06%)
RMS nagiba ceste, θ_{RMS} (°)	2.59	2.12 (-18.09%)	2.63 (+1.57%)	2.49 (-3.87%)
Korelacija brzine vozila i pozitivnog nagiba ceste, $\rho_{v\theta^+}$ (-)	0.06	-0.02	0.18	0.10
Srednja vrijednost frekvencijskog odziva (periodograma) specifične snage vozila, $\overline{PG}_{P_m}$ (kW^2/kg^2/Hz)	32.01	26.76 (-16.42%)	26.43 (-17.45%)	29.97 (-6.40%)
Prosječna potrošnja goriva, $\bar{V}_f$ (L/100 km)	39.84	36.52 (-8.34%)	41.71 (+4.69%)	38.95 (-2.24%)
Srednja specifična energija vožnje, $\bar{E}^+_{vc}$ (MJ/km)	4.32	4.00 (-7.44%)	4.01 (-7.05%)	4.47 (+3.52%)
Srednja brzina vozila, $\bar{v}$ (km/h)	23.03	19.85 (-13.78%)	23.33 (+1.31%)	24.05 (+4.43%)
Srednje pozitivno ubrzanje vozila, $\bar{a}^+$ (m/s^2)	0.45	0.43 (-5.29%)	0.46 (+1.20%)	0.43 (-3.97%)

* Vrijednosti u zagradama predstavljaju relativna odstupanja u odnosu na snimljene vrijednosti $\bar{\boldsymbol{\psi}}_z$, čije su minimalne vrijednosti označene zelenom bojom.

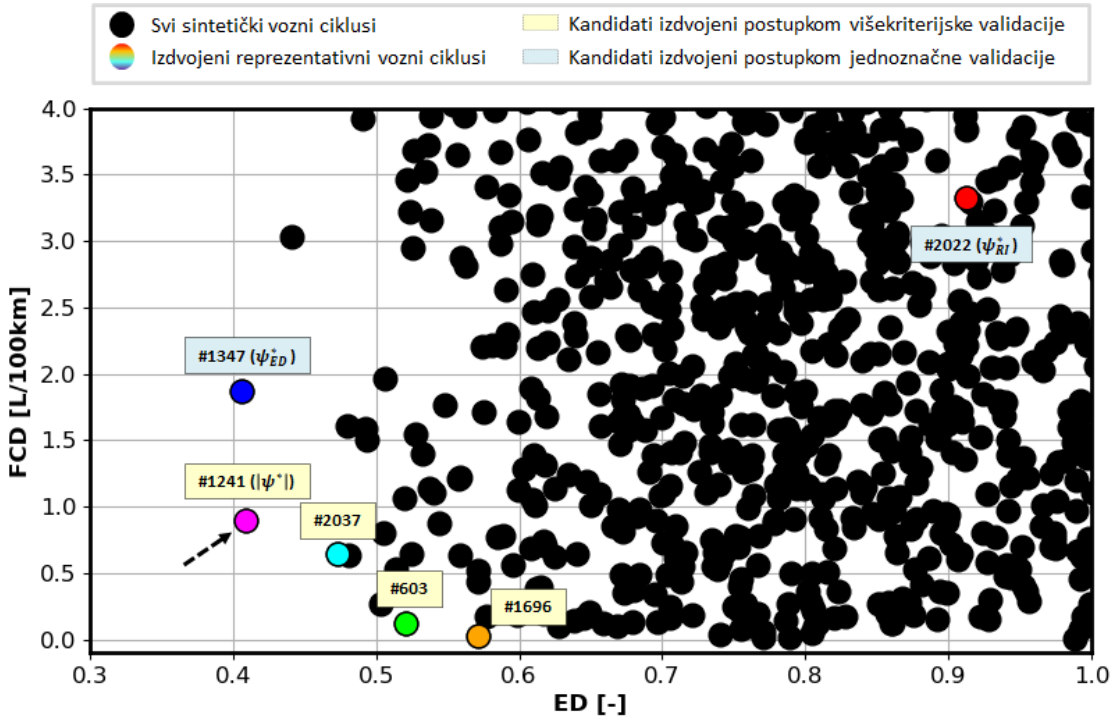
6.2.4. Usporedna analiza rezultata jednoznačne i višekriterijske validacije voznih ciklusa

S ciljem daljnje usporedbe reprezentativnih voznih ciklusa izdvojenih postupkom jednoznačne validacije u odnosu na višekriterijsku validaciju predstavljenu u poglavlju 5, provedena je usporedna analiza pokazatelja FCD i ED te statističkih značajki. Na slici 6-8 prikazano je svih šest izdvojenih reprezentativnih voznih ciklusa u ED-FCD koordinatnom sustavu (obojeni kružići), zajedno s ostalim sintetičkim voznim ciklusima (crni kružići). Pritom su redni brojevi reprezentativnih voznih ciklusa izdvojenih postupkom jednoznačne validacije označeni plavom pozadinom (ciklusi #1347, #2022), dok su oni izdvojeni temeljem višestrukih kriterija označeni

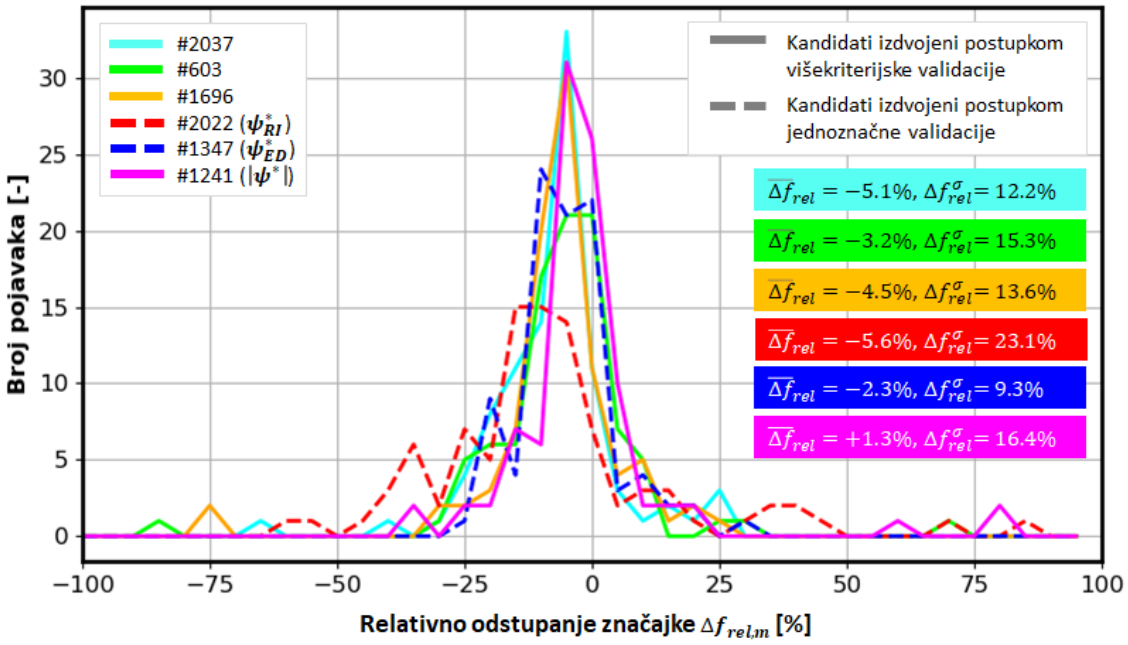
žutom pozadinom (ciklusi #1241, #2037, #603 te #1696). Vidljivo je kako se novoizdvojeni reprezentativni vozni ciklusi $\boldsymbol{\psi}_{\boldsymbol{ED}}^{*}$ (#1347), te $|\boldsymbol{\psi}^{*}|$ (#1241), uz vozne cikluse #2037, #603 te #1696, također nalaze na prednjoj/Pareto fronti ED-FCD pokazatelja (slika 6-8), te predstavljaju solidan odabir u smislu minimalnih odstupanja statističkih značajki (odraženih kroz niži ED od ≈0.4), uz određeni porast odstupanja potrošnje goriva (FCD u intervalu od 0.9-2 L/100 km naspram 0-0.7 L/100 km). S druge strane, vozni ciklus izdvojen postupkom jednoznačne validacije temeljem pokazatelja RI (#2022; $\boldsymbol{\psi}_{\boldsymbol{RI}}^{*}$**)** značajno je udaljen od ED-FCD Pareto fronte (slika 6-8), pa stoga ne predstavlja valjan odabir.

Na slici 6-9 prikazane su razdiobe relativnih odstupanja statističkih značajki $\Delta f_{rel,m}$ zajedno s prosječnim odstupanjima $\overline{\Delta f_{rel}}$ i pripadnim standandardnim devijacijama Δf_{rel}^{σ} za svaki izdvojeni reprezentativni vozni ciklus, za slučaj cjelokupnog nominiranog skupa $\mathcal{F}$ od 100 značajki. Vizualnom inspekcijom slike 6-9 može se ustanoviti kako se većina relativnih odstupanja značajki $\Delta f_{rel,m}$ nalazi unutar intervala ±10%, što je potkrijepljeno prosječnim odstupanjima značajki $\overline{\Delta f_{rel}}$ do ±5%. Minimalna vrijednost $\overline{\Delta f_{rel}}$ od +1.3% ostvarena je za vozni ciklus $|\boldsymbol{\psi}^{*}|$ (#1241; dodatno označen strelicom na slici 6-8), uz standardnu devijaciju Δf_{rel}^{σ} od 16.4%. Razlog nešto veće Δf_{rel}^{σ} za $|\boldsymbol{\psi}^{*}|$ u odnosu na vrijednosti dobivene za vozne cikluse #2037, #603 te #1696 (Δf_{rel}^{σ} redom 12.2%, 15.3% i 13.6%) je u tome što su prilikom višekriterijske validacije odgovarajućih voznih ciklusa uz ED dodatno korišteni pokazatelji MAE i STD. Najveća odstupanja statističkih značajki odnose se na $\boldsymbol{\psi}_{\boldsymbol{RI}}^{*}$ ($\overline{\Delta f_{rel}}$ te Δf_{rel}^{σ} redom -5.6% i 23.1%), ponovno potvrđujući lošija svojstva u odnosu na ostale kandidate (usp. tablicu 6-3).

Iz gore predstavljenih rezultata analize izdvojenih reprezentativnih voznih ciklusa može se zaključiti kako novopredloženi dvo-kriterijski postupak validacije temeljen na pokazateljima RI i ED (koji rezultira izdvajanjem voznog ciklusa $|\boldsymbol{\psi}^{*}|$) osigurava minimalna odstupanja prema gotovo svim karakterističnim statističkim značajkama, uključujući potrošnju goriva. Glavna praktična prednost ovog postupka validacije u odnosu na postupak višekriterijske validacije predložen u poglavlju 5 je u znatno jednostavnijem izračunavanju pokazatelja RI u odnosu na FCD. Konkretnije, pokazatelj RI indirektno odražava odstupanje potrošnje goriva (kao i FCD) kroz jednom određene vrijednosti regresijskih koeficijenata *LASSO* modela triju signifikantnih statističkih značajki (tablica 6-2), dok izračun FCD pokazatelja zahtjeva primjenu složenog *NN-H3D* modela definiranog u poglavlju 4.



Slika 6-8: Pozicije svih šest reprezentativnih sintetičkih voznih ciklusa izdvojenih temeljem postupka jednoznačne ili višekriterijske validacije u ED-FCD koordinatnom sustavu.



Slika 6-9: Razdiobe relativnih odstupanja statističkih značajki ($\Delta f_{rel,m}$) za svih šest reprezentativnih sintetičkih voznih ciklusa izdvojenih temeljem postupka jednoznačne ili višekriterijske validacije, zajedno s pripadnim prosječnim vrijednostima ($\overline{\Delta f_{rel}}$) i standandardnim devijacijama (Δf^{σ}_{rel}), izračunate za cjelokupan skup nominiranih statističkih značajki $\mathcal{F}$.

7. Predviđanje značajki voznih ciklusa

U ovom poglavlju razvijaju se prediktivni modeli predviđanja brzine vozila duž rute, temeljeni na unaprijednoj višeslojnoj neuronskoj mreži, te se provodi sveobuhvatna analiza točnosti predviđanja razvijenih prediktivnih modela temeljem odgovarajućih pokazatelja kvalitete predviđanja. U potpoglavlju 7.1 predlaže se statički stohastički model, koji predviđa razdiobe brzine vozila duž rute temeljem pozicije vozila kao jedine ulazne varijable u model. Točnost predviđanja statičkog stohastičkog modela ispituje se u smislu odabranih statističkih pokazatelja vezanih za razdiobe brzine vozila duž rute (srednja/očekivana vrijednost i standardna devijacija). U potpoglavlju 7.2 predlaže se deterministički model koji je znatno bolje uvjetovan ulaznim podacima te je namijenjen za precizno kratkoročno predviđanje profila brzine vozila na vremenskom horizontu u budućnosti. Točnost predviđanja detaljno se ispituje za razne kombinacije kandidiranih ulaznih varijabli tj. značajki, te za različite duljine vremenskog prozora povijesti brzina i horizonta predviđanja. Konačno, u potpoglavlju 7.3 predlaže se dinamički stohastički model koji posjeduje objedinjena svojstva dvaju prethodno definiranih modela.

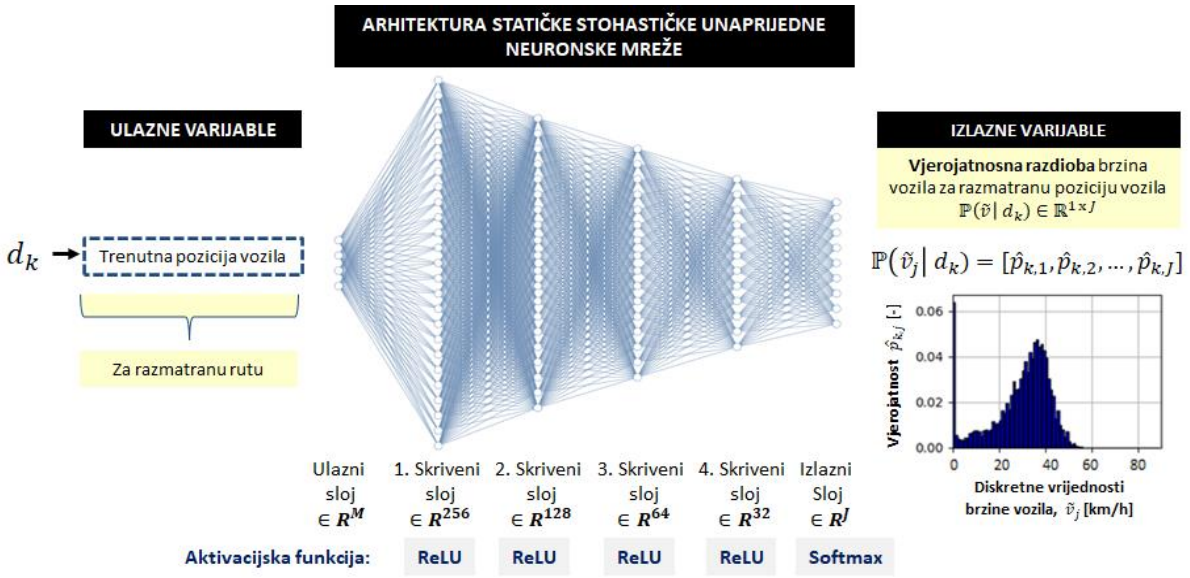
7.1. Statički stohastički model predviđanja brzine vozila

7.1.1. Modeliranje unaprijedne višeslojne neuronske mreže sa statičkim stohastičkim izlazom

Za potrebe modeliranja rasipanja brzine vozila duž određene rute predlaže se statički stohastički model temeljen na unaprijednoj višeslojnoj NN, koji je predstavljen arhitekturom prikazanom na slici 7-1 i dalje označavan kao NN-STC. Zbog velikog broja snimljenih podataka (tj. mnoštva prolazaka autobusa duž razmatrane rute), ovaj bazični model u stanju je mapirati karakteristike odgovarajuće rute, uključujući čitavu povijest prometnih uvjeta odraženu kroz srednje/očekivane vrijednosti i varijance brzine vozila duž rute.

Ulaz u NN-STC model uključuje podatak o trenutnoj poziciji vozila (d_k), koja predstavlja apsolutnu prijeđenu udaljenost u odnosu na referentnu točku definiranu krajnjom stanicom (Babin kuk, poglavlje 2). Izlaz iz NN-STC modela je vektor $\widehat{\boldsymbol{p}}_k = \left[\hat{p}_{k,j}\right]; j = 1,2,\dots,J$ koji predstavlja razdiobu uvjetnih vjerojatnosti diskretnih vrijednosti brzine vozila sadržanih u vektoru $\widetilde{\boldsymbol{v}} = \left[\tilde{v}_j\right]$ za razmatrani skup ulaznih podataka $\widehat{\boldsymbol{p}}_k = \mathbb{P}\left(\tilde{v}_j \middle| d_k\right)$, gdje indeks k označava trenutno stanje ulaza. Raspon brzina vozila $\widetilde{\boldsymbol{v}}$ postavljen je na interval od 0 do 90 km/h (maksimalna brzina autobusa), uz korak diskretizacije od 1 km/h, čime se dobiva ukupni broj

od $J = 91$ elemenata sadržanih u izlaznom vektoru $\hat{\boldsymbol{p}}_k \in \mathbb{R}^{1\,\mathrm{x}\,J}$. Struktura NN-STC modela definirana je s ulaznim slojem koji prima $M = 1$ ulaznu varijablu tj. značajku, četiri potpuno povezana skrivena sloja, koji redom sadrže 256, 128, 64 i 32 skrivena neurona, te izlaznog sloja sa $J = 91$ elementa (slika 7-1). Ukupni broj podesivih parametara NN-STC modela time iznosi 46.746. Za potrebe bolje generalizacije NN-STC modela, svaki skriveni sloj je regulariziran na način kako je opisano u poglavlju 4, uz faktor regularizacije λ postavljen na vrijednost $5 \cdot 10^{-4}$.



Slika 7-1: Predložena arhitektura statičke stohastičke unaprijedne višeslojne neuronske mreže za predviđanje rasipanja brzine vozila duž rute, $\tilde{v}_j$, na temelju pozicije vozila d_k (NN-STC).

Skriveni slojevi koriste ReLU aktivacijsku funkciju, dok izlazni sloj koristi *softmax* kao aktivacijsku funkciju. Razlog potonjem je što je izlaz iz NN-STC modela određen vjerojatnosnom razdiobom $\hat{p}_k$ (pri čemu mora vrijediti uvjet da je $\sum_{j=1}^{J} \hat{p}_{k,j} = 1$), a izlazne vrijednosti *softmax* aktivacijske funkcije predstavljaju upravo vjerojatnosti $\hat{p}_{k,j}$ da određeni skup ulaznih značajki pripada određenoj klasi tj. diskretnom intervalu brzine vozila $c_j \in [\tilde{v}_j, \tilde{v}_{j+1}]$:

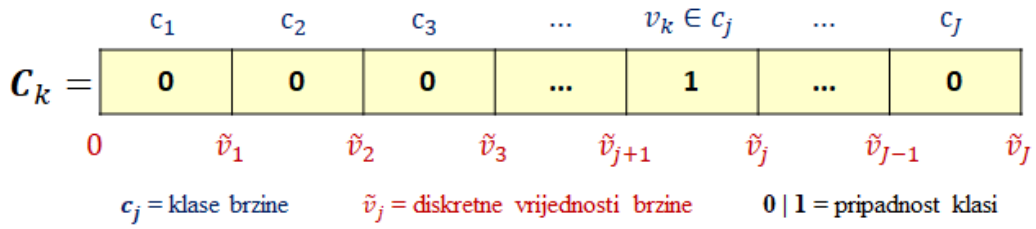
$$softmax(O_{k,j}) = \frac{e^{O_{k,j}}}{\sum_j e^{O_{k,j}}} \equiv \hat{p}_{k,j}, \tag{7-1}$$

gdje $O_{k,j}$ označava vrijednosti izlaznog sloja prije primjene *softmax* aktivacijske funkcije, dok je e baza prirodnog logaritma. Pritom se pripadnost stvarnih/snimljenih brzina vozila v_k klasi c_j unose u NN-STC model u obliku kodiranog (*engl. one-hot*) vektora dimenzije $\boldsymbol{C}_k \in \mathbb{R}^{1\,\mathrm{x}\,J}$, na način da se vrijednost j-tog elementa za kojeg vrijedi $v_k \in c_j$ postavi u 1, dok se svi ostali elementi vektora postavljaju na 0 (slika 7-2). Za potrebe rješavanja ovog više-klasnog

klasifikacijskog problema potrebno je odabrati parametre modela koji maksimiziraju vjerojatnost ispravno predviđenih klasa brzine vozila c_j za razmatrani skup snimljenih podataka, što se ostvaruje primjenom tzv. metode procjene maksimalne vjerojatnosti (*engl. maximum likelihood estimation*; MLE). U tu svrhu, kao funkcija gubitka tipično se koristi negativna log-vjerojatnost (*engl. negative log-likelihood*, NLL) [78], koja je definirana kao

$$NLL = \frac{1}{N}\sum_{k=1}^{N}\left(-\sum_{j=1}^{J} \boldsymbol{C}_{k,j} \cdot \log\left(\hat{p}_{k,j}\right)\right), \tag{7-2}$$

gdje $\hat{p}_{k,j}$ označava predviđene vjerojatnosti pripadnosti određenoj klasi brzine vozila c_j za k-to ulazno stanje (ovdje d_k), a N ukupni broj uzoraka u skupu podataka za učenje. Razlog uvođenja NLL naspram MLE je u tome što se funkcija gubitka tipično minimizira, pa se stoga minimiziranjem NLL izravno provodi maksimizacija MLE.



Slika 7-2: Ilustracija postupka kodiranja trenutne brzine vozila v_k u vektor $\boldsymbol{C}_k \in \mathbb{R}^{1\,x\,J}$ koji predstavlja oznaku tj. ciljanu vrijednost pri učenju NN-STC modela.

Realizacija NN-STC modela izvodi se unutar *Python* programskog okruženja korištenjem *Keras* modula s *Tensorflow* modulom kao pozadinom, isto kao i u slučaju NN-H3D modela definiranog u poglavlju 4. Pritom se za učenje modela također koristi ugrađeni optimizacijski algoritam ADAM sa zadanim postavkama, uz stopu učenja od $\eta = 10^{-3}$. Količina uzoraka iz koje se određuje gradijent postavljena je na 16, dok je broj epoha učenja postavljen na 150. Rezultati učenja, validacije i testiranja NN-SHC dani su u narednim odjeljcima.

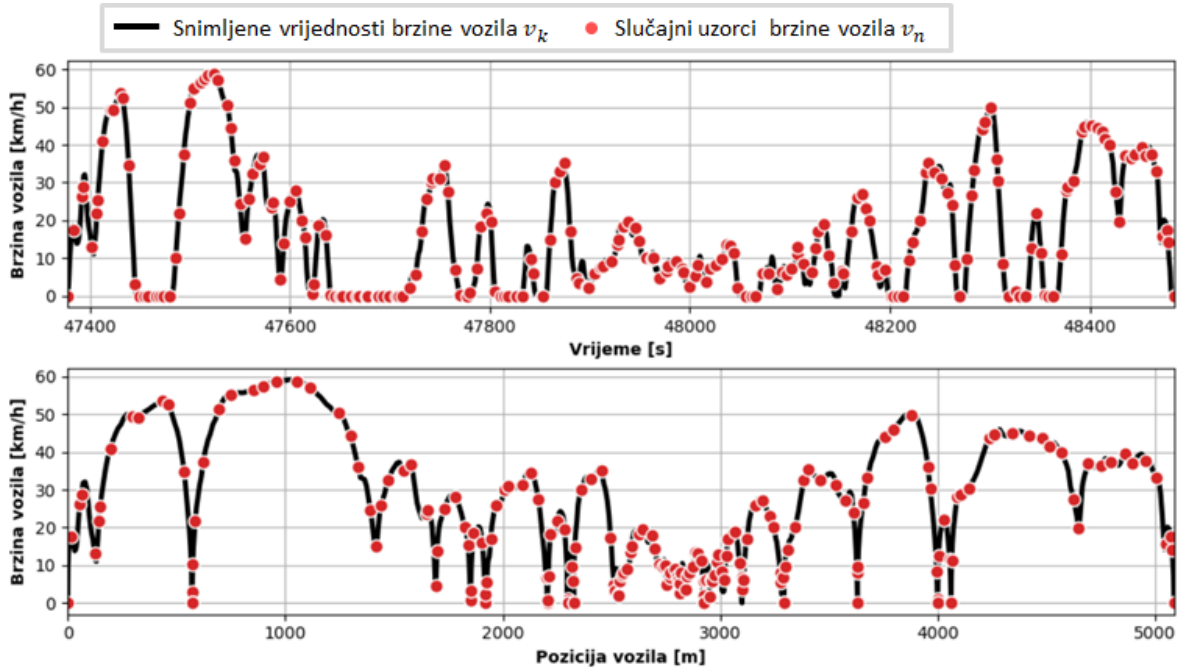
7.1.2. Priprema podataka za učenje, validaciju i testiranje prediktivnog modela

Za potrebe učenja, validacije i testiranja NN-STC modela korišten je puni skup od $N_{DC,BK-P}^{V} = 2313$ validnih voznih ciklusa, snimljenih na relaciji Babin kuk-Pile (poglavlje 2), što čini ukupni skup od 1,908.209 snimljenih uzoraka. Međutim, zbog složenosti samog zadatka predviđanja, učenje NN-STC modela na tako bogatom skupu snimljenih podataka (≈2 mil. uzoraka) može predstavljati vrlo zahtjevan zadatak u smislu vremena učenja NN. Stoga se predlaže metoda izdvajanja reduciranog skupa slučajnih uzoraka na sljedeći način: prolaskom kroz svaki peti uzastopni snimljeni vremenski trenutak $t_{i,k}; k = 1, 6, 11 \ldots, K$ za svaki i-ti

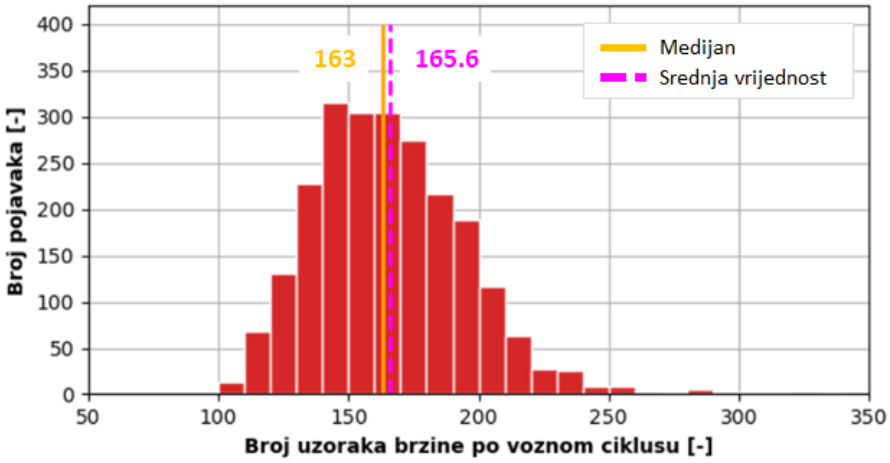
snimljeni vozni ciklus, na svakom uzastopnom vremenskom intervalu od $T_{i,k} \in [t_{i,k}, t_{i,k} + 5]$ sekundi uzima se metodom slučajnog odabira po jedan uzorak $\boldsymbol{u}_n$, opisan s diskretnom vrijednosti brzine vozila v_n i pozicije d_n, te se pohranjuje u skup $\mathcal{U} = \{\boldsymbol{u}_1, \boldsymbol{u}_2, \dots, \boldsymbol{u}_n\}$. Time se dobiva konačni/reducirani skup od 382.553 uzoraka, koji su dalje podijeljeni na skup za učenje (70% uzoraka), validaciju (15% uzoraka) i testiranje (15% uzoraka) NN-STC modela. Pritom se podrazumijeva da je svaki uzorak brzine vozila v_n kodiran u binarni vektor $\boldsymbol{C}_n$, koji predstavlja pripadnost odgovarajućoj klasi brzine c_j, na način kako je prikazano na slici 7-2.

Na slici 7-3 je za individualni vozni ciklus prikazan primjer snimljenog profila brzine vozila u domeni vremena i pozicije vozila (crne linije), zajedno s izdvojenim slučajnim uzorcima brzine vozila (crveni kružići). Na slici 7-4 prikazan je histogram ukupnog broja uzoraka brzine vozila po svakom snimljenom voznom ciklusu, iz kojeg se vidi da se broj uzoraka po ciklusu uglavnom kreće u rasponu od 100 do 250, dok prosječna vrijednost iznosi 165.6, te je približno jednaka medijanu koji iznosi 163.

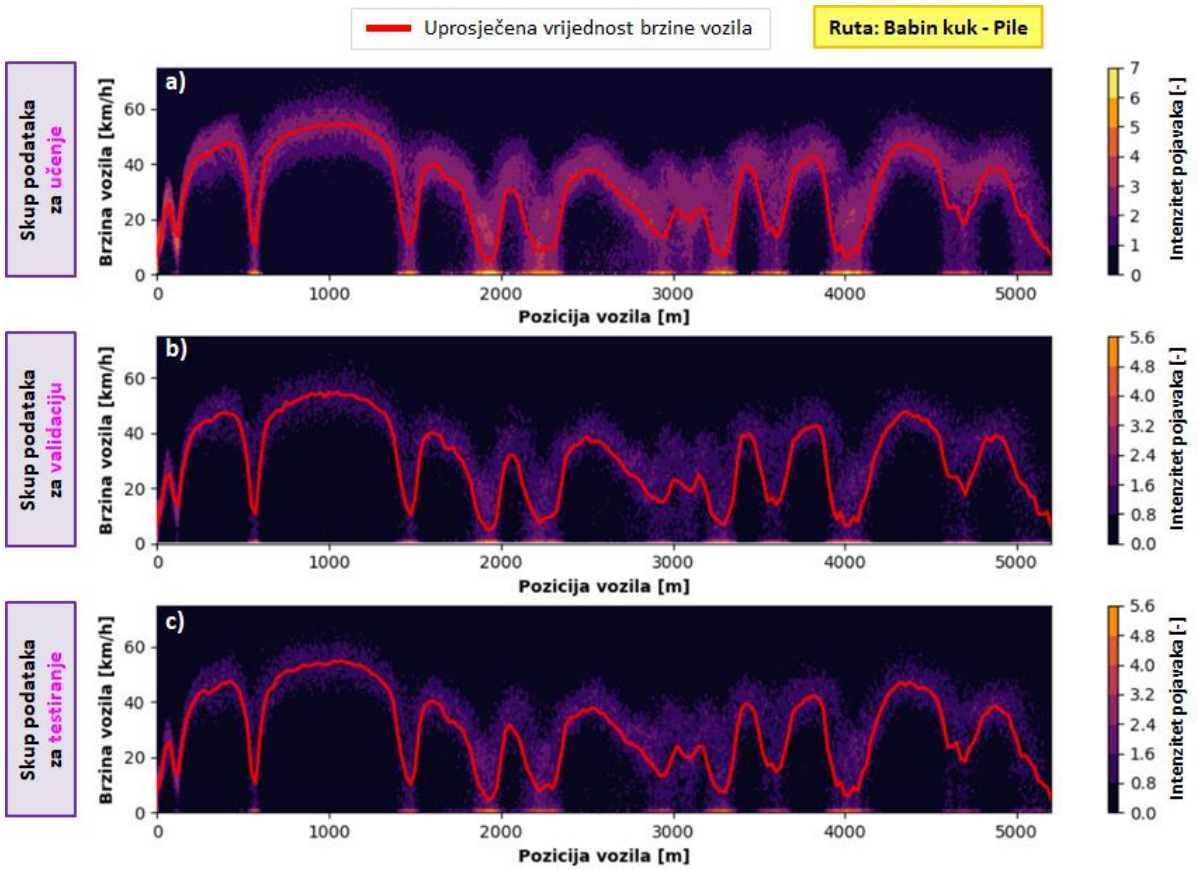
Nadalje, provodi se provjera pokrivaju li izdvojeni skupovi uzoraka za učenje, validaciju i testiranje NN-STC modela, koji su sačinjeni od nasumično izdvojenih uzoraka sadržanih u skupu $\mathcal{U}$, u dovoljnoj mjeri rasipanja brzina vozila duž razmatrane rute (tj. razmatrani prostor stanja). U tu su svrhu na slici 7-5 iscrtana rasipanja brzine vozila u ovisnosti o poziciji vozila duž razmatrane rute za sva tri skupa podataka, gdje je skalom u boji označen intenzitet/učestalost pojavaka određenih diskretnih vrijednosti brzine. Usporedbom slike 7-5a sa slikom 2-11a vidljivo je kako skup uzoraka za učenje NN-STC u najvećoj mjeri opisuje stohastiku prometnih uvjeta razmatrane rute, dok skupovi za validaciju i testiranje modela solidno odražavaju rasipanje snimljenih podataka. Time se potvrđuje valjanost rezultirajućih reduciranih skupova podataka i opravdava njihova daljnja primjena u svrhu daljnjeg razvoja i ispitivanja NN-STC modela.



Slika 7-3: Prikaz izdvojenih slučajnih uzoraka brzine vozila v_n na primjeru jednog snimljenog voznog ciklusa, odnosno pripadnom profilu brzine vozila u domeni a) vremena i b) pozicije.



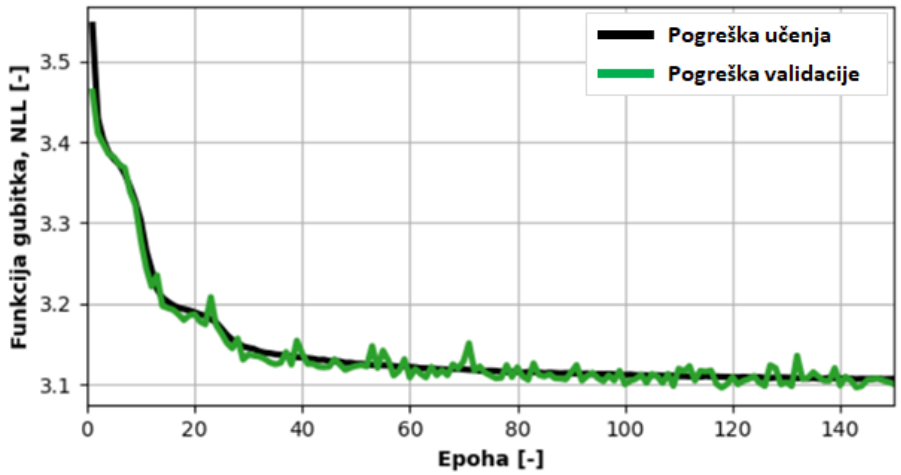
Slika 7-4: Histogram ukupnog broja izdvojenih uzoraka po voznom ciklusu, izračunat za skup svih validnih snimljenih voznih ciklusa.



Slika 7-5: Toplinske mape rasipanja snimljenih brzina vozila v_k duž rute u domeni pozicije vozila, sadržanih u skupu podataka za a) učenje, b) validaciju i c) testiranje NN-STC modela, uključujući po poziciji uprosječene vrijednosti brzine vozila (crvena linija).

7.1.3. Rezultati učenja, validacije i testiranja prediktivnog modela

Napredak funkcije gubitka NLL definirane izrazom (7-2), odnosno pogreške učenja i validacije kroz epohe/iteracije učenja NN-STC modela, prikazana je na slici 7-6. Odsutnost jaza između krivulje pogreške učenja i validacije ukazuje kako model posjeduje dobra generalizacijska svojstva, odnosno, u stanju je precizno predvidjeti izlazne vrijednosti i za prethodno neviđene podatke. Zasićenje validacijske pogreške nastupa već nakon 60. epohe, dok se minimalna vrijednost od 3.096 doseže u 117. epohi. Iako NLL kao funkcija gubitka nema direktnu interpretaciju, vrijedi pravilo da niže vrijednosti predstavljaju bolje rješenje (budući da je NLL = 0 za idealno predviđanje klase). Stoga se kao konačni parametri tj. težinski koeficijenti NN-STC modela usvajaju vrijednosti dobivene za slučaj minimuma validacijske pogreške.



Slika 7-6: Krivulje pogreške učenja (crna linija) i validacije (zelena linija) kroz epohe učenja za NN-STC model.

Nadalje, provedena je analiza točnosti predviđanja brzine vozila NN-STC modela na skupu podataka za testiranje. Pritom se za svako predviđanje uvjetne razdiobe brzine vozila duž prijeđenog puta $\hat{\boldsymbol{p}}_k = \mathbb{P}(\tilde{v}_j \mid d_k)$ izračunavaju srednje/očekivane vrijednosti ($\hat{\mu}_k$) i standardna odstupanja ($\hat{\sigma}_k$), a koje predstavljaju mjere kvantifikacije te su definirane kao [84]

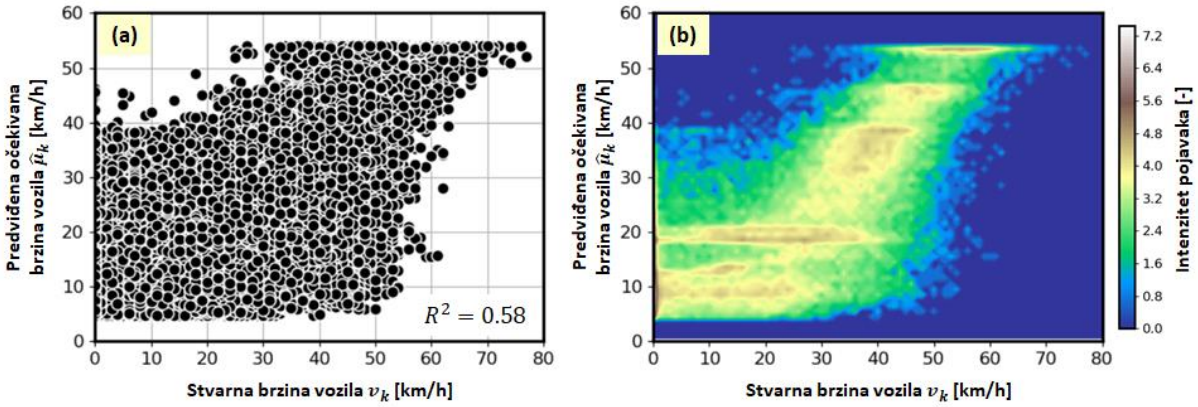
$$\hat{\mu}_k = E[\tilde{\boldsymbol{v}}] = \sum_j \tilde{v}_j \cdot \hat{p}_{k,j}, \tag{7-3}$$

$$\hat{\sigma}_k = \sqrt{E[(\tilde{\boldsymbol{v}} - \hat{\mu}_k)^2]} = \sqrt{\sum_j (\tilde{v}_j - \hat{\mu}_k)^2 \cdot \hat{p}_{k,j}}\,. \tag{7-4}$$

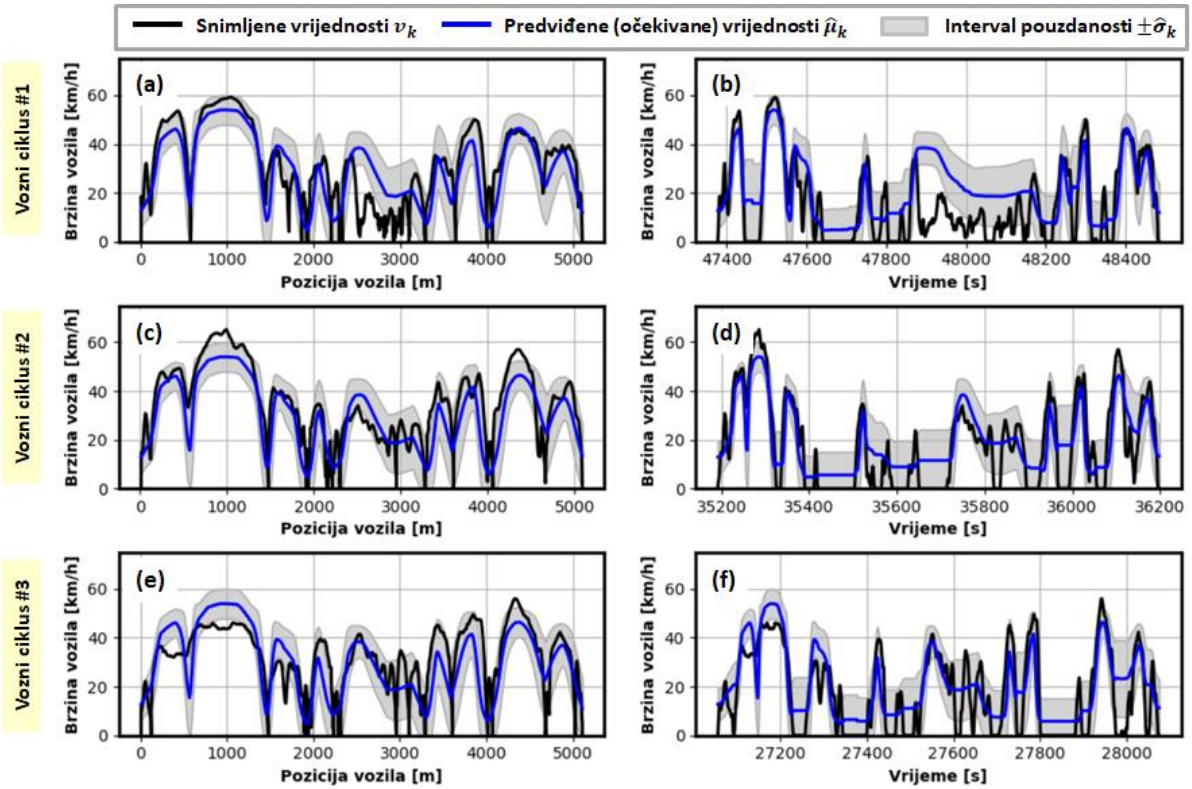
Pored toga, performanse NN-STC modela dodatno se kvantificiraju R^2 vrijednošću između snimljene vrijednosti brzine vozila v_k i očekivanja predviđanja razdiobe brzina $\hat{\mu}_k$ (vidi izraz (4-12)).

Slika 7-7 prikazuje ovisnost očekivanja predviđanja brzine vozila $\hat{\mu}_k$ u odnosu na stvarne/snimljene vrijednosti brzine vozila v_k, zajedno s pripadnom toplinskom mapom. Iz ovih rezultata je vidljivo kako postoje značajna rasipanja između stvarnih v_k i predviđenih (očekivanih) vrijednosti $\hat{\mu}_k$. Ovakve karakteristike NN-STC modela su i očekivane, budući da model predviđa čitavu vjerojatnosnu razdiobu brzine vozila $\hat{\boldsymbol{p}}_k$ za odgovarajuću poziciju vozila (slika 7-1), a ne individualne vrijednosti brzine vozila v_k koje se u određenoj mjeri razlikuju između snimljenih voznih ciklusa. Ovu tvrdnju dodatno potkrjepljuje rezultirajuća R^2 vrijednost od 0.58, koja ukazuje na loše performanse predviđanja individualnih profila brzine vozila. Primjeri predviđanja NN-STC modela prikazani su na slici 7-8 za slučaj tri nasumično

izdvojena snimljena vozna ciklusa, opisana pripadnim profilima brzine u domeni vremena (lijevi stupac) i pozicije (desni stupac).



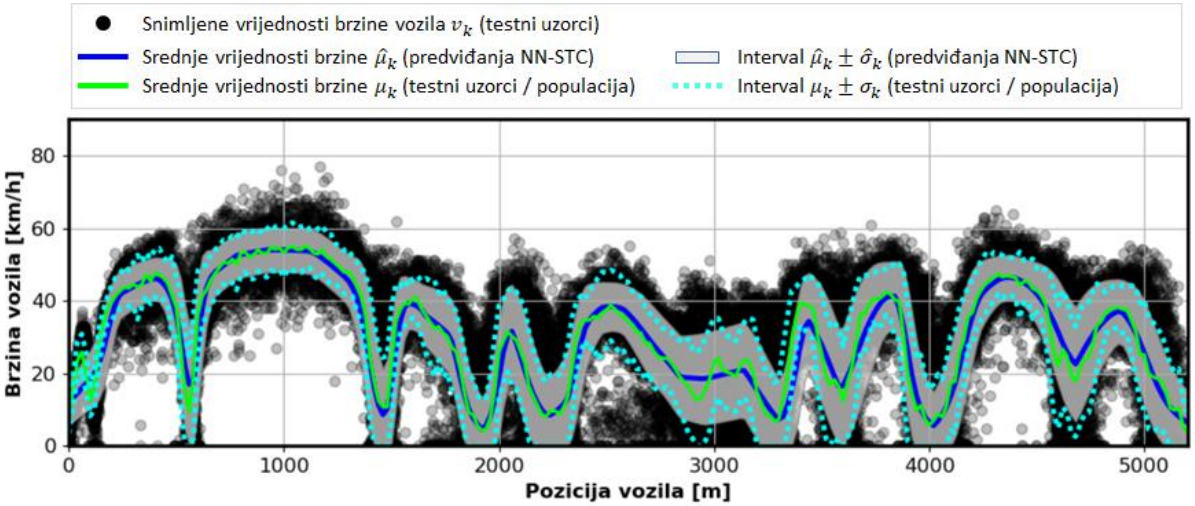
Slika 7-7: a) *Predviđene srednje vrijednosti brzine vozila μ_k dobivene primjenom NN-STC modela na temelju testnog skupa podataka u odnosu na snimljene vrijednosti brzine vozila v_k, te b) pripadna toplinska mapa.*



Slika 7-8: Primjeri usporedbe predviđenih i snimljenih profila brzine vozila za tri vozna ciklusa prikazani u domeni pozicije (lijevi stupac) i vremena (desni stupac).

Stoga je za potrebe daljnjeg ispitivanja NN-STC modela potrebno provesti analizu točnosti predviđanja u smislu odabranih statističkih indeksa vezanih za razdiobe brzine vozila duž rute. U tu se svrhu predviđene očekivane vrijednosti $\hat{\mu}_k$ i standardna odstupanja $\hat{\sigma}_k$ brzine vozila

uspoređuju s onima dobivenima za cjelokupni skup testnih uzoraka μ_k i σ_k, koji se izračunavaju zasebno za svaki uzastopni diskretni interval pozicije vozila duljine 20 m u rasponu od 0 do 5200 m. Time se dobiva ukupni broj od 260 parova očekivanih vrijednosti $(\hat{\mu}_k,\ \mu_k)$ i standardnih devijacija $(\hat{\sigma}_k, \sigma_k)$ brzine vozila, koji su prikazani na slici 7-9 zajedno s rasipanjem individualnih snimljenih vrijednosti brzine vozila v_k.



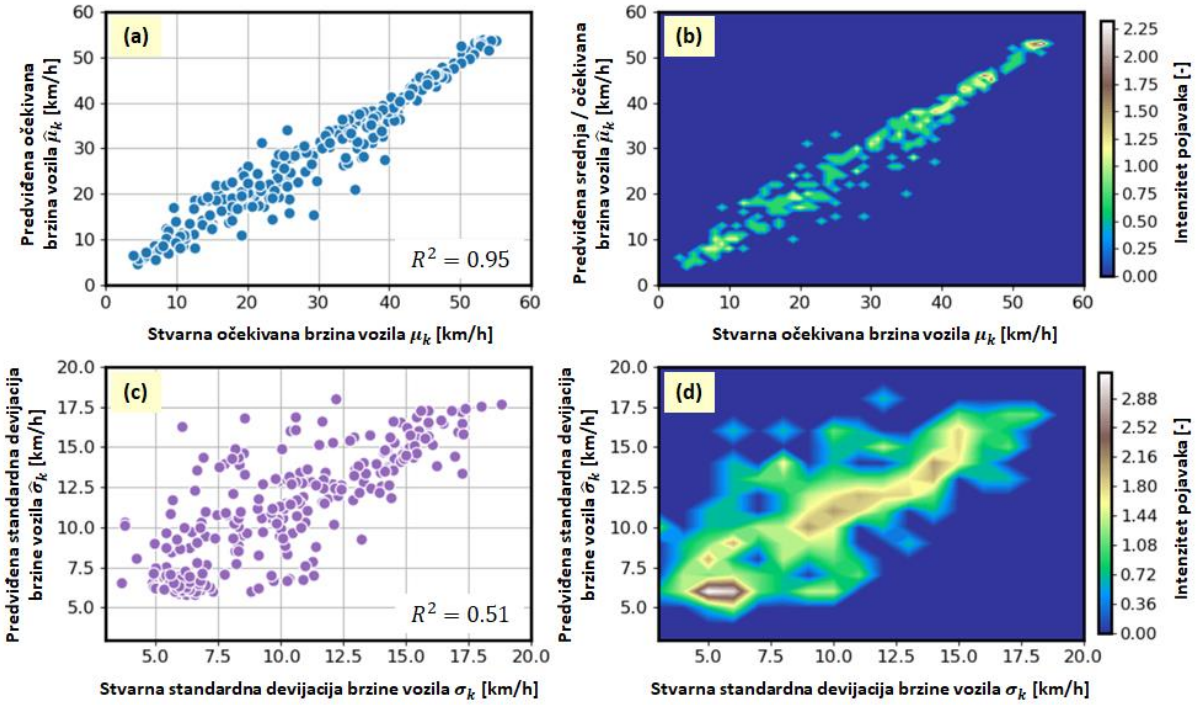
Slika 7-9: Predviđeni i stvarni profili srednje vrijednosti brzine vozila, zajedno s pripadnim intervalima pouzdanosti i rasipanjem snimljenih vrijednosti brzine vozila, sve dobiveno za testni skup podataka.

Vizualnom inspekcijom slike 7-9 može se preliminarno ustanoviti kako predviđeni profili $\hat{\mu}_k$ i $\hat{\sigma}_k$ u solidnoj mjeri rekonstruiraju profile μ_k i σ_k testne populacije (tj. snimljenih uzoraka). Međutim, značajno bolji, izravni uvid dobiva se iscrtavanjem parova $(\hat{\mu}_k, \mu_k)$ i $(\hat{\sigma}_k, \sigma_k)$, kako je dano na slici 7-10. U slučaju očekivanih vrijednosti $(\hat{\mu}_k, \mu_k)$ vidljiva je izvrsna podudarnost (slike 7-10a i 7-10b), koja se potvrđuje visokom R^2 vrijednošću od 0.95. S druge strane, kod standardnih devijacija $(\hat{\sigma}_k, \sigma_k)$ podudarnost je primjetno slabija (slike 7-10c i 7-10c), što se ogleda i u osjetno nižoj R^2 vrijednosti od 0.51.

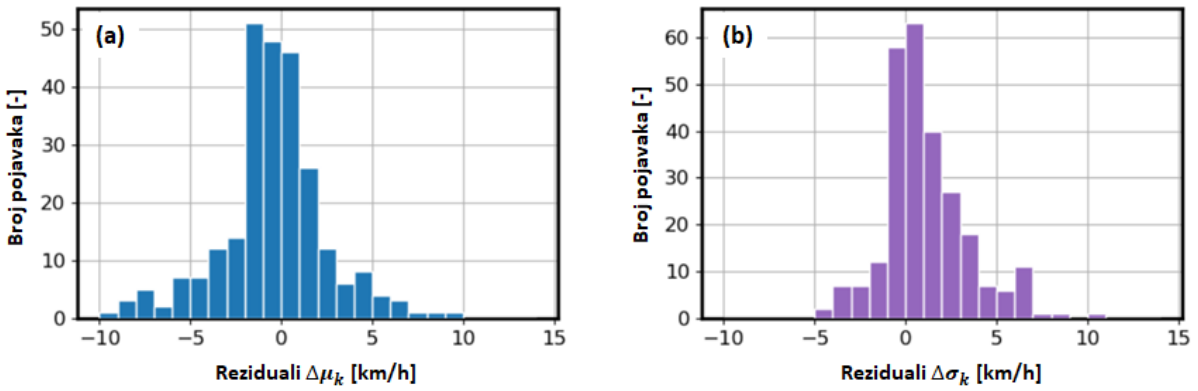
Nadalje, na slici 7-11 prikazani su histogrami odstupanja predviđenih u odnosu na stvarne vrijednosti: $\Delta\mu_k = \hat{\mu}_k - \mu_k$ i $\Delta\sigma_k = \hat{\sigma}_k - \sigma_k$, dok je pripadna statistika dana u tablici 7-1. Srednje vrijednosti ovih odstupanja $\Delta\mu_k$ i $\Delta\sigma_k$ nalaze se u intervalu ± 1 km/h, s rasipanjem od oko $2-3$ km/h, u čijem rasponu se nalazi 75% reziduala.

Dodatna točnost predviđanja razdiobe brzina mogla bi se ostvariti boljom uvjetovanošću ulaznih podataka, npr. dodavanjem trenutne brzine vozila (ili povijesti brzina), doba dana, dana u tjednu i sl. kao dodatnih ulaza u model. Međutim, takav stohastički model nije moguće jednoznačno validirati u odnosu na rasipanja snimljenih podataka. Stoga se u sljedećem

potpoglavlju predlaže deterministički model za precizno predviđanje profila brzine vozila u neposrednoj budućnosti, odnosno, na pomičnom diskretnom vremenskom horizontu, pri čemu se uzimaju u obzir spomenuti dodatni ulazi u model.



Slika 7-10: Ovisnost predviđenih očekivanih o stvarnim srednjim vrijednostima brzine vozila (gornji red) kao i predviđenih u odnosu na stvarne standardne devijacije (donji red).



Slika 7-11: Histogrami odstupanja predviđenih u odnosu na stvarne a) srednje brzine i b) standardne devijacije brzine.

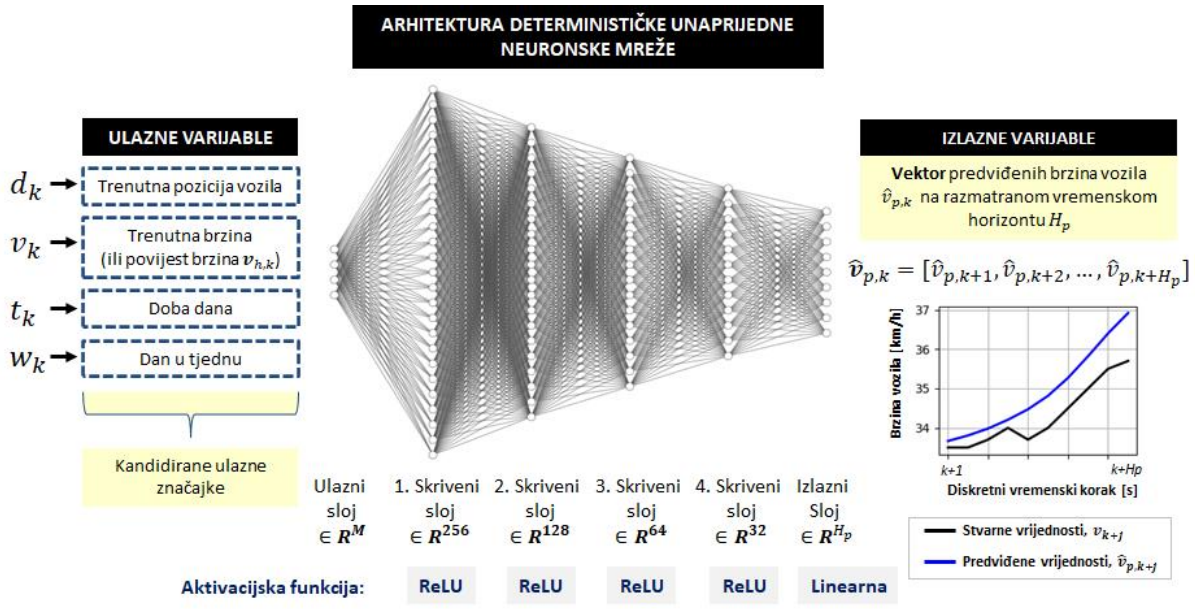
Tablica 7-1: Statistika odstupanja predviđenih u odnosu na stvarne srednje vrijednosti brzine i standardne devijacije brzine.

	Srednja vrijednost	**Standardna devijacija**	**25. percentil**	**50. percentil**	**75. percentil**
Reziduali $\Delta\mu_k$ [km/h]	-0.59	3.17	0.70	1.42	2.92
Reziduali $\Delta\sigma_k$ [km/h]	1.07	2.31	0.52	1.11	2.68

7.2. Deterministički model predviđanja brzine vozila

7.2.1. Modeliranje unaprijedne višeslojne neuronske mreže s deterističkim izlazom

Zbog ograničenih mogućnosti predviđanja individualnih vrijednosti brzine vozila tj. voznih ciklusa kod prethodno definiranog NN-STC modela, predlaže se deterministički model koji je znatno bolje uvjetovan ulaznim podacima te je namijenjen za precizno kratkoročno predviđanje profila brzine vozila na horizontu u budućnosti. Ovaj novopredloženi model, dalje označavan kao NN-DET, također se temelji na unaprijednoj višeslojnoj NN (poglavlje 4), čija je arhitektura prikazana na slici 7-12.



Slika 7-12: Predložena arhitektura determinističke unaprijedne višeslojne neuronske mreže (NN-DET) za predviđanje profila brzine vozila na horizontu u budućnosti $\hat{\boldsymbol{v}}_{p,k}$ na temelju trenutne pozicije vozila d_k, trenutne brzine v_k (ili povijesti brzina $\boldsymbol{v}_{h,k}$), doba dana t_k i dana u tjednu w_k.

Ulazi u NN-DET model uključuju podatke o trenutnoj poziciji vozila (d_k), trenutnoj brzini vozila v_k (ili povijesti brzina $\boldsymbol{v}_{h,k}$), doba dana t_k, te danu u tjednu w_k. Izlaz iz NN-DET modela je vektor predviđenih brzina vozila $\widehat{\boldsymbol{v}}_{p,k} = \left[\hat{v}_{p,k+j}\right]; j = 1,2, \dots, H_p$, na pomičnom vremenskom horizontu H_p. Pritom indeks k označava diskretni vremenski korak stanja ulaza $X_k = \{d_k, v_k, t_k, w_k\}$. NN-DET model sačinjen je od ulaznog sloja koji prima $M = 4$ ulazne varijable, četiri potpuno povezana skrivena sloja, koji redom sadrže 256, 128, 64 i 32 skrivena neurona, te izlaznog sloja od H_p elementa, koji predstavlja dodatan hiperparametar čiju je optimalnu vrijednost potrebno odrediti/ispitati zasebnom analizom. Ukoliko se kao drugi ulaz umjesto trenutne brzine v_k koristi vektor povijesnih brzina vozila $\boldsymbol{v}_{h,k} = \left[v_{h,k-H_h+\tilde{\jmath}}\right]; \tilde{\jmath} = 1,2, \dots, H_h$, tada se broj ulaza u NN-DET povećava za $H_h - 1$, čime se ujedno povećava i ukupni broj podesivih parametara modela. Za potrebe bolje generalizacije NN-DET modela, provedena je regularizacija parametara tj. težinskih koeficijenata modela na isti način kako je opisano za NN-STC model, uz faktor regularizacije $\lambda = 5 \cdot 10^{-4}$.

Skriveni slojevi koriste ReLU aktivacijsku funkciju, dok izlazni sloj koristi linearnu aktivacijsku funkciju. Kao funkcija gubitka koju treba minimizirati odabrana je MSE između predviđenih $\widehat{\boldsymbol{v}}_{p,k}$ i stvarnih/snimljenih $\boldsymbol{v}_{p,k}$ vektora brzina u budućnosti, odnosno, na diskretnom vremenskom horizontu H_p:

$$MSE_v = \frac{1}{N}\sum_{k=1}^{N}\frac{1}{H_p}\sum_{j=1}^{H_p}\left(\widehat{\boldsymbol{v}}_{p,k,j} - \boldsymbol{v}_{p,k,j}\right)^2, \tag{7-5}$$

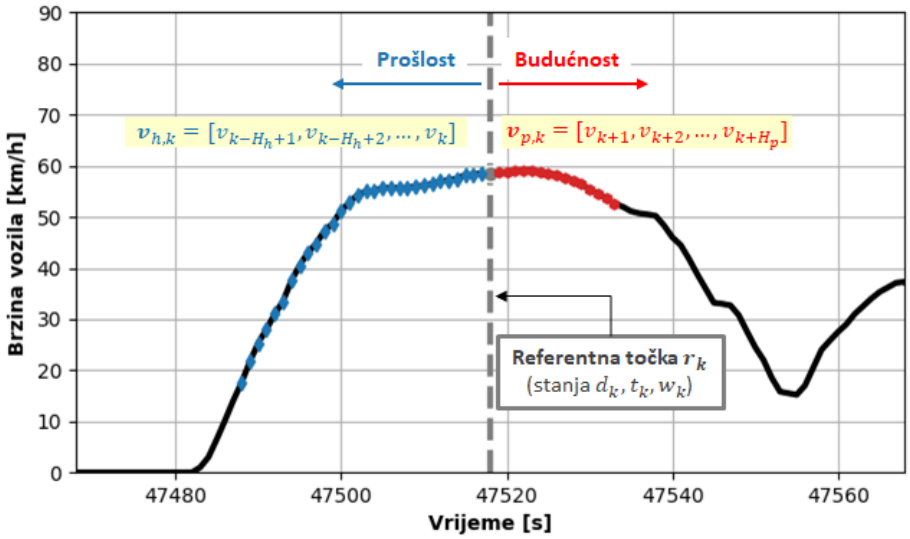
gdje N označava ukupan broj uzoraka korišten za učenje NN-DET modela.

Realizacija NN-DET modela izvodi se unutar *Python* programskog okruženja korištenjem *Keras* modula s *Tensorflow* modulom kao pozadinom, s istim postavkama kao i u slučaju NN-STC modela.

7.2.2. Priprema podataka za učenje, validaciju i testiranje prediktivnog modela

Za potrebe pripreme podataka za učenje, validaciju i testiranje NN-DET modela koristi se prethodno izdvojeni reducirani skup $\mathcal{U}$ od 382.553 uzoraka, čiji je postupak generiranja opisan u potpoglavlju 7.1. Pritom svaki izdvojeni uzorak vremena $t_k \in \boldsymbol{u}_k$ predstavlja referentnu točku r_k od koje se uzima po H_h brzina iz povijesti (vektor $\boldsymbol{v}_{h,k}$), te H_p brzina u budućnosti (vektor $\boldsymbol{v}_{p,k}$), na način kako je prikazano na slici 7-13. Postupak uzorkovanja povijesnih te budućih brzina ($\boldsymbol{v}_{h,k}$, $\boldsymbol{v}_{p,k}$) se ponavlja za svaku k-tu referentnu točku sadržanu u skupu $\mathcal{U}$, čime se dobiva konačni skup ulaznih podataka $\mathcal{X} \in \mathbb{R}^{382.553 \text{ x } (H_h+3)}$, i izlaznih vrijednosti brzine vozila

$\mathcal{Y} \in \mathbb{R}^{382.553 \times H_p}$. Skupovi $\mathcal{X}$ i $\mathcal{Y}$ dalje su skalirani prema principu min-max normalizacije i podijeljeni na skup za učenje (70% uzoraka), validaciju (15% uzoraka) i testiranje (15% uzoraka) NN-DET modela.



Slika 7-13: Ilustrativni prikaz postupka generiranja ulazno-izlaznih podataka (tj. skupova $\mathcal{X}$ i $\mathcal{Y}$) za učenje, validaciju i testiranje NN-DET modela.

7.2.3. Definiranje pokazatelja kvalitete predviđanja

U svrhu kvantificiranja točnosti predviđanja profila brzine vozila primjenom NN-DET modela uvode se razne varijante korijena srednje kvadratne pogreške (RMSE). Pritom se RMSE_k između individualno predviđenog vektora $\hat{\boldsymbol{v}}_{p,k}$ i stvarnog/snimljenog vektora $\boldsymbol{v}_{p,k}$ brzina vozila u budućnosti (duljine H_p) izračunava prema

$$\text{RMSE}_k = \sqrt{\frac{1}{H_p}\sum_{j=1}^{H_p}\left(\hat{v}_{p,k+j} - v_{p,k+j}\right)^2}, \tag{7-6}$$

gdje j označava diskretni vremenski korak predviđanja. Prosječna vrijednost $\overline{\text{RMSE}}_i$ za čitavi i-ti vozni ciklus izračunava se prema

$$\overline{\text{RMSE}}_i = \frac{1}{N_i}\sum_{k=1}^{N_i}\text{RMSE}_k, \tag{7-7}$$

gdje N_i predstavlja ukupni broj snimljenih uzoraka brzine vozila ($v_{p,1}$, $v_{p,2}$, ..., $v_{p,k}$) sadržanih u i-tom voznom ciklusu.

Ukoliko se razmatra točnost višestrukih (ili svih) predviđanja dobivenih za j-ti diskretni vremenski korak duž horizonta predviđanja H_p, onda se vrijednost $\widehat{\text{RMSE}}_j$ izračunava prema

$$\widehat{\mathrm{RMSE}}_j = \sqrt{\frac{1}{N_j}\sum_{k=1}^{N_j}\left(\hat{v}_{p,k+j} - v_{p,k+j}\right)^2}, \tag{7-8}$$

gdje N_j predstavlja ukupni broj razmatranih snimljenih uzoraka brzine vozila ($v_{p,k}$).

7.2.4. Određivanje konačnog skupa ulaznih značajki

U svrhu finog-podešavanja NN-DET modela provodi se analiza utjecaja različitih kombinacija ulaznih značajki na točnost predviđanja profila brzine vozila, za čije se ocjenjivanje koristi korijen vrijednosti funkcije gubitka MSE_v definirane izrazom (7-5) i izračunate za slučaj validacijskog skupa podataka (tj. koristi se $RMSE_v = \sqrt{MSE_v}$ kako bi se mjerna jedinica u prikazu rezultata svela na km/h). Glavni cilj ove analize je odrediti konačni podskup relevantnih ulaznih značajki za formiranje NN-DET modela. Također, od interesa je dobiti uvid u doprinos uključivanja trenutne pozicije vozila kao dodatnog ulaza pored podataka o brzini vozila, koji se tipično koriste u literaturi. U tu svrhu, razmatrane su sljedeće kombinacije ulaznih značajki:

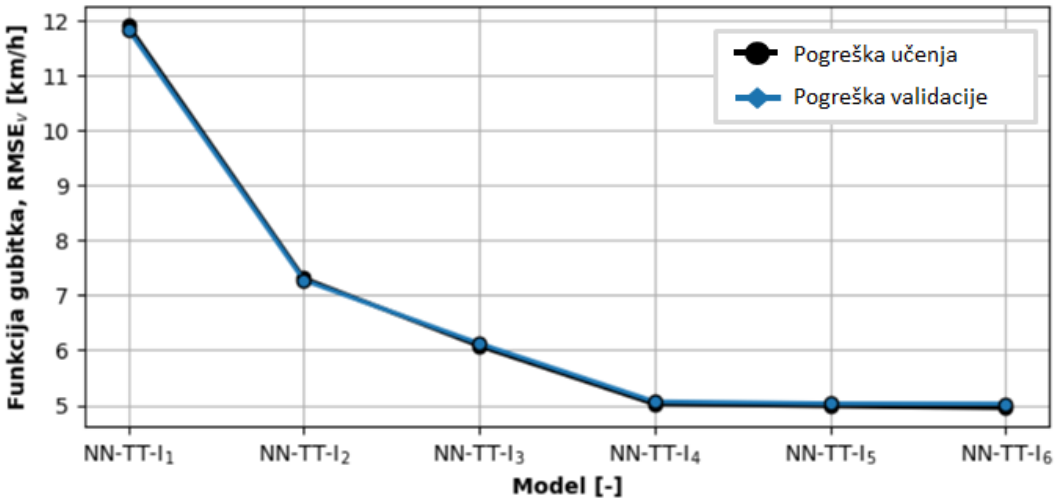
1) Trenutna pozicija vozila d_k,
2) Trenutna pozicija vozila d_k i trenutna brzina vozila v_k,
3) Povijest brzina vozila $\boldsymbol{v}_{h,k}$,
4) Trenutna pozicija vozila d_k i povijest brzina $\boldsymbol{v}_{h,k}$,
5) Trenutna pozicija vozila d_k, povijest brzina $\boldsymbol{v}_{h,k}$ i doba dana t_k,
6) Trenutna pozicija vozila d_k i povijest brzina $\boldsymbol{v}_{h,k}$, doba dana t_k i dan u tjednu w_k.

Za svaku od gore navedenih kombinaciju ulaznih značajki provedeno je zasebno učenje NN-DET modela prikazanog na slici 7-12, koji se označava kao NN-DET-I_n, gdje *n* označava redni broj ulazne kombinacije kako je navedeno gore. Pritom se primjenjuje fiksni horizont predviđanja $H_p = 10$ s i duljina intervala povijesti brzina $H_h = 20$ s.

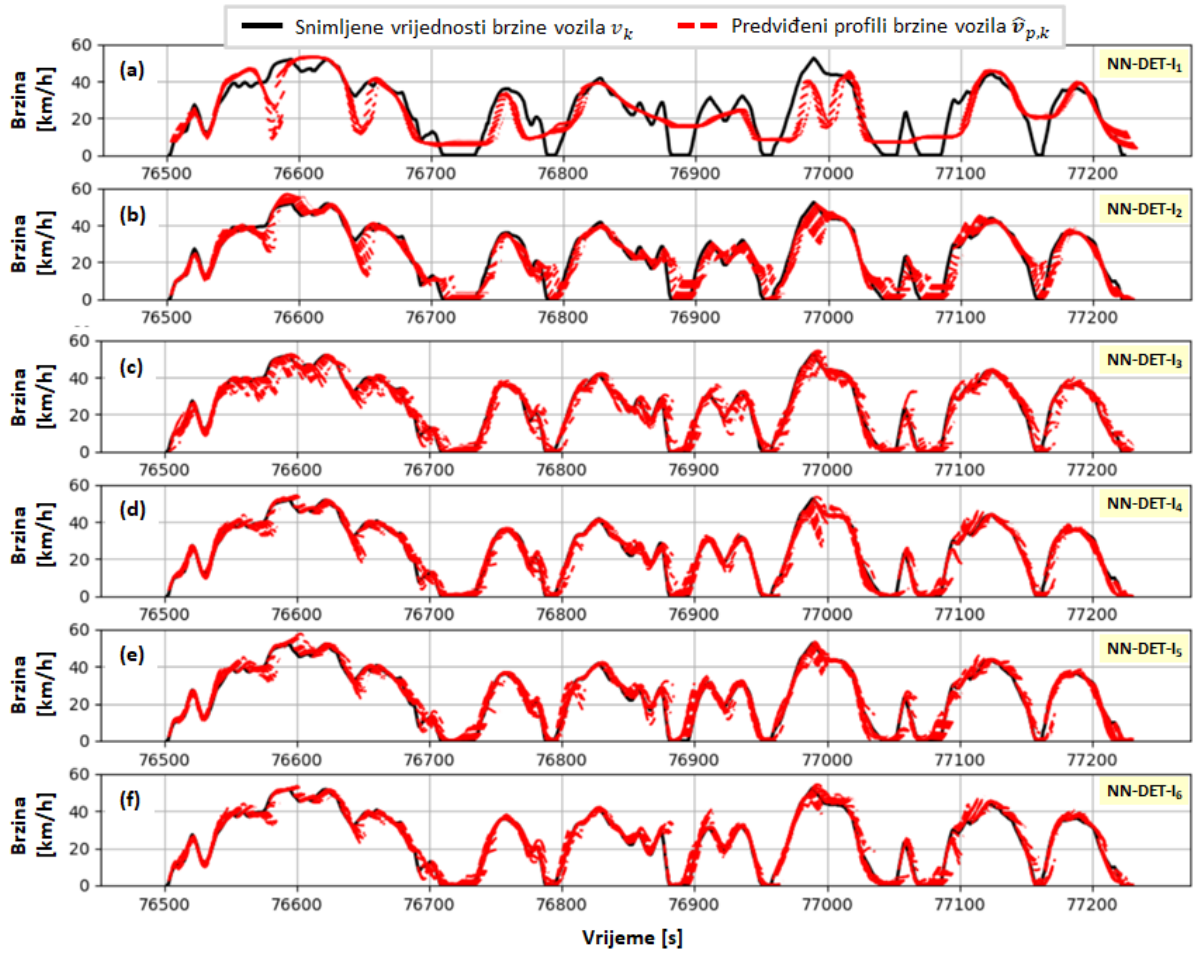
Na slici 7-14 prikazane su vrijednosti funkcije gubitka $RMSE_v$, gdje je crnom bojom označena pogreška učenja, a plavom pogreška validacije. Vizualnom inspekcijom slike 7-14 može se ustanoviti kako zasićenje točnosti predviđanja profila brzine vozila nastupa za slučaj NN-DET-I_4 modela ($RMSE_v = 5.066$ km/h), čime se jasno ukazuje na neznatni utjecaj doba dana t_k i dana u tjednu w_k kao dodatnih ulaznih značajki na samu točnost predviđanja modela. Preciznije, $RMSE_v$ doseže vrijednost od 5.030 km/h za NN-DET-I_5 te 5.027 km/h za NN-DET-I_6, što predstavlja relativno smanjenje pogreške tj. poboljšanje točnosti za manje od 1%. Najlošija točnost predviđanja dobiva se za slučaj NN-DET-I_1 model ($RMSE_v = 11.84$ km/h),

koji koristi samo trenutnu pozicija vozila d_k kao ulaznu značajku. Razlog tome je nedostatna uvjetovanost modela ulaznim podacima, kao što je prethodno potvrđeno kod predviđanja individualnih profila brzine vozila za slučaj NN-STC modela (potpoglavlje 7-1). Dodavanjem trenutne brzine v_k pored pozicije vozila d_k (NN-DET-I_2 model) ostvaruje se značajan, međutim i dalje nedovoljan pad $RMSE_v$ s 11.84 km/h na 7.28 km/h (≈40%). Zanimljiv rezultat dobiva se za slučaj NN-DET-I_3 modela, koji, unatoč korištenju same povijesti brzine vozila $\boldsymbol{v}_{h,k}$ kao ulaza, postiže točnost predviđanja blisku NN-DET-I_4 modelu koji dodatno koristi i poziciju vozila ($RMSE_v = 6.12$ km/h u odnosu na $RMSE_v = 5.066$ km/h). Usporedbom rezultata NN-DET-I_3 i NN-DET-I_4 modela otkriva se relativni utjecaj trenutne pozicije vozila d_k na poboljšanje točnosti predviđanja modela od 17.2%. Kako bi se dobio zorniji uvid u kvalitetu predviđanja profila brzine vozila svih NN-DET-I_n modela, na slici 7-15 iscrtani su rezultati individualnih predviđanja $\widehat{\boldsymbol{v}}_{p,k}$ na primjeru jednog testnog voznog ciklusa.

Temeljem provedene analize utjecajnosti ulaznih značajki na točnost predviđanja brzine vozila u neposrednoj budućnosti može se zaključiti kako NN-DET-I_6 model općenito predstavlja najbolji odabir zbog najniže vrijednosti $RMSE_v$. Međutim, kako očito postoje zanemarive varijacije brzine vozila duž rute u ovisnosti o doba dana t_k i danu u tjednu w_k (vidi također analizu u poglavlju 2), dobar praktični odabir predstavlja NN-DET-I_4 model, koji, pored gotovo identičnih performansi predviđanja kao i NN-DET-I_6 model, u manjoj mjeri reducira i složenost samog prediktivnog modela (ukupni broj podesivih parametara NN-DET-I_4 modela iznosi 48.938 naspram 49.194 za slučaj NN-DET-I_6 modela). Važno je za primijetiti da niti jedna od kombinacija ulaznih značajki ne rezultira prenaučenosti modela, što je vidljivo iz izostanka jaza između pogreške učenja i validacije (tj. plave i crne krivulje na slici 7-14).



Slika 7-14: Rezultirajuće vrijednosti funkcije gubitka $RMSE_v$ dobivene za svaku kombinaciju ulaznih značajki NN-DET-I_n modela te za skupove podataka za učenje (crna krivulja) i validaciju (plava krivulja).



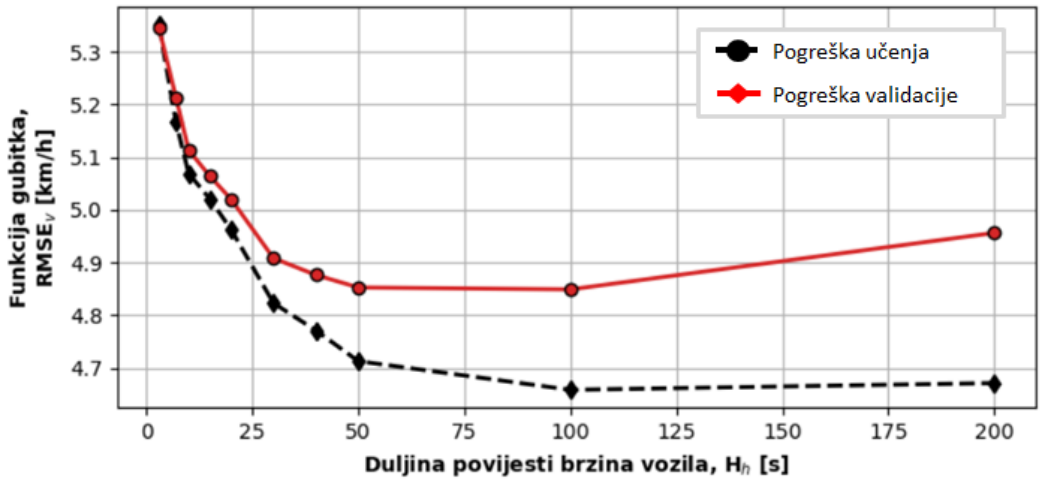
Slika 7-15: Predviđeni profili brzine vozila (crvena isprekidana linija) u odnosu na snimljene vrijednosti (crna puna linija), prikazani na primjeru jednog testnog voznog ciklusa i za svaki NN-DET-I_n model zasebno.

7.2.5. Određivanje optimalne duljine intervala povijesti brzine vozila

U svrhu određivanja optimalne duljine intervala povijesti brzina vozila $\boldsymbol{v}_{h,k} \in \mathbb{R}^{H_h}$, ispitana je točnost predviđanja NN-DET modela za sljedeće duljine H_h: 3, 7, 10, 15, 20, 30, 40, 50, 100, te 200 s. Rezultirajući modeli se dalje označavaju kao NN-DET-H_h, gdje h označava gore navedene vremenske intervale u sekundama. Za ocjenjivanje točnosti predviđanja NN-DET-H_h modela također se koriste vrijednosti funkcije gubitka $RMSE_v$ definirane izrazom (7-5) i izračunate za slučaj validacijskog skupa podataka. Za svaki od NN-DET-H_h modela provedeno je zasebno učenje NN-DET modela ilustriranog na slici 7-12, koristeći čitavi skup kandidiranih ulaznih značajki (d_k, $\boldsymbol{v}_{h,k}$, t_k i w_k) i fiksni horizont predviđanja $H_p = 10$ s.

Vrijednosti funkcije gubitka $RMSE_v$ prikazane su na slici 7-16 za svaki NN-DET-H_h model, gdje je crnom bojom označena pogreška učenja, a crvenom pogreška validacije. Rezultati analize utjecaja parametra H_h na točnost predviđanja pokazuju kako do zasićenja validacijske pogreške modela dolazi pri $H_h = 50$ s ($RMSE_v$ od 4.85 km/h), dok u slučaju vrlo bogate povijesti brzine ($H_h = 200$ s) dolazi do povećanja validacijske pogreške $RMSE_v$ s 4.85 km/h na 4.96 km/h. Razlog tome je u prevelikoj uvjetovanosti NN-DET-H_{200} modela ulaznim podacima uslijed čega dolazi do prenaučenosti modela, što pak rezultira lošijim generalizacijskim svojstvima modela (povećava se jaz između pogreške učenja i validacije s porastom duljine povijesti brzina; slika 7-16). Ova tvrdnja dodatno se potkrjepljuje sumarnim podacima danim za skupove podataka za učenje, validaciju i testiranje svih NN-DET-H_h modela (tablica 7-2).

Temeljem provedene analize može se zaključiti kako optimalna duljina vremenskog intervala povijesti brzina iznosi $H_h = 50$ s, dok vrijednosti H_h u rasponu od 30 do 50 s predstavljaju dobru praktičnu alternativu, koja u određenoj mjeri reducira složenost NN-DET modela (tablica 7-2) i daje bolja generalizacijska svojstva (slika 7-16).



Slika 7-16: Vrijednosti funkcije gubitka $RMSE_v$ dobivene za različite duljine ulaznog vektora povijesti brzina (H_h) NN-DET-H_h modela te za skupove podataka za učenje (crna krivulja) i validaciju (crvena krivulja).

Tablica 7-2: Rezultati učenja, validacije i testiranja NN-DET-H_h modela, uključujući broj podesivih parametara modela.

Model (duljina povijesti brzina; H_h)	**Vrijednost funkcije gubitka ($RMSE_v$) [km/h]**			**Broj podesivih parametara modela**
	Skup podataka za učenje	**Skup podataka za validaciju**	**Skup podataka za testiranje**	
NN-DET-H_3	5.32	5.35	5.38	45.354
NN-DET-H_7	5.13	5.21	5.23	46.378
NN-DET-H_{10}	5.00	5.11	5.10	47.146
NN-DET-H_{15}	4.93	5.06	5.05	48.426
NN-DET-H_{20}	4.91	5.02	5.01	49.706
NN-DET-H_{30}	4.76	4.91	4.92	52.266
NN-DET-H_{40}	4.70	4.88	4.86	54.826
NN-DET-H_{50}	4.61	4.85	4.84	57.386
NN-DET-H_{100}	4.57	4.85	4.89	70.186
NN-DET-H_{200}	4.53	4.96	4.93	95.786

7.2.6. Analiza točnosti predviđanja u ovisnosti o duljini horizonta predviđanja

Točnost predviđanja individualnih profila brzine vozila $\hat{\boldsymbol{v}}_{p,k} \in \mathbb{R}^{H_p}$ ispituje se za sljedeće duljine vremenskog horizonta predviđanja H_p: 3, 7, 10, 15, 20, 50, 100, te 200 s. Pritom se za svaki horizont predviđanja H_p provodi zasebno učenje NN-DET modela opisanog slikom 7-12 (dalje označavanih kao NN-DET-H_p), pri čemu se razmatraju sljedeće kombinacije ulaza:

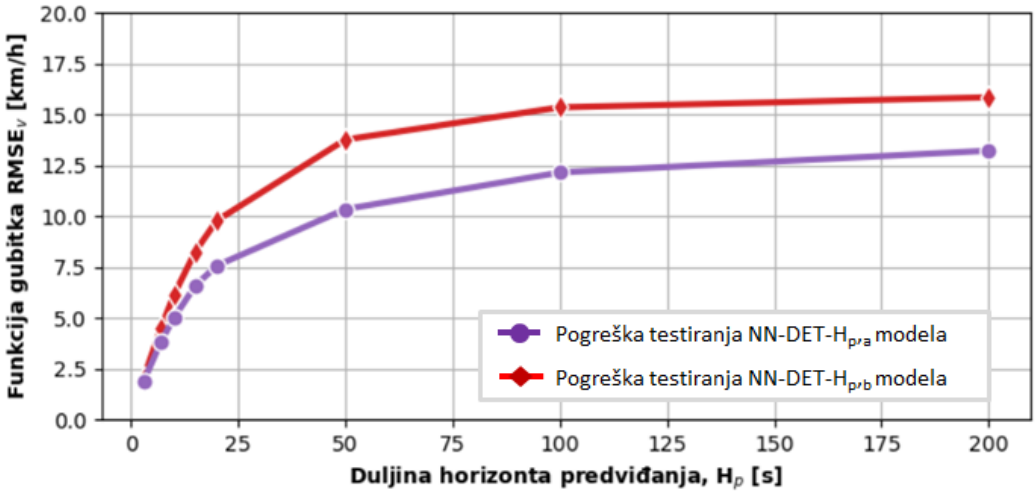
a. čitavi skup kandidiranih značajki (d_k, t_k, w_k i $\boldsymbol{v}_{h,k}$), te

b. samo povijest brzina $\boldsymbol{v}_{h,k}$,

uz fiksnu duljinu povijesti brzina $H_h = 20$ s. Ocjenjivanje točnosti predviđanja NN-DET-H_p modela provodi se temeljem:

i. vrijednosti funkcije gubitka $RMSE_v$ (korijen izraza (7-5)), izračunatih za slučaj testnog skupa podataka,
ii. pokazatelja kvalitete pojedinačnih predviđanja RMSE_k (izraz (7-6)),
iii. R^2 vrijednosti, izračunatih za svaki j-ti diskretni korak duž horizonta predviđanja ($j = 1, 2, \dots, H_p$).

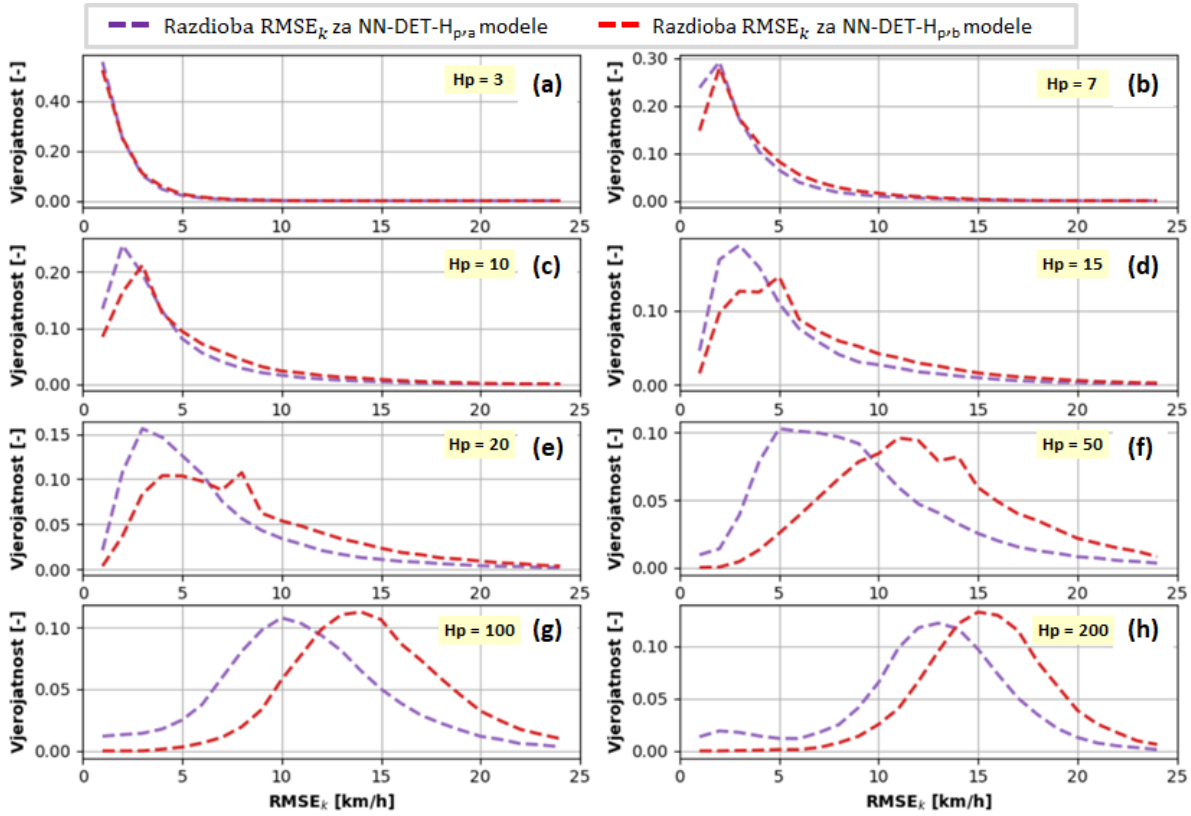
Na slici 7-17 prikazane su vrijednosti funkcije gubitka $RMSE_v$ za razne vrijednosti parametra H_p, pri čemu je ljubičastom bojom označena krivulja pogreške testiranja modela koji koristi čitav skup kandidiranih ulaznih značajki (NN-DET-$H_{p,a}$), dok je crvenom bojom označena krivulja vezana za model koji koristi samo povijest brzina kao ulaznu značajku (NN-DET-$H_{p,b}$). Ovi rezultati pokazuju kako obje funkcije gubitaka poprimaju gotovo linearni trend porasta sve do $H_p = 20$ s, te se zatim počinju zasićivati, sve dok ne postignu maksimalne vrijednosti od $RMSE_v = 13.22$ km/h u slučaju NN-DET-$H_{p,a}$ i $RMSE_v = 15.84$ km/h u slučaju NN-DET-$H_{p,b}$, u oba slučaja za $H_p = 200$ s (slika 7-17). Pritom NN-DET-$H_{p,b}$ modeli imaju u prosjeku 22.8% veći $RMSE_v$, ponovno potvrđujući značaj uključivanja trenutne pozicije vozila d_k kao dodatne ulazne značajke modela.



Slika 7-17: Vrijednosti pogreške testiranja $RMSE_v$ za različite duljine horizonta predviđanja H_p i različite prediktivne modele.

Na slici 7-18 prikazane su razdiobe izračunatih vrijednosti pokazatelja RMSE_k izračunatog za sva individualna predviđanja brzine vozila $\hat{\boldsymbol{v}}_{p,k}$ te testni skup podataka. Vidljivo je kako s

porastom horizonta predviđanja H_p razdiobe RMSE_k poprimaju sve veće očekivane vrijednosti $\mu(\mathrm{RMSE}_k)$, te uglavnom veće raspone rasipanja $\sigma(\mathrm{RMSE}_k)$, što je dodatno potkrijepljeno statistikom danom u tablici 7-3. Pored toga, model NN-DET-$H_{p,b}$, koji je slabije uvjetovan ulaznim podacima, rezultira većim individualnim pogreškama predviđanja RMSE_k u odnosu na bolje uvjetovani NN-DET-$H_{p,a}$ model, posebno za $H_p > 15$ s.



Slika 7-18: Razdiobe vrijednosti pokazatelja $RMSE_k$ dobivenih za sva individualna predviđanja profila brzine vozila i danih za različite duljine horizonta predviđanja H_p.

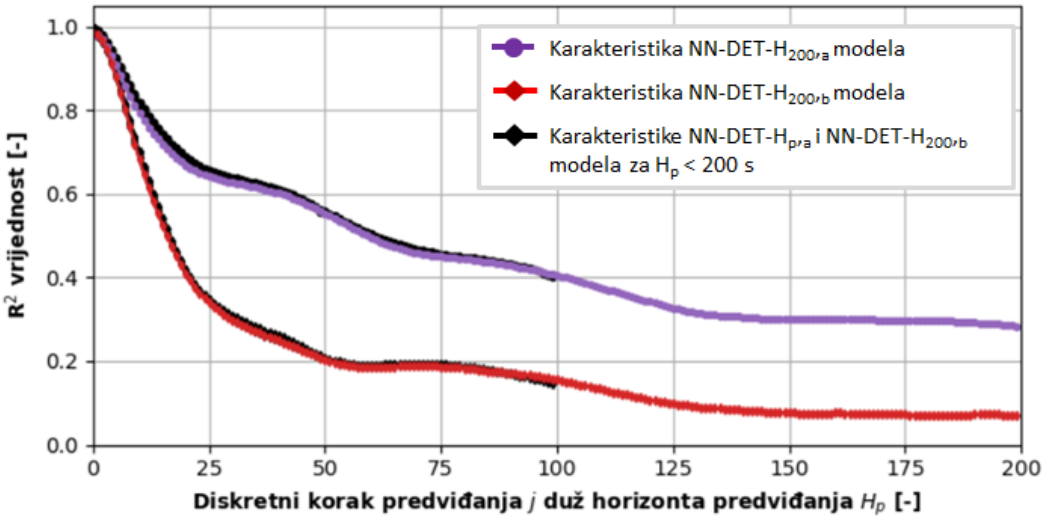
Konačno, provedena je dodatna usporedna analiza točnosti predviđanja NN-DET-$H_{p,a}$ i NN-DET-$H_{p,b}$ modela za svaki j-ti diskretni vremenski korak predviđanja duž horizonta predviđanja H_p. Karakteristične krivulje R^2 vrijednosti u ovisnosti o koraku predviđanja j, izračunate između svih predviđenih vrijednosti $\hat{\boldsymbol{v}}_{p,j}$ i stvarnih/snimljenih vrijednosti $\boldsymbol{v}_{p,j}$ brzine vozila za testni skup podataka, prikazane su na slici 7-19. Vizualnom inspekcijom slike 7-19 može se ustanoviti kako NN-DET-$H_{200,a}$ model zadržava solidna točnost predviđanja ($R^2 > 0.25$) i za predikcije na horizontu predviđanja $H_p > 100$ s, što nije slučaj kod NN-DET-$H_{200,b}$ modela, potvrđujući još jednom značaj trenutne pozicije vozila d_k u ulaznom skupu značajki. Također valja uočiti kako povećanje duljine horizonta predviđanja H_p ima sitan, gotovo neznatan utjecaj

na točnost predviđanja modela (vidi karakteristične krivulje R^2 vrijednosti dobivene učenjem odgovarajućih modela za slučajeve s $H_p < 200$ s, označene crnom bojom na slici 7-19).

Temeljem provedene analize točnosti predviđanja profila brzine vozila duž horizonta predviđanja može se zaključiti kako se za model koji koristi čitav skup kandidiranih ulaznih značajki (NN-DET-$H_{p,a}$) i $H_p < 10$ s dobiva vrlo visoka točnost predviđanja ($R^2 > 0.8$), te se dobra točnost predviđanja zadržava sve do $H_p < 40$ s ($R^2 > 0.6$).

Tablica 7-3: Statistika razdioba pokazatelja kvalitete individualnih predviđanja $RMSE_k$ za različite duljine horizonta predviđanja H_p, za slučaj NN-DET-$H_{p,a}$ modela (crna boja) i NN-DET-$H_{p,b}$ modela (crvena boja).

Model (duljina H_p)	**Srednja vrijednost $\mu(RMSE_k)$ [km/h]**	**Standardna devijacija $\sigma(RMSE_k)$ [km/h]**	**25. percentil [km/h]**	**50. percentil (medijan) [km/h]**	**75. percentil [km/h]**
NN-DET-H_3	1.28 | 1.42	1.31 | 1.41	0.42 | 0.56	0.86 | 0.94	1.68 | 1.84
NN-DET-H_7	2.67 | 3.32	2.51 | 2.83	1.04 | 1.46	1.87 | 2.36	3.41 | 4.31
NN-DET-H_{10}	3.54 | 4.54	3.15 | 3.77	1.48 | 2.01	2.58 | 3.30	4.47 | 5.97
NN-DET-H_{15}	4.84 | 6.44	3.94 | 4.67	2.19 | 3.12	3.61 | 4.97	6.13 | 8.56
NN-DET-H_{20}	5.74 | 7.96	4.32 | 5.05	2.80 | 4.24	4.53 | 6.87	7.25 | 10.4
NN-DET-H_{50}	8.53 | 12.31	4.79 | 5.10	5.09 | 8.73	7.61 | 11.58	10.81 | 14.97
NN-DET-H_{100}	10.71 | 14.30	4.45 | 4.19	7.91 | 11.49	10.37 | 13.83	13.17 | 16.60
NN-DET-H_{200}	12.07 | 15.06	4.19 | 3.28	10.15 | 12.96	12.38 | 14.96	14.60 | 17.02

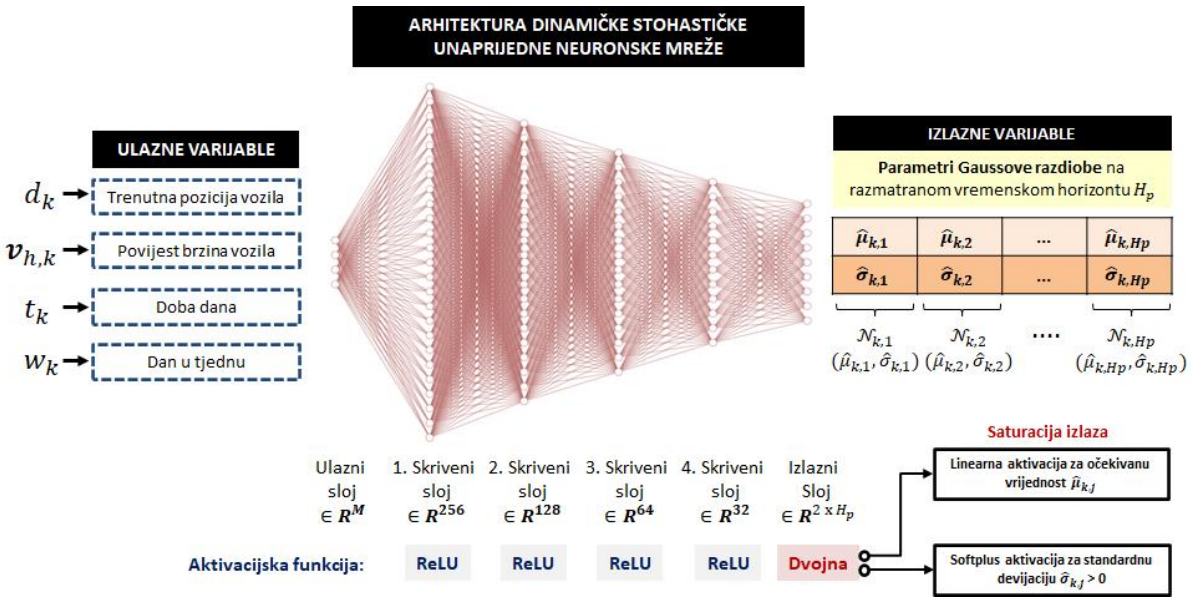


Slika 7-19: Karakteristične krivulje R^2 vrijednosti za svaki j-ti diskretni korak predviđanja duž horizonta duljine H_p i testni skup podataka.

7.3. Dinamički stohastički model predviđanja brzine vozila

7.3.1. Modeliranje, učenje i validacija unaprijedne višeslojne neuronske mreže s dinamičkim stohastičkim izlazom

S ciljem daljnjeg unaprjeđenja prethodno postavljenih NN-STC i NN-DET modela za predviđanje brzine vozila, kao ultimativno rješenje predlaže se dinamički stohastički model (dalje označavan kao NN-DSTC), koji NN-DET modelu daje dodatnu stohastičku komponentu u smislu modeliranja nesigurnosti predviđanja za svaki korak u budućnosti. Cjelokupna arhitektura NN-DSTC modela temeljena je na unaprijednoj višeslojnoj NN te je prikazana na slici 7-20. Na izlazu iz modela pretpostavljaju se normalne (Gaussove) razdiobe brzine vozila $\mathcal{N}_{k,j}(\hat{\mu}_{k,j}, \hat{\sigma}_{k,j})$, čiji se parametri $\hat{\mu}_{k,j}$ i $\hat{\sigma}_{k,j}$ podešavaju prilikom učenja modela, dok ulazi u model ostaju isti kao u slučaju NN-DET modela (vidi sliku 7-12). Time se provodi aproksimacija originalnih razdioba koje uči NN-STC (potpoglavlje 7.1) s normalnim razdiobama brzine vozila, čime se ukupna dimenzija izlaza smanjuje s $\mathbb{R}^{91 \text{ x } H_p}$ na $\mathbb{R}^{H_p \text{ x } 2}$, tj. postoje samo dva parametra po koraku predviđanja j duž pomičnog diskretnog vremenskog horizonta predviđanja $j = 1,2,\dots,H_p$. Suštinska razlika NN-DSTC u odnosu na NN-STC je u tome što osigurava predviđanje brzine vozila duž horizonta predviđanja, uključujući nesigurnost pojedinih uzoraka predviđanja uzduž horizonta.



Slika 7-20: Predložena arhitektura dinamičke stohastičke unaprijedne višeslojne neuronske mreže (NN-DSTC) za predviđanje razdioba brzine vozila na horizontu u budućnosti na temelju trenutne pozicije vozila d_k, povijesti brzina $\boldsymbol{v}_{h,k}$, doba dana t_k i dana u tjednu w_k.

Isto kao i u slučaju NN-STC i NN-DET modela, skriveni slojevi NN-DSTC modela koriste ReLU aktivacijsku funkciju, dok izlazni sloj koristi dvojnu aktivacijsku funkciju, odnosno

linearnu za očekivane vrijednosti $\hat{\mu}_{k,j}$ te *softplus* za standardne devijacije $\hat{\sigma}_{k,j}$. Razlog potonjem je potreba za zasićenjem izlaznih parametara razdiobe $\hat{\sigma}_{k,j}$ na vrijednosti veće ili jednake nuli

$$softplus\left(\hat{\sigma}_{k,j}\right) = \ln(1 + e^{\hat{\sigma}_{k,j}}) \geq 0. \tag{7-9}$$

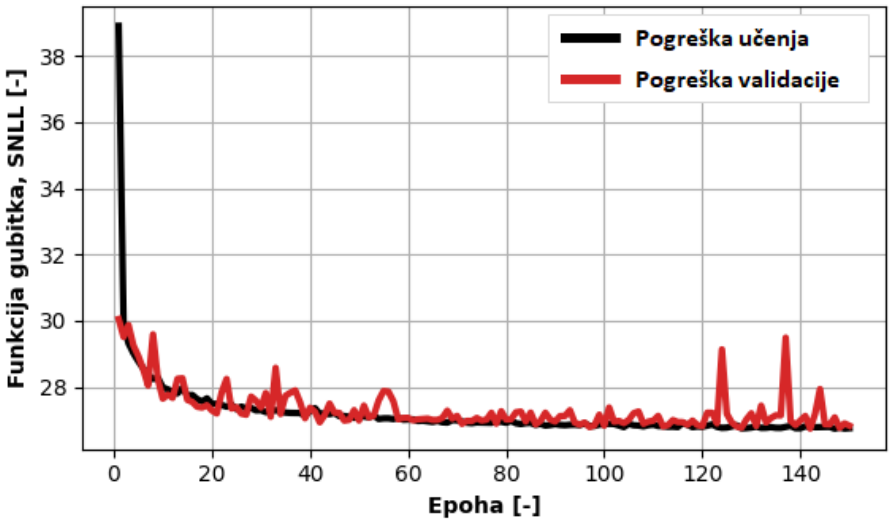
Glavni izazov kod učenja NN-DSTC modela predstavlja strukturiranje funkcije gubitka, koja se u ovom radu definira kao suma negativnih log-vjerojatnosti normalnih razdioba $\mathcal{N}_{k,j}(\hat{\mu}_{k,j}, \hat{\sigma}_{k,j})$ za svaki korak predviđanja $j \in [1,\ H_p]$:

$$\begin{aligned} SNLL &= \frac{1}{N}\sum_{k=1}^{N} - \sum_{j=1}^{H_p} \log\left(\mathcal{N}_{k,j}\left(\mathcal{Y}_k \mid \hat{\mu}_{k,j}, \hat{\sigma}_{k,j}\right)\right) \\ &= \frac{1}{N}\sum_{k=1}^{N} - \sum_{j=1}^{H_p} \log\left(\frac{1}{\sqrt{2\pi\hat{\sigma}_{k,j}^2}} \cdot e^{-\frac{1}{2}\left(\frac{(v_{k,j}-\hat{\mu}_{k,j})^2}{\hat{\sigma}_{k,j}^2}\right)}\right); \quad v_{k,j} \in \mathcal{Y}, \end{aligned} \tag{7-10}$$

gdje $\mathcal{Y}$ predstavlja skup izlaznih oznaka opisanih vrijednostima brzine vozila $v_{k,j}$ definiran u potpoglavlju 7.2, a N ukupan broj uzoraka korišten za učenje NN-DSTC modela.

Realizacija NN-DSTC modela izvodi se unutar *Python* programskog okruženja korištenjem *Keras* modula s *Tensorflow* modulom kao pozadinom, s istim postavkama kao i u slučaju NN-DET te NN-STC modela. Za potrebe učenja, validacije i testiranja NN-STC modela duljina pomičnog horizonta predviđanja H_p postavljena je na 10 s, a duljina intervala povijesti brzina H_h iznosi 20 s.

Napredak funkcije gubitka SNLL definirane izrazom (7-10), odnosno pogreške učenja i validacije kroz epohe/iteracije učenja NN-DSTC modela prikazana je na slici 7-21. Rezultati učenja pokazuju nešto veću oscilatornost pogreške validacije NN-DSTC modela u odnosu na NN-STC model (usp. slike 7-21 i 7-6). Razlog tome je veća složenost odgovarajuće funkcije cilja *SNLL*. Neovisno o tome, odsutnost jaza između krivulje pogreške učenja i validacije ukazuje kako model posjeduje dobra generalizacijska svojstva. Minimum validacijske pogreške od 26.73 doseže se u 141. epohi, čije se vrijednosti težinskih koeficijenata usvajaju kao konačne za parametriranje NN-DSTC modela.



Slika 7-21: Krivulje pogreške učenja i validacije kroz epohe učenja za NN-DSTC model.

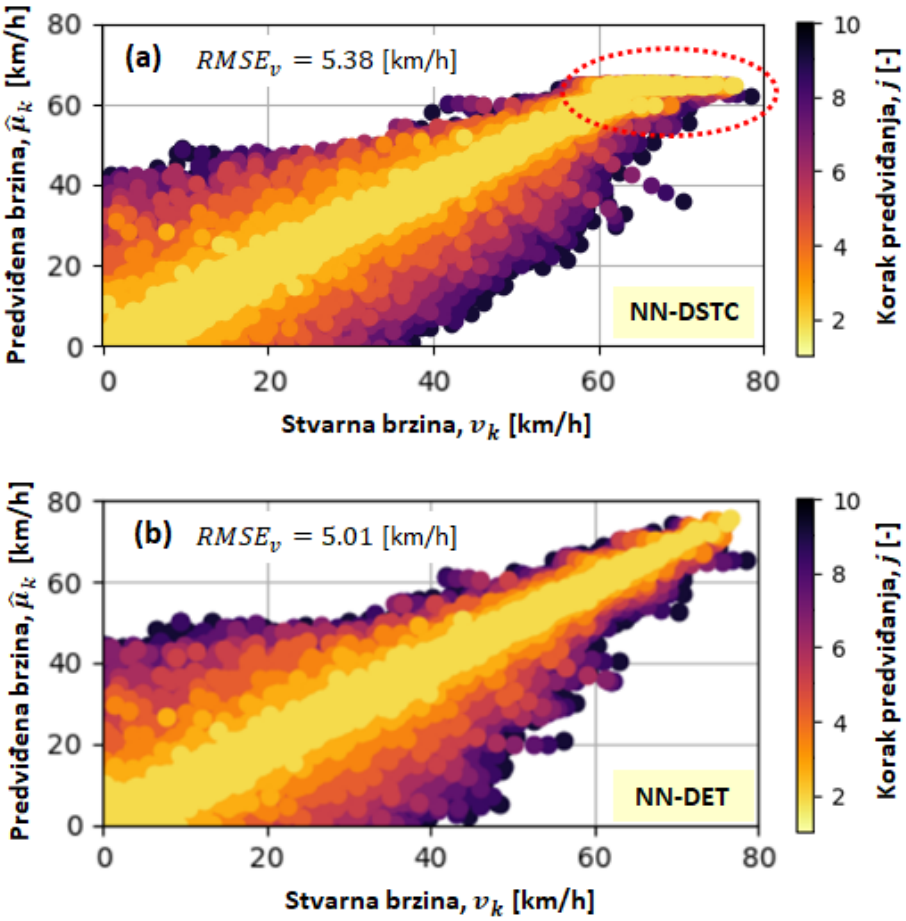
7.3.2. Analiza točnosti predviđanja prediktivnog modela

Sveobuhvatnu validaciju NN-DSTC modela nije moguće provesti jednoznačno u odnosu na snimljene podatke zbog stohastičke prirode samog modela. Razlog tome je što bi takva validacija modela zahtijevala poznavanje egzaktnih vrijednosti parametara originalnih razdioba brzine vozila za svaku kombinaciju ulaznih podataka, za čiji izračun je potrebno osigurati širu populaciju snimljenih uzoraka. Stoga se u nastavku daju preliminarni rezultati analize točnosti predviđanja NN-DSTC modela koji se temelje na usporednoj analizi:

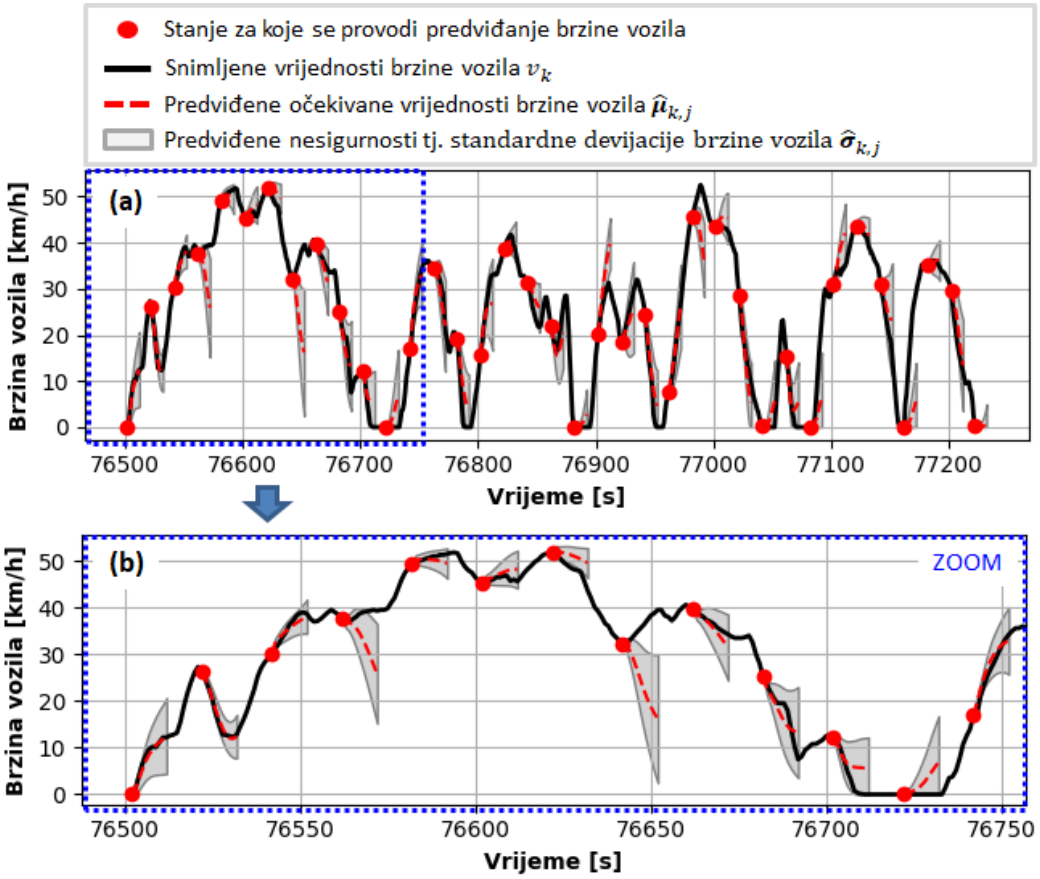
i. srednje kvadratne pogreške predviđenih očekivanih vrijednosti $\hat{\boldsymbol{\mu}}_k = [\hat{\mu}_{k,1}, \hat{\mu}_{k,2}, \dots, \hat{\mu}_{k,j}]$ u odnosu na snimljene vrijednosti brzine vozila v_k ($RMSE_v$ definiran korijenom izraza (7-5) i izračunat za testni skup podataka), te
ii. nesigurnosti predviđanja $\hat{\sigma}_{k,j}$ za svaki j-ti diskretni korak duž horizonta predviđanja u odnosu na prosječnu vrijednost pokazatelja $\widehat{\mathrm{RMSE}}_j$ i R^2 vrijednost dobivenu za NN-DET model (potpoglavlje 7.2).

Rezultirajuća vrijednost $RMSE_v$ za NN-DSTC, dobivena za čitav testni skup uzoraka, iznosi 5.38 km/h, što je samo za 7.4% veće u odnosu na NN-DET model iste konfiguracije ($RMSE_v =$ 5.01 km/h; tablica 7-2). Usporedna ilustracija kvalitete predviđanja (očekivanja) brzine prikazana je na slici 7-22. Razlozi nešto veće pogreške predviđanja NN-DSTC modela spram NN-DET modela su sljedeći: 1) znatno složeniji, stohastički izlaz kojim se predviđa znatno bogatija informacija o brzini vozila (usp. slike 7-20 i 7-12); te 2) nedostatni broj snimljenih brzina vozila većih od 60 km/h, što rezultira lošijim učenjem parametara $\hat{\mu}_{k,j}$ i $\hat{\sigma}_{k,j}$ u tom području brzina (vidi sliku 7-22). Radi zornije ilustracije složenosti zadatka predviđanja, na

slici 7-23 prikazana su predviđanja očekivanih vrijednosti brzine vozila $\hat{\mu}_{k,j}$ (crvene isprekidane linije) i pripadnih standardnih devijacija $\hat{\sigma}_{k,j}$ (područja označena sivom bojom) za primjer jednog izdvojenog testnog voznog ciklusa. Vidljivo je kako predviđene očekivane vrijednosti brzine $\hat{\mu}_{k,j}$ u slučajevima manje nesigurnosti prometnih uvjeta bolje prate snimljene vrijednosti v_k, uz uži interval pouzdanosti $\pm\hat{\sigma}_{k,j}$. Prema tome, rezultirajući porast pogreške predviđanja NN-DSTC modela može se smatrati zanemarivim s obzirom na prednosti koje ovaj stohastički model pruža u odnosu na NN-DET model.

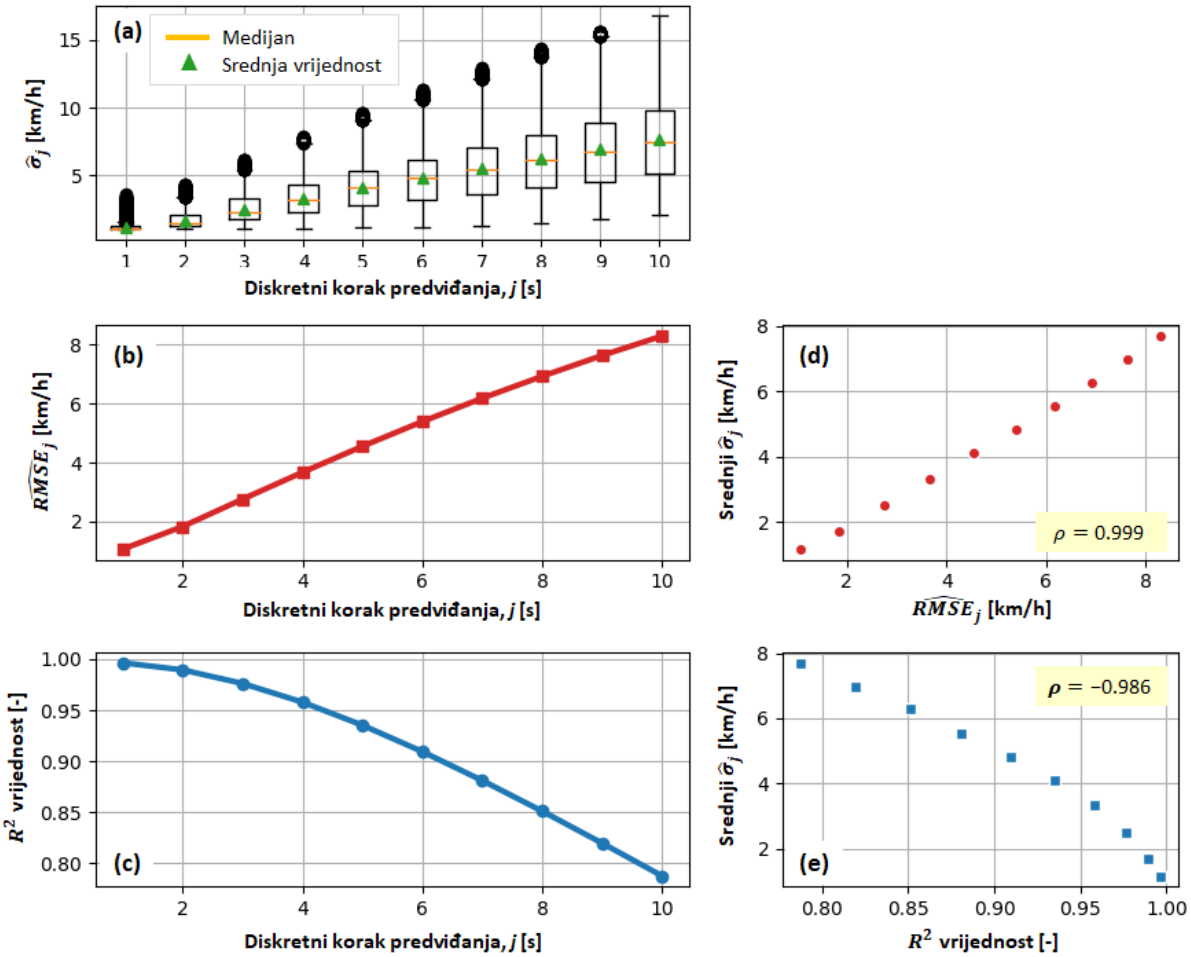


Slika 7-22: Predviđene očekivane vrijednosti $\hat{\mu}_k$ u odnosu na stvarne vrijednosti v_k brzine vozila za slučaj a) NN-DSTC i b) NN-DET modela. Predviđanja za svaki j-ti vremenski korak duž horizonta predviđanja označena su zasebnom bojom.



Slika 7-23: a) Predviđanja značajki brzine vozila ($\hat{\mu}_{k,j}$, $\hat{\sigma}_{k,j}$) za slučaj jednog testnog voznog ciklusa, uključujući b) uvećani segment profila brzine vozila.

Konačno, provedena je analiza točnosti predviđanja NN-DSTC modela o diskretnom vremenskom koraku predviđanja j duž horizonta predviđanja H_p. Pritom se, također, srednje vrijednosti predviđenih standardnih devijacija $\hat{\boldsymbol{\sigma}}_j$ dodatno uspoređuju s vrijednostima pokazatelja $\widehat{\mathrm{RMSE}}_j$ i R^2, kako bi se ustanovilo u kojoj mjeri se nesigurnosti predviđanja ($\hat{\boldsymbol{\sigma}}_j$) preklapaju s onima dobivenima analizom NN-DET modela. Rezultati sa slike 7-24a pokazuju kako nesigurnost predviđanja brzine vozila $\hat{\boldsymbol{\sigma}}_j$ očekivano raste s vremenskim odmakom u budućnosti. Isti trend prate i pokazatelji $\widehat{\mathrm{RMSE}}_j$ (slika 7-24b) te R^2 (slika 7-24c), čijom se korelacijskom analizom s obzirom na $\hat{\boldsymbol{\sigma}}_j$ potvrđuje dosljednost NN-DSTC modela (indeksi korelacije ρ iznose redom 0.999 i -0.986; slike 7-24d i 7-24e).



Slika 7-24: Ovisnost točnosti predviđanja NN-DSTC modela o diskretnom koraku duž horizonta predviđanja $j = 1,2,\ldots,H_p$, izražena kroz a) kutijasti dijagram standardne devijacije $\hat{\boldsymbol{\sigma}}_j$, b) prosječni $\widehat{RMSE}_j$, te c) R^2 vrijednost, uključujući međuovisnosti pokazatelja d) $\widehat{RMSE}$ i e) R^2 o $\hat{\boldsymbol{\sigma}}_j$; gdje ρ predstavlja odgovarajuće indekse korelacije.

8. Zaključak

Na početku rada opisan je postupak prikupljanja i predobrade podataka telemetrijskog praćenja flote gradskih autobusa. Snimljeni podaci vožnje najprije su segmentirani u vozne cikluse, gdje je svaki vozni ciklus definiran vremenskim profilima brzine vozila, prijeđene udaljenosti, geografskih koordinata i nadmorske visine između dvije krajnje stanice jedne karakteristične rute. Filtriranjem snimljenih podataka izdvojen je skup od 3216 validnih voznih ciklusa. Primjenom metoda temeljenih na algoritmu klasteriranja K-sredina i Gaussovim procesima na skup snimljenih podataka o poziciji vozila i prijeđenoj udaljenosti, rekonstruirani su precizni profili nagiba ceste te pridruženi svakom individualnom voznom ciklusu. Kako bi se uzeo u obzir stupanj zagušenja prometa, snimljeni vozni ciklusi kategorizirani su prema značajkama srednje brzine i broja zaustavljanja vozila po kilometru u tri klastera (lagana, srednja i teška zagušenost). Korelacijskom analizom snimljenih vrijednosti brzine vozila, ubrzanja vozila i nagiba ceste ustanovilo se da je profil nagiba ceste vrijedno sintetizirati zajedno s profilom brzine i ubrzanja vozila, kako bi se pravilno zabilježile njihove međuovisnosti ustanovljene u snimljenim voznim ciklusima. Također, pokazano je kako za vrlo ograničeni skup zabilježenih podataka o broju putnika u autobusu između mase putnika i brzine te akceleracije vozila postoji redom vrlo slaba ili nikakva korelacija.

Za potrebe sinteze voznih ciklusa uspostavljeno je sedam različitih modela voznih ciklusa zasnovanih na Markovljevim lancima, a koji su opisani matricama prijelaznih vjerojatnosti različitih dimenzionalnosti (TPM-2D do TPM-8D), ovisno o skupu varijabli korištenog za opisivanje stanja. Dodatno je predložen alternativni koncept realizacije modela voznih ciklusa koji razmatra metodu hijerarhijske sinteze voznih ciklusa, koja se temelji na činjenici da sljedeće stanje brzine i akceleracije vozila ovisi o nagibu ceste, dok suprotno ne vrijedi. Kako bi se poboljšala računalna učinkovitost i smanjili memorijski zahtjevi, predložena je implementacija TPM-a u obliku rijetke matrice temeljene na rječniku ključeva, te je sustavno uspoređena s tradicionalnom pristupom temeljenim na N-dimenzionalnom nizu u smislu zahtjeva za memorijom i računalne učinkovitosti. Rezultati analiza pokazali su kako u slučaju implementacije TPM-a u obliku rijetke matrice ne postoje restriktivna ograničenja za dimenzionalnost TPM-a, budući da je zahtjev za memorijom drastično smanjen (na ispod 60 MB), u usporedbi s konvencionalnim pristupom temeljenom na N-dimenzionalnom nizu gdje određeni slučajevi TPM-a nisu izvedivi čak i za Markovljeva stanja grube razlučivosti (zahtjev za memorijom je veći, nekad i znatno veći od 16 GB). Također, pokazano je kako TPM-ovi

implementirani u obliku rijetke matrice imaju superiornu računalnu učinkovitost u smislu da je vrijeme izvođenja praktički neovisno o dimenziji TPM-a, dok se kod konvencionalnog pristupa ono značajno povećava s dimenzijom TPM-a. Nadalje, ustanovljeno je kako modeli koji podrazumijevaju neizravno izračunavanje sljedećeg stanja brzine vozila ili hijerarhijsku sintezu nisu konkurentni izravnim postupcima, jer rezultiraju neuspjelim pokušajima generiranja sintetičkih ciklusa tj. usporenom sintezom. Stoga je kao ultimativni model za sintezu voznih ciklusa usvojen model izveden u obliku TPM-8D, koji koristi kombinaciju diskretnih vrijednosti brzine vozila, ubrzanja vozila, nagiba ceste i vremenske derivacije nagiba ceste kao stanja Markovljevog lanca. Za svaku kategoriju stupnja zagušenja prometa razvijena je zasebna 8D-TPM kojom je generirano po 1000 sintetičkih voznih ciklusa, te je demonstrirano kako je temeljem srednje brzine i broja zaustavljanja vozila po kilometru moguće izdvojiti reprezentativne sintetičke vozne cikluse za svaki stupanj zagušenosti prometa na nezamršen način. Za potrebe integracije promjenjive mase putnika u sintetičke vozne cikluse predložena je metoda sinteze profila mase putnika temeljena na događajima zaustavljanja autobusa. Usporednom analizom sintetiziranih profila mase putnika u odnosu na zabilježene profile pokazana je solidna međusobna podudarnost (odstupanja očekivane vrijednosti mase putnika manje od 10%), čime je potvrđena valjanost predložene metode.

U svrhu predviđanja potrošnje goriva na temelju poznatog (sintetičkog) voznog ciklusa, predloženi su i ispitani različiti pristupi temeljeni na polinomskim regresijskim modelima (linearnim u parametrima) te neuronskim mrežama (NN). Za potrebe dobro uvjetovanog učenja modela generiran je bogat skup kombiniranih voznih ciklusa nasumičnim spajanjem različitog broja snimljenih mikrociklusa, gdje je svaki mikrociklus određen potrošnjom goriva jednakoj razlučivosti mjerenja od 0.5 L. U slučaju NN modela, vozni ciklusi različite duljine transformirani su u histogram fiksne dimenzije, gdje svaki element histograma označava broj uočenih stanja opisanih kombinacijom diskretnih vrijednosti brzine i ubrzanja vozila (H2D), te dodatno nagiba ceste (H3D). Pokazalo se kako dodavanje informacije o nagibu ceste kao dodatnog ulaza u model značajno povećava točnost predviđanja, tj. standardna devijacija reziduala potrošnje goriva smanjuje se s 0.8 L na 0.5 L u slučaju polinomskih regresijskih modela, te s 0.5 L na 0.2 L u slučaju NN modela. Ultimativna izvedba NN modela temeljenog na 3D histogramu (NN-H3D) očituje se u sposobnosti hvatanja varijance u snimljenim podacima u mjeri koja je približno 95% veća u usporedbi s osnovnim linearnim regresijskim modelom ($Poly_{1D}$), koji se temelji samo na prijeđenoj udaljenosti. Dodatno su predložene metode za fino-podešavanje NN modela. Analizom glavnih komponenti pokazano je kako je

moguće reducirati dimenzionalnost ulaza NN-H3D s 13.325 značajki na 1976 glavnih komponenti (relativno smanjenje od -85%), uz zanemarivo pogoršanje točnosti predviđanja modela (standardna devijacija reziduala potrošnje goriva od 0.23 L naspram 0.19 L). Ispitivanje točnosti predviđanja za različite arhitekture NN-H3D modela u smislu broja skrivenih slojeva i neurona pokazalo je kako je optimalni broj skrivenih slojeva jednak 4, gdje svaki skriveni sloj (počevši od prvog) posjeduje redom 16, 8, 4 i 2 neurona. Provođenjem ispitivanja raspona stope učenja (engl. skrać. LRTT) pokazalo se kako se optimalna vrijednost stope učenja η nalazi u rasponu između 10^{-3} i 10^{-2}. Za potrebe bolje generalizacije NN-H3D modela provedena je L_2 regularizacija, te je pokazano kako se optimalna vrijednost faktora regularizacije λ kreće u rasponu od $1.96 \cdot 10^{-2}$ do $2.15 \cdot 10^{-2}$.

Generirani sintetički vozni ciklusi prvo su validirani postupkom višekriterijske validacije temeljene na bogatom skupu nominiranih statističkih značajki (ukupno 100), koje uz opće statističke indekse iz vremenske domene uključuju i značajke frekvencijske domene i unakrsne korelacije brzine i ubrzanja vozila te nagiba ceste. Za potrebe jednoznačne validacije izvedeni su različiti združeni pokazatelji reprezentativnosti voznog ciklusa temeljeni na statističkim značajkama (ED, MAE, STD) i sličnosti dvaju histograma (SI-HIST) ili sličnosti matrica prijelaznih vjerojatnosti (SI-TPM). Korelacijskom analizom u odnosu na odstupanje potrošnje goriva (FCD), dobiveno iz predviđanja primjenom NN-H3D modela, utvrđeno je kako su najutjecajniji pokazatelji za validaciju voznih ciklusa Euklidska udaljenost (ED) i srednja apsolutna pogreška odstupanja statističkih značajki (MAE). Primjenom predložene metode validacije temeljene na višestrukim kriterijima (FCD, MAE i ED) izolirano je ukupno 15 reprezentativnih sintetičkih voznih ciklusa, od kojih su nadalje izolirana tri najreprezentativnija u Pareto optimalnom smislu. Rezultati usporedne analize pokazali su kako su očekivanja osam karakterističnih statističkih značajki za kandidirane (najreprezentativnije) sintetičke vozne cikluse bliska srednjim vrijednostima kombiniranog snimljenog voznog ciklusa, tj. da se velika većina relativnih odstupanja statističkih značajki nalazi unutar ±10% tolerancije. K tome, sintetički profili nagiba ceste u velikoj mjeri nalikuju na snimljene profile kako u pogledu odstupanja od prosječnih vrijednosti i varijance tako i u frekvencijskom smislu, što je potkrijepljeno niskim relativnim standardnim odstupanjima nagiba ceste (manja od 3%).

Kako bi se pojednostavio cjelokupni postupak validacije višedimenzionalnih sintetičkih voznih ciklusa, relevantnost svake nominirane statističke značajke nadalje je ispitan u odnosu na kvalitetu predviđanja potrošnje goriva. Pritom je izdvojen podskup najsignifikantnijih statističkih značajki temeljem linearne regresijske analize, odnosno, operatera najmanjeg

apsolutnog skupljanja i odabira (engl. skrać. LASSO). Temeljem LASSO predviđene potrošnje goriva definiran je novopredloženi pokazatelj reprezentativnosti voznog ciklusa, tzv. regresijski indeks (RI), kojim se nadomješta složeniji FCD indeks. Rezultati LASSO regresijske analize pokazali su kako se visoka točnost predviđanja (konkurentna složenijoj neuronskoj mreži; NN-H3D) može postići korištenjem reduciranog skupa od samo tri (od ukupno 100) statističke značajke voznih ciklusa: srednja specifična energija vožnje, srednja brzina vozila te srednje pozitivno ubrzanje vozila. Korelacijska analiza ukazala je na značajno višu korelaciju predloženog LASSO regresijskog indeksa (RI) spram FCD u odnosu na ED spram FCD (koeficijent korelacije je 88% spram 38%). Nadalje, pokazano je kako postupak jednoznačne validacije voznih ciklusa temeljem isključivo pokazatelja RI ili ED ne garantira nužno i minimalno odstupanje potrošnje goriva FCD, odnosno, kako je reprezentativni sintetički vozni ciklus potrebno izdvojiti prema kriteriju minimalne udaljenosti RI i ED od ishodišta (RI, ED) koordinatnog sustava. Temeljem predloženog dualnog kriterija osiguravaju se minimalna odstupanja prema svim karakterističnim statističkim značajkama, uključujući potrošnju goriva.

Nadalje, razvijeni su stohastički i deterministički modeli za predviđanje razdiobe brzine vozila duž rute te profila brzine vozila na pomičnom vremenskom horizontu, koji su temeljeni na dubokoj unaprijednoj višeslojnoj neuronskoj mreži. Za potrebe pripreme podataka za učenje, validaciju i testiranje prediktivnih modela predložena je metoda izdvajanja reduciranog skupa slučajnih uzoraka kojom je inicijalni broj od 1,908.209 snimljenih uzorka sveden na 382.553 izdvojena uzorka. Rezultati testiranja statičkog stohastičkog modela (NN-STC) pokazali su kako model nije u mogućnosti kvalitetno predvidjeti individualne vrijednosti brzine vozila za svaki vozni ciklus (R^2 od 0.58), budući da mapira čitave razdiobe uvjetnih vjerojatnosti duž rute temeljem pozicije vozila kao jedine ulazne značajke. Međutim, analizom točnosti predviđanja NN-STC u smislu odabranih statističkih indeksa vezanih za razdiobe brzine vozila (predviđene očekivane vrijednosti $\hat{\mu}_k$ i standardna odstupanja $\hat{\sigma}_k$) u odnosu na one dobivene za skup snimljenih testnih uzoraka μ_k i σ_k ustanovljene su znatno bolje performanse (R^2 od 0.95 za očekivane vrijednosti ($\hat{\mu}_k$, μ_k), te 0.51 za standardne devijacije ($\hat{\sigma}_k$, σ_k), uz prosječna odstupanja srednjih vrijednosti od ± 1 km/h, s rasipanjem od oko $2-3$ km/h u 75% slučajeva).

Analizom utjecaja svake kandidirane ulazne varijable tj. značajke na točnost predviđanja determinističkog modela (NN-DET), ustanovljeno je kako su trenutna pozicija vozila i povijest brzina najznačajnije značajke, dok informacije o dobu dana i danu u tjednu imaju gotovo zanemariv utjecaj. Pored toga, rezultati ispitivanja točnosti predviđanja NN-DET modela za različite vrijednosti duljine vektora povijesti brzina H_h ukazuju kako je optimalna duljina H_h

jednaka 50 s, dok vrijednosti H_h u rasponu od 30 do 50 s predstavljaju dobru praktičnu alternativu, koja u određenoj mjeri reducira složenost modela i daje bolja generalizacijska svojstva. Usporednom analizom točnosti predviđanja NN-DET modela, koji pored povijesti brzina uključuje trenutnu poziciju vozila, pokazano je kako se solidna točnost predviđanja ($R^2 > 0.25$) zadržava čak i za pomične horizonte H_p dulje od 100 s u budućnosti. Time se potvrđuje značaj uključivanja trenutne pozicije vozila kao ulazne varijable u model, posebno za primjene gdje su vozni ciklusi vezani uz konstantne rute, poput razmatranog slučaja gradskog autobusnog prijevoza te određenih kategorija dostavnog prijevoza.

Konačno, predložen je ultimativni dinamički stohastički model (NN-DSTC), koji objedinjuje karakteristike NN-STC i NN-DET modela u smislu predviđanja očekivanja i standardne devijacije brzine u svakom koraku pomičnog horizonta. Usporedna analiza kvalitete predviđanja očekivanja brzine vozila pokazala je kako NN-DSTC model posjeduje nešto lošiju točnost u odnosu na NN-DET model iste konfiguracije (srednja kvadratna pogreška $RMSE_v$ veća samo za 7.4%), uz značajnu prednost uključivanja predviđanja nesigurnosti pojedinih uzoraka duž horizonta. Također, pokazano je kako očekivanja brzine predviđena primjenom NN-DSTC modela bolje prate snimljene vrijednosti u slučajevima manje nesigurnosti prometnih uvjeta, uz manja rasipanja standardnih devijacija brzine, te kako nesigurnost predviđanja NN-DSTC modela očekivano raste s vremenskim odmakom duž horizonta predviđanja i blisko prati vrijednosti srednje kvadratne pogreške dobivene za NN-DET model (indeks korelacije veći od 99%), čime se potvrđuje dosljednost NN-DSTC modela.

U ovom radu pokazano je kako je primjenom predložene metode sinteze i validacije višedimenzionalnih voznih ciklusa moguće izdvojiti reprezentativne vozne cikluse na nezamršen i računalno učinkovit način. Reprezentativnost voznih ciklusa odražena je kroz razmatranje bogatog skupa statističkih značajki izvedenih iz snimljenih obrazaca vožnje. Ovako generirani sintetički vozni ciklusi mogu se koristiti u različitim simulacijskim studijama (uključujući eksperimentalna, tipično dinamometarska ispitivanja), gdje je potrebno zamijeniti širok skup snimljenih voznih ciklusa s jednim ili nekoliko reprezentativnih voznih ciklusa (npr. za potrebe certifikacije, dizajna vozila i sl.) ili gdje postoji potreba za velikim brojem voznih ciklusa (npr. analize osjetljivosti potrošnje energije vozila i sl.). Povrh toga, predstavljeni LASSO linearni regresijski model može se koristiti za predviđanje potrošnje goriva te na sličan način i emisija štetnih ispušnih te stakleničkih plinova na temelju nekoliko najsignifikantnijih statističkih značajki, koje je moguće izračunati na temelju standardnih podataka o praćenju vozila niske razlučivosti (tipično odašiljanih svakih 20-40 s) ili čak i podataka transportnog

planiranja. Time se omogućuje predviđanje potrošnje energije na temelju lako dostupnih podataka na vrlo računski učinkovit način (tj. bez korištenja mikrosimulacija temeljenih na voznim ciklusima visoke rezolucije), kako bi se podržale različite studije planiranja i optimizacije transportnog sustava (npr. planiranje elektrifikacije gradskog autobusnog prijevoza). Konačno, predstavljeni modeli za predviđanje brzine vozila na pomičnom horizontu u budućnosti mogu se primijeniti za sintezu naprednih sustava upravljanja u svrhu minimiziranja potrošnje goriva i električne energije utičnog hibridnog električnog vozila (PHEV) u širokom rasponu voznih ciklusa, npr. kroz tradicionalni pristup modelskog prediktivnog upravljanja (MPC) pogonom vozila ukoliko se razmatra deterministički model predviđanja (NN-DET), ili stohastički MPC (SMPC) kada se koristi dinamički stohastički model predviđanja (NN-DSTC). Predmet budućeg rada u ovom području mogao bi se odnositi na primjenu ostalih metoda odabira značajki, poput metoda temeljenih na filtru, omotaču i ostalih ugrađenih metoda, te primjenu reprezentativnih voznih ciklusa zajedno s modelima za predviđanje brzine vozila za dizajn i optimizaciju naprednih prediktivnih i adaptivnih strategija za upravljanje pogonom PHEV.

Na temelju izloženog, mogu se izdvojiti sljedeći znanstveni doprinosi doktorskog rada:

1. Model voznih ciklusa temeljen na Markovljevim lancima, koji uz brzinu i akceleraciju kao stanje Markovljevog lanca uključuje i nagib ceste, masu vozila, te značajke zagušenja prometa.
2. Validacija modela voznih ciklusa korištenjem snimljenih podataka o potrošnji goriva i predviđanja potrošnje goriva uz primjenu neuronske mreže parametrirane temeljem značajki modela voznih ciklusa.
3. Stohastički model za predviđanje statističkih značajki voznih ciklusa prilagođen za potrebe stohastičkog modelskog prediktivnog upravljanja PHEV.

Literatura

1. European Commission. Available from: https://ec.europa.eu/clima/policies/strategies/2050_en [cited 30.7.2021.];
2. Carlson R, Lohse-Busch H, Duoba M, Shidore N. Drive Cycle Fuel Consumption Variability of Plug-In Hybrid Electric Vehicles Due to Aggressive Driving. SAE Technical Paper, no. 2009-01-1335, 2009. doi:10.4271/2009-01-1335.
3. Fontaras G, Franco V, Dilara P, Martini G, Manfredi U. Development and Review of Euro 5 Passenger Car Emission Factors Based on Experimental Results Over Various Driving Cycles. Science of The Total Environment. 2014; 468-469:1034–1042. doi:10.1016/j.scitotenv.2013.09.043.
4. Giakoumis EG. Driving and Engine Cycles. Cham: Springer; 2017.
5. Barlow TJ, Latham S, McCrae IS, Boulter PG. A Reference Book of Driving Cycles for Use in the Measurement of Road Vehicle Emissions. Workingham: Transport Research Laboratory; 2009.
6. Huertas JI, Giraldo M, Quirama LF, Díaz J. Driving Cycles Based on Fuel Consumption. Energies. 2018;11(11):3064. doi:10.3390/en11113064.
7. Andrade GMSd, Araújo FWCd, Santos MPMdN, Magnani FS. Standardized Comparison of 40 Local Driving Cycles: Energy and Kinematics. Energies. 2020;13(20):5434. doi:10.3390/en13205434.
8. Rajan BVP, McGordon A, Jennings PA. An Investigation on the Effect of Driver Style and Driving Events on Energy Demand of a PHEV. World Electr. Veh. J. 2012; 5(1):173-181. doi:10.3390/wevj5010173.
9. Davari MM, Jerrelind J, Stensson TA. Energy efficiency analyses of a vehicle in modal and transient driving cycles including longitudinal and vertical dynamics. Transportation Research Part D: Transport and Environment. 2017;53:263–275. doi:10.1016/j.trd.2017.04.019.
10. Chindamo D, Gadola M. What is the Most Representative Standard Driving Cycle to Estimate Diesel Emissions of a Light Commercial Vehicle?. IFAC-PapersOnLine. 2018; 51(5):73–78. doi:10.1016/j.ifacol.2018.06.213.
11. Borlaug B, Holden J, Wood E, Lee B, Fink J, Agnew S, Lustbader J. Estimating region-specific fuel economy in the United States from real-world driving cycles. Transportation Research Part D: Transport and Environment. 2020;86(3):102448. doi:10.1016/j.trd.2020.102448.
12. Lee H, Lee K. Comparative Evaluation of the Effect of Vehicle Parameters on Fuel Consumption under NEDC and WLTP. Energies. 2020;13(16):4245. doi: 10.3390/en13164245.
13. Cubito C, Millo F, Boccardo G, Di Pierro G, Ciuffo B, Fontaras G, Serra S, Otura Garcia M, Trentadue G. Impact of Different Driving Cycles and Operating Conditions on CO2 Emissions and Energy Management Strategies of a Euro-6 Hybrid Electric Vehicle. Energies. 2017; 10(10):1590. doi:10.3390/en10101590.

14. Tietge U, Diaz U, Mock P, German J, Bandivadekar A, Ligterink N. From Laboratory to Road - A 2016 Update of Official and 'Real-World' Fuel Consumption and CO2 Values for Passenger Cars in Europe, 2016. Available from: https://theicct.org/sites/default/files/publications/ICCT_LaboratoryToRoad_2016.pdf [cited 30.7.2021.];

15. Fontaras G, Zacharof NG, Ciuffoo B. Fuel consumption and CO2 emissions from passenger cars in Europe — Laboratory versus real-world emissions. Progress in Energy and Combustion Science. 2017;60:97–131. doi:10.1016/j.pecs.2016.12.004.

16. Peter M, Kuhlwein J, Tietge U, Franco V, Bandivadekar A, German J. The WLTP: How a new test procedure for cars will affect fuel consumption values in the EU. International Council on Clean Transportation (ICTT). 2014;9:35–47. Available from: https://theicct.org/publications/wltp-how-new-test-procedure-cars-will-affect-fuel-consumption-values-eu [cited 30.7.2021.];

17. Eisele A, Horlbeck L, Hann M, Jager B, Lienkamp M. Real-life Data Based Analysis of Driving Cycles Regarding their Insufficient Cycle Duration. 19th International Conference on Intelligent Transportation Systems (ITSC); Rio de Janeiro, Brazil, 1.-4.11.2013., p. 408-415. doi:10.1109/ITSC.2016.7795587.

18. Tutuianu M, Bonnel P, Ciuffo B, Haniu T, Ichikawa N, Marotta A, Pavlovic J, Steven H. Development of the World-wide harmonized Light duty Test Cycle (WLTC) and a Possible Pathway for its Introduction in the European legislation. Transportation Research Part D: Transport and Environment. 2015;40:61-75. doi:10.1016/j.trd.2015.07.011.

19. Zardini A, Bonnel P. Real Driving Emissions Regulation. EUR 30123 EN. Publications Office of the European Union, Luxembourg, 2020. ISBN 978-92-76-17157-7. doi:10.2760/176284. JRC119889.

20. Lee TK, Filipi ZS. Synthesis of Real-World Driving Cycles Using Stochastic Process and Statistical Methodology. International Journal of Vehicle Design. 2011;57(1):17-36. doi:10.1504/IJVD.2011.043590.

21. Škugor B, Deur J. Delivery Vehicle Fleet Data Collection, Analysis and Naturalistic Driving Cycles Synthesis. International Journal of Innovation and Sustainable Development. 2016;10(1):19-39. doi:10.1504/ IJISD.2016.073412.

22. Zeyu C, Qing Z, Jiahuan L, Jiangman B. Optimization-based method to develop practical driving cycle for application in electric vehicle power management: A case study in Shenyang, China. Energy. 2019;186:115766. doi:10.1016/j.energy.2019.07.096.

23. Ho S, Wong Y, Chang VW. Developing Singapore Driving Cycle for passenger cars to estimate fuel consumption and vehicular emissions. Atmos. Environ. 2014;97:353-362. doi:10.1016/j.atmosenv.2014.08.042.

24. Liessner R, Dietermann AM, Bäker B, Lüpkes K. Derivation of real-world driving cycles corresponding to traffic situation and driving style on the basis of Markov models and cluster analyses. 6th Hybrid and Electric Vehicles Conference; London, UK, 2.-3.11.2016., p. 1-7. doi:10.1049/cp.2016.0961.

25. Esser A, Zeller M, Foulard S, Rinderknecht S. Stochastic Synthesis of Representative and Multidimensional Driving Cycles. SAE Int. J. Alt. Power. 2018;7(3):263–272. doi:10.4271/2018-01-0095.
26. Peng J, Jiang J, Ding F, Tan H. Development of Driving Cycle Construction for Hybrid Electric Bus: A Case Study in Zhengzhou, China. Sustainability. 2020;12(7):7188. doi:10.3390/su12177188.
27. Lee TK, Filipi ZS. Synthesis and validation of representative real-world driving cycles for plug-in hybrid vehicles. IEEE Vehicle Power and Propulsion Conference; Lille, France, 1-3.9.2010., p. 1-6. doi:10.1109/VPPC.2010.5729040.
28. Liu Z, Ivanco A, Filipi ZS. Naturalistic driving cycle synthesis by Markov chain of different orders. International Journal of Powertrains 2018;6(4):307–322. doi:10.1504/IJPT.2017.10009484.
29. Conway G, Joshi A, Leach F, Garcia A, Senecal PK. A review of current and future powertrain technologies and trends in 2020. Transportation Engineering. 2021;5(1):100080. doi:10.1016/j.treng.2021.100080.
30. Kumar RR, Alok K. Adoption of electric vehicle: A literature review and prospects for sustainability. J. Clean. Prod. 2020; 253:119911. doi:10.1016/j.jclepro.2019.119911.
31. Das HS, Rahman MM, Li S, Tan CW. Electric vehicles standards, charging infrastructure, and impact on grid integration: A technological review. Renew. Sustain. Energy Rev. 2020;120:109618. doi:10.1016/j.rser.2019.109618.
32. Safoutin MJ. Predicting the Future Manufacturing Cost of Batteries for Plug-In Vehicles for the U.S. Environmental Protection Agency (EPA) 2017–2025 Light-Duty Greenhouse Gas Standards. World Electr. Veh. J. 2018;9(3):42. doi:10.3390/wevj9030042.
33. Sabri MFM, Danapalasingam KA, Rahmat MF. A review on hybrid electric vehicles architecture and energy management strategies. Renew. Sustain. Energy Rev. 2016;53:1433–1442. doi:10.1016/j.rser.2015.09.036.
34. Rezaei A, Burl JB. Effects of Time Horizon on Model Predictive Control for Hybrid Electric Vehicles. IFAC-PapersOnLine. 2015;48(15):252–256. doi:10.1016/j.ifacol.2015.10.036.
35. Martinez CM, Hu X, Cao D, Velenis E, Gao B, Wellers M. Energy Management in Plug-in Hybrid Electric Vehicles: Recent Progress and a Connected Vehicles Perspective. IEEE Transactions on Vehicular Technology. 2017;66(6):4534-4549. doi:10.1109/TVT.2016.2582721.
36. Van Duin JHR, Tavasszy LA, Quak HJ. Towards E(lectric) - Urban Freight: First Promising Steps in the Electric Vehicle Revolution. European Transport \ Trasporti Europei. 2013; 54:9. Available from: https://www.researchgate.net/publication/268509461_Towards_Electric-urban_freight_First_promising_steps_in_the_electric_vehicle_revolution [cited 30.7.2021.];

37. Yang Z, Alexandre R, Marie-Cécile P. A survey on driving prediction techniques for predictive energy management of plug-in hybrid electric vehicles. Journal of Power Sources. 2019;412:480 - 495. doi:10.1016/j.jpowsour.2018.11.085.

38. Hyeon E, Kim Y, Prakash N, Stefanopoulou AG. Influence of Speed Forecasting on the Performance of Ecological Adaptive Cruise Control, ASME 2019 Dynamic Systems and Control Conference; Park City, Utah, 8-11.102019., p. 1-8. doi:10.1115/DSCC2019-9046.

39. Shankar R, Marco J, Assadian F. The novel application of optimization and charge blended energy management control for component downsizing within a plug-in hybrid electric vehicle. Energies. 2012;5:4892–4923. doi:10.3390/en5124892.

40. Geller BM, Bradley TH. Analyzing Drive Cycles for Hybrid Electric Vehicle Simulation and Optimization. Journal of Mechanical Design. 2015;137(4):041401. doi:10.1115/1.4029583.

41. Brady J, O'Mahony M. Development of a driving cycle to evaluate the energy economy of electric vehicles in ubran areas. Applied Energy. 2016;177:165-178. doi:10.1016/j.apenergy.2016.05.094.

42. Patil C, Naghshtabrizi P, Verma R, Tang Z, et al. Optimal battery utilization over lifetime for parallel hybrid electric vehicle to maximize fuel economy. Proceedings of the American Control Conference; Boston, MA, USA, 6-8.7.2016., p. 1524-1529. doi:10.1109/ACC.2016.7525132.

43. Naranjo W, Camargo LEM, Pereda JE, Cortes C. Design of Electric Buses of Rapid Transit Using Hybrid Energy Storage and Local Traffic Parameters. IEEE Transactions on Vehicular Technology. 2017;66(7):5551-5563. doi:10.1109/TVT.2016.2616401.

44. Zhang F, Guo F, Huang H. A Study of Driving Cycle for Electric Special-purpose Vehicle in Beijing. Energy Procedia. 2017;105:4884-4889. doi:10.1016/j.egypro.2017.03.967.

45. Musardo C, Rizzoni G, Guezennec Y, Staccia B. A-ECMS: An Adaptive Algorithm for Hybrid Electric Vehicle Energy Management. European Journal of Control. 2005;11(4-5):509–524. doi:10.3166/ejc.11.509-524.

46. Onori S, Tribioli L. Adaptive Pontryagin's Minimum Principle supervisory controller design for the plug-in hybrid GM Chevrolet Volt. Applied Energy. 2015;147:224-234. doi: 10.1016/j.apenergy.2015.01.021.

47. Silvas E, Hereijgers K, Peng H, Hofman T, Steinbuch M. Synthesis of Realistic Driving Cycles with High Accuracy and Computational Speed, Including Slope Information. IEEE Transactions on Vehicular Technology. 2016;65(6):4118-4128. doi:10.1109/TVT.2016.2546338.

48. Zhao T, Hofman C, Lv M, Steinbuch M. Intelligent Synthesis of Driving Cycle for Advanced Design and Control of Powertrains. 2018 IEEE Intelligent Vehicles Symposium (IV); Changshu, Suzhou, China, 26.-30.6.2018., p. 1608-1613. doi:10.1109/IVS.2018.8500522.

49. Xiaodong L, Jian M, Xuan Z, Juan D, Yanfeng X. Study on Driving Cycle Synthesis Method for City Buses considering Random Passenger Load. Journal of Advanced Transportation. 2020;2020:3871703. doi:10.1155/2020/3871703.

50. Hereijgers K, Silvas E, Hofman T, Steinbuch M. Effects of using Synthesized Driving Cycles on Vehicle Fuel Consumption. IFAC-PapersOnLine. 2017;50(1):7505-7510. doi:10.1016/j.ifacol.2017.08.1183.

51. Esser A, Rinderknecht S. Process for the Validation of Using Synthetic Driving Cycles Based on Naturalistic Driving Data Sets. 23rd International Conference on Intelligent Transportation Systems (ITSC); Rhodes, Greece, 20.-23.9.2020., p. 1-6. doi:10.1109/ITSC45102.2020.9294369.

52. Topić J, Škugor B, Deur J. Synthesis and Validation of Multidimensional Driving Cycles. SAE Int. J. Adv. & Curr. Prac. in Mobility. 2021;3(4):1558-1568. doi:10.4271/2021-01-0125.

53. Guyon I, Elisseeff A. An Introduction to Variable and Feature Selection. Journal of Machine Learning Reasearch. 2003;3:1157-1182. doi:10.1162/153244303322753616.

54. Hastie T, Tibshirani R, Friedman JH. The elements of Statistical Learning: Data Mining, Inference, and Prediction. 2nd ed. New York: Springer: New York; 2009.

55. Fonti V, Belitser E. Feature Selection using LASSO. Res. Pap. Bus. Anal. VU Amst. 2017; 30:1-25. Available from: https://vdocument.in/feature-selection-using-lasso-vu-feature-selection-using-lasso-author-valeria.html [cited 30.7.2021.];

56. Chandrashekar G, Sahin F. A survey on feature selection methods. Comput. Electr. Eng. 2014;40(1):16–28. doi:10.1016/j.compeleceng.2013.11.024.

57. Tibshirani R. Regression Shrinkage and Selection via the Lasso. J. R. Stat. Soc. Ser. B. 1996;58(1):267–288. Available from: http://www.jstor.org/stable/2346178 [cited 30.7.2021.];

58. Škugor B, Deur J. Synthetic Driving Cycles-based Modelling of Extended Range Electric Vehicle Fleet Energy Demand. Electric Vehicle Symposium & Exhibition (EVS30); Stuttgart, Germany, 9.-11.10.2017.

59. Topić J, Škugor B, Deur J. Analysis of Markov Chain-based Methods for Synthesis of Driving Cycles of Different Dimensionality. 23rd IEEE International Conference on Intelligent Transportation Systems (ITTS), Rhodes, Greece, 20.-23.9.2020., p. 893-900. doi:10.1109/ITSC45102.2020.9294302.

60. Topić J, Škugor B, Deur J. Analysis of City Bus Driving Cycle Features for the Purpose of Multidimensional Driving Cycle Synthesis. SAE Technical Paper, no. 2020-01-1288, 2020. doi:10.4271/2020-01-1288.

61. Arata J, Leamy M, Cunefare K. Power-split HEV control strategy development with refined engine transients. SAE International Journal of Alternative Powertrains. 2012;1(1):119-133. doi:10.4271/2012-01-0629.

62. Jauch C, Tamilarasan S, Bovee K, Güvenc L, Rizzoni G. Modeling for drivability and drivability improving control of HEV. Control Engineering Practice. 2018;70:50-62. doi:10.1016/j.conengprac.2017.09.014.

63. Guzzella L, Sciarretta A. Vehicle Propulsion Systems: Introduction to modeling and optimization. 2nd ed. Berlin: Springer; 2007.

64. Rajamani R. Longitudinal Vehicle Dynamics. In: Vehicle Dynamics and Control. Mechanical Engineering Series. Boston: Springer; 2006., p. 87-111.

65. Karbowski D, Kim N, Rousseau A. Route-based online energy management of a PHEV and sensitivity to trip prediction. IEEE Vehicle Power Propuls. Conf. (VPPC); Coimbra, Portugal, 27.-30.10.2014., p. 1–6. doi:10.1109/VPPC.2014.7007126.

66. Huang Y, Wang H, Khajepour A, He H, Ji J. Model predictive control power management strategies for HEVs: A review. J. Power Sources. 2017;341:91-106. doi:10.1016/j.jpowsour.2016.11.106.

67. Lemieux J, Ma Y. Vehicle Speed Prediction Using Deep Learning. IEEE Vehicle Power and Propulsion Conference (VPPC); Montreal, QC, Canada, 19.-22.10.2015., p. 1-5. doi: 10.1109/VPPC.2015.7353037.

68. Zhou Y, Ravey A, Marion-Pera MC. A survey on driving prediction techniques for predictive energy management of plug-in hybrid electric vehicles. Journal of Power Sources. 2019;412(C):480-495. doi:10.1016/j.jpowsour.2018.11.085.

69. Lee TK, Filipi Z. Real-world driving pattern recognition for adaptive HEV supervisory control: Based on representative driving cycles in Midwestern US. SAE Technical Paper, no. 2012-01-1020, 2012. doi:10.4271/2012-01-1020.

70. Zhang S, Xiong R. Adaptive energy management of a plug-in hybrid electric vehicle based on driving pattern recognition and dynamic programming. Applied Energy. 2015;155:68-78. doi:10.1016/j.apenergy.2015.06.003.

71. Lei Z, Cheng D, Liu Y, Qin D, Zhang Y, Xie Q. A Dynamic Control Strategy for Hybrid Electric Vehicles Based on Parameter Optimization for Multiple Driving Cycles and Driving Pattern Recognition. Energies. 2017;10(1):54. doi:10.3390/en10010054.

72. Wei Z, Xu Z, Halim D. Study of HEV power management control strategy Based on driving pattern recognition. Energy Procedia. 2016;88:847-853. doi:10.1016/j.egypro.2016.06.062.

73. Yu H, Tseng F, McGee R. Driving pattern identification for EV range estimation. IEEE International Electric Vehicle Conference; Greenville, SC, USA, 4.-8.3.2012. doi:10.1109/IEVC.2012.6183207.

74. Park J, Chen Z, Kiliaris L, Kuang ML, Masrur MA, Phillips AM, et al. Intelligent vehicle power control based on machine learning of optimal control parameters and prediction of road type and traffic congestion. IEEE Transactions on Vehicular Technology. 2009;58(9):4741-4756. doi: 10.1109/TVT.2009.2027710.

75. Hrovat D, Cairano SD, Tseng HE, Kolmanovsky IV. The development of Model Predictive Control in automotive industry: A survey. IEEE International Conference on Control Applications; Dubrovnik, Croatia, 3.-5.10.2012., pp. 295-302. doi:10.1109/CCA.2012.6402735.

76. Sun C, Hu X, Moura SJ, Sun F. Velocity Predictors for Predictive Energy Management in Hybrid Electric Vehicles. IEEE Transactions on Control Systems Technology. 2015;23(3):1197-1204. doi:10.1109/TCST.2014.2359176.
77. Jiang B, Fei Y. Vehicle Speed Prediction by Two-Level Data Driven Models in Vehicular Networks. IEEE Transactions on Intelligent Transportation Systems. 2017;18(7):1793-1801. doi: 10.1109/TITS.2016.2620498.
78. Goodfellow I, Bengio Y, Courville A. Deep Learning. Cambridge: MIT Press; 2016.
79. Masters T. Designing Feedforward Network Architectures. In: Neural Network Recipes in C++. Cambridge: Academic Press; 1993., p. 173-180.
80. Zhang Z, He H, Guo J, Han R. Velocity prediction and profile optimization based real-time energy management strategy for Plug-in hybrid electric buses. Applied Energy. 2020;280:116001. doi:10.1016/j.apenergy.2020.116001.
81. Mozaffari L, Mozaffari A, Azad NL. Vehicle speed prediction via a sliding-window time series analysis and an evolutionary least learning machine: A case study on San Francisco urban roads. Engineering Science and Technology, an International Journal. 2015;18(2):150-162. doi:10.1016/j.jestch.2014.11.002.
82. Liu K, Asher Z, Gong X, Huang M, et al. Vehicle Velocity Prediction and Energy Management Strategy Part 1: Deterministic and Stochastic Vehicle Velocity Prediction Using Machine Learning. SAE Technical Paper, no. 2019-01-1051, 2019. doi:10.4271/2019-01-1051.
83. Li Y, Peng J, He H, Xie S. The Study on Multi-scale Prediction of Future Driving Cycle Based on Markov Chain. Energy Procedia. 2017;105:3219–3224. doi:10.1016/j.egypro.2017.03.709.
84. Zhang L, Liu W, Qi B. Combined Prediction for Vehicle Speed with Fixed Route. Chin. J. Mech. Eng. 2020;33:60. doi:10.1186/s10033-020-00472-0.
85. Shin J, Sunwoo M. Vehicle Speed Prediction Using a Markov Chain With Speed Constraints. IEEE Transactions on Intelligent Transportation Systems. 2019;20(9):3201-3211. doi: 10.1109/TITS.2018.2877785.
86. Gaikwad TD, Asher ZD, Liu K, Huang M, et al. Vehicle Velocity Prediction and Energy Management Strategy Part 2: Integration of Machine Learning Vehicle Velocity Prediction with Optimal Energy Management to Improve Fuel Economy. SAE Technical Paper, no. 2019-01-1212, 2019. doi:10.4271/2019-01-1212.
87. Škugor B, Hrgetić M, Deur J. GPS measurement-based road slope reconstruction with application to electric vehicle simulation and analysis. 8th Conference on Sustainable Development of Energy, Water and Environment Systems (SDEWES), Dubrovnik, Croatia, 27.9.-02.10.2015.
88. Python DB-API 2.0 interface for SQLite databases. Available from: https://docs.python.org/3/library/sqlite3.html [cited 19.8.2021.];
89. Li Y, Wu H. A Clustering Method Based on K-Means Algorithm. Physics Procedia. 2012;25:1104-1109. doi:10.1016/j.phpro.2012.03.206.

90. Rasmussen CE, Williams CKI. Gaussian Processes for Machine Learning. Cambridge; 2006.
91. ScikitLearn. Available from: https://scikit-learn.org/stable [cited 19.8.2021.];
92. Nanjundan S, Sankaran S, Arjun CR, Paavai AG. Identifying the number of clusters for K-Means: A hypersphere density based approach. International Conference on Computers, Communication and Signal Processing; Chennai, India, 12.-13.3.2019., p. 1-5. arXiv:1912.00643.
93. Numpy. Available from: https://numpy.org/ [cited 19.8.2021.];
94. Häggström O. Markov chains. In: Finite Markov Chains and Algorithmic Applications. 1st ed. Cambridge: Cambridge University Press; 2002. p. 8-16.
95. Privault N. Discrete Time Markov Chains. In: Understanding Markov Chains. Singapore: Springer; 2018. pp. 89-113.
96. Gagniuc PA. Markov Chains: From Theory to Implementation and Experimentation. 1st edition. New York: John Wiley & Sons; 2017.
97. Duff IS, Erisman AM, Reid JK. Direct Methods for Sparse Matrices. Oxford: Oxford Scholarship Online; 2017.
98. Scipy. Available from: https://docs.scipy.org/doc/scipy/reference/generated/scipy.sparse.dok_matrix.html [cited 8.9.2021.]
99. Topić J, Dabčević Z, Škugor B, Deur J. Generating high-resolution city bus driving cycle database from GPS data-derived low-resolution driving cycle database [technical report]. Zagreb: Faculty of Mechanical Engineering and Naval Arhitecture; 2021.
100. Wickramanayake S, Dilum Bandara HMN. Fuel consumption prediction of fleet vehicles using Machine Learning: A comparative study. Moratuwa Engineering Research Conference (MERCon); Moratuwa, Sri Lanka, 5.-6.4.2016., p. 90-95. doi:10.1109/MERCon.2016.7480121.
101. Ben-Chaim M, Shmerling E, Kuperman A. Analytic Modeling of Vehicle Fuel Consumption. Energies. 2013;6(1):117-127. doi:10.3390/en6010117.
102. Perrotta F, Parry T, Neves LC. Application of machine learning for fuel consumption modelling of trucks. IEEE International Conference on Big Data; Boston, MA, USA, 11.-14.12.2017., p. 3810-3815. doi:10.1109/BigData.2017.8258382
103. Suh B, Chang YH, Han SB, Chung YJ. Simulation of a powertrain system for the diesel hybrid electric bus. International Journal of Automotive Technology. 2012;13(5):701–711. doi:10.1007/s12239-012-0069-5.
104. Wang J, Rakha HA. Fuel consumption model for heavy duty diesel trucks: Model development and testing. Transportation Research Part D: Transport and Environment. 2017;55:127–141. doi:10.1016/j.trd.2017.06.011.
105. Capraz AG, Ozel P, Sevkli M, Beyca OF. Fuel Consumption Models Applied to Automobiles Using Real-Time Data: A Comparison of Statistical Models. Procedia Computer Science. 2017; 83:774-781. doi:10.1016/j.procs.2016.04.166.

106. Kim Y.-R., Jung M, Park J.-B. Development of a Fuel Consumption Prediction Model Based on Machine Learning Using Ship In-Service Data. J. Mar. Sci. Eng. 2021;9(2):137. doi:10.3390/jmse9020137.

107. Bhoraskar A. Prediction of Fuel Consumption of Long Haul Heavy Duty Vehicles using Machine Learning and Comparison of the Performance of Various Learning Techniques [dissertation]. Delft: Delft University of Technology; 2019.

108. Schoen A, Byerly A, Hendrix B, Bagwe RM, Santos Ecd, Miled ZB. A Machine Learning Model for Average Fuel Consumption in Heavy Vehicles. IEEE Transactions on Vehicular Technology. 2019;68(7):6343-6351. doi:10.1109/TVT.2019.2916299.

109. Li Y, Tang G, Du J, Zhou N, et al. Multilayer Perceptron Method to Estimate Real-World Fuel Consumption Rate of Light Duty Vehicles. IEEE Access. 2019;7:63395-63402. doi:10.1109/ACCESS.2019.2914378.

110. Delgado OF, Clark NN, Thompson GJ. Modeling Transit Bus Fuel Consumption on the Basis of Cycle Properties. Journal of the Air & Waste Management Association. 2011;61(4):443-452. doi:10.3155/1047-3289.61.4.443.

111. Bifulco GN, Galante F, Pariota L, Spena MR. A Linear Model for the Estimation of Fuel Consumption and the Impact Evaluation of Advanced Driving Assistance Systems. Sustainability. 2015;7:14326-14343. doi:10.3390/su71014326.

112. Topić J, Škugor B, Deur J. Neural Network-Based Modeling of Electric Vehicle Energy Demand and All Electric Range. Energies. 2019;12(7):1396. doi:10.3390/en12071396.

113. Hihlik M, Topić J, Škugor B, Deur J. Neural Network-based Prediction of Fuel Consumption of a Conventional Delivery Vehicle based on GPS-collected Tracking Data. 15th Conference on Sustainable Development of Energy, Water and Environment Systems (SDEWES); Cologne, Germany, 1.-5.9.2020.

114. Zeiler MD, Fergus R. Visualizing and Understanding Convolutional Neural Networks. 13th European Conference on Computer Vision; Zurich, Switzerland, 6.-12.9.2014., p. 818–833. arXiv:1311.2901

115. Szegedy C, Liu W, Jia Y, Sermanet P, Reed S, Anguelov D, Erhan D, Vanhoucke V, Rabinovich A. Going Deeper with Convolutions. IEEE Conference on Computer Vision and Pattern Recognition; Boston, MA, USA, 7.-12.6.2015. doi:10.1109/CVPR.2015.7298594.

116. Kedia RK, Naick BK. Review of Vehicle Route Optimisation. 2nd IEEE International Conference on Intelligent Transportation Engineering (ICITE), Singapore, Singapore, 1.-3.9.2017., p. 57-61. doi:10.1109/ICITE.2017.8056881.

117. Nielsen M. Neural Networks and Deep Learning. Determination Press; 2015. Available from: http://neuralnetworksanddeeplearning.com/ [cited 1.10.2021.];

118. Novaković B, Majetić D, Široki M. Umjetne neuronske mreže. Zagreb: Fakultet strojarstva i brodogradnje; 2011.

119. Abiodun OI, Jantan A, Omolara AE, Dada KV, Mohamed NA, Arshad H. State-of-the-art in artificial neural network applications: A survey. Heliyon. 2018;4(11):e00938. doi:10.1016/j.heliyon.2018.e0093.

120. Sutskever I, Martens J, Dahl G, Hinton G. On the importance of initialization and momentum in deep learning. 30th International Conference on Machine Learning; Atlanta, GA, USA, 16.21.6.2013., p. 1139-1147. doi:10.5555/3042817.3043064.
121. Keras. Available from: https://keras.io [cited 27.10.2021.];
122. Tensorflow. Available from: https://tensorflow.org [cited 27.10.2021.];
123. Smith LN. Cyclical Learning Rates for Training Neural Networks. IEEE Winter Conference on Applications of Computer Vision (WACV); Santa Rosa, CA, 24.-31.3.2017., p. 464-472. doi: 10.1109/WACV.2017.58.
124. Jolliffe IT. Principal Component Analysis. 2nd edition. New York: Springer; 2002.
125. Metcalf L, Casey W. Chapter 2 - Metrics, Similarity, and Sets. In: Cybersecurity and Applied Mathematics. Rockland: Syngress; 2016., p. 3-22.

Prilozi

Prilog A – Kandidirane značajke voznih ciklusa

A.1. Značajke vremenske domene

A.1.1. Pregled punog skupa značajki

Značajke vremenske domene uključuju opće indekse diskretnih vremenskih profila varijabli, poput minimalne vrijednosti ($x_{\min}$), maksimalne vrijednosti ($x_{\max}$), srednje vrijednosti ($\bar{x}$) i standardne devijacije (σ) te korijena srednjeg kvadrata (*engl. Root Mean Square*, RMS) (tablica A-1), kao i složenije indekse koji su nužni za opisivanje dinamike voznog ciklusa, karakteristika autobusne rute i međuovisnosti određenih varijabli (vidi odjeljak A.1.2 i tablicu A-2). Pritom se neke značajke poput srednje vrijednosti i standardne devijacije mogu računati samo za pozitivne odnosno negativne vrijednosti dane varijable. Ukupni broj značajki vremenske domene iznosi 77.

Minimalna i maksimalna vrijednost varijable x određuje se prema

$$x_{\min} = \min(x), \tag{A-1}$$

$$x_{\max} = \max(x). \tag{A-2}$$

Srednja vrijednost definirana je izrazom

$$\bar{x} = \frac{1}{N}\sum_{i=1}^{N} x_i, \tag{A-3}$$

gdje je N ukupni broj uzorkovanih vrijednosti u x.

Standardna devijacija σ definirana je izrazom

$$\sigma = \sqrt{\frac{1}{N-1}\sum_{i=1}^{N}(x_i - \bar{x})^2}. \tag{A-4}$$

Korijen srednje vrijednosti kvadrata signala (engl. skrać. RMS) određen je izrazom

$$RMS = \sqrt{\frac{1}{N}\sum_{i=1}^{N} x_i^2}. \tag{A-5}$$

Dodatno korišteni, izvedeni indeksi definirani su u narednim odjeljcima prema kategoriji kojoj pripadaju.

Tablica A-1: Opće statističke značajke povezane s diskretnim vremenskim profilima brzine vozila, ubrzanja vozila, nagiba ceste i njihovih izvedenica (znak + označava da se dana značajka primjenjuje).

Razmatrani diskretni vremenski profili (vidi poglavlje 5)	**Statistička značajka**								
	$x_{\min}$	$x_{\max}$	$\bar{x}$	$\bar{x}(x>0)$	$\bar{x}(x<0)$	σ	$\sigma(x>0)$	$\sigma(x<0)$	RMS
Brzina vozila v [km/h]		+	+	+		+			+
Ubrzanje vozila a [m/s^2]	+	+	+	+	+	+	+	+	+
Nagib ceste θ [deg]	+	+	+	+	+	+	+	+	+
Zahtijevana snaga na kotaču P_w [kW]	+	+	+	+	+	+	+	+	+
Horizontalna komponenta brzine vozila v_h [km/h]		+	+			+			+
Vertikalna komponenta brzine vozila v_v [km/h]	+	+	+	+	+	+	+	+	+
Specifična snaga vozila P_m [kW/kg]	+	+	+	+	+	+	+	+	+

Tablica A-2. Popis statističkih značajki povezanih s karakteristikama rute, dinamikom vožnje i faktorima unakrsne korelacije brzine vozila, ubrzanja vozila i nagiba ceste.

Statističke značajke po kategorijama		
Karakteristike rute	**Dinamika vožnje**	**Faktori unakrsne korelacije**
1) Postotak vremena vožnje pod negativnim ubrzanjem T_a^- [%]	1) Relativna pozitivna akceleracija (RPA) [m^2/s^2]	1) Korelacija brzine vozila i pozitivnog nagiba ceste ρ_{v,θ^+}
2) Postotak vremena vožnje pod pozitivnim ubrzanjem T_a^+ [%]	2) Pozitivna kinetička energija (PKE) [m^2/s^2]	2) Korelacija brzine vozila i negativnog nagiba ceste ρ_{v,θ^-}
3) Postotak vremena vožnje pod pozitivnom nagibom ceste T_θ^+ [%]	3) Srednja specifična energija voznog ciklusa $\bar{E}_{vc}$ [MJ/km]	3) Korelacija brzine i pozitivnog ubrzanja vozila ρ_{v,a^+}
4) Postotak vremena vožnje pod negativnim nagibom ceste T_θ^- [%]	4) Srednja specifična energija vožnje $\bar{E}_{vc}^+$ [MJ/km]	4) Korelacija brzine i negativnog ubrzanja vozila ρ_{v,a^-}
5) Postotak vremena mirovanja vozila T_m [%]	5) Srednja specifična energija kočenja $\bar{E}_{vc}^-$ [MJ/km]	5) Korelacija pozitivnog ubrzanja vozila i pozitivnog nagiba ceste ρ_{a^+,θ^+}
6) Broj zaustavljanja vozila po prijeđenom kilometru (*BZPK*) [km^{-1}]	6) Apsolutna (ukupna) potrošnja goriva V_f [L]	6) Korelacija pozitivnog ubrzanja vozila i negativnog nagiba ceste ρ_{a^+,θ^-}
7) Prosječno vrijeme mirovanja vozila $\bar{T}_m$ [s]	7) Prosječna (specifična) potrošnja goriva $\bar{V}_f$ [L/100 km]	7) Korelacija negativnog ubrzanja vozila i pozitivnog nagiba ceste ρ_{a^-,θ^+}
8) Prosječna udaljenost između dva zaustavljanja vozila $\bar{d}_z$ [m]		8) Korelacija negativnog ubrzanja vozila i negativnog nagiba ceste ρ_{a^-,θ^-}
9) Trajanje putovanja T_{vc} [min]		
10) Duljina rute d [km]		

A.1.2. Značajke dinamike vožnje

Za potrebe opisivanja dinamike vožnje koriste se izvedeni indeksi poput relativne pozitivne akceleracije (RPA) i pozitivne kinetičke energije (PKE), definirani kao [4]:

$$RPA = \frac{1}{d}\sum_{i}^{N} v_i \cdot a_i \cdot \Delta t_i\,; \quad a_i > 0, \tag{A-6}$$

$$PKE = \frac{1}{d}\sum_{i}^{N} v_i^2 - v_{i-1}^2; \quad v_i > v_{i-1}, \tag{A-7}$$

gdje je v_i brzina vozila, a_i ubrzanje vozila, d ukupna prijeđena udaljenost, a $\Delta t_i = t_{i+1} - t_i$ vremenska razlika između susjednih uzoraka u i-tom diskretnom vremenskom koraku. Vozni ciklusi s vrijednostima $RPA < 0.1$ m/s^2 smatraju se mekim (npr. vožnja autocestom), dok vrijednosti $RPA > 0.2$ m/s^2 ukazuju na dinamičnije/agresivnije vožnje s učestalim i izraženijim ubrzanjima. U slučaju PKE, tipične vrijednosti za urbani vozni ciklus kreću se oko 8 m/s^2.

Nadalje, iz zahtijevane snage na kotaču vozila definira se kumulativna energija voznog ciklusa E_{vc}, koja predstavlja značajnu varijablu iz energetske perspektive vozila [58]

$$E_{vc} = \sum_{i}^{N} P_{w,i} \cdot \Delta t_i. \tag{A-8}$$

Kumulativna energija voznog ciklusa E_{vc} dodatno se razlaže na pozitivnu komponentu E_{vc}^{+} (energija vožnje) i negativnu komponentu E_{vc}^{-} (energija kočenja) prema

$$E_{vc}^{+} = \sum_{i}^{N} P_{w,i} \cdot \Delta t_i\,; \;\; \forall\, \boldsymbol{P}_{w,i} > 0, \tag{A-9}$$

$$E_{vc}^{-} = \sum_{i}^{N} \boldsymbol{P}_{w,i} \cdot \Delta t_i\,; \;\; \forall\, \boldsymbol{P}_{w,i} < 0. \tag{A-10}$$

Iz (A-8), (A-9) i (A-10) se potom izvlače tri statističke značajke definirane kao (i) srednja specifična energija voznog ciklusa $\bar{E}_{vc}$ (MJ/km), (ii) srednja specifična energija vožnje $\bar{E}_{vc}^{+}$ (MJ/km), te (iii) srednja specifična energija kočenja $\bar{E}_{vc}^{-}$ (MJ/km), na način da se kumulativna enegija voznog ciklusa E_{vc}^{+} i pripadna pozitivna i negativna komponente (E_{vc}^{+}, E_{vc}^{-}) podijele s ukupnom prijeđenom udaljenošću d (uključujući dijeljenje s 10^6 kako bi se dobili MJ).

Apsolutna (ukupna) potrošnja goriva V_f (L) se u slučaju sintetičkih voznih ciklusa dobiva primjenom preciznog *NN-H3D* modela za predviđanje potrošnje goriva (definiranog u poglavlju 4), dok se dijeljenjem V_f s ukupnom prijeđenom udaljenošću d dobiva prosječna potrošnja goriva $\bar{V}_f$ (L/100 km).

A.1.3. Karakteristike rute

Statističke značajke kojima se opisuju karakteristike rute uključuju postotke vremena vožnje pod pozitivnim/negativnim ubrzanjem (T_a^+, T_a^-) i nagibom ceste (T_θ^+, T_θ^-), postotak vremena mirovanja vozila (T_m), prosječno vrijeme mirovanja vozila ($\bar{T}_m$), broj zaustavljanja vozila po prijeđenom kilometru (*BZPK*), prosječnu udaljenost između zaustavljanja vozila ($\bar{d}_z$), trajanje putovanja (tj. voznog ciklusa) T_{vc}, te duljinu rute, odnosno, ukupno prijeđenu udaljenost d.

Pritom se postotci vremena vožnje pod pozitivnim T_a^+ i negativnim T_a^- ubrzanjem vozila izračunavaju prema

$$T_a^+ = \frac{100\%}{T_{vc}} \cdot \sum_{i=1}^{N} \Delta t_i \,; \quad \forall\, a_i > 0, \tag{A-11}$$

$$T_a^- = \frac{100\%}{T_{vc}} \cdot \sum_{i=1}^{N} \Delta t_i \,; \quad \forall\, a_i < 0. \tag{A-12}$$

Na isti način dobivaju se i postotci vremena vožnje pod pozitivnim T_θ^+ i negativnim T_θ^- nagibom ceste, te postotak vremena mirovanja vozila T_m

$$T_\theta^+ = \frac{100\%}{T_{vc}} \cdot \sum_{i=1}^{N} \Delta t_i \,; \quad \forall\, \theta_i > 0, \tag{A-13}$$

$$T_\theta^- = \frac{100\%}{T_{vc}} \cdot \sum_{i=1}^{N} \Delta t_i \,; \quad \forall\, \theta_i < 0, \tag{A-14}$$

$$T_m = \frac{100\%}{T_{vc}} \cdot \sum_{i=1}^{N} \Delta t_i \,; \quad \forall\, v_i = 0. \tag{A-15}$$

Nadalje, zaustavljanje vozila registrira se za svaki i-ti diskretni vremenski trenutak za koji vrijedi uvjet da je brzina vozila jednaka $v_{i+1} = 0$ i $v_i > 0$, iz čega proizlazi ukupni broj zaustavljanja vozila (*BZ*) kao

$$BZ = \sum_{i}^{N-1} \begin{cases} 1, & v_i > 0 \,\wedge\, v_{i+1} = 0 \\ 0, & \text{u suprotnom} \end{cases}. \tag{A-16}$$

Broj zaustavljanja vozila po prijeđenom kilometru *BZPK* izračunava se dijeljenjem registriranog broja zaustavljanja vozila *BZ* s ukupnom prijeđenom udaljenošću d:

$$BZPK = \frac{BZ}{d}. \tag{A-17}$$

Prosječno vrijeme mirovanja vozila $\bar{T}_m$ prilikom zaustavljanja izračunava se dijeljenjem ukupnog vremena mirovanja $T_m/100$ s brojem zaustavljanja vozila *BZ*

$$\bar{T}_m = \frac{\frac{T_m}{100}}{BZ}. \tag{A-18}$$

Na sličan način se izračunava prosječna udaljenost između dva zaustavljanja vozila $\bar{d}_z$:

$$\bar{d}_z = \frac{d}{BZ}. \tag{A-19}$$

A.1.4. Značajke unakrsne korelacije brzine, ubrzanja i nagiba ceste

Za potrebe opisivanja međuovisnosti između brzine vozila, ubrzanja vozila i nagiba ceste, prema trećem stupcu tablice A-2, korišten je korelacijski indeks $\rho_{x,y}$, koji predstavlja mjeru sličnosti dva vektora $\boldsymbol{x}$ i $\boldsymbol{y}$, oba duljine *N*, a definiran je kao

$$\rho_{x,y} = \frac{\sum(\boldsymbol{x}, \boldsymbol{y})}{\sigma_x \cdot \sigma_y}, \tag{A-20}$$

gdje su σ_x i σ_y standardne devijacije $\boldsymbol{x}$ i $\boldsymbol{y}$ izračunate preko izraza (A-4), dok je kovarijanca $\sum(\boldsymbol{x}, \boldsymbol{y})$ definirana kako slijedi

$$\sum(\boldsymbol{x}, \boldsymbol{y}) = \frac{1}{N-1}\sum_{i=1}^{N}(\boldsymbol{x}_i - \bar{x}) \cdot (\boldsymbol{y}_i - \bar{y}). \tag{A-21}$$

Indeks $\rho_{x,y}$ poprima vrijednosti u intervalu [-1, 1], koja iznosi $\rho_{x,y} = 1$ u slučaju potpuno jednakih ulaznih vektora (puna sličnost), te $\rho_{x,y} = 0$ kada ne postoji korelacija, dok $\rho_{x,y} = -1$ označava antikorelaciju (ulazni vektori su potpuno jednaki, ali suprotnog smjera). U Pythonu se korelacija dvaju signala računa korištenjem funkcije *corrcoef* iz *numpy* modula [93].

A.2. Značajke frekvencijske domene

Pripadajuće statističke značajke izvedene su iz amplitudno-frekvencijskog odziva diskretnih vremenskih profila ključnih varijabli ($\boldsymbol{v}, \boldsymbol{a}, \boldsymbol{\theta}, \boldsymbol{v}_h, \boldsymbol{v}_v, \boldsymbol{P}_w, \boldsymbol{P}_m$; tablica A-1), koji su dobiveni primjenom brze Fourierove transformacije (*engl. fast Fourier transform*, FFT) primjenom *numpy.fft* modula u Pythonu [93].

FFT pretvara konačni niz od *N* jednako raspoređenih uzoraka $\{X_n\} \coloneqq X_0, X_1, \ldots, X_{N-1}$ u niz kompleksnih brojeva iste duljine $\{\hat{X}_k\} \coloneqq \hat{X}_0, \hat{X}_1, \ldots, \hat{X}_{N-1}$, koji su funkcija frekvencije:

$$\hat{X}_k = \sum_{n=0}^{N-1} X_n \cdot e^{-i2\pi k\frac{n}{N}} = \sum_{n=0}^{N-1} X_n \cdot \left[\cos\left(\frac{2\pi}{N}kn\right) - i \cdot \sin\left(\frac{2\pi}{N}kn\right)\right], \quad \text{(A-22)}$$

gdje je n indeks vremena, k je indeks frekvencije, a svaki $\hat{X}_k$ je kompleksan broj koji kodira amplitudu i fazu složene sinusoidne komponente $e^{-i2\pi k\frac{n}{N}}$ funkcije X_n.

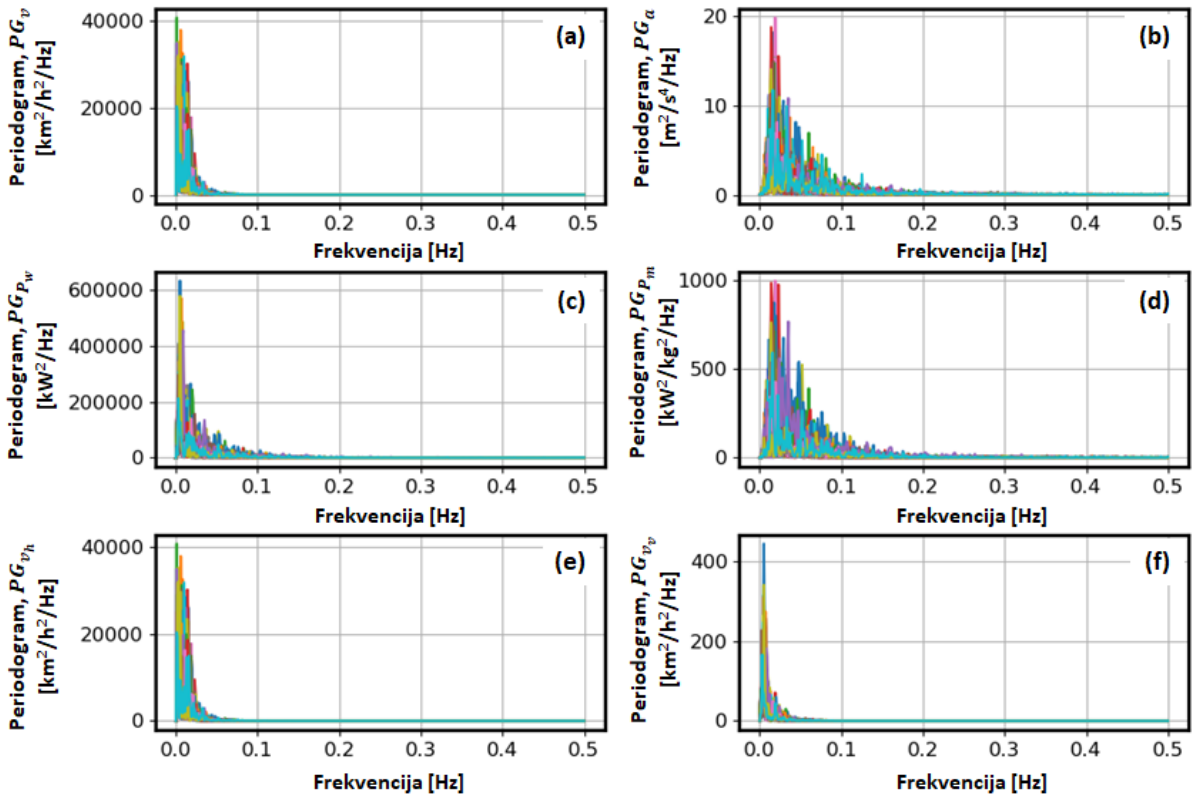
Za potrebe definiranja amplitudnih značajki frekvencijske domene iz frekvencijskog odziva $\hat{X}_k$, izračunava se periodogram PG definiran kao

$$PG(k) = \frac{1}{N}\left|\hat{X}(k)\right|^2 = \frac{1}{N}\left(\hat{X}_{re}^2(k) + \hat{X}_{im}^2(k)\right), \quad \text{(A-23)}$$

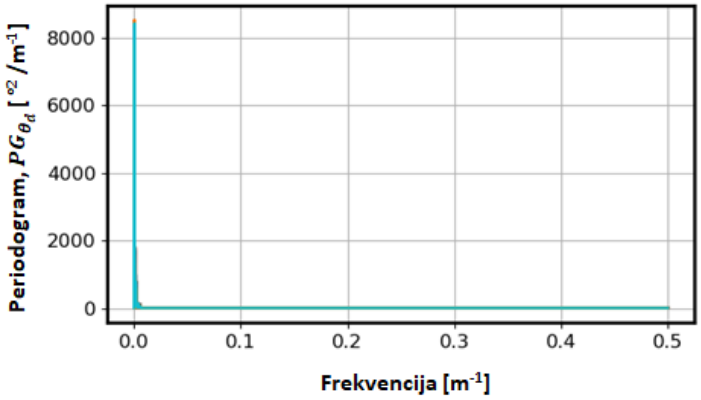
gdje $\hat{X}_{re}$ i $\hat{X}_{im}$ predstavljaju realnu i imaginarnu komponentu $\hat{X}_k$.

Primjenom (A-22) i (A-23) na diskretne vremenske profile $\boldsymbol{v}, \boldsymbol{a}, \boldsymbol{v}_h, \boldsymbol{v}_v, \boldsymbol{P}_w, \boldsymbol{P}_m$ dobivaju se pripadni periodogrami $\widehat{\boldsymbol{v}}, \widehat{\boldsymbol{a}}, \widehat{\boldsymbol{v}}_h, \widehat{\boldsymbol{v}}_v, \widehat{\boldsymbol{P}}_w, \widehat{\boldsymbol{P}}_m$, koji su za slučaj 100 snimljenih voznih ciklusa iscrtani na slici A-1 u različitim bojama. U slučaju brzine vozila $\widehat{\boldsymbol{v}}$, sve značajne frekvencijske komponente nalaze se unutar niskofrekventnog područja (<0.1 Hz; slika A-1a). Na slici A-1b je vidljivo kako se frekvencijske komponente ubrzanja vozila $\widehat{\boldsymbol{a}}$ također gomilaju u niskofrekventnom području (<0.2 Hz), uz značajnije vrhove na nekoliko frekvencija. Slični trend kao za $\widehat{\boldsymbol{a}}$ prate i frekvencijske komponente zahtjevane snage na kotaču $\widehat{\boldsymbol{P}}_w$ i specifične snage $\widehat{\boldsymbol{P}}_m$ (slike A-1c i A-1d). Konačno, budući da $\widehat{\boldsymbol{v}}_h$ i $\widehat{\boldsymbol{v}}_v$ predstavljaju horizontalnu i vertikalnu komponentu brzine, njihovi frekvencijski odzivi (slike A-1e i A-1f) nalikuju onom dobivenom za longitudinalnu brzinu vozila $\widehat{\boldsymbol{v}}$.

Transformacija profila nagiba ceste u frekvencijsku domenu $\widehat{\boldsymbol{\theta}}$ provodi se također primjenom FFT, ali u domeni prijeđene udaljenosti d. Prethodno se izvorni diskretni vremenski profili nagiba ceste $\boldsymbol{\theta}$ s varijabilnim prijeđenim putem između susjednih uzoraka ($d_k - d_{k-1} \neq d_{k+1} - d_k$) interpoliraju kako bi imali fiksni korak puta od 1 m. Na slici A-2 prikazan je frekvencijski odziv profila nagiba ceste $\widehat{\boldsymbol{\theta}}$ za slučaj 100 snimljenih voznih ciklusa. Vidljivo je kako se sve značajne frekvencijske komponente nalaze unutar niskofrekventnog područja (<0.005 m^{-1}; slika A-2), što je i očekivano budući da su periode oscilacija nagiba ceste kojim prometuju gradski autobusi duge barem par stotina metara.



Slika A-1: Amplitudno-frekvencijski odzivi (periodogrami) diskretnih vremenskih profila a) brzine, b) ubrzanja, c) snage na kotačima, d) specifične snage, te e) horizontalne i f) vertikalne komponente brzine.



Slika A-2: Amplitudno-frekvencijski odzivi (periodogrami) diskretnih profila nagiba ceste.

Za potrebe opisivanja značajki frekvencijskih odziva koriste se neki od općih statističkih indeksa, tj. srednja vrijednost, standardna devijacija i RMS vrijednost, koji su definirani u potpoglavlju A.1.1, te dva dodatna indeksa za slučaj nagiba ceste, koji se definiraju kako slijedi. Najprije se frekvencijska os f dijeli u dva intervala (moda), čiji su rasponi određeni vizualnom inspekcijom periodograma nagiba ceste na slici A-3, te iznose $[0, 0.0008\rangle$ m^{-1} za prvi mod i

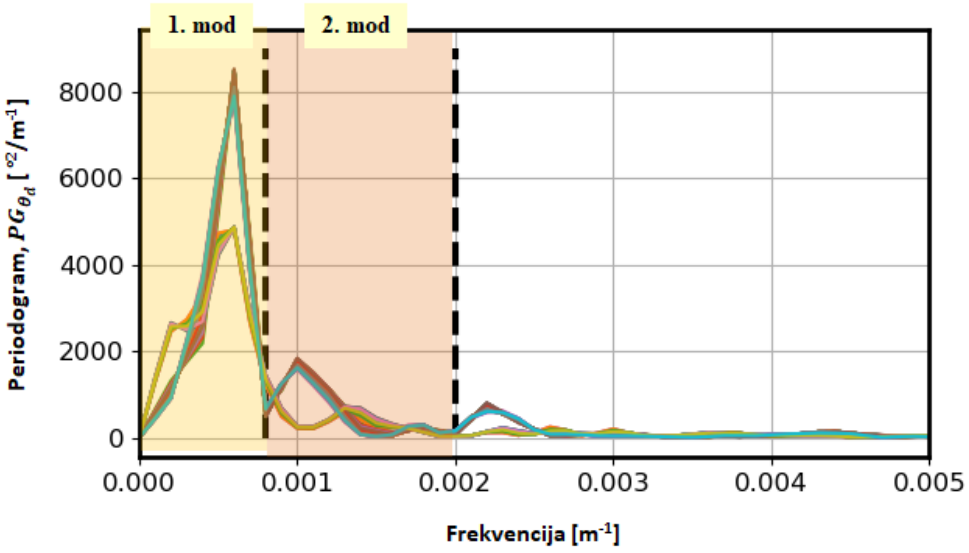
$[0.0008, 0.002\rangle$ m^{-1} za drugi mod. Zatim se izračunava omjer maksimalne i srednje vrijednosti periodograma $PG_{\hat{\boldsymbol{\theta}}_d}$ za svaki mod:

$$M_1 = \frac{\max\left(PG_{\hat{\boldsymbol{\theta}}_d}\right)}{\overline{PG}_{\hat{\boldsymbol{\theta}}_d}}; \quad \forall\, f \in [0, 0.0008\rangle\,, \tag{A-23}$$

$$M_2 = \frac{\max\left(PG_{\hat{\boldsymbol{\theta}}_d}\right)}{\overline{PG}_{\hat{\boldsymbol{\theta}}_d}}; \quad \forall\, f \in [0.0008, 0.002\rangle, \tag{A-24}$$

kao mjera izraženih oscilacija nagiba u danom frekvencijskom području, tj. za dani frekvencijski mod.

Ukupni broj značajki frekvencijske domene iznosi 23. FFT značajke individualnih varijabli izračunavaju se na frekvencijskom intervalu u kojem se nalaze značajne frekvencijske komponente, odnosno, na intervalu $f \in [0, 0.1\rangle$ Hz za $\hat{\boldsymbol{v}}, \hat{\boldsymbol{v}}_h$ i $\hat{\boldsymbol{v}}_v$, intervalu $f \in [0, 0.2\rangle$ Hz za $\hat{\boldsymbol{a}}, \hat{\boldsymbol{P}}_w$ i $\hat{\boldsymbol{P}}_m$, te intervalu $f \in [0, 0.002\rangle$ m^{-1} za $\hat{\boldsymbol{\theta}}$.



Slika A-3: Ilustracija odvajanja periodograma nagiba ceste u nisko- i srednje-frekvencijski mod za potrebe računanja značajki M_1 i M_2.

Životopis

Jakov Topić, rođen 27. studenog 1990., stekao je titulu magistra inženjera strojarstva 2017. godine na Fakultetu strojarstva i brodogradnje Sveučilišta u Zagrebu. Od veljače 2017. godine do travnja 2019. godine radi kao stručni suradnik na Fakultetu strojarstva i brodogradnje Sveučilišta u Zagrebu, gdje također pohađa doktorski studij u području mehatronike i robotike. Od svibnja 2019. radi kao mlađi istraživač na projektu „Napredne metode i tehnologije u znanosti o podatcima i kooperativnim sustavima“ (DATACROSS) Znanstvenog centra izvrsnosti za znanost o podatcima i kooperativne sustave. Sudjelovao je na projektu „Smart Solutions supporting Low Emission Zones and other low-carbon mobility policies in EU cities (SOLEZ)“ podržanom od strane Europske komisije kroz Interreg CE okvir, istraživačkom projektu „Adaptivno i prediktivno upravljanje utičnim hibridnim električnim vozilima (ACHIEVE)“ podržanom od strane Hrvatske zaklade za znanost, te na istraživačkom projektu „Safe Speed Anticipating for Encountering Pedestrian Crossing“ podržanom od tvrtke Ford Motor Company.

Njegovi istraživački interesi uključuju razvoj i implementaciju metoda sintezu višedimenzionalnih sintetičkih voznih ciklusa, planiranje elektrifikacije cestovnog prometa, te primjenu umjetne inteligencije u automobilskim i prometnim sustavima. Autor je ili koautor četiri radova u časopisima indeksiranim u CC/SCI bazama, jednog rada u ostalim međunarodnim časopisima referenciranim u SCOPUS bazi, osam radova publiciranih na međunarodnim kongresima i konferencijama, te niza studija i tehničkih izvještaja.

Popis javno objavljenih radova:

1. J. Topić, B. Škugor, J. Deur, "Neural Network based Prediction of Vehicle Fuel Consumption based on Driving Cycle Data", Sustainability, Vol. 14, No. 2, pp. 12, 2022.
2. J. Topić, B. Škugor, J. Deur, "Synthesis and Feature Selection-Supported Validation of Multidimensional Driving Cycles", Sustainability, Vol. 13, No. 9, pp. 21, 2021.
3. J. Topić, J. Soldo, F. Maletić, B. Škugor, J. Deur, "Virtual Simulation of Electric Bus Fleets for City Bus Transport Electrification Planning", Energies, Vol. 13, No. 13, pp. 23, 2020.
4. J. Topić, B. Škugor, J. Deur, "Neural Network-Based Modeling of Electric Vehicle Energy Demand and All Electric Range", Energies, Vol. 12, No. 7, pp. 1396-1416, 2019.

5. J. Topić, B. Škugor, J. Deur, "Synthesis and Validation of Multidimensional Driving Cycles", SAE Int. J. Adv. & Curr. Prac. in Mobility, Vol. 3, No. 4, pp. 1558-1568, 2021.
6. B. Škugor, J. Topić, J. Deur, V. Ivanović, E. Tseng, "Analysis of a Game Theory-based Model of Vehicle-Pedestrian Interaction at Uncontrolled Crosswalks", International Conference on Smart Systems and Technologies 2020 (SST 2020), Osijek, Croatia, 2020.
7. J. Topić, B. Škugor, J. Deur, "Analysis of Markov Chain-based Methods for Synthesis of Driving Cycles of Different Dimensionality", 2020 IEEE Intelligent Transportation Systems Conference (ITSC), Rhodes, Greece, 2020.
8. M. Hihlik, J. Topić, B. Škugor, J. Deur, "Neural Network-based Prediction of Fuel Consumption of a Conventional Delivery Vehicle based on GPS-collected Tracking Data", 15th Conference on Sustainable Development of Energy, Water and Environment Systems (SDEWES), Cologne, Germany, 2020.
9. J. Topić, B. Škugor, J. Deur, "Analysis of City Bus Driving Cycle Features for the Purpose of Multidimensional Driving Cycle Synthesis", SAE paper #2020-01-1288, 2020 SAE World Congress, Detroit, MI, 2020.
10. J. Deur, J. Topić, F. Maletić, "Electrification of City Bus Transport: An Overview and SOLEZ-gained Experience", International Conference on Smart solutions for urban and regional mobility in Europe, Brno, Czech Republic, 2019.
11. J. Topić, B. Škugor, J. Deur, "Neural Network-based Modelling of Extended Range Electric Vehicle Energy Demand and All Electric Range", 13th Conference on Sustainable Development of Energy, Water and Environment Systems (SDEWES), Palermo, Italy, 2018.
12. J. Topić, F. Maletić, B. Škugor, J. Deur, "Software Tool for Planning of City Bus Transport Electrification", Science and Traffic Development International Conference, Opatija, Croatia, 2018.
13. B. Škugor, J. Topić, J. Deur, “Quadratic Programming-based Electric Vehicle Charging Optimisation combining Charging Cost and Grid Power Peak Minimisation”, 12th Conference on Sustainable Development of Energy, Water and Environment Systems (SDEWES), Dubrovnik, Croatia, 2017.

Curriculum vitae

Jakov Topić, born on November 27, 1990, received his M.Sc. degree in Mechanical Engineering from the University of Zagreb, Croatia, in 2017. From February 2017 to April 2019, he has been working as an Expert Associate at the Faculty of Mechanical Engineering and Naval Architecture, University of Zagreb, where he also attends PhD study in the field of mechatronics and robotics. Since May 2019, he has been working as a Researcher on the project "Advanced Methods and Technologies in Data Science and Cooperative Systems" (DATACROSS) of The Centre of Research Excellence for Data Science and Advanced Cooperative Systems. He has participated on the project "Smart Solutions supporting Low Emission Zones and other low-carbon mobility policies in EU cities (SOLEZ)" supported by the European Commission through Interreg CE framework, the research project "Adaptive and predictive control of plug-in hybrid electric vehicles (ACHIEVE)" supported by Croatian Science Foundation, and on „Safe Speed Anticipating for Encountering Pedestrian Crossing“ project supported by Ford Motor Company.

His research interests include the synthesis of multidimensional driving cycles, electrification of land/road transport, and the application of artificial intelligence in automotive and traffic systems. He is author or co-author of four journal papers indexed in CC/SCI databases, one paper in other international journals referenced in SCOPUS database, eight papers presented at international congresses and conferences, and a number of studies and technical reports.