

UNIVERZA NA PRIMORSKEM  
FAKULTETA ZA MATEMATIKO, NARAVOSLOVJE IN  
INFORMACIJSKE TEHNOLOGIJE

DOKTORSKA DISERTACIJA

AKUSTIČNA KLASIFIKACIJA MORSKEGA DNA

SAŠO MOŠKON

KOPER, 2015



UNIVERZA NA PRIMORSKEM  
FAKULTETA ZA MATEMATIKO, NARAVOSLOVJE IN  
INFORMACIJSKE TEHNOLOGIJE

DOKTORSKA DISERTACIJA

AKUSTIČNA KLASIFIKACIJA MORSKEGA DNA

SAŠO MOŠKON

KOPER, 2015

MENTOR: IZR. PROF. DR. JANEZ ŽIBERT  
SOMENTOR: DOC. DR. BRANKO KAVŠEK



# Kazalo

<b>Kazalo slik</b>	<b>iii</b>
<b>Kazalo tabel</b>	<b>vi</b>
<b>Izvleček</b>	<b>vii</b>
<b>Abstract</b>	<b>x</b>
<b>Seznam kratic</b>	<b>xii</b>
<b>1 Uvod</b>	<b>1</b>
1.1 Metodologija . . . . .	2
1.1.1 Načrtovanje in izvedba meritev . . . . .	2
1.1.2 Preverjanje kakovosti in čiščenje podatkov . . . . .	3
1.1.3 Predobdelava podatkov . . . . .	3
1.1.4 Izračun in izbor značilk . . . . .	4
1.1.5 Zbiranje podatkov o tipih morskega dna . . . . .	5
1.1.6 Klasifikacija morskega dna . . . . .	5
1.1.7 Aplikacija v realni domeni in 3R vizualizacija rezultatov . . . . .	5
1.1.8 Povratna zanka . . . . .	6
1.2 Prispevki k znanosti . . . . .	6
1.3 Pregled vsebine . . . . .	8
<b>2 Osnovna znanja</b>	<b>11</b>
2.1 Slovensko morje . . . . .	11
2.2 Splošni opis sonarja . . . . .	12
2.3 Podvodna akustika . . . . .	13
2.4 Odkrivanje zakonitosti v podatkih . . . . .	17
2.5 Klasifikacija morskega dna . . . . .	19
2.5.1 Neposredno vzorčenje z metodo linijskih presekov . . . . .	20
2.5.2 Zračni in satelitski sistemi . . . . .	20
2.5.3 Akustična klasifikacija morskega dna . . . . .	21
2.5.4 Pristopi h klasifikaciji . . . . .	23

<b>3</b>	<b>Izvedba hidrografskih meritev</b>	<b>25</b>
3.1	Določanje položaja plovila in natančnost pridobljenih podatkov . . .	26
3.2	Mnogosnopni sonar Reson SeaBat 8125 . . . . .	28
3.3	Kalibracija batimetričnega sistema . . . . .	29
3.4	Izvedba meritev . . . . .	31
<b>4</b>	<b>Zbiranje podatkov o morskem dnu</b>	<b>35</b>
4.1	Točkovno vzorčenje s potapljači . . . . .	35
4.2	Letalski posnetki . . . . .	35
4.2.1	Opis uporabljene opreme in potek meritev . . . . .	37
4.3	Video snemanje . . . . .	38
<b>5</b>	<b>Predobdelava podatkov</b>	<b>43</b>
5.1	Opis podatkov . . . . .	43
5.2	Georeferenciranje podatkov . . . . .	45
5.3	Preverjanje kakovosti . . . . .	47
5.4	Izločanje vpliva nastavitve sistema . . . . .	51
5.4.1	Moč oddajnika in občutljivost sprejemnika . . . . .	51
5.4.2	Absorpcija in sferično širjenje zvoka . . . . .	53
5.4.3	Velikost akustičnega odtisa . . . . .	57
5.5	Uravnavanje kotnih odvisnosti . . . . .	57
<b>6</b>	<b>Opis podatkov in izbor značilk</b>	<b>63</b>
6.1	Definicija območja . . . . .	63
6.1.1	Podatki o razredu . . . . .	65
6.2	Opis značilk . . . . .	66
6.2.1	Osnovne značilke . . . . .	68
6.2.2	Asimetrija in sploščenost . . . . .	70
6.2.3	Teksturne značilke . . . . .	71
6.2.4	Pregled značilk . . . . .	72
6.2.5	Splošne lastnosti značilk . . . . .	73
6.3	Ovrednotenje različnih pristopov k uravnavanju kotnih odvisnosti . .	74
6.3.1	Opis podatkov in uporabljenih značilk . . . . .	74
6.3.2	Postopek testiranja . . . . .	76
6.3.3	Rezultati . . . . .	76
6.4	Ovrednotenje in izbor značilk . . . . .	78
6.4.1	Mere ocenjevanja značilk in uporabljene metode . . . . .	80
6.4.2	Izbira velikosti celic . . . . .	81
6.4.3	Izbir značilk . . . . .	83
6.5	Zaključek . . . . .	86
<b>7</b>	<b>Klasifikacija morskega dna</b>	<b>87</b>
7.1	Uporabljene metode in tehnike . . . . .	87
7.1.1	Uporabljeni algoritmi za klasifikacijo . . . . .	88
7.1.2	Metode ovrednotenja klasifikatorjev . . . . .	93
7.2	Množica podatkov in značilke . . . . .	96

---

7.2.1	Vizualizacija značilnk . . . . .	98
7.3	Rezultati . . . . .	100
7.4	Poobdelava rezultatov klasifikacije . . . . .	105
<b>8</b>	<b>Aplikacija metodologije v realni domeni</b>	<b>109</b>
8.1	Kartiranje morskih travnikov . . . . .	109
8.1.1	Uporabljena metodologija . . . . .	111
8.1.2	Rezultati in primerjava z obstoječimi kartami . . . . .	112
8.1.3	Predstavitev rezultatov v 3R GIS okolju . . . . .	121
<b>9</b>	<b>Sklep</b>	<b>123</b>
9.1	Smernice za nadaljnje delo . . . . .	126
	<b>Literatura</b>	<b>128</b>





# Slike

2.1	Primerjava različnih akustičnih sistemov. . . . .	12
2.2	Primer gibanja temperature morja za leto 2011. . . . .	14
2.3	Spreminjanje hitrosti zvoka v vodnem stolpcu Koprskega zaliva. Primer meritve z dne 4.9.2014. . . . .	15
2.4	Podatkovno rudarjenje po metodologiji CRISP-DM. . . . .	18
3.1	Prikaz ladje Lyra in delovanja mnogospnega sonarja. . . . .	26
3.2	Napaka pozicije globin zaradi odstopanja nagiba plovila. . . . .	30
3.3	Napaka globine zaradi odstopanja naklona plovila. . . . .	30
3.4	Napaka pozicije pri odstopanju smeri za $0,5^\circ$ . . . . .	30
3.5	Rotacije okoli osi plovila. . . . .	31
3.6	Merilne naprave na plovilu Lyra (KP-5058). . . . .	32
3.7	Batimetrični model slovenskega morja. . . . .	33
4.1	Letalska fotografija območja med Izolo in Koprrom. . . . .	36
4.2	Primer kartiranja morskih travnikov iz letalskih posnetkov. . . . .	39
4.3	Primer razporeditve linijskih presekov za določanje tipov morskega dna. . . . .	39
4.4	Transekt video snemanja iz plovila. . . . .	41
5.1	Stara maona, potopljena v Piranskem zalivu. . . . .	46
5.2	Potopljen rimski pomol v Simonovem zalivu, Izola. . . . .	46
5.3	Primer poteka georeferenciranja podatkov o globinah. . . . .	48
5.4	Prikaz orodja za ročno preverjanje kakovosti sonarskih podatkov. . . . .	51
5.5	Primerjava rezultatov ročnega in avtomatskega čiščenja podatkov. . . . .	52
5.6	Prikaz surovih podatkov mnogospnega sonarja in vpliva nastavitvev moči oddajnika na podatke. . . . .	54
5.7	Prikaz vpliva obdelave podatkov na sliko akustičnih odbojev. . . . .	55
5.8	Anomalije v podatkih zaradi kotnih odvisnosti akustičnih odbojev prikazane v 3R. . . . .	58
5.9	Primerjava različnih pristopov uravnavanja kotnih odvisnosti. . . . .	60
5.10	Podatki z odpravljenimi kotnimi odvisnostmi. . . . .	61
6.1	Območje izbrano za ovrednotenje metod strojnega učenja. . . . .	64
6.2	Fotografije različnih tipov morskega dna. . . . .	64
6.3	Prikaz tipov morskega dna na obravnavanem območju. . . . .	65
6.4	Prikaz batimetrij in slike akustičnih odbojev. . . . .	67

6.5	Histogram verjetnostnih porazdelitev globin znotraj celice za različne tipe morskega dna. . . . .	69
6.6	Histogram verjetnostnih porazdelitev intenzitet znotraj celice za različne tipe morskega dna. . . . .	69
6.7	Območje zajema podatkov. . . . .	74
6.8	Informacijski prispevek značilk. . . . .	79
6.9	Razmerje informacijskega prispevka za značilke izpeljane iz intenzitet odbojev. Primerjava različnih velikosti celic. . . . .	82
6.10	Razmerje informacijskega prispevka za značilke izpeljane iz globin. Primerjava različnih velikosti celic. . . . .	82
6.11	Razmerje informacijskega prispevka značilk. . . . .	85
6.12	Ocene značilk po metodi ReliefF. . . . .	85
6.13	Ocene značilk po metodi naključnih gozdov. . . . .	86
7.1	Prostorski prikaz značilk, pripravljenih za obdelavo z algoritmi strojnega učenja. . . . .	97
7.2	Graf medsebojne odvisnosti za značilki <i>i</i> in <i>z</i> ter tip morskega dna. . . . .	99
7.3	Graf medsebojne odvisnosti za značilki <i>z_korelacija</i> in <i>z_kontrast</i> ter tip morskega dna. . . . .	99
7.4	Graf medsebojne odvisnosti za značilki <i>z_std</i> in <i>z_homogenost</i> ter tip morskega dna. . . . .	100
7.5	Odločitveno drevo algoritma C4.5. . . . .	103
7.6	Prikaz napak algoritma naključnih gozdov. . . . .	104
7.7	Graf klasiifikacijskih točnosti pri uporabi različnih velikosti oken za interpolacijo. . . . .	106
7.8	Prikaz napak algoritma naključnih gozdov po uporabi medianinega filtra z velikostjo okna 3 metre. . . . .	107
7.9	Prikaz napak algoritma naključnih gozdov po uporabi medianinega filtra z velikostjo okna 6 metrov. . . . .	107
8.1	Diagram postopka nenadzorovane klasifikacije. . . . .	111
8.2	Karta tipov morskega dna. . . . .	113
8.3	Karta tipov morskega dna. . . . .	114
8.4	Karta tipov morskega dna. . . . .	115
8.5	Primerjava kart tipov morskega dna za območje med Izolo in Koprom. . . . .	117
8.6	Karta tipov morskega dna, prirejena po Kolega [74]. . . . .	118
8.7	Karta tipov morskega dna, prirejena po Kolega [74]. . . . .	119
8.8	Karta tipov morskega dna, prirejena po Kolega [74]. . . . .	120
8.9	Prikaz tipov morskega dna in točk vzorčenja v 3R GIS orodju. . . . .	122
8.10	Video posnetek s prikazom točne lokacije snemanja. . . . .	122

# Tabele

3.1	Tehnični podatki sonarja Reson Seabat 8125. . . . .	29
6.1	Seznam značilnk. . . . .	73
6.2	Primerjava klasifikacijskih točnosti na podatkih obdelanih z različnimi metodami uravnavanja kotnih odvisnosti. . . . .	77
6.3	Prikaz statistično značilnih razlik v klasifikacijski točnosti algoritma C4.5 na različno predobdelanih podatkih. . . . .	78
6.4	Vsota rangov značilnk po ocenah s tremi merami za ovrednostenje značilnk. . . . .	84
7.1	Primer tabele napačnih klasifikacij. . . . .	94
7.2	Klasifikacijska točnost algoritmov strojnega učenja na akustičnih podatkih. . . . .	101
7.3	Prikaz statistično značilnih razlik v klasifikacijski točnosti algoritmov pri $\alpha = 0.001$ . . . . .	101
7.4	Tabela napačnih klasifikacij algoritma naključnih gozdov. . . . .	104



# Izvleček

## *AKUSTIČNA KLASIFIKACIJA MORSKEGA DNA*

V disertaciji obravnavamo akustično klasifikacijo morskega dna. Razvili smo praktično uporabno metodo akustične klasifikacije morskega dna. V disertaciji podamo opis celotnega postopka, od zajema surovih podatkov, do karte tipov morskega dna.

Predstavimo uporabljeno opremo za izvajanje hidrografskih meritev in postopek zajema akustični podatkov o morskem dnu. Podrobneje opišemo metodo določanja položaja plovila in zajetih podatkov. Opišemo celoten postopek predobdelave akustičnih podatkov, od umeščanja podatkov v prostor in preverjanja kakovosti podatkov, do izločanja vplivov nastavitve sistema in uravnavanja kotnih odvisnosti. Predlagamo metodo za preverjanje kakovosti in čiščenje podatkov ter dve novi metodi uravnavanja kotnih odvisnosti. Različne pristope k uravnavanju vpliva kotnih odvisnosti ovrednotimo. Najboljše rezultate dobimo z uporabo pristopa `L_std` (lokalno pristop s standardnim odklonom) in našega predlaganega pristopa `G_L_std` (globalni in nato lokalni pristop s standardnim odklonom). Med ocenama pristopov `L_std` in `G_L_std` ni statistično značilnih razlik. Odločimo se za uporabo pristopa `L_std`, zaradi manjše časovne zahtevnosti.

Za izračun značilke podatke razdelimo na kvadratne celice s stranico dolžine 1 meter, 2 metra in 4 metre. Definiramo 20 značilke na treh velikostih celic, torej skupno 60 značilke. Na podlagi mer za ocenjevanje značilke posameznim značilkam določimo najustreznejšo velikost celice in nato določimo najboljše značilke za problem akustične klasifikacije morskega dna. Najbolj informativni značilki sta intenziteta odboja (za celico 2x2 metra) in povprečna globina (za celico 1x1 meter). Uporabimo še standardne odklone globin in intenzitet odbojev ter homogenost, kontrast, različnost in korelacijo podatkov o globinah.

Za ovrednotenje algoritmov strojnega učenja izberemo 200 metrov širok in 900 metrov dolg pas na območju med Izolo in Koprom. Za izbrano območje generiramo množico označenih podatkov. Z metodo 10-kratnega prečnega preverjanja ovrednotimo algoritme strojnega učenja in z enosmernim parnim T-testom analiziramo statistično značilnost razlik uspešnosti posameznih algoritmov. Najboljše rezultate dobimo z algoritmom odločitvenih dreves (97.64 % klasifikacijska točnost) in algoritmom naključnih gozdov (98.47 % klasifikacijska točnost), pri čemer so naključni gozdovi statistično značilno boljši od odločitvenih dreves pri parametru  $\alpha = 0.001$ . Na rezultatih klasifikacije uporabimo medianin filter, ki z upoštevanjem prostorske razporeditve podatkov še dodatno izboljša rezultate. Za potrebe akustične klasifikacije morskega dna priporočamo uporabo algoritma naključnih gozdov. V koli-

kor je potrebna interpretacija dobljenega modela, priporočamo uporabo algoritma odločitvenih dreves, saj ima slednji veliko bolj izrazen model. Izraznost dobljenega modela v nekaterih primerih odtehta nekoliko slabšo klasifikacijsko točnost.

Razvito metodologijo apliciramo na problem kartiranja morskih travnikov v slovenskem morju. Učinkovito kartiranje morskih travnikov je zelo aktualen problem, saj direktiva Evropske Unije določa ciklično spremljanje razširjenosti morskih travnikov. Podatke hidrografskih meritev obdelamo po razviti metodologiji in izdelamo karto tipov morskega dna za celotno slovensko obalo. Določimo 36 večjih morskih travnikov in prisotnost vseh potrdimo z metodo točkovnega vzorčenja. Rezultate predstavimo v 3-razsežnostnem geografskem informacijskem sistemu. Z uspešno aplikacijo v realni domeni potrdimo praktično uporabnost naše metodologije in pokažemo, da ima akustična klasifikacija morskega dna s podatki mnogospopnega sonarja velik potencial pri raziskavah morskega dna.

**Ključne besede:** Klasifikacija morskega dna, akustični podatki, podatkovno rudarjenje, strojno učenje, hidrografske meritve, mnogospopni sonar.

# Abstract

## *ACOUSTIC SEABED CLASSIFICATION*

This thesis deals with acoustic seabed classification. We have developed a practical method for acoustic seabed classification. In the thesis, we give a description of the whole process, from data acquisition to mapping seabed types.

We present the equipment used to carry out hydrographic measurements and the process of collecting acoustic data. We describe the method of determining the exact position of the vessel and captured data. We describe the whole process of pre-processing of acoustic data, from data verification and georeferencing, to elimination of the effects of system settings and angular dependence. We propose a method for data quality verification and elimination of low quality data. We also propose two new methods for elimination of angular dependencies in data. Different approaches to elimination of angular dependencies are evaluated. The best results are obtained with the use of the `L_std` (local approach with a standard deviation) and our proposed approach `G_L_std` (global and then local approach with a standard deviation). There are no statistically significant differences between approaches `L_std` and `G_L_std`. We decide to use the `L_std` approach, due to lower time complexity.

To calculate features data is divided into square cells with a side length of 1 meter, 2 meters and 4 meters. We define 20 features on three cell sizes, i.e. a total of 60 features. We evaluate each feature on different cell sizes and determine the most appropriate cell size. Then we calculate all features on corresponding cell sizes, evaluate them and determine the best features for acoustic seabed classification. The most informative feature is the backscatter intensity (for cell size 2x2 meters) and average depth (for cell size 1x1 meter). The depth standard deviations, intensity standard deviation, homogeneity, contrast, dissimilarity and correlation of depth data are also used.

For the evaluation of machine learning algorithms we use 200 meters wide and 900 meters long coastal zone in the area between Izola and Koper. For the selected area we have collected acoustic data and ground-truth data of seabed types. We've then used ground-truth data for evaluation of machine learning algorithms. We evaluate algorithms using 10-fold cross-validation and test statistical significance of obtained results with paired T-test. The best results are obtained by decision tree algorithm (97.64 % classification accuracy) and random forests algorithm (98.47 % classification accuracy), whereby the random forests are significantly better than decision trees for significance  $\alpha = 0.001$ . We post-process results using median fil-

ter, taking into account the spatial distribution of data, to further improve results. For the purposes of acoustic seabed classification we recommend using the random forests algorithm. If interpretation of the resulting model is important, we recommend a decision tree algorithm, which has a much more expressive model. The expressiveness of the resulting model in some cases outweighs the somewhat lower classification accuracy.

Developed methodology is applied to the problem of mapping of marine meadows in the Slovenian sea. Effective mapping of marine meadows is a very topical issue, since the directive of the European Union provides for cyclically monitoring of the prevalence of marine meadows. We processed hydrographic data using our developed methodology and mapped seabed types for the entire Slovenian coast. We found 36 large marine meadows and confirmed the presence of every marine meadow by underwater video method. The results are presented in 3D geographic information system. The successful application in the real domain confirms the practical applicability of our methodology and show that the acoustic classification of the seabed with multibeam sonar acoustic data has great potential in seabed research.

**Key words:** Seabed classification, acoustic data, data mining, machine learning, hydrographic survey, multibeam sonar.



# Seznam kratic

## **3R**

3-razsežnost

## **CSV**

(ang. Comma Separated Values); Datoteka, v kateri so vrednosti ločene z vejico

## **DMR**

Digitalni model reliefa

## **GDAL**

(ang. Geospatial Data Abstraction Library); Knjižnica za obdelavo geoprostorskih podatkov

## **GIS**

Geografski informacijski sistem

## **GLCM**

(ang. Grey-Level Cooccurrence Matrix); Matrika sopojsnosti sivin

## **GNSS**

(ang. Global Navigation Satellite System ); Globalni navigacijski satelitski sistem

## **GPS**

(and. Global Positioning System); Globalni sistem pozicioniranja

## **IHO**

(ang. International Hydrographic Organization); Mednarodna hidrografska organizacija

## **IMU**

(ang. Inertial Measurement Unit); Inercialni senzor premikov

## **INS**

(ang. Inertial Navigation System); Inercialni navigacijski sistem

## **KDD**

(ang. Knowledge Discovery in Databases); Odkrivanje zakonitosti v podatkih

**kNN**

(ang. k-Nearest Neighbours); Metoda k najbližjih sosedov

**LiDAR**

(ang. Laser illuminated Detection And Ranging); Laserska detekcija in rangiranje. Sistem za merjenje oddaljenosti z laserskimi žarki

**MBES**

(ang. Multibeam Echo Sounder); Mnogospopni sonar

 **$\mu$ Pa**

Mikro pascal

**NB**

(ang. Naïve Bayes); Naivni Bayes

**RADAR**

(ang. RAdio Detection And Ranging); Radijska detekcija in rangiranje. Sistem za merjenje oddaljenosti z elektromagnetnimi valovi

**RF**

(ang. Random Forest); Metoda naključnih gozdov

**RTK**

(ang. Real Time Kinematic); Metoda, ki z meritvijo faze nosilnega vala signala izračuna popravke GNSS pozicije in s tem omogoči centimetrsko natančnost pozicioniranja

**SB-AGDS**

(ang. Single-Beam Acoustic Ground Discrimination Systems); Enospopni akustični diskriminatorni sistem

**SBES**

(ang. Singlebeam echo sounder); Enospopni sonar

**SONAR**

(ang. SOund Navigation And Ranging); Tehnika, ki uporablja zvok za navigacijo, določanje položaja in komunikacijo

**SSS**

(ang. Side Scan Sonar); Bočni sonar

**SVM**

(ang. Support Vector Machine); Metoda podpornih vektorjev

**SVP**

(ang. Soung Velocity Profiler); Sonda za merjenje hitrosti zvoka v vodi

**TVG**

(ang. Time Varied Gain); Časovno spremenljiva občutljivost sprejemnika

**USB**

(ang. Universal Serial Bus); Univerzalno serijsko vodilo

**UTM**

(ang. Universal Transverse Mercator); Univerzalna prečna Mercatorjeva projekcija

**XTF**

(ang. eXtended Triton Format); Format podatkov za shranjevanje hidrografskih meritev



# Poglavje 1

## Uvod

V tem delu bomo predstavili postopek akustične klasifikacije morskega dna s podatki mnogospnega sonarja. Delo se uvršča v področji inteligentnih sistemov (2.07.07) in oceanografije (1.02.04). Klasifikacijo morskega dna lahko izvajamo vizualno, mehanično ali akustično. Vse vizualne metode (potapljači, video snemanje, fotografiranje) in mehanske metode (zajemanje vzorcev, pregledovanje) so počasne in zahtevajo veliko truda, posledično so tudi drage in neprimerne za pregledovanje velikih območij morskega dna. Alternativa so akustične metode, s katerimi lahko hitro obdelamo velika področja, saj lahko podatke zajemamo kar iz premikajočega se plovila, s pomočjo sonarjev. Meritve smo izvedli z mnogospnim sonarjem. Mnogospni sonar je aktiven senzor, ki izkorišča fizikalne lastnosti širjenja ultrazvočnih signalov v vodi za merjenje razdalje do morskega dna. Najpogosteje je nameščen na plovilu. Mnogospni sonar uporablja usmerjene snope s katerimi meritev ne izvaja le navpično, pač pa pod različnimi koti od  $-60^\circ$  do  $+60^\circ$  glede na normalo plovila. Mnogospni sonarji so se začeli uporabljati že v sedemdesetih letih prejšnjega stoletja, sprva v vojaške namene. Prvi komercialni sonar je bil izdelan leta 1977. V kombinaciji z GNSS (globalni navigacijski satelitski sistem) in INS (inercialni navigacijski sistemi) opazovanji je možno meritve sonarja zelo natančno umestiti v prostor. Meritve sodobnih sonarjev so dobra osnova za izdelavo podrobnejših navtičnih zemljevidov [68] in digitalnih globinskih modelov (batimetriji), hkrati pa so pomemben vir informacij za arheološke, naravovarstvene, oceanografske in geološke raziskave [46], [47].

V zadnjem desetletju smo bili priča velikemu napredku pri tehnologijah za kartiranje kopnega. Analize letalskih in satelitskih posnetkov omogočajo natančno kartiranje različnih habitatov, raziskave krčenja gozdov [81], [75] itd. Razvoj zračnih in satelitskih sistemov za daljinsko zaznavanje je v zadnjih nekaj desetletjih povečal dostopnost optičnih in laserskih daljinsko-zaznanih podatkov za ekološke raziskave širših območij. To je bistveno pripomoglo k boljšemu razumevanju prostorskih vzorcev kopnega [96], [44]. Oceani prekrivajo 71% površja zemlje in o njih vemo bistveno manj kot o kopnem. Podatke naj bi imeli le o 5% morskega dna, pa še ti se po kakovosti in ločljivosti ne morejo primerjati s satelitskimi podatki o kopnem [17]. Preučevanje morskega dna s pomočjo geofizikalnih metod (sonar - Sound Navigation And Ranging, LiDAR - Laser Illuminated Detection And Ranging) je v tujini že uveljavljena praksa, saj se batimetrični podatki zaradi svoje natančnosti uporabljajo tako pri povsem praktičnih (npr. vzpostavitev plinovodov, polaganje optičnih

kablov, vrtnje naftnih vrtin, izdelava navtičnih kart ...) kot tudi znanstveno-raziskovalnih dejavnostih (npr. podvodna arheologija, morska biologija, podvodna geologija ...). Sonarji predstavljajo ekvivalent danes v prostorskih študijah že skoraj nepogrešljivi metodi LiDAR na kopnem, pri obeh pa gre za daljinsko zaznavanje, ki predstavlja znanost pridobivanja informacij o površju brez da bi z njim prišli v neposredni stik in pri katerem zaznavamo odbito ali sevano svetlobo (LiDAR) ali zvok (sonar) in ju nato ustrezno obdelujemo. Iz digitalnega modela reliefa (DMR) morskega dna, ki ga pridobimo iz podatkov mnogospopnega sonarja lahko, podobno kot na kopnem iz podatkov LiDAR-ja, identificiramo posamezne geomorfološke elemente ter sklepamo na tektonske in sedimentacijske procese, ki so jih (so)oblikovali [123].

Glavni cilj našega dela je razvoj praktično uporabne metodologije za akustično klasifikacijo morskega dna. Večina obstoječih akustičnih metod temelji na interpretaciji digitalnega modela globin, sonarji pa beležijo tudi druge zanimive podatke, ki ostajajo neizkoriščeni. Sonarji delujejo tako, da merijo čas potovanja zvoka do morskega dna in nazaj. Na podlagi tega časa se, z upoštevanjem dodatnih informacij o trenutni temperaturi v vodnem stolpcu, izračuna globina morja. Stranski produkt meritev so intenzitete odbojev (ang. backscatter). Intenzitete odbojev so na prvi pogled zelo šumni in neuporabni podatki, vendar ustrezna predobdelava razkrije, da nosijo veliko informacije. Zvok se namreč od različnih podlag različno odbija. Groba podlaga povzroči večjo razpršitev zvoka, trda podlaga da močen odboj, mehka vpije zvočni snop. Moč odboja nam lahko torej marsikaj pove o sestavi morskega dna. Naša hipoteza je, da lahko na podlagi akustičnih podatkov učinkovito določimo tipe morskega dna.

## 1.1 Metodologija

Temeljne raziskave na področju pridobivanja informacij iz meritev mnogospopnega sonarja in možnosti uporabe teh informacij za določanje značilnosti morskega dna bodo omogočile razvoj novih visokotehnoških in inovativnih storitev. Raziskave bodo pokazale kakšne informacije se poleg informacije o globini še skrivajo v podatkih, pridobljenih z meritvami mnogospopnega sonarja. Raziskali bomo nove modele in metode za klasifikacijo morskega dna. Na podlagi izsledkov raziskav se bo nato lahko načrtovalo razvoj novih storitev za kartiranje morskega dna, oceanografske raziskave, raziskave habitatov na morskem dnu, ipd. Za doseg zadanih ciljev bo potreben razvoj sistema za obdelavo velikih količin podatkov, pridobljenih z meritvami mnogospopnega sonarja ter sistema za predobdelavo podatkov (izločanje slabih meritev, kompenzacija nastavitvev, normiranje podatkov ipd.). Prav tako bo potrebno preučevanje ustreznih statističnih metod in metod strojnega učenja za iskanje značilnk ter analiza, adaptacija in po potrebi razvoj novih metod strojnega učenja primernih za klasifikacijo morskega dna na osnovi zbranih značilnk.

### 1.1.1 Načrtovanje in izvedba meritev

Za uspešno izvedbo programa raziskovalnega dela bo potrebno zagotoviti zadostno količino ustreznih podatkov. V podjetju Harpha Sea, d.o.o. Koper imamo na razpo-

lago plovila, sonarje in druge merilne sonde ter v lastnih raziskovalnih skupinah tudi ustrezno usposobljene strokovnjake za izvajanje meritev. Meritve vodi hidrograf z mednarodno licenco, kar omogoča, da so meritve priznane s strani mednarodne hidrografske organizacije IHO [62]. V podjetju izdelujemo merilne sisteme za zajem podatkov na terenu ter za prenos podatkov v centralo, izvajamo meritve na terenu in izdelujemo zahtevne obdelave in analize zajetih podatkov. Z naprednimi globalnimi navigacijskimi satelitskimi sistemi (GNSS) in inercialnimi navigacijskimi sistemi (INS) v kombinaciji s sodobnim mnogosnopnim sonarjem bomo poskrbeli za zajem kakovostnih podatkov s centimetrsko natančnostjo. Meritve na morju bomo izvajali s plovilom Lyra dolžine 8 metrov na katerem je stalno vgrajen mnogosnopni sonar (MBES) SeaBat 8125, podpovršinski sonar SES2000-compact ter enosnopni sonar (SBES) HydroStar. Vsi sonarji so integrirani z GNSS sprejemnikom Javad Duo-G2D ter INS sistemom TSS Mahrs. Po potrebi bomo na plovilo integrirali tudi druge sonarje (bočne, slikovne) ali magnetometer.

Skupaj s strokovnjaki s področja hidrografije bomo sprva načrtovali in izvedli meritve za potrebe akustične klasifikacije morskega dna. Podatke bomo nato prečistili, analizirali in klasificirali. Po potrebi bomo meritve ponovili z optimiziranimi nastavitvami. Najprej se bomo ukvarjali s podatki, ki so posneti le za namen klasifikacije. Snemali bomo z različnimi nastavitvami, da raziščemo vplive nastavitvev na podatke. Razumevanje vplivov nastavitvev nam bo omogočilo uporabo enakih metod na starih podatkih – podatkih ki so bili posneti za izdelavo batimetričnih modelov. Slednjih imamo veliko. Na voljo imamo namreč podatke za celotno slovensko morje, posnete med letoma 2006 in 2008.

### 1.1.2 Preverjanje kakovosti in čiščenje podatkov

Med meritvami občasno prihaja do motenj, prav tako pa vsi podatki niso enako zanesljivi. Najbolj vprašljivi so navadno podatki na robu področja meritev posamezne linije, kjer so vpadni koti akustičnih signalov največji. Med načrtovanjem meritev bomo to ustrezno upoštevali in poskrbeli za prekrivanje podatkov. Prekrivanje je uporabno zaradi doseganja višje kakovosti izmerjenih podatkov in tudi zaradi preverjanja meritev. Med postopkom čiščenja bomo nezanesljive podatke na robu območja izločili. Izločili bomo tudi podatke kjer je pričakovana napaka velika in kjer je stanje odboja slabo. Trenutno čiščenje izvajamo ročno, v orodju PDS2000 [126], s pomočjo izurjenih hidrografov. Načrtujemo, da bomo v postopku razvoja sistema za klasifikacijo morskega dna razvili tudi modul za avtomatsko preverjanje kakovosti in čiščenje podatkov.

### 1.1.3 Predobdelava podatkov

Podatki o občutljivosti sprejemnika, moči oddanega pulza, količniku predvidene absorpcije zvoka v vodi in predvidenih izgub moči signala so pomembni, ker odražajo spremembe nastavitvev, ki jih opravlja hidrograf med meritvami. Za potrebe akustične klasifikacije morskega dna bi bilo bolje, da bi bila večina teh nastavitvev fiksna in bi torej lastnosti odbojev direktno odražale tip tal. Večina MBES meritev se izvaja z namenom merjenja globin, kjer pa se nastavitve prilagajajo, da se doseže

ustrezna natančnost meritev. Za potrebe akustične klasifikacije moramo take podatke ustrezno predobdelati in izločiti vpliv nastavitve sistema nanje. Z ustreznim načrtovanjem meritev bomo pridobili podatke o vplivu nastavitve na pridobljene podatke, vplive ustrezno analizirali in razvili ustrezne enačbe za izločanje vplivov nastavitve sistema na podatke.

Poleg zgoraj omenjenih nastavitve so intenzitete odboja močno odvisne tudi od vpadnega kota signala. Izločanje kotnih odvisnosti odbojev je eden izmed glavnih izzivov akustične klasifikacije morskega dna. V podatkih se pogosto pojavljajo nepravilnosti zaradi neustreznega pristopa pri uravnavanju kotnih odvisnosti [29]. V literaturi najpogosteje naletimo na dva različna pristopa - teoretičnega in empiričnega. Pri bočnih sonarjih se v nekaterih primerih uporablja kompenzacija glede na vnaprej pripravljen model (teoretični pristop), vendar tak pristop pri MBES ne deluje zaradi bistveno višje frekvence. Izkazuje se, da je praktično nemogoče izdelati model za vse vrste tal in vse kote. Uporabili bomo empirični pristop. Na testnih podatkih bomo uporabili več različnih metod, ki se bodo razlikovale predvsem glede na upoštevanje/neupoštevanje standardnega odklona in lokalno oziroma globalno povprečenje podatkov. Predlagali bomo tudi dva nova pristopa, katerih cilj je združiti prednosti obstoječih pristopov. Vse pristope bomo objektivno ovrednotili z oceno klasifikacijske točnosti po metodi 10-kratnega prečnega preverjanja in izbrali najboljši pristop.

#### 1.1.4 Izračun in izbor značilnk

Pri izračunu značilnk gre pravzaprav za določanje deskriptorjev akustičnih podatkov, ki bi lahko bili uporabni za razlikovanje med tipi morskega dna. Potrebno je izluščiti čim več značilnk, ki nam bodo kasneje pomagale pri klasifikaciji – torej izluščiti informacije iz surovih podatkov [104]. Značilke so lahko osnovne meritve, kot je intenziteta odboja, ali pa bolj sofisticirani deskriptorji, kot so rezultati statističnih ali prostorskih analiz. Izračun ustreznih značilnk je temelj za uspešno klasifikacijo morskega dna. Pri izračunu značilnk bomo sodelovali s strokovnjaki s področja hidrografije. Preučili bomo tudi pristope za klasifikacijo morskega dna, ki temeljijo na vizualni obravnavi modelov globin ali modelov intenzitet odbojev.

Prednost sodobnih MBES sistemov je možnost hkratnega merjenja globin in zajema podatkov o odboju. V kombinaciji s sodobnimi GNSS sistemi lahko podatke zelo natančno umestimo v prostor. Iz analiz globin je mogoče pridobiti veliko geomorfoloških podatkov o morskem dnu, na primer razločiti peščine, skalne gmote, itd. Poleg tega globina vpliva na razporeditev habitatov, saj se z globino hitro zmanjšuje osvetljenost morskega dna in tudi izpostavljenost valovanju morja. Iz modelov globin bomo pridobili tudi veliko drugih značilnk, kot so standardni odklon globin, asimetrija, sploščenost itd. Preučili bomo tudi možnost uporabe teksturnih značilnk, katerih cilj je posnemanje principov človeškega vida [7], ki razlikuje med različnimi toni in teksturami. V literaturi je opisanih nekaj orodij za računanje značilnk iz batimetričnih modelov [118]. Čeprav je analiza batimetrij učinkovita pri določanju geomorfoloških značilnosti morskega dna pa sama po sebi ni uspešna pri določanju tipov morskega dna če med posameznimi tipi ni pomembnih razlik v topografiji. Zato bomo vse značilke izračunali tako za podatke o batimetriji kot za podatke o



intenzitetah odbojev. Prav tako bomo značilke izračunali za manjša območja kot tudi za večje dele morskega dna in tako pridobili širše znanje o razgibanosti terena.

Vse izračunane značilke bomo ovrednotili z ustreznimi merami za ocenjevanje značilk in izbrali najboljše značilke za problem akustične klasifikacije morskega dna.

### 1.1.5 Zbiranje podatkov o tipih morskega dna

Poleg zbiranja akustičnih podatkov s sonarjem bo potrebno podatke zbirati tudi vizualno. Pri izvedbi eksperimentov in meritev si bomo, v primeru potrebe po specifičnih znanjih za katera nimamo ustreznih strokovnjakov ali opreme v okviru lastne raziskovalne infrastrukture, pomagali z zunanjimi sodelavci, na primer potapljači. Po potrebi bomo vzpostavili sodelovanje z ustreznimi raziskovalnimi institucijami, ki se ukvarjajo z enako ali sorodno problematiko kot so Morska Biološka postaja Piran, Geološki zavod v Ljubljani, OGS Trieste in Biološka postaja Rovinj. Sodelovanje z drugimi raziskovalnimi institucijami bo pripomoglo k hitrejši in kakovostnejši izvedbi raziskovalnega dela, pridobivanju dodatnih znanj in bo hkrati omogočilo sodelovanje s strokovnjaki s področji, ki niso pokrita v okviru lastnih raziskovalnih skupin podjetja. Rezultate klasifikacije bomo preverjali na različne načine. Na nekaterih delih bo možno za preverjanje uporabiti letalske posnetke, ki jih bomo posneli istočasno z akustičnimi podatki. V ostalih primerih bomo uporabljali potapljače in podvodne kamere. Dobljene rezultate bomo primerjali tudi z rezultati kartiranja drugih raziskovalnih institucij (MBP, OGS) in rezultati kartiranja objavljenimi v delu [74].

### 1.1.6 Klasifikacija morskega dna

Uporabili bomo fenomenološki pristop. Fenomenološki pristop predvideva, da obstaja korelacija med morfološkimi in fizičnimi lastnostmi morskega dna ter akustičnimi podatki. Podatke bomo razdelili v akustične razrede brez direktnega sklepanja na dejanske fizične lastnosti dna. S pomočjo preverjanja (potapljači, zajem vzorcev, kamere) bomo nato akustične razrede povezali z dejanskimi tipi tal. Pri fenomenološkem pristopu se po izpeljavi in izboru značilk uporabi nadzorovane ali nenadzorovane klasifikacijske algoritme. Pri nenadzorovanih algoritmih se podatke razvrsti v skupine statistično podobnih enot. Nadzorovana klasifikacija poteka tako, da se algoritem najprej na poznanem delu območja nauči razločiti tipe tal in nato to znanje uporabi na preostalem območju. Na pripravljenih podatkih bomo testirali in ovrednotili različne metode strojnega učenja, kot so naivni Bayes, drevesa, metodo podpornih vektorjev itd. Uporabili bomo okolje za strojno učenje ORANGE [30] v kombinaciji s programskim jezikom Python in okolje WEKA [54] v kombinaciji s programskim jezikom Java.

### 1.1.7 Aplikacija v realni domeni in 3R vizualizacija rezultatov

Razvito metodologijo bomo preizkusili na realni domeni s področja oceanografije in sicer na problemu kartiranja morskih travnikov slovenskega morja. Rezultate bomo primerjali z rezultati objavljenimi v delu rezultati kartiranja objavljenimi v delu [74].

V raziskovalni skupini imamo veliko izkušenj z izdelovanjem zahtevnih GIS sistemov in 3R vizualizacijo prostorskih podatkov [124], [136], [65]. Med delom bomo uporabljali prikazovalnik orodja PDS2000. Dobljene tipe morskega dna bomo vnesli in prikazali v 3-razsežnostnem geografskem informacijskem sistemu (3R GIS). Vizualizacijo bomo razvili v okolju OpenSceneGraph [20].

### 1.1.8 Povratna zanka

Celoten postopek ali del le-tega bomo po potrebi večkrat ponovili ter tako primerjali različne skupine značilk, različne metode in različne postopke predobdelave podatkov. Povratno zanko bomo uporabili tudi za optimizacijo izbora značilk in dodelavo razvitih metod. V povratni zanki bomo najprej izbrali ustrezne značilke, le te uporabili pri izgradnji modela, katerega bomo na koncu testirali na označenih podatkih. Glede na rezultate testiranja bomo lahko ponovno prilagajali izbor značilk ter parametre modela.

## 1.2 Prispevki k znanosti

1. **Razvoj sistema in metodologije za avtomatsko predobdelavo velikih količin podatkov, pridobljenih z meritvami mnogospnega sonarja**  
Razvili in opisali smo metodologijo za avtomatsko predobdelavo velikih količin podatkov, pridobljenih z meritvami mnogospnega sonarja. Opisana metodologija vsebuje 6 korakov:

- zajem in formatiranje surovih podatkov
- preverjanje kakovosti in izločanje slabih podatkov
- izločanje vpliva nastavitve moči oddajnika in občutljivosti sprejemnika
- izločanje vpliva absorpcije in sferičnega širjenja zvoka
- izločanje vpliva velikosti akustičnega odtisa
- uravnavanje vpliva kotnih odvisnosti

Formatiranje surovih podatkov služi predvsem razširljivosti metodologije na druge MBES sisteme. Avtomatsko preverjanje kakovosti in izločanje slabih podatkov, je korak, ki bistveno pohitri klasifikacijo morskega dna. V primerjavi z dosedanjim ročnim preverjanjem podatkov je bistveno hitrejši, še pomembneje pa je da se izognemo zamudnemu ročnemu delu in potrebi po dragi programski opremi. Koraki 3, 4 in 5 temeljijo na teoretičnih osnovah s področja podvodne akustike. Dodali smo interpolacijo prehoda pri spremembi nastavitve med meritvami, ko pride do zamika med časom spremembe določene nastavitve in dejanskim časom prilagoditve sistema. Pri izločanju vpliva velikosti akustičnega odtisa smo teoretične osnove le deloma upoštevali, saj je velikost akustičnega odtisa močno odvisna od vpadnega kota akustičnega signala, slednjega pa ni možno dovolj natančno izračunati. Tako smo del izločanja vpliva velikosti akustičnega odtisa združili z uravnavanjem vpliva kotnih odvisnosti in ga izvedli empirično. Uravnavanje vpliva kotnih odvisnosti temelji na empirično

izpeljanih enačbah, ki smo jih tudi ustrezno ovrednotili v poglavju 6.3. Za ovrednotenje smo uporabili klasifikacijsko točnost algoritma C4.5 na različno predobdelanih množicah enakih podatkov. Klasifikacijsko točnost smo ocenili po metodi 10 zagonov 10-kratnega prečnega preverjanja, značilnost razlik pa potrdili s popravljenim enosmernim parnim T-testom.

Vse korake smo združili v enoten postopek predobdelave podatkov ki je popolnoma avtomatiziran in ne potrebuje nobenih nastavitvev operaterja. Razvita metodologija je opisana v poglavju 5 in objavljena v [93].

## 2. Določitev ustreznih značilk, primernih za klasifikacijo morskega dna iz akustičnih podatkov

Definirali smo izračun 20 značilk na 3 velikostih celic, torej skupno 60 značilk za vsako celico. Uporabili smo osnovne značilke, ki se pogosto uporabljajo pri interpretaciji akustičnih podatkov ter na naš problem aplicirali nekatere značilke, ki se uspešno uporabljajo na področju analize slik [11]. Za ocenjevanje značilk smo uporabili razmerje informacijskega prispevka, ocenjevanje značilk z naključnimi gozdovi [15] in algoritem ReliefF [70]. Posameznim značilkam smo najprej določili najustreznejšo velikost celice za izračun značilke, nato pa značilke primerjali med seboj in določili najboljše značilke za problem akustične klasifikacije morskega dna. Postopek ovrednotenja značilk je opisan v poglavju 6 in objavljen v [93].

## 3. Generiranje učne množice podatkov ter analiza in ovrednotenje metod strojnega učenja na problemu klasifikacije morskega dna

Opisali in izbrali smo metode strojnega učenja, ki so ustrezne za akustično klasifikacijo morskega dna. Izvedli smo hidrografske meritve za pridobitev akustičnih podatkov o morskem dnu po postopku opisanem v poglavju 3 in zbrali podatke o tipih morskega dna z metodami opisanimi v poglavju 4. Iz zbranih podatkov smo generirali učno množico podatkov za uporabo v orodjih za strojno učenje. Z metodo prečnega preverjanja smo ovrednotili algoritme strojnega učenja in z enosmernim parnim T-testom analizirali statistično značilnost razlik uspešnosti posameznih algoritmov. Rezultate najuspešnejših algoritmov smo podrobneje analizirali in opisali. Izkoristili smo umeščenost podatkov v prostor in na rezultatih klasifikacije uporabili medianin filter ter s tem še dodatno izboljšali uspešnost klasifikacije.

Analiza in ovrednotenje metod strojnega učenja na problemu klasifikacije morskega dna so opisani v poglavju 7 in v delu [111].

## 4. Aplikacija razvite metodologije v realni domeni, t.j. oceanografiji, ovrednotenje na realnih podatkih in 3R predstavitev rezultatov

Razvito metodologijo smo aplicirali na problem kartiranja morskih travnikov v slovenskem morju. Učinkovito kartiranje morskih travnikov je zelo aktualen problem, saj direktiva Evropske Unije določa ciklično spremljanje razširjenosti morskih travnikov [40]. Izvedli smo meritve celotne obale slovenskega morja, do globine 10 metrov. Podatke smo obdelali po razviti metodologiji in izdelali karto tipov morskega dna za celotno slovensko obalo. Določili smo 36 večjih

morskih travnikov in prisotnost vseh potrdili z metodo točkovnega vzorčenja. Rezultate smo predstavili v 3R GIS okolju. Aplikacija razvite metodologije v realni domeni, t.j. oceanografiji, ovrednotenje na realnih podatkih in 3R predstavitev rezultatov so opisani v poglavju 8 in delih [111] in [93].

## 1.3 Pregled vsebine

Delo je organizirano v 9 poglavjih. Uvodnemu poglavju sledi poglavje 2, kjer so predstavljena osnovna znanja, potrebna za razumevanje tega dela. Predstavljene so osnovne značilnosti slovenskega morja, različni akustični sistemi, definirana je podvodna akustika in pojem odkrivanja zakonitosti v podatkih. V zadnjem delu poglavja so predstavljene različne metode klasifikacije morskega dna ter njihove prednosti oziroma slabosti.

V poglavju 3 je predstavljena uporabljena oprema za izvajanje hidrografskih meritev in postopek zajema akustični podatkov o morskem dnu. Podrobneje je opisana metoda določanja položaja plovila in zajetih podatkov, saj je natančen položaj plovila ključnega pomena za kakovostne podatke.

V poglavje 4 so opisane metode zbiranja podatkov o tipih morskega dna. Podrobneje so predstavljene tri metode, ki smo jih uporabljali za preverjanje rezultatov in tudi za generiranje množice podatkov za ovrednotenje metod strojnega učenja. V prvem delu je predstavljena metoda točkovnega vzorčenja s pomočjo potapljačev [90], v drugem delu uporaba letalskih posnetkov za določanje tipov morskega dna [92] in v tretjem delu metoda video snemanja morskega dna [110].

Poglavje 5 opisuje postopek obdelave akustični podatkov: izločanje slabih meritev ter izločanje vplivov nastavitve sistema na podatke. Predobdelava akustičnih podatkov je najpomembnejši del akustične klasifikacije morskega dna. Opisan je celoten postopek predobdelave podatkov, od umeščanja podatkov v prostor in preverjanja kakovosti podatkov, do izločanja vplivov nastavitve sistema in uravnavanja kotnih odvisnosti. Predstavljena je metoda za preverjanje kakovosti in čiščenje podatkov ter predlagani dve novi metodi uravnavanja kotnih odvisnosti.

V prvem delu poglavja 6 je definirano območje v slovenskem morju, kjer smo zbrali tako akustične podatke o morskem dnu, po postopku opisanem v poglavju 3, kot podatke o tipih morskega, po postopkih opisanih v poglavju 4. V drugem delu poglavja so predstavljene značilke, izračunane iz akustičnih podatkov o morskem dnu. V nadaljevanju je objektivno ovrednotenih šest pristopov k uravnavanju kotnih odvisnosti, opisanih v poglavju 5 in izbran najboljši pristop. V zadnjem delu poglavja so najprej izbrane najprimernejše velikosti celic za izračun posameznih značilk, nato pa ovrednotene in izbrane najbolj informativne značilke.

V poglavju 7 je opisan postopek klasifikacije morskega dna. Predstavljene so uporabljeni algoritmi za klasifikacijo in metode ovrednotenja klasifikatorjev. Na zbranih realnih podatkih so ovrednoteni algoritmi za klasifikacijo in podrobno predstavljene rezultati najbolje ocenjenih algoritmov. Predstavljena je uporaba medianinega filtra na prostorskih podatkih, ki še dodatno izboljša rezultate najboljših algoritmov.

Poglavje 8 predstavi uporabo razvite metodologije za akustično klasifikacijo morskega dna v realnih domeni. Metodologijo smo aplicirali na domeno kartiranja

morskih travnikov v slovenskem morju. Predstavljena je primerjava z obstoječimi metodami kartiranja in primerjava dobljene karte s karto pridobljeno po stari metodologiji. V zadnjem delu poglavja je opisan postopek prikaza rezultatov v 3 razsežnostnem GIS okolju.

V poglavju 9 komentiramo dobljene rezultate in predstavimo smernice nadaljnjih raziskav.



## Poglavje 2

# Osnovna znanja

Poglavje je namenjeno predstavitvi osnovnega znanja, ki je potrebno za razumevanje tega dela. V prvem podpoglavju predstavimo slovensko morje, nato v nadaljevanju na kratko opišemo sonarje, predstavimo področje podvodne akustike in področje odkrivanja zakonitosti v podatkih ter predstavimo najbolj uporabljane tehnike klasifikacije morskega dna.

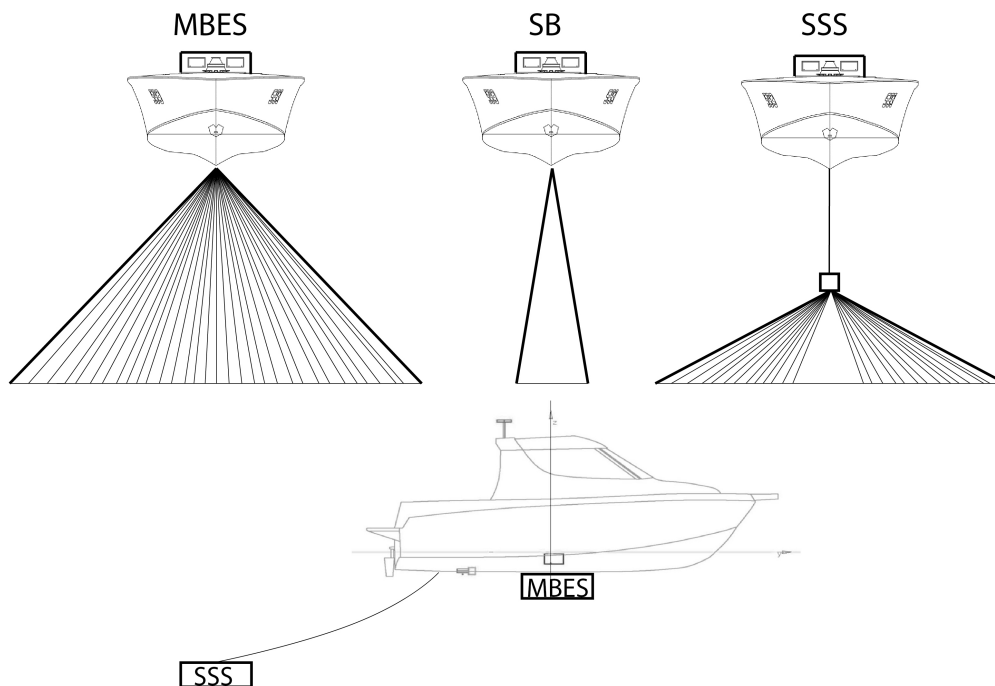
### 2.1 Slovensko morje

Dolžina morske obale, od zaliva Sv. Jernej na meji z Italijo do ustja reke Dragonje na meji s Hrvaško, je 46,6 km, medtem ko zračna razdalja znaša le 17 km. Najbolj se v obalo zajedata Koprski in Piranski zaliv, pomembnejša sta še manjša Strunjanski in Portoroški zaliv. Večina morske obale je pozidane oziroma spremenjene zaradi vplivov človeka, samo dobra petina je še naravne. Geodetski inštitut Slovenije in Harpha Sea d.o.o. sta v letih 2001 in 2002 za Ministrstvo za promet Republike Slovenije opravila detajlno izmero obale slovenskega morja s pomočjo tehnologije GPS. Izmerjena je bila zgornja temna linija (obalna linija aritmetrične sredine vseh visokih voda v letu), ki je tik nad mejo pojavljanja nekaterih tipičnih morskih organizmov oziroma rob zidane obale (pomoli). To nas kljub majhnosti oziroma kratkosti morske obale uvršča med redke države na svetu, ki imajo morsko obalo v celoti natančno premerjeno. K obali običajno štejemo še teritorialno morje Republike Slovenije, ki meri nekaj več kot 300 km<sup>2</sup>. Slovensko teritorialno morje spada v Tržaški zaliv. Povprečna globina Tržaškega zaliva je 16,4 metra [98], več kot petina zaliva je plitvejša od 10 m, 41 % globin se nahaja med 20 in 25 metri, največja globina zaliva je pri Piranu (37,5 m). Večina Tržaškega zaliva je prekrita z muljem, le v obalnem pasu se pojavljajo drugi tipi morskega dna. Kolega v delu [75] ocenjuje, da 77 % 500 metrskega obalnega pasu (morsko dno, do 500 metrov oddaljeno od obale) prekriva mulj, 13 % morski travniki, 2 % pesek, 2 % posamezne skale in 6 % kamenje. Torej mulj prekriva kar 77 % morskega dna v obalnem pasu. 500 metrski obalni pas predstavlja le cca. 10 % slovenskih teritorialnih voda, v preostanku slovenskih teritorialnih voda je delež muljastega dna še večji.

Zanimiva je predvsem obala, do cca. 10 metrov globine. Tam najdemo več kamnitega dna in morskih travnikov, ki nudijo zavetje ribam in ostalim morskim

organizmom. V tem delu se posvečamo predvsem metodologiji obdelave akustičnih podatkov o priobalnih delih morskega dna, kjer je raznolikost tipov morskega dna največja.

## 2.2 Splošni opis sonarja



Slika 2.1: Prikaz različnih tipov sonarjev: mnogospolni sonar (MBES), enospolni sonar (SB) in bočni sonar (SSS). Spodnji del slike prikazuje MBES sonar pritrjen na plovilo in SSS sonar, ki je pritrjen s kablom in ga plovilo vleče za seboj.

Sonar je zvočna različica radarja (RADAR - Radio Detection And Ranging), ki deluje tako, da pošilja pulze zvoka proti dnu, ti se od dna odbijejo in nam dajo natančne informacije o dnu. Tako kot je sonar zvočna različica radarja je LiDAR optična različica radarja. Gre za tri sorodne tehnike, ki se razlikujejo glede na to, kaj pošiljajo proti površju oziroma s pomočjo česa dobimo podatke (sliko) površja. Pri radarju so to mikrovalovi ali radijski valovi, pri LiDAR-ju laserski pulzi, pri sonarju pa pulzi zvoka. Sonarji se delijo na aktivne in pasivne. Za pasivne sonarje je značilno, da pulzov zvoka ne oddajajo sami, temveč poslušajo oziroma sprejemajo zvoke, ki jih oddajajo razni podvodni objekti. Aktivni sonarji, ki so pogostejše v uporabi in se jim bomo v nadaljevanju tudi natančneje posvetili, zvočne pulze oddajajo sami. Sestavljeni so iz oddajnika in sprejemnika. Oddajnik oddaja zvočne pulze, nato pa sprejemnik posluša odboj pulzov od dna. Zvočni pulz je ustvarjen elektronsko, ustvari ga sonarski projektor, tega pa sestavljajo generator signala, ojačevalec in elektroakustični pretvornik.



Razdaljo, torej globino, se meri z meritvijo časa potovanja pulza od oddajnika nazaj do sprejemnika. Da pa ta čas lahko preračunamo v globino, moramo poznati hitrost širjenja zvoka v vodi. Ker se ta spreminja, se jo običajno na začetku meritve izmeri. Signal, ki pride do sprejemnika, gre nato skupaj s šumi skozi različne oblike obdelave, da lahko iz časa potovanja signala in ob upoštevanju hitrosti zvoka v vodnem stolpcu izračunamo globino. Pulz ima lahko konstantno ali spreminjajočo se frekvenco, prav tako pa je lahko sestavljen iz enega ali več snopov. Več snopov nam omogoča pokritost večjega območja dna istočasno. Sonarji so običajno nameščeni na plovilih, velikost teh plovil pa je odvisna od tipa sonarja, kraja meritev in globin ki jih merimo. Za zanesljivejše rezultate je plovilo opremljeno še z GPS-navigacijo, sistemom za detekcijo nagibanja plovila ter drugimi pomožnimi napravami.

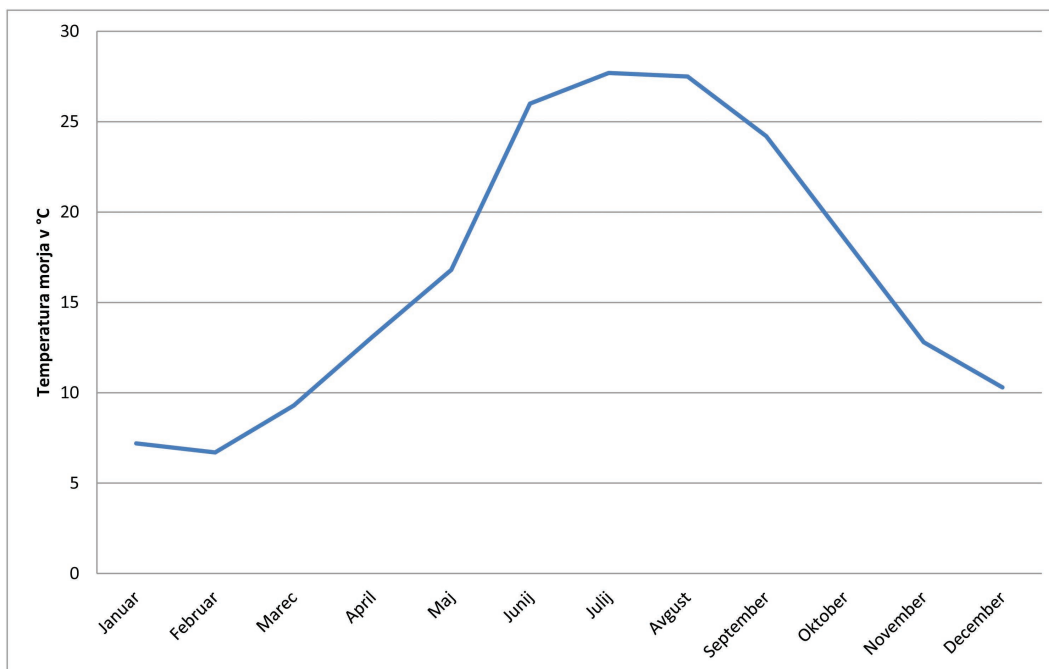
V grobem poznamo tri različne vrste sonarjev:

1. enosnopni sonar
2. bočni sonar
3. mnogosnopni sonar

Različni sonarji so skicirani na sliki 2.1. Enosnopni sonar, kot že ime pove, uporablja en sam zvočni snop. Snop je tipično širok med  $7^\circ$  in  $25^\circ$  in zbira le podatke navpično pod plovilom. Tak tip sonarja je zelo pogost in ga najdemo na skoraj vsakem plovilu. V splošnem se uporablja za merjenje globin in iskanje rib (ang. fish finder), bolj napredni enosnopni sistemi pa se uporabljajo tudi v raziskovalne namene. Bočni sonar se od enosnopnega razlikuje po dveh lastnostih - ni pritrjen direktno na plovilo in oddaja več zvočnih snopov. Kot je prikazano na sliki 2.1 je bočni sonar na plovilo pritrjen s kablom. Bočni sonar se navadno vleče na določeni višini nad morskim dnem. Oddaja dva široka zvočna snopa, enega levo in enega desno od plovila. Z uporabo prostorskega filtriranja (ang. spatial filtering) nato sprejemnik snope pretvori v visoko-ločljivostne podatke o morskem dnu. Bočne sonarje se pogosto uporablja za iskanje potopljenih predmetov in tudi iskanje utopljenec. V zadnjem času se bočni sonarji pojavljajo tudi v komercialni uporabi, v tem primeru gre za modele, ki so pritrjeni na plovilo in izrisujejo sliko morskega dna na zaslonu. Tretja vrsta sonarjev so mnogosnopni sonarji. Mnogosnopni sonarji delujejo podobno kot enosnopni, s tem da oddajajo veliko širši zvočni snop, širok tudi do  $120^\circ$  in nimajo le enega temveč mnogo usmerjenih sprejemnikov. Uporabljajo matriko usmerjenih sprejemnikov, tako da vsak sprejemnik pokriva le ozek del morskega dna, tipično med  $0.5^\circ$  in  $1^\circ$ . Tako mnogosnopni sonarji hkrati pridobivajo podatke o širokem pasu morskega dna pod plovilom.

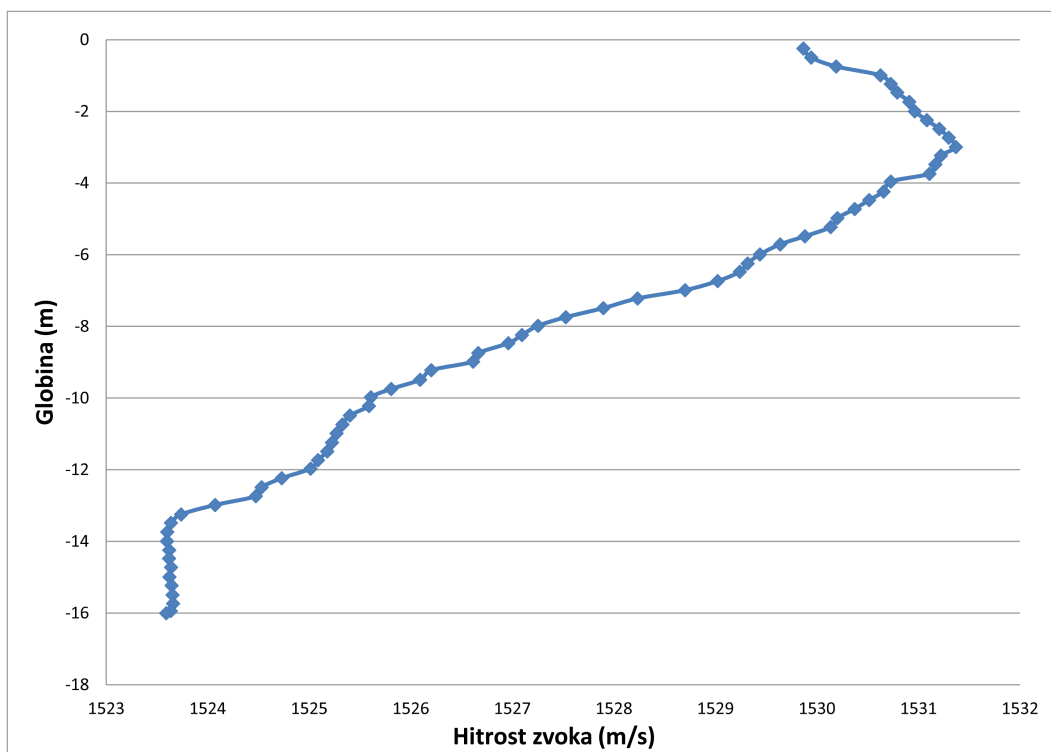
## 2.3 Podvodna akustika

Podvodna akustika je veda o širjenju zvoka v vodi in interakciji zvoka z vodo in njenimi mejami. Širjenje zvoka v vodi je pogojeno z izmeničnim stiskanjem in redčenjem vode. Hitrost zvoka določa longitudinalno gibanje valovnih front, povezano pa je s frekvenco in valovno dolžino. Hitrost širjenja zvoka skozi vodo ni konstantna, saj se



Slika 2.2: Primer gibanja temperature morja za leto 2011.

gostota vode, skozi katero potuje, spreminja glede na temperaturo, slanost in pritisk, ki je večinoma odvisen od globine [39]. Hitrost zvoka v vodi določa zmnožek frekvence in valovne dolžine. Ker v vodi hitrost zvoka ni odvisna od uporabljene frekvence, ampak od gostote medija, sta frekvenca in valovna dolžina medsebojno odvisni: z višanjem frekvence zmanjšamo valovno dolžino in obratno. Od izbire frekvence je odvisno delovanje sonarja in dobljeni rezultati [38]. Valovna dolžina je povezana z ločljivostjo dobljenih rezultatov, saj ne moremo zaznati predmetov, ki so manjši od uporabljene valovne dolžine [33]. Podvodni signali s širjenjem v vodi oslabijo sorazmerno s frekvenco signala in v obratnem sorazmerju s kvadratom razdalje od izvira signala [33]. V morju se del zvočne energije spremeni v toploto, del se je absorbira ali razprši ob različnih lebdečih delcih, zračnih mehurčkih, mikroorganizmih v vodi in drugih morskih organizmih. Na zvočno polje ali intenziteto zvoka vplivajo tudi odboji od gladine in dna ter lomljenje žarkov znotraj vodnega stolpca, ki ponekod intenziteto zvoka povečujejo, drugod pa znižujejo [3]. Absorpcija je postopno pojemanje zvoka oz. pretvorba dela zvočne energije v toploto. V morju je absorpcija zelo močna, saj vsebuje veliko magnezijevih sulfatov in drugih soli, ki zaradi svojih prožnostnih lastnosti pod vplivom zvočne energije močno spremenijo stisljivost medija in so tako odgovorne za izgubo zvočne energije v morski vodi. Frekvenca sonarja je parameter, od katerega je odvisen njegov doseg v globino in prodiranje zvoka dalje v sedimente. Zvočni signali med širjenjem v morski vodi namreč slabijo sorazmerno s frekvenco signala in obratno sorazmerno s kvadratom razdalje od izvora signala [33], čemur pravimo tudi sferična divergenca ali razhajanje



Slika 2.3: Spreminjanje hitrosti zvoka v vodnem stolpcu Koprskega zaliva. Primer meritve z dne 4.9.2014.

valovnega čela. Do slabenja prihaja tudi zaradi izgube energije med širjenjem valov zaradi absorpcije (prisotnost soli, še posebno  $\text{MgSO}_4$ , povzroča v morski vodi kar 25-krat višjo absorpcijo kot v sladki vodi [33]) in razpršitve ob različnih delcih v vodnem stolpcu [62]. Od izbire frekvence ni odvisen samo doseg ampak tudi ločljivost podatkov, saj z višjimi frekvencami dobimo natančnejše podobe dna (zato se jih npr. poslužujejo arheologi med iskanjem artefaktov na dnu). Za morja, plitvejša od 100 m, kot je slovensko, so za batimetrične meritve ustrezni sonarji s frekvencami, višjimi od 200 kHz. Odboj zvoka je osnova akustičnega detekcijskega sistema. Nanj vplivajo moč oddane zvočne energije, vpadni kot zvočnega žarka na dno ali oviro, odbojnost in absorpcija dna oz. ovire ter velikost obsijanega območja. Odboj in lom zvoka nista nikoli popolna. Vpadna zvočna energija se namreč razdeli na odbito, lomljeno, absorbirano (s strani vode in snovi, od katere se odbije) in razpršeno energijo. Tako dobimo nazaj pri sprejemanju odbitega zvočnega signala le majhen delček originalne oz. oddane energije. Vse omenjene vplive na intenziteto odbitega zvoka je potrebno upoštevati med predobdelavo podatkov in izničiti njihov vpliv na podatke o morskem dnu.

Intenziteto zvoka, odbitega od morskega dna, merimo v decibelih. Decibel (okrajšava dB) je enota brez dimenzije, s katero izražamo razmerje med spremenljivo količino in fiksno referenco. Uporabljamo ga pri meritvah v akustiki, fiziki, elektrotehniki in sorodnih področjih. Ker se za izračun uporablja logaritem, je z njim možno izražati zelo velik razpon razmerij z relativno majhnimi števili. Decibel se največkrat uporablja za izražanje glasnosti zvoka v primerjavi z referenčno vrednostjo 0 dB, ki predstavlja najnižjo intenziteto zvoka, ki ga še lahko zazna človek. Po standardu ANSI S1.1-1994, znaša referenčna intenziteta  $20 \mu\text{Pa}$  v zraku in  $1 \mu\text{Pa}$  v vodi.

Merjenje intenzitete odbitega zvoka ni poglavitni namen sonarjev na plovilih. Sonarji se uporabljajo predvsem za merjenje časa, ki ga zvok potrebuje, da pride do morskega dna in nazaj. Da je mogoče globino morja določati z merjenjem časa, ki ga potrebuje zvok od oddajnika do morskega dna in nazaj, je bilo ugotovljeno v začetku 19. stoletja, vendar so akustične meritve šele v dvajsetih letih 20. stoletja začele nadomeščati merjenje z utežmi [67]. Hitrost zvoka v mediju, skozi katerega zvočni signal potuje (npr. voda, sedimenti), je bistvena informacija pri računanju globin s sonarji, zato je nujno, da jo med izvajanjem hidrografskih meritev čim pogosteje izmerimo in dosledno upoštevamo. Hitrost zvoka v vodi merimo s sondo SVP (angl. Sound Velocity Profiler), ki (v našem primeru) na vsakega pol metra vodnega stolpca izmeri hitrost zvoka (slika 2.3). Povprečna hitrost zvoka v morski vodi je 1500 m/s, vendar je vodni stolpec v realnih razmerah nehomogen oz. stratificiran, saj se gostota vode spreminja glede na temperaturo, slanost in tlak. V splošnem velja, da se ob povišanju temperature vode za  $1^\circ \text{C}$  hitrost zvoka poviša za 3 m/s, pri povišanju slanosti za 1 promil za 1,3 m/s, na vsakih 100 m globine pa zaradi povišanja tlaka za 1,8 m/s [67]. Nehomogenost vodnega stolpca je posebej opazna v spomladanskem in jesenskem obdobju, zato so bile meritve SVP za zagotavljanje čim bolj kakovostnih rezultatov meritev takrat opravljane še pogosteje. Temperatura slovenskega morja se skozi leto spreminja, pade lahko tudi do  $5^\circ \text{C}$  pozimi in naraste na  $30^\circ \text{C}$  poleti. Primer temperature morja za leto 2011 je podan na sliki 2.2.  $25^\circ \text{C}$  razlike pa pomeni spremembo hitrosti zvoka tudi do 75 m/s. Upoštevanje razmer, kot sta temperatura

vode in globina je torej ključnega pomena za kakovost dobljenih podatkov.

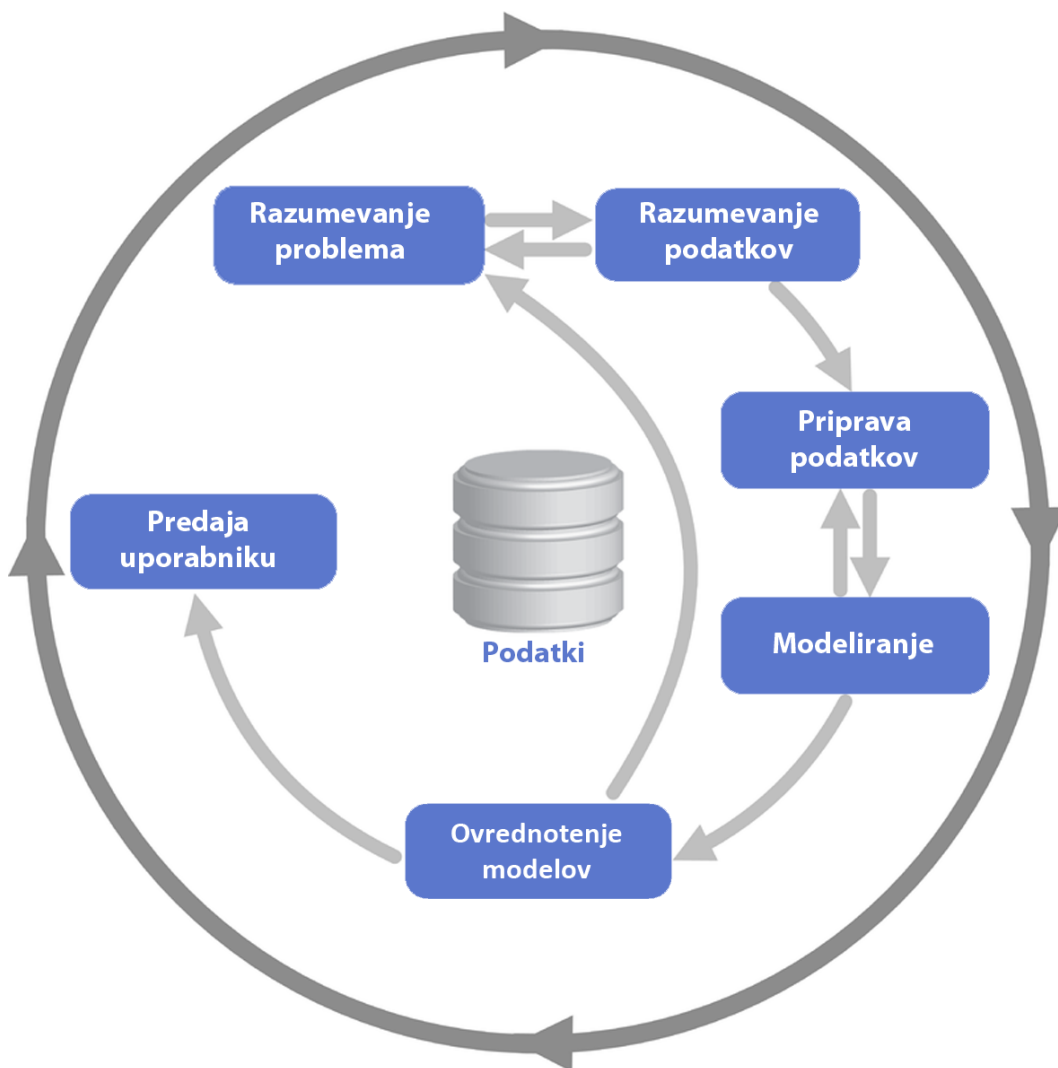
## 2.4 Odkrivanje zakonitosti v podatkih

Glavni cilj tega dela je iz akustičnih podatkov o morskem dnu izluščiti več kot le globine. Radi bi odkrili zakonitosti, skrite v akustičnih podatkih, na podlagi katerih bi bilo mogoče sklepati na tip morskega dna. Izraz “odkrivanje zakonitosti v podatkih” (Knowledge Discovery in Databases – KDD) se nanaša na širši proces iskanja znanja v podatkih. Eden od korakov v tem procesu je tudi podatkovno rudarjenje (data mining). Zanimiv je za raziskovalce na področjih strojnega učenja, razpoznavanja vzorcev, baz podatkov, statistike, umetne inteligence, ekspertnih sistemov in vizualizacije podatkov.

Skupni cilj odkrivanja zakonitosti v podatkih je luščenje znanja iz podatkov v kontekstu velikih baz podatkov. To dosega z uporabo metod in algoritmov strojnega učenja za luščenje (identifikacijo) znanja v obliki vzorcev in modelov, glede na izbor evalvacijskih mer, pragov, uporabo baz podatkov, vzorčenje ter raznih transformacij podatkov. Celoten proces iskanja in interpretacije vzorcev in modelov iz podatkov vključuje ponavljajočo se uporabo naslednjih korakov:

1. Razumevanje aplikacijske domene, relevantnega predznanja, ciljev končnega uporabnika;
2. Generiranje ciljne množice podatkov: izbira (pod)množice podatkov, spremenljivk ali vzorca podatkov, na katerem se bo izvajal proces odkrivanja zakonitosti;
3. Predobdelava/čiščenje podatkov: odstranitev šuma, zbiranje novih podatkov namesto šumnih, strategije za obravnavo manjkajočih podatkov, upoštevanje časovnih zaporedij;
4. Redukcija podatkov in projekcije: iskanje uporabnih značilnk za predstavitev podatkov v odvisnosti od cilja, uporaba tehnik redukcije dimenzionalnosti;
5. Izbira ustrezne metode podatkovnega rudarjenja: odločitev ali gre za klasifikacijo, regresijo, razvrščanje v skupine ali kak drug tip naloge;
6. Izbira ustreznega algoritma ter določitev parametrov algoritma;
7. Strojno učenje: uporaba izbranega algoritma za odkrivanje zakonitosti;
8. Interpretacija rezultatov;
9. Vrednotenje odkritega znanja ter uporaba v aplikacijski domeni;

Že leta 1999 se je oblikoval industrijski standard za odkrivanje zakonitosti v podatkih, imenovan CRISP-DM [22] (CRoss-Industry Standard Process for Data Mining). Standard natančno določa potek analize podatkov s pomočjo podatkovnega rudarjenja. CRISP-DM standard je neodvisen od problema in orodja, osredotočen na



Slika 2.4: Podatkovno rudarjenje po metodologiji CRISP-DM. Vir: [www.crisp-dm.org](http://www.crisp-dm.org)

poslovne probleme in je ogrodje, ki služi kot vodilo pri uporabi podatkovnega rudarjenja v industriji. Diagram metodologije CRISP-DM je prikazan na sliki 2.4

Metodologija CRISP-DM določa cikel 6-ih faz:

1. **Razumevanje problema (Business Understanding)**  
Razumevanje ciljev in potreb projekta ter definicija problema za podatkovno rudarjenje
2. **Razumevanje podatkov (Data Understanding)**  
Začetno zbiranje in spoznavanje podatkov, preverjanje kakovosti podatkov
3. **Priprava podatkov (Data Preparation)**  
Priprava in čiščenje podatkov, izbor primerov in značilk
4. **Modeliranje (Modeling)**  
Izbor ustreznih tehnik modeliranja in njihova aplikacija, nastavljanje parametrov
5. **Ovrednotenje modelov (Evaluation)**  
Ocena tehnik modeliranja, ocena v kolikšni meri smo zadovoljili cilje in potrebe projekta
6. **Predaja končemu uporabniku (Deployment)**  
Predaja/implementacija modela z razlago, tako da je celoten proces ponovljiv

Metodologiji CRIPS-DM smo sledili tudi pri obravnavi problema akustične klasifikacije morskega dna.

## 2.5 Klasifikacija morskega dna

Prvi poskusi kartiranja morskega dna segajo v 13. stoletje, ko so pomorski trgovci risali sheme mediteranskega morja [12]. Od tistih časov smo nadaljevali z risanjem batimetrični modelov morskega dna, še posebej v plitvih vodah, kjer spremenljive globine ogrožajo varnost plovbe. Geološke in biološke raziskave so se začele v 19. stoletju, najprej s pomočjo vzorcev vzeti iz morskega dna s primitivnimi orodji [34]. Kasneje so se uporabljale vse bolj napredne metode za raziskovanje morskega dna sublitoralnega pasu, kot so različna grabila, vlečne mreže in sani za zajem vzorcev, video sistemi itd. Težava vseh teh metod je zamudno in zahtevno vzorčenje, ki nas omejuje pri raziskovanju velikih območij morskega dna [5], [130]. Te metode zagotavljajo podrobne informacije o majhnem delu morskega dna, kjer vzamemo vzorce, pogosto pa je izredno zahtevno dobiti natančno predstavitev širšega območja prostorske razporeditve biofizičnih karakteristik morskega dna brez podrobnih in dragih raziskav z gosto mrežo vzorčenja ali transektov.

### 2.5.1 Neposredno vzorčenje z metodo linijskih presekov

Tehnika neposrednega vzorčenja vključuje uporabo bentoških grabljic ali strgal za vzorčenje bentoške vegetacije z namenom identifikacije vrst. S to tehniko lahko natančno določimo tipe morskega dna, vendar pa ne moremo dobiti natančnih podatkov o položaju in obsegu območja, ki ga posamezni tipi pokrivajo. Zato se sočasno uporablja tehnika opazovanja s potapljači, ki preučijo posebna mesta, prepoznavna iz fotografij površine [107]. Opazovanja s potapljači lahko izvedemo na več načinov. En izmed takšnih načinov je uporaba presekov, s katerimi dobimo natančne podatke o tipih morskega dna na obravnavanem preseku. Najprej je potrebno izbrati 3 ali več reprezentativnih območij, znotraj katerih se postavi merilne trakove, ki označujejo linijske preseke. Dolžine presekov se določi glede na spremembe vegetacije, morskega dna in globine. Na redne intervale (npr. na 10 ali 25 m) se v morsko dno zabije količek, ki predstavlja nadzorno točko. Vsakič, ko potapljači prečkajo presek, en potapljač položi merilni trak med 3 količke, medtem ko drugi posname spremembe vzdolž merilnega traku [71]. Po opisu vsakega intervala, se merilni trak prenese na naslednji interval, vse do konca preseka. Opis intervala vključuje popis vrst 50 cm na vsako stran od merilnega traku. Nastali zapis se prikaže v obliki profila ali preglednice, ki prikazuje spremembe od enega obdobja do drugega [1].

V novejšem času se namesto popisa ali s samim popisom s podvodno videokamero posname območje vzdolž preseka. Snemanja lahko izvajajo potapljači ali pa so daljinsko vodena preko računalniških programov in plovil. Dokler je pozicija posnetkov natančno določena, predstavljajo video-posnetki primerno terensko izmero, saj so hitreje in lažje izvedljivi [107].

Terenske meritve so bistvenega pomena za razlago prepoznavnih značilnosti tipov morskega dna iz daljinsko zaznanih slik ali sonarskih meritev, poleg tega pa služijo tudi kot oporna točka pri preverjanju interpretacije slik, npr. da se prepričamo, da nismo makroalge, školjke ali grebene zamenjali za morske travnike. Metoda zagotavlja natančnejše podatke o porazdelitvi morskega rastja, vendar so opazovanja in popisi izvedljivi le na majhnem območju zaradi česar je zamudna in dražja. Zato je delo na terenu samo po sebi pogosto preveč drago in neustrezno za kartiranje večjih obalnih območij [79], [1].

### 2.5.2 Zračni in satelitski sistemi

Razvoj zračnih in satelitskih sistemov za daljinsko zaznavanje je v zadnjih nekaj desetletjih povečal dostopnost optičnih in laserskih daljinsko-zaznanih podatkov za ekološke raziskave širših območij. To je bistveno pripomoglo k boljšemu razumevanju prostorskih vzorcev kopnega [96], [44]. V zadnjem desetletju smo bili priča velikemu napredku pri tehnologijah za kartiranje kopnega. Analize letalskih in satelitskih posnetkov omogočajo natančno kartiranje različnih habitatov, raziskave krčenja gozdov itd. Enake metode so deloma uporabne tudi za raziskave morskega dna [103], vendar je njihova uporabnost na morju omejena na plitve obalne predele zaradi omejene penetracije svetlobe skozi vodo [6]. Svetloba le v redkih primerih prodre globlje kot 20 metrov pod gladino morja, v našem primeru, kjer se ukvarjamo s Tržaškim zalivom, pa je ta številka še bistveno nižja. Po podatkih morske biologije postaje v Piranu, je povprečna Secchiejva globina [121] (globina, do katere je viden disk belo črne barve)



za slovensko morje okoli 8 metrov [106]. Podatek velja za zunanje dele slovenskega morja in je v zalivih še precej nižji. Za preučevanje morskega dna so ustreznejši batimetrični laserski sistemi [37] z zelenim laserjem, ki lahko dosežejo dva do trikratnik Secchijeve globine [121], vendar so še vedno zelo odvisni od motnosti vode. Največje težave so v zalivih, kjer so morski tokovi manj izraziti, vpliv valovanja večji in zamuljenost zelo izrazita.

### 2.5.3 Akustična klasifikacija morskega dna

Šele v zadnjih letih, z razvojem akustičnih sistemov, se znanstveniki, ki se ukvarjajo z morjem, približujejo kakovosti in ločljivosti kartiranja kolegov, ki se ukvarjajo s kopnim. Akustični sistemi uporabljajo zvočne signale, ki so neodvisni od prosojnosti vode in nam omogočajo zbiranje kakovostnih podatkov tudi na območjih, kjer optične in laserske metode odpovejo. Poleg tega akustični sistemi dosegajo bistveno višjo kakovost in ločljivost po kriterijih mednarodne hidrografske organizacije IHO (International Hydrographic Organization). Mnogosnopni akustični sistemi (multi-beam echo sounders - MBES) ustrezajo kriterijem po IHO standardu Special Order meritev, sistemi z batimetričnim laserjem pa le kriterijem po IHO standardu Order 1 meritev [53], [62].

Razvoj akustičnih raziskovalnih metod kot so enosnopni akustični diskriminatorni sistem (single-beam acoustic ground discrimination systems SB-AGDS), sistem bočnega sonarja (sidescan sonar systems -SSS) in v zadnjem času mnogosnopni akustični sistem zagotavlja orodja za raziskave širših območji morskega dna [25], [87], [88], [6]. Primerjava različnih sistemov je prikazana na sliki 2.1. Ločljivost akustičnih sistemov je odvisna od lastnosti le-teh, izvedbe meritev in v nekaterih primerih tudi od globine. Načeloma pa je od nekaj deset metrov pa vse do nekaj deset centimetrov [6], [69], [31]. S kombinacijo novih akustičnih metod in konvencionalnih metod z zajemanjem vzorcev se ponujajo nove možnosti raziskav in določanja bioloških in geoloških karakteristik morskega dna. V zadnjem desetletju smo bili priča mnogim poskusom visoko ločljivostnega kartiranja morskega dna zaradi bolj široko dostopnih in cenejših orodij skupaj z velikim napredkom v računski zmogljivosti računalnikov in geografskih informacijskih sistemov (GIS) [88].

#### Enosnopni sonarji

Najpreprostejši akustični sistem, ki beleži povratni signal odbit od morskega dna je SB-AGDS. SB-AGDS je sistem, ki beleži podatke le navpično pod plovilom. Uporablja enosnopni sonar s frekvenco med 30 in 200 kHz [87], [16]. Na voljo je nekaj lastniških/komercialnih sistemov ki se uporabljajo za študije morskih habitatov (RoxAnn, QTC-View, EchoPlus). Vsi ti sistemi beležijo globine in nekatere lastnosti odbitih signalov, povezanih s tipom morskega dna [16], [5]. Podatke, pridobljene s sistemom SB-AGDS, delimo v različne akustične razrede glede na lastnosti signalov iz povratnih ehogramov. Akustičnim razredom se nato določi tipe tal glede na podatke o posameznih delih morskega dna, ki so bili predhodno pridobljeni z vizualnimi metodami (kamera, potapljači, vzorčenje ...) Podatke enosnopnega sonarja je veliko lažje obdelovati v primerjavi s podatki SSS (sidescan sonar) ali MBES (multibeam

echo sounder). Pri SSS in MBES sistemih je interakcija akustičnega signala z morskim dnom, predvsem na območju ki ni navpično pod plovilom, zelo kompleksna. To precej oteži klasifikacijo. Navedeno je prednost SB-AGDS sistemov pred dražjimi in kompleksnejšimi MBES in SSS sistemi. Pomanjkljivost enosnopnega sistema se pokaže pri pokrivnosti oziroma natančnosti idealnih map klasificiranega morskega dna. Pri kartiranju s SB-AGDS sistemom je potrebno dobljene podatke v posameznih točkah interpolirati preko celotnega območja. To pri homogenih predelih morskega dna ne povzroča težav, lahko pa vodi do precejšnjih napak na nehomogenih predelih morskega dna, kjer nimamo dovolj akustičnih podatkov. Poleg tega z naraščanjem globine narašča velikost akustičnega odtisa (področje od katerega se odbije zvočni snop). Pri enosnopnih sonarjih je širina zvočnega žarka/snopa med  $15^\circ$  in  $25^\circ$ , kar je relativno veliko glede na mnogosnopne sisteme (okoli  $1^\circ$ ). To povzroči omejeno uporabnost enosnopnih sistemov na večjih globinah [43], [18], [16]. V večini primerov so SSS in MBES sistemi močno prekašali SB-AGDS sisteme [117], [97], [45], [82], [131], [52].

### Bočni sonarji

Signale akustičnih odbojev bočnih sonarjev (SSS) geologi že dolgo uporabljajo za klasifikacijo morskega dna v geološke razrede. Obstaja povezava med močjo akustičnih odbojev in geotehnoško sestavo morskega dna [27], [35], [41]. Bočni sonarji so bili razviti okoli leta 1940 [69], delujejo na relativno visokih frekvencah (med 100 in 500 kHz) in beležijo akustične odboje s širokega področja dna (levo in desno od plovila) ter tako ustvarijo teksturirano sliko morskega dna [87], [69], [5]. Z zajemanjem podatkov po paralelnih linijah in upoštevanju pokrivnosti sonarskega snopa pri načrtovanju razmika med linijami je z bočnim sonarjem enostavno doseči zvezno pokrivnost ciljnega področja. Tako ni težav z nenatančnostjo zaradi interpolacije, ki se pojavi pri uporabi enosnopnih sistemov. S pomočjo posnetkov morskega dna se je klasifikacija izvajala s pomočjo strokovnjakov, ki so glede na teksturo ali moči odbojev interpretirali rezultate meritev. Ročno določene akustične razrede se nato poveže z geološkimi razredi s pomočjo preverjanja – fizični zajem vzorcev, kamera, potapljači [127], [13]. Klasifikacija s pomočjo strokovnjakov se je uporabljala tudi za iskanje podobnih bioloških karakteristik v različnih okoljih [4], [95]. Kasneje je bilo več pozornosti usmerjene v avtomatske metode za klasifikacijo morskega dna s podatki bočnega sonarja. Prednost razvitih algoritmov pred interpretacijo strokovnjakov je predvsem v hitrosti in objektivnosti pri interpretaciji podatkov [86], [129]. Pri avtomatski klasifikaciji SSS podatkov se večinoma uporablja lastnosti akustičnih odbojev – slike odbojev (intenzitete, strukturo, lastnosti površja) [85], [133] nekaj pa je bilo tudi poskusov klasifikacije s pomočjo analize signalov. Klasifikacija s pomočjo analize signalov se je uveljavila pri enosnopnih sistemih, sledili pa so poskusi uporabe pri bočnih sonarjih [17]. Težave z analizo signalov pri SSS sistemih se pojavijo, ker večina teh sistemov ne beleži batimetrije morskega dna. Odsotnost geometrijskih podatkov, potrebnih za določitev kotnih odvisnosti signala pa močno oteži analizo odboja [87]. Na tem področju imajo bistveno prednost MBES sistemi, ki hkrati beležijo odboje in batimetrijo dna.

## Mnogosnopni sonarji

V zadnjih letih so se MBES sistemi izkazali za najprimernejše orodje za raziskovanje morskega dna, predvsem zaradi možnosti hkratnega beleženja batimetričnih podatkov in odbojev od dna na širokem področju levo in desno od plovila [88]. Podobno kot SSS nam tudi MBES sistemi omogočajo kontinuirano prekrivnost območja, ki ga raziskujemo. Potrebno je le načrtovati paralelne linije snemanja na primerni oddaljenosti, da zagotovimo ustrezno prekrivnost podatkov. MBES sistemi so zelo kompleksni in zahtevajo izredno sofisticirane GPS sisteme in sisteme za zaznavo premikov, da dobimo natančne lokacije vseh odbojev [87], [5], [129]. Na trgu je veliko različnih sistemov, ki se razlikujejo po namembnosti in globini na kateri lahko delujejo. Do nedavnega je bila uporaba MBES sistemov v akustični klasifikaciji morskega dna omejena zaradi visokih stroškov snemanja in tehničnih omejitev pri shranjevanju in obdelavi podatkov [69], [80]. Z večanjem računske moči sodobnih računalnikov, pocenitvijo shramb podatkov in večjo dostopnostjo MBES sistemov so se le-ti začeli pogosteje uporabljati kot raziskovalna orodja in orodja za snemanje batimetrij in odbojev od morskega dna. Večje število snemanj z MBES sistemi pa pomeni tudi več pridobljenih podatkov in omogoča več študij in raziskav na tem področju.

### 2.5.4 Pristopi h klasifikaciji

Bistvo akustične klasifikacije je segmentacija morskega dna v diskretne razrede na podlagi sonarskih podatkov. Kot je bilo poudarjeno, so se za najbolj primerno orodje zbiranja podatkov izkazali podatki posneti z MBES sistemi. V podjetju Harpha Sea imamo visoko usposobljen kader, ki je kvalificiran za delo z MBES sistemi ter tudi ustrezna plovila in sonarje za zajem podatkov. Na voljo imamo podatke za celotno slovensko morje, posnete med leti 2006 in 2008. Sonarski podatki posneti z MBES sistemi so zelo kompleksni in zahtevni za obdelavo. Podobno kot pri SSS sistemih se je tudi klasifikacija morskega dna na podlagi podatkov MBES sistemov začela ročno, s pomočjo strokovnjakov [108], [119]. Trenutno sta uveljavljena dva pristopa k analizi in obdelavi sonarskih podatkov za potrebe klasifikacije morskega dna – fenomenološki pristop in geoakustično modeliranje. Pri geoakustičnem modeliranju se poskuša akustične podatke transformirati v dejanske fizične lastnosti morskega dna, kot so velikost zrn, poroznost, itd. [10], [21], [42]. Tak pristop zahteva vhodni model za klasifikacijo odbojev, nekatere ročno dodane geoakustične parameter in kalibriran MBES sistem. V večini primerov, tako kot v našem, tak pristop ni primeren, ker so sonarske meritve izvedene z drugimi nameni (batimetrija) in ne izključno za potrebe analize podatkov in klasifikacije morskega dna. Fenomenološki pristop predvideva obstoj korelacije med morfološkimi in fizičnimi lastnostmi morskega dna ter akustičnimi podatki. Podatke razdeli v akustične razrede brez direktnega sklepanja na dejanske fizične lastnosti dna. S pomočjo vizualnega preverjanja (potapljači, zajem vzorcev, kamere) se nato akustične razrede poveže z dejanskimi tipi tal. Obstaja nekaj orodij, ki delujejo tako, vendar je težava v tem, da so večinoma komercialna in zato ne vemo točno kako delujejo. Primer takih orodij so QTC-Multiview (Preston 2004), RoxAnn [57], Echoplus [8]. Komercialno orodje QTC-view [112] uporablja metodo k-tih povprečij za razvrščanje v skupine (k-means clustering), v delu [61] je

opisan pristop z odločitvenimi drevesi v delu [21] pa pristop z uporabo nevronske mreže. Gre za komercialna orodja in veliko podrobnosti ni znanih - torej ni mogoče ponoviti postopka in ga preizkusiti, prav tako pa je izredno otežena primerjava rezultatov z rezultati omenjenih orodij. Poleg tega nekatera zahtevajo namenske sonarje, praviloma pa so za potrebe klasifikacije morskega dna potrebne namenske meritve.

## Poglavje 3

# Izvedba hidrografskih meritev

Za merjenje globin se uporabljajo različni tipi sonarjev, ki jih najpogosteje delimo glede na sevano število ultrazvočnih snopov ter kot oddajanja in sprejemanja. Tako ločimo enosnopne (singlebeam), sestavljene enosnopne (sweep system) ter mnogosnopne (multibeam) sonarje. Gre za naprave, pri katerih se meri čas, v katerem ultrazvočni signal skozi vodo prepotuje pot od oddajnika do morskega dna in nazaj [123]. V praksi pa je za izmero globin potrebno:

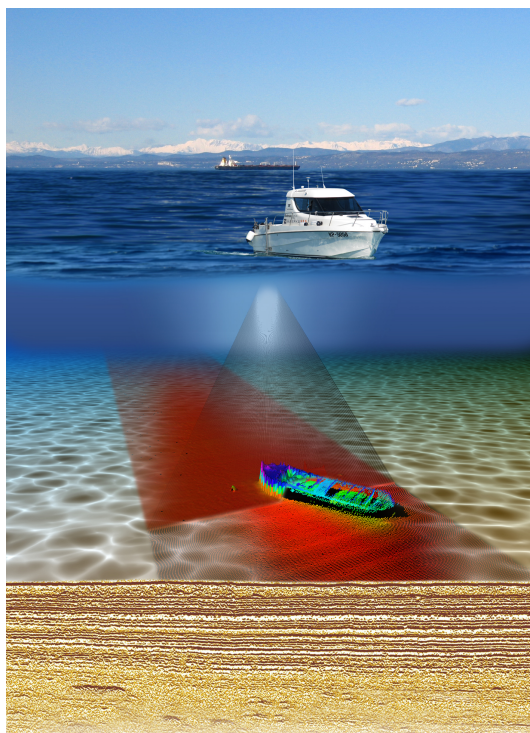
- proizvesti kratek impulz ultrazvoka in ga usmeriti k dnu,
- sprejeti in ojačiti odboj z dna,
- izmeriti čas od trenutka oddaje ultrazvočnega impulza do sprejema in
- avtomatsko spremeniti razliko (s poznano hitrostjo) v višino vodnega stolpca, skozi katerega je potoval zvok.

Meritve morskega dna je potrebno vselej obravnavati v funkcijski povezavi s parametri, ki vplivajo na hitrost zvoka v vodi. Med njimi moramo posebej izpostaviti:

- fizikalne lastnosti vode (temperatura, slanost ali prevodnost, gostota, pritisk) in
- hidrološke lastnosti vode (onesnaženost, delci organskega porekla).

Hitrost zvoka se spreminja glede na razmere v vodi. Izmerimo jo lahko s sondo SVP (sound velocity profiler), najpogosteje na vsake pol metra globine. V bolj neugodnih razmerah, ko se temperatura vode z globino bistveno spreminja, so potrebni pogostejši odčitki. Pred izvedbo izmere so potrebne natančne kalibracije instrumentov, pri katerih se določijo kompenzacijski koti glede na predpostavljene vrednosti izhodiščne horizontalne ravnine. Ob upoštevanju vseh parametrov kalibracije ter dovolj pogostih določitev hitrosti zvoka v vodi lahko dosežemo nekaj centimetrsko natančnost izmere globin morskega dna. Vse merske naprave morajo biti povezane med seboj prek računalnika in vodene z ustrezno programsko opremo, kot je na primer PDS2000 [126], HYPACK [60], QINSy [115]. Za neoporečne rezultate potrebujemo dobro podučeno in izurjeno ekipo merilcev, ki odgovorno spremlja potek meritev in podatke o njihovi kakovosti.

Za potrebe tega dela smo meritve opravili z mnogosnopnim sonarjem. Meritve je izvedel hidrograf z mednarodno licenco iz podjetja Harpha Sea, d.o.o., kar zagotavlja, da so meritve priznane s strani mednarodne hidrografske organizacije IHO. Z naprednimi GNSS sistemi v kombinaciji s sodobnim mnogosnopnim sonarjem smo poskrbeli za zajem kakovostnih podatkov s centimetrsko natančnostjo. Meritve morskega dna smo izvedli s plovilom Lyra (slika 3.1) dolžine 8 metrov, na katerem je vgrajen mnogosnopni sonar SeaBat 8125, podpovršinski sonar SES2000-compact ter enosnopni sonar HydroStar. Vsi sonarji so integrirani z GNSS sprejemnikom Javad Duo-G2D ter INS senzorjem TSS Mahrs.



Slika 3.1: Prikaz ladje Lyra in delovanja mnogosnopnega sonarja. Vir: Harpha Sea, do.o.o Koper

V kombinaciji s sistemom za zajem akustičnih podatkov se uporablja tudi SVP (Sound Velocity Profiler). SVP je orodje za meritve hitrosti zvoka v vodnem stolpcu. Zaradi različnih temperatur vode na različnih globinah pride do loma valovanj - v našem primeru zvoka. Lomne količnike moramo upoštevati, da lahko ustrezno interpretiramo meritve in dobimo natančno izmerjene globine.

### 3.1 Določanje položaja plovila in natančnost pridobljenih podatkov

Velik vpliv na rezultate pri opravljanju hidrografskih meritev ima sistem za določanje položaja merilne platforme - plovila. Od natančnosti določitve absolutnega položaja

plovila v prostoru je neposredno odvisna natančnost vseh nadaljnjih meritev, opravljenih s plovila. Ladja ima šest prostostnih stopenj. Tri so translacije:

- Premik naprej/nazaj (surge)
- Premik vstran (sway)
- Premik gor/dol (heave)

Tri pa rotacije:

- Nagib (roll), rotacija okoli osi surge
- Naklon (pitch), rotacija okoli osi sway
- Usmerjenost (yaw oz. heading), rotacija okoli osi heave

Javad duo-G2D je GNSS sprejemnik, ki nam določa zelo natančen položaj ladje. Sprejemnik skrbi za tri prostostne stopnje, z njim nadziramo premike naprej, vstran in spremembe višine zaradi plimovanja in valovanja morja. Sprejemnik pri izračunu natančne pozicije upošteva RTK (real time kinematic) korekcije [132]. Za določanje položaja plovila uporabljamo GNSS RTK metodo izmere z navezavo na mrežo SIGNAL [49]. V primeru plovila gre za zelo dinamično merilno platformo, ki se v času premika in rotira v vseh treh oseh. Metode za določanje položaja so zaradi dinamičnosti merskega okolja zelo omejene - edina praktično uporabna metoda je določanje položaja z uporabo globalnih navigacijskih satelitskih sistemov (GNSS) [123]. Princip delovanja GNSS temelji na več konstelacijah namenskih satelitov (ameriški GPS, ruski GLONASS, kitajski COMPASS, evropski GALILEO ...), ki preko radijskih valov oddajajo točen čas z vgrajene atomske ure ter svojo lokacijo. GNSS sprejemnik na podlagi zakasnitve signala s posameznega satelita in satelitove lokacije izračuna svoj položaj v globalnem koordinatnem sistemu. Zaradi atmosferskih vplivov na pot signala, netočnosti določitve položaja satelitov in napak ur na satelitih je samo z uporabo GNSS signala mogoče določiti položaj na približno 15 m natančno [73]. Takšna natančnost je za uporabo pri hidrografskih meritvah prenizka. Zadostno natančnost določanja položaja v realnem času lahko dobimo z uporabo RTK metode GNSS meritev. Pri tej metodi poleg GNSS sprejemnika na merilni platformi potrebujemo tudi statičen GNSS sprejemnik in medij za povezavo med njima. Statični GNSS sprejemnik mora imeti natančno določen položaj v prostoru in ne sme biti postavljen dlje kot cca. 50 km od delovišča. Na podlagi razlike meritev faze nosilnega valovanja signalov med obema GNSS sprejemnikoma se izračuna bazni vektor - prostorski vektor med obema sprejemnikoma - in določi položaj dinamičnega GNSS sprejemnika s centimetrsko natančnostjo [123]. Za prenos podatkov med sprejemnikoma se lahko uporabi več različnih medijev, najpogosteje omrežje mobilne telefonije ali radijska povezava. Prednost omrežij mobilne telefonije je splošna dostopnost na širšem območju priobalnega pasu, slabost pa omejen doseg na območjih, ki so bolj oddaljena od obale. Prednosti radijskih povezav so neodvisnost od tuje infrastrukture in večji doseg, slabost pa uporaba le enega oddajnika na fiksni lokaciji in potreba po vidni liniji med oddajnikom in sprejemnikom. Pri hidrografskih meritvah je smotrno uporabiti obe metodi in, glede na lokacijo

delovišča, izbrati tisto, ki nam v danem trenutku nudi zanesljivejšo povezavo. Izhodni podatek GNSS sprejemnika je položaj v globalnem elipsoidnem koordinatnem sistemu WGS84, v katerem je položaj vsake točke določen z geografsko širino in dolžino, ter elipsoidno višino. Za praktično uporabo je prikaz podatkov v elipsoidnih koordinatnih sistemih neprimeren, saj imata, razen v bližini ekvatorja, ločni enoti geografske dolžine in širine različno dolžino na površju Zemlje. Na območju Slovenije tako ena sekunda geografske širine pomeni dolžino 30,87 m, ena sekunda geografske dolžine pa 21,90 m. Podatke za praktično uporabo projiciramo na izbran ravninski koordinatni sistem, v našem primeru gre za 33 cono univerzalne prečne Mercatorjeve projekcije (angl. Universal Transverse Mercator – UTM33N). Projicirani podatki so razumljivejši, saj so pravokotne koordinate v dolžinskih merskih enotah - metrih, sistem pa je prav tako enostavno reprojicirati brez deformacij, če se podatki zahtevajo v drugem koordinatnem sistemu. Višina nad ploskvijo elipsoida je matematična količina, zaradi nepravilnosti zemeljskega površja in Zemljinega težnostnega polja pa nima enolične navezave na nadmorske višine, ki jih uporablja državni višinski koordinatni sistem. Za pretvorbo iz elipsoidnih višin v državni višinski koordinatni sistem se uporablja model geoida.

Ker z GNSS sistemom določamo samo lokacijo GNSS antene, merska platforma pa se med opravljanjem meritev tudi ziblje in vrti, potrebujemo za določitev njenega natančnega položaja še podatka o smernem kotu in kotnih hitrostih vrtenja okoli vseh treh osi (slika 3.5). Podatek o smernem kotu dobimo z dinamično optimiziranim inercialnim senzorjem, podatke o rotacijah okrog osi pa z dinamičnim senzorjem gibanja. Inercialni senzor TSS Mahrs meri nagib, naklon in zasuk oz. usmerjenost plovila. Natančnost izmerjene usmerjenosti (heading) je  $0,1^\circ$ , natančnost izmerjenega nagiba (roll) in naklona (pitch) je  $0,03^\circ$ . Kombinacija GNSS sprejemnika in inercialnega sensorja nam zagotavlja kakovostne podatke za natančno umeščanje v prostor – direktno georeferenciranje.

Globine, ki so uporabljene v nalogi, so vezane na uradni državni višinski koordinatni sistem (t. i. geodetsko ničlo) in se nekoliko razlikujejo od tistih, ki so uporabljene na navtičnih kartah, kjer je za varnost plovbe potrebno upoštevati najnižji nivo morske gladine. Hidrografska ničla je v Sloveniji umerjena na srednji nivo nižjih nizkih vod živih morskih men in se pri nas nahaja 0,48 m pod geodetsko ničlo [135]. Za namene plovbe oz. za primerljivost podatkov z navtičnimi bi torej morali vsem globinam odšteti 0,48 m.

## 3.2 Mnogospopni sonar Reson SeaBat 8125

Mnogospopni sonar Reson SeaBat 8125 zadošča vsem kriterijem po IHO standardu Special Order meritev [62]. Tehnični podatki sonarja so navedeni v tabeli 3.1. Z uporabo 240 dinamično usmerjenih sprejemnih snopov, SeaBat 8125 snema 120 stopinj morskega dna hkrati. Zaznava morskega dna s tem sonarjem je izredno zanesljiva, ločljivost globine je 6 mm, meritve pa lahko v realnem času spremljamo na zaslonu sonarja. Vgrajen napredni avtopilot omogoča lažji nadzor nad linijo meritve. Stabilizacija nagiba poskrbi za uspešno zaključene meritve tudi v slabem vremenu, brez izgube natančnosti. Sonar lahko istočasno pokrije širino pasu, ki je 3,5-krat večja



od globine vode. To pomeni, da so pri globini 10 m točke v povprečju izmerjene na vsakih 15 cm. Opisana naprava je uporabna za snemanje globin do 100 m. Preprosto povedano, je sonar sestavljen iz oddajnika in sprejemnika ter kontrolne enote, povezan pa je z računalnikom in monitorjem. Delovna frekvenca mnogospnega sonarja Reson SeaBat 8125 je 455 kHz. Širina enega žarka je 0,5° transversalno in 1° longitudinalno. Ločljivost sonarja znaša med 1 cm in 2 cm, zaradi učinkov nagibanja plovila pa je natančnost meritev okoli 5 cm. Največja možna operativna hitrost plovila s tovrstnim sonarjem je 8 vozlov. Število poslanih zvočnih impulzov na sekundo se glede na globino morja spreminja. Za izvedbo meritev mora biti mnogospni sonar zaradi svoje velikosti in teže, nameščen na večjem plovilu, dolžine 8 m. Z njim se običajno meri do približno 1 m globine oziroma odvisno od same oblikovanosti dna, v izjemnih primerih tudi do 0,5 m globine, vendar je snemanje pri takšnih globinah zelo zahtevno, zato se v takšnih primerih običajno uporablja enospni sonar.

Materiali	Sprejemnik in transmitter iz aluminija
Dimenzije	50 x 19 x 14 cm
Teža	24 kg v zraku 9kg v vodi
Delovna frekvenca	455 kHz
Širina žarka	0.5° trasversalno 1° longitudinalno
Število horizontalnih žarkov	245
Natančnost	1cm RMS centralni žarki, <2cm RMS zunanji žarki
Linearna ločljivost	6mm
Širina pokritja	120°
Max operativna hitrost	8 vozlov
Ping rate	Spremenljiva glede na globino, vsaj 10/s
Temperatura med delom	-5°C do +40°C

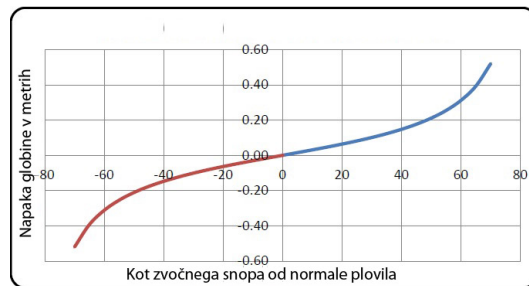
Tabela 3.1: Tehnični podatki sonarja Reson Seabat 8125.

### 3.3 Kalibracija batimetričnega sistema

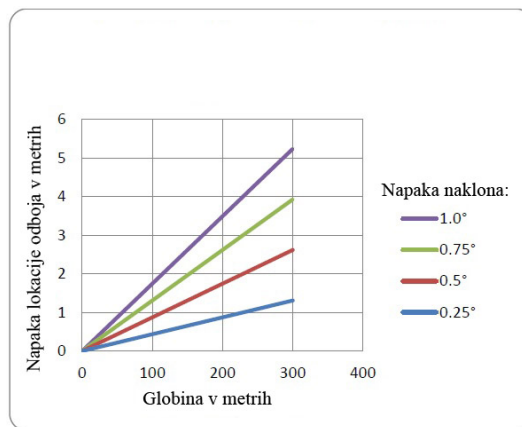
Sistem se kalibrira v rednih predvidenih intervalih glede na specifikacije, kot jih predpisuje proizvajalec globinmera (R2 Sonic) in programa za zajem in obdelavo (Reson). Sistem se dodatno obvezno kalibrira kadar pride do sprememb katerega od glavnih senzorjev (globinomer, GNSS, INS). Ponovna kalibracija je nujna že pri najmanjšem premikanju lege enega od glavnih senzorjev.

Kalibracija sistema poteka po metodi »Patch Test«, ki predvideva testne meritve preko specifičnega terena v različnih smereh. Program na podlagi podatkov ugotavlja odstopanja med posameznimi snemalnimi linijami in določi korekcijske faktorje, ki se upoštevajo med izvajanjem meritev. S testom se zagotovi popravke pri odstopanju pozicij, časa oz. latence ter kotnih popravkov (prečno - roll, vzdolžno - pitch in smer - yaw). Diagrami 3.2, 3.3 in 3.4 ponazarjajo stopnjo napake globine ali pozicije pri neujemanju kotnih popravkov.

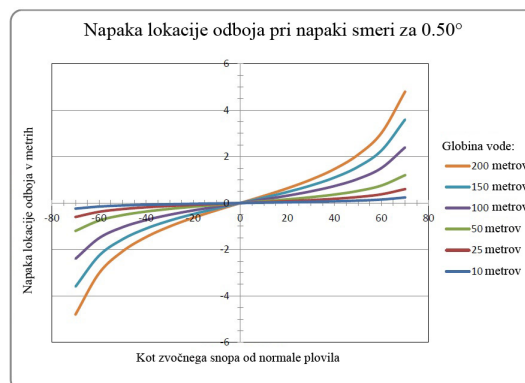
V podjetju imamo pridobljen certifikat ISO 9001, po katerem izvajamo kontrolo



Slika 3.2: Napaka pozicije globin zaradi odstopanja nagiba plovila.



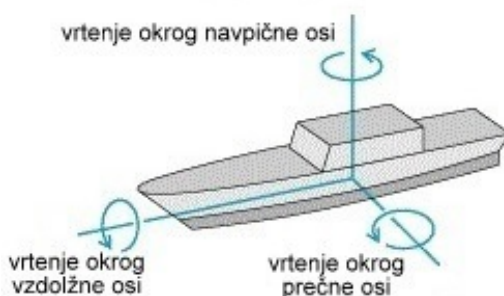
Slika 3.3: Napaka globine zaradi odstopanja naklona plovila.



Slika 3.4: Napaka pozicije pri odstopanju smeri za 0,5°.

sistema vodenja kakovosti, kjer so določeni termini za kalibracijo in umerjanje posameznih merilnih inštrumentov.

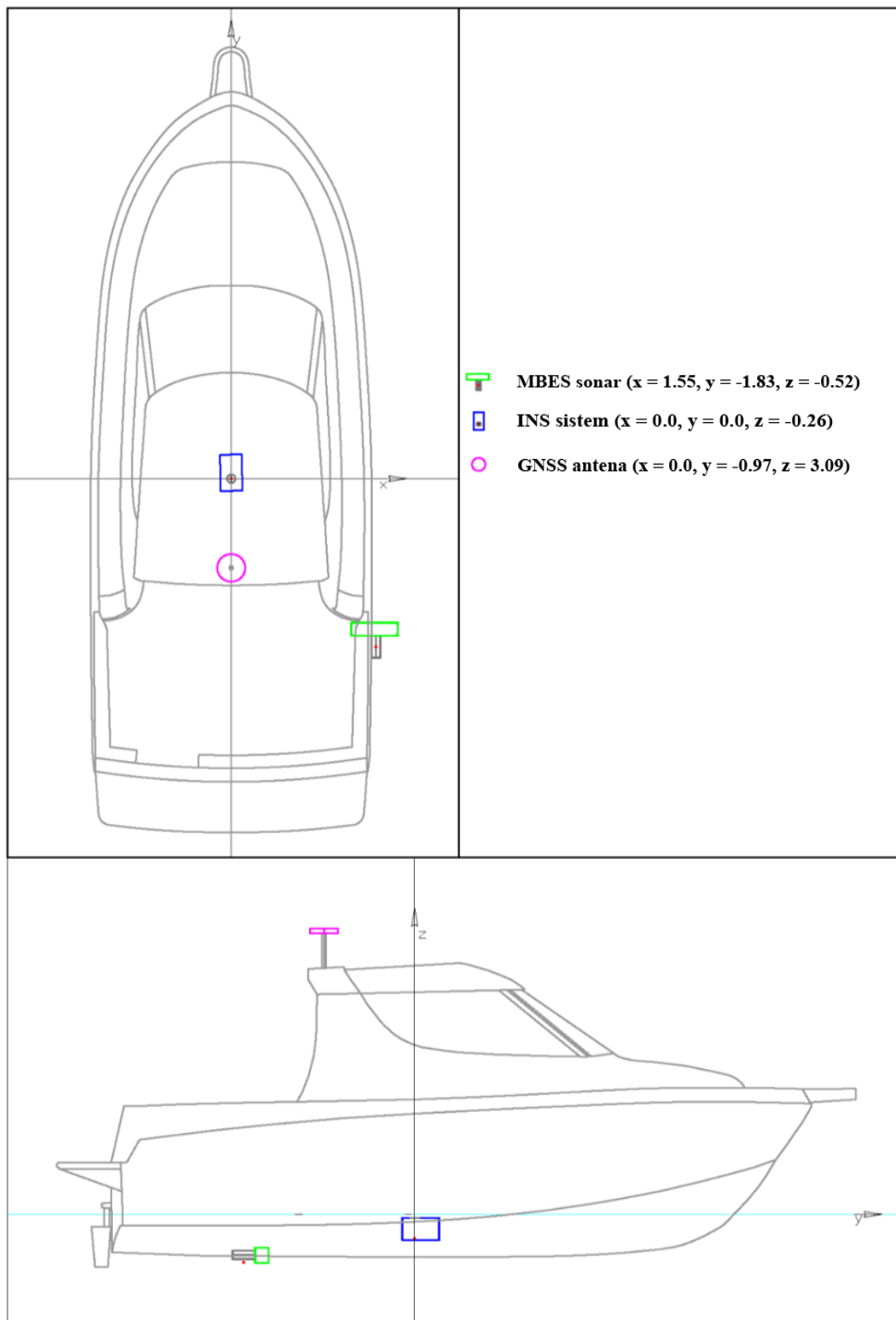
V postopku kalibracije opreme je potrebno za zagotovitev kar najbolj natančnih podatkov določiti začetna odstopanja kotov montaže sonde od glavnih osi inercialnega sensorja (slika 3.5): nagib okrog prečne (angl. pitch), vzdolžne (angl. roll) in navpične (angl. yaw) osi čolna, pa tudi časovni zamik med zabeleženimi podatki sonarja in položajem plovila, dobljenim iz GNSS sprejemnika. Prav tako je potrebno v program vnesti natančne lokacije merilnih naprav glede na lokalni koordinatni sistem plovila z izhodiščem v napravi TSS MAHRS (slika 3.6).



Slika 3.5: Rotacije okoli osi plovila.

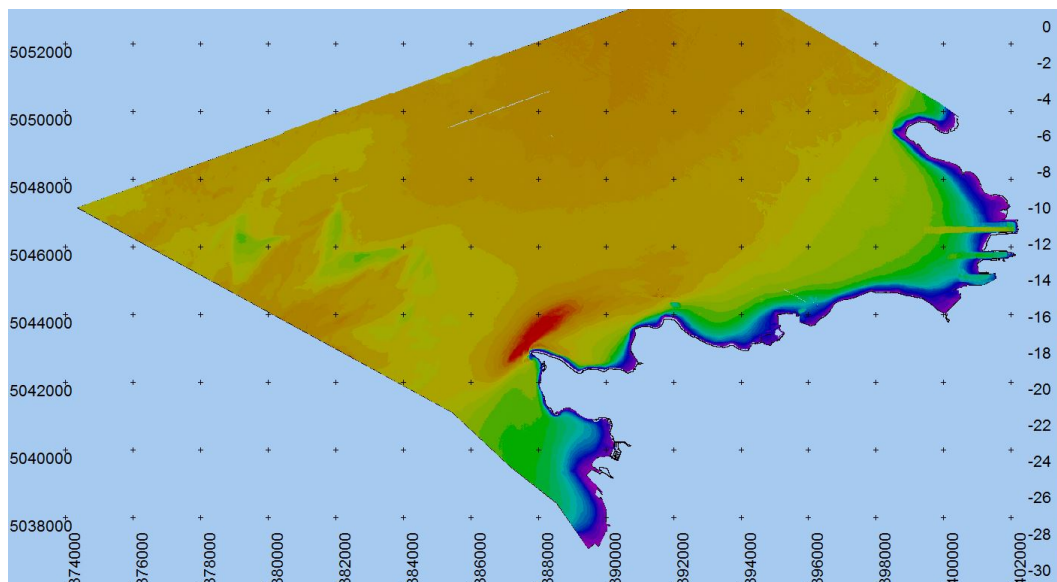
### 3.4 Izvedba meritev

Bistvenega pomena pri vseh meritvah, ki se izvajajo na morju, je natančno vnaprejšnje načrtovanje, saj na uspešnost izvedbe in kakovost rezultatov vplivajo številni dejavniki [62], [46]. Najprej je potrebno natančno opredeliti območje snemanja. Priprava na hidrografske meritve vključuje tudi določitev natančnega itinerarja, ki glede na navtične karte in ostale obstoječe podatke o globinah predvideva linije snemanja in morebitno potrebo po zamenjavi plovila in merilnih naprav v plitvih vodah. Itinerar vključuje tudi evidenco vseh objektov (npr. označevalne boje zaščitenih območij, ribogojnice, školjčišča, kopališča ...), ki zahtevajo dodatno prilagoditev načina in časa snemanja. Meritve na morju je potrebno izvajati v kar se da ugodnih vremenskih razmerah (čim manj vetra), saj lahko močno valovanje morja negativno vpliva na kakovost rezultatov. Pri meritvah v plitvih vodah morajo biti meritve časovno usklajene tudi s predvidenim plimovanjem. Predpogoja za uspešno opravljene meritve sta tudi delujoča, ustrezno nameščena ter kalibrirana merilna oprema in natančen navigacijski sistem. Glede na dostopnost signala je potrebno za posamezna območja snemanja določiti najustreznejši način prenosa podatkov, v našem primeru prek radijskih valov ali omrežja mobilne telefonije. Hitrost vožnje moramo prilagoditi vrsti meritev in razmeram na morju ter sprotno nadzorovati kakovost zajetih podatkov. Skozi celoten potek meritev je potrebno voditi zapisnike meritev, kamor si beležimo vremenske razmere, razmere na morju, morebitne težave z



Slika 3.6: Merilne naprave na plovilu Lyra (KP-5058).

merilno opremo ali pojav zanimivih struktur na posnetkih, ki nam pomagajo pri kasnejši obdelavi in interpretaciji podatkov.



Slika 3.7: Batimetrični model slovenskega morja. Globine so prikazane v metrih. Meritve je izvedlo podjetje Harpha sea, d.o.o. Koper med letoma 2006 in 2008.

Vnaprej načrtovane linije meritev se med samo plovbo prilagaja, da se doseže ustrezen preklap snemalnih pasov. Povprečna hitrost plovila med izvajanjem meritev je 5 vozlov (9 km/h). Sonar snema 60° levo in 60° desno od normale plovila, kar pri globini 20 metrov pomeni 70 metrov širok pas morskega dna. Če upoštevamo 20 % prekrivanja linij za kontrolo kakovosti meritev in 10 % porabe časa za obračanje plovila med linijami, sledi, da je možno v eni uri meritev posneti 50 hektarjev morskega dna. Pokrivenost se linearno zmanjšuje z manjšanjem globine. Dodatno je potrebno upoštevati več izgube časa pri manevriranju v plitvih obalnih predelih, vendar je hitrost zbiranja podatkov še vedno zelo velika v primerjavi z vizualnimi in mehničnimi metodami kartiranja morskega dna. Na sliki 3.7 je prikazan model batimertij celotnega slovenskega morja. Meritve je izvedel hidrograf z mednarodno licenco iz podjetja Harpha Sea, d.o.o. in ustrezajo kriterijem *Special order* meritev po standardih mednarodne hidrografske organizacije IHO [62]. Meritve so bile opravljene med letoma 2006 in 2008. Uporabljala se je oprema opisana v tem poglavju. Skupna količina zajetih podatkov mnogospopnega sonarja je 1070 GB in je bila zajeta v 171 merilnih dneh. Pokrito je bilo približno 205 km<sup>2</sup> veliko območje, hitrost vožnje med snemanjem pa se je, glede na razmere, gibala med 3 in 5 vozli.

Poleg uporabe podatkov iz omenjenih meritev smo za potrebe naše raziskave izvedli še veliko dodatnih meritev. Le-te so bile večinoma namenjene spoznavanju delovanja sonarja in vpliva različnih nastavitvev (moč oddajnika, občutljivost sprejemnika itd.) na akustične podatke.



## Poglavje 4

# Zbiranje podatkov o morskem dnu

Pri zbiranju podatkov o morskem dnu smo se oprli na doktorsko raziskavo dr. Mojce Poklar [110]. Poglavji o fotointerpretaciji letalskih posnetkov in video snemanju morskega dna sta povzeti po delih [75] in [110] omenjene avtorice in skupni raziskavi [111].

### 4.1 Točkovno vzorčenje s potapljači

Pri izvedbi točkovnega vzorčenja smo sodelovali s profesionalnimi potapljači na vdih šole potapljanja Apnea Center 2000, ki je podružnica mednarodne organizacije Apnea Academy [105]. Izvedba potopa na vdih je veliko hitrejša od klasičnega potopa z jeklenkami, poleg tega niso potrebni postanki za dekompresijo. Prednost je tudi boljša mobilnost zaradi bistveno manj opreme. Vse naštetu omogoči hitrejšo vzorčenje in večjo prilagodljivost.

Točke vzorčenja smo vnaprej določili in pripravili s pomočjo programa za načrtovanje hidrografskih meritev PDS2000 [126]. Za točkovno vzorčenje smo uporabljali plovilo Lyra in manjši gumenjak, dolžine 4.5 metra. Obe plovili sta opremljeni z natančnimi GNSS sistemi za določanje položaja plovila. Točko vzorčenja smo označili z utežjo in na označenem mestu izvedli potop. Opažanja potapljača je operater sproti vpisoval v zapisnik, potapljač je morsko dno na točkah vzorčenja fotografiral in posnel. Za podvodno snemanje smo uporabljali podvodno kamero Midland XTC-280 [91] in podvodni fotoaparatus Olympus Tough TG-2 [58].

### 4.2 Letalski posnetki

Primer letalske fotografije je slika 4.1. Pojem aerofotografije se nanaša na pridobivanje fotografij zemeljskega površja iz dvignjene pozicije in spada pod širšo disciplino daljinskega zaznavanja [99]. Slednje uporablja optične informacije, ki so lahko pridobljene iz zračnih (aerofotografija) ali satelitskih posnetkov. Fotografiranje iz zraka je močno orodje za prepoznavanje habitatov znotraj svetlobnega območja. Metoda je primerna za pridobivanje nekaterih vrst podatkov o bentoških habitatih, za druge



Slika 4.1: Letalska fotografija območja med Izolo in Koprom.

pa nekoliko manj. Na kopnem so optične metode zelo uporabne, v morju pa jih omejuje slaba penetracija svetlobe. Povprečna Secchijeva globina za slovensko morje je 8 metrov [106], optične metode pa so uporabne še nekoliko plitveje. Kljub temu lahko ob natančnem opazovanju in poznavanju dinamike motnosti morske vode ob pravem času uporabimo optične metode za določanje tipov morskega dna.

Pri aerofotografiji poteka pridobivanje podatkov o morskem dnu s fotografiranjem iz zraka. Preko obravnavanega območja se opravi prelet z letalom z namenom seznanitve z dimenzijami raziskovanega območja, določitve natančnosti, potrebne za posamezno nalogo, in opredelitve morebitnih težav, ki se utegnejo pojaviti. Namen preleta je tudi iskanje razpoznavnih točk za določanje pozicij za prostorsko usklajevanje (georeferenciranje) zračnih posnetkov z zemljevidi.

Postopek pridobivanja fotografij je sestavljen iz dveh delov in sicer nastavitve projektnih zahtev pridobivanja fotografij in obravnavanja okoljskih vidikov. Medtem ko prvi del vključuje določitev tehnični zahtev glede letala, fotoaparata, linij preletov in prekrivanja ter merila, drugi del obravnava določitev optimalnih okoljskih pogojev (prosojnost/motnost vode, plimovanje, veter in valovanje, kot sončnih žarkov) za uspešno izvedbo pridobivanja fotografij [36].

Za izvajanje fotografiranja iz zraka je primerna večina letal z enim ali dvema motorjema. Višina leta je poljubna, odvisna od zahtevane natančnosti. Na telo letala je pritrjen ustrezen fotoaparatus, tako da krila letala ne ovirajo fotografiranja. Glede na tehnične zahteve fotoaparata in višino leta oziroma zahtevano natančnost fotografij, se načrtuje linije preletov obravnavanega območja. Pri tem je pomembno tako čelno (prekrivanje sosednjih fotografij vzdolž linije preleta) kot bočno prekrivanje fotografij (prekrivanje sosednjih fotografij med vzporednimi linijami preleta). S 60 % čelnim prekrivanjem je olajšana interpretacija osrednjega dela fotografij, poleg tega pa se izravna izguba kritja zaradi odseva sonca na gladini morja. V primerih, ko se pričakuje, da bo sonce povzročalo težave zaradi odseva, se čelno prekrivanje poveča tudi na 80 %. Bočno prekrivanje 30 % zagotavlja neprekinjeno pokritost fotografij sosednjih linij preletov [36].

Ker je metoda v veliki meri odvisna od okoljskih pogojev, je značilnosti obravnavanega območja potrebno natančno poznati. Te vključujejo splošne značilnosti morskega dna, potencialno prisotne bentoške skupnosti, rastlinske vrste in njihove



morfologijo in fenologijo, globinske razpone in lokacije poznanih habitatov, sezonskost prosojnosti/motnosti morske vode, vremenske razmere, dnevne vzorce smeri in hitrosti vetra, kot sončnega sevanja preko dneva ter dnevne in sezonske režime plimovanja. Čas fotografiranja mora biti izbran glede na našete značilnosti, da tako dobimo kar najbolj optimalno vidljivost na fotografijah. Bistveno je, da se fotografije zajame ob oseki, kot sončnih žarkov pa mora biti večji od  $35^\circ$ . Pomembno je tudi, da zračno fotografiranje poteka ob sončnem vremenu, saj veter in oblaki vplivajo na poslabšano vidljivost. Veter z ustvarjanjem valov in dviganjem sedimenta zamegli oziroma prikrije morsko dno, medtem ko oblaki ustvarjajo sence na morsk gladini, ki jih lahko zamenjamo za morske travnike ali kak drug tip morskega dna. Kartograf lahko pridobljene fotografije interpretira subjektivno (pri tem gre le za njegovo oceno) ali s pomočjo računalniških programov [89]. Uporabljajo se predvsem različne tehnike obdelave in segmentacije fotografij [125].

Metoda je uporabna za obalna morja in zalive, ki so plitvejši, saj na fotografijah ne more zaznati morskega dna na večjih globinah. Večinoma se tudi uporablja za večja območja, saj omogoča hitro izmero, pri manjših območjih pa je uporaba relativno draga. Natančnost izdelanih kart tipov morskega dna iz zračnih posnetkov je odvisna od vrste in velikosti napake, prisotne pri zajemu, popravljanju in interpretaciji fotografij. Obseg in časovni razpored letalskih posnetkov lahko vplivata na napako. Napake se pojavljajo tudi v fazi interpretacije fotografij, saj je ta proces subjektiven.

#### 4.2.1 Opis uporabljene opreme in potek meritev

Fotografiranje iz zraka smo izvedli s profesionalnim fotoaparatom Nikon D700 s kalibriranim 50 mm objektivom z goriščnico 50 mm, kar najmanj popači sliko. Kljub temu smo objektiv še dodatno kalibrirali. Ločljivost fotoaparata znaša le 12 mega pikslov ( $4256 \times 2832$ ), vendar je od števila pikslov pomembnejša velikost in kakovost senzorja. Čim večji je senzor, manjša je gostota pikslov in posledično je fotografija bolj kakovostna. Najbolj je to opazno pri višjih občutljivostih (ISO – enota za občutljivost). Izbrani fotoaparat (D700) se po velikosti senzorja (FX) uvršča v profesionalni razred. Višina leta je odvisna od zahtevane ločljivosti posnetkov, v danem primeru pa so bile fotografije zajete na višini 1000 m. Posamezna fotografija v tem primeru zajame površino  $700 \times 460$  m in horizontalna ločljivost posnetka znaša 18 cm. Zaradi morebitnih tresljajev v času letenja je bil čas ekspozicije nastavljen na najkrajši še sprejemljiv čas in sicer  $1/1000$  s. Da bi zagotovili ostre posnetke tudi ob tresljajih, smo povečali svetlobno občutljivost (ISO 3200) in zaslonko (F5.6). Fotoaparat je (preko USB) povezan s prenosnim računalnikom. Program CameraControl (Nikon) med samim snemanjem omogoča nadzor vseh nastavitev in prenos fotografij. Za delovanje sistema smo razvili še dodatne programske module. Za proženje fotoaparata poskrbi poseben programski modul Trigger. Omenjeni programski modul prilagaja čase proženja glede na višino in hitrost letala. V našem primeru je povprečni čas med posameznimi fotografijami 3,8 s. Pri tem pride do 60% prekrivanja fotografij v smeri leta. Lokacijo in orientacijo fotoaparata spremljamo s pomočjo globalnega navigacijskega satelitskega sistema (Global Navigation Satellite Systems: GNSS) in inercialnega senzorja premikov (inertial measurement unit: IMU), s kate-

rim izračunavamo orientacijo v prostoru. S tem dobimo natančno pozicijo letala ob zajetju fotografij, kar nam omogoča, da so tudi posnetki natančno postavljeni v prostor. Za navigacijo po načrtovanih preletih smo razvili programski modul *AirbornePhotogrametry*. Zaradi pomanjkanja prostora v letalu (Cessna 172) smo uporabili prenosni računalnik, ki ga ima operater kar v naročju. Dodatni manjši ekran s podatki o trenutnem horizontalnem in vertikalnem odmiku od idealne smeri in višine pa se nahaja pred pilotom. Celoten sistem je zasnovan za avtonomno delovanje in ga operater le nadzira.

### Priprava zračnih fotografij

Da bi lahko zračne fotografije ustrezno interpretirali in iz njih pridobili podatke o porazdelitvi tipov morskega dna, smo morali fotografijam izboljšati kontrast in jih georeferencirati. Georeferenciranje in rektifikacija zračnih fotografij vodnih teles je predstavljalo izziv za kartiranje, saj celo na področjih, kjer so tipi morskega dna jasno vidni in kartirani, lahko pomanjkanje uporabnih nadzornih točk, ki fotografijo umestijo v prostor, resno ovira prizadevanja kartiranja morskih travnikov. Lokacijo fotografije smo tako pridobili iz druge, že georeferencirane fotografije ali kartografske podlage. Ker nismo imeli obstoječega vira nadzornih točk, smo po pregledu fotografij izbrali nadzorne točke, ki smo jih porazdelili čim bolj enakomerno. Nadzorne točke smo postavili na tista mesta, za katere je malo verjetno, da se bodo s časom spremenila. Fotogrametrične orientacijske rešitve zahtevajo najmanj tri točke za vzpostavitev modela, vendar smo izbrali še vsaj nekaj dodatnih točk (pet ali šest), da se omogoči zavrnitev točk, ki ne prispevajo k sprejemljivi rešitvi.

### Kartiranje tipov morskega dna iz zračnih posnetkov

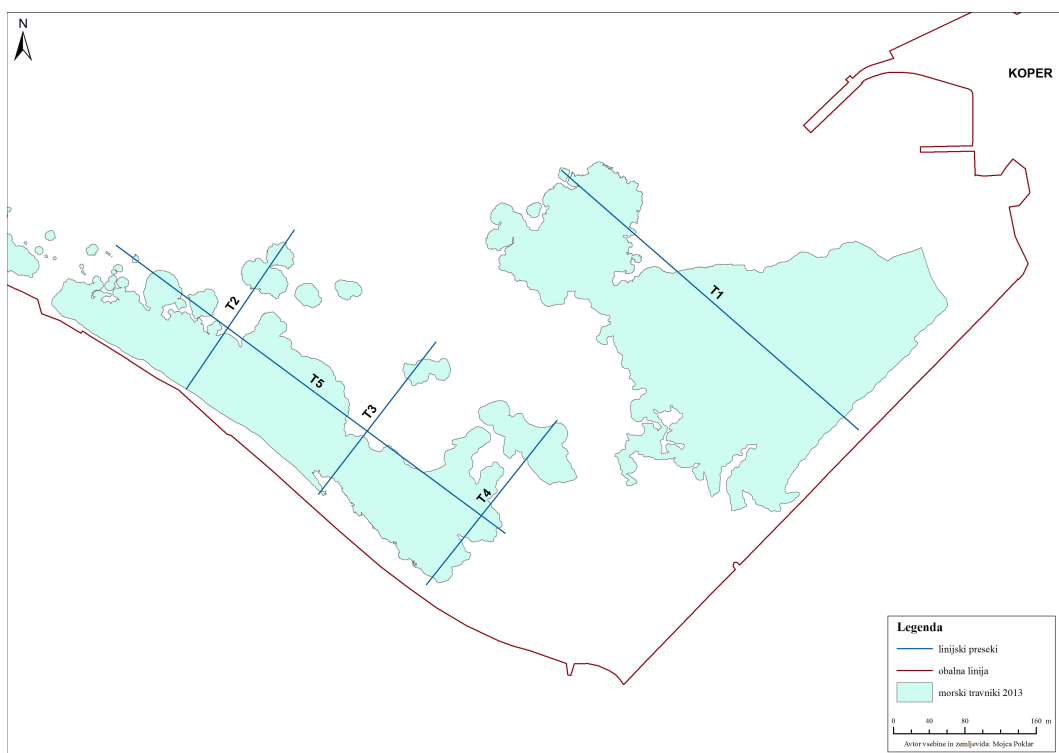
Da bi iz zračnih fotografij določili tipe morskega dna, smo vizualno ovrednotili temeljne elemente interpretacije fotografij (ton, barva, kontrast, tekstura, sence itd.), kar prispeva k veliki subjektivnosti metode. Po georeferenciranju zračnih fotografij in izboljšanju kontrasta, smo fotografije naložili v programski paket ArcGIS, kjer smo ustvarili vektorski sloj tipov morskega dna, tako da smo ročno označili posamezne področja. Na sliki 4.2 je prikazan primer zračne fotografije na kateri so z rdečo barvo obrobjeni morski travniki. Iz letalskih posnetkov so dobro vidne predvsem meje med morskimi travniki in muljem, kljub temu pa so se pojavljala območja, kjer iz fotografij ni bilo jasno ali gre za morski travnik ali ne. To še posebej velja za globlja območja, kjer je spodnja meja morskega travnika pogosto težko določljiva zaradi manjšega odstotka svetlobe, ki uspe prodreti do morskega dna oziroma do morskega travnika.

## 4.3 Video snemanje

Za preverjanje porazdelitve tipov morskega dna smo uporabili navpične linijske preseke, ki so pravokotni na obalo ter vodoravne, ki so z njo vzporedni. Na sliki 4.3) je primer s štirimi navpičnimi linijskimi preseki (T1, T2, T3, T4) in enim vodoravnim T5, postavljenim vzporedno z obalo s povprečno oddaljenostjo 100 metrov od obale.



Slika 4.2: Primer kartiranja morskih travnikov iz letalskih posnetkov. Vir: [111]



Slika 4.3: Primer razporeditve linijskih presekov za določanje tipov morskega dna. Vir: [110]

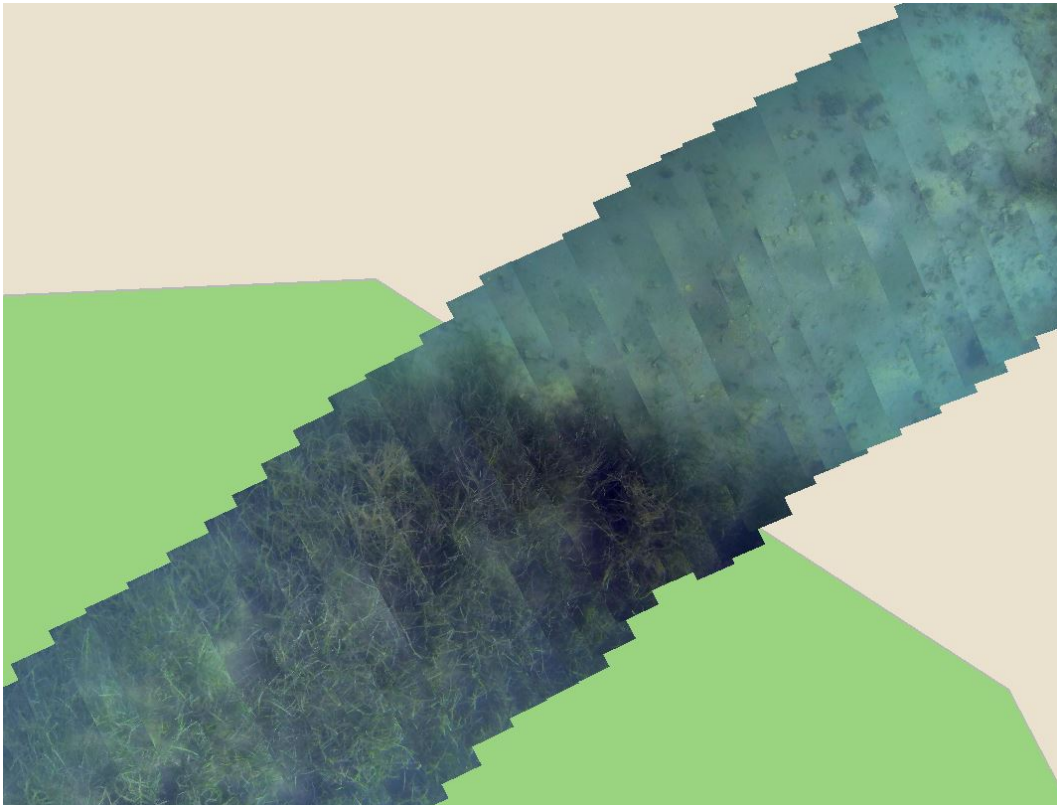
Preseki so bili izrisani le računalniško, saj za snemanje stanja levo in desno od preseka nismo uporabili metode s potapljači, temveč smo presek posneli neposredno s plovila. Podvodna snemanja smo izvedli z GoPro Hero3 black edition [51] kamero, ki še zdaleč ni profesionalna kamera, vendar smo se zanjo odločili zaradi nekaterih njenih dobrih lastnosti, kot so zelo širok zorni kot ( $170^\circ$  v zraku in  $128^\circ$  v morski vodi) ter kakovostno podvodno ohišje. Snemanja smo izvedli vzporedno z meritvijo mnogosnopnega sonarja. Pozicijo kamere na nosilcu smo predhodno natančno izmerili in jo vnesli v program (PDS) kot referenčno točko. Na ta način smo lahko z naknadno obdelavo podatkov iz GNSS sprejemnika in inercialnega senzorja določili natančno pozicijo (x,y,z) in kote (nagib, naklon, usmerjenost) vsakega posnetka. GoPro kamera žal nima možnosti sinhroniziranega proženja, zato je bila glavna težava sinhronizacija fotografij. Specifikacije kamere navajajo zaporedno proženje v intervalu 0.5 s, a se je izkazalo, da je ta interval zelo nenatančen in zato neuporaben za sinhronizacijo. Zato smo se odločili za uporabo video načina. Sinhronizacijo smo izvedli tako, da smo na začetku in koncu snemanja posneli zelo natančno uro (natančnost ms, sinhronizirana z GNSS). Na ta način smo uspeli zagotoviti časovno sinhronizacijo, žal smo pa s tem veliko izgubili na ločljivosti. V foto načinu kamera zmora ločljivost  $4000 \times 3000$ , v video načinu (2.7K) pa le  $2716 \times 1524$ , 30 okvirjev (frame) na sekundo. Kljub temu to zagotavlja še dovolj dobro ločljivost posameznih izrezanih okvirjev (frame). Uporabili smo 64 GB spominsko kartico, a se je izkazalo, da je omejujoči dejavnik čas delovanja baterije kamere in je tako najdaljši čas snemanja približno 2 uri.

Video posnetek smo računalniško pretvorili v fotografije in jim izboljšali kontrast s pomočjo programa ImageMagick [28]. Dodatno težavo pri obdelavi fotografij podvodne kamere predstavlja geometrijsko popačenje na robovih. Geometrijsko popačenje je tako veliko, da ga z običajnimi postopki ni bilo mogoče izničiti. Zaradi visoke frekvence snemanja (30 slik na sekundo) in nizke hitrosti med izvajanjem meritev se zaporedne fotografije v velikem delu prekrivajo. Prekrivanje zaporednih fotografij dosega preko 80 %. Iz vsake fotografije smo izrezali le ozek (20 % višine fotografije) sredinski del in tako skoraj v celoti izničili vpliv geometrijskega popačenja. Primerjali smo razdalje centroidov zaporednih fotografij in glede na razdaljo med fotografijami odrezali fotografijo tako, da se le-te še minimalno prekrivajo. Fotografijam smo nato obdelali z modulom za izboljšavo fotografij in jih združili.

Natančno prostorsko umestitev fotografij smo izvedli z enakimi orodji (GNSS sprejemnika in inercialni senzor) in z enakim postopkom kot za hidrografske podatke. Postopek je podrobneje opisan v poglavju 3.1. GNSS sprejemnik na plovilu zagotavlja podatke s frekvenco 5 Hz. Z isto frekvenco (vsake 0,2 s) programski paket PDS2000 omogoča izvoz podatkov za nosilec kamere in sicer pozicijo (x, y, z), kote (nagib, naklon, usmerjenost) in globino dna pod kamero. Izvoženi podatki so v formatu CSV. Razvili smo programski modul, ki v CSV datoteko doda podatke o nazivih fotografij, popravljene koordinate in orientacije centroidov fotografij. Podatke dodajamo na podlagi časovne sinhronizacije. Fotografije smo nato pretvorili v GeoTIFF format in jih umestili v prostor. Pretvorbo smo opravili na podlagi prostorskih podatkov v CSV datoteki in z uporabo GDAL razširitve za programski jezik Python. Tako obdelane fotografije so natančno umeščene v prostor in pripravljene za prostorske analize. Za prostorske analize in interpretacijo posnetkov smo uporabili

programa Qgis in ArcGis ter enake tehnike kot pri interpretaciji aerofotografij. Na sliki 4.4 so prikazane združene fotografije transektov. Zelena barva označuje območje morskoga travnika, bež barva pa območje muljnatega dna. Na združenih fotografijah transektov je razvidna meja med morskimi travniki in muljnatim dnom.

Vse omenjene tehnike obdelave video posnetkov in georeferenciranja fotografij smo razvili in implementirali v podjetju Harpha sea, d.o.o. Koper v sklopu del [75], [110], [111] in [93].



Slika 4.4: Georeferencirana fotografija transektov, generirana iz videoposnetka transektov s plovila.



## Poglavje 5

# Predobdelava podatkov

Namen poglavja je predstavitev razvitega sistema in metodologije za obdelavo in avtomatsko predobdelavo velikih količin podatkov, pridobljenih z meritvami mnogospnega sonarja. V poglavju je opisana celotna metodologija predobdelave podatkov mnogospnega sonarja, od surovih zajetih podatkov do podatkov, pripravljenih za akustično klasifikacijo morskega dna.

Izpostaviti je potrebno dva pomembna koraka predobdelave akustičnih podatkov - izločanje vpliva nastavitve sistema in uravnavanje vpliva kotnih odvisnosti. Izločanje vpliva nastavitve sistema je pomembno predvsem zaradi širše uporabnosti razvite metodologije akustične klasifikacije morskega dna. Meritve z mnogospnimi sonarji se v večini primerov izvajajo z namenom merjenja globin. Za natančno izmero globin mora hidrograf med meritvami prilagajati nastavitve sonarja razmeram na morju in globini morja. Spremenjene nastavitve, kot bomo videli v nadaljevanju, bistveno spremenijo intenzitete akustičnih odbojev. Izločanje vpliva nastavitve sistema je torej nujno, da lahko tudi take meritve uporabljamo za namene akustične klasifikacije morskega dna in da lahko še naprej delamo kakovostne batimetrične meritve in hkrati dobimo uporabne podatke za klasifikacijo. Izločanje vpliva nastavitve sistema je pomembno tudi zaradi velike količine starih podatkov meritev, ki bi jih lahko ob primerni predobdelavi uporabili za klasifikacijo morskega dna. Drugi pomemben korak je uravnavanje vpliva kotnih odvisnosti. Vpliv vpadnega kota na akustične podatke je zelo velik. Predstavili in ovrednotili bomo nekaj pristopov k reševanju tega problema.

### 5.1 Opis podatkov

Surovi podatki mnogospnega sonarja so shranjeni v formatu XTF. Format XTF (eXtendend Triton Format) je bil razvit z namenom shranjevanja veliko različnih tipov podatkov hidrografskih meritev. Format je enostavno razširljiv in se mu lahko v prihodnosti dodaja nove tipe podatkov, ne da bi to vplivalo na združljivost za nazaj. V XTF datotekah so shranjeni akustični podatki, podatki GNSS sistema in podatki INS senzorja.

MBES (mnogospni sonar) beleži veliko podatkov o nastavitvah sistema, razmerah med meritvami in akustičnih odbojih. Spodaj je naštetih nekaj podatkov, ki

so za nas pomembni pri akustični klasifikaciji morskega dna.

- Podatki o kakovosti meritev:
  - Status odboja (Beam Status)
  - Število zavrnjenih odbojev (Beams Rejected)
  - Stopnja zaupanja odboju (Beam uncertainty)
  - Ocena vertikalne napake (Vertical error)
  - Ocena horizontalne napake (Horizontal error)
- Lastnosti in nastavitve MBES sistema:
  - Frekvenca (Frequency)
  - Kot oddajnika (Projector angle)
  - Tip oddajnika (Projector Type)
  - Vzdolžna širina snopa sprejemnika (Tx Beam width along)
  - Prečna širina snopa sprejemnika (Tx beam width across)
  - Širina pulza (Ty pulse width)
  - Rotacija glave sonarja (Head rotation)
  - Širina snopa oddajnika (Projector beam width)
  - Parameter izgube zaradi širjenja signala (TVG Spread)
  - Parameter absorpcije akustičnega signala (TVG absorption)
  - Doseg (Range)
  - Oobčutljivost sprejemnika (Gain)
  - Moč oddajnika (Power)
- Podatki o razmerah in lastnostih odbojev:
  - Čas potovanja akustičnega signala (Two-Way travel times)
  - Globina (Depth)
  - Temperatura (Temperature)
  - Hitrost zvoka na površini (Sound velocity)
  - Globina pod plovilom (Nadir Depth)
  - Globina pod sonarjem (Depth below transducer)
  - Vpadni kot (Angle of incidence)
  - Usmerjenost (Azimuth angle)
- Podatki za georeferenciranje odbojev:
  - Višina (Heave)
  - Nagib (Pitch)



- Naklon (Roll)
- Usmerjenost (Heading)
- X koordinata odboja (Beam X)
- Y koordinata odboja (Beam Y)
- Z koordinata odboja (Beam Z)
- X koordinata plovila (Position x)
- Y koordinata plovila (Position Y)
- Z koordinata plovila (Position Z)
- Intenziteta odboja (Beam Intensity)

Podatki o kakovosti meritev služijo predvsem za preverjanje kakovosti meritev in izločanje slabih meritev, v kombinaciji z, v nadaljevanju opisanim, medianinim filtrom smo jih uporabili za izločanje slabih podatkov.

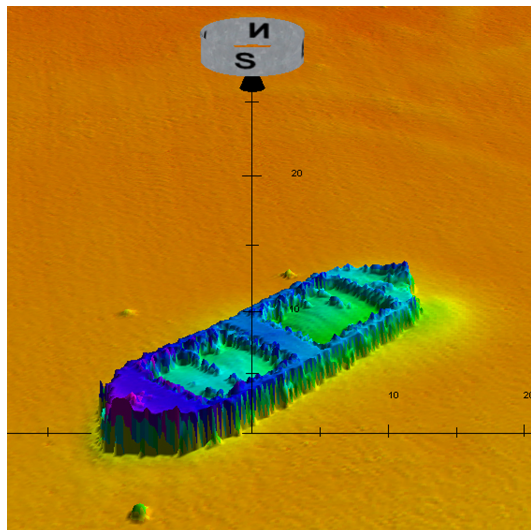
Lastnosti in nastavitve MBES sistema so pomembni, da vemo s kakšnim sistemom imamo opravka. Podatki o širinah pulza, frekvenci, nagibu sonde in rotaciji so v našem primeru fiksni, ker uporabljamo le en sistem. Kljub temu podatke ohranimo zaradi razširljivosti metodologije. Podatki pod oznakami »Range«, »Gain«, »Power« ter »TVG spread« in »TVG absorption« odražajo spremembe nastavitvev, ki jih opravlja hidrograf med meritvami. Za potrebe akustične klasifikacije morskega dna bi bilo bolje, da bi bila večina teh nastavitvev fiksna in bi torej lastnosti odbojev direktno odražale tip tal. Žal se večina MBES meritev izvaja z namenom merjenja globin, kjer pa se nastavitve prilagajajo, da se doseže ustrezna natančnost meritev. Za potrebe akustične klasifikacije moramo take podatke ustrezno predobdelati in izločiti vpliv nastavitvev sistema na podatke.

Iz podatkov o temperaturi vode, hitrosti zvoka v vodi in času potovanja akustičnega signala izračunamo kako daleč je signal potoval predno se je odbil. S tem dejansko izračunamo kako daleč je morsko dno od glave sonde. Nato z upoštevanjem vpadnega kota akustičnega signala, globine in pozicije sonde izračunamo relativno pozicijo odboja glede na plovilo.

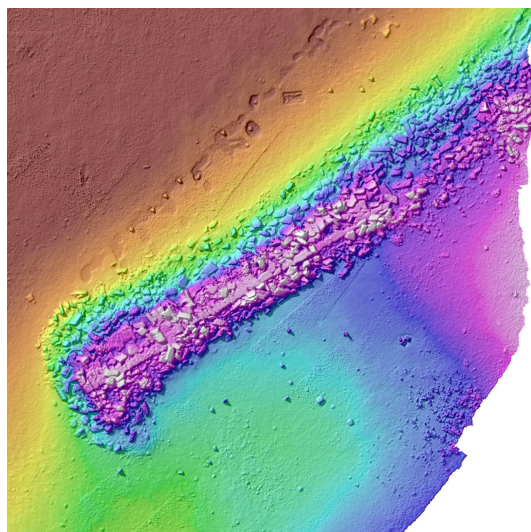
Podatki za georeferenciranje odbojev nam omogočajo georeferenciranje odbojev. MBES sistem shranjuje podatke o kotu sprejema odboja, času potovanja akustičnega signala, intenziteti odboja itd. Iz teh podatkov se najprej izračuna relativna pozicija odboja glede na plovilo, nato se z upoštevanjem koordinat (Position X,Y) in položaja plovila v času meritve (Heave, Roll, Pitch, Heading) izračuna natančne absolutne koordinate odbojev in ustrezne globine. Rezultat so georeferencirani podatki o odbojih akustičnega signala od morskega dna. Dobimo natančne batimetrične modele. Na sliki 5.1 je prikazan primer batimetričnega modela maone, potopljene na globini 23 metrov v Piranskem zalivu in na sliki 5.2 primer rimskega pomola v plitvini Simonovega zaliva.

## 5.2 Georeferenciranje podatkov

MBES sistem shranjuje podatke o kotu sprejema odboja, času potovanja akustičnega signala, intenziteti odboja, temperaturi vode ipd. Iz podatkov o temperaturi vode,



Slika 5.1: Stara maona, potopljena v Piranskem zalivu. Ločljivost batimetrije 20x20 cm.



Slika 5.2: Potopljen rimski pomol v Simonovem zalivu, Izola. Ločljivost batimetrije 10x10 cm.

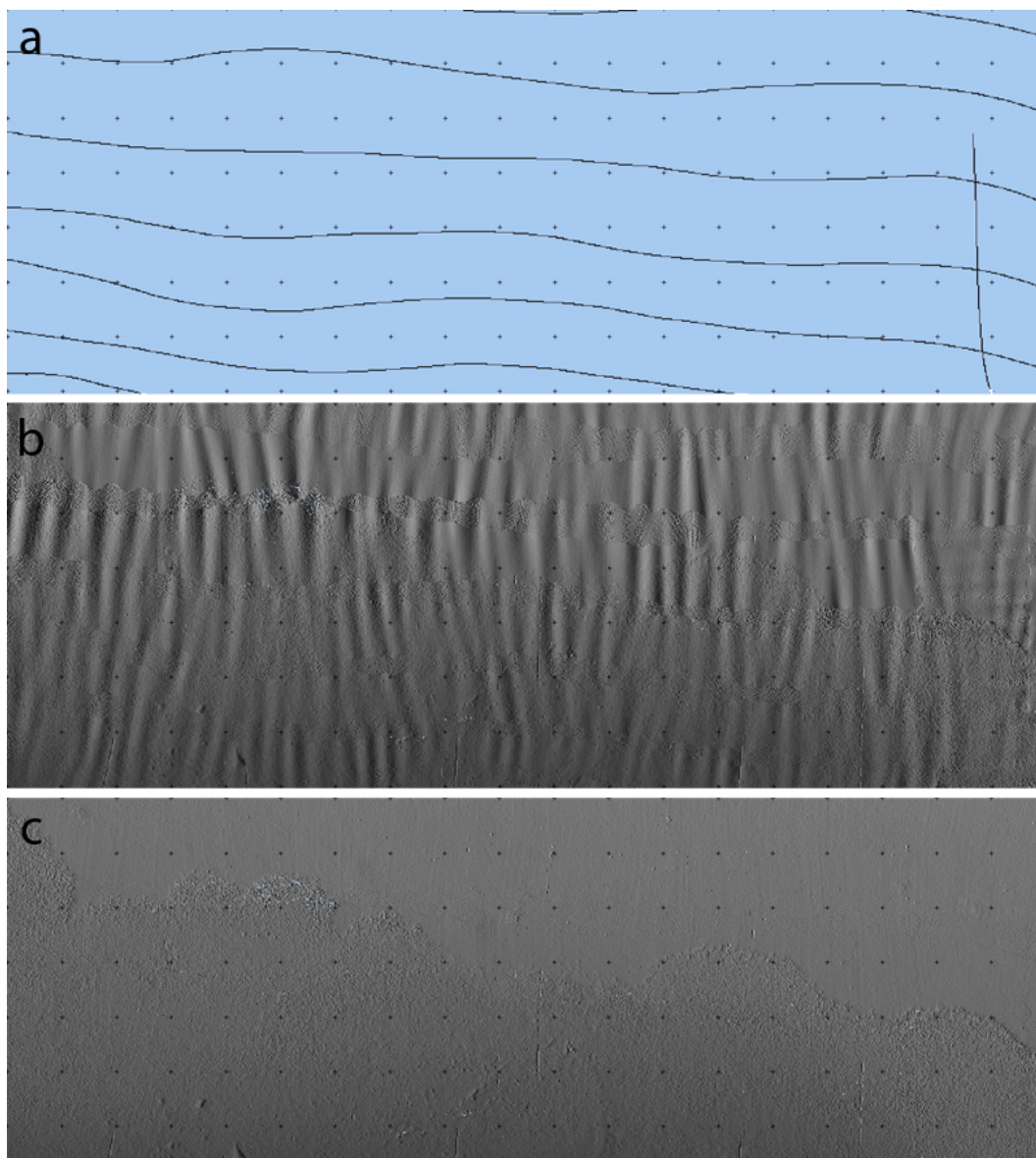
hitrosti zvoka v vodi in času potovanja akustičnega signala izračunamo kako daleč je signal potoval predno se je odbil. Do težave pride, ker temperatura vode v vodnem stolpcu, skozi katerega signal potuje, ni konstantna. Zaradi različnih temperatur vode se spreminja tudi njena gostota, kar povzroča lom zvoka. Efekt je podrobneje opisan v poglavju 2.3. Napakam se izognemo tako, da z napravo SVP (sound velocity profiler) izmerimo hitrosti zvoka v vodnem stolpcu in upoštevamo korekcije za variacije hitrosti zvoka v vodnem stolpcu, opisane v [87]. Tako natančno izračunamo oddaljenost morskega dna od glave sonde. Nato z upoštevanjem kota sprejemnika akustičnega signala in pozicije sonde na plovilu izračunamo relativno pozicijo odboja glede na plovilo.

Po izračunu natančnih relativnih pozicij odbojev je le-te potrebno preračunati v absolutne koordinate. Najprej vračunamo pozicijo plovila, ki jo dobimo in GNSS senzorja in nato smer, nagib in naklon plovila iz podatkov INS senzorja. Pri slednjih izračunih gre za preproste translacije in rotacije. Rezultat je natančen batimetrični model. Primer poteka georeferenciranja podatkov o globinah je prikazan na sliki 5.3. Na sliki 5.3a so prikazane linije poteka meritev, slika 5.3b prikazuje vpliv valovanja morja na podatke o globinah, na sliki 5.3c pa so prikazani končni georeferencirani podatki.

### 5.3 Preverjanje kakovosti

Pri razvoju metodologije in sistema za avtomatsko predobdelavo velikih količin podatkov, pridobljenih z meritvami mnogospopnega sonarja, smo se osredotočili predvsem na izločanje vpliva nastavitve sistema in uravnavanje odvisnosti odboja od vpadnega kota signala. Zelo pomemben dejavnik je tudi preverjanje kakovosti meritev in izločanje slabih podatkov. Običajno preverjanje kakovosti meritev in izločanje slabih podatkov izvajamo ročno s pomočjo izurjenih hidrografov. Surovi sonarski podatki, ki jih dobimo neposredno iz sonarja ob meritvi, še niso primerni za nadaljnjo uporabo. Podatke je potrebno obdelati, da so primerni za nadaljnjo uporabo. Prvi korak je preverjanje kakovosti in izločanje slabih podatkov. V podatkih se lahko pojavljajo razni šumi, ki so posledica odbojev, na primer od smeti v morju ali drugih plavajočih objektov. Čiščenje podatkov se izvaja v programu PDS2000. Izmerjene podatke je potrebno ročno pregledati po posameznih linijah meritev in iz njih odstraniti podatke, ki izstopajo, oziroma za katere je jasno vidno, da ne predstavljajo dna ali objektov na dnu. Slika 5.4 prikazuje program za ročno čiščenje sonarskih podatkov. Vidne so izstopajoče točke, nekatere tudi pod dnom, ki jih je potrebno odstraniti. Čiščenje podatkov je, kljub pomoči programa, v katerem poteka, (PDS2000) in raznim avtomatizacijam v njem, precej dolgotrajen postopek in lahko za bolj kompleksna območja, kjer se dno spreminja in je na njem več objektov, traja tudi še enkrat toliko časa kot meritve same. Očiščene podatke lahko nato pretvorimo v pravokotno mrežo zelene velikosti ali jih uporabljamo kot oblak točk. Pri izvajanju meritev je že med merjenjem upoštevan geoid za območje, na katerem se meritve izvajajo, tako so podatki, ki jih dobimo iz sonarja, umerjeni na geodetsko ničlo.

V postopku razvoja sistema za kartiranje morskega dna smo razvili modul za



Slika 5.3: Primer poteka georeferenciranja podatkov o globinah. Slika *a* prikazuje linije meritev, slika *b* podatke o globinah, georeferencirane samo z upoštevanjem GNSS pozicij in slika *c* podatke o globinah z izločenim vplivom valovanja morja.

avtomatsko čiščenje podatkov. Med meritvami občasno prihaja do motenj, prav tako pa vsi podatki niso enako zanesljivi. Najbolj vprašljivi so podatki na robu področja meritev posamezne linije, kjer so vpadni koti akustičnih signalov največji. Sam sistem beleži podatke o kakovosti odboja in kakovosti sprejetega signala ter tudi o morebitnih težavah GNSS sprejemnika in INS senzorja. Pri ročnem čiščenju podatkov se je izkazalo, da si lahko delo bistveno olajšamo z ustreznim načrtovanjem meritev. Med načrtovanjem meritev smo upoštevali preklap snemalnih pasov. Prekrivanje je uporabno zaradi doseganja višje kakovosti izmerjenih podatkov in tudi zaradi preverjanja meritev. S preklpom si zagotovimo dvojne podatke z različnim kotom vpada na najbolj kritičnih delih, kar nam omogoči preverjanje meritev. Za grobo izločanje slabih podatkov smo preizkusili veliko filtrov, ki uporabljajo statistične lastnosti odbojev. Za najbolj učinkovitega se je izkazal medianin filter [19]. S pomočjo medianinega filtra in z upoštevanjem medčetrtnskega razmika (razmik med četrtno najmanjših in četrtno največjih vrednosti) kot mere razpršenosti smo uspešno izločili slabe meritve in ohranili le kakovostne podatke. Kot je razvidno iz slike 5.4, prihaja pri akustičnih odbojih do velikih odstopanj. Filtri, ki uporabljajo povprečja in standardne odklone so zelo občutljiv na taka odstopanja. Z uporabo median namesto povprečja in medčetrtnskega razmika namesto standardnega odklona izničimo pretiran vpliv osamelcev na podatke. Primerjava rezultatov avtomatskega postopka čiščenja in podatkov, obdelanih s strani strokovnjaka, je prikazana na sliki 5.5. Na sliki 5.5a so prikazani neprečiščeni podatki. Slika 5.5b prikazuje podatke avtomatsko prečiščene z našim algoritmom, slika 5.5c pa ročno prečiščene podatke s strani hidrografa. Razvidno je, da ni bistvenih razlik med slikama 5.5b in 5.5c ter da so v avtomatsko prečiščenih podatkih izločene vse večje nepravilnosti. Učinkovitost je preverjal strokovnjak s področja hidrografije. Čiščenje podatkov hidrografskih meritev je v veliki meri odvisno od izkušenosti hidrografa, da zazna razliko med dejanskim odbojem od tal in odboji od smeti, jatami rib ipd. Cilj avtomatskega čiščenja podatkov ni bil 100 odstotno ujemanje s podatki, obdelanimi s strani hidrografa, pač pa grobo izločanje večjih napak s čim manjšo izgubo informacije. Po oceni hidrografa se je prirejen medianin filter najbolje izkazal s stališča izločanja slabih odbojev kot tudi s stališča ohranjanja informacije o strukturi morskega dna. V kombinaciji z medianimi filtrom smo uporabili medčetrtnski razmik kot mero razpršenosti.

Algoritem z uporabo median in medčetrtnskega razmika predstavljamo s psevdokodo Algoritem 1. Število snopov in širino okna smo določili empirično v sodelovanju s hidrografi. Na vsakem koraku postopka čiščenja smo izračunali medčetrtnski razmik sosednjih 50 snopov. Če so vrednosti obravnavanega snopa od mediane sosednjih snopov odstopale za več kot dvojno vrednost medčetrtnskega razmika smo podatke izločili. Opisani postopek smo najprej uporabili za celoten snop in nato še za vsakega izmed 240 odbojev v snopu. V tretji zanki algoritma smo enak postopek uporabili še znotraj snopa, opazovali smo odstopanje od mediane sosednjih 30 odbojev v istem snopu.

Pri obdelavi podatkov o priobalnem pasu slovenskega morja z našim postopkom se je izkazalo, da s filtrom izločimo med 0.5 % in 1 % podatkov. Več slabih podatkov se pojavlja v plitvem in razgibanem terenu, medtem ko je najmanj slabih podatkov na globljem, monotonem terenu. Na odprtem morju z globlinami okoli 20 metrov

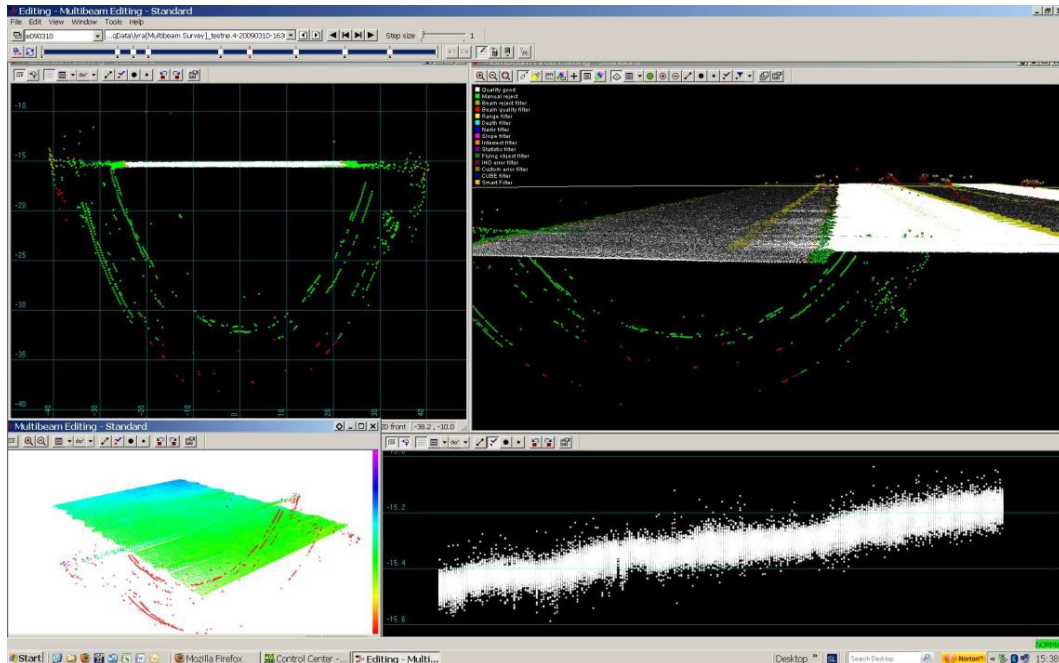
---

**Algoritem 1** Čiščenje podatkov

---

```
1: procedure ČIŠČENJE_PODATKOV(linija)
2:   Število_snopov = 50
3:   Širina_okna = 30
4:   for vsak snop v liniji meritev do
5:     primerjaj snop s sosednjimi Število_snopov snopi
6:     if odstopanje > 2× medčetrtnski razmik then
7:       označi snop a
8:     end if
9:     for i = 1 do 240 do ▷ 240 odbojev v snopu
10:      primerjaj odboj z mediano Število_snopov odbojev iz prejšnjih snopov
11:      if odstopanje > 2× medčetrtnski razmik then
12:        označi odboj i
13:      end if
14:    end for
15:    for i = 1 do 240 do ▷ 240 odbojev v snopu
16:      primerjaj odboj s Širina_okna sosednjimi odboji v istem snopu
17:      if odstopanje > 2× medčetrtnski razmik then
18:        označi odboj i
19:      end if
20:    end for
21:  end for
22: end procedure
```

---



Slika 5.4: Prikaz orodja za ročno preverjanje kakovosti sonarskih podatkov. Točke, ki izstopajo iz ravnine, so napake, ki jih mora hidrograf ročno izločiti iz podatkov.

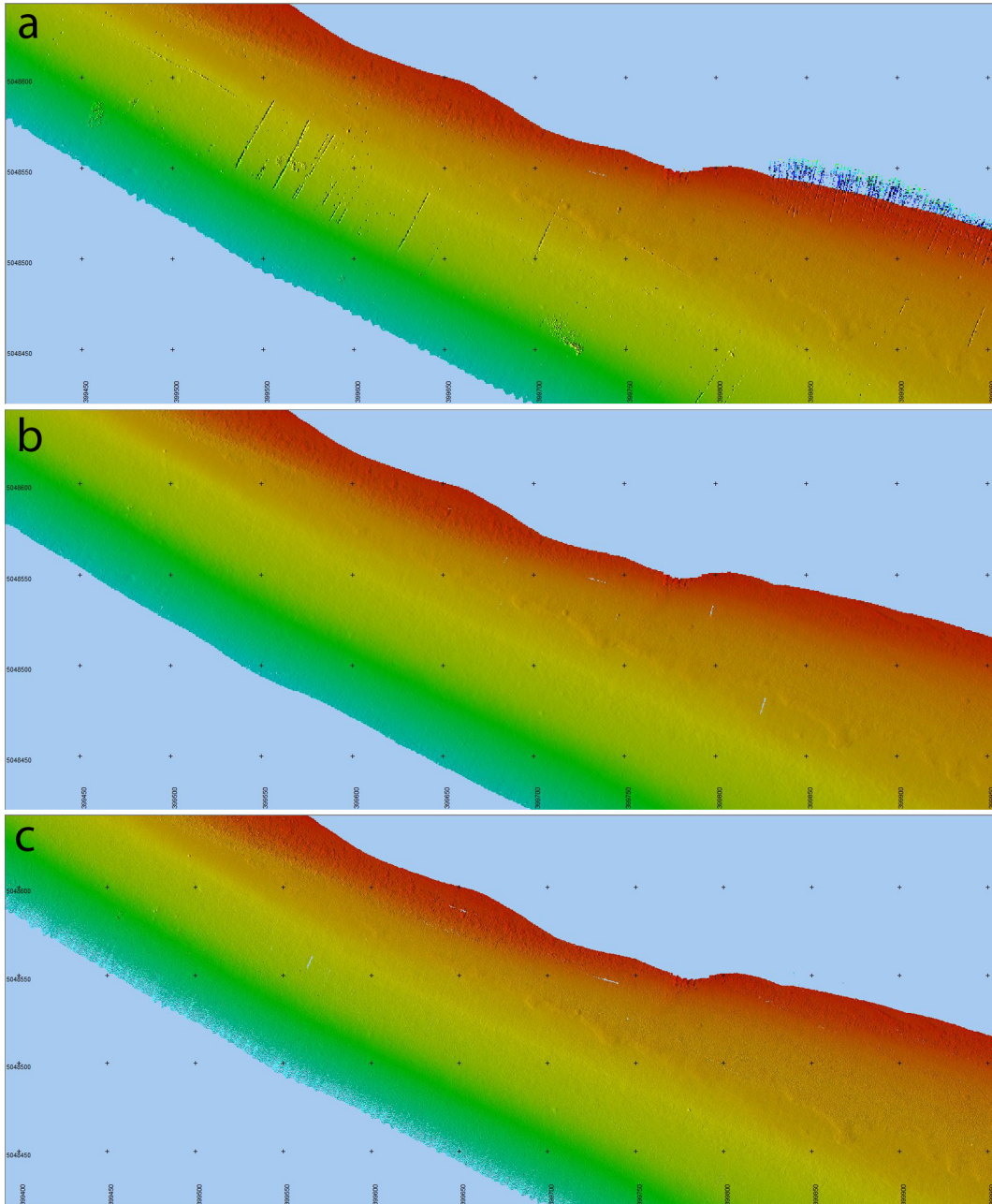
je slabih podatkov le okoli 0.1 %, medtem ko se je najvišja vrednost slabih podatkov, kar 2.9 % pojavila pri obdelavi meritev priobalnega območja med Piranom in Bernardinom. Za to območje je značilen zelo razgiban teren z umetno narejenim nasipom, ter velike variacije globin.

## 5.4 Izločanje vpliva nastavitve sistema

Podatki o občutljivosti sprejemnika, moči oddanega pulza, količniku predvidene absorpcije zvoka v vodi in predvidenih izgub moči signala so pomembni, ker odražajo spremembe nastavitve, ki jih opravlja hidrograf med meritvami. Za potrebe akustičnega kartiranja morskega dna bi bilo bolje, da bi bila večina teh nastavitve stalna in bi torej lastnosti odbojev direktno odražale tip tal. Večina meritev z mnogospornimi sonarji se izvaja z namenom merjenja globin, kjer se nastavitve prilagajajo, da se doseže ustrezna natančnost meritev na določeni globini morja. Za potrebe akustičnega kartiranja smo take podatke ustrezno obdelali in izločili vpliv nastavitve sistema nanje.

### 5.4.1 Moč oddajnika in občutljivost sprejemnika

Prvi korak predobdelave podatkov je izločitev vpliva nastavitve moči oddajnika in občutljivosti sprejemnika. Uporabili smo formulo:



Slika 5.5: Primerjava rezultatov ročnega in avtomatskega čiščenja podatkov. Slika a prikazuje surove podatke, slika b podatke obdelane z medianinim filtrom, slika c pa podatke obdelane s strani hidrografa v orodju PDS2000.



$$I_{norm} = \frac{I}{PG_o^2G_p^2}$$

Kjer je  $I$  intenziteta oddanega pulza,  $P$  moč oddajnika,  $G_o$  sistemska občutljivost sprejemnika in  $G_p$  programsko nastavljiva občutljivost sprejemnika. Manjše težave nastanejo v trenutku spremembe nastavitvev. Ko operater na plovilu spremeni nastavitvev Power ali Gain se to takoj zabeleži v podatkih, vendar traja nekaj časa da sistem dejansko prilagodi nastavitve. Če dosledno upoštevamo podatke o nastavitvah sistema pride do motenj v prehodu iz ene nastavitve na drugo. Težavo smo rešili tako, da smo interpolirali odboje od trenutka spremembe nastavitvev do trenutka stabilizacije sprejemnika oziroma oddajnika. Napaka ni tako pomembna, saj gre le za do največ tri problematične snope, ki bi jih lahko brez večje izgube natančnosti tudi izločili iz podatkov. Slika 5.6 prikazuje spremembe nastavitvev moči oddajnika med meritvami in vpliv na surove podatke. Na sliki 5.7 so prikazani popravljeni podatki po korakih, točka a prikazuje surove podatke, točka b pa podatke z izločenim vplivom nastavitvev sistema.

#### 5.4.2 Absorpcija in sferično širjenje zvoka

Izločanje vpliva nastavitve TVG (izgube moči signala med potjo do morskega dna in nazaj zaradi absorpcije in sferičnega širjenja signala) je naslednji pomemben korak predobdelave podatkov. Zvočni signal potuje skozi morsko vodo od oddajnika do morskega dna in nazaj do sprejemnika. Od opravljene poti je odvisno koliko signala se bo absorbiralo v morski vodi in koliko razpršilo v okolico.

MBES sistem sam računa približek TVG po formuli [126]:

$$TVG = 2\alpha R + Sp \text{Log}_{10} R$$

kjer  $R$  pomeni razdaljo do točke odboja, parametra  $\alpha$  in  $Sp$  pa nastavi uporabnik. Izračun po navedeni enačbi je dovolj dober za globine, ni pa dovolj natančen, da bi dobili ustrezne intenzitete odbojev. Iz sonarskih podatkov moremo najprej izločiti vračunan približek TVG. To naredimo po enačbi:

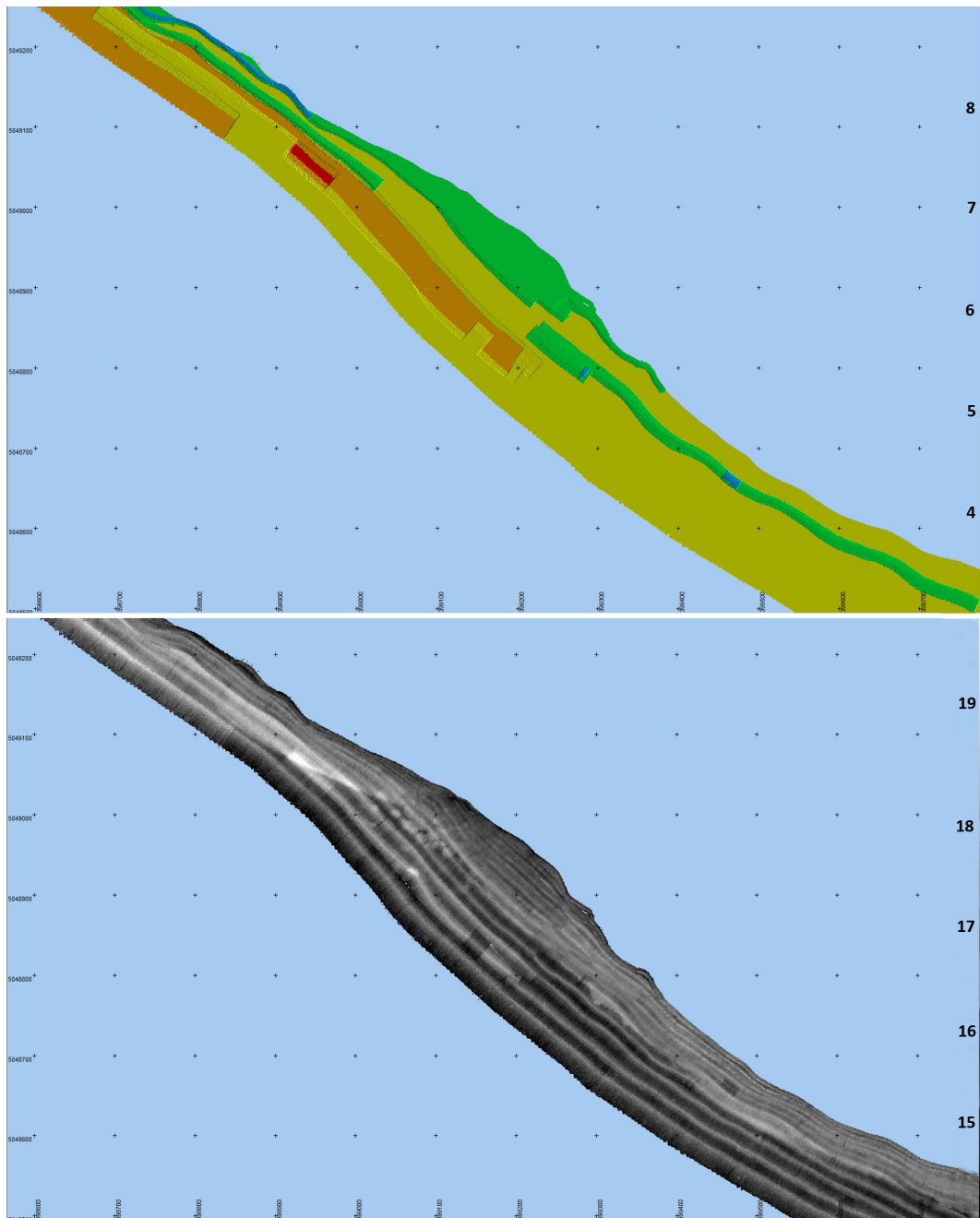
$$I_{noTVG} = \frac{I_{withTVG}}{10^{TVG/10}}$$

Nato izračunamo natančen TVG z uporabo formule

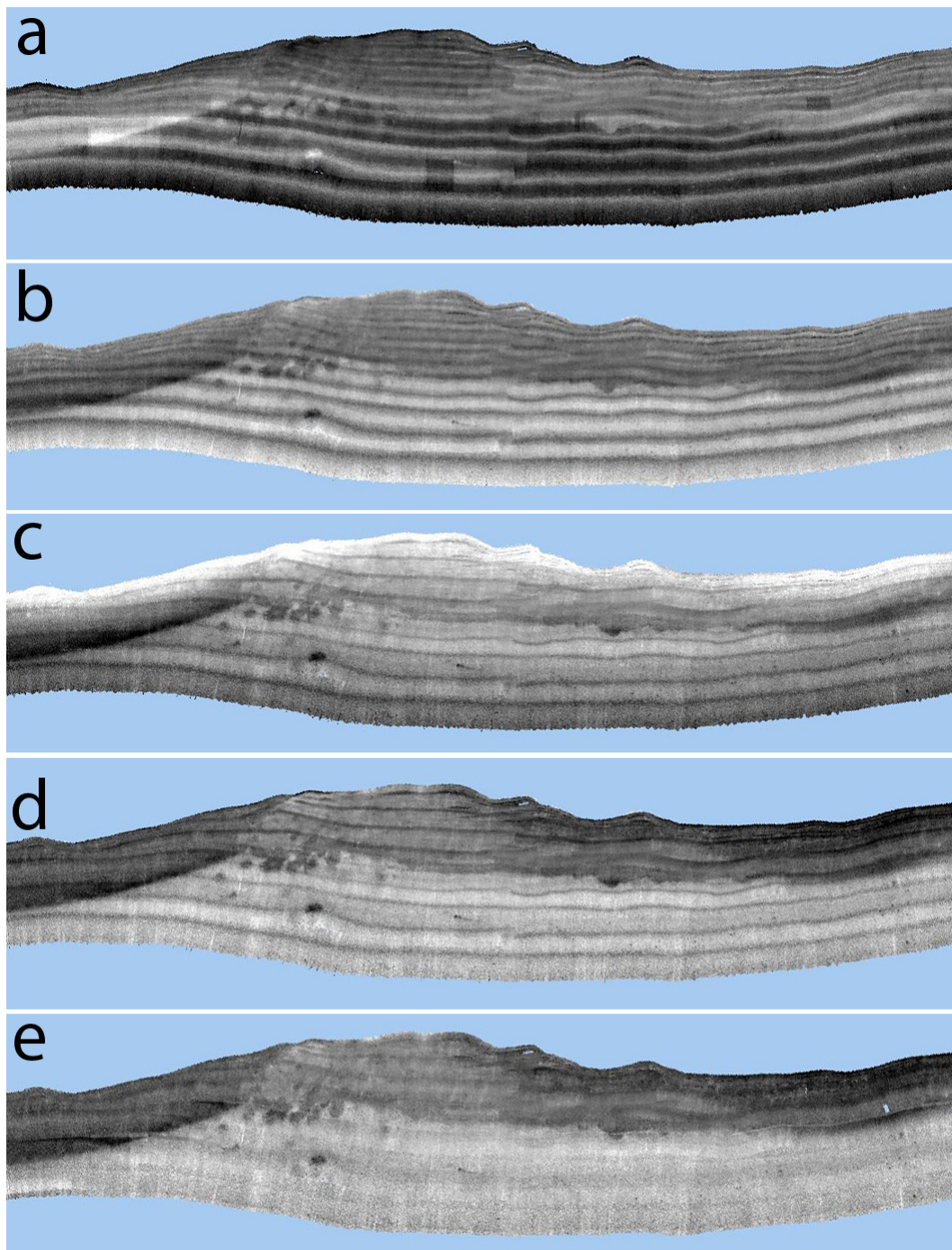
$$I_{TVG} = 10 \log(I_{noTVG}) + 20 \log(R) + \alpha R$$

kjer je  $R$  razdalja od sonarja do mesta odboja zvočnega signala. Formule izhajajo iz teorije podvodne akustike in so povzete po [39].

Izgube na poti od oddajnika do sprejemnika zaradi absorpcije v morski vodi so določene z enačbo  $\alpha R$ , izgube zaradi sferičnega širjenja signala pa z enačbo  $20 \log R$ . Izguba pri prenosu zaradi sferičnega širjenja signala je definirana kot zmanjšanje akustične intenzivnosti ko se val zvočnega tlaka širi navzven iz vira. V našem primeru sta oddajnik in sprejemnik na istem mestu, zvočni signal pa prepotuje pot



Slika 5.6: Na sliki zgoraj so prikazane uporabljene moči oddajnika. Moč oddajnika lahko operater nastavlja na lestvici od 1-14, pri čemer sprememba za eno enoto pomeni povečanje moči za 3 decibele. Spodaj so prikazani surovi podatki o intenzitetah odbojev v decibelih. Spremembe intenzitet odbojev sovpadajo s spremembami moči oddajnika. Zelo izrazito so vidne linije meritev.



Slika 5.7: 2R prikaz vpliva obdelave podatkov na sliko akustičnih odbojev. Pod točko a so prikazani surovi podatki. Točka b prikazuje podatke z normaliziranimi vpliv nastavitve moči in občutljivosti. Točka c in d prikazujeta podatke s popravljenim TVG in upoštevano velikostjo akustičnega odtisa. Točka e prikazuje podatke z izločenim vplivom kotnih odvisnosti.

do morskega dna in nazaj. Upoštevamo torej dvojno razdaljo do morskega dna  $R$ . Dobimo enačbo:

$$I_{TVG} = 10 \log(I_{noTVG}) + 40 \log(R) + 2\alpha R$$

Izmerjeni intenziteti torej prištejemo izgubo pri prenosu zaradi sferičnega širjenja signala in izgubo zaradi absorpcije zvoka v morski vodi.

Nekoliko bolj zahteven je izračun koeficienta  $\alpha$ , ki predstavlja absorpcijo zvoka v morski vodi. Absorpcije zvoka v morski vodi je del celotne izgube pri prenosu zvoka iz vira do sprejemnika. Odvisna je od lastnosti morske vode, kot so temperatura, slanost in kislost ter od frekvence zvoka. Podrobnosti o osnovnih fizikalnih lastnostih absorpcije zvočnih signalov v vodi so precej zapletene. Obstaja več enačb ki aproksimirajo izračun absorpcije zvoka v slani vodi. V tem delu smo uporabili enačbo povzeto po Ainslie-McColm [2]. Enačba upošteva tri dejavnike, ki vplivajo na absorpcijo zvoka v morski vodi: vpliv vode, vpliv borove kisline in vpliv magnezijevega sulfata. Koeficient absorpcije zvoka  $\alpha$  izračunamo kot vsoto vseh treh vplivov:

$$\alpha = \text{vpliv borove kisline} + \text{vpliv magnezijevega sulfata} + \text{vpliv vode}$$

Na koeficient absorpcije v vodi vplivata globina  $D$  in temperatura vode  $T$ :

$$\text{vpliv vode} = 0.00049 f^2 e^{-\left(\frac{T}{27} + \frac{D}{17}\right)}$$

Poleg temperature in globine je potrebno v morski vodi upoštevati še absorpcijo zaradi borove kisline in magnezijevega sulfata. Vpliv borove kisline je odvisen od temperature  $T$ , slanosti  $S$ , kislosti  $pH$  morske vode in delavne frekvence sonarja  $f$ . Izračunamo ga po formuli:

$$\text{vpliv borove kisline} = 0.106 \frac{f_1 f^2}{f_1^2 + f^2} e^{(pH-8)/0.56}$$

kjer je  $f_1$  definiran kot

$$f_1 = 0.78 \sqrt{\frac{S}{35}} e^{T/26}$$

Vpliv magnezijevega sulfata je odvisen od temperature vode  $T$ , slanosti  $S$  in globine  $D$

$$\text{vpliv magnezijevega sulfata} = 0.52 \left(1 + \frac{T}{43}\right) \left(\frac{S}{35}\right) \frac{f_2 f^2}{f_2^2 + f^2} e^{-D/6}$$

kjer je  $f_2$  definiran kot

$$f_2 = 42 e^{T/17}$$

Za skupno vrednost koeficienta absorpcije  $\alpha$  nato združimo vse tri vplive in dobimo formulo:

$$\alpha = 0.106 \frac{f_1 f^2}{f_1^2 + f^2} e^{(pH-8)/0.56} + 0.52 \left(1 + \frac{T}{43}\right) \left(\frac{S}{35}\right) \frac{f_2 f^2}{f_2^2 + f^2} e^{-D/6} + 0.00049 f^2 e^{-\left(\frac{T}{27} + \frac{D}{17}\right)} \quad (5.1)$$

Vrednost  $\alpha$  predstavlja koeficient absorpcije izražena pa je v enoti  $dB/km$ .

Na slikah 5.7b in 5.7c so prikazane intenzitete akustičnih odbojev pred in po popravku TVG.

### 5.4.3 Velikost akustičnega odtisa

Podatki so sedaj neodvisni od nastavitve sistema in opravljene poti signala, izločili smo tudi vpliv temperature vode, slanosti in kislosti. Na sliki 5.7c lahko še vedno opazimo, da se intenzitete odbojev spreminjajo z globino morja. Težava je v velikosti akustičnega odtisa. Akustični odtis je del morskega dna, od katerega se odbije akustični signal v sprejemnik. Gre za presek odtisa oddanega pulza in polja, ki ga pokriva sprejemnik. Površina akustičnega odtisa se veča z vpadnim kotom signala in globino morja. V literaturi je navedenih nekaj teoretičnih enačb za izločanje vpliva velikosti akustičnega odtisa, vendar nobena ne da zelenih rezultatov [102]. Težava je predvsem v tem, da je velikost akustičnega odtisa odvisna od vpadnega kota signala. Vpadni kot signala sicer lahko dokaj natančno določimo, vendar je morsko dno pogosto nagnjeno v več smereh, poleg tega pa enačbe predpostavljajo, da je morsko dno na celotni površini akustičnega odtisa ravno. Predvsem na razgibanem, skalnatem terenu take enačbe ne dajo zelenih rezultatov in celo popačijo sliko intenzitet akustičnih odbojev. Uporabili smo enotno formulo za vse vpadne kote akustičnega signala. S tem smo upoštevali spremembe velikosti akustičnega odtisa z globino morja, ne pa spremembe zaradi vpadnega kota signala. Vpliv vpadnega kota akustičnega signala je ločeno obravnavan v naslednjem poglavju.

Uporabili smo enačbo za velikost akustičnega odtisa  $P_{beam}$ :

$$P_{beam} = R^2 \times projector\_beam\_width \times \sin(Tx\_beam\_width\_across) \times \cos(30^\circ)$$

kjer je  $R$  razdalja do točke odboja,  $Tx\_beam\_width\_across$  prečna širina snopa sprejemnika in  $projector\_beam\_width$  vzdolžna širina snopa oddajnika. Vpadni kot smo nastavili na  $30^\circ$  in tako izločili le vpliv globine na površino akustičnega odboja. Popravljen intenzitete odbojev dobimo tako, da intenzitete delimo z velikostjo akustičnega odtisa  $P_{beam}$ :

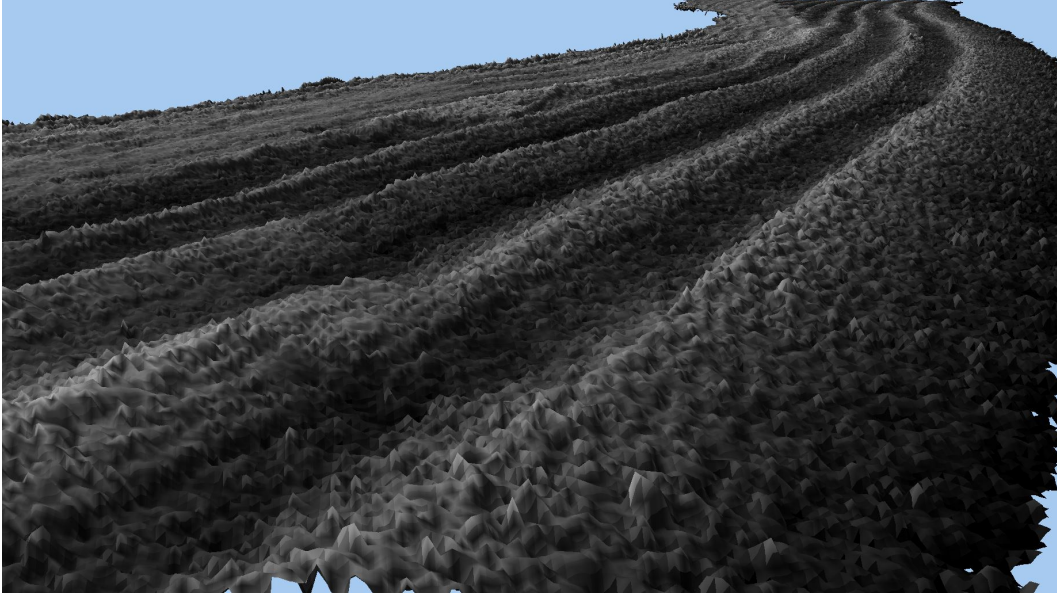
$$I = \frac{ITVG}{P_{beam}}$$

Popravljeni podatki so prikazani na sliki 5.7d. Na sliki je vidno, da navkljub vsem uravnavanjem in izločanjem različnih vplivov, podatki še vedno niso ustrezni. Lepo so razvidne linije po katerih je vozilo plovilo in razlike med meritvami pod plovilom ter levo in desno od plovila. Razlike nastanejo zaradi različnih vpadnih kotov signala. Vpadni kot in uravnavanje njegovega vpliva so obravnavani v poglavju 5.5.

## 5.5 Uravnavanje kotnih odvisnosti

Poleg prej omenjenih nastavitvev so intenzitete odboja močno odvisne tudi od vpadnega kota signala, čemur rečemo kotna odvisnost. Uravnavanje kotnih odvisnosti odbojev je eden izmed glavnih izzivov akustičnega kartiranja morskega dna. Primer anomalij v podatkih zaradi kotnih odvisnosti akustičnih odbojev je prikazan na sliki 5.8. Očitno je, da ima kot vpada velik vpliv na intenzitete akustičnega odboja. Z vpadnim kotom se spreminja velikost področja, od katerega se zvočni snop odbije, prav tako pa vpadni kot vpliva na razpršitev zvoka v različne smeri. Najmočnejši odboj dobimo, ko zvok zadane pravokotno ob tla, najšibkejšega pa, ko je vpadni

kot zelo velik. Ta vpliv moramo uravnati, sicer so podatki neuporabni za klasifikacijo. V literaturi najpogosteje naletimo na dva različna pristopa - teoretičnega in



Slika 5.8: Anomalije v podatkih zaradi kotnih odvisnosti akustičnih odbojev.

empiričnega [5]. V do sedaj opisanih postopkih predobdelave podatkov smo uporabljali enačbe, ki temeljijo na fizičnih lastnostih akustičnih odbojev in imajo trdne temelje v fiziki. Kotne odvisnosti so preveč zapletene, da bi jih bilo enostavno uravnati s fizičnim modelom. Teoretični pristop temelji na uravnavanju kotnih odvisnosti s pomočjo teoretičnih modelov izpeljanih iz fizikalnih zakonov širjenja signalov. Empirični pristop temelji na izvedbi praktičnih poskusov in izpeljavi formul na podlagi rezultatov poskusov. Pri bočnih sonarjih se v nekaterih primerih uporablja uravnavanje glede na vnaprej pripravljen fizikalni model odbojev signala pod različnimi koti in od različnih vrst morskega dna (teoretični pristop), vendar tak pristop pri mnogospornih sonarjih ne deluje zaradi bistveno višje frekvence [5]. Praktično nemogoče je izdelati model za vse vrste dna in vse kote. Uporabili smo empirični pristop.

Opravili smo poskusne meritve in na testnih podatkih preverili več različnih metod, ki so se razlikovale predvsem glede na upoštevanje/neupoštevanje standardnega odklona in lokalno oziroma globalno povprečenje podatkov. Uporabili smo štiri pristope, opisane v obstoječi literaturi in preizkusili dva nova pristopa. Zaradi preglednosti bomo ovrednotene pristope vključno s formulami navedli v skupnem seznamu ter jih podrobneje opisali v nadaljevanju. Seznam ovrednotenih pristopov:

**G:** globalni pristop [9]

$$I_G(\theta) = I(\theta) - \overline{I(\theta)} + \overline{I(30^\circ)}$$

**G\_std**: globalni pristop s standardnim odklonom [113]

$$I_{G\_std}(\theta) = \frac{I(\theta) - \overline{I(\theta)}}{I_{std}(\theta)} + \overline{I(30^\circ)}$$

**L**: Lokalni pristop [48]

$$I_L(\theta) = I(\theta) - \overline{I_{local}(\theta)} + \overline{I_{local}(30^\circ)}$$

**L\_std**: lokalni pristop s standardnim odklonom [101]

$$I_{L\_std}(\theta) = \frac{I(\theta) - \overline{I_{local}(\theta)}}{I_{local\_std}(\theta)} + \overline{I_{local}(30^\circ)}$$

**G\_L**: globalni in nato lokalni pristop

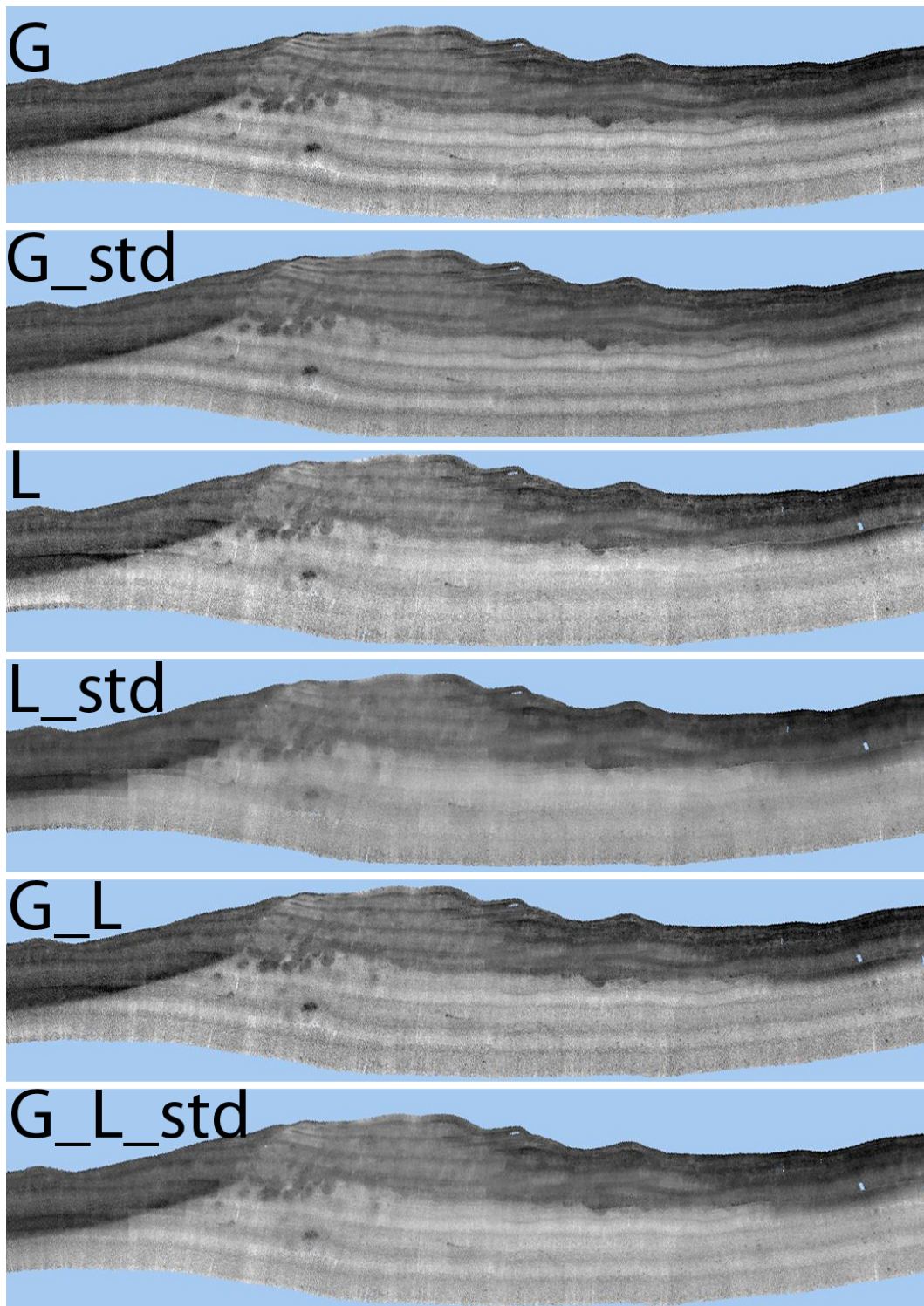
$$I_{G\_L}(\theta) = I_G(\theta) - \overline{I_{local}(\theta)} + \overline{I_{local}(30^\circ)}$$

**G\_L\_std**: globalni in nato lokalni pristop s standardnim odklonom

$$I_{G\_L\_std}(\theta) = \frac{I_G(\theta) - \overline{I_{local}(\theta)}}{I_{local\_std}(\theta)} + \overline{I_{local}(30^\circ)}$$

Pristop **G** od izmerjene intenzitete  $I(\theta)$  odšteje povprečno intenziteto  $\overline{I(\theta)}$  pri obravnavanem vpadnem kotu  $\theta$ . Pristop **G\_std** dobljeno razliko deli s standardnim odklonom intenzitet  $I_{std}(\theta)$  pri obravnavanem vpadnem kotu  $\theta$ . V obeh primerih prištejemo povprečne intenzitete pri kotu  $30^\circ$ ,  $\overline{I(30^\circ)}$ , zato da ohranimo razmerje med intenzitetami. Vpadni kot  $30^\circ$  je izbran le zato, ker je srednji kot iz razpona delovanja uporabljenega sonarja. Uporabili bi lahko katerikoli vpadni kot. Pristop **L**, za razliko od prvih dveh, deluje lokalno. Lokalna povprečna intenziteta  $\overline{I_{local}(\theta)}$  se prav tako izračuna za vsak vpadni kot posebej, vendar le v okolici obravnavanega odboja. Pristop **L\_std** dodatno upošteva še lokalni standardni odklon. Rezultati vseh pristopov so prikazani na sliki 5.9 pod ustreznimi oznakami. Rezultati s pristopom **G** kažejo, da kotne odvisnosti niso popolnoma odpravljene. Še vedno so v podatkih razločno vidne linije meritev. Pristop **L** veliko bolje odpravi kotne odvisnosti, vendar se pojavijo nove anomalije v podatkih zaradi lokalnega uravnavanja. Prednost globalnih pristopov je ohranjanje razmerij v podatkih, prednost lokalnih pristopov pa boljše uravnavanje kotnih odvisnosti.

Poleg omenjenih štirih pristopov smo razvili še dva nova pristopa s katerima želimo združiti prednosti lokalnih in globalnih pristopov. V pristopu **G\_L** predlagamo, da se na podatkih najprej izvede globalno uravnavanje in nato lokalno uravnavanje kotnih odvisnosti. Rezultate globalnega uravnavanja torej uporabimo za lokalno uravnavanje. Globalni pristop nekoliko slabše uravna kotne odvisnosti, vendar ne popači podatkov. Predpostavljamo, da bo lokalno uravnavanje na že nekoliko popravljenih podatkih dalo dobre rezultate z manj popačenja. Cilj je boljše uravnavanje kotnih odvisnosti in omilitev novih anomalij zaradi lokalnega uravnavanja. V pristopu **G\_L\_std** po globalnem uravnavanju uporabimo lokalni pristop s standardnim odklonom. Rezultata obeh pristopov sta prikazana na sliki 5.9.

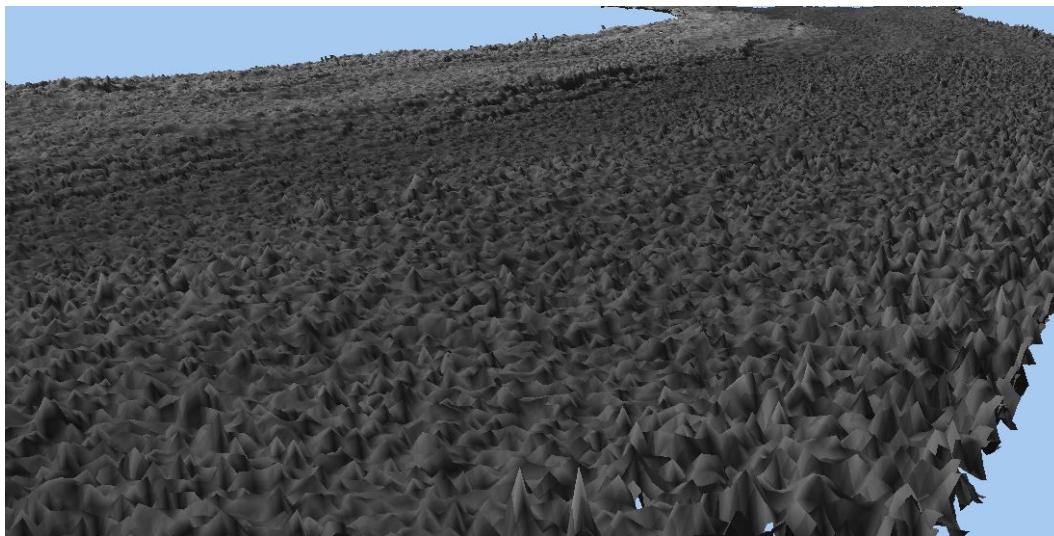


Slika 5.9: Primerjava različnih pristopov uravnavanja kotnih odvisnosti. Slike so označene z odgovarjajočimi oznakami uporabljenih pristopov.



Različne empirične pristope lahko primerjamo vizualno, kot je prikazano na sliki 5.9, vendar take primerjave niso objektivne. V delu [102] avtor primerja pristope **G**, **G\_std**, **L** in **L\_std**. Navaja slabosti posameznih metod in kot optimalen pristop priporoča povprečenje rezultatov globalnega uravnavanja na mrežo celic velikosti 5 metrov. S predlagano interpolacijo se izgubi veliko informacije in občutno zmanjša ločljivost podatkov, zato tak pristop ocenjujemo kot neprimeren za potrebe akustične klasifikacije morskega dna. Težavo vidimo tudi v tem da so metode ovrednotene le vizualno, avtorji primerjajo le slike uravnanih podatkov in na podlagi slik ocenjujejo kateri pristop je boljši. Veliko bolj pomemben je vpliv izbora metode na rezultate akustične klasifikacije. Primerjava različnih pristopov po sliki normaliziranih akustičnih odbojev ni objektivna. V tem poglavju je uporabljena le za prikaz delovanja pristopov. Pristope bomo ovrednotili in med sabo primerjali v poglavju 6.3, kjer bomo preučili vpliv uporabljenega pristopa za uravnavanja kotnih odvisnosti na rezultate klasifikacije akustičnih podatkov. Pristope bomo uporabili na realnih podatkih meritev slovenskega morja in ovrednotili njihov vpliv na klasifikacijsko točnost.

Kateregakoli od omenjenih pristopov uporabimo, so tako obdelani podatki neodvisni od nastavitve sistema, izločen je šum v podatkih in vpliv vpadnega kota na podatke. Popravljeni podatki so prikazani na sliki 5.10. Podatki so pripravljene za nadaljnjo obdelavo.



Slika 5.10: 3R prikaz podatkov z odpravljenimi kotnimi odvisnostmi po pristopu **G\_L\_std**.



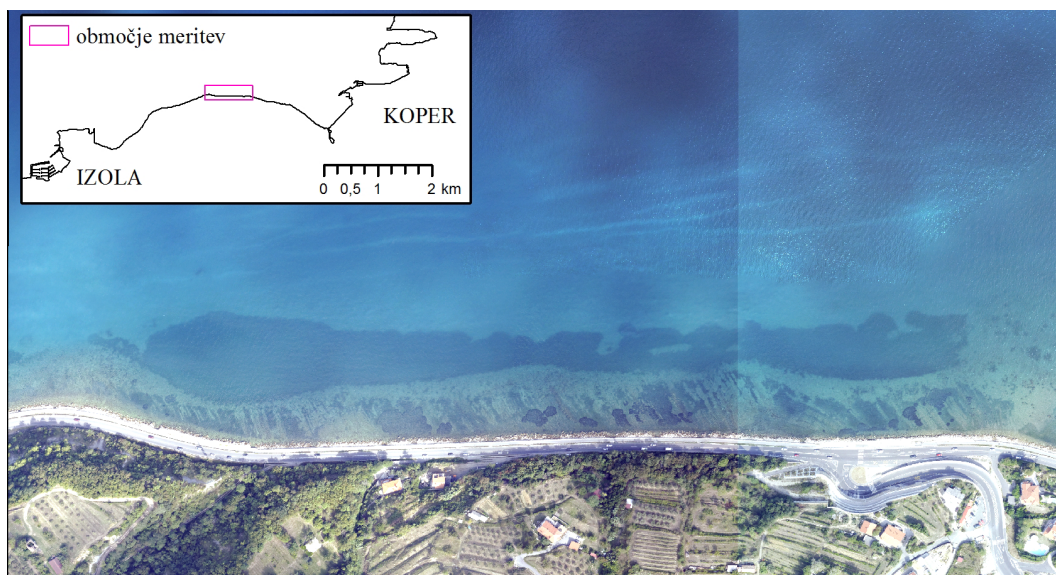
## Poglavje 6

# Opis podatkov in izbor značilk

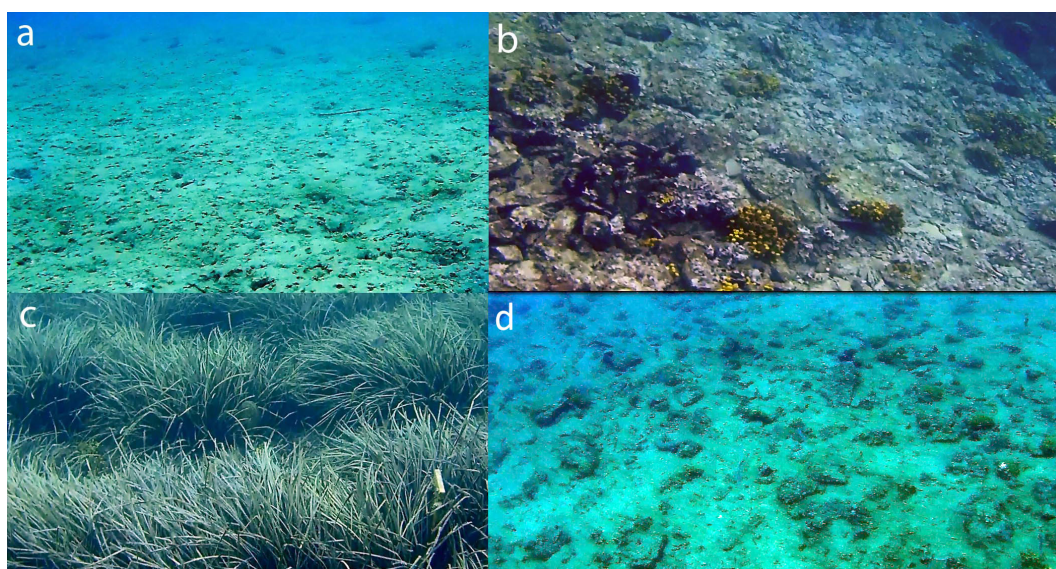
Za ovrednotenje klasifikatorjev in značilk za konkreten problem so ključnega pomena kakovostni podatki, na katerih lahko klasifikatorje in značilke ustrezno ovrednotimo. Okoljski podatki so na splošno znani kot zelo problematični za vrednotenje razvitih metodologij. Cilj naše metodologije je klasifikacija morskega dna. Največja težava, na katero smo naleteli, je pomanjkanje podatkov. Za slovensko morje imamo izdelano sedimentološko karto [72], ki pa je izdelana z interpolacijo maloštevilnih vzorcev sedimentov. Imamo tudi karto tipov morskega dna iz leta 2009 [74], ki temelji na subjektivni oceni strokovnjakov na podlagi batimetričnih podatkov o morskem dnu. Obstaja nekaj natančnejših kart morskih habitatov za manjša območja [106], ki pa so praviloma izrisane na podlagi linijskih transektov, z oddaljenostjo minimalno 50 metrov. Kar pomeni, da so vsi podatki za področje med dvema transektoma interpolirani in torej nezanesljivi. Za kakovostno in natančno ovrednotenje izbranih metod strojnega učenja smo tako morali izbrati dovolj veliko območje morskega dna in za to območje natančno določiti tipe morskega dna.

### 6.1 Definicija območja

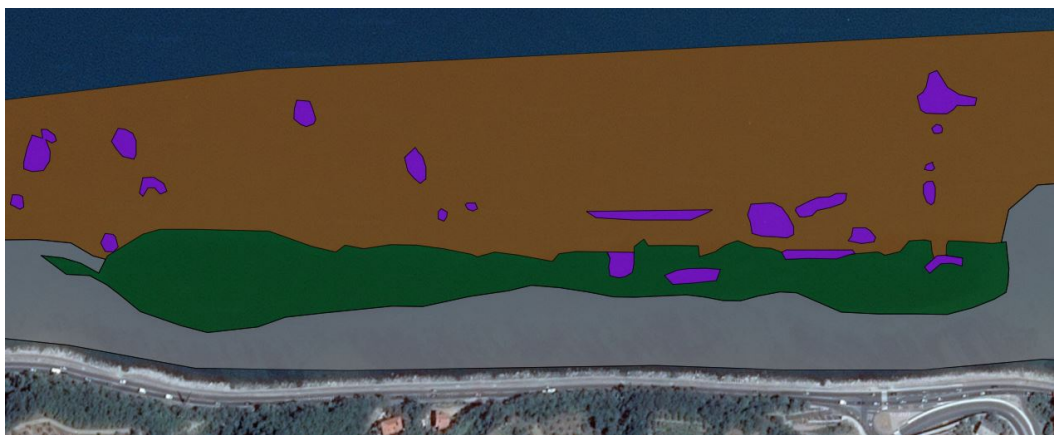
Hidrografske meritve smo izvedli na območju med Izolo in Koprrom. Območje je prikazano na sliki 6.1, ki prikazuje tudi aerofotografijo obravnavanega območja. Izbrano območje obsega 200 metrov širok in 900 metrov dolg obalni pas in pokriva 180.000 kvadratnih metrov morskega dna. Skupno je bilo zbranih 31 gigabajtov podatkov o preko 10.000.000 točkah morskega dna. Podatke smo prenesli na centralni strežnik in jih obdelali v razvitem modulu za avtomatsko predobdelavo podatkov, opisanem v poglavju 5, in izračunali značilke po postopku, opisanem v poglavju 6. Območje obale pred Žusterno smo izbrali, ker se na njem nahajajo štirje tipi morskega dna relativno blizu. Kot je bilo omenjeno že v uvodnem poglavju je dno slovenskega morja večinoma prekrito z muljem. Pogosto obalno kamnito dno direktno preide v mulj in je zaradi tega za naše raziskave manj zanimivo. Na izbranem območju najdemo tako kamnito dno in mulj kot tudi morske travnike in otoke izdankov kamnin sredi mulja.



Slika 6.1: Območje izbrano za ovrednotenje metod strojnega učenja.



Slika 6.2: Fotografije različnih tipov morskega dna. a) mulj, b) kamnito dno, c) morski travnik in d) izdanki kamnin



Slika 6.3: Prikaz tipov morskega dna na obravnavanem območju. S sivo je prikazano kamnito dno, zelena barva predstavlja morske travnike, rjava mulj in viola barva izdanke kamnin.

### 6.1.1 Podatki o razredu

Za ovrednotenje uspešnosti algoritmov strojnega učenja na problemu klasifikacije morskega dna potrebujemo podatke o tipih morskega dna. S temi podatki dopolnimo množico akustičnih podatkov. Informacij o morskem dnu dodamo v množico podatkov kot dodatno značilko, ki jo ustrezno označimo, da jo algoritmi strojnega učenja razpoznajo kot razred. Podatkom, ki imajo podan razred pravimo označeni podatki. Algoritmi za nadzorovano učenje potrebujejo podatke o razredu za gradnjo modelov. Dobljene modele nato uporabimo za klasifikacijo. V našem primeru so podatki o razredu pomembni tudi zaradi preverjanja uspešnosti algoritmov. Z vsakim od algoritmov bomo klasificirali določeno množico podatkov in rezultate primerjali z dejanskimi razredi.

Podatke o dejanskih tipih morskega dna smo pridobili po metodologijah opisanih v poglavju 4, ki so podrobneje predstavljene v delu [110]. Akustične podatke izbranega območja smo obdelali in jih klasificirali z metodo nenadzorovanega učenja. Uporabili smo metodo rojenja  $k$ -tih povprečij [56]. Tako smo dobili približen oris pestrosti iz razporeditve tipov morskega dna. Na podlagi slednjih informacij smo načrtovali video snemanje in točkovno vzorčenje. S kombinacijo podatkov video snemanje, točkovnega vzorčenja in fotointerpretacije aerofotografij smo izdelali poligone tipov morskega dna. Poligone smo izdelali v odprtokodnem programskem paketu za prostorske obdelave Qgis [114].

Na območju smo identificirali štiri različne tipe morskega dna, prikazane na sliki 6.2. Slika prikazuje:

- a) mulj
- b) kamnito dno
- c) morski travnik
- d) izdanke kamnin

Poligoni so prikazani na sliki 6.3. Pred pretvorbo prostorskih datotek v TAB in ARFF format smo poligonom izrezali en meter širok zunanji rob. Zunanji rob poligona predstavlja mejo med dvema tipoma morskega dna. Meje med različnimi tipi morskega dna so pogosto težko določljive. Morski travniki včasih zvezno prehajajo v muljnato dno, tudi meja med izdanki kamnin in muljem ni povsem natančno določljiva, četudi imamo jasne fotografije. Ker smo opazili da imajo strokovnjaki s področja oceanografije nemalo težav pri določanju mej med posameznimi tipi morskega dna, smo se odločili mejna območja izločiti iz množice podatkov namenjene ovrednotenju algoritmov strojnega učenja. S tem se želimo izogniti neželenemu vplivu nezanesljivih podatkov na rezultate vrednotenja algoritmov.

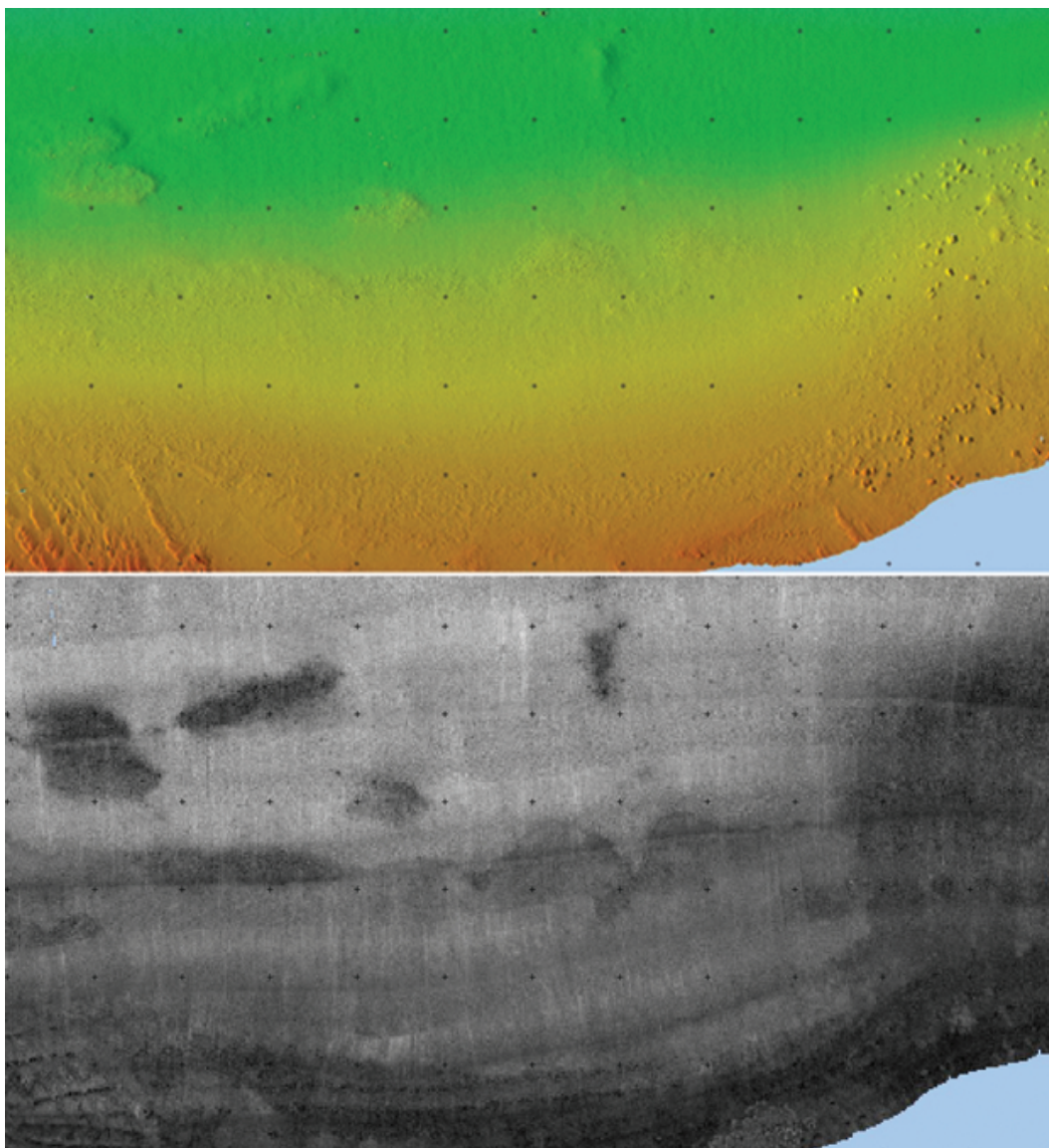
Dobljeni poligoni so izdelani v orodju Qgis v obliki vektorskega sloja. Za naše potrebe smo vektorski sloj pretvorili v rastrski sloj z metrsko natančnostjo. Uporabili smo format GeoTiff. Z uporabo knjižnice GDAL smo nato GeoTiff datoteke pretvorili v tekstovne datoteke z zapisanimi koordinatami in tipom morskega dna. Podatke smo na podlagi koordinat združili z akustičnimi podatki. Rezultat je množica podatkov, ki ima za vsak kvadratni meter morskega dna določeno vrednost razreda.

Velja omeniti tudi odkritje dveh novih živalskih vrst v slovenskem morju. Novi vrsti smo odkrili med zbiranjem podatkov o tipih morskega dna. Odkritje murene (*Muraena helena linnaeus*) smo opisali v delu [84], odkritje polža zaškrjarja vrste *Janolus cristatus* pa v delu [83]. V obeh primerih gre za prvi zabeležen primer pojavljanja vrste v slovenskih vodah.

## 6.2 Opis značilik

Značilke so neodvisne zvezne ali diskretne spremenljivke, ki opisujejo lastnosti podatkov. Cilj izračuna značilik je določanje tistih lastnosti akustičnih podatkov, ki bi lahko bile uporabne za razlikovanje med različnimi tipi morskega dna. Potrebno je izluščiti čim več kakovostnih značilik, uporabnih za klasifikacijo – želimo torej izluščiti informacije iz surovih podatkov [104]. Značilke so lahko osnovne meritve, kot je intenziteta odboja, ali pa bolj zapletene, kot so rezultati statističnih ali prostorskih analiz. Izračun ustreznih značilik je temelj za uspešno klasifikacijo morskega dna. Prednost sodobnih mnogospornih sonarjev je možnost hkratnega snemanja globin in podatkov o intenziteti odboja. V kombinaciji s sodobnimi GNSS lahko podatke zelo natančno umestimo v prostor. Rezultat sta dva modela – batimetrični model (digitalni model globin) in model akustičnih odbojev. Iz posameznega odboja ne dobimo dovolj informacije, da bi sklepali na tip tal. Ker so podatki umeščeni v prostor, lahko izkoristimo prostorsko komponento. Prostorska komponenta nam omogoči, da podatke združimo v celice in na njih izvajamo prostorske analize. Podatke smo razdelili na celice velikosti 1x1 meter, 2x2 metra in 4x4 metre ter iz podatkov o globinah izračunali v nadaljevanju opisane karakteristike terena. Podobno smo iz podatkov o intenzitetah akustičnih odbojev izračunali karakteristike slik akustičnih odbojev.

Na sliki 6.4 je zgoraj prikazan batimetrični model, spodaj pa slika intenzitet akustičnih odbojev. Opazimo, da je batimetrični model zelo čist, slika intenzitet akustičnih odbojev pa, kljub zahtevni predobdelavi, precej šumna. Iz analiz batimetričnega modela je mogoče pridobiti veliko geomorfoloških podatkov o morskem



Slika 6.4: Prikaz batimetričnega modela (zgoraj) in slike akustičnih odbojev (spodaj).

dnu, na primer razločiti peščine, skalne gmote, ipd. Poleg tega globina vpliva na razporeditev habitatov, saj se z globino hitro zmanjšuje osvetljenost morskega dna in tudi izpostavljenost valovanju morja. Čeprav je analiza batimetrij učinkovita pri določanju geomorfoloških značilnosti morskega dna, pa sama po sebi ni uspešna pri določanju habitatnih tipov, če ni pomembnih razlik v topografiji habitatov. Morski travniki, so na primer prisotni na ploskem terenu, brez večjih naklonov ali izstopajočih oblik in jih samo na podlagi batimetričnega modela zelo težko ločimo od muljastega ali peščenega dna. Za razliko od analize batimetrij je analiza slik akustičnih odbojev manj intuitivna, saj analiziramo moč odbojev in ne globin.

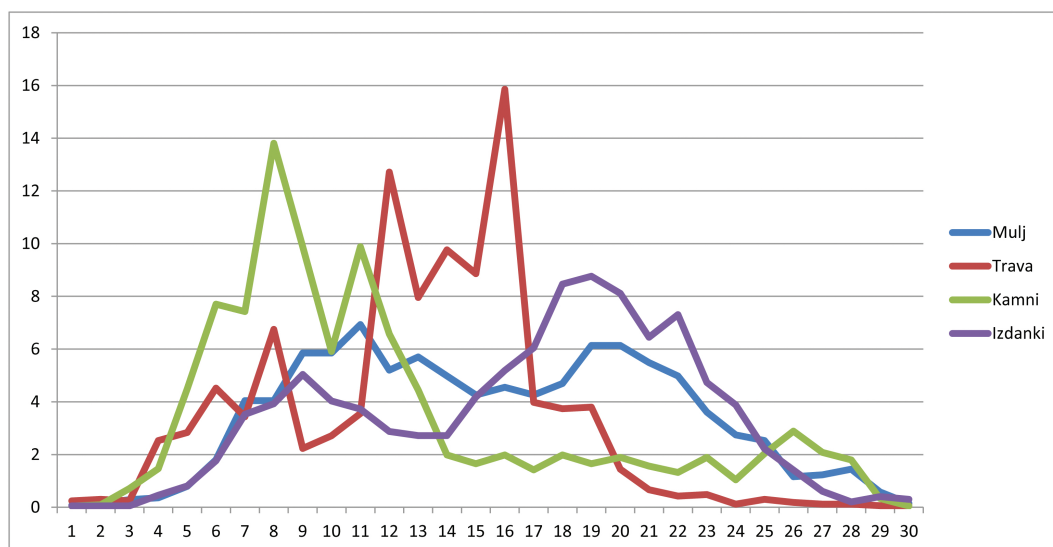
Analize slik akustičnih odbojev za kartiranje morskega dna so se začele z uporabo bočnih sonarjev. Metode z uporabo matrik sopojavnosti sivin (GLCM – Grey-Level Co-occurrence Matrix) Fourier-jevih analiz in analizo oscilacij so se izkazale za uspešne [7]. Cilj omenjenih metod je posnemanje principov človeškega vida, ki razlikuje med različnimi toni in teksturami. Značilke, ki smo jih uporabili, lahko grobo razdelimo v dva razreda: karakteristike terena ter karakteristike slik akustičnih odbojev. Uporabili smo povprečno vrednost, standardni odklon, višje momente (asimetrijo in sploščenost) ter lastnosti matrike sopojavnosti sivin - kontrast, različnost, homogenost, energijo in korelacijo [55]. Vse naštetje značilke smo izračunali iz obeh modelov, batimetričnega modela in modela akustičnih odbojev. Take podatke smo izračunali za manjša območja kot tudi za večje dele morskega dna in tako pridobili širše znanje o razgibanosti terena ter akustičnih lastnostih morskega dna. Vsi ti podatki so lahko poglavitnega pomena pri določanju habitatnih tipov. Zelo pomembno je, da smo pred izpeljevanjem značilk iz obeh modelov izločili vplive gibanja plovila, vplive nastavitve opreme in izgube zaradi širjenja zvočnega signala v morski vodi. Izračunane karakteristike smo pretvorili v ustrezen format za obdelavo v orodju za podatkovno rudarjenje Orange [30] in Weka [54].

V nadaljevanju poglavja opišemo uporabljene značilke, ovrednotimo pristope k uravnavanju kotnih odvisnosti in izberemo najboljši pristop. Nato značilke ponovno izračunamo iz podatkov obdelanih z uporabo izbranega pristopa. Za vsako izmed značilk določimo najprimernejšo velikost celice in iz reducirane množice značilk določimo tiste najbolj pomembne, ki jih bomo uporabljali v postopku klasifikacije.

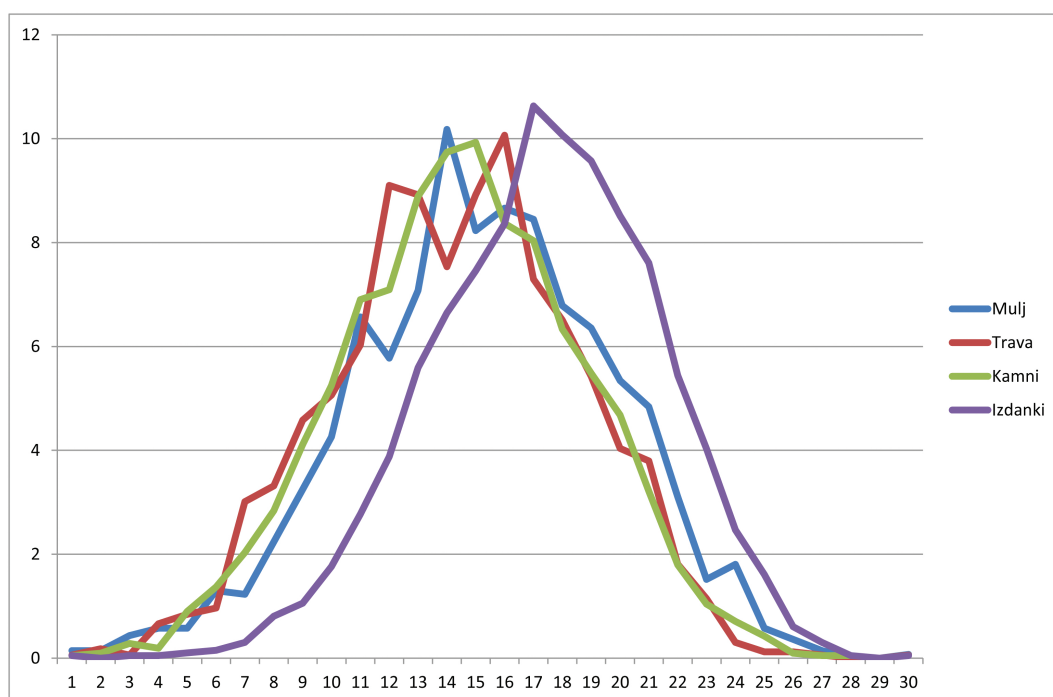
### 6.2.1 Osnovne značilke

Osnovni značilki, od katerih pričakujemo največ informacije, sta povprečna intenziteta akustičnega odboja in globina. Jasno je, da ima globina velik vpliv na razporeditev tipov morskega dna. Morski travniki v slovenskem morju ne uspevajo globlje od 8 metrov, zaradi pomanjkanja svetlobe [106], mulj najdemo nekoliko globlje saj ga valovanje in plimovanje izpirata stran od obale. Skalnato dno je načeloma lahko prisotno kjerkoli, vendar ga na globini večinoma prekrivajo debeli sedimenti [74]. Druga pomembna značilka je intenziteta odboja. Intenzitete odboja se že nekaj časa uporabljajo za določanje tipov morskega dna z interpretacijo slik akustičnih odbojev s strani strokovnjakov, ne nazadnje so to metodo uporabili leta 2015 v raziskavi dela obale v južni Italiji [63]. Intenzitete akustičnih odbojev so se izkazale za pomembne faktor pri razločanju različnih tipov morskega dna, zato pričakujemo, da bodo tudi naše analize potrdile dosedanje ugotovitve. Poleg povprečnih vredno-





Slika 6.5: Histogram verjetnostnih porazdelitev globin znotraj celice za različne tipe morskega dna.  $N = 7144$ .



Slika 6.6: Histogram verjetnostnih porazdelitev intenzitet znotraj celice za različne tipe morskega dna.  $N = 7144$ .

sti intenzitet in globin smo izračunali tudi standardne odklone. Standardni odklon batimetrij znotraj posamezne celice lahko interpretiramo kot razgibanost terena, v angleški literaturi pogosto naletimo na izraz 'roughness', kar bi pomenilo hrapavost. Iz slike 6.4 zgoraj vidimo, da je dno na nekaterih delih bolj razgibano kot drugje. Globlje, kjer je prisoten mulj, je dno skoraj popolnoma ravno. Desni del slike, kjer je morsko dno kamnito s posameznimi skalami, pa je bistveno bolj razgiban. V podatkih o intenzitetah je še vedno prisotnega precej šuma in razlike v hrapavosti niso takoj vidne. Poleg tega je interpretacija manj intuitivna, vendar bomo standardne odklone kljub temu izračunali in ovrednotili.

### 6.2.2 Asimetrija in sploščenost

Naslednji uporabljeni značilki sta asimetrija in sploščenost. Med analizo verjetnostnih porazdelitev globin in intenzitet smo opazili razlike med porazdelitvami za različne tipe tal. Razlike so opazne predvsem pri porazdelitvah globin (slika 6.5), pri porazdelitvah intenzitet te razlike niso tako očitne (slika 6.6). Prikazani so histogrami za celico velikosti 4x4 metre, za različne tipe tal. Histogram je ocenjen na 7144 odbojih od morskega dna. Točke za analizo smo določili z metodo točkovnega vzorčenja s pomočjo potapljačev. Primerjali smo sonarske podatke štirih tipov morskega dna - muljastega dna, kamnitega dna, morskih travnikov in izdankov kamnin. Na muljastem dnu so globine enakomerno porazdeljene, histogram je sploščen. Pri porazdelitvah globin kamnitega dna opazimo asimetrijo, vrh histograma je zamaknjen v levo. Sklepamo, da je to posledica večinoma sicer ravnega dna, z nekaj višjimi skalami. Histogram verjetnostnih porazdelitev globin za morski travnik je najbolj simetričen. Na območju izdanka kamnin dobimo nekoliko sploščen histogram z zamikom vrha v desno. Razlike v porazdelitvah so torej očitne. Histograma verjetnostnih porazdelitev ne moremo uporabiti kot značilke, lahko pa uporabimo mere ki opisujejo zgoraj omenjene lastnosti, taki meri sta asimetrija in sploščenost.

#### Asimetrija

Koeficient simetrije (včasih tudi koeficient asimetrije) je v teoriji verjetnosti in statistiki merilo, ki meri asimetrijo verjetnostne porazdelitve realne slučajne spremenljivke. Označujemo ga z  $\gamma$ .

Koeficient simetrije lahko zavzame naslednje vrednosti:

- negativni koeficient simetrije
- pozitivni koeficient simetrije
- koeficient simetrije je enak 0

Negativni koeficient simetrije ima porazdelitev takrat, ko je levi del porazdelitve daljši ali z drugimi besedami, kadar je v zbirni porazdelitveni funkciji večina porazdelitve skoncentrirana na desni strani. Pozitivni koeficient simetrije ima porazdelitev takrat, ko je desni del porazdelitve daljši ali z drugimi besedami kadar je v zbirni funkciji verjetnosti večina porazdelitve skoncentrirana na levi strani porazdelitve. Nesimetrična porazdelitev pomeni tudi, da ima srednja vrednost večjo vrednost kot

mediana. Pri simetričnih porazdelitvah je srednja vrednost enaka mediani in tudi modusu. Koeficient simetrije je enak 0 pri simetričnih porazdelitvah.

Simetrijo pri vzorcu  $n$  vrednosti izračunamo na naslednji način

$$\gamma = \frac{m_3}{m_2^{3/2}} = \frac{\frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})^3}{\left( \frac{1}{n-1} \sum_{i=1}^n (x_i - \bar{x})^2 \right)^{3/2}}$$

kjer je  $x_i$   $i$ -ta vrednost v vzorcu  $\bar{x}$  je srednja vrednost vzorca  $m_3$  je tretji centralni moment in  $m_2$  varianca vzorca.

### Sploščenost

Sploščenost (tudi koeficient ekscesa ali koeficient sploščenosti) je pri statistiki vrednost, ki meri koničastost (ostrota vrha) verjetnostne porazdelitve realne slučajne spremenljivke. Označimo jo z  $\gamma_2$ . Na splošno pomeni večja sploščenost tudi, da je večji del variance posledica izjemnih vrednosti.

Sploščenost je definirana kot razmerje med četrto kumulanto in kvadratom druge kumulante. Za vzorec  $n$  vrednosti izračunamo sploščenost vzorca na naslednji način:

$$\gamma_2 = \frac{\kappa_4}{\kappa_2^2} = \frac{\frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})^4}{\left( \frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})^2 \right)^2}$$

kjer je  $\bar{x}$  srednja vrednost vzorca  $x_i$  pa so posamezne vrednosti iz vzorca. Zaradi preglednosti v tem delu sploščenosti ne označujemo z oznako  $\gamma_2$  ampak kar z besedo sploščenost.

### 6.2.3 Teksturane značilke

Čeprav lahko teksturo intuitivno povežemo z različnimi lastnostmi površine nekega materiala - oziroma njene slike - kakršne so gladkost/hrapavost, grobost, usmerjenost ali periodičnost strukture, pa je sam pojem zelo težko natančno definirati. Gre za veličino, ki opisuje lokalno spreminjanje svetilnosti (sivinskih nivojev) pikselov digitalne slike, lahko pa jo razumemo tudi kot značilko, ki predstavlja prostorsko razporeditev vrednosti svetilnosti na določenem področju slike. Postopki analize teksture se v glavnem razlikujejo po načinih opisa teksture. Statistične metode opisujejo teksturo območij na sliki s pomočjo veličin, izračunanih na podlagi histogramov svetilnosti posameznih področij slike. Najbolj znana med njimi je metoda, ki temelji na generiranju t.i. matrike vezanih verjetnosti sivinskih nivojev (ang. Gray Level Cooccurrence Matrix - GLCM) [55]. Na podlagi GLCM je možno izračunati razne statistične mere oz. značilke (ang. texture measures). Haralick je v svojem izvornem članku o GLCM [55] predlagal 14 značilk, od katerih pa se jih danes v praksi uporablja le manjše število, zlasti kontrast (ang. contrast), korelacija (ang. correlation), energija (ang. energy), entropija (ang. entropy), homogenost (ang. homogeneity) in različnost (ang. dissimilarity).

Za izračun omenjenih mer moramo najprej pripraviti matriko vezanih verjetnosti sivinskih nivojev. Gre za tabelo, v katero se vpisujejo podatki o tem, kako pogosto se na sliki pojavljajo različne kombinacije vrednosti svetilnosti. Spremljajo se pari

pikslov, pri čemer se proučevana piksla nahajata na določeni medsebojni razdalji (1, 2, ... do največ 8 pikslov) in smeri (kot med piksloma običajno 0, 45, 90 ali 135°). Uporabili smo osnovne nastavitve kota med piksloma 0° in srednje razdalje 5 pikslov. Podatke smo razporedili v 10 nivojev svetilnosti (ang. levels). V matriki vezanih verjetnosti sivinskih nivojev vrednost  $P_{i,j}$  pomeni pogostost pojavitve para pikslov svetilnosti  $i$  in  $j$ . Uporabili smo naslednje formule za izračun značilik iz matrike vezanih verjetnosti sivinskih nivojev:

- Kontrast:

$$\sum_{i=0}^{N-1} \sum_{j=0}^{N-1} P_{i,j} (i - j)^2$$

- Različnost:

$$\sum_{i=0}^{N-1} \sum_{j=0}^{N-1} P_{i,j} |i - j|$$

- Homogenost:

$$\sum_{i=0}^{N-1} \sum_{j=0}^{N-1} \frac{P_{i,j}}{1 + (i - j)^2}$$

- Energija:

$$\sqrt{\sum_{i=0}^{N-1} \sum_{j=0}^{N-1} P_{i,j}^2}$$

- Korelacija:

$$\sum_{i=0}^{N-1} \sum_{j=0}^{N-1} P_{i,j} \left[ \frac{(i - \mu_i)(j - \mu_j)}{\sqrt{(\sigma_i^2)(\sigma_j^2)}} \right]$$

- Entropija:

$$- \sum_{i=0}^{N-1} \sum_{j=0}^{N-1} P_{i,j} \log_2 P_{i,j}$$

kjer je  $N$  število nivojev svetilnosti (v našem primeru 10),  $\mu$  označuje povprečno vrednost in  $\sigma$  standardni odklon. Vse našete značilke smo izračunali tako iz podatkov o intenzitetah kot iz podatkov o globinah.

#### 6.2.4 Pregled značilik

Pregled izračunanih značilik in oznak, ki jih bomo uporabljali v nadaljevanju je predstavljen v tabeli 6.1.

Skupno imamo torej 20 značilik, 10 izpeljanih iz globin in 10 izpeljanih iz intenzitet odbojev. Za vsako točko izračunamo 20 naštetih značilik za celice velikosti 1x1 meter, 2x2 metra in 4x4 metre, torej imamo skupaj 60 značilik.

Opis	Globine	Intenzitete
Povprečna vrednost	z	i
Standardni odklon	z_std	i_std
Simetričnost histograma	z_simetričnost	i_simetričnost
Sploščenost histograma	z_sploščenost	i_sploščenost
Kontrast po GLCM	z_kontrast	i_kontrast
Energija po GLCM	z_energija	i_energija
Različnost po GLCM	z_različnost	i_različnost
Entropija po GLCM	z_entropija	i_entropija
Korelacija po GLCM	z_korelacija	i_korelacija
Homogenost po GLCM	z_homogenost	i_homogenost

Tabela 6.1: Seznam značilk.

### 6.2.5 Splošne lastnosti značilk

Vsaka značilka ima različne lastnosti, od katerih je odvisna predobdelava učnih primerov in tudi izbira in nastavitvev klasifikacijskega ali regresijskega algoritma. Značilka z manjkajočimi vrednostmi je takšna značilka, pri kateri manjka vrednost pri enem ali več učnih primerih. Nekateri klasifikacijski in regresijski algoritmi lahko uporabljajo značilke z manjkajočimi vrednostmi, nekateri pa tega ne znajo. Med predobdelavo smo poskrbeli, da naši podatki ne vsebujejo manjkajočih vrednosti.

Šumna značilka vsebuje napake v podatkih bodisi zaradi napak pri meritvah ali zaradi napačnega vnosa podatkov. Tako so skoraj vsi realni podatki bolj ali manj šumni. Podobno kot pri značilkah z manjkajočimi vrednostmi tudi šumne značilke lahko nekateri klasifikacijski in regresijski algoritmi uporabljajo, nekateri pa v tem primeru odpovejo, oziroma postanejo nezanesljivi. Primer algoritma, ki uspešno deluje na šumnih podatkih, je odločitveno drevo. Z rezanjem spodnjih nivojev drevesa zmanjšuje nezanesljivost klasifikacije, ki je posledica šuma v podatkih. Primer šumne značilke so intenzitete odbojev brez predobdelave. V takšni značilki je prisotnega veliko šuma zaradi kotnih odvisnosti, nastavitvev sistema ipd.

Naključna značilka je nepovezana z ostalimi značilkami in s ciljno spremenljivko. Taka značilka je nepomembna in jo je najbolje odstraniti, saj samo znižuje zanesljivost učnih algoritmov. Primer naključnega značilke je zaporedna številka učnega primera, ki nima nobene povezave z razredom. V prejšnjem poglavju smo opisali postopke za izračun značilk iz slik akustičnih odbojev in batimetričnih podatkov. Povsem možno je, da bo katera izmed izračunanih značilk naključna in jo bomo morali odstraniti iz podatkov.

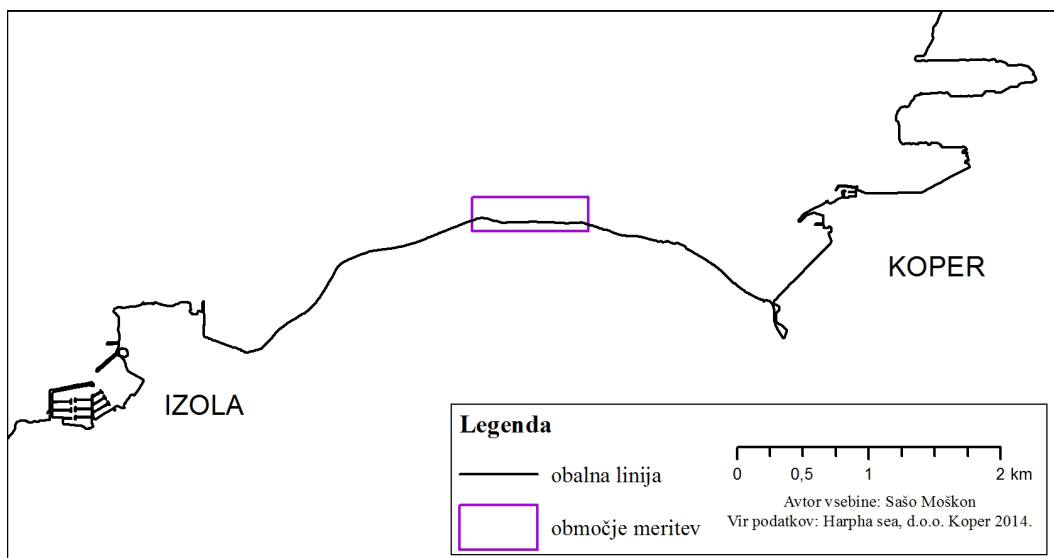
Redundantna značilka je značilka, katere informacijo vsebuje že neka druga značilka ali množica značilk. Tako značilko je najbolje odstraniti. Korelirana značilka je značilka, katere informacijo delno vsebuje že neka druga značilka ali množica značilk. Bolj kot so značilke korelirane med seboj, večje so njihove soodvisnosti in več je redundance. Vse značilke, ki so izračunane za metrske, dvometrske in štirimetrske celice so seveda močno korelirane, saj gre za enake značilke izračunane na podmnožicah istih podatkov. Namenoma smo uporabili več koreliranih značilk in jih kasneje, kot je opisano v poglavju 6.4 ovrednotili in izbrali le po eno iz vsake množice koreliranih značilk.

Če so značilke močno soodvisne glede na razred je ciljno funkcijo težko odkriti, saj se te značilke šele v kontekstu z drugimi pokažejo za pomembne. Primer močne soodvisnosti glede na razred sta značilki, ki sami ne povesta ničesar o razredu, skupaj pa ga zelo uspešno napovesta. Ekskluzivni ali je primer funkcije med dvema močno soodvisnima značilkama. Algoritem ReliefF, ki je opisan v poglavju 6.4 je sposoben uspešnega reševanja te vrste problemov.

### 6.3 Ovrednotenje različnih pristopov k uravnavanju kotnih odvisnosti

V poglavju 5.5 smo podali 6 različnih empiričnih pristopov za uravnavanje kotnih odvisnosti. Pristope smo primerjali vizualno, na surovih podatkih smo uporabili vseh šest pristopov, rezultate pretvorili v sliko intenzitet odbojev in ocenili posamezne pristope. Taka primerjava/ovrednotenje je prisotna tudi v literaturi [102], [6], [48]. Ker so take primerjave subjektivne in brez pravih argumentov za ali proti posameznemu pristopu, smo se odločili objektivno ovrednotiti pristope na primeru realnih podatkov posnetih v slovenskem morju in tako določiti najprimernejši pristop za nadaljnjo uporabo.

#### 6.3.1 Opis podatkov in uporabljenih značilk



Slika 6.7: Območje zajema podatkov.

Sonarske meritve za zajem podatkov smo opravili na območju obale pred Žusterno v koprskem zalivu. Območje je prikazano na sliki 6.7. Izmerjeno območje je dolgo 900 metrov. Najplitvejša izmerjena točka je globoka 1,3 metra, najgloblja pa 11,4 metra. Območje obale pred Žusterno je bilo izbrano, ker se na njem nahajajo štiri

tipi morskega dna relativno blizu. Območje in podatki so natančneje opisani na začetku poglavja 6. Kot je bilo omenjeno že v uvodnem poglavju je dno slovenskega morja večinoma prekrito z muljem. Pogosto obalno kamnito dno direktno preide v mulj in je zaradi tega za naše raziskave nezanimivo. Na izbranem območju najdemo tako kamnito dno in mulj kot tudi morske travnike in otoke izdankov kamnin sredi mulja. Surove sonarske podatke smo obdelali po razviti metodologiji opisani v poglavju 5, vendar le do koraka uravnavanja kotnih odvisnosti. Podatke smo nato obdelali na 7 različnih načinov in vsakemu pripisali ustrezno oznako. Dobili smo 7 množic podatkov:

- B: Neobdelani podatki. Kotne odvisnosti se ohranijo.
- G: Uporabljen pristop z globalnim uravnavanjem kotnih odvisnosti.
- L: Uporabljen pristop z lokalnim uravnavanjem kotnih odvisnosti.
- G\_L: Uporabljen pristop z globalnim in nato lokalnim uravnavanjem kotnih odvisnosti.
- G\_std: Uporabljen pristop z globalnim uravnavanjem kotnih odvisnosti in globalnim uravnavanjem standardnih odklonov.
- L\_std: Uporabljen pristop z lokalnim uravnavanjem kotnih odvisnosti in lokalnim uravnavanjem standardnih odklonov.
- G\_L\_std: Uporabljen pristop z globalnim in nato lokalnim uravnavanjem kotnih odvisnosti in lokalnim uravnavanjem standardnih odklonov.

Vsem primerom v vsaki od množic podatkov smo dodali tipe morskega dna. Ker uravnavanje kotnih odvisnosti vpliva le na značilke povezane z intenziteto odboja, smo vse ostale značilke izločili iz podatkov. V naslednjem koraku smo iz vsake množice podatkov izbrali podmnožico tako, da so vsi štirje tipi morskega dna enakomerno zastopani. Naš cilj je namreč določiti kateri pristop k uravnavanju kotnih odvisnosti je najboljši za razločevanje med različnimi tipi morskega dna. V originalnih podatkih mulj predstavlja kar 59% morskega dna, otoki izdankov kamnin pa le 3%. Ovrednotenje na takih podatkih bi lahko imelo za posledico predobro oceno za pristop, ki bi morda bolje razločil mulj od ostalih tipov morskega dna. Z enakomerno zastopanostjo vseh štirih razredov (vsak razred predstavlja 25% primerov v množici podatkov) se temu problemu izognemo. Rezultat obdelave je sedem množic podatkov, kjer:

- so le značilke povezane z intenzitetami
- so vsi štirje tipi morskega dna enakomerno zastopani
- je na vsaki množici uporabljen drugačen pristop uravnavanja kotnih odvisnosti

### 6.3.2 Postopek testiranja

Podatke smo najprej pretvorili v formata TAB in ARFF za uporabo v orodjih za podatkovno rudarjenje. Kakovost podatkov smo ovrednotili na podlagi klasifikacijske točnosti algoritma, naučenega na množici podatkov, opisani v prejšnjem razdelku. Klasifikacijska točnost nam pove kakšen delež primerov v testni množici je naš algoritem pravilno klasificiral in je definirana kot razmerje med pravilnimi rešitvami primerov problema in vsemi možnimi primeri. Klasifikacijsko točnost natančneje definiramo v poglavju 7.1.2.

Ker so vsi možni parametri, ki bi lahko vplivali na klasifikacijsko točnost, v vseh sedmih primerih enaki, bo na točnost vplivala zgolj kakovost podatkov. Bolj kakovostni podatki nam torej daje boljše rezultate. Poskrbeli smo, da je edino kar se med posameznimi množicami podatkov razlikuje, pristop k uravnavanju kotnih odvisnosti. Ko določimo najbolj kakovostne podatke bomo torej posledično dobili tudi najboljši pristop uravnoveženja kotnih odvisnosti. Na vseh sedmih množicah podatkov smo uporabili algoritem klasifikacije z odločitvenimi drevesi C4.5. Algoritem je podrobneje opisan v poglavju 7.1.1. Za ta algoritem smo se odločili ker je eden izmed najpogosteje uporabljenih algoritmov strojnega učenja, je zelo robusten in deluje tako z diskretnimi kot zveznimi spremenljivkami, zna obravnavati manjkajoče vrednosti in je robusten na šum. Za oceno klasifikacijskih točnosti smo uporabili 10-kratno prečno preverjanje. 10-kratno prečno preverjanje je splošno sprejet način vrednotenja točnosti algoritmov strojnega učenja. Deluje tako, da množico podatkov naključno razdeli na 10 enako velikih podmnožic, nato pa 10-krat uporabi 9 podmnožic za učenje in eno za testiranje naučenega modela. 10-kratno prečno preverjanje smo pognali 10 krat za vsako množico podatkov. Rezultat je povprečje klasifikacijskih točnosti vseh 100-ih tekov algoritma za posamezno množico podatkov. Rezultate 10 zagonov 10-kratnega prečnega preverjanja smo nato uporabili za test statistične značilnosti razlik. Za preverjanje razlik med rezultati na posameznih množicah podatkov smo uporabili parni T-test z Bonferroni popravki [26]. Test smo izvedli za vse možne pare množic podatkov, torej 21-krat. Meja za statistično značilnost je bila pri  $\alpha = 0.05$ . Metoda 10-kratnega prečnega preverjanja in parni T-test sta podrobneje opisani v poglavju 7.1.2.

Preverili smo tudi, kako se spreminjajo povezave med posameznimi značilkami in razredom z različnimi pristopi k uravnavanju kotnih odvisnosti. V ta namen smo za ovrednotenje značilk uporabili informacijski prispevek. Informacijski prispevek je definiran kot prispevana informacija značilke za določitev njene vrednosti. Informacijski prispevek in ostale mere za vrednotenje značilk so bolj podrobno opisane v poglavju 6.4.1.

### 6.3.3 Rezultati

V tabeli 6.2 so podane klasifikacijske točnosti algoritma C4.5 na vseh sedmih množicah podatkov, opisanih v razdelku 6.3.1. Klasifikacijske točnosti so izračunane kot povprečje 10-ih zagonov 10-kratnega prečnega preverjanja. Brez uravnavanja kotnih odvisnosti (množica B) je klasifikacijska točnost 67,38%, v vseh ostalih primerih pa med 73 % in 79 %. Razlika med rezultati na neobdelanih podatkih in vseh ostalih množicah podatkov je statistično značilna. Uravnavanje kotnih odvisnosti je torej



zelo pomemben faktor pri akustični klasifikaciji morskega dna in pomembno vpliva na rezultate.

V tabeli 6.3 so z znakom 'x' označene statistično značilne razlike med množicama podatkov. Opazimo lahko, da ni statistično značilnih razlik med klasifikacijskimi točnostmi pri uporabi podatkov z globalnim, lokalnim ali obema pristopoma k uravnavanju kotnih odvisnosti. Dobimo pa statistično značilne razlike med pristopoma z uporabo lokalnega uravnavanja standardnih odklonov in ostalimi pristopi. Klasifikacijska točnost se v obeh primerih uporabe lokalnega uravnavanja standardnih odklonov ( $L \rightarrow L\_std$  in  $G\_L \rightarrow G\_L\_std$ ) poveča za dobrih 5 %, v primerjavi z enako množico podatkov brez uravnavanja standardnih odklonov. V grobem lahko torej rezultate razdelimo na dva dela, skupino kjer je na podatkih uporabljeno le (poljubno) uravnavanje kotnih odvisnosti in skupino kjer je uporabljeno lokalno uravnavanje standardnih odklonov. Prva je statistično značilno boljša od množice podatkov brez izločanja kotnih odvisnosti, druga pa statistično značilno boljša od prve. V drugi skupini imamo podatke z oznako  $L\_std$  in  $G\_L\_std$ . Na obeh je uporabljen lokalni pristop k uravnavanju kotnih odvisnosti in lokalno uravnavanje standardnih odklonov, z razliko, da je pri  $G\_L\_std$  pred tem uporabljen še globalen pristop k uravnavanju kotnih odvisnosti. Slednja pristopa sta statistično značilno boljša od ostalih petih.  $G\_L\_std$  kljub drugačnim pričakovanjem ni dal boljših rezultatov od  $L\_std$ . Ker med pristopoma ni statistično značilnih razlik, pristop  $G\_L\_std$  pa je po klasifikacijskih točnostih nekoliko slabši poleg tega pa je počasnejši za izračun, bomo v nadaljevanju tega dela uporabljali pristop  $L\_std$ . Dodatna prednost pristopa  $L\_std$  je tudi v tem, da je primeren za morebitno implementacijo klasifikacije morskega dna v realnem času - že med izvajanjem meritev. V primeru take implementacije bi bilo lokalno delovanje velika prednost.

Množica podatkov	Klasifikacijska točnost
B	67.38 %
G	73.85 %
G_std	74.39 %
L	74.08 %
L_std	79.48 %
G_L	73.74 %
G_L_std	79.29 %

Tabela 6.2: Primerjava klasifikacijskih točnosti na podatkih obdelanih z različnimi metodami uravnavanja kotnih odvisnosti.

Na sliki 6.8 je prikazana primerjava informacijskega prispevka značilk v štirih različnih množicah podatkov B, G, L in  $L\_std$ . Informacijski prispevek se oceni enkrat na celotni množici podatkov, zato na tem mestu ne uporabljamo statističnih testov značilnosti razlik. Testiranje kakovosti značilk smo uporabili za oceno, kako pristop k uravnavanju kotnih odvisnosti vpliva na pomembnost posamezne značilke. V grafu vidimo, da globalno in lokalno uravnavanje vplivata le na značilko  $i$ , ki predstavlja intenziteto. Vse ostale značilke ostanejo bolj kot ne nespremenjene. Večja pomembnost značilke  $i$  je bila pričakovana, saj se z uravnavanjem kotnih odvisnosti izloči veliko šuma in tako značilka postane bolj informativna. Izboljšava

	B	G	G.L	G.std	L	L.std	G.L.std
B		x	x	x	x	x	x
G	x					x	x
G.L	x					x	x
G.std	x					x	x
L	x					x	x
L.std	x	x	x	x	x		
G.L.std	x	x	x	x	x		

Tabela 6.3: Prikaz statistično značilnih razlik v klasifikacijski točnosti algoritma C4.5 na različno predobdelanih podatkih. Oznaka G pomeni globalno uravnavanje kotnih odvisnosti, L lokalno uravnavanje kotnih odvisnosti, std pa upoštevanje standardnih odklonov.

je precejšnja. Informacijski prispevek značilke  $i$  se izboljša za kar 12 %, iz 0.375 na 0.42. (slika 6.8).

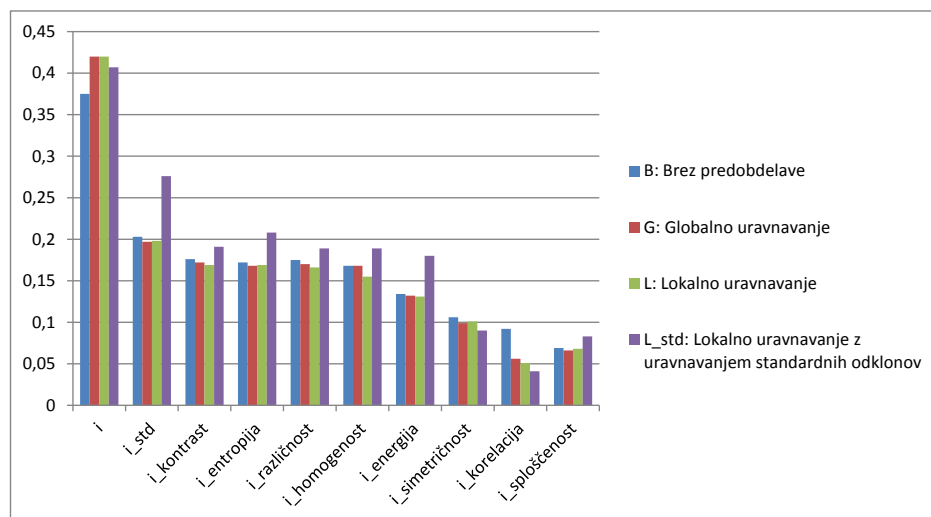
Pristop z lokalnim uravnavanjem kotnih odvisnosti in uravnavanjem standardnih odklonov izboljša večino značilnk. Od 10 ih ovrednotenih značilnk jih je 7 najbolj ocenjenih na množici podatkov L.std. Lokalno uravnavanje standardnih odklonov bistveno izboljša predvsem značilko  $i_{std}$ . Največjo razliko vidimo pri povečanju informacijskega prispevka slednje. Na sliki 6.8 je sicer razvidno, da lokalno uravnavanje standardnih odklonov rahlo negativno vpliva na informacijski prispevek značilke  $i$ , vendar so izboljšave pri ostalih značilnkah tako velike, da dobimo na množici podatkov L.std statistično značilno boljše rezultate kot na množicah B, G in L (tabela 6.3).

Iz dobljenih rezultatov lahko zaključimo, da je lokalno uravnavanje kotnih odvisnosti v kombinaciji z lokalnim uravnavanjem standardnih odklonov najprimernejši pristop za predobdelavo podatkov. Pokazali smo, da tak pristop statistično značilno izboljša rezultate klasifikacije in pokazali, da izboljša ocene večine značilnk.

## 6.4 Ovrednotenje in izbor značilnk

Za ovrednotenje in izbor značilnk smo uporabili enake podatke kot pri ovrednotenju različnih pristopov k uravnavanju kotnih odvisnosti. Območje je prikazano na sliki 6.7. Izmerjeno območje je dolgo 900 metrov. Najplitvejša izmerjena točka je globoka 1,3 metra, najgloblja pa 11,4 metra. Na območju se nahajajo štirje tipi morskega dna - mulj, kamnito dno, izdanki kamnin in morski travniki. Iz množice podatkov smo izbrali podmnožico tako, da so vsi štirje tipi morskega dna enakomerno zastopani. Z enakomerno zastopanostjo vseh štirih razredov (vsak razred predstavlja 25% primerov v množici podatkov) se izognemo prevelikemu vplivu posameznega tipa morskega dna. Za vsak tip tal smo naključno izbrali 2500 točk. Končna množica podatkov vsebuje podatke o 10.000 naključno izbranih točkah, kjer so vsi štirje tipi morskega dna enakomerno zastopani.

Za vsako točko smo izračunali vseh 20 značilnk, opisanih na začetku poglavja. Značilke smo izračunali za celice velikosti 1x1 meter, 2x2 metra in 4x4 metre. Zaradi različnih gostot podatkov in veliko šuma v podatkih smo predvidevali, da bi lahko nekatere značilke dale boljše rezultate, če bi pri izračunu upoštevali večjo



Slika 6.8: Informacijski prispevek značilk za 4 različno predobdelane množice podatkov.

okolico. Gostota podatkov pri sonarskih meritvah se spreminja z globino kot tudi z vpadnim kotom signala. Včasih se lahko zgodi da je v okolici neke točke zelo majhno število podatkov in šum prevlada nad dejansko informacijo v podatkih. Večja okolica zajame več podatkov in zmanjša vpliv šuma na rezultate. Za vsako koordinato (meterska natančnost) tako dobimo 60 značilk, ki jih bomo v nadaljevanju ovrednotili.

### 6.4.1 Mere ocenjevanja značilk in uporabljene metode

#### Informacijski prispevek

Informacijski prispevek je ena najosnovnejših mer pomembnosti značilke. Informacijski prispevek značilke je definiran kot prispevana informacija značilke za določitev njegove vrednosti. Da bi lahko izračunali informacijski prispevek, moramo najprej definirati entropijo in pogojno entropijo pri dani vrednosti značilke.  $H$ - entropija:

$$H = - \sum_k p(k) \log_2(p(k))$$

kjer je  $k$  število razredov in  $p(k)$  verjetnost  $k$ -tega razreda.  $H_{res}(A)$ - pogojna entropija razreda pri dani vrednosti značilke:

$$H_{res}(Z) = - \sum_j p(k) \sum_k p(k|j) \log_2 p(k|j)$$

kjer je  $A$  izbrana značilka,  $k$  število razredov,  $j$  število podmnožic pri delitvi značilke  $Z$  in  $p(k|j)$  verjetnost  $k$ -tega razreda pri dani vrednosti značilke  $Z$ . Sedaj lahko definiramo informacijski prispevek značilke  $A$  kot razliko med entropijo pred delitvijo in entropijo po delitvi podatkov glede na vrednosti značilke  $Z$ :

$$InfoGain(Z) = H - H_{res}(Z)$$

kjer je  $Z$  izbrana značilka.

Informacijski prispevek je navadno dobra mera pomembnosti značilke, ni pa idealna. Problem se pojavi pri značilkah z velikim številom različnih vrednosti, saj informacijski prispevek značilke kvečjemu raste s številom vrednosti. Zato se velikokrat namesto informacijskega prispevka uporablja razmerje informacijskega prispevka, ki je opisano v naslednjem razdelku.

#### Razmerje informacijskega prispevka

Razmerje informacijskega prispevka je normalizacija informacijskega prispevka z entropijo vrednosti značilke in odpravlja problem precenjevanja značilk z velikim številom različnih vrednosti, ki je slaba lastnost informacijskega prispevka.

$$H = - \sum_k p(j) \log_2(p(j))$$

kjer je  $j$  število vrednosti atributa  $Z$  in  $p(j)$  verjetnost  $j$ -te vrednosti atributa.

Nato lahko izračunamo razmerje informacijskega prispevka:

$$GainRatio(Z) = \frac{InfoGain(a)}{H(Z)}$$

kjer je  $InfoGain(Z)$  informacijski prispevek in  $H(Z)$  entropija vrednosti značilke.

## Ocenjevanje značilk z naključnimi gozdovi

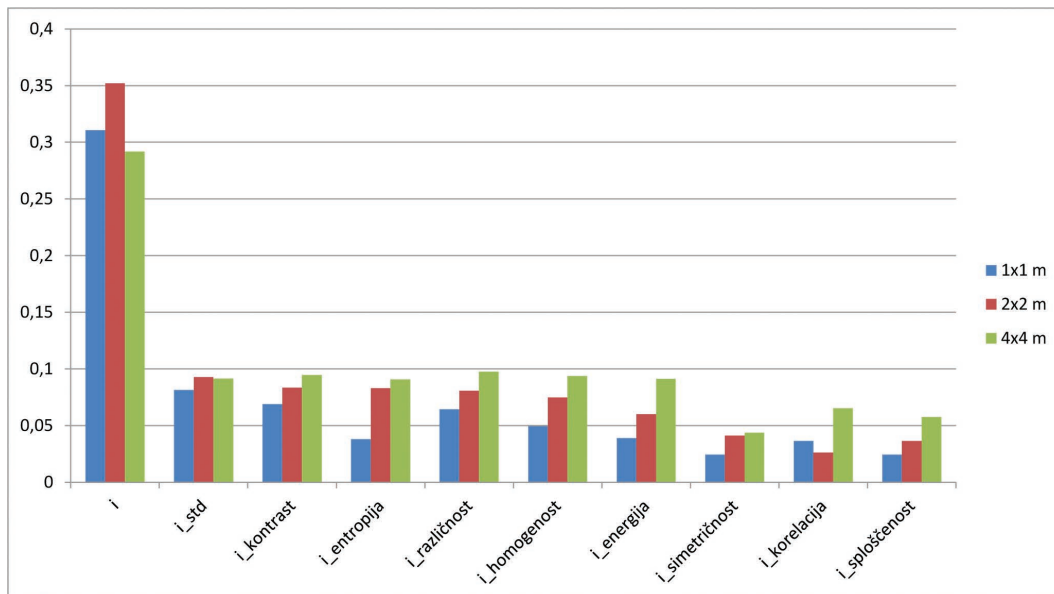
Razmerje informacijskega prispevka in sam informacijski prispevek sta tako imenovani kratkovidni meri ocenjevanja značilk. Kratkovidni sta zato, ker obravnavata vsako značilko posebej in predpostavljata pogojno neodvisnost značilk. L. Breiman [15] je predlagal možnost uporabe naključnih gozdov kot ne-kratkovidne mere pomembnosti značilke - torej mere, ki zna delno upoštevati tudi odvisnosti med značilkami. Ocena pomembnosti značilke z naključnimi gozdovi temelji na ideji, da naključno spreminjanje vrednosti pomembne značilke v veliki meri vpliva na razvrstitev primerov v razred, medtem ko spreminjanje vrednosti nepomembne značilke nanj ne vpliva. Algoritem ocenjevanja pomembnosti značilk z naključnimi gozdovi združi ocene značilk izbranega števila dreves. Pomembnosti značilk za posamezno drevo so izračunane kot: pravilno razvrščeni primeri minus pravilno razvrščeni primeri ko vrednost značilke naključno spremenimo. Vsota posameznih ocen se deli s številom dreves in množi s 100. Natančen postopek izračuna ocene značilk z naključnimi gozdovi je zapleten in presega okvire tega dela. V delu smo uporabili implementacijo v programskem paketu Weka [54], postopek pa je opisan v delu [15].

## ReliefF

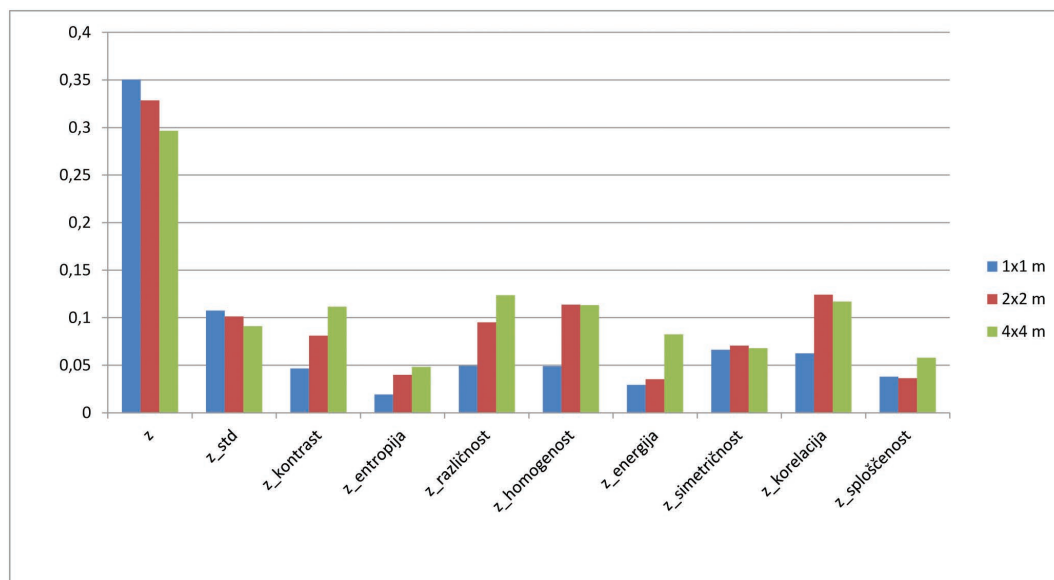
Relief je eden izmed boljših algoritmov za ocenjevanje značilk. Medtem ko večina mer za ocenjevanje značilk predpostavlja pogojno neodvisnost značilk pri danem razredu, Relief [70] učinkovito rešuje problem odvisnosti značilk. Podobno kot ocenjevanje z naključnimi gozdovi, je tudi ReliefF ne-kratkovidna mera ocenjevanja značilk. Osnovna ideja algoritma je, da za vsak učni primer poišče najbolj podobne primerme iz istega in najbolj podobne primere iz nasprotnega razreda ter tako oceni lokalno obnašanje posamezne značilke. Algoritem ReliefF, ki ga bomo uporabili v tem delu, je razširitev prvotnega algoritma Relief tako, da je uporaben za večrazredne probleme. V tem delu se ne bomo poglobljali v natančno definicijo algoritma ReliefF, le-ta je na voljo v delu [76].

### 6.4.2 Izbira velikosti celic

Kot smo omenili v uvodu poglavja, iz posameznega akustičnega odboja ne dobimo dovolj informacije, da bi sklepali na tip tal. Ker so podatki umeščeni v prostor, lahko izkoristimo prostorsko komponento. Prostorska komponenta nam omogoči, da podatke združimo v celice in na njih izvajamo prostorske analize. Velikost celice moramo določiti objektivno. Značilke smo izračunali za celice velikosti 1x1 meter, 2x2 metra in 4x4 metre. Za vse velikosti celic smo izračunali vse možne značilke in jih ocenili na podlagi razmerja informacijskega prispevka. Primerjave ocene posameznih značilk po velikostih celic so prikazane na slikah 6.9 in 6.10. Značilki  $i$  (intenziteta odboja) in  $z$  (globina) sta veliko bolje ocenjeni od ostalih značilk. Značilki  $i$  in  $i\_std$  sta najbolj ocenjeni pri uporabi celice velikosti 2x2 metra, značilki  $z$  in  $z\_std$  sta najbolj ocenjeni pri uporabi celice velikosti 1x1 metra, vse ostale izpeljane značilke pa pri celici velikosti 4x4 metre. Glavni cilj hidrografskih meritev so natančne izmere globin. Ker so globine zelo natančno izmerjene z večanjem celice le izgubljammo informacijo. Tudi standardni odklon globin nam da kakovostno informacijo o raz-



Slika 6.9: Razmerje informacijskega prispevka za značilke izpeljane iz intenzitet odbojev. Primerjava različnih velikosti celic.



Slika 6.10: Razmerje informacijskega prispevka za značilke izpeljane iz globin. Primerjava različnih velikosti celic.

gibanosti terena. Po drugi strani so intenzitete stranski produkt meritev in so zelo šumne. Veliko truda smo vložili v izločanje šuma in odpravljanje odvisnosti tako nastavitvev sistema kot vpliva vpadnega kota na intenzitete odbojev, vendar šuma v celoti ni mogoče odpraviti. Rezultat tega je ocena značilk na sliki 6.9, kjer sta značilki  $i$  in  $i_{std}$  najbolje ocenili pri uporabi celice velikosti 2x2 metra. Z večjim vzorcem (večjo celico), dobimo več podatkov in imamo manj vpliva šuma na rezultat izračuna. Pri celici velikosti 4x4 metre pa so podatki že preveč posplošeni, upoštevana okolica je prevelika, da bi nam dala natančno informacijo o obravnavani točki. Vse ostale kompleksnejše značilke so pričakovano nekoliko boljše pri izračunu na večjem vzorcu. Pri velikosti celice 1x1 meter se pogosto zgodi da računamo značilke na le nekaj (tudi manj kot 10) odbojih, kar je bistveno premalo za smiselne izračune matrike sopojavnosti sivin ali analize histogramov. Velikost celice 4x4 metre nam zagotavlja od 100 do 500 odbojev na celico in tako veliko bolj zanesljive izračune.

V nadaljevanju bomo izračune globin in standardnih odklonov globin izvajali na celicah velikosti 1x1 meter, izračune povprečnih intenzitet in standardnih odklonov intenzitet na celicah velikosti 2x2 metra, vse ostale izračune pa na celicah velikosti 4x4 metre.

### 6.4.3 Izbor značilk

Z izborom značilk iz množice značilk izločimo tiste, ki so manj pomembne in za katere pričakujemo, da ne bodo bistveno izboljšale rezultatov klasifikacije. Prednost manjše množice značilk je tudi bolj pregleden model (pri algoritmih ki omogočajo razlago svojih odločitev) in predvsem hitrejša gradnja modela. Ker imamo v našem primeru opravka z velikimi količinami podatkov, je izbor značilk pomemben predvsem s stališča hitrejše gradnje modela pri uporabi metodologije v praksi. Za izbor značilk lahko uporabimo dve metodi:

- Metodo na osnovi notranje optimizacije
- Filtrske metodo

Metode na osnovi notranje optimizacije uporabljajo klasifikacijski algoritem kot podprogram. Izbirni postopek za vsako podmnožico množice značilk zgradi klasifikacijski model na podlagi katerega klasificira podatke. Uspešnost klasifikacije da oceno podmnožice značilk. Pogosto se uporablja požrešno metodo za preiskovanje stanj zaradi manjše časovne zahtevnosti. Najbolj pereča slabost metode je velika računaska zahtevnost.

Filtrske metode izberejo značilke na podlagi mere za ocenjevanje značilk. Značilke se ovrednoti na podlagi izbrane mere in izbere določeno število najbolje ovrednotenih značilk. Filtrske metode so časovno manj zahtevne in so splošne - neodvisne od klasifikacijskega algoritma.

V tem delu smo uporabili filtrske metode za izbor značilk. Uporabili smo tri mere za ocenjevanje značilk in značilke razvrstili na podlagi vsote rangov po ocenah z vsako od treh mer. Vsoto rangov dobimo tako da po vsaki ob uporabljenih mer značilke razvrstimo od najboljše (rang 1) do najslabše (rang 20), nato pa seštejemo

vse tri range. Če je značilka po vseh merah za ovrednotenje značilk uvrščena na tretje mesto bo torej njena vsota rangov  $3 + 3 + 3 = 9$ .

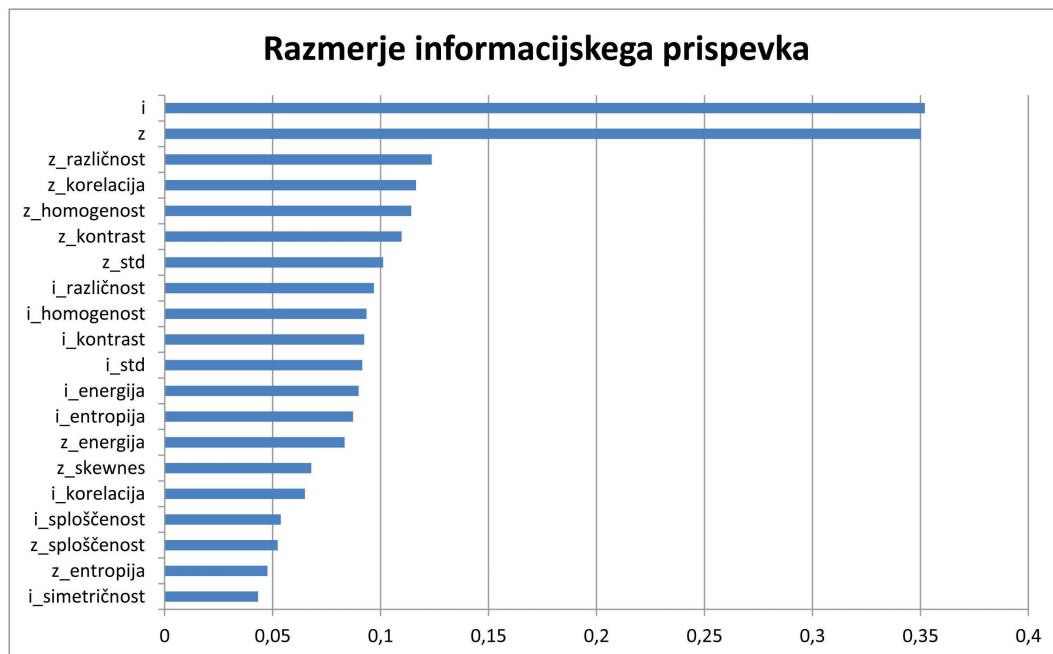
Osnovna mera za ocenjevanje pomembnosti značilk je informacijski prispevek, oziroma razmerje informacijskega prispevka. Poleg razmerja informacijskega prispevka smo značilke primerjali še na podlagi ocen algoritmov ReliefF in naključnih gozdov. Razmerja informacijskih prispevkov posameznih značilk so podana na sliki 6.11, ocene po metodi ReliefF na sliki 6.12 in ocene po metodi naključnih gozdov na sliki 6.13. Vsote rangov značilk so predstavljene v tabeli 6.4. Po vseh metodah sta daleč najboljše ocenjeni značilki *i* in *z*, torej povprečne intenzitete in povprečne globine. To se odraža tudi na vsoti rangov, kjer ima značilka *i* vsoto 4 in značilka *z* vsoto 5. S precej slabšimi ocenami jim sledijo standardni odkloni globin in intenzitet (*z\_std*, *i\_std*) in značilke dobljene iz matrike sopojavnosti sivin podatkov o globinah (*z\_homogenost*, *z\_kontrast*, *z\_korelacija*, *z\_različnost*). Značilke dobljene iz matrike sopojavnosti sivin podatkov o intenzitetah so po vseh uporabljenih kriterijih ocenjene slabo, zato jih bomo izločili iz podatkov. Prav tako so slabo ocenjeni višji momenti (asimetrija in sploščenost), tako pri izračunu iz globin kot intenzitet. Vse štiri značilke smo izločili iz nadaljnje obdelave. Zadnji dve značilki, ki nista dali zelenih rezultatov sta entropija globin in entropija intenzitet. Obe sta se po vseh ocenah uvrščali v zadnji del po pomembnosti značilk in bosta zato izločeni iz nadaljnje obdelave.

Ravno prvih osem značilk iz tabele 6.4 smo, poleg slabših ocen ostalih značilk, izbrali tudi iz praktičnih razlogov. Značilke, ki jim po ocenah sledijo, so teksturne značilke izračunane za intenzitete odbojev, za izračun slednjih pa moramo izračunati matriko sopojavnosti sivin. Z izločitvijo teh značilk torej prihranimo pri časovni zahtevnosti izračuna značilk.

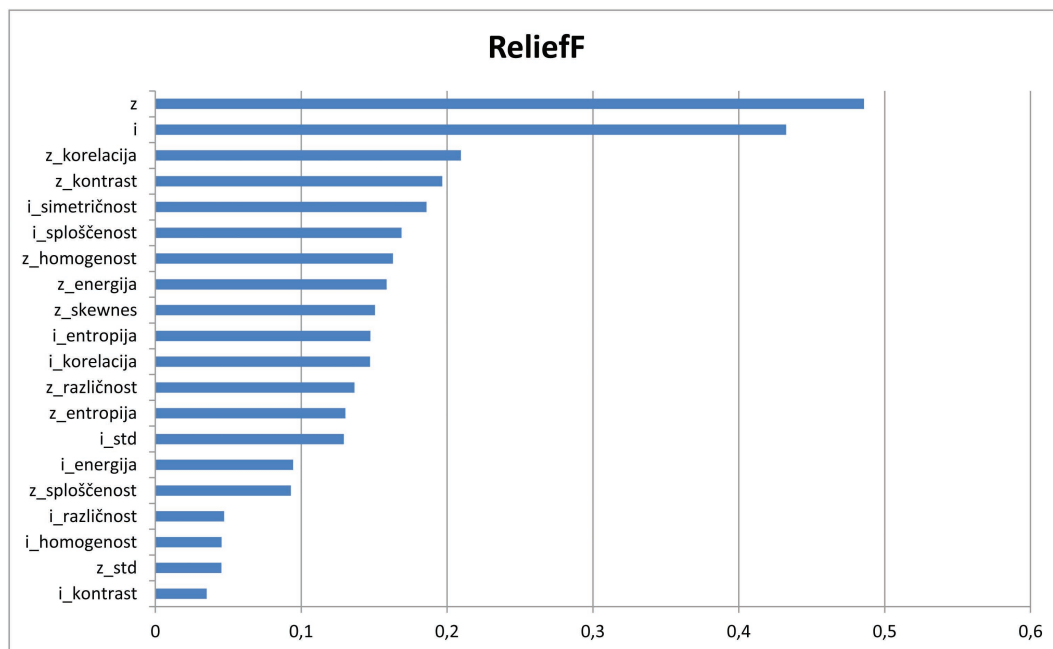
Značilka	Vsota rangov
<i>i</i>	4
<i>z</i>	5
<i>z_kontrast</i>	15
<i>z_homogenost</i>	22
<i>z_različnost</i>	27
G.L	27
<i>i_std</i>	29
<i>z_std</i>	29
<i>i_entropija</i>	31
<i>i_simetričnost</i>	32
<i>i_energija</i>	33
<i>i_različnost</i>	36
<i>i_sploščenost</i>	36
<i>z_energija</i>	41
<i>z_sploščenost</i>	41
<i>z_entropija</i>	41
<i>i_korelacija</i>	43
<i>i_homogenost</i>	45
<i>i_kontrast</i>	45
<i>z_sploščenost</i>	48

Tabela 6.4: Vsota rangov značilk po ocenah s tremi merami za ovrednotenje značilk.

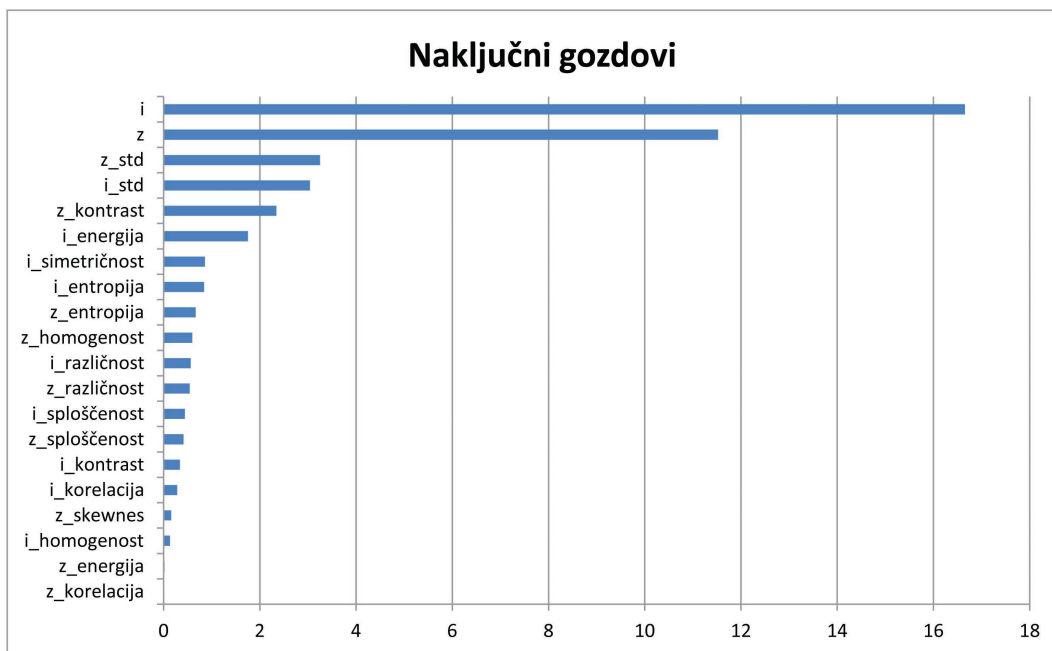




Slika 6.11: Razmerje informacijskega prispevka značilk.



Slika 6.12: Ocene značilk po metodi ReliefF.



Slika 6.13: Ocene značilk po metodi naključnih gozdov.

## 6.5 Zaključek

V začetku poglavja opisali podatke in značilke, nato smo ovrednotili pristope h kompenzaciji kotnih odvisnosti in za nadaljnje obdelave izbrali pristop L\_std. Za vsako točko v podatkih smo izračunali 60 značilk. Za vsako značilko smo ocenili najprimernejšo velikost celice in na podlagi ocene treh različnih mer pomembnosti značilk izbrali naslednje značilke:

- i : intenziteto odboja, izračunano na celici velikosti 2x2 metra
- z: povprečno globino, izračunano na celici velikosti 1x1 meter
- i\_std: standardni odklon intenzitete, izračunano na celici velikosti 2x2 metra
- z\_std: standardni odklon globine, izračunano na celici velikosti 1x1 metra
- lastnosti matrike sopoljavnosti sivin, izračunane na podatkih o globinah z velikostjo celice 4x4 metre:
  - z\_homogenost
  - z\_kontrast
  - z\_korelacija
  - z\_razlicnost

Skupno smo torej izbranih 8 značilk, ki jih bomo v nadaljevanju uporabljali za akustično klasifikacijo morskega dna.

## Poglavje 7

# Klasifikacija morskega dna

Eno najpogosteje uporabljenih področij strojnega učenja je klasifikacija ali razvrščanje. Naloga klasifikatorja je za objekt, opisan z množico značilk, določiti, kateremu izmed možnih razredov pripada. Zato, da lahko klasifikator določi razred, mora imeti na nek način predstavljeno diskretno funkcijo, ki preslika prostor značilk v razred. Ta funkcija je lahko podana vnaprej ali pa je naučena iz podatkov – primerov problemov rešenih v preteklosti. Naloga učnega algoritma je torej iz množice vzorcev z znanimi njim pripadajočimi razredi zgraditi model (izpeljati pravila, izpeljati preslikavo, izračunati funkcijo, zgraditi odločitveno drevo), ki ga lahko uporabimo pri klasifikaciji. Klasifikatorje ločimo glede na način predstavitve klasifikatorjeve funkcije. Najbolj pogosti klasifikatorji so: odločitvena drevesa, odločitvena pravila, naivni Bayes-ov klasifikator, Bayes-ove verjetnostne mreže, klasifikator z najbližjimi sosedi, linearna diskriminantna funkcija, logistična regresija ter klasifikator po metodi podpornih vektorjev (SVM). V našem primeru imamo osem značilk, predstavljenih v prejšnjem poglavju, razred pa je tip morskega dna. Za učenje potrebujemo množico podatkov, kjer so znane vrednosti značilk in vrednost razreda. Potrebujemo torej področje, kjer smo opravili hidrografske meritve in zbrali akustične podatke, poleg tega pa tudi z metodami opisanimi v poglavju 4 zbrali podatke o dejanskih tipih morskega dna. Na podlagi take učne množice lahko zgradimo model, ki ga nato uporabimo za klasifikacijo nepoznanega območja - torej območja za katerega imamo le akustične podatke.

### 7.1 Uporabljene metode in tehnike

Obstaja veliko algoritmov za klasifikacijo in vsak izmed njih ima svoje prednosti in slabosti, na podlagi katerih lahko predvidevamo, kateri algoritem je bolj in kateri manj primeren za naš problem. Objektivno oceno posameznih klasifikatorjev pa lahko dobimo le na podlagi uporabe ustrezne metode ovrednotenja klasifikatorjev. V nadaljevanju so opisani uporabljeni algoritmi za klasifikacijo, njihove prednosti oziroma slabosti pri aplikaciji na akustičnih podatkih in nenazadnje uporabljene metode ovrednotenja klasifikatorjev.

### 7.1.1 Uporabljeni algoritmi za klasifikacijo

Za klasifikacijo lahko uporabljamo nadzorovane ali nenadzorovane klasifikacijske algoritme. Pri nenadzorovanih algoritmih se podatke razvrsti v skupine statistično podobnih enot. Nadzorovana klasifikacija poteka tako, da se algoritem najprej na poznanem delu območja nauči razločiti tipe tal in nato to znanje uporabi na preostalem območju. V tem poglavju ovrednotimo algoritme nadzorovane klasifikacije na realnih podatkih o morskem dnu. Na pripravljenih podatkih ovrednotimo naslednje metode strojnega učenja:

- OneR klasifikator,
- Naivni Bayesov klasifikator NB,
- metoda podpornih vektorjev SVM,,
- metoda najbližjih sosedov kNN,
- Odločitven pravila CN2
- Odločitvena drevesa C4.5,
- metoda naključnih gozdov RF.

Uporabili smo implementacije algoritmov za strojno učenje v okolju ORANGE v kombinaciji s programskim jezikom Python in okolju WEKA v kombinaciji s programskim jezikom Java. Vsem algoritmom smo parametre nastavili pred ovrednotenjem in jih med postopkom ovrednotenja nismo več spreminjali.

#### OneR klasifikator

OneR [59] (okrajšava za ‘one rule’ - eno pravilo) je preprost, vendar natančen klasifikacijski algoritem, ki generira eno pravilo za vsako vrednost značilke v podatkih, nato pa izbere pravilo z najmanjšo skupno napako. Delovanja algoritma OneR je predstavljeno s psevdokodo Algoritem 2. Algoritem preizkuša vsako značilko posebej. Za vsako vrednost značilke izpelje eno pravilo in prešteje napake klasifikacije po izpeljanem pravilu. Izbrana značilka je značilka z najmanj napakami. Prednost algoritma je za človeka razumljiv model. Kljub temu pa da OneR algoritem zaradi enostavnosti običajno nekoliko slabše rezultate. Uporabili smo ga predvsem zato da vidimo katero značilko izbere kot najbolj informativno in za oceno, kakšno klasifikacijsko točnost lahko dosežemo z uporabo le ene značilke.

#### Naivni Bayesov klasifikator NB

Bayesov klasifikator [77] izračuna pogojne verjetnosti za vsak razred pri danih vrednostih značilk za dani novi primer, ki ga želimo klasificirati. Bayesov klasifikator, ki natančno izračuna pogojne verjetnosti razredov je optimalen, v tem smislu, da minimizira pričakovano napako. Ker natančno računanje vseh pogojnih verjetnosti ni vedno mogoče in bi bilo pogosto prezahtevno, se uporabljajo določene predpostavke. Naivni Bayesov klasifikator predpostavi pogojno neodvisnost značilk pri

**Algoritem 2** OneR klasifikator

---

```

1: procedure ONER
2:   for vsako značilko v naboru značilk do
3:     for vsako vrednost značilke do
4:       preštej število pojavitev posameznega razreda
5:       izberi najpogostejši razred
6:       naredi pravilo, ki dodeli izbrani razred obravnavani vrednosti značilke
7:     end for
8:     Izračunaj skupno napako pravil za obravnavano značilko
9:   end for
10:  Izberi značilko z najmanjšo skupno napako
11: end procedure

```

---

danem razredu. To omogoči, da učna množica običajno zadošča za zanesljivo oceno vseh potrebnih verjetnosti za izračun končne pogojne verjetnosti vsakega razreda. Uporabili bomo naslednje oznake:

$P(r_k)$  apriorna verjetnost razreda  $r_k$

$V = \langle v_1, \dots, v_a \rangle$  vektor vrednosti za vse značilke

$P(r_k|V)$  verjetnost razreda pri danih vrednostih značilk

$P(r_k|v_i)$  verjetnost razreda pri dani vrednosti  $i$ -te značilke

Verjetnosti razredov se računa po enačbi, ki predpostavlja neodvisnost značilk:

$$P(r_k|V) = P(r_k) \prod_{i=1}^a \frac{P(r_k|v_i)}{P(r_k)} \quad (7.1)$$

Naloga učnega algoritma je s pomočjo učne množice podatkov aproksimirati verjetnosti na desni strani enačbe. Znanje naivnega Bayesovega klasifikatorja je torej tabela aproksimacij apriornih verjetnosti razredov  $P(r_k)$ ,  $k = 1, \dots, n_0$  in tabela pogojnih verjetnosti razredov  $r_k$ ,  $k = 1, \dots, v_0$  pri dani vrednosti  $v_i$  značilke  $A_i$ ,  $i = 1, \dots, a$ :  $P(r_k|v_i)$ .

Za ocenjevanje apriornih verjetnosti se pogosto uporablja Laplaceov zakon zaporednosti

$$P(r_k) = \frac{N_k + 1}{N + n_0} \quad (7.2)$$

kjer je  $N_k$  število učnih primerov iz razreda  $r_k$ ,  $N$  število vseh učnih primerov in  $n_0$  število vrednosti razreda.

Za ocenjevanje pogojnih verjetnosti se pogosto uporablja  $m$ -ocena

$$P(r_k|v_i) = \frac{N_{k,i} + mP(r_k)}{N_i + m}, \quad (7.3)$$

kjer je  $N_{k,i}$  število učnih primerov iz razreda  $r_k$  in z vrednostjo  $i$ -te značilke  $v_i$ ,  $N_i$  število vseh učnih primerov z vrednostjo  $i$ -te značilke  $v_i$  in  $m$  parameter, ki ga nastavi uporabnik.

Odločanje naivnega Bayesovega klasifikatorja in vpliv vrednosti posamezne značilke na verjetnost ciljnega razreda je možno predstaviti z nomogrami [94].

Ker deluje Naivni Bayesov klasifikator le na diskretnih značilkah, smo značilke diskretizirali z metodo MDL diskretizacije [64]. Prednost Naivnega Bayesovega klasifikatorja je njegova hitrost in možnost predstavitve naučenega modela z nomogrami. Slabost je slabša klasifikacijska točnost, ko ne drži predpostavka o pogojni neodvisnosti značilk.

### Metoda podpornih vektorjev - SVM

Izhodišče za nastanek metode podpornih vektorjev [109] (SVM - *support vector machine*) je množica učnih primerov, za katere je znano kateremu razredu pripadajo. Vsak primer predstavimo z vektorjem v vektorskem prostoru. Naloga SVM je poiskati v tem  $n$  dimenzionalnem prostoru hiperravnino (hiperploskev), ki ločuje primere iz različnih razredov. Razdaljo vektorjev, ki ležijo najbližje hiperravnini, pri tem maksimiramo. To široko prazno območje med razredi nam kasneje omogoča, da lahko čim bolj zanesljivo razvrščamo tudi primere, ki niso čisto enaki učnim primerom.

Pri postavljanju hiperravnine ni nujno upoštevati vseh učnih vektorjev. Vektorji, ki so daleč od hiperravnine oz. so skriti za fronto ostalih ne vplivajo na lego hiperravnine. Torej je lega hiperravnine odvisna le od njej najbližjih vektorjev. Tem vektorjem rečemo podporni vektorji in tudi metoda ima od tod ime.

Pri uporabi hiperravnine je možno izvesti čisto ločitev le takrat, ko so predmeti linearno ločljivi. To pa v resnični uporabi običajno ni slučaj. Ko vektorji niso linearno ločljivi uporabimo zvijačo. Vektorje transformiramo tako, da jim povečamo dimenzijo. Če dimenzijo dovolj povečamo postanejo vsi razredi vektorjev linearno ločljivi. Pri tem nastaneta dve težavi:

1. Transformiranje vektorjev v prostor z višjo dimenzijo je računsko zahtevna operacija.
2. Inverzna transformacija hiperravnino v prostoru z višjo dimenzijo spremeni v zelo zapleteno hiperploskev v prostoru z nižjo dimenzijo.

Če za transformacijo uporabimo posebne funkcije, ki se jim reče jedro, se tem težavam izognemo. Transformacijo in inverzno transformacijo lahko opravimo ne da bi jo dejansko računsko izvedli. Tudi tukaj zadošča uporaba le enega dela učnih vektorjev (podporni vektorji) za popolni opis meje med razredi.

Tako linearno, kot tudi nelinearno ločljive vektorje lahko nadgradimo. S tem postanejo bolj prilagodljivi. Dodatne spremenljivke zavestno omogočijo klasifikatorju napačno razvrstitev nekaj primerov. Vsako tako napačno razvrstitev »kaznuje«. Na ta način se izognemo prekomernemu učenju in zmanjšamo število potrebnih podpornih vektorjev.

SVM se je že na veliko domenah izkazal za zelo učinkovit algoritem. Uporaben je predvsem ko potrebujemo le natančen klasifikator, ne zanima pa nas kako je leta prišel do odločitve. Slabost metode SVM je namreč v tem, da je predstavitev naučenega modela zelo otežena.

V tem delu smo uporabili algoritem SVM z RBF jedrom in nastavitvijo parametra  $\gamma = 0.01$  in paramtera  $C = 1$ [109].

### Algoritem k-najbližjih sosedov kNN

Algoritem kNN [77] (k-nearest neighbours) je klasifikator, ki klasificira na podlagi najbližjih sosedov. Najpreprostejša različica algoritma najbližjih sosedov kot znanje uporablja kar množico vseh učnih primerov. Učni algoritem si torej le zapomni vse primere. Ker učenja skorajda ni, pravimo tej vrsti učenja tudi leno učenje. Z lenim učenjem je prihranjen čas za gradnjo modela je pa zato bistveno počasnejše klasificiranje novih primerov. Pri klasifikaciji novega primera se iz učne množice poišče  $k$  najbolj podobnih (najbližjih) primerov. Nov primer klasificiramo v razred, ki mu pripada večina najbližjih primerov. Pri tem je potrebno zaradi ustrezne metrike v prostoru značilk normalizirati vrednosti zveznih značilk. Pri klasifikaciji se pogosto uporablja tudi uteževanje primerov glede na njihovo oddaljenost od primera, ki ga klasificiramo. Uporablja se različne razdalje, za zvezne značilke najpogosteje evklidsko. V našem primeru je vseh osem značilk zveznih zato smo uporabili smo evklidsko razdaljo.

Algoritem smo vključili na seznam algoritmov za ovrednotenje predvsem zato ker, za razliko od Naivnega Bayesa, nima težav z odvisnostmi med značilkami. Problem algoritma kNN je šum v podatkih, predvsem če je izbrani parameter  $k$  (število sosedov, ki se jih upošteva pri določanju razreda novega primera) zelo majhen. Uporabili smo nastavek  $k = 50$ , torej je o razredu vsakega primera odločalo 50 najbližjih sosedov.

### Odločitvena pravila CN2

Učenje pravil najpogosteje uporabljamo v kontekstu učenja klasifikacijskih in povezovalnih pravil. Pri učenju klasifikacijskih pravil gre običajno za nadzorovano učenje (supervised learning) usmerjeno k odkrivanju množic pravil, ki jih kasneje lahko uporabimo za klasifikacijo novih primerov in/ali napovedovanje. Takemu pristopu pravimo tudi napovedna indukcija (predictive induction).

CN2 [24] je algoritem za indukcijo odločitvenih pravil oblike “če Cond potem Class”, kjer predstavlja Cond konjunkcijo značilk oblike značilka–vrednost1, Class pa vrednost razreda.

CN2 sestavljata dve proceduri: nizko–nivojska iskalna procedura, ki uporablja iskanje v snopu (beam search) za iskanje najboljšega pravila ter visoko–nivojska kontrolna procedura, ki v zanki izvaja nizko–nivojsko iskanje in tako sestavi končno množico pravil. Nizko–nivojska iskalna procedura uporablja iskanje v snopu ter klasifikacijsko točnost kot mero kakovosti posameznega pravila. V primeru indukcije neurejene množice pravil se kontrolna procedura iterira za vsako vrednost razredne spremenljivke posebej. Za vsako inducirano pravilo se izločijo le pokriti primeri, ki pripadajo trenutni vrednosti razredne spremenljivke (pozitivni primeri). Negativni pokriti primeri v obeh primerih ostanejo v učni množici do konca. Kontrolni proceduri se razlikujeta tudi po tem, da “neurejena” CN2 inducira pravila od splošnega proti specifičnemu tako, da specializira le pravila v snopu (najboljša pravila) z iterativnim dodajanjem novih členov (konjunktov) v pogojni del pravila Cond. Procedura se ustavi, ko ni možno več dodati pravila v snop, ker nobena od možnih specializacij ni boljša od pravil, ki so trenutno v snopu.

Glavna prednost odločitvenih pravil je jasna predstavitev naučenega modela - v obliki pravil. Pravila so lahko razumljiva za človeka in tako lahko razberemo, kaj se je algoritem naučil in na podlagi katerih značilnik ter na podlagi kakšnih vrednosti značilnik se algoritem odloča.

### Odločitvena drevesa C4.5

Algoritmi za gradnjo odločitvenih dreves glede na izbrano oceno informativnosti posameznih značilnik izbiraajo značilke in ustrezne podmnožice njihovih vrednosti za gradnjo odločitvenega drevesa. Na vsakem koraku se oceni vse razpoložljive značilke in izbere najboljšo, na podlagi vrednosti katere se razdeli podatke. Algoritmi za gradnjo odločitvenih dreves ocenjujejo značilke na podlagi mer za ocenjevanje značilnik, kot so informacijski prispevek, razmerje informacijskega prispevka ali ReliefF. Mere za ocenjevanje značilnik so podrobneje opisane v poglavju 6.4.1. Vsaka veja predstavlja neko vrednost ali več vrednosti zbrane spremenljivke. V odločitvenem drevesu:

- notranja vozlišča predstavljajo značilke
- veje predstavljajo vrednosti značilke v vozlišču
- listi drevesa predstavljajo razrede
- v vsakem vozlišču je ena značilka. Na podlagi vrednosti te značilke se podatki razdelijo po vejah drevesa. Z delitvijo moramo pridobiti čim več informacije o razredu.
- Nov primer klasificiramo tako, da sledimo poti od korena drevesa do lista. Usmerjajo nas vrednosti značilnik novega primera.

Prednost odločitvenih dreves je jasna predstavitev naučenega modela v obliki odločitvenega drevesa. Slabost odločitvenih dreves je pretirano prilagajanje učni množici (ang. overfitting). Pogosto se zgodi, da se odločitveno drevo preveč prilagodi podatkom v učni množici, kar ima za posledico bistveno slabše rezultate na novih primerih. Težavo se rešuje s pravili za ustavitev gradnje drevesa (na podlagi števila primerov v vozlišču, zastopanosti večinskega razreda ali pa števila primerov v listih) ali pa z obrezovanjem po izgradnji celotnega drevesa (na podlagi m-ocene ali združevanja vozlišč z enakim večinskim razredom).

V tem delu smo uporabili algoritem za odločitvena drevesa C4.5 [116]. Algoritem C4.5 zna obravnavati zvezne značilke, manjkajoče vrednosti in je robusten na šum v podatkih. C4.5 je eden izmed najbolj poznanih in najpogosteje uporabljenih algoritmov strojnega učenja.

### Naključni gozdovi RF

Metoda naključnih gozdov [15] (ang. random forest) je namenjena izboljšanju napovedne točnosti drevesnih algoritmov. Razvita je bila z namenom preprečevanja klasične težave odločitvenih dreves - pretiranega prilagajanja učnim podatkom (ang. overfitting). Ideja je generirati množico dreves (gozd), običajno 100, tako da se



pri zbiru značilke v posameznem vozlišču izbor omeji na naključno generirano podmnožico vseh značilk. Po [15] vsakič generiramo podmnožico velikosti logaritma števila značilk plus ena. Vsako drevo, zgrajeno na učni množici, se nato uporabi za klasifikacijo novega primera po metodi glasovanja - vsako drevo ima en glas, ki ga nameni razredu, v katerega bi to drevo klasificiralo nov primer.

Metoda naključnih gozdov je robustna, saj zmanjša varianco drevesnih algoritmov. S to metodo drevesni algoritem doseže napovedno točnost, primerljivo z najboljšimi algoritmi. Slaba stran metode je, da je razlaga odločitev otežena, saj dobimo veliko množico dreves, ki je nepregledna in nerazumljiva za uporabnika.

### Metode za kombiniranje klasifikatorjev z glasovanjem VOTE

Metode za kombiniranje klasifikatorjev lahko razdelimo v dve skupini. V prvi so metode, ki za gradnjo osnovnih klasifikatorjev uporabljajo en sam učni algoritem, njihovo raznolikost pa dosežemo na primer s spreminjanjem učne množice. V to kategorijo spada prej opisana metoda naključnih gozdov. V drugi skupini so metode, ki kombinirajo klasifikatorje zgrajene z različnimi učnimi algoritmi (heterogene klasifikatorje). Najbolj znana metoda iz te skupine je metoda skladanja klasifikatorjev (ang. stacking), pri kateri poteka kombiniranje v dveh korakih. V prvem koraku z različnimi učnimi algoritmi na osnovni učni množici zgradimo osnovne klasifikatorje. Njihove napovedi opišemo s tako imenovanimi metaatributi in jih sestavimo v meta učno množico. V drugem koraku na tako dobljeni meta učni množici zgradimo metaklasifikator, ki združi klasifikacije osnovnih klasifikatorjev v končno klasifikacijo [134]. Na našem problemu bomo preizkusili metodo za kombiniranje klasifikatorjev z glasovanjem imenovano VOTE [128]. Kombinirali bomo vseh sedem prej opisanih metod strojnega učenja, ki bodo glasovale o razredu vsakega posameznega primera. Napovedan razred bo razred z največ glasovi.

#### 7.1.2 Metode ovrednotenja klasifikatorjev

Kakovost algoritmov strojnega učenja se ocenjuje z merami za ocenjevanje učenja. Algoritmi strojnega učenja so večinoma zgrajeni avtomatsko na podlagi nekega postopka, ki večinoma ni odvisen od narave problema, ki ga rešujemo. Zato nikoli ne moremo vnaprej vedeti, kako uspešno bo zgrajeni model reševal določen problem. Tudi v našem primeru ne moremo vnaprej vedeti, kateri algoritem je najboljši za problem akustične klasifikacije morskega dna. V prejšnjem poglavju smo predstavili nekaj klasifikatorjev. Nekatere smo izbrali zaradi preprostosti in sposobnosti interpretacije rezultatov, nekatere ker se so se na podobnih problemih izkazali za uspešne [6]. Vendar le na podlagi splošno znanih prednosti in pomanjkljivosti izbranih algoritmov ne moremo določiti tistega, ki je najbolj primeren za naš problem. Da lažje ugotovimo, kateri modeli so dobri in kateri so slabi, si pomagamo z merami za ocenjevanje učenja. Seveda so nam tudi mere za ocenjevanje učenja le v pomoč. Dober algoritem je tisti, ki je uporaben tudi v praksi, kar pa je spet odvisno od konkretnega problema. Če je nek algoritem dobro ocenjen ampak časovno zelo zahteven, mi pa potrebujemo klasifikacijo v realnem času, se bomo raje odločili za slabše ocenjen a hiter algoritem. Pri klasifikacijskih problemih želimo vedeti, kako

uspešna bo klasifikacija na novih primerih. Pri ocenjevanju uspešnosti avtomatsko zgrajenega modela praviloma ločimo podatke v dve množici. Prvo množico imenujemo učna množica. Učno množico podatkov uporablja algoritem pri učenju. Drugo množico imenujemo testna množica in se uporablja za testiranje avtomatsko zgrajenega modela. S takšnim postopkom ocenjevanja uspešnosti se izognemo problemu, da se bi naučeni model preveč prilegal testnim primerom, in bi tako dobili napačno oceno uspešnosti napovedovanja. Vendar pa moramo paziti, da sta tako učna kot tudi testna množica reprezentativni podmnožici danega problema. Za ocenjevanje uspešnosti klasifikacije smo v tem delu uporabili naslednje mere:

- klasifikacijska točnost,
- tabela napačnih klasifikacij.

### Mere ocenjevanja učenja

Klasifikacijska točnost je najbolj preprosta mera, ki nam pove kakšen delež primerov v testni množici je naš algoritem pravilno klasificiral. Rešitev vsakega primera klasifikacijskega problema je enolično določen razred iz množice možnih razredov. Tako lahko uspešnost reševanja klasifikacijskih problemov definiramo kot razmerje med pravilnimi klasifikacijami primerov problema in vsemi možnimi primeri. To zapišemo kot:

$$\text{Klasifikacijska točnost} = \frac{N_p}{N} \times 100\%$$

kjer je  $N_p$  število pravilnih klasifikacij in  $N$  število vseh možnih primerov problema. Razmerje množimo s 100, da dobimo rezultat v odstotkih.

Klasifikacijska točnost je preprosta mera ocenjevanja učenja. Pogosto si želimo vedeti več podrobnosti o uspešnosti posameznega algoritma. Zanima nas lahko na primer s katerimi razredi ima več težav, pri katerih vrednostih razreda je bolj uspešen, ali pa katere vrednosti najpogosteje zamenjuje ipd. To nam omogoča tabela napačnih klasifikacij. Tabela napačnih klasifikacij uporabljamo takrat, ko želimo vedeti kako dobro so klasificirani primeri iz posameznih razredov, česar nam pa klasifikacijska točnost ne pove, saj je povprečena preko vseh razredov. V praksi nas zanima uspešnost za vsak razred posebej.

Pravi razred	Klasificiran kot			Vsota
	Mulj	Trava	Kamnito	
Mulj	<b>190</b>	10	15	215
Trava	10	<b>60</b>	20	90
Kamnito	0	30	<b>250</b>	280
Vsota	200	100	285	585

Tabela 7.1: Primer tabele napačnih klasifikacij.

Tabela 7.1 je primer tabele napačnih klasifikacij za trirazredni klasifikacijski problem. V tabeli so prikazana števila pravilno in napačno klasificiranih primerov. V tabeli lahko na enak način prikažemo tudi odstotke. Na diagonali (notranje 3x3 tabele) je število pravilno klasificiranih primerov. To so primeri kjer se pravi in

napovedani razred ujemata. V vseh ostalih poljih imamo število napačno klasificiranih primerov. Vsota vrstic pomeni število primerov ki dejansko pripadajo razredu. Vsota stolpcev predstavlja število primerov ki so bili klasificirani v dotični razred. Število 10 v drugem stolpcu prve vrstice pomeni, da je bilo 10 primerov klasificiranih v razred 'Trava', dejansko pa pripadajo razredu 'Mulj'.

Tabela napačnih klasifikacij nam torej omogoči podrobnejši vpogled v točnost klasifikatorja. Iz nje lahko razberemo, kje je posamezen klasifikator dober in kje je njegova točnost slabša ter kateri razredi mu povzročajo največ težav.

### Prečno preverjanje

Kadar imamo dovolj veliko množico primerov z znano vrednostjo odvisne spremenljivke, lahko to množico razdelimo na testno in učno množico. Učno množico uporabimo za učenje algoritma in naučen model preverimo na testni množici. Za ocenjevanje kakovosti uporabimo eno izmed mer, opisanih v prejšnjem razdelku. Če pa imamo na voljo malo vhodnih podatkov, si ne moremo privoščiti, da bi postopek učenja prikrajšali za primere testne množice. Najzanesljivejša metoda za ocenjevanje uspešnosti modela z malo primeri danega problema je "izloči enega" (angl. leave-one-out). Po tej metodi iz celotne množice izločimo en primer in ga uporabimo za testiranje uspešnosti modela, ki ga zgradimo iz preostalih primerov. Postopek ponovimo za vse učne primere. Tako zgradimo  $N$  modelov, ki jih testiramo na enem učnem primeru. Uspešnost modela dobimo tako, da izračunamo povprečje uspešnosti vseh modelov na izločenem primeru. Zavedati se moramo, da smo pri tem uporabili  $N$  različnih modelov, ki so si med seboj zelo podobni, saj so bili zgrajeni na zelo podobni učni množici in imajo zato zelo podobno uspešnost. Metoda izloči enega je v našem primeru časovno nesprejemljiva, saj moramo zgraditi toliko modelov, kolikor imamo učnih primerov. Metodo izloči enega lahko posplošimo na izloči  $N/K$  primerov, ki ji pravimo  $K$ -kratno prečno preverjanje (angl.  $K$ -fold cross validation).  $K$  je število modelov, ki jih moramo zgraditi. Najprej množico razpoložljivih primerov razdelimo na  $K$  približno enako močnih podmnožic. Za vsako podmnožico zgradimo model na uniji preostalih podmnožic. Nato dobljeni model uporabimo za reševanje primerov na dani podmnožici. Uspešnost končnega modela, ki ga zgradimo iz vseh razpoložljivih primerov, ocenimo kot povprečno uspešnost vseh  $K$  zgrajenih modelov na celotni množici testnih primerov.

### Statistični test značilnosti razlik uspešnosti algoritmov

Za test značilnosti rezultatov prečnega preverjanja smo uporabili popravljene enosmerni parni  $t$ -test z Bonferroni popravki [26]. Omenjeni test je primeren za primerjavo uspešnosti več algoritmov na isti domeni [77]. Algoritme primerjamo paroma. Pomembno je, da pri uporabi prečnega preverjanja vse algoritme ovrednotimo na podlagi enake delitve podatkov. Standardni  $t$ -test predpostavlja neodvisnost poskusov, pri prečnem preverjanju pa poskusi niso neodvisni. Zato smo uporabili parni  $t$ -test. Potek testiranja za  $K$  poskusov je: najprej izračunamo razliko uspešnosti  $U$  obeh algoritmov:

$$raz_i = U_1 - U_2; i = 1, \dots, K$$

Preizkusiti želimo hipotezo, da oba algoritma dosejata enako uspešnost, torej  $\overline{raz} = 0$ . Izračunamo povprečje razlike:

$$\overline{raz} = \frac{1}{K} \sum_{i=1}^K raz_i$$

in standardno odstopanje

$$s = \sqrt{\frac{1}{K-1} \sum_{i=1}^K (raz_i - \overline{raz})^2}$$

Naslednja statistika  $t$  je porazdeljena po Studentovem zakonu:

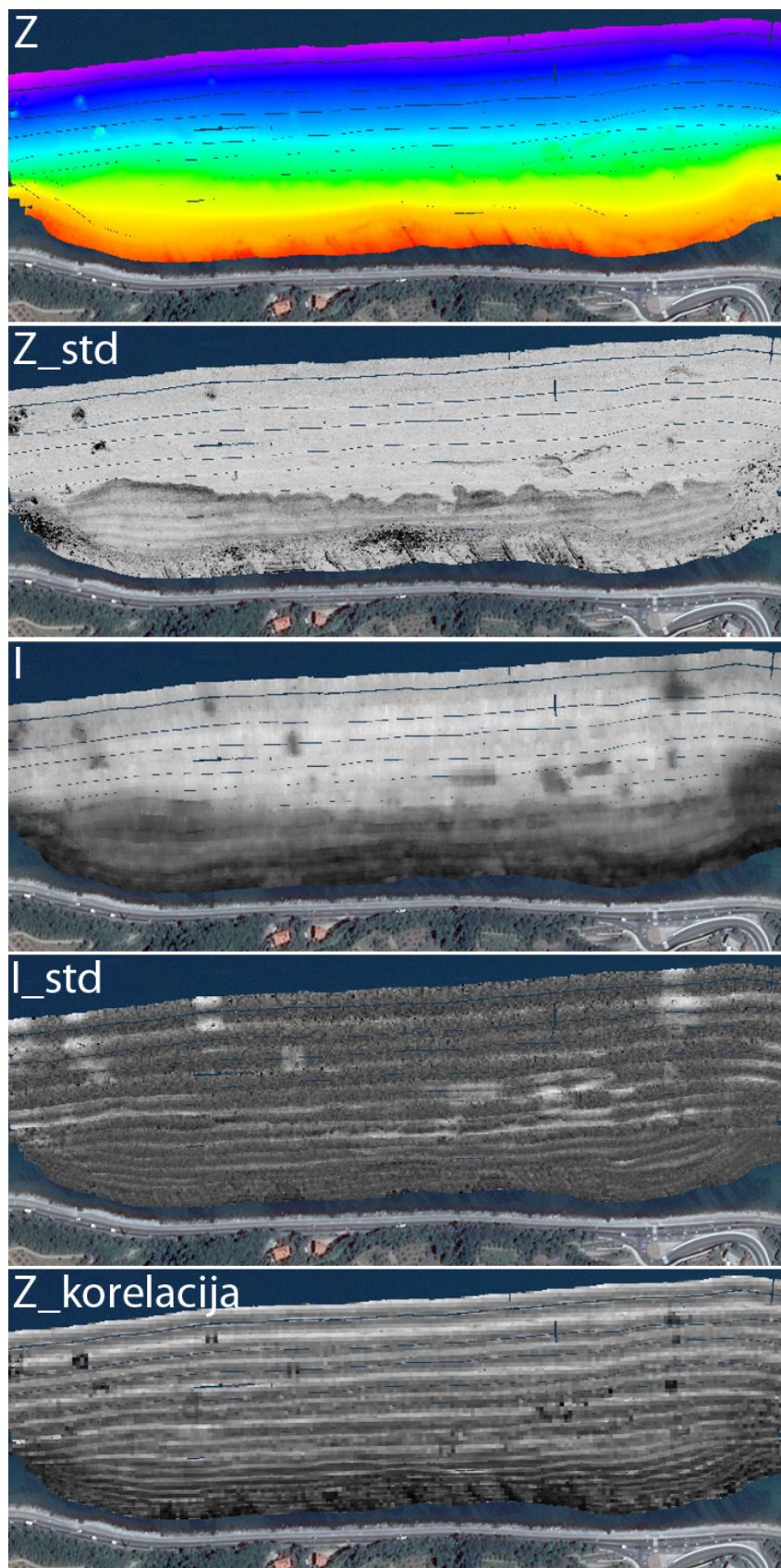
$$t = \frac{\overline{raz}}{s/\sqrt{K}}$$

Vrstni red primerjanih algoritmov izberemo tako da je  $\overline{raz} > 0$ . Če je vrednost statistike  $t < t(\alpha, K-1)$ , potem ničlo hipotezo, da so razlike enake 0, lahko zavrnemo, ob stopnji značilnosti  $\alpha$ .

## 7.2 Množica podatkov in značilke

Za ovrednotenje klasifikatorjev smo uporabili podatke iz območja med Izolo in Kopriv. Območje je natančneje opisano v poglavju 6.1. Surove sonarske podatke smo obdelali po postopku, opisanem v poglavju 5. Uporabili smo pristop z lokalnim uravnoteženjem kotnih odvisnosti in lokalnim uravnavanjem standardnih odklonov, ki smo ga ovrednotili kot najbolj primerne za potrebe akustične klasifikacije morskega dna (poglavje 6.3). Uporabili smo 8 najbolj ocenjenih značilk, izračunanih na ustreznih velikostih celic, podanih v poglavju 6.4. Uporabljene značilke:

- $i$  : intenziteto odboja, izračunano na celici velikosti 2x2 metra
- $z$  : povprečno globino, izračunano na celici velikosti 1x1 meter
- $i\_std$ : standardni odklon intenzitet, izračunan na celici velikosti 2x2 metra
- $z\_std$ : standardni odklon globine, izračunan na celici velikosti 1x1 metra
- lastnosti matrike sopojavnosti sivin, izračunane na podatkih o globinah z velikostjo celice 4x4 metre:
  - $z\_homogost$
  - $z\_kontrast$
  - $z\_korelacija$
  - $z\_različnost$



Slika 7.1: Prostorski prikaz značilnik, pripravljenih za obdelavo z algoritmi strojnega učenja.

Modul za predobdelavo na vходу sprejme surove sonarske podatke in nam vrne množico značilk v formatih ARFF in TAB, pripravljeno za strojno učenje. Modulu ni potrebno nastaviti nobenih parametrov, postopek je povsem avtomatiziran in sledi postopkom, opisanim v prejšnjih poglavjih. Za nadzorovano učenje podatkom dodamo še dodatno značilko - informacijo o razredu.

Na sliki 7.1 je prostorski prikaz izračunanih značilk. Vsaka izmed značilk nosi del informacije o morskem dnu. Pomen globin smo omenili že v začetku poglavja, iz standardnih odklonov globin ( $I_{std}$ ) so lepo razvidni robovi morskih travnikov, intenzitete ( $I$ ) razkrijejo področje kamnitega dna, standardni odkloni intenzitet poudarijo izdanke kamnin dlje od obale. Od značilk, izpeljanih iz GLCM smo za občutek vključili le sliko korelacij.

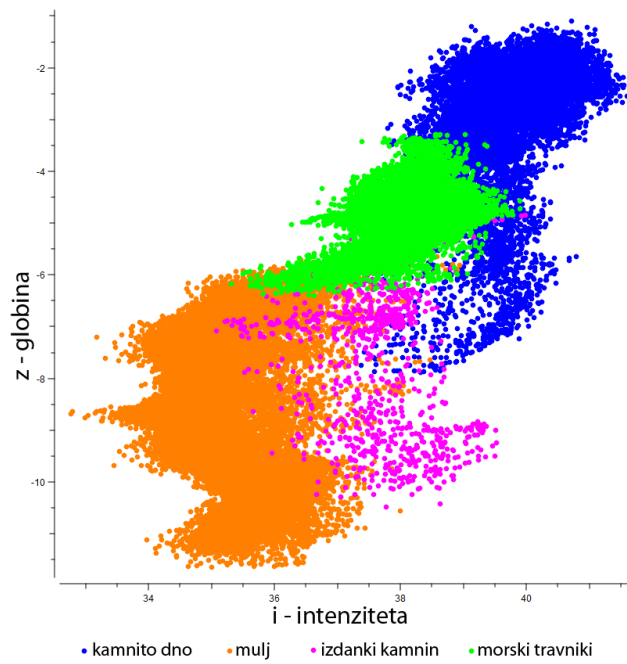
Za potrebe ovrednotenja algoritmov strojnega učenja smo zbrali in obdelali podatke o 182.686  $m^2$  morskega dna. 47.065  $m^2$  je kamnitega dna, 101.107  $m^2$  mulja, 30.111  $m^2$  morskih travnikov in 4.403  $m^2$  izdankov kamnin.

Glavni prednosti naše metode sta hitrost in zvezna pokrivnost obravnavanega področja. Hidrografske meritve obravnavanega območja smo izvedli v dveh urah, obdelava podatkov je trajala približno tri ure. Tako lahko območje velikosti 180.000 kvadratnih metrov natančno klasificiramo v enem delovnem dnevu, kar je bistveno manj časa kot bi ga porabili z vizualnimi ali mehanskimi metodami. Druga prednost je zvezna pokrivnost. Ločljivost naših podatkov na obravnavanem območju je v povprečju okoli 100 točk na kvadratni meter. Takšno ločljivost, od metod omenjenih v uvodu, dosega le bočni sonar, ki pa ima slabšo pozicijsko točnost in ne omogoča hkratnega snemanja globin in intenzitet odbojev. Dobro načrtovanje meritev in lastna napredna programska oprema za izvajanje meritev nam omogočita pokritje celotnega obravnavanega območja brez lukenj. Tako ni potrebe po interpolaciji dobljenih rezultatov, kot je to potrebno pri kartiranju z vizualnimi metodami, mehanskimi metodami in enosnopnimi sonarji [100], [120].

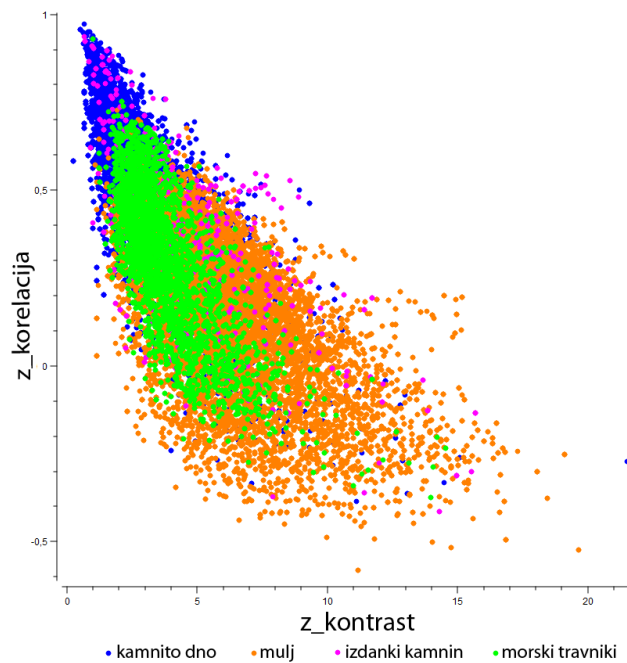
### 7.2.1 Vizualizacija značilk

Zelo je pomembno razumeti in interpretirati vmesne in končne rezultate strojnega učenja. Vizualizacija je pomembna, saj nam pohitri razumevanje smiselnosti in pomena naučenega ter razumevanja pomena posameznih značilk.

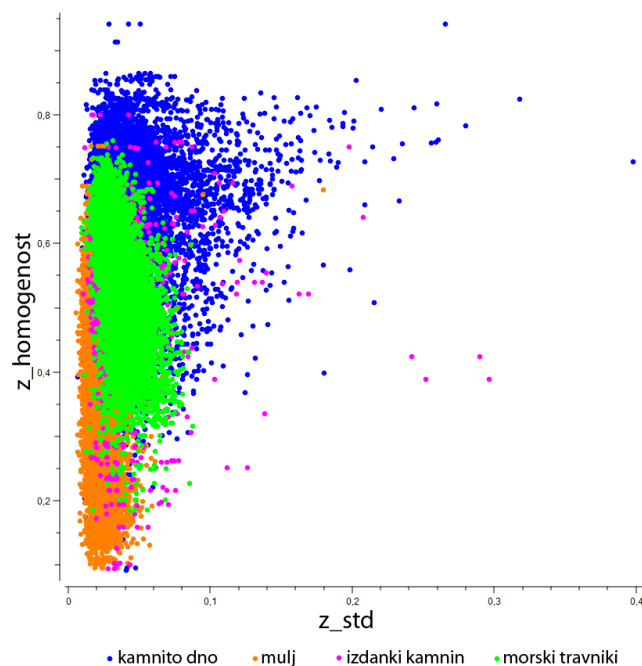
Na sliki 7.2 je graf medsebojne odvisnosti iz katerega razločno vidimo vpliv značilke  $i$  (intenziteta odboja) in  $z$  (globina) na tip morskega dna. Intenziteta odboja je najvišja na kamnitem dnu in najnižja na mulju. Trava in izdanki kamnin so si po intenziteti zelo podobni, ločita se po globini. Trava je pogostejša v bolj plitvi vodi. Najgloblje najdemo muljnato dno. Slika 7.3 prikazuje razmerje med korelacijo globin ( $z\_korelacija$ ), kontrastom ( $z\_kontrast$ ) in tipom morskega dna. Kamnito dno ima najmanjši kontrast in največjo korelacijo. Trava ima nekoliko nižjo korelacijo in višji kontrast. Slika 7.4 prikazuje značilki  $z\_homogenost$  in  $z\_std$ . Izdanki kamnin nimajo značilnih vrednosti omenjenih značilk in so razpršeni po celotnem grafu. Bolj enotne so vrednosti za travo, kamnito dno in mulj. Mulj ima zelo nizko homogenost in nizek standardni odklon globin, kar je pričakovano zaradi ravnega dna. Odklon globin je najvišja na kamnitem dnu, kjer je najvišja tudi vrednost značilke homogenost. Trava je ponovno neke vmes, ampak lepo razločna.



Slika 7.2: Graf medsebojne odvisnosti za značilki  $i$  in  $z$  ter tip morskega dna.



Slika 7.3: Graf medsebojne odvisnosti za značilki  $z\_korelacija$  in  $z\_kontrast$  ter tip morskega dna.



Slika 7.4: Graf medsebojne odvisnosti za značilki  $z\_std$  in  $z\_homogenost$  ter tip morskega dna.

Za travnato dno dobimo vrednosti značilke  $z\_homogenost$  med 0.3 in 0.7, vrednost standardnega odklona globin pa med 0.02 in 0.08.

### 7.3 Rezultati

Za ovrednotenje algoritmov smo uporabili klasifikacijsko točnost, ki smo jo ocenili na podlagi 10-ih ponovitev 10-kratnega prečnega preverjanja. Statistično značilnost razlik uspešnosti posameznih algoritmov smo paroma testirali s popravljenim parnim T-testom z nastavljenjo stopnjo zaupanja 99.9 % ( $\alpha = 0.001$ ). Rezultati uspešnosti klasifikacijskih algoritmov na problemu klasifikacije morskega dna so prikazani v tabeli 7.2. Podana klasifikacijska točnost je povprečna klasifikacijska točnost 10-ih zagonov 10-kratnega prečnega preverjanja. Za primerjavo je v prvi vrstici dodana verjetnost večinskega razreda, v našem primeru je to muljnato dno, ki prekriva 55.39% obravnavanega območja. Že algoritem OneR doseže klasifikacijsko točnost preko 90%. Algoritem OneR se odloča na podlagi ene sama, najbolj informativne značilke. V našem primeru je OneR kot najboljšo izbral značilko  $i$  in dosegel 91.50 % klasifikacijsko točnost. Po pričakovanjih se naivni Bayes odreže dobro, a ne odlično. Naivni Bayes je enostaven in hiter algoritem za klasifikacijo, vendar ima težave z odvisnostjo značilk. V našem primeru ne drži postavka o neodvisnosti značilk. Veliko značilk je izpeljanih in Naivni Bayes na takih podatkih ne deluje najbolje. Algoritmi SVM, kNN in CN2 dosežejo podobne klasifikacijske točnosti, vsi okoli 97 %. Med njihovimi rezultati ni statistično značilnih razlik (tabela 7.3).



Za najuspešnejša se izkažeta algoritma z uporabo odločitvenih dreves - C4.5 in naključni gozdovi RF ter metoda kombiniranja klasifikatorjev z glasovanjem VOTE. Algoritem C4.5 ima povprečno klasifikacijsko točnost 97.64 %, metoda VOTE 97.90 %, algoritem RF pa kar 98.47 %. Algoritem RF je statistično značilno boljši od algoritma C4.5 in metode VOTE (tabela 7.3). Najbolje se je torej odrezal algoritem naključnih gozdov. Algoritem odpravlja pomanjkljivost odločitvenih dreves - to je pretirano prilagajanje na podatke v učni množici.

Algoritem	Klasifikacijska točnost
Večinski razred	55.39
OneR	91.44
Naivni Bayes	93.97
SVM	96.84
kNN	97.01
CN2	97.12
C4.5	97.64
VOTE	97.90
RF	98.47

Tabela 7.2: Klasifikacijska točnost algoritmov strojnega učenja na akustičnih podatkih. Podana klasifikacijska točnost je povprečna klasifikacijska točnost 10-ih zagonov 10-kratnega prečnega preverjanja.

	OneR	NB	SVM	kNN	CN2	C4.5	VOTE	RF
OneR		x	x	x	x	x	x	x
NB	x		x	x	x	x	x	x
SVM	x	x				x	x	x
kNN	x	x				x	x	x
CN2	x	x				x	x	x
C4.5	x	x	x	x	x			x
VOTE	x	x	x	x	x			x
RF	x	x	x	x	x	x	x	

Tabela 7.3: Prikaz statistično značilnih razlik v klasifikacijski točnosti algoritmov pri  $\alpha = 0.001$ .

Kljub temu, da je algoritem naključnih gozdov dosegel statistično značilno najboljši rezultat, ga ne moremo enolično razglasiti za najbolj primerne za akustično klasifikacijo morskega dna. V kolikor želimo le dobiti karto tipov morskega dna je algoritem naključnih gozdov prava izbira. Če bi radi razumeli katere značilke in katere vrednosti posameznih značilk pripeljejo algoritem do odločitve je bolje izbrati algoritem C4.5. Algoritem C4.5 je tudi hitrejši za klasifikacijo novih primerov, saj nove primere klasificira na podlagi (enega) odločitvenega drevesa. Naključni gozdovi zgradijo 100, sicer manjših, odločitvenih dreves. Pri klasifikaciji novega primera mora vsako drevo klasificirati ta primer in glasovati o njegovem razredu. Uporabo metode VOTE za problem klasifikacije morskega dna odsvetujemo. V delu [32] avtorji pokažejo, da lahko z združevanjem heterogenih klasifikatorjev dosežemo kvečjemu enako dobre rezultate kot z izbiro najboljšega klasifikatorja. Poleg tega je metoda zelo počasna, saj moramo zgraditi modele za vse klasifikatorje, nato pa

mora vsak posamezen klasifikator glasovati za razred novega primera. Podobno kot pri algoritmu RF, je zaradi velikega števila klasifikatorjev interpretacija odločanja po metodi VOTE praktično nemogoča.

Algoritem C4.5 je torej relativno hiter in omogoča vizualizacijo in razlago svojih odločitev, algoritem naključnih gozdov pa je natančnejši. Odločitev, katerega izbrati, je odvisna od potreb raziskave.

### Graf odločitvenega drevesa

Algoritem C4.5 nam poleg napovedovanja razredov za nove primere omogoča tudi vizualizacijo modela, na podlagi katerega klasificira nove primere. Model predstavimo z odločitvenim drevesom. Odločitveno drevo algoritma C4.5 za našo testno domeno je prikazano na sliki 7.5 Drevo ima 363 listov, 725 vozlišč in povprečno 9 nivojev. Na sliki 7.5 je zaradi preglednosti prikazanih le prvih 5 nivojev drevesa. Iz odločitvenega drevesa lahko razberemo kako se algoritem odloča, katere značilke so najpomembnejše pri določanju razreda in kako vrednosti značilke vplivajo na verjetnosti razreda. V vsakemu vozlišču drevesa je najprej zapisan večinski razred, nato pa njegova verjetnost. V korenu drevesa je tako večinski razred *mulj*, zastopa pa 55.4 % primerov. Pod črto je v vozlišču zapisana značilka, na podlagi katere se vozlišče deli v nova vozlišča, na vejah drevesa pa ustrezne vrednosti. V korenskem vozlišču se primeri delijo glede na vrednost spremenljivke *i*. V levo vejo gredo primeri z  $i \leq 36.972$  v desno pa primeri z  $i > 36.972$ . Iz odločitvenega drevesa lahko razberemo, da nam že vrednost značilke  $i \leq 36.972$  s točnostjo 97.6% napove *mulj*-nato dno. Vrednost značilke večja od omenjene pa pomeni da gre najverjetneje za enega od ostalih tipov morskega dna.

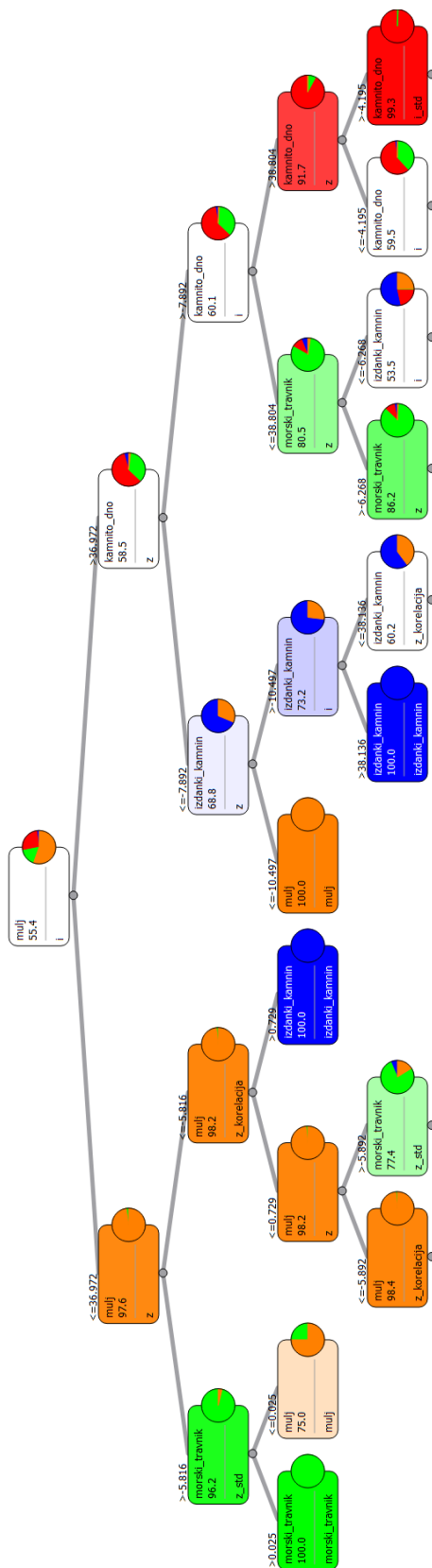
Algoritem nove primere klasificira tako, da se sprehodi po drevesu od korena do lista, na poti pa izbira veje ki ustrezajo vrednostim značilke obravnavanega primera. Primer klasificiramo v večinski razred lista drevesa. Za vsak list odločitvenega drevesa lahko razberemo pravilo, ki nas pripelje do tega lista. Primer pravil:

- če je  $i < 36.97$  in  $z \leq -5.82$  in  $z\_korelacija > 0.73 \Rightarrow$  tip dna = izdanki kamnin
- če je  $i < 36.97$  in  $z > -5.82$  in  $z\_std > 0.025 \Rightarrow$  tip dna = morski travnik
- če je  $i \geq 38.80$  in  $z > -4.20 \Rightarrow$  tip dna = kamnito dno

Razberemo lahko torej, da je za kamnito dno tipična nižja globina in visoka intenziteta odboja. Izdanki kamnin in morski travniki so povezani z nižjo intenziteto akustičnega odboja. Na območju izdankov kamnin je korelacija globin višja, na območju morskih travnikov pa je višji standardni odklon globin. Takih pravil lahko iz dobljenega odločitvenega drevesa izpeljemo 363, kolikor je listov. Za strokovnjake s področja obravnavanega problema so navadno zanimiva pravila ki so krajša in pokrijejo veliko primerov.

### Podrobnejša analiza napovednih točnosti naključnih gozdov

Najnatančnejši algoritem na naši domeni je algoritem naključnih gozdov, ki je v 10-ih zagonih 10-kratnega prečnega preverjanja dosegel klasifikacijsko točnost 98.47%. Re-

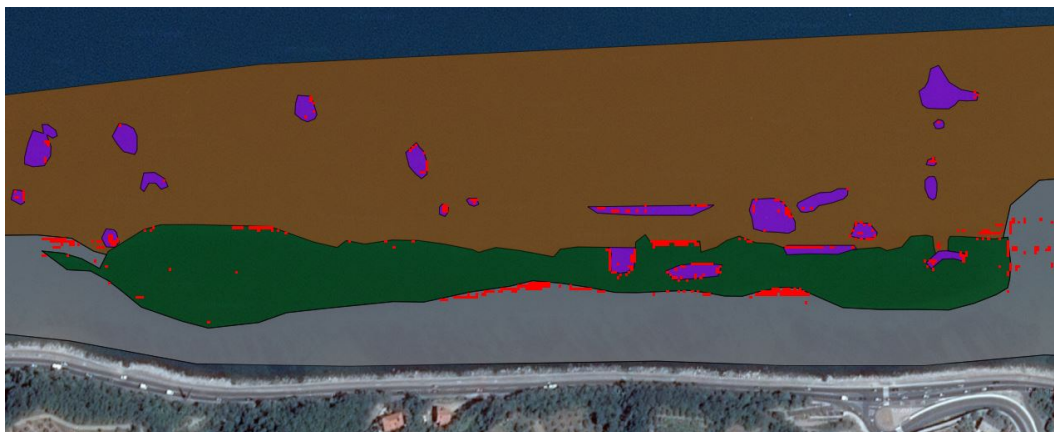


Slika 7.5: Prvih pet nivojev odločitvenega drevesa C4.5. V vozliščih je prikazan večinski razred in njegova verjetnost. Slika je narejena v programskem paketu Orange.

zultati algoritma naključnih gozdov so podrobneje predstavljeni z matriko napačnih klasifikacij, prikazano v tabeli 7.4. Rezultati prikazujejo število pravilno in nepravilno klasificiranih primerov po razredih. Rezultati so pridobljeni z metodo 10-kratnega prečnega preverjanja. Metoda podatke razdeli na 10 podmnožic in vsako podmnožico klasificira enkrat, torej tudi vsak primer klasificira natanko enkrat. Tako lahko preštejemo koliko primerov je pravilno klasificiranih in koliko napačno ter tudi za kakšne napake gre. Algoritem je pravilno klasificiral 180.204 od 183.020 primerov. Storil je 2.816 napak. Napake so prostorsko prikazane na sliki 7.6. Območja napak so podvojena zaradi nazornosti.

Pravi razred	Klasificiran kot				Vsota
	Kamnito dno	Izdanki kamnin	Mulj	Morski travnik	
Kamnito dno	<b>46380</b>	40	16	692	47128
Izdanki kamnin	84	<b>3300</b>	456	484	4324
Mulj	12	120	<b>101152</b>	84	101368
Morski Travnik	416	24	288	<b>29372</b>	30200
Vsota	46992	3484	101912	30632	183020

Tabela 7.4: Tabela napačnih klasifikacij algoritma naključnih gozdov.



Slika 7.6: Prikaz napak algoritma naključnih gozdov. Z rdečimi kvadrati so označena mesta, kjer je bila napoved tipa morskega dna napačna.

Večina napak se pojavlja na prehodih med različnimi tipi morskega dna. Dve veliki napačno klasificirani območja sta na prehodu med kamnitim dnom (siva barva) in morskimi travniki (zelena barva). Precej napak se pojavlja tudi med morskimi travniki in izdanki kamnin (viola barva). Mulj ne predstavlja težav za klasifikator naključnih gozdov, saj, kot smo videli že pri vizualizaciji značilk, imajo akustični odboji na muljnatem dnu manjšo intenziteto kot na ostalih tipih morskega dna. Iz matrike napačnih klasifikacij razberemo, da je algoritem pravilno klasificiral kar 101.152 od 101.368 celic muljnatega dna, kar je 99.8 %. Le 84-krat je muljnato dno zamenjal s travo, 120-krat z izdanki kamnin in 12-krat s kamnitim dnom. Največkrat je napačno

klasificirano kamnito dno in sicer kot morski travnik (692 primerov). Klasifikator je pogosto napačno napovedal mulj namesto izdankov kamnin (456 primerov), morski travnik namesto izdankov kamnin (484 primerov) in kamnito dno namesto morskega travnika (416 primerov). Slabše rezultate pri klasifikaciji izdankov kamnin smo do neke mere pričakovali. Izdanki kamnin so manjši kamni pomešani z muljem. Akustični odtisi na takem dnu so zelo nehomogeni. Na nekaterih delih je več mulja in manj kamnin in obratno. Zato se pogosto dogaja, da je področje označeno kot izdanki kamnin, klasificirano kot mulj.

Zelo pomembna je ugotovitev, da, razen na mejah med različnimi tipi morskega dna, ni opaziti večjih napak v klasifikaciji. Razvita metodologija, ki združuje postopke od zbiranja surovih podatkov, predobdelave podatkov in klasifikacije morskega dna, je pokazala velik potencial za uporabo akustičnih podatkov za klasifikacijo morskega dna.

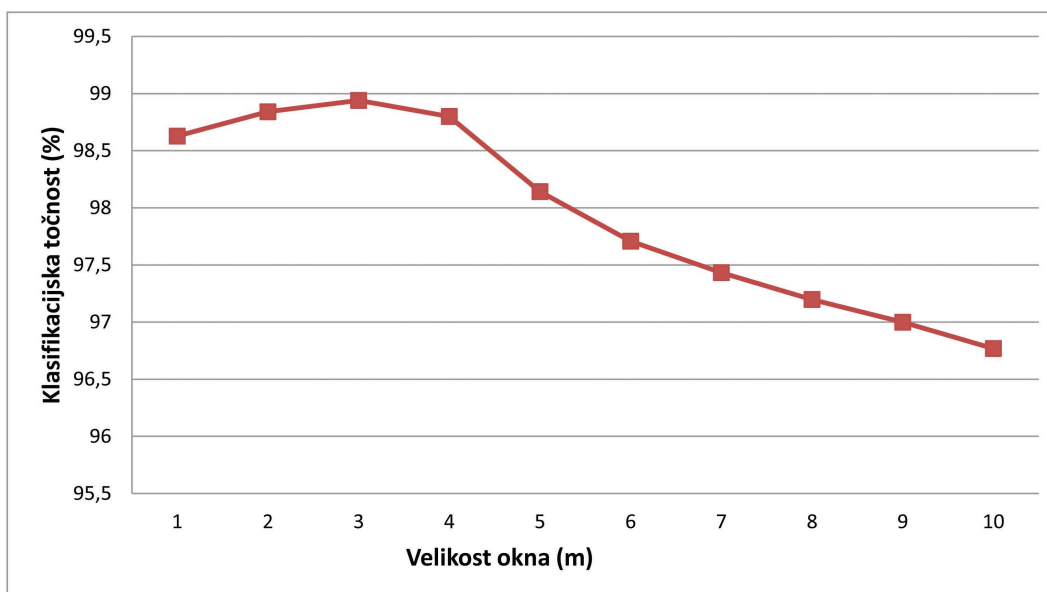
## 7.4 Poobdelava rezultatov klasifikacije

Na področju strojnega učenja se običajno srečujemo s problemi, kjer edino relacijo med posameznimi primeri predstavljajo vrednosti značilnk. Algoritmi strojnega učenja na podlagi vrednosti značilnk zgradijo klasifikacijske modele in razvrščajo primere v razrede. Kot smo omenili že v uvodu, je ena izmed glavnih prednosti mnogospopnega sonarja hkratno snemanje globin in intenzitet odbojev ter njihova natančna umestitev v prostor. Za vsak akustični odboj torej poznamo natančne koordinate dela morskega dna, od katerega se je odbil. Umeščenost podatkov v prostor nam omogoča, da podatke in rezultate grafično prikažemo na zemljevidu, kot smo to storili na sliki 7.6. Poleg grafičnega prikaza rezultatov nam umeščenost v prostor omogoča tudi, da rezultate akustične klasifikacije obravnavamo kot sliko. Na sliki 7.6 so napačno klasificirani primeri označeni z rdečo barvo. Poleg nekaj večjih območij napak, se veliko napak pojavlja posamično ali v manjših skupinah. Na te napake lahko gledamo kot na šum v sliki. Obstaja veliko število filtrov, katerih namen je odstranjevanje šuma na slikah, med najbolj poznanimi in najbolj učinkovitimi je medianin filter [23]. Naša ideja je, da bi z medianinim filtrom izboljšali rezultate klasifikacije.

Medianin filter je nelinearni digitalni filter, ki se pogosto uporablja za odstranjevanje točkastega šuma. Odstranjevanje šuma je tipičen korak predobdelave podatkov, s katerim izboljšamo kakovost podatkov in rezultate nadaljnjih obdelav. Nekoliko prirejen medianin filter smo v tem delu že uporabili za izločanje slabih podatkov. Glavna ideja medianinega filtra je, da se sprehodi skozi sliko, celico po celico in vrednost vsake celice zamenja z mediano sosednjih celic. Vzorcju, po katerem določamo sosednje celice, rečemo okno. Pri obdelavi dvodimenzionalnih signalov, kot so slike, je lahko okno različnih oblik. Najpogosteje se uporablja okna v obliki kvadrata, križa ali pa se aproksimira krog z uporabo radija.

Aplikacijo medianinega filtra na problem akustične klasifikacije morskega dna smo preizkusili na kumuliranih rezultatih metode naključnih gozdov. Rezultate klasifikacije z algoritmom naključnih gozdov smo pretvorili v matrično obliko in na njih aplicirali medianin filter za odstranjevanje šuma v podatkih. Popravljenе podatke

smo nato primerjali s podatki o dejanskih tipih morskega dna in izračunali ujemanje, ki mu zaradi nazornosti v nadaljevanju rečemo kar klasifikacijska točnost. Uporabili smo okno z različnimi radiji od 1 do 10. Uporabljamo celice velikosti 1 x 1 meter, posledično je tudi radij izražen v metrih. Na sliki 7.7 so prikazane klasifikacijske točnosti po uporabi medianinega filtra z različnimi velikostmi okna. Iz grafa lahko razberemo, da je največja izboljšava dosežena pri uporabi okna velikosti 3 metre, klasifikacijska točnost je kar 98.94 %, kar je za 0.47 % bolje od rezultata brez uporabe medianinega filtra. Z večanjem okna nad 3 metre začne klasifikacijska točnost padati in že pri velikosti okna 5 metrov pade pod začetno klasifikacijsko točnost algoritma naključnih gozdov. Uporaba prevelikega okna torej poslabša rezultate.

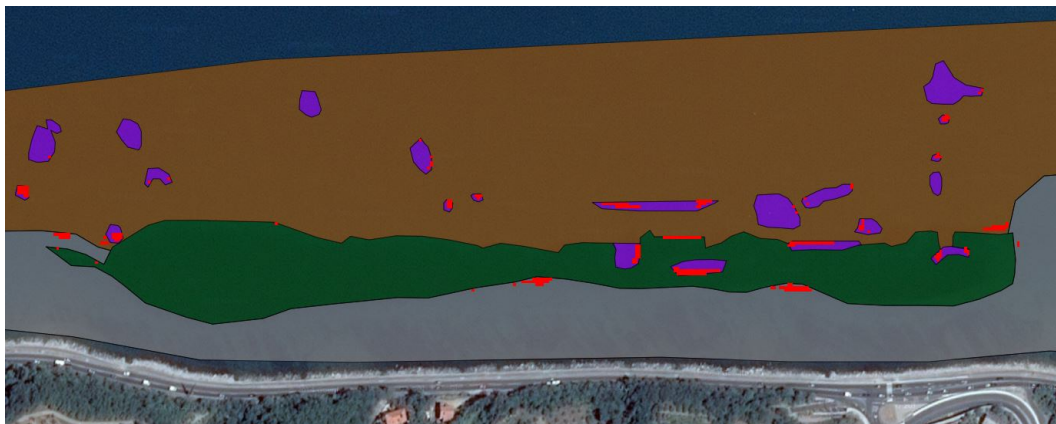


Slika 7.7: Graf klasifikacijskih točnosti pri uporabi različnih velikosti oken za interpolacijo.

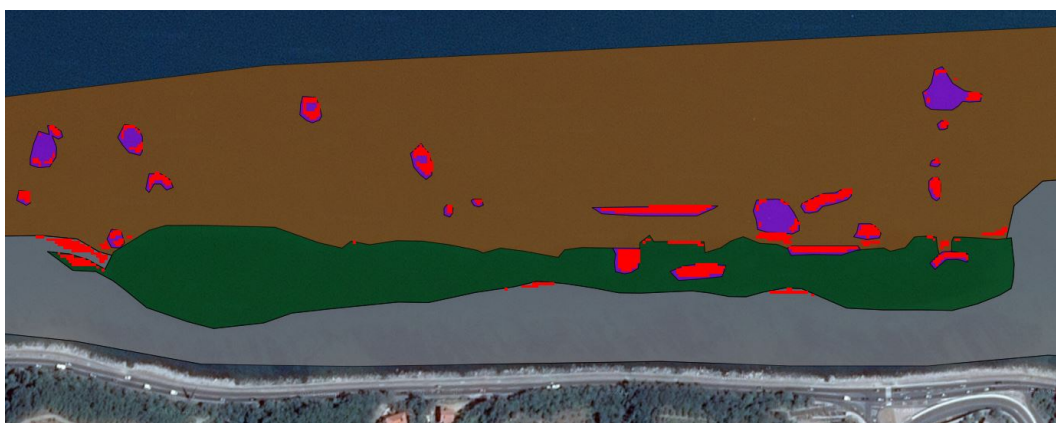
Na sliki 7.8 so prikazani rezultati poobdelave z medianinim filtrom z velikostjo okna 3 metre in na sliki 7.9 rezultati poobdelave z medianinim filtrom z velikostjo okna 6 metrov. Slika 7.8 je bolj čista od slike 7.6. Rdečih pik, ki označujejo napake je bistveno manj, območja napak pa so praviloma manjša in lepše zaključena. Medianin filter je uspešno popravil predvsem posamezne napačno klasificirane celice in manjše skupine napačno klasificiranih celic. Medianin filter z velikostjo okna 6 metrov sicer tudi odstrani manjše napake, vendar zaradi prevelikega okna povzroči veliko novih napak. Težave se pojavijo predvsem pri manjših in ožjih otokih izdankov kamnin, ki jih medianin filter zazna kot šum in jim posledično spremeni razred v razred širše okolice.

Medianin filter je učinkovit za poobdelavo rezultatov akustične klasifikacije morskega dna, saj iz dobljene karte tipov morskega dna izloči šum, ki ga predstavljajo napačno razvrščene celice in tako izboljša klasifikacijsko točnost. Na naši množici podatkov znaša napaka po klasifikaciji z algoritmom naključnih gozdov in poobdelavi

z medianinim filtrom zgolj 1.06 %.



Slika 7.8: Prikaz napak algoritma naključnih gozdov po uporabi medianinega filtra z velikostjo okna 3 metre.



Slika 7.9: Prikaz napak algoritma naključnih gozdov po uporabi medianinega filtra z velikostjo okna 6 metrov.





## Poglavje 8

# Aplikacija metodologije v realni domeni

V tem poglavju bomo prikazali praktično uporabnost razvite metodologije. Cilj našega raziskovalnega dela je bil razviti metodologijo, ki bo uporabna v praksi in na podlagi katere bo mogoče v podjetju razviti produkte in storitve z visoko dodano vrednostjo. Metodologijo smo preizkusili na realni domeni - kartiranju morskih travnikov v slovenskem morju.

### 8.1 Kartiranje morskih travnikov

Razvito metodologijo akustične klasifikacije morskega dna smo uporabil za kartiranje morskih travnikov. Morski travniki so izredno pomembni habitati za življenje nevretenčarjev, rib in mnogih ptic, poleg tega varujejo obalno regijo pred erozijo, s produkcijo kisika ključno prispevajo k oksigenaciji morske vode in so pomemben bio-indikator stanja obalnega morja. V slovenskem morju morski travniki ne dosegajo večjih globin zaradi motnosti vode, najdemo jih do globine 8 metrov [78]. Produktivnost morskih travnikov je izredna, saj bistveno prispevajo k primarni produkciji našega planeta. V severnih predelih Sredozemlja rastejo štiri vrste: prava morska trava *Zostera marina* in mala morska trava *Zostera nana*, ki raste do globine 15 metrov, ter kolenčasta cimodoceja *Cymodocea nodosa* in pozejdonka *Posidonia oceanica*, ki v čistih predelih lahko sežeta tudi do 50 metrov globine [14]. V Sredozemlju pozejdonka prerašča nekje med 25.000 – 50.000 km<sup>2</sup> površine, pri nas pa le kak kilometer dolg in 50 m širok pas med Koprom in Izolo, ki pa je tudi edino rastišče v Tržaškem zalivu in zahodni Istri [66]. Pri nas je zavarovana in na rdečem seznamu ogroženih vrst. V Sloveniji je najbolj razširjena cimodoceja. Morske trave v Sloveniji pokrivajo znatne površine sedimentnega dna, na katerem večina alg nima možnosti razvoja. Pri tem jih omejuje količina svetlobe, ki mora za omogočanje rasti dosegati vsaj 11% vpadne svetlobe [14]. S korenikami uspešno stabilizirajo sediment in tako ustvarjajo specifično okolje za naselitev številnih živalskih in rastlinskih vrst, listi pa omogočajo zaščito in skrivališče ter površino za naselitev epibiontov (organizmi, ki živijo na površini drugih organizmov). S produkcijo kisika ključno prispevajo k oksigenaciji morske vode, še posebej v zaprtih zalivih. Pozejdonka predstavlja klimaksni

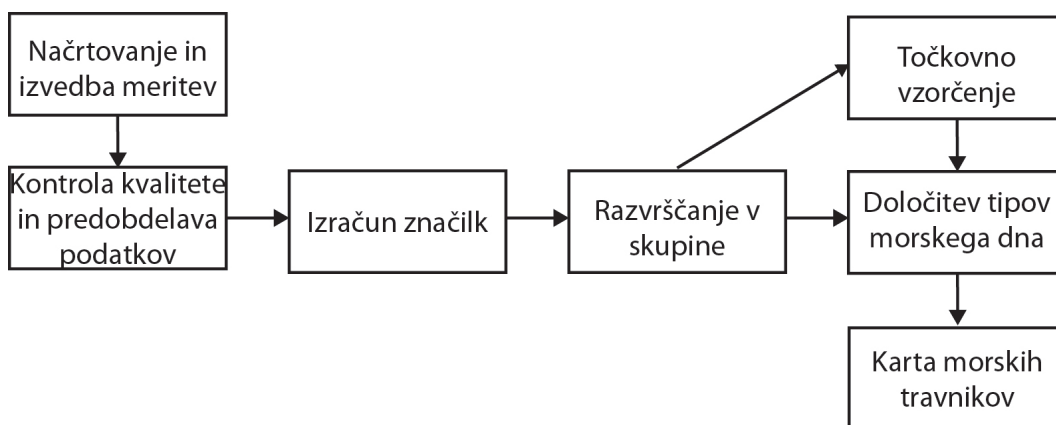
makrofitski sistem v stabilnem okolju, cimodoceja pa v bolj stresnem. Na splošno pri nas morski travniki ne dosegajo večjih globin zaradi motnosti vode (pozejdonka raste na primer le do 4 m globine, cimodoceja pa do 8 m), vendar so kljub temu izrednega pomena kot bivališča, prehranjevališča in skrivališča za morske organizme [66].

Morski travniki so torej izjemno pomemben element evropskih obalnih voda, hkrati pa so ranljivi in podvrženi propadanju zaradi človeških dejavnosti. Prekomerna količina hranil, zamuljevanje in mehanske (fizične) poškodbe dna so eni glavnih vzrokov za njihovo degradacijo in izginjanje. Mnoge države Evropske unije so že sprejele nacionalno zakonodajo in regulativo za zaščito morskih travnikov pred antropogenimi vplivi. Prav tako je bila v okviru vodne direktive (Water Framework Directive) ustanovljena skupna platforma držav članic za zagotavljanje "dobrega ekološkega statusa" preko minimalnega antropogenega vpliva na vse površinske vode, se pravi tudi obalno morje. Morske trave so dober pokazatelj razmer, ker odražajo kumulativen vpliv na okolje, se pravi kažejo negativne vplive na okolje, ki so se kopičili skozi čas. Zato so odlični indikatorski organizem, na katerega se lahko naslantajo okoljski monitoring in strategije upravljanja obalnega morja po celem svetu. V Sloveniji so bili ugotovljeni kot prevladujoče obremenitve in vplivi mnogi od faktorjev, ki vplivajo na morske travnike (fizične poškodbe, eutrofikacija) in stanje bližje obali v infralitoralju je bilo mestoma ocenjeno kot zmerno ali slabo stanje, kar kaže na pomembne lokalne obremenitve z negativnim vplivom na obalni pas [106]. Monitoring razširjenosti in porazdelitve morskih trav z oceno prisotnosti/odsotnosti ali prostorsko razporeditvijo je na večjih območjih mogoč s pomočjo akustičnih sonarjev, ki dobro ločijo morske trave od okolnega dna in so že del obsežnega monitoringa morskih travnikov nekaterih držav [122]. Uporablja se predvsem interpretacija akustičnih podatkov s strani strokovnjakov. Metoda je podobna fotointerpretaciji iz letalskih posnetkov. Strokovnjaki s področja hidrografije, oceanografije ali morske biologije iz podatkov o globinah in v zadnjem času tudi podatkov o intenzitetah akustičnih odbojev, določijo prisotnost morskih travnikov [63].

Razširjenost morskih travnikov je indikator, ki kaže na kakovost vode/onesnaženje morskega okolja in je, še posebej spodnja meja razširjenosti, tudi eden najbolj uporabljenih indikatorjev [14]. Spodnja meja razširjenosti travnika in gostota trave v globlji vodi sta indikatorja, neposredno povezana s prozornostjo vode, ker sta primarno regulirana s količino svetlobe. Zato imata visoko prioriteto v programih monitoringa, ki ocenjujejo vpliv eutrofikacije, onesnaženje s hranili in zamuljevanja, ki sta lahko posledica izpustov odpadne vode, školjčič, usedanja delcev iz ozračja in nanosov celinskih voda. Zgornja meja travnika je bolj občutljiva na mehanske (fizične) vplive in spremembe morskega dna, na primer zaradi ribiških dejavnosti ali sidranja. Vplivi, ki povzročijo izginjanje travnikov, se navadno tudi najprej odražajo v zgornji in spodnji meji razširjenosti. Čeprav so te meritve bolj robustne kot pregledovanje dna s potapljači, je njihova prednost pregled celotnih populacij na večjih območjih in v večjih časovnih obdobjih na enostaven in cenovno ugodnejši način.

### 8.1.1 Uporabljena metodologija

Ker o morskem dnu na obravnavanem območju nismo imeli podatkov, smo uporabili fenomenološki pristop. Fenomenološki pristop predvideva obstoj korelacije med morfološki in fizičnimi lastnostmi morskega dna ter akustičnimi podatki. Pri fenomenološkem pristopu se po izpeljavi in izboru značilk uporabi nenadzorovano klasifikacijo. Nenadzorovana klasifikacija podatke razvrsti v skupine statistično podobnih enot. Podatke smo razdelili v akustične razrede brez direktnega sklepanja na dejanske fizične lastnosti dna. S pomočjo preverjanja (potapljači, zajem vzorcev, kamere, aerofotografije, ocena strokovnjakov) smo nato akustične razrede povezali z dejanskimi tipi tal. Uporabili smo postopek razvrščanja v skupine z metodo k-tih povprečij (k-means clustering) [56]. Algoritem podatke razdeli v skupine tako, da je razdalja med posameznimi podatki znotraj skupine čim manjša, razdalja med skupinami pa čim večja. Za delovanje algoritma so zelo pomembne izbira razdalje, števila zelenih skupin in načina začetne razporeditve po skupinah. Za razdaljo smo uporabili evklidsko razdaljo. Algoritmu smo določili naključno začetno razporeditev. Diagram postopka je prikazan na sliki 8.1. Karto morskega dna, dobljeno z nenadzorovano klasifikacijo smo uporabili za določitev točk točkovnega vzorčenja. Nato smo omenjene točke uporabili za učenje algoritma naključnih gozdov in dobili natančno karto morskega dna.



Slika 8.1: Koraki klasifikacije morskega dna, od zajema podatkov do karte morskih travnikov.

Nenadzorovana klasifikacija je primerna predvsem za nepoznana področja. Omogoči nam da dobimo okvirno sliko o pestrosti morskega dna in da lažje načrtujemo zbiranje podatkov o morskem dnu. Predobdelava podatkov in izračun značilk potekata avtomatsko in nespremenjeno, ne glede na izbrano metodo klasifikacije.

Podatke o morskem dnu smo zbirali v začetku leta 2014. Opravljenih je bilo 34 dni meritev. Meritve so bile nekoliko bolj zamudne, ker smo merili morskno dno ob obali. Na odprtem morju in večjih globinah meritve potekajo bistveno hitreje. Za našo raziskavo je obalno področje zaradi večje pestrosti tipov morskega dna bolj zanimivo. Skupno smo obdelali podatke o preko 8.000.000 kvadratnih metrih mor-

skega dna, za celotno slovensko morje do globine 10 metrov. Ker smo postopek predobdelave podatkov in izračuna značilk povsem avtomatizirali, smo podatke obdelali v le 5 delovnih dneh. Avtomatski postopek nam omogoča obdelavo podatkov tudi ponoči in med vikendi. Predobdelavo podatkov smo implementirali v programskem jeziku Python. Postopek je možno pohitriti z združevanjem posameznih korakov, ki ji trenutno izvajamo zaporedno, na primer georeferenciranja podatkov in preverjanja kakovosti. Med samim razvojem metodologije smo posamezne korake predobdelave podatkov implementirali ločeno, zaradi lažjega nadzora nad tokom podatkov in lažjega odpravljanja napak. Za lažje razumevanje hitrosti predobdelave podajamo primer trajanja obdelave podatkov za konkreten dan meritev. Za primer smo vzeli meritev z dne 18.1.2014, ki je bila opravljena v Portoroškem zalivu. V skupnem modulu za predobdelavo podatkov so integrirani vsi postopki opisani v tem delu. Modul poskrbi za georeferenciranje, preverjanje kakovosti podatkov, izločanje nastavitvev sistema, uravnavanje kotnih odvisnosti in izračun značilk. V spodnjem seznamu podajamo čase izvajanja posameznih sklopov:

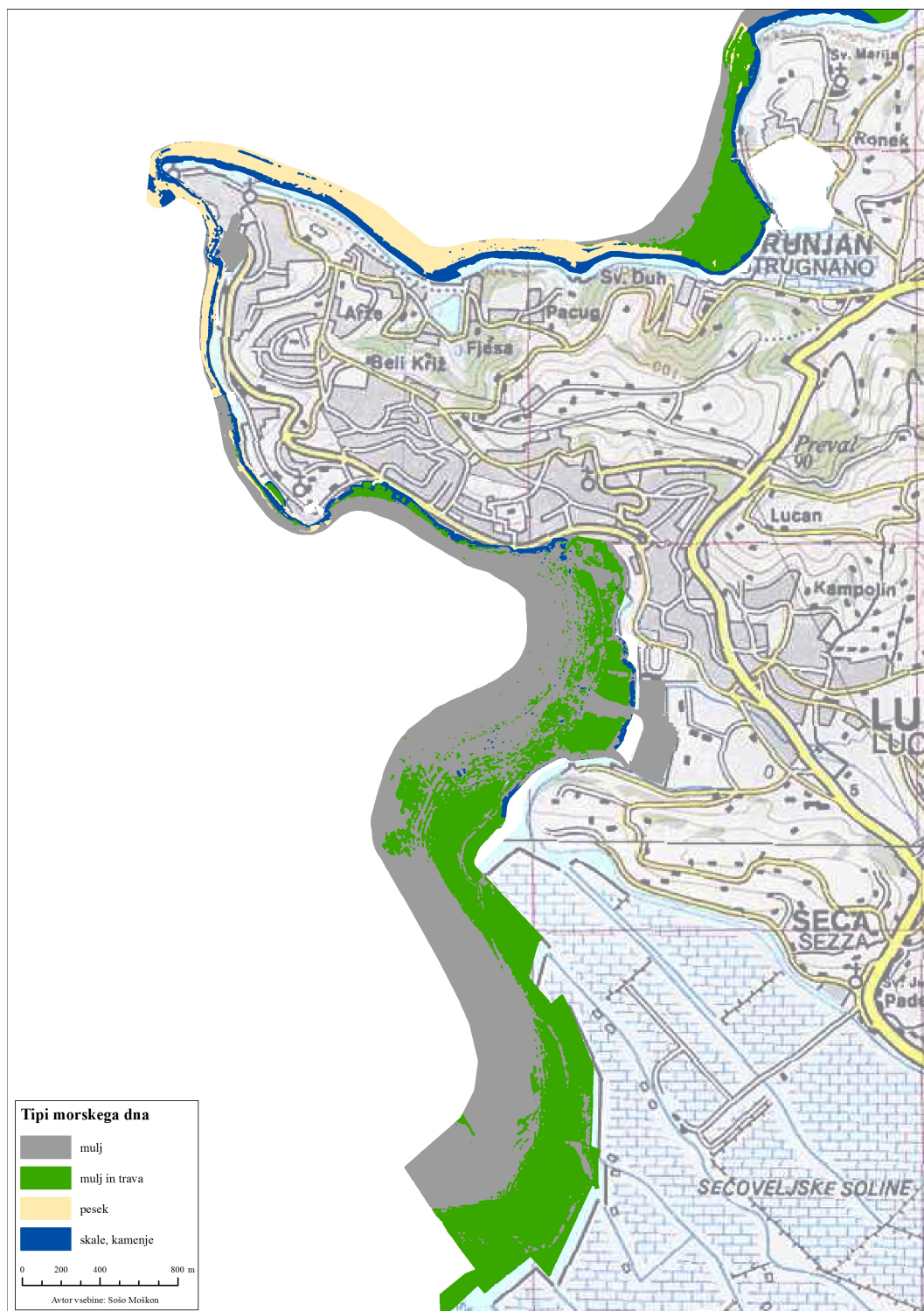
- Formatiranje in georeferenciranje podatkov: 27 minut in 35 sekund
- Izločanje nastavitvev sistema: 43 minut in 49 sekund
- Preverjanje kakovosti podatkov: 15 minut in 50 sekund
- Uravnavanje kotnih odvisnosti: 15 minut in 23 sekund
- Izračun značilk: 12 minut in 28 sekund

Skupno se je koda za predobdelavo podatkov izvajala 115 minut in 5 sekund, torej slabi dve uri. V tem času so bili obdelani podatki o 281.568 kvadratnih metrih morskega dna v Portoroškem zalivu. Skupno je bilo obdelanih 93.361.532 akustičnih odbojev, delež izločenih podatkov med preverjanjem kakovosti pa je znašal 1.5 %.

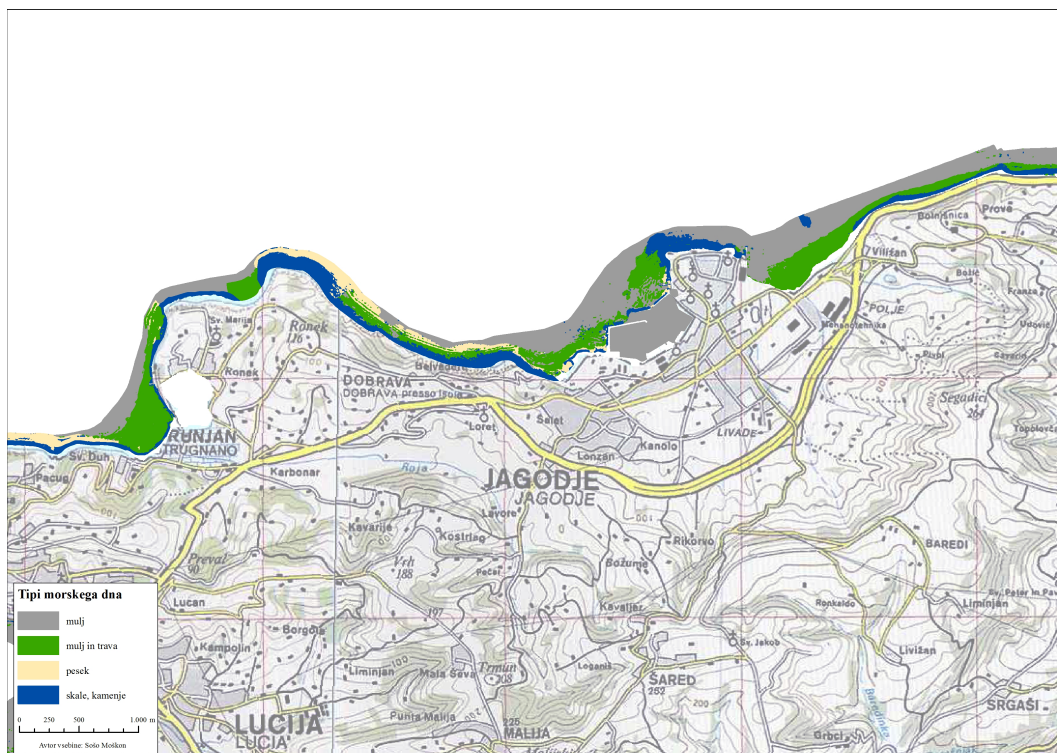
### 8.1.2 Rezultati in primerjava z obstoječimi kartami

Ker za Slovensko morje nimamo karte tipov morskega dna in niti karte morskih travnikov, smo za preverjanje rezultatov, kjer je bilo to mogoče, uporabili letalske posnetke sicer pa točkovno vzorčenje s pomočjo potapljačev. S pomočjo programa za prostorske obdelave Qgis smo označili vse morske travnike, večje od 300 kvadratnih metrov. Dobili smo 36 točk, na katerih smo izvedli točkovno vzorčenje. Na vsaki izmed označenih točk smo fotografirali morsko dno. Vzorčenje smo izvedli v enem delovnem dnevu. Na ta način smo potrdili prisotnost travnikov in ujemanje naših rezultatov z dejanskimi tipi dna. Izdelana karta morskih travnikov je zaradi preglednosti prikazana na treh slikah, Piranski akvatorij na sliki 8.2, Izolski na sliki 8.3 in Koprski na sliki 8.4. Čeprav so nas zanimali predvsem morski travniki, smo določili tudi druge tipe morskega dna. Morsko dno smo razdelili na kamnito dno, mulj, travnike in peščeno dno.

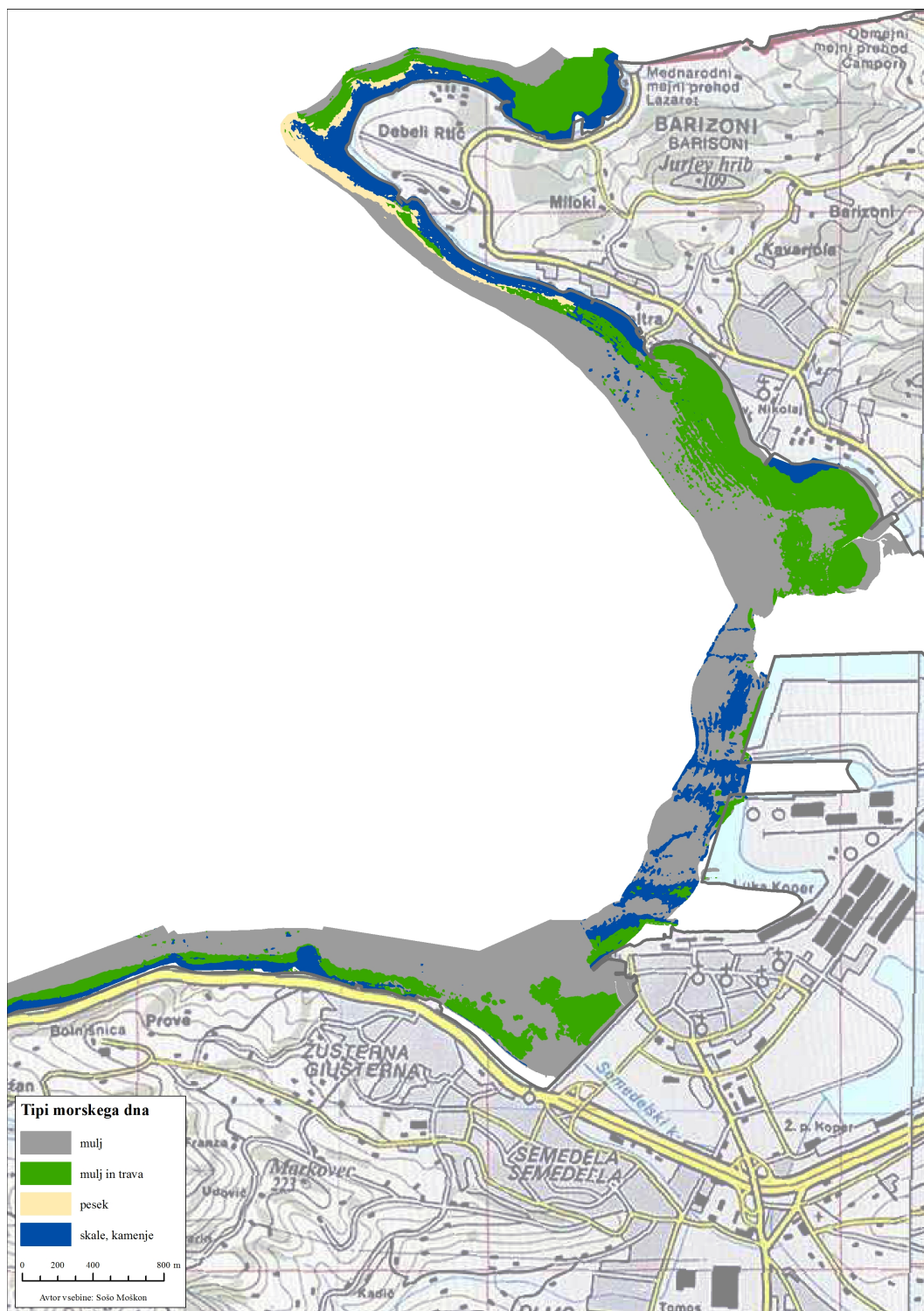
Podrobno analizo podvodnega reliefa v 500-metrskem pasu od obale slovenskega morja je sicer izdelala že Kolega [74]. Na podlagi digitalnega modela reliefa ločljivosti 0,5 x 0,5 m, pridobljenega s sonarskim snemanjem, je izračunala naklone in ekspozicije površja ter določila značilnosti stika med morjem in kopnim, ki je bilo v ta namen



Slika 8.2: Karta tipov morskega dna Piranskega akvatorija, dobljena po metodi akustične klasifikacije morskega dna. Podlaga: Geodetski zavod Slovenije in Mladinska knjiga, 1996, TK50 [50]



Slika 8.3: Karta tipov morskega dna Izolskega akvatorija, dobljena po metodi akustične klasifikacije morskega dna. Podlaga: Geodetski zavod Slovenije in Mladinska knjiga, 1996, TK50 [50]



Slika 8.4: Karta tipov morskoga dna Koprskega akvatorija, dobljena po metodi akustične klasifikacije morskoga dna. Podlaga: Geodetski zavod Slovenije in Mladinska knjiga, 1996, TK 50[50]

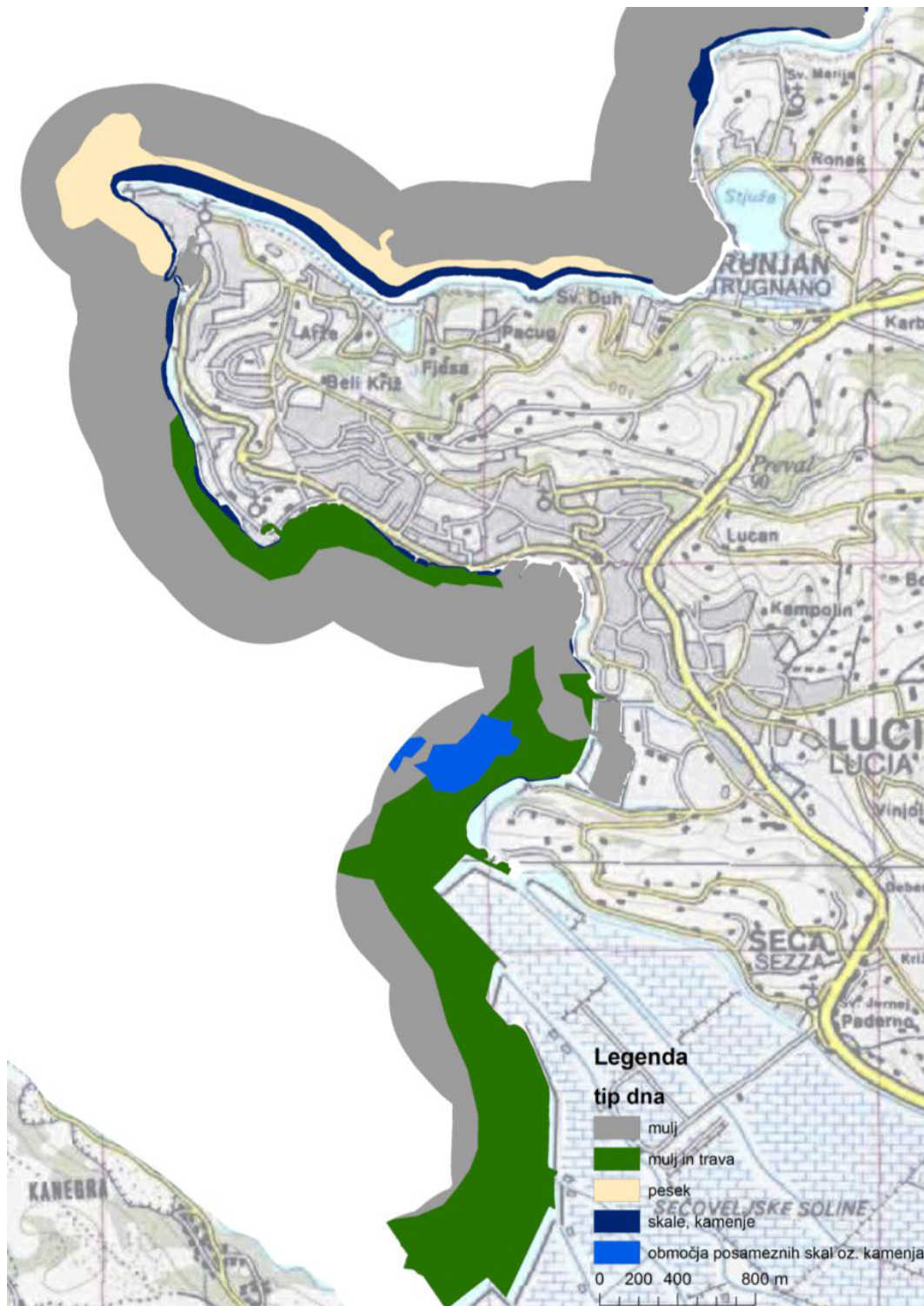
v 500-metrskem pasu od obale na kopno posneto s tehnologijo LiDAR. Določila je 5 različnih tipov morskega dna, od katerih po njenih ugotovitvah prevladuje muljasto dno (77 %), 13 % prekrivata mulj in trava, 6 % dna je skalnatega oz. kamnitega, 2 % peščenega, prav toliko pa ga prekrivajo tudi posamezna območja skal in kamenja. Natančno je popisala tudi človekove posege v morsko dno, pri čemer ima s plovnimi kanali in pristaniško infrastrukturo največji vpliv Luka Koper (291,5 ha), omenja pa še kanalizacijske cevi, marine, arheološke ostanke (npr. odtis ladje Rex), pomole na stebrih, sledi sider in verig, školjčičišča, ipd. Kolega v delu [74] uporablja podatke iz leta 2007. Tipe morskega dna določa ročno na podlagi zgolj vizualne tipizacije širšega morskega pasu na podlagi podatkov meritev z mnogosnopnim sonarjem. Na osnovnem rastru morskega dna z ločljivostjo 0,5 m x 0,5 m naredi tri različne prostorske analize (naklon površja, senčenje, azimut 150°, višina 15° in rang vrednosti v kvadratu 1 m x 1 m) za vsako celico posebej, ter nato s pregledovanjem in kombiniranjem omenjenih rastrov poskuša prepoznati ključne vzorce za različne tipe dna. Uporabljen je torej samo model globin in subjektivna ocena tipov morskega dna, postopek pa je ročen in torej precej zamuden.

Naše rezultate smo primerjali z rezultati omenjene študije. Karta, povzeta po Kolega [74] je predstavljena na treh slikah 8.6, 8.7 in 8.8. Prva prikazuje Piranski akvatorij, druga Izolskega in treja Koprškega. Ob primerjavi kart tipov morskega dna na podlagi podatkov iz let 2007 in 2014 opazimo precej razlik. Eno večjih neskladij je opazno na območju med Koprom in Izolo. Primerjava tega področja je prikazana na sliki 8.5. Z metodo akustične klasifikacije na podatkih iz leta 2014 smo na tem področju zaznali velik morski travnik. Na podatkih iz leta 2007 na istem področju morski travnik ni prisoten. Izkaže se, da morski travnik ni nov, ampak gre za pomanjkljivost stare metodologije. Z analizo batimetričnih modelov smo ugotovili, da je na podlagi stare metodologije (ročne klasifikacije s pomočjo naklonov in senčenj) zelo težko zaznati morski travnik na tem območju. Težava je predvsem v tem, da na tem območju morski travnik z globino postaja redkejši in počasi preide v mulj. Tako ni jasnega roba, kjer se morski travnik zaključi in ga je skoraj nemogoče zaznati na batimetričnem modelu. Nasprotno z metodo akustične klasifikacije tak morski travnik zaznamo, saj uporabljamo veliko več značilk (intenzitete, povprečja, standardni odkloni, višji momenti, itd.) ki jih poleg batimetričnega modela izračunamo tudi za slike akustičnih odbojev. Uporaba velikega števila značilk je mogoča zaradi avtomatiziranega procesa akustične klasifikacije s pomočjo metod strojnega učenja. Razvili smo metodologijo, s katero lahko učinkovito in v relativno kratkem času določimo tipe morskega dna in to zvezno na celotnem obravnavanem območju. Za spremljanje razširjenosti morskih travnikov bi bile zanimive ciklične meritve. Z razvito metodologijo lahko učinkovito spremljamo spremembe razširjenosti morskih travnikov v različnih časovnih intervalih. Kljub temu, da so pri primerjavi podatkov med letoma 2007 in 2014 opazne spremembe predvsem posledica drugačne oziroma izboljšane metodologije, bi s cikličnim snemanjem (na eno leto) lahko prišli do ugotovitev o stanju morskih travnikov ter trendih njihovega morebitnega upadanja oziroma rasti.

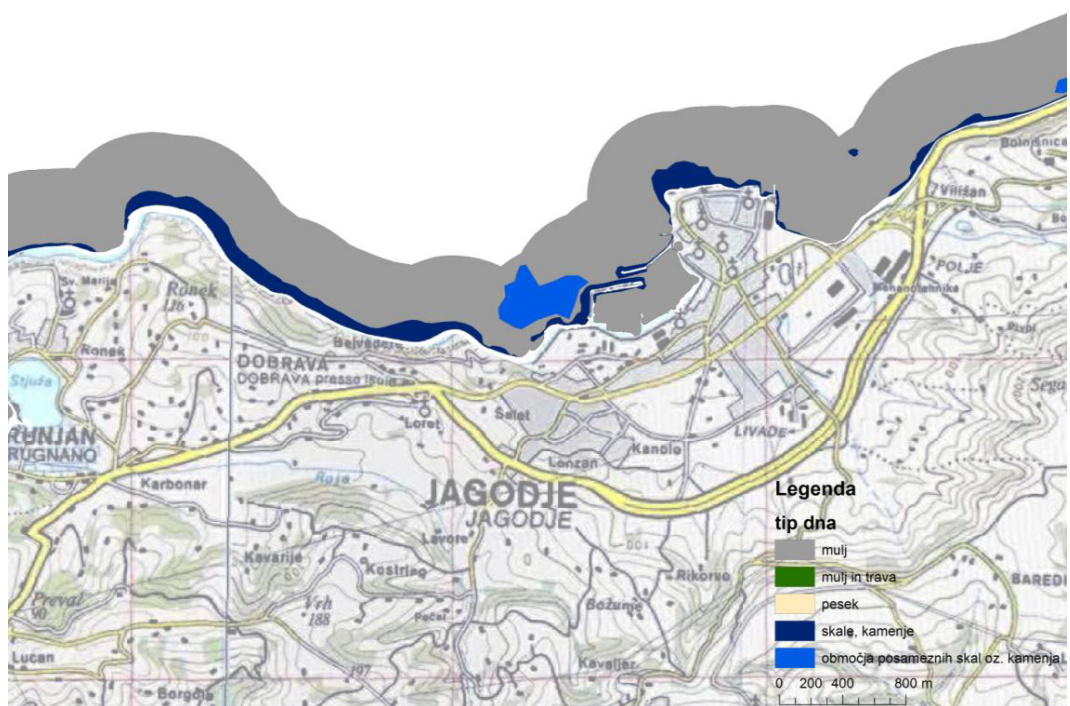




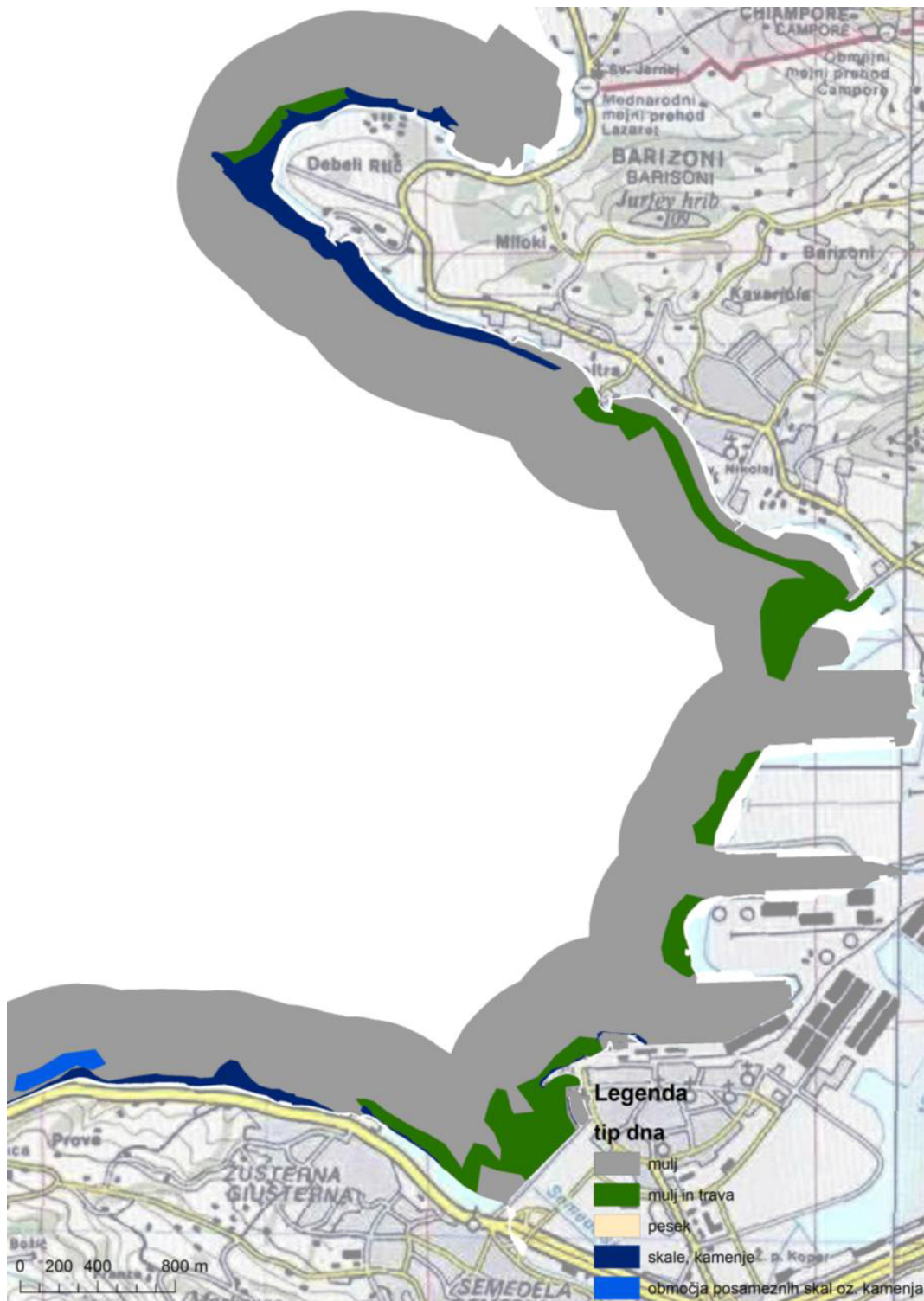
Slika 8.5: Primerjava kart tipov morskega dna za območje med Izolo in Koperom. Zgoraj je prikazana karta prirejena po Kolega [74], spodaj pa naša karta tipov morskega dna.



Slika 8.6: Karta tipov morskega dna, prirejena po Kolega [74].



Slika 8.7: Karta tipov morskega dna, prirejena po Kolega [74].



Slika 8.8: Karta tipov morskega dna, prirejena po Kolega [74].

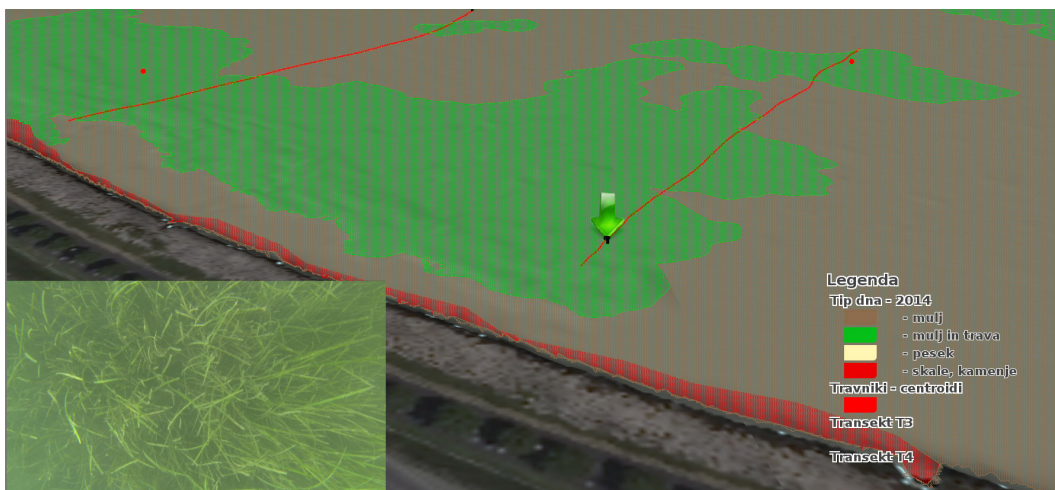
### 8.1.3 Predstavitev rezultatov v 3R GIS okolju

Za prikaz rezultatov v 3R tehniki smo razvili aplikacijo v okolju OpenSceneGraph viewer. Podatke predstavimo v treh dimenzijah, tako da najprej izdelamo digitalni model višin iz podatkov o globinah, nato pa nanj projiciramo karto tipov morskega dna. Aplikacija prikazuje tudi teren na kopnem. 3R prikaz tipov morskega dna je uporaben za načrtovanje upravljanja z morskim dnom, načrtovanje posegov v morsko dno in spremljanje sprememb na morskem dnu (posegi, krčenje ali širjenje morskih travnikov, zamuljevanje, ipd.). Geografski informacijski sistemi so v zadnjih dveh desetletjih postali nepogrešljiv pripomoček vsakogar, ki se ukvarja s načrtovanjem posegov v prostor, prostorskim modeliranjem in upravljanjem s kopnim in morjem. Pri tem so dobrodošla pomoč trirazsežnostna (3R) okolja, ki omogočajo prikaz stanja pred in po posegu. Z združitvijo uporabnosti 3R GIS okolja in realistično ponazoritvijo je uporabnikom olajšana izkušnja dojetanja stanja v prostoru in odločanja o bodočih projektih in analiza njihovega vpliva na okolje. Okolje omogoča prikaz in analize poljubnih rastrskih in vektorskih podatkovnih slojev, prav tako pa je mogoče v istem okolju hkrati prikazovati tudi 3R modele ter gibanje in spreminjanje posameznih modelov v prostoru skozi čas. 3R GIS orodje smo podrobneje opisali v delih [124], [136] in [65].

V 3R GIS okolju smo prikazali rezultate akustične klasifikacije morskega dna. Dodatno smo razvili prikaz video posnetkov in slik vzorčenja s potapljači in snemanja s plovilom. Okolje omogoča prikaz točk vzorčenja, ob kliku nanje pa prikaže fotografije morskega dna. Omogoča tudi prikaz linij video snemanja s plovilom in prostorsko sinhroniziran prikaz videoposnetka. Na sliki 8.9 je primer prikaza 3R podatkov. Prikazan je digitalni model višin kopnega s projiciranimi letalskimi posnetki za realistični prikaz. Za morsko dno smo uporabili digitalni model globin in nanj projicirali poligone tipov morskega dna. Rdeče pike predstavljajo točke vzorčenja, rdeče črte pa linije video snemanja s plovilom. Ob kliku na rdeče pike prikažemo fotografijo odgovarjajočega morskega dna. Ob kliku na zeleno puščico pričnemo v spodnjem levem kotu predvajati video posnetek, hkrati pa pomikamo zeleno puščico skladno s potjo plovila v času zajema video posnetka. Primer prikaza video posnetka je na sliki 8.10. Video posnetek je prikazan v spodnjem levem kotu. Zelena puščica se nahaja na območju, označenem kot morski travnik in tudi na video posnetku je vidno, da gre za morski travnik.



Slika 8.9: Prikaz tipov morskoga dna in točk vzorčenja v 3R GIS orodju.



Slika 8.10: Video posnetek s prikazom točne lokacije snemanja. Spodaj levo se predvaja video posnetek, hkrati pa se zelena puščica pomika skladno s potjo plovila v času zajema video posnetka.

## Poglavje 9

# Sklep

Osnovni cilj disertacije je razvoj praktično uporabne metode za klasifikacijo morskega dna. Razvili smo metodologijo, ki vsebuje vse korake od izvedbe meritev in predobdelave akustičnih podatkov do izračuna značilik in uporabe metod strojnega učenja za klasifikacijo morskega dna. Poudarek je na praktični uporabnosti razvite metodologije, saj je delo financirano v okviru operativnega programa razvoja človeških virov s prioriteto spodbujanja podjetništva in prilagodljivosti ter razvoja strokovnjakov in raziskovalcev za konkurenčnost podjetij in je torej aplikativno usmerjeno.

Razvili in opisali smo metodologijo za avtomatsko predobdelavo velikih količin podatkov, pridobljenih z meritvami mnogospopnega sonarja. Opisana metodologija vsebuje 6 korakov:

1. zajem in formatiranje surovih podatkov
2. preverjanje kakovosti in izločanje slabih podatkov
3. izločanje vpliva nastavitve moči oddajnika in občutljivosti sprejemnika
4. izločanje vpliva absorpcije in sferičnega širjenja zvoka
5. izločanje vpliva velikosti akustičnega odtisa
6. uravnavanje vpliva kotnih odvisnosti

Formatiranje surovih podatkov služi predvsem razširljivosti metodologije na druge MBES sisteme. Avtomatsko preverjanje kakovosti in izločanje slabih podatkov, je korak, ki bistveno pohitri klasifikacijo morskega dna. V primerjavi z ročnim preverjanjem podatkov je bistveno hitrejši, prečiščeni podatki pa so dovolj kakovostni za potrebe akustične klasifikacije morskega dna. Koraki 3, 4 in 5 temeljijo na teoretičnih osnovah s področja podvodne akustike. Dodali smo interpolacijo prehoda pri spremembi nastavitve med meritvami, ko pride do zamika med časom spremembe določene nastavitve in dejanskim časom prilagoditve sistema. Pri izločanju vpliva velikosti akustičnega odtisa smo teoretične osnove le deloma upoštevali, saj je velikost akustičnega odtisa močno odvisna od vpadnega kota akustičnega signala, slednjega pa ni možno dovolj natančno izračunati. Tako smo del izločanja vpliva

velikosti akustičnega odtisa združili z uravnavanjem vpliva kotnih odvisnosti in ga izvedli empirično. Uravnavanje vpliva kotnih odvisnosti temelji na empirično izpeljanih enačbah, ki smo jih tudi ustrezno ovrednotili. Poleg štirih pristopov, povzetih po obstoječi literaturi smo predlagali in ovrednotili dva nova pristopa, katerih cilj je združitev prednosti obstoječih pristopov. Za ovrednotenje smo uporabili klasifikacijsko točnost algoritma C4.5 na različno predobdelanih množicah enakih podatkov. Klasifikacijsko točnost smo ocenili po metodi 10 zagonov 10-kratnega prečnega preverjanja, značilnost razlik pa potrdili z enosmernim parnim T-testom. Najboljše rezultate smo dobili z uporabo pristopa `L_std` (lokalno pristop s standardnim odklonom) in našega predlaganega pristopa `G_L_std` (globalni in nato lokalni pristop s standardnim odklonom). Med ocenama pristopov `L_std` in `G_L_std` ni statistično značilnih razlik. Odločili smo se za uporabo pristopa `L_std`, zaradi manjše časovne zahtevnosti. Vse korake smo združili v enoten postopek predobdelave podatkov ki je popolnoma avtomatiziran in ne potrebuje nobenih nastavitvev operaterja.

Pri izračunu značilk smo podatke razdelili na celice različnih velikosti. Uporabili smo kvadratne celice s stranico dolžine 1 meter, 2 metra in 4 metre. Definirali smo izračun 20 značilk na 3 velikostih celic, torej skupno 60 značilk. Za ocenjevanje značilk smo uporabili razmerje informacijskega prispevka, ocenjevanje značilk z naključnimi gozdovi in algoritem ReliefF. Posameznim značilkam smo najprej določili najustreznejšo velikost celice za izračun značilke, nato pa značilke primerjali med seboj in določili najboljše značilke za problem akustične klasifikacije morskega dna. Izbrali smo naslednje značilke:

- `i` : intenziteto odboja, izračunano na celici velikosti 2x2 metra
- `z`: povprečno globino, izračunano na celici velikosti 1x1 meter
- `i_std`: standardni odklon intenzitete, izračunano na celici velikosti 2x2 metra
- `z_std`: standardni odklon globine, izračunano na celici velikosti 1x1 metra
- lastnosti matrike sopojavnosti sivin, izračunane na podatkih o globinah z velikostjo celice 4x4 metre:
  - `z_homogenost`
  - `z_kontrast`
  - `z_korelacija`
  - `z_razlicnost`

Skupno smo torej izbranih 8 značilk, ki smo jih v nadaljevanju uporabljali za akustično klasifikacijo morskega dna.

Opisali in izbrali smo algoritme strojnega učenja, ki so ustrezni za akustično klasifikacijo morskega dna. Ovrednotili smo algoritme:

- OneR klasifikator
- Naivni Bayesov klasifikator NB
- metoda podpornih vektorjev SVM



- metoda najbližjih sosedov kNN
- Odločitven pravila CN2
- Odločitvena drevesa C4.5
- metoda naključnih gozdov RF
- kombiniranje klasifikatorjev z glasovanjem VOTE

Uporabili smo implementacije algoritmov za strojno učenje v okolju ORANGE v kombinaciji s programskim jezikom Python in okolju WEKA v kombinaciji s programskim jezikom Java. Za ovrednotenje algoritmov smo potrebovali množico označenih podatkov, to so podatki, v katerih imamo izračunane vse prej omenjene značilke in označen razred - v našem primeru tip morskega dna. Izvedli smo hidrografske meritve na območju med Izolo in Koprom. Izbrano območje obsega 200 metrov širok in 900 metrov dolg obalni pas in pokriva 180.000 kvadratnih metrov morskega dna. Skupno je bilo zbranih 31 gigabajtov podatkov o preko 10.000.000 točkah morskega dna. Iz akustičnih podatkov o morskem dnu smo po razviti metodologiji izračunali vse izbrane značilke. Za določitev dejanskih tipov morskega dna smo uporabili tri metode - metodo točkovnega vzorčenja s pomočjo potapljačev, fotointerpretacijo letalskih posnetkov za določanje tipov morskega dna in metodo video snemanja morskega dna. Iz zbranih podatkov smo generirali učno množico podatkov v ustreznem formatu za uporabo v orodjih za strojno učenje. Z metodo 10-kratnega prečnega preverjanja smo ovrednotili algoritme strojnega učenja in z enosmernim parnim T-testom analizirali statistično značilnost razlik uspešnosti posameznih algoritmov. Rezultate najuspešnejših algoritmov smo podrobneje analizirali in opisali. Najboljše rezultate smo dobili z algoritmom odločitvenih dreves (97.64 % klasifikacijska točnost) in algoritmom naključnih gozdov (98.47 % klasifikacijska točnost), pri čemer so naključni gozdovi statistično značilno boljši od odločitvenih dreves pri parametru  $\alpha = 0.001$ . Rezultate klasifikatorjev smo še dodatno izboljšali z uporabo medianinega filtra na prostorskih podatkih o napovedanih razredih, ter tako izkoristili umeščenost podatkov v prostor. Z algoritmom naključnih gozdov v kombinaciji z medianinim filtrom smo dosegli klasifikacijsko točnost 98.94 %. Za potrebe akustične klasifikacije morskega dna priporočamo uporabo algoritma naključnih gozdov. V kolikor je potrebna interpretacija dobljenega modela, priporočamo uporabo algoritma odločitvenih dreves, saj ima slednji veliko bolj izrazen model. Izraznost dobljenega modela v nekaterih primerih odtehta nekoliko slabšo klasifikacijsko točnost.

Razvito metodologijo smo aplicirali na problem kartiranja morskih travnikov v slovenskem morju. Učinkovito kartiranje morskih travnikov je zelo aktualen problem, saj direktiva Evropske Unije določa ciklično spremljanje razširjenosti morskih travnikov [40]. Izvedli smo meritve celotne obale slovenskega morja, do globine 10 metrov. Podatke smo obdelali po razviti metodologiji in izdelali karto tipov morskega dna za celotno slovensko obalo. Določili smo 36 večjih morskih travnikov in prisotnost vseh potrdili z metodo točkovnega vzorčenja. Rezultate smo predstavili v 3R GIS okolju. Naše rezultate smo primerjali z rezultati študije iz leta 2007 [74], v kateri so avtorji za klasifikacijo morskega dna uporabili interpretacijo batimetričnega modela. Ob primerjavi kart tipov morskega dna na podlagi podatkov iz let

2007 in 2014 smo opazili precej razlik, ki smo jih podrobneje analizirali in ugotovili, da je do razlik prišlo zaradi pomanjkljivosti stare metodologije. Z aplikacijo razvite metodologije na problem kartiranja morskih travnikov smo potrdili praktično uporabnost naše metode in pokazali, da ima akustična klasifikacija morskega dna s podatki mnogospopnega sonarja velik potencial pri raziskavah morskega dna.

Pomembna je ugotovitev, da metoda omogoča hitro kartiranje in zvezno pokrivnost velikih območij morskega dna. Z mnogospopnim sonarjem lahko v enem dnevu izvedemo meritve in kartiramo območje veliko nekaj kvadratnih kilometrov. Prednost naše metode je tudi v robustnosti. Zajem podatkov z mnogospopnim sonarjem je neodvisen od prosojnosti vode in sončne svetlobe. Meritve je možno izvajati tudi v plitvih zalivih z zelo motno vodo. Prav tako nam natančni GNSS in INS sprejemniki omogočajo kakovostne meritve v vzvalovanem morju.

## 9.1 Smernice za nadaljnje delo

Zavedamo se, da naša metoda še ni popolna in da je še veliko prostora za raziskave. Opazili smo težave na prehodih med različnimi tipi morskega dna, kar je bilo pričakovano. V prihodnosti načrtujemo izboljšanje predobdelave podatkov, ki se je izkazala za ključen korak pri akustični klasifikaciji morskega dna. Veliko prostora za izboljšave vidimo predvsem pri uravnavanju vpliva kotnih odvisnosti, ki še vedno močno vplivajo na kakovost nekaterih značilk. Velik potencial razvite metodologije vidimo tudi v klasifikaciji sedimentov in izdelavi karte sedimentov. Karta sedimentov slovenskega morja je trenutno izdelana z interpolacijo na podlagi malega števila vzorcev morskega dna. Vzorce morskega dna bi lahko uporabili kot vhodni podatek in klasificirali sedimente na podlagi akustičnih podatkov.





# Literatura

- [1] K. Abukawa, M. Yamamuro, Z. Kikvidze, A. Asada, C. Xu in K. Sugimoto. Assessing the biomass and distribution of submerged aquatic vegetation using multibeam echo sounding in lake towada, japan. *Limnology*, 14(2013), 39–42.
- [2] M. A. Ainslie in J. G. McColm. A simplified formula for viscous and chemical absorption in sea water. *The Journal of the Acoustical Society of America*, 103(1998), 1671–1672.
- [3] V. M. Albers. Underwater acoustics handbook. *Physics Today*, 14(1965), 51.
- [4] Y. Allen, C. Wilson, H. Roberts in J. Supan. High resolution mapping and classification of oyster habitats in nearshore louisiana using sidescan sonar. *Estuaries*, 28(2005), 435–446.
- [5] J. T. Anderson. *Acoustic seabed classification of marine physical and biological landscapes*. International Council for the Exploration of the Sea, Copenhagen, 2007.
- [6] J. T. Anderson, D. Van Holliday, R. Kloser, D. G. Reid in Y. Simard. Acoustic seabed classification: current practice and future directions. *ICES Journal of Marine Science: Journal du Conseil*, 65(2008), 1004–1011.
- [7] L. Atallah, P. Probert Smith in C. Bates. Wavelet analysis of bathymetric sidescan sonar data for the classification of seafloor sediments in hopvågen bay - norway. *Marine Geophysical Researches*, 23(2002), 431–442.
- [8] C. Bates in E. Whitehead. Echoplus measurements in hopavagen bay, norway. *V Biennial Scientific Meeting*, volume 42, strani 34–43. The Oceanography Society, 2001.
- [9] J. Beaudoin, J. Hughes Clarke, E. Van Den Aemele in J. Gardner. Geometric and radiometric correction of multibeam backscatter derived from reson 8101 systems. *V Canadian Hydrographic Conference Proceedings*, strani 1–22. Canadian Hydrographic Association Toronto, Canada, 2002.
- [10] F. W. Bentrem, W. E. Avera in J. Sample. Estimating surface sediments using multibeam sonar. *arXiv preprint physics*, 47(2006), 37.
- [11] M. H. Bharati, J. J. Liu in J. F. MacGregor. Image texture analysis: methods and comparisons. *Chemometrics and intelligent laboratory systems*, 72(2004), 57–71.

- 
- [12] J. Blake. *The sea chart: the illustrated history of nautical maps and navigational charts*. Anova Books, New York, 2009.
- [13] P. Blondel. *The handbook of sidescan sonar*. Springer, Berlin, 2009.
- [14] J. Borum, C. M. Duarte, T. M. Greve in D. Krause-Jensen. *European seagrasses: an introduction to monitoring and management*. M&MS Project, Berlin, 2004.
- [15] L. Breiman. Random forests. *Machine learning*, 45(2001), 5–32.
- [16] C. Brown. Seafloor imagery, remote sensing and bathymetry: Acoustic ground discrimination system(agds). *Geological Association of Canada*, 47(2007), 3–10.
- [17] C. J. Brown in P. Blondel. Developments in the application of multibeam sonar backscatter for seafloor habitat mapping. *Applied Acoustics*, 70(2009), 1242–1247.
- [18] C. J. Brown, A. Mitchell, D. S. Limpenny, M. R. Robertson in N. Golding. Mapping seabed habitats in the firth of lorn off the west coast of scotland: evaluation and comparison of habitat maps produced using the acoustic ground-discrimination system, roxann, and sidescan sonar. *ICES Journal of Marine Science: Journal du Conseil*, 62(2005), 790–802.
- [19] D. Brownrigg. The weighted median filter. *Communications of the ACM*, 27(1984), 807–818.
- [20] D. Burns in R. Osfield. *Open scene graph a: Introduction, b: Examples and applications*. IEEE, Chicago, 2004.
- [21] B. Chakraborty, V. Kodagali in J. Baracho. Sea-floor classification using multi-beam echo-sounding angular backscatter data: a real-time approach employing hybrid neural network architecture. *IEEE Journal of Oceanic Engineering*, 28(2003), 121–128.
- [22] P. Chapman, J. Clinton, R. Kerber, T. Khabaza, T. Reinartz, C. Shearer in R. Wirth. *CRISP-DM 1.0 Step-by-step data mining guide*. IBM Corporation, New York, 2010.
- [23] T. Chen, K.-K. Ma in L.-H. Chen. Tri-state median filter for image denoising. *IEEE Transactions on Image Processing*, 8(1999), 1834–1838.
- [24] P. Clark in T. Niblett. The cn2 induction algorithm. *Machine learning*, 3(1989), 261–283.
- [25] J. E. H. Clarke, L. A. Mayer in D. E. Wells. Shallow-water imaging multibeam sonars: a new tool for investigating seafloor processes in the coastal zone and on the continental shelf. *Marine Geophysical Researches*, 18(1996), 607–629.

- [26] T. J. Cleophas in A. H. Zwinderman. Bonferroni t-test. V *Statistical Analysis of Clinical Data on a Pocket Calculator*, strani 41–42. Springer, 2011.
- [27] J. Collier in C. Brown. Correlation of sidescan backscatter with grain size distribution of surficial seabed sediments. *Marine Geology*, 214(2005), 431–449.
- [28] J. Cristy. *ImageMagick User Guide*. E.I. du Pont de Nemours and Company Incorporated, Wilmington, 1999.
- [29] P. Dartnell in J. V. Gardner. Predicting seafloor facies from multibeam bathymetry and backscatter data. *Photogrammetric Engineering & Remote Sensing*, 70(2004), 1081–1091.
- [30] J. Demšar, T. Curk, A. Erjavec, Č. Gorup, T. Hočevar, M. Milutinovič, M. Možina, M. Polajnar, M. Toplak, A. Starič, et al. Orange: data mining toolbox in python. *The Journal of Machine Learning Research*, 14(2013), 2349–2353.
- [31] R. J. Diaz, M. Solan in R. M. Valente. A review of approaches for classifying benthic habitats and evaluating habitat quality. *Journal of environmental management*, 73(2004), 165–181.
- [32] S. Džeroski in B. Ženko. Is combining classifiers with stacking better than selecting the best one? *Machine learning*, 54(2004), 255–273.
- [33] H. E. Edgerton. *Sonar images*. Prentice-Hall, Englewood, 1986.
- [34] A. Eleftheriou. *Methods for the study of marine benthos*. John Wiley & Sons, Chicester, 2013.
- [35] V. L. Ferrini in R. D. Flood. The effects of fine-scale surface roughness and grain size on 300 khz multibeam backscatter intensity in sandy marine sedimentary environments. *Marine Geology*, 228(2006), 153–172.
- [36] M. Finkbeiner, B. Stevenson in R. Seaman. *Guidance for benthic habitat mapping: an aerial photographic approach*. National Ocean Service/Coastal Services Center, Charleston, 2001.
- [37] C. W. Finkl, L. Benedet in J. L. Andrews. Interpretation of seabed geomorphology based on spatial analysis of high-density airborne laser bathymetry. *Journal of Coastal Research*, 21(2005), 501–514.
- [38] J. P. Fish in H. A. Carr. *Sound underwater images: a guide to the generation and interpretation of side scan sonar data*. Lower Cape Pub, Cape May, 1990.
- [39] F. Fisher in V. Simmons. Sound absorption in sea water. *The Journal of the Acoustical Society of America*, 62(1977), 558–564.
- [40] J. Foden in D. Brazier. Angiosperms (seagrass) within the eu water framework directive: A uk perspective. *Marine Pollution Bulletin*, 55(2007), 181–195.

- [41] L. Fonseca in L. Mayer. Remote estimation of surficial seafloor properties through the application angular range analysis to multibeam sonar data. *Marine Geophysical Researches*, 28(2007), 119–126.
- [42] L. Fonseca, L. Mayer in B. Kraft. Seafloor characterization through the application of avo analysis to multibeam sonar data. V *Conference proceedings: boundary influences in high frequency shallow water acoustics-University of Bath*, strani 4–9, 2005.
- [43] R. Foster-Smith in I. Sotheran. Mapping marine benthic biotopes using acoustic ground discrimination systems. *International Journal of Remote Sensing*, 24(2003), 2761–2784.
- [44] J. Franklin. *Mapping species distributions: spatial inference and prediction*. Cambridge University Press, Cambridge, 2010.
- [45] R. Freitas, A. M. Rodrigues, E. Morris, J. L. Perez-Llorens in V. Quintino. Single-beam acoustic ground discrimination of shallow water habitats: 50khz or 200khz frequency survey? *Estuarine, Coastal and Shelf Science*, 78(2008), 613–622.
- [46] J. Fridl, N. Kolega in A. Žerjal. Pomen digitalnega batimetričnega modela za trajnostni razvoj morja/the significance of digital bathymetric model for sustainable development of the sea. *Geodetski vestnik*, 52(2008), 854–866.
- [47] A. Gaspari, M. Erič in A. Maver. *Potopljena preteklost: arheologija vodnih okolij in raziskovanje podvodne kulturne dediščine v Sloveniji : zbornik ob 128-letnici Dežmanovih raziskav Ljubljani na Vrhniki (1884-2012)*, strani 81–90. Didakta, 2012.
- [48] A. Gavrilov, P. Siwabessy in I. Parnum. Multibeam echo sounder backscatter analysis: theory review, methods and application to sydney harbour swath data. Tehnicno poročilo, Centre for Marine Science and Technology, Curtin University of Technology, Kent St, 2005.
- [49] Geodetski inštitut Slovenije, Ministrstvo za okolje in prostor. *Državno omrežje stalnih GNSS-postaj SIGNAL*, 2015.
- [50] Geodetski zavod in Mladinska knjiga. *Topografska karta 1 : 50.000*, 1996.
- [51] GoPro. *GoPro Hero 3 black edition User Manual*. GoPro, Washington, 2014.
- [52] S. P. Greenstreet, G. J. Holland, E. J. Guirey, E. Armstrong, H. M. Fraser in I. M. Gibb. Combining hydroacoustic seabed survey and grab sampling techniques to assess local sandeel population abundance. *ICES Journal of Marine Science: Journal du Conseil*, 67(2010), 971–984.
- [53] G. C. Guenther, A. G. Cunningham, P. E. LaRocque in D. J. Reid. Meeting the accuracy challenge in airborne lidar bathymetry. *EARSel, Dresden*, (2000).



- [54] M. Hall, E. Frank, G. Holmes, B. Pfahringer, P. Reutemann in I. H. Witten. The weka data mining software: an update. *ACM SIGKDD explorations newsletter*, 11(2009), 10–18.
- [55] R. M. Haralick, K. Shanmugam in I. H. Dinstein. Textural features for image classification. *IEEE Transactions on Systems, Man and Cybernetics*, 3(1973), 610–621.
- [56] J. A. Hartigan in M. A. Wong. Algorithm as 136: A k-means clustering algorithm. *Applied statistics*, 28(1979), 100–108.
- [57] H. C. Hass. Seafloor classification using the roxann ground-discrimination system and a new tool for visualization and classification. V *GV Sediment Meeting, Hamburg*, strani 102–124, 2012.
- [58] S. Hiroyuki. *Olympus TG-2 Instruction Manual*. Olympus Imaging Corp., Shinjuku, 2013.
- [59] R. C. Holte. Very simple classification rules perform well on most commonly used datasets. *Machine learning*, 11(1993), 63–90.
- [60] Hypack. *HYPACK MAX, hydrographic survey software user manual*. Hypack, Middletown, 2013.
- [61] D. Ierodiaconou, L. Laurensen, S. Burq in M. Reston. Marine benthic habitat mapping using multibeam data, georeferenced video and image classification techniques in victoria, australia. *Journal of spatial science*, 52(2007), 93–104.
- [62] IHO. *IHO Standards for Hydrographic Surveys*. International Hydrographic Bureau, Monaco, 2008.
- [63] S. Innangi, M. Barra, G. Di Martino, I. Parnum, R. Tonielli in S. Mazzola. Reson seabat 8125 backscatter data as a tool for seabed characterization (central mediterranean, southern italy): Results from different processing approaches. *Applied Acoustics*, 87(2015), 109–122.
- [64] K. B. Irani. Multi-interval discretization of continuous-valued attributes for classification learning. V *Proceedings of the International Joint Conference on Uncertainty in AI*, strani 1022–1027, Chambery, 1993.
- [65] D. Jezeršek, B. Krapež in D. Presl Radi. Sistem za podporo gasilcem na intervenciji. V *Geografski informacijski sistemi v Sloveniji*, strani 249–255, Ljubljana, 2012.
- [66] N. Jogan, M. Kaligarič, I. Leskovar-Štamcar, A. Seliškar in J. Dobravec. *Habitatni tipi Slovenije: tipologija*. Ministrstvo za okolje, prostor in energijo, Agencija RS za okolje, Ljubljana, 2004.
- [67] J. L. Jones. *Mapping a flood... before it happens*. US Department of the Interior, US Geological Survey, Washington, 2004.

- [68] I. Karničnik, M. Klanjšček in D. Radovan. Sodobno hidrografsko kartiranje in dokumentiranje slovenskega morja. *Geodetski vestnik*, 50(2006), 37–43.
- [69] A. Kenny, I. Cato, M. Desprez, G. Fader, R. Schüttenhelm in J. Side. An overview of seabed-mapping technologies in the context of marine habitat classification. *ICES Journal of Marine Science: Journal du Conseil*, 60(2003), 411–418.
- [70] K. Kira in L. A. Rendell. The feature selection problem: Traditional methods and a new algorithm. V *AAAI*, strani 129–134, San Jose, 1992.
- [71] H. Kirkman. Baseline and monitoring methods for seagrass meadows. *Journal of Environmental Management*, 47(1996), 191–201.
- [72] M. Klanjšček. *Navtični vodnik slovenskega morja in obale*. Ministrstvo za promet Republike Slovenije, Geodetski inštitut Slovenije, Ljubljana, 2005.
- [73] A. Köhne in M. Wössner. Waas and egnos, [http://www.kowoma.de/en/gps/waas\\_egnos.htm](http://www.kowoma.de/en/gps/waas_egnos.htm), 2011.
- [74] N. Kolega, A. Gosar in M. O. Adamič. *Medsebojno vplivanje kopnega in morja:(določanje značilnosti stika med kopnim in morjem s pomočjo lidar-skih in sonarskih snemanj): doktorska disertacija*. Univerza na primorskem, Fakulteta za humanistične študije, Koper, 2009.
- [75] N. Kolega in M. Poklar. Morphological analysis of the slovenian coast with data from lidar and sonar ranging. *Geografski Zbornik/Acta Geographica Slovenica*, 52(2012), 121–140.
- [76] I. Kononenko. Estimating attributes: analysis and extensions of relief. V *Machine Learning: ECML-94*, strani 171–182. Springer, 1994.
- [77] I. Kononenko. *Strojno učenje*. Fakulteta za računalništvo in informatiko, Ljubljana, 2005.
- [78] M. Kotarac, A. Lešnik, D. Erjavec, M. Jakopič, B. Rozman, B. Trčak, N. Jogan, O. Urbanc-Berčič, M. Germ in M. Kaligarič. *Opredelitev območij evropsko pomembnih negozdnih habitatnih tipov s pomočjo razširjenosti značilnih rastlinskih vrst: končno poročilo*. Center za kartografijo favne in flore, Ljubljana, 2004.
- [79] D. Krause-Jensen, E. Diaz Almela, A. H. Cunha in T. Greve. Have seagrass distribution and abundance changed. V *European seagrasses: an introduction to monitoring and management*, strani 33–40. Seagrasses.org, 2004.
- [80] T. Le Bas in V. Huvenne. Acquisition and processing of backscatter data for habitat mapping—comparison of multibeam and sidescan systems. *Applied Acoustics*, 70(2009), 1248–1257.

- [81] M. A. Lefsky, W. B. Cohen, G. G. Parker in D. J. Harding. Lidar remote sensing for ecosystem studies lidar, an emerging remote sensing technology that directly measures the three-dimensional distribution of plant canopies, can accurately estimate vegetation structural attributes and should be of particular interest to forest, landscape, and global ecologists. *BioScience*, 52(2002), 19–30.
- [82] C. Lindenbaum, J. Bennell, E. Rees, D. McClean, W. Cook, A. Wheeler in W. Sanderson. Small-scale variation within a modiolus modiolus (mollusca: Bivalvia) reef in the irish sea: I. seabed mapping and reef morphology. *Journal of the Marine Biological Association of the UK*, 88(2008), 133–141.
- [83] L. Lipej, B. Mavrič in S. Moškon. New recordings of opisthobranch mollusks (mollusca: Opisthobranchia) in the slovenian portion of the adriatic sea. *Annales, Series Historia Naturalis*, 22(2012), 133–136.
- [84] L. Lipej in S. Moškon. On the record of the moray eel (*muraena helena* linnaeus, 1758) in slovenian coastal waters (gulf of trieste, northern adriatic). *Annales, Series Historia Naturalis*, 21(2011), 157–160.
- [85] V. Lucieer. The application of automated segmentation methods and fragmentation statistics to characterise rocky reef habitat. *Journal of spatial science*, 52(2007), 81–91.
- [86] V. L. Lucieer. Object-oriented classification of sidescan sonar data for mapping benthic marine habitats. *International Journal of Remote Sensing*, 29(2008), 905–921.
- [87] X. Lurton. *An introduction to underwater acoustics: principles and applications*. Springer, Berlin, 2002.
- [88] L. A. Mayer. Frontiers in seafloor mapping and visualization. *Marine Geophysical Researches*, 27(2006), 7–17.
- [89] A. Meehan, R. Williams in F. Watford. Detecting trends in seagrass abundance using aerial photograph interpretation: problems arising with the evolution of mapping methods. *Estuaries*, 28(2005), 462–472.
- [90] H. W. Menard, R. Dill, E. Hamilton, D. Moore, G. Shumway, M. Silverman in H. Stewart. Underwater mapping by diving geologists. *AAPG Bulletin*, 38(1954), 129–147.
- [91] Midland. *Midland XTC-280 manual*. Midland radio corporation, Kansas City, 2015.
- [92] R. Mount. The application of digital aerial photography to shallow water seabed mapping and monitoring-how deep can you see. V *Proceedings of coastal GIS*. University of Victoria, 2003.
- [93] S. Moškon, J. Žibert in B. Kavšek. Kartiranje morskih travnikov s pomočjo akustičnih podatkov. *Geographical Bulletin*, (2015 - sprejeto v objavo).

- [94] M. Možina, J. Demšar, M. Kattan in B. Zupan. *Nomograms for visualization of naive Bayesian classifier*. Springer, Berlin, 2004.
- [95] F. Nitsche, W. Ryan, S. Carbotte, R. Bell, A. Slagle, C. Bertinado, R. Flood, T. Kenna in C. McHugh. Regional patterns and local variations of sediment distribution in the hudson river estuary. *Estuarine, Coastal and Shelf Science*, 71(2007), 259–277.
- [96] R. V. O’neill, K. H. Riitters, J. Wickham in K. B. Jones. Landscape pattern metrics and regional assessment. *Ecosystem health*, 5(1999), 225–233.
- [97] A. Orłowski. Acoustic seabed classification applied to baltic benthic habitat studies: a new approach. *Oceanologia*, 49(2007), 229–243.
- [98] M. Orožen Adamič. Prispevek k poznavanju izoblikovanosti podvodnega reliefa slovenske obale. *Geografski vestnik*, 53(1981), 39–46.
- [99] D. P. Paine in J. D. Kiser. *Aerial photography and image interpretation*. John Wiley & Sons, Chicester, 2012.
- [100] P. Pandian, J. Ruscoe, M. Shields, J. Side, R. Harris, S. Kerr in C. Bullen. Seabed habitat mapping techniques: an overview of the performance of various systems. *Mediterranean Marine Science*, 10(2009), 29–44.
- [101] I. Parnum, A. Gavrilov, P. Siwabessy in A. Duncan. Analysis of high-frequency multibeam backscatter statistics from different seafloor habitats. V *Proceedings of the Eighth European Conference on Underwater Acoustics (Carvoeiro, Portugal)*, strani 775–780, 2006.
- [102] I. M. Parnum. *Benthic habitat mapping using multibeam sonar systems*. Doktorska disertacija, Curtin University of Technology, 2007.
- [103] V. Pasqualini, C. Pergent-Martini, G. Pergent, M. Agreil, G. Skoufas, L. So-urbes in A. Tsirika. Use of spot 5 for mapping seagrasses: An application to posidonia oceanica. *Remote Sensing of Environment*, 94(2005), 39–45.
- [104] N. Pavešič. *Razpoznavanje vzorcev: uvod v analizo in razumevanje vidnih in slušnih signalov*. Fakulteta za elektrotehniko, Ljubljana, 2000.
- [105] U. Pelizzari in S. Tovaglieri. *Manual of Freediving: Underwater on a single breath*. Idelson Gnocchi Pub, Napoli, 2004.
- [106] M. Peterlin. Numo - začetna presoja morskih voda v pristojnosti rs, bistvene lastnosti in značilnosti morskih voda. Tehnicno poročilo inštitut za vode republike slovenije, Ljubljana, 2013.
- [107] R. C. Phillips in C. P. McRoy. *Seagrass research methods*. Unesco, Paris, 1990.
- [108] R. A. Pickrill in B. J. Todd. The multiple roles of acoustic mapping in integrated ocean management, canadian atlantic continental margin. *Ocean and Coastal Management*, 46(2003), 601–614.

- [109] J. Platt. Fast training of support vector machines using sequential minimal optimization. V *Advances in Kernel Methods - Support Vector Learning*, MIT Press, Cambridge, 1999.
- [110] M. Poklar. *Metodologija geografskega raziskovanja stika sladkovodnega in morskega okolja na primeru Semedelskega zaliva*. Doktorska disertacija, Fakulteta za humanistične študije, Koper, 2015.
- [111] M. Poklar, S. Moškon in M. Berden Zrimec. Klasifikacija tipov morskega dna z določitvijo obsežnejšega morskega rastja na morskem dnu. Tehnicno poročilo, Harpha sea, d.o.o. Koper, 2014.
- [112] J. Preston. Automated acoustic seabed classification of multibeam images of stanton banks. *Applied Acoustics*, 70(2009), 1277 – 1287.
- [113] J. M. Preston in A. C. Christney. Compensation of sonar image data primarily for seabed classification, Mar. 15 2005. US Patent 6.868.041.
- [114] QDC Team. Qgis geographic information system. *Open Source Geospatial Foundation Project. Vienna, Austria: <http://qgis.osgeo.org>*, (2013).
- [115] QPS. *QINSy User Manual*. QPS, Zeist, 2006.
- [116] J. R. Quinlan. *C4. 5: programs for machine learning*. Elsevier, Amsterdam, 2014.
- [117] B. M. Riegl in S. J. Purkis. Detection of shallow subtidal corals from ikonos satellite and qtc view (50, 200 khz) single-beam sonar data (arabian gulf; dubai, uae). *Remote Sensing of Environment*, 95(2005), 96–114.
- [118] R. Rinehart, D. J. Wright, E. R. Lundblad, E. M. Larkin, J. Murphy in L. Cary-Kothera. Arcgis 8. x benthic terrain modeler: Analysis in american samoa. V *Proceedings of the 24th Annual ESRI User Conference*, strani 1124–1138, San Diego, 2004.
- [119] J. Roberts, C. Brown, D. Long in C. Bates. Acoustic mapping using a multibeam echosounder reveals cold-water coral reefs and surrounding habitats. *Coral Reefs*, 24(2005), 654–669.
- [120] C. N. Rooper. *Marine Habitat Mapping Technology for Alaska*, poglavje Underwater video sleds: Versatile and cost effective tools for habitat mapping, strani 99–107. Alaska Sea Grant, Fairbanks, 2008.
- [121] A. Secchi. *Sul moto ondoso del mare e su le correnti di esso specialmente su quelle littorali*, poglavje Esperimente per determinare la trasparenza del mare, strani 258–288. Accademia delle Scienze, Roma, 1866.
- [122] P. J. Siwabessy. *Overview of coastal water habitat mapping research for the Coastal CRC*. Cooperative Research Centre for Coastal Zone, Estuary and Waterway Management, Curtin, 2006.

- [123] P. Slavec. Analiza morfologije morskega dna slovenskega morja. Magistrsko delo, Univerza v Ljubljani, Naravoslovnotehniška fakuleta, Ljubljana, 2012.
- [124] D. Sopič, v. Rudman, A. Žerjal, D. Presl Radi in J. Grahor. 3R GIS okolje za načrtovanje in prikaz posegov v prostoru. V *Geografski informacijski sistemi v Sloveniji*, strani 183–190, Ljubljana, 2012.
- [125] I. Sotheran, R. Foster-Smith in J. Davies. Mapping of marine benthic habitats using image processing techniques within a raster-based geographic information system. *Estuarine, Coastal and Shelf Science*, 44(1997), 25–31.
- [126] H. Stoelinga. *PDS2000 V3.9 release notes*. Teledyn reson B.V., Rotterdam, 2015.
- [127] B. J. Todd, G. B. Fader, R. C. Courtney in R. A. Pickrill. Quaternary geology and surficial sediment processes, Browns Bank, Scotian Shelf based on multibeam bathymetry. *Marine Geology*, 162(1999), 165–214.
- [128] G. Tsoumakas, I. Katakis in I. Vlahavas. Effective voting of heterogeneous classifiers. V *Machine Learning: ECML 2004*, strani 465–476. Springer, 2004.
- [129] R. van Overmeeren, J. Craeymeersch, J. van Dalssen, F. Fey, S. van Heteren in E. Meesters. Acoustic habitat and shellfish mapping and monitoring in shallow coastal water—sidescan sonar experiences in the netherlands. *Estuarine, Coastal and Shelf Science*, 85(2009), 437–448.
- [130] H. Van Rein, C. Brown, R. Quinn in J. Breen. A review of sublittoral monitoring methods in temperate waters: a focus on scale. *Underwater Technology*, 28(2009), 99.
- [131] B. K. Walker, B. Riegl in R. E. Dodge. Mapping coral reef habitats in southeast florida using a combined technique approach. *Journal of Coastal Research*, 24(2008), 1138–1150.
- [132] K.-Y. Wong, K.-L. Man in W.-Y. Chan. Real-time kinematic spans the gap. *GPS world*, 12(2001), 10–19.
- [133] C. Yeung in R. A. McConnaughey. acoustic backscatter from a sidescan sonar to explain fish and invertebrate distributions: a case study in Bristol Bay, Alaska. *ICES Journal of Marine Science: Journal du Conseil*, 65(2008), 242–254.
- [134] B. Ženko. Izboljšave metode skladanja klasifikatorjev. Magistrsko delo, Fakulteta za računalništvo in informatiko, Ljubljana, 2003.
- [135] A. Žerjal, S. Poglajen in I. Rant. Hidrografske meritve slovenskega morja na območju koprskega zaliva. Tehnicno poročilo, Harpha sea, d.o.o. Koper, 2006.
- [136] T. Žerjal, N. Kolega in M. Poklar. Izdelava GIS orodja za podporo strokovnjakom za ohranjanje stavbne dediščine. V *Geografski informacijski sistemi v Sloveniji*, strani 225–232, Ljubljana, 2012.







# Zahvala

Zahvalil bi se mojim najbližjim, ki ste mi v vsem tem času potrpežljivo stali ob strani in na različne načine prispevali svoj kamenček v mozaik mojega izobraževanja in raziskovanja in posebej Mateji, ki mi je vedno stala ob strani. Predvsem pa bi se zahvalil Nalu, ker je moj sonček in me vedno spravlja v dobro voljo, tudi v najtežjih trenutkih ;)

Zahvalil bi se podjetju Harpha Sea, d.o.o. Koper, Mojci in ostalim sodelavcem, ki so mi pomagali pri raziskovalnem delu in pripravi doktorske disertacije. Zahvalil bi se tudi dr. Branku Kavšku in dr. Janezu Žibertu za njuno strokovno in ostalo pomoč pri pripravi doktorske disertacije.

Operacijo delno financira Evropska unija, in sicer iz Evropskega socialnega sklada. Operacija se izvaja v okviru Operativnega programa razvoja človeških virov za obdobje 2007 – 2013, 1. razvojne prioritete: Spodbujanje podjetništva in prilagodljivosti, prednostne usmeritve 1.1.: Strokovnjaki in raziskovalci za konkurenčnost podjetij.





