

Венцеслав Никифоров Димитров

Оценяване на тематичната точност на пространствени данни за земното покритие получени от спътникови изображения чрез качествено-количествен подход

А В Т О Р Е Ф Е Р А Т

на дисертационен труд
за присъждане на образователна и научна степен
„ДОКТОР“

Област на висше образование: 4. „Природни науки, математика и информатика“
Професионално направление: 4.4. „Науки за Земята“
Научна специалност: „Дистанционни изследвания на Земята и планетите“

Научен ръководител:

Проф. д-р Румен Недков Проф. д-р Георги Желев

София
2022 г.

Настоящото изследване е извършено в рамките на докторантура на самостоятелна подготовка в секция „Системи за дистанционни изследвания“ към Института за космически изследвания и технологии при Българската академия на науките (ИКИТ-БАН). Дисертационният труд е обсъден на заседание на еднократно разширен Научен семинар на секция „Системи за дистанционни изследвания“ при ИКИТ-БАН, проведено на 25.05.2022 г. и е разкрита процедура за публична защита с решение на Научния съвет на ИКИТ-БАН пред Научно жури в състав:

Вътрешни членове:

1. Проф. д-р Деница Борисова – ИКИТ-БАН
2. Доц. д-р Христо Николов – ИКИТ-БАН
3. Доц. д-р Даниела Аветисян – ИКИТ-БАН (резервен)

Външни членове:

1. Проф. д-р Никола Колев - Русенски университет "Ангел Кънчев"
2. Проф. д-р Юлин Тепелиев - Лесотехнически университет - София
3. Доц. д-р Радка Колева - Лесотехнически университет - София
4. Доц. д-р Веселина Господинова - Минногеоложки университет "Св. Ив. Рилски", (резервен)

Дисертационният труд включва 134 страници, 26 таблици, 50 фигури и 133 литературни източника (13 - на кирилица и 120 - на латиница).

Номерацията на фигурите, таблиците и цитираната литература, в рамките на автореферата, отговаря на тази в пълния дисертационен труд.

СЪДЪРЖАНИЕ

СЪДЪРЖАНИЕ.....	3
ИЗПОЛЗВАНИ СЪКРАЩЕНИЯ.....	5
ВЪВЕДЕНИЕ.....	6
Актуалност	6
Цел и задачи.....	7
ГЛАВА 1.....	8
Съвременно състояние на проблема за оценяване на тематичната точност на пространствени данни за земното покритие, получени от спътникови изображения.....	8
1.1. Въведение в проблема за оценка на тематичната точност	8
1.2. Верификация и валидация на продукти за земното покритие	8
1.3. Количествено оценяване на тематичната точност	9
1.4. Методи за качествено оценяване на тематичната точност на земното покритие	11
Изводи от Глава 1	11
ГЛАВА 2.....	12
Комбиниран качествено-количествен подход за оценяване на тематични продукти за земното покритие - методологични аспекти.	12
тематични продукти за земното покритие - методологични аспекти.	12
2.1. Същност на комбинирания качествено-количествен подход	12
2.2. Входни продукти и референтни данни.....	13
2.2.1. Слоеве с висока разделителна способност	13
2.2.2. Продукти от Локалния компонент	14
2.2.3. Референтни и спомагателни данни	14
2.3. Методи за оценяване на тематичната точност на продукти за земното покритие.....	15
2.3.1 Пространствена стратификация	15
2.3.2. Компютърно подпомогнато визуално дешифриране	16
2.3.3. Верификация на HRL чрез качествено оценяване	16
2.3.4. Вероятностно оценяване на бинарен слой с единичен клас земно покритие	17
2.3.5. Подходи към проблема за оценяване на HRL от непрекъснат тип	18
2.3.6. Оценяване на вероятността за приемане на продукти за земното покритие	20
2.3.7. Сравняване на тематични продукти на основата на тяхната точност	21
2.3.8. Вероятностно оценяване на слоеве на земното покритие, получени от VHR спътникови изображения.....	22
2.4. Методика за изследване и оценяване на тематичната точност на продукти за земното покритие, получено от спътникови изображения	23

Изводи към Глава 2	25
ГЛАВА 3.....	25
Резултати от прилагането на методиката за оценяване на продукти на земното покритие	25
3.1. Резултати от верификацията на HRL за горите TCD и DLT от 2012 и 2015 г.	25
3.1.1. Оценка на горските HRL 2012	25
3.1.2. Резултати от верификацията на горските HRL 2015.....	28
3.2. Резултати от оценяването с количествени методи на непрекъснати и бинарни слоеве с единичен клас земно покритие.....	33
3.2.1. Резултати количественото оценяване на продукта PWB2012	33
3.2.2. Изследване на слой от непрекъснат тип IMD2012 100m чрез обективно измерване на отговора.....	35
3.2.3. Резултати от анализа на точностите в бинарния слой на застрояването	36
3.2.4. Резултати от сравняване на HRL на основата на тематичната точност	40
3.3. Резултати от оценяването на продукти базирани върху VHR спътникови изображения - Локален компонент	42
3.3.1. Резултати от верификацията на Natura 2000 - тревни зони 2012	42
3.3.2. Резултати от верификацията на Атлас урбанизирани територии	45
Изводи към Глава 3	49
ЗАКЛЮЧЕНИЕ	49
ПРИНОСИ	49
Публикации на автора, свързани с дисертационния труд.....	50
Благодарности	50
Използвана литература	51

ИЗПОЛЗВАНИ СЪКРАЩЕНИЯ

БЛА	Безпилотни летателни апарати
ГИС	Географска информационна система
ДИ	Дистанционни изследвания
ЕАОС	Европейска агенция по околна среда
ИАОС	Изпълнителна агенция по околна среда
ИКИТ	Институт за космически изследвания и технологии
ФАО	Организация по прехрана и земеделие
НТП	Начин на трайно ползване
ПРС	Пространствена разделителна способност
СИЗП	Система за идентификация на земеделските парцели
CLC	CORINE Land Cover
CLMS	Copernicus Land Monitoring Service
ESA	European Space Agency
ETM+	Enhanced thematic mapper plus
FAO	Food and Agriculture Organization
FUA	Functional Urban Areas
GIS	Geographic Information System
GLC	Global Land Cover
GMES	Global Monitoring for Environment and Security
Eionet	European Environment Information and Observation Network
GSE	GMES Service Element
LCLU	Land cover and land use
LPIS	Land Parcel Identification System
MAES	Mapping and Assessment of Ecosystems and their Services
SP	Service Providers
SSL	Soil Sealing Layer
SSLE	Soil Sealing Layer Enhanced
TM	Thematic mapper
VHR	Very High Resolution

ВЪВЕДЕНИЕ

Земното покритие е основна променлива, която влияе върху и свързва много части от физическата и човешката среда. Промяната на земното покритие, например, се счита за единствената най-важна променлива на глобалната промяна, засягаща екологичните системи. Въпреки значението на земното покритие като променлива на околната среда, нашите познания за земното покритие и неговата динамика са непълни. Дистанционните методи са атрактивен източник на тематични пространствени данни като тези, изобразяващи земното покритие. От широкия спектър от съществуващи методи, информацията за земното покритие днес обикновено се извлича от дистанционно получени данни чрез класификация и често се представя под формата на тематични бази данни или карти. Поради това данните за земното покритие са централни за многобройни научни и практически приложения.

Актуалност

Качеството на масивите от пространствени данни е обширен въпрос, който може да се отнася до различни свойства, но често и тук обектът на интерес е точността на тематичната карта или на класификацията на тематични пространствени данни. За да бъдат тези данни или карти полезни, е необходима информация за тяхната точност. Оценката на тематичната точност на базите данни за земното покритие е важен елемент на качеството на тези данни и затова от години е обект на голямо внимание в изследванията.

Актуалността на представения за дисертационен труд се определя от няколко фактора. (1) Оценката на тематичната точност и други елементи на качеството на пространствените данни за земното покритие са от голямо значение за информираността на потребителите за свойствата на тези данни, респективно, за тяхното правилното използване; (2) Оценката на тематичната точност в днешно време е неразделна част от всяка задача или процедура по тематично картографиране; (3) Параметрите на качеството са неразделна част от метаданните на съвременните пространствени данни, като този факт е отразен и в съществуващите международни стандарти.

Цел и задачи

Цел на дисертацията е изследване на възможностите за реализиране на комбиниран подход за оценяване на точностни характеристики на продукти за земното покритие на базата на качествени и количествени методи.

За постигане на целта са поставени за изпълнение следните **задачи**:

1. Формулиране и обосноваване на качествено-количествен подход за оценяване на тематичното качество на продукти за земното покритие;
2. Проучване и избор на конкретни качествени и количествени методи за оценяване на тематичната точност.
3. Прилагане на методи за качествено оценяване на тематичната точност с коригиращи функции;
4. Оценяване на за тематичната точност на продукти по Програмата Коперник за България чрез количествени методи, ориентирани към различните видове входни продукти за земното покритие ;
5. Сравняване на продукти за земното покритие по тематична точност. Избор на критерии за сравняване.
6. Разработване на методика на базата на проучените и изпробвани методи за оценка на тематичната точност на продукти за земното покритие

ГЛАВА 1.

Съвременно състояние на проблема за оценяване на тематичната точност на пространствени данни за земното покритие, получени от спътникови изображения

1.1. Въведение в проблема за оценка на тематичната точност

Ние осъществяваме взаимодействието си с обектите от реалния свят по различни начини, като можем и да ги моделираме, прилагайки различни подходи. Например, чрез технологията на географските информационни системи (ГИС) (Longley et al., 1999) изследователите, а и други потребители, компилират слоеве от пространствени данни в „теми“, т.е., в информационни масиви, които имат обща характеристика или атрибут. Темата може да е от качествен тип (като в случая на класовете на земното покритие) или от количествен тип (например, температура на въздуха или количество на валежите).

С помощта на дистанционните изследвания (ДИ) част от биофизичните величини могат да бъдат измерени директно и температурата е един пример за променлива от тази група (Jensen, 2016). Втората основна група променливи са тези, които се получават чрез анализ на повече от една биофизична променлива. Към тях се отнася и земното покритие, което е един от най-важните продукти на технологията на ДИ. За извличане на географска информация за земното покритие, като основен метод се използва класифицирането на изображения (Tso and Mather, 2009).

Когато дадени решения се базират на картографска информация, за постигане на очакваните резултати е нужно да се знаят точностите на картите, респективно, на свързаните с тях пространствени данни. В противен случай, изпълнението на каквито и да е решения въз основа на тези продукти¹ може да доведе до неприемливи или непредвидими резултати (Congalton and Green, 2019).

Оценката на точността на тематичните пространствени данни, респективно, на картите², генерирани от тях, изисква оценка както на позиционната, така и на тематичната точност (Janssen, 1994; Congalton et al., 2007). Позиционната точност се отнася до определяне на местоположението на даден картографски обект (Bolstad, 2016). Тематичната точност се занимава с атрибутите (етикетите) и измерва, дали картографираните етикети са различни от истинския или референтния етикет.

Пространствените данни представят различни аспекти от реалния свят и тяхното качество може да се разглежда като степен на различие между тази представа и реалността (Devillers, 2006). Според стандарта ISO 19157:2013 "Geographic information — Data quality", позиционната и тематичната точност са два от шестте елемента на качеството на данните (Yang, 2013). Там тематичната точност се определя като точност на количествените атрибути и правилност на неколичествените атрибути и на класификациите на обектите и техните взаимоотношения.

1.2. Верификация и валидация на продукти за земното покритие

¹ Под продукт отгук нататък ще разбираме набор или масив от пространствени данни.

² Както в цитираната литература, така и в настоящия текстов документ, иначе имащият строго и точно значение в картографията термин "тематична карта" или само "карта", се използва в по-широк смисъл - като синоним на тематичен слой, тематични пространствени данни, или тематичен продукт.

Верификация и валидация³ в много случаи неправилно се считат за еднакви и се употребяват като синоними (McCaffrey, 2006). Разлика между тях може да бъде описана чрез следните въпроси (Pham, 1999): Верификация: „Правилно ли изграждаме продукта?“ Валидация: „Изграждаме ли правилния продукт?“

При създаването на тематични продукти за земното покритие, верификацията се прилага с коригираща цел, за подобряване на крайния резултат. Валидацията няма коригираща цел, а задачата е, крайният продукт да бъде независимо оценен, одобрен или отхвърлен, както и да се създадат неговите метаданни.

1.3. Количествено оценяване на тематичната точност

В една от базовите работи по въпроса за тематичната точност на земното покритие (Stehman and Czaplewski, 1998), са представени трите основни компонента на оценяването: (1) модел за вземането проби, (2) модел за регистрирането на отговора и (3) протокол за оценка и анализ.

Прилагането на вероятностен модел за вземането на проби допринася за научно обоснована оценка на точността (Stehman and Czaplewski, 1998). Вероятностното вземане на проби се дефинира посредством ненулеви вероятности за включване на пробните единици в извадката, От многото на брой модели за вероятностно вземане на проби най-често използваните са прост случаен подбор, стратифициран случаен подбор, систематичен и клъстерен (гнездови) случаен подбор (Калинов, 2002; Congalton, 1988).

Стратификацията е разделяне на изследваната територия на неприпокриващи се части (единици), при което всяка единица за оценка (assessment unit) се приписва към една страта. Обикновено обемът на извадката се оценява на базата на формулата за биномния тест, която може да се приложи към общата точност за проста случайна извадка или към точността на потребителя за стратифицирана случайна извадка (Foody, 2009a; Rossiter, 2014) - формула 1:

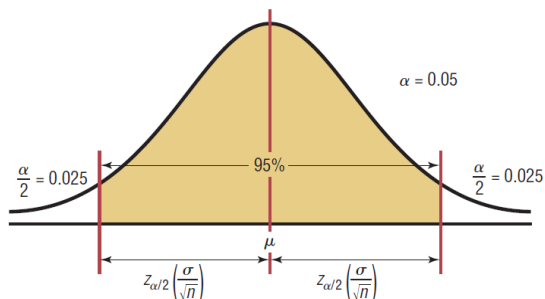
$$n = \frac{z_{\alpha/2}^2 P(1-P)}{d^2}, \quad (1)$$

където n е обемът на извадката, $z_{\alpha/2}$ е критичната стойност на нормалното разпределение за двустранно ниво на значимост α , P е очакваната обща точност на популацията, а d е очакваната стандартна грешка за измерване на P , равнявайки се на половината от желания доверителен интервал $\pm d$ (фиг. 1.2). Известно е, че при ниво на значимост $\alpha = 0.05$ критичната стойност $z_{\alpha/2}$ е равна на 1.96 като тази стойност съответства на доверителна вероятност 95 (Bluman, 2009). Нагледно връзката между n и d е показана на Фиг. 1.3 (CORINE, 2000).

³ Терминът "верификация" е утвърден в българския език и намира широко приложение.

Верификация - установяване, удостоверяване на истината; свидетелство

(<http://www.onlinerechnik.com/duma/верификация>). Относно превода на термина "validation", на български вероятно е по-правилно е да се казва "валидиране", но ще използваме "валидация" за съгласуваност и съзвучие с "верификация".



Фиг. 1.2. Доверителен интервал и ниво на значимост при нормалното разпределение

Моделът за регистриране на отговора (response design) изяснява въпроса за съответствието между класифицираните (прогнозните) и референтните данни (еталона) и обхваща всички стъпки от установения за целта протокол (Olofsson, 2014). Липсва консенсус по въпроса, коя единица за вземане на проби (пиксел, полигон или блок) е най-добра (Stehman and Czaplewski, 1998).

При анализа на събраните данни централна роля играе матрицата на грешките заедно с мерките на точността, получени от нея, а именно общата точност, точностите на потребителя и производителя (Olofsson, 2013). В съвременен изпълнение матрицата на грешките, вместо просто брой проби n_{ij} , включва данни, необходими за оценка на относителните дялове⁴ на площите на всеки клас, въз основа на референтната класификация (Таблица 1.2).

Таблица 1.2. Матрица на грешките с относителни дялове на площите на класовете

		Reference class						Total
		1	2	...	k	...	r	
Map class	1	p_{11}	p_{12}	...	p_{1k}	...	p_{1r}	p_{1+}
	2	p_{21}	p_{22}	...	p_{2k}	...	p_{2r}	p_{2+}
	:	:	:		:		:	:
	k	p_{k1}	p_{k2}	...	p_{kk}	...	p_{kr}	p_{k+}
	:	:	:		:		:	:
	r	p_{r1}	p_{r2}	...	p_{rk}	...	p_{rr}	p_{r+}
	Total	p_{+1}	p_{+2}	...	p_{+k}	...	p_{+r}	1

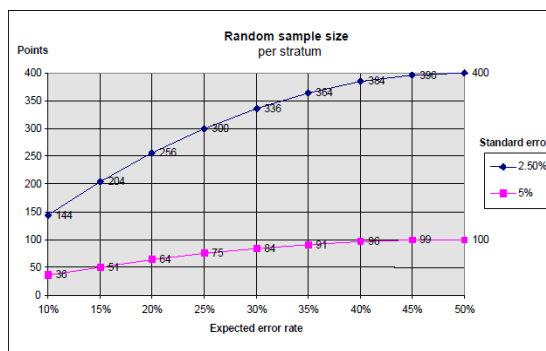
При решаване на приложни задачи се работи с данните от извадката и вероятностите от матрицата се заместват с оценки от извадката \hat{p}_{ij} на относителните дялове p_{ij} на площите, като приемаме, че стратите съвпадат с класифицираните класове:

$$\hat{p}_{ij} = W_i \frac{n_{ij}}{n_{i+}}, \quad (5)$$

където W_i е относителният площен дял на класифицирания клас i .

Често срещан проблем, както за потребителите, така и за производителите на тематични продукти, е сравняването им за установяване, кои имат по-висока точност

⁴ В статистиката, като синоними на термина "относителен дял", се срещат също "относителна честота" и "пропорция".



Фиг. 1.3. Обем на извадката (sample size) n в зависимост от очакваната стандартна грешка d .

(Foody, 2009a) или кой алгоритъм за класификация е по-добър (Stehman, 1999b). Проблемът е, че това се извършва чрез директно сравнение на извлечените оценки за точността, а не по статистически строг начин (Foody, 2004). Това може да се постигне чрез стандартен z-тест за статистическа значимост на разликата (Stehman and Foody, 2019; Foody, 2009b).

В проучванията с дистанционни изследвания често се използва един и същ референтен набор от данни за дадена целева територия при оценката на различни класификация. Такива свързани или „сдвоени“ проби не могат да се считат независими и трябва да се използва алтернативен метод, като теста на McNemar (Fleiss et al., 2003; Foody, 2004).

1.4. Методи за качествено оценяване на тематичната точност на земното покритие

Съществуват два основни типа оценяване на картата: качествени и количествени (Congalton and Green, 2019). В контекста на проблема за валидиране на глобални продукти за земното покритие (Morissette et al., 2004), подходите за оценяване също са разделени на два типа: мерки за изграждане на доверие и статистически подходи.

J. Morissette et al. (Morissette et al., 2004), обобщавайки описанието на методологичните елементи за валидиране на карти на земното покритие, а именно (1) мерките за изграждане на доверие, (2) извода, базиран на модел и (3) извода, базиран на дизайн, заключават, че правилният план за валидиране изисква наличието на всичките три елемента.

Метод за визуално-качествено оценяване

Масивите от данни за земно покритие HRL 2012, 2015 и 2018 са верифицирани с методи за качествено оценяване и с използването на подходящи референтни данни и по насоки, изготвени под ръководството на ЕАОС (ЕЕА, 2018). Използвана е процедура за качествено оценяване при верификацията, състояща се от две части: общ преглед на качеството на HRL данните и оценка по метода за визуално-качествено оценяване Look-and-feel. С прилагането на процедурата за качествено оценяване се цели откриването и анализирането два типа грешки от класификацията - пропуск и погрешна класификация - с цел по-нататъшна корекция и подобряване на продуктите.

Изводи от Глава 1

Оценяването на тематичната точност на продуктите за земното покритие е задължителен етап от тяхното създаване. Тематичната точност е неразделен елемент на качеството на пространствени данни, който факт е отразен и в международните стандарти. В резултат на оценяването се попълват метаданни и по този начин се информират потребителите на продуктите.

Верификация и валидация са две приличащи си дейности, но първата се извършва с коригираща цел, а втората - за определяне на точностите с цел приемане (одобряване) на готовия продукт. При верификацията на тематични продукти е важно прилагането на качествен подход. Той генерира полезна информация за разбирането на основната структура на грешките и да има коригираща роля в създаването на тематичните продукти.

Методът за визуално-качествено оценяване, състоящ се от общ преглед на качеството и оценяване Look-and-feel е подходящ за целите на верификацията с

коригиращи функции. Той позволява да се оцени стабилността на работата на класификатора в рамките на отделни тематични класове от обекти и е щадящ откъм ресурсно натоварване.

Количественото (статистическо) оценяване на тематичната точност е полезно, както за етапа на верификация, така и за валидацията на данните за земното покритие. При него се прилага вероятностен модел на вземане на проби, генерират се неотместени оценки и по извадката могат да се правят изводи за целия продукт. При избора на параметрите на модела на извадката е възможно да се наложи съобразяване с наличните ресурси за изпълнение на оценяването.

Сравняването на тематични данни на основата на тяхната тематична точност трябва да става с отчитане на статистическата значимост на различията.

Правилният подход за оценяване на продукти на земното покритие следва да интегрира три типа методи, последните два - статистически: за качествено оценяване, за извод, базиран на модел и за извод, базиран на дизайн.

ГЛАВА 2.

Комбиниран качествено-количествен подход за оценяване на тематични продукти за земното покритие - методологични аспекти.

тематични продукти за земното покритие - методологични аспекти.

2.1. Същност на комбинирания качествено-количествен подход

От изводите към Глава 1 става ясно, че е нужно е обединителният подход да включва качествен и количествен компонент, като количествената част се състои от две групи статистически методи - за извод, базиран на модел и извод, базиран на дизайн. Изводът, базиран на дизайн, се опира на статистически принципи, по които се събират проби и се извеждат характеристики на цялата популация. Изводът, базиран на дизайн, генерира неотместена оценка на точността на тематичната карта, на базата на последователни оценители и се реализира практически чрез някой от вероятностните модели за събиране на проби.

При извода, базиран на модел, вниманието е насочено към самата класификация като процес, а не към самия тематичен продукт. Последният се разглежда като реализация, съпътстващ продукт, на процеса на класификация, който е обект на грешка, а точността на тематичния продукт се характеризира с оценки на грешките в процеса на класификация, който го е създал. Така, изводът, базиран на модел, може да дава на потребителите количествена оценка за решението на класификатора за всеки отделен пиксел от входното спътниково изображение (някакъв показател за качеството на класификацията, например ниво на увереност).

Методи от групата на извода, базиран на модел, не се разглеждат в настоящата работа, защото, както стана ясно, този подход се реализира в процеса на самата класификация, ако е предвиден. От друга страна, в състава на изследваните от нас продукти отсъстват такива данни, както и сведения за създаването на такива в рамките на CLMS.

За реализирането на качествения подход от полза е информацията относно проучените няколко метода за качествено оценяване на тематичната точност: мерки за изграждане на доверие, систематичен контрол на качеството, качествени изследвания и метод за визуално-качествено оценяване. Тези методи се прилагат за получаване на евтини и бързи качествени оценки и друга информация на етапа на верификацията, като имат коригиращи функции. Методът за визуално-качествено оценяване, състоящ се от (1) общ преглед на качеството и (2) оценяване Look-and-feel, позволява ефективно дефиниране на тематичните и пространствени области за търсене и определяне на грешки на класификацията от системен характер и е широко използван в проекти на CLMS. Поради тези и други характеристики, този метод е включен и приложен като основен за реализиране на качествените оценки.

2.2. Входни продукти и референтни данни

Представянето на входните продукти е важно, тъй като техните свойства предпоставят изискванията към методите, с които да бъдат оценявани. От произведените по услугата CLMS няколко групи масиви от данни за земно покритие и земеползване (LCLU - Land cover and land use) (CLMS, 2022), към настоящото изследване отношение имат част от Слоевете с висока разделителна способност и Локалния компонент.

2.2.1. Слоеве с висока разделителна способност

Слоевете с висока разделителна способност (HRL) са растерни продукти, предоставящи информация за специфични характеристики на земното покритие, целящи допълване и гарантиране на непрекъснатостта на времевите серии на CORINE Land Cover. В таблица 2.1 съвместно са представени специфични свойства на продуктите на състоянието (status layers) и допълнителните продукти на HRL2012 (GIO, 2016).

Таблица 2.1. Слоеве с висока разделителна способност 2012 – специфични свойства

Продукти		Съдържание на HRL	Тип на слоя
Степен на непроницаемост/ Degree of Imperviousness (20m и 100m)		Степен на непроницаемост 0–100%	Плътност (непрекъснат)
Склопеност/ Tree Cover Density (TCD)		Склопеност (плътност на дървесното покритие), стойности от 0 – 100% .	Плътност (непрекъснат)
Вид гора/ Forest Type (FTY)	Преобладаващ дървесен тип/ Dominant Leaf Type (DLT)	Широколистни/Иглолистни/Не-гора ММУ от 0,5 ha и приложен 10% праг за плътността на дървесното покритие.	Номинален (категориен)
	Допълнителен слой/ Additional Support layer	Дървета в селскостопански зони (CLC 222) и в населени места (CLC 141)	Бинарен
Пасища и ливади/ Permanent Grassland		Пасища и ливади	Бинарен
Допълнителен слой към слой пасища и ливади		Тревни площи в населени места (CLC 141), летища (CLC 124), места за спорт и отдых (CLC 142)	Бинарен
Влажни зони/ Wetlands		Влажна зона/Не-влажна зона	Бинарен
Постоянни водни тела/ Permanent Water bodies		Вода/Не-вода. Постоянни водни тела, включително малки водни тела	Бинарен

Чрез продукта "Степен на непроницаемост" е картографирана степента на непроницаемост (почвеното запечатване) заемаща целочислени стойности 1-100 % (Фиг. 2.1) (Langanke et al., 2016). Под застроени площи (built-up area), по дефиниция, се разбира

район, запълнен с къщи, магазини, офиси и други сгради, с много малко открито пространство⁵.



Фиг. 2.1. а) Слойт Степен на непроницаемост (IMD) в покрайнините на гр. Ген. Тошево, наложен върху ортофото.

Фиг. 2.1. б) IMD в покрайнините на гр. Ген. Тошево, върху ортофото, в полупрозрачен режим на визуализация.

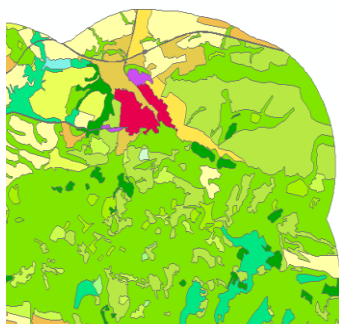
2.2.2. Продукти от Локалния компонент

Продуктите за земното покритие от Локалния компонент не осигуряват пълно покритие на територията на страната, но предоставят подробна информация за LCLU за конкретни области на интерес, така наречените „горещи точки“, с референтна година 2012. Създадени са три основни продукта за земното покритие:

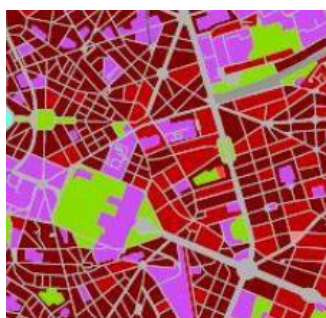
Атлас урбанизирани територии 2012 (Urban Atlas, UA),
Натура 2000 - тревни зони 2012 (Natura 2000 – grasslands, N2K)
Крайречни зони (Riparian Zones).

В настоящата работа се оценяват първите два продукта - UA и N2K.

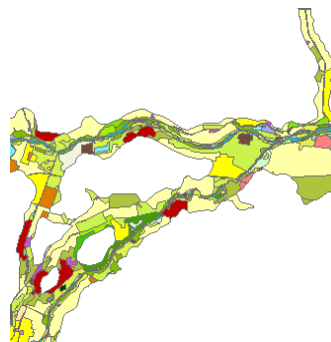
Документацията за продуктите е достъпна в техническата библиотека на CLMS⁶.



а)



б)



с)

Фиг. 2.2. Примерни фрагменти от трите локални компонента: а) Natura 2000; б) Urban Atlas; с) Riparian Zones.

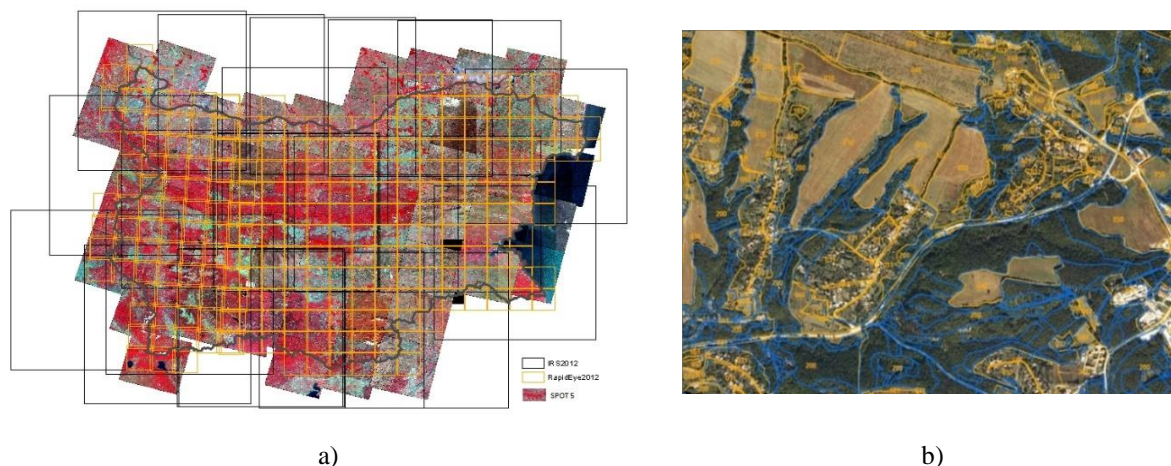
2.2.3. Референтни и спомагателни данни

⁵ Glossary: List of environmental terms used by EEA, https://www.eea.europa.eu/help/glossary#c4=10&c0=all&b_start=0&c2=built-up

⁶ Technical library <https://land.copernicus.eu/user-corner/technical-library>

При оценяването на HRL, като основни референтни и *in situ* (местни, локални) данни е използвана цифровата ортофото карта от годините 2011 и 2013-2016 г. VHR сателитни изображения също са използвани за проследяване на промените във времето (Фиг. 2.3). Използвани са и *in situ* векторни бази данни, както и онлайн ресурси.

Всички спътникови изображения са получени във времевия интервал 2011-2012 г. Основните им характеристики са представени в Таблица 2.3.



Фиг. 2.3. Примери за референтни данни: а) Покритие на страната със спътникови изображения IMAGE2012; б) Ортофото с *in situ* тематични слоеве върху него.

Таблица 2.3. Свойства на наборите от данни от спътникови изображения IMAGE2012

Характеристика/Спътник	IRS LISS III (coverage-1)	RapidEye (coverage-2)	IRS AWiFS	SPOT-5 pan-sharpened
Брой доставени канали	4	5	4	3
Спектрални канали	green, red, NIR, SWIR	blue, green, red, red-edge, NIR	green, red, NIR, SWIR	green, red, NIR,
Размер на терения елемент [m]	23.5	6.5	56-70	2.5 or 10
Дълбочина в битове	7	up to 12	10	8
ПРС на доставените данни	20 m	20 m	60 m	2.5

Наборите от референтни от данни за оценяването на Локалния компонент се състоят от две основни части: данни, осигурявани централизирано и *in situ*. Комплектите от данни от първата група, много от които са свободно достъпни, са следните: изображения от Bing maps; мозайка GioLand/VeryHighResolution 2012; GoogleEarth и StreetView; OpenStreetMap; изображения World Imagery на ArcGIS online.

Първите четири източника бяха достъпни чрез веб-базирания инструмент LACO-Wiki⁷, докато последните четири бяха използвани в средата на настолни ГИС програми.

2.3. Методи за оценяване на тематичната точност на продукти за земното покритие

2.3.1 Пространствена стратификация

Двете основни цели за прилагане на стратификация са получаване на оценки по определен териториален принцип или за подобряване на прецизността на оценките. В настоящото изследване, при оценяване на грешките за пропуските при тематични класове с относително малки площи, стратификацията се използва и за повишаване на

⁷ Laco-Wiki, <https://laco-wiki.net/en/Welcome>

ефективността на работата чрез изключване на територии с малка вероятност на присъствие на проверявания клас.

2.3.2. Компютърно подпомогнато визуално дешифриране

Под дешифриране на изображения се разбира разкриване и разпознаване на обектите от земната повърхност, както и установяване на качествената и количествената им характеристика (Желязков и Тепелиев, 1992). Една от области на приложение на тематичното дешифриране е горското стопанство. При верификацията на горските продукти от HRL2012 и HRL2015, трябва само да разпознаем и определим територията на разпространение на гората и вида гора (иглолистна или широколистна).

При работа с продукта Степен на непроницаемост главно се дешифрират елементи на урбанизираните територии. За верификацията на продукти за земното покритие от Локалния компонент важна роля имат класификационните им схеми, представени в техническите документации.

2.3.3. Верификация на HRL чрез качествено оценяване

От описаните в Глава 1 подходи и методи за качествено оценяване на продукти на земното покритие, за целите на разглеждания тук комбиниран подход е използван двустъпков качествен метод, разработени по задание на ЕАОС (ЕЕА, 2012) и приложен в кампании по верификация на HRL2012 и HRL2015 в 39 страни. Насоките - документ, описващ метода - в определена степен, имат препоръчителен характер и всяка страна-участничка е свободна да модифицира и подобрява тези методологии в зависимост от своите нужди, налични национални референтни данни и експертен опит (Congedo et al., 2016). Верификацията с качествено оценяване включва следните две задължителни стъпки (ЕЕА, 2012):

- общ преглед на качеството на HRL-данните;
- верификация на HRL по метода Look-and-feel.

Общият преглед на качеството има за цел да формира у експерта общо впечатление за качеството на данните и да породи разбиране, къде да се съсредоточат следващите, по-подробни действия.

Верификацията при Look-and-feel се дефинира като проверка, насочена към идентифициране на систематични грешки в класификацията, които могат да бъдат подобрени при бъдещите актуализации на продуктите. В резултат, на пробния полигон се присвоява оценка по петобална скала (Таблица 2.4). Трябва да се проверят достатъчен брой проби за всяка страта, за да може да ѝ се даде оценка по същата петобална скала.

Таблица 2.4. Степени за качествена оценка на HRL.

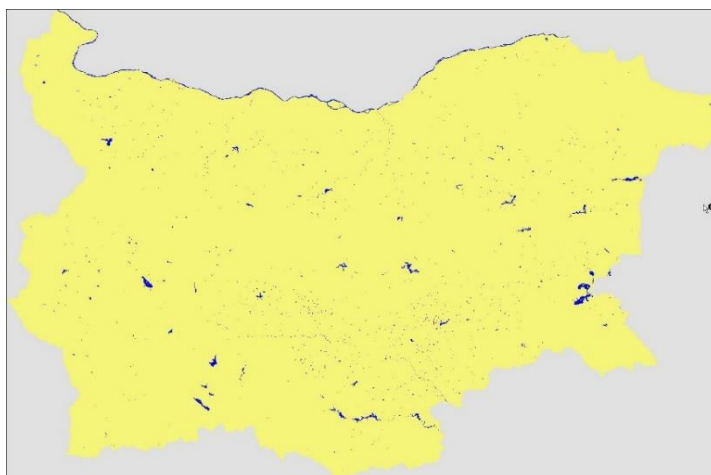
Оценка/Grade	Описание
отлична/excellent	При класификацията на HRL е постигната точност близка до 100%; не са открити грешки във верифицираните области.
добра/good	Интерпретаторът е сигурен, че точността в този HRL е най-малко 85 %; открити са само спорадични грешки във верифицираните области.
приемлива/acceptable	Интерпретаторът преценява, че точността на този HRL в повечето от верифицирани области е 85 %; откриват се само някои малки грешки.
неудовлетворителна/insufficient	Интерпретаторът не очаква точността на този HRL да достигне минимално допустимата точност от 85 %; срещнати са грешки в различни райони.

много лоша/*very poor*

Интерпретаторът е сигурен, че точността в HRL е лоша и много под 85%, мнозинството от верифицираните области са лошо картографирани.

2.3.4. Вероятностно оценяване на бинарен слой с единичен клас земно покритие

Настоящия раздел разглежда метод за количествена оценка на тематичната точност, с цел верификация, на един отделно взет клас земно покритие. Оценяваме слоя с висока разделителна способност за постоянните водни тела (СВР-ПВТ) - Permanent Water Bodies (PWB) 2012, но постановката е приложима за всички бинарни HRL с единичен клас. Стойностите в растера са, както следва: 1- при наличие на постоянно водно тяло; 0 – за всички други типове земно покритие. СВР-ПВТ има ПРС 20x20 m, също като на спътниковите изображения, от които е получен (Фиг. 2.6).



Фиг. 2.6. Слой с висока разделителна способност „Постоянни водни тела” – общ вид

Стратификация при оценяването на PWB

Водните тела покриват сравнително малка част от територията на страната – площта на класифицирания входен СВР-ПВТ е 0.81% от територията България. Тъй като оценяваме единичен клас - СВР-ПВТ, прилагаме на стратификация с две страти: едната за грешки „включване” (commission), а другата - за грешки „изключване” (omission) по отношение на въпросния клас. Първата страта се формира от множеството (маската) на пикселите, класифицирани като постоянно водно тяло. Втората страта обхваща определени територии извън маската.

Грешки на включване и изключване

За изчисляване на оценките на грешките, работим с отношения на площи. Истинската площ на оценявания СВР, е неизвестна, като разполагаме само с нейната оценка – сумарната площ на пикселите със стойност 1. Затова, практически, грешката „включване”, подобно като в (ЕЕА, 2012), се изчислява по формулата:

$$E_{comm} = \frac{A_{comm(PWB)}}{A_{(PWB)}} \approx \frac{N_{comm}}{N}, \quad (8)$$

където,

E_{comm} – грешка „включване” за класа СВР-ПВТ; $A_{(PWB)}$ – площ, на (класифицирана като) СВР-ПВТ; $A_{comm(PWB)}$ – площ на грешката „включване” за класа

СВР-ПВТ; N_{comm} – брой измервания (пиксели) с грешка „включване“; N – общ брой измервания, случайно разпределени в площта на класифицирания СВР.

Грешката „изключване“ се изчислява чрез грешката „включване“ за класа, обратен на СВР-ПВТ, т.е., незаетата с водни тела територия. Означението „non-PWB“ във формулите се заменя със „SFO“ (Strata For Omissions):

$$E_{omis(PWB)} = E_{comm(SFO)} \frac{A_{(SFO)}}{A_{(PWB)}}, \quad (10)$$

където,

$E_{omis(PWB)}$ – грешката на „изключване“ за СВР-ПВТ; $E_{comm(SFO)}$ – грешка „включване“ за класа non-PWB на базата на стратата SFO; $A_{(SFO)}$ – площ на стратата за грешките „изключване“ по отношение на класа СВР-ПВТ; $A_{(PWB)}$ – площ, на (класифицирана като) СВР-ПВТ.

Оценките на грешките „включване“ и „изключване“ и изчисляваните на тяхна основа точности са най-информативните параметри за характеризиране на тематичната точност на единичен клас с цел верификация. Заедно с тях се изчислява и стойността на съответната несигурност на оценяването. За нашите цели едно прагматично решение е броят на измерванията да бъде 250, а с допускане и на 30 невалидни проби - 280.

2.3.5. Подходи към проблема за оценяване на HRL от непрекъснат тип

Изследването на свойствата на HRL от непрекъснат тип за степента на непроницаемост IMD2012 100m (GIO, 2016) преследва следните няколко цели: (1) Търсене на отговор на въпроса за разминаването между високите оценки на доставчиците и резултатите от нашите и на други изследователи наблюдения за IMD продуктите, например, (Lajčáková, 2015); (2) Сравняване с резултатите от валидацията на IMD2012 100m от страна Европейския екип; (3) Практическо изследване на свойствата на IMD чрез прилагане на адекватни количествени методи и инструменти.

С цел вероятностно оценяване на тематичната точност, от многоградационния IMD слой се получава бинарен слой - т. нар., слой-карта на застрояването (built-up map) с два класа - "Плътно застрояване" (Built-up) и "Всички останали територии" (Non-built-up), на основата на следното правило за прилагане на праг (Таблица 2.6):

Таблица 2.6. Правила за получаване на стойностите в слоя Built-up map

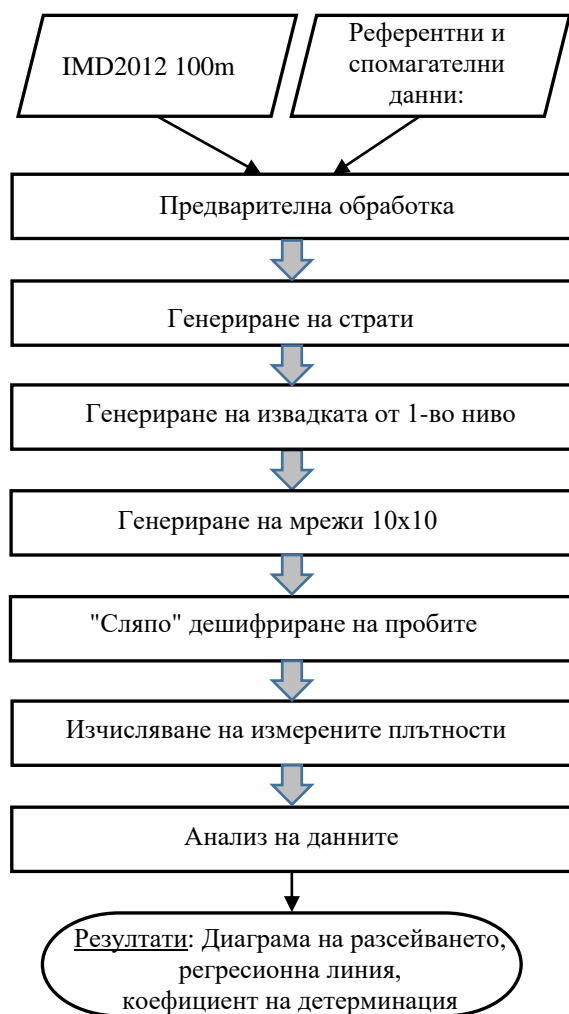
Клас	Стойност	Правило за стойностите на IMD
Плътно застрояване (Built-up)	1	Степен на непроницаемост в диапазона 30%-100%
Всички останали територии (Non-built-up)	0	Степен на непроницаемост <30%

Цялостната оценка на продукта IMD2012 100m се постига чрез изследвания и тестове в две направления: а) на самия непрекъснат слой, базирайки се на идеи, изложени в (Maucha et. al, 2011); б) на бинарния слой-карта на застрояването Built-up map.

Оценяване на непрекъснат слой

Един вариант за обективно определяне на референтните стойности на степента на непроницаемост е чрез измерването им с налагане на равномерна решетка от 10x10 точки върху всеки оценяван 100 m пиксел. За реализирането на този метод, е разработена и изпълнена алгоритмична схема в универсален софтуер за ГИС. На Фиг. 2.7 е представена

съответната ѝ блок-схема на алгоритъма за оценяване на параметри на качеството на IMD2012 100m чрез обективно регистриране на отговора.



Фиг. 2.7. Блок-схема на процеса на оценяване с обективно измерване

В резултат от анализа на гореспоменатия първи пакет от данни - измерени референтните стойности от извадката и стойности от непрекъснатия слой IMD, за последния са получават следните информационни показатели:

- Корелационно поле (диаграма на разсейването);
- Параметри на регресионната права (линеен регресионен модел);
- Коефициент на детерминация R^2 .

Цялата тази информация е ценна и от гледна точка на използването ѝ за калибрирането на продуктите от тип плътност, както е IMD.

Оценяване на бинарния слой на застрояването

Във връзка с изследване на свойствата на бинарния слой на застрояването (Built-up map) по отношение на тематичната точност, той се подлага на следните видове оценяване:

- оценяване на тематичната точност на единичен клас built-up в две страти - по една за двата типа грешки - commission и omission, подобно на случая с PWB, както и с интервална оценка;

- оценяване чрез стратифицирана случайна извадка с три страти, две от които за оценяване на грешките от тип пропуск - с използване на матрицата на грешките;
- оценяване на вероятността за приемане на продукта.

При определяне на точността на единичен клас се прилага методичната схема, като при постоянните водни тела РВВ. Двете страти тук са, както следва:

- Страта1, страта за грешки погрешна класификация - включва териториите със стойности на степента на непроницаемост 30% – 100%, т.е., площите с плътно застрояване, според класификатора;
- Страта2, страта за грешките от тип пропуск - обхваща териториите със стойности 1% - 29%.

По Интервалната оценка на даден параметър на една популация е по-точна от точковата (Христов, 1964; Bluman, 2009). Интервалната оценка за грешката в HRL като относителен дял по случайна извадка от n на брой пиксели, от които p' са с грешка, е представена чрез формулата:

$$p' - z_{\alpha/2} \sqrt{\frac{p'q'}{n}} < p < p' + z_{\alpha/2} \sqrt{\frac{p'q'}{n}}, \quad (13)$$

където: p – действителна грешка (в популацията), n – обем на извадката, p' – относителен дял на точките с грешка от погрешна класификация в извадката, $q' = (1 - p')$ – относителен дял на благоприятните случаи, α – ниво на значимост, $z_{\alpha/2}$ – стандартна стойност, съответстваща на $\alpha/2$.

От гледна точка на компромис между ресурса време и надеждността на оценката, са определени обеми на извадките от 250 и 200 случайни точки за Страта1 и Страта2 съответно.

Следващ етап в изследването на built-up map е прилагането на анализ на тематичната точност със стратифицирана случайна извадка в три страти. Планирано и проведено оценяване на грешки тип пропуск в още една, допълнителна страта (Страта3), базирана на териториите със степен на непроницаемост 0%. За анализ на получените от извадките резултати за бинарния слой-карта от трите страти, е използван свободно-достъпният в интернет онлайн инструмент Map Accuracy Tools (Salk et al., 2018).

2.3.6. *Оценяване на вероятността за приемане на продукти за земното покритие*

Вероятност за приемане на продукта

Имайки пред вид вече споменатата преди изисквана точност 85%, решението за приемане на един бинарен продукт би било оправдано, ако сме сигурни, че действителната му точност е над 85% (т.е., действителната грешка е под 15%). И обратно, за да отхвърлим продукта, трябва да сме сигурни, че действителната грешка е по-голяма или равна на 15%.

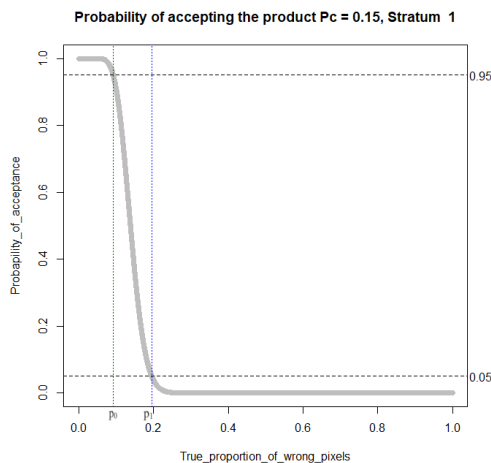
Ако допустимата действителна грешка в слоя обозначим с p , вероятността в извадка от n пиксела да присъстват k грешни, има вида, като във формула 14. Вероятността действителната грешка в слоя да е по-малка от p при установени c сгрешени пиксела, се представя чрез формула 15 (Aronoff, 1982b; Maucha and Büttner, 2008). Величината $L(p,n,c)$ е вероятността за приемане (одобряване) на продукта при предварително зададен праг p за действителната грешка. В случай на тематичен слой с

ниско качество, който е с недопустимо голяма грешка, същата величина L представлява вероятността за грешка от първи род или риск за потребителя.

$$P_k \approx C_n^k \cdot p^k \cdot (1-p)^{n-k} \quad (14)$$

$$L(p, n, c) = \sum_{k=0}^c C_n^k p^k (1-p)^{n-k} \quad (15)$$

На базата на аналитичния модел от формула 15 в (Maucha and Büttner, 2008) е предложен компютърен модел във вид на скрипт за софтуерната среда R⁸, с цел практическо приложение при валидацията на слой земно покритие с единичен клас, чрез изчисляване на вероятността за приемане на продукта. Скриптът бе модифициран с цел обработка на нашите данни и графично извеждане на част от резултатите. На фиг. 2.11 е показан примерен вариант на графичното представяне на вероятността за приемане. Относителният дял на сгрешените елементи в извадката $P_c = c/n$ е постоянно число, фиксирано на стойност 0.15. Обозначени са стойностите $\alpha_0 = \alpha_1 = 0.05$, както и получаваните като резултат от модела граници p_0 и p_1 .



Фиг. 2.11. Вероятност за приемане на продукта при $P_c = 0.15$

2.3.7. Сравняване на тематични продукти на основата на тяхната точност

Твърде ниските резултати за точността на слоя земно покритие IMD2012 100m, измерена върху бинарния продукт за застрояването (Димитров, 2016b), предизвикват нашия интерес към тяхното проверяване и потвърждаване, чрез сравняване с продукт с доказани точностни характеристики. Като такъв е възприет коригирания през 2015 г., заедно с останалите HRL 2012, слой IMD2012 100m. Резултатите (сравнително високи) от неговата валидация са публикувани през януари 2017 (HRL, 2017).

В случая на независими извадки, реализирани поотделно и независимо в двата сравнявани продукта, значимостта на разликата между двата относителни дяла се оценява чрез статистически z-тест по следния начин (Bluman, 2009; Foody, 2004):

$$z = \frac{\frac{x_1}{n_1} - \frac{x_2}{n_2}}{\sqrt{p(1-p)\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}}, \quad (16)$$

⁸ R Core Team (2021). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL <https://www.R-project.org/>

където x_1 и x_2 са брой на правилно определените случаи в две независими извадки с обеми n_1 и n_2 , съответно $p = (x_1 + x_2)/(n_1 + n_2)$.

Нулевата хипотеза (H_0) за липса на значима разлика, ще бъде отхвърлена при широко използваното ниво на значимост от 5%, ако $z > 1.96$.

За бинарния слой-карта на застрояването ние разполагаме с извадки в три страти: Страта1, Страта2 и Страта3, описани по-горе, с общо брой 550 наши проби. Извличаме референтните данни за територията на България от отчета на ЕАОС за валидацията на коригирания IMD2012 100m, при която пробите също се вземат от зоните на застрояването (с прилагане на 30% праг).

Вторият вариант за сравняване по точност е този със зависими (сдвоени от типа "преди и след") извадки. Ползвайки един и същ набор от 250 пробни пиксели на Страта1 реализираме зависими извадки в двата различни (сравнявани) растера Built-up - преди и след корекцията. Статистическата значимост на разликата между два относителни дяла оценяваме с теста на McNemar с матрица 2x2 при проучване от типа „преди и след“ (Fleiss et al., 2003, Foody, 2004):

$$\chi^2 = \frac{(f_{12} - f_{21})^2}{f_{12} + f_{21}} \quad (17)$$

където f_{ij} $i=1,2$; $j=1,2$ - брой на елементите във от главния диагонал и отразяват различаващите се стойности „преди и след“.

Както за формула 16, така и за формула17 можем да приложим и корекция за непрекъснатост (Fleiss et al., 2003).

2.3.8. Вероятностно оценяване на слоеве на земното покритие, получени от VHR спътникови изображения

При използване на аерофото и VHR спътникови изображения за извличане на информация за земното покритие, в редица случаи, за пробна единица се използва полигон, вместо точка или пиксел (Ye et al., 2018). При по-големите полигони спрямо MMU, се прилага пробонабиране на следващо ниво - вътре в полигона (Lea and Curtis, 2010).

Всичките гореспоменати съображения намират отражение в методологията, която ние прилагаме при верификацията на продукти от Локалния компонент - пространствени данни за земното покритие от тип полигон, получени от VHR изображения (Maucha et al., 2017). За реализиране на преобладаващата част от задачите по верификацията е използван онлайн ресурсът LACO-Wiki⁹. Приложена е комбинирана количествена и качествена процедура за получаване на информация за качеството на тези продукти с цел подобряването им.

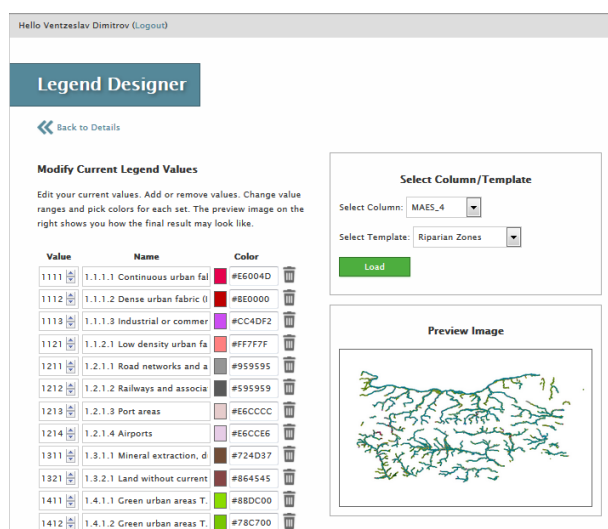
Процесът на верификация включва следните работни стъпки:

- Подготовка на входни и референтни данни;
- Генериране на извадки от LCLU полигони;
- Визуална интерпретация на пробите, с възможност за предоставяне на коментари и обратна връзка;
- Оценка на резултатите.

⁹ <https://laco-wiki.net/en/Welcome>

Данните за даден слой от Локалния компоненти, подлежащ на верификация, се зареждат в интернет-базирания инструмент LACO-Wiki. Следва създаване на легенда, съгласно класификационната схема на продукта (Фиг. 2.12) и генерирането на стратифицирана случайна извадка, на базата на класифицираните класове, с наличните в LACO-Wiki средства.

Референтните данни се организират в рамките на локални ГИС работни места.



Фиг. 2.12. Редактиране на легендата в диалогов режим в LACO-Wiki

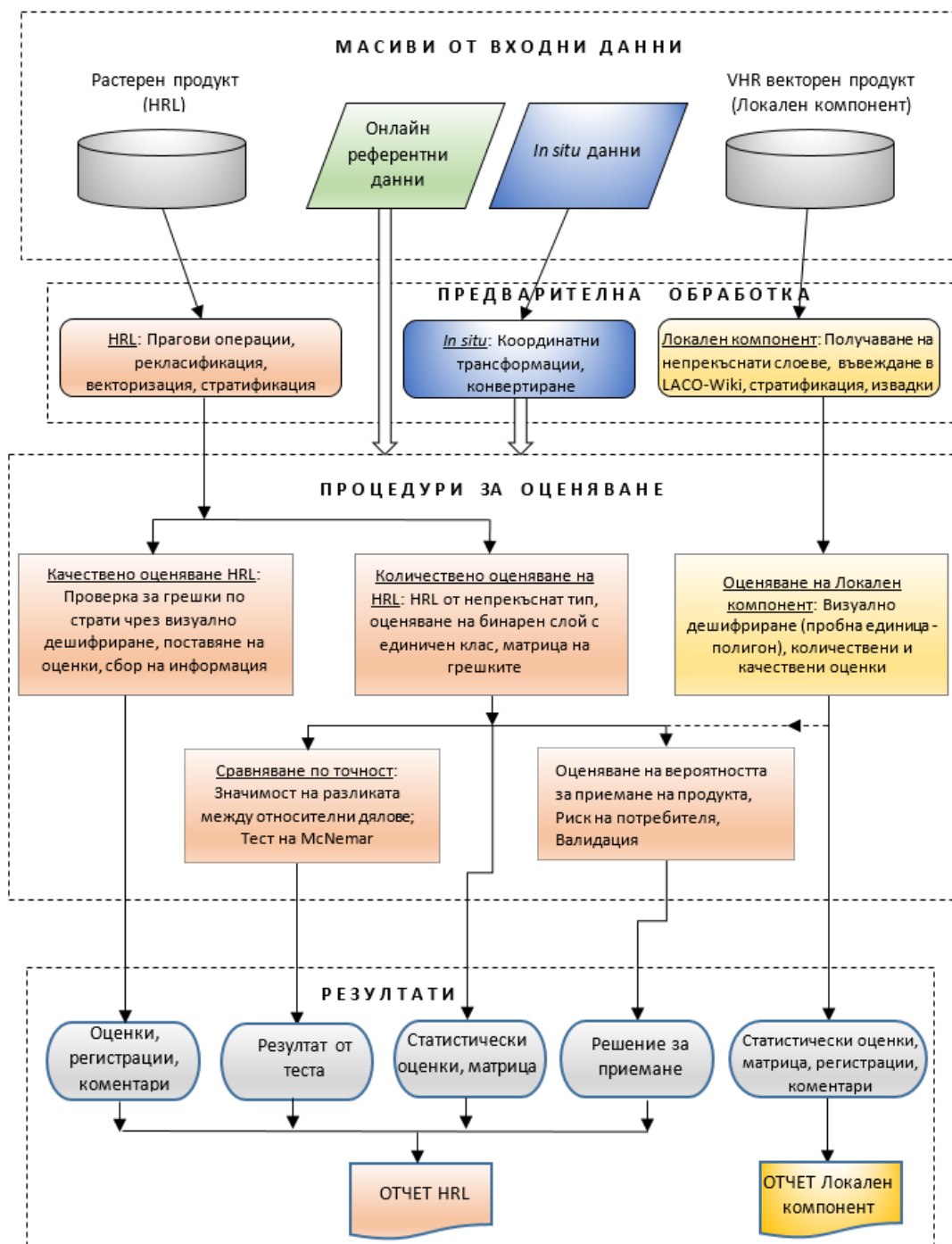
Взема се решение относно верността на тематичния клас за всеки пробен полигон, а чрез падащи менюта се дават оценки за още три характеристики от качествен тип за полигона: детайлност на очертаването, върност (по състав) на очертаната зона, позиционна точност.

След приключване на текущата сесия в LACO-Wiki, автоматично се създава и предоставя отчет в табличен формат - таблица на грешките с количествени оценки.

2.4. Методика за изследване и оценяване на тематичната точност на продукти за земното покритие, получено от спътникови изображения

За реализирането на Комбинирания качествен и количествен подход, на базата на подобрите и развити за целите на настоящото изследване методи, описани по-горе, бе разработена методика, чийто модел във вид на структурна схема, е показан на Фиг. 2.13.

Методиката, като последователност от стъпки, работни процедури и операции, с връзките между тях, има предназначението да служи като носеща методико-алгоритмична платформа за разрешаването на възможните ситуации по изпълнението на поставените пред изследването задачи, а и бъдещи такива. Състои се от четири блока - Масиви от входни данни, Предварителна обработка, Процедури за оценяване и Резултати - всеки от тях включващ в себе си семантично сходни блокове, изобразяващи масиви от данни, операции или резултати.



Фиг. 2.13. Блок - схема на методиката за оценяване на продукти на земното покритие

На блок-схемата могат да се проследят две относително независими потока на движение от входните данни към финалните документи-отчети - този на растрните слоеве с висока разделителна способност и другия - на векторните продукти от тип полигон на Локалния компонент, получени от VHR спътникови изображения. Текстовите описания на отделните блокове говорят достатъчно за тяхното съдържание. Връзката след блока за оценяване на Локалния компонент към блоковете за оценяване на вероятността за приемане и сравняване по точност, обозначена с пунктирна линия, показва възможност за прилагане на тези два процедурни блока към резултатите от оценяването на локалния компонент.

Изводи към Глава 2

Изводът, базиран на дизайн, една от количествените методологии в качествено-количествения подход, се опира на статистически принципи, по които се събират проби и се извеждат характеристики на цялата популация. Изводът, базиран на модел, може да дава на потребителите количествена оценка за решението на класификатора за всеки отделен пиксел от входното спътниково изображение. Методите за качествени оценяване се прилагат за получаване на евтини и бързи качествени оценки и друга информация на етапа на верификацията, като имат коригиращи функции.

Пространствена стратификация и компютърно подпомогнато визуално дешифриране са базови методи, използвани както от качествените, така и от количествените методи за оценяване.

Качествените методи, като този за визуално-качествено оценяване могат да разкриват причини и връзки, които не биха се разкрили чрез количествените методи, респективно, причини и връзки относно грешките в тематичните продукти.

Проектираната и приложена технологична схема за измерване на плътностите на непрекъснатия слой за степента на проницаемост чрез обективно регистриране на отговора ще осигури надеждни резултати за референтните стойности.

Най-важните характеристики на тематичната точност на единичен клас са оценките за двата типа грешки - *commission* и *omission*, изчислени за оценявания клас.

Прилагането на стратифицираната случайна извадка на базата на класифицираните класове земно покритие гарантира, че редките класове са достатъчно добре представени в извадката, за да се отговори на целите на оценяването на точностите на потребителя. Удобство в този смисъл предоставят средствата на LACO-Wiki за настройване на броя полигони за всеки клас от извадката в диалогов режим.

ГЛАВА 3.

Резултати от прилагането на методиката за оценяване на продукти на земното покритие

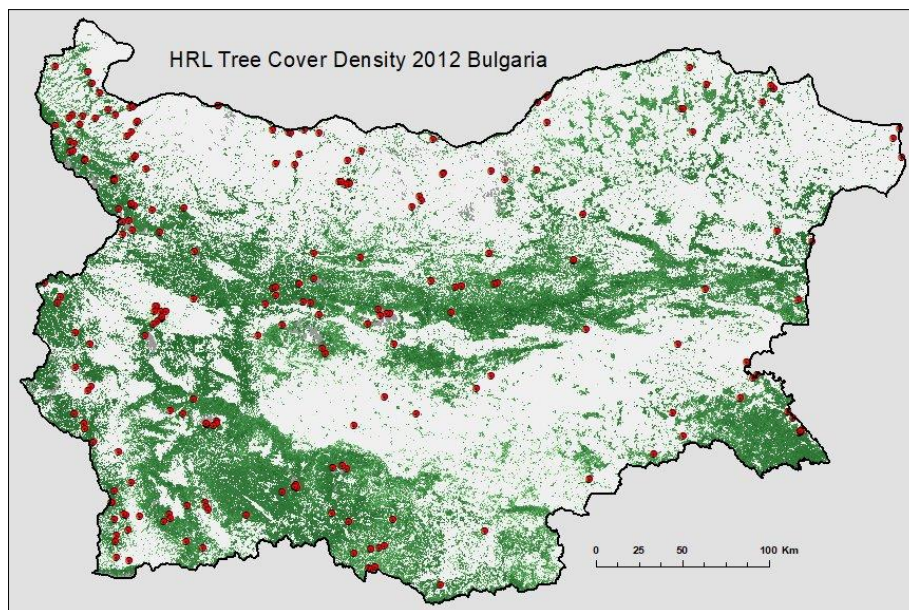
3.1. Резултати от верификацията на HRL за горите TCD и DLT от 2012 и 2015 г.

3.1.1. Оценка на горските HRL 2012

Плътност на дървесното покритие TCD 2012

В слоя Плътност на дървесното покритие погрешна класификация (*commission*) се среща при тревните площи край горите, тревни площи с храсти или при тревни и рядко залесени територии (Dimitrov, 2013a).

На фигура 3.1 е представен общ вид на слоя "Плътност на дървесното покритие 2012" с разпределението на пробните точки. Събирането на пробите е целенасочено, по преценка, затова не следва да се очаква равномерно покритие на горските територии.

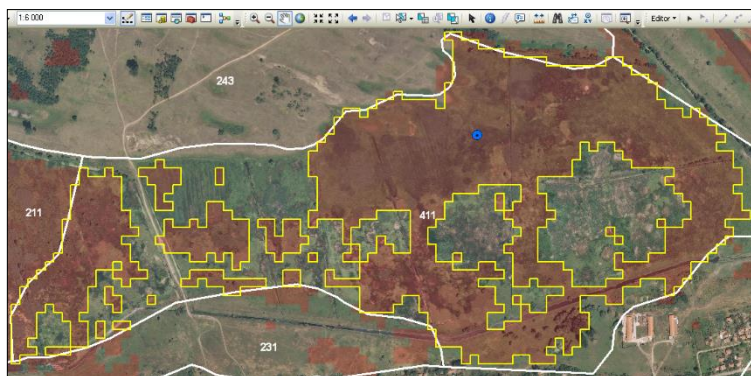


Фиг. 3.1. Картохема на слоя "Плътност на дървесното покритие 2012" с пробните точки, показани в червен цвят

Слоят е верифициран в 15 страти - 5 за погрешна класификация и 10 - за пропуски. последвани от коментари за типични проблемни случаи. Общата оценка е "добра", но в стратите за погрешна класификация резултатите са по-слаби.

Коментари от верификацията на слоя "Карта на дървесното покритие 2012"

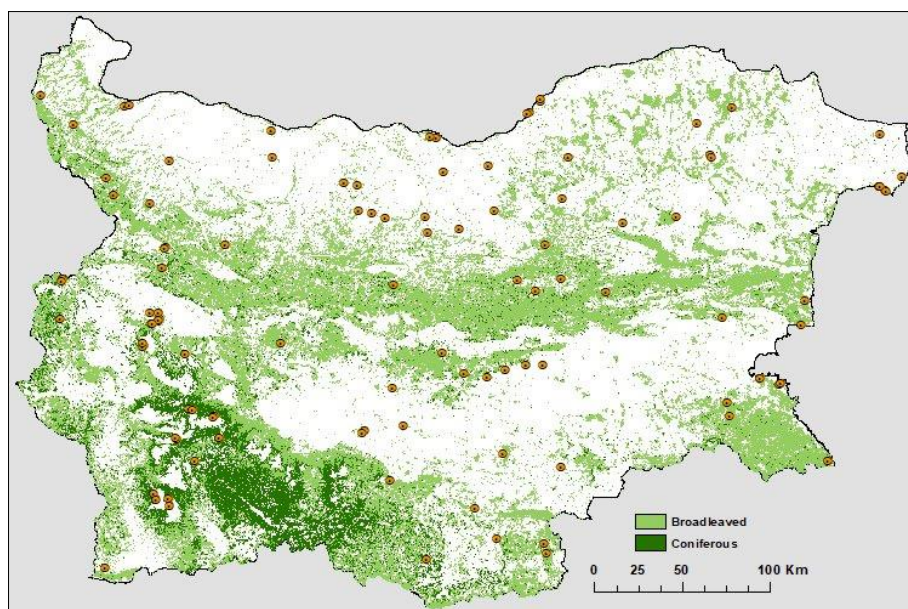
- Грешки от тип погрешна класификация (commission) възникват в територии с тревни зони с храсти в полигони CLC 324, както и в полигони CLC 243 с горски фрагменти;
- Големи площи с грешки от погрешна класификация от са налице в алпийските части на планините Рила и Пирин, заети предимно от клек. Някои случаи на грешки от погрешна класификация възникват и в безлесни затревени зони;
- В рамките на стратата Wetlands (Влажни зони) блата или затревени части от влажна зона (предимно близо до река Дунав и морския бряг) са неправилно картографирани като гора (Фиг. 3.2);
- В горските територии възникват грешки от погрешна класификация, в резултат на картографирането тревни площи като гора. Производителността на класификатора се влошава в случай на поява на негорски свойства на ситуацията в големи горски площи, например, пътни завои и скали, водещи в резултат до незначителни локални грешки от тип пропуск.



Фиг. 3.2. Блатиста местност, класифицирана като дървесно покритие в рамките на CLC полигон 411. В жълто е границата на стратата, а в бяло – границата на CLC полигона. Синята точка маркира пробата и е носител на данните за нея.

Преобладаващ дървесен тип DLT 2012

Фиг. 3.3 дава представа за разпределението на пробите от верификацията по територията на страната за слоя „Преобладаващ дървесен тип“.



Фиг. 3.3. Картосхема на слоя " Преобладаващ дървесен тип 2012" с пробните места в оранжев цвят

Основните тематични класове са три и стойности, които могат да имат пикселите са: 0 - всички негорски площи, 1 - широколистна гора, 2 - иглолистна гора (Dimitrov, 2013b). Коментарите разкриват големи проблеми, възникнали по време на проверката на DLT 2012.

Слоят DLT 2012 е проверен в 12 страти, от които 5 commission и 7 omission. За всички omission-страти оценката е "добра", докато при commission-стратите преобладава "неудовлетворителна".

Коментари за верификацията на слоя Преобладаващ дървесен тип 2012

- Повечето от проблемите са открити в стратите за грешки от тип погрешна класификация;

- Много от овощните градини, базирани на CLC2006 клас 222, са изоставени и/или залесени и не съществуват във физическите блокове на LPIS. Такива случаи водят до погрешна класификация;
- Големи площи в алпийските части на Рила и Пирин планини, покрити с клек, са картографирани като гори, причиняващи грешки от тип погрешна класификация. Някои случаи на този тип грешки се срещат и в безлесни затревени зони;
- Изискването за минимална ширина от 20 m за картографирани горски полета по принцип не се покрива от класификатора. Ветрозащитните пояси, широки над 30 m, просто се пропускат или по-късно се филтрират, поради изискването за 4-свързаност.

3.1.2. Резултати от верификацията на горските HRL 2015

Склопеност (TCD) 2015 - общ преглед на качеството

Слоят Плътност на дървесното покритие (Фиг. 3.4), респективно слойта Карта на дървесното покритие 2015, е изследван спрямо повечето от наборите от *in situ* данни (Димитров и др., 2019).

Грешки от тип погрешна класификация възникват при тревни фрагменти в близост до гора, треви с храсти, рядко залесени площи и голи сечи в CLC класове 243, 324, 211, 231, както и в класове с НТП 040, 101, 050 от физически блокове на СИЗП. Поляни и други открити площи в гори и някои крайбрежни влажни зони – вътрешни блата, солени и солени блата - са картографирани като дървесно покритие.

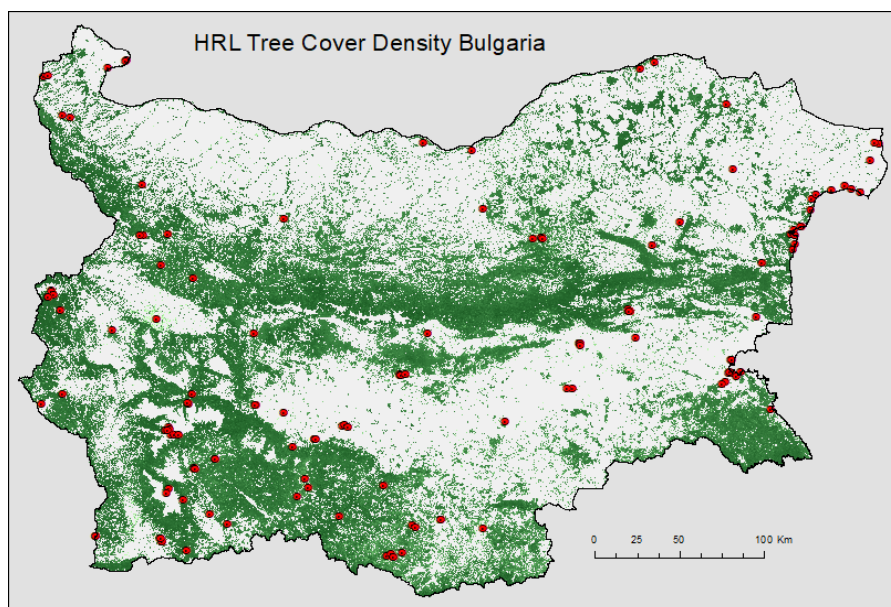


Fig. 3.4. Картохема с общ вид на слоя Плътност на дървесното покритие за България с разпространението на пробните точки за верификация.

Работата на класификатора се влошава в случаите, когато градските райони, напр. пътища, дворове и къщи, се класифицират като дървесни покрития. Такъв е и случаят с лозята. Най-разпространената грешка от погрешна класификация е класифицирането на районите с клек като дървесно покритие.

Look-and-feel верификация на слоя TCD 2015

Качественото оценяване на многоградационният слой TCD 2015 се извършва чрез верификация на бинарния слой Карта на дървесното покритие 2015 (TCM). Стратите, броят на пробите и присвоените оценки по страти, както и общата оценка за слоя са представени в таблица 3.3.

Таблица 3.3. Верификация Look-and-feel на слоя TCM 2015.

Страта №	Име на стратата	Брой проби	Оценки по страти
Commission			
1	Transitional woodland-shrub	12	<i>добра</i>
2	Moors and heathland and sclerophyllous vegetation	11	<i>неудовлетворителна</i>
3	Shrubs and grasslands LPIS101	10	<i>приемлива</i>
4	Wetland	11	<i>неудовлетворителна</i>
Omission			
5	Urban vegetation	10	<i>добра</i>
6	Trees in sport and recreation areas	9	<i>добра</i>
7	Orchards, fruit trees	12	<i>неудовлетворителна</i>
8	Lowland forests, broadleaved	12	<i>добра</i>
9	Lowland forests, coniferous	10	<i>добра</i>
10	Mountain forests, coniferous	11	<i>добра</i>
11	Mountain forests, broadleaved	10	<i>добра</i>
12	Coastal forests	10	<i>добра</i>
13	Forest along rivers & lakes	10	<i>добра</i>
Обща оценка			<i>добра</i>

Най-често срещани грешки от погрешна класификация в TCM 2015

Страта Transitional woodland-shrub (Преходна дървесно-храстова растителност) – пасища, класифицирани като гора, увредена от корояд гора и сечи в иглолистни и широколистни гори, неправомерно са включени в слоя Плътност на дървесното покритие.

Страта Moors and heathland and sclerophyllous vegetation (Растителни съобщества на храсти и треви и склерофилна растителност) – най-често срещаната грешка от тип погрешна класификация е картографирането на районите с клек като дървесно покритие (фиг. 3.5).

Wetland (Влажна зона) – като гора се погрешно се дават предимно вътрешни блата в близост до Черно море и река Дунав. В някои случаи част от солници и солени блата са включени в слоя Плътност на дървесната покривка.

Shrubs and grasslands LPIS101 (Храсти и тревни зони, съгласно клас 101 от СИЗП) – широколистни и иглолистни горски дървостои или група дървета са включени в храсти и пасища;

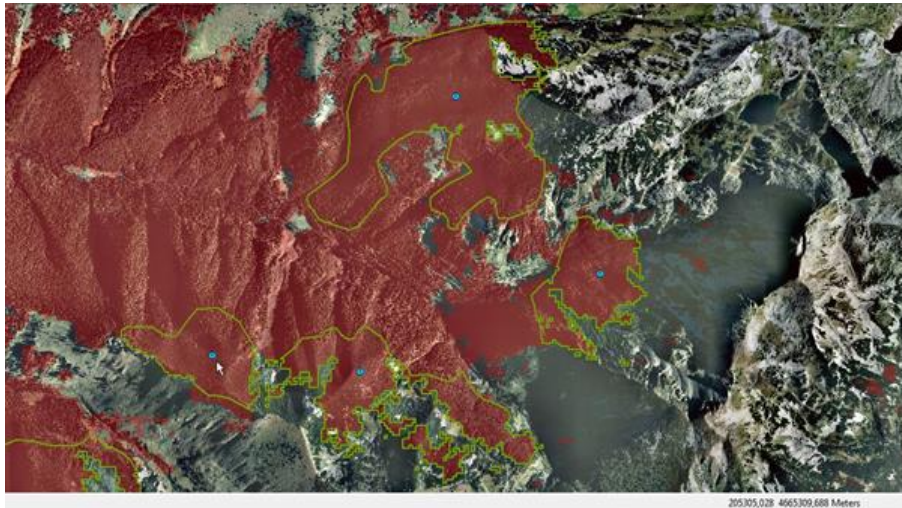


Fig. 3.5. Страта Moors and heathland and sclerophyllous vegetation, съдържаща само клек, който не се третира като дървесно покритие. Четири примера са маркирани със сини точкови символи – клек, неправилно класифициран като гора (полупрозрачен червено запълване).

Грешки от тип пропуск в TCD 2015

Градската растителност и зони за спорт и отдих, като цяло, са правилно картографирани. Типичните открити пропуски по страти са, както следва:

- Градска растителност – групи от дървета в паркове и градски дървета; Дървета в зони за спорт и отдих – групи дървета в морски, голф и ски курорти;
- Овощни градини – липсват насаждения с овощни дървета (CLC222);
- Низинни гори, широколистни – липсващи широколистни насаждения (дъб, космат дъб, акация); Низинни гори, иглолистни – липсват култури от австрийски бор и бял бор;
- Планински гори, иглолистни – липсващи култури от черен бор, бял бор и бяла ела (фиг. 3.6); Планински гори, широколистни – липсват дървостои от бук и космат дъб;
- Крайбрежни гори – липсват както широколистни, така и иглолистни гори; Гора покрай реки и езера – липсващи групи дървета, широколистни насаждения и култури – топола, дъб, акация и трепетлика.

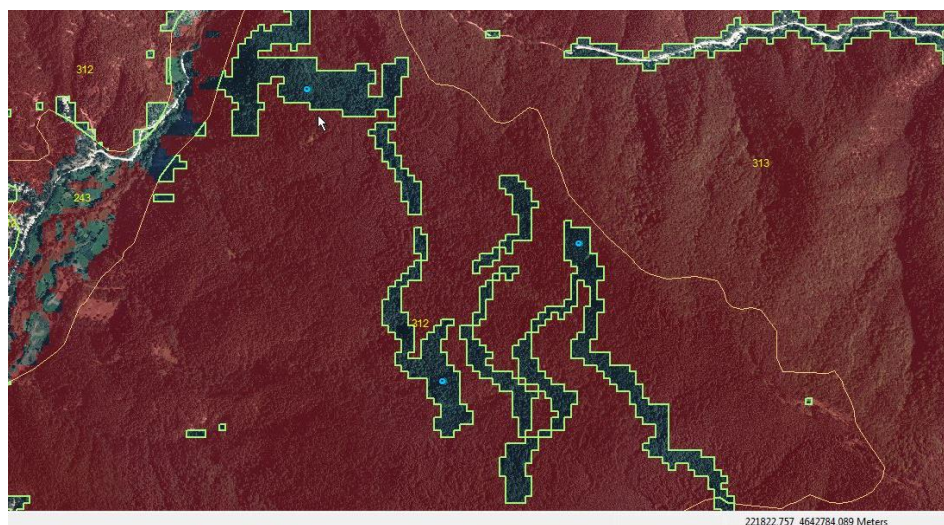


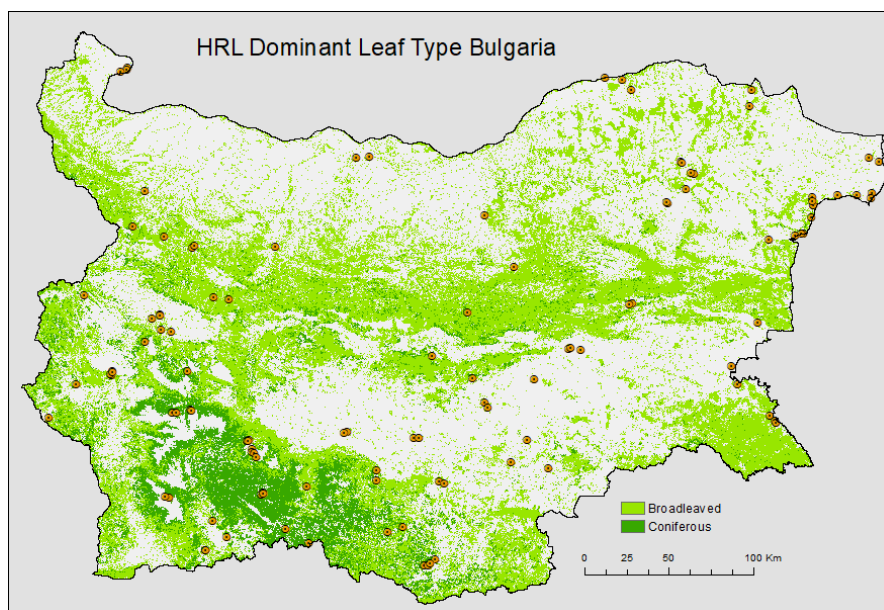
Fig. 3.6. Дървостои от бяла ела са пропуснати при класификацията на слоя TCD (полупрозрачно запълване с червен цвят).

Забележка: Тези грешки от тип пропуск са идентифицирани с помощта на базата данни за инвентаризация на горите, с нейната подробна таблица с атрибути, използвана като референтни *in situ* данни. Три такива случая са маркирани със сини точки.

Въпреки откритите грешки, както слоят DLT, така и слоят TCD от 2015 г. получават „добра“ оценка като обща оценка. Същата оценка преобладава и по страти.

Преобладаващ дървесен тип DLT 2015

На Фиг. 3.7 е показана обзорната картосхема на слоя Преобладаващ дървесен тип (DLT) 2015 с наложените върху него пробни точки.



Фиг. 3.7. Картосхема с общ вид на слоя "Преобладаващ дървесен тип 2015" с пробните точки в оранжев цвят

Преобладаващият дървесен тип, като цяло, е картографиран вярно. Чести грешки от тип погрешна класификация са широколистни фрагменти по сенчестите склонове на дълбоки долини и близо до пътища и реки, които се класифицират като иглолистни.

Верификация на слоя DLT 2015 по метода Look-and-feel

Стратите за слоя Преобладаващи дървесен тип (DLT) са разделени на две групи, за съответния вид грешки (погрешна класификация и пропуск), като имената на стратите, броя на пробите и оценките са представени в Таблица 3.4.

Таблица 3.4. Look-and-feel анализ на слоя DLT 2015.

Страта №	Име на стратата	Брой проби	Оценки по страти
Commission			
1	Transitional woodland-shrub	10	<i>добра</i>
2	Moors and heathland and sclerophyllous vegetation	6	<i>много лоша</i>
3	Shrubs and grasslands LPIS101	7	<i>добра</i>
4	Wetland	6	<i>приемлива</i>
Omission			
5	Urban vegetation	8	<i>добра</i>
6	Trees in sport and recreation areas	8	<i>добра</i>

7	Orchards, fruit trees	10	<i>приемлива</i>
8	Lowland forests, broadleaved	11	<i>добра</i>
9	Lowland forests, coniferous	9	<i>добра</i>
10	Mountain forests, coniferous	10	<i>добра</i>
11	Mountain forests, broadleaved	10	<i>добра</i>
12	Coastal forests	10	<i>добра</i>
13	Forest along rivers & lakes	10	<i>добра</i>
Обща оценка			<i>добра</i>

Слоят DLT, като цяло, е правилно картографиран. От списъка на стратите по-горе (колона „Оценки по страти“) се вижда, че повечето от проблемите са открити в стратите за търсене на грешки Comission.

Най-често срещани грешки от тип погрешна класификация (comission)

Страта Transitional woodland-shrub (Преходна горско-храстова растителност) – култури от бял и черен бор са дадени като широколистни (фиг. 3.8).

Wetland (Влажна зона) – широколистна, дадена като иглолистна гора. Безлесните затревени и/или влажни части от влажните зони често се картират като гора.

Страта Moors and heathland and sclerophyllous vegetation (Растителни съобщества на храсти и треви и склерофилна растителност) – тази страта е изцяло базирана на CLC клас 322, който за България включва само клек, който не се третира като дървесно покритие.

Типични грешки от тип пропуск в DLT 2015 по страти

Резултатите от верификацията за слоеве Urban vegetation и Trees in sport and recreation areas са добри и в много случаи - отлични. Бяха открити горски фрагменти, които не са картографирани (пропуснати са) в слоя Преобладаващ дървесен тип 2015. Много самозалесявания в овощни градини, базирани на клас 222 в CLC2018 са пропуснати.

Lowland forests, broadleaved (Низинни гори, широколистни) – издънкови широколистни гори са пропуснати, както и Lowland forests, coniferous (Низинни гори, иглолистни) – пропуснати култури от чер и бял бор.

В страта Mountain forests, coniferous (Планински гори, иглолистни) са пропуснати части от насаждения от иглолистни гори, а в страта Mountain forests, broadleaved (Планински гори, широколистни) са пропуснати части от насаждения от букови и дъбови гори.

Пропуски са налице и в страти Coastal forests (Крайбрежни гори) и Forest along rivers & lakes (Гора покрай реки и езера).

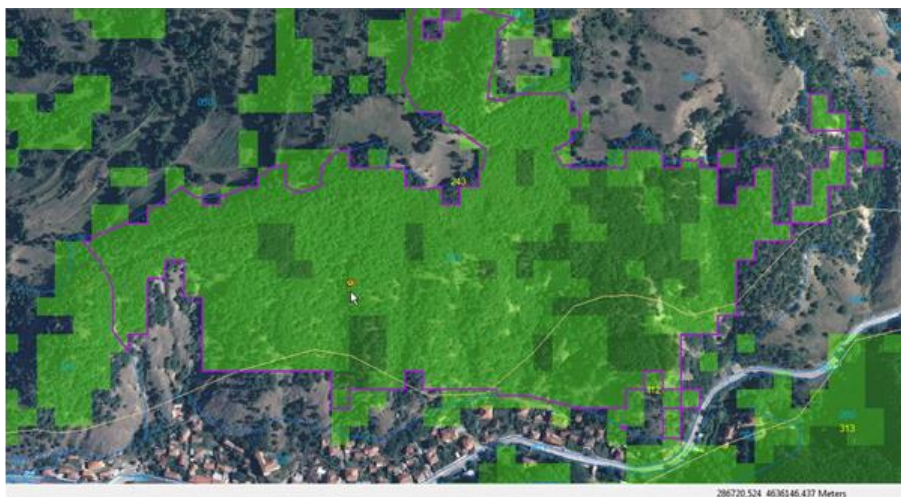


Fig. 3.8. Страта Храсти и треви LPIS101 – части от широколистни гори са класифицирани като иглолистни, а култури от бял и черен бор са дадени като широколистни гори.

Изводи относно верификацията на HRL за горите

С верификацията на горските слоеве HRL 2012 и 2015 е реализирана методология за оценка на качеството, насочена към идентифициране на типични проблеми, с цел коригиране на продуктите.

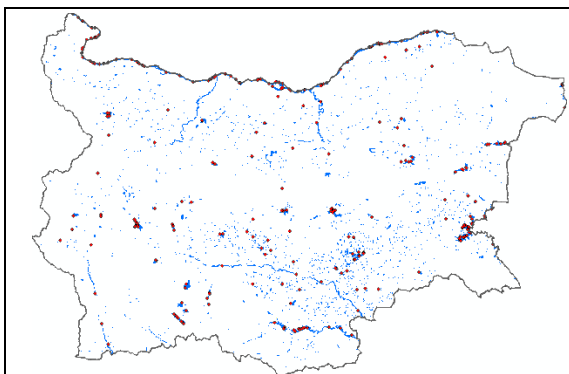
Въпреки регистрираните и описани грешки от пропуски и погрешна класификация, можем да заключим, че слоевете с висока разделителна способност Плътност на дървесното покритие и Преобладаващ дървесен тип показват последователно добри резултати.

Общите резултати от задачата за проверка на HRL показват добра обща способност на автоматизиран анализ на разновременни сателитни изображения за извличане на определени типове земно покритие.

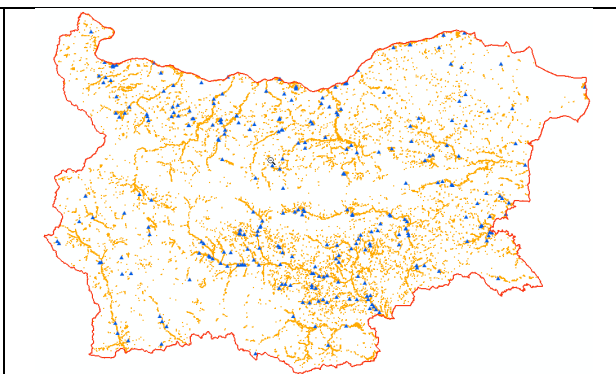
3.2. Резултати от оценяването с количествени методи на непрекъснати и бинарни слоеве с единичен клас земно покритие

3.2.1. Резултати количественото оценяване на продукта PWB2012

Генерирани са по 280 случайни точки за всяка от стратите commission и omission за проверка по двата типа грешки (Димитров, 2016а). Стратите са показани в общ вид, съответно, на фигури 3.9 и 3.10.



Фиг. 3.9. Страта за грешките commission за PWB.

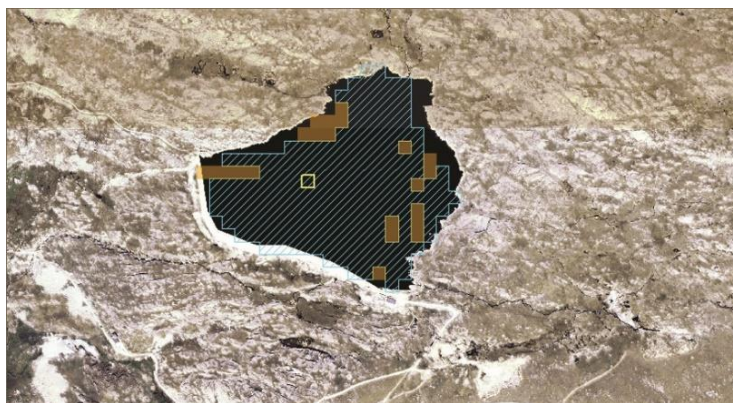


Фиг. 3.10. Страта за грешките omission за PWB.

На базата на двете страти и формулите от Глава 2, са измерени и изчислени следните площи, необходими за пресмятане на точностите и грешките:

- Площ на стратата omission – 53585 ha;
- Площ на класа PWB HRL (страта omission commission) – 89598 ha;
- Оценка за площта за грешките– 5359 ha.

Данните за стратификацията са подготвени на базата на in situ референтни данни с висока детайлност, позволявайки генерирането на точни и неголеми по размер тестови полигони. На Фиг. 3.11 щрихованият полигон върху езеро в Рила е от стратата на пропуските.



Фиг. 3.11. Регистриране на пропуск в рилско езеро.

Изчислените грешки и точности за слоя PWB2012 са представени в таблица 3.7.

Таблица 3.7. Резултати от оценяването на класа PWB HRL

Измерена величина	Стойност
COMMISSION	
Брой валидни проби	280
Проби, показващи грешка	19
Грешка неправилна класификация (commission error)	6.79 %
Точност на потребителя	93.21 %
Несигурност	1.50 %
OMISSION	
Брой валидни проби	280
Проби, показващи грешка	28
Отношение на грешката	10%
Точност	90%
Точност на производителя	94.02 %
Несигурност	1.07 %
Грешка от тип пропуск	5.98 %

Трябва да се отбележи, че частта OMISSION в таблицата се състои от две групи данни. Първите три реда се отнасят за стратата на пропуските (omission) и представляват набора от данни от измерването в тази страта. Данните във вторите три реда (Точност на производителя и надолу) характеризират слоя PWB (териториално заемащ стратата commission) откъм грешките за пропуските.

Резултатите от пробите показват висока за надеждност на оценките.

3.2.2. Изследване на слой от непрекъснат тип IMD2012 100m чрез обективно измерване на отговора

С оглед на големия интерес към зоните с по-висока степен на почвеното запечатване, т.е., плътното застрояване (built-up), оценяването на IMD като непрекъснат слой е извършено в рамките на Страта 1 (т.е., за $IMD > 30\%$). Следван е алгоритъмът от фиг. 2.7. от Глава 2. Събирането на пробите е изпълнено чрез обективното измерване на степента на непроницаемост с мрежа 10x10 (фиг. 3.12) за 250-те първични проби.

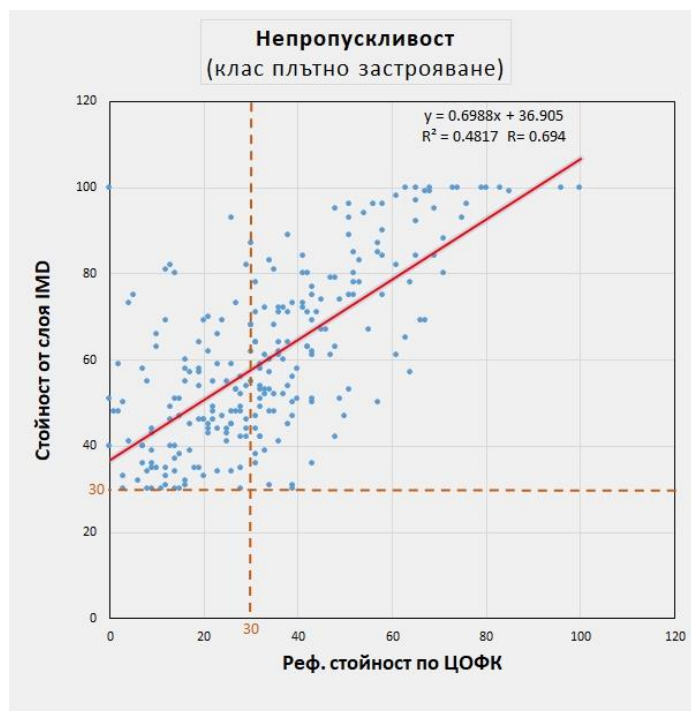


Фиг. 3.12 Измерване с равномерна мрежа 10x10 в населено място

След обработката и анализа на данните от извадката са получени следните три вида резултати: диаграма на разсейването, линия на регресията с дефиниращото я уравнение и коефициент на детерминация R^2 (фиг. 3.13). Тъй като по дефиниция Страта 1 се формира от пиксели със стойности от 30 до 100, то е логично и разполагането на точките в диаграмата над хоризонталната линия $IMD = 30$.

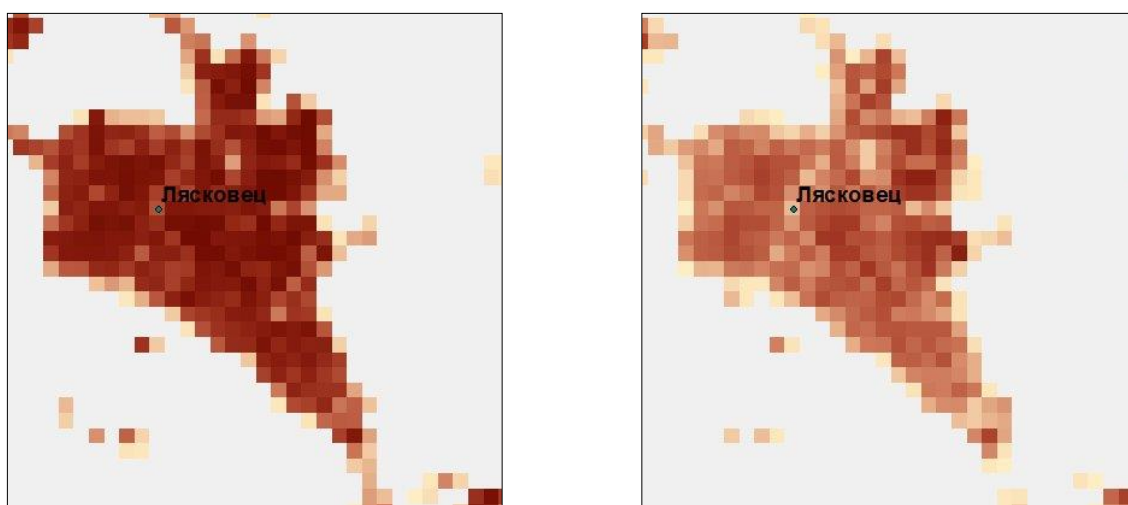
От данните на фигурата също става ясно, че корелационният коефициент R е 0.694, което, по принцип, говори за значителна корелация между предсказаните от класификатора стойности на степента на непроницаемост в Страта 1 и съответните им референтни стойности, измерени по ЦОФК. От друга страна, коефициентът на детерминация $R^2 = 0.482$ показва, че по-малко от половината (48%) от дисперсията в IMD се дължи на вариациите в референтните данни.

Следва да се отбележи фактът, че много голяма част от точките в диаграмата на разсейването с абсциси в диапазона 0-30 имат ординати значително над стойността 30. Това е отразено в стойността 36.91 на свободния член на уравнението и в отместването на регресионната права и целия модел към по-високите стойности на IMD. Налице е очевидно надценяване на степента на непроницаемост от страна на данните за слоя IMD 2012, обект на настоящото изследване. Тази констатация се потвърждава и визуално от сравняването на изображенията на IMD 2012 от времето преди и след повторната обработка/корекция на продуктите HRL 2012 (Фиг. 3.14).



Фиг. 3.13 Диаграма на разсейването и данни за регресията за IMD 2012 100 m

Накрая е редно да споменем, че параметрите на линията на регресията, съответстваща на множеството отчети предоставя информация за евентуално калибриране на непрекъснатия слой.



Фиг. 3.14. Изображения IMD2012 100m на гр. Лясковец: преди (ляво) и след (дясно) коригирането на продукта.

3.2.3. Резултати от анализа на точностите в бинарния слой на застрояването

Тематична точност на класа на плътното застрояване Built-up

Първият резултат относно бинарния слой на застрояването е за частта от него, отнасяща се за плътно застроените територии (клас Built-up) - тези със стойности на непроницаемостта 30% - 100%. Както бе обосновано в раздел Оценяване на бинарния слой на застрояването на Глава 1, стратите се определят, както следва:

- Страта1, страта за грешки commission - стойности на непроницаемост 30% – 100% според класификатора, т.е., площите с плътно застрояване (клас Built-up);
- Страта2, страта за търсене на грешки от тип пропуск - обхваща териториите със стойности 1% - 29%, т.е., липса на плътно застрояване.

Резултатите от събиране на пробите и анализа на данните са представени в таблица 3.8:

Таблица 3.8. Резултати от оценяването на единичен клас Built-up

Измерена величина	Стойност
COMMISSION	
Брой валидни проби	250
Проби, показващи грешка	116
Грешка неправилна класификация (commission error)	46.40%
Точност на потребителя	53.60%
Несигурност	3.15%
OMISSION	
Брой валидни проби	200
Проби, показващи грешка	7
Отношение на грешката	3.50%
Точност	96.50%
Точност на производителя	96.28%
Несигурност	1.07%
Грешка от тип пропуск	3.72%

Прави впечатление твърде високата стойност на грешката от неправилна класификация (commission error), респективно, ниската точност на потребителя, който резултат е в съзвучие с резултатите от предишния раздел за Изследване на слой от непрекъснат тип IMD 2012 100 m.

Величината "Несигурност" = 3.15% в раздела COMMISSION на горната таблица е изчислена при доверителна вероятност 68.26%, т.е., $z\alpha/2 = 1$, визирайки формула (6) от глава 2. Повишавайки изискванията към доверителната вероятност на 90%, води до увеличаване на несигурността 1.65 пъти ($z\alpha/2 = 1.65$). Така получаваме доверителен интервал CI за commission-грешката $p \text{ CI} = \pm 5.2\%$, което може да бъде записано като:

$$p = 46.4 \pm 5.2\% \text{ или } 41.2\% < p < 51.8\%.$$

Прилагане на критерия на вероятността за приемане с цел валидация на продукти за земното покритие

В експериментите по използването на вероятността за приемане с цел валидация на продукти за земното покритие, данните от вероятностните извадки и техния анализ се въвеждат в компютърен модел във вид на скрипт за софтуерната среда R.

Резултантните параметри, извеждани във вид на текст поотделно за стратите 1 (commission) и 2 (omission).

Отделно резултатите се извеждат и в графичен вид, подобно на примерната графика от глава 2 - фиг. 2.11.

Проведени са експерименти с данните от оценяване на два HRL продукта с единични класове - слоя на застрояването и слоя на постоянните водни тела. Резултатите, генерирани като изход от гореспоменатия програмен модел са, както следва:

Листинг за слоя на застрояването 2012 - страти Built-up и Non-built-up

Confidence interval: 90 %
 Excluded area: 94.95 % of the total area

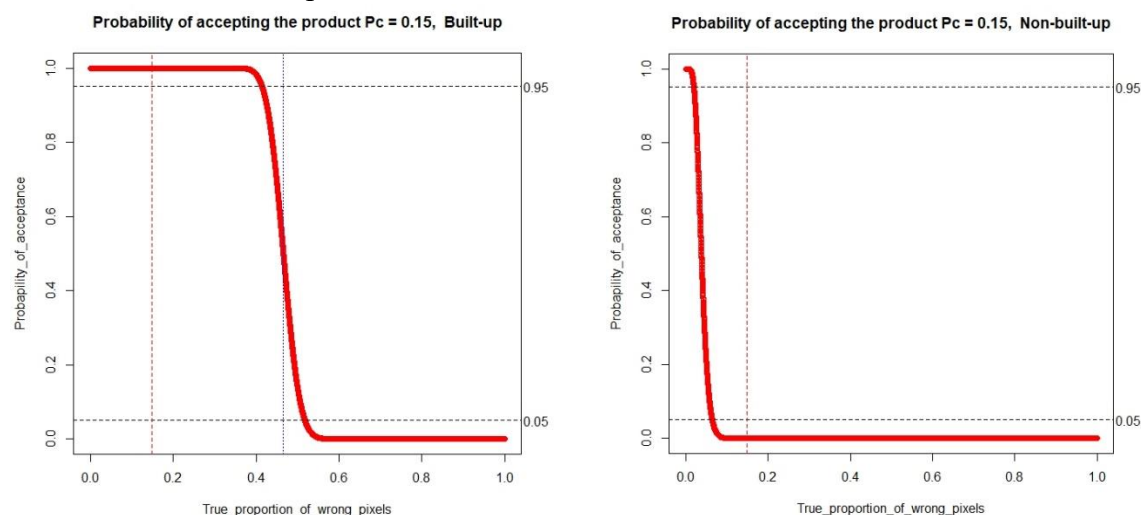
Results for Stratum 1 - Built-up

(2.63 % of the total area)
 Number of samples: 250
 Number of samples containing error: 116
 With 95 % probability the Built-up class contains more than $p_0= 41.45$ % commission error
 With 95 % probability
 the Built-up class contains less than $p_1= 51.8$ % commission error
 Mean commission error: 46.62 %
 Reliability of commission error: 5.17 %
 Probability of having more than 15% commission error in
 the database: 100 %

Results for Stratum 2 - Non-built-up

(2.42 % of the total area)
 Number of samples: 250
 Number of samples containing error: 22
 With 95 % probability the Built-up class contains more than $p_0= 5.85216730038023$ % omission error
 With 95 % probability
 the Non-built-up class contains less than $p_1= 11.3454752851711$ % omission error
 Mean omission error: 8.6 %
 Reliability of omission error: 2.75 %
 Probability of having more than 15% omission error in
 the database: 0.04 %

Графичният изход от скрипта е показан на фиг. 3.14, резултатите също са поотделно за двете страти.



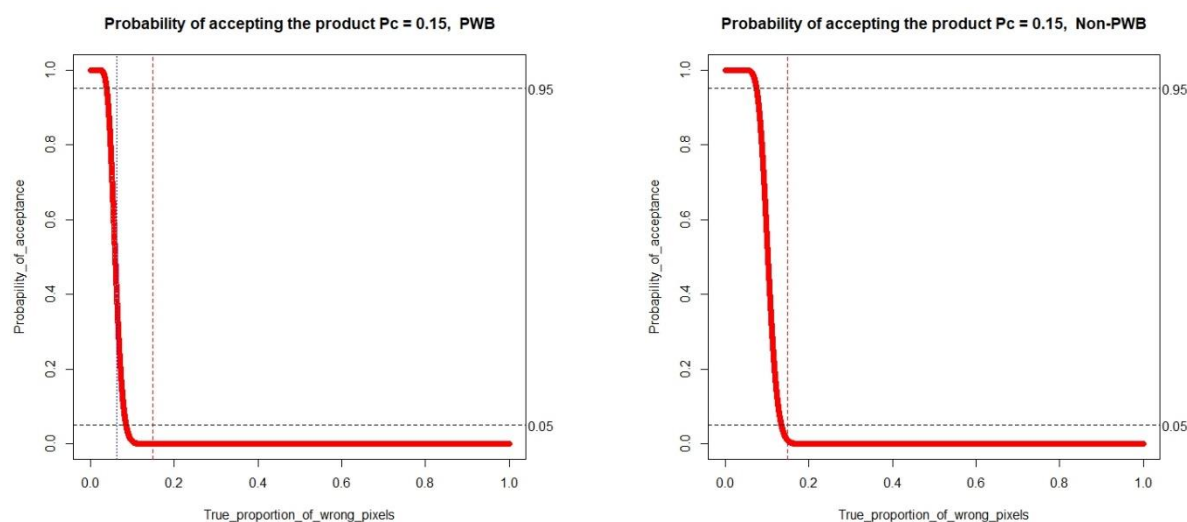
Фиг. 3.14. Вероятност за приемане за слоя на застрояването по страти: страта Built-up (ляво) и страта Non-Built-up (дясно)

За бинарния слой на застрояването резултатите са следните (Димитров, 2016b): с вероятност 95% може да се твърди, че грешката от неправилна класификация в Страта 1 е

46.62 %. Това е стойността на абсцисата на средата на наклонения участък на лявата графика на Фиг. 3.14. Със същата вероятност се твърди, че същата грешка е над $p_0=41.45\%$ и под $p_1=51.8\%$ - абсциси на пресечните точки на графиката с хоризонталните граници 0.05 и 0.95. Вероятността, грешката от неправилна класификация в Страта 1 да надхвърли прага от 15%, е 100%. Това се вижда от факта, че вертикалният червен пунктир с абсциса 0.15 пресича графиката в хоризонталния участък, стремящ се към ордината 1.0. В стратата за пропуските ситуацията е значително по-добра - средна грешка от пропуски 8.6%, долна и горна граница на доверителния интервал, съответно, 5.85% и 11.35%. Вероятността за надхвърляне на 15% праг е 0.04 %.

В заключение, относно резултатите от горния експеримент с бинарния слой на застрояването, оценяван в две страти, следва да се направи изводът, че слойът не издържа валидационния тест, при който и двете грешки - пропуск и погрешна класификация - трябва да са под 15%.

Показателите на слоя на постоянните водни тела са отлични, което е в съгласие с обявените в по-горен раздел резултати от количествения тест на този слой. Средните стойности на грешките commission и omission, съответно в страти PWB и Non-PWB: 6.23% и 6.27%. Вероятностите за надхвърляне на 15% праг са нулеви и за двете страти. Графичните резултати (фиг. 3.15) потвърждават ниските стойности на грешките.



Фиг. 3.15. Вероятност за приемане за слоя на постоянните водни тела по страти: страта PWB (отляво) и страта Non-PWB (отдясно)

Изследване на тематичната точност на бинарния слой-карта на застрояването в три страти

В досегашното изложение бе споменато, че обемите на извадките използвани при събирането на проби в слоя IMD, респективно, в бинарния слой на застрояването, за двете страти Страта1 (Commission) и Страта2 (Omission) са съответно 250 и 200. Добавянето в изследването на втора страта за пропуските със 100 проби, с което стратите стават три, повишава точността на оценяването на тематичната точност за слоя "Плътно застрояване" и то в посока на грешките от тип пропуск.

Както вече бе споменато, на етапа на анализа на данните от събраните проби по трите извадки, е използван онлайн инструментът Map Accuracy Tools, който експортира отчет в Ексел формат, показан в Таблица 3.9.

Таблица 3.9. Матрица на грешките за слой "Карта на застрояването " за оценяване с три страти

Confusion Matrix

		Reference Data			UserAccuracy	UserAccuracyVariance
		Omission_Low	Omission	Commission		
Classif. Data	Omission_Low	51	48	1	51.00%	0.098474
	Omission	27	166	7	83.00%	0.052191
	Commission	4	112	134	53.60%	0.061944
	Weights	164155	310675	292111		
	ProducerAccuracy	64.23%	55.15%	92.60%		
	ProducerAccuracyVariance	0.088069	0.032531	0.047545		
	OverallAccuracy	0.649529				
OverallAccuracyVariance	0.038051					

На фона на сумарната площ на слоя "Карта на застрояването" в рамките на стратите, използвани при оценяването, общата точност е умерено ниска, като причина за това, естествено, е много ниската точност на потребителя в класа/стратата Commission, тоест, класа на плътното застрояване Built-up. Дисперсиите на всички оценки в матрицата са ниски, което е потвърждение за правилното проектиране на модела за събиране на пробите.

3.2.4. Резултати от сравняване на HRL на основата на тематичната точност

Експериментите по изпълнение на планираните в глава 2 тестове за сравняване с цел определяне на статистическата значимост на разликите между версиите на продуктите IMD2012 100m преди и след корекциите от 2015 г. (HRL, 2017) са изпълнени с помощта на съответните им математически апарат - z-тест тест на McNemar.

Прилагане и резултати на z-тест за случая на независими извадки

За изпълнение на z-теста подготвяме данните, необходими за прилагана на формула 16 от Глава 2.

От документа (HRL, 2017) използваме таблица Table 2-2 Distribution of sample units per main strata and substrata на стр. 14, като сумираме стойностите за броя на пробите Omission и Commission за стратите High и Low. След умножаването на горните бройки проби с точностите User's и Producer's (на потребителя и производителя) от таблица 5-3, получаваме референтните данни, представени в таблица 3.10 .

Таблица 3.10. Данни за IMD 2012 от отчет на ЕАОС за валидацията

	Commission stratum	Omission startum
Commission stratum	106.92	18.827
Omission startum	58.08	262.173

За прилагане на теста пресмятаме необходимите числени стойности поотделно от страна на проверяваните и референтните данни:

- За обема на извадката на проверяваните данни (получени в хода на настоящото изследване) и за броя на правилно определените случаи, след сумиране на броя на съответните проби, получаваме:

$$n1 = 550$$

$$x1 = 426$$

- За референтните данни (от отчета на ЕАОС) за съответните елементи на формулата на теста получаваме следното:

$$n2 = 446$$

$$x2 = 369.093$$

След прилагането на формула 16 от Глава 2, за тестовата стойност се получават следните резултати без и с корекция за непрекъснатост (Foody, 2004):

$$Z_{\text{No Correction}} = -2.073$$

$$Z_{\text{Continuity Correction}} = 1.994$$

И в двата случая нулевата хипотеза не се изпълнява, което води до заключението за значимост на разликата в стойностите на точностите на двете версии на продукта IMD2012. Този резултат показва, че причината за ниските точности на продукта IMD2012 от времето преди корекциите от 2015 г., демонстрирани в хода на нашите измервания, действително се корени в недоброто му качество.

Резултати за случая на зависими извадки (тест на McNemar)

Извършваме сравняване на двете версии на HRL IMD2012 100m - преди и след корекциите от страна на ЕАОС, на основата на свързани, сдвоени извадки - един и същ набор от 250 пробни пиксели (точки) в два различни (сравнявани) растера built-up. Извличането на данните от референтния растер се изпълнява специално за целите на теста на McNemar. Създава се следната таблица (таблица 3.11) от типа "преди и след" за въвеждане в теста:

Таблица 3.11. Входни данни за тест на McNemar

		RASTERVALU		SUM
		Y	N	
RASTVAL_OLD	Y	99	40	139
	N	93	18	111
	SUM	192	58	250

Изчисленията по теста на McNemar са изпълнени с помощта на един от достъпните онлайн ресурси¹⁰. И този резултат потвърждава статистическата значимост на разликата в точностите на двата сравнявани слоя IMD2012 100m преди и след корекциите от 2015 г., този път - на базата на едни и същи пробни места.

Получената стойност за p-value е значително по-малка от стандартното ниво на значимост 0.05 (Фиг. 3.16), използвано и преди от нас в други експерименти. Следователно, имаме достатъчно доказателства, за да отхвърлим нулевата хипотеза –

¹⁰ Omniculator <https://www.omniculator.com/statistics/mcnemars-test>

нашите констатации относно разликата в точностите на двата слоя са статистически значими. Последното означава, че точността на IMD2012 100m преди корекциите от 2015 г. е статистически значимо по-ниска от тази на валидирания от Европейския екип слой след тези корекции. Отново получаваме потвърждение за правилността на нашите измервания и заключения.

	Test 2 positive	Test 2 negative
Test 1 positive	a	b
Test 1 negative	c	d
a		99
b		40
c		93
d		18
Test type	Standard McNemar's test	
Test statistic χ^2	21.12	
p-value	0.000004313344	
Your results are significant at the standard significance level of 5%! You can reject the null hypothesis and accept the alternative hypothesis.		

Фиг. 3.16. Резултатт от теста на McNemar

В резюме, относно резултатите от прилагането на частта от Методиката за количествено оценяване - верификация или валидация - на растерни слоеве на земното покритие с единичен клас, можем да обобщим следното. Бинарният слой за постоянните водни тела показва високи нива на тематичната точност, което е обяснимо, имайки пред вид добрата спектрална разделимост на водата с другите видове земно покритие. Първоначално произведеният като Европейска мозайка слой IMD2012 100m, за територията на България, показва висока стойност за грешки commission, респективно, ниска точност на потребителя. Този резултат е потвърден чрез оценяване като непрекъснат слой, като бинарен слой в две и три страти, както и чрез сравняване с коригирания и взет за референтен слой IMD2012 100 m, на основата на тематичната точност с използване на тестове за статистическа значимост на разликата за независима и зависима извадка.

3.3. Резултати от оценяването на продукти базирани върху VHR спътникови изображения - Локален компонент

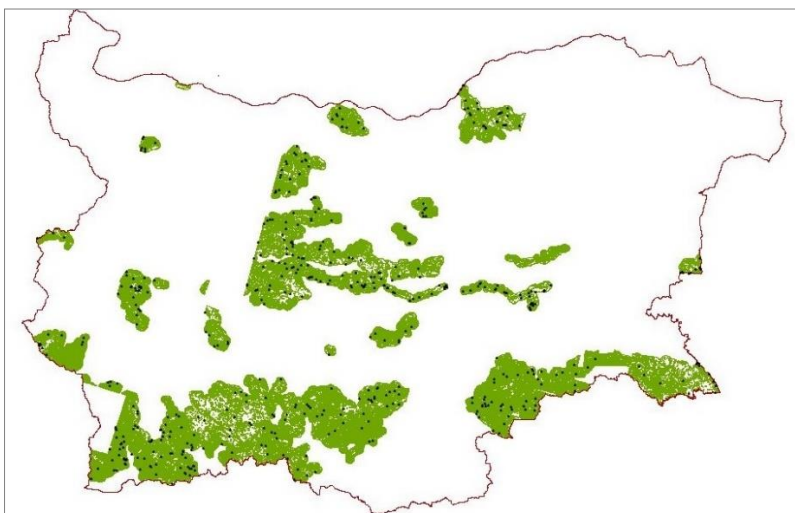
За получаването на информация за качеството на базите данни за продуктите от Локалния компонент на CLMS е приложен комбиниран подход по два метода - чрез количествено и чрез визуално-качествено оценяване. Основната част от дейностите са реализирани с помощта на уеб-базирания инструмент LACO-Wiki.

3.3.1. Резултати от верификацията на Натура 2000 - тревни зони 2012

Прилагаме количествен подход на ниво полигон, за да оценим тематичната точност на слоя Натура 2000 (N2K) и визуално-качествен метод за оценяване на някои геометрични характеристики (Димитров и др., 2019). Нашите цели са да предоставим статистически резултати, сравними с други страни и продукти и да направим качествена

оценка въз основа на местен опит и *in situ* данни. Типичните грешки, открити от местните екипи, могат да се използват за подобрения на технологиите. Друга цел е да се подпомогне идентифицирането на потенциални случаи на използване на N2K.

Общата площ на N2K за страната е 2 203 278 ха, тоест 19.87 % от територията на България (фиг. 3.17).



Фиг. 3.17. Разпределение на верифицираните територии и пробните точки за продукта от Локалния компонент Natura 2000 в България

В резултат на верификацията на продукта LCLU N2K 465 пробни полигона от общо 500 в извадката са причислени към правилния клас. Следователно, общата точност от 93.0 % е резултат, който е по-висок от изискваното ниво от 85 %. За 24 от 47-те класа, характерни за обекти от N2K в България, бе установено, че кодът на всички валидирани LCLU проби е 100 % верен. В по-голямата си част това са линейни обекти, на които трудно може да се присвои друг код от номенклатурата на MAES: пътни мрежи и свързаните с тях земи, железопътни линии и свързани земи, редове от дървета и храсти, речни брегове, водни течения и силно модифицирани водни течения и канали.

Грешки от неправилна класификация

По време на верификацията бяха открити 24 погрешно класифицирани полигона.

На фиг. 3.18 кодът за полигона е 3.2.3.1 Други естествени и полуестествени иглолистни гори. След внимателен преглед на референтните изображения, както и на характеристиките на насажденията в лесоустройствения план, се установи, че това са чисти иглолистни насаждения и кодът, предложен от нас е 3.2.4.1 Силно изкуствени иглолистни насаждения.

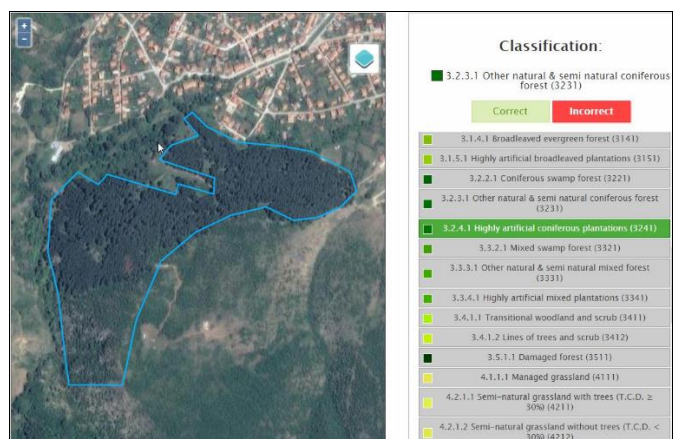


Fig. 3.18. Неправилна класификация 3.2.3.1 Други естествени и полуестествени иглолистни гори – Правилно е 3.2.4.1 Силно изкуствени иглолистни насаждения.

Неправилно очертани зони

В много случаи границите на полигоните са очертани твърде грубо; включват ненужни части или липсват части, а често и двете. В други случаи границите са изместени (фиг. 3.20).



Fig. 3.20. Прекалено грубо очертаване, липсващи и ненужни части, изместени граници.

Обобщени резултати от верификацията на N2K

Като цяло точностите от отчета от LACO-Wiki са добри, като точността на потребителя варира между 50% и 100 %, а точността на производителя попада в диапазона 76.47 % – 100 %. Общата точност е 93 % с доверителен интервал ± 2.06 %. В таблица 3.13 по-долу най-ниска стойност има показателят от качествен тип Вярност на очертаната зона (т.е., включени ненужни или пък, пропуснати части от полигони), което говори за необходимост от подобряване на методите и алгоритмите, въввлечени във формирането на полигоните от отделните сегменти.

Таблица 3.13. Оценени точностни характеристики на N2K.

Оценки на точността	Стойност
Обща точност	0.93
СИ на общата точност	± 0.0206
Капа	0.928246

СИ на капа	± 0.0229
Геометрично съгласуване	
Вярност на очертаната зона	0.69
Детайлност на очертаването	0.854
Позиционна точност	0.974

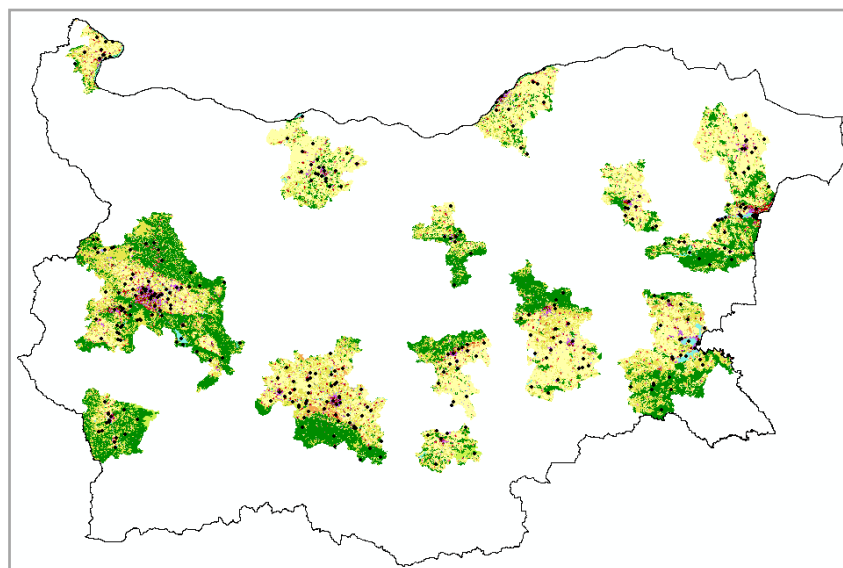
3.3.2 Резултати от верификацията на Атлас урбанизирани територии

В хода на верификацията на продукта UA 2012 бяха проверени 16 налични FUA (Таблица 3.14), покриващи площ от 2 719 508 ha, което представлява 24.5% от територията на страната (Димитров и др., 2019).

Таблица 3.14: Верифицирани FUA от UA2012

Град	Площ, ha	Град	Площ, ha
Благоевград	135485	Шумен	94515
Бургас	294941	Сливен	136664
Добрич	140394	София	571690
Хасково	73967	Стара Загора	132067
Пазарджик	63684	Варна	204054
Плевен	179161	Велико Търново	88510
Пловдив	277227	Видин	62242
Русе	133950	Ямбол	130958

Разпределението на 16-те FUA в страната е показано на фигура 3.21, като, като териториално съседните FUA изглеждат като едно общо петно на картосхемата.



Фиг. 3.21. Общ вид на статус слоя на Urban Atlas 2012 с пробни места (черни точки).

От 329 проби за масива от данни на UA, 279 имат правилен клас LCLU, което води до обща претеглена точност от 0.839 с ± 0.088 СИ. Подробни количествени резултати за тематична точност на всички класове LC/LU на UA са представени в таблица 3.15.

Таблица 3.15. Характеристики за оценка на точността на статус слоя UA 2012 – България

Class Code	Samples	Weight	UAc, %	PAc, %	Class Code	Samples	Weight	UAc%	PAc, %
11100	13	0.005613	61.54	100.00	13300	13	0.00033	61.54	100.00

11210	13	0.014381	61.54	83.67	13400	13	0.000724	84.62	44.75
11220	13	0.008331	76.92	65.88	14100	13	0.002459	76.92	100.00
11230	13	0.003766	100.00	61.33	14200	13	0.002025	100.00	100.00
11240	13	0.001469	53.85	25.85	21000	12	0.359437	91.67	89.73
11300	13	0.001019	76.92	100.00	22000	12	0.00982	83.33	100.00
12100	13	0.015797	92.31	94.24	23000	12	0.221461	50.00	95.96
12210	13	0.000374	100.00	100.00	24000	12	0.000543	100.00	2.86
12220	13	0.010024	100.00	100.00	31000	12	0.299789	100.00	93.63
12230	13	0.001256	100.00	94.12	32000	12	0.02233	75.00	20.00
12300	13	0.000229	92.31	100.00	33000	12	0.000758	100.00	77.55
12400	13	0.001417	100.00	53.83	40000	12	0.00147	83.33	97.97
13100	13	0.002524	84.62	100.00	50000	12	0.012656	100.00	100.00

В таблицата класовете са обозначени само с кодовете си, с цел избягване на претоварване на представянето с още колони, но връзката може да се направи с помощта на документацията за картографиране (UA2012, 2016). Вижда се, че десет класа показват точност на потребителя 100%, докато 11 класа имат 100% точност на производителя. Минимална точност на производителя се получава за клас 2.4. Complex and mixed cultivation patterns, неправилно класифициран в една от пробите като клас 2.3. Pastures, който има висока стойност за теглото, поради относително голямата си площ. От друга страна, класът 2.3. Pastures показва минимална точност на потребителя 50%, поради объркване с няколко други класа, съдържащи растителност.

Проблеми, свързани със степента на непроницаемост в клас Изкуствени повърхности

Класовете от групата на 1.1. Urban Fabric, ниво 3 и ниво 4, са свързани с различна степен на непроницаемост на урбанизираните структури и понякога се класифицират погрешно един спрямо друг. Виждаме възможни причини във включването на ранна версия на HRL IMD2012, който слой, съгласно технологията за картографиране, участва в процеса на класификация на UA. Поради тези причини типичните грешки за клас 1.1.1.0 най-често са свързани с погрешни класификации към класове с по-нисък процент на почвено запечатване – 1.1.2.1 (Фиг. 3.24) и 1.1.2.2.



Фиг. 3.24. Клас 1.1.2.1 Discontinuous dense urban fabric (IMD 50% - 80%) погрешно класифициран като клас 1.1.1.0 Continuous urban fabric (IMD \geq 80%) (ляво изображение); Правилна класификация на клас 1.1.2.3 Непрекъснат урбанизирани структури с ниска плътност (S.L. 10% - 30%) със срещане на липсващи части - жълта стрелка (дясно изображение).

Проблеми, свързани с използването на спомагателни данни по време на визуална интерпретация

Ръководството за картографиране на (UA2012, 2016) посочва, че за групата класове 1. Artificial surfaces (Изкуствени повърхности): 1.2.3. Port areas (Пристанищни зони), 1.2.4. Airports (Летища) и 1.3.1. Mineral extraction and dump sites (Места за добив на минерали и сметища), се препоръчват допълнителни данни (кадастър и/или въздушни снимки) за идентифициране на административната граница на зоната на класа. На Фиг. 3.25 първият пример е свързан с клас 1.2.3 Port areas, а вторият случай е проблем с военно летище с пропуснати зони. Допускаме, че съответните спомагателни данни (кадастър и/или въздушни снимки) вероятно, не са били налични и не са били използвани към момента на очертаване на обекта.

В таблица 3.16, заедно с характеристики на продукта UA2012, са показани в обобщен вид главните количествени резултати от верификацията. Доверителният интервал на общата точност е сравнително голям, за преодоляването на което следва разпределението на броят проби по тясно да съответства на теглата на класовете. Общата точност покрива изискванията, а геометричните показатели са високо оценени.



Фиг. 3.25. Проблеми при дешифрирането, свързани със спомагателните данни. (1) Неправилен клас 1.2.3 Port areas - правилният клас е 1.2.1 Industrial, commercial, public, military and private units (ляво изображение); (2) Правилен код на класа 1.2.4 Airports с пропуснати части от територията, обозначени с червени стрелки (дясно изображение).

Таблица 3.16. Характеристики на UA2012.

Характеристика	Стойност
Площ в рамките на страната	2 719 508 ha
Брой на валидните класове в страната	26
Брой проверени проби	329
Вярност на кода LCLU	
Брой правилно интерпретирани	279
Обща претеглена точност	83.89%
Доверителен интервал на точността	± 8.83%
Геометрично съгласуване	
Вярност на очертаната зона	89.7%
Детайлност на очертаването	98.2%
Позиционна точност	97.3%

В процеса на верификация на набора от данни Атлас урбанизирани територии 2012 са идентифицирани редица проблеми и причинно-следствени връзки, систематизирани са типични грешки при определяне на тематичния клас и слабости в очертанятия на полигоните. Използването за верификацията на подходящи масиви от референтни данни допринася за постигане на целите на тази задача. Внимателната стратификация и планиране на събирането на пробите, изпълнени в съответствие с целта и наличните ресурси, способства за получаването на достатъчно прецизни оценки за тематична точност. Въз основа на получените количествени резултати може да се заключи, че продуктът Атлас урбанизирани територии 2012 притежава много добри характеристики на тематична точност, като общата му точност от 83,9% е над изискваните 80%. Показаната добра точност, богатата LCLU легенди и детайлната MMU правят продукта UA2012 от Локалния компонент ценен фактор за различни регионални и местни приложения.

Изводи към Глава 3

Прилагането на наборите от методи и алгоритми, предвидени в предложената Методика за оценяване на продукти на земното покритие, показва, че планираните и предвидени методични структури за оценяване са работещи, което може да бъде отнесено и към качествено-количествения подход, като цяло.

Резултатите от оценяването на HRL за горите показват добрата обща способност на автоматизирания анализ на разновременни сателитни изображения за извличане на данни за горското земно покритие. Въпреки регистрираните и описани грешки от пропуски и погрешна класификация, можем да заключим, че слоевете с висока разделителна способност Плътност на дървесното покритие и Преобладаващ дървесен тип показват последователно добри резултати.

Бинарният HRL за постоянните водни тела показва високи нива на тематичната точност - над 90%, което е резонно за този тип земно покритие. От своя страна, слойът на застрояването от 2013-2014 г., IMD2012 100m, се характеризира с високи нива на грешките от тип неправилна класификация (commission). Този резултат е потвърден чрез прилагане на разнообразни, взаимно-допълващи се количествени методи, проучени и технически внедрени за целите на изследването - оценяване като непрекъснат слой, като бинарен слой в две и три страти, чрез вероятност за приемане, както и чрез сравняване на базата на тематичната точност.

Продуктите от Локалния компонент са с високи точности на потребителя и производителя, като, както Атласът урбанизирани територии, така и Натура 2000 показват по-висока обща точност от изискваната в техните спецификации.

ЗАКЛЮЧЕНИЕ

Настоящото изследване бе мотивирано от актуалността на темата и проблема за оценяване на тематичната точност като елемент на качеството и важна характеристика на всеки набор от пространствени данни за земното покритие. Формулирането и реализирането в практически вид на вариант на комбиниран качествено-количествен подход за изследване на проблема, е и опит той да бъде обхванат в неговата комплексност.

Получените резултати от оценяването на актуални и разнообразни по характеристики продукти на земното покритие за територията на страната, от една страна дават важна и конкретна информация за изследваните продукти, а от друга - демонстрират на практика един комплексен инструментариум за практическо използване.

ПРИНОСИ

1. Формулиран е и е обоснован комбиниран, качествено-количествен подход за оценяване на тематичната точност на пространствени данни за земното покритие. Той включва качествен и количествен подход, като количествената част се състои от методи за извод, базиран на модел и извод, базиран на дизайн. Включеният в подхода Метод за визуално-качествено оценяване, позволява ефективно дефиниране на

тематичните и пространствени области за търсене и определяне на грешки на класификацията от системен характер.

2. Разработена и предложена е методика за оценяване на продукти на земното покритие, състояща се от четири блока - Масиви от входни данни, Предварителна обработка, Процедури за оценяване и Резултати.
3. Разработена е и е изпълнена практическа реализация на метод за обективно измерване на степента на непроницаемост в слой от непрекъснат тип.
4. Извършено е оценяване на тематичната точност на продукти за земното покритие за територията на цялата страна, някои за първи път, с използване на уникални масиви от *in situ* данни.

Публикации на автора, свързани с дисертационния труд

1. Dimitrov, V. Evaluation of Forest High Resolution Layers 2012 for Bulgaria. CD Proceedings XXIII-d International Symposium Modern Technologies, Education and Professional Practice in Geodesy and Related Fields, November 7 – 8, 2013, Sofia.
2. Dimitrov, V. Verification of forest and water high resolution layers 2012 for Bulgaria. Proceedings of Ninth scientific conference with international participation “Space, Ecology, Safety” (SES 2013), , Space Research and Technology Institute – Bulgarian Academy of Sciences, 2013, ISSN 1313 – 3888, pp. 562 - 567.
3. Димитров, В. Оценка на точността на тематичен слой, получен от спътникови изображения. Списание „Геодезия, картография, земеустройство“, 1-2'2016: 16–20., ISSN:0324-1610.
4. Димитров, В. Осигуряване и контрол на качеството на данни за земното покритие, получени от спътникови изображения. Proceedings Twelfth Scientific Conference with International Participation “Space, Ecology, Safety – SES 2016”, 2 and 3 November 2016, Sofia, pp. 192 - 198.
5. Димитров, В., Ю. Тепелиев, Р. Колева, Ю. Крумова, Н. Илиева, Т. Любенов, Б. Борисова. Мониторинг на земната повърхност по програмата „Коперник“ през 2017–2018 г. за България - основни резултати. Proceedings of Fifteenth International Scientific Conference Space Ecology Safety, SES'2019, Space Research and Technology Institute – Bulgarian Academy of Sciences, 2019, ISSN:p-ISSN 2603 – 3313, e-ISSN 2603 – 3321, 146-151.

Благодарности

Благодарен съм на моите ръководители, проф. Румен Недков - светла му памет! - и проф. Георги Желев, за ценните съвети и напътствия, както и на колегите ми от секция СДИ на ИКИТ, за оказаното съдействие и дружеска подкрепа в хода на работата и

написването на дисертацията! Благодарен съм също на колегите-партньори от ИАОС, за възможността да ползвам в изследванията данни от съвместни с Агенцията проекти!

Използвана литература

1. Аковецкий В. И. Дешифрирование снимков, М.: Недра, 1983. - 374 с.
2. Димитров, В., 2016а. Оценка на точността на тематичен слой, получен от спътникови изображения. Списание „Геодезия, картография, земеустройство“, 1-2'2016: 16–20., ISSN:0324-1610.
3. Димитров, В., 2016б, Осигуряване и контрол на качеството на данни за земното покритие, получени от спътникови изображения. Proceedings Twelfth Scientific Conference with International Participation “Space, Ecology, Safety – SES 2016”, 2 and 3 November 2016, Sofia, pp. 192 - 198.
4. Димитров, В., Ю. Тепелиев, Р. Колева, Ю. Крумова, Н. Илиева, Т. Любенов, Б. Борисова. Мониторинг на земната повърхност по програмата „Коперник“ през 2017–2018 г. за България - основни резултати. Proceedings of Fifteenth International Scientific Conference Space Ecology Safety, SES 2019, Space Research and Technology Institute – Bulgarian Academy of Sciences, 2019, ISSN:p-ISSN 2603 – 3313, e-ISSN 2603 – 3321, 146-151.
5. Желязков, Й., Ю. Тепелиев. Фотограмметрия и дистанционни методи, Мартилен, 1992, 250 с.
6. Калинов, К. Практическа статистика за археолози и антрополози, Нов български университет, София, София, 2002.
7. Aronoff, S. (1982 b) The map accuracy report: A user's view. Photogrammetric Engineering and Remote Sensing, 48(8):1309-1312.
8. Bolstad, P. GIS Fundamentals: A First Text on Geographic Information Systems . 5th Edition. Eider Press, White Bear Lake, MN. 2016, 769 p.
9. Bluman, A. Elementary statistics: a step by step approach. 7th ed., McGraw-Hill, 2009.
10. CLMS. Copernicus Land Monitoring Service, 2022. <https://land.copernicus.eu/> (посетен на 10.02.2022 г.).
11. Congalton, R. G. 1988. Using spatial autocorrelation analysis to explore errors in maps generated from remotely sensed data. Photogrammetric Engineering and Remote Sensing. Vol. 54, No. 5, pp. 587– 592.
12. Congalton, R., Thematic and Positional Accuracy Assessment of Digital Remotely Sensed Data, In: McRoberts, Ronald E.; Reams, Gregory A.; Van Deusen, Paul C.; McWilliams, William H., eds. 2007. Proceedings of the seventh annual forest inventory and analysis symposium. Gen. Tech. Report WO-77. Washington, DC: U.S. Department of Agriculture, Forest Service, pp. 149-154.
13. Congalton R. and K. Green. Assessing the Accuracy of Remotely Sensed Data: Principles and Practices, 3rd Edition, CRC Press, 2019, 328 p.
14. CORINE, 2000. CORINE land cover technical guide - Addendum 2000, Technical Report No 40, EEA, Copenhagen, 2000, 105 p.
15. Devillers, R., R. Jeansoulin, (eds.). Fundamentals of Spatial Data Quality, ISTE Ltd., 2006
16. Dimitrov, V., 2013b. Verification of forest and water high resolution layers 2012 for Bulgaria. Proceedings of Ninth scientific conference with international participation “Space, Ecology, Safety” (SES 2013), , Space Research and Technology Institute – Bulgarian Academy of Sciences, 2013, ISSN 1313 – 3888, pp. 562 - 567.

17. EEA, 2012. Guidelines for verification of high-resolution layers produced under GMES/Copernicus Initial Operations (GIO) Land Monitoring 2011–2013, 58 p.
18. EEA, 2018. Guidelines for verification of High Resolution Layers produced by the CLMS (Copernicus Land Monitoring Service) as part of the 2015 reference year production, Version 1.4. Copenhagen: European Environment Agency, 2018, 59 p.
19. Fleiss, J.L.; Levin, B.; Paik, M.C. *Statistical Methods for Rates and Proportions*, 3rd ed.; John Wiley & Sons: Hoboken, NJ, USA; 2003.
20. Foody G. M., Thematic map comparison: Evaluating the statistical significance of differences in classification accuracy, *Photogramm. Eng. Remote Sens.*, vol. 70, no. 5, pp. 627–633, May 2004.
21. Foody, G.M., 2009a. Sample size determination for image classification accuracy assessment and comparison. *Int. J. Remote Sens.* 30, 5273–5291, 2009.
22. Foody, G.M., 2009b. Classification accuracy comparison: hypothesis tests and the use of confidence intervals in evaluations of difference, equivalence and non-inferiority. *Remote Sens. Environ.* 113, 1658–1663.
23. GIO. GIO land High Resolution Layers (HRLs) – summary of product specifications - Prepared by: T. Langanke, EEA, 2016, 14 p.
24. HRL Imperviousness Degree 2012 Validation Report, Document: GIOVal_Lot1_Validation_Report, Issue: 1.7, Date: 06/01/2017, EEA, 57 p.
25. Janssen, L.L.F., van der Wel, F.J.M. Accuracy assessment of satellite derived land-cover data: a review. *Photogramm. Eng. Remote Sens.* 60, 1994, 419–426.
26. Jensen J. *Introductory digital image processing: a remote sensing perspective*. 4th edition. Pearson Education, Glenview, IL. 2016, 623 pp.
27. Langanke, T., M. Steidl, C. Schleicher; C. Sannier. Copernicus Land Monitoring Service – High Resolution Layer Imperviousness: Product Specifications Document, Copernicus team at EEA, 2016, 39 p.
28. Longley Paul A., Michael F. Goodchild, David J. Maguire, and David W Rhind (eds.). *Geographical Information Systems*, John Wiley & Sons, Inc., 1999, 1046 p.
29. Maucha, G., G Büttner. Recommendations. Quantitative assessment. High-resolution soil sealing layer, v. 2.1. EEA and European Topic Centre for Land Use and Spatial Information, 2008, 31 p.
30. Maucha, G., R. Pataki, R. Lehoczki, S. Kleeschulte, C. Schröder, D. Abdul Malak, C. Perger. Guidelines for verification of Local component products 2012, ETC/ULS report, EEA, 2017.
31. McCaffrey J. D. Validation vs. Verification, 2006, <https://jamesmccaffrey.wordpress.com/2006/04/28/validation-vs-verification/> , последно посетен на 21.09.2021.
32. Morisette, J., J. Privette, A. Strahler, P. Mayaux and C. Justice. Validation of Global Land-Cover Products by the Committee on Earth Observing Satellites. , In Ross S. Lunetta, John G. Lyon, (eds.), *Remote sensing and GIS accuracy assessment*, CRC Press, 2004, pp. 31-40.
33. Olofsson, P., Foody, G.M., Stehman, S.V., Woodcock, C.E. Making better use of accuracy data in land change studies: estimating accuracy and area and quantifying uncertainty using stratified estimation. *Remote Sens. Environ.* 129, 2013, 122–131.
34. Pham, H. *Software Reliability*. John Wiley & Sons, Inc. 1999, p. 567.
35. Rossiter D. G. Technical Note: Statistical methods for accuracy assesment of classified thematic maps. University of Twente, ITC, Enschede, Netherlands, 2014.
36. Salk, C., C. Dresel, S. Fritz, L. See, I. McCallum. An exploration of some pitfalls of map accuracy assessment using the new Map Tools resource. *Remote Sensing*, 2018.

37. Stehman, S. 1999b. Comparing thematic maps based on map value, *Int. J. Remote Sens.*, vol. 20, no. 12, pp. 2347–2366, Jan. 1999.
38. Stehman, S.V., Czaplewski, R.L. Design and analysis for thematic map accuracy assessment: fundamental principles. *Remote Sens. Environ.* 64, 1998, pp. 331–344.
39. Stehman S. V., G. M. Foody. Key issues in rigorous accuracy assessment of land cover products, *Remote Sensing of Environment* 231, 2019, Article 111199.
40. Tso, B., P. Mather. *Classification Methods for Remotely Sensed Data*. 2nd ed. CRC press, 2009.
41. UA2012. Mapping Guide v4.7 for a European Urban Atlas 2012, EC DG Regional and Urban Policy, European Union, 2016. <https://land.copernicus.eu/user-corner/technical-library/urban-atlas-2012-mapping-guide-new>
42. Yang X, Blower JD, Bastin L, Lush V, Zabala A, Masy J, Cornford D, DHAZ P, Lumsden J. An integrated view of data quality in Earth observation. *Phil Trans R Soc A* 371: 20120072. 2013. <http://dx.doi.org/10.1098/rsta.2012.0072>.