



Българска академия на науките



Институт за космически изследвания и технологии

Илина Боянова Каменова

Моделиране и картографиране на биофизични и биохимични параметри на посеви от зимна пшеница на базата на сателитни изображения

АВТОРЕФЕРАТ

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

Научен ръководител:

Доц. д-р Петър Димитров

София • 2022

Дисертационният труд е разработен в рамките на задочна докторантура в Секция „Дистанционни изследвания и ГИС“ към Институт за космически изследвания и технологии (ИКИТ) на Българска академия на науките (БАН). Той е обсъден на заседание на разширен Научен семинар на секция „Дистанционни изследвания и ГИС“ на ИКИТ - БАН, проведено на 28.07.2022 г., и е насочен за защита пред специализирано жури в състав:

1. Доц. д-р. Петър Димитров – ИКИТ – Българска академия на науките
2. Проф. д-р. Лъчезар Филчев – ИКИТ – Българска академия на науките
3. Доц. д-р Милена Керчева – ИПАЗР “Никола Пушкарров” – Селскостопанска академия
4. Проф. д-р инж. Мартин Банов – ИПАЗР “Никола Пушкарров” – Селскостопанска академия
5. Доц. д-р Георги Железов – НИГГГ – Българска академия на науките

Дисертационният труд е в обем от 150 страници, включващи 22 таблици, 49 фигури, библиография от 150 заглавия, от които 11 на кирилица и 139 на латиница, както и 4 приложения (хистограми и диаграми, таблици с резултати от валидирането по фенофази на общите модели, диаграми за диагностика на регресионните модели, карти на биофизични показатели)

Работата има следната структура: Въведение – 4 стр., Глава 1. Основни теоретични и методични въпроси – 19 стр., Глава 2. Материали и методи на изследването – 31 стр., Глава 3. Резултати от изследването – 55 стр., Заключение – 3 стр., Приноси и списък на публикациите свързани с дисертацията и благодарности – 2 стр., Литература – 11 стр., Приложения – 25 стр.

Материалите по защитата са публикувани в интернет на страницата на ИКИТ - БАН (<http://www.space.bas.bg>) и са на разположение на интересуващите се в канцеларията на института - ст. 413, ет. IV бл.1, ул. Акад. Г. Бончев.

ББП – Биофизични и биохимични параметри

ВИ – Вегетационен индекс

ГИС – Географски информационни системи

ДИ – Дистанционни изследвания

ИСАК – Интегрирана система за администриране и контрол

ССС – Canopy Chlorophyll Content (Общо хлорофилно съдържание)

CGMS – Crop Growth Monitoring System (Системата за мониторинг на растежа на посевите)

CSDM – Canopy structural dynamic model (Модел на динамиката на структурата на растителността)

fAPAR – Fraction of Absorbed Photosynthetically Active Radiation (Дял на погълнатата фотосинтетична активна радиация)

fCover – Fraction of vegetation Cover (Проективно площно покритие)

LAI – Leaf Area Index (Индекс на листната повърхност)

LOOCV – метода на кръстосано валидиране с един независим отчет

MXL – Maximum Likelihood (Метод на максималното правдоподобие)

NIR – Near Infrared (Близко инфрачервен спектрален диапазон)

RE – Red Edge (Червен ръб)

RF – Random Forest (Метод на гора от дървета на решенията)

SVM – Support Vector Machines (Метод на опорните вектори)

Въведение

Земеделieto е един от отраслите с най-голямо влияние върху земеползването в световен мащаб, в момента около 1.2 – 1.5 милиарда хектара са заети от земеделски култури (FAO, 2021). За да се отговори на прогнозирания растеж на човешкото население и нуждата от продукция на хранителен ресурс, историческите темпове на увеличение в производството трябва да продължат (Godfray et al., 2010). Трябва да отбележим че земеделieto е основен фактор за промените на климата и замърсяването на околната среда, но също така е и най-уязвимият икономически сектор от самите промени на климата (Agovino et al., 2019).

Уязвимостта на земеделieto дава сериозни последиствия върху количеството и качеството на земеделска продукция, което подчертава необходимостта от разработване и налагане на по-устойчиви земеделски практики и методи (Tilman et al., 2002). В този контекст Европейският Съюз ЕС показва силен ангажимент като залага на политики, които в съответствие с Парижкото споразумение, налагат реформи за всички икономически сектори (Horowitz, 2016), което ще наложи промени и в земеделските практики. Devos et al., (2018) представя концептуална рамка за мониторинг, базираща се на регулярно и систематично наблюдение и оценка на земеделските дейности, като основна роля в този подход играят автоматизирани процедури за извличане на информация от сателитите Sentinel.

Редица изследвания подчертават ползата от използване на сателитни многоканални данни за: картографиране на площи заети със земеделски култури (Jin et al., 2019; Phiri et al., 2020); за наблюдение на биофизични параметри на посева (Kamenova et al., 2017; Toscano, 2020; Verrelst et al., 2015; Wiegand et al., 1991); за прогнозиране на добиви (Han et al., 2020); за наблюдение на екосистемни услуги и консервация на биоразнообразие (Cord et al., 2017; Dantas de Paula et al., 2019). Сред множеството налични източници на данни, ЕС предоставя достъп до безплатни сателитни данни от програмата Коперник и постепенно свързва политиките си с използването на този ресурс. За да се генерират продукти с площна информация за оценка на състоянието на земеделските култури е нужно количествено да се оценят растителни параметри, които са информативни за развитието на културите.

Такива параметри са: Индекса на листната повърхност (LAI), биомасата, общо съдържание на хлорофил в растителната покривка (CCC), дял на погълнатата фотосинтетично-активна радиация (fAPAR), проективно площно покритие (fCover) и др. Тези параметри са определени от системата за наблюдения на климата (GCOS) като съществени климатични променливи (ECVs). В България има разработена и патентована методика за оценка на състоянието на посеви на земеделски култури базирана на биофизични параметри (Е. Руменина et al.,

2019). В Европейският Съюз, зимната пшеница (*Triticum aestivum* L.) е с най-голяма засята площ в, 22 милиона хектара (Eurostat, 2021). В България зимната пшеница е традиционна култура с дълги традиции в отглеждането и заетите площи са 1.2 милиона хектара (Eurostat, 2021). Зимната пшеница е предимно конвенционално отглеждана култура с влагане на азотни торове, хербициди и други препарати, за осигуряване на по-голяма продукция.

Поради сериозният екологичен натиск, който земеделието и по-конкретно конвенционалното отглеждане на житни култури прилага, е нужно да се прилагат методи, осигуряващи по-ефективното и отглеждане (Shewry, 2009). Настоящата дисертация разглежда приложението на сателитни изображения за осигуряване на пространствено обвързана, площно (т.е. картографски) представена информация за нуждите на производството на зимна пшеница, акцентирайки върху три основни аспекта: разпространение на посевите, оценка на важни техни биофизични параметри и проследяване на динамиката им във времето.

Обект и предмет на изследването

Обект на изследване в дисертацията са посевите от зимна пшеница в два избрани тестови участъка на територията на България. Тестовите участъци се намират в землищата на град Кнежа и град София, квартал Требиш. Интересът към тази култура е продиктуван от значителното площно разпространение на пшеницата, което прави икономическото и екологично значение на тази култура за територията на България изключително голямо.

Предмет на изследването са комплекс от биофизични и биохимични параметри, съответно: индекс на листната повърхност (LAI), дял на погълнатата фотосинтетично активна радиация (fAPAR), проективно площно покритие (fCover) и общо съдържание на хлорофил в растителната покривка (ССС), както и техните пространствена вариация и времева динамика. Тези параметри могат да се използват за характеризирание на физиологичното състояние на посевите във своята връзка със спектралните отражателни характеристики регистрирани от сателитни сензори.

Цел и задачи на изследването

Целта на изследването е да се моделират и картографират пространствената и времева динамика на набор от биофизични и биохимични параметри на посеви от зимна пшеница на базата на многоканални сателитни изображения. За да бъде постигната тази цел е необходимо решаването на следните научни задачи:

1. Да се оценят възможностите за разпознаване на посеви от зимна пшеница в различни периоди на вегетационния сезон чрез класифициране на сателитни изображения и техни продукти по различни класификационни методи.
2. Да се състави карта на полетата със зимна пшеница , на база на най-ефективния класификационен метод.
3. Да се планират и проведат полеви кампании за събиране на наземни данни за биофизични и биохимични параметри в посеви от зимна пшеница.
4. Да се създадат и валидират регресионни модели за определяне на биофизични и биохимични параметри на посеви от зимна пшеница посредством многоканални сателитни изображения.
5. Да се направи сравнение между общите регресионни модели и тези разработени за отделните сортове и фенофази и да се оценят техните относителни предимства.
6. Да се предложи и тества подход за моделиране на ежедневната динамика на биофизичните параметри в течение на вегетационния сезон върху определена територия.

1 Основни теоретични и методични въпроси

1.1 Мониторинг на биофизични и биохимични параметри на посева

В земеделието и по конкретно в растениевъдството, системите за мониторинг, дават възможност за наблюдение и анализ на възлови показатели за състоянието на посевите и се стремят да облагодетелстват по ефективното и устойчиво производство на растителни култури. Понастоящем много държави и световни организации използват системи за мониторинг на посевите, за да наблюдават и прогнозираят развитието и добивите в национален, регионален или глобален мащаб. Мониторингът в селското стопанство на регионално и национално ниво е в приложение от десетилетия: напр. Глобалната система за информация и ранно предупреждение (GIEWS) на ООН; Системата за ранно предупреждение за опасност от недостиг на храни (FEWS NET) на Агенцията за международно развитие на САЩ (USAID); CropWatch в Китай; Системата за мониторинг на земеделието с дистанционни изследвания (MARS) на Европейската комисия (ЕК).

Навременната и пространствено точна информация за биофизичните и биохимичните параметри (ББП) на посевите, като например: индекс на листната повърхност (от англ. Leaf Area Index, съкр. LAI), Фотосинтетична активна радиация (от англ. Fraction of Absorbed Photosynthetically Active Radiation, съкр. fAPAR), проективно площно покритие (от англ. Fraction of vegetation Cover, съкр. fCover) и общо съдържание на хлорофил в растителната покривка (от англ. Chlorophyll

Canopy Content, съкр. CCC) са важни фактори за определяне на общото фитосанитарно състояние на посевите и съответно за взимане на навременни управленски решения от агрономите. С помощта на съвременните технологии като: географски информационни системи (ГИС) и дистанционни изследвания (ДИ) пространствената информация за БП се използва ефективно в прецизното земеделие, където се изисква картографиране на даден параметър за цялата площ, засята с определена култура.

1.2 Биофизични и биохимични параметри (ББП)

Индексът на листната повърхност (LAI) е мярка за количеството листа и тяхната площ в растителната покривка и е често използвана величина в изследванията на растителността, агрономическите и екосистемните изследвания. Понятието индекс на листната повърхност е въведено за първи път от Watson, (1937) той го определя като, общата листна повърхност на единица земна площ и следователно представлява безразмерна величина. Дял на погълнатата фотосинтетично активна радиация (fAPAR) е дялът на слънчевата радиация в спектралната област на фотосинтетично активната радиация PAR (400 - 700 nm), която се поглъща от растителните организми за нуждите на процеса фотосинтеза. Количеството слънчева енергия, погълната от растението за процеса на фотосинтеза, се определя като абсорбирана фотосинтетично активна радиация APAR, разделена на PAR, абсорбиран от растението, изразен като фракция е fAPAR. fAPAR е важна биофизична характеристика в моделите, оценяващи първичната продуктивност на растителността и в моделите на въглеродния цикъл между земният слой и атмосферата (Gitelson et al., 2005). Проективно площно покритие (fCover) е процентът почвена повърхност, покрита със зелена растителност. Това е важна мярка за установяване ранното развитие на културите и тяхната жизненост. Тази променлива има отношение към количеството погълната слънчевата енергия от растителната покривка и по този начин описва тяхната потенциална производителност (Pask et al., 2012). Фотосинтезата е основният двигател на глобалният въглероден цикъл, тясно свързан с климата, чрез множество процеси и обратна връзка (Sellers et al., 2018). Основна роля в процеса фотосинтезата играят растителните пигменти, изпълняващи ролята на първични получатели на светлинните кванти и осъществяващи по-нататъшно им превръщане в химичната енергия. Количеството хлорофил в клетката може да даде непряка оценка за физиологичното състояние на растенията и освен това съдържанието на хлорофил в листата е тясно свързано със стреса и стареенето на растенията (Merzlyak et al., 1999). Хлорофилните съединения могат да дадат косвена оценка за статуса на нутриентите, тъй като част от азотното съдържание в листата е именно в хлорофилните съединения.

1.2 Стратегии за получаване на информация от многоканални сателитни изображения

1.2.1 Класификация с обучение

Общата цел на процедурите за класификация на сателитни изображения е автоматично да се категоризират всички пиксели от изображението в отделни класове земно покритие. Терминът спектрална класификация се отнася до семейството на класификационни процедури, които използват т. нар. „pixel-by-pixel“ подход, при който спектралната информация на индивидуалният пиксел е основа за автоматизирана класификация на земното покритие. Сред най-популярните класификационни методи с обучение са: метода на максималното правдоподобие (MXL), класификаторите служещи си с невронни мрежи и непараметричните класификатори, като метода на опорните вектори (SVM) и метод с гора от дървета на решенията (RF).

1.2.2 Вегетационни индекси

Вегетационните индекси (ВИ) са математически преобразуваните стойности на оригиналната спектрална отразителна информация на заснетите обекти. Вегетационните индекси са предназначени да намалят адитивните и мултипликативни грешки, свързани с атмосферните ефекти, слънчевото осветление, фоновите ефекти на почвата и топологичните грешки (Huete, 1988). ВИ са често използван метод в дистанционните изследвания (ДИ) за установяване на корелация между сателитни многоканални изображения и „*in situ*“ данни за параметри на посевите. Разработени са голям брой ВИ които се използват за прогнозиране на LAI и fAPAR. Класическите широко канални ВИ, използващи канал в червения и близко инфрачервен (NIR) спектрален регион, са най-често използвани, но когато LAI се увеличи тези ВИ, стават нечувствителни за гъсти растителни покривки (Haboudane et al., 2004). С хлорофилното съдържание добре се корелират индексите които съдържат канали от спектралният диапазон на червеният ръб (RE) (Clevers & Kooistra, 2012). Поради силната корелация между хлорофил и азот, тези индекси са подходящи и за оценка на азотното съдържание.

1.2.3 Характеризиране на динамиката на биофизичните параметри на посевите

Посевите от земеделски култури се намират в процес на непрекъсната промяна в течение на вегетационния период, обусловена от натрупването на биомаса, фенологичното развитие, въздействието на метеорологични и биотични фактори и агротехническите мероприятия. Сателитните дистанционни наблюдения предоставят регулярни данни позволяващи да се характеризира тази динамика при това върху големи по площ територии. Апаратурни комплекси като Sentinel-3/OLCI, MODIS, PROBA-V и др. предоставят данни с времева разделителна способност от порядъка на 1-2 дни. Тези данни се използват за изготвяне в близко

до реалното време, на разнообразие от продукти (БИ, LAI, Хлорофилно съдържание и др.). Времеви серии от тези продукти намират широко приложение в изследванията на динамиката на растителността, включително в земеделските територии. Основен проблем при приложението на времеви серии от спътникови данни е, че в тях обикновено има липсващи данни поради облачната покривка и др. влияния. Baret (1986), демонстрира подход за възстановяване (интерполиране) на липсващи данни във времева. Методът е наречен CSDM (Canopy Structural Dynamic Model) и е приложен за интерполиране на стойностите на LAI във времето. Методът е използван за възстановяване на времеви серии от пшеница (Duveiller et al., 2011), за посеви от царевица (Koetz et al., 2005), също и за посеви от рапица (Jiao et al., 2021).

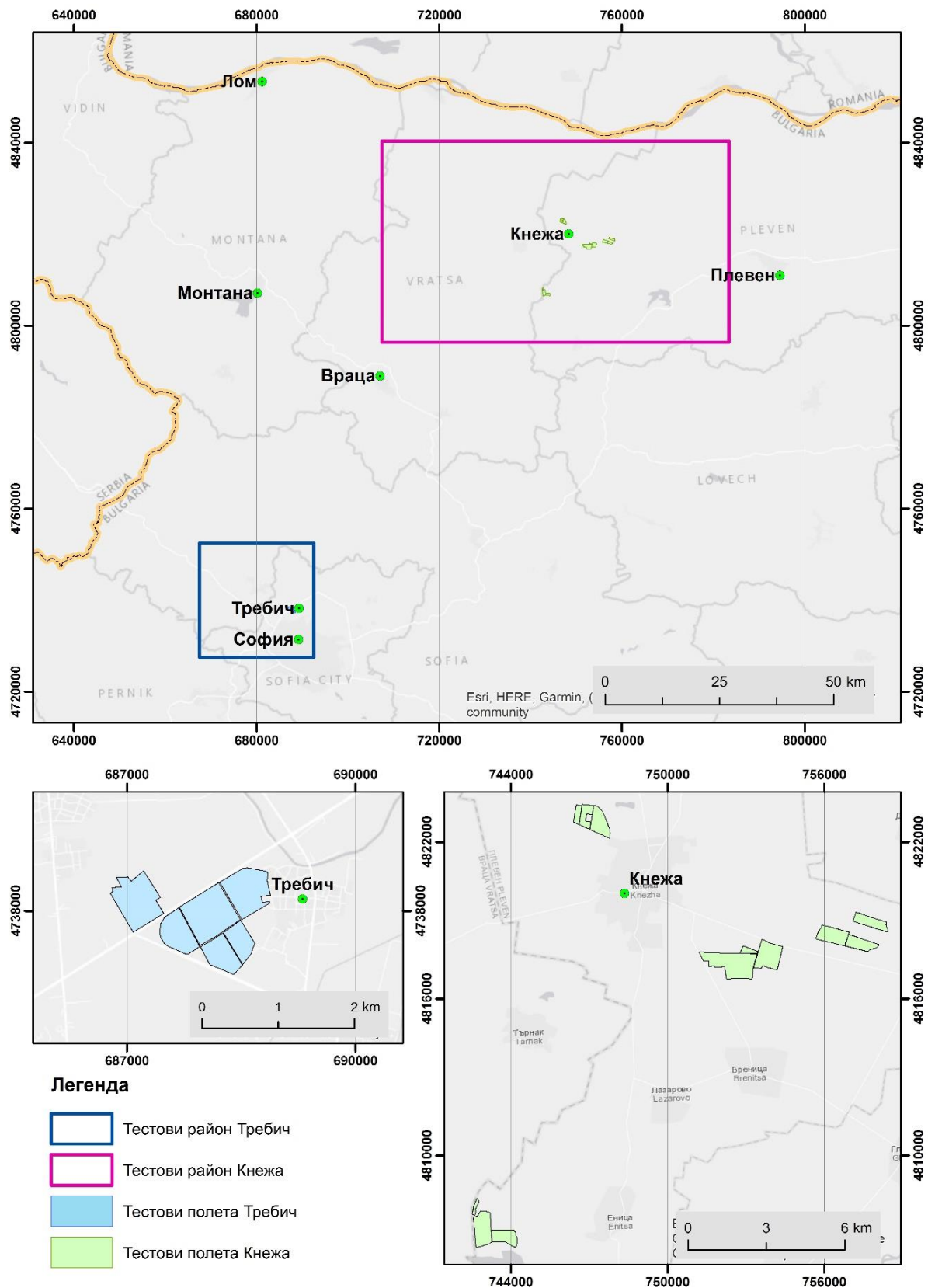
Изводи

Наблюдението по дистанционни данни има редица предимства пред традиционните полеви методи за определяне на ББП на посевите преодолявайки трудностите за събиране на информация за големи райони. Съществуващите карти, особено тези които покриват големи райони са съставени основно от статистическа информация на ниво държава. За България тази информация е трудно достъпна и с недоказана надеждност. Има разработени системи за мониторинг в земеделието по сателитни данни, които са наложени и са с добра прогнозна точност, като системата MARS Crop Yield Forecasting System (MCYFS). Тези системи, обаче дават твърде обобщена информация за ББП и редица локални тенденции са пренебрегнати, защото системите са унифицирани, за да са общо валидни. Затова ние съставяме методика за наблюдение на полета със зимна пшеница за България. Тази методика е валидирана за местните условия, сортове и цикъл на развитие.

2. Материали и методи на изследването

2.1. Тестови райони

Два тестови района са изследвани в настоящата дисертационна работа, район „Кнежа“ и район „Требич“ (Фигура 1). Тези два района са част от аерокосмическите полигони на територията на България, на чиято територия са проведени редица подспътникови експерименти (Roumenina et al., 2009; Е. Руменина & Желев, 2020). В рамките на тези два района се намират тестови полета в които са събрани наземни данни за изследваните ББП на зимна пшеница в рамките на няколко вегетационни сезона.



Фигура 1 Карта на тестовите райони и тестовите полета

2.2. Полеви данни

Инструментите използвани за събирането на полевите данните са:

- AccuPAR PAR/LAI Ceptometer, модел LP-80, Decagon Devices

Уредът измерва контактно, прихванатото слънчево лъчение от растителната покривка и се използва ръчно в полеви условия. Състои се от 80 независими сензора, които са на 1 cm разстояние един от друг, всеки сензор измерва PAR (400 nm-700 nm), мерната единица показвана от уреда е $\mu\text{mol m}^{-2}\text{s}^{-1}$. От PAR се изчисляват LAI, fAPAR и fCover посредством уравнения (DecagonDevices, 2014).

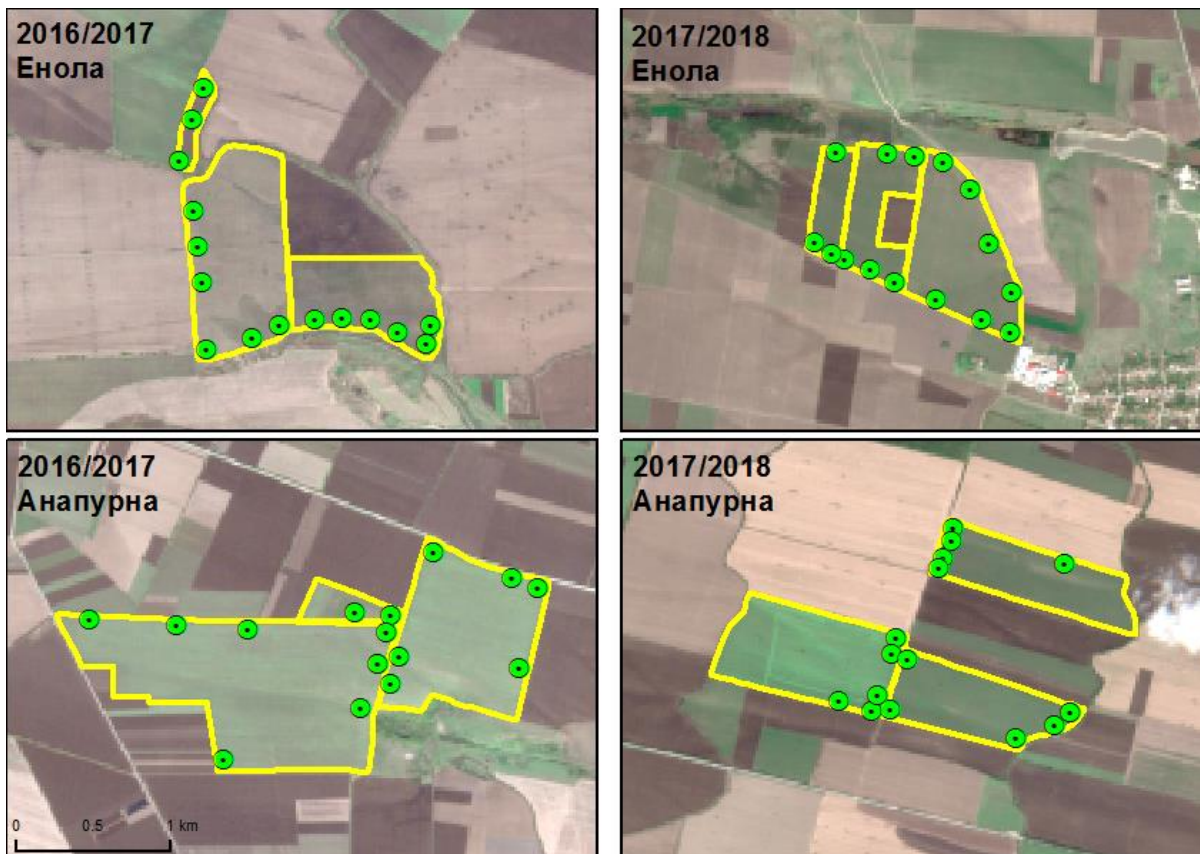
- CCC-300, Opti Sciences

Уредът представлява контактен уред за измерване на хлорофилна концентрация (CC) в растителен лист, мерната единица показвана от уреда е mg/m^2 . Измерването се извършва посредством оптичен кабел който се закрепя към растителният лист с щипка. Данните се преизчисляват от хлорофилна концентрация на площ растителен лист (CC; mg/m^2) в общо количество хлорофил в растителната покривка (CCC; g/m^2), посредством LAI (Kamenova et al., 2018).

Полевите данни за тестови район „Кнежа“ са събрани по проект “Тестване на вегетационни индекси от Sentinel-2 за оценка на състоянието на зимни култури в България (TS2AgroBg)” финансиран от ЕКА. В рамките на този проект са осъществени 6 полеви кампании, за две селскостопански години, 2016-2017 и 2017-2018 г. На Фигура 2 са представени тестовите полета в които са в направени наземни измервания. Тестовите площадки в които е извършено пробовземането са 30 на брой, но не при всяка полева кампания бе възможен достъпа до всяка една от тях (Таблица 1). Големината на всяка тестова площадка е 20 x 20 m, размер който е възприет, за да съвпада с пространствената разделителна способност на спектралните канали от Sentinel-2. Тестваните полета са засети със сортове „Енола“ и „Анапурна“.

Таблица 1 Полеви кампании за пробовземане в района „Кнежа“

Номер на полева кампания	Дата на пробовземане	Фенологична фаза на пшеницата (по Zadoks et al. 1974)	Брой на тестовите площадки
1	7-11 ноември 2016	Братене (Z20)	20
2	20-24 март 2017	Братене (Z21-26)	15
3	24-28 април 2017	Вретенене (Z31-34)	30
4	15-19 май 2017	Цъфтеж (Z65-69)	28
5	6-11 ноември 2017	Братене (Z20-23)	28
6	2-5 април 2018	Вретенене (Z30-31)	30



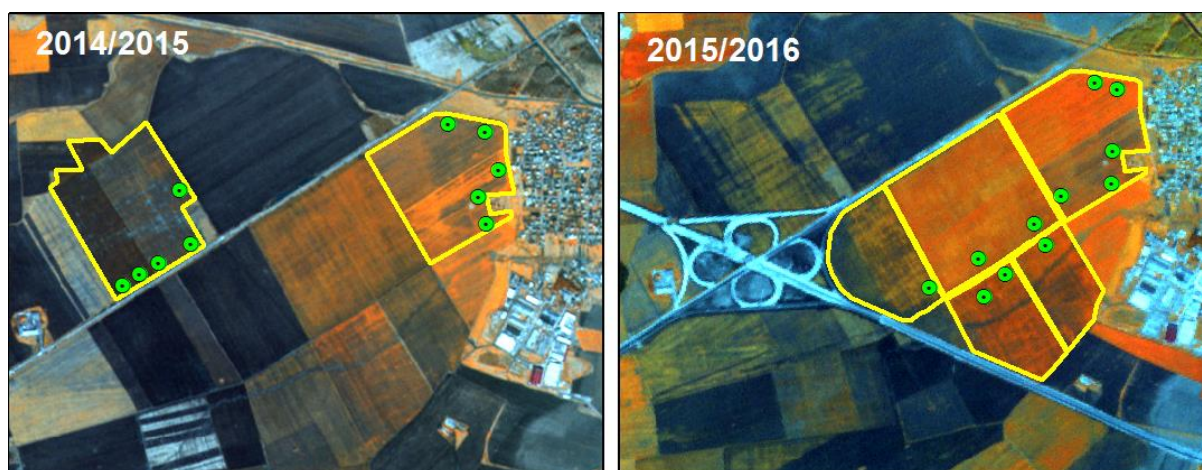
Фигура 2 Наземни данни за ББП, район „Кнежа“

Полевите данни за тестови район „Требич“, са събрани в рамките на 2 селскостопански години, 2014-2015 и 2015-2016 г. Броят на тестовите площадки е 10 за първата година и 11 за втората (Фигура 3). Размера на всяка тестова площадка е 5 x 5 m, като размера отговаря на пространствената разделителна способност на сателита RapidEye. Датите на пробовземания са съобразени с фенологичните фази на пшеницата (Таблица 2), но също така и с датите на заснемания на сателитни изображения. На тестовите полетата се отглежда сорт „Енола“.

Биохимичният параметър ССС е изследван и дискутиран в рамките на настоящето изследване само за тестови район „Требич“. Подобен и подробен статистически анализ на ССС е извършен и за тестовият район „Кнежа“, но в рамките на друго изследване, резултатите са публикувани от Dimitrov et al. (2019).

Таблица 2 Полеви кампании за пробовземане, район "Требич"

Номер на полева кампания	Дата на полева кампания	Фенологична фаза на пшеницата
1	02 април 2015	братене
2	15 април 2015	вретенене
3	04 май 2015	вретенене
4	15 май 2015	вретенене/изкласяване
5	20 май 2015	изкласяване/цъфтеж
6	13 април 2016	вретенене
7	18 май 2016	изкласяване



Фигура 3 Наземни данни за ББП, район "Требич"

2.3 Сателитни данни

Многоканалните сателитни изображения от Sentinel-2, които са използвани в настоящето изследване не са атмосферно коригирани, те са продукт Level-1C и са атмосферно коригирани в модула Sen2Cor в софтуера SNAP. Изображенията които отговарят на критерия за безоблачност (под 10%) и съвпадат приблизително във времето с датата на провеждане на полевите кампании са общо 4. За останалите полеви кампании не са налични изображения с достатъчно добро качество, затова посредством данните от полеви спектрометър FieldSpec4 са симулирани изображения съответстващи на изображения от Sentinel-2. Тези изображения (Таблица 3) са използвани за регресионно моделиране на ББП.

За класификациите на полета със зимна пшеница за селскостопанската 2017-2018 г., 8 изображения отговарят на критерия за безоблачност. От наличните

многоканални сателитни изображения са създадени и разновременни изображения, които включват данни за каналите от няколко дати. Съставено е едно изображение маска, съдържащо информацията за облачните и повредени пиксели. Тази маска е приложена върху изображенията от всички отделни дати и разновременните изображения. Друг подход, който е възприет за отстраняване на облачна покривка, сенки и други не съдържащи информация пиксели е съставяне на композитни изображения от Sentinel-2. Композитните изображения по своята същност също представляват разновременни продукт който обхваща определен период от време. При този подход, за всеки пиксел се изчислява средна стойност на база на изображенията налични за този период. Средната стойност която е взета в случая е медианата на стойностите за периода. Композитните изображения са съставени в Google Earth Engine. Общо 13 изображения, разновременни изображения и композитни изображения са използвани за класификациите (Таблица 4).

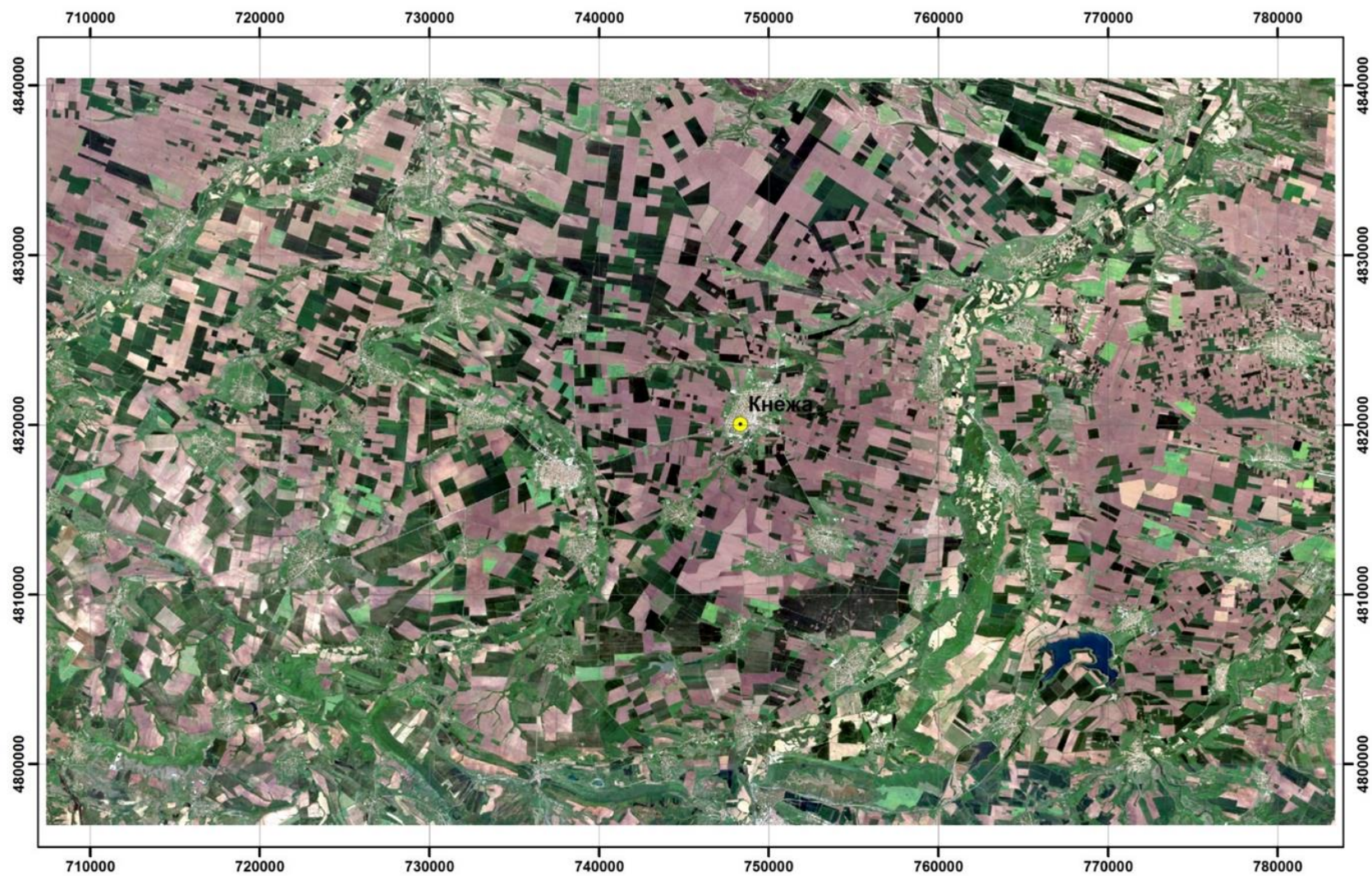
Таблица 3 Източник на спектрални данни и дата на заснемане в съответствие с кампаниите за пробовземане в тестови район „Кнежа“

Източник на спектрални данни	Дата на заснемане/измерване	Номер на полева кампания
Sentinel-2 A	31 октомври 2016	1
Полеви спектрометър ASD	20-24 март 2017	2
Sentinel-2 A	19 април 2017	3
Полеви спектрометър ASD	15-19 май 2017	4
Sentinel-2 B	31 октомври 2017	5
Sentinel-2 B	30 март 2018	6

Таблица 4 Източник на спектрални данни и дата на заснемане на изображенията използвани за процеса на класификация

№	Дати и периоди на заснемане на изображенията
1	31.10.17
2	25.12.17
3	30.03.18
4	09.04.18
5	14.04.18
6	24.04.18
7	29.04.18
8	08.06.18
9	31.10.17; 25.12.17
10	31.10.17; 25.12.17; 30.03.18
11	31.10.17; 25.12.17; 30.03.18; 14.04.18
12	31.10.17; 25.12.17; 30.03.18; 14.04.18; 29.04.18
13	31.10.17; 25.12.17; 30.03.18; 14.04.18; 29.04.18; 08.06.18

Композитно изображение от "Sentinel-2" за периода 01-04-2018 ; 31-05-2018



Фигура 4 Композитно изображение от Sentinel-2

2.4 Агrometeorологични данни за района на Кнежа

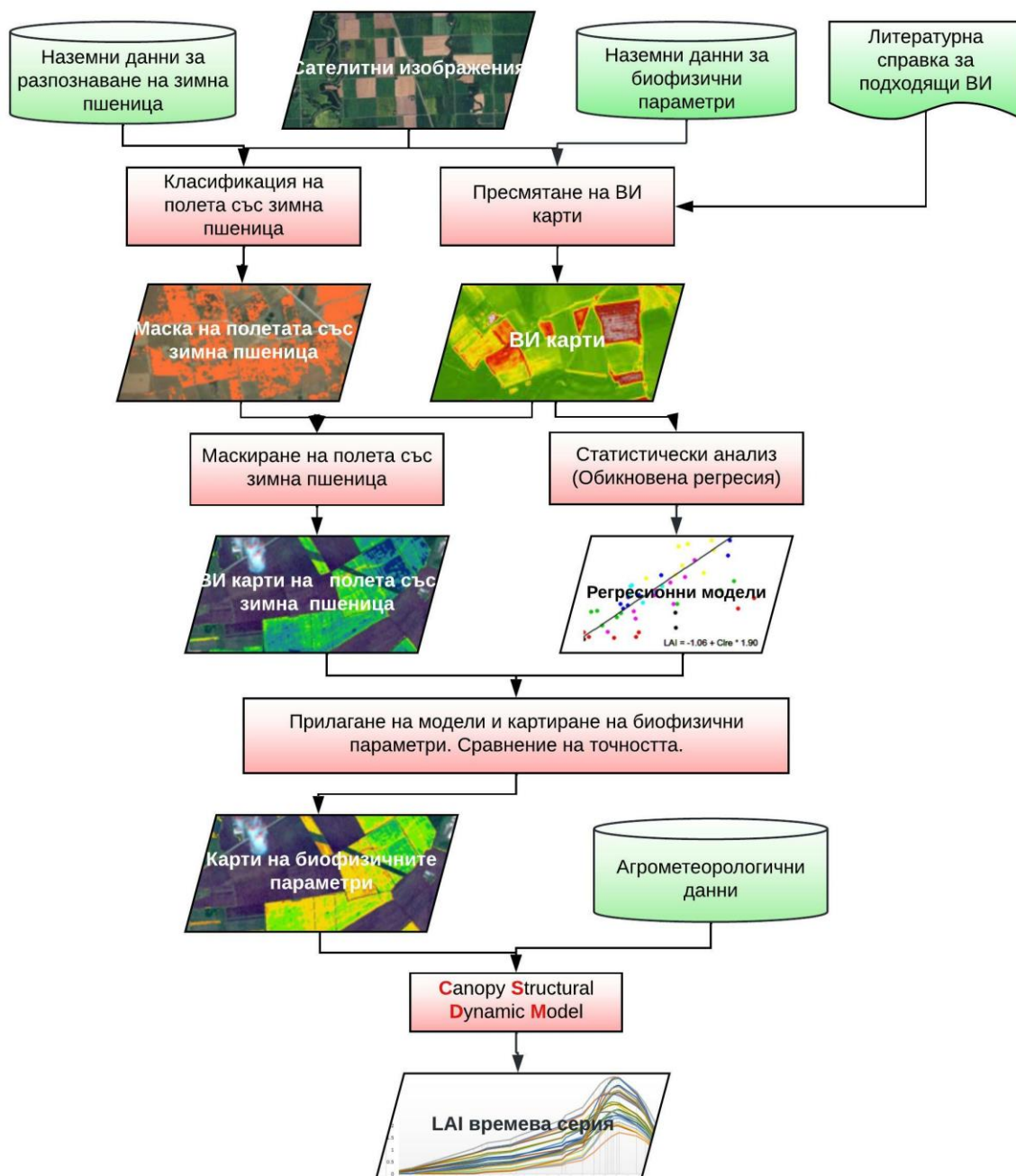
Базата данни на системата за мониторинг на растежа (CGMS), подържана от JRC съдържа информация за метеорологични параметри от климатични станции, като данните са интерполирани в грид от 25x25 km. Метеорологичните данни са достъпни от 1975 г., на дневна база и обхващат държавите членки на ЕС, съседни европейски държави и средиземноморските държави. Променливите които са събрани включват температура на въздуха, количество валежи, изпарение, влажност на въздуха, скорост на вятъра и други. Всички входящи данни се проверяват за грешки и съмнителни стойности, като напр. температури на въздуха извън реалистични граници. Променливите, получени от метеорологичните станции, имат неравномерно разпределение в пространството. Данните от една метеорологична станция са представителни само за местоположението на тази станция. Данните са интерполирани от системата за да се получи равномерен грид с метеорологични данни.

2.5 Референтни данни от системата ИСАК

За обучение на класификационните алгоритми са необходими обучаващи множества. За нуждите на настоящето изследване обучаващи данни са набавени от „Интегрирана система за администриране и контрол (ИСАК)“. Наборът от данни от ИСАК представлява ежегодно актуализирана векторна база данни за земеделските култури. Той е генериран и подържан от българското „Министерство на земеделието, храните и горите“. ИСАК съдържа и информация за границите на земеделските парцели (обработваеми полета, пасища и трайни насаждения), придружена от информация за типа култура или земно покритие във всеки парцел, съгласно декларациите подадени от фермерите. Тази база данни се поддържа на национално ниво и е инструмент за контрол на субсидиите, в рамките на Общата селскостопанска политика на Европейският съюз (ОСП).

2.6 Обработка и анализ на данните

Обобщена схема на методиката за обработка и анализ на данните и последователността на работа е представена на Фигура 5. В следващите подраздели са представени подробно отделните етапи.



Фигура 5 Диаграма на работния процес

2.6.1 Класификации на полета със зимна пшеница

За решаване на поставените научни Задачи 1 и Задача 2, са приложени и оценени следните класифициращи алгоритми: метод на опорните вектори (SVM), метод на гора от дървета на решенията (RF), метод на максималното правдоподобие (MXL). И трите класификационни метода са използвани за класифициране и картографиране на полетата със зимна пшеница и в двата района на изследване. Метод на опорните вектори (от англ. Support vector

machines, съкр. SVM) представлява група от непараметрични, управлявани алгоритми, които са приложими както към класификационни, така и към регресионни процедури. Теоретичните основи и математическата формулировка на този метод са предложени от (Varnik, 1995).

Метод на гора от дървета на (от англ. Random forest, съкр. RF) разпределя много сложен проблем при класификацията на множество отделни етапи, от прости процеси на вземане на решения (Belgiu & Drăgu, 2016). Алгоритъмът RF създава произволна извадка от множество дървета на решенията и ги обединява заедно, за да получи по стабилна и точна прогноза чрез кръстосано валидиране. Метод на максималното правдоподобие (MXL) представлява параметричен класификатор, основан на статистическа теория.

За изпълнение на заложените научни задачи трябва да се картографират полетата засети със зимна пшеница в рамките на тестовите райони, посредством трите класификационни методи. В класификационната схема има два класа: клас „Зимна пшеница“ и клас „Други култури и почви“. За съставяне на обучителни множества е използван слой от системата „Интегрирана система за администриране и контрол (ИСАК)“ съдържащ пространствената и атрибутивна информация за земеделските площи. От този слой са взети полигоните които са заети със селскостопански полета за периода 2017-2018 г. в тестовите райони. Според класификационната схема, която съдържа два класа, са изведени като една група полигоните с атрибутивни данни за „Мека зимна пшеница“. Неземеделските райони са изрязани и за двата тестови района.

За оценката на точността на класификационните методи приложени върху изображенията от всички отделни дати, разновремените изображения и композитни изображения е подготвен отделен слой, отново базиран на базата данни ИСАК. На базата на матрицата на грешките са изчислени няколко показателя за точност. Именно: обща точност, *каппа* статистика и F статистика.

2.6.2 Регресионно моделиране на биофизични и биохимични параметри

Регресионният анализ в настоящето изследване е извършен като избрани 40 ВИ (Kamenova & Dimitrov, 2020), изчислени от спектралните канали на Sentinel-2 и 18 ВИ (Kamenova et al., 2018), изчислени за RapidEye са използвани като независима променлива за предвиждане на стойностите на търсените ББП. За изпълнението на научна Задача 4 и Задача 5 са изчислени линейни и експоненциални регресионни модели. Регресията е широко използвана техника за анализ на експериментални данни, която предполага, че са изпълнени някои предварителни предположения (нормално разпределение на остатъците и еднаква дисперсия на грешките т.нар. *хомоскедастичност*).

Регресионният анализ на данните е извършен с езика R, използваните библиотеки са stats, nls и ggplot2. За оптималните модели са генерирани диаграми на разсейването, които да спомогнат и визуалната диагностика на регресионните модели. Оценка на точността на всички модели е направена по метода на кръстосано валидиране с един независим отчет (LOOCV). На база на изчислените остатъци се пресмятат средноквадратична грешка (RMSE). За диагностика на моделите по отношение на направените предположения за грешките на предвиждането са изчислени Q-Q плотове и диаграми на разсейването на стандартизираните остатъци срещу предвидените (моделирани) стойности.

Във връзка с втората поставена Задача 5 са определени 5 възможни сценария за анализ на наличните данни за тестови район „Кнежа“. За изпълнението им са подготвени 5 отделни набора от данни, върху които да се изпълни регресионният анализ. Първият набор от данни включва всички налични данни, независимо от сорта на зимната пшеница и фенофазата, в която е направено измерването. Други два набора от данни включват по отделно данните за фенофаза братене и фенофаза вретенене, без значение от сорта пшеница. Последните два набора от данни съдържат данни разделен на база на двата налични сорта: „Енола“ и „Анапурна“.

2.6.3 Времева интерполация

За решаване на проблема заложен в Задача 6 е възприет подхода за „реконструкция“ на времева серия на LAI за вегетационният период – 1/10/2017 до 30/06/2018 г. Нуждата от подобен тип методика се налага поради недостатъчната наличност на наземни, а също така и на сателитни данни през вегетационният период от които да се изчисли ежедневна стойност на например LAI. Методът за това изчисление който е приложен в настоящето изследване за изчисление на времева серия с дневна честота е CSDM (Canopy Structural Dynamic Model), който представлява прост, семи-механистичен модел, който описва комбинираният ефект от растежа и пожълтяването, във връзка със сумата на активната температура, под формата на математически израз (Уравнение 1).

Уравнение 1

$$LAI(tt) = k \left[\frac{1}{(1 + e^{-a(tt-T_0-Ta)})^c} - e^{b(tt-T_0-Tb)} \right]$$

В уравнението параметрите a и b определят скоростта на растеж и стареене, c е параметър, позволяващ известна пластичност на формата на кривата, k е коефициент на мащабиране и T_0 , T_a и T_b са сумите на ефективната температура на поникване на растението, на достигане на зрялост, и на край на пожълтяването. CSDM допълнително се параметризира, така че да се получи нулев LAI след приключване на пожълтяването. Променливата (предикторът), която обуславя целият процес, е сумата на ефективната температура (tt). Тя се изчислява като сума от средно денонощните температури над базовата температура на развитие на дадена земеделска култура, за определен период от време. Базовата температура е температурната стойност под която процесът на растеж на дадена култура спира.

Базовата температура, използвана в настоящето изследване за зимна пшеница, е $0\text{ }^\circ\text{C}$, а началната дата за температурната сума е 1 октомври 2017 г. За изчисляване на tt , средно денонощните температури на въздуха са изтеглени от базата данни JRC MARS Meteorological Database. За изчисляването на времева серия с дневна честота от модела CGMS освен сумата на ефективните температури за целият период като независима променлива, е нужно и калибриране на модела с входни данни за LAI. Данни за LAI, са изчислени от наличните изображения от Sentinel-2 за периода на вегетация на пшеницата, на база на регресионните уравнения.

Трябва да се отбележи че CSDM моделът може да се използва в рамките на един определен вегетационен сезон. Преди да се изчисли или „реконструира“ времева серия за LAI е нужно да бъде събиран целият обем от данни, включващ сумата на ефективните температури и данните от LAI, нужни за калибрирането на модела. Иначе казано, моделът не може да се използва по време на вегетационния сезон, а след неговото приключване. Тестовият район използван за времева интерполация е по-малък като обхват поради съображения за скоростта на изчисление от модела.

Избрана е една четвърт от тестови район „Кнежа“, на границата между две ивици на заснемане от Sentinel-2. Така може да се демонстрира „реконструирания“ чрез модела CSDM LAI за тази част от района, която не е заснета от Sentinel-2 в конкретния ден (при липсващи данни в едната ивица). За предварителната обработка на сателитните изображения и прилагането на регресионните модели и самият CSDM модел е написан скрипт с Python.

3. Резултати от изследването

3.1 Картографиране на посеви от зимна пшеница

Тестови район „Кнежа“

Според наличността на безоблачни изображения и съобразено с фенологичните фази на развитие на зимната пшеница са избрани изображения със следните дати на заснемане: 31.10.17, 25.12.17, 30.03.18, 09.04.18, 14.04.18, 24.04.18, 08.06.18. Върху тези изображения и техни продукти (разновремени и композитни изображения) са приложени трите класификационни алгоритми (SVM, RF, MXL). Резултатите от проверката на точността на класификациите са обобщени чрез показателите обща точност, коефициент *карпа* и F1 точност за клас „Зимна пшеница“ които са представени в Таблица 6.

Избраните дати от есенно-зимния период на 2017 година, съвпадат с фенологичната фаза братене преди презимуване. Идентифицирането на посевите със зимна пшеница с висока точност в този период е от възлово значение поради възможността да се вземат ранни управленски решения от фермерите, при евентуални нарушения, които могат да бъдат установени с дистанционни методи. Резултатите от изследването показват че от втората дата на заснемане, именно 25.12.2017 се постига висока обща точност от 92.58 % с SVM, за разлика от по-ранното изображение от 31.10.2017, което постига обща точност от 83.17 %.

След презимуване по време на пролетния период, има две подходящи безоблачни изображения, заснети в периода когато пшеницата все още е във фаза братене. Те са заснети на 30.03.2018 и 09.04.2018. При класифицирането на изображенията от тези дати е постигнатата точност от $\approx 93\%$ за SVM. Следващите три изображения са заснети по време на фенологична фаза вретенене, датите са 14.04.2018, 24.04.2018, 29.04.2018 г., отчетените точности са съответно: 93.1%, 94.56%, 95.27%. Резултатът от последната дата е с най-висока точност, постигната и от трите класификационни алгоритъма за изображение от една дата.

Разновремето изображение № 13, от Таблица 6, което включва изображения от фенологични фази: братене, вретенене и фаза въсърсна зрялост, постига най-висока точност - 96.79 %, за алгоритъма SVM. Това е най-високата точност постигната от всички изображения от единични дати, разновремени и композитни изображения.

В Таблица 6, следват трите композитни изображения, които са съставени посредством платформата “Google Earth Engine”. Тези изображения, са съставени за три периода. Композитно изображение от Период 1, постига сравнително ниски резултати от 79.51% от SVM , поради това че в периода октомври и ноември,

пшеницата все още не е поникнала достатъчно и голата почва доминира полетата със зимна пшеница, също така има и локални появи на снежна покривка. Най-висока точност при класифицирането на композитните изображения е получена за Период 2 (94,5% обща точност с метода RF), което се дължи на факта, че през месеците април и май посевите от зимна пшеница са спектрално добре различими от масовите пролетни култури, като слънчоглед и царевица, при които почвата все още не е напълно покрита от растенията.

Резултатите постигнати от трите класификационни алгоритми са приложени върху 8 изображения от единични дати; 5 разновременни изображения и 3 композитни изображения. Между трите класификационни алгоритми, няма такъв който да се откроява пред останалите методи по отношение на точността. Разликите в общата точност, коефициентът карра и F1 точност между трите метода е в рамките на 2-3% независимо от класифицирания тип изображение Таблица 6. В най-много от случаите алгоритъма с най-висока обща точност при валидацията е SVM. На Фигура 6 е показана съставената карта на земеделските полета засети със зимна пшеница в тестови район „Кнежа“. Картата е съставена на база на класификацията на композитно изображение от Период 2 по метод на гора от дървета на решенията. Избрано е именно това изображение и метод в съответствие със заложената първоначално цел, а именно да осигури ефективно висока точност, възможно най-рано през вегетационният период. Според данните представени от Интегрираната система за администриране и контрол за 2017/2018 стопанска година, площите засети със зимна пшеница са общо 80863,40 ha., а според класификацията площите са 84964.60 ha. Това прави общо 4101.20 ha разлика или 5.07% разлика от тази представена чрез системата. Приложеният класификационен алгоритъм надценява площите, но разликата е малка (Таблица 5).

Таблица 5 Сравнение на площите заети със зимна пшеница през стопанска година 2017/2018 за тестови район „Кнежа“ според класификацията на композитното изображение от Период 2 по метода RF и данните от ИСАК

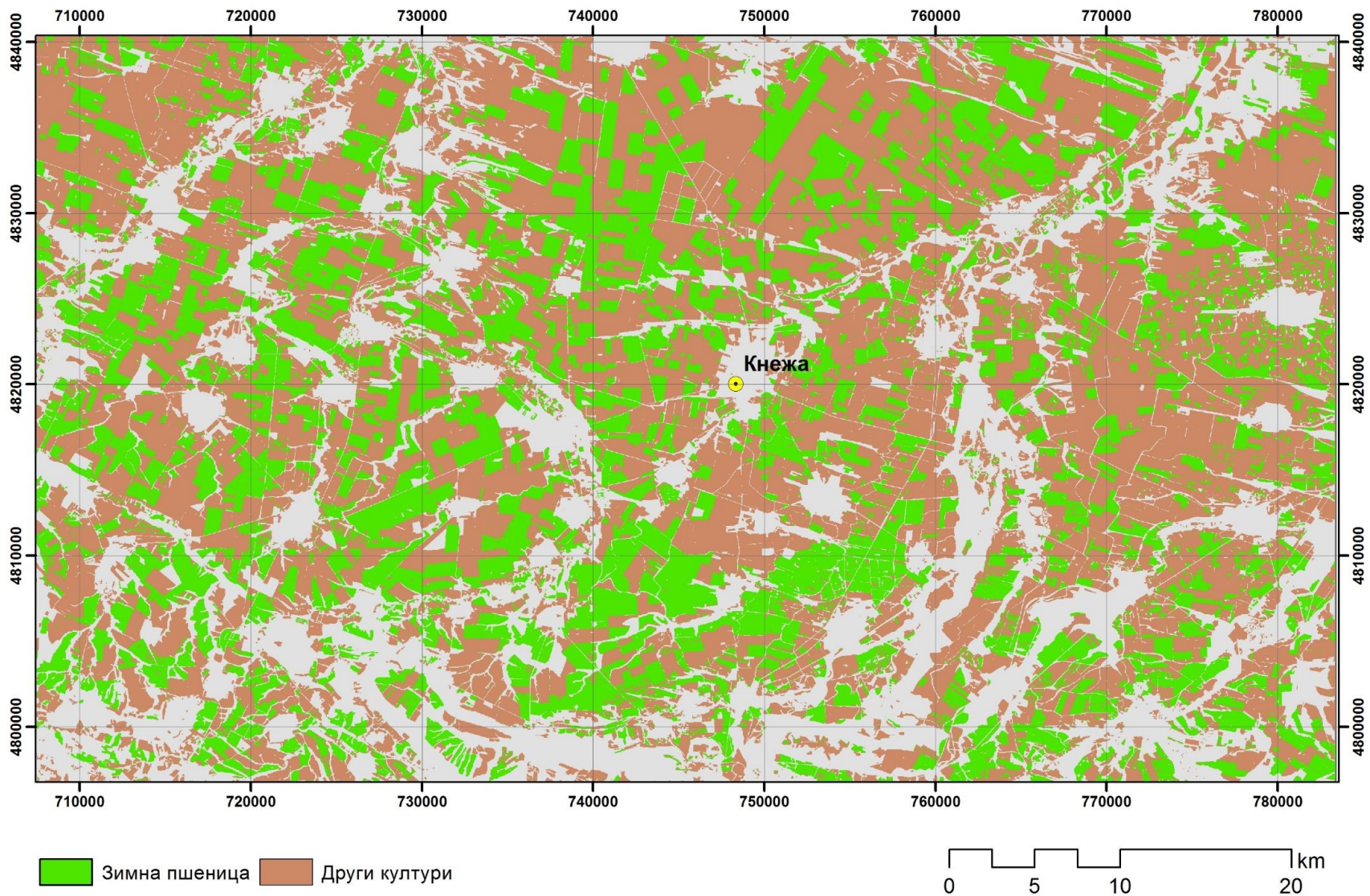
Площ зимна пшеница - класификация RF	84964.60	ha
Площ зимна пшеница ИСАК	80863.40	ha
Разлика	4101.20	ha
Разлика в проценти	5.07	%

Таблица 6 Обща точност, карра коефициент и F1 точност за класа "Зимна пшеница", за трите класификационни метода приложени върху района на тестови район „Кнежа“

№	Дати и периоди на заснемане на изображенията	Метод на опорните вектори			Метод на случайна гора на решенията			Метод на максималното правдоподобие		
		Обща точност, %	Коефициент карра, %	F1 точности за клас „зимна пшеница“, %	Обща точност, %	Коефициент карра, %	F1 точности за клас „зимна пшеница“, %	Обща точност, %	Коефициент карра, %	F1 точности за клас „зимна пшеница“, %
1	31.10.17 б	83.17	64.4	77.79	80.51	59.19	74.82	83.21	63.68	76.79
2	25.12.17 б	92.58	84.02	89.85	91.96	82.58	88.85	91.01	80.07	86.92
3	30.03.18 б	93.22	85.32	90.62	92.49	83.74	89.61	91.95	82.38	88.61
4	09.04.18 б	92.72	84.27	89.97	92.79	84.39	90.03	92.36	83.32	89.24
5	14.04.18 в	93.1	85.2	90.63	93.45	85.81	90.91	92.95	84.62	90.08
6	24.04.18 в	94.56	88.25	92.51	94.51	88.08	92.35	93.25	85.28	90.51
7	29.04.18 в	95.27	89.74	93.43	94.95	89.03	92.96	93.49	85.8	90.85
8	08.06.18 восьмична з.	94.24	87.46	91.94	92.44	83.51	89.37	93.88	86.52	91.22
9	31.10.17; 25.12.17	94.13	87.24	91.81	92.94	84.73	90.24	93.02	84.57	89.9
10	31.10.17; 25.12.17; 30.03.18	94.57	88.18	92.39	93.57	86.02	91.02	93.71	86.13	90.95
11	31.10.17; 25.12.17; 30.03.18; 14.04.18	95.82	90.86	94.09	95.39	89.9	93.46	94.54	87.94	92.11
12	31.10.17; 25.12.17; 30.03.18; 14.04.18; 29.04.18	96.5	92.37	95.08	96.26	91.79	94.67	95.28	89.56	93.16
13	31.10.17; 25.12.17; 30.03.18; 14.04.18; 29.04.18; 08.06.18	96.79	92.97	95.45	96.58	92.47	95.1	96.06	91.24	94.23
14	Композитно изображение, Период 1*	79.51	55.67	71.68	77.8	53.63	71.46	77.2	52.48	70.81
15	Композитно изображение, Период 2*	94.44	87.95	92.28	94.5	88.02	92.28	94.23	87.29	91.73
16	Композитно изображение, Период 3*	91.38	81.26	87.97	91.46	81.45	88.1	92.39	83.43	89.34

* Период 1 = '2017-10-01' - '2017-11-30'; Период 2 = '2018-04-01' - '2018-05-31'; Период 3 = '2018-06-01' - '2018-07-31'

Карта на полетата със зимна пшеница в тестови район "Кнежа", съставена чрез класификация по метода гора от дървета на решенията (2017-2018г).



Фигура 6 Карта на посевите със зимна пшеница за тестови район „Кнежа“, съставена чрез класифициране на комозитното изображение от Период 2 по метода RF.

Тестови район „Требич“

За картографиране на площите заети със зимна пшеница в тестови район „Требич“ са използвани пет изображения от спътника RapidEye заснети през периода март - юни 2015 г. По едно изображение е налично за месеците март, април и юни (31.03.2015, 20.04.2015 и 14.06.2015) и две за месец май (18.05.2015 и 21.05.2015). Заснемането на изображенията е съобразено с датите на кампаниите за пробовземане (във връзка със задачата за моделиране на биофизични параметри), както и с фенологичните фази на пшеницата. Поради ограничения брой на наличните изображения подходът за съставяне на времеви композитни изображения не е приложен за този тестови район, а са класифицирани само изображенията от отделните дати и разновремевите изображения съставени от тях. Резултатите са обобщени в Таблица 8.

Постигнатите общи точности след валидацията на класификациите, имат обща тенденция за повишаване с напредване на сезона. При разновремевите изображения се наблюдава очакваната тенденция с добавянето на данни от повече дати точността да се повишава, като най-високата постигната точност бива тази за разновременно изображение, включващо данни от 4 дати: 31.03.2015, 20.04.2015, 21.05.2015 и 14.06.2015. Постигнатата обща точност по метода SVM е от 91.4% и е най-висока измежду трите класификационни алгоритми. На Фигура 7 е илюстрирана съставената карта на посевите със зимна пшеница за тестови район „Требич“, класифицирани по метода на SVM с данни от разновремево изображение, включващо 4 дати. Като площи класифицираните пиксели в клас „Зимна пшеница“ за тестовия район представляват 5423.01 ha, което в сравнение с декларираните площи за 2015 година от 4624.59, дава 17.25% грешка (Таблица 7).

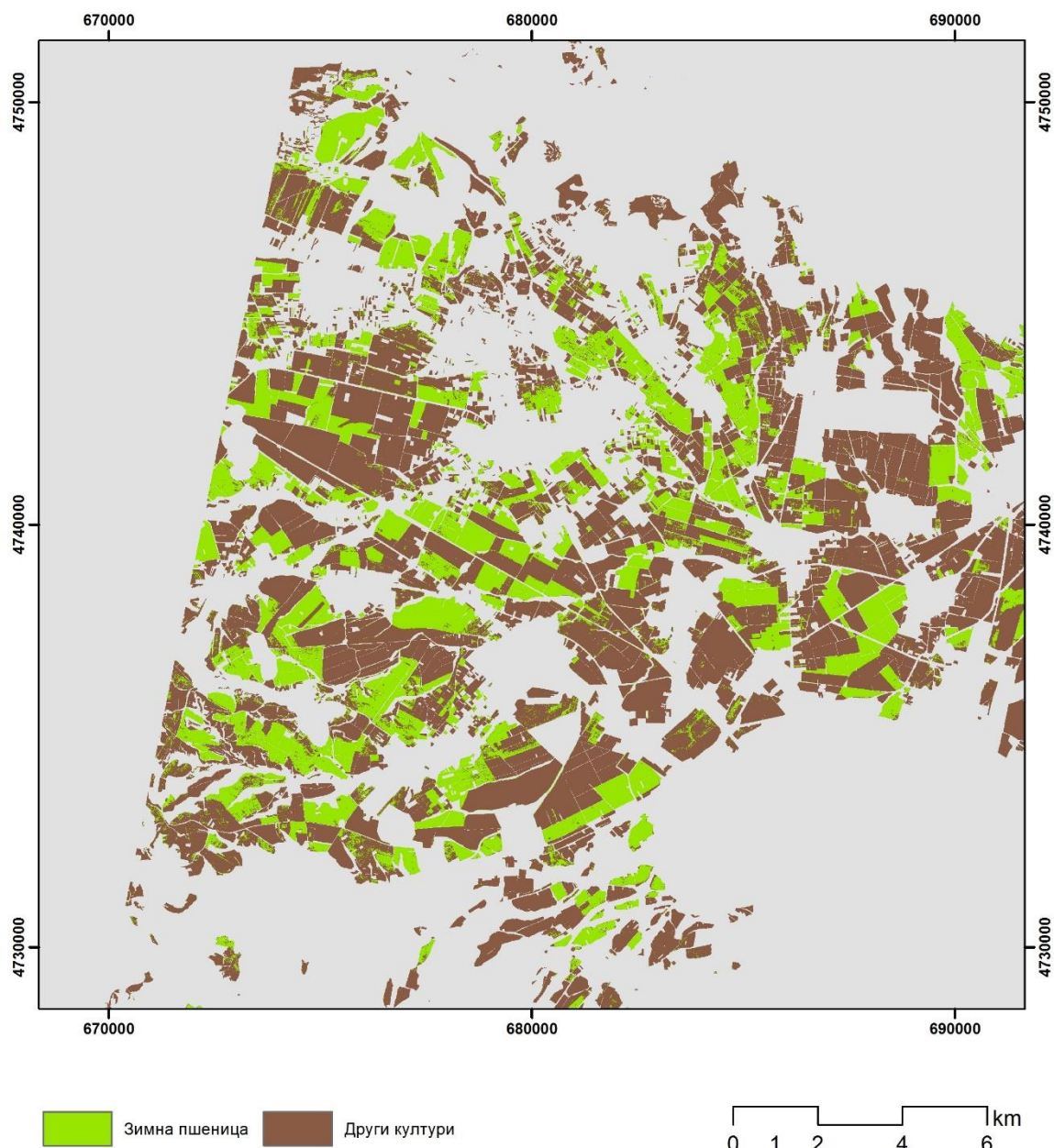
Таблица 7 Сравнение на площите зимна пшеница според МОВ и ИСАК, за район Требич

Площ зимна пшеница класификация - SVM		
„Требич“	5423.01	ha
Площ зимна пшеница - ИСАК	4624.59	ha
Разлика	798.42	ha
Разлика в проценти	17.26	%

Таблица 8 Обща точност, карра коефициент и F1 точност за класа "Зимна пшеница", за трите класификационни метода приложени върху тестови район „Требич“

№	Дати на заснемане на изображенията	Метод на опорните вектори			Метод на случайна гора на решенията			Метод на максималното правдоподобие		
		Обща точност, %	Коефициент карра, %	F1 точности за клас „Зимна пшеница“, %	Обща точност, %	Коефициент карра, %	F1 точности за клас „Зимна пшеница“, %	Обща точност, %	Коефициент карра, %	F1 точности за клас „Зимна пшеница“, %
1	31.03.2015	79.9	54.59	68.95	78.25	52.07	67.63	71.46	39.38	59.8
2	04.20.2015	78.65	50.26	65.43	75.74	46.5	63.87	71.95	33.58	53.43
3	18.05.2015	82.65	61.53	73.95	83.73	63.05	74.66	78.93	53.7	68.78
4	21.05.2015	82.32	60.54	73.19	83.26	61.88	73.82	76.79	50.9	67.49
5	14.06.2015	84.01	63.43	74.83	82.67	60.78	73.16	82.86	60.91	73.14
6	31.03.2015; 04.20.2015	82.3	59.93	72.57	81.21	58.01	71.45	77.03	48.09	64.5
7	31.03.2015; 04.20.2015; 21.05.2015	88.86	73.93	81.84	88.65	73.48	81.54	85.82	67.59	77.7
8	31.03.2015; 04.20.2015; 21.05.2015; 14.06.2015	91.4	79.52	85.59	91.14	78.82	85.07	90.03	76.35	83.4

Карта на полетата със зимна пшеница в тестови район "Требич",
съставена чрез класификация по метод на опорните вектори (2015-2016г.)



Фигура 7 Карта на посевите със зимна пшеница за тестови район „Требич“, класифицирани по метода на МОВ с данни от разновремено изображение

Дискусия и изводи и от картографирането на посеви от зимна пшеница

Трите тествани класификационни метода показват сходни точности. Макар че не може да бъде открит само един от тях се забелязва, че непараметричните методи RF и SVM като правило превъзхождат параметричния MXL. Методът на максималното правдоподобие дава приемлив резултат в двата тестови района на това изследване, но предвид изискванията към разпределението на данните които поставя трябва да се прилага внимателно в други райони, където доминиращите култури и фенологията на посевите са различни от тези в двата тестови района.

Резултатите от тестови район „Кнежа“ показват, че клас „Зимна пшеница“ може да бъде идентифициран с висока точност (F1 точност > 90%) чрез изображения от Sentinel-2 заснети в различни етапи от вегетацията. На базата на използваните в това изследване изображения най-ранната дата, за която е получена такава точност е 30 март (F1 точност = 90,62% по метода SVM; фаза братене). Точността се увеличава незначително в течение на вегетационния сезон и достига максимум за изображението от 29 април (F1 точност = 93,43% по метода SVM; фаза вретене). Резултатите показват също, че в периода преди презимуване на зимната пшеница също е възможно получаването на класификация с висока точност, макар и под 90%.

Използването на разновременни изображения подобрява резултатите от класификацията, но сравнително слабо (обикновено повишението на точността е до 1-2%). Това означава че, имайки предвид допълнителния обем данни, който трябва да бъде съхраняван и/или обработван, класификационният подход използващ разновременни изображения може не винаги да е оправдан.

Това затруднение може да се преодолее чрез съставяне на времево композитно изображение, при това без съществена загуба на точност - F1 има стойност 92.28% за композитното изображение *Период 2* (№15, Таблица 6) на фона на 95.45% за най-добрата класификация на разновременното изображение от шест дати (№13, Таблица 6). Резултатите от тестови район „Требич“ потвърждават, че класификаторите RF и SVM са по-точни от MXL. Подобно на тестови район „Кнежа“ се наблюдава сравнително стабилен резултат при класифицирането на изображения заснети от края на март до средата на юни, като все пак оптимални се явяват датите през втората половина на този период. Потвърждава се също изводът за предимство на разновременните класификации спрямо тези използващи данни само от една дата, като за разлика от тестови район „Кнежа“ разликата в F1 достига до 10%.

Резултатите получени при класифицирането на данни от Sentinel-2 в тестови район „Кнежа“ са сходни с публикуваните изследвания от някои други райони използващи същия или сходен тип данни. Li et al., (2022) получават обща точност от 86.7% идентифицирайки пет различни класа, включително зимна пшеница в земеделски район в Китай.

Сравним резултат е получен в настоящото изследване използвайки само оптични данни и по-проста двоична класификационна схема. Използвайки класифициране с един клас по метода SVM (one-class support vector machine) Yang et al., (2021) постигат обща точност от 92.61%. Qu et al., (2021) картографират зимната пшеница в три тестови района в северен Китай с F1 точност от 88.6% до 92.5% използвайки специално разработен индекс на зимната пшеница (winter wheat index; WWI) изчислен чрез времева серия от Landsat. Dong et al., (2020) използват метода TWDTW (time-weighted dynamic time warping) за да картографират зимната пшеница в 11 провинции на Китай получавайки обща точност в границите от 84.97 % до 95.85 %.

Авторите установяват, че точността на идентифициране на зимната пшеница нараства в течение на вегетационния период (от октомври до юни), като от април нататък почти не се променя. Това потвърждава наблюденията в нашите тестови райони. Резултатите получени в настоящото изследване могат да бъдат подобрени в бъдеще усъвършенствайки приложените методи. От особено значение е по-доброто разграничаване на зимната пшеница от останалите зимни култури, като зимната рапица. В литературата има примери в това отношение, основаващи се на използването на радарни данни (Dong et al., 2020), на разликите в отражението в червения и зеления канал между двете култури в определени фенофази (Yang et al., 2021) и др.

3.2 Регресионно моделиране на биофизични и биохимични параметри на зимна пшеница

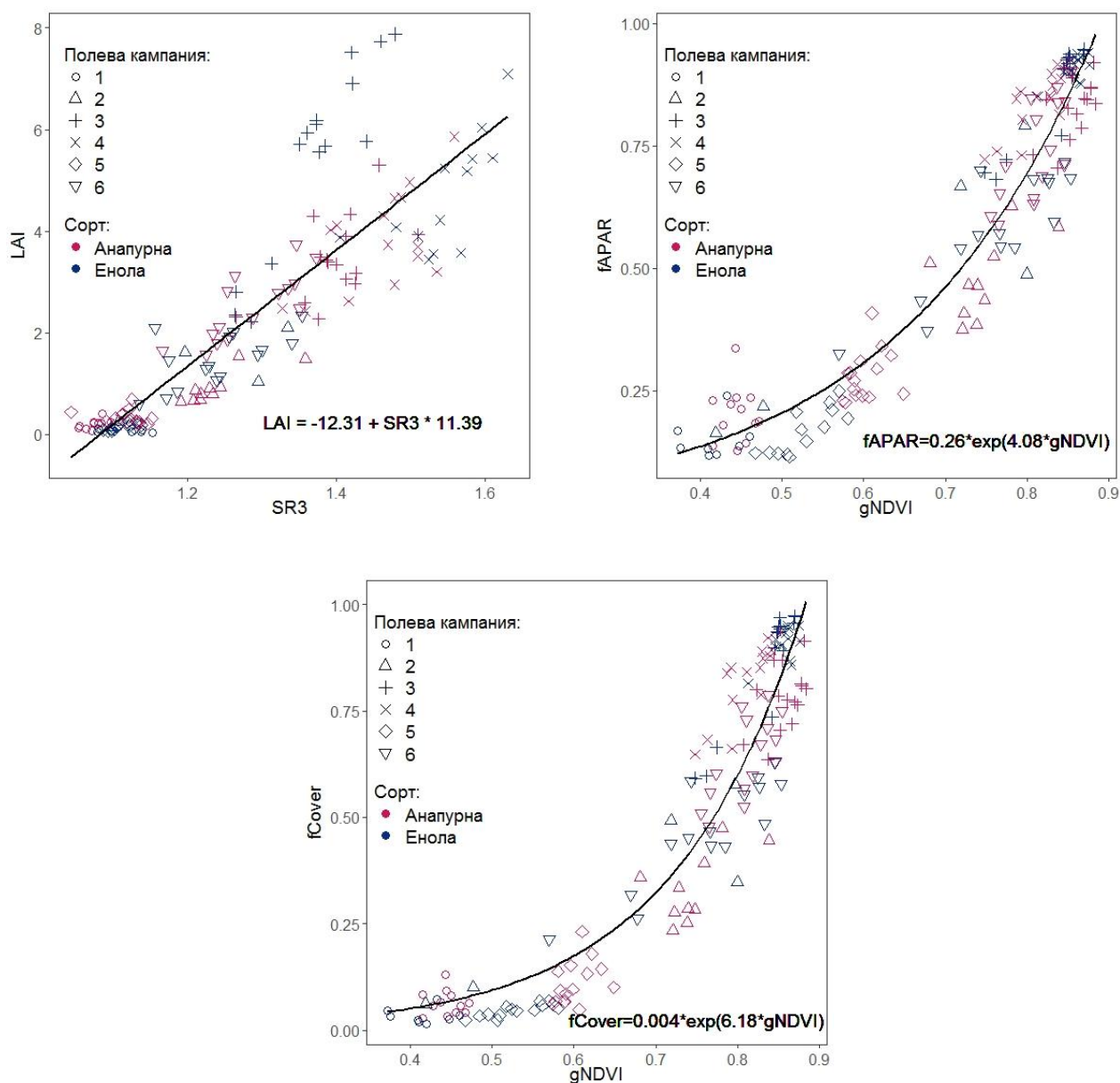
3.2.1 Резултати от регресионното моделиране за тестови район „Кнежа“

За тестови район „Кнежа“ е извършен регресионен анализ по пет възможни сценария, в зависимост от използваните входни данни за биофизичните параметри. Тези сценарии са описани в точка 2.6.2 Регресионно моделиране на биофизични параметри. Всеки един от 40 ВИ (Kamenova & Dimitrov, 2020) е систематично оценен, чрез съставяне на линейни и експоненциални регресионни модели. Индексите са използвани като независима променлива за прогнозиране на всеки един от търсените биофизични параметри. Като следващ етап е направено сравнение на точността на отделните модели на базата на изчисление

на средната квадратична грешка (RMSE). Подготовката на данните, съставянето и валидацията на регресионните модели, както и изготвянето на графики е изпълнено с езика R, използваните библиотеки са: stats, nls, ggplot2, ggpubr.

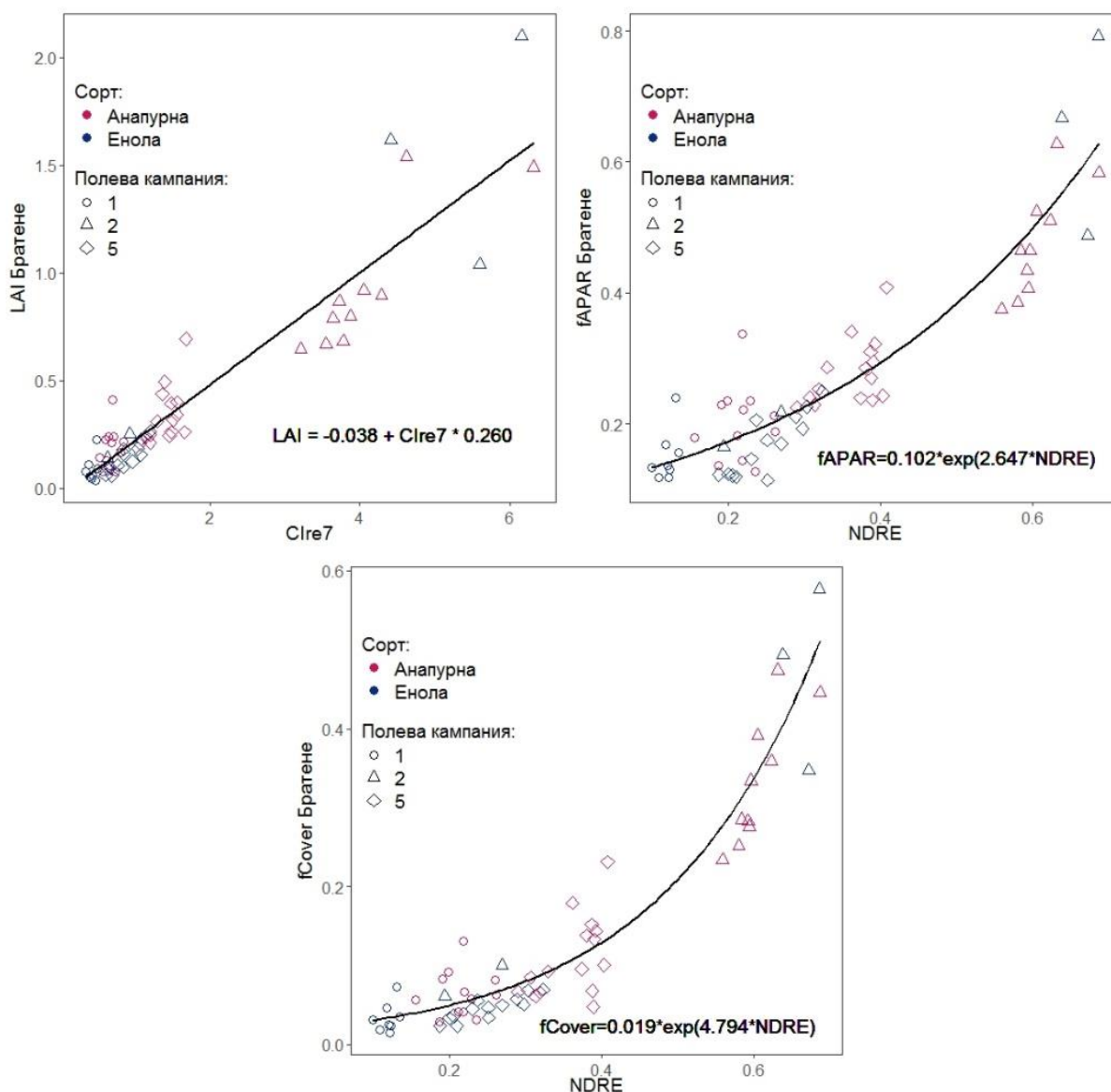
При прилагане на първия работен сценарий (с включени всички налични данни от район „Кнежа“), RMSE за LAI показват големи разлики в зависимост от ВИ, който е използван като зависима променлива (предиктор) в регресионния модел. Като най-оптимален в предсказването на стойности за LAI се представя вегетационният индекс SR3 ($rRMSE = 45\%$; $RMSE=1 \text{ m}^2 \text{ m}^{-2}$). Зависимостта на този индекс и LAI е линейна и моделът е значим ($p < .001$) (Фигура 8). Като цяло моделите на fAPAR показват по-висока прогнозна точност в сравнение с тези на LAI съдейки по RMSE. На Фигура 8 е показана връзката между fAPAR и gNDVI ($rRMSE=14\%$; $RMSE=0.08$), заедно със съставения експоненциален модел. При тази връзка, наблюдаваме по-постоянно разпределение на данните в ранните фенофази от полеви кампании: №1, №2, №5, отколкото в по-късните фенофази. Много сходни резултати са получени и за параметъра fCover, по отношение на индексите, които биват най-добри предиктори спрямо изчислената средна квадратична грешка. Все пак грешката е по-висока отколкото при fAPAR. %. Както и при fAPAR, най-силна връзка fCover показва с gNDVI, NDRE1 и GIPVI ($rRMSE=23\%$; $RMSE=0.11$).

Обсъдените по горе RMSE показват цялостната производителност на регресионните модели, и могат да се считат за обобщени грешки за целия вегетационен сезон. Представените до тук модели са съставени с измервания получени в течение на почти целия вегетационен период и следователно могат да се прилагат в различни фенофази. В идеалния случай точността на модела би трябвало да е еднаква независимо кога през вегетационния период се прилага. В действителност това може да не е така и представените по-горе стойности на RMSE да не отразяват реалната точност в дадена фенофаза.

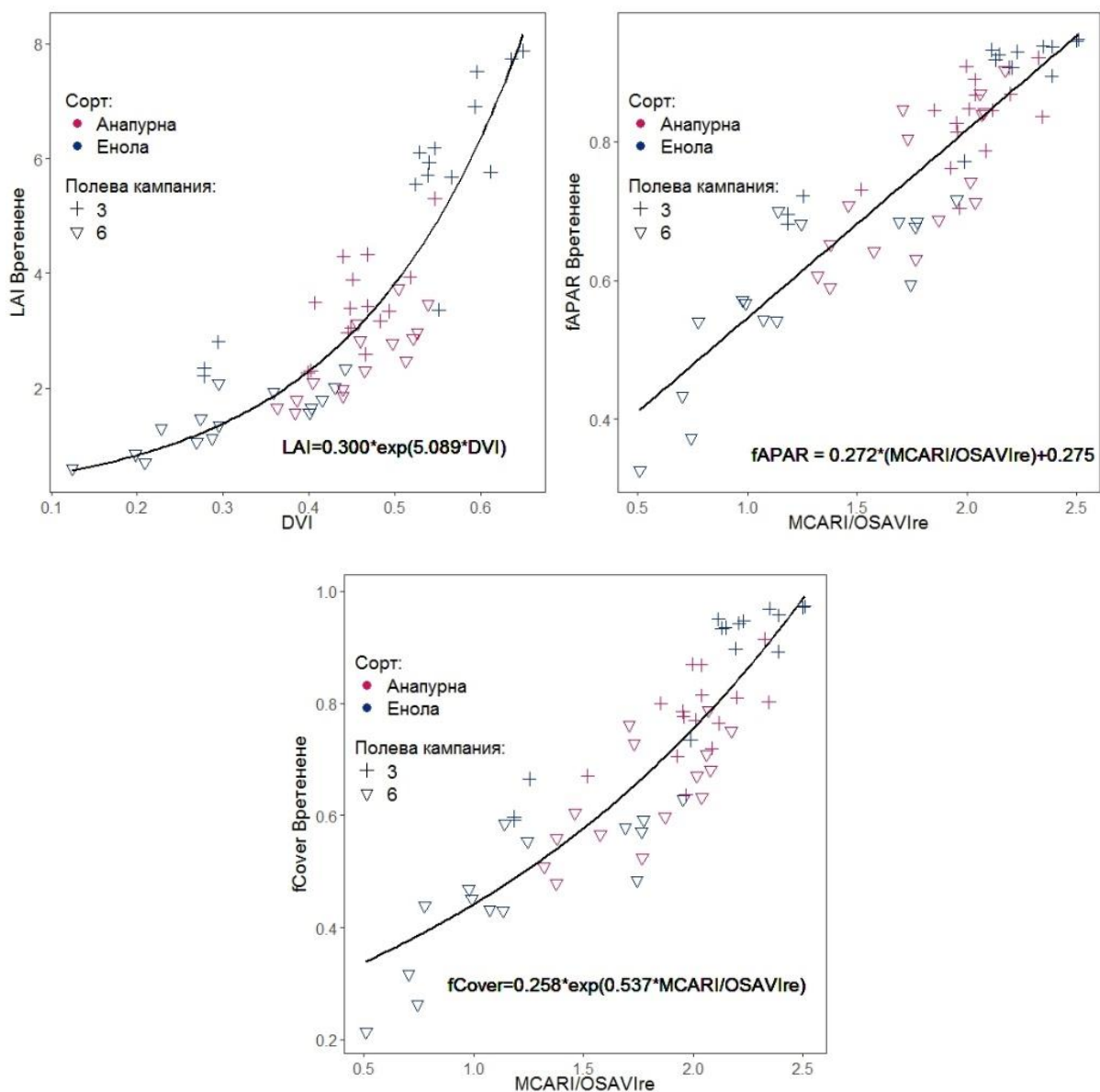


Фигура 8 Диаграми на разсейването на трите изследвани биофизични показатели (LAI, fAPAR, fCover), спрямо съответния избран оптимален вегетационен индекс за общия набор от данни (всички сортове и фенофази). Показани са и регресионната линия и уравнение на оптималния (т.е. с най-ниска rRMSE) регресионен модел. Номерата на полевите кампании отговарят на тези в Таблица 1

Следващите два сценария разгледани при регресионния анализ включват съставяне и валидиране на модели с данни от отделни фенофази. Разгледани са два случая: братене и вретенене. Моделът за фенофаза братене с най-ниска относителна средна квадратична грешка за LAI е базиран на Clre7 ($rRMSE=43\%$; $RMSE=0,16 \text{ m}^2 \text{ m}^{-2}$) (Фигура 9). На Фигура 10 е показан оптималният модел за LAI за фенофаза вретенене, е този базиран на DVI ($rRMSE=26\%$; $RMSE=0.85 \text{ m}^2 \text{ m}^{-2}$); връзката между двете променливи е експоненциална.



Фигура 9 Диаграми на разсейването за фаза братене, за трите изследвани биофизични показатели (LAI, fAPAR, fCover), спрямо избраният оптимален вегетационен индекс. Показани са и регресионната линия и уравнение на оптималния (т.е. с най-ниска $rRMSE$) регресионен модел. Номерата на полевите кампании отговарят на тази в Таблица 1



Фигура 10 Диаграми на разсейването за фаза вретенене, на трите изследвани биофизични показатели (LAI, fAPAR, fCover), спрямо избраният оптимален вегетационен индекс. Показани са и регресионната линия и уравнение на оптималния (т.е. с най-ниска rRMSE) регресионен модел. Номерата на полевите кампании отговарят на тази в Таблица 1

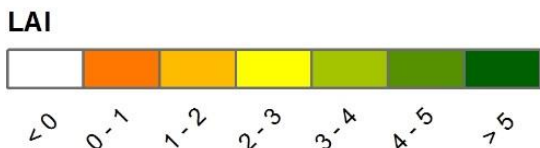
Най-ниската грешка при моделирането на параметъра fAPAR за фенофаза братене е постигната от индекса NDRE с експоненциален модел (rRMSE=22%; RMSE=0,06) (Фигура 9). С близка до тази грешка се представят и индексите NDRE1, WDRVI и CIre7. За фенофаза вретенене, линейната връзка на индекса MCARI/OSAVIre е с най-добър резултат (rRMSE=9%; RMSE=0,07) (Фигура 10). Експоненциалният модел с индекса NDRE постига най-ниската средна квадратична грешка за параметъра fCover във фенологичната фаза братене

($rRMSE=33\%$; $RMSE=0,04$), последван от индекса NDRE1. Във фенофаза вретенене, като най-точен предиктор се отличава MCARI/OSAVI_{re} с експоненциална функция ($rRMSE=12.6\%$; $RMSE=0.09$).

Последните два сценария на проведения регресионен анализ включват съставяне на регресионни модели, в които няма смесване на данни от различни сортове зимна пшеница. Тоест създадени са отделни набори от данни за сорт „Анапурна“ и сорт „Енола“ и са съставени отделни модели с всеки от тях. За сорт „Анапурна“, индексът който показва най-малката средна квадратична грешка е SR3 ($rRMSE = 30.6\%$). Този индекс е оптимален също така и при общите модели, връзката е също така линейна. За сорт „Енола“, като оптимален индекс се отличава Clg7 ($rRMSE = 42.9\%$), тази грешка е значително по-висока отколкото при сорт „Анапурна“. Следващите по точност вегетационни индекси са gNDVI, MCARI/OSAVI_{re} и OSAVI; при всички тях прави впечатление, че функцията е експоненциална. gNDVI и GIPVI ($rRMSE = 15.3\%$) са индексите които показват най-добри резултати при сорт „Анапурна“, спрямо fAPAR.). За сорт „Енола“ индексът NDRE1 дава най-точен резултат ($rRMSE = 13.6\%$) спрямо fAPAR. Резултатите за fCover при сорт „Анапурна“ показват, че индексите SR3, NDRE1, Clg8A, gNDVI и GIPVI имат най-точен резултат, и същевременно и петте индекса имат линейна връзка с fCover. При сорт „Енола“ индексът с най-добър резултат е Clg7 ($rRMSE = 19.1\%$) и също показва линейна зависимост спрямо fCover.

Съставяне на карти на биофизични параметри за район „Кнежа“

На база на изчислените регресионни модели показващи най-ниски средни квадратични грешки са изчислени и карти на биофизичните параметри за район „Кнежа“. Извадка от картите, изчислени на база на регресионните уравнения, е показана на Фигура 11. Картите са изготвени, като за всеки пиксел е изчислена стойността на оптималния вегетационен индекс по данните от сателитно изображение от Sentinel-2. След това, стойността на индекса се замества в регресионното уравнение с оптималните регресионни параметри изчислени за модела. По този начин (пиксел по пиксел) се изчислява карта на разпространението на търсеният биофизичен параметър за цялата територия. Картата на полетата със зимна пшеница получена в резултат от класификация (Фигура 6) се използва за да се моделират само стойностите, които попадат в полета със зимна пшеница.



Фигура 11 Извадка от карта на LAI за 09-04-2018 г., моделиран с регресионен модел, индекс SR3
 *за стойности под 0, виж текста

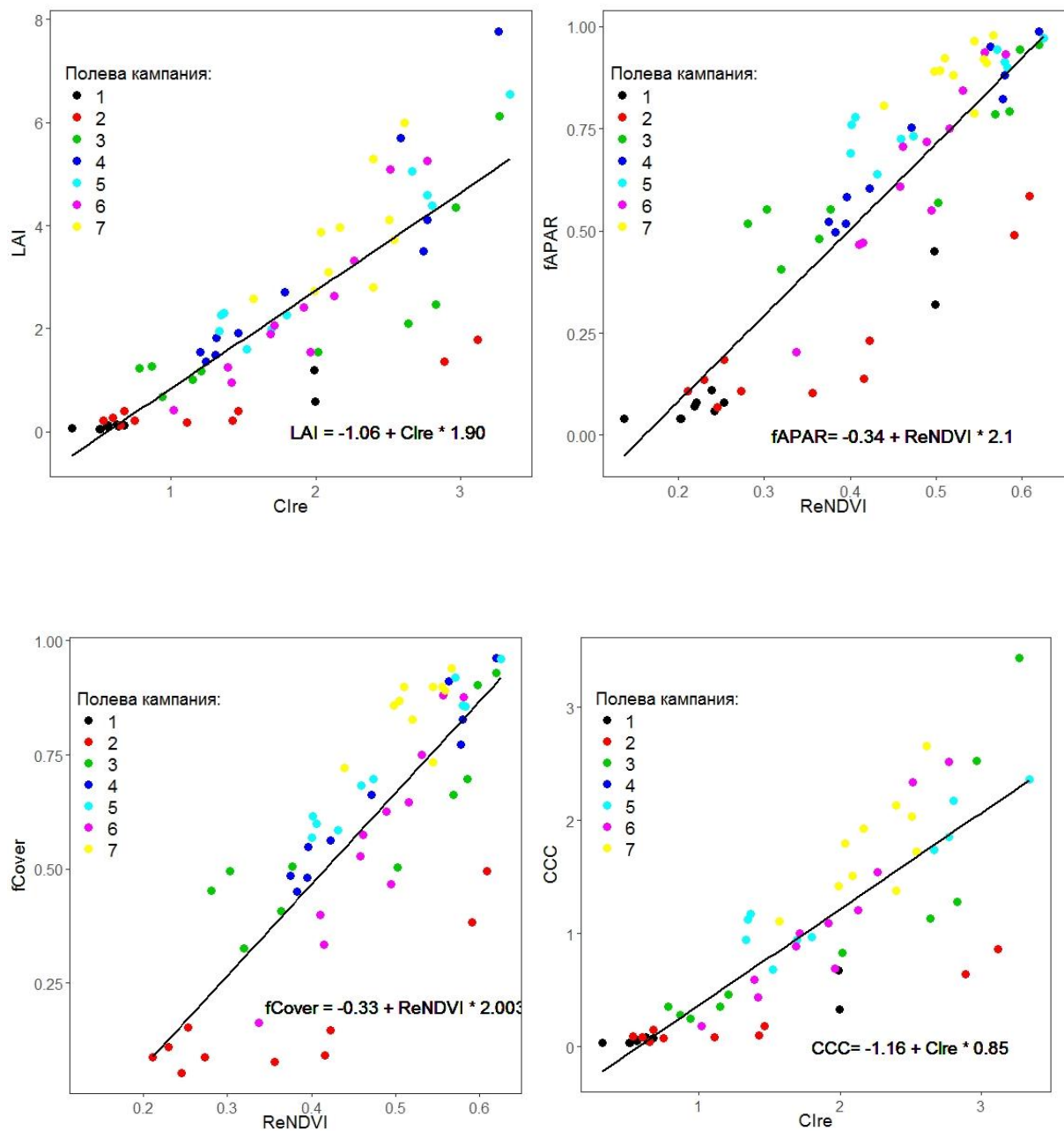
При изчислението на стойностите за LAI има моделирани и негативни стойности, които най-често попадат в пиксели заети с почва. Тези стойности са практически невъзможни, но поради естеството на регресионните модели които се базират на уравнение, което е калибрирано за дадени стойности, могат да се калкулират и негативни стойности. Затова, екстраполирани стойности извън обхвата на калибрационните данни не бива да се считат за достоверни.

3.2.2 Резултати от регресионното моделиране на тестови район „Требич“

За тестови район „Требич“ е проведен статистически регресионен анализ, за чието изпълнение са използвани наземни данни за съставяне на моделите (Таблица 2). Вегетационните индекси използвани в анализа са 18 на брой (Kamenova et al., 2018). Валидацията на резултатите е осъществена на базата на изчисление на средната квадратична грешка посредством кръстосано валидиране с един независим отчет (LOOCV). Поради по-малкия обем данни и наличието само на един сорт пшеница (сорт „Енола“) за този тестови район не са провеждани регресионни анализи по фенофази и сортове. Моделите с най-висока прогнозна точност (т.е. най-ниска rRMSE) за четирите ББП (LAI, fAPAR, fCover и CCC) са разгледани подробно на база на диаграми на разсейването.

Резултатите показват, че индексите с най-ниска средна квадратична грешка, спрямо LAI са NIR/Re и Clre (rRMSE = 46.73%). Дори за тези индекси грешката е твърде висока поради слабата им корелация с LAI както се вижда от Фигура 12. Резултатите за fAPAR, показват по-ниски средни квадратични грешки, в сравнение с резултатите за LAI. Най-добър резултат показва индекса reNDVI (rRMSE = 26.56%) и индекса NDVI (rRMSE = 28.10%). Въпреки това, дисперсията на стойностите около регресионната линия е значителна (Фигура 12). Като обща тенденция може да се отбележи, че линейните модели дават по-добри резултати от експоненциалните регресионни модели. Моделите за fCover дават подобни резултати като тези за fAPAR, като най-добър модел се получава използвайки индекса ReNDVI (rRMSE=27.02%).

Най-добрите модели са използвани за съставяне на карти на изследваните биофизични и биохимични параметри. за тяхното съставяне са използвани коефициентите от регресионното уравнение, и то е приложено върху наличните за всяка дата изображения от RapidEye. Използваните ВИ са както следва: за LAI и CCC е използван индексът Clre, а за fAPAR и fCover индексът ReNDVI. Картите са съставени за всяко от наличните изображения от RapidEye, като на Фигура 13 е показана извадка от картата за LAI за 18 май 2015 г.



Фигура 12 Диаграми на разсейването за район „Требич“ на изследваните биофизични и биохимични показатели (LAI, fAPAR, fCover, CCC), спрямо избраният оптимален вегетационен индекс. Показани са регресионната линия и уравнението на оптималния регресионен модел. Номерата на полевите кампании отговарят на тези в Таблица 2



Легенда

LAI



0-1 1-2 2-3 3-4 4-5 7-8

Фигура 13 Извадка от карта на LAI, за 18-05-2015 г., съставена чрез регресионен модел с индекс *Cire*

Дискусия и изводи от регресионното моделиране

В това проучване ние оценихме 40 вегетационни индекса, изчислени с данни от спектрални канали от Sentinel-2 (комбинация от изображения и симулирани данни) за техните възможности за прогнозиране на LAI, fAPAR и fCover на зимна пшеница. Съставени са линейни и експоненциални регресионни модели за всеки ВИ, за три отделни периода и за два отделни сорта: (i) през целия сезон, включващ всички налични данни, независимо от етапа на развитие (от братене до цъфтеж), (ii) за фаза братене, (iii) за фаза вретене, (iv) за сорт „Анапура“, (v) за сорт „Енола“.

Резултатите показват, че нито един от разглежданите ВИ не е оптимален за всички биофизични променливи в различните фенофази. За всеки биофизичен параметър и период от време се открояват групи от индекси с най-добри резултати вместо единичен оптимален индекс. В същото време групите от най-добри ВИ имат много сходен състав между LAI, fAPAR и fCover, т.е. ако някой индекс е добър предиктор за LAI, той обикновено се представя добре и за fAPAR и fCover и обратно. Индексите, които имат добри резултати в настоящето изследване, могат да бъдат разделени в няколко групи в зависимост от спектралните канали, използвани във формулата им:

- Индекси, използващи зелен канал (B03) в комбинация с канал от близко инфрачервения спектрален диапазон B08 - Clg8, GIPVI и gNDVI. Те са полезни при прогнозиране на биофизичните променливи с общи модели, включващи всички данни.

- ВИ, които използват канали от червения ръб (B05, B06 и B07) - Clre7, NDRE, NDRE1, SR3 и SR4. Те също са полезни при прогнозиране на биофизичните променливи с общи модели, но в допълнение те представляват доминиращата част от най-добрите индекси за фенофаза братене.

- Индекси, които използват червен канал (B04) в комбинация с канал от близко инфрачервения спектрален диапазон (B08) - DVI, EVI, PVI, OSAVI, TSAVI, PVI и TSAVI. От тази категория има ВИ, които използват параметрите на почвената линия в своите формули. Тези индекси са особено полезни при прогнозиране на биофизичните параметри във фаза вретене.

- Индексът MCARI/OSAVI_{re} - този индекс има същата полезност като предишната категория и е най-добрият предиктор за fAPAR и fCover във фаза вретене.

- Други индекси, които могат да бъдат разграничени, са Clre8A и Clg7, които използват червен канал в комбинация с близко инфрачервен канал или зелен канал.

Значението на каналите от спектралната област RE на Sentinel-2 за прогнозиране на LAI е демонстрирано и преди както на база на моделиране с ВИ, така и на физически базирани техники. Проучване на Pan et al. (2019) показва, че вегетационните индекси, които използват тези канали, като reNDVI и REP,

превъзхождат по-традиционните индекси като NDVI и EVI при емпирично базирано извличане на LAI. В съответствие с това са и резултатите от Delloye et al. (2018), които установяват, че интегрирането на канали от червения ръб на Sentinel-2 повишава точността на изчисляване на Green Area Index (параметър, тясно свързан с LAI) чрез инверсия на физически базиран модел.

Най-точните резултати за район „Кнежа“ са тези за fAPAR, следвани от fCover и LAI. Общите модели за fAPAR и fCover имат добра точност, особено моделът за fAPAR. Резултатите за LAI не са задоволителни при използване на общ модел. Предишни проучвания, използващи данни от Sentinel-2, съобщават за по-добра точност. Например Pan et al. (2019) използват експоненциална регресия и reNDVI, за да извлекат LAI на зимна пшеница с RMSE от $0,85 \text{ m}^2 \text{ m}^{-2}$ (rRMSE=22%). В настоящото изследване се установи, че за предпочитане е моделирането на LAI да се извършва по отделни фенофази, в случая братене или вретенене. Dahms et al. (2016) също установяват, че точността на моделиране е свързана с фенологичните фази.

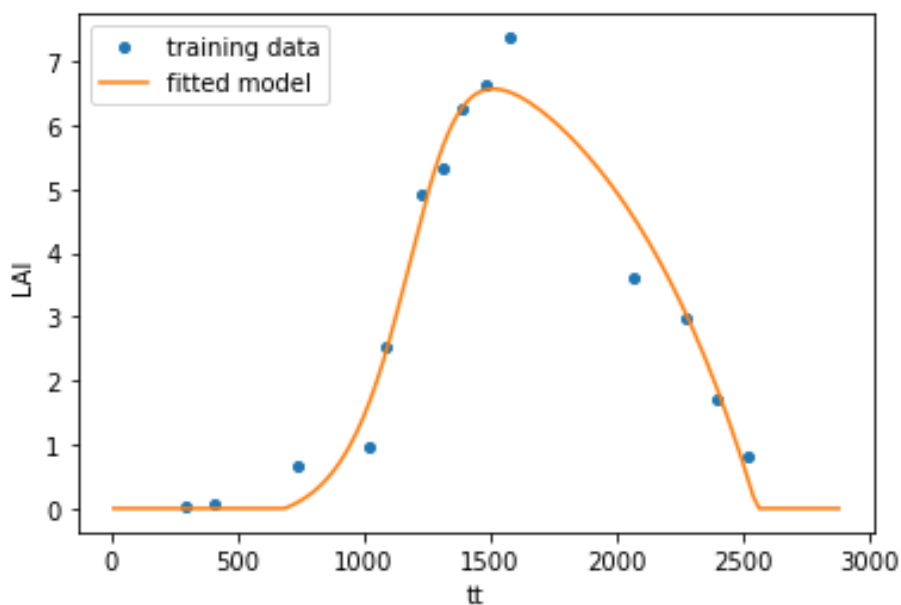
Регресионните модели могат да се използват за картографиране на биофизичните променливи в локален или регионален мащаб, когато е налична карта на полетата със зимната пшеница. На Фигура 11 и Фигура 13 са показани извадки от карти на биофизични променливи, получени от регресионен модел, съставен за фаза вретенене. В картите са получени и нереалистични стойности прогнозираны от модела, като например отрицателен LAI или fAPAR над 1,0. Такива резултати са възможни поради емпиричния характер на регресионните модели. Статистическите модели са обвързани с диапазона си на калибриране и могат да предвиждат резултати само в рамките на пространството от данни, в което са калибрирани.

От практическа гледна точка общият модел е по-удобен, защото може да се използва, когато културите са във различни фенофази. Въпреки това, в общия случай моделите за братене и вретенене са по-точни и тяхното използване трябва да се препоръча особено за LAI. Но за изчисление на голяма територия, потребителят не винаги е наясно в какъв етап на развитие се намира зимната пшеница. На една и съща дата някои полета може да са в етап вретенене докато други може да са все още в братене. Изборът кой модел, братене или вретенене да се използва не е лесен на границата на двата етапа на развитие. Въз основа на наличните полеви данни, моделът за фенофаза братене може да се използва до 15 март. След тази дата се препоръчва използването на модела за вретенене. Въпреки това, тази прагова дата е условна и трябва да се разглежда с голямо внимание, тъй като може да не е подходяща за други региони и вегетационни сезони.

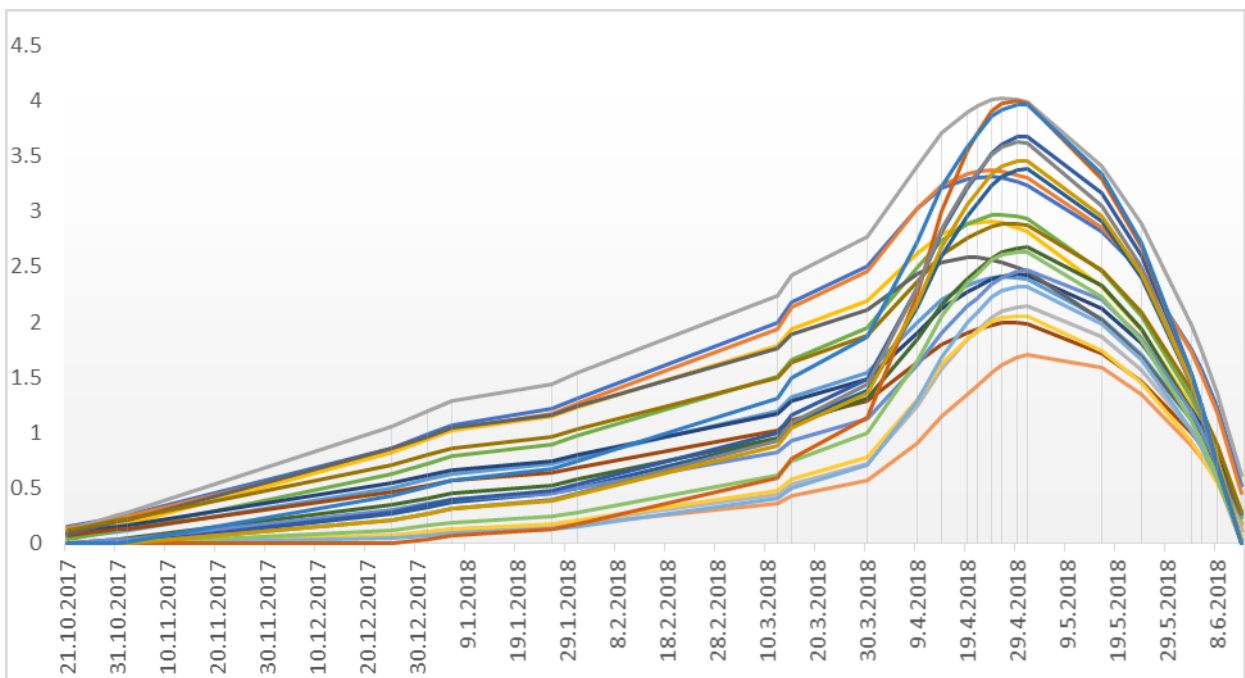
3.3 Резултати от времевата интерполация

Реконструирането на LAI е осъществено чрез времева интерполация с модела CSDM (Canopy Structural Dynamic Model). В хода на вегетационния сезон в настоящото изследване, функцията на модела е приложена върху цялата изследвана територия. Така може да се оцени разпределението на LAI и във времеви и в пространствени аспекти. CSDM е калибриран с данни за LAI, които са получени чрез регресионните модели изведени в предходния етап на настоящото изследване. Независимата променлива (tt) (от Уравнение 1) представлява сумата на ефективните температури от датата приета за начало на вегетационния сезон до конкретна дата. Тези данни също са запазени в растерен формат, с пространствена разделителна способност 2 km. Моделът прогнозира стойността на LAI ежедневно на базата на достигнатата сума на ефективните температури, през вегетационния сезон.

На Фигура 14 е показан моделът CSDM с жълта линия, за произволен пиксел в рамките на пшеничните полета. Той е калибриран с данни за LAI, които са генерирани с наличните сателитни данни за периода (отбелязани със сини точки на графиката). На Фигура 15 са показани времеви серии за избрани пиксели със зимна пшеница съставени чрез модела CSDM.



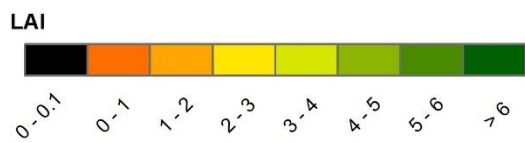
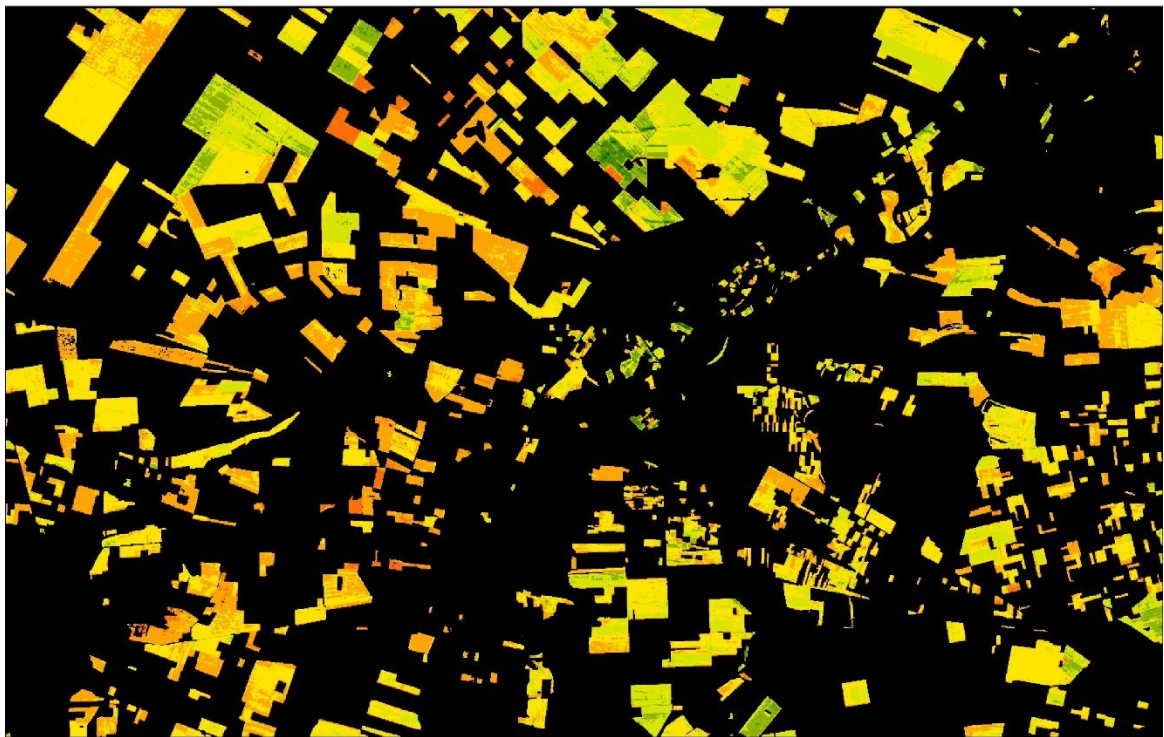
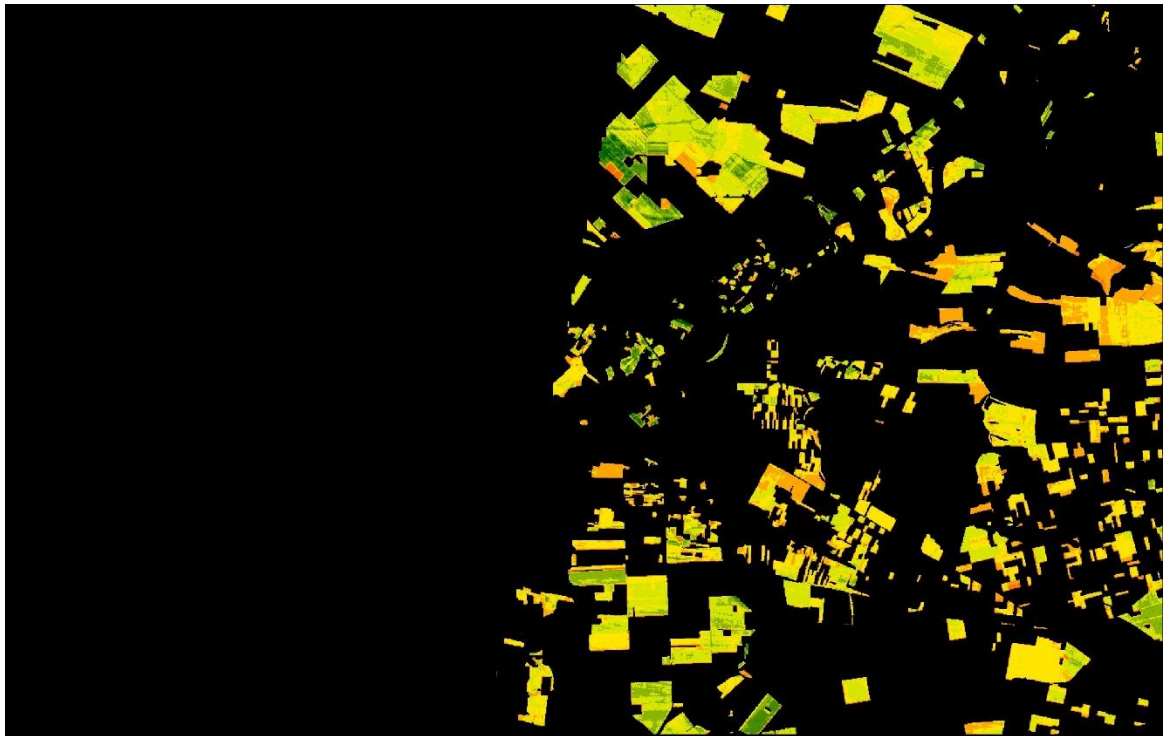
Фигура 14 CSDM модел, калибриран за произволен пиксел, tt - сума на ефективната температура.



Фигура 15 Времеви серии на LAI, за вегетационен сезон 2017 – 2018 получени чрез модела CSDM за точките на пробовземане от полевите кампании в тестови район „Кнежа“.

До голяма степен добавената стойност от моделирането с CSDM е, че можем да „реконструираме“ LAI, във всеки един момент за изследвания вегетационният период, за който нямаме спътникови данни. Това е демонстрирано на Фигура 16, където са представени карти на LAI за 01-05-2018 г. Картата в горната част на фигурата е получена посредством регресионно моделиране и на нея има липсващи стойности за голяма част от територията, поради липсващо изображение за тази дата от едната ивица на заснемане от Sentinel-2. В другата ивица, в която има налични данни, стойности за LAI са изчислени на база регресия.

На Фигура 16 (долу) е показана карта на LAI, която е изчислена посредством CSDM за същата дата. На тази карта няма територии без стойности за LAI, тъй като CSDM използва като предиктор сумата на ефективната температура и липсващите сателитни данни, нямат отношение към изчислението. Така реконструираният стойности на LAI може да запълват липсващи данни за минали периоди. Визуалната инспекция на картите показва, че вариацията в стойността на LAI в рамките на реконструирания растер отразява реалистично и съответства добре на вариацията в оригиналния растер, получен чрез регресионния модел.



Фигура 16 Карта на LAI за 01/05/2018, на част от територията на район „Кнежа“. Горе LAI, моделиран посредством регресия; Долу LAI моделиран посредством CSDM

Дискусия и изводи от времевата интерполация

На базата на резултатите от проведеното изследване върху територията на тестови район „Кнежа“ може да се заключи, че представеният подход за създаване на времева серия от пространствени данни за LAI базиран на модела CSDM е от една страна технически приложим, а от друга дава резултати в съответствие с очаквания ход на стойностите на LAI през вегетационния сезон на зимната пшеница. Основен момент при адаптирането на модела CSDM, така че да се използва в пространствен аспект, е наличието на пространствени метеорологични данни и тяхното съчетаване със спътниковите данни. Въпреки че използваните метеорологични данни от JRS са във вид на гريد с голям размер на клетката (25 km) те могат успешно да се използват като входна информация след като бъдат интерполирани до растер с по-малък размер на клетката. В настоящата работа интерполираните метеорологични данни са с разделителна способност от 2 km, което е добър компромис между желанието за по-голяма детайлност и нуждата да се моделира реалистично пространствената вариация на метеорологичните показатели.

Получените в резултат от приложението на метода времеви серии на LAI могат да намерят приложение като входна информация в други модели. Например, един от подходите за подобряване на прогнозната точност на моделите на растежа на земеделските култури и прогнозиране на добивите от тях се основава на калибрирането им чрез времева серия от LAI получен по сателитни данни (Dorigo et al., 2007; Huang et al., 2019). Методът е използван в България от Kercheva, (2018) за калибриране на модела WOFOST в отделни пунктове като времевата серия от LAI е създадена именно чрез CSDM. С помощта на предложения тук подход се осигуряват необходимите данни, така че този метод на калибриране на WOFOST или друг подобен модел да бъде приложен върху определена територия. Mirasi et al., (2021) използват линеен регресионен модел ($R^2=0.858$) за да изчислят добива от пшеница на базата на сумата на NDVI, регистриран в няколко дати. По аналогия с тези и други резултати може да се очаква, че съществува връзка и между добива и сумата на LAI.

С помощта на предложения метод сумата на LAI може да бъде лесно изчислена върху дадена територия. На този етап ограничените данни за добива, с които разполагаме не позволи да се разработи статистически значим регресионен модел и са необходими допълнителни проучвания в тази посока. Изготвеният скрипт значително улеснява приложението на метода. На този етап той използва предварително подготвени входни данни, тоест интерполирани метеорологични данни и данни за LAI генерирани чрез изображения от Sentinel-2. Като част от продължението на работата в тази област в бъдеще е възможно изцяло да се автоматизира процедура, включително и предварителната обработка на входните данни. Необходимо е също така да се извършат допълнителни тестове върху по-голяма територия за да се прецени ефективността на скрипта и при необходимост да се оптимизира с цел намаляване на времето за изчисление.

Заклучение

Целта на изследването е да се моделират и картографират пространствената и времева динамика на набор от биофизични и биохимични параметри (LAI, fAPAR, fCover и CCC) на посеви от зимна пшеница на базата на многоканални сателитни изображения. За постигането на целта беше необходимо да се изгради цялостна методика Фигура 5, която поетапно първо да картографира посева и след това да проследи неговото развитие, моделирайки биофизичните показатели на базата на регресионни и механистични модели. Методиката е цялостно тествана в тестови район „Кнежа“ използвайки сателитни данни от Sentinel-2 и частично (поради ограничените данни) в тестови район „Требич“ използвайки сателитни данни от RapidEye.

Първата стъпка е картографирането на полетата със зимна пшеница. Тествани са три класификационни алгоритъма (SVM, RF и MXL). Те показват сходни точности, като точностите постигнати по данни от сателита Sentinel-2 са по-високи от тези постигнати от сателита RapidEye. Забелязва се и че непараметричните класификатори RF и SVM имат по-добро представяне пред класическия MXL. Резултатите от тестови район „Кнежа“ показват, че клас „Зимна пшеница“ може да бъде идентифициран с висока точност (F1 над 90%) чрез изображения от Sentinel-2 заснети в различни етапи от вегетацията. Резултатите от класификациите могат да бъдат подобрени в бъдеще. От особено значение е по-доброто разграничаване на зимната пшеница от останалите зимни култури, като зимна рапица.

Втора стъпка е моделирането на биофизични параметри върху двата тестови района. За район „Кнежа“, набор от 40 вегетационни индекса е тестван за тяхната полезност при предвиждане на LAI, fAPAR и fCover на зимна пшеница с помощта на данни от Sentinel-2. Сравнението е направено по 5 възможни сценария, зависещи от входните данни за моделиране. Съставени и валидирани са: общи модели (всички данни), модели за фаза братене, модели за фаза вретене, модели за сорт „Анапурна“ и модели за сорт „Енола“. Най-полезните ВИ могат да бъдат групирани в няколко типа: 1. индекси, които използват зелен и близко инфрачервен канал (CIg8, GIPVI и gNDVI); 2. индекси, базирани единствено на канали от диапазона червен ръб (CIre7, NDRE, NDRE1, SR3 и SR4); 3. индекси, които използват червен и близко инфрачервен канал (DVI, EVI, PVI, OSAVI, TSAVI) и 4. индексът MCARI/OSAVIre.

Във всички тези случаи, най-добре представящият се ВИ е придружен от няколко други ВИ, чиито модели показват подобна точност. Важно е да се отбележи, че най-добре представящите се ВИ не са еднакви за различните етапи на развитие на културата. В същото време трябва да се обърне внимание, че често няколко ВИ показват подобни един на друг резултати. Необходими са допълнителни изследвания, които да покажат дали настоящите резултати са валидни за различни райони и години и дали даден ВИ трябва да бъде предпочитан пред останалите. При всички случаи групите от ВИ показващи добри резултати в настоящото изследване могат да се използват като отправна точка при

бъдещите изследвания стеснявайки огромния диапазон от съществуващи индекси до по-малка група индекси с най-добър потенциал.

Установено е, че fAPAR и fCover могат да бъдат предвидени с относително добра точност, като се използват общи модели (rRMSE от 14 % и 23 % съответно). От друга страна, най-добрият общ модел за LAI не е достатъчно точен (rRMSE=45%), което показва ограничен потенциал за предвиждане на LAI на зимна пшеница с помощта на общ регресионен модел за всички фенофази. Установено е, че моделът на LAI за фенофаза братене е много по-добър при предвиждане на стойности на LAI под 1 и следва да се препоръча за използване в ранните етапи на развитие на културата. Също така моделът на LAI за фаза вретене има добра точност (rRMSE=26%) и не е повлиян от сорта зимна пшеница както общия модел.

За район „Требич“ е тестван набор от 18 ВИ изчислени по данни от сателита RapidEye. Най-полезните индекси за регресионно моделиране на LAI и CCC според резултатите в този тестови район са Clre и reNDVI. Въпреки това, дори тези индекси не позволиха съставянето на достатъчно точни, практически приложими регресионни модели за тези два показателя. Установихме, че reNDVI е най-добрият предиктор за fAPAR и fCover, но и няколко други индекса се представят почти еднакво добре. Както reNDVI, така и Clre използват канал от диапазона червен ръб, това потвърждава важността на спектралната област на червения ръб за характеризирание на състоянието на културата. Точността на регресионните модели за район „Требич“ е относително по-ниска. Необходима е допълнителна експериментална работа, за да се разберат по-добре източниците на несигурност.

При използване на продукти базирани на сателитни данни, чест проблем е липсата на данни поради облачна покривка, грешки в измерванията и други. Това налага използване на процедури за „реконструиране“ на данните. В изследването приложихме метода CSDM за интерполация на времева серия на LAI с дневна честота, моделът е изпълнен в Python. Той показва обещаващи резултати за запълване на липсващи данни за LAI, върху част от територията на тестови район „Кнежа“, но трябва да се приложи върху по-голяма територия и да се оптимизира кода, който е съставен от автора за по-бързо изчисление.

В резултат от приложението на съставената методика и включените в нея процедури по класифициране, моделиране и времева интерполация се получават няколко вида цифрови пространствени данни и карти, които служат за входни данни в отделните стъпки на методиката, но и сами по себе си са приложими като продукти за целите на мониторинг на посевите в национален мащаб. Продуктите получени чрез настоящата методика могат да бъдат използвани за по-нататъшно моделиране на добивите, дългосрочни анализи на вариацията на параметрите на посева в рамките на отделните полета и тяхното зонироване, актуални оценки на състоянието на посевите за оперативно вземане на решения и, като цяло, да бъдат интегрирани в една функционираща система за мониторинг използваща пълноценно сателитните данни.

Приноси

1. Демонстриран е подход за картографиране на посевите от зимна пшеница чрез двоична класификация на времеви композитни изображения от сателита Sentinel-2.
2. Съставени и валидирани са регресионни модели за оценка на биофизичните показатели LAI, fAPAR и fCover при зимна пшеница въз основа на спектрални данни от сателитите Sentinel-2 и RapidEye.
3. Предложен, тестван и софтуерно приложен е подход за проследяване на ежедневната динамика на LAI в течение на вегетационния сезон върху определена територия на базата на познатия от литературата модел CSDM.
4. Съставени са карти на посевите от зимна пшеница и техни биофизични и биохимични показатели в различни етапи от вегетационния сезон в тестови район „Кнежа“.

Публикации по дисертационния труд:

Kamenova, I., L. Filchev, I. Ilieva. 2017. **Review of spectral vegetation indices and methods for estimation of crop biophysical variables.** Aerospace Research in Bulgaria, Aerospace Research in Bulgaria, 29, Space Research Technology Institute - Bulgarian Academy of Sciences, 2017 , pp. 72-82

Kamenova, I., P. Dimitrov, R. Yordanova. 2018. **Evaluation of RapidEye vegetation indices for prediction of biophysical/biochemical variables of winter wheat.** Aerospace Research in Bulgaria, 30, Space Research Technology Institute - Bulgarian Academy of Sciences, 2018, pp. 63-74

Каменова, И. **Сравнение на класификационни алгоритми за идентифициране и картографиране на полета със зимна пшеница по данни от сателитите RapidEye.** Сборник доклади от конференция SES'2022 (приета за печат)

Благодарности

Изказвам най-искрена благодарност на своя научен ръководител доц. д-р Петър Димитров за всеотдайната подкрепа и методическата помощ, както и на всички колеги от секция “Дистанционни изследвания и ГИС” на ИКИТ–БАН за тяхното съдействие, напътствия и добро настроение. Благодаря на д-р Русина Йорданова за безвъзмездната помощ при събиране на полеви данни и лабораторна работа. Благодаря на проф. д-р Лъчезар Филчев, за предоставените данни от полеви спектрометър. Част от полевите изследвания за район „Кнежа“ са финансирани по проект “Тестване на вегетационни индекси от Sentinel-2 за оценка на състоянието на зимни култури в България (TS2AgroBg)” финансиран от ЕКА. Използваните в дисертационната работа сателитни изображения от RapidEye са предоставени от ЕКА по програмата за достъп до данни от трети страни (ESA's Third Party Missions programme), за което съм дълбоко признателна. Благодаря на семейството си за подкрепата и търпението.

Списък на цитираната в автореферата литература:

- Agovino, M., Casaccia, M., Ciommi, M., Ferrara, M., & Marchesano, K. (2019). Agriculture, climate change and sustainability: The case of EU-28. *Ecological Indicators*, 105, 525–543. <https://doi.org/10.1016/J.ECOLIND.2018.04.064>
- Baret, F. (1986). Contribution au suivi radiométrique de cultures de céréales. <http://www.theses.fr/1986PA112090>
- Belgiu, M., & Drăgu, L. (2016). Random forest in remote sensing: A review of applications and future directions. *ISPRS Journal of Photogrammetry and Remote Sensing*, 114, 24–31. <https://doi.org/10.1016/J.ISPRSJPRS.2016.01.011>
- Clevers, J., & Kooistra, L. (2012). Using hyperspectral remote sensing data for retrieving canopy chlorophyll and nitrogen content. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 5(2), 574–583. <https://doi.org/10.1109/JSTARS.2011.2176468>
- Cord, A. F., Brauman, K. A., Chaplin-Kramer, R., Huth, A., Ziv, G., & Seppelt, R. (2017). Priorities to Advance Monitoring of Ecosystem Services Using Earth Observation. *Trends in Ecology & Evolution*, 32(6), 416–428. <https://doi.org/10.1016/J.TREE.2017.03.003>
- Dahms, T., Seissiger, S., Borg, E., Vajen, H., Fichtelmann, B., & Conrad, C. (2016). Important variables of a RapidEye time series for modelling biophysical parameters of winter wheat. *Photogrammetrie, Fernerkundung, Geoinformation*, 2016(5–6), 285–299. <https://doi.org/10.1127/pfg/2016/0303>
- Dantas de Paula, M., Gómez Giménez, M., Niamir, A., Thurner, M., & Hickler, T. (2019). Combining European Earth Observation products with Dynamic Global Vegetation Models for estimating Essential Biodiversity Variables. <https://doi.org/10.1080/17538947.2019.1597187>, 13(2), 262–277. <https://doi.org/10.1080/17538947.2019.1597187>
- Delloye, C., Weiss, M., & Defourny, P. (2018). Retrieval of the canopy chlorophyll content from Sentinel-2 spectral bands to estimate nitrogen uptake in intensive winter wheat cropping systems. *Remote Sensing of Environment*. <https://doi.org/10.1016/j.rse.2018.06.037>
- Devos W., Lemoine G., & Milenov P. (2018). *Technical guidance on the decision to go for substitution of OTSC by monitoring*. <https://doi.org/10.2760/693101>
- Dimitrov, P., Kamenova, I., Roumenina, E., Filchev, L., Ilieva, I., Jelev, G., Gikov, A., Banov, M., Krasteva, V., Kolchakov, V., Kercheva, M., Dimitrov, E., & Miteva, N. (2019). Estimation of biophysical and biochemical variables of winter wheat through sentinel-2 vegetation indices. *Bulgarian Journal of Agricultural Science*, 25(5), 819–832.
- Dong, J., Fu, Y., Wang, J., Tian, H., Fu, S., Niu, Z., Han, W., Zheng, Y., Huang, J., & Yuan, W. (2020). Early-season mapping of winter wheat in China based on Landsat and Sentinel images. *Earth System Science Data*, 12(4), 3081–3095. <https://doi.org/10.5194/ESSD-12-3081-2020>
- Dorigo, W. A., Zurita-Milla, R., de Wit, A. J. W., Brazile, J., Singh, R., & Schaepman, M. E. (2007). A review on reflective remote sensing and data assimilation techniques for enhanced agroecosystem modeling. *International Journal of Applied Earth Observation and Geoinformation*, 9(2), 165–193. <https://doi.org/10.1016/j.jag.2006.05.003>
- Duveiller, G., Baret, F., & Defourny, P. (2011). Crop specific green area index retrieval from MODIS data at regional scale by controlling pixel-target adequacy. *Remote Sensing of Environment*, 115(10), 2686–2701. <https://doi.org/10.1016/J.RSE.2011.05.026>
- Eurostat. (2021). *Eurostat - Data Explorer*. <http://appsso.eurostat.ec.europa.eu/nui/submitViewTableAction.do>
- FAO. (2021). World Food and Agriculture – Statistical Yearbook 2021. *World Food and Agriculture – Statistical Yearbook 2021*. <https://doi.org/10.4060/CB4477EN>
- Gitelson, A. A., Viña, A., Ciganda, V., Rundquist, D. C., & Arkebauer, T. J. (2005). *Remote estimation of canopy chlorophyll content in crops*. 32(8), 1–4. <https://doi.org/10.1029/2005GL022688>
- Godfray, H. C. J., Beddington, J. R., Crute, I. R., Haddad, L., Lawrence, D., Muir, J. F., Pretty, J., Robinson, S., Thomas, S. M., & Toulmin, C. (2010). Food security: The challenge of feeding 9 billion people. *Science*, 327(5967), 812–818. <https://doi.org/10.1126/SCIENCE.1185383>
- Haboudane, D., Miller, J. R., Pattey, E., Zarco-Tejada, P. J., & Strachan, I. B. (2004). Hyperspectral vegetation indices and novel algorithms for predicting green LAI of crop canopies: Modeling and validation in the context of precision agriculture. *Remote Sensing of Environment*, 90(3), 337–352. <https://doi.org/10.1016/j.rse.2003.12.013>
- Han, J., Zhang, Z., Cao, J., Luo, Y., Zhang, L., Li, Z., & Zhang, J. (2020). Prediction of Winter Wheat Yield Based on Multi-Source Data and Machine Learning in China. *Remote Sensing 2020, Vol. 12, Page 236*, 12(2), 236.

- <https://doi.org/10.3390/RS12020236>
- Horowitz, C. A. (2016). Paris Agreement. *International Legal Materials*, 55(4), 740–755. <https://doi.org/10.1017/S0020782900004253>
- Huang, J., Gómez-Dans, J. L., Huang, H., Ma, H., Wu, Q., Lewis, P. E., Liang, S., Chen, Z., Xue, J. H., Wu, Y., Zhao, F., Wang, J., & Xie, X. (2019). Assimilation of remote sensing into crop growth models: Current status and perspectives. *Agricultural and Forest Meteorology*, 276–277, 107609. <https://doi.org/10.1016/J.AGRFORMET.2019.06.008>
- Huete, A. R. (1988). A soil-adjusted vegetation index (SAVI). *Remote Sensing of Environment*, 25(3), 295–309. [https://doi.org/10.1016/0034-4257\(88\)90106-X](https://doi.org/10.1016/0034-4257(88)90106-X)
- Jiao, X., McNairn, H., & Dingle Robertson, L. (2021). Monitoring crop growth using a canopy structure dynamic model and time series of synthetic aperture radar (SAR) data. *Remote Sensing of Environment*, 252, 112207. <https://doi.org/10.1016/j.rse.2021.112207>
- Jin, Z., Azzari, G., You, C., Di Tommaso, S., Aston, S., Burke, M., & Lobell, D. B. (2019). Smallholder maize area and yield mapping at national scales with Google Earth Engine. *Remote Sensing of Environment*, 228, 115–128. <https://doi.org/10.1016/J.RSE.2019.04.016>
- Kamenova, I., & Dimitrov, P. (2020). Evaluation of Sentinel-2 vegetation indices for prediction of LAI, fAPAR and fCover of winter wheat in Bulgaria. *European Journal of Remote Sensing*, 1–19. <https://doi.org/10.1080/22797254.2020.1839359>
- Kamenova, I., Dimitrov, P., & Yordanova, R. (2018). Evaluation of RapidEye vegetation indices for prediction of biophysical/biochemical variables of winter wheat. *Aerospace Research in Bulgaria*, 30, 63–74. <https://doi.org/10.3897/ARB.V30.E06>
- Kamenova, I., Filchev, L., & Ilieva, I. (2017). Review of spectral vegetation indices and methods for estimation of crop biophysical variables. *Aerospace Research in Bulgaria*, 29, 72–82. <https://doi.org/10.7546/AeReBu.29.18.01.06>
- Kercheva, M. (2018). *Report on yield prediction. Deliverable D4.3. TS2AgroBG project.*
- Koetz, B., Baret, F., Poilvé, H., & Hill, J. (2005). Use of coupled canopy structure dynamic and radiative transfer models to estimate biophysical canopy characteristics. *Remote Sensing of Environment*, 95(1), 115–124. <https://doi.org/10.1016/J.RSE.2004.11.017>
- Li, C., Chen, W., Wang, Y., Ma, C., Li, Y., Li, J., Zhai, W., Doughty, R., Bajgain, R., Wang, J., Li, C., Chen, W., Wang, Y., Wang, Y., Ma, C., Li, Y., Li, J., & Zhai, W. (2022). Mapping Winter Wheat with Optical and SAR Images Based on Google Earth Engine in Henan Province, China. *Remote Sensing*, 14(2), 284. <https://doi.org/10.3390/RS14020284>
- Merzlyak, M. N., Gitelson, A. A., Chivkunova, O. B., & Rakin, V. Y. (1999). Non-destructive optical detection of pigment changes during leaf senescence and fruit ripening. *Physiologia Plantarum*, 106(1), 135–141. <https://doi.org/10.1034/j.1399-3054.1999.106119.x>
- Mirasi, A., Mahmoudi, A., Navid, H., Valizadeh Kamran, K., & Asoodar, M. A. (2021). Evaluation of sum-NDVI values to estimate wheat grain yields using multi-temporal Landsat OLI data. *Geocarto International*, 36(12), 1309–1324. <https://doi.org/10.1080/10106049.2019.1641561>
- Pan, H., Chen, Z., Ren, J., Li, H., & Wu, S. (2019). Modeling Winter Wheat Leaf Area Index and Canopy Water Content With Three Different Approaches Using Sentinel-2 Multispectral Instrument Data. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 12(2), 482–492. <https://doi.org/10.1109/JSTARS.2018.2855564>
- Pask, A., Pietragalla, J., & Mullan, D. (2012). Physiological Breeding II: A Field Guide to Wheat Phenotyping. In *Chemistry & ...*. <https://doi.org/10.1017/CBO9781107415324.004>
- Phiri, D., Simwanda, M., Salekin, S., Nyirenda, V. R., Murayama, Y., & Ranagalage, M. (2020). Sentinel-2 data for land cover/use mapping: A review. *Remote Sensing*, 12(14), 2311. <https://doi.org/10.3390/RS12142291>
- Qu, C., Li, P., & Zhang, C. (2021). A spectral index for winter wheat mapping using multi-temporal Landsat NDVI data of key growth stages. *ISPRS Journal of Photogrammetry and Remote Sensing*, 175, 431–447. <https://doi.org/10.1016/J.ISPRSIPRS.2021.03.015>
- Roumenina, E., Naydenova, V., Jelev, G., Vassilev, V., & Krалева, L. (2009). Aerospace test sites in Bulgaria - state and prospects. *Aerospace Research in Bulgaria*, 23, 59–69. <https://ui.adsabs.harvard.edu/abs/2009ARBI...23...59R/abstract>
- Sellers, P. J., Schimel, D. S., Moore, B., Liu, J., & Eldering, A. (2018). Observing carbon cycle–climate feedbacks from space. *Proceedings of the National Academy of Sciences of the United States of America*, 115(31), 7860. <https://doi.org/10.1073/PNAS.1716613115>

- Shewry, P. R. (2009). Wheat. In *Journal of Experimental Botany* (Vol. 60, Issue 6, pp. 1537–1553). Oxford Academic. <https://doi.org/10.1093/jxb/erp058>
- Tilman, D., Cassman, K. G., Matson, P. A., Naylor, R., & Polasky, S. (2002). Agricultural sustainability and intensive production practices. *Nature* 2002 418:6898, 418(6898), 671–677. <https://doi.org/10.1038/nature01014>
- Toscano, P. (2020). Remote Sensing Applications for Agriculture and Crop Modelling. In *Remote Sensing Applications for Agriculture and Crop Modelling*. MDPI. <https://doi.org/10.3390/books978-3-03928-227-2>
- Vapnik, V. N. (1995). The Nature of Statistical Learning Theory. *The Nature of Statistical Learning Theory*. <https://doi.org/10.1007/978-1-4757-3264-1>
- Verrelst, J., Rivera, J. P., Veroustraete, F., Muñoz-Marí, J., Clevers, J. G. P. W., Camps-Valls, G., & Moreno, J. (2015). Experimental Sentinel-2 LAI estimation using parametric, non-parametric and physical retrieval methods – A comparison. *ISPRS Journal of Photogrammetry and Remote Sensing*, 108, 260–272. <https://doi.org/10.1016/j.isprsjprs.2015.04.013>
- Watson, D. J. (1937). The estimation of leaf area in field crops. *The Journal of Agricultural Science*, 27(3), 474–483. <https://doi.org/10.1017/S002185960005173X>
- Wiegand, C. L., Richardson, A. J., Escobar, D. E., & Gerbermann, A. H. (1991). Vegetation indices in crop assessments. *Remote Sensing of Environment*, 35(2–3), 105–119. [https://doi.org/10.1016/0034-4257\(91\)90004-P](https://doi.org/10.1016/0034-4257(91)90004-P)
- Yang, G., Yu, W., Yao, X., Zheng, H., Cao, Q., Zhu, Y., Cao, W., & Cheng, T. (2021). AGTOC: A novel approach to winter wheat mapping by automatic generation of training samples and one-class classification on Google Earth Engine. *International Journal of Applied Earth Observation and Geoinformation*, 102, 102446. <https://doi.org/10.1016/J.JAG.2021.102446>
- Руменина, Е., Желев, Г., Димитров, П., Филчев, Л., Каменова, И., Гиков, А., Банов, М., Кръстева, В., Керчева, М., & Колчаков, В. (2019). *Интегрирана система за дистанционно определяне на състоянието на посеви на земеделски култури* (Patent No. № 3283 U1 / 27.09.2019). Официален бюлетин на патентното ведомство на Република България. 11.1/15.11.2019.
- Руменина, Е., & Желев, Г. (2020). *Развитие на дистанционните изследвания, провеждани на аерокосмическите полигони в България*. 187–206.

Modeling and mapping biophysical and biochemical parameters of winter wheat with satellite images

Ilina Boyanova Kamenova, PhD thesis

Space Research and Technology Institute, Bulgarian Academy of Sciences; Acad. G. Bonchev St., bl. 1, Sofia 1113, Bulgaria, e-mail: ilina.kamenova@space.bas.bg
PhD supervisor: Assoc. Prof. Petar Dimitrov

Summary

The objective of this study is to estimate and map the spatial and temporal dynamics of a set of biophysical and biochemical parameters of winter wheat crops based on multispectral satellite images. For that purpose Sentinel-2 and RapidEye data will be used for prediction of set of parameters: Leaf Area Index (LAI), fraction of Absorbed Photosynthetically Active Radiation (fAPAR), fraction of vegetation Cover (fCover), leaf Chlorophyll Concentration (CC) and Canopy Chlorophyll Content (CCC) of winter wheat. Study areas are situated in Bulgaria. Knezha test site is part of Danubian plane, this area is one of the most intensively used agricultural regions in the country. The main soil type is chernozem and the winter wheat is the most prominent crop in the area. Trebich test site is located in the north-western part of Sofia municipality. The fields in both test sites are relatively large and are suitable for monitoring with satellite data.

In Trebich test site were conducted seven field campaigns during the growing seasons of 2014-2015 and 2015-2016. During the 2014-2015 growing season, 5 plots in two fields were sampled. During the 2015-2016 growing season, five fields were sampled in 11 plots. In both seasons, the fields were sown with winter wheat *Enola* variety. In Knezha test site were conducted four field campaigns in 2016-2017 growing season over six fields sown with *Annapurna* and *Enola* varieties. In the next growing season (2017-2018), another six fields sown with the same varieties were sampled during two field campaigns. We evaluated the utility of a selection of vegetation indices (VI) derived from Sentinel-2 and RapidEye multispectral data to derive Leaf Area Index (LAI), fraction of Absorbed Photosynthetically Active Radiation (fAPAR) and fraction of vegetation Cover (fCover) of winter wheat fields using the regression method.

In order to discriminate, the winter wheat fields from the rest of the agricultural fields we performed image classification procedure. Three classification methods (Support vector machines (SVM); Random Forest (RF) and maximum likelihood (MXL) were tested. They have shown similar accuracies, with slightly better results achieved from the Sentinel-2 data, than the Rapid Eye data. It is also noticed that the non-parametric classifiers RF and SVM have better performance than the classic parametric MXL. Results from Knezha test area show that the winter wheat class can be identified with high accuracy (F1 over 90%) using Sentinel-2 images acquired at different phenological stages. The earliest date when such an accuracy was obtained was March 30 (90.62% by the SVM method; tillering phase). The results also show

that in the pre-wintering period of winter wheat it is also possible to obtain a classification with high accuracy, although below 90%.

The next step in the study is modeling of biophysical and biochemical parameters in the two test areas. For Knezha region, a set of 40 vegetation indices was tested for their utility in predicting LAI, fAPAR and fCover of winter wheat using Sentinel-2 data. The comparison is made under 5 possible scenarios depending on the modeling inputs. We created and validated: general models (all data), tillering phase models, stem elongation phase models, *Annapurna* cultivar models and *Enola* cultivar models.

The most useful VIs can be grouped into several types:

- 1) VIs that use green and near-infrared bands (CIg8, GIPVI and gNDVI),
- 2) VIs based solely on red edge range bands (CIre7, NDRE, NDRE1, SR3 and SR4),
- 3) VIs that use red and near infrared bands (DVI, EVI, PVI, OSAVI, TSAVI) and
- 4) the MCARI/OSAVIre VI

For "Trebich" test site, a set of 18 RapidEye VIs was tested. They have been systematically evaluated and compared for their potential to predict LAI, fAPAR, fCover and CCC for winter wheat fields. The most useful indices for regression modeling of LAI and CCC according to the results in this test area are CIre and reNDVI. However, even these indices did not allow the construction of sufficiently accurate, practically applicable regression models for these indicators. We found reNDVI to be the best predictor of fAPAR and fCover.

When using products derived from satellite data, a common problem is missing data due to cloud cover, measurement errors, and others. This necessitates the use of data "reconstruction" procedures. In the study, we applied the CSDM method to interpolate LAI time series with daily frequency, the model is implemented in Python. It shows promising results for filling missing LAI data. The agrometeorological data used to feed the model show good applicability for this type of modeling. This reveals the potential for the integration of satellite data with other types of publicly available spatial data to improve the modeling of various aspects of agricultural crops.

As a result of the methodology application, which comprises image classification phase, modeling parameters phase and temporal interpolation procedures, several types of digital spatial data and maps are obtained. Each result serves as input data in the individual steps of the methodology, but they are also applicable as products in themselves for the purposes of monitoring crops on a national scale. The products obtained through the present methodology can be used for further modeling of yields, long-term analyzes of the variation of crop parameters within individual fields and their zoning, up-to-date assessments of crop conditions for operational decision-making and, in general, to be integrated into a functioning monitoring system making full use of satellite data.