

КОРЕЛАЦИОНЕН АНАЛИЗ НА NDVI-ВРЕМЕВИ СЕРИИ ЗА КАРТОГРАФИРАНЕ НА ЗЕМЕДЕЛСКИ КУЛТУРИ

Петър Димитров, Илина Каменова, Георги Желев

*Институт за космически изследвания и технологии – Българска академия на науките
e-mail: petar.dimitrov@space.bas.bg;*

Ключови думи: *разпознаване на земеделски култури, годишна динамика на NDVI, PROBA-V*

Резюме: *Обработката на времеви серии от спътникови NDVI изображения е ефективен метод за картографиране на земеделските култури. Той разчита на възможността за разпознаване на културите на базата на фенологичните им особености, отразени в годишния ход на NDVI. Докладът има за цел да тества управляван класификационен подход, при който времевата серия от NDVI изображения се сравнява, пиксел по пиксел, с набор от референтни NDVI-времеве серии на търсените класове. Пикселът приема този клас, до чиято референтна времева серия е най-близък, като критерият за близост е коефициентът на корелация. Методът е приложен върху времева серия съставена от 14 еднодневни PROBA-V изображения заснети в периода март-октомври 2014 г. с цел да бъдат картографирани следните класове: пшеница, рапица, царевица, слънчоглед, тютюн и рапица плюс втора култура царевица. Общата точност на класификацията е 70,1%.*

CORRELATION ANALYSIS OF TIME SERIES NDVI DATA FOR CROP MAPPING

Petar Dimitrov, Iliana Kamenova, Georgi Jeleu

*Space Research and Technology Institute – Bulgarian Academy of Sciences
e-mail: petar.dimitrov@space.bas.bg;*

Keywords: *crop identification, within season NDVI variation, PROBA-V*

Abstract: *NDVI time series analysis, using satellite imagery is an effective tool for crop mapping. It is based on the possibility to distinguish between crops using their phenological differences as represented in the annual NDVI profile. The aim of this paper is to test a supervised classification approach where the NDVI time series is compared, pixel wise, with a set of reference NDVI time series of the classes to be mapped. Each pixel is assigned to the most similar reference class. The measure of similarity is the coefficient of correlation. The method is applied to map six classes (wheat, rapeseed, maize, sunflower, tobacco and double crop) making use of 14 daily PROBA-V images collected between March and October 2014. The overall accuracy of the classification is 70.1%.*

Въведение

Възможността за получаване на изображения на земната повърхност през кратки интервали от време е едно от ключовите предимства на спътниковите дистанционни методи, позволяващо наблюдението на сезонната динамика на растителността. Оптичните спътникови сензори позволяват да се проследи изменението на количеството зелена биомаса през вегетационния период. Традиционно използвана мярка за количеството на зелената биомаса са спектралните индекси, като например индекса NDVI (Normalized Difference Vegetation Index, Rouse et al., 1974; Roumenina et al., 2010). Сензори като MODIS, AVHRR и PROBA-VEGETATION предоставят NDVI данни през интервали от порядъка на десетдневие или месец, което освен за фенологични изследвания е полезно за картографиране на типовете земно покритие и растителност. Годишният ход на NDVI, т.нар. времева траектория, в много случаи има характерна форма при различните растителни съобщества или видове, включително при

отделните земеделски култури, и може да се използва за тяхното разпознаване по спътниковите времеви серии (Evans, Geerken 2006; Doraiswamy et al., 2007).

В литературата се срещат различни подходи използващи времеви серии от NDVI за разпознаване на типове растителност, земно покритие или земеделски култури. Например Vassilev (2013) прилага алгоритъма за клъстеризация ISODATA върху NDVI времева серия от десетдневни композитни изображения от сензора SPOT-VEGETATION за да разграничи четири основни култури на територията на България. Като предварителна стъпка при класификация на земното покритие в глобален мащаб Loveland et al. (2000) използват неуправляван алгоритъм за клъстеризация на месечни композитни NDVI изображения от AVHRR. Формираните в резултат на клъстеризацията спектрални класове (т.нар. „greenness classes“) представляват райони, хомогенни по отношение на сезонния ход на развитие на растителността. Друг подход за анализ на времеви серии от NDVI е да се извлекат „фенологични“ променливи като например максимален годишен NDVI, кумулативен годишен NDVI, момент на настъпване на максималния NDVI, NDVI на най-топлия месец и др., които да се използват в последваща класификация. Подобни променливи, калкулирани по месечни AVHRR данни, използват Hansen et al. (2000) в своята класификация на земното покритие в света. Сходна стратегия прилагат Moody и Johnson (2001) използвайки неуправлявана класификация, но вместо самите AVHRR NDVI изображения класифицират фазата и амплитудата на първите два хармоника получени при преобразуването на времевата серия чрез анализ на Фурие. Сред прилаганите методи за обработка на времеви серии от NDVI могат да се споменат още анализ на главните компоненти и управлявана класификация по метода на максималното правдоподобие (Townshend et al., 1987).

Geerken et al. (2005) предлагат подход за класифициране на времеви серии, в основата на който стои коефициентът на линейна корелация. Корелацията се използва за да се определи сходството на всеки пиксел с предварително определени референтни времеви серии на търсените класове. Методът е приложен от авторите за полу-пустинна територия с тревно-хростова растителност. Използвани са 36 десетдневни NDVI изображения, като времевите серии са филтрирани чрез анализ на Фурие преди класифицирането. Настоящото изследване е мотивирано от въпроса дали опростен вариант на корелационния подход на Geerken et al. (2005) може успешно да бъде приложен за класифициране на земеделски култури, както и дали методът е приложим при времева серия с малък брой изображения. За целта е използвана NDVI-времева серия от спътника PROBA-V съставена от избрани, безоблачни изображения в рамките на вегетационния сезон на 2014 г., с помощта на която са класифицирани шест земеделски типа земно покритие: „пшеница“, „рапица“, „царевица“, „слънчоглед“, „тютюн“ и „рапица/царевица“.

Материали и методи

Район на изследване и полеви данни. Използваните в настоящото изследване полеви данни са събрани в рамките на проект, осъществен през 2014 г., целящ да демонстрира възможностите на изображенията от спътника PROBA-V с пространствена разделителна способност (ПРС) 100 m за разпознаване на земеделски култури (Roumenina et al., 2015). Данните включват ГИС слой със 105 полигона с атрибутивна информация за засетите култури през селскостопанската година 2013/2014 (съседни полета засети с една и съща култура са обединени в един полигон). Основната част от полетата са засети с пшеница, рапица, слънчоглед и царевица. Две полета са засадени с тютюн, а в едно поле са отгледани две култури през сезона – рапица и царевица. Полетата засети с люцерна са с много малки размери, поради което тези култура е изключена от по-нататъшния анализ. Полетата се намират в западната част на Дунавската равнина в района на гр. Кнежа. Информация за почвените и климатичните условия в района може да се намери в Roumenina et al. (2015).

Спътникови NDVI данни. Основната цел на програмата PROBA-V (Project for On-Board Autonomy – Vegetation) е осъществяването на непрекъснат мониторинг на растителността и земното покритие в глобален мащаб (Dierckx et al., 2014). За териториите северно и южно от приблизително 40° г. ш. спътникът PROBA-V предоставя пълно покритие с данни със 100 m ПРС на всеки пет дни (Wolters, Dierckx, Swinnen, 2015). Наскоро бяха демонстрирани предимствата на тези данни при картографирането на земеделски култури (Roumenina et al., 2015), поради което те са използвани и в настоящото изследване. Използваните NDVI изображения са изтеглени във вид на готов продукт (S1 TOC NDVI 100 m) от портала на Фламандския институт за технологични изследвания (VITO)¹. Избрани са общо 14 изображения получени в периода от края на март до началото на октомври 2014 г. (Фиг. 1), които са

¹ VITO Product Distribution Portal - <http://www.vito-eodata.be/PDF/portal/Application.html#Home>

безоблачни или с незначителна облачност над изследвания район. Те са обединени в един общ многослоен файл представляващ времева серия с неравни интервали между изображенията. Пикселите заети от облаци дори в една от датите са маскирани и изключени от анализа. Териториите извън обработваемите земи също са маскирани.

Анализ. Многослойният NDVI файл е класифициран с цел да бъдат отделени следните шест класа: „пшеница“, „рапица“, „царевица“, „слънчоглед“, „тютюн“ и „рапица/царевица“ (рапица плюс царевица като втора култура). За целта е приложен метод сходен с тази на Geerken et al. (2005), при който се анализира степента на корелация между NDVI времевата серия и референтни NDVI времеви серии на търсените класове. Процедурата включва следните стъпки:

1. *Създаване на референтни NDVI времеви серии.* В многослойния NDVI файл са селектирани пиксели - представители на търсените класове. За целта са използвани границите на полетата, за които е налична наземна информация за земеделската култура. Чрез данните от тези „обучаващи“ пиксели са получени осреднени NDVI времеви серии, които служат като референтни за шестте класа.

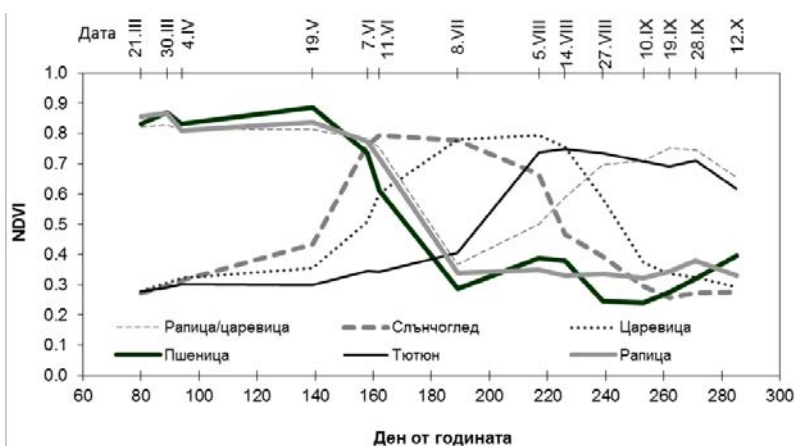
2. *Корелационен анализ.* Всеки пиксел в многослойния NDVI файл се разглежда като самостоятелна времева серия като последователно се изчислява нейният коефициент на корелация (r) с шестте референтни времеви серии. Получават се шест растера – по един за всеки клас, като стойностите в тях представляват корелационни коефициенти вариращи от -1 до +1. Тези растери показват доколко годишният ход на NDVI на всеки пиксел се доближава до референтния годишен ход на NDVI на съответния клас.

3. *Класификация.* Принадлежността на всеки пиксел към един от класовете се определя като се избира класът, с чиято референтна NDVI времева серия корелацията е най-висока (тоест r е най-близо до +1).

За оценката на точността на класификацията е използвана случайна извадка от пиксели. Принадлежността им към даден клас е определена чрез визуално дешифриране на спътникови изображения от Landsat 8. Използвани са общо осем изображения заснети в периода 23 март – 1 октомври 2014 г., което позволява при дешифрирането да се отчетат фенологичните особености на класовете. За класа „рапица/царевица“ е ползвана и информацията от времевата серия от PROBA-V NDVI изображения, като към този клас са отнесени референтни пиксели, за които вторият пик през сезона има NDVI над 0,6.

Резултати

Успешното приложение на корелационния метод за картографиране на земеделските култури в изследвания район зависи от това доколко шестте класа имат характерен, отличителен годишен ход на NDVI. Референтните времеви серии на класовете са представени графично на Фиг. 1. Четиринадесетте PROBA-V изображения са сравнително равномерно разпределени през вегетационния период. Това позволява да се наблюдават най-общите особености в годишния ход на NDVI при отделните класове. За двете зимни култури - пшеница и рапица, времевите серии обхващат само периода на вегетация след презимуването им. При тях NDVI има високи стойности в началото на времевата серия (края на март) тъй като те продължават вегетационното си развитие започнало през предходната есен. Кривите на изменение на NDVI за полетата с пшеница и рапица имат сходна форма. Разликите наблюдавани в двете криви в периода след като тези култури са вече прибрали се дължат на различните състояния, в които се намират ожънатите полета – стърнища, плевели, угар и др. и следователно не са показател за вида на предшестващата култура. Слънчогледът и царевицата имат NDVI криви с типична за пролетните култури форма. При първата култура обаче нарастването на NDVI е с по-бързи темпове, максималните му стойности се достигат по-рано през годината и понижението на NDVI свързано с приключването на вегетационното развитие започва по-рано. Характерната форма на кривата на NDVI при тютюна е свързана с късните дати на засаждане на тази култура. Класът „рапица/царевица“ има два ясно изразени максимума на NDVI през сезона. От Фиг. 1 се вижда, че повечето класове се отличават с характерен NDVI профил. Трудности могат да се очакват при разграничаването на двете зимни култури. Техните референтни времеви серии са силно корелирани помежду си ($r=0,98$). Корелацията между референтните времеви серии на слънчогледа и царевицата е значително по-ниска ($r=0,77$).



Фиг. 1. Референтни NDVI времеви серии на шестте класа по данни от PROBA-V 100 m. В горната част на графиката са показани датите на използваните 14 изображения от 2014 г.

Резултатът от приложението на корелационния метод за класификация върху времевата серия от PROBA-V е представен на Фиг. 2.A. Според класификацията най-голямо разпространение имат класовете „пшеница“, „слънчоглед“ и „царевица“. Към тях се отнасят 78% от всички класифицирани пиксели. Сравнително голяма площ заема и клас „рапица“. Останалите два класа заемат само 6% от територията включена в класификацията. Това съотношение между класовете е реалистично с оглед на специализацията на растениевъдството в района (МЗХ 2014). Резултатите от оценката на точността на класификацията са представени в Табл. 1. Общата точност на класификацията е 70,1 %. Точността по класове е най-висока за царевицата, слънчогледа и класа „рапица/царевица“, като варира от 74% до 84% (Табл. 1). Най-ниска е точността за пшеницата и рапицата, което е свързано с посоченото вече сходство в референтните времеви серии на двете култури. По-малко от половината пиксели класифицирани като рапица реално се отнасят към този клас (Табл. 1). Грешката се дължи на смесването с клас „пшеница“. Следователно разпространението на рапицата е преувеличено за сметка на пшеницата. Точността на клас „тютюн“ не е оценена поради липсата на референтни данни. Това се дължи на слабото разпространение на класа в района на изследване.

Трябва да се има предвид, че използването на данни получени чрез визуална интерпретация, а не такива събрани на терена, винаги крие риск от неточности при оценката на точността на класификацията. При липсата на други референтни данни обаче, този подход е добра алтернатива и ни позволява да получим поне ориентировъчна количествена представа за надеждността на класификацията и да я съпоставим с други подобни изследвания. Общата точност на класификацията получена чрез тествания корелационен метод е близка до точността от 73,9% постигната от Geerken et al. (2005) в тяхното изследване използващо аналогичен корелационен подход. Roumenina et al. (2015) постигат точности от 75% до 77% при класификация на земеделски култури използвайки данни от PROBA-V с разделителна способност 100 m и традиционни класификационни подходи като метод на максималното правдоподобие и ISODATA.

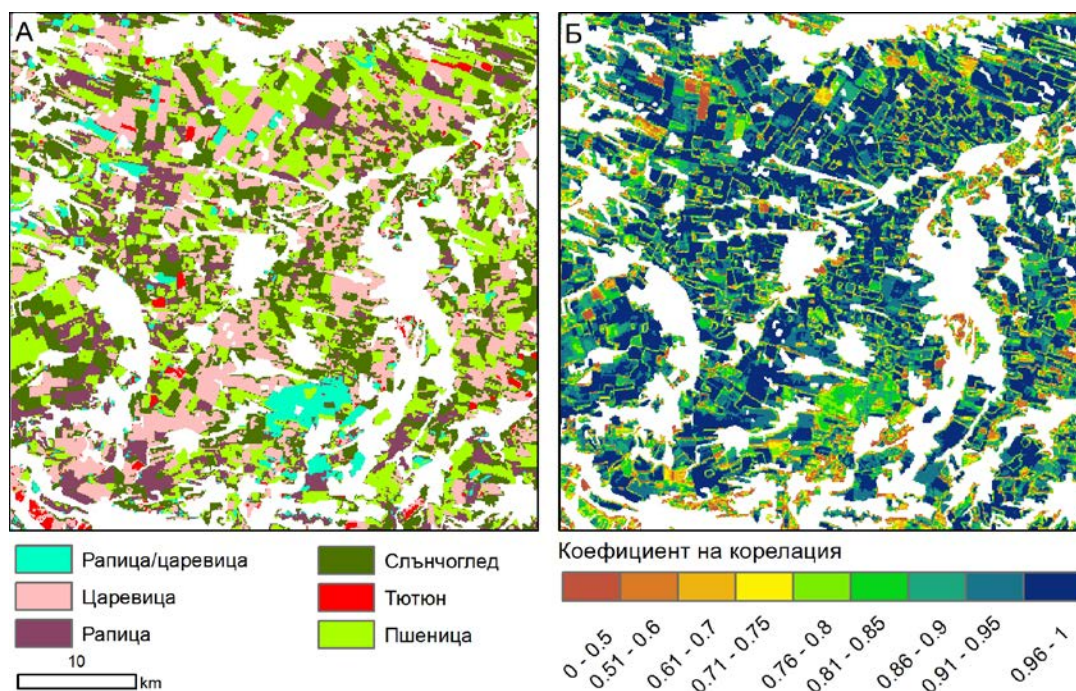
За целите на настоящото изследване е прието, че всички обработваеми земи в района могат да бъдат отнесени към един от шестте класа. В действителност в района се отглеждат и други земеделски култури, макар и на не големи площи. За това говори и фактът, че има полета, в които пикселите не са силно положително корелирани с нито една от референтните времеви серии (Фиг. 2.B.). Тези полета имат годишен ход на NDVI значително различаващ се от референтните времеви серии и вероятно част от тях са засети с култури или типове земно покритие различни от разглежданите в това изследване. При липсата на полеви данни за тези култури те не могат да бъдат идентифицирани и внасят грешка в останалите класове. Този проблем е избегнат от Geerken et al. (2005), които първоначално създават отделни карти на разпространението на всеки клас. Това става чрез определяне на минимални стойности на корелационния коефициент, при които даден пиксел е позволено да бъде причислен към съответен клас. Тези прагове се прилагат към растерите получени в стъпка 2 от настоящата процедура (виж „Материали и методи. Анализ“), за да бъдат маскирани пикселите не отговарящи на критериите за принадлежност към съответния клас. Отделните растери в последствие се комбинират, при което пикселите които са маскирани във всичките растери остават не класифицирани. Ако даден пиксел удовлетворява критерия на повече от един клас се причисляват към класа с най-висока корелация.

Таблица 1. Резултати от проверката на точността на класификацията базирана на корелацията между времеви серии

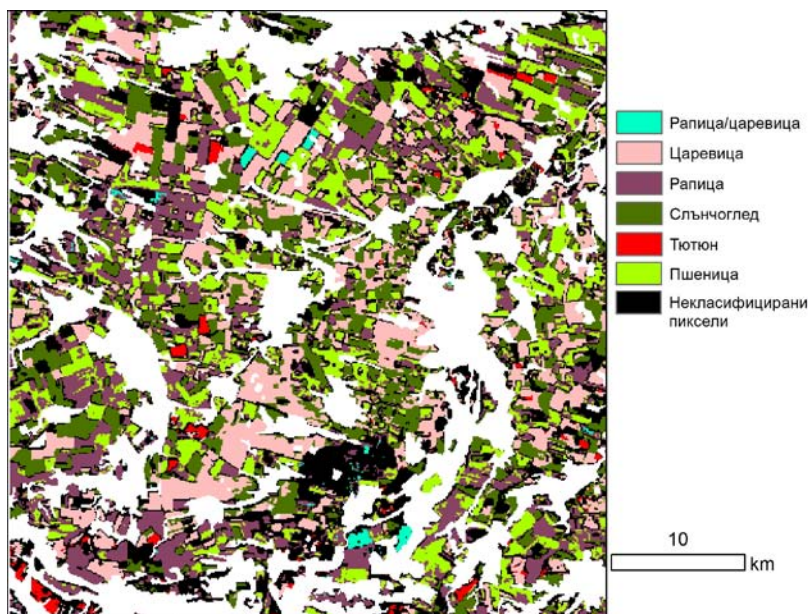
Класификация	Референтни данни							User Acc.(%)
	Царевица	Рапица	Слънчоглед	Рапица/царевица	Пшеница	Тютюн	Общо	
Царевица	62	0	8	0	4	0	74	83,8
Рапица	0	17	4	0	19	0	40	42,5
Слънчоглед	8	1	54	1	9	0	73	74,0
Рапица/царевица	0	1	0	10	2	0	13	76,9
Пшеница	9	8	1	1	52	0	71	73,2
Тютюн	4	0	3	0	0	0	7	
Общо	83	27	70	12	86	0		
Prod. Acc.(%)	74,7	63,0	77,1	83,3	60,5			

Трудностите при този подход са свързани с избора на най-подходящи прагови стойности на корелационния коефициент. В настоящото изследване беше направен опит за определяне на такива стойности използвайки набор от тестови пиксели съставен чрез полевите данни. Целта беше да се избере стойност на корелационния коефициент при която над 90% от тестовите пиксели на съответния клас да са „разпознати“. Резултатите от тази класификация са показани на Фиг. 3. При този вариант на класификацията една четвърт от пикселите не са класифицирани, тоест не са достатъчно силно корелирани с нито една от референтните NDVI времеви серии. Вижда се, че освен не класифицирани полета (компактни групи от пиксели) има и множество отделни пиксели или ивици от пиксели по границите на полетата които също не са класифицирани. Това са смесени пиксели, които нямат еднозначен фенологичен сигнал.

Броят на не класифицираните пиксели зависи от избраните прагови стойности и е толкова по-голям, колкото по-високи са те. Ако бъдат избрани прекалено високи прагове на корелационния коефициент има опасност пиксели, които реално принадлежат към даден клас да бъдат пропуснати поради естествената вариация на времевите серии в рамките на класа. Колкото по-ниски са праговете, толкова повече резултатите от класификацията ще се доближават до тези на Фиг. 2.А. В конкретния случай е трудно да преценим дали избраните прагове са подходящи. Тъй като референтните данни са събрани допускайки наличието в изследвания район само на шестте класа те не могат да се използват за проверка на точността на класификацията получена с прилагане на прагови стойности.



Фиг. 2. (А) Класификация на NDVI времева серия от спътника PROBA-V с ПРС 100 m на базата на корелацията с референтни NDVI времеви серии. (Б) Максимална стойност на коефициента на корелацията получена при корелацията с шестте референтни NDVI времеви серии.



Фиг. 3. Класификация на NDVI времева серия от спътника PROBA-V с ПРС 100 m на базата на корелацията с референтни NDVI времеви серии с приложени прагови стойности на корелационния коефициент.

Заклучение

Резултатите от изследването показват, че класификационният метод основан на корелацията между NDVI-времеви серии има значителен потенциал при картографиране на земеделски култури и стопански практики като отглеждане на две култури през сезона. Може да се счита, че коефициентът на корелация е достатъчно ефективна мярка за сходството между NDVI-времеви серии дори когато броят на наблюденията в тях е малък. Именно такъв е случаят в това изследване, където само 14 безоблачни изображения са налични през вегетационния период. Въпреки това трябва да се има предвид, че освен броят на изображенията от значение е и доколко те са равномерно разпределени във времето и отразяват характерния годишен ход на NDVI на различните култури. Липсата на данни в продължителни интервали от време или в ключови за разпознаването на културите моменти от годината може да намали приложимостта на корелационния метода на класификация. Получената в конкретния случай обща точност на класификацията от 70,1 % не е висока, но е съизмерима с точността постигната от други класификационни методи при използването на същите или сходни спътникови данни. Необходими са допълнителни изследвания, които да разкрият пълните възможности на метода. Това може да включва сравнение с други класификационни методи приложени при едни и същи условия.

Благодарности

Изказваме своята благодарност на проф. Евгения Руменина и г-н Илчовски за предоставените полеви данни.

Литература:

1. Dierckx, W., S. Sterckx, I. Benhadj, S. Livens, G. Duhoux, T. Van Achteren, M. Francois, K. Mellab, G. Saint. PROBA-V mission for global vegetation monitoring: standard products and image quality. *International Journal of Remote Sensing*, 35 (7), 2014, pp. 2589-2614
2. Doraiswamy, P., B. Akhmedov, A. Stern. Crop classification in the U.S. corn belt using MODIS imagery. In *Proceedings of the International Geoscience and Remote Sensing Symposium (IGARSS)*, July 2007.
3. Evans, J.P., R. Geerken. Classifying rangeland vegetation type and coverage using a Fourier component based similarity measure. *Remote Sensing of Environment* 105, 2006, pp.1-8
4. Fleiss, J.L., B. Levin, & M.C. Paik. *Statistical methods for rates and proportions*, 3rd edition. New Jersey: Wiley, 2003
5. Foody, G.M. Thematic map comparison: Evaluating the statistical significance of differences in classification accuracy. *Photogrammetric Engineering and Remote Sensing*, 70, 2004, pp. 627-633.
6. Geerken, R., B. Zaitchik and J. P. Evans. Classifying rangeland vegetation type and coverage from NDVI time series using Fourier Filtered Cycle Similarity. *International Journal of Remote Sensing*, 26 (24), 2005, 5535-5554

7. Hansen, M.C., DeFries, R.S., Townshend, J.R.G., Sohlberg, R. Global land cover classification at 1 km spatial resolution using a classification tree approach. *International Journal of Remote Sensing* 21 (6_7), 2000, 1331–1364.
8. Loveland, T. R., Reed, B. C., Brown, J. F., Ohlen, D. O., Zhu, Z., Yang, L., Merchant, J.W. Development of a global land cover characteristics database and IGBP DISCover from 1 km AVHRR data. *International Journal of Remote Sensing*, 21(6–7), 2000, 1303–1330.
9. Moody, A., D.M. Johnson. Land-surface phenologies from AVHRR using the discrete fourier transform. *Remote Sensing of Environment* 75, 2001, 305–323
10. Richards, J.A., X. Jia. *Remote sensing digital image analysis. An introduction.* 4th Edition. Springer, 2006, 439p.
11. Roumenina E., C. Atzberger, V. Vassilev, P. Dimitrov, I. Kamenova, M. Banov, L. Filchev and G. Jelev. Single- and Multi-Date Crop Identification Using PROBA-V 100 and 300 m S1 Products on Zlatia Test Site, Bulgaria. *Remote Sens.* 7, 2015, pp.13843-13862
12. Roumenina, E., L. Filchev, V. Naydenova, G. Jelev, P. Dimitrov, V. Vassilev, L. Krалева. 2010. Monitoring of Winter Crop Status in Bulgaria Using a Series of NOAA AVHRR NDVI Images. *Canadian Journal of Remote Sensing*, Vol. 36, Suppl. 2, pp. 224-230, ISSN 1712-7971
13. Rouse, J.W., R.H. Haas, J.A. Schell, D.W. Deering. Monitoring vegetation systems in the Great Plains with ERTS. In Proc. of the 3rd. ERTS-1 Symposium (Eds. S. Freden, E. Mercanti & M. Becker), 1974, Vol.1, Sect. A, 309–317, NASA, Washington, D.C., USA.
14. Swain, P. H., S.M. Davis, Eds., *Remote Sensing: The Quantitative Approach.* New York: McGraw-Hill, 1978
15. Townshend, J. R. G., Justice, C. O., & Kalb, V. Characterization and classification of South American land cover types using satellite data. *International Journal of Remote Sensing*, 8, 1987, 1189–1207
16. Vassilev, V. Crop monitoring using SPOT-VGT NDVIs S10 time-series product for the arable land of Bulgaria. *Aerospace Research in Bulgaria*, 25, 2013, pp. 172-182
17. Wolters, E., W. Dierckx, E. Swinnen. *PROBA-V Products User Manual v1.3.* VITO, 2015, 81p.
18. МЗХ (Министерство на земеделието и храните). 2014. *Агростатистически справочник 2000-2012.* 209с.