

**Univerzita Palackého v Olomouci  
Přírodovědecká fakulta  
Katedra geoinformatiky**

**GEOINFORMATICKÉ PŘÍSTUPY PŘI  
VÝZKUMU DIVERZITY PLANÝCH  
PŘEDCHŮDCŮ KULTURNÍCH  
BOBOVITÝCH ROSTLIN**

**Diplomová práce**

**Bc. Přemysl DRATVA**

**Vedoucí práce RNDr. Jan Brus, Ph.D.**

**Olomouc 2020  
Geoinformatika**

## **ANOTACE**

Práce se zabývá aplikací geoinformatických přístupů při výzkumu diverzity planých předchůdců bobovitých rostlin, konkrétně cizrny, čočky a hrachu v oblasti Úrodného půlměsíce. Hlavním cílem práce je zjištění vlivu environmentálních charakteristik na výskyt planých předchůdců cizrny, čočky a hrachu. Teoretická část práce obsahuje popis environmentálních charakteristik oblasti Úrodného půlměsíce a Turecka, popis zkoumaných rostlin a datových sad použitých v práci, popis platformy Google Earth Engine, ze které byla získána velká část použitých dat, a nakonec je zde obsažen také současný stav řešené problematiky.

V praktické části práce jsou využity dva typy dat. Prvním typem jsou data o výskytu zkoumaných rostlin. Tato data byla získána ze světových genobank a dále terénním sběrem pracovníky Katedry botaniky UP, který probíhal v letech 2013–2015 v jihovýchodní části Turecka. Tato data nejprve prošla selekcí, během které byly některé záznamy odstraněny. Druhým typem použitých dat jsou globální sady environmentálních dat, mezi které patří WorldClim, SRTM30 a několik datových sad získaných prostřednictvím platformy Google Earth Engine. Jedním z výstupů práce jsou vytvořené vzorové skripty, které slouží k získání dat z této platformy pro účely studií s podobným zaměřením.

Ze získaných datových sad byly pro každý druh rostlin vybrány environmentální faktory, které nepodléhají vzájemné korelaci a zároveň mají v daných datech největší zastoupení. Na základě těchto vybraných faktorů byly následně vytvořeny shluky rostlin, které byly popsány na základě geografického rozmístění a také jejich ovlivnění environmentálními faktory. U dat čočky bylo navíc možné tyto shluky porovnat s výsledky genotypování. Práce obsahuje také zjištění vlivu přesnosti bodových dat rostlin na základě environmentálních datových sad různé kvality.

## **KLÍČOVÁ SLOVA**

Plané rostliny; bobovité rostliny; Google Earth Engine; PCA

Počet stran práce: 77

Počet příloh: 8 (z toho 2 volné)

## **ANOTATION**

The work deals with the application of geoinformatic approaches in the research of the diversity of wild ancestors of legumes, specifically chickpeas, lentils and peas in the Fertile Crescent. The main goal of this work is to determine the influence of environmental characteristics on the occurrence of wild precursors of chickpeas, lentils and peas. The theoretical part of the thesis contains a description of the environmental characteristics of the Fertile Crescent and Turkey, a description of the investigated plants and datasets used in the work as well as a description of the Google Earth Engine platform, from which a large part of the data was obtained.

In the practical part of the work, two types of data are used. The first type is data on the occurrence of the studied plants. These data were obtained from world genobanks and also by field collection by the staff of the UP Department of Botany, which took place in 2013–2015 in the southeastern part of Turkey. This data first went through a selection, during which some records were deleted. The second type of data used are global environmental data sets, which include WorldClim, SRTM30, and several data sets obtained through the Google Earth Engine platform. One of the outputs of the work are the sample scripts created, which are used to obtain data from this platform for the purposes of studies with a similar focus.

From the obtained data sets, for each plant species environmental factors were selected that are not subject to mutual correlation and at the same time are the most frequent ones in the given data. Based on these selected factors, plant clusters were subsequently created, which were described on the basis of geographical distribution and their influence by environmental factors. In addition, for lentil data, it was possible to compare these clusters with genotyping results. The work also contains a determination of the influence of the accuracy of plant point data based on environmental data sets of varying quality.

## **KEYWORDS**

Leguminous plants; wild plants; Google Earth Engine; PCA

Number of pages: 77

Number of appendixes: 8

**Prohlašuji, že**

- diplomovou práci včetně příloh, jsem vypracoval samostatně a uvedl jsem všechny použité podklady a literaturu.
- jsem si vědom, že na moji diplomovou práci se plně vztahuje zákon č.121/2000 Sb. - autorský zákon, zejména § 35 – využití díla v rámci občanských a náboženských obřadů, v rámci školních představení a využití díla školního a § 60 – školní dílo,
- beru na vědomí, že Univerzita Palackého v Olomouci (dále UP Olomouc) má právo nevýdělečně, ke své vnitřní potřebě, diplomovou práci užívat (§ 35 odst. 3),
- souhlasím, aby jeden výtisk diplomové práce byl uložen v Knihovně UP k prezenčnímu nahlédnutí,
- souhlasím, že údaje o mé diplomové práci budou zveřejněny ve Studijním informačním systému UP,
- v případě zájmu UP Olomouc uzavřu licenční smlouvu s oprávněním užití výsledky a výstupy mé diplomové práce v rozsahu § 12 odst. 4 autorského zákona,
- použít výsledky a výstupy mé diplomové práce nebo poskytnout licenci k jejímu využití mohu jen se souhlasem UP Olomouc, která je oprávněna v takovém případě ode mne požadovat přiměřený příspěvek na úhradu nákladů, které byly UP Olomouc na vytvoření díla vynaloženy (až do jejich skutečné výše).

V Olomouci dne \_\_\_\_\_

Bc. Přemysl Dratva

\_\_\_\_\_  
podpis autora

Děkuji vedoucímu práce RNDr. Janu Brusovi, Ph.D. za skvělé vedení, přínosné konzultace a motivující přístup při vypracování této diplomové práce. Dále bych rád poděkoval doc. Ing. Petru Smýkalovi, Ph.D. za poskytnutá data a za odborné připomínky týkající se planých bobovitých rostlin a také Mgr. Karlovi Macků za užitečné rady při konzultacích ohledně statistických analýz.

UNIVERZITA PALACKÉHO V OLOMOUCI  
Přírodovědecká fakulta  
Akademický rok: 2018/2019

## ZADÁNÍ DIPLOMOVÉ PRÁCE

(PROJEKTU, UMĚLECKÉHO DÍLA, UMĚLECKÉHO VÝKONU)

Jméno a příjmení: **Bc. Přemysl DRATVA**  
Osobní číslo: **R18862**  
Studijní program: **N1301 Geografie**  
Studijní obor: **Geoinformatika**  
Název tématu: **Geoinformatické přístupy při výzkumu diverzity planých předchůdců kulturních bobovitých rostlin**  
Zadávající katedra: **Katedra geoinformatiky**

### Z á s a d y p r o v y p r a c o v á n í :

Cílem diplomové práce je aplikovat geoinformatické přístupy při výzkumu diverzity planých předchůdců bobovitých rostlin. Hlavním cílem práce bude analyzovat dostupná environmentální prostorová data a provést syntézu s výsledky genotypování. Konkrétně se bude jednat o hrách, čočku a cizrnu s lokalitami výskytu východního Turecka v oblasti "úrodného půlměsíce". Dílčím cílem práce bude zjištění vlivu kvality historických a vstupních dat na výsledné prostorové analýzy a predikci předpokládané tolerance. Student se zaměří na možnosti využití lokálních globálních datasetů a získání dat pomocí služby Google EarthEngine. Výstupem práce budou vybrané prostorové analýzy založené na integraci získaných dat, které budou mít za cíl potvrdit prostorovou a genotypovou diverzifikaci. Součástí práce budou také grafy a mapové výstupy.

Student vyplní údaje o všech datových sadách, které vytvořil nebo získal v rámci práce, do Metainformačního systému katedry geoinformatiky a současně vytvoří zálohu údajů ve formě validovaného XML souboru. Celá práce (text, přílohy, výstupy, zdrojová a vytvořená data, XML soubor) se odevzdá v digitální podobě na CD (DVD) a text práce s vybranými přílohami bude odevzdán ve dvou svázaných výtiscích na sekretariát katedry. O diplomové práci student vytvoří webovou stránku v souladu s pravidly dostupnými na stránkách katedry. Práce bude zpracována podle zásad dle Voženílek (2002) a závazné šablony pro diplomové práce na KGI. Povinnou přílohou práce bude poster formátu A2.

Rozsah grafických prací: **dle potřeby**  
Rozsah pracovní zprávy: **max. 50 stran**  
Forma zpracování diplomové práce: **tištěná**  
Seznam odborné literatury: **viz příloha**

Vedoucí diplomové práce: **RNDr. Jan Brus, Ph.D.**  
Katedra geoinformatiky

Datum zadání diplomové práce: **1. listopadu 2018**  
Termín odevzdání diplomové práce: **5. května 2020**

doc. RNDr. Martin Kubala, Ph.D.  
děkan

L.S.

UNIVERZITA PALACKÉHO V OLOMOUCI  
PŘÍRODOVĚDECKÁ FAKULTA  
KATEDRA GEONFORMATIKY  
17. listopadu 50, 771 46 Olomouc

prof. RNDr. Vít Voženílek, CSc.  
vedoucí katedry

V Olomouci dne 10. prosince 2018

## Příloha zadání diplomové práce

Seznam odborné literatury:

- AWCHI, T. A. AND M. M. KALYANA Meteorological drought analysis in northern Iraq using SPI and GIS. *Sustainable Water Resources Management*, 2017, 3(4), 451-463.
- BARI, A., H. KHAZAEI, F. L. STODDARD, K. STREET, et al. Insilico evaluation of plant genetic resources to search for traits for adaptation to climate change. *Climatic change*, 2016, 134(4), 667-680.
- BROWN, T. A., M. K. JONES, W. POWELL AND R. G. ALLABY The complex origins of domesticated crops in the Fertile Crescent. *Trends in Ecology & Evolution*, 2009, 24(2), 103-109.
- BRUNAZZI, A., D. SCAGLIONE, R. F. TALINI, M. MICULAN, et al. Molecular diversity and landscape genomics of the crop wild relative *Triticum urartu* across the Fertile Crescent. *The Plant Journal*, 2018, 94(4), 670-684.
- KATINAS, L. AND J. V. CRISCI Agriculture Biogeography: An emerging discipline in search of a conceptual framework. *Progress in Physical Geography: Earth and Environment*, 2018, 0309133318776493.
- PLEKHANOVA, E., M. A. VISHNYAKOVA, S. BULYNTSEV, P. L. CHANG, et al. Genomic and phenotypic analysis of Vavilov's historic landraces reveals the impact of environment and genomic islands of agronomic traits. *Scientific Reports*, 2017, 7(1), 4816.
- SMYKAL, P., O. TRNENY, J. BRUS, P. HANACEK, et al. Genetic structure of wild pea (*Pisum sativum* subsp. *elatius*) populations in the northern part of the Fertile Crescent reflects moderate cross-pollination and strong effect of geographic but not environmental distance. *PloS one*, Mar 26 2018, 13(3).
- VON WETTBERG, E. J., P. L. CHANG, F. BAŞDEMİR, N. CARRASQUILA-GARCIA, et al. Ecology and genomics of an important crop wild relative as a prelude to agricultural innovation. *Nature communications*, 2018, 9(1), 649.
- ZAIR, W., N. MAXTED AND A. AMRI Setting conservation priorities for crop wild relatives in the Fertile Crescent. *Genetic Resources and Crop Evolution*, 2018, 65(3), 855-863.



# OBSAH

<b>SEZNAM POUŽITÝCH ZKRATEK .....</b>	<b>10</b>
<b>ÚVOD .....</b>	<b>11</b>
<b>1 CÍLE PRÁCE.....</b>	<b>12</b>
<b>2 METODY A POSTUPY ZPRACOVÁNÍ.....</b>	<b>13</b>
2.1 Použité metody .....	13
2.2 Použitá data .....	16
2.3 Použité programy .....	23
2.4 Postup zpracování.....	23
<b>3 STAV ŘEŠENÉ PROBLEMATIKY .....</b>	<b>25</b>
3.1 Oblast blízkého východu – Úrodný půlměsíc .....	25
3.2 Popis zkoumaných rostlin .....	27
3.3 Řešená problematika v odborných studiích.....	28
<b>4 GOOGLE EARTH ENGINE .....</b>	<b>32</b>
4.1 Práce v Google Earth Engine .....	33
<b>5 VLASTNÍ ŘEŠENÍ .....</b>	<b>36</b>
5.1 Postup zpracování bodových dat rostlin.....	36
5.2 Postup zpracování environmetálních dat .....	38
5.3 Postup statistických analýz.....	41
5.3.1 Zjištění důležitých environmentálních faktorů.....	41
5.3.2 Proces shlukování .....	45
5.3.3 Porovnání přesnosti .....	49
<b>6 INTERPRETACE STATISTICKÝCH ANALÝZ.....</b>	<b>50</b>
6.1 Výběr environmentálních faktorů .....	50
6.1.1 Cizrna ( <i>Cicer reticulatum</i> ).....	50
6.1.2 Cizrna ( <i>Cicer echinospermum</i> ) .....	52
6.1.3 Čočka ( <i>Lens orientalis</i> ).....	54
6.1.4 Hrách ( <i>Pisum elatius</i> ).....	57
6.2 Shlukování .....	59
6.2.1 Cizrna ( <i>Cicer reticulatum</i> ).....	59
6.2.2 Cizrna ( <i>Cicer echinospermum</i> ) .....	62
6.2.3 Čočka ( <i>Lens orientalis</i> ).....	63
6.2.4 Hrách ( <i>Pisum elatius</i> ).....	68
6.3 Polohová přesnost.....	70
<b>7 SHRUTÍ VÝSLEDKŮ .....</b>	<b>73</b>
<b>8 DISKUZE.....</b>	<b>75</b>
<b>9 ZÁVĚR .....</b>	<b>77</b>
<b>SEZNAM OBRÁZKŮ</b>	
<b>SEZNAM TABULEK</b>	
<b>POUŽITÁ LITERATURA A INFORMAČNÍ ZDROJE</b>	
<b>PŘÍLOHY</b>	

# SEZNAM POUŽITÝCH ZKRATEK

<b>Zkratka</b>	<b>Význam</b>
ASTER	Advanced Spaceborne Thermal Emission and Reflection Radiometer
AVHRR	Advanced Very-High-Resolution Radiometer
BIL	Band Interleaved by Line
CHILI	Continuous Heat-Insolation Load Index
CRU	Climatic Research Unit
CSV	Comma Separated Values
CTI	Compound Topographic Index
CWR	Crop Wild Relatives
DAPC	Discriminant analysis of principal components
DTED	Digital Terrain Elevation Data
EPSG	European Petroleum Survey Group
ESA	European Space Agency
EVI	Enhanced Vegetation Index
FEWS NET	Famine Early Warning Systems Network
FLDAS	Famine Land Data Assimilation System
GEE	Google Earth Engine
GeoJSON	Geo JavaScript Object Notation
GeoTIFF	Georeferenced Tagged Image File Format
HDF-EOS	Hierarchical Data Format - Earth Observing System
HLI	Heat Load Index
IDE	Integrated Development Environment
IMI	Integrated Moisture Index
JRA55	Japanese 55-year Reanalysis
KML	Keyhole Markup Language
MODIS	The Moderate Resolution Imaging Spectroradiometer
NASA	Národní úřad pro letectví a kosmonautiku
NAIP	The National Agriculture Imagery Program
NDVI	Normalized Difference Vegetation Index
NGA	National Geospatial Intelligence Agency
PCA	Analýza hlavních komponent
PDSI	Palmerův index intenzity sucha
SEI	Site Exposure Index
SHP	Esri Shapefile
SRTM30	Shuttle Radar Topography Mission
TIFF	Tagged Image File Format
TRASP	Topographic Radiation Aspect Index
TWI	Topographic Wetness Index
USGS	United States Geological Survey
VIR	The Vavilov Institute of Plant Genetic Resources
WGS84	World Geodetic System
X-SAR	X-Band Synthetic Aperture Radar

## ÚVOD

Kvůli omezování dostupných zdrojů orné půdy a vody vlivem rostoucí lidské populace a změnami klimatu rostou neustále požadavky na vylepšení zemědělství, které by bylo schopné produkce i během změn klimatu či odolávat tlaku chorob a škůdců (Plekhanova a kol., 2017). Podle predikcí bude do roku 2050 na světě 9 mld. lidí. Aby bylo možné je všechny nasytit, měla by se zemědělská produkce zvyšovat o 6 % ročně (Alexandratos a Bruinsma, 2012). Jedním z důsledků tzv. zelené revoluce, byla eliminace genetické rozmanitosti pěstovaných plodin. Přestože v procesu šlechtění dochází ke kombinaci vlastností, je výchozí genetická diverzita limitována diverzitou použitou při počátečních fázích domestikace. Následkem tohoto zúžení nejsou domestikované plodiny schopny se adaptovat na změny v zemědělství (Plekhanova a kol., 2017; Berger a kol., 2003). Existují však stále v přírodě se vyskytující plané formy, předchůdci dnešních kulturních plodin (Crop Wild Relatives, CWR). Ty je stále možné křížit s kulturními plodinami a využít tak jejich genetické diverzity.

Bohužel jsou však tyto rostliny ohroženy v jejich přirozeném prostředí, a to především vlivem urbanizace, intenzivního zemědělství, odlesňováním, znečištěním půdy a vody a také změnou klimatu (El-Beltagy, 2006). Plané formy plodin jsou geneticky blízké pěstovaným plodinám a obsahují genetickou diverzitu, kterou lze použít pro zlepšení plodin. Důvodem jsou geny, díky kterým se přizpůsobily nejrůznějším stresům v jejich přirozeném prostředí, jako je zasolení, sucho či extrémní teploty (Maxted a kol. 2006). Proto se v dnešní době řada studií zaměřuje na plané rostliny, které mohou přinést řešení pro zvyšování zemědělské produkce. Tyto studie ovšem nevyužívají většiny přístupných environmentálních datových sad, které by mohly ve výzkumu pomoci. Proto se tato práce zaměřuje na analýzu globálních environmentálních prostorových datových sad především z platformy Google Earth Engine a snaží se definovat prostorovou distribuci druhů napříč geografickým prostředím s využitím vícerozměrných analýz a globálních datových sad.

# 1 CÍLE PRÁCE

Hlavním cílem diplomové práce je aplikovat geoinformatické přístupy při výzkumu diverzity planých předchůdců bobovitých rostlin. Budou analyzována bodová data rostlin získaná z Katedry botaniky UP a světových genobank, konkrétně se bude jednat o data cizrny, čočky a hrachu s lokalitami výskytu jihovýchodního Turecka a v oblasti Úrodného půlměsíce. K dosažení hlavního cíle vede několik dílčích cílů. Prvním z nich je získání a analýza dostupných globálních environmentálních prostorových dat s důrazem na datové sady z platformy Google Earth Engine.

Dalším dílčím cílem je nalezení environmentálních faktorů, které nejvíce ovlivňují jednotlivé druhy zkoumaných rostlin. K tomu poslouží statistické metody, především analýza hlavních komponent. Následovat bude vytvoření shluků rostlin na základě vybraných environmentálních faktorů. Tyto shluky budou vizualizovány tak, aby byla možná interpretace jejich závislosti na zkoumaných environmentálních faktorech a budou promítnuty také do mapových výstupů.

Dále bude u dat, která obsahují informace o přiřazení ke genetické skupině, zjištěna závislost těchto skupin na environmentálních faktorech a dále budou genetické skupiny porovnány s výsledky shlukování.

Posledním dílčím cílem je zjištění vlivu přesnosti bodových dat na výsledky analýz z hlediska využití prostorových environmentálních datových sad různé kvality.

Výstupem práce budou grafy, tabulky a mapy zobrazující výsledky provedených analýz založených na datech rostlin a stažených environmentálních datech.

Výsledky a výstupy této práce budou dále sloužit výzkumníkům Katedry botaniky UP ve výzkumu zaměřeném na plané předchůdce kulturních bobovitých rostlin. Velkým přínosem práce je využití globálních environmentálních datových sad, které nebyly v podobných výzkumech doposud brány v potaz. Získávání dalších dat bude usnadněno pomocí vytvořených skriptů pro stahování dat z platformy Google Earth Engine.

## 2 METODY A POSTUPY ZPRACOVÁNÍ

Tato kapitola se zaměřuje na popis metod použitých v této práci, na popis dat o rostlinách a environmentálních datových sad, které vstupovaly do analýz, dále na softwary, které byly během řešení práce využity a postup práce, který vedl ke splnění cílů práce.

### 2.1 Použité metody

#### Zpracování dat

Před samotnými analýzami bylo nutné získaná data zpracovat. U nálezových dat rostlin se jednalo o vytvoření bodové vrstvy na základě obdržných souřadnic, dále mazání bodů s nízkou přesností, duplicitních bodů a bodů s nevhodným umístěním. Při výpočtu indexů z výškových dat SRTM30 musela být vrstva rozdělena do UTM pásem, aby nedocházelo ke zkreslení výpočtů vlivem nevhodně zvoleného souřadnicového systému. U datových sad stažených z Google Earth Engine muselo dojít ke zprůměrování některých datových sad za celé období, aby korespondovaly s ostatními environmentálními datovými sadami. Hodnoty ze všech environmentálních datových sad byly extrahovány každému bodu výskytu rostlin, aby byly umožněny statistické analýzy v softwaru RStudio. Navíc musely být smazány všechny záznamy, které obsahovaly NULL hodnoty, jinak nebylo možné vytvořit korelační matici, která vyžaduje kompletní sadu dat.

#### Indexy odvozené ze SRTM30

##### Compound Topographic Index (CTI) také označovaný jako Topographic Wetness Index (TWI)

Složený topografický index (CTI) je index vlhkosti v ustáleném stavu, do jehož výpočtu vstupují akumulace průtoku a sklon svahu. Počítá se pomocí vzorce:

$$cti = \ln\left(\frac{\alpha}{\tan(\theta)}\right) \quad [1]$$

$\alpha$  = oblast zachycení [(akumulace průtoku + 1) × (plocha pixelů v m<sup>2</sup>)]

$\theta$  = úhel sklonu v radiánech (Evans a kol., 2014)

Výsledkem je hodnota indikátoru vlhkosti, která se obvykle pohybuje v rozmezí -3 až 30. Nižší hodnoty indikují oblasti s nejstrmějším sklonem svahu, tedy s nejmenším potenciálem akumulace vody. Naopak vyšší hodnoty jsou identifikovány mírným sklonem svahu se zvýšenou akumulací odtoku (Ballerine, 2017).

#### Heat Load Index (HLI)

Heat Load Index neboli Index tepelné zátěže ve výpočtu počítá se zeměpisnou šířkou, sklonem a orientací svahu. HLI se počítá pomocí vzorce:

$$hli = 0.039 + [0.808 * \cos(l) * \cos(\theta)] - [0.196 * \sin(\theta)] - [0.482 * \cos(f(\alpha)) * \sin(f(\alpha))] \quad [2]$$

$$f(\alpha) = \left| \pi - \left| \alpha - \frac{5\pi}{4} \right| \right| \quad [3]$$

$$\begin{aligned}
 l &= \text{zeměpisná šířka} \\
 \theta &= \text{sklon svahu} \\
 f(\alpha) &= \text{složená orientace svahu} \\
 \alpha &= \text{sklon svahu}
 \end{aligned}$$

Výsledné hodnoty se pohybují od 0 pro nejchladnější oblasti do 1 pro nejteplejší oblasti (Evans a kol., 2014). Vyšších hodnot dosahují svahy orientované na jihozápad, naopak nejnižší hodnoty mají svahy orientované na severovýchod (McCune a Keon 2002).

### **Integrated Moisture Index (IMI)**

Integrated Moisture Index odhaduje vlhkost půdy v topograficky heterogenní krajině, v závislosti na intenzitě stínu, zakřivení svahu a akumulace odtoku (Evans a kol., 2014). Počítá se pomocí vzorce:

$$imi = [\text{hillshade} * 0.5] + [\text{curvature}(\theta) * 0.15] + [\text{Flow Accumulation} * 0.35] \quad [4]$$

Hodnoty IMI se mohou pohybovat od 0 do 100, kdy s narůstající hodnotou narůstá hodnota vlhkosti (Iverson a kol., 1997).

### **Site Exposure Index (SEI)**

Site Exposure Index rozděluje svah podle orientace na dvě části, a to na severní a jižní. Zároveň zohledňuje strmost svahu. Počítá se pomocí vzorce:

$$sei = \theta \cos\left(\pi \frac{\alpha - 180}{180}\right) \quad [5]$$

Výsledky mohou dosahovat hodnot od -100 pro nejchladnější až po 100 pro nejteplejší oblast (Evans a kol., 2014).

### **Topographic Radiation Aspect Index (TRASP)**

Topographic Radiation Aspect Index počítá s kruhovou orientací svahu tak, že chladnějším a mokřejším svahům orientovaným na severovýchod přiřazuje hodnotu 0 a teplejším a sušším svahům jihozápadním přiřazuje hodnotu 1. Výsledné hodnoty se pohybují v rozmezí 0 až 1 (Roberts a Cooper, 1989; Evans a kol., 2014). Počítá se jako:

$$trasp = \frac{1 - \cos\left(\frac{\pi}{180}(\alpha - 30)\right)}{2} \quad [6]$$

$\alpha$  = orientace svahu ve stupních

### **Analýza hlavních komponent (PCA)**

Analýza hlavních komponent je statistická metoda pro zjednodušení popisu vzájemně korelovaných znaků. Cílem PCA je přeměna původních korelovaných znaků na nové nezávislé proměnné, tzv. komponenty. Komponenty jsou lineární kombinací původních proměnných a měly by co nejlépe vysvětlovat variabilitu těchto proměnných (Tonhauserová, 2013). Hlavní komponenty jsou seřazeny podle množství rozptylu

obsažených dat, přičemž první komponenta vždy obsahuje největší množství informace o variabilitě původních dat, druhá hlavní komponenta obsahuje největší část variability původních dat, která není obsažena v první komponentě atd. (Meloun, 2014).

Pro objektivní posouzení, kolik hlavních komponent má do následných analýz vstupovat, se nejčastěji používá scree plot. Jedná se o sloupcový diagram, který zobrazuje relativní velikost vlastních čísel. Vyberou se komponenty, které jsou od zbytku odděleny jasným zlomovým místem. Pokud zlom v diagramu není zřetelný, lze použít také Keiserovo kritérium, kdy se vyberou pouze ty komponenty, které mají vlastní číslo větší než 1 (Meloun, 2014).

Výsledky PCA lze kromě scree plotu zobrazit v několika dalších diagramech. Těmi jsou graf komponentních vah, scatterplot a biplot. Graf komponentních vah (graf komponentních zátěží) zobrazuje komponentní váhy pro první dvě hlavní komponenty a porovnává, jak spolu souvisí jednotlivé znaky (v této práci environmentální faktory). Čím blíže sobě se jednotlivé faktory nacházejí, tím více spolu korelují. Opačný směr faktorů v grafu značí negativní korelaci a nulovou korelaci vykazují faktory v grafu na sebe kolmé. Čím dále se faktory nachází od počátku grafu, tím větší mají na komponenty vliv.

Scatterplot neboli rozptylový diagram komponentního skóre zobrazuje u prvních dvou komponent hodnoty všech prvků. Prvky nacházející se blíže sobě jsou si podobnější než prvky vzdálené. Prvky, které jsou umístěny blíže středu jsou nejtypičtější, v opačném případě se jedná o extrémny. Biplot je kombinací grafu komponentních vah a scatterplotu (Meloun, 2014).

## **Metoda korelace**

Korelace vyjadřuje míru závislosti mezi několika proměnnými. Může dosahovat hodnot od -1 do 1, přičemž hodnoty blízké 1 vykazují vysokou pozitivní korelaci, hodnoty blízké -1 vykazují negativní korelaci a hodnoty okolo 0 značí, že spolu proměnné nekorelují. Pro posouzení míry korelace mezi větším množstvím proměnných lze využít korelační matice, která zobrazuje korelaci mezi všemi vstupními proměnnými navzájem. Hodnoty korelace lze rozdělit do několika kategorií. Například de Vaus (2002) uvádí následující rozdělení:

korelace (0; 0,10) → žádná závislost

korelace (0,10; 0,30) → nízká až střední závislost

korelace (0,30; 0,50) → střední závislost

korelace (0,50; 0,70) → podstatná závislost

korelace (0,70; 0,90) → velmi silná závislost

korelace (0,90; 1) → skoro perfektní závislost

## **Shluková analýza k-means**

Shlukové analýzy slouží k vymezení skupin, ve kterých jsou si data nejvíce podobná a zároveň jsou tyto skupiny navzájem odlišné. Jednou z metod shlukových analýz je metoda k-means. Před začátkem shlukování je vhodné provést standardizaci vstupních dat, aby se zabránilo ovlivnění výsledků neporovnatelnými hodnotami atributů. Metoda k-means je založena na postupném přesouvání objektů mezi shluky. Na začátku algoritmu se vždy zvolí počet shluků a počáteční hodnoty jejich průměrů. Následuje klasifikace, při které se všechny objekty přiřadí k nejbližšímu průměru určenému

v předchozím kroku. Následně se v každém shluku vypočítá nový střed jako těžiště shluku a přiřadí se mu všechny objekty, které se nachází nejbližší. To se opakuje, dokud nedochází k výraznému zlepšení (Ivan, 2016; Dvořák, 2008). Výhodou této metody je její výpočetní nenáročnost, nevýhodou je naopak to, že musí uživatel dopředu určit počet shluků. Aby byl počet shluků zvolen správně, lze použít některou z metod, která dokáže optimální počet shluků určit. Nejčastěji používaná je metoda elbow, kdy je podobně jako u scree plotu pomocí grafu vizuálně posouzeno místo zlomu, které je bráno jako optimální pro výběr počtu shluků (Ivan, 2016).

### Graf paralelních souřadnic

Tato metoda se používá pro zobrazení multidimenzionálních dat, kdy pro každou dimenzi je využita jedna osa rovnoběžná s ostatními. Za dimenze jsou v případě této práce považovány jednotlivé environmentální faktory. Tyto osy jsou rozmístěny paralelně za sebou. Každý bod je zobrazen jako lomená čára, jejíž vrcholy se nachází na jednotlivých osách a jsou dány souřadnicí. Tato souřadnice je úměrná hodnotě, kterou je zastoupena v každé dimenzi (Mrkvička, 2017).

## 2.2 Použitá data

V práci byly použity dva typy dat. Těmi jsou data o výskytu planých předchůdců cizrny, čočky a hrachu v oblasti Úrodného půlměsíce a Turecka a environmentální globální datasety, které byly využity k charakteristice výše zmíněných rostlin ve zkoumané lokalitě. Všechna použitá data jsou popsána níže.

### Data o rostlinách

Veškerá data o rostlinách použitá v této práci byla dodána doc. Smýkalem z Katedry botaniky UP. Poskytnutá data byla dodána v souborech .xls, jejichž obsahem bylo identifikační číslo rostliny, název rostliny, stát výskytu, datový zdroj a zeměpisné souřadnice. Pouze 133 záznamů čočky (*Lens orientalis*) obsahovalo navíc údaj o genetické informaci rostliny. Celkový počet jednotlivých druhů rostlin je znázorněn v tabulce 1.

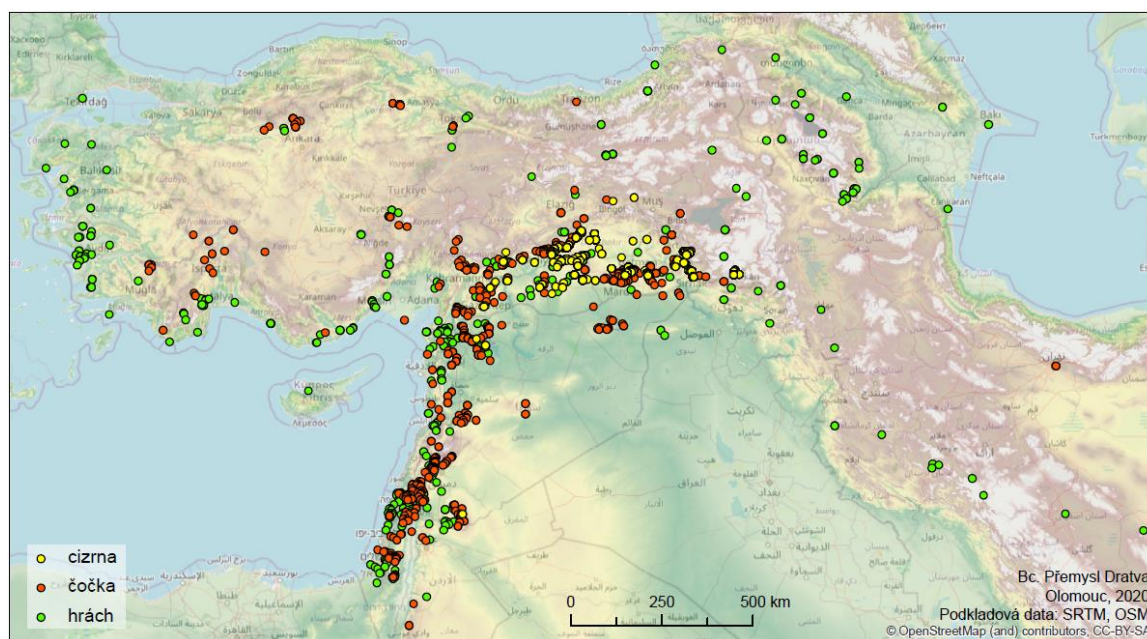
Tabulka 1 Počet rostlin z původních dat

Rostlina	Latinský název	Počet záznamů	Celkem
Cizrna	<i>Cicer bijugum</i>	203	3226
	<i>Cicer echinospermum</i>	750	
	<i>Cicer pinnatifidum</i>	586	
	<i>Cicer reticulatum</i>	1676	
	neoznačená cizrna	11	
Čočka	<i>Lens orientalis</i>	863	863
Hrách	<i>Pisum elatius</i>	907	907

Vzorky byly získány z několika zdrojů. Většina dat pocházela z terénního sběru Jense Bergera v jihovýchodní části Turecka, který probíhal v letech 2013–2015. Zbytek dat byl z několika světových genobank a dalších institucí. Data prošla v rámci zpracování selekcí, během které byla část odstraněna, přičemž z dat cizrny byly ponechány pouze dva druhy – *Cicer echinospermum* a *Cicer reticulatum*, jelikož se jedná



o sesterské druhy, které jsou nejbližší přímým předchůdcům kulturní cizrny. Lokalizaci dat lze vidět na obr. 1. Původ dat je znázorněn v tabulce 2, která obsahuje údaje z již vyřizovaných dat (proces třídění je popsán v kapitole 5).



Obr. 1 Lokalizace výskytu rostlin.

Tabulka 2 Zdroje dat rostlin, která prošla selekcí

Rostlina	Latinský název	Zdroje dat	Počet	Součet	Celkem
Cizrna	<i>Cicer echinospermum</i>	GBIF – Global Biodiversity Information Facility	59	730	2397
		Jens – Jens Berger, CSIRO, Australia	671		
	<i>Cicer reticulatum</i>	GBIF – Global Biodiversity Information Facility	64	1667	
		Jens – Jens Berger, CSIRO, Australia	1603		
Čočka	<i>Lens orientalis</i>	Australian genebank	10	710	710
		GBIF – Global Biodiversity Information Facility	234		
		GBR004 – Millennium Seed Bank Kew	25		
		Israel database – Israel Gene Bank	58		
		Jens – Jens Berger, CSIRO, Australia	192		
		SYR002 – ICARDA (International Center for Agriculture Research in Dry Areas)	114		
		PaulGepts-Alo2011data – Data z publikace autora (Alo a kol., 2011) využívající ICARDA	55		
USDA – Americké ministerstvo zemědělství	22				
Hrách	<i>Pisum elatius</i>	GBIS – Genebank Information System	273	716	716
		Jens Berger, CSIRO, Australia	119		
		NewOne – O. Kosterin, Institute of Cytogenetics	1		
		PeerJ – Data z publikace autora (Hradilová a kol., 2019) využívající USDA, John Innes Center, ICARDA a další	45		
		PLOS One – Data z publikace autora (Smýkal a kol., 2018)	14		
		SciRep – Data z publikace autora (Smýkal a kol., 2017) využívající USDA, John Innes Center, ICARDA a další	232		
		Table_list – USDA, John Innes Center, ICARDA a další	32		

## WorldClim

WorldClim je globální dataset rastrových klimatických vrstev, které byly vytvořeny Robertem Hijmansem a jeho kolegy. Data WorldClim jsou volně dostupná ke stažení pod licencí Creative Commons Attribution-ShareAlike 4.0 International License. Nejnovější verze k dispozici je Version2, která byla použita v této práci. Tato data jsou průměrem pro každý měsíc z let 1970 až 2000. Datové vrstvy lze stahovat ve čtyřech rozlišeních. Nejpodrobnější je 30 úhlových vteřin, dále dva a půl, 5 a 10 úhlových minut, což je velikost jednoho pixlu přibližně 1 km až 18,5 km.

Datasets lze stáhnout ve formátu .zip, který obsahuje textový soubor readme a 12 vrstev ve formátu GeoTIFF (Georeferenced Tagged Image File Format), kdy každá vrstva obsahuje údaje za jednotlivý měsíc. Ke stažení jsou tyto základní rastrové vrstvy: minimální, průměrná a maximální teplota (vše v °C), úhrn srážek (mm), sluneční radiace ( $\text{kJ m}^{-2} \text{den}^{-1}$ ), rychlost větru ( $\text{m s}^{-1}$ ) a tlak vodní páry (kPa). Dále je možné bezplatně získat bioklimatické proměnné, kterých je 19. Bioklimatické proměnné jsou rastrové vrstvy odvozené z měsíčních hodnot srážek a teplot. Označují se jako BIO1 až BIO19 (viz tabulka 3) (WorldClim).

Tabulka 3 Bioklimatické proměnné (zdroj: WorldClim)

BIO1	Průměrná roční teplota
BIO2	Průměrný denní rozsah teplot
BIO3	Isotermalita (BIO2-BIO7) $\times$ 100
BIO4	Sezónnost teploty (směrodatná odchylka $\times$ 100)
BIO5	Maximální teplota nejteplejšího měsíce
BIO6	Minimální teplota nejchladnějšího měsíce
BIO7	Roční rozsah teplot (BIO5-BIO6)
BIO8	Průměrná teplota nejvlhčího čtvrtletí
BIO9	Průměrná teplota nejsuššího čtvrtletí
BIO10	Průměrná teplota nejteplejšího čtvrtletí
BIO11	Průměrná teplota nejchladnějšího čtvrtletí
BIO12	Roční úhrn srážek
BIO13	Srážkový úhrn v nejvlhčím měsíci
BIO14	Srážkový úhrn v nejsušším měsíci
BIO15	Sezónnost srážek (variační koeficient)
BIO16	Srážkový úhrn v nejvlhčím čtvrtletí
BIO17	Srážkový úhrn v nejsušším čtvrtletí
BIO18	Srážkový úhrn v nejteplejším čtvrtletí
BIO19	Srážkový úhrn v nejchladnějším čtvrtletí

## **SRTM**

SRTM (Shuttle Radar Topography Mission) je globální datová sada obsahující informace o nadmořské výšce v rozlišení jedné úhlové vteřiny (přibližně 30 m). Data byla získána z radarového snímání, které probíhalo 11. až 22. února 2000 z raketoplánu Endeavour. Mise byla pod záštitou Národního úřadu pro letectví a kosmonautiku (NASA) a National Geospatial Intelligence Agency (NGA). Data SRTM, která mají geografickou projekci WGS84 (World Geodetic System), pokrývají 80 % zemského povrchu mezi 60° severní a 56° jižní šířky.

Snímkování probíhalo s využitím technologie interferometrického radaru X-Band Synthetic Aperture Radar (X-SAR). Princip spočívá v porovnání dvou radarových snímků pořízených ze dvou úhlů. Toho bylo docíleno pomocí 60 metrů dlouhého stožáru s anténou upevněného na raketoplánu a druhé antény umístěné v raketoplánu. Rozdíl mezi signály z těchto dvou antén posloužil k výpočtu nadmořské výšky.

Existuje několik výškových dat SRTM, například:

- **SRTM Non-Void Filled** jsou výšková data zpracovaná z C-band radarových signálů o intervalech 1 úhlová vteřina (30 m) pro USA. Tato verze obsahuje úpravy pobřežních linií a dalších nedostatků, jako je odstranění vertikálních chyb a doplnění chybějících dat v malých oblastech. Data mimo USA mají rozlišení 3 úhlové vteřiny (90 m). Jako metoda převzorkování byla použita kubická konvoluce.
- **SRTM Void Filled** je datová sada, která je výsledkem dodatečného upravování oblastí s chybějícími daty v SRTM Non-Void Filled. Data jsou doplněna pomocí interpolace nebo pomocí výškových dat z jiných zdrojů. Prostorové rozlišení je stejné jak u Non-Void Filled, tedy 30 m pro USA a 90 m pro zbytek světa.
- **SRTM 1 Arc-Second Global** jsou výšková data dostupná v rozlišení 1 úhlová vteřina pro celý svět. Mohou se zde vyskytovat malé oblasti bez dat (USGS b).

Data jsou dostupná ke stažení po registraci na internetové stránce <https://earthexplorer.usgs.gov/>. Zde je možné stáhnout také datové vrstvy SRTM Water Body Dataset, což je vektorová vrstva (Esri Shapefile) pobřežních linií, jezer, řek, která vznikla jako vedlejší produkt při tvorbě výškových dat SRTM. Vrstva je aktuální k datu, kdy probíhalo snímání SRTM (USGS a).

Výšková data SRTM lze stáhnout ve třech formátech BIL, DTED a GeoTIFF. Formát BIL (Band Interleaved by Line) je binární rastrový formát, který navíc obsahuje záhlaví (samostatný soubor), ve kterém je uloženo rozdělení a formátování rastru. Formát DTED (Digital Terrain Elevation Data) je vytvořený agenturou NGA. Obsahuje matici vertikálních výšek, které jsou zjišťovány pomocí zeměpisné šířky a délky a jsou rozděleny do pravidelných horizontálních intervalů. GeoTIFF je postaven na formátu TIFF, oproti kterému navíc obsahuje geografické informace. Tento standardizovaný výměnný formát je hojně využíván pro letecké či satelitní snímky nebo také pro výšková data (EARTHDATA, 2019; USGS b).

## **Datové sady z Google Earth Engine**

Z platformy Google Earth Engine bylo použito pět datových sad, které byly vybrány na základě konzultace s odborníky na danou problematiku. Soupis všech použitých vrstev spolu s jejich charakteristikou se nachází v tabulce 5.

### **1) MOD13Q1.006 Terra Vegetation Indices 16-Day Global 250m**

*MOD13Q1.006 Terra Vegetation Indices 16-Day Global 250m* je datová sada od Národního úřadu pro letectví a kosmonautiku (NASA) LP DAAC v USGS EROS Center (Earth Engine Data Catalog d). Terra Moderate Resolution Imaging Spectroradiometer (MODIS) produkuje data vegetačních indexů pod označením MOD13Q1 každých 16 dnů, která jsou k dispozici v časovém rozmezí od 2. února 2000 do současnosti. Datová sada má globální pokrytí s rozlišením 250 metrů. Data jsou publikována ve formátu HDF-EOS (Hierarchical Data Format - Earth Observing System). Datová sada obsahuje 12 vrstev, ze kterých si uživatel může vybírat, jako je například odrazivost v různých vlnových délkách, úhly zenitu, kvalita vegetačních indexů a další (USGS c).

Hlavními vrstvami jsou ovšem dva vegetační indexy. Prvním je NDVI (Normalized Difference Vegetation Index), který se označuje jako index kontinuity. Druhým indexem je EVI (Enhanced Vegetation Index), který má zlepšenou citlivost pro oblasti s vysokou biomasou a také snižuje vlivy atmosféry (Didan a kol. 2015). V této práci byl použit

NDVI index, což je poměrový vegetační index užívající blízké infračervené a viditelné červené spektrum. Je počítán pomocí rovnice:

$$NDVI = \frac{NIR - RED}{NIR + RED} \quad [7]$$

Výsledkem rovnice jsou hodnoty pohybující se v intervalu od -1 do 1 v závislosti na obsahu zelené hmoty. Kladných hodnot vyšších než 0,5 nabývá hustá zdravá vegetace. Hodnoty blízké nule dosahuje holá a kamennitá půda a záporné hodnoty má voda, sníh či led (Dobrovolný, 1998). Hodnoty NDVI se v této datové sadě MOD13Q1.006 pohybují v rozmezí -2 000 až 10 000 a proto je potřeba je vynásobit koeficientem 0,0001 (USGS c).

## 2) TerraClimate: Monthly Climate and Climatic Water Balance for Global Terrestrial Surfaces, University of Idaho

University of Idaho poskytuje datovou sadu *TerraClimate: Monthly Climate and Climatic Water Balance for Global Terrestrial Surfaces*. Ta obsahuje měsíční data od 1. 1. 1958 do konce roku 2015, přičemž dochází k postupným aktualizacím, které zpřístupňují předběžná data až do roku 2018. Prostorové rozlišení datové sady je 2,5 úhlových minut, což je v přepočtu přibližně 4,5 km. Díky časovému i prostorovému rozlišení jsou tato data vhodná pro ekologické a hydrologické studie v globálním měřítku.

Data z TerraClimate vychází z dat WorldClim s vysokým prostorovým rozlišením, které kombinuje s časově proměnlivými daty z CRU Ts4.0 (Climatic Research Unit) a JRA55 (Japanese 55-year Reanalysis). TerraClimate produkuje primární klimatické faktory, ze kterých odvozuje další faktory. Primárními faktory jsou maximální a minimální teplota, tlak páry, srážkový úhrn, krátkovlnné záření a rychlost větru. Odvozenými faktory jsou například aktuální evapotranspirace, vlhkost půdy, Palmerův index intenzity sucha a další (Climatology Lab).

V této práci jsou použity faktory aktuální evapotranspirace a půdní vlhkost, které jsou odvozeny pomocí jednorozměrného modelu rovnováhy půdní vody, dále úhrn srážek a Palmerův index intenzity sucha (PDSI). PDSI je index vyvinutý v 60. letech 20. století jako jeden z prvních indexů, které do výpočtů nezahrnovaly pouze data o srážkách. Do výpočtu PDSI vstupují měsíční data o teplotě a srážkách a informace o schopnosti půdy zadržovat vodu (Alley, 1984). Výstupní hodnoty indexu PDSI jsou znázorněny v tabulce 4.

Tabulka 4 Hodnoty PDSI (zdroj: Rožnovský, 2012)

Hodnota PDSI	Charakteristika oblasti
≤ -4,00	Extrémní sucho
-3,99 až -3,00	Drsné sucho
-2,99 až -2,00	Mírné sucho
-1,99 až -1,00	Slabé sucho
-0,99 až -0,50	Období začínajícího sucha
-0,49 až 0,49	Blízká normálu
0,50 až 0,99	Období začínajícího vlhka
1,00 až 1,99	Slabě vlhká
2,00 až 2,99	Mírně vlhká
3,00 až 3,99	Velmi vlhká
≥ 4,00	Extrémně vlhká

Hodnoty PDSI v použitém datasetu se pohybují od -4 317 do 3 418, pro správné hodnoty je třeba je vynásobit koeficientem 0,01, hodnoty dalších dvou použitých vrstev (aktuální evapotranspirace a půdní vlhkost) je potřeba vynásobit koeficientem 0,1 (Earth Engine Data Catalog e).

### **3) FLDAS: Famine Early Warning Systems Network (FEWS NET) Land Data Assimilation System**

*FLDAS: Famine Early Warning Systems Network (FEWS NET) Land Data Assimilation System* je jednou z datových sad od NASA. Datasety slouží pro monitorování a předpovědi potravinové bezpečnosti ve světě. Cílem FLDAS je efektivní využití dostupných hydroklimatických dat, která slouží pro podporu rozhodování FEWS NET. Jedná se o globální dataset s prostorovým rozlišením 0,1 úhlového stupně (cca 11 km). Dataset poskytuje data v měsíčním intervalu od 1. 1. 1982 do současnosti. Obsahuje 28 vrstev, například evapotranspirace, povrchový tlak, měrná vlhkost, celková míra srážek, teplota půdy, vlhkost půdy a další. V této práci jsou použity vrstvy evapotranspirace, celková míra srážek a vlhkost půdy v hloubce 10–40 cm (NASA, 2020).

### **4) OpenLandMap**

*OpenLandMap* je soubor patnácti datových sad od poskytovatele EnvirometriX. V této práci byly použity datové sady:

- *OpenLandMap Soil bulk density* – hustota půdy,
- *OpenLandMap Soil pH in H<sub>2</sub>O* – pH vody v půdě,
- *OpenLandMap Soil texture class (USDA system)* – typ textury půdy (USDA systém),
- *OpenLandMap Soil organic carbon content* – obsah organického uhlíku v půdě,
- *OpenLandMap Soil water content at 33kPa (field capacity)* – obsah půdní vody v 33 kPa.

Použité datasety byly vytvořeny na základě dat z několika zdrojů (například USDA National Cooperative Soil Characterization Database, WHRC global mangrove soil carbon dataset nebo LUCAS Soil database (GitHub 2019)). Datasety jsou tvořeny průměrnými hodnotami z časového období 1. 1. 1950 až 1. 1. 2018. Tyto datasety mají prostorové rozlišení 250 metrů a všechny obsahují šest vrstev s hloubkami půdy od 0 po 200 cm. Data *Soil bulk density* a *Soil pH in H<sub>2</sub>O* je pro získání správných hodnot nutné vynásobit koeficientem 10 a data *Soil organic carbon content* koeficientem 5 (Earth Engine Data Catalog f).

### **5) Global SRTM CHILI (Continuous Heat-Insolation Load Index)**

Poslední datovou sadou z Google Earth Engine použitou v této práci je *Global SRTM CHILI (Continuous Heat-Insolation Load Index)*. Datová sada obsahuje index CHILI odvozený ze SRTM v rozlišení 30 metrů, přičemž rozlišení datové sady je 90 metrů. Výpočet indexu je založený na slunečním záření brzy odpoledne a výška Slunce je brána v době rovnodennosti. Výsledkem je hodnota účinku slunečního záření a topografického zastínění na evapotranspiraci v rozsahu od 0 (velmi chladný) do 255 (velmi teplý). Poskytovatelem této datové sady jsou Conservation Science Partners (Earth Engine Data Catalog g).

Tabulka 5 Popis vrstev z Google Earth Engine

Název datové sady	Rozišení	Časová dostupnost	Granularita	Vrstva	Název v analýzách	Jednotka	Min	Max	Popis vrstvy
MOD13Q1.006 Terra Vegetation Indices 16-Day Global 250m	250 m	18 .2 .2000 – nyní	1 × za 16 dnů	NDVI	NDVI		-2000	10000	Normalized Difference Vegetation Index
TerraClimate: Monthly Climate and Climatic Water Balance for Global Terrestrial Surfaces, University of Idaho	2,5 úhlových min, 4,625 km	1. 1. 1958 – 31. 12. 2015	1 × za měsíc	aet	aet	mm	0	3140	Aktuální evapotranspirace, odvozená pomocí jednorozměrného modelu rovnováhy půdní vody
				pdsi	pdsi		-4317	3418	Palmerův index intenzity sucha
				pr	pr	mm	0	7245	Úhrn srážek
				soil	soil	mm	0	8882	Půdní vlhkost odvozená pomocí jednorozměrného modelu rovnováhy půdní vody
FLDAS: Famine Early Warning Systems Network (FEWS NET) Land Data Assimilation System	0,1 úhlových stupňů, 11,1 km	1. 1. 1982 – nyní	1 × za měsíc	Evap_tavg	Evap	kg m <sup>-2</sup> s <sup>-1</sup>			Evapotranspirace
				Rainf_f_tavg	Rainf	kg m <sup>-2</sup> s <sup>-1</sup>			Celkový úhrn srážek
				SoilMoi10_40cm_tavg	SoilMoi	m <sup>3</sup> m <sup>-3</sup>			Vlhkost půdy v hloubce 10–40 cm
OpenLandMap Soil bulk density	250 m	1. 1. 1950 – 1. 1. 2018	1 vrstva	b0		kg / m <sup>3</sup>	5	185	Hustota půdy v hloubce 0 cm
				b10	SBulkb10	kg / m <sup>3</sup>	5	185	Hustota půdy v hloubce 10 cm
				b30		kg / m <sup>3</sup>	5	185	Hustota půdy v hloubce 30 cm
				b60		kg / m <sup>3</sup>	5	185	Hustota půdy v hloubce 60 cm
OpenLandMap Soil pH in H2O	250 m	1. 1. 1950 – 1. 1. 2018	1 vrstva	b0			42	110	pH H <sub>2</sub> O v půdě v hloubce 0 cm
				b10	SpHH2Ob10		42	110	pH H <sub>2</sub> O v půdě v hloubce 10 cm
				b30			42	110	pH H <sub>2</sub> O v půdě v hloubce 30 cm
				b60			42	110	pH H <sub>2</sub> O v půdě v hloubce 60 cm
OpenLandMap Soil texture class (USDA system)	250 m	1. 1. 1950 – 1. 1. 2018	1 vrstva	b0					Typ textury půdy (USDA) v hloubce 0 cm
				b10	STexturb10				Typ textury půdy (USDA) v hloubce 10 cm
				b30					Typ textury půdy (USDA) v hloubce 30 cm
				b60					Typ textury půdy (USDA) v hloubce 60 cm
OpenLandMap Soil organic carbon content	250 m	1. 1. 1950 – 1. 1. 2018	1 vrstva	b0		g / kg	0	120	Obsah organického uhlíku v půdě v hloubce 0 cm
				b10	SCarbonb10	g / kg	0	120	Obsah organického uhlíku v půdě v hloubce 10 cm
				b30		g / kg	0	120	Obsah organického uhlíku v půdě v hloubce 30 cm
				b60		g / kg	0	120	Obsah organického uhlíku v půdě v hloubce 60 cm
OpenLandMap Soil water content at 33kPa (field capacity)	250 m	1. 1. 1950 – 1. 1. 2018	1 vrstva	b0		%	0	52,974	Obsah půdní vody v 33 kPa v hloubce 0 cm
				b10	SWatConb10	%	0	52,974	Obsah půdní vody v 33 kPa v hloubce 10 cm
				b30		%	0	52,974	Obsah půdní vody v 33 kPa v hloubce 30 cm
				b60		%	0	52,974	Obsah půdní vody v 33 kPa v hloubce 60 cm
Global SRTM CHILI (Continuous Heat- Insolation Load Index)	90 m	24. 1. 2006 – 13. 5. 2011	1 vrstva	constant	CHILI		0	255	Index CHILI odvozený od SRTM v rozsahu od 0 (velmi chladný) do 255 (velmi teplý)

## 2.3 Použité programy

Pro splnění dílčích cílů práce bylo využito několika softwarů. Pro výběr a zpracování bodových dat rostlin byl použit software ArcGIS Desktop 10.6 od společnosti Esri, ve kterém byla vytvářena také část mapových výstupů. V rámci analýz byl použit toolset Temperature and Moisture z toolboxu Geomorphometry & Gradient Metrics od Evanse a kol. (2014), který obsahuje šest nástrojů:

- 2nd Derivate Slope – počítá druhou derivaci sklonu svahu
- Compound Topographic Index
- Heat Load Index
- Integrated Moisture Index
- Site Exposure Index
- Slope/Aspect Transformation

Pro stažení dat SRTM30 z webu Earth Explorer od USGS bylo zapotřebí využití softwaru Bulk Download Application 1.4.1, který umožnil stažení všech vybraných snímků zároveň. Dále byl využit software Excel z kancelářského balíčku MS Office 2016 od společnosti Microsoft, ve kterém probíhalo třízení, úprava a zpracování dat a také tvorba tabulek. Veškeré statistické analýzy probíhaly v open source softwaru RStudio, který prostřednictvím programovacího kódu R a nainstalovaných balíčků umožňuje rozsáhlé statistické analýzy a vizualizace. Balíčky využitě v této práci jsou car, factextra, FactoMineR, corrplot, writexl, NbClust a GGally. Pro finalizaci grafických výstupů byl využit grafický editor Adobe Illustrator CC.

## 2.4 Postup zpracování

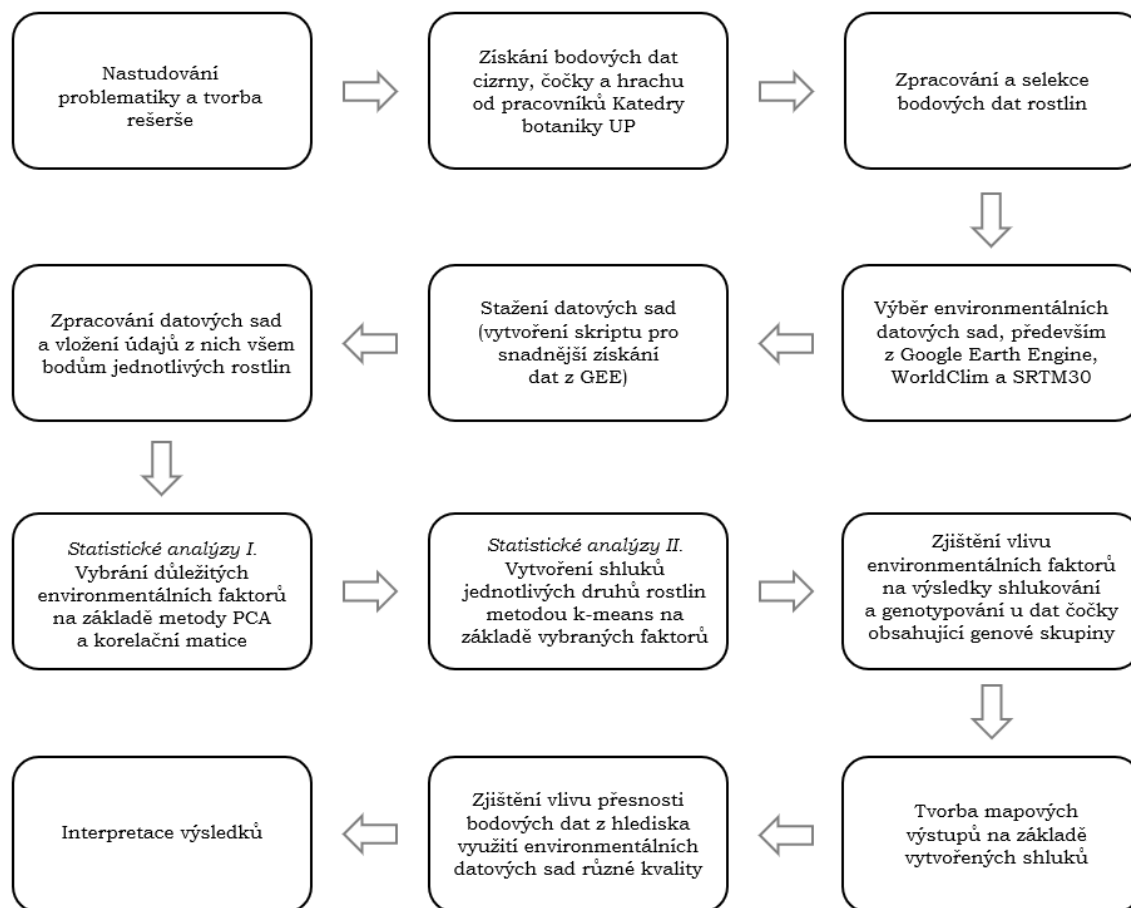
Postup práce koresponduje s vytyčenými cíli práce a je znázorněn na obr. 2. Nejprve byla sepsána rešerše, která je zaměřená na charakteristiku zkoumaných rostlin, popis klimatu lokality výskytu těchto rostlin a dále na současný stav řešené problematiky. Dalším krokem bylo získání bodových dat cizrný, čočky a hrachu od pracovníků Katedry botaniky UP, která byla následně zpracována a vytržena. Třízení probíhalo na základě přesnosti dat, duplicity či nevhodného umístění bodů. Tento proces je blíže popsán v kapitole 5.1.

Dále byly na základě konzultace s odborníky vybrány environmentální datové sady, které by mohly napomoci při charakteristice výskytu rostlin. Jedná se především o data z platformy Google Earth Engine, WorldClim a SRTM30. Vybrané datové sady byly následně staženy, u platformy Google Earth Engine musely být vytvořeny skripty, pomocí kterých lze data získat. Postup stahování dat je popsán v kapitole 5.2 a vytvořené skripty se nachází v přílohách 1 a 2. Následně bylo třeba environmentální datasety zpracovat a extrahovat jejich hodnoty všem bodům rostlin, aby bylo možné provádět další analýzy.

Jakmile byla data zpracována, bylo možné přistoupit ke statistickým analýzám, které byly prováděny v softwaru RStudio. První částí je výběr důležitých environmentálních faktorů, které mají nejzásadnější vliv na výskyt zkoumaných rostlin. To bylo provedeno na základě analýzy hlavních komponent a korelační matice (viz kapitola 5.3.1). Na základě vybraných faktorů byly následně vytvořeny shluky jednotlivých druhů rostlin, aby bylo možné popsat jejich závislost na těchto faktorech. Shlukování proběhlo pomocí metody k-means (viz kapitola 5.3.2). Součástí shlukové analýzy je také porovnání výsledků shlukování u dat čočky obsahujících informaci o zařazení ke genetickým skupinám s výsledky genotypování.

Dalším krokem je vizualizace vytvořených shluků, a to pomocí grafu biplot a pomocí mapových výstupů. Poslední analýzou je zjištění vlivu přesnosti bodových dat z hlediska využití environmentálních datových sad různé kvality. Zjištění vlivu přesnosti dat je popsáno v kapitole 5.3.3.

Posledním krokem, který je rozepsán v kapitole 6 je interpretace výsledků a výstupů.



Obr. 2 Diagram postupu práce.



### 3 STAV ŘEŠENÉ PROBLEMATIKY

Kvůli důležitosti planých rostlin a jejich ohrožení je nutné zachovat jejich přirozená stanoviště v Úrodném půlměsíci (Zair a kol., 2018) a také uchovat jejich genetické zdroje v ex situ kolekcích v genových bankách. Tuto důležitost si uvědomoval už Alphonse Pyramus de Candolle, který vydal v roce 1885 knihu *Origin of Cultivated Plants*, ve které zkoumal důvody geografického rozšíření pěstovaných rostlin a jejich původ (de Candolle, 1885). V roce 1926 ruský vědec Nikolaj Ivanovič Vavilov přišel s teorií geografických center rozmanitostí plodin. Tato teorie spočívá ve skutečnosti, že každá pěstovaná plodina pochází z určitého geografického místa, ze kterého se následně rozšířila do dalších oblastí. Tato centra se podle Vavilova nachází tam, kde má daná plodina nejvyšší genetickou rozmanitost (Vavilov, 1926).

#### 3.1 Oblast blízkého východu – Úrodný půlměsíc

Můžeme hovořit o čtyřech hlavních oblastech světa, ve kterých se začalo nezávisle na sobě vyvíjet zemědělství, a to přibližně ve stejnou dobu. Těmito oblastmi jsou okolí řeky Jang-c'-t'iang v oblasti dnešní Číny, kde byla domestikována rýže a sója, Jižní Amerika, přibližně oblast dnešního Peru a Bolívie, s domestikací bramboru a manioku, Střední Amerika, dnešní Mexiko, Honduras, s domestikací kukuřice a fazole a dále Úrodný půlměsíc, ve kterém byly domestikovány pšenice, ječmen, len, čočka, hrách a cizrna (Brown a kol., 2009).

V Úrodném půlměsíci došlo k domestikaci těchto plodin před více než 10 tisíci lety (Zeder, 2011) a jedná se o jeden z regionů s nejvyšší koncentrací CWR (Crop Wild Relatives – rostliny příbuzné divokým odrůdám) na jednotku plochy (Vincent a kol. 2013). Podle Zaira a kol. (2018) se v Úrodném půlměsíci nachází na 21 tisíc taxonů, z čehož 835 taxonů jsou CWR.

Úrodný půlměsíc je oblast v jihozápadní Asii a rozkládá se od Středoziemního moře až po Perský záliv. Zasahuje do území Egyptu, Izraele, Jordánska, Libanonu, Sýrie, Turecka, Iráku a Íránu. V Turecku je Úrodný půlměsíc ohraničen pohořím Antitaurus a od jihovýchodní části Turecka až po Perský záliv je lemován pohořím Zagros (viz obr. 3).

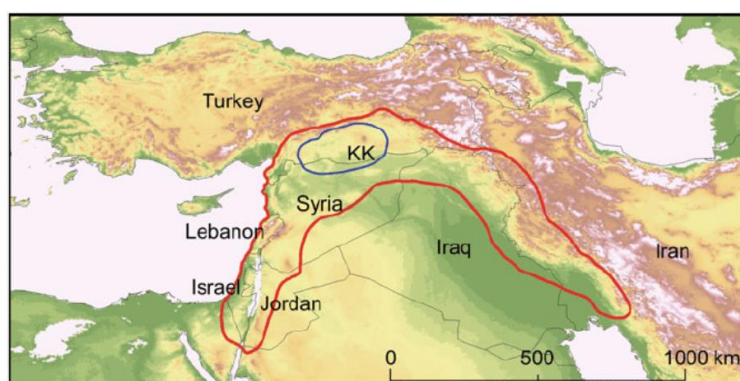
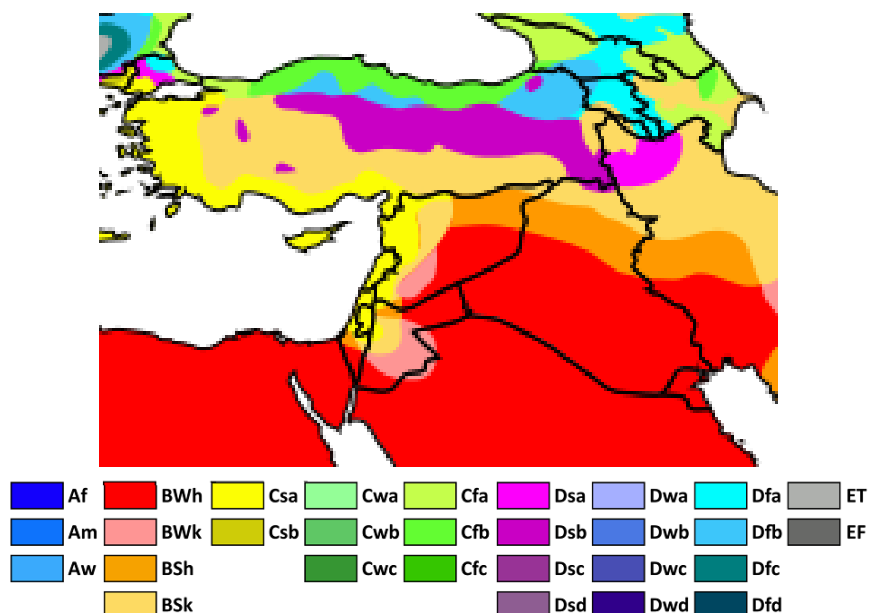


Fig. 1 Fertile Crescent and “core area” of plant domestication within the Fertile Crescent. The Fertile Crescent is indicated with a red line and the “core area” is shown with a blue line. KK Karacadag mountain range in south-eastern Turkey

Obr. 3 Oblast Úrodného půlměsíce (zdroj: Kilian, 2010).

Oblast Úrodného pŕľmĕsice je charakteristick vpencovm podlzm s malmi oblastmi tvořenmi ědiěem (Smykal a kol., 2018). Dky sv rozloze lze rodn pŕľmĕsc rozdlit na nkolik oblast podle klimatickch pomĕrŕ s pouztm Kppen-Geigerovy klimatick klasifikace (viz obr. 4).



Obr. 4 Kppen-Geiger klimatick klasifikace (zdroj: Physical Geography, 2020).

Kppen-Geigerova klimatick klasifikace obsahuje 29 klimatickch oblast (Murray, 2012; Beck, 2018), pŕicemz v rodnm pŕľmĕsci se jich nachz 6:

#### **Csa:**

- **Mrn klima** (teplota nejteplejho mĕsce je vĕt nebo rovna 10 °C a teplota nejchladnĕjho mĕsce je vy než 0 °C a men než 18 °C) **se suchmi** (v nejsum letnm mĕsci spadne mĕnĕ než 40 mm srzek a srzky nejsuho mĕsce v ltĕ jsou men než tŕetina srzek nejvlhěho mĕsce v zimĕ) **a horkmi** (teplota nejteplejho mĕsce je vy nebo rovna 22 °C) **lty**.
- Mrn klima se suchmi a horkmi lty se nachz na pobŕez Stŕedozemnho moŕe v Turecku, Libanonu, Srii, Izraeli, na Kypru a z ěsti tak v Jordnsku.

#### **BSk:**

- **Such** (prŕmĕrn roěn srzky jsou men než 10 × prh; prh = 2 × prŕmĕrn roěn teplota, pokud vce jak 70 % roěnch srzek spadne v zimĕ; prh = 2 × prŕmĕrn roěn teplota + 28, pokud 70 % roěnch srzek spadne v ltĕ; jinak prh = 2 × prŕmĕrn roěn teplota + 14) **stepn** (prŕmĕrn roěn srzky jsou vy nebo rovny 5 × prh (viz ve)) **chladn** (prŕmĕrn roěn teplota je men než 18 °C) **oblasti**.
- Such stepn chladn oblasti le dl ve vnitrozemi nĕz Csa. Nachz se v centlnm a jihovchodnm Turecku, dle ve stŕedozpadn ěsti Srie a v severn ěsti Irku.

**BSh:**

- **Suché** (viz BSk) **stepní** (viz BSk) **horké** (průměrná roční teplota je větší nebo rovna 18 °C) **oblasti**.
- Suché stepní horké oblasti leží v severovýchodní části Sýrie, v severní části Iráku a ve středozápadní části Íránu.

**BWh:**

- **Suché** (viz BSk) **pouštní** (průměrné roční srážky jsou menší než 5 × práh (viz BSk)) **horké** (viz BSh) **oblasti**.
- Suché pouštní horké oblasti se rozprostírají v jihovýchodní Sýrii, v jižní polovině Iráku a v jihozápadní části Íránu.

**BWk:**

- **Suché** (viz BSk) **pouštní** (viz BWh) **chladné** (viz BSk) **oblasti**.
- Suché pouštní chladné oblasti se nachází v jižní části Sýrie a Jordánska.

**Dsb:**

- **Chladné klima** (teplota nejteplejšího měsíce je větší nebo rovna 10 °C a teplota nejchladnějšího měsíce je menší nebo rovna 0 °C) **se suchým** (viz Csa) **a teplým** (když teplota nejteplejšího měsíce není větší nebo rovna 22 °C a počet měsíců s teplotou vyšší nebo rovnou 10 °C je větší nebo rovno 4) **létem**.
- Chladné klima se suchým a teplým létem leží ve střední a východní oblasti Turecka.

Podle Zaira a kol. (2018) bude pravděpodobně docházet k úbytku rozmanitosti CWR z důvodu nepříznivých změn klimatu, a proto je třeba využít maximální potenciál současné genetiky a GIS metod, které pomohou při vývoji strategie pro zachování CWR v Úrodném půlměsíci.

### 3.2 Popis zkoumaných rostlin

V 60. letech minulého století byly upřednostňovány obiloviny, které byly pěstovány na vysoce úrodné půdě na úkor bobovitých rostlin, které byly upozaděny a byly pěstovány v méně úrodných oblastech (Coba de la Peña a Pueyo, 2012), ačkoliv jsou primárním zdrojem nutričních bílkovin pro asi 30 % obyvatel světa (Afshin a kol., 2014). V této práci jsou zkoumány plané příbuzné druhy hrachu, cizrny a čočky.

#### **Hrách (*Pisum sp.*)**

Divoký *Pisum elatius* pochází z evropsko-středomořské oblasti a ze střední a severozápadní Asie, s největší rozmanitostí na Blízkém východě (Smýkal a kol., 2017). Semena divokého hrachu klíčí většinou na podzim po horkém a suchém létě. Ke kvetení dochází časně z jara, díky čemuž vysoké teploty negativně neovlivňují rostlinu během květu. Teplota a srážky jsou tedy nejdůležitějšími environmentálními faktory pro dormanci a klíčení semen (Probert, 2000). CWR jako je divoký hrách, čočka či cizrna jsou rostliny dlouhodobní, které potřebují delší den (min. 12 hodin) k tomu, aby se zahájila fáze kvetení (Weller a kol., 2012).

### **Cizrna (*Cicer sp.*)**

Rod *Cicer* zahrnuje celkem 42 druhů z nichž některé jsou víceleté nebo jednoleté. Planý předchůdce kulturní cizrny je *Cicer reticulatum* a sesterský druh *Cicer echinospermum*. V současné době je znám výskyt obou těchto druhů pouze na území jihovýchodního Turecka. *Cicer echinospermum* se vyskytuje především na čedičových substrátech na otevřených pastvinách s nižší nadmořskou výškou, zatímco *Cicer reticulatum* roste na pískovcových nebo na žulových substrátech ve smíšených pastvinách. *Cicer reticulatum* má ve svém přirozeném prostředí dlouhou vegetační sezónu a obvykle dozrává později než domestikovaná cizrna (Coyne a kol., 2020).

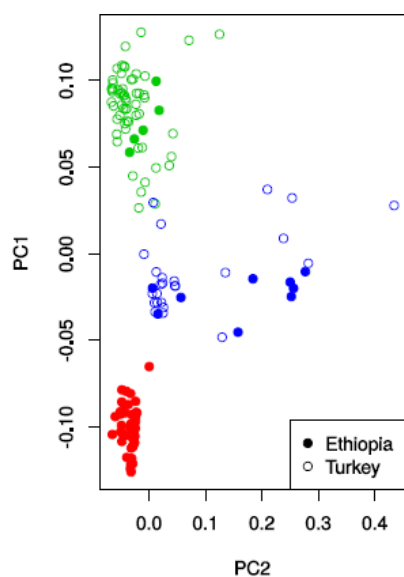
### **Čočka (*Lens sp.*)**

Je několik druhů divoce rostoucích čoček a v této práci byla analyzována *Lens orientalis*, jako přímý předchůdce kulturní čočky. Ta se v přirozeném prostředí vyskytuje v Úrodném půlměsíci a sahá na východ až do Tádžikistánu a na sever do Arménie. Roste na kamenitých vápenatých až čedičových půdách na otevřených nebo částečně zastíněných stanovištích. *Lens orientalis* má oproti jiným druhům čočky hlubší kořeny díky čemuž se lépe vyrovnává s nedostatkem vody (Coyne a kol., 2020).

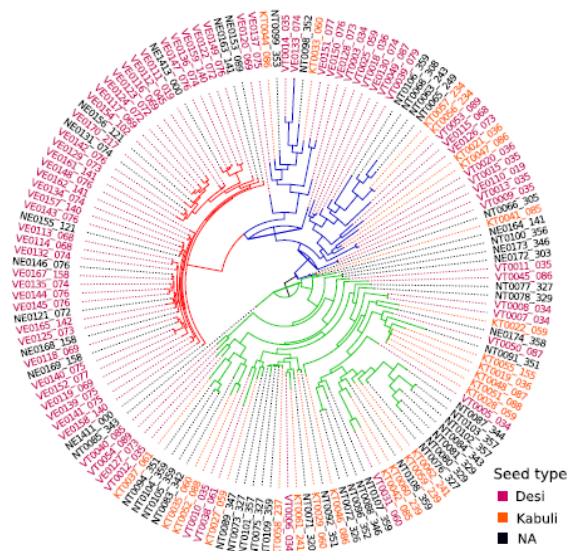
## **3.3 Řešená problematika v odborných studiích**

Vzhledem k důležitosti zachování a možnému využití genetické rozmanitosti divokých předchůdců plodin je věnována pozornost jejich studiu, ať už více z pohledu genetiky nebo environmentálních charakteristik.

Jednou z nich je studie od výzkumníků z ústavu – The Vavilov Institute of Plant Genetic Resources (VIR) v Petrohradu. VIR má genetickou banku založenou již během působení N. I. Vavilova, která odráží genetickou a funkční různorodost plodin. Zaměřují se například na využití historických dat, jak bylo nedávno publikováno na příkladu cizrny (Plekhanova a kol. (2017)). Autoři zkombinovali genomiku, bioinformatiku a fenotypovou charakterizaci pro analýzu vzorků cizrny pro oblast Turecka a Etiopie. Pomocí genotypizace určili přes 14 tisíc polymorfismů, jejichž podíl mezi přístupy je silným prediktorem fenotypové a environmentální podobnosti. Podobnost rostlin zkoumali mimo jiné pomocí analýzy hlavních komponent (PCA), která odhalila tři shluky podle výskytu rostlin (viz obr. 5). Dále vytvořila tři shluky, které jsou shodné s genetickou strukturou podle dendrogramu (viz obr. 6).



Obr. 5 Shluky podle metody PCA (zdroj: Plekhanova a kol., 2017).



Obr. 6 Dendrogram shluků (zdroj: Plekhanova a kol., 2017).

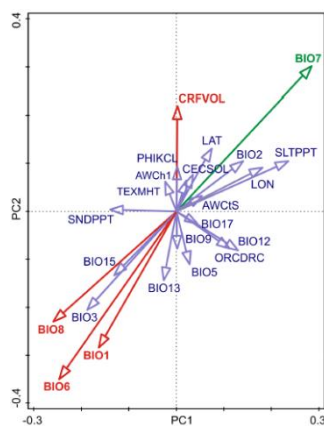
Dva ze zmíněných shluků jsou cizrny pocházející z Turecka a další shluk odděluje vzorky pocházející z Etiopie. Zjistili mírnou negativní korelaci teploty a doby květu a pozitivní korelaci s ročním úhrnem srážek a nadmořskou výškou. Dále byla zjištěna mírná korelace mezi teplotou a reprodukčními vlastnostmi (počet lusků, doba do květu) a negativní korelace mezi nadmořskou výškou, srážkami a reprodukčními vlastnostmi.

Další výzkum byl zaměřen na genetickou strukturu divokého hrachu v severní části Úrodného půlměsíce. Smýkal a kol. (2018) analyzovali vzorky 14 populací divokého hrachu v JV části Turecka, přičemž jedna populace obsahovala 5-22 jedinců. Nejčastěji se jednalo o rostliny rostoucí samostatně, kdy nejbližší rostlina byla vzdálena do 10 metrů, nebo se nacházely ve shlucích dvou až pěti rostlin.

Mezi vybrané klimatické aspekty, které použili z databáze WorldClim z let 1970-2000 patří například roční průměrná teplota, maximální a minimální teplota, maximální a minimální teplota nejchladnějšího a nejteplejšího měsíce, roční srážky, srážky v nejteplejším a nejchladnějším čtvrtletí. Pro morfometrické charakteristiky využili digitální model terénu ASTER GDEM, z něhož vypočítali sklon, orientaci a další indexy. Data byla převedena do ESRI gridu v měsíčních intervalech o prostorovém rozlišení 30 úhlových vteřin (cca 1 km) v souřadnicovém systému WGS84. Pro vizualizaci genetické struktury v prostoru provedli prostorovou analýzu hlavních komponent, pomocí níž byly odhaleny prostorové shluky a shrnuta genetická rozmanitost. Pomocí Bayesovské analýzy, DAPC (Discriminant analysis of principal components), PCA a SplitsTree zjistili, že genetická diverzita a velikost populace vykazují korelaci, což vede k závěru, že čím je populace dané rostliny větší, tím je vyšší genetická rozmanitost populace. Dále bylo zjištěno, že populace divokého hrachu mají podobnou genetickou strukturu až do vzdálenosti 60 km. Z toho vyplývá menší pravděpodobnost nalezení podobných jedinců u vzdálenějších populací. Autoři došli k závěru, že v budoucnu kvůli změně klimatu ubude vhodných míst pro růst divokého hrachu a současné stanoviště se přesunou do vhodnějších oblastí. Dá se očekávat, že v budoucnu bude mít změna klimatu vliv na růst a reprodukci populací planého hrachu, a to buď fyziologickým poškozením (teplotní stres během kvetení)

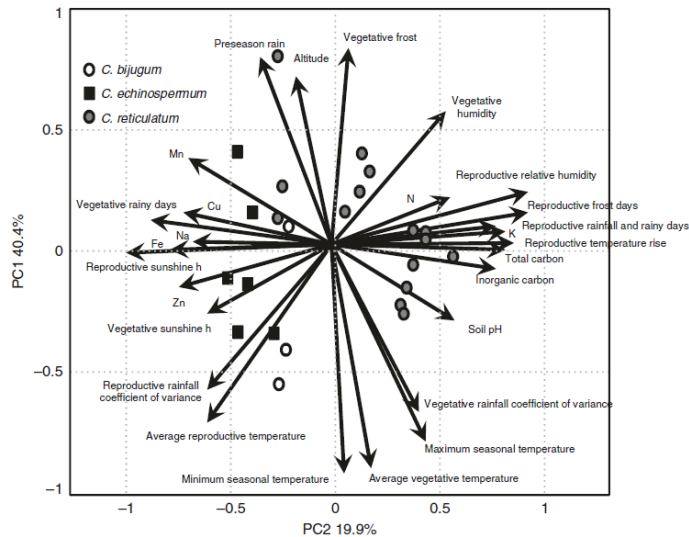
nebo úbytkem opylovačů v dané oblasti. Výsledky výzkumu divokého hrachu v oblasti Úrodného půlměsíce v Turecku ukazují, že rozdíly mezi populacemi jsou dány více geografickou vzdáleností než environmentální vzdáleností. Genetickou rozmanitost dané populace více ovlivňuje genetické křížení uvnitř populací a mezi příbuznými populacemi, které jsou v blízké vzdálenosti než vliv okolního prostředí.

Hradilová a kol (2019) popisovali vliv prostředí na dormanci semen divokého hrachu. Pro získání environmentálních charakteristik použili data o klimatu z WorldClim, konkrétně bioklimatické faktory BIO1 až BIO19 a data o půdě ze SoilGrids, která byla v rozlišení 1 km a měřena v 5 cm hloubce v půdě. Po získání všech potřebných dat bylo potřeba určit, zda se vstupní proměnné navzájem ovlivňují, a právě kvůli silné kovarianci bylo vyloučeno 8 půdních a 9 klimatických proměnných. Pro každou zbývající proměnnou byla vypočtena prostorová korelace pomocí metody Moranovo I. Zkoumané charakteristiky hrachu významně korelovaly pouze s bioklimatickými proměnnými BIO1, BIO6, BIO7, BIO8 a s obsahem uhlíku v půdě. S klíčovostí, tedy ukončením dormance semen, která byla hlavním předmětem výzkumu významně korelovala pouze proměnná BIO7 (viz obr. 7).



Obr. 7 Graf komponentních vah  
(zdroj: Hradilová a kol., 2019).

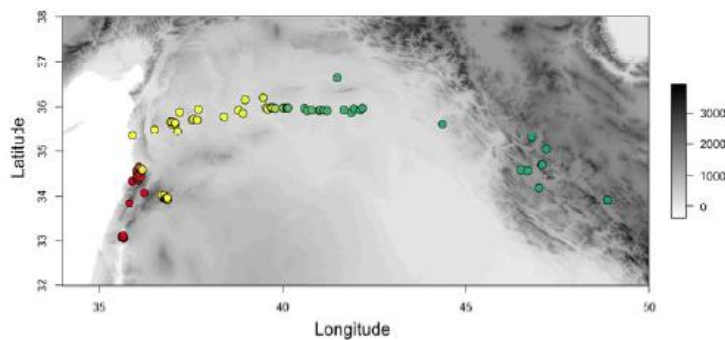
Předmětem studie von Wettberga a kol. (2018) byla planá cizrna (*C. reticulatum*) jejíž stanoviště porovnávali na základě genetických a environmentálních proměnných. Pro vyhodnocení variace lokalit bylo pracováno s teplotou, relativní vlhkostí, texturou půdy, chemickým složením půdy a dalšími faktory. Tyto environmentální faktory byly následně vyhodnoceny pomocí metody PCA (viz obr. 8).



Obr. 8 Biplot (zdroj: von Wettberg a kol., 2018).

Při vyhodnocení genetických vlastností rostlin na základě geografických proměnných zjistili, že nadmořská výška koreluje s genomovými podpisy více než geografická vzdálenost. Podle jejich výpočtů se u jednotlivých divokých populací jeden metr nadmořské výšky rovná okolo 11,2 km zeměpisné vzdálenosti.

Podobně pak Brunazzi a kol. (2018) zkoumali divokou pšenici v oblasti Úrodného pŕlměsíce. Pro geografickou analýzu využili 19 bioklimatických proměnných z WorldClim, na které aplikovali metodu PCA a Bayesovskou analýzu. Použili data o divoce rostoucí pšenici napříč Úrodným pŕlměsícem. Pomocí Bayesovské analýzy vyčlenili dva hlavní shluky na základě genetické informace. Tyto dva shluky rozdělávaly pšenici podle geografické polohy na západní a východní. Následně provedli prostorovou analýzu hlavních komponent, ze které vzešly tři hlavní shluky frekvencí alel lineárně rozmístěné od západu na východ (viz obr. 9), což je dovedlo k závěru, že genetická vzdálenost souvisí s geografickou vzdáleností.



Obr. 9 Shluky plané pšenice (zdroj: Brunazzi a kol. 2018).

Zair a kol., 2018 se ve své studii zabývali zvýšením ochrany CWR v oblasti Úrodného pŕlměsíce. Vytvořili seznam a inventář CWR a navrhli strategii pro zachování CWR v této oblasti, kterou lze využít na národní, regionální i globální úrovni.



## 4 GOOGLE EARTH ENGINE

Google Earth Engine (GEE) je platforma sloužící pro vizualizaci, stahování a analýzu geoprostorových dat v cloudovém prostředí. Platforma obsahuje rozsáhlý katalog satelitních snímků a geoprostorových datasetů nejrůznějších fyzicko-geografických a environmentálních témat. GEE pomáhá výzkumníkům, vývojářům, neziskovým organizacím, vládám i široké veřejnosti zkoumat povrch Země, detekovat změny v čase a vytvářet komplexní výstupy (Google Earth Engine). Využití GEE je zdarma pro výzkum, neziskové organizace a pro studium. V případě komerčního využití je potřeba kontaktovat příslušné oddělení. Hlavní součásti GEE jsou:

- datové sady,
- výpočetní infrastruktura schopná paralelního výpočtu geoprostorových dat,
- API pro JavaScript a Python, pomocí kterých se zadávají požadavky na servery Earth Engine,
- editor kódu, což je online vývojové prostředí (IDE – Integrated Development Environment) pro tvorbu kódů, pro práci s daty a vizualizaci prostorových analýz pomocí API Java Scriptu (Google Earth Engine, 2019).

GEE obsahuje archiv satelitních snímků a datových sad starých i více než 40 let. Datové sady se dělí do tří základních skupin, a to na podnebí a počasí, satelitní snímky a geofyzikální data.

### Podnebí a počasí

První podskupinou ze skupiny *Podnebí a počasí* je *Povrchová teplota*, která v současnosti obsahuje 31 datových sad. Povrchová teplota se získává tepelnými satelitními senzory, které detekují povrchovou teplotu a emisivitu. Mimo to je možné získat i povrchovou teplotu moří a oceánů z odvozených dat ze senzorů MODIS (The Moderate Resolution Imaging Spectroradiometer), ASTER (Advanced Spaceborne Thermal Emission and Reflection Radiometer) a AVHRR (Advanced Very-High-Resolution Radiometer).

Druhou podskupinou je *Klíma* obsahující 50 datových sad. Jejich součástí jsou data s dlouhodobými předpověďmi, současným nebo historickým stavem klimatu. Příkladem mohou být datové sady FLDAS nebo TerraClimate.

Třetí podskupinou je *Atmosféra* s šesti datovými sadami. Tyto datové sady mohou sloužit jako korekční data k opravě dat z dálkového průzkumu Země. Příkladem těchto dat jsou údaje o ozonu nebo vodní páře v atmosféře.

Poslední podskupinou je *Počasí*, která obsahuje 27 datových sad. Tato data obsahují měřené i předpovídané údaje o srážkách, teplotě, větru, vlhkosti a dalších proměnných. V této podskupině lze nalézt například datové sady WorldClim verze 1, Global Forecast System nebo datovou sadu NOAA NHC HURDAT2 s historickými trajektoriemi hurikánů v Atlantském či Tichém oceánu (Earth Engine Data Catalog c).

### Satelitní snímky

Skupina *Satelitní snímky* obsahuje snímky ze systému Landsat, Sentinel, MODIS a snímky s vysokým rozlišením.

Landsat je program vedený agenturami USGS (United States Geological Survey) a NASA, jehož satelitní snímky jsou k dispozici od roku 1972 do současnosti.



Tyto snímky z družic Landsat 1 až Landsat 8 lze v Google Earth Engine získat v rozlišení až 30 metrů s granularitou dvou týdnů. V současnosti jsou satelitní snímky včetně multispektrálních a termálních dat dostupné pro celou Zemi. Data jsou v Google Earth Engine k dispozici v nezpracované podobě, s atmosférickými korekcemi nebo s připravenými vrstvami jako například vypočítané vegetační indexy NDVI a EVI (Earth Engine Data Catalog a).

Sentinel je seskupení satelitů vyvinutých agenturou ESA (European Space Agency) v rámci programu Copernicus. V Google Earth Engine lze získat radarová data ze Sentinel-1 dostupná od roku 2014, multispektrální snímky s vysokým rozlišením ze Sentinel-2 od roku 2015, které primárně slouží pro monitoring vegetace, půdy a vodních ploch, dále data o oceánech, ledovcích a atmosféře ze Sentinel-3 od roku 2016 a nakonec složení troposféry monitorované z družice Sentinel-5P od roku 2018 (Earth Engine Data Catalog b).

MODIS jsou senzory na satelitech Terra a Aqua od agentury NASA, které od roku 1999 pořizují satelitní snímky s 16denním časovým rozlišením, které obsahují odrazivost povrchu a odvozená data, jako například vegetační indexy a sněhovou pokrývku Země. Google Earth Engine nabízí 25 datových sad z MODIS.

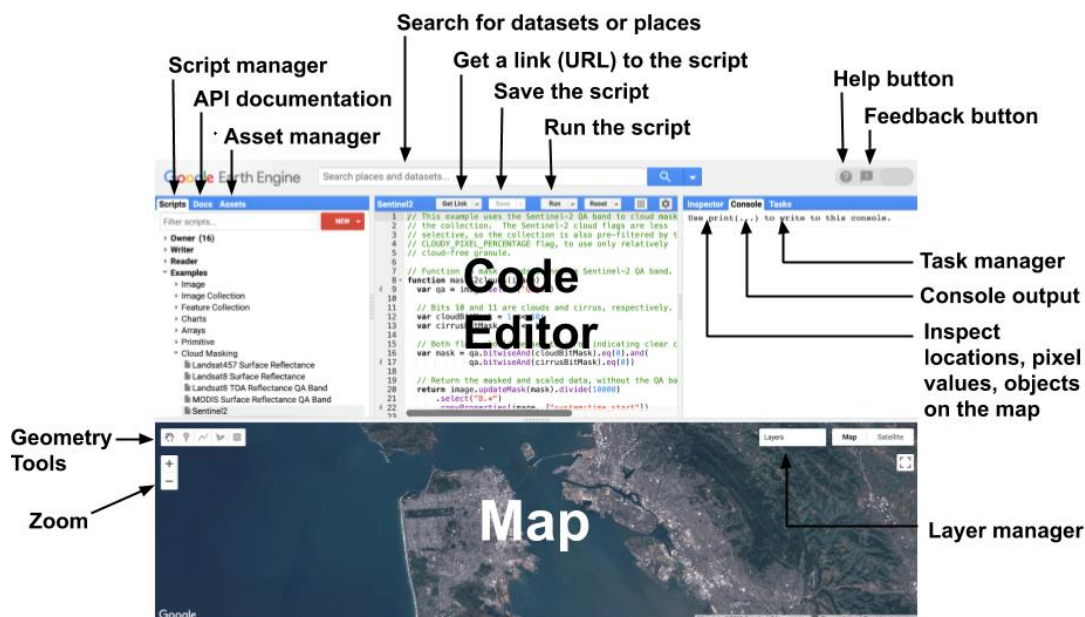
Podskupina *snímky vysokého rozlišení* obsahuje tři datové sady. Dvě z nich jsou satelitní snímky Planet SkySat od společnosti Planet Labs Inc. a třetí jsou letecké snímky Spojených států od NAIP (The National Agriculture Imagery Program).

### **Geofyzikální data**

První podskupinou je *Terén*, ve kterém jsou obsaženy digitální výškové modely Země. Katalog Google Earth Engine obsahuje 36 výškových modelů, včetně několika globálních, jako je SRTM30 nebo GTOPO30. Druhou podskupinou je *Krajinný pokryv*, jehož součástí je 20 datových sad, mezi které se řadí CORINE Land Cover, GlobCover nebo MODIS Land Cover. Dalšími geofyzikálními daty v GEE jsou *Pole* (cropland) obsahující tři datasety z USDA NASS a GFSAD1000 od agentury NASA. Kromě dat z těchto tří podskupin lze z Google Earth Engine získat i jiná geofyzikální data, jako jsou noční satelitní snímky znázorňující noční osvětlení (Earth Engine Data Catalog c).

## **4.1 Práce v Google Earth Engine**

Pro přístup do editoru Google Earth Engine je nutné se přihlásit pomocí Google účtu na oficiálních webových stránkách <https://earthengine.google.com/> a v záložce Platform se nachází přístup do Code Editoru, ve kterém probíhá veškerá práce. Editor se skládá z pěti hlavních částí (panelů) (viz obr. 10). Nahoře se nachází vyhledávač datasetů, lokalit a tabulek.

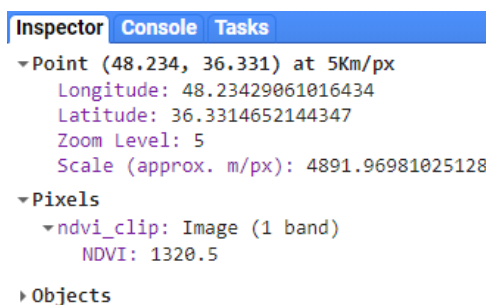


Obr. 10 Rozhraní Google Earth Engine (zdroj: Google Earth Engine, 2020).

Levý panel obsahuje tři záložky, a to Scripts, Docs a Assets. Záložka Scripts obsahuje uložené vlastní skripty, sdílené skripty a vybrané ukázkové skripty od společnosti Google. Je zde možné vytvářet vlastní složky a zřehlednit či seřadit skripty podle vlastního uvážení. Skripty je možné přejmenovat, mazat nebo se vrátit v historii daného skriptu podle data uložení. Záložka Docs obsahuje JavaScript API dokumentaci obsahující jednotlivé příkazy a jejich popis. Záložka Assets slouží k nahrávání a správě vlastních dat nahraných do GEE. Do GEE lze nahrávat snímky ve formátech GeoTIFF (.tif, .tiff) a TFRecord (.tfrecord + .json), dále vektorové vrstvy ve formátu Esri Shapefile (.shp, .shx, .dbf, .prj nebo .zip) a textové soubory ve formátu CSV (Comma Separated Values). V záložce Assets je možné si vytvořit vlastní kolekci snímků, popřípadě vlastní složky pro větší přehlednost. Dále je možné vlastní nahraná data sdílet nebo je velmi jednoduše vložit do skriptu a následně s nimi dále pracovat.

Prostřední panel je editor skriptů, ve kterém probíhá veškeré skriptování kódu. Jsou zde tlačítka na uložení, spuštění, resetování a sdílení vlastního skriptu.

Pravý panel obsahuje záložky Inspector, Console a Tasks. Záložka Inspector umožňuje interaktivní dotazování do mapy. Například pokud je spuštěn skript, který zobrazuje danou vrstvu v mapě, tak po najetí na záložku Inspector se změní kurzor šipky na kříž, pomocí něhož lze kliknutím do mapy zobrazit hodnoty pixelů, objektů a lokací v mapě (viz obr. 11).

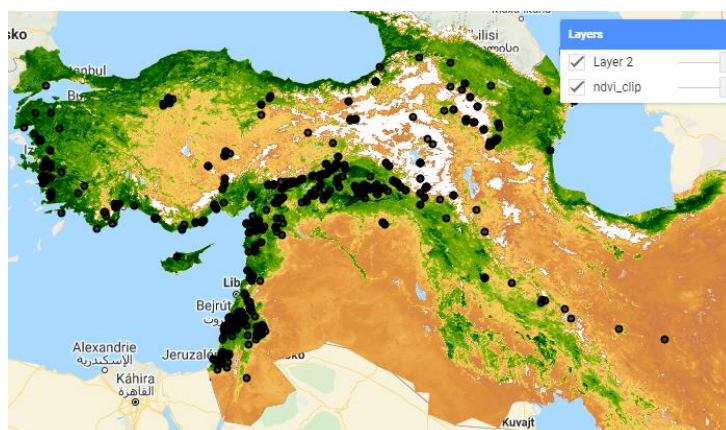


Obr. 11 Záložka Inspektor v GEE (zdroj: vlastní).

Záložka Console slouží pro zobrazení hodnot, objektů, textů nebo grafů, pokud je ve skriptu použit příkaz *print()* pro požadované zobrazení (například se zobrazí kolik je dostupných snímků v daném časovém období a kolik záznamů je v souboru Shapefile. V záložce Console se také zobrazují chybové hlášky kódu.

Další záložka tohoto panelu je Tasks, která slouží pro stahování souborů vytvořených skriptem. Doba stahování souborů závisí na velikosti a počtu procházených snímků v cloudovém uložení Googlu, na komplexnosti požadavku a na množství vlastních vstupních dat jako je například počet prvků v Shapefile. Soubory je možné uložit na Google Drive, na Cloud Storage nebo do EE Assetu. Pokud je vybráno uložení do Google Drive, je zde možnost vložit název složky z Drive a uložit výsledek přímo do konkrétní složky. Na Google Drive a Cloud Storage je možné soubory uložit do formátů CSV, GeoJSON, KML, KMZ, SHP a TFRecord.

Poslední částí online vývojového prostředí je mapa. Tento panel obsahuje geometrické nástroje pro rychlé vytváření bodů, linií, obdélníků či jiných polygonů. Výhodou těchto geometrických nástrojů je, že se nově vytvořená geometrie ihned automaticky zapíše do skriptu a lze s ní dále jednoduše pracovat. Dále se v mapě zobrazují jednotlivé vrstvy, mezi kterými lze jednoduše přepínat (viz obr. 12), ale to pouze pokud je v daném skriptu nastaveno zobrazení vrstev do mapy.



Obr. 12 Zobrazení datasetu MOD13Q1.006 Terra Vegetation Indices 16-Day Global 250m a bodové vrstvy hrachu (zdroj: vlastní).

## 5 VLASTNÍ ŘEŠENÍ

V praktické části bylo nejprve nutné získat potřebná data a zpracovat je do vhodné podoby, aby bylo možné nad nimi provádět analýzy. Tato kapitola obsahuje postup získání a zpracování dat a také postup pro tvorbu analýz, jejichž výsledky jsou popsány v kapitole 6.

### 5.1 Postup zpracování bodových dat rostlin

Nejprve bylo třeba pročistit data o rostlinách obdržena od doc. Smýkala. Jak již bylo zmíněno, jednalo se o .xls soubory, které bylo třeba upravit tak, aby bylo možné s nimi dále pracovat v dalších analýzách. Z tabulek bylo nutné odstranit redundantní záznamy, záznamy rostlin s chybějící polohou nebo s nízkou polohovou přesností a sjednotit názvy tabulek a atributů.

Nejprve byly tabulky upraveny v softwaru Excel, kde byly smazány záznamy s chybějící polohou a upraveny atributy pro možnost práce v softwaru ArcGIS Desktop. Do tohoto softwaru byla tabulka vložena pomocí nástroje *Make XY Event Layer*, která umožňuje vytvořit bodovou vrstvu na základě zeměpisných souřadnic. Jako souřadnicový systém byl nastaven GCS\_WGS\_1984 (EPSG kód: 4326). Pomocí atributové tabulky byly zjištěny duplicitní záznamy, které měly stejný identifikační název (atribut „Name“). Pomocí nástroje *Summarize*, který byl spuštěn nad atributem Name, bylo zjištěno, kolik rostlin má v tabulce více než jeden záznam. Duplicitní údaje se nacházely pouze v tabulce hrachu, odkud bylo smazáno 43 duplicitních záznamů. Duplicita byla způsobena tím, že byla data agregována z několika zdrojů, které mohou obsahovat stejné údaje o rostlinách. Pro zajištění relevantních výsledků této práce bylo provedeno třídění dat na základě jejich polohové přesnosti. Smazány byly všechny body, které měly přesnost alespoň jedné ze souřadnic na méně než dvě desetinná místa, což značilo polohovou nepřesnost přibližně 11,1 až 111 km (viz tabulka 6).

Tabulka 6 Přibližná polohová přesnost GNSS

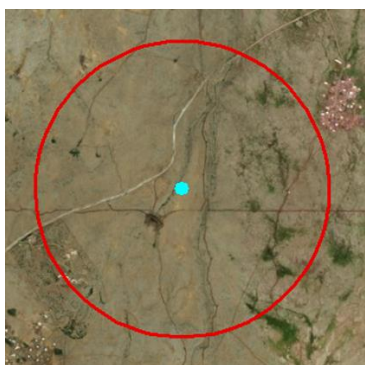
GNSS přesnost
1° = 111 km
0,1° = 11,1 km
0,01° = 1,11 km
0,001° = 0,111 km
0,0001° = 0,011 km

Zbývajícím záznamům byly podle přesnosti uděleny hodnoty 1–3 v nově vytvořeném atributu „Přesnost“. Význam těchto hodnot je znázorněn v tabulce 7. Toto rozdělení sloužilo k tomu, aby bylo možné v následujících analýzách pracovat buď se všemi záznamy nebo pouze s daty s vysokou polohovou přesností. Největší nepřesnost dosahovala hodnot maximálně okolo 1,5 km vzdušnou čarou, a to u přesnosti 3. Přesnosti 2 a 3 se objevovaly pouze u dat převzatých z genobank, data z terénního sběru byla nejpřesnější. Na základě konzultace s vedoucím práce a doc. Smýkalem vstupují do výsledných analýz pouze data s přesností 1.

Tabulka 7 Klasifikace dat dle přesnosti

		příklad	
		lat	lon
<b>přesnost 1</b>	obě souřadnice (lat, lon) mají přesnost alespoň na tři desetinná místa	35,682	36,532
<b>přesnost 2</b>	jedna souřadnice má přesnost na alespoň dvě desetinná místa a druhá alespoň na tři desetinná místa	37,75	39,766
<b>přesnost 3</b>	obě souřadnice mají přesnost alespoň na dvě desetinná místa	35,42	36,28

Na základě atributu přesnosti byly okolo každého bodu s přesností 2 nebo 3 vytvořeny buffery o poloměru 1,11 km. U bodů s přesností 2 (s vyšší nepřesností pouze u jedné souřadnice) se polohová chyba může vždy vyskytnout buď po rovnoběžce nebo poledníku, pro zjednodušení byla ovšem vytvořena kruhová oblast, která zahrnuje obě tyto možnosti (viz obr. 13). U přesnosti 3 může být chyba u zeměpisné šířky i zeměpisné délky 1,11 km, což vytvoří čtvercové území, ve kterém se bod může skutečně nacházet. Proto byly u bodů s přesností 3 okolo vytvořených bufferů pomocí nástroje *Minimum Bounding Geometry* vytvořeny ohraničující čtverce (viz obr. 14). Tyto ohraničující oblasti napomohly při následné selekci bodů, při které bylo vizuálně kontrolováno, zda se bod nachází v území zástavby či vodních ploch. V takovém případě byl bod smazán.



Obr. 13 Oblast chyby s přesností 2.



Obr. 14 Oblast chyby s přesností 3.

Dále bylo kontrolováno, zda se body vyskytovaly ve státech, které měly přiřazeny v původních datech a pokud ne, byla tato informace opravena. Počet bodů dle přesnosti je znázorněn v tabulce 8.

Tabulka 8 Počet rostlin podle přesnosti

Rostlina	Latinský název	Přesnost		
		1	2	3
Cizrna	<i>Cicer echinospermum</i>	710	12	8
	<i>Cicer reticulatum</i>	1652	13	2
Čočka	<i>Lens orientalis</i>	614	53	43
Hrách	<i>Pisum elatius</i>	534	52	130

## 5.2 Postup zpracování environmetálních dat

### Stažení dat

Jedním z důležitých kroků bylo stažení globálních datasetů environmentálních dat. Nejjednodušeji proběhlo stažení dat WorldClim, kdy stačilo datové vrstvy stáhnout z internetové stránky poskytovatele. Důležitou datovou sadou je výškový datový model SRTM30, který obsahuje informace o nadmořské výšce v rozlišení 30 metrů. Tato vrstva sloužila jako vstupní vrstva pro výpočet sklonu, orientace svahu a indexů TRASP, SEI, IMI, HLI a CTI.

Stažení SRTM dat proběhlo ze stránky <https://earthexplorer.usgs.gov/>, do které bylo nejprve třeba se registrovat. V záložce Search Criteria byla manuálně vybrána zájmová oblast a v záložce Data Sets byla jako datová sada vybrána SRTM 1 Arc-Second Global. V záložce Results byly vygenerovány všechny dlaždice zasahující do vybrané zájmové oblasti a každá tato dlaždice byla pomocí ikony Add to Bulk Download přidána do „košíku“ Item Basket. V Item Basket se po potvrzení požadavku vygeneroval kód, který bylo třeba vložit do softwaru Bulk Download Application. Výhodou stažení dat pomocí tohoto postupu je úspora času, protože lze všechna data stáhnout zároveň.

### Postup práce v GEE

Prvním krokem bylo v záložce Assets nahrání bodové vrstvy rostlin ve formátu Shapefile ve zkomprimované formě .zip. Vrstva byla následně pomocí Import into script vložena do skriptu.

Poté byly vytvořeny dva vzorové skripty. Jeden pro tzv. *ImageCollection*, což je dataset obsahující několik snímků z různého časového období (viz příloha 1). Skript byl vytvořen tak, aby byly pro jednotlivé body ze vstupní vrstvy získány informace ze snímků pro každý měsíc ze zadaného časového rozpětí. Výsledkem je nová tabulka ve formátu CSV, která je exportována do uživatelského úložiště na Google Drive. Tato tabulka obsahuje všechny atributy ze vstupní vrstvy, ke kterým jsou přidány atributy s průměrnými měsíčními hodnotami ze snímků pro každý bod.

Druhý skript (viz příloha 2) slouží pro získání informací z tzv. *Image*, což je pouze jeden snímek většinou obsahující průměrné hodnoty za delší časové období. Tento skript je jednodušší než předchozí zmíněný, jelikož pouze získává hodnoty z jednoho snímku pro body ze vstupní vrstvy. Výsledkem je opět nová tabulka CSV exportovaná do Google Drive stejně jako u předchozího skriptu. Po konzultaci s odborníky na dané téma byly vybrány datasey, které jsou vypsány v tabulce 5.

Dalším krokem po vytvoření funkčních vzorových skriptů bylo samotné stahování potřebných dat. Ze vzorových skriptů bylo nutné změnit následující (ve skriptech v přílohách označeno červeně):

- vstupní data (bodové vrstvy) v proměnné *inputData*,
- vstupní dataset (snímky), ze kterého se berou informace, v proměnné *dataset*,
- název atributu (tzv. bands) ze vstupního datasetu v proměnné *selectBand*,
- název výstupní tabulky v příkazu *Export.Table.toDrive*.

Pro *ImageCollection* bylo navíc třeba změnit:

- nastavení počátečního a koncového data v proměnných *startDate* a *endDate*.

Jakmile byly tyto příkazy nastaveny, pomocí tlačítka Run byl skript spuštěn a v pravé části v záložce Console byly vypsány všechny údaje z příkazů *print()*. V záložce Tasks byla připravena žádost na vytvoření požadované tabulky k exportu do Google Drive, která byla potvrzena pomocí tlačítka RUN. V následujícím dialogu bylo možné vybrat způsob exportu, přejmenovat název tabulky a vybrat požadovaný formát. Doba exportování se pohybovala v řádu desítek sekund pro Image a od půl hodiny do tři hodin pro *ImageCollection* v závislosti na velikosti vstupních dat.

Výsledná tabulka byla z uložště Google Drive stažena do počítače. Hodnoty v tabulce obsahovaly desetinné tečky, což způsobovalo problémy v softwaru Excel, kdy byly například číselné hodnoty automaticky předělány na datum. Pro zamezení těchto problémů při použití desetinné tečky bylo nutné soubor CSV nejprve otevřít v textovém editoru, kde byly desetinné tečky nahrazeny čárkami, čemuž muselo předcházet nahrazení čárek středníky pro možnost oddělení sloupců v tabulce. Po uložení byl předělaný soubor otevřen v softwaru Excel přes záložku *Data* → *Načíst externí data* → *Z textu*, kde byl jako oddělovač zvolen středník. Při očištění dat byl vymazán první a poslední sloupec s nerelevantními systémovými informacemi z Google Earth Engine. Tento postup byl následně aplikován na další dvě datové sady s rostlinami a na každý dataset s vybranými atributy ze snímků. Celkem bylo tedy z Google Earth Engine staženo 42 nových tabulek, z toho 18 tabulek bylo staženo pomocí prvního skriptu a 24 tabulek pomocí druhého skriptu.

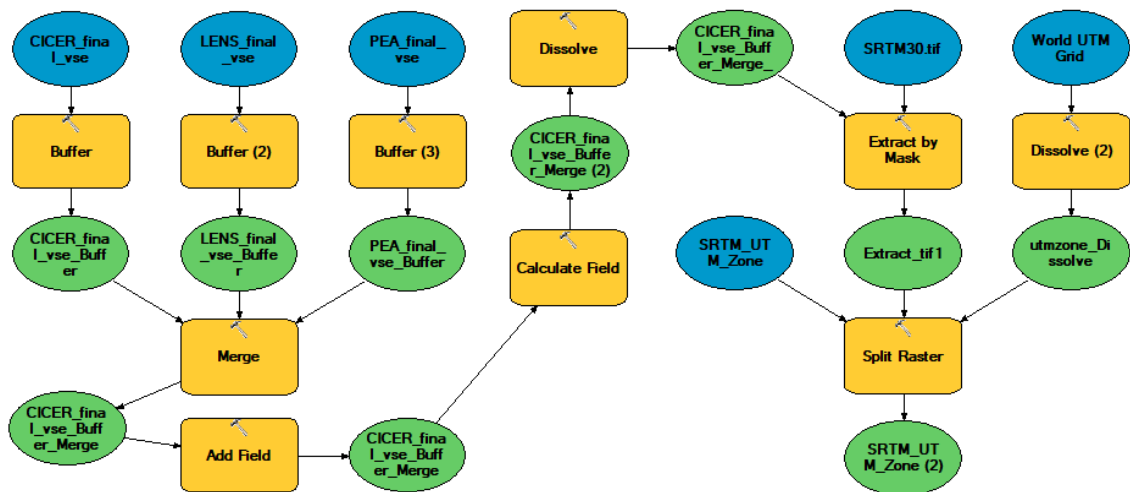
## **Zpracování dat**

Pro jednodušší získání informací z výškových dat SRTM30 a pro následnou práci s touto datovou sadou byly všechny dlaždice pomocí nástroje *Mosaic To New Raster* v softwaru ArcGIS Desktop spojeny do jednoho rastru. Pro snížení velikosti rastrových vrstev bylo provedeno jejich ořezání na oblast zájmu.

Pro usnadnění následující práce s rastrovými vrstvami byl v softwaru ArcGIS Desktop vytvořen toolbox obsahující nástroje vytvořené v Model Builderu. Nejprve byl vytvořen model, který slouží pro ořezání rastrů na zájmovou oblast z důvodu snížení velikosti vrstev a rychlejšího procesu. Do modelu vstupuje složka s rastry, které jsou všechny automaticky postupně ořezány na vybranou zájmovou oblast a výstupy jsou uloženy do nové složky.

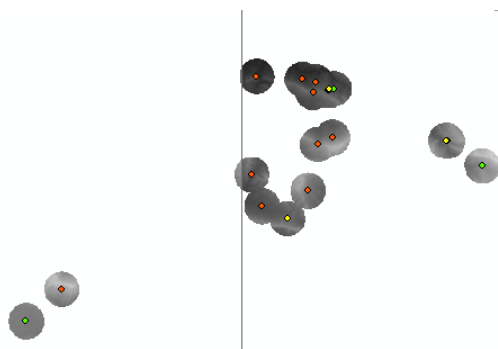
Další model je určen pro ořezání rastrové vrstvy SRTM30 a její následné rozdělení podle UTM zón (viz obr. 15). Vrstva UTM zón byla získána pomocí *Add Data From ArcGIS Online* a následně byla exportována do nové vrstvy. V tomto modelu jsou vytvořeny buffery o poloměru 2 km okolo každého bodu ze všech tří vrstev se zkoumanými rostlinami a ty jsou následně spojeny do jedné vrstvy. Do této vrstvy je přidán atribut, do kterého je u všech záznamů vyplněna hodnota 1, aby bylo možné ze všech záznamů ve vrstvě udělat pouze jeden. Tato vrstva slouží jako ořezová vrstva pro rastrovou vrstvu SRTM30, která je následně rozdělena podle UTM zón. Před tímto krokem jsou ještě UTM zóny spojeny do pásů nástrojem *Dissolve* podle atributu *Zone*. Rozdělení SRTM30 podle UTM zón bylo provedeno z toho důvodu, aby nedocházelo ke zkreslení vlivem nesprávného souřadnicového systému.





Obr. 15 Model pro úpravu SRTM30.

Výsledkem tohoto modelu je šest ořezaných rastrových vrstev SRTM podle bufferů rostlin, rozřezaných podle UTM zón, do kterých spadají (viz obr. 16).



Obr. 16 SRTM30 rozdělené podle UTM zón a ořezané podle bufferů.

Dalším krokem bylo nastavení souřadnicových systémů všem šesti nově vytvořeným vrstvám SRTM pomocí nástroje *Project Raster*. EPSG kód pro UTM začíná vždy 326xx a poslední dvě číslice jsou hodnoty daného pásu. V tomto případě 35 až 40.

Následně byly vytvořeny vrstvy sklonu svahu ve stupních a orientace svahu ze SRTM30. Pro výpočet se použily nástroje *Slope* a *Aspect*, které byly spuštěny pro všech šest UTM pásem rastrové vrstvy SRTM30.

Dále byly vypočítány indexy TRASP, SEI, IMI, HLI a CTI pomocí toolboxu *Geomorphometry & Gradient Metrics*. Vstupní vrstvou všech indexů bylo vždy pouze SRTM30 a nebylo třeba dalšího nastavení. Pouze u nástroje *Slope/Aspect Transformation* byl navíc jako *Transformation Type* zvolen TRASP.

Pro následné analýzy bylo nutné vložit informace ze všech environmentálních vrstev do datových sad rostlin. Pro tyto účely byl vytvořen model, který iteračně pomocí nástroje *Extract Multi Values to Points* vložil hodnoty pixelů z rastrů ke každému bodu. Aby proces proběhl správně, bylo nutné, aby měly rastrové vrstvy i vrstvy s rostlinami stejný souřadnicový systém.

Každá z datových sad s plodinami tak měla po tomto procesu vytvořeno několik nových sloupců v atributové tabulce, které obsahovaly hodnoty z rastrů SRTM30,



Bio1 až Bio19, TRASP, SEI, IMI, HLI, CTI, Aspect a Slope, spolu s atributy z Google Earth Engine.

Vzhledem k tomu, že některé datové sady z GEE byly staženy v měsíčním intervalu, bylo nutné pro vzájemné porovnání s ostatními environmentálními faktory tato data zprůměrovat za celé období, ve kterém byla stažena. U dat OpenLandMap, která obsahují hodnoty pro hloubky půdy 0, 10, 30 a 60 cm, byly vybrány pouze hodnoty z 10 cm hloubky, protože data ze všech hloubek spolu vysoce korelovala.

Na závěr zpracování dat byly smazány záznamy, ve kterých se objevovaly NULL hodnoty u atributů environmentálních faktorů, jelikož vytvoření korelační matice v softwaru RStudio nelze provést na datech s NULL hodnotami. Dále bylo smazáno několik záznamů kvůli hodnotám -1 u atributů se sklonem a orientací svahu. Po tomto třídění **do následných analýz vstupovalo 710 bodů cizrny (*Cicer echinospermum*), 1649 bodů cizrny (*Cicer reticulatum*), 604 bodů čočky, 109 bodů čočky obsahující genové skupiny a 524 bodů hrachu.**

## 5.3 Postup statistických analýz

Vyhodnocení vlivu environmentálních faktorů na výskyt planých předchůdců cizrny, čočky a hrachu bylo provedeno pomocí statistických analýz. Tato část práce probíhala v softwaru RStudio, ve kterém lze pomocí programovacího jazyka R vypočítat a vizualizovat různé typy statistických analýz. Pro statistické analýzy byly vytyčeny dva hlavní cíle. Prvním je výběr environmentálních faktorů, které nejvíce ovlivňují výskyt daných rostlin a druhým je vytvoření shluků rostlin na základě těchto vybraných environmentálních faktorů. Dále byl zjišťován vliv přesnosti polohy bodů rostlin v závislosti na prostorovém rozlišení environmentálních datových sad.

### 5.3.1 Zjištění důležitých environmentálních faktorů

Aby bylo možné vytvořit shluky rostlin na základě environmentálních faktorů, bylo nutné vybrat takové faktory, které mají na výskyt největší vliv a zároveň spolu navzájem nekorelují. Při vstupu všech faktorů do výsledného shlukování by mohlo dojít k výraznému zkreslení kvůli vysoké korelaci některých faktorů. Vyhodnocení bylo provedeno pomocí analýzy hlavních komponent (PCA) a korelační matice.

Do RStudia byly postupně nahrány předpřipravené tabulky s daty pomocí volby *Import Dataset* → *From Excel*. Každá tabulka obsahuje data za každou rostlinu pro 41 environmentálních faktorů. Těmi jsou SRTM30, bio01, bio02, bio03, bio04, bio05, bio06, bio07, bio08, bio09, bio10, bio11, bio12, bio13, bio14, bio15, bio16, bio17, bio18, bio19, TRASP, SEI, IMI, HLI, CTI, Aspect, Slope, Evap, Rainf, SoilMoi, CHILL, NDVI, SBulkb10, SCarbonb10, pHH2Ob10, STexturb10, SWatConb10, aet, pdsi, pr a soil. Pro vysvětlení názvů datových sad z Google Earth Engine viz tabulka 5.

Nejprve bylo nutné do RStudia nainstalovat balíčky FactoMineR, factoextra, corrplot a car. První dva zmíněné jsou určeny pro tvorbu PCA, další dva jsou zaměřeny na tvorbu korelační matice.

Tabulka s daty byla vložena do proměnné data, aby byla další práce usnadněna.

```
data <- CICER_echinospermum_presnost1
```

Následně byla provedena PCA z proměnné data. Pomocí části kódu `scale.unit = TRUE` se původní data přímo standardizují, aby nedošlo ke zkreslení výsledku. Údaj `data[,9:49]` určuje, se kterými atributy z dané tabulky se bude pracovat. Tento kód ve výchozím nastavení zobrazuje pouze 8 hlavních komponent, proto byla do kódu přidána část `ncp = 10`, která určuje počet hlavních komponent, které mají být zobrazeny.

```
data.pca <- PCA(data[,9:49],scale.unit = TRUE, graph = TRUE, ncp = 10)
print(data.pca)
```

Dalším příkazem se vypíše procentuální zastoupení všech dat v každé komponentě. Výsledkem je tabulka, pomocí níž lze zjistit, kolik procent dat je celkem obsaženo při využití určitého počtu komponent (viz obr. 17).

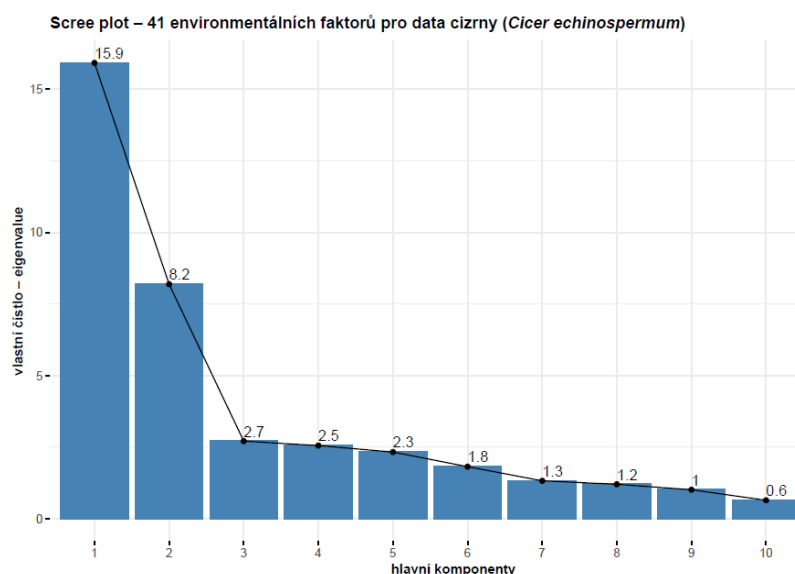
```
eig.val <- get_eigenvalue(data.pca)
print(eig.val)
```

	eigenvalue	variance.percent	cumulative.variance.percent
Dim.1	1.592365e+01	3.883817e+01	38.83817
Dim.2	8.189696e+00	1.997487e+01	58.81304
Dim.3	2.705991e+00	6.599977e+00	65.41302
Dim.4	2.549324e+00	6.217862e+00	71.63088
Dim.5	2.323644e+00	5.667425e+00	77.29830
Dim.6	1.812237e+00	4.420091e+00	81.71839
Dim.7	1.318306e+00	3.215380e+00	84.93377
Dim.8	1.198033e+00	2.922032e+00	87.85581
Dim.9	1.005844e+00	2.453277e+00	90.30908
Dim.10	6.388825e-01	1.558250e+00	91.86733

Obr. 17 Ukázka tabulky zobrazující procenta dat v prvních 10 dimenzích dat *Cicer echinospermum*.

Dále se vytvoří scree plot, což je sloupcový graf, který znázorňuje vlastní hodnoty (eigenvalues) na základě, kterých lze pomocí Keiserova kritéria (vlastní číslo je větší než 1) určit počet důležitých dimenzí (hlavních komponent) (viz obr. 18).

```
fviz_screplot(data.pca,addlabels = TRUE, geom = c("bar", "line"), choice = "eigenvalue")
summary(data.pca)
```



Obr. 18 Ukázka scree plotu pro data *Cicer echinospermum*.

V této fázi je znám počet dimenzí, se kterými se bude dále pracovat. Dále je třeba určit zastoupení všech environmentálních faktorů v každé dimenzi. Pomocí využití `cos2` se zastoupení faktorů v každé dimenzi navíc provádí důležitostí dimenze (větší váhu mají faktory zastoupené v 1. dimenzi).

```
var <- get_pca_var(data.pca)
head(var$cos2, 41)
```

Z tohoto výsledku se do softwaru Excel vloží hodnoty pro počet dimenzí určených pomocí scree plotu. Pro každý environmentální faktor se udělá součet hodnot ze všech těchto dimenzí a tyto součty se seřadí sestupně dle velikosti (viz obr. 19). Díky tomu lze určit, které faktory mají na výskyt zkoumaných rostlin největší vliv.

	1	2	3	4	5	6	7	8	Součet
bio12	0,892402826	0,044783271	0,017765870	0,008328012	0,000192620	0,000593288	0,019288750	0,003890748	0,987245385
bio01	0,950226174	0,007932419	0,002337718	0,003119587	0,007124236	0,003107338	0,012390530	0,000059992	0,986297994
bio11	0,925981757	0,015302385	0,000445382	0,031014541	0,006041612	0,001468849	0,005765109	0,000181314	0,986200949
bio16	0,631261888	0,246313824	0,013303330	0,040580053	0,002115623	0,007870168	0,036818720	0,002278907	0,980542513
bio19	0,622435555	0,254044381	0,011738270	0,041331304	0,002446024	0,009802879	0,036615750	0,001326085	0,979740248
bio05	0,507660629	0,313554748	0,036032540	0,012357140	0,044278100	0,022797909	0,039646370	0,002546175	0,978873610
bio06	0,914849703	0,000174918	0,002195950	0,045717084	0,000005535	0,000358865	0,015280690	0,000024247	0,978606992
pr	0,810220296	0,073176782	0,000000049	0,038285268	0,000038994	0,008868749	0,031797140	0,011536400	0,973923678
bio08	0,906297782	0,013532710	0,001447753	0,030834150	0,007532007	0,003729910	0,006345322	0,001554893	0,971274527
SRTM30	0,909323529	0,041735573	0,002714727	0,004805357	0,001413704	0,005550320	0,000006603	0,001936699	0,967486512
bio13	0,573823731	0,250491023	0,026000920	0,065974394	0,000496607	0,007521364	0,040495020	0,000610327	0,965413385
bio15	0,053050642	0,756488563	0,009682974	0,052890002	0,022810260	0,053263882	0,008711432	0,004614762	0,961512517
bio10	0,859443629	0,005484313	0,013097440	0,005317434	0,014872310	0,011167375	0,048630840	0,000136212	0,958149552
bio02	0,006472145	0,815448659	0,037930670	0,007134833	0,054389470	0,020975482	0,002636872	0,009733122	0,954721253

Obr. 19 Ukázka zastoupení faktorů v datech *Cicer echinospermum*.

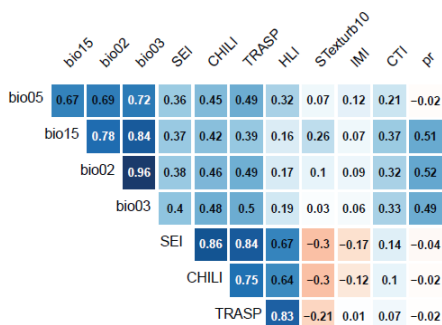
Některé z faktorů jsou si ovšem podobné a vzájemně spolu korelují, což může negativně ovlivnit následnou tvorbu shluků. Proto bylo třeba tyto korelace co nejvíce omezit odstraněním méně důležitých faktorů. Pro tyto účely byla vytvořena korelační matice, která znázorňuje vzájemnou korelaci mezi všemi 41 environmentálními faktory.

Pomocí následujícího příkazu se vytvoří korelační matice spearmanovou metodou a následně se vizualizuje (viz obr. 20).

```
cormatrix <- cor(data[, 9:49], method = "spearman")

mat <- as.matrix(data[, 9:49])
cor.mtest <- function(mat) {
  mat <- as.matrix(mat)
  n <- ncol(mat)
  p.mat <- matrix(NA, n, n)
  diag(p.mat) <- 0
  for (i in 1:(n - 1)) {
    for (j in (i + 1):n) {
      tmp <- cor.test(mat[, i], mat[, j], method = "spearman")
      p.mat[i, j] <- p.mat[j, i] <- tmp$p.value
    }
  }
  colnames(p.mat) <- rownames(p.mat) <- colnames(mat)
  p.mat
}
```

```
p.mat <- cor.mtest(mat)
corrplot(cormatrix,p.mat = p.mat,sig.level = 0.05,insig = "pch",pch.col =
"grey", method="color", type="upper", order = "hclust", tl.col="black",
tl.srt=45, diag=FALSE, addCoef.col = "black", outline = "white", tl.cex =
0.8)
```



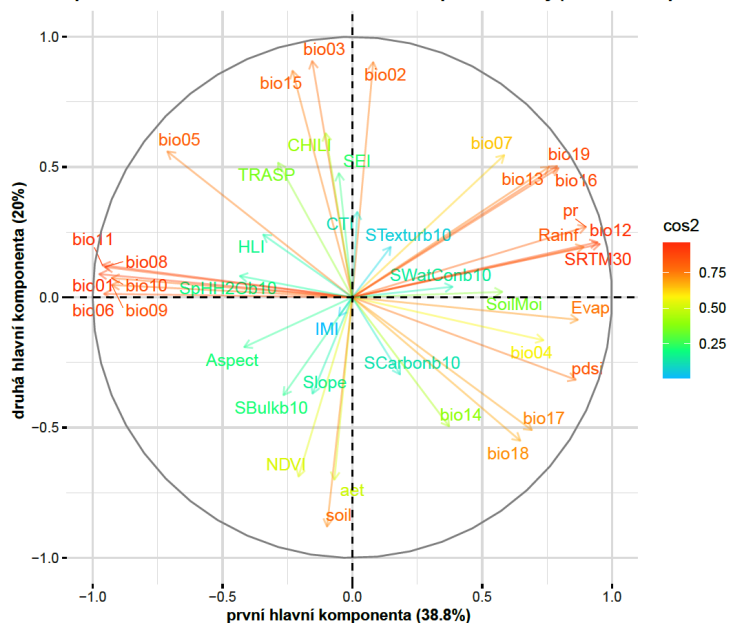
Obr. 20 Ukázka části korelační matice dat *Cicer echinospermum*.

Následně byly porovnávány ty environmentální faktory, které měly v korelační matici vzájemnou hodnotu korelace vyšší než 0,7, nebo nižší než -0,7. Tyto hodnoty dle de Vause (2002) značí velmi silnou pozitivní nebo negativní závislost. Porovnání faktorů proběhlo na základě hodnot zastoupení faktorů v dimenzích zjištěných v předchozím kroku, kdy byl ponechán faktor s vyšším součtem hodnot. Všechny faktory, které měly celkové zastoupení ve vybraných dimezních menší než 0,7, byly z dalších analýz také vyřazeny.

Posléze byl vytvořen graf komponentních vah pro data se všemi environmentálními faktory (viz obr. 21).

```
fviz_pca_var(data.pca, col.var= "cos2", gradient.cols =
c("#33DAFF", "#33FFB2", "#BEFF33", "#FFF933", "#FFA833", "#FF6833"), repel
= TRUE, alpha.var = 0.3)
```

Graf komponentních vah – 41 environmentálních faktorů pro data cizny (*Cicer echinospermum*)



Obr. 21 Ukázka grafu komponentních vah pro data *Cicer echinospermum*.

Tento postup byl proveden pro data cizrny (*Cicer echinospermum*), cizrny (*Cicer reticulatum*), čočky, čočky obsahující genové skupiny a hrachu. Graf komponentních vah byl navíc vytvořen i pro data s pouze vybranými environmentálními faktory, v tomto případě bylo třeba vždy znovu provést příkaz PCA a příkaz pro vytvoření grafu komponentních vah.

### 5.3.2 Proces shlukování

Po vybrání environmentálních faktorů se mohlo přistoupit ke shlukování. Pomocí metody k-means byly vytvářeny shluky rostlin na základě vybraných faktorů. Pro tyto účely musely být do softwaru RStudio navíc nainstalovány balíčky NbClust a writexl.

Nejprve byla tabulka s daty vložena do proměnné data.

```
data <- CICER_echinospermum_presnost1_15_faktoru
```

Dále bylo nutné data standardizovat, aby nedošlo ke zkreslení výsledků. Pomocí `data[,9:23]` jsou vybrány atributy, které mají projít standardizací a budou vstupovat do následného shlukování.

```
data_standard <- scale(data[,9:23])
```

Následně se pomocí nainstalovaného balíčku NbClust zjistil ideální počet shluků. Tento balíček slouží k zobrazení vhodného počtu shluků pro různé shlukové metody pomocí 30 různých indexů. Do příkazu vstupují standardizovaná data z předchozího kroku. Bylo nastaveno, že se má hledat optimální počet shluků mezi třemi a deseti shluky. Výsledkem tohoto příkazu je přehled, který udává kolik indexů doporučuje určitý počet shluků (viz obr. 22).

```
pocetshluku <- NbClust(data = data_standard, distance = "euclidean",  
min.nc = 3, max.nc = 10, method = "kmeans", index = "all", alphaBeale =  
0.1)
```

```
Among all indices:  
* 6 proposed 3 as the best number of clusters  
* 2 proposed 4 as the best number of clusters  
* 4 proposed 5 as the best number of clusters  
* 1 proposed 7 as the best number of clusters  
* 1 proposed 8 as the best number of clusters  
* 2 proposed 9 as the best number of clusters  
* 8 proposed 10 as the best number of clusters
```

```
***** Conclusion *****
```

```
* According to the majority rule, the best number of clusters is 10
```

Obr. 22 Vyhodnocení ideálního počtu shluků pro data *Cicer echinospermum*.

Následoval mezikrok, který zajistí, že náhodná čísla vytvářená při metodě k-means budou vždy stejná.

```
set.seed(1234)
```

Poté byly vytvořeny shluky bodů rostlin. Na základě příkazu NbClust byl vždy vytvořen počet shluků pro tři nejlepší výsledky, tzn. pro každou rostlinu bylo

shlukování provedeno třikrát pro jiný počet shluků. Pro výpočet bylo vždy zvoleno 200 iterací. Příkaz byl zrovna vypočítán do nové proměnné.

```
kmeansshluky3 <- kmeans(data_standard, 3, iter.max = 200, nstart = 50)
kmeansshluky5 <- kmeans(data_standard, 5, iter.max = 200, nstart = 50)
kmeansshluky10 <- kmeans(data_standard, 10, iter.max = 200, nstart = 50)
```

Dalším krokem bylo vytvoření nového atributu „shluk(počet shluků)“ do kopie tabulky původních dat s názvem „export“. Do tohoto atributu se vložily údaje o zařazení do shluku z atributu cluster proměnné „kmeansshluky(počet shluků)“. Aby bylo možné vytvořit biplot a graf paralelních souřadnic s nově vytvořenými shluky, bylo nutné číselné hodnoty nově vytvořeného atributu převést na faktor.

```
export <- data
export$shluk3 <- kmeansshluky3$cluster
export$shluk3 <- as.factor(export$shluk3)
export$shluk5 <- kmeansshluky5$cluster
export$shluk5 <- as.factor(export$shluk5)
export$shluk10 <- kmeansshluky10$cluster
export$shluk10 <- as.factor(export$shluk10)
```

Tabulka export byla pomocí balíčku writexl exportována do softwaru Excel, se kterou se mohlo dále pracovat.

```
write_xlsx(export, path = "CICER_reti_shluky.xlsx", col_names = TRUE,
format_headers = TRUE)
```

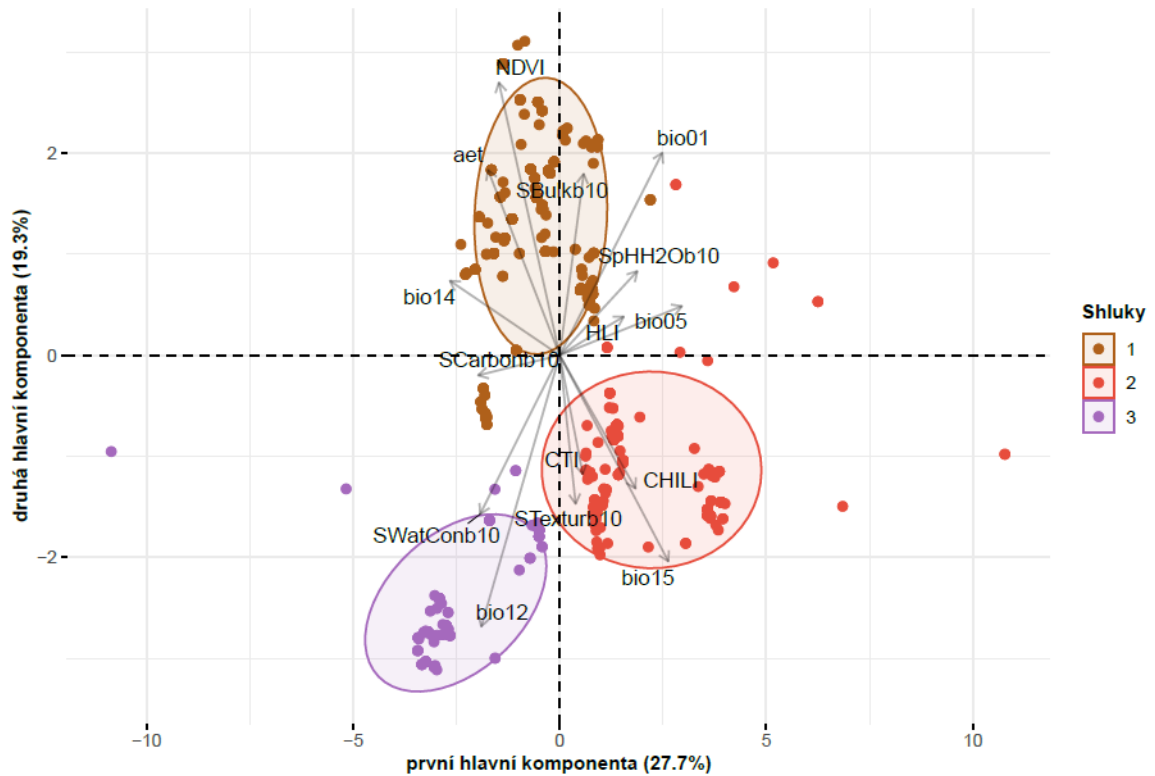
Následovala opět analýza hlavních komponent pro vizualizaci shluků pomocí biplotu. Do analýzy opět vstupují původní data s vybranými faktory.

```
data.pca1 <- PCA(data[,9:23], scale.unit = TRUE, graph = TRUE)
```

Výsledné shluky se vizualizovaly pomocí biplotu, kde je možné posoudit jejich závislost na vybraných environmentálních faktorech ve dvou prvních dimenzích (viz obr. 23).

```
fviz_pca_biplot(data.pca1,
  pointshape = 21, pointsize = 2,
  fill.ind = export$shluk10,
  col.ind = export$shluk10,
  palette = c("#AF601A", "#E74C3C", "#A569BD", "#3498DB",
    "#2ECC71", "#F9E79F", "#F39C12", "#FA05C3", "#99A3A4", "#117864"),
  addEllipses = TRUE,
  ellipse.level = 0.75, # velikost elips
  mean.point = FALSE, # odstranění středového bodu elips
  label = "var",
  col.var = "black",
  alpha.var = 0.3, # průhlednost sipek
  repel = TRUE, # zamezení překryvu popisku
  legend.title = "shluky")
```

Biplot – shluky podle k-means na základě 15 environmentálních faktorů pro data cizrny (*Cicer echinospermum*)

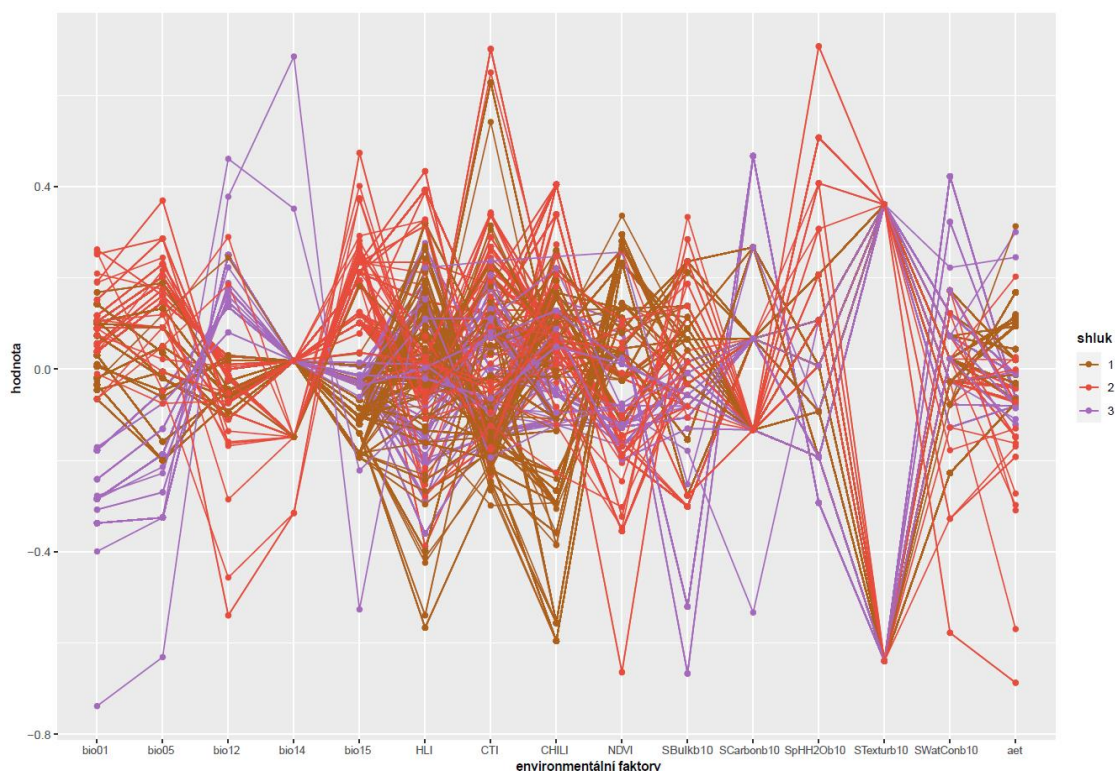


Obr. 23 Ukázka biplotu pro data *Cicer echinospermum*.

Další možnost, jak vizualizovat vliv jednotlivých environmentálních faktorů na vytvořené shluky, je graf paralelních souřadnic (viz obr. 24). Do příkazu opět vstupuje tabulka export, jejíž hodnoty atributů s vytvořenými shluky jsou převedeny na faktor. V části *columns* jsou zvoleny atributy, které mají být v grafu znázorněny, v tomto případě environmentální faktory, *groupColumn* se odkazuje na atribut obsahující přiřazení bodů ke shlukům. Pomocí *scale="center"* je možné data standardizovat tak, aby se hodnoty pohybovaly od -1 do 1.

```
ggparcoord(export, columns = 9:22, groupColumn = 23,
  showPoints = TRUE,
  scale="center", alphaLines = 1,
  title = "Parallel Coordinate Plot for the Iris Data")+
  scale_color_manual(values=c("#AF601A", "#E74C3C", "#A569BD", "#3498DB"))
```

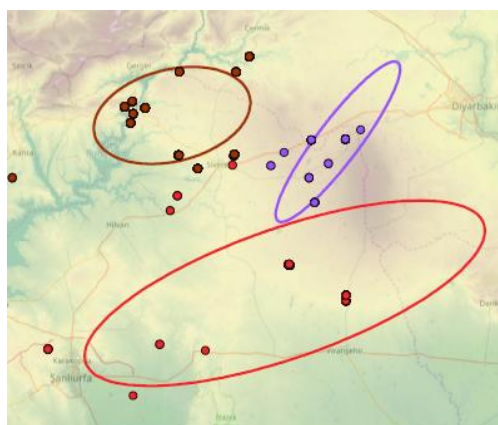




Obr. 24 Ukázka grafu paralelních souřadnic pro data *Cicer echinospermum*.

Tento postup byl proveden pro všechny zkoumané druhy rostlin. U dat čocky, která navíc obsahuje informaci o genetických skupinách, byl biplot vytvořen na základě genetických skupin.

Díky možnosti exportování tabulky s výsledkem shlukování z RStudia do Excelu bylo možné informaci o zařazení bodu do shluku převést do vektorové vrstvy v softwaru ArcGIS Desktop. K tomu posloužily nástroje *Excel to Table* a *Make XY Event Layer*. Dále byly pomocí nástroje *Directional Distribution (Standard Deviational Ellipse)* nad těmito shluky vytvořeny elipsy pro lepší vizualizaci. Elipsy představují směrovou distribuci, pomocí které lze vyhodnotit prostorové charakteristiky geografických prvků. Shluky byly i s elipsami následně obarveny stejně jako v biplotu, aby se daly lépe porovnat (viz obr. 25).



Obr. 25 Ukázka shluků dat *Cicer echinospermum* v mapě.



### 5.3.3 Porovnání přesnosti

Poslední analýzou prováděnou v této práci je zjištění vlivu přesnosti bodových dat rostlin na výsledné analýzy. Polohová přesnost bodů má různý vliv v závislosti na prostorovém rozlišení environmentálních dat, například datasety s rozlišením desítek kilometrů by neměly vyžadovat bodová data s přesností v řádu jednotek či desítek metrů. Porovnání bylo provedeno nad environmentálními proměnnými z WorldClim v rozlišení 1 km a nad výškovými daty SRTM30 a indexy z něj odvozenými v rozlišení 30 m.

Pro tyto účely bylo vybráno 455 bodů čičky s přesností 1 nacházející se v UTM pásmu 37. Tyto body byly porovnány s náhodně vytvořenými body s přesností 3. Okolo bodů s přesností 1 byl vytvořen buffer o poloměru 1,11 km, což odpovídá chybovosti bodů s přesností 3. Okolo těchto bufferů byly nástrojem *Minimum Bounding Geometry* vytvořeny ohraničující čtverce, v jejichž oblasti byly následně nástrojem *Create Random Points* náhodně vytvořeny nové body, které simulují náhodnost polohové přesnosti 3. Těmto nově vytvořeným bodům byly přiřazeny hodnoty z rastrů bioklimatických proměnných, SRTM30 a indexů z něj odvozených pomocí nástroje *Extract Multi Values to Points*. Atributové tabulky přesných i nepřesných bodů byly exportovány do softwaru Excel, kde byly následně vypočítány procentuální rozdíly mezi přesnými a nepřesnými daty v aritmetickém průměru, mediánu a rozpětí hodnot. Tabulky byly také vloženy do softwaru RStudio, kde byla provedena analýza hlavních komponent a byly vytvořeny grafy komponentních vah pro obě přesnosti.

## 6 INTERPRETACE STATISTICKÝCH ANALÝZ

Posledním krokem praktické části této práce je interpretace statistických analýz. Tato kapitola obsahuje výsledky a výstupy dosažené postupem popsáním v kapitole 5 spolu s jejich interpretací. Jsou zde popsány výsledky výběru environmentálních faktorů, shlukování a zjištění vlivu přesnosti bodových dat, a to pro každý druh zkoumaných rostlin zvlášť.

### 6.1 Výběr environmentálních faktorů

Jedním z hlavních cílů práce je zjištění vlivu environmentálních faktorů na výskyt zkoumaných rostlin. Pro zkoumání vlivu na výskyt rostlin bylo vybráno 41 environmentálních faktorů na základě konzultace s odborníky z Katedry botaniky a s vedoucím práce. Z těchto faktorů byly vybírány pouze ty, které jsou na sobě nezávislé a mají v datech rostlin největší zastoupení. Tento výběr byl proveden kvůli zamezení zkreslení shlukových analýz a také pro další výzkum pracovníků Katedry botaniky. Proces výběru faktorů je popsán v kapitole 5.3.1 Zjištění důležitých environmentálních faktorů a jeho výsledky jsou níže popsány pro každou rostlinu zvlášť.

#### 6.1.1 Cizrna (*Cicer reticulatum*)

Nejprve byl určen počet dimenzí (hlavních komponent), které jsou důležité pro vysvětlení největšího množství variability dat. Pomocí scree plotu bylo na základě Kaiserova kritéria, které říká, že důležité jsou ty komponenty, které mají vlastní číslo větší než 1, vybráno 8 hlavních komponent. V případě *Cicer reticulatum* je v 8 hlavních komponentách zahrnuto 86,6 % dat.

Na základě výpisu zastoupení environmentálních faktorů v osmi dimenzích bylo zjištěno, že na *Cicer reticulatum* mají největší vliv faktory týkající se teploty a srážek, menší zastoupení mají indexy vycházející z výškového modelu a nejméně jsou zastoupeny faktory, které se týkají vlastností půdy. Pro odstranění faktorů, které spolu vysoce korelují byla využita korelační matice (viz příloha 3). Vždy byly porovnávány faktory, které vykazovaly korelaci vyšší než 0,7 nebo nižší než -0,7. Vysokou pozitivní korelaci ve většině případů vykazují faktory týkající se stejného tématu, např. faktory s teplotou, faktory se srážkami apod. Vysokou negativní korelaci vykazuje nadmořská výška s teplotními faktory. Naopak nízkou nebo téměř žádnou korelaci s ostatními faktory vykazují vlastnosti půdy a indexy odvozené z výškového modelu. Díky tomuto porovnání bylo nakonec vybráno 15 environmentálních faktorů, které se vzájemně významně neovlivňují a zároveň mají v datech co největší zastoupení (viz tabulka 9).

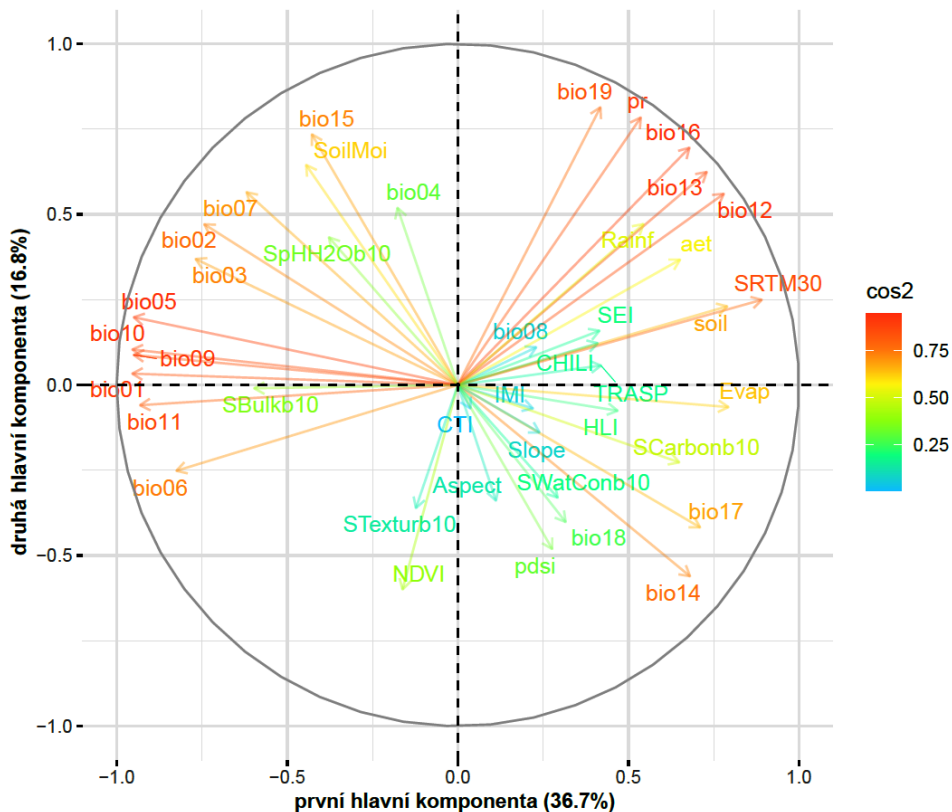
Podobnost faktorů lze vyvodit také z vytvořeného grafu komponentních vah (viz obr. 26). Z grafu je patrné, že největší vliv v prvních dvou komponentách mají bioklimatické proměnné z datové sady WorldClim, jelikož se nejvíce blíží jednotkové kružnici. Naopak malé zastoupení mají indexy odvozené z výškového modelu a faktory týkající se vlastností půdy. Zároveň lze pozorovat shluky bioklimatických proměnných týkající se teploty na jedné straně a shluky srážkových faktorů na druhé straně. Je ovšem nutné si uvědomit, že graf komponentních vah je zobrazením pouze prvních dvou dimenzí.

Pro porovnání byl vytvořen také graf komponentních vah s vybranými faktory (viz obr. 27). Směr polopřímek faktorů se od prvního grafu příliš nezměnil, ovšem jejich zastoupení v prvních dvou dimenzích je rozdílné. To je způsobeno tím, že do analýzy hlavních komponent vstupuje méně faktorů, které spolu výrazně nekorelují.

Tabulka 9 Vybrané environmentální faktory pro *Cicer reticulatum*

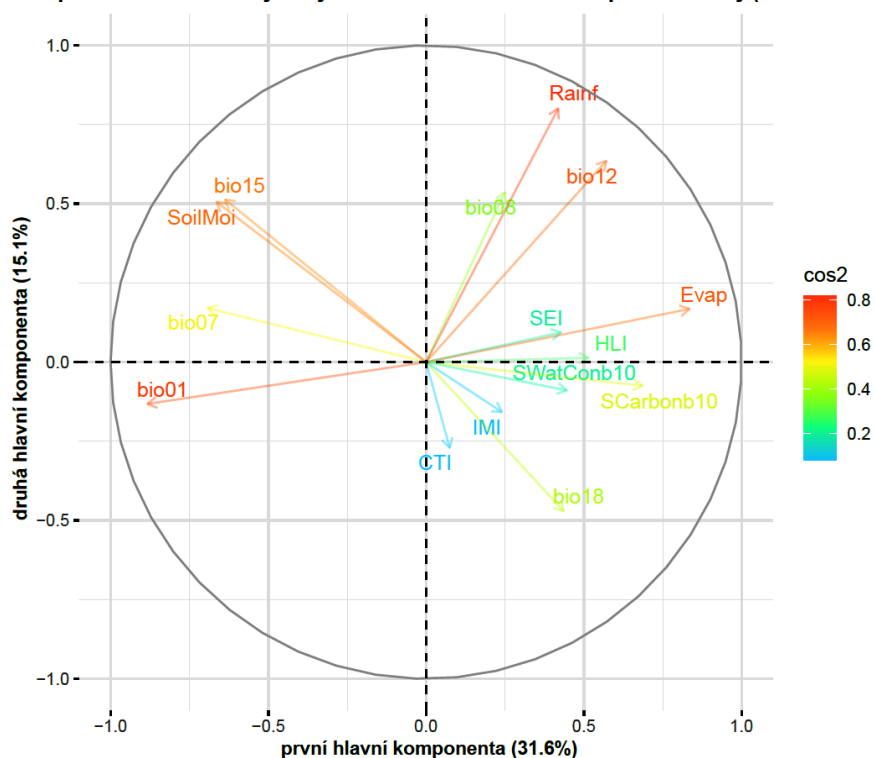
bio01	Průměrná roční teplota
bio07	Roční rozsah teplot (BIO5-BIO6)
bio08	Průměrná teplota nejvlhčího čtvrtletí
bio12	Roční úhrn srážek
bio15	Sezónnost srážek (variační koeficient)
bio18	Srážkový úhrn v nejteplejším čtvrtletí
SEI	Site Exposure Index
IMI	Integrated Moisture Index
CTI	Compound Topographic Index
HLI	Heat Load Index
Evap	Evapotranspirace
Rainf	Celkový úhrn srážek
SoilMoi	Vlhkost půdy v hloubce 10–40 cm
SCarbonb10	Obsah organického uhlíku v půdě v hloubce 10 cm
SWatConb10	Obsah půdní vody v 33 kPa v hloubce 10 cm

Graf komponentních vah – 41 environmentálních faktorů pro data cizmy (*Cicer reticulatum*)



Obr. 26 Graf komponentních vah všech faktorů pro data cizry (*Cicer reticulatum*).

Graf komponentních vah – 15 vybraných environmentálních faktorů pro data cizrny (*Cicer reticulatum*)



Obr. 27 Graf komponentních vah vybraných faktorů pro data cizrny (*Cicer reticulatum*).

### 6.1.2 Cizrna (*Cicer echinospermum*)

Pomocí Kaiserova kritéria bylo zjištěno, že by mělo být využito 9 hlavních komponent. Vzhledem k tomu, že při vysokém počtu komponent nelze považovat analýzu hlavních komponent za příliš úspěšnou, a protože je v osmi hlavních komponentách obsaženo 87,9 % dat, což je více než v případě osmi komponent *Cicer reticulatum*, bylo nakonec opět zvoleno 8 hlavních komponent.

Největší zastoupení v těchto osmi dimenzích mají podobně jako u předchozího druhu cizrny faktory týkající se teploty a srážek a nejmenší zastoupení mají faktory s vlastnostmi půdy. Z korelační matice (viz příloha 4) je opět zřejmá korelace mezi teplotními faktory navzájem a srážkovými faktory navzájem. U druhu *echinospermum* se více projevila negativní korelace mezi teplotami s vlhkostí půdy a teplotami s indexem sucha. Při porovnání korelační matice a zastoupení faktorů v osmi dimenzích bylo stejně jako u druhu *reticulatum* vybráno 15 environmentálních faktorů (viz tabulka 10), ačkoli se od faktorů *reticulatu* liší.

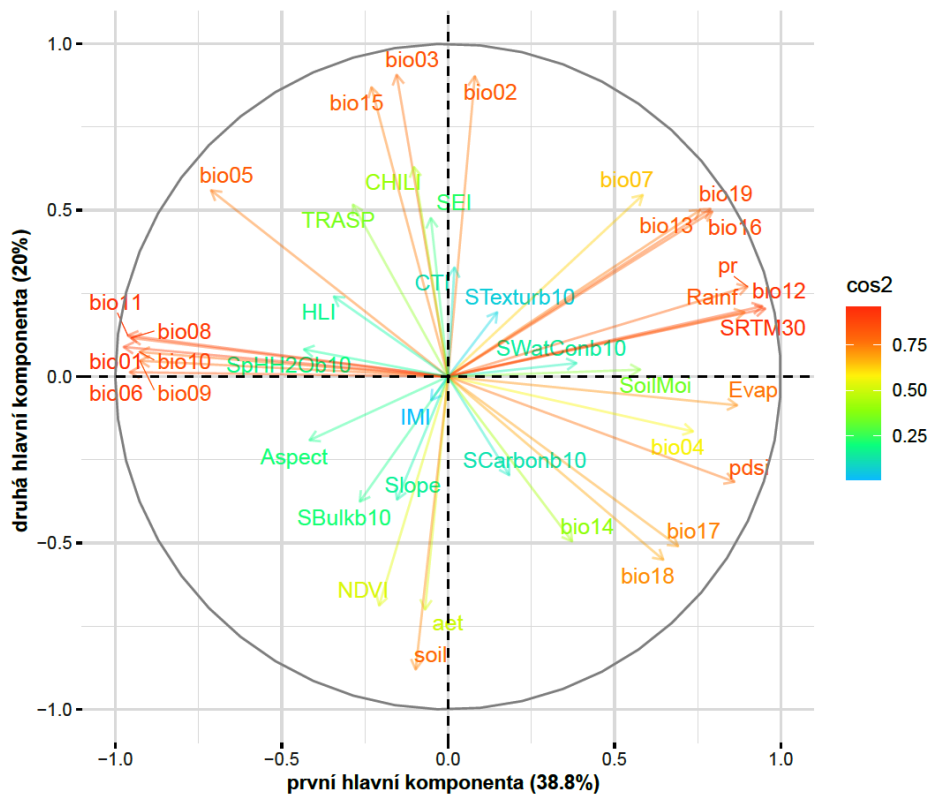
Tabulka 10 Vybrané environmentální faktory pro *Cicer echinospermum*

bio01	Průměrná roční teplota
bio05	Maximální teplota nejteplejšího měsíce
bio12	Roční úhrn srážek
bio14	Srážkový úhrn v nejsušším měsíci
bio15	Sezónnost srážek (variační koeficient)
CTI	Compound Topographic Index
CHILI	Index CHILI
HLI	Heat Load Index
aet	Skutečná evapotranspirace
NDVI	Normalized Difference Vegetation Index
SBulkb10	Hustota půdy v hloubce 10 cm
SCarbonb10	Obsah organického uhlíku v půdě v hloubce 10 cm
SWatConb10	Obsah půdní vody v 33 kPa v hloubce 10 cm
SpHH2Ob10	pH půdy v H <sub>2</sub> O v hloubce 10 cm
STexturb10	Typ textury půdy (USDA) v hloubce 10 cm

Z grafu komponentních vah pro 41 faktorů lze vyčíst, že největší zastoupení v prvních dvou komponentách mají opět bioklimatické proměnné z WorldClim, ale také index sucha, vlhkost půdy a evapotranspirace (viz obr. 28). Nízké zastoupení mají indexy odvozené z výškového modelu a ostatní faktory týkající se vlastností půdy. Z grafu je patrný výrazný shluk teplotních faktorů u první hlavní komponenty.

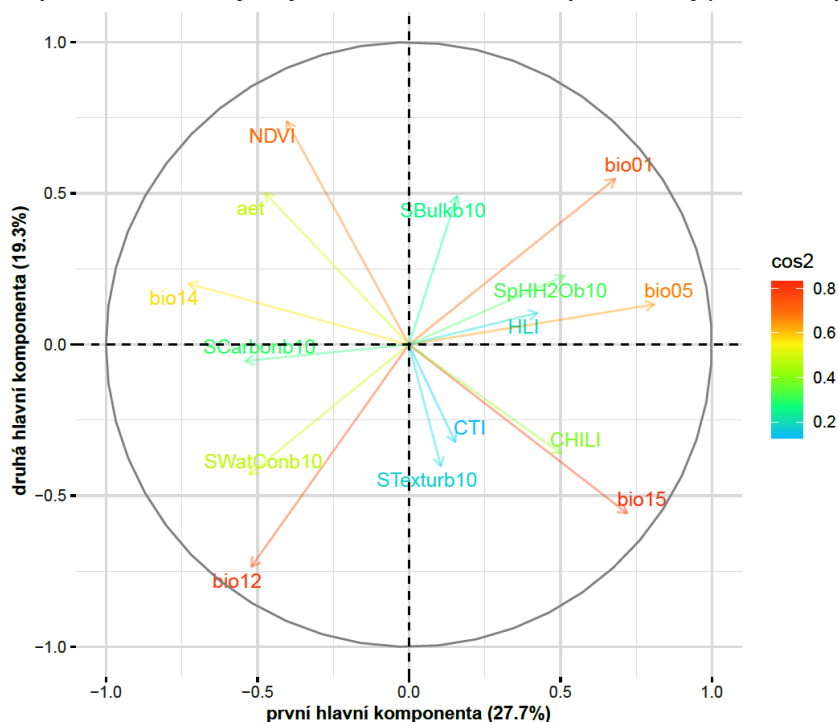
Graf komponentních vah s 15 vybranými faktory je v případě *echinospermu* odlišný od grafu se všemi faktory (viz obr. 29). Směr průvodičů faktorů je ve většině případů velice odlišný a lze pozorovat pootočení průvodičů.

Graf komponentních vah – 41 environmentálních faktorů pro data cizrny (*Cicer echinospermum*)



Obr. 28 Graf komponentních vah všech faktorů pro data cizrny (*Cicer echinospermum*).

Graf komponentních vah – 15 vybraných environmentálních faktorů pro data cizrny (*Cicer echinospermum*)



Obr. 29 Graf komponentních vah vybraných faktorů pro data cizrny (*Cicer echinospermum*).

### 6.1.3 Čočka (*Lens orientalis*)

U čočky bylo pomocí scree plotu a Kaiserova kritéria opět vybráno 8 hlavních komponent. Tyto komponenty dohromady obsahují 83,6 % dat.

V případě čočky mají největší zastoupení v prvních osmi komponentách teplotní faktory, následují faktory srážkové, poté indexy vypočítané z výškového modelu a nejmenší zastoupení mají vlastnosti půdy. Korelační matice (viz příloha 5) vykazuje podobné hodnoty jako v předchozích případech. Vysoce spolu korelují faktory, které se týkají srážek, evapotranspirace a vlhkosti půdy. K významné korelaci dochází i u teplotních faktorů. Negativní korelaci vykazuje sezónnost srážek se srážkovým úhrnem v nejteplejších a nejsušších částech roku. Na základě korelační matice a zastoupení faktorů v komponentách bylo pro čočku vybráno 14 environmentálních faktorů (viz tabulka 11).

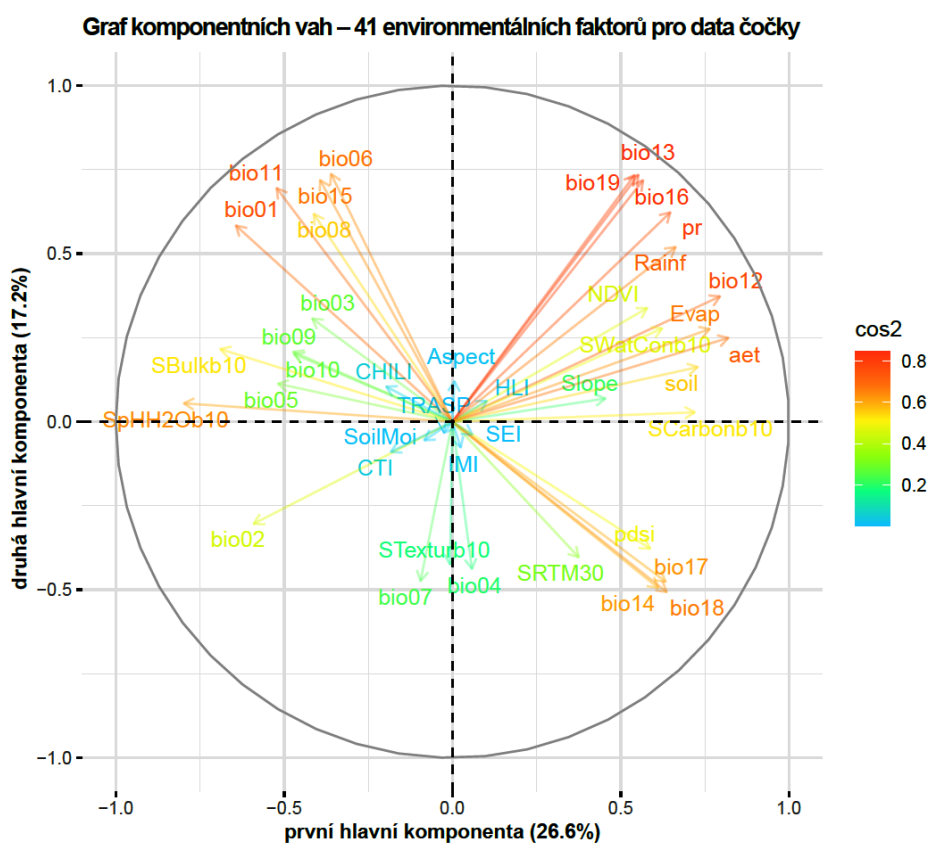
Tabulka 11 Vybrané environmentální faktory pro čočku

bio01	Průměrná roční teplota
bio02	Průměrný denní rozsah teplot
bio07	Roční rozsah teplot (BIO5-BIO6)
bio10	Průměrná teplota nejteplejšího čtvrtletí
bio12	Roční úhrn srážek
bio18	Srážkový úhrn v nejteplejším čtvrtletí
HLI	Heat Load Index
SEI	Site Exposure Index
CTI	Compound Topographic Index
IMI	Integrated Moisture Index
CHILI	Index CHILI
pdsi	Palmerův index intenzity sucha
NDVI	Normalized Difference Vegetation Index
SpHH2Ob10	pH půdy v H <sub>2</sub> O v hloubce 10 cm

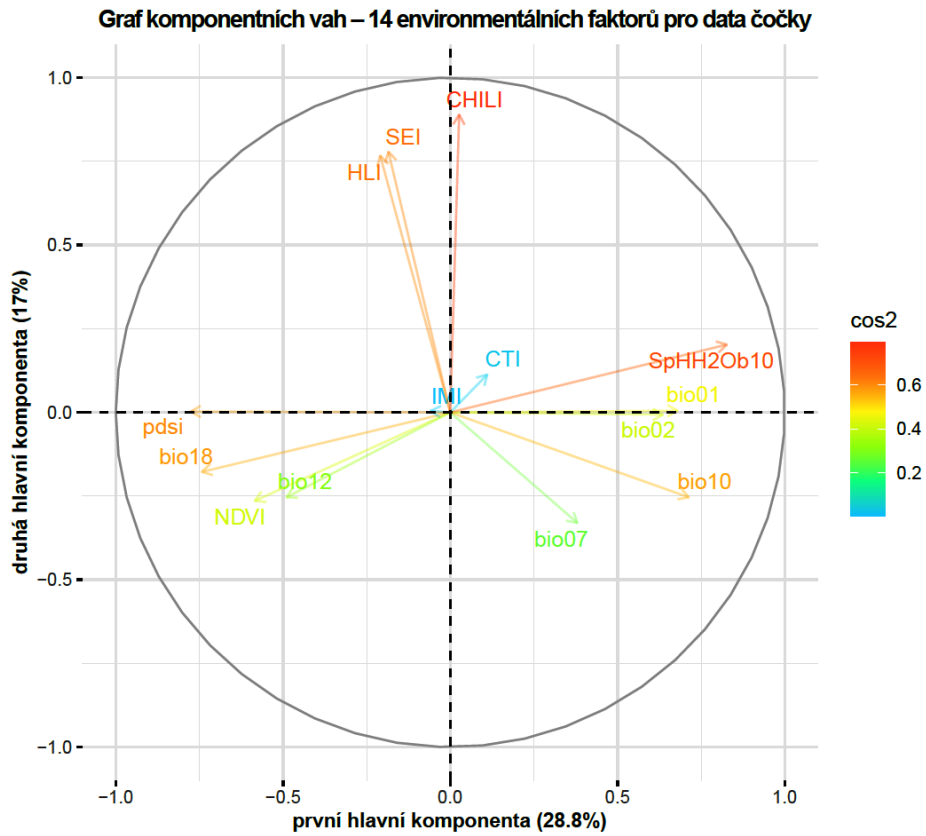
Z grafu komponentních vah čocky se všemi zkoumanými environmentálními faktory lze vyčíst, že některé bioklimatické faktory zůstaly upozaděny, naopak se více projevují vlastnosti půdy (viz obr. 30). Z grafu je patrná velmi nízká korelace mezi částí teplotních a srážkových faktorů. Srážkové faktory poměrně výrazně ovlivňují obě hlavní komponenty, naopak většina teplotních faktorů tak výrazný vliv na komponenty nemá. Nejméně jsou komponenty ovlivněny indexy vypočítanými z výškového modelu.

Z grafu komponentních vah pro 14 vybraných faktorů je patrné rozdělení těchto faktorů do tří shluků (viz obr. 31). První shluk obsahuje teplotní faktory, obsah pH vody v půdě a negativně koreluje s druhým shlukem obsahujícím srážkové faktory, NDVI a dva indexy. Nízkou korelaci k těmto dvěma shlukům má třetí shluk obsahující tři indexy vypočítané z výškového modelu.

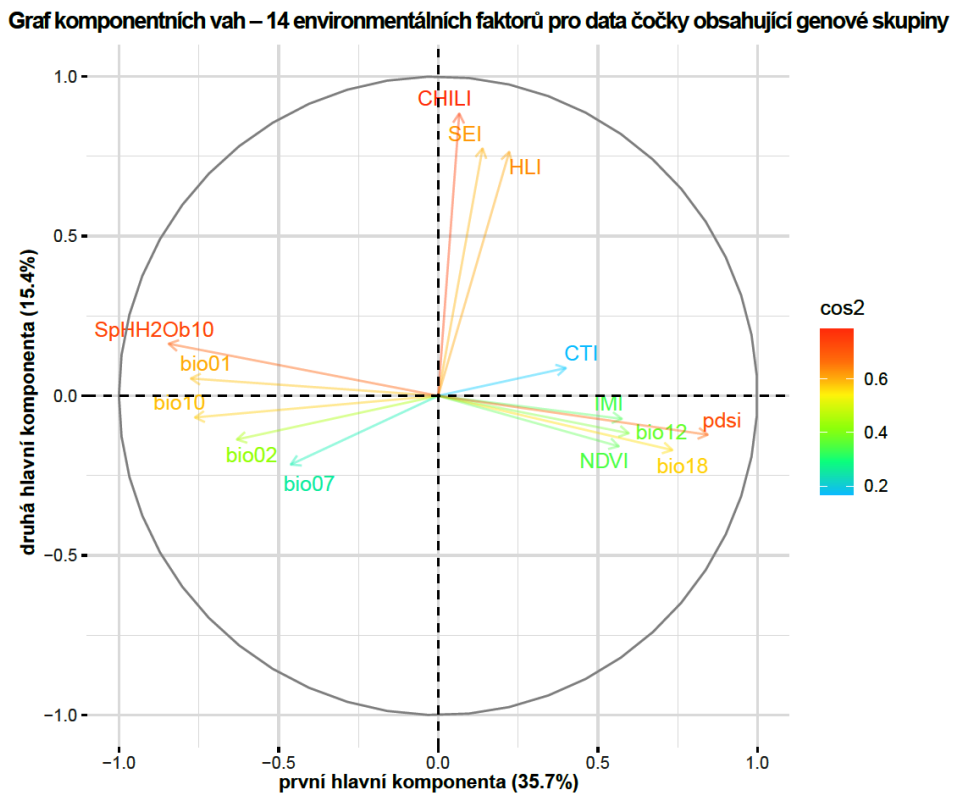
Kromě grafů se všemi faktory a se 14 faktory pro všechna data čocky byl vytvořen ještě graf komponentních vah se 14 vybranými faktory pouze pro data čocky obsahující informaci o genetické skupině (viz obr. 32). Tento graf je podobný grafu s vybranými faktory pro všechna data čocky, ovšem zmíněné shluky jsou mnohem výraznější.



Obr. 30 Graf komponentních vah všech faktorů pro data čocky.



Obr. 31 Graf komponentních vah vybraných faktorů pro data čocky.



Obr. 32 Graf komponentních vah vybraných faktorů pro data čocky obsahující genové skupiny.



### 6.1.4 Hrách (*Pisum elatius*)

V případě hrachu Kaiserovo kritérium ve scree plotu doporučilo využít 10 hlavních komponent, ovšem stejně jako u předchozích druhů bylo vybráno 8 hlavních komponent, jelikož tyto komponenty obsahují celkem 80 % dat, což je dostatečné zastoupení.

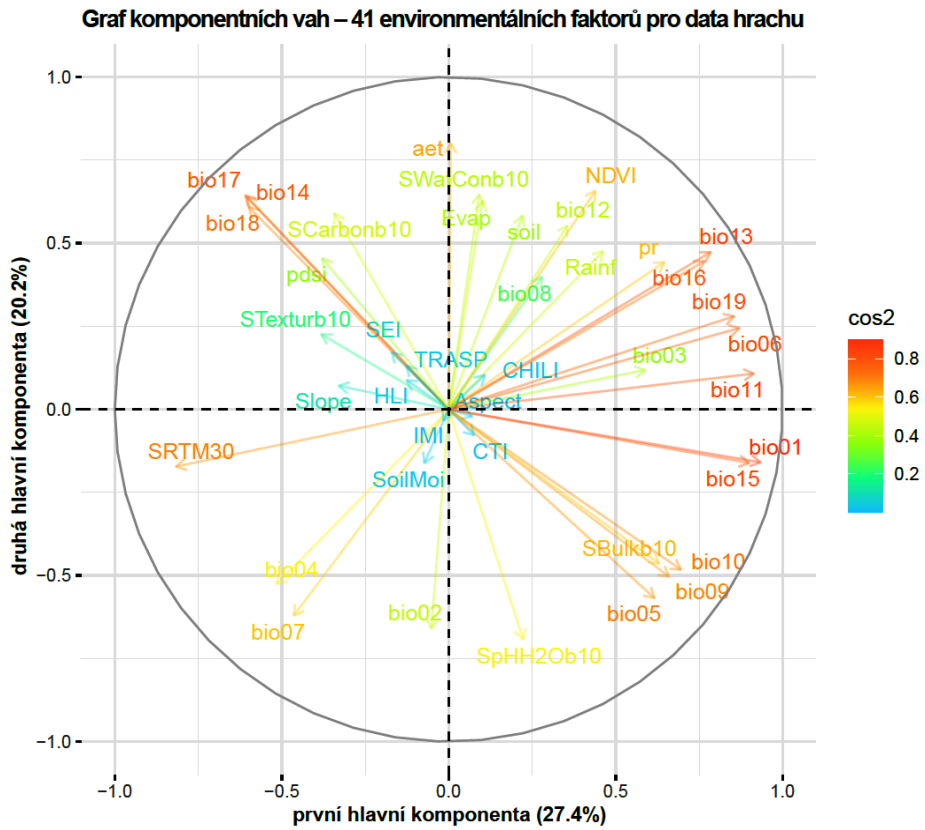
U hrachu mají v prvních osmi komponentách největší zastoupení bioklimatické proměnné z WorldClim, větší význam mají teplotní faktory. Poměrně významný vliv má také SRTM30. Nejnížší zastoupení mají vlastnosti půdy a index IMI, které byly kvůli nízkému vlivu přímo vyřazeny bez ohledu na korelaci s ostatními faktory. Korelační matice 41 environmentálních faktorů (viz příloha 6) pro data hrachu se od předchozích příliš neliší a stále se opakuje stejný vzorec podobnosti teplotních faktorů navzájem a srážkových faktorů navzájem. Celkově lze ovšem pozorovat, že ke korelaci vyšší než 0,7 nedochází v tak časté míře jako u předchozích druhů rostlin. Po porovnání korelací s mírou zastoupení faktorů v osmi hlavních komponentách bylo vybráno 13 environmentálních faktorů (viz tabulka 12). Toto číslo je ze všech 4 druhů rostlin nejnížší, což je navzdory ne příliš časté korelaci způsobeno právě vyřazením faktorů s nízkým zastoupením v komponentách.

Tabulka 12 Vybrané environmentální faktory pro hrách

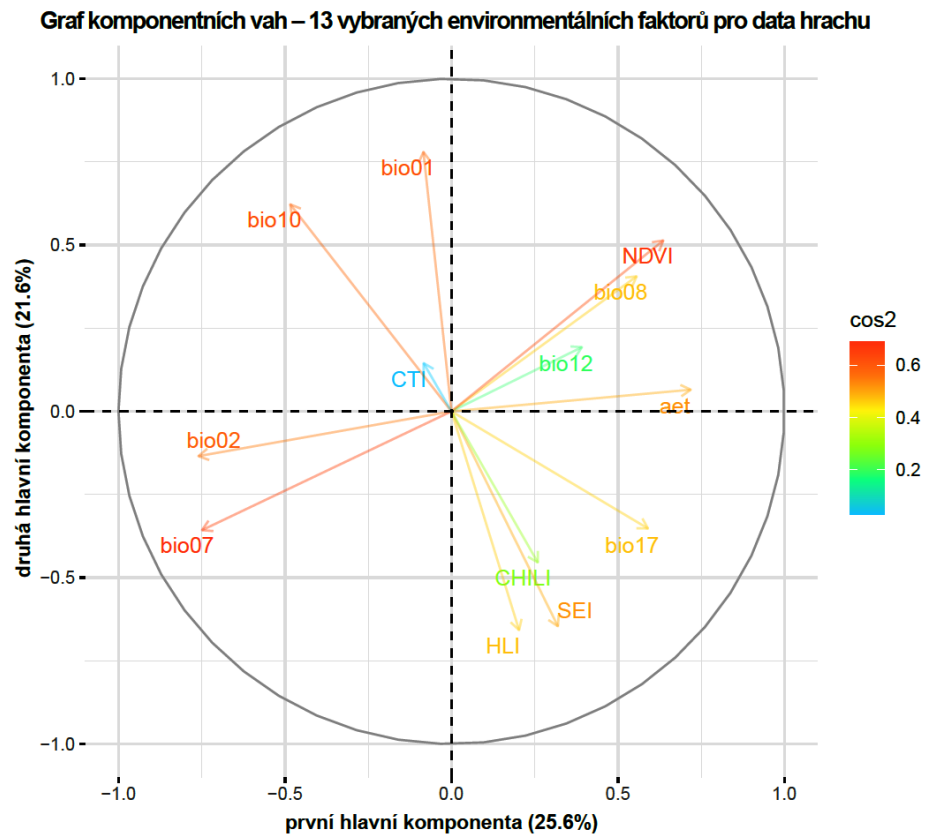
bio01	Průměrná roční teplota
bio02	Průměrný denní rozsah teplot
bio07	Roční rozsah teplot (BIO5-BIO6)
bio08	Průměrná teplota nejvlhčího čtvrtletí
bio10	Průměrná teplota nejteplejšího čtvrtletí
bio12	Roční úhrn srážek
bio17	Srážkový úhrn v nejsušším čtvrtletí
CTI	Compound Topographic Index
SEI	Site Exposure Index
HLI	Heat Load Index
NDVI	Normalized Difference Vegetation Index
aet	Skutečná evapotranspirace
CHILI	Index CHILI

Z grafu komponentních vah lze opět vyčíst velice nízký vliv indexů odvozených z výškového modelu na první dvě komponenty (viz obr. 33). Faktory mají mezi komponentami poměrně rovnoměrné zastoupení, nedochází k významnému shlukování faktorů k některé z komponent. Velký vliv má opět většina bioklimatických proměnných z WorldClim a následují je vlastnosti půdy.

V grafu komponentních vah se 13 vybranými faktory bylo rovnoměrné rozložení faktorů zachováno a nedochází zde k výrazným shlukům (viz obr. 34). Více podobné jsou si pouze indexy CHILI, HLI a SEI a také NDVI, bio08 a bio12.



Obr. 33 Graf komponentních vah všech faktorů pro data hrachu.



Obr. 34 Graf komponentních vah vybraných faktorů pro data hrachu.

## 6.2 Shlukování

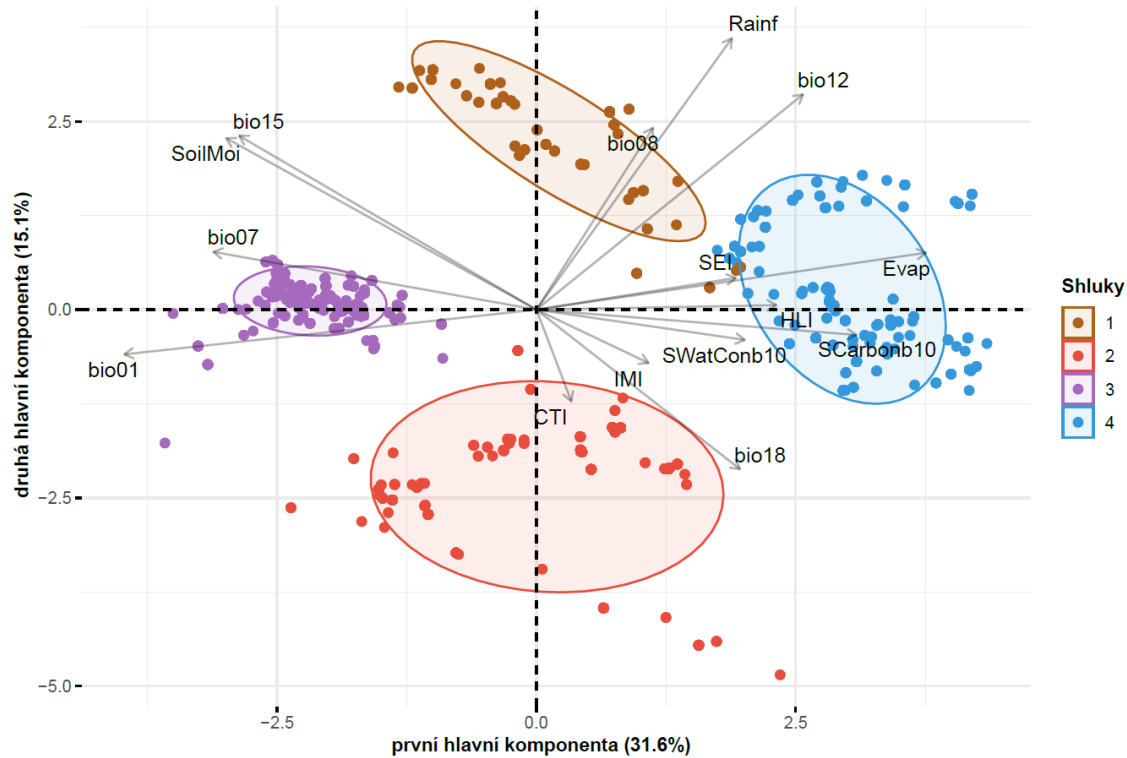
Pro zjištění podobnosti výskytu jednotlivců rostlin na základě environmentálních faktorů byla provedena shluková analýza. Shluky byly nejprve vytvořeny v softwaru RStudio, aby bylo možné zjistit, jak jsou ovlivněny environmentálními faktory. Následně byly shluky rostlin převedeny také do softwaru ArcGIS Desktop pro konkretizaci z hlediska jejich geografického rozmístění. Pro část dat čočky, která obsahovala informaci o genetických skupinách, byl navíc zjišťován vliv environmentálních faktorů na tyto skupiny. V softwaru RStudio byl na základě 30 různých indexů zjištěn ideální počet shluků pro každý druh zkoumaných rostlin. Vždy byly vybrány tři nejlepší možnosti, které byly následně konzultovány s doc. Smýkalem. Na základě této konzultace byla pro každý druh rostliny vybrána pouze jedna varianta počtu shluků, které se mají vytvořit. Níže jsou popsány výsledky shlukové analýzy pro každý druh zkoumaných rostlin zvlášť.

### 6.2.1 Cizrna (*Cicer reticulatum*)

Statistická analýza provedená pro zjištění optimálního počtu shluků doporučila využít počtu 9, 4 nebo 3 shluků, z nichž byl pro následné analýzy vybrán počet 4 shluků. Vytvořené shluky bodů rostlin a jejich závislost na environmentálních faktorech byly vizualizovány pomocí biplotu (viz obr. 35). Z biplotu je patrné, že se shluky mezi sebou poměrně výrazně vyčlenily a každý závisí na jiných environmentálních faktorech. Závislost bodů na faktorech lze odvodit z blízkosti faktorů k danému shluku. Shluk 1 výrazněji přispívá do druhé hlavní komponenty a faktory, na nichž nejvíce závisí, jsou bio08 (průměrná teplota nejvlhčího čtvrtletí), Rainf (celkový úhrn srážek), bio12 (roční úhrn srážek) a index SEI. Shluk 2 je nevíce ovlivňován indexy CTI, IMI a bio18 (srážkový úhrn v nejteplejším měsíci). Shluk 3 má nejhustější uspořádání ze všech čtyř vytvořených shluků. Rozléhá se těsně kolem první hlavní komponenty a je nejvíce ovlivňován bioklimatickými proměnnými bio01 (průměrná roční teplota) a bio07 (roční rozsah teplot). Z části je ovlivňován také faktorem SoilMoi (vlhkost půdy) a bio15 (sezónnost srážek). Shluk 4 leží na opačné straně první hlavní komponenty a nejvíce na něj působí faktory Evap (evapotranspirace), SCarbonb10 (obsah organického uhlíku v půdě), SWatConb10 (obsah půdní vody v 33 kPa) a indexy SEI a HLI.

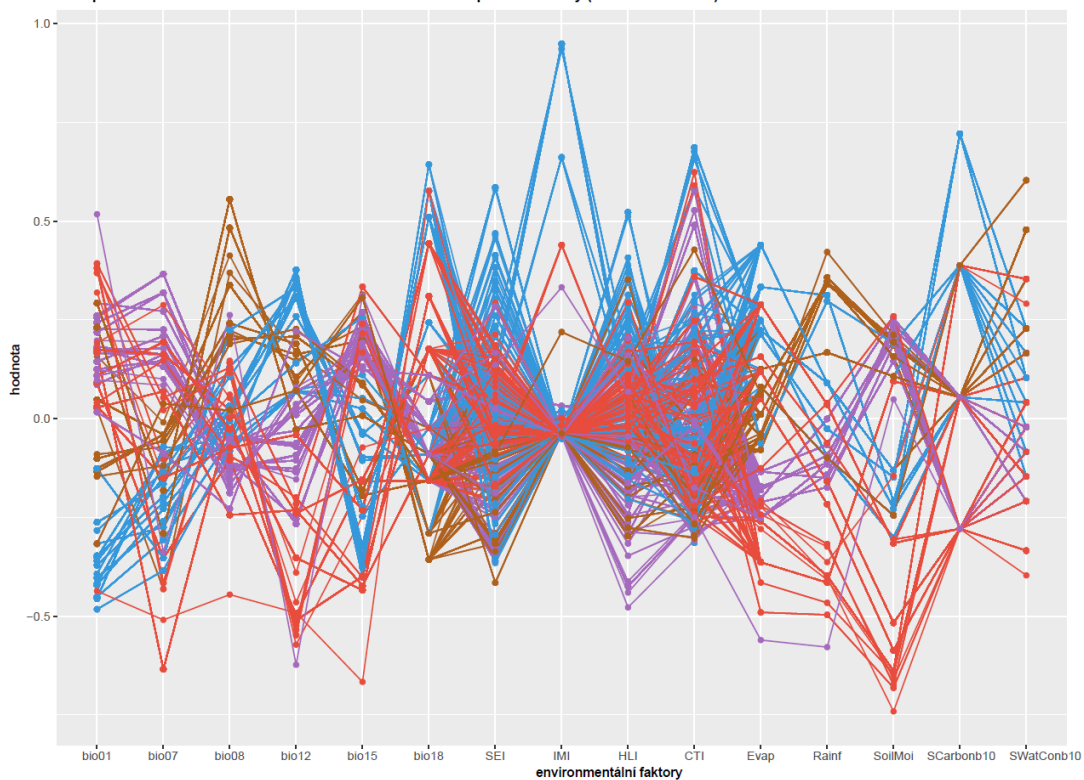
Jelikož je biplot zobrazením pouze první a druhé hlavní komponenty, je třeba si závěry z něj vyvozené ověřit. K tomuto účelu posloužil graf paralelních souřadnic, který vizualizuje zastoupení jednotlivých bodů ze shluků v každém environmentálním faktoru (viz obr. 36). Z tohoto grafu lze vyvozovat stejné závěry jako z biplotu. Shluk 1 dosahuje vysokých hodnot u faktoru bio08, bio12, bio15 a Rainf. Shluk 2 dosahuje vysokých hodnot u bio18, a CTI, a naopak záporných hodnot dosahuje u bio07, bio12, Evap, Rainf a SoilMoi (stejně tak v biplotu jsou tyto faktory umístěny opačným směrem). Shluk 3 má vyšší hodnoty u bio01, bio07, bio15 a SoilMoi a u ostatních faktorů dochází spíše k negativnímu ovlivnění. Shluk 4 je kladně ovlivňován bio12, bio18, SEI, HLI, Evap a SCarbonb10.

**Biplot – shluky podle k-means na základě 15 environmentálních faktorů pro data cizny (*Cicer reticulatum*)**



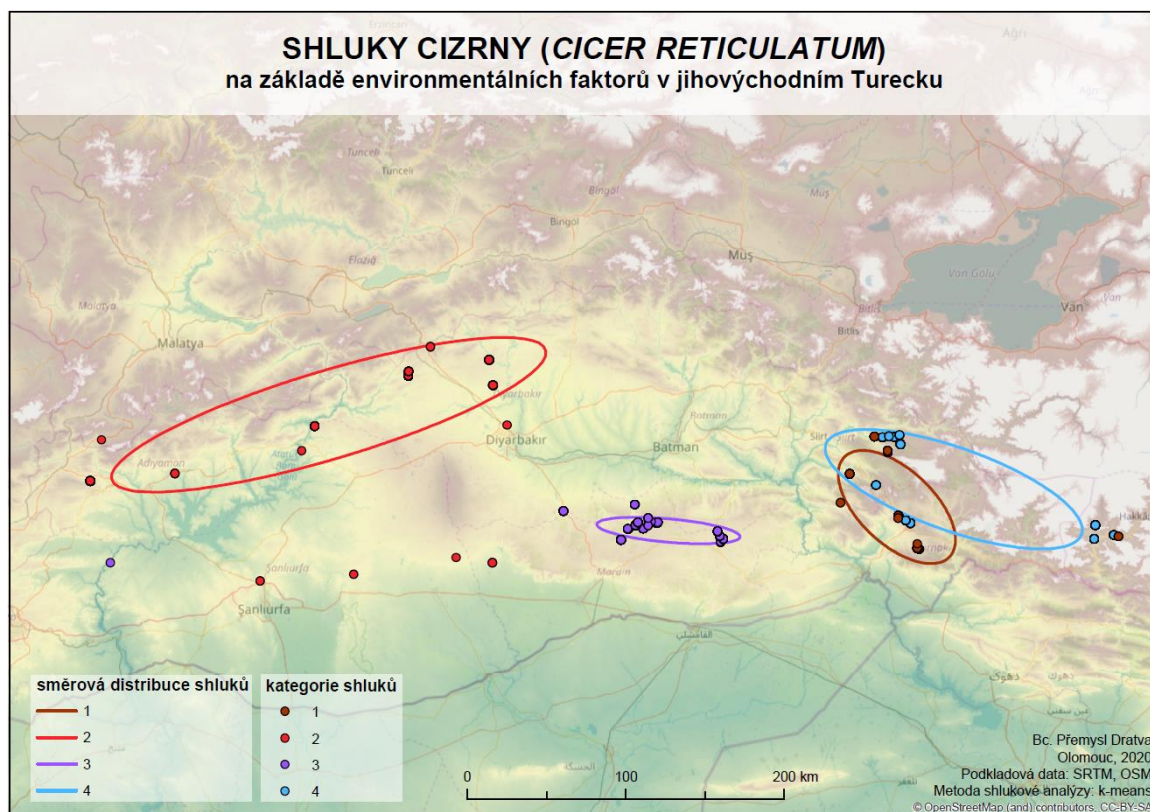
Obr. 35 Biplot se shluky pro *Cicer reticulatum*.

**Graf paralelních souřadnic – 15 environmentálních faktorů pro data cizny (*Cicer reticulatum*)**



Obr. 36 Graf paralelních souřadnic se shluky pro *Cicer reticulatum*.

Po zobrazení shluků do mapy (viz obr. 37) je patrné, že se body ze shluků 1 a 4 mísi a nejsou striktně odděleny geografickou vzdáleností, stejně tak jako je jejich mírné mísení znatelné i z výše popsaného biplotu. Body z těchto dvou shluků se nachází v hornaté oblasti, kde je vyšší nadmořská výška, která negativně koreluje s průměrnou roční teplotou (bio01), což je patrné i z biplotu, kde se body těchto dvou shluků nacházejí na opačné straně než faktor bio01. Shluk 2 je podobně jako v biplotu poměrně rozlehlý a nachází se ve větší geografické vzdálenosti od ostatních shluků. Body se v hodně případech nachází poměrně blízko vodním tokům v ne příliš vysoké nadmořské výšce. Body ve shluku 3 se až na jeden odlehlý bod nachází v blízké vzdálenosti a relativně konstantní nadmořské výšce.



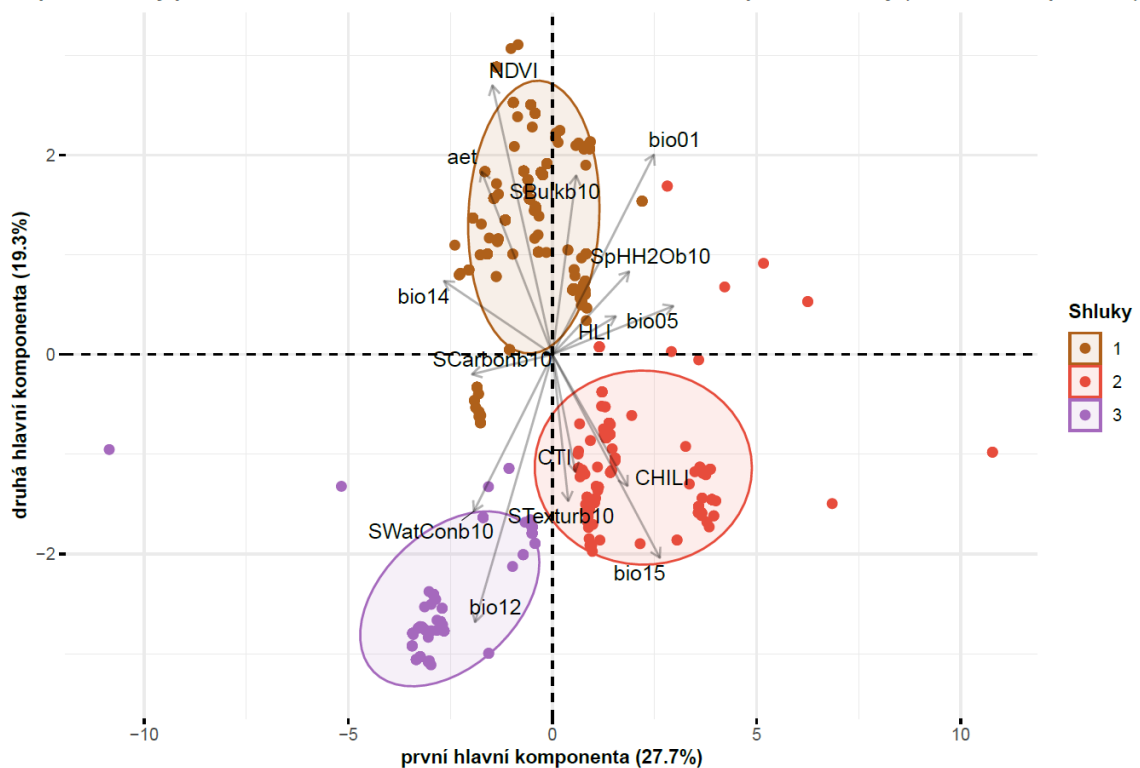
Obr. 37 Mapa shluků *Cicer reticulatum*.

## 6.2.2 Cizrna (*Cicer echinospermum*)

Pro *Cicer echinospermum* byly statistickou analýzou vyčleněny počty 10, 5 a 3 shluků. Pro následnou analýzu byl vybrán počet tří shluků. Vzhledem k tomu, že graf paralelních souřadnic u všech druhů rostlin vykazuje stejné výsledky jako biplot, je k dalšímu popisu výsledků využít pouze biplot, jelikož je přehlednější a lépe interpretovatelný.

V grafu lze pozorovat (viz obr. 38), že shluk 1 je dobře separovaný v kladných hodnotách druhé hlavní komponenty a nejvíce je ovlivněn faktory NDVI, aet (aktuální evapotranspirace), SBulkb10 (hustota půdy), bio01 (průměrná roční teplota), bio14 (srážkový úhrn v nejsušším měsíci), SpHH2Ob10 (pH vody v půdě) a HLI. Negativně je ovlivňován indexy CHILI a CTI, dále bio15 (sezónnost srážek), STexturb10 (textura půdy) a bio12 (roční úhrn srážek).

Biplot – shluky podle k-means na základě 15 environmentálních faktorů pro data cizrny (*Cicer echinospermum*)



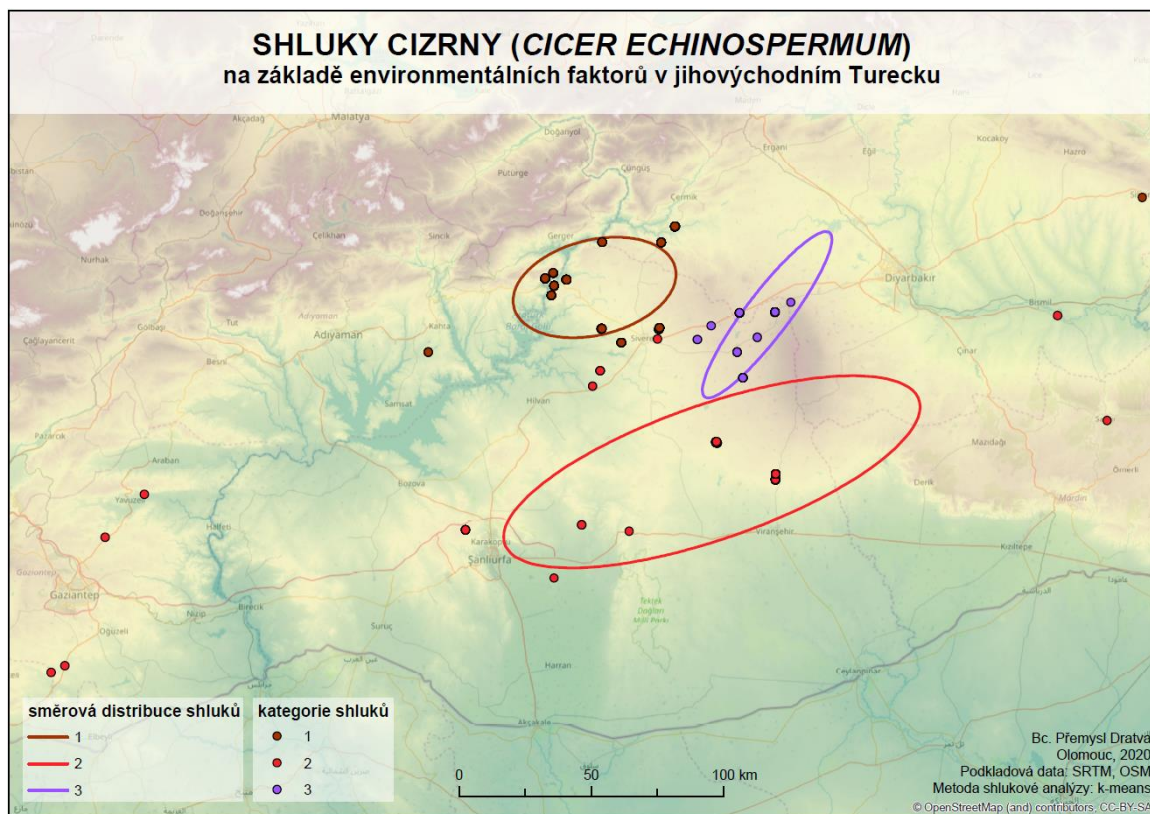
Obr. 38 Biplot se shluky pro *Cicer echinospermum*.

Shluky 2 a 3 jsou naopak separovány v záporných hodnotách druhé hlavní komponenty a měly by tak být více odlišné od shluku 1. Shluk 2 je formován v blízkosti faktorů CHILI, bio15 (sezónnost srážek), STexturb10 (textura půdy) a CTI, zatímco shluk 3 je více ovlivňován bio12 (roční úhrn srážek), a SWatConb10 (obsah půdní vody v 33 kPa).

Shluky zobrazené v mapě jsou od sebe relativně dobře oddělené (viz obr. 39). Většina bodů shluku 1 se nachází v blízkosti vodního toku. Shluk 2 lze charakterizovat velkou geografickou vzdáleností mezi body. Jsou položeny spíše v jižnější části než body ostatních shluků. Shluk 3 je rozmístěn v poměrně malé oblasti v porovnání s ostatními



shluky. Tyto body se nachází ve vyšší nadmořské výšce a v porovnání s biplotem lze říci, že nejsou příliš závislé na teplotě.

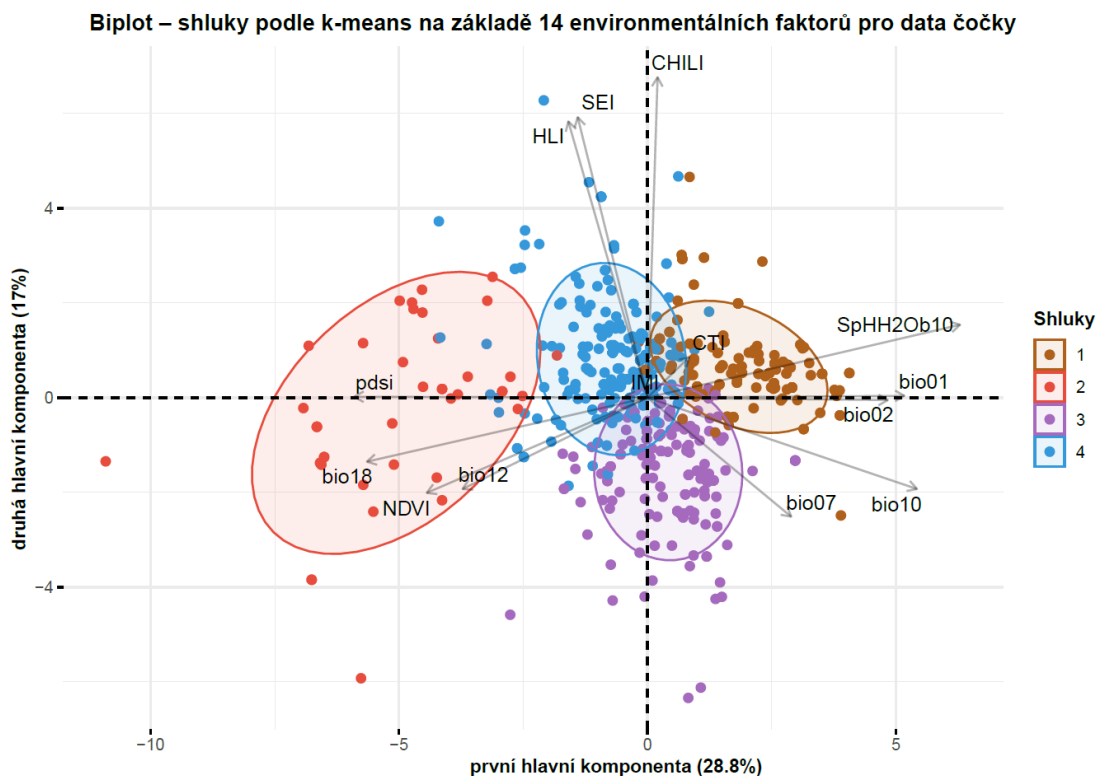


Obr. 39 Mapa shluků *Cicer echinospermum*.

### 6.2.3 Čočka (*Lens orientalis*)

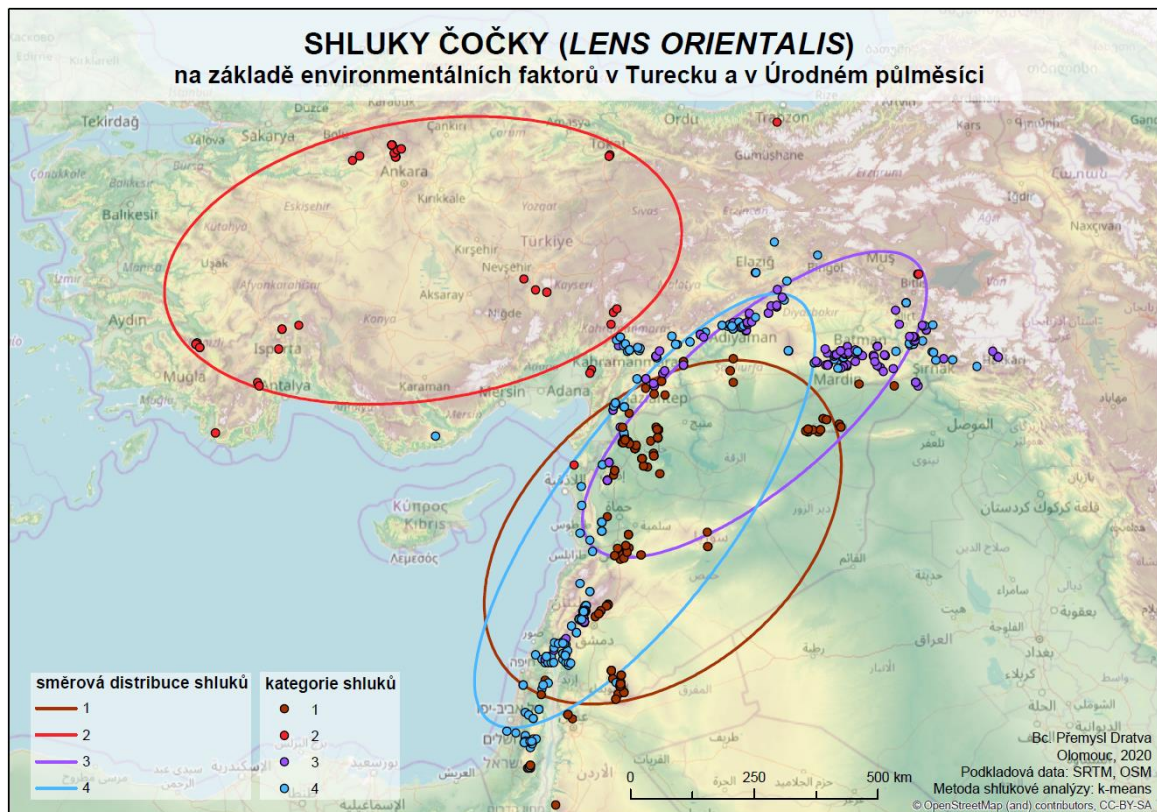
U čočky byl statistickou analýzou určen počet 10, 4 a 3 shluků, přičemž k následnému postupu byl vybrán počet čtyř shluků. Z biplotu znázorňujícího tyto shluky spolu s environmentálními faktory (viz obr. 40) lze vyčíst, že shluky nejsou vyčleněny tak výrazně jako v předchozích případech a dochází zde k jejich překryvu. Nejvíce se vzájemně překrývají shluky 1, 3 a 4, přičemž jejich vzájemný střed se nachází přibližně v místě styku hlavních komponent. Na shluk 1 nejvíce působí environmentální faktory CTI, SpHH2Ob10 (pH vody v půdě), bio01 (průměrná roční teplota), bio02 (průměrný denní rozsah teplot) a bio10 (průměrná teplota nejteplejšího čtvrtletí). Shluk 2 je od ostatních vyčleněn nejvýrazněji, ačkoliv se jeho body z části mísí s body ze shluku 4. Shluk 2 závisí na pdsí (Palmerův index intenzity sucha), bio18 (srážkový úhrn v nejteplejším čtvrtletí), NDVI a bio12 (roční úhrn srážek). Shluk 3 je nejlépe odlišitelný v záporných hodnotách druhé hlavní komponenty a je ovlivněn faktory bio10 (průměrná teplota nejteplejšího čtvrtletí), bio07 (roční rozsah teplot) a z části také NDVI, bio12 (roční úhrn srážek), bio01 (průměrná roční teplota) a bio02 (průměrný denní rozsah teplot). Naopak spíše záporně jej ovlivňují indexy HLI, SEI a CHILL, na rozdíl od shluku 4, na který mají tyto faktory velký vliv. Tento shluk je z části ovlivněn všemi sledovanými faktory, jelikož se nachází v kladné i záporné části obou hlavních komponent, ačkoli faktory nacházející se v kladné části první hlavní komponenty mají na většinu bodů z tohoto shluku spíše negativní vliv.

V mapě lze pozorovat stejnou tendenci jako v grafu, a to prolínání shluků 1, 3 a 4 a jejich oddělení od shluku 2 (viz obr. 41). Jednotlivé shluky lze i navzdory jejich mísení relativně dobře prostorově charakterizovat. Nejvíce je od ostatních vyčleněný shluk 2, jehož body jsou rozmístěny především v hornaté části Turecka. Rozmístění bodů ostatních shluků kopíruje oblast Úrodného půlměsíce. Nejblíže ke Středomořnímu moři se vyskytují body ze shluku 4. V jejich relativní blízkosti, ale již více vzdálené od pobřeží jsou body ze shluku 1 a body ze shluku 3 se nachází převážně v jihovýchodním Turecku.



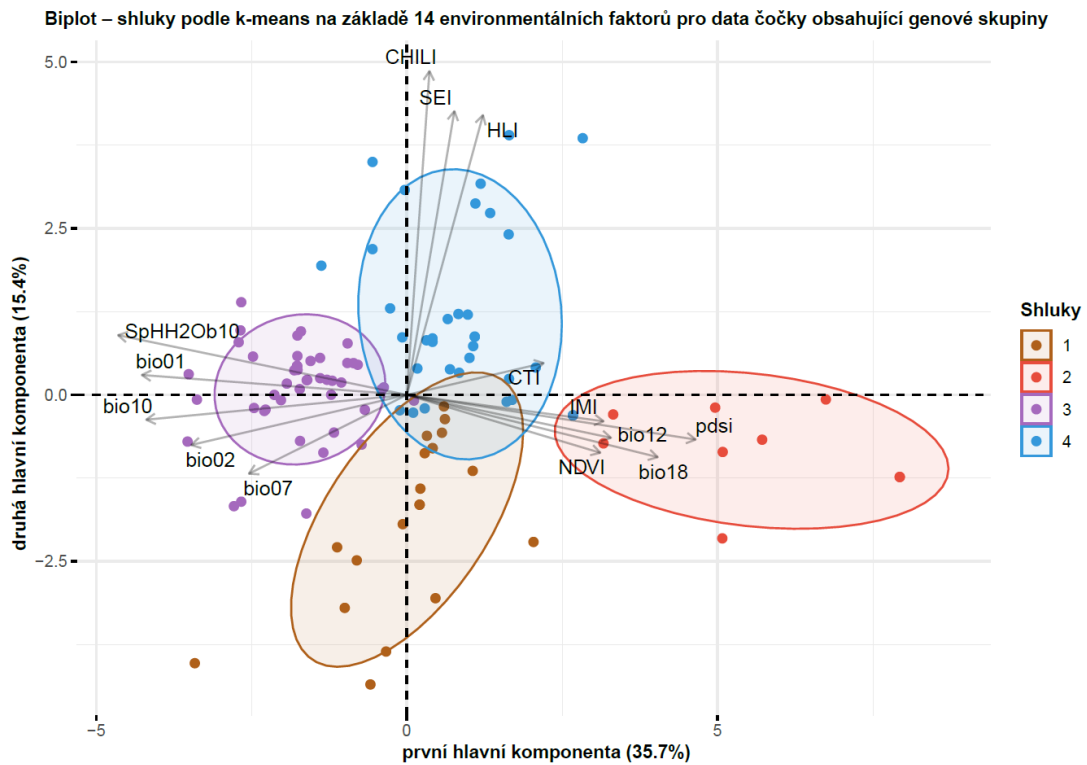
Obr. 40 Biplot se shluky pro čocky.





Obr. 41 Mapa shluků čočky.

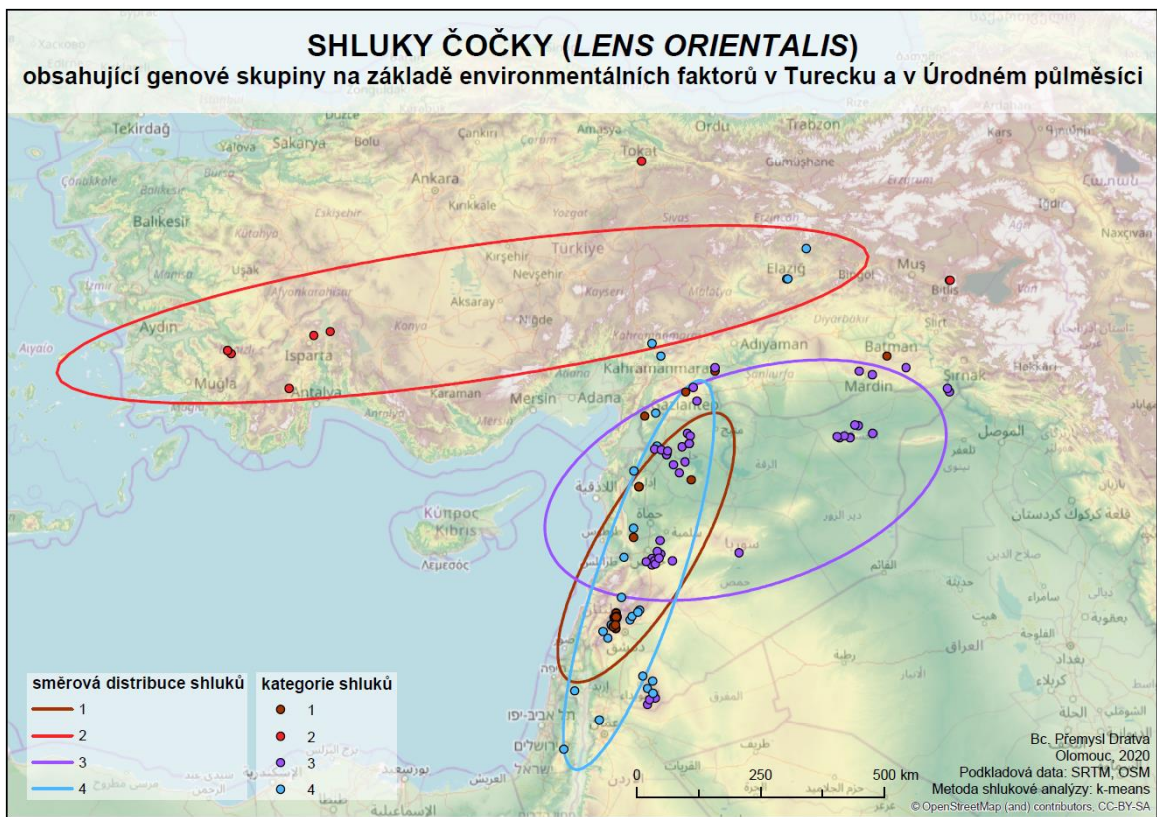
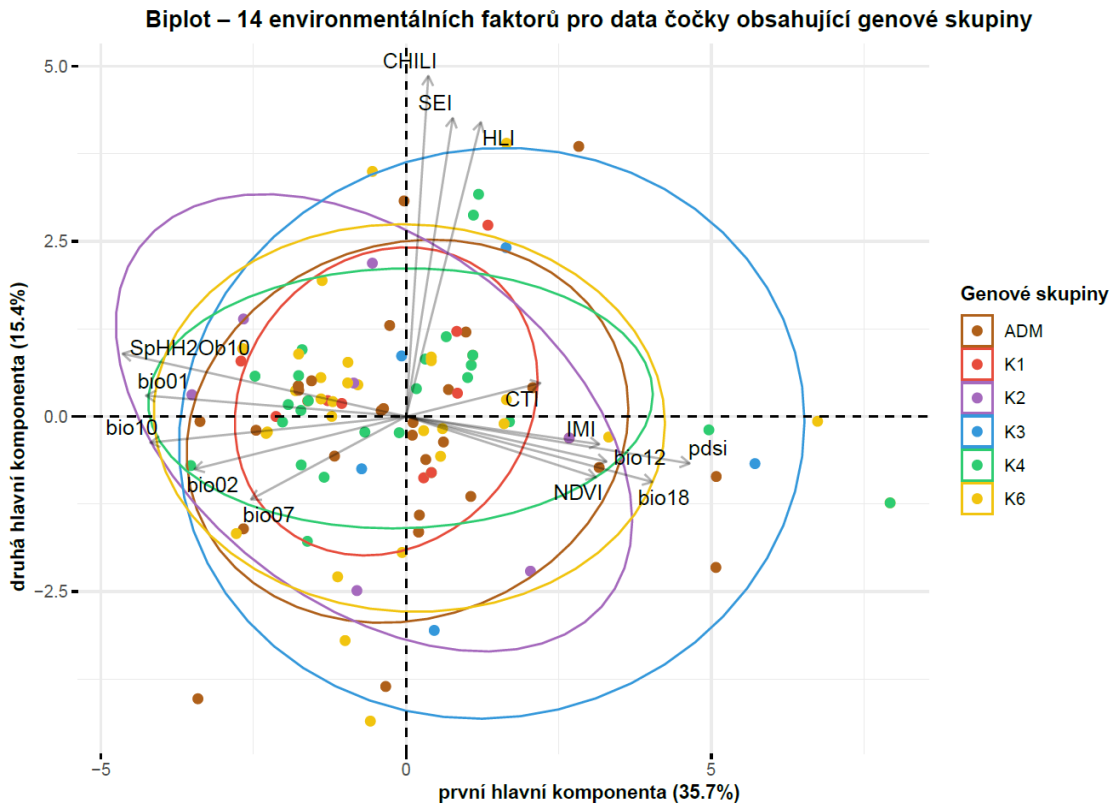
Následně byly tyto výstupy vytvořeny také pro 109 bodů čočky, které obsahují informace o zařazení do genetických skupin, a to ve dvou způsobech. První je vytvoření nových shluků na základě environmentálních charakteristik a druhý je vizualizace genetických skupin v závislosti na environmentálních faktorech. Biplot pro nově vytvořené shluky je podobný biplotu se všemi body čočky (viz obr. 42). Shluk 4 se částečně překrývá se shluky 1 a 3, shluk 2 je od nich více separovaný. Shluk 2 je ovlivňován faktory IMI, pdsi (Palmerův index intenzity sucha), bio12 (roční úhrn srážek), bio18 (srážkový úhrn v nejteplejším čtvrtletí) a NDVI, ostatní faktory na něj působí spíše negativně a indexy CHILI, SEI a HLI vůbec. Shluk 1 není výrazně ovlivňován žádným ze zkoumaných environmentálních faktorů, částečně na něj působí faktory, působící na shluk 2 a také bio07 (roční rozsah teplot). Shluk 3 se nachází na opačné straně od shluku 2, je tedy negativně ovlivňován faktory, které pozitivně působí na shluk 2 a sám je pozitivně ovlivňován bioklimatickými proměnnými bio01 (průměrná roční teplota), bio02 (průměrný denní rozsah teplot), bio07 (roční rozsah teplot) a bio10 (průměrná teplota nejteplejšího čtvrtletí) a také faktorem SpHH2Ob10 (pH vody v půdě). Shluk 4 je v tomto případě nejvýrazněji ovlivňován indexy CHILI, SEI, HLI a CTI, částečný vliv mohou mít také faktory působící na shluk 2.



Obr. 42 Biplot se shluky pro data čocky obsahující genetické skupiny.

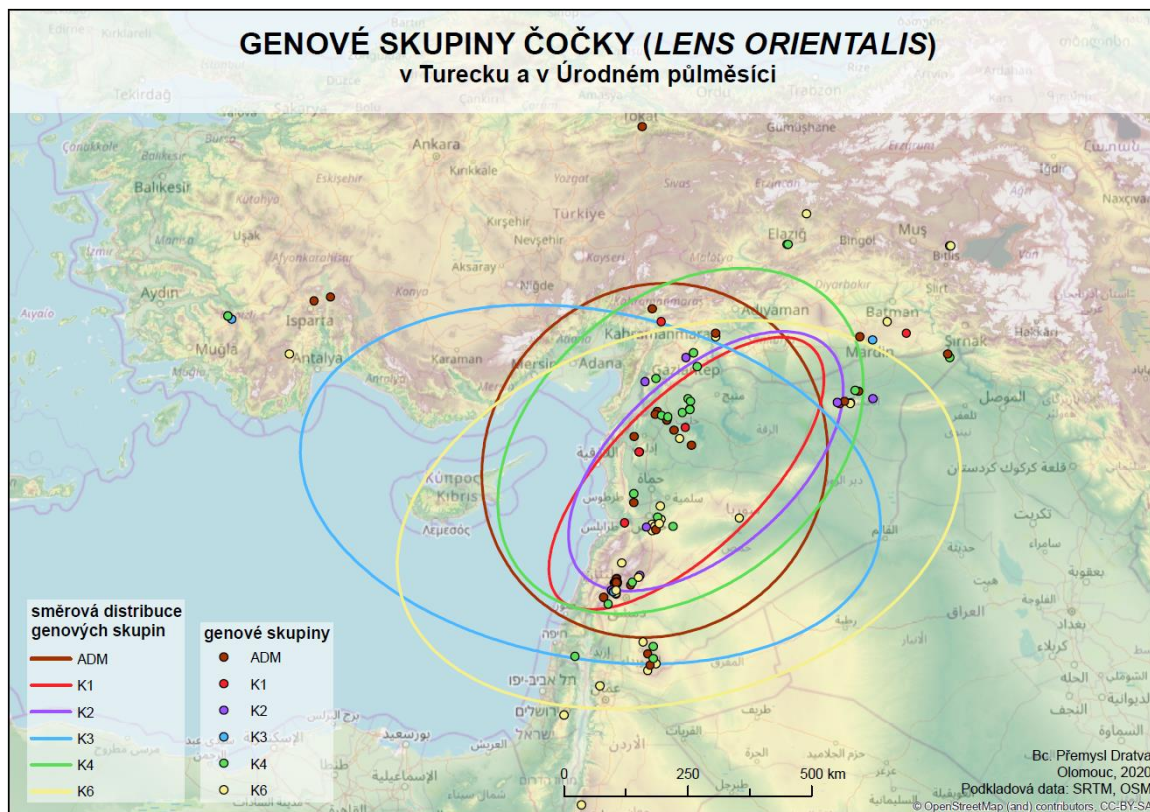
V biplotu znázorňujícím genetické skupiny nebyly tyto skupiny nijak výrazně vyčleněny (viz obr. 43). Body ze všech skupin se mísí, u žádné skupiny nelze striktně vymezit její odlišení od skupin ostatních.

Mapa zobrazující vytvořené shluky čocky s informací o genetických skupinách (viz obr. 44) je opět značně podobná mapě se všemi body čocky. Body ze shluku 2 se rozprostírají po Turecku ve vyšších nadmořských výškách. Shluky 1 a 4 jsou formovány spíše v blízkosti východního pobřeží Středozemního moře a shluk 3 směřuje více do vnitrozemí.





Mapa shluků genetických skupin čočky nepotvrzuje odlišnost genetických skupin rostlin na základě geografické vzdálenosti (viz obr. 45).

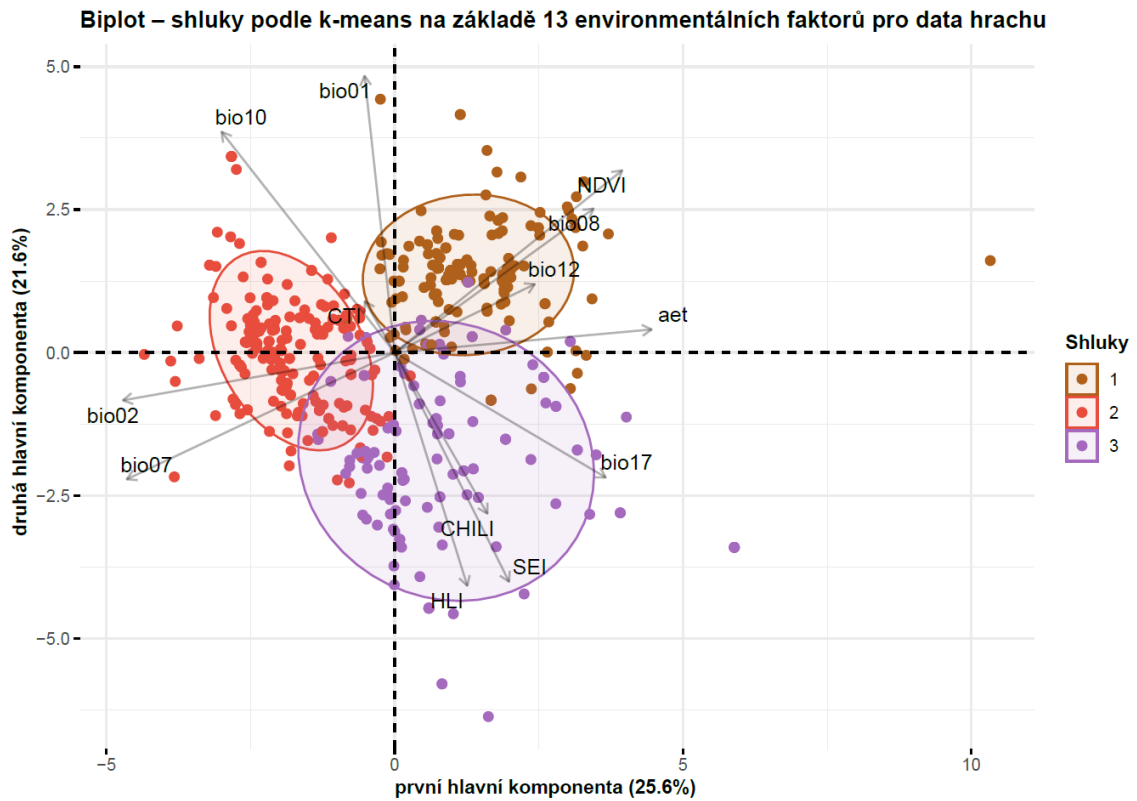


Obr. 45 Mapa shluků genetických skupin čočky.

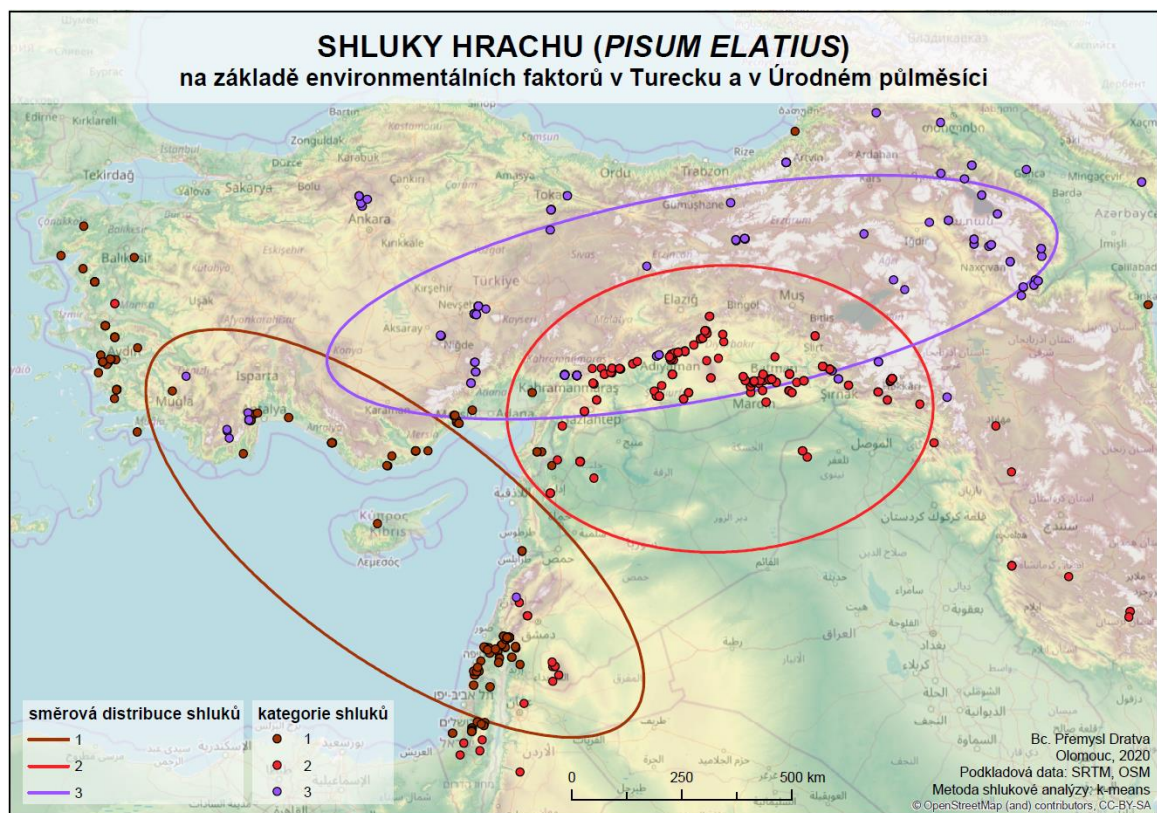
#### 6.2.4 Hrách (*Pisum elatius*)

Pro data hrachu byl určen počet tří z doporučených 9, 6 a 3 shluků. Z biplotu vytvořeného pro tyto tři shluky je patrné, že se od sebe shluky příliš výrazně nesePARovaly (viz obr. 46). Body ze shluku 3 se částečně mísí s body dalších dvou shluků. Shluk 1 je formován faktory NDVI, bio08 (průměrná teplota nejvlhčího čtvrtletí), bio12 (roční úhrn srážek), aet (aktuální evapotranspirace) a bio01 (průměrná roční teplota). Shluk 2 je v opačném směru ovlivňován bioklimatickými proměnnými bio02 (průměrný denní rozsah teplot) a bio07 (roční rozsah teplot), z části také bio10 (průměrná teplota nejteplejšího čtvrtletí) a indexem CTI. Body ze shluku 3 zasahují do všech částí grafu a některé z nich tam mohou být kladně ovlivněny téměř všemi faktory. Velký vliv na nejvíce těchto bodů mají ovšem indexy HLI, SEI a CHILI a bioklimatická proměnná bio17 (srážkový úhrn v nejsušším čtvrtletí).

Z mapy znázorňující shluky rostlin hrachu (viz obr. 47) je na první pohled patrné, že body ze shluku 1 se nachází v blízkosti pobřeží Středozemního moře, popřípadě jiných vodních ploch a toků v nížinných oblastech. Body ze shluku 2 se nachází ve vyšších nadmořských výškách především v jihovýchodním Turecku a body ze shluku 3 se vyskytují ve vyšších nadmořských výškách v pohoří Taurus.



Obr. 46 Biplot se shluky pro hrách.



Obr. 47 Mapa shluků hrachu.

## 6.3 Polohová přesnost

Pro zjištění vlivu polohové přesnosti bodů rostlin byla vybrána data hrachu s přesností 1, ke kterým byly náhodně vytvořeny body v oblasti polohové chybovosti bodů s přesností 3, tedy ve čtvercové oblasti o délce hrany 2,22 km. Tyto dvě přesnosti byly porovnány nad bioklimatickými proměnnými s rozlišením 1 km a nad výškovými daty SRTM30 a indexy z něj odvozenými s rozlišením 30 m.

Na základě vypočítaného procentuálního rozdílu aritmetického průměru, mediánu a rozpětí hodnot obou přesností bylo zjištěno, že při využití bioklimatických proměnných s rozlišením 1 km se body s oběma přesnostmi téměř neliší (viz tabulka 13). U aritmetického průměru dosahuje procentuální rozdíl průměrně 0,2 %, u mediánu je průměrný rozdíl 0,9 % a u rozpětí hodnot je to 1,8 %. Tyto hodnoty jsou zanedbatelné, a proto by nepřesnost dat s přesností 3 neměla výsledné analýzy nad datovými sadami WorldClim s rozlišením 1 km výrazně ovlivňovat. To dokazují i grafy komponentních vah vytvořené z dat čocky s přesnostmi 1 a 3 (viz obr. 48 a obr. 49). Úhly mezi jednotlivými průvodiči jsou téměř totožné, stejně tak délky všech průvodičů se v obou grafech neliší. Rozdíl je pouze v „pootočení“ bioklimatických proměnných mezi prvními dvěma dimenzemi a také nepatrný rozdíl v zastoupení dat v těchto dvou dimenzích.

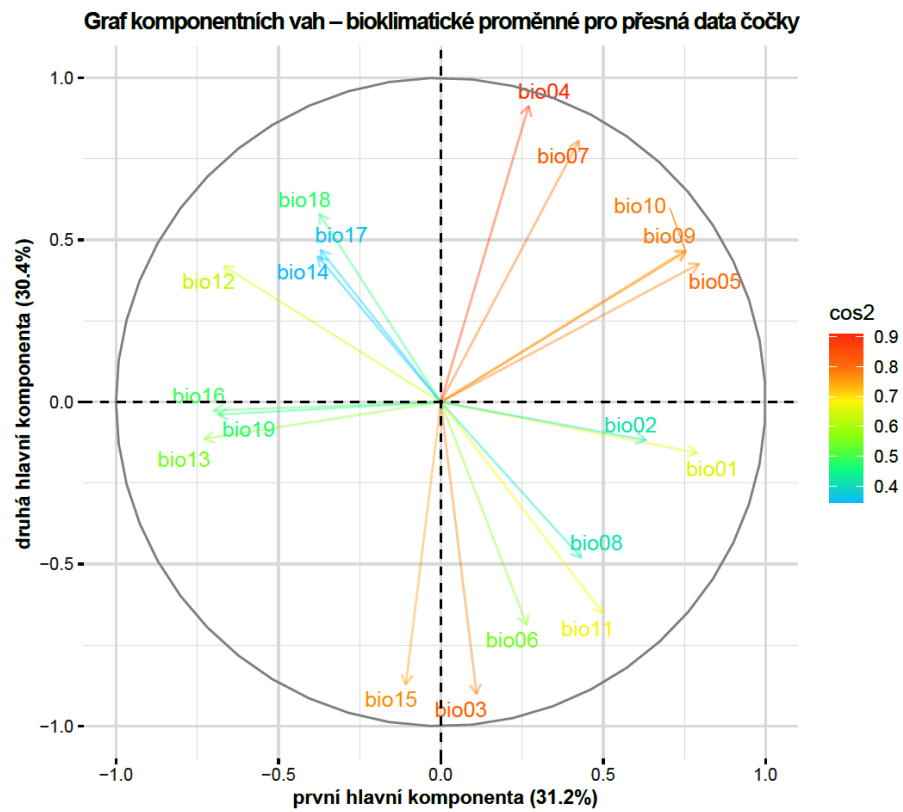
Tabulka 13 Procentuální rozdíl mezi přesnými a méně přesnými daty u dat z WorldClim

	bioklimatická data z WorldClim v rozlišení 1 km																		
	bio01	bio02	bio03	bio04	bio05	bio06	bio07	bio08	bio09	bio10	bio11	bio12	bio13	bio14	bio15	bio16	bio17	bio18	bio19
rozpětí hodnot	0,7	4,1	1,4	1,4	4,7	1,2	4,4	0,0	1,3	0,5	2,0	0,2	0,5	2,4	1,6	0,6	3,3	3,3	0,6
median	0,3	0,6	0,6	0,4	0,0	3,1	0,0	2,1	0,3	0,1	2,0	0,2	2,0	0,0	0,4	1,1	0,0	0,0	3,2
aritmetický průměr	0,2	0,2	0,1	0,0	0,1	0,2	0,1	0,1	0,1	0,1	0,4	0,2	0,0	0,6	0,1	0,1	0,7	0,1	0,1

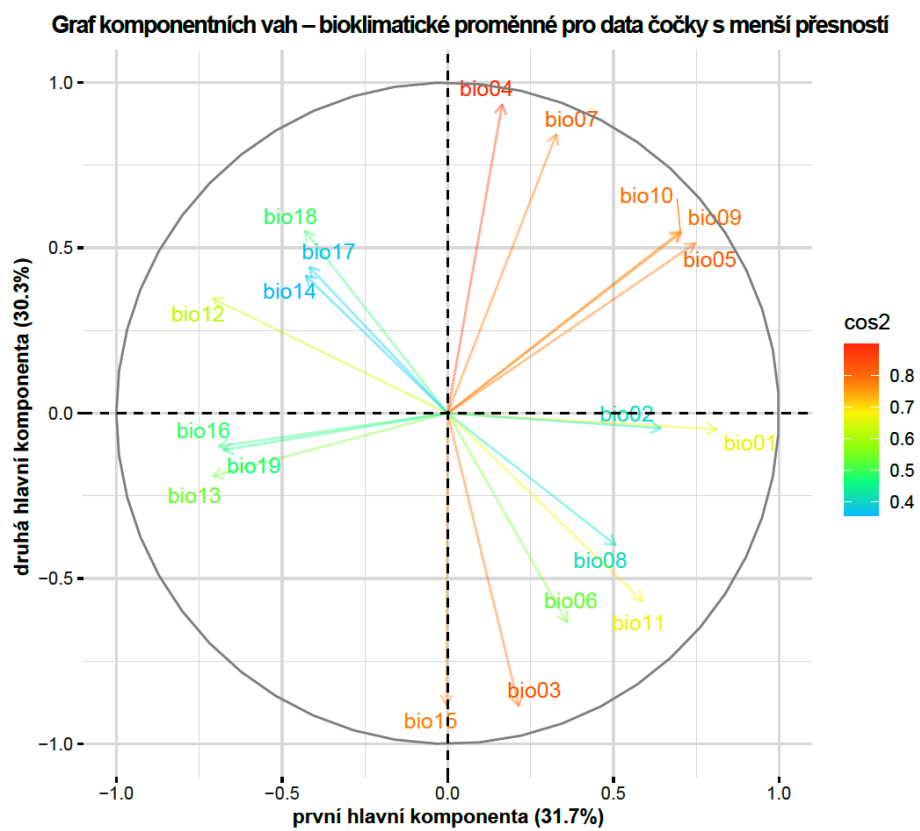
Výraznější rozdíly nastávají u datových sad s rozlišením 30 metrů (viz tabulka 14). Procentuální rozdíl rozpětí hodnot je v tomto případě v průměru 15,6 %, u mediánu je průměrný rozdíl 15,3 % a u aritmetického průměru je to 10,4 %. Tyto hodnoty už mohou výsledné analýzy poměrně zkreslovat, proto je vhodné pro datové sady s rozlišením 30 metrů používat pouze data s lepší polohovou přesností. Rozdíl je znatelný i při porovnání grafů komponentních vah prováděných nad těmito daty (viz obr. 50 a obr. 51). Největší rozdíl je patrný u SRTM30, kdy u dat s menší přesností má tato sada větší zastoupení v obou dimenzích a na rozdíl od přesných dat směřuje do kladné části první hlavní komponenty. Dále lze v grafu pozorovat mírnou změnu směru všech indexů.

Tabulka 14 Procentuální rozdíl mezi přesnými a méně přesnými daty u dat SRTM30 a odvozených indexů

	SRTM30 a z něj odvozené indexy v rozlišení 30 m					
	SRTM30	CTI	HLI	IMI	SEI	TRASP
rozpětí hodnot	22,3	1,3	21,8	15,3	32,8	0,0
median	2,2	0,8	0,3	1,3	66,6	20,4
aritmetický průměr	1,0	0,2	0,6	8,9	40,3	11,4



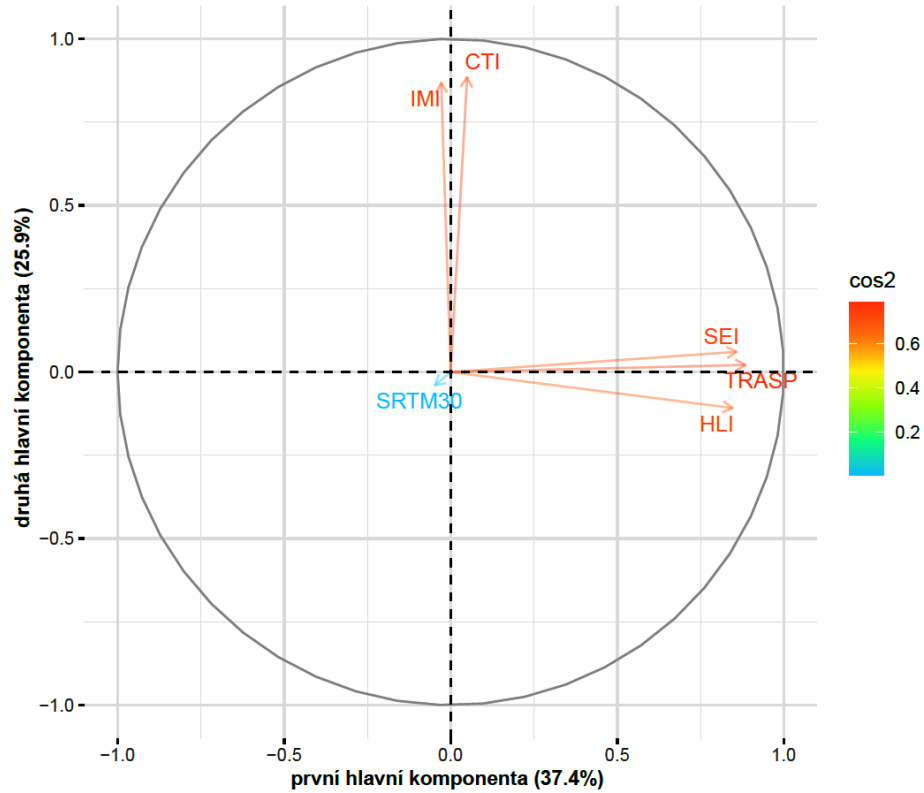
Obr. 48 Graf komponentních vah bioklimatických proměnných pro data čocky s přesností 1.



Obr. 49 Graf komponentních vah bioklimatických proměnných pro data čocky s přesností 3.

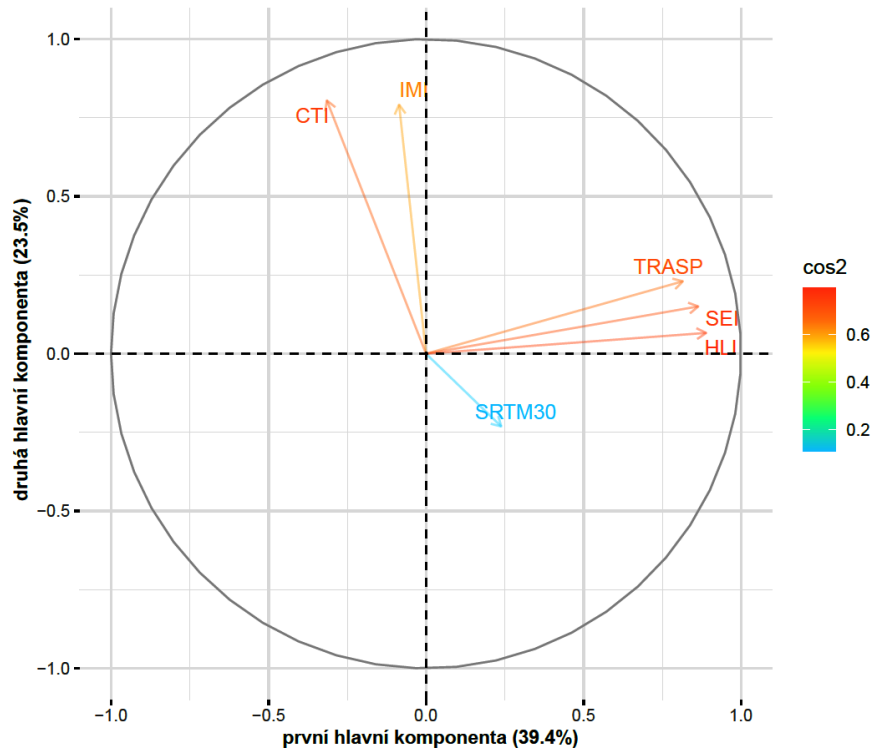


Graf komponentních vah – výškový model SRTM30 a odvozené indexy pro přesná data čocky



Obr. 50 Graf komponentních vah SRTM30 a odvozených indexů pro data čocky s přesností 1.

Graf komponentních vah – výškový model SRTM30 a odvozené indexy pro data čocky s menší přesností



Obr. 51 Graf komponentních vah SRTM30 a odvozených indexů pro data čocky s přesností 3.



## 7 SHRNU TÍ VÝSLEDKŮ

Diplomová práce se zabývá zjišťováním vlivu environmentálních charakteristik na výskyt planých předchůdců cizrny, čočky a hrachu. Toho je docíleno analýzou prostorových environmentálních datových sad v kombinaci s daty o výskytu jednotlivých rostlin.

Prvním z důležitých výstupů práce jsou dva skripty vytvořené pro stahování datových sad z platformy Google Earth Engine. Tyto skripty jsou koncipovány tak, aby byly hodnoty ze snímků extrahovány jednotlivým bodům rostlin a dále tak není třeba pracovat s rastrovými daty, ale pouze jen s bodovými daty rostlin obsahujícími hodnoty ze všech vybraných datasetů. Jeden skript slouží ke stažení průměrných měsíčních dat z datasetu, který obsahuje několik snímků z různého časového období. Druhý skript je určen pro získání dat z jednoho snímku, který obsahuje průměrné hodnoty za delší časové období.

Dále byly statistickou analýzou vybrány environmentální faktory, které ze statistického hlediska nejlépe popisují výskyt každého ze zkoumaných druhů rostlin. Toho bylo docíleno pomocí analýzy hlavních komponent a korelační matice, kdy byly vybrány faktory, které jsou na sobě vzájemně nezávislé a zároveň mají v datech rostlin největší zastoupení. Pro cizrnu (*Cicer reticulatum*), která je přímým předchůdcem kulturní cizrny, bylo vybráno 15 environmentálních faktorů (viz tabulka 9), do nichž v největší míře spadají faktory týkající se srážek a vlhkosti, teploty a indexy vypočítané z výškového modelu. U dalšího druhu cizrny (*Cicer echinospermum*) bylo opět vybráno 15 environmentálních faktorů (viz tabulka 10). Mezi nimi jsou ve větší míře zastoupeny faktory týkající se vlastností půdy než u předchozího druhu, teplotní a srážkové faktory byly vybrány také. Pro čočku (*Lens orientalis*) bylo vybráno 14 faktorů (viz tabulka 11). V tomto případě se jedná více o faktory týkající se teploty a indexů vypočítaných z výškového modelu. U hrachu (*Pisum elatius*) bylo vybráno 13 environmentálních faktorů (viz tabulka 12), které stejně jako v případě čočky zahrnují teplotní proměnné, indexy vypočítané z výškového modelu a srážkové faktory. Vliv všech i vybraných environmentálních faktorů je možné vidět v jejich vizualizaci formou grafů komponentních vah v kapitole 6.1.

Hlavními výsledky práce jsou výsledky shlukové analýzy. Ta byla provedena pro každý druh rostliny zvlášť na základě vybraných environmentálních faktorů. U každé rostliny byly vytvořeny shluky, které jsou popsány pomocí environmentálních faktorů a také podle geografické polohy. Vytvořené shluky byly pro interpretaci promítnuty do mapy, ale také do biplotů a grafů paralelních souřadnic. Jelikož není graf paralelních souřadnic pro data použitá v této práci příliš přehledný a jeho výsledky odpovídají výsledkům biplotu, byl pro interpretaci závislosti shluků na environmentálních faktorech využit pouze biplot.

Z dat cizrny (*Cicer reticulatum*) byly vytvořeny čtyři shluky, které se od sebe v grafu poměrně dobře separovaly. Co se týče geografického rozmístění shluků, body ze shluků 1 a 4 jsou si geograficky blízké a nachází se v severovýchodní části Úrodného půlměsíce ve vyšších nadmořských výškách než zbylé dva shluky. Na výskyt rostlin v této oblasti mají velký vliv faktory týkající se srážek, indexy týkající se oslunění svahů a z části také vlastnosti půdy. Shluk 2, který se nachází okolo vodních toků na úpatí hor Antitaurus v poměrně velké geografické vzdálenosti, je formován indexy vlhkosti a srážkami. Shluk 3, který je v grafu i v mapě nejvýrazněji separován a nachází se jižně od města

Batman v relativně konstantní nadmořské výšce, je nejvíce závislý na teplotních faktorech.

U cizrny (*Cicer echinospermum*) byly vytvořeny tři shluky. Body ze shluku 1 se nachází převážně v okolí vodního toku a jsou ovlivňovány vlastnostmi půdy, NDVI, teplotou atd. Shluk 2 má největší prostorovou distribuci z těchto tří shluků, nachází se jak v nížinách, tak i vyšších nadmořských výškách. Je ovlivňován sezónností srážek, texturou půdy a indexem CHILI a CTI. Shluk 3 se nachází na severozápadní části sopky Karacadag a je formován obsahem půdní vody a ročními srážkami.

Pro čočku (*Lens orientalis*) byly vytvořeny 4 shluky, přičemž nejvíce separovaný je shluk 2 ležící v hornaté části Turecka. Tento shluk je nejvíce ovlivněn srážkami, indexy sucha a NDVI. Body z dalších tří shluků se mísí, a to jak z hlediska geografické polohy, tak i v závislosti na environmentálních faktorech v biplotu a nejsou od sebe příliš dobře odlišitelné. Body nacházející se u pobřeží Středozemního moře jsou nejvíce ovlivňovány indexy zaměřenými na oslunění svahu a také srážkami. Body ze shluků 1 a 3 se nachází více ve vnitrozemí a jsou tak více ovlivněny teplotními faktory.

Při porovnání výsledků shlukování na základě environmentálních faktorů s genetickými skupinami nebyla nalezena žádná shoda. Zároveň nedošlo k výrazné separaci bodů jednotlivých skupin, a to jak v biplotu tak ani v mapě.

Data hrachu (*Pisum elatius*) jsou rozdělena do tří shluků. Body prvního shluku jsou rozmístěny podél pobřeží Středozemního moře a jsou ovlivněny teplotními faktory, průměrnými srážkami a indexem NDVI. Body ze shluku 2 se nachází v severní části Úrodného půlměsíce na úpatí pohoří Antitaurus a jsou ovlivňovány především teplotními faktory. Shluk 3 je rozprostřen v hornatých oblastech Turecka až po Arménskou vysočinu. Tento shluk je formován indexy vypočtenými z výškového modelu a srážkami.

Bylo zjištěno, že při porovnání přesností 1 a 3 nemá polohová přesnost bodových dat rostlin téměř žádný vliv při použití bioklimatických proměnných s prostorovým rozlišením 1 km. Naopak při použití datové sady SRTM30 a indexů z ní odvozených, které mají rozlišení 30 metrů, se vliv nepřesnosti odrazuje na statistickém vyhodnocení a autor proto doporučuje pro data s tímto rozlišením využít pouze bodová data s vyšší přesností.

## 8 DISKUZE

Tato práce měla několik vytyčených cílů. Těmi jsou získání a analýza dostupných globálních environmentálních prostorových dat, dále nalezení faktorů ovlivňující výskyt planých předchůdců cizrny, čočky a hrachu, vytvoření shluků těchto rostlin na základě nalezených faktorů a zjištění vlivu přesnosti vstupních dat o rostlinách na výsledné analýzy. Všechny tyto cíle byly v průběhu řešení práce naplněny.

Pro řešení práce bylo z Katedry botaniky UP obdrženo celkem 4 996 bodových výskytů rostlin. Z těchto bodů bylo ovšem použito jen 3 487 bodů kvůli selekci, která obnášela mazání duplicitních bodů, bodů s nízkou přesností, s nevhodnou polohou nebo smazání záznamů obsahující NULL hodnoty, ale také z důvodu provádění analýz pouze pro data s přesností 1. Zároveň bylo použito 109 bodů čočky obsahující informaci o zařazení do genetických skupin, čímž pádem nebylo možné efektivně posoudit vytvořené shluky s výsledky genotypování a u ostatních druhů rostlin to nemohlo být vůbec provedeno.

Z platformy Google Earth Engine je možné stáhnout velké množství rozmanitých datových sad, některé byly ovšem z výběru vyřazeny kvůli nízkému prostorovému rozlišení, nejnižší rozlišení datové sady, která byla použita, bylo 11 km. Zbylé datové sady byly vybírány na základě konzultace s doc. Smýkalem a vedoucím práce. Některé z datových sad byly staženy jako průměr z delšího časového období, některá data byla stažena jako měsíční průměry za celé časové období, ze kterého jsou data dostupná. Pro další postup práce byla ovšem i tato podrobnější data zprůměrována, do analýz tedy vstupovala průměrná hodnota za celé časové období. Průměrné hodnoty byly vytvářeny proto, aby data korespondovala s ostatními datovými sadami využitými v této práci. Měsíční data dále poslouží pracovníkům z Katedry botaniky UP pro jejich výzkum. Právě v možnosti využít podrobná měsíční data vidí autor velký potenciál pro budoucí rozšíření této práce a dalšího výzkumu. Výskyt rostlin by mohl být zkoumán v závislosti na environmentálních charakteristikách v různých obdobích roku, nebo by se mohl porovnat vývoj bioklimatických a ostatních environmentálních proměnných v čase.

Datové sady lze z Google Earth Engine stáhnout i v rastrovém formátu a data by tak mohla být využita i pro jiné účely. Pro účely této práce bylo však výhodnější hodnoty z rastrů extrahovat do bodových vrstev, čímž je ušetřeno množství času a nestahuje se velký objem dat, což by byla z hlediska využití úložiště Google Drive komplikace.

V průběhu tvorby statistických analýz bylo zkoušeno několik alternativ softwaru RStudio. Mezi ně se řadí softwary GeoDa, BioVinci, a nadstavba do softwaru Excel Analyse-it. Nakonec byl pro svoji obsáhlost a velké možnosti vybrán software RStudio.

Pro vizualizaci závislosti shluků na vybraných environmentálních faktorech je možné využít několik statistických metod. V této práci byla využita vizualizace pomocí biplotu a grafu paralelních souřadnic. Graf paralelních souřadnic by se mohl jevit jako lepší varianta vzhledem k tomu, že zobrazuje všechna data, a nejen určité procento, jak je tomu u biplotu se dvěma hlavními komponentami. Z obou těchto vizualizací ovšem v případě použitých dat plynou stejné závěry a jelikož je graf paralelních souřadnic pro použitá data nepřehledný, byla interpretace prováděna pouze na základě biplotu. V textu práce je tak graf paralelních souřadnic zařazen pouze na ukázkou u cizrny a u dalších druhů rostlin není popisován.

Co se týče zjištění vlivu polohové přesnosti bodových dat rostlin na výsledné analýzy, ověření bylo prováděno pomocí náhodně vytvořených bodů v zóně možného výskytu

v rámci polohové nepřesnosti rostlin. Právě kvůli náhodnému generování bodů je třeba brát v potaz, že tento krok může pokaždé vyjít jinak a výsledky se tak mohou lišit.

Přesnosti byly porovnávány pouze nad datovými sadami WorldClim, SRTM30 a také nad indexy odvozených z této datové sady. Možností by bylo také zkoumat přesnost nad daty staženými z Google Earth Engine, ale kvůli velkému rozsahu práce a časové náročnosti postupu stahování dat z Google Earth Engine nebylo porovnání nad těmito daty realizováno.

## 9 ZÁVĚR

Hlavním cílem diplomové práce bylo aplikovat geoinformatické přístupy při výzkumu diverzity planých předchůdců cizrny, čočky a hrachu. Jinými slovy zjistit vliv environmentálních faktorů na výskyt zkoumaných rostlin za pomoci prostorových dat. Součástí práce je rešerše na téma charakteristiky zkoumaných rostlin, Úrodného půlměsíce a současného stavu řešené problematiky. Praktická část je zaměřena na získání a zpracování potřebných dat, na statistické analýzy vedoucí ke zjištění vlivu environmentálních faktorů na výskyt zkoumaných rostlin a na interpretaci výsledků.

V diplomové práci byly použity dva druhy dat. Prvním z nich jsou data o rostlinách obdržena z Katedry botaniky UP, která byla obdržena v tabulkové podobě s údaji o poloze každé rostliny. Tato data bylo nutné vyselektovat a zpracovat do podoby vhodné k dalším analýzám. Druhým typem použitých dat jsou data environmentální, konkrétně se jedná o datové sady z platformy Google Earth Engine, bioklimatické proměnné z datové sady WorldClim a výšková data SRTM30. Významná část praktické části práce se zabývá právě získáním prostorových dat z Google Earth Engine, pro jejíž účely byly vytvořeny dva skripty usnadňující získání těchto dat.

Po zpracování všech potřebných dat do vhodné podoby následovala analytická část, ve které byly vytyčeny dva hlavní cíle, a to nalezení environmentálních faktorů, které nejvíce ovlivňují výskyt zkoumaných rostlin a vytvoření shluků rostlin na základě těchto vybraných faktorů. Pro dosažení obou těchto cílů byly využity statistické metody. Environmentální faktory byly hledány pro každý druh zkoumaných rostlin zvlášť a posloužila k tomu analýza hlavních komponent v kombinaci s korelační maticí. Ke každému druhu rostlin byly vybrány faktory, které spolu vzájemně nekorelují a zároveň mají na data největší vliv. Nalezení těchto faktorů bylo nutné pro zajištění správnosti dalšího kroku, jelikož při tvorbě shluků s mnoha faktory, které se vzájemně ovlivňují, mohou být výsledky shlukování výrazně zkresleny.

Pro účely shlukování byla vybrána metoda k-means, pomocí které byly pro každý druh rostlin vytvořeny shluky bodů rostlin formované pomocí vybraných environmentálních faktorů. Na základě těchto shluků lze u jednotlivých rostlin říci, které faktory nejvíce ovlivňují jejich výskyt. Pro vizualizaci vytvořených shluků byl zvolen biplot, pomocí kterého byla určena závislost shluků na daných faktorech a shluky byly promítnuty také do mapy pro zjištění jejich prostorové distribuce.

Posledním krokem praktické části práce je vyhodnocení vlivu polohové přesnosti dat rostlin na výsledné analýzy. Toto vyhodnocení je založeno na porovnání vysoce přesných dat s náhodně vytvořenými body v chybové oblasti méně přesných dat nad environmentálními daty různého prostorového rozlišení. Bylo zjištěno, že u dat s prostorovým rozlišením 1 km nemá nižší polohová přesnost (chybovost 1,11 km) na výsledky téměř žádný vliv. U dat s rozlišením 30 metrů je vliv nepřesnosti podstatně znatelnější.

Výsledky této práce a prostorová environmentální data, která byla v rámci práce stažená, poslouží pracovníkům Katedry botaniky UP v jejich výzkumu planých předchůdců kulturních bobovitých rostlin. Výsledky shlukování spolu s popisem vytvořených shluků z hlediska jejich závislosti na konkrétních environmentálních faktorech mohou napomoci ve výzkumech zaměřených na plané předchůdce bobovitých rostlin, které si kladou za cíl zvýšit genetickou rozmanitost plodin křížením jejich planých předchůdců, čímž bude navýšen jejich výnos.

Skripty vytvořené pro stažení dat z platformy Google Earth Engine mohou napomoci k většímu využití těchto datových sad v rámci podobných výzkumů, které s podobnými datovými sadami doposud téměř nepracovaly.

# SEZNAM OBRÁZKŮ

Obr. 1 Lokalizace výskytu rostlin .....	17
Obr. 2 Diagram postupu práce .....	24
Obr. 3 Oblast Úrodného půlměsíce .....	25
Obr. 4 Köppen-Geiger klimatická klasifikace .....	26
Obr. 5 Shluky podle metody PCA .....	29
Obr. 6 Dendrogram shluků .....	29
Obr. 7 Graf komponentních vah .....	30
Obr. 8 Biplot .....	31
Obr. 9 Shluky plané pšenice .....	31
Obr. 10 Rozhraní Google Earth Engine .....	34
Obr. 11 Záložka Inspektor v GEE .....	34
Obr. 12 Zobrazení datasetu MOD13Q1.006 Terra Vegetation Indices 16-Day Global 250m a bodové vrstvy hrachu .....	35
Obr. 13 Oblast chyby s přesností 2 .....	37
Obr. 14 Oblast chyby s přesností 3 .....	37
Obr. 15 Model pro úpravu SRTM30 .....	40
Obr. 16 SRTM30 rozdělené podle UTM zón a ořezané podle bufferů .....	40
Obr. 17 Ukázka tabulky zobrazující procenta dat v prvních 10 dimenzích dat <i>Cicer echinospermum</i> .....	42
Obr. 18 Ukázka scree plotu pro data <i>Cicer echinospermum</i> .....	42
Obr. 19 Ukázka zastoupení faktorů v datech <i>Cicer echinospermum</i> .....	43
Obr. 20 Ukázka části korelační matice dat <i>Cicer echinospermum</i> .....	44
Obr. 21 Ukázka grafu komponentních vah pro data <i>Cicer echinospermum</i> .....	44
Obr. 22 Vyhodnocení ideálního počtu shluků pro data <i>Cicer echinospermum</i> .....	45
Obr. 23 Ukázka biplotu pro data <i>Cicer echinospermum</i> .....	47
Obr. 24 Ukázka grafu paralelních souřadnic pro data <i>Cicer echinospermum</i> .....	48
Obr. 25 Ukázka shluků dat <i>Cicer echinospermum</i> v mapě .....	48
Obr. 26 Graf komponentních vah všech faktorů pro data cizrny ( <i>Cicer reticulatum</i> ) .....	51
Obr. 27 Graf komponentních vah vybraných faktorů pro data cizrny ( <i>Cicer reticulatum</i> ) .....	52
Obr. 28 Graf komponentních vah všech faktorů pro data cizrny ( <i>Cicer echinospermum</i> ) .....	53
Obr. 29 Graf komponentních vah vybraných faktorů pro data cizrny ( <i>Cicer echinospermum</i> ) .....	54
Obr. 30 Graf komponentních vah všech faktorů pro data čočky .....	55
Obr. 31 Graf komponentních vah vybraných faktorů pro data čočky .....	56
Obr. 32 Graf komponentních vah vybraných faktorů pro data čočky obsahující genové skupiny .....	56
Obr. 33 Graf komponentních vah všech faktorů pro data hrachu .....	58
Obr. 34 Graf komponentních vah vybraných faktorů pro data hrachu .....	58
Obr. 35 Biplot se shluky pro <i>Cicer reticulatum</i> .....	60
Obr. 36 Graf paralelních souřadnic se shluky pro <i>Cicer reticulatum</i> .....	60
Obr. 37 Mapa shluků <i>Cicer reticulatum</i> .....	61
Obr. 38 Biplot se shluky pro <i>Cicer echinospermum</i> .....	62
Obr. 39 Mapa shluků <i>Cicer echinospermum</i> .....	63
Obr. 40 Biplot se shluky pro čočku .....	64
Obr. 41 Mapa shluků čočky .....	65
Obr. 42 Biplot se shluky pro data čočky obsahující genetické skupiny .....	66
Obr. 43 Biplot s genetickými skupinami dat čočky .....	67
Obr. 44 Mapa shluků dat čočky obsahující genetické skupiny .....	67
Obr. 45 Mapa shluků genetických skupin čočky .....	68

Obr. 46 Biplot se shluky pro hrách .....	69
Obr. 47 Mapa shluků hrachu .....	69
Obr. 48 Graf komponentních vah bioklimatických proměnných pro data čočky s přesností 1 ....	71
Obr. 49 Graf komponentních vah bioklimatických proměnných pro data čočky s přesností 3 ....	71
Obr. 50 Graf komponentních vah SRTM30 a odvozených indexů pro data čočky s přesností 1 ..	72
Obr. 51 Graf komponentních vah SRTM30 a odvozených indexů pro data čočky s přesností 3 ..	72

## SEZNAM TABULEK

Tabulka 1 Počet rostlin z původních dat .....	16
Tabulka 2 Zdroje dat rostlin, která prošla selekcí .....	17
Tabulka 3 Bioklimatické proměnné .....	18
Tabulka 4 Hodnoty PDSI .....	20
Tabulka 5 Popis vrstev z Google Earth Engine .....	22
Tabulka 6 Přibližná polohová přesnost GNSS .....	36
Tabulka 7 Klasifikace dat dle přesnosti .....	37
Tabulka 8 Počet rostlin podle přesnosti .....	37
Tabulka 9 Vybrané environmentální faktory pro <i>Cicer reticulatum</i> .....	51
Tabulka 10 Vybrané environmentální faktory pro <i>Cicer echinospermum</i> .....	53
Tabulka 11 Vybrané environmentální faktory pro čočku .....	54
Tabulka 12 Vybrané environmentální faktory pro hrách .....	57
Tabulka 13 Procentuální rozdíl mezi přesnými a méně přesnými daty u dat z WorldClim .....	70
Tabulka 14 Procentuální rozdíl mezi přesnými a méně přesnými daty u dat SRTM30 a odvozených indexů .....	70

## POUŽITÁ LITERATURA A INFORMAČNÍ ZDROJE

AFSHIN, Ashkan a kol. Consumption of nuts and legumes and risk of incident ischemic heart disease, stroke, and diabetes: a systematic review and meta-analysis. *The American Journal of Clinical Nutrition* [online]. 2014, 100(1), 278-288 [cit. 2020-02-17]. DOI: 10.3945/ajcn.113.076901. ISSN 0002-9165. Dostupné z: <https://academic.oup.com/ajcn/article/100/1/278/4576571>

ALEXANDRATOS, N. a J. BRUINSMA. 2012. World agriculture towards 2030/2050: the 2012 revision. *ESA Working paper* No. 12-03. Rome, FAO

ALLEY, William M. *The Palmer Drought Severity Index: Limitations and Assumptions* [online]. 1984 [cit. 2020-03-01]. Dostupné z: <https://journals.ametsoc.org/doi/pdf/10.1175/1520-0450%281984%29023%3C1100%3ATPDSIL%3E2.0.CO%3B2>

A planetary-scale platform for Earth science data & analysis. *Earth Engine Data Catalog* c [online]. [cit. 2020-03-01]. Dostupné z: <https://developers.google.com/earth-engine/datasets/>

BALLERINE, Clayton. *Topographic Wetness Index Urban Flooding Awareness Act Action Support: Will and DuPage Counties* [online]. Illinois, 2017 [cit. 2020-03-21]. Dostupné z: <https://www.isws.illinois.edu/pubdoc/CR/ISWSCR2017-02.pdf>

BECK, Hylke a kol. Present and future Köppen-Geiger climate classification maps at 1-km resolution. *Scientific Data* [online]. 2018, 5(1) [cit. 2020-02-17]. DOI: 10.1038/sdata.2018.214. ISSN 2052-4463. Dostupné z: <http://www.nature.com/articles/sdata2018214>

BERGER, Jens a kol. Ecogeography of Annual Wild Cicer Species. *Crop Science* [online]. 2003, 43(3), 1076-1090 [cit. 2020-02-17]. DOI: 10.2135/cropsci2003.1076. ISSN 0011-183X. Dostupné z: <https://onlinelibrary.wiley.com/doi/abs/10.2135/cropsci2003.1076>

BROWN, Terence a kol. The complex origins of domesticated crops in the Fertile Crescent. *Trends in Ecology & Evolution* [online]. 2009, 24(2), 103-109 [cit. 2020-02-17]. DOI: 10.1016/j.tree.2008.09.008. ISSN 01695347. Dostupné z: <https://linkinghub.elsevier.com/retrieve/pii/S0169534708003479>

BRUNAZZI, Alice a kol. Molecular diversity and landscape genomics of the crop wild relative *Triticum urartu* across the Fertile Crescent. *The Plant Journal* [online]. 2018, 94(4), 670-684 [cit. 2020-02-17]. DOI: 10.1111/tpj.13888. ISSN 0960-7412. Dostupné z: <https://onlinelibrary.wiley.com/doi/abs/10.1111/tpj.13888>

COBA DE LA PEÑA, Teodoro a José J. PUEYO. Legumes in the reclamation of marginal soils, from cultivar and inoculant selection to transgenic approaches. *Agronomy for Sustainable Development* [online]. 2012, 32(1), 65-91 [cit. 2020-02-17]. DOI: 10.1007/s13593-011-0024-2. ISSN 1774-0746. Dostupné z: <http://link.springer.com/10.1007/s13593-011-0024-2>



COYNE, Clarice J. a kol. *Potential and limits of exploitation of crop wild relatives for pea, lentil, and chickpea improvement*. Legume Science [online]. [cit. 2020-04-12]. DOI: 10.1002/leg3.36. ISSN 26396181. Dostupné z: <http://doi.wiley.com/10.1002/leg3.36>

Datasets tagged openlandmap in Earth Engine. *Earth Engine Data Catalog f* [online]. [cit. 2020-03-01]. Dostupné z: <https://developers.google.com/earth-engine/datasets/tags/openlandmap>

DE CANDOLLE AP (1885) Origin of Cultivated Plants. *New York: International Scientific Series*, D. Appleton and Co.

DE VAUS, D. A. Surveys in Social Research. Psychology Press, 2002, 379 s. ISBN 9780415268578.

DIDAN, Kamel a kol. *MODIS Vegetation Index User's Guide* [online]. 2015 [cit. 2020-03-01]. Dostupné z: [https://lpdaac.usgs.gov/documents/103/MOD13\\_User\\_Guide\\_V6.pdf](https://lpdaac.usgs.gov/documents/103/MOD13_User_Guide_V6.pdf)

DOBROVOLNÝ, Petr. *Dálkový průzkum Země: digitální zpracování obrazu*. Brno: Masarykova univerzita, 1998. ISBN isbn80-210-1812-7.

DVOŘÁK, Marek. *Metody shlukové analýzy a jejich aplikace v marketingu*. 2008. Diplomová práce. Univerzita Karlova v Praze.

Earth Engine Code Editor. *Google Earth Engine* [online]. 2020 [cit. 2020-03-01]. Dostupné z: <https://developers.google.com/earth-engine/playground>

EL-BELTAGY, A (2006) Central and West Asia and North Africa: challenges and opportunities. *ICARDA*, Aleppo

Envirometrix / LandGISmaps. *GitHub* [online]. 2019 [cit. 2020-03-01]. Dostupné z: <https://github.com/Envirometrix/LandGISmaps#soil-properties-and-classes>

EVANS, Jeffrey a kol. *An ArcGIS Toolbox for Surface Gradient and Geomorphometric Modeling.: version 2.0* [online]. 2014 [cit. 2020-03-21]. Dostupné z: <https://github.com/jeffrejevans/GradientMetrics>

FLDAS: Project Goals. *NASA* [online]. 2020 [cit. 2020-03-01]. Dostupné z: <https://ldas.gsfc.nasa.gov/fldas>

GeoTIFF. *EARTHDATA* [online]. 2019 [cit. 2020-03-01]. Dostupné z: <https://earthdata.nasa.gov/esdis/eso/standards-and-references/geotiff>

Global SRTM CHILI (Continuous Heat-Insolation Load Index). *Earth Engine Data Catalog g* [online]. [cit. 2020-03-01]. Dostupné z: [https://developers.google.com/earth-engine/datasets/catalog/CSP\\_ERGo\\_1\\_0\\_Global\\_SRTM\\_CHILI](https://developers.google.com/earth-engine/datasets/catalog/CSP_ERGo_1_0_Global_SRTM_CHILI)

*Google Earth Engine* [online]. [cit. 2020-03-01]. Dostupné z: <https://earthengine.google.com/>

HRADILOVÁ, Iveta a kol. Variation in wild pea (*Pisum sativum* subsp. *elatius*) seed dormancy and its relationship to the environment and seed coat traits. *PeerJ* [online]. 2019, 7 [cit. 2020-02-17]. DOI: 10.7717/peerj.6263. ISSN 2167-8359. Dostupné z: <https://peerj.com/articles/6263>

IVERSON, Louis R. a kol. A GIS-derived integrated moisture index to predict forest composition and productivity of Ohio forests (U.S.A.). *Landscape Ecology*. 1997, 12(5), 331-348. DOI: 10.1023/A:1007989813501. ISSN 09212973. Dostupné také z: <http://link.springer.com/10.1023/A:1007989813501>

KILIAN, Benjamin a kol. Genetic Diversity, Evolution and Domestication of Wheat and Barley in the Fertile Crescent. GLAUBRECHT, Matthias, ed. *Evolution in Action* [online]. Berlin, Heidelberg: Springer Berlin Heidelberg, 2010, 2010-6-20, s. 137-166 [cit. 2020-02-17]. DOI: 10.1007/978-3-642-12425-9\_8. ISBN 978-3-642-12424-2. Dostupné z: [http://link.springer.com/10.1007/978-3-642-12425-9\\_8](http://link.springer.com/10.1007/978-3-642-12425-9_8)

Landsat Collections. *Earth Engine Data Catalog a* [online]. [cit. 2020-03-01]. Dostupné z: <https://developers.google.com/earth-engine/datasets/catalog/landsat>

MAXTED, Nigel a kol. Towards a definition of a crop wild relative. *Biodiversity and Conservation* [online]. 2006, 15(8), 2673-2685 [cit. 2020-02-17]. DOI: 10.1007/s10531-005-5409-6. ISSN 0960-3115. Dostupné z: <http://link.springer.com/10.1007/s10531-005-5409-6>

MCCUNE, Bruce a KEON Dylan. Equations for potential annual direct incident radiation and heat load. *Journal of Vegetation Science* [online]. 2002, 13(4), 603-606 [cit. 2020-03-21]. DOI: 10.1111/j.1654-1103.2002.tb02087.x. ISSN 1100-9233. Dostupné z: <https://onlinelibrary.wiley.com/doi/abs/10.1111/j.1654-1103.2002.tb02087.x>

MELOUN, Milan. *Vícerozměrná analýza dat metodou hlavních komponent a shluků* [online]. Univerzita Pardubice, 2014 [cit. 2020-03-21]. Dostupné z: [https://www.researchgate.net/publication/325217736\\_314\\_Vicerozmerna\\_analyza\\_dat\\_metodou\\_hlavnich\\_komponent\\_a\\_shluku](https://www.researchgate.net/publication/325217736_314_Vicerozmerna_analyza_dat_metodou_hlavnich_komponent_a_shluku)

MOD13Q1.006 Terra Vegetation Indices 16-Day Global 250m. *Earth Engine Data Catalog d* [online]. [cit. 2020-03-01]. Dostupné z: [https://developers.google.com/earth-engine/datasets/catalog/MODIS\\_006\\_MOD13Q1](https://developers.google.com/earth-engine/datasets/catalog/MODIS_006_MOD13Q1)

MOD13Q1 v006. USGS c [online]. [cit. 2020-03-01]. Dostupné z: <https://lpdaac.usgs.gov/products/mod13q1v006/>

MRKVIČKA, Daniel. *Clustering mraku bodů*. Brno, 2017. Bakalářská práce. Vysoké učení technické v Brně.

MURRAY, Peel. *Updated Köppen-Geiger climate map of the world* [online]. The University of Melbourne, 2012 [cit. 2020-02-17]. Dostupné z: <https://people.eng.unimelb.edu.au/mpeel/koppen.html>

OER services. *Physical Geography* [online]. 2020 [cit. 2020-02-17]. Dostupné z: <https://courses.lumenlearning.com/suny-geophysical/chapter/climate-zones-and-biomes/>

PLEKHANOVA, Elena a kol. Genomic and phenotypic analysis of Vavilov's historic landraces reveals the impact of environment and genomic islands of agronomic traits. *Scientific Reports* [online]. 2017, 7(1) [cit. 2020-02-17]. DOI: 10.1038/s41598-017-05087-5. ISSN 2045-2322. Dostupné z: <http://www.nature.com/articles/s41598-017-05087-5>

PROBERT, R. J. The role of temperature in the regulation of seed dormancy and germination. FENNER, M., ed. *Seeds: the ecology of regeneration in plant communities* [online]. Wallingford: CABI, 2000, s. 261-292 [cit. 2020-02-17]. DOI: 10.1079/9780851994321.0261. ISBN 9780851994321. Dostupné z: <http://www.cabi.org/cabebooks/ebook/20083076657>

ROBERTS, D. W a S. V COOPER. Concepts and techniques of vegetation mapping. *Applications for Resource Management. USDA Forest Service GTR INT-257*, Ogden, UT, pp 90-96. In *Land Classifications Based on Vegetation*, 1989.

ROŽNOVSKÝ, Jaroslav a kol. *Sucho na území ČR a jeho dopady: Český hydrometeorologický ústav* [online]. 2012 [cit. 2020-03-01]. Dostupné z: [http://portal.chmi.cz/files/portal/docs/katastrofy/26zasedani/Roznovsky\\_sucho\\_230412.pdf](http://portal.chmi.cz/files/portal/docs/katastrofy/26zasedani/Roznovsky_sucho_230412.pdf)

Sentinel Collections. *Earth Engine Data Catalog* b [online]. [cit. 2020-03-01]. Dostupné z: <https://developers.google.com/earth-engine/datasets/catalog/sentinel>

Shluková analýza: Metoda k-means. *Výukové materiály – Igor Ivan* [online]. 2016 [cit. 2020-03-21]. Dostupné z: [http://spatial.vsb.cz/?page\\_id=2193](http://spatial.vsb.cz/?page_id=2193)

SMÝKAL, Petr a kol. Genetic structure of wild pea (*Pisum sativum* subsp. *elatius*) populations in the northern part of the Fertile Crescent reflects moderate cross-pollination and strong effect of geographic but not environmental distance. *PLOS ONE* [online]. 2018, 13(3) [cit. 2020-02-17]. DOI: 10.1371/journal.pone.0194056. ISSN 1932-6203. Dostupné z: <http://dx.plos.org/10.1371/journal.pone.0194056>

SMÝKAL, Petr a kol. Genomic diversity and macroecology of the crop wild relatives of domesticated pea. *Scientific Reports* [online]. 2017, 7(1) [cit. 2020-02-17]. DOI: 10.1038/s41598-017-17623-4. ISSN 2045-2322. Dostupné z: <http://www.nature.com/articles/s41598-017-17623-4>

TERRACLIMATE. *Climatology Lab* [online]. [cit. 2020-03-01]. Dostupné z: <http://www.climatologylab.org/terraclimate.html>

TerraClimate: Monthly Climate and Climatic Water Balance for Global Terrestrial Surfaces, University of Idaho. *Earth Engine Data Catalog e* [online]. [cit. 2020-03-01]. Dostupné z: [https://developers.google.com/earth-engine/datasets/catalog/IDAHO\\_EPSCOR\\_TERRACLIMATE](https://developers.google.com/earth-engine/datasets/catalog/IDAHO_EPSCOR_TERRACLIMATE)

TONHAUSEROVÁ, Zuzana. *Metoda hlavních komponent a její aplikace*. Olomouc, 2013. Diplomová práce. Přírodovědecká fakulta Univerzita Palackého.

USGS EROS Archive - Digital Elevation - Shuttle Radar Topography Mission Water Body Dataset. *USGS a* [online]. [cit. 2020-03-01]. Dostupné z: [https://www.usgs.gov/centers/eros/science/usgs-eros-archive-digital-elevation-shuttle-radar-topography-mission-water-body?qt-science\\_center\\_objects=0#qt-science\\_center\\_objects](https://www.usgs.gov/centers/eros/science/usgs-eros-archive-digital-elevation-shuttle-radar-topography-mission-water-body?qt-science_center_objects=0#qt-science_center_objects)

USGS EROS Archive - Digital Elevation - Shuttle Radar Topography Mission (SRTM) 1 Arc-Second Global. *USGS b* [online]. [cit. 2020-03-01]. Dostupné z: [https://www.usgs.gov/centers/eros/science/usgs-eros-archive-digital-elevation-shuttle-radar-topography-mission-srtm-1-arc?qt-science\\_center\\_objects=0#qt-science\\_center\\_objects](https://www.usgs.gov/centers/eros/science/usgs-eros-archive-digital-elevation-shuttle-radar-topography-mission-srtm-1-arc?qt-science_center_objects=0#qt-science_center_objects)

VAVILOV, NI (1926) Studies on the origin of cultivated plants. *Bulletin of Applied Botany and Plant Breeding* 14: 1–245

VINCENT, Holly a kol. A prioritized crop wild relative inventory to help underpin global food security. *Biological Conservation* [online]. 2013, 167, 265-275 [cit. 2020-02-17]. DOI: 10.1016/j.biocon.2013.08.011. ISSN 00063207. Dostupné z: <https://linkinghub.elsevier.com/retrieve/pii/S0006320713002851>

VON, WETTBERG a kol. Ecology and genomics of an important crop wild relative as a prelude to agricultural innovation. *Nature Communications* [online]. 2018, 9(1) [cit. 2020-02-17]. DOI: 10.1038/s41467-018-02867-z. ISSN 2041-1723. Dostupné z: <http://www.nature.com/articles/s41467-018-02867-z>

WELLER, J. L. a kol. A conserved molecular basis for photoperiod adaptation in two temperate legumes. *Proceedings of the National Academy of Sciences* [online]. 2012, 109(51), 21158-21163 [cit. 2020-02-17]. DOI: 10.1073/pnas.1207943110. ISSN 0027-8424. Dostupné z: <http://www.pnas.org/cgi/doi/10.1073/pnas.1207943110>

WorldClim Version2. *WorldClim - Global Climate Data* [online]. [cit. 2020-03-01]. Dostupné z: <http://worldclim.org/version2>

ZAIR, Wathek a kol. Setting conservation priorities for crop wild relatives in the Fertile Crescent. *Genetic Resources and Crop Evolution* [online]. 2018, 65(3), 855-863 [cit. 2020-02-17]. DOI: 10.1007/s10722-017-0576-3. ISSN 0925-9864. Dostupné z: <http://link.springer.com/10.1007/s10722-017-0576-3>

ZEDER, Melinda A. The Origins of Agriculture in the Near East. *Current Anthropology* [online]. 2011, 52(S4), S221-S235 [cit. 2020-02-17]. DOI: 10.1086/659307. ISSN 0011-3204. Dostupné z: <https://www.journals.uchicago.edu/doi/10.1086/659307>

# **PŘÍLOHY**

# SEZNAM PŘÍLOH

## Vázané přílohy:

- Příloha 1 Vzorový skript 1 pro extrahování hodnot bodům rostlin z více snímků
- Příloha 2 Vzorový skript 2 pro extrahování hodnot bodům rostlin z jednoho snímku
- Příloha 3 Korelační matice – 41 environmentálních faktorů pro data cizrny (*Cicer reticulatum*)
- Příloha 4 Korelační matice – 41 environmentálních faktorů pro data cizrny (*Cicer echinospermum*)
- Příloha 5 Korelační matice – 41 environmentálních faktorů pro data čočky
- Příloha 6 Korelační matice – 41 environmentálních faktorů pro data hrachu

## Volné přílohy

- Příloha 7 Poster
- Příloha 8 DVD

## Struktura DVD

Adresáře:

- poster
- text\_prace
- data
- web

## Příloha 1

### Vzorový skript 1 pro extrahování hodnot bodům rostlin z více snímků.

```
// vložení vstupních dat (body s rostlinami)
var inputData = ee.FeatureCollection("users/user/CICER_final_GEE");

// vložení datasetu, se kterým se bude pracovat
var dataset = ee.ImageCollection('MODIS/006/MOD13Q1');

// vybraný dataset obsahuje několik tzv. bands, ze kterých se může vybírat.
// Zde jsou např. spočítané indexy NDVI a EVI. Tady se vybere NDVI.
var selectBand = dataset.select('NDVI');

// znovu se vybere band, z něhož je třeba získat hodnoty, odkazuje se na
// proměnnou selectBand (ve které je jeden band → za get musí být 0)
var bandName = ee.Image(selectBand.mean()).bandNames().get(0);
print ("Selected Band",bandName);

// nastavení počátečního data
var startDate = ee.Date('2000-2-18');

// nastavení koncového data
var endDate = ee.Date('2020-1-31');

// spočítá počet měsíců ke zpracování
var countMonths = ee.Number(endDate.difference(startDate,'month')).round();
print ("Count of months",countMonths);

// získá seznam (list) měsíců ve formátu YYYYMMDD, které jsou použity do
// později vytvořeného slovníku jako názvy sloupců v tabulce
var monthList = ee.List(ee.List.sequence(0,countMonths).map(function (n){
  return startDate.advance(n,'month').format('YYYYMMdd');
}));
print("Months",monthList);

// počítá se výsledek, do kterého vstupují vstupní data
var result = inputData.map(function(feature){
  // počítá se pro každý měsíc
  var timeSeries = ee.List.sequence(0,countMonths).map(function (n){
    // zvolení počátečního data do výpočtu ze startDate
    var ini = startDate.advance(n,'month');
    // počítá se po jednom měsíci
    var end = ini.advance(1,'month');
    // filtruje se časové rozpětí na základě předchozího nastavení, pokud je
    // více hodnot v jenom měsíci, tak se zprůměrují
    // reduceRegion je příkaz pro přidělení hodnot ze snímku do vstupních dat
    // (pokud je více snímků, tak se zprůměrují)
    var data = selectBand.filterDate(ini,end).mean().reduceRegion({
      reducer: ee.Reducer.mean(),
      geometry: feature.geometry(),
      scale: 30
    });
  });
});
```

```

// získá hodnoty na základě zvoleného Band v bandName, (value,value,"") -
// "" pro null hodnoty je prázdný záznam
var value = ee.Number(data.get(bandName));
value = ee.Number(ee.Algorithms.If(value,value,""));
// vrátí průměr (zonal mean) pro zvolené měřítko (scale: 30)
return value;
});
// vytvoří se nový slovník, do kterého je vložen seznam měsíců s hodnotami
var timeDictionary =
ee.Dictionary.fromLists(monthList,ee.List(timeSeries));
// vrátí prvky s hodnotami snímku pro každý měsíc ze slovníku
return feature.set(timeDictionary);
});

// vytiskne pro ověření výsledek
print("Result",result);

// exportuje data se získanými hodnotami do nově vytvořené tabulky (CSV),
// která se uloží na Google Drive
Export.table.toDrive({
  collection:result,
  description:"MOD13Q1_006_NDVI_2000_2_18_2020_1_31_CICER",
  fileFormat:"CSV",
})

```



## Příloha 2

### Vzorový skript 2 pro extrahování hodnot bodům rostlin z jednoho snímku.

```
// vložení vstupních dat (body s rostlinami)
var inputData = ee.FeatureCollection("users/user/CICER_final_GEE");

// vložení snímku, se kterým se bude pracovat
var dataset = ee.Image('OpenLandMap/SOL/SOL_ORGANIC-CARBON_USDA-6A1C_M/v02');

// vybraný dataset obsahuje několik tzv. bands, ze kterých se může vybírat.
// Zde se vybere hloubka 0-60 ('b0','b10','b30','b60')
var selectBand = dataset.select('b0','b10','b30','b60');

// reduceRegion je příkaz pro přidělení hodnot ze snímku do vstupních dat
// (pokud je více snímků, tak se zprůměrují), 30 znamená měřítko
var output_table = selectBand.reduceRegions(inputData, ee.Reducer.mean(),30);

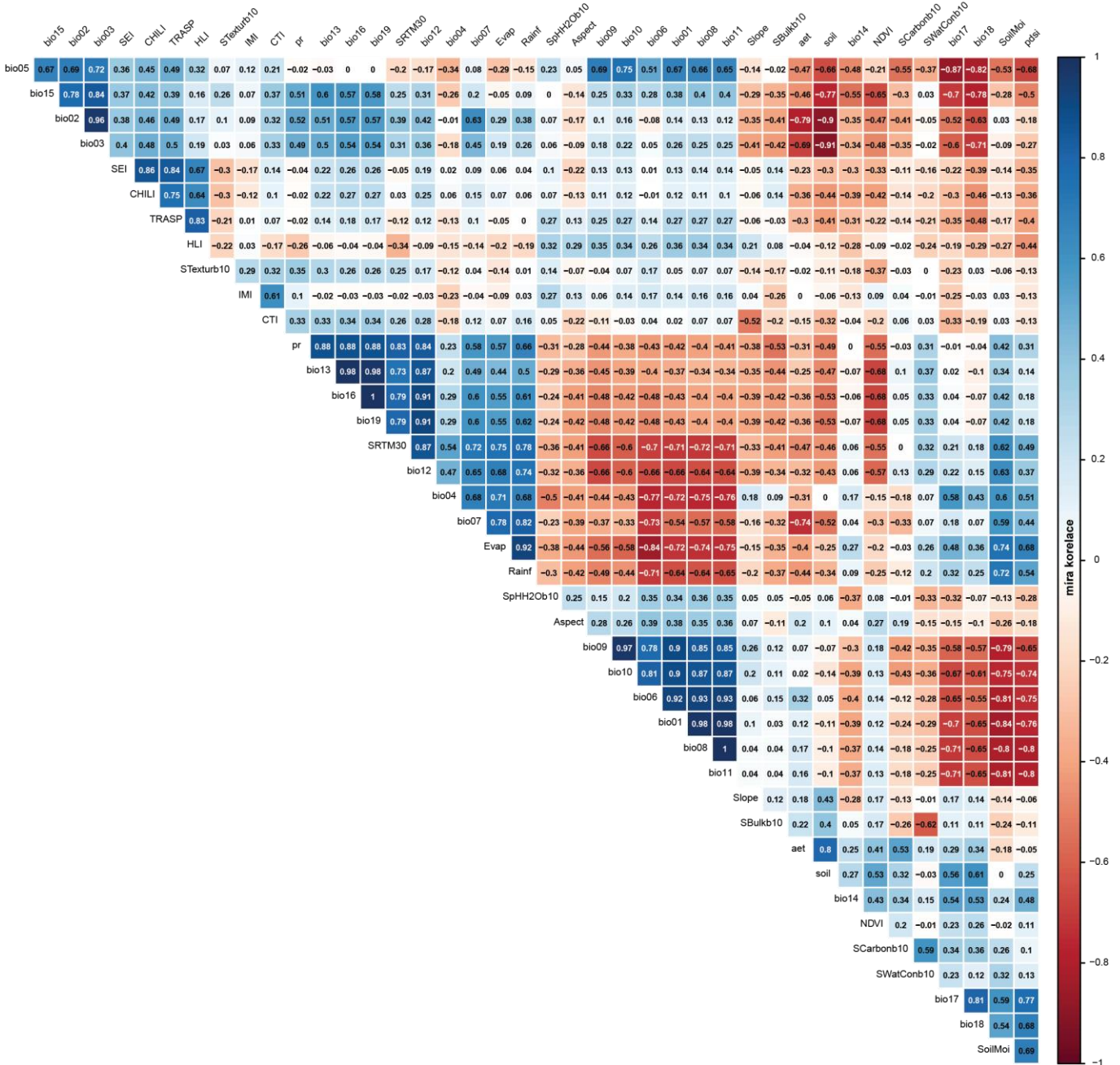
// vypíše počet prvků v tabulce (jen informativně)
var count = output_table.size();
print("tabulka",count);

// exportuje data se získanými hodnotami do nově vytvořené tabulky (CSV),
// která se uloží na Google Drive
Export.table.toDrive(output_table,"MOD13Q1_006_NDVI_");
```



# Příloha 4

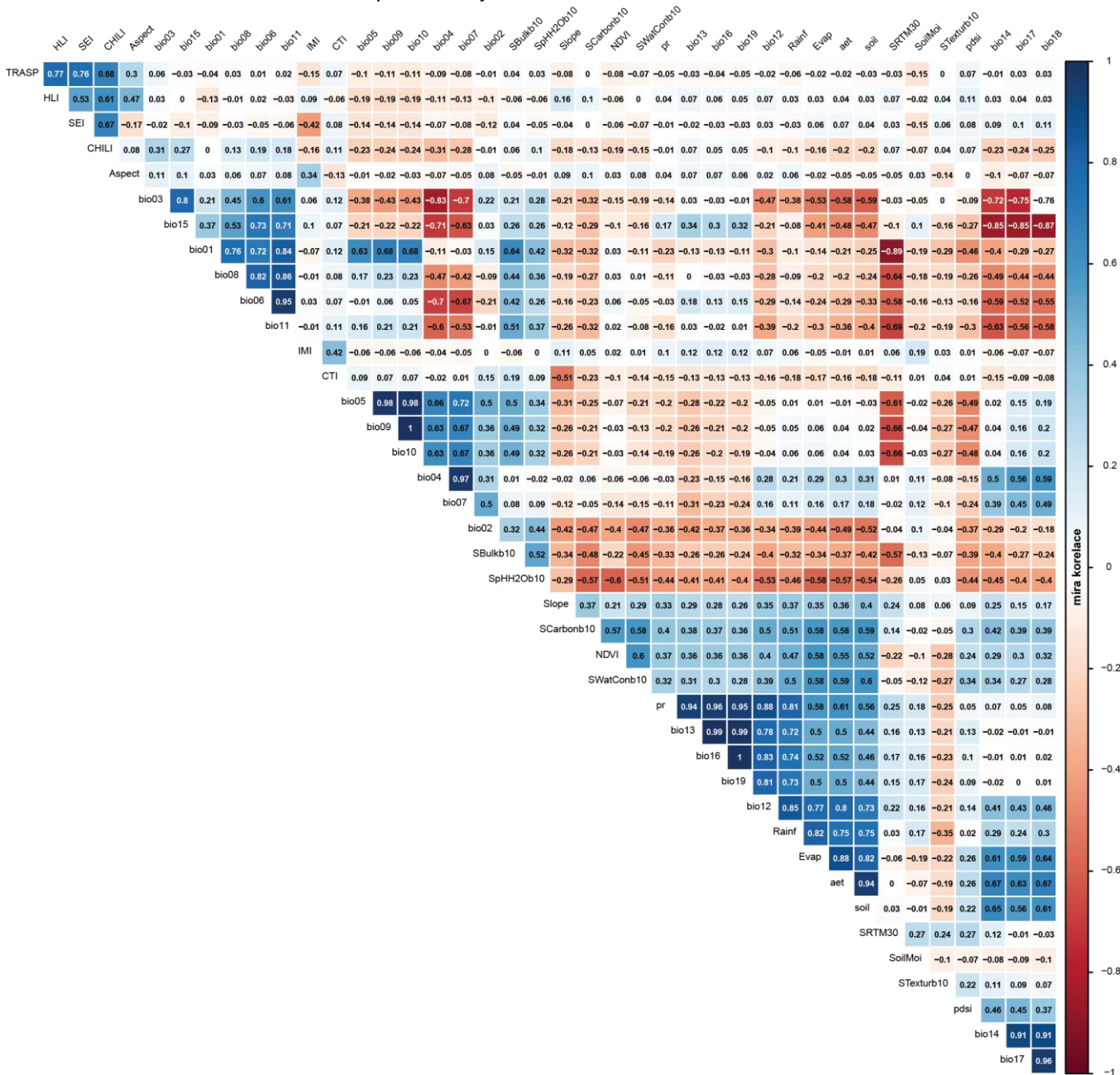
Korelační matice – 41 environmentálních faktorů pro data cizrná (*Cicer echinospermum*)





# Příloha 5

Korelační matice – 41 environmentálních faktorů pro data čocky



# Příloha 6

Korelační matice – 41 environmentálních faktorů pro data hrachu

