

# Aplikace pro automatizované zpracování dat DPZ a vyhodnocení hydrologických poměrů

## Uživatelská dokumentace

Aplikace zahrnuje celkem pět modulů, jejichž běh na sebe logicky navazuje:

- 1. downloader** slouží pro automatizované stahování dat. Pracuje se Sentinel-2, další datové zdroje lze snadno přidat.
- 2. indexer** slouží pro výpočet indexů (lineárních kombinací, případně poměrů pásem), které následně mohou vstupovat do klasifikace, resp. analýzy časových řad. Sada indexů je předdefinována, další indexy lze velmi snadno přidat.
- 3. classifier** slouží pro řízenou klasifikaci snímků s možností zadání celé řady parametrů, zejména automatickému opakování klasifikace a generalizace výsledku. Použitelné metody: Maximum Likelihood, Random Trees, Support Vector Machine.
- 4. ts\_analyzer** (ts = time series) slouží pro analýzu vývoje území v čase. Pracuje s vektorovými i rastrovými daty, s bodovou informací o změně či s celou plochou. Produkuje informace určené k dalšímu snadnému uživatelskému zpracování a náhledy grafů.
- 5. LC2hydro** určí vodní bilanci lokality po jednotlivých typech krajinného pokryvu, Výpočet: založeno na dlouhodobých simulacích jednotlivých složek hydrologického cyklu (výpar, odtok) na předem definovaných plochách (listnatý/jehličnatý les, louka, pastvina, holina apod.). Při změně krajinného pokryvu pak lze usuzovat na dlouhodobé změny ve vodním režimu. Lze využít i na what-if analýzy, s tím, že lze odhadovat budoucí stav hydrologie například na základě analýzy hyperspektrálních dat. V místě, kde v analýze dat DPZ vyjde les poškozený, v následné what-if analýze budou tato místa zadána jako bezlesí.

**pro správné fungování je potřeba:** OS Windows, ArcGIS Pro verze 3.2 a novější, Python s běžnými moduly plus arcpy, kvalitní připojení k internetu v případě modulu 1

**technologie:** sada skriptů pro příkazovou řádku. Parametry jsou popsány a vysvětleny níže. Příklady praktického využití viz též v manuálu.

## Downloader

Stáhne scény Sentinel-2 Level 2-A ze zájmového území podle parametrů zadaných uživatelem, vybere a převzorkuje relevantní pásma pro další výpočty a výsledek ořízne na zájmové území.

Syntaxe volání:

```
01_downloader.py [-h] -area AREA -start START -end END -max_cloud
                  MAX_CLOUD -out_dir OUT_DIR -work_dir WORK_DIR -user
                  USER -password PASSWORD
```

-h: vypíše nápovědu k syntaxi skriptu

-area: zájmové území. Čtveřice čísel oddělených čárkou bez mezer, po řadě zeměpisná délka levého dolního rohu zájmového území, zeměpisná šířka levého dolního rohu, zeměpisná délka pravého horního rohu, zeměpisná šířka pravého horního rohu  
-start: počátek časového období, ve formátu YYYYMMDD  
-end: konec časového období, ve formátu YYYYMMDD. Budou staženy všechny snímky spadající do období mezi start a end  
-max\_cloud: číslo od 0 do 100, maximální přípustné pokrytí scény oblačností  
-out\_dir: složka, kam bude uložen výstup  
-work\_dir: složka, kam se budou ukládat dočasné soubory  
-user: uživatelské jméno pro DataHub, viz poznámky níže  
-password: heslo po DataHub, viz poznámky níže

#### Příklad použití:

```
01_downloader.py -area "15.6247100,50.6852986,15.7927664,50.7487722" -start 20240801 -end 20240831 -max_cloud 20 -out_dir S:\images -work_dir S:\temp -user myuser -password mypass
```

#### Poznámky:

- pro použití modulu je nutná registrace na <https://documentation.dataspace.copernicus.eu/Registration.html>, uživatelské jméno a heslo je nutné zadat jako parametr skriptu
- stahování dat může trvat v závislosti na parametrech dotazu a kvalitě síťového připojení i několik hodin, obvyklá velikost scény jsou vyšší stovky MB
- je nutné mít dostatek volného místa na disku, bez něj může skript spadnout
- je nutné respektovat politiku Copernicus Data Space Ecosystem týkající se množství stahovaných scén
- skript používá k vyhledávání a stahování scén OData API <https://documentation.dataspace.copernicus.eu/APIs/OData.html>, původně zamýšlený modul sentinelsat (uvažovaný v zadání projektu) přestal být v 11/2023 podporovaný, viz <https://scihub.copernicus.eu/> a <https://github.com/sentinelsat/sentinelsat>
- ve výsledném souboru (\*\_clip.tif) jsou všechna relevantní pásma převzorkována na nejvyšší rozlišení (tj. 10 m), a pojmenována Band + ('\_B01\_20m', '\_B02\_10m', '\_B03\_10m', '\_B04\_10m', '\_B05\_20m', '\_B06\_20m', '\_B07\_20m', '\_B08\_10m', '\_B8A\_20m', '\_B09\_60m', '\_B11\_20m', '\_B12\_20m')

## Indexer

Pro zadaný rastr, který je (a musí být) výstupem procesu Downloader, spočítá jeden či více vegetačních indexů dle výběru uživatele a uloží je jako samostatné soubory do zadané složky.

#### Syntaxe volání:

```
02_indexer.py [-h] -i INPUT_FILE -idx_def INDEX_DEFINITION -idx INDICES -out_dir OUTPUT_DIRECTORY
```

- h: vypíše nápovědu k syntaxi skriptu
- i: vstupní rastr

-idx\_def: soubor s definicí indexů. Jeho syntaxi viz dále v poznámkách  
-idx: seznam indexů oddělených čárkami bez mezer. Tyto indexy se spočítají. Jejich výpočet musí být definován v souboru zadaném jako argument -idx\_def  
-out\_dir: složka, kam se uloží výstup. Výstupní soubory jsou pojmenovány stejně jako vstup s přidanou koncovkou `_{název-indexu}.tif`

Příklad použití:

```
02_indexer.py -i
s:\temp\S2A_MSIL2A_20230907T100031_N0509_R122_T33UWS_20230907T173654.SAFE_c
lip.tif -idx_def veg_index_sentinel2.txt -idx NDVI,EVI,GNDVI,NDRE -out_dir
S:\indices
```

Poznámky:

- skript podporuje definování indexu uživatelem bez nutnosti programování
- součástí odevzdání je ukázkový soubor `veg_index_sentinel2.txt`, jehož součástí je definice 12 nejběžnějších indexů
- definice každého indexu se skládá z řádku, kde je:
  - název indexu
  - středník
  - vzorec pro výpočet
- vzorec pro výpočet může obsahovat libovolné výrazy používané v ArcGIS v rastrové algebře (viz <https://pro.arcgis.com/en/pro-app/latest/arcpy/spatial-analyst/an-overview-of-the-map-algebra-operators.htm>), s tím, že proměnná reprezentující každé pásmo je označena `b01`, `b02`, ... `b12` a navíc je k dispozici `b8a`
- prázdné řádky a řádky začínající `#` se v souboru s definicemi indexů ignorují, lze je použít pro přehlednější rozčlenění a komentáře
- k vyhodnocení se používá funkce `eval` s bezpečnostní omezením na proměnné `b01`, `b02`, ... `b12` a `b8a`

## Classifier

Pro zadaný vstupní rastr a referenční data provede sérii řízených klasifikací, opraví jejich výsledek a tyto produkty zkombinuje a uloží do výsledného rastru. Zadáním celé řady parametrů má uživatel možnost podrobně ovlivnit průběh celého procesu.

Syntaxe volání:

```
03_classifier.py -i input_file -out output_file -work work_dir -tv
training_validation_data [-ratio tv_ratio] [-aprob in_prob] [-special
special_classes] [-repeat repeat_classification] [-min_patch pixels] [-
simpl repeat_simpl] [-select select_for_result]
```

kde

- `input_file` je vstupní rastr,
- `output_file` je výstupní rastr s výsledkem klasifikace,
- `work_dir` je složka pro uložení dočasných souborů,
- `training_validation_data` jsou trénovací a validační data (shapefile),
- `tv_ratio` je poměr, jakým se `training_validation_data` dělí na trénovací a validační, trénovacích dat je `tv_ratio * celkový počet záznamů pro danou třídu v training_validation_data`; číslo z intervalu (0, 1), nepovinné, výchozí hodnota 0,33;

- `in_prob` je seznam apriorních pravděpodobností výskytu tříd v podobě pythonovského slovníku, kde klíč je název třídy a hodnota je pravděpodobnost, součet hodnot ve slovníku musí být 1, názvy tříd musí odpovídat atributu `Classname` v `training_validation_data`; nepovinné, v případě, že není zadáno, proces zjistí sám s využitím všech trénovacích a validačních dat metodou MLC,
- `special_classes` je seznam tříd, se kterými je při stratifikovaném výběru třeba zacházet speciálně, protože se vyskytují vzácně, jejich zastoupení v trénovacích datech bude tedy mírně nadhodnoceno, aby výsledky byly relevantnější,
- `repeat_classification` je počet klasifikací, které se mají provést, číslo, nepovinné, výchozí hodnota 15, střídají se po řadě metody MLC, SVM a RandomTrees (s `ntrees=100`), vždy s náhodně stratifikovaně vybranými daty podle parametrů `training_validation_data` a `tv_ratio`,
- `pixels` je minimální velikost výsledné plošky v pixelech, menší plošky jsou agregovány se sousedními, číslo, nepovinné, výchozí hodnota 5,
- `repeat_simpl` je maximální počet iterací, kterými se eliminují malé plošky, číslo, nepovinné, výchozí hodnota 4; zvýšení hodnot velmi prodlužuje výpočet,
- `select_for_result` je počet klasifikací, které se mají zkombinovat do finálního výsledku, vybírá se pro danou třídu vždy `select_for_result` klasifikací s nejlepším F1 skóre; nepovinné, výchozí hodnota 5, musí být menší než `repeat_classification`

Příklad použití i se všemi nepovinnými parametry:

```
03_classifier.py -i S:\classification\2021_07_13_BL_MSSRGB_3cm_envi.tif -
out S:\classification\2021_07_13_BL_MSSRGB_3cm_envi_classified.tif -work
S:\classification -tv S:\classification\tv_data_pnt_aligned.shp -ratio 0.33
-aprob "{ 'afs': 0.09212, 'cv': 0.17971, 'cxbig': 0.02820, 'desch': 0.11637,
'mol': 0.17491, 'nard': 0.40632, 'smrk': 0.00238}" -special
"cxbig,smrk,afs" -repeat 30 -min_patch 10 -simpl 4 -select 10
```

Poznámky:

- `training_validation_data` je shapefile minimálně s atributy `classname` a `classvalue` (není case sensitive). Pokud je vrstva polygonová, převede se na body zarovnané se středy pixelů vstupního rastru. Pokud je vrstva bodová, použije se tak, jak je a je zodpovědností uživatele si zarovnání zajistit.
- pro trénování se použije podíl určený `tv_ratio` pro každou třídu bez ohledu na apriorní pravděpodobnosti. Ty se uplatní až při tvorbě validačních dat, kde výběr dat pro validaci (a tím i celkový výsledek klasifikace ve smyslu OA) respektuje pravděpodobnost jejich výskytu. Nasbíraná data tedy nepřicházejí při trénování “na zmar” a zároveň jejich nasbírání v poměru výrazně jiném než je relativní zastoupení jednotlivých tříd neovlivní celkový výsledek klasifikace.
- ve `work_dir` vzniknou jako mezivýsledek soubory `result_C.tif` pro každé C obsažené jako hodnota `classname` v `training_validation_data`. Hodnota v pixelech těchto rastrů pak představuje pravděpodobnosti výskytu jednotlivých tříd (zprůměrované f1 skóre). Výsledný rastr je tvořen třídou, která má největší hodnotu. Tyto rastry slouží jako zobecněná chybová matice, protože umožní uživateli pro každý pixel zjistit pravděpodobnosti výskytu každé třídy v něm.
- v závislosti na volbě parametrů (zejména počet opakování klasifikací a generalizace) může běh skriptu trvat i velmi dlouho (jednotky až desítky hodin).

# TS Analyzer

Slouží k analýze vývoje dat v čase. Uvažují se následující scénáře použití:

- změna v ploše mezi n-časovými obdobími → proces `ts_analyzer_area`
- změna v místě (bodě) mezi n-časovými obdobími → proces `ts_analyzer_pnt`

Pro přehlednost je každý ze scénářů realizován samostatným procesem, popsáním samostatně dále.

## ts\_analyzer\_area

Pomocí překrytí série vektorových dat vytvoří informace vhodné pro analýzu změn v pohodlné uživatelské podobě. Zahrnuje též možnost generalizace výsledku s inteligentní eliminací štěpin.

Syntaxe volání:

```
04_ts_analyzer_area.py [-h] -i INPUT_FILE -og OUTPUT_GEODATABASE -of
                        OUTPUT_FEATURE_CLASS [-size SIZE_LIMIT]
                        [-iters MAX_ITERS] [-sep SEPARATOR]
```

kde

- `INPUT_FILE` je existující soubor se zadáním vstupních vektorových vrstev. Textový soubor, na každém řádku název vektorové vrstvy včetně cesty, za ním znak středník, za ním název atributu, kde je v dané vrstvě uložena klasifikace jako číselná hodnota. Předpokládá se, že číselník hodnot je pro všechna vektorová data stejný. Pokud ne, je potřeba před spuštěním procesu údaje sjednotit.
- `OUTPUT_GEODATABASE` je ESRI file geodatabáze pro uložení výsledku,
- `OUTPUT_FEATURE_CLASS` je feature class pro uložení výsledku. Data mohou být značně velká, proto se povinně ukládají do geodatabáze a pomocí dvojice parametrů je uživatel donucen jí zadat.
- `SIZE_LIMIT` je minimální velikost mapovací jednotky (v jednotkách souřadnicového systému), číslo, nepovinné, výchozí hodnota 5,
- `MAX_ITERS` je maximální počet iterací pro eliminaci podměrečných objektů, číslo, nepovinné, výchozí hodnota 10, vyšší hodnoty mohou prodlužovat dobu běhu,
- `SEPARATOR` je oddělovač ve sloupci `CHANGE` ve výsledné feature class. Využití oddělovače umožňuje elegantním způsobem provádět (s využitím SQL) uživatelsky definované statistické a analytické dotazy, viz též poznámky. Znak, nepovinné, výchozí hodnota `-`.

Příklad použití:

```
04_ts_analyzer_area.py -i S:\changes\classification_list.txt -og
S:\changes\result.gdb -of result -size 5 -iters 4 -sep :
```

Poznámky:

- pro smysluplné použití se musí všechny vektorové vrstvy překrývat, resp. výsledek je definovaný jen v místech, kde se překrývají všechny

- neslouží pro rastrová data; pro ta je určen následující nástroj. Chcete-li použít, převedte nejprve rastrová data na polygon s využitím funkce Raster to Polygon
- příklad využití oddělovače. Předpokládejme 4 časové řezy, kódy tříd 1-9, oddělovač - . Ve sloupci CHANGE pak vzniknou záznamy typu 1-1-1-1. Pomocí SQL příkazů LIKE (ze zástupnými znaky % a ?) a IN lze pak velmi snadno vybrat například stabilní plochy či určitý typ změny a s výběrem pak dále pracovat.

## ts\_analyzer\_pnt

Extrahuje ze série libovolných rastrů informace o vývoji v čase na zadaném místě a uloží výsledek jako data k dalšímu zpracování, volitelně též jako graf.

Syntaxe volání:

```
04_ts_analyzer_pnt.py [-h] -i INPUT_FILE -c COORDS -o OUTPUT_DATA
                        [-g OUTPUT_CHART]
```

kde

- INPUT\_FILE je soubor se zadáním vstupních rastrů. Textový soubor, na každém řádku název rastru včetně cesty, za ním znak hvězdička, a číslo pásma (číslováno od 1)
- COORDS jsou zeměpisné souřadnice zadané jako lat,lon, tedy oddělené čárkou, bez mezery mezi souřadnicemi
- OUTPUT\_DATA je výsledný csv soubor s hodnotami rastrů, který se vytvoří. Výsledná data lze dále zpracovávat v libovolném softwaru
- OUTPUT\_CHART je výsledný soubor ve formátu PDF, kam bude uložen graf ukazující vývoj v daném místě. Parametr je nepovinný a graf je potřeba chápat jako rychlý a potenciálně vizuálně nedokonalý náhled vývoje v daném místě. Pro tvorbu vizuálně zdařilejšího grafu slouží primárně data v OUTPUT\_DATA

Příklad použití:

```
04_ts_analyzer_pnt.py -i S:\change\list.txt -c 15.7073686,50.7200598 -o
S:\change\output_data.csv -g S:\change\output_chart.pdf
```

Poznámky:

- předpokládá se, že se všechny rastry v zadaném bodě prostorově překrývají
- lze použít mimo jiné pro stažená data sentinelu, z nich spočítané indexy či výsledky klasifikací (tedy výstupy z předchozích procesů)
- neslouží pro vektorová data; pro ta je určen předchozí nástroj. Chcete-li použít, převedte vektorová data na rastr s využitím funkce Polygon to Raster

# LC2hydro

Pro danou klasifikaci krajinného pokryvu určí dlouhodobou vodní bilanci území. Klasifikace krajinného pokryvu je založena na výše prezentovaných skriptech a vodní bilance je odvozena pomocí hydrologického modelu SWAT. Model SWAT rozděluje modelované území na tzv. hydrotopy, tj. území se stejnou kombinací krajinného pokryvu, půdy a sklonu svahu. Následně je vypočten průměr jednotlivých složek vodní bilance vážený plochou daných hydrotopů za definované dílčí povodí (na povodí horní Úpy bylo celkem 490 hydrotopů a v povodí Čisté 147 hydrotopů), které vždy ústí do určitého segmentu koryta vodního toku. V korytě je odtok hydrologicky transformován na výsledný odtok v uzávěrovém profilu.

Dlouhodobá simulace vychází z kalibrace modelu v letech 1993-1997 a celková délka simulace zahrnuje roky 1991-2021. Za toto období byla stanovena vodní bilance pro všechny hydrotopy v povodí Horní Úpy a Černé.

Pro každý typ klasifikovaného krajinného pokryvu byla vyčíslena množství výparu (AET) a odtoku (Q), kdy odtok byl rozdělen na přímý (SURQ), laterální odtok půdou (LATQ) a podzemní odtok (GWQ). Analyzované kategorie krajinného pokryvu zahrnovaly louky (PAST), listnaté (FRSD), jehličnaté (FRSE) a smíšené lesy (FRST), křoviny (SHRB), kombinaci holé půdy s roztroušenou vegetací (BSVG) a území s holou půdou úplně bez vegetace (BARR). Pro tyto kategorie byly stanoveny dlouhodobé průměry vodní bilance tak, aby mohlo následně dojít k odhadu změn ve vodní bilanci při změně zastoupení jednotlivých kategorií land cover.

Změna ve vodní bilanci celého území může být odhadnuta ze změny plošného zastoupení jednotlivých kategorií krajinného pokryvu pomocí změny jejich podílu na celkové ploše území, která je získána z jejich klasifikace. Hodnoty jednotlivých složek vodní bilance území jsou uvedeny v tabulce níže (viz Tab. 1). Podíl celkové plochy jednotlivých krajinných pokryvů na celkové ploše zkoumaného území ( $a_i$ ) tvoří zásadní vstup pro odhad celkové vodní bilance na základě aditivního vztahu, který vychází z principu fungování modelu SWAT, kdy v požadovaném časovém kroku dochází ke sčítání jednotlivých složek vodní bilance získaných kaskádovitě za unikátní kombinace krajinného pokryvu, půdy a topografických charakteristik. Aplikovaný postup vychází z následujících rovnic:

$$AET = \sum_{i=1}^7 a_i AET_i$$
$$SURQ = \sum_{i=1}^7 a_i SURQ_i$$
$$LATQ = \sum_{i=1}^7 a_i LATQ_i$$
$$GWQ = \sum_{i=1}^7 a_i GWQ_i$$

kde  $i$  reprezentuje jednotlivé kategorie land coveru a  $AET_i$ ,  $SURQ_i$ ,  $LATQ_i$  a  $GWQ_i$  reprezentují dlouhodobé hodnoty vodní bilance získané na základě panujících půdních, topografických a klimatických podmínek.

Pozornost lze zaměřit i separátně na vlhké a teplé periody, kdy se jen změní průměrné hodnoty jednotlivých složek vodní bilance tak, aby korespondovaly s požadovanou variantou vývoje. Hodnoty používané pro suché a vlhké periody byly stanoveny na základě průměrů let 2000-2002 pro vlhké období (viz Tab. 2) a 2014-2016 pro suché období (viz Tab. 3).

Tabulka 1: Průměrné hodnoty jednotlivých složek vodní bilance získané simulací modelu SWAT v období 1991-2021 (mm.rok<sup>-1</sup>)

	AET	SURQ	LATQ	GWQ	Q
BARR	314	632	138	180	949
BSVG	327	289	350	316	955
FRSD	335	241	295	477	1013
FRSE	321	260	252	438	950
FRST	327	197	221	554	972
PAST	329	246	413	299	958
SHRB	327	259	327	358	943

Tabulka 2: Průměrné hodnoty jednotlivých složek vodní bilance získané simulací modelu SWAT v období 2000-2003 (mm.rok<sup>-1</sup>)

	AET	SURQ	LATQ	GWQ	Q
BARR	193	198	829	154	1181
BSVG	203	327	495	365	1187
FRSD	210	274	426	558	1258
FRSE	198	300	363	518	1180
FRST	203	229	321	662	1211
PAST	204	275	575	343	1194
SHRB	203	294	463	418	1174

Tabulka 3: Průměrné hodnoty jednotlivých složek vodní bilance získané simulací modelu SWAT v období 2014-2015 (mm.rok<sup>-1</sup>)

	AET	SURQ	LATQ	GWQ	Q
BARR	200	143	408	114	665
BSVG	212	225	203	243	671



FRSD	218	188	165	364	717
FRSE	207	198	142	327	667
FRST	210	151	118	416	686
PAST	214	193	247	233	673
SHRB	212	200	189	274	662

Poznámky:

- vyčíslení vodní bilance neobsahuje informaci o rozložení odtoku v čase, protože není známé budoucí rozložení ani výše srážkových úhrnů
- výsledky lze využít zejména pro odhad rozdělení odtoku na jednotlivé komponenty – rychlý přímý odtok (SURQ), zpomalený laterální odtok (LATQ) a dotaci podzemní vody, která následně tvoří podzemní odtok (GWQ)
- velmi podobné hodnoty výparu (AET) jsou dány vysokou nadmořskou výškou zkoumaného území, a tudíž velmi nízkou potenciální evapotranspirací, která neumožňuje projevení výraznějších rozdílů mezi krajinnými pokryvy

Uživatelská dokumentace k výsledku V5 Software „Aplikace pro automatizované zpracování dat DPZ a vyhodnocení hydrologických poměrů“ projektu TAČR SS05010124 „Hodnocení vlivu změn krajinného pokryvu na lokální hydrologii a klima v Krkonošském národním parku s využitím dálkového průzkumu Země a hydrologického modelování“. Web projektu: [www.lucc4hydro.cz](http://www.lucc4hydro.cz). Hlavní řešitel projektu: Lucie Kupková (PřF UK). Autoři výsledku V5: Jakub Lysák, Katedra aplikované geoinformatiky a kartografie, Přírodovědecká fakulta Univerzity Karlovy a Václav Šípek, Ústav pro hydrodynamiku AV ČR, v. i. Praha, prosinec 2024.