

Shrnutí výsledků

27 Jan
28 2025



CIIRC CTU
Prague, CZ

GeoAI
CAMP PRG 2025

Karel Charvát

v.z. zpravodajů
podle příslušných
pracovních skupin

Innovate and shape the future of Geospatial AI
through joint development of GeoAI solutions.



Plan4all



FACULTY
OF INFORMATION
TECHNOLOGY
CTU IN PRAGUE



CZECH INSTITUTE
OF INFORMATICS
ROBOTICS AND
CYBERNETICS
CTU IN PRAGUE



RICAIP

Supported by:



EDIH CTU



Co-funded by
the European Union



Funded by
the European Union
NextGenerationEU



CZECH
RECOVERY PLAN

PODPORA PŘI VÝBĚRU A ANALÝZE DOKUMENTŮ DOSTUPNÝCH V HUBU, KTERÉ MAJÍ POMOCI PŘI PROVÁDĚNÍ TEMATICKÉ ANALÝZY

Úkol se skládá ze dvou na sebe navazujících fází:

Indexování a výběr dokumentů

Využití pokročilých technik indexování ke zpracování a klasifikaci dokumentů v rámci centra. Využití algoritmů sémantického vyhledávání a řazení podle relevance k identifikaci a upřednostnění dokumentů na základě předem definovaných témat a dotazů specifických pro uživatele.

Analýza dokumentů a optimalizace dotazů

Analyzujte vybrané dokumenty pomocí nejmodernějších modelů zpracování přirozeného jazyka (NLP), abyste získali užitečné poznatky. Dále zdokonalte schopnosti generování odpovědí rozsáhlých jazykových modelů (LLM), abyste zlepšili jejich kontextové porozumění a přesnost při řešení uživatelských dotazů.

VÝZVY

lokalizace / jazyk -> problém s parsováním a vkládáním pdf souborů

pokud výsledky nejsou optimální -> nasazení metody řazení

délka chunkingu -> sémantická vs. pevná délka

čištění textu po parsování -> odstranění artefaktů metodou založenou na pravidlech vs. čištění LLM

PROCESY

parsování vstupních dat do prostého textu / markdown / json

chunking - rozdělení textu na menší fragmenty

vkládání kousků do vektorů

ukládání vektorů do vektorové databáze - ChromaDB

VÝSLEDKY

datový zdroj, ke kterému má JackDaw přístup:

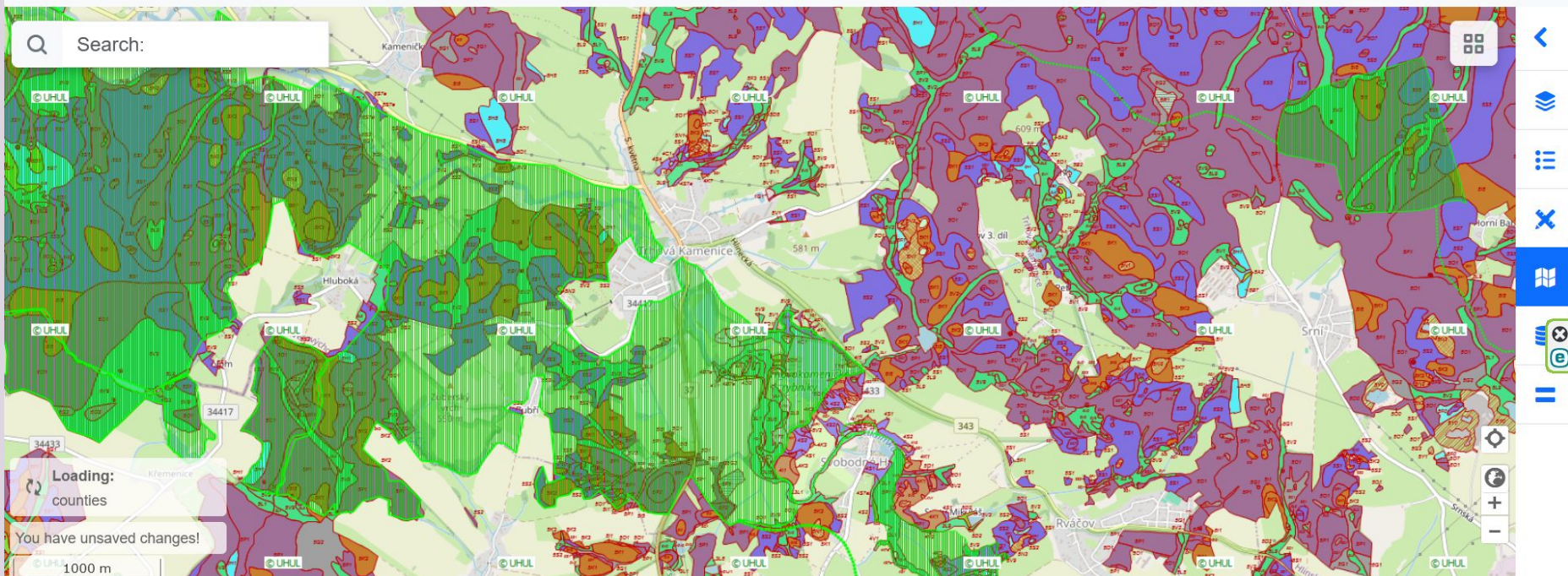
zlepšit znalosti

snížení pravděpodobnosti halucinací

ON-LINE INTERPRETACE SLUŽEB OGC WMS

- Nastavení čtení dostupných dat z instance služby Catalog Services for the Web (CSW), například MICKA. Takový požadavek vrátí dostupné zdroje dat:
[MICKA_instance_URL]/csw?request=getrecords&typenames=records
- Aby z něj bylo možné získat důležité metainformace, jako např:
Název, Abstrakt, Geografický rozsah (geo entita nebo bbox), Kvalita dat, Související datové sady (nadřazené, podřazené atd.), Časový rozsah.
- Použijte službu GetStyle WMS k získání stylu SLD pro vrstvu zájmu, abyste ji mohli použít v našich nástrojích AI. SLD obvykle obsahuje název třídy s jejím barevným kódovým znázorněním v mapě.
- Sestavte sémantické grafy tříd objektů z různých dostupných zdrojů dat: Inspire Registry, jako např:
 - <https://inspire.ec.europa.eu/codelist/HILUCSValue/> ,
 - WikiData: <https://www.wikidata.org/wiki/Q515905>,
 - OpenStreetMap: <https://www.openstreetmap.org/relation/435514>
- Propojte tyto grafy, částečně napůl ručně, částečně automaticky, pokud již existují vazby mezi objekty.

- Zahrňte do kódu funkci, která vypočítá požadovanou velikost obrázku, aby odpovídal potřebnému měřítku v dané projekci.
- Zahrňte funkci pro maskování údajů pro dané zeměpisné oblasti (statistické jednotky, místa, přírodní parky atd.).
- Vytvořte obecnější nástroj pro práci s příslušnými službami WMS: typickým úkolem je výpočet podílu různých prvků v zájmové oblasti (tj. kolik je v oblasti lesa? Jaký je podíl jehličnatých stromů v dané oblasti? Jaký byl podíl jehličnatých stromů v této oblasti před 10 lety? Jaký je podíl jezer v dané oblasti? V podstatě vytvořit nástroj, který bude schopen automaticky pracovat se všemi těmito otázkami, vybrat vhodný zdroj, automaticky získat vhodná data, správně je interpretovat pomocí požadavku `getstyles`.



VSTUPY

- Katalog CSW
- Instance Mapserveru
- Různé zdroje sémantických informací:
- WikiData, OpenStreetMap, registry Inspire atd.

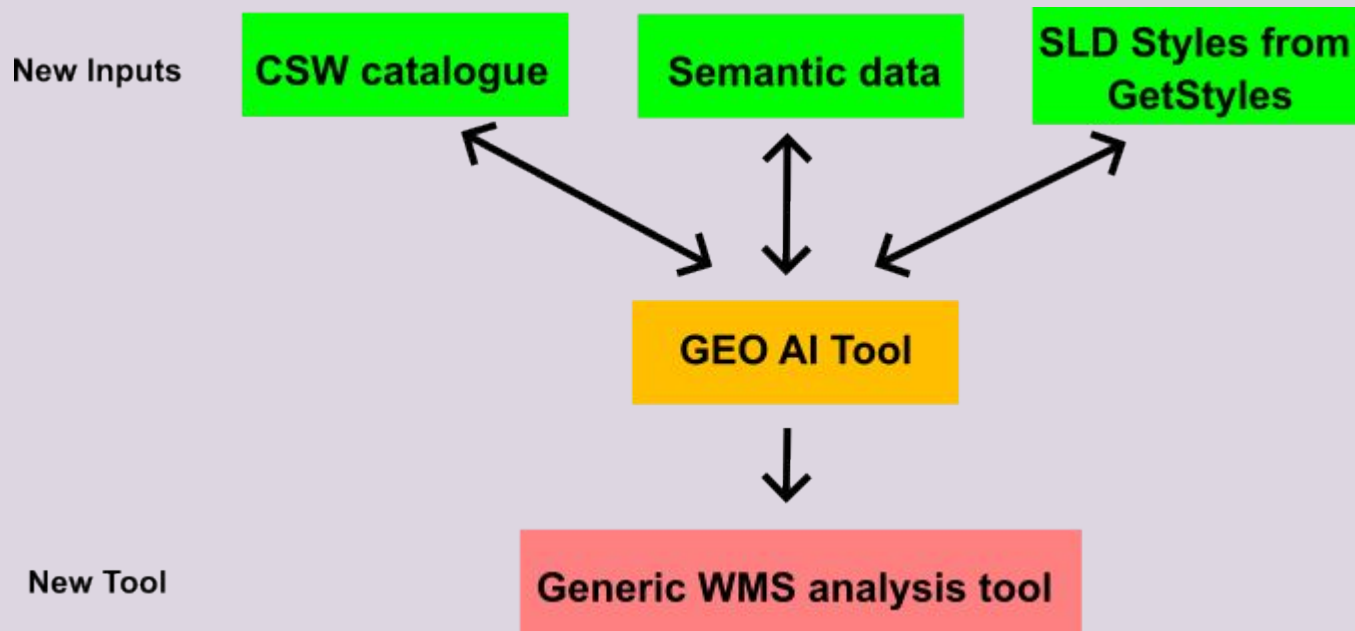
VÝZVY

- Kódování čtení a předzpracování dostupných dat z katalogu CSW
- Vytvořit grafy entit ze sémantických registrů.
- Propojit koncepty z grafu s třídami v souboru SLD (Style Layer Descriptor) vráceném požadavkem GetStyles.
- Vytvořit obecný nástroj pro zpracování a analýzu obrazových dat vrácených službami WMS

PROCESY

- Čtení dat z katalogu CSW
- Čtení sémantických dat
- Propojení záznamů z katalogu CSW se sémantickými daty
- Propojit data s výzvou uživatele
- Stahování stylů a tříd
- Propojit třídy se sémantickými daty
- Zpřesnění výzvy
- V případě WMS použít nový generický nástroj pro zpracování požadavku

VÝSLEDKY



NÁVRH A TESTOVÁNÍ ROZHRAŇÍ API PRO METADATA PRO INTEGRACI NOVÝCH ANALÝZ DO SYSTÉMU JACKDAW

Cílem je definovat standardizovaný přístup k popisu nových prostorových a neprostorových analýz, který umožní bezproblémovou integraci jejich API do systému JackDaw. To zahrnuje návrh a testování schématu metadat, aby bylo zajištěno, že analýzy jsou vhodně charakterizovány příslušnými atributy, včetně vstupních požadavků, možností zpracování a očekávaných výstupů.

Současná implementace systému JackDaw umožňuje vybírat z předem definovaných analýz a datových sad, identifikovat ty nejrelevantnější a zahrnout poznatky z těchto analýz do odpovědí generovaných rozsáhlými jazykovými modely (LLM). Rozšíření této funkce o podporu dynamické integrace nových analýz zvýší adaptabilitu a škálovatelnost systému a zajistí jeho schopnost efektivně využívat různé analytické pracovní postupy a datové sady.



Magpie

Metadata Aggregation and Geospatial Processing Intelligence Engine



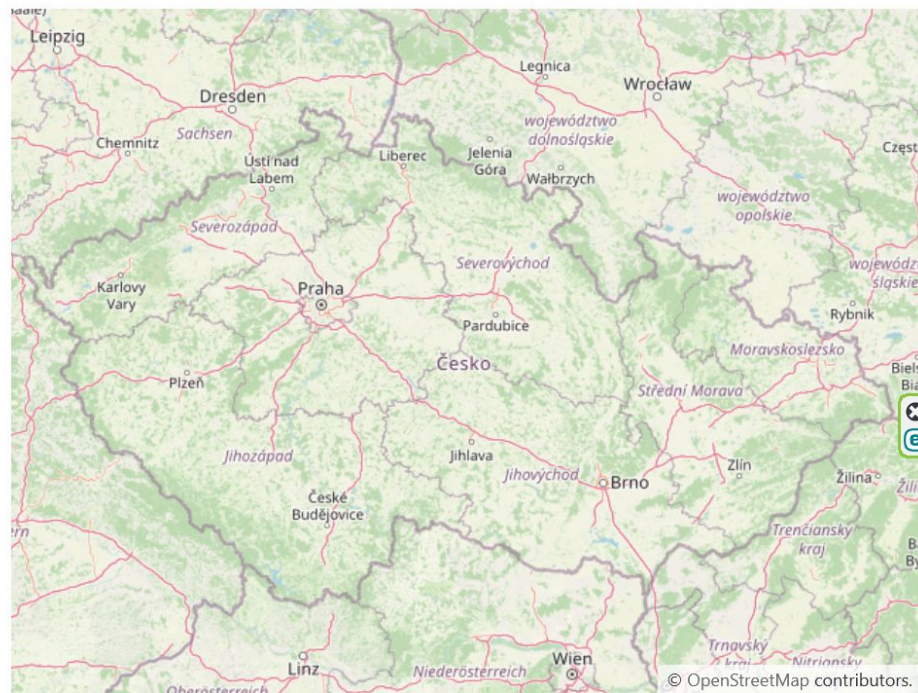
How can I help you?

I can find map information for you in the
Czech metadata catalog

Jaka data mohou byt uzitecna pro majitele lesa



[Sample 1](#) [Sample 2](#)





Jaka data mohou byt uzitecna pro majitele lesa



[Sample 1](#) [Sample 2](#)

INSPIRE prohlížečící služba pro téma Využití území - Královéhradecký kraj, RozvOsy_p

RozvOsy_p

wms

+ Add to map

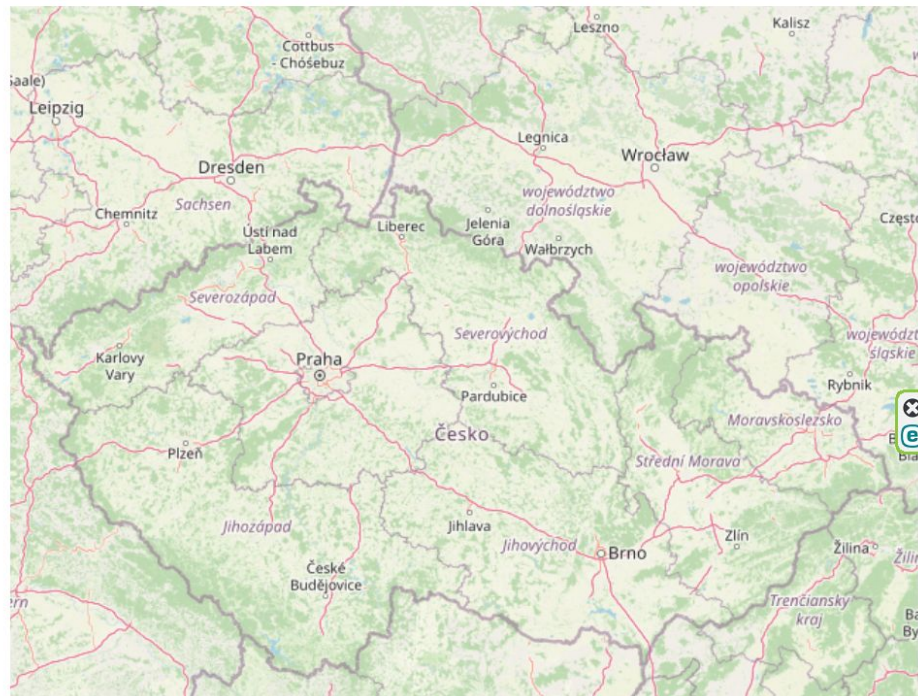
INSPIRE prohlížečící služba pro téma Využití území - Královéhradecký kraj, služby v oblasti rozvodu elektřiny, plynu a tepelné energie

služby v oblasti rozvodu elektřiny, plynu a tepelné energie

wms

+ Add to map

INSPIRE prohlížečící služba pro téma Využití území - Královéhradecký



VSTUPY

- stávající metadatové standardy pro katalogy (prostorové i neprostorové):
 - DCAT - katalog metadat pro všeobecné účely
 - GeoDCAT - geoprostorové rozšíření
 - DQV - rozšíření kvality
 - STAC - prostorově-časové katalogy aktiv
 - Záznamy OGC
 - V souladu s ISO 19115 (CSW)
 - Záznamy OGC jsou v souladu s ISO
 - STAC je v souladu se záznamy OGC
 - DCAT a STAC:
- <https://docs.google.com/spreadsheets/d/1XgYuHoYygcvxatltG1FqN8cNKINOCWFMSUkxtSxAFoo/edit?pli=1&gid=0#gid=0>

VSTUPY

- Vyvinuté nástroje pro agenta LLM, každý nástroj má specifický zdroj dat
 - ohraničení oblasti zájmu vložené jako argument do každého nástroje
 - některé nástroje mají argumenty generované agentem LLM
- Nástroje/koncové body
 - Sémantický tripplestore
 - Příklady API
 - EUROSTAT
<https://grlc-dpi-enabler-demeter.apps.paas-dev.psnc.pl/api-git/ad4gd/statistics>
 - SOIL (lucas)
<https://grlc-dpi-enabler-demeter.apps.paas-dev.psnc.pl/api-git/glosis-ld/api>
 - Rozhraní
 - <https://www.foodie-cloud.org/sparql>
 - <https://www.foodie-cloud.org/fct>
 - <http://metaphactory.foodie-cloud.org/>
 - zpracování rastrových obrazů ze služeb WMS (mapserver)
 - Land use / Land cover
 - Stažení png obrázku
 - zjišťování zón na základě barev (pomocí legendy barev k typům zón / typům půdy atd.)
 - SPOI – koncový bod WFS (mapserver)
 - Rozbor názvů a kategorií SPOI z JSON
 - Analýza počasí koncový bod WMS (mapserver) (mapserver)
 - pomocí vygenerovaných argumentů určujících časový rámeček (měsíc, rok)
 - stahování tiff obrázku
 - souhrnné výpočty (průměrná teplota)
 - Předpověď počasí / nowcasting – openmeteo API
 - argument pro požadovaný časový rámeček
 - parsování teplotních dat z JSON

VÝZVY

- přístup ke správným a vhodným údajům
- dohodnout se na
 - základní metadata (povinná), která jackdaw potřebuje k výběru zdroje dat a/nebo datové sady ve zdroji dat.
 - název, popis, prostorové, časové, přístupové informace (parametry, ověření).
 - u datových zdrojů také informace o tom, jak filtrovat/přístupovat.
 - nepovinná metadata, ze kterých může jackdaw těžit
 - (standardní) sledované/měřené proměnné/vlastnosti, informace týkající se kvality atd.
- výběr relevantních nástrojů a mapování na typy datových sad, aby bylo možné přístupovat k datům/zpracovávat je.

PROCESY

- jackdaw potřebuje mít přístup k metadatům z různých zdrojů dat
 - jednotlivé datové sady lze přidávat např. pomocí nějakého tlačítka pro přidání, které zobrazí zkrácený formulář (relevantní?)
 - koncový bod/API k nějakému katalogu s více datovými sadami
 - jackdaw bude potřebovat metadata o katalogu a informace o tom, jak je filtrovat/dotazovat se na ně
 - LLM může vygenerovat dotaz na filtrování katalogu, aby získal pouze relevantní metadata o datasetech
- LLM se používá k výběru relevantních datasetů na základě jejich metadat
- jackdaw získá z metadat datasetu/zdroje informace o tom, jak získat aktuální data
- jackdaw získá data – pomocí určité sady příslušných nástrojů
 - standardizuje je*

graph example:

VÝSLEDKY







```
{
  "@context": [
    "https://w3id.org/iliad/oim/metadata-context.jsonld",
    "https://w3id.org/iliad/oim-context.jsonld"
  ],
  "@graph": [
    {
      "@type": "Dataset",
      "description": "Total cultivated agricultural and fallow land; Crops on arable land; Garden area; Areas under trees; Vines (grapes and raisins); Fallow land",
      "title": "Land usage in Greece",
      "distribution": {
        "@type": "Distribution",
        "accessURL": "http://example.org/endpoint",
        "format": "http://publications.europa.eu/resource/authority/file-type/CSV"
      },
      "spatial": {
        "@type": "Location",
        "bbox": "POLYGON((20.85452 41.74851, 26.60419 41.74851, 26.60419 34.91999, 23.40065 34.69139, 20.01422 36.22057, 20.01422 41.74851, 20.85452 41.74851))"
      },
      "temporal": {
        "@type": "PeriodOfTime",
        "startDate": "2022-01-01",
        "endDate": "2022-12-31"
      },
      "structure": {
        "@type": "DataStructureDefinition",
        "component": {
          "measure": [
            {
              "@type": "MeasureProperty",
              "concept": "http://w3id.org/glois/model/codelists/landUseClassValueCode",
              "unit": "http://www.qudt.org/vocab/unit/HA"
            }
          ]
        }
      }
    }
  ]
}
```

VÝZVY K AKCI

- poskytnout příklady
- provádět experimenty
- pokusit se najít možné styčné body s prací z CZ/PL

INTEGRACE ATLASU OSVĚDČENÝCH POSTUPŮ S APLIKACÍ JACKDAW

The screenshot displays the website <https://sumavaprodukt.regionalnispéciality.cz/atlas/>. The interface includes a search bar labeled "Vyhledávání", a "Kategorie" dropdown menu, and an "E-shop" toggle. The main content area features a grid of product cards, each with a representative image and a title. The map on the left shows the Sumava region with various location markers.

Image	Product Name
	Americká zahrada - Chudence
	Angus Nejdí
	Ateliér Ivety Kúsově
	Barokní lékárna U bílého
	Bejkovna - Měčin -
	Biofarma Skřivánek

VSTUPY

Integrace chatbota do aplikace

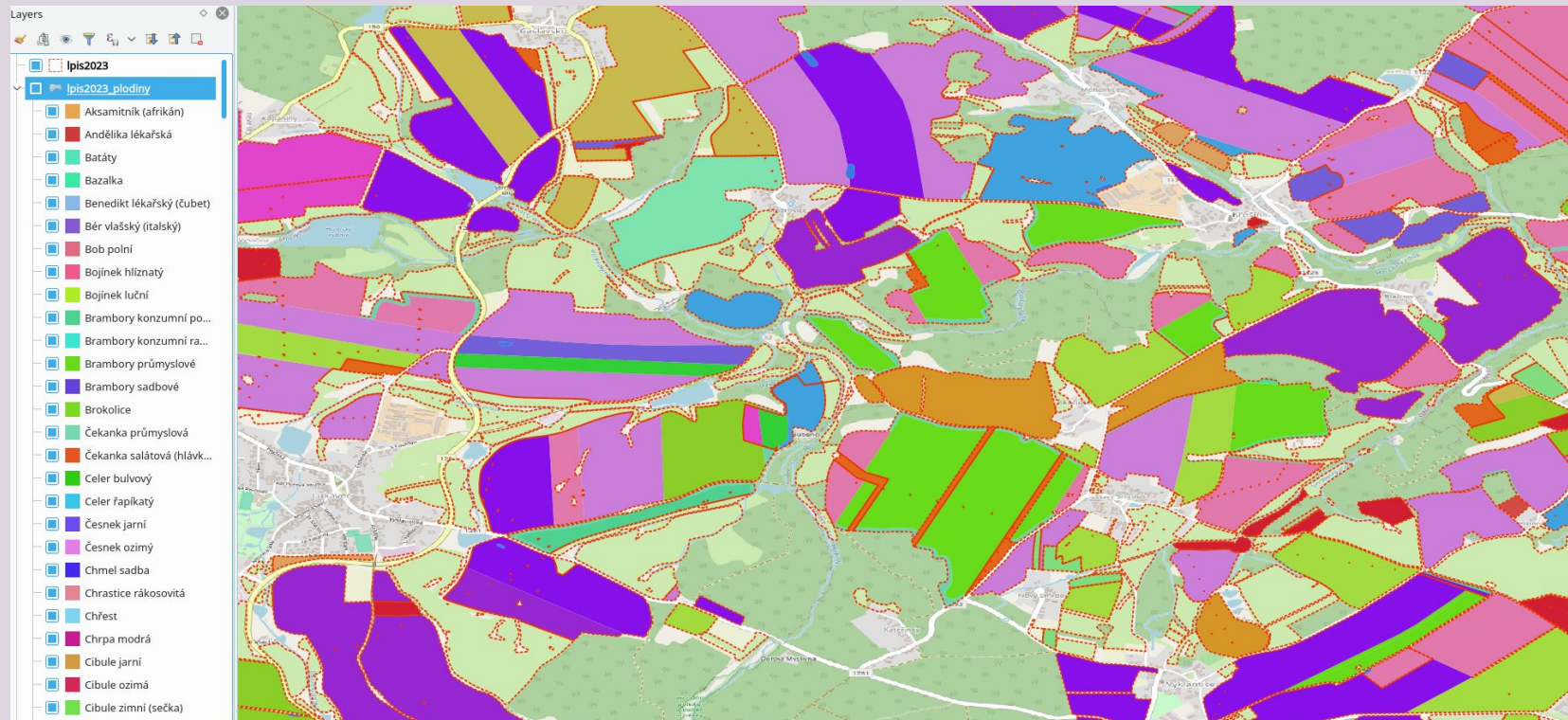
Vyhledávání dalších dat

ANALÝZA DAT LPIS PRO REGIONÁLNÍ ENVIRONMENTÁLNÍ UDRŽITELNOST A POTENCIÁL KRÁTKÉHO DODAVATELSKÉHO ŘETĚZCE

Kombinovat informace z veřejné verze XML LPIS a export kultur na jednotlivá pole a využít tyto informace při odhadu turistické atraktivity regionu.

VSTUPY

LPIS data: XML + SHP (kuřtury)



VÝZVY

Připojit a analyzovat obě dostupné datové sady LPIS (soubory XML pro katastrální území obsahující zemědělské bloky a samostatná pole).

PROCESY

Získat data a jejich načtení do databáze pomocí služeb.

Provést počáteční vizualizaci dat.

Analyzovat data.

Rozhodnout, jak je doručit do systému JackDaw.

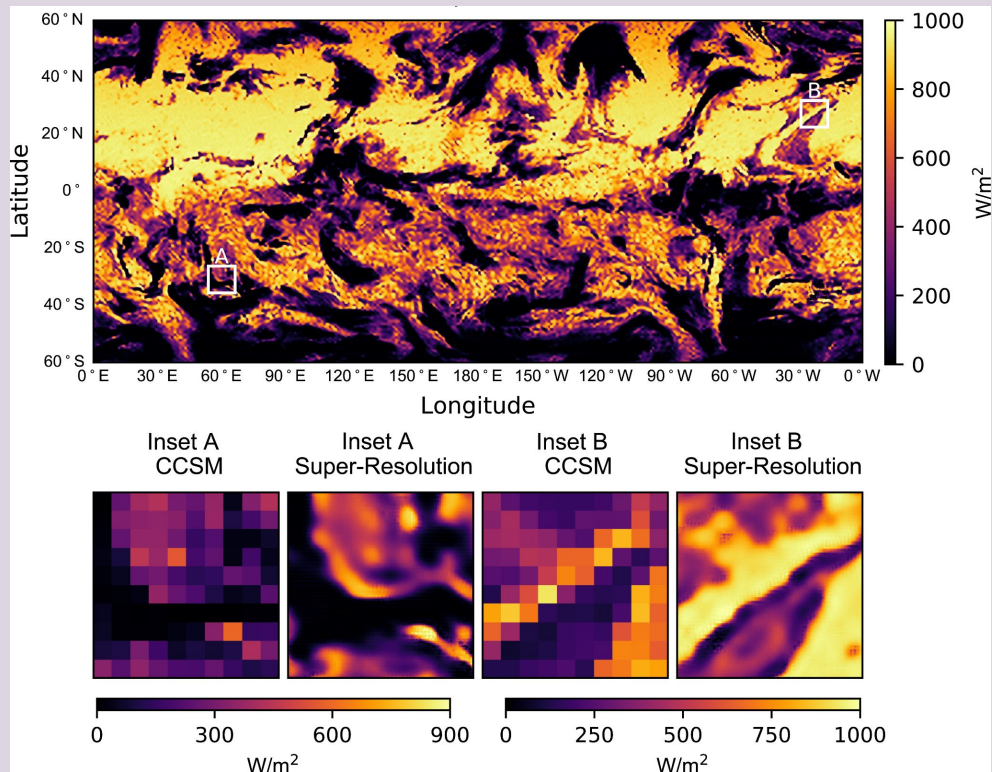
VÝSLEDKY

- Sběr dat pro rok 2024
- Zahájení analýzy důležitých atributů pro cestovní ruch

VÝZVY K AKCI

- Spojit 2 veřejné datové sady LPIS XML a export polí a kultur na nich
- Využijte dostupné atributy k odhadu turistické atraktivity regionů (tj. malá pole jsou pravděpodobně atraktivnější, levandule je vhodnější než brambory atd.).
- Vložte data do vhodného nástroje (např. JackDaw).

FOCAL: ZVÝŠENÍ ROZLIŠENÍ KLIMATICKÝCH DAT



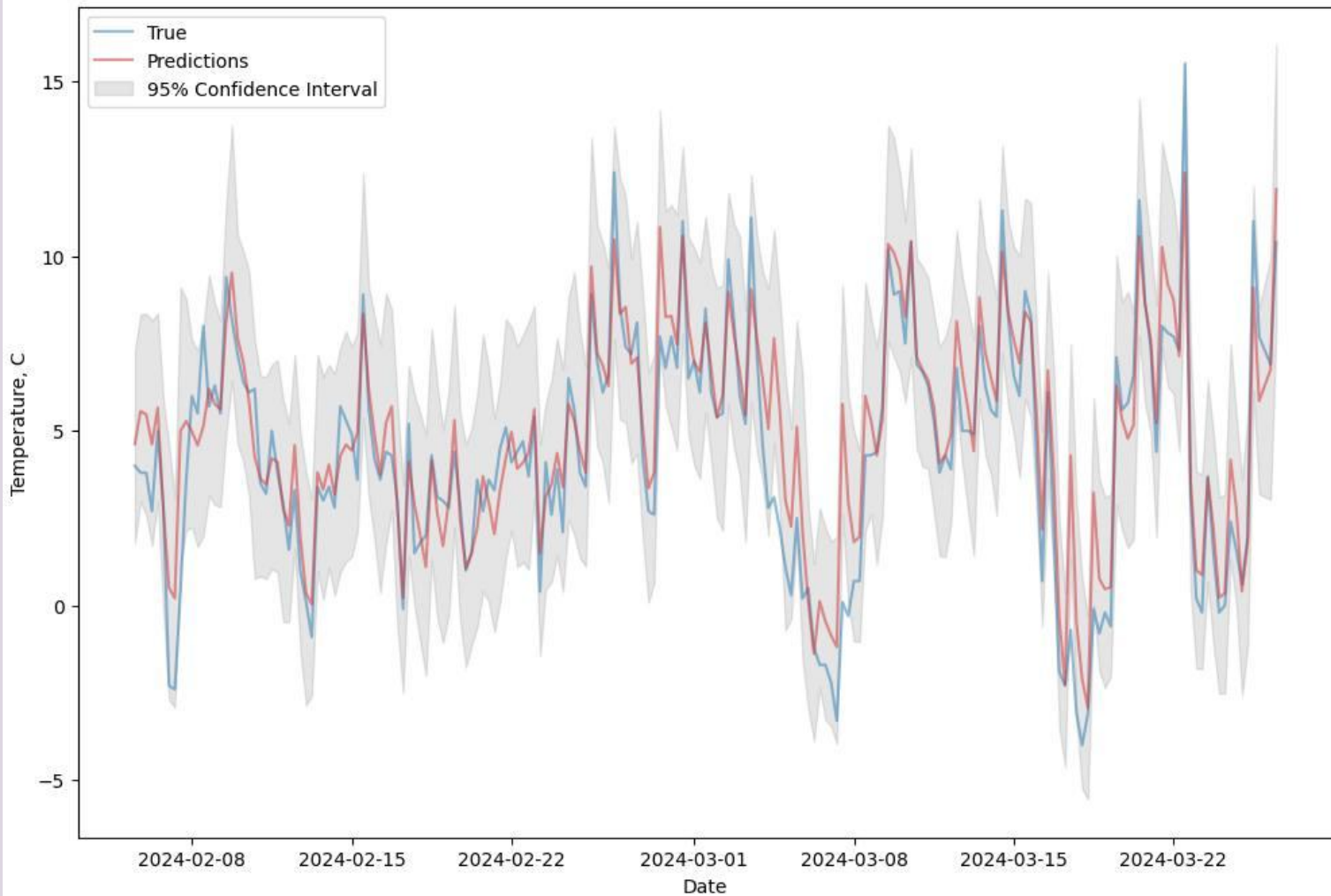
VSTUPY

- Klimatické modely
- Satelitní data z geostacionárních družic s vysokým časovým rozlišením
- Družicová data z geosynchronních družic s vysokým prostorovým rozlišením
- Data z meteostanic
- Topografické údaje
- Globální předpovědi
- Údaje ze senzorů

VÝZVY

- Vysoce kvalitní validační data!
- Získat značné množství dat z různých míst v Evropě i mimo ni.
- Superrozlišení klimatických dat se může lišit od superrozlišení snímků
- Dynamika klimatu v různých částech světa se (pravděpodobně) liší

VÝSLEDKY

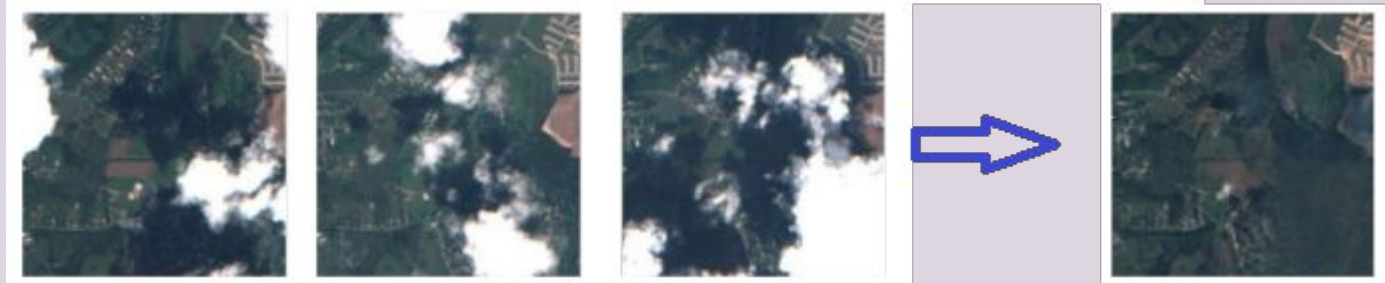


VÝZVY K AKCI

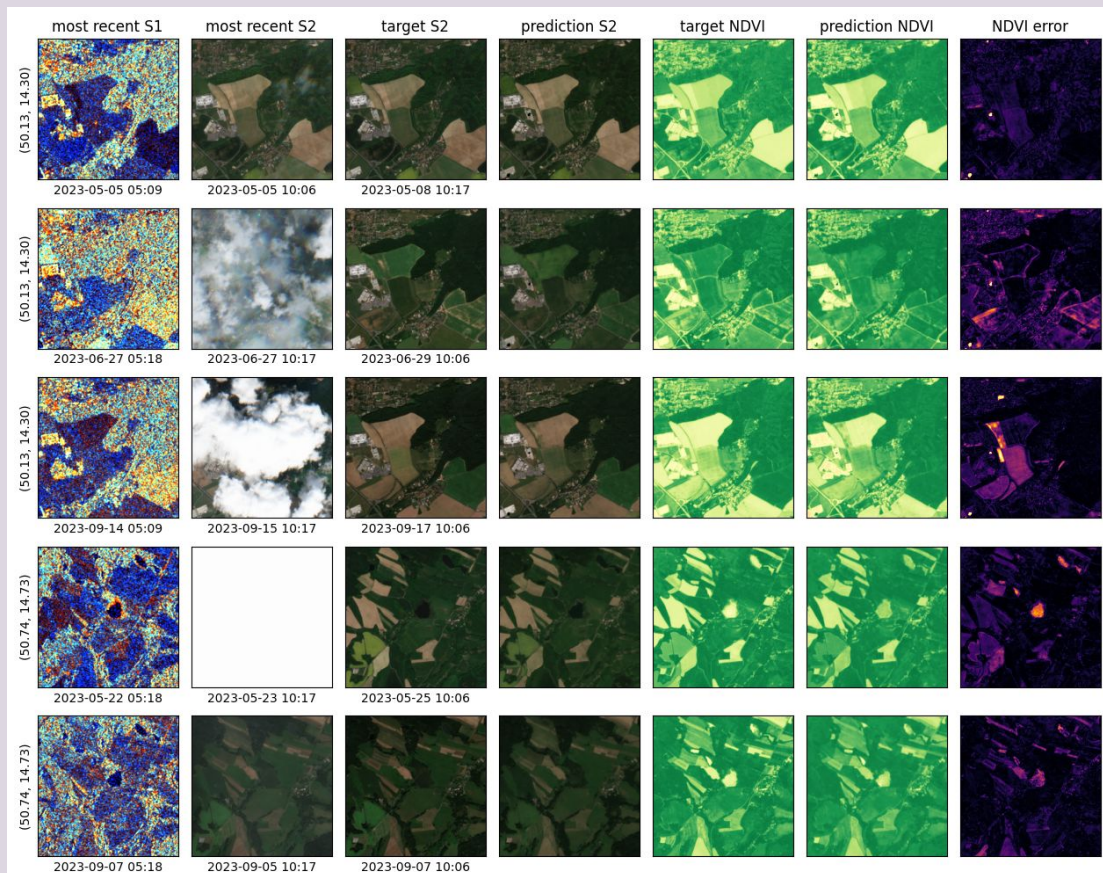
- Stahování dat z různých částí Evropy (ERA5/Landsat)
- Získávání dat ze senzorů, meteostanic atd.
- Vyhledejte potenciální uživatele pro úlohy superrozlišení
- Aplikace pro lesy/zahrady/farmy/sady.

VÝZVA PRO DRUŽICOVÉ SNÍMKY ALIANCE

- Odstranění oblačnosti předpovídáním bezoblačného snímku na základě sekvence posledních snímků
- Vstupy: Sentinel-1 a Sentinel-2



VÝZVA PRO DRUŽICOVÉ SNÍMKY ALIANCE



VÝSLEDKY

- Spolupráce se společností LESPROJEKT-slужby
- Brainstorming řešení pro nasazení umělé inteligence pro odstranění oblačnosti
- Seznámení se s rozhraním EO API pro stahování dat Sentinel

UKÁZKA FRONTENDU

The screenshot displays a web-based farm management interface. On the left is a dark blue sidebar with navigation options: MANAGEMENT (My projects, New project), CURRENT PROJECT (Fields, Sensors, Sim), and PLANNING (Place, Crop, Fertilisation). The main area is split into two panes. The left pane shows a map of 'Under Sun Farm' with a large green field area outlined in blue. The right pane is titled 'Manage' and shows a list of simulation tiles. At the top right of the right pane is a 'Jobs' section with a slider for 'SIM.MANAGE.cloudCover'. Below this is a search bar for 'SIM.MANAGE.dateRange' and a '1 item' indicator. The list of tiles includes IDs, thumbnails, and labels like 'Tile: T38TMK'. A 'SIM.MANAGE.markAsUnusable' button is visible at the bottom right of the tile list.

MANAGEMENT

- My projects
- New project

CURRENT PROJECT

- Fields
- Sensors
- Sim

PLANNING

- Place
- Crop
- Fertilisation

Farm Under Sun Farm

Manage Jobs

SIM.MANAGE.dateRange

SIM.MANAGE.cloudCover

1 item SIM.MANAGE.deselectAll

20240403 Tile: T38TMK

20240405 Tile: T38TMK

20240408 Tile: T38TMK

20240410 Tile: T38TLK

20240413 Tile: T38TLK

20240415 Tile: T38TLK

SIM.MANAGE.markAsUnusable

500 m

Logged in as: Filip

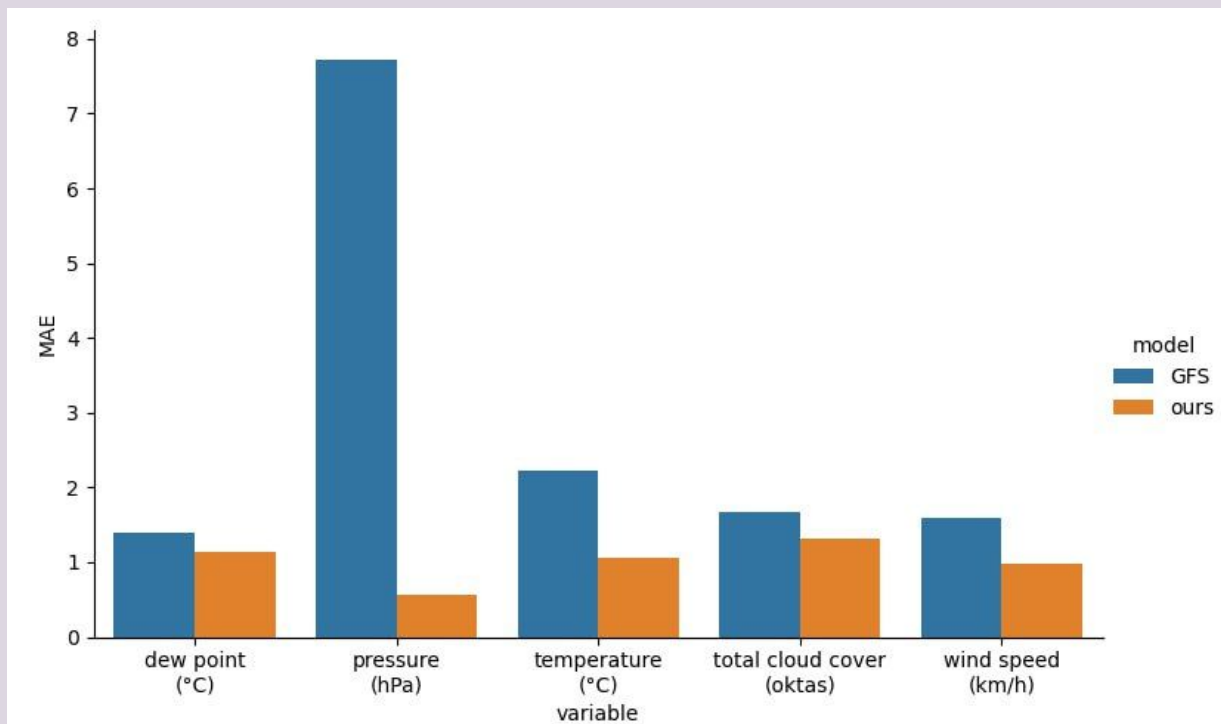
UKÁZKA FRONTENDU

The screenshot displays a web-based farm management application. On the left is a dark blue sidebar with navigation options: MANAGEMENT (My projects, New project), CURRENT PROJECT (Fields, Sensors, Sim), and PLANNING (Place, Crop, Fertilisation). The main area shows a map of a farm with green field boundaries and a 500m scale bar. The top right of the map area includes 'Farm' and 'Under Sun Farm' buttons. On the right side, a 'Manage' tab is active, showing a 'Satellite Image Jobs' list. The list includes a 'DOWNLOAD (10)' section with one 'PENDING' job and eight 'COMPLETED' jobs, and a 'TCI (15)' section with one 'EVI (1)' job. Each job entry has a status button and a link to 'info about date'.

Job Status	Action	Info
DOWNLOAD (10)		
PENDING		info about date
COMPLETED		info about date
COMPLETED		info about date
COMPLETED		info about date
IN-PROGRESS		info about date
COMPLETED		info about date
COMPLETED		info about date
COMPLETED		info about date
COMPLETED		info about date
COMPLETED		info about date
COMPLETED		info about date
TCI (15)		
EVI (1)		

VÝZVA PRO DRUŽICOVÉ SNÍMKY ALIANCE

Hyperlokální předpověď počasí pro zemědělství.



VSTUPY

1. Údaje z meteorologických stanic, globální předpovědi.
2. Měření přímo na stromech.

Vstupy pro budoucí zlepšení::

- topografická data
- půdní typ

VÝZVY

- Vytvořit předpověď počasí nebo popsat aktuální počasí (zejména teplotu) pro ovocné sady, zaměřit se na různé ovocné stromy a posoudit riziko zmrznutí v různých částech sadu.
- Které rostliny jsou vhodné do znečištěných oblastí?
- Můžeme použít kamery k posouzení stavu rostlin?

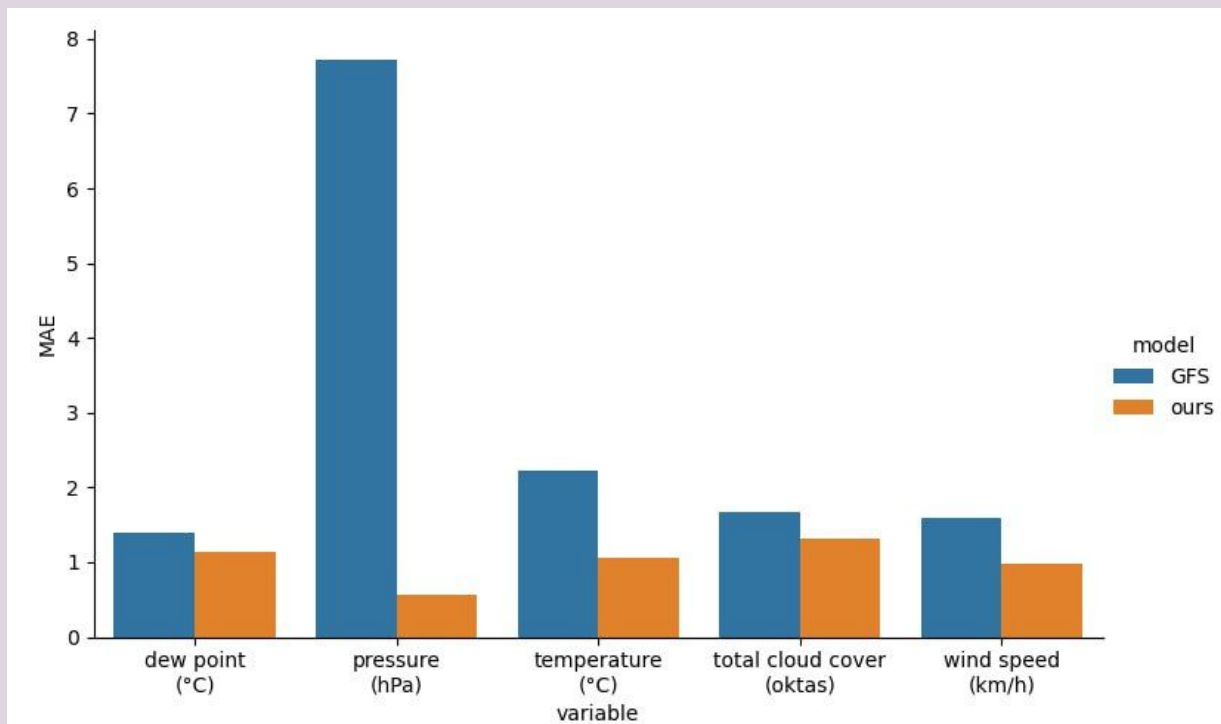
PROCESY

Jak použít naměřená data, provést předpovědi a interpolovat je tak, abychom měli teploty kdekoli v sadu.

Později je třeba provést rozhodnutí pro konkrétní oblast na základě typu stromu.

VÝSLEDKY

Zatím ne v sadu, ale víme, že to jde.



VÝZVY K AKCI

Příprava dalších kroků mezi FIT ČVUT a Národním centrem zemědělského a potravinářského výzkumu:

1. Rozhodnout, který sad má vhodná data
2. Přesně definovat, která data se budou předpovídat a jak se budou vyhodnocovat výsledky.
3. Rozhodnout o způsobu nasazení.

Také... kontakt mezi NCZPV a FIT Improlab pro analýzu videozáznamů rostlin.