

Lze získat optické vegetační indexy pro monitorování lesa ze SAR dat pomocí AutoML?

Daniel Paluba (EO4Landscape, PřF UK, Univerzita Karlova),
Bertrand Le Saux (ESA Φ -lab), Francesco Sarti (ESA EOP-SD)



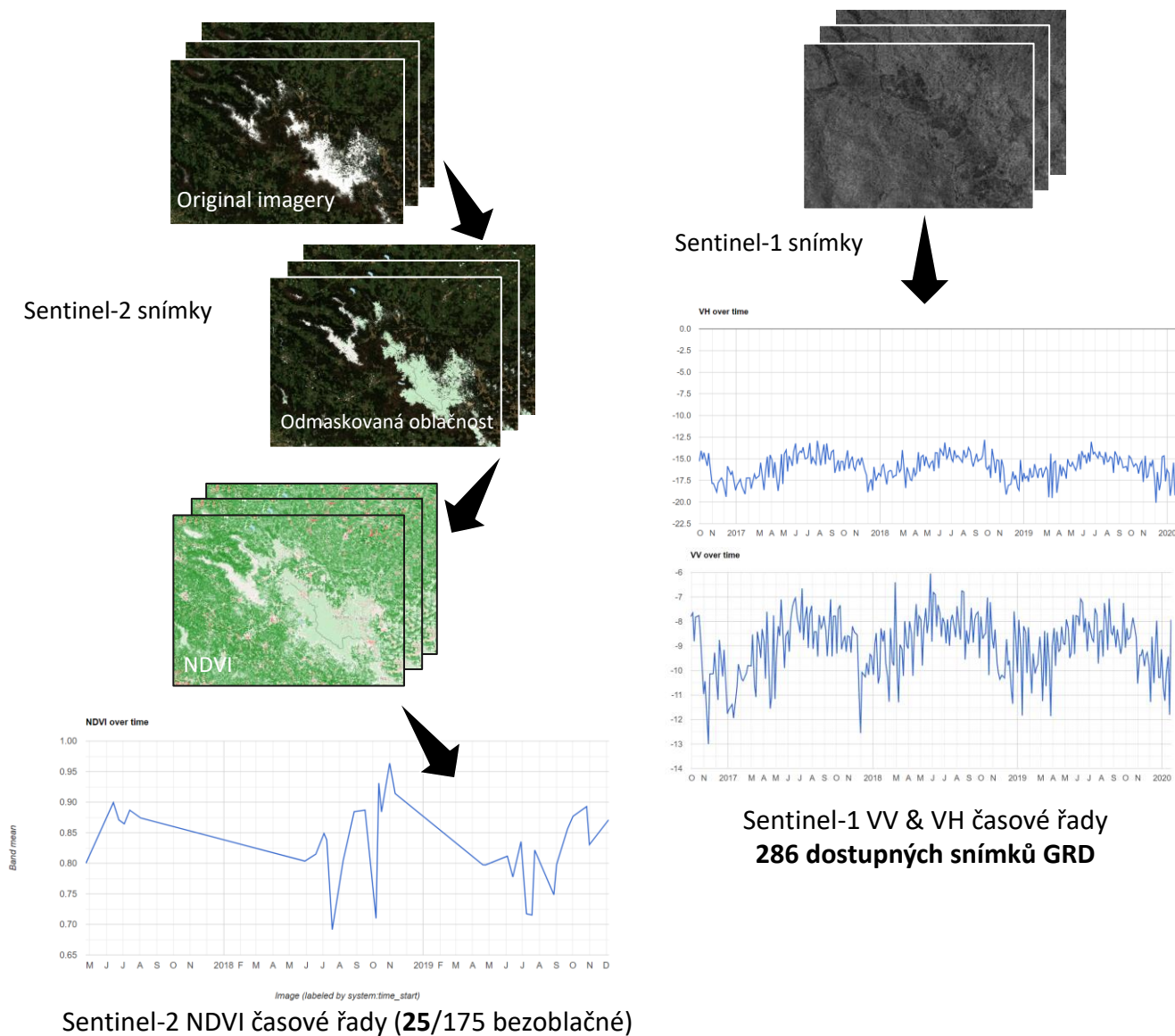
UNIVERZITA KARLOVA
Přírodovědecká fakulta



Motivace

- **Indikátory zdravotního stavu lesů pocházejí většinou z optických družicových snímků:** Techniky monitorování vegetace pomocí optických snímků jsou dobře zavedené a často je využívají osoby s rozhodovací pravomocí v lesnictví, zemědělství atd.
- **Problém:** Oblačnost - nejen v tropech → Středoevropské lesy se nacházejí převážně v horských oblastech.
- **Řešení:** Snímky SAR naopak nabízejí konzistentní časové řady!

Časová řada starého, zdravého lesa v Česku (24.04.2017 - 19.01.2020)



Cíle

1) **Optické vegetační indexy ze SAR**

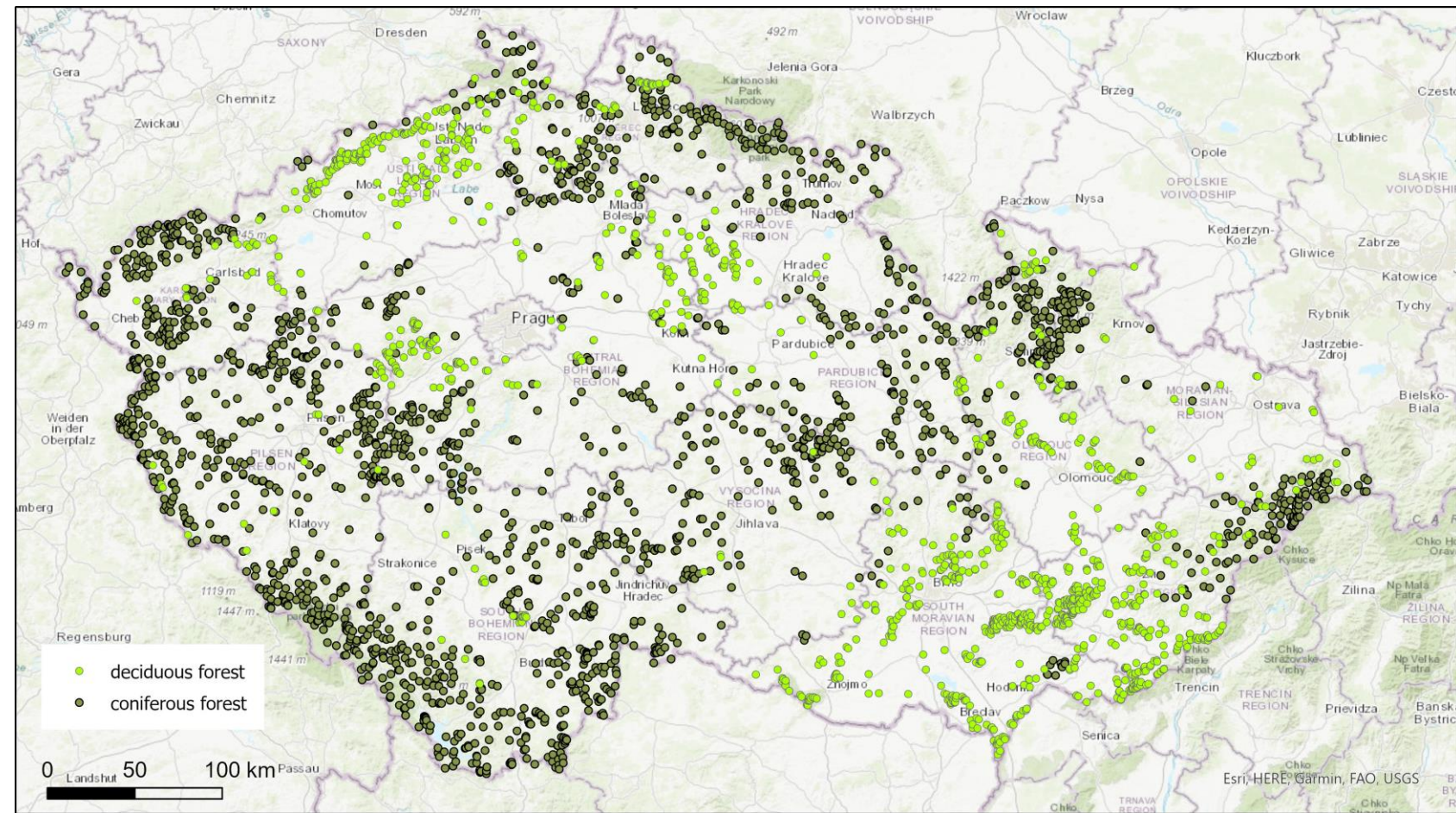
- Lze pomocí strojového učení (Machine Learning - ML) odhadnout standardní vegetační ukazatele (včetně optických vegetačních indexů) z dat SAR v pásmu C (Sentinel-1) pro různé typy lesů?

2) Podporovat současné **Essential Climate Variables (ECV)**: Použít vyvinutou metodiku při generování ECV **ve vyšším časovém i lepším prostorovém rozlišení.**

3) **Vytvořit materiály pro budoucí vzdělávací akce a školení v oblasti DPZ.**

Zájmové území

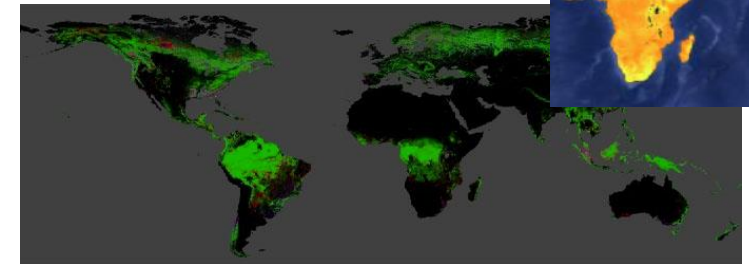
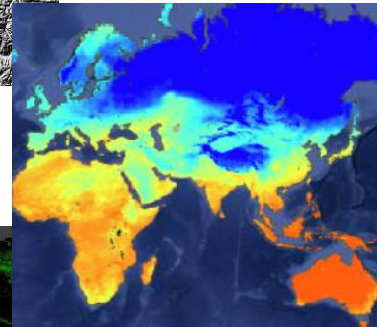
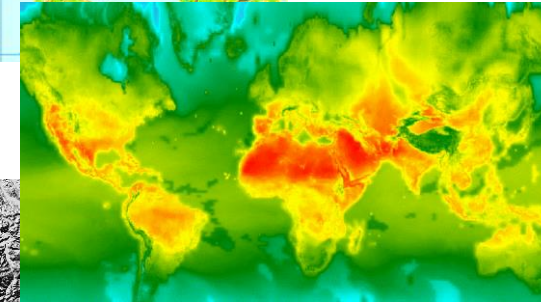
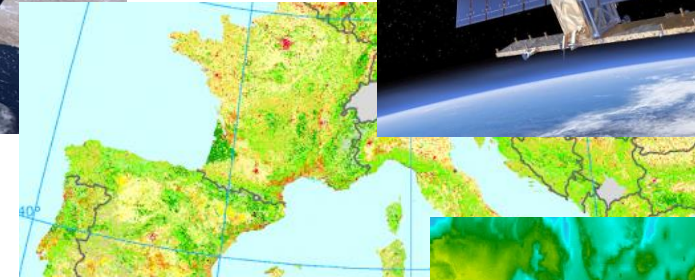
- 2100 jehličnatých lesů
- 1300 listnatých lesů



- Průnik 4 databází land cover / forest v Česku → generování náhodných bodů
 - Hansen Global Forest Change → 2000 základní vrstva (lesnatost >50 %), úbytek lesů (2000-2021) byl maskován.
 - Copernicus Global Land Cover Layer 2019 - **listnaté vs. jehličnaté porosty**,
 - CORINE Land Cover 2018 - **listnaté vs. jehličnaté**,
 - ESA WorldCover 2021

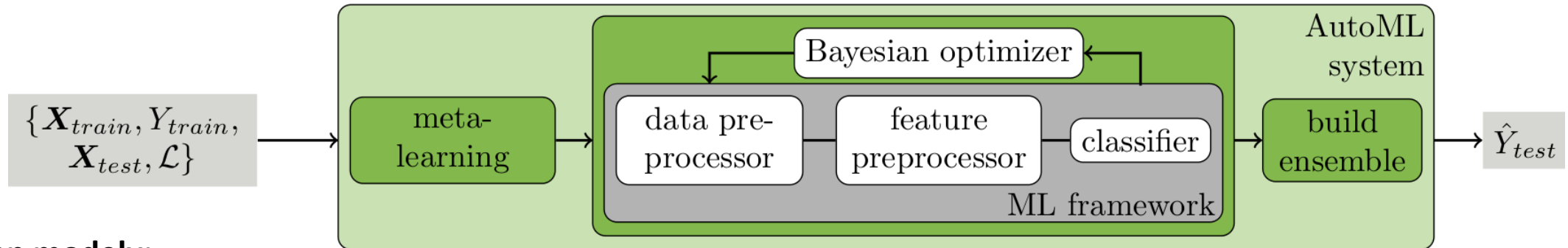
Data

- Hlavní zaměření na open access data:
- C-pásmové SAR Sentinel-1, multispectral Sentinel-2,
- Copernicus DEM (nadmořská výška, sklon)
- land cover datasets: ESA World Cover, Copernicus Global Land Cover, Hansen Global Forest Change, Copernicus CORINE Land Cover
- Datasets o počasí (srážky, teplota) z ERA5-Land
- Předzpracování časových řad dat v aplikaci Google Earth Engine (GEE);
- **Přístup AutoML (Auto-sklearn a Auto-PyTorch)**



Přístup AutoML - regresní analýza pomocí auto-sklearn

- Out-of-the-box strojové učení - snížení bariéry pro použití ML
- **Výběr algoritmu a ladění hyperparametrů**
- 15 klasifikačních a 11 regresních algoritmů, 14 algoritmů pro předzpracování dat (skicit-learn)
- Na konci se **sestaví „ensemble“** → kombinace modelů s nejlepšími výsledky
- Využívá meta-learning k identifikaci podobných předtrénovaných datasetů
- **Auto-PyTorch** → Neural Architecture Search (NAS) - Fully automated deep learning (AutoDL)



Regression modely:

- adaboost
- ard_regression
- decision_tree
- extra_trees
- gaussian_process
- gradient_boosting
- k_nearest_neighbors
- liblinear_svr
- libsvm_svr
- mlp
- random_forest
- SGD

Tréning:

- 30% na testování
- 10-fold cross-validation
- Loss function: MAE

Auto-sklearn výsledky

Statistické výsledky (pro LAI odhad):

- Listnaté lesy

- MAE = 0.296
- RMSE = 0.163

- Jehličnaté lesy

- MAE = 0.256
- RMSE = 0.113

- Smíšený model - LAI FAPAR**

- MAE = 0.268
- RMSE = 0.137

- MAE = 0.054
- RMSE = 0.005

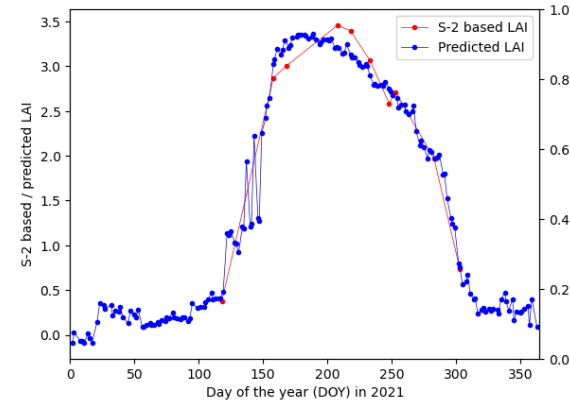
- NDVI_{red-edge}

- MAE = 0.037
- RMSE = 0.003

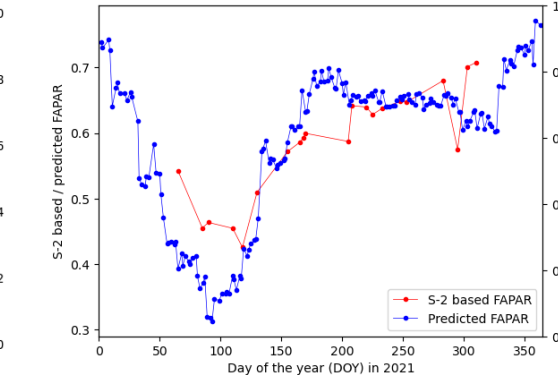
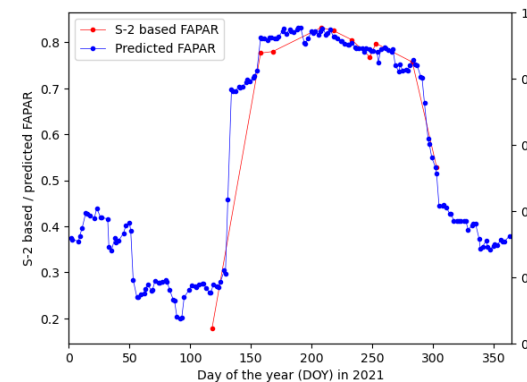
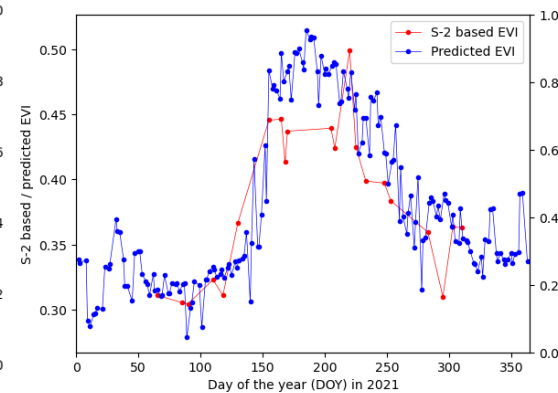
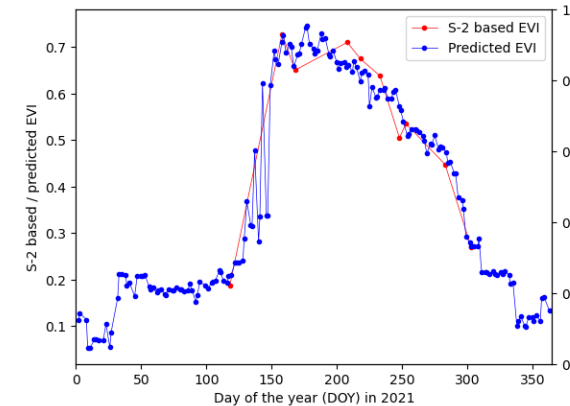
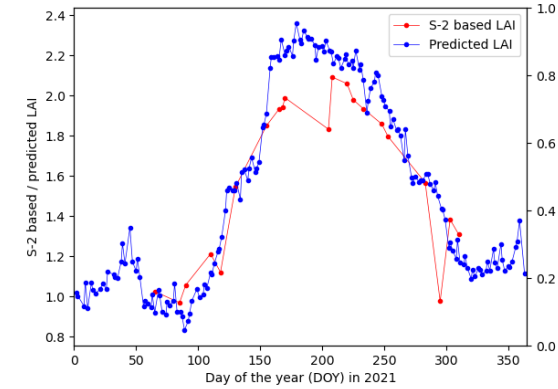
Optické: 9-25 v 2021
SAR: 70-170 v 2021

Smíšený model

Listnaté lesy

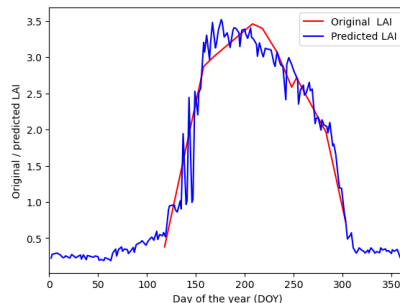


Jehličnaté lesy

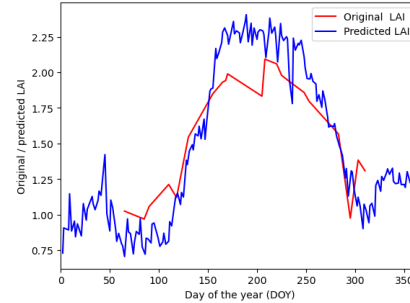


Separate models

Deciduous forests



Coniferous forests



Auto-PyTorch results

Statistické výsledky:

- LAI (all features)
 - MAE = 0.259
 - RMSE = 0.125
- FAPAR
 - MAE = 0.052
 - RMSE = 0.005

LAI (only SAR features)

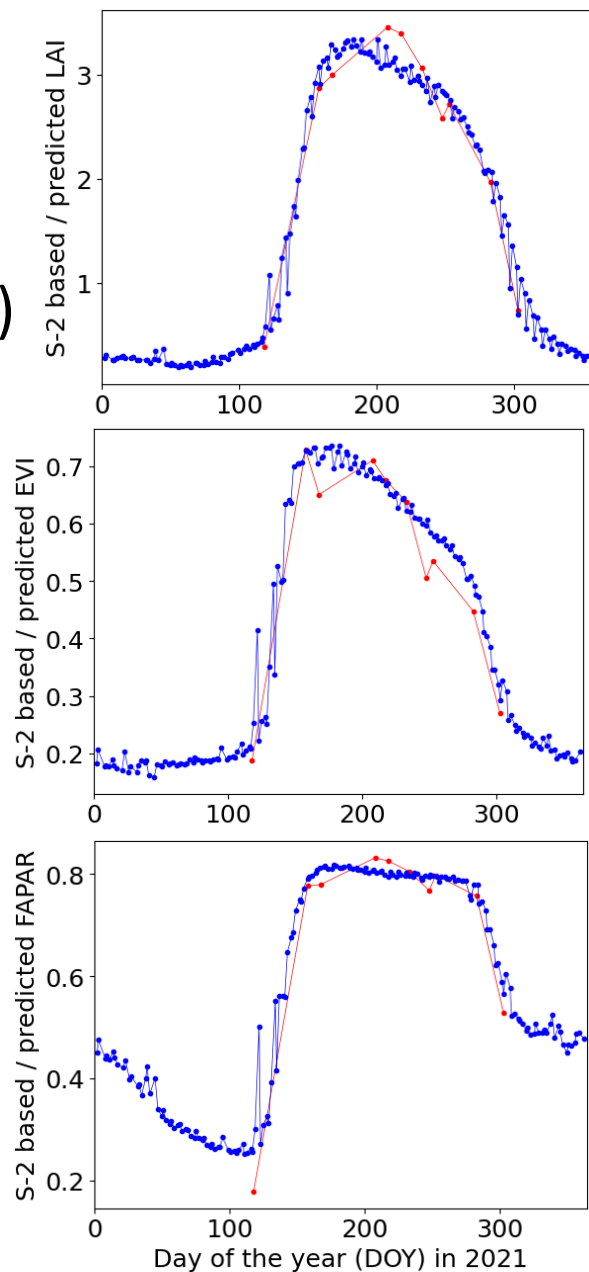
- MAE = 0.272
- RMSE = 0.136

EVI

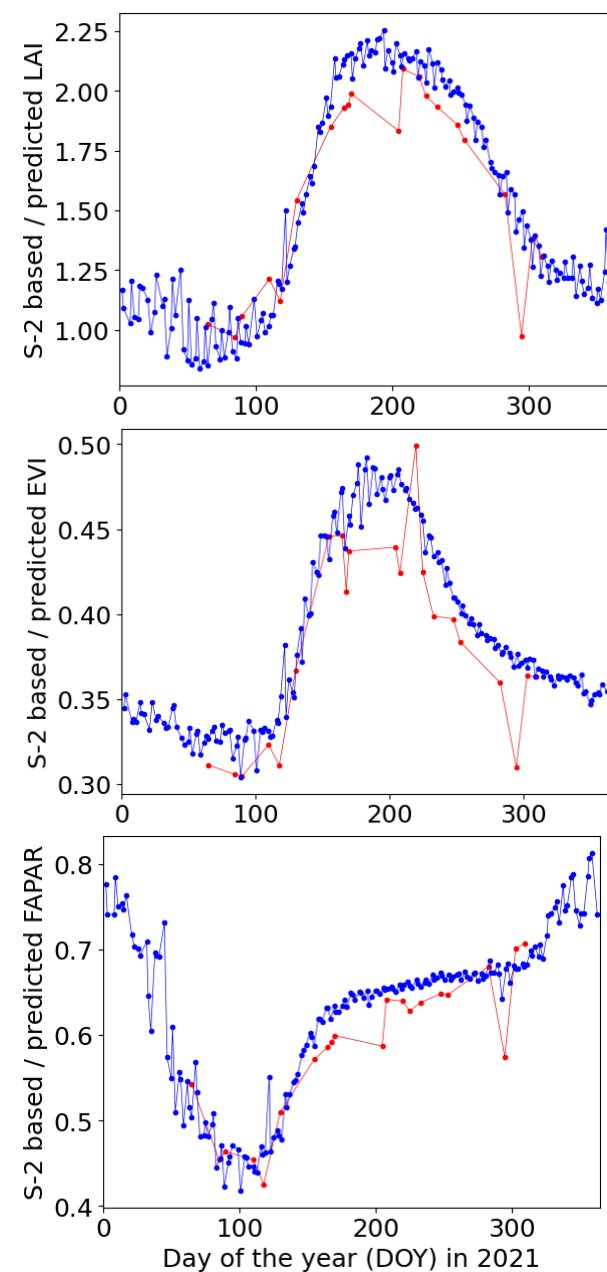
- MAE = 0.053
- RMSE = 0.006

Optické: 9-25 v 2021
SAR: 70-170 v 2021

Deciduous forests



Coniferous forests



Závěr

- Mírně lepší výsledky při použití Auto-PyTorch
 - Vyžaduje delší trénink (několik hodin) na výkonném počítači
- Auto-Sklearn lze spustit na "běžném" počítači se 4-8 jádry, přičemž kratší doba trénování stačí k nalezení dostatečného počtu úspěšných modelů → **užitečné pro vzdělávací účely**
- Ve srovnání se Sentinel-2 a jinými produkty (např. Copernicus GLMS) bylo dosaženo **lepšího časového rozlišení**.
 - Konzistentní časové řady vegetačních indexů - až **170 měření/rok**.
- **Lepší prostorové rozlišení** (20 m) ve srovnání např. s Copernicus GLMS (300 m).