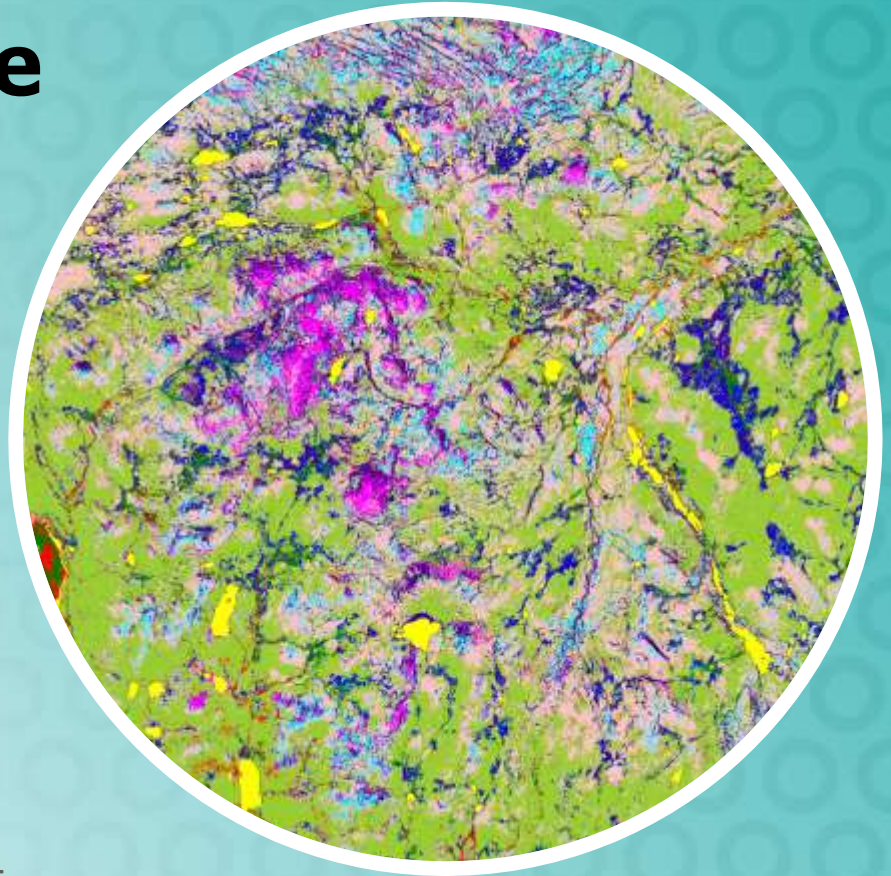


Modern machine learning in land cover classification



Mikko Impiö

Finnish Environment Institute SYKE

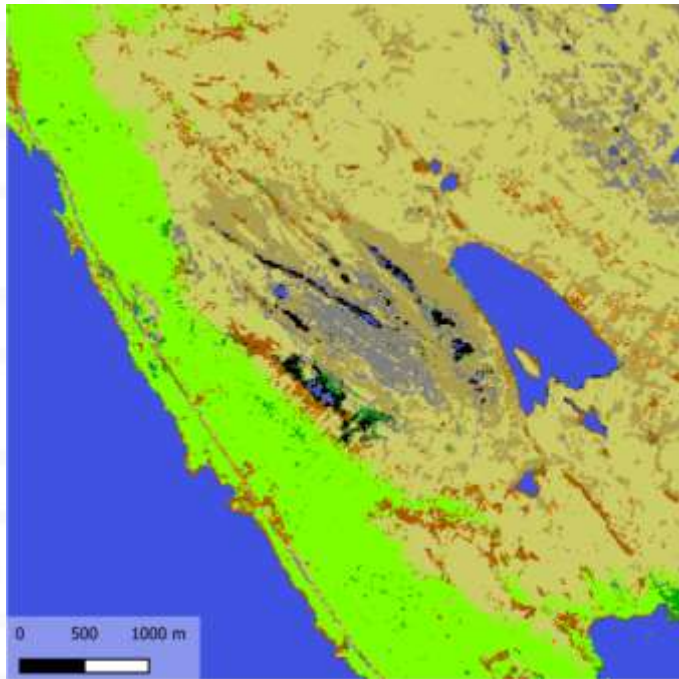
CSC Geoportti Workshop 16.5.2022

Today's topics:

Overview of the current machine learning and geospatial python ecosystem

Hands on demo for land cover classification

Land cover classification



Source: SYKE

- Classifying land cover to discrete classes is a common and well-studied problem
- Common tools
 - ArcGIS
 - ERDAS IMAGINE
 - QGIS
 - R
 - Python

Python machine learning and geospatial ecosystem



Modern tools for increasing scale



Flexible parallelization
dask.org



Hyperparameter optimization,
Industrial-level scaling
ray.io



GPUs and deep learning
pytorchlightning.ai

Interfaces make scaling trivial

Multi-GPU training with
PyTorch lightning

```
trainer = pl.Trainer(gpus=4,  
                    num_nodes=2,  
                    max_epochs=100,  
                    accelerator='gpu',  
                    strategy='ddp')
```

Multi-node parallelization
with Dask

```
from dask_jobqueue import SLURMcluster  
cluster = SLURMcluster(  
    queue = "small",  
    project = project_name,  
    cores = 1,  
    memory = "8GB",  
    walltime = "00:10:00",  
    interface = 'ib0',  
    local_directory = "/scratch/<YOUR-PROJECT>/temp"  
)
```

Code from: <https://docs.csc.fi/support/tutorials/dask-python/>

CSC tutorials and guides

- Parallelization, Dask:

<https://docs.csc.fi/support/tutorials/dask-python/>

- Distributed GPU compute, PyTorch lightning:

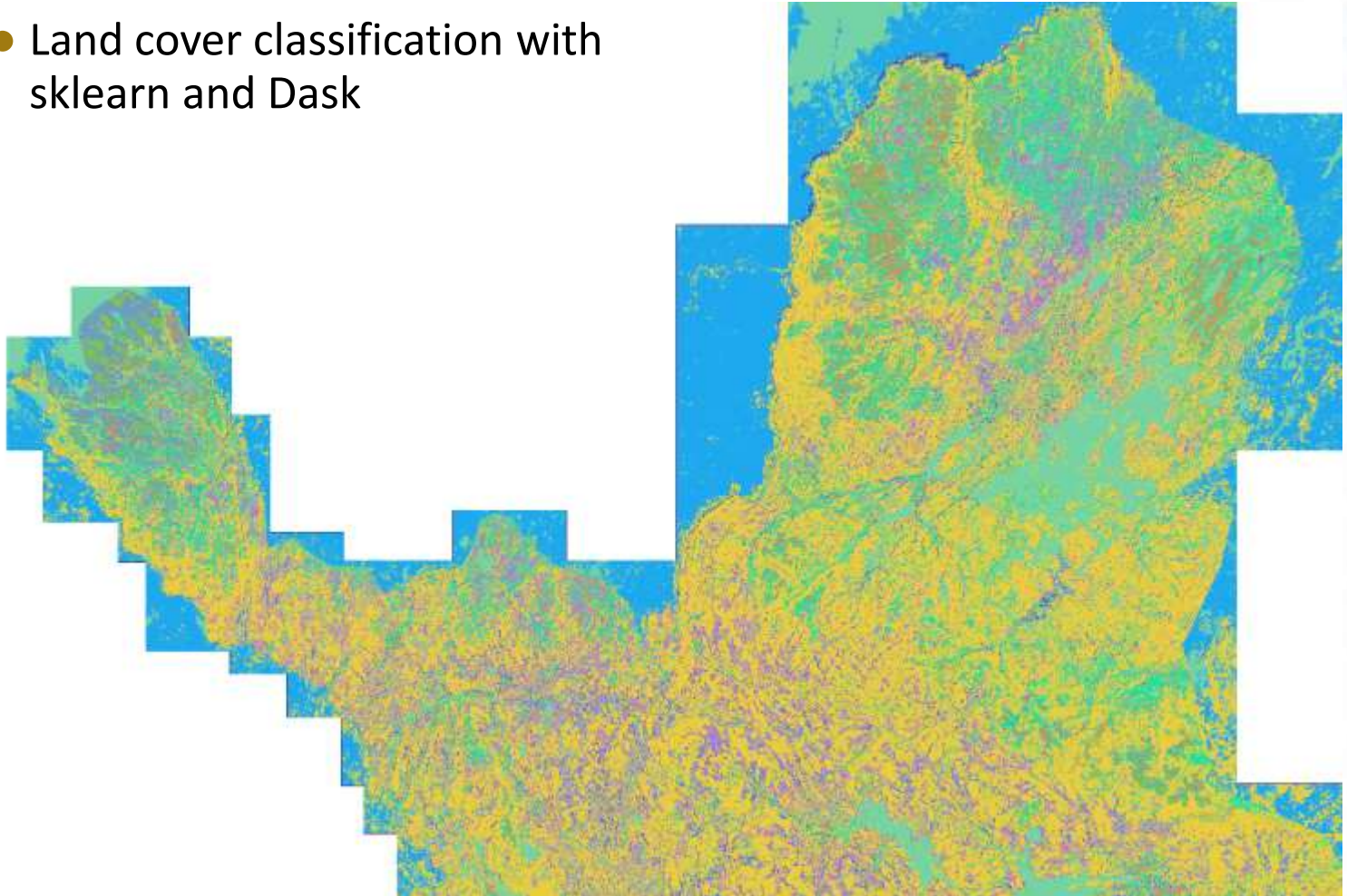
<https://github.com/CSCfi/pytorch-ddp-examples>

- Hyperparameter search, Ray:

https://docs.csc.fi/support/tutorials/hyperparameter_search/

Concrete example and demo

- Land cover classification with sklearn and Dask



Source: SYKE