Doc. ref: ESA AO/1-10552/21/I-DT

ESA Contract No.4000137253/22/I-DT

Predicting Forest Height with Self-Supervised Learning: MoCo and MAML

Lloyd Hughes (EPFL)

Marc Rußwurm (EPFL)

Oleg Antropov (VTT)

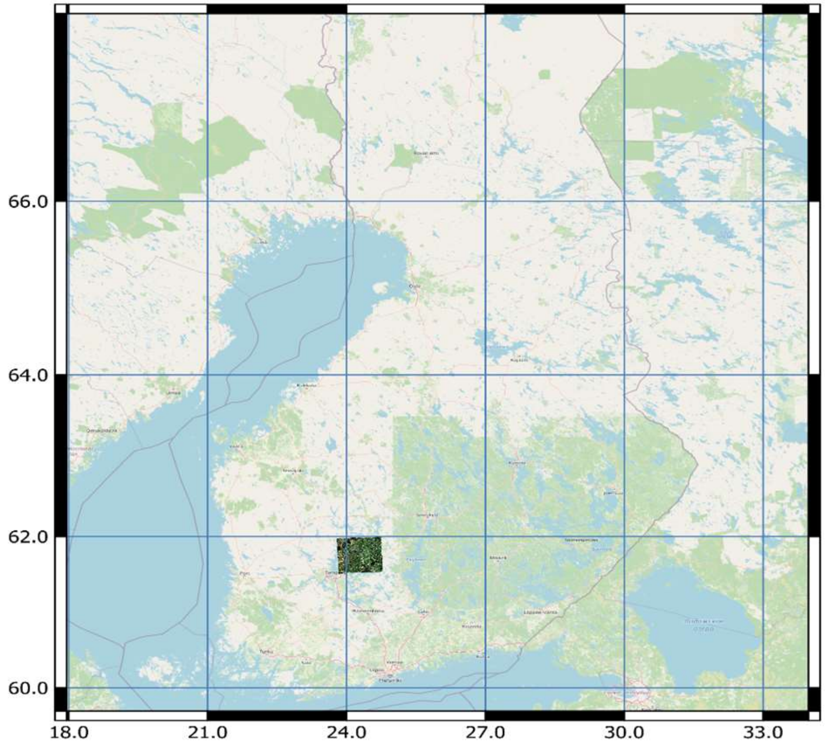
Background and Motivation

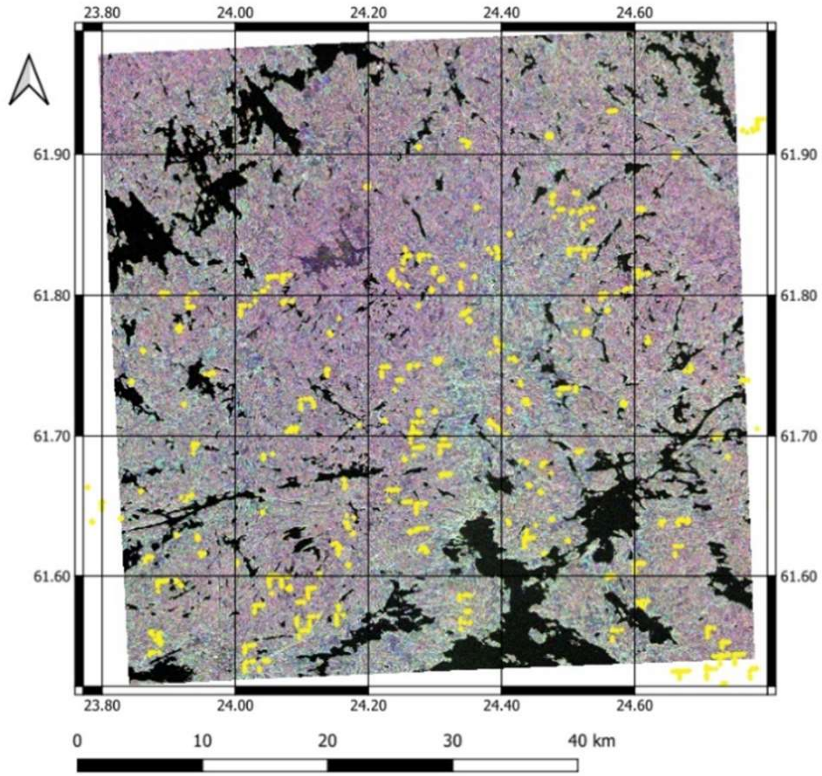
- Increased demand for timely, accurate, and large-scale forest inventories
  - Climate change
  - Deforestation
  - Urbanization

- Forest height is a key indicator of:
  - forest health
  - productivity
  - above ground biomass

- Traditional methods:
  - Ground plot survey data (Labour intensive)
  - Analysis of airborne laser scanning data (Expensive)
  - Aerial photography and expert analysis (Interpretation bias)

Study site: Southern Finland



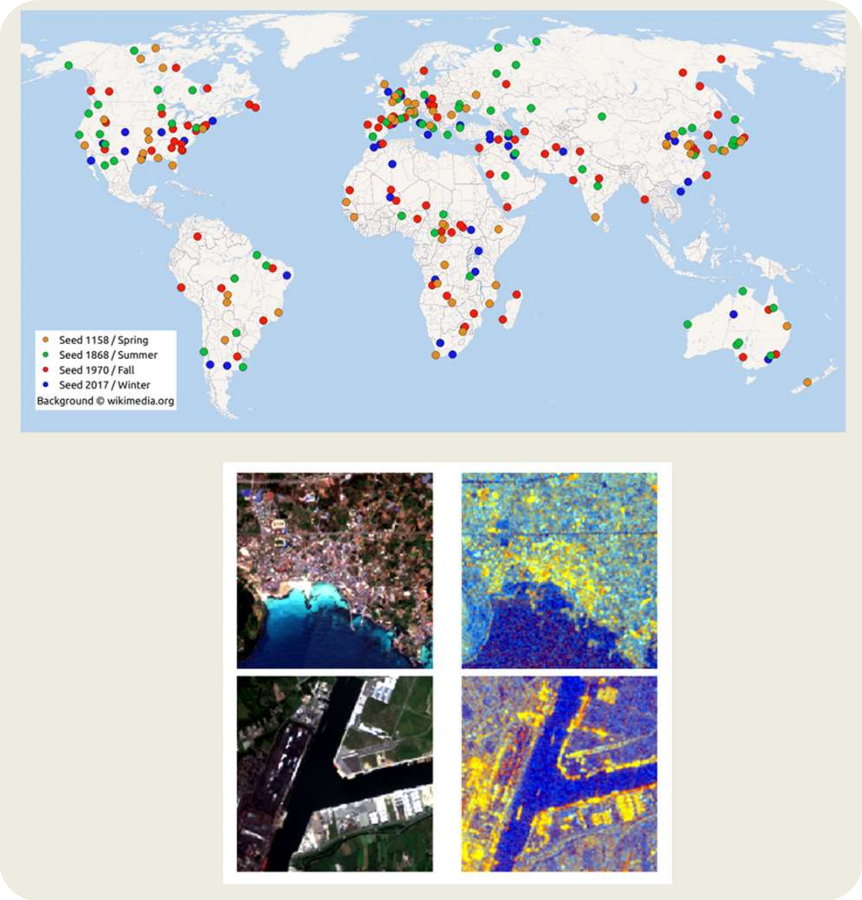


- 50 km x 50 km boreal forest in southern Finland
- 897 unique surveyed forest plots
- 15-25m diameter per plot
- Reference heights from airborne LiDAR and ground surveys

Can **unlabelled data** be exploited to enable **locally accurate** predictions of **forest plot height** from a small survey dataset?

Self Supervised Learning

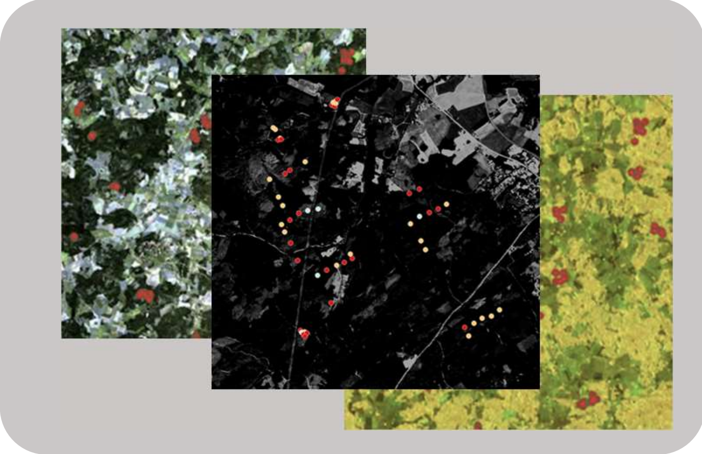
Large-scale Unlabelled Dataset



Pretrain

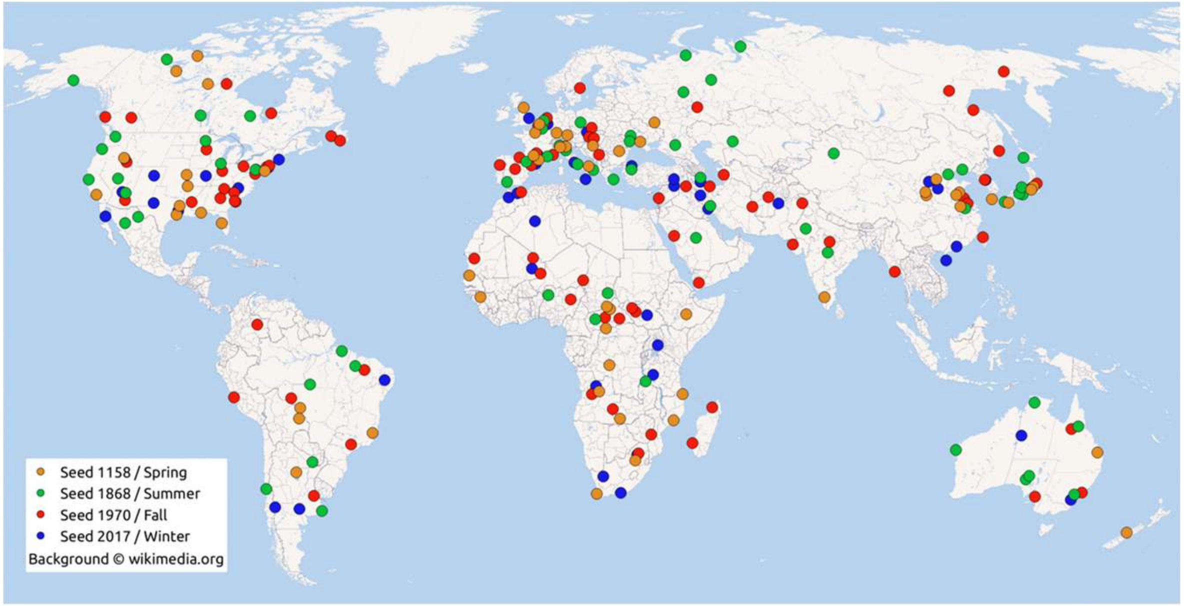
Backbone Model

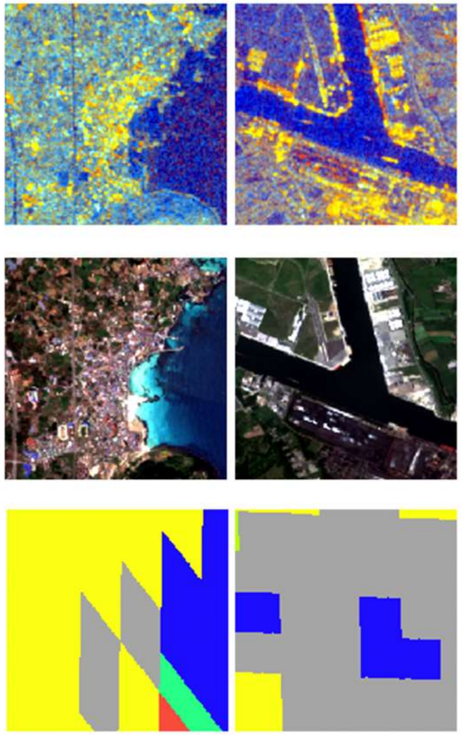
Finetune



Task/Location specific dataset

SSL: Pretraining Dataset



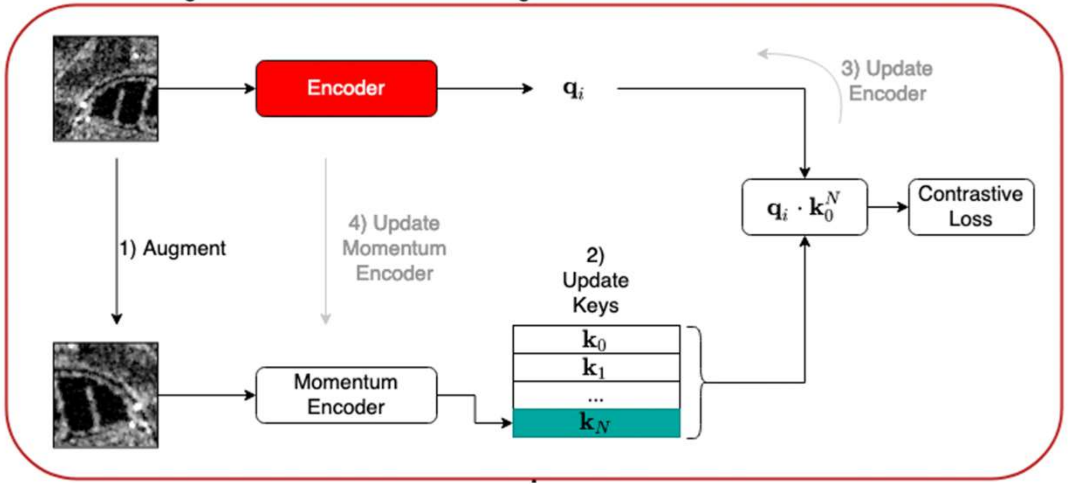


Sen12MS Dataset [Schmitt et al. 2019]:

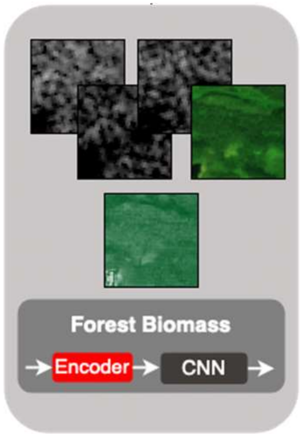
- 180,662 triplets (S1, S2, MODIS IGBP)
- 90/10 train/validation split

SSL: Momentum Contrast (MoCo)

Pretraining on unlabelled remote sensing dataset



Finetune on plot survey data

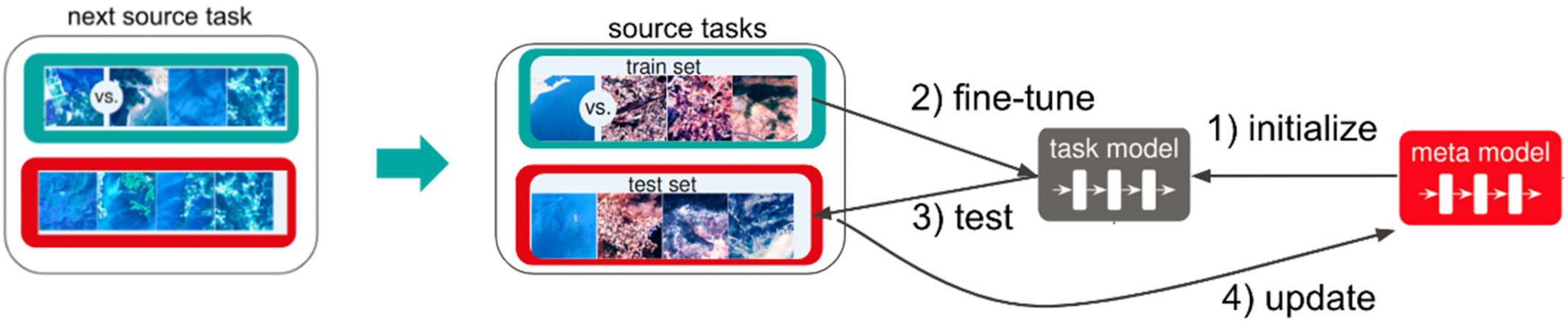


Sentinel 1	Sentinel 2
Random Cropping (128 x 128px)	
Vertical & Horizontal Flipping (p=0.25)	
Speck Augmentation (p=0.5)	Gaussian Blur (p=0.5)

- ResNet18 backbone per data modality (S1 and S2)
- Augmentations drive representation learning

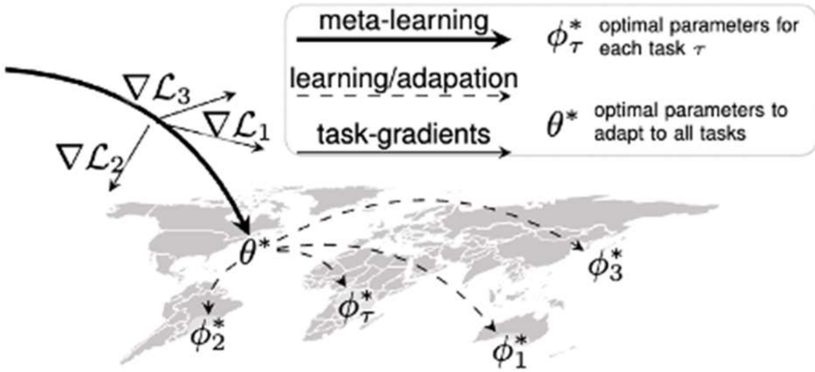
We learn transferrable representations of the data

SSL: Model-agnostic Meta-learning (MAML)



- ResNet-12 backbone per modality
- Diverse set of binary 1-v-all tasks drive learning of a meta-model

We learn a model to learn new problems



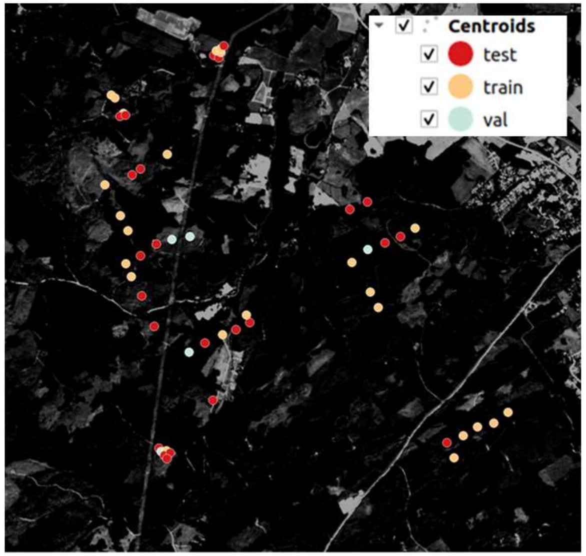
Dataset Creation

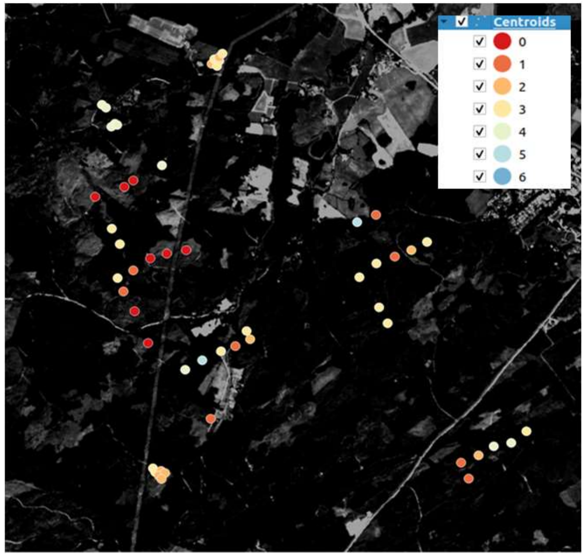
Data split:

- Plot level forest heights partitioned into 6 classes
- 50/50 Stratified split into test/train subsets

Training patches:

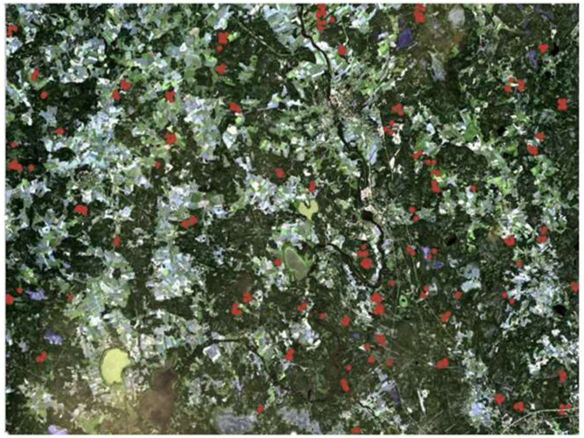
- Single Sentinel-2 (L1C)
- Median multi-temporal Sentinel-1 (VV+VH)
- Cropped patches centered on forest plot

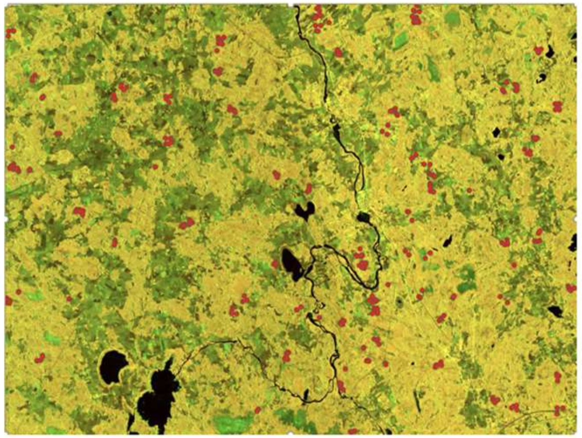




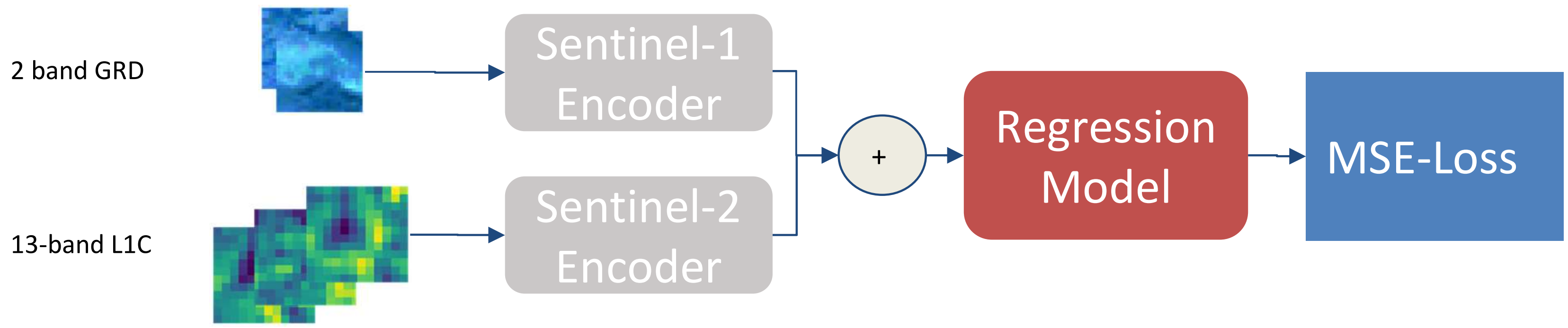
Sentinel-2

Sentinel-1





MoCo: Fine-Tuning



- Fixed encoder networks, learn regression model
- Fine-tuned using:
  - Sentinel-2 only
  - Sentinel-1 and Sentinel-2 (concatenated representations)

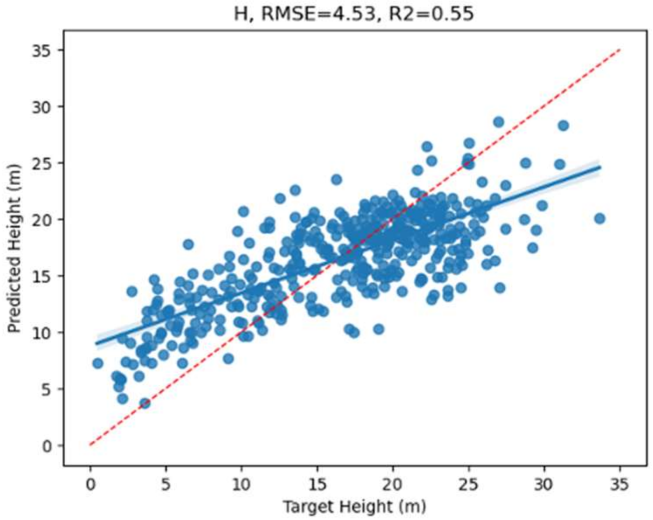
- Two patch sizes:
  - 4x4 pixels
  - 12x12 pixels

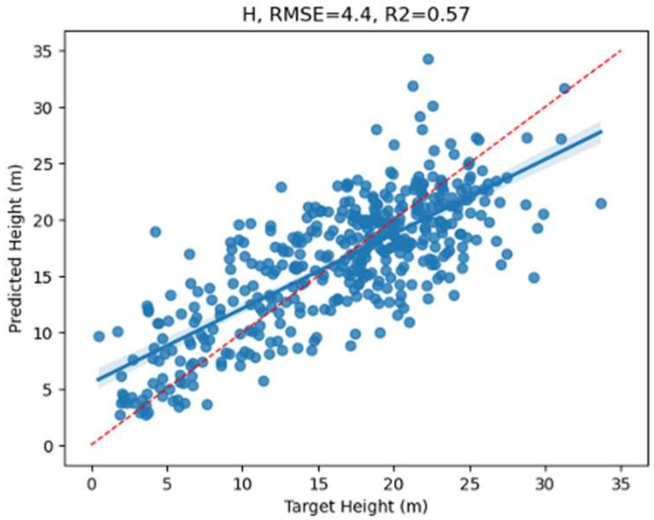
MoCo: Fine-Tuning

4x4px

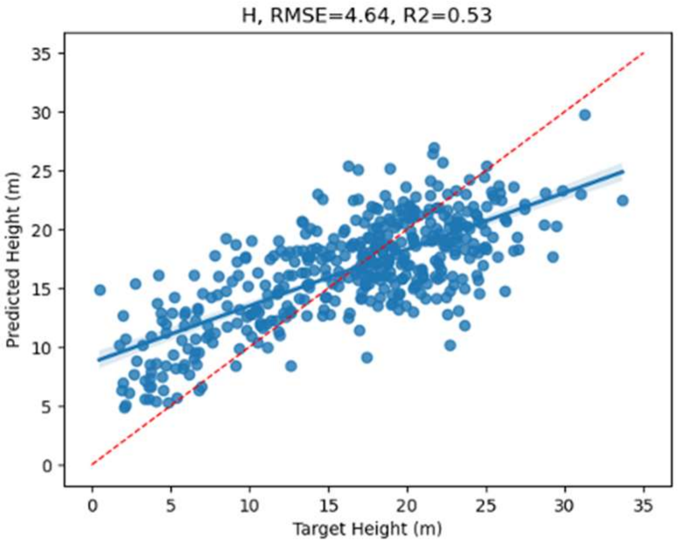
12x12px

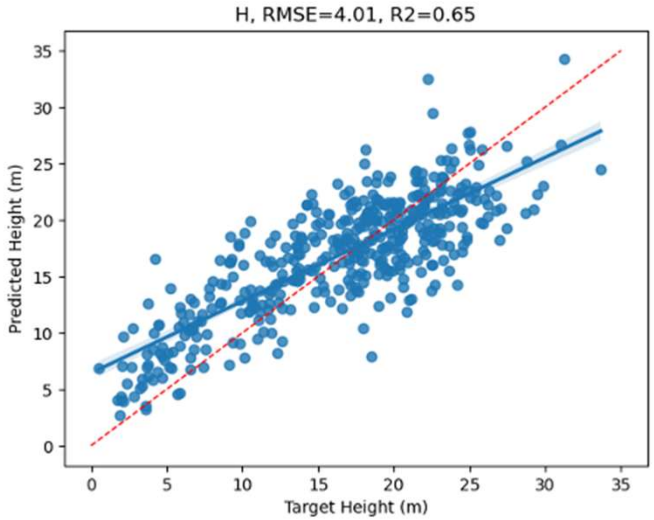
S2 Only





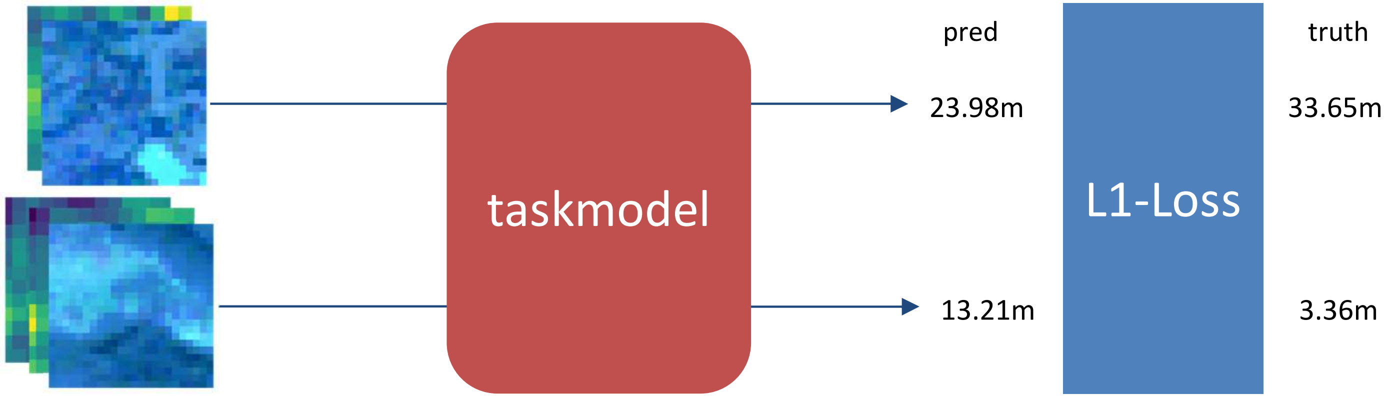
S1 + S2



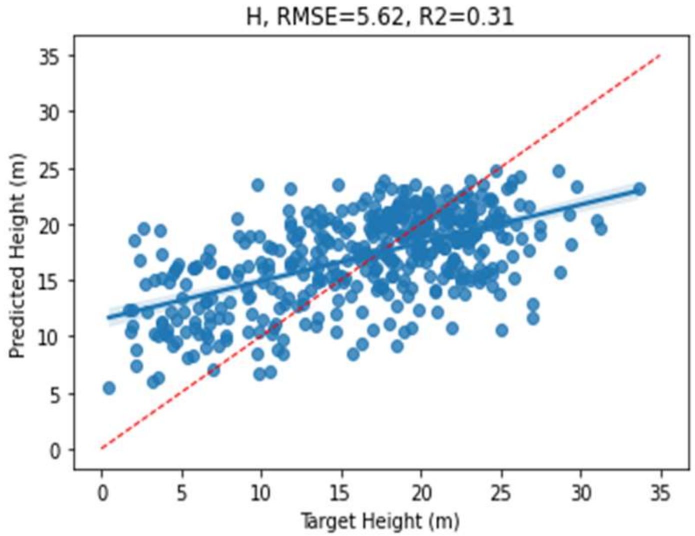


UC5: MAML fine-tuning

MAML fit
L1 Loss



- ResNet-12 task model
- Fine-tune task model using L1 loss
- Sentinel-1 and Sentinel-2 stacked at input
- 24x24 pixel patches



Results

Model	**R^2**	**RMSE (meters)**
kNN Baseline	0.44	5.07
MAML (24px)	0.31	5.62
MoCo S1+S2 (4px)	0.53	4.64
MoCo S1+S2 (12px)	**0.65**	**4.01**
MoCo S2 only (4px)	0.55	4.53
MoCo S2 only (12px)	0.57	4.4

- MoCo better capture the spectral information
- Sentinel-1 has little impact on regression accuracy
- MoCo S1+S2 with 12x12px patches exploits spatial autocorrelations
- Model is reasonable for interpolation, but is unlikely to be predictive
- Application to other regions and times will require further fine-tuning
- Models can be applied to predicting other forest variables

RepreSent **project**

Thank you!

DLR