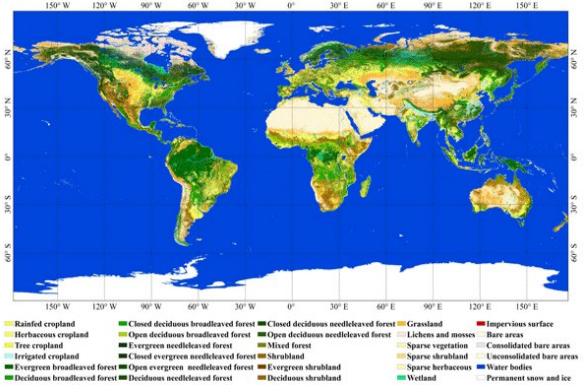
A framework for global ensemble land cover mapping at 30 m resolution (2000–2025+)

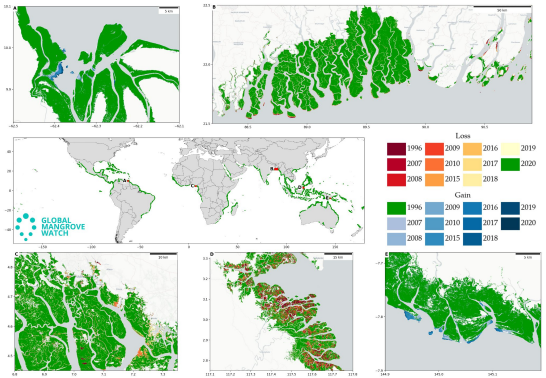
Mateo Moreno, Rolf Simoes and Tomislav Hengl
OpenGeoHub Foundation, the Netherlands

Context

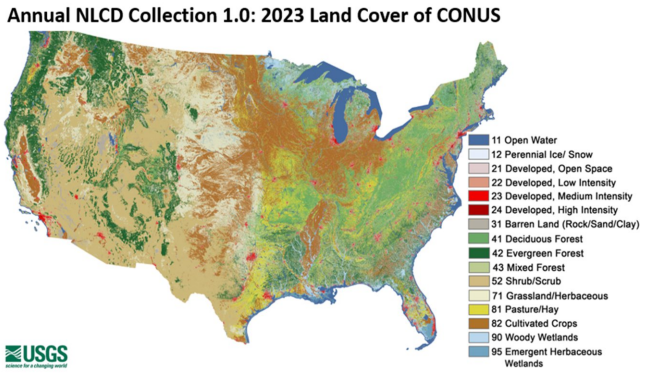
Today there are many independent active land cover mapping projects in the world.



GLC-FCS30D (Zhang et al., 2021)



GMW (Bunting et al., 2022)



NLCD (USGS)

- LC maps have varying accuracy and different legend systems
- Detailed national LC maps
- Specialized LC maps
- Probabilistic vs categorical formats

Context and challenge

Most existing products are great, yet each has limitations. Some are redundant or outdated, while others excel only in specific aspects.

Context and challenge

Most existing products are great, yet each has limitations. Some are redundant or outdated, while others excel only in specific aspects.

A fully reproducible, open-data framework for Global Ensemble Land Cover (GELC) mapping.

Context and challenge

Most existing products are great, yet each has limitations. Some are redundant or outdated, while others excel only in specific aspects.

A fully reproducible, open-data framework for Global Ensemble Land Cover (GELC) mapping.

- Systematic method to harmonize diverse LC products into a consistent output.
- Adaptability to local and regional contexts through flexible reference data integration.
- Full transparency and reproducibility. Open data, open methods.

Global multisource ref. points

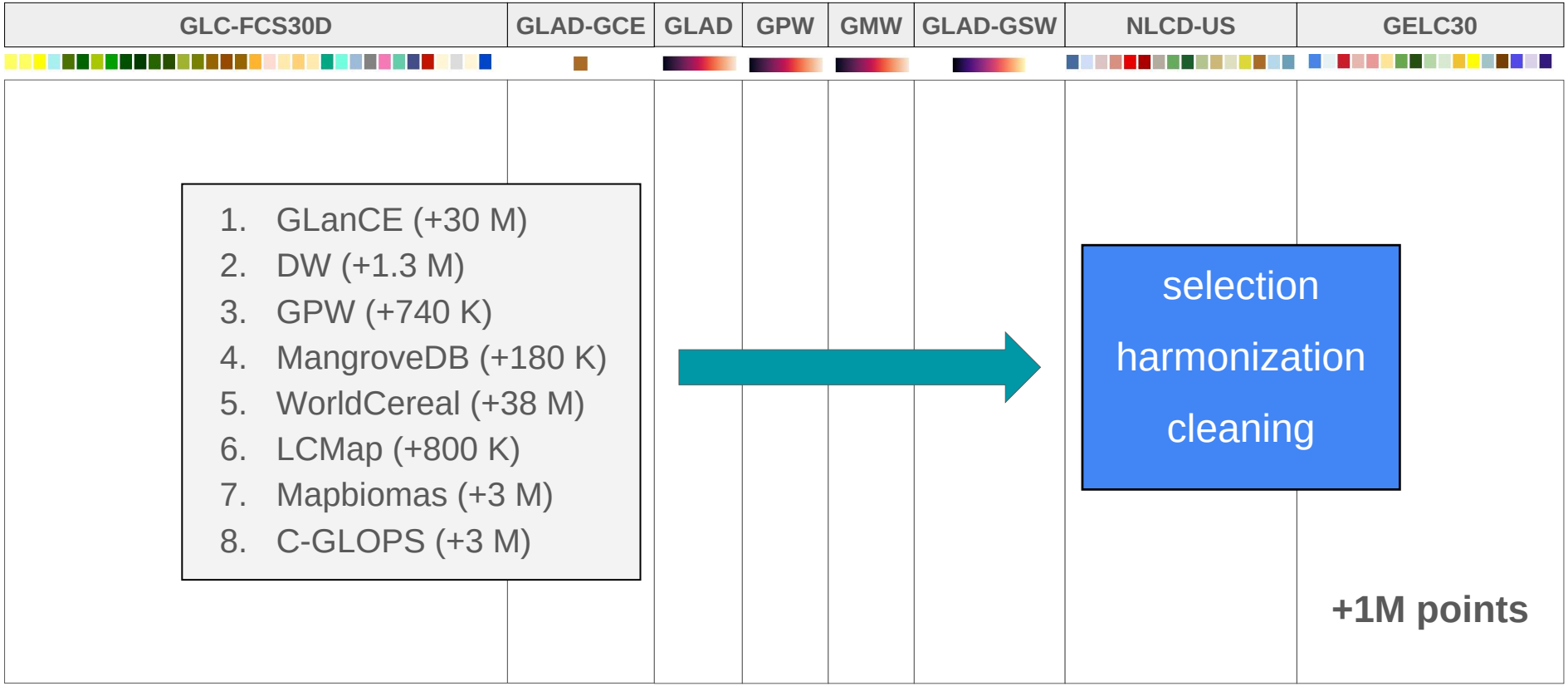
1. GLanCE (+30 M)
2. DW (+1.3 M)
3. GPW (+740 K)
4. MangroveDB (+180 K)
5. WorldCereal (+38 M)
6. LCMap (+800 K)
7. Mapbiomas (+3 M)
8. C-GLOPS (+3 M)

Compilation
+77 M points

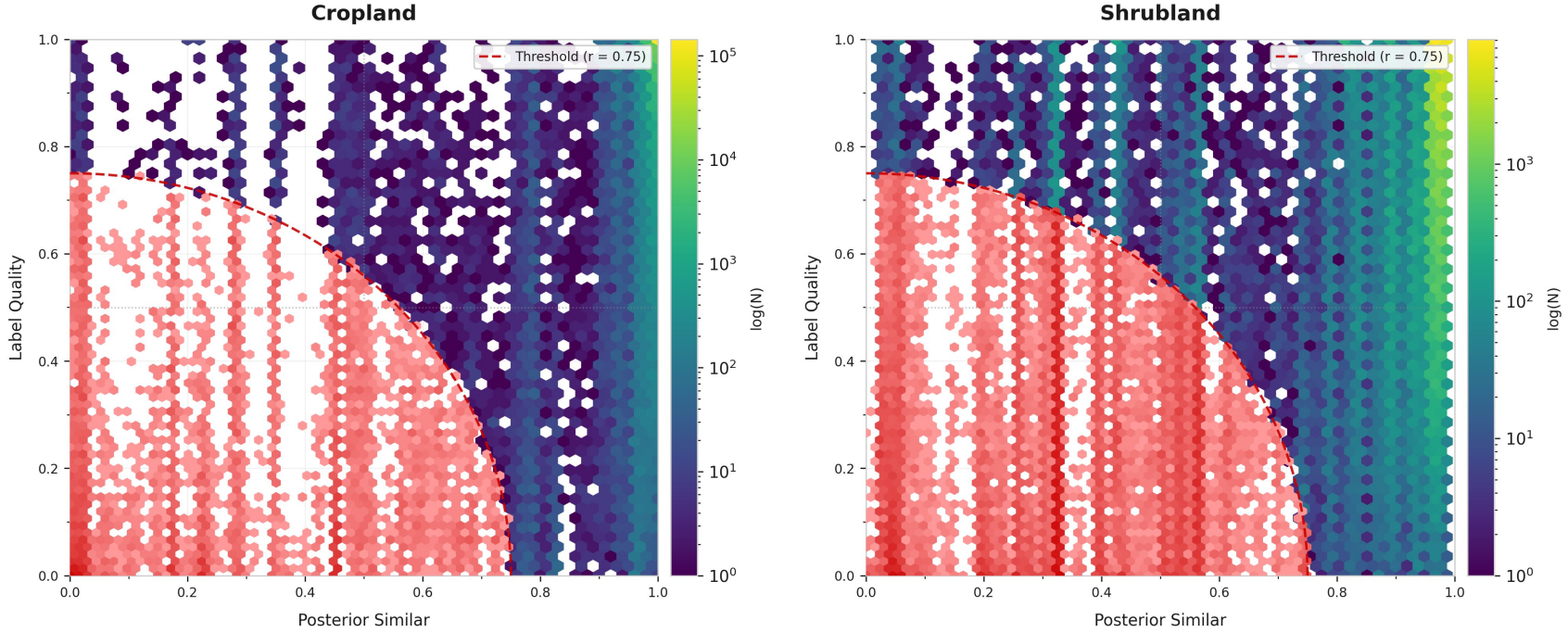
Selection & harmonization
+29 M points

https://doi.org/10.5281/zenodo.15594681

Reference points – CONUS pilot



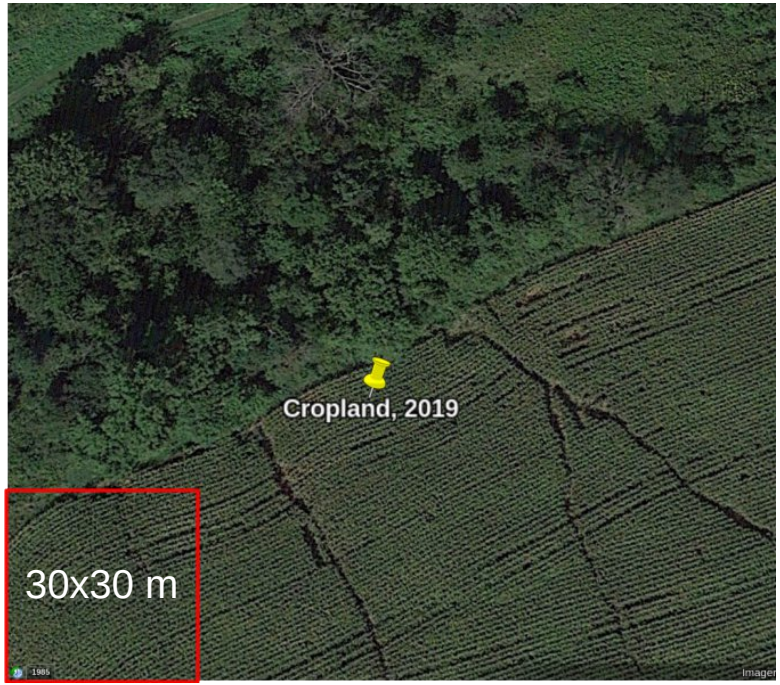
Data cleaning – CONUS pilot



- **Cleanlab:** self-confidence (quality metric)
- **Bayesian:** naive posterior probability

Euclidean magnitude < 0.75 → low quality points ~7% of total sample (CONUS)

Data cleaning - examples



- Issues with class boundary
- Seasonality (e.g., water bodies and permanent snow)
- Issues with class labels

Results – CONUS pilot

	UA	PA	F1-score
1. Developed/urban	0.775	0.674	0.721
2. Cropland	0.852	0.885	0.868
3. Grassland and shrubland			
3.1 Natural semi-natural grassland	0.856	0.556	0.674
3.2 Cultivated grassland/pasture	0.641	0.576	0.607
3.3. Shrubland	0.536	0.643	0.584
4. Forest			
4.1 Deciduous forest	0.622	0.754	0.682
4.2 Evergreen forest	0.641	0.728	0.682
4.3 Mixed forest	0.632	0.472	0.540
5. Barren			
5.1 Bare soil	0.798	0.777	0.787
5.2. Rock outcrop	0.959	0.875	0.915
5.3 Shifting sand	0.848	0.894	0.870
5.4 Permanent snow/ice	0.829	0.829	0.829
6. Wetland	0.968	0.713	0.821
7. Water	0.834	0.976	0.900

Random forest classifier
Test set of 10% of sample size

Level 1

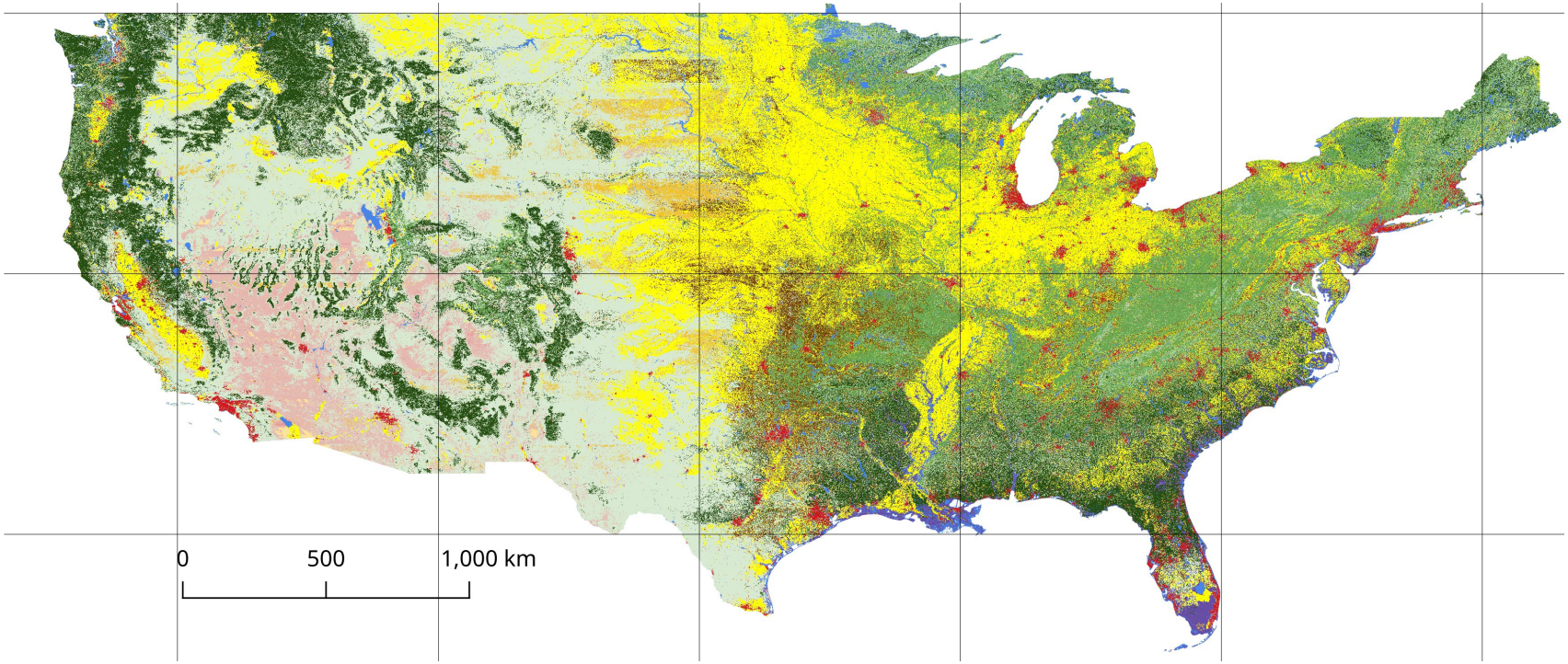
Accuracy 0.86
F1-score macro 0.83

Level 2

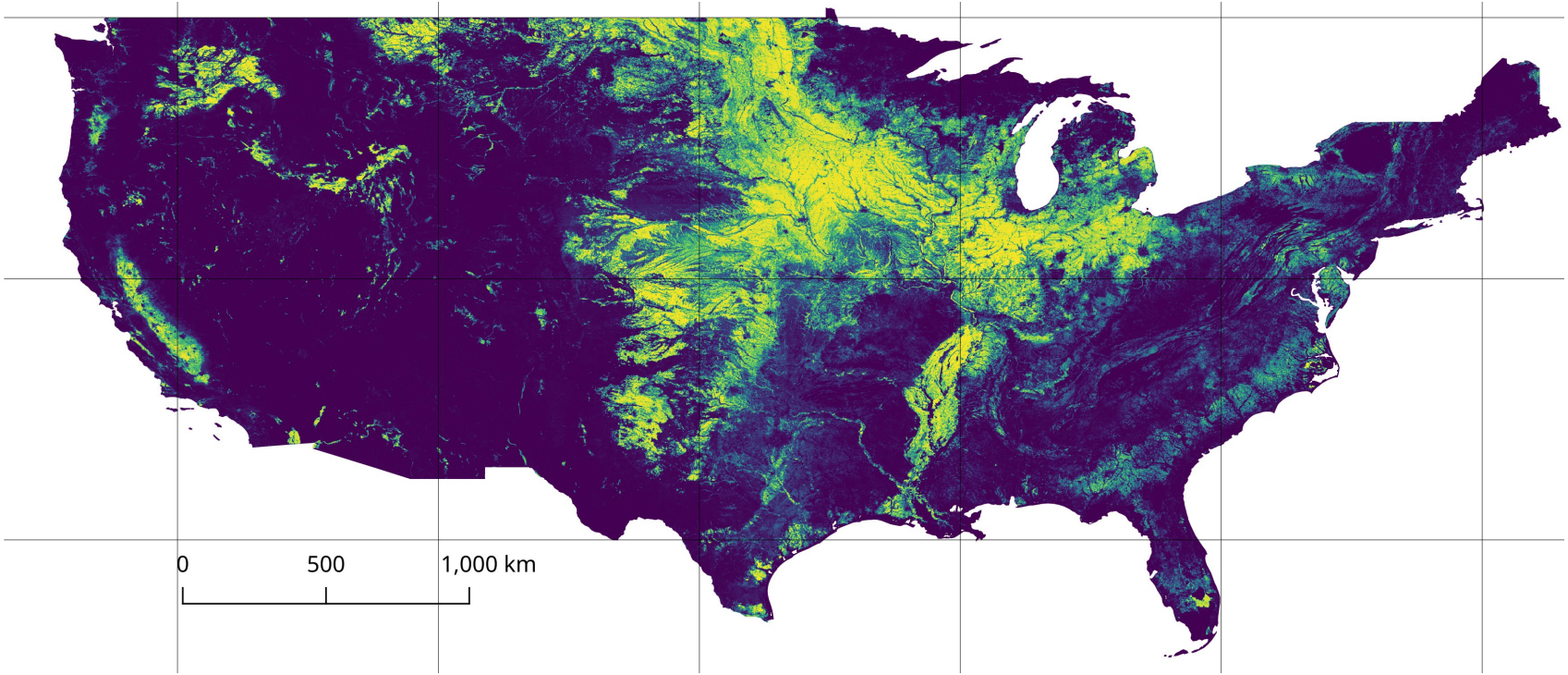
Accuracy 0.74
F1-score macro 0.77

Computing time
1.5 h/year/server

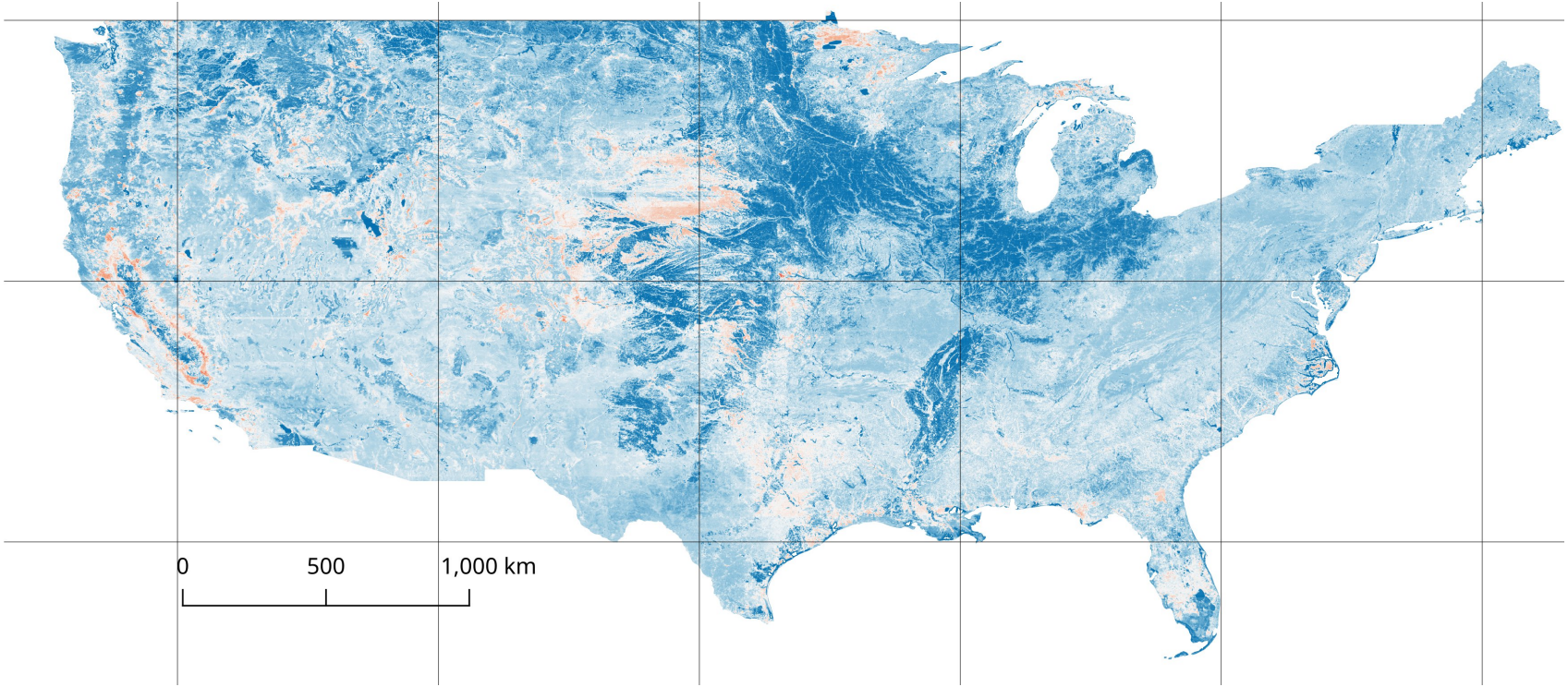
Results – CONUS pilot (2020)



Results – CONUS pilot (2020)



Results – CONUS pilot (2020)



Next steps

- Apply the trained framework to continental Europe (30m, 2000–2025+), leveraging the expanded harmonized reference dataset.
- Eventually extend the pipeline to global scale, assessing model transferability across bioclimatic zones and legend systems.
- Pipeline LC maps with different spatial and temporal resolutions to support multi-scale applications.
- Release the harmonized reference dataset and model outputs as open-access resources to support community validation and reproducibility.

Thank you!

mateo.moreno@opengeohub.org

mmorenoz

https://opengeohub.org

Open Geospatial Carbon Registry

AI 4 Soil Health

OPEN EARTH MONITOR