Funded by project

NFDI4Earth

Learn more at nfdi4earth.de

Is maximizing spatial resolution worth the computational cost?

A Case Study in the Cerrado Biome (Brasília), Brazil

Yomna Eid [1], Edzer Pebesma [1]

[1] University of Münster, Department of Geoinformatics (ifgi)

Contact Me

via e-mail *via LinkedIn*

WWU MÜNSTER

ifgi Institut für Geoinformatik Universität Münster

Motivation

Remote sensing products demand high storage-capacities, with imagery archives spanning ***petabytes***. High- and very high-resolution remote sensing imagery has emerged as an important source of data for various geoscientific analyses, most of which are **highly computationally taxing**. With a trend of increasing spatial and temporal resolution, a crucial question remains: is the **accuracy** and **overall quality** of the analysis significantly impacted when the high-resolution product is substituted with a less computationally-intensive, lower-resolution one?

A generally accepted attitude is that developing products at higher resolutions is a legitimate scientific goal. However, the interest is often not *which* 10 m pixel changes land use and *when* exactly things happen, but rather how many pixels change land use over a larger area (a country, or basin) and over a larger time period (e.g. by year over a decade).

For 10-meter resolution images from the **Google Earth Engine Sentinel-2 Harmonized Data Catalog**, an **NDVI-classification** is carried out, splitting pixels into two classes: Forest and Non-forest. We evaluate how time-series of aggregated **Forest fraction** computed at **progressively lower spatial resolution** data changes in quality (accuracy), and which lower resolutions still seem acceptable. We use systematic sampling, which corresponds to down-sampling with "nearest" strategy.

Research Question

1. **How** do estimates of Forest fraction change when the classified image is gradually down-sampled from a 10-m resolution to 10000-m resolution?
2. **How** does the **standard error** of this fraction vary with down-sampling when using systematic non-random sampling? [1]
3. **When** does lowering the resolution stop being acceptable?

Results

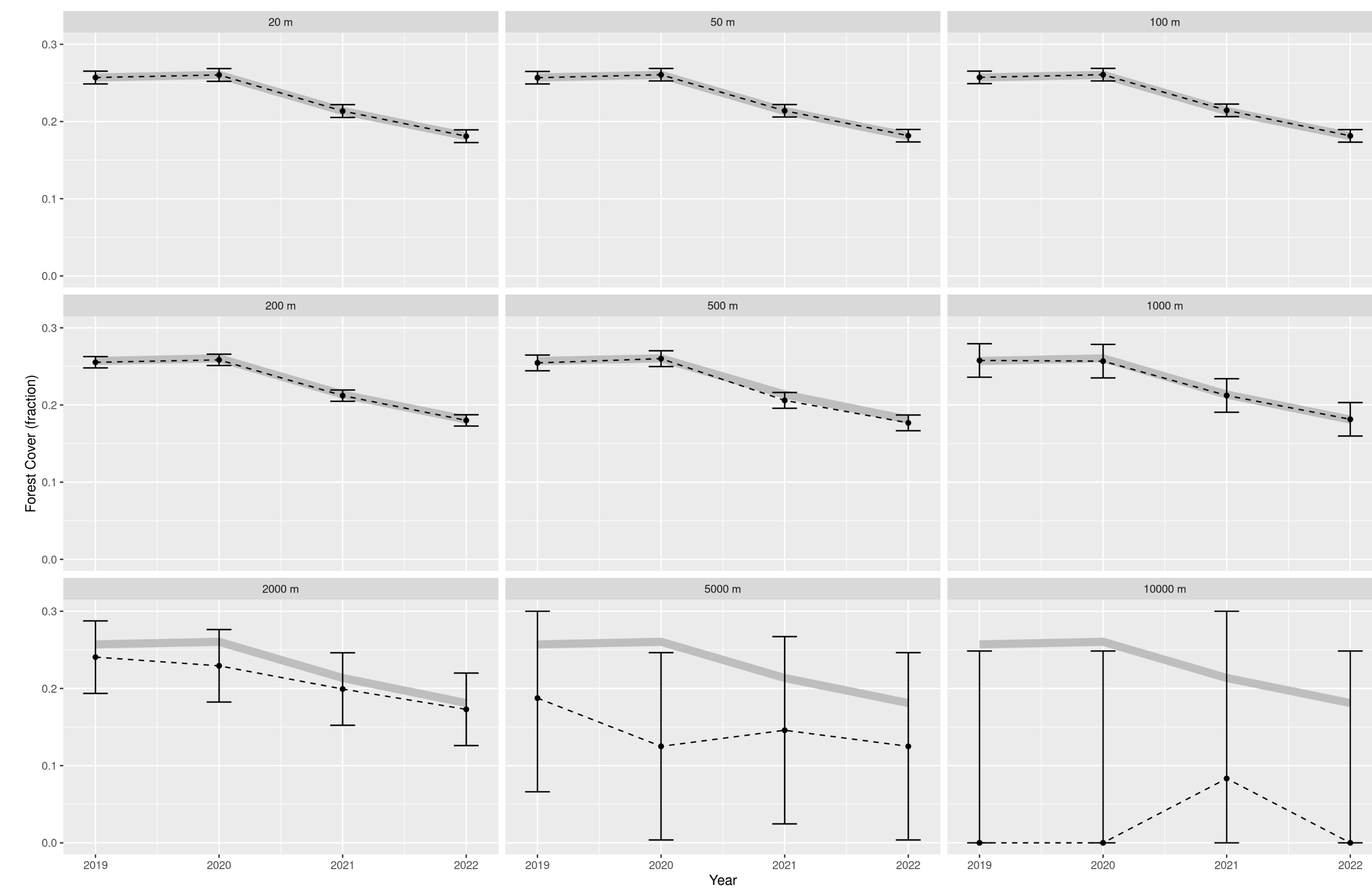


Figure A

The Forest fraction for the ***"true"*** 10-m resolution is displayed in grey for reference in the plots. The standard **error bar** remains unchanged until 200-m resolution, after which it rapidly oscillates in magnitude.

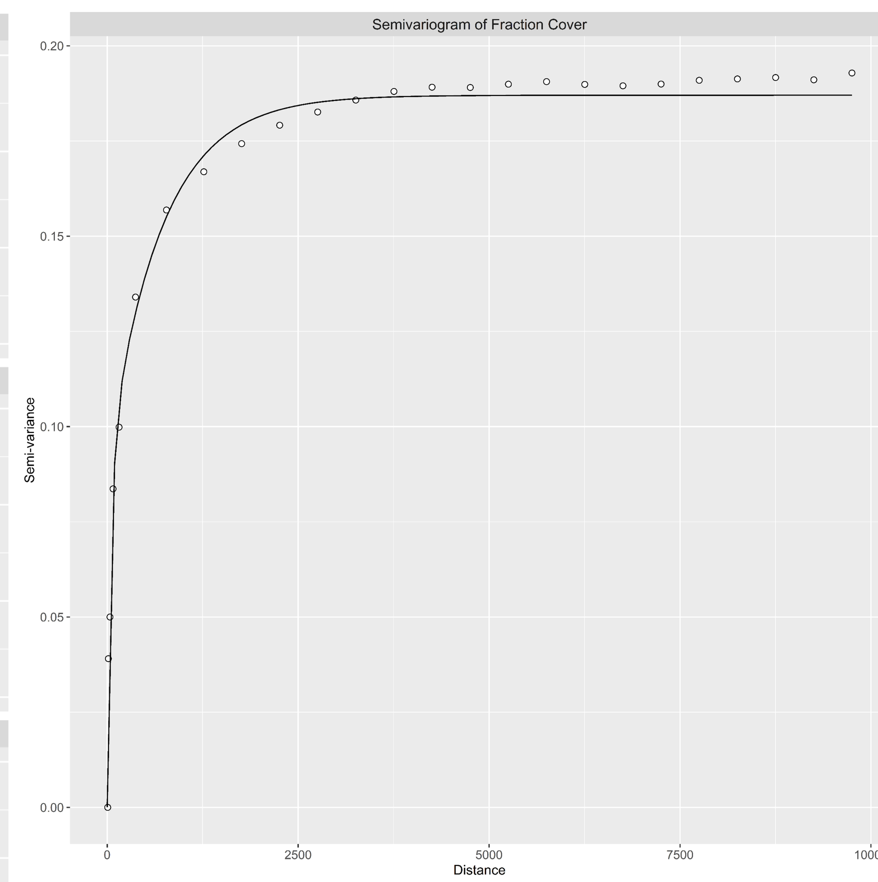
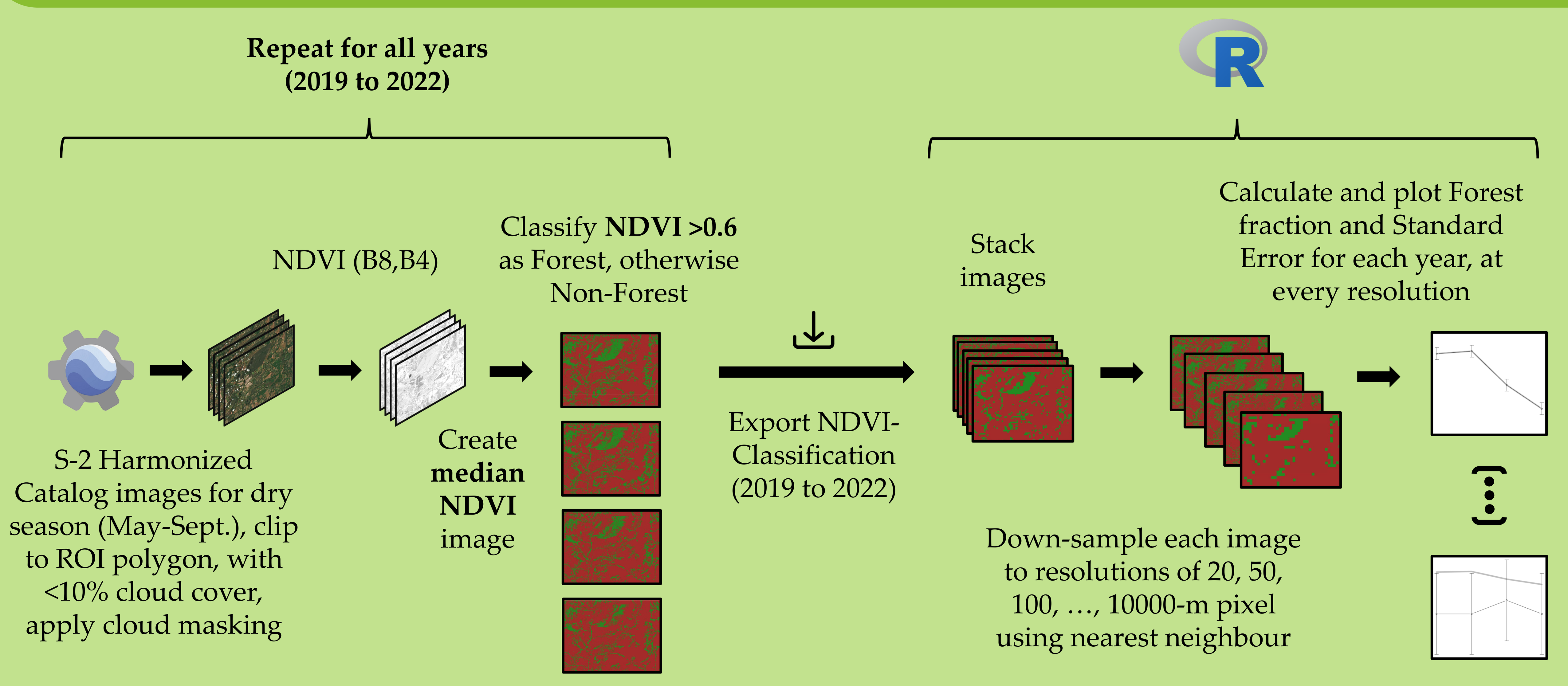


Figure B

The semi-variance follows a **double exponential model**. It rapidly increases until roughly 1000-m, after which it plateaus.

check out the code

Methodology



Down-sampling

We used systematic sampling to down-sample images to 20-, 50-, 100-, 200-, 500-, 1000-, 2000-, 5000- and 10000-m resolution. The original, 10-m image is taken as the "truth", or population value. Down-sampled images take the upper-left (N-W) corner pixel as the new value.

Standard error of mean

Since we used non-random sampling, we used Ripley (1981) Eq. 3.4 [1] *below*, a model-based estimate of the sampling error, which takes the spatial covariance function ***C(u,v)*** of the forest fraction variable as input. We used Monte Carlo integration to estimate the point-block and block-block average covariances.

$$var\left(\bar{Z}-\tilde{Z}(A)\right)=\frac{1}{n^2}\sum_{\boldsymbol{u},\boldsymbol{v}}C(\boldsymbol{u},\boldsymbol{v})-2\sum_{\boldsymbol{u}}\frac{1}{an}\int_A C(\boldsymbol{u},\boldsymbol{y})d\boldsymbol{y}$$
$$+\frac{1}{a^2}\int_A\int_A C(\boldsymbol{x},\boldsymbol{y})d\boldsymbol{x}d\boldsymbol{y}$$

Discussion & Conclusion

- **Down-sampling**, or **systematic sampling**, can give estimates for spatial means that are hardly distinguishable from the full resolution estimates, for our case study for 10m to 1000m resolution, which implies a reduction of the computations with a factor 10^4
- Software to compute associated **standard errors** is not easily available
- We hypothesise that a lot of studies are carried out on full resolution not because it is needed, but because the **consequences** of choosing a lower resolution are not clear
- We tried to carry out the down-sampling in **Google Earth Engine** but failed to get realistic results in reasonable time

References

1. See Eq. 3.4, page 23 in Ripley, B.D. (1981). Spatial Sampling. In Spatial Statistics, B.D. Ripley (Ed.). https://doi.org/10.1002/0471725218.ch3

EGU Abstract