**Supplementary Material S1**

**Input variables selection**

The input variable selection for the vegetation bias removal algorithm comprised two sequential steps. In the first step, 50 variables were selected for initial testing, described in Table S1. All variables are available in the Google Earth Engine (GEE) public catalog (earthengine.google.com)

**Table S1.** Candidate variables analyzed for vegetation bias removal in South America.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Source**  **GEE address** | **Acronym** | **Variable name** | **Spatial resolution (meters)** | **Temporal Aggregation** |
| LANDSAT 8  LANDSAT/LC08/C02/T2\_L2 | SR\_B1 | Ultra blue, coastal aerosol | 30 | Median |
| SR\_B2 | Blue | 30 | Median |
| SR\_B3 | Green | 30 | Median |
| SR\_B4 | Red | 30 | Median |
| SR\_B5 | Near infrared | 30 | Median |
| SR\_B6 | Shortwave infrared 1 | 30 | Median |
| SR\_B7 | Shortwave infrared 2 | 30 | Median |
| ST\_B10 | Surface temperature | 100 | Median |
| NDVI | Normalized difference vegetation index | 30 | Median, interquartile range |
| NDWI | Normalized difference water index | 30 | Median, interquartile range |
| NDMI | Normalized difference moisture index | 30 | Median, interquartile range |
| EVI | Enhanced vegetation index | 30 | Median, interquartile range |
| SR | Simple ratio | 30 | Median, interquartile range |
| MSR | Modified simple ratio | 30 | Median, interquartile range |
| SAVI | Soil adjusted vegetation index | 30 | Median, interquartile range |
| MSAVI | Modified soil adjusted vegetation index | 30 | Median, interquartile range |
| SENTINEL 2  COPERNICUS/S2\_HARMONIZED | B1 | Aerosols | 60 | Median |
| B2 | Blue | 10 | Median |
| B3 | Green | 10 | Median |
| B4 | Red | 10 | Median |
| B5 | Red edge 1 | 20 | Median |
| B6 | Red edge 2 | 20 | Median |
| B7 | Red edge 3 | 20 | Median |
| B8 | NIR | 10 | Median |
| B8A | Red edge 4 | 20 | Median |
| B9 | Water vapor | 60 | Median |
| B10 | Cirrus | 60 | Median |
| B11 | SWIR 1 | 20 | Median |
| B12 | SWIR 2 | 20 | Median |
| NDVI | Normalized difference vegetation index | 10 | Median, interquartile range |
| NDWI | Normalized difference water index | 10 | Median, interquartile range |
| NDMI | Normalized difference moisture index | 20 | Median, interquartile range |
| EVI | Enhanced vegetation index | 10 | Median, interquartile range |
| SR | Simple ratio | 10 | Median, interquartile range |
| MSR | Modified simple ratio | 10 | Median, interquartile range |
| SAVI | Soil adjusted vegetation index | 10 | Median, interquartile range |
| MSAVI | Modified soil adjusted vegetation index | 10 | Median, interquartile range |
| Copernicus DEM GLO-30 (COPDEM)  COPERNICUS/DEM/GLO30 | z | Surface elevation | 30 | None |
| s | Surface slope | 30 | None |
| PALSAR  JAXA/ALOS/PALSAR/YEARLY/SAR | HH | HH polarization backscattering coefficient | 25 | Median |
| HV | HV polarization backscattering coefficient | 25 | Median |
| HV/HH | HV divided by HH | 25 | Median |
| SENTINEL  COPERNICUS/S1\_GRD | VV | Single co-polarization, vertical transmit/vertical receive | 10 | Median |
| VH | Dual-band cross-polarization, vertical transmit/horizontal receive | 10 | Median |

Table S2 presents the index equations used in this study, as well as the input bands for both Landsat 8 and Sentinel 2 images.

**Table S2.** Index formulas and inputs.

|  |  |  |  |
| --- | --- | --- | --- |
| **Index** | **Formula** | **Landsat 8 bands** | **Sentinel 2 bands** |
| NDVI |  | NIR: B5  R: B4 | NIR: B8  R: B4 |
| NDWI |  | NIR: B5  B: B2 | NIR: B8  B: B2 |
| NDMI |  | NIR: B5  SWIR1: B6 | NIR: B8  SWIR1: B11 |
| EVI |  | NIR: B5  R: B4  B: B2 | NIR: B8  R: B4  B: B2 |
| SR |  | NIR: B5  R: B4 | NIR: B8  R: B4 |
| MSR |  | NIR: B5  R: B4 | NIR: B8  R: B4 |
| SAVI |  | NIR: B5  R: B4 | NIR: B8  R: B4 |
| MSAVI |  | NIR: B5  R: B4 | NIR: B8  R: B4 |

The importance of variables was ranked according to the Boruta algorithm (Kursa & Rudnicki, 2010), which uses random forest classification iteratively and selects variables that are statistically shown to be more relevant than random data. The Boruta algorithm was selected due to its low computation time, parsimonious framework, and higher overall efficiency (Speiser et al., 2019). Based on these results, twenty variables were selected as the most statistically relevant for the estimation of vegetation bias from the Copernicus DEM. Among these variables, correlation analysis was performed to identify independent variables (Table S3).

**Table S3.** Selected variables resulting from the Boruta algorithm correlation analysis application.

|  |  |  |  |
| --- | --- | --- | --- |
| Source | Variable | Spatial resolution | Temporal aggregation |
| Landsat 8 | EVI | 30 | Median, interquartile range |
| MSAVI | 30 | Median, interquartile range |
| MSR | 30 | Median, interquartile range |
| NDMI | 30 | Median, interquartile range |
| Sentinel 2 | EVI | 10 | Median, interquartile range |
| MSAVI | 10 | Median, interquartile range |
| MSR | 10 | Median, interquartile range |
| NDMI | 20 | Median, interquartile range |

Finally, visual inspection of several locations in South America was used to assess the combination of variables that yielded acceptable performance within a reasonable computation time.

**References**

Kursa, M. B., & Rudnicki, W. R. (2010). Feature selection with the boruta package. *Journal of Statistical Software*, *36*(11). https://doi.org/10.18637/jss.v036.i11

Speiser, J. L., Miller, M. E., Tooze, J., & Ip, E. (2019). A comparison of random forest variable selection methods for classification prediction modeling. In *Expert Systems with Applications* (Vol. 134). https://doi.org/10.1016/j.eswa.2019.05.028