S2BAVG: EXAMPLE OF SAMPLING DESIGN APPLICATION

1. Sampling Design Methodology

As an example of how S2BAVG can be used, in this document we illustrate a global sampling design for the year 2019 (implemented with the *sampling.py* code). We employ stratified random sampling to represent the diverse conditions influencing the accuracy of BA products. Furthermore, we ensure that the selected units meet minimum cloud cover and land surface criteria to enable the accurate creation of reference perimeters. For clarity, throughout this document we refer to the Sentinel-2 Scene Areas (S2SAs) as the basic spatial units of the S2BAVG framework. These represent the initial population units from which a probability sample is drawn. The units selected through the sampling design are referred to as sample units.

1.1 Stratification: Biome and fire activity

Two levels of stratification were applied to all S2SAs in 2019 using S2BAVG attributes. The first level was based on the predominant aggregated biome, following the methodology of Franquesa et al. (2022), while the second level was determined by the percentage of burned area on the land surface, derived from the FireCCI51 product. Following Padilla et al. (2014), we sorted the units within each biome in ascending order based on their burned area percentage. Those exceeding the 80th percentile were assigned to the high fire activity stratum, while the remaining units were assigned to the low fire activity stratum.

1.2 Population units: Cloud and land cover filtering

To ensure optimal conditions for generating reference perimeters, S2SAs were filtered based on land area and cloud cover criteria to exclude those that could hinder land observation and burned area detection. Units with more than 50% land area and a median interval of 10 days or less between consecutive cloud-free images were considered valid for sampling, while the rest were discarded. The proportion of the area of discarded units should be reported as this area is no longer included in the population represented by the sample estimates.

1.3 Sample size allocation

The total number of units to be sampled for each stratum was determined using the allocation formula (Eq. 1) proposed by Padilla et al. (2017):

$$n_h \propto N_h \sqrt{\overline{BA}_h} \quad (1)$$

where for each stratum h the number n_h of units to be sampled is proportional to the total number of units N_h available for sampling after the filtering (section 1.2), multiplied by the square root of the mean burned area percentage $\overline{BA}_h$ of those units. For our sampling practical demonstration using the S2BAVG grid, we initially set the total sample size to n = 100, consistent with previous global BA validation (Franquesa, Stehman, et al., 2022; Padilla et al., 2017)and based on the practical effort required for reference data generation. A minimum of $n_h = 2$ was imposed per stratum to allow variance estimation for each group which could increase the final sample size beyond 100.

2. Sampling Design Example

The 19,263 S2SAs were stratified for the year 2019. Table 1 shows the 80th percentile threshold value used to assign each unit into its corresponding fire activity stratum within each biome.

Table 1. The threshold for stratifying high/low fire activity within each biome defined by the 80th percentile of the BA percentage according to FireCCI51 product across the terrestrial surface.

Biome	**Threshold (BA %)**
Boreal Forest	0.22
Deserts & Xeric Shrublands	0.02
Mediterranean	0.60
Temperate Forest	0.41
Temperate Savanna	0.73
Tropical Forest	1.65
Tropical Savanna	26.44
Tundra	0.00

Table 2 provides the total number of units per stratum in the S2BAVG for 2019, along with the number of units remaining after land and cloud cover filtering and the final number of selected units following sample allocation. After applying the filtering criteria, 11,301 units were designated as population units, representing 59% of the total, while 41% were excluded due to excessive cloud cover or insufficient land area. Although the initial sample size was set at n=100, a total of 111 sample units were ultimately selected to meet the requirement of at least nh=2 per stratum.

Table 2. The total number of units per stratum before and after land and cloud cover filtering (Nh) and at the conclusion of the sampling (nh). The percentages in parentheses indicate the proportion of S2SAs discarded after applying land and cloud cover filters, relative to the initial number of units in each stratum.

Biome	Before filtering		After filtering, N_h		n_h	
	high	low	high	low	high	low
Boreal Forest	388	1,558	135 (65%)	559 (64%)	2	2
Deserts & Xeric Shrublands	694	2,949	479 (31%)	2,298 (22%)	5	2
Mediterranean	122	489	69 (43%)	264 (46%)	2	2
Temperate Forest	441	1,797	286 (35%)	771 (57%)	4	2
Temperate Savanna	392	1,565	253 (35%)	1,029 (34%)	4	2
Tropical Forest	842	3,384	578 (31%)	1,235 (63%)	11	5
Tropical Savanna	629	2,517	585 (7%)	2,035 (19%)	28	36
Tundra	142	1,354	29 (80%)	696 (49%)	2	2
Global	**3,650**	**15,613**	**2,414 (34%)**	**8,887 (43%)**	**58**	**53**

Figure 1 shows the S2SAs that remained after applying the land and cloud-cover filtering, along with the 111 units selected for the global sample in 2019. Consistent with Table 2, it highlights extensive regions without valid S2SAs in Boreal, Tundra and Tropical Forest biomes. The sample allocation intentionally increases the sample in strata with greater burned area, but

the stratified estimators are unbiased as the unequal inclusion probabilities associated with this allocation are incorporated in the estimators.

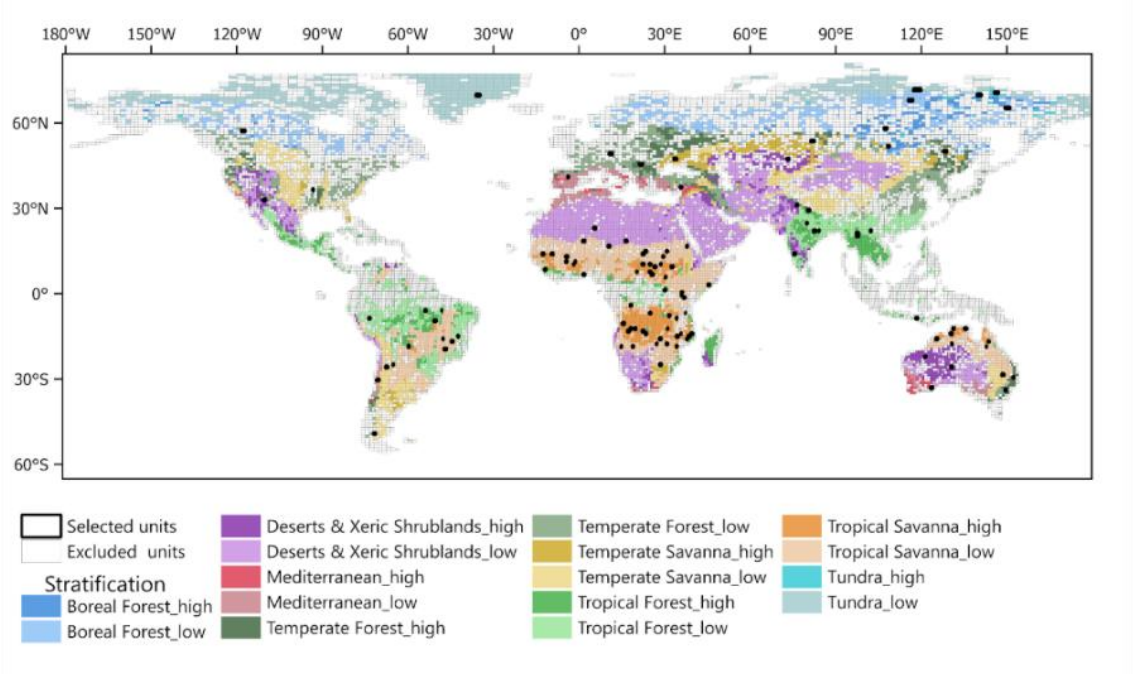

Figure 1. The global validation grid for Sentinel-2 (S2BAVG), stratified by biomes and fire activity level, for the year 2019 after land and cloud cover filtering. S2SAs that do not meet these criteria are classified as excluded (grey), while the final selected sample units are displayed in black.

The stratified, filtered and selected S2SAs in this application example are available in the *sampling_firecci51_2019.gpkg* file.

REFERENCES

Franquesa, M., Lizundia-Loiola, J., Stehman, S. V., & Chuvieco, E. (2022). Using long temporal reference units to assess the spatial accuracy of global satellite-derived burned area products. *Remote Sensing of Environment*, *269*. https://doi.org/10.1016/j.rse.2021.112823

Franquesa, M., Stehman, S. V., & Chuvieco, E. (2022). Assessment and characterization of sources of error impacting the accuracy of global burned area products. *Remote Sensing of Environment*, *280*. https://doi.org/10.1016/j.rse.2022.113214

Padilla, M., Olofsson, P., Stehman, S. V., Tansey, K., & Chuvieco, E. (2017). Stratification and sample allocation for reference burned area data. *Remote Sensing of Environment*, *203*, 240–255. https://doi.org/10.1016/j.rse.2017.06.041

Padilla, M., Stehman, S. V., & Chuvieco, E. (2014). Validation of the 2008 MODIS-MCD45 global burned area product using stratified random sampling. *Remote Sensing of Environment*, *144*, 187–196. https://doi.org/10.1016/j.rse.2014.01.008