

This manuscript presents a validation dataset developed for the Land Cover Map of Europe 2017 within the Sentinel-2 Global Land Cover project. The dataset contains more than 50,000 reference samples distributed across Europe, aligned with 10 m Sentinel-2 pixels and classified into 13 land-cover/land-use classes. The study used a two-stage sampling strategy, first selecting Sentinel-2 tiles by country and then applying stratified random sampling within the selected tiles using CORINE Land Cover data. Candidate samples were subsequently interpreted using independent, higher-resolution imagery and checked by additional experts.

The dataset is potentially useful for validating European land-cover products, comparing classification methods, and supporting broader reference-data compilations. The topic is suitable for ESSD, and the public release of an independent validation dataset is valuable. However, several aspects of interpretation protocol, quality assessment, and data usability require substantial clarification or additional analysis before the manuscript can fully support the reliability and reuse claims made by the authors. I recommend major revision.

Major comments:

1. Expert interpretation and inter-interpreter consistency are insufficiently documented.

The study states that candidate samples were visually interpreted by experienced and trained interpreters and were subsequently verified by independent specialists. It also states that uncertain samples were removed. However, the manuscript does not provide sufficient information about the actual interpretation and verification process. The study should clarify the following points:

How many interpreters participated?

Did each interpreter classify a separate subset of samples, or were some samples interpreted independently by multiple experts?

Was the independent verification applied to all retained samples or only to a subset?

What criteria were used to accept, revise, or reject an initial label?

How were disagreements between interpreters resolved?

Was a senior expert responsible for final adjudication?

Most importantly, the manuscript does not report any quantitative measure of agreement among experts. Since the reliability of the dataset depends strongly on visual interpretation, the authors should provide an inter-interpreter consistency assessment. This could include the percentage agreement and an appropriate agreement statistic, such as Cohen's kappa or Fleiss' kappa, based on a common subset classified independently by multiple interpreters.

2. The available attributes of data is limited. In the table 4, the data only has three attributes: S2GLC, TILE, and NAME_ENG. These fields allow basic use, but they provide limited traceability and constrain reproducibility. For a reference dataset intended for broad reuse, the authors should consider adding or releasing supplementary metadata, such as reference image date or date range, country code, confidence level, geographic coordinates, etc.

3. The comparison with current datasets. The manuscript briefly discusses limitations of existing reference datasets, but the comparison is general. Since the main contribution is the dataset rather than a new methodological framework, the manuscript should provide a more explicit comparison with existing European or global reference datasets (Lesiv et al., 2025, <https://doi.org/10.5194/essd-17-6149-2025>).

Minor comments:

Line 12. "In this first step, validation sites (Sentinel-2 tiles)" The first step is to select the tiles instead of the site samples. Please clarify this terminology through the manuscript (Line 42,..). The manuscript alternates among "validation sites," "samples," "sample units," "validation samples," "tiles," and "granules." These terms should be defined once and used consistently. primary sampling unit is Sentinel-2 tile. secondary sampling unit is at 10 m pixel;

Lines 17. Please add the URL of dataset at the end of abstract easy downloading.

Line 46. What is the EO data ?

Lines 87. Typo-fromthe. Same issue in Figure 2 caption – validationssites.

Lines 93-94: Błąd! Nie można odnaleźć źródła odwołania. Błąd! Nie można

odnaleźć źródła odwołania. Błąd! Nie można odnaleźć źródła odwołania. Correct this.

Table 1. shows the selected tiles instead of sites. Clarify this. Also define the “base” site.

Line 123. Figure 4, Add the legend showing the land cover class of each color.

In Table 3. Does “validation sites” mean Sentinel-2 tiles?