**Readme file for the resampled Soil Spectral Libraries**

This package contains ten soil spectral datasets that simulated two current multispectral (Landsat 8 OLI, Sentinel-2 MSI, both 2a and 2b) and two forthcoming hyperspectral (EnMAP, CHIME) satellite sensors. The LUCAS15 and Brazilian soil spectral libraries (SSLs) were utilized to produce two different sets of “resampled” spectra. A first set of spectra was obtained by re-sampling the SSLs, using convolution procedures, to the specific spectral response of the Sentinel-2 (a&b) and Landsat8. This process was carried out using the associated spectral response functions. On the other hand, for EnMAP we utilized Gaussian functions.

1 Satellite sensors

The main technical specifications of the satellite spectral sensors are summarized, in Table 1.

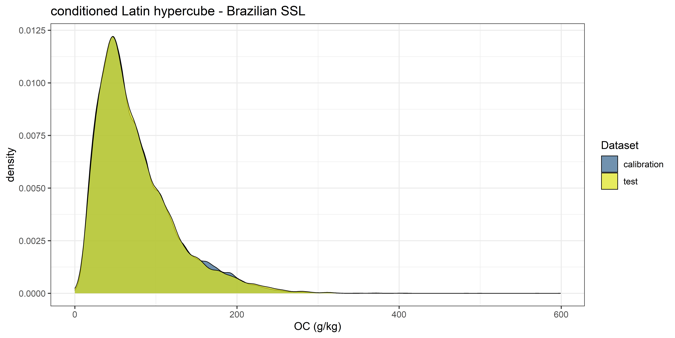
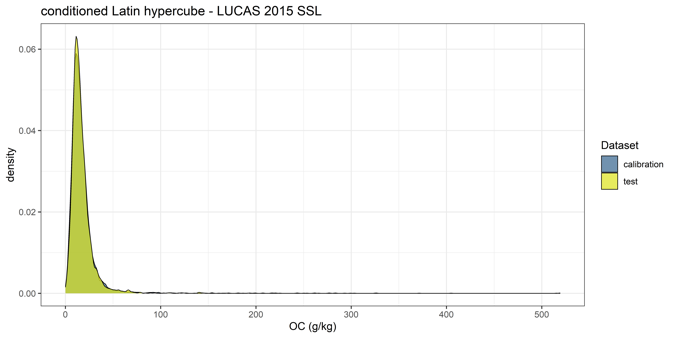
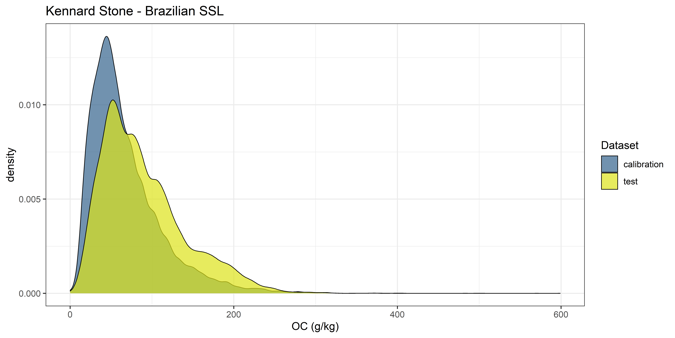
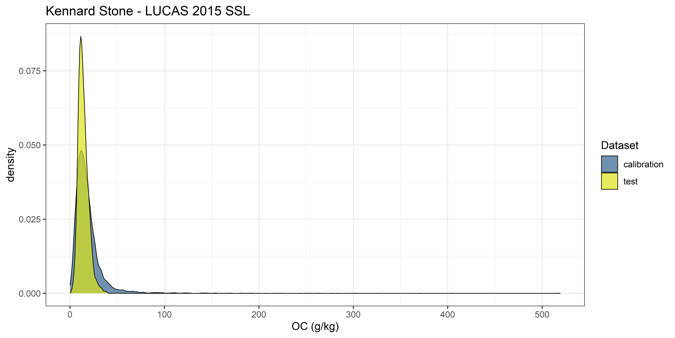
Table 1: Main technical characteristics of the spaceborne sensors considered in this study. The FWHM column indicates the minimum and the maximum of the sensor's bandwidth

|  |  |  |  |
| --- | --- | --- | --- |
| Sensor | Spectral bands | Spectral range | FHWM (nm) |
| Sentinel-2MSI | 12 | 9VNIR-3SWIR | 10-60 |
| Landsat 8 OLI | 7 | 5VNIR-2SWIR | 20-200 |
| EnMAP | 242 | 420-2450 | 10 |
| CHIME | 295 | 400-2500 | ~7nm |

The resampling analyses were performed using the hsdar (Lehnert et al., 2019) package developed in R software (R Development Core Team, 2013).

2 Data partitions

The LUCAS 2015 and Brazilian datasets were split into calibration (66%) and test (33%) sets, using the Conditioned Latin Hypercube Sampling algorithm (Minasny & McBratney, 2006), which is an effective way to replicate the multivariate distribution of the input space in the calibration set. This will enable the learning algorithms to construct their models using data which are representative in the spectral space, compared to Kennard Stone split.



3 Description of Datasets

lucas15: [1] EnMAP\_resampled\_lucas15.csv, [2] L8\_resampled\_lucas15.csv, [3] S2a\_resampled\_lucas15.csv, [4] S2b\_resampled\_lucas15.csv, [5] CHIME\_resampled\_lucas15.csv

|  |  |  |
| --- | --- | --- |
| **Field** | **Description** | **Units/**  **Values** |
| PointID | Unique identifier of the LUCAS survey point | 8 digits number |
| OC | Organic carbon content | g/kg |
| [1] B\_423.03…B\_2438.6 | Re-sampled reflectance data with ~10nm resolution | Reflectance (%) |
| [2] B\_444.5… B\_2194.5 | Re-sampled reflectance data with ~20-200nm resolution | Reflectance (%) |
| [3] B\_442.7 … B\_2202.4 | Re-sampled reflectance data with ~10-60nm resolution | Reflectance (%) |
| [4] B\_442.3 … B\_2185.7 | Re-sampled reflectance data with ~10-60nm resolution | Reflectance (%) |
| [5] 421…2479 | Re-sampled reflectance data with ~7nm resolution | Reflectance (%) |

BSSL: [1] EnMAP\_resampled\_bssl.csv, [2] L8\_resampled\_bssl.csv, [3] S2a\_resampled\_bssl.csv, [4] S2b\_resampled\_bssl.csv, [5] CHIME\_resampled\_bssl.csv

|  |  |  |
| --- | --- | --- |
| **Field** | **Description** | **Units/**  **Values** |
| PointID | Unique identifier of the Brazilian survey point | 4-5 digits number |
| OC | Organic carbon content | % |
| [1] B\_423.03…B\_2438.6 | Re-sampled reflectance data with ~10nm resolution | Reflectance (%) |
| [2] B\_444.5… B\_2194.5 | Re-sampled reflectance data with ~20-200nm resolution | Reflectance (%) |
| [3] B\_442.7 … B\_2202.4 | Re-sampled reflectance data with ~10-60nm resolution | Reflectance (%) |
| [4] B\_442.3 … B\_2185.7 | Re-sampled reflectance data with ~10-60nm resolution | Reflectance (%) |
| [5] 421…2479 | Re-sampled reflectance data with ~10nm resolution | Reflectance (%) |

Description of the fields in the clhs\_lucas15.csv file:

|  |  |  |
| --- | --- | --- |
| **Field** | **Description** | **Units/**  **Values** |
| PointID | Unique identifier of the LUCAS survey point | 8 digits number |
| class | Class indicating where each sample belongs cal or val datasets | Calibration/test |

Description of the fields in the clhs\_brazilian.csv file:

|  |  |  |
| --- | --- | --- |
| **Field** | **Description** | **Units/**  **Values** |
| PointID | Unique identifier of the Brazilian survey point | 4-5 digits number |
| class | Class indicating where each sample belongs cal or val datasets | Calibration/test |

References

Lehnert, L. W., Meyer, H., Obermeier, W. A., Silva, B., Regeling, B., Thies, B., & Bendix, J. (2019). Hyperspectral Data Analysis in {R}: The {hsdar} Package. *Journal of Statistical Software*, *89*(12), 1–23. https://doi.org/10.18637/jss.v089.i12

Minasny, B., & McBratney, A. B. (2006). A conditioned Latin hypercube method for sampling in the presence of ancillary information. *Computers & Geosciences*, *32*(9), 1378–1388. https://doi.org/https://doi.org/10.1016/j.cageo.2005.12.009

R Development Core Team. (2013). R: A language and environment for statistical computing. *R Foundation for Statistical Computing, Vienna, Austria*.