“**GSFC-Cloud validation dataset for Sentinel-2 and Landsat 8 satellites: proof of concept**”

Sergii Skakun a,b,c, Eric F. Vermote c, Andres Eduardo Santamaria Artigas a,c,  
William H. Rountree a,c, Jean-Claude Roger a,c

a*Department of Geographical Sciences, University of Maryland, College Park, MD 20742, USA*

b*College of Information Studies (iSchool), University of Maryland, College Park, MD 20742, USA*

c *NASA Goddard Space Flight Center Code 619, 8800 Greenbelt Road, Greenbelt, MD 20771, USA*

**Description of data**

Reference cloud data are provided in the vector form as shapefiles. Shapefiles are in the geographical projection:

epsg:4326 or proj4= +proj=longlat +ellps=WGS84 +datum=WGS84 +no\_defs

Each shapefile has the following naming convention:

{sat\_id}\_{date}.shp,

where sat\_id can be s2a, s2b and lc8 for Sentinel-2A, Sentinel-2B and Landsat 8, respectively, date is in the for YYYYMMDD, for example:

s2a\_20170923.shp

Each file contains polygons corresponding to one of the following classes:

* 1 — clear observations (no clouds)
* 2 — clouds
* 3 — cirrus/transparent clouds
* 4 — shadows