Supporting the ESA SAR4Citizen and Urban-TIP projects, the German Aerospace Center (DLR) has lately produced the World Settlement Footprint 2015 (WSF2015), i.e. a beam-on-the-ground extent of human settlements globally derived by means of 2014-2015 multitemporal Landsat-8 and Sentinel-1 IW GRD imagery acquired at 30m and 10m resolution, respectively.

For quantitatively assessing the high accuracy and reliability of the layer DHR has recently carried out in collaboration with Google an unprecedented validation exercise based on a huge amount of ground-truth samples labeled by crow-sourcing photointerpretation. In particular, to this purpose a statistically robust and transparent protocol has been defined following the state-of-the-art practices currently recommended in the literature.

Furthermore, the assessment figures have been compared to those derived in previous efforts currently employed similar global datasets, including: the Global Urban Footprint – GUF (available at 12m resolution and referring to the year 2012), the Global Human Settlement Layer – GHSL (available at 30m resolution and referring to the year 2014), and the Global Settlement Layer – GSL (available at 30m resolution and referring to the year 2015).

INTRODUCTION

RESPONSE DESIGN

The response design encompasses all steps leading to a decision regarding agreement of the reference and map classifications. The four major features of the designed protocol are:

- **Source of Reference Data:** Google Earth (GE) satellite/aerial VHR imagery has been used. The spatial resolution varies depending on the specific data source; in the case of SPOT imagery it is ~1.5m, for Digital Globe’s WorldView-2/3 series, GeoEye-1, and Airbus’ WorldView, it is in the order of ~0.5m resolution, whereas for airborne data (primarily available for North America, Europe and Japan) it is about 0.15m.

- **Spatial Assessment Units:** for North America, Europe and Japan it is about 0.15m - classification: majority rule over entire 3x3 block

- **Assessment Criterion:**
  - **Settlements = Buildings if the given cell intersects any building:**
  - **Settlements = Roads/Paved Surfaces:** if the given cell intersects any road/paved surface or no buildings, or buildings.

- **Definition of Agreement:** Given the classification and the reference labels derived as described above, 4 different agreement criteria have been defined and tested against 3 different settlement definitions.

- **Crowd-sourcing was performed internally at Google.** Specifically, by means of an an ad-hoc tool operating has been created and tested a 3x3 grid on top of the available GE reference VHR scene closest in time to the year 2015 and given the possibility of assigning of the 4 labels to each cell.

- **Settlements = Buildings if the given cell intersects any building:**

- **Settlements = Roads/Paved Surfaces:** if the given cell intersects any road/paved surface or no buildings. or buildings.

As the settlement class covers a small sensible proportion of area compared to the merger of all other non-settlement classes (~1% of Earth’s emerged surface), an equal allocation reduces the standard error of its class-specific accuracy. Moreover, such an approach allows to best address user's accuracy estimation, which corresponds to the map "reliability" and is indicative of the probability that a pixel classified on the map actually represents the corresponding category on the ground.

Accordingly, for each of the 51 selected tiles we randomly extracted 1000 settlement and 1000 non-settlement samples from the WSF2015 and used these as centre cells of the 3x3 block assessment units to be labelled by photointerpretation.

Such a strategy resulted in an overall amount of (1000 + 1000) x 9 x 50 = 900,000 cells labelled by the crowd.

SANGLING DESIGN

Stratified random sampling design has been applied since it satisfies the basic accuracy assessment objectives and most of the desirable design criteria. In particular, stratified random sampling is a probability sampling design and it is one of the easier to implement; indeed, it involves first the division of the population into strata within which random sampling is performed afterwards. To include a representative population of settlement patterns, 50 tiles of 1°x1° degree size (out of the ~18,000 comprising the WSF2015) have been selected based on the ratio between the number of estimated settlements (i.e., digitized clusters of polygons categorized as settlement in the WSF2015) and their area. In particular, the nth selected tile has been chosen randomly among those whose ratio belongs to the interval [1/n² - 1/(n+1)² x 1/(518 x 11)] (where n denotes the n-th percentile of the ratio).

The settlement class covers a small sensible proportion of area compared to the merger of all other non-settlement classes (~1% of Earth’s emerged surface), an equal allocation reduces the standard error of its class-specific accuracy. Moreover, such an approach allows to best address user's accuracy estimation, which corresponds to the map "reliability" and is indicative of the probability that a pixel classified on the map actually represents the corresponding category on the ground.

Accordingly, for each of the 51 selected tiles we randomly extracted 1000 settlement and 1000 non-settlement samples from the WSF2015 and used these as centre cells of the 3x3 block assessment units to be labelled by photointerpretation.