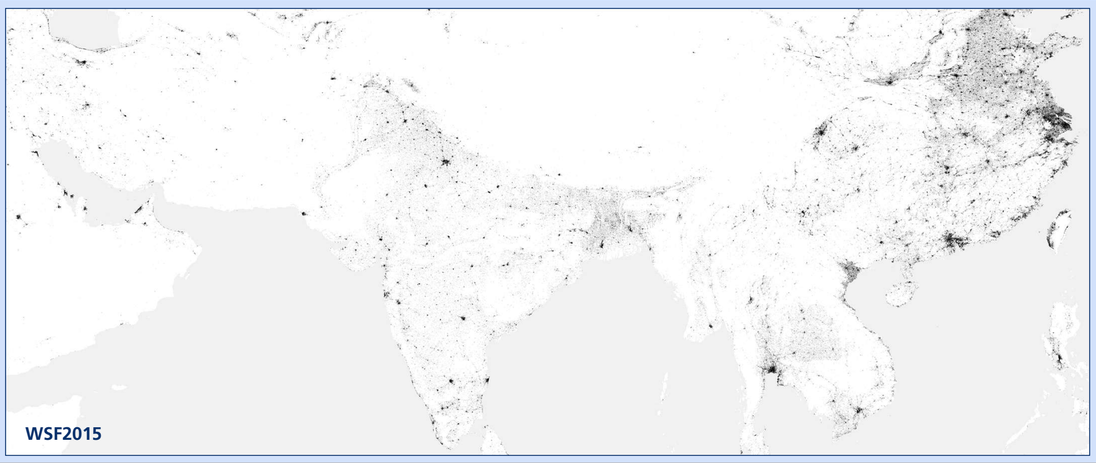
On the Quality Assessment of the World Settlement Footprint 2015

Mattia Marconcini[1], Daniela Palacios-Lopez[1], Wiebke Vogt[1], Felix Bachofer[1], Annekatrin Metz-Marconcini[1], Thomas Esch[1], Allison Lieber[2], Ashwin Kakarla[2], Noel Gorelick[2]

[1] German Aerospace Center (DLR), Weßling, Germany

[2] Google, Inc., Mountain View, CA, USA



INTRODUCTION

Supported by the ESA SAR4Urban and Urban-TEP projects, the German Aerospace Center (DLR) has lately produced the World Settlement Footprint (WSF) 2015, i.e. a binary mask outlining the extent of human settlements globally derived by means of 2014-2015 multitemporal Landsat-8 and Sentinel-1 IW GRDH imagery acquired at 30m and 10m resolution, respectively.

For quantitatively assessing the high accuracy and reliability of the layer DLR has recently carried out in collaboration with Google an unprecedented validation exercise based on a huge amount of ground-truth samples labelled 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.

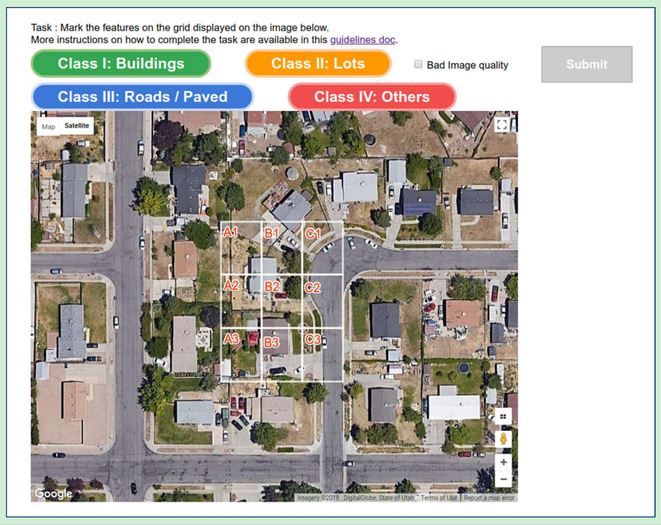
Furthermore, the assessment figures have been compared to those derived for other currently employed similar global datasets, including: the Global Urban Footprint – GUF (available at 12m resolution and referring to the year 2012), the Global Human Settlement Layer – GHSL (available at 30m resolution and referring to the year 2014), and the GLOBELAND30 – GLC30 (available at 30m resolution and referring to the year 2010).

RESPONSE DESIGN

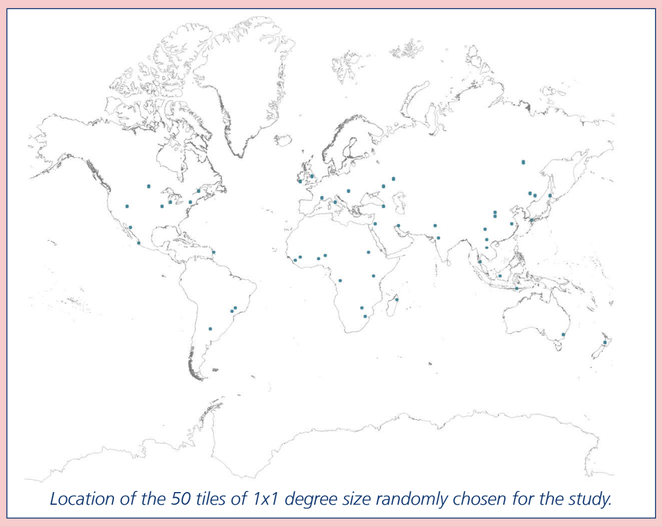
The response design encompasses all steps leading to a decision regarding agreement of the reference and map classifications. The four major features of the designed protocol are:

- **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-1/2 series, GeoEye-1, and Airbus' Pleiades it is in the order of ~0.5m resolution, whereas for airborne data (mostly available for North America, Europe and Japan) it is about 0.15m.
- **Spatial Assessment Unit:** since input data with different spatial resolutions have been employed to generate the WSF2015 (i.e., 30m Landsat-8 and 10m Sentinel-1 IW GRDH), a **3x3 block spatial assessment unit composed of 9 cells of 10x10m** size was used.
- **Reference Labelling Protocol:** to cope with the different existing definitions of settlement, 4 possible labels were collected, namely:
  - → ***Buildings*** if the given cell intersects any building;
  - → ***Building Lots***: if the given cell intersects any building lot and no buildings;
  - → ***Roads/Paved-Surfaces:*** if the given cell intersects any road/paved surface and no buildings or building lots;
  - → ***None-of-the-previous***.
- **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 *ad-hoc* tool, operators have been iteratively prompted a 3x3 unit on top of the available GE reference VHR scene closest in time to the year 2015 and given the possibility of assigning any of the 4 labels to each cell.



credits: Google



Location of the 50 tiles of 1x1 degree size randomly chosen for the study.

SAMPLING 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 1x1 degree size** (out of the ~18.000 composing the WSF2015) have been selected based on the ratio between the number of estimated settlements (i.e., disjoint clusters of pixels categorized as settlement in the WSF2015) and their area. In particular, the *i*-th selected tile has been chosen randomly among those whose ratio belongs to the interval $]P_{2(i-1)}; P_{2i}], i \in [1; 50] \subset \mathbb{N}$ (where P_x denotes the *x*-th percentile of the ratio).

As the settlement class covers a sensibly small 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 50 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.

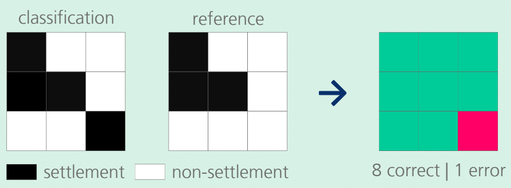
Such a strategy resulted in an overall amount of

$$(1000 + 1000) \times 9 \times 50 = \mathbf{900.000}$$

cells labelled by the crowd.

ANALYSIS

Assessment Criterion 1

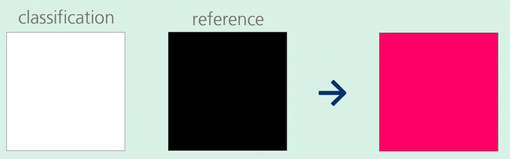


per-cell matching

Assessment Criterion 2

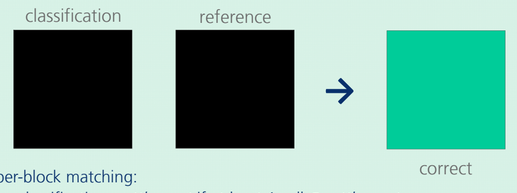
per-block matching:
- classification: majority rule over entire 3x3 block
- reference: majority rule over entire 3x3 block

Assessment Criterion 3



per-block matching:
- classification: majority rule over entire 3x3 block
- reference: *settlement* if at least 1 cell ∈ *settlement*

Assessment Criterion 4



per-block matching:
- classification: *settlement* if at least 1 cell ∈ *settlement*
- reference: *settlement* if at least 1 cell ∈ *settlement*

Settlement = Buildings

Global Layer	Average Accuracy %	Kappa	Producer's Accuracy %		User's Accuracy %	
			Settlement	Non-Settlement	Settlement	Non-Settlement
WSF2015	**83.27**	0.5486	91.99	74.56	59.41	96.40
GUF	79.52	**0.5702**	72.41	86.64	65.31	90.04
GHSL	70.38	0.4430	49.54	91.23	66.24	83.89
GLC30	67.41	0.3921	42.76	92.05	65.14	82.24

Global Layer	Average Accuracy %	Kappa	Producer's Accuracy %		User's Accuracy %	
			Settlement	Non-Settlement	Settlement	Non-Settlement
WSF2015	**84.49**	0.5750	93.82	75.16	57.72	97.11
GUF	81.82	**0.6177**	75.80	87.84	69.25	90.94
GHSL	72.34	0.4775	52.39	92.30	72.46	83.98
GLC30	68.64	0.4218	44.26	93.02	69.62	82.20

Global Layer	Average Accuracy %	Kappa	Producer's Accuracy %		User's Accuracy %	
			Settlement	Non-Settlement	Settlement	Non-Settlement
WSF2015	**88.55**	**0.7651**	88.60	88.50	84.28	91.77
GUF	79.60	0.6194	63.95	95.24	90.34	79.15
GHSL	69.43	0.4226	42.18	96.68	89.84	70.60
GLC30	66.38	0.3597	36.19	96.57	88.01	68.49

Global Layer	Average Accuracy %	Kappa	Producer's Accuracy %		User's Accuracy %	
			Settlement	Non-Settlement	Settlement	Non-Settlement
WSF2015	**89.04**	**0.7711**	90.46	87.62	83.57	92.95
GUF	82.16	0.6635	70.70	93.62	88.52	82.11
GHSL	73.19	0.4958	50.98	95.40	88.53	73.65
GLC30	68.74	0.4069	41.84	95.65	87.00	70.26

Settlement = Buildings + Building Lots

Global Layer	Average Accuracy %	Kappa	Producer's Accuracy %		User's Accuracy %	
			Settlement	Non-Settlement	Settlement	Non-Settlement
WSF2015	**84.14**	**0.5996**	91.08	77.20	61.99	95.49
GUF	79.68	0.5961	70.72	86.63	71.75	88.11
GHSL	70.46	0.4543	48.34	92.59	72.71	81.44
GLC30	67.59	0.4011	41.89	93.28	71.79	79.72

Global Layer	Average Accuracy %	Kappa	Producer's Accuracy %		User's Accuracy %	
			Settlement	Non-Settlement	Settlement	Non-Settlement
WSF2015	**85.37**	0.6278	92.73	78.01	64.29	96.17
GUF	81.63	**0.6378**	73.39	89.87	75.56	88.78
GHSL	71.62	0.4816	49.58	93.67	76.97	81.32
GLC30	68.56	0.4241	42.88	94.23	76.02	79.45

Global Layer	Average Accuracy %	Kappa	Producer's Accuracy %		User's Accuracy %	
			Settlement	Non-Settlement	Settlement	Non-Settlement
WSF2015	**88.75**	**0.7737**	87.67	89.84	86.47	90.77
GUF	79.37	0.6125	62.79	95.94	91.98	77.68
GHSL	69.42	0.4180	41.57	97.26	91.84	69.20
GLC30	66.47	0.3572	35.79	97.14	90.27	67.12

Global Layer	Average Accuracy %	Kappa	Producer's Accuracy %		User's Accuracy %	
			Settlement	Non-Settlement	Settlement	Non-Settlement
WSF2015	**89.33**	**0.7822**	89.61	89.06	85.85	92.04
GUF	82.08	0.6613	69.63	94.53	90.41	80.77
GHSL	73.25	0.4935	50.35	96.16	90.67	72.33
GLC30	68.90	0.4062	41.45	96.35	89.37	68.95

Settlement = Buildings + Building Lots + Roads / Paved Surfaces

Global Layer	Average Accuracy %	Kappa	Producer's Accuracy %		User's Accuracy %	
			Settlement	Non-Settlement	Settlement	Non-Settlement
WSF2015	**83.65**	**0.6429**	86.13	81.17	71.41	91.46
GUF	78.52	0.5963	65.48	91.57	80.92	82.93
GHSL	70.64	0.4603	45.98	95.31	84.26	76.36
GLC30	67.86	0.4051	40.03	95.70	83.55	74.50

Global Layer	Average Accuracy %	Kappa	Producer's Accuracy %		User's Accuracy %	
			Settlement	Non-Settlement	Settlement	Non-Settlement
WSF2015	**85.26**	**0.6801**	87.77	82.75	74.70	92.10
GUF	80.24	0.6325	67.32	93.15	85.09	83.09
GHSL	71.45	0.4765	46.41	96.48	88.45	75.62
GLC30	68.50	0.4174	40.29	96.71	87.68	73.62

Global Layer	Average Accuracy %	Kappa	Producer's Accuracy %		User's Accuracy %	
			Settlement	Non-Settlement	Settlement	Non-Settlement
WSF2015	**86.85**	**0.7396**	81.23	92.48	90.99	84.04
GUF	76.98	0.5466	56.94	97.02	94.71	70.66
GHSL	68.50	0.3774	38.39	98.62	96.29	63.11
GLC30	65.91	0.3250	33.32	98.50	95.42	61.23

Global Layer	Average Accuracy %	Kappa	Producer's Accuracy %		User's Accuracy %	
			Settlement	Non-Settlement	Settlement	Non-Settlement
WSF2015	**87.69**	**0.7558**	83.38	92.01	90.71	85.54
GUF	79.86	0.6034	63.63	96.08	93.82	73.85
GHSL	72.37	0.4549	46.75	97.99	95.61	66.30
GLC30	68.40	0.3752	38.76	98.05	94.90	63.12

DLR