European Commission

JRC SCIENTIFIC AND POLICY REPORTS

Global Human Settlement Layer/ Urban Atlas Integration: Feasibility Report

Application of the JRC GHSL Image Information Extraction Protocol in the frame of the Urban Atlas product specifications

Martino Pesaresi
Matina Halkia

2012

Report EUR 25328 EN

Joint Research Centre

European Commission
Joint Research Centre
Institute for the Protection and Security of the Citizen

Contact information
Martino Pesaresi
Address: Joint Research Centre, Via Enrico Fermi 2749, TP 267, 21027 Ispra (VA), Italy
E-mail: martino.pesaresi@jrc.ec.europa.eu
Tel.: +39 0332 78 9524
Fax: +39 0332 78 5154

http://isferea.jrc.ec.europa.eu/
http://www.jrc.ec.europa.eu/

JRC69962

EUR 25328 EN

ISBN 978-92-79-24933-4

ISSN 1831-9424

doi:10.2788/26820

Luxembourg: Publications Office of the European Union, 2012

Printed in Italy

European Commission, The Joint Research Centre

GHSL/UA Integration: Feasibility Report

Application of the JRC GHSL Image Information Extraction Protocol in the frame of the Urban Atlas Product Specifications.

Authors: Martino Pesaresi, Matina Halkia –
Contact: martino.pesaresi@jrc.ec.europa.eu
3/21/2012

FEASIBILITY REPORT

APPLICATION OF THE JRC GHSL IMAGE INFORMATION EXTRACTION PROTOCOL IN THE FRAME OF THE URBAN ATLAS PRODUCT SPECIFICATIONS.

March 21, 2012
[2012_GHSL_UA_v7.docx]

Table of Contents

1 Executive Summary

The Joint Research Centre, in pursuing its mission of scientific and technical support for policy making in Brussels, has planned and carried out a series of tests in collaboration with DG Regional Policy, in view of a possible integration of its in-house Global Human Settlement Layer (GHSL) technology to the European Urban Atlas (UA). The support is based on specific parameters derived from the application of the satellite-based methodology developed by the JRC for human settlement analysis. The Urban Atlas provides detailed and cost-effective digital mapping, ensuring that city planners have the most up-to-date and accurate data available, offering new tools to assess risks and opportunities, ranging from threat of natural disasters and impact of climate change, to identifying new infrastructure and public transport needs. The GHSL/UA integration would contribute to population disaggregation and risk and disaster management applications, as well as support regional planning in general.

The premises of this test were: i) to evaluate the computational feasibility for a wall-to-wall (pan-European) coverage of a GHSL layer for the UA; ii) to control the reliability of the output in relation to a well-known and accepted reference layer, used in UA, namely the the European Soil Sealing Layer (SSL); iii) to estimate the cost of a possible integration of GHSL in the UA.

The main findings of this experiment were that: the integration of GHSL in UA is computationally feasible; it is thematically reliable when compared to the SSL; and it is cost-effective. As a result, we believe that the GHSL integration in UA could potentially provide a technological means for the continuous monitoring of European urban areas (or to be precise, any European inhabited area of any size) in addition to the existing information provided by the SSL.

More specifically, our findings emphasize that the agreement between the GHSL and SSL outputs is more than 90%. When discrepancies are reported, these pertain to inherent technical characteristics of each system, like scale and generalization parameters, in addition to semantic differences in the measurements reported. In practice, where one system errs the other detects and vice versa. Although the GHSL layer reports on built-up areas and SSL reports on sealed surface areas, and therefore one might argue that the former is semantically more appropriate for monitoring urban change in support to regional planning, we sustain that the two layers should complement each other.

Currently the GHSL output has been evaluated in terms of its performance in reporting built-up areas only. We sustain that the added value of GHSL integration in UA will become more evident once the characterization of the built-up areas detected is complete, a test which is under way at the moment by the JRC team.

In addition, what makes the GHSL technology unique is that it requires little human intervention, given that it involves automatic extraction of information from satellite images. More specifically, the costs incurred for a possible full integration of the GHSL technology in UA would involve infrastructure provisions (data acquisition, hardware and memory storage); the costs for data input quality management, ingestion, and system integration (preparation and data processing); and last but not least, validation and integration of output for appropriate applications (existing or envisioned) in UA. This latter would actually involve more than 90% of the budget commitment.

The above serve only as an outline of what a full GHSL/UA integration might include in its technical specifications document. The JRC is looking forward to define in concomitance with the needs and requirements of DG Regional Policy the details of the technical steps required for the full integration of the GHSL in UA.

2 Foreword

JRC started the design of the global human settlement layer (GHSL) concept during 2010-2011, together with the development of an image query (IQ) system able to generate and manage geo-information in an integrated way. The IQ system aggregated the experiences related to automatic information extraction from meter and sub-metre resolution satellite image data in the disaster and crisis management scenarios supported by JRC since 2003-2004. The first alpha-test of the IQ system was delivered in Dec 2011, performing a GHSL image information query task over high and very-high resolution satellite image data covering more than 615 billions of square kilometres of global earth surface, mostly placed in populated regions of Europe, Africa, Asia and South America.

During 2011, first contacts with DGREGIO were made in order to understand if the JRC IQ technology and the derived GHSL information layers may be of interest in the context of the "European Urban Atlas" (UA) implementation and in general, in pan-European mapping and characterization of European settlements.

This feasibility report describes the application of the GHSL protocol according to the Urban Atlas product specifications and more specifically the comparison between SSL output information with the GHSL built-up information extraction in the context of the Urban Atlas 2012-2013.

The objectives of the work described in this report were i) to test the processing capacity of the JRC IQ system in order to assess the feasibility of a pan-European GHSL coverage or "built-up-areas detection" using the image data prepared for the UA 2012-2013, ii) to assess the reliability and added value of the automatic image information retrieval by systematic comparison of the automatic output with a known reference layer reporting about similar information, namely, the European soil sealing layer.

3 Image data under test

The image data potentially available for the current test is made of 793 satellite scenes collected by the Spot 5 platform during the years 2003-2009. The data was made available by the GMES ESA platform and is the same image data used for the creation of the UA 2006.

The satellite scenes are projected in various local UTM projections, and they include also geographical lat-lon projection (EPSG 4326). These scenes have different radiometric characteristics including multispectral, panchromatic and "pan-sharpened" products (
Table 1). Moreover, these input scenes show various levels of cloud cover. 260 scenes show cloud coverage less than 0.1%, while all the others show cloud coverage up to 50% (
Table 2).

The expected spatial displacement or tolerance admitted is of 5m RSM, as specified in the UA technical specifications.

Count of imageid	Column Labels			
Row Labels	MUL	PAN	PSH	Grand Total
4326			23	23
32628			2	2
32629	13	12	26	51
32630	13	7	122	142
32631	4	8	63	75
32632	10	23	136	169
32633		19	126	145
32634	13	9	124	146
32635	5	6	29	40
Grand Total	58	84	651	793

Table 1 - Number of scenes by Scene Projection and Type of image data

less than 0.1%	260
from 0.1% to 10%	331
from 10% to 20%	18
from 20% to 30%	8
from 20% to 40%	4
from 40% to 50%	2
more than 50%	1

Table 2 - Number of scenes by Cloud Cover percentage

Figure 1 shows the spatial distribution of the available image data for sensor characteristics. We can observe that while the majority of areas are covered by PSH scenes, in some areas only PAN image data are available, while occasionally some PAN+ MUL bundle images are also available. In order to simplify the design of the test, only PSH images having a UTM projection were introduced in the processing list. In total they are 628 scenes scattered all around the European territory (
Figure 2).

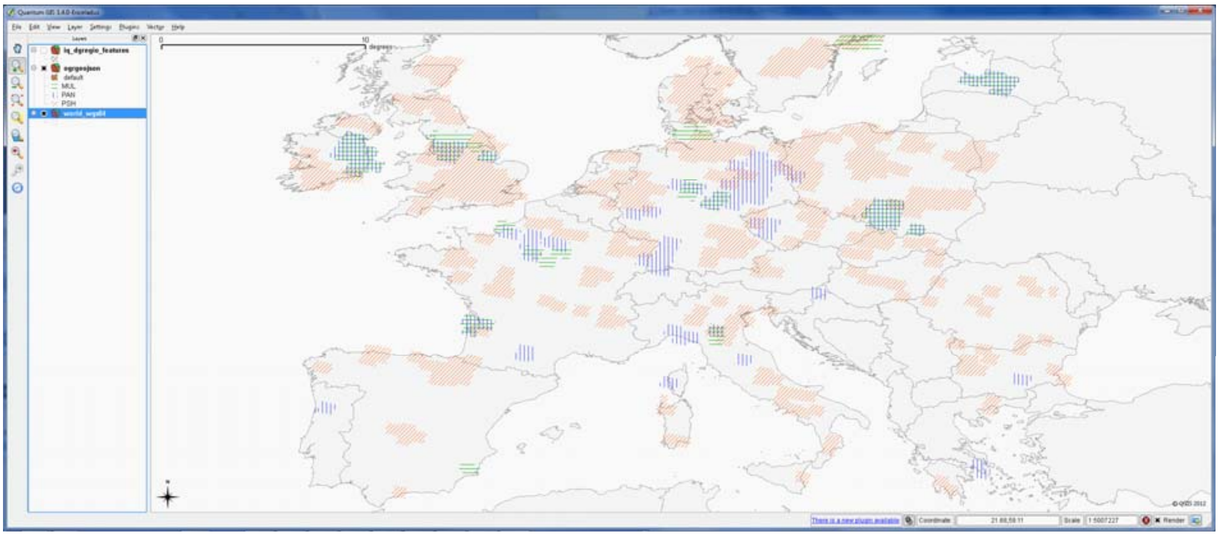

Figure 1 - Available image data input by type of image data

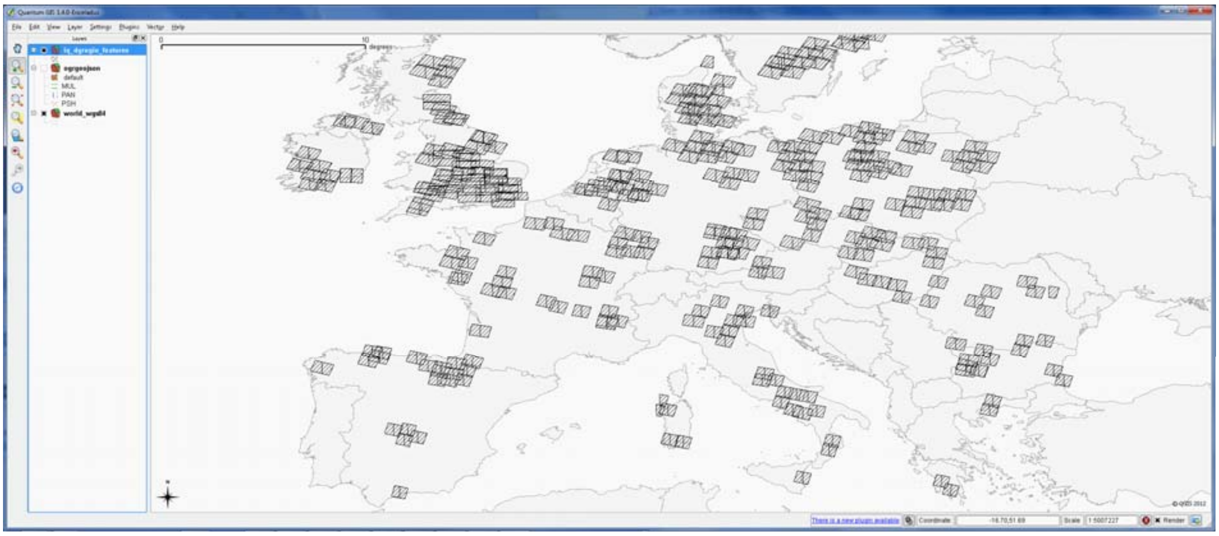

Figure 2 - Image data input under test

In respect to the standard IQ GHSL workflow the available image data for this test is placed between the 1:10K and 1:50K GHSL specifications. In particular, the geo-coding quality would allow 1:10K GHSL products, but the input image resolution is not enough to calculate the morphological/shape criteria that are necessary to recognize and describe the built-up structures at 1:10K scale. Moreover, the "pan-sharpening" processing adopted for the fusion of multispectral and panchromatic data, also if producing suitable results for visual inspection and interpretation, it has some drawbacks on the side of the automatic exploitation of the same input image data. In particular, image "pan-sharpening" degrades the image radiometric criteria then degrading the performances of automatic shadows and vegetation analysis. Moreover, according to evidences collected in the available images, the same process included a "visual enhancing" step performed probably with high-band-pass filtering by convolution. This kind of filtering introduces image artifacts on the borders of the image structures (objects) and increases the overall image noise, then decreasing the signal/noise ratio. All these facts have a negative impact on the performances of the textural and morphological/shape image-derived criteria needed for the automatic discrimination and characterization of the built-up structures.

3.1 Reference data

In order to maximize consistency of the output, the IQ GHSL image information workflow includes automatic optimization of some processing parameters by systematic comparison with known reference information sources. In particular, in this version of the IQ system are included two standard reference layers: a global population density layer (LANDSCAN2010[1]) and a global land use map of urban extents derived from remote sensing data (MODIS500[2]). They have respectively 1 kilometre and 500 meters nominal resolution.

During the current test we introduced in the workflow a new reference layer available only in European countries: the EEA Fast Track Service Precursor on Land Monitoring – "Degree of soil sealing 100m"[3]. This layer is used for the production of the UA "urban fabric" classes discriminating them in different "soil seal" percentage thresholds. Because already well known by DGREGIO and the service providers it is used here for benchmarking the GHSL product accuracy in detection of built-up areas. This layer was produced by a hybrid procedure intersecting satellite-image-derived land cover information with COTS data reporting on roads and settlements.

4 Adopted image information extraction workflow

During the current experiment, a mix of the standard 1:10K and 1:50K IQ GHSL queries was implemented in order to cope with the intermediate scale (between 1:10K and 1:50K) of the available input images respect to the GHSL specifications and data requirements. The output of the two queries is then integrated (by sum) and evaluated using the reference sealed soil surface layer at 100 meters of resolution (
Figure 3).

In particular, the recognition of built-up areas are done at 1:50K scale, while a subset if the 1:10K GHSL workflow was applied for detection and characterization of single built-up structures inside these detected built-up areas. The characterization was based only on the estimated area (in plant) of the built-up structures. Other parameters available in the standard GHSL workflow 1:10K, like building height and shape, were not processed because of the poor quality of the input images not allowing reliable estimation of them.

[1] LANDSCAN 2010 http://www.ornl.gov/sci/landscan/index.shtml

[2] The MODIS 500-m map of global urban extent made by the Center for Sustainability and the Global Environment, University of Wisconsin-Madison http://sage.wisc.edu/people/schneider/research/data_readme.html

[3] Raster data set of built-up and non built-up areas including continuous degree of soil sealing ranging from 0 - 100% in aggregated spatial resolution (100 x 100 m). http://www.eea.europa.eu/data-and-maps/data/eea-fast-track-service-precursor-on-land-monitoring-degree-of-soil-sealing-100m-1

At 1:50K scale, it is considered "built-up" any spatial reference unit (pixel) of 50x50 meters, covering one built-up structure or part of it. Analogously, at 1:10K scale it is considered "built-up" any spatial unit of 10x10 meters, covering a building or a part of it. At 2.5-m input spatial resolution, only 1:50K built-up areas are reliable: moreover, the tolerance in characterization of the area of built-up structures is of 6.25 meters (surface of one input image pixel).

The two GHSL output scales are then integrated by sum at the 1:10K scale resolution: in this way we create a hybrid 1:10K-1:50K product that is evaluated by systematic comparison with the European SSL.

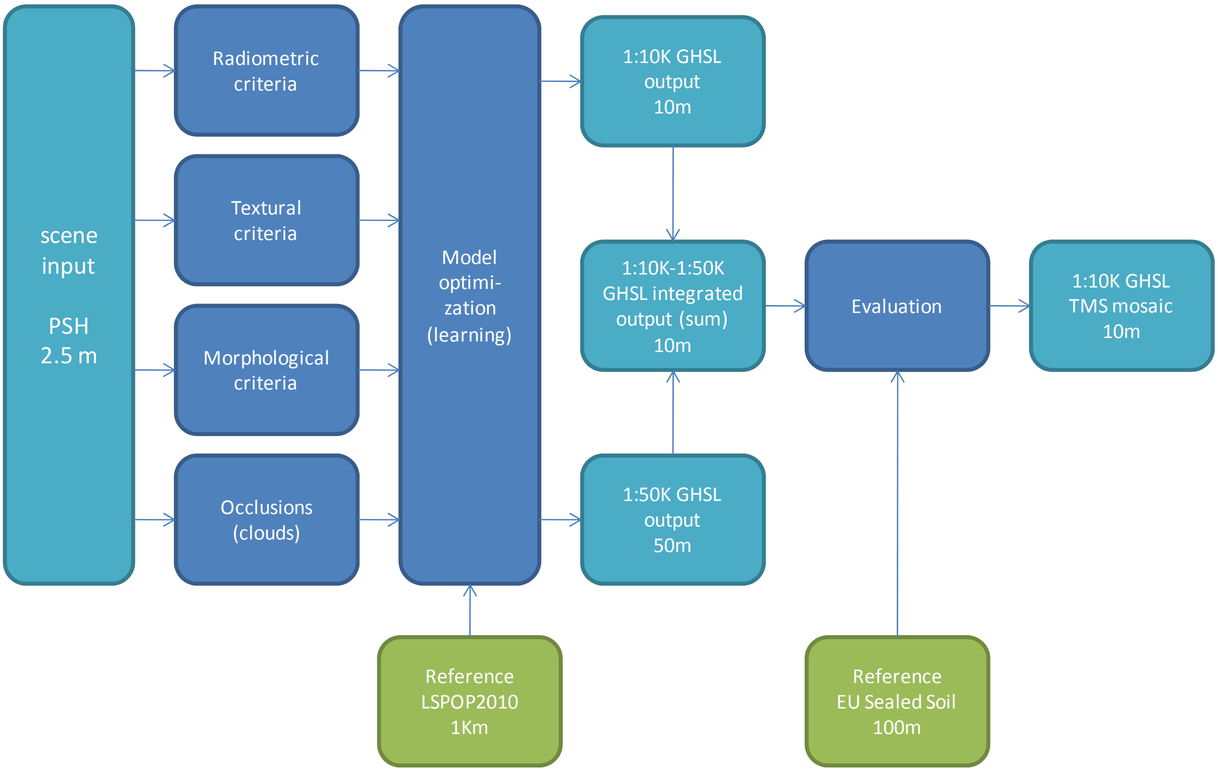

Figure 3 – automatic image information workflow applied during the experiment

5 Evaluation parameters

Both the reference layer and the GHSL output under test are continuous information layers: the reference layer reports about the percentage of "sealed surfaces" in respect to the spatial unit of 100x100 meters, while the GHSL output reports about the presence of built-up structures in the spatial units of 10x10 and 50x50 meters, integrated by sum.

The "built-up presence" information released by the IQ system at any scale is a membership value to the class "built-up" as formalized by the image information query. For the purpose of this report, this membership value can be interpreted both i) as probability (or possibility) score made by the system that this cell corresponds to a built-up structure on the ground, and ii) as a percentage of the unit cell corresponding to a built-up structure. The two statements can be merged in one by interpreting the membership value as "the estimated possibility that the *whole* output unit cell is covering a built-up structure on the ground".

In order to measure the agreement between the GHS output and the reference, we adopted two strategies: i) directly measure the agreement of the continuous information layers by linear regression techniques, and ii) dichotomize the continuous information layers by a given threshold, and then calculate agreement measurements based on confusion matrix. In this experiment, a threshold of 25% in both layers was considered to discriminate between built-up (BU) and not-built-up (NBU) classes. In the dichotomic-classification case the GHSL output under test was aggregated at the same resolution provided with the reference layer, then 100x100 meters. In the continuous case two different resolutions were put under test: 50x50 meters and 500x500 meters.

During the evaluation, the following measurements were collected per each input scene processed: overall accuracy, built-up accuracy, built-up agreement, and x_fit.

i. The "overall accuracy" is the number of pixels with agreement on the BU/NBU classification divided by the sum of all the pixels analysed in the scene.
ii. The "built-up accuracy" is the number of pixels with agreement on the BU class divided by the sum of all BU reference pixels analysed in the scenes.
iii. The "built-up agreement" is a per-scene global measurement expressing the agreement on the total surface classified as BU class.
    a. A = 1 – abs(pixBUclass – pixBUref) / (pixBUclass + pixBUref)
    b. With "pixBUclass" be the number of BU pixels estimated by the classification under test and the "pixBUref" the number of BU pixels estimated by the reference layer
iv. The "x_fit" measurement reports about the per-pixel R-square linear regression fit (correlation) between the GHSL output and the reference layer. It was calculated using two different scales or spatial resolutions: 50x50 meters, and 500x500 meters.

6 Results

6.1 Qualitative inspection

Figure 6 shows the standard output obtained by the GHSL workflow under test over the city of Luneburg, Germany. Brighter grey levels mean higher membership output value to the class "built-up" which is automatically generated by the IQ system.
Figure 4 shows the same region as represented by GE imageries, while
Figure 5 shows the representation made by the SSL reference layer. Also in this case brighter grey level means higher value of "sealed surface" percentage.

Figure 7 shows the GHSL output concerning the variable: "size of built-up structures". For "size" here it is understood the estimated area (in plant) of the built-up structure. The original GHSL information is represented in square meters or "scale": for visualization purposes the colour palette applied ranges from green (low) values to red (high) values.

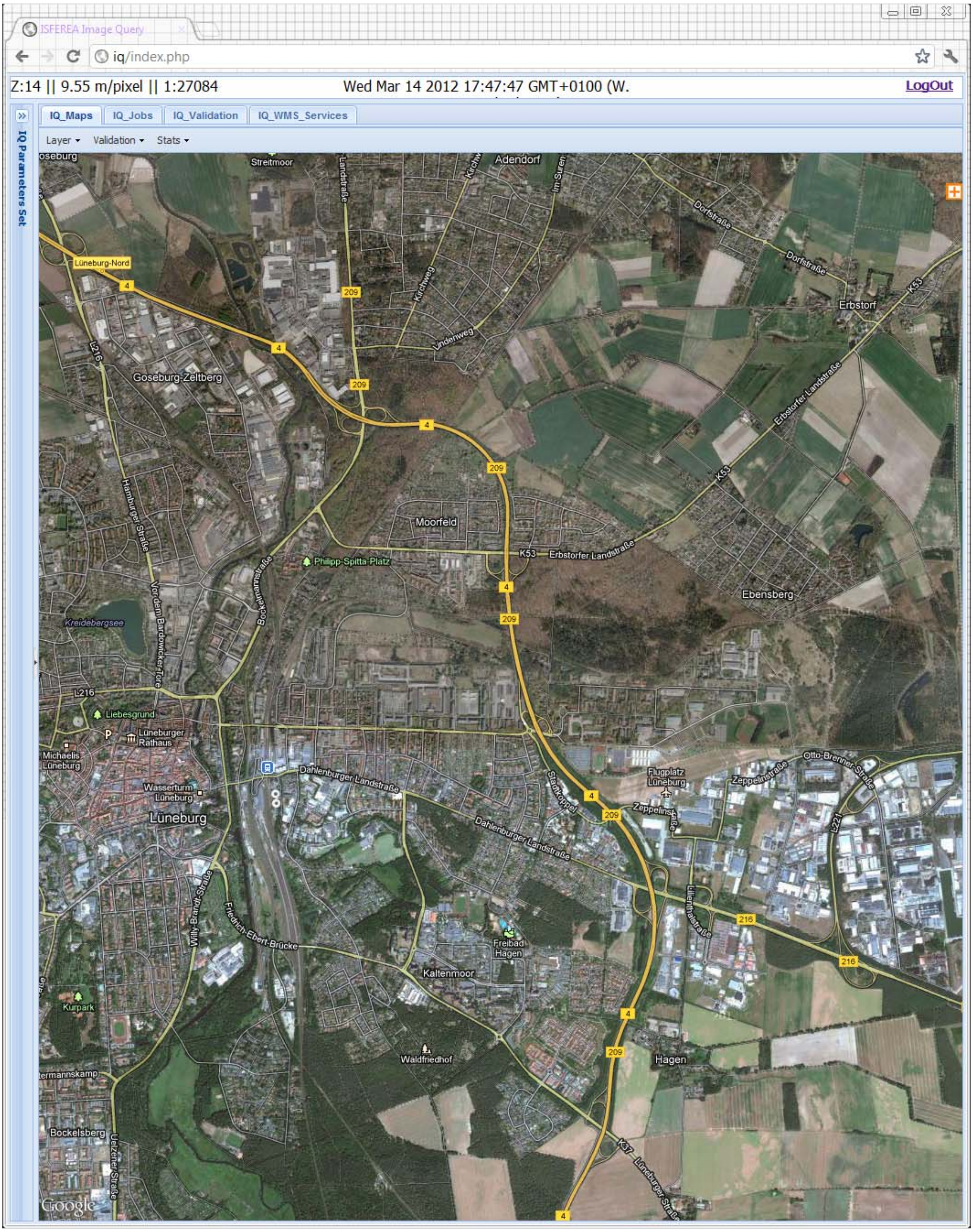

Figure 4 – The city of Luneburg, Germany at 10-m-resolution

Figure 5 – the same region as represented by the SSL reference layer

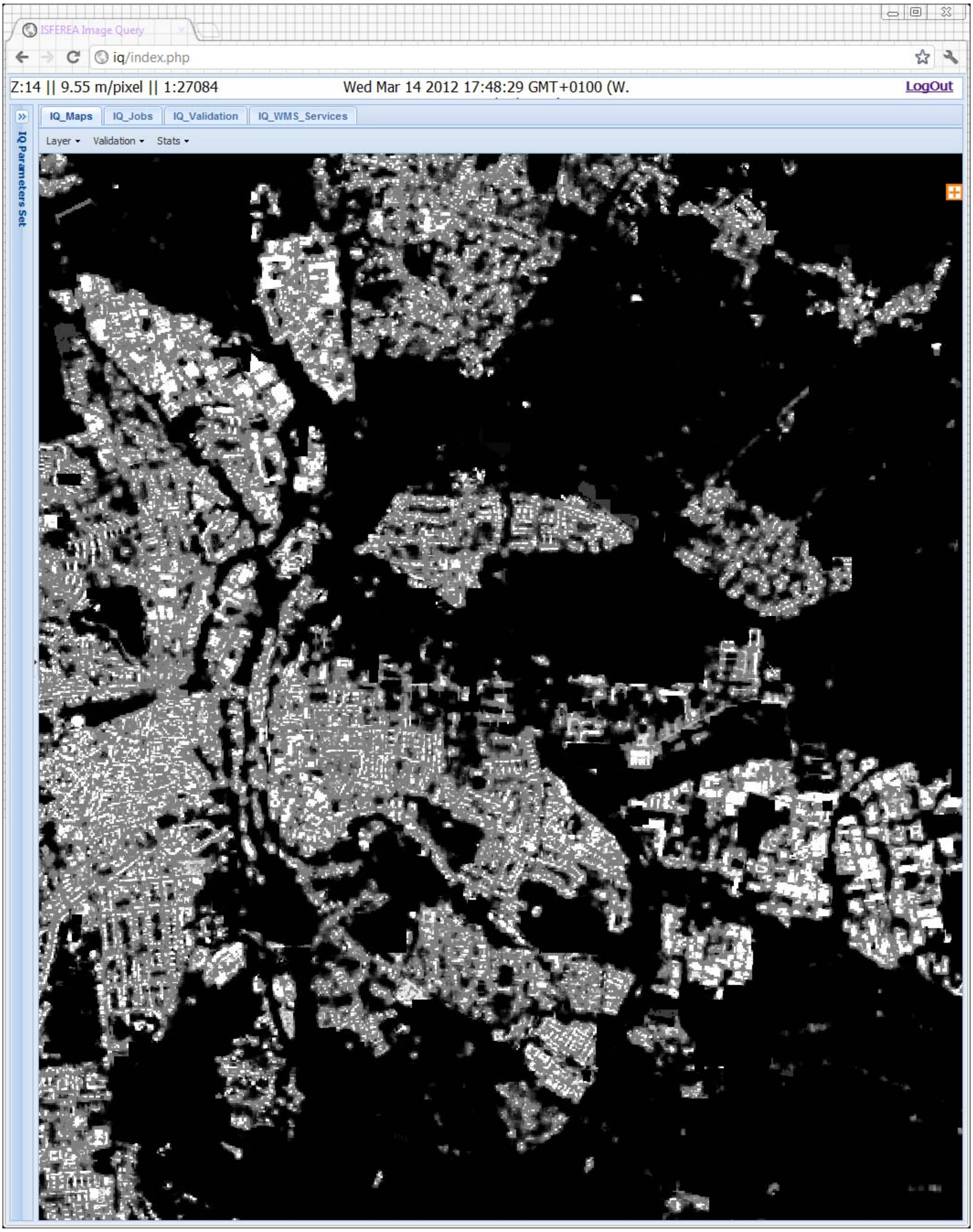

Figure 6 – the same region as represented by the "built-up areas" GHSL output

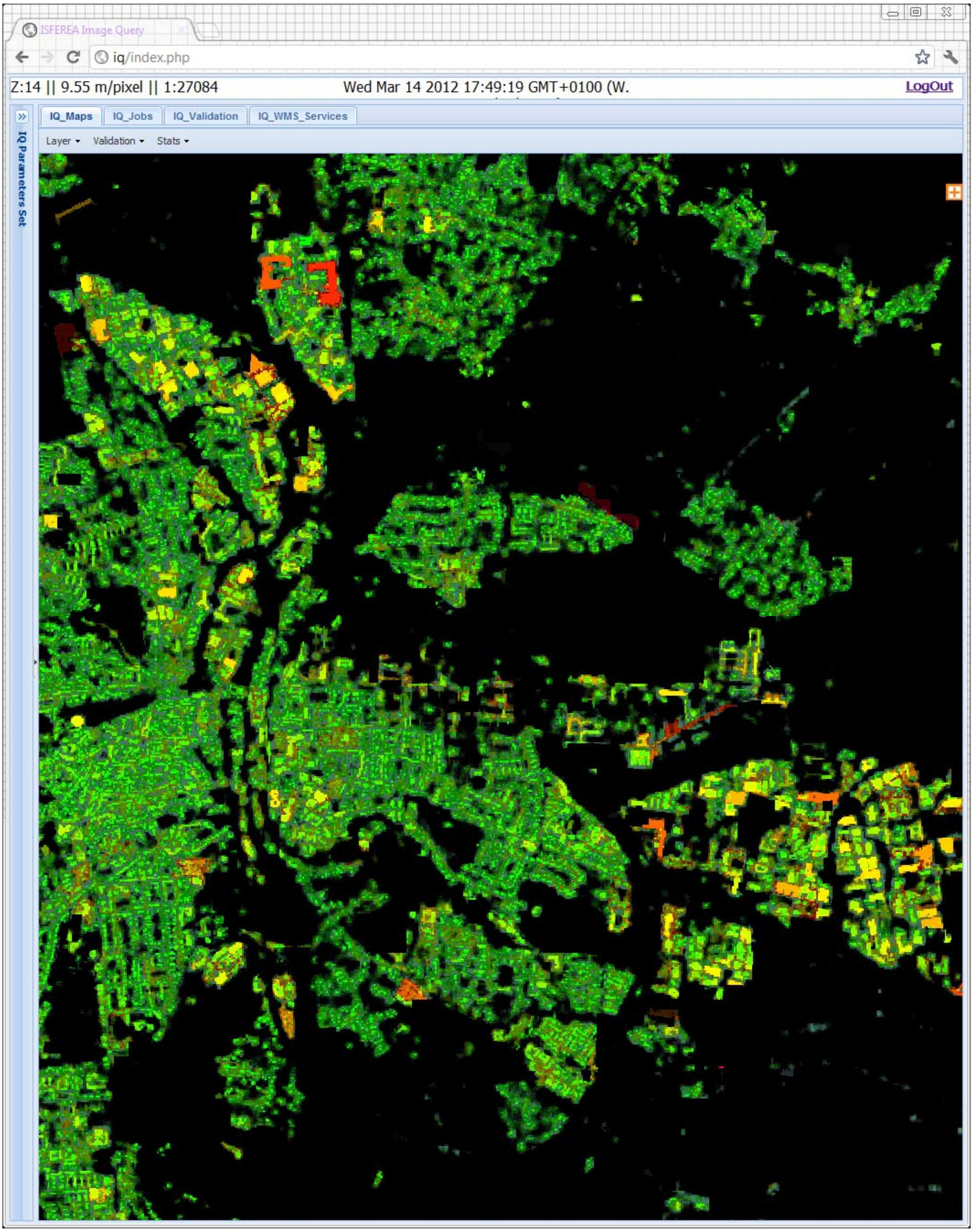

Figure 7 - the same region as represented by the "size of built-up areas" GHSL output

6.2 Quantitative results

In this test, only the "built-up areas" detection GHSL output was evaluated in quantitative way. Other GHSL output information layers including characterization of built-up structures inside the built-up areas (discrimination by size) were not taken in to account. The reason of this choice is twofold: i) "built-up areas" detection is hierarchically precedent (dominant) respect to all the other GHSL output measurements, including built-up areas characterization. Then this is the most important variable logically driving the reliability of all the others, and ii) the impossibility to implement a consistent quantitative evaluation protocol without having at disposal a reference layer of a suitable scale – as a single building footprints at 1:10K scale.

The results of this first quantitative test show a good level of agreement between the "built-up areas" automatically generated by the JRC IQ system following a customized GHSL task and the reference SSL layer.

The agreement is good in both dichotomic and continuous evaluation schemas. In particular, the 628 satellite scenes under test were showing a 90.8 ± 3.9 average "overall accuracy" rate and 86.6 ± 7.1 average rate of "built-up accuracy". Moreover, the correlation fit is estimated as 84.7±8.0 and 96.1±5.1 at 50 and 500 meters of spatial generalization, respectively.

	Overall accuracy	BU_accuracy	BU_agreement	x_fit(50)	x_fit(500)
mean	90.82%	86.64%	87.46%	84.75%	96.17%
stdv	3.99%	7.10%	12.49%	8.09%	5.11%

Table 3 – agreement measures between the GHSL "built-up areas" output and the reference layer

Because of the slight different nature of the two information layers under comparison, these measurements include both "errors" in the classical term but also the different semantics and scale/generalization parameters embedded in the two different information layers. In other words, the 10% of disagreement between the two information layers can be originated by the following phenomena:

i) Wrong "built-up area" value in the GHSL output if compared with reality (input imagery)
ii) Wrong "sealed percentage" value in the reference layer if compared with reality (input imagery)
iii) Different scale and generalization parameters
iv) Different semantic definition of "built-up area" and "sealed surface area".

While the first two items are more classically considered "errors" scores, the last two are more related to relative differences in the definition of the geographical information in the two layers under comparison. The discrimination between these different phenomena is quite challenging and may require expensive additional independent assessments following the two distinct geographic information collection protocols.

Some attempts can be made by observing that the factors related to the points iii) and iv) have more chance to take the behaviour of as a systematic bias effect, while factors i) and ii) are in principle randomly distributed[4].

Figure 8 shows the correlation between the total (per-scene) surface estimated as "built-up" by the SSL and the GHSL layers, for all the scenes under test. It can be noticed a slight positive slope in the estimated linear regression (y = 1.15*x) meaning that on the whole set under study the GHSL estimation of BU areas tends to be systematically more abundant than the one made by the SSL layer.

This is a counterintuitive result, because in principle "built-up" surfaces should be considered logically a sub-set of the "sealed" surfaces, then one would expect to find systematically more sealed surface than built-up one. Then the expected slope factor in the linear regression would be smaller than 1. This result can be partially explained by observing the linear regression in the subset of scenes having low density settlement pattern.

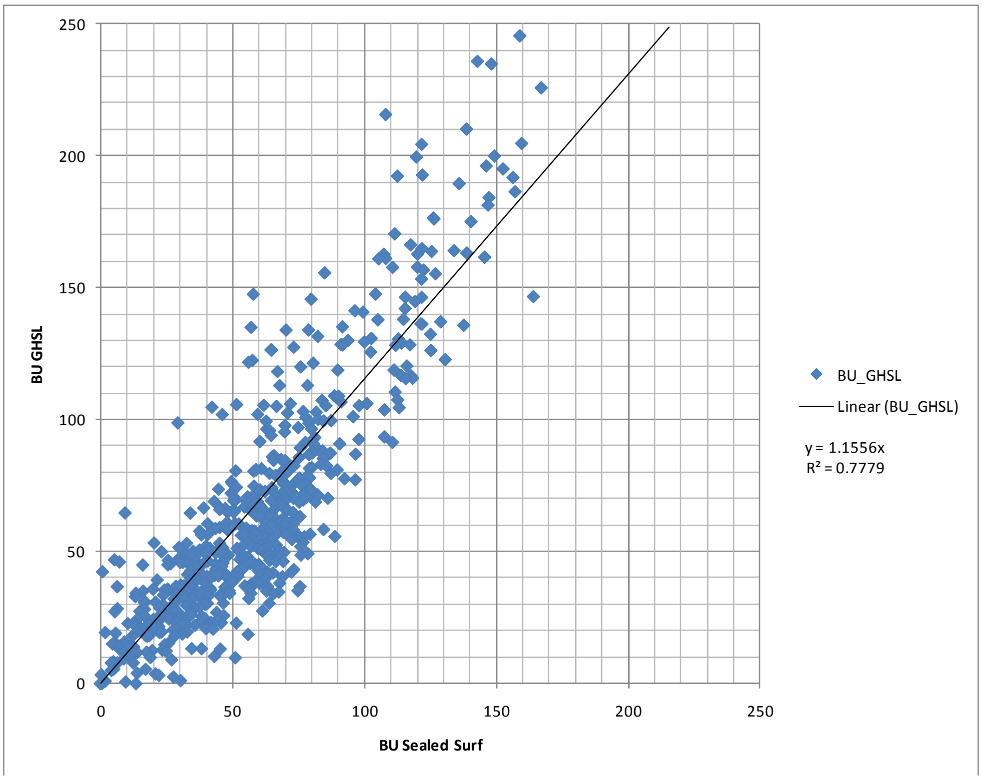

Figure 8 – correlation between the total estimation of BU surface per scene in the reference and in the GHSL layers: all satellite scenes under test

[4] In reality this may be more complicate and should be further inquired. For example arbitrary local bias errors in the soil surface layer can be systematically associated to the different service producers adopting slightly different production workflows in different parts of Europe. – see for example Pavol Hurbanek and others, "Accuracy of Built-up Area Mapping in Europe at Varying Scales and Thresholds" in proceedings of Accuracy 2010 Symposium, July 20-23, Leicester, UK. Also in http://data.geoinovace.quonia.cz/tmp/prednasky/2010_12_07_presentation_BrnoOlomouc_3_.pdf

For example, in
Figure 9 the regression is estimated only in the subset of scenes having less than 100 square kilometres of built-up surface estimation, and then excluding all the scenes dominated by big urban centres. In this case, the linear regression takes the expected slope inferior than 1 (y = 0.95*x), then showing a systematic underestimation of GHSL "built-up" surfaces respect the "sealed" surface.

Even if requiring further analysis, we think that these empirical evidences are the combined effect of the two "disagreement" factors described above: the iii) Different scale and generalization parameters and the iv) Different semantic definition of "built-up area" and "sealed surface area".

The working hypothesis is that while the factor (iv) is always on the background, the factor (iii) becomes dominant in case of presence of compact large urban nucleus where the surface of open spaces and roads between buildings is very relevant respect to the total area under analysis. This "interstitial" surface is then changing BU/NBU class according to the different scale and generalization rules adopted in the two layers under comparison. As known, the impact of scale and generalization factors is heavily related to the spatial pattern of the represented information.

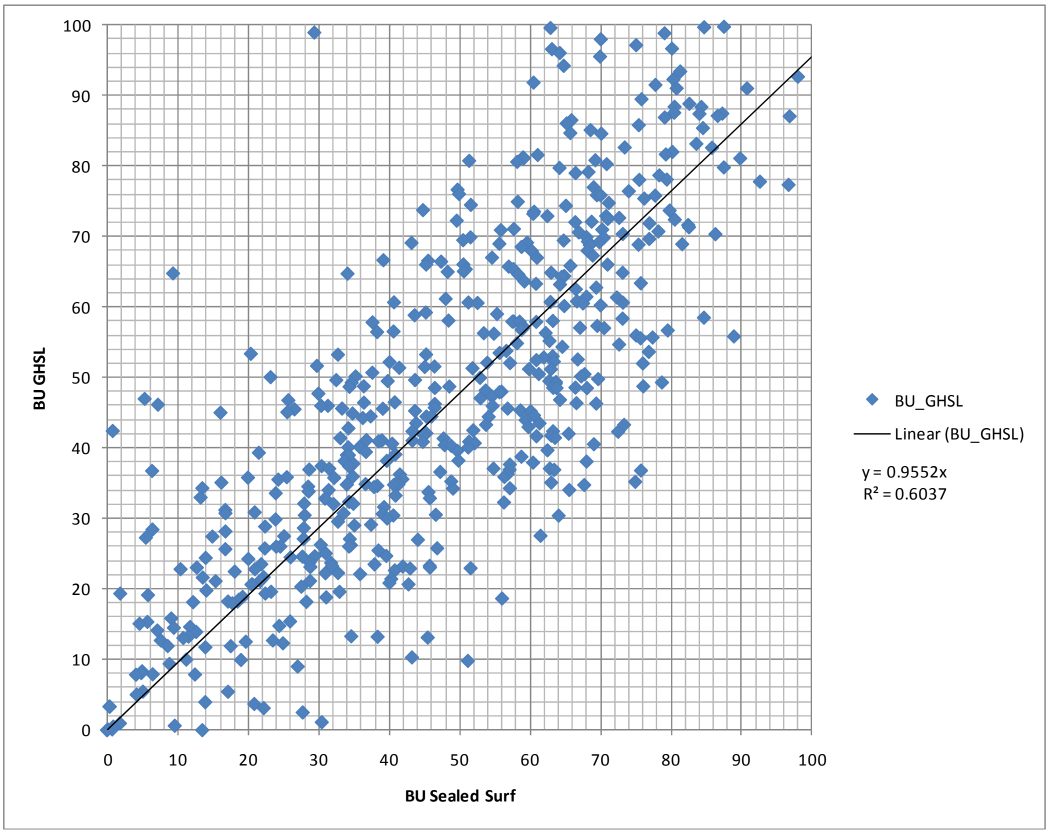

Figure 9 - correlation between the total estimation of BU surface per scene in the reference and in the GHSL layers: it is considered only the subset of satellite scenes having less than 100 square kilometers of BU surface

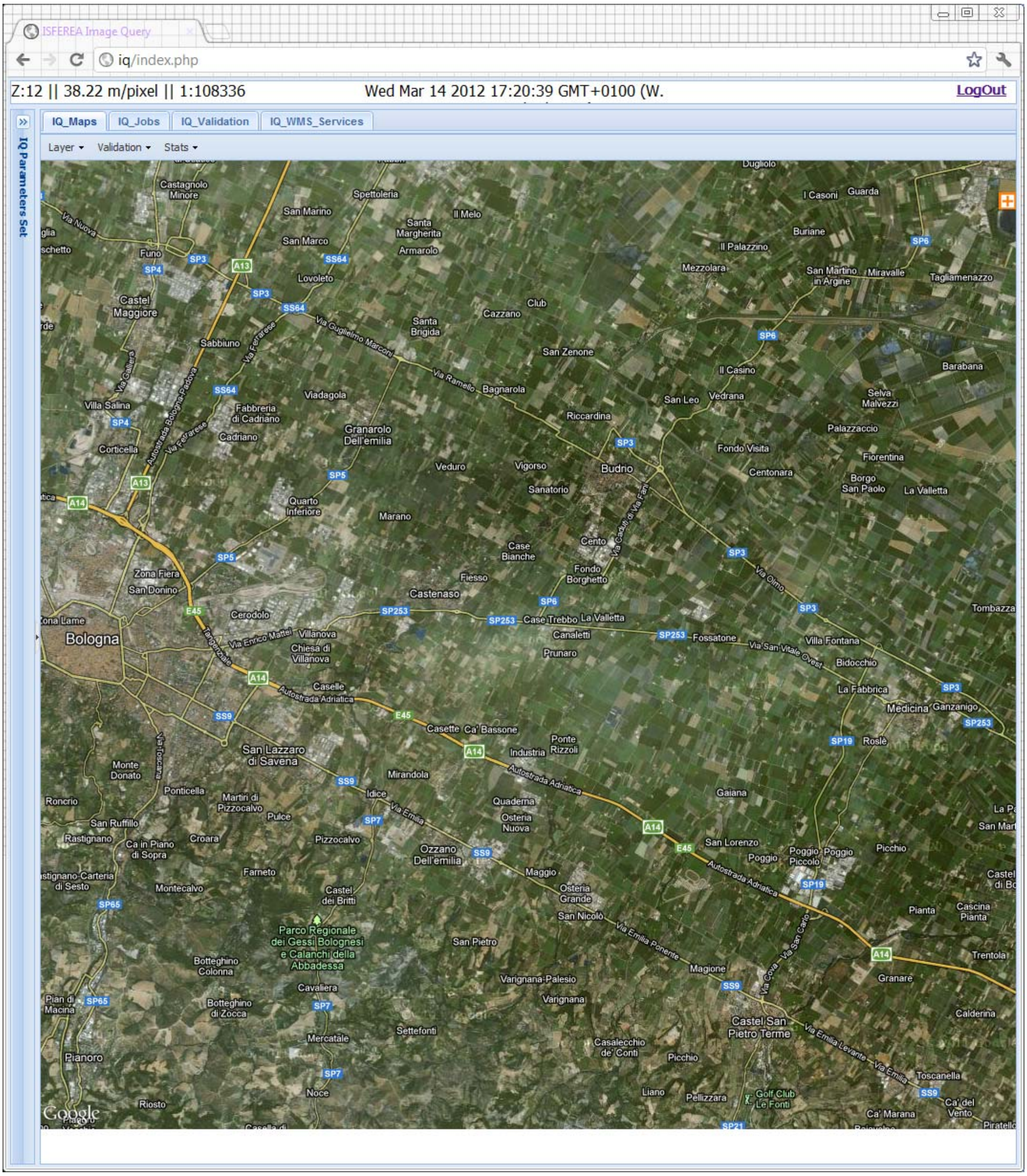

Figure 10 - example region S-E of Bologna, Italy

Figure 11 - the same region as represented by the SSL layer

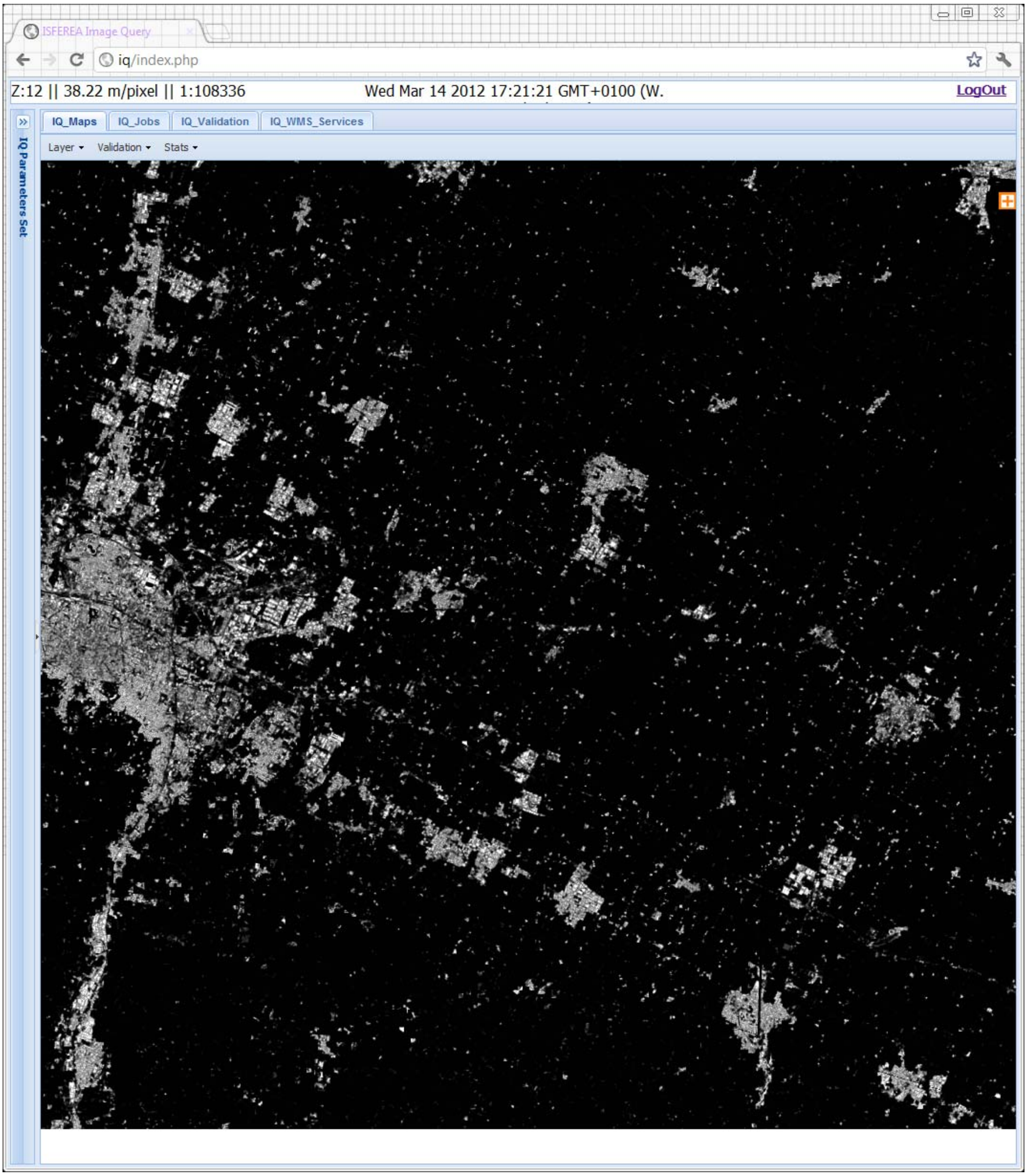

Figure 12 - the same region as represented by the "built-up areas" GHSL output

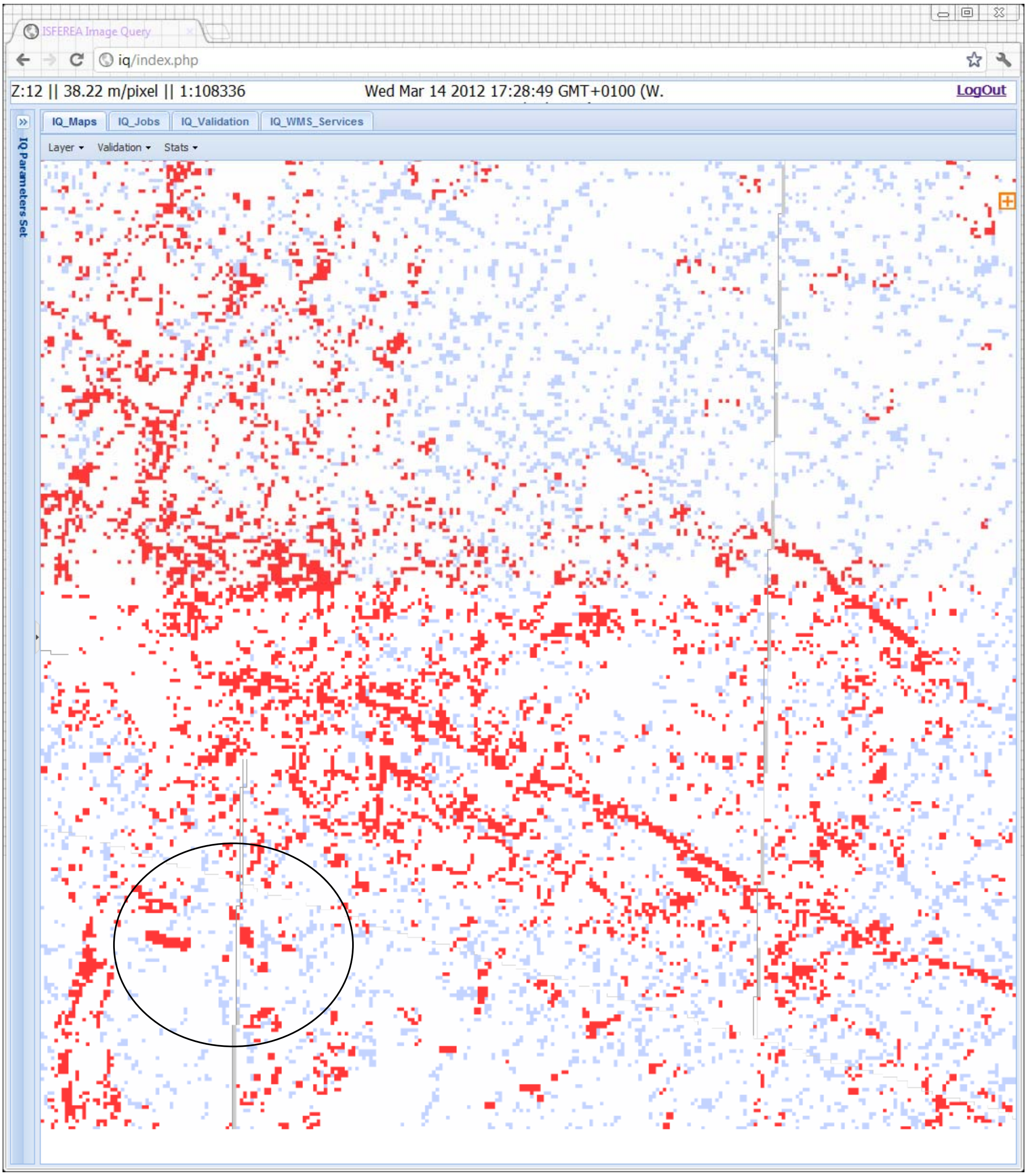

Figure 13 – "Difference map" between the dichotomic "built-up areas" and "sealed surface" classes. In red and light blue the positive, negative differences, respectively, in the Sealed - GHSL comparison.

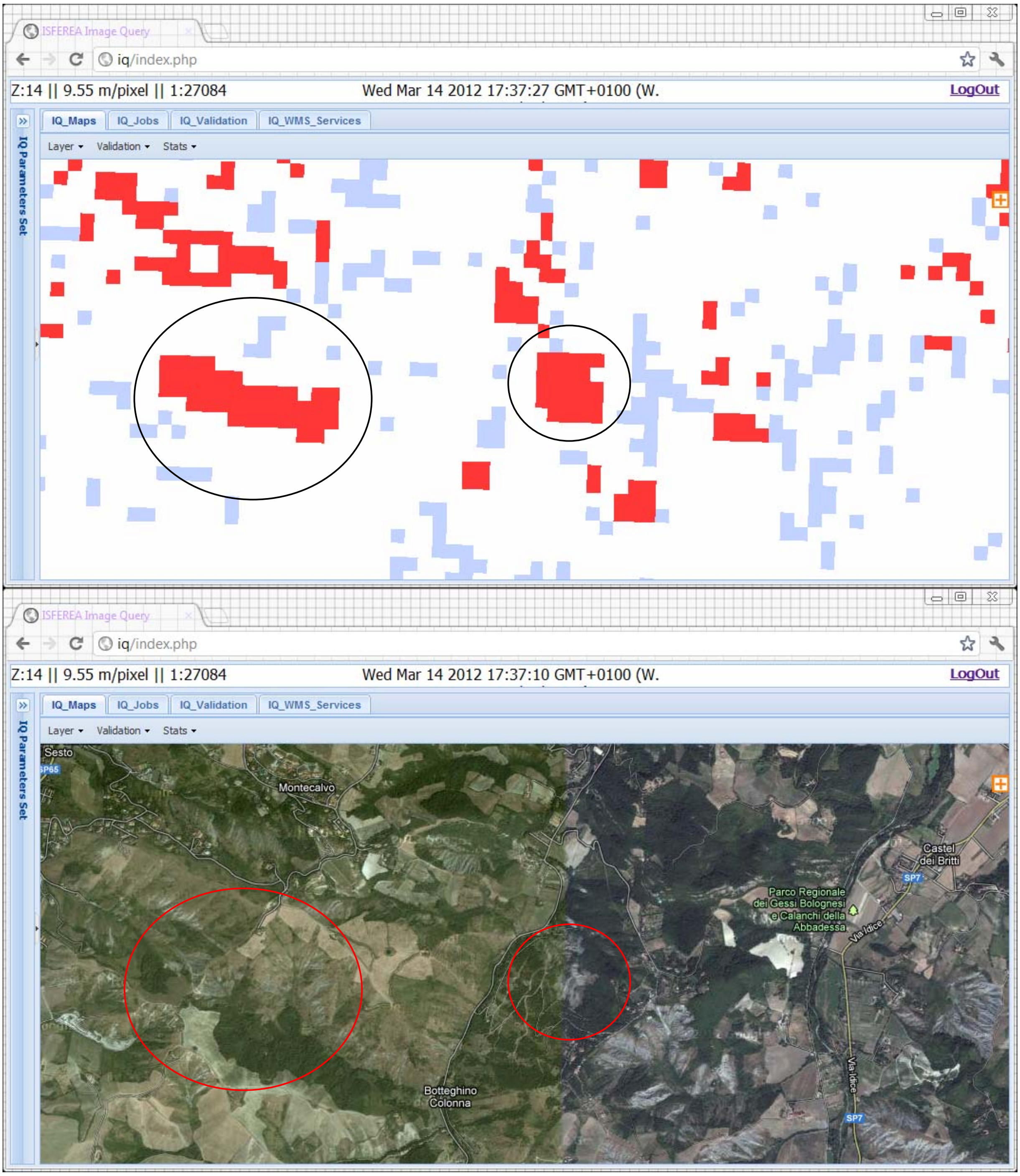

Figure 14 – zoom in the "difference map" – some errors spots in the SSL reference layer

6.3 Difference maps

During the test “difference maps” were systematically calculated for all the scenes under analysis in order to help the identification and understanding of the agreement/disagreement of spatial patterns between the two information layers. In this first test, only dichotomic difference map where calculated using the same spatial resolution and thresholds used for the accuracy analysis discussed above.

Figure 13 shows an example of the “difference map” obtained in region S-E of Bologna Italy. Red and blue pixels represent areas where the SSL overestimates and underestimates, respectively, the GHSL built-up areas.
Figure 10 shows the same region from GE imageries for visual inspection, while
Figure 11 and
Figure 12 represent, respectively, the SSL and the GHSL information layers.

Some patterns observable from
Figure 13 are well explained by the different semantic embedded in the two products: for example the highway is present in the SSL layer while it is not reported by the GHSL output, because not matching with the “built-up” definition.

Other patterns (usually large red patches, relatively isolated) of sealed overestimation respect to the GHSL output can be explained instead as errors in the SSL detection. See for example
Figure 14 showing two “sealed surface” patches in reality corresponding to agricultural fields.

7 Conclusion and next steps

7.1 Assessment

A test of the JRC GHSL production workflow and output was done by JRC during February 2012 with the purpose of:

i. testing the processing capacity of the JRC IQ system in order to assess the feasibility of a whole European GHSL coverage or “built-up areas detection” using the image data prepared for the UA 2012-2013,
ii. assessing the reliability and added value of the automatic image information retrieval by systematic comparison of the automatic output with a known reference layer reporting similar information, namely, the European sealed soil surface layer.

During the test 628 satellite scenes used for the production of the UA were successfully processed by the JRC IQ system, performing automatically a customized version of the standard GHSL query including i) automatic recognition of built-up areas and ii) automatic characterization of built-up structures based on size (area). Automatic detection of occlusions and no-image-data areas (clouds) and hierarchical mosaic-ing were also activated in order to test the production of seamless information layers from any arbitrary set of partially-overlapping, cloud-covered, input satellite scenes. Other standard GHSL descriptors (building height, vegetation) were not calculated because of technical limitations of the input images available.

The computational and production test was successful and showed the capacity to produce seamless European layers using the same JRC technology and similar input data.

Limiting factors are substantially linked only to the available storage space capacity of the current IQ system that should be up-scaled in order to i) improve the spatial resolution (then the scale) of the output information layer now set to 10 meters ii) improve the number of information layers that can be produced (settlement characterization)

The results of the automatic recognition of built-up areas made by the IQ system were systematically compared in all the 628 processed satellite scenes with the SSL layer used for the production of the UA.

According to the results of this comparison, the IQ GHSL automatic output showed an overall high degree of agreement in respect to the SSL layer. The agreement is good in both dichotomic and continuous evaluation schemas. In particular, the 628 satellite scenes under test were showing a 90.8 ± 3.9 average "overall accuracy" rate and 86.6 ± 7.1 average rate of "built-up accuracy". Moreover, the correlation fit is estimated as 84.7±8.0 and 96.1±5.1 at 50 and 500 meters of spatial generalization, respectively.

7.2 Next steps

The analysis of the 10-20% thematic differences between the two layers has only begun. We can observe the contribution of four main factors: i) commission errors in GHSL; ii) commission errors in SSL; iii) different scale and generalization parameters; and iv) different semantic definition of "built-up area" and "sealed-surface area".

Better understanding of the contribution of these factors would require the design of another experiment including:

i. Independent collection of reference information from the same input images by an explicit protocol including visual interpretation rules and a clear definition of "built-up" areas.
ii. Increased resolution of the GHSL output under test in order to better describe the settlement components including built-up structure morphological characteristics and the space between built-up structures (road, open spaces)
iii. Production of increased number of descriptors of settlement patterns including different levels of generalization and different constraints in the available GHSL information layers (morphological characteristics, patterns)

8 ANNEX

8.1 GHSL Rationale

Information about human settlements is relevant for several issues including i) understanding where are placed buildings and population, ii) understanding the physical characteristics of the settlement and infer the social level / vulnerability. This information is relevant input of any impact or risk modelling (monitoring, decision, planning) exercise having implementation in the spatial domain.

Global/regional actors need global/regional assessments or local assessments with globally consistent approaches (comparable). There are gaps between these needs and the available (global/regional) information layers about human settlements. These gaps can be summarized mainly by i) inconsistencies between different scales of representation eventually available in different places of the world ii) too coarse detail of globally available layers in detection of built-up structures ii) semantic abstraction difficult to make interoperable, iv) too expensive production process at detailed scales (1:10K-1:50K).

The global human settlement layer (GHSL) notion tries to address these issues by the following steps: i) drastic revision of the semantic embedded in the product, in particular the adoption of quantitative spatial measures vs. dichotomic mutual-exclusive classes (as dominant in the LULC paradigm), and the preference low vs. high class abstraction, improving semantic interoperability and ii) strong push on automatic image information extraction.

8.2 What is GHSL

8.2.1 It is an inclusive and collaborative concept

The main objective is to coordinate the efforts enhancing global awareness and objective fact findings on the Human Settlement phenomena. This will include a multi-scale vision including globally and locally consistent representations, and then exploiting advanced multi-sensor and multiple-criteria automatic image information retrieval technology. Moreover, the JRC GHSL initiative aims to foster a wider, inclusive and dynamic concept of human settlements including in the general picture informal, poor and temporary settlements (as for example refugees, IDP camps), underrepresented in the similar available information layers.

8.2.2 It is a work-in-progress with a scalable architecture

Because of the characteristics of the technology used for image information extraction and aggregation at different scales, mapping the human settlement can become a collective exercise improving piece by piece the global picture at the specific sensor resolution available for different places of the world. Interested users of the derived information layers will be allowed to upload image data and download the related image information contents describing the human settlements, in this way improving the reliability of the general picture, for a better collective awareness.

8.2.3 It is designed for supporting crisis management

All the crisis management cycle - including damage assessment, recovery, reconstruction and planning should benefit from an improved and globally-consistent description of human settlements. From this point of view, the applied notion of “human settlement” must be a proxy to population and assets at risk, and consequently must provide quantitative input for impact and risk modelling exercises. These characteristics make the GHSL concept different from other available global geo-information layers: “sealed”, “impervious”, land cover surface, the “urban” land use land cover (LU/LC) surface are not applicable directly to GHSL concept.

8.2.4 It is designed in order to optimize image information extraction efforts improving information sharing and re-using

For this purpose the GHSL will be open to a wider range of applications. it is simple in semantic and shows a lower abstraction level than the standard land use land cover (LU/LC) model.
This fact makes it easier to semantically interoperate the GHSL across applications and easier to make consistent multi-scale information layers by mathematically consistent aggregation, generalization operators. These logical characteristics will make it easier to integrate GHSL with other information layers coming from heterogeneous sources and at specific scales (slope, vegetation, nightlight, income, vulnerability …)

8.2.5 It is measured globally from multi-scale remotely sensed image data

The GHSL contains globally-consistent physical measurements describing human settlements at different scales, and then must be generated systematically and automatically.
The image information extraction technology applied for the GHSL production works with different sensor resolution / sensor-method and reliability conditions.

8.2.6 It is validated globally with multi-scale approach

The vision is that GHSL will be validated with global reference repositories, eventually improved by collective reference information collection by collaborative mapping mechanisms.
GHSL is validated using multi-scale reference information collection and aggregation with consistent automatic quality measurements and benchmarking.

8.3 GHSL technical specifications

8.3.1 Format

GHSL data it is released trough standard protocols defined by the Open Geospatial Consortium (OGC). In particular, Web Map Service (WMS[5]) and Tile Map Service (TMS[6]) platforms release the GHSL product as output of a specific query to a spatial DB, then including time and dynamic information queries. The native storage format and structure of the GHSL information is tile-based: the basic Spatial Unit is a surface “tile” that is the representation of a given portion of the earth surface with a given size and projection. The tile entity is organized in a hierarchical multi-scale structure following the TMS standard.

[5] See http://www.opengeospatial.org/standards/wms/
[6] See OSGEO Tile Map Service (TMS) Specification http://wiki.osgeo.org/wiki/Tile_Map_Service_Specification

While satellite-derived image features are stored and managed in the local (UTM) metric projection, the global mosaic, classification and GHSL representation of them are adopting a global metric projection, that is the Spherical Mercator EPSG:900913 with the WGS84 Datum.

8.3.2 Scale

GHSL information layers it is provided with three nominal scales of reference, namely “local”, “regional” and “continental” scales. They correspond to specific parameters regarding the TMS zoom level, the spatial unit of reference and the tolerance admitted in the geo-coding of the information.

I. “local” nominal scale ~ 1:10.000
   a. TMS Z = 22
   b. Spatial Unit = 10 meters
   c. Spatial RMS tolerance = 5 meters
II. “regional” nominal scale ~ 1:50.000
   a. TMS Z = 19
   b. Spatial Unit = 50 meters
   c. Spatial RMS tolerance = 25 meters
III. “continental” or “global” nominal scale ~ 1:500.000 TMS Z = 17
   a. TMS Z = 16
   b. Spatial Unit = 500 meters
   c. Spatial RMS tolerance = 250 meters

8.3.3 GHSL information production

GHSL information is produced by automatic image information extraction techniques, using in input several different digital images coming different satellite platforms. The GHSL image information extraction workflow has been tested on several meter and sub-meter optical and radar satellite and airborne imageries. Moreover, some high resolution optical sensors ranging from 2.5-5.0-10.0 meter resolution have been also tested including Spot 4-5, CBERS and RapidEye imageries.

The standard image query building the GHSL product is composed by the merging of three main image-derived characteristics (or image features, also called “criteria” of the image query): namely, radiometric, textural, and morphological/shape criteria.

Of course, there is a relation linking the quality of the data in input (resolution, geo-coding) and the scale and nature of the GHSL information that can be produced. In this relation play a role the different requirements related to the different techniques used for image information retrieval. In particular, the different image-derived criteria require different input resolution in order to fit in the same GHSL scale (
Figure 15). For example, in the local 1:10K scale 5.0, 2.5 and 0.5 meters of spatial resolution are required for radiometric, textural, and morphological image feature extraction, respectively. The 1:50K GHSL scale is less demanding requiring 20, 2.5-5.0, 2.5 meters of spatial resolution for computation of the same set of image-derived features.

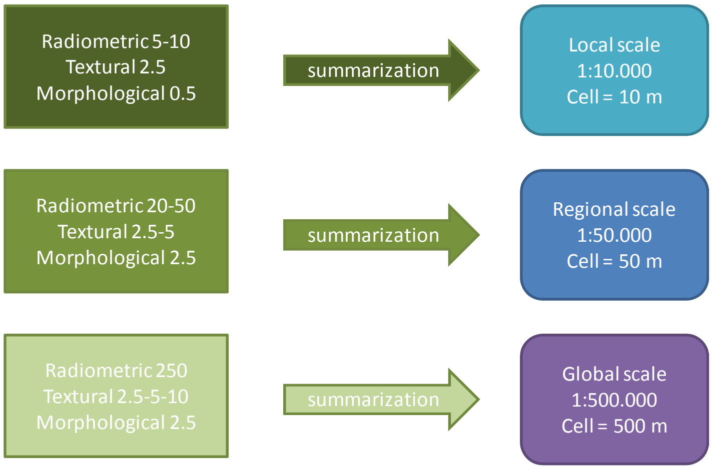

Figure 15 – Characteristics of the image data input and different GHSL scales

Because of the multi-scale nature of the GHSL product, image data filling in one specific scale contributes also to the more generalized scales (

Figure 16).

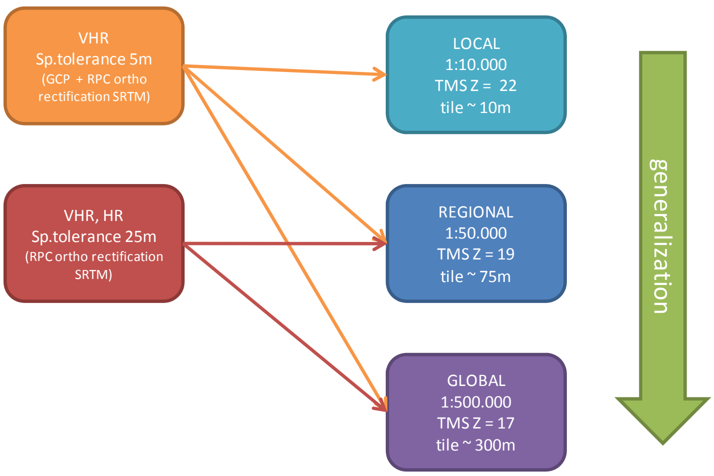

Figure 16 – Generalization from local to the global GHSL scale for different image data input qualities

8.3.4 Quality

8.3.4.1 Spatial consistency

GHSL information layers are produced using various satellite image data in input, having different sensor and platform characteristics, and different pre-processing standards including different geo-coding and ortho-rectification protocols. Consequently, geo-coding quality of the input image data will be checked and the derived image information will feed the GHSL scale corresponding to the assessed spatial tolerance. The input image spatial tolerance admitted is always less than half of the GHSL spatial unit output.

8.3.4.2 Quality control and Validation

The quality of the product is tested by applying a public and reproducible protocol including i) visual image reference data collection and ii) systematic statistical distance measurements respect to known reference layers available in any scale (local, regional, continental).
Only products passing the quality test will be shared with external public and users.
The quality measures are embedded in the output metadata.
The protocol will be used for drafting validation guidelines in case the product will be generated by third parties (GMES services for example)

8.4 UA specifications

The Urban Atlas is providing pan-European comparable land use and land cover data for more than 300 Large Urban Zones (LUZ) each with more than 100.000 inhabitants as defined by the Urban Audit. Its thematic classes are based on Corine LC nomenclature and GUS Legend. Its input data sources are EO data with 2.5m spatial resolution, they are multispectral or pan-sharpened. Topographic maps used are at scale 1:50.000 or larger. COTS navigation data for the road network are embedded. The AOIs are determined by DG Regio. The soil sealing layer based on FTS specifications are used for the degree of sealing. Ancillary data used, additionally to COTS navigation data are, Google Earth, local city maps, local zoning data, very high resolution imagery and field checks. The geometric resolution is 1:10.000 and the minimum mapping unit is 0.25 hectare. The positional accuracy is +/-5m. Minimum thematic accuracy for class “artificial surfaces” is 85%. For all classes, minimum thematic accuracy is 80%. The minimum overall accuracy for “artificial surfaces” must include both omission and commission errors. The data type is vector.

European Commission
EUR 25328 EN – Joint Research Centre – Institute for the Protection and Security of the Citizen

Title: **GHSL/UA Integration: Feasibility Report. Application of the JRC GHSL Image Information Extraction Protocol in the frame of the Urban Atlas product specifications**

Authors: Martino Pesaresi, Stamatia Halkia

Luxembourg: Publications Office of the European Union

2012 – 33 pp. – 21.0 x 29.7 cm

EUR – Scientific and Technical Research series – ISSN 1831-9424 (online), ISSN 1018-5593 (print)

ISBN 978-92-79-24933-4

doi:10.2788/26820

Abstract

The Joint Research Centre, in pursuing its mission of scientific and technical support for policy making in Brussels, has planned and carried out a series of tests in collaboration with DG Regional Policy, in view of a possible integration of its in-house Global Human Settlement Layer (GHSL) technology to the European Urban Atlas (UA). The support is based on specific parameters derived from the application of the satellite-based methodology developed by the JRC for human settlement analysis. The Urban Atlas provides detailed and cost-effective digital mapping, ensuring that city planners have the most up-to-date and accurate data available, offering new tools to assess risks and opportunities, ranging from threat of natural disasters and impact of climate change, to identifying new infrastructure and public transport needs. The GHSL/UA integration would contribute to population disaggregation and risk and disaster management applications, as well as support regional planning in general.

LB-NA-25328-EN-N

As the Commission's in-house science service, the Joint Research Centre's mission is to provide EU policies with independent, evidence-based scientific and technical support throughout the whole policy cycle.

Working in close cooperation with policy Directorates-General, the JRC addresses key societal challenges while stimulating innovation through developing new standards, methods and tools, and sharing and transferring its know-how to the Member States and international community.

Key policy areas include: environment and climate change; energy and transport; agriculture and food security; health and consumer protection; information society and digital agenda; safety and security including nuclear; all supported through a cross-cutting and multi-disciplinary approach.

Publications Office

ISBN 978-92-79-24933-4