



HORIZON 2020

The EU Framework Programme for Research and Innovation

H2020-EO-1-2014

Urban Surface morphology land cover/use and characteristics

Deliverable D3.1



LEAD AUTHOR

Fabio Del Frate (GEO-K)

DATE

1 January 2016

ISSUE

1.0

GRANT AGREEMENT

no 637519

DISSEMINATION LEVEL

PU

AUTHORS

Daniele Latini (GEO-K), Wieke Heldens (DLR),
Fredrik Lindberg (UoG), Zina Mitraka (FORTH),
Sue Grimmond (UoR), Andrew Mark Gabey (UoR)
Frans Olofson (UoG), Christian Feigenwinter
(UNIBAS), Ahmad Albitar (CESBIO)

CONTRIBUTORS

Nektarios Chrysoulakis (FORTH)



CONTENTS

1	Introduction	3
1.1	Purpose of the document	3
1.2	Definitions and acronyms.....	3
1.3	Document references	4
2	Project Overview	6
3	Generated Products	8
3.1	Morphology Products	10
3.1.1	Input	11
3.1.2	Geoprocessing.....	11
3.1.3	Extrapolation.....	13
3.1.4	Accuracies	13
3.2	Land Cover Products.....	14
3.2.1	Input	14
3.2.2	Output.....	15
3.2.3	Accuracies	18
3.3	Biophysical Products.....	22
3.3.1	Input.....	22
3.3.2	Output and Accuracies.....	24
4	Organization of the Data Base	28
4.1	Data Description	28
4.2	Metadata Description.....	28
4.3	Data Directories	29

1 INTRODUCTION

1.1 Purpose of the document

This document aims at describing the products regarding land surface yielded within the URBANFLUXES project. In fact, a quantitative description of surface properties has a significant impact on the estimation of the urban energy flux. The document also explains the organization of the Data Base including original EO data, ancillary data and derived products. The activity has been performed in WP3 of the project and received as input the user requirements. For the three test urban areas considered (Basel, London, and Heraklion) two classes of parameters have been considered: morphology parameters and surface parameters. More in detail, surface parameters can be distinguished in land cover fractions and geophysical parameters. The EO products described in this document are important input for WPs 4, 5 and 6 of the project.

After a general introduction of the URBANFLUXES project, the document describes the land products based on three main categories: morphology, land cover, biophysical characteristics. For each category the products are specified in terms of the input/output schemes necessary to yield them and of their accuracies. A final section will focus on the organization of the public access data base generated within the project.

All the products this documents describes, are publically available through the project's web site www.urbanfluxes.eu after simple registration.

1.2 Definitions and acronyms

Acronyms

ASTER	Advanced Spaceborne Thermal Emission and Reflection Radiometer
BOA	Bottom Of Atmosphere
CATDS	Centre Aval de Traitement des Données SMOS
CNES	Centre National d'Etudes Spatiales
CoP	Community of Practice
DEM	Digital Elevation Model
DNB	Day Night Band
DSM	Digital Surface Model
DSS	Decision Support System
ESA	European Space Agency
EO	Earth Observation
GUF	Global Urban Footprint
GIS	Geographic Information System
QGIS	Quantum GIS
HR	High Resolution
IFREMER	French Research Institute for Exploitation of the Sea

LAI	Leaf Area Index
LCZ	Local Climate Zones
LIDAR	Light Detection and Ranging
MERIS	MEdium Resolution Imaging Spectrometer
MIRAS	MIR Infrared Atmospheric Spectrometer
NA	Not Applicable
NDVI	Normalized Difference Vegetation Index
RADAR	Radio Detection and Ranging
RMSE	Root Mean Square Error
SMOS	Soil Moisture Ocean Salinity
SPOT	Satellite Pour l'Observation de la Terre
SRTM	Shuttle Radar Topography Mission
SVF	Sky View Factor
SWIR	Short Wave InfraRed
TIFF	Tag Image File Format
UEB	Urban Energy Budget
UHI	Urban Heat Island
UMEP	Urban Multi-scale Environmental Predictor
URBANFLUXES	URBan ANthropogenic heat FLUX from Earth observation Satellites
UTM	Universe Transverse Mercator
VHR	Very High Resolution
WP	Work Package
XML	Extensible Markup Language

1.3 Document references

Del Frate F., Pacifici F., Schiavon G., Solimini C. (2007), "Use of neural networks for automatic classification from high resolution imagery," IEEE Transactions on Geoscience and Remote Sensing, vol. 45, n. 4, pp. 800-809

Liao, L. B.; Weiss, Stephanie; Mills, Steve; Hauss, Bruce (2013): Suomi NPP VIIRS day-night band on-orbit performance. In: J. Geophys. Res. Atmos. 118 (22), S. 12,705. DOI: 10.1002/2013JD020475.

Lindberg, F. and Grimmond, C. (2011a) Nature of vegetation and building morphology characteristics across a city: Influence on shadow patterns and mean radiant temperatures in London. Urban Ecosystems 14:4, 617-634.

Lindberg, F. and Grimmond, C. (2011b) The influence of vegetation and building morphology on shadow patterns and mean radiant temperature in urban areas: model development and evaluation. Theoretical and Applied Climatology 105:3, 311-323.

Lindberg, F., Grimmond, CSB., Onomura, S., Järvi, L. and Ward, H. (2015a) UMEP - An integrated tool for urban climatology and climate-sensitive planning applications. ICUC9 – 9

the International Conference on Urban Climate jointly with 12th Symposium on the Urban Environment, TUKUP7 (cont): Warning plans & Decision support tools. Toulouse, July 2015.

Lindberg, F., Grimmond, C.S.B. and Martilli, A. (2015b) Sunlit fractions on urban facets - Impact of spatial resolution and approach. *Urban Climate*. 12:65-84.

Lindberg, F. and Grimmond, C. (2010) Continuous sky view factor from high resolution urban digital elevation models. *Climate Research* 42:3, 177-183.

Martilli, A., (2009). On the derivation of input parameters for urban canopy models from urban morphological datasets. *Bound.-Layer Meteorol.* 130, 301–306.

Mitraka, Z., Chrysoulakis, N., Doxani, G., Del Frate, F., & Berger, M. (2015). Urban Surface Temperature Time Series Estimation at the Local Scale by Spatial-Spectral Unmixing of Satellite Observations. *Remote Sensing*, 7(4), 4139–4156. doi:10.3390/rs70404139

Mitraka, Z., Chrysoulakis, N., Kamarianakis, Y., Partsinevelos, P., & Tsouchlaraki, A. (2012). Improving the estimation of urban surface emissivity based on sub-pixel classification of high resolution satellite imagery. *Remote Sensing of Environment*, 117, 125–134. doi:10.1016/j.rse.2011.06.025

Richter, R. (1998): Correction of satellite imagery over mountainous terrain. In: *Appl. Opt.* 37 (18), S. 4004–4015. DOI: 10.1364/AO.37.004004.

Richter, R. and Schläpfer, D. (2002): Geo-atmospheric processing of airborne imaging spectrometry data. Part 2: Atmospheric/topographic correction. In: *Int. J. Remote Sens.* 23 (13), S. 2631–2649. DOI: 10.1080/01431160110115834.

Richter R. and Schläpfer D. (2015): ATCOR-2/3 User Guide. Version 9.0.0. DLR & ReSe Applications.

Tum, M.; Böttcher, M.; Baret, F.; Günther, K. P.; Bittner, M.; Brockmann, C.; Weiss, M. (2016): Global gap-free MERIS LAI time series (2002-2012). In: *Remote Sensing*, 8(1), 69. DOI. 10.3390/rs8010069 UF_WP3_requirements

Watson ID, Johnson GT (1987) Graphical estimation of skyview factors in urban environments. *J Climatol* 7:193–197

Zhu, Z.; Wang, S.; Woodcock, C. E. (2015): Improvement and expansion of the Fmask algorithm: cloud, cloud shadow, and snow detection for Landsats 4–7, 8, and Sentinel 2 images. In: *Remote Sens. Environ.* 159, S. 269–277. DOI: 10.1016/j.rse.2014.12.014.

2 PROJECT OVERVIEW

The anthropogenic heat flux (Q_F) is the heat flux resulting from vehicular emissions, space heating and cooling of buildings, industrial processing and the metabolic heat release by people. Both urban planning and Earth system science communities need spatially disaggregated Q_F data, at local (neighbourhood, or areas larger than the order of 100 m x 100 m) and city scales. Such information is practically impossible to derive by point *in-situ* fluxes measurements, while satellite remote sensing is a valuable tool for estimating Urban Energy Budget (UEB) parameters exploiting Earth Observation (EO) data. While the estimation of Q_F spatial patterns by current EO systems is a scientific challenge, the major challenge lies on the innovative exploitation of the Copernicus Sentinels synergistic observations to estimate the spatiotemporal patterns of Q_F and all other UEB fluxes.

The main goal of URBANFLUXES is to investigate the potential of EO to retrieve Q_F , supported by simple meteorological measurements. The main research question addresses whether EO is able to provide reliable estimates of Q_F for the time of the satellite acquisition. URBANFLUXES answers this question by investigating the potential of EO to retrieve Q_F spatial patterns, by developing a method capable of deriving Q_F from current and future EO systems. URBANFLUXES aims to develop an EO-based methodology easily transferable to any urban area and capable of providing Q_F benchmark data for different applications. URBANFLUXES is expected to increase the value of EO data for scientific analyses and future emerging applications (such as urban planning and local/regional level climate change mitigation/adaptation), by exploiting the improved data quality, coverage and revisit times of the Copernicus Sentinels data. To this end, the specific objectives of the project are:

- to improve the accuracy of the radiation balance spatial distribution calculation;
- to develop EO-based methods to estimate the flux of heat storage in the urban fabric, as well as the turbulent sensible and latent heat fluxes at local scale;
- to employ energy budget closure to estimate the anthropogenic heat flux patterns;
- to specify and analyse the uncertainties associated with the derived products;
- to evaluate the products by comparisons with Q_F estimations by independent methods;
- to improve the understanding of the impact of Q_F on urban climate; and to communicate this understanding to the urban planning community, which will in turn lead to a better understanding of what new knowledge is needed on the ground;
- to exploit Sentinels 2 and 3 synergistic observations to retrieve UEB fluxes at the local scale, with the frequency of the Sentinel 3 series acquisitions.
- to standardise the resulting products, and by organizing an effective dissemination mechanism, to enhance their use by urban planners and decision makers in cities, as well as by EO scientists, Earth system modellers and urban climatologists.

The duration of URBANFLUXES is three years and it is divided into two main phases: during the 1st Phase an analysis method is being developed to estimate Q_F spatial patterns using currently available satellite data; during the 2nd Phase the developed method will be adapted to Sentinels synergy to derive Q_F spatiotemporal patterns. Three different urban areas are selected in URBANFLUXES as case studies: a highly urbanized mega city (London); a typical central European medium size city, that requires a substantial amount of energy for heating (Basel); and a smaller, low latitude Mediterranean city that requires a substantial amount of energy for cooling (Heraklion). The project uses a Community of Practice (CoP) approach, which means that in the case studies, local stakeholders and scientists meet on a regular basis to learn from each other and to make clear what aspects are important for the future users of the URBANFLUXES products.

URBANFLUXES is expected to generate a novel analysis method for estimation of UEB components from Copernicus data, enabling its integration into applications and operational services; for example to: develop rules of thumb for density and green space ratio, distinguish between insulated and non-insulated buildings and evaluate the implementation of climate change mitigation technologies, such as solar-screening and green-belting.

Despite its local importance, Q_F is omitted from climate models simulations. Observations of global temperature evolution indicate a pronounced warming over the last 150 years, with an increase in the occurrence of heat waves. The added value and benefit expected to emerge from URBANFLUXES is therefore related to quality of life, because it is expected to improve our understanding of the contribution of Q_F to heat wave intensity and thus to allow insight into strategies for mitigation. Q_F estimates are needed for all cities to be able to document the magnitude of the fluxes effects on urban climate so that the impact of Q_F can be included in climate modelling. URBANFLUXES is therefore expected to advance the current knowledge of the impacts of Q_F on urban heat island and hence on urban climate, and consequently on energy consumption in cities. This will lead to the development of tools and strategies to mitigate these effects, improving thermal comfort (social benefit) and energy efficiency (economic benefit). The long term operation of the Sentinels series guarantees the future supply of satellite observations, providing the means for the development and realization of the URBANFLUXES methodology.

URBANFLUXES is expected to support sustainable planning strategies relevant to climate change mitigation and adaptation in cities, because knowledge of Q_F spatio-temporal patterns is important for urban planning (e.g. to reduce or prevent Q_F hot spots), health (e.g. to estimate the impact on thermal comfort) and future proofing (e.g. to plan and implement interventions towards Q_F reduction in these areas). Planning tools, such as Urban Climatic Maps and Climatope Maps, should be enriched with information on Q_F patterns. Mapping provides visualization of assessments of these phenomena to help planners, developers and policy makers make better decisions on mitigation and adaptation.

3 GENERATED PRODUCTS

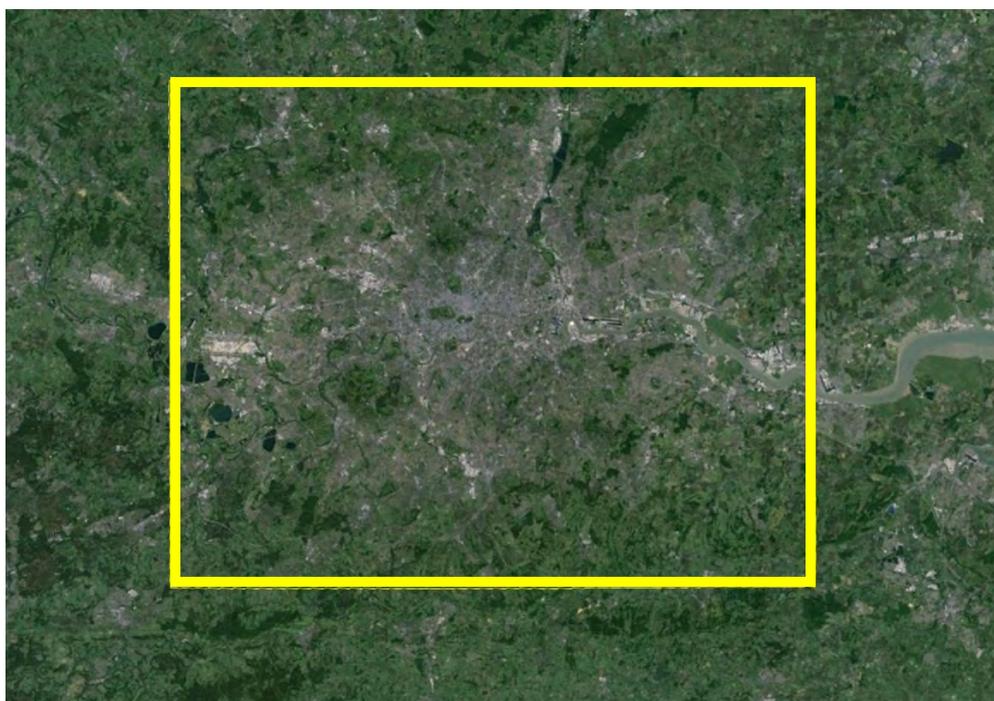
This section describes the base data used to calculate the fluxes in WPs 4, 5, and 6. All output data have these common characteristics, in particular:

- UTM projection
- TIFF Output format

For the three cities a focus area of interest has been selected. The three areas are shown in Figure 1 within the yellow boxes using Google Earth as source of images.



(a)



(b)



(c)

Figure 1: Areas of interest for (a) Basel (about 760 km²), (b) London (about 2700 km²) and (c) Heraklion (about 130 km²)

Other aspects depend on the specific type of product and will be addressed in the next subsections.

3.1 Morphology Products

The morphology product for the three study areas are derived from high resolution surface models. The surface models are generally derived from either datasets such as LiDAR (Light Detecting and Ranging) or vector based polygon and/or point data. Surface information on building morphology as well as vegetation is required. The parameters derived are given in Table 1. All parameters except for the H/W-ratio are calculated for both vegetation and buildings separately. Other obstacles are ignored as the digital surface models are not detailed enough.

Table 1: Urban morphometrical parameters derived. Above ground level (agl)

Symbol	Name	Definition	Units
$\lambda_p = f_b$	Plan area index (fraction of roof area)	Building plan area per total plan area	0-1
f_v	Fraction vegetated	Plan area vegetated per total plan area	0-1
f_i	Fraction impervious	Plan area of roads, sidewalks etc (excluding building) per total area	0-1
f_p	Fraction pervious	$f_p = f_w + f_v$	
f_w	Fraction water	Plan area of water (rivers, lake, fountain etc) per total area	0-1
λ_f	Frontal area index	Area of windward building walls relative to total plan area	
λ_w	Wall area fraction (2 wall infinite canyon)	fraction of the wall area relative to the total plan area	
z_B/W	Height to width ratio	Mean height of buildings to the mean width of "street" canyons. All manufactures taller than 3 m have been considered as buildings in the calculations.	
z_H	Mean obstacle height	includes buildings, trees, power poles etc that are resolved in the data source	m agl
$z_{H\sigma}$	Standard deviation of obstacle height		m
z_{Hmax}	Maximum obstacle height		m agl
SVF	Sky View Factor	Fraction of sky hemisphere not blocked (0 fully blocked, 1 complete sky).	0-1

3.1.1 Input

A Digital Surface Model (DSM) with both ground and object (building) heights needs to accompany a Digital Elevation Model (DEM) to relate all heights to the ground surface. Where available (London and Basel), LiDAR data has been used to derive the DSMs according to Lindberg and Grimmond (2011a) whereas for Heraklion the DSM was produced from analysis of aerial stereo-imagery. Figure 2 shows an example of surface models needed for the morphology products.

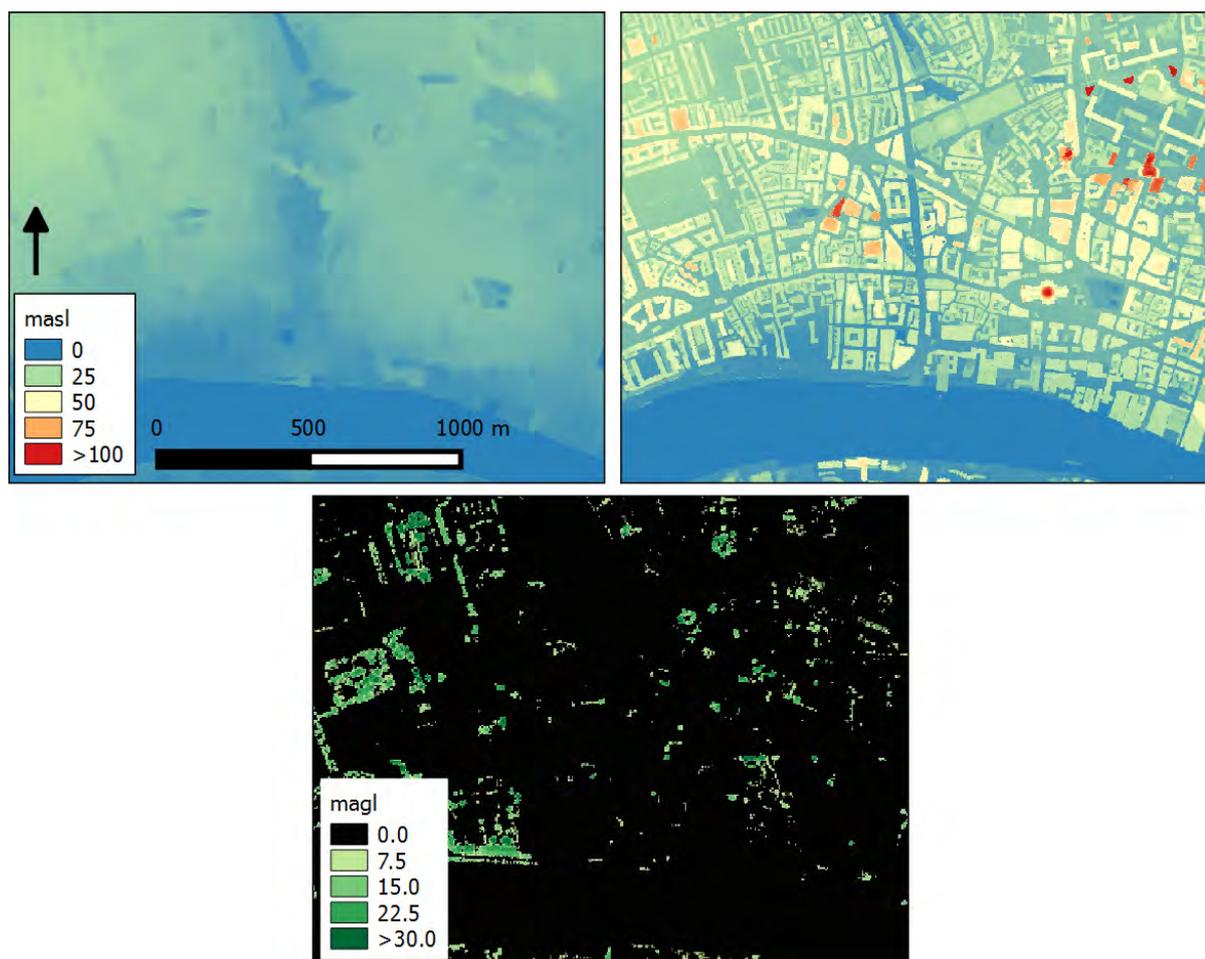


Figure 2: Examples of surface models from central London; a) DEM, b) ground and building DSM, c) Vegetation canopy DSM. Heights are expressed in metres above sea level (masl) and metres above ground level (magl)

3.1.2 Geoprocessing

Most of the morphometric parameters are derived using a newly developed tool, the Urban Multi-scale Environmental Predictor, UMEP (Lindberg et al., 2015a). UMEP is designed to be a

climate service tool, designed for researchers, architects and urban planners. The tool can be used for a variety of applications related to outdoor thermal comfort, urban energy consumption, climate change mitigation etc. UMEP consists of a coupled modelling system which combines “state of the art” 1D and 2D models related to the processes essential for scale independent urban climate estimations. UMEP is a community, open source model, where users can contribute as well as extend the tool to improve modelling capabilities. A major feature is the ability for a user to interact with spatial information to determine model parameters. The spatial data across a range of scales and sources are accessed through QGIS – a cross-platform, free, open source desktop geographic information systems (GIS) application - that provides data viewing, editing, and analysis capabilities. The tool includes a sub-plugin (Image Morphometric Calculator) to calculate the morphometric parameters on a grid, H/W-ratio and SVF excluded (see below).

The morphometric parameters are considered isotropically except for λ_F that is dependent on wind direction. Parameters such as z_H , and λ_P are easily derived with a zonal statistic approach for each grid (100 m x 100 m). λ_F is derived using a vector from north to south throughout the grid. A filter separating out windward facing walls, along the vector, is used in the calculation. To consider all wind directions the grid is rotated based on a user defined interval (here 5°). The grid is rotated instead of the vector since the number of pixels considered would differ depending on direction.

In addition to using UMEP approach, z_H is also calculated directly on the building DSM using an aggregation approach. The information of the building DSM is aggregated per 100 m grid cell into a mean value (z_H). At the same time, also additional statistics over the pixels contributing to this grid cell is saved. These statistics include the minimum and maximum building height, the standard deviation, and the 25th and 75th percentile. If the data are available, this morphological information can also be calculated for the vegetation height (this is the case for London).

Here the transformation of the real 3D urban morphology is simplified following the techniques used for urban land surface modelling where vegetation is not integrated. The surface is simplified to only consist of three facets (roof, walls and ground facets). This are a variety of techniques to do this (Lindberg et al., 2015b). The method used conserves heat and momentum fluxes by assuming there is an infinitely long canyon with one road, one roof and two walls (Martilli, 2009):

$$H/W = \frac{\lambda_W}{2(1-\lambda_P)}$$

The plan area index (λ_P) is derived from a high resolution building DSM in conjunction with a DEM. Vector data can also be used, e.g. a polygon building footprint dataset. Deriving the fraction of wall area is based on the data format (i.e. vector or raster), different approaches

need to be applied. For a vector dataset, the topological structure or the accuracy of the data are of absolute importance. Vector data ideally consist of a building footprint polygon (or polyline) layer with height information embedded in the object structure alternatively included in appurtenant attribute tables. Using vector data with full roof structures description (i.e. including objects such as chimneys etc.) makes it very complicated to derive wall areas and conversion to a raster dataset is recommended. The accuracy of the vector data becomes especially important where different building segments are located at the same position (e.g. two attached buildings). Those segments need to represent the difference in height between the two building roofs. This type of information is extremely rare and needs to be derived using geoprocessing techniques. A direct conversion of linear vector walls will result in an overestimation of wall areas (see Appendix in Lindberg et al. 2015). Here, a 4-directional 3 x 3 kernel majority filter on a DSM is applied and then the differences between the original DSM and the raster produced by the filtering process are identified. By setting a threshold limit of the height that should represent a wall (e.g. 3 m), wall pixels are identified. The wall area is calculated in a sub-plugin in UMEP (Wall Aspect and Height).

The Sky View Factor (SVF) is calculated based on Lindberg and Grimmond (2010) and Lindberg and Grimmond (2011b). By definition, SVF is the ratio of the radiation received (or emitted) by a planar surface to the radiation emitted (or received) by the entire hemispheric environment (Watson and Johnson 1987). It is a dimensionless measure between zero and one, representing totally obstructed and free spaces, respectively.

The morphology products are all aggregated to the common 100 m grid which is used for the final products in URBANFLUXES.

3.1.3 Extrapolation

Currently, morphology parameters can only be calculated for areas and dates that a high resolution DSM is available for. To extend this, a (LiDAR or RaDAR) DSM for a larger area or a new time step could be used. Potential data sets to be explored are: TerraSar-X based GUF® and/or building volume map that covers a larger area than the LiDAR data sets. Also Sentinel 1 will provide valuable information on the surface morphology. The relationship between the standard morphology parameters from high resolution data and radar based parameters will therefore be assessed using statistical techniques.

3.1.4 Accuracies

The quality of the output data is only related to the precision and accuracy of the surface model used. A precision (pixel resolution) of 1 meter is used for all three case study areas.

3.2 Land Cover Products

The land cover products are obtained considering EO data at both HR and VHR spatial resolutions. Ancillary data provided by the “Urban Atlas” repository (<http://www.eea.europa.eu/data-and-maps/data/urban-atlas>) have also been used. The Urban Atlas is a joint initiative of the European Commission Directorate-General for Regional and Urban Policy and the Directorate-General for Enterprise and Industry with the support of the European Space Agency and the European Environment Agency. The initiative is providing pan-European comparable land use and land cover data for Functional Urban Areas (FUA) at very high spatial resolution (<http://land.copernicus.eu/local/urban-atlas>)

3.2.1 Input

For the HR case, with a pixel size of 30 m, the project has taken as reference input Landsat images while SPOT products and Urban Atlas have been used for the VHR maps. For Landsat data, Landsat 8 acquisitions have been considered due to their ready availability and wide coverage. The used data, atmospherically calibrated, are characterized by the 7 following spectral bands:

- Coastal erosion (0.44 μm)
- Blue (0.48 μm)
- Green (0.56 μm)
- Red (0.65 μm)
- Near Infrared (0.86 μm)
- SWIR 1 (1.6 μm)
- SWIR 2 (2.2 μm)

ASTER and Sentinel-2 data can be also used to generate the same output products described in the next subsection while Sentinel-1 data are exploited for the updates of the classification maps. It has to be noted that acquisitions taken at different seasons of the year have been considered to improve the discrimination between some particular classes such as bares soils and agricultural areas. For the VHR case, with a pixel size of 2.5 m, SPOT 5 data have been selected as the reference EO data. As a matter of fact, these were the only VHR data available for the three areas of interest. The SPOT provided bands are:

- Green: (0.56 μm)
- Red: (0.65 μm)
- Near Infrared (0.84 μm)

As before for ASTER and Sentinel 1, also in this case the procedures have been designed to easily incorporate different new image products, possibly characterized by better spatial or spectral resolutions (for example Worldview data).

3.2.2 Output

As far as the land cover products are concerned, the project has yielded the following types of outputs for each of the three city areas of Basel, London and Heraklion considered in the project:

- HR (30 m spatial resolution) Land cover classification maps with the following classes considered: Bare Soil, Grassland, Agriculture, Woodland, Low Residential, High Residential, Industrial, Water
- VHR (2.5 spatial resolution) Land cover classification maps, with the following classes considered: Bare Soil, Grassland, Agriculture, Woodland, Roads, Built Residential, Built Industrial, Water
- Abundances maps computed on a cell grid of 100 m for the following classes: Bare Soil, Agriculture, Grassland, Woodland, Urban (Roads + Built Residential in the VHR maps), Industrial, Water
- Imperviousness maps computed on a 100 m grid. The imperviousness is defined as percentage of impervious surface, the sum of the fraction of urban (built residential in the VHR maps) and industrial areas as resulting from the land cover maps

The spatial resolution of 100 m for the third and fourth product has been found the most suitable for the subsequent radiative transfer models. The land cover classification maps have to be understood as "hard" classifications so the pixel is assigned with the prevailing class in the pixel area. It has to be underlined that abundances maps and imperviousness maps stem directly from land cover classification maps. Conversely, the generation of the land cover classification maps consists of two processing steps: the first one is automatic and based on neural networks algorithms (Del Frate et al., 2007), the second one manually refines the previous output using photo-interpretation. Some comments should be reported regarding the definition of the land cover classes. For example, in the HR (Landsat) case, the discrimination between Low Residential and High Residential is based on the assumption that the difference is represented by the average number of stories characterizing the buildings in that pixel. This is consistent with the guidelines applied in the Urban Atlas products. When roads such as highways or motorways are visible in the HR image they are assigned to the Low Residential class. Both in HR and VHR case vineyards and olive yards are assigned with the class Agriculture. The definition of the other classes is rather straightforward. The generation of land cover classification maps at two different spatial resolutions is motivated by at least a couple of reasons. First of all it is important to evaluate the impact of a coarser spatial resolution of the EO data on the estimation of the anthropogenic heat flux Q_F . Moreover lower spatial resolutions can be more appropriate for detecting changed areas when the generation of updated products is required. In fact, for this latter task the project intends to fully exploit Sentinel missions.

One or more examples for each type of the aforementioned products are shown in the following figures.

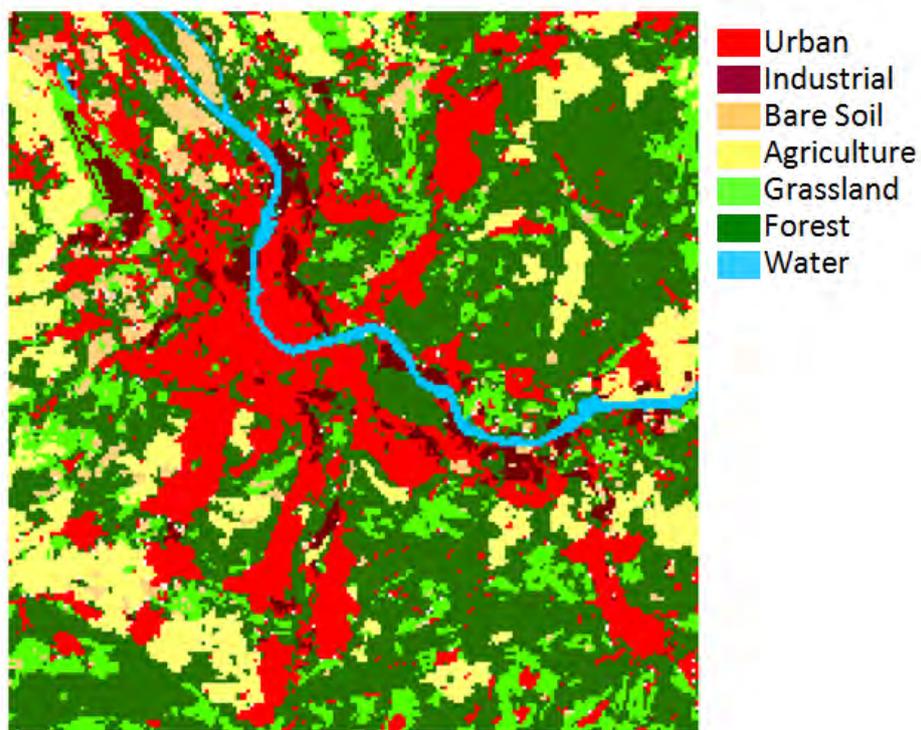


Figure 3: Land cover classification map at HR for the city of Basel

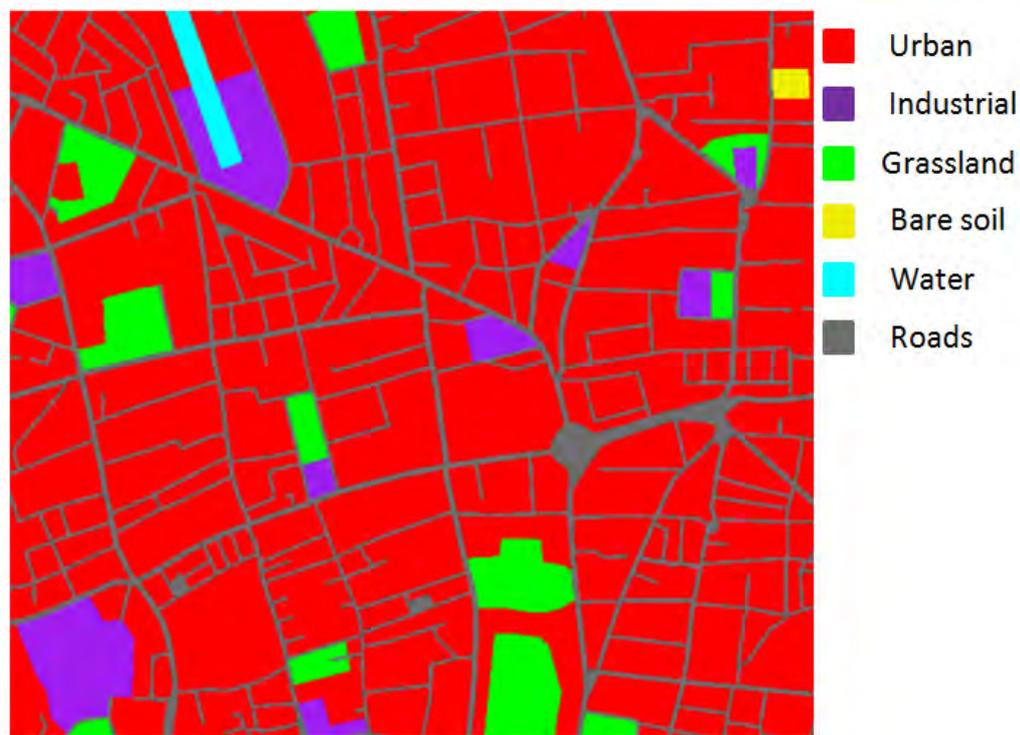


Figure 4: Land cover classification map at VHR for the city of London (detail)

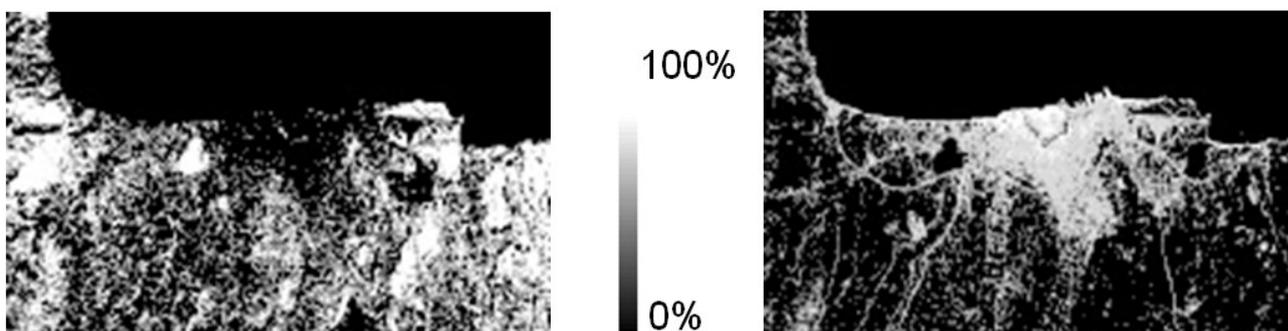


Figure 5: Abundancies maps for Heraklion: Bare Soil (left), Urban (right)

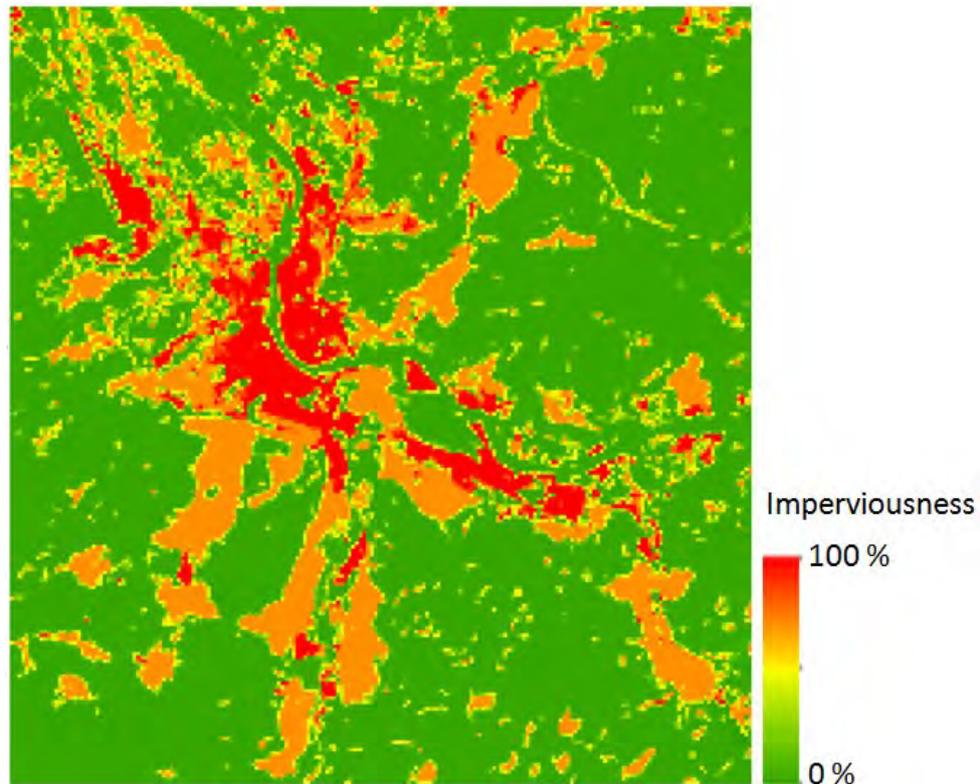


Figure 6: Example of imperviousness product for the city of Basel

3.2.3 Accuracies

For each of the product reported in the previous subsections accuracies figures have been computed. For land cover classification maps validation has been carried out by means of confusion matrices. To assure independent evaluation of the results, the validation activities have been performed by a team not involved in the generation of the maps. A statistically significant number of points have been randomly generated over the area of interest on Google Earth. For each point the ground truth has been estimated by photo-interpretation on Google Earth. The ground-truth has then been compared with the corresponding classification resulting from the produced map. As far as the number of the necessary validating points is concerned we used the following scheme: we started with a certain number of N points and computed the elements of the confusion matrix. After that we generated a new set of N points and computed the new confusion matrix on the overall $2N$ points. We iterated this procedure until the new creation of N points generated an accuracy variation greater than 0.5% on the overall accuracy and/or at least on one element of the confusion matrix. When the impact of a new set of points was less than 0.5% either for the overall accuracy or for each element of the confusion matrix no additional validation points have been considered. According to the described procedure the following confusion matrices have been obtained for the three cities and for the 2 types of products:

Basel HR

CLASSIFIED AS	TRUE							
	Water	Grassland	Woodland	Agriculture	Bare Soil	Urban	Dense Urban	Industrial
Water	22	1	0	0	0	0	0	0
Grassland	0	37	10	2	1	6	0	4
Woodland	0	5	291	19	0	5	1	1
Agriculture	0	5	11	295	0	4	0	2
Bare Soil	0	0	0	0	8	0	0	0
Urban	0	5	11	9	2	102	0	5
Dense Urban	0	0	1	0	0	1	12	4
industrial	0	0	2	0	1	1	0	59
Overall Accuracy: 87.4%								

London HR

CLASSIFIED AS	TRUE							
	Water	Grassland	Woodland	Agriculture	Bare Soil	Urban	Dense Urban	Industrial
Water	19	1	0	0	0	0	0	0
Grassland	0	149	3	6	0	5	0	7
Woodland	1	42	54	14	1	11	0	3
Agriculture	0	4	0	171	0	0	0	2
Bare Soil	0	0	0	0	0	0	0	0
Urban	0	10	0	0	0	194	4	1
Dense Urban	0	0	0	0	0	2	30	1
industrial	0	4	0	0	0	1	0	52
Overall Accuracy: 84.5%								

Heraklion HR

CLASSIFIED AS	TRUE							
	Water	Grassland	Woodland	Agriculture	Bare Soil	Urban	Dense Urban	Industrial
Water	119	0	0	0	0	0	0	0
Grassland	0	0	0	0	0	0	0	0
Woodland	0	0	27	4	4	0	0	0
Agriculture	0	0	40	54	34	0	0	1
Bare Soil	0	0	0	6	37	4	0	0
Urban	0	0	3	2	4	37	1	1
Dense Urban	0	0	0	0	0	0	5	0
industrial	0	0	0	0	0	5	0	4
Overall Accuracy: 70.8								



Basel VHR

CLASSIFIED AS	TRUE							
	Water	Grassland	Woodland	Agriculture	Bare Soil	Urban	Asphalt	Industrial
Water	28	0	0	1	0	0	0	0
Grassland	0	55	27	3	2	4	5	6
Woodland	0	19	278	12	1	5	5	2
Agriculture	0	1	6	143	0	0	0	2
Bare Soil	0	0	0	1	2	0	0	0
Urban	0	6	15	1	0	143	10	14
Asphalt	0	2	2	1	0	12	37	3
industrial	0	1	4	0	0	0	2	109
Overall Accuracy: 81.6%								

London VHR

CLASSIFIED AS	TRUE							
	Water	Grassland	Woodland	Agriculture	Bare Soil	Urban	Asphalt	Industrial
Water	30	0	0	0	0	0	0	2
Grassland	0	105	10	0	0	15	5	2
Woodland	0	16	29	0	0	11	3	0
Agriculture	0	0	0	0	0	0	0	0
Bare Soil	0	0	0	0	0	0	0	0
Urban	0	10	5	0	1	394	39	8
Asphalt	0	8	1	0	0	33	61	2
industrial	0	1	1	0	1	10	8	80
Overall Accuracy: 78.5%								

Heraklion VHR

CLASSIFIED AS	TRUE							
	Water	Grassland	Woodland	Agriculture	Bare Soil	Urban	Asphalt	Industrial
Water	73	0	0	0	0	0	0	0
Grassland	0	0	0	0	0	0	0	0
Woodland	0	0	4	3	1	0	0	0
Agriculture	0	1	6	51	23	1	0	2
Bare Soil	0	0	2	33	2	2	1	0
Urban	0	0	0	2	3	10	3	0
Asphalt	0	0	1	3	1	3	8	0
industrial	0	0	0	0	0	2	2	1
Overall Accuracy: 61.1								

The accuracy values of the abundances and of the imperviousness maps are strictly linked to the classification accuracies above as these parameters are computed with a straightforward method from the classification maps. The label NA is used when the presence of the specific land cover class is not statistically significant in the considered test area.

BASEL ABUNDANCIES

Class	Average Error
Water	0.1%
Grassland	1.9%
Woodland	1.0%
Agriculture	1.0%
Bare Soil	0.2%
Urban	2.4%
Industrial	2.1%

LONDON ABUNDANCIES

Class	Average Error
Water	0.2%
Grassland	0.3%
Woodland	1.5%
Agriculture	NA
Bare Soil	NA
Urban	0.3%
Industrial	1.7%

HERAKLION ABUNDANCES

Class	Average Error
Water	0.0%
Grassland	NA
Woodland	2.0%
Agriculture	3.3%
Bare Soil	4.1%
Urban	0.8%
Industrial	0.8%

IMPERVIOUSNESS

AREA	Average Error
Basel	0.5%
London	0.1%
Heraklion	0.8%

3.3 Biophysical Products

For the LCZ mapping and the modelling of the urban fluxes information on biophysical properties in the case study areas is required. Especially for vegetation related products, time series are provided, to document the seasonal changes. Seven products have been derived and are described in the following sections:

- Reflectance
- Albedo
- Surface temperature
- NDVI
- LAI
- Soil moisture
- Nightlights

3.3.1 Input

Because of the specific requirements, biophysical products are derived from a range of sensors.

3.3.1.1 Reflectance

The reflectance image is currently derived from selected, cloud free Landsat 8 scenes. The absence of clouds is an important requirement for the selection. For London, the reflectance has been calculated for two scenes, July 17, 2013 and October 2, 2015. For Basel, two scenes April 24, 2015 and August 30, 2015 are used. For Heraklion, one scene from November 7, 2015 has been used.

3.3.1.2 Albedo

The broadband albedo is currently derived from selected Landsat 8 scenes (see section 3.3.1.2).

3.3.1.3 Surface temperature

The surface temperature is currently derived from the thermal bands from Landsat 8 (day time) and from ASTER (night time). The Landsat 8 scenes used for the surface temperature calculation are the same as those for the reflectance calculation. The available ASTER scenes are recorded in 2015 on August 17 and 25 (Basel), July 19, August 2 and 18 (Heraklion) and

August 6 (London) in the evening. To achieve the high temporal resolution of 4 images per day, data from MODIS are considered as well.

3.3.1.4 NDVI

The NDVI is derived from a time series of Landsat 8 data. Depending on the cloud cover, the number of images varies for each case study. For example for 2014, for Heraklion, 18 scenes are available. For Basel there are 24 scenes available, whereas for London only 10 scenes. Figure 7 shows the acquisitions of the available scenes and their quality regarding cloud cover.

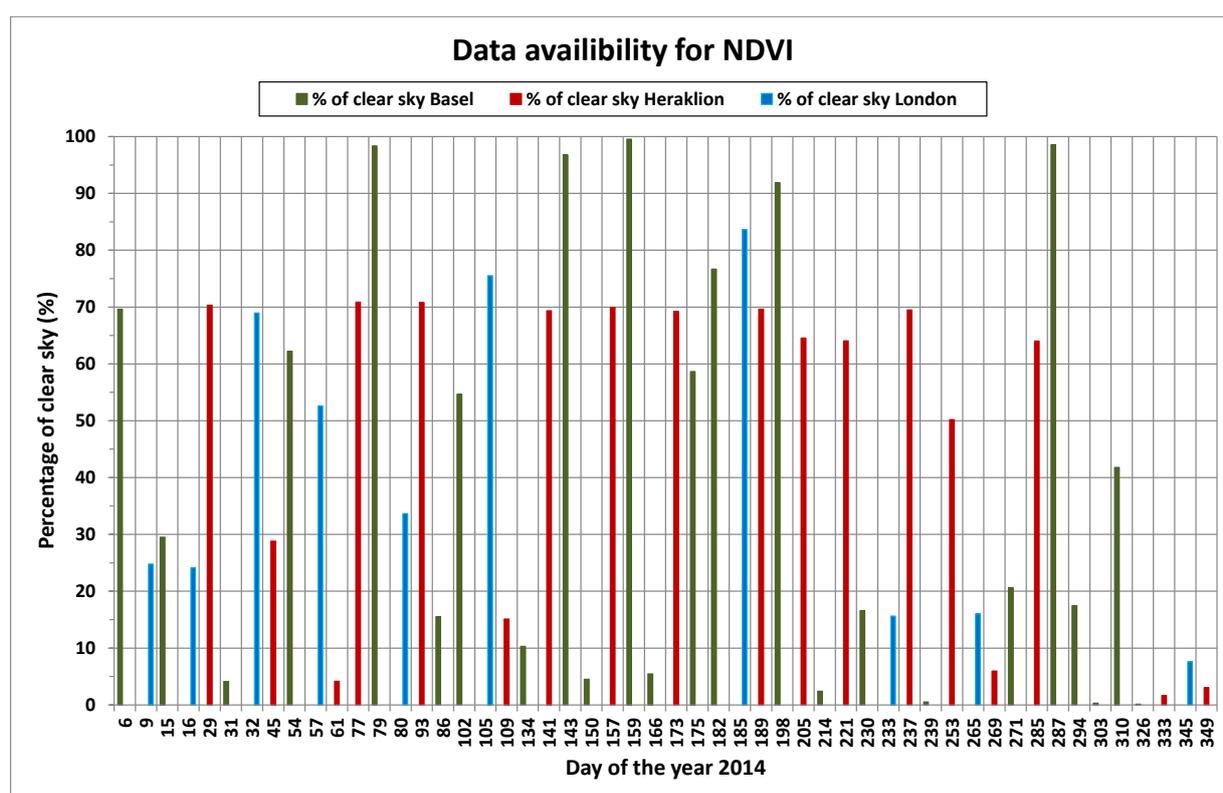


Figure 7: Data availability of Landsat 8 scenes used for NDVI calculation.

3.3.1.5 LAI

The LAI is derived from MERIS time series, making use of all available 10-day composites and available at eoweb.dlr.de (DOI: 10.15489/ak90g1wty909). Additionally, LAI is calculated for selected Landsat 8 scenes, accompanying the reflectance, LST and albedo images.

3.3.1.6 Soil Moisture

The soil moisture is a product of the Soil Moisture and Ocean Salinity (SMOS) mission of the European Space Agency (ESA) obtained from the instrument MIRAS. Data availability is assured through the ground segment CATDS (Centre Aval de Traitement des Données SMOS) operated for the "Centre National d'Etudes Spatiales" (CNES, France) by IFREMER (Brest, France). The data are currently derived from level 3 monthly averaged products. The monthly

resolution is obtained from aggregated daily global maps, whose number can vary from month to month.

3.3.1.7 Nightlights

The Nightlights images consist of Version 1 monthly averaged radiance composites using nighttime data from the Day/Night Band of the Visible Infrared Imaging Radiometer Suite (VIIRS). The data are obtained from the NPP satellite and produced by the Earth Observation Group (EOG) at NOAA/NGDC (National Geophysical Data Center). The Day/Night Band can detect the lighting from cities, towns, villages, combustion sources and lit fishing boats (low light imaging data). The number of data for every pixel that are used to construct the monthly radiance average can vary between months as the data are filtered to avoid data impacted by stray light, lightning, lunar illumination, and cloud-cover. To determine cloud cover the VIIRS Cloud Mask Product (VCM) was used. Therefore a second data set is provided indicating the NPP integer counts of the number of cloud-free coverages or observations that are used for the average. Additionally, data near the edges of the swath are not included in the composites (aggregation zones 29-32).

3.3.2 Output and Accuracies

For each of the six biophysical products, different methods are used to derive the information from the sensor data. This approach, the output (definition, resolution etc.) and the accuracy are described in the following sections.

3.3.2.1 Reflectance

The reflectance is derived from the Landsat 8 radiance images by using ATCOR-2 (Richter und Schläpfer 2015). This atmospheric correction software supports the estimation of aerosols and visibility. Additionally, topographic influences are corrected for by the use of an SRTM DEM. Using these estimations and the correction coefficients provided in the scene metadata, atmospherically corrected reflectance is calculated (BOA reflectance). ATCOR uses the bands 1-7 and 9 (Cirrus) of Landsat. The thermal bands are not considered here. The panchromatic band is corrected separately. In the reflectance images, the bands are sorted after wavelength. The cirrus channel then becomes band 6. Since this band is within the atmospheric water vapour window, it is replaced by an interpolated value between band 5 and 7. The unit of the data is reflectance in %. The original spatial resolution of 30 m is kept. The atmospheric correction of the panchromatic band is carried out with the same aerosol type as the other bands of that acquisition, the mean visibility and a 15 m DEM, resulting in a reflectance image with a spatial resolution of 15 m.

The accuracy of the atmospheric correction cannot be expressed in a single figure. The radiometric accuracy depends on the calibration accuracy of the sensor, the quality of geometric co-registration of the spectral bands, the accuracy of the radiative transfer code (MODTRAN 5), the correct choice of atmospheric input parameters, the terrain type (flat or

rugged), and the surface cover (Richter und Schläpfer 2015). For flat terrain the accuracy of the retrieved reflectance is between +/- 2 % and +/- 4% (Richter und Schläpfer 2002). For rugged terrain, which is the case for the URBANFLUXES case study areas, the accuracy depends strongly on the quality and matching of the DEM (Richter 1998). The accuracy varies with the terrain; the largest errors should be expected in areas with critical geometries, such as mountain ridges. Luckily, such extreme ridges are not present in the URBANFLUXES case study areas.

3.3.2.2 *Surface temperature*

Also for the calculation of surface temperature ATCOR can be used (Richter and Schläpfer 2015). ATCOR allows the use of variable emissivity based on land cover. This option, available for single band or multispectral thermal sensors, assumes various fixed emissivity values to solve the temperature-emissivity separation. This is the preferred approach when analyzing urban areas, because the emissivity of urban materials (asphalt, brick, metals) covers a larger range of values than the emissivity of natural materials (vegetation, water, soil). However, when using Landsat 8 data, ATCOR has to make use of a split window approach on band 10 and 11. For the calculation, an overall emissivity of 0.98 is assumed. A variable emissivity depending on land cover is not available within ATCOR for Landsat 8. However, an offset can be applied, e.g. to adjust the surface temperature to measured temperatures. The resulting images show the surface temperature in Celsius with a spatial resolution of 30 m. Additionally, a surface temperature map resampled to the 100 m grid is provided.

The accuracy of the surface temperature depends on the selected surface emissivity. If the selected emissivity deviates less than 2% from the real emissivity, the accuracy of the surface temperature can reach 1-2 K [15]. Each error of 1% emissivity will result in an additional error of 0.5-0.8 K surface temperature (Richter und Schläpfer 2015). In addition to the uncertainty induced by the temperature retrieval method, the TIRS data of Landsat 8 is currently suffering calibration issues (http://landsat.usgs.gov/calibration_notices.php and https://landsat.usgs.gov/l8handbook_appendixa.php). These calibration issues can result in an overestimation of the surface temperature by 2 K or more.

An approach to downscale low spatial resolution thermal measurements (i.e. MODIS or Sentinel-3) to local scale (hundreds of meters) will be applied within URBANFLUXES (Mitraka et al., 2015). The methodology to be used is a multi-step procedure for enhancing the spatial resolution of satellite thermal observations and estimating high-resolution emissivity, with an ultimate goal of deriving LST in high spatial and temporal resolution. The surface cover information is derived from high spatial resolution VNIR observations (tens of meters), and it is then used to improve the low spatial resolution of thermal measurements (hundreds of meters). Emissivity is estimated using the high spatial resolution VNIR data and information from spectral libraries, using spectral unmixing (Mitraka et al., 2012). The estimated emissivity and the downscaled thermal observations are then combined with atmospheric

information in a split-window algorithm, and they provide frequent LST estimations at the local scale. The method has already been tested with MODIS and Landsat data for Heraklion (Mitraka et al., 2015) and it will be expanded for and adjusted to Sentinel-2 and Sentinel-3 data for URBANFLUXES within WP8.

3.3.2.3 Albedo

The albedo is calculated for the same scenes as the reflectance using ATCOR (Richter and Schläpfer 2015). Within ATCOR the integrated hemispherical- directional reflectance weighted with the global flux on the ground (E_g) is used as a substitute for surface albedo (i.e. actually bi-hemispherical reflectance). This is calculated according to the following equation:

$$a = \frac{\int_{0.3\mu m}^{2.5\mu m} \rho(\lambda) E_g(\lambda) d\lambda}{\int_{0.3\mu m}^{2.5\mu m} E_g(\lambda) d\lambda}$$

Where a is albedo and ρ is reflectance at wavelength λ . Missing wavelength regions are interpolated before integration. For Landsat 8 this is the case for the 0.3-0.4 μm region. The reflectance for this region ($\rho_{0.3-0.4\mu m}$) is calculated as follows:

$$\rho_{0.3-0.4\mu m} = 0.8\rho_{0.45-0.50\mu m}$$

3.3.2.4 NDVI

The Normalized Differentiated Vegetation Index (NDVI) is defined as (NIR-red)/(NIR+red). For Landsat 8 this means in band numbers: (B5-B4)/(B5+B4). The employed processing chain at DLR first calculates a cloud mask for each image using Fmask (Zhu et al. 2015). Subsequently, NDVI is calculated. The resulting NDVI maps are masked according to the cloud mask, clipped to the study areas and resampled to the 100 m grid.

The accuracy of the NDVI map depends on the radiometric quality of the input data. However, discontinuous data availability due to cloud cover has a larger influence on the quality of this product.

3.3.2.5 LAI

The leaf area index (LAI) is calculated with a spatial resolution of 300 m from the MERIS level 1B data. For the calculation of the LAI, the BEAM MERIS vegetation processor was used. After the LAI processing, a time series analysis was applied to fill data gaps and filter outliers (Tum et al., under review). The result is a continuous data set. For URBANFLUXES this data set was clipped for the three study areas as well as resampled to the 100 m grid. This data set provides information on the phenology during the year. However, because of the course

resolution, also LAI is derived from the same images as the reflectance to have an up-to-date complete high resolution data set for the selected dates. This LAI is derived using ATCOR according to the following equation (Richter and Schläpfer 2015):

$$LAI = -\frac{1}{a_2} \ln\left(\frac{a_0 - VI}{a_1}\right)$$

Where $a_0 = 0.820$, $a_1 = 0.780$, $a_2 = 0.600$ and VI is a vegetation index, in this case SAVI:

$$SAVI = \frac{(\rho_{850} - \rho_{650}) * 1.5}{(\rho_{850} + \rho_{650} + 0.5)}$$

Where ρ_{850} and ρ_{650} is the reflectance at 850 nm and 650 nm, respectively.

Because here empirically derived parameters are used, the absolute LAI values may not be correct, but the approach allows capturing the seasonal trend.

3.3.2.6 Soil Moisture

The target of the SMOS mission is to provide a global image of soil moisture every one to three days with an accuracy of 4 % at a spatial resolution of 50 km. The unit is m^3/m^3 . As the original data has a different coordinate reference system (EASE-Grid, EPSG:6933) than used for this project, a transformation to UTM WGS84 has been performed. This data set was clipped according to the study areas and resampled to the 100 m grid, resulting in monthly SMOS data for the year 2014 for each study area. For the estimation of the accuracy, it is relevant to note that the number of daily maps included in the monthly average varies, depending on the satellite cover. However, there was no information available on these coverages.

3.3.2.7 Nightlights

The Nightlights data files contain floating point radiance values with units in $nW/(cm^2/sr)$ with a spatial resolution of 750 m. The original DNB radiance values have been multiplied by $1E9$. This was done to alleviate issues some software packages were having with the very small numbers in the original units. The DNB is panchromatic and its wavelength range starts from 500 nm up to 900 nm (Liao, L. B. et al. 2013). As the global data are divided into 6 tiles only tile number 2 (75N/060W) was needed for the extent of this project. The Nightlights data are clipped to the study area in each city and resampled to the 100 m grid.

The absolute low light calibration uncertainty was estimated at a moderate 15% because it relies on solar signal which is more than seven orders of magnitudes brighter than the low lights from fishing vessels. The long-term stability has yet to be characterized (Liao, L. B. et al. 2013).

4 ORGANIZATION OF THE DATA BASE

4.1 Data Description

All EO products yielded within the project are available in the URBANFLUXES website and made publicly accessible to registered users through <http://urbanfluxes.eu/data/>. Every file has an xml metadata file (with the same name) describing the data, describing the properties of the data file. The structure of the metadata files complies with the INSPIRE guidelines. Thus, XML metadata files are created and accompanying the data. The XML metadata files can be loaded and viewed through the INSPIRE Metadata Editor (<http://inspire-geoportal.ec.europa.eu/editor/>). In the next two subsections the metadata description and the directories organization inside the private ftp will be summarized.

4.2 Metadata Description

Table 2 lists all the fields that are used for the correct classification and description of the URBANFLUXES products, and the respective fields of the INSPIRE directive.

Table 2. List of mandatory for URBANFLUXES metadata fields

	Name of field	Name of the respective INSPIRE field
1	Owner/Publisher	Metadata → Organization name + email Responsible Party → Organization name + email + role
2	Title	Identification → Resource Title
3	File name	Identification → Identifier → Code
4	Short Description	Identification → Resource abstract + Resource locator
5	Topic category	Classification → Topic category
6	INSPIRE keyword	Keyword → Keyword from INSPIRE Data themes
7	Keywords	Keyword → Free keyword → Keyword value
8	Geographic location	Geographic → Geographic bounding box
9	Temporal Extent	Temporal → Temporal Extent
10	Reference Dates	Temporal → Date of Creation, Publication, last revision
11	Process history	Quality&Validity → Lineage
12	Spatial Resolution	Quality&Validity → Resolution distance + Unit of measure
13	Access and use	Constraints → Conditions applying to access and use + Limitations on public access
14	File size	<i>(automatic)</i>

4.3 Data Directories

Figure 10 represents the tree of the directories considered for the public data in the website

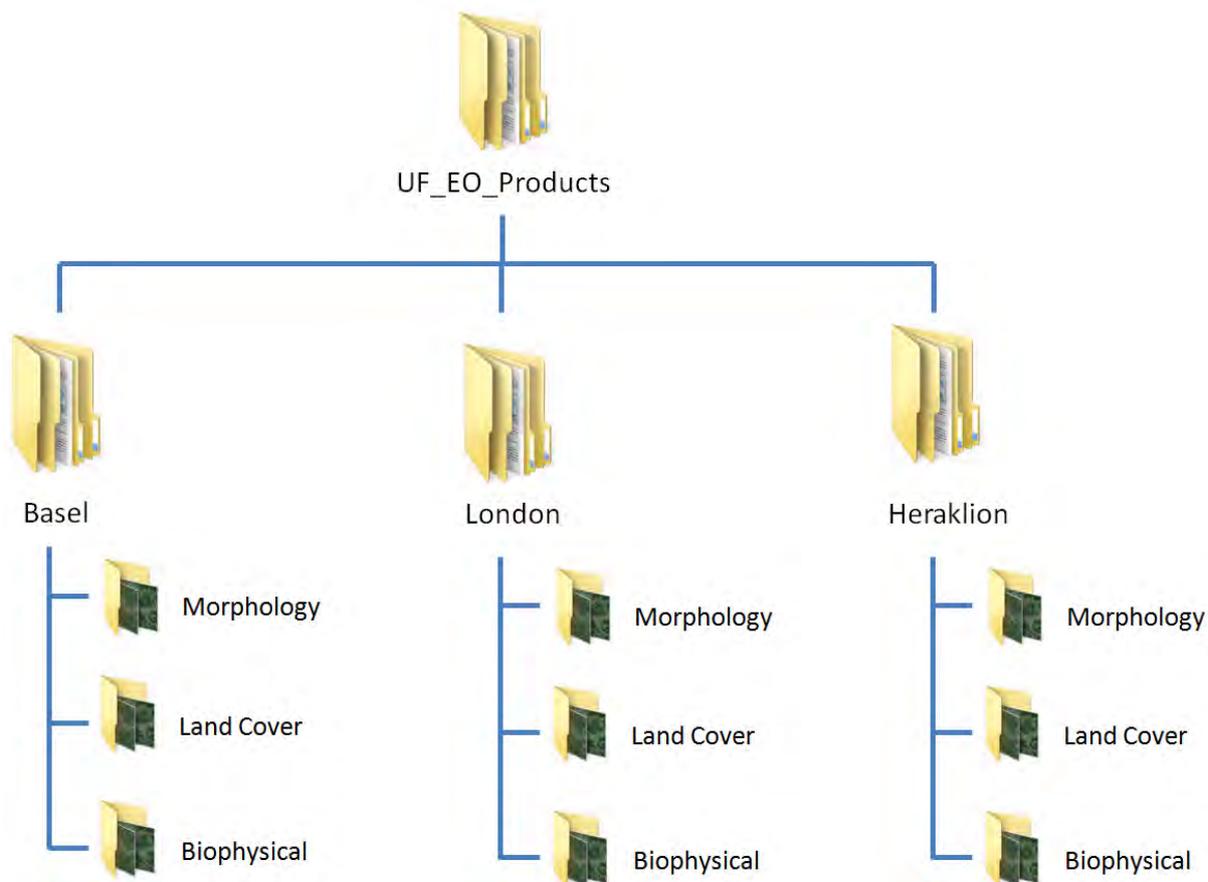


Figure 10: Tree structure for EO products (data and metadata) directories