Description of Environmental variables available through Exposome Maps

Version 1.0 09 November 2023

Table of contents

1. Introduction

This document provides information about the environmental exposure variables that are available through the EXPOPSOME maps application.

2. Physico-chemical environment

2.1 Air Pollution

2.1.1 From EXPANSE

Previous European land-use regression (LUR) models assumed fixed linear relationships between air pollution concentrations and predictors such as traffic and land use. We evaluated whether including spatially-varying relationships could improve European LUR models by using geographically weighted regression (GWR) and random forest (RF). We built separate LUR models for each year from 2000 to 2019 for NO2, O3, PM2.5 and PM10 using annual average monitoring observations across Europe. Potential predictors included satellite retrievals, chemical transport model estimates and land-use variables. Supervised linear regression (SLR) was used to select predictors, and then GWR estimated the potentially spatially-varying coefficients. We developed multi-year models using geographically and temporally weighted regression (GTWR). Five-fold cross-validation per year showed that GWR and GTWR explained similar spatial variations in annual average concentrations (average R2 = NO2: 0.66; O3: 0.58; PM10: 0.62; PM2.5: 0.77), which are better than SLR assuming fixed coefficients (average R2 = NO2: 0.61; O3: 0.46; PM10: 0.51; PM2.5: 0.75) and RF (average R2 = NO2: 0.64; O3: 0.53; PM10: 0.56; PM2.5: 0.67). The GTWR predictions and a previously-used method of back-extrapolating 2010 model predictions using CTM were overall highly correlated (R2 > 0.8) for all pollutants. Including spatially-varying relationships using GWR modestly improved European air pollution annual LUR models, allowing time-varying exposure-health risk models (Shen et al., 2022).

2.1.2 From ELAPSE

West-European land use regression models (LUR) were developed for 2010 estimating annual mean PM2.5, NO2, BC and O3 concentrations (including cold and warm season estimates for O3 (de Hoogh et al., 2018)). The models were based on AirBase routine monitoring data (PM2.5, NO2 and O3) and ESCAPE monitoring data (BC), and incorporated satellite observations, dispersion model estimates, land use and traffic

data. Kriging was performed on the residual spatial variation from the LUR models and added to the exposure estimates. One model was developed using all sites (100%). Robustness of the models was evaluated by performing a five-fold hold-out validation and for PM2.5 and NO2 additionally with independent comparison at ESCAPE measurements. To evaluate the stability of each model's spatial structure over time, separate models w2.1ere developed for different years (NO2 and O3: 2000 and 2005; PM2.5: 2013). Results: The PM2.5, BC, NO2, O3 annual, O3 warm season and O3 cold season models explained respectively 72%, 54%, 59%, 65%, 69% and 83% of spatial variation in the measured concentrations. Kriging proved an efficient technique to explain a part of residual spatial variation for the pollutants with a strong regional component explaining respectively 10%, 24% and 16% of the R2 in the PM2.5, O3 warm and O3 cold models. Explained variance at fully independent sites vs the internal hold-out validation was slightly lower for PM2.5 (65% vs 66%) and lower for NO2 (49% vs 57%). Predictions from the 2010 model correlated highly with models developed in other years at the overall European scale.

2.1.3 RUN – UFP National Model

Reference: Kerckhoffs, J; Hoek, G; Gehring, U; Vermeulen, R.C.H. Modelling Nationwide Spatial Variation of Ultrafine Particles based on Mobile Monitoring. Environment International. 2021, 154 (9).https://doi.org/10.1016/j.envint.2021.106569

This map shows the spatial variability of UFP in the entire Netherlands. It combined a regional background monitoring campaign with 3 times 2-week measurements at 20 different sites across the Netherlands (REF) with a dedicated mobile monitoring campaign in several cities and villages in the Netherlands.

2.1.4 Google Air View – Amsterdam Model

Reference: Kerckhoffs, J; Jibran Khan, Gerard Hoek, Zhendong Yuan, Ole Hertel, Matthias Ketzel, Steen Solvang Jensen, Fares Al Hasan, Kees Meliefste, Roel Vermeulen, Hyperlocal variation of nitrogen dioxide, black carbon, and ultrafine particles measured with Google Street View cars in Amsterdam and Copenhagen, Environment International, Volume 170, 2022, 107575
https://doi.org/10.1016/j.envint.2022.107575

Maps were developed using Google Street-View cars equipped with air quality instruments. Every street in Amsterdam was driven 7 times on average and all streets were used in mixed-model approach to develop hyperlocal maps of the city for UFP, NO2 and BC. This means all street segments were used to develop a LUR model but can those street segments can also alter the prediction of the LUR model based on the

within and between-segment variation. Maps are also available via the Google Insights website: https://insights.sustainability.google/labs/airquality

2.1.5 Particle Elemental Composition

We developed Europe-wide models of long-term exposure to eight elements (copper, iron, potassium, nickel, sulfur, silicon, vanadium, and zinc) in particulate matter with diameter <2.5 μm (PM2.5) using standardized measurements for one-year periods between October 2008 and April 2011 in 19 study areas across Europe, with supervised linear regression (SLR) and random forest (RF) algorithms. Potential predictor variables were obtained from satellites, chemical transport models, land-use, traffic, and industrial point source databases to represent different sources. Overall model performance across Europe was moderate to good for all elements with hold-out-validation R-squared ranging from 0.41 to 0.90. RF consistently outperformed SLR. Models explained within-area variation much less than the overall variation, with similar performance for RF and SLR. Maps proved a useful additional model evaluation tool. Models differed substantially between elements regarding major predictor variables, broadly reflecting known sources. Agreement between the two algorithm predictions was generally high at the overall European level and varied substantially at the national level. Applying the two models in epidemiological studies could lead to different associations with health. If both between- and within-area exposure variability are exploited, RF may be preferred. If only within-area variability is used, both methods should be interpreted equally.

Citation: Chen J, de Hoogh K, Gulliver J, Hoffmann B, Hertel O, Ketzel M, Weinmayr G, Bauwelinck M, van Donkelaar A, Hvidtfeldt UA, Atkinson R, Janssen NAH, Martin RV, Samoli E, Andersen ZJ, Oftedal BM, Stafoggia M, Bellander T, Strak M, Wolf K, Vienneau D, Brunekreef B, Hoek G. Development of Europe-Wide Models for Particle Elemental Composition Using Supervised Linear Regression and Random Forest. Environ Sci Technol. 2020 Dec 15;54(24):15698-15709. doi: 10.1021/acs.est.0c06595. Epub 2020 Nov 25. PMID: 33237771; PMCID: PMC7745532.

2.2 Biodiversity

2.2.1 Species diversity

Species diversity in the Netherlands
Type: TIF

Description: The map shows the estimated species diversity per kilometer loft in classes for the whole of the Netherlands, based on data from the National Database Flora and Fauna for the last 10 years.
The eight most important and best-researched species groups are combined here. Corrections have been made in the map for the effect that the observation is not equally intense at every location.

Temporal resolution: 2017
spatial resolution: 1000X1000 m

2= '101 - 200 soorten',
3='201 - 300 soorten',
4='301 - 400 soorten',
5= '401 - 500 soorten',
8='> 700 soorten'.

Use restrictionsAtlas including: use restrictions: the map shows relative and not absolute species numbers in classes, distinguishing relatively species-rich areas from relatively species-poor areas on the basis of eight important species groups.
Date of the source (created)15/11/2017
Technical information
Source Identification22833e30-8815-4b41-9af1-2a6b99b75130
Code reference system28992
General description of provenanceAtlas including: restrictions on use: the map shows relative and not absolute numbers of species in classes, distinguishing relatively species-rich areas from relatively species-poor areas on the basis of eight important species groups. The information for species distribution and research coverage is taken from and generated with data from the National Flora and Fauna Database (NDFF). . Other restrictions: none.

Link:
https://nationaalgeoregister.nl/geonetwork/srv/dut/catalog.search#/metadata/22833e30-8815-4b41-9af1-2a6b99b75130

2.3 Light at Night

2.3.1 Light intensity at Night from EXPANSE

Light-at-night (LAN) have been assessed using harmonized data produced with DMSP (Defense Meteorological Satellite Program) and VIIRS (Version 1 VIIRS Day/Night Band Night-time Lights) data sources for the period 1992-2020 at annual level (Figure 2).

VIIRS (Elvidge et al., 2017; Mills et al., 2013) data is available at monthly or annual levels since 2012 and it is composed by average radiance values and numbers of available observations. Data is provided at 15 arc resolution (around 0.7x0.7 km) by the Earth Observations Group (EOG) (*Earth Observation Group - Defense Meteorological Satellite Progam, Boulder | Ngdc.Noaa.Gov*, n.d.).

Version 4 of the DMSP-OLS Night-time Lights Time Series consists of cloud-free composites made using all the available archived DMSP-OLS fine/smooth resolution data (1km / 2.7km) for calendar years. In cases where two satellites were collecting data, two composites were produced.

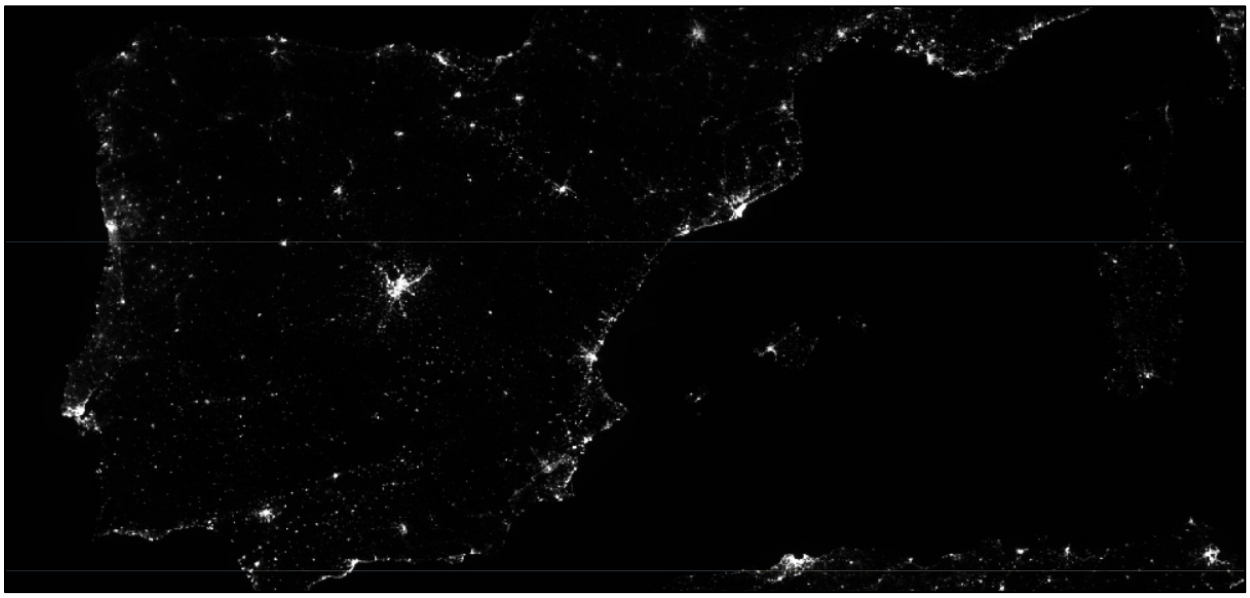

Figure 2. Monthly VIIRS Night Band (DNB) in 2019 in Spanish region.

The night-time light intensity in each cell is represented by the "digital number", an integer from 0 to 63, that it is "*not exactly proportional to the physical amount of light received (called true radiance*)," quoted from p. 999 of Henderson et al. (2012). After identifying inconsistencies between DMSP-OLS and VIIRS (Zhao et al., 2019) a long time series of consistent and comparable night-time light dataset was produced from 1992 to present (Li et al., 2020). Data for the entire world is freely accessible (*Harmonization of DMSP and VIIRS Nighttime Light Data from 1992-2018 at the Global Scale*, n.d.) for annual levels with a spatial resolution of 30 arc-seconds (~1km). Each data set consists of temporally calibrated DMSP-OLS NTL time-series data from 1992-2013 and converted NTL time series from the VIIRS data (2014-2018) (Li & Zhou, 2017).

For EXPANSE, LAN exposures have been calculated within 300m, 500m and 1000m at 5 different time-windows (2000 to 2020) using harmonized datasets. As a result, we have delivered fifteen raster datasets (5 time-windows and 3 area-level) with the amount of radiation (as median or mean, also possible to derive other statistics) per each 100m cell (see table S4 in Annex). To do that, a focal statistics tool (see Methodology section) have been applied to get the averaged radiance values.

2.3.2 Light at Night VNL V2

Description:
Annual VNL V2

A new consistently processed time series of annual global VIIRS nighttime lights has been produced from monthly cloud-free average radiance grids spanning 2012* to 2020. The new methodology is a modification of the original method based on nightly data (Annual VNL V1).

* For 2012 annual VNL V2, there are two sets. (A) 201204-201212, and (B) 201204-201303. Only set (B) has masked median and average, as well as lit area mask.

In both methods there is an initial filtering to remove sunlit, moonlit and cloudy pixels, leading to rough composites that contains lights, fires, aurora and background. In the original method, the rough annual composites are made from a full year of nightly DNB data. In the new method, the rough composites are made on monthly increments and then combined to form rough annual composites. Both methods employ outlier removal to discard biomass burning pixels and isolate the background.

In the original method the outlier removal is performed on scattergrams generated for each 15 arc second grid cell, with outliers clipped off from both the high and low radiance sides of the scattergram. The discard of outlier pixels proceeds until the scattergram's standard deviation stabilizes. The new method uses the twelve-month median radiance to discard high and low radiance outliers, filtering out most fires and isolating the background. Background areas are zeroed out in both methods using the data range (DR) calculated from 3x3 grid cells. In both methods, the DR threshold for background is indexed to cloud-cover levels, with higher DR thresholds in areas having low numbers of cloud-free coverages. In the new method, particular attention is given to setting a single DR threshold for distinguishing lit grid cells from background for each 15 arc second grid cell. This is achieved by setting the DR threshold from a multiyear maximum median and a corresponding multiyear percent cloud-cover grids. The multiyear approach makes it possible to detect lighting present in each 15 arc second grid cell with a single DR threshold across all the years in the series.

Delivery File Tyle: gz (gzipped)
Delivery File Content : average-masked
Image File Type : GeoTIFF
Unit: nW/cm2/sr
CRS: EPSG:4326 (Geographic Latitude/Longitude)
Resolution: 15 arc second (~500m at the Equator)

Coverage : 180W, 75N, 180E, 65S
Temporal coverage: 2012 to 2021

Link: https://eogdata.mines.edu/products/vnl/

2.4 Noise

2.4.1 Nationwide noise level

The file rivm_20210201_g_soundmap_lden_alle_sources_v3.tif is a raster file for use in GIS software, and contains nationwide the noise level in 1 dB classes. The resolution is 10 x 10 m.

The noise classes relate to the cumulative noise exposure in Lden (year) as caused by
- national highways (2017)
- municipal and provincial roads (2017)
- rail traffic (2016)
- aviation (2016)
- industry (key estimate)
- wind turbines (2020)

The card only gives a rough indication of the sound quality and is not suitable to be able to test noise levels in the context of the standards from the Noise Abatement Act to 'the noise load on the facade'. The map does, however, offer a good picture of the sound quality on a more global scale of focus areas such as residential areas, nature and quiet areas.

The contribution to the cumulative map by road and rail traffic is determined using a standard calculation method (RMV2012).
This method uses digital files that include the location of roads and railways.
The traffic data of national highways and railways are based on data for the benefit of Directive 2002/49/EC for 2016 from Rijkswaterstaat and Prorail(1).
The indication of the noise load from provincial and municipal roads is based on traffic estimates from 2011 (www.emissieregistratie.nl).
The noise impact from wind turbines is based on 2015 data from wind energy news(2).
For industrial noise, an (indicative) sound quality has been estimated on the basis of key figures for various types of industrial sites.
The noise pollution from aviation is based on noise maps drawn up by the National Aerospace Laboratory (2011).
The cumulative map is a combined total of the sources mentioned and gives an indicative picture of the sound quality. No rights can be derived.
Link:
https://www.nationaalgeoregister.nl/geonetwork/srv/api/records/68711fca-7589-4b83-829c-42550803c287

2.4.2 Traffic noise

Three Geotiff files for the Netherlands with trafficnoise from 2011 ,2017 and 2020.
These maps have generally an update-frequency of every 5 years.
Cell Size X: 10 m

Cell Size Y: 10 m
Coordinate System: RD New
The units are cB, which means that you need to divide the values by 10 in order to get dB values.
It is a bit tricky to compare between two years, because the model has been altered during the years.
Source of the data is RIVM (Harm van Wijnen : harm.van.wijnen@rivm.nl)

2.5 Pesticides

2.6 Temperature

2.7 Weather_micro

2.7.1 Heat islands

Title of the map: Stedelijk_hitte_eiland_effect_01062022_v2

Geographical coverage: The Netherlands

Description / abstract / note:
This map shows the temperature difference between the city and the countryside in the Netherlands.
This is also known as the urban heat island effect or urban heat island effect (UHI).
The map shows the average temperature difference in degrees Celsius (°C).

Temporal coverage: 01062022

Additional information:

Source: The maker is the National Institute for Public Health and the Environment (RIVM). View exactly how the map was made in the National Georegister.

reslution: 10 m
File type:Geotiff
URL: https://www.atlasleefomgeving.nl/kaarten

2.7.2 Local climate zones

This EU LCZ map readme.txt file was generated on 2020-12-02 by Matthias Demuzere

Versions:

v0.0 | 2019-04-24 | Initial release of the raw data
v1.0 | 2020-12-03 | Gaussian filtered version + Q/Arc-GIS packages and color information.
v2.0 | 2021-06-09 | Adding LCZ map in EPSG 4326 + fixing some typos.

v3.0 | 2022-05-10 | Fixing LAEA projection, removing epsg4236 projection file.
v4.0 | 2022-05-10 | Properly update readme

GENERAL INFORMATION

Title of Dataset: European LCZ map

Principal Authors Information:

Principal Investigator: Matthias Demuzere (matthias.demuzere ~at~ rub.de)
Alternate Contact: Benjamin Bechtel (benjamin.bechtel ~at~ rub.de)

Summary:
A European Local Climate Zone map at a 100 m spatial resolution,
derived from multiple Earth Observation datasets and expert LCZ class labels.
There are 10 urban LCZ types, each associated with a set of relevant variables such that the map
represent a valuable database of urban properties.

=> Please read the paper carefully for more details on methods, underlying datasets, quality control,
accuracy, its relevance and usage notes.

IMPORTANT: The .geotif version in this archive has been filtered using the Gaussian filter procedure
as elaborated in Demuzere et al. (2020).

SHARING/ACCESS INFORMATION

Recommended citations for the data:

- Demuzere, M., Bechtel, B., Middel, A., & Mills, G. (2019). Mapping Europe into local climate zones. PLOS ONE, 14(4),
e0214474. https://doi.org/10.1371/journal.pone.0214474
- Demuzere, M., Bechtel, B., Middel, A., & Mills, G. (2020). European LCZ map.
10.6084/m9.figshare.13322450

Other Notes:

to use these data without restriction, subject to the following:

- Understand that these data are created by crowd sourcing and machine learning and thus will naturally contain some errors.

It comes as it is without any warranty.

- The above copyright notice and this permission notice shall be included in all copies or substantial portions of the data.
- The authors (or generically WUDAPT contributors) are acknowledged.
- For scientific use, the following papers could be cited to introduce the LCZ concepts:

* Stewart, I. D., & Oke, T. R. (2012). Local Climate Zones for Urban Temperature Studies. Bulletin of the American Meteorological Society, 93(12), 1879–1900. https://doi.org/10.1175/BAMS-D-11-00019.1

* Bechtel B, Daneke C (2012) Classification of Local Climate Zones based on multiple Earth Observation data. IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing 5:1191-1202

* Bechtel B, Alexander PJ, Böhner J, Ching J, Conrad O, Feddema J, Mills G, See L, Stewart I (2015) Mapping Local Climate Zones for a Worldwide Database of the Form and Function of Cities. ISPRS International Journal of Geo-Information 4:199-219

* Ching, J., Mills, G., Bechtel, B., See, L., Feddema, J., Wang, X., … Theeuwes, N. (2018). WUDAPT: An Urban Weather, Climate, and Environmental Modeling Infrastructure for the Anthropocene. Bulletin of the American Meteorological Society, 99(9), 1907–1924. https://doi.org/10.1175/BAMS-D-16-0236.1

* Bechtel, B., Alexander, P. J., Beck, C., Böhner, J., Brousse, O., Ching, J., … Xu, Y. (2019). Generating WUDAPT Level 0 data – Current status of production and evaluation. Urban Climate, 27, 24–45. https://doi.org/10.1016/j.uclim.2018.10.001

* Demuzere, M., Bechtel, B., & Mills, G. (2019). Global transferability of local climate zone models. Urban Climate, 27, 46–63. https://doi.org/10.1016/j.uclim.2018.11.001

* Demuzere, M., Hankey, S., Mills, G., Zhang, W., Lu, T., & Bechtel, B. (2020). Combining expert and crowd-sourced training data to map urban form and functions for the continental US. Scientific Data, 7(1), 264. https://doi.org/10.1038/s41597-020-00605-z

* Demuzere M, Kittner J, Bechtel B. LCZ Generator: A Web Application to Create Local Climate Zone Maps. Front Environ Sci. 2021;9. doi:10.3389/fenvs.2021.637455

DATA & FILE OVERVIEW

Files and brief description:

1. EU_LCZ_map.tif

This is a compressed geotif of the European LCZ map, with classes indicated by numbers 1-17.

The table below describes the class number | official class description | official hex color.

LCZ 1	Compact highrise	'#910613'
LCZ 2	Compact midrise	'#D9081C'
LCZ 3	Compact lowrise	'#FF0A22'
LCZ 4	Open highrise	'#C54F1E'
LCZ 5	Open midrise	'#FF6628'
LCZ 6	Open lowrise	'#FF985E'
LCZ 7	Lightweight low-rise	'#FDED3F'
LCZ 8	Large lowrise	'#BBBBBB'
LCZ 9	Sparsely built	'#FFCBAB'
LCZ 10	Heavy Industry	'#565656'
LCZ 11 (A)	Dense trees	'#006A18'
LCZ 12 (B)	Scattered trees	'#00A926'
LCZ 13 (C)	Bush, scrub	'#628432'
LCZ 14 (D)	Low plants	'#B5DA7F'
LCZ 15 (E)	Bare rock or paved	'#000000'
LCZ 16 (F)	Bare soil or sand	'#FCF7B1'
LCZ 17 (G)	Water	'#656BFA'

Other characteristics:

- Projection: ETRS89 / ETRS-LAEA (EPSG:3035) (matching the projection used by the European Environmental Agency - https://www.eea.europa.eu/data-and-maps/data/eea-reference-grids-1/about-the-eea-reference-grid/eea_reference_grid_v1.pdf/at_download/file)
- Spatial Resolution: 100 m
- Representative for 2016
- Overall average accuracy (%): 80 (see Fig. 6 in Demuzere et al. (2019) for more info)

2. cmap_Qgis_WUDAPT_LCZ.txt

A QGIS (https://qgis.org/) generated LCZ Color Map Export File to display the geotif in the corresponding LCZ colors.

3. EU_LCZ_map.qgz

A QGIS (https://qgis.org/) package file that displays the LCZ map together with the appropriate colors.

When opening the project file, please adjust the filepath pointing to the LCZ map .tif file.

Tested on QGIS > 3.12

4. EU_LCZ_map.mpk

An ArcGIS (https://www.arcgis.com/) package file that displays the LCZ map together with the appropriate colors.
Just drag and drop the file into ArcMap's map window to open it.

3. Built environment

Considering these issues at the European level, we have integrated the built environment characterization into a common raster of 100m resolution across Europe (except for NDVI), defined within the EXPANSE project framework. For each raster cell, we have calculated the required exposures at the different buffer sizes of 300, 500 and 1000m to better characterize the surrounding of people's residential addresses. NDVI exposure have been delivered at 250 m cell resolution to preserve the spatial resolution of the data source (see table 1). To reduce the data storage and computation for partners, the final datasets have been cropped and provided per each country in Europe.

Table 1: Summary for the built-up datasets created in EXPANSE.

Variable	Source	Data type	Year	Resolution	indicators
NDVI	MODIS	Raster	2000, 2005, 2010, 2015, 2020	250m	Surrounding greenness within 300, 500 and 1000m
MSAVI2	MODIS	Raster	2000, 2005, 2010, 2015, 2020	250m	Surrounding greenness within 300, 500 and 1000m
Blue space	EU-Hydro	Raster	2011-2013	100	Distance to nearest blue space

Imperviousness	ESM	Raster	2006. 2009. 2012. 2015. 2018	100m	Percentage of imperviousness within 300, 500 and 1000m
Light-at-night (LAN)	VIIRS / DMSP-OLS	Raster	2000, 2005, 2010, 2015, 2020	1 km	Average radiation

3.1 Green spaces

Prior research has evidenced positive associations between greenness and physical and mental health outcomes. While surrounding greenness is the commonly used metric, quantification of green spaces and accessibility using land cover datasets have been less popular (Fong et al., 2018).

As a general guideline, WHO has recommended that urban residents should be able to access public green spaces of at least 0.5–1 hectare located within 300 metres' linear distance (around 5 minutes' walk) of their homes (WHO, 2017).

In EXPANSE, we have characterized green spaces using two different indicators: i) surrounding greenness using satellite-based indices including normalized difference vegetation index (NDVI) and modified soil-adjusted vegetation index (mSAVI) and ii) accessibility to green spaces using land use/land cover GIS datasets.

3.1.1 Residential surrounding greenness

Residential surrounding greenness have been assessed using satellite-derived NDVI and mSAVI imagery data.

Normalized Difference Vegetation Index (NDVI)

NDVI quantifies photosynthetically active vegetation by measuring the difference between near-infrared (which vegetation strongly reflects) and red light (which vegetation absorbs). NDVI values range from +1.0 to -1.0. Areas of barren rock, sand, or snow usually show very low NDVI values (for example, 0.1 or less). Sparse vegetation such as shrubs and grasslands or senescing crops may result in moderate NDVI values (approximately 0.2 to 0.5). High NDVI values (approximately 0.6 to 0.9) correspond to dense vegetation such as that found in temperate and tropical forests or crops at their peak growth stage. Negative values of NDVI (values approaching -1) correspond mainly to water and ice and snow cover.

In EXPANSE, NDVI have been derived from the Vegetation Indices (MOD13Q1) product of the Terra Moderate Resolution Imaging Spectroradiometer (MODIS) with 250 m x 250 m resolution (K. Didan, 2015). MOD13Q1 16-day composite vegetation indices provide one NDVI value every 16 days allowing the composition of gap-free temporal-series. The per-pixel compositing algorithm is described in the MOD13Q1 product user guide (Kamel Didan et al., 2015). The MOD13Q1 product includes the following relevant bands, among others: a) NDVI and b) quality assessment (QA) information. More information could be found elsewhere (https://lpdaac.usgs.gov/products/mod13q1v006/). Per-pixel Quality Assessment (QA) band provides information about how reliable the acquired data is and allows for removal of pixels with less accuracies. QA values of 0 and 1 mean a good quality level, whereas values of 2 and 3 represent pixels covered by ice, snow or clouds, which are discarded. Water features have typically negative signal values for NDVI, so they add noise to the negative edge when conducting spatial analysis. In order to identify and remove the influence of water bodies in the analysis, the Land Water Mask derived from MODIS and SRTM (MOD44W) have been used (Carroll et al., 2017) to filter out the water bodies from the NDVI maps. This mask is a well-described satellite imagery with the same spatial resolution (250m) provided by the U.S. Geological Survey (USGS). MOD44W is a raster binary mask with land pixels represented as 0 and water as 1.

For EXPANSE, surrounding greenness for the last 20 years have been calculated using 5-year time points (2000, 2005, 2010, 2015 and 2020) at different buffers (300, 500, 1000 m). The following statistics were calculated for the pilot stage: min, p10, p25, p50, mean, p75, p90, max, and sd. Based on this experience, the final datasets were derived for the mean, median and sd for each buffer and temporal window. As a result, a total of 45 raster datasets with 250m resolution for NDVI exposure have been delivered. Each of the bands represents a selected statistic (see table S4 in Annex). In brief, the data processing strategy includes the following steps:

- All MOD13Q1 imagery within a year have been loaded and processed using Google Earth Engine. A maximum of 23 dates are available each year.
- Processed imagery have been clipped to the defined study area within the project.
- Quality assessment (QA) information have been applied to obtain good quality data.
- Water bodies have been masked out, using MOD44W. See an example in figure 1.
- Selected statistics (explored in the pilot study) have been calculated from the previous processed image collection to result in a multi-band raster dataset.

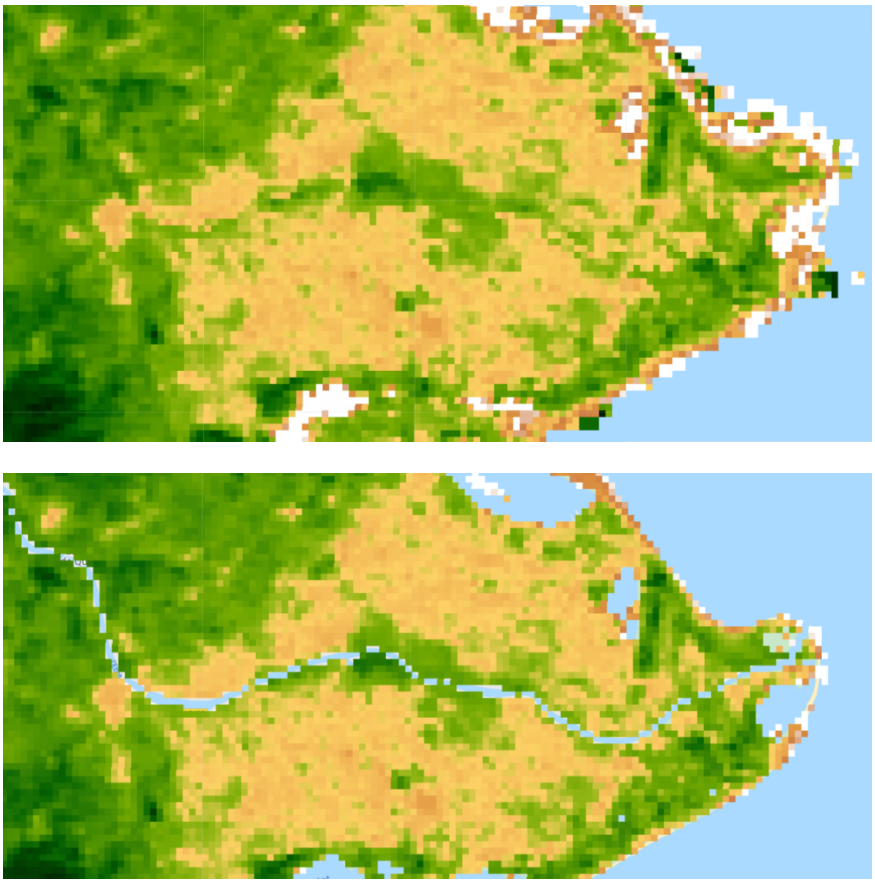

Figure 1. Delta de l'Ebre (Catalunya, Spain). A) Original NDVI MODIS raster and B) NDVI MODIS raster with water mask applied.

Modified Soil Adjusted Vegetation Index (MSAVI)

MSAVI and its later revision, MSAVI2 (MSAVI2 is generally the version that is used when people refer to MSAVI), are soil adjusted vegetation indices that seek to address some of the limitation of NDVI when applied to areas with a high degree of exposed soil surface. Compared to NDVI, MSAVI has the advantage of minimizing the effect of background soil. Qi et al. (1994a) developed the MSAVI, and later the MSAVI2 (Qi et al. 1994b) to more reliably and simply calculate a soil brightness correction factor. The output of MSAVI or MSAVI2 is a new image layer representing vegetation greenness with values ranging from -1 to +1. One significant limitation of the MSAVI is that it sacrifices some overall sensitivity to changes in vegetation amount/cover to correct for the soil surface brightness. Hence, MSAVI may not be as sensitive to vegetation change as other indices such as NDVI. MSAVI would also be more sensitive to differences in atmospheric conditions between areas or times (*Modified Soil-Adjusted Vegetation Index*, n.d.). MSAVI2 requires a red and a near infrared band to be calculated, following the next equation:

$$MSAVI2 = \frac{\left(2 * NIR + 1 - \sqrt{(2 * NIR + 1)^2 - 8 * (NIR - RED)}\right)}{2}$$

For EXPANSE, MSAVI2 have been derived using MOD13Q1 product, following the same procedure described in the previous section on NDVI. Instead of using NDVI band within the product, RED and NIR bands have been used to calculate the MSAVI following the previous equation.

As a result, 45 raster datasets have been created (5 time-point,3 distance scales and 3 statistics) including the same statistic parameters as selected for NDVI (mean, median and sd) (see table S4 in Annex).

3.1.2 Accessibility to green spaces

Accessibility to green spaces have been characterized as the distance to the nearest green space. To characterize the accessibility to green spaces, we have used Corine Land Cover (CLC) and Urban Atlas (Copernicus, see table 2). Based on standardized methods and sources, comparable land categories have been selected to define green spaces (see Tables S1-S3 in the Annex).

The scale of CORINE LC is 1:100.000 and the minimum mapping unit is 25 ha. The scale of Urban atlas is 1:10.000 and the minimum mapping unit is 0.25 ha for the artificial surfaces and 1 ha for the other surfaces. While Urban Atlas provides a better spatial resolution, it is restricted to urban agglomeration whereas CLC covers the entire Europe. In addition, CLC has a greater temporal extent (since 1990) than Urban Atlas (starting in 2006). Both land products have experienced improvements along time. To benefit from both data sources CLC and Urban Atlas products have been separately produced.

Seven raster datasets with the averaged distance to nearest green space have been derived. Four-time windows for CLC and three time windows for Urban Atlas characterized (i.e., 2000 for CLC and 2006, 2012 and 2018 for both), (see table S4 in Annex).

Distance to nearest green space

Straight (Euclidean) distances from the center of each 100m grid cell to the nearest green space have been calculated according to the relevant land use/cover categories (see Tables S1-S3 from Annex).

As a result, four 100m raster datasets (one for each time window) have been produced for each data source (i.e., CLC and Urban Atlas) as the mean distance to a nearest green spaces (see table 4 in Annex).

To this end, we have used the distance raster analysis (see Methodology section), which measures the linear distance from each cell in the raster to the closest green space feature (Figure 4). This have provided a convenient approach since the distance is calculated from the center of the grid cell, instead of the border, minimizing the effect of people living in one border of the grid cell but the shortest distance to the nearest green space being calculated from the opposite edge.

Table 2: Data description on land uses/land cover products from CLC and Urban Atlas.

data source	year	scale (Ha)	nr urban agglomerations	extension	categories used as green spaces
Urban Atlas	2006	0.25 / 1	319	EU28	14100, 14200, 20000, 30000
Urban Atlas	2012	0.25 / 1	785	EEA39	14100, 14200, 21000, 22000, 23000, 24000, 31000, 32000
Urban Atlas	2018	0.25 / 1	785	EEA39	14100, 14200, 21000, 22000, 23000, 24000, 31000, 32000
CORINE	2000	25	all	EEA39	141, 142, 211, 212, 213, 221, 222, 223, 231, 241, 242, 243, 244, 311, 312, 313, 321, 322, 323, 324
CORINE	2006	25	all	EEA39	141, 142, 211, 212, 213, 221, 222, 223, 231, 241, 242, 243, 244, 311, 312, 313, 321, 322, 323, 324
CORINE	2012	25	all	EEA39	141, 142, 211, 212, 213, 221, 222, 223, 231, 241, 242, 243, 244, 311, 312, 313, 321, 322, 323, 324
CORINE	2018	25	all	EEA39	141, 142, 211, 212, 213, 221, 222, 223, 231, 241, 242, 243, 244, 311, 312, 313, 321, 322, 323, 324

3.2 Blue spaces

Distance to nearest blue space

Due to the complexity in water bodies features (i.e. narrow linear features such as rivers, canals, coastal path, and beaches) distance metrics are often preferred to coverage metrics (e.g. (surface areas) in research concerning blue spaces (Elliott et al., 2020).

Exposure to blue spaces have been assessed using the EU-Hydro map developed by the CLMS (*Copernicus Land Monitoring Service*, 2019). This baseline map (2011-2013) for all EEA39 countries provides high-resolution data on the river network, water bodies (e.g., lakes and wide rivers), drainage network with catchment areas, drainage lines and sea/ocean water.

We have calculated the Euclidean distance from the center of each cell to the nearest blue space separately for sea or ocean and inland freshwater features, and use it as a global indicator for all types of water bodies.

As a result, we have delivered three 100m raster datasets, representing the entire period assessed, based on the temporal stability of large water bodies along the time (see table S4 in Annex). As explained for green spaces indicators, we have used distance raster analysis (see Methodology section).

3.3 Grey spaces

Grey (i.e. built up) spaces have been characterized using imperviousness density (IMD) maps (*Copernicus Land Monitoring Service*, 2020). IMD maps are high-resolution maps (100m) that represent the percentage of soil sealing per area unit. Sealed/Impervious areas are characterized by the substitution of the original (semi-) natural land cover or water surface with an artificial, often impervious cover. These artificial surfaces are usually maintained over long periods of time. IMD status maps are available at different reference years (2006, 2009, 2012, 2015 and 2018).

IMD within 300m, 500m and 1000m have been calculated as an indicator of surrounding built-up areas. As a result, we have delivered fifteen raster datasets (5 time-windows and 3 area-level) with the degree of imperviousness (%) per each 100m cell (see table S4 in Annex). To do that, we have applied a focal statistics tool (see Methodology section) to get the average built-up area in each defined scenario (year/scale)

4. Social environment

4.1.1 Road accidents

Title:
Road accidents
Discription:
Total number of traffic accidents during 2003 to 2012 and 2013 to 2021 in the Netherlands The information is based on data from Rijkswaterstaat (http://www.rijkswaterstaat.nl/apps/geoservices/geodata/dmc/bron/).

Temporal coverage:
2003 -2012
2013 -2021
Spatial resolution:
62.5*62.5 m

Trafic accidents on highway was removed.

4.1.2 Leefbaarometer

Description: With the help of the Leefbaarometer, the quality of life in all inhabited districts, neighborhoods and streets in the Netherlands can be monitored.

The Leefbaarometer shows how the quality of life is and how it has developed in recent years.

Geographical extent: The netherlands

Temporal extent: 2002-2008-2012-2014-2016-2018-2020

Spatial resultion : municipality, district, neighbourhood,Post code(PC4), Grid 100*100 meter.

Link: https://www.leefbaarometer.nl/kaart/#kaart

https://www.leefbaarometer.nl/page/FAQ#grid

5. Background information

5.1 Methodology

In this section we provide a detailed information about the GIS tools and software used to calculate the previous built environment characterization The following tools and functions are described below:

5.1.1 Focal statistics

The Focal Statistics tool performs a neighbourhood operation that computes an output raster, where the value for each output cell is a function of the values of all the input cells that are in a specified neighbourhood around that location.

For each cell in the raster, the tool allows to calculate the specified statistic (i.e. min, p10, p25, p50, mean, p75, p90, max, and sd) for the defined neighbourhood (i.e. 300m, 500m, 1000m buffer). The neighbourhood could be determined with user-defined shape (i.e. donut, square, circle, irregular) and size (width or diameter) as number of cells. Neighbourhoods can overlap so that cells in one neighbourhood may also be included in a neighbourhood of another processing cell. Figure 3 demonstrates how "sum" focal statistics is applied. Raster cell processing (value = 5) turns into 24 after the operation. Nine cells have been included into the analysis (3x3). Further information canbe found elsewhere (*ArcGIS for Desktop*, n.d.)

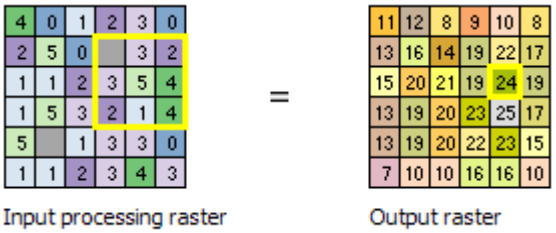

Figure 3. Focal statistics example.

5.1.2 Distance in raster analysis

The Euclidean distance tools in raster analysis describe each cell's relationship to a source or a set of sources based on the straight-line distance. Euclidean distance is calculated from the centre of the source cell to the centre of each of the surrounding cells. Conceptually, the Euclidean algorithm works as follows: for each cell, the distance to each source cell is determined by calculating the hypotenuse with x_max and y_max as the other two legs of the triangle. This calculation derives the

true Euclidean distance, rather than the cell distance. The shortest distance to a source is determined, and if it is less than the specified maximum distance, the value is assigned to the cell location on the output raster.

To this purpose, an open-source library for Python will be used (GDAL/OGR contributors, 2021)

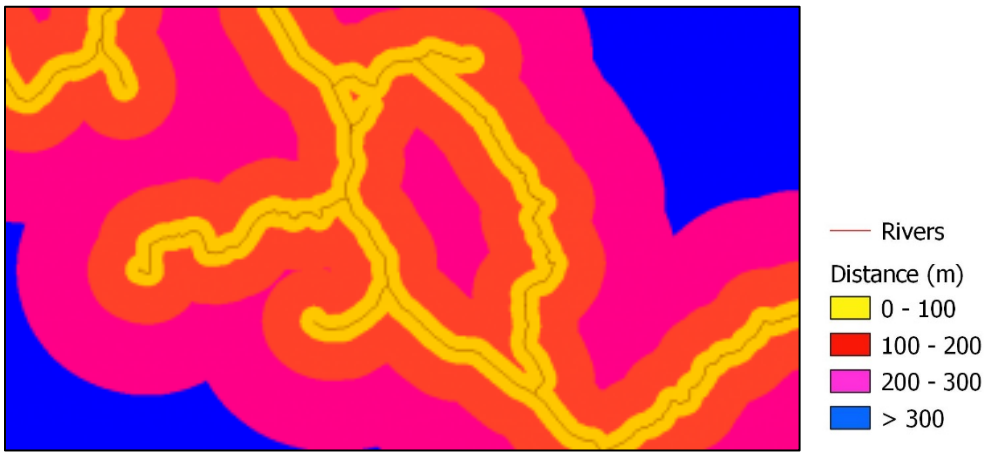

Figure 4. An example for a distance raster analysis between a raster grid cells and a linear feature.

5.2 Software

We present the potential software and programmes that we will be applied to carry out the spatial analysis described in this protocol.

5.2.1 Python

Python is an interpreted object-oriented programming language. Detailed information can be found everywhere (Python Software Foundation, n.d.)

5.2.2 Google Earth Engine

Google Earth Engine platform (Gorelick et al., 2017) will be used to undertake the processes involving huge amount of satellite imagery (i.e., NDVI and MSAVI2). Earth Engine is a platform for searching, visualizing and analysing geospatial datasets, for academic, non-profit, business and government users. Earth Engine hosts satellite imagery and stores it in a public data archive that includes historical earth images going back more than forty years.

5.2.3 QGIS

QGIS is a free and open-source GIS software that supports viewing, editing, and analysis of geospatial data (*QGIS Geographic Information System.*, 2009).

6. References

de Hoogh, K., Chen, J., Gulliver, J., Hoffmann, B., Hertel, O., Ketzel, M., . . . Hoek, G. (2018). Spatial PM2.5, NO2, O3 and BC models for Western Europe – Evaluation of spatiotemporal stability. *Environment International, 120*, 81-92. doi:https://doi.org/10.1016/j.envint.2018.07.036

Shen, Y., de Hoogh, K., Schmitz, O., Clinton, N., Tuxen-Bettman, K., Brandt, J., . . . Hoek, G. (2022). Europe-wide air pollution modeling from 2000 to 2019 using geographically weighted regression. *Environment International, 168*, 107485. doi:https://doi.org/10.1016/j.envint.2022.107485

ArcGIS for Desktop. (n.d.). How Focal Statistics Works. Retrieved September 18, 2020, from https://desktop.arcgis.com/en/arcmap/10.3/tools/spatial-analyst-toolbox/how-focal-statistics-works.htm

Carroll, M. L., DiMiceli, C. M., Townshend, J. R. G., Sohlberg, R. A., Elders, A. I., Devadiga, S., Sayer, A. M., & Levy, R. C. (2017). Development of an operational land water mask for MODIS Collection 6, and influence on downstream data products. *International Journal of Digital Earth*, *10*(2), 207–218. https://doi.org/10.1080/17538947.2016.1232756

Didan, K. (2015). MOD13Q1 MODIS/Terra Vegetation Indices 16-Day L3 Global 250m SIN Grid V006. NASA EOSDIS Land Processes DAAC. In *USGS* (Vol. 5, pp. 2002–2015). https://doi.org/10.5067/MODIS

Didan, Kamel, Barreto Munoz, A., Solano, R., & Huete, A. (2015). *MODIS Vegetation Index User's Guide (MOD13 Series)*. http://vip.arizona.edu

Earth Observation Group - Defense Meteorological Satellite Progam, Boulder | ngdc.noaa.gov. (n.d.). Retrieved September 21, 2020, from https://ngdc.noaa.gov/eog/dmsp/downloadV4composites.html

Elliott, L. R., White, M. P., Grellier, J., Garrett, J. K., Cirach, M., Wheeler, B. W., Bratman, G. N., van den Bosch, M. A., Ojala, A., Roiko, A., Lima, M. L., O'Connor, A., Gascon, M., Nieuwenhuijsen, M., & Fleming, L. E. (2020). Research Note: Residential distance and recreational visits to coastal and inland blue spaces in eighteen countries. *Landscape and Urban Planning*, *198*, 103800. https://doi.org/10.1016/j.landurbplan.2020.103800

Elvidge, C. D., Baugh, K., Zhizhin, M., Hsu, F. C., & Ghosh, T. (2017). VIIRS night-time lights. *International Journal of Remote Sensing*, *38*(21), 5860–5879. https://doi.org/10.1080/01431161.2017.1342050

EU-Hydro — Copernicus Land Monitoring Service. (2019). https://land.copernicus.eu/imagery-in-situ/eu-hydro

Fong, K. C., Hart, J. E., & James, P. (2018). A Review of Epidemiologic Studies on Greenness and Health: Updated Literature Through 2017. In *Current environmental health reports* (Vol. 5, Issue 1, pp. 77–87). Springer. https://doi.org/10.1007/s40572-018-0179-y

GDAL/OGR contributors. (2021). *GDAL/OGR Geospatial Data Abstraction software Library. Open Source Geospatial Foundation.* .

Gorelick, N., Hancher, M., Dixon, M., Ilyushchenko, S., Thau, D., & Moore, R. (2017). Google Earth Engine: Planetary-scale geospatial analysis for everyone. *Remote Sensing of Environment*, *202*,

18–27. https://doi.org/10.1016/j.rse.2017.06.031

Harmonization of DMSP and VIIRS nighttime light data from 1992-2018 at the global scale. (n.d.). Retrieved March 11, 2021, from https://figshare.com/articles/dataset/Harmonization_of_DMSP_and_VIIRS_nighttime_light_data_from_1992-2018_at_the_global_scale/9828827/2

Li, X., & Zhou, Y. (2017). A Stepwise Calibration of Global DMSP/OLS Stable Nighttime Light Data (1992–2013). *Remote Sensing*, *9*(6), 637. https://doi.org/10.3390/rs9060637

Li, X., Zhou, Y., Zhao, M., & Zhao, X. (2020). A harmonized global nighttime light dataset 1992–2018. *Scientific Data*, *7*(1), 1–9. https://doi.org/10.1038/s41597-020-0510-y

Mills, S., Weiss, S., & Liang, C. (2013). VIIRS day/night band (DNB) stray light characterization and correction. In J. J. Butler, X. (Jack) Xiong, & X. Gu (Eds.), *Earth Observing Systems XVIII* (Vol. 8866, p. 88661P). SPIE. https://doi.org/10.1117/12.2023107

Modified Soil-adjusted Vegetation Index. (n.d.). Retrieved March 11, 2021, from https://wiki.landscapetoolbox.org/doku.php/remote_sensing_methods:modified_soil-adjusted_vegetation_index

Python Software Foundation (Python Language Reference, version 3.7). (n.d.). https://www.python.org/

QGIS Geographic Information System. (2009). Open Source Geospatial Foundation. QGIS Development Team. http://qgis.org

Status Maps — Copernicus Land Monitoring Service. (2020). https://land.copernicus.eu/pan-european/high-resolution-layers/imperviousness/status-maps

Su, J. G., Dadvand, P., Nieuwenhuijsen, M. J., Bartoll, X., & Jerrett, M. (2019). Associations of green space metrics with health and behavior outcomes at different buffer sizes and remote sensing sensor resolutions. *Environment International*, *126*, 162–170. https://doi.org/10.1016/j.envint.2019.02.008

USGS. (n.d.). *LP DAAC - MOD44W*. Retrieved September 18, 2020, from https://lpdaac.usgs.gov/products/mod44wv006/

Wagtendonk, A., & Lakerveld, J. (2019). *Walkability score Netherlands version 1.0*.

Walk and the City | Digital Earth. (n.d.). Retrieved March 11, 2021, from https://digitalearthlab.jrc.ec.europa.eu/app/walk-and-city

WHO. (2017). *Urban green spaces: a brief for action*.

Zhao, Zhou, Li, Cao, He, Yu, Li, Elvidge, Cheng, & Zhou. (2019). Applications of Satellite Remote Sensing of Nighttime Light Observations: Advances, Challenges, and Perspectives. *Remote Sensing*, *11*(17), 1971. https://doi.org/10.3390/rs11171971

ANNEX

Table S1. Land use categories in Urban Atlas (version 2006). Green-colored classes represent selected green spaces within EXPANSE project.

CODE	ITEM
11100	Continuous Urban Fabric (S.L. > 80%)
11210	Discontinuous Dense Urban Fabric (S.L. : 50% - 80%)
11220	Discontinuous Medium Density Urban Fabric (S.L. : 30% - 50%)
11230	Discontinuous Low Density Urban Fabric (S.L. : 10% - 30%)
11240	Discontinuous Very Low Density Urban Fabric (S.L. < 10%)
11300	Isolated Structures
12100	Industrial, commercial, public, military and private units
12210	Fast transit roads and associated land
12220	Other roads and associated land
12230	Railways and associated land
12300	Port areas
12400	Airports
13100	Mineral extraction and dump sites
13300	Construction sites
13400	Land without current use
14100	Green urban areas
14200	Sports and leisure facilities
20000	Agricultural + Semi-natural areas + Wetlands
30000	Forests
40000	Wetlands
50000	Water bodies

Table S2. Land use categories in Urban Atlas (version 2012 and 2018). Green-colored classes represent selected green spaces within EXPANSE project.

CODE	ITEM
11100	Continuous urban fabric (S.L.: > 80%)
11210	Discontinuous dense urban fabric (S.L.: 50% - 80%)
11220	Discontinuous medium density urban fabric (S.L.: 30% - 50%)
11230	Discontinuous low density urban fabric (S.L.: 10% - 30%)
11240	Discontinuous very low density urban fabric (S.L.: < 10%)
11300	Isolated structures
12100	Industrial, commercial, public, military and private units
12210	Fast transit roads and associated land
12220	Other roads and associated land
12230	Railways and associated land
12300	Port areas
12400	Airports
13100	Mineral extraction and dump sites
13300	Construction sites
13400	Land without current use
14100	Green urban areas
14200	Sports and leisure facilities
21000	Arable land (annual crops)
22000	Permanent crops
23000	Pastures
24000	Complex and mixed cultivation patterns
25000	Orchards at the fringe of urban classes
31000	Forests
32000	Herbaceous vegetation associations
33000	Open spaces with little or no vegetation
40000	Wetlands
50000	Water
91000	No data (Clouds and shadows)
92000	No data (Missing imagery)

Table S3. Land cover classification for Corine Land Cover (CLC). Green-coloured classes represent selected green spaces within EXPANSE project.

CLC_CODE	LABEL1	LABEL2	LABEL3
111	Artificial surfaces	Urban fabric	Continuous urban fabric
112	Artificial surfaces	Urban fabric	Discontinuous urban fabric
121	Artificial surfaces	Industrial, commercial and transport units	Industrial or commercial units
122	Artificial surfaces	Industrial, commercial and transport units	Road and rail networks and associated land
123	Artificial surfaces	Industrial, commercial and transport units	Port areas
124	Artificial surfaces	Industrial, commercial and transport units	Airports
131	Artificial surfaces	Mine, dump and construction sites	Mineral extraction sites
132	Artificial surfaces	Mine, dump and construction sites	Dump sites
133	Artificial surfaces	Mine, dump and construction sites	Construction sites
141	Artificial surfaces	Artificial, non-agricultural vegetated areas	Green urban areas
142	Artificial surfaces	Artificial, non-agricultural vegetated areas	Sport and leisure facilities
211	Agricultural areas	Arable land	Non-irrigated arable land
212	Agricultural areas	Arable land	Permanently irrigated land
213	Agricultural areas	Arable land	Rice fields
221	Agricultural areas	Permanent crops	Vineyards
222	Agricultural areas	Permanent crops	Fruit trees and berry plantations
223	Agricultural areas	Permanent crops	Olive groves
231	Agricultural areas	Pastures	Pastures
241	Agricultural areas	Heterogeneous agricultural areas	Annual crops associated with permanent crops
242	Agricultural areas	Heterogeneous agricultural areas	Complex cultivation patterns
243	Agricultural areas	Heterogeneous agricultural areas	Land principally occupied by agriculture, with significant areas of natural vegetation
244	Agricultural areas	Heterogeneous agricultural areas	Agro-forestry areas
311	Forest and semi natural areas	Forests	Broad-leaved forest
312	Forest and semi natural areas	Forests	Coniferous forest
313	Forest and semi natural areas	Forests	Mixed forest
321	Forest and semi natural areas	Scrub and/or herbaceous vegetation associations	Natural grasslands
322	Forest and semi natural areas	Scrub and/or herbaceous vegetation associations	Moors and heathland
323	Forest and semi natural areas	Scrub and/or herbaceous vegetation associations	Sclerophyllous vegetation
324	Forest and semi natural areas	Scrub and/or herbaceous vegetation associations	Transitional woodland-shrub
331	Forest and semi natural areas	Open spaces with little or no vegetation	Beaches, dunes, sands

332	Forest and semi natural areas	Open spaces with little or no vegetation	Bare rocks
333	Forest and semi natural areas	Open spaces with little or no vegetation	Sparsely vegetated areas
334	Forest and semi natural areas	Open spaces with little or no vegetation	Burnt areas
335	Forest and semi natural areas	Open spaces with little or no vegetation	Glaciers and perpetual snow
411	Wetlands	Inland wetlands	Inland marshes
412	Wetlands	Inland wetlands	Peat bogs
421	Wetlands	Maritime wetlands	Salt marshes
422	Wetlands	Maritime wetlands	Salines
423	Wetlands	Maritime wetlands	Intertidal flats
511	Water bodies	Inland waters	Water courses
512	Water bodies	Inland waters	Water bodies
521	Water bodies	Marine waters	Coastal lagoons
522	Water bodies	Marine waters	Estuaries
523	Water bodies	Marine waters	Sea and ocean
999	NODATA	NODATA	NODATA
990	UNCLASSIFIED	UNCLASSIFIED LAND SURFACE	UNCLASSIFIED LAND SURFACE
995	UNCLASSIFIED	UNCLASSIFIED WATER BODIES	UNCLASSIFIED WATER BODIES