EO Clinic project:

Nature-Based Solutions in Cities

Work Order Report

Support requested by:
World Bank Group (WBG)

e-geos AN ASI / TELESPAZIO COMPANY | DHI GRAS | indra

Reference: EOC0020_WOR_v01
Date: 2021 July 30

TABLE OF CONTENTS

LIST OF FIGURES

ABOUT THIS DOCUMENT

This publication was prepared in the framework of the EO Clinic (Earth Observation Clinic, see below), in partnership between ESA (European Space Agency), the World Bank Group (WBG) and a team of service providers contracted by ESA: e-GEOS S.p.A. (Italy) as Prime with support from DHI GRAS (Denmark) and INDRA (Spain).

This Work Order Report (WOR) describes the context of the team activities on Nature-Based Solutions in Cities, the geoinformation requirements of the activities and finally, the EO products and services delivered by the EO Clinic service providers in support of those activities.

ABOUT THE EO CLINIC

The EO Clinic (Earth Observation Clinic) is an ESA (European Space Agency) initiative to create a rapid-response mechanism for small-scale and exploratory uses of satellite EO information in support of a wide range of International Development projects and activities. The EO Clinic consists of "on-call" technically pre-qualified teams of EO service suppliers and satellite remote sensing experts in ESA member states. These teams are ready to demonstrate the utility of satellite data for the development sector, using their wide range of geospatial data skills and experience with a large variety of satellite data types.

The support teams are ready to meet the short delivery timescales often required by the development sector, targeting a maximum of 3 months from request to solution.

The EO Clinic is also an opportunity to explore more innovative EO products related to developing or improving methodologies for deriving socio-economic and environmental parameters and indicators.

The EO Clinic was launched in March 2019 and is open to support requests by key development banks and agencies during the 3 years project duration.

AUTHORS

The present document was prepared and coordinated by Valeria Donzelli (project manager, e-GEOS) with main contributions from Mads Christensen (Geoinformatics specialist, DHI GRAS), Radoslaw Guzinski (Senior remote sensing specialist, DHI GRAS), Daniel Druce (Remote sensing expert, DHI GRAS), Cecilia Helena Valdés (GIS expert and Remote Sensing Specialist, INDRA) and Cesar Morate Urquiaga (GIS Senior Consultant, INDRA).

ACKNOWLEDGEMENTS

The following colleagues provided valuable inputs, insights and evaluation feedback on the work performed: Boris van Zanten, (Senior Consultant Nature-based Solutions & Disaster Risk Management, WBG), and Zoltan Bartalis (ESA Coordinator and Technical Officer).

For further information	Please contact Valeria Donzelli, project manager, e-GEOS S.p.A. (valeria.donzelli@e-geos.it) with copy to Zoltan.Bartalis@esa.int if you have questions or comments with respect to content or if you wish to obtain permission for using the material in this report. Visit the ESA EO Clinic: https://eo4society.esa.int/eo_clinic.

1 DEVELOPMENT CONTEXT AND BACKGROUND

Rapid urbanisation in recent decades has led to significant growth of urban areas worldwide, and as populations continue to grow, and urban areas expand, pressure on all types of infrastructure and development objectives is mounting. At the same time, communities find themselves increasingly exposed to the growing challenges and impacts caused by climate change with more frequent flooding's, droughts and intense heat waves. Consequently, cities around the world are adapting to climate change and enhancing their resilience, including mitigating the risk of flooding and rising temperatures.

Nature Based Solutions (NBS) are increasingly used in urban areas to enhance the quality of life and address the concomitant challenges caused by a changing climate and the exacerbated problems caused by urbanisation processes. Complementary integration of both grey, blue and green infrastructure into urban planning processes, such as turning impermeable surfaces into green spaces, tree plantings, water features and storage areas for excess run-off are amongst such NBS activities that contribute to increase the resilience of urban areas, e.g., by lowering temperatures to reduce the impact of urban heat islands, providing green recreational spaces and strengthening existing flood management infrastructures. [1]

Scaled-up implementation of urban NBS is supported by improved and targeted planning, monitoring and evaluation tools to determine the potential and suitability of NBS locations and assess underlying conditions. Documentation of past extreme events, such as information on historic flooding's and flood zones as well as insight pertaining to land use and land cover dynamics and seasonality patterns, climate information and temperature variation, all contribute to identify NBS hot spot zones and qualify the potential impact of NBS activities. [2,3]

Earth Observation (EO) solutions provides a cost-efficient and effective approach to inform urban NBS planning, by providing information on the underlying and historic conditions of urban areas at scale, thereby contributing to identify NBS hot spot zones by determining spatial suitability and effectiveness.

1.1 Objective

The overall objective of this activity was to develop a set of EO-derived datasets for the World Bank Group, to pilot a novel geospatial suitability methodology to assess NBS potential in urban and urbanising regions and identify spatially suitable and effective locations for NBS activities, including flood risk management and urban heat island reduction and co-benefits. Within the proposed framework, the following EO based mapping components were conducted:

- Land use/land cover mapping and climate information
    - High resolution (10 m) land use and land cover (LULC) mapping and mapping of associated dynamics of green areas and water bodies.
    - Provision of climate parameter timeseries, including daily minimum, maximum, mean and range of air and surface temperatures; daily maximum and mean relative humidity; and daily total rainfall.
- Urban flooding
    - High resolution (10 m) flood extent mapping of the January 2020 flooding's in Antananarivo, Madagascar, and flood hazard mapping through timeseries analysis of surface water frequency.

[1] UN Environment-DHI, UN Environment and IUCN 2018. Nature-Based Solutions for Water Management: A Primer
[2] UN Environment-DHI, UN Environment and IUCN 2018. Nature-Based Solutions for Water Management: A Primer
[3] Asian Development Bank. Nature-based solutions for building resilience in towns and cities: Case studies from the Greater Mekong Subregion. Mandaluyong City, Philippines: Asian Development Bank, 2016

    - Provision of Digital Elevation Model (DEM) at 30 m spatial resolution and derived parameters such as slope, Topographic Wetness Index (TWI) and Height Above Nearest Drainage (HAND) to describe terrain characteristics as well as soil data to describe soil properties.
- Urban heat
    - Timeseries of bimonthly land surface temperature maps of Tashkent, Uzbekistan in high resolution (30 m) covering the period 2015-2020 in addition to urban heat island indicators.

For each data layer, a product sheet was developed containing key metadata descriptions as well as information about data usage and interpretation, limitations, processing steps and feasibility/applicability in urban areas, globally.

2 PROPOSED WORK LOGIC FOR EO-BASED SOLUTIONS

The project was structured into three tasks or Work Packages (WP).

- WP1 covered all the management activities and lasts for the whole project duration.
- WP2 dealt with acquisition and processing of data required to realise services.
- WP3 generated Service 1 – Consolidated NBS-Relevant EO Mapping Products, which is described in detail in section 3.1.

The data and information layers delivered through this service are all derived from open and non-commercial input data with global coverage, to enable scaling and cost-effective replication in any other World Bank client country. Consequently, the data layers demonstrate the feasibility for using EO datasets as a low-cost supplement, or alternative, to existing very high-resolution aerial and in situ datasets. The diagram in Figure 1 describes the rationale for the implementation of the overall service workflow.

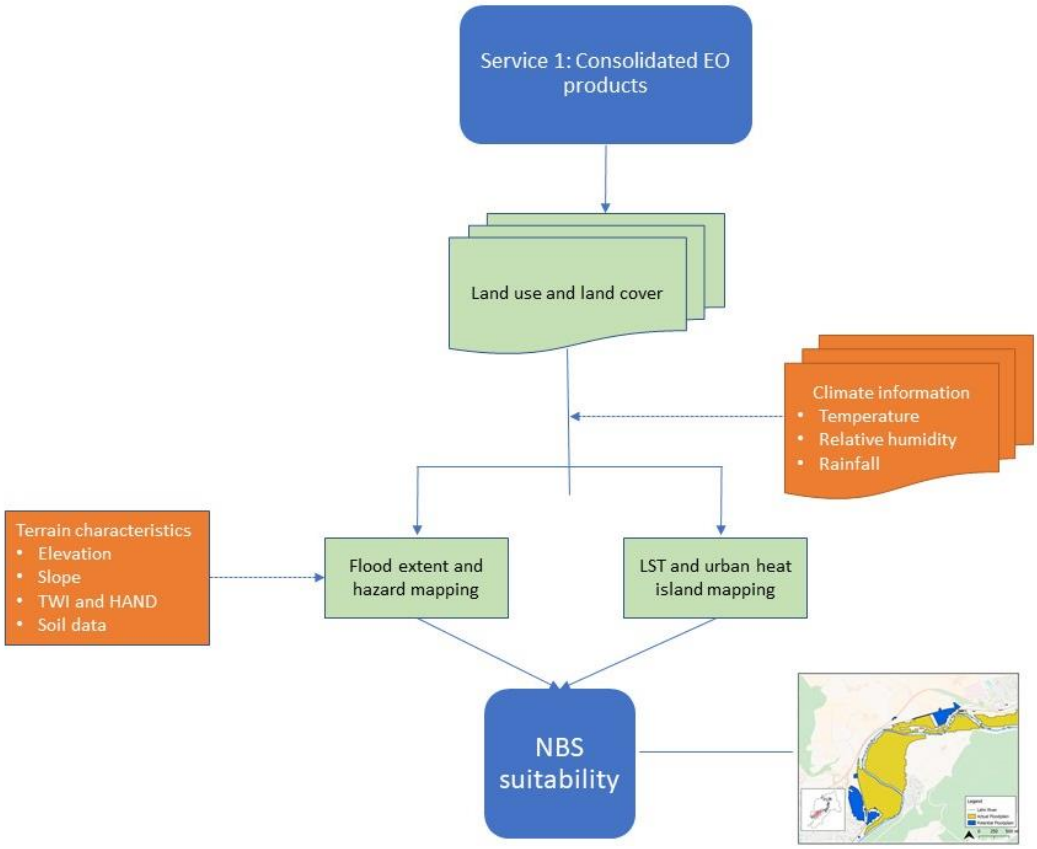

Figure 1 Central building blocks and workflow logic

2.1 Area of interest

The area of interests (AoI # 1 - Figure 2 and AoI # 2 - Figure 3) fully encompasses the cities of Antananarivo, Madagascar and Tashkent, Uzbekistan, plus a buffer of 2 km around each city. The buffer around the urban areas was added to capture trends and variability in peri urban areas, relevant for urban NBS activities. Accordingly, AoI # 1 covers an extent of 180 km² around Antananarivo and AoI # 2 an extent of 568 km² around Tashkent.

Antananarivo and Tashkent were chosen as preferred AoIs as both are sizeable cities with dense urban centres and diverse and fragmented peri urban areas, thus provide challenging, yet representative samples for delivering the proposed services and scaling these to other cities worldwide.

LULC mapping and climate information covers both AOIs, while urban flooding and urban heat covers AoI # 1 and AOI # 2, respectively.

Figure 2 Service area of interest # 1, Antananarivo, Madagascar

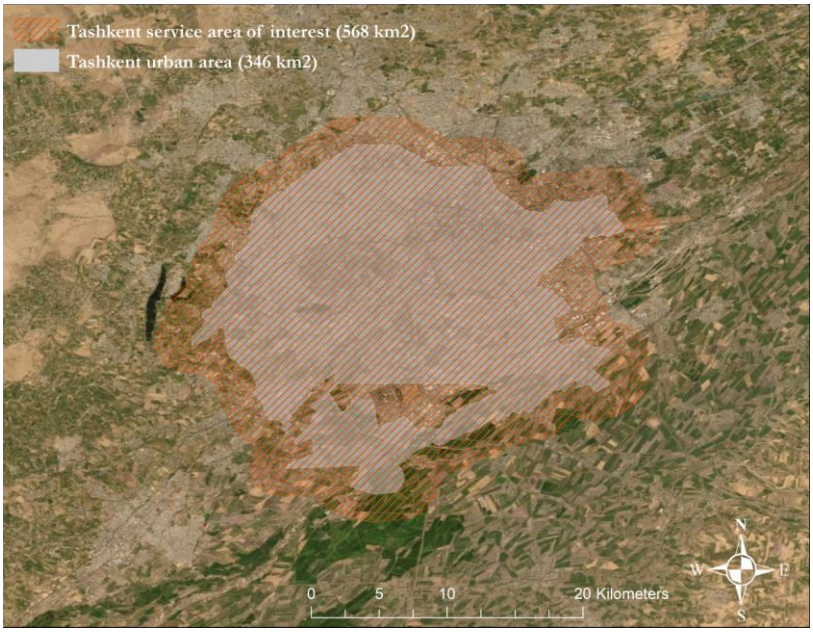

Figure 3 Service area of interest # 1, Tashkent, Uzbekistan

The results presented in this report are final i.e., they represent the stage of development and production achieved at the end of the procurement. Further, refinements and potential additional datasets will be considered in response to comments and/or clarifications from ESA and World Bank counterparts.

3 DELIVERED EO-BASED PRODUCTS AND SERVICES

3.1 Service 1: Consolidated NBS-Relevant EO Mapping Products

Service 1 delivered a series of consolidated NBS-relevant EO data products for extended AoIs around the cities of Antananarivo and Tashkent.

- Land use/land cover mapping and climate information for both AoIs.
- Urban flooding datasets and terrain characteristics for AoI # 1 - Antananarivo
- Land surface temperature and urban heat island indicators for AoI # 2 – Tashkent.

3.1.1 Data layers for both urban flooding and urban heat

3.1.1.1 General Land Use and Land Cover

Climate change, which is becoming increasingly noticeable, is having a very direct impact on cities and their inhabitants. Their resilience to adapt to the new circumstances and to mitigate the effects of climate change is being put to the test.

In the more specific subject of concern (heat island and floods), to begin any study and to understand any phenomenon, it is essential to have a General Land Use Land Cover (LULC) map as the base product, categorized by representative classes in the urban environment and its surroundings.

To this end, the LULC product was generated by a completely automatic process, that can be extrapolated to any region of the world and is replicable in any given city. The product provides discrete information on land use classes extracted from Sentinel-2 imagery, focused on urban land use classes, which are the ones that change more rapidly.

The following 13 classes were extracted for both Antananarivo and Tashkent for the year 2017, 2018, 2019 and 2020:

Table 1. Land Use Land Cover classification.

Code	Description
1	Airports
2	Mineral Extraction and Construction sites
3	Continuous urban fabric (SS.L.: >30%)
4	Discontinuous urban fabric (S.L.: <=30%)
5	Forests
6	Sports and leisure facilities and Green urban areas
7	Industrial, Commercial, Public, Military and Private Units
8	Port Areas
9	Water
10	Isolates Structures
11	Crops
12	Orchards and Pastures
0	Other Land Covers

The final results for each LULC per year are as shown in the following figure:

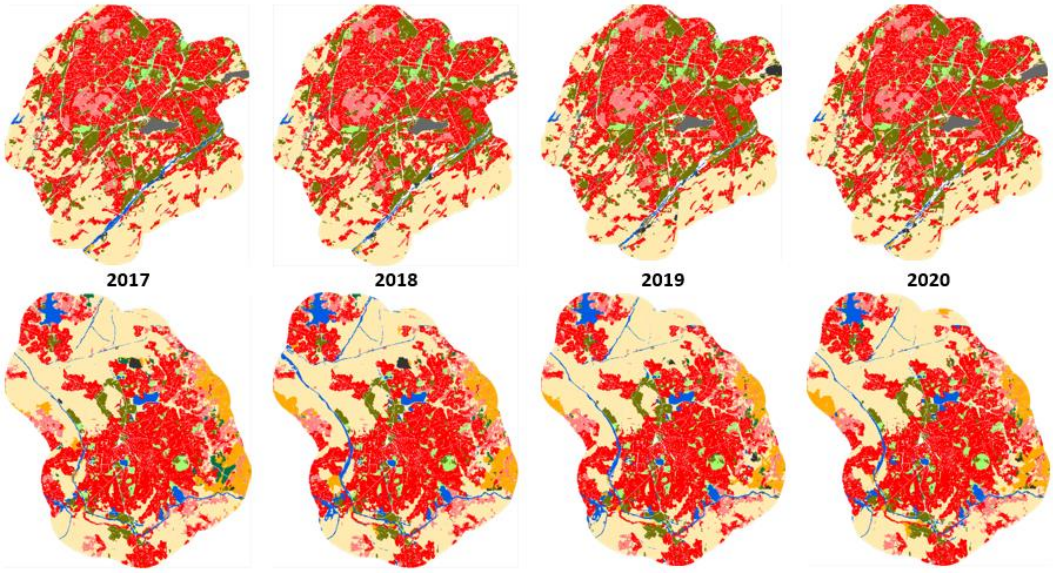

Figure 4 Land Use / Land Cover evolution 2017-2020. Tashkent (up), Antananarivo (bottom)

Additionally, seeing as in the scope of this project vegetation dynamics is an important issue and is not completely covered by the LULC, a more exhaustive study was performed by calculating the Normalized Difference Vegetation Index (NDVI) per Building Block (BB) for both AOIs (two seasons per year). The NDVI gives complementary information to the uses reflected in the LULC, thus adding insight to the product in hand. The data generated by this index is easy to compare between seasons and can be used to estimate and monitor the vegetation variability in the AOIs. An example of the results obtained is shown below:

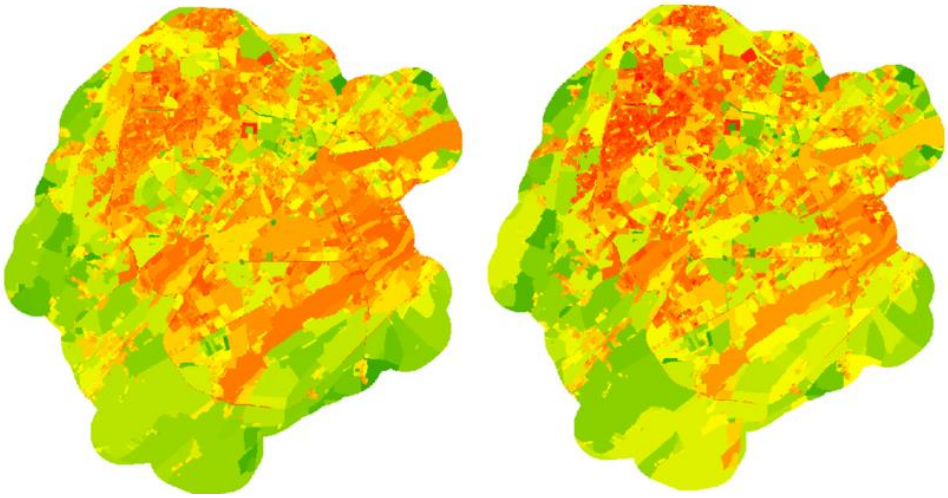

Figure 5: Seasonal NDVI depicted by building block, Tashkent (2020). Dry season (left), wet season (right)

3.1.1.1.1 Specifications

The technical process gone through to obtain both the LULC predictions and the NDVI is specified in the table below:

EO input data: Sentinel-2

Other input data: OSM roads

Method: The methodology used consists in applying Artificial Intelligence (AI) automated models for the extraction of the classification of **Land Use and Land Cover (LULC)** entities. The resulting product has 13 categories of land uses per year. To achieve more complete and accurate results, the Open Street Map (OSM) transport network for each AOI was downloaded and added to the LULC, using in house python developments. OSM is a free source platform available worldwide and can therefore be applied in any given area.

Additionally to the LULC, the **NDVI** was then calculated for both **wet and dry season** per year. The NDVI was calculated with S-2 images by an automated process that took into account the mean per season. Results are depicted by **Building Blocks (BB),** which were calculated using the OSM road network, representing the mean NDVI on each BB.

Output indicators: Land Use and Land Cover (LULC); Mean NDVI per Building Block (NDVI per BB)

Map legend: LULC classes (13 different classes); Mean NDVI, continuous raster map

Spatial resolution: 10 m pixel resolution

Temporal resolution: 2017, 2018, 2019, 2020

Delivery format: Esri Shapefile (LULC), GeoTIFF (NDVI) and PNG for QC matrix; additional information or other data formats upon request.

3.1.1.1.2 Mapping process

The LULC prediction and NDVI calculations generated rely on a series of steps and automated processes which are depicted in the following sections.

3.1.1.1.3 Image processing

The most important dataset needed for the generation of these products is the high-resolution (10m) Sentinel-2 optical imagery. For this purpose, a separate tool was developed to select the S-2 imagery (CREODIAS) through an automatic process, in accordance to the specified time stamp and AOI.

The first search obtains images with 2% cloud coverage, increasing its reach every 2% until filling the AOI, seeking to cover as much area as possible as to minimize the number of images used for the mosaic. Once this process is done, it extracts a co-registered, mosaicked and UTM reprojected image with its corresponding Top Of Atmosphere (TOA) corrections made.

3.1.1.1.4 LULC mapping

Once the images were obtained, the methodology used consists in applying an automated Deep Learning Model for the extraction of the different Land Use and Land Cover (LULC) classes. The main technique for the extraction of these classes is via segmentation by implementing a Convolutional Neural Network and is a model that can be scaled and applied worldwide.

The automated AI model used for the extraction of the LULC classes has been trained using mainly Urban Atlas (UA) layers available for Europe, as well as manually generated datasets by digitalization in other areas of the world. For this project, seeing as the terrain characteristics for both Tashkent and Antananarivo vary from Europe, specific training datasets were created in order to obtain more accurate results. The AI model was then retrained with these new datasets in order to obtain the LULC layers. The resulting product has 13 categories of land uses per year, based on those of UA and adapted to Sentinel-2 resolution.

As to enrich and complete these LULC results, auxiliary data from OSM transport network for each AOI was automatically merged with the resulting LULC. In house python scripting for filtering, classified and merged the data regardless the region under study.

As a post-processing step, in order to have clean results and avoid unnecessary noise, polygons with a 0.25ha Minimum Mapping Unit (MMU) were automatically detected and eliminated.

3.1.1.1.5 NDVI

In addition, by using the auxiliary OSM roads data extracted in the previous process, Building Blocks (polygons enclosed within the road network) were defined for each AOI. BB allow the comparison and analysis of the evolution over a Time Series (TS). These can be very useful for urban planners, seeing as they are based on the urban structure of cities.

Finally, the NDVI was obtained by an automated process that first calculated the index for each month and then took into account the mean per season (wet season Nov-April, dry season May-Oct) for every year (2017, 2018, 2019 and 2020). This seasonal mean was then assigned to each BB throughout zonal statistic, thus obtaining a layer of vegetation mean per season for each year in both AOIs.

Figure 6: Example of Building Blocks created with OSM road network, Antananarivo (Madagascar)

3.1.1.1.6 Quality Control and Validation

The Quality Control (QC) carried out involved both topological and thematic controls to ensure the final quality of the product. After production, an external operator who was not implicated in the production process visually reviewed the layers.

For each LULC, manual quality control has been carried out due to the absence of a ground truth with which to carry out an automatic QC of the product. The methodology used consists of evaluating via photointerpretation ≈1000 random points in a stratified sample, giving more weight to the classes based on their representation in the area of study. It was therefore possible to evaluate the degree of agreement between the layers obtained and the Sentinel-2 images used as reference. As a result, the degree of accuracy represented in the map on the image was appropriate. For the current service, 4 out of the 8 LULC layers have been reviewed and their results are shown in the following figure:

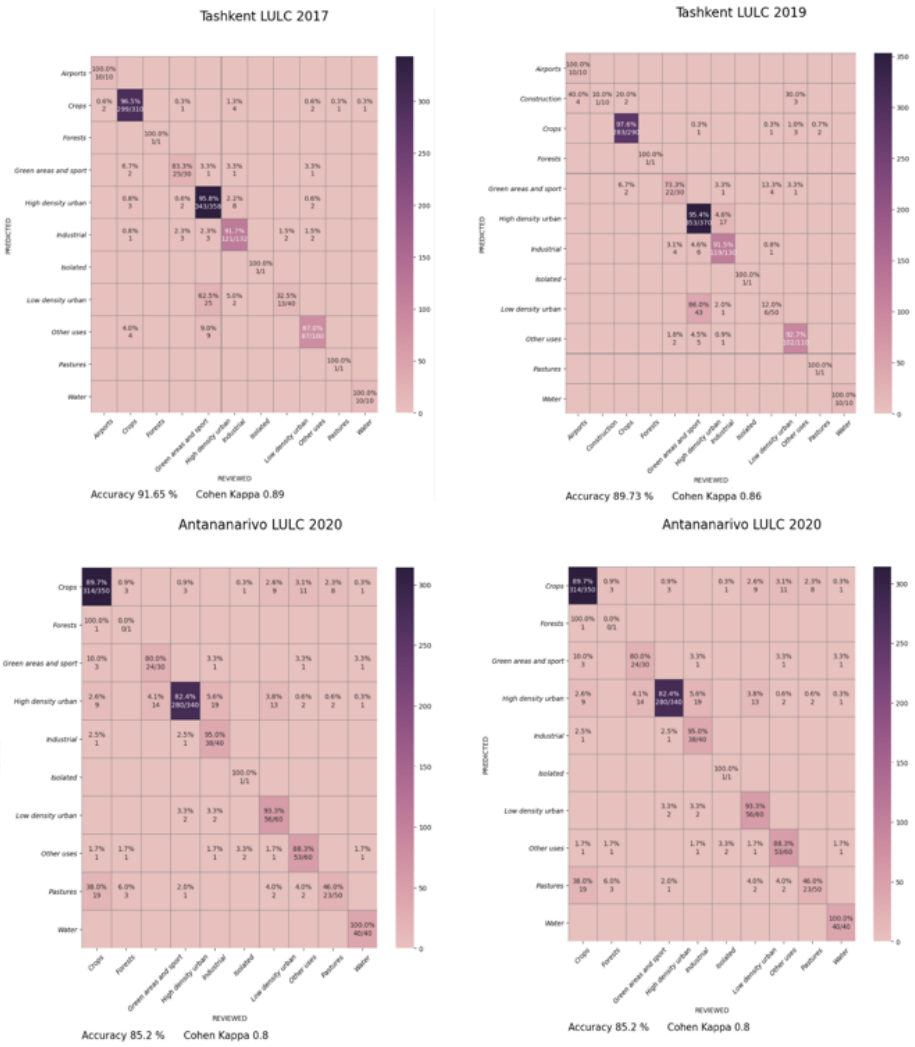

Figure 7 : Quality Control confusion matrix results

3.1.1.1.7 Usage, Limitations and Constraints

Two important limitations of using Sentinel-2 are:

- The resolution of 10m limits the extraction of detail for the LULC classes.
- The presence of clouds can affect the predictions. In order to avoid any imprecisions, an in-house python script was used for this purpose, minimizing its impact.

Other limitations:

- OSM can be an incomplete data source with geometrical errors affecting the results.
- Seeing as the model was mostly trained for Europe, the different landscapes and morphologies of both AOIs might create some inconsistencies in the predictions.

3.1.1.2 General climate information

Knowledge about the spatiotemporal dynamics of climatological parameters, as well as up to date information on current/recent weather patterns provides key information to inform NBS activities in urban areas.

The Copernicus Climate Change Service provides access to high-quality climate reanalysis data through the Copernicus Climate Data Store (CDS).[4] The reanalysis data is based on advanced climate models driven by, and corrected with, in situ and remotely sensed measurements. The data is produced by the European Centre for Medium-Range Weather (ECMWF) forecasts using the ERA5 model.[5] ERA5 is run globally on a 0.25° grid (around 31 km) and provides hourly estimates of several atmospheric, land and oceanic climate variables covering the period from January 1950 to present. A subset of ERA5, called ERA5-Land, provides hourly information in a 9 km spatial resolution covering the period from 1981 to 2-3 months before the present. However, this product does not cover coastal areas, and is thus limited to areas which are at least 20-30 km inland. Both ERA5 and ERA5-Land provide datasets on land surface parameters such as air temperature, surface temperature, humidity, and rainfall.

A timeseries of climate parameters for the cities of Tashkent and Antananarivo (and their surroundings) based on ERA5-Land dataset was sourced as part of Service 1. The following parameters were included: daily minimum, maximum, mean and range of air and surface temperatures; daily maximum and mean relative humidity; and daily total rainfall. In addition, daily timeseries of the Universal Thermal Climate Index (UTCI) was provided. UTCI represents current state-of-the-art bioclimatology data and describes how atmospheric conditions, specifically air temperature, humidity, ventilation and radiation are experienced by the human body. UTCI is derived from ERA5 data and therefore provided with a resolution of roughly 31 km.

[4] Available online at https://cds.climate.copernicus.eu/#!/home
[5] More information about ERA5 at https://www.ecmwf.int/en/forecasts/dataset/ecmwf-reanalysis-v5

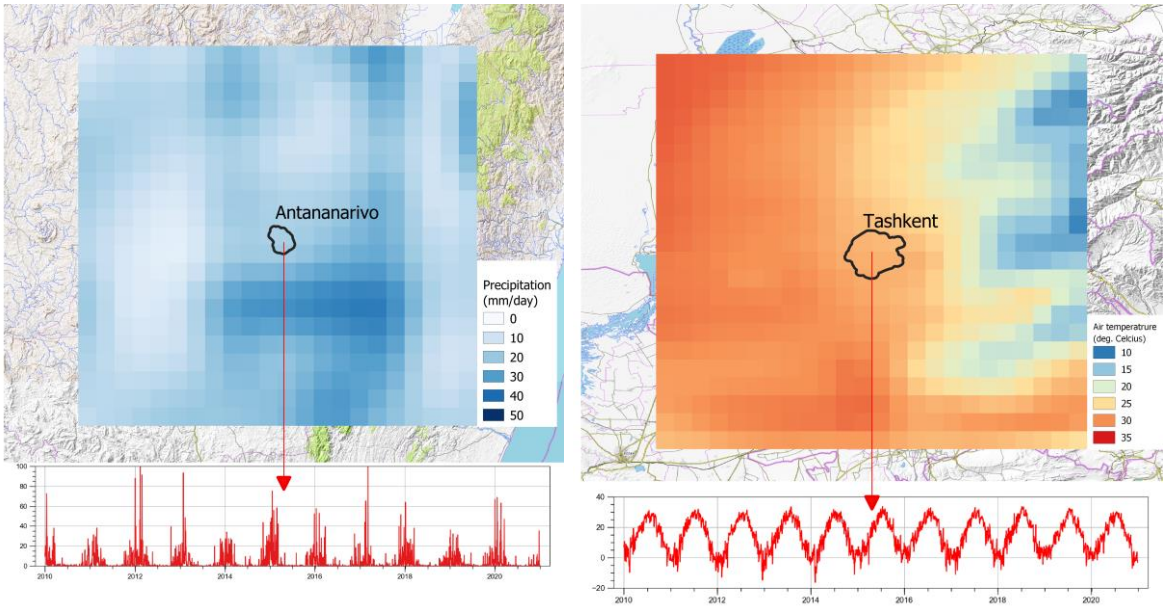

Figure 8: Example ERA5-land derived time-series data of daily precipitation in Antananarivo, Madagascar (left), example ERA5-land derived time-series data on air temperature in Tashkent, Uzbekistan (right)

3.1.1.2.1 Specifications

Technical specifications of the climate data products are summarised below.

EO input data: RTTOV-11, all-sky for various components

Family of ERA5 data used: ERA5-Land[6] and UTCI

Method: Details of the production of the ERA5 datasets is available in the ERA5 documentation. Further details on UTCI are available in the UTCI documentation.

Further postprocessing of the data was conducted to generate daily composites, as the hourly timesteps provided by ERA5 are too frequent for the purposes of long-term urban climate mapping. Therefore, the data was aggregated to daily timesteps (local midnight to local midnight).

Output indicators: Mean, maximum, minimum and range values for air and surface temperatures; mean and maximum for dew-point temperature; sum for precipitation; and numerical representation of the thermal comfort/discomfort for UTCI

Units: Degree Celsius for temperatures; mm/day for precipitation; and Kelvin for UTCI.

Spatial resolution: 9 km for ERA5-Land and 31 km for UTCI.

Temporal resolution: Daily composites from 2015 - 2021

Delivery format: NetCDF, additional information or other data formats upon request; added to webportal

[6] Muñoz Sabater, J., (2019): ERA5-Land hourly data from 1981 to present. Copernicus Climate Change Service (C3S) Climate Data Store (CDS). (Accessed on < 30-06-2021 >), 10.24381/cds.e2161bac

3.1.1.2.2 Quality Control and Validation

Information and documentation on quality assessment of ERA5-Land data are available from CDS[7] with further information provided in the ERA5 documentation in the section '*Accuracy and uncertainty*'[8].

3.1.1.2.3 Usage, Limitations and Constraints

The temporal and spatial resolutions of ERA5-Land make the dataset useful for various land surface applications, such as flooding or drought forecasting as well as assessments of temperature trends. UTCI provides information and insight into thermal stress as experienced in the human body, which is useful to track and monitor the effect of heat stress and temperature rise in urban environments.

For further information please see the ERA5-Land documentation[9] and the UTCI documentation.[10]

3.1.2 Data layers for urban flooding

3.1.2.1 Flood hazard

Flood hazard can be estimated through geospatial insight on historical trends and dynamics, as a proxy indicator for potential future flooding. By applying time-series analysis on continuous observations from both Sentinel-1 and -2 satellites, surface water dynamics can be mapped. The surface water frequency (SWF) product derived at yearly and seasonal basis for Antananarivo from 2017-2021, provides an effective way to track inter- and intra-annual changes and dynamics of surface water bodies, and thus providing insight into the frequency (or occurrence) of flood events. While the annual SWF products provide insight into the overall dynamics of the hydrological regime and longer-term trends and changes since 2017, the seasonal products produced for the dry seasons (May-Oct.) and wet seasons (Nov.-Apr.) since 2017 allow for intra-annual comparison to assess seasonal dynamics and shorter-term trends and variation.

An example of the annual SWF products for the year 2020 and 2017, as mapped from a synergistic use of Sentinel-1 and Sentinel-2 imagery, is shown in Figure 9 and 10, respectively. The maps show consistent spatial patterns of surface water and flood regimes but also high inter-annual variation in terms of surface water extent and frequency, which indicates the importance of understanding this variability better in order to support flood protection and mitigation.

[7] Available online at https://cds.climate.copernicus.eu/cdsapp#!/dataset/reanalysis-era5-land?tab=eqc
[8] Available online at https://confluence.ecmwf.int/display/CKB/ERA5%3A+data+documentation#ERA5:datadocumentation-Accuracy-anduncertainty
[9] Available online at https://confluence.ecmwf.int/display/CKB/ERA5-Land%3A+data+documentation
[10] Available online at https://datastore.copernicus-climate.eu/documents/utci/UTCIMRTdatabase_UserGuide.pdf

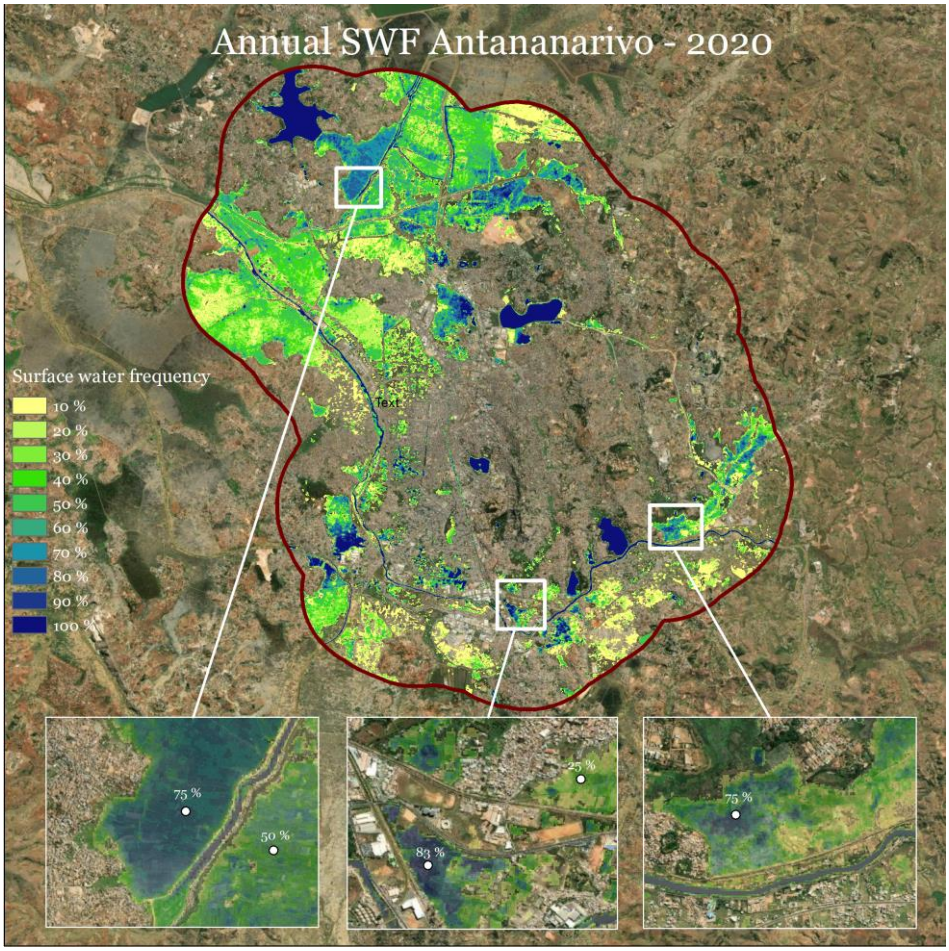

Figure 9: Annual SWF in Antananarivo in 2020

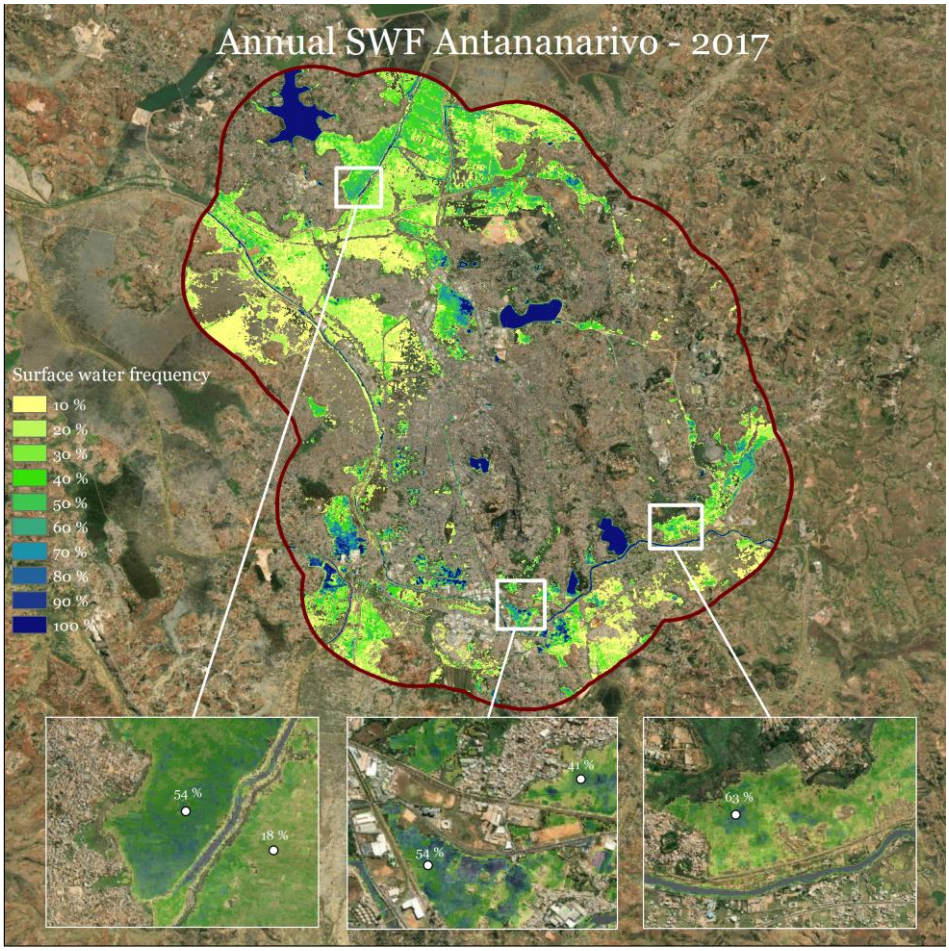

Figure 10: Annual SWF in Antananarivo in 2017

3.1.2.1.1 Specifications

Technical specifications for the SWF product are summarized below and with the details of the image processing and applied classification model provided in the following sections.

EO input data: Sentinel-1, Sentinel-2

Other input data: Digital Elevation Model (30 m resolution)

Method: The water extent maps are derived using a supervised machine learning algorithm (i.e., Logistic Regression) that takes a set of training data to establish the relationship between the response variable (i.e., water class) and the explanatory variables (cf. the satellite imagery). The model uses the full temporal resolution of the Sentinels to generate monthly water masks. Water frequency parameters (seasonal and annual frequency) are derived subsequently to give a comprehensive representation of the surface water variations between seasons and from year to year. A final urban mask was created from the Sentinel-2 imagery to remove inaccuracies in urban areas.

Output indicators: Surface Water frequency

Map legend: Water frequency [%]

Spatial resolution: 10 m pixel resolution.

Temporal resolution: Annual and seasonal, 2017-2021

Delivery format: GeoTiff, additional information or other data formats upon request; added to webportal

3.1.2.1.2 Image processing

The significant datasets used for this project are optical and SAR satellite data from Sentinel-2 and Sentinel-1. Before conducting the image analyses, essential pre-processing of the acquired data was conducted. The basic outline of the processing steps is given in below.

The Multi-Spectral Instrument (MSI) onboard Sentinel-2 acquires 13 spectral bands ranging from visible and near-infrared (VNIR) to shortwave infrared (SWIR) wavelengths along a 290-km orbital swath and a spatial resolution of 10 m (four visible and near-infrared bands), 20 m (six red edge and shortwave infrared bands) and 60 m (three atmospheric correction bands) (cf. Figure 11)

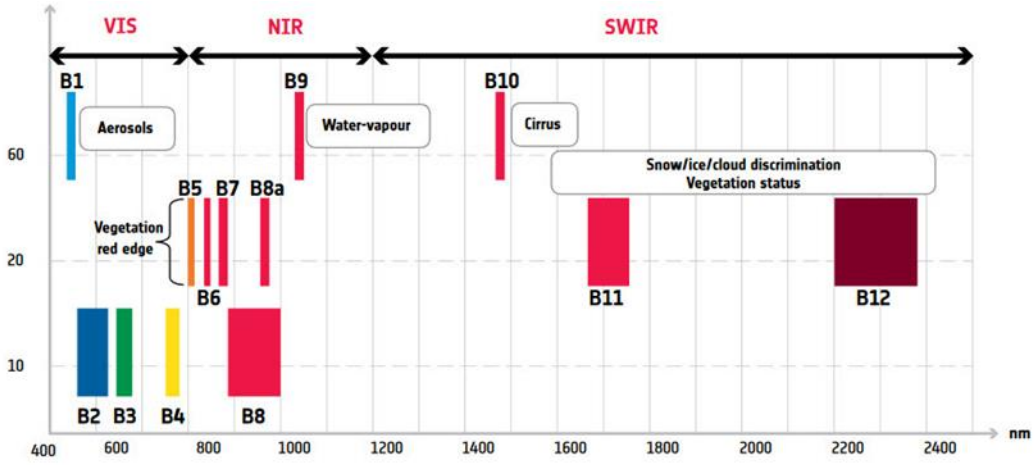

Figure 11. Spatial resolution versus wavelength of the Multi-Spectral Instrument (MSI) onboard Sentinel-2 (source: ESA).

We used all the bands with 10 and 20 m spatial resolution. We applied Top of the Atmosphere (TOA) correction to the data, and the 20 m bands were resampled to 10 m. In order to remove the cloudiest images from the time series, only images with a cloud cover less than 90 % (according to internal quality flags) were retained for the analysis. Masking of remaining clouds and cirrus as well as cloud shadows and snow were done by s2cloudless.

Thereafter, spectral indices were calculated based on combinations of the Sentinel-2 reflectance bands. Spectral indices are useful to highlight specific properties relevant for water detection (cf.Table 2). The data was then composited into monthly images.

Table 2. Sentinel-2 spectral indices for water presence prediction

Index	Short name	Equation [Sentinel-2 bands]
Normalized Difference Vegetation Index	NDVI	("B8") - ("B4") / ("B8") + ("B4")
Normalized Difference Water Index	NDWI	("B8") - ("B11") / ("B8") + ("B11")
Modified Normalized Difference Water Index	mNDWI	("B11") - ("B3") / ("B11") + ("B3")

From the Sentinel-1 dataset, we used Level-1 Informetric Wide Swath (IW) and Ground Range Detected (GRD) data. The data were processed to generate a calibrated, ortho-corrected product with a 10 m spatial resolution. The VV and VH backscattering values were used for the water classifications. The data was composited by monthly averages to remove some of the inherent speckle in SAR imagery. An exclusion mask was applied to

reduce the number of false positive predictions in SAR imagery which are common in bare and sandy soils because they exhibit a similar backscatter response as water. The exclusion layer was achieved by excluding pixels where there has been no evidence of water in the optical imagery in the preceding 3 months.

3.1.2.1.3 Water mapping

Predicting the extent of water using EO data relies on 3 key components: 1) training data, 2) machine learning, and 3) post-processing. The approach uses all available data from Sentinel-1, Sentinel-2, and feeds the information into a Logistic Regression model to predict water probability.

3.1.2.1.3.1 Training data

The main requirement of a well-trained classification model is a set of training samples that represent the classes of interest. In this case we had a binary (i.e. 2 classes) classification model which required samples of water and non-water locations, respectively. The training samples were compiled from the JRC Global Surface Water Explorer – GSWE.[11] A systematic random sampling approach was used to generate the training data for the model prediction.

3.1.2.1.3.2 Machine learning

There are various machine learning methods available that have been shown to work well for surface water mapping. Whilst accuracy should be the foremost concern when choosing a mapping approach, simplicity and ease of implementation are also important to consider to increase understanding, maintainability, and potential scalability. We therefore chose to implement a simple logistic regression model, in combination with logic-based masking, to ensure sufficient accuracy whilst maintaining computational efficiency and simplicity that facilitates analysis and understanding at scale. Logistic regression is named for the function used at the core of the method, the logistic function (aka the sigmoid function) defined mathematically as 1/ (1 + e^-value) where e is the base of the natural logarithms and value is the numerical value that you want to transform. Logistic regression is defined using an equation, where input values (x) are combined linearly using weights/coefficients (b) to predict an output value (y), e.g. y = e^(b0 + b1*x) / (1 + e^(b0 + b1*x)). The optimal coefficients of the logistic regression algorithm are estimated from the training data using maximum likelihood estimation.

3.1.2.1.3.3 Post-processing

A few post-processing routines were implemented to convert the water probability map secondly into annual/seasonal water frequencies. The outcome of the Logistic Regression equation is a probability estimate (0-100 %) for water presence in each 10x10 m pixel and for each month. These are converted into monthly binary products by using a probability threshold of 70 % to separate water from non-water. Water frequencies are then derived by taking the total number binary water predictions over a time period and dividing by the total number of observations over the same period and thereby returning a 0-1 percentage of the water frequency within the given period.

3.1.2.1.4 Quality Control and Validation

The quality control was done qualitatively since no ground data was available to validate the findings, meaning that all quality control and validation efforts are tied to visual inspection and evaluation of the data.

The data was also compared with JRC GSWE data for the same areas and visually inspected against samples of the source imagery used in the production. As seen from the examples visualised in **Errore. L'origine riferimento non è stata trovata.** and **Errore. L'origine riferimento non è stata trovata.**, the SWF product from EO-Clinic shows a well-defined and detailed insight into the hydrological regime and surface water dynamics, closely aligned with the dynamics seen in the high resolution (10 m) source imagery. Compared to global level alternatives, such as JRC GWSE data, which is 30 m spatial resolution, the 10 m resolution EO-Clinic SWF product is able to capture much finer details on surface water change and dynamics and more

[11] Pekel, J. F., Cottam, A., Gorelick, N., & Belward, A. S. (2016). High-resolution mapping of global surface water and its long-term changes. Nature, 540(7633), 418-422.

well defined flood plains and water bodies. These observations clearly show the benefit of using Sentinel-1 and Sentinel-2 10m data compared to Landsat 30m data.

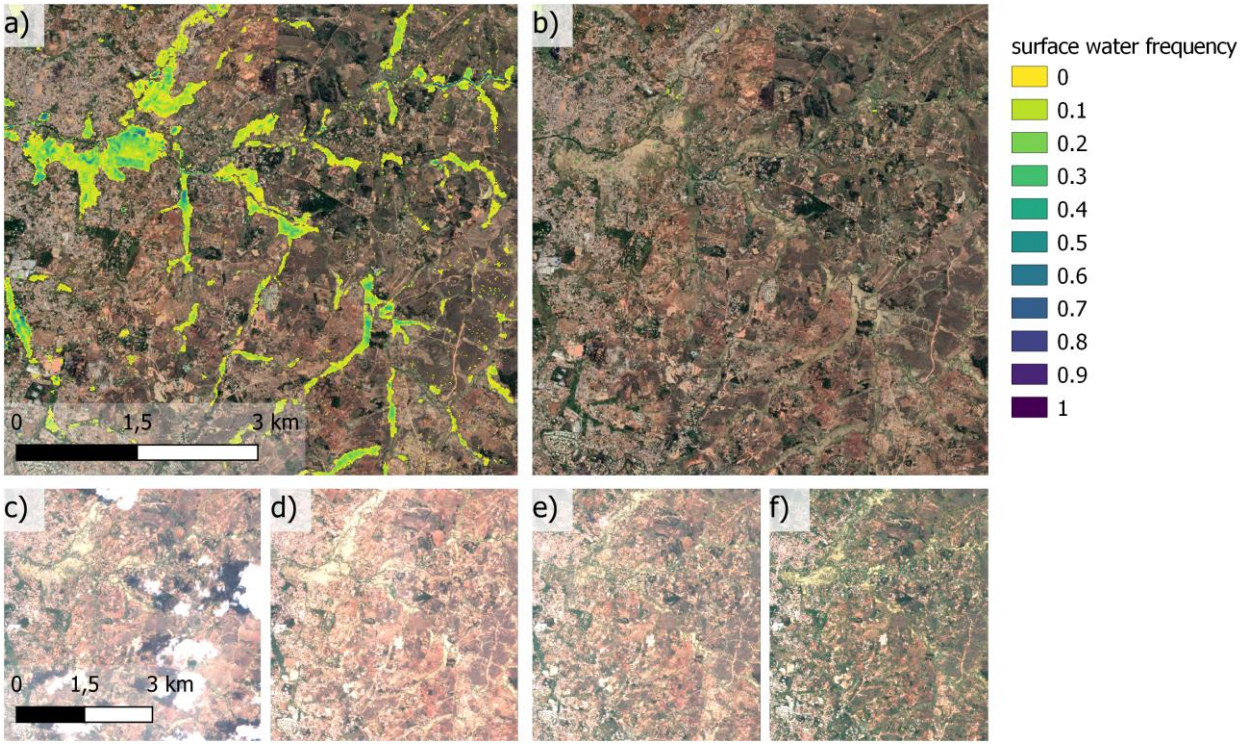

Figure 12: a) EO-Clinic SWF product, b) JRC GSWE, c-f) sentinel-2 RGB 01/01/2017, 12/07/2017, 11/07/2019, 01/11/2020

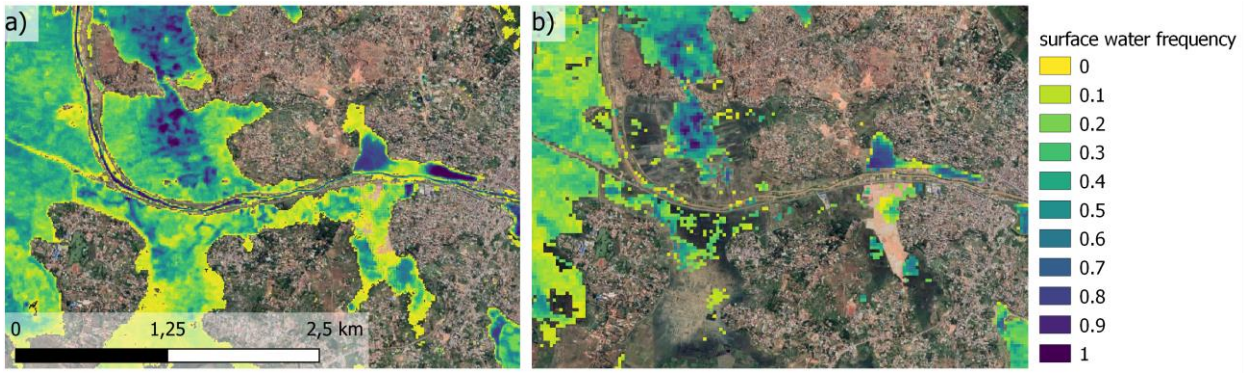

Figure 13: a) EO-Clinic SWF, b) JRC GSWE

3.1.2.1.5 Usage, Limitations and Constraints

The SWF product contains a detailed mapping of the annual and seasonal variations of surface water in and around Antananarivo, Madagascar.

This product serves to characterize the inter- and intra-annual variations of the water extent, and consequently to monitor the dynamics of water retention and flow, and to assess how these changes of water dynamics may affect the overall flooding regime. The product provides a good indication of exposed and vulnerable areas, and consequently enable hazard or risk maps to be determined. These maps are useful for identifying locations where urban expansion could be problematic, or areas where mitigation measures, such as NBS, may be useful.

As the algorithm is weighted towards optical predictions due to higher accuracies using optical data compared with SAR data, it may produce slight underestimations in water extent during cloudy periods (i.e. during the

rainy season). This may lead to a smaller bias in comparison to seasonal composites for less cloudy seasons (i.e. the dry season). The algorithm could also be applied without utilising the SAR exclusion layer, which would likely provide a larger water extent in each map, but with reduced overall accuracy.

A DEM is required to remove areas affected by shadow (urban, terrain, radar shadow etc.). The DEM has a spatial resolution of 30 m and may contain artefacts; therefore it does not necessarily capture the necessary topography required for 10 m products in complex terrains. Nevertheless, we believe this is a minor issue, and if higher resolution DEMs become accessible, they are easily integrable. Lastly, dense urban areas can cause several false positives due to their highly heterogenous nature, resulting in mixed pixels, as well as low albedo effects from building shadows and dark roads. Urban areas that are not affected by these issues can still be mapped effectively. Higher resolution commercial sensors (<10 m pixel resolution) have shown promising results in mapping these areas more accurately because they are less affected by mixed pixels.

3.1.2.2 Flood extent

Flood extent can be accurately mapped using both multispectral optical data as well as Synthetic Aperture Radar (SAR). In most cases, optical data will provide the most accurate results, however, it is limited by cloud cover and time of day, thus conditions need to be right. On the other hand, the strength of SAR data lies in its ability to penetrate clouds and provide imagery during night-time conditions. This ensures a continuous, gap free and consistent data source, even when the use of optical data is limited by unfavourable conditions. By using both optical and radar satellite data, the best of both methods is combined, to provide high precision flood extent mapping regardless of weather conditions, cloud cover and time of day. Flood maps were produced for three separate timesteps (January 11, 2020; January 14, 2020; and January 26, 2020) to capture the extent and dynamics of the January 2020 flooding's in Antananarivo.

The three flood maps are shown in Figure 14. The first (January 11, 2020) flood map was produced using optical Sentinel-2 data while the two latter maps were produced using Sentinel-1 SAR data, due to cloud coverage. The three flood maps documents well the growth of the flood extent from January 11 to 26, covering significantly larger land areas at the end of the month compared to the beginning. However, while the EO-based flood mapping works well in rural and peri-urban areas, the method reaches its limits in dense urban centres, where flood mapping is more difficult.

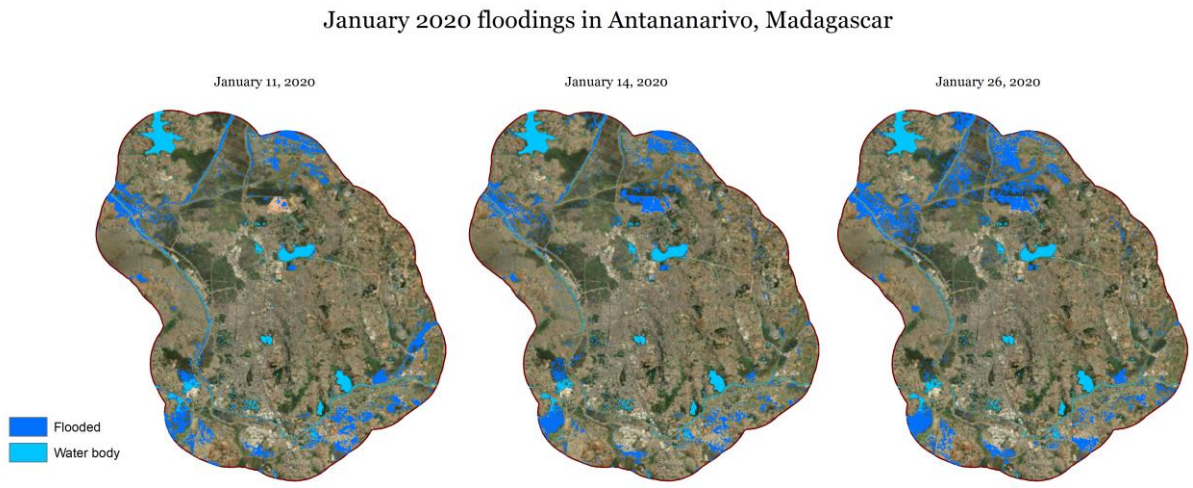

Figure 14: Flood extent maps for three different dates during the January 2020 flooding in Antananarivo, Madagascar

3.1.2.2.1 Specifications

Technical specifications of the flood extent product are summarized below.

EO input data: Sentinel-1, Sentinel-2

Other input data: Digital Elevation Model (30 m resolution), HAND

Method: Flood extent mapping using Sentinel-2 imagery is done by thresholding the Normalized difference water index (NDWI), which is created from the green and Near Infrared (NIR) wavelengths of Sentinel-2, to effectively separate water from non-water. Known sources of error, such as false positives, caused by clouds, cloud/terrain/urban shadow, and dense vegetation as well as turbidity (which can reduce separability of water/non-water) is removed or reduced, to the extent possible, in the pre-processing phase to ensure the most accurate result. Flood mapping using Sentinel-1 SAR imagery is done through a sequence of various pre-processing and processing steps. Initially, essential pre-processing of the Ground Range Detected (GRD) product includes GRD border noise removal, thermal noise removal, radiometric calibration, and terrain correction. Subsequently, a threshold is applied to the co-polarization VV band, to effectively separate water (returning very low backscatter values) from non-water (returning higher backscatter). The VV band is preferred as it is more sensitive to surface scatterers such as water bodies. However, other land characteristics such as sandy soils, flat impermeable surfaces such as roads and airports, and geometric distortions such as radar shadow, can also return similar values leading to errors. A change detection approach is applied, involving multiple images to reduce many of these errors, still with the assumption that water typically results in lower backscatter values, and the water is transient i.e., it must have changed from non-water to water in the observed period. Eventually, HAND data is used to limit false positives at higher altitudes, and thus improve the final flood extent product.

Output indicators: Flooded areas, permanent water bodies and land.

Map legend: Categorical classification (dry, permanent water and flooded areas).

Spatial resolution: 10 m pixel resolution.

Temporal resolution: January 11, 2020; January 14, 2020; January 26, 2020

Delivery format: GeoTiff, additional information or other data formats upon request; added to webportal

3.1.2.2.2 Quality Control and Validation

The quality control was done qualitatively since no ground data was available to validate the findings, meaning that all quality control and validation efforts are tied to visual inspection and evaluation of the data using the source imagery and pre flood imagery as reference. See example below in **Errore. L'origine riferimento non è stata trovata.**15.

Figure 15: QC approach for the flood extent product. Reference image to the left, source imagery used for mapping in the middle and flood map to the right

3.1.2.2.3 Usage, Limitations and Constraints

The flood extent product contains a detailed delineation of flooded areas in and around Antananarivo, Madagascar during the January 2020 flood event. Satellite-based flood maps provide a synoptic view of the impact of ongoing or historical flooding events, contributing key knowledge about changes in the extent of water bodies and occurrence of floods. This information contributes to a better understanding of the relationship between river discharge and flood extents in floodplains/wetlands/peri-urban areas, as well as the spatial identification of floodplains and flood hotspots. Such insight into the water body extent, hydrological regime and areas impacted by flooding's contributes to assess the potential for NBS activities and track and monitor the effect of these.

Cloud cover impedes flood mapping using optical data only, however, a multisensory approach using both optical and SAR data mitigates some of the limitations caused by this. However, particularly in urban areas, SAR based flood mapping is challenged by "noise" in the SAR signal, caused by buildings as well as land characteristics such as flat impermeable surfaces (e.g., roads and airports), and geometric distortions such as radar shadow, which all exhibit similar characteristics as water, leading to false positives (pixels wrongly classified as flooded). Most of these wrongly classified pixels can be corrected during pre-processing, however, some might persist. Furthermore, as a side looking sensor, SAR imagery is taken at an angle, thus water bodies can be "hidden" by buildings and building shadows. Additionally, flooded areas underneath trees and dense vegetation cannot be mapped using either optical Sentinel-2 or Sentinel-1 SAR data.

3.1.2.3 Terrain characteristics

The terrain characteristics of a region, as characterized by its topography and soil characteristics, have a strong impact on surface water dynamics. Steep slopes with low infiltration rate facilitate surface runoff to lower regions, where the water may either infiltrate if the permeability of the surface allows or continue as surface runoff. The topography of a region can be determined using a digital elevation model (DEM), which is a pixel-based modelled representation of the Earth's surface, where each pixel represents an elevation value. Soil composition and characteristics is derived from maps based on interpolations between in situ measurements. The highest spatial resolution, free and open and globally available DEM comes with a 30 m pixel resolution, and the most accurate solution is provided through the Copernicus DEM GLO-30. The Copernicus DEM GLO-30 can be used to calculate additional terrain characteristics including slope, Topographic Wetness Index (TWI) and Height Above the Nearest Drainage (HAND) which all provides useful information on the terrain and its characteristics (further explained in section 3.1.2.3.3). Global soil data is produced by the Food and Agriculture Organization of the United Nations (FAO) and publicly available as the Harmonized World Soil Database (HWSD 1.21). HWSD contains information about various soil characteristics at 30 arc-second spatial resolution. Combined, the topographic and soil layers provide an overview of which areas are likely to experience flooding during precipitation events. The DEM GLO-30 was sourced for Antananarivo to provide information layers on elevation, slope, TWI and HAND, while HWSD data was sourced to provide information layers on major soil group, available water storage capacity and drainage group.

The seven datasets described above are shown in Figure 16. The four topographic layers provide detailed insight into the topographic characteristics of the terrain and the hydrological processes, while the HWSD datasets provide coarse level insight into the soil characteristics, water storage capacity and drainage.

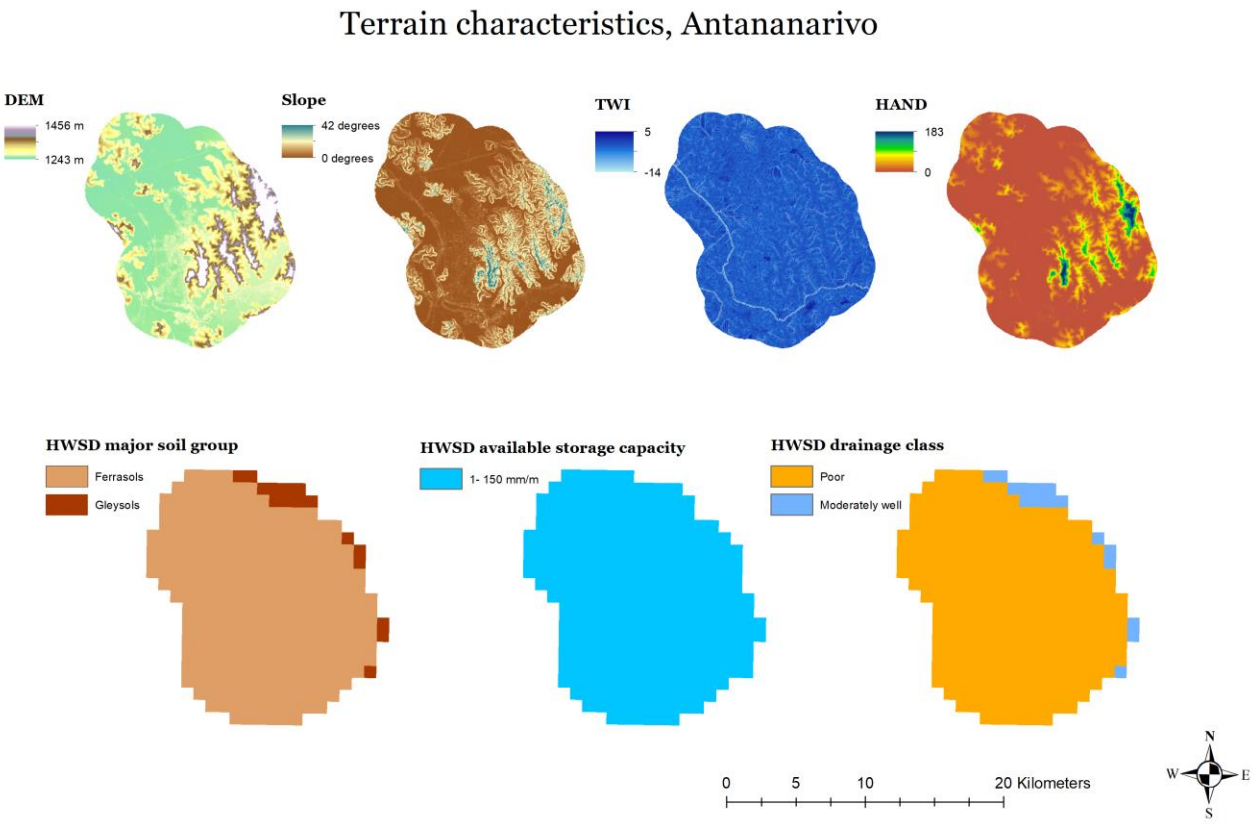

Figure 16: Terrain characteristics for Antananarivo

3.1.2.3.1 Specifications

Technical specifications for the terrain characteristic products are summarized below.

EO input data: Copernicus GLO-30 and HWSD 1.21

Method: See the Copernicus DEM Product Handbook (*Section 3 - Copernicus DEM: Product Generation and Editing Process*) for information on the processing steps of the DEM GLO-30. See the HWSD 1.21 product documentation for information about the compilation of the HWSD datasets.

The **Slope** is processed from the DEM GLO-30 elevation model. The slope is calculated using the maximum rate of change in the horizontal (dz/dx) and vertical (dz/dy) directions from one cell to its neighbours.

The **TWI** has been produced from the slope, flow direction, flow accumulation, tan of slope, and eventually TWI as TWI=ln(A/tanβ) where A is the local upslope contributing area and tanβ is the local slope. The contributing area known as the upslope area or flow accumulation, determines the size of the upslope area draining into a cell. The flow direction is used to understand how the water flows on the land surface. The direction of flow is determined by the direction of steepest descent, or maximum drop, from each cell. Subsequently, the flow accumulation is determined by how many cells flow through that cell using the flow direction raster. Areas prone to water accumulation and characterized by low slope angle are

linked to high TWI values while, well-drained dry areas are associated to low TWI values.

The computation of the **HAND** index requires initial pre-processing of the DEM data. First, the Copernicus GLO-30 dataset was hydrologically corrected to remove artifacts, topographic depressions, and flat areas. A second set of processing steps was applied to get local drain directions and drainage network, to define flow paths, and delineate the drainage channel. The drainage network is then converted into a normalized topographic reference so that the HAND model does not hold a reference to sea level. Each cell of the final HAND model represents the difference in level to its respective nearest drainage cell.

Output indicators: Elevation, slope, TWI, HAND, HWSD major soil group, HWSD available storage capacity and HWSD drainage class.

Spatial resolution: Elevation and derived products (30 m) and HWSD data products (1 km).

Temporal resolution: Copernicus GLO-30: 2010-2015; HWSD: composed from four different source databases described in section 2.1 in the product documentation

Delivery format: GeoTiff, additional information or other data formats upon request; added to webportal

3.1.2.3.2 Quality Control and Validation

Further information about GLO-30 DEM product quality and accuracy is provided in Section 2 (*Copernicus DEM: Product Quality Performance*) in the Copernicus DEM Product Handbook. The derived DEM products (Slope, TWI and HAND) have all been produced using standardized approaches and the results have all been manually quality controlled by technical experts with expertise in DEMs and derived parameters for hydrological modelling.

3.1.2.3.3 Usage, Limitations and Constraints

The DEM provides information on topography and elevation levels, thus can be used to define the pathway of potential flooding's in both the upstream catchment (river channel and floodplain) as well as urban areas (complex urban topography of buildings and streets). Slope is used to infer direction and estimate velocity of surface runoff. TWI provides information on the effect of local topography on hydrological processes and can be used to locate areas which are susceptible to local flooding events. HAND provides additional information about the vertical distance between a location and its nearest stream surface or bed, thus indicating flood vulnerability and potential. Soil characteristics, such as texture and bulk density, can be used to estimate infiltration rates around the urban centre, where the surface is permeable.

The GLO-30 DEM is one of the best and most accurate DEMs currently available, with free and open data access. Although the spatial resolution of 30 m is considerably higher than most global level alternatives, it can still be considered coarse for very localized studies, such as application in smaller urban areas or individual districts in larger cities, where highly granular information is required, e.g., at building level, smaller patches of green infrastructure and individual roads. At city level, however, GLO-30 provides sufficient details for large-scale screening of NBS potential. The same applies for the HWSD data, which has a spatial resolution of 30 arc-second (approximately 1 km). HWSD data is useful for larger screening analyses at city level, but unfit for localized studies requiring higher levels of detail.

For further information on Copernicus GLO-30 see the Product Handbook and for further information on HWSD see the product documentation.

3.1.3 Data layers for urban heat

The Urban heat island (UHI) effect is a well-known phenomenon allusive to the temperature rise experienced in many urbanized areas. The heat island phenomenon has been commonly associated to cities because their

surfaces are characterized by low albedo, high impermeability and favourable thermal properties for the energy storage and heat release. Besides, many cities have narrow urban canyons with reduced sky view that tend to absorb and reemit the radiated energy from their surfaces. These factors contribute to higher temperatures in urbanized areas compares to rural suburbs which are usually more vegetated, and therefore moderate the temperatures mainly through evapotranspiration, shading, and solar radiation interception.

3.1.3.1 Land Surface Temperature

The most commonly used indicator to describe and categorize Urban Heat Islands is land surface temperature (LST), and LST variability between green areas, water bodies and impervious surfaces. Such variability provides an indication of heat pockets in urban areas as well as the effect of green infrastructure. In order to evaluate the potential for Nature Based Solutions and UHI mitigation measures (such as urban greening, water features, etc.), LST data is needed in sufficient spatial resolution to capture major urban landscape features, such as parks, street blocks, plazas, etc. Currently, the only available satellite-based data which fulfils this condition are the thermal observations by NASA's Landsat satellites however future thermal missions, like the Copernicus Land Surface Temperature Monitoring (LSTM) mission from ESA, are also planned to provide high resolution LST in the near future. LST data is produced by the United States Geological Survey (USGS) based on Landsat-7 and Landsat-8 thermal observations and delivered in a 30 m spatial resolution and 8-day temporal resolution.

Bimonthly LST composites (for late spring, summer and early autumn) from 2015 to 2021 were produced from Landsat data to assess the potential impacts and effects of UHI in Tashkent, Uzbekistan.

An example LST product from the months July/August 2018 is shown in Figure 17. The data illustrates the temperature dynamics and variations as reflected by the composition of the urban landscape. As indicated in the data points, the highest temperatures are experienced in dense industrial areas, individual production facilities as well as the international airport. The lower left extract indicates the temperature difference between the neighbouring Chilanzar district and the densely populated Shaykhontohur district. The density of urban trees, the green infrastructure and the spacing between buildings in the Chilanzar district contributes to significantly reduce temperatures compared to the denser urban landscape with fewer trees and green areas in the Shaykhontohur district. The lowest temperatures are found outside the core urban area in the agricultural areas south of the city, while the highest are found in dense industrial areas.

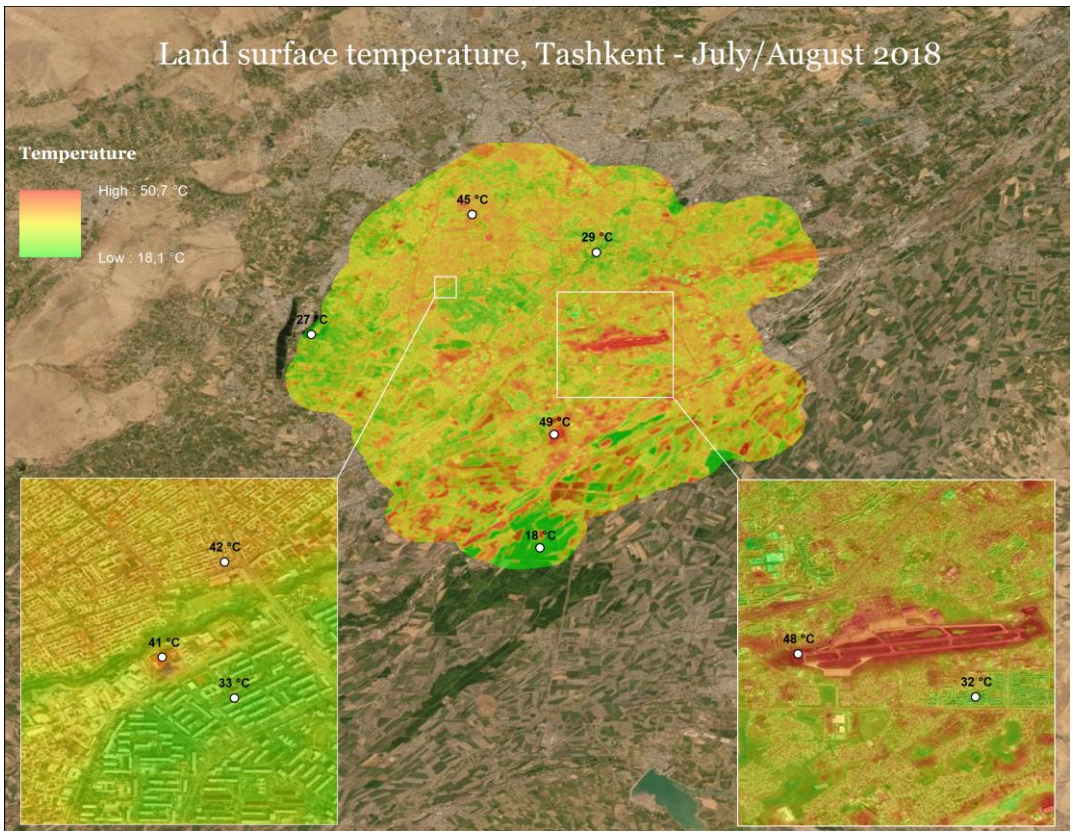

Figure 17: Example LST and UHI indicators in Tashkent during July/August 2018

3.1.3.1.1 Specifications

Technical specifications for the LST product are summarized below.

EO input data: Landsat 7 and Landsat 8

Method: For information about the Landsat LST product see the Landsat Level-2 Surface Temperature Science Product courtesy of the U.S. Geological Survey.

Additional postprocessing of the data was conducted to remove data gaps, primarily caused by cloud cover. To address this issue bimonthly multitemporal composites were created to produce cloud free and coherent datasets. Subsequently, the dataset is checked for additional gaps, and larger potentially remaining gaps in the Landsat LST composites can be filled using analogous timeseries of highly temporal data (like Sentinel-3 based LST). The compositing was performed using maximum LST criteria, meaning that the highest LST value recorded during the period was selected from the time-series. This is preferred when assessing UHI, where insight into the range and differences between peak temperatures is needed. Finally, the lowest 1% and highest 0.1% of values are removed and gap-filled based on neighbouring pixels to minimize artifacts introduced by undetected clouds and fires.

Output indicators: Land surface temperature

Spatial resolution: 30 m

Temporal resolution: Bimonthly (for late spring, summer, and early autumn) from 2015 to 2021

Delivery format: GeoTiff, additional information or other data formats upon request; added to webportal

3.1.3.1.2 Quality Control and Validation

For more information about the Landsat LST product, see the documentation documents at https://www.usgs.gov/core-science-systems/nli/landsat/landsat-collection-2-surface-temperature.

The bimonthly composites were manually quality controlled by an experienced technician to assess data coherency and overall product quality.

3.1.3.1.3 Usage, Limitations and Constraints

High resolution LST data is used in conjunction with high resolution land cover maps to describe temperature variation and dynamics in urban settings. The data is used to interpret how the urban landscape affects temperature and temperature variation, such as monitoring the impact of dense urban structures and impervious surfaces on temperature extremes. After NBS activities, such as green roofing or vegetation corridors, LST data can be used to measure the overall impact in terms of temperature reductions on the surrounding urban neighbourhoods.

Currently, satellite derived LST-data is available in either high resolution (30 m)/lower temporality (8 days) or low resolution (1 km.)/high temporality (daily). The former is made available through Landsat 7+8, while the latter is provided through e.g., the Sentinel 3 and Terra and Aqua satellites. To study UHI, high resolution LST data is needed, however, currently only the Landsat satellites provides LST data in resolutions below 100m, and even at 30 m resolution (such as Landsat derived LST), subtle differences and discrete changes in the urban landscape might not be detectable. Furthermore, with a temporal resolution of 8 days, cloud cover is an issue and may limit opportunities to detect trends and changes within seasons and month to month, depending on cloud conditions. Highly temporal resolution LST data, such as from Sentinel 3, can be used as a supplement to address this issue, however at 1 km spatial resolution, it can only be used to assess coarse overall patterns at broader city level. Additionally, most thermal satellites overpass around 10-11 am, meaning

that assumed peak temperatures in the afternoon are not captured. Lastly, Landsat 7 derived LST data occasionally has stripe gaps caused by a failure in the satellite's Scan Line Corrector. This issue will be resolved later in 2021 with the launch of the new Landsat 9 satellite. Additional limitations are described in Section 3 (*Caveats and Constraints*) in the Landsat 8 Collection 2 Level 2 Science Product Guide

3.1.4 Output delivery

3.1.4.1 WEBGIS publication

As a final deliverable a web viewer has been delivered for easy display of the data and information products produced under this procurement (cf. Figure 18).

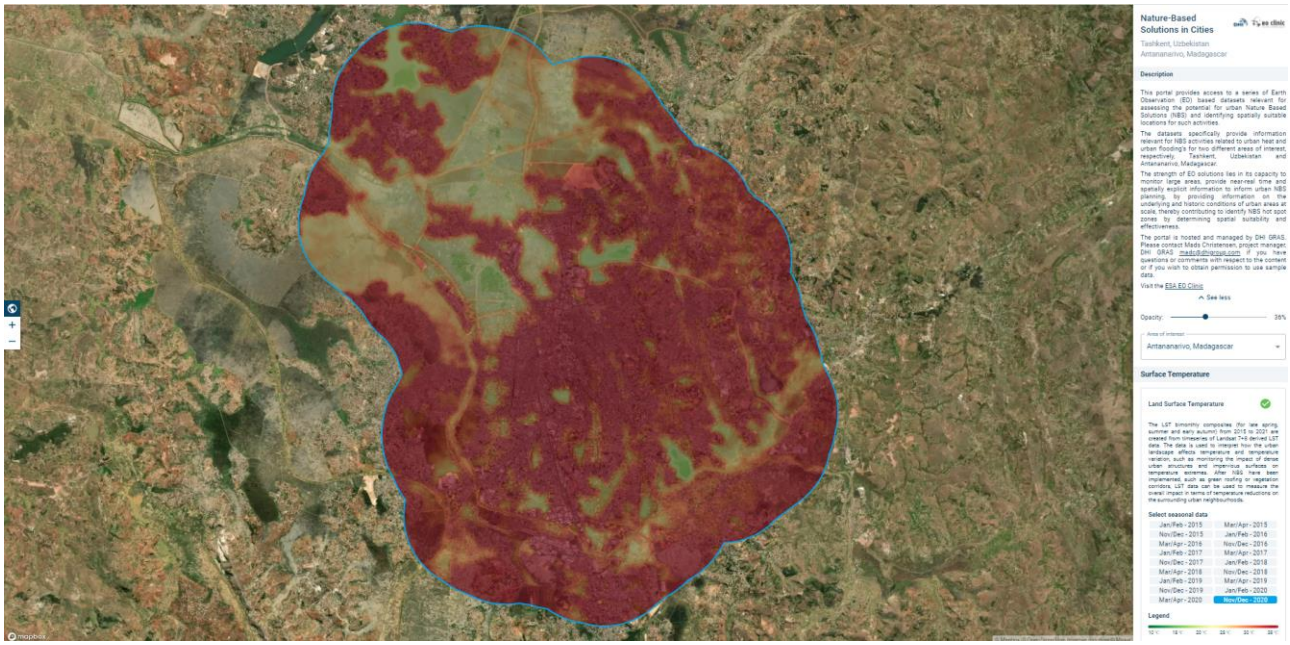

Figure 18: Screen dump of the web viewer for publishing the results.

The web viewer can be accessed by clicking the following link: https://labs.dhi-gras.com/eo-clinic-nbs

APPENDIX A: BIBLIOGRAPHY

Asian Development Bank. Nature-based solutions for building resilience in towns and cities: Case studies from the Greater Mekong Subregion. Mandaluyong City, Philippines: Asian Development Bank, 2016

Muñoz Sabater, J., (2019): ERA5-Land hourly data from 1981 to present. Copernicus Climate Change Service (C3S) Climate Data Store (CDS). (Accessed on < 30-06-2021 >), 10.24381/cds.e2161bac

Pekel, J. F., Cottam, A., Gorelick, N., & Belward, A. S. (2016). High-resolution mapping of global surface water and its long-term changes. Nature, 540(7633), 418-422.

UN Environment-DHI, UN Environment and IUCN 2018. Nature-Based Solutions for Water Management: A Primer