



CLIMATE ANALYSIS IN AFRICAN CITIES (CAIAC)

D1.3 Inventory of relevant datasets and tools

Study accomplished under the authority of the European Space Agency

Reference: 2025/EI/R/3452

September 2025

D1.3 Inventory of relevant datasets and tools

VITO

Boeretang 200

2400 MOL

Belgium

VAT No: BE0244.195.916

vito@vito.be – www.vito.be

IBAN BE34 3751 1173 5490 BBRUBEBB

Nele Veldeman

Project Manager

nele.veldeman@vito.be

Koen De Ridder

Science Lead

koen.deridder@vito.be

AUTHORS

VITO

Benjamin Lanssens

Tomas Crols

Koen De Ridder

Jonathan Leon Tavares

Niels Souverijns

Inge Uljee

Nele Veldeman

University of Ottawa

Mohammad Alobaidi

Hossein Bonakdari

Ousmane Seidou

DISTRIBUTION LIST

Clement Albergel, ESA

SUMMARY

Deliverable D1.3 compiles a comprehensive inventory of Earth Observation (EO) datasets, atmospheric and hydrological records, and modelling frameworks to support the Climate Analysis in African Cities (CAIAC) project. The primary objective is to enable high-resolution (300 m) projections of extreme heat and flood risk across 54 African cities, covering both the historical baseline and future horizons under CMIP6 climate scenarios.

The report catalogues major EO sources such as ESA CCI Land Cover, WorldCover, the Global Human Settlement Layer, Sentinel-1 flood monitoring, Sentinel-2 NDVI, and climate variables including land surface temperature, soil moisture, and river discharge. These are complemented by socio-economic and ground-based datasets from ACMAD and national meteorological services. Together, they provide the empirical basis for modelling urban climate dynamics, land-use change, and hydrological hazards.

Three core modelling tools form the project's backbone: UrbClim, for simulating urban climate and heat stress; GeoDynamix, for land-use and population growth projections; and SAHEL, a machine-learning framework designed to assess flood risks by integrating EO and hydrological inputs. These models are supported by dedicated processing pipelines that harmonise multiscale datasets and ensure readiness for simulation and AI workflows.

Finally, the document highlights the research infrastructure underpinning the work: VITO's high-performance computing facilities and uOttawa's geospatial and machine-learning resources. By consolidating all available resources, D1.3 provides the foundation for scientifically robust analyses of climate impacts in African urban environments.

TABLE OF CONTENTS

Authors	1
Distribution list.....	2
Summary	3
Table of contents	4
List of acronyms.....	6
1 Introduction	8
2 EO data	9
2.1 ESA CCI Land Cover	9
2.2 ESA WorldCover.....	10
2.3 Global Human Settlement Layer (GHSL) – Built-Up & Population.....	11
2.4 Sentinel-1 Global Flood Monitoring	12
2.5 Sentinel-2 NDVI	14
2.6 ESA CCI Land Surface Temperature - L3S.....	15
2.7 ESA CCI Soil Moisture	16
2.8 ESA CCI River Discharge	17
2.9 Copernicus Global Digital Elevation Model	19
2.10 Global Local Climate Zones (LCZ) Map	20
2.11 Large-scale Slum Probability Maps.....	21
3 Atmospheric (re-analysis / climate) data	23
3.1 ACMAD In-Situ Meteorological Data (Africa).....	23
3.2 ERA5 Atmospheric Reanalysis	24
3.3 CMIP6 Climate Projections	25
4 Modelling tools	28
4.1 UrbClim.....	28
4.2 GeoDynamix	29
4.3 SAHEL.....	30
5 Processing tools.....	33
5.1 Input for urban growth / climate models	33
5.2 Input for flooding model	36
6 Research infrastructure	38
6.1 VITO	38
6.2 uOttawa	39

7	Conclusion	40
	References	41

LIST OF ACRONYMS

AATSR	Advanced Along-Track Scanning Radiometer
ACMAD	African Centre of Meteorological Applications for Development
AHF	Anthropogenic Heat Flux
AOI	Area Of Interest
ATSR-2	Along-Track Scanning Radiometer-2
AVHRR	Advanced Very High Resolution Radiometer
CCI	Climate Change Initiative
CDSE	Copernicus Data Space Ecosystem
CEMS	Copernicus Emergency Management Service
CEOS	Committee on Earth Observation Satellites
CMIP6	Coupled Model Intercomparison Project Phase 6
CORDEX	Coordinated Regional Downscaling Experiment
CSV	Comma-Separated Values
C3S	Copernicus Climate Change Service
DEM	Digital Elevation Model
DHS	Demographic and Health Surveys
DSM	Digital Surface Model
ECMWF	European Centre for Medium-Range Weather Forecasts
EO	Earth-Observation
ERA5	Fifth-generation ECMWF atmospheric reanalysis
ESA	European Space Agency
ETH	Swiss Federal Institute of Technology
EVI	Enhanced Vegetation Index
FAO	Food and Agriculture Organization
GEDI	Global Ecosystem Dynamics Investigation
GFM	Global Flood Monitoring
GHSL	Global Human Settlement Layer

GHS-BUILT-S	GHSL Built-up Surface
GloFAS	Global Flood Awareness System
GSHHS	Global Self-consistent, Hierarchical, High-resolution Geography Database
GTS	Global Telecommunication System
IFS	Integrated Forecast System
JRC	Joint Research Centre
LCCS	Land Cover Classification System
LCZ	Local Climate Zone
LPV	Land Product Validation
LST	Land Surface Temperature
MERIS	Medium Resolution Imaging Spectrometer
MODIS	Moderate Resolution Imaging Spectroradiometer
MSI	Multispectral Instrument
NDVI	Normalised Difference Vegetation Index
NRT	Near Real-Time
OSM	OpenStreetMap
PSU	Primary Sampling Unit
SAR	Synthetic Aperture Radar
SCL	Scene Classification Layer
SLSTR	Sea and Land Surface Temperature Radiometer
SWIR	Short-Wave Infrared
SSP	Shared Socioeconomic Pathway
UTM	Universal Transverse Mercator
VRT	Virtual Raster (Tile)
WCRP	World Climate Research Programme
WGS 84	World Geodetic System 1984 (EPSG:4326)
WMO	World Meteorological Organization

1 INTRODUCTION

The D1.3 Inventory Document is an essential step in the CAIAC project. It brings together the key datasets, models, and processing tools that will be used to study climate and flood risks in African cities. This deliverable turns scientific priorities into a practical overview of resources that will guide the next phases of the project.

The focus on Africa is both urgent and timely. Cities on the continent are growing fast and facing increasing threats from extreme heat and flooding, while reliable climate data at the city scale are often missing. Earth Observation (EO) data therefore play a central role: ESA's Essential Climate Variables, Copernicus products, and other satellite sources provide the basis for model inputs, for checking results, and for assessing uncertainties. In addition, ground-based observations provided by ACMAD and hydrological or socio-economic datasets will help to strengthen flood and urban growth modelling.

The inventory also describes the tools that will be used. UrbClim, GeoDynamix, and SAHEL form the core modelling framework, while dedicated EO processing methods make it possible to turn raw satellite observations into usable data layers. Taken together, this document gives all partners a clear and shared view of the resources available, ensuring that the project can move forward in a consistent way.

2 EO DATA

2.1 ESA CCI Land Cover

This dataset provides global maps describing the land surface into 22 classes, which have been defined using the United Nations Food and Agriculture Organization's (UN FAO) Land Cover Classification System (LCCS).

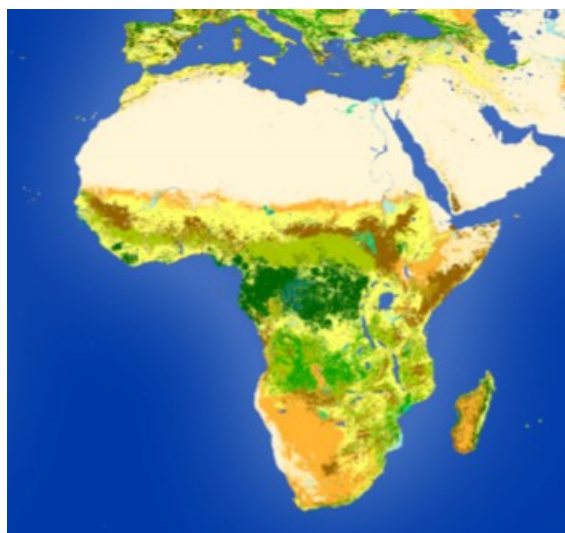


Figure 1: Global ESA CCI Land Cover map (year 2015). Colours correspond to FAO LCCS land cover categories. Source: CEOS (<https://ceos.org/gst/cci-lc.html>).

2.1.1 Core metadata

Source	ESA Climate Change Initiative
Spatial resolution	300 m (0.002778°)
Native CRS	WGS 84 (EPSG: 4326)
Temporal coverage	Yearly 1992-2022
Update frequency	Yearly (T-1 year release lag)
Sensor lineage	Classification chain uses a MERIS FR/RR baseline map (2003–2012) combined with change detection from AVHRR, SPOT-VGT, PROBA-V and Sentinel-3 time series
Variables	Map (22 LCCS classes) plus four quality flags: (1) pixel processed, (2) pre-processing status, (3) number of valid observations, (4) number of Land Cover changes
File format	NetCDF (global), optional GeoTIFF per year
Access / DOI	https://climate.esa.int/en/projects/land-cover/data/
Licence	CC-BY 4.0

2.1.2 Validation & Uncertainty

Validation follows the CEOS LPV protocol using a two-stage stratified sample (~13 000 plots). For the 2001-2020 window, global overall accuracy ranges 72–75 %. User's Accuracy by class: forests > 80 %, croplands ≈ 70 % and urban ≈ 55 %. Pixels with confidence level below 40 % should be treated as low reliability. Because 2001-2002 layers inherit 1 km spatial detail, it should be avoided to make sub-kilometer interpretations in those years.

2.2 ESA WorldCover

The WorldCover product delivers global land-cover maps at 10 m resolution for 2020 and 2021. It uses a classification chain driven by Sentinel-1 C-band radar and Sentinel-2 multispectral imagery to assign each pixel to one of 11 UN-LCCS classes and provides a per-pixel quality score.

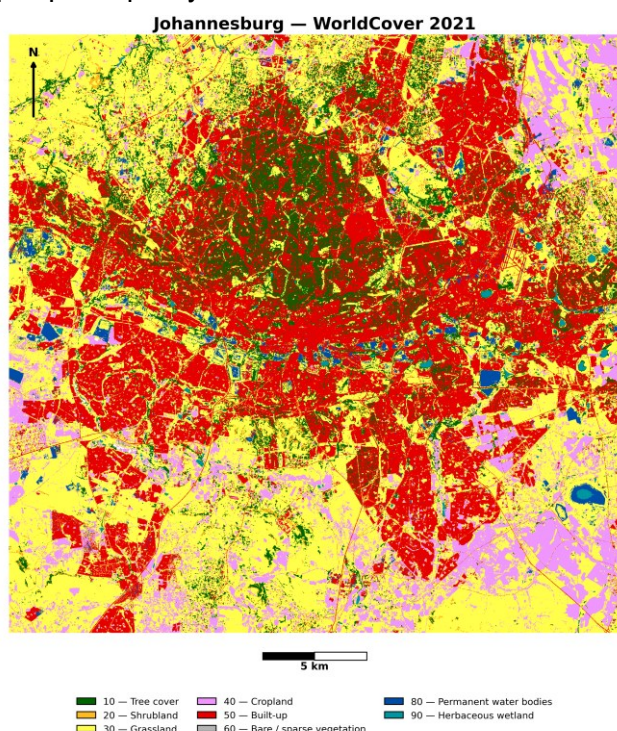


Figure 2: ESA WorldCover map of Johannesburg (2021).

2.2.1 Core metadata

Source	ESA; Copernicus Programme
Spatial resolution	10 m ($\approx 0.00009^\circ$)
Native CRS	WGS 84 (EPSG: 4326)
Temporal coverage	Yearly 2020-2021
Update frequency	Annual releases; current versions v100 (2020) and v200 (2021)

Sensor lineage	Sentinel-1 (IW-mode, VV polarisation) and Sentinel-2 MSI
Variables	Map (11 LCCS classes) and InputQuality score (0–100)
File format	GeoTIFF (3° × 3° tiles)
Access / DOI	https://worldcover2020.esa.int/
Licence	CC-BY 4.0

2.2.2 Validation & Uncertainty

For the 2020 map, the global overall accuracy is 74.4 ± 0.1 %. Accuracies exceed 80 % for tree cover, snow/ice, cropland, water and bare classes, but are lower for shrubs and herbaceous wetlands. Accuracy varies by continent, from ~81 % in Asia to ~67 % in Oceania. The InputQuality layer provides a simple quality indicator: values below 50 % should be treated cautiously. It should be noticed that narrow water bodies and small built-up areas may be missed at 10 m resolution. No uncertainty quantification is provided for the 2021 map.

2.3 Global Human Settlement Layer (GHSL) – Built-Up & Population

The Global Human Settlement Layer (GHSL) is an initiative of the European Commission's Joint Research Centre to provide consistent global information on human presence. It includes built-up proportion maps, population grids and settlement models.

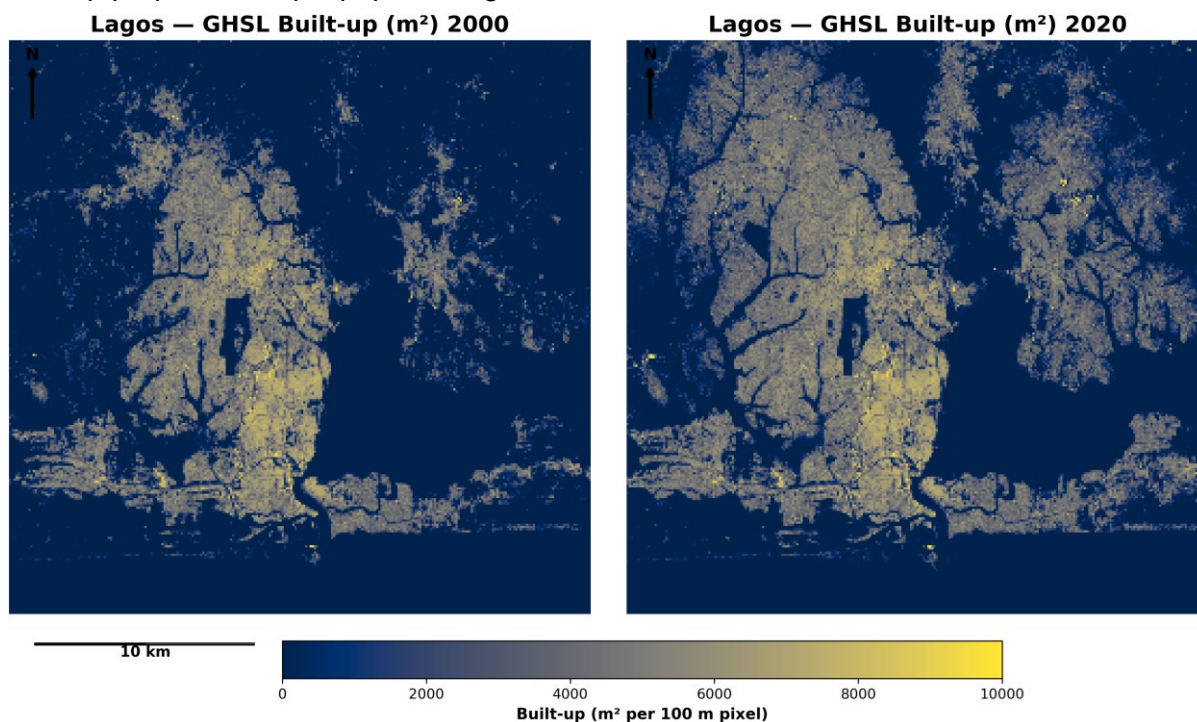


Figure 3: Copernicus Global Human Settlement Layer (GHSL) built-up surface (Lagos, 2000 & 2020).

2.3.1 Core metadata

Source	European Commission; Joint Research Centre (JRC)
Spatial resolution	Built-up proportion at 10 m; population at 100 m
Native CRS	UTM zones per tile for built-up product; WGS-84 geographic grid for population
Temporal coverage	Built-up: reference year 2018 (Sentinel-2) and 1975/1990/2000/2014 (Landsat); population: 1975-2020
Update frequency	Discrete releases
Sensor lineage	Sentinel-2 MSI for built-up proportion; Landsat MSS/ETM+ for historical built-up; census and ancillary data for population
Variables	Built-up proportion (0–100); population counts; additional settlement classes (urban center, urban cluster, rural)
File format	GeoTIFF tiles for built-up; GeoTIFF/NetCDF for population
Access / DOI	https://ghsl.jrc.ec.europa.eu/
Licence	Free for non-commercial use; Copernicus data licence

2.3.2 Validation & Uncertainty

The GHSL team reports that the Sentinel-2 built-up product performs comparably or better than existing global datasets for 2018. Nonetheless, there is no official global accuracy figure. For the Landsat-based built-up products, evaluation suggests good agreement with other datasets for 2018 and 2020 but a positive bias in built-up change rates after 2000.

2.4 Sentinel-1 Global Flood Monitoring

The Copernicus Global Flood Monitoring (GFM) service provides near-real-time maps of surface water and flooding derived from Sentinel-1 Synthetic Aperture Radar (SAR) imagery. It is part of the Global Flood Awareness System (GloFAS) and processes all incoming Sentinel-1 acquisitions to detect water extent and flood events worldwide.

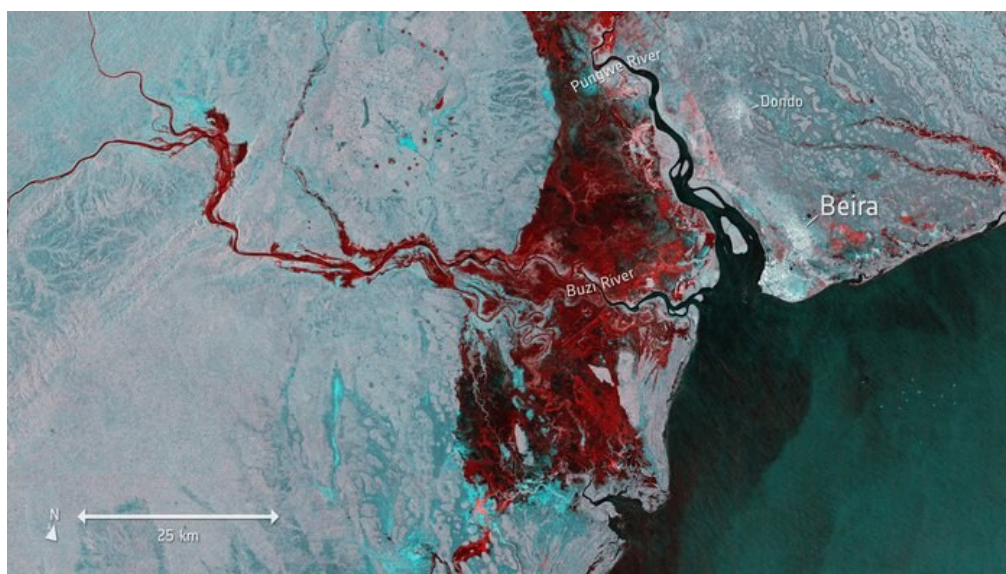


Figure 4: Image from Copernicus Sentinel-1 (mars 2019) shows the extent of flooding around the port town of Beira (Mozambique) after the Cyclone Idai. Source: ESA/CC BY-SA 3.0 IGO, via esa.int.

2.4.1 Core metadata

Source	Copernicus Emergency Management Service (CEMS); JRC
Spatial resolution	≈ 20 m
Native CRS	WGS 84 (EPSG: 4326)
Temporal coverage	Continuous since 2021; historical archive from Sentinel-1 (2014)
Update frequency	Near-real-time: products available within hours of acquisition; global coverage every 6–12 days depending on Sentinel-1 pass
Sensor lineage	Sentinel-1 C-band SAR in Interferometric Wide-swath mode; VV or VH polarisation
Variables	Observed flood extent (binary flood mask); observed water extent; reference water mask; exclusion mask; uncertainty values (0-100); ancillary data.
File format	GeoTIFF and vector shapefiles
Access / DOI	Copernicus Emergency Management Service
Licence	Copernicus free and open data licence

2.4.2 Validation & Uncertainty

No global accuracy statistics have been published. Errors may arise from radar layover, shadow and speckle noise, leading to omission or commission errors especially in mountainous areas and densely vegetated floodplains. It is advised to cross-validate GFM maps with optical imagery or ground reports where possible. Because the product is binary,

no per-pixel probability or uncertainty layer is provided; however, the consensus approach reduces algorithm-specific biases.

2.5 Sentinel-2 NDVI

Vegetation indices such as the Normalised Difference Vegetation Index (NDVI) can be computed from the red and near-infrared bands of Sentinel-2 twin-satellite constellation sensors to monitor vegetation status and dynamics.

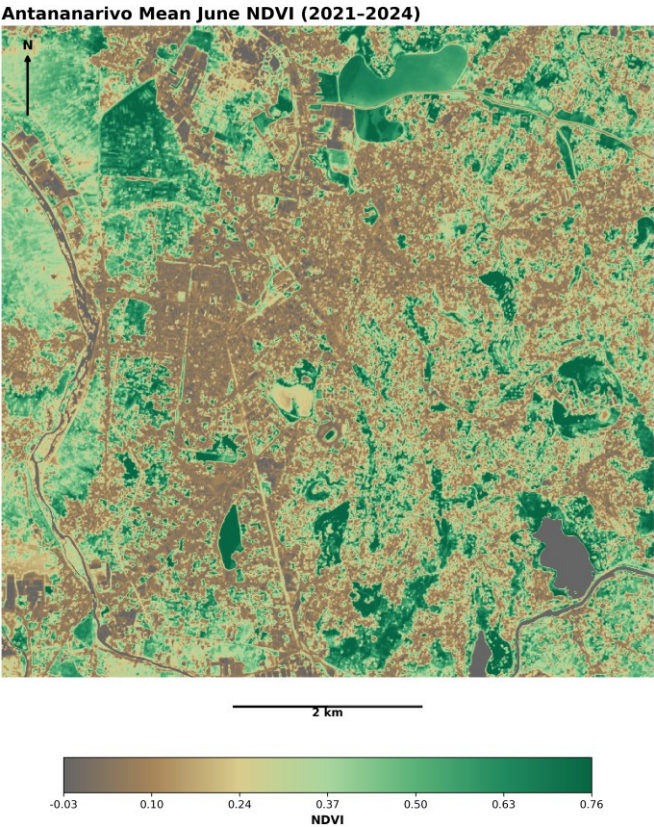


Figure 5: Normalized Difference Vegetation Index (NDVI) map for the central area of Antananarivo, averaged for June (2021–2024).

2.5.1 Core metadata

Source	ESA; Copernicus Programme
Spatial resolution	10 m for VIS and NIR bands; 20 m for red-edge and SWIR bands
Native CRS	UTM zone grid per tile
Temporal coverage	Continuous since 2021
Update frequency	Level-2A global archive since December 2018
Sensor lineage	MSI on Sentinel-2A/B: 13 spectral bands from 443 to 2190 nm
Variables	Surface reflectance for 13 bands; $NDVI = (NIR - red)/(NIR + red)$

File format	JP2 tiles packaged in SAFE products
Access / DOI	https://dataspace.copernicus.eu/
Licence	Copernicus free and open data licence

2.5.2 Validation & Uncertainty

The absolute radiometric calibration accuracy is < 3 %. NDVI is sensitive to atmospheric correction residuals, viewing geometry and canopy structure. Cloud contamination is the primary source of error. NDVI also saturates in dense vegetation (NDVI > 0.8), so alternative indices such as EVI or red-edge-based metrics may offer improved sensitivity in tropical Africa region.

2.6 ESA CCI Land Surface Temperature - L3S

The ESA Climate Change Initiative Land Surface Temperature (LST) project provides a long, consistent record of land surface temperature. In its Level-3 Supercollated version (L3S), the dataset offers daily day-time and night-time measurements by merging multiple low Earth orbit infrared sensors.

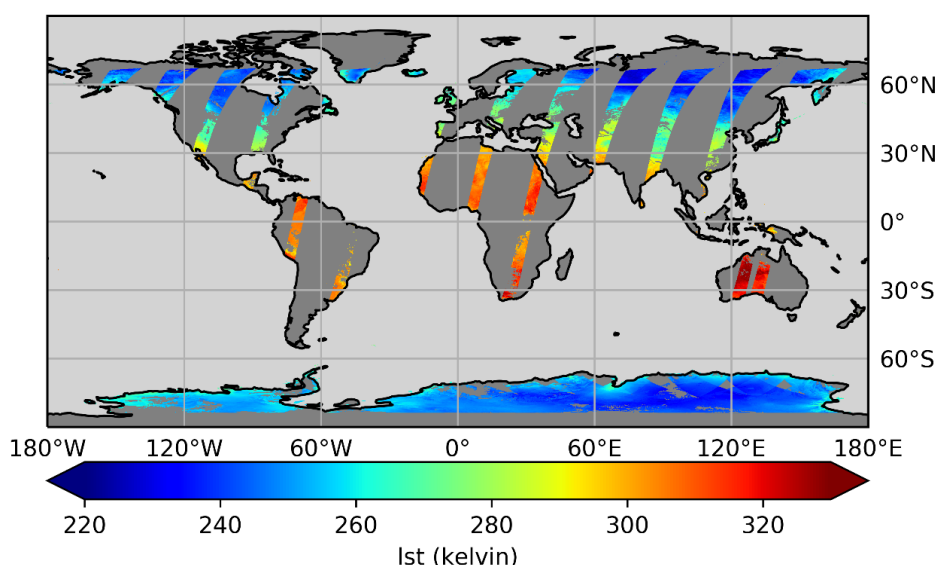


Figure 6: Example of MULTISENSOR_IRCDR_L3S_0.01 daily daytime data for 1st January 2010. Source: LST-CCI Product User Guide (v2.0).

2.6.1 Core metadata

Source	ESA Climate Change Initiative
Spatial resolution	0.01° × 0.01° (~1 km at the equator)
Native CRS	WGS 84 (EPSG: 4326)
Temporal coverage	Daily. 1995-2020

Update frequency	Twice daily
Sensor lineage	Merged infrared from ATSR-2, AATSR, MODIS and SLSTR-A
Variables	Land Surface Temperature (K); uncertainty estimates; observation time; viewing & solar geometry angles; quality flags; land cover class.
File format	NetCDF
Access / DOI	CEDA catalogue
Licence	Open data; CC-BY 4.0

2.6.2 Validation & Uncertainty

Validation against ground-based radiometers shows that LST meets the target 1 K accuracy both night and day. Each pixel is accompanied by quality flags and uncertainty estimates reflecting retrieval noise and merging errors. Infrared retrievals are cloud-limited and sensitive to atmospheric dust and aerosols, which can affect retrievals over parts of Africa.

2.7 ESA CCI Soil Moisture

The Soil Moisture product merges passive and active microwave observations from numerous satellites to produce a long-term, globally consistent record of surface soil moisture.

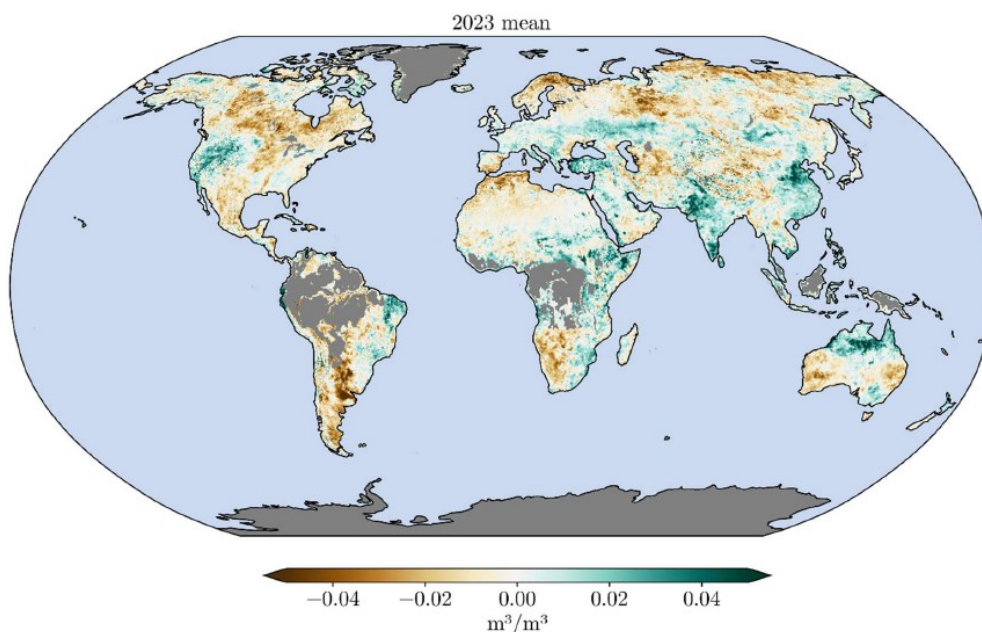


Figure 7: Mean Soil Moisture anomaly for 2023, with respect to the 1991-2020 baseline. Source: SM-CCI Product User Guide (v09.1).

2.7.1 Core metadata

Source	ESA Climate Change Initiative
--------	-------------------------------

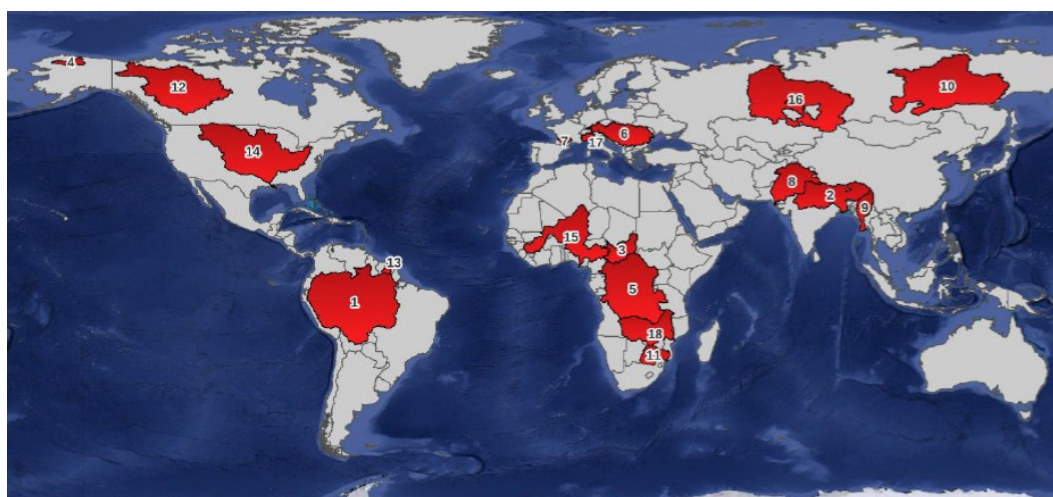
Spatial resolution	0.25° (~25 km)
Native CRS	WGS 84 (EPSG: 4326)
Temporal coverage	Daily 1978-2023
Update frequency	Climate data record updated annually; near-real-time extension available
Sensor lineage	Merged passive microwave radiometers (SMMR, SSM/I, TMI, AMSR-E, WindSat, AMSR2, SMOS, SMAP) and active scatterometers (ASCAT) using triple-collocation
Variables	Surface volumetric soil moisture ($\text{m}^3 \text{m}^{-3}$); per-sensor data sets; ancillary flags (frozen soil, snow, vegetation, desert)
File format	NetCDF
Access / DOI	CEDA catalogue
Licence	Open data; CC-BY 4.0

2.7.2 Validation & Uncertainty

Validation and intercomparison against in-situ networks show that uncertainty has decreased over time and is below $0.01 \text{ m}^3 \text{m}^{-3}$ in recent years. Retrievals are less reliable over dense vegetation, frozen or snow-covered ground and deserts; flags indicate these conditions. The product represents only the top few centimeters of soil and may not capture deeper moisture storage. African arid regions require careful interpretation due to sparse vegetation and the barren ground flag.

2.8 ESA CCI River Discharge

The River Discharge project produces long-term time series of river discharge by combining satellite altimetry with in-situ measurements.



Selected basins

1 - AMAZON	4 - CONGO	7 - GARONNE	10 - LENA	13 - MARONI	16 - OB
2 - CHAD	5 - DANUBE	8 - INDUS	11 - LIMPOPO	14 - MISSISSIPPI	17 - PO
3 - COLVILLE	6 - GANGA-BRAHMAPUTRA	9 - IRRRAWADDY	12 - MACKENZIE	15 - NIGER	18 - ZAMBEZI

Figure 8: All the selected 18 basins of the ESA CCI River Discharge data. Source: Selection of river basin document (v1.1).

2.8.1 Core metadata

Source	ESA Climate Change Initiative
Spatial resolution	Point measurements at 54 gauging stations across 18 river basins
Native CRS	WGS 84 (EPSG: 4326)
Temporal coverage	2002–2022; monthly or daily discharge series depending on station
Update frequency	Climate data record updated periodically
Sensor lineage	Satellite radar altimetry (Envisat, Jason-1/2/3, Sentinel-3) providing water surface elevations; in-situ discharge used to derive rating curves
Variables	Water surface elevation (m); discharge ($\text{m}^3 \text{s}^{-1}$); rating-curve parameters; quality flags
File format	NetCDF and CSV per station
Access / DOI	https://climate.esa.int/en/projects/river-discharge/
Licence	Open data; CC-BY 4.0

2.8.2 Validation & Uncertainty

The accuracy of satellite-derived discharge depends on the stability of rating curves and the quality of both altimetric water levels and in-situ calibration. No single global accuracy figure is available. Uncertainties increase during extreme events when rating curves may change

due to overbank flow. Only 54 stations are available, so many African rivers and urban catchments are not represented.

2.9 Copernicus Global Digital Elevation Model

The Copernicus DEM is a Digital Surface Model (DSM) that represents the surface of the Earth including buildings, infrastructure and vegetation. The Copernicus DEM is provided in 3 different instances: EEA-10, GLO-30 and GLO-90. Data were acquired through the TanDEM-X mission between 2011 and 2015. In this project we use GLO-30 and GLO-90 data.

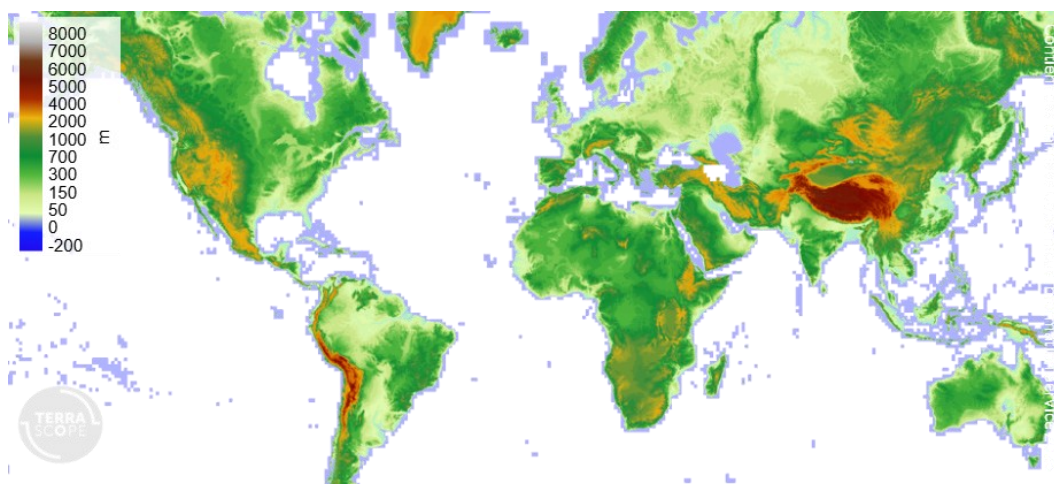


Figure 9: GLO-30 global map visualize in Terrascope (<https://terrascope.be/fr>) platform.

2.9.1 Core metadata

Source	DLR and Airbus, provided under the Copernicus Programme
Spatial resolution	30 m, 90 m
Native CRS	WGS 84 (EPSG: 4326)
Temporal coverage	Between 2011 and 2015
Update frequency	None
Sensor lineage	TanDEM-X Mission
Variables	Height (m)
File format	GeoTIFF or DTED
Access / DOI	https://dataspace.copernicus.eu/explore-data/data-collections/copernicus-contributing-missions/collections-description/COP-DEM
Licence	CC-BY 4.0

2.9.2 Validation & Uncertainty

The absolute vertical accuracy is < 4 m, the relative vertical accuracy is < 2 m for slopes <= 20%, or < 4 m for slopes > 20%. The absolute horizontal accuracy is < 6 m.

2.10 Global Local Climate Zones (LCZ) Map

The LCZ map provides a globally consistent classification of urban form and function. It applies the LCZ typology (10 built and 7 natural classes), derived using EO data and machine learning algorithm.

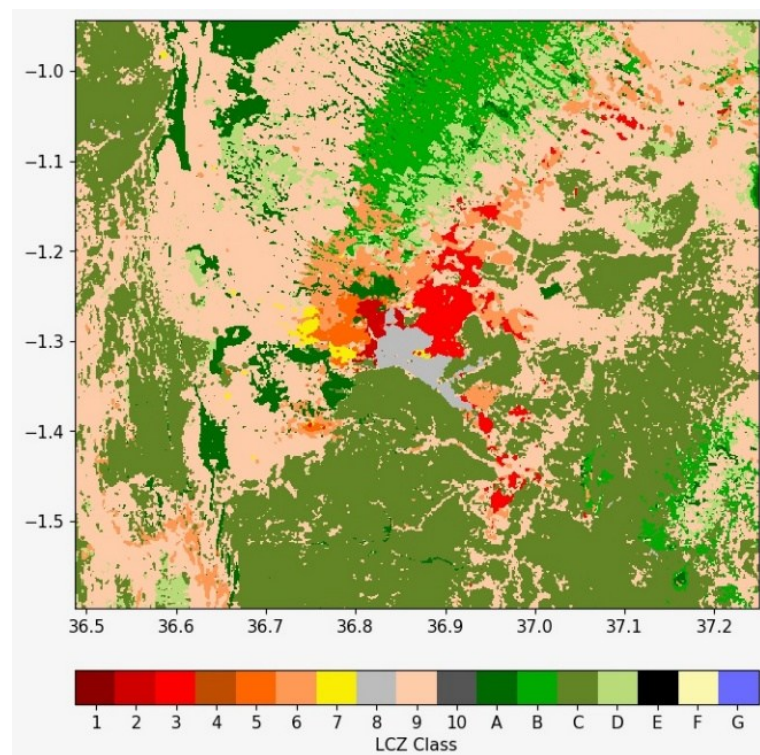


Figure 10: LCZ map of Nairobi (Kenya, Republic of). Source: LCZ Generator, a web application to create Local Climate Zone maps (Demuzere et al., 2021).

2.10.1 Core metadata

Source	Demuzere et al. 2022, Ruhr-Univ. Bochum, WUDAPT project
Spatial resolution	~100 m ($\approx 0.0009^\circ$)
Native CRS	WGS 84 (EPSG: 4326)
Temporal coverage	Nominal year 2018
Update frequency	One global release published in 2022.
Sensor lineage	Landsat-8, Sentinel-1, Sentinel-2, ALOS PALSAR, VIIRS NTL, GEDI + auxiliary DEM/DSM datasets
Variables	LCZ class (17 categories), Classification probability layer (%)

File format	GeoTIFF
Access / DOI	https://doi.org/10.5281/zenodo.6364594
Licence	CC-BY 4.0

2.10.2 Validation & Uncertainty

The Global LCZ map (2018) was validated through extensive benchmarking against labelled training polygons and independent reference datasets. The overall mean accuracy is about 74–75 %, with performance varying by class. Built LCZ classes show F1-scores ranging from ~50 % for compact high-rise areas to ~78 % for open low-rise neighbourhoods, while natural LCZs generally achieve higher scores.

Brousse et al., 2019 applied LCZ data creation workflow to map Kampala (Uganda) and Dakar (Senegal). The LCZ lightweight low-rise accuracy they measured was very low (F1 score below 0.3 and overall accuracy below 40%). They showed LCZ lightweight class was often misclassified with LCZ compact low-rise class because of the same spectral signature.

2.11 Large-scale Slum Probability Maps

This dataset provides probability maps of slums across 529 major cities in 44 sub-Saharan African countries. A deep learning model was trained on ground-truth data from 16 cities and applied to Sentinel-2 imagery.

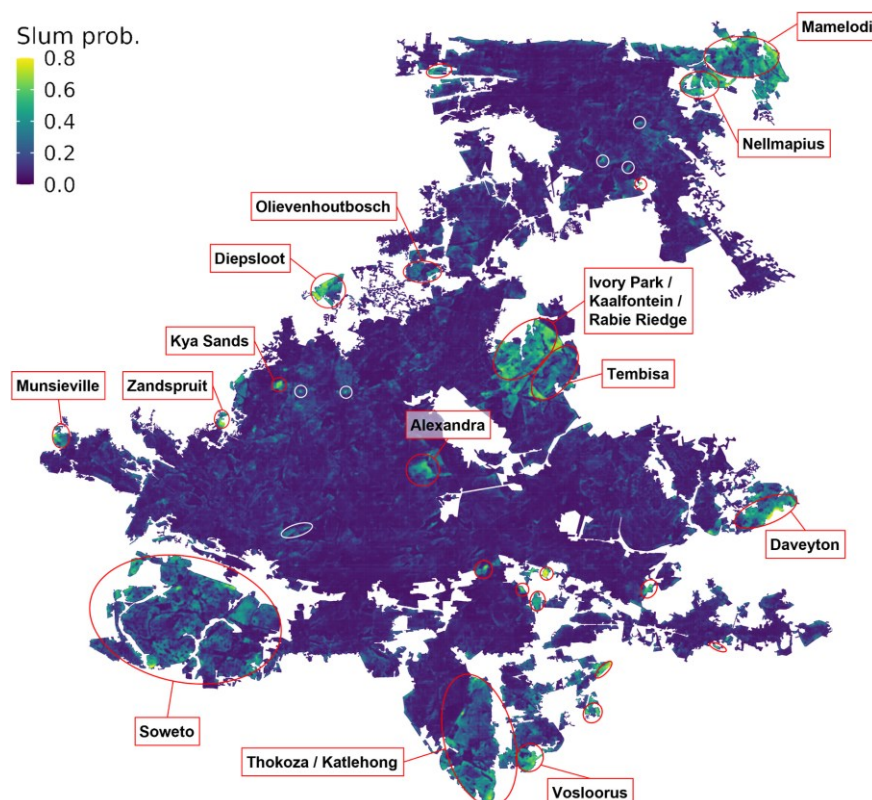


Figure 11. Predicted slum locations in Johannesburg/Pretoria, South Africa (2022). Source: Büttner et al., 2025.

2.11.1 Core metadata

Source	Büttner et al., 2025, ETH Zurich, Swiss Data Science Center, University College London.
Spatial resolution	10 m
Native CRS	WGS 84 (EPSG: 4326)
Temporal coverage	Yearly (2016-2022)
Update frequency	One-off release published in 2025.
Sensor lineage	Sentinel-2 MSI (10 m bands), Africapolis urban polygons, WorldPop population data.
Variables	Slum probability (0-1); binary slum masks; urban population share in slums.
File format	GeoTIFF
Access / DOI	https://doi.org/10.1016/j.habitatint.2025.103403
Licence	CC-BY 4.0

2.11.2 Validation & Uncertainty

The model achieved recall ~70 % and precision ~59 % in cities with ground-truth data (Tanzania, Kenya, South Africa), giving F1-scores around 63 %. Accuracy is higher for large, contiguous slum areas but decreases in cities with small, scattered settlements. It is advised to interpret results conservatively in peri-urban zones where probability values are near threshold levels.

3 ATMOSPHERIC (RE-ANALYSIS / CLIMATE) DATA

3.1 ACMAD In-Situ Meteorological Data (Africa)

The African Centre of Meteorological Applications for Development (ACMAD) operates the Regional Climate Centre for Africa and maintains a database of in-situ meteorological observations across the continent.

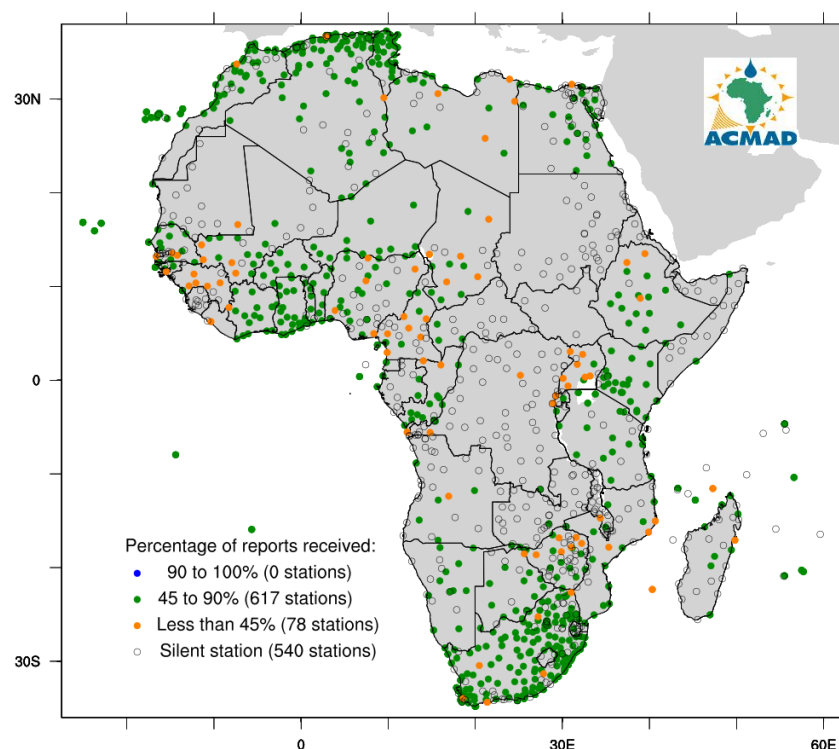


Figure 12: Monthly monitoring of synoptic data report showing the distribution of the climate stations over Africa and the number of reports received from those stations in August 2025. Source: [ACMAD Data Services](#).

3.1.1 Core metadata

Source	ACMAD
Spatial resolution	Point observations at synoptic and climate stations across Africa
Native CRS	WGS 84 (EPSG: 4326)
Temporal coverage	Station records span multiple decades (mainly since 1980)
Update frequency	Daily updates via GTS-SYNOP
Sensor lineage	In-situ instruments at synoptic weather stations; reanalysis products (ARC2 precipitation, ERA5 temperature/humidity) supplement gaps
Variables	Daily minimum, maximum and mean temperature, daily precipitation, humidity, wind speed and direction (where available)

File format	exports available as CSV or NetCDF on request
Access / DOI	Data access via ACMAD upon request; not openly distributed
Licence	Data sharing agreements with national meteorological services

3.1.2 Validation & Uncertainty

Observations follow World Meteorological Organization standards, but accuracy and continuity vary by station and country. Sparse network density in parts of Africa and inconsistent temporal records can limit regional analyses. Data quality control is performed at national and regional levels, with flags for missing or suspect values. Reanalysis estimates (ARC2, ERA5) are used to fill gaps, introducing their own uncertainties.

3.2 ERA5 Atmospheric Reanalysis

ERA5 is the fifth-generation global atmospheric reanalysis produced by the European Centre for Medium-Range Weather Forecasts (ECMWF) under the Copernicus Climate Change Service (C3S). It assimilates observations from surface stations, radiosondes, aircraft and satellites into the Integrated Forecast System (IFS) to provide hourly estimates of atmospheric, land-surface and ocean-wave variables.

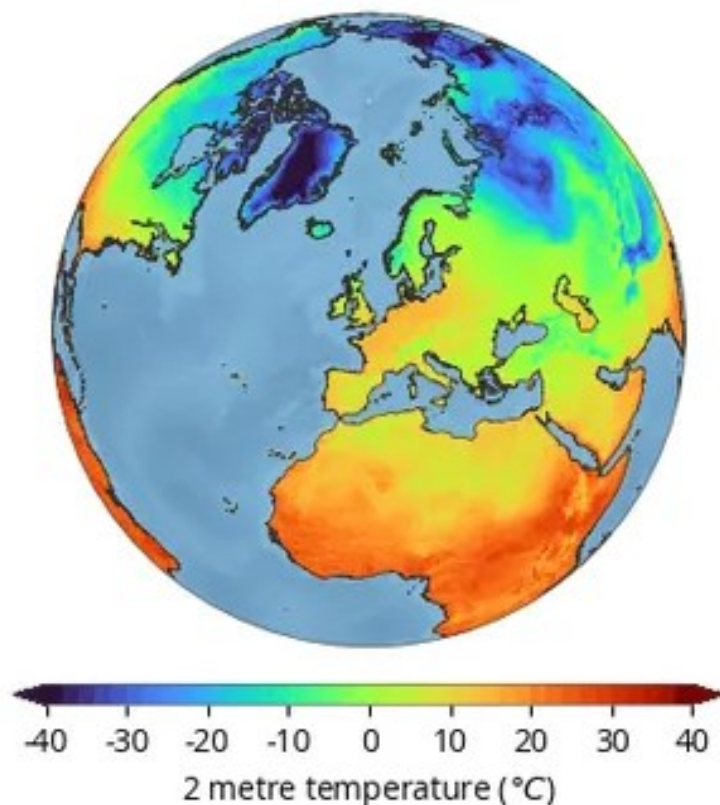


Figure 13: ERA5-Land air temperature on 1 January 2023 showing global temperature distribution. Source: [CDS](#).

3.2.1 Core metadata

Source	ECMWF; C3S
Spatial resolution	0.25° × 0.25° (~31 km)
Native CRS	WGS 84 (EPSG: 4326)
Temporal coverage	Hourly estimates from 1940 to the present (5-day latency)
Update frequency	Preliminary data daily with ~5-day latency; final data released 2–3 months later
Sensor lineage	Assimilation of surface, radiosonde, aircraft and satellite observations into the ECMWF IFS via 4D-Var
Variables	Atmospheric fields (temperature, pressure, winds, humidity), surface variables (precipitation, soil moisture, land temperature), sea-state variables
File format	NetCDF, GRIB
Access / DOI	https://cds.climate.copernicus.eu/
Licence	Copernicus free and open data licence

3.2.2 Validation & Uncertainty

ERA5 has been extensively evaluated against radiosondes, surface observations and previous reanalyses. The 10-member ensemble provides an estimate of analysis uncertainty. Uncertainties are higher prior to the satellite era (pre-1979), and biases in precipitation and surface fluxes vary regionally.

3.3 CMIP6 Climate Projections

The Coupled Model Intercomparison Project Phase 6 (CMIP6) coordinates climate simulations from modelling teams worldwide. Historical experiments (1850-2014) and future scenarios (2015–2100) driven by Shared Socioeconomic Pathways (SSPs) provide daily and monthly fields of key climate variables.

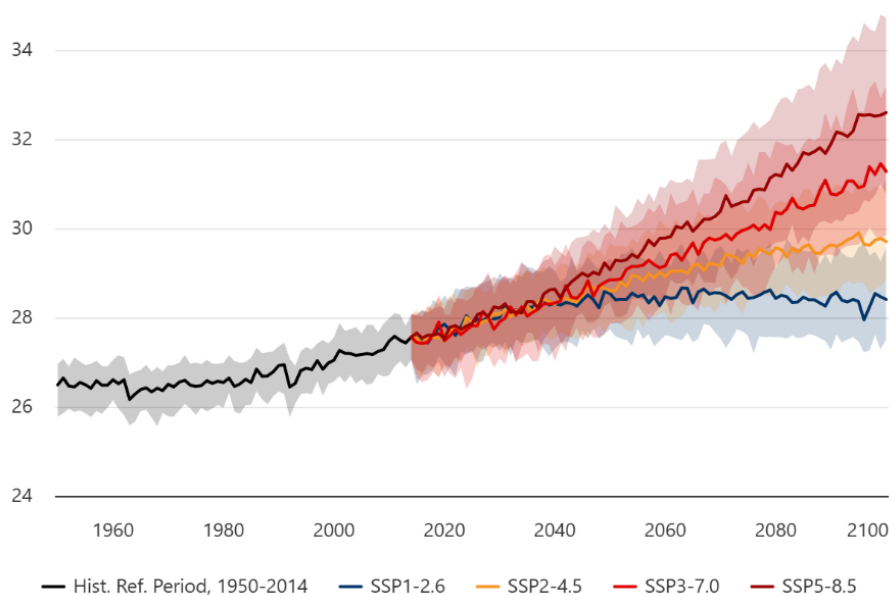


Figure 14: projected ensemble annual mean surface air temperature under different SSPs (Chad, 1950-2100). Shading indicate the 10th and 90th percentile ranges among the different climate models. Source: World Bank, [Climate Change Knowledge Portal](#).

3.3.1 Core metadata

Source	World Climate Research Programme (WCRP); CMIP6 participating modelling centres
Spatial resolution	Varies by model; typical grid spacing 0.5–2.5° (~50–250 km)
Native CRS	WGS 84 (EPSG: 4326)
Temporal coverage	Historical simulations 1850–2014; scenario simulations 2015–2100
Update frequency	One-off model runs; new experiments released periodically; daily and monthly output
Sensor lineage	Numerical climate models representing atmosphere, ocean, land and sea ice; no direct observations
Variables	Temperature, precipitation, winds, humidity, radiation, soil moisture and hundreds of other variables; daily and monthly time steps
File format	NetCDF
Access / DOI	https://cds.climate.copernicus.eu/
Licence	Varies by modelling center; generally free for research use

3.3.2 Validation & Uncertainty

Historical runs are evaluated against observed climate; performance varies by region and variable. Ensemble means reproduce large-scale climate reasonably, but individual models

exhibit biases in temperature and precipitation. Uncertainty arises from model structure, parameterization and scenario assumptions.

4 MODELLING TOOLS

The modelling tools described in this section are pre-existing tools developed by the consortium partners to simulate urban climate, urban growth, and flooding. Below, a brief description of the salient features and examples of previous applications of these models is provided, including a description of the use of Earth Observation data.

4.1 UrbClim

UrbClim (De Ridder et al., 2015) is a dedicated urban climate model developed to simulate the specific atmospheric conditions of cities at medium-high resolution (100-300 m). Unlike computationally demanding mesoscale models, UrbClim achieves a unique balance between physical realism and computational efficiency (García-Díez et al., 2016) by coupling a simplified but physically consistent representation of the urban boundary layer with a detailed land surface scheme. This allows simulations to be performed rapidly over large numbers of cities while still capturing the fine-scale climate processes that shape the urban environment, most notably the urban heat island and its implications for thermal comfort and health.

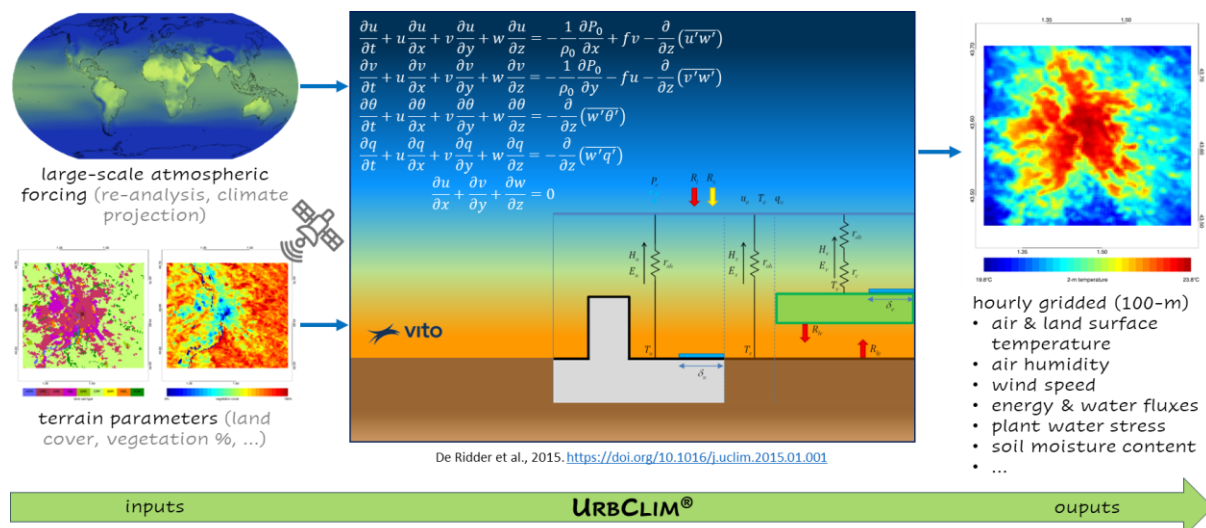


Figure 15. Schematic overview of the UrbClim model.

A defining strength of UrbClim is its reliance on satellite Earth Observation data to characterise the urban surface. Global and regional EO products provide information on land cover, impervious surfaces, and vegetation abundance, which are essential for representing the heterogeneity of cities and their surroundings. Fractional vegetation cover and related indices such as NDVI are used to describe the amount and seasonal cycle of green infrastructure, directly informing evapotranspiration and shading in the model. In addition, EO-driven terrain parameters such as soil type, albedo and emissivity are integrated to ensure a realistic simulation of surface-atmosphere exchanges. This satellite-based specification of input data makes UrbClim particularly suitable for applications in regions where ground-based information is limited, providing a consistent and transferable framework for urban climate analysis worldwide.

The model has been applied extensively across Europe, where it was used to provide detailed assessments of the urban heat island in over 100 cities, demonstrating its capacity to capture both the spatial distribution and intensity of urban warming at scales relevant to local planning (Lauwaet et al., 2024). Its value has also been demonstrated in African contexts, where it has been used to quantify current and future heat stress in Niamey and to evaluate the cooling potential of urban vegetation (Souverein et al., 2023) in Johannesburg, where nested metre-scale simulations have been combined with vulnerability data to assess exposure and risk (Souverein et al., 2022). Similar applications in Asia, such as the analysis of heat risks in Delhi (Sharma et al., 2019), further illustrate the transferability of the approach to diverse climatic and urban settings.

UrbClim can be driven by global and regional climate projections such as those from the CMIP and CORDEX initiatives, enabling robust assessments of how urban climates will evolve under different greenhouse gas scenarios. This capacity to dynamically downscale large-scale climate information to the scale of individual cities, while making explicit use of satellite-derived surface data, provides an essential tool for understanding how future warming, extreme heat events, and urban growth will interact. Beyond the study of climate hazards, UrbClim has also been applied to explore adaptation options, including the effectiveness of nature-based solutions such as parks and tree planting, as well as urban design measures aimed at reducing heat stress and improving outdoor comfort (Lauwaet et al., 2018).

4.2 GeoDynamix

GeoDynamix was developed at VITO for spatial-dynamic land-use modelling (White et al., 2012; Crols, 2017). GeoDynamix is a cellular automata (CA) land-use model that allows to allocate the expected growth in different socio-economic scenarios into a projected population distribution and a set of different land-use categories, taking into account spatial attractivity rules, physical suitability, restricted areas and transportation. Unlike most CA models, it is not a purely deterministic model by including stochastic perturbation, which makes Monte Carlo analysis (addressing spatial uncertainty) possible.

The GeoDynamix model needs to be calibrated with historical land-use and population maps to derive values for its parameters that determine the transition rules of each land-use category. In order to allocate population and urban growth (or other land-use change), a transition potential is calculated at each annual time step for each land-use category for all cells in the study area. The transition potential is calculated based on four elements:

1. **Transition rules of spatial attraction and repulsion:** assess the quality of its neighbourhood – which attractive and/or repulsive land uses are present in the surrounding neighbourhood?
2. **Suitability:** quantifies the degree to which a cell is fit to support a particular land-use function and is calculated on the basis of a number of factor maps determining the physical and environmental appropriateness of cells (such as slope and wetlands).
3. **Zoning:** maps that identify for each land-use category whether there are *no-go* zones which are not allowed to be occupied by a certain land use (e.g. protected

nature areas, military areas).

4. **Accessibility:** quantifies the level of accessibility by the transport mode(s) relevant for each specific land use (e.g. accessibility to roads and/or public transport stations).

Based on these four elements – CA transition rules, suitability, zoning and accessibility – the model calculates for every time step the transition potential for each cell (see Figure 16). In the course of time, population densities will be updated, and until overall land claims are satisfied, cells will change to the land use for which they have the highest transition potential. Only changes towards built-up categories are actively modelled. Changes from built-up to natural can rarely appear in cells with low potentials for existing built-up land covers.

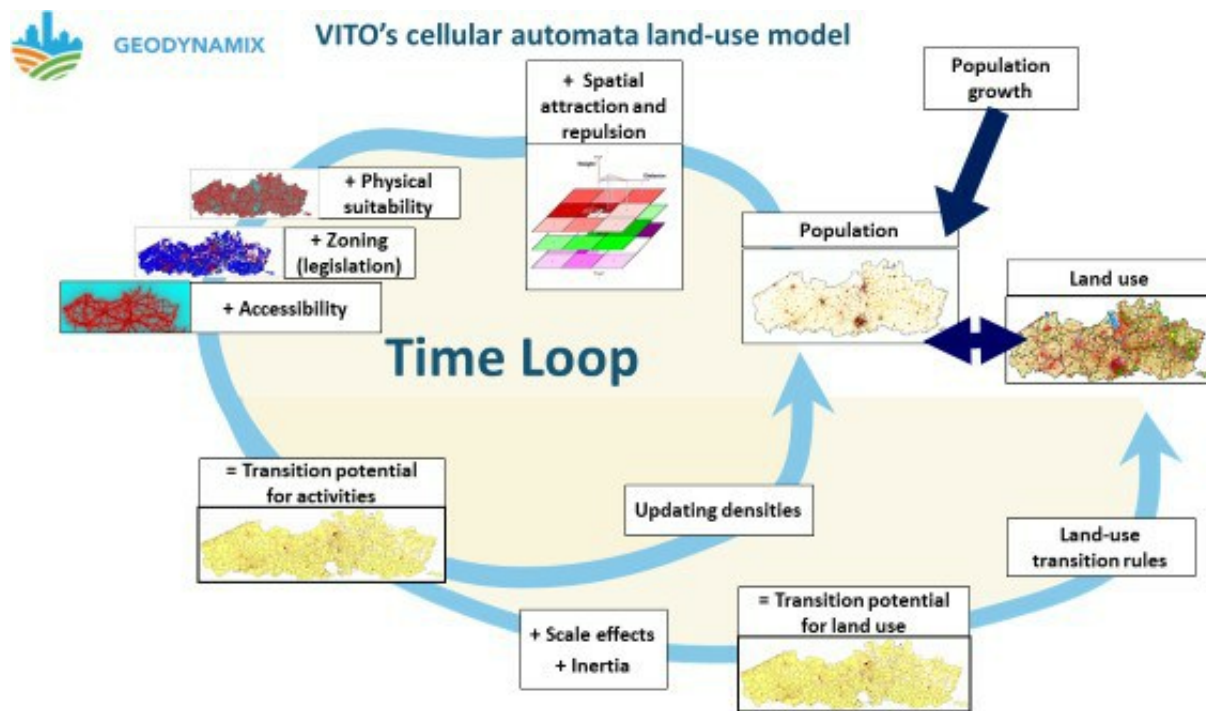


Figure 16. The components of the GeoDynamix land-use change model

The model has been applied to numerous projects in different countries with varying scales, ranging from small cities to entire countries. Some recent applications include the calculation of urban sprawl costs in Flanders, Belgium (Vermeiren et al., 2022), but also coupling the results to urban climate modelling in Niamey (Souverijns et al., 2023) and several cities in India (e.g. Bano et al., 2025).

4.3 SAHEL

The Satellite-based African Hydrological Ensemble Learner (SAHEL) is an advanced machine learning framework that utilizes ECVs and features derived from the SWAT rainfall-runoff model. SAHEL is designed to predict inundation risk in African cities under future climate conditions. The model is also designed to address the dual challenge of data scarcity and rapid urbanization in Africa. SAHEL leverages EO derived hydrometeorological variables driving flood patterns, and urban growth maps which depict the varying rate and type of

expansion in different African cities. SAHEL's primary objective is to estimate pluvial and fluvial floods and their impacts under evolving climate and land-use scenarios, supporting evidence-based adaptation strategies.

SAHEL is structured as an ensemble of deep learning models, with Convolutional Neural Networks (CNNs) as its core building blocks. The inputs of SAHEL's sub-ensemble models comprise EO features (ECV-aligned) and non-EO features describing the hydrologic identity of the city, its urban growth and topology. Some features such as surface runoff and streamflow may be co-generated through a SWAT hydrological model to ensure physical consistency of the model and validate other input features. CNNs are particularly suited for distributed data such as flood maps and satellite imagery, making them a great solution across Hydroinformatics research (Frame et al., 2024; Jamali et al., 2024; Sharma & Kumari, 2024). They enable SAHEL to capture complex spatiotemporal dependencies between multiscale hydrometeorological drivers and inundation patterns (Jamali et al., 2025). The ensemble approach enhances robustness and generalization by combining multiple machine learning architectures trained on diverse subsets of data and features (Sharma & Saharia, 2025).

Key features of the SAHEL mechanism include:

- Multiscale Input Configuration: Inputs are processed at varying spatial and temporal resolutions of EO data and climate projections.
- Feature Integration: Input space includes precipitation, soil moisture, land cover, digital elevation models (DEMs), SWAT-derived features and urban growth layers, harmonized into a common grid.
- Adversarial Learning: some of the sub-ensemble CNN models may incorporate adversarial training to improve resilience against data variability and noise.
- Uncertainty Quantification: Ensemble predictions are aggregated to provide probabilistic flood risk estimates, accounting for model and input uncertainties.

The Satellite-based African Hydrological Ensemble Learner (SAHEL) is a collection of machine learning models which predict inundation risk in African cities. SAHEL utilizes hydrometeorologic and climate profiles as well as their urban growth projection of the cities to infer the pluvial and fluvial contribution to the inundation risk. The main building blocks of SAHEL are deep convolutional neural networks (CNNs) which have been recently proven to be effective across hydroinformatics research (Frame et al., 2024; Jamali et al., 2024; Sharma & Kumari, 2024).

The individual CNN models are designed to ingest a multiscale input configuration to address the varying spatiotemporal resolution of the input space (Jamali et al., 2025). Developed by uOttawa, the ensemble-based deep learning framework integrates multiple data streams to ensure comprehensive representation of flood drivers. These are EO data (ECVs), non-EO data (DEM), with emphasis on explicit utilization of VITO's GeoDynamix urban growth model. In addition, uOttawa is developing a novel rapid inundation labelling (RIL) approach which utilizes dynamic ECV and static flood risk data to create dynamic, instantaneous inundation maps. This approach is the first to be used and eliminated the adverse effects of directly utilizing SAR maps which has limited insights on urban flood settings. RIL is depicted in Figure 3, as an integral part of the Framework to create SAHEL.

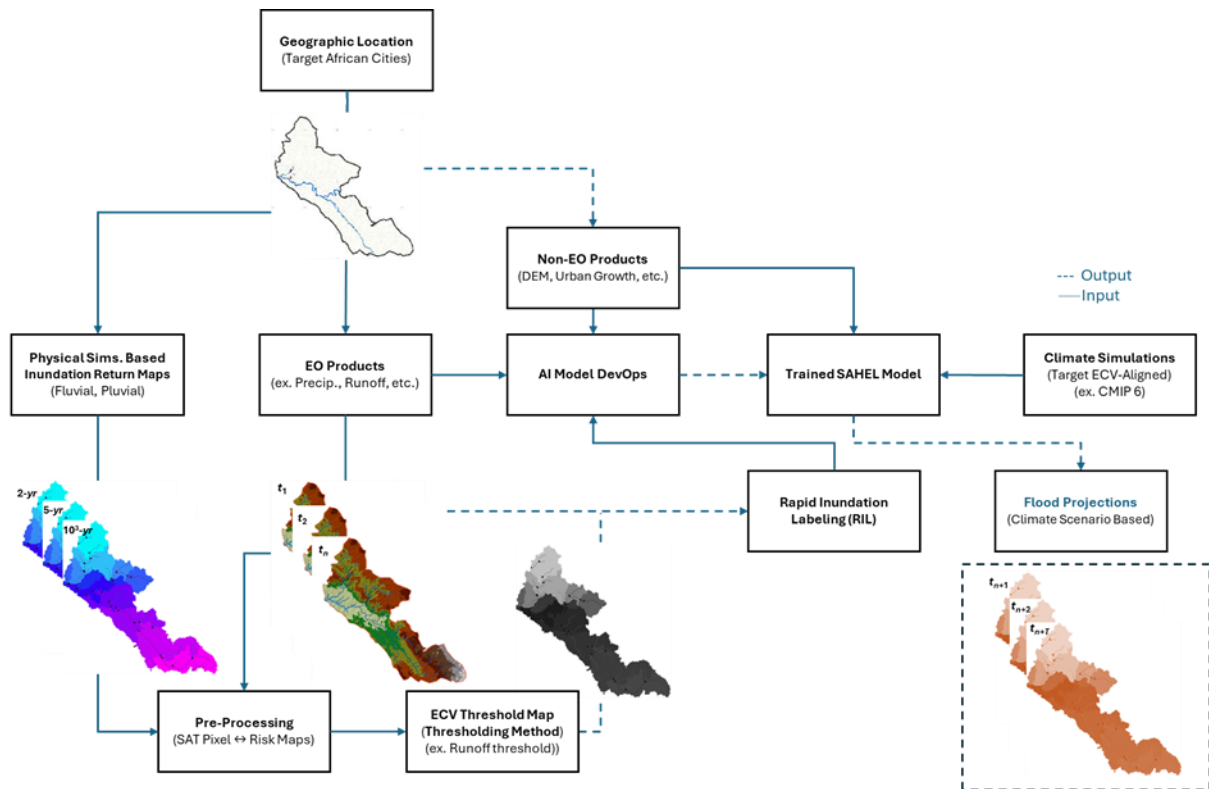


Figure 17: The components of the SAHEL development framework, with emphasis on the RIL approach.

5 PROCESSING TOOLS

5.1 Input for urban growth / climate models

5.1.1 UrbClim Meteo Processing

Variable	Description	Unit
3D		
U	Zonal wind component (east–west)	m/s
V	Meridional wind component (north–south)	m/s
T	Temperature	K
Q	Specific humidity	kg/kg
2D		
sp	Surface pressure	Pa
z	Geopotential	m ² /s ²
sst	Sea surface temperature	K
tp	Total precipitation	m
ssrd	Surface solar radiation downwards	J/m ²
strd	Surface thermal radiation downwards	J/m ²
stl1–4	Soil temperature levels 1-4 (Depths: 0-7, 7-28, 28-100, 100-289 cm)	K
swvl1–4	Soil water volumetric content levels 1–4	m ³ /m ³

Table 1: Description of the ERA5 variables used in the UrbClim model.

For each year, ERA5 3D (U, V, T, Q) and 2D (sp, z, sst, tp, ssrd, strd, stl1–4, swvl1–4) datasets (3.2) subset to the target city domain are read, duplicate/wrapped coordinates are corrected, and all variables are interpolated to the city center.

Full-level pressure and geopotential are computed from the ECMWF hybrid coefficients, and potential temperature is derived from temperature and pressure. Vertical profiles are then linearly interpolated to the UrbClim output heights, and the time axis is resampled to the model step.

Soil and surface initial states are constructed on the UrbClim grid from ERA5 soil layers and a multi-depth soil-texture map (using Clapp–Hornberger relations with light smoothing). Sea-surface temperature (SST) is obtained from an area-mean over a $0.5^\circ \times 0.5^\circ$ box around the city. The result is one UrbClim forcing NetCDF per year containing center-point meteorological profiles, surface fluxes/precipitation time series, and grid-based soil/vegetation initialization.

Future climate projections are generated by combining UrbClim baseline output with large-scale climate change signals from CMIP6 (3.3). First, UrbClim outputs are aggregated into

daily series at a rural reference grid cell and each day is classified into monthly percentiles. Independently, CMIP6 bias-corrected model data are used to calculate delta matrices, which quantify the mean change between a baseline period and a future period for each month and percentile. These deltas are then applied to the UrbClim daily baseline series: each baseline day inherits the change associated with its month and percentile. This methodology preserves UrbClim’s fine-scale spatial patterns and realistic daily variability, while consistently shifting the climate towards the CMIP6 future projections. Processing is done in python using xarray, numpy, pandas, netCDF4, matplotlib, cartopy, scipy and GDAL libraries.

5.1.2 Sentinel-2 NDVI

For each year, Sentinel-2 products from the Copernicus Data Space Ecosystem (CDSE) are queried quarterly over a $1^\circ \times 1^\circ$ area of interest ($\pm 0.5^\circ$ about the city center). For every returned product, NDVI is computed from bands B08 and B04, and all product-level NDVI rasters in a given quarter are mosaicked to produce a single quarterly NDVI at 10 m for that year. Quarterly means are then computed across years (Q1–Q4), and a monthly NDVI climatology is derived by linearly combining adjacent quarters with fixed 2/3–1/3 weights around seasonal boundaries. Processing is done in python using requests, geopandas, shapely, rasterio, xarray, rioarray, numpy, tqdm and subprocess libraries.

5.1.3 Land-use maps

WorldCover 10 m data are clipped over three large African AOIs (north-west, north-east, south), and built-up pixels are masked to create a “natural-only” layer. Both the original and the natural-only rasters are resampled to the reference 300 m CCI Land Cover grid (same EPSG, pixel size, and origin). Any 300 m cell classified as built-up is then replaced by the corresponding value from the natural-only map.

A land–sea mask is derived by clipping GSHHS to each AOI. Open-ocean water is distinguished from inland water by recoding pixels with WorldCover water class and land–sea mask indicates sea to a dedicated ocean class. The three regional outputs are finally mosaicked into a single, continent-wide 300 m GeoTIFF.

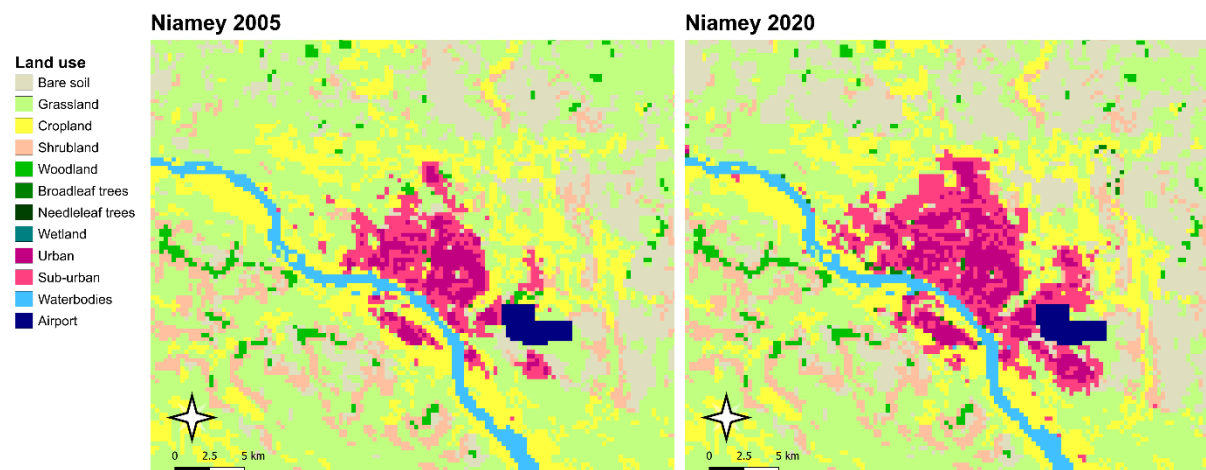


Figure 18. Example of resulting land-use maps for Niamey, Niger for the years 2005 and 2020.

This processed WorldCover layer is then integrated with other datasets to construct the final land-use maps for the 54 modelled cities. Specifically, the land-use maps combine ESA CCI Land Cover (1.1), the Copernicus Global Human Settlement Layer – Built-up (GHS-BUILT-S, see 1.3), airports from OpenStreetMap, and the WorldCover-derived data described above.

The GHS-BUILT-S is reprojected to 300 m resolution using a sum. The map starts with a reclassification of all non-urban CCI classes (see Table 2). Next, urban and suburban are added on top of this classification, defined as cells with a GHSL Built-up proportion above specific thresholds. Thirdly, OSM airports are added. Unclassified cells after these three first steps are filled with the most present natural (non-urban) WorldCover class within the 300-m cell. Processing is done in python using GDAL, rasterio and numpy libraries.

Table 22. Reclassification from ESA CCI Land Cover, GHSL Built-up and OpenStreetMap into 12 classes of land use.

ID	Land Use	ESA CCI Land-Cover classes, GHSL Built-up, OpenStreetMap (OSM)
1	Bare soil	150 - Sparse vegetation (tree, shrub, herbaceous cover) (<15%) 200 - Bare areas
2	Grassland	40 - Mosaic natural vegetation (tree, shrub, herbaceous cover) (>50%) / cropland (<50%) 110 - Mosaic herbaceous cover (>50%) / tree and shrub (<50%) 130 - Grassland 140 - Lichens and mosses
3	Cropland	10 - Cropland, rainfed 20 - Cropland, irrigated or post-flooding 30 - Mosaic cropland (>50%) / natural vegetation (tree, shrub, herbaceous cover) (<50%)
4	Shrubland	120 - Shrubland
5	Woodland	100 - Mosaic tree and shrub (>50%) / herbaceous cover (<50%)
6	Broadleaf trees	50 - Tree cover, broadleaved, evergreen, closed to open (>15%) 60 - Tree cover, broadleaved, deciduous, closed to open (>15%) 90 - Tree cover, mixed leaf type (broadleaved and needleleaved)
7	Needleleaf trees	70 - Tree cover, needleleaved, evergreen, closed to open (>15%) 80 - Tree cover, needleleaved, deciduous, closed to open (>15%)
8	Wetland	160 - Tree cover, flooded, fresh or brakish water 170 - Tree cover, flooded, saline water 180 - Shrub or herbaceous cover, flooded, fresh/saline/brakish water
9	Urban	GHS-BUILT-S >= 40%
10	Sub-urban	20% <= GHS-BUILT-S < 40%
11	Water bodies	210 - Water bodies
12	Airport	OSM: aeroway = aerodrome

5.1.4 UrbClim terrain processing

Terrain and surface fields for UrbClim are generated for all 54 cities. The land-use map (Figure 18) is used both directly by the model and indirectly to assign parameter fields (albedo, emissivity, roughness length, root depth, and plant resistance).

Land mask and sea mask are derived from the land-use map. Vegetation cover is computed month-by-month from the NDVI composites (5.1.2) and rescaled to [0, 1] (Gutman and Ignatov, 1998). Urban aerodynamic parameters are obtained by combining GHSL Built-up fraction (2.3) with building-height information. The topography is taken from the Copernicus Global Digital Elevation Model (2.9). Soil sealing follows the Global Impervious Surface Area (GISA) dataset (Huang et al., 2022). Anthropogenic heat flux is taken from Wang et al. (2022). Soil texture classes are converted to hydraulic parameters, default values are applied first, then class-specific Clapp & Hornberger parameters are mapped and lightly smoothed with a Gaussian filter to suppress pixel-scale noise. Processing is done in python using GDAL, numpy, netCDF4, and scipy libraries.

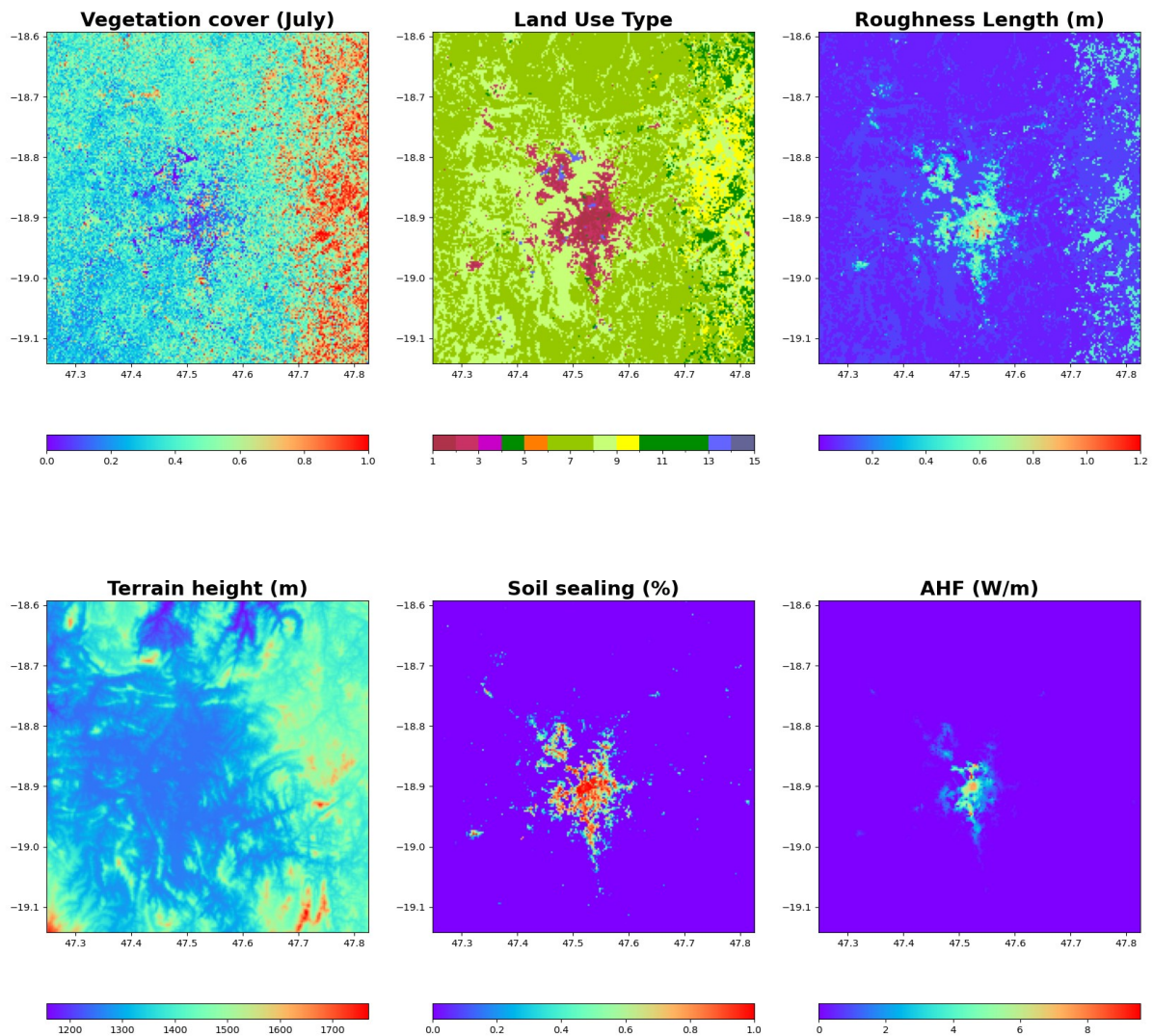


Figure 19: Output example of UrbClim terrain processor for Antananarivo (Mozambique).

5.2 Input for flooding model

Credible flood analysis requires inputs that are high-quality, complete, and internally consistent. An ESA-aligned inventory was therefore assembled around the Essential Climate

Variables (ECVs). Mature ESA products were treated as primary references where available, and additional datasets from reputable providers were added to broaden coverage and support continuity. All layers were acquired through scripted pipelines and standardized prior to integration, yielding a reproducible basis for downstream hydrologic analyses and AI workflows.

5.2.1 Acquisition and domain strategy

Acquisition and preparation were automated in Python and executed in parallel across several machines due to data volume. To avoid schedule slippage while city selections are finalized, downloads were performed for the entire African domain, using the earliest available start date for each product and extending to the present. Processing was made in Python using xarray, pandas, cdsapi, netCDF4 and related libraries.

5.2.2 Missing-data handling

A conservative, documented procedure was applied. Short temporal gaps (sub-daily to 1-2 days) were bridged via appropriate interpolation (zero-filling for daily/sub-daily precipitation, linear for other variables) or forward filling when event statistics were not distorted; longer gaps were infilled from physically consistent companion products after bias alignment. Spatial gaps were handled case by case: GRDC station gaps were provisionally supported using nearby stations within the same river system where hydrologically defensible, and SRTM voids were resolved with the official void-filled DEM. All imputed values were flagged at the attribute level, and basic sanity checks were performed prior to release for modelling.

5.2.3 Multiscale outputs and AI-readiness

The input portfolio is intrinsically multiscale: reanalysis fields operate at coarse resolutions relative to local hydraulics; precipitation products vary by resolution and sampling frequency; discharge is point-based; and terrain data are orders of magnitude finer. A multiscale AI architecture is adopted to jointly learn local high-resolution signals and broad-scale climatic controls. Accordingly, analysis-ready harmonized layers are maintained alongside native-resolution tiles, with consistent spatial and temporal indexing. This configuration enables immediate use in preliminary analyses while providing an ESA-aligned, reproducible basis for subsequent training and validation.

6 RESEARCH INFRASTRUCTURE

6.1 VITO

VITO's Environmental Intelligence group operates a state-of-the-art computing infrastructure housed in an energy-efficient Tier 2 data centre. The facility is designed with a cold corridor concept for optimal cooling and redundant power supply for reliability.

6.1.1 Computing power & network performance

- 1024 CPU threads, 8 TB RAM, and 500 TB high-performance storage, ensuring robust capabilities for intensive computational tasks.
- Switched copper-based gigabit Ethernet for core networking, with fiber links supporting high-traffic nodes such as the storage engine and virtualization farm for efficient data transfer.
- Segmented network infrastructure for development, testing, and production, ensuring security and operational efficiency.
- 10 Gbit/s internet connection via Belnet and GEANT, with an upgraded internal core network featuring 40 Gbit/s switching to the central firewall and 10 Gbit/s connections to critical data centre switches.

6.1.2 Hybrid cloud & scalable processing for big data

- A hybrid scale-out storage system (SAN/NAS) supports flexible data storage and access across research and operational workflows.
- OpenStack private cloud provides on-demand software-defined resources for internal and external users, enabling seamless scaling to hybrid cloud solutions like DIAS.
- Hadoop cluster with 8,000+ cores and 30+ TB RAM, integrated with Spark, supports high-performance big data analytics, particularly for Earth observation and large-scale time-series processing.

6.1.3 Algorithm development & workflow automation

- Pre-configured virtual machines (VMs) allow researchers to develop, test, and deploy algorithms in an optimised environment with full access to data archives.
- Apache Airflow enables robust workflow orchestration and monitoring, ensuring transparent and efficient data processing.

6.1.4 Comprehensive IT infrastructure overview

- Network: 1/10 Gbit/s LAN, 16/32 Gbit/s FibreChannel SAN, and redundant 10 Gbit/s firewall with advanced security features.
- Servers: OpenStack private cloud (300+ VMs), VMware cluster, and Hadoop-based parallel satellite data processing.
- Storage: 11 PB NetApp storage for live datasets and 7 PB tape storage for long-term archiving.

- Cloud Computing: Integration with CreoDIAS, WeKEO, Open Telekom Cloud, and AWS for scalable processing.
- Collaboration Tools: Teams & Zoom with hybrid conferencing technology for seamless virtual engagement.

6.1.5 Relevance to the planned work

This infrastructure provides the scalability, security, and computational power necessary for processing large-scale datasets, including intensive environmental modelling, big data analytics, and machine learning applications. The combination of high-performance computing, large-scale storage, and cloud integration ensures that VITO can handle the demands of Earth observation, climate research, and operational data workflows with efficiency and reliability. This setup is fully equipped to support the planned work, providing a highly capable and future-proof platform for computational research and large-scale processing.

6.2 uOttawa

uOttawa provides access to academic computing infrastructure as well as specialized geospatial and machine learning tools. The team has access to workstations with industry-standard software such as ArcGIS and ENVI for remote sensing data analysis and visualization, enabling high-resolution environmental mapping and climate change modelling. For large-scale computations, uOttawa leverages High-Performance Computing clusters through the Digital Research Alliance of Canada, which supports parallel processing, big data analytics, and machine learning workflows. This includes access to GPU resources and cloud services via the Rapid Access Service (RAS). The university also provides global access to a wide array of research literature and databases that are very useful to the project.

7 CONCLUSION

The D1.3 Inventory Document brings together the data and tools that CAIAC needs to move from planning to implementation. By listing available datasets, highlighting their uncertainties, and linking them to the main models, the document ensures that the following work packages can be carried out with confidence and transparency.

The inventory shows how central EO data are to understanding Africa's urban climate challenges, but it also recognises the limits of current products and the importance of combining them with ground-based and socio-economic information. It also points to the complementarity of the models: UrbClim for high-resolution heat mapping, GeoDynamix for projecting urban growth and SAHEL for flood risk.

In conclusion, the D1.3 Inventory Document paves the way for robust simulations of extreme heat and flooding in the targeted African cities.

REFERENCES

- Bano, F., Sanyal, A., Sehgal, V., Gulati, R., Singh, G., Broeckx, J., Rehman, I. H., Theunissen, R., Crols, T., Yang, S. (2025). Building urban climate resilience: An inclusive approach to heat stress mapping and capacity building in Ayodhya, India. *Green Technologies and Sustainability*, 3(3), 100203. <https://doi.org/10.1016/j.grets.2025.100203>
- Brousse, O., Georganos, S., Demuzere, M., Vanhuyse, S., Wouters, H., Wolff, E., Linard, C., Van Lipzig, N.P.-M., Dujardin, S., 2019. Using Local Climate Zones in Sub-Saharan Africa to tackle urban health issues. *Urban Climate* 27, 227–242. <https://doi.org/10.1016/j.uclim.2018.12.004>
- Büttner, N., Stalder, S., Volpi, M., Suel, E., Harttgen, K., 2025. Large-scale slum mapping in sub-Saharan Africa's major cities: Remote sensing and deep learning reveal strong slum growth in the urban periphery between 2016 and 2022. *Habitat International* 161, 103403. <https://doi.org/10.1016/j.habitatint.2025.103403>
- Crols, T., 2017. Integrating network distances into an activity based cellular automata land-use model. Semi-automated calibration and application to Flanders, Belgium. PhD thesis, Vrije Universiteit Brussel and VITO, Brussels, Belgium, and Mol, Belgium. https://geodynamix.eu/sites/geodynamix/files/PhD_thesis_Tomas_Crols_Fin.pdf
- De Ridder, K., Lauwaet, D., Maiheu, B., 2015. UrbClim – A fast urban boundary layer climate model. *Urban Climate* 12, 21–48. <https://doi.org/10.1016/j.uclim.2015.01.001>
- Demuzere, M., Kittner, J., Bechtel, B., 2021. LCZ Generator: A Web Application to Create Local Climate Zone Maps. *Front. Environ. Sci.* 9, 637455. <https://doi.org/10.3389/fenvs.2021.637455>
- Demuzere, M., Kittner, J., Martilli, A., Mills, G., Moede, C., Stewart, I.D., Van Vliet, J., Bechtel, B., 2022. A global map of local climate zones to support earth system modelling and urban-scale environmental science. *Earth Syst. Sci. Data* 14, 3835–3873. <https://doi.org/10.5194/essd-14-3835-2022>
- Frame, J.M., Nair, T., Sunkara, V., Popien, P., Chakrabarti, S., Anderson, T., Leach, N.R., Doyle, C., Thomas, M., Tellman, B., 2024. Rapid Inundation Mapping Using the US National Water Model, Satellite Observations, and a Convolutional Neural Network. *Geophysical Research Letters* 51, e2024GL109424. <https://doi.org/10.1029/2024GL109424>
- García-Díez, M., Lauwaet, D., Hooyberghs, H., Ballester, J., De Ridder, K., Rodó, X., 2016. Advantages of using a fast urban boundary layer model as compared to a full mesoscale model to simulate the urban heat island of Barcelona. *Geosci. Model Dev.* 9, 4439–4450. <https://doi.org/10.5194/gmd-9-4439-2016>
- Gutman, G., Ignatov, A., 1998. The derivation of the green vegetation fraction from NOAA/AVHRR data for use in numerical weather prediction models. *International Journal of Remote Sensing* 19, 1533–154. <https://doi.org/10.1080/014311698215333>

- Huang, X., Yang, J., Wang, W., Liu, Z., 2022. Mapping 10 m global impervious surface area (GISA-10m) using multi-source geospatial data. *Earth Syst. Sci. Data* 14, 3649–3672. <https://doi.org/10.5194/essd-14-3649-2022>
- Jamali, A., Roy, S.K., Hashemi Beni, L., Pradhan, B., Li, J., Ghamisi, P., 2024. Residual wave vision U-Net for flood mapping using dual polarization Sentinel-1 SAR imagery. *International Journal of Applied Earth Observation and Geoinformation* 127, 103662. <https://doi.org/10.1016/j.jag.2024.103662>
- Jamali, A., Roy, S.K., Lu, B., Beni, L.H., Kakhani, N., Ghamisi, P., 2025. MSHCCT: A Multiscale Compact Convolutional Network for High-Resolution Aerial Scene Classification. *IEEE Geosci. Remote Sensing Lett.* 22, 1–5. <https://doi.org/10.1109/LGRS.2025.3556373>
- Lauwaet, D., De Nijs, T., Liekens, I., Hooyberghs, H., Verachtert, E., Lefebvre, W., De Ridder, K., Remme, R., Broekx, S., 2018. A new method for fine-scale assessments of the average urban Heat island over large areas and the effectiveness of nature-based solutions. *OE* 3, e24880. <https://doi.org/10.3897/oneeco.3.e24880>
- Lauwaet, D., Berckmans, J., Hooyberghs, H., Wouters, H., Driesen, G., Lefebvre, F., De Ridder, K., 2024. High resolution modelling of the urban heat island of 100 European cities. *Urban Climate*, 54, 101850. <https://doi.org/10.1016/j.uclim.2024.101850>
- Martinez, G.S., Diaz, J., Hooyberghs, H., Lauwaet, D., De Ridder, K., Linares, C., Carmona, R., Ortiz, C., Kendrovski, V., Aerts, R., Van Nieuwenhuyse, A., Bekker-Nielsen Dunbar, M., 2018. Heat and health in Antwerp under climate change: Projected impacts and implications for prevention. *Environment International* 111, 135–143. <https://doi.org/10.1016/j.envint.2017.11.012>
- Sharma, R., Hooyberghs, H., Lauwaet, D., De Ridder, K., 2019. Urban Heat Island and Future Climate Change—Implications for Delhi’s Heat. *J Urban Health* 96, 235–251. <https://doi.org/10.1007/s11524-018-0322-y>
- Sharma, S., Kumari, S., 2024. Comparison of machine learning models for flood forecasting in the Mahanadi River Basin, India. *Journal of Water and Climate Change* 15, 1629–1652. <https://doi.org/10.2166/wcc.2024.517>
- Sharma, N.K., Saharia, M., 2025. DeepSARFlood: Rapid and automated SAR-based flood inundation mapping using vision transformer-based deep ensembles with uncertainty estimates. *Science of Remote Sensing* 11, 100203. <https://doi.org/10.1016/j.srs.2025.100203>
- Souverijns, N., De Ridder, K., Takacs, S., Veldeman, N., Michielsens, M., Crols, T., Foamouhoue, A. K., Nshimirimana, G., Dan Djé, I., & Tidjani, H., 2023. High resolution heat stress over a Sahelian city: Present and future impact assessment and urban green effectiveness. *International Journal of Climatology*, 1–19. <https://doi.org/10.1002/joc.8268>
- Souverijns, N., K. De Ridder, N. Veldeman, F. Lefebvre, F. Kusambiza-Kiingi, W. Memela, N. K.W. Jones, 2022. Urban heat in Johannesburg and Ekurhuleni, South Africa: A meter-scale assessment and vulnerability analysis. *Urban Climate*, 46, 101331, <https://doi.org/10.1016/j.uclim.2022.101331>.

- Vermeiren, K., Crols, T., Uljee, I., De Nocker, L., Beckx, C., Pisman, A., Broekx, S., Poelmans, L., 2022. Modelling urban sprawl and assessing its costs in the planning process: A case study in Flanders, Belgium. *Land Use Policy*, 113, 105902.
<https://doi.org/10.1016/j.landusepol.2021.105902>
- Wang, S., Hu, D., Yu, C., Wang, Y., Chen, S., 2022. Global mapping of surface 500-m anthropogenic heat flux supported by multi-source data. *Urban Climate* 43, 101175.
<https://doi.org/10.1016/j.uclim.2022.101175>
- White, R., Uljee, I., Engelen, G., 2012. Integrated modelling of population, employment and land-use change with a multiple activity-based variable grid cellular automaton. *International Journal of Geographical Information Science*, 26(7), 1251–1280.
<https://doi.org/10.1080/13658816.2011.635146>

**vision on technology
for a better world**

