D2R Nowcasting

Release Development v0.8.0

Data2Resilience Project Team

Jul 23, 2025

INTRODUCTION

Extreme heat poses risks to human health, well-being, and public spaces. The ICLEI Action Fund 2.0 project Data2Resilience aims to enhance Dortmund's heat resilience through a three-part approach: measurement, modelling, and communication. D2R is collaborating with the city of Dortmund to co-deploy a **state-of-the-art biometeorological sensor network**, that measures the air temperature, relative humidity, wind speed, and mean radiant temperature. To monitor thermal discomfort across the city, D2R developed a nowcasting service that models the **Universal Thermal Climate Index (UTCI) at street level resolution**, using as a basis the biometeorological weather station data and numerical weather predictions from DWD. At its core, the nowcasting service uses the Urban Multi-scale Environmental Predictor (UMEP). Specifically, the SOLWEIG model is utilized for thermal comfort calculations. The final product of the nowcasting service is an interactive dashboard (Fig. 1) whose features and maps are developed according to the collected requirements and feedback from stakeholders.

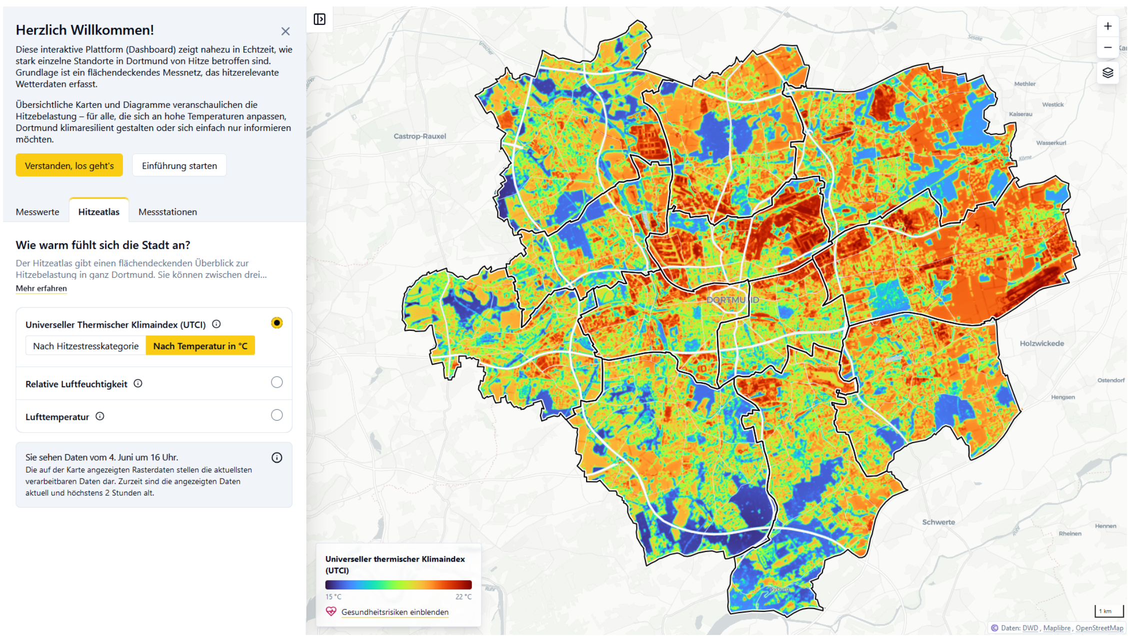

Fig. 1: A screenshot of the developed Dashboard for the city of Dortmund.

This document provides an overview of the D2R Nowcasting Service, including the data sources, processing steps, algorithms, and other relevant information.

Get started in the *Overview* section.

CHAPTER

ONE

OVERVIEW

1.1 Introduction

The data-driven urban climate adaptation project Data2Resilience (D2R) implements data-driven actions and services to enhance Dortmund's resilience to extreme heat. It deploys a state-of-the-art biometeorological sensor network and develops a nowcasting service for monitoring thermal comfort throughout the city. This real-time information aims to guide on-ground actions and services for smarter urban climate comfort planning, pioneering the integration of thermal comfort data within Dortmund's smart city ecosystem.

The nowcasting service provides real-time information about the weather and thermal comfort conditions in the city of Dortmund. This information is retrieved from the biometeorological weather stations and from Numerical Weather Predictions (NWP) from the ICON-D2 model, which is DWD's regional numerical weather prediction model. The processing of the data is done in the backend of the service and includes the following processes: simulation of the radiation fluxes and the mean radiant temperature (MRT); geostatistical interpolation of the network data; and calculation of a thermal comfort index, namely the UTCI, at street level resolution. The service's frontend provides an interactive webmap and relevant plots that present the thermal comfort in the city. An example of the UTCI data provided by D2R's nowcasting service over Dortmund's city center is given in Fig. 1.1.

The D2R nowcasting service is open source. At its core it uses the Urban Multi-scale Environmental Predictor (UMEP) climate service tool, which is designed for researchers and service providers (e.g. architects, climatologists, energy, health and urban planners), and in-house software developed by the Bochum Urban Climate Lab.

Fig. 1.1: An example of the UTCI data provided by D2R's nowcasting service over Dortmund's city center for 2024-06-24 12:00 UTC.

1.2 What is the problem D2R solves?

Extreme heat endangers human health and well-being and impairs the use of public spaces. Dortmund's Integrated Climate Adaptation Master Plan (MiKaDo) puts a priority on actions and measures that improve heat resilience. D2R implemented data-driven actions and services to support officials from the City of Dortmund in improving the city's resilience to extreme heat. We view D2R as a key step towards achieving the goal of climate-resilient and sustainable urban development that is set in MiKaDo. MiKaDo forms the basis for the implementation of local climate adaptation measures and recognizes extreme heat and thermal discomfort as major risks for the health and well-being of its citizens. It promotes the adoption of measures that protect climate-sensitive facilities (hospitals, nursing homes, etc.) from extreme temperatures and reduce the exposure of vulnerable groups. To achieve these objectives, spatially detailed data about the excess heat and thermal comfort—defined as the state-of-mind in which a person feels satisfied with the surrounding thermal environment (i.e., not too hot or too cold)—, across Dortmund are required, which are currently missing. The D2R project addresses this gap and provides the necessary spatial and meteorological data by co-deploying a state-of-the-art network of biometeorological weather stations designed to measure thermal comfort and a service that provides fine resolution nowcasts of thermal discomfort. More details can be found on the project website.

CHAPTER

TWO

INSTALL AND USE

2.1 Setup

The backend is developed in Python 3.12

For running the pipeline (see `src/process_next_timestep.sh`) as the backend service the following steps need to be taken:

1. Setup the structure

   a. The project runs in Python 3.12, which can be seen in the base image used in the Dockerfile for the backend, `backend.Dockerfile`.

   b. We import modules from a separate project called `UMEP-processing-fork`. It keeps the standalone files of the required UMEP tool along with the original code of the toolbox. (See also Section 4 in the documentation for a more detailed description of the methods and processes.)

   c. For frequently downloading NWP GRIB2 data from the DWD file server, the DWD `downloader` has to be installed (current version 0.2.0). This happens within `backend.Dockerfile` but requires a clone of that project next to `d2r-nowcast` (see folder structure below).

   d. For downloading relevant ICON-D2 data and cropping it to the city extent, we use the functions placed in `src/icon_d2`.

> **Note**
>
> DWD announced a change in accessing DWD open data Release Notes in April 2025, which also affects the `downloader` tool and expected data format used in this project. At some point the `downloader` will not be functional anymore due to changes in the data format and this step will have to be adjusted. The tool's repository was already converted to a public archive on github and no further changes or updates are expected.

2. The project structure now has to look like:

```
Common Parent Folder
├── d2r-nowcast
├── downloader
└── UMEP-processing-fork
```

3. Build the Docker image

   Build the repository into a Docker image using the following command from the parent folder of `d2r-nowcast`:

   `docker build -f d2r-nowcast/src/backend.Dockerfile -t d2r-backend:<version> .`

Note

The flag `-f` allows to define a specific Dockerfile to be used. The container tag indicated with the `-t` flag can be set individually but needs to be consistent throughout the references in later steps.

4. Setup a script to run the Docker container

   Create a script to run the Docker container at an arbitrary path `<path>/<to>/cronscript_<version>.sh`. To do this, copy and adjust the Docker run and Docker log commands in `src/cronscript_template.sh` to your needs. Adjust the paths to the datasets on your machine, set up available memory and CPU, etc.

   Note

   The SOLWEIG processes are run in multiprocessing mode, and the number of processes should match the defined available CPUs. Within the code, the number of processes is fixed to 32 in `umep_wrapper/solweig_multi_processing.py` with this line:

   ```
   with Pool(processes=32, initializer=worker_init, initargs=[q]) as pool
   ```

   The number of available CPUs is set with the `docker run` command in the above-mentioned script, using the `--cpus` flag. For our current setup, the following configuration is useful: `--memory=128gb --cpus=32`, but this can be adjusted according to your requirements and limitations.

5. Edit the crontab

   Edit the crontab for your user on your machine using `crontab -e` and set up an hourly cron job by adding the following line:

   ```
   0 * * * * <path>/<to>/cronscript_<version>.sh 1> /dev/null 2> <path>/<to>/<error_file>.err
   ```

   This command executes the prepared script from the previous step every hour, redirecting standard output (1>) to `/dev/null` and standard errors (2>) to `<path>/<to>/<error_file>.err`.

   There is a `src/cronscript_template.sh` as guide for the specific docker run command executed on cron call.

2.2 Structure

The project is structured as follows:

`src/umep_wrapper`:
Contains scripts that can be run separately on the server to generate new data and corresponding Dockerfiles.

`src/umep_wrapper/config-templates`:
Contains config templates to run algorithms of the UMEP toolbox: Wall, SVF, SOLWEIG.

`src/icon_d2`:
Contains scripts to download and pre-process ICON-D2 NWP data from the DWD for our purposes.

`src/interpolate`:
Contains scripts to interpolate data from the measurement network for our purposes.

`src/utils`:
Contains utility scripts.

`tests`:
Contains some tests for ensuring the functionality of the pipeline. Tests can be executed via `pytest -s -v` in a properly set up virtual environment.

CHAPTER

THREE

INPUT DATA

To model thermal discomfort, the nowcasting service utilizes raster data describing the urban form and meteorological conditions. The urban form data include various Digital Surface Models that represent the ground, building, and tree canopy heights, as well as the land cover and the Sky View Factor (SVF). The weather data should provide the meteorological conditions at a height of 2 meters above ground, including the near-surface air temperature, the relative humidity, and the wind speed. All input raster data have to be formatted according to the specifications used by UMEP's SOLWEIG model, a core component of the nowcasting service.

Note

Details of SOLWEIG's model inputs and outputs can be found here.

Warning

All the input variables should have the same extent and pixel resolution and be provided in a projected Coordinate Reference System (CRS).

The sections below provide a detailed list of the input data, as well as how they are organized into folders and subfolders.

3.1 Data Describing the Urban Form

The urban form refers to the physical layout of the city, including its three-dimensional structure and land cover. The urban form data used by D2R's nowcasting service as provided in Table 3.1:

Table 3.1: The urban form data used by D2R's nowcasting service

	Name	Description	Units	File type
1	DEM	Digital Elevation Model	meters a.s.l.	GeoTIFF
2	Building DSM	A DSM consisting of ground and building heights	meters a.s.l.	GeoTIFF
3	Canopy DSM	A DSM consisting of pixels with vegetation heights	meters a.g.l.	GeoTIFF
4	Land Cover	Categorizes the urban surface in the seven classes expected by UMEP (see here)		GeoTIFF
5	SVF*	The sky view factor (SVF) is the ratio of the visible sky area of a point in space to the total sky area. A dimensionless measure between zero and one, representing totally obstructed and free spaces, respectively		GeoTIFF
6	Wall Height*	The height of the wall pixels	meters a.g.l.	GeoTIFF
7	Wall Aspect*	The aspect of the wall pixels relative to the ground level. Wall aspect is given in degrees where a north facing wall pixel has a value of zero	degrees where a north facing wall pixel has a value of zero	GeoTIFF

Note

* The SVF, the Wall Height, and the Wall Aspect datasets should be created using the corresponding pre-processors from UMEP that require as input the using the ground and building, and canopy DSM (see here for the SVF and here for the Wall Height & Aspect).

The DEM, DSM, and land cover data used in Data2Resilience were sourced from GEOportal NRW (see DSM here and land cover here). For the building DSM, the heights were extracted from the availabe DSM at building sites according to the land survey register (ALKIS) and given in meter above sea level (a.s.l.). For canopy DSM, we extracted tree heights from the DSM with a mask generated from the tree map of Google's Environmental Insights Explorer and transformed the heights to be above ground level (a.g.l.), as SOLWEIG expects this. For land cover we manually reduced the 15 land cover classes for the dataset into those expected by SOLWEIG. These datasets have been resampled to a 3-meter regular grid, clipped to the same boundaries, and saved as GeoTIFFs. The Coordinate Reference System (CRS) is ETRS89 / UTM zone 32N (EPSG: 25832). The SVF, wall height, wall aspect, and daily shadow maps were derived using the appropriate UMEP methods, with the building and canopy DSM as inputs. To facilitate data processing and speed up calculations, the resulting files have been tiled using GDAL into 120 tiles, each with a size of 1000 x 1000 pixels and a 20% overlap. The tiling scheme used is shown in Fig. 3.1.

3.2 Data Describing the Weather Conditions

Assessing the outdoor human thermal comfort requires detailed information on radiation, wind, temperature, and humidity in the exposure environment. Acquiring this information with high spatial resolution is extremely difficult since the ambient environment is dynamic and highly variable over very short distances.

D2R has been collecting weather data from two different data sources.

3.2.1 Numerical Weather Predictions

It uses Numerical Weather Predictions (NWP) from DWD's ICOsahedral Nonhydrostatic model (ICON).

ICON is the global and regional numerical weather prediction model at DWD. D2R uses the ICON-D2 convection-permitting model setup, which is DWD's finest resolution weather model (see here). It covers Germany, Switzerland, Austria, and other neighboring countries with a grid cell resolution of approximately 2 km. Forecasts of ICON-D2 are performed 8 times a day (namely at 00, 03, 06, 09, 12, 15, 18, 21 UTC), with a forecast range of 48 h. The ICON-D2

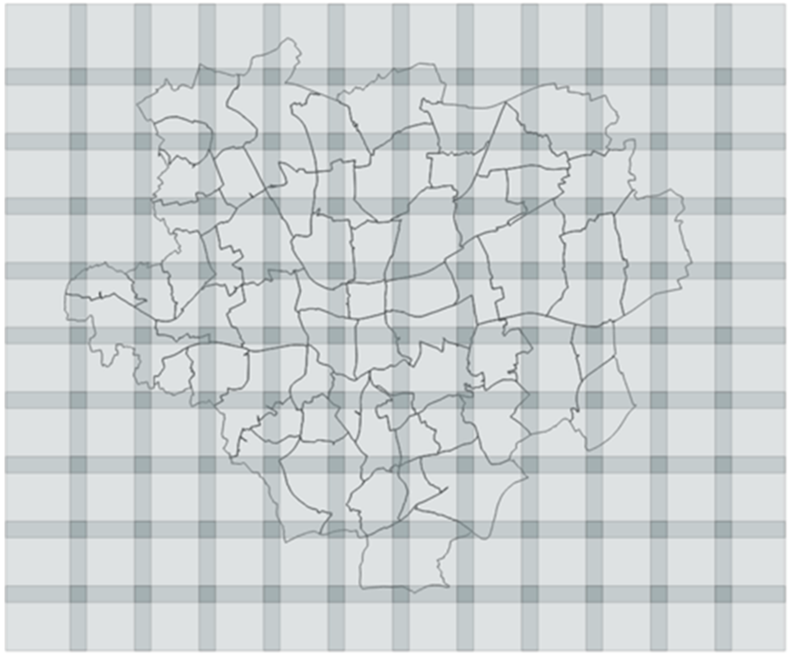

Fig. 3.1: The employed tiling scheme.

forecast data for each weather element are made available via opendata.dwd.de, both on a rotated grid and on a regular grid. D2R uses the following eight elements:

Table 3.2: Used variables of ICON-D2 model

Variable	Description
`t_2m`	The air temperature at a height of 2 m above ground (see Fig. 3.2).
`relhum_2m`	The relative humidity at a height of 2 m above ground (see Fig. 3.2).
`u_10m`	The East-West wind components
`v_10m`	The North-South wind components
`tot_prec`	The total precipitation
`asob_s`	The shortwave net radiation flux
`ASWDIR_S`	The downward solar direct radiation flux
`ASWDIFD_S`	The downward solar diffuse radiation flux

The `u_10m` and `v_10m` variables (both in m/s) in the ICON model represent the zonal (east-west) and meridional (north-south) wind speeds at 10 meters above the ground, respectively. A positive value for `u_10m` indicates that the wind is blowing from west to east (in the direction of increasing longitude), while a negative value for `u_10m` indicates that the wind is blowing from east to west (in the direction of decreasing longitude). Similarly, a positive value for `v_10m` indicates that the wind is blowing from south to north (in the direction of increasing latitude), while a negative value for `v_10m` indicates that the wind is blowing from north to south (in the direction of decreasing latitude). The wind speed is calculated from the `u_10m` and `v_10m` using the Pythagorean theorem.

To download the ICON-D2 data, a software tool has been developed using Python-3 that automatically downloads and pre-processes the ICON-D2 forecasts for Dortmund. In particular, the tool starts by downloading all the GRIB2 files for the selected fields (e.g., the 2 m air temperature, the relative humidity, etc.). It then clips the raster data to the city bounds, and stacks the corresponding forecasts into datacubes (one per field), which then saves as netCDF files in D2R's dedicated data repository.

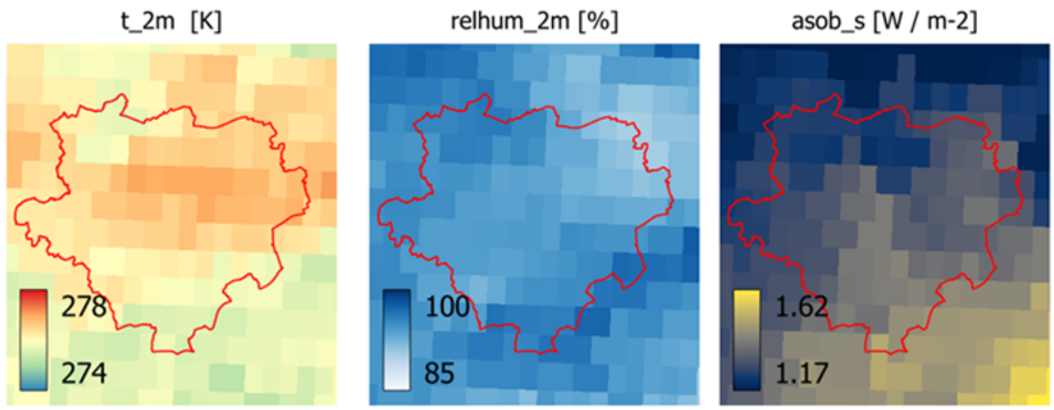

Fig. 3.2: An example of the `t_2m`, `relhum_2m`, and `asob_s` ICON-D2 data for a random day in Dortmund.

3.2.2 In-Situ Measurements

The second source of meteorological data is the weather station network in the city of Dortmund, also developed during the Data2Resilience project. This is a state-of-the-art meteorological weather station network with 80 stations at 76 locations installed throughout the city and at a height of around 3.30 meters. 55 stations measure air temperature and relative humidity while 25 measure a comprehensive set of biometeorological parameters including also radiation, wind and precipitation etc. Four locations are equipped with both sensor types to assess the accuracy of the different devices over time. See more details on the stations at the D2R Dashboard. With a measurement frequency of five minutes, the network allows to assess current weather situations in Dortmund with a high spatial and temporal resolution and therefore is the preferred source for this service.

From the network's api air temperature (°C) and relative humidity (%) information is retrieved at the beginning of each pipeline run (every hour).

The stations send their measurements via the LoRaWAN technology which is by design affected by packet loss. In case not enough data is available for forcing the SOLWEIG model, the NWP data is the back-up source.

Furthermore, the stations are not in sync regarding their measurement and sending time. To ensure that the data is up-to-date, latest data not older than 15 minutes is requested from all stations, which results in a list of a single measurement per station which has data available. To consider shadow movements (i.e. some passing of time) in the SOLWEIG radiation modeling later, the concept *latest data* is used but for the *previous hour*. For this, again data for a time window of 15 minutes is requested, but from the previous hour (for hourly model runs) and in case of multiple measurements available, only the latest of these is kept per station.

A short example:

- Current hour to model is 12:00, the latest data is requested via the `latest_data` endpoint of the network's API, which is not older than 11:45
- Previous hour is 11:00 and the data for this time is requested from the window 10:45 - 11:00 per station via the `data` endpoint.
- Note, that all times are given in UTC.

The measurement data is interpolated to generate rasters that reflect the distribution of the two parameters across the city (see Fig. 3.3). More information on the use of meteorologial data, the interpolation process and further processing is given in Section 4.

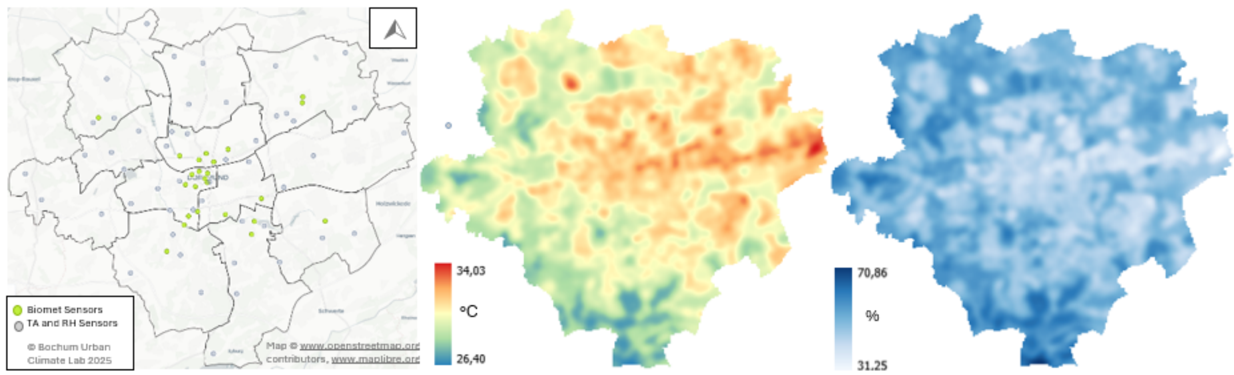

Fig. 3.3: A map of the station distribution across dortmund and the derived `air temperature` and `relative humidity` rasters for a random day in Dortmund.

3.3 File organization

The input files are organized into subfolders as shown in the directory tree below. The root directory includes two folders: the folder “dynamic” contains the variables that are continuously gathered or generated by the project, like the ICON-D2 NWP, while the folder “static” contains the variables that remain constant over time, such as the DSM. For some folders exemplary names of subdirectories or files are added.

```
Common Parent Folder
├── dynamic data
│       ├── icon-d2-data
│       │       └── YYYYMMDDHH
│       ├── metadata
│       ├── metfiles
│       ├── MRT
│       ├── RH
│       ├── SOLWEIG_3m_1000+200
│       ├── station_data
│       ├── TA
│       ├── UTCI
│       │       └── DO_UTCI_YYYY_DOY_HH_vX.Y.Z_cog.tif
│       └── UTCI_CLASS
└── static data
        ├── Building DSM
        ├── Canopy DSM
        ├── DTM
        ├── Land Cover
        └── SOLWEIG preprocess (SVF, Wall Height, Wall Aspect per tile)
```

CHAPTER

FOUR

METHODS AND PROCESSES

The approach to model thermal comfort is based on the UMEP-Toolbox developed by Lindberg et al.[1]. For thermal comfort assessment we selected the well-known and well-researched UTCI (Universal Thermal Climate Index, see Section 6.2), which is driven by the following four meteorological parameters:

- Near Surface Air Temperature (°C)
- Relative Humidity (%)
- Mean Radiant Temperature (°C)
- Wind (10m above ground) (m/s)

From NWPs and weather observations we can derive three of these parameters; Mean Radiant Temperature is driven by radiative forcing and calculated in our case via the SOLWEIG model[2] from the UMEP-Toolbox. SOLWEIG is applied in a tile-wise manner due to the large data volume, following the UMEP's recommendation of tiling the input data before running the SOLWEIG model:

> *This plugin is computationally intensive i.e. large grids will take a lot of time and very large grids will not be possible to use. Large grids e.g. larger than 4000000 pixels should preferably be tiled before.*[3]

For running the tool in our targeted resolution we tile the city using a regular overlapping grid (see Fig. 3.1). The overlap is needed because neighboring buildings create shadows which are necessary to consider during the radiation modeling and might be split in two or more tiles. The workflow of the service's backend is shown in Fig. 4.1. In the following subsections, the various processes (orange boxes in Fig. 4.1) are described in the order of application.

4.1 Processes

4.1.1 Preprocessing

For running SOLWEIG two steps of preprocessing are required, namely the creation of the SVF and wall height and aspect maps, which are later used to calculate the 3D radiation fluxes. Both processes can be run via the UMEP-Toolbox as well. Both need the combined Building and Ground DSM to derive the respective information.

See their respective manuals from UMEP for details:

- SkyViewFactor Manual
- Wall Height and Aspect

[1] Fredrik Lindberg, Ting Sun, Sue Grimmond, Yihao Tang, Nils Wallenberg. 2018 - today, Public Repository https://github.com/UMEP-dev/UMEP-processing, and documentation https://umep-docs.readthedocs.io/en/latest/

[2] Fredrik Lindberg, Björn Holmer, Sofia Thorsson, SOLWEIG 1.0 – Modelling spatial variations of 3D radiant fluxes and mean radiant temperature in complex urban settings, 2008, International Journal of Biometeorology

[3] UMEP Manual. Outdoor Thermal Comfort: SOLWEIG, UMEP developers, https://umep-docs.readthedocs.io/en/latest/processor/Outdoor%20Thermal%20Comfort%20SOLWEIG.html

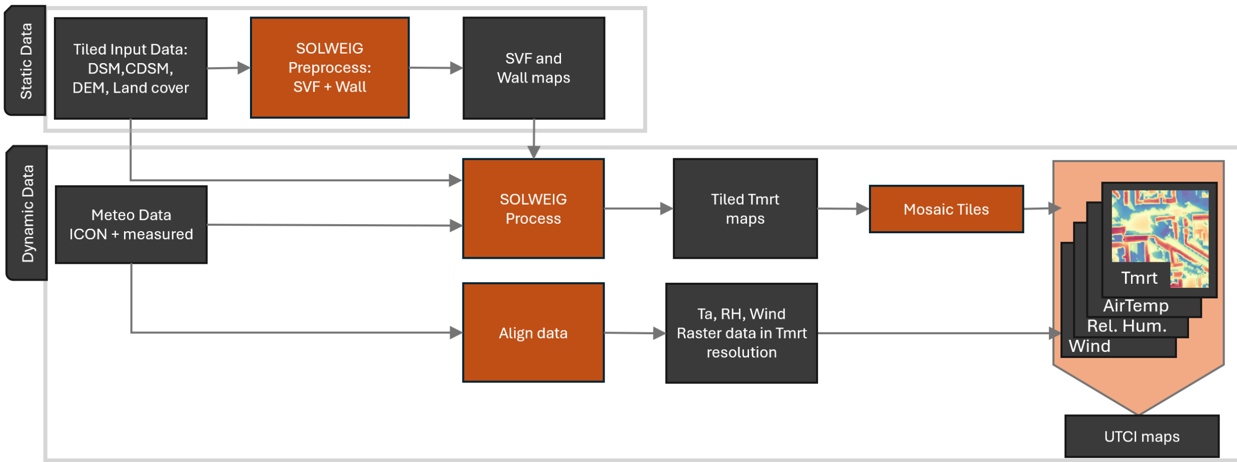

Fig. 4.1: The backend workflow. The grey color is used to represent data and the orange is used for processes.

4.1.2 Alignment of Weather Data

As described in the previous chapter (see Section 3.2) and also in Fig. 4.1 two sources of weather information namely NWP and in-situ measurements are used.

For Numerical Weather Predictions (NWP) we use the data by the ICON-D2 model of the DWD ICON-D2 creates a horizontal resolution of ~2 km for Germany and to some extent covers regions across the German borders. Since February 5 2025 the ICON-D2 NWP product is running with the TERRA_URB module and therefore incorporates effects of the urban canopy layer (see Change log, section 2). The NWP data of the listed parameters in Table 3.2 (see Section 3.2) is therefore cropped to the city boundaries and supersampled to match the resolution of the surface models. Since the 2 km resolution only gives average values for the area, local in-situ measurements are additionally considered to get a more precise view.

The meteorological measurement network is installed in the city area and measures weather parameters at 76 locations. To generate a sensible rasters that considers measurements in context of the urban form we developed a interpolation module which is explained in detail in the next section. In short, this module checks the availability of enough recent measurements from the network and if so follows a method to generate these meaningful rasters of **Air Temperature and Relative Humidity** (see Fig. 3.3). If not, the ICON-D2 data for these two parameters is used as a back-up option, for which the parameter values at the different station locations are extracted from the supersampled ICON-D2 rasters.

The other relevant meteorological parameters for forcing SOLWEIG are taken from the ICON-D2 data.

4.1.3 Geostatistical Interpolation Module

D2R's nowcasting service relies on air temperature (Tair) and relative humidity (RH) measurements from the biometeorological weather station network to calculate the UTCI rasters. To accomplish this, the station data are converted into a continuous raster surface covering the entire city, following the workflow outlined in Fig. 4.2. The primary method used is Regression Kriging, which is a geostatistical method that combines two steps: a regression model that relates the target variable (e.g., Tair) to auxiliary variables (predictors), and kriging to interpolate the residuals of that regression, enhancing the spatial accuracy of the final surface.

4.1.3.1 Design of the Interpolation Module

To ensure robust, uninterrupted operation, even in cases of data failure or degraded conditions, the interpolation module incorporates 3.5 distinct processing paths. These paths act as fallback options, guaranteeing that a UTCI raster can always be generated. The primary methods are Path 1.0 and Path 1.5 (both shown in pink in Fig. 4.2). Path 1.0 performs a direct regression from station data to a continuous surface, while Path 1.5 extends this approach by kriging the residuals of the regression. Path 2.0 (blue path) skips the regression step if its performance is subpar and applies

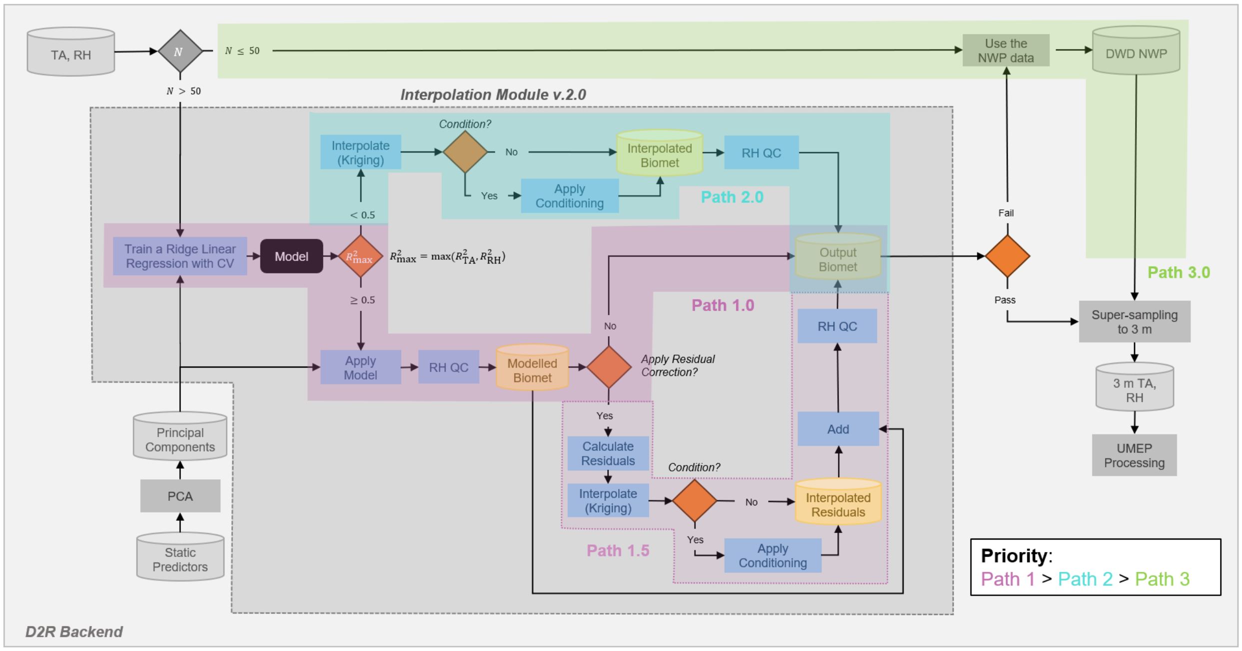

Fig. 4.2: The workflow of D2R's geostatistical interpolation module.

kriging directly to the station data. As a last resort, Path 3.0 (green path) bypasses station data entirely and instead supersamples the Tair and RH fields from the ICON-D2 numerical weather prediction model provided by the German Weather Service (DWD).

The selection of the most appropriate path is determined by a series of internal checks placed at different workflow steps, illustrated as orange diamonds in Fig. 4.2. These checks primarily rely on accuracy statistics, specifically the R2 value, calculated using a 4-fold cross-validation scheme. Cross-validation is a statistical technique used to evaluate a model's reliability by repeatedly splitting the dataset into training and validation subsets. Given the limited number of stations (N = 80), a separate verification subset is not created. The workflow follows a predefined priority order: Paths 1.0 and 1.5 are the preferred methods, followed by Path 2.0, with Path 3.0 serving as the final fallback. The decision between Paths 1.0 and 1.5 is controlled by the *RESCORR* flag—if set to true, the workflow adopts Path 1.5, utilizing regression kriging instead of simple regression. This flexible, multi-path approach ensures high reliability and adaptability of the UTCI production process, even under varying data conditions.

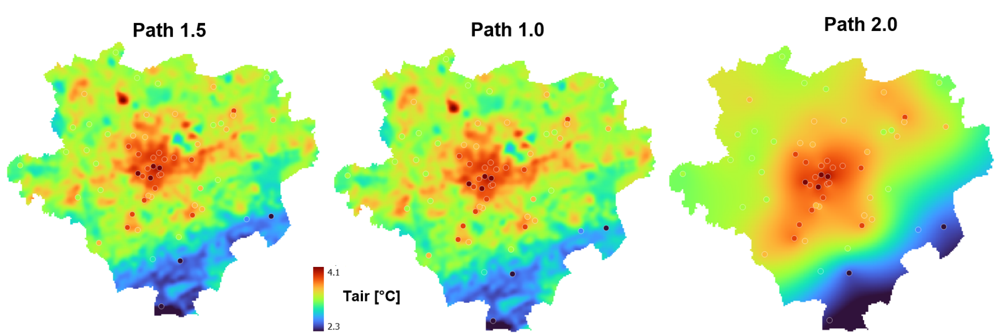

Fig. 4.3: Example of the Tair predictions from Paths 1.5 (left), 1.0 (middle), and 2.0 (right) of D2R's interpolation module. The point markers indicate the corresponding weather station measurements that have been used as input (observation date: February 6, 2025, at 06:30 UTC).

An example of the output generated by Paths 1.5, 1.0, and 2.0 is presented in Fig. 4.3. It is evident that Paths 1.0

and 1.5, which incorporate auxiliary surface information, such as building heights, produce spatial fields that are more physically realistic and better aligned with expected urban microclimate patterns. In contrast, the output from Path 2.0 appears noticeably smoother, lacking the finer spatial variability introduced by the auxiliary predictors. Despite these differences in spatial detail, all three methods show strong agreement with the input station data, as indicated by the point markers in Fig. 4.3.

The auxiliary information used in the interpolation process, shown in Fig. 4.4, consists of 23 predictors that capture key characteristics of the urban environment. These include variables describing the built-up and vegetation fractions, the building heights, the topography, the sky-view factor (SVF), the climatopes (i.e., areas with similar microclimatic conditions), the cold-airflow patterns, and the geographic coordinates (longitude and latitude) of each grid cell. All predictors have been precomputed at a spatial resolution of 100 meters using high-resolution geospatial data from NRW or generated by the D2R consortium.

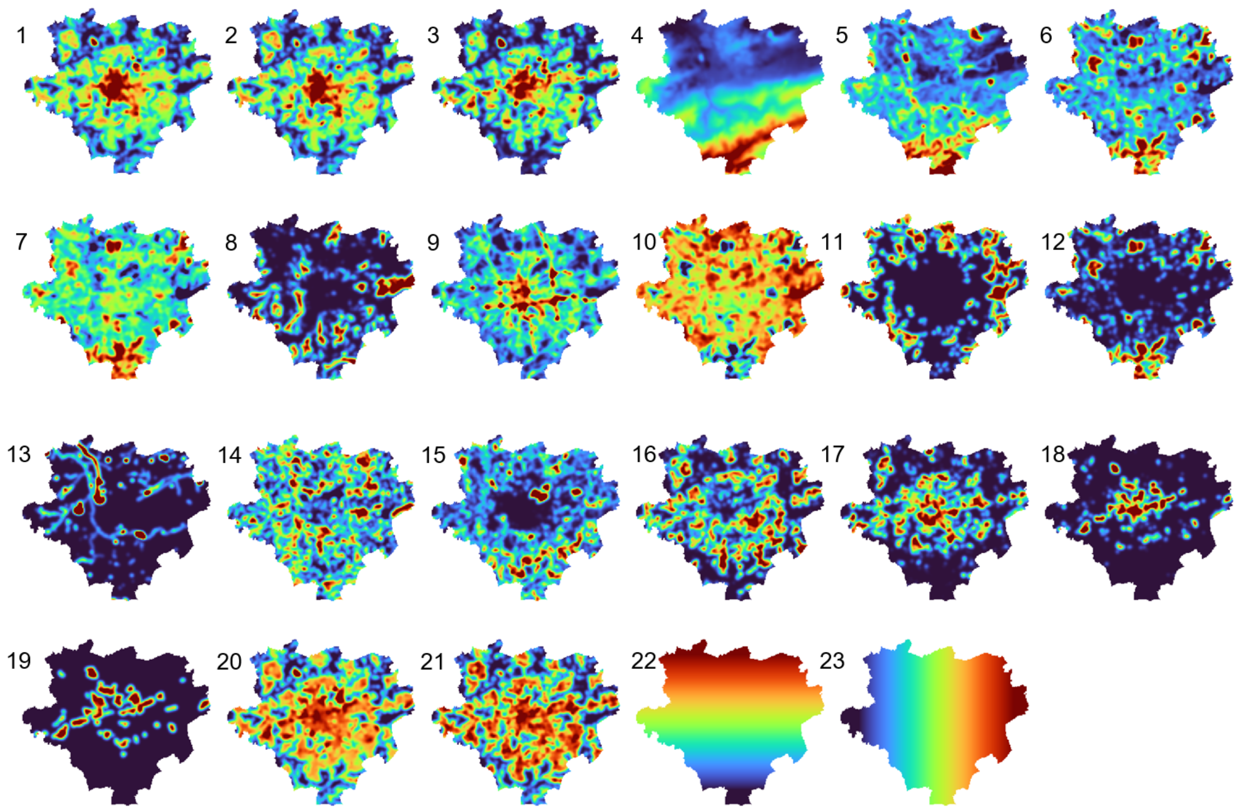

Fig. 4.4: The 23 predictors used by D2R's interpolation module, namely: (1) mean building height; (2) SD of building heights; (3) build-up fraction; (4) mean elevation; (5) SD of elevation; (6) vegetation fraction; (7) mean vegetation height; (8) cold airflow pool fraction; (9) road fraction; (10) mean SVF of urban canyons; (11-18) climatopes; (19) industrial areas fraction; (20) mean building area; (21) SD of building area; (22) latitude; and (23) longitude.

4.1.3.2 Workflow Description

The interpolation workflow (Fig. 4.2) operates as follows: when at least n weather stations (default: 50) report valid measurements, the module begins by extracting the corresponding auxiliary predictor values at each station location. These data are then used to train two separate linear machine learning (ML) models—one for Tair and one for RH. The selected model is a Ridge regressor, a form of linear regression that applies L2-norm regularization to reduce the risk of overfitting. This approach is well-suited for the relatively small dataset available (up to 80 stations), where more complex models would likely underperform due to insufficient training data.

After training, both models are evaluated using 4-fold cross-validation, and their performance is quantified using the R^2 score. The first decision point in the workflow is based on this evaluation: if at least one of the two models achieves an R2 score above a defined threshold (default: 0.5), the workflow proceeds to the next step of the Path 1.x sequence. If neither model meets the threshold, the workflow bypasses regression-based methods and defaults to Path 2.0, which uses kriging alone for spatial interpolation. If the initial check is passed, the next decision depends on the value of the RESCORR flag. If RESCORR is False, the Ridge regression output is saved directly as a GeoTIFF, and the process terminates. If RESCORR is True, the module continues with Path 1.5.

In Path 1.5, the workflow calculates the residuals as the difference between the Ridge model predictions and the actual station observations and then interpolates these residuals using Universal Kriging. The kriged residuals are added back to the initial predictions, improving their agreement with the input station data, and the resulting file is saved as a GeoTIFF file. An example illustrating the initial Ridge predictions, the interpolated residuals, and the final corrected output produced by Path 1.5 is shown in Fig. 4.5. For both Path 1.0 and Path 1.5, the output GeoTIFFs are accompanied by a JSON file that includes the computed R2 scores for both models, the identifier of the path used, and a quality flag indicating whether the result is considered reliable or not.

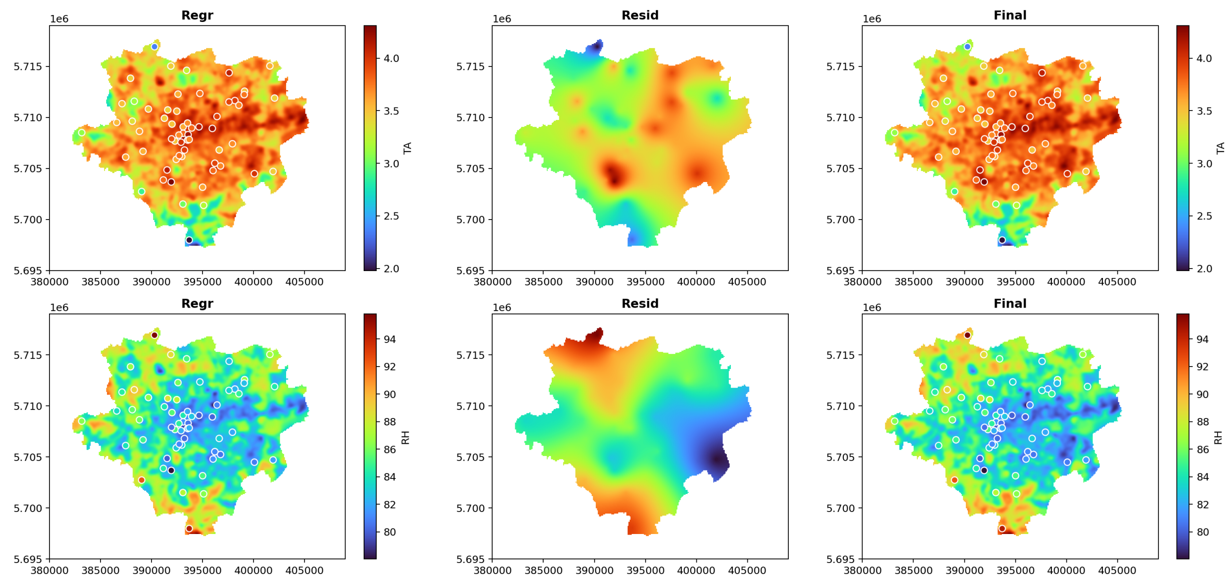

Fig. 4.5: Example outputs from Path 1.5 for Tair (top row) and RH (bottom row). Each row illustrates: the initial Ridge regression predictions (left), the interpolated residuals obtained via Universal Kriging (center), and the final corrected output (right).

The Universal Kriging workflow used in both Path 1.5 and Path 2.0 follows the same three-step process; the only difference lies in the input data provided. In the first step, the empirical semivariogram is computed separately for each variable (i.e., Tair and RH) based on the input data. The semivariogram quantifies how the similarity between data points decreases with increasing distance. In the second step, an Exponential model with a nugget effect is automatically fitted to the empirical semivariogram, which provides a smooth mathematical representation of the spatial variability. In the final step, the fitted model is used to perform the spatial interpolation across the defined spatial grid, generating the continuous surface data. D2R's interpolation module includes support to condition the kriging result, which allows for the generation of multiple plausible spatial fields that honor both the input data and the underlying spatial structure. This feature is controlled by the N_REALIZ flag, which by default is set to 0, meaning no conditional simulations are performed unless explicitly requested.

4.1.4 SOLWEIG Processing

At this point all the static and dynamic data required to run SOLWEIG are available. The dynamic aspect of the sun's position is handled internally in SOLWEIG by deriving it from the Geolocation of an input layer, the given Day of the Year (DOY) and time of the day (obtained from the weather data). The UMEP-Toolbox QGIS-plugin which includes SOLWEIG was adjusted for this processing step to run it in a standalone way and to enable automation. Due to computational limitations of SOLWEIG the input data is tiled (see Fig. 3.1). Each tile is associated with specific input data, i.e. snippets from the full surface models and respective preprocessed data sets. These form an input data set for SOLWEIG which is described in a structured YAML-formatted configuration file. It also contains further information on setting up SOLWEIG, which is identical to the settings made via manual input in QGIS' UMEP module. Based on the row and column of the tiling grid (see Fig. 3.1) each tile gets assigned a unique ID , which allows automatic filling of the fields in the configuration file and following automated setup and running of SOLWEIG.

Further information on the SOLWEIG model and it's handling in QGIS can be found here:

- Manual: Outdoor Thermal Comfort (UMEP team)
- Tutorial: Introduction to SOLWEIG (UMEP team)
- UMEP Video Series on Youtube (Carlos Bartesaghi Koc)

4.1.5 Mosaicing Tiles

Mosaicing of tiles is basically the reverse process of tiling, all tile-wise results are stitched together to a complete, city-wide raster. For this process the initial overlap was designed to ensure the correct representation of shadows and radiation in the overlapping areas, i.e. edge cases.

4.1.6 Calculating Thermal Comfort

Finally, for calculating Universal Thermal Climate Index as Thermal Comfort index the python package thermal-comfort is used.

> *The thermal-comfort package wraps a few common thermal-comfort functions from official sources such as ISO-norms or VDI-Guidelines in python. The underlying functions are implemented in fortran to achieve blazingly fast performance on large arrays.*

Learn more in the related documentation.

Generally, the raster for Air Temperature, Relative Humidity, Wind and Mean Radiant Temperature are read as arrays, UTCI is calculated for each array position and finally stored as a separate UTCI raster.

Note, that wind values are not simulated separately for the area but supersampled from the ICON-D2 NWP data.

4.2 API

The pipeline results are converted to Cloud Optimized GeoTIFFs (COG, see the project's website) and stored in a terracotta database which also is run as a tile server. The tiles are made available via the Data2Resilience project's d2r-api. See also the terracotta API description hosted there as well, to learn more about how to access the data.

For the key `param` the options are:

- `TA`: Air Temperature raster depending on processing path (see Section 4.1.3 above) either supersampled from ICON-D2 or interpolated from in-situ measurements,
- `RH`: Relative Humidity raster depending on processing path (see Section 4.1.3 above) either supersampled from ICON-D2 or interpolated from in-situ measurements,
- `MRT`: Mean Radiant Temperature raster mosaiced from SOLWEIG model outputs,
- `UTCI`: UTCI raster resulting from thermal-comfort step (see Section 4.1.6) and
- `UTCI_CLASS`: Classified UTCI values into heat stress categories.

For `year`, `doy` (Day of year) and `hour` timestamps in UTC are used.

CHAPTER

FIVE

OUTPUT FORMAT

This section describes the technical characteristics of the output data. D2R's nowcasting service generates raster data that represent various weather variables and a thermal discomfort index. The weather variables are the: - Near Surface Air Temperature (°C) - Relative Humidity (%) - Mean Radiant Temperature (°C) while the discomfort index is: - Universal Thermal Climate Index (UTCI, °C), which is also given in a classified form. Each variable and index are stored as a separate GeoTIFF file

5.1 Filenames

The filenames of the results files adhere to the following template:

`DO_{varname}_{year}_{doy}_{hour}_v{version}{suffix}.{extension}`

This template includes the following keyholders:

- `{varname}`: The variable name, e.g. MRT, UTCI, etc.
- `{year}`: The year, e.g. 2024, 2025, etc.
- `{doy}`: The day-of-year, e.g. 001, 002, … 365
- `{hour}`: The hour of the day in UTC time, e.g. 00, 01, 02, … 23
- `{version}`: The backend code version, e.g. v0.0.0, v1.1.2, etc.
- `{suffix}`: An optional underscore-prefixed suffix for the file such as *_cog* for Cloud Optimized GeoTIFFs
- `{extension}`: The file extension, e.g. tif

For example:

`DO_MRT_2025_152_12_v0.8.0_cog.tif` is the Dortmund Mean Radiant Temperature GeoTIFF image generated for 2025-06-01 (day-of-year 152) at 12:00 UTC with the code version 0.8.0 converted to a Cloud Optimized GeoTIFF.

5.2 Spatial Coverage

The resulting data cover the administrative boundary of the city of Dortmund, as shown below:

The bounding box coordinates in ETRS89 / UTM zone 32N are given in the table below.

Table 5.1: Bounding Box Coordinates of the input data in ETRS89 / UTM zone 32N

Min X [m]	Min Y [m]	Max X [m]	Max Y [m]
380150.0	5694550.0	406850.0	5718850.0

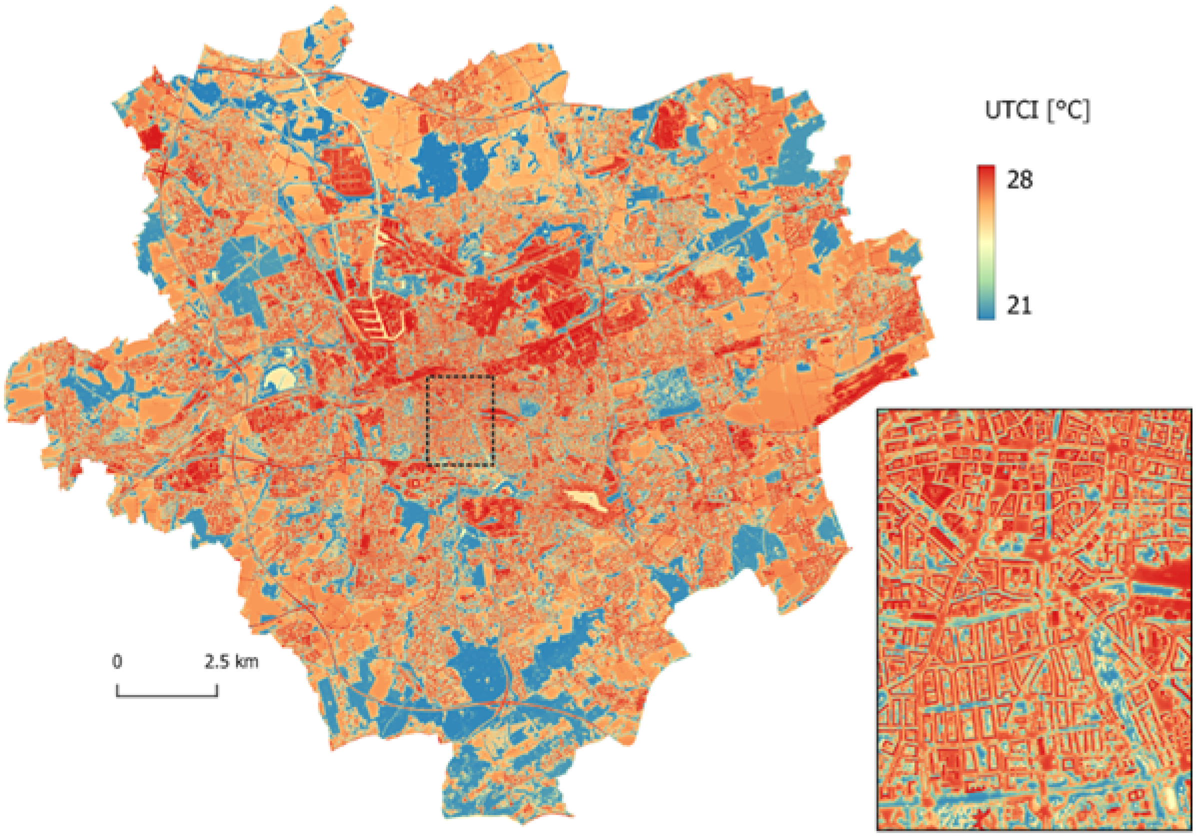

Fig. 5.1: The spatial coverage of the data produced by D2R's nowcasting service.

5.3 Spatial Resolution

D2R aims to map thermal comfort across the city at the street level. To achieve this, the resulting files have a very high spatial resolution of 3 square meters per pixel. The spatial resolution of gridded data refers to the dimension of the grid cell size representing the area covered on the ground. As such, it determines the level of detail that can be represented by the raster data: the finer the resolution of a raster, the smaller the cell size and, thus, the greater the detail. This fine resolution level allows for a very detailed representation of the urban features, such as buildings and trees, that govern the thermal comfort of humans at street level.

5.4 Temporal Resolution and Coverage

The resulting data are generated in almost real-time in hourly steps both for daytime and nighttime. The data are generated only for the warm period of the year, namely from April (~ Day of year: 106) to October (~ Day of year: 300).

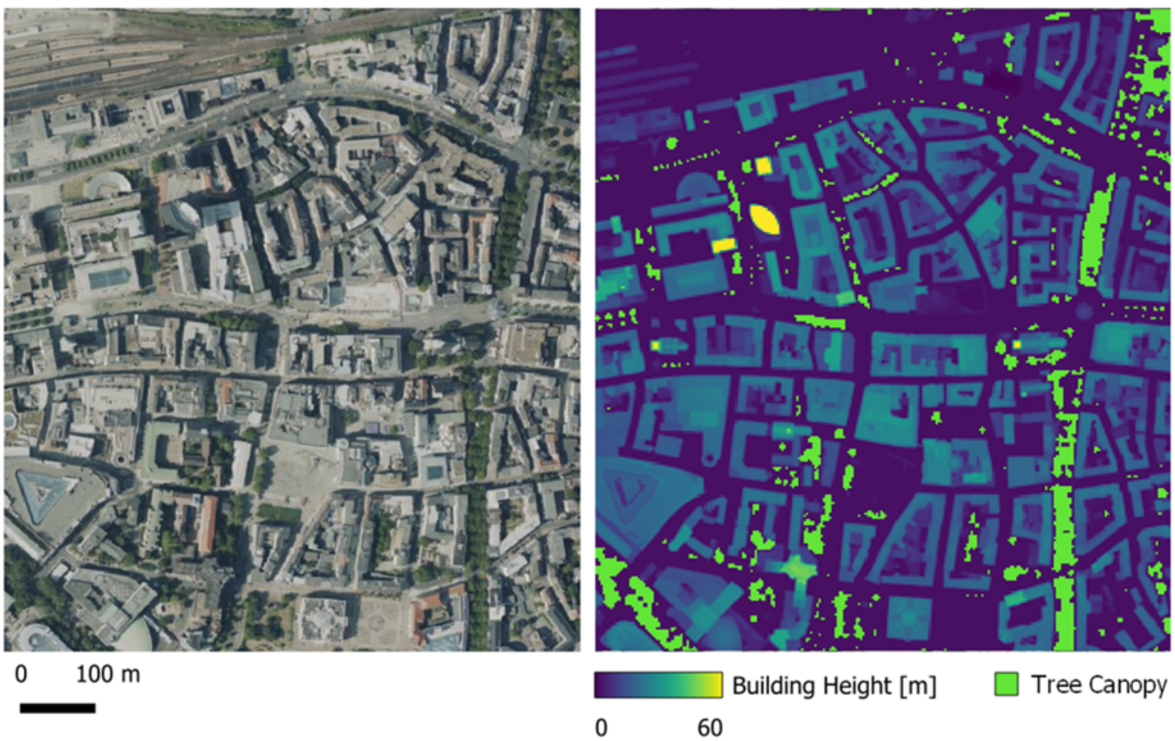

Fig. 5.2: Example of how the buildings and trees in Dortmund's city center are represented as raster data with a 3-meter spatial resolution (left: Google, right: derived from GeoPortal NRW).

CHAPTER

SIX

SCIENTIFIC BACKGROUND

This section covers the topic of scientific background, what is thermal comfort, what drives it and how it can be modeled.

6.1 Thermal Comfort

The ANSI/ASHRAE Standard[1] defines Thermal Comfort as follows:

> "Thermal comfort is defined as the condition of mind that expresses satisfaction with the thermal environment and is assessed by subjective evaluation."

Oke et al.[2] complement this definition by emphasizing the individual, psychological, and cultural aspects of comfort assessment, which further complicate an objective valuation of this condition. This becomes even more difficult when considering the different conditions of spaces where people spend time, such as a constant indoor climate or changing microclimates outdoors. The complexity resulted in a variety of indices, which were comprehensively reviewed in the literature in general[3][4][5] and for specific regions[6][7][8].

One of the most prominent indices is the Universal Thermal Climate Index (UTCI)[9]. It is driven by the four meteorological variables: Air Temperature, Relative Humidity, Wind Speed, and Radiation. Human behavior also plays a role in the sensation of thermal comfort in terms of clothing selection, health condition, activity, etc. These factors are considered and parameterized differently in different indices.

6.2 UTCI

The UTCI is categorized as a rational model[2] and considers the Fiala model for the biophysical processes, but uses an *outdoor reference* situation and a walking person[2][9] .

"The UTCI is defined as the air temperature [..] of the reference condition causing the same model response as actual conditions."[9]

[1] "ANSI/ASHRAE Standard 55-2010, Thermal Environmental Conditions for Human Occupancy"

[2] "T. R. Oke, G. Mills, A. Christen, and J. A. Voogt, Urban Climates. Cambridge: Cambridge University Press, 2017. https://doi.org/10.1017/9781139016476"

[3] "I. Charalampopoulos, A comparative sensitivity analysis of human thermal comfort indices with generalized additive models. Theor Appl Climatol 137, 1605–1622, 2019. https://doi.org/10.1007/s00704-019-02900-1 "

[4] "M. Migliari, R. Babut, C. De Gaulmyn, L. Chesne, O. Baverel, The Metamatrix of Thermal Comfort: A compendious graphical methodology for appropriate selection of outdoor thermal comfort indices and thermo-physiological models for human-biometeorology research and urban planning, Sustainable Cities and Society, Volume 81, 2022, https://doi.org/10.1016/j.scs.2022.103852. "

[5] "J. Shaeri, A Comparative Study of Outdoor and Indoor Thermal Comfort Indices, 2023, http://dx.doi.org/10.2139/ssrn.4489293 (preprint)"

[6] "F. Binarti, M. D. Koerniawan, S. Triyadi, S. Sesotya Utami, A. Matzarakis, A review of outdoor thermal comfort indices and neutral ranges for hot-humid regions, Urban Climate, Volume 31, 2020, https://doi.org/10.1016/j.uclim.2019.100531. "

[7] "S. Patle, V.V. Ghuge, Evolution and performance analysis of thermal comfort indices for tropical and subtropical region: a comprehensive literature review. Int. J. Environ. Sci. Technol., 2024. https://doi.org/10.1007/s13762-024-05703-8. "

[8] "Z. Tao, X. Zhu, G. Xu, D. Zou, G. Li, A Comparative Analysis of Outdoor Thermal Comfort Indicators Applied in China and Other Countries. Sustainability, 2023, 15, https://doi.org/10.3390/su152216029. "

[9] "K. Blazejczyk, G. Jendritzky, P. Bröde, J. Baranowski, D. Fiala, G. Havenith,Y. Epstein, A. Psikuta, B. Kampmann, An introduction to the Universal Thermal Climate Index (UTCI), Geographia Polonica, 86, 2013, https://doi.org/10.7163/GPol.2013.1. "

So the approach of calculating UTCI is similar to that of PET, but the reference conditions and the underlying energy model differ. For UTCI, the reference conditions are[Page 23, 2]:

- Outdoor reference where the individual is walking.
- Tair = TMRT (Air Temperature equals Mean Radiant Temperature).
- Wind Speed (V) is 0.5 m/s.
- Relative Humidity is 50% (capped at 20 hPa).

UTCI comes with a classification of the temperature ranges into thermal (dis)comfort categories. See the table below for the different ranges:

Table 6.1: UTCI classification into thermal stress levels (derived from[Page 23, 9]).

UTCI [°C]	Description
[< -40]	Extreme Cold Stress
[-40 to -27]	Very Strong Cold Stress
[-27 to -13]	Strong Cold Stress
[-13 to 0]	Moderate Cold Stress
[0 to 9]	Slight Cold Stress
[9 to 26]	No Thermal Stress
[26 to 32]	Slight Heat Stress
[32 to 38]	Moderate Heat Stress
[38 to 46]	Strong Heat Stress
[> 46]	Extreme Heat Stress

CHAPTER

SEVEN

LIMITATIONS

The developed tool consists of several processing steps, relies on preprocessed data and existing tools. The selection of tools and design of the process allows an execution in near real-time with also limited computing capacities.

Therefore the resulting rasters have some limitations and have to be treated with respect to the different accuracies:

- The interpolation of in-situ measurements considers urban topography but does not model microclimatic processes of air temperature and relative humidity.
- Urban wind fields are not given in a high resolution, therefore the given wind speed from ICON-D2 is used.
- ICON-D2 incorporates urban effects with the TERRA_URB module, but has a resolution of about 2 km.
- For the original SOLWEIG an RMSE of 4.8K was repoorted in 2008[1]; since then multiple upgrades were incorporated, such as the vegetation scheme, which reduced the RMSE to 3.1K[2].
- The processing pipeline is run once per hour on a 3m grid.
- Information on the urban form is rarely available and only some land cover classes can be considered with SOLWEIG.

The final accuracies for this service are currently being investigated.

[1] Fredrik Lindberg, Björn Holmer, Sofia Thorsson, SOLWEIG 1.0 – Modelling spatial variations of 3D radiant fluxes and mean radiant temperature in complex urban settings, 2008, International Journal of Biometeorology

[2] Fredrik Lindberg, C. Sue Grimmond, The influence of vegetation and building morphology on shadow patterns and mean radiant temperatures in urban areas: model development and evaluation, 2011, Theoretical and Applied Climatology

