https://doi.org/10.5194/gmd-2023-150
Preprint. Discussion started: 22 August 2023

Geoscientific Model Development Discussions

GEO4PALM v1.1: an open-source geospatial data processing toolkit for the PALM model system

Dongqi Lin[1], Jiawei Zhang[2], Basit Khan[3], Marwan Katurji[1], and Laura E. Revell[4]

[1]School of Earth and Environment, University of Canterbury, Christchurch, New Zealand
[2]Scion, New Zealand Forest Research Institute Limited, Christchurch, New Zealand
[3]Institute of Meteorology and Climate Research, Atmospheric Environmental Research (IMK-IFU), Karlsruhe Institute of Technology (KIT), Garmisch-Partenkirchen, 82467, Germany
[4]School of Physical and Chemical Sciences, University of Canterbury, Christchurch, New Zealand

Correspondence: Dongqi Lin (dongqi.lin@canterbury.ac.nz)

Abstract. A geospatial data processing tool, GEO4PALM, has been developed to generate geospatial static input for the Parallelised Large Eddy Simulation (PALM) model system. PALM is a community-driven large eddy simulation model used for atmospheric and environmental research. Throughout PALM's 20-year development, research interests have been increasing in its application in realistic conditions, especially for urban areas. For such applications, geospatial static input is essential. Although abundant geospatial data are accessible worldwide, currently, no tools are available in the community to automatically generate PALM's geospatial static input for most areas of the world. In addition, geospatial data availability and quality are highly variable and inconsistent. Therefore, PALM users face significant challenges, particularly regarding data acquisition, data pre-processing, and conversion of all information and metadata into PALM Input Data Standard (PIDS). This paper presents and describes a free and open-source tool, GEO4PALM, to help users generate the PALM static input with a simple and standardized process. GEO4PALM is compatible with geospatial data obtained from any source provided that the data sets comply with common geo-information formats. Users can either provide existing geospatial data sets or use the embedded data interfaces to download geo-information data from free online sources for any global geographic area of interest. All online data sets incorporated in GEO4PALM are globally available with several data sets having the finest resolution of 1 m. Two application examples are presented to demonstrate successful PALM simulations driven by geospatial input generated by GEO4PALM using different geospatial data sources for Berlin, Germany and Christchurch, New Zealand. GEO4PALM provides an easy and efficient way for PALM users to configure and conduct PALM simulations for applications and investigations such as urban heat island effects, air pollution dispersion, renewable energy resourcing, and weather-related hazard forecasting.

1 Introduction

The Parallelised Large Eddy Simulation (PALM) model is a Large Eddy Simulation (LES) model that has been used for Atmospheric Boundary Layer (ABL) research for over 20 years (Maronga et al., 2015, 2020; Raasch and Schröter, 2001). PALM is a free and open-source model that has high scalability to simulate atmospheric flows from the mesoscale (10 to 200 km) to microscale (1 cm to 1 km). To resolve land surface physics, PALM provides several features including the Radiative

https://doi.org/10.5194/gmd-2023-150
Preprint. Discussion started: 22 August 2023

Transfer Model (RTM; Krč et al., 2021), Land Surface Model (LSM; Gehrke et al., 2021), Urban Surface Model (USM; Resler et al., 2017), and Plant Canopy Model (PCM; Maronga et al., 2020). Over the last few years, PALM has been extensively developed for various microscale and mesoscale applications, especially for wind energy and urban applications. With the implementation of new modules, PALM has been used to study wind turbine wake in a German wind farm (Vollmer et al., 2017) and urban environments, such as ventilation in the city of Hong Kong (Gronemeier et al., 2017), urban air quality and pollutant dispersion in Cambridge, United Kingdom (Kurppa et al., 2019), Helsinki, Finland (Kurppa et al., 2020), and Bergen, Norway (Wolf et al., 2020, 2021). In addition, several studies (e.g. Belda et al., 2021; Resler et al., 2021; Salim et al., 2022) have carried out sensitivity analysis and simulations to validate PALM in urban environments. Geospatial data sets that describe the land surface characteristics and provide ground surface boundary conditions to the atmospheric model are critical for realistic simulations especially for microscale and/or urban climate studies.

The exchanges of energy and moisture between the surface and the atmosphere are impacted by physical characteristics of land surface, such as urban canopy, plant canopy, topography, and land use, across a spectrum of spatial and temporal scales throughout the atmospheric boundary layer (ABL; e.g. Bou-Zeid et al., 2004; Maronga et al., 2014; Rihani et al., 2015; Srivastava et al., 2020). To include, resolve, and realise the near-surface physical characteristics, PALM's current initialization setup allows users to provide a netCDF static driver strictly formatted in PALM Input Data Standard (PIDS) as an input (hereafter referred to as the static driver). In PALM, geospatial data should be processed and stored in a static driver to carry out simulations. Heldens et al. (2020) described the data requirements of the static driver for PALM and provided the PALM Create Static Driver tool (hereafter PALM CSD) to generate a static driver for PALM using geospatial data. However, the data processing routine provided by Heldens et al. (2020) can only be applied to the specific data sets prepared for three cities in Germany (Stuttgart, Berlin, and Hamburg) described in their study. Big hurdles still exist to apply PALM with realistic land surface characteristics in other regions. Numerous geospatial data are compatible and can be used to generate the PALM static driver while the spatial coverage, resolution, data quality, and data format could vary. In addition, the final conversion to PALM-readable formats requires extra processing. These issues go beyond the understanding and knowledge of physical processes and may have prevented further applications of PALM in the community. Furthermore, the lack of a common static driver preparation tool likely hinders the reproducibility of scientific results across different research groups.

Looking at the history of Numerical Weather Prediction (NWP) models, the Weather Research and Forecasting (WRF) model is arguably one of the most popular numerical atmospheric models in the world, while its broader development has been community driven (Powers et al., 2017). The community effort has empowered WRF users towards more advanced research and operational applications. In the WRF community, tools and packages have been developed with community contributions (e.g. as cited in Meyer and Riechert, 2019; Powers et al., 2017, and references therein), while the supporting tools for the PALM community are still limited. As discussed in Maronga et al. (2020), more development of PALM is still needed to broaden its applicability and accessibility, and to strengthen its position within the boundary layer and urban climate scientific community. Through the implementation of a sustainable community model governance structure, PALM has the opportunity to become as popular and widely-used as WRF.

Geoscientific Model Development Discussions

We have developed a widely applicable geo-information toolkit which contains a set of routines written in the Python programming language designed to process geospatial data for PALM simulations. This tool is hereafter referred to as GEO4PALM. GEO4PALM can interface with several free online application programming interfaces (APIs) providing users with the ability to obtain domain-specific information from globally available databases. For users who have obtained their own geospatial data, GEO4PALM can process any geospatial data in geotiff format regardless of the data source. With GEO4PALM, users can include land surface characteristics such as topography, land use, and building and plant canopy information in their PALM simulations. Here we describe and document and GEO4PALM toolkit. The input standard of the PALM static driver is described in Section 2. Along with the framework and workflow of GEO4PALM, Section 3 presents detailed descriptions and requirements of the input data and the online data interfaces for GEO4PALM. Two application examples of GEO4PALM are given in Section 4. Conclusions and an outlook of GEO4PALM are presented in Section 5.

2 PALM static input

The hierarchy and data format of the variables in the static driver of PALM are described in the PIDS (https://palm.muk.uni-hannover.de/trac/wiki/doc/app/iofiles/pids, last access: 20 June 2023) and by Heldens et al. (2020). Depending on the application, PALM simulations can include surface features such as buildings, pavements, and plants. In GEO4PALM, two settings are available for users to choose. One is the default or the minimum setting, and the other allows users to add additional features for urban surface and plant canopy. For the minimum setting, the static driver is incorporated with all the required variables to conduct a PALM simulation, while all the additional features are optional. Although GEO4PALM does not cover all the available features presented in PIDS, it includes most of the available features including basic urban features such as buildings, pavements, and streets, which are sufficient to represent most urban and plant canopies. The variables covered by GEO4PALM are presented in Table 1.

3 GEO4PALM framework

An early version of GEO4PALM was applied in simulations presented by Lin et al. (2021) and Lin et al. (2022). However, this version (GEOPALM v1.0) was only applicable for geospatial data for Christchurch, New Zealand. In this paper, several new features have been added to GEO4PALM v1.1. The new design of GEO4PALM v1.1 aims to: 1) allow users to create static drivers for PALM simulations regardless of the geospatial data sources and 2) simplify the workflow of generating the static driver. The file steering structure of GEO4PALM is shown in Figure 1. The main source code of GEO4PALM is stored in the `util` folder with the main executable Python script (`run_config_static.py`) located in the main directory. The `JOBS` folder allows users to create static driver files for multiple different jobs (jobs x, y, z, etc.). In each job directory, users must have an `INPUT` folder, which includes a namelist file and all input geospatial data files for the static driver. The `TMP` folder stores all temporary files, and all static driver files are created and stored in the `OUTPUT` folder.

https://doi.org/10.5194/gmd-2023-150
Preprint. Discussion started: 22 August 2023

Table 1. List of variables that GEO4PALM can include in the PALM static driver. More detailed descriptions of the variables refer to Heldens et al. (2020) and the PIDS (https://palm.muk.uni-hannover.de/trac/wiki/doc/app/iofiles/pids, last access: 20 June 2023).

Variable name	**Feature**	**Description**
`zt`	Required	Terrain height above sea level in meters
`vegetation_type`	Required	Classification of vegetation types at land surface
`pavement_type`	Required	Classification of pavement types at land surface
`water_type`	Required	Classification of water bodies
`soil_type`	Required	Classification of soil types, usually specified for corresponding vegetation types
`surface_fraction`	Required	Relative fraction of the respective surface type given - depending on `vegetation_type`, `pavement_type`, and `water_type`
`albedo_type`	Optional	Optional classification of albedo for land surface
`water_pars`	Optional	Optional parameters for water bodies, including water temperature, roughness length for momentum, emissivity, etc.
`lad`	Optional	Three-dimensional leaf area density in $\mathrm{m^2 m^{-3}}$. Required for plant canopy model to resolve vegetation canopy
`street_type`	Optional	Optional classification of street type. Required for application of the parameterized traffic emissions and for the multi-agent system
`building_type`	Optional	Classification of building types. Required for buildings in urban surface model
`buildings_2d`	Optional	Heights of buildings relative to the underlying terrain. Required for buildings in urban surface model
`building_id`	Optional	Building ID to identify individual building envelopes. Required for buildings in urban surface model
`buildings_3d`	Optional	Three-dimensional building topology relative to the underlying terrain. Required for buildings in urban surface model

https://doi.org/10.5194/gmd-2023-150
Preprint. Discussion started: 22 August 2023

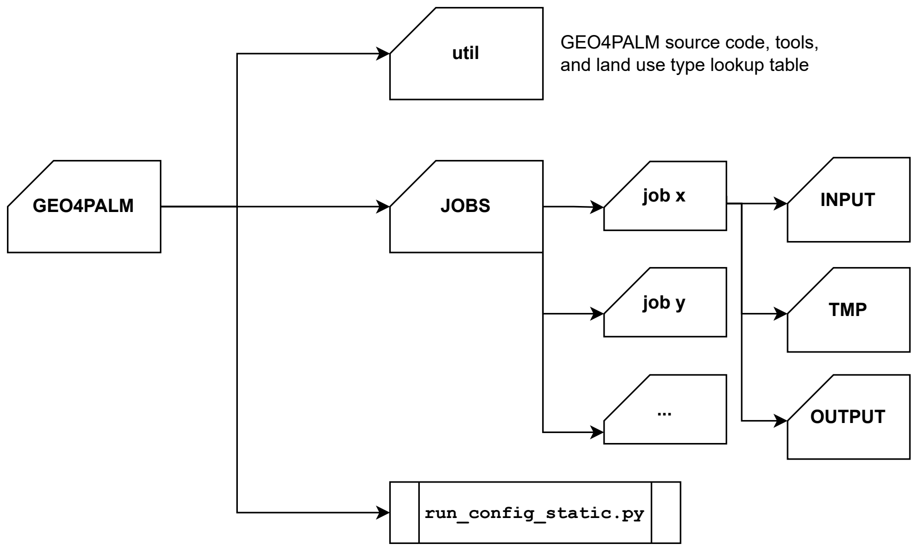

Figure 1. GEO4PALM file steering outline. The main excutable script is `run_config_static.py`. The `util` folder contains all GEO4PALM utility including source code, tools and land use classification lookup table. Within the `JOBS` folder, each user-specified job is stored separately with input files in `INPUT`, temporary files in `TMP`, and output files in `OUTPUT`.

The framework of GEO4PALM is shown in Figure 2. Users must at least provide a namelist file, which contains PALM domain configuration details and the data sources of static driver input. GEO4PALM uses Python packages including xarray, rasterio, rioxarray, geopandas, and geocube (Hoyer and Hamman, 2017; Gillies et al., 2019; Jordahl et al., 2020; Snow et al., 2022a, b) to process the geo-information data. When converting geospatial data into static driver, GEO4PALM requires input files in geotiff format, while no requirement is set regarding the spatial resolution and projection of the geotiff files. The geospatial input data are interpolated or extrapolated automatically based on the grid spacing values specified in the namelist. Users have the freedom to choose the resampling method for interpolation and/or extrapolation depending on their simulation needs. For details of the available options, users are referred to the rasterio documentation (https://rasterio.readthedocs.io/en/stable/api/rasterio.enums.html#rasterio.enums.Resampling; last access: 23 June 2023). In addition to the geotiff format, the shapefile format is one of the most common file types for geospatial data. Therefore, a script `shp2tif.py` is provided for users to convert shapefiles to geotiff format of the desired resolution (finest in the input namelist by default). Table 2 explains all variables contained in the namelist. A step-by-step guide for GEO4PALM and a sample namelist are provided in Appendix A.

In the namelist, users need to specify the desired geographic projection, domain configuration, and geospatial input data source for PALM simulations. If the desired projection of PALM simulations is not specified, GEO4PALM would use the nearest Universal Transverse Mercator (UTM) projection based on latitude and longitude of the PALM domain centre given in the namelist. To visualise PALM domain locations and to help users build the domain configuration easily, GEO4PALM is incorporated with a web-based interactive graphical user interface (GUI; Figure 3). This GUI is generated by the `palm_domain_utility.py`

https://doi.org/10.5194/gmd-2023-150
Preprint. Discussion started: 22 August 2023

Geoscientific Model Development
Discussions

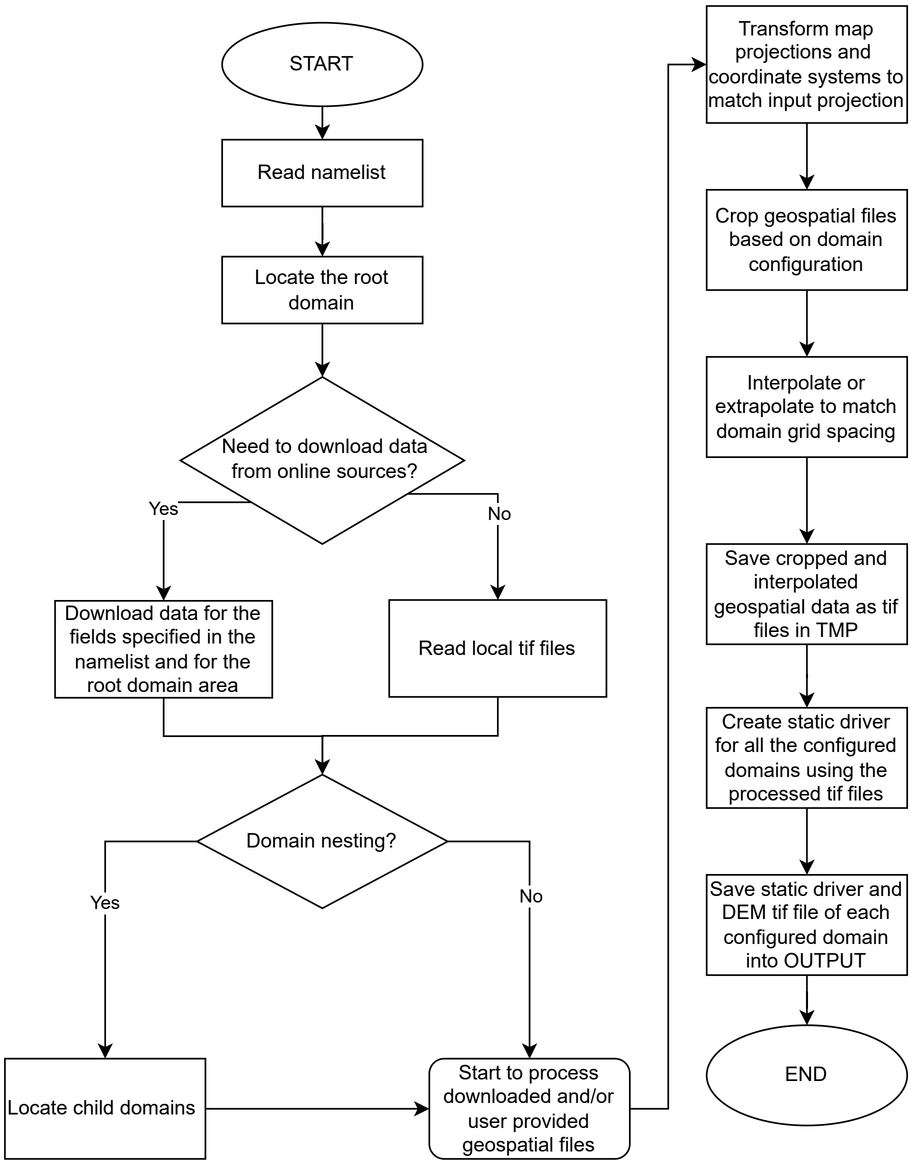

Figure 2. Flowchart showing the code structure of GEO4PALM.

https://doi.org/10.5194/gmd-2023-150
Preprint. Discussion started: 22 August 2023

Geoscientific Model Development
Discussions

script. This PALM domain utility allows users to render grid boxes of PALM domains over satellite and aerial imagery obtained from Environmental Systems Research Institute (ESRI) World Imagery (https://geoviews.org/gallery/bokeh/tile_sources.html; last access: 19 June 2023) through open-source Python libraries including Panel and Geoviews (https://panel.holoviz.org/ and https://geoviews.org/; last access: 19 June 2023). The utility automatically checks and adjusts the domain configuration to avoid violations of rules for PALM domain nesting such as overlapping of domain boundaries and mismatching of grid between parent and child domains. More details for using the PALM domain GUI are described in the GEO4PALM user manual (https://github.com/dongqi-DQ/GEO4PALM; last access: 21 June 2023).

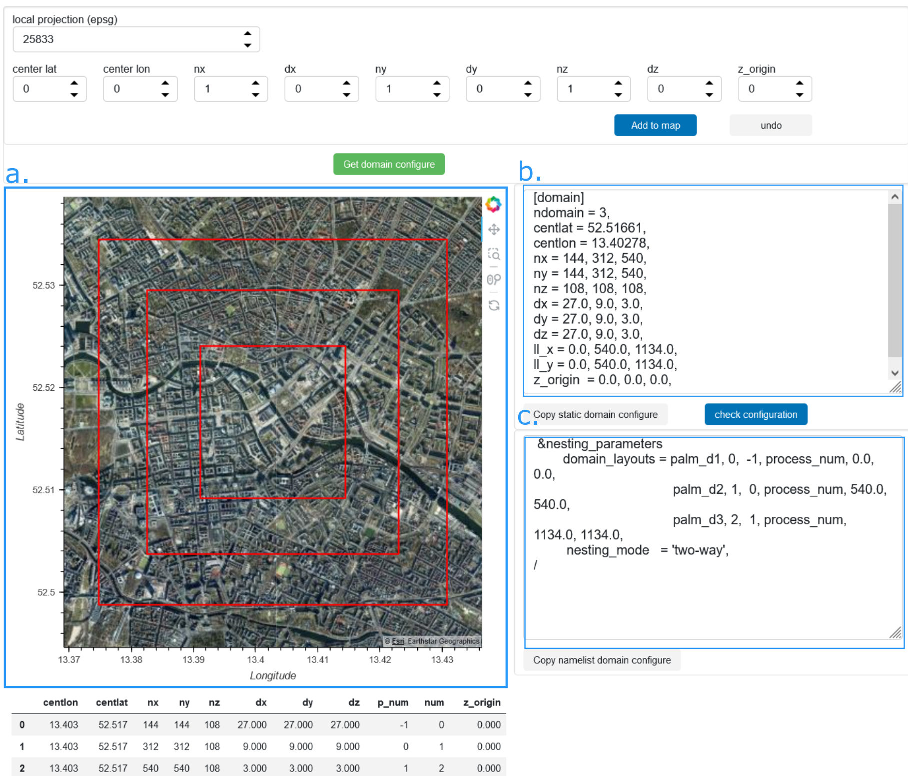

Figure 3. Screenshot of the web-based PALM domain utility GUI showing domain configuration for the Berlin case described in Section 4.1. PALM domains are drawn over the satellite imagery in the interactive subtab (a) on the left. The GUI generates namelists for both GEO4PALM and PALM domain configuration in the subtabs (b) and (c) on the right, respectively.

Users are not required to have a full set of input data. For each geospatial input field, users can either provide their own data in geotiff or shapefile format or choose to download data from online sources. For geospatial files provided by users, one needs to specify the file name in the namelist. If one desires to use online data, the data source should be specified. For the time being, we provide several interfaces to download globally available data sets. Sea surface temperature (SST), digital elevation model (DEM) and land use are the three mandatory elements to create a static driver. Water temperature is an important factor which can have impact on boundary layer structure (e.g. Mahrt and Hristov, 2017), while by default PALM prescribes a water

https://doi.org/10.5194/gmd-2023-150
Preprint. Discussion started: 22 August 2023

temperature of 283.0 K for all water bodies. Therefore, water temperature is required in GEO4PALM. As SST is widely available across various global data sets, we provide an interface to download SST data to represent water temperature for water bodies in the GEO4PALM static driver. The SST data (resolution of 0.01°) are obtained using the Earthdata Common Metadata Repository (CMR) API operated by the National Aeronautics and Space Administration (NASA) which is linked to the OPeNDAP interface (Open-source Project for a Network Data Access Protocol; https://lpdaac.usgs.gov/tools/opendap/; last access: 20 May 2023). The NASA data sets are available globally and users are required to register an account to use the data freely. By default, GEO4PALM downloads the version 4.1 Multiscale Ultrahigh Resolution (MUR) of a Group for High Resolution Sea Surface Temperature (GHRSST) Level 4 analysis provided by NASA Jet Propulsion Laboratory (NASA/JPL, 2015). To download and use this data set, users must specify "`online`" in the namelist for the variable "`sst`". If users have spatial water temperature data available for water bodies in geotiff format, they can specify the data file name in the namelist for the variable "`sst`".

The DEM (spatial resolution of 30 m) and land use classification are available to download from the Application for Extracting and Exploring Analysis Ready Samples (A$\rho\rho$EEARS, https://appeears.earthdatacloud.nasa.gov/; last access: 21 June 2023) operated by NASA. The DEM is the product of NASA Shuttle Radar Topography Mission 1 arc second NetCDF V003 (SRTMGL1_NC.003) acquired by space borne radar (Rabus et al., 2003). For the NASA data, users may provide the start and end date for the data acquisition such that the land use is more representative for the simulation period. GEO4PALM source code provides a lookup table for the MODIS Land Cover Type Product (A$\rho\rho$EEARS product code: LC_type1), which converts the land use classification to PALM recognisable values. The MODIS Land Cover Type Product supplies global land cover maps at 500 m spatial resolution dated from 2001 (Sulla-Menashe and Friedl, 2018). More options for land use classification data source provided by NASA refer to A$\rho\rho$EEARS online documentation (https://appeears.earthdatacloud.nasa.gov/; last access: 21 June 2023). If users desire to download and use A$\rho\rho$EEARS data sets, they must specify "`nasa`" in the namelist for the variable "`dem`" and/or "`lu`".

In addition to the A$\rho\rho$EEARS interface, GEO4PALM has incorporated with the application programming interface that connects to the worldwide land cover mapping (WorldCover; https://esa-worldcover.org/en; last access 21 June 2023) operated by European Space Agency (ESA). The ESA WorldCover data have a spatial resolution of 10 m (Zanaga et al., 2021, 2022). Users must register a free account to obtain data from ESA. A lookup table for PALM readable conversion is provided in GEO4PALM source code. For the usage of ESA data, users are required to specify "`esa`" in the namelist for the variable "`lu`". In addition to the lookup tables for NASA and ESA data, GEO4PALM source code provides a lookup table for New Zealand land cover database (LCDB) V5.0 (Landcare Research, 2020). All the lookup tables are presented in Appendix B.

All urban and plant fields can be left blank (""), in case users do not require such features in the static driver. If users desire to have other land surface features, GEO4PALM can process data for urban surface model and plant canopy model. In PALM, urban surfaces include pavements, buildings, and streets. Although according to PIDS v1.12, the typology of streets is represented by the variable `pavement_type` and the variable `street_type` is only used for the chemistry model in PALM, GEO4PALM still includes the variable `street_type` such that the static driver can be used for simulations that require chemistry model and/or the multi-agent system. Like the DEM and land use data, users can either provide their

https://doi.org/10.5194/gmd-2023-150
Preprint. Discussion started: 22 August 2023

Geoscientific Model Development
Discussions
Open Access
EGU

own geotiff data, or choose to download online. If users wish to provide their own data, they must provide geotiff files with building height at the building location, building ID for each building, PALM-recognisable pavement types for each pavement, and/or PALM-recognisable street types for each street, separately. Otherwise, users can specify "`osm`" for the urban variables in the namelist to download data from OpenStreetMap (OSM; https://www.openstreetmap.org/; last access: 21 June 2023). GEO4PALM uses the OSMnx package (Boeing, 2017) to obtain data from OSM. Downloaded OSM data sets are converted to PALM recognisable data by GEO4PALM using the conversion described by Heldens et al. (2020).

For plant canopy, GEO4PALM currently only supports leaf area density (LAD; Lalic and Mihailovic, 2004) calculation based on vegetation height provided in the geotiff files. To avoid noise from other surface geometry, GEO4PALM applies an automatic process in which surface objects with height less than 1.5 m are removed such that objects like cars or fences are not included as vegetation. This height limit can be adjusted in the source code depending on the quality of the geospatial data input. With high quality data, this noise filter can be removed such that low objects like grass, long grass, and bushes, can be included and represented in PALM simulations. The LAD calculation is adopted from PALM CSD (Heldens et al., 2020) and is based on equation proposed by Lalic and Mihailovic (2004) as follows,

$$LAD(z) = LAD_m \frac{h - z_m}{h - z}^n e^{(1 - \frac{h - z_m}{h - z})n} \qquad [\mathrm{m}^2\ \mathrm{m}^{-3}] \tag{1}$$

where L_m is the maximum LAD and the z_m is the height where the leaf area index (LAI) reaches L_m, $n = 6$ when $z < z_m$ and $n = 0.5$ when $z > z_m$. According to Lalic and Mihailovic (2004), the normalised value of z_m range from 0.2 to 0.4 depending on the tree type. L_m is derived based on LAI and z_m. Currently, GEO4PALM only allows users to provide fixed LAI (`tree_lai_max` in Table 2) and z_m (`lad_max_height` in Table 2) values as input. GEO4PALM automatically scales the LAI based on the height of the vegetation canopy at individual grid points. This approach does not take account of spatial variation in LAD resulted from different tree species, while it is still useful in cases where no LAD data are available. For cases which LAD or even LAI information is available, users are advised to adjust the code to directly read the spatial LAD/LAI information. However, to our best knowledge, no globally available vegetation height along with plant canopy data are currently free to obtain from any online sources. Therefore, GEO4PALM does not provide any online interface for this purpose. One possible solution to obtain vegetation height is to calculate surface objects height using digital surface model (DSM) and DEM. In addition to the information of ground surface altitude contained in DEM, DSM supplies the heights of all surface objects such as buildings and trees.

Once users have provided all required information in the namelist file along with their own geospatial data where applicable, GEO4PALM downloads and/or processes the input data to create the static driver. GEO4PALM allows users to configure nested domains. To reduce the learning curve, the domain nesting configuration is similar to PALM's nesting module, in which the nested domain location is determined by the distance of the lower left corners between the root domain and child domains. PALM's own input and output files do not contain any geospatial coordinate reference, while geospatial coordinates could be important in real-world applications. To overcome this potential issue, along with each static driver, a geotiff file with coordinate information is created by GEO4PALM.

https://doi.org/10.5194/gmd-2023-150
Preprint. Discussion started: 22 August 2023

Table 2. Variables descriptions of GEO4PALM input namelist.

Variables	Descriptions	Comment
case		
case_name	Name of the case, identical to the job folder name.	
origin_time	Date and time to start the PALM simulation in "YYYY-MM-DD HH:mm:ss +HH" format. For example, 1200 UTC on 21st June 2019 is "2019-06-21 12:00:00 +00".	Refer to origin_date_time in the PALM input namelist for more details.
default_proj	The default projection (EPSG:4326) for GEO4PALM to use latitudes and longitudes to locate domains.	For most users, no changes are needed.
config_proj	The desired projection of the static driver.	Local projection with units in metres is recommended, for example, EPSG:2193 for New Zealand.
domain		
ndomain	Maximum number of domains.	If ndomain $\geq$ 2, domain nesting is enabled.
centlat	Centre latitude of the root domain.	Not required for nested child domains.
centlon	Centre longitude of the root domain.	Not required for nested child domains.
nx	Number of grid points along x-axis.	The actual number of grid points along x-axis (nx). Note that this is different from the PALM namelist (nx-1).
ny	Number of grid points along y-axis.	The actual number of grid points along y-axis (ny). Note that this is different from the PALM namelist (ny-1).
nz	Number of grid points along z-axis.	
dx	Grid spacing in metres along x-axis.	
dy	Grid spacing in metres along y-axis.	
dz	Grid spacing in metres along z-axis.	
z_origin	Mean elevated terrain grid position in metres.	Default is 0.0 m.

https://doi.org/10.5194/gmd-2023-150
Preprint. Discussion started: 22 August 2023

Geoscientific Model Development
Open Access
Discussions
EGU

Table 2. Continued.

Variables	Descriptions	Comment
ll_x	Lower left corner distance to the first domain in metres along x-axis; only use when nesting is required.	
ll_y	Lower left corner distance to the first domain in metres along y-axis; only use when nesting is required.	
geotiff		
sst	Input data source for water temperature. Users need to specify the input file name or data can be downloaded from OPeNDAP with the option "online".	
dem	Input data source for topographical height. Users need to specify the file name or data can be downloaded from NASA A$\rho\rho$RAS with the option "nasa".	
lu	Input data source for land use classification. Users need to specify the file name or data can be downloaded from NASA A$\rho\rho$RAS with the option option "nasa" and/or from ESA WorldCover with the option "esa".	A lookup table to convert land use typologies to PALM recognisable values is required.
resample_method	Method to resample geotiff files when interpolating/extrapolating to desired grid spacing.	Default is "nearest".
If A$\rho\rho$RAS interface is used		
dem_start_date	DEM data start date in YYYY-MM-DD format.	Default is 2000-02-12 and no need to change if SRTMGL1_NC.003 data set is used.
dem_end_date	DEM data end date in YYYY-MM-DD format.	Default is 2000-02-20, and no need to change if SRTMGL1_NC.003 data set is used.

https://doi.org/10.5194/gmd-2023-150
Preprint. Discussion started: 22 August 2023

Table 2. Continued.

Variables	**Descriptions**	**Comment**
`lu_start_date`	Land use data start date in YYYY-MM-DD format.	Default is `2020-10-01` for product LC_Type01. Should be changed upon users' needs.
`lu_end_date`	Land use data end date in YYYY-MM-DD format.	Default is `2020-10-30` for product LC_Type01 Should be changed upon users' needs.
urban		
`bldh`	Input data source for building height. Users need to specify the file name or data can be downloaded from OSM with the option "`online`".	
`bldid`	Input data source for building ID. Users need to specify the file name or data can be downloaded from OSM with the option "`online`".	
`pavement`	Input data source for pavement types. Users need to specify the file name or data can be downloaded from OSM with the option "`online`".	
`street`	Input data source for street types. Users need to specify the file name or data can be downloaded from OSM with the option "`online`".	
plant		
`tree_lai_max`	Input value for maximum leaf area index.	
`lad_max_height`	Input value for z_m (range between 0.2 and 0.4) as described in Equation 1.	
`sfch`	Input data source for plant height above the surface. Users need to specify the file name.	Currently no online data source/interface available.

https://doi.org/10.5194/gmd-2023-150
Preprint. Discussion started: 22 August 2023

Geoscientific Model Development Discussions

4 Examples for real-world applications

As mentioned in Section 3, the aim of GEO4PALM is to allow users to seamlessly use geospatial data for their PALM simulations. Users may have data ready locally in geotiff or shapefile format, or they can download all geospatial data freely for any location in the world. The input data described in this article and the earlier versions of GEO4PALM have been used in several other studies. For example, the case studies carried out by Lin et al. (2021) used static input generated by GEO4PALM for Christchurch International Airport, which demonstrated applications of the WRF4PALM offline nesting tool. Lin et al. (2022) have used GEO4PALM generated static input for fog research over Christchurch city. In this section, we present two examples to demonstrate the performance and compatibility of GEO4PALM. One example is Berlin, Germany (52.516615°N, 13.402782°E), and the other is Christchurch, New Zealand (43.529599°S, 172.596928°E). For both cases, we have used the online API to obtain geospatial input. For Berlin, we have geospatial data sets prepared and stored locally containing topography, streets, buildings, water bodies, vegetation, etc., at a 2 m resolution. These data have been processed by the German Space Agency (DLR), similar to those described by Khan et al. (2021) and Heldens et al. (2020). Hereafter, we refer to the local data sets for Berlin as the DLR data sets. We have used the same DLR data sets to run both PALM CSD described by Heldens et al. (2020) and GEO4PALM with map projection of EPSG:25833 (ETRS89 / UTM zone 33N) to match the projection coded in PALM CSD. At present, PALM CSD only supports ten UTM projections between zone 28N and zone 37N. Therefore, for Christchurch, although local data sets are available, only GEO4PALM was used with the map projection of EPSG:2193 (New Zealand Transvers Mercator). We do not have data for Christchurch that are compatible with PALM CSD.

Table 3 gives an overview of the simulations conducted for Berlin and Christchurch. Overall, three simulations were conducted for Berlin, and two simulations were carried out for Christchurch. The three simulations for Berlin are denoted as Berlin_CSD, Berlin_GEO, and Berlin_OL, which used static drivers generated by PALM CSD, GEO4PALM with the DLR data sets, and GEO4PALM with online data sets, respectively. The two simulations for Christchurch are denoted as Chch_GEO and Chch_OL, which used static drivers generated by GEO4PALM with local data sets, and GEO4PALM with online data sets, respectively. For demonstration purposes, both examples have identical domain dimensions and grid spacing (shown in Table 4). All geospatial input data were interpolated or extrapolated to match the grid spacing of the simulations. To demonstrate the LAD calculation in GEO4PALM, we used `tree_lai_max`$=5.0$ and `lad_max_height`$=0.4$, corresponding to pine trees, for both Berlin_GEO and Chch_GEO. The initialisation time was set to 0000 UTC 1 January 2021 for Christchurch and 1200 UTC 1 January for Berlin, which is midday summer time for Christchurch and around midday winter time for Berlin. Simulation time is 6 hours for both simulations. Due to the high computation cost with fine grid spacing, here we only performed simulations using domain nesting with finest grid spacing at 3 m, while GEO4PALM can generate static driver with grid spacing finer than 1 m, depending on the data source.

4.1 Case study for Berlin

This section demonstrates a case to compare the static drivers created by PALM CSD, GEO4PALM with local data, and GEO4PALM with online data. The input data sources used in GEO4PALM are listed in Table 5. For Berlin_CSD and Berlin_GEO,

Table 3. Overview of simulations for Berlin and Christchurch.

Simulation case name	Data source	Static driver generation tools
Berlin		
Berlin_CSD	DLR data set (resolution of 2 m)	PALM CSD
Berlin_GEO	DLR data set (resolution of 2 m)	GEO4PALM
Berlin_OL	Online data sets included in GEO4PALM (refer to Table 5 for details)	GEO4PALM
Christchurch		
Chch_GEO	Christchurch local data sets (refer to Table 6 for details)	GEO4PALM
Chch_OL	Online data sets included in GEO4PALM (refer to Table 6 for details)	GEO4PALM

Table 4. Nested domain dimension summary. Here x refers to the west-east coordinate, y refers to the south-north coordinate, and z refers to the vertical coordinate.

Domain	Number of grid points (x, y, z)	Domain size (x, y, z)	Horizontal grid spacing (dx, dy)	Vertical grid spacing (dz)
N01	144*144*108	3888*3888*2916	27 m	27 m
N02	312*312*108	2808*2808*972	9 m	9 m
N03	540*540*108	1620*1620*324	3 m	3 m

the data were processed by DLR. For details of the DLR data sets and PALM CSD, refer to Khan et al. (2021) and Heldens et al. (2020). As water temperature was not included in the DLR data set, we used GEO4PALM to obtain water temperature via the Earthdata API for Berlin_GEO. Regarding Berlin_OL, the only user input is the namelist file. GEO4PALM handled all data downloading from online sources including the global GHRSST data, NASA 30 m DEM data (SRTMGL1_NC.003), ESA WorldCover land use classification data, and OSM urban data sets.

The domain location and static driver data for Berlin_CSD, Berlin_GEO, and Berlin_OL are shown in Figures 4 and 5. In Berlin_OL, several buildings are included in the simulation domain, however OSM data (used in the Berlin_OL simulation) do not include building height for most of the buildings. Therefore, buildings with no height data available were assigned with a height of 3 m for demonstration purposes. Leaf area index, the vertical integrated LAD, shows areas with plant canopy in Figures 4b and 4c. For the Berlin_OL case, LAD was not included in the static driver input, since we do not have any DEM or DSM with spatial resolution finer than 30 m. The LAI calculated by GEO4PALM (Figure 4b) is higher than the one calculated by PALM CSD (Figure 4c). In Berlin_GEO, the vegetation patch height data were directly used by GEO4PALM for the estimation of LAD, without taking vegetation type into account. In Berlin_CSD, however, vegetation type and the simulation season were considered. A great amount of data processing is required to have vegetation data pre-processed for PALM CSD. Considering the inconsistency in data sources and quality worldwide, GEO4PALM adopts the simplified method to calculate LAD, while it can be modified further by users depending on their modelling needs.

https://doi.org/10.5194/gmd-2023-150
Preprint. Discussion started: 22 August 2023

Geoscientific Model Development Discussions
Open Access EGU

Table 5. Geospatial input data for GEO4PALM used in the Berlin case study. Refer to Khan et al. (2021) and Heldens et al. (2020) for details of data sources for Berlin_CSD.

GEO4PALM variables	Data set	Source
Berlin_GEO		
`sst`	GHRSST Level 4 MUR product	OPeNDAP via Earthdata (https://lpdaac.usgs.gov/tools/opendap/; last access: 21 June 2023)
`dem`	DLR data sets	Refer to Khan et al. (2021) and Heldens et al. (2020).
`lu`	DLR data sets	Refer to Khan et al. (2021) and Heldens et al. (2020).
`bldh`	DLR data sets	Refer to Khan et al. (2021) and Heldens et al. (2020).
`bldid`	DLR data sets	Refer to Khan et al. (2021) and Heldens et al. (2020).
`pavement`	DLR data sets	Refer to Khan et al. (2021) and Heldens et al. (2020).
`street`	DLR data sets	Refer to Khan et al. (2021) and Heldens et al. (2020).
`sfch`	DLR data sets	Refer to Khan et al. (2021) and Heldens et al. (2020).
Berlin_OL		
`sst`	GHRSST Level 4 MUR product	OPeNDAP via Earthdata (https://lpdaac.usgs.gov/tools/opendap/; last access: 21 June 2023)
`dem`	NASA Shuttle Radar Topography Mission 1 arc second NetCDF V003 (SRTMGL1_NC.003)	A$\rho\rho$EEARS (https://appeears.earthdatacloud.nasa.gov/; last access: 21 June 2023)
`lu`	ESA WorldCover 10m (2020 V1)	TerraCatalogueClient (https://vitobelgium.github.io/terracatalogueclient; last access: 21 June 2023)
`bldh`	OSM	OpenStreetMap via osmnx (Boeing, 2017)
`bldid`	OSM	OpenStreetMap via osmnx (Boeing, 2017)
`pavement`	OSM	OpenStreetMap via osmnx (Boeing, 2017)
`street`	OSM	OpenStreetMap via osmnx (Boeing, 2017)
`sfch`	Not available	Not applicable

Berlin_GEO and Berlin_CSD present similar features in terrain heights (Figures 4e and 4f) in which the shapes of rivers are distinguishable. However, the terrain heights in Berlin_OL (Figure 4d) do not contain many details due to the coarse resolution of its data source. The online DEM data only have a spatial resolution of 30 m. Although Berlin_GEO and Berlin_CSD used the same geospatial input, the detailed topography is different. This is because PALM CSD asks users to input the vertical

https://doi.org/10.5194/gmd-2023-150
Preprint. Discussion started: 22 August 2023

Geoscientific Model Development
Discussions
Open Access
EGU

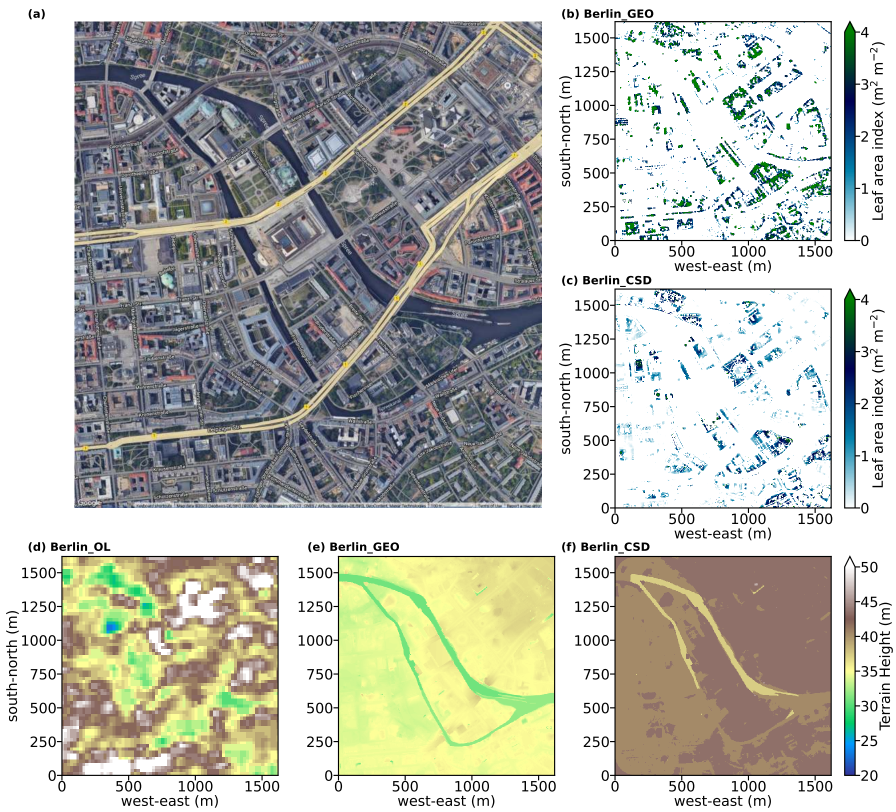

Figure 4. (a) Satellite images showing domain location of the nested domain N03 for the Berlin case (©Google Earth, last access: 19 May 2023). Domain centre is located near Humboldt Forum in central Berlin, Germany. The horizontal cross sections of static input data: (b-c) leaf area index (vertically integrated LAD), and (d-f) terrain height. Data sources refer to Table 4. No LAI data are displayed for Berlin_OL, as no online data source is available for the estimation of LAD.

grid spacing for static drivers and processes the topography onto vertical model levels before feeding data into PALM, while GEO4PALM adopts DEM directly and lets PALM itself to process and convert topography into simulation grid.

Vegetation type in Berlin_OL (Figure 5a) also has a coarser resolution compared to Berlin_GEO (Figure 5b) and Berlin_CSD (Figure 5c), because the spatial resolution for ESA land use data is 10 m. The difference in the vegetation type classification between Berlin_OL and the other two simulations could be due to the conversion between the ESA worldcover data set and PALM. Users are referred to Table B3 for the conversion and this can be edited depending on users' own needs. Comparing Berlin_GEO (Figure 5b) to Berlin_CSD (Figure 5c), one can notice that vegetation patches classified as type 7 (deciduous broadleaf trees; brown patches in Figure 5b) in Berlin_GEO are classified as type 3 (short grass; light purple in Figures 5c) in Berlin CSD. This is caused by the adjustment applied in PALM CSD. It corrects vegetation type when a vegetation height is available and is indicative of low-laying plant cover. This feature is currently not available in GEO4PALM due to lack of data.

https://doi.org/10.5194/gmd-2023-150
Preprint. Discussion started: 22 August 2023

Geoscientific Model Development
Open Access
Discussions

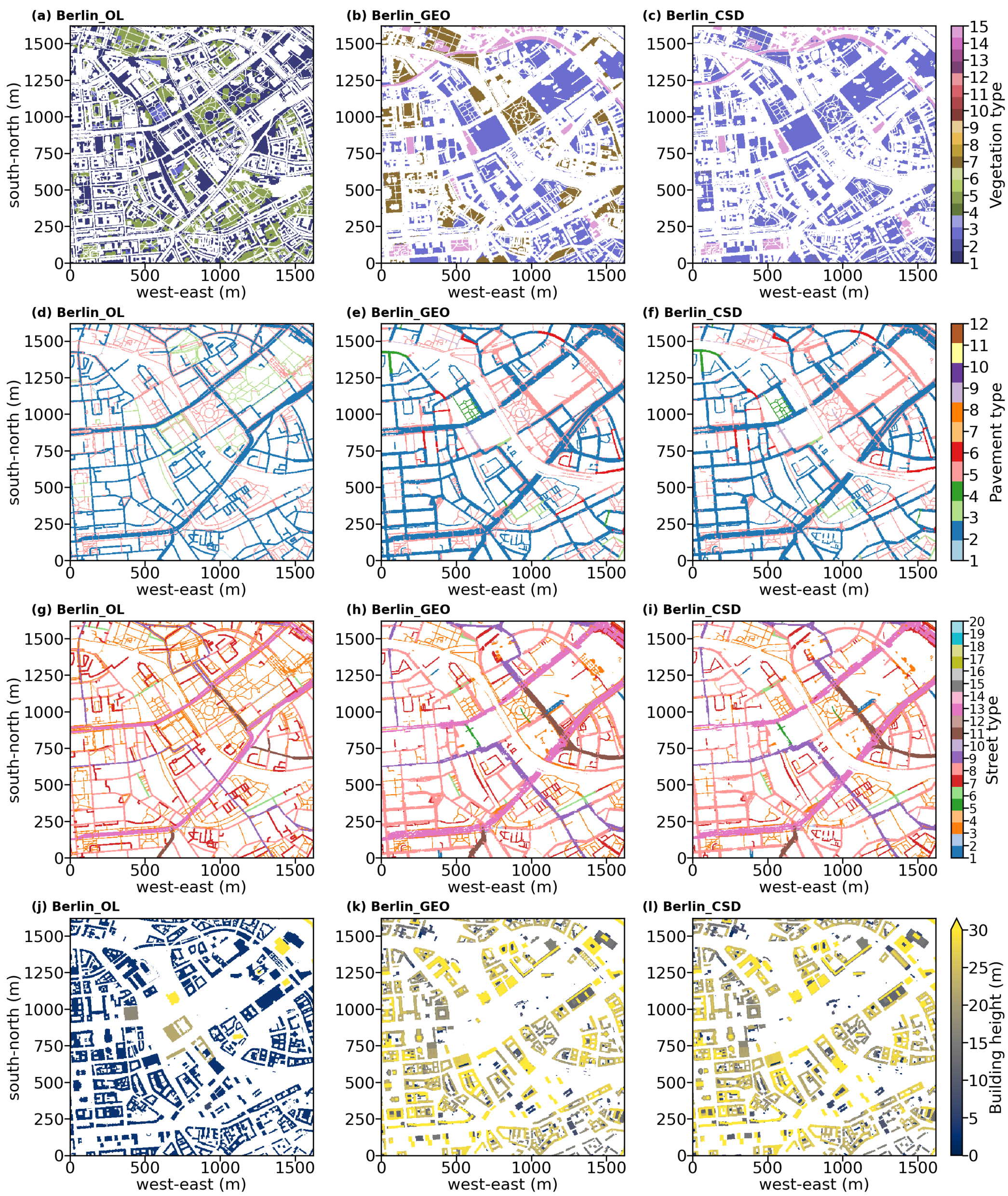

Figure 5. Static input data of the nested domain N03 for the Berlin case: (a-c) vegetation type, (d-f) pavement type, (g-i) street type, and (j-l) building height. Refer to panel label for the corresponding simulation. Data sources refer to Table 5.

https://doi.org/10.5194/gmd-2023-150
Preprint. Discussion started: 22 August 2023

In general, the pavement type and street type presented in Berlin_OL are similar to Berlin_GEO and Berlin_CSD. As the widths of pavements and streets were generated automatically in Berlin_OL by GEO4PALM, some of the details, such as width and type of each pavement and street, are of lower fidelity in Berlin_OL compared to those in Berlin_GEO and Berlin_CSD. This is similar regarding buildings. The structures and locations of buildings in Berlin_OL align with those in Berlin_GEO and Berlin_CSD. However, the DLR geospatial data do not include Humboldt Forum (as shown in the centre of Figures 4a, and 5j-5l), while this building is presented in Berlin_OL (Figure 5j). This missing building in Berlin_GEO and Berlin_CSD may be caused by errors in pre-processing of the original data.

The simulation results for the three cases are illustrated in Figure 6. The results include 2 m potential temperature (θ), surface temperature (T_{sfc}), surface net radiation (R_{net}), and 10 m wind speed (WS) at the last time step of the simulations. In all simulations, R_{net} and T_{sfc} are strongly dependent on the land surface type and surface canopy. Comparing Berlin_OL to Berlin_GEO, the differences show the impact of the geospatial data input. Due to lack of building height data, buildings are generally lower in Berlin_OL. Therefore, both θ at 2m and T_{sfc} over the built-up area are lower in Berlin_OL (Figures 6a and 6d) than those in Berlin_GEO (Figures 6b and 6e). The reduction of building heights and lack of LAD in Berlin_OL leads to a decrease in surface friction and hence the 10 m WS in Berlin_OL (Figure 6g) does not present as much wind speed variation caused by buildings as in Berlin_GEO (Figure 6h). In addition, as the land use input data have a grid spacing of 10 m, the water bodies were not presented with good fidelity in Berlin_OL, compared to Berlin_GEO. This is reflected in the simulated R_{net} in Figures 6j and 6k. In both Berlin_OL and Berlin_GEO, the water bodies coincide with strong positive R_{net}, while the outlines of the rivers are coarse in Berlin_OL.

The results of Berlin_GEO and Berlin_CSD present a comparison between GEO4PALM and PALM CSD. Compared to Berlin_CSD, Berlin_GEO presents a lower 2 m θ (Figures 6b and 6c), a higher T_{sfc} over buildings (Figures 6e and 6f) and stronger winds (Figures 6h and 6i). This is a result of different adoptions of topography data. Whether GEO4PALM or PALM CSD is more accurate regarding the topography data processing requires further investigation, which is out of scope of this paper. One drawback of PALM CSD is that it does not provide water feature rendering. As the simulations presented here are for winter, the water temperature obtained from the GHRSST data set is 275.0 K while Berlin_CSD has the default water temperature of 283.0 K. Such a difference in water temperature leads to significant contrasts in R_{net} between Berlin_GEO (strong positive; Figure 6k) and Berlin_CSD (weak negative; Figure 6l). Overall, the static driver generated by GEO4PALM and subsequently the simulation fidelity of PALM is highly dependent on the input geospatial data quality. With high quality data, GEO4PALM managed to create static drivers (Belin_GEO) that are comparable to PALM CSD (Berlin_CSD). Without any local data, GEO4PALM can still represent the simulated environment with reasonable details as presented in Berlin_OL.

4.2 Case study for Christchurch

In this section, we present a case study for Christchurch, New Zealand, to demonstrate the application of GEO4PALM when the simulation location changes. Local geospatial data sets are used to demonstrate the suitability and applicability of GEO4PALM for the case Chch_GEO. Similar to Berlin_OL, Chch_OL used all geospatial data downloaded by GEO4PALM rather than local data sets. PALM CSD was not used for Christchurch due to incompatibility of data and map projections. The input data for

https://doi.org/10.5194/gmd-2023-150
Preprint. Discussion started: 22 August 2023

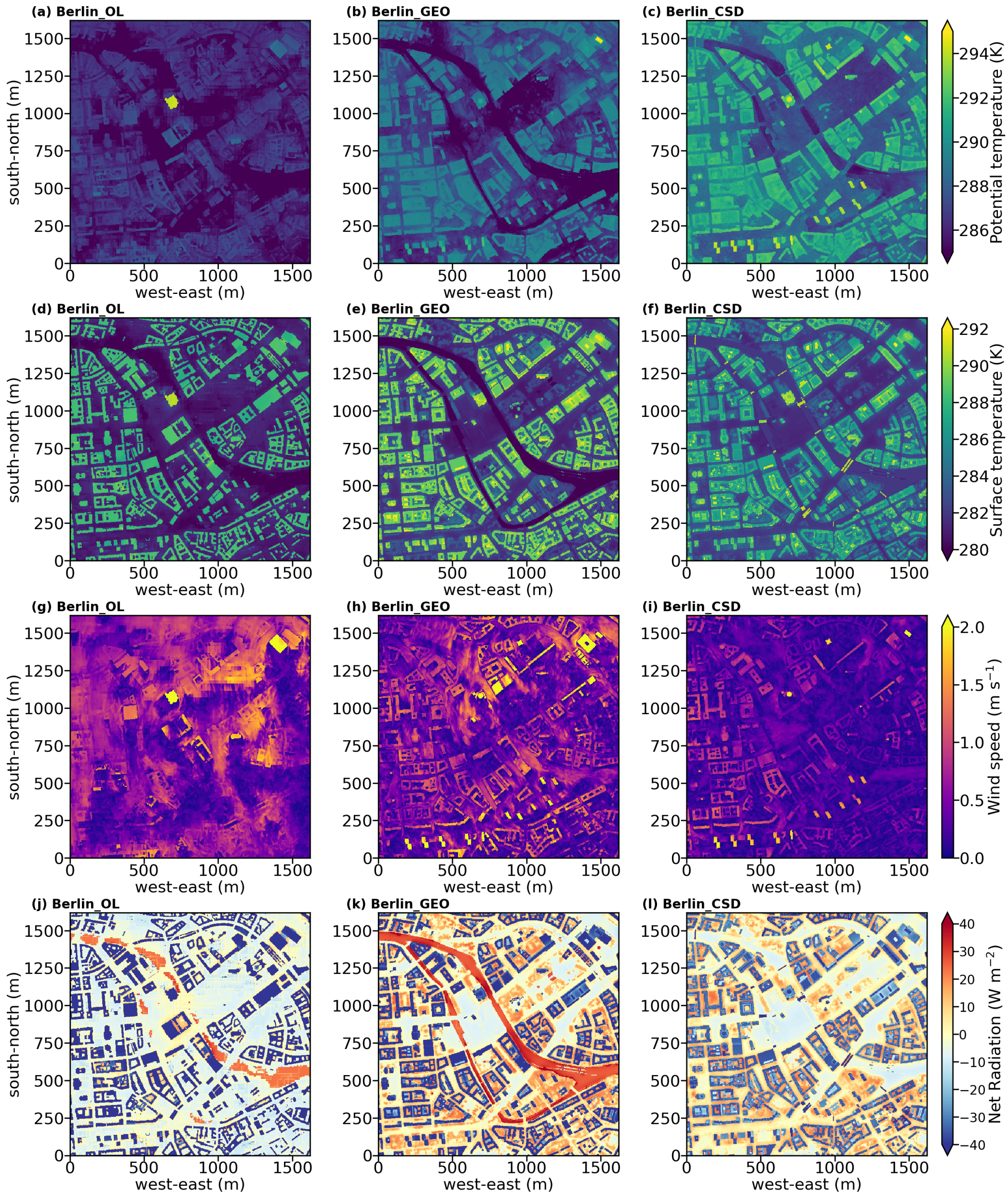

Figure 6. Horizontal cross-sections of simulation results at the last time step of the simulation for the Berlin case: (a-c) 2 m potential temperature, (d-f) surface temperature, (g-i) 10 m wind speed, and (j-l) surface net radiation. Refer to panel label for the corresponding simulation.

https://doi.org/10.5194/gmd-2023-150
Preprint. Discussion started: 22 August 2023

Geoscientific Model Development Discussions
Open Access
EGU

this case study are listed in Table 6. For both simulations, water temperature was obtained from the GHRSST data sets. For Chch_GEO, the DEM was obtained from Envirionment Canterbury Regional Council (2020), the land use classification was
285 obtained from New Zealand LCDB v5.0 (Landcare Research, 2020), the building footprint, location, and building ID were derived using OSM data and New Zealand building outlines data set (Land Information New Zealand, 2020), the building height was calculated using the difference between DSM and DEM at the building locations, the pavement and street data were obtained from OSM and converted to PALM recognisable data using the conversion provided by Heldens et al. (2020), and the tree height was derived using the difference between DSM and DEM with buildings excluded.

Table 6. Geospatial input data for GEO4PALM used in the Christchurch case study.

GEO4PALM variables	**Data set**	**Source**
Chch_GEO		
`sst`	GHRSST Level 4 MUR product	OPeNDAP via Earthdata (https://lpdaac.usgs.gov/tools/opendap/; last access: 21 June 2023)
`dem`	Christchurch DEM with spatial resolution of 1 m	Envirionment Canterbury Regional Council (2020)
`lu`	New Zealand LCDB V5.0	Refer to Landcare Research (2020)
`bldh`	– OSM – Christchurch DEM with spatial resolution of 1 m – Christchurch DSM with spatial resolution of 1 m – New Zealand building outlines data set	– OpenStreetMap – Envirionment Canterbury Regional Council (2020) – Land Information New Zealand (2020)
`bldid`	OSM	OpenStreetMap
`pavement`	OSM	OpenStreetMap
`street`	OSM	OpenStreetMap
`sfch`	– Christchurch DEM with spatial resolution of 1 m – Christchurch DSM with spatial resolution of 1 m	Envirionment Canterbury Regional Council (2020)

https://doi.org/10.5194/gmd-2023-150
Preprint. Discussion started: 22 August 2023

Table 6. Continued.

GEO4PALM variables	Data set	Source
Chch_OL		
`sst`	GHRSST Level 4 MUR product	OPeNDAP via Earthdata (https://lpdaac.usgs.gov/tools/opendap/; last access: 21 June 2023)
`dem`	NASA Shuttle Radar Topography Mission 1 arc second NetCDF V003 (SRTMGL1_NC.003)	A$\rho\rho$EEARS (https://appeears.earthdatacloud.nasa.gov/; last access: 21 June 2023)
`lu`	ESA WorldCover 10m (2020 V1)	TerraCatalogueClient (https://vitobelgium.github.io/terracatalogueclient; last access: 21 June 2023)
`bldh`	OSM	OpenStreetMap via osmnx (Boeing, 2017)
`bldid`	OSM	OpenStreetMap via osmnx (Boeing, 2017)
`pavement`	OSM	OpenStreetMap via osmnx (Boeing, 2017)
`street`	OSM	OpenStreetMap via osmnx (Boeing, 2017)
`sfch`	Not available	Not applicable

Figures 7 and 8 show the domain location and static input data for the nested domain of 3 m grid spacing (domain N03) for Chch_GEO and Chch_OL, respectively. Riccarton bush is located near the centre of domain N03 coinciding with high LAI over an area of approximately 98,000 m^2 as shown in Figure 7b. Again, due to coarse resolution of the online DEM data, Chch_OL does not present good details in topography height (Figure 7c) compared to Chch_GEO (Figure 7d). However, Chch_OL does capture the decrease in topography from west to east of the domain. As the New Zealand land use data set (New Zealand LCDB v5.0) classifies all urban area as only one type, most area within Christchurch city was identified as bare soil by GEO4PALM (Figure 8b). This is different in ESA Worldcover data set that Chch_OL shows more areas with vegetation type 5 (deciduous needleleaf trees; green in Figure 8a), which align with the satellite image shown in Figure 7a. Pavement type and street type are almost identical in Chch_OL (Figures 8c and 8e) and Chch_GEO (Figures 8d and 8f) as they both used OSM data, while Chch_OL used online data directly obtained by GEO4PALM and Chch_GEO used local data that were processed manually. The differences between Chch_OL and Chch_GEO could be caused by the different data acquisition date and different input data for buildings. Comparing Figure 8g to Figure 8h, Chch_OL has several buildings missing. Since Chch_GEO was sourced from both OSM and New Zealand building outlines data set (Land Information New Zealand, 2020), its building information is more comprehensive and accurate. In addition, similar to the Berlin case, OSM online data do not provide much building height information and hence most buildings in Chch_OL have a height of 3 m (Figure 8g).

https://doi.org/10.5194/gmd-2023-150
Preprint. Discussion started: 22 August 2023

Geoscientific Model Development
Discussions

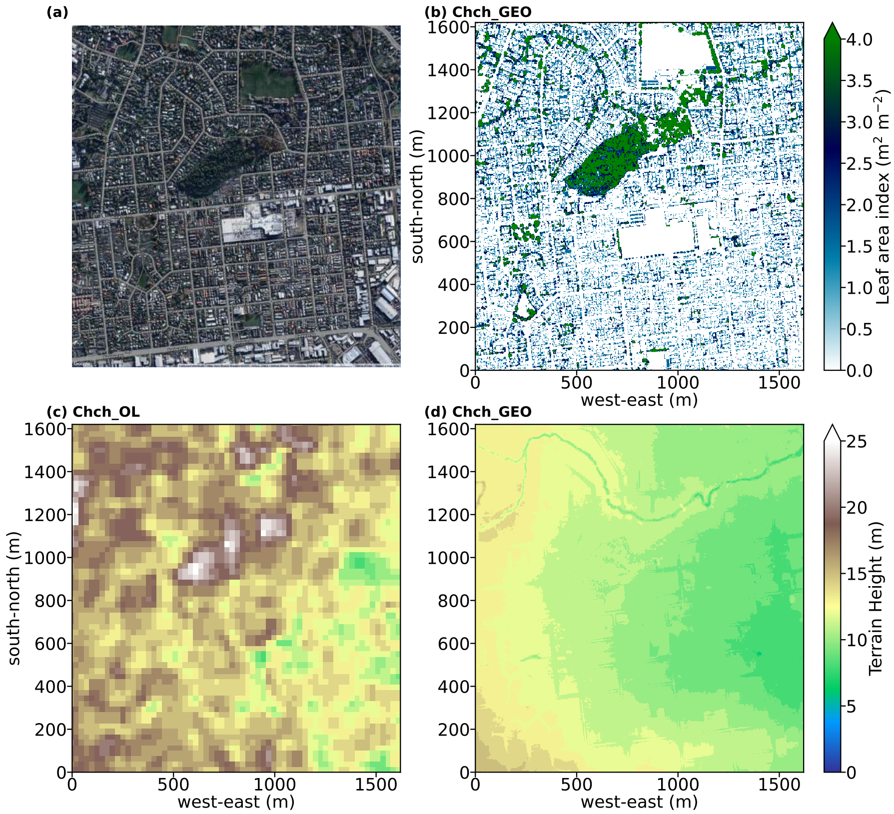

Figure 7. Similar to Figure 4, but for the Christchurch case. (a) Domain centre is located near Riccarton Bush in Christchurch, New Zealand (satellite image (©Google Earth, last access: 19 May 2023). LAI is only shown for Chch_GEO in panel (b). No LAI data are displayed for Chch_OL, as no online data source is available for the estimation of LAD. Panels (c-d) display terrain height. Refer to panel label for the corresponding simulation. Data sources refer to Table 6.

Simulation results for the Christchurch case are shown in Figure 9. The impact of plant canopy is noticeable in Chch_GEO. In contrast to Chch_OL, Chch_GEO presents lower θ at 2 m (Figures 9a and 9b) and T_{sfc} (Figures 9c and 9d) over the plant canopy. The presence of the dense plant canopy over Riccarton bush leads to considerable decrease in wind speed in Chch_GEO (Figures 9e and 9f) coinciding with a lower positive R_{net} at the surface. With a more comprehensive building data set, Chch_GEO illustrates more details in the simulated results in association with buildings. This is noticeable in the simulated R_{net} that shadows of buildings lead to a negative R_{net} (blue in Figure 9h). Due to coarse grid spacing of the online DEM data (30 m), the Chch_OL results coincide with patchy structures, especially in 2 m θ (Figure 9a), T_{sfc}(Figure 9b), and R_{net} (Figure 9g). Although all input data have been interpolated and processed by GEO4PALM, mismatch could still exist when resolution varies between the input data sets. Despite the quality of the geospatial input data, these results agree with those shown in Heldens et al. (2020).

https://doi.org/10.5194/gmd-2023-150
Preprint. Discussion started: 22 August 2023

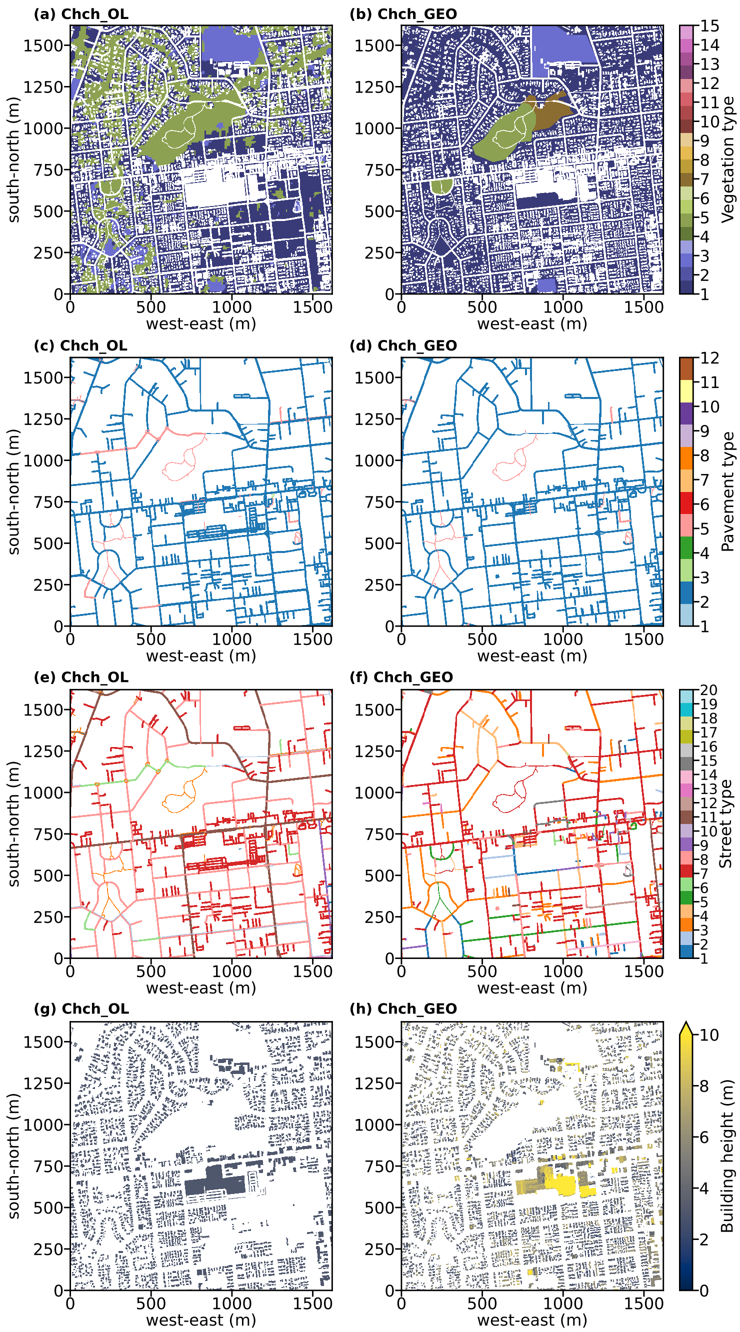

Figure 8. Similar to Figure 5, but for the Christchurch case: (a-b) vegetation type, (c-d) pavement type, (e-f) street type, and (f-h) building height. Refer to panel label for the corresponding simulation. Data sources refer to Table 6.

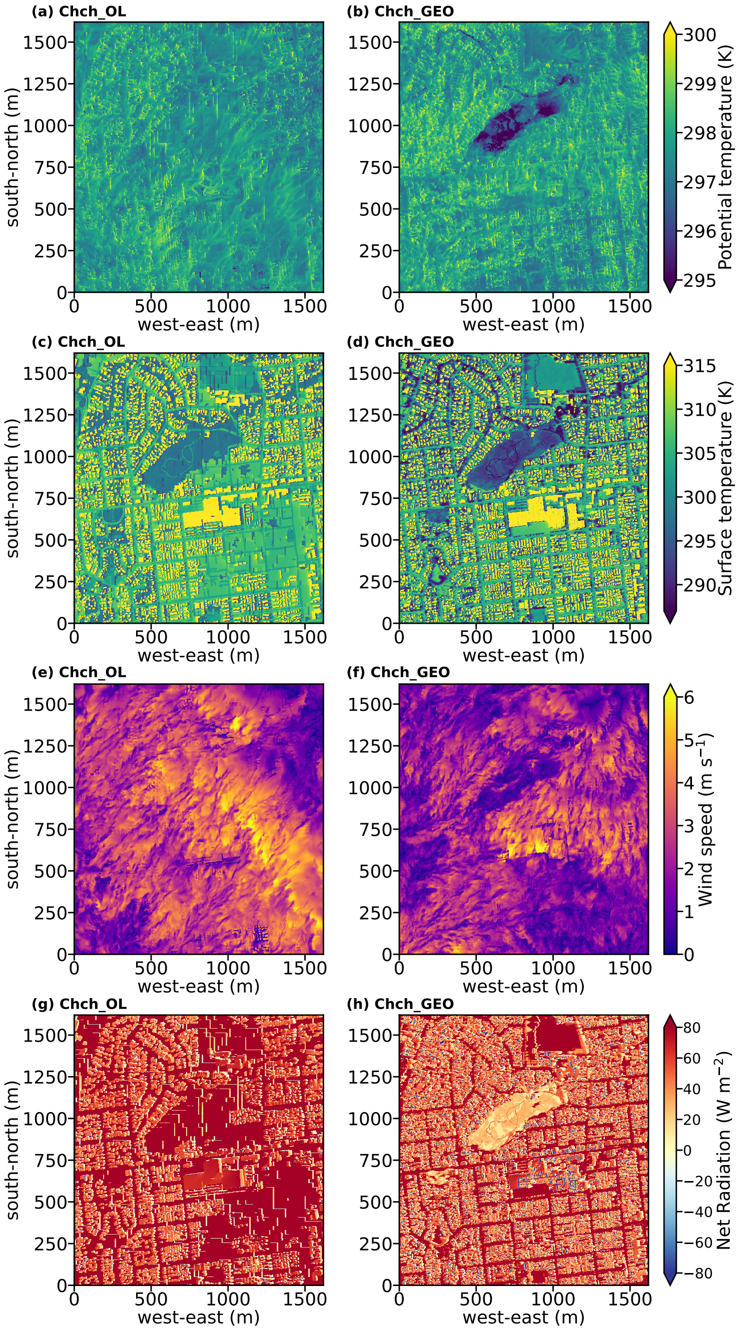

Figure 9. Similar to Figure 6, but for the Christchurch case: (a-b) 2 m potential temperature, (c-d) surface temperature, (e-f) 10 m wind speed, and (g-h) surface net radiation. Refer to panel label for the corresponding simulation.

https://doi.org/10.5194/gmd-2023-150
Preprint. Discussion started: 22 August 2023

Geoscientific Model Development Discussions

5 Conclusions

This study presents GEO4PALM, a utility written in Python that generates PALM static driver input for PALM model system. With GEO4PALM, PALM users now have freedom to generate static driver with any geospatial data they have had and/or the online data provided. Two application examples were given to demonstrate the applicability of GEO4PALM. The results show that GEO4PALM is a useful tool to realise near-surface microscale structures. When the same geospatial input data are used, the GEO4PALM static drivers can present simulation results comparable to the PALM CSD static drivers. There are differences between the results generated by the two tools, while the optimal simulation setup needs to be investigated in future simulations and experiments. To our best knowledge, no widely applicable geospatial data processing tool was available in the PALM community. GEO4PALM simplifies the processes of data acquisition, data pre-processing, and data conversion, and provides a standardized process such that tedious and time-consuming tasks are significantly reduced.

GEO4PALM is developed as a free, open-source, and community-driven tool and is distributed on GitHub (https://github.com/dongqi-DQ/GEO4PALM; last access: 21 June 2023). As discussed, GEO4PALM does not cover all the variables in static driver of PALM model system. The fidelity and features of static driver input depend on the input geospatial data quality and availability, which have been a particular challenge when conducting microscale simulations. GEO4PALM has provided several interfaces for users to download global geospatial data sets, which includes the basic features of PALM static input, such as topography and land use typology. However, many regions in the world still do not have high resolution geospatial data to realise simulations with high fidelity. High resolution building height data are still not widely available making simulations over human settlement difficult. Fortunately, with increasing efforts towards research at the microscale, especially urban climate research, new data sets have been developed for microscale simulations. For example, high resolution geospatial data can be obtained from Geoscape (https://geoscape.com.au/; last access: 21 June 2023) for applications in Australia. Microsoft provides AI-assisted building footprints mapping (https://www.microsoft.com/en-us/maps/building-footprints; last access: 21 June 2023). Esch et al. (2022) presented World Settlement Footprint 3D, which provides three-dimensional morphology and density of buildings worldwide. Some of these data sets are not freely available or need to be acquired based on individual requests, while GEO4PALM accepts all geospatial data in geotiff format. Once users have obtained the data sets they desired, GEO4PALM is able to process such data for PALM simulations. Furthermore, another common challenge in the land use data is that many land use classification data sets only classify urban areas into a limited number of typologies. This could lead to loss of fidelity. Lipson et al. (2022) has described a data transformation method to make the urban land use data more descriptive. This may potentially improve the quality of land use classification data sets, for example, for applications in New Zealand, where only one type of land use was classified for urban areas. For plant canopy, GEO4PALM currently only accepts vegetation heights as input due to lack of geospatial data. In the future, we aim to improve this feature based on the PALM CSD tool (Heldens et al., 2020) and to include more data sources if available.

GEO4PALM accepts any geospatial data sets as input and is easily adaptive to new data downloading interface. With the development of geospatial data sets towards better spatial coverage and data quality, GEO4PALM can be improved and extended.

https://doi.org/10.5194/gmd-2023-150
Preprint. Discussion started: 22 August 2023

Geoscientific Model Development
Discussions

All PALM users are encouraged to provide feedback, and report bugs and issues via the issue system provided by GitHub. Any optimisation, modification, or contribution made to the code are welcome and much appreciated.

Code availability. The PALM model system 6.0 is a free and open-source numerical model distributed on GitLab (https://gitlab.palm-model.org/releases/palm_model_system/-/releases; last access: 21 June 2023) under the GNU General Public License v3.0. The exact PALM model source code used for this study is release 22.10 (https://gitlab.palm-model.org/releases/palm_model_system/-/releases/v22.10; last access: 21 June 2023). GEO4PALM code is freely available at https://doi.org/10.5281/zenodo.8062322 and https://github.com/dongqi-DQ/GEO4PALM (last access: 21 June 2023) under the GNU General Public License v3.0. Details of Python packages and environment used for GEO4PALM are given on the GitHub repository. PALM CSD code is included in the PALM source code with technical information available at https://palm.muk.uni-hannover.de/trac/wiki/doc/app/iofiles/pids/palm_csd (last access: 21 June 2023).

Data availability. All PALM input files for the Christchurch case described in Section 4.2 are available in the supplement. The GEO4PALM namelists for both the Christchurch and Berlin cases are included in the supplement. Geospatial data availability for the application examples refer to the main text. Other data sets can be provided upon request.

Appendix A: GEO4PALM step-by-step guide

A more detailed user manual is available at https://github.com/dongqi-DQ/GEO4PALM (last access: 21 June 2023).

A1 Step 1: prepare namelist

Namelist should be provided in `./JOBS/case_name/` folder as follows:

```
[case]
case_name          -  name of the case
origin_time        -  date and time at model start*
default_proj       -  default is EPSG:4326. This projection uses lat/lon to
                      locate domain. This may not be changed.
config_proj        -  projection of input tif files. GEO4PALM will automatically assign
                      the UTM zone if not provided.
                      We recommend users use local projection with units in metre,
                      e.g. for New Zealand users, EPSG:2193 is a recommended choice.

[domain]
ndomain            -  maximum number of domains, when >=2, domain nesting is enabled
centlat, centlon   -  centre latitude and longitude of the first domain. Note this is
```

https://doi.org/10.5194/gmd-2023-150
Preprint. Discussion started: 22 August 2023


```
                                not required for nested domains
nx                   -   number of grid points along x-axis
ny                   -   number of grid points along y-axis
nz                   -   number of grid points along z-axis
dx                   -   grid spacing in meters along x-axis
dy                   -   grid spacing in meters along y-axis
dz                   -   grid spacing in meters along z-axis
z_origin             -   elevated terrain mean grid position in meters
                         (leave as 0.0 if unknown)
ll_x                 -   lower left corner distance to the first domain in
                          meters along x-axis
ll_y                 -   lower left corner distance to the first domain in
                         meters along y-axis

[geotif]             -   required input from user; can be provided by users
                         in the INPUT folder or "online"
sst                  -   input for water temperature
dem                  -   digital elevation model input for topography
lu                   -   land use classification
resample_method      -   method to resample geotiff files for interpolation/extrapolation

# if NASA API is used format in YYYY-MM-DD
# SST date should be the same as the orignin_time

## No need to change start/end dates for NASA SRTMGL1_NC.003
dem_start_date = '2000-02-12',
dem_end_date = '2000-02-20',
## start/end dates for land use data set
lu_start_date = '2020-10-01',
lu_end_date = '2020-10-30',

[urban]              - input for urban canopy model; can leave as "" if this
                       feature is not included in the simulations, or provided by
                       users; or online from OSM
bldh                 - input for building height
```

https://doi.org/10.5194/gmd-2023-150
Preprint. Discussion started: 22 August 2023


```
bldid               - input for building ID
pavement            - input for pavement type
street              - input for building ID

[plant]             - input for plant canopy model; can leave as "" if this
                      feature is not included in the simulations,
                      or provided by users
tree_lai_max        - input value for maximum leaf area index (LAI)
lad_max_height      - input value for the height where the leaf area index (LAI)
                       reaches leave area density (LAD)
sfch                - input for plant height; this is for leave area density (LAD)
```

A2 Step 2: link land use classification lookup table

A lookup table is required to convert typologies of land use obtained from a geospatial data set to PALM compatible land use classification. Before running GEO4PALM, link the corresponding csv file to `util/lu_2_PALM_num.csv`:

```
# In util/ directory
ln -sf lu_csv/{your_csv} lu_2_PALM_num.csv
```

A3 Step 3: run the main script

```
Python run_config_static.py case_name
```

Appendix B: Land use type lookup table

https://doi.org/10.5194/gmd-2023-150
Preprint. Discussion started: 22 August 2023

Table B1. Lookup table to convert NZLCDB v5.0 classes to PALM vegetation type, pavement type, building type, water type, and soil type.

Class code	Class name	vegetation type	pavement type	building type	water type	soil type
1	Built-up Area (settlement)			3		
2	Urban Parkland/Open Space	3				3
5	Transport Infrastructure		3			1
6	Surface Mine or Dump	1				1
10	Sand or Gravel	1				1
11	River and Lakeshore Gravel and Rock				2	
12	Landslide	1				1
13	Alpine Gravel and Rock	1				1
14	Permanent Snow and Ice	1				1
15	Alpine Grass/Herbfield	3				6
16	Gravel or Rock	1				1
20	Lake or Pond				1	
21	River				2	
22	Estuarine Open Water				2	
30	Short-rotation Cropland	2				6
33	Orchards, Vineyards or Other Perennial Crops	11				6
40	High Producing Exotic Grassland	3				6

https://doi.org/10.5194/gmd-2023-150
Preprint. Discussion started: 22 August 2023

Table B1. Continued.

Class code	Class name	vegetation type	pavement type	building type	water type	soil type
41	Low Producing Grassland	3				6
43	Tall Tussock Grassland	8				6
44	Depleted Grassland	3				1
45	Herbaceous Freshwater Vegetation	14				6
46	Herbaceous Saline Vegetation	16				6
47	Flaxland	8				6
50	Fernland	3				6
51	Gorse and/or Broom	8				6
52	Manuka and/or Kanuka	8				6
54	Broadleaved Indigenous Hardwoods	7				6
55	Sub Alpine Shrubland	15				6
56	Mixed Exotic Shrubland	15				6
58	Matagouri or Grey Scrub	8				6
60	Minor Shelterbelts	18				6
61	Major Shelterbelts	18				6
64	Forest - Harvested	17				6
68	Deciduous Hardwoods	5				6
69	Indigenous Forest	6				6
70	Mangrove	14				6
71	Exotic Forest	7				6
-9999	Not a number or no data				2	

https://doi.org/10.5194/gmd-2023-150
Preprint. Discussion started: 22 August 2023

Geoscientific Model Development
Discussions
Open Access
EGU

Table B2. Lookup table to convert NASA LC_Type01 classes to PALM vegetation type, pavement type, building type, water type, and soil type.

Class code	Class name	vegetation type	pavement type	building type	water type	soil type
1	Evergreen Needleleaf Forests	4				5
2	Evergreen Broadleaf Forests	6				5
3	Deciduous Needleleaf Forests	5				5
4	Deciduous Broadleaf Forests	7				5
5	Mixed Forests	17				5
6	Closed Shrublands	15				5
7	Open Shrublands	16				5
8	Woody Savannas	17				4
9	Savannas	17				4
10	Grasslands	8				3
11	Permanent Wetlands	14				6
12	Croplands	11				4
13	Urban and Built-up Lands			3		
14	Cropland/Natural Vegetation Mosaics	2				5
15	Permanent Snow and Ice	13				1
16	Barren	12				1
17	Water Bodies				2	
255	Unclassified				2	
-9999	Not a number or no data				2	

https://doi.org/10.5194/gmd-2023-150
Preprint. Discussion started: 22 August 2023

Geoscientific Model Development Discussions
Open Access
EGU

Table B3. Lookup table to convert ESA WorldCover classes to PALM vegetation type, pavement type, building type, water type, and soil type.

Class code	Class name	vegetation type	pavement type	building type	water type	soil type
10	Tree cover	5				4
20	Shrubland	15				4
30	Grassland	3				3
40	Cropland	2				3
50	Built-up			3		
60	Bare / sparse vegetation	1				1
70	Snow and ice	13				1
80	Permanent water bodies				2	
90	Herbaceous wetland	14				6
95	Mangroves	6				6
100	Moss and lichen	3				6
-9999	Not a number or no data				2	
255	Not a number or no data				2	

Author contributions. DL was responsible for the data acquisition, conceptualisation of the GEO4PALM tool, initial and major development of GEO4PALM, GEO4PALM code distribution and documentation, conducting PALM simulations, formal analysis, and visualisation. JZ contributed to conceptualisation of GEO4PALM v1.1, major GEO4PALM development, GEO4PALM documentation, and developed the PALM domain utility. BK provided the DLR data sets and contributed to conceptualisation of case studies. DL wrote the manuscript with contributions from JZ, BK, MK, and LER. DL, JZ, BK, MK, and LER reviewed the manuscript.

Competing interests. The authors declare that they have no conflict of interest.

Acknowledgements. The contributions of Dongqi Lin and Jiawei Zhang were funded by the New Zealand Ministry of Business, Innovation and Employment (MBIE) prjoect 'Extreme wildfire: Our new reality – are we ready?' (Grant No. C04X2103). Marwan Katurji was supported by the MBIE extreme wildfire project (Grant No. C04X2103), and the Royal Society of New Zealand (Grant No. RDF-UOC1701). The contribution of Basit Khan was supported by the MOSAIK and MOSAIK-2 projects, which are funded by the German Federal Ministry of Education and Research (BMBF) (Grant Nos. 01LP1601A and 01LP1911H), within the framework of Research for Sustainable Development (FONA; http://www.fona.de, last access: 21 June, 2023). We performed PALM simulations presented in this study on New Zealand eScience Infrastructure (NeSI) high-performance computing facilities. GEO4PALM development was conducted on the School of Earth and Environment (SEE) computing cluster and the University of Canterbury high-performance research computing cluster (RCC). The early development of GEO4PALM was inspired by WRF2PALM code (now replaced by WRF4PALM toolkit) developed by Ricardo Faria from the Oceanic Observatory of Madeira.

https://doi.org/10.5194/gmd-2023-150
Preprint. Discussion started: 22 August 2023

Geoscientific Model Development
Discussions

References

Belda, M., Resler, J., Geletič, J., Krč, P., Maronga, B., Sühring, M., Kurppa, M., Kanani-Sühring, F., Fuka, V., Eben, K., Benešová, N., and Auvinen, M.: Sensitivity analysis of the PALM model system 6.0 in the urban environment, Geoscientific Model Development, 14, 4443–4464, https://doi.org/10.5194/gmd-14-4443-2021, publisher: Copernicus GmbH, 2021.

Boeing, G.: OSMnx: New methods for acquiring, constructing, analyzing, and visualizing complex street networks, Computers, Environment and Urban Systems, 65, 126–139, iSBN: 0198-9715 Publisher: Elsevier, 2017.

Bou-Zeid, E., Meneveau, C., and Parlange, M. B.: Large-eddy simulation of neutral atmospheric boundary layer flow over heterogeneous surfaces: Blending height and effective surface roughness, Water Resources Research, 40, 2004.

Envirionment Canterbury Regional Council: Christchurch and Ashley River, Canterbury, New Zealand 2018, https://doi.org/https://doi.org/10.5069/G91J97WQ, 2020.

Esch, T., Brzoska, E., Dech, S., Leutner, B., Palacios-Lopez, D., Metz-Marconcini, A., Marconcini, M., Roth, A., and Zeidler, J.: World Settlement Footprint 3D - A first three-dimensional survey of the global building stock, Remote Sensing of Environment, 270, 112 877, https://doi.org/10.1016/j.rse.2021.112877, 2022.

Gehrke, K. F., Sühring, M., and Maronga, B.: Modeling of land–surface interactions in the PALM model system 6.0: land surface model description, first evaluation, and sensitivity to model parameters, Geoscientific Model Development, 14, 5307–5329, https://doi.org/10.5194/gmd-14-5307-2021, publisher: Copernicus GmbH, 2021.

Gillies, S. et al.: Rasterio: geospatial raster I/O for Python programmers [Software], https://github.com/rasterio/rasterio, 2019.

Gronemeier, T., Raasch, S., and Ng, E.: Effects of Unstable Stratification on Ventilation in Hong Kong, Atmosphere, 8, https://doi.org/10.3390/atmos8090168, 2017.

Heldens, W., Burmeister, C., Kanani-Sühring, F., Maronga, B., Pavlik, D., Sühring, M., Zeidler, J., and Esch, T.: Geospatial input data for the PALM model system 6.0: model requirements, data sources and processing, Geoscientific Model Development, 13, 5833–5873, 2020.

Hoyer, S. and Hamman, J.: xarray: ND labeled arrays and datasets in Python, Journal of Open Research Software, 5, 2017.

Jordahl, K., den Bossche, J. V., Fleischmann, M., Wasserman, J., McBride, J., Gerard, J., Tratner, J., Perry, M., Badaracco, A. G., Farmer, C., Hjelle, G. A., Snow, A. D., Cochran, M., Gillies, S., Culbertson, L., Bartos, M., Eubank, N., maxalbert, Bilogur, A., Rey, S., Ren, C., Arribas-Bel, D., Wasser, L., Wolf, L. J., Journois, M., Wilson, J., Greenhall, A., Holdgraf, C., Filipe, and Leblanc, F.: geopandas/geopandas: v0.8.1 [Software], https://doi.org/10.5281/zenodo.3946761, 2020.

Khan, B., Banzhaf, S., Chan, E. C., Forkel, R., Kanani-Sühring, F., Ketelsen, K., Kurppa, M., Maronga, B., Mauder, M., Raasch, S., Russo, E., Schaap, M., and Sühring, M.: Development of an atmospheric chemistry model coupled to the PALM model system 6.0: implementation and first applications, Geoscientific Model Development, 14, 1171–1193, https://doi.org/10.5194/gmd-14-1171-2021, 2021.

Krč, P., Resler, J., Sühring, M., Schubert, S., Salim, M. H., and Fuka, V.: Radiative Transfer Model 3.0 integrated into the PALM model system 6.0, Geoscientific Model Development, 14, 3095–3120, 2021.

Kurppa, M., Hellsten, A., Roldin, P., Kokkola, H., Tonttila, J., Auvinen, M., Kent, C., Kumar, P., Maronga, B., and Järvi, L.: Implementation of the sectional aerosol module SALSA2.0 into the PALM model system 6.0: model development and first evaluation, Geoscientific Model Development, 12, 1403–1422, https://doi.org/10.5194/gmd-12-1403-2019, publisher: Copernicus GmbH, 2019.

Kurppa, M., Roldin, P., Strömberg, J., Balling, A., Karttunen, S., Kuuluvainen, H., Niemi, J. V., Pirjola, L., Rönkkö, T., Timonen, H., et al.: Sensitivity of spatial aerosol particle distributions to the boundary conditions in the PALM model system 6.0, Geoscientific Model Development, 13, 5663–5685, 2020.

https://doi.org/10.5194/gmd-2023-150
Preprint. Discussion started: 22 August 2023

Lalic, B. and Mihailovic, D. T.: An empirical relation describing leaf-area density inside the forest for environmental modeling, Journal of Applied Meteorology, 43, 641–645, 2004.

Land Information New Zealand: New Zealand Building Outlines (All Sources), https://data.linz.govt.nz/layer/101292-nz-building-outlines-all-sources/, 2020.

Landcare Research: LCDB v5.0 - Land Cover Database version 5.0, Mainland New Zealand, https://lris.scinfo.org.nz/layer/104400-lcdb-v50-land-cover-database-version-50-mainland-new-zealand/, 2020.

Lin, D., Khan, B., Katurji, M., Bird, L., Faria, R., and Revell, L. E.: WRF4PALM v1. 0: a mesoscale dynamical driver for the microscale PALM model system 6.0, Geoscientific Model Development, 14, 2503–2524, https://doi.org/https://doi.org/10.5194/gmd-14-2503-2021, 2021.

Lin, D., Katurji, M., Revell, L. E., Khan, B., and Sturman, A.: Multiscale meteorological controls and impact of soil moisture heterogeneity on radiation fog in complex terrain, EGUsphere, 2022, 1–30, https://doi.org/10.5194/egusphere-2022-1229, 2022.

Lipson, M. J., Nazarian, N., Hart, M. A., Nice, K. A., and Conroy, B.: A Transformation in City-Descriptive Input Data for Urban Climate Models, Frontiers in Environmental Science, 10, https://www.frontiersin.org/articles/10.3389/fenvs.2022.866398, 2022.

Mahrt, L. and Hristov, T.: Is the Influence of Stability on the Sea Surface Heat Flux Important?, Journal of Physical Oceanography, 47, 689–699, https://doi.org/10.1175/JPO-D-16-0228.1, 2017.

Maronga, B., Hartogensis, O. K., Raasch, S., and Beyrich, F.: The effect of surface heterogeneity on the structure parameters of temperature and specific humidity: A large-eddy simulation case study for the LITFASS-2003 experiment, Boundary-layer meteorology, 153, 441–470, iSBN: 1573-1472 Publisher: Springer, 2014.

Maronga, B., Gryschka, M., Heinze, R., Hoffmann, F., Kanani-Sühring, F., Keck, M., Ketelsen, K., Letzel, M. O., Sühring, M., and Raasch, S.: The Parallelized Large-Eddy Simulation Model (PALM) version 4.0 for atmospheric and oceanic flows: model formulation, recent developments, and future perspectives, Geoscientific Model Development, 8, 2515–2551, https://doi.org/10.5194/gmd-8-2515-2015, 2015.

Maronga, B., Banzhaf, S., Burmeister, C., Esch, T., Forkel, R., Fröhlich, D., Fuka, V., Gehrke, K. F., Geletič, J., and Russo, E.: Overview of the PALM model system 6.0, Geoscientific Model Development, 13, 1335–1372, https://doi.org/10.5194/gmd-13-1335-2020, 2020.

Meyer, D. and Riechert, M.: Open source QGIS toolkit for the Advanced Research WRF modelling system, Environmental Modelling & Software, 112, 166–178, https://doi.org/10.1016/j.envsoft.2018.10.018, 2019.

NASA/JPL: GHRSST Level 4 MUR Global Foundation Sea Surface Temperature Analysis (v4.1), https://doi.org/10.5067/GHGMR-4FJ04, 2015.

Powers, J. G., Klemp, J. B., Skamarock, W. C., Davis, C. A., Dudhia, J., Gill, D. O., Coen, J. L., Gochis, D. J., Ahmadov, R., Peckham, S. E., Grell, G. A., Michalakes, J., Trahan, S., Benjamin, S. G., Alexander, C. R., Dimego, G. J., Wang, W., Schwartz, C. S., Romine, G. S., Liu, Z., Snyder, C., Chen, F., Barlage, M. J., Yu, W., and Duda, M. G.: The Weather Research and Forecasting Model: Overview, System Efforts, and Future Directions, Bulletin of the American Meteorological Society, 98, 1717–1737, https://doi.org/10.1175/BAMS-D-15-00308.1, 2017.

Raasch, S. and Schröter, M.: PALM - A large-eddy simulation model performing on massively parallel computers, Meteorologische Zeitschrift, 10, 363–372, https://doi.org/10.1127/0941-2948/2001/0010-0363, 2001.

Rabus, B., Eineder, M., Roth, A., and Bamler, R.: The shuttle radar topography mission—a new class of digital elevation models acquired by spaceborne radar, ISPRS Journal of Photogrammetry and Remote Sensing, 57, 241–262, https://doi.org/10.1016/S0924-2716(02)00124-7, 2003.

https://doi.org/10.5194/gmd-2023-150
Preprint. Discussion started: 22 August 2023

Geoscientific Model Development Discussions

Resler, J., Krč, P., Belda, M., Juruš, P., Benešová, N., Lopata, J., Vlček, O., Damašková, D., Eben, K., Derbek, P., Maronga, B., and Kanani-Sühring, F.: PALM-USM v1.0: A new urban surface model integrated into the PALM large-eddy simulation model, Geoscientific Model Development, 10, 3635–3659, https://doi.org/10.5194/gmd-10-3635-2017, publisher: Copernicus GmbH, 2017.

Resler, J., Eben, K., Geletič, J., Krč, P., Roseckỳ, M., Sühring, M., Belda, M., Fuka, V., Halenka, T., Huszár, P., et al.: Validation of the PALM model system 6.0 in a real urban environment: a case study in Dejvice, Prague, the Czech Republic, Geoscientific Model Development, 14, 4797–4842, 2021.

Rihani, J. F., Chow, F. K., and Maxwell, R. M.: Isolating effects of terrain and soil moisture heterogeneity on the atmospheric boundary layer: Idealized simulations to diagnose land-atmosphere feedbacks, Journal of Advances in Modeling Earth Systems, 7, 915–937, iSBN: 1942-2466 Publisher: Wiley Online Library, 2015.

Salim, M. H., Schubert, S., Resler, J., Krč, P., Maronga, B., Kanani-Sühring, F., Sühring, M., and Schneider, C.: Importance of radiative transfer processes in urban climate models: a study based on the PALM 6.0 model system, Geoscientific Model Development, 15, 145–171, 2022.

Snow, A., BENR0, Cook, J., LiamRMoore, Taves, M., and Pierrick, R.: corteva/geocube: 0.2.0 [Software], https://doi.org/10.5281/zenodo.6399307, 2022a.

Snow, A., Brochart, D., Raspaud, M., Bell, R., et al.: corteva/rioxarray: 0.11.1 [Software], https://doi.org/10.5281/zenodo.6478182, 2022b.

Srivastava, A., Kumari, N., and Maza, M.: Hydrological Response to Agricultural Land Use Heterogeneity Using Variable Infiltration Capacity Model, Water Resources Management, 34, 3779–3794, https://doi.org/10.1007/s11269-020-02630-4, 2020.

Sulla-Menashe, D. and Friedl, M. A.: User guide to collection 6 MODIS land cover (MCD12Q1 and MCD12C1) product, USGS: Reston, VA, USA, 1, 18, 2018.

Vollmer, L., Lee, J. C.-Y., Steinfeld, G., and Lundquist, J. K.: A wind turbine wake in changing atmospheric conditions: LES and lidar measurements, Journal of Physics: Conference Series, 854, 012 050, https://doi.org/10.1088/1742-6596/854/1/012050, publisher: IOP Publishing, 2017.

Wolf, T., Pettersson, L. H., and Esau, I.: A very high-resolution assessment and modelling of<? xmltex\break?> urban air quality, Atmospheric Chemistry and Physics, 20, 625–647, 2020.

Wolf, T., Pettersson, L. H., and Esau, I.: Dispersion of particulate matter ($PM_{2.5}$) from wood combustion for residential heating: optimization of mitigation actions based on large-eddy simulations, Atmospheric Chemistry and Physics, 21, 12 463–12 477, https://doi.org/10.5194/acp-21-12463-2021, publisher: Copernicus GmbH, 2021.

Zanaga, D., Van De Kerchove, R., De Keersmaecker, W., Souverijns, N., Brockmann, C.and Quast, R., Wevers, J., Grosu, A., Paccini, A., Vergnaud, S., Cartus, O., Santoro, M., Fritz, S., Georgieva, I., Lesiv, M., Carter, S., Herold, M., Li, L., Tsendbazar, N., Ramoino, F., and Arino, O.: ESA WorldCover 10 m 2020 v100, https://doi.org/10.5281/zenodo.5571936, 2021.

Zanaga, D., Van De Kerchove, R., Daems, D., De Keersmaecker, W., Brockmann, C., Kirches, G., Wevers, J., Cartus, O., Santoro, M., Fritz, S., Lesiv, M., Herold, M., Tsendbazar, N., Xu, P., Ramoino, F., and Arino, O.: ESA WorldCover 10 m 2021 v200, https://doi.org/10.5281/zenodo.7254221, 2022.