https://doi.org/10.5194/egusphere-2025-5521
Preprint. Discussion started: 22 December 2025

Towards standardising output datasets using the numerical obstacle-resolving model MITRAS as an example

Vivien Voss[1], K. Heinke Schlünzen[1], David Grawe[1], and Karolin S. Samsel[1,2]

[1]Meteorological Institute, CEN, Universität Hamburg, Hamburg, Germany
[2]Climate Service Center Germany (GERICS), Helmholtz-Zentrum Hereon, Chilehaus, Fischertwiete 1, 20095 Hamburg, Germany

Correspondence: Vivien Voss (vivien.voss@uni-hamburg.de)

Abstract. The publication of well-described FAIR datasets is an important part of atmospheric modelling and research. Data standards ensure that datasets are delivered in a consistent way that is easy to understand for a data user. Standards define how the data is described, i.e. which variable names, descriptions and data formats are used. However, existing model data standards such as the CF conventions are mainly adapted for global or regional scale models. For atmospheric micro-scale obstacle-resolving (urban) models (ORMs), there is no discipline-specific model data standard and the existing ones are not fully suitable to adequately describe ORM datasets. To overcome the lack of standardisation processes, the ATMODAT STANDARD has been developed to promote the publication of FAIR datasets when no discipline-specific standard is available. This paper describes the process of producing standardised model results. The processing for ORM MITRAS serves as an example to show possible ways for the publication of FAIR datasets. The adaptation of the model's post-processing routine M2CDF and the development of a new post-processing routine called NC2ATMODAT are shown. The last may be applicable by other ORM modellers, its limitations, challenges and further use cases are discussed. Application of the two post-processors allows the preparation of datasets according to the requirements of the CF convention and the ATMODAT STANDARD. The first standardised MITRAS datasets are successfully processed and published.

1 Introduction

Numerical modelling is a commonly used tool to investigate complex physical processes in the atmosphere. Models can simulate processes on different scales, ranging from the global scale (weather forecasts or climate simulations) via regional scales (investigating regional effects) to the urban and neighbourhood scale, where processes in urban areas are analysed. An important aspect of atmospheric modelling is the publication and exchange of model results. Especially in the climate modelling community, model datasets have been exchanged for decades, e.g. with the Coupled Model Intercomparison Project (CMIP) (Eyring et al., 2016; Touzé-Peiffer et al., 2020) as a prominent example. The need for standardised model output data was established. The awareness and amount of publication of research data are increasing, and storing such data is becoming more and more important (Bermudez, 2017; De Vos et al., 2020; Eggleton and Winfield, 2020). Ideally, published model datasets should fulfil the FAIR principles (Wilkinson et al., 2016), meaning that the datasets should be "findable, accessible, interoperable and reusable". FAIR data ensures that the handling of model datasets is convenient and easy for data users. One way

https://doi.org/10.5194/egusphere-2025-5521
Preprint. Discussion started: 22 December 2025

to achieve and ensure FAIR datasets is to apply data standards to datasets, as they provide structure and guidelines for the production of well-described and FAIR-compliant datasets.

In atmospheric and climate science, the *network common data form (netCDF)* (Rew and Davis, 1990) is a commonly used data format and well established within the community. It is a self-describing data format for scientific data that allows a multi-dimensional structure of the data (Rew and Davis, 1990). It works well for storing time and space dependent data. NetCDF is often used alongside the Climate and Forecast (CF) conventions, which provide guidelines for the provision of (meta)data information in the netCDF data structure (Eaton et al., 2024). These conventions ensure that NetCDF files are well-described, and easily understood and usable by data users. An important feature of the CF conventions is the provision and usage of a standardised vocabulary for variable names[1]. The netCDF data format, together with the CF conventions, is widely used and frequently adapted by global and regional scale models (e.g. models within CMIP), as well as in other applications that produce large and or multi-dimensional datasets.

However, the publication of FAIR datasets is not trivial in research disciplines, where no data standardisation processes are established. Such a discipline is the high-resolution urban climate modelling, carried out with obstacle-resolving models (ORMs). The existing standards, such as the CF conventions, are not sufficient to meet the needs of these model types, since they currently do not cover discipline-specific variables and features, whereas no overarching discipline-specific standardisation is established. This hinders the publication of well-described model datasets in this discipline and impairs the reusability of such model datasets.

ORMs are obstacle-resolving atmospheric micro-scale models, focusing on processes in small domains, like urban neighbourhoods and city quarters. They use grid sizes of metres and provide results in a high spatial and temporal resolution (Oke et al., 2017; Baklanov et al., 2018). The model domains are typically not much larger than a few km^2, and obstacles, like buildings or trees, are resolved. ORMs can be either numerical models or physical models, e.g. wind tunnels or water tanks (Oke et al., 2017).

Each ORM prepares and provides data in a model-specific way. Many of them use their own model-specific data formats (such as binary or text files), whereas some ORMs also provide their model data as netCDF files (Jänicke et al., 2021; Voss et al., 2024). Models such as MISKAM (Eichhorn and Kniffka, 2010; Eichhorn, 2011), OPENFOAM (OpenFOAM, 2025), ANSYS FLUENT (Poole et al., 1995), ENVI-MET (Bruse and Fleer, 1998) or MITRAS (Schlünzen et al., 2003; Salim et al., 2018) provide their data mainly in their model-specific way, but also allow a conversion to netCDF with tools provided by the model developer. The ORM PALM-4U uses netCDF and applies a model-specific data standard to the netCDF files that follows the CF conventions (Scherer et al., 2019, 2022). For MITRAS, not all variables were initially adapted to the CF conventions. Therefore, the use of published data is difficult, as identifying variables without the use of standardised variable names is only possible when using the technical documentation (e.g. Schlünzen et al. (2018b) or Scherer et al. (2022)). In general, if a model provides data in their model-specific formats, the use or access of the data might be hindered and complicated, especially if special tools or programs to access or convert the data are required. Due to the lack of general standards for ORM results, the extent to which ORM datasets are FAIR varies. As ORMs continue to evolve and become more widely used, the

[1]https://cfconventions.org/Data/cf-standard-names/current/build/cf-standard-name-table.html

https://doi.org/10.5194/egusphere-2025-5521
Preprint. Discussion started: 22 December 2025

need for standardised datasets and a common vocabulary for ORM variables will increase. In obstacle-resolving modelling, the reusability and inter-comparison of model data will become an upcoming and important aspect in this discipline (WMO, 2023). With urban areas now specifically addressed in IPCC [2], it seems even more important that ORM datasets are published FAIR and become reusable.

To overcome the lack of standardisation, e.g. in the ORM community, and to guide the creation of well-described and FAIR model datasets, the Atmospheric Model Data Standard (ATMODAT STANDARD) (Ganske et al., 2021) was developed. This standard is intended to be applied to model datasets from different atmospheric modelling disciplines, especially for those that do not have a discipline-specific standard. First datasets prepared according to the ATMODAT STANDARD were already published. The ATMODAT STANDARD was applied to datasets containing reanalysis data of observations to classify weather types over the Baltic and North Sea (Loewe, 2022; Loewe and Schade, 2024), to datasets for sensitivity studies of aerosol and cloud microphysics processes simulated with the ECHAM6-HAM2 model (Mülmenstädt, 2022), to datasets from urban atmospheric MITRAS simulations focusing on dispersions of particles (Voss, 2023) and precipitation (Ferner et al., 2024; Samsel et al., 2025) and datasets of the urban climate in Hamburg for winter and summer generated with the mesoscale model METRAS (Boettcher et al., 2024a, b).

In this paper, the challenges in creating a FAIR ORM data publication that fulfils the ATMODAT STANDARD, are discussed. For the description of a successful path to such a publication, the ORM MITRAS will serve as an example. As a first step, the application of the CF convention on MITRAS datasets are investigated. Therefore, CF standard names and attributes need to be retrieved and assigned to the model specific variables, hereby revealing issues with ORM specific variables (Section 2). Modifications of the post-processing of the ORM MITRAS are described, and details of the necessary steps for applying the ATMODAT STANDARD to ORM data are provided (Section 3). The application of the new standardisation processing is applied to datasets from the ORM MITRAS that serves throughout this paper as an example (Section 4).

2 Determination of ORM variables

The CF conventions provide a collection of standardised variable names (CF Standard Name Table[3]), the so-called standard names, which make the meaning of a netCDF variable self-explanatory. Without a standardised vocabulary, each modelling group would apply their model-specific and individual naming conventions to the variables. Data user unfamiliar with the model might be unsure if those self-named variables describe the same properties they assume and that they want to use. Standardised variable names are already defined in the CF conventions for many common atmospheric variables such as wind speed, temperature, humidity, etc. Hereby, a `standard_name` should only be assigned to a variable if the model variable fits the definition of the corresponding entry in the CF Standard Name Table. If no matching standard name is available, a variable

[2](https://www.ipcc.ch/report/special-report-on-climate-change-and-cities/; last acccess 17.07.2025)
[3]https://cfconventions.org/Data/cf-standard-names/current/build/cf-standard-name-table.html

should be assigned with a `long_name` that contains a descriptive variable name to indicate the purpose of the variable (Eaton et al., 2024). This is useful if (discipline-specific) variables are not represented by existing standard names.

Many ORM-specific properties are currently neither standardised within the community nor available in the CF conventions. Such non-standardised ORM variables are e.g. related to the obstacle and its surfaces, such as wall temperatures, radiation fluxes from surfaces, the amount of rainwater on roofs or walls, thermal indices or obstacles themselves, if they are provided as an own variable.

To illustrate what is required to apply the CF conventions and the ATMODAT Standard, the output of ORM MITRAS is used. As a first step, all variables available in the model code, in the post-processor code and in the model documentation (Schlünzen et al., 2018b) were collected and summarised. Based on this collection, it became visible that variable names provided for reuse needed to be more self-explanatory and required adjustments.

To align the variable names with those used in other ORMs and to gain an understanding of the ORM-specific variables typically used, the variable names from PALM-4U (Maronga et al., 2019, 2020) were considered. Hereby, PALM-4U was selected for comparison because it is a state-of-the-art ORM that uses netCDF as output data format, uses standard names for its variable descriptions and follows the CF conventions to describe the datasets (Scherer et al., 2019, 2022).

The comparison of variables in the ORMs MITRAS and PALM-4U was based on matching the provided `long names`, units and descriptions in the respective model documentations. In addition, the MITRAS variables were checked against the CF Standard Name Table[4], and fitting `standard names` were added to the MITRAS model variables. Furthermore, typical ORM variables which are not covered by the existing standard names were identified.

Based on the comparison of variable long names between MITRAS and PALM-4U, it became clear that the mapping of variables to each other is not trivial. It is not always clear whether similar model variable names represent the same properties, which would only be conclusive if the respective model routines were checked. The meaning was predominantly unclear for variables with slightly differently constructed variable names.

The assignment of CF standard names to MITRAS variables was also not always trivial, as some aspects of the MITRAS variables do not always fully match the definitions of existing standard names. Here, it was necessary to consider whether the MITRAS variable in question needs to be adjusted in favour of the existing CF definition or whether to define a new standard name that takes the missing aspects into account. For example, MITRAS provides the variable 'water_temperature' for the different water bodies (e.g. sea water, freshwater), which describes the temperature for all water bodies in the model domain. In urban modelling, it can be assumed that freshwater is more common, except if coastal cities are investigated. In the current CF conventions, only standard names for saline or seawater are available. Therefore, the handling of such property needs to be optimised.

Without clearly defined standard names and definitions for such variables, using ORM datasets becomes difficult, as the meaning of variables might differ from model to model. Therefore, ORM variables need to be identified across different models. Standard names with corresponding definitions and units should be determined and then proposed as new standard name candidates to the CF conventions.

[4]https://cfconventions.org/Data/cf-standard-names/current/build/cf-standard-name-table.html

https://doi.org/10.5194/egusphere-2025-5521
Preprint. Discussion started: 22 December 2025

Figure 1. Data post-processing workflow for the FAIR publication of MITRAS model results.

3 Adjustments to the post-processing for MITRAS

Data provided in proprietary data formats, such as model-specific output formats, may not be accessible to external users. It is therefore advisable to convert the data into more commonly used data formats. MITRAS stores simulation results in model-specific binary data files, which ensures a fast writing of the output during the run time while integrating the model (Example in Fig. 1, MITRAS (green), direct output is binary).

The direct binary model output is transferred into netCDF with the model post-processor program M2CDF (Figure 1, step 1; details in Section 3.1). However, not all necessary information and metadata may be included at this point, so further information is needed to describe the data adequately. A second step is introduced, where a newly developed post-processor called NC2ATMODAT is used (Figure 1, step 2). This program adjusts the netCDF model data from step 1 to make it FAIR and to fulfil the requirements of the ATMODAT STANDARD.

NC2ATMODAT adds or corrects information missing in the netCDF datasets from step 1, adds georeferencing to the MITRAS data, allows to select the data to be published domain-wise and variable-wise and adds global metadata that are required by the ATMODAT STANDARD (details in Section 3.2). In step 3 these processed datasets are then checked for the fulfilment of the ATMODAT STANDARD, using the ATMODAT STANDARD CHECKER (Figure 1, details in Section 3.3). They are then ready for being published as a FAIR dataset.

3.1 Step 1: Adaptation of the model post-processor

To create accessible model results, the binary model output data of MITRAS is transferred to netCDF, using the model-specific post-processor M2CDF. The program adds descriptions and attributes to the variables as well as generic global metadata to the netCDF dataset. Based on a review of MITRAS variables (Section 2), it became visible that variable names and corresponding attributes needed to be added in the MITRAS post-processor M2CDF to become CF compliant. The post-processor M2CDF received an update to create CF-compliant netCDF datasets, for which the major part of the update involved the addition of variable attributes with corresponding entries following the CF conventions.

As an example, Fig. 2 shows a comparison of the header from two MITRAS netCDF files for the variable wind speed (ff). In the original version before the update of the post-processor (left) no `long_name` and `standard_name` attributes were assigned. Instead, the variable was described by an attribute `name`. In the update, the attribute `name` was renamed to `long_name`, and for each MITRAS variable a self-describing long name was then either taken from the previous entry in

https://doi.org/10.5194/egusphere-2025-5521
Preprint. Discussion started: 22 December 2025

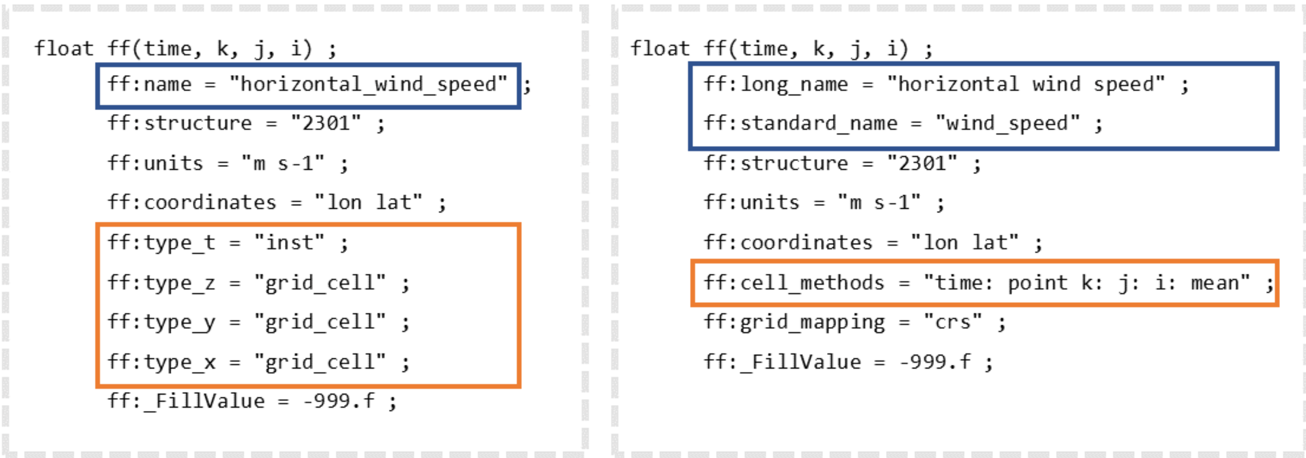

Figure 2. Example for the netCDF header of MITRAS model results. Left: Layout of the header before the adaptations in the post-processor M2CDF. Right: Header with the updated version of M2CDF consistent with CF conventions.

the attribute `name` or a new self-describing long name was defined. The attribute `standard_name` was introduced as a new attribute to M2CDF and, if applicable, for each variable the corresponding standard name from the CF convention was added.

According to the CF convention, the characteristics of a grid cell can be provided with the attribute `cell_methods` (Eaton et al., 2024). The attribute `cell_methods` replaces the former used attributes `type_t`, `type_x`, `type_y`, and `type_z` (Figure 2) that were previously used to provide this information.

Besides the adaptation to the CF conventions, M2CDF received an update regarding the provision of information on obstacles (e.g. buildings) and obstacle-related variables. Obstacles are a key element and important feature of ORMs as they affect the flow within the model domain. As an ORM, MITRAS deals with explicitly resolved obstacles in the model domain. Previously, in MITRAS simulation results, the raw output did not highlight the building locations but showed the variable values continuously (Figure 3a), whereas the location of buildings was written as a separate variable. Note, the values in the buildings are not used in the calculation. M2CDF sets a Fillvalue (-999.f) to the cells matching with the building positions to ensure that exact building locations are visible directly in the variable (Figure 3b). As the building mask is now included in all variables and, if needed, can be indirectly derived, it is no longer necessary to explicitly provide it. This makes it possible to reduce the size of the output. In the case that each output variable should be provided in an individual file, the building information is now guaranteed to always be provided.

Further changes in M2CDF concern the representation of building surface-related variables. Building surface-related variables such as the building wall surface temperature or amount of rain on roofs are given in an index-variable structure as common within the code to safe storage space for the model. However, these arrays are not intuitively understandable for users unfamiliar with MITRAS. MITRAS model equations are solved on an Arakawa-C grid (Arakawa and Lamb, 1977; Schlünzen et al., 2018b). On this grid, scalar quantities like the temperature are defined at grid cell centres, while vector variables, like

https://doi.org/10.5194/egusphere-2025-5521
Preprint. Discussion started: 22 December 2025

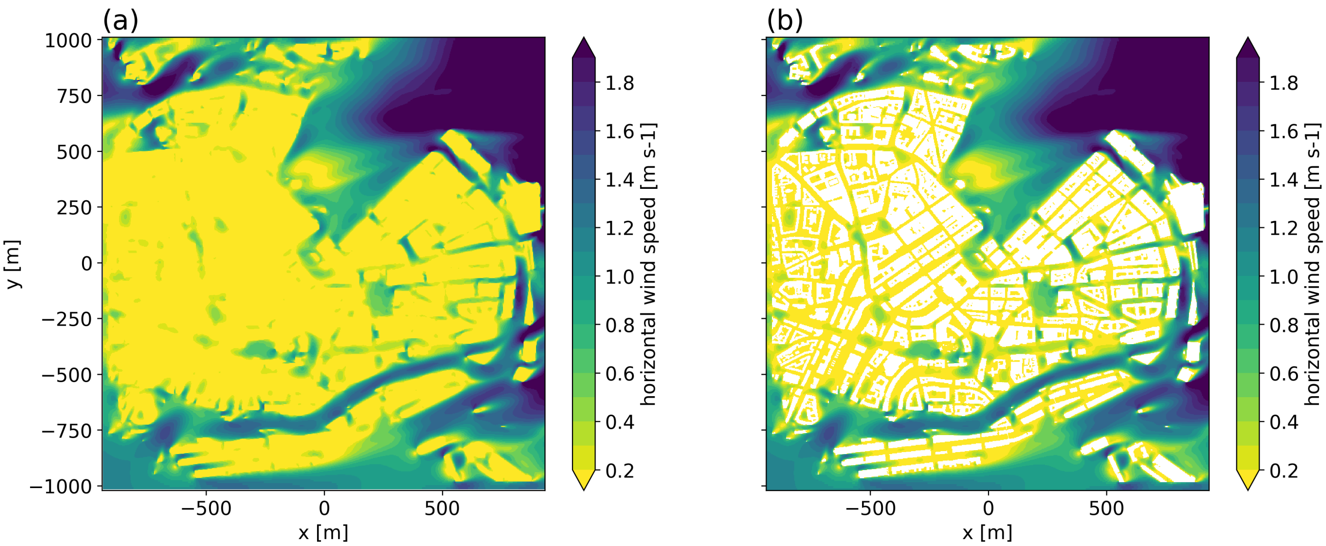

Figure 3. Horizontal wind speed in 2.5 m height above ground in an urban area (here: Depicting the city centre of Hamburg, Germany) without (a) and with marking the buildings with missing values (b).

the wind components, are located at grid cell boundaries. To represent a vector variable at the grid cell centre or a scalar variable at the grid cell boundary, it must be interpolated. Fig. 4 provides an example of the building mask vertical cross-section. Values of atmospheric variables are set to zero inside buildings (grey cells) during the calculations. As MITRAS explicitly resolves obstacles in the model domain, walls are treated as unpermeable obstacles for flow, with wall functions used at the rigid boundaries. Variables assigned to the building surfaces, e.g. wall temperatures, can be assigned or calculated solving the surface energy budget equation. The model used wall temperatures are provided in the output for further use.

The spatial dimension of a building surface-related variable is represented in an array structure, as a $N_{surfcells} \times 3$- matrix with $N_{surfcells}$ denoting the number of obstacle-adjacent grid cells. These arrays contain, based on an auxiliary grid, the grid cell index, at which a cell with an adjacent building wall is located. Further array variables provide the orientation of the building surface (Schlünzen et al., 2018a, b), representing the direction of the obstacle face (1: east/west, 2: north/-south, 3:top/bottom). The structure results from how the building surface-related variable are calculated within the model code (Appendix A) as it improves the computational efficiency. Without further knowledge of these model properties, the grid information is not trivial to derive and therefore not automatically accessible to a data user. To better access those variables, it is advisable to represent building surface-related variables in a Cartesian grid similar to atmospheric variables like temperature or wind speed.

Building surfaces are located at the grid cell boundaries (black dots in Fig. 4). It is preferred to provide variables on the grid cell centres, like done for the wind speed components in M2CDF. However, this approach would give a wrong impression of the positioning of the building boundaries. Furthermore, it is not feasible to interpolate the building surface-related variables from the grid cell centre to the boundary or vice versa. It is preferable to provide the data at the positions of the building boundaries

https://doi.org/10.5194/egusphere-2025-5521
Preprint. Discussion started: 22 December 2025

Table 1. List of variables derived from the obstacle surface temperature.

Variable name	Description
tbuisurf_e(z,y,xv)	Obstacle surface temperature east wall
tbuisurf_w(z,y,xv)	Obstacle surface temperature west wall
tbuisurf_n(z,yv,x)	Obstacle surface temperature north wall
tbuisurf_s(z,yv,x)	Obstacle surface temperature south wall
tbuisurf_t(zv,y,x)	Obstacle surface temperature top wall (ceiling)
tbuisurf_b(zv,y,x)	Obstacle surface temperature bottom wall (floor/roof)
tbuisurf_p(y,x)	Obstacle surface temperature on all roofs (plane view)

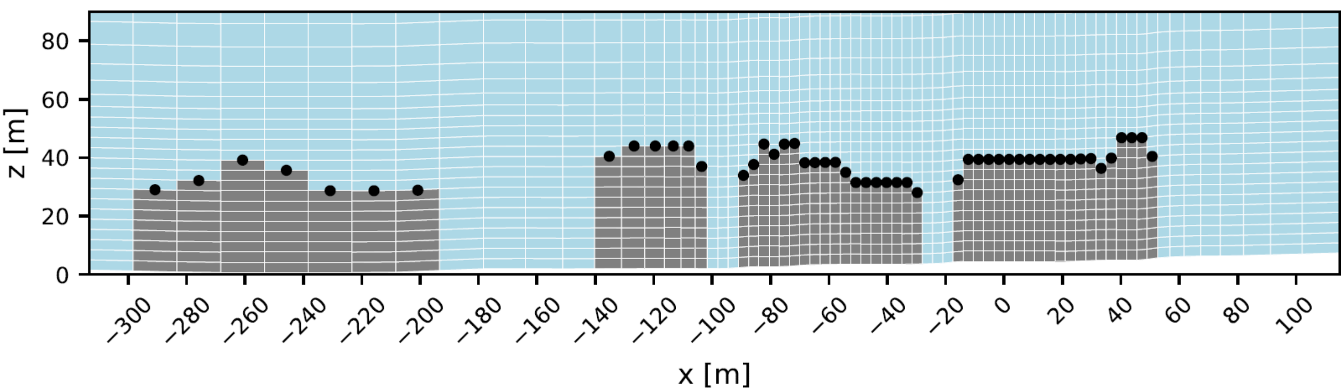

Figure 4. Vertical cross section of building mask at $y = -20\,\mathrm{m}$ (dashed line in Fig. 8) with buildings in grey and atmospheric grid cells in blue. Black dots denote the location of surfaces (roofs).

(walls or roof), as these are located at the actual boundaries of the grid cells. Therefore, the 3D grid positioned at the cell boundaries of the model grid was introduced. This grid uses the already established vector coordinate indices iv, jv and kv. A similar approach for surface-related variables by introducing a secondary coordinate system, that describes the location of the wall properties in the model domain, is utilised by the model PALM-4U (Scherer et al., 2022).

To keep the orientation of each surface-related variable, the variables are allocated to the coordinates of the corresponding direction (east, west, north, south, top and bottom), leading to six building surface-related variables for each direction as listed for the building surface temperature (tbuisurf) in Table 1.

An example of the application of this feature is described in Section 4.3.

https://doi.org/10.5194/egusphere-2025-5521
Preprint. Discussion started: 22 December 2025

Completing this post-processing step makes the netCDF datasets of MITRAS better described, understandable, user-friendly and amenable to open-access programs (e.g. ncview (https://cirrus.ucsd.edu/ncview/), panoply (https://www.giss.nasa.gov/tools/panoply/)). The drawback is a more than 6 times larger output for all surface-related variables.

3.2 Step 2: Development of nc2atmodat to create FAIR netCDF files for ORMs

The post-processor M2CDF transfers the model output to a netCDF file and adds variable names and descriptions consistent with the CF conventions to the netCDF file. Also, it provides very generic global metadata to the datasets. However, the data is not fully adapted to the ATMODAT STANDARD and still miss some additional features relevant to the understanding of the variables.

To fulfil the ATMODAT STANDARD, to modify, correct or add information to the netCDF file, and add the georeferences, the program NC2ATMODAT was developed. NC2ATMODAT provides a quick and convenient way for MITRAS data producers to add additional information, modifications or corrections to the dataset without altering the hardcoded entries in the model core or the post-processor. A re-translation of the model results into netCDF, which can be rather time-consuming, is therefore not needed. Especially, generating properly georeferenced MITRAS model datasets was added and enabled by this program. Adding global attributes required for fulfilling the ATMODAT STANDARD is easier compared to the provision with the existing structures in M2CDF. Last but not least, program NC2ATMODAT allows to select the data to be published, since in general not all data are meant to be published, but only some variables or only a part of the model domain.

Furthermore, since the program modifies netCDF files, it might be usable by other ORM modellers with a few adaptations towards their model properties. The code of NC2ATMODAT is written in Python (https://www.python.org/) and uses the modules Xarray (Hoyer and Hamman, 2017), Pandas (pandas development team, 2020), Numpy Harris et al. (2020) and Pyproj (Whitaker et al., 2019). The program is open-source and available via GitHub (Voss, 2025). Figure 5 shows the workflow of NC2ATMODAT, whereas the main functions of NC2ATMODAT are described in the following section.

3.2.1 User Input

NC2ATMODAT includes a User Input section, where the user provides the path and name to the netCDF file and the AtMoDat metadata table. This metadata table is provided with the program and contains the global attributes entries for the final resulting datasets. The information within the metadata table has to be filled in by the program user. The User Input section provides three control options that can be activated: spatially and temporal cropping the dataset, defining a custom "georeference point" and selecting output variables.

https://doi.org/10.5194/egusphere-2025-5521
Preprint. Discussion started: 22 December 2025

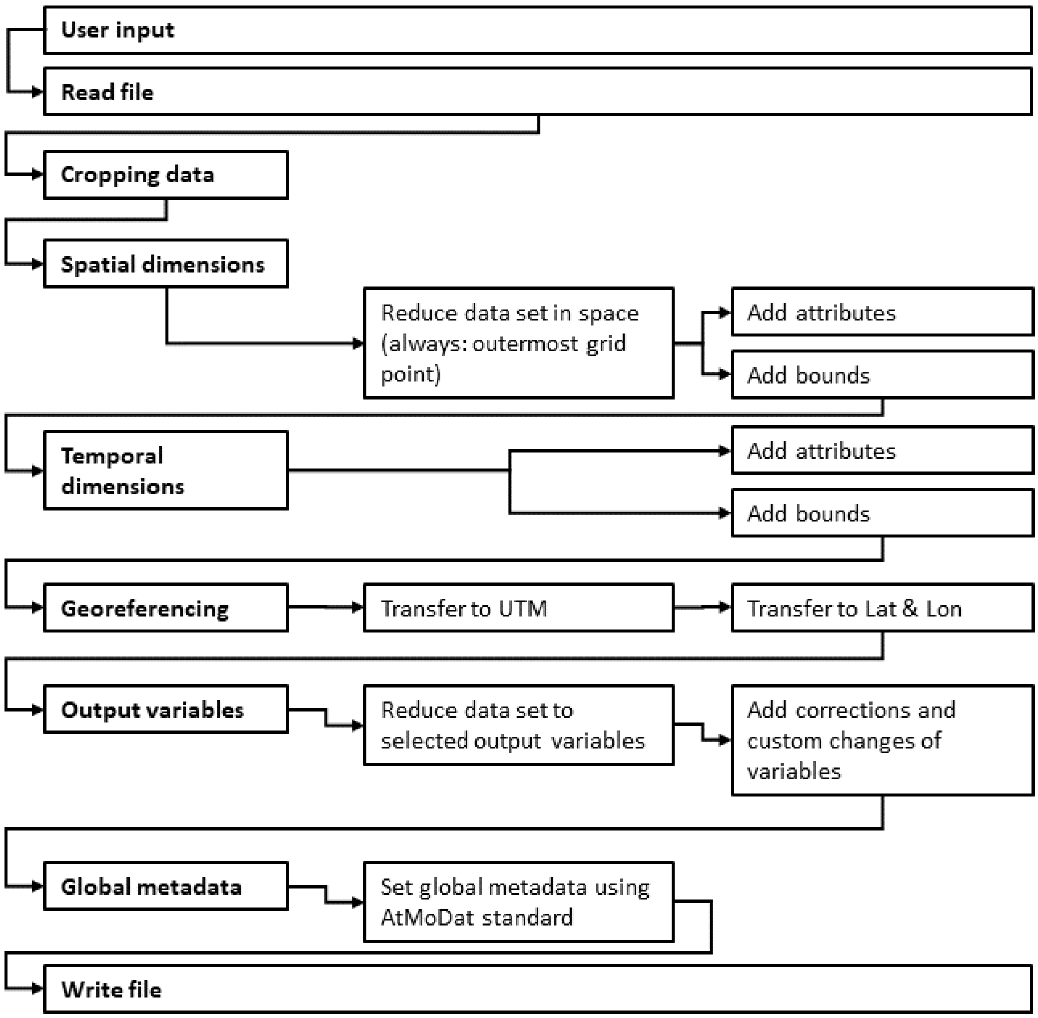

Figure 5. Overview of the workflow of the program nc2atmodat.

https://doi.org/10.5194/egusphere-2025-5521
Preprint. Discussion started: 22 December 2025

3.2.2 Cropping data

NC2ATMODAT crops the datasets in the temporal and spatial dimensions if requested by the user (section 3.2.3). Otherwise, in the spatial dimension, the outermost grid cells are removed, which is required to assign the correct grid cell bounds to each grid cell in the MITRAS case.

3.2.3 Spatial and temporal dimensions

Spatial coordinates with corresponding attributes and information are assigned. The current version of NC2ATMODAT requires that all variables are located on a scalar grid, described by indices i and j in horizontal and k in vertical direction. An exception are building surface-related variables defined at grid cell boundaries (Section 3.1), which use indices iv, jv, kv, that are initially used to describe vector coordinates.

MITRAS uses a Cartesian coordinate system in the horizontal direction (Schlünzen et al., 2018b). The use of geographic coordinates is not common for ORMs, since with grid sizes of 1 to 10 m, the grid sizes in geographical coordinates will differ by only $0.03''$ to $0.3''$. While the addition of a georeferenced grid is not straightforward, it is essential for using the data in e.g. graphical systems used in meteorology (e.g. GIS applications). For further processing reasons, NC2ATMODAT introduces the dimensions x, y, z (xv, yv, zv; for surface properties), given in metres. If applicable, the indices are replaced by the newly introduced coordinates x, y, z, xv, yv and zv in the final AtMoDat conformal netCDF model results, otherwise, they remain as indices. The temporal dimension coordinate (MITRAS variable name: "time") is set and assigned with variable attributes.

For spatial and temporal coordinate variables, the attribute `bounds` was introduced. Bounds are used to indicate the extent of a grid cell and provide the location of the cell boundary (Eaton et al., 2024). After cropping, the spatial and temporal variables are then assigned attributes and new variable names.

3.2.4 Georeferencing

This is a feature that is mainly implemented for the modification of MITRAS model datasets but adjustable for other ORMs that use similar projections. MITRAS provides the coordinates as distances of each grid point to the origin of the model domain. The origin is provided as latitude / longitude information. To be consistent with the CF conventions and to publish georeferenced datasets, geographical coordinates have to be provided with the dataset. NC2ATMODAT calculates and adds geographic coordinates to the dataset using the projection used for the creating the grid in Cartesian coordinates. For MITRAS simulations, the model-specific pre-processor generates the model domain using Geographic coordinates and metre coordinates (specifially: Gauss-Kruger coordinates). Geographical coordinates, provided in latitude and longitude are required to calculate the solar radiation, whereas metre coordinates are used to select and construct the model domain from a set of raster-grid files that contain information on building heights, land surface cover classes and terrain heights. Currently, only the geographic coordinates at the origin, further named as "reference point coordinates", are provided with the model domain input and output files and used in the model simulations. In both files, metre coordinates are available for each grid point.

With the Python module Pyproj (Whitaker et al., 2019) and the reference point coordinates, geographic coordinates can be

https://doi.org/10.5194/egusphere-2025-5521
Preprint. Discussion started: 22 December 2025

calculated for each grid point and are assigned to the dataset. Thus, netCDF files processed with NC2ATMODAT provide two sets of coordinates, Universal Transverse Mercator (UTM) coordinates (x_utm, y_utm) and geographical coordinates (named as lat and lon). The reference point coordinates can be either taken from the model results or provided via the User Input. With these adaptations, the dataset is then properly georeferenced.

3.2.5 Output variables

Most likely not all variables in the model output are analysed in depth and relevant for the publication. The function 'output variables' of the program NC2ATMODAT allows to reduce the dataset to those variables that were picked for publication. These output variables can be either pre-defined via an additional file ("dict_OUTPUT_variables"), provided with the program or submitted via the User Input.

ORM post-processors M2CDF sufficiently describe the majority of variables by providing hardcoded variable attributes. The second post-processor NC2ATMODAT provides the possibility to modify or add attributes, if necessary. To give an example, MITRAS might be used to calculate the transport and concentration of air pollutants as passive tracers. However, the tracer variable in the model is not specified as a certain air pollutant in the output data, thus M2CDF adds a generic long name and no standard name to this variable. The CF convention provides a large list of standard names for different air pollutants, that could be applied to the variable. To retain the flexible usage of the variable in the models but to be able to add standard names to this variable, fitting long and standard names can then be added via NC2ATMODAT. Furthermore, any additional attribute can be added using NC2ATMODAT. For example, if no standard name was available, but additional information or further explanations were helpful to clarify the meaning or purpose of a variable, a `comment` attribute can be added, making the variable sufficiently self-explanatory. Such additional attributes and corrections can be stored in the input file ("dict_custom_variable_attributes").

3.2.6 Global metadata

The ATMODAT STANDARD requires that the dataset has to be sufficiently described and contain as much information as needed to be FAIR. Therefore, the dataset shall contain the global metadata attributes summarised and explained in the ATMODAT STANDARD (Ganske et al., 2021). The global metadata attributes are categorised as mandatory, recommended and optional. To fulfil the requirements for the ATMODAT STANDARD, the data producer shall fill attributes and describe their datasets with as much detail as possible. As sometimes it might not be possible to add all attributes for each dataset, at least the mandatory attributes must be provided. To guide and help the data producer to provide global metadata to their dataset to fulfil the ATMODAT STANDARD, an Excel file ("Metadata_for_atmodat_standard.xlsx") is provided with the program NC2ATMODAT. With the NC2ATMODAT the global metadata can then be added to the dataset. Explanations, examples of how to fill out the form and further information for each attribute are given within the document.

3.2.7 Write output file

Lastly, the data are saved as a new netCDF file. This is then ready for checking if it fulfils the ATMODAT STANDARD.

https://doi.org/10.5194/egusphere-2025-5521
Preprint. Discussion started: 22 December 2025

3.3 Step 3: Application of ATMODAT Standard Checker

With the publication of the ATMODAT STANDARD (Ganske et al., 2021), the ATMODAT STANDARD CHECKER (Kretzschmar et al., 2022; Ganske et al., 2022) was released. It is a tool for checking if a dataset fulfils the requirements and its compliance with the ATMODAT STANDARD. If required, the ATMODAT STANDARD CHECKER also provides the possibility to add or correct metadata information to the dataset header. The output of the ATMODAT STANDARD CHECKER is machine-readable, which makes it possible to evaluate multiple files at a time automatically. Instructions for installing and using the checker are available on GitHub (Kretzschmar et al., 2022). The checker helps any person interested in publishing model results to determine missing information (Section 4.2 and Fig. 7).

4 Application of suggested standardisation to ORM datasets

In this Section, the post-processing described in Section 3 is applied to two datasets from the ORM MITRAS: one dataset includes dispersion (Voss, 2023), hereafter named V23, and the second dataset considers precipitation (Ferner et al., 2024) further denoted as F24. Both model outputs were converted to netCDF by using M2CDF (Section 3.1), further post-processed using NC2ATMODAT(Section 3.2) and are already successfully published in the World Data Center for Climate (WDCC). Both datasets are based on the same domain (Section 4.1). Hereby, in Section 4.2, the preparation of a georeferenced MITRAS dataset is shown with V23. In Section 4.3 the standardisation of model data with ORM-specific obstacle variables is shown using F24.

4.1 Model domain

Both datasets are generated by the ORM MITRAS (Schlünzen et al., 2003; Salim et al., 2018), which is part of the M-SYS model system (Trukenmüller et al., 2004; Schatzmann et al., 2006). The building data for both datasets are taken from Salim et al. (2015). They are based on the Digital Terrain Model, the data of the German geoinformation system ATKIS (official topographic-cartographic information system) (Müller and Seyfert, 2000), and on the 3D-urban model data (LoD 2) for building details. Salim et al. (2015) used a model domain representing the City Centre of Hamburg, Germany. The status of the City is from 2014, with the Hamburg City Hall (Rathaus) in the centre of the domain.

4.2 Dataset with dispersion

The dataset V23 depicts the domain around Hamburg City Hall as described in Section 4.1 without orography. It contains one simulation with constant temperature and humidity and with stable atmospheric conditions. The domain extents in the horizontal direction to $1,9 \times 2,0\,\mathrm{km}^2$ and reaches the top of the model domain at a height of $5.4\,\mathrm{km}$. It has a horizontal resolution of $2.5\,\mathrm{m}$ over the whole model domain. The vertical resolution begins with $5\,\mathrm{m}$ within the lowest model levels (positioned at $2.5\,\mathrm{m}$ above ground) and increases above the domain height of $152.5\,\mathrm{m}$ to $200\,\mathrm{m}$. The dataset was mainly created to test and

https://doi.org/10.5194/egusphere-2025-5521
Preprint. Discussion started: 22 December 2025

apply the ATMODAT STANDARD to an ORM dataset; therefore, no further complexity was considered in the simulation.
This model simulation includes the emission and transport of passive tracers from areas with a specific surface cover class (small green spaces within the city). These tracers were assumed to be air pollutants, with no chemical reactions, deposition or sedimentation processes occurring. V23 provides a variable named “conc”, for which no defined long or standard name is assigned in M2CDF, as the variable can be denoted to different air pollutants of the same qualities. In this particular dataset, the variable “conc” should be described as particulate matter with particle sizes of 10 μm (PM10). For this variable the CF standard name “mass_concentration_of_pm10_ambient_aerosol_particles_in_air” is selected and the corresponding long name is defined as “concentration of pm10”. These attributes and the corresponding unit are then added to the variable with the program NC2ATMODAT.
Proper geographical coordinates had to be added to the dataset. Geographical coordinates can be calculated from the existing grid coordinates, as they are provided as distances from the model origin in metres (Section 3.2). In this dataset, the model origin of the coordinate system is located in the centre of the model domain at the Hamburg Rathaus (Figure 6). Therefore, the geographical coordinates for the model origin ($x = 0, y = 0$), the so-called reference point, are required. The reference point coordinates are mainly used during the model domain initialisation and were initially not provided with the model output. M2CDF includes the reference point coordinates in the model output. NC2ATMODAT then calculates the geographical coordinates from the reference point and adds them to the dataset. With these features added to both post-processing routines, the dataset can now be assigned geographical coordinates and is correctly georeferenced. This is visible in Fig. 6, where the building positions in the model are given in latitude/longitude coordinates and align with the building positions on the underlying map.

After applying M2CDF and NC2ATMODAT on the model dataset, the ATMODAT STANDARD CHECKER (Section 3.3) was used. Figure 7 shows an excerpt of the ATMODAT STANDARD CHECKER output of the same dataset before and after the post-processing with NC2ATMODAT. The left-hand side shows the checker results of the dataset after the use of M2CDF but without the use of NC2ATMODAT. The ATMODAT STANDARD CHECKER raised several warnings, as for example, the variable “conc01” was previously missing the `standard name` or a `long name` attribute. The right-hand side of Fig. 7 shows the checker output for the dataset, after processing with the updated version of M2CDF and with NC2ATMODAT. After V23 passed the ATMODAT STANDARD CHECKER, the dataset is thereby considered as fulfilling the ATMODAT STANDARD and thereby standardised.

4.3 Dataset with precipitation

F24 (Ferner et al., 2024) was created for studies on the heterogeneity of rain in an urban neighbourhood (Ferner et al., 2023). It contains precipitation quantities defined on building surfaces providing a suitable test case for the representation of variables specific to ORMs. The example given here can be applied to other building surface-related variables or results created by other ORMs.

https://doi.org/10.5194/egusphere-2025-5521
Preprint. Discussion started: 22 December 2025

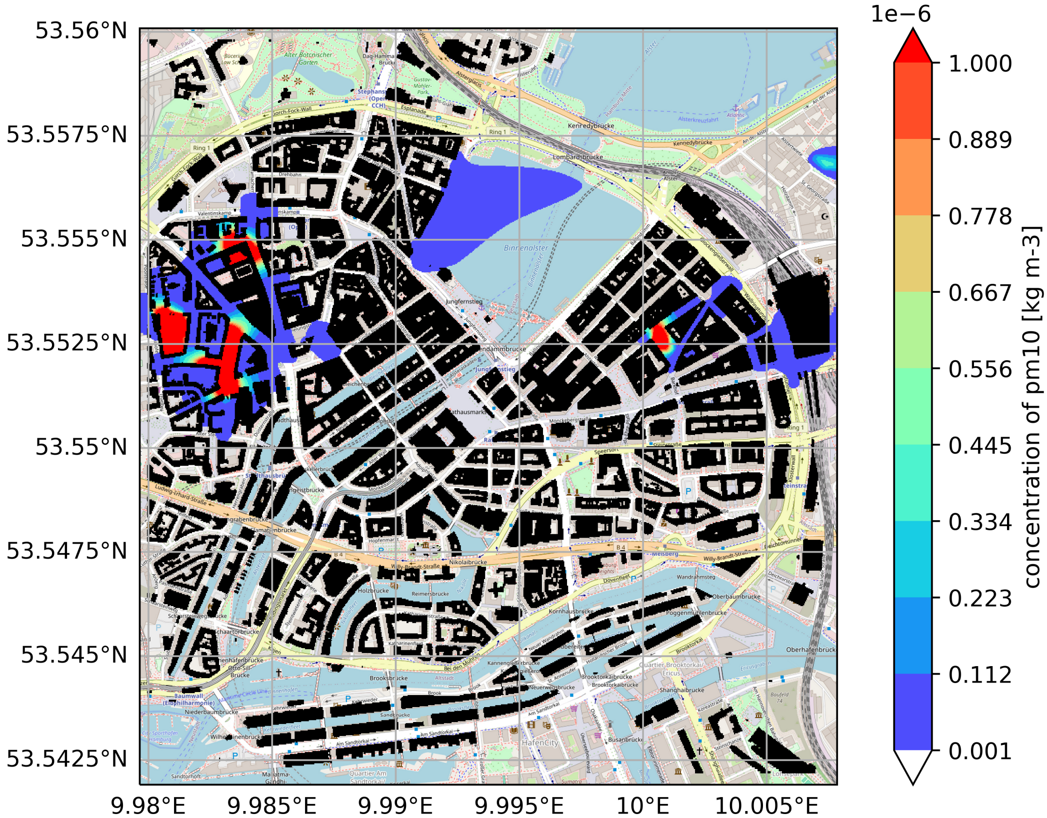

Figure 6. Model domain V23 of the city centre of Hamburg, Germany, with buildings in black, use of horizontally uniform grid. Coloured areas denote concentration of PM10. The underlying map was taken from © OpenStreetMap (openstreetmap.org). Plot was generated with Cartopy (Hrisko, 2020; Met Office, 2010 - 2015).

A stretched grid is used with a highly resolved area with grid sizes of 3.5 m near Hamburg City Hall (white box in Fig. 8). The grid sizes increase by a factor of 1.175 resulting in resolutions of 15 m horizontally and 150 m vertically at the domain boundaries. The domain extends to $1.6 \times 1.8\,\mathrm{km}^2$ around Hamburg City Hall reaching up to 5 km. A decrease in grid resolution was necessary to perform computationally expensive simulations including cloud processes within a reasonable time. Due to
355 the irregular grid, the obstacles become coarser towards the edges of the domain. This can be seen by comparing the model domain of F24 in Fig. 8 with the domain of V23 (Figure 6). The building block west of the lake Alster (9.99°E and 53.5565°N in Fig. 6 and x=-400 m and y=700 m in Fig. 8) is much finer resolved in the domain of V23. The stretched grid is also visible in the vertical cross-section of the building mask (Figure 4). In contrast to V23, in this model domain orography is considered (Figure 8).

360 As mentioned in section 3.1, the structure of building surface-related variables as provided initially in the MITRAS output was changed. Each surface-related variable is now written as six 4D variables for each wall with corresponding coordinates

https://doi.org/10.5194/egusphere-2025-5521
Preprint. Discussion started: 22 December 2025


```
CHECKING NetCDF FILE:
=====================
WARN: Cannot determine CF version from the Conventions
attribute; checking against latest CF version: CF-1.8
Using CF Checker Version 4.1.0
Checking against CF Version CF-1.8
Using Standard Name Table Version 82 (2023-07-06T13:17:07Z)
Using Area Type Table Version 11 (06 July 2023)
Using Standardized Region Name Table Version 4 (18 December
2018)

WARN: (2.6.1): No 'Conventions' attribute present


------------------
Checking variable: conc01
------------------
WARN: (3): No standard_name or long_name attribute specified
INFO: (3.1): No units attribute set.  Please consider adding
units attribute for completeness.

------------------
Checking variable: tket
------------------
WARN: (3): No standard_name or long_name attribute specified
ERROR: (3.1): Invalid units: ?

------------------
Checking variable: time
------------------
WARN: (3): No standard_name or long_name attribute specified
WARN: (4.4.1): Use of the calendar and/or month_lengths
attributes is recommended for time coordinate variables
```

```
=== Short summary ===

ATMODAT Standard Compliance Checker Version: 1.3.2
Checking against: ATMODAT Standard 3.0, CF Version 1.8
Checked at: 2023-05-24T10:29:55

Number of checked netCDF files: 1

Mandatory ATMODAT Standard checks passed: 0/4 (4 missing, 0
error(s))
Recommended ATMODAT Standard checks passed: 1/20 (19
missing, 0 error(s))
Optional ATMODAT Standard checks passed: 0/9 (9 missing, 0
error(s))

CF checker errors: 2
CF checker warnings: 79
```

```
CHECKING NetCDF FILE:
=====================


Using CF Checker Version 4.1.0
Checking against CF Version CF-1.8
Using Standard Name Table Version 82 (2023-07-06T13:17:07Z)
Using Area Type Table Version 11 (06 July 2023)
Using Standardized Region Name Table Version 4 (18 December
2018)




------------------
Checking variable: conc01
------------------




------------------
Checking variable: tket
------------------



------------------
Checking variable: time
------------------
```

```
=== Short summary ===

ATMODAT Standard Compliance Checker Version: 1.3.2
Checking against: ATMODAT Standard 3.0, CF Version 1.8
Checked at: 2023-05-24T10:29:55

Number of checked netCDF files: 1

Mandatory ATMODAT Standard checks passed: 4/4 (0 missing, 0
error(s))
Recommended ATMODAT Standard checks passed: 16/20 (4
missing, 0 error(s))
Optional ATMODAT Standard checks passed: 7/9 (2 missing, 0
error(s))

CF checker errors: 0
CF checker warnings: 0
```

Figure 7. Example output of the ATMODAT STANDARD CHECKER for MITRAS model data V23 before (left) and after (right) the application of NC2ATMODAT.

https://doi.org/10.5194/egusphere-2025-5521
Preprint. Discussion started: 22 December 2025

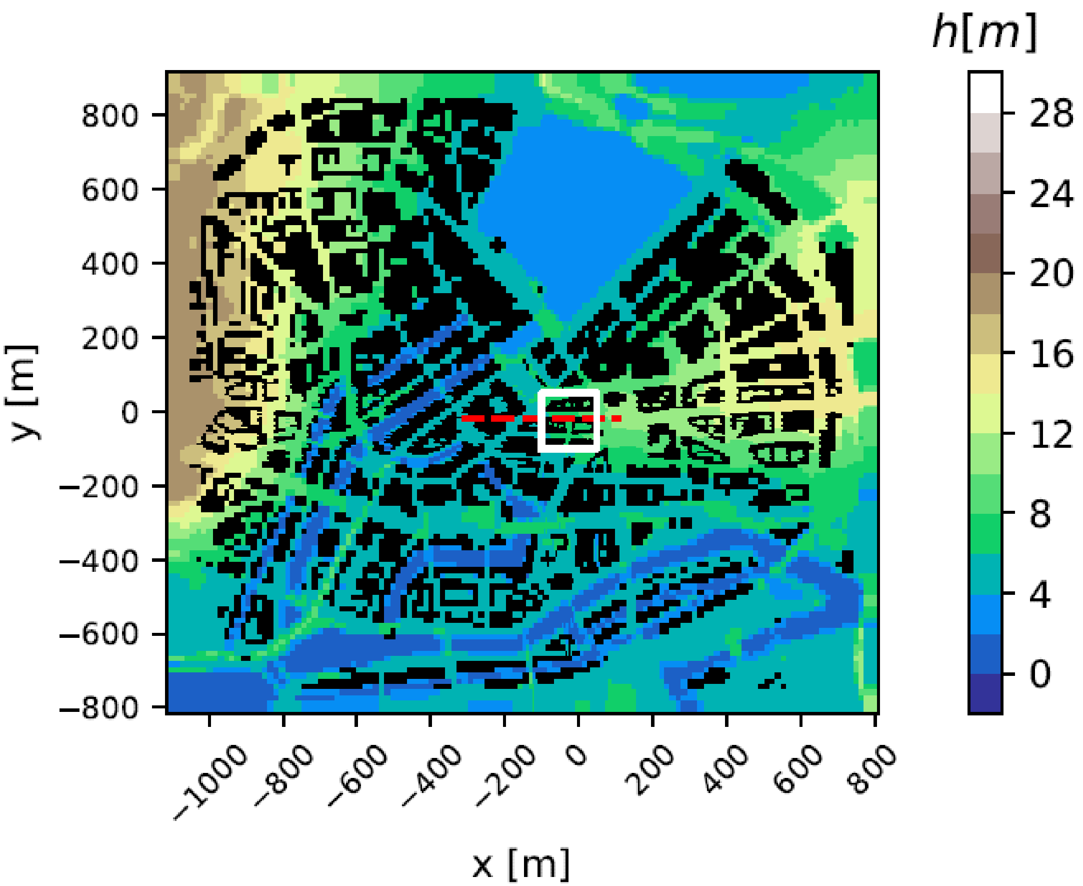

Figure 8. The model domain of Hamburg city centre using a horizontally non-uniform grid with buildings in black and the orography height h in m. The white box denotes the area with equally sized grid cells (equidistant area), and the red dashed line denotes the location of the cross-section in Fig. 4.

and one 3D variable holding a plane-view of all building surface roof cells. For tbuisurf on east walls (tbuisurf_e), for instance, scalar vertical coordinates (z) and y-coordinates (y) are used, but for the x-direction, the location of the grid cell boundaries (xv) is used. Figure 9 shows results for precipitation at roofs (left) and the different roof heights (right). While the precipitation field has a general increase towards south-east in the domain, it is also visible that high buildings might receive higher precipitation values than their surroundings, as visible at the highest building ($x \approx -800\,m$, $y \approx 750\,m$).

For the temperature on bottom surfaces, i.e. roofs (tbuisurf_t), the vertical coordinate equals the height of the grid cell boundaries, while the horizontal coordinates still represent the grid cell centres. With the variable tbuisurf_p a plane view of all values on roofs is provided, which is shown as an example for the rain amount on roofs in Fig. 9a. The corresponding roof heights can be seen in Fig. 9b.

Figure 10 shows a visualisation of the building surface-related variables and the location at the building surface. The building surface-related variables are located at the vector grid that overlaps with the scalar grid, indicated by the transparent colours

https://doi.org/10.5194/egusphere-2025-5521
Preprint. Discussion started: 22 December 2025

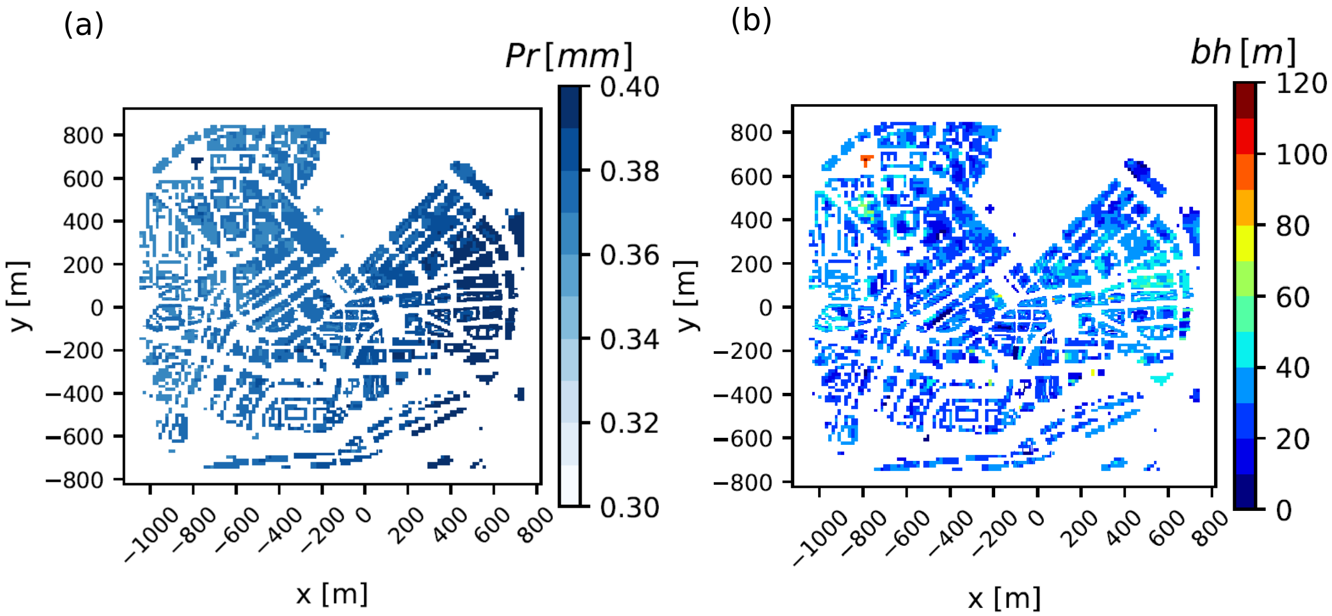

Figure 9. Amount of rainwater Pr on all roofs (a) and all roof heights bh (b).

overlapping with the building mask. Figure 11 shows the netCDF header for these building surface-related variables.

5 Conclusions

To foster the publication of FAIR model data created by micro-scale obstacle-resolving atmospheric models (ORM), ORM datasets should be well-described, self-explanatory and easy to understand by model data users. This can be achieved by the application of data standards to model datasets. Existing data standards, such as the CF conventions, are widely used for numerical model data, but are mainly designed for climate model data and are therefore not fully applicable to ORMs, as they miss features that are common in ORMs. Therefore, the standardisation of ORM data is not trivial as general discipline-specific standards are not established. To overcome the lack of standardised datasets in this discipline, a standardisation routine for model data was developed using results of the ORM MITRAS as an example. The aim was to standardise ORM MITRAS results so that they comply with the ATMODAT STANDARD, since the the ATMODAT STANDARD (Ganske et al., 2021) is a generic data standard for atmospheric model datasets.

The ATMODAT STANDARD requires the use of the netCDF data format together with the CF convention and requires a set of global metadata that needs to be provided with the dataset. To fulfil the requirements of the ATMODAT STANDARD and the CF conventions, the post-processing of MITRAS received several adaptations and extensions. The description of variables and ORM features provided by the post-processor M2CDF was improved to be better understandable for model data users not

https://doi.org/10.5194/egusphere-2025-5521
Preprint. Discussion started: 22 December 2025

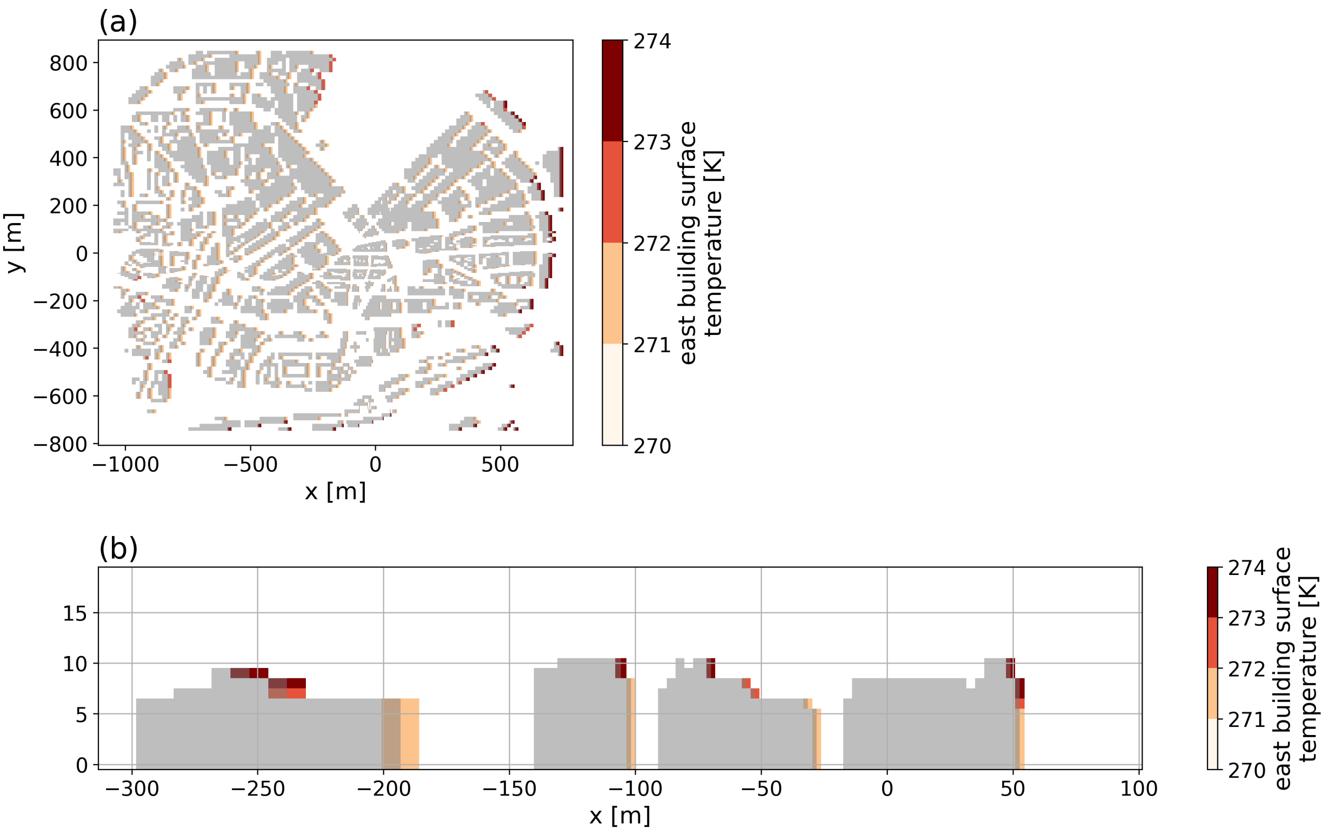

Figure 10. Example for the 3D representation of building surface-related variables representing the eastern building wall surface temperature as (a) horizontal cross-section view of the whole domain at a height of 10 m above ground and (b) vertical cross-section view located at the position denoted in Fig. 8. Grey areas denote the building positions.

familiar with the model. This includes the extension and addition of descriptive attributes and information to each MITRAS variable with their corresponding entries and, if applicable, standard names. Furthermore, the post-processing was extended to include a more accessible representation of building surface-related variables. Previously, building surface-related variables, e.g. wall temperatures, radiation fluxes or rain water amounts on building walls, were stored in a vector array structure. To view these variables, a data user would require further knowledge about the in-model representation of the variables. The new storage in 3D (time dependent 4D) array structures is more intuitive for data users and improves the provision of building surface-related variables. It comes, however, with the drawback of needing much more storage. This will become even more relevant in future work, when further building surface-related properties will be implemented to study e.g. effects of adaptation or heat mitigation measures. This includes, for example, different facade structures, representation of bridges or balconies or green facades.

https://doi.org/10.5194/egusphere-2025-5521
Preprint. Discussion started: 22 December 2025


```
float tbuisurf_e(time, k, y, xv) ;
    tbuisurf_e::_FillValue = -999.f ;
    tbuisurf_e:long_name = "east building surface temperature" ;
    tbuisurf_e:structure = "5901 " ;
    tbuisurf_e:units = "K" ;
    tbuisurf_e:cell_methods = "time: point k: y: xv: mean" ;
    tbuisurf_e:comments = "Surface temperature at eastern building wall" ;
    tbuisurf_e:coordinates = "lonu latu" ;
float tbuisurf_n(time, k, yv, x)  ;
    tbuisurf_n::_FillValue = -999.f ;
    tbuisurf_n:long_name = "north building surface temperature" ;
    tbuisurf_n:structure = "5901 " ;
    tbuisurf_n:units = "K" ;
    tbuisurf_n:cell_methods = "time: point k: yv: x: mean" ;
    tbuisurf_n:comments = "Surface temperature at northern building wall" ;
    tbuisurf_n:coordinates = "lonv latv" ;
float tbuisurf_t(time, kv, y, x)  ;
    tbuisurf_t::_FillValue = -999.f ;
    tbuisurf_t:long_name = "top building surface temperature" ;
    tbuisurf_t:structure = "5901 " ;
    tbuisurf_t:units = "K" ;
    tbuisurf_t:cell_methods = "time: point kv: y: x: mean" ;
    tbuisurf_t:comments = "Surface temperature at building roof" ;
    tbuisurf_t:coordinates = "lon lat" ;
```

Figure 11. Header of F24 dataset with focus on the description for the newly added building surface-related variables.

Additionally, a new post-processing program called NC2ATMODAT was introduced. This routine summarises the required steps for the preparation of MITRAS netCDF datasets. It provides the option of small changes, corrections or additions to variables and metadata, without the need to adapt the models specific post-processing routines. If required, it allows the cropping of the dataset in space and time and a variable-based selection of the data. This new post-processing has been successfully applied to two ORM MITRAS datasets, which were subsequently released as FAIR datasets in the World Data Center for Climate (WDCC). The program NC2ATMODAT is written in Python and available as open source (Voss, 2025).

First tests on the application of NC2ATMODAT to model data beside ORMs were carried out on model data from the mesoscale model METRAS (Schluenzen, 1990) and the chemistry model CITYCHEM (Karl, 2024).

The restructuring, organisation and assignment of MITRAS variables revealed that many variables, which are typically occurring in ORMs are not covered by existing standard names. For a better identification and understanding across different

https://doi.org/10.5194/egusphere-2025-5521
Preprint. Discussion started: 22 December 2025

ORMs and to avoids confusion or misinterpretation in data handling, they need to become standardised This would need a community approach to collect variables between different models, the definition and discussion of new standard names for typical ORM variables and the proposal of new ORM variables to the CF conventions. This would foster and increase the amount of well-described ORM datasets and improve their usability.

Code and data availability. The datasets V23 (Voss, 2023) (https://doi.org/10.26050/WDCC/MITRAS_ATMODAT) and F24 (Ferner et al., 2024) (https://doi.org/10.26050/WDCC/MitrasWinterHamburg) are published via the World Data Center for Climate (WDCC). The M2CDF source code is distributed upon request under the terms of a user agreement with the Mesoscale and Microscale Modeling (MeMi) working group at the Meteorological Institute, University of Hamburg (metras@uni-hamburg.de, https//www.mi.uni-hamburg.de/memi). The program NC2ATMODAT is published and available via GitHub (https://doi.org/10.5281/zenodo.17035812). The ATMODAT STANDARD CHECKER is available via GitHub (https://doi.org/10.5281/zenodo.6701508) as well.

Appendix A: Representation of a building surface variable in the model code

In MITRAS, obstacles are explicitly resolved using the mask method (Briscolini and Santangelo, 1989; Salim et al., 2018). There, a grid cell is either an atmospheric grid cell surrounded by atmospheric grid cells, an obstacle grid cell, or an atmospheric grid cell adjacent to an obstacle grid cell. The number of obstacle adjacent grid cells is given with `nsurfcells` ($N_{surfcells}$).

An example code for the calculation of an obstacle surface variable (`var_building`) is provided in Fig. A1. The first do loop runs through every obstacle adjacent grid cell. The location in the Cartesian grid is provided by `nxobst`, `nyobst`, and `nzobst`. In order to address an obstacle surface variable, it is not necessary to go through every grid cell and check whether it is obstacle adjacent. Instead, the variable is defined for every obstacle adjacent grid cell, which reduces the number of iterations in loops and increases the efficiency of the algorithm.

The number of obstacle faces at an obstacle adjacent cell is given with `nsurfcount`. Up to three obstacles faces can be represented at each obstacle adjacent grid cell. Again, the efficiency of the algorithm is increased by only including `nsurfcount` iterations instead of three regardless of surface count (second do loop in Fig. A1).

The information whether a wall is facing east, west, north, etc. is contained in `nsurftype` and `nsurfdir` as shown in Table A1. Assuming for example an obstacle adjacent grid cell with a wall north and a wall below, `nsurfcount` would be 2, `nsurftype` and `nsurfdir` 2 and 1 for the first wall and 3 and -1 for the second wall. Note, that even though `nsurftype`, `nsurfdir`, and `var_building` have the same spatial dimensions (`nsurfcells`×3), the last dimension has a different meaning in the obstacle surface variable.

In order to calculate $var\_building$, all necessary information on the type of wall and the location of the grid cell are provided. Unfortunately, there is still an if statement in the do loops, which slows down the algorithm. This statement is necessary, as different calculations are performed depending on the type of wall (`foo1`, `foo2`, `foo3` in Fig. A1). However, an additional if statement to differentiate between wall directions is avoided by utilising the fact, that `nsurfdir` is either one or minus one.

https://doi.org/10.5194/egusphere-2025-5521
Preprint. Discussion started: 22 December 2025


```
DO jb=1,nsurfcells
   ji                 = nxobst(jb)
   jj                 = nyobst(jb)
   jk                 = nzobst(jb)
   DO jm=1,nsurfcount(jb)
      jr              = nsurftype(jb,jm)
      jd              = nsurfdir(jb,jm)
      IF (jr .EQ. 1_ni) THEN ! east(jd=1) / west(jd=-1)
         var_building(jb,jr) =   (1 + jd) / 2 * foo1(jk  ,jj  ,ji-1) &
                   &           + (1 - jd) / 2 * foo1(jk  ,jj  ,ji  )
      ELSEIF (jr .EQ. 2_ni) THEN ! north(jd=1) / south(jd=-1)
         var_building(jb,jr) =   (1 + jd) / 2 * foo2(jk  ,jj-1,ji  ) &
                   &           + (1 - jd) / 2 * foo2(jk  ,jj  ,ji  )
      ELSEIF (jr .EQ. 3_ni) THEN ! top(jd=1) / bottom(jd=-1)
         var_building(jb,jr) =   (1 + jd) / 2 * foo3(jk-1,jj  ,ji  ) &
                   &           + (1 - jd) / 2 * foo3(jk  ,jj  ,ji  )
```

Figure A1. Example Fortran95 code for the calculation a building surface variable var_building.

https://doi.org/10.5194/egusphere-2025-5521
Preprint. Discussion started: 22 December 2025

Table A1. Representation of obstacle face directions in MITRAS.

wall	`nsurftype`	`nsurfdir`
east	1	1
west	1	-1
north	2	1
south	2	-1
top	3	1
bottom	3	-1

Author contributions. K. Heinke Schlünzen has contributed to the conceptualising. Karolin Samsel contributed to this paper by running simulations for precipitation in Hamburg, providing the datasets for application of the described methods and wrote about the building mask treatment in MITRAS (Section 3.1) and the datasets in Section 4.3. All authors contributed to the discussion of the results.

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

Acknowledgements. This work contributes to and uses results of project "AtMoDat", funded by the Federal Ministry of Education and Research under the funding number 16QK02C. The creation of the dataset including precipitation was financed within the framework of the Helmholtz Institute for Climate Service Science (HICSS), a cooperation between Climate Service Center Germany (GERICS) and Universität Hamburg, Germany, and the corresponding work conducted as part of the WINTER project (Investigating climate change related impacts on the urban winter climate of Hamburg). This work contributes to and is partly funded by the Cluster of Excellence 'CLICCS - Climate, Climatic Change, and Society', and is a contribution to the Center for Earth System Research and Sustainability (CEN) of Universität Hamburg. CLICCS is funded by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) under Germany's Excellence Strategy – EXC 2037 'CLICCS - Climate, Climatic Change, and Society' – Project Number: 390683824, contribution to the Center for Earth System Research and Sustainability (CEN) of Universität Hamburg. Responsibility for the content of this publication lies with the authors.

https://doi.org/10.5194/egusphere-2025-5521
Preprint. Discussion started: 22 December 2025

References

Arakawa, A. and Lamb, V. R.: Computational Design of the Basic Dynamical Processes of the UCLA General Circulation Model, Method. Comput. Phys., 17, 173–265, https://doi.org/10.1016/b978-0-12-460817-7.50009-4, 1977.

Baklanov, A., Grimmond, C. S., Carlson, D., Terblanche, D., Tang, X., Bouchet, V., Lee, B., Langendijk, G., Kolli, R. K., and Hovsepyan, A.: From urban meteorology, climate and environment research to integrated city services, Urban Climate, 23, 330–341, https://doi.org/10.1016/j.uclim.2017.05.004, 2018.

Bermudez, L.: New frontiers on open standards for geo-spatial science, Geo-Spatial Information Science, 20, 126–133, https://doi.org/10.1080/10095020.2017.1325613, 2017.

Boettcher, M., Flagg, D. D., Grawe, D., Hoffmann, P., Schoetter, R., Petrik, R., and Schlünzen, K. H.: Hamburg urban summer climate data on 250 m resolution (Version 1.0), https://doi.org/10.26050/WDCC/METRAS_1981-2010_summer_v1, dataset, DOI: https://doi.org/10.26050/WDCC/METRAS_1981-2010_summer_v1, 2024a.

Boettcher, M., Flagg, D. D., Grawe, D., Hoffmann, P., Schoetter, R., Petrik, R., and Schlünzen, K. H.: Hamburg urban winter climate data on 250 m resolution (Version 1.0), https://doi.org/10.26050/WDCC/METRAS_1981-2010_winter_v1, dataset, DOI: https://doi.org/10.26050/WDCC/METRAS_1981-2010_winter_v1, 2024b.

Briscolini, M. and Santangelo, P.: Development of the mask method for incompressible unsteady flows, Journal of Computational Physics, 84, 57–75, https://doi.org/10.1016/0021-9991(89)90181-2, 1989.

Bruse, M. and Fleer, H.: Simulating surface–plant–air interactions inside urban environments with a three dimensional numerical model, Environmental Modelling & Software, 13, 373–384, https://doi.org/10.1016/S1364-8152(98)00042-5, 1998.

De Vos, M. G., Hazeleger, W., Bari, D., Behrens, J., Bendoukha, S., Garcia-Marti, I., Van Haren, R., Haupt, S. E., Hut, R., Jansson, F., Mueller, A., Neilley, P., Van Den Oord, G., Pelupessy, I., Ruti, P., Schultz, M. G., and Walton, J.: Open weather and climate science in the digital era, Geoscience Communication, 3, https://doi.org/10.5194/gc-3-191-2020, 2020.

Eaton, B., Gregory, J., Drach, B., Taylor, K., Hankin, S., Caron, J., Signell, R., Bentley, P., Rappa, G., Höck, H., Pamment, A., Juckes, M., Raspaud, M., Blower, J., Horne, R., Whiteaker, T., Blodgett, D., Zender, C., Lee, D., Hassell, D., Snow, A. D., Kölling, T., Allured, D., Jelenak, A., Soerensen, A. M., Gaultier, L., Herlédan, S., Manzano, F., Bärring, L., Barker, C., and Bartholomew, S. L.: NetCDF Climate and Forecast (CF) Metadata Conventions, https://doi.org/10.5281/zenodo.14275599, 2024.

Eggleton, F. and Winfield, K.: Open data challenges in climate science, Data Science Journal, 19, https://doi.org/10.5334/dsj-2020-052, 2020.

Eichhorn, J.: MISKAM Manual for Version 6, uRL: https://www.lohmeyer.de/site/assets/files/4840/manual_miskam_63.pdf, accessed: 2025-06-23, 2011.

Eichhorn, J. and Kniffka, A.: The numerical flow model MISKAM: State of development and evaluation of the basic version, Meteorologische Zeitschrift, 19, 81–90, https://doi.org/10.1127/0941-2948/2010/0425, 2010.

Eyring, V., Bony, S., Meehl, G. A., Senior, C. A., Stevens, B., Stouffer, R. J., and Taylor, K. E.: Overview of the Coupled Model Intercomparison Project Phase 6 (CMIP6) experimental design and organization, Geoscientific Model Development, 9, https://doi.org/10.5194/gmd-9-1937-2016, 2016.

Ferner, K. S., Boettcher, M., and Schlünzen, K. H.: Modelling the heterogeneity of rain in an urban neighbourhood with an obstacle-resolving model, Meteorologische Zeitschrift, 32, 67–81, https://doi.org/10.1127/metz/2022/1149, 2023.

Ferner, K. S., Boettcher, M., Sieck, K., and Voss, V.: Sensitivity studies on rain heterogeneity in Hamburg city centre using the obstacle-resolving model MITRAS, https://doi.org/10.26050/WDCC/MitrasWinterHamburg, 2024.

Ganske, A., Kraft, A., Kaiser, A., Heydebreck, D., Lammert, A., Höck, H., Thiemann, H., Voss, V., Grawe, D., Leitl, B., Schlünzen, K. H., Kretzschmar, J., and Quaas, J.: ATMODAT Standard (v3.0), World Data Center for Climate (WDCC) at DKRZ, https://doi.org/10.35095/WDCC/atmodat_standard_en_v3_0, 2021.

Ganske, A., Heil, A., Lammert, A., Kretzschmar, J., and Quaas, J.: Publication of Atmospheric Model Data using the ATMODAT Standard, Meteorologische Zeitschrift, 31, 493–504, https://doi.org/10.1127/metz/2022/1118, 2022.

Harris, C. R., Millman, K. J., van der Walt, S. J., Gommers, R., Virtanen, P., Cournapeau, D., Wieser, E., Taylor, J., Berg, S., Smith, N. J., Kern, R., Picus, M., Hoyer, S., van Kerkwijk, M. H., Brett, M., Haldane, A., del Río, J. F., Wiebe, M., Peterson, P., Gérard-Marchant, P., Sheppard, K., Reddy, T., Weckesser, W., Abbasi, H., Gohlke, C., and Oliphant, T. E.: Array programming with NumPy, Nature, 585, 357–362, https://doi.org/10.1038/s41586-020-2649-2, 2020.

Hoyer, S. and Hamman, J.: xarray: N-D labeled arrays and datasets in Python, Journal of Open Research Software, 5, https://doi.org/10.5334/jors.148, 2017.

Hrisko, J.: Geographic Visualizations in Python with Cartopy. Maker Portal., uRL: https://makersportal.com/blog/2020/4/24/geographic-visualizations-in-python-with-cartopy, accessed: 2024-02-15, 2020.

Jänicke, B., Milošević, D., and Manavvi, S.: Review of User-Friendly Models to Improve the Urban Micro-Climate, Atmosphere, 12, https://doi.org/10.3390/atmos12101291, 2021.

Karl, M.: City-scale Chemistry Transport Model EPISODE-CityChem, https://doi.org/10.5281/ZENODO.14361973, dOI: https://doi.org/10.5281/ZENODO.14361973, 2024.

Kretzschmar, J., atmodatcode, am kaiser, and wachsylon: AtMoDat, atmodat_data_checker: Version 1.3.1, https://doi.org/10.5281/zenodo.6701508, dOI: https://doi.org/10.5281/zenodo.6701508, 2022.

Loewe, P.: Lamb weather types (reduced set) and gale days over the North Sea since 1948 based on NCEP/NCAR Reanalysis 1 daily mean sea level pressure fields, https://doi.org/10.26050/WDCC/LambWTyRSetAndGaleDaysOverTheNo, dataset, DOI: https://doi.org/10.26050/WDCC/LambWTyRSetAndGaleDaysOverTheNo, 2022.

Loewe, P. and Schade, N.: Lamb weather types (reduced set) and gale days over the Baltic Sea since 1948 based on NCEP/NCAR Reanalysis 1 daily mean sea level pressure fields, https://doi.org/10.26050/WDCC/LambWTyRSetAndGaleDaysOverTheBa, dataset, DOI: https://doi.org/10.26050/WDCC/LambWTyRSetAndGaleDaysOverTheBa, 2024.

Maronga, B., Gross, G., Raasch, S., Banzhaf, S., Forkel, R., Heldens, W., Kanani-Sühring, F., Matzarakis, A., Mauder, M., Pavlik, D., Pfafferott, J., Schubert, S., Seckmeyer, G., Sieker, H., and Winderlich, K.: Development of a new urban climate model based on the model PALM – Project overview, planned work, and first achievements, Meteorologische Zeitschrift, 28, 105–119, https://doi.org/10.1127/metz/2019/0909, 2019.

Maronga, B., Banzhaf, S., Burmeister, C., Esch, T., Forkel, R., Fröhlich, D., Fuka, V., Gehrke, K. F., Geletič, J., Giersch, S., Gronemeier, T., Groß, G., Heldens, W., Hellsten, A., Hoffmann, F., Inagaki, A., Kadasch, E., Kanani-Sühring, F., Ketelsen, K., Khan, B. A., Knigge, C., Knoop, H., Krč, P., Kurppa, M., Maamari, H., Matzarakis, A., Mauder, M., Pallasch, M., Pavlik, D., Pfafferott, J., Resler, J., Rissmann, S., Russo, E., Salim, M., Schrempf, M., Schwenkel, J., Seckmeyer, G., Schubert, S., Sühring, M., Von Tils, R., Vollmer, L., Ward, S., Witha, B., Wurps, H., Zeidler, J., and Raasch, S.: Overview of the PALM model system 6.0, Geoscientific Model Development, 13, 1335–1372, https://doi.org/10.5194/gmd-13-1335-2020, 2020.

Met Office: Cartopy: a cartographic python library with a Matplotlib interface, https://scitools.org.uk/cartopy, 2010 - 2015.

https://doi.org/10.5194/egusphere-2025-5521
Preprint. Discussion started: 22 December 2025

Mülmenstädt, J.: ECHAM6-HAM2 model experiments to characterize the effect of parameterized autoconversion/warm rain on cloud lifetime, https://doi.org/10.26050/WDCC/Aerocom_ECHAM6-HAM2_rain, dataset, DOI: https://doi.org/10.26050/WDCC/Aerocom_ECHAM6-HAM2_rain, 2022.

Müller, W. and Seyfert, E.: Atkis (R) data base revision and generation of digital topographic base maps, 2000.

Oke, T. R., Mills, G., Christen, A., and Voogt, J. A.: Methods. In Urban Climates, Cambridge University Press, p. 44–76, https://doi.org/10.1017/9781139016476.004, 2017.

OpenFOAM: Documentation, uRL: https://www.openfoam.com/, 2025.

pandas development team, T.: pandas-dev/pandas: Pandas, https://doi.org/10.5281/zenodo.3509134, 2020.

Poole, E. L., Heroux, M. A., Vaidya, P., and Joshi, A.: Performance of iterative methods in ANSYS on cray parallel/vector supercomputers, Computing Systems in Engineering, 6, 251–259, https://doi.org/10.1016/0956-0521(95)00016-S, 1995.

Rew, R. and Davis, G.: NetCDF: an interface for scientific data access, IEEE Computer Graphics and Applications, 10, 76–82, https://doi.org/10.1109/38.56302, 1990.

Salim, M. H., Schlünzen, K. H., and Grawe, D.: Including trees in the numerical simulations of the wind flow in urban areas: Should we care?, Journal of Wind Engineering and Industrial Aerodynamics, 144, 84–95, https://doi.org/10.1016/j.jweia.2015.05.004, 2015.

Salim, M. H., Schlünzen, K. H., Grawe, D., Boettcher, M., Gierisch, A. M., and Fock, B. H.: The microscale obstacle-resolving meteorological model MITRAS v2.0: Model theory, Geoscientific Model Development, 11, 3427–3445, https://doi.org/10.5194/gmd-11-3427-2018, 2018.

Samsel, K. S., Boettcher, M., Grawe, D., Schlünzen, K. H., and Sieck, K.: Simulations for assessing model extensions using the obstacle-resolving model MITRAS, https://doi.org/10.26050/WDCC/WINTER_HAM_MitrasModEx, dataset, DOI: https://doi.org/10.26050/WDCC/WINTER_HAM_MitrasModEx, 2025.

Schatzmann, M., Bächlin, W., Emeis, S., Kühlwein, J., Leitl, B., Müller, W. J., Schäfer, K., and Schlünzen, K. H.: Development and validation of tools for the implementation of European air quality policy in Germany (Project VALIUM), Atmospheric Chemistry and Physics, 6, 3077–3083, https://doi.org/10.5194/acp-6-3077-2006, 2006.

Scherer, D., Antretter, F., Bender, S., Cortekar, J., Emeis, S., Fehrenbach, U., Gross, G., Halbig, G., Hasse, J., Maronga, B., Raasch, S., and Scherber, K.: Urban climate under change [UC]2 - A national research programme for developing a building-resolving atmospheric model for entire city regions, Meteorologische Zeitschrift, 28, 95–104, https://doi.org/10.1127/metz/2019/0913, 2019.

Scherer, D., Fehrenbach, U., Grassmann, T., Holtmann, A., Meier, F., Scherber, K., Pavlik, D., Höhne, T., Kanani-Sühring, F., Maronga, B., Ament, F., Banzhaf, S., Langer, I., Halbig, G., Kohler, M., Queck, R., Stratbücker, S., Winkler, M., Wegener, R., and Zeeman, M.: [UC]2 -Datenstandard "Stadtklima im Wandel", Version 1.5.2, URL: https://uc2-program.org/sites/default/files/inline-files/uc2_datenstandard.pdf, accessed: 2024-09-01, 2022.

Schluenzen, K.: Numerical studies on the inland penetration of sea breeze fronts at a coastline with tidally flooded mudflats, Beiträge zur Physik der Atmosphäre, 63 (3/4), 243-256., 1990.

Schlünzen, K. H., Hinneburg, D., Knoth, O., Lambrecht, M., Leitl, B., López, S., Lüpkes, C., Panskus, H., Renner, E., Schatzmann, M., Schoenemeyer, T., Trepte, S., and Wolke, R.: Flow and Transport in the Obstacle Layer: First Results of the Micro-Scale Model MITRAS, Journal of Atmospheric Chemistry, 44, 113–130, https://doi.org/10.1023/A:1022420130032, 2003.

Schlünzen, K. H., Boettcher, M., Fock, B. H., Gierisch, A., Grawe, D., and Salim, M.: Scientific Documentation of the Multiscale Model System M-SYS, MEMI Technical Report 4, URL: https://www.mi.uni-hamburg.de/en/arbeitsgruppen/memi/modelle/dokumentation/msys-scientific-documentation-20180706.pdf, accessed: 2024-06-22, 2018a.

https://doi.org/10.5194/egusphere-2025-5521
Preprint. Discussion started: 22 December 2025

Schlünzen, K. H., Boettcher, M., Fock, B. H., Gierisch, A., Grawe, D., and Salim, M.: Technical Documentation of the Multiscale Model System M-SYS, MEMI Technical Report 4, URL: https://www.mi.uni-hamburg.de/en/arbeitsgruppen/memi/modelle/dokumentation/msys-technical-documentation-20180706.pdf, accessed: 2024-06-22, 2018b.

Touzé-Peiffer, L., Barberousse, A., and Le Treut, H.: The Coupled Model Intercomparison Project: History, uses, and structural effects on climate research, Wiley Interdisciplinary Reviews: Climate Change, 11, https://doi.org/10.1002/wcc.648, 2020.

Trukenmüller, A., Grawe, D., and Schlünzen, K. H.: A model system for the assessment of ambient air quality conforming to EC directives, Meteorologische Zeitschrift, 13, 387–394, https://doi.org/10.1127/0941-2948/2004/0013-0387, 2004.

Voss, V.: High resolution obstacle-resolved model results for the city centre of Hamburg, Germany, from the microscale transport and stream model MITRAS on 21 June 2000, https://doi.org/10.26050/WDCC/MITRAS_ATMODAT, dataset, DOI: https://doi.org/10.26050/WDCC/MITRAS_ATMODAT, 2023.

Voss, V.: AtMoDat/nc2atmodat: v1.0.2, https://doi.org/10.5281/zenodo.17035812, 2025.

Voss, V., Schlünzen, K. H., and Grawe, D.: Survey On Output Data From Obstacle Resolving Atmospheric Models, Meteorologische Zeitschrift, 33, 243–253, https://doi.org/10.1127/metz/2024/1217, 2024.

Whitaker, J., Snow, A. D., Cochran, M., Karney, C., Ouzounoudis, G., Dearing, J., de Kloe, J., Heitor, Filipe, May, R., Itkin, M., Eubank, N., Dunphy, M., Brett, M., da Costa, M. A., de Maeyer, J., Gohlke, C., Willoughby, C., Barker, C., Wiedemann, B. M., and Couwenberg, B.: jswhit/pyproj: version 2.0.2 release, https://doi.org/10.5281/zenodo.2592233, 2019.

Wilkinson, M. D., Dumontier, M., Aalbersberg, I. J., Appleton, G., Axton, M., Baak, A., Blomberg, N., Boiten, J.-W., da Silva Santos, L. B., Bourne, P. E., Bouwman, J., Brookes, A. J., Clark, T., Crosas, M., Dillo, I., Dumon, O., Edmunds, S., Evelo, C. T., Finkers, R., Gonzalez-Beltran, A., Gray, A. J., Groth, P., Goble, C., Grethe, J. S., Heringa, J., 't Hoen, P. A., Hooft, R., Kuhn, T., Kok, R., Kok, J., Lusher, S. J., Martone, M. E., Mons, A., Packer, A. L., Persson, B., Rocca-Serra, P., Roos, M., van Schaik, R., Sansone, S.-A., Schultes, E., Sengstag, T., Slater, T., Strawn, G., Swertz, M. A., Thompson, M., van der Lei, J., van Mulligen, E., Velterop, J., Waagmeester, A., Wittenburg, P., Wolstencroft, K., Zhao, J., and Mons, B.: The FAIR Guiding Principles for scientific data management and stewardship, Scientific Data, 3, 160 018, https://doi.org/10.1038/sdata.2016.18, 2016.

WMO: Guidance on Measuring, Modelling and Monitoring the Canopy Layer Urban Heat Island (CL-UHI). WMO-No. 1292, World Meteorological Organisation (WMO), CH-1211 Geneva 2, Switzerland., https://library.wmo.int/idurl/4/58410, uRL: https://library.wmo.int/idurl/4/58410, accessed: 2025-06-16, 2023.