First results of a model user survey on a micro-scale model data standard

Vivien Voss[1], K. Heinke Schlünzen[1], David Grawe[2], Daniel Heydebreck[3], and Anette Ganske[4]

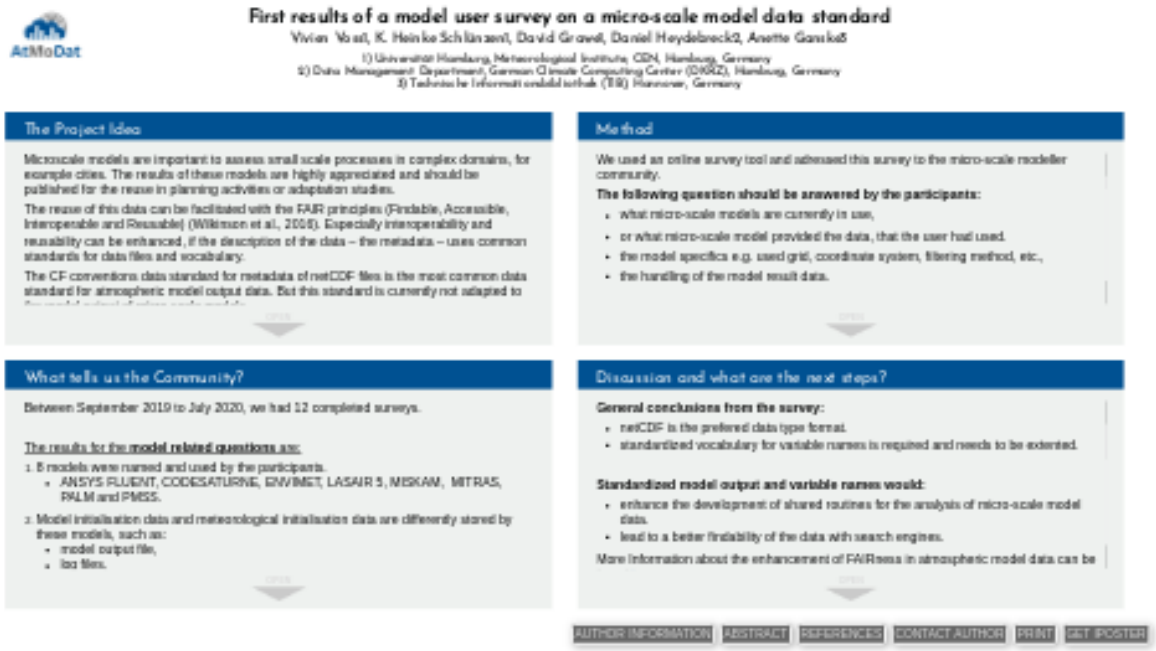
[1]Universität Hamburg
[2]CEN
[3]Deutsches Klimarechenzentrum GmbH
[4]Technische Informationsbibliothek TIB

November 21, 2022

Abstract

Micro-scale models are important to assess processes in complex domains, for example cities. The most common data standard for atmospheric model output data are the CF-conventions, a data standard for netCDF files, but this standard is not adapted to the model output of micro-scale models. As a part of the project AtMoDat (Atmospheric Model Data) we want to develop a model data standard for obstacle resolving models (ORM), including the additional variables (i.e. building structures, wall temperatures) used by these models. In order to involve the micro-scale modeller community in this process, a web based survey was developed and distributed in the modeller community via conferences and email. With this survey we want to find out which micro-scale ORMs are currently in use, their model specifics (e.g. used grid, coordinate system), and the handling of the model result data. Furthermore, the survey provides the opportunity to include suggestions and ideas, what we should consider in the development of the standard. Between September 2019 and July 2020, the survey was accessed 29 times, but only 12 surveys were completed. The finished surveys refer to eight different models and their corresponding model information. Results show that these different models use different output formats and processing tools, which results in different model result handling routines. The participants suggested to use the netCDF data format and to provide information on model initialization, model settings and model input along with the model output data. This would enable an easier intercomparison between different models and repetition of model simulations. Standardized model output and variable names would also enhance the development of shared routines for the analysis of micro-scale model data and a better findability of the data with search engines. This survey will remain open with regular assessments of contents (i.e. November 2020, May 2021; https://uhh.de/orm-survey).

First results of a model user survey on a micro-scale model data standard



Vivien Voss1, K. Heinke Schlünzen1, David Grawe1, Daniel Heydebreck2, Anette Ganske3

1) Universität Hamburg, Meteorological Institute, CEN, Hamburg, Germany
2) Data Management Department, German Climate Computing Center (DKRZ), Hamburg, Germany
3) Technische Informationsbibliothek (TIB), Hannover, Germany

PRESENTED AT:

AGU FALL MEETING
Online Everywhere | 1–17 December 2020

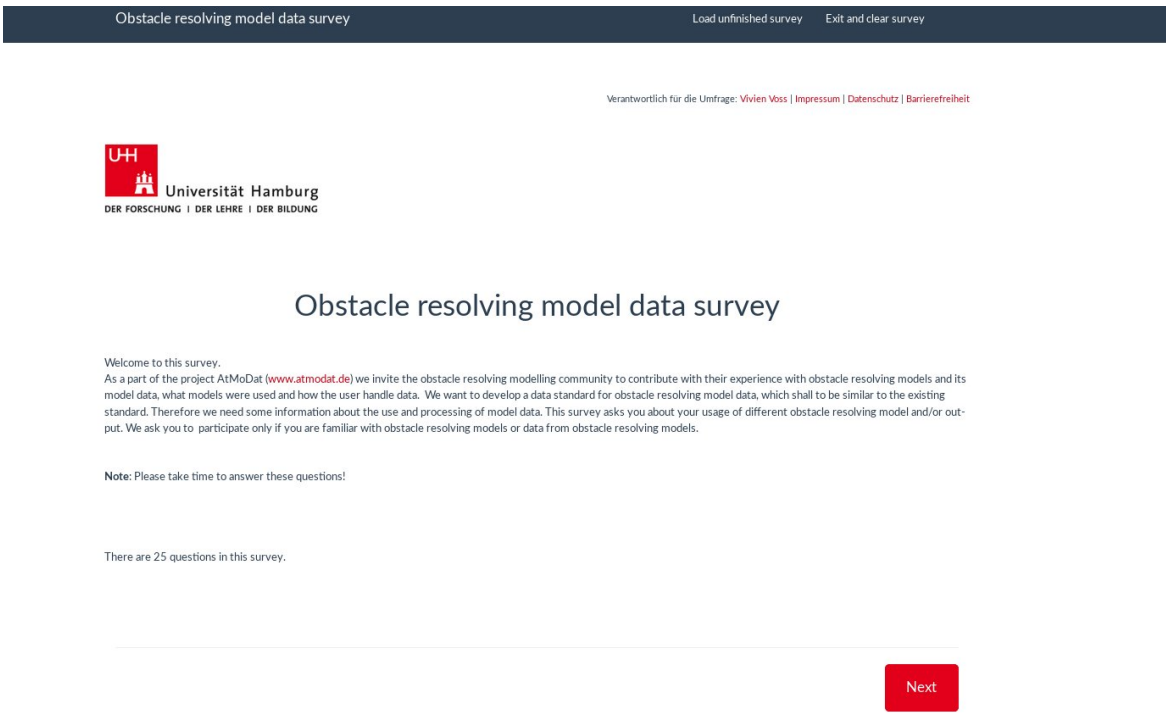
THE PROJECT IDEA

Microscale models are important to assess small scale processes in complex domains, for example cities. The results of these models are highly appreciated and should be published for the reuse in planning activities or adaptation studies.

The reuse of this data can be facilitated with the FAIR principles (Findable, Accessible, Interoperable and Reusable) (Wilkinson et al., 2016). Especially interoperability and reusability can be enhanced, if the description of the data – the metadata – uses common standards for data files and vocabulary.

The CF conventions data standard for metadata of netCDF files is the most common data standard for atmospheric model output data. But this standard is currently not adapted to the model output of micro-scale models.

Within the project AtMoDat (Atmospheric model data) we want to enhance the publication FAIR and open atmospheric model data. As there are no common standards for microscale models, we started a web-based survey about these models in September 2019. The results will be used to involve the community in the development of a new standard for microscale models. The survey is still open and further input is appreciated.



WHAT TELLS US THE COMMUNITY?

Between September 2019 to July 2020, we had 12 completed surveys.

The results for the **model related questions** are:

1. 8 models were named and used by the participants.
   - ANSYS FLUENT, CODESATURNE, ENVIMET, LASAIR 5, MISKAM, MITRAS, PALM and PMSS.
2. Model initialisation data and meteorological initialisation data are differently stored by these models, such as:
   - model output file,
   - log files,
   - external input files, i.e. namelists, Excel,
   - report files.
3. Used coordinate system and grids are:
   - Horizontal coordinates (Figure 1):
     - Cartesian,
     - Lamber conformal,
     - rotated lat. lon.,
     - UTM.
   - Vertical coordinates (Figure 2):
     - z coordinate,
     - surface fitted grid.
   - Grid type (Figure 3):
     - Arakawa-C,
     - non-uniform,
     - uniform.
4. All three filtering methods (RANS, LES, DNS) were selected.
   The majority used the RANS Filtering Method (Figure 4).
5. Used output formats were (Figure 5):
   - binary data format (exept netCDF),
   - ascii data format,
   - netCDF.

The results for the **user related questions** are:

1. Different knowledge on existing data standards.
2. A variety of analysing tools were used.
   i.e. CDO, NCO, NCL, Python, R, Matlab,...

Additional annotation of the participants are:

- netCDF prefered output format.
- Treatment of different coordinate systems seems to be difficult.
- Usage of consistent variable names is very important.
- Data should be easily read and processed by different applications and skripting languages (e.g. python).

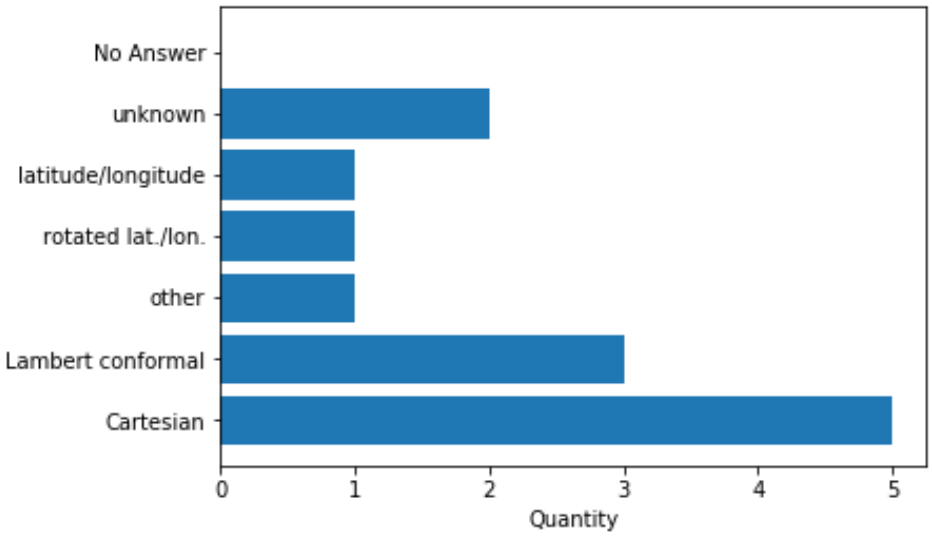


Figure 1: Selected answers regarding the horizontal coordinate system, multiple choice.

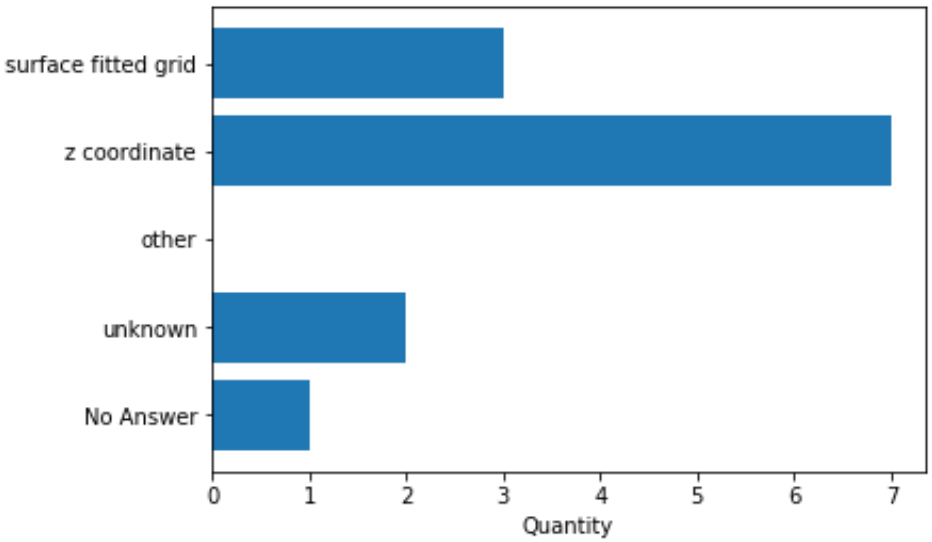


Figure 2: Selected answers regarding the vertical coordinate system, multiple choice.

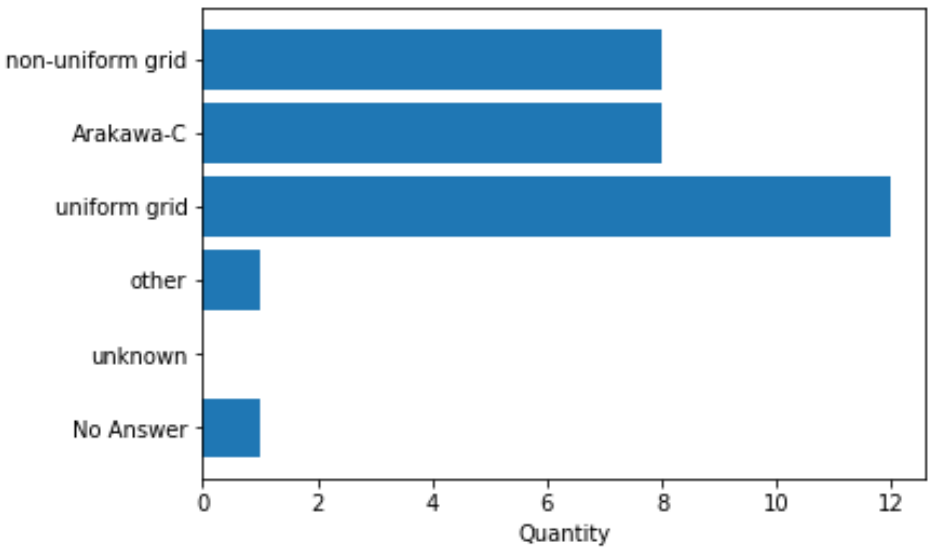


Figure 3: Selected answers regarding the grid, multiple choice.

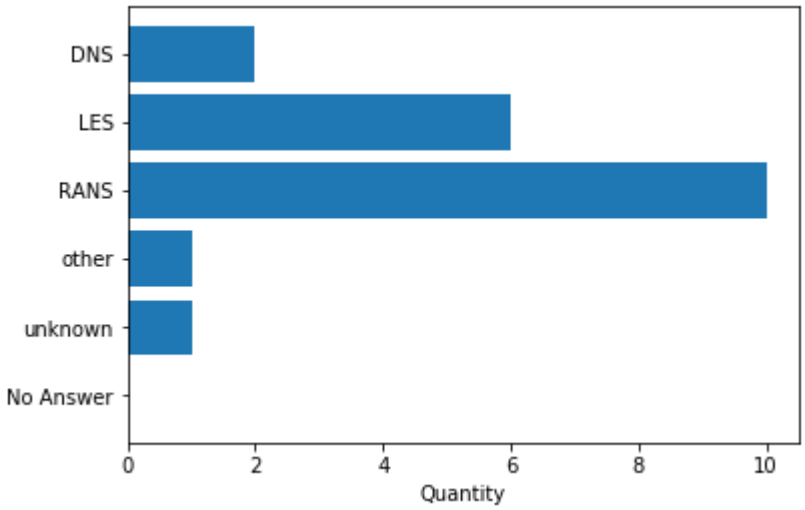


Figure 4: Selected answers regarding the filtering method, multiple choice.

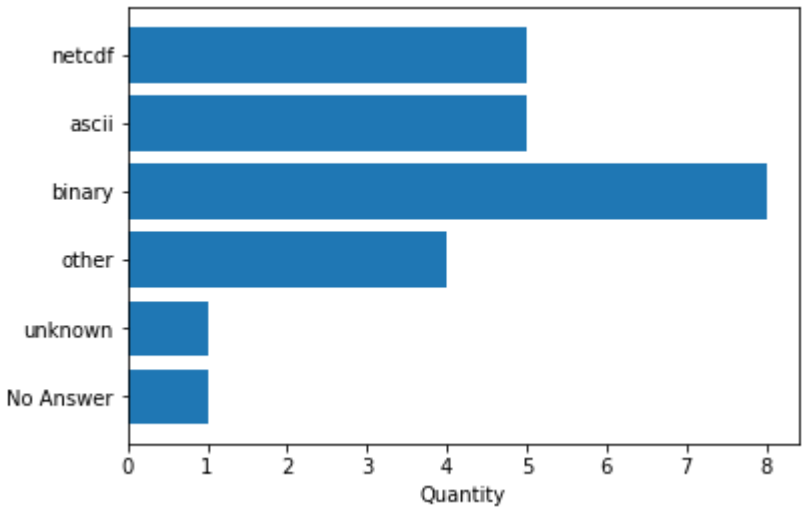


Figure 5: Selected answers regarding the output format, multiple choice.

METHOD

We used an online survey tool and adressed this survey to the micro-scale modeller community.

The following question should be answered by the participants:

- what micro-scale models are currently in use,
- or what micro-scale model provided the data, that the user had used.
- the model specifics e.g. used grid, coordinate system, filtering method, etc.,
- the handling of the model result data.

Also it provided the opportuinity to give us:

- suggestions and ideas, what we should consider in the development of the standard.

The participant can name up to 12 models and answer several questions for each model individually (Figure 1). Some questions provide the opportunity to more detailed answers (Figure 2) or multiple choice.

Model Usage

✱ Are you currently working with a obstacle resolving model?

Yes | No

Please list the model(s) you use.

Model 1	MITRAS
Model 2	
Model 3	
Model 4	
Model 5	
Model 6	

Did you ever use data from an obstacle resolving model, where you did not prepare the data yourself?

Yes | No | No answer

ModelSetup

In this section we want to ask you about the features of the models you mentioned in the section before.

Are the meteorological data of the initialisation stored in the output?

	Yes	Uncertain	No	No answer
MITRAS	●	○	○	○

How does the model store the initialisation data? (Please write your answer)

MITRAS: written in the first model output

Are the initial settings of the model simulations stored in the output?

	Yes	Uncertain	No	No answer
MITRAS	○	○	●	○

DISCUSSION AND WHAT ARE THE NEXT STEPS?

General conclusions from the survey:

- netCDF is the prefered data type format.
- standardized vocabulary for variable names is required and needs to be extented.

Standardized model output and variable names would:

- enhance the development of shared routines for the analysis of micro-scale model data.
- lead to a better findability of the data with search engines.

More Information about the enhancement of FAIRness in atmospheric model data can be found **here. (/default.aspx?s=36-FA-69-AF-15-1C-17-CB-33-CA-5E-68-19-FA-CA-5C&guestview=true)**

In order to enhance the reusability of model data and especially model data from micro-scale models, we invite you to tell us:

- **what should the data look like?**
- **what information do you need from the data for the analysis?**
- **what is missing in the current state of the provided data?**

This survey will remain open with regular assessments of contents (i.e. January 2021) and is accessable via: **https://uhh.de/orm-survey (https://uhh.de/orm-survey)** or via the QR-Code.

(https://uhh.de/orm-survey)

www.atmodat.de

AUTHOR INFORMATION

Vivien Voss has graduated with a master degree in meteorology in June 2019 at the University of Hamburg, Germany. In her master thesis she dealed with micro scale modelling of air pollution in urban areas. Currently, Vivien is a scientific researcher at the Meteorological Institute of the University of Hamburg in the working group „mesoscale and microscale modelling (MEMI)“, conducted by K. Heinke Schlünzen. She participates in the Project AtMoDat (Atmospheric model data), which aims to create a data standard for microscale models and smaller MIPs.

ABSTRACT

Micro-scale models are important to assess processes in complex domains, for example cities. The most common data standard for atmospheric model output data are the CF-conventions, a data standard for netCDF files, but this standard is not adapted to the model output of micro-scale models. As a part of the project AtMoDat (Atmospheric Model Data) we want to develop a model data standard for obstacle resolving models (ORM), including the additional variables (i.e. building structures, wall temperatures) used by these models. In order to involve the micro-scale modeller community in this process, a web based survey was developed and distributed in the modeller community via conferences and email.

With this survey we want to find out which micro-scale ORMs are currently in use, their model specifics (e.g. used grid, coordinate system), and the handling of the model result data. Furthermore, the survey provides the opportunity to include suggestions and ideas, what we should consider in the development of the standard.

Between September 2019 and July 2020, the survey was accessed 29 times, but only 12 surveys were completed. The finished surveys refer to eight different models and their corresponding model information. Results show that these different models use different output formats and processing tools, which results in different model result handling routines. The participants suggested to use the netCDF data format and to provide information on model initialization, model settings and model input along with the model output data. This would enable an easier intercomparison between different models and repetition of model simulations.

Standardized model output and variable names would also enhance the development of shared routines for the analysis of micro-scale model data and a better findability of the data with search engines.

This survey will remain open with regular assessments of contents (i.e. November 2020, May 2021; https://uhh.de/orm-survey (https://uhh.de/orm-survey)).

REFERENCES

Wilkinson, M., Dumontier, M., Aalbersberg, I. et al. The FAIR Guiding Principles for scientific data management and stewardship. Sci Data 3, 160018 (2016). https://doi.org/10.1038/sdata.2016.18