(Atmospheric) Clouds of Data:
Ruisdael observatory and beyond

Ruisdael team

Jordi Vilà-Guerau de Arellano

Wageningen University

Ruisdael observatory

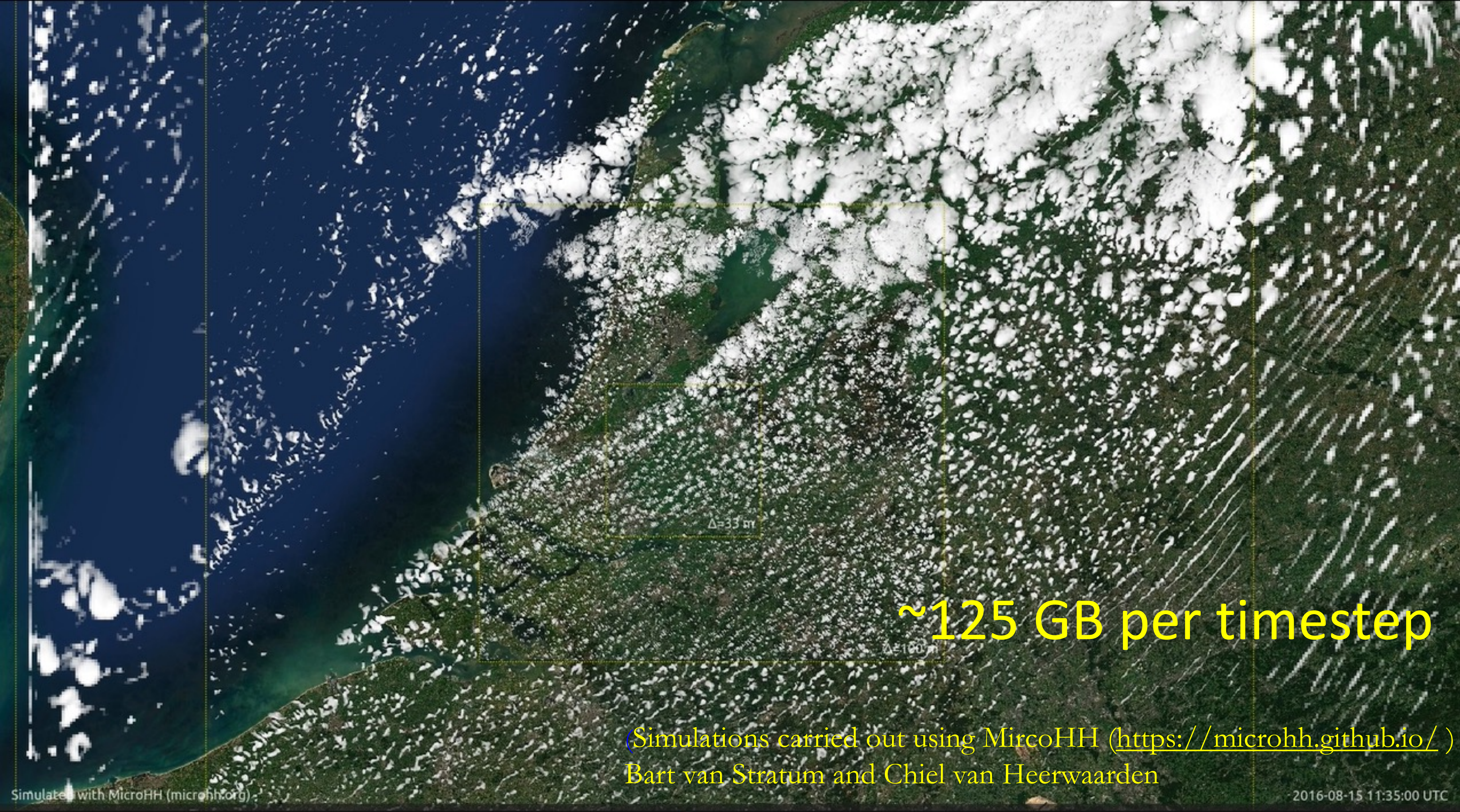
Advanced and comprehensive in-situ and remote sensing observations

High-resolution simulation interconnecting scales and processes

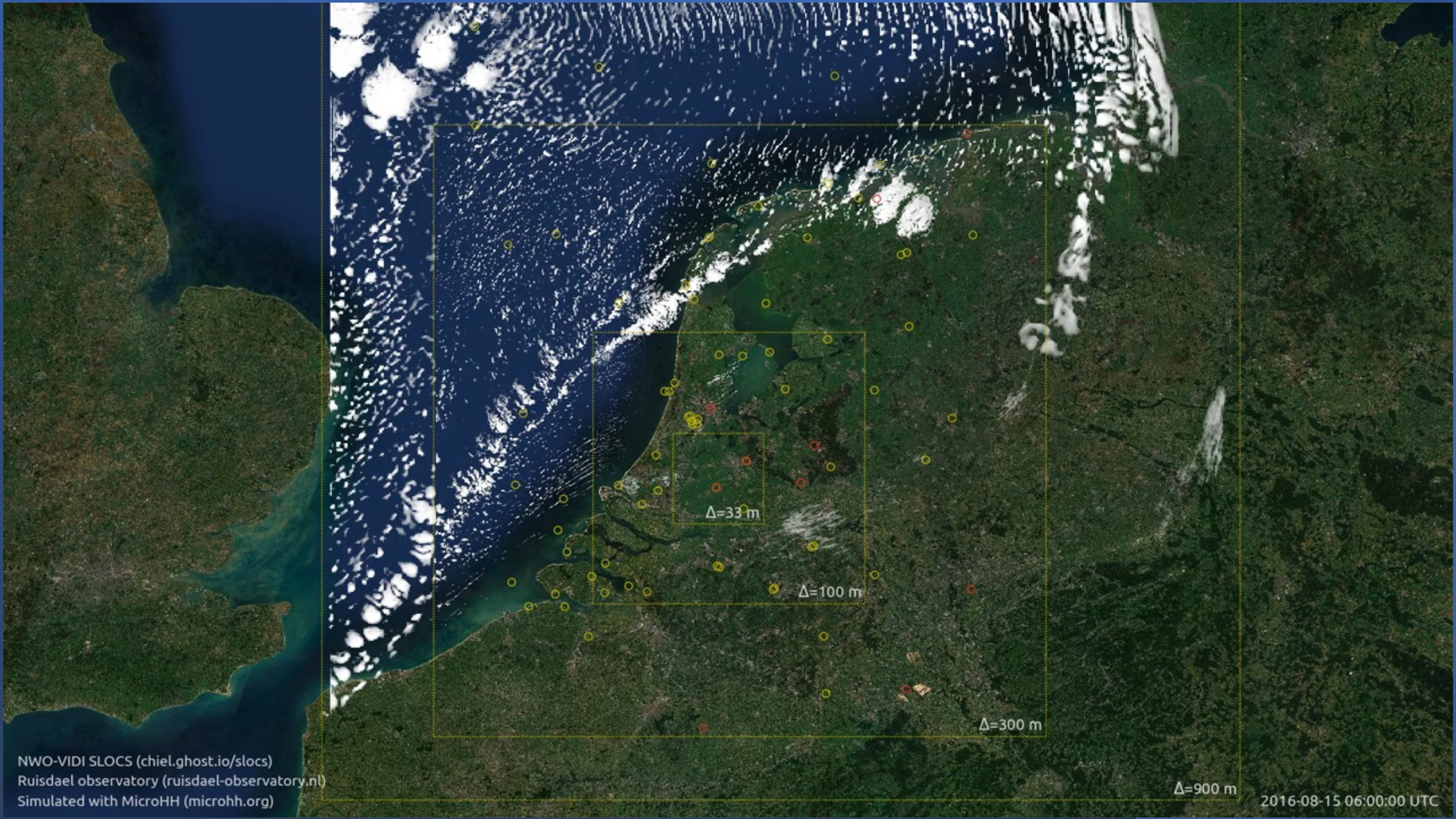
TestBed

Weather Forecast
Air Pollution Forecast
Climate Projections

But also,
Atlantic Trade Winds, Western Europe and Amazon Basin



~125 GB per timestep
(Simulations carried out using MircoHH (https://microhh.github.io/)
Bart van Stratum and Chiel van Heerwaarden
Simulated with MicroHH (microhh.org)
2016-08-15 11:35:00 UTC
Δ=33 m


Δ=33 m
Δ=100 m
Δ=300 m
Δ=900 m
NWO-VIDI SLOCS (chiel.ghost.io/slocs)
Ruisdael observatory (ruisdael-observatory.nl)
Simulated with MicroHH (microhh.org)
2016-08-15 06:00:00 UTC

Ruisdael observatory and beyond
TestBed

4D Numerical Simulations
- Continuum scales
- Wide range of land/sea & atmospheric conditions
- Explicit resolving physical processes (turbulence, clouds)

Integrating Observations
- Remote sensing
- In-situ
- Data Assimilation

Visualize to Quantify
- Understanding processes
- Representations for climate projections

Challenge I

Large Amount Data Numerical Simulations

- Minimize and optimize the time to treat and visualize the data
- Dealing with different formats and frequencies
- Easley handling of time steps, specific heights and variables
- InZarr format with modular access
- Python scripts

⇒ *Example*

Cloud Botany of Shallow Cumulus under a wider range of environmental conditions (Jansson et al., 2024)

Simulations carried out by DALES (https://github.com/dalesteam/dales)

Data management Cloud Botany

Roadmap

1. Intake catalog

2. Zarr format: storing chunked, compressed, N-dimensional array

3. Swift object storage

Powerful, efficient and flexible way to host, access and treat atmospheric data

Easily accessible with Python

Visualization systems that runs in browser enabling you visualize Zarr data

1. Intake catalog

Aim: Describes and points the data sets

Format YAML: human-readable data serialization language

In the Atlantic trade wind project EUREC4A:

- ✓ Hosted and developed by GitHub
- ✓ Packaged as a Python module: import eurec4a
- ✓ Add new datasets by pull requests
- ✓ Automatic testing – notice if a dataset has become unreachable
- ✓ When adding to the catalog, require documentation and examples

2. Zarr format

Zarr is similar to netCDF

- Gridded dada
- Metdata: variable name, unit and coordinates

Xarray (dimensions, coordinates and attributes) has good Zarr support

Latest netCDF library can handle Zarr

Zarr data is stored in multiple miles, **chunks**

Typically 20 MB
Good for online access => Download only needs the required chunks

3. Swift object storage

“Swift” is one way to host Zarr data.

Optimized for handling many “small” files (20 Mb)

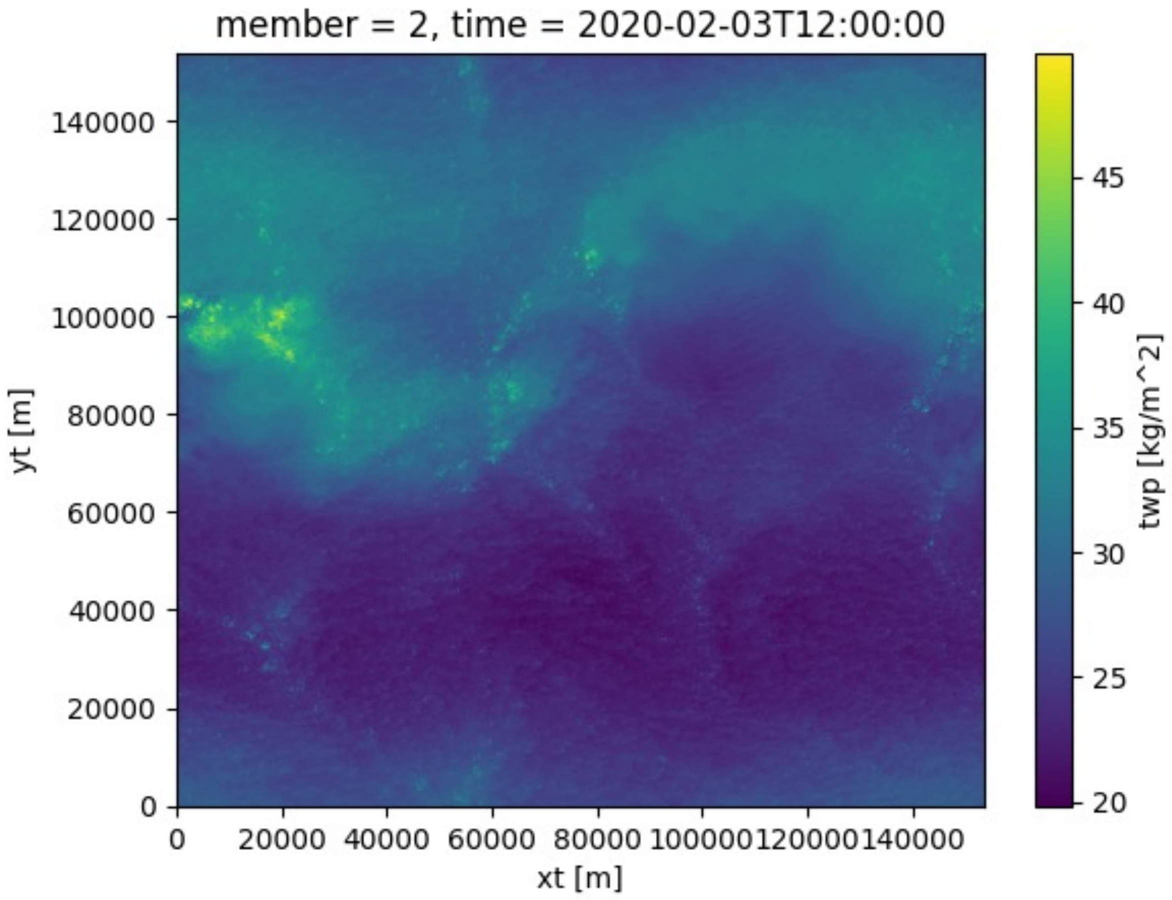
Current possibilities of hosting atmospheric data:

- ✓ EUREC4A data is hosted at DKRZ (German Climate Computing Center)
- ✓ SURF also has a Swift storage service, but requires financial support
- ✓ Commercial cloud providers have these possibilities: Microsoft and Amazon

A representative example: Total Water Path (twp)

[1]:
```python
import eurec4a # <- get the package on top of the intake catalog
cat = eurec4a.get_intake_catalog() # <- establish a connection to the catalog
ds_2D = cat.simulations.DALES.botany.dx100m.nx1536['2D'].to_dask() # <- establish a connection to the 2D data from Botany, and store it in a dask
twp = ds_2D['twp'].isel(member=1, time=-1) # Get final time slice from first member simulation
twp.plot() # Plot it
# Time to plot: 1m 35s
```

[1]: <matplotlib.collections.QuadMesh at 0x7fff0c4086a0>



Benefits

Minimize the time of visualizing and treating data: easy, quick and continuous

Online access => work on the data in Python without downloading data

Python (xarray) integration

Easy for new users

Challenges

User support (for those uploading data sets): Administration and logistics

Coordination, requesting/enforcing documentation and consistency

Where to host the data in short and long term?
Financial support

Challenge II

Integrating Observations and Numerical Simulations

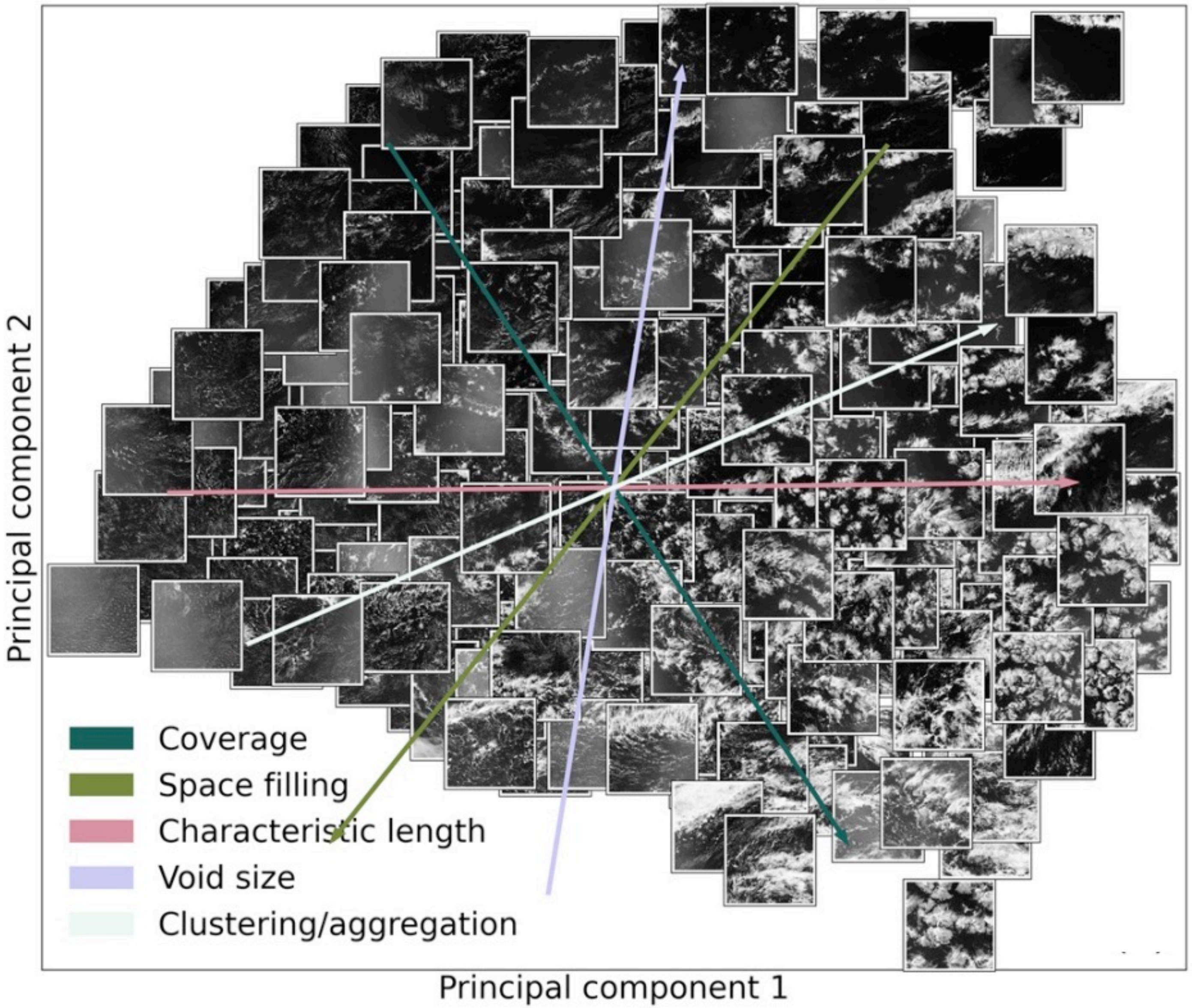
- Facilitate the access of large amount of data
- Advancing on understanding of physics from combing observations and simulations
- Evaluation simulation performance and selecting systematical challenging situations, and with statistically robustness simulations flaws
- Integration with data in an efficient and useful way

⇒ *Examples*

1.Cloud Patterns in the Trades from multiple remote sensing scenes (Janssens et al., 2020)

2.Photosynthesis in Amazonia; large amount of data and dependencies (Pedruzo-Bagazgoitia et al, 2023)

Wide variety of remote sensing cloud patterns at the Trade Winds

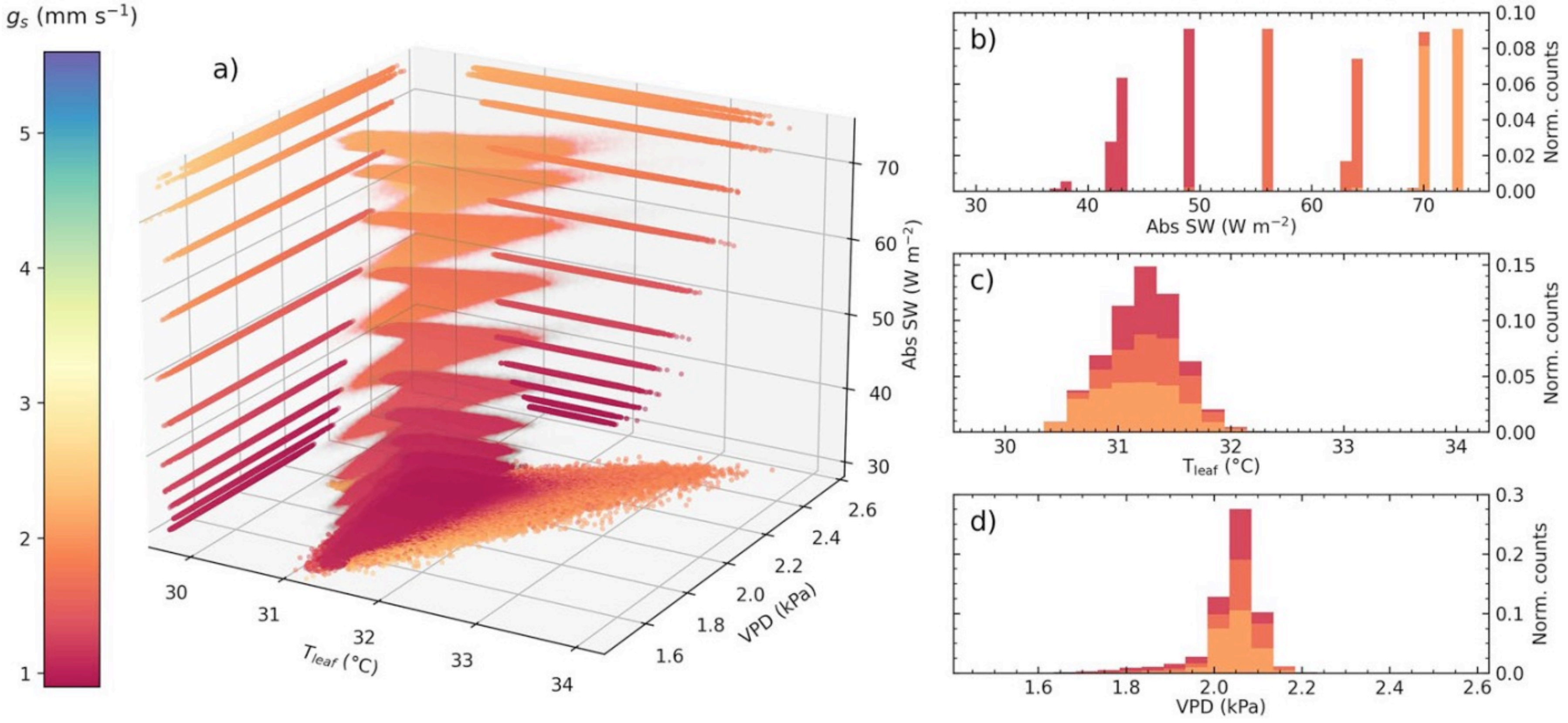


(Janssens et al., 2020)

LES simulation Amazon rainforest photosynthesis as a function of environmental variables

1.10^9 points taken instantaneously at 12 local time

g_s stomatal aperture



Shortwave radiation

Leaf temperature

Water vapor pressure deficit

(Pedruzo-Bagazgoitioanet al., 2020)

Challenge III

Visualization to quantify

- Volume, velocity and variety

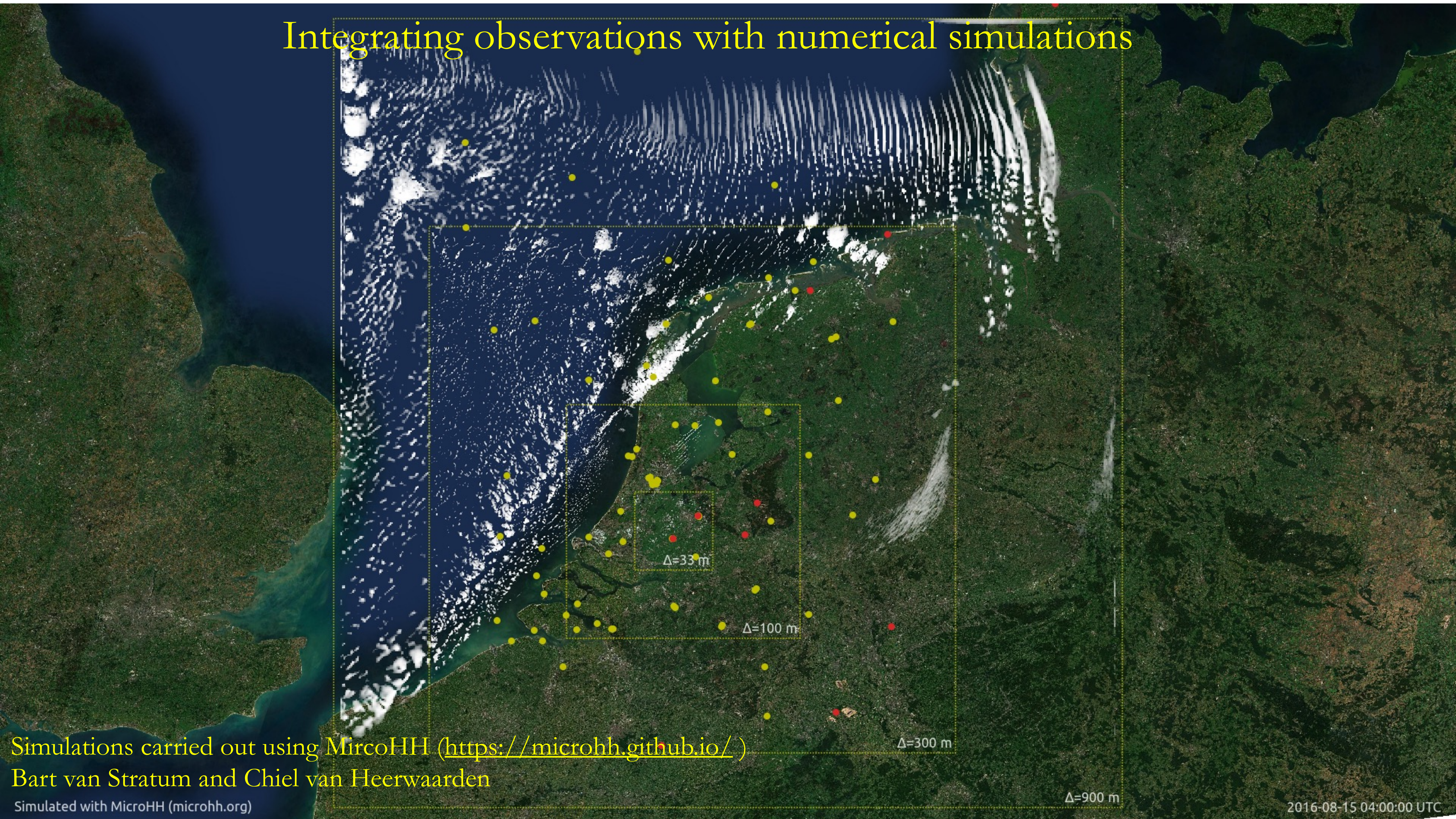
- Visualization that enables you to quantify to advance in physic understanding and provides clarity of the atmospheric complexity

- Improve the outreach/societal impact of your reserach

⇒ *Example*:

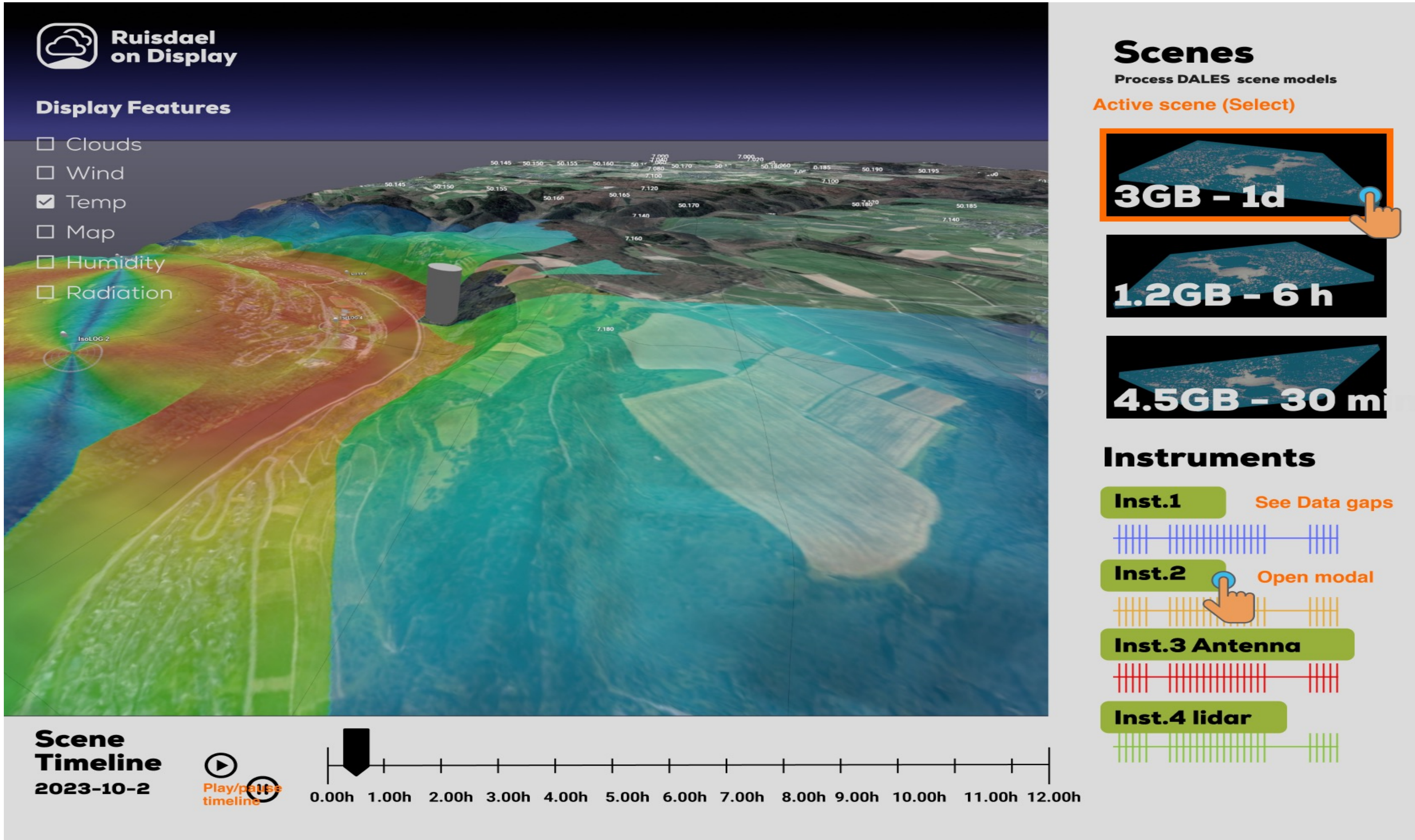
Towards comprehensive, integrative, complete and user friendly TestBed

Integrating observations with numerical simulations



Simulations carried out using MircoHH (https://microhh.github.io/)
Bart van Stratum and Chiel van Heerwaarden

Prototype of Ruisdael in-browser data visualization system



Simulations carried out using DALES
(https://github.com/dalesteam/dales)

Jesus Garcia Gonzalez
eScience project

Current Atmospheric Projects dealing with BigData of The Dutch Atmospheric Community

- Ruisdael observatory: integrative and comprehensive atmospheric research laboratory
https://ruisdael-observatory.nl/cesar/

- EUREC4A: Clouds in the trade winds
https://eurec4a.eu/ and https://howto.eurec4a.eu/bacardi.html

- SLOCS: Clouds and three-dimensional radiation
https://chiel.ghost.io/slocs/ and vimeo.com/channels/microhh

- CloudRoots: Clouds and rainforest
https://cloudroots.wur.nl/

Ruisdael observatory and beyond
TestBed

4D Numerical Simulations

- Continuum scales
- Wide range of land/sea atmospheric conditions
- Explicit resolving physical processes (turbulence, clouds)

Integrating Observations

- Remote sensing
- In-situ
- Data Assimilation

Visualize to Quantify

- Understanding processes
- Representations for climate projections