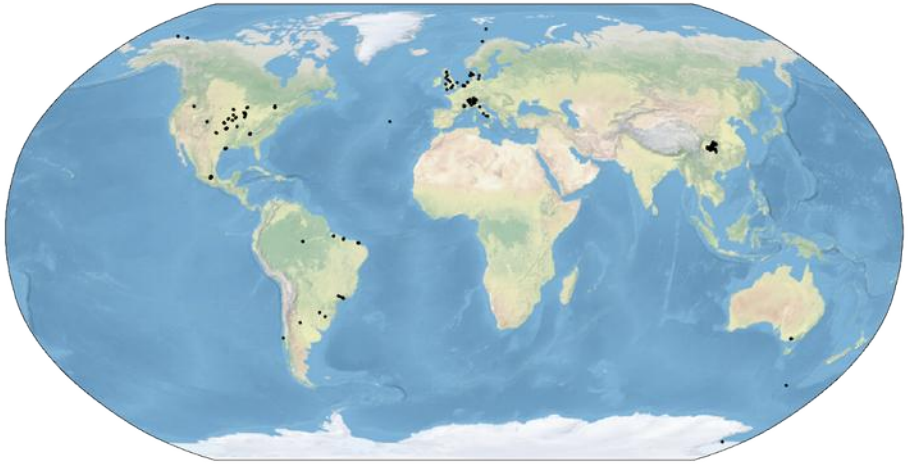
DISDRODB

A global disdrometer archive of raindrop size distribution observations

Gionata Ghiggi[1], Anne Claire Billault-Roux[1], Kim Candolfi[1], Leo Pillac-Mage[1], Christine Unal[3], Marc Schleiss[3], Remko Uijlenhoet[2] , Tim Raupach[4], Alexis Berne[1]

(1) Environmental Remote Sensing Laboratory, EPFL, Lausanne, Switzerland
(2) Department of Water Management, TU Delft, Netherlands
(3) Geoscience & Remote Sensing Department, TU Delft, Netherlands
(4) Institute for Climate Risk and Response, University of New South Wales, Sydney, Australia



Motivation

- Disdrometer data are scattered across institutions
- Access to the data is often very difficult
- Data documentation is often lacking
- Older datasets are at risk of getting lost !

- Different sensors
- Different data formats

Motivation

- Disdrometer data are scattered across institutions
- Access to the data is often very difficult
- Data documentation is often lacking
- Older datasets are at risk of getting lost !

- Different sensors
- Different data formats

No standards for
- Data Exchange
- Data Processing
- Data Quality Checks (QC)

Limited
- Open Source Software
- Analysis-Ready-Data (ARD)
- Global DSD Studies

Motivation

- Disdrometer data are scattered across institutio
- Access to the data is often very difficult
- Data documentation is often lacking
- Older datasets are at risk of getti

- Different sensors
- Different data form

No standard

- Data Exch
- Data Proces g
- Data Quality Checks (QC)

Limited

- Open Source Software
- Analysis-Ready-Data (ARD)
- Global DSD Studies

Weeks are wasted in searching and wrangling the data

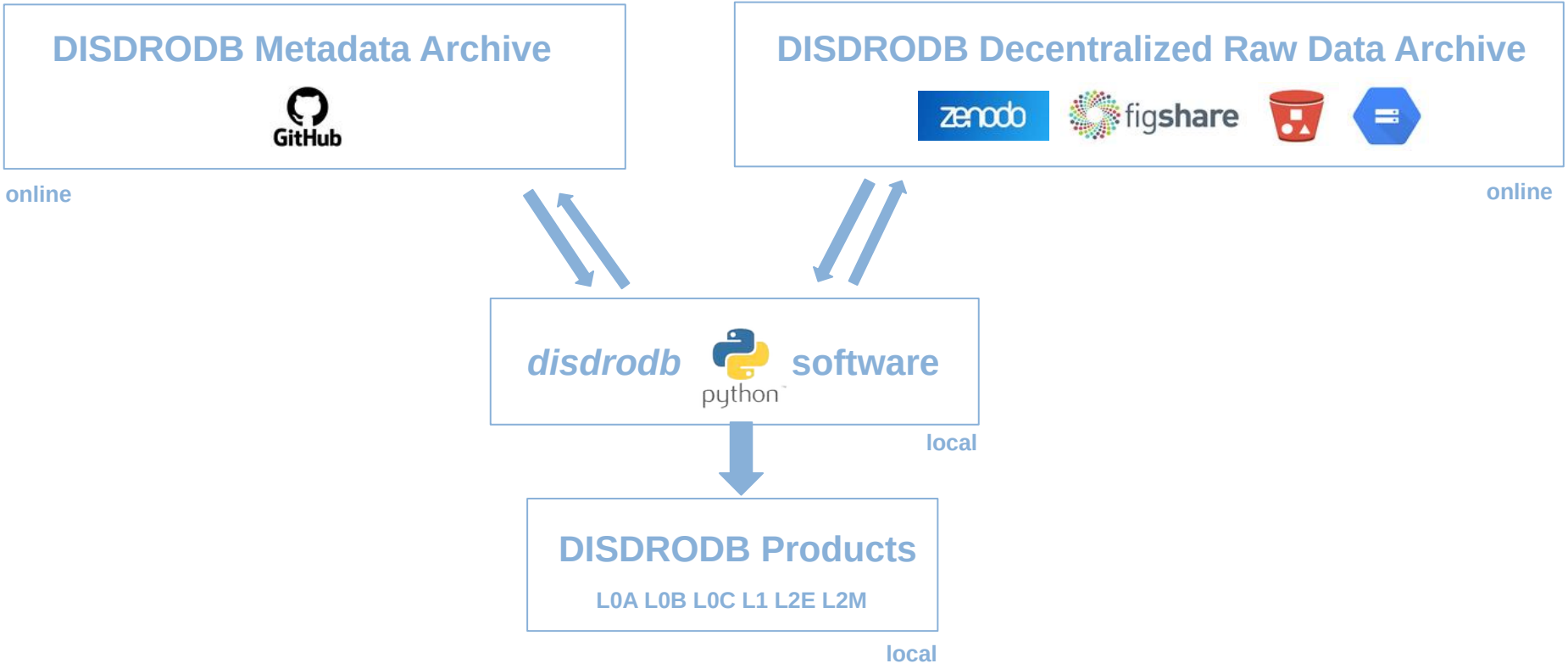
Vision

- A **global archive** of **homogenized** and **quality checked** disdrometer data
- A public **infrastructure** for **data exchange**, **sharing** and **standardization**
- An **open source software** for **data management**, **data processing** and generation of **analysis-ready DSD products**

Goals

- Make very convenient to join the DISDRODB initiative
- Establish a research community automatically sharing their data
- Promote mobilization and standardization of existing data archives

DISDRODB Infrastructure



DISDRODB Metadata Archive

- Keeps track of available disdrometer stations and remote data repositories with raw data
- Each station has a file with a **standardized set of metadata** fields informing about geolocation, authorship, deployment status, issues, … ← Summary of ACTRIS, ARM, DIVEN, OceanRAIN metadata
- **Hosted on GitHub**, enables collaborative and recursive metadata improvements while keeping a **transparent and fully reproducible history**. Changes can be tracked over time !

DISDRODB Decentralized Raw Data Archive

- Remove the cost (IT + staff) associated to maintaining a centralized data archive
- Data contributors do not need to wait for “external” actions
- Contributors upload directly their raw data on a remote repository (i.e. Zenodo)
- Contributors maintain control and ownership of their data !

DISDRODB Data Sharing Mechanism

- To share their data through DISDRODB, contributors just need to:
  1) Upload the station raw data on a remote repository
  2) Add the station reader (if not already available) to the *disdrodb* software
  3) Specify the reader name and the url where the raw data are stored in the metadata
  4) Upload the station metadata file to the DISDRODB Metadata Archive
- Users can download the raw data of stations of interest using the *disdrodb* software:

```
> disdrodb_download_archive --data_sources <data_sources> --campaign_names
<campaign_names>
```

DISDRODB Products Generation

- Users can generate the DISDRODB products of interest using the *disdrodb* software:

```
disdrodb_run_l
disdrodb_run_l
disdrodb_run_l
```

DISDRODB Products

DISDRODB L0

- Raw data standardized in netCDF format

DISDRODB L1

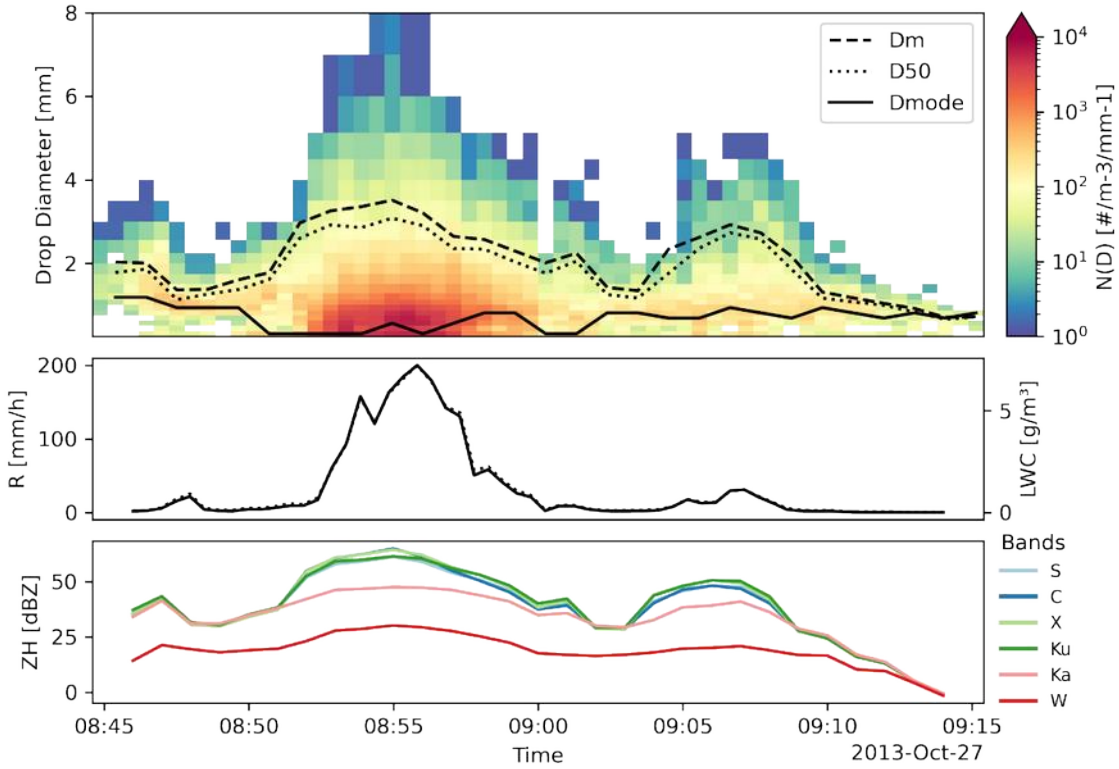
- Quality-checked and homogenized DISDRODB L0 product

DISDRODB L2

- Empirical DSD parameters (L2E)
- Model-based DSD parameters (L2M)
- Polarimetric radar variables at S, C, X, Ku, Ka and W bands.

DISDRODB L2 product

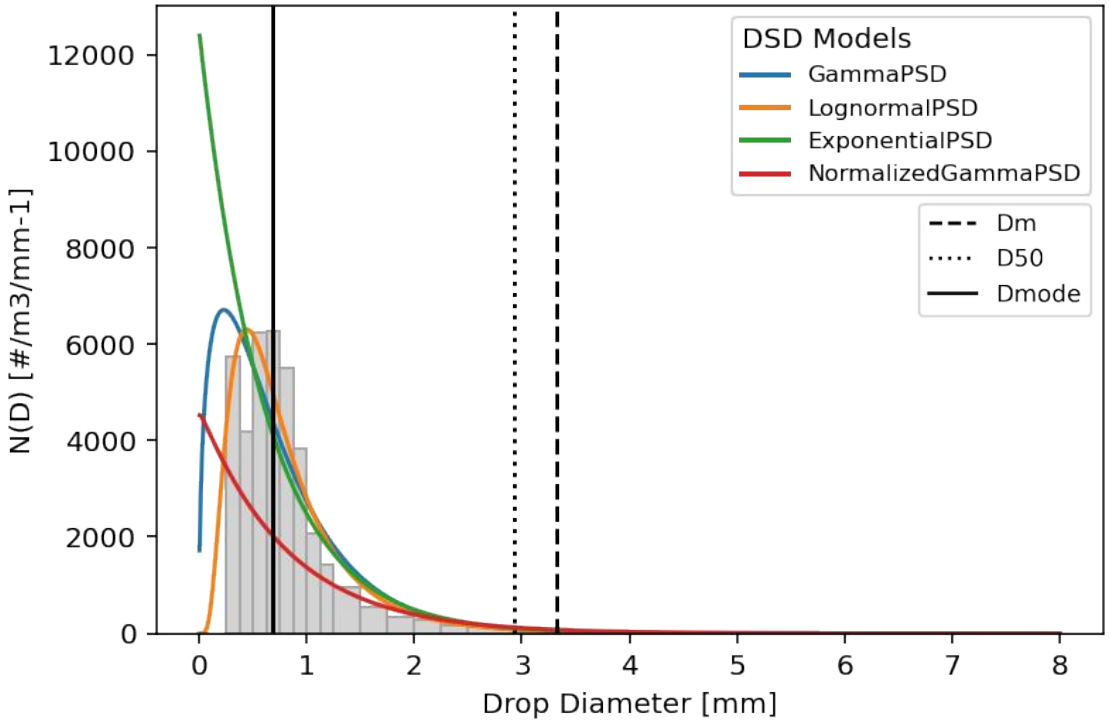
- Multiple temporal aggregations
  → …, 1MIN, 5MIN, 10MIN, …
- Optional resampling using sliding windows (ROLL)
- Polarimetric radar variables simulated using T-Matrix
- Empirical DSD variables are provided in L2E product



disdrodb_run_l2

DISDRODB L2 product

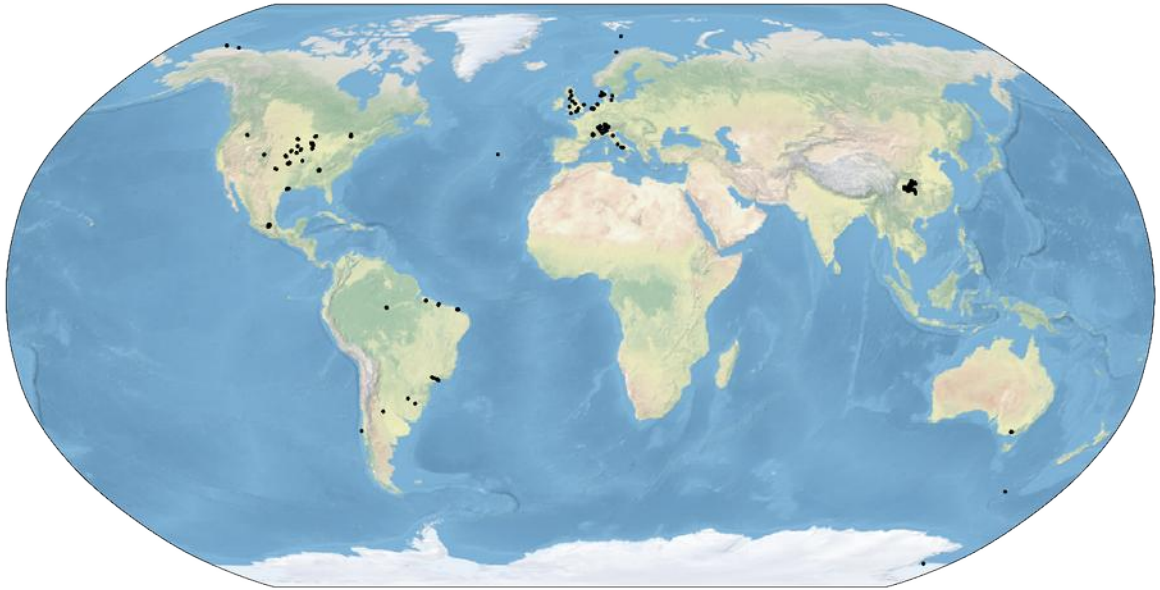
- Multiple temporal aggregations
  → …, 1MIN, 5MIN, 10MIN, …
- Optional resampling using sliding windows (ROLL)
- Polarimetric radar variables simulated using T-Matrix
- Empirical DSD variables are provided in L2E product
- Modelled DSD parameters and variables are provided in L2M product



disdrodb_run_l2

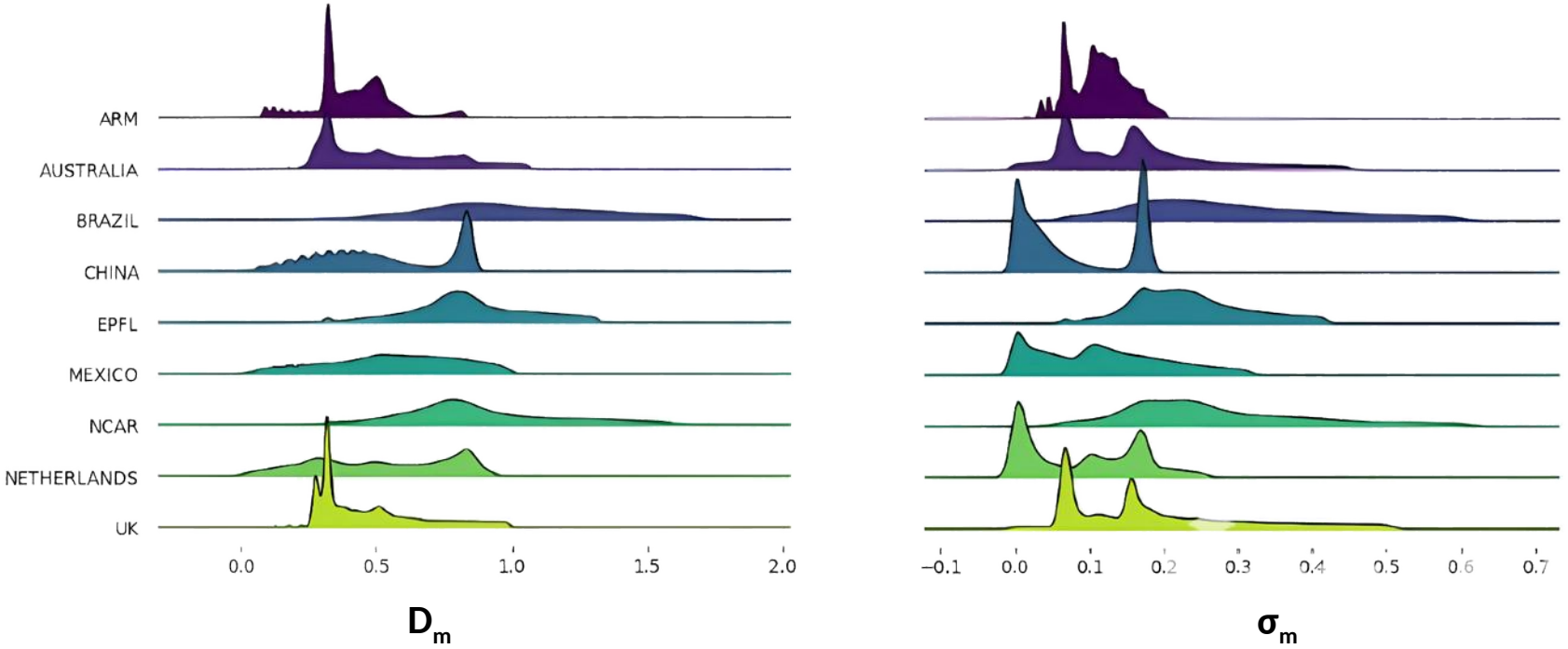
DISDRODB Global Archive

- We initiated the effort toward the creation of a global DSD archive
- Over 400 stations from 12 institutions have already been included
- We need your help to grow the database!



DISDRODB Opportunities

- The analysis-ready data format makes exploration straightforward
- Lot of DSD patterns to be revealed and analyzed



Summary

- We developed a suite of tools to standardize DSD data management and processing
- The DISDRODB metadata archive documents and keeps tracks of the disdrometer stations !
- The DISDRODB decentralized data archive allows the exchange of the raw data
  Everyone can contribute his data !
- The *disdrodb* software allows to easily download the raw data and produce the DISDRODB products
- Modular configuration files allow to customize every step of the DISDRODB processing chain
- The DISDRODB products provides standardized scientific analysis-ready data
- We initiated the effort by including > 400 stations worldwide (and additional > 300 in the making…)

- Potential to open new opportunities for global DSD studies
- Further product development and improvements are possible and encouraged !
- We look forward for feedbacks, new data and software contributors !

References

- **Metadata Archive**: https://github.com/ltelab/disdrodb-data
- **Documentation**: https://disdrodb.readthedocs.io
- **Software**: https://github.com/ltelab/disdrodb
- **Installation**: conda install disdrodb
- **Contact**: gionata.ghiggi@epfl.ch

Thanks for your attention

DISDRODB L0 product

- Raw data converted and stored into daily **standardized CF-compliant netCDF4 files**
- Standardized names for :
  - Dimensions
  - Coordinates
  - Attributes
  - Metadata

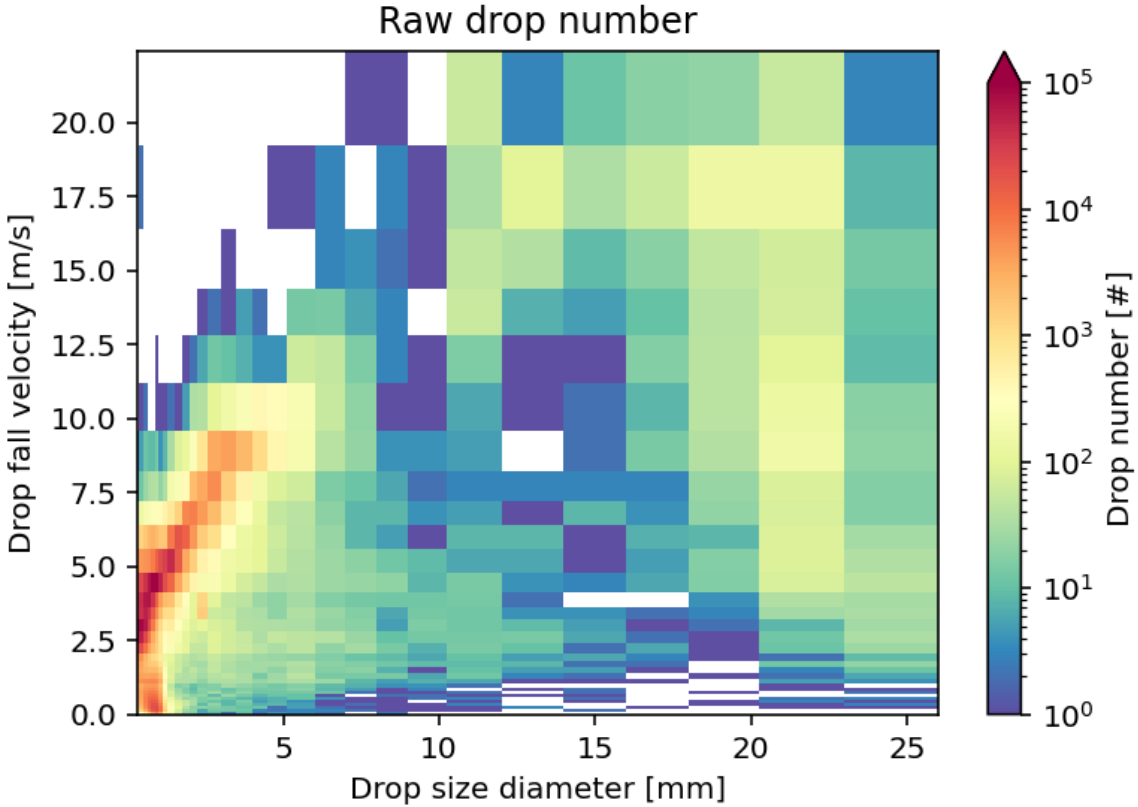
- Duplicated timesteps are removed
- No raw data is discarded or masked
- No quality check of the data is performed

```
<xarray.Dataset>
Dimensions:                          (time: 820714, diameter_bin_center: 32, velocity_bin_center:
32)
Coordinates: (12/13)
  * diameter_bin_center               (diameter_bin_center) float64 0.062 ... 24.5
    diameter_bin_lower                (diameter_bin_center) float64 ...
    diameter_bin_upper                (diameter_bin_center) float64 ...
    diameter_bin_width                (diameter_bin_center) float64 ...
  * velocity_bin_center               (velocity_bin_center) float64 0.05 ... 20.8
    velocity_bin_lower                (velocity_bin_center) float64 ...
    ...                                ...
    velocity_bin_width                (velocity_bin_center) float64 ...
  * time                              (time) datetime64[ns] 2012-08-28T08:10:00...
    crs                               object ...
    latitude                          float64 ...
    longitude                         float64 ...
    altitude                          int64 ...
Data variables: (12/16)
    raw_drop_concentration            (time, diameter_bin_center) float32 ...
    raw_drop_average_velocity         (time, velocity_bin_center) float32 ...
    raw_drop_number                   (time, diameter_bin_center, velocity_bin_center) float64 ...
    rainfall_rate_32bit               (time) float32 ...
    rainfall_accumulated_32bit        (time) float32 ...
    weather_code_synop_4680           (time) float32 ...
    ...                                ...
    number_particles                  (time) float64 ...
    sensor_temperature                (time) float32 ...
    sensor_heating_current            (time) float32 ...
    sensor_battery_voltage            (time) float32 ...
    sensor_status                     (time) float32 ...
    rainfall_amount_absolute_32bit    (time) float32 ...
Attributes: (12/60)
    campaign_name:                     HYMEX_2012
    description:                       HYdrological cycle in the Mediterranean ...
    station_name:                      Mirabel
    deployment_status:                 terminated
    deployment_mode:                   land
    platform_type:                     fixed
    ...                                ...
    authors:                           Alexis Berne, Jacopo Grazioli
    authors_url:                       Alexis Berne, Jacopo Grazioli
    contact:                           alexis.berne@epfl.ch
    contact_information:               http://lte.epfl.ch
    institution:                       Laboratoire de Teledetection Environneme...
    website:                           https://hymex.org/
```

disdrodb_run_l

DISDRODB L1 product

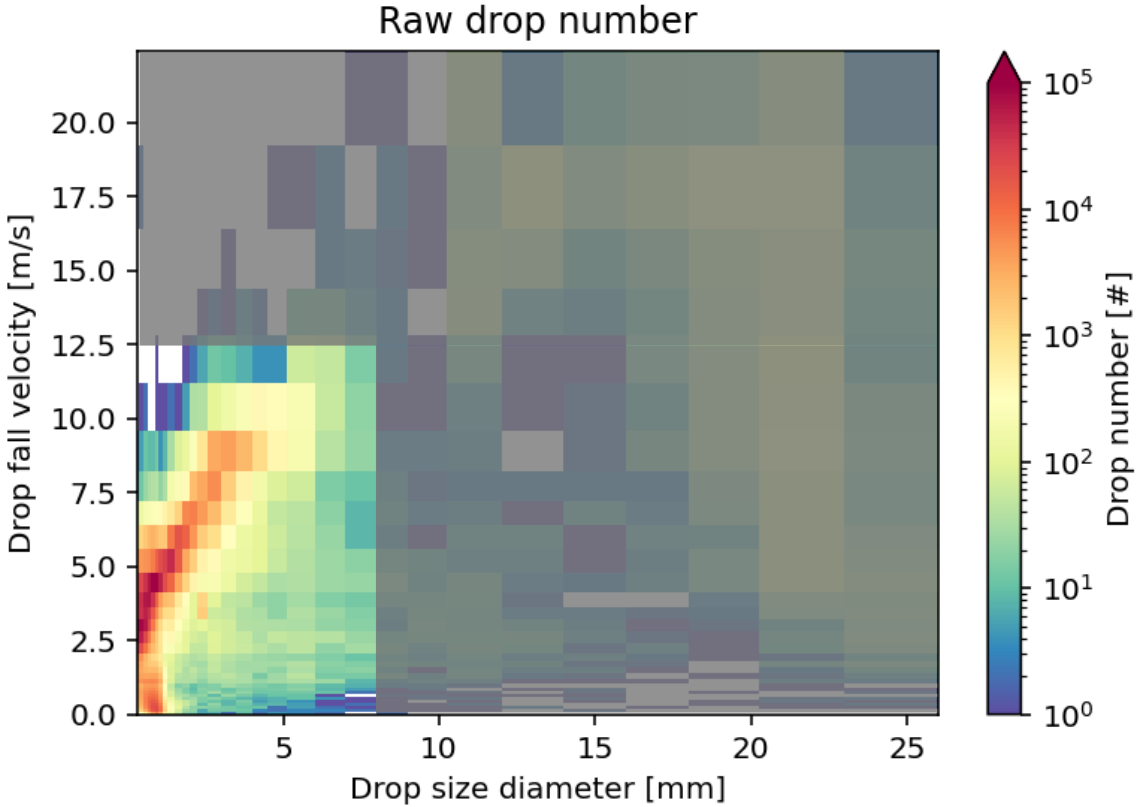
- Filtered DSD spectrum
- A configuration file allows to customize the filtering criteria and the techniques for estimating drop fall velocity



disdrodb_run_l

DISDRODB L1 product

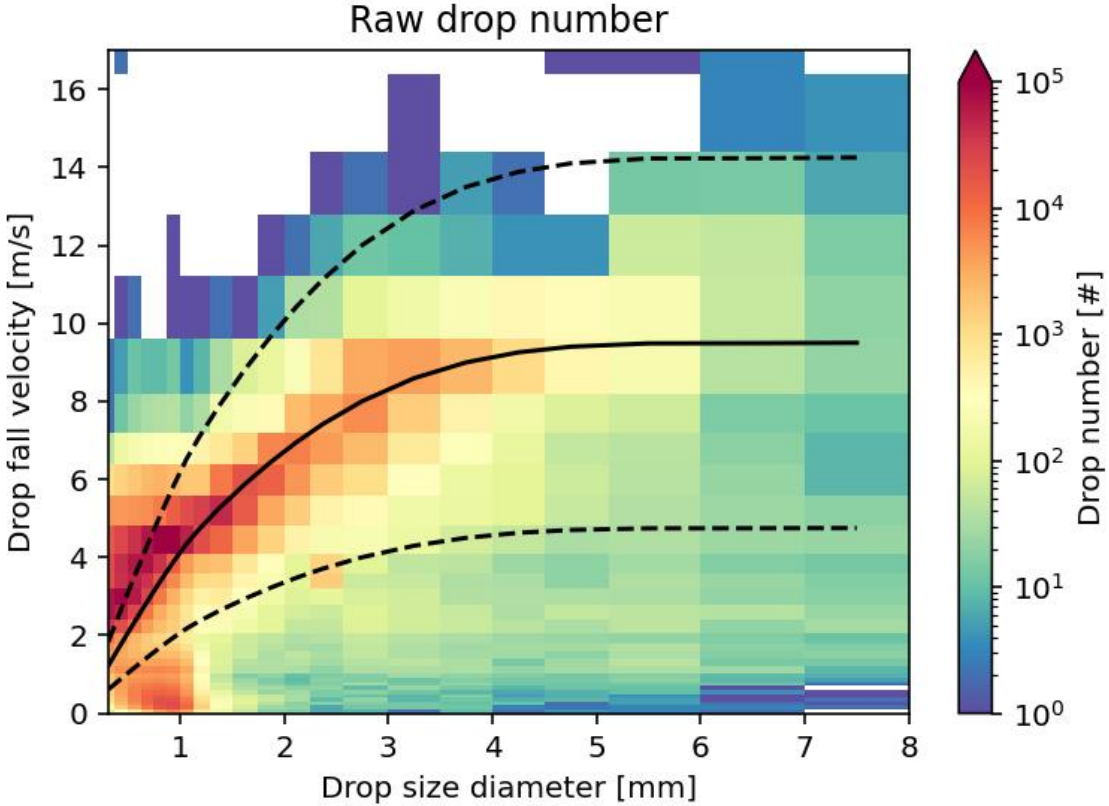
- Filtered DSD spectrum
- A configuration file allows to customize the filtering criteria and the techniques for estimating drop fall velocity



disdrodb_run_l

DISDRODB L1 product

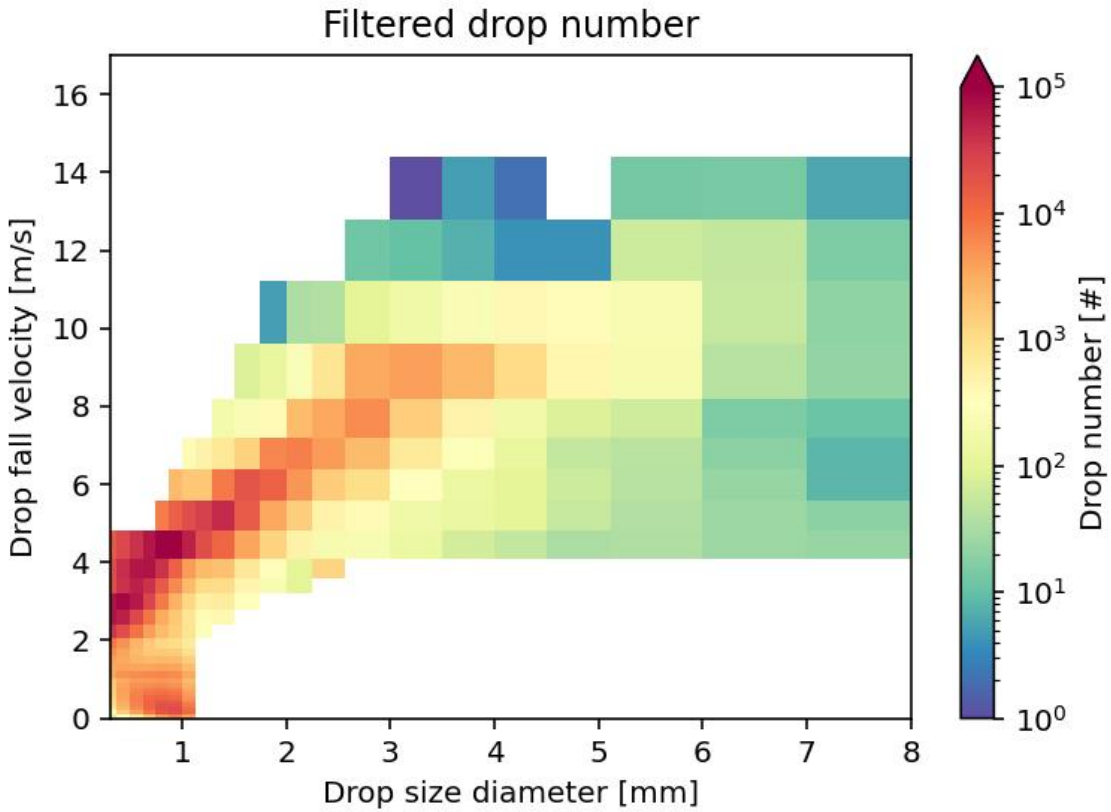
- Filtered DSD spectrum
- A configuration file allows to customize the filtering criteria and the techniques for estimating drop fall velocity



disdrodb_run_l

DISDRODB L1 product

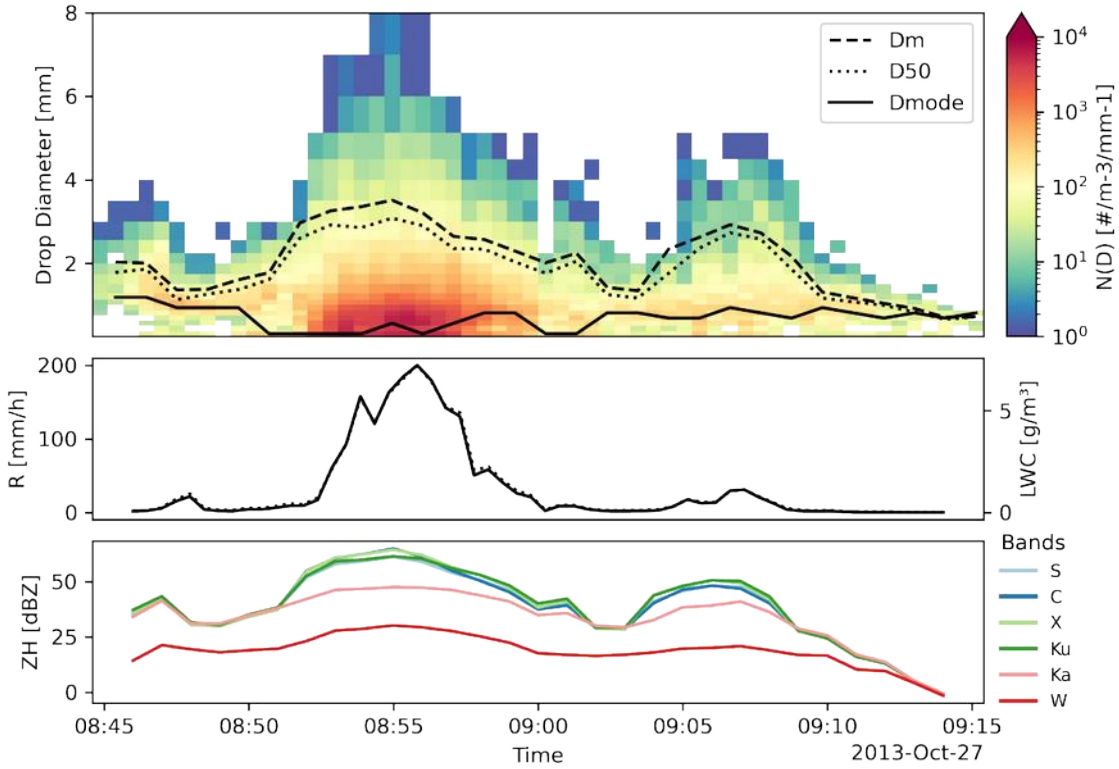
- Filtered DSD spectrum
- A configuration file allows to customize the filtering criteria and the techniques for estimating drop fall velocity



disdrodb_run_l

DISDRODB L2 product

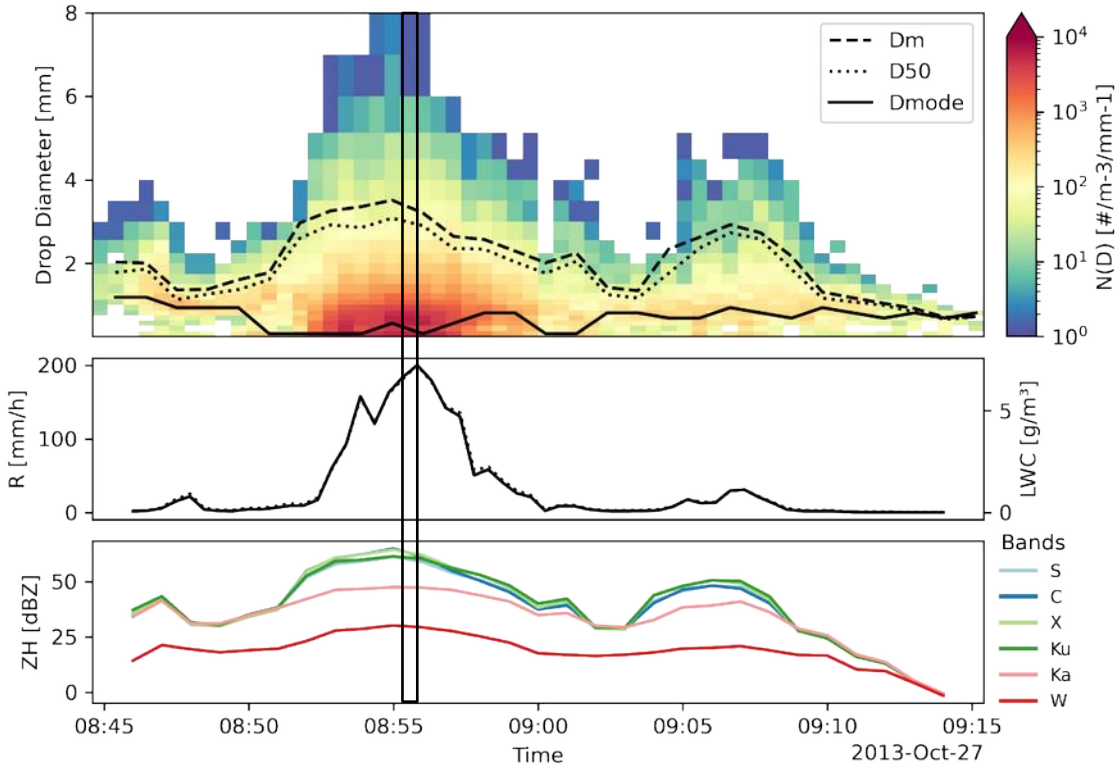
- Multiple temporal aggregations
  → …, 1MIN, 5MIN, 10MIN, …
- Optional resampling using sliding windows (ROLL)
- Polarimetric radar variables simulated using T-Matrix
- Empirical DSD variables are provided in L2E product



disdrodb_run_l2

DISDRODB L2 product

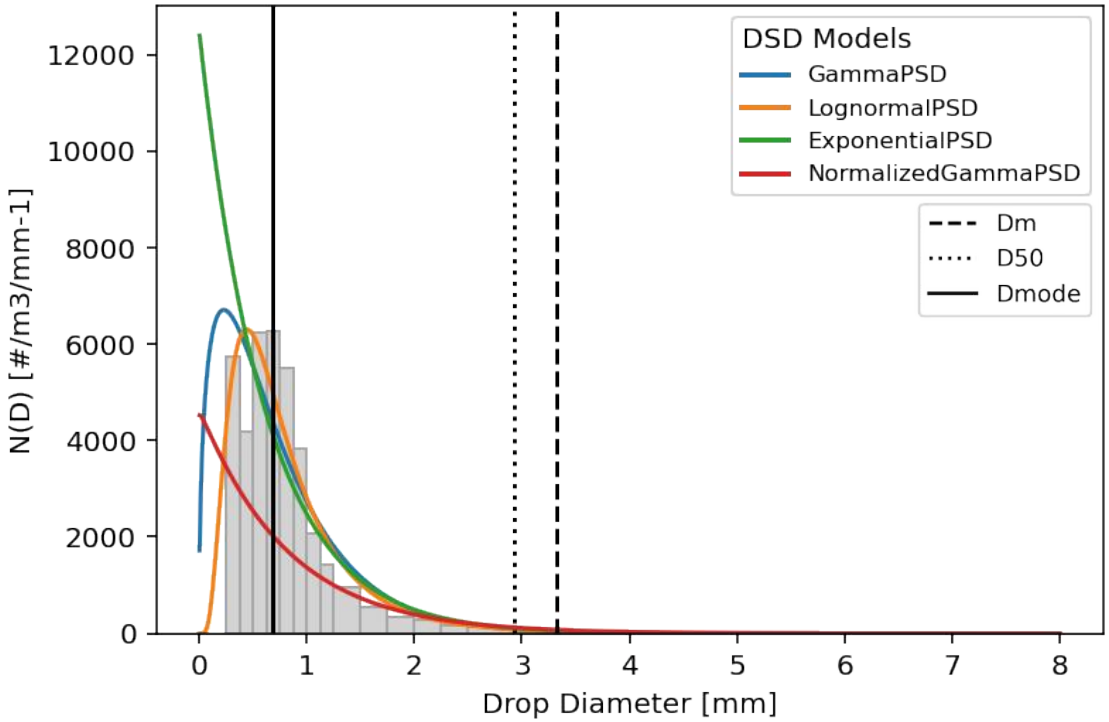
- Multiple temporal aggregations
  → …, 1MIN, 5MIN, 10MIN, …
- Optional resampling using sliding windows (ROLL)
- Polarimetric radar variables simulated using T-Matrix
- Empirical DSD variables are provided in L2E product



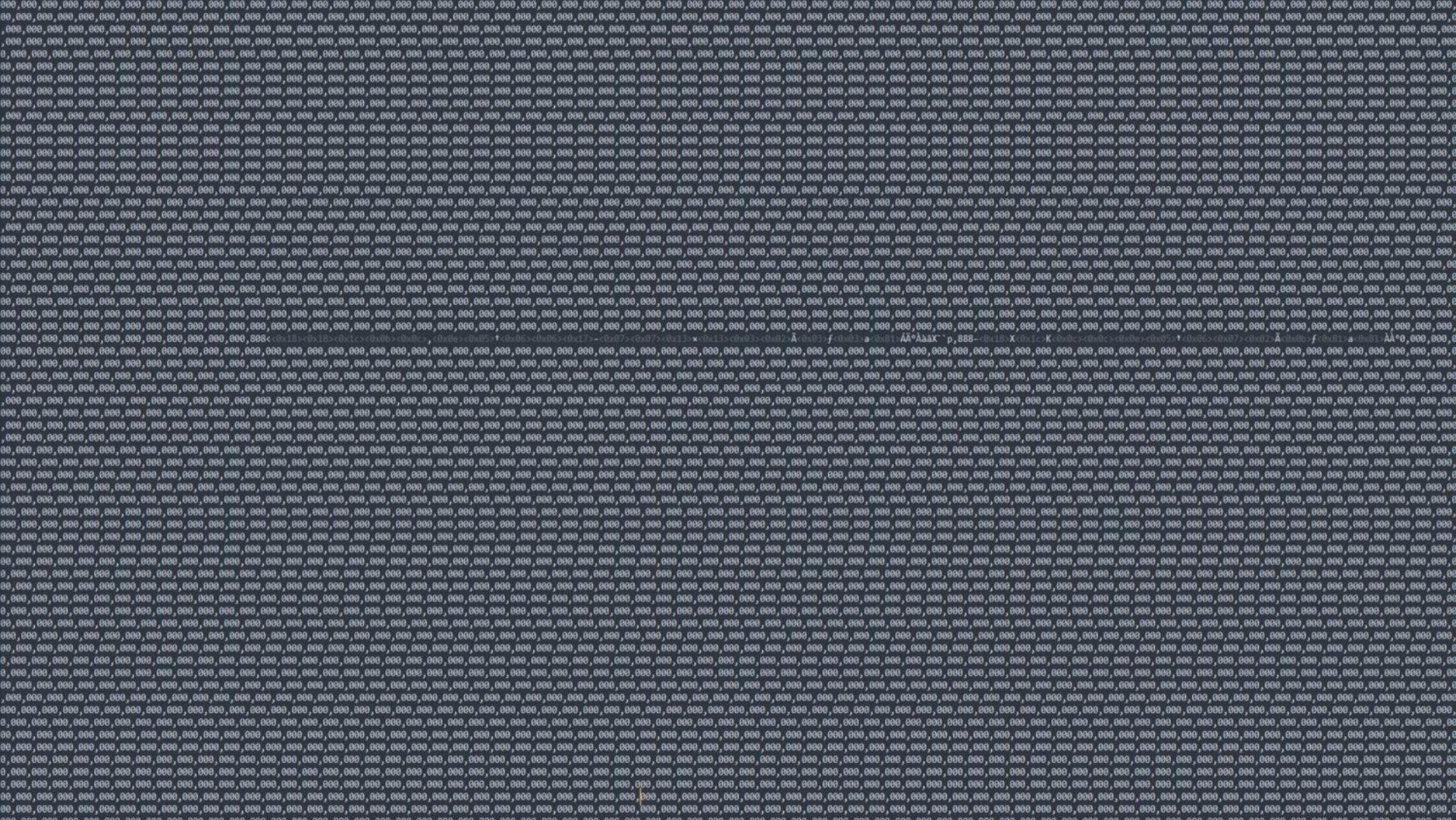
disdrodb_run_l2

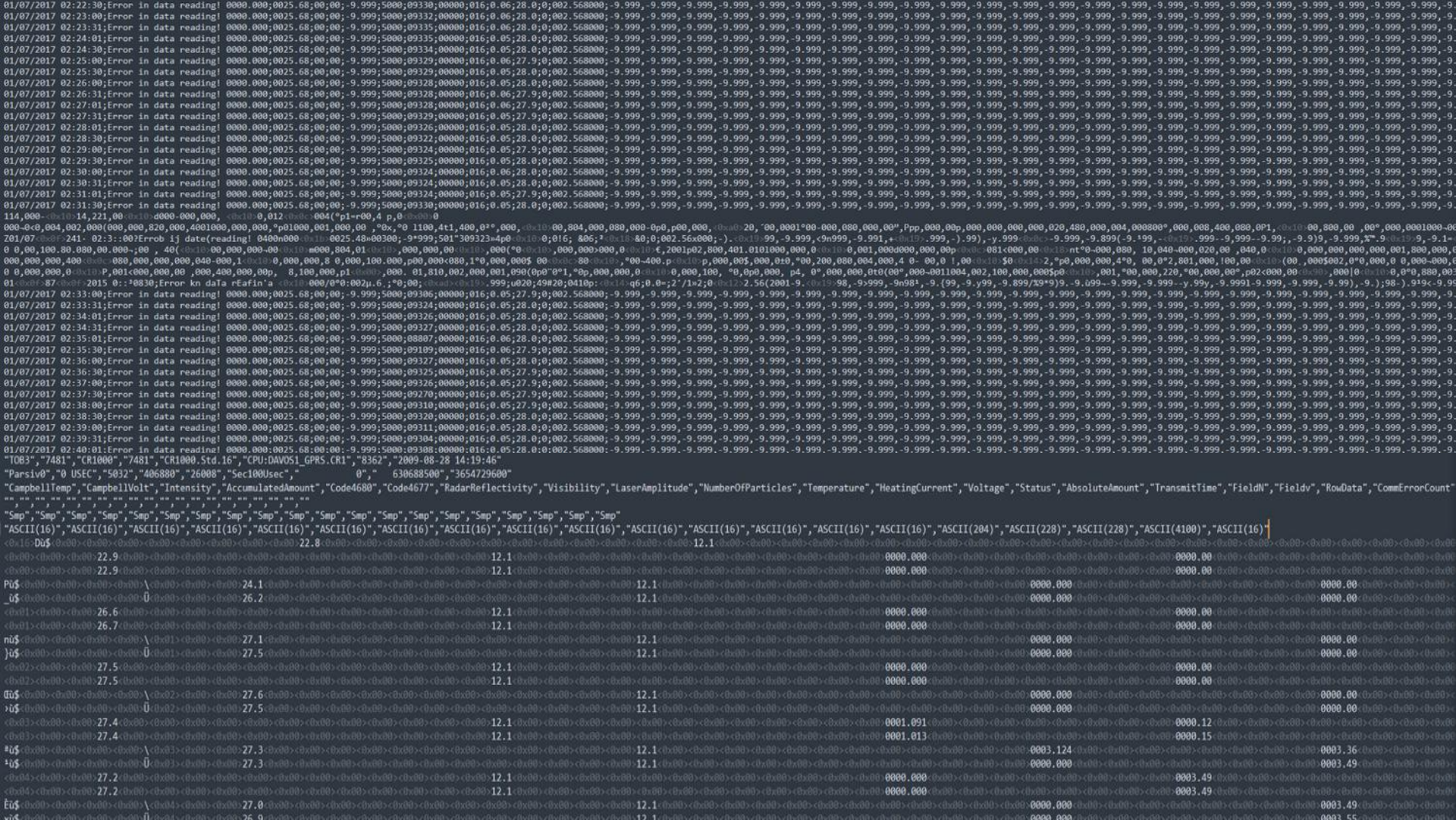
DISDRODB L2 product

- Multiple temporal aggregations
  → …, 1MIN, 5MIN, 10MIN, …
- Optional resampling using sliding windows (ROLL)
- Polarimetric radar variables simulated using T-Matrix
- Empirical DSD variables are provided in L2E product
- Modelled DSD parameters and variables are provided in L2M product

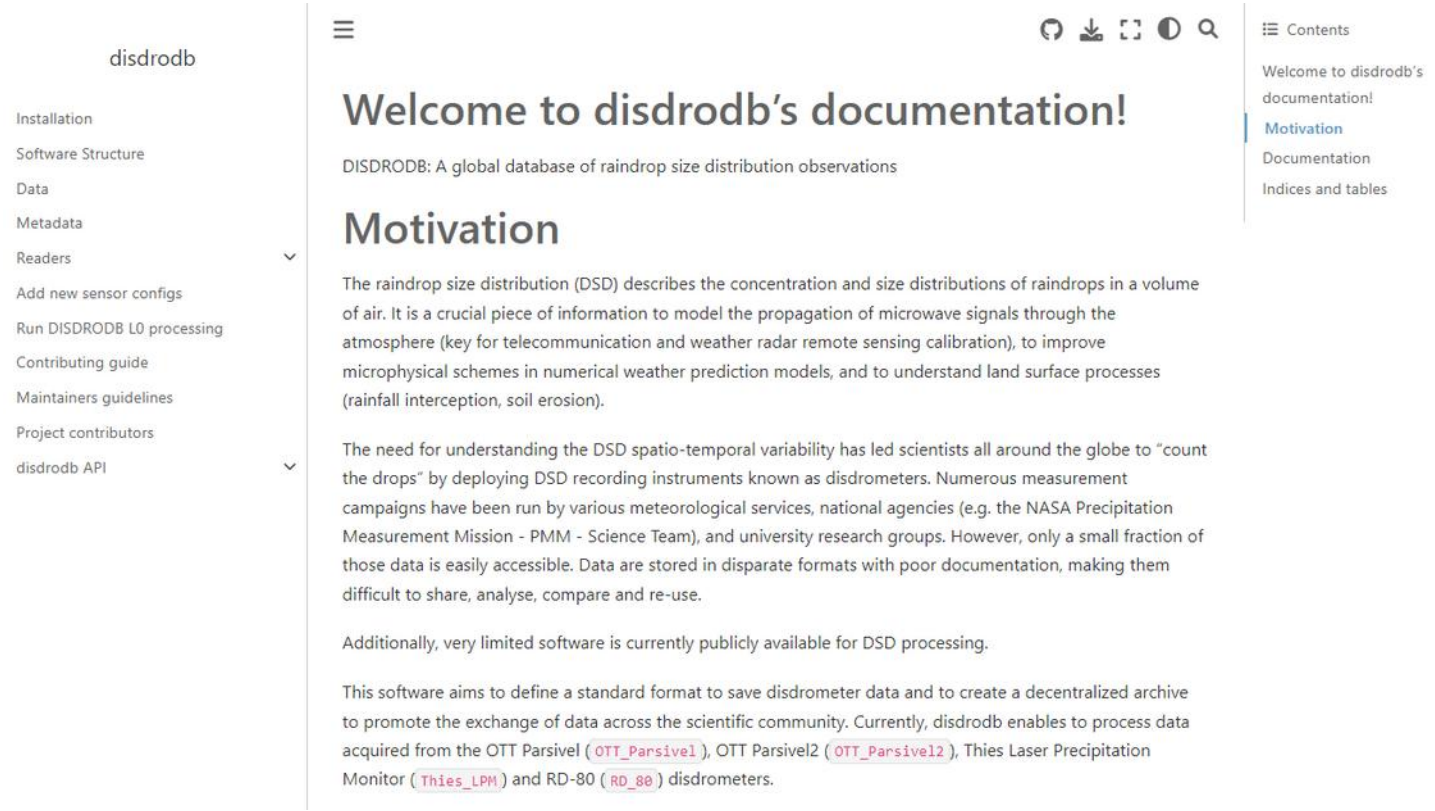


disdrodb_run_l2





DISDRODB - Documentation



DISDRODB – Data Contribution System

1) Add your station information to the DISDRODB Metadata Archive
2) Design the reader for your data and update the metadata **disdrodb_reader** field
3) Ensure that your metadata and reader pass the DISDRODB Quality Screening Check
4) Upload your data to a remote repository and update the metadata **disdrodb_data_url** field

DISDRODB L0 products

DISDRODB L0A

- Conversion of raw text-format data into Apache Parquet binary format.
- Each input file produces one output file.

DISDRODB L0B

- Conversion of raw netCDF-format data into DISDRODB L0B netCDF format.
- Conversion of DISDRODB L0A files into DISDRODB L0B netCDF format.
- Each input file produces one output file.

DISDRODB L0C

- Data homogenization products.
- DISDRODB L0B files are grouped into daily files.
- Duplicate timestamps are removed, and timestamps are adjusted to correct for sensor offsets (e.g., 00:02 becomes 00:00).
- A quality indicator is associated with each time step to indicate the availability of neighboring data, time adjustments, etc.
- For sensors with variable sampling intervals, a separate file is generated for each interval, ensuring correct resampling for L2 products.

DISDRODB L1 product

DISDRODB L1

- Provides the filtered drop size distribution (DSD) spectrum.
- For each DISDRODB L0C file, a DISDRODB L1 file is generated.
- A configuration file allows customization of filtering criteria and techniques for estimating drop fall velocity.
- All calculations are vectorized and parallelized.

DISDRODB L2 products

DISDRODB L2E

- Provides empirical DSD parameters and polarimetric radar variables at W, Ka, Ku, X, C, and S simulated using T-Matrix
- All calculations and radar simulations are vectorized and parallelized.
- A file is generated for each "event."
- An event is defined by a *start_time* and *end_time*.
- Users can choose to define an event as a group of consecutive days, months, or years (or the entire dataset) or, alternatively, based on configurable criteria that identify rain events.
- The processing chain allows multiple temporal aggregations (1MIN, 5MIN, etc.) and resampling using sliding windows (e.g., "ROLL1MIN").

DISDRODB L2M

- Estimates PSD parameters and subsequently computes integrated DSD variables and polarimetric radar variables.
- Various PSD models and parameter estimation methods available.
- A L2M file is generated for each L2E file.

DISDRODB – PSD models and parameter estimation

PSD models

- ExponentialPSD
- GammaPSD
- NormalizedGammaPSD
- LognormalPSD

Parameter estimation methods

ML (Maximum Likelihood)

✓ Options for likelihood: Poisson vs Multinomial
✓ Options for use of Truncated Likelihood.
✓ Options to use MOM estimates as initial values for ML optimization.
✓ Options for numerical optimization algorithms

MOM (Method of Moments)

✓ Multiple formulations

GS (Grid Search)

✓ Minimizes a user-defined cost function with targets like ND, R, LWC, or Z, using different error metrics (i.e. MAE, MSE, …).
✓ Options to fix parameters (i.e. μ of NormalizedGammaPSD)