Algorithm Theoretical Basis Document for Version 1.0 of the WegenerNet 3D Observing System L1 and L2 Data Products

WegenerNet Technical Report No. 2/2024

Andreas Kvas, Jürgen Fuchsberger, Gottfried Kirchengast

December 2024

Wegener Center for Climate and Global Change (WEGC), University of Graz, Austria

Contents

1. Introduction

This document presents the theoretical basis behind the algorithms (ATBD) utilized in the generation of the WegenerNet 3D Observing System (WEGN3D) v1.0 data cubes. It is supplementary to the v1.0 release notes document[1] which provides a more general high-level overview of the data products and is aimed towards geophysical and atmospheric science applications. The algorithms and processing flows described here apply to the Level 1b data (L1b) with the Digital Object Identifier (DOI) 10.25364/WEGC/WPS3D-L1b-10 and the Level 2 (L2) datasets with the DOI 10.25364/WEGC/WPS3D-L2-10.

The ATBD is structured as follows: section 2 provides an overview of the WegenerNet 3D Processing System (WPS3D). The versioning scheme of the data products is given in section 3, while section 4 details the WegenerNet 3D Quality Control System (QCS3D), which implements consistency checks, outlier detection, and time/space resampling. In section 5 the WegenerNet 3D Data Product Generator (DPG3D), responsible for generating georeferenced, gridded, and gap-filled data cubes is described. Finally, section 6 outlines the strategy to georeference observations and transform between different coordinate frames.

2. WegenerNet 3D Processing System (WPS3D)

The WegenerNet 3D Processing System (WPS3D), is an extension of the WegenerNet Processing System (WPS, Fuchsberger, Kirchengast, and Kabas 2021), and forms the foundation for the generation of the WegenerNet 3D Observing System (WEGN3D) data cubes. In analogy to the WPS, WPS3D is split into different application parts which will be outlined in the following paragraphs.

Command Receive Archiving System (CRAS)

The Command Receive Archiving System (CRAS), shared with the WPS, is responsible for retrieving raw data from the different WegenerNet 3D sensors, and saves these sensor- and vendor-specific data files in the WegenerNet storage infrastructure as L0 data. As per definition of the L0 data, no processing is applied in this step.

WegenerNet 3D Quality Control System (QCS3D)

After acquisition of the L0 data via the CRAS, the WegenerNet 3D Quality Control System (QCS3D) transforms the sensor- and vendor specific data file formats into CF-compliant NetCDF files following the WEGN3D data model. The WEGN3D data model comprises definitions and metadata for all physical quantities observed in the WegenerNet 3D Open-Air Laboratory, derived quantities, and auxiliary data such as housekeeping flags and quality flags and their interrelations. This encompasses properties such as units, climate bounds, human-readable descriptions, and application specific parameters related to storage, encoding, and compression. It follows the CF metadata conventions for all public-facing interfaces where application-related parameters are typically hidden.

[1] https://wegenernet.org/downloads/Kvas_et_al_2024_WEGN3D_v1_release_notes-WEGN-TR-1-2024.pdf

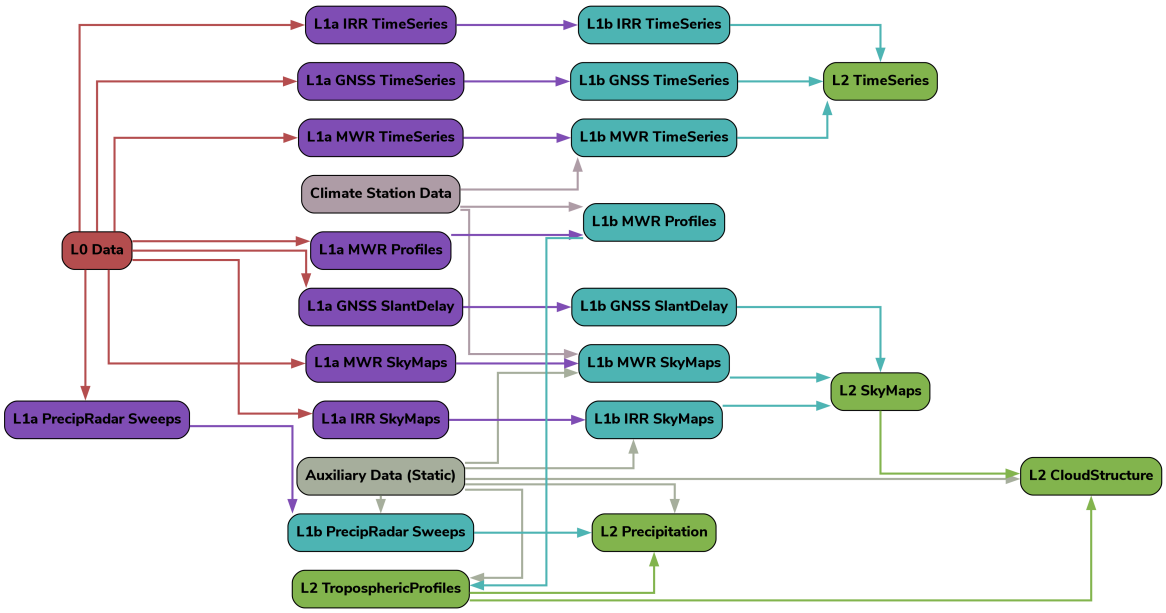

Figure 1: WegenerNet 3D Processing System (WPS3D) data flow.

These L1a files are then further processed, by aggregating and resampling L1a data into time series with regular sampling, and regular spatial coordinates. For these L1b data files QCS3D further derives quality flags, which are attached as additional data variables.

WegenerNet 3D Data Product Generator (DPG3D)

The WegenerNet 3D Data Product Generator (DPG3D) processes these quality-flagged files to compute L2 products, which are gap-filled and geolocated data cubes. L2 data are generally multisensor products, which combine multiple L1b data files to derive new added-value data variables.

WPS3D Data Flow

The data flow within WPS3D is shown in Figure 1. It starts with L0 data collected by the CRAS, which, together with auxiliary static data and WegenerNet climate station data are transformed into L1b data by the QCS3D, with L1a data products as an intermediate step. L1a data products are typically not stored to reduce the storage footprint. L1b are then used to generate L2 data in the DPG3D, again using auxiliary static data. Auxiliary static data comprises datasets used in transformations between different reference frames, such as geoid models and digital elevation models (see subsection 6.2). Auxiliary static data also includes pre-trained classification models for use in hydrometeor classification and neural network weights for statistical retrieval of various atmospheric state variables from microwave brightness temperatures. While the data flow is generally hierarchical, there are also intraconnections within L2 data present, such as the use of tropospheric profiles in the generation of precipitation data products. These intraconnections are detailed for each data product in the corresponding processing description in section 5.

3. Data Product Versioning

Version numbering of these data collections adheres the semantic versioning guidelines[2], albeit with adaptions tailored to data products. The primary deviation from the semantic versioning guidelines is that version numbers within WEGN3D are restricted to major and minor versions, with the possibility to add descriptive labels for intermediate or experimental data product versions. Thus, all main data product releases that have been assigned a Digital Object Identifier carry a version number `MAJOR.MINOR` which may be prefixed with a `v` to be easily identifiable, for example, `v1.0`. Special data releases, such as experimental processing chains or release candidates for evaluation and testing, are suffixed with a hyphen, followed by a descriptive identifier, as in `v2.0-rc1`.

How the version number is incremented depends on the type of changes introduced into the processing chain. In general, small changes that do not change the meaning and content of output variables may be introduced to the WPS3D during operational processing without a version increment. Historic data is not changed when such small changes are introduced. To ensure traceability in this case, the software version with which each data file is created is stored in its metadata via the corresponding Git commit hash.

The minor version of a dataset is incremented when changes not related to the data model are introduced. This means that the values of variables change compared to the previous version while the data structure itself remains the same. This typically is a result of changes and improvements in the processing chain of existing data variables. A new minor version is expected to be a drop-in replacement for previous versions of the same major version release and not require the adaption of downstream software processing chains.

The major version of a dataset is incremented when backwards incompatible changes in a data product are introduced. Such changes may include different coordinate variable values, changes to variable or product names, or changes in variable units. A new major version is also expected to require software adaptions of downstream processing chains.

Any minor or major version increment leads to a reprocessing of the whole time series to ensure a consistent data record.

The WN3D version string can be parsed by the following regular expression:

```
^v?(?P<major>0|[1-9]\d*)\.(?P<minor>0|[1-9]\d*)
(?:-(?P<prerelease>(?:0|[1-9]\d*|\d*[a-zA-Z-][0-9a-zA-Z-]*)
(?:(?:0|[1-9]\d*|\d*[a-zA-Z-][0-9a-zA-Z-]*))*))?\$
```

In case the data version is included in a file name, it is converted into a string according to

```
# get regex groups <major>, <minor>, <prerelease>
# from semantic version string

version_string = 'v' + <major>
if <minor> != 0:
    version_string += 'm' + <minor>
if <prerelease> != NULL:
    version_string += <prerelease>
```

which, for example, converts `v2.0-rc1` into `v2rc1`.

[2] https://semver.org, retrieved 2024-09-02

4. QCS3D L1b Processing

4.1. Generic QCS3D Checks

The WegenerNet 3D Quality Control System (QCS3D) performs a number of quality checks denoted *layers* to identify and flag outliers and other spurious data values. The output of the QCS3D is represented as a bitmask for each data variable <variable_name>_qcs_flag and added to each of the L1b data products. Version 1.0 of QCS3D supports 8 layers, thus the bitmask is represented by an 8-bit integer. If the checks within one layer detect spurious data, the corresponding bit is set Fuchsberger, Kirchengast, and Kabas 2021.

Due to the variety of the the WEGN3D sensor infrastructures, not all layers and quality check algorithm can be applied to all input data in the same fashion. Thus, checks are implemented to be as generic and flexible as possible and specialized to specific sensors when needed. The following list describes the algorithms for the generic QCS3D layer implementation.

- **Layer 0: Operations Check**

  Checks whether the sensor is operational (i.e. online) for a given time stamp t and not manually shut down or in maintenance. Manual sensor shutdown events are tracked and documented, thus a simple check whether t lies within an offline time period is performed.

- **Layer 1: Availability Check**

  Checks whether the sensor delivers data as expected. If the sensor is not manually shut down for a given time stamp t, but no or malformed data is received, bit 1 is set. Since all L1b products do have a well-defined temporal base sampling, this is checked by counting the non-`NaN` data values in each time interval.

- **Layer 2: Sensor Bounds Check**

  Checks whether the sensor values lies within the sensor specification. Sensor bounds are typically defined within the sensor specification documentation. Where no such information is available, physical bounds are taken instead. Sensor bounds are uniquely defined for each sensor and data variable and are time invariant. The corresponding check tests whether a data value is within the sensor bounds.

  For elevation-dependent quantities which are affected by the path length through the atmosphere in direction of the measurement ray, sensor bounds are defined in zenith direction and mapped to the actual elevation. This mapping is done via a sinusoidal function

  $$b_{\text{slant}} = \frac{b_{\text{zenith}}}{\sin \epsilon}. \tag{1}$$

- **Layer 3: Climate Bounds Check**

  Checks whether the sensor values lies within physically sensible values for the WegenerNet Feldbach region at the given day of year. Climate bounds are derived by analyzing long time series of model output and in some cases observations. To avoid accidentally flagging extreme events, a buffer of 15 % is added (or reduced) to (from) the upper (lower) bounds. The resulting bound values are defined for each data variable and are sensor independent. Elevation dependent quantities are treated in the same way as described in the sensor bounds check.

- **Layer 4: Variability Check**

  Checks whether the spatial and temporal variability of the sensor values lie below an empirically determined threshold. The temporal variability in this context is defined as the absolute value of the temporal gradient which is compared to a predefined threshold. This detects and flags spurious jumps in the data. Additionally, temporal variability is checked by comparing each data point to the median of its neighborhood. This moving window check facilitates detecting consecutive outliers where the gradient check would fail. For multidimensional data cubes, for example, all-sky scans, all temporal variability checks are performed on each data value individually.

  Spatial variability is checked by comparing each point of the spatial field for each time stamp to the median of surrounding pixels. Similar to the temporal variability check, this identifies large localized outliers. The size of the median filter is chosen based on the expected spatial variability and adapted for each data variable.

- **Layer 5: Intrastation Check**

  Checks whether values of a multi-channel sensor a physically consistent or are not already flagged by the sensor software.

- **Layer 6: Interstation Check**

  Checks whether values of spatially distributed sensors are physically consistent. The check is realized through computing robust anomalies by subtracting the median computed over all distributed sensor locations from the measurement values, taking the absolute value, and comparing the result against a predefined threshold.

- **Layer 7: Reference Check**

  Checks the sensor values against external independent reference values. This layer is currently unused, because currently no external reference dataset with sufficient spatial and temporal resolution for comparison exists.

4.2. WN3D_L1b_v1_GNSS_TimeSeries

WN3D_L1b_v1_GNSS_TimeSeries	
Dependencies	-
Input Files	GNSS TROPOSINEX Files (**.zpd*) GNSS TROPOSINEX Files (**.tro*)
Output	WN3D_L1b_v1_GNSS_TimeSeries (NetCDF)

- Read GNSS L0 TROPOSINEX files, apply WEGN3D data model, and convert to L1a

  Data variables read in this step are tropospheric path delay in zenith direction (apd_nh_zenith, apd_h_zenith), integrated water vapor (iwv_zenith), and total tropospheric gradients (tropograd_tot_north, tropograd_tot_east)

- Apply QCS3D Layers to flag suspicious data

  No non-standard QCS3D checks are applied to the data.

- Append long-term averages of the GNSS StarNet station positions are to the data product for geolocation

Data variables containing the geographic station longitude, latitude, and ellipsoidal_height are added to the data product.

4.3. WN3D_L1b_v1_GNSS_SlantDelay

WN3D_L1b_v1_GNSS_SlantDelay	
Dependencies	-
Input Files	Raw GNSS slant zenith time series files (L0)
Output	WN3D_L1b_v1_GNSS_SlantDelay (NetCDF)

- Read GNSS L0 Slant Delay SINEX files, apply WEGN3D data model and convert to L1a

  The slant delay values are converted to stored as time, station_number, azimuth data cubes. The observation geometry (i.e. the direction of the measurement ray between the station/satellite pairs) is stored in two additional data variables, namely elevation and azimuth.

  In addition to slant delay values, mapped total delay in zenith direction (apd_tot_zenith) for each slant delay value is also added to the **WN3D_L1b_v1_GNSS_SlantDelay** data product.

- Apply QCS3D Layer to flag suspicious data

  QCS3D checks are performed on the apd_tot_zenith values.

- Append long-term averages of the GNSS StarNet station positions are to the data product for geolocation

  Data variables containing the geographic station longitude, latitude, and ellipsoidal_height are added to the data product.

4.4. WN3D_L1b_v1_IRR_TimeSeries

WN3D_L1b_v1_IRR_TimeSeries	
Dependencies	-
Input Files	Raw IRR zenith time series files (L0)
Output	WN3D_L1b_v1_IRR_TimeSeries (NetCDF)

- Read IRR L0 zenith time series data, apply WEGN3D data model and convert to L1a

  Variables read in this step are infrared brightness temperature T_ir_zenith in the 8 µm - 14 µm band, and cloud_binary_mask which is produced by the onboard sensor software.

- Aggregate data to a 10 min time series

- Apply QCS3D Layers to flag suspicious data values

  No special QCS3D are applied to **WN3D_L1b_v1_IRR_TimeSeries**

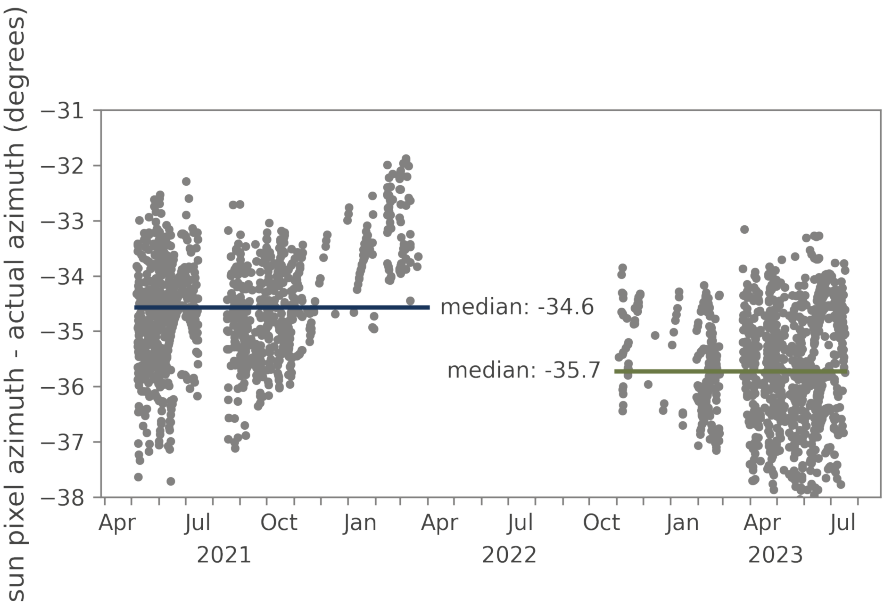

Figure 2: Azimuth offset determined by comparison with brightness temperature spikes and sun position.

4.5. WN3D_L1b_v1_IRR_SkyMaps

WN3D_L1b_v1_IRR_SkyMaps	
Dependencies	-
Input Files	Raw IRR all-sky scan data files (L0)
Additional input	Azimuth offset data file IRR clutter map
Output	WN3D_L1b_v1_IRR_SkyMaps (NetCDF)

- Read IRR L0 all-sky scan data, apply WEGN3D data model and convert to L1a

  Target variables read in this step are infrared brightness temperature T_ir in the 8 µm - 14 µm band, and cloud_type produced by the onboard sensor software.

- Apply azimuth offset $\Delta\alpha$ to properly orient SkyMaps (cf. Figure 2)

$$\alpha_{\text{L1b}} = \alpha_{\text{L1a}} - \Delta\alpha \tag{2}$$

- Resample all-sky scans to a regular 10 min time series

  This leads to the assumption, the one full scan is representative its assigned 10 min window. No aggregation is required, because one all-sky scan takes about 7 min.

- Apply QCS3D Layers to flag suspicious data values

  Specialized QCS3D checks for **WN3D_L1b_v1_IRR_SkyMaps** include:

  - Layer 5: Apply an elevation/azimuth clutter map to mask buildings in low elevations
  - Layer 5: Compute the current sun position and mask corresponding pixels

4.6. WN3D_L1b_v1_MWR_TimeSeries

WN3D_L1b_v1_MWR_TimeSeries	
Dependencies	-
Input Files	Raw MWR data files (L0)
Additional Input	WegenerNet Climate Station Network data (station 156) MWR housekeeping Data
Output	WN3D_L1b_v1_MWR_TimeSeries (NetCDF)

- Read MWR raw data files (L0), apply WEGN3D data model and convert to L1a
- Read MWR housekeeping data and generate quality flags for each 10 min interval

  The quality flag is extended 30 min after precipitation is detected at the radiometer site to account for the drying time of the radiometer membrane.
- Apply QCS3D Layers to flag suspicious data values

  No special QCS3D are applied to **WN3D_L1b_v1_MWR_TimeSeries**.

4.7. WN3D_L1b_v1_MWR_Profiles

WN3D_L1b_v1_MWR_Profiles	
Dependencies	-
Input Files	Raw MWR data files (L0)
Additional Input	WegenerNet Climate Station Network data (station 156) MWR housekeeping Data
Output	WN3D_L1b_v1_MWR_Profiles (NetCDF)

- Read MWR L0 profile data, apply WEGN3D data model and convert to L1a
- Read MWR housekeeping data and generate quality flags for each 10 min interval

  The quality flag is extended 30 min after precipitation is detected at the radiometer site to account for the drying time of the radiometer membrane.
- Aggregate data to a 10 min time series
- Apply QCS3D Layers to flag suspicious data values

  No special QCS3D are applied to **WN3D_L1b_v1_MWR_Profiles**.
- Compute barometric pressure profile P_profile

  The barometric pressure profile is computed from the density of humid air

$$\rho_H = \frac{p_d}{R_d T} + \frac{e}{R_v T}, \tag{3}$$

  which depends on the partial pressure of dry air p_d, the air temperature T in Kelvin, the specific gas constant of dry air R_d and water vapor R_v, and the pressure of water vapor e. The pressure of water vapor e can be computed from relative humidity R in percent and the saturation vapor pressure e_s via

$$e = \frac{R}{100} e_s. \tag{4}$$

The saturation vapor pressure e_s is computed using the formula of Bolton 1980. To compute the partial pressure of dry air p_d, the previously computed vapor pressure e is subtracted from the total pressure P as in

$$p_d = P - e. \tag{5}$$

This introduces a circular dependency on P, so the air density needs to be computed iteratively with an initial guess for P. Assuming hydrostatic pressure, the computed density can simply be integrated along the vertical column with

$$P = P_0 + \int_{h_0}^{h} -\rho_H g dz, \tag{6}$$

with the initial value P_0 taken from the surface air pressure measurements. To simplify the integral, $g\ dz$ is substituted with the infinitesimal geopotential change dV.

- The QCS3D flags for P_profile is derived from the quality flags of the input profiles (temperature and relative humidity) by flagging all data points where either of the input quantities is flagged

4.8. WN3D_L1b_v1_MWR_SkyMaps

WN3D_L1b_v1_MWR_SkyMaps	
Dependencies	-
Input Files	Raw MWR data files (L0)
Additional Input	WegenerNet Climate Station Network data (station 156) MWR housekeeping Data MWR clutter map
Output	WN3D_L1b_v1_MWR_SkyMaps (NetCDF)

- Read MWR L0 scan data, apply WEGN3D data model and convert to L1a
- Read MWR housekeeping data and generate quality flags for each 10 min interval

  The quality flag is extended 30 min after precipitation is detected at the radiometer site to account for the drying time of the radiometer membrane.
- Resample all-sky scans to a regular 10 min time series

  This leads to the assumption, the one full scan is representative its assigned 10 min window. No aggregation is required, because one all-sky scan takes about 5 min.
- Apply QCS3D Layers to flag suspicious data values

  Specialized QCS3D checks for **WN3D_L1b_v1_MWR_SkyMaps** include:

  - Layer 5: Apply an elevation/azimuth clutter map to mask buildings in low elevations
  - Layer 5: Compute the current sun position and mask corresponding pixels

4.9. WN3D_L1b_v1_PrecipRadar_Sweeps

WN3D_L1b_v1_PrecipRadar_Sweeps	
Dependencies	-
Input Files	Raw radar PPI data files (L0)
Additional Input	Radar housekeeping files
Output	WN3D_L1b_v1_PrecipRadar_Sweeps data cube (NetCDF)

- Read PrecipRadar L0 volume scans, apply WEGN3D data model and convert to L1a
- Remap azimuth to a regular 0.5° grid within $[-90°, 90°]$

  To remap the observations to regular azimuth values, first, an interpolation grid is created. Then, for each elevation level independently, IDW(1, 4) is performed using the observation and grid azimuth values as sample and interpolation points respectively. As an additional constraint, only rays whose azimuth value lies in the respective 0.5° bin are used for the interpolation. This means that while the remapped value might be a weighted average, no correlations between neighbouring azimuth bins is introduced
- Compute attenuation-corrected horizontal reflectivity and differential reflectivity

  Horizontal reflectivity Z_{h} and differential reflectivity Z_{DR} are attenuation corrected by utilizing the specific differential phase K_{DP}. Specfically,

$$Z_{\mathrm{h,\,corr}}(r) = Z_{\mathrm{h}}(r) + 2\int_0^r a \cdot K_{\mathrm{DP}}(r')^b dr', \tag{7}$$

  where $a = 0.233$ and $b = 1.02$ For Z_{DR} the same formulation is used, with

$$Z_{\mathrm{DR,\,corr}}(r) = Z_{\mathrm{DR}}(r) + 2\int_0^r c \cdot K_{\mathrm{DP}}(r')^d dr', \tag{8}$$

  where $c = 0.128$ and $d = 1.156$.
- Compute precipitation rate from attenuation-corrected reflectivity from a variable $Z - R$-relationship:

$$R = a(Z_{\mathrm{h,\,corr}}, S) z_{\mathrm{h,\,corr}}^{b(Z_{\mathrm{h,\,corr}}, S)}, \tag{9}$$

  where $z_{\mathrm{h,\,corr}} = 10^{0.1 Z_{\mathrm{h,\,corr}}}$ and S is defined as the sum of the absolute $z_{\mathrm{h,\,corr}}$ differences in a 3 km $\times$ 3 km window around the computation point[3]. The values for a and b are then taken from the following table:

$Z_{\mathrm{h,\,corr}}$ (dBZ)	< 36.5			36.5 – 44	> 44
S	< 3.5	3.5 – 7.5	> 7.5		
a	125	200	320	200	77
b	1.4	1.6	1.4	1.6	1.9

- Compute second trip echo (STE) probability following Park et al. 2016
- Merge STE probability and radar housekeeping flags for QCS3D Layer 5 (intrastation check)

  Radar echos with an STE probability greater than 0.55 are flagged during the QCS3D check.

[3] https://www.dwd.de/DE/fachnutzer/wasserwirtschaft/unsere_leistungen/radarniederschlagsprodukte/niederschlagsbestimmung_pdf.pdf?__blob=publicationFile&v=2, last accessed 2024-09-09

5. DPG3D L2 Processing

5.1. WN3D_L2_BD_v1_Precipitation

WN3D_L2_BD_v1_Precipitation	
Dependencies	WN3D_L2_BD_v1_TroposphericProfiles WN3D_L2_BD_v1_Precipitation
Input Files	WN3D_L1b_v1_PrecipRadar_Sweeps WN3D_L2_BD_v1_TroposphericProfiles
Auxiliary Input	ERA5 Temperature Profiles
Output	WN3D_L2_BD_v1_Precipitation data cube (NetCDF)

- Read T_profile_cmp from **WN3D_L2_BD_v1_TroposphericProfiles**

  If no valid T_profile_cmp values are available, use ERA5 temperature profile instead.

- Read polarimetric radar variables from **WN3D_L1b_v1_PrecipRadar_Sweeps**

- Generate attenuation correction variable attn_corr from Z_h and Z_h_corr

$$A = Z_{\text{h, corr}} - Z_{\text{h}} \tag{10}$$

- Compute precipitation amount precip_amount from precip_rate

  The precipitation amount is computed from the instantaneous precipitation rate by numerical integration using the trapezoidal rule, followed by an epoch-wise differentiation resulting in

$$P(t_k) = [R(t_{k-1}) + R(t_k)]\frac{\Delta t}{2}. \tag{11}$$

- Geolocate volume scan and interpolate to data cube output grid

  Geolocation of the radar volume is performed by first transforming the data cube coordinates given in UTM33N coordinates (X, Y) and height above sea level h_{msl} into the local radar ENU reference frame.

  Once the target points are available in the radar reference frame, an adaptive Barnes interpolator (Askelson, Aubagnac, and Straka 2000) is used to interpolate the polarimetric variables. The adaptive Barnes interpolator is similar to IDW in the sense that it is also based on a weighted average of the sample points as given in Equation 24. The weights are however not simply defined by a power of the inverse distance, rather an exponential function with smoothing parameters

$$w_i(\epsilon, \alpha, \rho) = \exp\left(-\frac{(\rho-\rho_i)^2}{\kappa_\rho^2} - \frac{(\alpha-\alpha_i)^2}{\kappa_\alpha^2} - \frac{(\epsilon-\epsilon_i)^2}{\kappa_\epsilon^2}\right) \tag{12}$$

  is used to determine the sample point weights. The smoothing parameters are $\kappa_\rho = 150$, $\kappa_\alpha = 3$, and $\kappa_\epsilon = 3$. To reduced the computational load, only the 16 closest neighbors are used in the weighted average.

  To interpolate discrete categorical variables, the value of the data point i with the maximum weight is assigned instead of the weighted average.

- Compute signal extinction flag from geolocated attn_corr

  Values where attn_corr exceeds 20 dBZ are flagged and the resulting data variable signal_attenuation_flag is added to the data cube.

- Classify hydrometeors following Zrnić et al. 2001 using membership functions based on Straka, Zrnić, and Ryzhkov 2000 adapted for X-band by Evaristo et al. 2013 resulting in the data variable hydrometeor_type

- Classify convective precipitation areas following Powell, Houze, and Brodzik 2016 (PHB16)

  This decision tree base classifier, flags areas where convective precipitation occurs based on the attenuation corrected horizontal reflectivity. The resulting data variable precipitation_type is added to the data cube. The parameter values used in the algorithm can be found in the following table:

a	8 dBZ	Z_{weak}	15 dBZ
b	64 dBZ	Z_{shallow}	28 dBZ
Z_{th}	42 dBZ	A_{low}	0.6 km^2
R_{bg}	3.3 km	A_{med}	5 km^2
R_{conv}	2.1 km	A_{high}	200 km^2
Z_{conv}	48 dBZ		

- Compute <hydrometeor_type>_area_fraction data variables

  Based on the hydrometeor_type flag values, the area fraction covered by each hydrometeor_type is computed. This is done for each height level independently, resulting in (time, h_msl) profiles.

5.2. WN3D_L2_BD_v1_CloudStructure

WN3D_L2_BD_v1_CloudStructure	
Dependencies	WN3D_L2_BD_v1_TroposphericProfiles WN3D_L2_BD_v1_SkyMaps
Input Files	WN3D_L2_BD_v1_TroposphericProfiles WN3D_L2_BD_v1_SkyMaps
Auxiliary Input	ERA5 Temperature Profiles
Output	WN3D_L2_BD_v1_CloudStructure data cube (NetCDF)

The **WN3D_L2_BD_v1_CloudStructure** combines infrared (IR) brightness temperature all-sky scans (T_ir or T_ir_narrowband) and air temperature profiles (T_profile_cmp) to determine the cloud base heights for all identified cloud pixels following the approach of Brede et al. 2017. IR brightness temperatures are converted to temperature using empirical functions for emissivity dependent on cloud liquid water path (lwp), which can be found in Figure 3. Combining profiles with IR brightness temperature all-sky scans

A caveat of using IR brightness temperatures from a radiometer is that along each measurement ray, only the temperature of the first detected cloud can be observed, thus no information beyond this point can be retrieved.

Still,

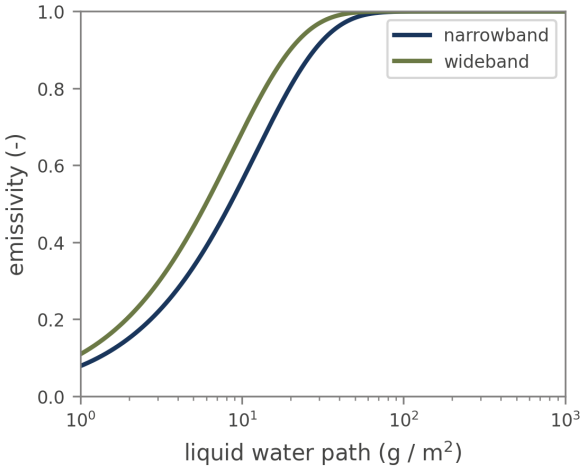

Figure 3: Modeled infrared cloud emissivity depending on liquid water path and measurement bandwidth for 8 µm - 14 µm (wideband) and 9.6 µm - 11.5 µm (narrowband).

- Read T_profile_cmp from **WN3D_L2_BD_v1_TroposphericProfiles**

  If no valid T_profile_cmp values are available, use ERA5 temperature profile instead.

- Read T_ir from **WN3D_L2_BD_v1_SkyMaps**

  If no valid T_ir values are available, use T_ir_narrowband.

- Compute cloud emissivity using lwp (liquid water path) from **WN3D_L2_BD_v1_SkyMaps**

  Infrared cloud emissivity is computed from liquid water path using empirical models from Stephens 1978 and Chylek and Ramaswamy 1982, depending on the bandwidth (cf. also Figure 3). The models express the cloud emissivity e as an exponential function depending on the liquid water path as

  $$e = 1 - \exp(-a \cdot LWP), \tag{13}$$

  where $a = 0.116$ (Stephens 1978) for wideband (T_ir) and $a = 0.082$ (Chylek and Ramaswamy 1982) for narrowband (T_ir_narrowband) infrared brightness temperatures. In case no valid lwp values are available, an emissivity of $e = 1$ is assumed.

- Compute cloud base temperature T from brightness temperature and emissivity via $T = T_{\text{ir}}/e$

- Read cloud_mask from **WN3D_L2_BD_v1_SkyMaps**

  If no valid cloud_mask values are available, use cloud temperature T and liquid water path LWP to find cloudy pixels. A pixel is assumed to be cloudy when

  $$\frac{T^2}{a_T^2} + \frac{(LWP \sin^2 \epsilon)^2}{a_{LWP}^2} > 1, \tag{14}$$

  with a_T = 260 K and a_{LWP} = 25 g m^2, i.e., the (T, LWP) tuple lies outside an empirically determined ellipsoidal region.

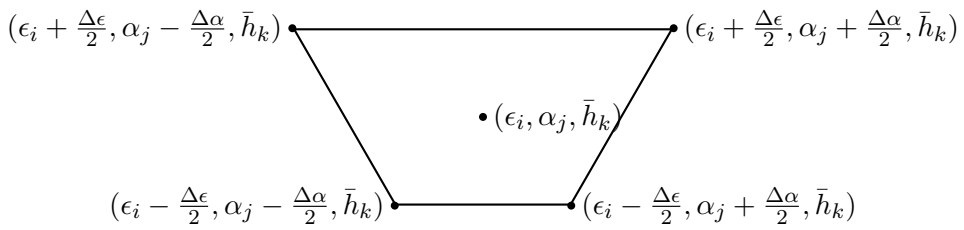

Figure 4: Base trapezoid of a 3D cloud grid voxel.

- Compute cloud base height for cloudy elevation/azimuth pixels by comparing T and profile temperatures T_{prof} in T_profile_cmp

  $T_{\text{prof}}(h)$ is assumed to be a piecewise linear function with nodal points represented by T_profile_cmp. The cloud base height h_{cb} is determined by applying Newton's root finding method to the function $f(h) = T - T_{\text{prof}}(h)$, with an initial guess of 1 km.

- Geolocate cloud pixels to get height above sea level (h_{msl}) coordinates for each cloudy pixel

  - Compute the distance ρ from the cloud structure radiometer for each pixel via

$$\rho(\epsilon, \alpha) = \frac{1}{\sin \epsilon} h_{\text{cb}}(\epsilon, \alpha). \tag{15}$$

  - Compute ENU coordinates from (ϵ, α, ρ) and transform them into LLH via ECEF (cf. subsection 6.1.
  - Transform the resulting ellipsoidal height into height above sea level h_{msl} (cf. subsection 6.2)

- Aggregate (bin) cloudy pixels based on height

  Each cloud pixel is assigned a height level $\bar{h}_k$ by binning the observed h_{msl}. Binning the observed cloud base heights into discrete levels means that each detected cloud can now be assigned as precomputed cloud grid voxel in the shape of a trapezoid as shown in Figure 4.

- Classify each voxel, resulting in one of three states: clear sky, cloudy, and unobservable

  The unobservable state is a consequence of the IR radiometer only being able to detect the first cloud along each measurement ray. Thus, once a cloud is detected, no information about cloud presence beyond this point can be acquired.

- Transform voxels into gridded data cubes by assigning the respective flag value to each of the (X, Y, h_{msl}) grid points within the voxel

  The transformation from voxel domain to geolocated grid domain results in the cloud mask variable cloud_structure.

- Generate cloud base height map H_cb, by reducing the cloud_structure data cube

  For each grid point, the height level value of the lowest detected cloud is taken as the cloud base height, resulting in a 2-D map.

6. Georeferencing and Gridding

Gridding denotes the interpolation of irregularly spaced observation data onto a regular coordinate grid, which allows the representation of information in data (hyper-)cubes. Gridding happens on both horizontal and vertical axes, as well as in spherical coordinate systems. This section gives the theoretical basis of the gridding operations performed in the WEGN3D data cubes.

6.1. Spatial Coordinate Transformation and Projection

When transforming observations from raw measurements from a local sensor frame into a common, georeferenced coordinate system a few intermediate transformation steps and coordinate systems are necessary. These coordinate systems and the transformations between them are outlined in the following.

EAR (Elevation, Azimuth, Range)

This coordinate system's origin is the sensor reference point, for example, the radome or radiometer mounting platform. The azimuth α is defined as the angle between due north and the line of sight and counted clockwise positive. Its range is between 0° and 360° or −180° and 180°. Elevation ϵ is the angle between the measurement ray line of sight and the horizontal plane counted positive upwards (i.e. 90° elevation points towards the zenith direction, 0° elevation points towards the horizon) and lies between 0° and 90°. Range ρ is the euclidian distance along the line of sight to the range gate center and thus always positive.

The corresponding coordinate frame is realized either on a software basis through solar angle comparisons or by physical alignment of the sensor platform.

ENU (East, North, Up)

A cartesian local-level coordinate system with the same axis alignment as EAR, that is "up" points toward the (geometric) zenith direction, "north" to due north, and "east" complements the orthogonal frame. Its origin is the intersection of a line that is normal to the reference ellipsoid (the ellipsoid approximating Earth's figure) and goes through the sensor reference point. Thus, the EAR origin is a signed distance h shifted along the ellipsoid normal away from the ENU origin. Conversion between EAR and ENU is given by

$$\begin{aligned} E &= \rho \cos\epsilon \sin\alpha \\ N &= \rho \cos\epsilon \cos\alpha \\ U &= \rho \sin\epsilon + h \end{aligned} \tag{16}$$

and

$$\begin{aligned} \epsilon &= \arctan \frac{\sqrt{N^2 + E^2}}{U} \\ \alpha &= \arctan \frac{N}{E} \\ \rho &= \sqrt{E^2 + N^2 + U^2}. \end{aligned} \tag{17}$$

Note that the 2-argument arctangent function (i.e. `atan2`) should be used in Equation 17 to get the correct and unambiguous angle.

LLH Longitude, Latitude, Height

LLH is a Geographic Coordinate System (GCS) with geodetic longitude λ, latitude φ, and ellipsoidal height h. A GCS is always attached to a reference ellipsoid, which is WGS84 throughout this report. Longitude measures the rotational angle between the reference meridian and the point of interest. The geodetic latitude is the angle between the equatorial plane and a line that is normal to the reference ellipsoid and goes through the point of interest. Ellipsoidal height is defined as the signed distance along the normal vector between the point of interest and the reference ellipsoid, with positive heights above and negative heights below the ellipsoidal surface.

ECEF (Earth Centered, Earth Fixed)

ECEF is a global, cartesian coordinate system with its origin in the geocenter, the x-axis points towards the prime meridian, the z-axis towards the north pole, and the y-axis complements the orthogonal right handed coordinate system. Transformation between ENU and ECEF coordinates requires the origin of the ENU-frame expressed in geographic (geodetic) longitude and latitude λ_0, φ_0. Then, global coordinates ECEF can be computed from a local ENU-triple via

$$\begin{aligned} x &= x_0 + s_\lambda E - c_\lambda s_\varphi N + c_\lambda c_\varphi U \\ y &= y_0 + c_\lambda E - s_\lambda s_\varphi N + s_\lambda c_\varphi U \\ z &= z_0 + c_\varphi N + s_\varphi U, \end{aligned} \tag{18}$$

where $s_\lambda = \sin\lambda_0, c_\lambda = \cos\lambda_0, s_\varphi = \sin\varphi_0, c_\varphi = \cos\varphi_0$, and x_0, y_0, z_0 is the point in ECEF coordinates computed from the LLH-triple $(\lambda_0, \varphi_0, 0)$. Generally, ECEF coordinates can be computed from LLH by evaluating

$$\begin{aligned} x &= (N(\varphi) + h)\cos\varphi\cos\lambda \\ y &= (N(\varphi) + h)\cos\varphi\sin\lambda \\ z &= ((1 - e^2)N(\varphi) + h)\sin\varphi, \end{aligned} \tag{19}$$

where $e^2 = 2f - f^2$ is the squared first eccentricity and $N(\varphi) = \frac{a}{\sqrt{1-e^2\sin^2\varphi}}$ is the radius of curvature, depending on the reference ellipsoid flattening f and semi-major axis a. There is no exact closed form conversion from ECEF to LLH, however, multiple iterative approaches exist. In WEGN3D, Bowring's irrational geodetic-latitude equation is solved iteratively to compute (λ, φ, h)-triples Bowring 1976.

Equation 18 represents a translation and rotation, so the inverse conversion (ECEF to ENU) is given by

$$\begin{aligned} E &= -s_\lambda(x - x_0) + c_\lambda(y - y_0) \\ N &= -c_\lambda s_\varphi(x - x_0) - s_\lambda s_\varphi(y - y_0) + c_\varphi(z - z_0) \\ U &= c_\lambda c_\varphi(x - x_0) + s_\lambda c_\varphi(y - y_0) + s_\varphi(z - z_0). \end{aligned} \tag{20}$$

PRJ (Projected Coordinate Systems)

Horizontal grids in WEGN3D are typically expressed in projected geographical coordinates (X, Y). PRJ coordinates are a purely horizontal representation, thus height only needs to be projected if the reference ellipsoids in source and target projections differ. Thus, in

almost all cases, a LLH-triple is converted into a PRJ-triple (X, Y, h) by simply appending the height to the projection result. Projections in WEGN3D are performed via the *PROJ* library documented by PROJ contributors 2023.

6.2. Height Coordinate Transformation

Orthometric Height (Height Above Mean Sea Level) $\leftrightarrow$ Ellipsoidal Height

Transformation between orthometric height (height above mean sea level) h_{msl} and ellipsoidal height h_{ell} is performed by making use of the height correction grids of the Austrian Federal Federal Office of Metrology and Surveying.

The difference between *Höhen-Grid plus Geoid v1.0.1*[4] and *Höhen-Grid v1.0.1*[5] yields a height correction grid Δh_{corr}, that can be used to convert between h_{msl} and h_{ell}, like

$$h_{\mathsf{ell}} = h_{\mathsf{msl}} + \Delta h_{\mathsf{corr}}. \tag{21}$$

Since the height correction Δh_{corr} is location dependent, the coordinates of the point whose height shall be transformed are also required. The coordinates are projected into the native height correction projection (*EPSG:4312*), then IDW(1, 4) is performed to obtain the height correction value at the target point.

Orthometric Height (Height Above Mean Sea Level) $\leftrightarrow$ Height Above Ground Level

Conversion between orthometric height (height above mean sea level) h_{msl} and height above ground level h_{gnd} is performed by applying the *EU-DEM v1.1*[6]. With the EU-DEM, the height above ground level can be computed from orthometric heights via

$$h_{\mathsf{gnd}} = h_{\mathsf{msl}} - \Delta h_{\mathsf{DEM}} \tag{22}$$

In analogy to the $h_{\mathsf{ell}} \leftrightarrow h_{\mathsf{msl}}$ transformation, the DEM correction is location dependent, and is therefore interpolated to the target points via a bilinear interpolation.

6.3. Spatial Interpolation and Gridding

Within WEGN3D, generally inverse distance weighting (IDW) based interpolators are used for spatial interpolation. These interpolators form a weighted average of the sample S, where the weights are based on a power p of the inverse euclidian distance, as in

$$w_i(\mathbf{x}) = \frac{1}{\|\mathbf{x} - \mathbf{x}_i\|^p}. \tag{23}$$

Typically, to improve computational performance, not the full sample S but a subset $N \subset S$ of k-nearest points around the interpolation is used in the weighted average, as far away points

[4] https://www.bev.gv.at/Services/Produkte/Grundlagenvermessung/Hoehen-Grid-plus-Geoid.html, last accessed 2024-09-09

[5] https://www.bev.gv.at/Services/Produkte/Grundlagenvermessung/Hoehen-Grid.html, last accessed 2024-09-09

[6] https://sdi.eea.europa.eu/catalogue/srv/api/records/d08852bc-7b5f-4835-a776-08362e2fbf4b, last accessed 204-09-09

have limited contribution to the interpolation result anyway. With these weights, the value at the interpolation point $\mathbf{x}$ is then computed via

$$f(\mathbf{x}) = \begin{cases} \frac{\sum_{i \in N} w_i f(\mathbf{x}_i)}{\sum_{i \in N} w_i} & \text{if } \|\mathbf{x} - \mathbf{x}_i\| \neq 0 \; \forall i \in N \\ f(\mathbf{x}_i) & \text{if } \|\mathbf{x} - \mathbf{x}_i\| = 0 \text{ for any } i \in N \end{cases}. \tag{24}$$

Since the IDW is fully defined by the power parameter p and the amount of nearest neighbours k used in the interpolation, we introduce IDW(k, p) as a descriptor. IDW interpolators cannot only be applied to geographical data, but are in principle applicable to any quantity where sensible coordinate values are available. One use case within WEGN3D is the interpolation and resampling of all-sky scans (SkyMaps) given as elevation/azimuth grids. Since lines of constant azimuth converge towards the zenith, the interpolation domain varies within the elevation/azimuth grid. If this behavior is to be avoided, the 2D elevation/azimuth coordinates can be be transformed in 3D directional unit vectors

$$\mathbf{u} = \begin{bmatrix} \cos\epsilon \cos\alpha \\ \cos\epsilon \sin\alpha \\ \sin\epsilon \end{bmatrix}, \tag{25}$$

which can be used as input coordinates for IDW interpolation. Sometimes it is useful to increase the interpolation weights in either elevation or azimuth direction. This can be achieved by deforming the sphere represented by Equation 25, through scaling the z-component of the unit vector as in

$$\mathbf{v}(s) = \begin{bmatrix} \cos\epsilon \cos\alpha \\ \cos\epsilon \sin\alpha \\ s \cdot \sin\epsilon \end{bmatrix}. \tag{26}$$

A factor $s < 1$ gives more weight to neighbouring elevations, while a factor $s > 1$ weighs neighbouring azimuth observation rays stronger.

Acronyms

ATBD Algorithm Theoretical Basis Document. 3

CRAS Command Receive Archiving System. 3, 4

DPG3D WegenerNet 3D Data Product Generator. 3, 4

EAR Elevation, Azimuth, Range coordinate system. 17

ECEF Earth Centered Earth Fixed coordinate system. 18

ENU East, North, Up, coordinate system. 17, 18

GCS Geographic Coordinate System. 18

LLH Longitude, Latitude, Height coordinate system. 18, 19

LWP Liquid Water Path. 23

MWR Microwave Radiometer. 23

PRJ Projected Coordinate System. 18, 19

QCS3D WegenerNet 3D Quality Control System. 3, 4, 6

UTM33N Universal Transverse Mercator Projection, Zone 33N. 13

WEGN3D WegenerNet 3D Observing System. 3, 5–12, 17, 19, 20

WPS3D WegenerNet 3D Processing System. 3–5

References

Bowring, B. R. (July 1976). "TRANSFORMATION FROM SPATIAL TO GEOGRAPHICAL COORDINATES". In: *Survey Review* 23.181. Publisher: Taylor & Francis, pp. 323–327. ISSN: 0039-6265. DOI: 10.1179/sre.1976.23.181.323. URL: https://doi.org/10.1179/sre.1976.23.181.323.

Stephens, G. L. (Nov. 1978). "Radiation Profiles in Extended Water Clouds. II: Parameterization Schemes". English. In: *Journal of Atmospheric Sciences* 35.11. Place: Boston MA, USA Publisher: American Meteorological Society, pp. 2123–2132. DOI: 10.1175/1520-0469(1978)035<2123:RPIEWC>2.0.CO;2. URL: https://journals.ametsoc.org/view/journals/atsc/35/11/1520-0469_1978_035_2123_rpiewc_2_0_co_2.xml.

Bolton, David (1980). "The Computation of Equivalent Potential Temperature". In: *Monthly Weather Review* 108.7. Place: Boston MA, USA Publisher: American Meteorological Society, pp. 1046–1053. DOI: 10.1175/1520-0493(1980)108<1046:TCOEPT>2.0.CO;2. URL: https://journals.ametsoc.org/view/journals/mwre/108/7/1520-0493_1980_108_1046_tcoept_2_0_co_2.xml.

Chylek, Petr and V. Ramaswamy (Jan. 1982). "Simple Approximation for Infrared Emissivity of Water Clouds". English. In: *Journal of Atmospheric Sciences* 39.1. Place: Boston MA, USA Publisher: American Meteorological Society, pp. 171–177. DOI: 10.1175/1520-0469(1982)039<0171:SAFIEO>2.0.CO;2. URL: https://journals.ametsoc.org/view/journals/atsc/39/1/1520-0469_1982_039_0171_safieo_2_0_co_2.xml.

Askelson, Mark A., Jean-Pierre Aubagnac, and Jerry M. Straka (2000). "An Adaptation of the Barnes Filter Applied to the Objective Analysis of Radar Data". In: *Monthly Weather Review* 128.9. Place: Boston MA, USA Publisher: American Meteorological Society, pp. 3050–3082. DOI: 10.1175/1520-0493(2000)128<3050:AAOTBF>2.0.CO;2. URL: https://journals.ametsoc.org/view/journals/mwre/128/9/1520-0493_2000_128_3050_aaotbf_2.0.co_2.xml.

Straka, Jerry M., Dusan S. Zrnić, and Alexander V. Ryzhkov (2000). "Bulk Hydrometeor Classification and Quantification Using Polarimetric Radar Data: Synthesis of Relations". In: *Journal of Applied Meteorology* 39.8. Place: Boston MA, USA Publisher: American Meteorological Society, pp. 1341–1372. DOI: 10.1175/1520-0450(2000)039<1341:BHCAQU>2.0.CO;2. URL: https://journals.ametsoc.org/view/journals/apme/39/8/1520-0450_2000_039_1341_bhcaqu_2.0.co_2.xml.

Zrnić, Dusan S. et al. (2001). “Testing a Procedure for Automatic Classification of Hydrometeor Types”. In: *Journal of Atmospheric and Oceanic Technology* 18.6. Place: Boston MA, USA Publisher: American Meteorological Society, pp. 892–913. DOI: 10.1175/1520-0426(2001)018<0892:TAPFAC>2.0.CO;2. URL: https://journals.ametsoc.org/view/journals/atot/18/6/1520-0426_2001_018_0892_tapfac_2_0_co_2.xml.

Evaristo, R. M. et al. (Sept. 2013). “A holistic view of precipitation systems from macro- and microscopic perspective”. In.

Park, S.-G. et al. (2016). “Identification of Range Overlaid Echoes Using Polarimetric Radar Measurements Based on a Fuzzy Logic Approach”. In: *Journal of Atmospheric and Oceanic Technology* 33.1. Place: Boston MA, USA Publisher: American Meteorological Society, pp. 61–80. DOI: 10.1175/JTECH-D-15-0042.1. URL: https://journals.ametsoc.org/view/journals/atot/33/1/jtech-d-15-0042_1.xml.

Powell, Scott W., Robert A. Houze, and Stella R. Brodzik (2016). “Rainfall-Type Categorization of Radar Echoes Using Polar Coordinate Reflectivity Data”. In: *Journal of Atmospheric and Oceanic Technology* 33.3. Place: Boston MA, USA Publisher: American Meteorological Society, pp. 523–538. DOI: 10.1175/JTECH-D-15-0135.1. URL: https://journals.ametsoc.org/view/journals/atot/33/3/jtech-d-15-0135_1.xml.

Brede, Benjamin et al. (Oct. 2017). “Spatiotemporal High-Resolution Cloud Mapping with a Ground-Based IR Scanner”. In: *Advances in Meteorology* 2017, pp. 1–11. DOI: 10.1155/2017/6149831.

Fuchsberger, J., G. Kirchengast, and T. Kabas (2021). “WegenerNet high-resolution weather and climate data from 2007 to 2020”. In: *Earth System Science Data* 13.3, pp. 1307–1334. DOI: 10.5194/essd-13-1307-2021. URL: https://essd.copernicus.org/articles/13/1307/2021/.

PROJ contributors (2023). *PROJ coordinate transformation software library*. Open Source Geospatial Foundation. DOI: 10.5281/zenodo.5884394. URL: https://proj.org/.

A. Recomputation of Radiometer Retrievals

The HATPRO G5 Software package uses neural networks to derive atmospheric quantities, such as temperature and humidity profiles, Liquid Water Path (LWP), and tropospheric path delay from measured brightness temperatures. Each neural network has the same basic structure, consisting of input scaling and shifting, two dense layers, each followed by a non-linear activation functions, and output scaling and shifting. This can be written as

$$\mathbf{y} = \mathbf{f}(\mathbf{x}) = \mathbf{h}_2\Big(\mathbf{W}_2\mathbf{h}_1\big(\mathbf{W}_1\left(\mathbf{x} - \mathbf{b}_x\right)/\mathbf{s}_x + \mathbf{b}_1, \alpha_1\big) + \mathbf{b}_2, \alpha_2\Big)\mathbf{s}_y + \mathbf{b}_y. \tag{27}$$

The layer coefficients $\mathbf{W}_1, \mathbf{W}_2$ and bias vectors $\mathbf{b}_1, \mathbf{b}_2$, in addition to the input/output scaling and offset vectors $\mathbf{s}_{\{x,y\}}, \mathbf{b}_{\{x,y\}}$ and activation function type h and smoothing parameter α are given in retrieval files and fully describe the neural network. Throughout the retrievals in WEGN3D the tangens hyperbolicus is used as an activation function with the parameter α governing the slope of the function, as in

$$h(x, \alpha) = \frac{2}{1 + e^{-2\alpha x}} - 1. \tag{28}$$

Observational input into the retrieval networks can be put into three categories:

- Zenith Scan

  MWR brightness temperatures from a zenith-pointing scan with longer integration time

- Boundary Layer Scan

  MWR brightness temperatures from a boundary layer scan along a constant azimuth profile with longer integration time. The scan starts at an elevation of 4.2° and ends in zenith direction with an irregular sampling.

- All-Sky Scan

  MWR brightness temperatures from an all-sky scan with a regular elevation/azimuth sampling. Due to the large number of sampled points, this scan exhibits a shorter integration time of approximately 1 s.

In addition to observed brightness temperatures, surface meteorological measurements (air temperature, relative humidity and air pressure), and information about the day of year at which the measurements were taken with

$$\begin{aligned} c_{\mathsf{DOY}}(t) &= \cos 2\pi \frac{\mathsf{DOY}(t)}{365} \\ s_{\mathsf{DOY}}(t) &= \sin 2\pi \frac{\mathsf{DOY}(t)}{365}. \end{aligned} \tag{29}$$