Facilitating cloud radar and lidar algorithms: the Cloudnet Instrument Synergy/Target Categorization product

Robin J. Hogan* and Ewan J. O'Connor

August 17, 2004

1 Introduction

There is a growing recognition of the usefulness of cloud radar for evaluating numerous aspects of the representation of clouds in numerical forecast models, but to make the best use of such data it should be combined with information from other sources such as lidar, radiometers, rain gauge and a forecast model. For cloud retrievals to be useful for evaluating models they must be applied to large volumes of data, rather than just a few case studies, which necessitates the development of robust algorithms for performing essential preprocessing such as identifying cloud types and hence to which data certain algorithms can be applied. Differences in data formats and conventions can make it a major undertaking to adapt an algorithm developed for the instruments at one site to work with data from another.

This document describes an intermediate data product designed to facilitate the application of multi-sensor algorithms by performing most of the typical preprocessing that such algorithms require and providing the results in a common format for all sites. It was developed for the *Cloudnet* project[1] involving three European remote sensing sites, but is also applicable to data from the similar Atmospheric Radiation Measurement (ARM) sites[2]. In the context of *Cloudnet* this product is designated "Level 1c", lying between calibrated instrumental data in NetCDF format at Level 1b and meteorological products at Level 2. The key procedures that are carried out are:

Standardization of conventions: The output data are provided in a common format for all sites, with common units and with the same conventions such as height being above mean sea level, Doppler velocity being positive upwards and so on.

Ingestion of model data: Many algorithms require temperature and horizontal wind speed, and unless there are regular 6-hourly radiosonde ascents from close to the site, these are best obtained from a model analysis or forecast.

Regridding: The lidar and model data are interpolated on to the same time-height grid as the radar. Likewise, the rain rate and liquid water path are interpolated on to the time axis of the radar.

Target categorization: Each pixel is categorized in terms of the presence of liquid droplets, ice, insects, aerosol etc., thereby allowing algorithms specific to one type of target to be applied.

Gaseous attenuation correction: Model temperature, pressure and humidity are used to correct radar reflectivity for gaseous attenuation.

Liquid attenuation correction: The radar can be significantly attenuated by the presence of liquid water, but liquid water path from the microwave radiometer, in combination with the location of the liquid water in the profile from lidar and radar, allows this effect to be corrected.

Instrument errors: For radar and lidar, variables representing both random error (due to the finite number of samples averaged and the accuracy of any correction for radar attenuation) and systematic error (an indication of the accuracy of the calibration) are added, allowing subsequent retrieval algorithms to make realistic estimates of their associated error.

Instrument sensitivity: Knowledge of the minimum detectable signal of the radar allows one to take account of clouds that may not be detected in comparisons with a model.

Data quality flags: These inform the user when signals are contaminated by ground clutter, unknown radar attenuation in rain and other effects.

In section 2 the essential and optional sources of input data are described, and in section 3 the details of the algorithm

* *Corresponding author address:* Department of Meteorology, Earley Gate, PO Box 243, Reading RG6 6BB, UK.
E-mail: r.j.hogan@reading.ac.uk.

[1] http://www.met.rdg.ac.uk/radar/cloudnet/

[2] http://www.arm.gov/

are provided. The format of the data is outlined in the appendix.

2 Input datasets

The essential instruments that must be present are vertically pointing cloud radar and backscatter lidar, and hourly model forecast data must be available over the site. Recommended but non-essential instruments are microwave radiometer (for providing liquid water path) and rain gauge. Processing is done one day at a time. Ideally the instruments operate continuously, but if at any time in the day radar, lidar or model data are missing then there will be a gap in the output product. The absence of data from the non-essential instruments may mean that correction for radar attenuation is not possible, but data quality flags will indicate that this is the case.

The decision regarding which instruments and auxiliary data to include in this product was made by considering those that are most used in retrieval algorithms and those that are necessary to correct for radar attenuation and flag bad data. Radar and lidar are complementary due to their very different dependence on particle size; this means that the combination of the two offers the most accurate estimates of cloud occurrence and cloud fraction (Mace et al. 1998, Hogan et al. 2001), and can also be used to retrieve particle size (Donovan et al. 2001, O'Connor et al. 2004). However, it is not the intention that all instruments at a site that could conceivably be used in a retrieval algorithm (or that might be useful for model comparison, such as broadband fluxes) should be combined into this product, as it would be an unending task to have to cope with all possible instruments and their own idiosyncrasies. A possible exception might be when two radars of different frequencies are available, as there are a number of dual-wavelength algorithms that offer distinct advantages over what is possible with a single radar and lidar (e.g. Sekelsky et al. 1999, Hogan et al. 2000, 2004) and the analysis of such data requires much of the preprocessing provided by this product to have been performed.

The algorithm is implemented in Matlab[3], although future implementation in C is possible. The input datasets are read into the processing algorithm using separate functions that may make modifications depending on the particular instrument and site. It is convenient, but not essential, if the format of the input datasets is NetCDF.

2.1 Cloud radar

The vertically pointing radar would typically operate continuously at 35 or 94 GHz and provides profile with a time resolution of around 30 seconds and a height resolution of better than 100 m. The radar resolution is used as a master grid on to which all other datasets are interpolated. The temporal resolution of 30 s is long enough that the datasets are conveniently small to use, but short enough that cloud fraction and other parameters to be used to evaluate models are of sufficiently high precision.

Essential parameters that the radar must provide are radar reflectivity factor Z and Doppler velocity v[4]. Radar reflectivity factor should have had the following processing applied to it:

- Linear averaging to the resolution required.
- Noise subtraction: the reported Z should be that of the atmospheric targets without the contribution from instrument noise and thermal emission that is present in the raw measurements.
- Clear sky clearing: areas with no detectable atmospheric signal should should be flagged (typically using the `_FillValue` facility in NetCDF) such that Matlab can treat the value as "Not a Number" (`NaN`). Any remaining speckle noise should be removed using masking algorithms.
- Elimination of artifacts, such as the near field effect for instruments with large antennas, overlap effects for some bistatic systems and spurious instrument-specific echos.

The quality control algorithm attempts to flag likely ground clutter, but since the clutter behaviour can be very different for different radars it is better if it can be removed prior to being read in. Likewise, it is beneficial but not necessary for insects to have been removed. The radar should report values as low as possible above the ground. No attempt should be made to correct for attenuation.

The radar should obviously be calibrated as well as possible and an indication of the likely accuracy of the calibration should be available. If necessary the appropriate factor will be applied to Z in order to conform to the following calibration convention, defined for a distribution of Rayleigh-scattering liquid water droplets in the absence of attenuation:

$$Z = \frac{|K|^2}{|K_0|^2} \int_0^\infty n(D) D^6 dD,$$

where $|K|^2$ is the dielectric factor of liquid water, $|K_0|^2$ is the same but at 0°C and $n(D)dD$ is the number concentration of droplets in the diameter range D to $D + dD$. This defines Z

[3] http://www.mathworks.com

[4] *A large volume of legacy data without Doppler velocity are available from Chilbolton, so the algorithm may be modified in future to cope without Doppler velocity.*

in linear units, but it is conventional to use logarithmic dBZ units defined thus:

$$Z[\mathrm{dBZ}] = 10 \log_{10} \left(Z[\mathrm{mm}^6\ \mathrm{m}^{-3}] \right).$$

This calibration convention ensures that in the absence of attenuation, a cloud at 0°C containing one million 100 μm (i.e. Rayleigh scattering) droplets per cubic metre will have a reflectivity of 0 dBZ at all frequencies. If several reflectivity channels are available from a particular site, such as the four specialized modes used by the MMCR radars on the ARM sites (Clothiaux et al. 1999), then the one most free from artifacts will be taken, even if this is not the most sensitive. This is because the Instrument Synergy/Target Categorization product is intended to be used in automated algorithms that operate on large volumes of data without the need for significant human quality control.

The Doppler velocity should have had clear sky pixels removed in the same way as Z. If necessary the values will be inverted to ensure that the convention of positive velocities upwards is adhered to. The velocities need not be unfolded, but the folding velocity v_{fold} should be known to the algorithm. It is possible that the algorithm for locating the melting layer (see section 3.4.1) will perform poorly if the folding velocity is too small.[5]

Other radar parameters, such as spectral width (σ_v), will be transfered into the output dataset. If available, the 30-s standard deviation of the 1-s mean velocities, $\sigma_{\bar{v}}$(used to estimate turbulence levels; Bouniol et al. 2003), will be used to assist in the diagnosis of ground clutter.

2.2 Cloud lidar

The cloud lidar is principally used to identify the base of liquid water clouds. Most commonly this instrument will be a near-infrared lidar ceilometer reporting only attenuated backscatter coefficient β'. The instrument should be operated pointing between 2° and 5° from zenith to avoid specular reflection from horizontally aligned pristine crystals, which could be mistaken by the algorithm for the presence of supercooled liquid water (Hogan et al. 2003b). The lidar should be calibrated as well as possible (e.g. using the technique of O'Connor et al. 2004), although as few algorithms require an accurate absolute β', this is not as important as for Z. β' is converted into units of m^{-1} sr^{-1}. The algorithm assumes that all signals from the lidar are due to atmospheric particulates such as cloud and aerosol, rather than Rayleigh scattering from air molecules. This is generally valid in the near infrared (e.g. 905 nm used by the Vaisala CT25K and CT75K instruments) but for the 532 nm micropulse lidars used on the ARM sites an additional processing step is necessary to distinguish between molecular and particulate echos, as well as to remove instrument background noise and any negative values.[6]

[5] *It is hoped that a future version will include an algorithm to unfold the velocities, although in convective conditions this may be difficult.*

2.3 Model parameters

At least four radiosonde profiles per day would be required from close to the site to provide adequate dynamic and thermodynamic data above the instruments. Because most sites do not have this information, we use hourly profiles from short-range model forecasts. These models assimilate the data from the radiosonde network so the temperature (T), pressure (p), humidity (q) and horizontal wind (u and v) will usually be accurate enough for our purposes. We do not make use of the cloud variables as these are to be evaluated, and will generally be much less accurate.

Before it can be used the model output must be converted into the *Cloudnet* single-site model format (Openshaw 2004), which includes calculation of a number of radar propagation and scattering parameters at 35 GHz and 94 GHz that depend on thermodynamic state. The variables T, p, u and v are then interpolated on to the time grid of the radar and provided as part of the Instrument Synergy/Target Categorization product. To conserve space, they are not interpolated in height; subsequent processing algorithms will need to do this. The following additional parameters are loaded in and used by the algorithm, but are not output to the final product: q, specific gas attenuation κ_g (dB km^{-1}) (predicted from model q, T and p using the line-by-line model of Liebe 1985), specific gas attenuation for saturation with respect to liquid water κ_{gs} (dB km^{-1}) and specific liquid water attenuation κ_l (dB km^{-1} [g m^{-3}]$^{-1}$), using the formulation given by Liebe et al. (1989). For details of their use, see sections 3.4.1 and 3.5.

In *Cloudnet* we use the Met Office 6–11 hour forecast data as this model has the highest horizontal resolution (around 12 km) and the other models provide data only from 12 hours onwards. When Met Office data are unavailable we use ECMWF model output.

2.4 Microwave radiometer

A recommended but optional input dataset is liquid water path (LWP) derived using multiple wavelength radiometers. This is principally important for the correction of radar attenuation attenuation in liquid water clouds; at 94 GHz it can exceed 5 dB and so if correction is not performed then

[6] *Depolarisation and Raman lidars may be operational continuously from some sites, and as the depolarisation ratio and the extinction are extremely useful parameters for liquid water diagnosis and cloud retrievals, these may be provided in the Instrument Synergy/Target Categorization product.*

reliable retrievals in ice clouds above liquid clouds based on the value of Z are impossible. LWP is also used in retrieval algorithms for liquid water clouds.

Ideally LWP will not be derived using climatologically tuned coefficients, as these can result in errors exceeding 50 g m^{-2}, including retrieved values going negative. The preferred method is described by Gaussiat (2004) which makes use of other sources of data to provide more accurate coefficients; lidar is used to locate the height (and therefore the temperature) of liquid cloud, and the model to provide a more appropriate humidity distribution than climatology. Also, profiles identified by the lidar to be free of liquid water cloud are used to estimate the zero offset and by interpolation across periods of cloud, to provide considerably more accurate LWP in thin cloud. This approach also has the advantage that it can produce accurate LWP even if with poorly calibrated radiometer channels.

2.5 Rain gauge

The presence of rain on the dish or radome of a radar can result in a large and variable attenuation (Hogan et al. 2003a), of order 10 dB, making it impossible to use the absolute value of Z with any reliability. Rain itself also extinguishes the signal and can be difficult to correct for as microwave radiometers tend not to provide accurate LWP estimates in rain, and also the attenuation predicted by LWP may not be valid for large drops if their extinction is outside the Rayleigh regime.

To diagnose the presence of rain we use either a rain gauge or the radar itself. Ideally the rain gauge should have a fairly high sensitivity; tipping-bucket gauges in particular can be very slow to register the start of light rain events. The radar parameters Z and σ_v in the lowest few gates can also be used to diagnose probable rain on the ground. This approach tends to be more sensitive than a rain gauge, so rain gauge is not treated as an essential input dataset.

3 The algorithm

3.1 Standardization of conventions

The data read in consist of the individual fields and the associated meta-data describing the fields, typically lifted directly from the NetCDF attributes. A number of minor changes are made before further analysis is performed or attributes are copied into the final product.

Firstly the vertical coordinate of each input dataset is converted to height above mean sea level in metres. This correction is necessary as most instruments report range from the instrument in kilometres, and they may have been mounted at different heights above the ground. As the lidar often points several degrees from zenith, the range reported by this instrument is multiplied by the sine of the zenith angle to obtain height.

Variable and attribute names are standardized, in particular radar reflectivity being always denoted by `Z` (rather than `Zh`) and comment attributes by `comment` (rather than `comments`). Variables represented by two-byte integers, with the NetCDF attributes `scale_factor` and `add_offset` to convert them into their correct units, are "expanded" into floating-point representation; this applies to lidar backscatter β'.

3.2 Regridding

A "universal" grid is defined which consists essentially of the radar time and height grid. However, if there are any temporal gaps in the model or lidar data, or if the model or lidar height grids start higher or end lower than the radar height grid, then the universal grid will be reduced so that there is always a model and lidar pixel in the vicinity of a radar pixel. The lidar and model variables (β', T, p, q, κ_g, κ_{gs}, κ_l) are then interpolated on to the universal grid. In the case of the model data the interpolation is linear but for the lidar we wish to conserve the integrated backscatter as it is useful for calibration (O'Connor et al. 2004) and estimating optical depth in certain situations (Hogan et al. 2003b). This is done by using nearest-neighbour interpolation in time followed by integration in height, interpolation of the integral on to the universal grid, and differentiation.

The one-dimensional fields rain rate and liquid water path are interpolated linearly on to the radar grid. Missing values are assigned as `NaN`. In the case of missing rain rates the radar is used to indicate the presence of rain at the ground and hence when the reflectivity values may be unreliable (see section 3.3.1), while in the case of missing liquid water path the attenuation flags are used to indicate pixels when liquid water below them is likely to have caused attenuation that has not been corrected (see sections 3.3.3 and 3.5.2). The result is a set of 2D fields that share the same grid, and a few 1D fields that share the time grid of the 2D fields.

3.3 Data quality flags

Before the target categorization algorithm can be applied a certain amount of quality control is performed, including diagnosis of the likelyhood of rain at the ground and identifying radar pixels affected by ground clutter. The resulting data quality information is presented as the time-height bit field `quality_bits` in the final product, where the bits have the following interpretation:

Bit 0: An echo is detected by the radar.

Bit 1: An echo is detected by the lidar.

Bit 2: The apparent echo detected by the radar is ground clutter or some other non-atmospheric artifact.

Bit 3: The echo detected by the lidar is due to clear-air molecular scattering.

Bit 4: Liquid water cloud or rainfall below this pixel will have caused radar and lidar attenuation; if bit 5 is set then a correction for the radar attenuation has been performed; otherwise do not trust the absolute values of reflectivity factor. No correction is performed for lidar attenuation.

Bit 5: Radar reflectivity has been corrected for liquid-water attenuation using the microwave radiometer measurements of liquid water path and the lidar estimation of the location of liquid water cloud; be aware that errors in reflectivity may result.

We now describe the criteria that are used to determine the quality of each pixel.

3.3.1 1-D Rain bit

Rain at the ground can wet the radar dish or radome, causing strong (and unknown) attenuation and therefore rendering the reflectivity values above unreliable. Additionally, the raindrops themselves can cause attenuation that is not well estimated using the technique for liquid clouds described in section 3.5.2, as the attenuating particles will be outside the Rayleigh size regime and attenuation may no longer be proportional to liquid water content. Furthermore, it is difficult to partition the measured liquid water path (which may itself be affected by water on the instrument) with height.

To facilitate the formulation of the 2-D attenuation flags which indicate regions of corrected and uncorrected attenuation, a 1-D bit is first generated to indicate that the reflectivity values in the profile may be unreliable due to rain at the ground. If rain gauge data are available then the bit is set to unity for all times when the rain gauge indicates non-zero rain rate. If rain gauge data are not available then the rain rate variable is instead created from the radar information, and consists of zeros when the radar indicates no rain at the ground (when Z is less than 0 dBZ in the third range gate above the ground) and `NaN` when the radar indicates the likelihood of rain at the ground but is unable to accurately estimate the rate (when Z is greater than 0 dBZ, corresponding to a rain rate of around 0.05 mm h^{-1}). The 1-D rain bit is then set to unity whenever rain rate is `NaN`.

Finally, all pixels within 2 minutes of rain are also deemed to be raining in the 1-D rain bit, and additionally any clear spells between rain bits that are less than 2 minutes long are also set to raining.

3.3.2 Ground clutter

The characteristics of the ground clutter are strongly instrument dependent, and two algorithms are available to identify ground clutter. Hereafter Z refers to radar reflectivity with clutter pixels removed, and Z_c refers to reflectivity with clutter untouched. Note that in the final product it is Z_c that is reported, but Bit 2 of the `quality_bits` bit field indicates the location of probable radar clutter so that it may be removed.

The most robust ground clutter algorithm we use makes use of the fact that ground clutter tends to have a Doppler velocity close to zero. The whole day is analysed at a time, but only the first 10 gates are considered for ground-clutter removal. If rain is detected at the ground then it is assumed that the return from the lowest gates will be dominated by rain and no attempt is made to identify clutter in these profiles. At each height starting with the lowest, pixels are deemed to be dominated by clutter whenever the rain bit is not set, v lies between -0.05 and $+0.05$ m s^{-1}, *and* $\sigma_{\bar{v}}$ is less than 0.2 m s^{-1}. If a height is reached where no clutter is found at any time in the day then the gates above are not analysed.

If $\sigma_{\bar{v}}$ is not present then an alternative algorithm is used, although it is so ugly that it will not be described here.

Note that the 905-nm Vaisala lidars do not detect molecular scattering so Bit 3 of `quality_bits` is not set by them. However, the ARM lidars operate at 532 nm and do detect molecular scattering to some extent.

3.3.3 Attenuation bits

Two bits are provided to indicate the radar pixels that have been affected by attenuation and those for which a correction has been made. The first (Bit 4 of `quality_bits`) indicates that the radar and lidar returns have been attenuated by liquid water cloud, rain and/or melting ice in the intervening pixels, while the second (Bit 5 of `quality_bits`) indicates that microwave radiometer estimates of liquid water path have been used to correct this attenuation, but that errors in reflectivity may result. These bits are calculated after liquid attenuation correction has been performed, so we defer further descussion until section 3.5.2.

3.4 Target categorization

The type of target(s) present in each pixel is important for the application of subsequent algorithms, and is diagnosed using all the data available. As several target types may be present in a given pixel, this information is also presented in the form of a bit field, with each bit typically representing a different type of target. Of course, some types of target can never be identified when others are present (such as aerosol when liquid water is present), but this format is intended to allow the maximum flexibility for when different types can be identified simultaneously.

There are separate bits diagnosing the presence of liquid cloud droplets, melting ice particles, aerosols and insects. Two further bits are used to define liquid precipitation and ice: `falling_bit`, indicating particles that have appreciable terminal velocity, and `cold_bit`, indicating whether such falling particles are likely to be composed of ice or liquid water. In this simple scheme we are accepting that there is no distinct difference between ice cloud and ice precipitation (at least, not one that can be discerned from the observations; see Hogan et al. 2001), and also that we are unable to distinguish supercooled drizzle from ice, although this may be possible in future. We also do not distinguish between rain originating from melting ice, and drizzle originating from the warm rain process, although it would be a simple matter to do that from the final target categorization bit field. The format of the resulting `category_bits` variable is as follows:

Bit 0: Small liquid droplets are present (`droplet_bit`).

Bit 1: Falling hydrometeors are present; if Bit 2 is set then these are most likely to be ice particles otherwise they are drizzle or rain drops (`falling_bit`).

Bit 2: Wet-bulb temperature is less than 0°C, implying the phase of Bit-1 particles (`cold_bit`).

Bit 3: Melting ice particles are present (`melting_bit`).

Bit 4: Aerosol particles are present and visible to the lidar (`aerosol_bit`).

Bit 5: Insects are present and visible to the radar (`insect_bit`).

One would use `category_bits` to diagnose the presence of cloud as when Bit 0 is set (indicating the presence of liquid water droplets), or both Bits 1 and 2 were present (indicating the presence of ice particles). In C this would be implemented as follows:

```
char droplet_bit = ((category_bits & (1<<0)) > 0);
char falling_bit = ((category_bits & (1<<1)) > 0);
char cold_bit    = ((category_bits & (1<<2)) > 0);
char is_cloud = droplet_bit || (falling_bit && cold_bit);
```

For convenience, *Cloudnet* also provides a simpler "classification" product at Level 2, consisting of numbers from 0 to 10 indicating the main combinations possible in the more complex "target categorization" bit field described above. It is important to stress that the intention with these two products is not to provide automated cloud classification in the classical sense (e.g. to distinguish cirrocumulus from cirrostratus), but to use criteria that are objectively defined from the measurements, useful for subsequent retrieval algorithms and which match some of the distinctions that are made in numerical forecast models.

We now describe the categorization algorithm in more detail, considering each bit in turn. The following sections describe the separate functions (some of which calculate more than one bit) and are headed by the input variables that they make use of. It should be noted that many of the procedures contain seemingly arbitrary parameters; these have been chosen to produce the best agreement with a subjective analysis of real cases. The order in which the bits are described matches the order in which they are calculated in the program.

3.4.1 Cold bit, melting bit

Input fields: Model wet bulb temperature T_w, Doppler velocity v, folding velocity v_{fold}.

The purpose of `cold_bit` to indicate where "falling" particles are likely to composed of ice rather than liquid. It is initially defined to be where the wet-bulb temperature T_w (calculated from model temperature, pressure and humidity) is less than 0°C (note that falling ice melts when T_w, rather than T, becomes positive). To cope with isothermal layers, this is actually implemented such that all pixels below the highest 0°C isotherm in the profile are deemed to be "warm" (`cold_bit` equals zero), since melted ice precipitation is unlikely to refreeze.

This field is then refined using the radar to locate the melting layer more precisely in stratiform precipitation. The radar reflectivity profile usually provides a distinct step at 94 GHz where ice particles melt (see Mittermaier and Illingworth 2003) and a bright band at lower frequencies. However, the Doppler profile provides a more distinct signal, with a large and sharp increase in fall speed at the point of melting, so this is the parameter that we use in our algorithm.[7]

Firstly, to produce the most likely velocity field for rain, any pixels with an updraft greater than 1 m s^{-1} have $2v_{fold}$ subtracted from them temporarily. Using centred finite differences, the vertical "divergence" of the Doppler velocity (i.e. $\partial v/\partial z$) is calculated for each pixel in the range $-5°C < T_w < +5°C$. Any pairs of gates (two pixels apart for centred differencing) that have a difference in v greater than v_{fold} or less than $-v_{fold}$ are assumed to be affected by folding, so $2v_{fold}$ is added to or subtracted from the difference, in order that the divergence lies within the most probable range. Hence this algorithm should still work for radars with low folding velocities (less than 4 m s^{-1}).

[7] *Future changes: have a reflectivity version of melting layer detection if Doppler velocity is not present or unreliable, and an LDR version if such capability is available.*

The height of the maximum divergence in this T_w range is found, and all contiguous pixels for which the divergence exceeds 0.0075 s^{-1} within ±150 m of this height are deemed to be "melting" (`melting_bit` set to unity). As the maximum divergence exceeding 0.0075 s^{-1} will not always mark out the height of the correct melting layer, spurious jumps in the height of the diagnosed melting layer are sometimes evident. We therefore apply other criteria in an attempt to reject these spurious rays, and interpolate between the rays for which the diagnosis is most confident. Rays are rejected for which the Doppler velocity at the peak in divergence is less than 0.5 m s^{-1} downwards, or if the rays to either side do not register the presence of a melting layer. Neighbouring profiles and next-to-neighbouring profiles are compared, and where the height of the peak in divergence differs by more than 150 m then both are rejected.

In those profiles in which melting pixels still remain, the `cold_bit` profile is adjusted such that the topmost "melting" pixel is the highest pixel where `cold_bit` equals zero, and all those above are unity. The height at which `cold_bit` changes in the remaining profiles is determined as follows. For those within 1 hour of a profile containing melting pixels, the height is interpolated between the two nearest rays on each side containing melting pixels. For the next two hours there is a linear relaxation to the height where $T_w = 0°C$ in the model, and beyond 3 hours the model value is used. Finally, whenever the radar has a signal in the highest "warm" pixel, it is assigned to be melting.

3.4.2 Droplet bit

Input fields: Attenuated lidar backscatter coefficient β′, radar reflectivity factor Z, cold bit (defined above), temperature T.

Each lidar ray is examined in turn, and searched for one or more liquid layers. We utilise the fact that to lidar the base of liquid clouds appears as a strong echo that is confined over only a few hundred metres. Note that Hogan et al. (2003b) used the integrated backscatter through the layers to assist in their diagnosis; while this enabled the optical thickness necessary to trigger their algorithm to be defined, it only allowed one liquid layer to be identified in any given profile. Here we are interested in identifying pixels that, on the balance of probabilities, contain liquid droplets, so need not be so restrictive in our criteria.

The first liquid layer is found by locating the lowest pixel in the ray where both $\beta' > 2 \times 10^{-5}$ m^{-1} sr^{-1} and the β' value 250 m higher up is a factor of 10 lower; this is denoted the "pivot" value from now on. The maximum gate-to-gate increase in β' in the 100 m below the pivot, $\Delta\beta'$, is calculated. Liquid cloud base is defined as the lowest pixel in this 100 m range for which the difference in β' between it and the pixel above exceeds $\Delta\beta'/4$.

Lidar cloud top is defined as follows. If the lidar return falls to zero within 300 m above the pivot then the top is defined to be the last non-zero pixel just below this point. Otherwise the procedure is similar to cloud base; $\Delta\beta'$ is calculated as the maximum decrease in the gate-to-gate β' in the 300 m above the pivot, and cloud top is declared to be the highest pixel in this range where the decrease in β' from the pixel below exceeds $\Delta\beta'/4$. Then the radar profile is analysed to determine cloud top in the case that the lidar has been extinguished while the radar still has a signal. If we are in a sub-zero region (as determined by `cold_bit`), the radar is only searched a further 300 m above the lidar-diagnosed top; otherwise we search up to the last pixel where `cold_bit` is zero. If there are any radar pixels in this region in which no signal is detected, then cloud top is changed to be the pixel immediately beneath the first pixel where no radar signal is detected. If on the other hand there is a radar signal throughout this region, then it is regarded as ice or drizzle falling from further up in the profile. The use of rather arbitrary search distances sometimes results in erroneous liquid water profiles. This simply represents the difficulty in locating the tops of clouds when the radar is dominated by larger particles falling through them and the lidar signal has been extinguished.

The `droplet_bit` is then set to unity in all pixels between cloud base and top. The next layer is diagnosed by repeating this process, with the lidar profile searched above the pivot for the next pixel with $\beta' > 2 \times 10^{-5}$ m^{-1} s^{-1} to act as a new pivot. Note that if several contiguous pixels have β' exceeding this value then the pixels in the vicinity may be analysed several times to determine if they qualify for `droplet_bit` being set. Finally, the droplet bit is set to zero for any pixels with $T < -40°C$, as liquid water cannot persist at these temperatures.

3.4.3 Falling bit, insect bit

Input fields: Radar reflectivity without clutter Z, radar reflectivity with clutter Z_c, attenuated backscatter coefficient β′, cold bit and droplet bit (defined above), 1D rain bit.

The `falling_bit` incorporates rain, drizzle and all ice particles. Discrimination between ice and liquid is then possible using `cold_bit` defined in the previous section. The procedure is basically to assign all radar echos that have not been identified as clutter as "falling", then to remove drizzle-free liquid clouds and to reassign those due to insects using `insect_bit`.

We first consider profiles containing no liquid water droplets at any height and no radar echo in the lowest sub-zero pixel (the first in the profile with `cold_bit` = 1, indicating ice about to melt). In these profiles all pixels with a finite radar echo (using Z with clutter removed) that have `cold_bit` set are assigned to be "falling", while those in the warm region are assigned as "insects". However, if rain is detected at the ground (as indicated by the 1-D rain bit) then all radar echos in the profile are assigned as falling.

We next consider profiles containing liquid water droplets as indicated by `droplet_bit`. For the purposes of this algorithm, if a radar echo is received from the lowest pixel in the profile with `cold_bit` set, then this is also considered to be a (single-pixel) liquid cloud. The reason is that both liquid clouds and ice just on the verge of melting can be considered as sources of liquid precipitation, and the same methodology is used on the pixels below them to distinguish precipitation from insects. Thus the bases and tops of each of the liquid clouds in the profile are determined.

A simple method is then used to discriminate insects from drizzle beneath the first cloud base. If radar echos are recorded continuously between the ground and the first cloud base then the minimum reflectivity with clutter, Z_c, is found, and pixels above this are designated "falling" while those below are "insects". If the radar echo is not continuous then `falling_bit` is set only in the contiguous finite radar echos below cloud base; all finite radar echos below this are assigned as insects.

Next the pixels within the cloud are considered; the aim is to remove from `falling_bit` those for which the radar echo is predominantly due to liquid droplets. Firstly cloud top is examined. If there is a radar signal in the pixel above it then this implies that precipitation particles are falling into the liquid cloud, so all radar pixels within the cloud are deemed to be falling. If there is no radar signal immediately above cloud top then we must decide whether any radar echo within the cloud is predominantly due to the cloud itself or to growing drizzle or ice. As precipitation-free liquid water clouds tend to have liquid water content increasing with height, the reflectivity of such clouds also increases with height. Conversely, when precipitation is present it tends to grow due to accretion of liquid water, and Z decreases with height. We therefore compare the Z values 20% above cloud base and 20% below cloud top. If from these two pixels Z increases with height then the entire cloud is deemed to be free of precipitation and `falling_bit` is set to zero (although note that Z might be finite just below cloud base and so `falling_bit` might still be set here). If Z decreases with height then drizzle or ice is deemed to be present in the profile, and `falling_bit` is set to unity between cloud base and the highest pixel below cloud top where Z exceeds -30 dBZ.

This procedure is repeated for all liquid layers, although the insect/drizzle distinction is only performed for the first layer. This has the unfortunate consequence that any pixel with a finite radar echo above the first cloud base that is not within not within a cloud is declared to be "falling" even though subjectively it might be identified as insects.[8]

Some tenuous ice clouds are only detected by the lidar. These are added by declaring any pixel above 6000 m for which a lidar echo is received to be "falling", provided that `cold_bit` is set and `droplet_bit` is not set. Note that those pixels that satisfy these criteria but are below 6000 m will be set to aerosol.[9]

Finally a modification is made to `melting_bit`, which previously was set for all radar echos immediately beneath the lowest `cold_bit`. For those pixels that have been reassigned to insects, `melting_bit` is unset.

3.4.4 Aerosol bit

Input fields: Attenuated lidar backscatter coefficient β falling bit, droplet bit, cold bit.

A pixel is deemed to be aerosol (`aerosol_bit` is set) if a finite lidar signal is present, and it is at or below 6000 m (or `cold_bit` is not set), and neither `droplet_bit` nor `falling_bit` have been set.

3.5 Attenuation correction

The radar reflectivity profile is corrected for the effects of both liquid water and gas (predominantly water vapour and oxygen) attenuation. Ice attenuation in vertical profiles can be considered negligible at frequencies below 100 GHz. Two fields are generated on the same universal grid as used by the radar; the two-way gas attenuation and the two-way liquid attenuation, both in dB. These are simply added to the reflectivity field (in dBZ). They are also recorded as part of the product, so that the user may recover the actual measured Z field if desired.

3.5.1 Gas attenuation

The *Cloudnet* model data includes κ_g, the specific gas attenuation at the frequency of the radar calculated from the model temperature, pressure and humidity, and also κ_{gs}, the specific gas attenuation for 100% humidity with respect to liquid water. A "most likely" specific gas attenuation field is generated as a combination of the two, using `droplet_bit` to dictate when κ_{gs} should be used rather than κ_g. The result is integrated with height to obtain the total 2-way

[8] *Possible improvement would be to use radar LDR if available, or perhaps spectral width which also can be used to identify individual insects.*

[9] *Occasionally contiguous lidar-only signals can span the 6000 m level resulting in the unlikely situation of a plume of aerosol immediately beneath a tenuous ice cloud. In these situations some attempt should probably made to declare the phenomenon as entirely aerosol or entirely ice.*

gaseous attenuation, designated radar_gas_atten in the output NetCDF file.

3.5.2 Liquid water attenuation

Liquid water attenuation is estimated by partitioning the LWP measured by microwave radiometer with height amongst those pixels with droplet_bit set to obtain the "most likely" profile of liquid water content (LWC). This is done by considering each individual liquid water cloud (i.e. each contiguous sequence of pixels with droplet_bit = 1) and calculating the adiabatic profile of LWC using T and p at each cloud base (specifically LWC is assumed to increase linearly with height from the base of each liquid layer). Then the entire profile is scaled to match LWP. The resulting LWC is then multiplied by $2\kappa_l$ and integrated with height to obtain the cumulative 2-way liquid attenuation in that profile, designated radar_liquid_atten in the output NetCDF file.

If LWP reported by the microwave radiometers is zero or negative then no liquid water correction is performed; even if droplets are present then it is assumed that the LWC is so small that liquid attenuation is negligible. If no LWP is available then the liquid attenuation in the lowest pixel containing droplets, and all those above, will be set to NaN. The radar Z will not then be corrected for liquid attenuation.

At this point in the algorithm enough information is available to populate the two data quality bits referring to attenuation, discussed briefly in sections 3.3. The first (Bit 4) indicates if any radar pixel has been attenuated by either liquid water or rain; this bit is set if radar_liquid_atten is non-zero (either positive or NaN) or if the 1-D rain bit is set for that profile. The second (Bit 5) indicates if a correction has been performed for liquid attenuation, and is set if radar_liquid_atten is non-zero and non-NaN, *and* the 1-D rain bit is *not* set for that profile. Hence if no LWP measurements are available, pixels within and above liquid cloud would have the first of these bits set and the second unset, indicating that liquid attenuation correction has not been performed and the Z values should not be relied upon by subsequent retrieval algorithms.

3.6 Radar sensitivity

It is important to know the sensitivity of the radar in order to estimate the fraction of clouds that might not be detected. The variable Z_sensitivity indicates the approximate minimum detectable reflectivity, after correction for gas attenuation, as a function of height, estimated as follows. Firstly, the range normalisation of the radar is removed, i.e. $20\log_{10}(r)$ is subtracted from Z in dBZ (where r is range in km) to yield the return signal at 1 km. The minimum non-NaN value in this matrix provides an estimate of the minimum detectable signal, which is added to $20\log_{10}(r)$ to provide minimum detectable Z versus height. Then the effect of gas attenuation to reduce sensitivity is accounted for by adding the mean radar_gas_atten at each height. Finally the effect of ground clutter is included by identifying the lowest few gates in which ground clutter was detected (see section 3.3.2) and replacing the value obtained by the analysis so far with the median reflectivity of the clutter pixels at that height.

3.7 Instrument errors

The variables Z and β' are assigned a *bias* and an *error*. These should be used to generate a bias and an error for all subsequently retrieved parameters. Only an error is reported for LWP, since the temporal correlation of errors this variable means that it is not really meaningful to distinguish bias from random error. Biases and errors on other parameters are not reported.

The *bias* indicates the expected calibration accuracy of the parameter, i.e. the possible systematic errors on all values. This is estimated given knowledge of the calibration method for that particular instrument, and is typically a single value that applies to all Z or β' values recorded in a day. Typically the *Cloudnet* radars are calibrated by inter-comparison with a polarimetric weather radar, which is itself calibrated to 0.5 dB using the method of Goddard et al. (1994). Depending on the directness of the comparison, the bias will be between 1 and 2 dB. Radars at 94 GHz may also be calibrated in rain using the fact that 250 m from the antenna the combined effects of attenuation and Mie scattering render the reflectivity on average 19 dBZ for rain-rates between 3 and 10 mm h^{-1} (Hogan et al. 2003a). The accuracy in calibration depends on the number of rain events averaged over, and whether or not the radar was kept dry. The *Cloudnet* lidar ceilometers operate in the near infrared so cannot be calibrated using the molecular return. Instead we use the technique of O'Connor et al. (2004) and adjust the calibration such that the integral of the β' return through an optically thick liquid water cloud, averaged over many cases, equals the theoretical value. This approach is believed to be accurate to 5-10%.

The *error* indicates the expected one-standard-deviation random error, or precision, of a measurement. In the case of β' it is difficult to estimate as the internal processing carried out by the comercial Vaisala ceilometers is not known in detail, so a single value (typically 0.5 dB) is reported that applies to all pixels. In the case of LWP it is calculated as the root-mean-squared combination of a linear error (e.g. due to errors in radiometer calibration) and a fractional error (e.g. due to an error in the assumed cloud temperature). Typically the fractional error is taken to be 25% while the linear error is 20 g m^{-2} for retrievals that use lidar to correct offsets using clear sky regions and 50 g m^{-2} otherwise.

In the case of Z the error is calculated separately for each pixel, as the square-root of the sum of the variances from three different sources of error:

The natural precision of a Z measurement. Due to the fluctuating echo from pulse to pulse, it is necessary to average a large number of pulses from a cloud radar in order to reduce the error on the mean. In the case of a simple pulsed radar that performs only incoherent averages (i.e. no spectral processing) and no pulse compression, the expected error is easy to calculate (e.g. Doviak and Zrnić 1993), provided certain parameters of the radar and processing are known. When spectral processing is performed, this becomes rather more difficult and many more of the details of the processing need to be considered. For simplicity, we therefore calculate the error assuming that simple incoherent averaging is performed followed by noise subtraction and thresholding at 3 standard deviations above the noise. This yields the following error (e.g. Hogan 1998)

$$\Delta Z[\mathrm{dB}] = \frac{4.343}{\sqrt{N_\mathrm{i}}}\left[1 + \frac{10^{0.1(Z_\mathrm{min}[\mathrm{dBZ}]-Z[\mathrm{dBZ}])}}{3}\right],$$

where Z_{min} is the minimum detectable signal at that height and N_i is the total number of independent pulses measured in the dwell time τ_d (usually 30 s), which is given by (Atlas 1964)

$$N_\mathrm{i} = \frac{4\pi^{\frac{1}{2}}\tau_\mathrm{d}\sigma_v}{\lambda}, \qquad (1)$$

where λ is the radar wavelength. This has the property that at large signal-to-noise ratios (i.e. $Z \gg Z_\mathrm{min}$) the precision is determined by the number of independent pulses, but at low signal-to-noise ratios, the error rapidly increases.

The error in gas attenuation. The humidity profile from the model will not be perfect, resulting in an error due to the correction of the radar reflectivity for gas attenuation. We assume the error in Z from this source to be 10% of the 2-way cumulative gas attenuation to that height.

The error in liquid attenuation. The combination of error in LWP and the partitioning of liquid water with height results in an error in Z due to correction for liquid attenuation. This is estimated by following the same steps for computing the liquid water attenuation as was carried out in section 3.5.2, but using the *error* in LWP to scale the LWC profile rather than LWP itself. To account for the possibility that LWC does not increase linearly with height but has more of a symmetric profile, this calculation assumes constant LWC with height in each liquid cloud. As liquid water attenuation is much less at 35 GHz than 94 GHz, radars at 94 GHz have a considerably higher error in retrievals from ice clouds above liquid water clouds.

4 Conclusions

A data product has been described that performs much of the preprocessing that is necessary before synergetic radar/lidar algorithms can be applied, and should facilitate the development of cloud retrieval algorithms that can be applied to large volumes of data. The target categorization component should be straightforward to use in simpler studies just concerned with cloud boundaries, such as evaluation of cloud fraction in models and investigations into cloud overlap.

Appendix

A Data format

The dataset is stored in the self-describing format "NetCDF"[10], and conforms to both the *Cloudnet* conventions[11] and the "Climate and Forecasting" (CF) conventions[12]. This appendix outlines the format by presenting the meta-data from a particular day at one of the *Cloudnet* sites in the form of pseudo-`ncdump` output, with annotation where appropriate. For those not familiar with NetCDF, a NetCDF file first defines a number of *dimensions*, and the non-scalar *variables* must be defined in terms of these. Variables may be of a number of different types, although only 4-byte floating-point values (designated "`float`") and bytes are used in this product. Each variable may provide descriptive information about itself using *attributes*, which may be in the form of strings or vectors of numbers. The numerical attributes used in this product are either floating-point numbers (indicated by the suffix "`f`") or 2-byte signed integers (indicated by the suffix "`s`"). In addition, *global attributes* may be defined, containing information about the whole file.

A.1 Dimensions

Three dimensions are used:

[10] http://my.unidata.ucar.edu/content/software/netcdf/index.html
[11] http://www.met.rdg.ac.uk/radar/cloudnet/data/data_format/netcdf.html
[12] http://www.cgd.ucar.edu/cms/eaton/cf-metadata/

time = 2764
height = 191
model_height = 25

The dimension model_height is only used for the four variables taken directly from the model, temperature, pressure, uwind and vwind.

A.2 Variables

The following scalar variables are defined:

float altitude

altitude:units = *"m"*
altitude:long_name = *"Height of radar above mean sea level"*

float latitude

latitude:units = *"degrees_north"*
latitude:long_name = *"Latitude of site"*
latitude:standard_name = *"latitude"*

float longitude

longitude:units = *"degrees_east"*
longitude:long_name = *"Longitude of site"*
longitude:standard_name = *"longitude"*

float radar_frequency

radar_frequency:units = *"GHz"*
radar_frequency:long_name = *"Radar frequency"*

The following are the three "coordinate variables", one for each dimension:

float time(time)

time:axis = *"T"*
time:units = *"hours since 2004-07-13 00:00:00 +0:00"*
time:long_name = *"Time UTC"*
time:standard_name = *"time"*

float height(height)

height:axis = *"Z"*
height:units = *"m"*
height:long_name = *"Height above mean sea level"*
height:standard_name = *"height"*

float model_height(model_height)

model_height:units = *"m"*
model_height:long_name = *"Height of model variables above mean sea level"*

The following are the meteorological measurements and their associated errors and biases, with correction for attenuation where appropriate:

float rainrate(time)

rainrate:long_name = *"Rain rate"*
rainrate:units = *"mm h-1"*
rainrate:units_html = *"mm h⁻¹"*
rainrate:comment = *"Converted to rainrate from raingauge c drop counts"*
rainrate:plot_range = 0.f, 50.f
rainrate:plot_scale = *"linear"*
rainrate:missing_value = -999.f
rainrate:_FillValue = -999.f

float lwp(time)

lwp:long_name = *"Liquid water path"*
lwp:units = *"g m-2"*
lwp:units_html = *"g m⁻²"*
lwp:missing_value = -999.f
lwp:plot_range = -50.f, 500.f
lwp:plot_scale = *"linear"*

float Z(time, height)

Z:units = *"dBZ"*
Z:long_name = *"Radar reflectivity factor"*
Z:missing_value = -999.f
Z:calibration_applied = -20.5f
Z:comment = *"This variable has been corrected for attenuation by gaseous attenuation (using the thermodynamic variables from a forecast model; see the radar_gas_atten variable), but attenuation by liquid water clouds, rain and the melting layer has not been corrected. Calibration convention: in the absence of attenuation, a cloud at 273 K containing one million 100-micron droplets per cubic metre will have a reflectivity of 0 dBZ at all frequencies. Original comment: Calibration convention: in the absence of attenuation, a cloud at 273 K containing one million 100-micron droplets per cubic metre will have a reflectivity of 0 dBZ at all frequencies. To reduce speckle noise, any cloudy pixel that had cloud-free pixels to each side of it in range was removed."*
Z:source = *"Chilbolton 35-GHz Radar (Copernicus)"*
Z:error_variable = *"Z_error"*
Z:bias_variable = *"Z_bias"*
Z:plot_range = -40.f, 20.f
Z:plot_scale = *"linear"*

float v(time, height)

v:units = *"m s-1"*
v:long_name = *"Doppler velocity"*
v:units_html = *"m s⁻¹"*
v:missing_value = -999.f
v:folding_velocity = 5.f
v:comment = *"This parameter is the radial component of the velocity, with positive velocities towards the radar."*
v:source = *"Chilbolton 35-GHz Radar (Copernicus)"*
v:plot_range = -4.f, 2.f
v:plot_scale = *"linear"*

float width(time, height)

width:units = *"m s-1"*
width:long_name = *"Doppler spectral width"*
width:units_html = *"m s⁻¹"*
width:missing_value = -999.f
width:comment = *"This parameter is the standard deviation of the reflectivity-weighted velocities in the radar pulse volume."*
width:source = *"Chilbolton 35-GHz Radar (Copernicus)"*
width:plot_range = 0.03f, 3.f
width:plot_scale = *"logarithmic"*

float sigma_v(time, height)

sigma_v:units = *"m s-1"*
sigma_v:long_name = *"Standard deviation of mean velocity"*
sigma_v:units_html = *"m s⁻¹"*
sigma_v:missing_value = -999.f
sigma_v:comment = *"The data in this file are at a lower resolution than the raw data, and this parameter is the standard deviation of the 30 raw Doppler velocities measured during in each output gate and ray."*
sigma_v:source = *"Chilbolton 35-GHz Radar (Copernicus)"*
sigma_v:plot_range = 0.001f, 1.f
sigma_v:plot_scale = *"logarithmic"*

float Z_bias

Z_bias:units = *"dB"*
Z_bias:long_name = *"Calibration error in Z, one standard deviation"*

`Z_bias:comment` = *"This variable is an estimate of the one-standard-deviation calibration error (i.e. the likely systematic error) in radar reflectivity factor."*

`float Z_error(time, height)`

- `Z_error:units` = *"dB"*
- `Z_error:long_name` = *"Random error in Z, one standard deviation"*
- `Z_error:missing_value = -999.f`
- `Z_error:plot_range = 0.f, 3.f`
- `Z_error:plot_scale` = *"linear"*
- `Z_error:comment` = *"This variable is an estimate of the one-standard-deviation random error in radar reflectivity factor. It originates from the following independent sources of error:*
  1) Precision in reflectivity estimate due to finite signal to noise and finite number of pulses
  2) 10% uncertainty in gaseous attenuation correction (mainly due to error in model humidity field).
  Note that liquid water attenuation has not been corrected for."

`float beta(time, height)`

- `beta:units` = *"m-1 sr-1"*
- `beta:units_html` = *"m⁻¹ sr⁻¹"*
- `beta:long_name` = *"Attenuated backscatter coefficient"*
- `beta:comment` = *"This variable has not been corrected for attenuation."*
- `beta:missing_value = 0.f`
- `beta:_FillValue = 0.f`
- `beta:source` = *"Chilbolton Vaisala 905-nm CT75K lidar ceilometer"*
- `beta:error_variable` = *"beta_error"*
- `beta:bias_variable` = *"beta_bias"*
- `beta:plot_range = 1.e-07f, 1.e-04f`
- `beta:plot_scale` = *"logarithmic"*

`float beta_bias`

- `beta_bias:units` = *"dB"*
- `beta_bias:long_name` = *"Calibration error in beta, one standard deviation"*
- `beta_bias:comment` = *"This variable is an estimate of the one-standard-deviation calibration error (i.e. the likely systematic error) in attenuated lidar backscatter coefficient."*

`float beta_error`

- `beta_error:units` = *"dB"*
- `beta_error:long_name` = *"Random error in beta, one standard deviation"*
- `beta_error:comment` = *"This variable is a very approximate esimate of the one-standard-deviation random error in attenuated lidar backscatter coefficient. It should really take account of signal-to-noise ratio, number of pulses averaged and so on, but the exact algorithm used to calculate the reported backscatter values is proprietary."*

The following are the parameters taken directly from the model, but averaged on to the universal time axis:

`float temperature(time, model_height)`

- `temperature:units` = *"K"*
- `temperature:long_name` = *"Temperature"*
- `temperature:standard_name` = *"air_temperature"*
- `temperature:C_format` = *"%.2f"*
- `temperature:_FillValue = -999.f`
- `temperature:missing_value = -999.f`
- `temperature:source` = *"UK Met Office Unified Model (Mesoscale)"*
- `temperature:plot_range = 200.f, 300.f`
- `temperature:plot_scale` = *"linear"*

`float pressure(time, model_height)`

- `pressure:units` = *"Pa"*
- `pressure:long_name` = *"Pressure"*
- `pressure:standard_name` = *"air_pressure"*
- `pressure:C_format` = *"%.0f"*
- `pressure:_FillValue = -999.f`
- `pressure:missing_value = -999.f`
- `pressure:source` = *"UK Met Office Unified Model (Mesoscale)"*
- `pressure:plot_range = 0.f, 110000.f`
- `pressure:plot_scale` = *"linear"*

`float uwind(time, model_height)`

- `uwind:units` = *"m s-1"*
- `uwind:long_name` = *"Zonal wind"*
- `uwind:standard_name` = *"eastward_wind"*
- `uwind:C_format` = *"%.6f"*
- `uwind:_FillValue = -999.f`
- `uwind:missing_value = -999.f`
- `uwind:source` = *"UK Met Office Unified Model (Mesoscale)"*
- `uwind:plot_range = -50.f, 50.f`
- `uwind:plot_scale` = *"linear"*

`float vwind(time, model_height)`

- `vwind:units` = *"m s-1"*
- `vwind:long_name` = *"Meridional wind"*
- `vwind:standard_name` = *"northward_wind"*
- `vwind:C_format` = *"%.6f"*
- `vwind:_FillValue = -999.f`
- `vwind:missing_value = -999.f`
- `vwind:source` = *"UK Met Office Unified Model (Mesoscale)"*
- `vwind:plot_range = -50.f, 50.f`
- `vwind:plot_scale` = *"linear"*

The following are the attenuations described in section 3.5 that have been applied to radar reflectivity factor:

`float radar_gas_atten(time, height)`

- `radar_gas_atten:units` = *"dB"*
- `radar_gas_atten:long_name` = *"Two-way radar attenuation due to atmospheric gases"*
- `radar_gas_atten:plot_range = 0.f, 4.f`
- `radar_gas_atten:plot_scale` = *"linear"*
- `radar_gas_atten:comment` = *"This variable was calculated from the model temperature, pressure and humidity, using the millimeter-wave propagation model of Liebe (1985, Radio Sci. 20(5), 1069-1089). It has been used to correct Z."*

`float radar_liquid_atten(time, height)`

- `radar_liquid_atten:units` = *"dB"*
- `radar_liquid_atten:missing_value = -999.f`
- `radar_liquid_atten:plot_range = 0.f, 4.f`
- `radar_liquid_atten:plot_scale` = *"linear"*
- `radar_liquid_atten:long_name` = *"Approximate two-way radar attenuation due to liquid water"*
- `radar_liquid_atten:comment` = *"This variable was calculated from the liquid water path measured by microwave radiometer, using the lidar and radar returns to perform an approximate partitioning of the liquid water content with height. The quality_bits variable indicates where a correction for liquid water attenuation has been performed. The dielectric parameters of liquid water were calculated using the double-Debye formulation of Manabe, Liebe and Hufford (1987, Conf. Dig. 12th Int. Conf Infrared & Millimeter Waves, Lake Buena Vista, Dec. 14-18); see also Liebe, Manabe and Hufford (1989, IEEE Trans. AP 37(12), 1617-1623)."*

The following are the two bit fields providing target categorization and data quality information:

`byte category_bits(time, height)`

`category_bits:long_name =` *"Target categorization bits"*

`category_bits:comment =` *"This variable contains information on the nature of the targets at each pixel, thereby facilitating the application of algorithms that work with only one type of target. The information is in the form of an array of bits, each of which states either whether a certain type of particle is present (e.g. aerosols), or the whether some of the target particles have a particular property. The definitions of each bit are given in the definition attribute. Bit 0 is the least significant."*

`category_bits:definition =`

"Bit 0: Small liquid droplets are present.
Bit 1: Falling hydrometeors are present; if Bit 2 is set then these are most likely to be ice particles, otherwise they are drizzle or rain drops.
Bit 2: Wet-bulb temperature is less than 0 degrees C, implying the phase of Bit-1 particles.
Bit 3: Melting ice particles are present.
Bit 4: Aerosol particles are present and visible to the lidar.
Bit 5: Insects are present and visible to the radar."

`byte quality_bits(time, height)`

`quality_bits:long_name =` *"Data quality bits"*

`quality_bits:comment =` *"This variable contains information on the quality of the data at each pixel. The information is in the form of an array of bits, and the definitions of each bit are given in the definition attribute. Bit 0 is the least significant."*

`quality_bits:definition =`

"Bit 0: An echo is detected by the radar.
Bit 1: An echo is detected by the lidar.
Bit 2: The apparent echo detected by the radar is ground clutter or some other non-atmospheric artifact.
Bit 3: The echo detected by the lidar is due to clear-air molecular scattering.
Bit 4: Liquid water cloud or rainfall below this pixel will have caused radar and lidar attenuation; if bit 5 is set then a correction for the radar attenuation has been performed; otherwise do not trust the absolute values of reflectivity factor. No correction is performed for lidar attenuation.
Bit 5: Radar reflectivity has been corrected for liquid-water attenuation using the microwave radiometer measurements of liquid water path and the lidar estimation of the location of liquid water cloud; be aware that errors in reflectivity may result."

A.3 Global attributes

The following are the global attributes contained in the NetCDF file. Note that the `history` and `source` attributes essentially consist of the contents of these attributes in the various source datasets.

`:Conventions =` *"CF-1.0"*

`:location =` *"Chilbolton"*

`:title =` *"Cloud categorization products from Chilbolton, 2004-07-13"*

`:day = 13s`

`:month = 7s`

`:year = 2004s`

`:history =` *"16 Jul 2004 13:16:15 - Generated from level 1 data by Ewan O'Connor <e.j.oconnor@reading.ac.uk>*
Radar history: Wed Jul 14 01:45:02 2004 - NetCDF generated from original data by dnl on wilma
Lidar history: Fri Jul 16 12:26:27 2004 - NetCDF generated from original data by Ewan O'Connor <e.j.oconnor@reading.ac.uk> on hogwarts
Model history: Thu Jul 15 02:06:02 BST 2004 - NetCDF generated from original data by Ewan O'Connor <e.j.oconnor@reading.ac.uk> using cnmodel2nc on hogwarts
Thu Jul 15 02:06:05 2004 - Comments added by radar on hogwarts
Gauge history: Recorded using Microlink 3000 series DAQ
2004-07-16 04:28:21 : converted to netCDF from FORMAT5 using Matlab 6.5.0.180913a (R13) running on GLNX86"

`:source =` *"Chilbolton 35-GHz Radar (Copernicus)*
Frequency: 34.960 GHz
Antenna diameter: 2.4 m
Peak power: 1.5 kW
Pulse width: 0.5 us
Pulse repetition frequency: 5000 Hz
Beam width: 0.25 degrees;
Chilbolton Vaisala 905-nm CT75K lidar ceilometer
Wavelength: 905 nm
Half-angle beam divergence: 0.75 mrad
Half-angle field of view: 0.66 mrad;
UK Met Office Unified Model (Mesoscale);
meteorological sensors"

`:institution =` *"Data produced at Department of Meteorology, University of Reading, UK"*

`:reference =` *"Documentation may be found at http://www.met.rdg.ac.uk/radar/doc/categorization.html"*

`:software_version =` *"0.4"*

`:comment =` *"This dataset is an aggregation of data from cloud radar, lidar, rain gauge, a numerical forecast model and optionally a microwave radiometer. It is intended to facilitate the application of synergistic cloud-retrieval algorithms by performing a number of the preprocessing tasks that are common to these algorithms. Each of the observational datasets has been interpolated on to the same grid, although the model data are provided on a reduced height grid. Radar reflectivity has been corrected for attenuation,where possible, and two additional fields have been added: "category_bits" contains a categorization of the targets in each pixel and "quality_bits" indicates the quality of the data at each pixel. Finally, estimates of the random and systematic errors in reflectivity factor and attenuated backscatter are provided."*

References

Atlas, D., 1964: Advances in radar meteorology. *Advances in Geophys.*, **10,** 318–478.

Bouniol, D., A. J. Illingworth and R. J. Hogan, 2003: Deriving turbulent kinetic energy dissipation rate within clouds using ground based 94 GHz radar. *Proc. 31st AMS Conf. on Radar Meteorology, Seattle,* 193–196.

Clothiaux, E. E., K. P. Moran, B. E. Martner, T. P. Ackerman, G. G. Mace, T. Uttal, J. H. Mather, K. B. Widener, M. A. Miller and D. J. Rodriguez, 1999: The Atmospheric Radiation Measurement program cloud radars: Operational modes. *J. Atmos. Oceanic Technol.,* **16,** 819–827.

Donovan, D. P., A. C. A. P. van Lammeren, H. W. J. Russchenberg, A. Apituley, R. J. Hogan, P. N. Francis, J. Testud, J. Pelon, M. Quante and J. W. F. Goddard, 2001: Cloud effective particle size and water content profi le retrievals using combined lidar and radar observations - 2. Comparison with IR radiometer and in situ measurements of ice clouds. *J. Geophys. Res.,* **106,** 27 449–27 464.

Doviak, R. J., and D. S. Zrnić, 1993: *Doppler radar and weather observations.* 2nd Ed. Academic Press.

Gaussiat, N., and co-authors, 2004: Refinement of microwave radiometer retrievals of liquid water path using additional information from lidar and forecast model. *Document in preparation*.

Goddard, J. W. F., J. Tan and M. Thurai, 1994: Technique for calibration of meteorological radars using differential phase. *Electronics Letters,* **30,** 166–167.

Hogan, R. J., 1998: *Dual-wavelength radar studies of clouds.* PhD Thesis, University of Reading, UK.

Hogan, R. J., A. J. Illingworth and H. Sauvageot, 2000: Measuring crystal size in cirrus using 35 and 94-GHz radars. *J. Atmos. Oceanic Technol.,* **17,** 27-37.

Hogan, R. J., C. Jakob and A. J. Illingworth, 2001: Comparison of ECMWF cloud fraction with radar-derived values. *J. Appl. Meteorol.,* **40,** 513–525.

Hogan, R. J., D. Bouniol, D. N. Ladd, E. J. O'Connor and A. J. Illingworth, 2003a: Absolute calibration of 94/95-GHz radars using rain. *J. Atmos. Oceanic Technol.,* **20,** 572–580.

Hogan, R. J., A. J. Illingworth, E. J. O'Connor and J. P. V. Poiares Baptista, 2003b: Characteristics of mixed-phase clouds - 2. A climatology from ground-based lidar. *Quart. J. Roy. Meteorol. Soc.,* **129,** 2117–2134.

Hogan, R. J., N. Gaussiat and A. J. Illingworth, 2004: Stratocumulus liquid water content from dual-wavelength radar. *Submitted to J. Atmos. Oceanic Technol*.

Liebe, H. J., 1985: An updated model for millimeter-wave propagation in moist air. *Radio Science,* **20,** 1069–1089.

Liebe, H. J., T. Manabe and G. A. Hufford, 1989: Millimeter-wave attenuation and delay rates due to fog/cloud conditions. *IEEE AP,* **37,** 1617–1623.

Mace, G. G., C. Jakob and K. P. Moran, 1998: Validation of hydrometeor occurrence predicted by the ECMWF using millimeter wave radar data. *Geophys. Res. Lett.,* **25,** 1645–1648.

Mittermaier, M. P., and A. J. Illingworth, 2003: Comparison of model-derived and radar-observed freezing level heights: Implications for vertical reflectivity profile correction schemes. *Quart. J. Roy. Meteorol. Soc.,* **129,** 83–96.

O'Connor, E. J., R. J. Hogan and A. J. Illingworth, 2004: Retrieving stratocumulus drizzle parameters using Doppler radar and lidar. *J. Appl. Meteorol.,* in press.

Openshaw, A., 2004: The Cloudnet common model format. *University of Reading Radar Group internal document*.

Sekelsky, S. M., W. L. Ecklund, J. M. Firda, K. S. Gage and R. E. McIntosh, 1999: Particle size estimation in ice-phase clouds using multi-frequency radar reflectivity measurements at 95, 33 and 2.8 GHz. *J. Appl. Meteorol.,* **38,** 5–28.