Milestone MS10.10: Description of model data format and interface for model data evaluation

Ewan O'Connor (FMI)

Work package no	**WP10**
Milestone no.	**MS10.10**
Lead beneficiary	**FMI**
Deliverable type	☑ R (Document, report) ☐ DEC (Websites, patent fillings, videos, etc.) ☐ OTHER: please specify ....
Dissemination level	☑ PU (public) ☐ CO (confidential, only for members of the Consortium, incl Commission)
Estimated delivery date	**Month 24**
Actual delivery date	**30/04/2017**
Version	
Comments	

ACTRIS (www.actris.eu) is supported by the European Commission under the Horizon 2020 – Research and Innovation Framework Programme, H2020-INFRAIA-2014-2015, Grant Agreement number: 654109

Purpose and objective

A major objective for Cloudnet is the routine automated evaluation of the representation of clouds in numerical models using observations derived from a combination of ground-based remote-sensing instruments (Illingworth et al., 2007). Evaluating the representation of clouds in climate and numerical weather prediction (NWP) models is not straightforward. For NWP models, this task is compounded by the expectation of a good forecast, as well as the reliable representation of the specific cloud parameters themselves. Cloudnet has developed and implemented a comprehensive suite set of objective metrics for the evaluation of model cloud parameters, in continual joint collaboration with operational modellers (Morcrette et al., 2012). The set of evaluation metrics is designed to investigate both the climatological aspects required of a climate model, and the ability to forecast the correct cloud at the right time, a necessary validation for NWP.

Routine evaluation requires that input data adheres rigorously to a pre-defined format; a documented description of the input and interface is necessary to enable automation and reliability. Within ACTRIS-2, the aim is to build an interface that will be capable of near-real-time (NRT) evaluation so that NWP modellers receive rapid feedback. In addition, the interface will support the evaluation of multiple model versions to investigate the impact of different physical parametrizations undergoing testing. Due to the rapid update schedule that NWP employs, evaluating multiple model versions and providing feedback must be responsive on similar timescales to NWP update schedules if it is to provide the expected benefit.

The purpose of this document is to describe the interface, model data format and required model parameters to enable routine automated evaluation.

ACTRIS (www.actris.eu) is supported by the European Commission under the Horizon 2020 – Research and Innovation Framework Programme, H2020-INFRAIA-2014-2015, Grant Agreement number: 654109

Cloudnet scheme

The Cloudnet concept is outlined in Fig. 1. **Level 1** deals with the processing of observations from different instruments and their subsequent combination to provide a single synergistic product **(Level 1c**) on a well-defined time-height grid. All individual observations are pre-processed, quality-checked, and Cloudnet-formatted at stage **Level 1b**.

Level 1c is the basis from which all Cloudnet products are created. High resolution products, at the native instrument resolution if possible, are created in **Level 2a**, and are used for all scientific studies.

Specific products for model evaluation are created in **Level 2b**, where the high resolution products are averaged onto the grid of each individual model. Model evaluation products are produced in **level 3** and beyond. Ingestion of NWP model data in **Level 2b** requires that it has already been transformed into Cloudnet format, in **Level 1b**.

In **Level 2b**, temporal averaging is used to create the equivalent of a two-dimensional slice through the three-dimensional model grid box. This temporal averaging can be performed in a number of ways, including a simple temporal averaging of one hour for the entire profile, and more sophisticated methods. As discussed in Illingworth et al. (2007), an appropriate amount of averaging time is given by the advective timescale, which describes how long it takes for a cloud structure to advect through the grid box and is given by the horizontal wind speed. The advective timescale is obtained as a function of height from the vertical profile of horizontal wind speed taken from the model itself, radiosonde, or a combination of radar and Doppler lidar wind profilers. Thus, given the 16 km horizontal resolution of the current ECMWF model grid, for example, a wind speed of 20 m s^{-1} corresponds to an averaging time of 800 seconds (13 minutes), centred on the model timestep.

Note that, since the wind varies with both height and time, the advective timescale, and thus the number of points averaged, changes with height and time. As each model has different horizontal and vertical resolutions, a separate level-2b product is produced for each model.

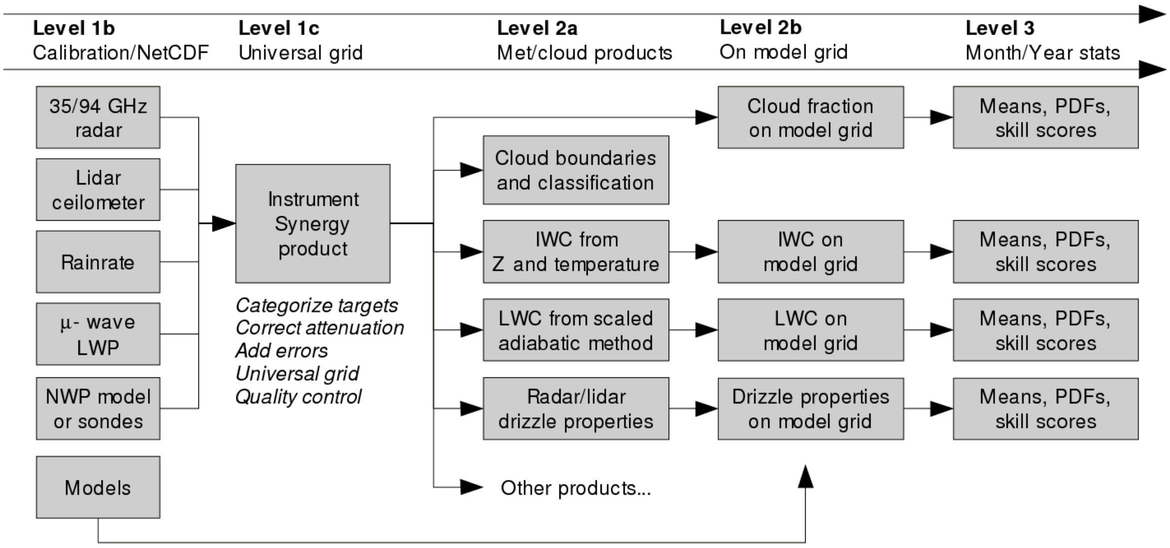

Figure 1: Overview of Cloudnet processing chain

ACTRIS (www.actris.eu) is supported by the European Commission under the Horizon 2020 – Research and Innovation Framework Programme, H2020-INFRAIA-2014-2015, Grant Agreement number: 654109

Cloudnet model data format

There are a number of evaluation methodologies that have been developed within the Cloudnet community including: climatological means, distributions; seasonal and diurnal composites; joint-pdfs for creating the contingency tables used for deriving the skill score of choice. Thus, from the files created in **Level 2b**, a wide range of metrics can be routinely plotted and analysed.

A standardised format for model data is a prerequisite to enable these evaluation methods to be employed; hence a standardised format was designed for model data within Cloudnet (see Appendix A). Post-processing of model data ingested within Cloudnet ensures harmonisation of variable names, units, and meta-data. Forecast concatenation is also performed **within** Cloudnet.

Forecast concatenation

An important part of the post-processing is the ability to concatenate a set of sequential forecasts to enable rapid evaluation of how the forecast degrades with time (Hogan et al., 2009), together with investigating the influence of model spin-up, and data assimilation/analysis-preferred climate versus free-running model climate. Note that the ability to create these particular metrics depends on the particular forecast output supplied by the modelling centre, and how long the model forecast runs for.

The concatenation scheme outlined in Fig. 2 shows how multiple forecast runs are combined to create daily files with specific forecast lead-times. The example in Fig. 3 shows a concatenation scheme for a model initialised 4 times a day (at 00Z, 06Z, 12Z, 18Z) and providing forecast output out to t+60 hours. This enables a set of daily files to be created containing data from a 6-hour forecast window. Certain models are initialised twice a day with each forecast running out to 72 hours and more, whereas others, especially high-resolution models designed to capture high-impact weather, are often initialised more often (maybe 6-8 times a day) but only run for 24 hours.

Model data provide in hourly or three-hourly temporal resolution is currently catered for, with higher resolution (15 minutes) also possible.

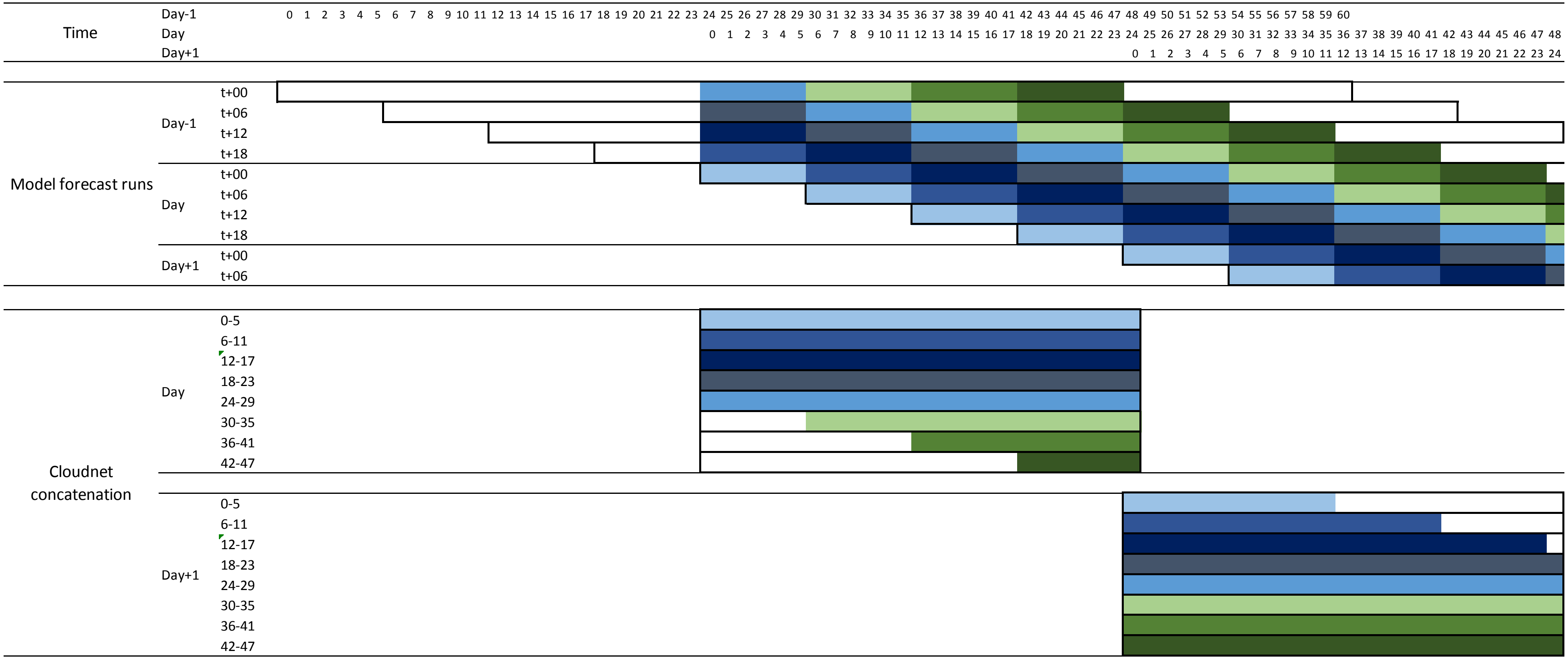

Figure 2: Model forecast concatenation within Cloudnet (for NWP model initialised 4 times a day: 00Z, 06Z, 12Z, 18Z). Each colour refers to a particular forecast window of 6 hours duration given in the 3rd column for the "Cloudnet concatenation" rows. These forecast windows are selected from sequential model forecast runs initialised every 6 hours as shown in the third column for rows in "Model forecast runs".

Model data requirements

For ingestion into the Cloudnet processing scheme, and subsequent evaluation, the input model data must meet the following requirements. Certain information must be present in the files or, if static, be communicated to Cloudnet.

Data structure comprises a set of vertical profiles of model variables from a forecast run, i.e. dimensions of time (*time* since forecast initialised) and height (*level*). It is permissible to combine multiple locations within one file using an additional dimension (*location*), as long as locations are specified – either by an accompanying latitude/longitude variable, or a location string.

Filenames should be consistent and use a set of appropriate descriptors:

[YYYYMMDD]	Date
[HH]	Forecast initialisation time
[model_id]	Model descriptor
[version_id]	Version descriptor
[param_id]	Parametrization descriptor
[site]	Site descriptor

in an agreed order. Filenames can contain other text as long as it cannot be mistaken for one of these descriptors. The use of all descriptors is not mandatory.

Mandatory information consists of:

scalar:	time of model initialisation
	horizontal resolution
vector:	forecast time
	surface pressure or geopotential
profile:	temperature
	pressure
	specific humidity
	horizontal wind

Horizontal winds and horizontal resolution are necessary to generate the advective timescale for observation averaging (Hogan et al., 2001; Brooks et al., 2005). Most NWP models use some form of scaled pressure as the vertical coordinate, typically a hybrid-sigma coordinate system which is terrain-following near the surface, transitioning to pressure levels aloft. This means that the level heights change from profile to profile. If the height variable is not present, it can be calculated using the profile of pressure, temperature and specific humidity, and the surface pressure. If not present, the surface height above mean sea level can be calculated from the surface geopotential.

Different NWP models may not always contain the same variables, and often employ different physical parametrizations. Therefore, some of these variables may not exist or exist as a combination of two more variables.

Variables for evaluation include cloud fraction and the various condensates, typically existing in mixing ratios: liquid water, ice water, rain, and potentially snow, graupel and hail. Note that precipitating quantities (rain, snow, graupel, hail) may be present in flux form rather than mixing ratios, especially in global models with lower horizontal resolution. Cloud fraction is a necessary quantity for evaluating water contents. Whether snow is included when calculating cloud fraction should be communicated to Cloudnet, as should sub-grid-scale quantities if they feature as separate entities within the physical parametrization scheme. Schemes which separate large-scale and convectively-produced cloud quantities can potentially be evaluated with respect to both parametrizations, including a precipitation fraction quantity if applicable.

Microphysical quantities (size, number) if applicable/available should be included together with the bulk water contents. Future enhancements will enable the investigation of humidity and relative humidity profiles, together with the ability of new physical parametrisations to generate mixed-phase and other multiple hydrometeor populations. Winds, turbulent properties and the boundary-layer should also be included as there are major developments in observing these parameters.

Surface variables may also be evaluated.

Translator scripts

Since many NWP centres already have their own format for distributing model profile data, Cloudnet has created a number of translator scripts for converting particular NWP centre formats into the internal Cloudnet format. Since post-processing of the model data is often necessary to create the concatenated set of forecast leadtimes suitable for skill score evaluation, it is appropriate and simpler to provide data in these NWP centre formats.

Translators are available for certain formats from the following NWP centres:

ECMWF (ASCII format, MARS output, ARM netcdf)

Met Office (ASCII format, MOLTS netcdf)

MeteoFrance (ASCII format)

DWD (University of Köln netcdf, ICON meteogram netcdf)

RACMO (netcdf)

HARMONIE (netcdf)

ACCESS (netcdf)

SCAM (netcdf)

NCEP (binary, ARM netcdf)

A translator for the standard WRF (Weather Research and Forecasting, Skamarock et al., 2008) model format is also in development and will be available soon.

Cloudnet should be contacted in the first instance, as we may have, or know of, a translator that requires minimal modification to be used for a new NWP model.

ACTRIS (www.actris.eu) is supported by the European Commission under the Horizon 2020 – Research and Innovation Framework Programme, H2020-INFRAIA-2014-2015, Grant Agreement number: 654109

Data workflow

The general workflow is summarised in Fig. 3. Model data from each NWP centre is transferred to the Cloudnet incoming FTP server. Model data is then archived internally within the Cloudnet database and post-processed to the Cloudnet format. The model data from each centre is ingested within the Cloudnet scheme and processed to create the evaluation metrics at each available site. The results are made available on the Cloudnet website. Due to the distributed nature of Cloudnet processing, it is possible to process either within the central Cloudnet server, or on-site at the observation location. The model data in Cloudnet format can also be used by the Doppler lidar processor, and by the EARLINET SCC processor, and an outgoing FTP server has been setup for this purpose.

Transfer method:
Protocol: sftp
Server: cloudnet.fmi.fi
Port: 2222
Authentication: ssh key
Username: [*assigned on request*]

Procedure:
Create local ssh key (e.g. ssh-keygen -t rsa -f ~/.ssh/id_rsa.cloudnet)
Email ewan.oconnor@fmi.fi with the public key (e.g. ~/.ssh/id_rsa.cloudnet.pub)
You will be allocated a username (e.g. *nwpuser*) and a directory (e.g. *nwpcentre*)

Connection:
ssh –i ~/.ssh/id_rsa.cloudnet –P 2222 *nwpuser*@cloudnet.fmi.fi
cd incoming/*nwpcentre*

All files from one model from one NWP centre should be placed in the same directory, with suitable filenames as discussed in section "Model data requirements". For NWP centres that provide more than one model (version, domain, resolution, parametrization), directories for each individual model should be created with descriptive names.

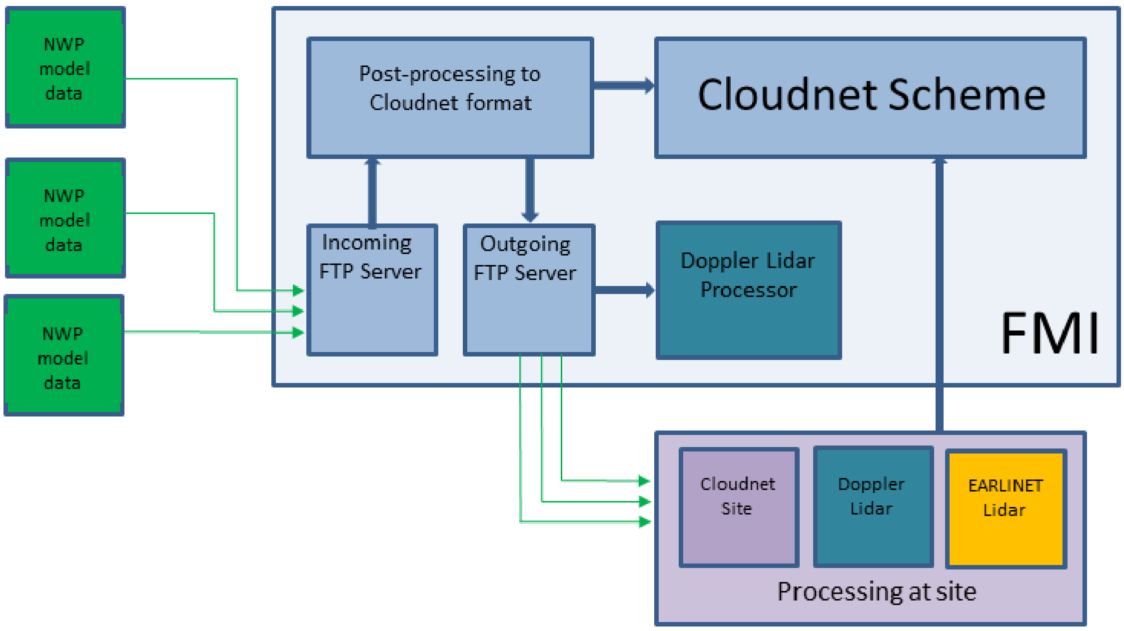

Figure 3: Model data workflow within Cloudnet

ACTRIS (www.actris.eu) is supported by the European Commission under the Horizon 2020 – Research and Innovation Framework Programme, H2020-INFRAIA-2014-2015, Grant Agreement number: 654109

References

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

Hogan, R. J., E. J. O'Connor and A. J. Illingworth, 2009: Verification of cloud-fraction forecasts*. Q. J. R. Meteorol*. Soc., **135**, 1494-1511.

Illingworth, A. J., R. J. Hogan, E. J. O'Connor et al., 2007: Cloudnet – continuous evaluation of cloud profiles in seven operational models using ground-based observations. *Bull. Am. Meteorol. Soc*., **88**, 883-898.

Morcrette, C. J., O'Connor, E. J. and Petch, J. C., 2012: Evaluation of two cloud parametrization schemes using ARM and Cloud-Net observations. *Q.J.R. Meteorol. Soc*., **138**, 964–979.

Skamarock, W. C., J. B. Klemp, J. Dudhia, D. O. Gill, D. M. Barker, M. G Duda, X.-Y. Huang, W. Wang, and J. G. Powers, 2008: A Description of the Advanced Research WRF Version 3. NCAR Tech. Note NCAR/TN-475+STR, 113 pp. doi:10.5065/D68S4MVH

Appendix A

This appendix describes the contents of a typical model dataset in Cloudnet format. The file format is netcdf (currently netcdf version 3).

For all files float: single-precision floating-point vector
int: signed two-byte integer vector

Dimensions and coordinate variables

Dimensions	Units	Type	Description
time	hours since YYYY-MM-DD 00:00:00 +00:00	float	hours UTC
level		int	Model level
flux_level		int	Model flux level
frequency	GHz	float	Microwave frequency

All model files in Cloudnet format contain one set of profiles sequential in time for one specific location covering a period of one day. If there is more than one forecast per day, then there will be separate files containing each concatenated forecast set. The file contents comprise dimensions, variables, variable attributes and global attributes.

There are 4 dimensions; an additional dimension, soil_level, may also be present in some files. Note that each dimension is also a coordinate variable (although this is not a netcdf requirement).

There are two vertical dimensions, level and flux_level (three vertical dimensions if soil_level is present). This is because certain variables are held on full model levels (level), while others are held on half model levels (flux_level) for computational reasons. Most fluxes are on half levels, hence the choice of dimension name.

Note the special form for the units attribute for the variable time, which contains the current date and start time. Time is always relative to 0000 UTC in Cloudnet files, so the time portion of the units attribute string is always 0. The units attribute string is also important for plotting purposes.

The variable forecast_time contains the time elapsed since the initialization time of the forecast from which it was taken (profiles may be taken from more than one forecast).

ACTRIS (www.actris.eu) is supported by the European Commission under the Horizon 2020 – Research and Innovation Framework Programme, H2020-INFRAIA-2014-2015, Grant Agreement number: 654109

Scalars, vectors, and parameters on model levels

Parameter Name	Dimensions	Units	Type	Description
latitude		degrees_N	float	Latitude of model gridpoint
longitude		degrees_E	float	Longitude of model gridpoint
horizontal_resolution		km	float	Horizontal resolution of model
level	level		int	Model level
flux_level	flux_level		int	Model flux level
frequency	frequency	GHz	float	Microwave frequency
time	time	* hours	float	Hours UTC
forecast_time	time	hours	float	Time since initialization of forecast
pressure	time level	Pa	float	Pressure
temperature	time level	K	float	Temperature
uwind	time level	m s-1	float	Zonal wind
vwind	time level	m s-1	float	Meridional wind
omega	time level	Pa s-1	float	Vertical wind in pressure coordinates
wwind	time level	m s-1	float	Vertical wind
q	time level		float	Specific humidity
rh	time level		float	Relative humidity. With respect to liquid above 0 degrees C and with respect to ice below 0 degrees C.
ql	time level		float	Gridbox-mean liquid water mixing ratio
qi	time level		float	Gridbox-mean ice water mixing ratio
qr	time level		float	Gridbox-mean rain water mixing ratio
qs	time level		float	Gridbox-mean snow water mixing ratio
qq	time level		float	Gridbox-mean graupel water mixing ratio
cloud fraction	time level		float	Cloud fraction
height	time level		float	Height above ground. The heights have been calculated using pressure, temperature and specific humidity.

* see description in "Dimensions and coordinate variables"

ACTRIS (www.actris.eu) is supported by the European Commission under the Horizon 2020 – Research and Innovation Framework Programme, H2020-INFRAIA-2014-2015, Grant Agreement number: 654109

Parameters on flux levels

Parameters on flux levels (also known as half levels) are denoted with the prefix ‘*flx_*’
Note that winds are usually calculated on flux levels but are then interpolated to the model levels.

Parameter Name	Dimensions	Units	Type	Description
flx_ls_rain	time flux_level	kg m-2 s-1	float	Large-scale rainfall flux
flx_ls_snow	time flux_level	kg m-2 s-1	float	Large-scale snowfall flux
flx_conv_rain	time flux_level	kg m-2 s-1	float	Convective rainfall flux
flx_conv_snow	time flux_level	kg m-2 s-1	float	Convective snowfall flux
flx_net_sw	time flux_level	W m-2	float	Net shortwave flux
flx_net_lw	time flux_level	W m-2	float	Net longwave flux
flx_down_sens_heat	time flux_level	W m-2	float	Sensible heat flux
flx_turb_moist	time flux_level	W m-2	float	Turbulent moisture flux
flx_turb_mom_u	time flux_level	kg m-1 s-2	float	Zonal turbulent momentum flux
flx_turb_mom_v	time flux_level	kg m-1 s-2	float	Meridional turbulent momentum
flx_height	time flux_level	m	float	Height above ground

ACTRIS (www.actris.eu) is supported by the European Commission under the Horizon 2020 – Research and Innovation Framework Programme, H2020-INFRAIA-2014-2015, Grant Agreement number: 654109

Surface parameters

Surface parameters on flux levels are denoted with the prefix '*sfc_*' .

Parameter Name	Dimensions	Units	Type	Description
sfc_pressure	time	kg m-2 s-1	float	Surface pressure
sfc_geopotential	time	m2 s-2	float	Geopotential
sfc_height_amsl	time	m	float	Surface height above mean sea level
sfc_temp_2m	time	K	float	Temperature at 2m
sfc_q_2m	time		float	Specific humidity at 2m
sfc_wind_u_10m	time	m s-1	float	Zonal wind at 10m
sfc_wind_v_10m	time	m s-1	float	Meridional wind at 10m
sfc_net_sw	time	W m-2	float	Surface net downward shortwave flux
sfc_net_lw	time	W m-2	float	Surface net downward shortwave flux
sfc_down_sw	time	W m-2	float	Surface downwelling shortwave flux
sfc_down_lw	time	W m-2	float	Surface downwelling longwave flux
sfc_cs_down_sw	time	W m-2	float	Clear sky downwelling shortwave flux
sfc_cs_down_lw	time	W m-2	float	Clear sky downwelling longwave flux
sfc_down_lat_heat_flx	time	W m-2	float	Latent heat flux
sfc_down_sens_heat_flx	time	W m-2	float	Sensible heat flux
sfc_albedo	time		float	Surface albedo
sfc_ls_rain	time	kg m-2	float	Large-scale rainfall amount
sfc_ls_snow	time	kg m-2	float	Large-scale snowfall amount
sfc_conv_rain	time	kg m-2	float	Convective rainfall amount
sfc_conv_snow	time	kg m-2	float	Convective snowfall amount
sfc_ls_precip_fraction	time		float	Large-scale precipitation fraction
sfc_bl_height	time	m	float	Boundary layer height
sfc_rough_mom	time	m	float	Surface roughness for momentum
sfc_rough_heat	time	m	float	Surface roughness for heat

ACTRIS (www.actris.eu) is supported by the European Commission under the Horizon 2020 – Research and Innovation Framework Programme, H2020-INFRAIA-2014-2015, Grant Agreement number: 654109

Radar attenuation parameters

In addition, the internal Cloudnet model data format may contain these parameters. Note that these are produced **within** the Cloudnet suite, using the *propagation* utility, and are not requested in file submission. These parameters are used to estimate the radar attenuation due to gases and liquid water, and are used in conjunction with the observed liquid water path, if available. Note that all parameters may be calculated for more than one radar frequency, hence the frequency dimension. The frequencies selected are typically 35 and 94 GHz, but this is flexible, and so the frequency parameter should be checked before selecting the appropriate index.

Parameter Name	Dimensions	Units	Type	Description
gas_atten	frequency time level	dB	float	Two-way attenuation from the ground due to atmospheric gases
specific_gas_atten	frequency time level	dB km-1	float	Specific one-way attenuation due to atmospheric gases
specific_saturated_gas_atten	frequency time level	dB km-1	float	Specific one-way attenuation due to atmospheric gases for saturated air (saturated with respect to ice below 0 degrees C)
specific_dry_gas_atten	frequency time level	dB km-1	float	Specific one-way attenuation due to atmospheric gases for dry air (no water vapour)
K2	frequency time level		float	Dielectric parameter (|K|^2) of liquid water
specific_liquid_atten	frequency time level	(dB km-1) / (g m-3)	float	Specific one-way attenuation due to liquid water, per unit liquid water content

ACTRIS (www.actris.eu) is supported by the European Commission under the Horizon 2020 – Research and Innovation Framework Programme, H2020-INFRAIA-2014-2015, Grant Agreement number: 654109

Attributes

All parameters have the following attributes:

long_name
units
_FillValue
missing_value

They may also have the following attributes:

original_name
original_format
short_name
standard_name
comment

The file also contains at least the following global attributes:

Conventions = "CF-1.0" ;
title
location
source
institution
history

E.g. :Conventions = "CF-1.0" ;
:title = "ECMWF single-site output over Chilbolton" ;
:location = "Chilbolton" ;
:source = "ECMWF Integrated Forecast System (IFS)" ;
:institution = :"European Centre for Medium-Range Weather Forecasting" ;
:initialization_time = "2016-03-05 00:00:00 +00:00" ;
:history = "Mon Mar 7 07:38:39 GMT 2016 - NetCDF generated from original data by update_files.pl using cnmodel2nc on anvil2" ;

ACTRIS (www.actris.eu) is supported by the European Commission under the Horizon 2020 – Research and Innovation Framework Programme, H2020-INFRAIA-2014-2015, Grant Agreement number: 654109