AATSR CLOUD CLEARING ROUND ROBIN PROTOCOL

Table of Content

List of Figures

List of Tables

1. Introduction

The AATSR Cloud Clearing Round Robin (CCRR) is designed to facilitate cloud detection algorithm development for implementation in remote sensing applications for both land and atmospheric variables. Undetected cloud remains the largest source of uncertainty in the retrieval of many geophysical parameters, with data providers using a range of methods to screen or identify cloud prior to completing their retrieval. The CCRR is an independent scientific exercise, designed to identify the best cloud detection algorithm over land for the Advanced Along Track Scanning Radiometer (AATSR) by encouraging competition between algorithm developers and comparing cloud detection skill across the globe. This document describes the protocol to be used in running the CCRR.

1.1. Reference documents

Table 1: List of reference documents

Reference Number	Reference
RD1	Hocking, J., Rayer, P., Rundle, D., Saunders, R., Matricardi, M., Geer, A., Brunel, P. and Vidot, J (2014). RTTOV v11 Users Guide, EUMETSAT Satellite Application Facility on Numerical Weather Prediction (NWP SAF), NWPSAF-MO-UD-028.
RD2	Arino, O., Leroy, M., Bicheron, P., Brockman, C., Defourney, P., Vancutsem, C., Achard, F., Durieux, L., Bourg, L., Latham, J., Di Gregorio, A., Witt, R., Herold, M., Sambale, J., Plummer, S. and Weber, J-L. 2007. GlobCover: ESA service for Global Land Cover from MERIS. 1-4244-1212-9/07 IEEE, 2412-2415.
RD3	Borbas, E., and Ruston, B. 2010. The RTTOV UWiremis IR land surface emissivity module. EUMETSAT Satellite Application Facility on Numerical Weather Prediction. NWPSAF-MO-VS-042. Version 1.
RD4	Vidot, J., and Borbas, E. 2012. RTTOV land surface VIS/NIR BRDF module.
RD5	Gesch, D., and Greenlee, S. 1996. GTOPO30 Documentation. US Department of the Interior U.S. Geological Survey. http://eros.usgs.gov//Find_Data/Products_and_Data_Available/GTOPO30.
RD6	Group for High Resolution Sea Surface Temperature (GHRSST), 2011. NAVOCEANO Land/Sea/Lake Mask. www.ghrsst.org/products-and-services/tools/navo-ghrsst-pp-land-sea-mask/.

1.2. Glossary

AATSR---------- Advance Along Track Scanning Radiometer
ATSR---------- Along Track Scanning Radiometer
CCRR---------- Cloud Clearing Round Robin
CDR---------- Climate Data Record
DUE---------- Data user Element
ESA ---------- European Space Agency
FTP---------- File Transfer Protocol
LST---------- Land Surface Temperature
NWP---------- Numerical Weather Prediction.

RRDP---------- Round Robin Data Package

2. Participation

2.1. Who is invited to participate?

Anyone is welcome to participate in the Cloud Clearing Round Robin (CCRR) and this has been advertised to the scientific community through user consultation meetings, a user requirements questionnaire, conferences and other scientific meetings. Personal invites will be extended to those in the scientific community known to be developing cloud detection algorithms to encourage as much participation as possible.

2.2. What are the benefits of participation?

The benefits of participating in the cloud clearing round robin include:

- The opportunity to develop your cloud detection algorithm using training data with a manually classified cloud mask provided as reference data, and performance metric tools provided within the Round Robin Data Package (RRDP).
- The opportunity to compare the output of your cloud detection algorithm with other leading algorithms, identifying the relative strengths and weaknesses. Comparisons will be made available within the CCRR Report (DEL-33).
- Contribution to reducing one of the largest sources of uncertainty in remote sensing of geophysical parameters.
- The opportunity to be a co-author on a peer-reviewed paper.
- Potentially providing the algorithm most suited to Land Surface Temperature (LST) retrieval, which could be implemented within the Along Track Scanning Radiometer (ATSR) LST Climate Data Record (CDR).

2.3. What are my commitments as a participant?

As a participant in the CCRR you are committing to the following:

- Providing a cloud mask generated using your algorithm for each of the selection scenes (details on the round robin process and data submission are given in the following sections). Data must be provided in the specified format and by the submission deadline. If a contributing algorithm is applicable only under day or night-time conditions, submissions for just one category of scenes will be accepted.
- Providing a description of your cloud-clearing algorithm including reference to peer-reviewed documents where available using the template provided. This description should include the channel combinations used, the views required, applicability under different solar illumination, swath constraints where applicable and sensitivity to stratospheric aerosol (if known). The size and nature of both static and dynamic auxiliary files used should be stated along with the main steps of the algorithm indicating where significant processing is required or external models are used.
- Providing a statement about the potential for further development of the algorithm used in the round robin submission.

- Providing my own resources to cover participation.
- Registering my intent to participate with the project team in order to gain access to the data (see section on contacts and submission support).
- Agreeing to the GlobTemperature CCRR Conditions of use:
  - To use the RRDP only for round robin participation.
  - To not redistribute the RRDP data to other parties without the permission of ESA or the original data provider as appropriate.
  - To acknowledge the assistance of the ESA GlobTemperature Project in any publication that is based on round robin data.

Agreement to these conditions is implicit upon registration.

- To download the RRDP and associated documentation from the secure FTP site.
- Give permission for the cloud masks and calculated performance metrics to be made publically available.

3. Round Robin Design and Selection

3.1. Design

The Cloud Clearing Round Robin design is illustrated in Figure 1. This is an objective procedure designed to facilitate a fair comparison between all submissions. Any participants from within the GlobTemperature consortium will be subject to the same process and will not have access to data prior to its distribution to all participants. There will be three data releases. The first data release will consist of a set of training data. The details of the data content are provided in the data section, but there will be enough information provided to run threshold based, Bayesian and neural network cloud detection algorithms. A manually screened cloud mask will also be provided with these data to enable evaluation of algorithm performance. To maintain objectivity, once the manually screened cloud mask has been developed by the consortium it will not be available to them for development purposes aside from within the CCRR data releases. The performance metric tools that will be used to compare all final submissions will also be made available for development purposes. The scenes provided will cover key land surface types across the globe.

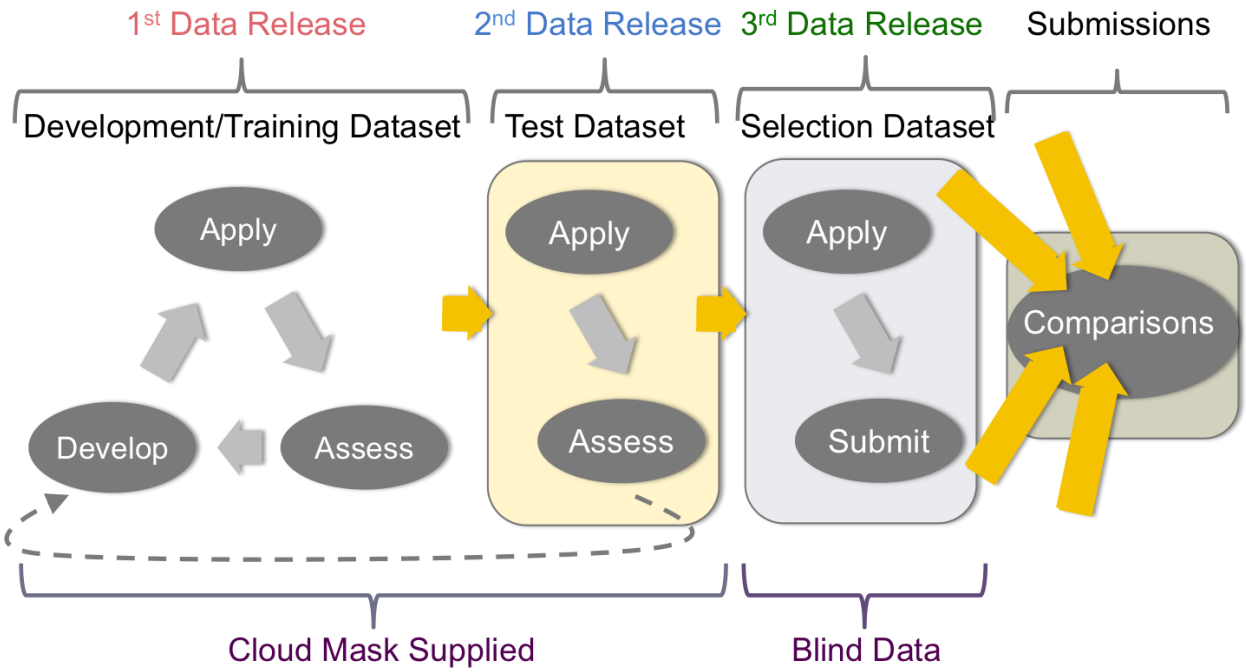

Figure 1: Cloud Clearing Round Robin Design

The second data release will consist of some test data. The purpose of these data is to enable participants to evaluate their algorithm performance on data that has not been used for training and development purposes. If issues are identified, the training data can then be used to refine the algorithm. The test data are only available for a short time prior to the submission date as these are not intended for training and development purposes. The test data will be provided with a manually screened cloud mask.

The selection data are 'blind' (ie. provided without a reference cloud mask) and these are the data intended for use in the submission process. All participants will be required to run their algorithm on all

scenes in the selection dataset and submit their cloud masks by the deadline specified. Due to the schedule of this exercise, submissions after this deadline will not be considered.

3.2. Selection Criteria

The selection criteria for the best cloud detection algorithm fall into the following categories.

- Performance metrics including the percentage of perfect classification, hit rate, false alarm rate and true skill score. The percentage of perfect classification is the percentage of all pixels classified correctly as either cloud or clear. The hit rate is the percentage of cloudy pixels correctly identified, and the false alarm rate the percentage of clear-sky pixels falsely flagged as cloud. The true skill score is the hit rate minus the false alarm rate.

  These assess how many pixels the submitted algorithm correctly identify as clear or cloud, as well as the number of pixels missed and those screened unnecessarily.

- Land Surface Temperature impacts.

  These metrics will give an indication of the effect of over/under-screening by the submitted algorithm on the retrieved land surface temperature.

- Performance under clear-sky atmospheric conditions (low aerosol loading).

  Cloud detection performance will be assessed separately for clear-sky conditions.

- Performance in the presence of heavy aerosol loading.

  Classification will be assessed separately for high atmospheric aerosol loadings.

- Performance over cold and or reflective surfaces eg. deserts and snow/ice.

  Cloud detection over bright surfaces with high reflectance is particularly difficult. Algorithm performance will be evaluated over deserts and snow/ice where the cold temperatures also increase the challenge of accurate cloud detection.

- Performance under different solar illuminations (day, night and twilight).

  Cloud detection algorithms can use observations in different channels during the day and at night. Performance will be assessed separately under these conditions, as the most appropriate algorithm for a particular application may be different for day and night time conditions. Submission to a single category (day or night) would be acceptable.

- Performance in cloud/clear transition zones eg. cloud edges, thin cirrus and pixel/sub-pixel cloud.

  Cloud detection along transition zones is also particularly difficult. These metrics will analyse performance along coastlines, at cloud edges, in regions of thin cirrus and in the presence of pixel or sub-pixel cloud.

- Improvability or the potential for further development of the algorithm.

  As new observations are made and more data become available, there is the potential to use these in cloud detection algorithms. This qualitative metric assesses the potential for algorithm

development with the provision of additional data and is based on the technical note submitted by the participant.

- Difficulty of implementation of the algorithm.

  The difficulty of algorithm implementation is assessed from the details of the required static and dynamic auxiliary files required and the computing/processing power needed, plus details of any external processing such as running models. The criteria for this will be assessed based on the participant's technical note describing the algorithm.

For all criteria where metrics are calculated from the submitted cloud mask, code will be supplied during the training and development phase to enable participants to see how well their algorithm is performing.

4. Schedule

4.1. Round Robin Process

The round robin process will commence on the 1st June 2015 and run for six months. On June 1st 2015 the training dataset will be disseminated via FTP with the Round Robin Data Package describing the data content and usage. Participants will then have a five-month period in which to develop their cloud detection algorithms, running as many tests as they require on the training dataset. Throughout the entire six-month process, code to calculate the performance metrics that will be used in the selection process will be made available for participants to use.

On the 2nd November 2015 the second data release containing the test dataset with a manually classified cloud mask will be made available. Participants will be able to test their algorithm making updates where necessary using the training data during this one month period. The selection dataset will be available from the 2nd November on a request basis. Once a participant has completed their testing with the test dataset they may request access to the selection dataset by email. Once access has been provided, the participant will have a 24-hour period in which to upload their submissions. This 24-hour period will only be extended in the event of IT issues. Problems of this nature should be communicated to the round robin team at the earliest opportunity.

Cloud mask submissions are required by the 30th November. Any submissions after this date will not be included in the selection process.

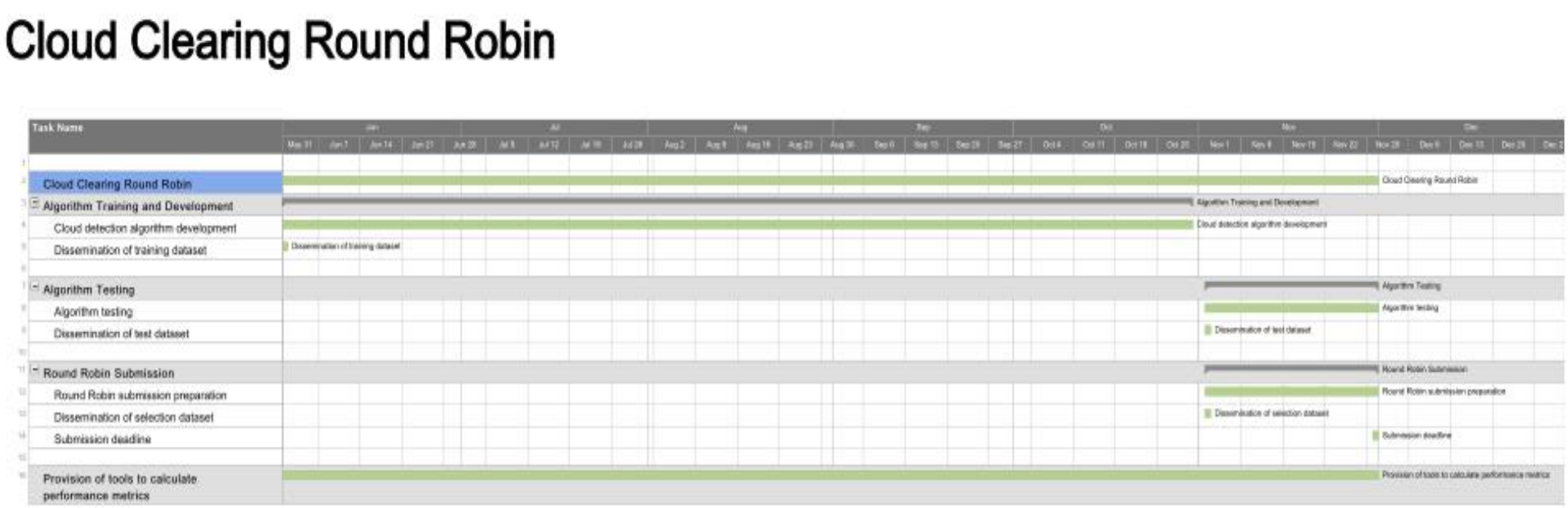

Figure 2: Cloud Clearing Round Robin Schedule

4.2. Deadline for Submissions

The deadline for cloud mask submissions for the selection dataset is the 30th November 2015. Due to the schedule of this exercise submissions after this date will not be included in the selection process. The description of your algorithm and statement about potential future development (as detailed under participant commitments) must also be submitted by this deadline.

4.3. What happens after submission?

Following the submission deadline, the science team will assess all contributions. Submissions will be compared using the performance metric tools and selection criteria detailed above. A full report of the comparisons will be generated and made publically available on the GlobTemperature website within one year of the submission deadline. Individual communications with participants or communication of comparisons via upload of plots to the GlobTemperature website may occur at regular intervals during the comparison period. The GlobTemperature website can be accessed at www.globtemperature.info.

Towards the end of the analysis period following the round robin exercise, a paper will be drafted to describe the round robin process and outcomes. The paper will be drafted in consultation with participants and co-authored appropriately.

Once the full report has been compiled and the paper submitted for peer-review the round robin data will be made publically available via the GlobTemperature website. The complete dataset (selection scenes and submitted cloudmasks) will be freely available but should be appropriately referenced if used in further publications.

The most appropriate algorithm for land surface temperature retrieval will have the potential for inclusion in the GlobTemperature LST Climate Data Record with the consent of the participant. If time constraints on the part of both the participant and the GlobTemperature project team allow the algorithm to implemented in the processing chain, further checks will be carried out in a large-scale application to ensure that no spurious features occur when applied globally.

5. Data

5.1. Round Robin Data Distribution

The round robin datasets (training, test and selection) will be distributed in NetCDF format. For each scene the following data will be provided.

- AATSR nadir view brightness temperatures.
- AATSR forward view brightness temperatures.
- AATSR nadir view reflectance.
- AATSR forward view reflectance.
- Numerical Weather Prediction (NWP) surface temperature (provided from ECMWF ERA-Interim).
- NWP total column water vapour (provided from ECMWF ERA-Interim).
- Reference cloud mask generated by expert manual inspection.
- Land/sea mask (1/120th degree global distance to land (0-24 km) for each cell with lakes/inland water treated as land [adapted from RD6]).
- Simulated brightness temperatures in both nadir and forward views using RTTOV 11.2 [RD1].
- Simulated reflectance in both nadir and forward views using RTTOV 11.2.
- Data quality information from AATSR nadir and forward view confidence flags.
- Land surface cover (near 1 km variant of the GlobCover land classification [RD2]).
- Emissivity (used to simulate brightness temperatures in RTTOV 11.2, taken from the UWIREMIS atlas [RD3]).
- Surface reflectance (used to simulate top of atmosphere reflectance in RTTOV 11.3, taken from the BRDF atlas [RD4]).
- Scene date and location.
- Retrieved land surface temperature (all pixels, for LST impact metrics).
- Elevation (provided from the GTOPO30 digital elevation map [RD5]).

Three datasets will be disseminated during the Cloud Clearing Round Robin:

1). The training dataset.

This is available throughout the entirety of the Cloud Clearing Round Robin. For the first five months it is the only dataset available and is intended for algorithm training and/or development.

2). The test dataset.

This is released at the beginning of month six of the CCRR. This scene selection is intended to allow participants to check that they haven't over-tuned their algorithm to the training dataset. If

performance updates are required at this stage it is recommended that developments are made using the training data and tested using the test dataset to prevent alternative tuning to this dataset.

3). The selection dataset.

This is available from the time of the test dataset release. It is the dataset on which final algorithm performance will be assessed. No manual mask is supplied with this dataset to maintain objectivity in the selection procedure. To access this dataset, the participant must contact Claire Bulgin via email (see contacts section). Data will be then be made available for download and submissions must be uploaded within 24 hours.

All three datasets contain the same variables with the exception of the selection dataset, which is 'blind' with no manual cloud mask supplied. The same number of scenes will be available in each set covering the same range of conditions – different surface types (vegetated, desert, snow and ice), different atmospheric states (high/low water vapour), different aerosol loadings and differing solar illumination.

5.2. Round Robin Data Submission

All Cloud Clearing Round Robin data submissions should consist of two components; the calculated cloud mask for the selection scenes and a brief technical note describing the algorithm. Without both aspects, the submission cannot be included in the selection process.

5.2.1. Cloud Mask Submission

Data submissions to the Cloud Clearing Round Robin should be in NetCDF format, using NetCDF3 (Classic) as a minimum requirement. The following fields must be included in the submission:

Variable Name	Description
scene.id	Unique ID provided with each selection scene.
scene.cloudmask_binary	Binary cloud mask, clear = 0, cloud = 1.

Table 2: Mandatory fields in NetCDF submission files.

For participants who generate cloud probability fields, an appropriate threshold for LST retrieval will need to be applied to generate a binary mask. Optionally, the following fields can also be included:

Variable Name	Description
scene.cloudmask	Cloudmask with up to four possible classifications – clear = 0, cloud = 1, probably clear = 2, probably cloud = 3.
scene.quality_flag	Quality flag mask with up to three possible classifications – poor quality = 0, acceptable = 1, excellent = 2.

Table 3: Optional fields in NetCDF submission files.

Submissions will need to be made via the FTP site and an example submission file will be provided to all participants. Resource and time constraints within this exercise will compel the selection team to adopt a zero tolerance policy to submissions with an incorrect format specification. Please liaise with the main round robin contact if you have any questions with regard to the submission of data.

5.2.2. Technical Note Submission

The technical note describing the algorithm can be brief but must cover the following points. A template will be provided to all participants for this.

- A description of the algorithm including reference to peer-reviewed documents where available.
- The data views required for the algorithm to work.
- The channel combinations required by the algorithm.
- Applicability of the algorithm under different solar illuminations.
- Swath constraints on the algorithm (if any).
- Sensitivity to aerosol including stratospheric aerosol (if known).
- The size and nature of static auxiliary files used by the algorithm.
- The size and nature of dynamic auxiliary files used by the algorithm.
- The main steps of the algorithm including an indication of those requiring significant processing or the running of external processes eg. models.

6. Contacts and Submission Support

6.1. Project Support and Troubleshooting

The main contact throughout the Cloud Clearing Round Robin is Claire Bulgin at the University of Reading (c.e.bulgin@reading.ac.uk). Queries regarding data access and submission should in the first instance be directed to Claire who will be able to provide support to participants. An online record of frequently asked questions will be maintained throughout the round robin to facilitate troubleshooting.

If necessary queries can also be copied to the work package manager, Chris Merchant (c.j.merchant@reading.ac.uk) or the GlobTemperature Science Leader, Darren Ghent (djg20@leicester.ac.uk).

End of document