LAND COVER CCI

ALGORITHM THEORETICAL BASIS DOCUMENT VERSION 2

DOCUMENT REF:	CCI-LC-ATBD
DELIVERABLE REF:	D3.6-ATBD
VERSION:	2
ISSUE:	2.3
CREATION DATE:	2012-03-01
LAST MODIFIED:	2013-11-28

Document Signature Table

	NAME	FUNCTION	COMPANY	SIGNATURE	DATE
PREPARED	G. Kirches		BC		
PREPARED	O. Krueger		BC		
PREPARED	M. Boettcher		BC		
PREPARED	S. Bontemps		UCL		
PREPARED	C. Lamarche		UCL		
PREPARED	A. Verheggen		UCL		
PREPARED	C. Lembrée		UCL		
PREPARED	J. Radoux		UCL		
VERIFIED	P. Defourny	PI	UCL		23.05.2013
APPROVED					

Document Change Record

VERSION	DATE	DESCRIPTION	APPROVED
1.0	2012-03-28	First version of the ATBDv1, adapted from the ATBDv0	
1.1	2012-06-01	Updated version, according to RIDs and PM7	
1.2	2012-07-31	Corrected version w.r.t numbering, images and style	
2.0	2012-10-12	ATBD v2, adapted from the ATBD v1.2	
2.1	2012-11.30	Updated version, according to RIDs and PM8	
2.2	2013-04-30	Updated version, which is aligned with Epoch1 v1.1, Epoch2 v1.0 and Epoch3 v1.0	
2.3	2013-11-28	Updated version according to the final LC products of phase I	

From version 1.0 to version 2.2

RID	SECTION	COMMENTS
ATBDv1_1.0/1	2	Figure has been updated
ATBDv1_1.0/2	2	Figure has been updated
ATBDv1_1.0/3	3.1.3	The merging of the composites (performed in the land cover map and condition processing chain) requires a collocation of FR and RR products. This will be performed with the scene collocation described here. More details in the document have been included.

ATBDv1_1.0/4	3.2.2	Explanation has been included (Method developed by M. Bouvet, documentation is available).
ATBDv1_1.0/5	3.2.1	Phrase has been deleted.
ATBDv1_1.0/7	3.3.5 and 3.3.6	The latest development of the cloud screening and an update of the error treatment have been included
ATBDv1_1.0/8	3.3.9	Phrase has been deleted
ATBDv1_1.0/10	3.4.2	This information has been added.
ATBDv1_1.0/12	3.5.3	False reference has been corrected
ATBDv1_1.0/13	3.5.4.2	The document has been harmonised; the finally chosen interval is consistently used throughout the text.
ATBDv1_1.0/14	-	This section has been removed.
ATBDv1_1.0/15	4.1.4 and 4.4	A new section about the multi-sensor approach has been added in the land cover mapping processing chain. This section contains all useful information about this technique
ATBDv1_1.0/18	3.7.3.2	This section has been updated. Option 1 “in-situ” has been used and results are included in the PVASR Option 2 “simulated” has not been used, but we would like to keep it here as something which someone else might do A new option 3 “model intercomparison” will be added. We did this in the PVASR by comparing with SCAPE-M.
ATBDv1_1.0/19	-	Section has been deleted
ATBDv1_1.0/20	4.2.2	Document has been corrected
Action PM7-8	All sections	Processing logic scheme in the ATBD has been aligned with the DPM
Action PM7-9	3.5.3 and 3.6.3	Description oft he time compositing fort he weekly product has been included
Progress	4 and 5	Processing chains developed to generate land cover map and condition products are now presented in separate sections Each section starts with a general presentation (in the form of logical models) and then presents the detailed algorithms
Progress	4.1.1 and 4.2.3	New sections have been added to include the reference land cover database preparation as a preliminary step of the land cover mapping processing chain
Progress	4.1.1 and 4.2.4	New sections have been added to include the stratification layer preparation as a preliminary step of the land cover mapping processing chain

Progress	4.1.3 and 4.3	New sections have been added to include the multi-year approach in the land cover mapping processing chain. In addition, the impact of the multi-year approach on the classification algorithms are included in the classification algorithm description
Progress	4.1.4 and 4.4	New sections have been added to include the multi-sensor approach in the land cover mapping processing chain In addition, the impact of the multi-sensor approach on the classification algorithms are included in the classification algorithm description
Progress	4.1.5	A new section has been added to illustrate and clarify how the algorithms developed in the framework of the multi-year and multi-sensor approache are linked with the classification algorithms
Progress	5	The sequence of the condition products was changed to reflect the two versions of the products. The order is now: NDVI, Burnt Area, Snow and Water conditions. All references to the fire condition were replaced by the burnt area.
Progress	5.1	In the NDVI condition, the input MERIS FR data are replaced by the MERIS RR dataset (see PM7 conclusions)
Progress	5.1.1	In the NDVI condition, the preliminary step – Resampling was removed from the processing chain as the sources of input data (MERIS RR and SPOT-VGT) share the same spatial resolution.
Progress	5.1.3	In the NDVI condition, “Step 3 – Trend detection and removal” was given up due to the unrealistic processing time needed
Progress	5.1.4	In the NDVI condition, “Step 6 – smoothing” was moved before the temporal aggregation step in order to obtain a smoothed standard deviation required by the climate modelling community. Smoothing the standard deviation after the temporal aggregation appears to be statistically incorrect.
Progress	5.3	All references to the fire condition were replaced by the burnt area.
Progress	5.3	Section related to the ATSR WFA was removed and replaced by the GlobCarbon.v2 product. The GFED.v3 was added as input to the Burnt area land cover condition. The processing chain was modified accordingly.
Action PM8_7	4.1	Alignement of the deliverable to the LC Map V1.0. New paragraph with the description of the changes made in the production of the LC Map V1.0 compared to the described processing chain.
Action PM8_7	5.2 and 5.3	Alignement of the deliverable to the LC NDVI and BA condition processing chains.
ATBDv2/1.	5.3.3	The changes will be made for the BA condition product V1.1 and therefore on the corresponding deliverable.

ATBDv2/2.	5.2	This section has been updated to be in line with NDVI condition V1.0. As it was produced with SPOT-VEGETATION data, this RID is no longer valid.
ATBDv2/3.	5.2.1	The resulted reflectance values are more coherent in space after this spatial filtering. An illustration of the surface reflectances before and after this step has been added in the document under section 5.2.1
Progress ATBDv2.1/1.	3.5.3	updating the definition of status_invalid
Progress ATBDv2.1/2.	3.5.3	updating the aggregation rules w.r.t cloud status
Progress ATBDv2.1/3.	3.5.5	updating of the product description of the LC-CCI pre-processing products w.r.t requirements of Data Standard Working Group
Progress ATBDv2.1/4.	2	Updating to include the LC condition module
Progress ATBDv2.1/5.	4.1.2, 4.1.3, 4.1.4, 4.1.5	Updating w.r.t. the chain run to generate the product Epoch 1 v1.1, Epoch 2 v1.0 and Epoch 3 v1.0 (mainly: no temporal classification, no use of SWIR band from SPOT-VGT)
Progress ATBDv2.1/6.	4.3, 4.5	Section about the temporal classification has been removed
Progress ATBDv2.1/7.	4.4	Section about the use of the SWIR band of SPOT-VGT has been removed Section about the fusion between MERIS FR and RR has been updated
Progress ATBDv2.1/8.	4.5.4.2, 4.5.5, 4.5.7	Section about the generation of the “baseline” land cover map and the way to derive from this baseline 3 LC maps specific to epochs has been added
Progress ATBDv2.1/9.	5.3	This section has been changed to take into account the modified processing chain that was applied to build the burnt areas condition.
Progress ATBDv2.1/10.	5.4	Major changes were applied to this section te depict the actual processing of the snow condition.
Progress ATBDv2.3/1.	5.1, 5.6	The logical model was updated to take into account a post processing step of compliance between the condition products. Section 5.6 was added for the same reason.
Progress ATBDv2.3/2.	5.2	This section was updated to take into account new aspects of the NDVI condition processing chain : the period extension to 2012 instead of 2011 and the removal of step 1 spatial filtering
Progress ATBDv2.3/3.	5.3, 5.3.4	This section was updated to take into account new aspects of the BA condition processing chain: new processing step that increases the spatial consistency of the BA condition product.
Progress ATBDv2.3/4.	5.4, 5.4.5	This section was updated to take into account new aspects of the snow condition processing chain: new processing step that reduces the commission errors in the snow condition product.

Document Diffusion List

ORGANISATION	NAME	QUANTITY
ESA	O. Arino, V. Kalogirou, F.M. Seifert	

TABLE OF CONTENT

LIST OF FIGURES

LIST OF TABLES

1 INTRODUCTION

1.1 Purpose and scope

This document presents the Technical Specification (TS) of the Land Cover project belonging to the European Space Agency (ESA) Climate Change Initiative (CCI). It is intended to provide sufficient information to be able to understand the whole processing chain.

The overall objective of the CCI Land Cover (CCI-LC) project is to deliver, in a consistent way over years and from various Earth Observation (EO) instruments, global land cover products matching the needs of key users belonging to the Climate Modelling Community (CMC).

Pre-processing and classification algorithms are developed in order to design and demonstrate a prototype system able to generate products that meet, as best as possible, the specifications described in the Product Specification Document (PSD). Using the Database for Task 2 (DBT2), different algorithms have been developed, tested and validated, starting from Level 1B imagery through the round-robin exercise. For each step of the processing chain, specific algorithms have been selected (the selection process being fully documented in the Product Validation and Algorithm Selection Report (PVASR).

The Algorithm Theoretical Basis Document (ATBD) version 1 aims at providing a detailed description of the algorithms selected in the PVASR. More precisely, it includes a scientific background and theoretical justification for each implemented algorithm [AD-1].

The workflow of these activities is presented in Figure 1-1.

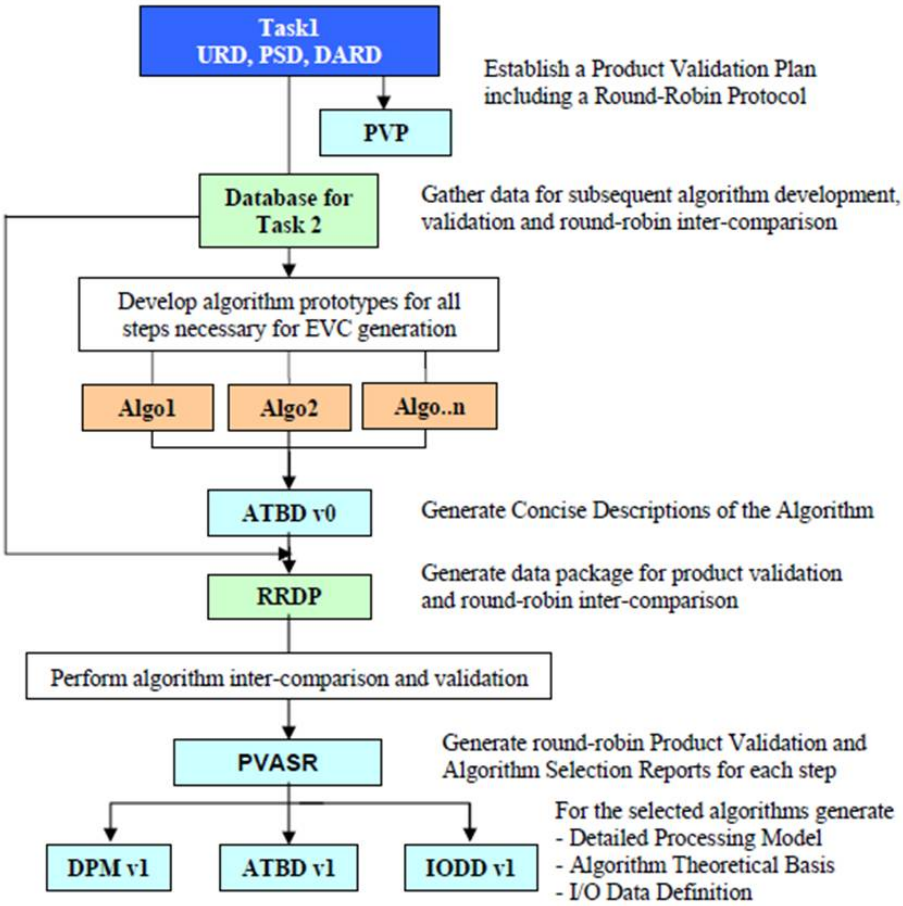


Figure 1-1: Concept of the ATBD version 1 as part of the Task 2 (Algorithm Development, Inter-comparison and Selection) of the CCI-LC project

1.2 References

1.2.1 Applicable documents

This document is in compliance with the applicable documents listed in Table 1-1.

Table 1-1: Applicable documents

REF	TITLE	DOCUMENT CODE	VERSION	DATE
[AD 1]	ESA Climate Change Initiative Phase I - Scientific User Consultation and Detailed Specification Statement of Work (SoW)		1.4	09.11.2009
[AD 2]	ESA Climate Change Initiative – Projects guidelines (EOP-DTEX-EOPS-SW-10-0002)		1.0	05.11.2010
[AD 3]	CCI Land Cover project: Technical proposal		1.0	05.03.2011

REF	TITLE	DOCUMENT CODE	VERSION	DATE
[AD 4]	CCI Land Cover project: Product Specification Document		1.4	27.08.2011
[AD 5]	CCI Land Cover project: Technical Note on Geolocation		1.0	06.2011
[AD 6]	CCI Land Cover project: Round-Robin Protocol		1.2	21.04.2011
[AD 7]	CCI Land Cover project: Algorithm Theoretical Basis Document version 0		0.6	12.12.2011
[AD 8]	CCI Land Cover project: Product Validation Algorithm Selection Report		1.0	28.03.2012
[AD 9]	CCI Land Cover project: Algorithm Theoretical Basis Document version 0		1.0	28.09.2012
[AD 10]	CCI Land Cover project: Detailed Processing Model		1.0	28.03.2012
[AD 11]	CCI Land Cover project: Input and Output Data Description		1.0	27.03.2012
[AD 12]	CCI Land Cover project: Detailed Processing Model		1.0	27.03.2012

1.2.2 Reference documents

The reference documents taken into account in preparing this document are listed in Table 1-2.

Table 1-2: Reference documents

REF	TITLE	VERSION	DATE
[RD-1]	GlobCover 2005 project: Technical Specification Document, End-User meeting 3	4.0	27.11.2008
[RD-2]	GlobCover 2005 project, Design Justification File, I3.8	3.8	11.12.2008
[RD-3]	The AMORGOS MERIS CFI (Accurate MERIS Ortho-Rectified Geo-location Operational Software) RR Processing – Validation report (PO-RP-ACR-GS-0014; L. Bourg; ACRI-ST)	1.0	15/07/2011
[RD-4]	Goshtasby, A., 2005, "2-D and 3-D Image Registration for Medical, Remote Sensing, and Industrial Applications", Wiley Press, 2005	/	/
[RD-5]	Zitová, B. and Flusser, J., 2003, "Image registration methods: a survey", Image Vision Computing 21(11) (2003), pp.: 977-1000	/	/
[RD-6]	Lowe, D. G., 1999, "Object recognition from local scale-invariant features", Proceedings of the International Conference on Computer Vision. 2. pp. 1150–1157,1999	/	/
[RD-7]	Mikolajczyk, K. and Schmid, C.A., 2005, "Performance evaluation of local descriptors", IEEE Transactions on Pattern Analysis and Machine	/	/

REF	TITLE	VERSION	DATE
	Intelligence 27 (10): 1615–1630		
[RD-8]	Bauer, J., Sünderhauf, N. and Protzel, P., 2007, "Comparing Several Implementations of Two Recently Published Feature Detectors". In Proc. of the International Conference on Intelligent and Autonomous Systems, IAV, Toulouse, France, 2007	/	/
[RD-9]	Bay, H., Tuytelaars, T. and Van Gool, L., 2006, "Surf: Speeded up robust features". European Conference on Computer Vision, 1:404-417	/	/
[RD-10]	Evans, C., 2009, "Notes on the opensurf library". Technical Report CSTR-09-001, University of Bristol, http://www.cs.bris.ac.uk/Publications/pub_master.jsp?id=2000970	/	/
[RD-11]	Paperin, M. and Brockmann, C., 2007, "Atmospheric Properties from Optical/Infrared, Cloud Structures". In the ENVISAT Symposium Montreux (CH), 23-27 April 2007, Conference Proceedings 3P5.8	/	/
[RD-12]	Wylie, D., 1998, "Cirrus and Weather: A satellite perspective". OSA Meeting, pp.66-68, Baltimore, U.S.A., 6-8, Oct. 1998	/	/
[RD-13]	Rossow, W.B., and Schiffer, R.A., 1999, "Advances in Understanding Clouds from ISCCP". Bull. Amer. Meteor. Soc., 80, 2261-2288	/	/
[RD-14]	Liou, K.N., 1992, "Radiation and Cloud Processes in the Atmosphere". Oxford University Press, Oxford, 487pp.	/	/
[RD-15]	Luo et al, 2008, "Developing clear-sky cloud and cloud shadow mask for producing clear-sky composites", Remote Sensing of Environment	/	/
[RD-16]	GlobAlbedo project - Preusker, R., Hühnerbein, A., Fischer, J., 2006, "MERIS Global Land Surface Albedo Maps", ATBD Cloud Detection. http://www.brockmann-consult.de/albedomap/pdf/MERIS-AlbedoMap-ATBD-CloudDetection1.0.pdf	1.0	2006
[RD-17]	Eumetsat, Final Report for the Study on Visual Scenes Analysis of AVHRR Data.	2.0	January 2006
[RD-18]	Brockmann et al, 2008, "Cloud Screening Algorithms". Proceedings in MERIS - AATSR Workshop, 2008. ESRIN, Frascati, Italy	/	/
[RD-19]	Schiller et al, 2008, "A method for detection and classification of clouds over water", Proceedings in MERIS - AATSR Workshop, 2008. ESRIN, Frascati, Italy	/	/
[RD-20]	Gomez-Chova et al, 2007, "Cloud-Screening Algorithm for ENVISAT MERIS Multispectral Images", IEEE Transactions on Geoscience and remote sensing, 45, 12	/	/
[RD-21]	Merchant et al, 2006, "Sea surface temperature for climate from the ATSRS, ESA publication, 2006.	/	/

REF	TITLE	VERSION	DATE
[RD-22]	Fischer, J. et al, 2000, "MERIS ESL Cloud Cloud Top Pressure ATBD" http://envisat.esa.int/instruments/meris/pdf/atbd_2_03.pdf	2.03	/
[RD-23]	Fischer, J. et al, 2000, "MERIS ESL Cloud Albedo and Cloud Optical Thickness ATBD" http://envisat.esa.int/instruments/meris/pdf/atbd_2_01.pdf	2.01	/
[RD-24]	Muller, J.-P. and Fischer, J., 2007, "The EU-CLOUDMAP project: Cirrus and contrail cloud-top maps from satellites for weather forecasting climate change analysis". Int. J. Remote Sensing, 28, 1915– 1919.	/	/
[RD-25]	Rathke et al, 2002, "Evaluation of four approximate methods for calculating infrared radiances in cloudy atmospheres", Journal of Quantitative Spectroscopy & Radiative Transfer ,75, 297– 321	/	/
[RD-26]	Brenguier, J.-L., Pawlowska, H., Schüller, L., Preusker, R., Fischer, J. and Fouquart, Y., 2000, "Radiative properties of boundary layer clouds: droplet effective radius versus droplet concentration". J. Atmos. Sci., 57, 803-821	/	/
[RD-27]	Pawlowska, H., Brenguier, J.-L., Fouquart, Y., Armbruster, W., Bakan, S., Descloitres, J., Fischer, J., Fouilloux, A., Gayet, J.-F., Ghosh, S., Jonas, P., Parol, F., Pelon, J., Schüller, L., 2000, "Microphysics/ radiation interaction in a stratocumulus cloud: The EUCREX mission 206 case study". Atmos. Res., 55, 85-102	/	/
[RD-28]	Phulpin T., Derrien M., and Brard A., 1983, "A two dimensional histogram procedure to analyse cloud cover from NOAA satellite high resolution imagery". J. Clim. Appl. Meteorol. 22, 1332-1345	/	/
[RD-29]	Saunders R.W., and Kriebel K.T., 1988, "An improved method for detecting clear sky and cloudy radiances from AVHRR data".Int. J. Rem. Sens. 9: 123-150	/	/
[RD-30]	Garant L., and Weinman J. A., 1986, "A structural stochastic model for the analysis and synthesis of cloud images". J. Clim. Appl. Meteorol. 25, 1052-1068	/	/
[RD-31]	Goodman et al, 1988, "Cloud Detection and Analysis A Review of Recent Progress", Atmospheric Research, 21, 203-228	/	/
[RD-32]	Birks et al, 2007, "Improvements to the AATSR IPF relating to Land Surface Temperature Retrieval and Cloud Clearing over Land", Rutherford Appleton Laboratory	/	/
[RD-33]	Santer,R., Ramon D., Vidot J., Dilligeard E., A., 2007, "Surface Reflectance Model for Aerosol Remote Sensing over Land". International Journal of Remote Sensing, International Journal of Remote Sensing, 28, 3 & 4, pp.: 737 – 760	/	/

REF	TITLE	VERSION	DATE
[RD-34]	Ackerman, S., Strabala, K., Menzel, P., Frey, R., Moeller, C., Gumley, L., Baum, B., Wetzel Seemann, S. and Zhang, H., 2006, "Discriminating clear sky from cloud with MODIS algorithm theoretical basis document (MOD35)". Available at: http://modis.gsfc.nasa.gov/data/atbd/atbd_mod06.pdf	/	/
[RD-35]	Aoki, T., Hori, M., Motoyoshi, H., et al., 2007, "ADEOS-II/GLI snow/ice products — Part II: Validation results using GLI and MODIS data". Remote Sensing Environ., 111, 274-290	/	/
[RD-36]	Hori, M., Aoki, T., Stamnes, K., Li, W., 2007, ADEOS-II/GLI snow/ice products — Part III: Retrieved results, Remote Sensing of Environment, 111, 291-336	/	/
[RD-37]	Stamnes, K., Tsay, S. C., Wiscombe, W., and Jayaweera K., 2007, "Numerically stable algorithm for discrete-ordinate method radiative transfer in multiple scattering and emitting layered media", Appl. Optics, 27, 2502-2510	/	/
[RD-38]	Cervino et al, 2000, "Cloud fraction within GOME footprint using a refined cloud clearing algorithm", Adv. SpaceRes, 25, 5, pp. 993-996	/	/
[RD-39]	Kokhanovsky et al, 2008, "The determination of the cloud fraction in the fraction in the sciamachy pixels using MERIS". Proceedings in the MERIS - AATSR Workshop, 2008. ESRIN, Frascati, Italy	/	/
[RD-40]	Lavanant et al, 2005, "A global Cloud detection scheme for high spectral resolution instrument", ITSC14 Conference presentation	/	/
[RD-41]	McNally et al, 2003, "A cloud detection algorithm for high-spectral-resolution infrared sounders", Q. J. R. Meteorol. Soc., 129, 3411–3423	/	/
[RD-42]	Rathke, C., Fischer, J., Neshyba, S. and Shupe, M., "Improving IR cloud phase determination with 20 microns spectral observations", Geophys. Res. Lett., 29 (8)	/	/
[RD-43]	Susskind et al, 1998, "Determination of ATM and surface parameters from simulated Air/AMS/HSB sounding data retrieval and clouds clearing methodology", Adv. Space Re, 21, 3. 36%384	/	/
[RD-44]	Merchant et al, 2005, "Probabilistic physically based cloud screening of satellite infrared imagery for operational sea surface temperature retrieval", Q. J. R. Meteorol. Soc., 131, pp. 2735–2755	/	/
[RD-45]	Colapicchioni et al, 2004, "Information Mining in Remote Sensing Image - The KIM, KES and KIMV projects", ESA ESRIN Publication	/	/
[RD-46]	Baret, F. et al, 2007, "LAI, fAPAR and fCover CYCLOPES global products derived from VEGETATION Part 1: Principles of the algorithm". Remote	/	/

REF	TITLE	VERSION	DATE
	Sensing of Environment, 110, 3-15, 275-286		
[RD-47]	Shi et al, 2007, "Detection of daytime arctic clouds using MISR and MODIS", Remote Sensing of Environment 107, 172–184	/	/
[RD-48]	Lissens, G., Kempeneers, P., Fierens, F. and Van Rensbergen, J., 2000, "Development of cloud, snow, and shadow masking algorithms for VEGETATION Imagery", IEEE International Geoscience and Remote Sensing Symposium, Honolulu, Hawaii, USA	/	/
[RD-49]	Santer, R., 1997, MERIS ATBD Pixel identification http://envisat.esa.int/instruments/meris/pdf/atbd_2_17.pdf	/	/
[RD-50]	Gomez-Chova et al, 2009, "Cloud screening with combined MERIS and AATSR images", Geoscience and Remote Sensing Symposium, IEEE International, IGARSS 2009, 4, pp.: IV-761 - IV-764	/	/
[RD-51]	Atmospheric products over land for MERIS level 2 ATBD 2.15	2.15	09.2006
[RD-52]	Santer, R., Carrère, V., Dubuisson, P., and Roger, J. C., 1999, "Atmospheric corrections over land for MERIS", International Journal of Remote Sensing, 20,1819-1840	/	/
[RD-53]	GlobAlbedo Project: GlobAlbedo_Albedo_Aer_ATBD	1.0	2010
[RD-54]	GlobAlbedo Project: GlobAlbedo_BBDR_ATBD	1.0	2010
[RD-55]	SYN Detailed Processing Model	2.3	09/2010
[RD-56]	AEROSOL CORRECTION ATBD 1.3	1.3	04/2005
[RD-57]	Fischer J., and H. Grassl, 1984, "Radiative transfer in an atmosphere-ocean system: an azimuthally dependent matrix-operator approach", Applied Optics, 23, 1032-1039	/	/
[RD-58]	Fell F., and J. Fischer, 2001, "Numerical simulation of the light field in the atmosphereocean system using the matrix-operator method", Journal of Quantitative Spectroscopy & Radiative Transfer, 69, 351-388	/	/
[RD-59]	Rummelhart D. and J. McClelland, 1986, "Parallel Distributed Processing", MIT Press, Cambridge, Massachusetts	/	/
[RD-60]	GlobAlbedo Project: GlobAlbedo_Albedo_ATBD	1.0	2010
[RD-61]	Hu, B., Lucht, W., and Strahler, A. H., 1999, "The interrrelationship of atmospheric correction of reflectances and surface BRDF retrieval: a sensitivity study". IEEE Transactions on Geoscience and Remote Sensing, 37:724–738	/	/
[RD-62]	Mekler, Y. and Kaufman, Y. J., 1982, "Contrast reduction by atmosphere and retrieval of nonuniform surface reflectance". Applied Optics, 21:310–316	/	/

REF	TITLE	VERSION	DATE
[RD-63]	Kaufman, Y. J., 1989, "The atmospheric effect on remote sensing and its correction". In Asrar, G., editor, Theory and Applications of optical Remote Sensing, pages 336–428. Wiley and Sons, New York.	/	/
[RD-64]	Vermote, E. F., El-Saleous, N., Justice, C. O., Kaufman, Y. J., Privette, J. L., Remer, L., Roger, J. C., and Tanr´e, D., 1997, "Atmospheric correction of visible to middle infrared EOS-MODIS data over land surface: Background, operational algorithm and validation". Journal of Geophysical Research, 102:17131–17141	/	/
[RD-65]	Vermote, E. F., Tanr´e, D., Deuz´e, J. L., Herman, M., and Morcrette, J. J., 1997, "Second Simulation of the Satellite Signal in the Solar Spectrum, 6S: An overview". IEEE Transactions on Geoscience and Remote Sensing, 35:675–686.	/	/
[RD-66]	Rahman H., Dedieu G., 1994, "SMAC : a simplified method for the atmospheric correction of satellite measurements in the solar spectrum", Int. J. Remote Sensing, 15, 1, 123-143.	/	/
[RD-67]	Tanré, D., Holben, B. N., and Kaufman, Y. J., 1992, "Atmospheric correction algorithmfor NOAA-AVHRR products: theory and application". IEEE Transactions onGeoscience and Remote Sensing, 30, 231–250.	/	/
[RD-68]	Data processing of SPOT-VGT. Available at: http://www.spot-vegetation.com/pages/VegetationSystem/dataprocessing.htm	/	25.07.2011
[RD-69]	IOCCG, 2004, Reports of the International Ocean-Colour Coordinating Group, IOCCG Report Number 4	/	/
[RD-70]	Campbell et al.1995, Level-3 SeaWiFS data Products: Spatial and Temporal Binning Algorithms, SeaWiFS Technical Report Series Volume 32. Available at: http://oceancolor.gsfc.nasa.gov/SeaWiFS/TECH_REPORTS/PreLPDF/PreLVol32.pdf	/	/
[RD-71]	Viovy N., Arino, O., Belward, A., 1992, "The Best Index Slope Extraction (BISE): a method for reducing noise in NDVI time series", Int. J. of Rem. Sen., 13, 8, 1585-1590	/	/
[RD-72]	Vancutsem C., Peckel, J-F., Boagert, P., Defourny, P., 2004, "Assessment of the mean compositing strategy for SPOT VEGETATION time serie", Int. J. of Rem. Sen	/	/
[RD-73]	Hagolle et al. 2004, "Quality assessment and improvement of temporally composited products of remotely sensed imagery by combination of VEGETATION 1 and 2", Rem. Sen. Env., 94, 172-186	/	/
[RD-74]	Vermote, E., Justice, C. and Bréon, F.M., 2009, "Toward Towards a generalized approach for correction of the BRDF effect in MODIS directional reflectances", IEEE Transactions in Geoscience and Remote	/	/

REF	TITLE	VERSION	DATE
	Sensing, 47,3, 898-908.		
[RD-75]	Berthelot, B., Adam, S., Kergoat, L., Cabot, F., Dedieu, G. and Maisongrande, P., 1997, "A global dataset of surface reflectances and vegetation indices derived from AVHRR/GVI time series for 1989 1990: The LAnd SUrface Reflectances (LASUR) data". In Proc. Commun. 7th Int. Symp. Physical Measurements and Signatures in Remote Sens. Courchevel, France, Apr. 7 –11, 1997.	/	/
[RD-76]	Duchemin B. and Maisongrande P., 2002, "Normalisation of directional effects in 10-day global syntheses derived from VEGETATION/SPOT : I. Investigation of concepts based on simulation", Remote Sensing of Environment 81: 90-100	/	/
[RD-77]	Duchemin et al, 2002, "Normalisation of directional effects in 10-day global syntheses derived from VEGETATION/SPOT : II. Validation of an operational method on actual data sets", Remote Sensing of Environment 81:101-113	/	/
[RD-78]	Roujean, J.-L., Leroy, M. and Deschamps, P. Y., 1992, "A bidirectional reflectance model of the Earth's surface for the correction of remote sensing data," J. Geophys. Res. 97, 20455–20468	/	/
[RD-79]	Geiger et al, 2003, Land Surface Albedo from Meteosat Second Generation (MSG) Observations, Geoscience and Remote Sensing Symposium, IGARSS '03. Proceedings. 2003 IEEE International, volume 4, pp.: 2617 – 2619	/	/
[RD-80]	Strahler, A.H., Muller, J.P., Lucht, W., Schaaf, C.B., Tsang, T., Gao, F., Li, X., Lewis P. and Barnsley, M.J., 1999, MODIS BRDF/Albedo Product: Algorithm Theoretical Basis Document Version 5.0 - MODIS Product ID: MOD43 http://modis.gsfc.nasa.gov/data/atbd/atbd_mod09.pdf	/	/
[RD-81]	Bréon, F.M., Maignan, F., Leroy M. and Grant, I., 2002, "Analysis of hot spot directional signatures measured from space", Journal of Geophysical Research, 107, D16, 4282, 1-1 -1-15	/	/
[RD-82]	Maignan, F., Bréon, F.M. and Lacaze, R., 2004, "Bidirectional reflectance of Earth targets: evaluation of analytical models using a large set of spaceborne measurements with emphasis on the Hot Spot", Remote Sensing of Environment, Volume 90, Issue 2, 30, 210-220	/	/
[RD-83]	Snyder, W.C. and Wan, Z., 1998, "BRDF Models to Predict Spectral Reflectance and Emissivity in the Thermal Infrared", IEEE Transactions on Geosciences and Remote Sensing, 36, 1, 214-225	/	/
[RD-84]	Jacquemoud, S., Verhoef, W., Baret, F., Bacour, C., Zarco-Tejada, P.J., Asner, G.P., François, C. and Ustin, S.L., 2009, "PROSPECT + SAIL models:	/	/

REF	TITLE	VERSION	DATE
	a review of use for vegetation characterization", Remote Sensing of Environment, 113, S56-S66.		
[RD-85]	Krippendorff, ., 2007, "Computing Krippendorff's Alpha Reliability", Annenberg School for Communication - Departmental Papers (ASC) - University of Pennsylvania. http://repository.upenn.edu/asc papers/43	/	/
[RD-86]	Krippendorff, K., 2004, "Reliability in Content Analysis: Some Common Misconceptions and Recommendations", Human Communication Research 30, 3, 411-433	/	/
[RD-87]	Scott, W. A., 1955, "Reliability of Content Analysis: The Case of Nominal Scale Coding", Public Opinion Quarterly, 19, 3, 321-325	/	/
[RD-88]	Hilgers, R.D., Bauer, P., Scheiber, V., 2006, „Einführung in die Medizinische Statistik", Springer Verlag, Berlin; Auflage: 2., (5. Dezember 2006), 346 pp, ISBN-10: 9783540339434 ISBN-13: 978-3540339434	/	/
[RD-89]	Gong, P., 1999, "Notes to the courses ESPM 172 - Photogrammetry and Remote Sensing and the ESPM 271 - Advanced Remote Sensing for Natural Resources, http://www.cnr.berkeley.edu/~gong/textbook/	/	/
[RD-90]	IASI/AVHRR Cloud Detection and Characterisation, Technical report	1.1	07/2010
[RD-91]	IASI/AVHRR Visual Scences Analysis and Cloud Detection – IAVISA-, Technical report	3.4	11/2008
[RD-92]	GlobCover 2009 project - Product Description Manual	1.0	21/12/2010
[RD-93]	BEAM 4.8 http://www.brockmann-consult.de/beam		
[RD-94]	BEAM 4.8 VISAT Help, http://www.brockmann-consult.de/beam/doc/help/index.html		
[RD-95]	Getasse 3.0 Digital Elevation Model http://earth.esa.int/services/amorgos/download/		
[RD-96]	Köppen, W.P., 1936, "Das geographische System der Klimate", Berlin	/	/
[RD-97]	Bartholomé, E. and Belward, A., 2005, GLC2000: a new approach to global land cover mapping from Earth observation data, International Journal of Remote Sensing, 26, 9, 1959-1977	/	/
[RD-98]	Vancutsem, C. and Defourny, P., 2009, "A decision support tool for the optimization of compositing parameters", International Journal of Remote Sensing, 30, 1, 41-56	/	/
[RD-99]	Baatz, M. and Schäpe, A., 2000, "Multiresolution segmentation: an optimization approach for high quality multi-scale image segmentation". In J. Strbl, and T. Blaschke (Eds.), Angewandte Geographische	/	/

REF	TITLE	VERSION	DATE
	Informationsverarbeitung (pp. 12-23). Heidelberg: Wichmann.		
[RD-100]	Clevers, J.G.P.W., Schaepman, M.E., Mucher, C.A., de Wit, A.J.W., Zurita-Milla, R. and Bartholomeus H.M., 2007, "Using MERIS on Envisat for land cover mapping in The Netherlands". International Journal of Remote Sensing, 28, 3-4, 637-652.	/	/
[RD-101]	Ifarraguerri, A., 2004., Visual method for spectral band selection, Geoscience and Remote Sensing Letters, 1, 101-106.	/	/
[RD-102]	Bruzzone L., 1995., "An Extension of the Jeffreys-Matusita Distance to Multiclass Cases for Feature Selection". IEEE Transactions on geoscience and remote sensing, 33, 6, 1318-1321.	/	/
[RD-103]	GlobCover 2005 Project - Validation Report (Report editors: Bicheron P. and Leroy, M, End-User Meeting 3	2.1	November 2008
[RD-104]	GlobCorine 2005 Project –Description and Validation Report (Report editors: Defourny, P., Bontemps, S., Van Bogaert E., Weber, J.L., Soukup, T., Kalogirou, V. and Arino, O.)	2.2	2010
[RD-105]	SRTM Water Body Data Product Specific Guidance.	2.0	2003
[RD-106]	Geoland 2 project, ATBD Small Water Bodies	1.0	2010
[RD-107]	Nyoungui, A. D., Tonye, E. and Akono, A., 2002, "Evaluation of speckle filtering and texture analysis methods for land coverclassification from SAR images". International Journal of Remote Sensing, 23, 9, 1895-1925	/	/
[RD-108]	Dell' Acqua, F. and Gamba, P., 2003, "Texture-based characterization of urban environments on satellite SAR images", IEEE Transactions on Geoscience and Remote Sensing", 41, 1, 153-159	/	/
[RD-109]	Dekker, R. J., 2003, "Texture analysis and classification of ERS SAR images for map updating of urban areas in the Netherlands". IEEE Transactions on Geoscience and Remote Sensing, 41, 9, 1950-1958	/	/
[RD-110]	Riedel, T., Thiel, C. and Schmullius, C., 2008, "An object-based classification procedure for the derivation of broad land cover classes using optical and SAR data". Proceedings of 1st International Conference on Object-based Image Analysis OBIA, Salzburg, 4 – 5 July, CD-ROM.	/	/
[RD-111]	SRTM 90m Digital Elevation Data – User Guide http://srtm.csi.cgiar.org/	/	/
[RD-112]	Digital Elevation Data over specific areas http://www.viewfinderpanoramas.org/dem3.html	/	/
[RD-113]	Corine Land Cover data (European Environment Agency) http://www.eea.europa.eu/publications/COR0-landcover	/	/

REF	TITLE	VERSION	DATE
[RD-114]	Global Lake and Wetlands Database http://www.worldwildlife.org/science/data/item1872.html	/	/
[RD-115]	Lehner, B. and Döll, P., 2004, "Development and validation of a global database of lakes,reservoirs and wetlands", Journal of Hydrology, vol. 296, pp. 1-22	/	/
[RD-116]	Ginevan, M. E., 1979, "Testing land-use map accuracy: another look", Photogrammetric Engineering and Remote Sensing, 45, 1371-1377	/	/
[RD-117]	Werner, C., Wegmüller, U., Strozzi, T. and Wiesmann, A., 2000, "GAMMA SAR and interferometric processing software". Proc. ERS-Envisat Symposium, Gothenburg, 16-20 October, 2000	/	/
[RD-118]	Anon., 2007, "ASAR Product Handbook", ESA ESRIN Issue 2.2, 27 February 2007	/	/
[RD-119]	Wegmüller, U., 1999, "Automated terrain corrected SAR geocoding", Proc. IGARSS'99, Hamburg, 28 June - 2 July 1999	/	/
[RD-120]	Quegan, S. and Yu, J. J., 2001, "Filtering of multichannel SAR images", IEEE Trans. Geosci. Remote Sensing, 39, 2373-2379	/	/
[RD-121]	Wiesmann, A., Wegmüller, U., Santoro, M., Strozzi, T. and Werner, C., 2004, "Multi-temporal and multi-incidence angle ASAR Wide Swath data for land cover information", Proc. 4th International Symposium on Retrieval of Bio- and Geophysical Parameters from SAR Data for Land Applications, Innsbruck, 16-19 November 2004	/	/
[RD-122]	Wegmüller, U., Werner, C., Strozzi, T. and Wiesmann, A., 2002, "Automated and precise image registration procedures", in Analysis of Multi-temporal remote sensing images, vol. 2, Series in Remote Sensing, Bruzzone and Smits, Eds. Trento, 13-14 September 2001: World Scientific, 2002, pp. 37-49	/	/
[RD-123]	Lopes, A., Nezry, E., Touzi, R. and Laur, H., 1993, "Structure detection and statistical adaptive speckle filtering in SAR images", Int. J. Remote Sens., vol. 14, pp. 1735-1758	/	/
[RD-124]	Oliver, C. and Quegan, S., 1998, "Understanding Synthetic Aperture Radar Images". Boston: Artech House	/	/
[RD-125]	Ulander, L.M.H., 1996, "Radiometric slope correction of synthetic-aperture radar images", IEEE Trans. Geosci. Remote Sensing, 34, 1115-1122	/	/
[RD-126]	Castel, T., Beaudoin, A., Stach, N., Stussi, N., Le Toan, T. and Durand, P., 2001, "Sensitivity of space-borne SAR data to forest parameters over sloping terrain. Theory and experiment", Int. J. Remote Sens., 22, 2351-2376	/	/

REF	TITLE	VERSION	DATE
[RD-127]	Santoro, M., Fransson, J.E.S., Eriksson, L.E.B., Magnusson, M., Ulander, L.M.H. and Olsson, H., 2009, "Signatures of ALOS PALSAR L-band backscatter in Swedish forest", IEEE Trans. Geosci. Remote Sensing, 47, 4001-4019	/	/
[RD-128]	Thiel, C. J., Thiel, C. and Schmullius, C., 2009, "Operational large-area forest monitoring in Siberia using ALOS PALSAR summer intensities and winter coherence", IEEE Trans. Geosci. Remote Sensing, 47, 3993-4000	/	/
[RD-129]	Santoro, M., Beer, C., Cartus, O., Schmullius, C., Shvidenko, A., McCallum, I., Wegmüller, U. and Wiesmann, A., 2011, "Retrieval of growing stock volume in boreal forest using hyper-temporal series of Envisat ASAR ScanSAR backscatter measurements", Remote Sens. Environ., 115, 490-507	/	/
[RD-130]	Swain, P.H. and Davis, S. M., 1978, "Remote Sensing: The Quantitative Approach". New York: McGraw-Hill	/	/
[RD-131]	Story, M. and Congalton, R.G., 1986, "Accuracy assessment: a user's perspective"; Photogrammetric Engineering and Remote Sensing, 52, 397-399	/	/
[RD-132]	Radoux, J. and Defourny, P., 2010, "Automated Image-to-Map Discrepancy Detection using Iterative Trimming", Photogrammetric Engineering & Remote Sensing, 76, 2, 173-181	/	/
[RD-133]	Desclée, B., Bogaert, P. and Defourny, P., 2006, "Forest change detection by statistical object-based method", Remote Sensing of Environment, 102, 1-11	/	/
[RD-134]	Langner, A., Miettinen, J. and Siegert, F., 2007, „Land cover change 2002–2005 in Borneo and the role of fire derived from MODIS imagery", Global Change Biology, 13, 2329–2340.	/	/
[RD-135]	Bontemps, S., Langner, A. and Defourny, P., 2012, "Monitoring forest changes in Borneo on a yearly basis by an object-based change detection algorithm using SPOT-VEGETATION time series", 33, 15, 4673-4699	/	/
[RD-136]	Cihlar, J., Ly, H., Li, Z., Chen, J., Pokrant, H. and Huang, F., 1997, "Multitemporal, multichannel AVHRR data sets for land biosphere studies-artefacts and corrections", Remote Sensing of Environment, 60, 35–57.	/	/
[RD-137]	Lillesand, T. M. and Kiefer, R. W., 2000, "Remote sensing and image interpretation", New York: John Wiley & Sons ,Inc.	4th edition	
[RD-138]	de Beurs & Henebry, 2005de Beurs, K. M. & Henebry, G. M. (2005), 'A statistical framework for the analysis of long image time series', International Journal Of Remote Sensing 26(8), 1551–1573.	/	

REF	TITLE	VERSION	DATE
[RD-139]	Hirsch et al., 1982Hirsch, R. M., Slack, J. R. & Smith, R. A. (1982), 'Techniques of trend analysis for monthly water quality data', Water Resour. Res. 18(1), 107–121. http://dx.doi.org/10.1029/WR018i001p00107	/	
[RD-140]	Gilbert, 1987Gilbert, R. (1987), Statistical Methods for Environmental Pollution Monitoring.	/	
[RD-141]	Atzberger & Eilers, 2011aAtzberger, C. & Eilers, P. H. (2011a), 'A time series for monitoring vegetation activity and phenology at 10-daily time steps covering large parts of south america', International Journal of Digital Earth 4(5), 365–386. http://www.tandfonline.com/doi/abs/10.1080/17538947.2010.505664	/	
[RD-142]	Atzberger, C. & Eilers, P. H. C. (2011b), 'Evaluating the effectiveness of smoothing algorithms in the absence of ground reference measurements', International Journal of Remote Sensing 32(13), 3689–3709. http://www.tandfonline.com/doi/abs/10.1080/01431161003762405	/	
[RD-143]	Eilers, 2003Eilers, P. H. C. (2003), 'A perfect smoother', Analytical Chemistry 75(14), 3631–3636. http://pubs.acs.org/doi/abs/10.1021/ac034173t	/	
[RD-144]	Whittaker, E. T., On a new method of graduation, Proc. Edinburgh Math. Soc., 1923, pp 63-75.	/	
[RD-145]	Riggs, G.A., Hall, D.K. and Salomonson, V.V., 2006. MODIS Snow Products, User Guide to Collection 5.	/	
[RD-146]	Saint, G.,(1994), "VEGETATION" onboard SPOT 4. Products specifications version 2	/	
[RD-147]	Chuvieco, E, Calado, T, Oliva, P (2011), ESA CCI ECV Fire Disturbance (fire_cci) Product Specification Document (PSD), version 2.0	/	
[RD-148]	Arino, O., S. Plummer, and D. Defrenne, (2005) Fire disturbance: the ten years time series of the ATSR World Fire Atlas, Proceedings of the MERIS-AATSR workshop 2005, Frascati, Italy, September 2005	/	
[RD-149]	Defourny, P., Lederer, D., Bontemps, S. (2011), LC-CCI Contract Change Notice 1. Water bodies detection using Envisat ASAR data	/	
[RD-150]	d'Andrimont, R., Pekel, JF., Bartholomé, E., Defourny, P. (2012); A continental scale near real time water body monitoring method based on MODIS daily data – application to Africa, Transactions on Geoscience and Remote Sensing, (Submitted)	/	
[RD-151]	Bouvet M., Ramoino F. (2010), Equalization of MERIS L1b products from	/	

REF	TITLE	VERSION	DATE
	the 2nd reprocessing, ESA TN TEC-EEP/2009.521/MB (available on demand at mbouvet@esa.int).		
[RD-152]	Giglio, L. , Loboda, T., Roy, D.P., Quayle, B., Justice, O.J., An active-fire based burned area mapping algorithm for the MODIS sensor, Remote Sensing of Environment, Volume 113, Issue 2, 16 February 2009, Pages 408-420, ISSN 0034-4257, 10.1016/j.rse.2008.10.006.		
[RD-153]	Hall, D. K., G. A. Riggs, J. L. Foster, and S. V. Kumar (2010), Development and evaluation of a cloud-gap-filled MODIS daily snow-cover product, Remote Sensing of Environment, 114(3), 496-503.		
[RD-154]	Klein, A. (2003), Validation of daily MODIS snow cover maps of the Upper Rio Grande River Basin for the 2000–2001 snow year, Remote Sensing of Environment, 86(2), 162-176.		
[RD-155]	Liang, T., X. Zhang, H. Xie, C. Wu, Q. Feng, X. Huang, and Q. Chen (2008), Toward improved daily snow cover mapping with advanced combination of MODIS and AMSR-E measurements, Remote Sensing of Environment, 112(10), 3750-3761.		
[RD-156]	Paudel, K. P., and P. Andersen (2011), Monitoring snow cover variability in an agropastoral area in the Trans Himalayan region of Nepal using MODIS data with improved cloud removal methodology, Remote Sensing of Environment, 115(5), 1234-1246.		
[RD-157]	Tekeli, A. E., Z. Akyürek, A. Arda Şorman, A. Şensoy, and A. Ünal Şorman (2005), Using MODIS snow cover maps in modeling snowmelt runoff process in the eastern part of Turkey, Remote Sensing of Environment, 97(2), 216-230.		
[RD-158]	Zhou, X., H. Xie, and J. M. H. Hendrickx (2005), Statistical evaluation of remotely sensed snow-cover products with constraints from streamflow and SNOTEL measurements, Remote Sensing of Environment, 94(2), 214-231.		
[RD-159]	Dozier, J., T. H. Painter, K. Rittger, and J. E. Frew (2008), Time–space continuity of daily maps of fractional snow cover and albedo from MODIS, Advances in Water Resources, 31(11), 1515-1526.		
[RD-160]	Simic, A., R. Fernandes, R. Brown, P. Romanov, and W. Park (2004), Validation of VEGETATION, MODIS, and GOES+ SSM/I snow-cover products over Canada based on surface snow depth observations, Hydrological Processes, 18(6), 1089-1104.		
[RD-161]	Maurer, E. P., J. D. Rhoads, R. O. Dubayah, and D. P. Lettenmaier (2003), Evaluation of the snow-covered area data product from MODIS, Hydrological Processes, 17(1), 59-71.		
[RD-162]	Arendt, A., T. Bolch, J.G. Cogley, A. Gardner, J.-O. Hagen, R. Hock, G.		

REF	TITLE	VERSION	DATE
	Kaser, W.T. Pfeffer, G. Moholdt, F. Paul, V. Radić, L. Andreassen, S. Bajracharya, N. Barrand, M. Beedle, E. Berthier, R. Bhambri, A. Bliss, I. Brown, D. Burgess, E. Burgess, F. Cawkwell, T. Chinn, L. Copland, B. Davies, H. De Angelis,E. Dolgova, K. Filbert, R.R.Forester, A. Fountain, H. Frey, B. Giffen, N. Glasser, S. Gurney, W. Hagg, D. Hall, U.K. Haritashya, G. Hartmann, C. Helm, S. Herreid, I. Howat, G. Kapustin, T. Khromova, C. Kienholz, M. Köonig, J. Kohler, D. Kriegel, S. Kutuzov, I. Lavrentiev, R. LeBris, J. Lund, W. Manley, C. Mayer, E. Miles, X. Li, B. Menounos, A. Mercer, N. Mölg,P. Mool, G. Nosenko,A. Negrete, C. Nuth, R. Pettersson, A. Racoviteanu, R. Ranzi, P. Rastner, F. Rau, B. Raup, J. Rich, H. Rott, C. Schneider, Y. Seliverstov, M. Sharp, O. Sigurðsson, C. Stokes, R. Wheate, S. Winsvold, G. Wolken, F. Wyatt, N. Zheltyhina. 2012, Randolph Glacier Inventory – A Dataset of Global Glacier Outlines: Version 3.2. Global Land Ice Measurements from Space, Boulder Colorado, USA. Digital Media		
[RD-163]	http://www.add.scar.org/index.jsp		

1.3 Acronyms

The acronyms used in this document are listed in Table 1-3.

Table 1-3: Acronyms

ACRONYM	MEANING
AATSR	Advance Along Track Scanning Radiometer
ALOS	Advanced Land Observing Satellite
AMORGOS	Accurate MERIS Ortho Rectified Geo-location Operational Software
ANN	Artificial Neural Network
AOD	Aerosol Optical Depth
AOT	Aerosol Optical Thickness
APP	Alternating Polarisation Mode Precision Image
ARVI	Atmospheric Resistant Vegetation Index
ASAR	Advanced Synthetic Aperture Radar
ASCII	American Standard Code for Information Interchange
ATBD	Algorithm Theoretical Basis Document
ATSR	Along track scanning radiometer
AVHRR	Advanced Very High Resolution Radiometer
BA	Burnt Area
BAE	Burnt Area Estimates

ACRONYM	MEANING
BEAM	Basic Envisat Tool for AATSR & MERIS (http://envisat.esa.int/services/beam/)
Bha	Bhattacharya
BOA	Bottom-Of-Atmosphere
BRDF	Bidirectional Reflectance Distribution Function
BRF	Bidirectional Reflectance Factor
CESBIO	Center for the Study of the Biosphere from Space
CCI	Climate Change Initiative
CCI-LC	CCI Land Cover
CDED	Canadian Digitial Elevation Dataset
CLC	Corine Land Cover
CLiC	Climate and Cryosphere Project
CLC2000	Corine Land Cover 2000
CMC	Climate Modelling Community
DBT2	Database for Task 2
DEM	Digital Elevation Model
DLM	Direct Location Model
DOM	Dark Object Method
DOY	Day of year
ECV	Essential Climate Variable
EEA	European Environment Agency
EGD	Effective Grain Diameter
ENL	Equivalent Number of Looks
ENVISAT	Environnement Satellite (http://envisat.esa.int)
EO	Earth Observation
ERS	European Remote Sensing Satellite
ESA	European Space Agency
EUMETSAT	European Meteorological Satellites Agency
FIREM3	The Fire M3 Hotspots product from the Canadian Wildland Fire Information System
FR	Full Resolution
FRS	Full Resolution Swath
FSC	Fractional Snow Cover
FSG	Full Swath Geo-corrected

ACRONYM	MEANING
G2	Geoland 2
GCOS	Global Climate Observing System
GDAL	Geospatial Data Abstraction Library
GFED.v3	Global Fire Emissions Database version 3
GLC2000	Global Land Cover 2000
GLWD	Global Lakes and Wetland Database
GMES	Global Monitoring for Environment and Security
HDF	Hierarchical Data Format
H-SAF	Hydrology and Water Management SAF
ICSU	International Council for Science
IFI	Irkutsk Forest Institute
IGOS	Integrated Global Observing Strategy
ILM	Inverse Location Model
IOC	Intergovernmental Oceanographic Commission
IODD	Input/Output Data Description.
JAXA	Japan Aerospace Exploration Agency
JM	Jeffries-Matsushita
KO	Kick-Off
L0, L1, L2, L3	Level 0, Level 1, Level 2, Level 3
LARS	Land Aerosol Remote Sensing
LC	Land Cover
LC-condition	Land Cover condition
LCCS	Land Cover Classification System
LSA	Land Surface Application SAF
LUT	Look-Up Table
MC	Mean Compositing
MCD64A1	MODIS Direct Broadcast Monthly Burned Area
MERIS	Medium Resolution Imaging Spectrometer
MHB	Minimum Hyper-temporal Backscatter
ML	Maximum Likelihood
MLC	Maximum Likelihood Classification
MODIS	Moderate Resolution Imaging Spectroradiometer

ACRONYM	MEANING
MSI	Multi-Spectral Imager
MTR	Mid-Term Review
MVA	Mean Annual Variability
NDII	Normalized Difference Ice Index
NDVI	Normalized Difference Vegetation Index
NDSI	Normalized Difference Snow Index
NGA	National Geospatial-Intelligence Agency
NGLD	Neighbourhood Grey level Dependency matrix
NIR	Near InfraRed
NOAA	National Oceanic and Atmospheric Administration
NRT	Near Real Time
OA	Overall Accuracy
OGR	OpenGIS Simple Features Reference Implementation
OLCI	Ocean and Land Colour Instrument
PA	Producer Accuracy
PALSAR	Phased Array-type L-band Synthetic Aperture Radar
PAR	Preliminary Analysis Report
PDF	Probability Density Function
PM	Progress Meeting
POLDER	POLarization and Directionality of the Earth's Reflectances
PSD	Product Specification Document
PVAS	Product Validation and Algorithm Selection Report
RB	Requirement Baseline Document
RMS	Root Mean Square
ROI	Region of Interest
RR	Reduced Resolution
RTC	Radiative Transfer Code
RTE	Radiative Transfer Equation
S-2	GMES Sentinel-2 (http://www.esa.int/esaLP/LPgmes.html)
S-3	GMES Sentinel-3 (http://www.esa.int/esaLP/LPgmes.html)
SAF	Satellite Application Facility
SAJF	Sensitivity Analysis Justification File

ACRONYM	MEANING
SAR	Synthetic Aperture Radar
SCA	Snow Covered Area
SCAR	Scientific Committee for Antarctic Research
SCIAMACHY	Scanning Imaging Spectrometer for Atmospheric CHartographY
SDR	Surface Directional Reflectance
SIFT	Scale-Invariant Feature Transform
SLSTR	Sea and Land Surface Temperature Radiometer
SMAC	Simplified Method for Atmospheric Correction
SoW	Statement of Work
SPOT	Satellite Pour l'Observation de la Terre
SPOT-VGT	SPOT-VEGETATION
SRTM	Shuttle Radar Topography Mission
SSA	Single Scattering Albedo
SSE	Sum of Squared Error
SSW	Snow Surface Wetness
STS	Snow Temperature for Surface
STSE	Support to Science Element
SURF	Speeded-Up Robust Features
SWBD	SRTM Water Body Data
SWE	Snow Water Equivalent
SWIR	Short-Wave InfraRed
TIR	Thermal InfraRed
TOA	Top Of Atmosphere
TRMM-VIRS	Tropical Rainfall Measuring Mission Visible and Infrared Scanner
TS	Technical Specification
TUL	Technical University of Lisbon
UA	User Accuracy
UADP	Urban Area Detecting Parameter
UNEP	United Nations Environment Programme
UNESCO	United Nations Educational, Scientific and Cultural Organization
UNFCCC	United Nations Framework Convention on Climate Change
UR	Utility Report

ACRONYM	MEANING
URD	Users Requirement Document
USGS	United States Geological Survey
UV	Ultra Violet
VNIR	Visible Near InfraRed
WB	Water Bodies
WCRP	World Climate Research Programme
WFA	World Fire Atlas
WFSC	Weekly aggregated Fractional Snow Cover
WMO	World Meteorological Organization
WS	Wide Swath
WSM	Wide Swath Mode

1.4 Organisation

This document includes the following sections:

- **Section 1** gives the scope, purpose, reference and applicable documents and list of acronyms, notations and conventions used in this document.
- **Section 2** provides an overview of the overall CCI-LC processing chain and its main processing steps.
- **Section 3** describes in detail the logical flow and individual processing steps carried out in the pre-processing part of the system.
- **Section 4** describes in detail the logical flow and individual processing steps carried out in the classification part of the system dedicated to the production of land cover maps.
- **Section 5** describes in detail the logical flow and individual processing steps carried out in the classification part of the system dedicated to the production of land cover conditions.

2 LC PROCESSING OVERVIEW

The CCI-LC project will deliver to the climate community 3 global land cover databases made of LC state for three epochs - centered around 2000, 2005 and 2010 - and of LC condition products.

The project builds upon the state-of-art established in the GlobCover project. As a result, the developed processing chain will be based on the GlobCover operational processing system [RD-1]. The classification module will be articulated to the pre-processing one in the way illustrated in Figure 2-1. The pre-processing module produces global composited mosaics of land surface reflectance values at a "fundamental" temporal resolution, which will constitute the input to the classification module, which interprets them into land cover classes. An additional module will be developed to generate the LC condition products.

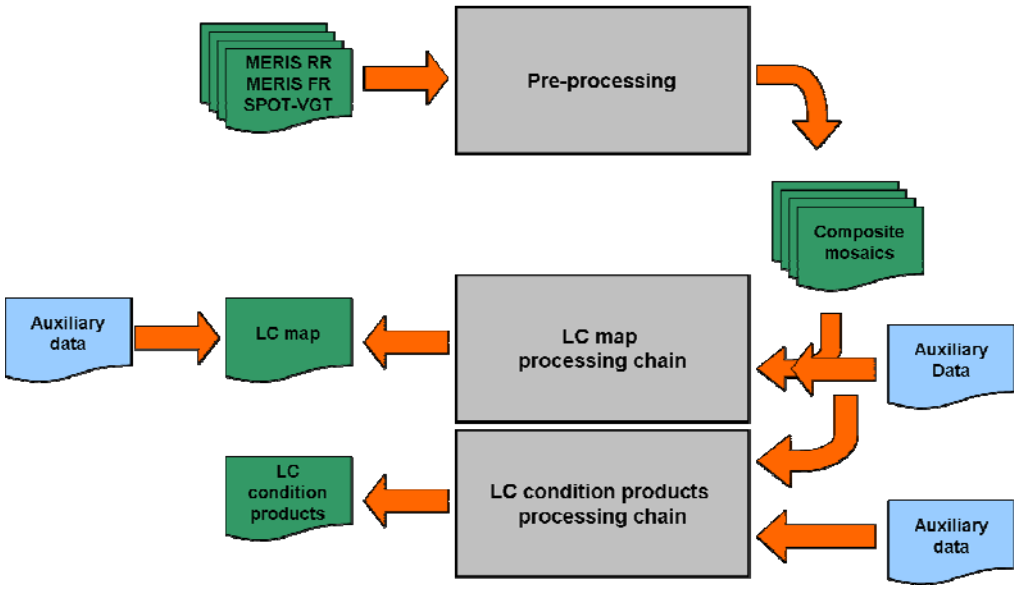


Figure 2-1: Schematic illustration of the pre-processing and classification modules (taken from [AD-8]).

The completed automated pre-processing chain performs the following operations (Figure 2-2): radiometric, geometric correction, pixel identification, atmospheric correction with aerosol retrieval as well as compositing and mosaicing (possibly including sensor merging). Improved algorithms will be developed and validated for each of these pre-processing steps. They are exhaustively detailed in section 3.

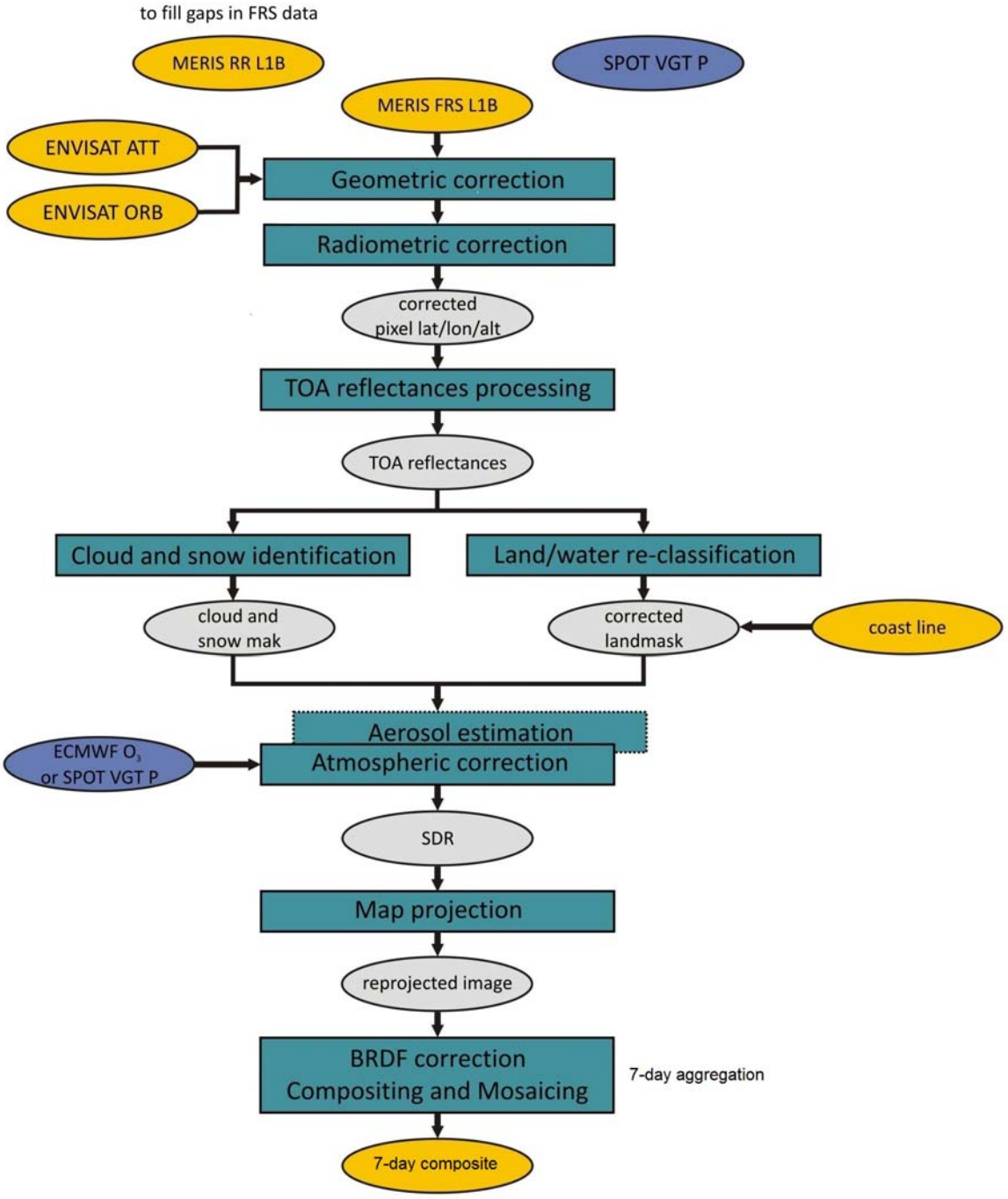


Figure 2-2: Schematic representation of the CCI-LC pre-processing chain including input (pre-processing chain based on the GlobAlbedo chain)

The classification process aims at transforming the cloud-free reflectance mosaics generated by the pre-processing module into a land cover map. It is organized in 4 main steps (Figure 2-3). First, a per-pixel spectral classification delineates spatially consistent clusters which show similar spectral properties. This spectral classification combines supervised and unsupervised approaches. The first one is used when available training data are reliable while the second one is used in the opposite case.

The two approaches are combined based on an objective decision criterion. Second, a temporal classification will be run to improve the accuracy and the homogeneity of some specific classes. This second step is optional. Third, spectral and temporal land cover maps are merged to deliver a first global land cover map. Post-classification editions are finally foreseen to deal with specific issues encountered in the classification or to add high-value external dataset (such as the Water Bodies product based on SAR data that is achieved in the CCI-LC project. Before the classification process, an a priori stratification step is applied in order to allow adjusting the classification parameters to sub-regions specificities (seasonality, landscape patterns, etc.). Section 4 focuses on this topic.

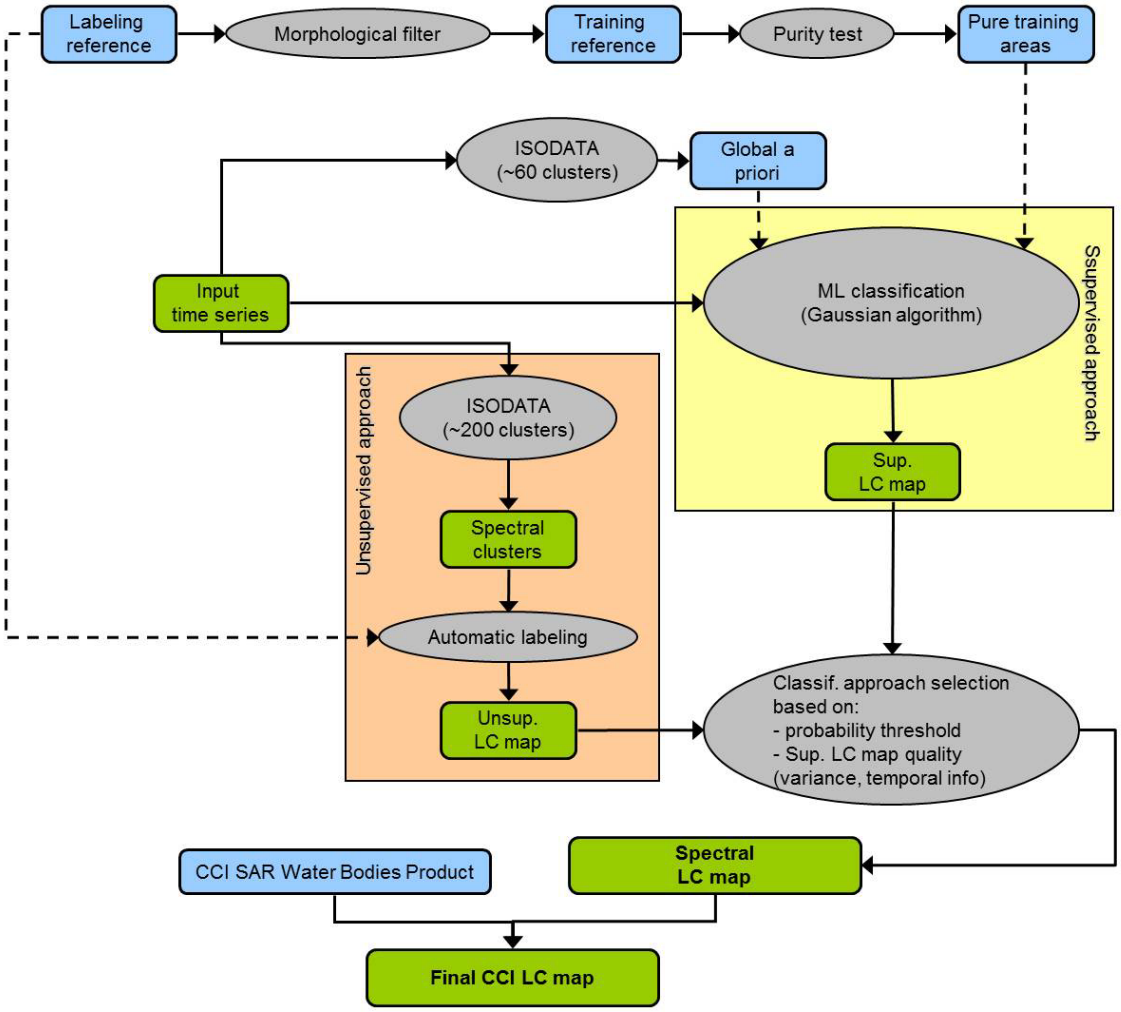


Figure 2-3: Scheme showing the principle of the CCI-LC classification algorithm (from [AD-8])

3 LAND COVER PRE-PROCESSING CHAIN

The CCI-LC project will significantly contribute to the Global Climate Observing System (GCOS) Land Cover Essential Climate Variable (ECV), by responding to the need for Product T.5.1 of GCOS-107, and actions T26, T27 of GCOS-92.

As already mentioned, the completed automated pre-processing chain generates global various temporal resolution composites of surface reflectances of MERIS and SPOT-VGT by a series of pre-processing steps (Figure 2-2). Improved and new algorithms developed for each of these steps are presented in the sections below.

3.1 Geometric correction

With AMORGOS (Accurate MERIS Ortho Rectified Geo-location Operational Software), ESA has made available a tool which improves MERIS FR geolocation better than 70 RMS. AMORGOS has been developed by ACRI-ST. The accuracy achieved by AMORGOS is considered to be sufficient for the CCI-LC requirements. However, AMORGOS is known to be computationally intensive – i.e. requiring long processing time – and having problems in high latitudes. Both are not on the critical path for this phase I of the CCI-LC project, but should be mentioned and brought forward to phase II. Further, it should be pointed out that AMORGOS needed an update for working with the MERIS data acquired after the orbital change manoeuvre in November 2010. An updated version for Linux operating systems (version 3.0p1) is now available, and provides compatibility with the new ENVISAT Orbit. This version has to be integrated in the pre-processing chain.

With regard to the inter-sensor geometric compatibility, the geometric accuracy of SPOT-VGT and MERIS RR, which are available on a 1-km grid, is less good. AMORGOS 3.0 did not work with MERIS RR products, and by definition is not applicable to any other sensors. The analysis of geolocation accuracy brought a misalignment of MERIS FR and RR forward, which is separately described in a technical note [AD-5]. In brief, in all probability, it is a combined effect of on-board resampling of CCD coordinates and on-ground geometric L0 → L1 processing. The updated version AMORGOS 4.0 is addressing this geolocation issue and therefore improves the matching between MERIS FR and RR data.

In order to achieve consistent geolocation accuracy across all used instrument data, a new module will have to be developed if the geometric accuracy of the SPOT-VGT products does not fulfil the CCI-LC requirements with respect to the geolocation. MERIS FRS products after application of AMORGOS will be taken as reference, and all SPOT-VGT shall be geo-referenced with respect to this MERIS RR.

3.1.1 AMORGOS 3.0

AMORGOS includes a precise orbit determination, instrument pointing and performs an ortho-rectification. BEAM Ortho-rectification only includes the ortho-rectification. It is therefore clear, that

the preferred solution is AMORGOS while BEAM ortho-rectification is being considered as a fall-back solution in case that AMORGOS totally fails. The criteria, which lead to the choice of the final tool, are not only the achieved geometric accuracy but also the:

- availability of the software;
- usability in the context of the CCI-LC processing chain;
- completeness of the available software, e.g. with respect to auxiliary data;
- processing performance.

If both candidates fail on these criteria, an alternative has to be found. Such an alternative could be an automated co-registration, as been proposed by Magellium. Figure 3-1 summarizes the decision path.

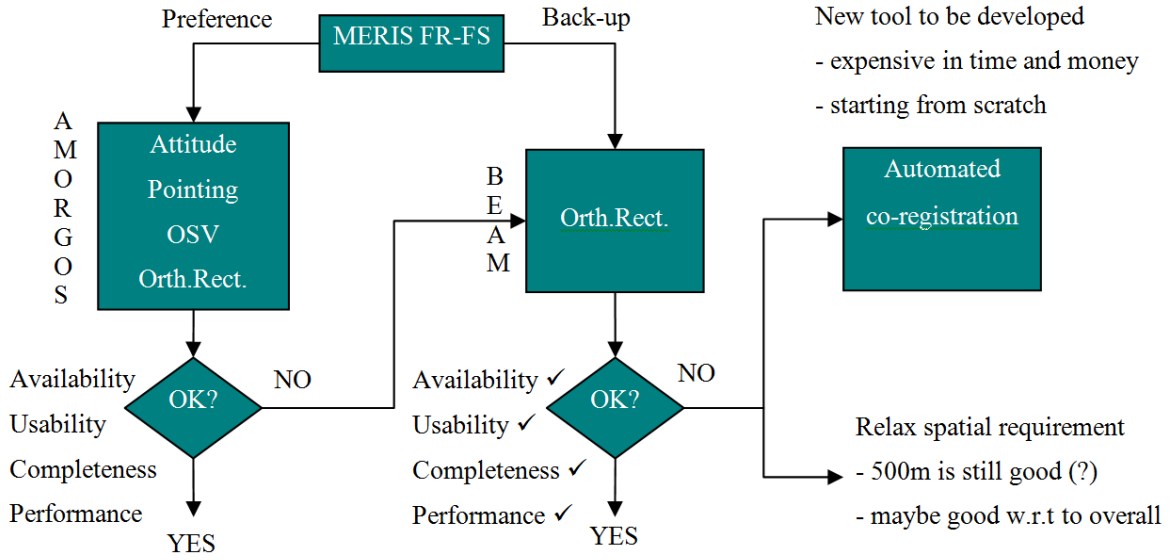


Figure 3-1: Decision logic for the selection of the geometric correction tool (from [RD-2])

The AMORGOS Software User Manual and Interface Control Document as well as the Software Transfer Document (Issue 2.0) were made available by ESA to the GlobCover team on 23.12.2005. This document is included in the GlobCover Design Justification File [RD-2] as Amorgos_ICD-SUM_2.0.doc and Amorgos_STD_i1r0.doc.

AMORGOS is run once for every MERIS_FRS_1P file and generates the corresponding MER_FSO_1P file. Re-build of the instrument projection is done using MER_FRS_1P internal data, the so-called MERIS detector index data set, and some additional MERIS Level 1b auxiliary data, the FR frame offset, extracted from appropriate MER_INS_AX file. For each MERIS pixel restored in instrument projection, an ortho-rectification algorithm computes the first intersection between the pixel's line of sight and the Earth surface, represented by interpolation of the Digital Elevation Model (DEM) cells elevations on top of the reference ellipsoid. Line of sight is determined using its pointing vector expressed relative to the satellite, the satellite location and attitude, which are in turn determined from Orbit and Attitude files using the appropriate CFI routines. Location of the intersection is expressed as longitude, geodetic latitude and geodetic altitude.

AMORGOS is producing a so called FSG (Full Swath Geo-corrected) product. This is structurally identical with the FRS product but with 3 additional bands for the precise and corrected latitude, longitude and altitude.

This version of AMORGOS has been applied to a large number of different FRS products, which partly were processed several times, in order to get an estimate of the processing performance. The tests were further done on two different computers and network environments in order to separate the influence of the HW from the pure AMORGOS performance. These tests and its results are described in a dedicated technical note (GLBC-AMORGOS-ProcessingTime.doc). The overall result of these tests is that AMORGOS processes 32000 pixel/sec on a 3GHz Xeon with Linux OS, 2GB RAM and data on local disks or accessible via a fast network. Extrapolated to the full CCI-LC dataset the processing would require 159 days on a single machine.

It should be noted that the processing performance depends on the product size. The smaller the product size the better is the performance as shown in the Figure 3-2.

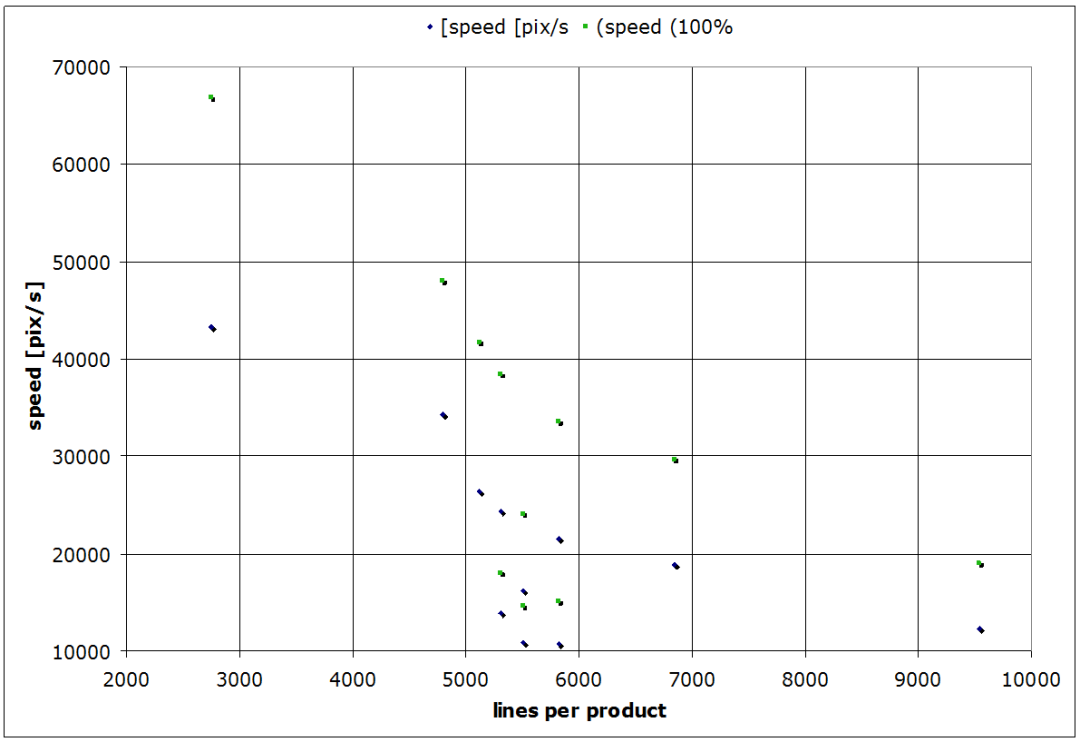


Figure 3-2: Processing performance (from the GlobCover 2005 project – [RD-2])

AMORGOS is executable software, which takes a MERIS FRS product as input and generates an FSG product as output. AMORGOS performs an improved navigation and pointing calculation and orthocorrects the data using an external digital elevation model. The FSG product in principle is an FRS product with three additional bands: corrected latitude and longitude, and the altitude. An important issue is the limitation of frame numbers for the FSG product. Due to output product time limits, the number of lines of an FSG product cannot exceed the value of 10305. Because the input FRS product may consist of up to 12865 lines, the processing of such a product has to be performed in two steps. AMORGOS provides input switches in the configuration file to handle this constraint.

Since AMORGOS is a binary distribution for Linux, it runs only under this operating system. It expects a certain directory structure, as depicted in the following Figure 3-3, which is taken from the AMORGOS Software Transfer Document. The figure does not contain the DEM directory, where the tiles of the digital elevation model reside.

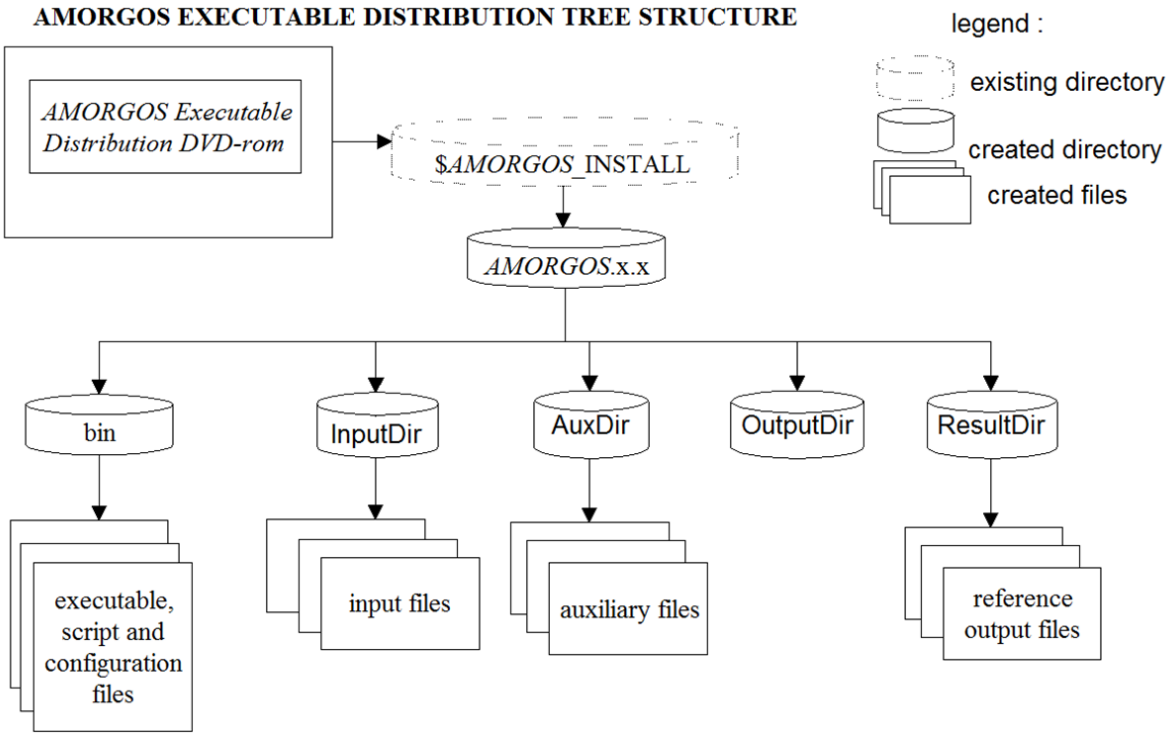


Figure 3-3: AMORGOS directory structure (from the GlobCover 2005 project – [RD-2])

3.1.2 AMORGOS 4.0

The improvements of AMORGOS V4.0 with respect to previous version (V3.0p1) are:

- the processing of MERIS RR L1b products (MER_RR__1P) into geo-referenced RR L1B (MER_RRG_1P);
- the handling of the both reprocessing 2 and reprocessing 3 data (the difference affecting the format of an L1b processor auxiliary data file also used by AMORGOS, not the MERIS L1b products).

Consequently, the testing philosophy addresses the two changes listed above by verifying that: MERIS RR georeferenced products generated by AMORGOS V4.0 give results equivalent to:

- MERIS FR georeferenced products with same coverage generated by AMORGOS V4.0, the FR mode being already validated;
- MERIS RR georeferenced products are generated by MEGS, the MERIS Ground Segment test platform, in which the georeference option has been implemented and validated for both RR and FR.

- MERIS FR georeferenced products generated by AMORGOS V4.0 give results equivalent to products with same coverage generated by AMORGOS V3 (or MEGS, the latter having been validated against AMORGOS V3);
- AMORGOS V4.0 results are equivalent when obtained from MERIS 2nd re-processing data and MERIS 3rd reprocessing data with identical coverage, and that for both RR and FR.

Test scenes have been selected so as to offer various types of terrain, in particular including mountainous areas with a high variability of elevation. In all cases, validation is assessed through comparisons of (a) pixel altitudes with respect to the reference and (b) pixel geolocation through computation of ground distance with respect to the reference geolocation [RD-3].

Short summary of validation results:

- AMORGOS V4.0 RR versus FR - test is fully satisfactory;
- AMORGOS V4.0 RR versus MEGS RR - test is fully satisfactory;
- AMORGOS V4.0 FR versus MEGS FR- test is fully satisfactory;
- AMORGOS V4.0 from re-processing 3 versus AMORGOS V4 from re-processing 2 - test is fully satisfactory for both resolutions.

3.1.3 Scene collocation

The investigation of the geolocation accuracy of the MERIS FSG and RRG as well as the SPOT VGT P products revealed that the geometric accuracy of all products fulfils the CCI-LC requirements [AD 8]. The merging of the composites requires a collocation of MERIS FSG and MERIS RRG as well as SPOT VGT P or S1 products. This will be performed with the scene collocation described here. The collocation of geo-corrected MERIS FSG scenes with geo-corrected MERIS RRG and SPOT-VGT P or S1 scenes solely based on their geocodings (image geometry) is applied.

The Collocation Tool allows collocating two spatially overlapping products. Collocating two products implies that the pixel values of one product (the slave) are resampled into the geographical raster of the other (the master). For flag bands and bands where a valid-pixel expression is defined, the Nearest Neighbour method is used. In order to establish a mapping between the samples in the master and the slave rasters, the geographical position of a master sample is used to find the corresponding sample in the slave raster. If there is no sample for a requested geographical position, the master sample is set to the no-data value which was defined for the slave band. The collocation algorithm requires accurate geopositioning information for both master and slave products.

For MERIS RRG and SPOT-VGT P scenes, we can assume a sub-pixel accuracy of their provided geocodings. In contrast to MERIS FSG scenes, these geocodings yield for each pixel location a geographical coordinate of a given reference ellipsoid (e.g. WGS-82) that closely matches the Earth's surface. For a collocation of all three sensors, these geographical coordinates need to be corrected for an underlying DEM. Ideally, the DEM used for the generation of the MERIS FSG products is also used for the geo-correction of the MERIS RRG.

A geocoding comprises a direct location model (DLM) and an inverse location model (ILM). The DLM transforms image pixel coordinates into geographical coordinates, g = DLM(p) while the ILM does the opposite: p = ILM(g).

The DLM is usually issued from a well-known satellite viewing model. In many cases it is given as a set of tie-points, each point associating a pixel position with a geographical coordinate. For MERIS, pixel positions are given for the sub-sampled scene image and organised in rectangular tie-point grids. The ILM is most often only a predictor, for example estimated by polynomial (rational) functions or a search algorithm that finds the closest pixels positions for a given geographical coordinate. The latter is true for the MERIS FSG and RRG products – they provide a unique DEM-corrected geographical coordinate for each individual pixel in the image.

For the collocation of a MERIS FSG image with, for example, an SPOT VGT-P image, the SPOT VGT-P pixel for a given MERIS FSG pixel is found by (eq. 3-1):

$$g = ILM_{MERIS\ FSG}(p_{MERIS\ FSG})$$
$$p_{SPOT\ VGT\text{-}P} = DLM^{*}_{SPOT\ VGT\text{-}P}(g)$$

where DLM^{*} is the direct location model of an orthorectifying geocoding that is a function of $ILM_{SPOT\ VGT\text{-}P}$, $DLM_{SPOT\ VGT\text{-}P}$ and a DEM. An implementation and documentation of such a geo-coding is available in the BEAM toolbox (as part of the orthorectification tool) [RD-93, RD-94]. Similar to AMORGOS, BEAM uses the GETASSE DEM [RD-95].

3.2 Calibration – MERIS calibration and smile correction

3.2.1 MERIS calibration

Spectral and radiometric calibrations are key processes for all quantitative algorithms which exploit the spectral shape and magnitude. The spectral calibration of MERIS is performed by using measurements of a coloured diffuser and by a spectral matching technique using the measurements of the $O_{2\text{-}A}$ absorption band. The spectral calibration has been proved to be within 0.1nm accuracy. This is considered to be adequate for the generation of CCI-LC products. The radiometric accuracy is more critical. MERIS has two diffusers onboard, the first one being used regularly to monitor the radiometric response and the second one being exposed to sun light only 3 to 4 times per year in order to monitor the ageing of the first diffuser. With this technique it was possible to derive a new degradation model of the instrument, which is applied for the third reprocessing, which is taking place in Winter 2009/Spring 2010. This reprocessing quality is required for CCI-LC. The reprocessing, however, concerns only Reduced Resolution products. In spring 2011, also the IPF (the operational processor) will be upgraded with the new degradation model and from that point onwards, also FRS products will be processed with the improved calibration. But all archived FRS products will remain at old calibration status. Hence, an algorithm was developed to correct 2nd re-processing products to 3rd reprocessing radiometric quality.

The radiometric calibration as described below is a non-linear process including several steps. The radiometric gains are the second last step before the L1b are written. However, the last step is the

straylight correction which is a non-linear process and not revertible from L1b product. The Improved Calibration is therefore only an approximation.

The valid MERIS samples are digital counts resulting from the detection and acquisition by MERIS of a bi-dimensional field of spectral radiance in front of the instrument. The objective of the radiometric processing, together with the stray light correction, is to estimate that spectral radiance. An inverse model of the MERIS processing is used for that purpose, using parameters stored in the Characterisation and Radiometric Calibration data bases and the MERIS samples themselves. The MERIS acquisition model may be described as (eq. 3-2):

$$X_{b,k,m,f} = \mathrm{NonLin}_{b,m}\left[g(T_f^{VEU}).\left[A_{b,k,m}.\left(L_{b,k,m,f} + G_{b,k,m}(L_{*,*,*,f}) \right) + Sm_{b,k,m,f}(L_{b,k,m,*}) \right] + g_c(T_f^{CCD}).C^0_{b,k,m} \right] + \varepsilon$$

where

- M is the camera (or module); b is the spectral band; k is the pixel column; f is a frame (processing unit of number of image lines)
- $X_{b,k,m,f}$ is the MERIS raw sample
- $\mathrm{NonLim}_{b,m}$ is a non-linear function, representing the non-linear transformations which take place in the CCD, amplifier and A/D converter; NonLin depends on band and gain settings
- T_f^{VEU} is the temperature of the MERIS amplifiers (VEUs) at the time of frame f
- T_f^{CCD} is the temperature of the MERIS detectors (CCDs) at the time of frame f
- g and g_c are (dimensionless) temperature correction functions
- $AL_{b,k,m}$ the "absolute radiometric gain" in counts/radiance unit; AL depends on band & gain settings
- $L_{b,k,m,f}$ the spectral radiance distribution in front of MERIS
- $Sm_{b,k,m,f}$ the smear signal, due to continuous sensing of light by MERIS
- $C^0_{b,k,m}$ the calibrated dark signal (possibly including an on-board compensation), dependent on band and gain settings
- $G_{b,k,m}$ a linear operator (weighted sum) representing the stray light contribution to the signal. For a given sample, some stray light is expected from all the other samples in the module, spread into the sample by specular (ghost image) or scattering processes
- ε is a random process representative of the noise and measurement errors

This model is inverted during processing: The inverse of the absolute instrument gain ALb,k,m is applied to the valid samples of all bands after dark and smear signal subtraction, with a compensation for the estimated temperature which is expressed as a function of time (eq. 3-3):

$$R_{b,k,m,f} = \left(AL_{b,k,m}\right)^{-1}.\left\{ \left(X'_{b,k,m,f} - S_{b,k,m,f} \right).\left[g_0 + g_1(t_f - t_{ref}) + g_2(t_f - t_{ref})^2 \right] - C_{b,k,m,f} \right\}$$

where $R_{b,k,m,f}$ are the spectral radiances before the straylight correction. The CoastColour Improved calibration is assuming the final straylight corrected radiances as equal to this $R_{b,k,m,f}$. The 2nd

reprocessing radiometric gains (AL) are multiplied to R, and then the inverse of the 3rd reprocessing gains are multiplied to give an estimate of the 3rd reprocessing radiances.

3.2.2 Coherent noise equalisation

The MERIS equalization module performs a radiometric equalisation of the MERIS L1b products. It reduces detector-to-detector and camera-to-camera systematic radiometric differences and results into a diminution of the vertical stripping observed on MERIS L1b products. The MERIS swath is imaged by a CCD. The radiance at each pixel of a MERIS L1b products results from the measurements of 5 cameras spread across the swath, each one imaging a part of the swath with 740 so-called detectors in FR (corresponding to 185 mean detectors in RR). This results into 3700 detectors imaging the swath of MERIS FRS product (925 in RR). The response of each one of these detectors is calibrated during the routine operation of the instrument. Residual uncertainties in the calibration process result into detector-to-detector and camera-to-camera systematic radiometric differences. The equalisation corrects for these radiometric differences via a set of detector dependant coefficients correcting for the residual uncertainties in the calibration process. These coefficients are retrieved via a methodology described in Bouvet and Ramino, 2010, based on observations of the Antarctica plateau spread out throughout the MERIS mission lifetime.

The coefficients are different for MERIS Reduced and Full Resolution products. While the compilation of a full mission data set in reduced resolution, and subsequently the calculation of the noise equalisation coefficients, this process is not yet completed for full resolution data. Consequently the coherent noise equalisation is not included in the CoastColour Preliminary Processing [RD-151].

3.2.3 Smile correction

MERIS is measuring the reflected sun light using CCD technquie. A CCD is measuring in one of its dimensions one image line, and in the other dimension the spectrally dispersed radiance for each pixel along the image line. I.e., the spectral measurements of each pixel along an image line is made by it's own set of sensors of the CCD. This causes small variations of the spectral wavelength of each pixel along the image. This is called the "smile effect".

The MERIS instrument is composed of 5 cameras, each equipped with it is own CCD sensor. The variation of the wavelength per pixel is in order of 1nm from one camera to another, while they are in the order of 0.1nm within one camera.

Even though this variation is small compared to the spectral bandwidth of a band, which is typicall 10nm, and can hardly be seen in an image, it can cause disturbances in processing algorithms which require very precise measurements, for example the retrieval of chlorophyll in the ocean. These disturbances can result in visual artefact ("camera borders") or reduced accuracy of the Level 2 products. Therefore, the MERIS Level 2 processor corrects the smile effect. The Level 1b product is not smile corrected, because this product shall provide the user exactly what the instrument is measuring, and that is in fact the radiance at the given wavelength of each pixel.

The smile correction consists actually of two parts: a radiance correction and a reflectance correction.

The reflectance correction consists of an estimation of the reflectance spectral slope from the measurements in two neighbouring bands, and the interpolation of the reflectance along this slope from the actual pixel wavelength to the nominal wavelength of the considered band.

It is defined in an external configuration data file, which bands will be corrected, and which bands are used for the slope estimation. The corrected reflectance is calculated according to (eq. 3-4):

$$\rho_{slope} = \frac{\rho_{brr}(b_2) - \rho_{brr}(b_0)}{\lambda_{pix}(b_2, detector(j)) - \lambda_{pix}(b_0, detector(j))}$$

$$\rho_{sm-corr}(b_1) = \rho_{brr}(b_1) + \rho_{slope}(\lambda_{nom}(b_1) - \lambda_{pix}(b_1, detector(j))$$

with ρ_{brr} the Bottom-of-Rayleigh reflectance, λ_{pix} the wavelength per pixel and $\rho_{sm-corr}$ the smile corrected reflectance.

Table 3-1: Smile correction

PARAMETER	DESCRIPTION	INTENT (IN, OUT, INOUT)	PHYSICAL UNIT
ρ_{brr}	Bottom-of-Rayleigh reflectances	IN	none
λ_{pix}	wavelengths per pixel	IN	nm
λ_{nom}	nominal wavelengths	IN	nm
$detector(j)$	the detector index at the pixel location of the spectra	IN	none
b_1	the band index of reflectance to be corrected	IN	none
b_2	upper band index for interpolating the slope	IN	none
b_0	lower band index for interpolating the slope	IN	none
ρ_{slope}	slope of reflectance	IN	none
$\rho_{sm-corr}$	smile corrected reflectance per band and pixel	OUT	none

3.3 Land water delineation and cloud screening - pixel identification

The Multi Sensor Pixel Identification approach (which is also applied in the GlobAlbedo project) classifies each pixel to be processed according to a series of pixel categories, which include cloud, clear-land, clear-water and clear-snow. Cloudy pixels are not processed in CCI-LC project, while land, water and snow pixels must be distinguished because of the particular processing steps associated to each surface type. In particular, water pixels must be separated from land surfaces even in the case of continental water bodies, as these are flagged in the final pre-processed product. Snow and snow-free surfaces will also be considered separately in the final pre-processed product.

3.3.1 Pixel identification

The term “Pixel identification” refers to a classification of a measurement made by a space borne radiometer, for the purpose of identifying properties of the measurement which are influencing further algorithmic processing steps. Most importantly is the classification of a measurement as being made over cloud, a clear sky land surface or a clear sky ocean surface. The term “pixel” is often used for such a measurement in order to express it being part of a spatially oriented collection of many measurements, which all are geo- located and which form, as a whole, an image of the earth below the satellite.

While the information if a pixel is made over water or land can be taken from a static map, provided the geo-location of the pixel is better than the size of the pixel, as a good first guess, the cloud coverage is spatially and temporally highly variable and needs to be derived from the measurement itself. After knowing whether a pixel is cloudy or clear, in the clear sky case the land-water information can be refined using the measurement. This is particularly necessary in the coastal zone where the actual land-water boundary is changing due to tides, when the pixel size is small enough to resolve this difference. Also maps are not always correct so that a radiometric refinement is advisable.

3.3.2 Background

A large portion of the earth surface is covered by clouds [RD-11]. Consequently, most earth observation images in the visible spectral domain include a significant amount of cloudy pixels. Such measurements are treated in two opposite ways: either cloud properties are retrieved , e.g. for weather forecast or climate studies [RD-12, RD-13, RD-14], or the focus of the interest is the earth surface – being it land or water – which is then masked by the cloud [RD-15]. In the latter case, the presence of the cloud needs to be identified, and the change of the surface reflectance due to the cloud has to be estimated.

An image pixel can be cloud free, totally cloudy, or partly cloudy. In the cloud free case there are no water droplets or ice crystals in the atmosphere which change the surface reflectance. In the totally cloudy case the optical thickness is so high that the portion of surface reflectance at the signal measured by the satellite is negligible. The partly cloudy case comprises all intermediate situations where the measured reflectance is a mixture of a significant portion of the surface reflectance, but modified due to the presence of a cloud. This can be either due to an optically thin cloud, or the cloud is covering only a fraction of a pixel in the field of view of the sensor [RD-16].

Cloud free and totally cloudy pixels can be identified rather easily, and most of the tests used in earth observation processing systems for cloud identification today, assign either of these two stages, and hence also partly cloudy cases have to be assigned to either of these two classes [RD-17]. For spatial high resolution instruments such a binary cloud flag is not appropriate if several different higher level processing algorithms are applied, each of which having a different robustness to partial cloudy pixel [RD-18]. Some novel algorithms therefore deliver a graduated scale, as an indicator of the extent to which a signal is influenced by the presence of clouds [RD-19, RD-20 and RD-21]. Such an indicator can be related to cloud properties, e.g. apparent cloud optical thickness, the atmospheric transmission [RD-19], or cloud features [RD-20].

Clouds have certain characteristics which can be used for their identification and characterisation [RD-15]: brightness, whiteness, cold temperature and high altitude. However, none of these characteristics is always given if a pixel is cloudy; this is the main problem of cloud identification.

Thin clouds are difficult to differentiate from bright land surfaces and clouds in the mountains can be lower than the mountain ridges around. Then other methods not based on the features given below must be used. In particular, the clouds can be also detected using the spatial and temporal variability of the reflected radiation. In addition, clouds screen the tropospheric gases. This leads to the increase in the reflection inside corresponding gaseous absorption bands (e.g. H_2O, CO_2, O_2) which is routinely used for the cloud top height monitoring [RD-22].

One way to detect clouds would be to work directly with optical measurements. Further, derived cloud physical properties can be used to characterise clouds and assess their impact on the retrieved signal. This includes, amongst others, cloud fraction, cloud top temperature, cloud top pressure, cloud type, cloud phase, cloud optical depths and cloud effective particle size. Such properties can be studied using the radiative transfer modelling. Extensive work in this respect over the past years can be found in the literature [RD-22, RD-23, RD-24, RD-25, RD-26 and RD-27]. They have developed the MERIS algorithms for cloud top pressure, cloud optical thickness, cloud albedo and cloud type retrieval and have translated this knowledge into a probability based cloud detection algorithm [RD-16].

Cloud detection became important with the systematic processing of the NOAA AVHRR instrument in the 1980s. Statistical histogram analysis methods were developed [RD-28]. Most common used were threshold algorithms (e.g. in [RD-29]). Large scale textures were identified using pattern recognition techniques [RD-30]. These methods worked quite well over the ocean but exposed problems in Polar Regions (separation of clouds from ice and snow) and in the tropics (low level, warm clouds). A good overview of the cloud screening techniques at the late 80s is given in [RD-31]. Improved methods are later proposed for the AVHRR and for ATSR.

The cloud screening algorithms for the ATSR 1 and 2 in the 1990s were mainly based on previous work for AVHRR and use spectral threshold tests [RD-32]. The thermal band at 12µm is used as main tool to identify the cold cloud surface by a threshold, supported by other thresholds on band differences and on the histogram of the radiance distribution in the image. The unique feature of two views under different angles of the same pixel and the spatial coherence of the radiance are also exploited. The cloud screening of the AATSR is basically the same with refined and additional tests due to additional bands. Recently, tests on vegetation and snow indices have been introduced [RD-32]. However, application oriented projects are not satisfied with the standard cloud screening and are proposing alternative methods, for example for the GlobCarbon processing.

The MERIS Level 2 cloud screening is a combination of 8 different tests [RD-33]. Three of those are classical threshold tests on spectral radiances or differences, and five are connected with the pressure estimates derived from the differential oxygen A-band absorption measurements. However, due to the current insufficient quality of the standard pressure products derived from the measurements, these tests are not used.

The potential of the O2A feature has been addressed recently in ESA funded projects, namely “Exploitation of the oxygen absorption band” and “MERIS AATSR Synergy”. The result of these activities has lead to an upgrade of the operational MERIS pixel classification in the third

reprocessing. Major improvement is due to including dedicated pressure algorithms for detection of the height of the scattering surface over land and ocean.

The strong water vapour absorption at 1.38µm can be used to detect the presence of high clouds, including thin cirrus under daytime viewing conditions. With sufficient atmospheric water vapour in the beam path, almost no upwelling reflected radiance from the earth's surface reaches the satellite which is in particular handy for snow covered surfaces. However, precipitable water is often less than 1 cm over polar and in high elevation regions. The 1.38 µm reflectance threshold is set to 0.03 for MODIS [RD-34].

A big problem is the distinction between clouds and snow/ice, in particular for instruments which do not have spectral bands in the NIR and SWIR. An extensive study including the cloud screening over snow and ice has been undertaken [RD-35, RD-36, RD-37]. Snow and ice are less reflective in the NIR spectral region, and the so called normalized differentiation ice index (NDII) and the corresponding snow index (NDSI) is a good tool to differentiate clouds from snow and ice. These indices are defined as follows (eq. 3-5):

$$NDII = \frac{R(0.545\mu m) - R(1.05\mu m)}{R(0.545\mu m) + R(1.05\mu m)}, \quad NDSI = \frac{R(0.545\mu m) - R(1.64\mu m)}{R(0.545\mu m) + R(1.64\mu m)}$$

The reflectance for ice decreases with the wavelength must faster as compared to snow. Therefore, large values of NDII signify the bare ice case.

Also measurements of trace gas vertical columns (e.g., SCIAMACHY onboard ENVISAT) are disturbed by cloud presence because corresponding instruments have large fields of view to enhance the sensitivity to small gaseous concentrations. Cloud clearing algorithms are described for GOME in [RD-38], and [RD-39] reports on using MERIS to support the cloud screening for SCIAMACHY.

3.3.3 Methods

Cloud detection methods can be categorized in the following classes [RD-18]:

Spectral threshold methods

Spectral characteristics, such as temperature, brightness, whiteness or height of the scatterer are tested against a threshold value. The threshold can be parameterized by viewing geometry, location or time. Most cloud screening algorithms given in the reference list include such tests.

A special subsection of these tests concern spectral high resolution methods. Feature selection or PCA pre-processing is sometimes applied to reduce the dimensionality of the dataset [RD-40, RD-41, RD-42 and RD-43].

Feature extraction and classification

The spectral data space, if transformed into a feature space, can be statically or dynamically (i.e. scene dependent) separated into cloud or clear classes. This group of algorithms also includes spatial structure based algorithms. Examples are given [RD-20].

Learning algorithms

The Bayesian probability approach and general data mining techniques are employed. Cloud probability or cloudiness index values are generated after training the algorithm with simulated or measured data. Examples are given in [RD-44, RD-20] for AATSR and in [RD-19] for MERIS. A generic approach of a learning algorithm has been developed [RD-45].

Multitemporal analysis

Pixels are not always cloud covered and a time series of data is used to separate cloudy from clear cases. For example, such kind of method is applied in the Cyclopes processing [RD-47].

Multi sensor approach

In cases, where multiple sensors are on the same platform and perform simultaneous measurements, the synergetic algorithms can be used to better identify clouds. This was considered, for example, in the case of MERIS and SCIAMACHY [RD-39] and MISR and MODIS [RD-47].

As it follows from the discussion given above, the screening procedures are of great importance for successful retrievals of snow properties from space.

3.3.4 Theoretical description

The current operational detection of clouds in SPOT-VGT data relies on spectral threshold tests using the reflectance in the blue and SWIR bands [RD-48]. Different threshold combinations are used to identify cloudy and clear pixels. A pixel which does not pass either test is declared uncertain. A snow mask is calculated using spectral threshold tests on the red and MIR channels combined with 3 spectral slope tests which exploit the lower scattering of snow in MIR and SWIR bands compared with clouds. A cloud shadow is finally added based on an estimation of a potential cloud shadow and a test on the NDVI of concerned pixels.

Current MERIS cloud screening uses spectral thresholds on shortwave bands, complemented by spectral slope tests in order to recover bright land surface and snow [RD-49]. In the current reprocessing of MERIS these cloud and snow tests are significantly changed and improved (Brockmann and Santer, in preparation) by adding tests on the height of the scattering surface (based on the oxygen absorption measurements in MERIS band 11), and new tests for snow and ice detection using the MERIS Differential Snow Index (MDSI), based on the ratio of bands 13 (865nm) and 14 (885nm).

Both MERIS and AATSR cloud screening are not optimal because of missing spectral information in each of the single instrument (SWIR and TIR bands in MERIS, O2 and water vapour bands in AATRS). In the framework of the MERIS – AATSR synergy project an algorithm has been developed that combines the data from both instruments [RD-50]. A thorough analysis has been undertaken on the information content w.r.t to cloud detection in both instruments, and a set of 19 features has been identified as optimal with respect to the number of features (which should be kept low) and information content. These features include the spectral reflectances of the two instruments, and a number of band combinations. The cloud screening algorithm is a combination of feature extraction and supervised classification and spectral unmixing. The training vectors for the supervised classification have been obtained from a database of radiative transfer calculations. The results of the

algorithm are a binary cloud mask resulting from the feature tests and a cloud abundance values (between 0 and 1) from the unmixing. These two values can be used by subsequent algorithms to decide if a pixel can be processed, or a final logic is applied to conclude on the pixel status.

A critical step for the synergistic use of MERIS and AATSR is the collocation of the products. Due to the high spatial and temporal dynamic of clouds, misalignment of the two data sets would impact the cloud retrieval. Figure 3-4 shows the steps included in the pre-processing of the data of the two instruments.

Neither of the single instrument algorithms discussed in the previous section is considered of sufficient quality for the purpose of the GlobAlbedo project. The MERIS and AATSR cloud screening algorithms have been criticised on the MERIS-AATSR user workshops, and also the SPOT-VGT algorithm does not screen out sufficiently the doubtful cloudy pixels. The reason is that the global Level 2 algorithms cannot be too severe in order to permit analysis of single Level 2 products. On the contrary, for a Level 3 product such as the albedo, a clear sky conservative (i.e. severe cloud screening) approach is required. Even a small number of undetected clouds can significantly impact the final albedo product. This has been demonstrated in the MERIS GlobAlbedo and the GlobCover projects.

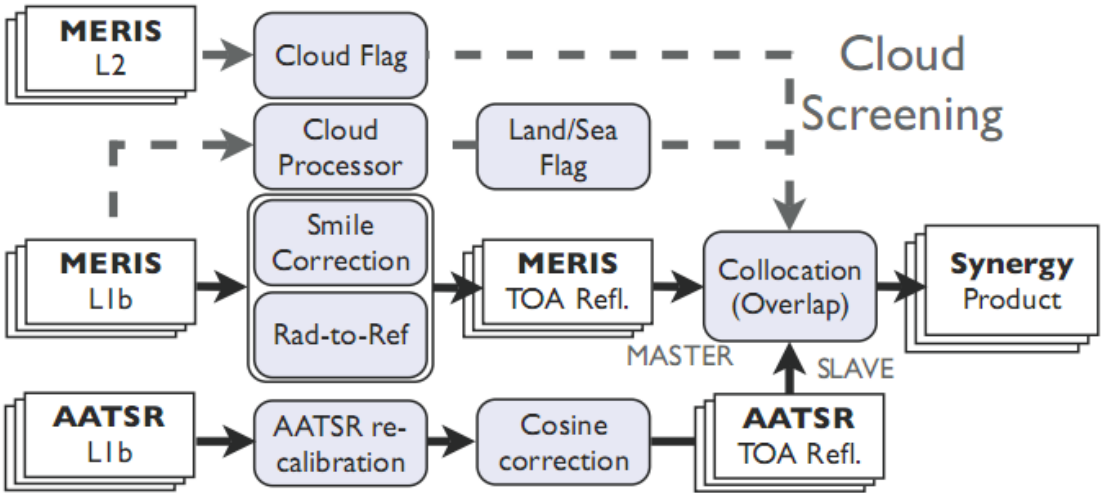


Figure 3-4: Flowchart of the MERIS and AATSR synergy preprocessing module

The synergistic method presented in [RD-50] is a significant improvement. It exploits optimally all features available in both instruments and combines these in a non-linear, self trained mathematical way. The comparison shown in the MERIS-AATSR-Synergy Project demonstrates the improvements compared to the standard MERIS and AATSR algorithms.

However, for the CCI-LC a synergistic use of MERIS and AATSR is not foreseen because of (1) the limitation of AATSR to the centre of the MERIS swath, thus not providing data for half of the MERIS swath and (2) the overall design of the processing architecture which does not foresee multi-sensor processing at this stage. The first point will change with Sentinel 3, and the latter point could of course be resolved if AATSR would be included.

The CCI-LC pixel classification will be a unique method, adapting the principles and mathematical implementation of several of the above approaches and features provided by the MERIS and SPOT-VGT, each treated separately.

In a first step the features will be combined in a logical order of a sequence of threshold tests, but working on and resulting in probability values between [0 … 1]. The combination will be done by arithmetic operations, addition and multiplication. This extends the Boolean logic into a probabilistic space. If tests result in the extreme 0 and 1, the probabilistic calculations are identical with the Boolean expressions.

3.3.5 Practical consideration

3.3.5.1 Overall principles

The CCI-LC pixel identification is a unique classification for the MERIS FR, RR and SPOT-VGT data considered (Figure 3-5). The uniqueness consists of a certain set of features, which are calculated for each instrument and probabilistic combination of these features in order to calculate a set of pixel classification attributes. The implementation how the features are calculated is instrument specific. This approach has the advantage to be easily extendable to other instruments, and it is also applied in the GlobAlbedo project, where additionally AATSR is used.

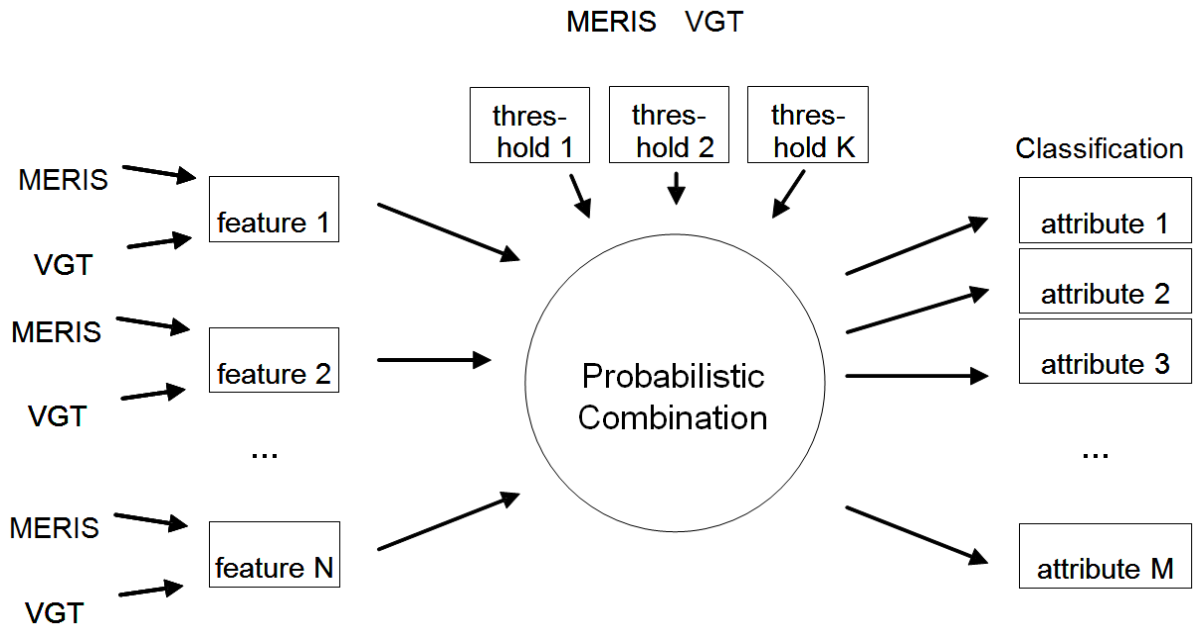


Figure 3-5: Unified Pixel Classification Scheme

3.3.5.2 Probabilistic arithmetic

A feature is a probabilistic quantity with a value between 0 and 1, with the following meaning:

Table 3-2: Definition of values for classification features.

VALUE	MEANING
0	the feature is definitively not true
0.5	status of the feature not known
1	feature is definitively true

Features are combined by simple arithmetic averaging. Let's assume, as an example, two features, f1 and f2, which do have no dependency from each other, and both being an indication that a third feature, f3, is true. Then, f3 is the average of f1 and f2: f3 = (f1 + f2)/2

As a first example let's assume the case where we are 100% sure that both f1 and f2 are true, i.e. f1=f2=1.0 then also f3=1.0. Another example is when we are 100% about f1, but less sure about f2, let's say f2= 0.8. Then the resulting probability of f3 is f3=0.9. When don't know anything about f1, i.e. f1=0.5 but are sure that f2 is false, i.e. f2=0, then f3=0.25, i.e. quite likely that it is false. If f1=0 and f2=1, i.e. a contradiction, then f3=0.5, i.e. we can't say if it is true or false.

The introduction of the probability scale [0 … 1] has further the advantage that it enable decoupling of feature values from the instruments. It doesn't matter how a physical quantity is derived because it will be mapped to the interval [0 … 1]. For example, the brightness feature will be calculated from top of atmosphere radiances in the case of MERIS, whereas it will be calculated from reflectances in the case of SPOT-VGT. These are different physical quantities, but they are both scaled to [0 … 1].

The scaling from a physical quantity, such as radiance or temperature, to a probability value may include a non-linear mapping. This can express the (un-)certainty that we have in value ranges in the physical data space. For example, very low temperatures have a very high probability to be a cloud, whereas above a certain temperature value the probability decreases exponentially.

Not every feature can be calculated for every instrument. In such cases the feature value is constant equal to 0.5. This convention allows formulating the logical combination of features even if a feature is not available for a certain instrument, and hence the logical combination can be formulated instrument independently.

3.3.5.3 Features

The following features are used in the probabilistic combination, and how they are calculated for each instrument.

Table 3-3: Feature - Pixel identification

FEATURE	EXPLANATION
Pressure	Indicating a high altitude from where the photons are scattered. Can be derived from measurements in gaseous absorption bands, e.g. O2A or water vapour
NDVI	A high vegetation index is an indication of a (semi-) transparent atmosphere
NDSI	The NDSI is a meaningful quantity only above bright surfaces. Then it can be used to separate snow/ice from clouds
White	A bright and spectrally flat signal; can be a cloud or snow/ice
Spectral Flatness	A spectrally flat signal; The colour can be anything from black over grey to white.
Temperature	Temperature of the emitting surface; clouds can be very cold1.
Bright	Brightness of the scattering surface
Glint Risk	The glint risk can be calculated from the observation geometry and wind speed, assuming a certain wave distribution (e.g. Cox and Munk). Glint and clouds are hardly separable and hence it is useful to identify glint risk in addition with the cloud/water classification2.
Radiometric Land Value	A classification of the surface type as land, provided that the pixel is clear and the measurement can be used to assess the surface type.
Radiometric Water Value	A classification of the surface type as water, provided that the pixel is clear and the measurement can be used to assess the surface type.
A priori Land Value	Classification of the pixel using a static background map and the geolocation of the pixel.
A priori Water Value	Classification of the pixel using a static background map and the geolocation of the pixel.

[1] Temperature information is not available from MERIS and VGT, but it is included as a feature in the logic because the overall approach is to be sensor independent; temperature is generally a very powerful criterion for cloud detection.

[2] CCI-LC requires atmospheric correction over water bodies; this implies cloud detection over water, which is problematic under sun glint conditions.

Feature Definition for MERIS

The index used in array notation below is starting with 0.

Table 3-4: Feature Definition for MERIS

FEATURE	MERIS	COMMENT

Pressure	if (isLand()) {
	press_value = (pbaro - p1)/1000.0;
	} else if (isWater()) {
	press_value =(pbaro - pscatt)/1000.0;
	}
		p1 and pscatt are apparent pressure products, defined in dedicated ATBDs from R. Santer, developed within the ESA O2 project. pbaro is the barometric pressure.

NDVI	value = brr[b753]-brr[b620])/ (brr[b753]+brr[b620]	brr is the reflectance corrected for gaseous absorption and Rayleigh scattering
NDSI	value = brr[b865]-brr[b885]) / (brr[b865]+brr[b885]	

White	if (brightValue()>BRIGHT_FOR_WHITE_THRESH) {
	white_value=spectralFlatnessValue();
	} else {
	white_value = 0;
	}	

Spectral Flatness	slope0 = spectralSlope(refl[490], refl[412]) slope1 = spectralSlope(refl[560], refl[620]) slope2 = spectralSlope(refl[665], refl[753]) spectralFlatness = 1.0f - Math.abs(1000.0 * (slope0 + slope1 + slope2) / 3.0);	
Temperature	0.5	
Bright	bright_value = brr[b442] /(6.0 * brr442Thresh)	brr442Thresh is a value read from a LUT; the LUT is a theoretical maximal reflectance for a given geometry and a bright land surface. It has been calculated by R. Santer and is available from the auxiliary data of the MERIS operational processor.

FEATURE	MERIS	COMMENT
Glint Risk	Shall be calculated from geometry and wind speed from tie points. Not available in current IDEPIX version	currently set to FALSE
Radiometric Land Value	if { refl[b753] >= refl[620] && refl[620] > refl620_Land_Thresh } radiom_land_value = 1.0 } else { radiom_land_value = 0.5 }	in case of cloudy pixel: radiom_land_value = 0.5
Radiometric Water Value	if { refl[b753] < refl[620] && refl[620] < refl620_Land_Thresh } radiom_water_value = 1.0 } else { value = 0.5 }	in case of cloudy pixel: radiom_water_value = 0.5
A priori Land Value	if (l1FlagLand) { return 1.0f; } else { return 0.0f; }	
A priori Water Value	if (!l1FlagLand) { return 1.0f; } else { return 0.0f; }	

Feature Definition for VGT

refl[0] … reflectance in band 0 at 450nm

refl[1] … reflectance in band 2 at 645nm

refl[2] … reflectance in band 3 at 835nm

refl[3] … reflectance in band MIR at 1.6nm

Table 3-5: Feature Definition for SPOT-VGT

FEATURE	VGT	COMMENT
Pressure	0.5	
NDVI	ndvi_value = (refl[2] - refl[1])/ (refl[2] + refl[1]	
NDSI	ndsi_value = refl[2] - refl[3])/ (refl[2] + refl[3])	

White	if (brightValue()>BRIGHT_FOR_WHITE_THRESH) {
	return spectralFlatnessValue();
	} else {
	return 0f;
	}
		The spectral flatness gives a high value even if the spectrum is black. In order to be an indicator for white, a minimum brightness is required.

Spectral Flatness	slope0 = spectralSlope(refl[0], refl[1]) slope1 = spectralSlope(refl[1], refl[2]) spectralFlatness = 1.0f - Math.abs(1000.0 * (slope0 + slope1));	
Temperature	0.5	

Bright	if (isLand()) {
	return (refl[0] + refl[1])/2.0f;
	} else if (isWater()) {
	return (refl[1] + refl[2]);
	} else {
	return (refl[0] + refl[1])/2.0f;
	}	

Glint Risk	isWater() && isCloud() && spectralSlope(refl[0], refl[1], 450, 645) > GLINT_THRESH)	

FEATURE	VGT	COMMENT

Radiometric Land Value	if (refl[2] > refl[1] && refl[2] > REFL835_LAND_THRESH) {
	radiom_land_value = 1.0;
	} else if (
	refl[2] > REFL835_LAND_THRESH) {
	radiom_land_value = 0.75f;
	} else {
	radiom_land_value = 0.25;
	}
		in case of cloudy pixel: radiom_land_value = 0.5

Radiometric Water Value	if (refl[0] > refl[1] && refl[1] > refl[2] && refl[2]<REFL835_WATER_THRESH) {
	value=1.0;
	} else {
	value = 0.25;
	}
		in case of cloudy pixel: radiom_water_value = 0.5

A priori Land Value	if (smLand) {
	return 1.0f;
	} else {
	return 0.0f;
	}	

A priori Water Value	if (!smLand) {
	return 1.0f;
	} else {
	return 0.0f;
	}	

3.3.5.4 Processing logic

The following classification attributes (binary values, also named flags) are derived from the features with given logic. This logic is no longer instrument dependent.

Table 3-6: Processing logic

CLASSIFICATION ATTRIBUTE	DEFINITION	COMMENT
isCloud	return (whiteValue() + brightValue() + pressureValue() + temperatureValue() > CLOUD_THRESH && !isClearSnow())	The final binary cloud flag. A pixel is either Cloud, ClearLand or ClearWater.or ClearSnow
isClearSnow	return (isBrightWhite() && ndsiValue() > NDSI_THRESH)	isBrightWhite is defined below

isClearLand	if (radiometricLandValue() != 0.5) {
	landValue = radiometricLandValue();
	} else if (aPrioriLandValue() > 0.5) {
	landValue = aPrioriLandValue();
	} else {
	return false;
	// this means: if we have no information about land, we return isClearLand = false
	}
	return (!isCloud() && landValue > LAND_THRESH)
		If a radiometric land value is available, i.e. it is not the uncertainty value of 0.5, than this is used in the subsequent test. Otherwsise the a priori land value is used in the test.
		The test simply compares the value with a threshold. The choice of the threshold depends on the user. If he wants to be really sure he should use a value close to 1.

isClearWater	if (radiometricWaterValue() !=0.5) {
	waterValue = radiometricWaterValue();
	} else if (aPrioriWaterValue() > 0.5) {
	waterValue = aPrioriWaterValue();
	} else {
	return false;
	// this means: if we have no information about water, we return isClearWater = false
	}
	return (!isCloud() && waterValue > WATER_THRESH);
		same logic as for the ClearLand test

isBrightWhite	return (whiteValue() + brightValue() > BRIGHTWHITE_THRESH)	A pixel that has one of the two characteristics, bright or white, has a potential to be cloudy. The stronger both features are the higher the probability.

CLASSIFICATION ATTRIBUTE	DEFINITION	COMMENT
isLand	return (aPrioriLandValue() > LAND_THRESH)	This is the surface type of the pixel, regardless if it is snow covered or if a cloud is above during measurement.
isWater	return (aPrioriWaterValue() > WATER_THRESH)	
isBright	return (brightValue() > BRIGHT_THRESH)	These tests map the real values of the features to binary flags
isWhite	return (whiteValue() > WHITE_THRESH)	
isCold	return (temperatureValue > TEMPERATURE_THRESH)	
isVegRisk	return (ndviValue() > NDVI_THRESH)	
isGlintRisk	return (glintRiskValue > GLINT_RISK_THRESH)	
isHigh	return (pressureValue() > PRESSURE_THRESH)	

3.3.5.5 Thresholds

The following table lists the nominal values for the scalar thresholds. It should be noted that these are subject to tuning and may change for the version used for product mass production.

Table 3-7: Thresholds

THRESHOLD	VALUE	
	MERIS	VGT
BRIGHTWHITE_THRESH	1.5	0.65
NDSI_THRESH	0.68	0.5
PRESSURE_THRESH	0.9	0.9
CLOUD_THRESH	1.65	1.65
UNCERTAINTY_VALUE	0.5	0.5
LAND_THRESH	0.9	0.9
WATER_THRESH	0.9	0.9
BRIGHT_THRESH	0.25	0.3
WHITE_THRESH	0.9	0.5

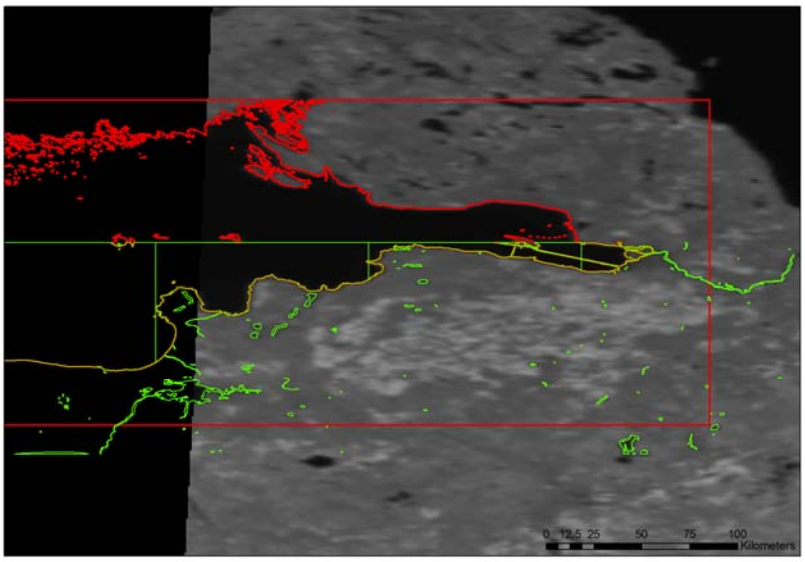
THRESHOLD	VALUE	
	MERIS	VGT
BRIGHT_FOR_WHITE_THRESH	0.4	0.2
NDVI_THRESH	0.7	0.4
TEMPERATURE_THRESH	0.9	0.9
REFL835_LAND_THRESH	n.a	0.15
REFL835_WATER_THRESH	n.a	0.1
GLINT_THRESH	n.a	0.000365

This is chosen based on the logic that it should be BRIGHTWHITE_THRESH + 2*0.5, when the pressureValue and temperatureValue = 0.5, i.e. uncertain.

3.3.5.6 Water/land mask: SWBD mask

Both classes, the a priori Land Value and the a priori Water Value are determined by a static background map as well as the geo-location of the corresponding pixel.

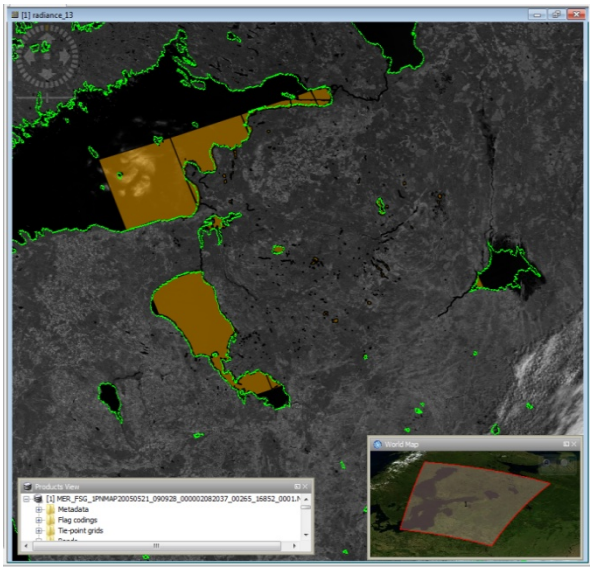
For refinement of the classification a new static map has been developed and the implementation is currently ongoing. ISCIENCES Water land mask mentioned in the proposal is improper because it only takes the coastline into account see Figure 3-6 . Therefore SRTM SWBD map (NGA) is used as the best alternative from http://dds.cr.usgs.gov/srtm/version2_1/SWBD/ or http://eros.usgs.gov/#/Find_Data/Products_and_Data_Available/Shuttle_Radar_Topography_Mission_Water_Body_Dataset.

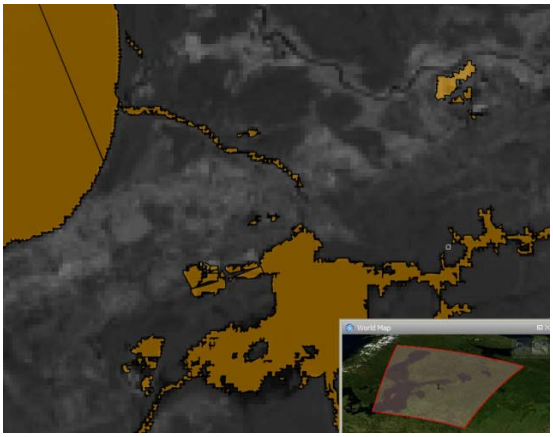


SPOT VGT - V2KRNP____20050521F196.ZIP

Figure 3-6: Comparison of SRTM SWBD and ISCIENCES map

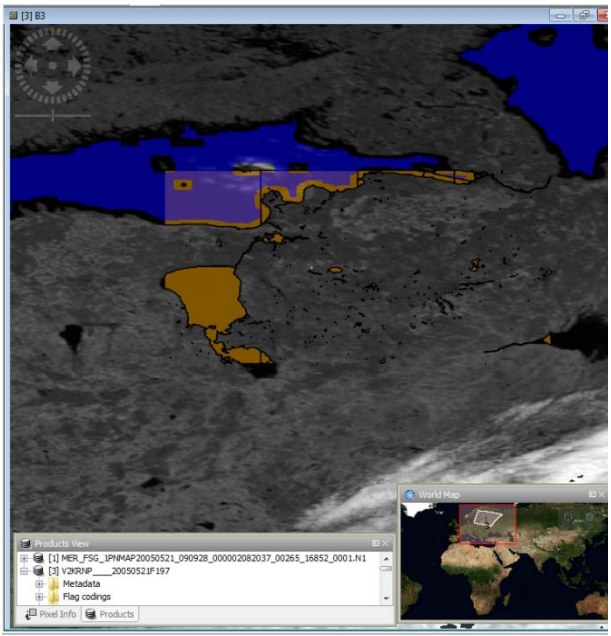
Figure 3-7 shows the origin land/water mask (green) and the application of the new SRTM SWBD map (brown). The improvement is clearly visible.

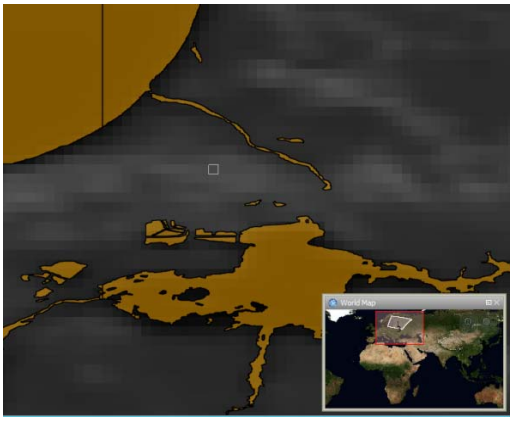




MERIS

MER_FSG_1PNMAP20050521_090928_000002082037_00265_16852_0001.N1





SPOT-VGT

V2KRNP____20050521F197.ZIP

Figure 3-7: Origin land/water mask and SRTM SWBD map

3.3.6 Adaptation of IdePix for MERIS FRS and RR

The assessment of MERIS FR and RR SR products using the former IdePix shows that the data quality has not yet reached the expected level, mainly due to clouds. A better pixel classification w.r.t. the class “clear land” is required. Therefore new pixel identification features for the CCI-LC processing has been developed and applied.

Temporal Features

Further improvements of pixel identification are accessible by analysing their temporal characteristic. Based on this approach a temporal filter has also been applied and is described in detail in the section 3.5.4.

Spectral Features

The following new spectral features have been considered, investigated und applied:

- blue band feature;
- cloud value feature;
- clear land feature.

The **blue band feature** based on the Blue Band Cloud Screening of the GlobCover Project and the entire description may be found in the GlobCover Design Justification File [RD-2]. A simple blue band test has been adopted for MERIS by using the 412 nm channel. The developed cloud screening method is applied to reflectances. A first threshold for band 1 reflectances is used to detect the most brilliant dense clouds. The clear pixels are tested by a filter which performs the reflectance ratio of band11 and band10 related to the altitude of the scattering surface. An optimised threshold permits identifying thin clouds which are not detected by the first blue band test. Three states are supposed: 0=Out of Orbit, 1=Clear and 2=Cloud. The optimised cloud mask is globally coherent with the bright flag of standard MERIS products. However, a higher performance of the cloud screening over semi-transparent clouds can be observed. Large areas are not detected by the bright flag or probability algorithms. The analysis shows that the blue band algorithm is more able to detect semi transparent clouds.

The **cloud value feature** is predicted by two neuronal network models or to put it more succinctly, they are back-propagation neuronal networks with one input and one output layer. The NNs have been trained with a back-propagation learning algorithm by using the PixBox data whereas the training dataset was splitted w.r.t. the surface properties land and water and these two datasets have been used separately for the training. Therefore we get a so called LandNN and a WaterNN for the prediction of the cloud value. The PixBox data are described in the PVASR [AD-8]. The decision which NN will be used is only depending on the classified surface type through IdePix. Afterwards the cloud value is calculated through the NN by using the MERIS FSG respectively and RRG TOA reflectance spectrum as input. The cloud value has a range from 0 through 2. The user defined thresholds - 1.25 for the LandNN and 1.35 for the WaterNN are used for the verification of the current pixel status as clear land or clear water. If the cloud value is greater than the particular threshold then the pixel status is set to cloud. The PixBox data based only on those of the MERIS, therefore the cloud value feature is not applicable to SPOT VGT P data.

The **clear land feature** of pixel identification bases on the analysis of the spectral characteristic in the HSV (Hue Saturation Value) - colour space. The clear land feature is described in detail in the section 3.5.4.

3.3.7 Error budget estimates

The adaption of IdePix by including the new features has influenced the error estimation. Now the error cannot simply be calculated automatically by the probabilistic arithmetic. Therefore as a first guess we will introduce a measure of certainty that the inspected pixel is covered by a cloud. This measure of certainty for each pixel is calculated by counting the number of the true cloud features. So the measure of certainty for a pixel is very high, if all cloud feature has been classified this pixel as cloudy.

3.3.8 Assumptions and limitations

None

3.3.9 Improvements

Spatial Features

Further improvements of pixel identification are accessible if features which are not only calculated by using the radiometric characteristics of each single pixel but analysing their individual neighbourhood.

Spectral Features

The definition of this cloud feature based on a training sample. The elements of this training set are built on basis of visually detected clouds in the data of MERIS L1b products. With the support of training data, the following condition and definition for the cloud feature could be defined (eq. 3-6):

```
if {
   rad_1  ϵ [110.954698614776; 489.892221887595]   UND  rad_1  ∉ [109.012356602586; 120.686167307198]   UND
   rad_2  ϵ [110.697029320523; 544.088897639885]   UND  rad_2  ∉ [111.778583999723; 126.789050581865]   UND
   rad_3  ϵ [101.590829581022; 573.706123851239]   UND  rad_3  ∉ [105.853863149881; 126.284532736986    UND
   rad_4  ϵ [94.6231306791305; 569.974756628274]   UND  rad_4  ∉ [99.397863805294; 121.631598353385]    UND
   rad_5  ϵ [79.5122078303247; 520.784201537258]   UND  rad_5  ∉ [85.8748179972171; 109.901390730403]   UND
   rad_6  ϵ [68.6891217874363; 480.18538188748]    UND  rad_6  ∉ [76.6152256401255; 100.739536187611]   UND
   rad_7  ϵ [65.6249167695641; 449.788562081754]   UND  rad_7  ∉ [73.8768465369939; 99.3999824300408]   UND
   rad_8  ϵ [63.8769699875265; 455.679951677098]   UND  rad_8  ∉ [72.1886596512049; 97.8748938012868]   UND
   rad_9  ϵ [58.3485030834563; 414.458504260051]   UND  rad_9  ∉ [66.2284095734357; 90.2341051995754]   UND
   rad_10 ϵ [55.2074812687933; 401.418734680861]   UND  rad_10 ∉ [63.430219009518600; 87.5943529829382]  UND
   rad_12 ϵ [.7855027711484; 239.06330769439251]   UND  rad_12 ∉ [59.610151657601800; 83.0062435523141]  UND
   rad_13 ϵ [41.9968677135184; 233.933762553147]   UND  rad_13 ∉ [48.664812148548600; 68.5467036478221]  UND
   rad_14 ϵ [40.4503375631757; 295.254060312639]   UND  rad_14 ∉ [47.011832612566600; 66.3158683939836]
} then cloud_feature=1 (true)
else cloud_feature=0 (false)
```

with rad_x is the MERIS L1b radiance of band x (x=1-14 without 11).

3.3.10 Cloud shadow and cloud edge detection

A more comprehensive representation of clouds is provided by two additional pixel properties, 'cloud edge' and 'cloud shadow'. These two properties are also kept in the pixel identification flag and can be considered as a kind of "postprocessing" in the cloud detection.

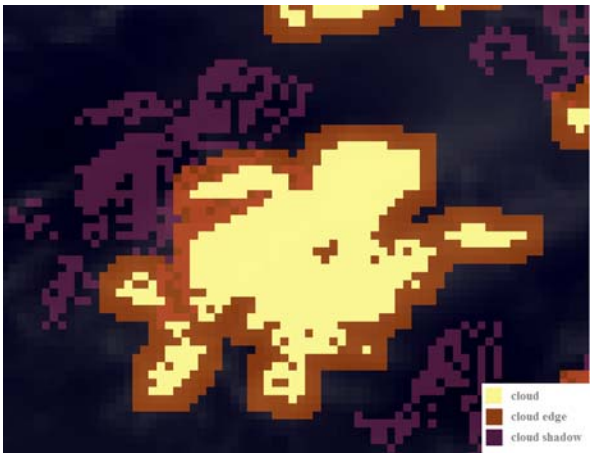


Figure 3-8: Cloud, cloud edges and cloud shadow

To detect dark cloud shadow pixels -whose spectra are polluted by the shadow- the position of the shadow will be determined by projecting the 'cloud pixels' onto the ground using the sun position, the pixel's altitude on Earth, and the cloud height estimated from cloud top pressure.

Cloud edge pixels are in principle regarded as neighbour pixels of a 'cloud' as identified before in the pixel classification. The width of this edge (in number of pixels) can be set by the user. In brief, the algorithm to identify cloud edge pixels works as follows:

- use a 2x2 square with reference pixel in upper left
- move this square row-by-row over the given tile
- if reference pixel was not cloud, don't do anything
- if reference pixel is cloudy:
- if 2x2 square only has cloud pixels, then set cloud buffer of two pixels in both x and y direction of reference pixel;
- if 2x2 square also has non-cloudy pixels, do the same but with cloud buffer of only 1.

3.4 Atmospheric correction - aerosol and spectral directional reflectance retrieval

The atmospheric correction includes the correction for the absorbing and scattering effects of atmospheric gases, in particular ozone, oxygen and water vapour, of the scattering of air molecules (Rayleigh scattering) and the correction of absorption and scattering due to aerosol particles. All components except aerosols can be rather easily corrected. Ozone can be taken from other satellites, available from met services, and oxygen and water vapour can be taken from MERIS measurements thanks to dedicated spectral bands. However, aerosols are spatially and temporally highly variable and do not have a distinct spectral absorption features, so that they cannot be measured easily from MERIS and SPOT-VGT measurements. Techniques exists to derive aerosol information from measurements where the surface signal is low (LARS technique, developed by R. Santer and implemented in the MERIS ground segment), but the areas where the surface is dark enough are sparse and spatial-temporal interpolation is necessary. Also, these derived aerosols optical properties are connected with a substantial error.

The aerosol correction is the largest contributor the error of the atmospheric correction. The challenge for CCI-LC is to minimize this error and to quantify it reliably. The method should ideally work across different instruments.

The aerosol correction typically consists of two parts: the calculation of these aerosol properties and secondly the actual correction of the reflectance (after Rayleigh and gas absorption correction). This can be a sequential process where the reflectance correction takes an aerosol optical depth and its spectral dependency as input, or a one step approach where aerosol properties and reflectance corrections are one implicit step. In the current GlobCover processing the aerosol property retrieval is done offline using the aerosol retrieved from the corresponding MERIS L2 product, as well as monthly and yearly aerosol climatology derived from MERIS, and finally climatological background values. This gives a gap free product, which is important for the correction step. Although it has been demonstrated that this approach is by far advantageous compared to any other source for aerosols available at the time of development, the results were far from being perfect. It seems obvious that the quality needs to be improved to meet GCOS requirements.

A new approach for retrieval of aerosol properties for MERIS and AATSR has been developed throughout recent years by the University Swansea for the GlobAlbedo project. It is based on an inverse modelling approach, working on a larger area of 8 x 8km to estimate the aerosol properties, which is then used to correct each single pixel. This method provides significantly more aerosol retrievals than the LARS method, which is the baseline for the GlobCover aerosols. Also the quality of the retrieved aerosols has been proven to be good. This method has proven its strength with AATSR data and also with MERIS and AATSR in synergistic mode. It has further been applied successfully to Chris/Proba, and it works also well with MERIS data alone.

This new atmospheric correction is fully documented and prototyped, but still in a stage of validation and fine tuning. The involved researchers and developers of the GlobAlbedo project prefer reviewing it with respect to the terms of robustness of the retrievals.

3.4.1 Aerosol retrieval

The aerosol retrieval developed under the GlobAlbedo project for Surface Directional Reflectance (SDR) estimation; this is fully discussed in [RD-53, RD-54] and here we give a brief summary of points relevant to aerosol estimation. The iterative retrieval of atmospheric aerosol requires a fast approximation of atmospheric radiative transfer, to relate TOA to surface reflectance for varying aerosol loading.

To retrieve estimates of aerosol properties from measured satellite radiances, we need to solve the inverse problem to separate the atmospheric and surface scattering contributions to the observed signal. This normally requires some assumptions to be made on the land surface brightness. Within the proposed framework, these assumptions are expressed as constraints defined by error of fit to a parameterized model describing the surface angular or spectral reflectance. For the single view instruments we apply constraints based on the dark object method (DOM).

The aerosol retrieval includes the calculation of the aerosol optical depth (AOD). Estimates of aerosol extinction are needed for the conversion from top-of-atmosphere measurements to surface reflectance. Aerosol optical depth and aerosol model plus an estimate of the uncertainty in AOD are derived by the Aerosol Retrieval processor from every data set to be processed. It must be noted that AOD and aerosol model are assumed to sufficiently account for the variability in the atmospheric conditions to calculate these terms, while water vapour and ozone column contents are needed in addition to aerosol parameters in order to retrieve the most accurate Lambertian equivalent reflectance. All other atmospheric constituents are just set to climatology values in the algorithm. This selection is justified by the relatively higher impact of aerosol extinction in the spectral channels of MERIS and SPOT-VGT instruments, particularly in the visible.

The problem of surface reflectance and aerosol retrieval can essentially be formulated as one of multivariate optimisation subject to multiple constraints:

- Given a set of satellite TOA reflectance and an initial guess of atmospheric profile, the corresponding set of surface reflectance is estimated
- Application of the observed set to the estimated set of reflectance results in an error metric, where a lower value of the metric corresponds to a set of surface reflectance (and hence atmospheric profile) that is more realistic

These two steps are repeated with refined atmospheric profiles until convergence at an optimal solution. Given a certain aerosol model, the logical flow of the pixel-wise retrieval of aerosol properties is illustrated in Figure 3-10. It consists of the following steps:

- Atmospheric correction of averaged TOA radiance using an estimate of aerosol optical thickness. The resulting estimate of surface reflectance for each input channel is calculated by means of lookup tables. Besides the aerosol optical depth the atmospheric correction depends on the aerosol model, viewing and illumination geometry, surface pressure, and ozone content
- Error metric calculation
- Optimisation of the estimated aerosol optical thickness by means of minimisation of the error metric as implicit function of aerosol optical thickness.

- The implicit dependency of aerosol optical thickness is given by the surface reflectance estimated in Step 1

The algorithm components are therefore (i) design of an efficient and accurate scheme for deriving surface reflectance for known atmospheric profile, and (ii) formulation of constraints on the land surface reflectance suitable spectral sampling of the instruments used (SPOT-VGT and MERIS). The method is applied to estimate aerosol at a more coarse spatial resolution (8x8 km) than the underlying surface reflectance, and a subsequent interpolation step is used to obtain per-pixel values.

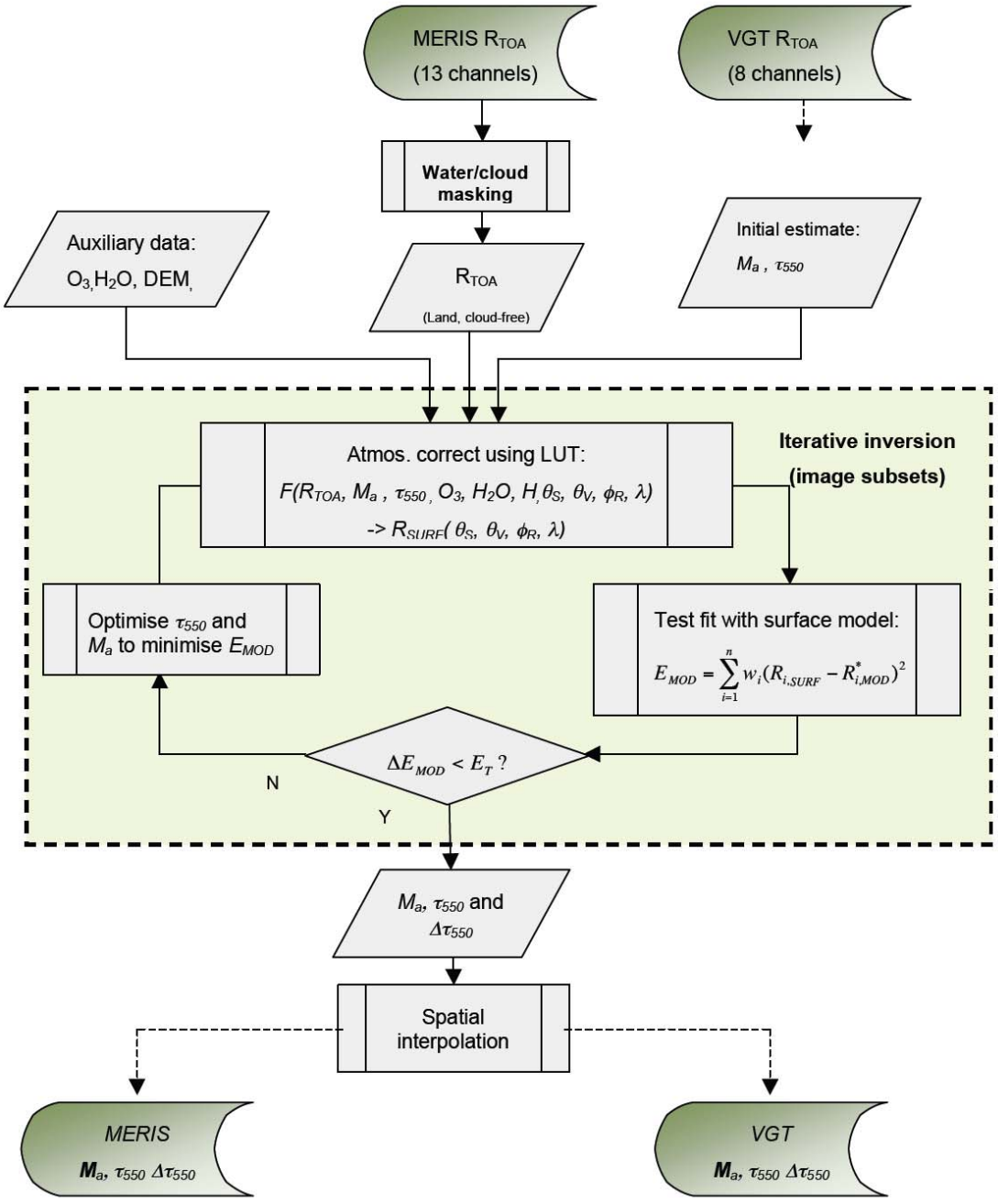


Figure 3-9: Scheme of process, graphic adopted from [RD-53]

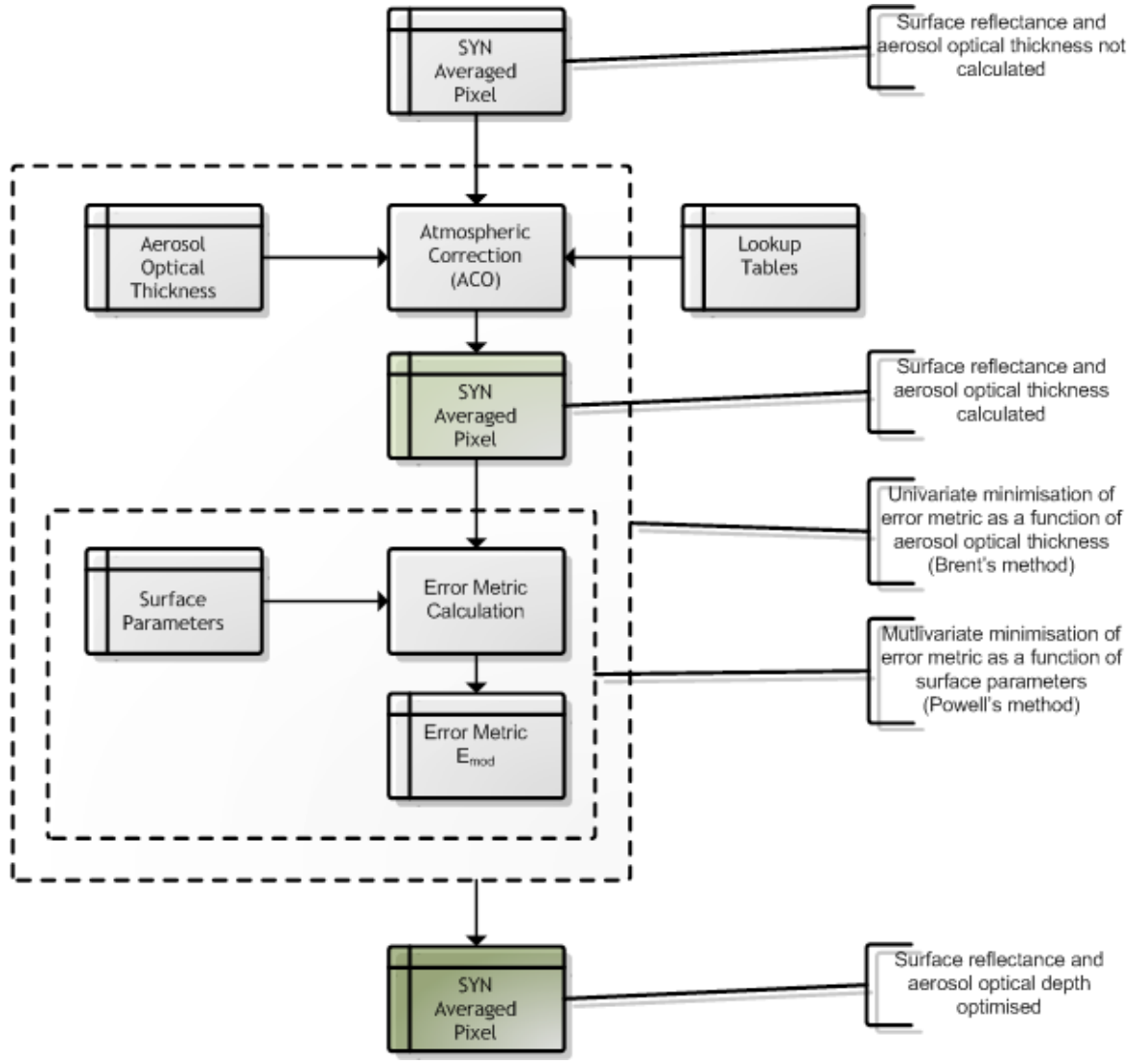


Figure 3-10: Logical flow of the aerosol retrieval and the atmospheric correction scheme. Figure from [RD-55]

3.4.2 Atmospheric correction

The purpose of this paragraph is to provide an overview of the theoretical background of the SDR retrieval, developed by the Institute for Space Sciences, Free University of Berlin, fully documented in [RD-2] and an improved version documented in [RD-54].

The atmospheric correction scheme is designed for the MERIS Level2 data above land surfaces and is based on inverse modelling of radiative transfer simulations by using artificial neural network (ANN) techniques. The inverse model consists of a Multi-Layer-Perceptron with one hidden layer trained by the back-propagation algorithm [RD-59]. Figure 3-11 illustrates the flowchart of the aerosol correction scheme, which is divided into two main parts: development and operation of the algorithm. The development of the algorithm is sub-divided further into the sections for forward and inverse modelling.

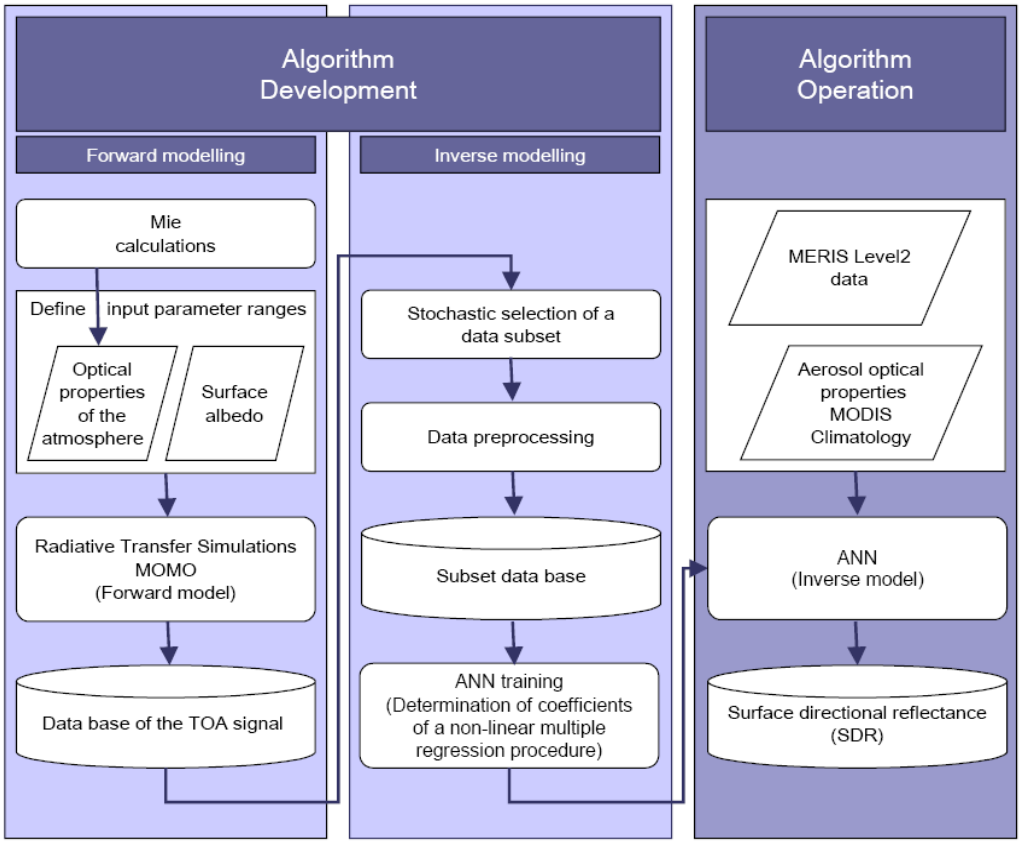


Figure 3-11: Flowchart of the proposed neural network based correction algorithm for the retrieval of the surface directional reflectance (from [RD-56])

As forward model we used the established radiative transfer code MOMO based on the Matrix-Operator-Method [RD-57, RD-58], to simulate a database, describing the functional relationship between the surface reflectance and the MERIS Level2 signal at TOA depending on the atmospheric properties and the surface reflectance. The simulated data is then used to adapt the inverse model. The inverse model consists of a Multi-Layer-Perceptron with one hidden layer trained by the backpropagation algorithm [RD-59] on a stochastic selected and preprocessed subset of the complete database. The proposed aerosol correction scheme derives the SDR at the Bottom-Of-Atmosphere (BOA) from the cloud screened TOA MERIS Level2 reflectance subsequently for each MERIS band on a pixel-by-pixel basis. To correct the influence of the aerosols the algorithm requires additional input information like aerosol optical thickness (AOT) at two wavelengths and the associated Angstroem coefficient. In the operational mode these data will be taken from MODIS measurements or from an external database or from the aerosol retrival. Moreover, the algorithm gives an error estimate of the derived reflectances when introducing the typical uncertainty of the MODIS aerosol retrieval to the ANN input.

The improved version of the SDR retrieval, fully documented in [RD-54], should fulfil some requirements

- It must be capable of deriving SDR products from top-of-atmosphere radiance data measured by MERIS and SPOT-VGT;

- SDR products must include a reliable tracking of uncertainties within the entire processing including error covariances between the different spectral ranges.

The algorithm is designed to produce SDR (including uncertainties) from TOA radiance data. The algorithm requires inputs from the pixel classification and aerosol retrieval.

The apparent spectral directional reflectance R* λ in the waveband λ of the coupled surface-atmosphere system is related to the radiance L_λ measured by a satellite at the TOA by (eq. 3-7):

$$R^*_\lambda(\Omega_v, \Omega_s) = \frac{\pi L_\lambda(\Omega_v, \Omega_s)}{\mu_s E_{s,\lambda}}$$

where Es,λ is the extraterrestrial irradiance at the time of the measurement. The view and solar vectors are denoted by Ω_s and Ω_v respectively, while μ_s and μ_v denotes cosines of solar and view zenith.

A uniform, Lambertian surface is normally assumed as a basis for the modelling of the atmosphere-surface radiative transfer for operational atmospheric correction algorithms of single-view instruments. Under that assumption, the relationship between top of atmosphere reflectance R* and the surface directional reflectance R λ can be approximated by the equation (eq. 3-8):

$$R^*_\lambda(\Omega_v, \Omega_s) = R_{atm,\lambda}(\Omega_v, \Omega_s) + \gamma_\lambda(-\mu_s)\gamma_\lambda(\mu_v)\frac{R_\lambda}{1 - \bar{\bar{\rho}}_\lambda R_\lambda}$$

Where R_{atm} means the atmospheric scattering term (TOA reflectance for zero surface reflectance), γ_λ denotes atmospheric transmission for either sensor to ground or ground to sensor for waveband λ, and s denotes atmospheric bi-hemispherical albedo with respect to the surface. The view and solar vectors are denoted by Ω_s and Ω_v respectively, while μ_s and μ_v denotes cosines of solar and view zenith. Spectral directional reflectance R is derived from TOA reflectance by means of the analytical inversion of eq. 3-8, and the coresponding equaation is (eq. 3-9):

$$R_\lambda(\Omega_v, \Omega_s) = \frac{R^*_\lambda(\Omega_v, \Omega_s) - R_{atm,\lambda}(\Omega_v, \Omega_s)}{\gamma_\lambda(-\mu_s)\gamma_\lambda(\mu_v) + \bar{\bar{\rho}}_\lambda[R^*_\lambda(\Omega_v, \Omega_s) - R_{atm,\lambda}(\Omega_v, \Omega_s)]}$$

As discussed in [RD-60], the Lambertian equivalent reflectance R_λ derived from eq. 3-9, which is taken as SDR in the GlobAlbedo processing chain, represents a smoothed version of the surface BRDF, with errors up to 15% for turbid atmospheres ([RD-61] - see Figure 3-12).

R* is calculated from TOA radiance with Eq. 3-7 and first corrected from gaseous absorption with a look-up table for a set of input AMF, CWV and OZO. AMF and CWV are considered pixel-wise, while a mean value representative of the imaged area is used for OZO. The atmospheric parameters $R_{atm,\lambda}$, γ_λ and $\bar{\bar{\rho}}_\lambda$ are provided by a look-up table per pixel, being the input VZA, SZA, RAA, ELEV and AOD550 also provided on a per-pixel basis. No correction of adjacency effects [RD-62, RD-63] has been found to be necessary.

The look-up tables intended for aerosol and SDR retrieval and for error propagation have been compiled with the MOMO radiative transfer code [RD-57, RD-58]. MOMO is a well introduced and

widely accepted radiative transfer code which provides all the features required for atmospheric radiative transfer simulations in CCI-LC project. The look-up tables intended for gas correction and also for error propagation for are calculated with the 6S (Second Simulation of the Satellite Signal in the Solar Spectrum) radiative transfer code [RD-64, RD-65]. This code is intended for fast radiative transfer calculations at a relatively coarse internal spectral sampling of 2.5 nm.

Tracking of uncertainties is one of the specific requirements for the CCI-LC processing. The purpose in this part of the processing is to generate a reliable estimation of the errors in the SDR products including the error covariance between the different spectral ranges. Different error sources are considered: instrumental noise, atmosphere and directional reflectances effects.

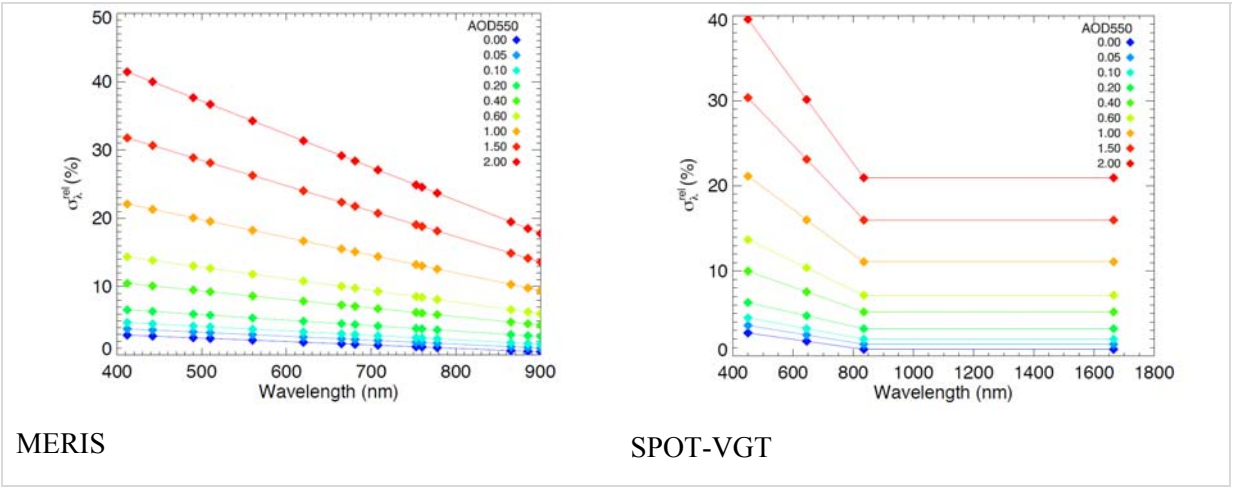


Figure 3-12: Estimation of the relative error in reflectance retrieval caused by the assumption of a Lambertian surface [RD-61] after resampling to the GlobAlbedo AOD550 grid and the spectral response (from [RD-54])

3.4.3 Atmospheric correction of SPOT-VGT S1 products

The atmospheric corrections are based on the use of the SMAC software developed by the Center for the Study of the Biosphere from Space (CESBIO) [RD-66], which is a simplified implementation of the 6S method [RD-64, RD-65, RD-67]. The SMAC software has been tuned for the SPOT-VGT spectral bands. The physical data used as input of SMAC are the following:

- The water vapour data are global short term forecasts, downloaded 4 times/day from Meteo-France. The elementary data cell corresponds to 1.5° x 1.5° in longitude and latitude. A geometrical interpolation is performed to obtain 8/112° x 8/112° cells, and a temporal interpolation is performed between the two nearest measures, taking into account the viewing time.
- A climatology of ozone obtained from the CESBIO
- A simple static models for aerosols obtained from the CESBIO
- A 8/112° resolution DEM for atmospheric pressure estimation

The atmospheric parameters (optical thickness) are calculated on 8km x 8km cells, then interpolated for every pixel, and used for the final reflectance calculation [RD-68].

3.5 L3-Processing: projection and compositing

The processing steps to retrieve the final Level 3 products from an input set of single satellite observations (i.e., SDR and pixel classification data as derived in the previous sections) are (i) the reprojection of the input products onto a Plate Carrée grid, (ii) the aggregation of the single satellite observations for given binning cells (tiles), and (iii) the mosaicing of the binning cells to a (usually global) Level 3 product with the aggregation results. This processing flow is illustrated in Figure 3-13. The Level 3 product will in return serve as input for the classification algorithms to derive a final landcover map [AD-8].

3.5.1 Plate-Carrée projection

The plate carrée projection or geographic projection or equirectangular projection, is a very simple map projection that has been in use since the earliest days of spherical cartography. The name is from the French for "flat and square". It is a special case of the equidistant cylindrical projection in which the horizontal coordinate is the longitude and the vertical coordinate is the latitude.

The spherical earth can only be mapped onto a developable surface by allowing distortion, so certain geometric properties on the sphere are not preserved. The Platte Carrée projection is a cylindrical projection but unlike the Mercator projection, the entire sphere, including the poles can be represented on a finite sized map. The projection is not a conformal map so angles are not preserved.

Because of the distortions introduced by this projection, it has little use in navigation or cadastral mapping and finds its main use in thematic mapping. It has also become a de-facto standard for computer applications that process global maps, such as Celestia, because a given co-ordinate is very easily identifiable in an image file.

The following equations describe the mapping of geographic coordinates in terms of latitude φ and longitude λ onto the *x* and *y* coordinates of a point on the map. From its latitude φ and longitude λ (with φ_0 and λ_0 being the latitude and longitude in the centre of map) and *k* being an appropriate scale factor at the equator (eq. 3-10):

$$x = k(\lambda - \lambda_0)$$

$$y = k(\varphi - \varphi_0)$$

For the CCI-LC products, the geoid WGS84 is used.

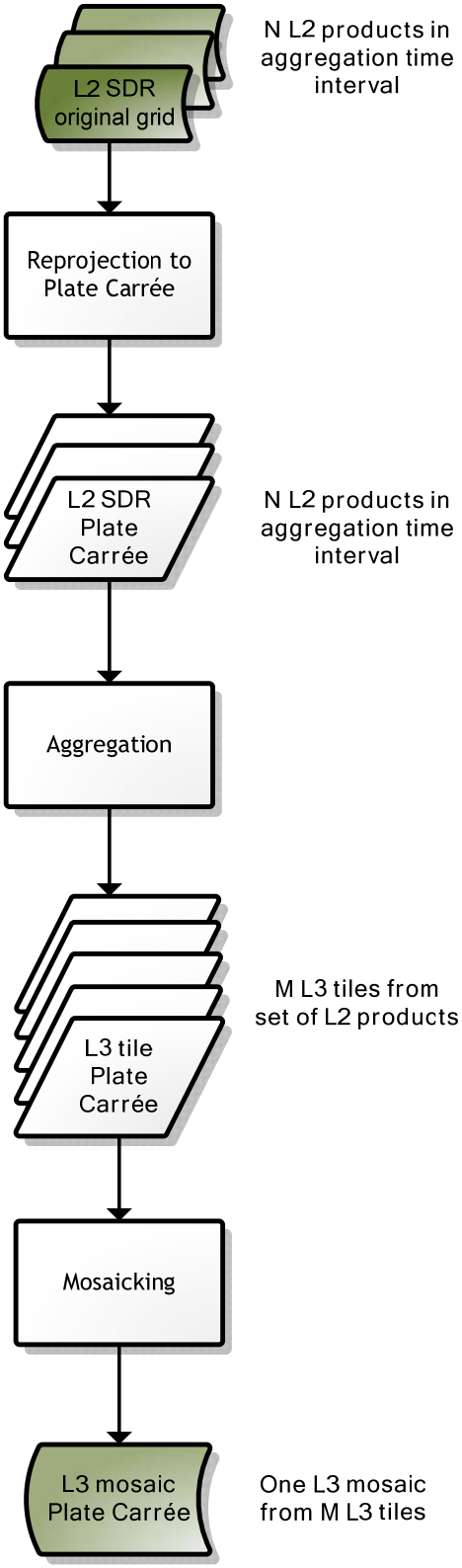


Figure 3-13: Overall logical flow of the CCI-LC L3 processing as final step of the pre-processing part (from [AD-8])

3.5.2 Resampling

If a product is projected it comes up that the pixel centres of the target product generally not correspond to the centres of the pixels of the input product. Resampling entitles the process of determination and interpolation of pixels in the source product for computation of the pixel values in the target product. The effects of resampling will especially be visible if the pixels in the target product are larger than the source pixels. Three different resampling methods for this computation are described here.

3.5.2.1 Nearest neighbour

Every pixel value in the output product is set to the nearest input pixel value.

Table 3-8: Advantages and disadvantages of the nearest neighbour

PROS	CONS
Very simple, fast	Some pixels get lost and others are duplicated
No new values are calculated by interpolation	Loss of sharpness
Fast, compared to Cubic Convolution resampling	

Figure 3-14 demonstrates the calculation of the new pixel value.

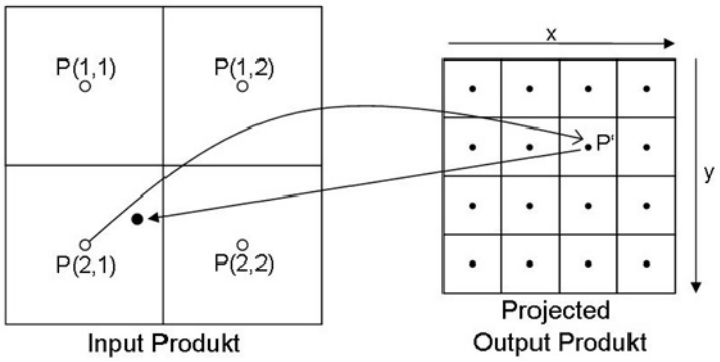


Figure 3-14: Nearest Neighbour

3.5.2.2 Bi-linear interpolation

Calculation of the new pixel value is performed by the weight of the four surrounding pixels.

Table 3-9: Advantages and disadvantages of the bilinear interpolation

PROS	CONS
Extremas are balanced	Less contrast compared to Nearest Neighbour
Image losses sharpness compared to Nearest Neighbour	New values are calculated which are not present in the input product

Figure 3-15 demonstrates the calculation of the new pixel value.

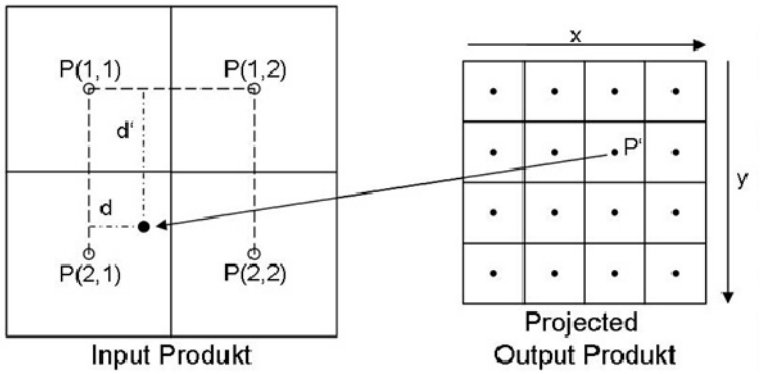


Figure 3-15: Bilinear Interpolation

The bilinear interpolation is performed by the following equation (eq. 3-11):

$$P'(x,y) = P(1,1)\cdot(1-d)\cdot(1-d') + P(1,2)\cdot d\cdot(1-d') + P(2,1)\cdot d'\cdot(1-d) + P(2,2)\cdot d\cdot d'$$

3.5.2.3 Cubic convolution

Calculation of the new pixel value is performed by weighting the 16 surrounding pixels.

Table 3-10: Advantages and disadvantages of the cubic convolution

PROS	CONS
Extremas are balanced	Less contrast compared to Nearest Neighbour
Image is sharper compared to Bi-linear Interpolation	New values are calculated which are not present in the input product
	Slow, compared to Nearest Neighbour resampling

Figure 3-16 demonstrates the calculation of the new pixel value.

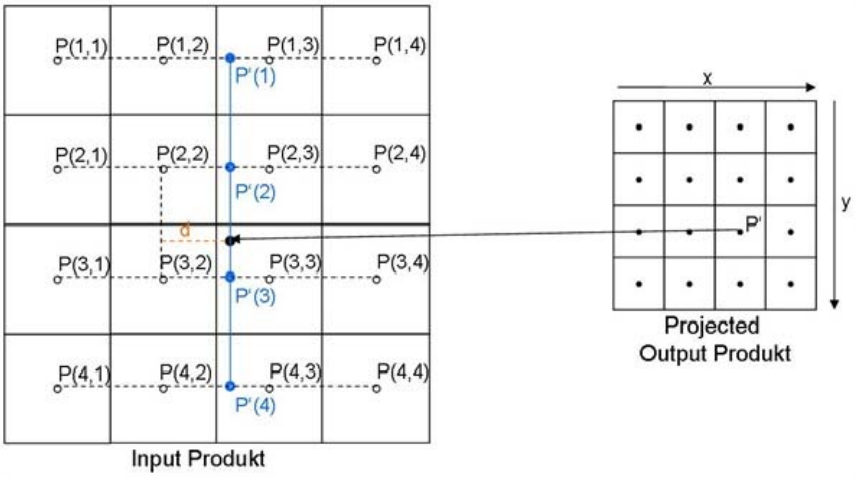


Figure 3-16: Cubic Convolution

The cubic interpolation is performed by the following equation (eq. 3-12):

$$P'(k) = P(k,1)\cdot\left(4-8(1+d)+5(1-d^2)-(1+d)^3\right)+ P(k,2)\cdot(1-2d^2+d^3)+ P(k,3)\cdot\left(1-2(1-d)^2+(1-d)^3\right)+ P(k,4)\cdot\left(4-8(2-d)+5(2-d)^2-(2-d)^3\right)$$

In the first step the average value for each line is calculated, afterwards the new pixel value is calculated with the four new average values P'(1) - P'(4) similar to the preceding calculation.

3.5.2.4 Visual comparison of the resampling methods

Nearest Neighbour | Bi-linear Interpolation | Cubic Convolution

Figure 3-17: Visual comparison

Nearest Neighbour

Each output cell value in the nearest neighbour method is the unmodified value from the closest input cell. Less computation is involved than in other methods, leading to a speed advantage for large input raster files. Preservation of the original cell values can also be an advantage if the resampled raster will be used in later quantitative analysis, such as automatic classification. However, nearest neighbour resampling can cause feature edges to be offset by distances up to half of the input cell size. If the raster is resampled to a different cell size, a blocky appearance can result from the duplication (smaller output cell size) or dropping (large cell size) of input cell values.

Bilinear Interpolation

An output cell value in the bilinear interpolation method is the weighted average of the four closest input cell values, with weighting factors determined by the linear distance between output and input cells. This method produces a smoother appearance than the nearest neighbour approach, but it can diminish the contrast and sharpness of feature edges. It works best when resampling to a smaller output cell size.

Cubic Convolution

The cubic convolution method calculates an output cell value from a 4 X 4 block of surrounding input cells. The output value is a distance-weighted average, but the weight values vary as a nonlinear

function of distance. This method produces sharper, less blurry images than bilinear interpolation, but it is the most computationally intensive resampling method. It is the preferred method when resampling to a larger output cell size.

However, the nearest neighbour method has been selected as resampling method for the projection caused by its preservation of the original physical values.

3.5.3 Temporal sample aggregation

Basically, the temporal sample aggregation is, for a given pixel in a given aggregation grid cell (a spatial tile), an integration over all input data, providing aggregated values of SDR and SDR uncertainties, the number of counts from all observations for each possible status, and a final status flag derived from a scheme which analyses these number of counts. Finally, SDR and SDR uncertainties are normalized by the number of occurrences of the "final status" for the given pixel. Table 3-11 summarizes the types of final status flag [AD-8].

Table 3-11: Final status flag

FLAG NAME	DESCRIPTION
CURRENT_PIXEL_STATE	final pixel classification after aggregation (one of the following flags)
CLEAR_LAND	pixel was classified as land
CLEAR_WATER	pixel was classified as water
CLEAR_SNOW_ICE	pixel was classified as cloudy
CLOUD	pixel was classified as cloud or pixel was classified as 'temporal' cloud (a posteriori cloud check within the aggregation scheme)
CLOUD_SHADOW	pixel was classified as cloud shadow
INVALID	pixel was classified as invalid or cosmetic

The temporal sample aggregation sums up all measurements to the intermediate variables and saved different status counts for a pixel [RD-94]. The 7-day compositing period is adapted on a fixed basis: every January 1st as starting point and a continuous 7 day intervals for each year has been proposed. The last composite of each year and the last February composite of a leap year include 8 days instead 7days.

Figure 3-18 and the following pseudo-code (eq. 3-13) describe and illustrate the temporal sample aggregation [AD-8].

pixel classification status **after** aggregation (X=LAND, CLOUD etc.)

STATUS_X

surface reflectances for N_λ wavelengths of given sensor

for j=1 ... N_λ and $N_{obs} := N_{obs,STATUS\_X}$

$$S_{agg(j,i)} := \overline{SDR_j}(i) = \frac{1}{N_{obs}(i)} \sum_{m=1}^{N_{obs}(i)} SDR_j(i,m)$$

NDVI

for $Nobs := Nobs_{,STATUS\_X}$

$$S_{agg(N\lambda+1,i)} := \overline{NDVI}(i) = \frac{1}{N_{obs}(i)} \sum_{m=1}^{Nobs(i)} NDVI(i,m)$$

error in surface reflectances for N_λ wavelengths of given sensor

for j=N_λ +2 ... 2*N_λ +2 and $N_{obs} := N_{obs,STATUS\_X}$

$$S_{agg(j,i)} := \overline{\Delta SDR_j}(i) = \frac{1}{N_{obs}(i)} \sum_{m=1}^{N_{obs}(i)} \Delta SDR_j(i,m)$$

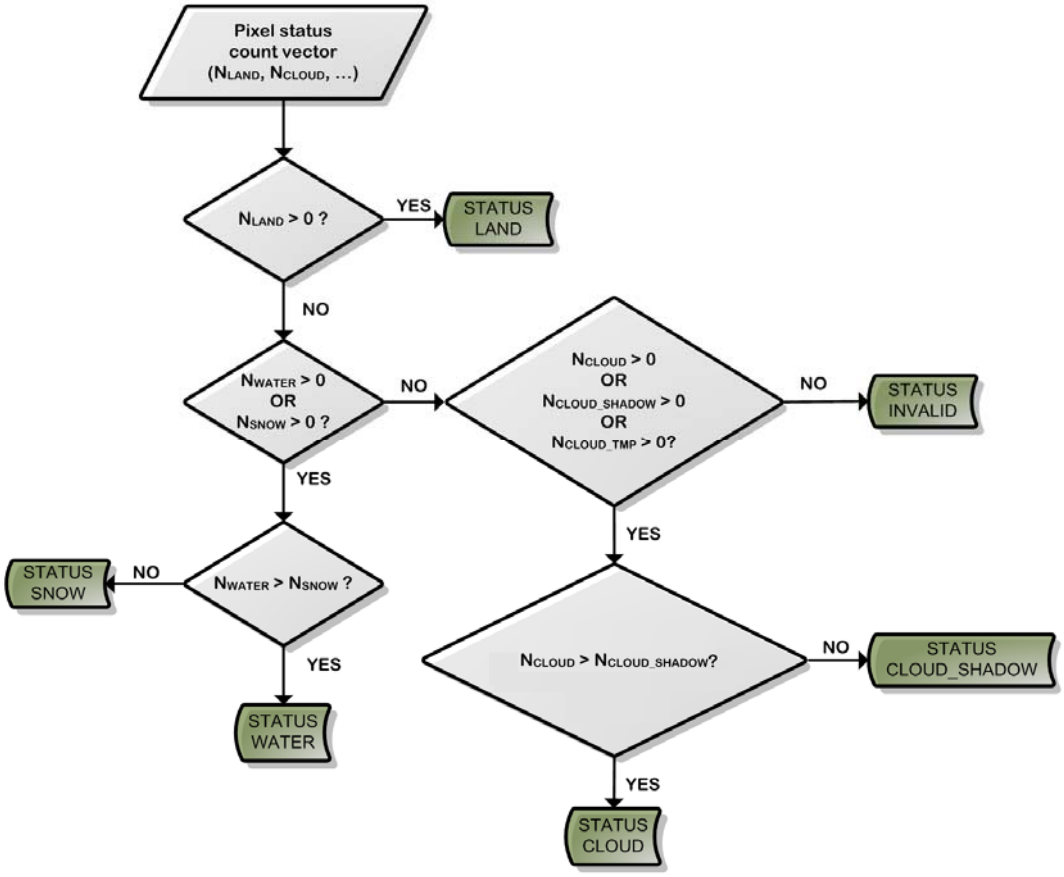


Figure 3-18: Logical flow of the pixel classification after aggregation. NX (with STATUS_X from Table 3-11) represents the number of single observations of a pixel Pi with status X, the green boxes show the final.

3.5.4 Clear land filtering

3.5.4.1 Clear land feautre of IdePix

An important step before or in the compositing algorithm is the application of the clear land feature, which was developed to deal with a significant amount of residual clouds and haze not detected by the other IdePix features and which are thus present in the L2 products. The clear land feature is applied to all pixels classified as "clear land" by IdePix. The steps of this clear land pixel filter applied to each pixel of the SDR image can be summarized as follow:

- Conversion from RGB to HSV
- Retrieving of probability of land and cloud based on prior probability densities of each class
- Applying of threshold to a combination of the retrieved probability of land and cloud

This algorithm is based on surface reflectance values (i.e. on the final L2 products, just before the re-projection and compositing steps).

To convert from RGB to HSV, the values of H (hue), S (saturation) and V (value) could been computed from the values of R (red), G (green) and B (blue) for each pixel by applying the following equations (eq. 3-14):

$$V = MAX(R,G,B)$$

$$S = \begin{cases} 0 & if\ R = G = B = 0 \\ \dfrac{MAX(R,G,B) - MIN(R,G,B)}{MAX(R,G,B)} \end{cases}$$

$$H = \begin{cases} 0° & if\ R = G = B \\ 60° \cdot \left(0 + \dfrac{G - B}{MAX(R,G,B) - MIN(R,G,B)}\right) & if\ MAX(R,G,B) = R \\ 60° \cdot \left(2 + \dfrac{B - R}{MAX(R,G,B) - MIN(R,G,B)}\right) & if\ MAX(R,G,B) = G \\ 60° \cdot \left(4 + \dfrac{R - G}{MAX(R,G,B) - MIN(R,G,B)}\right) & if\ MAX(R,G,B) = B \end{cases}$$

$$if\ H < 0°\ then\ H = H + 360°$$

The clear land filter uses the following bands coding the false colour RGB.

R = SDR of MERIS band 7

G = SDR of MERIS band 14

B = SDR of MERIS band 3

The computed values of hue, saturation and value of each pixel allows determining the probability of cloud, town and land on each pixel. The value of the probabilities is determined for this purpose by the prior probability density. This prior probability density can be derived from manually classified cloud, town and land pixel in satellite images. Here different MERIS FR L2 SDR pre-processing

products has been used. The Figure 3-19 shows the exemplified prior probability density for cloud, town and land.

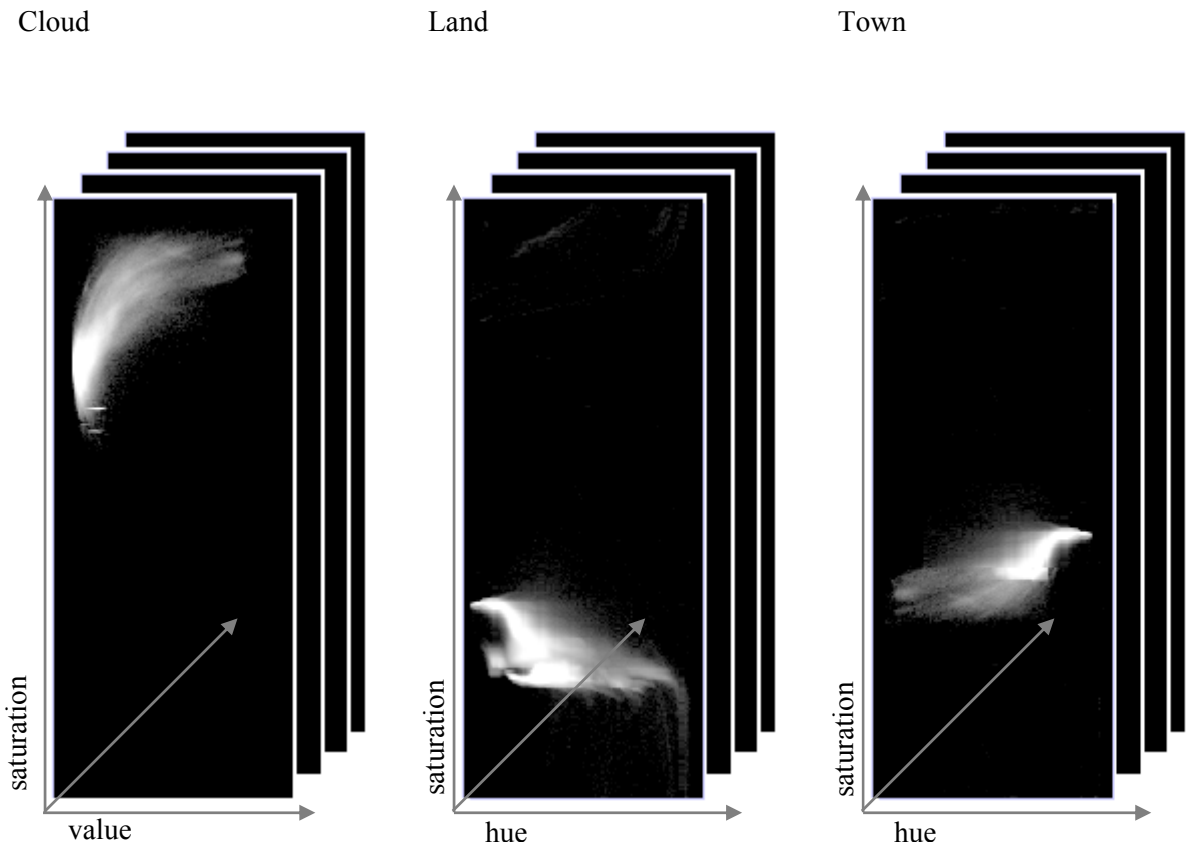


Figure 3-19: Exemplified prior probability density for cloud, town and land derived from MERIS FR L2- SDR pre-processing products

Based on the probability of clear land (land+ town) and cloud pixel, the final status for each pixel has been defined by the following conditions (eq. 3-15):

$$p_{cloud} = NaN \text{ and } p_{land} = NaN \text{ and } p_{town} = NaN \Longrightarrow Status_{final} = cloud$$

$$p_{cloud} = NaN \text{ and } p_{land} \neq NaN \text{ or } p_{town} \neq NaN \Longrightarrow status_{final} = land$$

$$p_{cloud} \neq NaN \text{ and } p_{land} = NaN \text{ and } p_{town} = NaN \Longrightarrow status_{final} = cloud$$

$$p_{cloud} \neq NaN \text{ and } p_{land} \neq NaN \text{ and } p_{town} = NaN \Longrightarrow status_{final} = \begin{cases} land & if\ p_{cloud} < p_{land} \\ cloud & if\ p_{cloud} \geq p_{land} \end{cases}$$

$$p_{cloud} \neq NaN \text{ and } p_{land} = NaN \text{ and } p_{town} \neq NaN \Longrightarrow status_{final} = \begin{cases} land & if\ p_{cloud} < p_{town} \\ cloud & if\ p_{cloud} \geq p_{town} \end{cases}$$

$$p_{cloud} \neq NaN \text{ and } p_{land} \neq NaN \text{ and } p_{town} \neq NaN \Longrightarrow status_{final} = \begin{cases} cloud & if\ p_{cloud} \geq p_{town} \text{ and } p_{cloud} \geq p_{land} \\ land & otherwise \end{cases}$$

The combination of the other IdePix feature and clear land feature leads to a balancing of their strengths and weaknesses. Although it does not completely solve the problem of remaining clouds, it allows significant improvement of the quality of the detection of clouds.

3.5.4.2 Temporal filtering

An important step in the compositing algorithm is the temporal pixel classification, which was developed to deal with a significant amount of residual clouds and haze not detected by Idepix and the clear land filter and which are thus present in the L2 products. The temporal pixel classification amounts to a temporal filtering. It is applied to all pixels classified as "clear land" by IdePix and consists in computing the mean μ and standard deviation σ of the MERIS band 8 ρ_8 over a certain time interval. This time interval does not only consist of the compositing period but has been extended by 10 days before and after the compositing period, thus corresponding to 27 days for the compositing periods of 7. Such enlargement was found to be necessary to ensure statistic values (μ and σ) computed based on a sufficient amount of data. Based on these μ and σ values, the three following thresholds τ_1, τ_2 and τ_3 have been defined (eq. 3-16):

$\tau_1 = \sigma / \mu$

$\tau_2 = \mu + \sigma$

$\tau_3 = 1.35\ \mu$

The threshold τ_3 has been inferred from investigations about BRDF achieved in this project, in which a "variability" of the SDR measurements between 5 to 31% was found. The threshold value was defined using the maximum variability (i.e. 31%) plus an additional safety factor.

A pixel is flagged as cloud (i.e. not clear land) if it satisfies the following conditions (eq. 3-17)

$\tau_1 > 0.2$

$\rho_8 > \text{Min}(\tau_2; \tau_3)$

Accordingly, the number of clear land of observation in the L3 products are a combined function of the composite period but also of the temporal filter efficiency. The temporal filter efficiency depends on three main factors: (i) the length of the filter interval, (ii) the quality of IdePix results and (iii) the pixels seasonality.

With regard to the length of the filter interval (first factor), it is related to the statistic values computation: a longer filter interval will allow more robust mean and standard deviation values (μ and σ) and hence, a higher efficiency of the filter. The IdePix performance (second factor) is also important since it will influence the proportions of correct and misclassified clear land pixels in the set of clear land pixels to reclassify through the temporal filter. If the fraction of misclassified clear land pixel is high, the corresponding SDRs lead to an increasing of the standard deviation of the calculated threshold respective and prevent an effective filtering. Finally, the pixels seasonality (third factor) also plays a role in decreasing the robustness of the mean value (μ).

The Figure 3-20 shows the dependence of temporal filtering on the distribution of the SDRs for 3 possible cases:

- **ideal case**: a lot of correct classified clear land pixels and only a few misclassified pixels leads to the following result after applying the temporal filter: only correct classified clear land pixels exist;
- **seasonal case**: only correct classified clear land pixels, but with a seasonal characteristic leads to the following result after applying the temporal filter: a lot of correct classified clear land pixels are reclassified as cloud;
- **worst case**: a lot of misclassified clear land pixels and only a few correct classified pixel leads to the following result after applying the temporal filter: a lot of misclassified clear land pixels still remain.

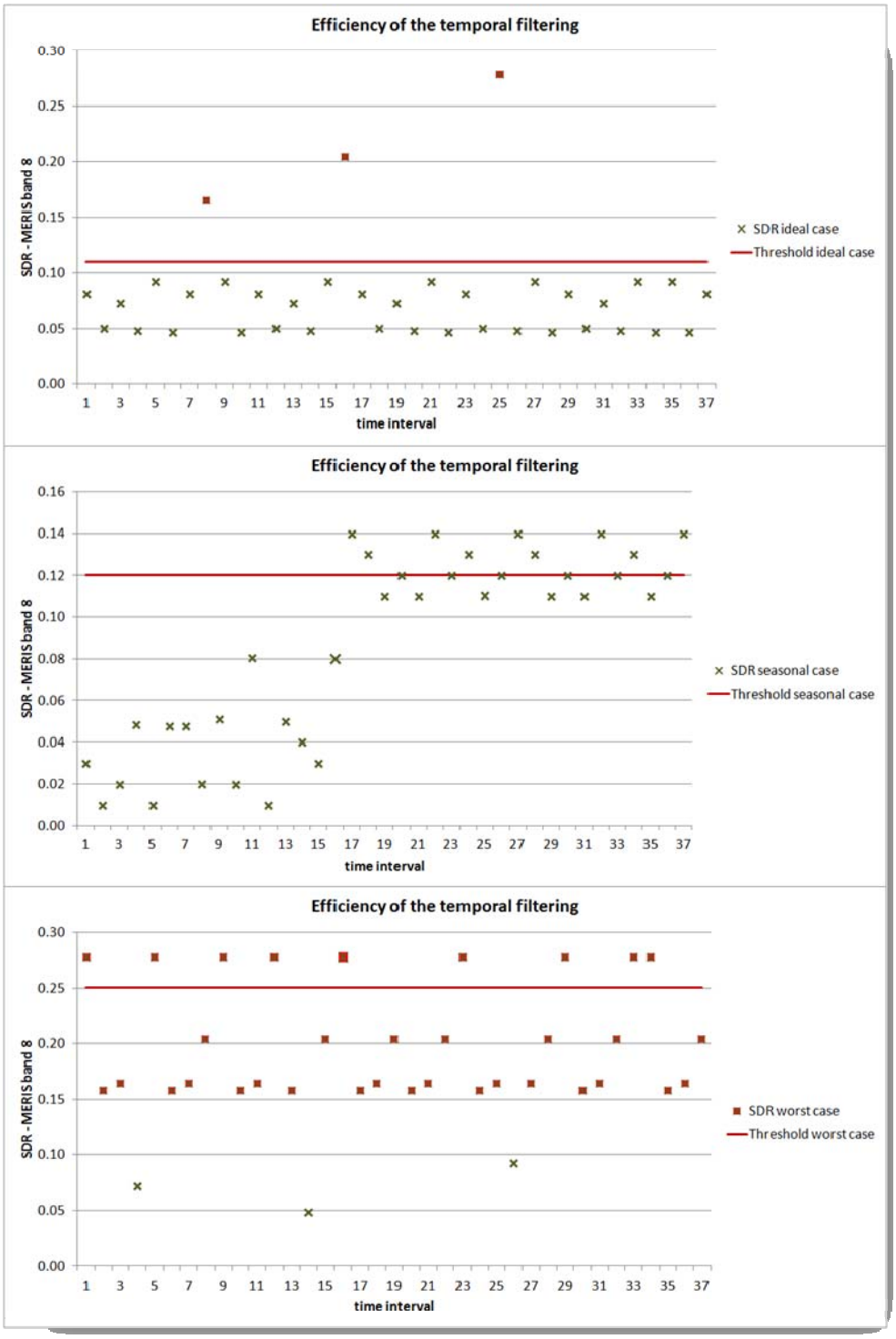


Figure 3-20: Efficiency of temporal filtering depends on the quality of the classification by IdePix and seasonal variability of the surface

3.5.5 Generation of mosaics

As final step, all grid cells (tiles) which were part of one or more input products for the aggregation and which were processed as described above are merged into a single Level 3 result product. Usually, this product shall be a global mosaic. The Table 3-12 summarizes the bands, which will be included in the final L3 MERIS pre-processing products [AD-8].

Table 3-12: Variables used in Level 3 mosaicing.

PARAMETER	DESCRIPTION
SR of band x; x=1 ... 10,12 ... 14	surface reflectance of band x; x=1 ... 10,12 ... 14 for MERIS
ε_{SR} of band x; x=1 ... 10,12 ... 14	uncertainties of the surface reflectance of band x; x=1 ... 10,12 ... 14 for MERIS
Ndvi	NDVI
Status	current pixel status classified by IdePix, clear land filter and temporal filter
Status_counts of status y; y=land, snow/ice, water, cloud, etc.	number of observation over pixel covered by clear land clear snow/ice clear water cloud cloud shadow

3.6 BRDF - Correction - averaging of the surface directional reflectances

To produce MERIS composited images corrected for angular effects (so called L3A), the base line is to start from the atmospherically corrected daily images and through an appropriate compositing technique, at the same time actually produce composited images at various time steps, and correct the reflectances for sun and view directions. The temporal frequency of these composites is variable according to various needs. The classification module needs "fundamental" composites which could be averaged over dedicated periods according to each equal reasoning area.

Three methods were implemented, tested and intercompared during the GlobCover project: the BISE method [RD-71], the Mean Compositing algorithm [RD-72] and an algorithm developed by [RD-73]. These methods and the results and conclusions of the analysis (which are still applicable), are summarized below in the section.

In addition, during the Round Robin phase of the CCI-LC project, a further investigation is planned to test a fourth method [RD-74]. This method is presented in the section 3.6.1.4 and its performance will be pointed out in the Product Validation and Algorithm Selection Report of the CCI-LC project.

3.6.1 GlobCover experience

3.6.1.1 Method 1: BISE algorithm

A widely used technique is that of the Best Index Slope Extraction (BISE) [RD-71]. It has been used for the generation of LASUR (LAnd SUrface Reflectance) data set [RD-75] that includes top of atmosphere and « surface » (atmospherically corrected) reflectances, temperatures and NDVI, viewing and sun angles, plus a channel (flag channel) at 1/7° x 1/7° and 1° x 1° spatial resolution. The BISE algorithm is therefore well suited for residual noise removal on NDVI temporal series. But the need here is to produce noise free reflectance time series, in addition to NDVI time series. The noise associated with reflectance time series has therefore to be carefully assessed with this method.

From the first date of the NDVI time series, for a given pixel, the BISE algorithm searches forward and considers the succeeding point if it is of higher value than the preceding point. When the NDVI value decreases from one day to the next, this decrease is accepted only if there is no point ahead in a pre-defined period of time (called a sliding period, typically 30 days) with a value greater than a percentage threshold of the difference between the two days in consideration. If such a high value is encountered it is selected and the low point is ignored. The original threshold (20%) has been empirically determined for West African conditions.

The drawback of such method lies in the fact that the first point of the series can be kept even if it is cloudy. We have therefore used a slightly modified version allowing reducing the errors generated at the beginning of the cycle. In this version (Figure 3-21), the date of the maximum NDVI value is firstly searched. From this date, the temporal profile is divided in two parts: one corresponding to the growing vegetation phase and another corresponding to the senescence phase.

For each part, BISE scans the points searching if the previous point is greater than the current point. To do that, the slope between two successive points is processed. If this slope is lower than a predefined percentage (20%) then the points are kept. Moreover, some irregularities may still be present (considered as errors) in the shape of the NDVI profile. To detect at best these errors, it is supposed than in each sliding window, the difference between two kept points do not exceed a threshold following the relationship [NDVI (t) – NDVI (t+ti)]/NDVI (t) < 0.15. Following the Figure 3-21, the low NDVI value (shown by the red triangle) is not kept. Therefore, the smaller the sliding window is the larger the number of kept points is.

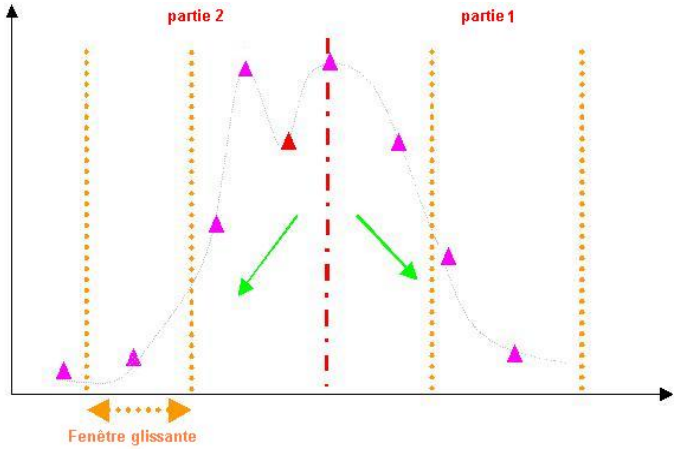


Figure 3-21: Scheme for the BISE method. The triangles represent points of a standard profile. The profile is divided in two parts from the point where NDVI is maximal. In each part (part 1 and part 2), successive iterations on the NDVI points start from the highest point of the profile (green arrows). The time period of the sliding window is draw in vertical dashed points (figure taken from [RD-2])

3.6.1.2 Method 2: Vancutsem algorithm

Another way to avoid the inconsistencies due to geometric and atmospheric variations while being flexible enough is to use a statistical approach. We propose to consider here the Mean Compositing (MC) method [RD-72]. The methodology (Figure 3-22) consists of averaging for each pixel and in each spectral band all the valid reflectance values acquired during the chosen compositing period. Vegetation indices are computed on the basis of these reflectance averages.

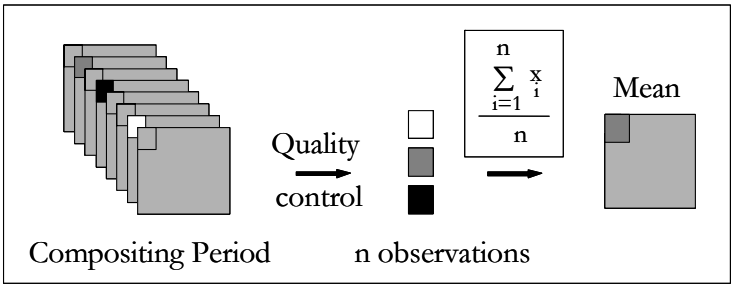


Figure 3-22: Quality control of all reflectance values and mean estimation for each band (figure taken from [RD-2])

This compositing strategy requires two conditions to be applicable. First, the daily images have to be superimposed with an accuracy corresponding to the sub-pixel level, which is the case for the SPOT-VGT images and which will be the case for MERIS images after the geometric corrections. Second, in case of a cyclic variation of the viewing zenith angle, the compositing period must be a multiple of this cycle. In this way, the different compositing periods will have a similar distribution of the viewing angle.

The first step of the work consists in determining the sub-cycle of the viewing angle which a factor independent from the pixel location. It has been determined plotting a temporal profile of MERIS RR L2 reflectances for two sites extracted from the BELMANIP database (see Figure 3-23). A sub-cycle

of around 17 days has been found leading to a temporal period synthesis of around 34 days or 51 days.

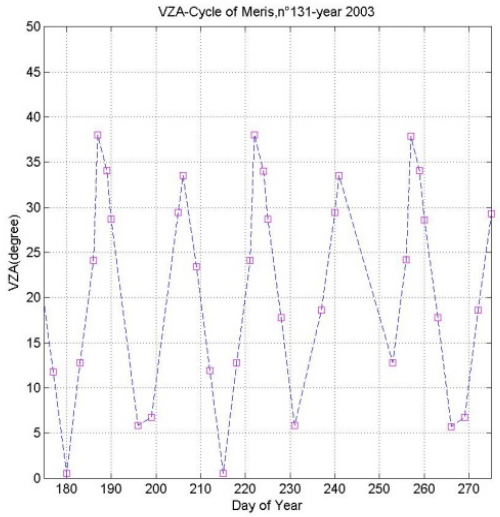


Figure 3-23: Subcycle of the viewing zenith angle (figure taken from [RD-2])

3.6.1.3 Method 3: Hagolle algorithm

The Hagolle algorithm (CYC) is based on an improvement of the method called BDC [RD-76, RD-77] and has been tested and applied so far on SPOT-VGT data.

Its principle is to fit a linear bidirectional reflectance model on the atmospherically corrected reflectance data over a temporal window, then use the model to retrieve a reflectance corrected for sun and view effects (see Figure 3-24). The BRDF model used is that proposed in [RD-78]. It encompasses three unknown coefficients, $k_{i=0,1,2}$, weighting the Lambertian (or isotropic) component, a geometric kernel, and a volume-scattering kernel. The inversion of the BRDF model is operated on a pixel-per-pixel basis for each of the spectral bands of the sensor, provided that at least 3 cloud-free scenes are available within the temporal window. The $k_{i=0,1,2}$, are retrieved by minimizing the cost function C defined by (eq. 3 18):

$$C = \left(\sum_{i=1}^{N} \frac{(\rho_i - \hat{\rho}_i)^2}{\sigma_i^2} \right) + \frac{(k_1 - C1(\lambda))^2}{\sigma_{k1}^2} + \frac{(k_2 - C2(\lambda))^2}{\sigma_{k2}^2}$$

In that equation, $C1(\lambda)$ and $C2(\lambda)$ are constants for a given spectral band. They do not depend on the pixel not on the date. They have been determined by prior experiments from the inversion of the model for various types of landscapes, which has provided averaged k_1 and k_2 values. $C1(\lambda)$ and $C2(\lambda)$ have then been assigned to these averaged k_1 and k_2 values. N is the number of observations available in the temporal window, ρ_i are the measured reflectance data, and $\hat{\rho}_i$ are modelled reflectances. $1/\sigma_i^2$ are the weights of the observations and $1/\sigma_{k1}^2$ and $1/\sigma_{k2}^2$ are the weights of the constraints, they have been determined by previous work [RD-79].

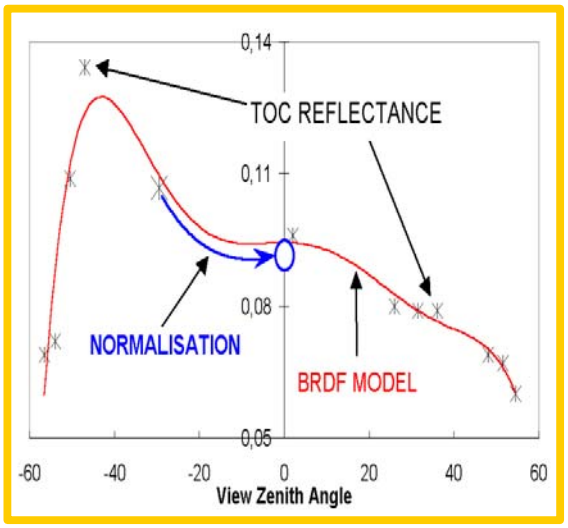


Figure 3-24: Scheme of the angular effects normalization procedure (figure taken from [RD-2])

The use of C1(λ) and C2(λ) in the cost function brings some a priori information in the system, which constrains and influences the result (Figure 3-25). This makes easier to detect and discard residual cloud contaminated pixels. The detection is operated using the blue band because it enables the best distinction between clouds and continental surfaces. If the standard deviation of errors σ of the model fit is above a threshold, our algorithm discards all the observations that are greater that the adjusted model values plus σ. At this stage, many previously undetected cloudy pixels are discarded. Then a second fit and a new σ value is computed (smaller than the previous one); all the pixels with a difference to the fitted model greater than 1.5* σ are discarded whatever the sign of the difference is. This last step is iterated until no pixels are discarded anymore, or until the number of remaining pixels is lower than 3.

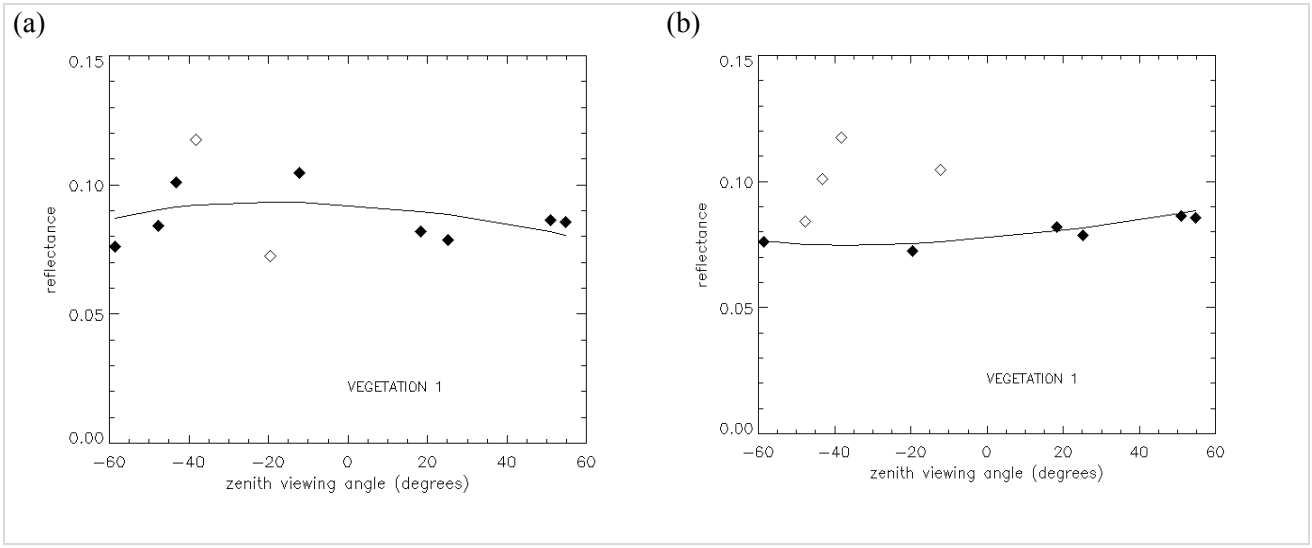


Figure 3-25: BRDF model fit for VEGETATION data with (a) no a priori information and (b) a priori information. Unfilled diamonds are discarded pixels, only filled diamonds are used for the BDRF estimate (from [RD-2])

The 3 coefficients $k_{i=0,1,2}$ are used to normalize the cloudless reflectance dataset entering the 10-day period to nadir view and solar observation observed at 10h00 local time for the median day of the compositing period. Besides, the method supplies for each composite product the following information for each pixel and each spectral band: the normalized reflectance, the number of observations used to assess the BRDF and a quality flag including information about the processed pixel. This algorithm has been tested and evaluated with a method described at the end of this Section. It is really efficient when there are a sufficient number of valid observations during the temporal compositing period (Figure 3-26). As for the Vancutsem et al. algorithm, the fact that MERIS data are about three times less frequent than SPOT-VGT data may induce some limitations of the method.

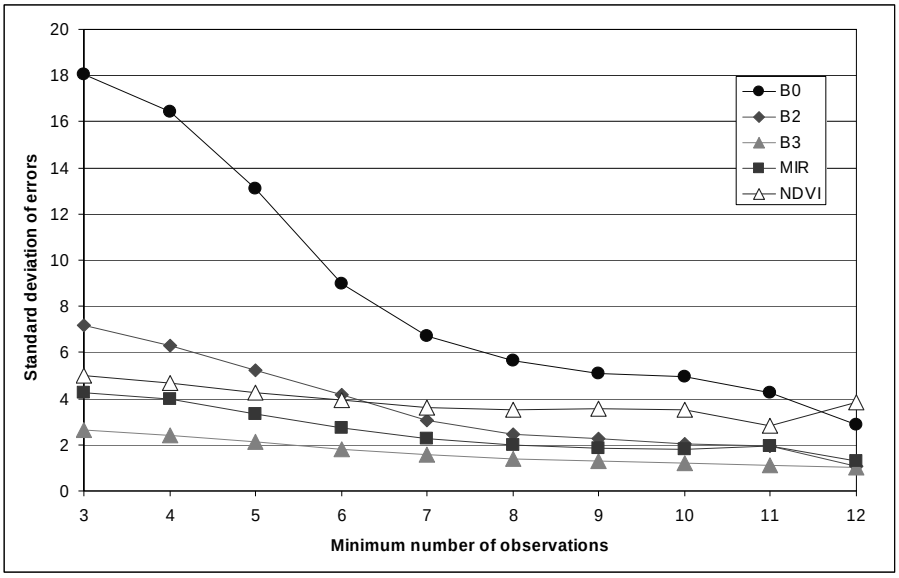


Figure 3-26 : Standard deviation of the errors calculated as a function of the minimum number of observations used in the BRDF model inversion (from [RD-2])

3.6.1.4 GlobCover analyses and decisions for the BRDF correction

The work of the GlobCover project has addressed the problem of estimating the performance of global reflectance composited products of land surfaces. Two criteria were used to perform such estimations: a qualitative criterion relying on temporal MERIS RR profile over pixels extracted from BELMANIP sites. A quantitative criterion providing a way to estimate the standard deviation of the noise added to time series of reflectance data [RD-2].

The qualitative criteria have allowed observing that all the methods were able to reproduce the temporal trends of the temporal vegetation evolution. They presented rather similar behaviours. Using the quantitative criterion, two situations may be distinguished. When the number of input data is relatively sufficient (favourable cases), the performances of BISE, MC, CYC are rather similar. When this number of data is low (unfavourable conditions but a regular case for MERIS), the BISE method is not so efficient as MC or CYC. The MC method restores a mean of the values observed during the temporal period of acquisition. The values derived from BISE are directly linked to the chosen sliding window. The selection of this sliding window is a function of the phenological stage of the vegetation. Therefore, this method requires having complementary information on this phenology. To have better

results with BISE, a geographical mapping of this parameter should be applied. This dependence makes this method non generic at the global scale. A further investigation was done to assess the performance of MC and CYC. Both methods have very similar results but with always a slight advantage for MC.

The GlobCover project finally selected the MC algorithm as BRDF correction and compositing method. However, a slight modification was adopted, which consisted in coupling the MC method with the filtering module of the CYC algorithm. The MC method was first applied on the observations present in a 51-day compositing period (defined as a good trade off between a small temporal period suitable for monitoring the temporal evolution of the vegetation and a large period to be able to produce composites). Then, the filtering module of the CYC method was performed to remove outliers. This method was applied on the 51 days period to flag "valid" observations which were used to perform composites over reduced time period (15-day composites in GlobCover).

3.6.2 Method 4: improved BRDF inversion developed by Vermote et al. 2009

The proposed method for the estimation and the correction of directional effects on satellite derived surface reflectance time series based on the BRDF inversion technique used for the MODIS BRDF/albedo product [RD-74, RD-80, RD-81, RD-82, RD-83] with the objective of minimisation the differences between the observed and modelled directional reflectances over a given time period. The fundamental BRDF model, the Ross-Li-Maignan model, computes the surface reflectances as the sum of three terms, which describe the isotropic, volumetric and geometric parts (eq. 3-18)

$$\rho(\theta_s,\theta_v,\Phi) = k_0 + k_1F_1(\theta_s,\theta_v,\Phi) + k_2F_2(\theta_s,\theta_v,\Phi) = k_0\left[1+\frac{k_1}{k_0}F_1(\theta_s,\theta_v,\Phi)+\frac{k_2}{k_0}F_2(\theta_s,\theta_v,\Phi)\right]$$
$$= k_0[1+VF_1(\theta_s,\theta_v,\Phi)+RF_2(\theta_s,\theta_v,\Phi)] \ where\ V=\frac{k_1}{k_0}\ and\ R=\frac{k_2}{k_0}$$

where θ_s, θ_v, Φ are the solar zenith, view zenith and the relative azimuth angles, F_i are the kernel and the k_i are the free weighting parameter. The kernel F_1 as the volumetric scattering kernel accounting for the hot spot and F_2 as the geometric scattering kernel, are function of the observation geometry and given by (eq. 3-19):

$$F_1(\theta_s,\theta_v,\Phi) = \frac{4}{3\pi}\frac{1}{\cos\theta_s+\cos\theta_V}\left[\left(\frac{\pi}{2}-\xi\right)\cos\xi+\sin\xi\right]\left[1+\frac{1}{1+{}^{\xi}/_{\xi_0}}\right]-\frac{1}{3}$$

$$F_2(\theta_s,\theta_v,\Phi) = \frac{m}{t}(t-\sin t\cos t-\pi)+\frac{1+\cos\xi}{2\cos\theta_s\cos\theta_V}$$

$$m = \frac{1}{\cos\theta_s}+\frac{1}{\cos\theta_v}$$

$$\xi = \cos^{-1}(\cos\theta_s\cos\theta_v+\sin\theta_s\sin\theta_v\cos\Phi)$$

$$\xi_0 = \frac{LD}{H}\frac{(\cos\theta_s + \cos\theta_v)^2}{4\cos\theta_s\cos\theta_v} \approx \frac{LD}{H}$$

$$\Delta(\theta_s, \theta_v, \Phi) = \sqrt{\tan^2\theta_s + \tan^2\theta_v - 2\tan\theta_s\tan\theta_{sv}\cos\Phi}$$

$$\cos t = \frac{2}{m}\sqrt{\Delta^2 + (\tan\theta_s\tan\theta_v\sin\phi)^2}$$

where ξ is the scattering angle, ξ_0 is a characteristic angle that can be related to the ratio of scattering element size and the canopy vertical density, H is the canopy depth, L is the leaf area index and D define the typical scale of the leaves. To avoid the addition of a free parameter, [RD-82] suggests taking a constant value of $\xi_0 = 1.5°$.

The alternative BRDF inversion technique assumes no temporal variation of shape of the BRDF, but allows the change of target reflectances during the compositing period in the contrast to the MODIS standard algorithm. Based on this hypothesis and on a set of N observations the parameter V and R are retrieved by minimizing the merit function M on the classical way of computation of the first derivation of the merit function, which is defined by (eq. 3-20):

$$\frac{\partial M}{\partial V} = \frac{\partial M}{\partial R} = 0$$

$$M = \sum_{i=1}^{N-1}\frac{\left(\rho_{i+1}[1 + VF_1^i + RF_2^i] - \rho_i[1 + VF_1^{i+1} + RF_2^{i+2}]\right)^2}{day^{i+1} - day^i + 1}$$

$$\begin{pmatrix} \sum_{i=1}^{N-1}\Delta^i\rho F_1\,\Delta^i\rho F_1 & \sum_{i=1}^{N-1}\Delta^i\rho F_1\Delta^i\rho F_2 \\ \sum_{i=1}^{N-1}\Delta^i\rho F_1\Delta^i\rho F_2 & \sum_{i=1}^{N-1}\Delta^i\rho F_2\Delta^i\rho F_2 \end{pmatrix}\begin{pmatrix} V \\ R \end{pmatrix} = \begin{pmatrix} -\sum_{i=1}^{N-1}\Delta^i\rho\Delta^i\rho F_1 \\ -\sum_{i=1}^{N-1}\Delta^i\rho\Delta^i\rho F_2 \end{pmatrix}$$

where

$$\Delta^i d = day_{i+1} - day_i + 1$$

$$\Delta^i\rho = \frac{(\rho_{i+1} - \rho_i)}{\sqrt{\Delta^i d}}$$

$$\Delta^i\rho F_1 = \frac{\left(\rho_{i+1}F_1^i - \rho_i F_1^{i+1}\right)}{\sqrt{\Delta^i d}}$$

$$\Delta^i\rho F_2 = \frac{\left(\rho_{i+1}F_2^i - \rho_i F_2^{i+1}\right)}{\sqrt{\Delta^i d}}$$

A further refinement accounting for a change of the BRDF shape with the vegetation cover based on the clustering of the data with respect to the NDVI. For each of these classes the parameter R and V were retrieved and the relationship to the NDVI can be derived by a linear fit among the different classes.

Therefore, the full BRDF correction process includes the following steps

- first guess estimate of V and R by inverting the model using the method described before
- correction of the reflectances with these estimates of V and R
- computation of the NDVI
- clustering based on the NDVI
- inversion of the model for each class to get V and R
- linear fit between NDVI and the parameter V and R
- final BRDF-correction

The proposed method doesn't assume an invariance of the surface during the compositing period and allows the monitoring of the vegetation with a period smaller than the compositing method. Therefore the correction can be performed with the same frequency as the observation. The used parameter V and R can be derived from previous observations, which avoids the limitation by data availability.

The important step of the work consists in the verification of the proposed algorithm for the CCI-LC project and its requirements by a study based on synthetic data. Because the proposed approach was evaluated on a global MODIS data set at a 0.05° resolution for two channels and the NDVI. Therefore the widely used PROSPECT/SAIL model has been implemented and different BRDF could be retrieved [RD-84]. The following shows an example of the retrieved BRDF.

Based on these modelled BRDF's the retrieving of the parameter V and R will be investigated. The typical observation und illumination geometry is derived from the MERIS Round Robin database. It has been determined plotting a temporal profile of MERIS FSG L1 viewing and sun angles for six sites (see Figure 3-28 and Figure 3-29). A sub cycle of viewing zenith angle of around 17 days has been found.

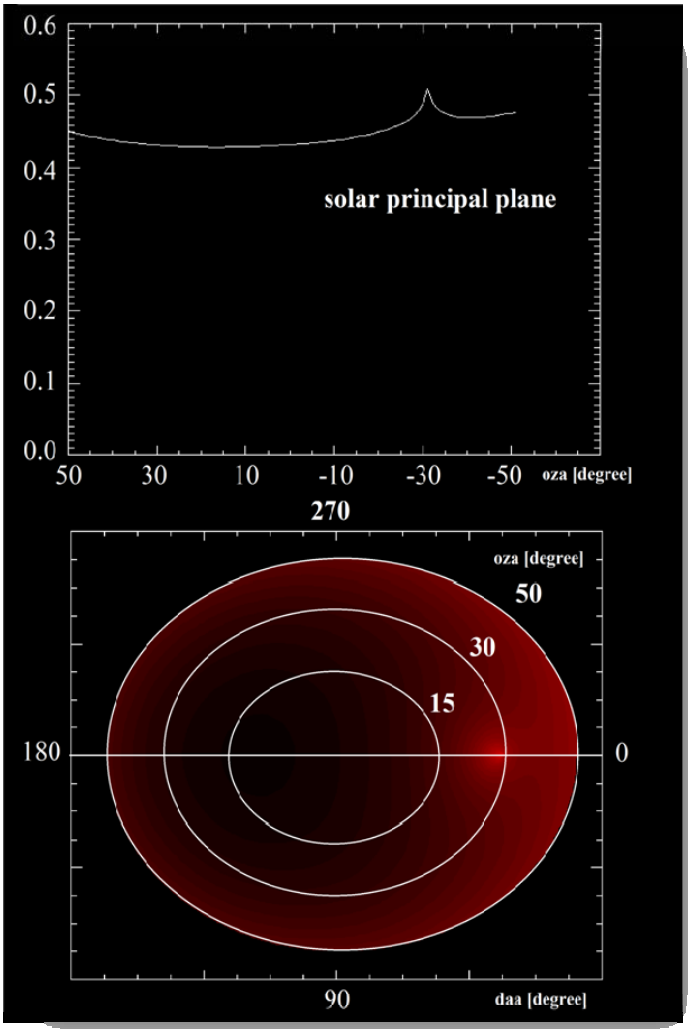


Figure 3-27: Example of modelled BRDF using PROSPECT/SAIL model (parameter: Cab = 30.0 (chlorophyll content [µg cm-2]); Car = 10.0 (carotenoid content [µg cm-2]); Cbrown = 0.0 (brown pigment content [arbitrary units]; Cw = 0.015 (equivalent water thickness EWT [cm]); Cm = 0.009 (leaf mass per area LMA [g.cm-2]); N = 1.5 (structure coefficient); LAI = 4 (leaf area index); angl = 50.0 (average leaf angle [degree]); psoil = 1(soil coefficient); skyl = 70 (% diffuse/direct radiation); hspot = 0.05 (hot spot - leaf size relative to canopy height)); ihot = 1.0 (hot spot flag); tts = 30.0 (solar zenith angle [degree])). daa: difference azimuth angle [degree]; oza: observer zenith angle [degree]

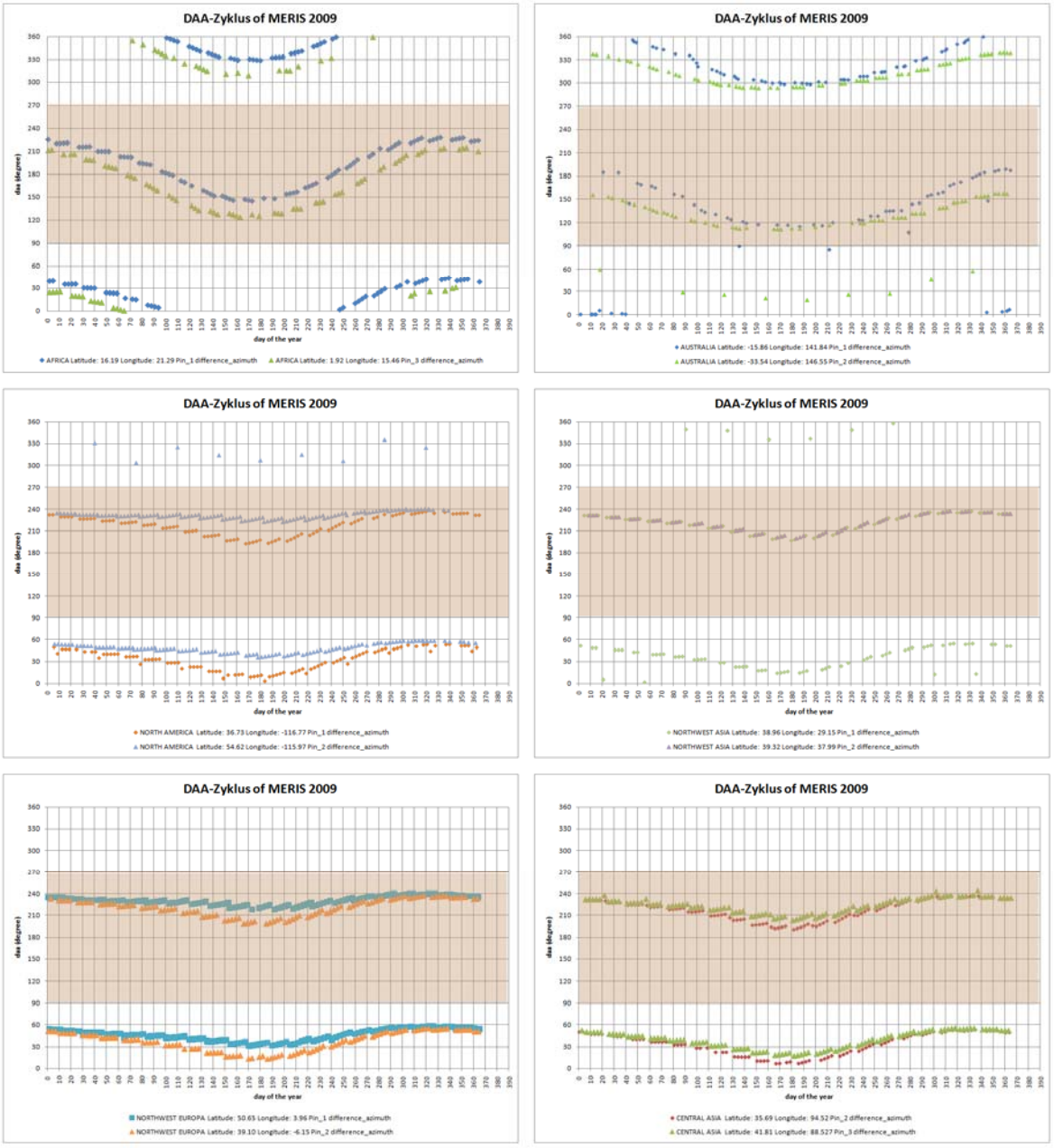


Figure 3-28: Sub cycle of the difference azimuth angle

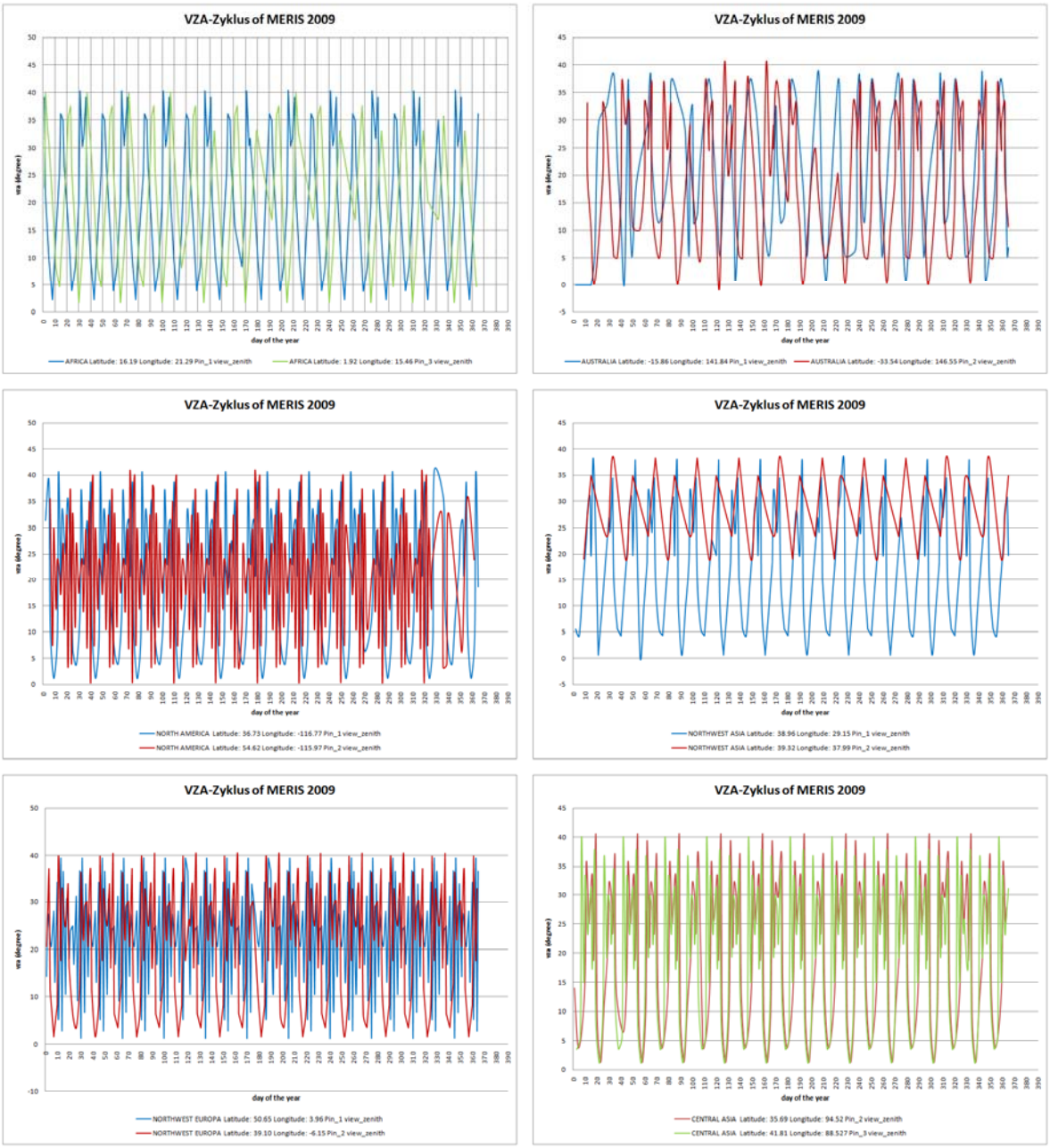


Figure 3-29: Sub cycle of the viewing zenith angle

3.6.3 Compositing process

Finally, it is worth noting that since the BRDF correction is linked with the compositing process, the pre-processing module does not deliver daily products but global surface reflectance "fundamental" composites 7-day composite (i.e. generated over a "fundamental" period). The 7-day compositing period is adapted on a fixed basis. So every January 1st as starting point and a continuous 7 day

intervals for each year has been proposed. The last composite of each year and the last February composite of a leap year include 8 days instead 7days.

These "fundamental" composites will serve as input to the classification module, and thus for the generation of longer (seasonal and annual) composites. This articulation is illustrated in Figure 3-30.

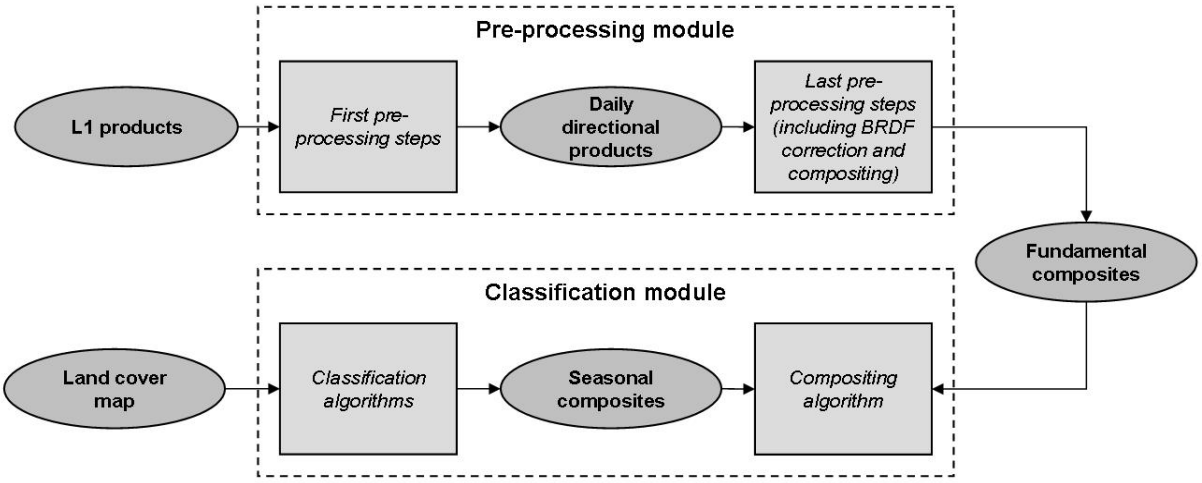


Figure 3-30: Articulation between the fundamental composites delivered by the pre-processing module and the seasonal composites used in the classification algorithms

3.7 Validation of algorithms

Validation is needed in order to assess the effectiveness and quality of the different algorithms. Several mathematical and other empirical methods of validation are available. However, the validation of an algorithm is fairly complex and the validation procedure proposed for the pre-processing algorithms is to run through an iterative process with following major points:

- Testing the algorithms on (real) remote sensing data for the use cases
- Comparison between independent in situ data and outputs created by the algorithms if possible

3.7.1 Validation plan

The validation plan provides the different factors influencing the requirements and the selection process for suitable validation methods should be examined.

Therefore in the following steps the influence of several disturbances as noise and clouds are analyzed and the sensitivity due to the choice of thresholds is determined. The validation and optimisation of the algorithms will be an iterative process and consists of:

- Testing the algorithm with remote sensing data using case studies
- Validation of intermediate products with in-situ data if possible
- Analyzing the dependency of thresholds used in the algorithms

- Quality measure
- Documentation
- At the end we get an overview about the individual thresholds and quality measure of pre-processing intermediate products.

3.7.2 Quality measures

Before processing steps such as validation or assimilation take place it is necessary to understand the error statistics of pre-processing intermediate products. A standard method solving this problem is the comparison of pre-processing intermediate products with in-situ data with a statistical approach. A typical tool solving this problem is the usage of the confusion matrix, which allows statistical measures of the performance of classification [RD-85, RD-86, RD-87].

3.7.2.1 Accuracy and precision in binary classification

A number of accuracy measures are available for assessing the quality of a binary classification.

Table 3-13: Table of confusion, also known as a confusion matrix

		Condition		
		True	False	
Test Result	Positive	True positive	False positive Type I error	→ Positive predictive value
	Negative	False negative Type II error	True negative	→ Negative predictive value
		↓ Sensitivity	↓ Specificity	Accuracy

Thus, the accuracy is the proportion of true results (both true positives and true negatives) in the population. The accuracy is used as a statistical measure of how well a binary classification test correctly identifies or excludes a condition [RD-88, RD-89] (eq. 3 22):

$$accuracy = \frac{number\ of\ true\ positives + number\ of\ true\ negatives}{number\ of\ true\ pos. + number\ of\ false\ pos. + number\ of\ false\ neg. + number\ of\ true\ neg.}$$

On the other hand, precision or relevance (positive predictive value) is defined as the proportion of the true positives against all the positive results (both true positives and false positives) (eq. 3-21):

$$precision = \frac{number\ of\ true\ positives}{number\ of\ true\ positives + number\ of\ false\ positives}$$

The false positive rate is equal to the significance level and is defined as proportion of the false positives against the false positives and true negatives results (eq. 3-22):

$$false\ positive\ rate = \frac{number\ of\ false\ positives}{number\ of\ false\ positives + number\ of\ true\ negatives}$$

The false negative rate is defined as proportion of the false negatives against the false negatives and true positives results (eq. 3 25):

$$false\ negative\ rate = \frac{number\ of\ false\ negatives}{number\ of\ false\ negatives + number\ of\ true\ positives}$$

The sensitivity is defined as proportion of the false negatives against the false negatives and true positives results (eq. 3 26):

$$sensitivity = \frac{number\ of\ true\ positives}{number\ of\ false\ negatives + number\ of\ true\ positives}$$

The specificity is defined as proportion of the true negatives against the false positives and true negatives results (eq. 3-23):

$$specificity = \frac{number\ of\ true\ negatives}{number\ of\ false\ positives + number\ of\ true\ negatives}$$

The segregance (negative predictive value) is defined as proportion of the true negatives against the false negatives and true negatives results (eq. 3-24):

$$segregance = \frac{number\ of\ true\ negatives}{number\ of\ false\ negatives + number\ of\ true\ negatives}$$

The false-accuracy is the proportion of false results (both false positives and false negatives) in the population (eq. 3-25):

$$false - accuracy = \frac{number\ of\ false\ positives + number\ of\ false\ negatives}{number\ of\ true\ pos. + number\ of\ false\ pos. + number\ of\ false\ neg. + number\ of\ true\ neg}$$

An accuracy of 100% means that the measured values are exactly the same as the given values. Accuracy may be determined from Sensitivity and Specificity, provided Prevalence is known, using the equation (eq. 3-26):

$$accuracy = (sensitivity)(prevalence) + (specificity)(1 - prevalence)$$

The prevalence is the proportion of positives to the positives and negatives in the population (eq. 3-27):

$$Prevalence = \frac{number\ of\ true\ positives + number\ of\ false\ negatives}{number\ of\ true\ pos. + number\ of\ false\ pos. + number\ of\ false\ neg. + number\ of\ true\ neg}$$

3.7.2.2 Accuracy and precision in classification

A number of accuracy measures are available for assessing the quality of a classification. The above table is an example confusion matrix. The diagonal elements in this matrix indicate numbers of sample for which the classification results agree with the reference data. The matrix contains the complete information on the categorical accuracy. Off diagonal elements in each row present the numbers of sample that has been misclassified by the classifier, i.e., the classifier is committing a label to those samples which actually belong to other labels. The misclassification error is called commission error [RD-88, RD-89].

The off-diagonal elements in each column are those samples being omitted by the classifier. Therefore, the misclassification error is also called omission error.

		Conditions				Users Accuracy
		Class I	Class II	...	Class N	
Algorithm	Class I	a_{11}	a_{12}	...	a_{1n}	$\frac{a_{11}}{\sum_{i=1}^{n} a_{1i}}$
	Class II	a_{21}	a_{22}	...	a_{2n}	$\frac{a_{22}}{\sum_{i=1}^{n} a_{2i}}$
	⋮	⋮	⋮	⋱	⋮	...
	Class N	a_{n1}	a_{n2}	...	a_{nn}	$\frac{a_{nn}}{\sum_{i=1}^{n} a_{ni}}$
Producers Accuracy		$\frac{a_{11}}{\sum_{i=1}^{n} a_{i1}}$	$\frac{a_{22}}{\sum_{i=1}^{n} a_{i2}}$	...	$\frac{a_{nn}}{\sum_{i=1}^{n} a_{in}}$	Overall Accuracy $\sum_{i=1}^{n} a_{ii}$ for $\sum_{i=1}^{n}\sum_{j=1}^{n} a_{ij} = 1$

In order to summarize the classification results, the most commonly used accuracy measure is the overall accuracy. More specific measures are needed because the overall accuracy does not indicate how the accuracy is distributed across the individual categories. The categories could, and frequently

do, exhibit drastically differing accuracies but overall accuracy method considers these categories as having equivalent or similar accuracies.

By examining the confusion matrix, it can be seen that at least two methods can be used to determine individual category accuracies:

- The ratio between the number of correctly classified and the row total is called the user's accuracy because users are concerned about what percentage of the classes has been correctly classified.
- The ratio between the number of correctly classified and the column total is called the producer's accuracy. The producer is more interested in this accuracy because it tells how correctly the reference samples are classified.

However, there is a more appropriate way of presenting the individual classification accuracies. This is through the use of commission error and omission error.

- Commission error = 1 - user's accuracy
- Omission error = 1 - producer's accuracy

3.7.2.3 *Cohen's kappa, Scott's pi and Krippendorf's alpha coefficient for a table of confusion*

Cohen's kappa and Scott's pi coefficient measure the proportion of observed agreement to the agreement by chance and the maximum agreement attainable to chance agreement considering pair wise agreement and is a statistical measure of concordance for qualitative items. The equation for Cohen's kappa measures κ is defined as the proportion of agreement after expected or chance agreement is removed from consideration [RD-85, RD-86, RD-87].

Cohen's kappa coefficient for a table of confusion ***K***

Cohen's kappa coefficient for a table of confusion K is given by (eq. 3-28):

$$\kappa = \frac{p_0 - p_e}{1 - p_e}$$

where

$$p_0 = \frac{number\ of\ true\ positives + number\ of\ true\ negatives}{number\ of\ true\ pos. + number\ of\ false\ pos. + number\ of\ false\ neg. + number\ of\ true\ neg}$$

$$p_e = \frac{positive\ predictive\ value * sensitivity + negative\ predictive\ value * specificity}{(number\ of\ true\ pos. + number\ of\ false\ pos. + number\ of\ false\ neg. + number\ of\ true\ neg.)^2}$$

or more general :

$$\mathbb{K} = \begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{n1} & a_{n2} & \cdots & a_{nn} \end{bmatrix} \; with \sum_{i=1}^{n}\sum_{j=1}^{n} a_{ij} = 1$$

$$\kappa = \frac{p_0 - p_e}{1 - p_e}$$

where

$$p_0 = \sum_{i=1}^{n} a_{ii}$$

$$p_e = \sum_{i=1}^{n} h_{i\diamond} \cdot h_{\diamond i} \;\; with \; h_{i\diamond} = \sum_{j=1}^{n} a_{ij} \; and \; h_{\diamond j} = \sum_{i=1}^{n} a_{ij}$$

The observed agreement p_0 is the proportion of true results and the expected agreement p_e is the proportion of items for which agreement is expected by chance when the items are random.

Scott's PI coefficient for a table of confusion ***K***

Scott's pi measures similar statistic as Cohen's kappa but differs in terms of how p_e is calculated and is given by (eq. 3-29):

$$\mathbb{K} = \begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{n1} & a_{n2} & \cdots & a_{nn} \end{bmatrix} \; with \sum_{i=1}^{n}\sum_{j=1}^{n} a_{ij} = 1$$

$$\kappa = \frac{p_0 - p_e}{1 - p_e}$$

where

$$p_0 = \sum_{i=1}^{n} a_{ii}$$

$$p_e = \sum_{i=1}^{n} \left(\frac{h_{i\diamond} + h_{\diamond i}}{2}\right)^2 \; with \; h_{i\diamond} = \sum_{j=1}^{n} a_{ij} \; and \; h_{\diamond j} = \sum_{i=1}^{n} a_{ij}$$

Krippendorff's Alpha coefficient for a coincidence matrix ***C***

Krippendorff's Alpha coefficient is also a statistical measure of the agreement between observers or measuring instruments and its general form is (eq. 3-30):

$$\mathbb{K} = \begin{bmatrix} a_{11} & a_{12} & \dots & a_{1n} \\ a_{21} & a_{22} & \dots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{n1} & a_{n2} & \dots & a_{nn} \end{bmatrix} with \sum_{i=1}^{n}\sum_{j=1}^{n} a_{ij} = N$$

$$\mathbb{C} = \mathbb{K} + \mathbb{K}^T = \begin{bmatrix} a_{11} & a_{12} & \dots & a_{1n} \\ a_{21} & a_{22} & \dots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{n1} & a_{n2} & \dots & a_{nn} \end{bmatrix} with \sum_{i=1}^{n}\sum_{j=1}^{n} a_{ij} = 2N$$

$$\kappa = 1 - \frac{D_0}{D_e}$$

where

$$D_0 = \frac{1}{N}\sum_{i=1}^{n}\sum_{j=1}^{n} a_{ij} \cdot_{metric} \delta_{ij}^2$$

$$D_e = \frac{1}{N(N-1)}\sum_{i=1}^{n}\sum_{j=1}^{n} h_{i\circ} \cdot h_{\circ j} \cdot_{metric} \delta_{ij}^2$$

$$with\ h_{i\circ} = \sum_{j=1}^{n} a_{ij}\ and\ h_{\circ j} = \sum_{i=1}^{n} a_{ij}$$

If

$$_{metric}\ \delta_{ij}^2 = _{nominal}\ \delta_{ij}^2 = \begin{cases} 0\ if\ and\ only\ if\ (iff) j = i \\ 1\ if\ and\ only\ if\ (iff) j \neq i \end{cases}$$

then

$$D_0 = \frac{1}{N}\sum_{i=1}^{n}\sum_{j=1}^{n} a_{ij}\ \ with\ i \neq j$$

$$D_e = \frac{1}{N(N-1)}\sum_{i=1}^{n}\sum_{j=1}^{n} h_{i\circ} \cdot h_{\circ j}\ \ with\ i \neq j$$

$$with\ h_{i\circ} = \sum_{j=1}^{n} a_{ij}\ and\ h_{\circ j} = \sum_{i=1}^{n} a_{ij}$$

D_o is the observed disagreement and D_e is the expected disagreement. The coincidence matrix can be retrieved from the confusion matrix (see above).

3.7.3 Validation of pre-processing algorithms

3.7.3.1 Land water delineation and cloud screening - pixel identification

The main tool for validating the pixel identification is the comparison of the results of the algorithm with manually characterised pixels. This requires a large data set of pixels from a globally distributed imagery. Each collected validation sample point (selected sensor pixel) will be classified concerning its characteristics based on the visual inspection of the appropriate sensor pixels and then will be compared with the output of the Pixel Classification [RD-90, RD-91].

In order to ensure the representativeness of the data-set for the natural variability of those parameters which affect the radiation measurement from a space borne instrument additional classification are performed for each validation sample point: the surface type, the climate zone, the season as well as day and night discrimination.

Pixel type characteristics

Cloud characterisation

- totally cloudy
- (desert) dust over clouds
- specular reflection ("glint")
- cloud type (stratus, cumulus, convective cloud, cirrus, unknown)
- semi-transparent cloud (problem: cannot be separated from partly cloudy; can be distinguished if semi-transparent cloud is spatially large)

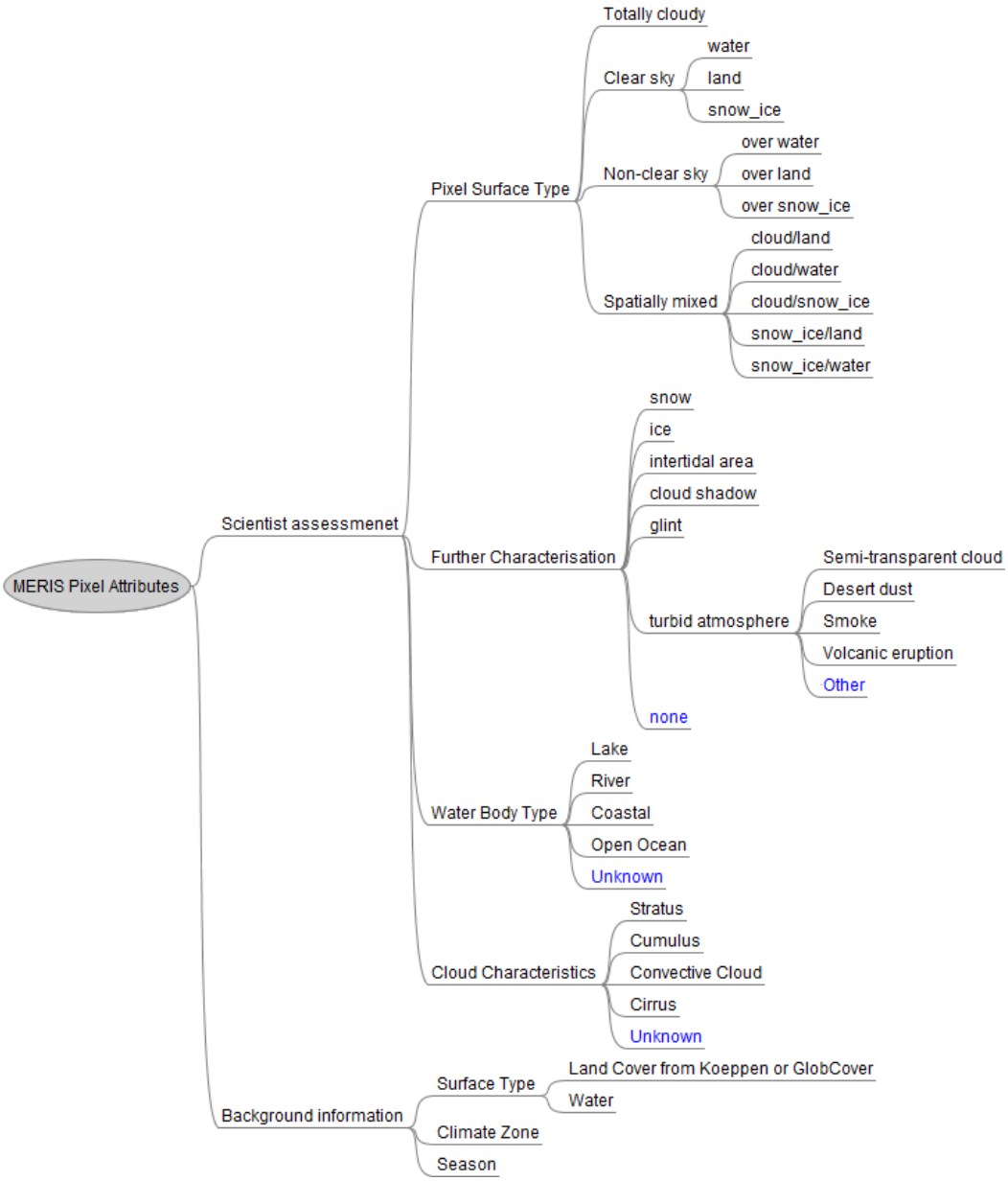
Land / water

- clear sky land / water
- spatially mixed pixel (land/water/cloud) (problem: cloud mixed pixel cannot be separated semi-transparent clouds)
- turbid atmosphere (dust, high aerosol loading, volcanic eruption, desert dust, East Asian dust)
- sun glint
- cloud shadow
- intertidal areas
- water body type: lake, river, coastal, open ocean, unknown

Snow

Ice

Logical structure



Pixel collection requirements

- global distribution
- all seasons equal
- surface types corresponding to their natural occurrence, but all should be covered
- all climate zones corresponding to their natural occurrence, but all should be covered
- day and night (AATSR only) equal
- 30% totally cloudy, 30% clear sky, 30% not clear sky, 10% spatially mixed pixel

- 60% over land, 40% over water
- snow covered land: greenland, Antarctic, mountains, high latitudes (in addition to Greenland and Antarctic), Artic (snow covered ice)
- ice: Arctic, Antarctic
- water types (include all types)
- all viewing zenith angles (sun geometry is covered by all seasons)

Data source requirements

- MERIS FR and RR L1b
- AATSR L1b
- SPOT-VGT L1

Dataset selection

- final size of database shall be ~40 000 samples; this shall be completed in summer 2011
- first database shall have ~5000 samples, it shall be ready mid January; the dataset shall be primarily over ocean, but should include for demo purpose also land (for CCI-LC progress meeting). The distribution shall be ~4.500 ocean and ~500 land.
- totally cloudy 30%
- clear sky 30%
- not clear sky 20%
- spatially mixed pixel 20%
- over land 60%
- over water 40%

Surface Types

The surface type database which is used as ancillary information in this project is the GlobCover land cover map (Figure 3-31). GlobCover land cover maps are available for the epochs 2005 and 2009, both being freely available from the GlobCover ESA website - http://ionia1.esrin.esa.int/. A technical description of the dataset is given in [RD-92]].

This surface type map includes 22 categories (Table 3-14).

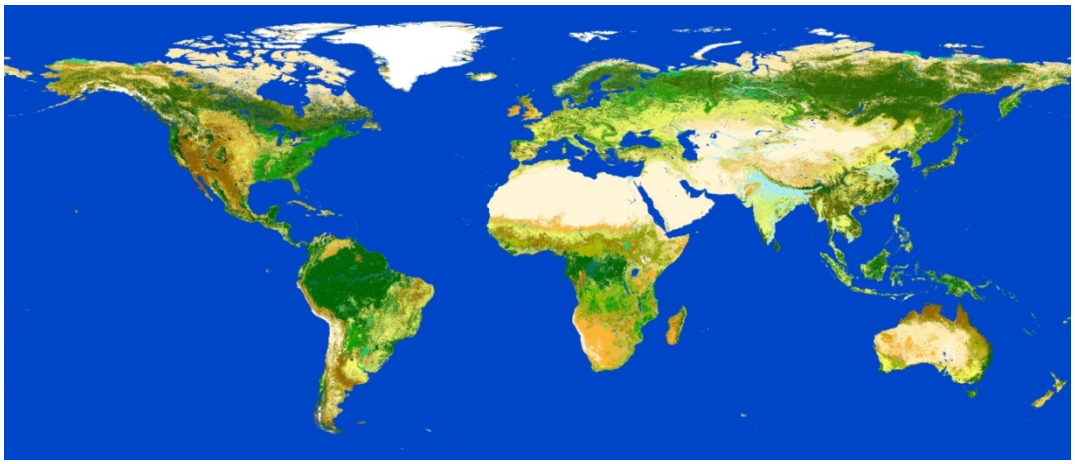


Figure 3-31: GlobCover 2009 land cover map

Table 3-14: Classes of the GlobCover 2005 and GlobCover 2009 land cover maps

1	Post-flooding or irrigated croplands	ClassID 11
2	Rainfed croplands	ClassID 14
3	Mosaic cropland (50-70%) / vegetation (grassland/shrubland/forest) (20-50%)	ClassID 20
4	Mosaic vegetation (grassland/shrubland/forest) (50-70%) / cropland (20-50%)	ClassID 30
5	Closed to open (>15%) broadleaved evergreen and/or semi-deciduous forest (>5m)	ClassID 40
6	Closed (>40%) broadleaved deciduous forest (>5m)	ClassID 50
7	Open (15-40%) broadleaved deciduous forest/woodland (>5m)	ClassID 60
8	Closed (>40%) needle-leaved evergreen forest (>5m)	ClassID 70
9	Open (15-40%) needle-leaved deciduous or evergreen forest (>5m)	ClassID 90
10	Closed to open (>15%) mixed broadleaved and needleaved forest	ClassID 100
11	Mosaic forest or shrubland (50-70%) and grassland (20-50%)	ClassID 110
12	Mosaic grassland (50-70%) and forest or shrubland (20-50%)	ClassID 120
13	Closed to open (>15%) shrubland (<5m)	ClassID 130
14	Closed to open (>15%) grassland	ClassID 140
15	Sparse (<15%) vegetation	ClassID 150
16	Closed (>40%) broadleaved forest regularly flooded, fresh water	ClassID 160
17	Closed (>40%) broadleaved semi-deciduous and/or evergreen forest regularly flooded, saline water	ClassID 170
18	Closed to open (>15%) grassland or shrubland or woody vgt on regularly flooded or waterlogged soil, fresh, brakish or saline water	ClassID 180
19	Artificial surfaces and associated areas (Urban areas >50%)	ClassID 190
20	Bare areas	ClassID 200

21	Water bodies	ClassID 210
22	Permanent Snow and Ice	ClassID 220
-	No data (burnt areas, clouds, …	ClassID 230
-	No data	ClassID 254

For the initial phase of the CCI-LC project, a simplified land cover map, which is the IGBP map (Figure 3-32), is used because of its small size and handy format. After the round-robin phase, it is planned to replace it with the GlobCover map. The IGBP map data are freely available from the IGBP Website http://www-surf.larc.nasa.gov/surf/pages/ftp-page.html). A technical description of the dataset is given in the beginning of this chapter. This surface type map includes the categories listed in Table 3-15.

Table 3-15: IGBP surface types

1 - Evergreen Needleleaf Forest	7 - Open Shrubland	13 - Urban
2 - Evergreen Broadleaf Forest	8 - Woody Savanna	14 - Crop/Natural Veg - Mosaic
3 - Deciduous Needleleaf Forest	9 - Savanna	15 - Permanent Snow/Ice
4 - Deciduous Broadleaf Forest	10 - Grassland	16 - Barren/Desert
5 - Mixed Deciduous Forest	11 - Permanent Wetland	17 - Water Bodies
6 - Closed Shrubland	12 - Cropland	18 - Tundra

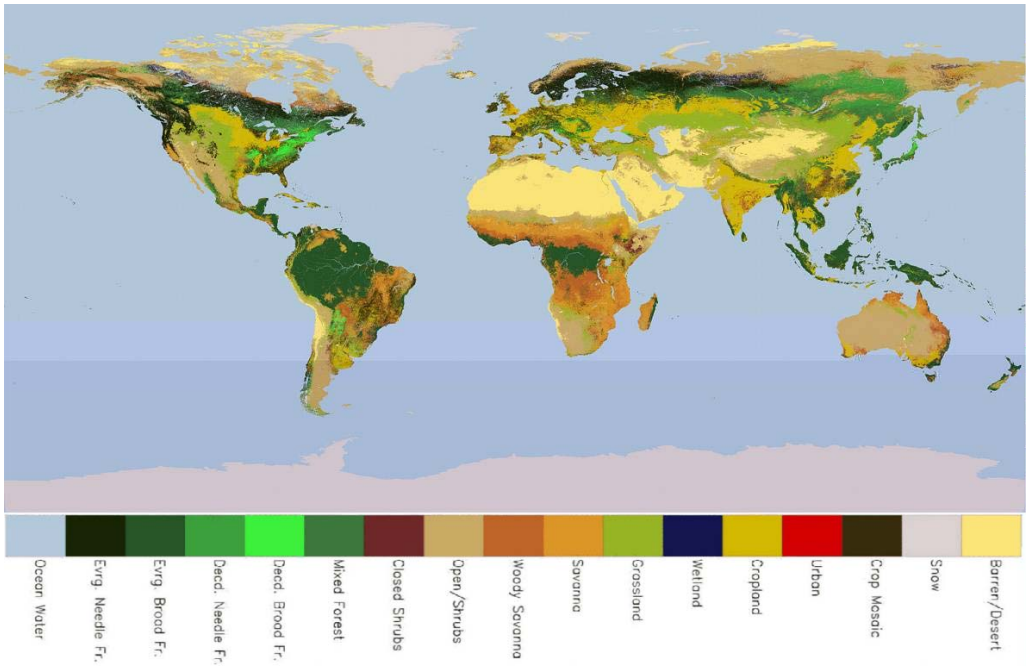


Figure 3-32: IGBP surface types map (http://modis-atmos.gsfc.nasa.gov/ECOSYSTEM/index.html).

Climate Zones

The most widely used system of climate classification is that developed by the German climatologist Köppen virtually all more recent classifications are refinements or variantions of the "Köppen system" [RD-96]. The Köppen climate classification distribution grid derived by the Food and Agriculture Organization (FAO) from the International Institute for Applied Systems Analysis (IIASA) data sets enables a determination of the general climate of any geographical entity (national or sub-national). Köppen categories are based on the annual and monthly averages of temperatures and precipitation.

Five major climatic types are recognized in this system, each type being designated by a capital letter (Figure 3-33):

A -	Moist Tropical Climates are known for their high temperatures all year round and for their large amount of annual rainfall.
B -	Dry Climates are characterized by little rain and a large daily temperature range. Two subgroups, S – semi-arid or steppe, and W - arid or desert, are used with the B climates.
C -	In Humid Middle Latitude Climates land/water differences play a large part. These climates have warm, dry summers and cool, wet winters.
D -	Continental Climates can be found in the interior regions of large land masses. Total precipitation is not very high and seasonal temperatures vary widely.
E -	Cold Climates describe this climate type perfectly. These climates are part of areas where permanent ice and tundra are always present, and only about four months of the year have above freezing temperatures.

The data are freely available from http://www.fao.org/SD/EIdirect/climate/EIsp0002.htm

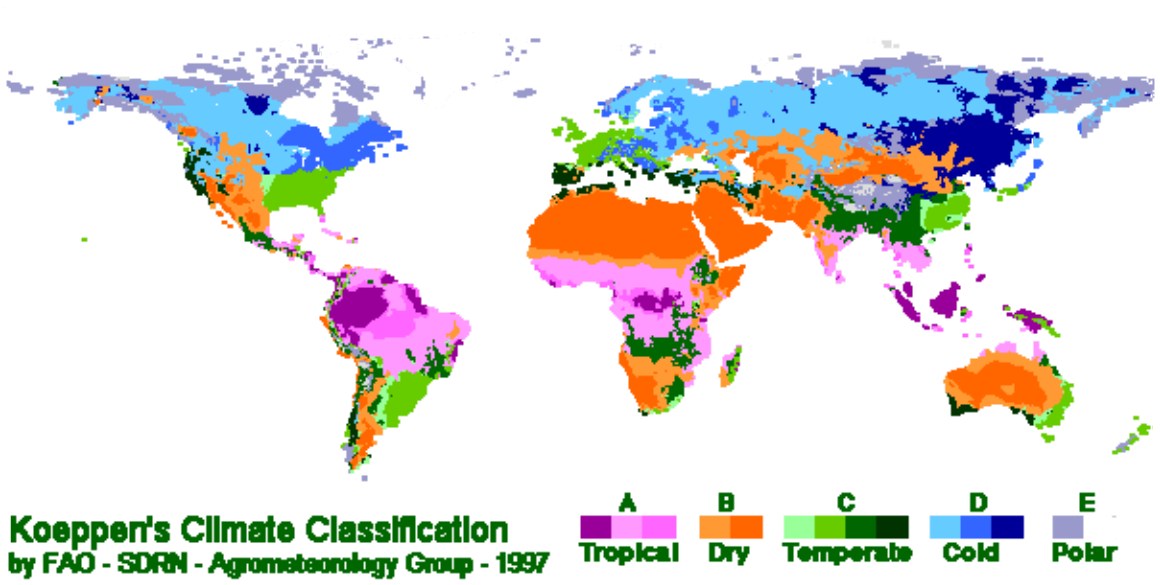


Figure 3-33: Climatological regions according to the Köppen classification (from http://www.blueplanetbiomes.org/climate.htm).

No-data gaps in the Köppen map are filled by the software with fallback values according to the following scheme based on the pixel's latitude value:

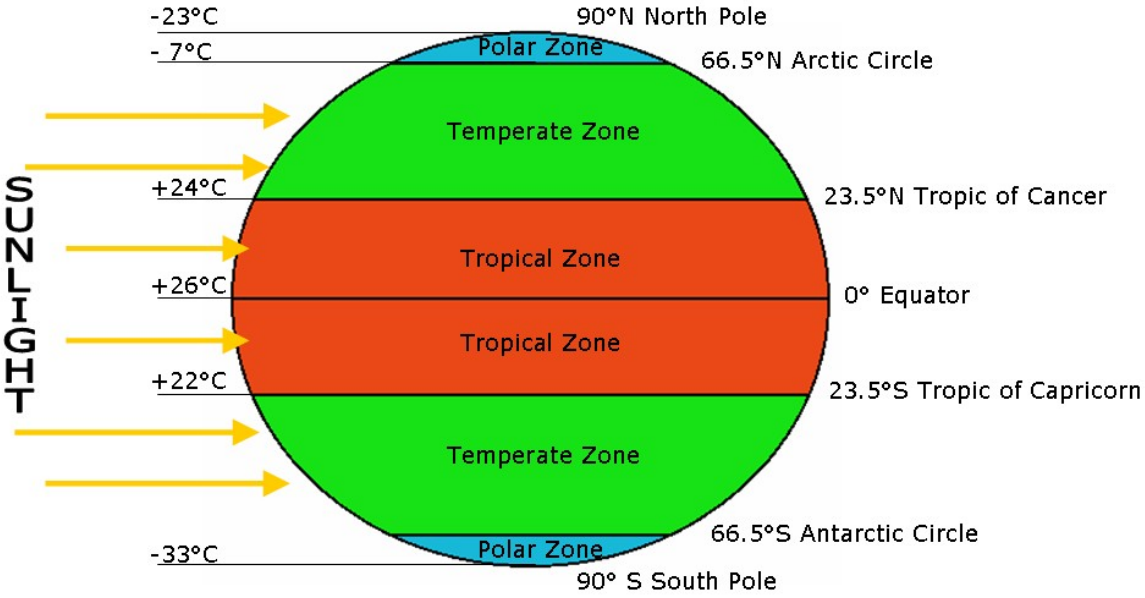


Figure 3-34: Fallback climate zone classification (from http://www.m-forkel.de/klima/klimazonen.html)

LATITUDE RANGE	REGIONS	KÖPPEN CODE
-90° … -66.5°	South Pole → Antarctic Circle	E - Polar
-66.5° … -23.5°	Antarctic Circle → Tropic of Capricorn	C - Temperate
-23.5° … +23.5°	Tropic of Capricorn → Tropic of Cancer	A - Tropical
+23.5° … +66.5°	Tropic of Cancer → Arctic Circle	C - Temperate
+66.5° … + 90.0°	Arctic Circle → North Pole	E - Polar

Seasons

In the context of the current study, climatological seasons are used. i.e., for the Northern hemisphere, spring begins by convention on 1 March, summer on 1 June, autumn on 1 September and winter on 1 December.

START DATE	DAY OF YEAR	NORTHERN	SOUTHERN
1 March	59	Spring	Autumn
1 June	151	Summer	Winter
1 September	243	Autumn	Spring
1 December	334	Winter	Summer

3.7.3.2 Atmospheric correction - aerosol and spectral directional reflectance retrieval

Aerosol

The data of the AERONET program might be used for the validation of the aerosol retrieval results. The presentation of the validation results are planned in form of the confusion matrix and the corresponding quality measures – see chapter 3.7.2.

The AERONET (AErosol RObotic NETwork) program is a federation of ground-based remote sensing aerosol networks established by NASA and PHOTONS (Univ. of Lille 1, CNES, and CNRS-INSU). The program provides a long-term, continuous and readily accessible public domain database of aerosol optical, mircrophysical and radiative properties for aerosol research and characterization, validation of satellite retrievals, and synergism with other databases. The network imposes standardization of instruments, calibration, processing and distribution.

Surface directional reflectances

The validation of the surface directional reflectances retrieval is a challenge. Three options can be following up:

- comparing with in situ data;
- model intercomparison;
- generating synthetic data: surface reflectance data transferred to TOA via radiation transfer model.

Option “in-situ”and “model intercomparison” have been used and results are included in the PVASR. The option “simulated” has not been used, but for the sake of completeness it shall be mentioned that this option can be also used for the validation.

3.7.3.3 BRDF correction - averaging of the surface directional reflectances

The validation of the BRDF correction – at present averaging of the surface directional reflectance is very difficult. The use of in situ data for the review of the averaging approach can be not proposed because the measurement and the successive retrieval of the BRDF is afflicted with different uncertainties e.g. measurement of the diffuse incident radiation. The other possibility is the use of synthetic BRDF data retrieved from canopy radiative transfer models.

Furthermore, the generated surface reflectance time series will be investigated with regard to the temporal variation of the mean and variance over selected areas (including “reference” areas such as pixels covered by dark vegetation or desert).

3.7.3.4 Multi-sensor global land cover map

The geolocation accuracy of the final products would be verified by using a set of Landsat TM images for reference. If the development of a tool for automated pattern matching of optical earth observation data is required then it can be further used to assess the geo-location accuracy of the final products.

4 LAND COVER MAP PROCESSING CHAIN

4.1 Classification logical model

The classification process transforms the L3 7-day aggregated 5°x5° surface reflectance composites produced by the pre-processing step into meaningful global land cover products.

4.1.1 Preliminary steps

Before the classification process, four preliminary steps are performed:

- the **format conversion** and **mosaicking** of the NetCDF 7-day aggregated 5°x5° surface reflectance composites produced by the pre-processing chain into GeoTiff global 7-day surface reflectance composites;
- the **generation of seasonal composites** : the spectral classification step is not directly run on 7-day products but uses as input seasonal composites (i.e. composites characterized by longer compositing periods). The seasonal compositing period is a classification parameter adjusted at the stratum level. For each equal-reasoning area, 1 or 2 periods within the year are defined to allow the optimal discrimination between the expected land cover classes;
- the **preparation of the reference land cover database**: a key auxiliary dataset in the classification chain – both for spectral and temporal classification algorithms (see section 4.2.3) – is a global reference land cover database. This database consists of a set of existing global, regional and local land cover maps which are merged together after preparatory transformations and according to specfic merging rules;
- the **preparation of the stratification layer**: before the classification process, the world is stratified in equal-reasoning areas from an ecological and a remote sensing point of view. The stratification objectives are twofold: (1) reducing the land surface reflectance variability in the dataset in order to improve the classification efficiency and (2) allowing a regional tuning of the classification parameters to take into account the regional characteristics (vegetation seasonality, cloud coverage, etc). The great but much controlled flexibility of this strategy allows defining a classification process valid at global scale while tackling both the regional heterogeneity of the land cover characteristics.

4.1.2 Classification steps

The classification process in itself is organised in 4 major processing steps (Figure 4-1), which rely on data generated in the preliminary steps and thus run independently for each equal-reasoning area with specific parameters.

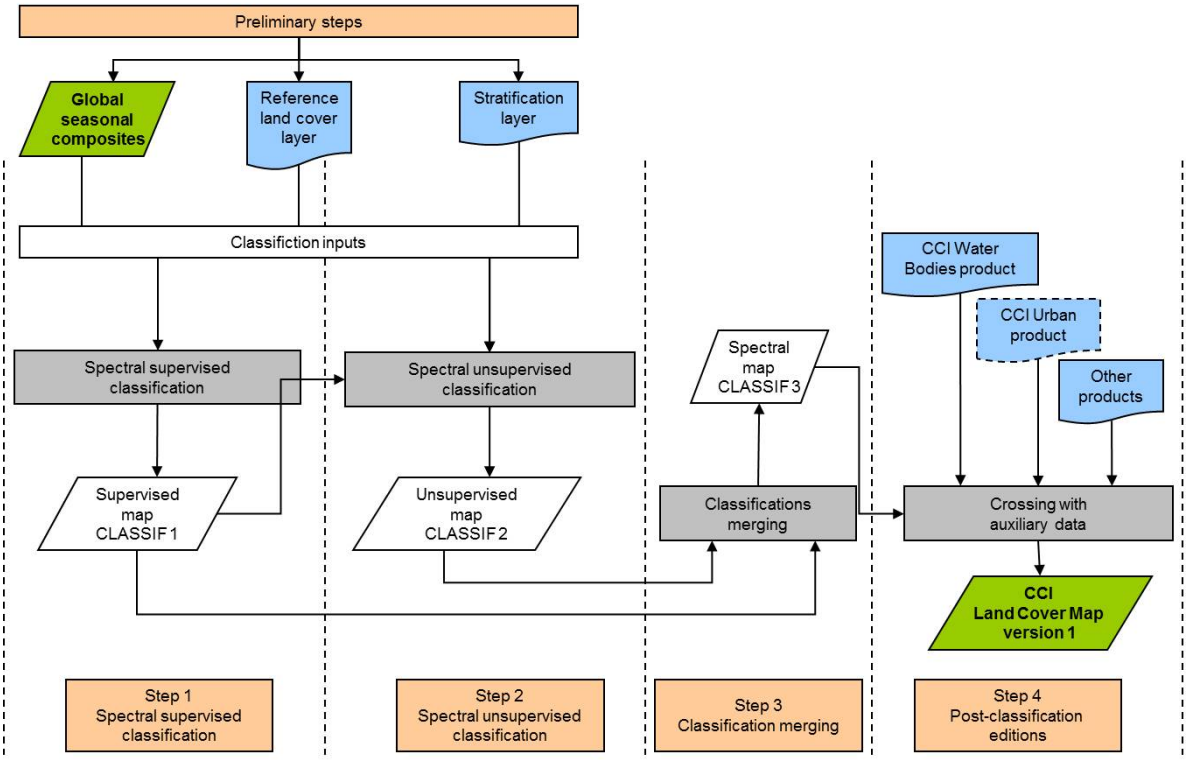


Figure 4-1: Activity diagram for the classification processing chain

In the first and second steps, supervised and unsupervised classification algorithms are run using the spectral properties of seasonal composites as input, resulting in two different maps "CLASSIF1" and "CLASSIF2". In the third step, these two maps are merged to produce a spectral map "CLASSIF3". This map is finalized in the fourth step through post-classification editions. The main one consists in adding the CCI Water Bodies product. The different algorithmic steps are presented in detail below in the sections 4.5.1, 4.5.2, 4.5.3 and 4.5.6. The classification parameters for each equal-reasoning area are included in look-up tables (LUT) associated with each step. They are also presented here after.

4.1.3 Multi-year approach

One requirement expressed by the Climate Modelling Community is the need to have successive global land cover maps stable over time. To this end, tests have been conducted during the round-robin exercises and have demonstrated the interest of using several years of Earth Observation dataset to generate a land cover map [AD 8]. The CCI-LC project will therefore deliver land cover maps which are not related to a single year but which are representative of 5-year epochs: 1998-2002, 2003-2007 and 2008-2012 [AD 4]. These 3 maps will be derived from a "baseline" land cover map generated using the full MERIS archive (i.e. 10 years of data from 2003 to 2012). These principles are illustrated in Figure 4-3.

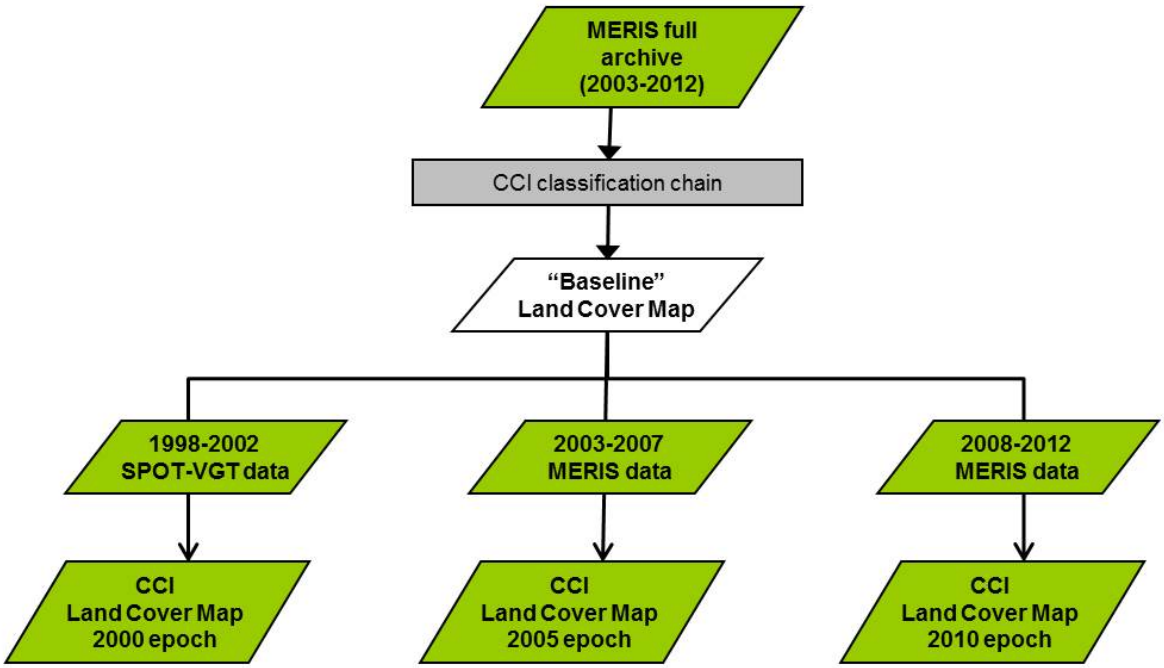


Figure 4-2: Workflow of the approach developed to ensure stable LC maps using multiple years of EO data

Two different strategies for handling multi-year dataset in the classification chain have been found efficient to increase the products accuracy and stability (Figure 4-3). First, multiple years are combined during the seasonal compositing procedure (which is one of the preliminary steps – see section 4.1.1) and the classification chain is thus run once on multi-year seasonal composites, directly resulting in a multi-year land cover map. Second, the classification chain is run once over each single year, resulting in a set of annual land cover maps. These annual maps are then combined in a multi-year land cover map according to specific aggregation rules. Either the strategy 1 or the strategy 2 is applied depending on the stratum.

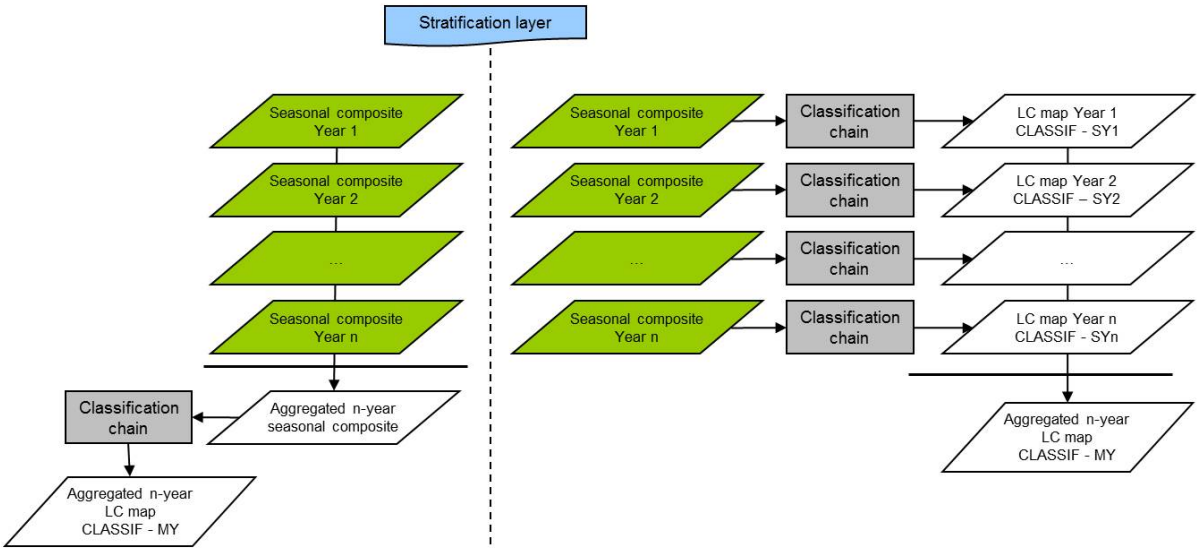


Figure 4-3: Logical flow for the multi-year approach in the classification chain

4.1.4 Multi-sensor approach

As already demonstrated through the GlobCover experiences [RD-103] and described in the ATBDv0 [AD 7], the main issue when working with MERIS FR as main input data source is the lack of acquisition. Complementing their spatial coverage is a reason that fully justifies the development of multi-sensor approaches in the CCI-LC project.

The lack of acquisitions is addressed by using images acquired in the MERIS RR mode obeying the following consideration: priority is given to the best available EO observation. In other terms, the merging is not systematic: when and where the MERIS FR coverage is high enough to allow producing accurate and stable land cover maps, MERIS RR is not used. As a result, the decision to operate the merging is made based on the quality of (Figure 4-4).

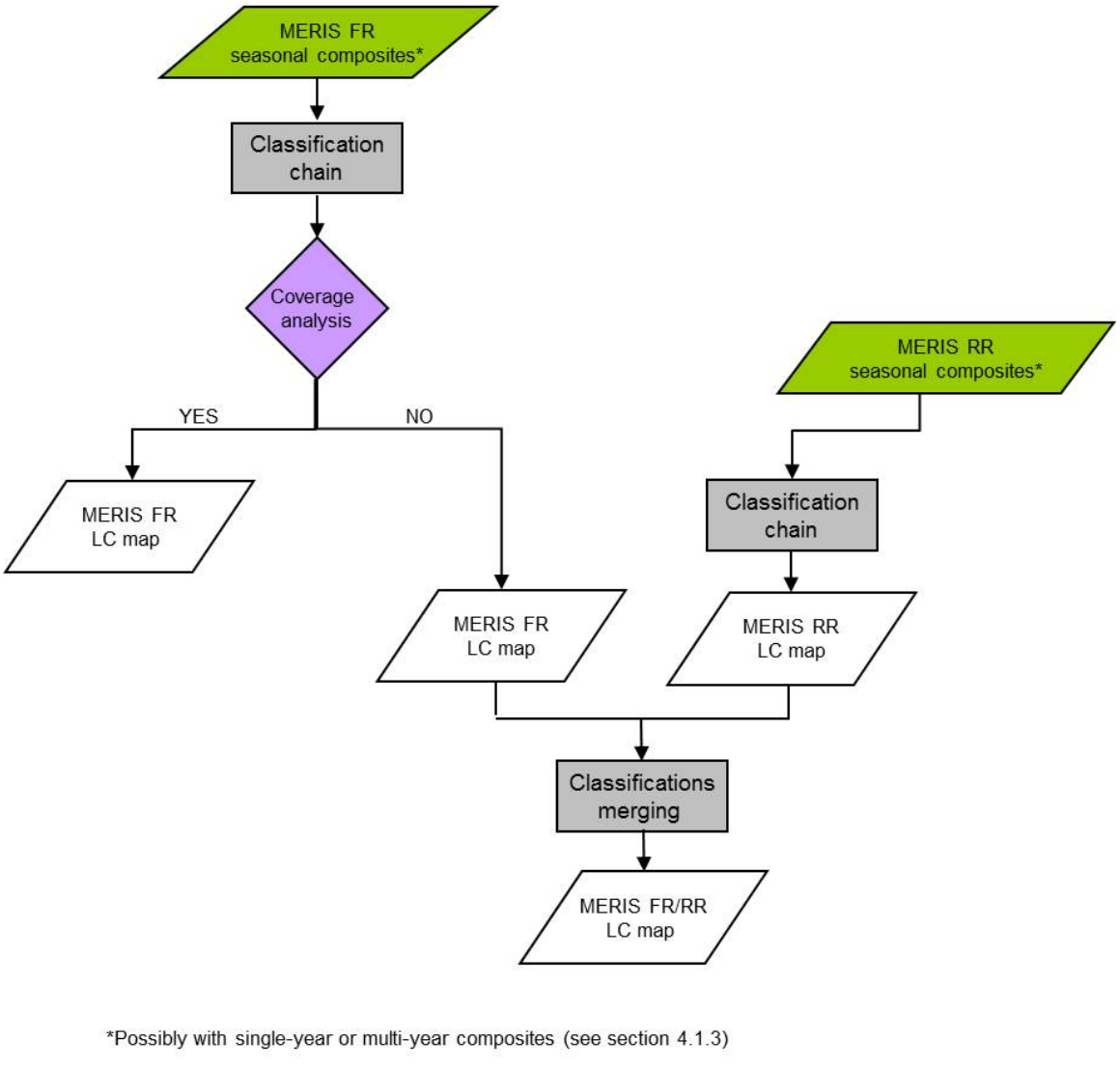


Figure 4-4: Logical flow for the multi-sensor approach developed to increase the MERIS FR coverage

4.1.5 Summary

Figure 4-5 illustrates the articulation between the classification steps (section 4.1.2) and the approaches developed to (i) handle multi-year dataset (section 4.1.3) and (ii) increase the MERIS temporal and spatial coverage (section 4.1.4).

Sections below present in detail the whole land cover mapping processing chain, starting with the preliminary steps (4.2), then presenting the methods underlying the multi-year (4.3) and multi-sensor (4.4) approaches and ending by detailing the classification steps in section 4.5.

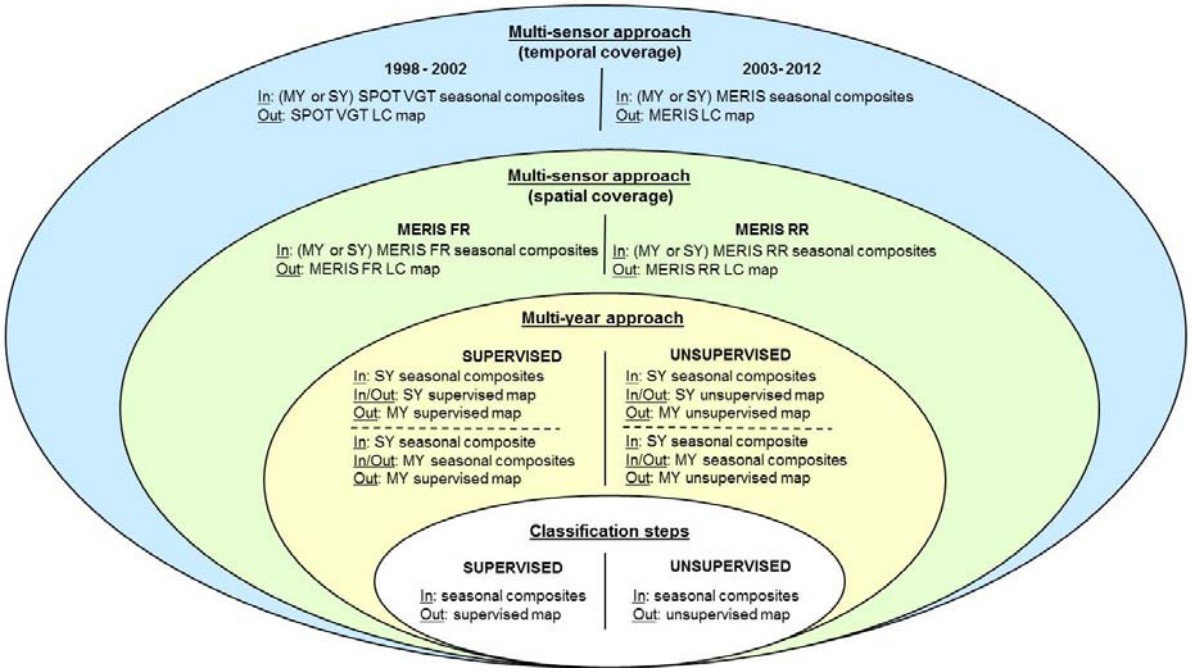


Figure 4-5: Logical flow for the articulation between the classification steps and the multi-year and multi-sensor approaches (MY = multi-year, SY = single-year, FR = full resolution, RR = reduced resolution)

4.2 Detailed processing scheme of the preliminary steps

4.2.1 Format conversion and mosaicking of pre-processing outputs

All the classification chain relies on the Geospatial Data Abstraction Library (GDAL). GDAL is a library for reading and writing raster geospatial data formats and is released under the permissive X/MIT style free software license by the Open Source Geospatial Foundation. As a library, it presents a single abstract data model to the calling application for all supported formats. It may also be built with a variety of useful command-line utilities for data translation and processing. The classification chain also makes use of the related OpenGIS Simple Features Reference Implementation (OGR) library (which is part of the GDAL source tree), which provides a similar capability for simple features vector data.

The GDAL/OGR library is considered as major free software as it ensures the compatibility with numerous commercial software packages and allows reading and writing multiple common GIS

formats. With regard to the NetCDF and GeoTiff formats required in this CCI-LC project, both can be read by GDAL. However, the library processes GeoTiff files more quickly and more efficiently than the NetCDF ones.

Since the input of the classification chain consists of 7-day surface reflectance 5°x5° NetCDF tiles, a first preliminary step of file format conversion (from NetCDF to GeoTiff) thus needs to be operated. This is done using the "GDAL_Translate" function which allows converting raster data between different formats. It uses as parameters the output format (which is GeoTiff in our case), the compression type, the image structure information and the names of the input and output files.

A second preliminary step consists in mosaicking the 5°*5° tiles in global raster files. This is achieved by running the "GDAL_Merge" function. "GDAL_Merge" allows appending several adjacent raster files together to form a unique raster file. It uses as parameters the name of the output file, the output format (which is GeoTiff by default), the compression type, the image structure information and the names of the input tiles.

Algorithm assumptions and limitations

The "GDAL_Merge" function makes the hypotheses that all tiles share the same coordinate system and have a matching number of bands. In case of overlap between tiles, the last images have priority over earlier ones.

Input and output data

Input and output data associated with these two preliminary steps are described in Table 4-1. Associated parameters are presented in Table 4-2 and Table 4-3.

Table 4-1: Input and output data for the format conversion and mosaicking procedure in the preliminary steps

DATA	DESCRIPTION	INTENT (IN, OUT, INOUT)	PHYSICAL UNIT	RANGE
L3_7d_SR_<n>_mean_tiles n = 1,..,15 (MERIS) // n = B0, B2, B3, MIR (SPOT-VGT)	Aggregated 5°*5° tiles of surface reflectance in each band of the 7-day composites NetCDF format	IN	None	[0 … 1]
L3_7d_SR_<n>_sigma_tiles n = 1,..,15 (MERIS) // n = B0, B2, B3, MIR (SPOT-VGT)	Aggregated 5°*5° tiles of uncertainties values on surface reflectance in each band of the 7-day composites NetCDF format	IN	None	[0 … 0.5]
L3_7d_NDVI_mean_tiles	Aggregated 5°*5° tiles of NDVI of the 7-day composites NetCDF format	IN	None	[-1 … 1]

DATA	DESCRIPTION	INTENT (IN, OUT, INOUT)	PHYSICAL UNIT	RANGE
L3_7d_NDVI_sigma_tiles	Aggregated 5°*5° tiles of uncertainty values on NDVI of the 7-day composites NetCDF format	IN	None	[0 … 0.5]
L3_7d_Status_tiles	Aggregated 5°*5° tiles of pixel status associated with the surface reflectance in the 7-day composites NetCDF format	IN	None	[1 … 7]
L3_7d_ValidCount_tiles	Aggregated 5°*5° tiles counting the contributing observations in the 7-day composites NetCDF format	IN	None	[1 … 150]
L3_7d_LandCount_tiles	Aggregated 5°*5° tiles counting the number of land observations in the 7-day composites NetCDF format	IN	None	[0 … 150]
L3_7d_WaterCount_tiles	Aggregated 5°*5° tiles counting the number of water observations in the 7-day composites NetCDF format	IN	None	[0 … 150]
L3_7d_SnowIceCount_tiles	Aggregated 5°*5° tiles counting the number of snow and ice observations in the 7-day composites NetCDF format	IN	None	[0 … 150]
L3_7d_CloudCount_tiles	Aggregated 5°*5° tiles counting the number of cloud observations in the 7-day composites NetCDF format	IN	None	[0 … 150]
L3_7d_CloudShadowCount_tiles	Aggregated 5°*5° tiles counting the number of cloud shadow observations in the 7-day composites NetCDF format	IN	None	[0 … 150]

DATA	DESCRIPTION	INTENT (IN, OUT, INOUT)	PHYSICAL UNIT	RANGE
L3_7d_SR_<n>_mean n = 1,..,15 (MERIS) // n = B0, B2, B3, MIR (SPOT-VGT)	Global raster of surface reflectance in each band of the 7-day composites GeoTiff format	OUT	None	[0 … 1]
L3_7d_SR_<n>_sigma n = 1,..,15 (MERIS) // n = B0, B2, B3, MIR (SPOT-VGT)	Global raster of uncertainties values on surface reflectance in each band of the 7-day composites GeoTiff format	OUT	None	[0 … 0.5]
L3_7d_NDVI_mean	Global raster of NDVI of the 7-day composites GeoTiff format	OUT	None	[-1 … 1]
L3_7d_NDVI_sigma	Global raster of uncertainty values on NDVI of the 7-day composites GeoTiff format	OUT	None	[0 … 0.5]
L3_7d_Status	Global raster of pixel status associated with the surface reflectance in the 7-day composites GeoTiff format	OUT	None	[1 … 7]
L3_7d_ValidCount	Global raster counting the contributing observations in the 7-day composites GeoTiff format	OUT	None	[1 … 150]
L3_7d_LandCount	Global raster counting the number of land observations in the 7-day composites GeoTiff format	OUT	None	[0 … 150]
L3_7d_WaterCount	Global raster counting the number of water observations in the 7-day composites GeoTiff format	OUT	None	[0 … 150]
L3_7d_SnowIceCount	Global raster counting the number of snow and ice observations in the 7-day composites GeoTiff format	OUT	None	[0 … 150]

DATA	DESCRIPTION	INTENT (IN, OUT, INOUT)	PHYSICAL UNIT	RANGE
L3_7d_CloudCount	Global raster counting the number of cloud observations in the 7-day composites GeoTiff format	OUT	None	[0 ... 150]
L3_7d_CloudShadowCount	Global raster counting the number of cloud shadow observations in the 7-day composites GeoTiff format	OUT	None	[0 ... 150]

Table 4-2: Parameters needed for running the format conversion procedure in the preliminary steps

PARAMETERS	DESCRIPTION	INTENT (IN, OUT, INOUT)	FORMAT	RANGE
Format	Output format	IN	String	/
Compress	Compression type	IN	String	/
Interleave	Image structure information	IN	String	/
Tiled	Image structure information	IN	String	/
Input	Name of the input file	IN	String	/
Output	Name of the output file	IN	String	/

Table 4-3: Parameters needed for running the mosaicking procedure in the preliminary steps

PARAMETERS	DESCRIPTION	INTENT (IN, OUT, INOUT)	FORMAT	RANGE
Output	Name of the output file	IN	String	/
Compress	Compression type	IN	String	/
Interleave	Image structure information	IN	String	/
Tiled	Image structure information	IN	String	/
Tiles	Name of the input tiles	IN	String	/

4.2.2 Generation of seasonal composites

Global 7-day composites are composited into seasonal composites (i.e. composites characterized by longer compositing periods ranging from 1.5 months to 1 year) using the Mean Compositing (MC) algorithm [RD-72].

The compositing process aims at minimizing the cloud contamination and reducing variations in reflectance values due to image acquisition with varying geometries [RD-136]. The MC algorithm averages quality controlled reflectance values over the compositing period. It has proved to significantly reduce the BRDF and atmospheric artefacts while resulting in spatially homogeneous cloud-free composites with good radiometric consistency even in humid tropical regions [RD-72, RD-134 and RD-135]. In addition, MC algorithm does not require any model adjustment or additional parameterizations, hence contributing to the generalization and automation of the process.

The MC algorithm uses as input a set of superimposed 7-day composites included in a given time interval referred to as "compositing period". The quality control of the reflectance values to composite has already been achieved in the pre-processing chain when generating the 7-day composites (see section 3.5). In this step of seasonal composites generation, the aggregation rule relies on the status associated with each pixel. On one hand, the aggregation is operated on pixels associated with a same status. On the other hand, rules exist to define the priorities between statuses. Priority is given to the "land" status and then, come the "snow", "water" and "cloud" statuses in this sequence. The average is weighted by the number of observations associated with the selected status.

Based on these two principles, reflectance values from the 7-day composites are averaged by pixel and by band to generate the seasonal composite, as indicated in Figure 4-6.

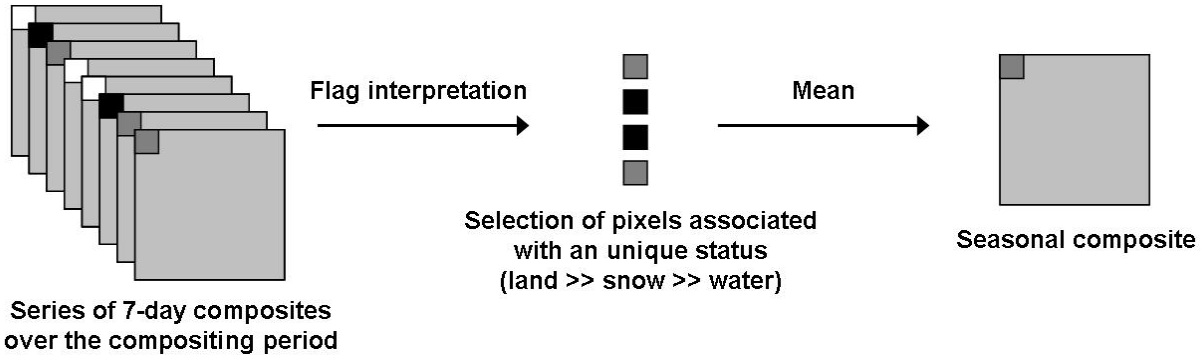


Figure 4-6: Mean compositing workflow to generate seasonal composites starting from 7-day composites

The pixel status associated with the surface reflectance in the seasonal composite is recorded in a dedicated layer, as well as the contributing number of observations over the aggregation period.

Algorithm assumptions and limitations

None: the same processing chain is applicable to any type of satellite data and for any length of compositing period.

Input and output data

Input and output data associated with this compositing process are described in Table 4-4. Associated parameters are presented in Table 4-5.

Table 4-4: Input and output data of the seasonal composites generation in the preliminary steps

DATA	DESCRIPTION	INTENT (IN, OUT, INOUT)	PHYSICAL UNIT	RANGE
L3_7d_StartDate_EndDate_ SR_<n>_mean n = 1,..,15 (MERIS) // n = B0, B2, B3, MIR (SPOT-VGT)	Global raster of surface reflectance in each band of the 7-day composites GeoTiff format	IN	None	[0 … 1]
L3_7d_StartDate_EndDate_ NDVI_mean	Global raster of NDVI of the 7-day composites GeoTiff format	IN	None	[-1 … 1]
L3_StartDate_EndDate_ 7d_Status	Global raster of pixel status associated with the surface reflectance in the 7-day composites GeoTiff format	IN	None	[1 ... 7]
L3_7d_StartDate_EndDate_ ValidCount	Global raster counting the contributing observations in the 7-day composites GeoTiff format	IN	None	[1…150]
LUT 1	Look-Up-Table describing the parameters for the seasonal compositing process (see Table 4-5)	IN	/	/
L3_<sensor>_StartDate_EndDate_SR_<n> n = 1,..,15 (MERIS) // n = B0, B2, B3, MIR (SPOT-VGT) sensor = FR for MERIS FR, RR for MERIS RR, VGT for SPOT-VGT	Global raster of surface reflectance in each bands of the seasonal composites GeoTiff format	OUT	None	[0 … 1]
L3_<sensor>_StartDate_EndDate_NDVI sensor = FR for MERIS FR, RR for MERIS RR, VGT for SPOT-VGT	Global raster of NDVI of the seasonal composites GeoTiff format	OUT	None	[-1 … 1]
L3_<sensor>_StartDate_EndDate_Status sensor = FR for MERIS FR, RR for MERIS RR, VGT for SPOT-VGT	Global raster of pixel status associated with the surface reflectance in the seasonal composites GeoTiff format	OUT	None	[1 ... 7]

DATA	DESCRIPTION	INTENT (IN, OUT, INOUT)	PHYSICAL UNIT	RANGE
L3_<sensor>_StartDate_EndDate_NOBS sensor = FR for MERIS FR, RR for MERIS RR, VGT for SPOT-VGT	Global raster counting the contributing observations in the seasonal composites GeoTiff format	OUT	None	[0 ... 500]

Table 4-5: Parameters needed in the seasonal composites generation in the preliminary steps (seasonal composites generation) contained in LUT 1

PARAMETERS	DESCRIPTION	INTENT (IN, OUT, INOUT)	FORMAT	RANGE
StartDate_Ni	Number of the 7-day composite which marks the start of the seasonal composite i	IN	Short	[1 ... 52]
StartDate_Si	Date which marks the start of the seasonal composite i	IN	Date (YYYYMMDD)	
EndDate_Ni	Number of the 7-day composite which marks the end of the seasonal composite i	IN	Short	[1 ... 52]
EndDate_Ni	Date which marks the end of the seasonal composite i	IN	Date (YYYYMMDD)	

4.2.3 Preparation of the reference land cover database

A key auxiliary dataset in the classification chain, for the spectral algorithms, is a global reference land cover database. The reference database consists of a set of existing global, regional and local land cover maps which have been selected as the most accurate ones available for a given region, with the highest spatial resolution and with a CCI-compatible legend.

Each dataset is re-projected, re-sampled and translated into the CCI-LC legend, generating a set of GeoTiff intermediate products. These operations are performed with the ArcGIS software, using three different toolboxes: “Project_Raster”, “Resample” and “Reclassify”. The generated intermediate products are finally merged, using the “Raster_Calculator” toolbox, to constitute the final reference land cover database (Figure 4-7). The set of original maps and of intermediate products are documented in the IODD [AD 11].

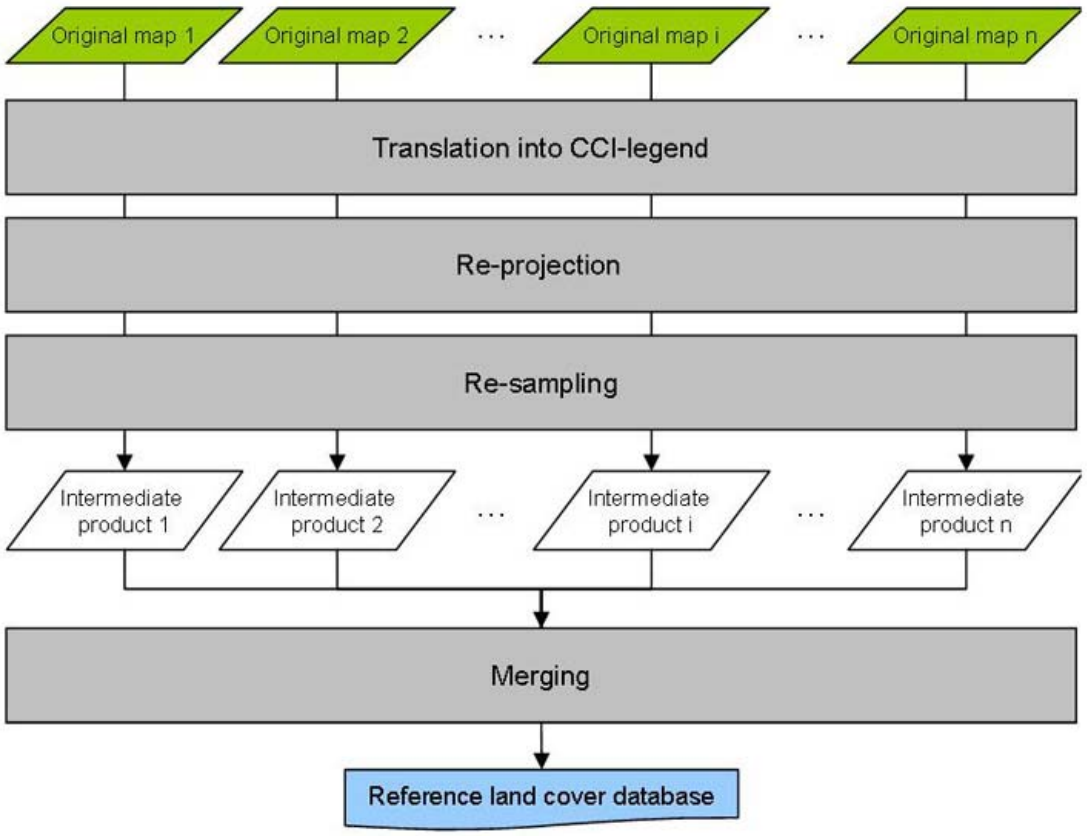


Figure 4-7: Logical flow for generation of the reference land cover database

The "Reclassify" toolbox changes the values in a raster. It uses as parameters the field denoting the values that will be reclassified, the conversion table specifying how to reclassify values of the input raster and the information denoting whether missing values in the conversion table retain their value or get mapped to NoData.

The "Project_Raster" toolbox transforms a raster dataset from one projection to another. It uses as parameters the output raster name and format (which is GeoTiff in our case), the output coordinate system (which is a Plate-Carrée in our case) and the transformation method used between two geographic systems or datums.

The "Resample" toolbox alters the raster dataset by changing the cell size and resampling method. It uses as parameters the output raster name and format (which is GeoTiff in our case), the cell size for the new raster dataset and the resampling algorithm to be used.

The "Raster_Calculator" toolbox builds and executes a single map algebra expression using Python syntax in a calculator-like interface. It uses as parameters the map algebra expression allowing combining the different maps in a single final one.

Input and output data

Input and output data associated with the "Reclassify", "Project_Raster" and "Resample" toolboxes are presented in Table 4-6. Associated parameters are presented in Table 4-7.

Table 4-6: Input and output data for the "Reclassify", "Project_Raster" and "Resample" toolboxes in the preliminary step preparing the reference land cover database

DATA	DESCRIPTION	INTENT (IN, OUT, INOUT)	PHYSICAL UNIT	RANGE
OriginalMap_<n>	Original land cover map as a raster file	IN	None	[0 … 255]
OriginalMap_<n>_CCI	Original land cover map translated into the CCI LCCS legend as a GeoTiff raster file	INOUT	None	[0 … 255]
OriginalMap_<n>_CCI_WGS84	Original land cover map translated into the CCI LCCS legend and projected in the Plate-Carrée projection as a GeoTiff raster file	INOUT	None	[0 … 255]
OriginalMap_<n>_CCI_WGS84_300m	Original land cover map translated into the CCI LCCS legend, projected in the Plate-Carrée projection and resampled at 300m as a GeoTiff raster file	OUT	None	[0 … 255]

Table 4-7: Parameters needed for running the "Reclassify", "Project_Raster" and "Resample" toolboxes in the preliminary step preparing the reference land cover database

PARAMETERS	DESCRIPTION	INTENT (IN, OUT, INOUT)	FORMAT	RANGE
Reclass_Field	Field containing the raster values to modify	IN	String	/
Remap	Table associating old raster values to new values	IN	String	/
Missing_Values	Information about how dealing with missing values in the reclassification table	IN	Boolean	[DATA, NODATA]
Out_coor_system	Output coordinate System	IN	String	/
Geographic_transform	Transformation method used between two geographic systems or datums (optional when the input and output coordinate systems have the same datum)	IN	String	/
Cell_size	Cell size for the new raster dataset	IN	Short	/
Resampling_Type	Resampling algorithm to be used	IN	String	

PARAMETERS	DESCRIPTION	INTENT (IN, OUT, INOUT)	FORMAT	RANGE
Out_raster	Output raster name and format	IN	String	/

Input and output data associated with the "Raster_Calculator" toolbox are presented in Table 4-8. Associated parameters are presented in Table 4-9.

Table 4-8: Input and output data for the "Raster_Calculator" toolbox in the preliminary step preparing the reference land cover database

DATA	DESCRIPTION	INTENT (IN, OUT, INOUT)	PHYSICAL UNIT	RANGE
OriginalMap_<n>_CCI_WGS84_300m	Original land cover maps translated into the CCI LCCS legend, projected in the Plate-Carrée projection and resampled at 300m as a GeoTiff raster file	IN	None	[0 … 255]
ReferenceLC_CCI	Reference land cover database, as a GeoTiff raster file	OUT	None	[0 … 255]

Table 4-9: Parameters needed for running the "Raster_Calculator" toolbox in the preliminary step preparing the reference land cover database

PARAMETERS	DESCRIPTION	INTENT (IN, OUT, INOUT)	FORMAT	RANGE
Expression	Map Algebra expression defining how to combine the different raster files used as input	IN	String	/

The LUT 2 describing the CCI land cover legend based on the Land Cover Classification System (LCCS) is presented hereafter in Table 4-10. It includes the parameters describing the CCI LCCS land cover legend and contains the following fields:

- **NB_LAB**, which lists the numbers (IDs) ranging between 0 and 255 and corresponding to each LCCS land cover class;
- **LABEL**, which gives the names of each LCCS land cover class;
- **R, G and B**, which provide the red, green and blue components (from 0 to 255) of the color codes associated with each LCCS land cover class.

Table 4-10: Parameters describing the CCI LCCS land cover legend (contained in LUT 2)

NB_LAB	LABEL	R	G	B
0	No data	0	0	0
10	Cropland, rainfed	255	255	100
11	Cropland, rainfed, herbaceous cover	255	255	100
12	Cropland, rainfed, shrub and tree cover	255	255	0
20	Cropland, irrigated or post-flooding	170	240	240
30	Mosaic cropland (>50%) / natural vegetation (tree, shrub, herbaceous cover) (<50%)	220	240	100
40	Mosaic natural vegetation (tree, shrub, herbaceous cover) (>50%) / cropland (< 50%)	200	200	100
50	Tree cover, broadleaved, evergreen, closed to open (>15%)	0	100	0
60	Tree cover, broadleaved, deciduous, closed to open (> 15%)	0	160	0
61	Tree cover, broadleaved, deciduous, closed (>40%)	0	160	0
62	Tree cover, broadleaved, deciduous, open (15-40%)	170	200	0
70	Tree cover, needleleaved, evergreen, closed to open (> 15%)	0	60	0
71	Tree cover, needleleaved, evergreen, closed (>40%)	0	60	0
72	Tree cover, needleleaved, evergreen, open (15-40%)	0	80	0
80	Tree cover, needleleaved, deciduous, closed to open (> 15%)	40	80	0
81	Tree cover, needleleaved, deciduous, closed (>40%)	40	80	0
82	Tree cover, needleleaved, deciduous, open (15-40%)	40	100	0
90	Tree cover, mixed leaf type (broadleaved and needleleaved)	120	130	0
100	Mosaic tree and shrub (>50%) / herbaceous cover (< 50%)	140	160	0
110	Mosaic herbaceous cover (>50%) / tree and shrub (<50%)	190	150	0
120	Shrubland	150	100	0
130	Grassland	255	180	0
140	Lichens and mosses	255	180	0
150	Sparse vegetation (tree, shrub, herbaceous cover)	255	235	175
160	Tree cover, flooded, fresh or brakish water	0	120	90
170	Tree cover, flooded, saline water	0	150	120
180	Shrub or herbaceous cover, flooded, fresh-saline or brakish water	0	220	130
190	Urban	195	20	0
200	Bare areas	255	245	215
210	Water	0	70	200

NB_LAB	LABEL	R	G	B
220	Snow and ice	255	255	100

4.2.4 Preparation of the stratification layer

Using an a priori stratification in the land cover mapping processing chain allows increasing the classification algorithms performance but, at the same time, can also induce local artefacts in the final map [AD 8]. It has therefore to be used with cautions. The location of strata limits has to be decided in order to separate regions characterized by different climatic conditions, seasonal behaviours and remote sensing conditions. Figure 4-8 illustrates the GlobCover stratification layer, in which strata were delineated manually and which is used as starting point in this project.

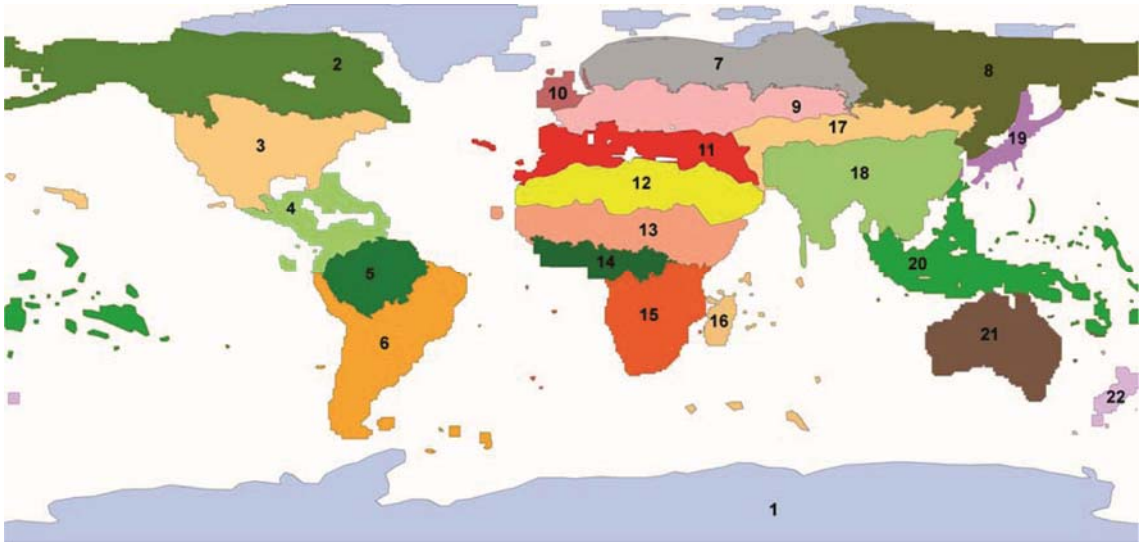


Figure 4-8: GlobCover stratification, made of 22 equal-reasoning areas

The LUT 3 describing the 22 strata included in the GlobCover stratification layer is presented hereafter in Table 4-11. It contains the following fields:

- **NB_ST**, which lists the numbers (IDs) ranging between 1 and 22 and corresponding to each equal-reasoning area;
- **Name_ST**, which gives the names of each equal-reasoning area.

Table 4-11: Parameters describing the stratification layer (contained in LUT 3)

NB_ST	NAME_ST
1	Polar areas
2	Canada
3	US
4	Central-America
5	Amazon

NB_ST	NAME_ST
6	South-America
7	North-west Eurasia
8	North-east Eurasia
9	Eurasia
10	England
11	Mediterranea
12	Africa desert
13	North-Africa
14	Central Africa
15	South-Africa
16	Madagascar
17	Asia desert
18	South-east Asia
19	Japan
20	Indonesia
21	Australia
22	New-Zeland

The CCI-LC stratification layer is designed to be an improved version of the GlobCover stratification layer in the sense that some strata limits have been adjusted to make them less visible in the final map[3].

GlobCover strata limits (i.e. strata limits delineated manually) are retained over sharp interfaces and over bare areas. Where strata limits need to be located in regions with continuous vegetation, a segmentation procedure is applied to exactly locate the limit according to the spectral, temporal and textural properties of the image to classify. The segmentation procedure is performed using the e-Cognition software. For each strata limit, two successive segmentation algorithms are applied to the seasonal composites that have been defined to classify the two strata of interest.

First, a multiresolution segmentation algorithm is applied, which is based on a bottom-up region-merging technique [RD-99]. One-pixel objects are grouped according to an optimization function that specifies h_{scale}, the maximal admissible within-object heterogeneity in terms of spectral and spatial properties (eq. 4-1):

[3] In this current version (land cover maps Epoch 1 V1.1, Epoch 2 V1.0 and Epoch 3 V1.0), no refinement has been made for the stratification. We used the same approach than in GlobCover. It is planned to improve this step in a future update of the products.

$$h_{scale} \geq w_{spectral} \,.\, h_{spectral} + (1 - w_{spectral}) \,.\, h_{spatial}$$

where $h_{spectral}$ measures the spectral variability of the object; $h_{spatial}$ characterizes the object shape and $w_{spectral}$ inversely weights $h_{spectral}$ and $h_{spatial}$. The merging procedure continues until the within-object heterogeneity exceeds the user-defined threshold h_{scale}.

The multiresolution segmentation algorithm uses as main parameter the scale parameter, which is a kind of homogeneity criterion. It is indeed used to determine the upper limit for a permitted change of heterogeneity throughout the segmentation process. The scale parameter determines the average image object size higher values for the scale parameter resulting in larger image objects and smaller values in smaller image objects).

The homogeneity criterion of the multiresolution segmentation algorithm measures how homogeneous or heterogeneous an image object is within itself. The homogeneity criterion is calculated as a combination of color and shape properties of both the initial and the resulting image objects of the intended merging. The color homogeneity is based on the standard deviation of the spectral colors and the shape homogeneity is based on the deviation of a compact or a smooth shape. The homogeneity criterion can be customized by weighting shape and compactness criteria, which are two additional parameters to define.

A spectral difference segmentation algorithm is then applied, on the objects defined by the multi-resolution segmentation process. This segmentation algorithm merge neighboring image objects if the difference between their layer mean intensities is below the value given by the maximum spectral difference threshold. It is designed to refine existing segmentation results, by merging spectrally similar image objects produced by previous segmentations and therefore is a bottom-up segmentation. It uses as parameter the maximum spectral difference threshold.

Input and output data

Input and output data associated with the segmentation procedure are presented in Table 4-12. Associated parameters (LUT 4) are presented in Table 4-13. The seasonal compositing periods associated with each stratum in the classification chain (LUT 5) are presented in Table 4-14.

Table 4-12: Input and output data for the segmentation procedure run in the preliminary step preparing the stratification layer

DATA	DESCRIPTION	INTENT (IN, OUT, INOUT)	PHYSICAL UNIT	RANGE

L3_<sensor>_StartDate_EndDate_SR_<n>
n = 1,..,15 (MERIS) // n = B0, B2, B3, MIR (SPOT-VGT)
sensor = FR for MERIS FR, RR for MERIS RR, VGT for SPOT-VGT
	Global raster of surface reflectance in each bands of the seasonal composites
	GeoTiff format
		IN	None	[0 … 1]

LUT 4	LUT specifying the input parameters of the segmentation procedure (see Table 4-13)	IN	None	[0 … 1]

DATA	DESCRIPTION	INTENT (IN, OUT, INOUT)	PHYSICAL UNIT	RANGE
LUT 5	LUT specifying the seasonal composites to use in the segmentation procedure (those defined to classify the two strata of interest) (see Table 4-14)	IN	None	[0 … 1]
Stratification_ST<a>_ST<b> a and b being the 2 strata ID to separate	Stratification layer, as a shapefile	OUT	None	[0 … 255]

Table 4-13: Parameters needed for the segmentation procedure run in the preliminary step preparing the stratification layer (contained in LUT 4)

PARAMETERS	DESCRIPTION	INTENT (IN, OUT, INOUT)	FORMAT	RANGE
Limit	List of all possible limits between adjacent strata	IN	String	/
NB_STa	ID of the first adjacent stratum (see LUT 3 presented in Table 4-11)	IN	Short	[1 … 22]
NB_STb	ID of the second adjacent stratum (see LUT 3 presented in Table 4-11)	IN	Short	[1 … 22]
Segm	Stratification method (manual or segmentation-based) for each strata limit, expressed as a boolean expression (1 indicating that a segmentation needs to be applied)	IN	Boolean	[0 1]
SP (for Scale Parameter)	Homogeneity criterion used to determine the upper limit for a permitted change of heterogeneity throughout the multiresolution segmentation process	IN (multiresolution)	Short	/
Shape criterion	Shape homogeneity, based on the deviation of a compact or a smooth shape. The definition of this parameter automatically defines the color property since color and shape criteria shall sum up to 1.	IN (multiresolution)	Short	[0 … 1]
Compactness criterion	Refinement of the shape criterion, by defining its compact character. It automatically defines its smooth character since compactness and smoothness criteria shall sum up to 1	IN (multiresolution)	Short	[0 … 1]
W1	Specifies the weight of the different spectral channels in the multiresolution segmentation procedure	IN (multiresolution)	Short	[0 … 1]

PARAMETERS	DESCRIPTION	INTENT (IN, OUT, INOUT)	FORMAT	RANGE
SD (for Spectral Difference)	Threshold value used which guides the merging between neighboring image objects in the spectral difference segmentation process	IN (spectral difference)	Short	/
W2	Specifies the weight of the different spectral channels in the multiresolution segmentation procedure	IN (spectral difference)	Short	[0 … 1]

Table 4-14: Seasonal compositing periods associated with each stratum, used in the classification chain (contained in LUT 5)

PARAMETERS	DESCRIPTION	INTENT (IN, OUT, INOUT)	FORMAT	RANGE
NB_ST	Number of the stratum (see Table 4-11, LUT 3)	IN	Short	[1 … 22]
Name_ST	Name of the stratum (see Table 4-11, LUT 3)	IN	String	/
StartDate _Si	Start date of the seasonal composite(s) to be used, i being the composite number	IN	Date (YYYYMMDD)	19980101 … 20121231
EndDate_Si	End date of the seasonal composite(s) to be used, i being the composite number	IN	Date (YYYYMMDD)	19980101 … 20121231

4.3 Detailed processing scheme of the multi-year approach

As introduced in the classification logical models (4.1.3), it has been decided to use several years of Earth Observation dataset to generate a land cover map with the aim of increasing the products accuracy and stability. More precisely, the CCI-LC project will deliver land cover maps representative of 5-year epochs: 1998-2002, 2003-2007 and 2008-2012 [AD 4]. The developed methodology plans to derive these 3 maps from a "baseline" land cover map generated using the full MERIS archive (i.e. 10 years of data from 2003 to 2012).

Two different strategies have been developed to integrate multiple years of EO data in the land cover chain (see Figure 4-3 in section 4.1.3).

The first strategy (hereafter referred to as "MY_S1" for "Multi-Year Strategy 1") makes use of the multi-year dataset to increase the quality of the composites to classify. Several years are combined during the seasonal compositing procedure and each multi-year seasonal composite then serves of input to the classification chain, which is run only once to directly produce the multi-year land cover product. In this way, the "CLASSIF1" and "CLASSIF2" in Figure 4-1 thus stand for a multi-year product "CLASSIF1_MY" and "CLASSIF2_MY".

In the second strategy (hereafter referred to as "MY_S2" for "Multi-Year Strategy 2"), the multiple years are combined at the end of the classification chain, at the level of the land cover products. The supervised and unsupervised classification algorithms are run once for each year of interest, resulting

in multiple annual land cover product (CLASSIF1_SY and CLASSIF2_SY in Figure 4-1). The multiple single-year products are then combined in a multi-year land cover map "CLASSIF1_MY" and "CLASSIF2_MY" according to specific aggregation rules.

According to the stratum, either strategy 1 or 2 is applied (Table 4-15). Indeed, each equal-reasoning area represents specific climatic conditions, seasonal behaviours and remote sensing conditions and this influence the performance of both strategies. The main advantage of strategy 1 is to increase the quality of the composites to classify. It is therefore applied in strata where the data coverage and quality is poorer due to rather long snow and cloud periods. Conversely, in regions where the data coverage and quality is not problematic, the strategy 2 is applied since it allows better accounting for the inter-annual variability.

Table 4-15: Definition of the multi-year strategy for the spectral classification (contained in LUT 6)

NB_ST	NAME_ST	MY_STRATEGY
1	Polar areas	MY_S1
2	Canada	MY_S1
3	US	MY_S2
4	Central-America	MY_S1
5	Amazon	MY_S1
6	South-America	MY_S2
7	North-west Eurasia	MY_S1
8	North-east Eurasia	MY_S1
9	Eurasia	MY_S2
10	England	MY_S2
11	Mediterranea	MY_S2
12	Africa desert	MY_S2
13	North-Africa	MY_S2
14	Central Africa	MY_S1
15	South-Africa	MY_S2
16	Madagascar	MY_S1
17	Asia desert	MY_S2
18	South-east Asia	MY_S2
19	Japan	MY_S1
20	Indonesia	MY_S1
21	Australia	MY_S2

NB_ST	NAME_ST	MY_STRATEGY
22	New-Zeland	MY_S2

For strata concerned by the MY_S1, an additional preliminary step is required, which consists in generating the multi-year seasonal composites. For strata concerned by the MY_S2, there is no additional preliminary step but the classification algorithms have to be run multiple times (one for each year of interest) and the multiple single-year land cover maps have to be aggregated in a multi-year land cover map. These two procedures are detailed hereafter.

These operations are run for 4 multi-year periods: the 1998-2002 period for generating the 2000 epoch map (using SPOT-VGT data), the 2003-2007 period for generating the 2005 epoch map (using MERIS data), the 2008-2012 period for generating the 2010 epoch map (using MERIS data) and the 2003-2012 period for generating the baseline map.

4.3.1 MY_S1 – Generation of multi-year seasonal composites

Single-year seasonal composites are aggregated into multi-year seasonal composites through an average operation weighted by the number of observations.

The aggregation is operated on a pixel basis and with pixels associated with a same status. Priority is given to the "land" status and then, come the "snow", "water" and "cloud" statuses in this sequence. Pixel status has been recorded in a specific flag band named "L3_StartDate_EndDate_NOBS" during the single-year seasonal composite generation (see section 4.2.2). In addition, the average is weighted by the number of valid observations associated with each single-year seasonal composites (information stored in a specific flag band named "L3_StartDate_EndDate_NOBS" – see section 4.2.2)

Algorithm assumptions and limitations

None: the same processing chain is applicable to any type of satellite data and for any length of compositing period.

Input and output data

Input and output data associated with this compositing process are described in Table 4-16. Associated parameters are presented in Table 4-17.

Table 4-16: Input and output data of the multi-year seasonal composites generation (multi-year approach)

DATA	DESCRIPTION	INTENT (IN, OUT, INOUT)	PHYSICAL UNIT	RANGE

L3_<sensor>_StartDate_EndDate_SR_<n>
n = 1,..,15 (MERIS) // n = B0, B2, B3, MIR (SPOT-VGT)
sensor = FR for MERIS FR, RR for MERIS RR, VGT for SPOT-VGT
	Global raster of surface reflectance in each bands of the single-year seasonal composites (year information being included in the Date fields
	GeoTiff format
		IN	None	[0 … 1]

L3_<sensor>_StartDate_EndDate_Status sensor = FR for MERIS FR, RR for MERIS RR, VGT for SPOT-VGT	Global raster of pixel status associated with the surface reflectance in the single-year seasonal composites GeoTiff format	IN	None	[1 … 7]
L3_<sensor>_StartDate_EndDate_NOBS sensor = FR for MERIS FR, RR for MERIS RR, VGT for SPOT-VGT	Global raster counting the contributing observations in the single-year seasonal composites GeoTiff format	IN	None	[0 … 500]
LUT 6	Look-Up-Table indicating which strata require the generation of multi-year seasonal composites (see Table 4-15)	IN	/	/
LUT 5	Look-Up-Table indicating the seasonal compositing periods associated with each stratum in the classification chain (see Table 4-14)	IN	/	/
LUT 7	Look-Up-Table indicating the start and end years of each epoch (see Table 4-17)	IN	/	/

L3_<sensor>_MY_StartYear_EndYear
_StartMonthDay_EndMonthDay_SR_<n>
n = 1,..,15 (MERIS) // n = B0, B2, B3, MIR (SPOT-VGT)
sensor = FR for MERIS FR, RR for MERIS RR, VGT for SPOT-VGT
	Global raster of surface reflectance in each bands of the multi-year seasonal composites
	GeoTiff format
		OUT	None	[0 … 1]

DATA	DESCRIPTION	INTENT (IN, OUT, INOUT)	PHYSICAL UNIT	RANGE
L3_<sensor>_MY_StartYear_EndYear _StartMonthDay_EndMonthDay _Status sensor = FR for MERIS FR, RR for MERIS RR, VGT for SPOT-VGT	Global raster of pixel status associated with the surface reflectance in the multi-year seasonal composites GeoTiff format	OUT	None	[1 ... 7]
L3_<sensor>_MY_StartYear_EndYear _StartMonthDay_EndMonthDay _NOBS sensor = FR for MERIS FR, RR for MERIS RR, VGT for SPOT-VGT	Global raster counting the contributing observations in the multi-year seasonal composites GeoTiff format	OUT	None	[0 ... 500]

Table 4-17: Parameters needed in the multi-year seasonal composites generation (contained in LUT 7)

PARAMETERS	DESCRIPTION	INTENT (IN, OUT, INOUT)	FORMAT	RANGE
StartYear	Start year of the multi-year period	IN	Date (YYYY)	[1998 ... 2012]
EndYear	End year of the multi-year period	IN	Date (YYYY)	[1998 ... 2012]

4.3.2 MY_S2 – Aggregation of single-year land cover maps

A superposition between the single-year land cover products (raster files CLASSIF1_<Year> or CLASSIF2_<Year>) is operated. For each pixel, a histogram of class frequency is computed, which is then interpreted according to several decision rules. As a result, a multi-year land cover class (i.e. a unique class number NB_LAB and name LABEL) is associated with each pixel and an output multi-year land cover product (raster file CLASSIF1_MY or CLASSIF2_MY) is generated.

The histogram interpretation process also associates each pixel with an ambiguity code (CODE_MY) that quantifies the frequency of occurrence of the land cover class (NB_LAB and LABEL) finally associated with the pixel according to the deicision rules. This code can stand for an indicator of the land cover label.

Input and output data

Input and output data associated with the aggregation of single-year land cover maps are described in Table 4-18.

Table 4-18: Input and output data of the aggregation of single-year land cover maps (multi-year approach)

DATA	DESCRIPTION	INTENT (IN, OUT, INOUT)	PHYSICAL UNIT	RANGE
L4_<NB_ST>_CLASSIF1_<Year>	Raster at the stratum level resulting from the spectral classification algorithm run on single-year seasonal composites, where each pixel is associated with a land cover class ID (NB_LAB)	IN	None	[0 … 255]
LUT 2	Look-Up-Table describing the CCI LCCS land cover legend (see Table 4-10)	IN	/	/
L4_< NB_ST>_CLASSIF1_Histo	Text file (one for each equal-reasoning area) containing for each pixel the land cover classes frequency	INOUT	Long	[0 … 100]
L4_<NB_ST>_CLASSIF1_MY	Land cover map resulting from the aggregation of single-year spectral land cover maps, where each pixel is associated with a land cover class through an ID (see Table 4-10)	OUT	None	[0 … 255]
CODE_MY	Frequency of a same land cover class observed over the multiple aggregated years, thus reflecting, at the pixel level, the reliability of the CLASSIF1_MY or CLASSIF2_MY	OUT	None	0 … 10

4.4 Detailed processing scheme of the multi-sensor approach

The development of a multi-sensor approach is required by (i) the lack of acquisition in the MERIS FR acquisition mode and (ii) the temporal coverage of the MERIS sensor limited to the 10 year-period from 2003 to 2012.

4.4.1 Increasing MERIS FR spatial coverage

Using images acquired in the MERIS RR mode is considered as the most convenient approach to deal with a possible lack of MERIS FR acquisitions. Indeed, MERIS FR and RR time series are acquired by the same sensor on the same satellite, with the difference that FR data are less regularly downlinked. In any cases, it must obey to the following consideration: priority is given to the best available EO information and therefore, the merging between MERIS FR and RR is not systematic. In other terms, where and when the coverage of the MERIS FR dataset is high enough to allow producing consistent land cover maps, MERIS RR should not be used.

The merging between MERIS FR and RR dataset is needed where there are gaps in the MERIS FR seasonal composites (i.e. where the number of MERIS FR available observation is null). The merging is operated at the land cover map level, meaning that the classification chain is run for the MERIS FR and RR time series independently.

4.4.2 Increasing MERIS temporal coverage

In order to generate the 2000 epoch land cover map, SPOT-VGT time series need to be used. 5 years of SPOT-VGT time series from 1998 to 2002 are used to "update" the reference baseline and derive the 2000 epoch.

All preliminary and classification steps (including the multi-year approach) are applied to these 5 years of data.

4.5 Detailed processing scheme of the classification chain

As introduced in the introduction about classification logical models (section 4.1), the classification process in itself is organised in 4 major processing steps (Figure 4-1), which rely on data generated in the preliminary steps (seasonal composites, reference land cover database and stratification layers). The algorithms run independently for each equal-reasoning area with specific parameters.

In addition, the multi-year and multi-sensor approaches developed to increase the land cover products accuracy and stability over time require the application of additional preliminary steps in order to prepare multi-year seasonal composites from MERIS FR and RR and SPOT-VGT sensors. It also requires running additional steps during the classification process in itself and in particular, after the spectral classification in order to (1) aggregate single-year land cover maps into a multi-year product and to (2) complement MERIS FR land cover maps with the MERIS RR land cover maps.

The steps of the land cover mapping processing chain are detailed in the sections below (4.5.1 to 4.5.6). The link with the additional steps required by the multi-year and multi-sensor approaches is also highlighted.

4.5.1 Step 1 – Supervised spectral classification

Any supervised classification procedure relies on the following steps:

- Gathering training data (i.e. spectral signatures for each land cover class of interest) as representative of the "real-world" as possible;
- Determining the type of the classifier to use – the classifier corresponding to the function which (i) analyzes the training data and (ii) predicts the output class of any input pixel by generalizing the training data to "unseen" situations;
- Defining the classification parameters to optimize the algorithm's performance.

With regard to the training dataset collection, two refinements are brought in this project with respect to a classical approach. First, training data are defined locally in order to take into account that a spectral signature of a given land cover class is not necessarily reliable over large extents. Indeed, it

must be recognized that when working at large scales, each land cover label is probably associated with several spectral signatures (e.g. several spectral signatures for the crops to render maize or wheat for instance, etc). In order to face this problem, a locally-adjusted training dataset gathering (and thus classification approach) is developed. Each ***equal-reasoning area*** to classify is split into smaller moving windows which will stand for the ***classification areas***. The size of the classification areas varies according to the stratum. The regions inside which training dataset are gathered are called ***search areas*** and correspond to 240*240 km areas centred on each classification area. This decoupling between classification and search areas allows accounting for the imperfection of the auxiliary reference dataset from which training samples are collected (e.g. for a false absence of a given land cover class over a specific region). If one land cover class is not represented in significant proportions within the classification area, it will still be included in the training data thanks to the larger extent of the search area. Furthermore, the overlap of the search areas contributes to the seamless change of the training data to avoid artefacts at the boundaries between two classification areas. Illustrations and relationships between equal-reasoning, classification and search areas are presented in Figure 4-9.

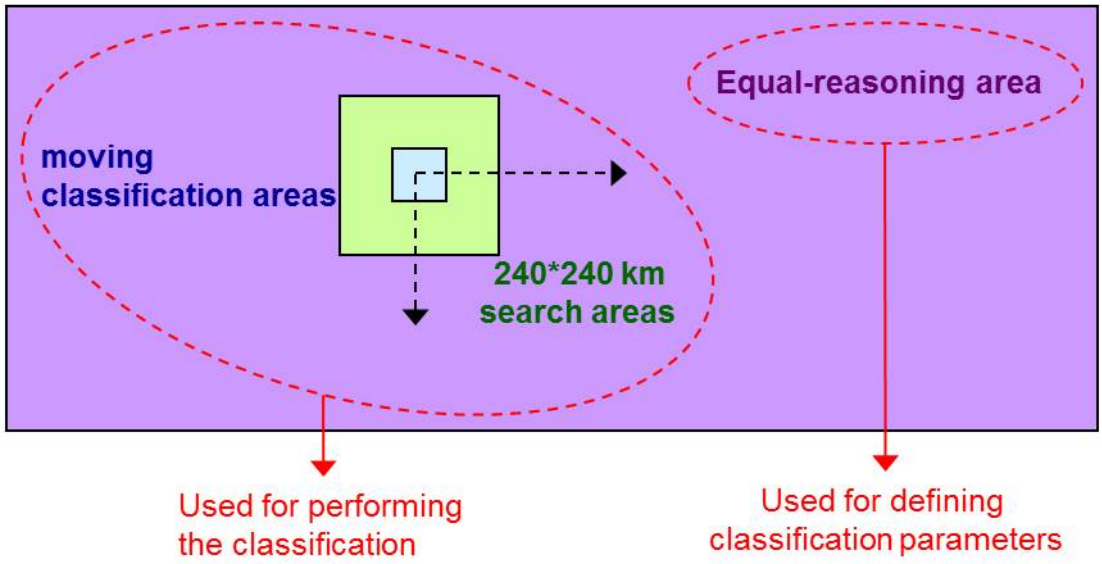


Figure 4-9: Illustration of the tree spatial units (equal-reasoning, classification and search areas) of the developed locally-adjusted supervised classification procedure

It should be clearly stated that these classification and search areas are completely distinct from the equal-reasoning areas defined in the stratification phase. Equal-reasoning areas allow adjusting the classification parameters while classification and search areas allow a local classification. The preparation of such training data is achieved in a ***first preliminary step*** of the supervised spectral classification procedure (see section 4.5.1.1).

The second refinement with regard to the training dataset collection consists in completing the classical training data (i.e. the representation of each land cover class through spectral signatures) with a priori information about the occurrence probability of each land cover class. This a priori information will be extracted from an auxiliary reference dataset at the spatial scale of the search areas. Indeed, while the local classification strategy allows accounting for local specificities, it also

makes the algorithm more sensitive to spatial inconsistency of the training dataset. The occurrence probabilities computation at an intermediate scale is performed in a ***second preliminary step*** of the supervised spectral classification procedure (see section 4.5.1.2).

As for the supervised classification algorithm in itself, it relies on the classical maximum likelihood principle. The algorithm and associated parameters are described in section 4.5.1.3, along with the entire procedure.

4.5.1.1 Preliminary step 1 – Training dataset preparation

The training dataset shall provide representative spectral signatures for each land cover class of interest for the 2400*240 km search areas centererd on the classification areas. It is derived from the reference land cover database, which is a key auxiliary dataset built in this project [AD-11] (see Figure 4-9). This reference database consists of a set of global, regional and local reference land cover maps selected as the most accurate ones available for a given region, with the highest spatial resolution and with a CCI-compatible legend.

The workflow of this preliminary step for training dataset preparation is provided in Figure 4-10. It is organized in two steps:

- The first step consists in applying a morphological filter to the reference database in order to ensure training dataset as “pure” as possible (i.e. not contaminated by “border effects”);
- The second step consists in extracting, for each eroded land cover class, representative spectral signatures at the search area level.

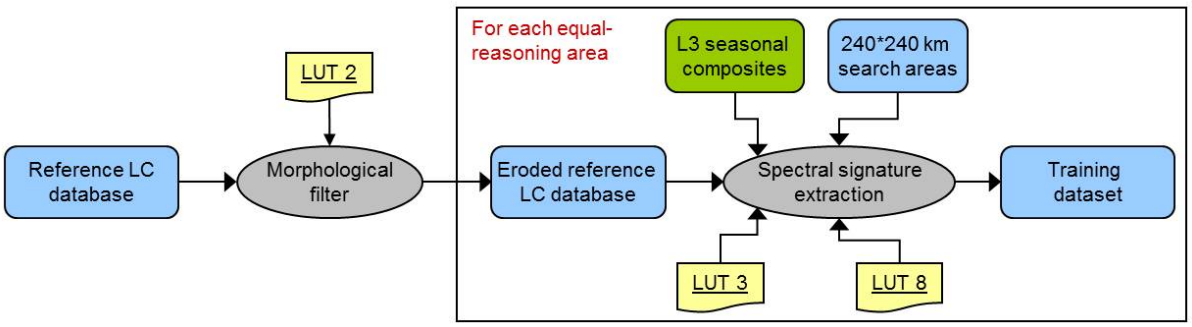


Figure 4-10: Activity diagram illustrating the training dataset preparation: (1) the reference land cover database is eroded and (2) spectral signatures are extracted or each eroded class

The morphological filter is based on the erosion principle, i.e removing pixels along the boundaries of each land cover class. However, eroding each of the class with the same number of pixels would erase the thin classes from the reference. The erosion was therefore designed in such a way that at least one pixel remains from each group of adjacent pixels belonging to the same class.

The morphological filter keeps only the pixels with the most neighbours of the same class. A fixed neighbourhood of 3*3 pixels was used for all strata. The filter is composed of two passes. The first pass counts the number of pixels from the same class than the central pixel (example on Figure 4-11 B). The second pass erases the label of the central pixel if it has less neighbours than another pixel of the same class in the neighbourhood (example on Figure 4-11D). This specificity allows conserving

isolated pixels in the reference layer if there is no larger group in the neighbourhood, which is not the case with a standard erosion filter (example on Figure 4-11 C).

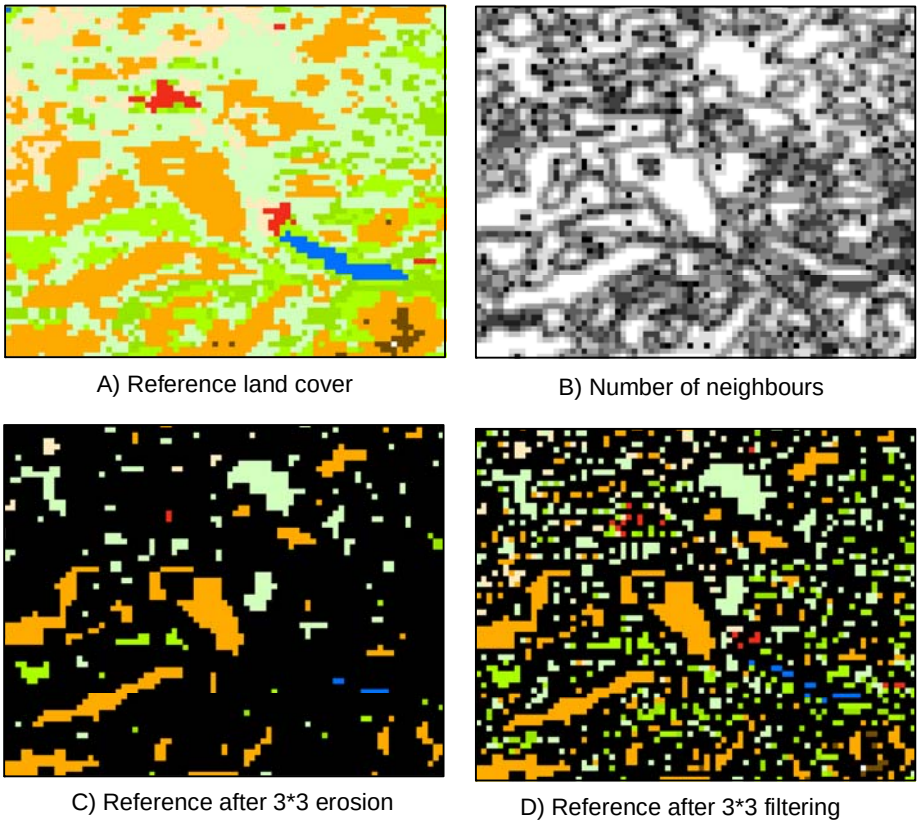


Figure 4-11: Results of morphological erosion and majority neighbours filtering in a rural landscape

This composite filter results in a new raster file which contains the new eroded reference land cover database. This new database is then used to extract spectral signature for each class and thus build training dataset.

This second step is operated at the 240*240 search area level while running with parameters defined at the stratum level. These parameters are contained in the LUT 8 and include, for each stratum, the L3 seasonal composites and the spectral channels from which extracting the spectral signatures. Inside each search area, for each class and from the specified seasonal composites and spectral channels, the algorithm extracts reflectance values to define representative spectral distributions that will serve as training dataset.

Algorithm assumptions and limitations

None

Input and output data

Input and output data associated with the application of the morphological filter on the reference are described in Table 4-19 while associated parameters are presented in Table 4-20.

Table 4-21 and Table 4-22 provide the same information for the generation of the training dataset (through the extraction of spectral signature by class).

Table 4-19: Input and output data of the 1st preliminary step of the supervised spectral classification, for the application of the morphological filter on the reference

DATA	DESCRIPTION	INTENT (IN, OUT, INOUT)	PHYSICAL UNIT	RANGE
Ref_LC	Global reference land cover database, where each pixel is associated with a land cover class of the LCCS CCI-legend through an ID (see LUT2 in Table 4-10)	IN	Short	[0 .. 255]
LUT 2	Look-up table describing the ID of each land cover class (see Table 4-10)	IN	/	/
Ref_LC_Training	Reference land cover database, processed with the morphological filter. Each pixel is associated with a land cover class of the LCCS CCI-legend or with a no data value through an ID (see LUT 2 in Table 4-10)	OUT	Short	[0 … 255]

Table 4-20: Parameter needed in the 1st preliminary step of the supervised spectral classification, for the application of the morphological filter on the reference

PARAMETERS	DESCRIPTION	INTENT (IN, OUT, INOUT)	FORMAT	RANGE
k	Size of the window inside which the filter is applied	IN	Short	[0 … 255]

Table 4-21: Input and output data of the 1st preliminary step of the supervised spectral classification, for the spectral signature extraction

DATA	DESCRIPTION	INTENT (IN, OUT, INOUT)	PHYSICAL UNIT	RANGE
L3_StartDate_EndDate_SR_<n> n = 1,..,15 (MERIS) // n = B0, B2, B3, MIR (SPOT-VGT)	Global raster of surface reflectance in each bands of the seasonal composites GeoTiff format	IN	None	[0 … 1]
Ref_LC_Training	Reference land cover database, processed with the morphological filter. Each pixel is associated with a land cover class of the LCCS CCI-legend or with a no data value through an ID (see LUT 2 in Table 4-10)	IN	None	[0 … 255]

DATA	DESCRIPTION	INTENT (IN, OUT, INOUT)	PHYSICAL UNIT	RANGE
GRID_240	Raster grid for each stratum defining the 240*240 km search areas localization	IN	None	[0 … ~60000]
LUT 3	Look-Up-Table describing the location of the 22 strata (see Table 4-11)	IN	/	/
LUT 8	Look-Up-Table describing the parameters of the supervised classification algorithm (see Table 4-22)	IN	/	/
ROI_<NB_LAB>_<GRID>_StartDate _EndDate_SR - NB_LAB representing each class of the LCCS legend - GRID representing the search area - StartDate and EndDate defining the source seasonal composite	Pure training dataset for each class ("NB_LAB") and for each search area ("GRID"), consisting in representative reflectance values distributions in specific seasonal composites (defined by "StartDate" and "EndDate")	OUT	None	[0 … 1]

Table 4-22: Parameters of the supervised spectral classification (contained in LUT 8)

PARAMETERS	DESCRIPTION	INTENT (IN, OUT, INOUT)	FORMAT	RANGE
NB_ST	Number of the stratum (see Table 4-11)	IN	Short	[1 … 22]
StartDate _Si	Start date of the seasonal composite(s) to be used, i being the number of composite used for each ROI	IN	Date (YYYYMMDD)	19980101 … 20121231
EndDate_Ei	End date of the seasonal composite(s) to be used, i being the number of composite used for each ROI	IN	Date (YYYYMMDD)	19980101 … 20121231
CHij	Number of the j spectral channels for the i composite to be used	IN	Short	[1 … 15] for MERIS [B0, B2, B3, MIR] for SPOT-VGT

4.5.1.2 Preliminary step 2 – Computation of land cover classes' occurrence probabilities at the stratum level

As already mentioned in the introduction, the supervised classification is locally adjusted: the algorithm is run within classification areas, using training dataset collected within 240*240 km search areas. In order to minimize the sensitivity of the algorithm to the possible spatial inconsistency of the training dataset, the algorithm also makes use of a priori occurrence probabilities for each land cover class defined at at the spatial scale of the search areas.

The occurrence of each land cover class is computed inside each search area through a comparison with the reference land cover database. These occurrences will serve as a priori information for the supervised classification algorithm and complete the classical training dataset made of spectral signatures.

Input and output data

Table 4-23 presents the input and output data associated with the step of land cover classes' occurrence computation.

Table 4-23: Input and output data of the 2nd preliminary step of the supervised spectral classification, for the LC classes occurrence computation

DATA	DESCRIPTION	INTENT (IN, OUT, INOUT)	PHYSICAL UNIT	RANGE
REF_LC	Reference land cover layer where each pixel is associated with a land cover class through an ID (see LUT 2 in Table 4-10)	IN	None	[0 … 255]
GRID_240	Raster grid for each stratum defining the 240*240 km search areas localization	IN	None	[0 … ~60000]
LUT 2	Look-Up-Table describing the CCI LCCS land cover legend (see Table 4-10)	IN	/	/
OCC_<NB_LAB>_<GRID> - NB_LAB representing each class of the LCCS legend - GRID representing the search area	Pure training dataset for each class ("NB_LAB") and for each search area ("GRID"), consisting in representative reflectance values distributions in specific seasonal composites (defined by "StartDate" and "EndDate")	OUT	None	[0 … 1]

4.5.1.3 Supervised spectral classification

This section describes the supervised classification algorithm in itself, and more precisely the classifier it uses to predict the output class of all pixels.

The algorithm makes use of the most common classifier, which is the Maximum Likelihood (ML), and relies on a statistical approach. The ML classifier assumes that each spectral class can be described by a multivariate normal distribution. The ML algorithm therefore takes advantage of both the mean vectors and the multivariate spreads of each class, and would be able to also identify elongated classes. More precisely, the probability density function (PDF) of each class is estimated under a Gaussian assumption. It is rather simple and rapid but it assumes that there is only one population per class (i.e. a unique spectral signature by land cover class). This assumption seems realistic since the algorithm is applied at a local scale.

Maximum likelihood Classification is a statistical decision criterion to assist in the classification of overlapping signatures; pixels are assigned to the class of highest probability.

As already mentioned, the supervised algorithm is run with parameters defined at a larger scale while being applied at a more local scale. On one hand, the input Earth Observation data (i.e. seasonal composites in a definite number of spectral channels) are defined at the equal-reasoning area level. On the other hand, the algorithm is processed using moving km classification areas using (i) training data valid within 240*240 km search areas centred on the classification areas (see section 4.5.1.1) and (ii) a priori land cover classes' occurrence probabilities defined at the same scale than the search areas (see section 4.5.1.2).

The principle of this supervised classification strategy is illustrated in Figure 4-12.

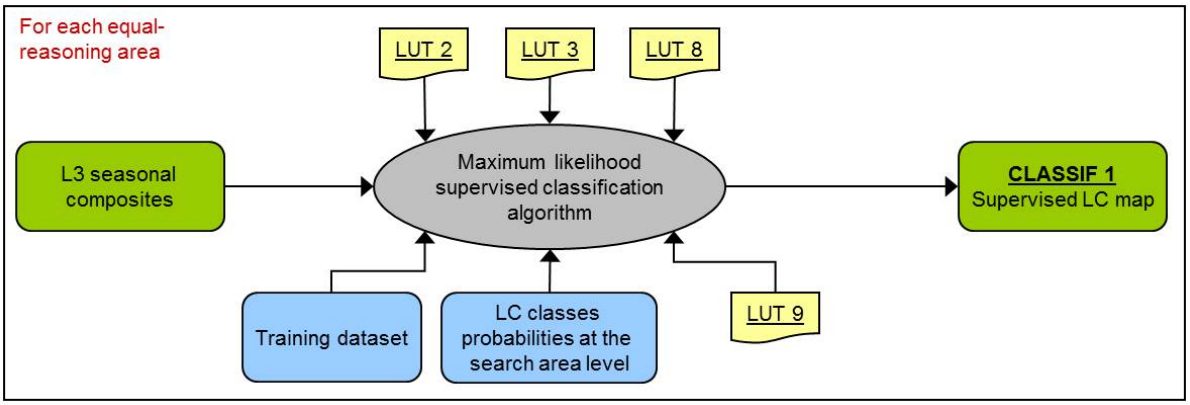


Figure 4-12: Activity diagram illustrating the Gaussian Maximum Likelihood supervised classification algorithm developed in the CCI-LC project

The supervised classification algorithm is applied on L3 seasonal composites independently for each equal-reasoning area (NB_ST).

The principle of the maximum likelihood classification relies on the Bayes theorem. In the context of the supervised classification of remote sensing data, the likelihood of a land cover class for a given set of spectral values can be derived from the a priori probability and the distribution of the spectral values for this class. The computation of the a priori probability is described in the second preliminary step; the parameters of the distribution of the spectral values are derived from the training datasets. For each pixel, the label is assigned to the class with the largest probability. It is worth noting that the largest probability is stored together with the label in order to be used as a quality flag.

As an output, the supervised classification algorithm creates, for each stratum (NB_ST), a raster file (CLASSIF 1) where each pixel is associated to a land cover class through the NB_LAB identifier (see LUT 2 in Table 4-10). It also produces an auxiliary layer providing the classification probability associated with each pixel (CL1_PROB). This probability informs about the confidence in the classification output.

Local classification could lead to tiling artefacts because the classifiers change for each classification area. However, this is mitigated by the use of a priori values computed from the entire equal-reasoning area as well as the use of a search area that is larger than the classification area. While the classification is locally based, the a priori brings some consistency and the classifiers parameters are seamlessly changing. The size of the classification area was set by tuning the method in order to remove tiling artefact with optimal processing time. In addition, the tiling artefacts can also be avoided by shifting the starting point of the classification areas year after year.

Algorithm assumptions and limitations

The Gaussian ML approach underlying the supervised classification algorithm makes the assumption that each spectral class can be described by a Gaussian distribution. It assumes that classes in the input data have a Gaussian distribution and that signatures were well selected; this is not always a safe assumption.

Input and output data

Input and output data associated with this supervised classification process are described in Table 4-24. Associated parameters are presented in Table 4-22 (LUT 8) and Table 4-25 (LUT 9). Parameters of the CCI land cover legend and of the stratification layer – which are used in this step but also in the following ones throughout the classification chain – are provided in Table 4-10 and Table 4-11.

Table 4-24: Input and output data of the step 1a of the classification chain, i.e. the spectral supervised classification algorithm

DATA	DESCRIPTION	INTENT (IN, OUT, INOUT)	PHYSICAL UNIT	RANGE
L3_StartDate_EndDate_SR_<n> n = 1,..,15 (MERIS) // n = B0, B2, B3, MIR (SPOT-VGT)	Global raster of surface reflectance in each bands of the seasonal composites GeoTiff format	IN	None	[0 … 1]
ROI_<NB_LAB>_<GRID>_StartDate _EndDate_SR - NB_LAB representing each class of the LCCS legend - GRID representing the search area - StartDate and EndDate defining the source seasonal composite	Pure training dataset for each class (“NB_LAB”) and for each search area (“GRID”), consisting in representative reflectance values distributions in specific seasonal composites (defined by “StartDate” and “EndDate”)	IN	None	[0 … 1]

DATA	DESCRIPTION	INTENT (IN, OUT, INOUT)	PHYSICAL UNIT	RANGE
OCC_<NB_LAB>_<GRID> - NB_LAB representing each class of the LCCS legend - GRID representing the search area	Pure training dataset for each class ("NB_LAB") and for each search area ("GRID"), consisting in representative reflectance values distributions in specific seasonal composites (defined by "StartDate" and "EndDate")	IN	None	[0 … 1]
LUT 2	Look-up table describing the ID of each land cover class (see Table 4-10)	IN	/	/
LUT 3	Look-Up-Table describing the location of the 22 strata (see Table 4-11)	IN	/	/
LUT 8	Look-Up-Table describing the parameters of the supervised classification algorithm (see Table 4-22)	IN	/	/
LUT 9	Look-Up-Table providing the classification areas for each stratum (see Table 4-25)	IN	/	/
CLASSIF_1	Land cover map resulting from the supervised classification algorithm, where each pixel is associated with a land cover class through an ID (see Table 4-10)	OUT	None	[0 … 255]
Code_Classif1	Classification probability associated with the label selected for each pixel	OUT	None	0 … 1

Table 4-25: Parameters defining the classification areas for each stratum in which running the step 1 of the classification chain, i.e. the spectral supervised classification algorithm (contained in LUT 9)

PARAMETERS	DESCRIPTION	INTENT (IN, OUT, INOUT)	FORMAT	RANGE
NB_ST	Number of the stratum (see Table 4-11)	IN	Short	[1 … 22]
Size	Size (in kilometres) of the classification and search areas	IN	Short	[1 … 255]

4.5.2 Step 2 – Unsupervised spectral classification

Using an unsupervised classification algorithm allows reaching a rather high degree of automation while reducing the processing time. These advantages were successfully demonstrated in the GlobCover experience [RD-103].

Unsupervised image classification is a classification process based solely on the image statistics, without availability of training data or other a priori knowledge of the area. The unsupervised algorithm used in this project relies on the clustering principle. Clustering is the task of assigning a set of pixels into clusters so that the pixels in the same cluster are more similar to each other than to those in other clusters. The assignation is based on natural groupings present in the reflectance values. The basic premise is that reflectance values within a given land cover class should be close together in the measurement space, whereas pixels belonging to different land cover classes should be comparatively well-separated [RD-137].

Clustering is not an algorithm in itself but a general task, which can be achieved by various algorithms that differ significantly in their notion of what constitutes a cluster and how to efficiently find them. The unsupervised classification algorithm used in this project relies on the ISODATA clustering technique, which represents each cluster by a single mean vector. More detail about this algorithm (and about all the classification procedure) is given hereafter in section 4.5.2.1.

Directly linked with the unsupervised algorithm is the labelling procedure, which aims at transforming the spectral clusters into land cover classes. The unsupervised algorithm indeed generates spectrally separable clusters, for which the land cover label is not known. The land cover class associated with each cluster needs to be determined in a further step, by comparing the cluster to some auxiliary reference dataset.

The requirements for continuity and consistency in the long-term require an objective and automated labelling procedure. This challenge was already successfully addressed in the framework of the GlobCover project thanks to the use a global reference dataset and the definition of a set of generic decision rules. This procedure is detailed in the section 4.5.2.2.

4.5.2.1 Unsupervised ISODATA algorithm

This section describes the unsupervised classification algorithm used in the project, which relies on the ISODATA clustering technique.

The ISODATA algorithm is an iterative optimization clustering procedure, also called the migrating means technique, presented by [RD-96]. It is based upon estimating some reasonable assignment of the pixel vectors into candidate clusters and then moving them from one cluster to another in such a way that the Sum of Squared Error (SSE) is progressively reduced.

The algorithm is implemented by the following set of basic steps:

- First, the procedure starts by randomly selecting C points in the multidimensional input data space that will serve as candidate cluster centres:

  $\mathbf{m}_i$, i = 1 … C;

- Second, each pixel in the image (or segment of image) to classify is assigned to the nearest candidate cluster. This assignment is based on the minimization of the Euclidean distance function between that pixel and the candidate cluster centers $\mathbf{m}_i$;
- Third, after each iteration, the new set of means that result from the grouping produced in step 2 are computed;
- Fourth, the entire process is repeated. After each iteration, a new mean is calculated for each cluster, based on the actual spectral locations of the pixels in the cluster. Then, these new means are used for defining clusters in the next iteration. The process continues until there is little change between iterations, i.e. if the normalised percentage of pixels whose assignments are unchanged since the last iteration reaches a convergence threshold or the maximum number of iterations is reached.

The ISODATA principle is illustrated in Figure 4-13.

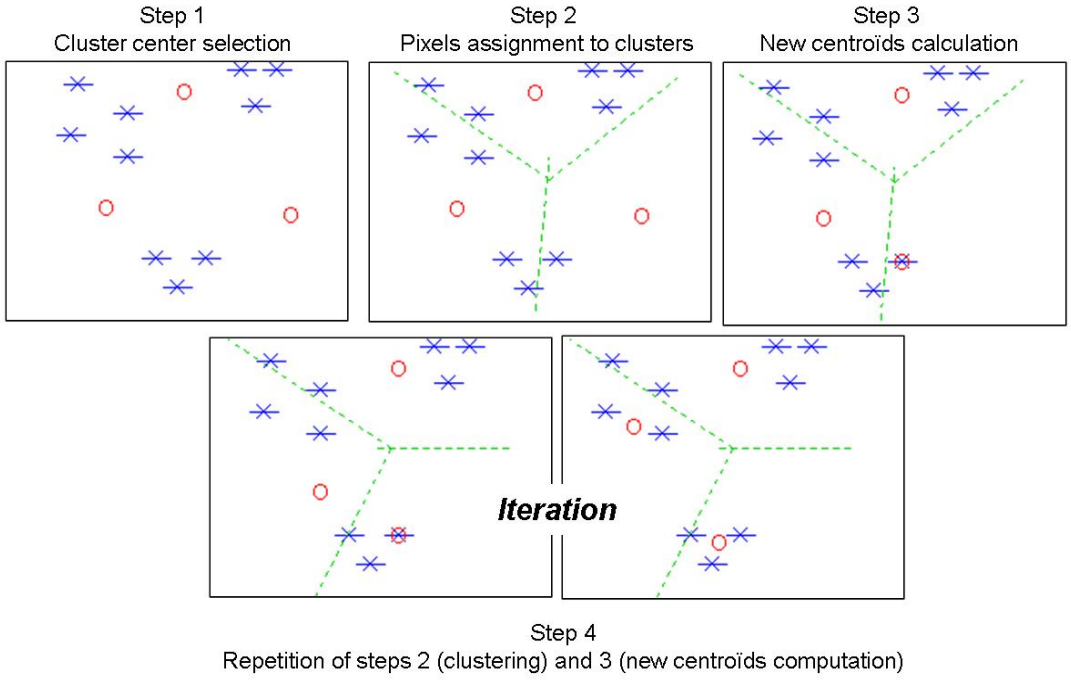


Figure 4-13: Principle of the ISODATA clustering technique

The ISODATA algorithm is applied to L3 seasonal composites independently for each equal-reasoning area (NB_ST). The input and output data involved in this algorithm are summarized in Table 4-26. The classification parameters are contained in the LUT 10 and listed in Table 4-27. As an output, the ISODATA algorithm creates, for each stratum (NB_ST), an output raster file (L4_<NB_ST>_Clusters_Spectral) where each pixel is associated to a spectrally homogeneous but unlabelled cluster (NB_Cluster).

Algorithm assumptions and limitations

The ISODATA clustering technique relies on the assumption that each land cover class can be well-represented by a single mean vector.

Input and output data

Input and output data associated with this unsupervised classification process are described in Table 4-26. Associated parameters are presented in Table 4-27.

Table 4-26: Input and output data of the step 1b of the classification chain, i.e. the spectral unsupervised (ISODATA) classification algorithm

DATA	DESCRIPTION	INTENT (IN, OUT, INOUT)	PHYSICAL UNIT	RANGE
L3_StartDate_EndDate_SR_<n> n = 1,..,15 (MERIS) // n = B0, B2, B3, MIR (SPOT-VGT)	Global raster of surface reflectance in each bands of the seasonal composites GeoTiff format	IN	None	[0 … 1]
LUT 3	Look-Up-Table describing the location of the 22 strata (see Table 4-11)	IN	/	/
LUT 10	Look-Up-Table describing the parameters of the unsupervised algorithm (see Table 4-27)	IN	/	/
L4_<NB_ST>_Clusters_Spectral	Raster at the stratum level resulting from the unsupervised classification algorithm where each pixel is associated with a cluster ID (NB_Cluster)	OUT	None	[0 … 255]

Table 4-27: Parameters needed in the step 2 of the classification chain, i.e. the spectral unsupervised (ISODATA) classification algorithm (contained in LUT 10)

PARAMETERS	DESCRIPTION	INTENT (IN, OUT, INOUT)	FORMAT	RANGE
NB_ST	Number of the stratum (see Table 4-11)	IN	Short	[1 … 22]
StartDate _Si	Start date of the seasonal composite(s) to be used, i being the composite number	IN	Date (YYYYMMDD)	19980101 … 20121231
EndDate_Si	End date of the seasonal composite(s) to be used, i being the composite number	IN	Date (YYYYMMDD)	19980101 … 20121231
CHij	Number of the j spectral channels for the i composite to be used	IN	Short	[1 … 15] for MERIS [B0, B2, B3, MIR] for SPOT-VGT
N	The maximum number of clusters (or classes) to generate	IN	None	[0..255]
NB_IT	The number of iterations to be performed	IN	None	[0..100]
NB_PIX	The minimum number of pixels needed to form a class	IN	None	[50…50000]

4.5.2.2 Automatic reference-based labelling

The ISODATA algorithm has interpreted, for each stratum, L3 seasonal composites into a set of clusters which have the property to be spectrally homogeneous but which are not identified by a land cover label. Transforming these clusters into land cover classes (identified by a number LB_LAB and a name LABEL – see LUT 2 in Table 4-10) is the objective of this labelling step. This is done through a comparison between the cluster raster file (L4_<NB_ST>_CLUSTERS_Spectral) and an auxiliary land cover reference.

The auxiliary data that is used as land cover reference is a key dataset of this project [AD-11]. This is the same data than the one used to define the training sample in the supervised classification approach (see section 4.5.1.1). As already mentioned, this reference database (REF_LC) consists of a set of global, regional and local reference land cover maps selected as the most accurate ones available for a given region, with the highest spatial resolution and with a CCI-compatible legend. It is associated with the CCI land cover legend, i.e. with the NB_LAB, LABEL and color codes contained in the LUT 2 (Table 4-10).

A superposition of the cluster raster (L4_<NB_ST>_CLUSTERS) with the reference layer (REF_LC) is operated. For each cluster, a histogram of class frequency is computed (Figure 4-14). The most

represented classes inside the cluster are identified and ranked using (i) the number of pixels they cover (NB_Pix1, NB_Pix2, etc.) and (ii) their label (NB_LAB1, NB_LAB2, etc.).

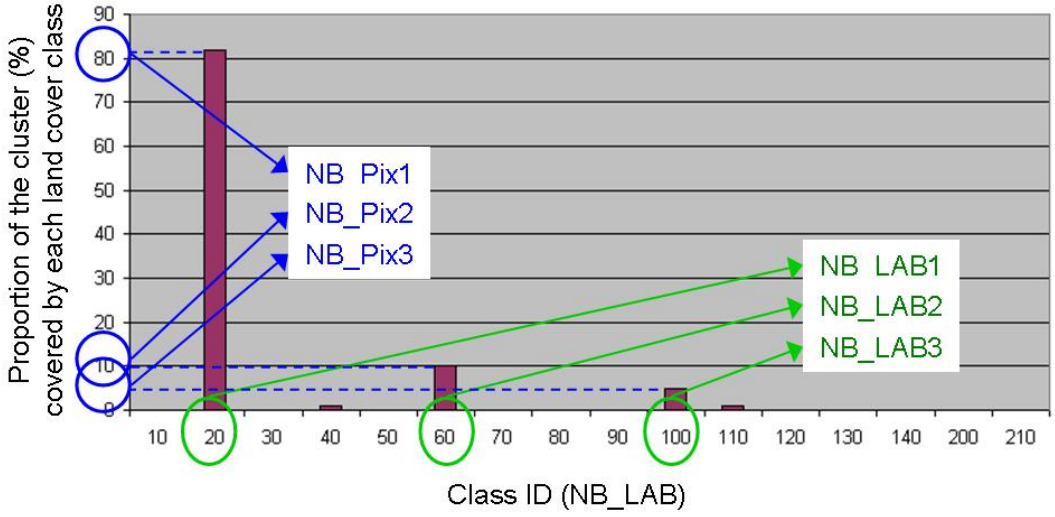


Figure 4-14: Histogram of class frequency interpretation: most represented classes are identified

The class frequency histogram is then interpreted according to several decision rules. As a result, a unique land cover class (i.e. a unique class number NB_LAB and name LABEL) is associated with each cluster (NB_Cluster) and an output raster file (CLASSIF_2) is generated. The histogram interpretation process also associates each cluster with an ambiguity code (CODE) that characterizes the ambiguity of the interpretation, and thus the reliability of the associated land cover label.

Input and output data

Input and output data associated with the automated labelling procedure are described in Table 4-28.

Table 4-28: Input and output data of the step 2 of the classification chain, which is the spectral unsupervised classification process, for the automated labelling procedure

DATA	DESCRIPTION	INTENT (IN, OUT, INOUT)	PHYSICAL UNIT	RANGE
L4_<NB_ST>_Clusters_Spectral	Raster at the stratum level resulting from the unsupervised classification algorithm where each pixel is associated with a cluster ID (NB_Cluster)	IN	None	[0 … 1024]
REF_LC	Reference land cover layer where each pixel is associated with a land cover class through an ID (see LUT 2 in Table 4-10)	IN	None	[0 … 255]

LUT 2	Look-Up-Table describing the CCI LCCS land cover legend (see Table 4-10)	IN	/	/
L4_<NB_ST>_Histo	Text file (one for each equal-reasoning area) containing for each cluster (NB_Cluster) the land cover classes frequency (with indices NB_Pixi and NB_LABi)	INOUT	Long	[0 … 100]
CLASSIF 2	Land cover map resulting from the unsupervised classification approach (ISODATA algorithm + labelling process), where each pixel is associated with a land cover class through an ID (see Table 4-10)	OUT	None	[0 … 255]
Code_Classif2	Ambiguity code that characterizes the ambiguity of the classification process, thus reflecting, at the pixel level, the reliability of the CLASSIF 1B	OUT	None	0 … 10

4.5.3 Step 3 – Merging of land cover maps obtained with the spectral supervised and unsupervised approaches

At this stage of the classification chain, two global land cover maps have been produced:

- the map called CLASSIF_1, resulting from the supervised classification approach detailed throughout the section 4.5.1;
- the map called CLASSIF_2, resulting from the unsupervised classification approach detailed throughout section the 4.5.2.

This step aims at combining the two maps according to objective decision criteria to generate a unique land cover map (CLASSIF_3). This unique land cover map is the output of the CCI classification chain. The decision criteria between CLASSIF_1 and CLASSIF_2 are class-specific and are provided in a set of decision rules.

Algorithm assumptions and limitations

None

Input and output data

Input and output data associated with this merging process are described in Table 4-29.

Table 4-29: Input and output data for the step 3 of the classification chain, i.e. the merging of land cover maps obtained by the supervised and unsupervised classification approaches

DATA	DESCRIPTION	INTENT (IN, OUT, INOUT)	PHYSICAL UNIT	RANGE
CLASSIF_1	Land cover map resulting from the supervised classification algorithm, where each pixel is associated with a land cover class through an ID (see Table 4-10)	IN	None	[0 … 255]
Code_Classif1	Classification probability associated with the label selected for each pixel	IN	None	[0 … 1]
CLASSIF 2	Land cover map resulting from the unsupervised classification approach (ISODATA algorithm + labelling process), where each pixel is associated with a land cover class through an ID (see Table 4-10)	IN	None	[0 … 255]
Code_Classif2	Ambiguity code that characterizes the ambiguity of the classification process, thus reflecting, at the pixel level, the reliability of the CLASSIF 1B	IN	None	[0 … 10]
CLASSIF_3	Land cover map resulting from the spectral classification approach (both supervised and unsupervised), where each pixel is associated with a land cover class through an ID (see Table 4-10)	OUT	None	[0 … 255]
Code_Classif3	Quality flag that characterizes, at the pixel level, the reliability of the CLASSIF_1	OUT	None	[0 … 100]
Source_Classif3	Flag that indicates, at the pixel level, if the land cover label is derived from the supervised or unsupervised classification algorithm	OUT	None	[1, 2]

4.5.4 Impact of multi-year approach

4.5.4.1 Multi-year strategies 1 and 2 in the classification chain

For the equal-reasoning areas concerned by the MY_S1 (see LUT 6 in Table 4-15), the seasonal composites used as input in the steps 1 (Table 4-21 and Table 4-24) and 2 (Table 4-26) need to be multi-year seasonal composites. To this end, the preliminary step described in section 4.3.1 consisting in generating multi-year seasonal composites need to be operated before running the steps 1 and 2.

For the equal-reasoning areas concerned by the MY_S2 (see LUT 6 in Table 4-15), the seasonal composites used as input in the steps 1 (Table 4-21 and Table 4-24) and 2 (Table 4-26) are single-year seasonal composites. Therefore, the classification algorithms (supervised in step 1 and unsupervised in step 2) need to be run several times (one for each year), thus producing several single-year spectral

land cover maps (CLASSIF1_Yn and CLASSIF2_Yn). Furthermore, the a posteriori step described in section 4.3.2 consisting in aggregating the single-year maps needs to be operated to produce the multi-year spectral land cover map.

4.5.4.2 Produding 3 maps and 1 "baseline" LC map

Knowing that the project aims at delivering 3 maps specific to epochs (centered on 2000, 2005 and 2010) and that these 2 maps are derived from a "baseline" land cover map, it means that the whole classification chain need to be run four times over 4 multi-year periods (Figure 4-15): the 1998-2002 period for generating the 2000 epoch map (using SPOT-VGT data), the 2003-2007 period for generating the 2005 epoch map (using MERIS data), the 2008-2012 period for generating the 2010 epoch map (using MERIS data) and the 2003-2012 period for generating the baseline map.

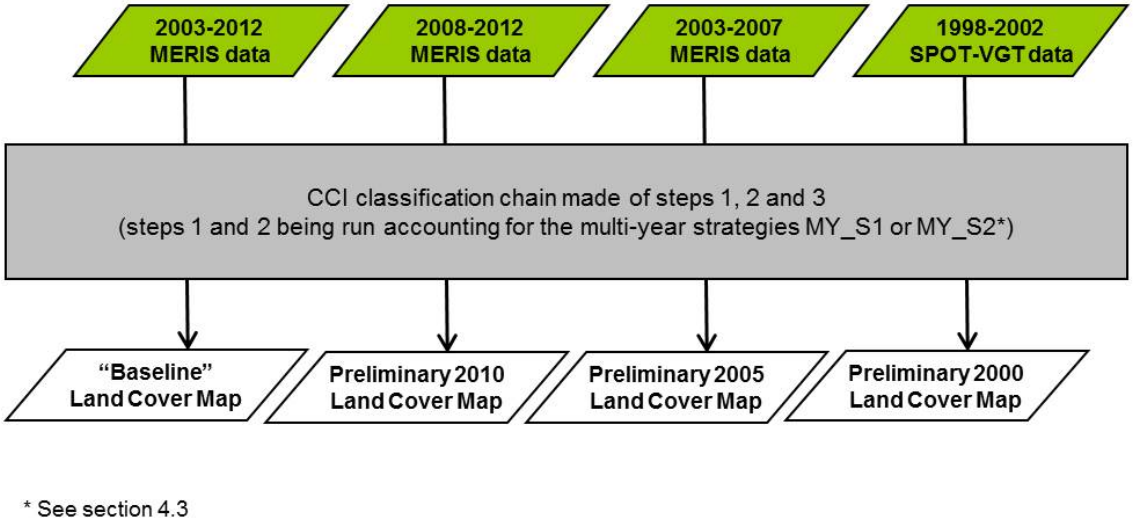


Figure 4-15: logic scheme illustrating the process developed to derive maps related to specific epochs from the "baseline" LC map

4.5.5 Fusion between MERIS FR and RR "baseline" land cover maps

If there are gaps in the MERIS FR "baseline" land cover map (i.e. the one generated from the full MERIS FR archive), they are filled in with the MERIS RR "baseline" land cover map.

Algorithm assumptions and limitations

None

Input and output data

Input and output data associated with this MERIS FR and RR fusion process are described in Table 4-30.

Table 4-30: Input and output data for the merging between the MERIS FR and RR "baseline" land cover maps

DATA	DESCRIPTION	INTENT (IN, OUT, INOUT)	PHYSICAL UNIT	RANGE
CLASSIF_3_ baseline_FR	Land cover map resulting CCI classification chain applied on MERIS FR seasonal composites, where each pixel is associated with a land cover class through an ID (see Table 4-10)	IN	None	[0 … 255]
Code_Classif3_ baseline_FR	Quality flag that characterizes, at the pixel level, the reliability of the CLASSIF_3_baseline_FR	IN	None	[0 … 100]
Source_Classif3_ baseline_FR	Flag that indicates, at the pixel level, if the land cover label of the CLASSIF_3_baseline_FR is derived from the supervised or unsupervised classification algorithm	IN	None	[1, 2]
CLASSIF_3_ baseline_RR	Land cover map resulting CCI classification chain applied on MERIS RR seasonal composites, where each pixel is associated with a land cover class through an ID (see Table 4-10)	IN	None	[0 … 255]
Code_Classif3_ baseline_RR	Quality flag that characterizes, at the pixel level, the reliability of the CLASSIF_3_baseline_RR	IN	None	[0 … 100]
Source_Classif3_ baseline_RR	Flag that indicates, at the pixel level, if the land cover label of the CLASSIF_3_baseline_RR is derived from the supervised or unsupervised classification algorithm	IN	None	[1, 2]
CLASSIF_3_ baseline_FR_RR	Land cover map resulting from the merging of the "baseline" LC maps obtained from the MERIS FR and RR seasonal composites, where each pixel is associated with a land cover class through an ID (see Table 4-10)	OUT	None	[0 … 255]
Code_Classif3_ baseline_FR_RR	Quality flag that characterizes, at the pixel level, the reliability of the CLASSIF_3_ baseline_FR_RR	OUT	None	[0 … 100]
Source_Classif3_ baseline_FR_RR	Flag that indicates, at the pixel level, if the land cover label is derived from the supervised or unsupervised classification algorithm and from the FR or RR seasonal composites	OUT	None	[1, 2]

4.5.6 Step 4 – Post-classification editions on the "baseline" LC map

The last step of the classification chain consists in bringing improvements thanks to existing external datasets. It should allow (i) correcting for some errors resulting from the algorithms and (ii) taking most benefit of existing high quality thematic products. With respect to the integration of existing external dataset, the main one concerns the integration of the CCI Water Bodies product, based on SAR dataset that will be generated in this project.

4.5.7 Deriving 3 epochs from the "baseline" land cover map

The "baseline" LC map has been generated using 10 years of data. It will serve as basis to derive LC maps representative of specific epochs. It is achieved through the comparison between the "baseline" LC map and the 5-year LC maps generated by the processing chain (see Figure 4-15), on a class-by-class basis.

In the current version, only the forest dynamics in the tropical regions have been detected and mapped. For the other LC classes, the 3 maps are identical to the "baseline" LC map.

Algorithm assumptions and limitations

None

Input and output data

Input and output data associated with this process are described in Table 4-31.

Table 4-31: Input and output data for the deriving a map specific to one epoch from the "baseline" LC map

DATA	DESCRIPTION	INTENT (IN, OUT, INOUT)	PHYSICAL UNIT	RANGE
CLASSIF_3_ baseline_FR_RR	Land cover map resulting from the merging of the "baseline" LC maps obtained from the MERIS FR and RR seasonal composites, where each pixel is associated with a land cover class through an ID (see Table 4-10)	IN	None	[0 … 255]
Code_Classif3_ baseline_FR_RR	Quality flag that characterizes, at the pixel level, the reliability of the CLASSIF_3_ baseline_FR_RR	IN	None	[0 … 100]
Source_Classif3_ baseline_FR_RR	Flag that indicates, at the pixel level, if the land cover label is derived from the supervised or unsupervised classification algorithm and from the FR or RR seasonal composites	IN	None	[1, 2]

DATA	DESCRIPTION	INTENT (IN, OUT, INOUT)	PHYSICAL UNIT	RANGE
CLASSIF_3_YYYY	Intermediate land cover map for the epoch YYYY resulting from the classification of MERIS FR seasonal composites (in the case of 2005 and 2010 epochs) or SPOT-VGT seasonal composites (in the case of 2000 epoch), where each pixel is associated with a land cover class through an ID (see Table 4-10)	IN	None	[0 … 255]
Code_Classif3_ YYYY	Quality flag that characterizes, at the pixel level, the reliability of the intermediate land cover map CLASSIF_3_YYYY	IN	None	[0 … 100]
Source_Classif3_ YYYY	Flag that indicates, at the pixel level, if the land cover label is derived from the supervised or unsupervised classification algorithm	IN	None	[1, 2]
CCI LAND COVER MAP EPOCH YYYY	Land cover map for the epoch derived from the “baseline” LC map, where each pixel is associated with a land cover class through an ID (see Table 4-10)	OUT	None	[0 … 255]
Code for CCI LAND COVER MAP EPOCH YYYY	Quality flag that characterizes, at the pixel level, the reliability of the CLASSIF_3_YYYY	OUT	None	[0 … 100]
Source for CCI LAND COVER MAP EPOCH YYYY	Flag that indicates, at the pixel level, if the land cover label is derived from the supervised or unsupervised classification algorithm	OUT	None	[1, 2]

5 LAND COVER CONDITION PRODUCTS PROCESSING CHAIN

5.1 Logical model for land cover condition products generation

In addition to the global land cover maps, the CCI-LC project will deliver four global land cover condition (LC-condition) products: the NDVI, snow, burnt area (BA) and water condition products over the 1999-2012 epoch. The water bodies condition is not delivered in this release version of the project.

On a per pixel basis, the LC-condition reflects, along the year, the average trajectory (or behaviour) and the intra-annual variability of a land surface feature (e.g. forest – the LC state) over a 14-year period. This variability is described by four observable variables: the green vegetation phenology, snow coverage, BA occurrence and open water presence, corresponding respectively to the four LC-condition products (NDVI, snow, BA and water). The LC-condition variables show how the LC state is temporarily modified (e.g. snow falls on forest) while the intrinsic characteristics of the LC state remain untouched (forest remains forest; although covered by snow).

The LC-condition products are expressed as time profiles of the mean and standard deviation for continuous variables (NDVI) or as percentages of occurrence for discrete variables (snow, BA and water).

The workflow of the creation of the LC-condition products is presented in Figure 5-1. The following sections describe in details the construction of each condition product and the way the compliance between them is tackled.

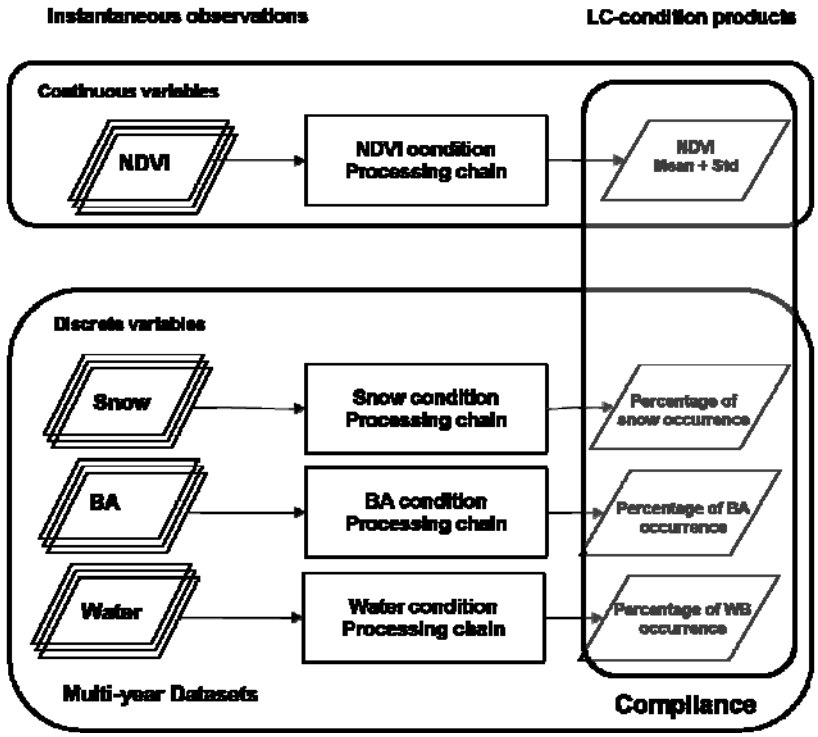


Figure 5-1: Schematic illustration of the construction of the LC-condition products chain for continuous (NDVI) and discrete variables (snow, BA and water).

5.2 Detailed processing scheme for NDVI LC-condition product

The production of the NDVI LC-condition is based on 14 years of SPOT-VGT reflectances. It covers the 1999-2012 period. The methodology consists in several steps as illustrated in Figure 5-2. In the first step of the product processing, the NDVI vegetation index is computed from the Red and NIR reflectance values. Second, the NDVI profiles are smoothed with a Wittaker filter. Finally, the resulting 14 years time series are aggregated over years, 7-day period by 7-day period. It results into a yearly mean and a yearly standard deviation profiles representative of a 1999-2012 years period.

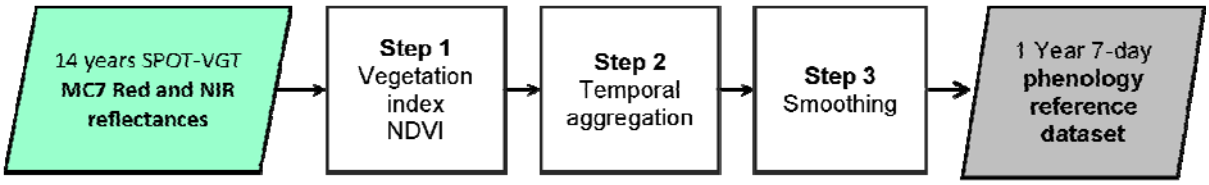


Figure 5-2: Schematic illustration of the processing for generating the NDVI LC-condition product.

This dataset consists of 728 7-day SPOT-VGT composites from January 1999 to December 2012 at a spatial resolution of 1 km. The NDVI LC-condition product is delivered at global scale.

5.2.1 Step 1 – Vegetation index NDVI

The NDVI is used as an indicator of the vegetation growing cycle. It is computed from the Red and NIR reflectance values of the 7-day time series. The status of the pixels does not change during this step. Only LAND status is processed.

Algorithm assumptions and limitations

It is assumed that the NDVI is the most appropriate index to evaluate the dynamic component of the vegetation at global scale.

Input and output data

Input and output data associated with this first step of vegetation index calculation are described in Table 5-1. Associated parameters are presented in Table 5-2.

Table 5-1: Input and output data of the step 1 of the NDVI LC product generation

DATA	DESCRIPTION	INTENT (IN, OUT, INOUT)	PHYSICAL UNIT	RANGE
L3_<source>_7d_mosaic_reflectance <source> = SPOT-VGT_1km	Normalized reflectance in each bands of the 7-day composites	IN	None	5*10e-4
L3_<source>_7d_mosaic_Status <source> = SPOT-VGT	Status of the 7-day composites	IN	None	[1 ... 5]
L3_<source>_7d_mosaic_Nobs <source> = SPOT-VGT_1km	Number of valid observations for the status LAND of the 7-day composites	ININ	None	[0 ... 7]
L3_<source>_7d_NDVI <source> = SPOT-VGT_1km	7-day NDVI composites	OUT	None	[0...1]
L3_<source>_7d_NDVI_Status <source> = SPOT-VGT_1km	Status of the 7-day NDVI composites	OUT	None	[1 ... 6]
L3_<source>_7d_NDVI_Nobs <source> = SPOT-VGT_1km	Number of valid observations for the status LAND of 7-day NDVI composites	OUT	None	[0 ... 7]

Table 5-2: Parameters needed in the step 2 of the NDVI LC product generation

PARAMETERS	DESCRIPTION	INTENT (IN, OUT, INOUT)	PHYSICAL UNIT	RANGE
b<n>	Spectral channels of the 7-day composite	IN	None	[1 … 15] for MERIS [B0, B2, B3, MIR] for SPOT-VGT

5.2.2 Step 3 – Temporal aggregation

This third step delivers the NDVI LC-condition product. The smoothed NDVI time series are temporally aggregated over the 13 years in order to define the reference annual trajectory of a pixel NDVI value and to explore its inter-annual variability.

Each pixel is characterized by (i) a synthetic reference profile, presenting the pattern of variation of NDVI for one year and (ii) a temporal standard deviation profile, representing the inter-annual variability during the investigated period. The standard deviation informs on how much variation in vegetation growth occurs over the investigated period.

The 13-year time series is decomposed year by year (Figure 5-3). Then for a specific 7-day period, the 13 NDVI values (corresponding to the 13 years) are temporally aggregated by taking the mean value. A temporal standard deviation is computed as well for each 7-day period.

As for the auxiliary information:

- the number of valid observations for each 7-day period is summed over the 13 years;
- the source of the NDVI value for each 7-day period is kept by adding two layers, one for MERIS and one for SPOT-VGT. Each layer contains a count by 7-day period of the number of years where the NDVI observation originates from the specific sensor.
- a layer with the count of valid years used in the final output is also produced.

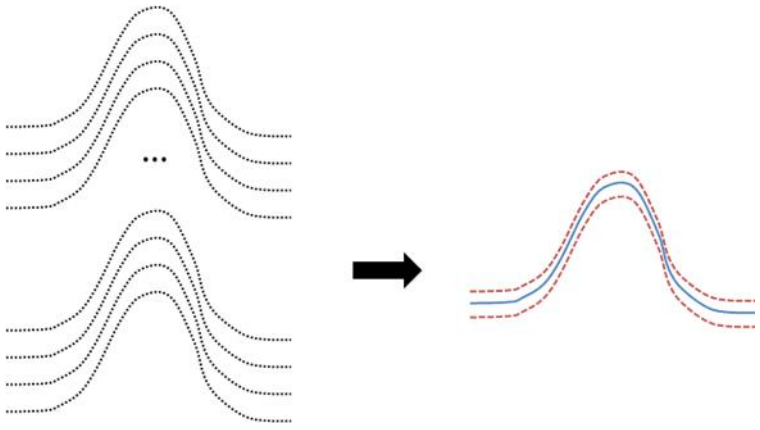


Figure 5-3: Mean and standard deviation profiles from a time series of 13 years

Algorithm assumptions and limitations

The assumption is made that pixels selected for the temporal aggregation are not affected by land cover change between 1999 and 2011. It is also assumed that there are enough values over 13 years to compute consistent mean and standard deviation statistics. Otherwise, a threshold will have to be set to determine which pixel is processed.

Input and output data

Input and output data associated with this last step of temporal aggregation are described in Table 5-3. Associated parameters are presented in Table 5-4.

Table 5-3: Input and output data of the step 3 of the NDVI LC product generation

DATA	DESCRIPTION	INTENT (IN, OUT, INOUT)	PHYSICAL UNIT	RANGE
L3_<source>_7d_NDVI <source> = SPOT-VGT_1km	7-day NDVI composites	IN	None	[0…1]
L3_<source>_7d_NDVI_Status <source> = SPOT-VGT_1km	Status of the 7-day NDVI composites	IN	None	[1 … 6]
L3_<source>_7d_NDVI_Nobs <source> = SPOT-VGT_1km	Number of valid observations for the status LAND of 7-day NDVI composites	IN	None	[0 … 7]
L4_NDVI_SPOT-VGT_1km_ 7d_Agg_Mean	NDVI 7-day reference mean value for one year	OUT	None	[0…1]
L4_NDVI_SPOT-VGT_1km_ 7d_Agg_Std	NDVI 7-day reference standard deviation value for one year	OUT	None	[0…1]

Table 5-4: Parameters needed in the step 3 of the NDVI LC product generation

PARAMETERS	DESCRIPTION	INTENT (IN, OUT, INOUT)	PHYSICAL UNIT	RANGE
Ybegin	Year of begin	IN	None	[1999…2011]
Yend	Year of end	IN	None	[1999…2011]

5.2.3 Step 4 – Smoothing

Even if the spatial and temporal aggregations reduce the noise and remove extreme observations, the 13 years of combined NDVI profiles are smoothed using the Wittaker smoother [RD-141]. The status and the valid observations number remain unchanged.

[RD-143] presented an actualized version of the algorithm published by [RD-144]. The Wittaker smoother fits a discrete series to discrete data and puts a penalty on the roughness of the smooth curve. The algorithm is extremely fast, gives continuous control over smoothness with only one parameter and automatically interpolates. It adapts well to boundaries (e.g. start and end of the time series) and does not assume periodicity such as the Fourier filtering or wavelets [RD-142]. A more detailed description of the algorithm can be found in the original work [RD-143; RD-141].

Algorithm assumptions and limitations

The assumption is made that the filter output is a realistic pattern of NDVI.

Input and output data

Input and output data associated with this fourth step of smoothing are described in Table 5-5. Associated parameters are presented in Table 5-6.

Table 5-5: Input and output data of the step 4 of the NDVI LC product generation

DATA	DESCRIPTION	INTENT (IN, OUT, INOUT)	PHYSICAL UNIT	RANGE
L4_NDVI_SPOT-VGT_1km_ 7d_Agg_Mean	NDVI 7-day reference mean value for one year	IN	None	[0…1]
L4_smoothed_7d_NDVI	13 years of smoothed 7-day NDVI profiles	OUT	None	[0…1]

Table 5-6: Parameters needed in the step 4 of the NDVI LC product generation

PARAMETERS	DESCRIPTION	INTENT (IN, OUT, INOUT)	PHYSICAL UNIT	RANGE
λ	Wittaker smoothing parameter	IN	None	[0…x]

5.3 Detailed processing scheme for the burnt area LC-condition product

The BA condition product presents the percentage at which burnt areas have been detected along the year on given pixel, based on observations over the 2000-2012 period. It provides the percentage of BA occurrence with a 7-day temporal resolution and depicts the seasonal dynamic of the fire impact on the surface, at 500m spatial resolution.

Two parallel processing chains leading to BA LC-condition products are detailed in Figure 5-4. The final BA LC-condition product is to be constructed based on the Level 3 CCI-Fire disturbance product generated by the ESA CCI Fire project [RD-147]. However, it will be available at the end of the CCI phase 1, which prevents its direct use for the moment. Meanwhile, the BA input data are currently derived from the MODIS Direct Broadcast Monthly Burned Area Product (MCD64A1) [RD-12] being part of the Global Fire Emissions Database version 3 (GFED.v3) products [RD-152].

Two preliminary steps are included in the product processing chain, consisting in format conversion and mosaicking. They are necessary for the application of further algorithms. Their representation, in dotted lines, shows their secondary importance. As the final BA LC-condition product is presented on a 7-day basis, a compositing algorithm is applied in step 1 to the monthly aggregated input data. Step 2 temporarily aggregates BA observations over the 13 years and leads to the BA condition product itself. Finally, the product is projected to a Plate Carrée projection with a geographic Lat/Long representation based on the WGS84 ellipsoid.

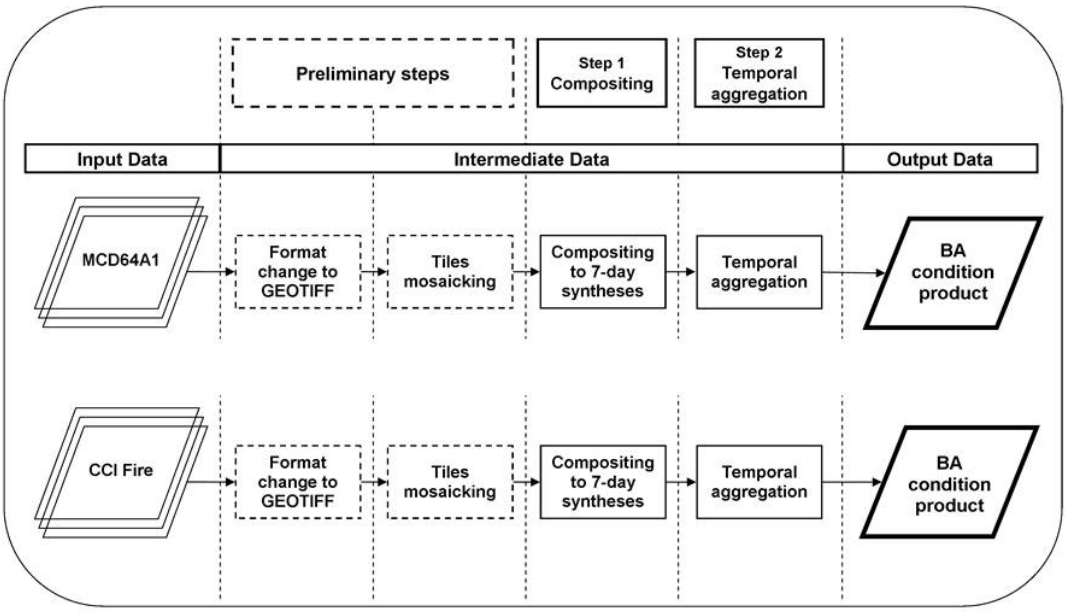


Figure 5-4: Schematic illustration of the processing chain of the BA LC-condition product construction.

5.3.1 Preliminary steps

The first preliminary step of the BA LC-condition product generation is the format harmonization of the input data to GeoTiff. This GeoTiff format is far more efficiently supported than the HDF one. The conversion applies to the MCD64A1 and L3 CCI Fire products are both in the Hierarchical Data Format (HDF).

The second preliminary step consists in mosaicking the tiles of the MCD64A1 and L3 CCI Fire (outputs of the previous preliminary step) to global coverage products.

The preliminary steps rely on the GDAL library (see details in section 4.2), and more precisely on the following the "GDAL_Translate" and "GDAL_Merge" functions. The first one allows converting raster data between different formats. It uses as parameters the output format (which is GeoTiff in our case), the compression type, the image structure information and the names of the input and output files. The second one allows appending several adjacent raster files together to form a unique raster file. It uses as parameters the name of the output file, the output format (which is GeoTiff by default), the compression type, the image structure information and the names of the input tiles.

Input and output data

Input and output data of the two preliminary steps are described in Table 5-7. The parameters associated with the format conversion and the mosaicking steps for the MCD64A1 and L3 CCI Fire products are presented in Table 5-8 and Table 5-9, respectively.

Table 5-7: Input and output data for the 2 preliminary steps (BA LC-condition product)

DATA	DESCRIPTION	INTENT (IN, OUT, INOUT)	PHYSICAL UNIT	RANGE

DATA	DESCRIPTION	INTENT (IN, OUT, INOUT)	PHYSICAL UNIT	RANGE
MCD64A1				
Date	Tile raster layer in HDF format with the ordinal day of BA detection	IN	DOY	[-1 … 366]
Fire CCI				
Band 1	Tile raster layer in HDF5 format with the date of the first BA detection	IN	DOY	[1 … 365]
Band 2	Tile raster layer in HDF5 format with the sensor detecting the BA	IN	dl	[0 … 7]
Band 3	Tile raster layer in HDF5 format with the confidence level of the detection	IN	dl	[0 … 100]
MCD64A1				
L3_BA_MCD64A1_500m_ 1m_mosaic_Date	Global Raster mosaic in GeoTiff format with the ordinal day of BA detection	OUT	DOY	[-1 … 366]
Fire CCI				
L3_BA_CCI_300m_ 1m_mosaic_Date	Global Raster mosaic in GeoTiff format with the date of the first BA detection	OUT	DOY	[1 … 365]
L3_BA_CCI_300m_ 1m_mosaic_s<sensor> with <sensor>: 01to 07 [AD-11]	Global Raster mosaic in GeoTiff format with the sensor detecting the BA	OUT	dl	[0 … 7]
L3_BA_CCI_300m_ 1m_mosaic_QA	Global Raster mosaic in GeoTiff format with the confidence level of the detection	OUT	dl	[0…100]

Table 5-8: Parameters needed to run the format conversion preliminary step for the MCD64A1 and L3 CCI Fire products (BA LC-condition product)

PARAMETERS	DESCRIPTION	INTENT (IN, OUT, INOUT)	FORMAT	RANGE
Format	Output format	IN	String	/
Compress	Compression type	IN	String	/

PARAMETERS	DESCRIPTION	INTENT (IN, OUT, INOUT)	FORMAT	RANGE
Interleave	Image structure information	IN	String	/
Tiled	Image structure information	IN	String	/

Table 5-9: Parameters needed to run the mosaicking preliminary step for the the MCD64A1 and L3 CCI Fire products (BA LC-condition product)

PARAMETERS	DESCRIPTION	INTENT (IN, OUT, INOUT)	FORMAT	RANGE
Input	Name of the input file	IN	String	/
Output	Name of the output file	IN	String	/
Compress	Compression type	IN	String	/
Interleave	Image structure information	IN	String	/
Tiled	Image structure information	IN	String	/
Tiles	Name of the input tiles	IN	String	/

5.3.2 Step 1 – Compositing

The first step of the construction of the BA LC-condition product aims at producing 7-day composites from monthly, global rasters. This step is identical for all BA products.

For each monthly raster, four empty rasters with the same charcteristics are created. These correspond to the division of the month into four weeks. Each pixel of the monthly raster files with the dates of the BA detections is scanned iteratively. According to the date, the date value is added in the right 7-day synthese.

Input and output data

Input and output data associated with this compositing step are described in Table 5-10. The parameters associated with this step are presented in Table 5-11. This table contains the following fields:

- $Month_i$ giving the monthly composite (input) to which the number of the 7-day BA composite i is related (output)
- $StartDate_i$ specifying the exact date which marks the start of the 7-day BA composite i
- $StartDOY_i$ giving the day of the year which marks the start of the 7-day BA composite i
- $EndDate_i$ specifying the exact date which marks the end of the 7-day BA composite i
- $EndDOY_i$ giving the day of the year which marks the end of the 7-day BA composite i

Table 5-10: Input and output data of the step 1 of the BA LC-condition product generation.

DATA	DESCRIPTION	INTENT (IN, OUT, INOUT)	PHYSICAL UNIT	RANGE

DATA	DESCRIPTION	INTENT (IN, OUT, INOUT)	PHYSICAL UNIT	RANGE
MCD64A1				
L3_BA_MCD64A1_500m_ 1m_mosaic_Date	Monthly raster mosaic in GeoTiff format with the ordinal day of BA detection	IN	DOY	[-1 … 366]
Fire CCI				
L3_BA_CCI_300m_ 1m_mosaic_Date	Monthly raster mosaic in GeoTiff format with the date of the first BA detection	IN	DOY	[1 … 365]
L3_BA_CCI_300m_ 1m_mosaic_s<sensor> with <sensor>: 01to 07 [AD-11]	Monthly raster mosaic in GeoTiff format with the sensor detecting the BA	IN	dl	[0 … 7]
L3_BA_CCI_300m_ 1m_mosaic_QA	Monthly raster mosaic in GeoTiff format with the confidence level of the detection	IN	dl	[0…100]
MCD64A1				
L3_BA_MCD64A1_500m_ 7d _comp_Date	7-day raster composite with the ordinal day of BA detection	OUT	DOY	[-1 … 366]
Fire CCI				
L3_BA_CCI_300m_ 7d_comp_Date	7-day raster composite with the date of the first BA detection	OUT	DOY	[1 … 365]
L3_BA_CCI_300m_ 7d _comp_s<sensor> with <sensor>: 01to 07 [AD-11]	7-day raster composite with the sensor detecting the BA	OUT	dl	[0 … 7]
L3_BA_CCI_300m_ 7d_comp_QA	7-day raster composite with the confidence level of the detection	OUT	dl	[0…100]

Table 5-11: Parameters needed in the compositing step of the BA LC-condition product generation

7-DAY BA COMPOSITE NUMBER	PARAMETERS				
	$MONTH_i$	$STARTDATE_i$	$STARTDOY_i$	$ENDDATE_i$	$ENDDOY_i$
01	1	YYYY0101	1	YYYY0107	7

7-DAY BA COMPOSITE NUMBER	PARAMETERS				
	$MONTH_i$	$STARTDATE_i$	$STARTDOY_i$	$ENDDATE_i$	$ENDDOY_i$
02	1	YYYY0108	8	YYYY0114	14
03	1	YYYY0115	15	YYYY0121	21
04	1	YYYY0122	22	YYYY0128	28
05	1, 2	YYYY0129	29	YYYY0204	35
06	2	YYYY0205	36	YYYY0211	42
07	2	YYYY0212	43	YYYY0218	49
08	2	YYYY0219	50	YYYY0225	56
09	2,3	YYYY0226	57	YYYY0304	63
10	3	YYYY0305	65	YYYY0311	71
11	3	YYYY0312	72	YYYY0318	78
12	3	YYYY0319	79	YYYY0325	85
13	3,4	YYYY0326	86	YYYY0401	92
14	4	YYYY0402	93	YYYY0408	99
15	4	YYYY0409	100	YYYY0415	106
16	4	YYYY0416	107	YYYY0422	113
17	4	YYYY0423	114	YYYY0429	120
18	4,5	YYYY0430	121	YYYY0506	127
19	5	YYYY0507	128	YYYY0513	134
20	5	YYYY0514	135	YYYY0520	141
21	5	YYYY0521	142	YYYY0527	148
22	5,6	YYYY0528	149	YYYY0603	155
23	6	YYYY0604	156	YYYY0610	162
24	6	YYYY0611	163	YYYY0617	169
25	6	YYYY0618	170	YYYY0624	176
26	6,7	YYYY0625	177	YYYY0701	183
27	7	YYYY0702	184	YYYY0708	190
28	7	YYYY0709	191	YYYY0715	197
29	7	YYYY0716	198	YYYY0722	204
30	7	YYYY0723	205	YYYY0729	211
31	7,8	YYYY0730	212	YYYY0805	218
32	8	YYYY0806	219	YYYY0812	225

7-DAY BA COMPOSITE NUMBER	PARAMETERS				
	$MONTH_i$	$STARTDATE_i$	$STARTDOY_i$	$ENDDATE_i$	$ENDDOY_i$
33	8	YYYY0813	226	YYYY0819	232
34	8	YYYY0820	233	YYYY0826	239
35	8,9	YYYY0827	240	YYYY0902	246
36	9	YYYY0903	247	YYYY0909	253
37	9	YYYY0910	254	YYYY0916	260
38	9	YYYY0917	261	YYYY0923	267
39	9,10	YYYY0924	268	YYYY1000	274
40	10	YYYY1001	275	YYYY1007	281
41	10	YYYY1008	282	YYYY1014	288
42	10	YYYY1015	289	YYYY1021	295
43	10	YYYY1022	296	YYYY1028	302
44	10,11	YYYY1029	303	YYYY1104	309
45	11	YYYY1105	310	YYYY1111	316
46	11	YYYY1112	317	YYYY1118	323
47	11	YYYY1119	324	YYYY1125	330
48	11,12	YYYY1126	331	YYYY1202	337
49	12	YYYY1203	338	YYYY1209	344
50	12	YYYY1210	345	YYYY1216	351
51	12	YYYY1217	352	YYYY1223	358
52	12	YYYY1224	359	YYYY1231	366

5.3.3 Step 2 – Temporal aggregation

The last step of the processing chain is the construction of the BA LC-condition product itself. The BA condition product presents the frequency at which burnt areas have been detected along the year on a given pixel, based on observations over the 2000-2012 period. It provides the percentage of BA occurrence with a 7-day temporal resolution and depicts the seasonal dynamic of the fire impact on the surface. This step applies to the BA intermediate poduct resulting from step 1 as well as to the CCI-Fire product.

At the end of step 2, the BA condition product is composed of two layers. The first one represents the percentage of BA occurrence, on a 7-day basis, calculated as the sum of BA detections over the number of years in the aggregation period (currently 13 years – 2000 to 2012). It is expressed between 0 and 100. The second layer gives, on a 7-day basis, the number of valid years contributing to each 7-day period. It stands for a quality indicator of the occurrence values and is expressed between 0 and

13. Then, the product is projected to a Plate Carrée projection with a geographic Lat/Long representation based on the WGS84 ellipsoid.

Input and output data

Input and output data associated with this temporal aggregation step are described in Table 5-12. No parameter is associated with this processing step.

Table 5-12: Input and output data of the step 2 of the BA LC-condition product generation

DATA	DESCRIPTION	INTENT (IN, OUT, INOUT)	PHYSICAL UNIT	RANGE
Product resulting from step 1				
L3_BA_MCD64A1_500m_ 7d _comp_Date	7-day raster composite with the ordinal day of BA detection	IN	dl	[1 … 366]
Fire CCI				
L3_BA_CCI_300m_ 7d_comp_Date	7-day raster composite with the date of the first BA detection	IN	DOY	[1 … 365]
L3_BA_CCI_300m_ 7d _comp_s<sensor> with <sensor>: 01to 07 [AD-11]	7-day raster composite with the sensor detecting the BA	IN	dl	[0 … 7]
L3_BA_CCI_300m_ 7d_comp_QA	7-day raster composite with the confidence level of the detection	IN	dl	[0…100]
Aggregated product				
L4_BA_ MCD64A1_500m_ 7d_AggOcc	Percentage of Burnt Areas occurrence as detected over the 2000-2012 period on a 7-day basis (ranging from 0 to 100). This describes a reference of the burnt areas dynamic at a 7-day frequency.	OUT	dl	[0 … 100]
L4_BA_ MCD64A1_500m_ 7d_ NYearObs	Number of valid years contributing to each 7-day period of the AggOcc series. It is a quality indicator of the occurrence values.	OUT	dl	[0 … 13]
Fire CCI				
L4_BA_CCI_300m_7d_Agg_Occ	Normalized occurrence of BA detections over the available 5 years (1999, 2000, 2002, 2003	OUT	dl	[0 … 100]

DATA	DESCRIPTION	INTENT (IN, OUT, INOUT)	PHYSICAL UNIT	RANGE
	and 2005)			
L4_BA_CCI_300m_7d_Agg_QA	Normalized level of condidence over the available 5 years (1999, 2000, 2002, 2003 and 2005)	OUT	dl	[0 … 100]

This has lead into extending the effect of the original BA detection for 4 weeks after the burnt and providing a condition product on a monthly basis (as in the input data). The number and valid cloud-free observations layer remain unchanged.

Then, the product is projected to a Plate Carrée projection with a geographic Lat/Long representation based on the WGS84 ellipsoid.

Input and output data

Input and output data associated with this temporal aggregation step are described in Table 5-12. No parameter is associated with this processing step.

Table 5-13: Input and output data of the step 3 of the BA LC-condition product generation

DATA	DESCRIPTION	INTENT (IN, OUT, INOUT)	PHYSICAL UNIT	RANGE
L4_BA_ MCD64A1_500m_ 7d_AggOcc	Percentage of Burnt Areas occurrence as detected over the 2000-2012 period on a 7-day basis (ranging from 0 to 100). This describes a reference of the burnt areas dynamics at a 7-day frequency.	IN	dl	[0 … 100]
L4_BA_ MCD64A1_500m_ 7d_AggOcc_4Weeks	Percentage of Burnt Areas occurrence as detected over the 2000-2012 period on a 7-day basis and on which a temporal filter increasing the duration of the burnt for 4 weeks has been applied (ranging from 0 to 100). This describes a spatially more consistent reference of the burnt areas dynamics at a 7-day frequency.	OUT	dl	[0 … 100]

5.4 Detailed processing scheme for the snow LC-condition product

The snow condition product presents the proportion of snow occurrence detected, along the year based on observations over the 2000-2012 period. It provides the seasonal dynamic behaviour of snow coverage with a 7-day temporal resolution at 500m spatial resolution.

The snow input data originates from the 8-day MOD10A2 product [RD-145]. Data are respectively available freely from 2000 up to now at global scale. All products come in tiles of 1200*1200km and share the same technical properties: a sinusoidal projection and HDF format.

The processing chain leading to the snow LC-condition product is detailed in Figure 5-5. The "Format harmonization" and "Mosaicking" steps are preliminary steps necessary for the application of further algorithms. Their representation, in dotted lines, shows their secondary importance. The first step, consisting in a spatial filtering, aims at reducing the number of NoData observations due to the presence of clouds in order to consolidate the estimators. Step 2 temporarily aggregates snow observations over 13 years to obtain the percentages of snow occurrence. Step 3 consists in a temporal resampling as the snow condition product needs to be presented on a 7-day basis but the original time series is made of 8-day periods. Then, a temperature filter is applied on the 7-day temporarily aggregated product in order to reduce the existing commission errors. Finally, the product was reprojected to Plate Carrée with a geographic Lat/Long representation based on the WGS84 ellipsoid.

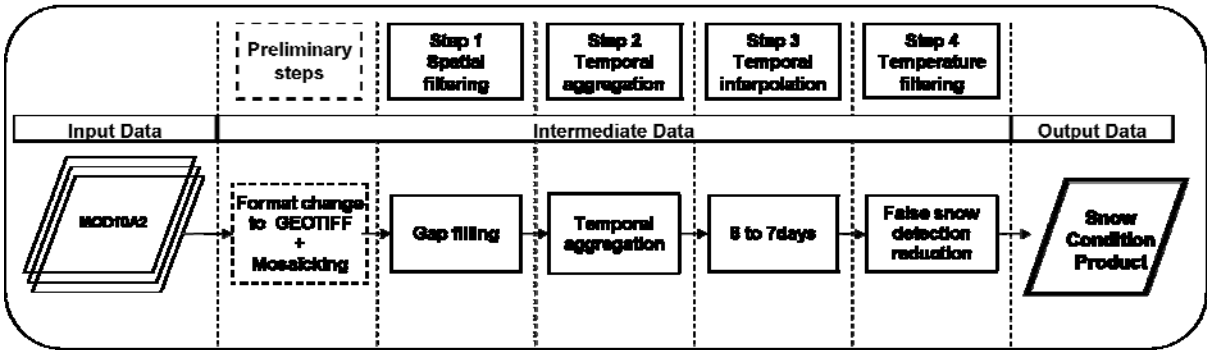


Figure 5-5: Schematic illustration of the processing chain of the snow LC-condition product construction.

5.4.1 Preliminary step

The first preliminary step of the snow LC-condition product generation is the harmonization of the format of input data from HDF to GeoTiff. This is due to the fact that GeoTiff is far more efficiently supported than the HDF. The second preliminary step consists in mosaicking the 1200*1200 km tiles (output of the format conversion step) in global raster files.

The two preliminary steps rely on the GDAL library (see detail in section 4.2), and more precisely on the following the "GDAL_Translate" and "GDAL_Merge" functions. The first one allows converting raster data between different formats. It uses as parameters the output format (which is GeoTiff in our case), the compression type, the image structure information and the names of the input and output files. The second one allows appending several adjacent raster files together to form a unique raster file. It uses as parameters the name of the output file, the output format (which is GeoTiff by default), the compression type, the image structure information and the names of the input tiles.

Algorithm assumptions and limitations

The "GDAL_merge" algorithm makes the hypotheses that all tiles share the same coordinate system and have a matching number of bands.

Input and output data

Input and output data of the two preliminary steps are described in Table 5-14. The parameters associated with the format conversion and the mosaicking steps are presented in Table 5-15 and Table 5-16, respectively.

Table 5-14: Input and output data for the 2 preliminary steps (snow LC-condition product)

DATA	DESCRIPTION	INTENT (IN, OUT, INOUT)	PHYSICAL UNIT	RANGE
Full archive of MOD10A2 tiles from 2000 to 2012 in HDF format.	It includes our science data set of interest : Maximum_Snow_Extent (MSE) [AD 11]	IN	dl	[0 … 255]
L3_MOD10A2_500m_8d_MSE.tif	Global raster layer in GeoTiff format with the Maximum Snow Extent	OUT	dl	[0…255]

Table 5-15: Parameters needed to run the format conversion preliminary step (snow LC-condition product)

PARAMETERS	DESCRIPTION	INTENT (IN, OUT, INOUT)	FORMAT	RANGE
Input	Name of the input file	IN	String	/
Output	Name of the output file	IN	String	/
Format	Output format	IN	String	/
Compress	Compression type	IN	String	/
Interleave	Image structure information	IN	String	/
Tiled	Image structure information	IN	String	/

Table 5-16: Parameters needed to run the mosaicking preliminary step (snow LC-condition product)

PARAMETERS	DESCRIPTION	INTENT (IN, OUT, INOUT)	FORMAT	RANGE
Tiles	Name of the input tiles	IN	String	/
Output	Name of the output file	IN	String	/
Compress	Compression type	IN	String	/
Interleave	Image structure information	IN	String	/
Tiled	Image structure information	IN	String	/

5.4.2 Step 1 – Spatial filtering

The first step of the processing chain is the filtering of the L3_MOD10A2_500m_8d_MSE products of the preliminary steps to conservatively remove small clouds.

Based on the preliminary results, it was decided to constraint the maximum size of filtered clouds to a 5 by 5 pixels square (~6.25 km²) and to set that the cloud must not exceed a given size and must be totally surrounded by the same class. In practice, this filter is based on two passes:

1. Assign a new status at the center of small clouds that are surrounded by the same non cloud status. This is done by verifying that the perimeter of the window does not include any cloudy pixel and that all non cloud pixels within the window belong to the same class. It is worth noting that the radius of the processing window has to be one pixel larger than the maximum size of the cloud to be removed so that the surrounding of the cloud can be analyzed.
2. Expand the new snow status values to all cloud pixels connected to a pixel with a new snow status. When a cloud pixel has been identified as the center of a cloud surrounded by a single class, all the pixels actually connected to this central pixels meet the conditions for being removed. The second pass of the filter therefore expands the central pixel in order to cover the full cloud extent.

The result of this filter is similar to a classical erosion filter, but keeps the no data values in ambiguous cases such as the border of a large cloud. Furthermore, due to the strict obedience to the rule of a single class around the central pixel, the filter prevents from guessing a class in case of heterogeneous landscape. However, it iteratively runs with increasing radius size in order to increase the chances of removal of small clouds. Indeed, the probability to have a single class inside a square neighbourhood decreases with the size of this neighbourhood.

Algorithm assumptions and limitations

It is assumed that the sow cover does not rapidly change in space. However, we are aware that small isolated patches could be observed on the ground. Those patches however remain extremely rare according to our analysis.

Input and output data

Input and output data associated with this compositing step are described in Table 5-17. The parameters associated with this step are presented in Table 5-18.

Table 5-17: Input and output data of the step 1 of the Snow LC-condition product generation

DATA	DESCRIPTION	INTENT (IN, OUT, INOUT)	PHYSICAL UNIT	RANGE
L3_MOD10A2_500m_8d_MSE.tif	Raster layer in GTiff format with classes snow and ice	IN	/	[0...255]

DATA	DESCRIPTION	INTENT (IN, OUT, INOUT)	PHYSICAL UNIT	RANGE
L3_MOD10A2_500m_8d_MSE_filtered.tif	Raster layer in GTiff format with classes snow and ice, filtered for small clouds	OUT	/	[0 ... 255]

Table 5-18: Parameters needed in the step 1 of the Snow LC-condition product generation

PARAMETERS	DESCRIPTION	INTENT (IN, OUT, INOUT)	PHYSICAL UNIT	RANGE
Input	Name of the input file	IN	String	/
Output	Name of the output file	OUT	String	/
Ni_eight	Date which marks the start of the 7-day period	IN	None	[1…46]
SRmax	Maximum size of the radius for the filtering	IN	None	[1 2 3]
variables	Variables created in the algorithm are defined in the pseudo-code representation	INOUT	dl	

5.4.3 Step 2 – Temporal aggregation

The second step of the processing chain is the construction of the snow LC-condition product itself. The snow LC-condition product is an interpreted snow indicator that gives, for each pixel, the percentage occurrence of snow coverage. It is composed of 8-day time series of 1 variable: the average annual reference trajectory of snow behaviour and one quality-oriented variable to inform on the robustness of the occurrence.

For each valid pixel of each 8-day period, the percentage of snow occurrence is first determined by the calculation of the sum of the number of snow covered days over the 13 years weighed by the number of valid observations. The output ranges from 0 (no snow was detected in 13 years) to 100 (permanent snow).

Algorithm assumptions and limitations

It is assumed that the valid obsevartions are correct and that at least one snow observation is needed to flag the period as “snow”.

Input and output data

Input and output data associated with this temporal aggregation step are described in Table 5-19. Associated parameters are presented in Table 5-20.

Table 5-19: Input and output data of the step 2 of the Snow LC-condition product generation

DATA	DESCRIPTION	INTENT (IN, OUT, INOUT)	PHYSICAL UNIT	RANGE
L3_MOD10A2_500m_8d_MSE_filtered.tif	Maximum snow extent classification in the 8-day filled snow product.	IN	dl	[0 ... 255]
L4_MOD10A2_500m_8d_WAverage13y	8-day synthesis of the average number of snow covered days weighed by the number of valid observations over the 13 years period	OUT	percent	[0 … 100]
L4_MOD10A2_500m_8d_NYearObs13y	8-day synthesis of the total number of years when snow was detected	OUT	dl	[0 … 13]

Table 5-20: Parameters needed in the step 2 of the Snow LC-condition product generation

PARAMETERS	DESCRIPTION	INTENT (IN, OUT, INOUT)	PHYSICAL UNIT	RANGE
Inputs	Name of the input files	IN	String	/
Outputs	Name of the output files	OUT	String	/
Ni_eight	Number of the eight day period	IN	None	[1 … 46]
variables	Variables created in the algorithm are defined in the pseudo-code representation [AD 12]	INOUT	dl	

5.4.4 Step 3 – Temporal interpolation

In the framework of the CCI-LC project, the convention for the aggregation of observations is set to 7 days. The temporal interpolation algorithm enables fulfilling this requirement. The 8-days time series are converted into 7-days time series using a linear interpolation.

This step can be split in two parts. The first part consists in finding the two eight-day periods that intersect each seven-day period. The second part is the temporal interpolation itself.

In order to find the eight day periods corresponding to the seven-day periods, the calendar day of the middle of the seven-day period is computed, then the two closest eight day periods are identified. There are two special cases for this rule: for the first week, the 7-day period is assumed to be identical to the first 8-day period; for the last week, the average between the last and the first 8-day periods is computed.

A linear interpolation was selected because it is robust and fast. Due to the small distance in time between the interpolated values (maximum 3.5 days to the closest date), little difference are expected between a linear interpolation and more advanced interpolation methods. Finally the product is reprojected to a Plate Carree projection with a geographic Lat/Long representation based on the WGS84 ellipsoid.

Table 5-28: Input and output data of the step 3 of the Snow LC-condition product generation

DATA	DESCRIPTION	INTENT (IN, OUT, INOUT)	PHYSICAL UNIT	RANGE
L4_Snow_500m_8d_WAverage13y	8-day synthesis of the average number of snow covered days weighed by the number of valid observations over the 13 years period	IN	percent	[0 … 100]
L4_Snow_500m_8d_NYearObs13y	8-day synthesis of the total number of years when snow was detected	IN	dl	[0 … 13]
L4_Snow_500m_7d_WAverage13y	7-day synthesis of the average number of snow covered days weighed by the number of valid observations over the 13 years period	OUT	percent	[0 … 100]
L4_Snow_500m_7d_NYearObs13y	7-day synthesis of the total number of years when snow was detected	OUT	dl	[0 … 13]

Table 5-29: Parameters needed in the step 3 of the Snow LC-condition product generation

PARAMETERS	DESCRIPTION	INTENT (IN, OUT, INOUT)	PHYSICAL UNIT	RANGE
Inputs	Name of the input files	IN	String	/
Outputs	Name of the output files	OUT	String	/
Ni_seven	Number of the week	IN	None	[1 … 52]
variables	Variables created in the algorithm are defined in the pseudo-code representation [AD 12]	INOUT	dl	

5.4.5 Step 4 – Temperature filter

Although the MODIS suite of snow products is qualified by high accuracy [RD-160, RD-161], it is also widely agreed that commission errors are present due to the high frequency of cloud cover and the presence of bright areas that are counfounded with snow ([RD-153, RD-154, RD-155, RD-156, RD-157, RD-158]. This is also visible in the 7-day synthesis of the average number of snow covered days resulting from step 3.

Therefore, a temperature filter was applied on these 7-day syntheses in order to reduce the percentage of snow occurrence in regions/periods mostly unprobable, as defined by the expected temperature in the area. The likelihood of a snow event has been evaluated by taking into consideration the probability of false detection and the probability of snow event conditionally to the temperature. Indeed, snow is less likely to be present when the temperature increases but it is also unlikely that the classification algorithm fails on the same pixel several times. For each week, the monthly averaged temperature from the 'WorldClim' database is used to detect potential commission errors. The candidate commission errors are then confirmed if the snow probability of the area is low in order to give the priority to the classification algorithm when there are multiple detections. In practice, a linear combination of temperature and snow probability (temperature – 20*probability) is compared with a threshold value of 13. This value has been defined globally based on qualitative and quantitative analysis of the raw products.

The application of such filter improves the spatial consistency of the snow condition product and removes the majority of the artefacts in cloudy areas.

Algorithm assumptions and limitations

It is assumed that the compiled monthly average temperatures over the 1950-2000 period can be reliably used on the 7-day syntheses of the percentage of snow occurrence although the divergence of spatio/temporal resolution of both series of products.

In addition, it is assumed that using data measured at weather stations from a large number of global, regional, national, and local sources that are then interpolated, gives more robust results than current EO-based land surface temperature data. Indeed, EO-based LST are affected by clouds and therefore invalid when clouds are confused with snow.

Input and output data

Table 5-28: Input and output data of the step 4 of the Snow LC-condition product generation

DATA	DESCRIPTION	INTENT (IN, OUT, INOUT)	PHYSICAL UNIT	RANGE
L4_Snow_500m_7d_ WAverage13y	7-day synthesis of the average number of snow covered days weighed by the number of valid observations over the 13 years period	IN	percent	[0 … 100]

DATA	DESCRIPTION	INTENT (IN, OUT, INOUT)	PHYSICAL UNIT	RANGE
WorldClimTemperature_ 1km_1m_Average50y	monthly average temperature compiled from 1950 to 2000	IN	10°C (scaling factor 0.1)	[-513 … 338]
L4_Snow_500m_7d_ WAverage13y_Cleaned	Cleaned 7-day synthesis of the average number of snow covered days weighed by the number of valid observations over the 13 years period, based on the WorldClim database.	OUT	percent	[0 … 100]

5.5 Detailed processing scheme for the water LC-condition product

The water LC-condition product gives, for each pixel, the annual average temporal profile of frequency of the water status during the 1998-2012 epoch at global scale and with a 300-m spatial resolution. It gives, week after week, the frequency at which a pixel is covered with water over the 15 years.

The processing chain leading to the water LC-condition product is detailed in Figure 5-6.

The input data include three existing water products [AD-11]: the L3 CCI Water Bodies (L3 CCI WB) product [RD-149], the L3 aggregated MERIS 5°x5° tiles and the Geoland 2 Water Bodies product version2 (G2 WB) [RD-150].

Three preliminary steps are included in the product processing chain, which tend to harmonize the shape of input data to GeoTiff files at global extent and Plate-Carrée projection. They consist in format conversion, mosaicking and re-projection processes and are necessary for the application of further algorithms. Their representation, in dotted lines, shows their secondary importance. As the final water LC-condition product is presented on a 7-day basis, a compositing algorithm is applied in step 1 to data sharing a different aggregation periods. Step 2 consists in the combination of the water products in order to reach a unique 7-day combined product at global scale with sufficient observations. In a concern for improving the robustness of the final LC-condition product, remaining invalid (NaN) values are filled in step 3. Finally, step 4 temporarily aggregates water observations over 15 years and leads to the water LC-condition product itself.

As all preliminary steps regarding the L3 aggregated MERIS 5°x5° tiles were described in the snow LC-condition processing chain, the reader is invited to refer to section 5.4.1 for supplemental information.

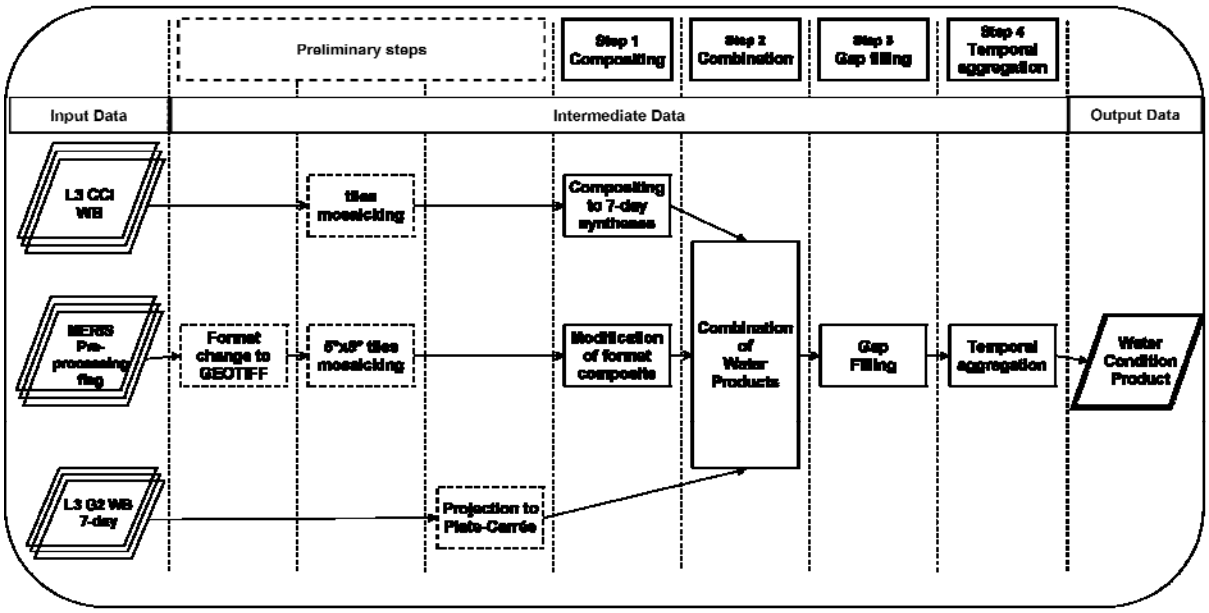


Figure 5-6: Schematic illustration of the processing chain of the water LC-condition product construction.

5.5.1 Preliminary steps

The mosaicking of tiled products concerns the L3 CCI WB while the re-projection into Plate-Carrée applies to the sinusoidal L3 G2 WB syntheses. This conversion of projection is mandatory to perform the combination of water products in step 3.

These two preliminary steps (mosaicking and re-projection) rely on the following "GDAL_Merge" and "GDAL_warp" functions. The first one allows appending several adjacent raster files together to form a unique raster file. It uses as parameters the name of the output file, the output format (which is GeoTiff by default), the compression type, the image structure information and the names of the input tiles. The second one can re-project any raster with a defined projection to any other supported projection. In our case, the conversion from sinusoidal to Plate-Carrée projection is operated. The nearest neighbour re-sampling method is used by default.

Input and output data

Input and output data of the two preliminary steps (mosaicking and re-projection) are described in Table 5-21. The parameters associated with the mosaicking and the re-projection steps are presented in Table 5-22.

Table 5-21: Input and output data for the mosaicking and re-projection preliminary steps (water LC-condition product)

DATA	DESCRIPTION	INTENT (IN, OUT, INOUT)	PHYSICAL UNIT	RANGE
L3_CCI_WB	Tiles of the CCI WB product	IN	dl	[0 … 2]
L3_Water_G2_250m_7d_Status	Water status of the 7-day	IN	dl	[0 … 2]

DATA	DESCRIPTION	INTENT (IN, OUT, INOUT)	PHYSICAL UNIT	RANGE
	period in "sinusoïdal" projection			
L3_Water_G2_250m_7d_NObs	Number of valid observations in the 7-day period in "sinusoïdal" projection	IN	dl	[0 … 7]
L3_Water_CCI_150m_mosaic_Status	Mosaic with the water status	OUT	dl	[0 … 2]
L3_Water_CCI_150m_mosaic_<var>Date with <var>: Start ; End	Mosaic of the dates when water was first/last detected	OUT	Day of the year	[1 ... 365]
L3_Water_G2_250m_7d_proj_Status	Water status of the 7-day period in the "Plate-Carrée" projection	OUT	dl	[0 … 2]
L3_Water_G2_250m_7d_ proj_Nobs	Number of valid observations in the 7-day period in the "Plate-Carrée" projection	OUT	dl	[0 … 7]

Table 5-22: Parameters needed to run the mosaicking and re-projection preliminary step (water LC-condition product)

PARAMETERS	DESCRIPTION	INTENT (IN, OUT, INOUT)	FORMAT	ALGORITHM
Input	Name of the input file	IN	String	Merge
Output	Name of the output file	IN	String	Merge
Compress	Compression type	IN	String	Merge
Interleave	Image structure information	IN	String	Merge
Tiled	Image structure information	IN	String	Merge
t_srs	target spatial reference set with "Proj" being Plate-Carrée and the "datum" being WGS84	IN	String	Wrap

5.5.2 Step 1 – Compositing

The first step of construction of the water LC-condition product is the construction of 7-day composite products. This step is described for the L3 CCI WB product (L3_Water_CCI_150m_mosaic layers). The aggregation is run in a way that keeps the water status of the 7-day period together with the number of valid observations on which it was defined.

For each 7-day periods, the status of each pixel of the L3 CCI WB product is included in the dedicated "Status" layer according to the "L3_Water_CCI_150m_mosaic_StartDate" and "L3_Water_CCI_150m_mosaic_EndDate". The water status is attributed to the 7-day period as soon as water was detected once in the period. The number of valid observation is counted and added in the Nobs layer.

Input and output data

Input and output data associated with this compositing step are described in Table 5-23. The parameters associated with this step are presented in Table 5-24.

Table 5-23: Input and output data of the step 1 of the Water LC-condition product generation

DATA	DESCRIPTION	INTENT (IN, OUT, INOUT)	PHYSICAL UNIT	RANGE
L3_Water_CCI_150m_mosaic_Status	Mosaic of the water status	IN	dl	[0 … 2]
L3_Water_CCI_150m_mosaic_<var>Date with <var>: Start ; End	Mosaic of the dates when water was first/last detected	IN	dl	[1 ... 365]
L3_Water_CCI_150m_7d_comp_Status	Water status of the 7-day period	OUT	dl	[0…1]
L3_Water_CCI_150m_7d_comp_Nobs	Number of valid observations in the 7-day period	OUT	dl	[0…7]

Table 5-24: Parameters needed in the step 1 of the Water LC-condition product generation

PARAMETERS	DESCRIPTION	INTENT (IN, OUT, INOUT)	PHYSICAL UNIT	RANGE
Input	Name of the input file	IN	String	/
Output	Name of the output file	IN	String	/
Ni_START	Date which marks the start of the 7-day period	IN	None	[YYYY0101 ... YYYY1222]
Ni_END	Date which marks the end of the 7-day period	IN	None	[YYYY0107 ... YYYY1231]
variables	created in the algorithm are defined in the pseudo-code representation [AD 12]	INOUT	dl	

5.5.3 Step 2 – Products combination

The second step of the water LC-condition product processing chain is the combination of the three 7-day water products into a unique combined product. This combination is applied for each 7-day periods for the 15 years.

The L3 CCI WB product is mainly derived from a water detection indicator generated using Synthetic Aperture Radar (SAR) data. It delineates, with a spatial resolution of 150 m, permanent and seasonal water observations along with the dates of detections. Although the global spatial coverage is satisfying, the number of measurements over time is expected to be limited (e.g Quebec). Moreover, SAR data might be unreliable in mountainous areas. This will certainly prevent from a clear discrimination of water bodies in such regions. To circumvent this issue, the L3 aggregated 5°x5° composites are the G2 WB product are integrated in reinforcement of the water detection.

An empty raster is firstly created with a global spatial extent and a Plate-Carrée Projection. The size of cells reaches 150m, the best spatial resolution amongst water products (L3 CCI WB). Water observations are assigned to the cells of the empty matrix according to the product from which it is originated. Priority is given to L3 CCI WB composites when they exist, then to the status of the L3 MERIS composites and finally to the G2 WB product.

Input and output data

Input and output data associated with this products combination step are described in Table 5-25. The parameters associated with this step are presented in Table 5-26.

Table 5-25: Input and output data of the step 2 of the Water LC-condition product generation

DATA	DESCRIPTION	INTENT (IN, OUT, INOUT)	PHYSICAL UNIT	RANGE
L3_Water_CCI_150m _7d_comp_Status	Water status of the 7-day period	IN	dl	[0…1]
L3_Water_CCI_150m _7d_comp_Nobs	Number of valid observations in the 7-day period	IN	dl	[0…7]
L3_Water_Meris_ 300m_7d_mosaic_Status	Mosaic of the number of observations with water in the 7-day period	IN	dl	[0 ... 2]
L3_Water_Meris_ 300m_7d_mosaic_Nobs	Mosaic of the number of valid observations in the 7-day period	IN	dl	[0 ... 7]
L3_Water_G2_ 250m_7d_proj_Status	Water status of the 7-day period in the “Plate-Carrée” projection	IN	dl	[0 ... 2]
L3_Water_G2_ 250m_7d_ proj_Nobs	Number of valid observations in the 7-day period in the “Plate-Carrée” projection	IN	dl	[0 ... 7]
L3_ Water_Combi_ 150m _7d_Status	Water status of the 7-day period	OUT	dl	[0 ... 2]
L3_ Water_ Combi_150m _7d_Nobs	Number of valid observations in the 7-day period	OUT	dl	[0 ... 7]
L3_ Water _Combi_ 150m _7d_Origine	Name of the product from which the water observation was taken.	OUT	/	[1 … 3] 1 = CCI 2 =

DATA	DESCRIPTION	INTENT (IN, OUT, INOUT)	PHYSICAL UNIT	RANGE
				MERIS 3 = G2

Table 5-26: Parameters needed in the step 2 of the Water LC-condition product generation

PARAMETERS	DESCRIPTION	INTENT (IN, OUT, INOUT)	PHYSICAL UNIT	RANGE
Input	Name of the input file	IN	String	/
Output	Name of the output file	IN	String	/
Ni_START	Date which marks the start of the 7-day period	IN	None	[YYYY0101 ... YYYY1222]
Ni_END	Date which marks the end of the 7-day period	IN	None	[YYYY0107 ... YYYY1231]
variables	Variables created in the algorithm are defined in the pseudo-code representation [AD 12]	INOUT	dl	

5.5.4 Step 3 – Gap filling

In order to enhance the robustness of the final water LC-condition product, step 3 tends to optimize the number of available observations by filling missing and invalid values remaining in the 7-day syntheses of the combined product. This gap filling is

Firstly, for each 7-day period, the combined product is scanned iteratively. When an invalid value (the gap) is detected, the values of its 8 neighbour pixels are observed. If at least 4 of them are valid and share the water status, then the water status is assigned to the gap. Consequently, the status "FILLED" is assigned to the corresponding pixel in the "L3_Water _Combi_7day_Origine" layer. Filled gaps are not taken into account as valid neighbour pixels for further gap filling.

Algorithm assumptions and limitations

The gap filling algorithm assumes auto-correlation of water presence.

Input and output data

Input and output data associated with this gap filling step are described in Table 5-27. The parameters associated with this step are presented in Table 5-28.

Table 5-27: Input and output data of the step 3 of the Water LC-condition product generation

DATA	DESCRIPTION	INTENT (IN, OUT, INOUT)	PHYSICAL UNIT	RANGE
L3_Water_Combi_ 150m _7d_Status	Water status of the 7-day period	IN	dl	[0 ... 2]
L3_ Water_Combi_ 150m _7d_Nobs	Number of valid observations in the 7-day period	IN	dl	[0 ... 7]
L3_Water_Combi_ 150m _7d_Origine	Name of the product from which the water observation was taken.	IN	/	[1 … 3] 1 = CCI 2 = MERIS 3 = G2
L3_ Water_ Filled _ 150m _7d_Status	Water status of the 7-day period after the gap filling	OUT	dl	[0 ... 7]
L3_ Water _ Filled _ 150m _7d_Nobs	Number of valid observations in the 7-day period	OUT	dl	[0 ... 7]
L3_ Water _ Filled_ 150m _7d_Origine	Name of the product from which the water observation was taken.	OUT	/	[1 … 4] 1 = CCI 2 = MERIS 3 = G2 4 = FILLED

Table 5-28: Parameters needed in the step 3 of the Water LC-condition product generation

PARAMETERS	DESCRIPTION	INTENT (IN, OUT, INOUT)	PHYSICAL UNIT	RANGE
Input	Name of the input file	IN	String	/
Output	Name of the output file	IN	String	/
Ni_START	Date which marks the start of the 7-day period	IN	None	[YYYY0101 ... YYYY1222]
Ni_END	Date which marks the end of the 7-day period	IN	None	[YYYY0107 ... YYYY1231]
variables	Variables created in the algorithm are defined in the pseudo-code representation [AD 12]	INOUT	dl	

5.5.5 Step 4 – Temporal aggregation

The last step of the construction of the water LC-condition product is the multi-annual aggregation of fire observations of the 15 years of 7-day syntheses into a unique annual and averaged water detection indicator. It is composed of 3 layers: a 15-year average number of water detections, the corresponding standard deviation and a quality flag layer.

For each valid pixel of each 7-day period, the average number of water detected over the 15 years and the corresponding standard deviation are calculated. A quality flag layer is then created summarizing the total number of valid yearly observation used in the water LC-condition calculation.

Input and output data

Input and output data associated with this temporal aggregation step are described in Table 5-29. The parameters associated with this step are presented in Table 5-30.

Table 5-29: Input and output data of the step 4 of the Water LC-condition product generation

DATA	DESCRIPTION	INTENT (IN, OUT, INOUT)	PHYSICAL UNIT	RANGE
L3_ Water _ Filled _150m _7d_Status	Water status of the 7-day period after the gap filling	IN	dl	[0 ... 7]
L3_ Water _ Filled _150m _7d_Nobs	Number of valid observations in the 7-day period	IN	dl	[0 ... 7]
L3_ Water _ Filled _150m _7d_Origine	Name of the product from which the water observation was taken.	IN	/	[1 … 4] 1 = CCI 2 = MERIS 3 = G2 4 = FILLED
L4_ Water _150m _7d_WAverage15y	Proportion of water status over the 15 years period	OUT	dl	[0 … 1]
L4_ Water _150m _7d_Std15y	Standard deviation related to the of 7-day period with the water status over the 15 years period	OUT	dl	[0 … x]
L4_ Water _150m _7d_NYearObs15y	Number of valid years included in the source products	OUT	dl	[0 … 15]
L4_ Water _150m _7d_15y_<source> with <source> = CCI; MERIS; G2; FILLED	7-day synthesis of the number of years where the water observation originates from the <source>	OUT	dl	[[1 … 4] 1 = CCI 2 = MERIS 3 = G2 4 = FILLED

Table 5-30: Parameters needed in the step 4 of the Water LC-condition product generation

PARAMETERS	DESCRIPTION	INTENT (IN, OUT, INOUT)	PHYSICAL UNIT	RANGE
Input	Name of the input file	IN	String	/
Output	Name of the output file	IN	String	/
Ni_START	Date which marks the start of the 7-day period	IN	None	[YYYY0101 ... YYYY1222]
Ni_END	Date which marks the end of the 7-day period	IN	None	[YYYY0107 ... YYYY1231]
variables	Variables created in the algorithm are defined in the pseudo-code representation [AD 12]	INOUT	dl	

5.6 Compliance between LC-condition products

As the condition products and the LC maps are built from various independent data sources, some discrepancies related to conformity or thematic aspects were visible. Therefore, some post-edition processing steps were applied on all condition products

Firstly, the spatial coverage was forced to be true global for all condition products and LC maps so that the extent is identical for all LC products.

Secondly, the land mask that spatially delineates the area concerned by the LC and therefore the coastline delineation was standardized for all conditions by superimposing the CCI-LC WB product. The percentages of BA and snow occurrences are forced to 0 under the water mask while the status of the NDVI condition is set to water (value 2). The value of the NDVI itself was set to NoData. The CCI-LC WB product is also applied in the LC maps.

Finally, the glaciers from the Randolph Glacier Inventory [RD-162] and the Antarctic coastline from the Scientific Committee on Antarctic Research Antarctic Digital Database [RD-163] were used to remove thematic discrepancies. Under these layers, the percentage of BA occurrence was set to zero, the percentage of Snow occurrence was set to 100%, and the status of the NDVI was set to filled ice (value 5) and the value of the NDVI itself to NoData. These layers are also applied in the LC maps.