fire_cci

D2.1 - Product Validation Plan

(PVP)

Project Name	ESA CCI ECV Fire Disturbance (fire_cci)
Contract N°	Nº 4000101779/10/I-NB
Project Manager	Arnd Berns-Silva
Last Change Date	05/07/2011
Version	3.1
State	Final
Author	Marc Padilla, Emilio Chuvieco, Stijn Hantson, Richard Theis, Christopher Sandow
Document Ref:	Fire_cci_Ph1_UAH_D2_1_PVP_v3_1
Document Type:	Public

Project Partners

Prime Contractor/ Scientific Lead	-	UAH - University of Alcalá de Henares (Spain)
Project Management	-	GAF AG, (Germany)
System Engineering Partners	-	GMV - Aerospace & Defence (Spain)
	-	DLR - German Aerospace Centre (Germany)
Earth Observation Partners	-	ISA - Instituto Superior de Agronomia (Portugal)
	-	UL - University of Leicester (United Kingdom)
	-	DLR – German Aerospace Centre (Germany)
Climate Modelling Partners	-	IRD-CNRS - L'Institut de Recherche pour le Développement - Centre National de la Recherche Scientifique (France)
	-	JÜLICH - Forschungszentrum Jülich GmbH (Germany)
	-	LSCE - Laboratoire des Sciences du Climat et l'Environnement (France)

Summary

This document is the Product Validation Plan (PVP) for the fire_cci project. It refers to the Task 2, Work Package 3100. The document establishes the validation protocols to test the performance of the algorithms including the round-robin protocol for inter-comparison of improved and data merging algorithms.

	Affiliation/Function	Name	Date
Prepared	UAH, GAF	Marc Padilla, Emilio Chuvieco, Stijn Hantson, Christopher Sandow	27/02/2011
Reviewed	UL, GAF	Kevin Tansey, Arnd Berns-Silva	08/06/2011 14/06/2011 05/07/2011
Authorized	UAH/ Prime Contractor	Emilio Chuvieco	04/07/2011
Accepted	ESA/ Project Manager	Stephen Plummer	15/06/2011

Signatures

	Name	Date	Signature
Signature of authorisation and overall approval	Emilio Chuvieco		
Signature of acceptance by ESA	Stephen Plummer		

Distribution

Affiliation	Name	Address	Copies
ESRIN	Stephen Plummer (ESA – ESRIN)	Stephen.Plummer@esa.int	electronic copy
Project Team	Emilio Chuvieco, (UAH) Patricia Oliva (UAH) María Teresa Calado (UAH) Stijn Hantson (UAH) Marc Padilla (UAH) Carlos Ricardo Cardozo (UAH) Dante Corti (UAH) Arnd Berns-Silva (GAF) Stefan Saradeth (GAF) Christopher Sandow (GAF) Federico González Alonso (INIA) Jose Miguel Pereira (ISA) Bernardo Moat (ISA) Kevin Tansey (UL) Andrew Bradley (UL) Máximo Fernández Cortizo(GMV) Oscar Pérez (GMV) Andreas Müller (DLR) Martin Bachmann (DLR) Kurt Guenther (DLR) Thomas Krauß (DLR) Eric Borg (DLR) Martin Schultz (JÜLICH) Angelika Heil (JÜLICH) Florent Mouillot (IRD) Philippe Ciais (LSCE) Patricia Cadule (LSCE) Chao YUE (LSCE)	emilio.chuvieco@uah.es Patri.ccaa@gmail.com teresa.calado@uah.es Hantson.stijn@gmail.com padilla.marc@gmail.com rickycardozo@gmail.com dantecorti@gmail.com Arnd.Berns-Silva@gaf.de Stefan.Saradeth@gaf.de Christopher.Sandow@gaf.de alonso@inia.es jmocpereira@gmail.com bmota@isa.utl.pt kjt7@le.ac.uk avb4@le.ac.uk maxfernandez@gmv.com operez@gmv.com andreas.mueller@dlr.de martin.bachmann@dlr.de kurt.guenther@dlr.de thomas.krauss@dlr.de eric.borg@dlr.de m.schultz@fz-juelich.de a.heil@fz-juelich.de florent.mouillot@ird.fr philippe.ciais@cea.fr patricia.cadule@lsce.ipsl.fr chaoyuejoy@yahoo.com.cn	electronic copy

Document Status Sheet

Issue	Date	Details
0.1	14/12/2010	First Draft
0.2	24/02/2011	First Document Issue
1.0	01/03/2011	First Document uploaded
2.0	20/04/2011	Revision
3.0	14/06/2011	Revision
3.1	14/06/2011	Minor adaption

Document Change Record

#	Date	Request	Location	Details
2.0	20/04/2011	UAH/GAF	Whole document	Revision following receipt of review comments from Stephen Plummer (ESA) and fire_cci team members
			Section 7	Introducing Reference Section
			p.8,Table 2,	Updated values
			p.9,Table 3,	Updated values
			p.10,Figure 2	Output processing example added
			Section 5.3.2	L3JRC-GLOBCARBON merged into one section
			Section 6.2.2.2	Various formulas added
			Whole document	Typo/formatting corrections
3.0	14/06/2011	UAH/GAF	Section 6.2.1.1	Vector to raster format conversion
			Section 6.2.1.2	Projection conversion introduced
			Section 6.5	Chapter 6.5 introduced addressing KSB review comments
			Annex 3	KSB formal review document included
			Whole document	Minor corrections and typo/formatting corrections
3.1	04/07/2011	UAH/GAF	Section 5.2.1.1	Minimum mapping unit of reference perimeters added,
			References	Updated

Table of Contents

List of Tables

List of Figures

List of Abbreviations

AATSR	Advanced Along-Track Scanning Radiometer
AD	Applicable Document
AFD	Active Fire Density
BA	Burned Area
CCI	Climate Change Initiative
CEOS	Committee on Earth Observation Satellites
CMUG	Climate Modelling User Group
DARD	Data Access Requirements Document
EFFIS	European Forest Fire Information System
ENVISAT	ENVIronmentalSATellite
ESA	European Space Agency
ECV	Essential Climate Variables
FAO	Food and Agriculture Organization
FR	Full Resolution
FTP	File Transfer Protocol
GAF	GAF AG
GCOS	Global Climate Observing System
GHG	GreenHouse Gases
GOFC-GOLD	Global Observation for Forest and Land Cover Dynamics
GTOS	Global Terrestrial Observing System
HDF	Hierarchical Data Format
IPCC	Intergovernmental Panel on Climate Change
MERIS	Medium Resolution Imaging Spectrometer
MODIS	Moderate Resolution Imaging Spectroradiometer
MIR	Middle InfraRed
NASA	National Aeronautics and Space Administration
NetCDF	NETwork Common Data Format
NIR	Near-InfraRed
PSD	Product Specification Document
PVR	Product Validation Report
R-R	Round Robin
RR	Reduced Resolution
SWIR	Short-Wave InfraRed
UN-REDD	United Nations Collaborative Programme on Reducing Emissions from Deforestation and Forest Degradation in Developing Countries
UNESCO	United Nations Educational, Scientific and Cultural Organization
URD	User Requirements Document
UTM	Universal Transverse Mercator
VEGETATION	CNES Earth’s observation sensor onboard SPOT-4
VIS	VISible
WFA	World Fire Atlas

1 Executive Summary

The *Product Validation Plan* is the top-level definition of the approach, protocols, tools, techniques and methods used for verifying that the results coming from the developed algorithms have geophysical meaning and that the developed climate relevant data products meet the user requirements. The plan shall ensure that the tests performed are representative of the full geophysical data space.

This document describes the main steps to follow through the different validation phases to select the best algorithm for producing BA maps and afterwards to assess its performance. For the different validation sites, BA and reference data will be generated and processed to produce comparable datasets with the outputs of the BA algorithm, with common temporal interval and spatial coverage. The most common validation methods were reviewed and the different pros and cons evaluated. Finally, the following were selected to be included in the fire_cci validation exercises: Cross-tabulation, linear regression and number of fires detected. These validation methods will firstly be applied during the R-R exercise and posterior for the validation of the different global BA products. To assess the spatial and temporal variability in quality of the BA product, validation metrics will be produced for different ecosystems and fire regimes, as well as the variably over time. Furthermore, a first trial for the integration of the different metrics in a single index has been analysed. Special attention is addressed to the R-R exercise, including a description of the requirements of the input data, the criteria used to integrate all the assessments of each algorithm and the access conditions of the external participation.

2 Introduction

2.1 Background

The final goal of any validation activity in Earth Observation image interpretation is to provide a quantitative assessment on the quality of the output product. The Committee on Earth Observing Satellites Working Group on Calibration and Validation (CEOS-WGCV) defines validation as: "The process of assessing, by independent means, the quality of the data products derived from the system outputs" (http://lpvs.gsfc.nasa.gov/). This implies to compare those results with a reference source, which is assumed to be the ground truth. Many authors distinguish between thematic and positional accuracy (Congalton & Green, 1999). The former refers to the agreement between the class assigned by the classification process and the reference document (considered as the ground truth), while the latter refers to the geographical matching between the polygons detected by both sources (map and reference). Disagreements between classification and reference data may be caused by these two factors, so they should be distinguished in the validation process.

The assessment of global Earth Observation products is particularly challenging (Chuvieco 2008), because of the great extension and diversity of the area covered, since accuracies may greatly vary between ecosystems and climatic conditions (Csiszar et al. 2009). However, the great demand of global variables as a result of the growing interest on global change studies has fostered in the last decades the creation of groups which try to establish standard strategies that could help local validation efforts to be applicable at global level. This is the case of the CEOS Land Product Validation Subgroup (LPVS), which is currently considered the reference source for data sets derived from remotely sensed data (http://lpvs.gsfc.nasa.gov/). In addition to establishing standards, global validation efforts are greatly help by the creation of regional networks of scientists. These networks promote the exchange of local knowledge and information that may be useful for compiling global databases or validating new products. An example is the regional networks included in the Global Observation of Forest Cover-Land Dynamics (GOFC-GOLD) programme (http://www.fao.org/gtos/gofc-gold/networks.html).

A critical step in the acceptance of the CCI products by external communities will be to provide some reliable measurements on the product quality, as well as its limitations and uncertainties. Accuracy in general terms indicates that the measurement has a small systematic uncertainty [AD-2]. Accuracy measurement implies comparing the derived results with a reference source that is assumed to be confident and analysing the differences. Most commonly, accuracy is measured in terms of global agreement between results and reference data, or in terms of omission and commission errors. For BA applications, the former imply that actual burned pixels are not included in the final product, while the latter are false positives, which are those pixels incorrectly detected as burned.

High precision indicates a measurement which has a small random uncertainty [AD-2]. It is commonly computed from some measure of the dispersion around the estimation mean (standard deviation for parametric variables, or interquartile range for non-parametric ones).

Finally, consistency can be defined as the temporal stability of accuracy over time. In other words, the term refers to whether the measured accuracy changes throughout time (year to year, for instance). For the BA product, this would imply to be sure that the BA product properly detects fire affected areas for fire seasons of high, medium and low occurrence.

On many occasions, ground measurements are not feasible or they are very costly. Therefore, many remote sensing validation projects rely on higher resolution images, which should be acquired simultaneously or close enough in time as to portray the same ground conditions as the input images from which the validating product is generated. This is the case of global burned maps, which have been commonly validated from a set of sampling sites where higher resolution images are available. Those sampling sites should be selected as to properly represent the variety of conditions that affect the accuracy of BA cartography, both in time and space. The most common options for validating global products (either active fires or burned area) have been ASTER, Landsat-TM/ETM+ and SPOT images (Boschetti et al. 2006; Fraser et al. 2004; Morisette et al. 2005; Roy et al. 2005).

The validation approach of the fire_cci project will follow the guidelines of the CEOS LPV sub-group (LPVS) and specifically the procedures proposed to assess radiance and global burned area products. Following LPVS terminology, four validation levels can be identified (http://lpvs.gsfc.nasa.gov/):

1. Product accuracy is assessed from a small (typically < 30) set of locations and time periods by comparison with in-situ or other suitable reference data.

2. Product accuracy is estimated over a significant set of locations and time periods by comparison with reference in situ or other suitable reference data. Spatial and temporal consistency of the product and with similar products has been evaluated over globally representative locations and time periods. Results are published in the peer-reviewed literature.

3. Uncertainties in the product and its associated structure are well quantified from comparison with reference in situ or other suitable reference data. Uncertainties are characterized in a statistically robust way over multiple locations and time periods representing global conditions. Spatial and temporal consistency of the product and with similar products has been evaluated over globally representative locations and periods. Results are published in the peer-reviewed literature.

4. Validation results for stage 3 are systematically updated when new product versions are released and as the time-series expands.

For the fire_cci project, the main validation effort will be concentrated on the BA product. Radiometric and geometric accuracy of the input images will be assessed based on the same validation sites as the burned area product, mostly for error tracing and temporal consistency.

2.2 Purpose of the document

The objective of the Product Validation Plan document is to establish the validation protocol to test the performance of the algorithms to be developed, including a formal round-robin protocol for inter-comparison of improved and data merging algorithms.

2.3 Applicable Documents

[AD-1]	ESA CCI Fire Disturbance - Proposal prepared by UAH University of Alcalá (Spain) in association with DLR (Germany), GAF (Germany), GMV (Spain), INIA (Spain), ISA (Portugal), UL (United Kingdom), IRD (CNRS) (France), JÜLICH (Germany), LSCE (France), 15 July, 2010, ESA Climate Change Initiative (CCI) Phase 1, Invitation to Tender AO/1-6207/09/I-LG (https://www.esa-fire-cci.org/webfm_send/110)
[AD-2]	ESA Climate Change Initiative. CCI Project Guidelines. Ref. EOP-DTEX-EOPS-SW-10-0002, issue 1, date of issue 05/11/2010. https://www.esa-fire-cci.org/webfm_send/117

2.4 Document Structure

The fire_cci PVP is structured as follows:

Section 1: Executive Summary

Section 2: Scope and introduction

Section 3: Review of existing validation methods for Burned Area products

Section 4: Validation criteria for the fire_cci project

Section 5: Description of reference data

Section 6: Description of validation activities, including the Round-Robin exercise. This section includes answers to the external validation of the PVP

Section 7: Contains the References made in this document

The appendices contain the metadata on validation data, the conditions to access the data and the external evaluation of the PVP.

3 Review of existing validation methods for Burned Area products

During the last decade, several validation efforts of BA products have been carried out within international projects (Table 1). The first one available was the validation of the GlobCarbon product (GlobCarbon 2007), using 72 pairs of Landsat images distributed globally, mostly for the year 2000, and a method based on the linear regression to assess the accuracy. A short time later, the L3JRC product was validated (Tansey et al. 2008), using the same methodology and reference data. The validation of the AQL2004 project (Chuvieco et al. 2008), besides linear regression, uses cross-tabulation methods and the number of burned land patches detected to assess accuracy of BA estimations, with Landsat and CBERS images as reference data in 28 sites form Latin America, from December 2003 to December 2004. A first evaluation of MODIS burned area products (Roy et al. 2008) has been produced using as reference data the MODIS active fire product, globally from July 2001 to June 2002, with linear regressions for the assessment. Most recently, Roy and Boschetti (2009) intercompared three products, GlobCarbon, L3JRC and MODIS, in Southern Africa, against eleven pairs of Landsat scenes using confusion matrix and linear regression methods. The comparison of the performances of those BA products reported by their publications is difficult. As shown in Table 1, there is no any common statistic that is used for all the studies, nor a common way to present the results.

BA validation efforts generally used three different methods: (1) the cross-tabulation, (2) the regression analysis and (3) the number of burned land patches detected, or variations on these.

The first method, cross-tabulation, aims to produce a confusion matrix accounting for the spatial coincidences and disagreements between a reference map and the target map. In our case, the BA product will be compared with reference perimeters extracted from higher resolution images. Several accuracy indices can be derived from the confusion matrix, such as global accuracy and omission and commission errors. They have been used in the last decade in variety of projects (Roy et al. 2009; Chuvieco et al., 2008; Padilla and Chuvieco 2009). An adaptation of this method to analyze the omission and commission errors of the confusion matrix is the Pareto Boundary, described in Boschetti et al. (2004).

Table 1 - Review of the most important works in BA validation, Ec refers to commission and Eo to omission error[1], respectively

Title	Citation	BA product	Reference data	Methodology	Results
GlobCarbon DPQRv4.2 Demonstration Products and Qualification Report (Results of the GlobCarbon validation)	GlobCarbon 2007	GlobCarbon	72 pairs of Landsat images distributed globally, mostly from 2000	Linear regression (in a hexagonal grid)	For global values r ranges from 0.34 to 0.85 depending on the algorithms Standard deviation[2] of 6.797 (IFI algorithm), 8.642 (UTL algorithm), 3.909 (GLOBSCAR algorithm) and 10.157 (logical OR algorithm)

[1] For a detailed definition of omission and commission errors, check section 6.2.2.1.

[2] This standard deviation refers to the ratio of the mean quadratic error and the sum of the differences between each error and the mean error (SD = MCE / Σ(Xi-Xmean): Globcarbon (2007). It measures consistency of the algorithm results. The larger the value the more scattered the fitting.

Title	Citation	BA product	Reference data	Methodology	Results
A new, global, multi-annual (2000–2007) burnt area product at 1 km resolution (Results of the L3JRC validation)	Tansey et al. 2008	L3JRC	72 pairs of Landsat images distributed globally, mostly from 2000	Linear regression (in a hexagonal grid)	Standard deviation (from the best-fit line) from 1.905 to 6.894, depending on the Continental Region
Global burned-land estimation in Latin America using MODIS composite data (Results of the AQL2004 Project)	Chuvieco et al. 2008	BA estimates from MODIS	Landsat and CBERS images in Latin America, from December 2003 to December 2004	Confusion matrix and linear regression (on a 5 x 5 km grid) and number of burned land patches detected	r^2 from 0.34 to 0.75 through the different study periods Ec[3] = 0.9 to 0.3 Eo[4] = 0.8 to 0.1
The collection 5 MODIS burned area product — Global evaluation by comparison with the MODIS active fire product (Results of an evaluation of the MODIS product)	Roy et al. 2008	MODIS	MODIS active fire product for the entire Globe, from July 2001 to June 2002	Linear regression (in a 40 x 40 km grid)	r from 0.38 to 0.80 depending on the tree cover and LAI
Southern Africa Validation of the MODIS, L3JRC, and GlobCarbon products	Roy and Boschetti 2009	MODIS L3JRC GlobCarbon	11 pairs of Landsat scenes in South Africa, in 2001	Confusion matrix and linear regression (on a 5 x 5 km grid)	L3JRC r^2 = 0.128 Ec = 1 to 0.026 Eo = 1 to 0.663 MODIS r^2 = 0.746 Ec = 0.655 – 0.086 Eo = 0.943 – 0.047 GlobCarbon r^2 = 0.509 Ec = 0.899 – 0.289 Eo = 0.964 – 0.256

The regression analysis, comparing the proportion of burned area detected by the reference and target dataset within a coarser grid than the BA product to be validated, is also widely used to assess the accuracy. The most important works using this methodology are from Silva et al. (2005), Boschetti et al. (2006), Roy and Boschetti (2009), Oliva (2009), Giglio et al. (2010) and the validations of past projects, such as AQL2004 (Chuvieco et al. 2008), GlobCarbon (GlobCarbon 2007), L3JRC (Tansey et al. 2008) and MODIS (Roy et al. 2008).

The number of burned land patches detected, considering their size and land cover affected has not been used widely, an example is the AQL2004 (Chuvieco et al. 2008) and more recently the study of Bastarrika et al. (2011).

[3] Ec = Commission error, defined as the ratio of false positives to the total number of BA detected pixels.

[4] Eo = Omission error, defined as the ratio of false negatives to the total number of true BA pixels.

4 Validation criteria to follow in the fire_cci project

Following the proposal presented to ESA [AD-1] and the guidelines of the first collocation meeting of the ESA CCI programme [AD-2], validation activities of the fire_cci project will take into account the following basic principles: objective assessment, independence, consistency, completeness. A further discussion is attached on methods to assure these principles.

4.1 Objective assessment

This criterion implies that the validation methods are objective and replicable by any researcher, and avoids subjective decisions that may bias the validation process. The simpler way of assuring objectivity is the use of metrics that are quantitatively and explicitly defined.

4.2 Independence

The independence is a critical characteristic of the validation process, since it assures that unbiased estimations are obtained. Independence implies that validation datasets are not used during the production of the BA product, not even during the generation of the geophysical variables of the pre-processing phase, and that the process is carried out by teams who are not involved in the BA production. For the fire_cci project, these conditions will be met, since BA reference data will be generated from different sources to those used for BA production (see section 5), and the teams involved in validation (GAF and UAH) will not work on algorithm development (while UAH is involved in both elements, the team will be split in two groups to ensure independence, one focused on developing the MERIS BA algorithm and the other one generating the validation datasets.

4.3 Consistency

Any validation effort should consider temporal variability of algorithm performance. In other words, the validation should provide a measure of whether results are dependent or not on temporal trends. For the fire_cci project, the reference datasets will include multi-temporal series for all 10 study sites. Therefore, the consistency of the BA product will be measured across a range of years, providing estimation on temporal variability of BA product accuracy. Validation metrics will include the calculation of standard deviation of accuracy measurements around the mean during the time series of the BA product.

4.4 Completeness

This concept refers to accounting for, in the validation, the main factors that affect accuracy. Consequently, the selection of reference sites should consider the potential variability of accuracy, both in space and time, and tackle that variability with an appropriate sampling. For the fire_cci project a statistical analysis of different factors affecting fire occurrence has been carried out to select the study sites and the reference images. Completeness was addressed both in terms of temporal and spatial diversity. The former was mentioned in the previous section, while the latter will be approached from a selection of 100 validation sites, to be generated within the fire_cci project from multi-temporal analysis of high resolution images (Landsat mostly). In addition to them, other BA reference information that is available from different national forest services and other international projects (Globcarbon, MODIS, Geoland, etc.) will be used to further extend the validation exercise. The following chapter explains this selection in more detail.

5 Description of reference data

5.1 Criteria to select validation sites

The spatial distribution of the reference data is very relevant to assure an unbiased and comprehensive validation. For global products, validation efforts require considering a wide set of fire conditions, land covers and biomes. The validation exercises consist of two steps, the first one is the selection of study areas to test the performance of the BA algorithms and the second one is the validation of the global BA products to be developed in the last phase of the project. Selection criteria for both validation steps are detailed hereafter.

5.1.1 Study sites

For the fire_cci project a selection of study sites was carried out to test the performance of the BA product in a wide range of fire and ecosystem conditions, considering time and data constraints, study sites were selected from the major biomes affected by fires:

- Tropical forest,
- Tropical savannas,
- Boreal forest, and
- Temperate-Mediterranean forest.

The selection considered historical records of fire occurrence, emissions from the most recent version of the Global Fire Emissions Database derived from the new MODIS Direct Broadcasat (DB) Burned Area algorithm (Giglio et al. 2010), and the suggestions of end-users and the GOFC-GOLD regional networks. Beside these two major criteria the following sub-criteria were crucial for the determination of the boundaries for every study site: their geographical position, type of climate, dominant vegetation cover, emission estimations, fire perimeters, local support and previous experience in the area. The sites were chosen taking into account the needs of the climate modelling community through the identification of those vegetation types whose contribution to the C cycle and emissions are considered to be relatively more important (e.g. boreal forests, tropical forests, peat fires). In addition, other factors were taken into account, such as availability of high resolution satellite data for every year of the time series, diversity with respect to fire behaviour, fire size distribution, affected vegetation, balance between low and high fire occurrence and presence of major biomes.

The ten selected study sites were approximately 500 km x 500 km in size and include the full time series of the fire_cci project: 1995 to 2009 to measure the consistency in the processing chain for the BA product outputs. The Data Access Requirements Document (Chuvieco et al. 2011a) includes a full description of all study sites.

Validation of the study sites will be done whenever possible in all years of the target time series using Landsat or other high resolution images. In this way, algorithm consistency for different years will be measured; including consideration of those regions with low and high fire occurrence (with higher potential for commission and omission errors, respectively). For each site, reference data (high resolution data and fire perimeters) and input data have been collected from different sources. The primary source of reference information is multi-temporal analysis of Landsat-TM/ETM+ images. The availability of these images in the historical archives of the USGS (http://glovis.usgs.gov/) for all Landsat scenes included in the ten study sites was surveyed. For each study site, the Landsat scene with more historical images available was selected and the suitable images (low cloud cover) downloaded (see Chuvieco et al. 2011a for the list of images available).

5.1.2 Sites for the validation of the global BA product

In terms of analyzing spatial variability of accuracy, an additional set of validation sites has been selected and will be generated within the fire_cci project. Each of these validation sites covers a Landsat scene (around 34,000 km^2). These additional validation sites will be used to assess the global BA product, and they will be selected independently from the ten study sites referred in the previous section.

The validation of the global BA product is mostly focused on determining its quality for different ecosystems, fire regimes and climatic conditions. These validation sites should be chosen randomly, but still cover at least the different types of fire regime present at a global scale. We performed a stratified random sample, selecting 20 samples for each of the 5 equal-frequency intervals of Average Fire density (AFD: see Figure 1). The AFD was computed from Terra-MODIS hotspot data and it was defined as the average number of hotspots detected in a 0.5° x 0.5º grid over a month (multiplied by 10^5 to avoid small numbers). The selected validation sites are presented in Figure 1, overlaid on the AFD map.

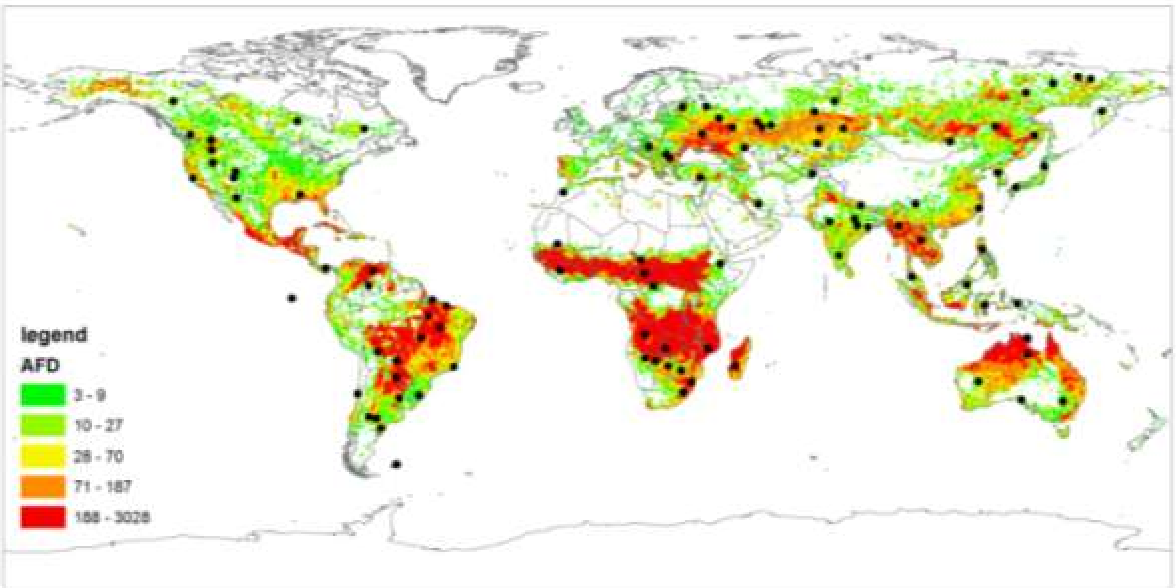

Figure 1 - Validation areas for the global BA products over a map of Average Fire Density, extracted from Terra-MODIS hotspot data

To evaluate the main type of vegetation corresponding to these sites, vegetation cover was extracted from the Globcover2005 dataset. The distribution of the validation sites over vegetation type is presented in Table 2. The main vegetation types are well represented in the sample, with a reasonable distribution between forest, pastures and agricultural areas.

Table 2 - Distribution of the validation areas over the different vegetation types (derived from Globcover2005)

Vegetation type	Nº areas
broadleaved evergreen	10
broadleaved deciduous	14
coniferous	20
shrub	12
rangeland	15
agriculture (mosaic)	29

5.2 Protocol for creating and documenting BA reference files

This section describes the proposed protocol to generate and document reference information for BA validation. This document is based in the CEOS-CalVal protocol for the validation of burned area products (Boschetti et al. 2009), and has been agreed by the partners involved in the fire_cci consortium, as well as by the group working on validating MODIS BA information (Boschetti, personal communication).

5.2.1.1 Generation of BA reference data

Reference perimeters will be primarily generated from multi-temporal comparison of high resolution satellite imagery (Landsat-TM or similar), acquired from before and after the fire(s), preferably from the same year. In case burned areas are unambiguously visible in satellite data for more than a single fire season, as it is the case in many boreal ecosystems, and the landscape is sufficiently homogeneous to map the burned area with confidence, the pre-fire image could be accepted when acquired 2-3 years before the fire.

The generation of fire perimeters from satellite images will be preferably performed by applying the ABAMS software, which implements an automatic BA detection algorithm widely tested in temperate fires (Bastarrika et al. 2011). The software is available for general use at http://www.ehu.es/aitor.bastarrika/.

After the automatic discrimination of perimeters, a systematic quality control will be performed through visual inspection. All fire perimeters should be reviewed by a different interpreter. This evaluation is done in two phases. In the first one, all perimeters are revised visually by another expert, and those polygons where interpretation is uncertain will be labelled. Afterwards, a meeting with the original interpreter is held to solve those uncertainties. Additionally, for a sample of scenes and dates a completely new interpretation is done by another interpreter, and cross-tabulation between the original and the new polygons is carried out.

When fire perimeters are derived from forest services (either by field observation, GPS or other methods) a Landsat TM image from after the fire will be used to verify and date the perimeter. In case, the post-fire image is not available, those fire perimeters will not be considered for validation purposes. To clarify the pre-fire date, an image acquired the same year of the fire will be used. In case that image is not available and the fire record does not include the specific date of the fire but only the year, the pre-fire date will be considered January 1st of that year, since that is the best estimation of fire time available.

The same rules for documenting reference burned area files apply even if fire perimeters are generated from other methods, in order to generate a uniform database of reference BA sites.

Since validation needs to discriminate between observed and not-observed areas, the BA reference files will include not only fire perimeters, but also the not-burned areas of the scene, as well as part of the scene that could not be observed, either by clouds or by sensor problems (i.e. SLC-off problems of ETM+), in both cases to assure that areas with no valid data are included in the validation process. Depending on the cloud characteristics, they will be masked out by a supervised classification or on-screen digitized (at least roughly, see an example in Figure 2).

The minimum mapping unit for the reference perimeters will be 1 ha, regardless what is the source of fire perimeters. This minimum size also applies to unburned patches inside burned areas and to non-observed areas by the high-resolution images (either by cloud or SLC-off problems).

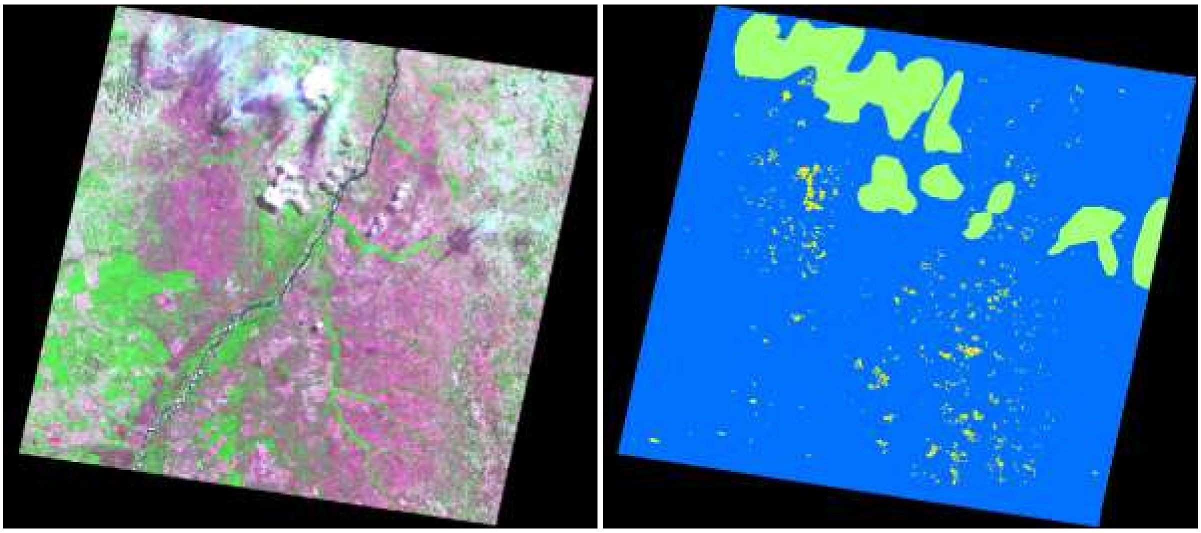

Figure 2 – Example of a Landsat post-fire image RGB (7, 4, 3) (left) and output of the processing steps described in the text (right) (yellow: BA, green: clouds and not observed, blue: Unburned)

5.2.1.2 Data structure and naming convention

Each burned area reference file will be an ArcGIS™ shape file (.shp), along with the auxiliary files required (.dbf, .prj, shx, .sbn, .xml). The projection will be UTM, WGS84, with the UTM zone being the zone that is covered by the major part of the scene. The following attribute fields will be included in the shape file (Figure 3):

- PreDate. Acquisition date of the image taken before the occurrence of the fire: yyyymmdd (year, month, day). When fire perimeters are obtained only from one post-fire image, pre-fire date will be established from images acquired the previous year. In case the fire perimeters are derived from forest services and the perimeters have the year but not the specific date/month, the pre-fire date will be the first of January of that year.
- PostDate. Acquisition date of the satellite image taken after the fire: yyyymmdd (year, month, day). The post-fire date will be the reference for validation purposes, regardless when the fire(s) actually occurred.
- PreImg. The pre-fire image name, following this format: satellite-code_Path_Row (i.e. LT5_201_032). The satellite codes are given in Table 3. When fire perimeters are obtained only from one post-fire image, the pre-fire image will be taken from images acquired the previous year, or when derived by forest services this attribute will be left in blank.

Table 3 - Satellite-sensor codes naming convention

Satellite-sensor	code
Landsat-4 MSS	LM4
Landsat-4 TM	LT4
Landsat-5 TM	LT5
Landsat-7 ETM+	LE7
Terra-Aster	AST
EO1-ALI	ALI
IRS1C-LISS-3	I1CL3
IRS1C-WIFS	I1CW
IRS-p6-AWIFS	I6A
CBERS1-CCD	CD1
CBERS2-CCD	CD2

- PostImg. The image used for BA detection, following the same naming convention as PreImg.
- Area (in square metres, m^2)
- Category (Observation category):
    - Burned area = 1. This area includes all polygons detected as burned.
    - Not-observed = 2. This area includes all polygons that were not observed by the sensor, either by clouds and/or cloud shadows, topographic shadows, smoke, or sensor errors (for instance, those caused by SLC-off problems of ETM+)
    - Unburned = 3. This area includes all polygons observed as not burned within the limits of the area covered by the image.

FID	Shape *	PreDate	PostDate	PreImg	PostImg	Area	Category
0	Polygon	20030630	20030801	LT5_223_066	LT5_223_066	1062000	1
1	Polygon	20030630	20030801	LT5_223_066	LT5_223_066	85500	1
2	Polygon	20030630	20030801	LT5_223_066	LT5_223_066	933300	1
3	Polygon	20030630	20030801	LT5_223_066	LT5_223_066	108000	1
4	Polygon	20030630	20030801	LT5_223_066	LT5_223_066	163800	1
5	Polygon	20030630	20030801	LT5_223_066	LT5_223_066	1454400	1
6	Polygon	20030630	20030801	LT5_223_066	LT5_223_066	38700	1
7	Polygon	20030630	20030801	LT5_223_066	LT5_223_066	12600	1
8	Polygon	20030630	20030801	LT5_223_066	LT5_223_066	55800	1
9	Polygon	20030630	20030801	LT5_223_066	LT5_223_066	244800	1
10	Polygon	20030630	20030801	LT5_223_066	LT5_223_066	332100	1

Figure 3 – Example of attribute table for BA reference data.

The name of the .shp and associated files will be defined as follows:

PRO_RD_YYYYMMDD_YYYYMMDD_PPPRRR

where:

PRO = Project where the fire perimeters were generated for. For the fire perimeters developed within the fire_cci project, PRO=fire_cci. When the perimeters come from another project, they will be referenced in abbreviation (for instance, GLOBC, for GLOBCARBON).

RD = stands for Reference Data

yyyymmdd (year, month, date). The first one is the pre-fire date, which is the date of the first image used for BA detection; the second one is the post-fire date, which is the date of the last image used for generating the reference fire perimeters. As previously said, in cases when fire perimeters are derived from forest services, and there is not a suitable pre-fire image available, the pre-fire date will be the 1st of January of the same year.

ppprrr represents the path and row of the scene (in the case where no Landsat imagery was used, the closest path-row will be selected): ppp=path; rrr=row

Whenever possible, along with the validation file, a compact version of the images used for the validation exercise will be included (Figure 4). They will be preferably in .jpg format, using a pseudo-true colour infrared composite (Landsat TM 7/4/3), and with a size of 400 x 400 pixels. These files will be named as the shape files, with the extension .jpg

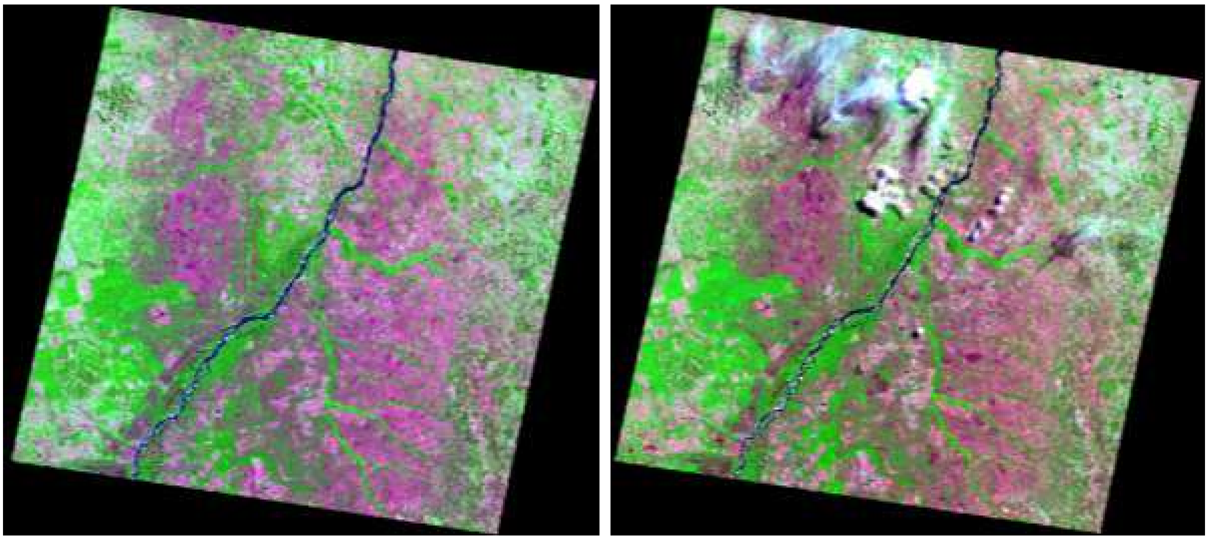

Figure 4 – Example of pre (2003/06/30) and post fire (2003/08/01) Landsat imagery (scene 223_066) in RGB (7, 4, 3)

5.2.1.3 Metadata convention

The metadata of the reference files will be delivered as a XML document following the international CSDGM and ISO 19115 standards, so that the metadata can be easily consulted by the majority of the GIS and metadata generation software packages. The metadata will contain the fields to cover all necessary information to be provided to external users. The metadata that will be included is detailed in Annex 1.

5.3 Sources of external reference data

5.3.1 Requirements for external reference files

Most reference data for validation within the fire_cci project will be generated by the consortium itself, but we will also try to gather as much reliable information as possible from other sources of data. These external sources should comply with the following requirements:

- Reference files should have a start and end date of BA detection so that it adequately can be compared at a temporal scale with BA products.
- Data should have a high reliability, with all present BA patched detected in the studied area and areas that could not be analysed satisfactory masked.
- Data should have a high quality, with low commission and omission errors, extracted from high resolution satellite imagery or field data.

In any case, those external reference files will be checked to comprise the same requirements as those explained in the Section 5.2, and whenever possible they will be documented in the same way as suggested by our protocol.

Potential sources of reference BA information are the following international global scale BA projects, the regional data derived from national forest services and regional analysis carried out by Earth Observation scientists.

5.3.2 L3JRC-GLOBCARBON

Validation areas used from previous BA products will be incorporated to our validation dataset. Those used for the L3JRC and GLOBCARBON BA (Globcarbon 2007) products have been provided by Dr. Tansey, from University of Leicester. They were sent in HDR raster format with the observation category encoded (1 = burned, 0 = not burned and 255 indicates areas outside of the Landsat WRS frame). These files were derived either from full resolution Landsat images or from quicklooks. They sum a total of 78 images, but those derived from quick looks will not be included. Presently, these files are being reprocessed at University of Leicester.

5.3.3 MODIS

MCD45 BA product has not yet been validated for the entire globe, with the most extensive validation exercise so far being the validation of the African continent by Roy and Boschetti (2009). Currently, additional reference data is being collected to perform a global validation. These data are generated by visual interpretation of Landsat imagery. Due to the fact that these data are not yet published, we are not sure this dataset will be available for validation purpose in our project. However, we are in contact with the responsible scientists to reach an agreement.

5.3.4 AQL2004

This project validated a MODIS BA algorithm developed for Latin America (Chuvieco et al. 2008). The validation data was generated by visual interpretation of Landsat, SPOT and CBERS imagery. High quality validation data are available for 28 sites distributed over Latin America for the year 2004 in shape format.

5.4 Access conditions

The reference files generated within the fire_cci project will be made available through the fire_cci website for public access and use, providing the source of data is properly acknowledged.

After the first phase of validation, including the some of the multi-temporal series of data and the Round Robin outputs, a first validation document will be made available. This document should be ready during the first three months of 2012.

After the second phase of validation the product validation and algorithm selection report will be released on the website but also send to every key user involved. The final product validation is documented in different reports like the Product Validation Report for the Time Series, the Product Validation Report for the Global BA Product and the Product Inter comparison Report. These reports will be also made available on the project website at the end of the project.

In addition, the data products for the validation sites will be provided by the fire_cci project to the CEOS Land Product Validation Fire sub-group for hosting on their web portal for free distribution to other interested scientists, providing proper acknowledgement of the source is made.

6 Description of validation activities

6.1 Validation of pre-processed data

Description of the error characterization and sensitivity analysis of the different phases of data pre-processing are described in the Algorithm Theoretical Basis Document – Volume I – Pre-processing (Bachmann et al. 2011).

6.2 Validation of the BA product

The validation of the BA product for the fire_cci project will be restricted to the pixel-based BA outputs. The grid-based product will not be validated, since they will be derived from the pixel products and will provide accumulated data of total burned area for each grid. No reference data are available for that level of detail. The only quality control of the grid-product will be consistency test between outputs of models running with grid fire_cci outputs and other properly documented sources of BA information.

6.2.1 Transformations of reference and BA product

The BA product derived from the fire_cci project will obviously have different spatial and temporal conditions than the reference BA information used for validation. Therefore, validation of the BA product involves different processing steps to properly relate the reference information with the BA output results. The following sections describe this process, which will also be included in the R-R software.

6.2.1.1 Processing steps for BA reference data

The following transformations are required for reference data:

1. Cleaning small polygons: All fire polygons smaller than the minimum size of BA detectable by the BA product will be excluded from further analysis. The definition of the minimum BA patch size is included in the PSD document (Chuvieco et al. 2011b). To avoid that clusters of small BA patches are not considered, a previous spatial analysis will be performed by merging together those BA patches that are within a 100m radius, and they will be considered one single BA patch.
2. Raster conversion: The shape file will be converted to raster format, keeping the original resolution of the images used for reference perimeters (30 m for Landsat-TM/ETM+).

6.2.1.2 Processing steps for the fire_cci BA product

The following transformations are required for each pixel-based BA product:

1. Projection conversion: The BA product will be re-projected to the standard projection used in the reference perimeter files (UTM). This process implies minor positional errors, below the pixel size, since the projection conversion is a mathematical transformation with well known parameters.
2. Clipping spatial subsets: The spatial coverage of the BA product must coincide with the extent of the reference data, i.e. the extent of the Landsat images used for generating the reference data. The limits of the Landsat scenes are defined by a vector comprising all categories included in the shape reference file (see 5.2.1.1).
3. Temporal selection: Pixels to be compared with reference information should be within the temporal range of the reference files. The pre and post-fire dates (YYYYMMDD) are included in two fields described in the reference file structure (again see 5.2.1.1).
4. Pixel resize: The pixels of the fire_cci BA product must be resized to the original resolution of the images from which reference fire perimeters were derived (by default 30 m), ensuring

proper overlay with pixels of the reference data. This is necessary for the pixel per pixel comparison of some validation methods, detailed below.

6.2.2 Methods to validate the BA product

Three validation methods cited in Section 3 that will be used in the validation of the BA product, are described below. These methods, as indicated above, are widely accepted and used by the scientific community, in peer reviewed publications and in the validation of BA products.

6.2.2.1 Cross-tabulation

A cross tabulation between the pixels assigned by in the BA Product and in the reference data will be computed to produce the confusion matrix (Table 4). The diagonal of this table will include burned or unburned areas mapped correctly; while the other cells will represent omission (actual burned areas not detected as such) or commission errors (detected burned areas that were not in fact burned). The information contained in the error matrix can be summarized by global metrics or by class specific metrics. Examples of global metrics are the Overall Accuracy (total percentage of correctly classified pixels) or the Kappa coefficient, which is widely used to measure the agreement between reference and classification data (Congalton & Green, 1999). The most commonly used class-specific metrics are the User's and Producer's Accuracy, which are the opposite of omission and commission Errors (Congalton & Green, 1999).

Table 4 - Structure of the confusion matrix

		Reference data		
		Burned	Not burned	Total
Burned area product	Burned	X_{11}	X_{12}	X_{1+}
	Not burned	X_{21}	X_{22}	X_{2+}
	Total	X_{+1}	X_{+2}	$\sum X_{ij}$

The statistics derived from the error matrix are calculated as follows:

1. The commission error (Ec): The ratio between the false BA positives and the total area classified as burned by the global product.

$$E_c = \frac{X_{12}}{X_{1+}} \quad (1)$$

2. The omission error (E_o): The ratio between the false BA negatives and the total area classified as burned by the reference data.

$$E_o = \frac{X_{21}}{X_{+1}} \quad (2)$$

3. The Kappa index (k) measures the agreement between the reference and BA data, by analyzing differences between observed and the expected classifications, assuming they were randomly distributed. A value of k equal to 1 indicates a full agreement between the reference and the target BA product, and a value close to 0 indicates that this agreement is purely random. The formula is.

$$k = \frac{N\sum_{i=1,2} X_{ii} - \sum_{i=1,2} X_{i+}X_{+i}}{N^2 - \sum_{i=1,2} X_{i+}X_{+i}} \qquad (3)$$

Where N is the total area analysed in the cross tabulation ($\sum_{i=1,2}\sum_{j=1,2} X_{ij}$) .

6.2.2.2 Regression analysis

Since the cross tabulation between high and coarse resolution data is affected by the different pixel sizes, it thus includes errors due solely to the spatial resolution (Boschetti et al. 2004). Different authors have suggested basing the accuracy assessment of BA product on comparison of the proportion of burned area detected by the reference and target dataset within a coarse grid (typically 5x5 km: Roy et al. 2005, Roy & Boschetti, 2009, although other authors suggest 15 x15 km: Plummer, 2004). The accuracy is then defined by the regression estimation (coefficients of determination and slope coefficients mainly) between the two proportions for the whole validation site.

Scattergraphs can be produced to show the proportion burned in each 5 x 5 km grid cell, by the global product (y-axis) and by the reference data (x-axis). As usually burned area proportions do not have a normal distribution, instead of using the least squares estimator, a nonparametric method based on Kendall’s rank correlation $\boldsymbol{\tau}$ may be utilized to compute linear regression estimates (Roy et al. 2008; Sen 1968; Theil 1950).

The regression line estimated, expressed as the slope and the intercept coefficient estimates, shows the over or under-estimation trend of the BA product. If the intercept value is close to 0, the slope of the estimated regression line shows over-estimation when higher than 1 or over-estimation when lower than 1. The correlation coefficient shows the lineal relationship strength, high if close to 1 and low if close to 0.

The statistics derived from the linear regression of Y (proportion of burned area detected by the global product) on X (proportion of burned area detected by the reference dataset), are calculated as follows:

$$y = a + bx \qquad (4)$$

where

$$b = \text{median of the } S_{ij}\text{s} = S_{(\text{median})} \qquad (5)$$

S_{ij}s are the “two-point slopes” for all pairs (X_i, Y_i) and (X_j, Y_j), such that $i < j$ and $X_i \neq X_j$

$$S_{ij} = \frac{Y_j - Y_i}{X_j - X_i} \qquad (6)$$

and

$$a = Y_{0.50} - bX_{0.50} \qquad (7)$$

$X_{0.50}$ and $Y_{0.50}$ refer to the sample medians.

On the other hand, the correlation coefficient between Y and X is computed as:

$$\tau = \frac{N_c - N_d}{N_c + N_d} \qquad (8)$$

Where N_c (number of concordant observations) and N_d (number of discordant observations) are computed as follows (for all pairs (X_i, Y_i) and (X_j, Y_j), such that $i < j$):

if $\frac{Y_j - Y_i}{X_j - X_i} > 0$, 1 is added to N_c

if $\frac{Y_j - Y_i}{X_j - X_i} < 0$, 1 is added to N_d

if $\frac{Y_j - Y_i}{X_j - X_i} = 0$, ½ is added to N_c and ½ to N_d

if $X_i = X_j$, no comparison is made.

Where X is the burned proportion assigned by the reference data and Y is the burned proportion assigned by the BA product of paired observations (5 x 5km cells).

6.2.2.3 Patches detected

The number of reference BA patches detected in the fire_cci BA product provides an estimation of the detection rate of the different BA algorithms. This approach tries to alleviate the problems associated with the different spatial resolution between the reference and the target products. Instead of focusing on a pixel by pixel analysis, this test measures whether a particular burned patch was detected or not by the BA algorithm. This index is calculated by computing if at least a pixel of the coarser resolution product is within a burned perimeter of the reference information. The exact shape of the burned patch may differ, because of spatial resolution or geometric displacements, but if the burned patch is detected we may assume that the algorithm correctly or at least partially detected the actual fire disturbance of the target site. This approach makes it possible as well to quantify the algorithm sensitivity to detect small fires.

6.2.3 Confidence intervals

All the above referred measurements of accuracy are in fact estimations of true accuracy based on samples. As in any other sampling exercise, those results are probability estimations with a certain confidence, depending on the significance level and sampling error, as follows:

$$\bar{x} = \hat{x} \pm z * SE \qquad (9)$$

where $\bar{x}$ is the mean of the population (in this case the actual accuracy or error in the image), $\hat{x}$ the estimated error from the sampling, z the level of probability given to the estimation (one minus de significance level), and SE the sampling error, calculated differently for each type of sampling. The term (z * SE) indicates the confidence interval, that is the value below and above the estimated mean in which one can expect to find the target parameter with a certain level of confidence. For the fire_cci project, all accuracy parameters will include confidence intervals with a 95% of probability, meaning that the estimations would have lower than 5% chance of being incorrectly estimated ($p<0.05$).

6.2.4 Controlling factors

Disagreements between the reference and the classification data will be analyzed globally, separately by years and considering different factors that affect algorithm accuracy. Therefore, the previous detailed tests will be assessed globally and considering the different factors, this will provide further information on what is influencing more the error of each algorithm performance. This analysis of errors is very relevant to improve algorithms and to provide to the BA product user an honest assessment of what constraints and limitations would the final product have.

The controlling factors considered in the validation of the fire_cci product will be:

1. Land cover types (derived from Globcover2005)
2. Vegetation density (extracted from the MODIS Vegetation Continuous Fields)
3. Climate-ecosystem zones (extracted from the Holdridge Life Zones and the Olson's 2001 biomes)
4. Burned patch size

6.2.5 Validation phases

The first use of validation data will be within the algorithm development phase. Each EO science team in charge of developing the algorithms will use the ten study sites to calibrate their algorithms. A limited number of reference files will be given to the algorithm developers to test the results and find potential problems. The processing of reference data for all of those sites has started during the first months of the project and first results have been uploaded to fire_cci website. This will help the algorithm developers in their efforts.

The second phase of the validation activity will be performed in the R-R exercise, when the study sites to be included will have all the reference information available for testing accuracy, precision and temporal consistency. Therefore, different reference data will be used for calibration and validation of the algorithms. The R-R software and database will include those reference data as well as the routines to compute assessment metrics for the results of the different algorithms to be tested. Consequently, the output of the exercise will provide objective criteria to select the most robust and reliable algorithm. A further description of the R-R exercise is included in section 6.3.

Once the BA algorithm is selected and the production phase begins, the rest of reference data not previously used for algorithm calibration and validation will be used. This will comprises a subset of reference files for the study sites (see section 5.1.1), as well as the additional 100 global validation sites described in section 5.1.2.

After processing the reference and the product data, the accuracy indices previously commented will be computed, and the Product Validation Report (PVR) will be written.

6.3 Round Robin (R-R) exercise

6.3.1 Framework for the R-R exercise

The project will develop a set of algorithms for burned area mapping based on previous experiences and literature review. Preliminary validation of those algorithms in the sample study sites will serve to adapt the algorithm to different environmental conditions. However, to decide the most convenient approach for global processing of the BA product, a Round Robin (R-R) exercise will be carried out at the middle of the project (15-18 month). The R-R will serve as a benchmark test of the BA algorithms proposed, as well as other algorithms that have been previously used or can be suggested during the exercise. The final goal is to select the best performing algorithm, mostly in terms of accuracy, although other aspects (i.e. processing time), may also be taken into account. As part of this exercise, reference information is critical, since the performance of different algorithms against a wide range of conditions (e.g. fire characteristics, ecosystems, land covers) needs to be evaluated. The Round Robin exercise will be carried out at four study sites with contrasting ecosystem and fire regime conditions. Reference data for those sites will also be part of the R-R data package.

6.3.2 Input data layers

Conceptually, the R-R package will include three modules: input, processing and assessment. The input module will include all necessary data layers to run the exercise, namely:

- Calibrated reflectance from ATSR, VGT and MERIS
- Reference BA perimeters
- Auxiliary information: globcover2005 data, hot spots, eco-regions, etc.

These input data layers will be taken from some of the sampling test sites used in the project, considering the full historical series of different sensors. Scientists willing to run the exercise with their own study sites will have software tools created by the fire_cci project to input their own set of data layers.

6.3.3 Algorithms tested

The processing module of the R-R package will implement the algorithms developed within the project and will provide computer tools to facilitate that external algorithms are easily connected, so any scientists interested to check his/her algorithm can easily connect with the R-R database.

The R-R package will provide precise instructions for external participants on:

- the datasets available for testing,
- the approach needed for reading each dataset,
- the format and information content for inclusion of any results generated by external participants,
- the grid and projection information needed,
- the result description,
- the time available for submission of results for inclusion in the project.

6.3.4 Validation criteria

The validation module of the R-R package will include a set of metrics that will be able to quantitatively validate the accuracy of the different algorithms. The set of metrics described in Section 6.2.2, namely the error matrices, scattergraphs (and their correspondent determination coefficients and regression coefficients), and Number of burned land patches detected, will be computed. They should be able to quantitatively validate the accuracy of the different algorithms.

In all cases, disagreements between the reference and the classification data will be analysed considering different factors that affect algorithm accuracy, as described in section 6.2.4. This assessment will provide further information on which factors of error are more relevant in each algorithm performance.

This will produce a high number of metrics, which would provide relevant information to select the most adequate algorithm. To summarize those metrics, we will derive a synthetic validation index, which will facilitate the comparison of performances of different algorithms. Further discussion is needed to develop the methodology for this integration. A first idea is a Multi-Criteria Analysis, were the weights of the different metrics could be given by the user community and extracted from the User Requirements Document (Schultz et al.2011).

At the end of the exercise, a Product Validation and Algorithm Selection Report (PVASR) will be written, which will serve to select the final algorithms to be used in Burned Area production for all the areas previously indicated (globally for five years and historical series for selected test sites).

6.3.5 Schedule for the Round Robin activities

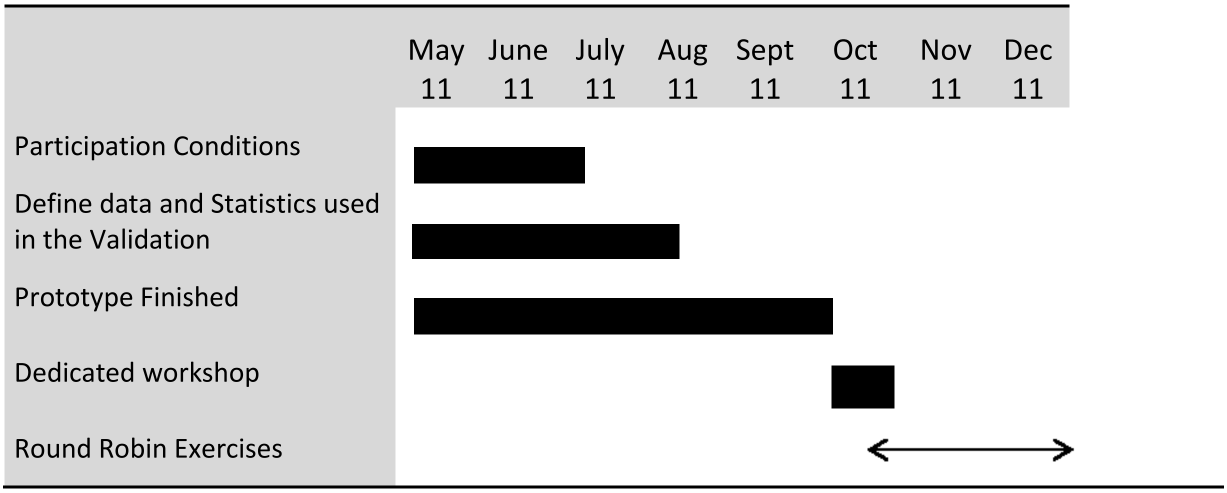

Figure 5 – Schedule for the Round Robin activities

6.4 Engagement of external scientists

The fire_cci project acknowledges the open policy of the ESA CCI programme with respect to access to information (data and methods) from external scientists. This is especially critical in the validation exercises, since one of the key criteria of validation is independence. Therefore, the BA product will be publicly available for external validation by interested scientists, once the proper quality control and internal validation has been accomplished.

The participation of external experts in the validation phases will be organized around two main phases, the R-R exercise and the final BA product validation.

6.4.1 During the R-R exercise

As previously said, the R-R Exercise will be open to all interested scientists. In principle, the main actors of this exercise will be the GOFC-GOLD regional networks experts that have shown interest to be involved in this project, and formally supported our proposal (see list of external scientists [AD-1[), but we will try to engage as well other networks of fire scientists from different communities (carbon modellers, atmospheric chemists, ecologists, natural hazard managers, etc.). During the R-R exercise, three regional workshops will be carried out to foster the participation of regional experts. They will be organized by the validation team of the project (UAH, GAF), along with the regional scientists that support the project. Each workshop will be attended by local scientists who will be able to identify potential problems in algorithm performance for different ecosystems.

The exercise will also be properly announced in specialised forums and professional societies, looking for the involvement of as many relevant scientists as possible in the exercise. A dedicated workshop will be held in Stresa (Italy) on 17th and 18th October, 2011, where the most relevant scientists (considering their publication records) working on BA algorithm development will be invited.

In addition to scientists attending these workshops, a web-based tool will allow easy access to the R-R package to other scientists working in BA mapping. We will try to involve as many relevant researchers as possible in the exercise. Everyone will have to register to access the data sets and provide feedback on algorithm performance, from the standard set of metrics implemented in the R-R system.

6.4.2 At the final validation phase

Once the final algorithm(s) is (are) selected and the BA production phase is performed, the first results of the project will be validated internally. After this internal validation is finished, the BA product will be made publicly available in two phases. The first one will be restricted to those experts working in this field, and will assure that an independent validation is carried out before the official release of the product. In addition to validation, intercomparison with other available BA products will be carried out. This exercise will be described in the Product Inter-comparison Report (PIR)

To ensure the validation process is endorsed by the whole community, external key scientists working actively on burned area mapping will be contacted and invited to participate in forums and specialised workshops. This integration of the community in this process provides the opportunity for this exercise to be a community effort and the partners, who are creating the products, can benefit from the input of these local experts. Relevant activities can be proposed for each end-user group participating in the project, while additional links with other communities through the CMUG may extend these modelling efforts.

6.5 Corrections after the internal (ESA officer) and external review of the PVP

Several amendments have been introduced from version 2.0 of the PVP, as a result of suggestions done by Stephen Plummer, ESA scientific officer for the fire_cci project, and the external evaluation of the PVP. The most remarkable one are the following:

1. An internal cross-checking procedure for the generation of fire reference data has been introduced (see 5.2).
2. The resampling of reference perimeter files from the original Landsat-TM spatial resolution of 30 m to 50 m pixel size has been removed (Comment 2 of external review). Since fire perimeters may be derived from other sensors (not necessarily from Landsat-TM/ETM+), the original resolution will be kept.
3. Non-parametric metrics have been introduced to avoid potential problems with non-normality of output results. Traditional linear regression analysis has been changed to non-parametric analysis (Kendall), as well as standard deviation to interquartile range.
4. Confidence intervals of all validation metrics will be computed.

Other suggestions of the external review of the PVP have not been considered. The main ones are the following:

1. Cleaning of small polygons related to the minimum mapping unit is a cartographic basis of any validation effort (comment 1 of external review). Validation implies to check whether the output product fulfils the requirements stated in the product specification. We consider meaningless to validate a product with a spatial detail that it is impossible to achieve because of the intrinsic limitations of the input dataset used. Any remote sensing classification states a minimum area under which objects will not be recognized. For instance, the CORINE land cover project considers 25 ha as the minimum mapping unit for rural areas and 5 ha for urban areas (both well above the Landsat-TM pixel size). Therefore, land cover patches below those thresholds are not used for validation. At this stage, the minimum burned patch for the fire_cci project is not fixed, but whenever is done, it should be consistent with the minimum patch considered for validation. Another issue is how much of the real burned area is covered in the output BA product, but this is not properly validation of the product. By analyzing the distribution of fire sizes from the reference information, we will try to establish a general estimation of the actual amount of burned area that may be missed as a result of the minimum BA mapping unit for different ecosystems.
2. Kappa is commonly used in remote sensing classification (Congalton & Green, 1999), since it provides synthetic estimations of omission and commission errors (comment 3 of external review). In addition, the variance of kappa can be computed and therefore, analysis of whether two classifications are significantly different or not accomplished. Certainly, since kappa considers the agreement of both burned and unburned areas, and the former will be most abundant, it may bias the agreement when the sites have a small proportion of burned areas. In any case, kappa is only one of the validation metrics being used and therefore its use will not significantly affect the BA assessment.

7 References

Asner, G.P., Scurlock, J.M.O., & Hicke, J.A. (2003), Global synthesis of leaf area index observations: implications for ecological and remote sensing studies. Global Ecology and Biogeography, 12, 191-205

Bachmann, M. et al. (2011), ESA CCI ECV Fire Disturbance - Algorithm Theoretical Basis Document – Volume I – Pre-processing: Fire_cci_Ph1_DLR_D2_3_1_ATBD_I_v0_2.pdf, https://www.esa-fire-cci.org/webfm_send/192

Bastarrika, A., Chuvieco, E., & Martin, M.P. (2011), Mapping burned areas from Landsat TM/ETM+ data with a two-phase algorithm: balancing omission and commission errors. Remote Sensing of Environment, 115, 1003-1012.

Boschetti, L., Flasse, S.p.P., & Brivio, P.A. (2004). Analysis of the conflict between omission and commission in low spatial resolution dichotomic thematic products: The Pareto Boundary. Remote Sensing of Environment, 91, 280-292..

Boschetti, L., Brivio, P.A., Eva, H.D., Gallego, J., Baraldi, A., & Gregoire, J.M. (2006), A sampling method for the retrospective validation of Global Burned Area products. IEEE Transactions on Geoscience and Remote Sensing, 44, 1765-1773.

Boschetti, L., Roy, D., & Justice, C. (2009), International Global Burned Area Satellite Product Validation Protocol. Part I – production and standardization of validation reference data. In CEOS-CalVal (Ed.) (pp. 1-11). USA: Committee on Earth Observation Satellites

Congalton, R.G., & Green, K. (1999), Assessing the Accuracy of Remotely Sensed Data: Principles and Applications. Boca Raton: Lewis Publishers.

Chuvieco, E. et al. (2011a), ESA CCI ECV Fire Disturbance - Data Access Requirements Document: Fire_cci_Ph1_UAH_DOC_D1_3_DARD_v1_7.pdf, https://www.esa-fire-cci.org/webfm_send/189

Chuvieco, E. et al. (2011b), ESA CCI ECV Fire Disturbance - Product Specification Document: Fire_cci_Ph1_UAH_DOC_D1_2_PSD_v2_0.pdf, https://www.esa-fire-cci.org/webfm_send/190

Chuvieco, E. (Ed.) (2008), Earth observation of global change: the role of satellite remote sensing in monitoring global environment. New York: Springer Berlin Heidelberg.

Chuvieco, E., Opazo, S., Sione, W., Del Valle, H., Anaya, J., Di Bella, C., Cruz, I., Manzo, L., López, G., Mari, N., González-Alonso, F., Morelli, F., Setzer, A., Csiszar, I., Kanpandegi, J.A., Bastarrika, A., & Libonati, R. (2008). Global Burned Land Estimation in Latin America using MODIS Composite Data. Ecological Applications, 18, 64-79.

Csiszar, I. A. et al. (2009), Assessment of the status of the development of the standards for the terrestrial ECV" T13 Report. Version 27, GTOS. Available online on: http://www.fao.org/gtos/ECV-T13.html

Fraser, R.H., Hall, R.J., Landry, R., Lynham, T., Raymond, D., Lee, B., & Li, Z. (2004), Validation and calibration of Canada-wide coarse-resolution satellite burned-area maps. Photogrammetric Engineering and Remote Sensing, 70, 451-460.

Giglio, L., Randerson, J.T., van der Werf, G.R., Kasibhatla, P.S., Collatz, G.J., Morton, D.C., & DeFries, R.S. (2010), Assessing variability and long-term trends in burned area by merging multiple satellite fire products. Biogeosciences Discuss., 7, 1171-1186, doi:1110.5194/bg-1177-1171-2010,

GlobCarbon (2007), Demonstration Products and Qualification Report. In (pp. 1-69): ESA

Morisette, J.T., Giglio, L., Csiszar, I., & Justice, C.O. (2005). Validation of the MODIS active fire product over Southern Africa with ASTER data. International Journal of Remote Sensing, 26, 4239-4264.

Padilla, M., & Chuvieco, E. (2009), Validación de productos globales de área quemada en la Península Ibérica. Revista de Teledetección, 31, 69-79.

Plummer, S. (2004), GLOBCARBON product validation plan. Frascati: ESA-ESRIN

Oliva, P. (2009), Métodos para la cartografía de Área quemada con imágenes del sensor ENVISAT-MERIS en la cuenca mediterránea. Alcalá de Henares, España: Departamento de Geografía, Universidad de Alcalá.

Randerson, J. T., G. R. van der Werf, L. Giglio, G. J. Collatz, & P. S. Kasibhatla. 2007. Global Fire Emissions Database, Version 2 (GFEDv2.1). Data set. Available on-line [http://daac.ornl.gov/] from Oak Ridge National Laboratory Distributed Active Archive Center, Oak Ridge, Tennessee, U.S.A. doi:10.3334/ORNLDAAC/849.

Roy, D., Frost, P., Justice, C., Landmann, T., Roux, J., Gumbo, K., Makungwa, S., Dunham, K., Du Toit, R., Mhwandagara, K., Zacarias, A., Tacheba, B., Dube, O., Pereira, J., Mushove, P., Morisette, J., Santhana, S., & Davies, D. (2005), The Southern Africa Fire Network (SAFNet) regional burned area product validation protocol. International Journal of Remote Sensing, 26, 4265-4292

Roy, D.P., Boschetti, L., Justice, C.O., & Ju, J. (2008). The collection 5 MODIS burned area product - Global evaluation by comparison with the MODIS active fire product. Remote Sensing of Environment, 112, 3690-3707

Roy, D.P., & Boschetti, L. (2009). Southern Africa validation of the MODIS, L3JRC, and GlobCarbon burned-area products. IEEE Transactions on Geoscience and Remote Sensing, 47, 1032-1044.

Schultz, M. et al. (2011), ESA CCI EVC Fire Disturbance - User Requirements Document: Fire_cci_Ph1_JUELICH_D1_1_URD_v2_0.pdf, https://www.esa-fire-cci.org/webfm_send/137

Sen, P.K. (1968). Estimates of the regression coefficient based on Kendall's tau. Journal of the American Statistical Association, 63, 1379-1389.

Tansey, K., Grégoire, J.-M., Defourny, P., Leigh, R., Pekel, J.-F., Bogaert, E., & Bartholome, E. (2008). A new, global, multi-annual (2000-2007) burnt area product at 1 km resolution. Geophysical Research Letters, 35, L01401, doi:10.1029/2007GL03156.

Theil, H. (1950), A rank-invariant method of linear and polynomial regression analysis. Nederlandse Akademie Wetenchappen Series A., 53, 386-392

Annexe 1: Metadata fields of the reference data

Section	Field to complete in ArcCatalog	How to complete the Field
Description		
	Keywords:	"Burned Area"
	Abstract:	Indicate the test site and the year of the data (e.g. "Burned Area for Kazakhstan Test site, 2002")
	Purpose:	"Validation and Reference"
Time period for which the data is relevant		
	Currentness Reference:	"ground condition"
	Beginning Date:	The beginning of the period of the fires included in this file
	Ending Date:	The end of the period of the fires included in this file
Publication Information and Data storage and access information		
	Title:	Name of the data file (e.g.: "20020613_20020731_165026")
	Originator:	Name of the organization who generated the file
	Publication Date:	Publication Date of the data
	Geospatial Data Presentation Form:	"vector digital data"
	Online Linkage:	http://www.esa-fire-cci.org/
Details about this document		
	Metadata Date:	Publication Date of the metadata
Who completed this document		
	Person:	Who completed the metadata file
	Organization:	The organization that the person belongs to
Standards used to create this document		
	Metadata Standard Name:	"FGDC Content Standards for Digital Geospatial Metadata"
	Metadata Standard Version:	FGDC-STD-001-1998

Section	Field to complete in ArcCatalog	How to complete de Field
Spatial		
Horizontal coordinate system		
	Geographic Coordinate System Name:	
	Projected Coordinate System Name:	
	Horizontal Datum Name:	
	Ellipsoid Name:	
	Planar Distance Units:	"metres"
	Encoding Type:	"Universal Transverse Mercator
	UTM Zone Number:	
Bounding coordinates		
	North:	
	South:	
	East:	
	West:	
Lineage		
	Source Contribution:	The use of each source used for generating perimeters, indicating the names of the quicklooks of the images if generated.
	Source Citation Abbreviation:	Name of each satellite image (e.g. LE71700762005159ASN00) or name of the fire database (e.g. Canadian Wildland Fire Large Database)
	Originator:	The organization responsible of each source
	Online Linkage:	Link of the webpage where is available each source (e.g. http://glovis.usgs.gov/)
	Process description:	Thresholds used in ABAMS for Phase 1 (Seeds) and Phase 2 (Contextual Region Growing)
	Process Software and Version:	"ABAMS"
Attributes		Explain each attribute of the attribute table
	Label:	Label of each attribute
	Type:	Type of the data of each attributes (string, integer, etc.)
	Definition:	Description of each attribute specifying the units used

Annexe 2: Conditions to access the fire_cci datasets

The fire_cci project acknowledges the open policy of the ESA CCI programme with respect to access to information (data and methods) from external scientists. However, since the validation exercises require tracking the sources of errors, scientists interested to use the validation or input datasets will be asked to fill a simple letter of agreement, which should assure adequate feedback to improve the project outputs and usability. This agreement should be signed by scientists participating in the Round Robin and validation exercises, as well as those interested in using the fire_cci results.

The following information will be collected for each person that intends to download data from the fire_cci project.

Personal information

Name:

Institution:

Contact details (email, web page, address, telephone)

Main field of activity (research, education, management)

Main application of BA products (carbon models, climate, emissions, terrestrial ecology, land cover change, landscape analysis, ….

Selected username and password

General agreement statement

1. I acknowledge that the attached datasets were generated under the Fire Disturbance (fire_cci) project, which is part of the ESA Climate Change Initiative (ESA contract Nº 4000101779/10/I-NB).
2. The origin of the data will be cited in all publications derived from the analysis of the attached datasets.
3. Feedback related to errors or inconsistencies in the datasets should be reported to the fire_cci contact points (http://www.esa-fire-cci.org/content/contact-points).

Specific agreement for participating in the Round Robin exercise

1. I accept to participate in the Round-Robin exercise of the fire_cci project to test the performance of alternative burned area algorithms in selected study sites or the standard algorithms in alternative study sites.
2. The assessment of alternative algorithms should be done on the following basis:
    a. The standard pre-processing data should be used as input to the exercise (no alternative atmospheric correction or cloud-water-snow masking should be included).
    b. The algorithms should be tested at least with 4 sites and at least with 3 full year of data each.
    c. The standard validation metrics should be used. The results of alternative algorithms should be uploaded to the internet interface of the R-R exercise to compute validation metrics with the standard BA reference files.
    d. Results should be sent in before January 15th, 2012.
3. Those scientists interested in running the standard algorithms on their own study sites, should upload the datasets required by the standard algorithms, as detailed in the internet interface of the R-R exercise.

Annexe 3: Review of the fire_cci PVP by the KSB

E-mail received from Joanne M. Nightingale, chairperson of the CEOS-LPV, with the evaluation of the PVP performed by Luigi Boschetti, co-chairperson of the Fire validation subgroup.

Dear Emilio,

Luigi has kindly provided a technical review of the Fire CCI validation plan on behalf of LPV. On another note, given you've witnessed firsthand the issues we are having with reviewing GCOS/CEOS documents, I was wondering if you have a formal review process in place for your document? I.e. documentation on who was asked to comment, as well as a journal type response to the comments you have accepted/rejected etc?

Cheers, Joanne

Hi Joanne,

I have a few comments on the latest version of the CCI validation plan. The most serious point is #1, which could results into inflated accuracy (and gets dangerously close to massaging the data).

1) section 6.2.1.1 is not entirely clear. As far as I have understood, step 1 (cleaning of the small polygons) applies to the reference data, i.e. all the polygons smaller than the patch size defined in the Product Specification Document are removed from the validation dataset. I believe that this step is quite inappropriate, and would artificially inflate the accuracy measures.

The function of the reference data is to provide the best possible approximation of the reality, and then the accuracy metrics describe the conformity between the low resolution product and this approximation of the reality. If a significant amount of burning happens in small patches, that cannot be detected at the scale of the Fire CCI product, the validation metrics should report such omission error, even if it is due solely to the minimum size of the detected burns, and not to actual misclassification.

Of course in the internal product QA phase it would be perfectly acceptable to remove the small patches, as the focus in the QA is on how well the available input data is classified (even if the input data has limitations), rather than on how well the output product represents the reality.

2) the transformation and reprojection of the Landsat data from 30 to 50 m will introduce a significant amount of aliasing errors. In the worst case of small, fragmented burns this could result in a significant bias in the areal estimates. An alternative strategy would be:

a) reprojection of the relevant spatial subset of fire CCI products into UTM

b) oversampling at 30 m of the fire CCI products.

This alternative procedure will still allow the pixel by pixel comparison, but will reduce the aliasing - the 30 to 50 metre resampling causes resampling problems in the majority of the pixels, while the 1000m to 30 metre oversampling will cause problem only in a small percentage of pixels.

3) validation metrics: I do not believe that kappa is an appropriate accuracy metric to describe the accuracy of moderate resolution burned area products; as discussed at length by Stehman, 1996 ("Selecting and Interpreting measures of thematic classification accuracy", RSE 62:77-89), the kappa coefficient compensates for chance agreement between two populations, in the hypothesis that the marginal distribution of the classes is know a priori. In our case, this translates in knowing a priori the total area burned in each Landsat scene! This is not at all the case, and a correct estimate of the total area burned would indicate alone, as a matter of fact, quite a good performance of the algorithm. Furthermore, quoting again Stehman (1996): "If the opbjective is to describe the accuracy of a final map product, the user of this particular map is probably not concerned with the hypotetical proportion of area classified correctly by random chance"