



**fire**  
cci

---

## ESA Climate Change Initiative – Fire\_cci

### D1.3 Comprehensive Error Characterization Report (CECR)

---

<b>Project Name</b>	ECV Fire Disturbance: Fire_cci Phase 2
<b>Contract N°</b>	4000115006/15/I-NB
<b>Issue Date</b>	18/05/2018
<b>Version</b>	2.0
<b>Author</b>	Philip Lewis, José Gomez-Dans, James Brennan, Maxim Chernetskiy
<b>Document Ref.</b>	Fire_cci_D3.1_CECR_v2.0
<b>Document type</b>	Public

*To be cited as: P. Lewis, J. Gomez-Dans, J. Brennan, M. Chernetskiy (2018) ESA CCI ECV  
Fire Disturbance: D3.1. Comprehensive Error Characterization Report, version 2.0.  
Available at: <http://www.esa-fire-cci.org/documents>*

	<b>Fire_cci</b>	Ref.:	Fire_cci_D1.3_CECR_v2.0		
	<b>Comprehensive Error Characterization</b>	Issue	2.0	Date	18/05/2018
	<b>Report</b>	Page	2		

## Project Partners

Prime Contractor/ Scientific Lead & Project Management	UAH – University of Alcalá (Spain)
Earth Observation Team	UAH – University of Alcalá (Spain)
	EHU – University of the Basque Country (Spain)
	UL – University of Leicester (United Kingdom)
	UCL – University College London (United Kingdom)
System Engineering	ISA – School of Agriculture, University of Lisbon (Portugal)
	BC – Brockmann Consult (Germany)
Climate Research Group	MPIC – Max Planck Institute for Chemistry (Germany)
	IRD - Research Institute for Development (France)
	LSCE - Climate and Environmental Sciences Laboratory (France)
	VUA - Vrije Universiteit Amsterdam (Netherlands)



## Distribution

Affiliation	Name	Address	Copies
ESA	Stephen Plummer (ESA)	stephen.plummer@esa.int	electronic copy
Project Team	Emilio Chuvieco (UAH)	emilio.chuvieco@uah.es	electronic copy
	M. Lucrecia Pettinari (UAH)	mlucrecia.pettinari@uah.es	
	Joshua Lizundia (UAH)	joshua.lizundia@uah.es	
	Gonzalo Otón (UAH)	gonzalo.oton@uah.es	
	Mihai Tanase (UAH)	mihai.tanase@uah.es	
	Miguel Ángel Belenguer (UAH)	miguel.belenguer@uah.es	
	Aitor Bastarrika (EHU)	aitor.bastarrika@ehu.es	
	Ekhi Roteta (EHU)	ekhi.roteta@gmail.com	
	Kevin Tansey (UL)	kjt7@leicester.ac.uk	
	Marc Padilla Parellada (UL)	mp489@leicester.ac.uk	
	James Wheeler (UL)	jemw3@leicester.ac.uk	
	Philip Lewis (UCL)	ucfalew@ucl.ac.uk	
	José Gómez Dans (UCL)	j.gomez-dans@ucl.ac.uk	
	James Brennan (UCL)	james.brennan.11@ucl.ac.uk	
	Maxim Chernetskiy (UCL)	m.chernetskiy@ucl.ac.uk	
	Jose Miguel Pereira (ISA)	jmocpereira@gmail.com	
	Duarte Oom (ISA)	duarte.oom@gmail.com	
	Manuel Campagnolo (ISA)	mlc@isa.ulisboa.pt	
	Thomas Storm (BC)	thomas.storm@brockmann-consult.de	
	Johannes Kaiser (MPIC)	j.kaiser@mpic.de	
Angelika Heil (MPIC)	a.heil@mpic.de		
Florent Mouillot (IRD)	florent.mouillot@cefe.cnrs.fr		
M. Vanesa Moreno (IRD)	marivanesa.morenodominguez@cefe...		
Philippe Ciais (LSCE)	philippe.ciais@lsce.ipsl.fr		
Chao Yue (LSCE)	chaoyuejoy@gmail.com		
Pierre Laurent (LSCE)	pierre.laurent@lsce.ipsl.fr		
Guido van der Werf (VUA)	guido.vander.werf@vu.nl		
Ioannis Bistinas (VUA)	i.bistinas@vu.nl		

	<b>Fire_cci</b> <b>Comprehensive Error Characterization</b> <b>Report</b>	Ref.:	Fire_cci_D1.3_CECR_v2.0		
		Issue	2.0	Date	18/05/2018
		Page	3		

## Summary

This document is the Comprehensive Error Characterisation Report, and provides an assessment of error characterisation in the Burned Area (BA) products.

	Affiliation/Function	Name	Date
Prepared	UCL	Philip Lewis José Gomez-Dans James Brennan Maxim Chernetskiy	01/05/2018
Reviewed	UAH – Project Manager	M. Lucrecia Pettinari	18/05/2018
Authorized	UAH - Science Leader	Emilio Chuvieco	18/05/2018
Accepted	ESA - Technical Officer	Stephen Plummer	

This document is not signed. It is provided as an electronic copy.

## Document Status Sheet

Issue	Date	Details
1.0	07/03/2017	First Document Issue
1.1	02/06/2017	Addressing comments of CCI-FIRE-EOPS-MM-17-0040.
2.0	18/05/2018	General update of the document, including results and discussion of the results

## Document Change Record

Issue	Date	Request	Location	Details
1.1	02/06/2017	ESA	Section 2.1	New section 'Example of uncertainty estimation through error propagation'.
2.0	18/05/2018	UCL	Section 1 Sections 2.1, 2.2  Sections 3, 4, 5, 6, 7, 8 and their corresponding sub-sections	Text updated. Text expanded and sub-divided into two sub-sections. Sections added.



**Table of Contents**

**1. Introduction .....8**

**2. General approach .....8**

    2.1. Example of uncertainty estimation through error propagation: BRDF effects..... 10

    2.2. Example of uncertainty estimation through error propagation: Burn indices ..... 16

**3. Methods ..... 18**

    3.1. Rationale..... 18

    3.2. Description of reference sites ..... 19

    3.3. Generation of surface reflectance realisations ..... 19

        3.3.1. Estimation of true surface reflectance ..... 19

        3.3.2. Simulation of observation uncertainty..... 23

        3.3.3. Simulation of observation opportunity ..... 24

        3.3.4. Estimation of true algorithm uncertainty *P<sub>bT</sub>* ..... 24

**4. Grid scale uncertainty aggregation .....25**

    4.1. Aggregation basics.....26

    4.2. Unreliable probability of burn estimates .....28

**5. Uncertainty descriptions for algorithms.....29**

    5.1. Fire\_cci v5.0: Uncertainty calculations .....29

    5.2. SLSTR uncertainty characterization.....29

**6. Results.....30**

    6.1. Uncertainty assessment for the Pixel Product .....30

        6.1.1. The MODIS Fire CCI v5.1 ..... 30

        6.1.2. The SLSTR algorithm ..... 33

    6.2. Uncertainty assessment for the Grid Product .....36

        6.2.1. The MODIS Fire CCI v5.1 ..... 36

        6.2.2. The SLSTR algorithm ..... 38

**7. Discussion .....40**

    7.1. MODIS Fire\_cci v5.1 .....40

    7.2. SLSTR algorithm.....41

**8. Conclusion .....42**

**9. References.....44**

**10. Annex 1: Acronyms and abbreviations .....46**

**List of Figures**

Figure 1: Probability Distribution Function of examples of classes B and U. .... 9

Figure 2: Cumulative Probability of class membership for examples of classes B and U. .... 9

Figure 3: Illustrative surface BRDF data as a function of time for MODIS land bands. Fire takes place on DoB 228 (confirmed by a thermal anomaly). Uncertainty is modelled as in the text with  $\alpha = 0.03$ ..... 12



Figure 4: Near infrared reflectance normalised with regularisation ( $\gamma=1e3$ ) (blue) with +/- 1 std. Original samples and uncertainty shown in green. .... 14

Figure 5: Near infrared reflectance normalised with edge-preserving regularisation ( $\gamma=1e3$ ) (blue) with +/- 1 std. Original samples and uncertainty shown in green..... 14

Figure 6: Zoom of near infrared reflectance normalised with edge-preserving regularisation ( $\gamma=1e3$ ) (blue) with +/- 1 std. Original samples and uncertainty shown in green. Non-edge-preserving result shown by red dashed line..... 15

Figure 7: Spectral reflectance (with uncertainty) pre- (red) and post-fire (black). .... 15

Figure 8: Reflectance (with uncertainty) in R (red) and NIR (black). .... 17

Figure 9: NDVI (with uncertainty). .... 17

Figure 10: Reflectance (with uncertainty) in band 6 (red) and band 4 (black). .... 17

Figure 11: NBR (with uncertainty). .... 17

Figure 12: Delta NDVI and its uncertainty ..... 18

Figure 13: Delta NBR and its uncertainty ..... 18

Figure 17: Location of the selected test sites..... 19

Figure 18: Cloud cleared surface reflectance from the MOD09GA and MYD09GA daily products for the two sites. Left: no fire, right: fire happening on day 247..... 22

Figure 19: Isotropic reflectance for the two sites. Left) no fire, Right) fire..... 23

Figure 20: (Top) One realisation of the surface reflectance over a pixel experiencing a fire. (Bottom) 20 realisations produce a sampling from the probability density function of the observations..... 24

Figure 14: The Poisson binomial PDF (green line) derived from a simulated set of independent samples (300, 100 with probabilities between 0.7 and 0.9, 100 with probabilities between 0.2-0.3 and 100 with probabilities between 0-0.1). A Gaussian approximation (red line) derived from calculating the mean (~110) and standard deviation (~39) is also shown. Skewness was ~0.01..... 28

Figure 15: Example of aggregation. See text for more details ..... 28

Figure 16: Example of applying the rescale approach to the example shown in Figure 15. .... 29

Figure 21: Fire CCI algorithm v5.1, Tropical site. Inferred probability of burn (left column), mean algorithm reported probability of burn (middle column) and standard deviation of algorithm reported probability of burn (right column). Different rows represent different noise levels: 0.5 nominal (top row), nominal (middle) and 1.5 nominal (bottom). All plots share the same scale..... 31

Figure 22: Fire CCI algorithm v5.1, Savannah site. Inferred probability of burn (left column), mean algorithm reported probability of burn (middle column) and standard deviation of algorithm reported probability of burn (right column). Different rows represent different noise levels: 0.5 nominal (top row), nominal (middle) and 1.5 nominal (bottom). All plots share the same scale..... 32

Figure 23: Fire CCI algorithm v5.1, Boreal site. Inferred probability of burn (left column), mean algorithm reported probability of burn (middle column) and standard deviation of algorithm reported probability of burn (right column). Different rows represent different noise levels: 0.5 nominal (top row), nominal (middle) and 1.5 nominal (bottom). All plots share the same scale..... 33

Figure 24: SLSTR algorithm, Tropical site. Inferred probability of burn (left column), mean algorithm reported probability of burn (middle column) and standard deviation of algorithm reported probability of burn (right column). Different rows represent different noise levels: 0.5 nominal (top row), nominal (middle) and 1.5 nominal (bottom). All plots share the same scale..... 34

	<b>Fire_cci</b> <b>Comprehensive Error Characterization</b> <b>Report</b>	Ref.:	Fire_cci_D1.3_CECR_v2.0		
		Issue	2.0	Date	18/05/2018
		Page	6		

Figure 25: SLSTR algorithm, Savannah site. Inferred probability of burn (left column), mean algorithm reported probability of burn (middle column) and standard deviation of algorithm reported probability of burn (right column). Different rows represent different noise levels: 0.5 nominal (top row), nominal (middle) and 1.5 nominal (bottom). All plots share the same scale. .... 35

Figure 26: SLSTR algorithm, boreal site. Inferred probability of burn (left column), mean algorithm reported probability of burn (middle column) and standard deviation of algorithm reported probability of burn (right column). Different rows represent different noise levels: 0.5 nominal (top row), nominal (middle) and 1.5 nominal (bottom). All plots share the same scale. .... 36

Figure 27: Fire CCI v5.1 Tropical site. Aggregation to CMG results. Each panel shows the aggregated BA PDF of a 100x100 pixel area in the original pixel product (x-axis: burned pixels per area, y-axis: frequency). The dashed lines represent the PDF calculated from the inferred pb value, and the full lines represent the PDF calculated from the mean of the algorithm reported pb across the 20 MC realisations. Different colours represent different noise levels (blue: 0.5 times nominal noise level, orange: nominal noise level and green 1.5 times nominal noise level). .... 37

Figure 28: Fire CCI v5.1 Savannah site. Aggregation to CMG results. Each panel shows the aggregated BA PDF of a 100x100 pixel area in the original pixel product (x-axis: burned pixels per area, y-axis: frequency). The savannah site only has four windows). The dashed lines represent the PDF calculated from the inferred pb value, and the full lines represent the PDF calculated from the mean of the algorithm reported pb across the 20 MC realisations. Different colours represent different noise levels (blue: 0.5 times nominal noise level, orange: nominal noise level and green 1.5 times nominal noise level). .... 37

Figure 29: Fire CCI v5.1 boreal site. Aggregation to CMG results. Each panel shows the aggregated BA PDF of a 100x100 pixel area in the original pixel product (x-axis: burned pixels per area, y-axis: frequency). The dashed lines represent the PDF calculated from the inferred pb value, and the full lines represent the PDF calculated from the mean of the algorithm reported pb across the 20 MC realisations. Different colours represent different noise levels (blue: 0.5 times nominal noise level, orange: nominal noise level and green 1.5 times nominal noise level). .... 38

Figure 30: SLSTR Tropical site. Aggregation to CMG results. Each panel shows the aggregated BA PDF of a 100x100 pixel area in the original pixel product (x-axis: burned pixels per area, y-axis: frequency). The dashed lines represent the PDF calculated from the inferred pb value, and the full lines represent the PDF calculated from the mean of the algorithm reported pb across the 20 MC realisations. Different colours represent different noise levels (blue: 0.5 times nominal noise level, orange: nominal noise level and green 1.5 times nominal noise level). .... 39

Figure 31: SLSTR Savannah site. Aggregation to CMG results. Each panel shows the aggregated BA PDF of a 100x100 pixel area in the original pixel product (x-axis: burned pixels per area, y-axis: frequency). The dashed lines represent the PDF calculated from the inferred pb value, and the full lines represent the PDF calculated from the mean of the algorithm reported pb across the 20 MC realisations. Different colours represent different noise levels (blue: 0.5 times nominal noise level, orange: nominal noise level and green 1.5 times nominal noise level). .... 39

Figure 32: SLSTR Boreal site. Aggregation to CMG results. Each panel shows the aggregated BA PDF of a 100x100 pixel area in the original pixel product (x-axis: burned pixels per area, y-axis: frequency). The dashed lines represent the PDF



**Fire\_cci  
Comprehensive Error Characterization  
Report**

Ref.:	Fire_cci_D1.3_CECR_v2.0		
Issue	2.0	Date	18/05/2018
Page			7

calculated from the inferred pb value, and the full lines represent the PDF calculated from the mean of the algorithm reported pb across the 20 MC realisations. Different colours represent different noise levels (blue: 0.5 times nominal noise level, orange: nominal noise level and green 1.5 times nominal noise level)..... 40

Figure 33: Uncertainty space for the Fire CCI v5.1. The plot shows how reported uncertainty is calculated as a function of near infrared reflectance of the pixel (*y* axis) and the relative drop in near infrared reflectance between the fire data and the previous observation (*x* axis). The contours show the values of the probability of burn calculation..... 41

Figure 34: Comparison of inferred (or expected) probability of burn versus algorithm reported mean probability of burn across MC realisations for boreal, nominal noise level. .... 42

Figure 35: Comparison of inferred (or expected) probability of burn versus algorithm reported mean probability of burn across MC realisations for subtropical site, nominal noise level. .... 42



## 1. Introduction

Error characterisation and validation are critical phases to generate any Essential Climate Variable (ECV), and therefore both have been included as key deliverables of the ESA CCI programme. All products generated by the CCI are required to have associated per pixel uncertainty characterisation. The details of how uncertainty characterisation is calculated are included in the relevant ATBD, though we will provide overviews in this document for clarity. Deriving burned area from spaceborne sensors is an indirect measurement, with a number of processes having an impact on how good the retrieval can be. Validation efforts compare products with a known reference to assess the overall quality of the product. However, validation exercises do not have information on whether a particular pixel is highly certain to be a burn or not. This is important and relevant information required by users. In the CECR, an assessment of the uncertainty characterisation of the BA product is included.

This is produced as deliverable D1.3 (WP2600), addressing task requirements 5 and 10 due for calendar completion M18 (draft) and M30 (final).

We start from the assumption that a per pixel estimate giving *the probability that a pixel should be labelled 'burned'* is provided. The main purpose of this report is to assess these uncertainty estimates.

Additionally, and as climate users often require data aggregated to coarser grid resolution (e.g. half a degree), an approach to aggregate both burned area data from the pixel product to the climate modeller's grid scale will be presented.

## 2. General approach

BA products are generally binary, i.e. they characterise a decision made in the algorithm to class a pixel as being 'fire affected' (burned) or not. The information is then generally used to estimate a total area affected by fire as effective pixel size times the number of 'fire affected' pixels (e.g. for fire emissions estimates).

In many cases we can trace the decision back to a probabilistic measure, i.e. the algorithm may define a class membership measure of whether a pixel should be classed as burned, which may sometimes be phrased as the probability of being burned, relative to not being burned. In such a case, the product will assign a pixel to the relevant class, based on some probabilistic threshold.

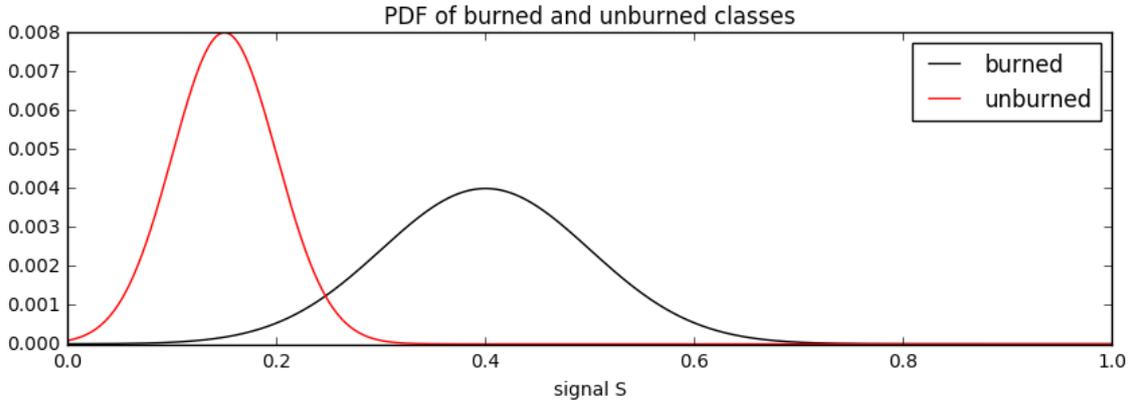
For example, we define a probability distribution function (PDF)  $P(s|B)$  that characterises the probability that signal  $s$  belongs to class  $B$  (burned) and  $P(s|U)$  that it belongs to class  $U$  (unburned) (illustrated in Figure 1). In algorithms, these are typically obtained from training data.

Since this is a two-class problem,  $P(s) = P(s|B) + P(s|U)$ . From Bayes theorem:

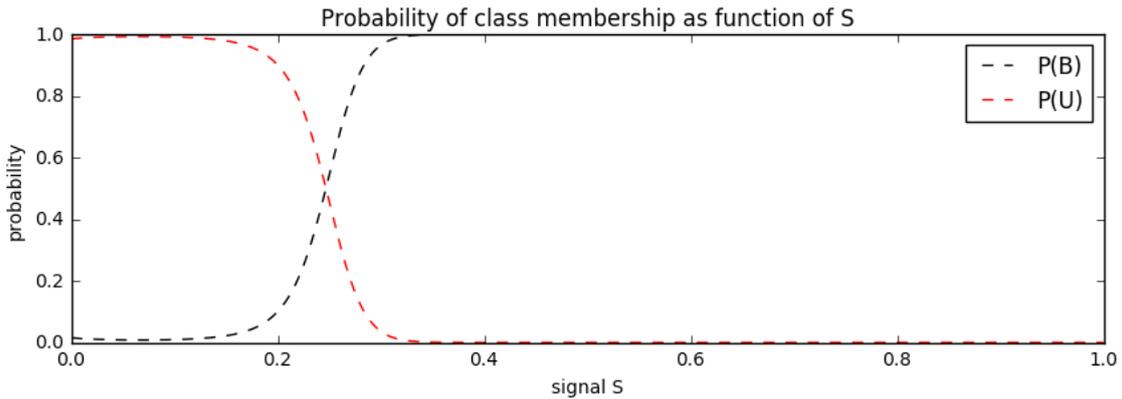
$$P(B|s) = \frac{P(s|B)P(B)}{P(s)}$$

Then the *relative probability* of the pixel belonging to the burned class, assuming equal prior probability of class membership is (Figure 2):

$$P_r(B) = \frac{P(B|s)}{P(B|s) + P(U|s)}$$



**Figure 1: Probability Distribution Function of examples of classes B and U.**



**Figure 2: Cumulative Probability of class membership for examples of classes B and U.**

This metric is related to the likelihood ratio:

$$\Lambda = \frac{P_r(B)}{P_r(U)} = \frac{P_r(B)}{1 - P_r(B)}$$

In this example, we use a simple linear discriminant. If the probability of class membership is greater than 0.5, then the probability of the pixel being correctly attributed to class *B* is greater than that of class *U*, as illustrated in Figure 2. In the example, this occurs at a signal threshold of around  $S = 0.25$ . Because of class overlap, applying a threshold to  $S$  at this value will tend to misclassify around 3% of class *U* as *B*, and around 6% of *B* as *U* (the difference in these percentages is apparent from the relative maximum magnitudes in Figure 1). For a lower threshold, we will increase the number of pixels classified as *B*, get fewer omission errors, but increase the commission errors. If  $s$  is between around 0.1 and 0.4 here, we will always get some non-trivial amount of error because of class overlap. Only beyond these limits can we be almost

	<b>Fire_cci</b>	Ref.:	Fire_cci_D1.3_CECR_v2.0		
	<b>Comprehensive Error Characterization</b>	Issue	2.0	Date	18/05/2018
	<b>Report</b>	Page	10		

certain of class membership. We use this as an example to illustrate the approach to error characterisation in the CCI.

In our approach to uncertainty characterisation, we suppose that the BA algorithm provides as input a signal  $s$  with an associated standard error  $\sigma_s$ , i.e.  $s$  is a PDF. The standard error estimate would normally be arrived at by error propagation from the inputs through all elements of the algorithm up to the calculation of  $s$  (e.g. noise in the input data, but also estimates of uncertainty in the definition of the burned/unburned classes, etc.).

So here,  $s$  provides the estimate of the mean signal and  $\sigma_s$  a measure of its corruption by input data noise and other effects. If we are sure of  $s$  (i.e.  $\sigma_s$  is low), we can just ‘read’ the probabilities off the graph in Figure 2: e.g. if  $s$  is 0.25, then the probability of it being in class  $B$  rather than  $U$  is 0.5. If  $\sigma_s$  is non-negligible, then we must marginalise  $P(B)$  over  $S$ , which will have the impact of broadening the effective PDFs of the class distribution and increasing the region of overlap.

For most current applications of a burned area product, the user will want a characterisation of the PDF of some aggregate area burned, not particularly caring about whether an individual pixel is correctly labelled or not. If BA per pixel is reported simply as a binary term, an aggregate sum can be simply calculated, but would take no account of uncertainty. If instead we base the aggregation directly on per pixel measures of  $P(B|s)$  (and assume statistics independent for each pixel), we instead arrive at a Poisson binomial distribution.  $P(B|s)$  is also useful for spatial filtering, e.g. region growing of fire patches, in which case the probabilities would have spatial dependencies and should probably be described as patch area and area uncertainty.

## 2.1. Example of uncertainty estimation through error propagation: BRDF effects

To illustrate this, consider a sequence of surface BRDF measurements over some spatial support at time sequence  $t$ ,  $\rho_t$ . We will use linear models to transform this to a burn signal  $s$ , and show how the uncertainty in  $s$  may be estimated using error propagation.

Assume each observation has an uncertainty that can be described with a Gaussian of variance  $\sigma_t^2$ . This uncertainty measure will take into account impacts of uncertainty in calibration, atmospheric correction and possibly other effects (such as IFOV issues). In general, we would expect the uncertainty to be higher for higher off-nadir viewing and illumination angles (longer path length, and IFOV effects) and for higher concentrations of atmospheric constituents. It may also scale with the absolute value of reflectance. For illustrative purposes then, let us assume:

$$\sigma_t^2 = \alpha^2 \frac{\rho_t}{|\cos\theta_{vt}\cos\theta_{st}|}$$

where  $\theta_{vt}$  is the viewing zenith angle,  $\theta_{st}$  the illumination zenith angle and  $\alpha$  a scalar that will be a function of atmospheric concentrations and wavelength. An example of such a dataset is illustrated below, with a fire event in the time series around the middle of the sequence. We assume in this illustration that uncertainties between wavebands are uncorrelated.

We fit a model  $K$  to describe the BRDF effects with:

$$\hat{\rho} = \rho + \varepsilon_\rho = Kx$$



with  $\varepsilon_\rho$  the residual in modelling of  $\rho$  and  $x$  a state vector describing the variation in BRDF with viewing and illumination angles. We will use  $x$  to normalise for angular effects in looking for a burn signal, specifically, the parameter  $x_0$ :

$$x_0 = Vx = [1 \ 0 \ 0]x$$

Let:

$$E(\varepsilon_\rho^2) = \sigma_t^2$$

We cannot solve for state  $x_0$  at all time steps as the problem stands, as there will generally be insufficient information in the data alone (ill-posed problem). To rectify this, we introduce a smoothness constraint. First, consider  $\delta x$  as a departure from the mean state  $\bar{x}$ :

$$x = \delta x + \bar{x}$$

Then assume  $\delta x$  is smooth so that:

$$\nabla \delta x = \varepsilon_d$$

where  $\nabla$  is a discrete time difference operator (the ‘process model’ in a data assimilation framework) and  $\varepsilon_d$  (the first order difference model here) residual.

Then:

$$E(\nabla \delta x) = 0$$

$$E(\varepsilon_d^2) = \frac{1}{\gamma}$$

In a Bayesian sense, the model used to describe the BRDF effects of the observations that includes the uncertainty due to residual atmospheric correction effects and changes in IFOV as an additive Gaussian term can be thought of as a likelihood function, whereas the term that encodes a belief in the smooth temporal trajectory of  $x$  is a prior PDF (again, Gaussian). Given that all operators are linear and all statistics are Gaussian, the posterior PDF is also Gaussian. The variance (or in this case, covariance matrix) of the posterior PDF is what we here term “uncertainty”. In what follows, we show a way of arriving at this result.

We can provide an estimate of  $x$  and its uncertainty now by phrasing as an optimal estimation problem. We form a cost function (proportional to the negative log posterior) which is the sum of two terms:

$$J = J_\rho + J_d$$

With

$$J_\rho = \frac{1}{2} z^T z,$$

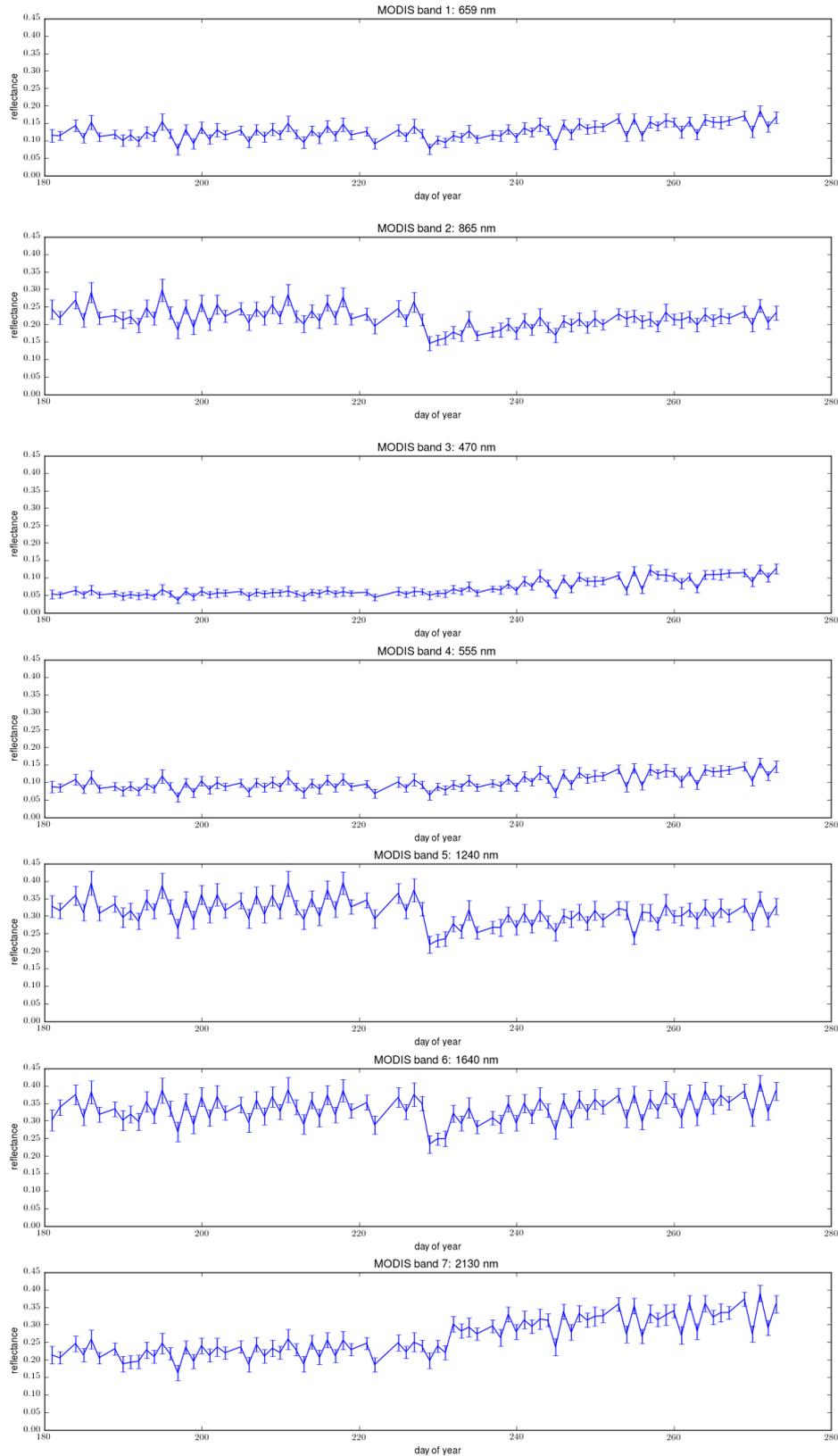
with

$$z = \frac{\rho - Kx}{\sigma_t}$$

and  $T$  the transpose operator. This term is the likelihood (or “fit to the observations” term). Further, we have the prior term:



$$J_d = \gamma \frac{1}{2} (\nabla \delta x)^T (\nabla \delta x) = \gamma \frac{1}{2} (\nabla x)^T (\nabla x)$$



**Figure 3: Illustrative surface BRDF data as a function of time for MODIS land bands. Fire takes place on DoB 228 (confirmed by a thermal anomaly). Uncertainty is modelled as in the text with  $\alpha = 0.03$ .**

Differentiating the cost functions with respect to state:

$$J' = J'_\rho + J'_d$$

$$J'_\rho = z^T z' = -\frac{K^T}{\sigma_t} \left( \frac{\rho - Kx}{\sigma_t} \right)$$

$$J'_d = \gamma \nabla^T \nabla x$$

The mean optimal solution is found when  $J' = 0$ . Remember that this is the negative log posterior, so this same point is the maximum *a posteriori* estimate:

$$-\frac{K^T}{\sigma_t} \left( \frac{\rho - Kx}{\sigma_t} \right) + \gamma \nabla^T \nabla x = 0$$

so:

$$\left( \frac{K^T K}{\sigma_t^2} + \gamma \nabla^T \nabla \right) x = \frac{K^T \rho}{\sigma_t^2}$$

As the posterior is the sum of the log of two Gaussians, it can be written as a single Gaussian, with the mean given by the solution of the previous linear system of equations, and the covariance matrix (the measure of uncertainty) is given by the inverse of the radius of curvature of the cost function at the minimum:

$$C_x^{-1} = \left( \frac{K^T K}{\sigma_t^2} + \gamma \nabla^T \nabla \right)^{-1}$$

so we can analytically calculate the uncertainty in  $x_0$  through (Lucht and Lewis, 2000):

$$\sigma^2 = V^T C_x^{-1} V$$

which is equivalent to the first element on the matrix leading diagonal in this case.

From

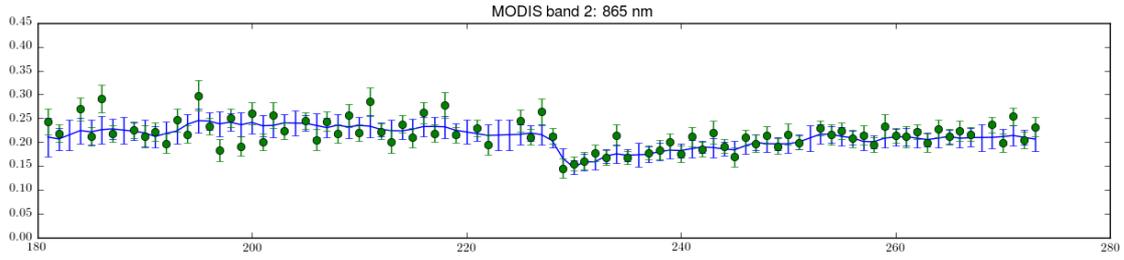
$$\left( \frac{K^T K}{\sigma_t^2} + \gamma \nabla^T \nabla \right) x = \frac{K^T \rho}{\sigma_t^2}$$

we write:

$$Ax = b$$

with  $A$  the 'information matrix' and  $b$  the 'information vector'. We can then use standard linear solvers to estimate  $x$ . The uncertainty in  $x$  is then given by  $A^{-1}[0,0]$ .

This is illustrated for the near infrared in Figure 4.



**Figure 4: Near infrared reflectance normalised with regularisation ( $\gamma=1e3$ ) (blue) with +/- 1 std. Original samples and uncertainty shown in green.**

The modelling effectively removes the artefacts from angular sampling, as well as providing an interpolation where samples are missing. Uncertainty estimates are given using the equations above, and are available for all time samples. Notice that the ‘interpolation’ uncertainty is higher than that when samples are available as we would expect. The magnitude of this is impacted by the regularisation term  $\gamma$ . One problem with this approach for the detection of burn signals is that the step edge apparent in the data is smoothed over in the result above. This is because we assume the ‘degree of smoothness’ (expectation of change) is constant over the entire signal. This issue is easily remedied, by introducing a relaxation vector  $R$  into the equation above:

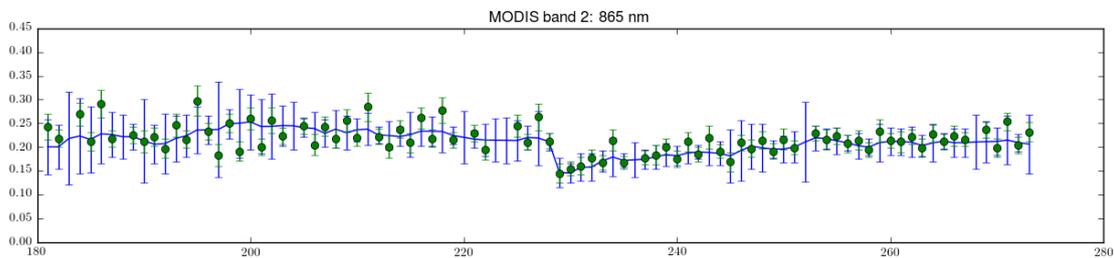
$$\left( \frac{K^T K}{\sigma_t^2} + \gamma \nabla^T R \nabla \right) x = \frac{K^T \rho}{\sigma_t^2}$$

where  $R$  is a diagonal matrix related to the derivative of  $x$ :

$$diag(R') = \frac{1}{1 + (\gamma \nabla x)^2}$$

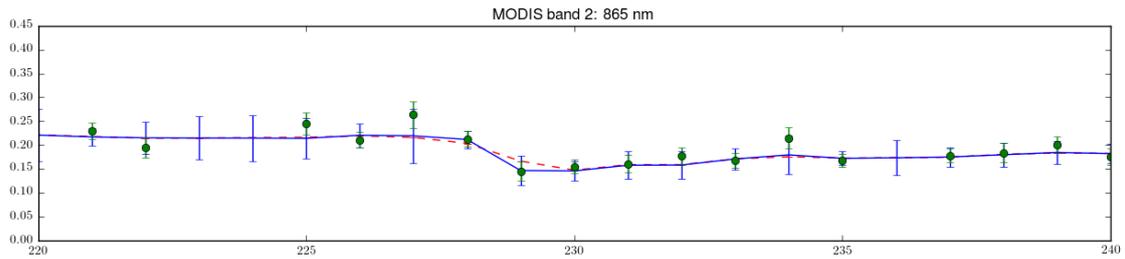
$$R = \eta R'$$

where  $\eta$  is a normalisation term  $\eta = n / Tr(R')$  with  $Tr(R')$  the trace of  $R'$  and  $n$  the number of time samples.  $R$  is approximately 1 when the derivative is low and approaches 0 as  $(\gamma \nabla x)^2$  increases. This effectively weakens the regularisation effect in regions of change. The impact of this is shown in Figure 5.



**Figure 5: Near infrared reflectance normalised with edge-preserving regularisation ( $\gamma=1e3$ ) (blue) with +/- 1 std. Original samples and uncertainty shown in green.**

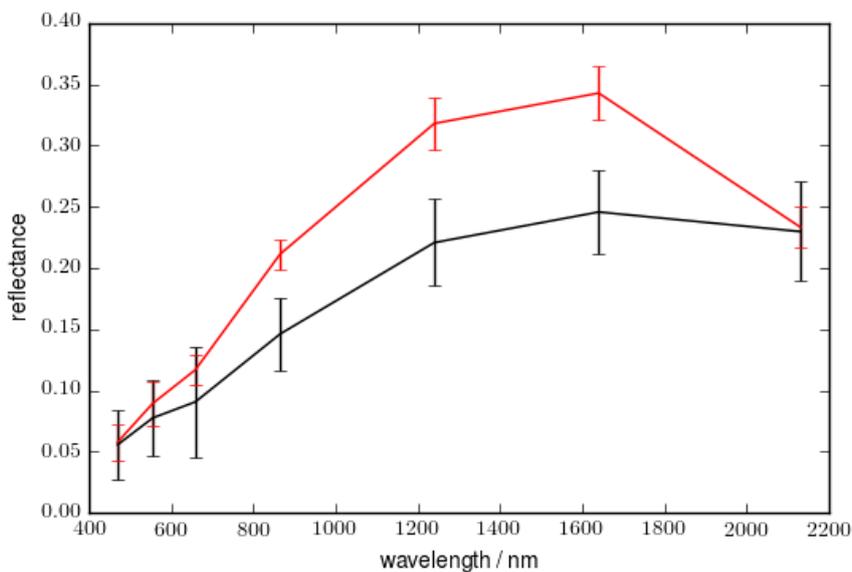
The edge-preserving regularisation also has the effect of slightly increasing the uncertainties in regions of apparent change, but the main impact is to ‘sharpen’ edges in the dataset, which is appropriate for signal change detection algorithms.



**Figure 6: Zoom of near infrared reflectance normalised with edge-preserving regularisation ( $\gamma=1e3$ ) (blue) with  $\pm 1$  std. Original samples and uncertainty shown in green. Non-edge-preserving result shown by red dashed line.**

The use of linear models up to this point has made the error propagation easy to achieve: at this point, we have described the uncertainty in the ‘composite’ (gap-filled) time series by propagating uncertainty from the original BDRF samples. The resultant uncertainty is responsive to variations in BDRF uncertainty, angular sampling and data gaps, as we would expect.

A burn is apparent in the data at day 228. The intention of this illustration is not to provide a competing BA algorithm to those developed in the project, but instead to illustrate error propagation so we do not concern ourselves with the detection here. Instead, we now examine the burn signal change from day 227 to 228.



**Figure 7: Spectral reflectance (with uncertainty) pre- (red) and post-fire (black).**

Figure 7 illustrates the spectral change in (angular normalised) reflectance before and after the fire, along with associated uncertainties. It is likely that the spectral uncertainties are correlated, but we generally have no information on this and so are forced to ignore it.



## 2.2. Example of uncertainty estimation through error propagation: Burn indices

Let us consider now some manifestations of the burn signal  $s$ , and show the error propagation through these. For this example, we shall use NDVI, though this is illustrative of all such ratio-difference transformations, and the results are directly applicable to other indices such as NBR, and the general approach can be generalised to any non-linear band transformation:

$$NDVI = \frac{NIR - R}{NIR + R}$$

where NIR is the near infrared reflectance and R the red reflectance. Let us first consider the error propagation of  $NIR + R$ : If we assume the errors uncorrelated, the variance of this sum is simply equal to the sum of the variances. The same applies to the difference  $NIR - R$ . It is not possible to directly apply linear propagation through the ratio, so instead it is common to approximate this by linear terms using a Taylor expansion:

$$NDVI \approx NDVI_0 + \frac{\partial NDVI}{\partial a} a + \frac{\partial NDVI}{\partial b} b$$

with  $a = NIR - R$  and  $b = NIR + R$ . Then:

$$\sigma_{NDVI}^2 \approx \left| \frac{\partial NDVI}{\partial a} \right|^2 \sigma_a^2 + \left| \frac{\partial NDVI}{\partial b} \right|^2 \sigma_b^2$$

we have seen that  $\sigma_a^2 = \sigma_b^2 = \sigma_{NIR}^2 + \sigma_R^2$  so:

$$\sigma_{NDVI}^2 \approx (\sigma_{NIR}^2 + \sigma_R^2) \left\{ \left| \frac{\partial NDVI}{\partial a} \right|^2 + \left| \frac{\partial NDVI}{\partial b} \right|^2 \right\}$$

$$\frac{\partial NDVI}{\partial a} = \frac{1}{NIR + R}$$

$$\frac{\partial NDVI}{\partial b} = \frac{-NDVI}{NIR + R}$$

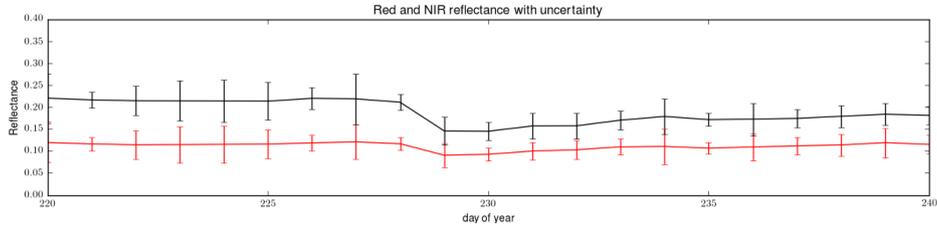
so

$$\sigma_{NDVI}^2 \approx (\sigma_{NIR}^2 + \sigma_R^2) f$$

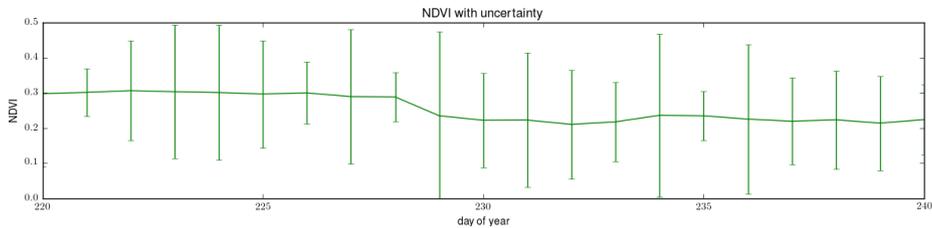
with

$$f = \frac{(1 + NDVI)^2}{(NIR + R)^2}$$

We can see that in this case, we have a mean signal-related component ( $f$ ) and the signal uncertainty component ( $\sigma_{NIR}^2 + \sigma_R^2$ ). For the cases here,  $\sqrt{f}$  is typically around 2 to 3, giving NDVI uncertainty (standard deviation) as between 2 to 3 times the combined reflectance uncertainty. In other words, the index behaves as an “uncertainty amplifier”.



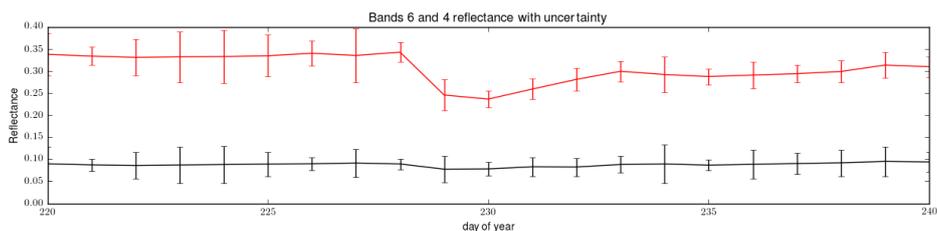
**Figure 8: Reflectance (with uncertainty) in R (red) and NIR (black).**



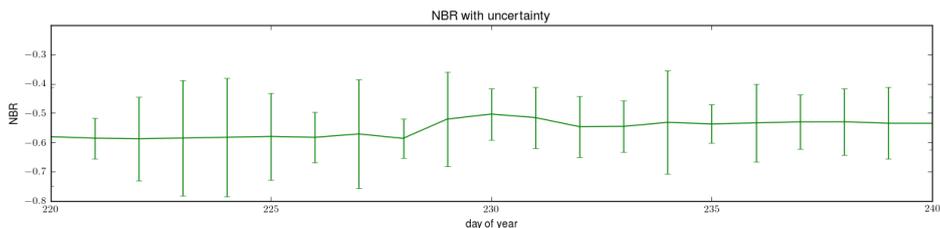
**Figure 9: NDVI (with uncertainty).**

If we examine the red and near infrared signals, we can see a relatively small uncertainty prior to the fire and slightly larger uncertainties after the fire. Because of this and the fact that the difference in the signal is lower after the fire, we see the NDVI go from low uncertainty prior to the fire, to high uncertainty after.

The same theory applies to e.g. the normalised burn ratio (NBR), using MODIS bands 4 and 6 rather than 2 and 1.



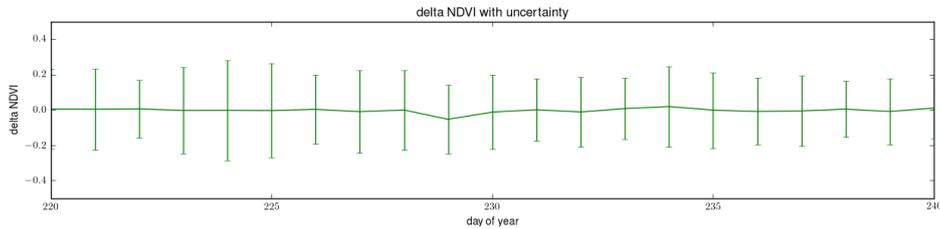
**Figure 10: Reflectance (with uncertainty) in band 6 (red) and band 4 (black).**



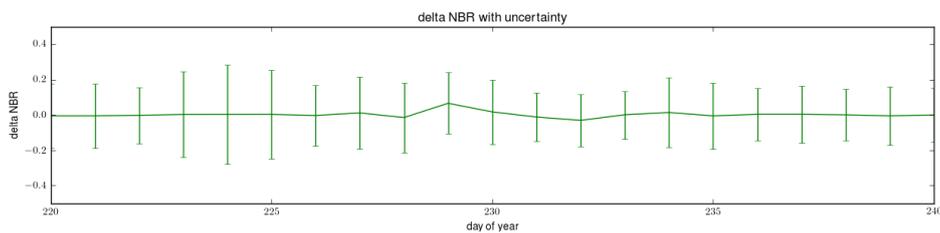
**Figure 11: NBR (with uncertainty).**

The post-fire uncertainty here is relatively smaller than that for NDVI, mainly because of the greater separation of the spectral signals.

An example burn signal, then, might be the change in NBR or change in NDVI. The uncertainty in this is simply the sum of the pre- and post-fire variances. We can see that the mean signal-related component ( $f$ ) here plays an important role in inflating the uncertainty in the change signal.



**Figure 12: Delta NDVI and its uncertainty**



**Figure 13: Delta NBR and its uncertainty**

## 3. Methods

### 3.1. Rationale

Once we provide estimates of burned area as distributions (at the pixel scale or aggregated) we need to confirm that these are reasonable estimates of the true distribution. The problem then is to ‘know’ the true distributions. In the Fire\_cci project, we have developed a testing framework as part of the round robin/uncertainty characterisation. This gives us measured or derived synthetic measurements of reflectance to use in testing BA characterisation, with a (mainly) objective concept of truth derived from higher spatial resolution datasets.

Our approach to testing the BA algorithm error characterisation is to use a set of Monte Carlo simulations to provide (synthetic) sets of input data instances over some given conditions (e.g. noise in reflectance). The sets of data, each of which have  $N$  instances, covering  $n_x$  by  $n_y$  spatial samples and  $n_t$  temporal samples, are provided to the algorithm developers as a spatio-temporal gridded sequence. The developers return the  $N$  outputs (over the  $n_x$  by  $n_y$  spatial samples), along with associated uncertainty information. The average of the  $N$  outputs then characterises the algorithm’s interpretation of mean burned area and uncertainty. We assume the uncertainty associated with these means to decrease as  $1/\sqrt{N}$ , so will need to determine an appropriate value of  $N$  after some initial experiments. A measure of the fidelity of the algorithm outputs is calculated by comparing the individual sample outputs to each other and the known truth.

This method provides a direct comparison of the algorithm-derived metrics of uncertainty to an inferred estimate of the true uncertainty. However, this exercise also presents an interesting opportunity to test the robustness of the algorithms. In effect, each input realisation from the set will provide an independent assessment of the BA,

	<b>Fire_cci</b> <b>Comprehensive Error Characterization</b> <b>Report</b>	Ref.:	Fire_cci_D1.3_CECR_v2.0		
		Issue	2.0	Date	18/05/2018
		Page	19		

and will provide relevant information on the effect of the different studied characteristics on the algorithm.

While the impact of the per-pixel uncertainty is important, it is also important to propagate this uncertainty into the coarser climate modeller's grid (see Section 4). Having the original product estimate as well as the inferred "true" uncertainty will shed some important information on how large the impact of the algorithm-provided uncertainty estimate is when aggregating to coarse resolution grids.

### 3.2. Description of reference sites

Three sites were selected for the exercise. The sites include a savannah site in Africa (MODIS tile h20v09), a boreal site in Siberia (MODIS tile h24v03) and a tropical rainforest site in South America (MODIS tile h11v10). All are shown in Figure 14.



**Figure 14: Location of the selected test sites.**

### 3.3. Generation of surface reflectance realisations

In order to proceed further with the validation of the reported uncertainty, it is necessary to produce a synthetic dataset that will be used to run the different algorithms and provide an estimate of the uncertainty. This necessitates the inferences of the land surface BRDF, which can be then used to forward model the actual observations from different sensors, including their imperfections (e.g. additive and/or multiplicative error arising from incomplete atmospheric correction, gaps due to clouds/orbital choices or changes in view/illumination geometries). Different realisations of these imperfections can be made to result in a realistic sampling of the algorithm inputs.

#### 3.3.1. Estimation of true surface reflectance

To generate realisations of the surface reflectance we first performed a BRDF correction scheme to provide an initial best estimate of the true state of the land surface. Observations of surface reflectance show considerable variability due to changes in the illumination and acquisition geometries. This bi-directional scattering anisotropy is an inherent property of the landscape and is described by the (BRDF), which is defined as:



$$BRDF_{\lambda} = \frac{dL_r(\theta_i, \phi_i, \theta_r, \phi_r)}{dE_i(\theta_i, \phi_i)}$$

where  $dL_r$  is the radiance of surface per unit steradian [ $\text{sr}^{-1}$ ] and  $dE_i$  the irradiance per unit steradian at the surface.  $\theta_i, \phi_i, \theta_r, \phi_r$  denote (respectively) the solar zenith, solar azimuth, view zenith and view azimuth angles. It is not possible to directly measure the BRDF. Instead remote sensing instruments collect measurements of the BRF (the bidirectional reflectance factor) of the land surface. Significant sampling of the BRDF is provided by wide swath polar orbiting instruments frequently used for BA detection. The variation in view/illumination geometry results in a variation in the measured BRF that is significantly higher than the uncertainty in the measurements. This means that BA algorithms can confuse these variations with changes in the land surface unless the BRDF effect is compensated for. A widely used approach to this correction is to express the shape of the BRDF as a sum of kernel functions that describe the bi-directional signature of the land surface. These kernels can be thought of as archetypes of BRDF behaviour. Under this approach, the measured BRF at a given time is written as the weighted sum of  $n$  kernels,  $K_i(\Omega, \Omega')$ , which are functions of the illumination and view geometries ( $\Omega$  and  $\Omega'$ , respectively)

$$\rho(\Omega, \Omega') = \sum_n a_i K_i(\Omega, \Omega')$$

where  $K_i$  are the kernel functions. The Ross-Thick-Li-Sparse family of kernel models has been found to be adequate for most moderate resolution data (Wanner et al. 1995, 1997). These kernels contain an isotropic component, a volumetric scattering component and a geometric scattering component. The volumetric component (the Ross Thick kernel) is an approximation to the BRDF behaviour of a homogenous layer with a random population of particles, and the geometric component encodes the shadowing of a background by geometric primitives. From pairings of observations of reflectance and information on the view/illumination geometries, the weights of the kernel functions ( $a_{iso}, a_{vol}$  and  $a_{geo}$ ) can be inferred by least squares. Although in theory, only three observations would be needed to invert the kernel weights, it is advisable to use at least seven or nine observations to provide a reasonable sampling of BRDF space, as well as counteract the inevitable uncertainty in the measurements. The least squares solution can be viewed as a minimization of a cost function  $J$  which is proportional to the squared difference between the predicted reflectance as given by the linear kernel model and the magnitude of the measured reflectance.

$$\text{minimise } J(f) \propto ||Kf - \rho||^2$$

We implemented a custom BRDF inversion algorithm which is better able to preserve the ‘edge signal’ of the fire process. We explored the use of Tikhonov regularisation (Quaife and Lewis, 2010; Gomez-Dans *et al*, 2013) which allowed for the solution to vary smoothly in time. We present here an adaption of this method based on the discrete cosine transform (DCT). The DCT is a real transform in which a vector of length  $N$  can be written as a linear combination of  $N$  orthogonal basis vectors  $v_k$  ( $k = 1, \dots, N$ ) defined as

$$v_k = s(k) \cos(j + 0.5)k \frac{\pi}{N}$$

Where  $j$  goes from 1 to  $N$ , and  $s(k)$  is a normalising constant given by



$$s(k = 1) = \frac{1}{\sqrt{2}}, \quad s(k > 1) = 1$$

Vectors  $v_i$  form an orthonormal basis. The utility of defining the DCT as orthonormal is that we can represent a K-dimensional real signal  $x$  as a linear combination of these basis vectors:

$$x = a_1 v_1 + a_2 v_2 \dots + a_k v_k$$

An important property of the DCT is that the basic functions used in the transform are the eigenvectors of Toeplitz-plus-Hankel matrices. The importance of this observation arises because the second difference matrix  $\mathbf{L}$  with symmetric boundary conditions is one these matrices. This has a number of computational advantages as computationally costly matrix inversions can be performed as multiplications in the transformed domain, but more importantly it suggests that one may be able to approximate the model above with a much smaller parametric model of just a handful of basic functions.

For the edge-preservation of changes related to fire we view that the solution can be decomposed into two parts: one smooth and low order and another which represents the rapid change in the state associated with a fire occurrence. We can phrase the model as the linear combination of these two components:

$$A = [D \ S]$$

Where the matrix  $A$  is composed of two block matrices.  $D$  is a matrix of low order DCT components which model the smooth evolution of the kernel weights:

$$D = [DCT_{iso} \ DCT_{geo} \ DCT_{vol}]$$

Quaife & Lewis (2010) indicate that less temporal variation should be expected in the geometric and volumetric kernel weights relative to the isotropic one. It is therefore proposed that the volumetric and geometric kernels can be modelled with only one DCT component. As more variation is expected in the isotropic kernel this needs to be modelled with  $k$  components. The block matrix  $S$  represents the edge process associated with a fire. Therefore each column of  $S$  represents the step model associated with a fire on day  $t$ :

$$S = [S_t(t, t_f = 1) \dots S_t(t, t_f = T)]$$

Where  $S_t$  is a discontinuous function defined as:

$$S_t(t, t_f) = \begin{cases} 1 & \text{if } t < t_f \\ -1 & \text{if } t \geq t_f \end{cases}$$

Where  $t_f$  is the day that the fire occurred. We then wish to solve the system:

$$Ax = \rho(\Omega, \Omega')$$

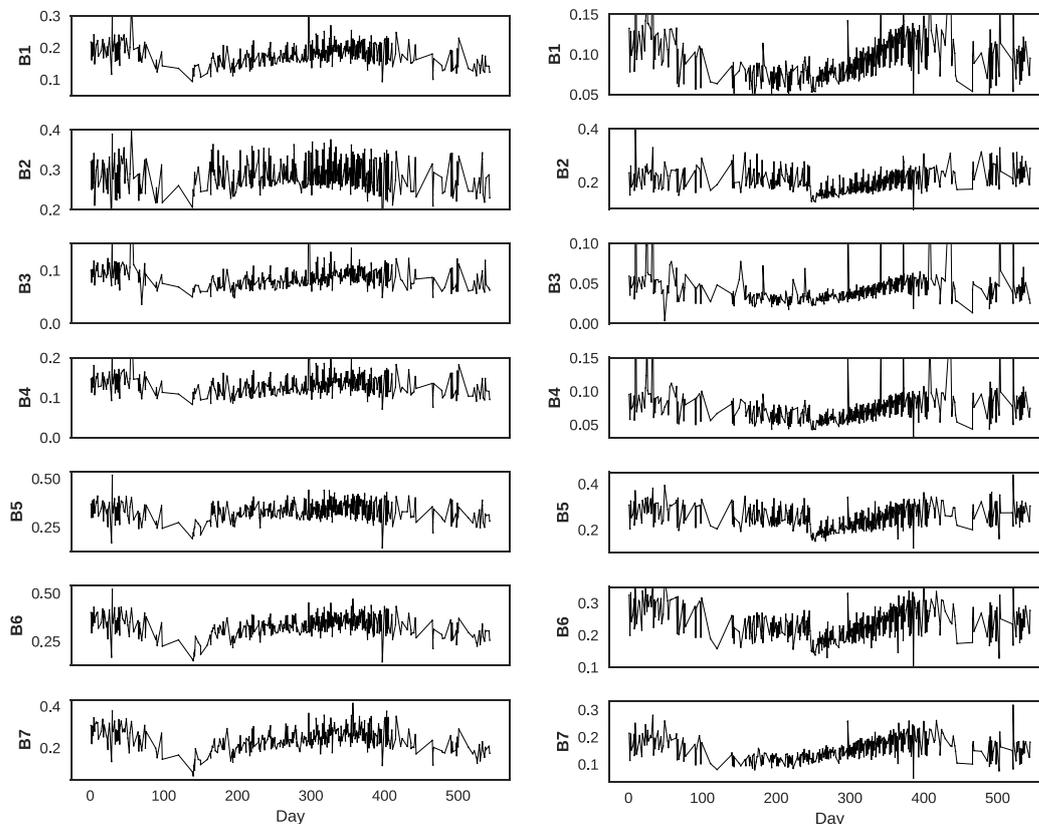
The solution  $x$  then corresponds to two components, the smooth evolution of the kernel weights and the magnitude of the step change related to the fire.

$$x = [x_D, x_S]$$

At present, the model is over-parametrised, with the number parameters in excess of the number of equations. But the interesting feature of  $x$  is that it should be very sparse. This is because we only expect sudden changes in the solution to be caused by a fire. Although we know the spatial extent of burning for each site we need to estimate the

DoB for burned pixels. To avoid implementing an additional burned area algorithm, we have used expert delineated burn scars derived from Landsat pairs to prescribe the fire. This raises the problem of timing: using the Landsat pairs, one can only say with any certainty that the fire occurred between the acquisitions. In order to more clearly define the DoB, a Heaviside step function is swept between the first and second Landsat acquisitions. The day with the best goodness of fit metric is chosen as the DoB.

We now show a simple demonstration of the method over two pixels in the Boreal reference site. Figure 15 shows the typical evolution of surface directional reflectance over a) a fire occurring on DoB 247 and b) a nearby pixel which does not experience a fire. The primary thing to note is the considerable noise in the data arising from the variations in view and illumination angles. The day to day variation in reflectance caused due to this effect is of a similar size to the separation between the pre and post fire reflectance in the case of site a). The other thing to note about site a) is that the spectral impact of the fire is not equal across all bands, with obvious changes occurring in bands 2, 4, 5 and 6, but only a small relative change occurring in bands 1, 3 and band 7. We only need to solve for the parameters of the model that determine the evolution of the surface reflectance. Once we have solved for the parameters for each model we can predict what the reflectance would be for a nadir viewing and solar by using the isotropic reflectance (see Figure 16). Primarily, we see results we would expect, the evolution of the surface reflectance is smooth and gradual apart from at day 247 in site A where there is a large change in reflectance associated with the fire, followed by a gradual recovery.



**Figure 15: Cloud cleared surface reflectance from the MOD09GA and MYD09GA daily products for the two sites. Left: no fire, right: fire happening on day 247.**

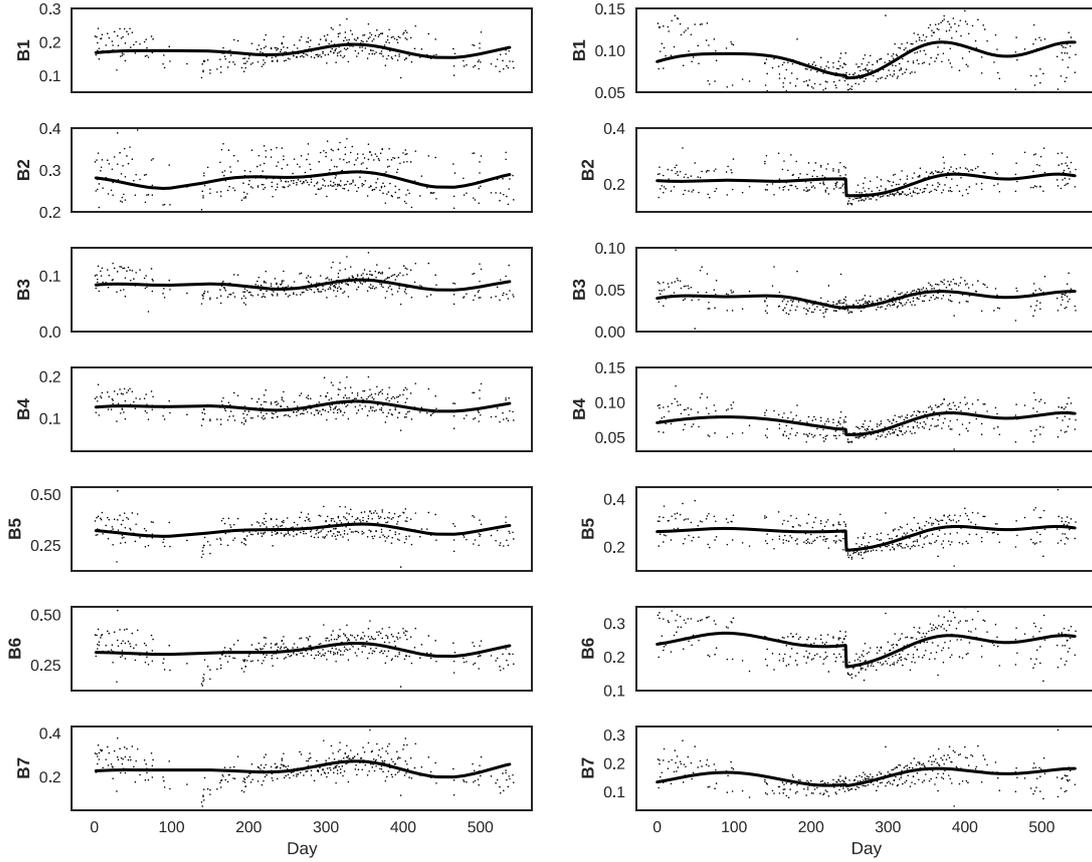


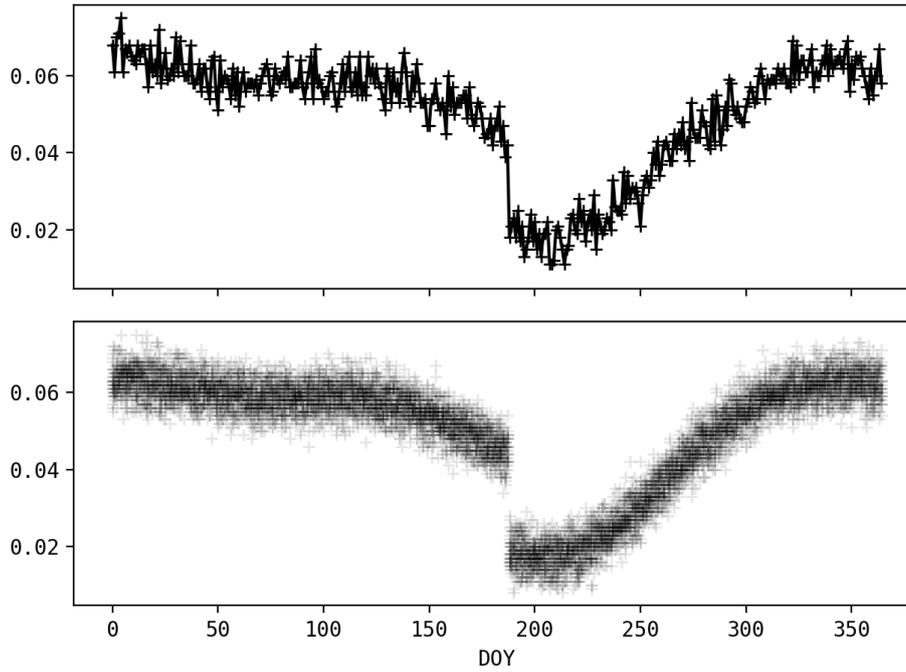
Figure 16: Isotropic reflectance for the two sites. Left) no fire, Right) fire

### 3.3.2. Simulation of observation uncertainty

The uncertainty associated with each observation is modelled with a normal distribution. This means that the recorded observation is realised as:

$$\rho_{observed} = \rho_{true} + N(0, \sigma_{obs})$$

Where  $N$  is a normal distribution. The uncertainty in the observation is encoded in the standard deviation of the observation  $\sigma_{obs}$ . To approximate the distribution of  $P(\rho_{observed})$  we draw  $N$  realisations from the distribution of  $\rho_{true} + N(0, \sigma_{obs})$ . Repeating this process for each observed observation leads to an approximation of the distribution for all of observations in the time series; which we can be seen in the figure below. In the exercise we considered three noise schemes. For MODIS,  $\sigma_{obs}$  varies with wavelength and so we characterise  $\sigma_{obs}^{\lambda}$  from the theoretical error estimates provided from the MODIS atmospheric correction.



**Figure 17: (Top) One realisation of the surface reflectance over a pixel experiencing a fire. (Bottom) 20 realisations produce a sampling from the probability density function of the observations.**

To consider different uncertainties we applied a scaling factor  $c$  so that the per-channel standard deviation was scaled as  $c \sigma_{obs}^\lambda$ . We considered three scaling factors, 0.5, 1.0, and 1.5 to provide a wider range of uncertainties.

### 3.3.3. Simulation of observation opportunity

To simulate a realistic sampling of the observations we use a simple first order Markov chain model. Let the state of whether the surface is observed be defined as  $S$ . We would expect that cloudy observations should be clustered. We define the first order Markov chain therefore as:

$$P(S_t = s | S_{t-1})$$

For each realisation of the surface reflectance dataset we draw from this Markov chain determined by the transition probabilities of  $P_c$  and  $P_s$  which are the cloud-cloud and cloud-clear transition probabilities.

### 3.3.4. Estimation of true algorithm uncertainty $P_b^T$

We can derive estimate of the true algorithm uncertainty  $P_b^T$  given the data based on a Bernoulli approximation. The output burnt area for the  $N$  runs represents a sampling of conditional probability density of the algorithm given the distribution of uncertainty in the observations.

At the pixel level we therefore have  $N$  samples of the distribution:

$$B = [1, 0, 1 \dots N]$$

where a 1 corresponds to the algorithm reporting the pixel burned.

	<b>Fire_cci</b> <b>Comprehensive Error Characterization</b> <b>Report</b>	Ref.:	Fire_cci_D1.3_CECR_v2.0		
		Issue	2.0	Date	18/05/2018
		Page	25		

These outcomes represent draws from a Bernoulli distribution with a probability characterised by  $P_b^T$ . The maximum likelihood estimate of  $P_b^T$  given  $B$  is simply:

$$\widehat{P}_b^T = \frac{n_b}{N}$$

where  $n_b$  is the number of burnt realisations and  $N$  the total number of realisations. Given the limited number of sample runs, this estimate may however be poor (Megill & Mladen, 2011). Instead if we assume that the confidence  $P_b^T$  ranges between 0 and 1, for  $n_b$  burnt draws out of  $N$  draws a better estimate of  $P_b^T$  is:

$$P_b^T = \frac{B(n_b + 2, N - n_b + 2)}{B(n_b + 1, N - n_b + 1)}$$

where  $B$  is the Beta function:

$$B = \int_0^1 p^{b-1}(1-p)^{n-1} dp$$

The above formulation provides a way of estimating the true algorithm uncertainty from a limited set of full algorithm runs. It provides thus a benchmark with which to compare the uncertainty estimates provided by the algorithm, and thus assess their correctness. It has to be noted however that this is based on a realistic sampling of the input data used to produce the ensemble of outputs required for estimating  $P_b^T$ . This means that simulations of input uncertainty (which in this case could include the effect of residual atmospheric correction, as well as thermal noise inherent to the nature of the sensors, and other effects such as cirrus or cloud or cloud shadow contamination), and observation opportunity (broadly speaking, cloudiness) should have realistic statistical descriptions. This is in itself a complex task as this is location, sensor and pre-processing chain dependent. In this work we aim to provide a simple generic testing based on some general assumptions described in the paragraphs above

#### 4. Grid scale uncertainty aggregation

The pixel level product provides a binary indicator of the presence of a fire within a certain time period (a month). This information is extended with the most likely date of the fire taken place if a fire has been detected. This information is available at high spatial resolution (10s-100s of m grid size), and is too detailed for climate users, where spatial resolution is usually of the order of 100s of km. In order to aggregate the high resolution information to the coarser climate grid, the procedure starts by defining what high resolution pixels fall within a given climate modelling grid (CMG) cell, and then counting the pixels in that set that have burned. This is either reported as an area (multiplying the sum of burned pixels by the individual high resolution pixel area), or as an area fraction.

A major development in the ESA Fire\_cci is the addition of uncertainty information. This means that each pixel is qualified by some uncertainty, a so-called probability of burn,  $p_b$ . This is a number between 0 (absolute certainty that the pixel **did not** burn in the temporal interval considered) to 1 (absolute certainty that the pixel **did** burn in the temporal interval considered). This metric should reflect the degree in which different factors limit the detection of fires (e.g. the observational opportunity, the inevitable limitations of the pre-processing chain, such as residual atmospheric effects not fully

	<b>Fire_cci</b> <b>Comprehensive Error Characterization</b> <b>Report</b>	Ref.:	Fire_cci_D1.3_CECR_v2.0		
		Issue	2.0	Date	18/05/2018
		Page	26		

corrected by atmospheric correction, gridding artefacts, the properties of the fire and its effect in the remote sensing signal, etc.) These developments affect how one does aggregation to a coarser resolution.

#### 4.1. Aggregation basics

From the point of view of the ESA Fire\_cci pixel-level product, we have two layers which are relevant: the date of first detection, and the confidence level. Additionally, some pixels will be labelled as non-burnable (e.g. ocean or lakes, deserts...), or may be deemed unobservable (insufficient number of observations). Those with insufficient observations form an important aspect of the grid cell information, either for use as a quality control measure or for attempting a correction for missing values. We note in the latter context that all pixels considered ‘non burnable’ should be labelled as such, even if they are not observed. Generally speaking, we can determine the burned area inside a CMG cell as the sum of pixels where the first date of detection is between 0 and 366 (both inclusive), multiplied by the area of the pixel. This is intuitive and in line with previous estimates. However, if we interpret the confidence layer as a probability of burn,  $p_b$ , (and in consequence, a probability of not being burned of  $1 - p_b$ ), then this information would need to be scaled up to the CMG, as a form of standard error. There are two common definitions relating to standard error (Weisstein et al., 2017): (i) the square root of the estimated error variance (standard deviation); (ii) the standard error of a sample of sample size  $n$  is the sample standard deviation divided by  $\sqrt{n}$ . We need to consider then which would be appropriate in this context.

In the PUG of the MERIS Fire\_cci v4.1 product (Pettinari et al., 2016), the standard error was defined as:

“The standard error is modelled and predicted with a regression model, calibrated with reference data. The response variable is the absolute observed error and the explicative variable is the burned area extent estimated for the grid cell. The standard error is positively related with the estimated extent of burned area in each grid cell. Even though those reference datasets were chosen to represent different fire regimes, they may be not fully representative of some regional fire conditions. More detail on the statistical models can be found in the ATBD III v2.3 (Tansey et al. 2014).”

However that is clearly not a ‘standard error’ according to either of the definitions, but just some monotonic function of burned area.

The sample variance  $\sigma^2$  of a sample set of size  $n$  is given by:

$$\sigma^2 = \frac{1}{n} \sum_{i=1}^n (y_i - \bar{y})^2,$$

where  $y_i$  is sample  $i$  and  $\bar{y}$  is the sample mean, given by

$$\bar{y} = \frac{1}{n} \sum_{i=1}^n y_i$$

The sample terms  $\bar{y}$  and  $\sigma^2$  are random variables, and the expected value of the variance  $\hat{\sigma}^2$  is given by

$$\hat{\sigma}^2 = \frac{n}{n-1} \sigma^2$$



Often,  $\sigma^2$  is the *biased sample variance* and  $\hat{\sigma}^2$  is the *unbiased sample variance*. Going back to our earlier discussion of definitions of standard error, we can say that the first definition  $\hat{\sigma}_1$  (square root of the estimated error variance) is thus

$$\hat{\sigma}_1 = \frac{n}{n-1} \sigma$$

where  $\sigma$  is the sample standard deviation. Using the second definition (sample standard deviation divided by  $\sqrt{n}$ ) we have that

$$\hat{\sigma}_2 = \frac{1}{\sqrt{n}} \sigma_1$$

The first definition is more consistent with many uses of standard error in the physical sciences, where it takes the role of an unbiased estimate of the standard deviation of a distribution. If the distribution is assumed normal and  $y$  is continuous (or effectively continuous if  $n$  is large), then the estimate of the mean ( $\bar{y}$ ) and standard deviation  $\sigma_1$  fully define the Probability Distribution Function (PDF) for BA.

The second definition is more directly related to the uncertainty of the mean and is used in the definition of probable error. The standard error of the mean is given by  $\sigma^2$ . So, with more samples (greater  $n$ ) we can better estimate the mean of the distribution.

In the light of this, we will use

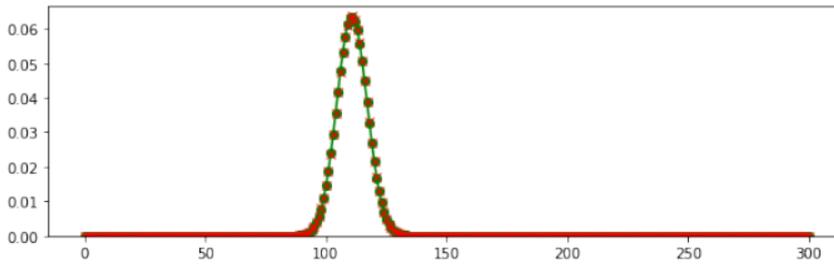
$$\hat{\sigma} = \frac{1}{\sqrt{n-1}} \sum_{i=1}^n (y_i - \bar{y})^2$$

which is a unbiased estimate of the likely variability in burned area.

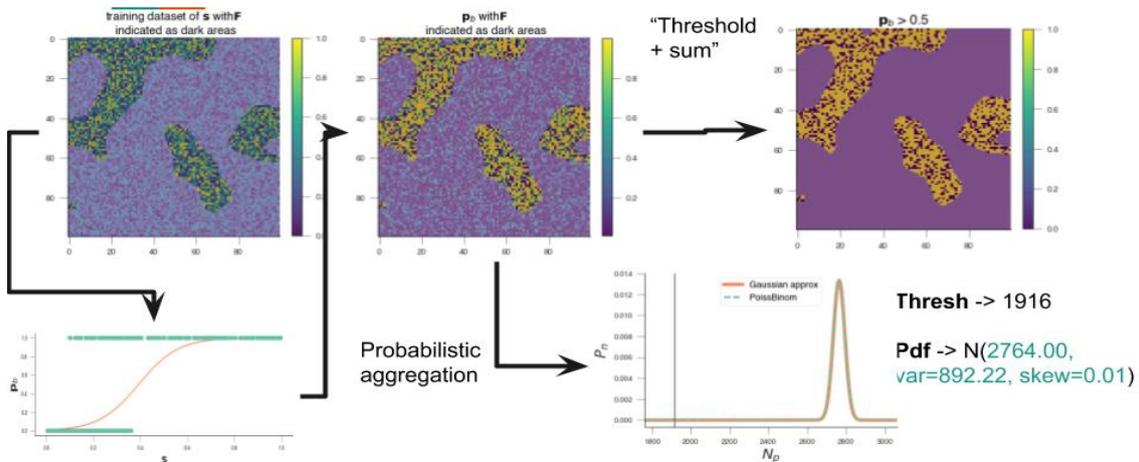
Assuming that each pixel has an independent probability of burn  $p_b$ , which can be different for every pixel, then the sum of these independent probabilities is given by a Poisson Binomial distribution. This distribution is only defined over positive integer numbers, and has first and second order statistics given by

$$\bar{N}_b = \sum_{i=1}^{N_p} p_{b,i}$$
$$\sigma_b^2 = \sum_{i=1}^{N_p} p_{b,i} (1 - p_{b,i})$$

In Figure 18, we show the full PDF derived from a set of samples each characterised by a different  $p_b$ . We calculate the PDF as a Poisson binomial, and also calculate the mean and variance using the equations above, and plot the normal approximation to the PDF. For a large number of samples, the skewness of the PDF is very low, and the PDF is acceptably approximated by a Gaussian distribution. This is of importance, as it means that one can parametrize the full PDF of BA using only the mean and the “standard error” (defined as the standard deviation in the discussion above), and in accordance to the product specification.



**Figure 18:** The Poisson binomial PDF (green line) derived from a simulated set of independent samples (300, 100 with probabilities between 0.7 and 0.9, 100 with probabilities between 0.2-0.3 and 100 with probabilities between 0-0.1). A Gaussian approximation (red line) derived from calculating the mean ( $\sim 110$ ) and standard deviation ( $\sim 39$ ) is also shown. Skewness was  $\sim 0.01$ .



**Figure 19:** Example of aggregation. See text for more details

In Figure 19, we show a typical workflow. We start at the top left panel, which shows the spatial distribution of some metric  $s$  that is somehow related to burned area (e.g. reflectance in some band, or a band combination). The burn scars are characterised by larger values in this space, and it is clear that there is some randomness. On the panel below that one, a mapping from the  $s$  to probability of burn is produced (in this case, we have used a simple logistic regression, but this is general). On the middle top panel we have the spatial distribution of the probability of burn, which shows a speckly but visually clear distribution of burn area. If a threshold is applied to the probability of burn map (so that e.g. any pixel with  $p_b > 0.5$  is assumed burned), we end up with a binary map with a value of 0 for values below 0.5 and 1 for values above. Finally, the aggregates from both approaches can be displayed. The thresholded approach results in a single value, the sum of pixels above the threshold (in this case, 1916). The proposed aggregation using a Poisson binomial results in a PDF (again we show the Poisson Binomial and the Gaussian approximation), with a mean of  $\sim 2760$  and a standard deviation of  $\sim 30$ . The actual number of burned pixels in this case was around 2780.

## 4.2. Unreliable probability of burn estimates

Since the quantification of probability of burn per pixel is still fairly new, and both users and product developers are more used to using the sum of pixels aggregation, it might be useful to consider how to make both approaches compatible. One simple approach might be to re-scale  $p_b$  based on the sum of pixels: if the mean of the Poisson binomial (or Gaussian approximation) is given by the sum of pixels (rather than by the sum of individual  $p_b$ ), then we can scale the individual values of  $p_b$  so that the mean is

identical to the sum of pixels, and then use that rescaled  $p_b$  to calculate the standard deviation and thus provide some form of uncertainty estimate. The effect of this “correction” can be seen in Figure 20.

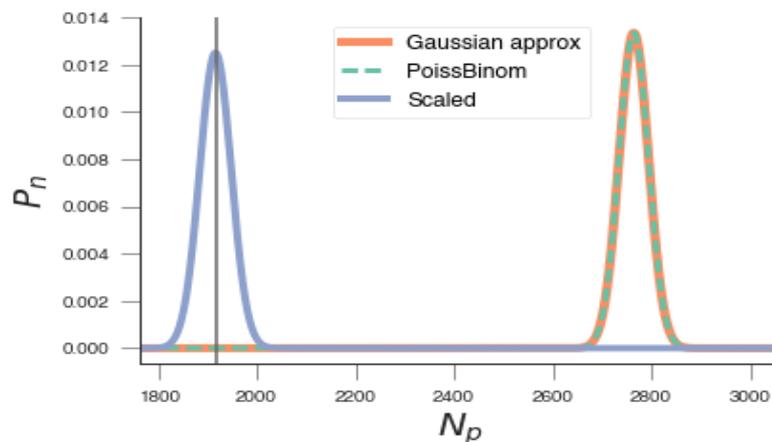


Figure 20: Example of applying the rescale approach to the example shown in Figure 19.

## 5. Uncertainty descriptions for algorithms

### 5.1. Fire\_cci v5.0: Uncertainty calculations

The Fire\_cci v5.0 and v5.1 products use a function driven by a number of observables to estimate the per pixel uncertainty. A full description of these calculations is provided in the relevant product ATBDs, but broadly speaking, a generic mapping (a logistic regression) is proposed in terms of a set of parameters (such as magnitude of the burn signal, distance to a thermal anomaly and other spatial context considerations). This mapping is then fitted to a set of burn/unburned events to elucidate the regression mapping hyperparameters (logistic regression coefficients) which can then be applied to all pixels.

### 5.2. SLSTR uncertainty characterization

For each pixel, the algorithm produces a score  $s$ . This is the standardized output of the MaxEnt algorithm applied to the classification variables, and ranges between 0 and 1: it gives a likelihood of burn at the pixel level. For a spatiotemporal patch to be classified as "burned", (1) its median score has to be larger than a data-derived threshold  $T$ , and (2) the patch has to contain at least one active fire.

The reported value of  $p_b$  was derived from the score in order to guarantee that  $p_b \geq 0.5$  if and only if the pixel is classified as burned. In order to get that condition, a linear transformation of scores over  $[0, T]$  and over  $[T, 1]$  is applied (overall the transformation is piecewise linear and continuous).

- (1) for pixels in patches that are classified as burned (which implies there is at least one active fire for the patch):  $p_b$  is 0.5 if  $s \leq T$ ;  $p_b = \text{linear transformation of } s$  if  $s > T$
- (2) for pixels in patches that contain active fires but the median score is below  $T$  (so they are classified as not burned):  $p_b$  is a linear function of  $s$  if  $s \leq T$ ;  $p_b = 0.5$  if  $s > T$
- (3) for pixels in patches that do not contain active fires (so they are classified as not burned):  $p_b$  takes values between 0 and 0.5 and is a linear function of  $s$

	<b>Fire_cci</b> <b>Comprehensive Error Characterization</b> <b>Report</b>	Ref.:	Fire_cci_D1.3_CECR_v2.0		
		Issue	2.0	Date	18/05/2018
		Page	30		

## 6. Results

### 6.1. Uncertainty assessment for the Pixel Product

The main strategy for assessing uncertainty is based on the dataset introduced in Section 4. A synthetic dataset has been prepared and different complicating factors were simulated as different scenarios. A number (20) of different realisations of a particular scenario have been used to infer the “true” probability of burn, a benchmark to compare the actual product uncertainty estimates against.

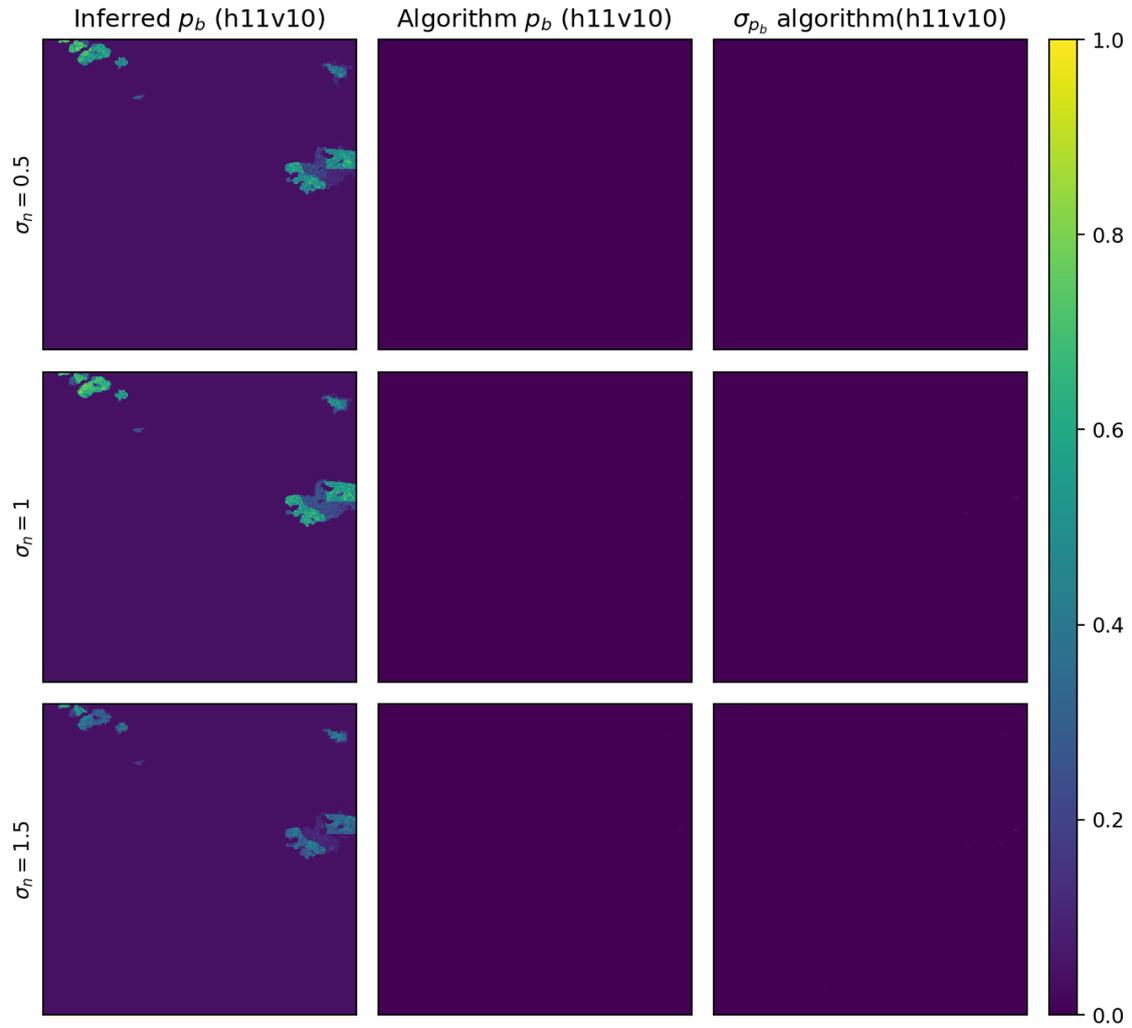
Given that, for each simulation scenario the statistical properties of the observations are identical, one would expect that the algorithms produce an estimate of pb which is largely identical from one realisation to the next for the same scenario. The opposite case would indicate that the uncertainty calculation is over-reliant on the actual observations, rather than on their statistical properties.

#### 6.1.1. The MODIS Fire CCI v5.1

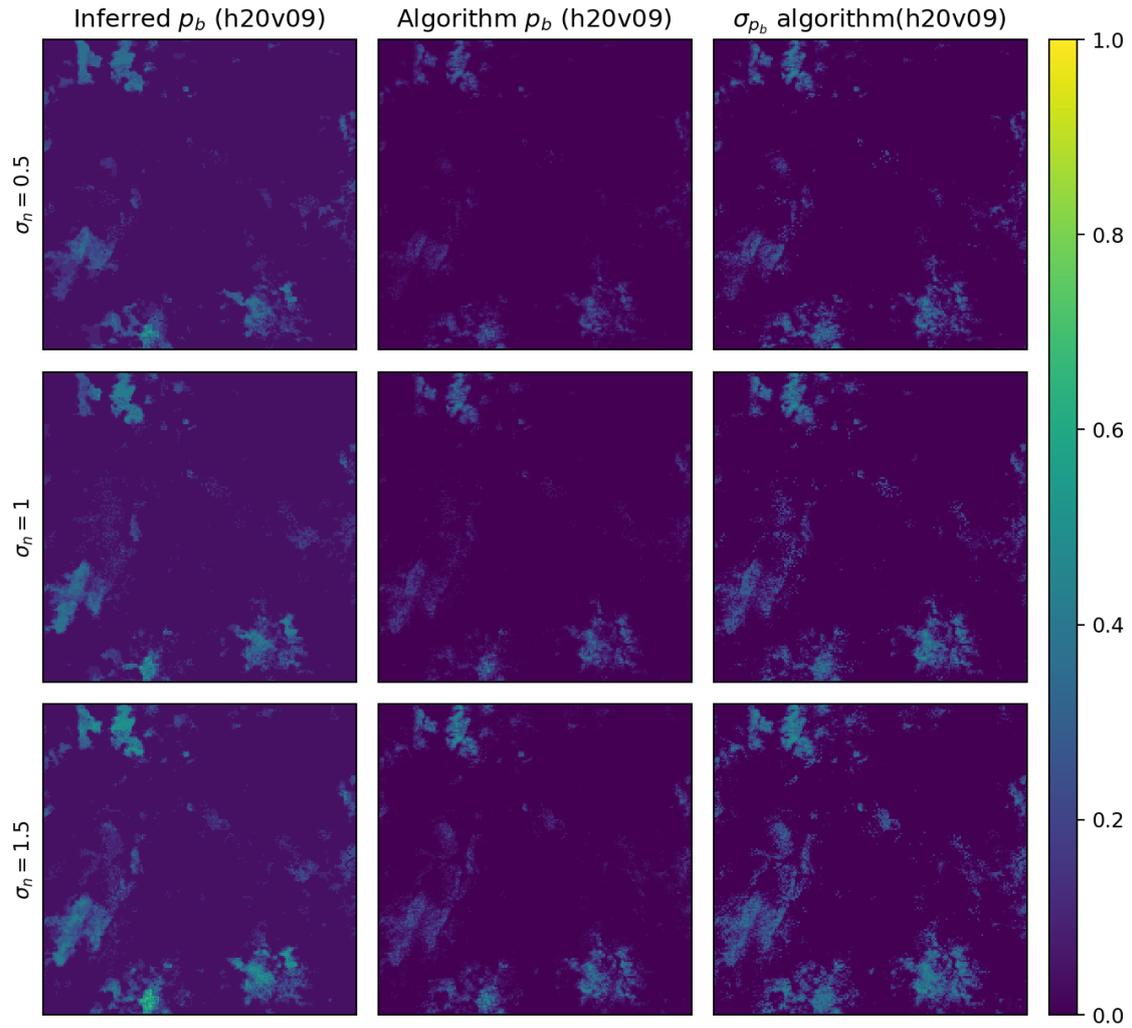
The results for the MODIS Fire CCI v5.1 BA dataset are shown in Figure 21, Figure 22 and Figure 23. The patterns of inferred pb are broadly indicative of higher probability of burn within the burn scars, often higher in the middle of the burn scar than on the edges. Additionally, both for the tropical and boreal sites, the inferred probability of burn decreases with increasing noise, as expected.

The algorithm reported pb is very low for the tropical forest site (values are less than 0.08 for all pixels). For the savannah and boreal sites, probabilities are generally speaking low and high (respectively). However, for these two sites, the probability of burn tends to be higher for the high noise case than for the low noise case, opposite of what the inferred value of pb suggests, and contrary to expectations.

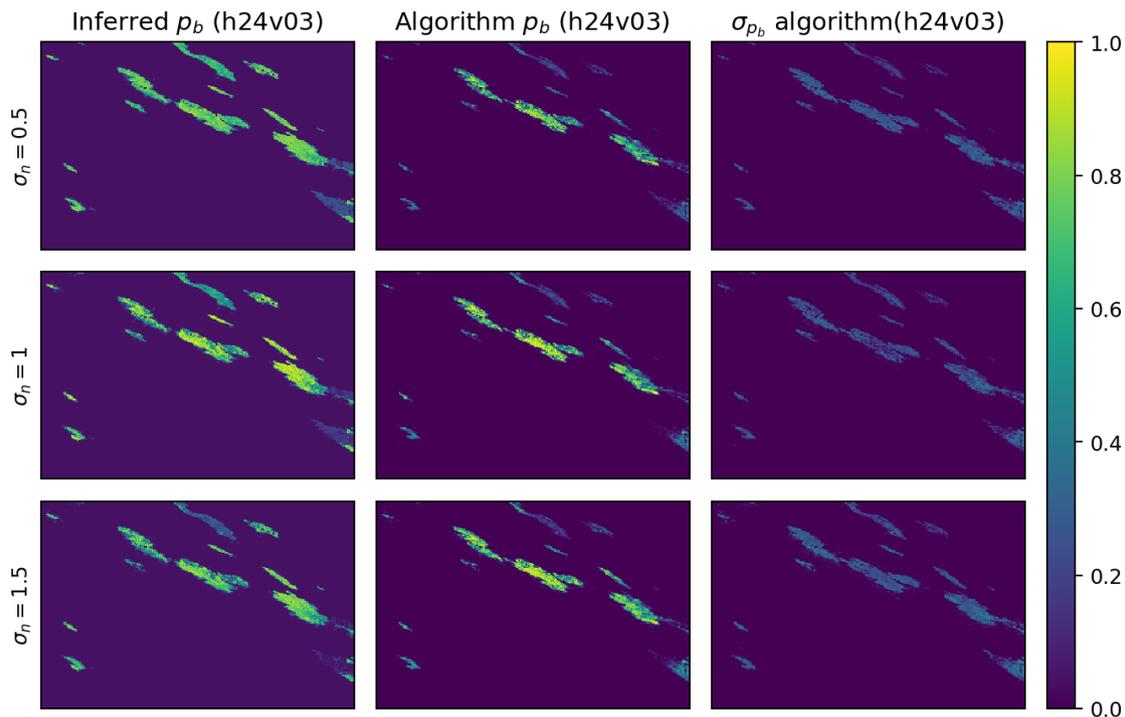
Across different realisations (see Section 3.3 for details), the algorithm shows a large variation of pb over burned areas (standard deviations of 0.3 and 0.4 for the savannah and boreal in probability of burn across the 20 realisations, irrespective of noise level). For unburned pixels, the probability of burn is always extremely low (less than 0.1, often less than 0.05). The large variability across burned pixels is indicative of the actual observations having a strong impact on the calculation of probability of burn, whereas over unburned pixels there is a strong belief by the algorithm that no omission errors are present.



**Figure 21: Fire CCI algorithm v5.1, Tropical site. Inferred probability of burn (left column), mean algorithm reported probability of burn (middle column) and standard deviation of algorithm reported probability of burn (right column). Different rows represent different noise levels: 0.5 nominal (top row), nominal (middle) and 1.5 nominal (bottom). All plots share the same scale.**



**Figure 22: Fire CCI algorithm v5.1, Savannah site. Inferred probability of burn (left column), mean algorithm reported probability of burn (middle column) and standard deviation of algorithm reported probability of burn (right column). Different rows represent different noise levels: 0.5 nominal (top row), nominal (middle) and 1.5 nominal (bottom). All plots share the same scale.**



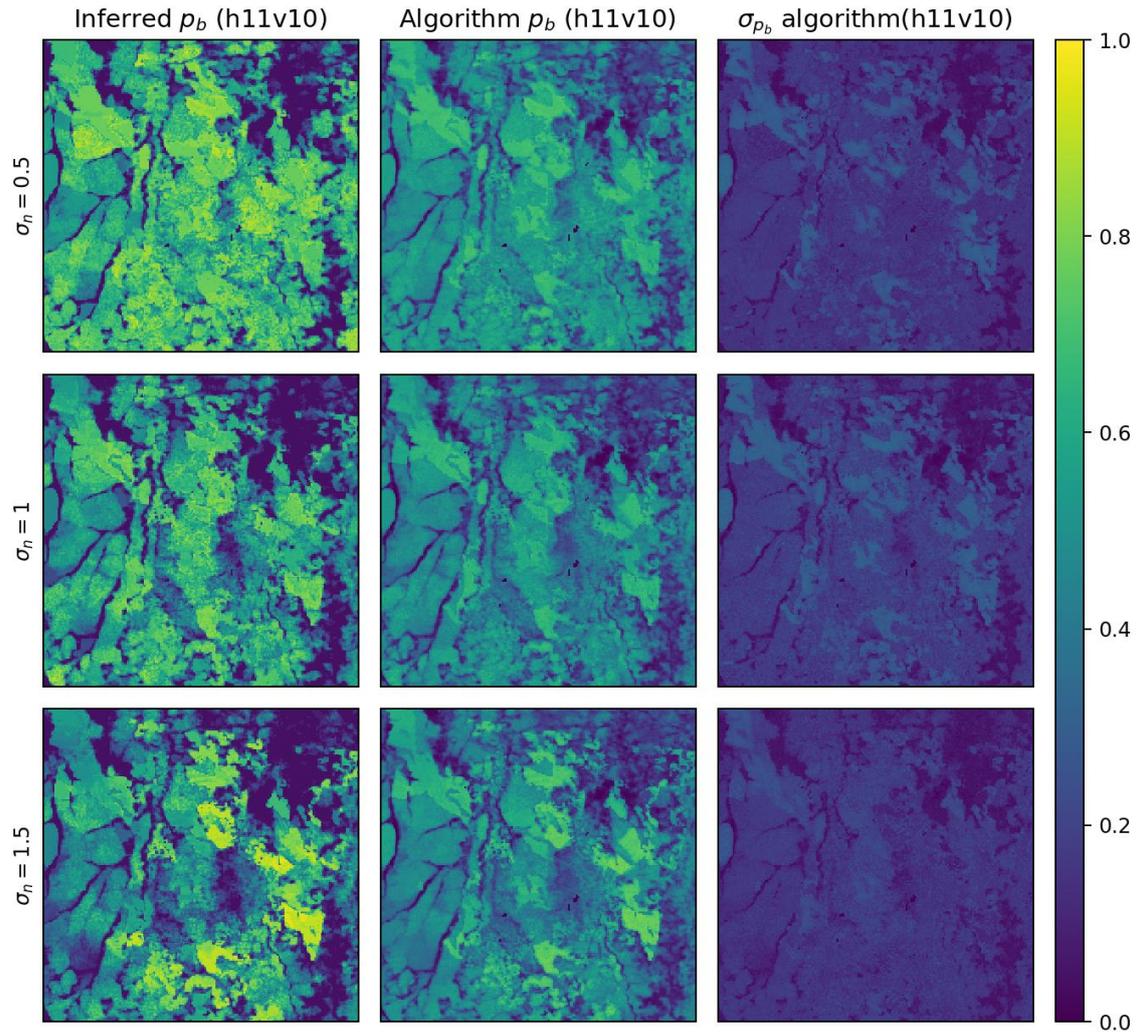
**Figure 23: Fire CCI algorithm v5.1, Boreal site. Inferred probability of burn (left column), mean algorithm reported probability of burn (middle column) and standard deviation of algorithm reported probability of burn (right column). Different rows represent different noise levels: 0.5 nominal (top row), nominal (middle) and 1.5 nominal (bottom). All plots share the same scale.**

### 6.1.2. The SLSTR algorithm

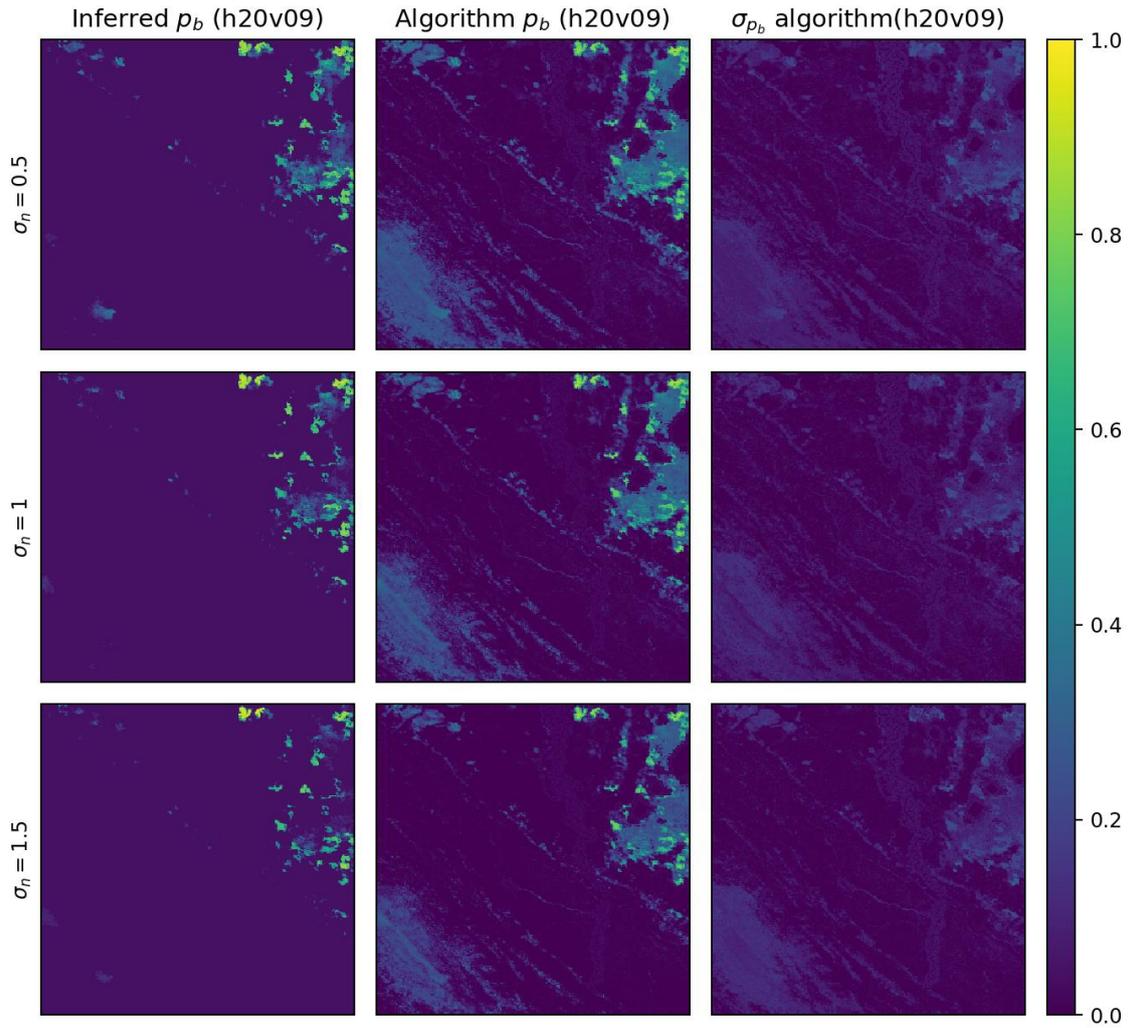
The results from the SLSTR algorithm estimation of probability of burn are shown in Figure 24, Figure 25 and Figure 26. As far as the Monte Carlo-inferred probability of burn is concerned, on the savannah site there is no clear trend with respect to noise level: some areas in the bottom left corner of Figure 25 increase their probability of burn with increasing noise. This could be explained by the algorithm relying on informative observations from LST and/or thermal anomalies, which were not modified in the MC simulations. In the tropical rainforest case, some medium probability of burn patches are lost with increasing noise levels (see Figure 24, centre left), so broadly speaking, the inferred  $p_b$  decreases with noise level, particularly when for low probability of burn fires. A similar effect is seen in the boreal: inferred  $p_b$  does not change inside large burn scars, but it does change or even shrink to values close to zero.

The algorithm-reported  $p_b$  values show very similar spatial patterns to the inferred  $p_b$  values, and algorithm-reported  $p_b$  values appear to be of the same order as inferred  $p_b$ . The algorithm reported  $p_b$  is slightly higher for unburned pixels, which might be due to poor MC sampling in the inferred calculations.

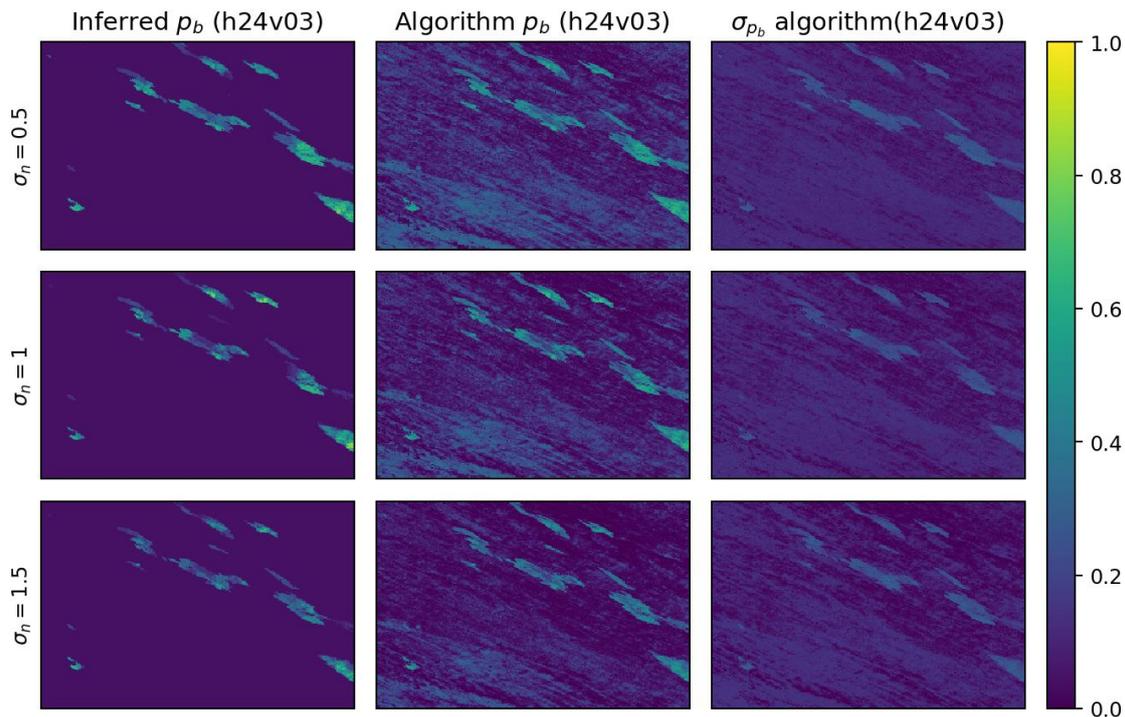
There is variation between the algorithm-reported  $p_b$  across different realisations, but this is generally speaking low (less than 0.2), which is positive. Larger values are reported over burns in the boreal site, but these are not particularly significant.



**Figure 24: SLSTR algorithm, Tropical site. Inferred probability of burn (left column), mean algorithm reported probability of burn (middle column) and standard deviation of algorithm reported probability of burn (right column). Different rows represent different noise levels: 0.5 nominal (top row), nominal (middle) and 1.5 nominal (bottom). All plots share the same scale.**



**Figure 25: SLSTR algorithm, Savannah site. Inferred probability of burn (left column), mean algorithm reported probability of burn (middle column) and standard deviation of algorithm reported probability of burn (right column). Different rows represent different noise levels: 0.5 nominal (top row), nominal (middle) and 1.5 nominal (bottom). All plots share the same scale.**



**Figure 26: SLSTR algorithm, boreal site. Inferred probability of burn (left column), mean algorithm reported probability of burn (middle column) and standard deviation of algorithm reported probability of burn (right column). Different rows represent different noise levels: 0.5 nominal (top row), nominal (middle) and 1.5 nominal (bottom). All plots share the same scale.**

## 6.2. Uncertainty assessment for the Grid Product

### 6.2.1. The MODIS Fire CCI v5.1

The previous datasets have been processed with the scheme introduced in Section 4 to a CMG. This has been done by aggregating pixels in 100x100 pixel windows, broadly a 50km square at the equator, and around the kind of coarse resolution grid cell size. For the different sites and noise levels, we have used the inferred pb (inferred from the MC realisations), and the mean algorithm provided pb across the different MC realisations to provide an estimate of the burned area (in units of pixels).

For each CMG cell there are two PDFs: a “true” PDF based on the inferred pb, and an “algorithm” PDF based on the algorithms pb value. One would expect both PDFs to overlap each other (or at least have a significant overlap), and also one would expect to see a broadening of the PDF as the noise level increases.

The results are shown in Figure 27, Figure 28 and Figure 29. The comparison between the PDFs shows little or no overlap with those calculated from the inferred pb calculations. The PDFs calculated from the inferred pb are distinct, with the scenario having 0.5 times the nominal noise and the scenario with the nominal noise showing important overlaps in many cases, but with the high noise scenario being further apart. Often, the high noise case PDF is broader, correctly suggesting a larger uncertainty. This suggests that there is a modicum of robustness to noise for lower noise levels, but that once the noise exceeds that amount, the changes in the estimated PDF are dramatic.

For the algorithm reported pb calculations, the picture is more confused, with PDFs behaving in largely different ways depending on the window and site. Although here we



show the results calculated with the mean pb from the algorithm, changing to individual realisations results in vast shifts of the PDFs, in line with previous comments of reliance of uncertainty calculations on the observational data.

Tile h11v10

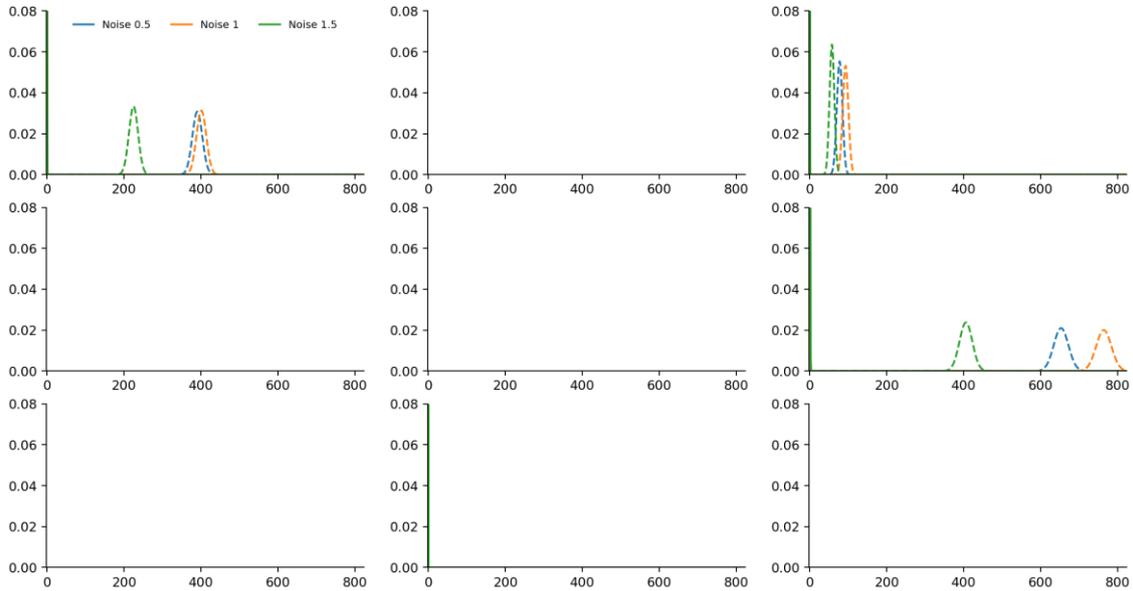


Figure 27: Fire CCI v5.1 Tropical site. Aggregation to CMG results. Each panel shows the aggregated BA PDF of a 100x100 pixel area in the original pixel product (x-axis: burned pixels per area, y-axis: frequency). The dashed lines represent the PDF calculated from the inferred pb value, and the full lines represent the PDF calculated from the mean of the algorithm reported pb across the 20 MC realisations. Different colours represent different noise levels (blue: 0.5 times nominal noise level, orange: nominal noise level and green 1.5 times nominal noise level).

Tile h20v09

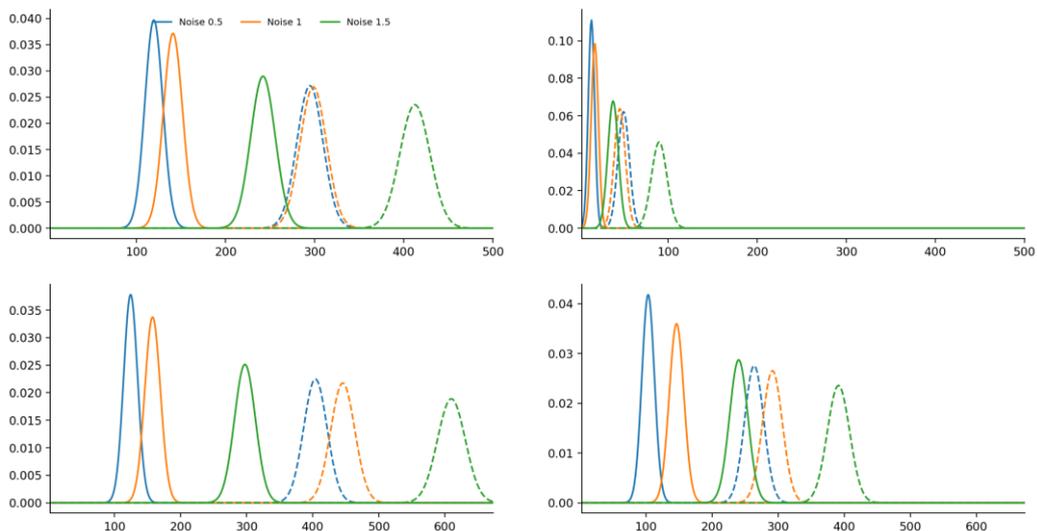
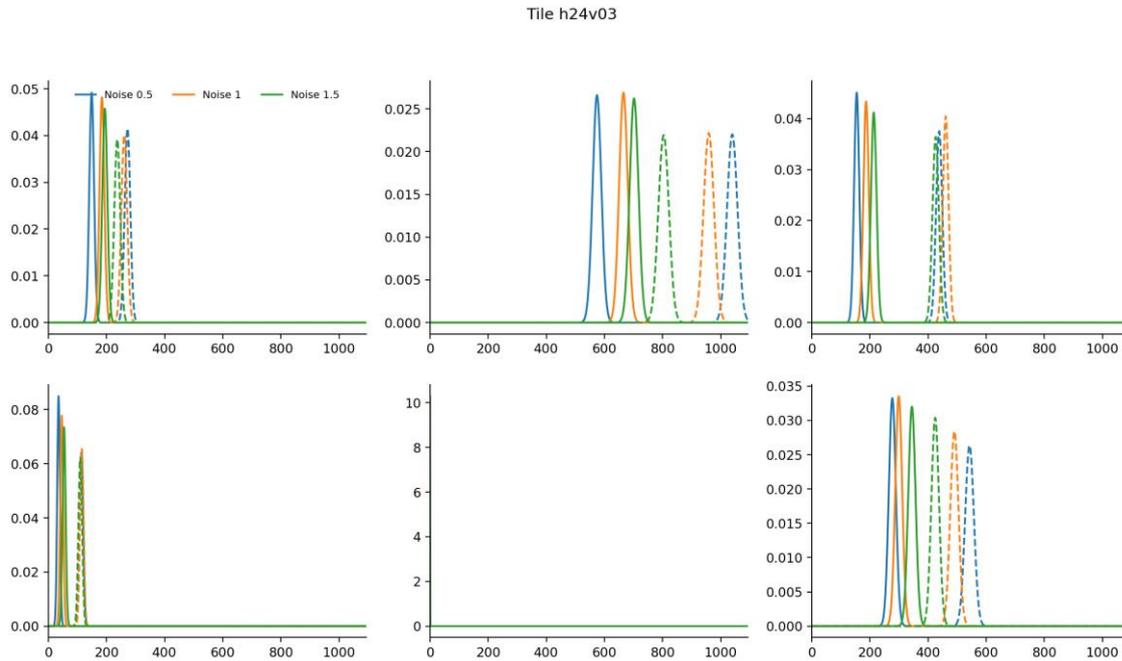


Figure 28: Fire CCI v5.1 Savannah site. Aggregation to CMG results. Each panel shows the aggregated BA PDF of a 100x100 pixel area in the original pixel product (x-axis: burned pixels per area, y-axis: frequency. The savannah site only has four windows). The dashed lines represent the PDF calculated from the inferred pb value, and the full lines represent the PDF calculated from the mean of the algorithm reported pb across the 20 MC realisations. Different colours represent



different noise levels (blue: 0.5 times nominal noise level, orange: nominal noise level and green 1.5 times nominal noise level).



**Figure 29: Fire CCI v5.1 boreal site. Aggregation to CMG results. Each panel shows the aggregated BA PDF of a 100x100 pixel area in the original pixel product (x-axis: burned pixels per area, y-axis: frequency). The dashed lines represent the PDF calculated from the inferred pb value, and the full lines represent the PDF calculated from the mean of the algorithm reported pb across the 20 MC realisations. Different colours represent different noise levels (blue: 0.5 times nominal noise level, orange: nominal noise level and green 1.5 times nominal noise level).**

### 6.2.2. The SLSTR algorithm

The results of aggregating the inferred pb and the algorithm-reported pb values to the CMG cell are shown in Figure 30, Figure 31 and Figure 32. For the tropical site (Figure 30), we can see that for the high noise case there is a large overlap of the benchmark and algorithm-reported PDFs, whereas for the other noise levels, values are generally non-overlapping. In general, PDFs from either the benchmark or reported calculations are well separated, with minimal overlap between noise levels, with a clearer separation between the reported dataset. For the savanna and boreal sites, the increased width of the PDF with increased noise is evident, and although the pattern is more marked for the benchmark case than for the reported one, it is still noticeable in both cases. Generally speaking, these results imply that the relative patterns of reported pb are in line with those in the benchmark dataset, but that their absolute values might still require some fine tuning.



Tile h11v10

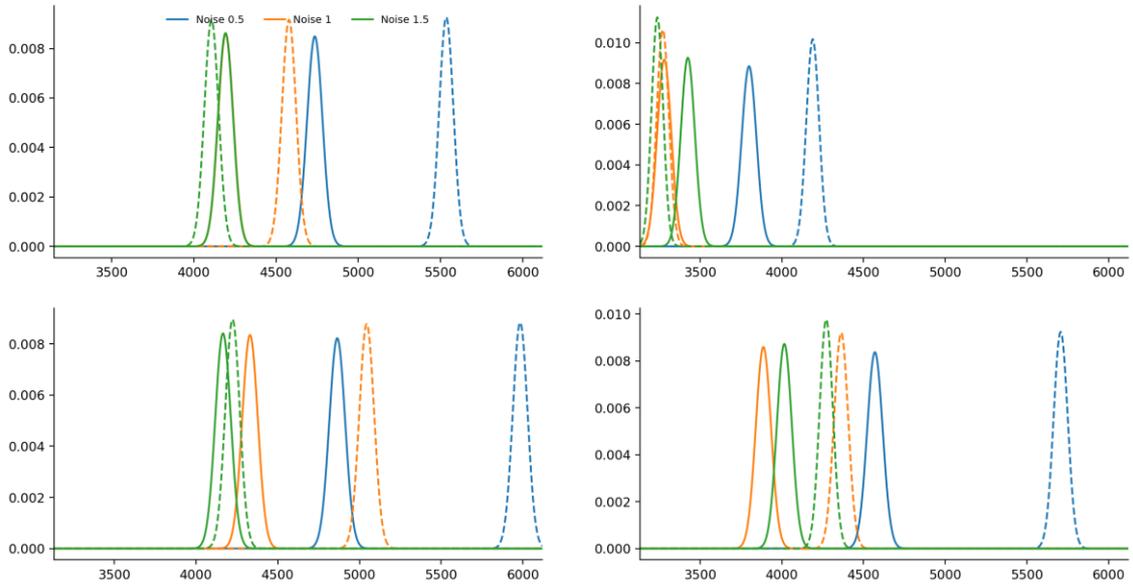


Figure 30: SLSTR Tropical site. Aggregation to CMG results. Each panel shows the aggregated BA PDF of a 100x100 pixel area in the original pixel product (x-axis: burned pixels per area, y-axis: frequency). The dashed lines represent the PDF calculated from the inferred pb value, and the full lines represent the PDF calculated from the mean of the algorithm reported pb across the 20 MC realisations. Different colours represent different noise levels (blue: 0.5 times nominal noise level, orange: nominal noise level and green 1.5 times nominal noise level).

Tile h20v09

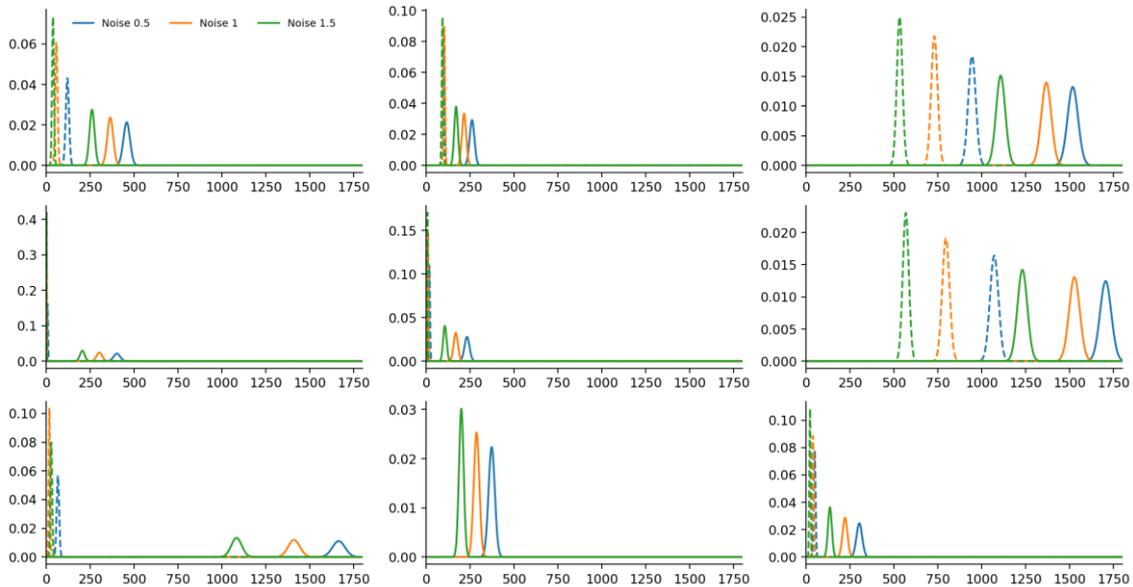
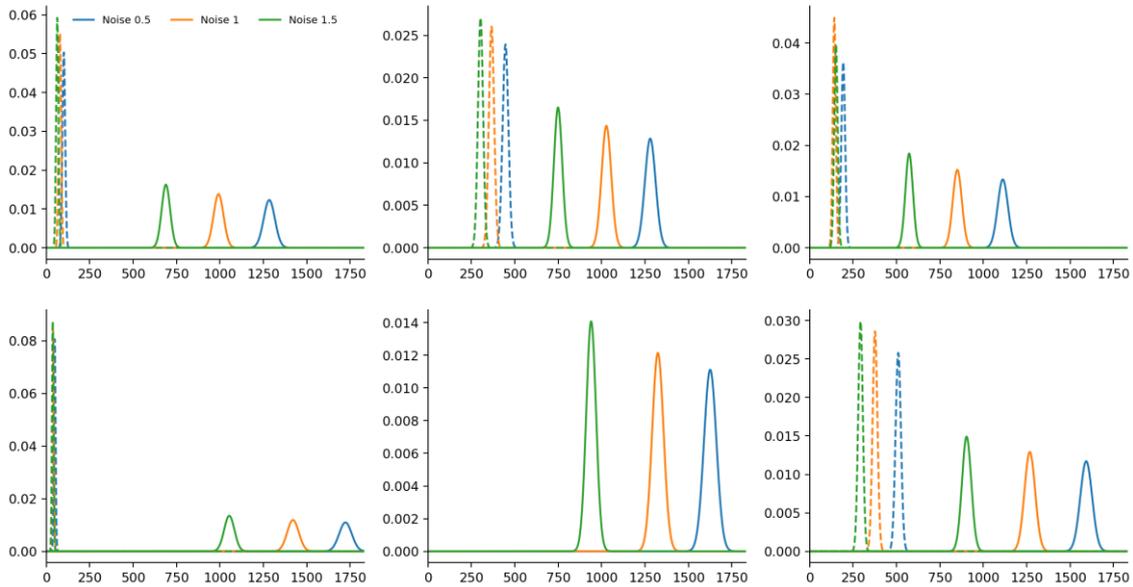


Figure 31: SLSTR Savannah site. Aggregation to CMG results. Each panel shows the aggregated BA PDF of a 100x100 pixel area in the original pixel product (x-axis: burned pixels per area, y-axis: frequency). The dashed lines represent the PDF calculated from the inferred pb value, and the full lines represent the PDF calculated from the mean of the algorithm reported pb across the 20 MC realisations. Different colours represent different noise levels (blue: 0.5 times nominal noise level, orange: nominal noise level and green 1.5 times nominal noise level).

Tile h24v03



**Figure 32: SLSTR Boreal site. Aggregation to CMG results. Each panel shows the aggregated BA PDF of a 100x100 pixel area in the original pixel product (x-axis: burned pixels per area, y-axis: frequency). The dashed lines represent the PDF calculated from the inferred pb value, and the full lines represent the PDF calculated from the mean of the algorithm reported pb across the 20 MC realisations. Different colours represent different noise levels (blue: 0.5 times nominal noise level, orange: nominal noise level and green 1.5 times nominal noise level).**

## 7. Discussion

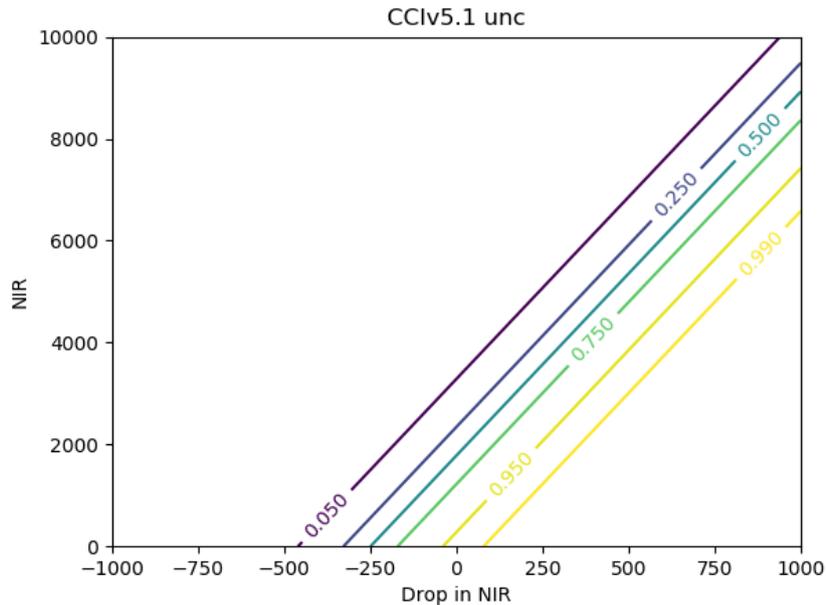
### 7.1. MODIS Fire\_cci v5.1

Overall, the algorithm reported uncertainty shows spatial patterns that are similar to the inferred one. However, the actual values are quite different, and the reported probability of burn increases with increasing noise, which is unexpected: it should decrease or stay the same if other sources of evidence still point to fire/no fire. Additionally, the reported probability of burn changes very substantially between realisations of the same scenario, indicative of a strong dependence on the actual data.

Given that other ancillary datasets are not modified in the MC scenario, one can look at the method used in this product to calculate uncertainty and try to understand how these effects appear. From the description of the algorithm in Section 5.1, a sensitivity analysis shows that there is little effect on the number of observations once there are more than 2. As the thermal anomalies have not changed, the distance term to anomaly term is constant, and the only sources of change are the NIR reflectance of the burned pixel, and the relative drop in NIR reflectance between the pre- and post-fire dates. A plot of the mapping is shown in Figure 33. It is clear that the transition zone is very narrow, so small changes in the relative drop of near infrared reflectance will precipitate large swings of the reported pb. For example, if the pixel NIR reflectance is 0.2, a relative drop of 0 will result in a calculated pb between 0.5 and 0.75. The different realisations of each scenario can then be seen as a random walk around the interface: for unburned pixels, there is very little chance of going above 0.1 (say), but there can be large variations for burned pixels, as small changes in e.g. drop in relative NIR reflectance due to noise, BRDF effects or different choices of day of burn at the



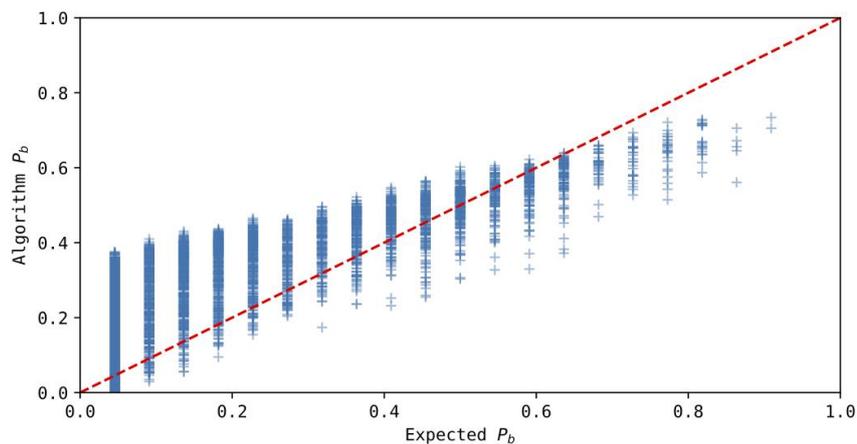
compositing stage can result in large unpredictable changes along the x axis in Figure 33. This suggests that while the chosen two magnitudes are clearly relevant, the estimator will be showing too much sensitivity to the actual magnitudes. Instead of using these individual values, it might be worth exploring the use of the average relative NIR drop for a period after the fire, as this will limit the variations introduced by BRDF effects and noise. It is possible that exploring other variables (such as the number of burned neighbours) might provide more insights, but assessing their usefulness is something that can only be done once the uncertainty quantification approach has been fixed and an experiment based on a synthetic realisations like that shown here takes place.



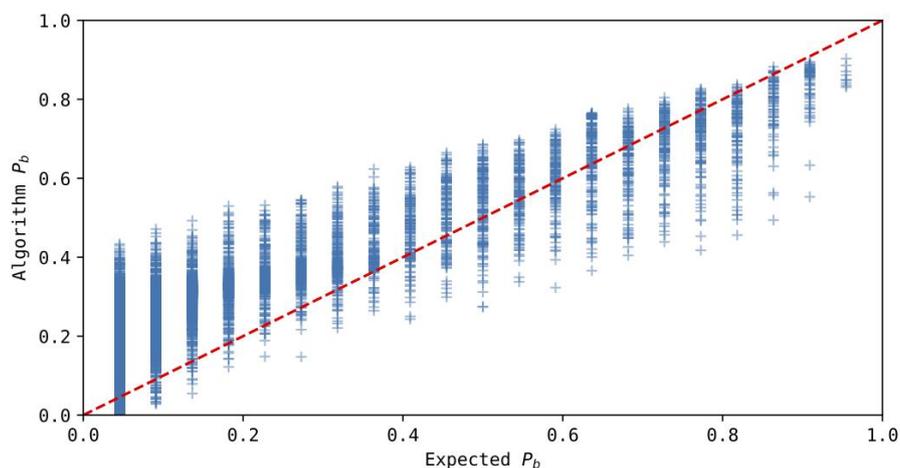
**Figure 33: Uncertainty space for the Fire CCI v5.1. The plot shows how reported uncertainty is calculated as a function of near infrared reflectance of the pixel (y axis) and the relative drop in near infrared reflectance between the fire data and the previous observation (x axis). The contours show the values of the probability of burn calculation.**

## 7.2. SLSTR algorithm

The reported probability of burn from the algorithm is broadly in line with the values inferred by the MC experiment. Spatial patterns are similar, with high probability of burn on clear burns, and lower probabilities of burn in unburnt patches. Comparisons between the average reported probability of burn across different realisations as shown in Figure 34 and Figure 35 demonstrate that while there is a large scatter, reported values are comparable to inferred ones, and the trend is sensible across biomes. While the reported algorithm probability calculations are only approximate, they result in a behaviour which is consistent with the benchmark dataset.



**Figure 34: Comparison of inferred (or expected) probability of burn versus algorithm reported mean probability of burn across MC realisations for boreal, nominal noise level.**



**Figure 35: Comparison of inferred (or expected) probability of burn versus algorithm reported mean probability of burn across MC realisations for subtropical site, nominal noise level.**

## 8. Conclusion

There is a strong requirement for uncertainty quantification in burned area products. The current report is a step in this direction, focusing on the efforts of the Fire\_cci project. The first Section introduces uncertainty propagation to a simple model. While not a practical endeavour in its own right, it shows the general approach that algorithm developers may want to follow for uncertainty propagation through arbitrary models. The aim of this uncertainty propagation when discussing BA products is to provide an estimate of the per pixel probability of burn,  $p_b$ . This value gives an indication of the strength of the evidence in labelling a pixel as burned or unburned, but it also assesses the effect of other nuisance factors, such as poor temporal sampling, residual atmospheric correction, issues derived from gridding and the modelling of the sensors instantaneous field of view (IFOV), etc. All these processes result in a change in the observations that are used as an input to the BA algorithms, but if they can be statistically characterised and the algorithm design follows some guidelines, they can be directly propagated into a probability of burn. In an ECV scenario, having these uncertainties for individual products derived from different sensors would allow a

	<b>Fire_cci</b> <b>Comprehensive Error Characterization</b> <b>Report</b>	Ref.:	Fire_cci_D1.3_CECR_v2.0		
		Issue	2.0	Date	18/05/2018
		Page	43		

consistent blending of products: the uncertainty characterises the strength of the evidence of each product, and as such, can be used as a weighting for this combination. This is however challenging both in terms of algorithm design, and in terms of input products (e.g. surface directional reflectance, thermal anomalies) being properly characterised by an uncertainty estimate.

As BA products are often used on global scales, it is important to consider the spatial aggregation of the original pixel-level product to a climate modeller's grid (typically, grid sizes will be of the order of 10s or 100s of km). For binary BA products, this is often done by summing the number of pixels within the CMG cell, but once the pixels are qualified by an uncertainty measurement, this needs to be propagated as well. We develop a method to achieve this, based on the assumption of individual pixels being independent Bernoulli trials. The aggregation of this set of trials results in a Poisson binomial distribution that can, in most practical cases, be approximated by a Gaussian distribution. In this respect, the CMG burned area is defined by a mean value and a standard deviation or variance. Note that if burned pixels have a probability of 1 and unburned pixels a probability of burn of 0, then the mean of the Gaussian is equal to the sum of burned pixels (consistent with previous approaches), but the variance is then zero (in effect, encoding no uncertainty).

As propagation of uncertainty through an established algorithm is hard, a method of inferring it has also been introduced. This method is based on Monte Carlo sampling, and as such is numerically intensive. Different realisations of the observations are presented to the algorithm. While all the datasets are different, their statistics (noise, number of available observations, angular sampling, etc.) are identical. The algorithm is run for these realisations, which allow an inference of the true per-pixel probability of burn. This provides a benchmark value to compare algorithm-reported probability of burn metrics with, and potentially a way to develop methods to provide an indication of probability of burn. Although the algorithm could be used to provide uncertainty estimates of arbitrary BA algorithms, computational complexity and thorough investigation of the sampling regime used for the Monte Carlo simulations make this approach impractical on a large scale.

Applying the benchmarking approach to the MODIS Fire\_cci v5.1 and SLSTR BA algorithms shows that the behaviour is expected: as noise in the input increases, the inferred probability of burn tends to drop, but the actual changes are different between sites and algorithm used.

Comparisons of the benchmark metric against the algorithm reported values show that the algorithm reported uncertainties still need development work. While both investigated algorithms (MODIS Fire\_cci v5.1 and SLSTR) show general probability of burn high over clear burn scars and low over unburned pixels, the actual values differ a lot from the benchmark. In the case of the MODIS Fire\_cci v5.1 algorithm, there is also a large variation of the uncertainty with realisation (sometimes up to 0.4-0.5 in units of probability of burn), suggestive of an important effect of the random variation introduced in the observations, and thus a low level of robustness. This is explained as an over-reliance of the calculation of the aggregation in the actual observations, rather than their statistical properties. However, some simple pre-processing of the magnitudes that go into the calculation might result in more robust estimates. For example, instead of using a single measure of post-fire NIR reflectance drop, this metric could be averaged over some time after the fire to minimise fluctuations from BRDF and other effects. Similarly, other spatial contextual metrics could be introduced (such as the

	<b>Fire_cci</b> <b>Comprehensive Error Characterization</b> <b>Report</b>	Ref.:	Fire_cci_D1.3_CECR_v2.0		
		Issue	2.0	Date	18/05/2018
		Page	44		

number of burned neighbours). Additionally, after fitting it would be important to consider the sensitivity of the different input magnitudes in the uncertainty calculations and trim those deemed to have a low predictive value.

The SLSTR case is generally more in line with the benchmark, although the actual value of the probability of burn might not be accurate. As the probability of burn of the SLSTR algorithm is in effect a scaled version of the MaxEnt score with some additional scaling to consider thermal anomalies and belonging to a patch, it is hard to assess the causes of uncertainty. However, the tapering of probability of burn towards edges as well as flagging of areas with low probability suggest that most of the uncertainty is coming from the contrast of the burn signal.

The effect of the discrepancy in probability of burn estimates results in very different estimates of the burned area PDF for the CMG case. This is a consequence of the calculations of the aggregation being driven by an assumption of the per pixel probability of burn being correct.

In summary, in this report we introduce methods to both propagate uncertainty, aggregation per pixel uncertainty to coarse resolutions, and infer per pixel uncertainty through runs of a binary output algorithm. These are general methods that can be used to test and learn more about uncertainty and how uncertainty is calculated in BA products. Additionally, we use these techniques to assess the calculations of uncertainty in the MODIS Fire\_cci v5.1 and SLSTR BA products, showing that reported uncertainties are very different between products, and while spatial pattern are largely similar to the benchmark, the actual values are very different. Clearly, more effort needs to go into uncertainty quantification of BA algorithms.

## 9. References

Lucht, W. and P. Lewis (2000) Theoretical noise sensitivity of BRDF and albedo retrieval from the EOS-MODIS and MISR sensors with respect to angular sampling. *International Journal of Remote Sensing*, 21(1) 81-89.

Gomez-Dans J, Lewis P, Disney M, Roy D, Quaife T, Wooster M. (2013) Edge-preserving data assimilation for fire monitoring using optical data. *Living Planet Symposium*. European Space Agency; doi:10.5281/zenodo.34634

Megill, Norman D., and Mladen Pavicic (2011). "Estimating Bernoulli trial probability from a small sample." *arXiv preprint arXiv:1105.1486*.

Pettinari, M.L., Chuvieco, E., Alonso-Canas, I., Storm, T., Padilla Parellada, M. (2016) *ESA CCI ECV Fire Disturbance: Product User Guide, versión 2.1*.

Quaife, Tristan, and Philip Lewis (2010). "Temporal constraints on linear BRDF model parameters." *IEEE Transactions on Geoscience and Remote Sensing* 48, no. 5: 2445-2450.

Tansey K., Bradley A. V. and Padilla M. (2014). "ESA CCI ECV Fire Disturbance: Algorithm Theoretical Basis Document - Volume III - BA Merging." *Fire\_cci\_Ph3\_UL\_D3\_6\_3\_ATBD\_III\_v2\_3*.

Wanner, W., Li, X., & Strahler, A. H. (1995). On the derivation of kernels for kernel-driven models of bidirectional reflectance. *Journal of Geophysical Research, D: Atmospheres*, 100(D10), 21077–21089.

	<b>Fire_cci</b> <b>Comprehensive Error Characterization</b> <b>Report</b>	Ref.:	Fire_cci_D1.3_CECR_v2.0		
		Issue	2.0	Date	18/05/2018
		Page	45		

Wanner, W., Strahler, A. H., Hu, B., Lewis, P., Muller, J.-P., Li, X., ... Barnsley, M. J. (1997). Global retrieval of bidirectional reflectance and albedo over land from EOS MODIS and MISR data: Theory and algorithm. *Journal of Geophysical Research*, 102(D14), 17143. <https://doi.org/10.1029/96JD03295>

Weisstein, Eric W. "Standard Error." From MathWorld--A Wolfram Web Resource. <http://mathworld.wolfram.com/StandardError.html>

## 10. Annex 1: Acronyms and abbreviations

ATBD	Algorithm Theoretical Basis Document
BA	Burned Area
BRF	Bidirectional Reflectance Factor
BRDF	Bidirectional Reflectance Distribution Function
CCI	Climate Change Initiative
CDF	Cumulative Distribution Function
CECR	Comprehensive Error Characterization Report
CMG	Climate modelling grid
DCT	Discrete Cosine Transformation
DoB	Day of Burn
ECV	Essential Climate Variables
ESA	European Space Agency
IFOV	Instantaneous Field of View
MaxEnt	Maximum Entropy
MODIS	Moderate Resolution Imaging Spectroradiometer
NBR	Normalized Burned Ratio
NDVI	Normalized Difference Vegetation Index
NIR	Near Infrared
PDF	Probability distribution function
SLSTR	Sea and Land Surface Temperature Radiometer