Climate Change Initiative Extension (CCI+) Phase 1
New Essential Climate Variables (NEW ECVS)
High Resolution Land Cover ECV (HR_LandCover_cci)

End-to-End ECV Uncertainty Budget

(E3UB)

Prepared by:

Università degli Studi di Trento
Fondazione Bruno Kessler
Università degli Studi di Pavia
Università degli Studi di Genova
Université Catholique de Louvain
Politecnico di Milano
Université de Versailles Saint Quentin
CREAF
e-GEOS s.p.a.
Planetek Italia
GeoVille

Changelog

Issue	Changes	Date
1.0	First version.	02/07/2019
2.0	Updated version.	03/01/2020

Detailed Change Record

Issue	RID	Description of discrepancy	Sections	Change

Contents

1 Introduction

1.1 Executive summary

This document provides an assessment of the end-to-end uncertainty budget the HRLC ECV products are associated with. HRLC products are based on a wide range of input data whose uncertainties propagate at different levels of dependency according to the data characteristics and the processing steps involved in the production. By taking into account the scarce availability of ground-measured reference information and the practical impossibility to collect physical measurements on wide areas as those selected for this project, the proposed uncertainty models will be, by necessity, theoretical.

In year 1 of the CCI HRLC Phase 1, the core activity concerning algorithm implementation relates to Algorithm development, System Development and a round-robin for the classification algorithms. Many steps of the processing chain (e.g., pre-processing, geolocation, etc.) are involving well-known best-performing algorithms that come with uncertainty models associated to them. For instance, the classification task is able to output probabilistic posteriors that can be managed at the fusion level to infer uncertainty score pixelwise.

Both uncertainties of input data sets and processing model-related ones must be considered, including error propagation dynamics. The nature of the input data sets (discrete classes vs. continuous variables) and the associated error characteristics (random error/ bias, error distribution), including potential correlations between errors of different input variables should be evaluated. Finally, uncertainties related to the spatial scales of data sets, scaling issues related to the validation activity must be accounted for as well.

1.2 Purpose and scope

This document provides an overview of the main sources of uncertainty for the HRLC ECV variables, i.e., LC and LCC. The global behaviour of error propagation to the uncertainty budget estimation is currently ongoing. Indeed, outcomes from internal benchmarking activities are key for drafting a first global model that integrates all the sources of error/uncertainty into one product. Therefore, this issue of the document presents a list of all potential sources of error, uncertainty and known correlations in the data that will contribute to the global model to be delivered in future issues. The structure of the document is thus sequential, listing all the items related to the different processing steps.

1.3 Applicable documents

Ref.	Title, Issue/Rev, Date, ID
[AD1]	CCI HR Technical Proposal, v1.1, 16/03/2018

1.4 Reference documents

Ref.	Title, Issue/Rev, Date, ID

1.5 Acronyms and abbreviations

SAR	Synthetic Aperture Radar
RMSE	Root mean square error
RST	Rotation scale translation
GT	Ground truth
LOGP	Logarithmic opinion pool
ATBD	Algorithm Theoretical Basis Document
MRF	Markov Random Field
MAP	Maximum a posteriori

MPM	Marginal a posteriori modes
LaSRC	Landsat-8 surface reflectance code
TOA	Top of Atmposphere
MODIS	Moderate Resolution Imaging Spectroradiometer
AOT	Aerosol optical thickness
LEDAPS	Landsat Ecosystem Disturbance Adaptive Processing System
OLI	Operational Land Imager
MSI	Multispectral Instrument
SWIR	Short-wave infrared
GRD	Ground Range Detected
GUM	Uncertainty in Measurement
DEM	Digital Elevation Model
UTM	Universal Transverse Mercator

2 Optical pre-processing

Detailed analysis of pre-processing errors/accuracy related to harmonized Sentinel-2 / Landsat products is given in [1]. Altough the final processing chain of the CCI HRLC project will not be identical to [1], this work provides the most complete reference for the prior modelling of the variables that contribute to pixel-level measurable uncertainty of the products coming from an integrated pre-processing stage.

2.1 Radiometric correction

LaSRC assumes a Lambertian, plane-parallel atmosphere, and uses the 6S radiative transfer model to invert directional surface spectral reflectance from observed top-of-atmosphere reflectance. Several atmospheric parameters are required for the inversion including surface pressure, column water vapor, ozone, and aerosol properties. LaSRC algorithm assumes two SR ratios, red to blue and red to ultra blue, and uses the difference between these assumed ratios and observed TOA reflectance ratios to invert for AOT and Angstrom exponent. The two fixed SR ratios for the globe are derived from MODIS and MISR data, and ex-pressed as a function of mid-infrared vegetation index.

Currently, uncertainty estimates for LaSRC are based on comparison with corrections based on in situ atmospheric parameters from the Aerosol Robotic Network [1]. These comparisons indicate improved performance compared to the LEDAPS algorithm or an alternative version of LEDAPS that used MODIS aerosol products as input. For Landsat 8 OLI, overall uncertainty varied from 0.11% absolute reflectance (SWIR1 band) to 0.85% absolute reflectance (blue band). For Sentinel-2/MSI, overall uncertainty varied from 0.3% absolute reflectance (SWIR1band) to 1.4% absolute reflectance (blue band).

2.2 Cloud and cloud-shadow detection / restoration

Cloud and cloud-shadow detection accuracy is intrinsically difficult because of the impossibility of directly measuring physical parameters related to clouds. Posterior evaluation of cloud detection accuracy can be figured by referencing to appropriate literature, but this does not provide a direct method to quantify uncertainty in a pixelwise manner. A possible strategy to mitigate this problem is to associate probabilities in the classification step related to cloud identification. This is similar to the idea implemented in processors like sen2cor.

2.3 Spectral filtering and harmonization

Given the differing solar and view angles associated with Landsat 8and Sentinel-2, normalizing the BRDF effects is desirable. Retrieving the BRDF information directly from medium resolution optical remote sensing data is not feasible with the current temporal and angular distribution of the data. Instead, the BRDF information needs to be ingested a priori. It is currently on-going revision of the most appropriate technique to achieve this so that effects of different illumination conditions may be included in the model.

3 SAR pre-processing

The spaceborne synthetic aperture is a powerful key Earth observation technique for large-area monitoring, and in recent decades, its all-weather capability and fine ground resolution facilitated the development of a great variety of applications, such as the land mapping, for example. Due to the nature of the SAR range mapping and reflectance functions, the measurements of multi-channel SAR system can be biased by error originating from many deleterious factors in which significantly degrade the quality of SAR image. In this section we quantify sources of uncertainty on SAR results, i.e. all those aspects that leading doubts about the validity of the result of a measurement (or processing).

1.1 SAR processing chain

Undoubtedly a first uncertainty cause of uncertainty in SAR processing is to be researched in the application of techniques for reconstructing the observed scene, such as Level-1 data produced as Single Look Complex (SLC) and Ground Range Detected (GRD) from raw data (also so-called Level-0), i.e. the backscattered wave. This may be modeled for instance by the Uncertainty in Measurement (GUM) framework, introduced in [2] and used to evaluate the uncertainties of the amplitude values pixel by pixel, on the bases of a statistical analysis.

1.2 Geometric processing

A second source of uncertainty refers to next steps applied to SAR data sets, such as geometric processing. Specifically, radar image processing requires the geometrical overlaying of the remotely data sensed from different sensors and/or geometries, in order to mitigate very severe distortions over elevated and sloping terrain [3]. Before the change detection and surface classification, these distortions must be mitigated by terrain correction. SAR imaging requires precise determination of the relative position and velocity of the radar platform with respect to its target at all times. However, this information is not available at the time of imaging, since the platform must be moving for the azimuth resolution technique to work. The basic solution is to assume that the platform has a uniform velocity over a smooth geoid [4]. To remove the geometric distortions due to terrain relief, the *radiometric calibration* could be applied for compensating all spatial and time dependent variation, as well as the cross-track and image to image intensity inconsistencies due to signal attenuation by distance [5]. The algorithm is based on a Digital Elevation Model (DEM) and creates a simulated SAR image based on an imaging radar model.

However, this process introduces a level of uncertainty because it simplifies the image creation considerably, introducing certain amount of distortions into the results. In fact, when there is significant relief in the area being imaged, for example, the DEM model that is being adopted will be based on a proper smooth geoid assumption and will lead to pixel placement, rendering it suitable (or not) for a quantitative analysis of terrain features.

Finally, to correct image orientations, SAR images are geocoded. This phase includes SAR image resampling to a spatial representation with known geometric properties. Standard map projections, like the Universal Transverse Mercator (UTM) mapping, are used. Processing involves the image rotation and scaling to properly transform it into the mapping coordinates chosen [6]. The geocoding represents another uncertainty source, since images are taken at varying pass angles, and each resulting image contains an approximately square rotated SAR image inside it, with unused image pixels set to black.

1.3 Speckle noise

In addition to geometrical features, another limitation of SAR data sets is the speckle noise. The speckle noise is an intrinsic feature of SAR data, and it is given by the consistent summation of signals from ground scatters randomly and loosely distributed within the scene. The existence of speckle noise in SAR images is an inherent and specific random characteristic. This noise has an impact on the interpretation of these images and introduces further limitations in applications exploiting SAR time series [7]. The speckle noise reduction has to strictly be carried out in order to preserve polarimetric properties, without introducing any image quality degradation and corrupting statistical characteristics. Consequently, not using a suitable speckle filtering involves a dramatic impact in terrain classification performance [8].

Instead, performing a multi-temporal analysis based on multiple images acquired over time, backscattered values in can be aggregated in both coherent and incoherent ways to reduce the effect of noise. For classification purposes, it has been proved [9] that the use of multitemporal sequences improves the accuracy of the final results, either thanks to the fusion at the decision level of the results for each image, or by combining multiple SAR images into a single input to the classification procedure.

4 Multi-sensor geolocation

Multi-sensor geolocation [10] aligns data collected from different sensors (a reference and still image and an input image to be transformed) in a common reference system in order to process them coherently further in the processing chain. Within the CCI+ HRLC processing chain, the two image data sources correspond to optical and SAR data. The result of the geolocation process may be more or less precise, yielding uncertainty associated with the output images. Such uncertainty generally also affects the subsequent blocks along the processing chain.

There exist different strategies for assessing the accuracy of the geolocation process: one of the possibilities is the computation of the root mean square error (RMSE) in pixel units [11]. In an experimental setup, where the correct transformation is known, the RMSE may be computed analytically. Otherwise, it is possible to estimate it by means of specific control points or landmarks.

The control points are identified in both images and the RMSE is computed based on the residual spatial distances. Ideally, in case the images are perfectly matched, the distances of the control points in the reference and registered images is equal to zero. In all the other cases, the control points may not be perfectly matched, and the distances are generally non-zero, although they can be smaller than one-pixel size, on average.

The following section describes the computation of the RMSE in case the transformation is known and is modelled as a "rotation-scale-translation" (RST) transformation. For lower-complexity transformation, like rigid or shift transformations, the same computation holds but it is necessary to fix unitary scale (rigid) and the rotation angle to zero degrees (shift).

4.1 An example of RMSE computation

Let $Ref(x,y)$ and $In(x,y)$, $(x,y) \in \Omega \subset \mathbb{R}^2$, where Ω is a region of interest, be two images called reference and input, respectively. If they are both of size $A \times B$ pixels, then $\Omega = [0,A] \times [0,B]$. In the RST case, $T_p(x,y)$ is the geometric transformation described by the parameter vector $p = \left(t_x, t_y, \theta, k\right)$ and has the form:

$$T_p(x,y) = \begin{pmatrix} k\cos(\theta) & k\sin(\theta) & t_x \\ -k\sin(\theta) & k\cos(\theta) & t_y \end{pmatrix} \begin{bmatrix} x \\ y \\ 1 \end{bmatrix}$$

where $\{t_x, t_y\}$ determine translations in the x and y directions, θ is the rotation angle, and k is the scaling factor. Thus, we can write $T_p(x,y) = Q_p \cdot [x,y,1]^T$, where Q_p is the RST transformation matrix given above, and the superscript "T" indicates the transpose operator. There is a one-to-one correspondence between Q_p and p. To register $Ref(x,y)$ and $In(x,y)$ it is necessary to find the value of p such that $In\left(T_p(x,y)\right)$, the input transformed by T_p, best matches the reference (see ATBD).

When accurate ground truth is available, such as when test images are created synthetically (a typical scenario when a geolocation method is developed and is being validated), a standard way of assessing registration accuracy is by using the RMSE $E(p_e)$ [11]. Suppose the ground truth (GT) transformation is given by $p_{GT} = \left(t_{x1}, t_{y1}, \theta_1, k_1\right)$ and the computed transformation is $p = \left(t_{x2}, t_{y2}, \theta_2, k_2\right)$, with the two RST matrices $Q_{p_{GT}}$ and Q_p respectively. It is possible to define the error transformation $p_e = \left(t_{xe}, t_{ye}, \theta_e, k_e\right)$, along with the corresponding RST matrix Q_{P_e}, and measure the discrepancy between p_{GT} and p.

According to the matrix formulation of the RST transformation, being Q_{P_e} the error transformation matrix, the following should hold [11]:

$$Q_{P_e} = Q_p \cdot Q_{p_{GT}}^{-1}$$

that yields:

$$\begin{cases} k_e = \dfrac{k_2}{k_1}, \quad \theta_e = \theta_2 - \theta_1 \\ t_{xe} = t_{x2} - k_e\big(t_{x1}\cos(\theta_e) + t_{y1}\sin(\theta_e)\big) \\ t_{ye} = t_{y2} - k_e\big(t_{y1}\cos(\theta_e) - t_{x1}\sin(\theta_e)\big) \end{cases}$$

Now, let $(x, y) \in \Omega$ and let $[x', y']^T = Q_{P_e} \cdot [x, y, 1]^T$. This can be equivalently written as:

$$\begin{bmatrix} x' \\ y' \end{bmatrix} = k_e \begin{pmatrix} \cos(\theta_e) & \sin(\theta_e) \\ -\sin(\theta_e) & \cos(\theta_e) \end{pmatrix} \begin{bmatrix} x \\ y \end{bmatrix} + \begin{bmatrix} t_{xe} \\ t_{ye} \end{bmatrix}$$

Then, the RMS error is defined as:

$$E(p_e) = \sqrt{\frac{1}{AB}\int_0^B \int_0^A (x' - x)^2 + (y' - y)^2 \, dx \, dy},$$

Substituting the formula for x' and y' and solving for $E^2(p_e)$ yields:

$$E^2(p_e) = \frac{\alpha}{3}(k_e^2 - 2k_e\cos(\theta_e) + 1) + \big(t_{xe}^2 + t_{ye}^2\big) - \big(At_{xe}^2 + Bt_{ye}^2\big)(1 - k_e\cos(\theta_e)) \\ - k_e\big(At_{ye} - Bt_{xe}\big)\sin(\theta_e)$$

where $\alpha = A^2 + B^2$. This formula is used in this research to measure registration accuracy when the ground truth transformation is available. This scenario holds when a geolocation method is being developed and semi-simulated data sets are used for its tuning and validation. It obviously does not apply to the case in which two real image data sets are available because the reference GT transformation is not known.

4.2 RMSE Computation without GT

The previous section describes how to compute analytically the RMSE in case ground truth data is available. In case no ground truth information is available, it is still possible to compute an estimate of the RMSE through control points (landmarks) [10].

It is possible to identify well-known control points in both the reference and the registered input image and estimate the accuracy of the registration process from them. Hence, a sample estimate of the registration RMSE can be computed by averaging the single RMSE computed with respect to each pair of control points, i.e., as a sample estimate of the RMSE functional.

4.3 From Registration Error to Uncertainty

Once the registration error is estimated, one can indirectly derive information about the uncertainty generated by the geolocation process within the overall land-cover mapping process. In particular, it is convenient to distinguish two scenarios based on the achieved error.

4.3.1 Sub-Pixel Registration Error

There are cases where the registration error is less than a single pixel. Thus, the grid of the two images after registration is almost perfectly matched. Indeed, this is the goal of most image registration efforts.

In this case, every pixel is matched with the corresponding pixel in the other image. Obviously, there may still be a residual error. However, achieving sub-pixel accuracy implies that the Earth region associated with a pixel in the reference image is almost the same as the Earth region associated with the registered image, i.e., spatial mismatch between the optical and SAR sources is smaller than the pixel size after registration.

This is the best possible achievement in multi-sensor geolocation, and in this case, no uncertainty is deemed to be forwarded to the following processing blocks of the chain. Every pixel is correctly located, and no uncertainty needs to be propagated along the overall processing chain.

4.3.2 Non-sub-pixel Registration Error

If the registration RMSE is larger than one pixel, the pixels in the reference image are not correctly matched with the pixels in the registered image, on average, and the mismatch implies an actual misalignment of the data associated with the two pixel grids. Non-sub-pixel errors may cause artefacts in boundary regions, where pixels of different classes are superimposed due to the residual shift between the reference and registered images. Conversely, in a flat homogeneous image region, registration error may not cause problems, as the mismatch may not influence the resulting classification map.

In this case, the uncertainty is propagated to the data fusion block. Operatively, the probabilistic fusion that takes place in the decision fusion module may generally be affected by the residual non-sub-pixel misregistration. The goal of the CCI+ HRLC processing chain is to estimate land cover. In principle, the impact of non-sub-pixel registration error on land cover uncertainty may be assessed, on each pixel, by making use of the probabilistic per-pixel information available to the decision fusion module within a local window, whose size will be determined as a function of the registration RMSE. However, residual misregistration is expected to intrinsically translate *per se* into increased uncertainty in the pixelwise posterior probabilities obtained by the HRLC processing chain on each pixel. Moreover, the aforementioned local-moving-window process may add significantly to the overall computational burden. Therefore, consistently with the goal of assessing the uncertainty in the output HRLC product, the impact of residual misregistration on the overall uncertainty will be characterized through the pixelwise posterior distribution. The possible use of moving-window processes will be considered methodologically or experimentally in a tradeoff with computational burden.

5 Classification

Uncertainty is unavoidable in all classification domains: a certain amount of uncertainty is always involved in deciding the class a sample is assigned to. The unanimously recognized framework to represent uncertainty is probability. Specifically, the Bayesian concept of maximum posterior probability encloses the amount of uncertainty (measurable from data) that generates in the probabilistic decision of a classifier. Since no specific classifier has been selected yet (both for optical and SAR processing classification), the treatment is given in this section in a general way. In Figure 1, the general workflow associated with the classification part of the processing chain is recalled.

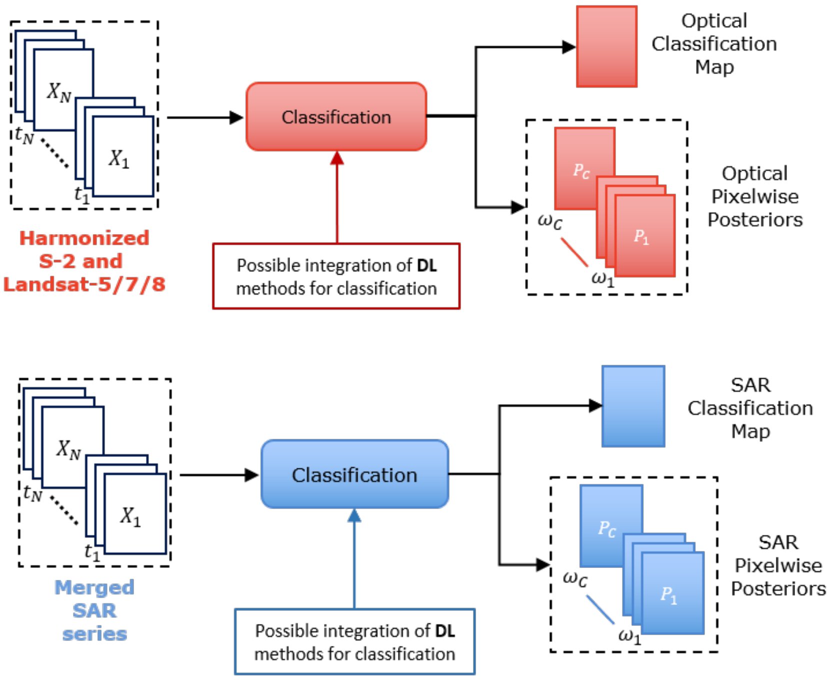

Figure 1. Workflow of the classification process for optical and SAR time series of images.

To model input of the Decision Fusion block dealing with integration of different sources of uncertainty, here we present a general framework to posterior probability definition that is algorithm independent [10]. We can model the posterior probabilities using $m$linear classifiers. Each linear classifier implements a hyperplane that separates its corresponding class from the other classes. The equation of each hyperplane is

$$f_j(\boldsymbol{x}) = \boldsymbol{w}_j \cdot \boldsymbol{x} + b_j$$

where $\boldsymbol{w}_j$ is a weight vector and b_j is a bias term. Each source-specific posterior probability is typically computed using the softmax function:

$$P(\omega_j | \boldsymbol{x}_i, \theta_h) = \frac{\exp(f_h(\boldsymbol{x}_i))}{\sum_{k=1}^{m} \exp(f_k(\boldsymbol{x}_i))}$$

Hence, the parameter set $\theta_h = (\boldsymbol{w}_h, b_h)$. For any probabilistic-based classifier, plugging the posterior probabilities into the cross-entropy function and solving the equation (gradient methods) is at the core of the well-known learning rules and or backpropagation algorithms.

As widely acknowledged, SAR images are an important source of information but the speckle noise gives SAR images a granular appearance that makes interpretation and analysis hard tasks. Furthermore, one of the major issues is the assessment of topographic information content in this kind of images that could be extrapolated by exploiting classification techniques. Classification accuracy values, which include user’s accuracy (UA), producer’s accuracy (PA), and overall accuracy (OA), are strongly influenced by the adopted input model, which could cause considerable errors in the model output.

An image contains an enormous amount of information, and the challenge is how to represent it in a more compact way, which is why features are originated. In other words, for a more compact and possibly more significant representation of the information embedded in image, it is usually decomposed into several features. Specifically, the extraction of spatial features from remotely sensed data and the use of this information as input to further processing steps has received considerable attention over the two last decades. Unfortunately, due to the complexity of the images and the existence of image noise and disturbances, the information derived from an image is always ambiguous. This source of uncertainty makes the following recognition/classification process more complex. The accuracy of spatial feature extraction can hardly formulated due to intra-class variation and inter-class similarity [12]. This uncertainty of spatial features extraction is of course on top of the other source of inaccuracy, common to any image element, such as positional uncertainty, attribute uncertainty, topological uncertainty, computational inaccuracy, imprecision/inexactitude, inconsistency, incompleteness, repetition, vagueness, omission, misinterpretation, misclassification, abnormalities and knowledge uncertainty.

Land cover maps are generated by the classification of remote sensing data, and are frequently used as input to spatially explicit environmental models, and its quality is generally assessed at the global or class-specific. It is therefore clear that another uncertainty factor is closely correlated with the used classification algorithms. However, a clear assessments of classification accuracy is not easily provided, and several work have been proposed to evaluate the technical classification uncertainty [13], [14]. These studies demonstrate that uncertainty assessment provides valuable information on the performance of land cover classification models, both in space and time.

Finally, it must be noted that supervised classifiers are always based on class assignment rules that derive from a set a multiclass training samples, which consequently introduces another uncertainty level. The training data are samples of individual classes and the class assignment rules are derived from the entire study area. A high quality training dataset is mandatory to train the classifier model. However, in practice, the label (class name) in a training dataset may not be correct (when it is generated by a human interpreter, for instance, mistakes are going to happen). Therefore, if the training dataset is not of high quality, it may lead to lower classification performances. Thus, classification accuracy is inherently associated with uncertainty [15].

6 Decision fusion

Data fusion methodologies should consider source-specific uncertainties in order to estimate the overall uncertainty of the classification result. More in detail, decision fusion combines the posterior probabilities associated with the outputs of single classifiers when applied to the single data inputs, here namely optical and SAR data. Therefore, multiple decisions are combined into a final result by taking into account the level of uncertainty associated with each source, which is intrinsically expressed by the corresponding pixelwise posterior probability distribution.

As described in ATBD-v2, the whole class legend Ω is divided into: Ω_O, the set of classes that are distinguished only by using optical data ("optical-exclusive"); Ω_S, the set of classes that are distinguished only by using SAR data ("SAR-exclusive"); and Ω_c, the set of classes that are discriminated by the classifiers operating with both data modalities (common classes). The optical classifier works on the set of classes $\Omega_O \cup \Omega_C$, the SAR classifier outputs posterior probabilities for the set of classes $\Omega_S \cup \Omega_C$. The decision fusion stage first merges the optical and SAR outputs on the common classes Ω_C, then it takes into account the presence of the exclusive classes Ω_O and Ω_S through a class-specific combination rule.

The following subsections discuss uncertainty modelling issues with regard to the families of decision fusion methods that are developed, i.e., weighted voting and consensus theory, and fusion based on Markovian modelling (both families are combined with the aforementioned class-specific combination rules).

6.1 Uncertainty in Consensus Theory

Consensus theory [16], [17] involves general procedures with the goal of combining multiple probability distributions to summarize their estimates. Since the use of consensus theory simply aims at fusing posterior probabilities coming from different classifiers, it is possible to obtain again a probability distribution. The source-specific uncertainties are therefore directly combined in the process, leading to an overall uncertainty.

In the case of the HRLC pipeline, the individual information sources correspond to the outputs of the optical and SAR processing chains. Since the two classifiers generally work on different sets of classes, this fusion is possible on the set of common classes Ω_C only. Considering this specific case, the two major consensus-theoretic approaches, i.e., linear and logarithmic opinion pool (LOP and LOGP) compute functionals $\mathcal{C}(\omega_j|\underline{x}, \Omega_C)$ and $\mathcal{L}(\omega_j|\underline{x}, \Omega_C)$ ($\omega_j \in \Omega$) that merge the pixelwise posteriors provided by the optical and SAR processing chains, conditioned on the common classes Ω_C and as a function of the multisensor feature vector $\underline{x}$ (see ATBD-v2).

$\mathcal{C}$ and $\mathcal{L}$ do not determine proper probability distributions *per se*, except in special cases (e.g., when the weights sum to one in the case of $\mathcal{C}$ or when they are uniform in the case of $\mathcal{L}$). However, both functionals can be normalized (linearly in the case of LOP and nonlinearly through a softmax operator in the case of LOGP) to derive a probability distribution $P_{\mathcal{F}}(\omega_j|\underline{x},\Omega_C)$, that expresses a pixelwise measure of uncertainty on the set of common classes. As proven in ATBD-v2, this measure of pixelwise uncertainty is extended to the whole set of classes as:

$$P_{\mathcal{F}}(\omega_j|\underline{x}) = P_{\mathcal{F}}(\omega_j|\underline{x},\Omega_C)[\lambda P(\Omega_C|\underline{O},\Omega_O \cup \Omega_C) + (1-\lambda)P(\Omega_C|\underline{S},\Omega_S \cup \Omega_C)]$$
$$+ \lambda P(\omega_j|\underline{O},\Omega_O)P(\Omega_O|\underline{O},\Omega_O \cup \Omega_C) + (1-\lambda)P(\omega_j|\underline{S},\Omega_S)P(\Omega_S|\underline{S},\Omega_S \cup \Omega_C),$$

where purple terms result from consensus-theoretic fusion on the common classes, blue terms are derived from the output of the optical chain and regard the optical-exclusive classes, red terms are similarly associated with SAR-exclusive classes, and $\lambda \in [0,1]$ is a weight computed as a function of the prior probabilities. The resulting $P_{\mathcal{F}}(\omega_j|x)$ yields a probability distribution that expresses a pixelwise measure of uncertainty after the consensus processing stage.

6.2 Uncertainty in Markov Random Fields

Markov random fields (MRFs) are probabilistic graphical models able to include contextual information in the form of class interactions between neighbouring pixels. As discussed in the ATBD-v2, an MRF is determined by an energy function, whose minimization with respect to the labels is equivalent to the application of the maximum *a-posteriori* (MAP) criterion [18]:

$$Y^{MAP} = \underset{Y}{\operatorname{argmax}}\, P(Y|X),$$

where Y and X indicate the random fields of all class labels and feature vectors, respectively, across the whole pixel grid I. In particular, the Hammersley-Clifford theorem specifies the relation between such energy $U(\cdot)$ and the posterior probability $P(Y|X)$:

$$P(Y|X) = \frac{1}{Z}\exp\big(-U(Y|X)\big), \qquad Z = \sum_{Y} \exp\big(-U(Y|X)\big),$$

where Z is the normalization constant (named partition function). Considering the MRF model in which only up to pairwise clique potentials are non-zero, then the energy can be written as:

$$U(Y|X) = -\sum_{i\in I} \alpha \log P_{\mathcal{F}}(y_i|\underline{x}_i) - \gamma \sum_{\substack{i\in I \\ j\in\partial i}} \delta(y_i, y_j),$$

where $y_i \in \Omega$ is the class label of the ith pixel, $\underline{x}_i$ is its feature vector, ∂i is its neighborhood, $P_{\mathcal{F}}(y_i|\underline{x}_i)$ is derived from pixelwise fusion (see above), α and γ are weight coefficients, $\delta(\cdot)$ is the Kronecker impulse, the first summation considers pixelwise contributions, and the second one represents the pairwise interactions.

The uncertainty associated with the class labels predicted according to such an MRF model can be computed as a function of the corresponding energy. Indeed, using the aforementioned Hammersley-Clifford theorem yields the global posterior probability $P(Y|X)$ that is not a pixel-wise measure of uncertainty and is generally hard to compute because the partition function Z is intractable except in special cases [18]. However, the local contextual pixelwise probability $P\left(y_i\middle|\underline{x}_i,\{y_j\}_{j\in\partial i}\right)$, i.e., the distribution of the class label of each pixel, conditioned to its observations from all sources and to the labels of the neighbouring pixels, is easily derived from the energy [18] and provides a spatial-contextual measure of uncertainty of the predicted land cover. In principle, a further possible measure of spatial-contextual pixel-wise uncertainty would be $P(y_i|X)$, i.e., the probability distribution of the label of each pixel, conditioned to all image observations used to compute prediction. However, the calculation, estimation, and optimization of $P(y_i|X)$ corresponds to the use of the marginal *a-posteriori* modes (MPM) criterion to MRF-based classification rather than to the MAP criterion. On one hand, MPM formulations for MRF-based classifiers are computationally convenient in the case of multiscale quadtree graphs. On the other hand, they are remarkably time-expensive in the case of planar graphs because of the need to iteratively run time-consuming stochastic samplers (Gibbs or Metropolis sampling) [19]. Accordingly, the use of this MPM-based uncertainty measure is deemed substantially disadvantageous in the HRLC pipeline. More generally, attention is being devoted to identifying the most appropriate uncertainty measure in the Markovian case in a compromise between accuracy and execution time.

6.3 Uncertainty with Deep Learning

Several formulations involving deep learning are discussed in the ATBD-v2 with regard to classification and fusion stages. Deep neural networks [20] allow the computation of pixelwise uncertainty measures. The output of the last layer may normally be interpreted probabilistically by using softmax activation functions. Let $\sigma(\boldsymbol{z})$ be the output quantities, where $\sigma(\cdot)$ is the softmax function and the vector $\boldsymbol{z} = [z_1, z_2, \dots, z_C]$ collects the inputs resulting from the last hidden layer, with C being the number of classes. The softmax output is computed as:

$$\sigma(\boldsymbol{z})_i = \frac{e^{z_i}}{\sum_{j=1}^{C} e^{z_j}}, \qquad i = 1,2,\dots,C.$$

The softmax function takes as input a vector of real numbers and normalizes it into a new vector of numbers that can be interpreted as expressing a probability distribution associated with the predicted label. After applying the softmax, each component of the input vector will be in the interval $(0, 1)$, and the components will sum up to 1, so that they can be interpreted as probabilities. Accordingly, picking the class that yields the largest value of $\sigma(\boldsymbol{z})_i$ $(i = 1,2 \dots, C)$ is interpreted as a formulation of the MAP criterion.

Such consideration allows the deep learning formulations discussed in the ATBD-v2 to be automatically associated with pixelwise uncertainty measures by inspecting the output values of the softmax activation function, before thresholding such values in order to determine the class labels.

7 Multitemporal change detection and trend analysis

As the last part of the CCI HRLC processing chain, the multitemporal change detection and trend analysis is highly influenced by the uncertainties coming from previous steps. In particular, uncertainty is co-related to: i) the decision fusion step, ii) the classification maps and iii) the multisensor geolocation part. In consequence, the analysis done in previous steps applies in the same way for this last step.

Since products are developed at pixel-level (see ATBD for further details), uncertainty will be associated in the same way but with the three contributions mentioned before. More specifically, it will be associated to the three products from this step of the processing chain, being: i) change HRLC maps (30m), ii) inter-annual CD maps (30m) and iii) seasonal CD maps, when required/possible.

8 References

[1] J. Le Moigne, N. S. Netanyahu, and R. D. Eastman, Eds., *Image Registration for Remote Sensing*. Cambridge: Cambridge University Press, 2011.

[2] S. Savastano and R. Guida, “Uncertainty Quantification in Synthetic Aperture Radar Remote Sensing Data Processing,” in *IGARSS 2018 - 2018 IEEE International Geoscience and Remote Sensing Symposium*, 2018, pp. 9193–9196, doi: 10.1109/IGARSS.2018.8518217.

[3] J. C. Curlander and R. N. McDonugh, *Synthetic Aperture Radar: Systems and Signal Processing*. 1991.

[4] T. L. Logan, “Terrain Correction of Synthetic Aperture Radar Imagery Using the Cray T3D,” 1995.

[5] C. E. Wivell, C. Olmsted, D. R. Steinwand, and C. S. Taylor, “An earth remote sensing satellite- 1 Synthetic Aperture Radar Mosaic of the Tanana River Basin in Alaska,” 1993.

[6] T. Bayer, R. Winter, and G. Schreier, “Terrain influences in SAR backscatter and attempts to their correction,” *IEEE Transactions on Geoscience and Remote Sensing*, vol. 29, no. 3, pp. 451–462, May 1991, doi: 10.1109/36.79436.

[7] T. Lillesand, R. W. Kiefer, and J. Chipman, “Remote Sensing and Image Interpretation, 7th Edition,” *Wiley.com*, 2015. [Online]. Available: https://www.wiley.com/en-it/Remote+Sensing+and+Image+Interpretation%2C+7th+Edition-p-9781118343289. [Accessed: 02-Jul-2019].

[8] Jong-Sen Lee, M. R. Grunes, and G. de Grandi, “Polarimetric SAR speckle filtering and its implication for classification,” *IEEE Transactions on Geoscience and Remote Sensing*, vol. 37, no. 5, pp. 2363–2373, Sep. 1999, doi: 10.1109/36.789635.

[9] G. Lisini, A. Salentinig, P. Du, and P. Gamba, “SAR-Based Urban Extents Extraction: From ENVISAT to Sentinel-1,” *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, vol. 11, no. 8, pp. 2683–2691, Aug. 2018, doi: 10.1109/JSTARS.2017.2782180.

[10] J. Le Moigne, N. S. Netanyahu, and R. D. Eastman, Eds., *Image Registration for Remote Sensing*. Cambridge: Cambridge University Press, 2011.

[11] I. Zavorin and J. Le Moigne, “Use of multiresolution wavelet feature pyramids for automatic registration of multisensor imagery,” *IEEE Transactions on Image Processing*, vol. 14, no. 6, pp. 770–782, Giugno 2005, doi: 10.1109/TIP.2005.847287.

[12] S. Ahmed and Md. M. Rahman, “Handling Uncertainty under Spatial Feature Extraction through Probabilistic Shape Model (PSM),” *International Journal of Science and Research (IJSR)*, vol. 9, no. 2, Sep. 2013.

[13] L. Loosvelt *et al.*, “Random Forests as a tool for estimating uncertainty at pixel-level in SAR image classification,” *International Journal of Applied Earth Observation and Geoinformation*, vol. 19, pp. 173–184, Oct. 2012, doi: 10.1016/j.jag.2012.05.011.

[14] F. Löw, P. Knöfel, and C. Conrad, “Analysis of uncertainty in multi-temporal object-based classification,” *ISPRS Journal of Photogrammetry and Remote Sensing*, vol. 105, pp. 91–106, Jul. 2015, doi: 10.1016/j.isprsjprs.2015.03.004.

[15] L.-H. Hsiao and K.-S. Cheng, “Assessing Uncertainty in LULC Classification Accuracy by Using Bootstrap Resampling,” *Remote Sensing*, vol. 8, no. 9, p. 705, Sep. 2016, doi: 10.3390/rs8090705.

[16] J. A. Benediktsson, “Hybrid consensus theoretic classification,” *IEEE Transactions on Geoscience and Remote Sensing*, vol. 35, no. 4, pp. 833–843, Jul. 1997, doi: 10.1109/36.602526.

[17] J. A. Benediktsson and P. H. Swain, “Consensus theoretic classification methods,” *IEEE Transactions on Systems, Man, and Cybernetics*, vol. 22, no. 4, pp. 688–704, 1992, doi: 10.1109/21.156582.

[18] S. Z. Li, *Markov random field modeling in image analysis*. Springer, 2009.

[19] I. Hedhli, G. Moser, J. Zerubia, and S. B. Serpico, “A New Cascade Model for the Hierarchical Joint Classification of Multitemporal and Multiresolution Remote Sensing Data,” *IEEE Transactions on Geoscience and Remote Sensing*, vol. 54, no. 11, pp. 6333–6348, Nov. 2016, doi: 10.1109/TGRS.2016.2580321.

[20] I. Goodfellow, Y. Bengio, and A. Courville, *Deep learning*. .