AutoCloud+, a "universal" physical and statistical model-based 2D spatial topology-preserving software toolbox for cloud/cloud-shadow detection in multi-sensor single-date Earth observation multi-spectral imagery eligible for systematic ESA EO Level 2 product generation

Andrea Baraldi [1,2,*] and Dirk Tiede [1]

[1] Department of Geoinformatics – Z_GIS, University of Salzburg, Schillerstr. 30, 5020 Salzburg, Austria; dirk.tiede@sbg.ac.at

[2] Italian Space Agency (ASI), Via del Politecnico, 00133 Rome RM, Italy.

* Correspondence: andrea6311@gmail.com; Tel.: +39-329-2648110

Abstract: The European Space Agency (ESA) defines as Earth observation (EO) Level 2 information product a single-date multi-spectral (MS) image corrected for atmospheric, adjacency and topographic effects, stacked with its data-derived scene classification map (SCM), whose thematic map legend includes quality layers cloud and cloud-shadow. ESA EO Level 2 product generation is an inherently ill-posed computer vision (CV) problem never accomplished to date in operating mode by any EO data provider at the ground segment. Herein, it is considered: (I) necessary not sufficient pre-condition for the yet-unaccomplished dependent problems of semantic content-based image retrieval (SCBIR) and semantics-enabled information/knowledge discovery (SEIKD) in multi-source EO *big data* cubes. (II) Synonym of EO Analysis Ready Data (ARD) format. (III) Equivalent to a horizontal policy for background developments in Space Economy 4.0. In compliance with the GEO-CEOS Quality Assurance Framework for EO Calibration/Validation guidelines, to contribute toward filling an analytic and pragmatic information gap from multi-sensor EO *big data* to timely, comprehensive and operational EO value-adding information products and services, this work presents an innovative AutoCloud+ CV software toolbox for cloud and cloud-shadow quality layer detection in ESA EO Level 2 product. In vision, spatial information dominates color information. Inspired by this true-fact, the inherently ill-posed AutoCloud+ CV software was conditioned, designed and implemented to be "universal", meaning fully automated (no human-machine interaction is required), near real-time, robust to changes in input data and scalable to changes in MS imaging sensor's spatial and spectral resolution specifications.

Keywords: artificial intelligence; color naming; color constancy; cognitive science; computer vision; object-based image analysis (OBIA); physical and statistical data models; radiometric calibration; semantic content-based image retrieval; spatial topological and spatial non-topological information components.

Acronyms

AI: Artificial Intelligence

AI4DIAS: Artificial Intelligence for Data and Information Access Services (at the ground segment)

45   AI4Space: Artificial Intelligence for Space (segment)
46   ARD: Analysis Ready Data (format)
47   ATCOR: Atmospheric/Topographic Correction commercial sofwtare product
48   AVHRR: Advanced Very High Resolution Radiometer
49   BC: Basic Color
50   BIVRFTAB: Bivariate Frequency Table
51   Cal: Calibration
52   Cal/Val: Calibration and Validation
53   CBIR: Content-Based Image Retrieval
54   CEOS: Committee on Earth Observation Satellites
55   CESBIO: Centre d'Etudes Spatiales de la Biosphère
56   CFMask: C (programming language version of) Function of Mask
57   CLC: CORINE Land Cover (taxonomy)
58   CNES: Centre national d'études spatiales
59   CNN: Convolutional Neural Network
60   CORINE: Coordination of Information on the Environment
61   CV: Computer Vision
62   DCNN: Deep Convolutional Neural Network
63   DEM: Digital Elevation Model
64   DIAS: Data and Information Access Services
65   DLR: Deutsches Zentrum für Luft- und Raumfahrt (German Aerospace Center)
66   DN: Digital Number
67   DP: Dichotomous Phase (in the FAO LCCS taxonomy)
68   DRIP: Data-Rich, Information-Poor (syndrome)
69   EO: Earth Observation
70   EO-IU: EO Image Understanding
71   EO-IU4SQ: EO Image Understanding for Semantic Querying
72   ESA: European Space Agency
73   FAO: Food and Agriculture Organization
74   FIEOS: Future Intelligent EO imaging Satellites
75   FMask: Function of Mask
76   GEO: Intergovernmental Group on Earth Observations
77   GEOSS: Global EO System of Systems
78   GIGO: Garbage In, Garbage Out principle of error propagation
79   GIS: Geographic Information System
80   GIScience: Geographic Information Science
81   GUI: Graphic User Interface
82   IGBP: International Global Biosphere Programme
83   IU: Image Understanding
84   LAI: Leaf Area Index
85   LC: Land Cover
86   LCC: Land Cover Change
87   LCCS: Land Cover Classification System (taxonomy)
88   LCLU: Land Cover Land Use
89   LEDAPS: Landsat Ecosystem Disturbance Adaptive Processing System
90   MAACS: Multisensor Atmospheric Correction and Cloud Screening
91   MAJA: Multisensor Atmospheric Correction and Cloud Screening (MACCS)-
92   Atmospheric/Topographic Correction (ATCOR) Joint Algorithm
93   mDMI: Minimally Dependent and Maximally Informative (set of quality indicators)
94   MHP: Modular Hierarchical Phase (in the FAO LCCS taxonomy)
95   MIR: Medium InfraRed
96   MODIS: Moderate Resolution Imaging Spectroradiometer

97   MS: Multi-Spectral
98   MSI: (Sentinel-2) Multi-Spectral Instrument
99   NASA: National Aeronautics and Space Administration
100   NIR: Near InfraRed
101   NLCD: National Land Cover Data
102   NOAA: National Oceanic and Atmospheric Administration
103   NP: Non-Polynomial
104   OBIA: Object-Based Image Analysis
105   OGC: Open Geospatial Consortium
106   OP: Outcome (product) and Process
107   OP-Q^{2}I: Outcome and Process Quantitative Quality Index
108   QA4EO: Quality Accuracy Framework for Earth Observation
109   Q^{2}I: Quantitative Quality Indicator
110   RGB: monitor-typical Red-Green-Blue data cube
111   RMSE: Root Mean Square Error
112   RS: Remote Sensing
113   RTD: Research and Technological Development
114   SCBIR: Semantic Content-Based Image Retrieval
115   SCM: Scene Classification Map
116   SEIKD: Semantics-Enabled Information/Knowledge Discovery
117   Sen2Cor: Sentinel 2 (atmospheric) Correction Prototype Processor
118   SIAM™: Satellite Image Automatic Mapper™
119   STRATCOR: Stratified Topographic Correction
120   SURF: Surface Reflectance
121   TIR: Thermal InfraRed
122   TM (superscript): (non-registered) Trademark
123   TMask: Temporal Function of Mask
124   TOA: Top-Of-Atmosphere
125   TOARF: TOA Reflectance
126   UAV: Unmanned Aerial Vehicle
127   UML: Unified Modeling Language
128   USGS: US Geological Survey
129   Val: Validation
130   VAPS: Value-Adding information Products and Services
131   VQ: Vector Quantization
132   WGCV: Working Group on Calibration and Validation

133   **1. Introduction**

134   Radiometric calibration (*Cal*) is the process of transforming remote sensing (RS) sensory
135   data, consisting of non-negative dimensionless digital numbers (DNs, with DN ≥ 0) provided
136   with no physical meaning, into a physical variable provided with a community-agreed
137   radiometric unit of measure, such as top-of-atmosphere reflectance (TOARF), surface
138   reflectance (SURF) or surface albedo values in range [0.0, 1.0] [1-3].
139   To cope with the five Vs characterizing *big data* analytics, specifically, volume, variety,
140   veracity, velocity and value [4], radiometric *Cal* of Earth observation (EO) *big data* is considered
141   mandatory by the intergovernmental Group on Earth Observations (GEO)-Committee on
142   Earth Observation Satellites (CEOS) Quality Accuracy Framework for Earth Observation
143   (QA4EO) Calibration/Validation (*Cal/Val*) guidelines [3]. In agreement with the yet-
144   unaccomplished visionary goal of a GEO's implementation plan for years 2005-2015 of a Global
145   Earth Observation System of Systems (GEOSS) [5], the ambitious goal of the GEO-CEOS
146   QA4EO *Cal/Val* guidelines is systematic transformation of EO *big data* cubes into timely,
147   comprehensive and operational EO value-adding information products and services (VAPS).

148 Despite being considered a well-known "prerequisite for physical model-based analysis of
149 airborne and satellite sensor measurements in the optical domain" [1], EO data radiometric *Cal*
150 is largely oversighted in the RS common practice. A lack of EO input data *Cal* requirements
151 means that statistical model-based data analytics and inductive learning-from-data algorithms
152 are dominant in the RS community, including (geographic) object-based image analysis
153 (GEOBIA) applications [6, 7] in the domain of geographic information science (GIScience). On
154 the one hand, statistical model-based and inductive learning-from-data algorithms require as
155 input DNs provided with no physical meaning. On the other hand, inductive learning-from-
156 data algorithms are inherently semi-automatic and site-specific [2]. In practice, they require no
157 radiometric *Cal* data pre-processing, but they typically gain in robustness when input with
158 radiometrically calibrated data.
159   In compliance with the GEO-CEOS QA4EO *Cal/Val* requirements and with the GEO's
160 visionary goal of a GEOSS, the European Space Agency (ESA) has recently defined an ESA EO
161 Level 2 information product as follows [8, 9].
162 (i)   A single-date multi-spectral (MS) image, radiometrically corrected for atmospheric,
163 adjacency and topographic effects,
164 (ii)   stacked with its data-derived scene classification map (SCM), whose general-purpose,
165 user- and application-independent thematic map legend includes quality layers cloud and
166 cloud-shadow,
167 (iii)   to be systematically generated at the ground segment, automatically (without human-
168 machine interaction) and in near real-time.
169
170   **Table 1.** Non-standard general-purpose, user- and application-independent ESA EO Level 2
171   SCM legend adopted by the sensor-specific Sentinel 2 (atmospheric, adjacency and
172   topographic) Correction (Sen2Cor) Prototype Processor [8, 9], developed and distributed free-
173   of-cost by ESA to be run on user side.

Label	Classification
0	NO_DATA
1	SATURATED_OR_DEFECTIVE
2	DARK_AREA_PIXELS
3	CLOUD_SHADOWS
4	VEGETATION
5	BARE_SOILS
6	WATER
7	CLOUD_LOW_PROBABILITY
8	CLOUD_MEDIUM_PROBABILITY
9	CLOUD_HIGH_PROBABILITY
10	THIN_CIRRUS
11	SNOW

174
175
176   **Table 2.** Standard general-purpose, user- and application-independent ESA Level 2 SCM
177   legend proposed in [10-12]. "Augmented" 3-level 8-class FAO LCCS Dichotomous Phase (DP)
178   taxonomy (identified as classes A11 to B48) + quality layers Cloud and Cloud-shadow + Others
179   = 8 land cover (LC) classes + 2 LC classes (Cloud-shadow, Others) + 1 non-LC class (Cloud).

		Pseudocolor
A11	1. Cultivated and Managed Terrestrial (non-aquatic) Vegetated Areas	
A12	2. Natural and Semi-Natural Terrestrial Vegetation	
A23	3. Cultivated Aquatic or Regularly Flooded Vegetated Areas	
A24	4. Natural and Semi-Natural Aquatic or Regularly Flooded Vegetation	
B35	5. Artificial Surfaces and Associated Areas	
B36	6. Bare Areas	
B47	7. Artificial Waterbodies, Snow and Ice	
B48	8. Natural Waterbodies, Snow and Ice.	
	9. Quality layer: Cloud	
	10. Quality layer: Cloud-shadow	
	11. Others	

180
181   Unlike the non-standard ESA EO Level 2 SCM legend adopted by the Sentinel 2 imaging
182   sensor-specific (atmospheric, adjacency and topographic) Correction Prototype Processor
183   (Sen2Cor), developed by ESA and distributed free-of-cost to be run on user side [8, 9], see Table
184   1, an alternative ESA EO Level 2 SCM legend proposed in [10-12], shown in Table 2, is selected
185   equal to an "augmented" 3-level 9-class Dichotomous Phase (DP) taxonomy of the Food and
186   Agriculture Organization of the United Nations (FAO) – Land Cover Classification System
187   (LCCS) [13], see Figure 1. Such an "augmented" land cover (LC) class taxonomy in the 4D
188   geospatial-temporal scene-domain encompasses the standard fully-nested 3-level 8-class FAO
189   LCCS-DP legend, , see Figure 1, in addition to a thematic layer "other" or "rest of the world",
190   which includes quality layers cloud and cloud-shadow. In traditional EO image classification
191   system design and implementation requirements [14], the presence of output class "unknown"
192   is considered mandatory to cope with uncertainty in inherently equivocal *information-as-data-*
193   *interpretation* (classification) tasks [15].
194

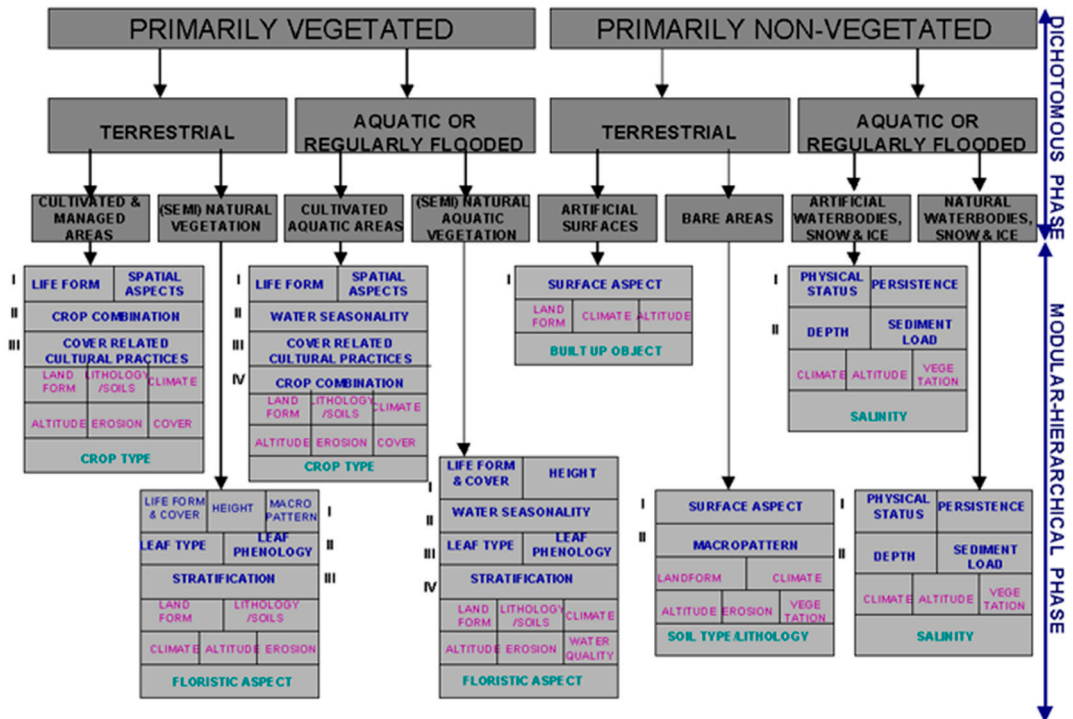



Figure 1. Two-stage fully-nested FAO Land Cover Classification System (LCCS) taxonomy
        [13]. The first-stage fully-nested 3-level 8-class FAO LCCS Dichotomous Phase (DP) taxonomy
        is general-purpose, user- and application-independent. It consists of a sorted set of 3
        dichotomous layers: (i) vegetation versus non-vegetation, (ii) terrestrial versus aquatic, and
        (iii) managed versus natural or semi-natural. These 3 dichotomous layers deliver as output the
        following 8-class FAO LCCS-DP taxonomy. (A11) Cultivated and Managed Terrestrial (non-

201 aquatic) Vegetated Areas. (A12) Natural and Semi-Natural Terrestrial Vegetation. (A23)
202 Cultivated Aquatic or Regularly Flooded Vegetated Areas. (A24) Natural and Semi-Natural
203 Aquatic or Regularly Flooded Vegetation. (B35) Artificial Surfaces and Associated Areas. (B36)
204 Bare Areas. (B47) Artificial Waterbodies, Snow and Ice. (B48) Natural Waterbodies, Snow and
205 Ice. The general-purpose user- and application-independent 3-level 8-class FAO LCCS-DP
206 taxonomy is preliminary to a second-stage FAO LCCS Modular Hierarchical Phase (MHP)
207 taxonomy, consisting of a battery of user- and application-specific one-class classifiers,
208 equivalent to one-class grammars (syntactic classifiers) [16].

209 Shown in Figure 1, the standard two-phase fully-nested FAO LCCS hierarchy consists a
210 first-stage fully-nested general-purpose, user- and application-independent 3-level 8-class
211 FAO LCCS-DP legend, preliminary to a second-stage application-dependent and user-specific
212 FAO LCCS Modular Hierarchical Phase (MHP) taxonomy, consisting of a hierarchical (deep)
213 battery of one-class classifiers [13]. The standard first-stage 3-level 8-class FAO LCCS-DP
214 hierarchy is "fully nested". It comprises three dichotomous LC class-specific information
215 layers, equivalent to a world ontology, world model or mental model of the world [10, 13, 16-
216 21]: DP Level 1 – Vegetation versus non-vegetation, DP Level 2 – Terrestrial versus aquatic and
217 DP Level 3 – Managed versus natural or semi-natural. In recent years, the two-phase FAO
218 LCCS taxonomy has become increasingly popular [22]. One reason of its popularity is that the
219 FAO LCCS hierarchy is "fully nested" while alternative LC class hierarchies, such as the
220 Coordination of Information on the Environment (CORINE) Land Cover (CLC) taxomony [23],
221 the U.S. Geological Survey (USGS) Land Cover Land Use (LCLU) taxonomy by J. Anderson
222 [24], the International Global Biosphere Programme (IGBP) DISCover Data Set Land Cover
223 Classification System [25] and the EO Image Librarian LC class legend [26], start from a Level
224 1 taxonomy which is already multi-class. In a hierarchical EO image understanding (EO-IU)
225 system architecture submitted to a garbage in, garbage out (GIGO) information principle,
226 synonym of error propagation through an information processing chain, the fully-nested two-
227 phase FAO LCCS hierarchy makes explicit the full dependence of high-level LC class estimates,
228 performed by any high-level (deep) LCCS-MHP data processing module, on the operational
229 quality (in accuracy, efficiency, robustness, etc.) of lower-level LCCS modules, starting from
230 the initial FAO LCCS-DP Level 1 vegetation/non-vegetation information layer whose relevance
231 in thematic mapping accuracy (vice versa, in error propagation) becomes paramount for all
232 subsequent LCCS layers. The GIGO commonsense principle applied to hierarchical semantic
233 dependence is neither trivial nor obvious to underline [27]. On the one hand, it agrees with a
234 minor portion of the RS literature where supervised data learning classification of EO image
235 datasets at continental or global spatial extent into binary LC class vegetation/non-vegetation
236 is considered very challenging [28]. On the other hand, it is at odd with the RS mainstream,
237 where the semantic information gap from sub-symbolic EO data to multi-class LC taxonomies
238 is typically filled in one step, implemented as a supervised data learning classifier [29, 30], e.g.,
239 a support vector machine, random forest or deep convolutional neural network (DCNN) [31],
240 which is equivalent to an unstructured black box [27], inherently semi-automatic and site-
241 specific [2] and whose opacity contradicts the well-known engineering principles of
242 modularity, regularity and hierarchy typical of scalable systems [32]. "No Free Lunch"
243 theorems have shown that inductive learning-from-data algorithms cannot be universally
244 good [33, 34].

245 Over land surfaces of the Earth, the global cloud cover is approximately 66% [35]. In the
246 ESA EO Level 2 product definition, cloud and cloud-shadow quality layer requirements
247 specification accounts for a well-known prerequisite of clear-sky multi-temporal EO image
248 compositing and understanding (classification) solutions proposed by the RS community,
249 where accurate masking of cloud and cloud-shadow phenomena is considered necessary not
250 sufficient pre-condition [9, 10, 36-56]. Intuitively, in single-date and multi-temporal MS image
251 analysis, cloud and cloud-shadow preliminary detection is a relevant problem, because
252 unflagged cloud and cloud-shadow phenomena may be mapped onto erroneous LC classes or

253   false LC change (LCC) occurrences. Noteworthy, cloud/cloud-shadow detection is a typical
254   example of physical model-based cause-effect relationship, expected to be very difficult to
255   solve by inductive machine learning-from-data algorithms, such as increasingly popular
256   DCNNs [31], suitable for learning complex correlations between input and output features, but
257   capable of no inherent representation of causality [27, 57], in agreement with the well-known
258   dictum that correlation does not imply causation and vice versa [10, 16, 27, 30, 57, 58].

259   In the last decade, many different cloud/cloud-shadow detection algorithms were
260   presented in the RS literature to run either on a single-date MS image or on an MS image time-
261   series, typically acquired by either one EO spaceborne/airborne MS imaging sensor or a single
262   family (e.g., Landsat) of MS imaging sensors [9, 10, 36-53]. To be accomplished in operating
263   mode at the ground segment (midstream) by EO data providers in support of the downstream
264   sector in an era of Space Economy 4.0 [59], systematic radiometric *Cal* of multi-source multi-
265   angular MS big image data cubes [1, 3, 10, 60-62], encompassing either single-date or multi-
266   temporal cloud and cloud-shadow detection as a necessary not sufficient pre-condition, is
267   regarded by the RS community as an open problem to date [54-56].

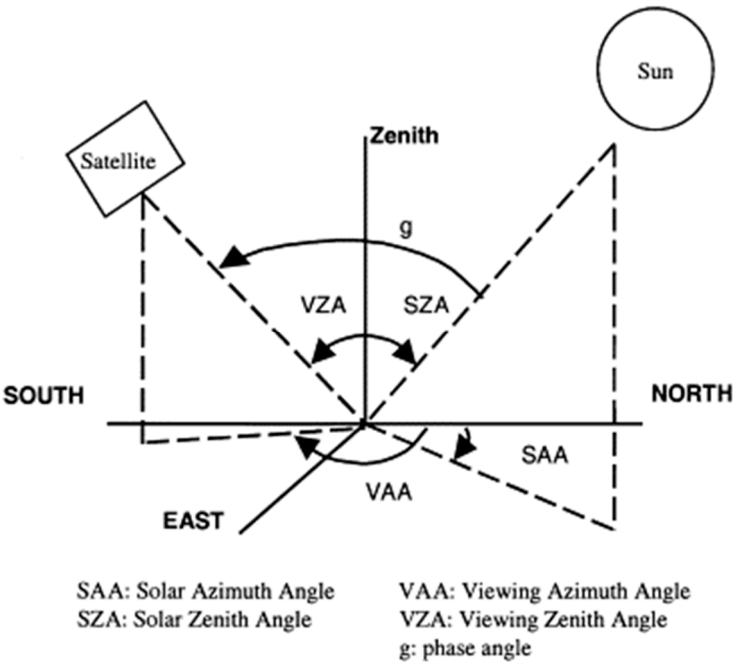
268   To overcome conceptual limitations and well-known failure modes of existing cloud and
269   cloud-shadow detection algorithms [37, 40, 54-56], such as the single-date open source
270   Function of Mask (FMask) algorithm [51, 52], the free-of-cost single-date ESA Sen2Cor software
271   toolbox to be run on user side [8, 9, 37] and the multi-date Multisensor Atmospheric Correction
272   and Cloud Screening (MACCS)-Atmospheric/Topographic Correction (ATCOR) Joint
273   Algorithm (MAJA) developed by the Centre national d'études spatiales (CNES)/ Centre
274   d'Etudes Spatiales de la Biosphère (CESBIO)/ Deutsches Zentrum für Luft- und Raumfahrt
275   (German Aerospace Center, DLR) [39-41], which incorporates capabilities of the ATCOR
276   commercial software toolbox [63-66], this work presents an innovative AutoCloud+ computer
277   vision (CV) software toolbox for cloud and cloud-shadow quality layer detection, suitable for
278   systematic ESA EO Level 2 product generation at the ground segment [36, 60, 67]), in
279   compliance with the GEO-CEOS QA4EO *Cal/Val* requirements [3].

280   Synonym of inherently ill-posed scene-from-image reconstruction and understanding [10,
281   20, 68, 69]), vision is a cognitive (*information-as-data-interpretation*) process [15], encompassing
282   both biological vision and CV, where CV is *subset-of* artificial (general) intelligence (AI) [70-74],
283   i.e., AI ⊃ CV. In vision, spatial information dominates color information [20]. This
284   unquestionable true-fact is familiar to all human beings wearing sunglasses: in perceptual
285   terms, human panchromatic and chromatic visions are nearly as effective [10]. Based on this
286   observation, an inherently ill-posed AutoCloud+ CV subsystem was constrained, designed and
287   implemented to be "universal" in operating mode, i.e., automated (no human-machine
288   interaction is required to run), near real-time, robust to changes in input data and scalable to
289   cope with changes in MS imaging sensor's spatial and spectral resolution specifications.

290   The rest of this paper is organized as follows. Provided with a relevant survey value,
291   Section 2 reviews the interdisciplinary problem background of systematic ESA EO Level 2
292   product generation, regarded as necessary not sufficient pre-condition for a yet-
293   unaccomplished GEOSS development. Related works in cloud/cloud-shadow detection are
294   critically revised in Section 3. Adopted AutoCloud+ methods and materials are proposed in
295   Section 4 and 5 respectively. Collected results are presented in Section 6 and discussed in
296   Section 7. Conclusions are reported in Section 8.

2. Problem background

298   Featuring a relevant survey value, this Section aims at providing a critical review of the
299   interdisciplinary problem background of systematic ESA EO Level 2 product generation.
300   According to our original working hypothesis, to be regarded as our first original contribution,
301   systematic ESA EO Level 2 product generation is postulated as necessary not sufficient pre-
302   condition for a yet-unaccomplished GEOSS development, in agreement with the GEO-CEOS
303   QA4EO *Cal/Val* requirements [3]. In this critical review, our second original contribution

304   highlights the several degrees of novelty of the ESA EO Level 2 product definition (refer to
305   Section 1) in comparison with, for example, alternative EO Analysis Ready Data (ARD) format
306   definitions proposed by the RS community. Next, as our third original contribution, structural
307   limitations in system design of existing open source or commercial software products for EO
308   image radiometric *Cal* (correction) are underlined, to be overcome by an alternative system
309   architecture.
310

311   **Figure 2.** Solar illumination geometries and viewpoint geometries in spaceborne and airborne
312   EO image acquisition.

313   To successfully cope with the five Vs of EO *big data* analytics, specifically, volume, variety,
314   veracity, velocity and value [4], multi-sensor analysis of multi-temporal multi-angular EO
315   sensory data cubes depends upon the ability to distinguish between relevant changes and no-
316   changes occurring at the Earth surface through time [61, 62]. Necessary not sufficient pre-
317   condition for EO *big data* transformation into timely, comprehensive and operational EO data-
318   derived VAPS, expected by GEO to be pursued by a GEOSS never accomplished to date [5], is
319   radiometric *Cal*, considered mandatory by the GEO-CEOS QA4EO *Cal/Val* requirements [3]. In
320   short, radiometric *Cal* guarantees EO sensory data interoperability (consistency,
321   "normalization") through time, geographic space and sensors. In greater detail, the capability
322   to detect and quantify change/no-change in terms of either qualitative (nominal, categorical)
323   Earth surface variables, such as LC classes belonging to a finite and discrete LC class taxonomy
324   (legend), or quantitative (numeric) Earth surface variables, such as biophysical variables, e.g.,
325   leaf area index (LAI), biomass, etc. [75], depends on the radiometric *Cal* of EO sensory data,
326   equivalent to non-negative dimensionless DNs ≥ 0, provided with no physical meaning and
327   typically affected at sensor-level by ever-varying atmospheric conditions, solar illumination
328   conditions, spaceborne/airborne viewing geometries and Earth surface topography, into a
329   physical variable provided with a community-agreed radiometric unit of measure, such as
330   SURF or surface albedo values in range [0, 1] [1-3]. Solar illumination conditions are typically
331   parameterized by metadata *Cal* parameters, such as image acquisition time, solar
332   exoatmospheric irradiance, solar zenith angle and solar azimuth angle, see Figure 2. Sensor
333   viewing characteristics are typical metadata *Cal* parameters, such as sensor zenith angle and
334   sensor azimuth angle, see Figure 2. Atmospheric conditions are described by categorical
335   variables, such as aerosol type, haze, cloud and cloud-shadow, and by numeric variables, such
336   as water vapor, temperature and aerosol optical thickness (AOT) [61, 62]. Finally, Earth surface

337 geometries must be inferred from ancillary data, such as a digital elevation model (DEM), in
338 combination with solar and viewing conditions [76], see Figure 2.
339   Adopted by different scientific disciplines, such as machine learning-from-data [30], AI as
340 *superset-of* CV [70], i.e., AI $\supset$ CV, and RS [2, 10], popular synonyms of deductive inference are
341 top-down inference, prior knowledge-based inference, learning-by-rule inference and physical
342 model-based inference. Synonyms of inductive inference are bottom-up inference, learning-
343 from-data inference, learning-from-examples inference and statistical model-based inference.
344   On the one hand, non-calibrated sensory data, provided with no physical meaning, can
345 be investigated by statistical data models and inductive inference algorithms, exclusively. On
346 the other hand, although they do not require physical variables as input, statistical data models
347 and inductive learning-from-data algorithms can benefit from input data *Cal* in terms of
348 augmented robustness to changes in the input data set acquired through time, space and
349 sensors. On the contrary, radiometrically calibrated data, provided with a physical meaning,
350 can be interpreted by either inductive, deductive (physical model-based) or hybrid (combined
351 deductive and inductive) inference algorithms. Although it is considered a well-known
352 "prerequisite for physical model-based (and hybrid) analysis of airborne and satellite sensor
353 measurements in the optical domain" [1, 10, 60-62], EO data *Cal* is largely neglected in the RS
354 common practice. For example, in major portions of the RS literature, including the GEOBIA
355 sub-domain of GIScience [6, 7], no reference to radiometric *Cal* issues is found. This lack of
356 input EO data *Cal* requirements proves that, to date, EO image analytics mainly consists of
357 inductive learning-from-data algorithms, starting from scratch because no *a priori* physical
358 knowledge is exploited in addition to data. This is in contrast with biological cognitive systems,
359 where "there is never an absolute beginning" [77], because *a priori* genotype provides initial
360 conditions (that reflect properties of the world, embodied through evolution, based on
361 evolutionary experience) to learning-from-examples phenotype, according to a hybrid
362 inference paradigm, where phenotype explores the neighborhood of genotype in a solution
363 space [10, 71]. Hybrid inference combines deductive and inductive inference to take advantage
364 of each and overcome their shortcomings [2]. Inductive inference is typically semi-automatic
365 and site-specific [2]. Deductive inference is static (non-adaptive to data) and typically lacks
366 flexibility to transform ever-varying sensory data (sensations) into stable percepts (concepts)
367 in a world model [10, 20, 78, 79].
368   In compliance with the GEO-CEOS QA4EO *Cal/Val* requirements and the visionary goal
369 of a GEOSS ([5]), ESA has recently provided an original ESA EO Level 2 information product
370 definition, refer to Section 1 [8, 9]. The ESA EO Level 2 product definition is non-trivial.
371 Noteworthy, it is more restrictive than the National Aeronautics and Space Administration
372 (NASA) EO Level 2 product definition of "a data-derived geophysical variable at the same
373 resolution and location as Level 1 source data" [80]. According to the standard Unified
374 Modeling Language (UML) for graphical modeling of object-oriented software [81], where
375 symbol '$\rightarrow$' denotes relationship *part-of* pointing from the supplier to the client (vice versa, it
376 would denote relationship *depend-on*), not to be confused with relationship *subset-of*, whose
377 symbol is '$\supset$', meaning specialization with inheritance from the superset to the subset,
378 dependence relationship 'NASA EO Level 2 product $\rightarrow$ ESA EO Level 2 product' holds true,
379 see Figure 3. It means that a NASA EO Level 2 product can be accomplished although no ESA
380 EO Level 2 product exists, whereas the vice versa does not hold.
381

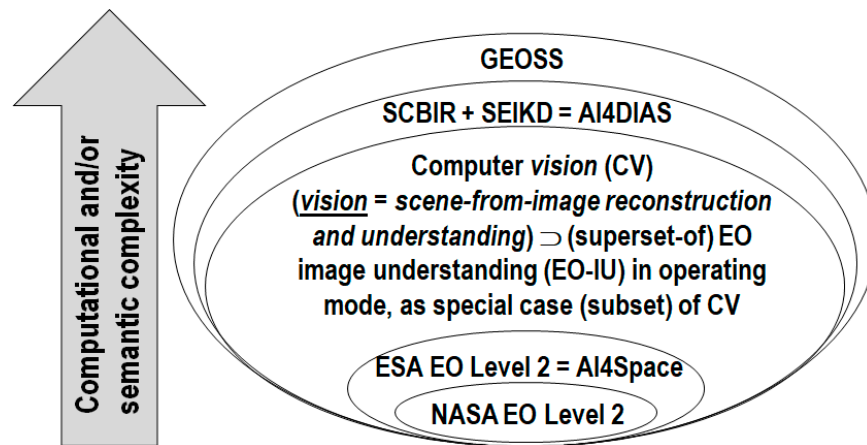
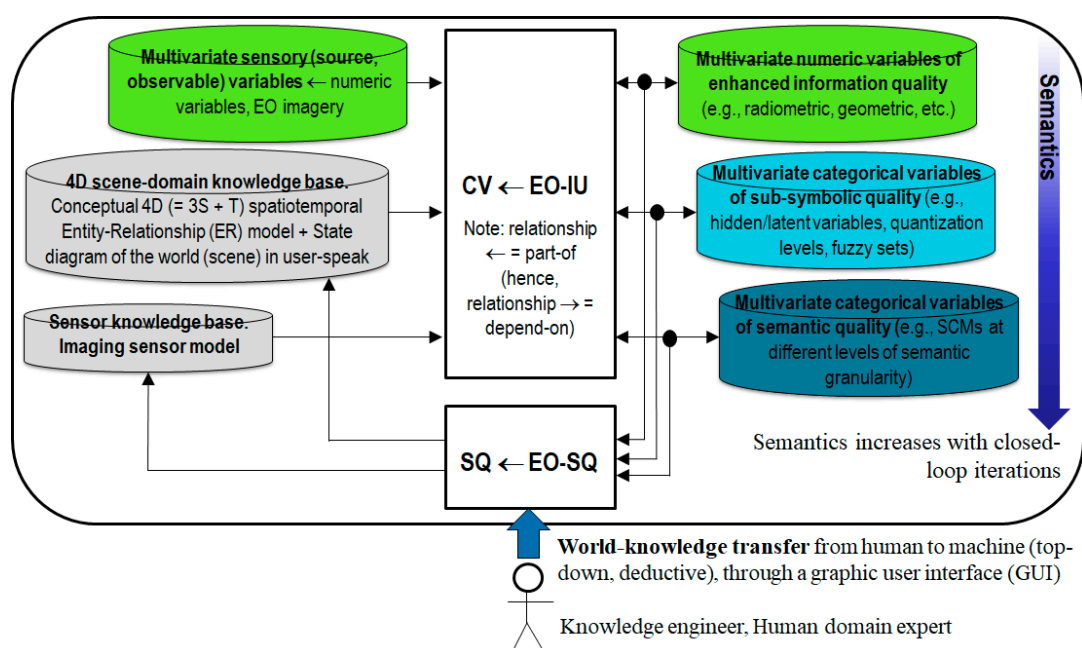



Figure 3. In agreement with the standard Unified Modeling Language (UML) for graphical modeling of object-oriented software ([81]), relationship *part-of*, denoted with symbol '→' pointing from the supplier to the client, should not to be confused with relationship *subset-of*, '⊃', meaning specialization with inheritance from the superset to the subset [81]. A NASA EO Level 2 product is defined as "a data-derived geophysical variable at the same resolution and location as Level 1 source data" [80]. Herein, it is considered *part-of* an ESA EO Level 2 product defined as [8, 9]: (a) a single-date multi-spectral (MS) image whose digital numbers (DNs) are radiometrically corrected into surface reflectance (SURF) values for atmospheric, adjacency and topographic effects, stacked with (b) its data-derived general-purpose, user- and application-independent scene classification map (SCM), whose thematic map legend includes quality layers cloud and cloud-shadow. In this paper, ESA EO Level 2 product is regarded as an information primitive to be accomplished by Artificial Intelligence for the Space segment (AI4Space), such as in future intelligent small satellite constellations, rather than at the ground segment in an AI for data and information access services (AI4DIAS) framework. In this graphical representation, additional acronyms of interest are computer vision (CV), whose special case is EO image understanding (EO-IU) in operating mode, semantic content-based image retrieval (SCBIR) [10, 82-87], semantics-enabled information/knowledge discovery (SEIKD), where SCIR + SEIKD is considered synonym of AI4DIAS, and Global Earth Observation System of Systems (GEOSS), defined by the Group on Earth Observations [5]. Our working hypothesis postulates that 'NASA EO Level 2 product → ESA EO Level 2 product = AI4Space ⊂ EO-IU in operating mode ⊂ CV → [EO-SCBIR + SEIKD = AI4DIAS] → GEOSS'. It means that GEOSS, whose *part-of* are the still-unsolved (open) problems of SCBIR and SEIKD, cannot be achieved until the necessary not sufficient pre-condition of CV in operating mode, specifically, systematic ESA EO Level 2 product generation, is accomplished in advance. Encompassing both biological vision and CV, vision is synonym of scene-from-image reconstruction and understanding. Vision is a cognitive *(information-as-data-interpretation)* problem [15] very difficult to solve because: (i) non-polynomial (NP)-hard in computational complexity [88, 89], (ii) inherently ill-posed in the Hadamard sense [90], i.e., vision admits no solution, multiple solutions or, if the solution exists, the solution's behavior changes continuously with the initial conditions [20, 68]. Vision is inherently ill-posed because affected by: (I) a 4D-to-2D data dimensionality reduction from the scene-domain to the image-domain, e.g., responsible of occlusion phenomena, and (II) a semantic information gap from ever-varying sub-symbolic sensory data (sensations) in the physical world to stable symbolic percepts in the mental model of the physical world (modeled world, world ontology, world model) [10, 16-21]. Since it is inherently ill-posed, vision requires *a priori* knowledge in addition to sensory data to become better posed for numerical solution [30]. If the aforementioned working hypothesis holds true, then the complexity of SCBIR + SEIKD is not inferior to the complexity of vision, acknowledged to be inherently ill-posed and NP-hard. To make the inherently-ill-posed CV problem better conditioned for numerical solution, a CV system is required to comply with human visual perception. In other words, a CV system is constrained to include a computational model of human vision, i.e., 'Human vision → CV'. Hence,

423 dependence relationship: 'Human vision → CV ⊃ EO-IU in operating mode ⊃ NASA EO Level
424 2 product → ESA EO Level 2 product → [EO-SCBIR + SEIKD = AI4DIAS] → GEOSS' is our
425 working hypothesis.

426 An improved ("augmented") and more restrictive ESA EO Level 2 product definition
427 could be defined to account for bidirectional reflectance distribution function (BRDF) effect
428 correction, in addition to atmospheric, topographic and adjacency effects correction, to model
429 surface anisotropy in multi-temporal multi-angular EO image data cubes [2, 61, 62, 66, 91-94].
430 A surface that reflects the incident energy equally in all directions is said to be Lambertian,
431 where reflectance is invariant with respect to illumination and viewing conditions, see Figure
432 2. On the contrary, a surface is said to be anisotropic when its reflectance varies with respect
433 to illumination and/or viewing geometries. These changes are driven by the optical and
434 structural properties of the surface material. In other words, in EO image pre-processing
435 (enhancement) for radiometric *Cal*, BRDF effect correction is LC class-specific [61, 62, 66, 91-
436 94]. The LC class-specific task of BRDF correction is to derive, for non-Lambertian surfaces,
437 spectral albedo (bi-hemispherical reflectance, BHR) values, defined over all directions [2, 66,
438 91-94], from either SURF or TOARF values where the Lambertian surface assumption holds
439 [63-66, 91].
440 Our first original contribution is to consider an ESA EO Level 2 information product a
441 baseline information unit (information primitive) whose systematic generation is of
442 paramount importance to contribute toward filling an analytic and pragmatic information gap
443 from multi-sensor, multi-temporal and multi-angular EO *big image data cubes* into timely,
444 comprehensive and operational EO data-derived VAPS, in compliance with the yet-
445 unaccomplished visionary goal of a GEOSS [3, 5]. In Figure 3, our working hypothesis
446 postulates that systematic ESA EO Level 2 product generation is an inherently ill-posed CV
447 problem, where CV ⊃ EO-IU, whose solution in operating mode is necessary not sufficient pre-
448 condition for the yet-unaccomplished dependent problems of semantic content-based image
449 retrieval (SCBIR) [10, 82-87] and semantics-enabled information/knowledge discovery (SEIKD)
450 in large-scale EO image data cubes, with SCBIR + SEIKD considered synonym of artificial
451 intelligence for data and information access services (AI4DIAS), where AI4DIAS is *part-of* a
452 yet-unaccomplished GEOSS. The closed-loop AI4DIAS system architecture, suitable for
453 semantics-enabled incremental learning [10, 87], is sketched in Figure 4 [10].
454

455 **Figure 4.** Artificial intelligence (AI) for data and information access services (AI4DIAS),
456 synonym of semantics-enabled DIAS or closed-loop EO image understanding (EO-IU) for

semantic querying (EO-IU4SQ) system architecture. At the Marr level of system understanding known as system design (architecture) [69], AI4DIAS is sketched as a closed-loop EO-IU4SQ system architecture, suitable for incremental semantic learning. It comprises a primary (dominant, necessary not sufficient) hybrid (combined deductive and inductive) feedback (provided with feedback loops) EO-IU subsystem in closed-loop with a secondary (dominated) hybrid feedback EO-SQ subsystem. *Subset-of* a computer vision (CV) system, where CV ⊃ EO-IU, the EO-IU subsystem is required to be automatic (no human-machine interaction is required by the CV system to run) and near real-time to provide the EO-SQ subsystem with useful information products, including thematic maps of symbolic quality, such as single-date ESA EO Level 2 Scene Classification Map (SCM) considered necessary not sufficient pre-condition to semantic querying, synonym of semantics-enabled information/knowledge discovery (SEIKD) in massive multi-source EO image databases. The EO-SQ subsystem is provided with a graphic user interface (GUI) to streamline: (i) top-down knowledge transfer from-human-to-machine of an *a priori* mental model of the 4D geospatial-temporal real-world, (ii) high-level user- and application-specific EO semantic content-based image retrieval (SCBIR) operations. Output products generated by the closed-loop EO-IU4SQ system are expected to monotonically increase their value-added with closed-loop iterations, according to Bayesian updating where Bayesian inference is applied iteratively [95, 96]: after observing some evidence, the resulting posterior probability can be treated as a prior probability and a new posterior probability computed from new evidence. One of Marr's legacies is the notion of computational constraints required to make the typically ill-posed non-polynomial (NP)-hard problem of intelligence, encompassing vision [89], better conditioned for numerical solution [30]. Marr's computational constraints that reflect properties of the world are embodied through evolution, equivalent to genotype [71], into the human visual complex system, structured as a hierarchical network of networks with feedback loops [89, 97-102, 181]). Marr's computational constraints are Bayesian priors in a Bayesian inference approach to vision [69, 95, 103], where ever-varying sensations (sensory data) are transformed into stable percepts (concepts) about the world in a world model [20], to perform successfully in the world [15].

To justify our working hypothesis sketched in Figure 3, specifically, 'Human vision → CV ⊃ EO-IU in operating mode ⊃ NASA EO Level 2 product → ESA EO Level 2 product → [EO-SCBIR + SEIKD = AI4DIAS] → GEOSS', let us introduce, first, the definition proposed for an EO-IU system to be considered in operating mode and, second, the background knowledge of vision in the multidisciplinary domain of cognitive science [70-74].

Based on scientific literature [10-12, 60, 78, 79, 104], a CV ⊃ EO-IU system is defined in operating mode if and only if it scores "high" in every index of a minimally dependent and maximally informative (mDMI) set of EO outcome and process (OP) quantitative quality indicators (Q^2Is), to be community-agreed upon to be used by members of the RS community, in agreement with the GEO-CEOS QA4EO *Cal/Val* guidelines [3]. A proposed instantiation of an mDMI set of EO OP-Q^2Is includes:

    (i)   degree of automation, inversely related to human-machine interaction, e.g., inversely related to the number of system's free-parameters to be user-defined based on heuristics,

    (ii)   effectiveness, e.g., thematic mapping accuracy,

    (iii)   efficiency in computation time and in run-time memory occupation,

    (iv)   robustness (vice versa, sensitivity) to changes in input data,

    (v)   robustness to changes in input parameters to be user-defined,

    (vi)   scalability to changes in user requirements and in sensor specifications,

    (vii) timeliness from data acquisition to information product generation,

    (viii) costs in manpower and computer power,

    (ix)   value, e.g., semantic value of output products, economic value of output services, etc.

509   According to the Pareto formal analysis of multi-objective optimization problems,
510 optimization of an mDMI set of OP-Q²Is is an inherently-ill posed problem in the Hadamard
511 sense [90], where many Pareto optimal solutions lying on the Pareto efficient frontier can be
512 considered equally good [105]. Any EO-IU system solution lying on the Pareto efficient frontier
513 can be considered in operating mode, therefore suitable to cope with the five Vs of spatial-
514 temporal EO *big data*, namely, volume, variety, veracity, velocity and value [4].
515

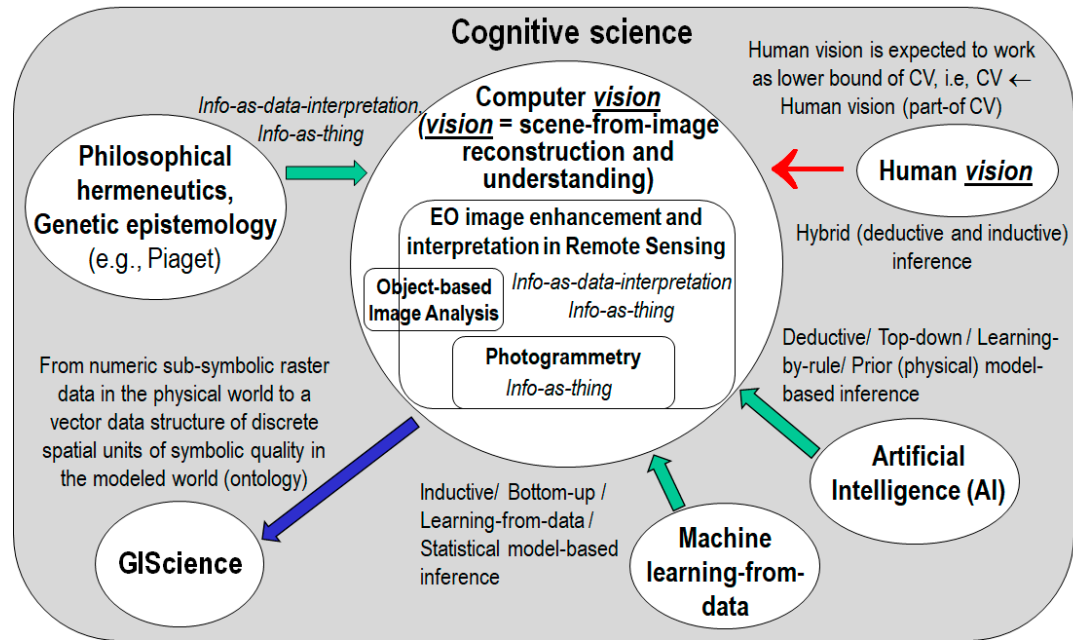



Figure 5. Multi-disciplinary cognitive science domain, where it is postulated that 'Human
vision → computer vision (CV)', where symbol '→' denotes relationship *part-of* pointing from
the supplier to the client, not to be confused with relationship *subset-of*, '⊃', meaning
specialization with inheritance from the superset to the subset, in agreement with the standard
Unified Modeling Language (UML) for graphical modeling of object-oriented software [81].
Working hypothesis 'Human vision → CV' means that human vision is expected to work as
lower bound of CV, i.e., a CV system is required to include as *part-of* a computational model
of human vision. In practice, to become better conditioned for numerical solution, an
inherently ill-posed CV system is required to comply with human visual perception
phenomena in the multi-disciplinary domain of cognitive science. Cognitive science is the
interdisciplinary scientific study of the mind and its processes. It examines what cognition
(learning, adaptation, self-organization) is, what it does and how it works [15, 70-74]. It
especially focuses on how information/knowledge is represented, acquired, processed and
transferred either in the neuro-cerebral apparatus of living organisms or in machines, e.g.,
computers. Like engineering, remote sensing (RS) is a meta-science [106], whose goal is to
transform knowledge of the world, provided by other scientific disciplines, into useful user-
and context-dependent solutions in the world. Neuroscience, in particular neurophysiology,
studies the neuro-cerebral apparatus of living organisms. Neural network (NN) is synonym of
distributed processing system, consisting of neurons as elementary processing elements and
synapses as lateral connections. Is it possible and even convenient to mimic biological mental
functions, e.g., human reasoning, by means of an artificial mind whose physical support is not
an electronic brain implemented as an artificial NN (ANN)? The answer is no according to the
"connectionists approach" promoted by traditional cybernetics, where a complex system
always comprises an "artificial mind-electronic brain" combination. This is alternative to a
traditional approach to artificial intelligence (AI), whose so-called symbolic approach
investigates an artificial mind independently of its physical support [70].

In the multidisciplinary domain of cognitive science (see Figure 5), vision is synonym of scene-from-image reconstruction and understanding [20], see Figure 6. Encompassing both biological vision and CV, vision is a cognitive (*information-as-data-interpretation*) problem [15], very difficult to solve because: (i) non-polynomial (NP)-hard in computational complexity [88, 89], (ii) inherently ill-posed in the Hadamard sense [90], i.e., vision admits no solution, multiple solutions or, if the solution exists, the solution's behavior changes continuously with the initial conditions [20, 68]. Vision is inherently ill-posed because affected by: (I) a 4D-to-2D data dimensionality reduction, from the spatial-temporal scene-domain to the (2D) image-domain, for example responsible of occlusion phenomena, and (II) a semantic information gap, from ever-varying sub-symbolic sensory data (sensations) in the physical world to stable symbolic percepts in the mental model of the physical world (modeled world, world ontology, world model) [16-21]. Since it is inherently ill-posed, vision requires *a priori* knowledge in addition to sensory data to become better posed for numerical solution [30]. For example, in inherently ill-posed CV systems, a valuable source of *a priori* knowledge is reverse engineering primate visual perception [89, 97-102], so that a CV system is constrained to include a computational model of human vision, i.e., 'Human vision → CV', see Figure 5. According to cognitive science, AI ⊃ CV is not a problem in statistics [107], which is tantamount to saying there is no (qualitative, equivocal, nominal) semantics in (quantitative, unequivocal, numeric) sensory data [10]. These principles are implicit in the dual meaning of word information, either quantitative (unequivocal) *information-as-thing*, typical of the Shannon data communication/transmission theory [108], or qualitative (equivocal) *information-as-data-interpretation*, typical of AI ⊃ CV tasks investigated by philosophical hermeneutics [15].

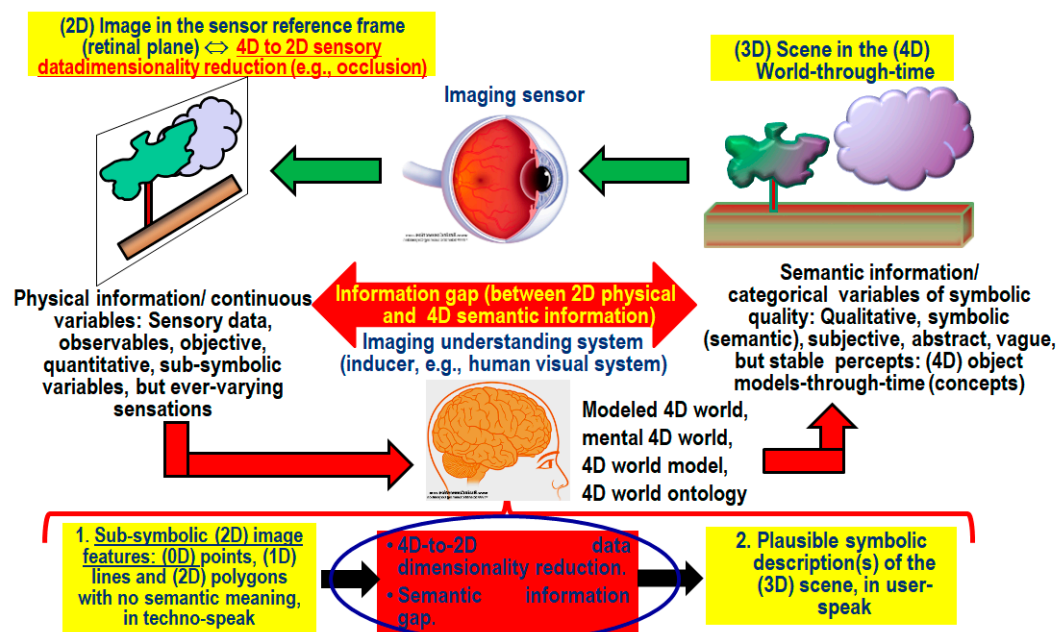



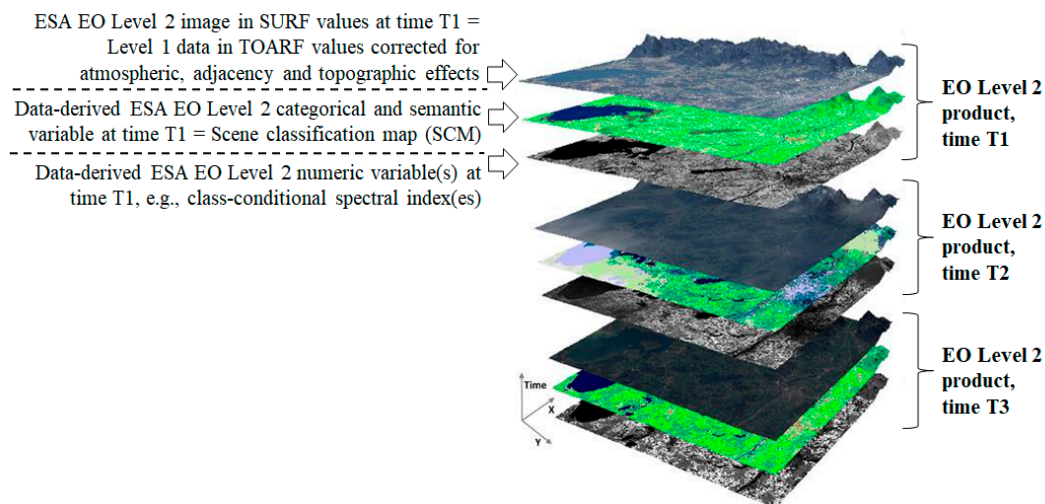
Figure 6. Synonym of scene-from-image reconstruction and understanding, vision is a cognitive (*information-as-data-interpretation*) problem [15] very difficult to solve because: (i) non-polynomial (NP)-hard in computational complexity [88, 89], and (ii) inherently ill-posed [20, 68] in the Hadamard sense [90]. Vision is inherently ill-posed because affected by: (I) a 4D-to-2D data dimensionality reduction from the scene-domain to the image-domain, e.g., responsible of occlusion phenomena, and (II) a semantic information gap from ever-varying sub-symbolic sensory data (sensations) in the physical world-domain to stable symbolic percepts in the mental model of the physical world (modeled world, world ontology, world model) [16, 21]. Since it is inherently ill-posed, vision requires *a priori* knowledge in addition to sensory data to become better posed for numerical solution [30].

575　　　　Largely oversighted by the RS and CV literature, an unquestionable true-fact
576　(observation) is that spatial information dominates color information in vision [20]. This
577　commonsense knowledge is obvious, but not trivial. For example, it is familiar to all human
578　beings wearing sunglasses: human panchromatic vision is nearly as effective as chromatic
579　vision in scene-from-image reconstruction and understanding [10]. This true-fact means that
580　spatial information dominates both the 4D geospatial-temporal scene-domain and the (2D)
581　image-domain involved with the cognitive task of vision, whose goal is scene-from-image
582　reconstruction and understanding, see Figure 6. This evidence is also acknowledged by the
583　Tobler's first law (TFL) of geography, familiar to geographers. The TFL of geography states
584　that "all things are related, but nearby things are more related than distant things"[109],
585　although certain phenomena clearly constitute exceptions [110]. Obscure to many geographers
586　familiar with the TFL formulation, the statistical concept of spatial autocorrelation is the
587　quantitative counterpart of the qualitative TFL of geography [10]. The relevance of spatial
588　autocorrelation in both the 4D geospatial-temporal scene-domain and the (2D) image domain
589　involved with vision is at the very foundation of the (GE)OBIA approach to CV, originally
590　conceived around year 2000 by the GIScience community as a viable alternative to traditional
591　1D spatial-context insensitive (pixel-based) image analysis [6, 7]. Unfortunately, rather than
592　starting from past background knowledge in the multi-disciplinary domain of cognitive
593　science, the GEOBIA approach was started from scratch by a self-referencing GEOBIA sub-
594　community within the GIScience domain, see Figure 5 [6, 7, 111, 112]. As a consequence of its
595　lack of interdisciplinarity, the GEOBIA community showed an increasing tendency to "re-
596　invent the wheel" in ever-varying implementations of the same sub-optimal EO-IU system
597　architecture, although the CV community clearly acknowledged that the linchpin of success of
598　any CV solution is system design together with information/knowledge representation, rather
599　than algorithm or implementation [16, 69]. Based on these observations, to enforce a 'Human
600　vision → CV' paradigm, see Figure 5, the following constraint can be adopted to make an
601　inherently ill-posed CV system better conditioned for numerical solution [10].

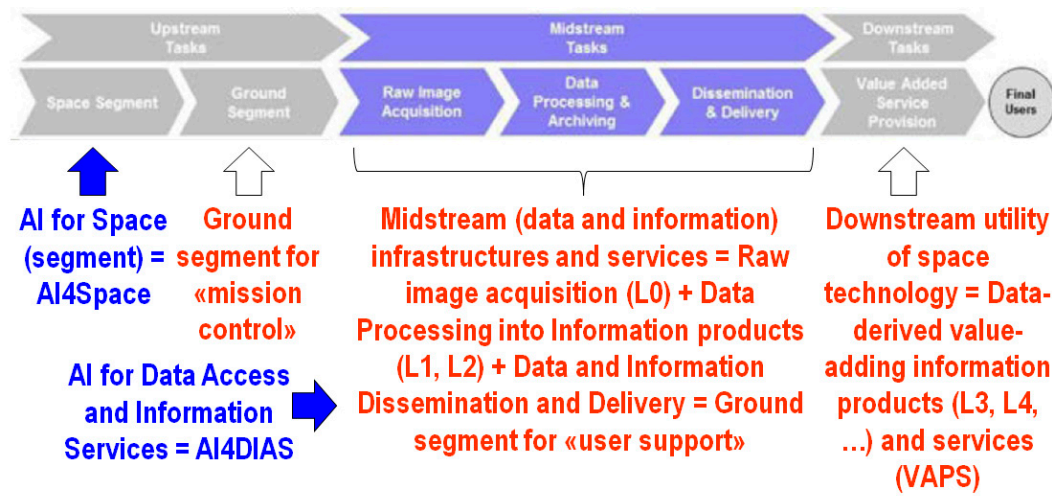
602　　　　If a chromatic CV ⊃ EO-IU system does not down-scale seamlessly to achromatic
603　image analysis, then it tends to ignore the paramount spatial information in favor of
604　subordinate (secondary) spatial context-insensitive color information, such as MS
605　signatures typically investigated in traditional pixel-based single-date or multi-
606　temporal EO-IU algorithms. In other words, a necessary and sufficient condition for
607　a CV ⊃ EO-IU system to fully exploit primary spatial topological information (e.g.,
608　adjacency, inclusion, etc.) and spatial non-topological information (e.g., spatial
609　distance, angle distance) components, in addition to secondary colorimetric
610　information, is to perform nearly as well when input with either panchromatic or
611　color imagery [10].

612　　　　Supported by this background knowledge about vision, a true-fact is that ESA EO Level
613　2 product generation is an inherently ill-posed CV problem (*chicken-and-egg* dilemma), whose
614　inherently ill-posed CV subtask is cloud and cloud-shadow quality layer detection. Since it is
615　ill-posed, ESA EO Level 2 product generation is, first, very difficult to solve; in fact, no ESA EO
616　Level 2 product has ever been accomplished in operating mode by any EO data provider at the
617　ground segment to date. Second, it requires *a priori* knowledge in addition to sensory data to
618　become better conditioned for numerical solution.

619　　　　Our conclusion is that systematic ESA EO Level 2 product generation, at the core of the
620　present work, is of potential interest to relevant portions of the RS community, involved with
621　EO *big data* transformation into timely, comprehensive and operational EO data-derived VAPS
622　[3, 5]. Regarded as necessary not sufficient pre-condition for a GEOSS to cope with the five Vs
623　of EO *big data* analytics, see Figure 3, systematic ESA EO Level 2 product generation is still
624　open for solution in operating mode.

625
626

627 **Figure 7.** Semantics-enabled EO data cube, synonym of artificial intelligence for data and
628   information access services (AI4DIAS). Each single-date EO Level 1 source image,
629   radiometrically calibrated into top-of-atmosphere reflectance (TOARF) values and stored in
630   the database, is automatically transformed into an ESA EO Level 2 product comprising: (i) a
631   single-date multi-spectral (MS) image radiometrically calibrated from top-of-atmosphere
632   reflectance (TOARF) into surface reflectance (SURF) values, corrected for atmospheric,
633   adjacency and topographic effects, stacked with (ii) its EO data-derived value-adding scene
634   classification map (SCM), equivalent to a sensory data-derived categorical/qualitative variable
635   of semantic quality, where the thematic map legend is general-purpose, user- and application-
636   independent and comprises quality layers such as cloud and cloud-shadow. It is eventually
637   stacked with (iii) its EO data-derived value-adding numeric variables, such as biophysical
638   variables, e.g., LAI [75], class-conditional spectral indexes, e.g., vegetation class-conditional
639   greenness index [113, 114], categorical variables of sub-symbolic quality (geographic field-
640   objects), e.g., fuzzy sets/discretization levels low/medium/high of a numeric variable, etc.
641   [115].

642   The several degrees of novelty of the ESA EO Level 2 product definition (refer to Section
643 1) are highlighted below.
644   First, the ESA EO Level 2 product definition is innovative because it overtakes the
645 traditional concept of EO data cube with an innovative EO data cube stacked with its data-
646 derived value-adding information cube, synonym of semantics-enabled data cube or AI4DIAS,
647 see Figure 7. A semantics-enabled EO data cube is alternative to existing EO data cubes,
648 affected by the so-called data-rich information-poor (DRIP) syndrome [116], such as the
649 existing first generation of the European Commission (EC) data and information access
650 services (DIAS) [117, 118]. Intuitively, EC-DIAS is affected by the DRIP syndrome because it is
651 provided with no CV system in operating mode as inference engine, capable of transforming
652 geospatial-temporal EO *big data*, characterized by the five Vs of volume, variety, veracity,
653 velocity and value [4], into VAPS, starting from semantic information products, such as the
654 ESA EO Level 2 SCM baseline product. Sketched in Figure 4, AI4DIAS complies with the
655 Marr's intuition that "vision goes symbolic almost immediately… without loss of information"
656 [69] (p. 343).
657   Second, in our understanding the ESA EO Level 2 product definition is synonym of EO
658 Analysis Ready Data (ARD) format. This ARD definition is in contrast with the Committee on
659 Earth Observation Satellites (CEOS) ARD for Land (CARD4L) products definition [119], where
660 atmospheric effect removal is required exclusively, whereas no data-derived SCM is
661 demanded as additional output. It is also in contrast with the U.S. Landsat ARD format [120,
662 121], where atmospheric effect removal is required, but no SCM is provided as output. In
663 practice, the CARD4L and U.S. Landsat ARD definitions are *part-of* the ESA EO Level 2 product

664   definition. The latter includes the former. When the latter is accomplished, so is the former, but
665   the vice versa does not hold.
666

667   **Figure 8.** Definitions in Space Economy 4.0: space segment, ground segment for «mission
668   control» = upstream, ground segment for «user support» = midstream (infrastructures and
669   services), downstream (utility of space technology) [59] (pp. 6, 57).

670   Third, we consider ESA EO Level 2 product generation a horizontal policy for background
671   developments in support of a "new" Space Economy 4.0 [59], see Figure 8. In Space 4.0, global
672   value chains will require both vertical and horizontal policies. Vertical policies are more
673   directional and 'active', focussing on directing change, often through mission-oriented policies
674   that require the active creation and shaping of markets. Horizontal policies are more focused
675   on the background conditions necessary for innovation, correcting for different types of market
676   and system failures [59].
677   Fourth, in an "old" mission-oriented (vertical) space economy [59], the ground segment
678   is typically defined as consisting of upstream and midstream, defined as the portion of ground
679   segment for mission support and user support respectively, see Figure 8. ESA expects ESA EO
680   Level 2 product generation to be accomplished by EO data providers at midstream. Actually,
681   ESA EO Level 2 product generation should occur as early as possible in the information
682   processing chain, e.g., in the space segment preliminary to the ground segment, in compliance
683   with the Marr's intuition that "vision goes symbolic almost immediately… without loss of
684   information" [69] (p. 343). Hence, in Figure 3, ESA EO Level 2 product generation becomes
685   synonym of AI applications for the space segment, AI4Space, where AI ⊃ CV. AI4Space comes
686   before the application of AI techniques to the ground segment, specifically, AI4DIAS. If a CV
687   ⊂ AI application in operating mode is implemented onboard a spaceborne platform of an EO
688   imaging sensor to provide imagery with intelligence (semantics), then Future Intelligent EO
689   imaging Satellites (FIEOS), conceived in the early 2000s [122], become realistic, such as future
690   intelligent EO small satellite constellations. In EO small satellite constellations provided with
691   no on-board radiometric *Cal* subsystem, improved time resolution is counterbalanced by
692   inferior radiometric *Cal* capabilities, considered mandatory by the GEO-CEOS QA4EO *Cal/Val*
693   guidelines [3] to guarantee interoperability of multiple platforms and sensors within and
694   across constellations. The visionary goal of AI4Space is realistic, based on the recent
695   announcement of a research and technological development (RTD) project focused on future
696   intelligent small satellite constellations. The quote is: "an Earth-i led consortium will develop
697   a number of new Earth Observation technologies that will enable processes, such as the
698   enhancement of image resolution, cloud-detection, change detection and video compression,
699   to take place on-board a small satellite rather than on the ground. This will accelerate the
700   delivery of high-quality images, video and information-rich analytics to end-users. On-board

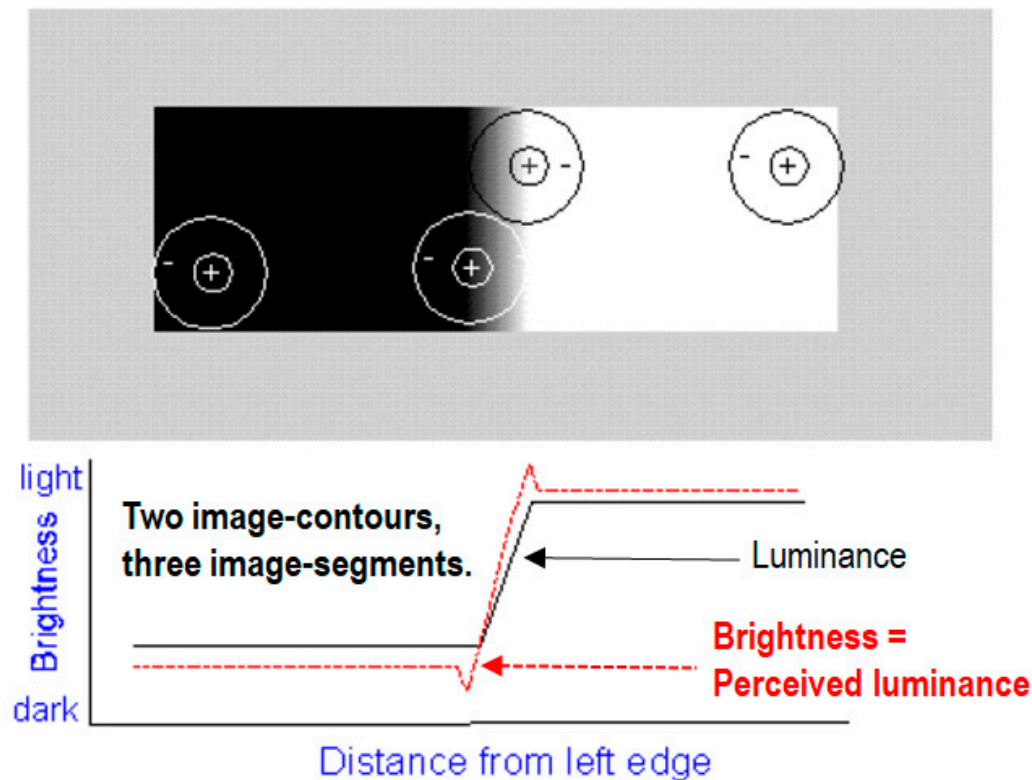
701 cloud detection will make the tasking of satellites more efficient and increase the probability
702 of capturing a usable and useful image or video. To achieve these goals, 'Project OVERPaSS'
703 will implement, test and demonstrate very high-resolution optical image analysis techniques,
704 involving both new software and dedicated hardware installed onboard small satellites to
705 radically increase their ability to process data in space. The project will also determine the
706 extent to which these capabilities could be routinely deployed on-board British optical imaging
707 satellites in the future" [123].

708 Fifth, in the RS common practice, the potential impact of the ESA EO Level 2 product
709 definition is relevant because it makes explicit once and for all the undisputable true-fact that
710 atmospheric effect correction [9, 49, 61-66, 124], adjacency and topographic effect corrections
711 [76, 125-129], BRDF effect correction [61, 62, 66, 91-94]; [92], in addition to cloud/cloud shadow
712 quality layer detection [9, 10, 36-39, 41-56], are inherently ill-posed CV problems [20, 68] in the
713 Hadamard sense [90], whose solution does not exist or, if it exists, it is not unique or it is not
714 robust to small changes in the initial condition, including changes to the input dataset. Since
715 they are inherently ill-posed, atmospheric, topographic, adjacency and BRDF effect
716 corrections, in addition to cloud/cloud shadow quality layer mapping, require *a priori*
717 knowledge in addition to sensory data to become better posed for numerical solution [30]. This
718 is tantamount to saying that radiometric *Cal* of EO imagery, encompassing ESA EO Level 2
719 product generation, is a *chicken-and-egg* dilemma [10, 125]. On the one hand, no EO image
720 understanding (classification) into a finite and discrete taxonomy of LC classes, in addition to
721 categorical layers cloud and cloud-shadow, is possible in operating mode if radiometric *Cal* is
722 not accomplished in advance, where dimensionless DNs are transformed into a physical unit
723 of radiometric measure to guarantee data interoperability through space, time and sensors, in
724 agreement with the GEO QA4EO *Cal/Val* requirements [5]. On the other hand, no radiometric
725 *Cal* of EO imagery is possible without knowing in advance LC classes and nominal quality
726 layers, such as cloud and cloud-shadow masks, since atmospheric, topographic and BRDF
727 effect corrections are LC class-dependent. This is tantamount to saying that, to become better
728 posed for automatic numerical solution (requiring no human-machine interaction), an
729 inherently ill-conditioned CV algorithm for EO image correction from atmospheric,
730 topographic, adjacency and BRDF effects ought to be run on a stratified (layered, class-
731 conditional) basis, i.e., it should run separately on informative EO image strata (masks, layers).
732 A stratified (layered, class-conditional) approach to CV complies with well-known criteria in
733 the equivocal domain of *information-as-data-interpretation* [15].

734 • Well known in statistics, the principle of statistic stratification. It states that "stratification
735 will always achieve greater precision provided that the strata have been chosen so that
736 members of the same stratum are as similar as possible in respect of the characteristic of
737 interest" [130].

738 • The popular problem solving criterion known as divide-and-conquer (*dividi-et-impera*) [29],
739 to be accomplished in agreement with the engineering principles of modularity, hierarchy
740 and regularity considered necessary for scalability in structured system design [32].

741 • A Bayesian approach to CV, where driven-without-knowledge (unconditional) data
742 analytics is replaced by driven-by-(prior) knowledge (class-conditional) data analytics [69,
743 95, 103, 131]. In the words of Quinlan: "one of David Marr's key is the notion of constraints.
744 The idea that the human visual system embodies constraints that reflect properties of the
745 world is foundational. Indeed, this general view seemed (to me) to provide a sensible way
746 of thinking about Bayesian approaches to vision. Accordingly, Bayesian priors are Marr's
747 constraints. The priors/constraints have been incorporated into the human visual system
748 over the course of its evolutionary history (according to the "levels of understanding of an
749 information processing system" manifesto proposed by Marr and extended by Tomaso
750 Poggio in 2012)" [103, 131]. In agreement with a Bayesian approach to CV, our working
751 hypothesis, sketched in Figure 3, postulates that CV includes a computational model of
752 human vision, i.e., 'Human vision → CV'. In practice, a CV system is constrained to comply

753   with human visual perception. This CV requirement agrees with common sense, although
754   it is largely oversighted in the RS and CV literature. In the words of Marcus: "there is no
755   need for machines to literally replicate the human mind, which is, after all, deeply error
756   prone, and far from perfect. But there remain many areas, from natural language
757   understanding to commonsense reasoning, in which humans still retain a clear advantage.
758   Learning the mechanisms underlying those human strengths could lead to advances in AI,
759   even if the goal is not, and should not be, an exact replica of human brain. For many people,
760   learning from humans means neuroscience; in my view, that may be premature. We don't
761   yet know enough about neuroscience to literally reverse engineer the brain, per se, and
762   may not for several decades, possibly until AI itself gets better. AI can help us to decipher
763   the brain, rather than the other way around. Either way, in the meantime, it should
764   certainly be possible to use techniques and insights drawn from cognitive and
765   developmental psychology, now, in order to build more robust and comprehensive AI,
766   building models that are motivated not just by mathematics but also by clues from the
767   strengths of human psychology" [27]. According to Iqbal and Aggarwal: "frequently, no
768   claim is made about the pertinence or adequacy of the digital models as embodied by
769   computer algorithms to the proper model of human visual perception... This enigmatic
770   situation arises because research and development in computer vision is often considered
771   quite separate from research into the functioning of human vision. A fact that is generally
772   ignored is that biological vision is currently the only measure of the incompleteness of the
773   current stage of computer vision, and illustrates that the problem is still open to solution"
774   [132]. For example, according to Pessoa, "if we require that a CV system should be able to
775   predict perceptual effects, such as the well-known Mach bands illusion where bright and
776   dark bands are seen at ramp edges, then the number of published vision models becomes
777   surprisingly small" [133], see Figure 9. In the CV and RS common practice, constraint
778   'Human vision → CV' is a viable alternative to heuristics typically adopted to constrain
779   inherently ill-posed inductive learning-from-data algorithms, where *a priori* knowledge is
780   typically encoded by design based on empirical criteria [27, 30, 57]. For example, state-of-
781   the-art DCNNs [31] encode *a priori* knowledge by design, where architectural
782   metaparameters must be user-defined based on heuristics. In inductive DCNNs, number
783   of layers, number of filters per layer, spatial filter size, inter-filter spatial stride, local filter
784   size for spatial pooling, spatial pooling filter stride, etc., are typically selected by users
785   based on trial-and-error strategies. As a consequence, inductive DCNNs work as black
786   boxes [27, 57], whose opacity contradicts the well-known engineering principles of
787   modularity, regularity and hierarchy typical of scalable systems [32]. In general, inductive
788   learning-from-data algorithms are inherently semi-automatic (requiring human-machine
789   interactions to define system's free-parameters, such as architectural metaparameters) and
790   site-specific (data-dependent) [2]. "No Free Lunch" theorems have shown that inductive
791   learning-from-data algorithms cannot be universally good [33, 34].
792

793   **Figure 9.** Mach bands illusion. In black: Ramp in luminance units across space. In red:
794   Brightness (perceived luminance) across space. One of the best-known brightness illusions,
795   where brightness is defined as a subjective aspect of vision, i.e., brightness is the perceived
796   luminance of a surface, is the psychophysical phenomenon of the Mach bands: where a
797   luminance (radiance, intensity) ramp meets a plateau, there are spikes of brightness, although
798   there is no discontinuity in the luminance profile. Hence, human vision detects two
799   boundaries, one at the beginning and one at the end of the ramp in luminance. Since there is
800   no discontinuity in luminance where brightness is spiking, the Mach bands effect is called a
801   visual "illusion". Along a ramp, no image-contour is perceived by human vision, irrespective
802   of the ramp's local contrast (gradient) in range $(0, +\infty)$. In the words of Pessoa, "if we require
803   that a brightness model should at least be able to predict Mach bands, the bright and dark
804   bands which are seen at ramp edges, the number of published models is surprisingly small"
805   [133]. In 2D signal (image) processing, the important lesson to be learned from the Mach
806   bands illusion is that local variance, contrast and first-order derivative (gradient) are statistical
807   features (data-derived numeric variables) computed locally in the (2D) image-domain NOT
808   suitable to detect image-objects (segments, closed contours) required to be perceptually
809   "uniform" ("homogeneous") in agreement with human vision. In other words, these popular
810   local statistics are not suitable visual features if detected image-segments/image-contours are
811   required to be consistent with human visual perception, including ramp-edge detection. This
812   straightforward (obvious), but not trivial observation is at odd with a large portion of the
813   existing computer vision (CV) and remote sensing (RS) literature, where many semi-
814   automatic image segmentation/image-contour detection algorithms are based on thresholding
815   the local variance, contrast or first-order gradient, e.g., [134-136], where a system's free-
816   parameter for thresholding image-objects or image-contours must be user-defined in range $\in$
817   $(0, +\infty)$ based on heuristics.

818   Although it is largely ignored by large portions of the RS community, the unequivocal
819   true-fact that inherently ill-posed CV $\supset$ EO-IU algorithms for radiometric *Cal* of DNs into SURF
820   or surface albedo values do require EO image classification to be performed in advance is
821   implicitly confirmed by existing open source or commercial software toolboxes for EO image
822   enhancement (pre-processing), reviewed hereafter.

823   Supported by NASA, the baseline of the U.S. Landsat ARD format [120, 121] is
824 atmospheric effect removal by the open source Landsat-4/5/7 Ecosystem Disturbance Adaptive
825 Processing System (LEDAPS). In LEDAPS, exclusion masks for water, cloud, shadow and
826 snow surface types were detected by an over-simplistic set of prior knowledge-based spectral
827 decision rules applied per pixel [124]. Quantitative analyses of LEDAPS products led by its
828 authors revealed that these exclusion masks were prone to errors, to be corrected in future
829 LEDAPS releases [124]. The same considerations hold for the Landsat 8 OLI/TIRS-specific
830 Landsat Surface Reflectance Code (LaSRC) adopted by the U.S. Landsat ARD format [137].
831 Unfortunately, suitable for testing or *Val* purposes, a multi-level image consisting of exclusion
832 masks has never been generated as output by either LEDAPS or LaSRC. To detect quality
833 layers cloud and cloud-shadow, in addition to snow/ice pixels, recent versions of LEDAPS and
834 LaSRC adopted the open source C Function of Mask (CFMask) algorithm [120, 121]. CFMask
835 was derived from the open source Function of Mask (FMask) algorithm [51, 52], translated into
836 the C programming language to facilitate its implementation in a production environment.
837 Unfortunately, to date, in a recent comparison of cloud and cloud-shadow detectors, those
838 implemented in LEDAPS scored low among alternative solutions [55]. By the way, potential
839 users of U.S. Landsat ARD imagery are informed by USGS in advance about typical CFMask
840 artifacts [56]. Like other cloud detection algorithms [54, 55], CFMask may have difficulties over
841 bright surface types such as building tops, beaches, snow/ice, sand dunes, and salt lakes.
842 Optically thin clouds will always be challenging to identify and have a higher probability of
843 being omitted by the U.S. Landsat ARD algorithm. In addition, the algorithm performance has
844 only been validated for cloud detection, and to a lesser extent for cloud shadows. No rigorous
845 evaluation of the snow/ice detection has ever been performed.
846   Transcoded into CFMask by the U.S. Landsat ARD processor, the open source Fmask
847 algorithm for cloud, cloud-shadow and snow/ice detection was originally developed for
848 single-date 30 m resolution 7-band (from visible blue, B, to thermal InfraRed, TIR) Landsat-
849 5/7/8 MS imagery, which includes a thermal band as key input data requirement [51]. In recent
850 years, FMask was extended to 10 m/20 m resolution Sentinel-2 MS imagery [52], featuring no
851 thermal band, and to Landsat image time-series (multiTemporal Mask, TMask) [53]. For more
852 details about the FMask software design and implementation, refer to the further Section 3.
853

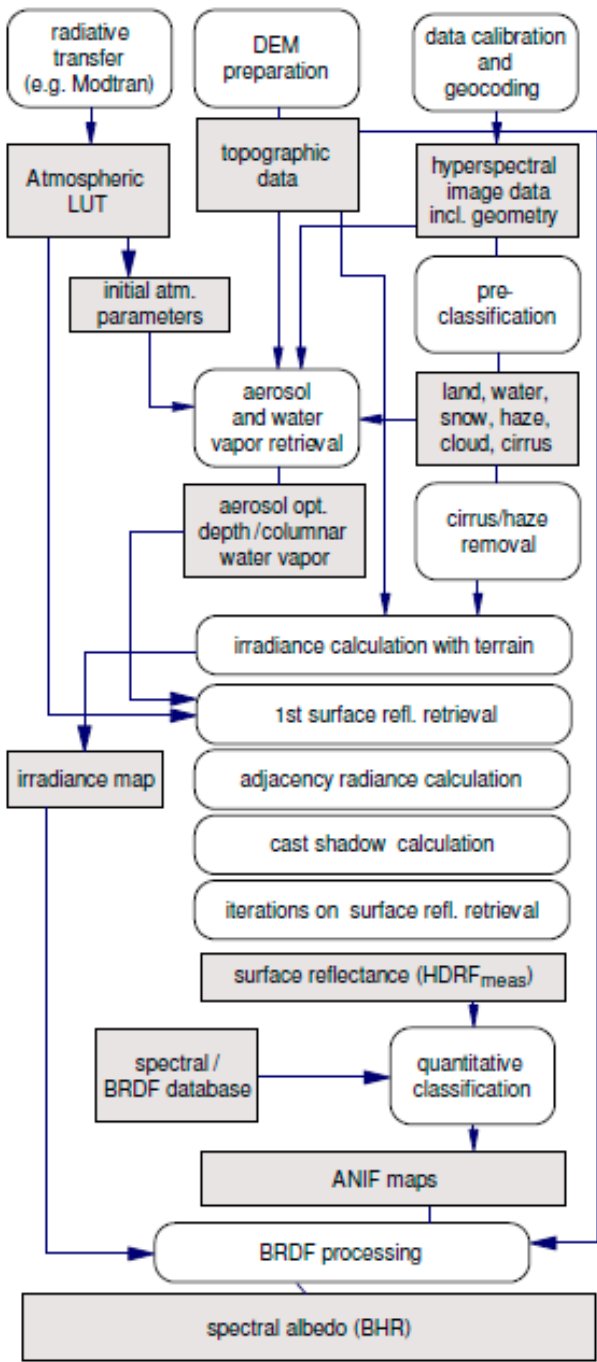



Figure 10. Same as in [66], courtesy of Daniel Schläpfer, ReSe Applications Schläpfer. A complete ("augmented") hybrid (combined deductive and inductive) inference workflow for MS image correction from atmospheric, adjacency and topographic effects. It combines a standard Atmospheric/Topographic Correction for Satellite Imagery (ATCOR) commercial software workflow [63, 64], with a bidirectional reflectance distribution function (BRDF) effect correction, which requires as input an image time-series of the same surface area acquired with different combinations of the sun and sensor positions. Processing blocks are represented as circles and output products as rectangles. This hybrid workflow alternates deductive/prior knowledge-based and inductive/learning-from-data inference units, starting from initial conditions provided by a first-stage prior knowledge-based decision tree for static (non-adaptive to data) color naming, such as the Spectral Classification of surface reflectance signatures (SPECL) decision tree [65] implemented within the ATCOR commercial software toolbox. Categorical variables generated as output by the two processing blocks identified as "pre-classification" and "classification" are employed as input by the subsequent processing

868   blocks to stratify (mask) unconditional numeric variable distributions, in line with the statistic
869   stratification principle [130]. Through statistic stratification, inherently ill-posed inductive
870   learning-from-data algorithms are provided with *a priori* knowledge required in addition to
871   data to become better posed for numerical solution, in agreement with the machine learning-
872   from-data literature [30].

873   In the Atmospheric/Topographic Correction for Satellite Imagery (ATCOR) commercial
874   software product, several per-pixel (spatial context-insensitive) deductive spectral rule-based
875   decision trees are implemented for use in different stages of an EO image enhancement
876   pipeline [63-66, 138, 139], see Figure 10. According to Richter and Schläpfer [63, 64], "pre-
877   classification as part of the atmospheric correction has a long history, e.g., in the NASA's
878   processing chain for MODIS" [50], also refer to [9, 49]. One of the ATCOR's prior knowledge-
879   based per-pixel decision trees delivers as output a haze/ cloud/ water (and snow) classification
880   mask file ("image_hcw.bsq"), see Table 3. In addition, ATCOR includes a so-called prior
881   knowledge-based decision tree for Spectral Classification of surface reflectance signatures
882   (SPECL) [65], see Table 4. Unfortunately, SPECL has never been tested by its authors in the RS
883   literature, although it has been validated by independent means [138, 139].

885   **Table 3.** Thematic map legend of the ATCOR-2/3/4 spectral pre-classification [63-66], whose
886   output product is identified as "image_hcw.bsq" (hcw = haze / cloud / water and snow) map.
887   According to Richter and Schläpfer, "pre-classification as part of the atmospheric correction
888   has a long history, e.g., as part of NASA's processing chain for MODIS", e.g., refer to [50].

Label	ATCOR-2/3/4 spectral pre-classification, Land Cover (LC) class definition	Order of detection
0	Background	
1	Cloud shadow	5
2	Cirrus – Thin over water	10
3	Cirrus – Medium over water	11
4	Cirrus – Thick over water	12
5	Land (if not 0 to 5 or 6 to 17)	17
6	Saturated (if (DN > 0.9 * DNmax), then saturated)	1
7	Snow / ice	6
8	Cirrus – Thin over land	7
9	Cirrus – Medium over land	8
10	Cirrus – Thick over land	9
11	Haze – Thin/medium over land	13
12	Haze – Thick/medium over land	14
13	Haze – Thin/medium over water	15
14	Haze – Thick/medium over water	16
15	Cloud over land	3
16	Cloud over water	4

892   **Table 4.** Rule set (structural knowledge) and order of presentation of the rule set (procedural
893   knowledge) adopted by the prior knowledge-based MS reflectance space quantizer, eligible
894   for MS reflectance space hyperpolyhedralization into MS color names, called Spectral
895   Classification of surface reflectance signatures (SPECL), implemented within the ATCOR
896   commercial software toolbox [63-65].

Index	Spectral Categories	Spectral Rule (based on reflectance measured at Landsat TM central wave bands: b1 is located at 0.48 μm, b2 at 0.56 μm, b3 at 0.66 μm, b4 at 0.83 μm, b5 at 1.6 μm, b7 at 2.2 μm)	Pseudo-color
1	Snow/ice	b4/b3 ≤ 1.3 AND b3 ≥ 0.2 AND b5 ≤ 0.12	
2	Cloud	b4 ≥ 0.25 AND 0.85 ≤ b1/b4 ≤ 1.15 AND b4/b5 ≥ 0.9 AND b5 ≥ 0.2	
3	Bright bare soil / sand / cloud	b4 ≥ 0.15 AND 1.3 ≤ b4/b3 ≤ 3.0	
4	Dark bare soil	b4 ≥ 0.15 AND 1.3 ≤ b4/b3 ≤ 3.0 AND b2 ≤ 0.10	
5	Average vegetation	b4/b3 ≥ 3.0 AND (b2/b3 ≥ 0.8 OR b3 ≤ 0.15) AND 0.28 ≤ b4 ≤ 0.45	

6	Bright vegetation	b4/b3 ≥ 3.0 AND (b2/b3 ≥ 0.8 OR b3 ≤ 0.15) AND b4 ≥ 0.45	
7	Dark vegetation	b4/b3 ≥ 3.0 AND (b2/b3 ≥ 0.8 OR b3 ≤ 0.15) AND b3 ≤ 0.08 AND b4 ≤ 0.28	
8	Yellow vegetation	b4/b3 ≥ 2.0 AND b2 ≥_b3 AND b3 ≥ 8.0 AND b4/b5 ≥ 1.5 [a]	
9	Mix of vegetation / soil	2.0 ≤ b4/b3 ≤ 3.0 AND 0.05 ≤ b3 ≤ 0.15 AND b4 ≥ 0.15	
10	Asphalt / dark sand	b4/b3 ≤ 1.6 AND 0.05 ≤ b3 ≤ 0.20 AND 0.05 ≤ b4 ≤ 0.20[a] AND 0.05 ≤ b5 ≤ 0.25 AND b5/b4 ≥ 0.7[a]	
11	Sand / bare soil / cloud	b4/b3 ≤ 2.0 AND b4 ≥ 0.15 AND b5 ≥ 0.15[a]	
12	Bright sand / bare soil / cloud	b4/b3 ≤ 2.0 AND b4 ≥ 0.15 AND (b4 ≥ 0.25b OR b5 ≥ 0.30[b])	
13	Dry vegetation / soil	(1.7 ≤ b4/b3 ≤ 2.0 AND b4 ≥ 0.25[c]) OR (1.4 ≤ b4/b3 ≤ 2.0 AND b7/b5 ≤ 0.83[c])	
14	Sparse veg. / soil	(1.4 ≤ b4/b3 ≤ 1.7 AND b4 ≥ 0.25[c]) OR (1.4 ≤ b4/b3 ≤ 2.0 AND b7/b5 ≤ 0.83 AND b5/b4 ≥ 1.2[c])	
15	Turbid water	b4 ≤ 0.11 AND b5 ≤ 0.05[a]	
16	Clear water	b4 ≤ 0.02 AND b5 ≤ 0.02[a]	
17	Clear water over sand	b3 ≥ 0.02 AND b3 ≥ b4 + 0.005 AND b5 ≤ 0.02[a]	
18	Shadow		
19	Not classified (outliers)		
[a] These expressions are optional and only used if b5 is present. [b] Decision rule depends on presence of b5. [c] Decision rule depends on presence of b7.			

Commissioned by ESA, the Sentinel-2 (atmospheric and topographic) Correction Prototype Processor (Sen2Cor) is not run systematically at the ESA ground segment. Rather, it can be downloaded free-of-cost from an ESA web site to be run on user side [8, 9, 37]. Hence, the Sen2Cor software toolbox does not satisfy the ESA EO Level 2 product requirements specification, refer to Section 1. The existing Sentinel-2 sensor-specific Sen2Cor prototype processor, sketched in Figure 11, adopts the same feedforward workflow of the popular ATCOR commercial software product [63-66]. The Sen2Cor software toolbox accomplishes, first, one SCM product generation from TOARF values by means of a per-pixel (spatial context-insensitive) prior spectral knowledge-based (static, non-adaptive to data) decision tree, whose SCM legend is shown in Table 1. Next, a stratified (class-conditional, driven-by-knowledge) MS image radiometric correction approach is adopted to transform TOARF into SURF values, where SURF values are sequentially corrected for atmospheric, adjacency and topographic effects stratified by the same SCM product generated at first stage from TOARF values.

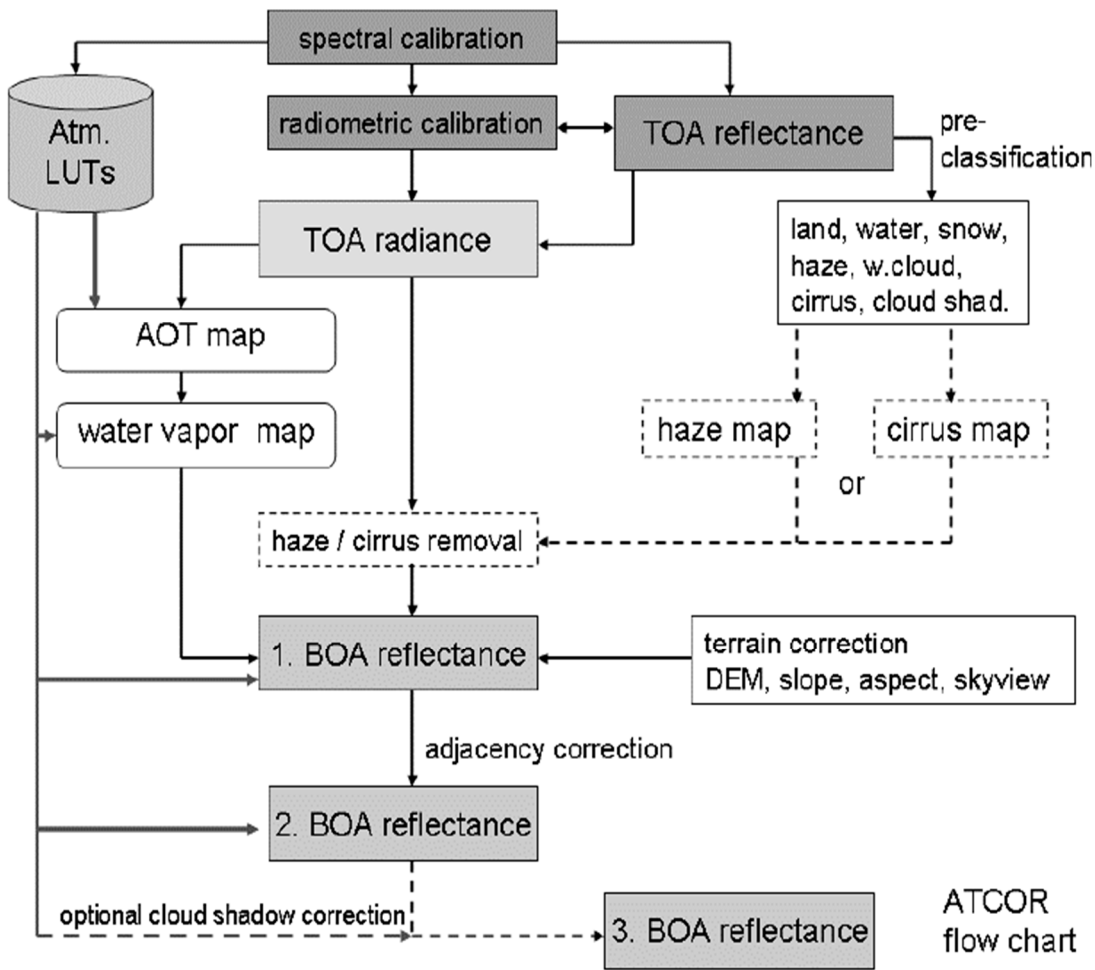
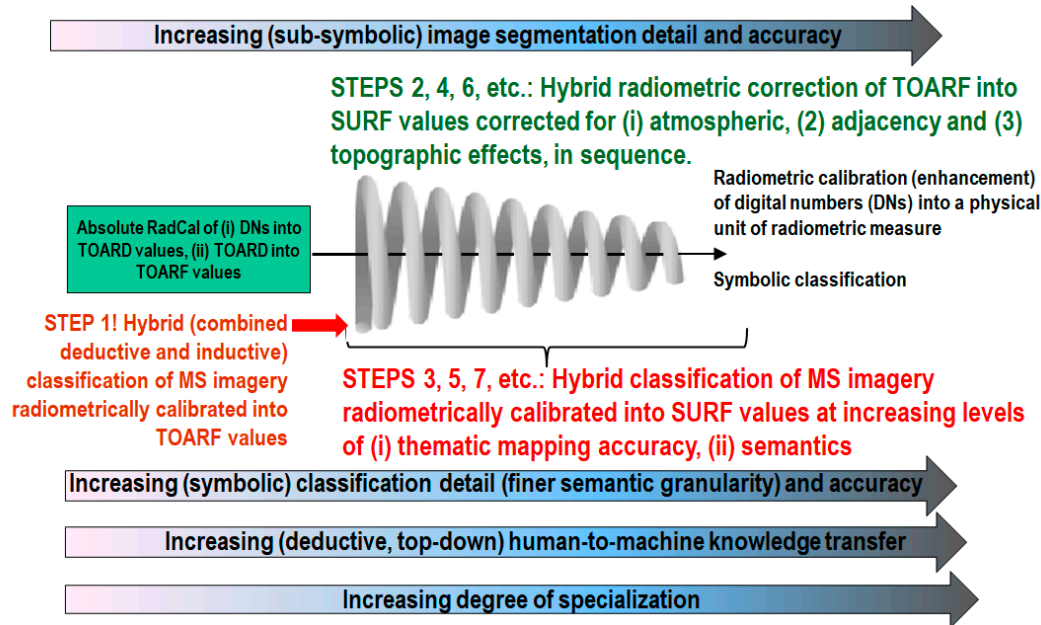
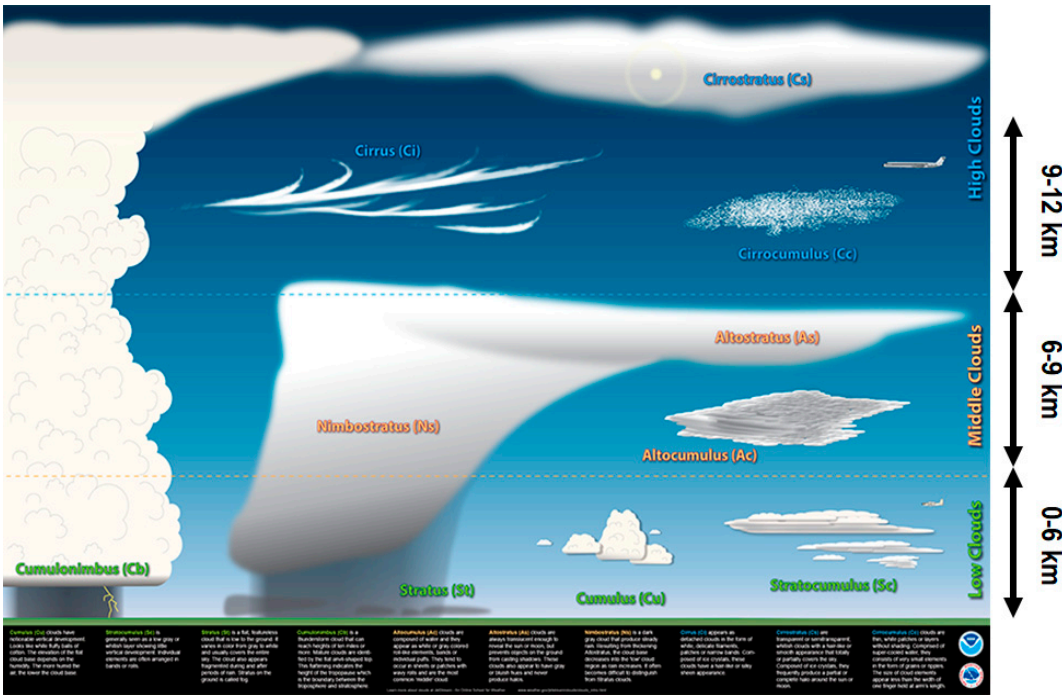



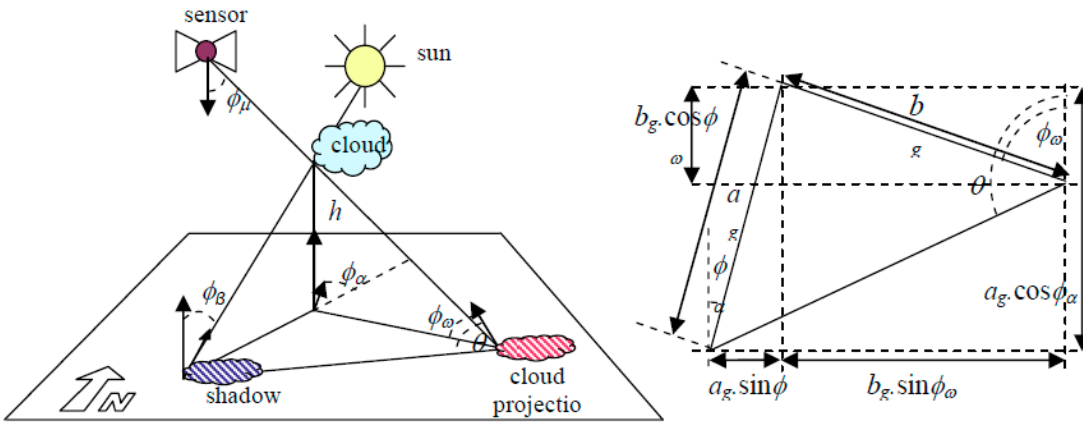
Figure 11. Sen2Cor flow chart for ESA Level 2 product generation from Sentinel-2 imagery [8, 9, 37], same as in the Atmospheric/Topographic Correction for Satellite Imagery (ATCOR) commercial software toolbox [63, 64]. While sharing the same system design, Sen2Cor and ATCOR differ at the levels of abstraction of algorithm and implementation [69]. First, a scene classification map (SCM) is generated from top-of-atmosphere reflectance (TOARF) values. Next, class-conditional MS image radiometric enhancement of TOARF into surface reflectance (SURF) values, synonym of bottom-of-atmosphere (BOA) reflectance values, corrected for atmospheric, adjacency and topographic effects is accomplished in sequence, stratified by the same SCM product generated at first stage from TOARF values. More acronyms in this figure: AOT = aerosol optical thickness, DEM = digital elevation model, LUT = look-up table.

Aa a viable alternative to the existing Sen2Cor, ATCOR, LEDAPS and LaSRC system design and implementation solutions, an inherently ill-posed atmospheric, topographic, adjacency and BRDF effect corrector for systematic ESA EO Level 2 product generation, where EO image classification, including cloud/cloud-shadow quality layer detection, is required in advance to be run on a LC class-specific basis [76, 125-127], is conceptually proposed as sketched in Figure 12 [10, 78, 79]. The estimation from an input numeric variable, starting from EO sensory data equivalent to DNs provided with no physical meaning, of an output numeric variable provided with a physical meaning of increasing quality, such as EO data-derived TOARF, SURF and spectral albedo values featuring increasing levels of radiometric *Cal* quality, requires as additional input, at each stage of the EO image enhancement flow chart, a categorical (nominal) variable belonging to a preliminary SCM automatically generated from the EO image radiometrically corrected at the previous stage in the workflow. In other words, equivalent to two sides of the same coin, categorical variables (e.g., SCMs at increasing levels of mapping accuracy and semantics) and continuous variables (e.g., TOARF, SURF and

936 spectral albedo values) should be estimated from raw EO imagery (coded as dimensionless
937 DNs, provided with no physical meaning) alternately and iteratively. It means that, in an EO
938 image pre-processing workflow conceived as a hybrid inference feedback system, such as the
939 stratified topographic correction (STRATCOR) algorithm proposed in [76], numeric variables
940 are estimated on a categorical (stratified, masked, class-conditional) basis while (enhanced)
941 categorical variables are estimated from (enhanced) continuous variables, in addition to *a priori*
942 knowledge, which is required in addition to data to make the inherently ill-posed CV problem
943 of radiometric *Cal* better posed for numerical solution [10, 30, 91], see Figure 12.
944

945 **Figure 12.** Ideal ESA EO Level 2 product generation design as a hierarchical alternating
946 sequence of: (A) hybrid (combined deductive and inductive) radiometric enhancement of
947 multi-spectral (MS) dimensionless digital numbers (DNs) into top-of-atmosphere reflectance
948 (TOARF), surface reflectance (SURF) values and spectral albedo values corrected in sequence
949 for (1) atmospheric, (2) adjacency, (3) topographic and (4) BRDF effects, and (B) hybrid
950 (combined deductive and inductive) classification of TOARF, SURF and spectral albedo values
951 into a sequence of ESA EO Level 2 scene classification maps (SCMs), whose legend (taxonomy)
952 of community-agreed land cover (LC) class names, in addition to quality layers cloud and
953 cloud-shadow, increases hierarchically. An implementation in operating mode of this EO
954 image pre-processing system design for stratified topographic correction (STRATCOR) is
955 presented and discussed in [10, 76, 78, 79]. In comparison with this desirable system design,
956 let us consider that, for example, the existing Sen2Cor software toolbox, developed by ESA to
957 support a Sentinel-2 sensor-specific Level 2 product generation on user side [8, 9, 37]), adopts
958 no hierarchical alternating approach between MS image classification and MS image
959 radiometric enhancement. Rather, Sen2Cor accomplishes, first, one SCM generation from
960 TOARF values based on a per-pixel (spatial context-insensitive) prior spectral knowledge-
961 based decision tree. Next, a class-conditional MS image radiometric enhancement of TOARF
962 into SURF values corrected for atmospheric, adjacency and topographic effects is
963 accomplished in sequence, stratified by the same SCM product generated at first stage from
964 TOARF values, see Figure 11.

965 As our fourth original contribution, it is worth mentioning here that TOARF, SURF and surface
966 albedo values whose physical domain of change is [0.0, 1.0], e.g., estimated as intermediate products or
967 final output product by an ESA EO Level 2 product generator, can be rescaled to the discrete and finite
968 range {0, 255} according to an unsigned char data coding, e.g., byte(SURF / 255. + 0.5), affected by a
969 quantization (discretization) error equal to (Max − Min) / number of bins / 2 (due to rounding to the
970 closest integer, either above or below) = (1.0 − 0.) / 255. / 2. = 0.002 = 0.2%. This observation, to be

971 considered straightforward, which does not mean either obvious or trivial, means that, for example, the
972 Planet Surface Reflectance (SR) Product [182], coded as a 16-bit unsigned short integer, can be transcoded
973 into an 8-bit unsigned char, affected by a quantization error as low as 0.2%, with a 50% save in memory
974 storage.

975 **Figure 13.** Cloud classification according to the U.S. National Weather Service
976 (www.srh.noaa.gov/jetstream/clouds/corefour.htm).

977

978 **Figure 14.** Courtesy of [48]. Sun-cloud-satellite geometry for arbitrary viewing and
979 illumination conditions. Left: Actual 3D representation of the Sun/ Cloud/ Cloud-shadow
980 geometry. Cloud height, h, is a typical unknown variable. Right: Apparent Sun/ Cloud/ Cloud-
981 shadow geometry in a 2D soil projection, with $a_g = h \cdot \tan\varphi_\beta$, $b_g = h \cdot \tan\varphi_\mu$.

3. Related works

983 To contribute toward filling an analytic and pragmatic information gap from multi-
984 sensor, multi-temporal and multi-angular EO *big data* cubes to systematic ESA EO Level 2
985 product generation never accomplished to date by any EO data provider at the ground
986 segment (refer to Section 1), our attention focuses on the RTD of an inherently ill-posed CV
987 subsystem for cloud and cloud-shadow quality layer detection (see Figure 13 and Figure 14) in
988 operating mode (refer to Section 2), considered an open problem to date by the RS community

989   [54-56]. Provided with a relevant survey value, this Section revises standard cloud/cloud-
990   shadow detectors, available either open source or free of cost. To highlight their structural
991   drawbacks and failure modes, if any, as our fifth original contribution, standard cloud/cloud-
992   shadow detectors are critically reviewed at the Marr's four levels of system abstraction, known
993   as information/knowledge representation, system design, algorithm and implementation [10,
994   16, 69].

995   In the last decade, many different cloud/cloud-shadow detection algorithms were
996   presented in the RS literature, to be input with either a single-date MS image or an MS image
997   time-series typically acquired by either one EO spaceborne/airborne MS imaging sensor or a
998   single family of MS imaging sensors, e.g., spaceborne Landsat [9, 10, 36-39, 41-53].

999   Predated by a long history of deductive convolutional neural networks (CNNs), consisting
1000   of "handcrafted" (physical model-based) multi-scale 2D spatial filter banks, developed since
1001   the early 1980s for multi-scale image analysis (encoding, decomposition), synthesis (decoding,
1002   reconstruction) [140] and classification [141], the recent hype about inductive end-to-end
1003   learning-from-data Deep CNNs (DCNNs) in CV applications [31] is progressively affecting the
1004   RS discipline, in spite of increasing disillusionment in the AI community [10, 27, 57, 97, 107,
1005   142-145]. For example, an increasing number of DCNN applications for cloud detection can be
1006   found in the RS literature, e.g., [35]. However, to our best knowledge, only one application of
1007   DCNNs to the simultaneous detection of cloud and cloud-shadow phenomena exists to date
1008   [146].

1009   To explain this fact, a non-trivial understanding of the information/knowledge
1010   representation in DCNNs is required, according to the Marr's four levels of understanding of
1011   an information processing system, specifically, information/knowledge representation, system
1012   design (architecture), algorithm per architectural block and implementation per algorithm [10,
1013   69, 78, 79]. In inductive learning-from-data DCNNs, 2D spatial filter profiles are learned end-
1014   to-end from supervised image examples whereas *a priori* knowledge is encoded by design
1015   based on heuristics, a.k.a. trial-and-error. In greater detail, the whole set of DCNN architectural
1016   metaparameters, specifically, number of layers, number of convolutional filters per layer, filter
1017   size, inter-filter stride, subsampling filter size, subsampling inter-filter stride, etc., is user-
1018   defined based on empirical criteria. Unfortunately, inductive learning-from-data algorithms,
1019   including DCNNs, are affected by inherent limitations [2, 10, 27, 30, 33, 34, 57, 97, 107, 142, 143]
1020   and known failure modes [144, 145], including the following. Suitable for learning complex
1021   correlations between input and output features, inductive learning-from-data systems are
1022   unable to learn representations of causality (cause-effect relationships, dependency), known
1023   that "correlation does not imply causation and vice versa" [10, 16, 27, 30, 57, 58]. This limitation
1024   implies that, first, it not straightforward to integrate *a priori* knowledge about the physical
1025   world onto a DCNN architecture, whose information representation pertains mainly to
1026   (largely opaque) correlations between input and output features. Second, DCNNs thus far
1027   have shown no self-organizing capability, e.g., based on combinations of cooperative with
1028   competitive learning-from-data policies [147], suitable for developing a syntactic hierarchical
1029   system [16, 27], structured as a network of specialized subnetworks [10], in agreement with the
1030   engineering principles of modularity, hierarchy and regularity considered necessary for
1031   scalability in structured system design [32]. In the words of Marcus, "the core problem, at least
1032   at present, is that deep learning learns correlations between sets of features that are themselves
1033   "flat" or non-hierarchical, as if in a simple, unstructured list, with every feature on equal
1034   footing. Hierarchical structures (e.g., syntactic trees that distinguish between main clauses and
1035   embedded clauses in a sentence) are not inherently or directly represented in such systems,
1036   and as a result deep learning systems are forced to use a variety of proxies that are ultimately
1037   inadequate, such as the sequential position of a word presented in a sequences " [27].

1038   Our straightforward, but not trivial conjecture is that cloud/cloud-shadow detection is a
1039   typical example of physical model-based cause-effect relationship expected to be very difficult
1040   to solve by inductive machine learning-from-examples algorithms, including DCNNs, suitable

1041 for learning complex correlations between input and output features, but unable to discover
1042 inherent representations of causality.

1043   This conjecture is perfectly confirmed by experimental conclusions reported in [146] (pp.
1044 32, 33): "The main problem of DCNN-based classification is cloud-shadow, e.g., in an image
1045 object-based Intersection over Union, IOU, classification quality measure assessment in range
1046 [0, 1], class cloud-shadow scored as low as IOU = 0.0212 in the validation image set at hand".
1047 Predicted by theory, these experimental results prove *a posteriori* the well-groundedness of our
1048 first-principles in the AI ⊃ CV domains. These expected results reveal that, based on
1049 knowledge of the first-principles in inference, specifically, induction, deduction and abduction
1050 [16, 19, 21], and in agreement with the increasing disillusionment of the AI community about
1051 deep learning-from-data [10, 27, 57, 97, 107, 142-145], the RS community should avoid
1052 exploring dead-end solutions where AI ⊃ CV algorithms, developed by other communities in
1053 the cognitive science domain (see Figure 5), are adopted as black-boxes by RS scientists and
1054 practitioners, such as increasingly popular DCNNs applied for detecting cloud and cloud-
1055 shadow cause-effect-relationships in EO imagery. Based on this rationale, although considered
1056 state-of-the-art by the mainstream CV and RS audience, DCNN solutions are excluded from
1057 any further review in this paper, focused on jointly cloud/cloud-shadow detection.

1058   For comparison purposes, three popular computer programs available free-of-cost for
1059 cloud/cloud-shadow detection in spaceborne MS imagery are critically revised hereafter, see
1060 Table 5.

1061 •   The open source single-date FMask algorithm [51], originally developed for single-date 30
1062 m resolution 7-band (from visible blue, B, to thermal InfraRed, TIR) Landsat-5/7/8 MS imagery,
1063 which includes a thermal band as key input data requirement. FMask was recently extended
1064 to 10 m/20 m resolution Sentinel-2 MS imagery [52], featuring no thermal band, and to Landsat
1065 image time-series (multiTemporal Mask, TMask) [53]. The potential relevance of FMask is
1066 augmented by considering that CFmask, an Fmask program version transcoded into the C
1067 programming language for increased efficiency, is adopted for cloud, cloud-shadow and
1068 snow/ice classification by the LEDAPS and LaSRC algorithms for atmospheric correction in the
1069 U.S. Landsat ARD product [120, 121, 137]. Unfortunately, CFMask is affected by known
1070 artifacts [56]. Moreover, in a recent comparison of cloud and cloud-shadow detectors, those
1071 implemented in LEDAPS scored low among alternative solutions [55].

1072 •   The single-date Sentinel-2 Multi-Spectral Instrument (MSI) sensor-specific *Sen2Cor*
1073 prototype processor, capable of automated atmospheric, adjacency and topographic effect
1074 correction and SCM product generation, including cloud and cloud-shadow detection, refer to
1075 the SCM legend shown in Table 1. The Sen2Cor prototype processor is distributed free-of-cost
1076 by ESA to be run on user side [8, 9]. Hence, it does not satisfy the ESA EO Level 2 product
1077 requirements specification proposed in Section 1. Sen2Cor incorporates capabilities of the
1078 ATCOR commercial software toolbox [63-65], see Figure 11. Sen2Cor is affected by known
1079 artifacts [37, 40, 54].

1080 •   The multi-date CNES/CESBIO/DLR Multisensor Atmospheric Correction and Cloud
1081 Screening (MACCS)-Atmospheric/Topographic Correction (ATCOR) Joint Algorithm (MAJA)
1082 [39, 41], which incorporates capabilities of the ATCOR commercial software toolbox [63-65]. It
1083 is affected by known artifacts [40, 54].

1084   Table 5 summarizes the Sen2cor, MAJA and FMask computer programs at the Marr levels
1085 of system understanding known as information/knowledge representation, system design
1086 (architecture) and algorithm [10, 69, 78, 79].

1087
1088   **Table 5.** Comparison of alternative cloud/cloud-shadow detection algorithms at the Marr
1089   levels of system understanding of information/knowledge representation, system design
1090   (architecture) and algorithm [10, 69, 78, 79]

Acronyms	Multi-spectral (MS)	Radio metric	Spatial res	Spectral resolution	Single-date (S) or	Land cover (LC) class detection, in addition to	Cloud detection	Cloud-shadow detection

Digital number (featuring no physical meaning) = **DN** Top-of-atmosphere reflectance = **TOARF**. Surface reflectance = **SURF**, where TOARF ⊃ SURF, because TOARF ≈ SURF + atmospheric noise + topographic noise. Inductive = **I**. Deductive = **D**. Hybrid = Combined I + D = **H**. 1D Pixel-based (spatial context-insensitive and spatial topology non-preserving) = **P**. 1D Object-based (spatial context-sensitive, but spatial topology non-preserving) = **O**. 2D image analysis (spatial context-sensitive and spatial topology-preserving) = **2D**. Yes: **Y**. No: **N**.	Sensor(s)	calibration	olution		Multi-temporal (MT)	classes and cloud-shadow, if any					Spatial search of cloud-shadow pixels starting from cloud candidates
		Acronyms: **TOARF** or **SURF**, refer to cell at top left		Acronyms: Visible Blue, Green, Red = **B**, **G**, **R**. Near InfraRed (IR) = **NIR**. Medium IR = **MIR.** Thermal IR = **TIR.**		Acronyms: **I, D** or **H**, refer to cell at top left	Acronyms: **P, O** or **2D**, refer to cell at top left	Acronyms: **I, D** or **H**, refer to cell at top left	Acronyms: **P, O** or **2D**, refer to cell at top left	Acronyms: **I, D** or **H**, refer to cell at top left	Acronyms: **P, O** or **2D**, refer to cell at top left	
Proposed approach, AutoCloud+	All MS past, present and future, airborne or spaceborne, whether or not provided with radiometric calibration metadata files	DN or TOARF or SURF or surface albedo	Any	From B, G, R, NIR, MIR to TIR, including Cirrus band (depending on the available spectral channels).	S	H, with LC classes: Water, Shadow, Bare soil, Built-up, Vegetation, Snow, Ice, Fire, Others	P + O + 2D	H	P + O + 2D	H	P + O + 2D	Y
Sen2Cor (including all cloud probability classes)	Sentinel-2 MSI	TOARF	10 m	B, G, R, NIR, MIR to TIR, including Cirrus band.	S	D, with LC classes: Water, Bare soil, Vegetation, Snow	P	D	P	D	P	Y
MAJA	Formosat2, LANDSAT 5/7/8, SPOT 4/5, Sentinel 2, VENµS.	TOARF	5.3 to 30 m	From B, G, R, NIR, MIR to TIR, including Cirrus band (depending on the available spectral channels)	MT	D, with LC classes: Water, Bare soil, Vegetation, Snow	P	D	P	D	P	Y
FMask	Landsat-7/8 (with thermal band) and Sentinel-2 (without thermal band)	TOARF	10 to 30 m	From B, G, R, NIR to MIR, plus TIR when available (in Landsat imagery)	S	D, with LC classes: Clear land, clear water, snow.	P	D	P + contextual to remove isolated pixels	H	P + O, due to Segmentation of the cloud layer	Y

ATCOR-4 [64]	Airborne, spaceborne	TOARF or SURF or spectral albedo	Any	From B, G, R, NIR to MIR, including Cirrus band (depending on the available spectral channels). No TIR is exploited.	S	D, with LC classes: Water, Land, Haze, Snow/Ice (depending on the available spectral channels)	P	H (D + I, e.g., I = image-wide histogram-based analytics)	P	H	P	N

1091

1092   The first observation stemming from Table 5 is that the Sen2cor and MAJA computer
1093   programs employ deductive (top-down, prior knowledge-based) inference, exclusively, to
1094   map MS imagery onto LC classes, in addition to quality layers cloud and cloud-shadow. This
1095   is in contrast with biological cognitive systems [77], where hybrid inference combines
1096   deductive and inductive inference to take advantage of each and overcome their shortcomings
1097   [2, 10, 33, 34, 78, 79], refer to Section 2.
1098   The second observation driven from Table 5 is that the Sen2cor, MAJA and FMask
1099   computer programs employ 1D image analysis algorithms exclusively, either pixel-based, i.e.,
1100   spatial context-insensitive and spatial topology non-preserving, or spatial context-sensitive
1101   (e.g., image object-based or local window-based), but spatial topology non-preserving (non-
1102   retinotopic) [10, 89, 97, 147]. In CV programs, 1D image analysis is a methodological
1103   (structural) drawback because perceptual evidence proves that, in vision, primary spatial
1104   topological information (e.g., adjacency, inclusion, etc.) and spatial non-topological
1105   information (e.g., spatial distance, angle measure) components dominate secondary color
1106   information, which is the sole information available at the sensor's spatial resolution, i.e., at
1107   the pixel level of spatial analysis, refer to Section 2 [10]. Intuitively, 1D image analysis
1108   algorithms are invariant to permutations in the 1D vector data sequence generated from a (2D)
1109   image, where image is synonym of 2D gridded data set, see Figure 15 [31]. In short, 1D analysis
1110   of (2D) imagery is affected by a loss in data dimensionality.

1111

Dali's "Gala contemplating the Mediterranean Sea" at 20 mtr becomes the Lincoln portrait, RGB image, ≈ 18 × 13 = 234 Dali's square, 504 × 669 digital pixels

Pnchrmtc image.

Median filter, 35 × 35 pixels in size

2D grid, 35 × 35 pixels per local window

1D image analysis is invariant to permutations in the input sequence, i.e., insensitive to changes in the order of presentation of the input sequence!

1D image analysis = vector sequence / vector series analysis, where each vector data is either pixel-based or local window-based

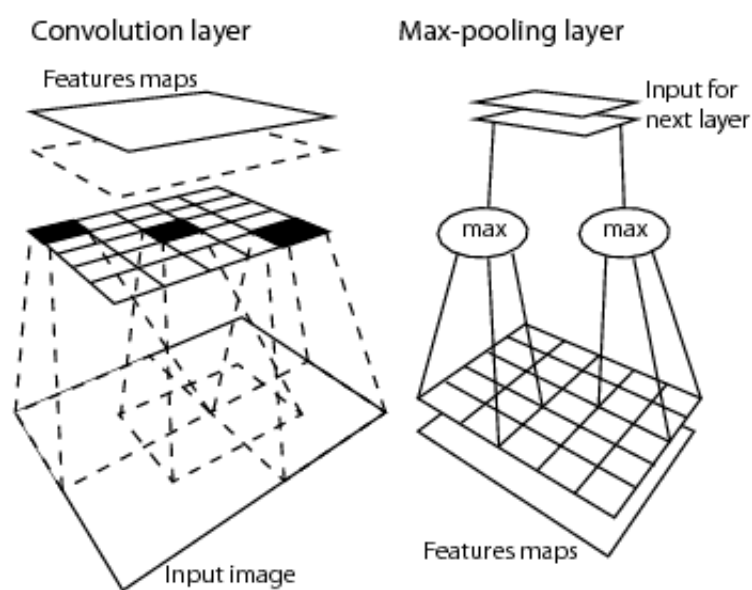
Order of presentation in the input vector sequence

This is WHAT a traditional inductive 1D data-sequence learning classifier actually sees when watching the (2D) image at left.

1D image analysis = either local window-based (insensitive to spatial topological info) OR pixel-based (spatial context-insensitive, insensitive to spatial topological AND spatial non-topological information)

1112   **Figure 15.** Example of 1D image analysis, which is spatial topology non-preserving (non-
1113   retinotopic) in a (2D) image-domain [10, 89, 97, 147]. Intuitively, 1D image analysis is
1114   insensitive to permutations in the input data set [31]. Synonym of 1D analysis of a 2D gridded
1115   data set, 1D image analysis is affected by spatial data dimensionality reduction. The (2D)
1116   image at left is transformed into the 1D vector data stream (sequence) shown at bottom, where

1117   vector data are either pixel-based or spatial context-sensitive, e.g., local window-based. This
1118   1D vector data stream means nothing to a human photointerpreter. When it is input to either
1119   an inductive learning-from-data classifier or a deductive learning-by-rule classifier, the 1D
1120   vector data sequence is what the classifier actually sees when watching the (2D) image at left.
1121   Undoubtedly, computers are more successful than humans in 1D image analysis. Nonetheless,
1122   humans are still far more successful than computers in 2D image analysis (see Figure 16).

1123   Alternative to 1D image analysis is 2D image analysis, which is spatial topology-
1124 preserving (retinotopic) [10, 89, 97, 147], i.e., it is sensitive to permutations in the order of
1125 presentation of the input data set [31], see Figure 16. In our understanding, 2D image analysis
1126 is the linchpin of success of multi-scale 2D spatial filter banks for image analysis (encoding,
1127 decomposition), synthesis (decoding, reconstruction) and classification, either
1128 deductive/physical model-based [140, 141] or end-to-end inductive learning-from-data, such
1129 as DCNNs [31].
1130

1131   **Figure 16.** 2D image analysis as synonym of spatial topology-preserving (retinotopic) feature
1132   mapping in a (2D) image-domain [10, 89, 97, 147]. Intuitively, 2D image analysis is sensitive to
1133   permutations in the input data set [31]. Activation domains of physically adjacent processing
1134   units in the 2D array of convolutional spatial filters are spatially adjacent regions in the 2D
1135   visual field. Provided with a superior degree of biological plausibility in modelling 2D spatial
1136   topological and spatial non-topological information components, distributed processing
1137   systems capable of 2D image analysis, such as deep convolutional neural networks (DCNNs),
1138   typically outperform traditional 1D image analysis approaches. Will computers ever become
1139   as good as humans in 2D image analysis?

1140   The fundamental difference between 1D image analysis and 2D image analysis with
1141 regard to spatial topology-preserving (retinotopic) feature mapping [10, 89, 97, 147] is at the
1142 basis of a proposed revision of the loose GE(OBIA) paradigm [7] into a more restrictive
1143 definition of EO imagery for Geographical Sciences (EO4GEO), synonym of 2D analysis of EO
1144 (2D) imagery for GISCience applications [111, 112], see Figure 17.
1145

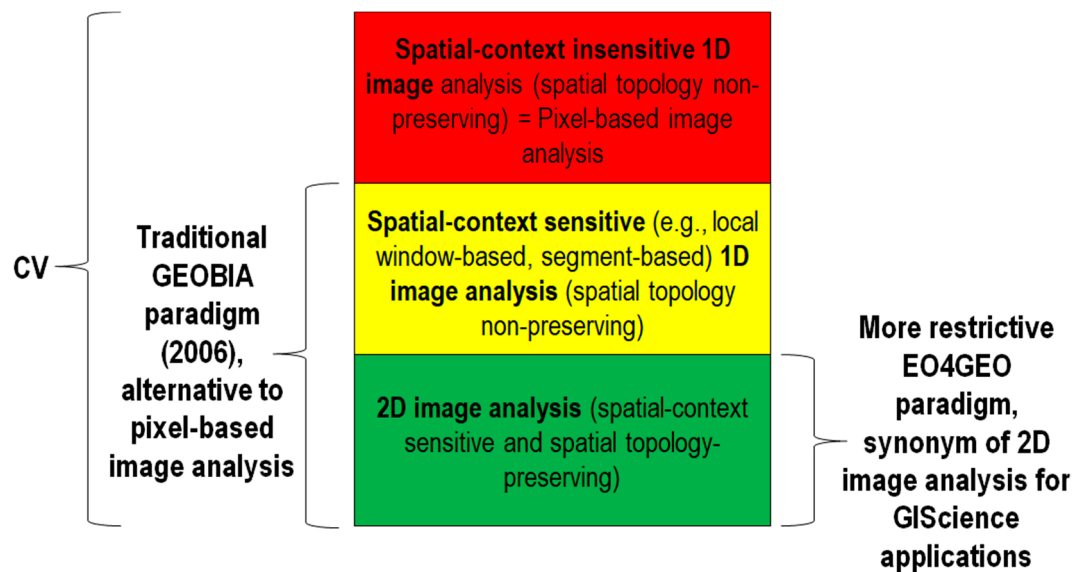



Figure 17. EO for Geographical Sciences (EO4GEO, EO4GIScience) framework, meaning EO big data analytics in operating mode for GIScience applications, constrained by 2D (retinotopic, spatial topology-preserving) image analysis in cognitive science [111]. EO4GEO is more restrictive than the traditional GEOBIA paradigm, formalized in 2006 and 2014 as a viable alternative to 1D spatial context-insensitive (pixel-based) image analysis [6, 7, 112].

At the Marr's less abstract levels of understanding of an information processing system, known as algorithm and implementation [10, 69, 78, 79], subsequent to the two more abstract levels of understanding of information/knowledge representation and system design (architecture), it is interesting to investigate the Sen2Cor's static (non-adaptive to data, prior knowledge-based) per-pixel spectral rule-based decision tree implementation for automated ESA EO Level 2 SCM product generation [9]. Noteworthy, the same considerations driven from the Sen2Cor's spectral knowledge-based decision tree implementation for SCM generation hold for the LEDAPS, LaSRC and ATCOR spectral knowledge-based decision trees, implemented for pixel-based pre-classification as necessary not sufficient pre-condition to make inherently ill-posed atmospheric and/or topographic correction algorithms better posed for numerical solution [30], see Section 2. According to Richter and Schläpfer [63, 64], "pre-classification as part of the atmospheric correction has a long history, e.g., in the NASA's processing chain for MODIS" [50].

First, in the interdisciplinary domain of cognitive science [11, 12], it was highlighted that static (non-adaptive to data, prior knowledge-based) spectral rule-based decision trees aim at a mutually exclusive and totally exhaustive (hyper)polyhedralization of a MS reflectance (hyper)space, equivalent to (numeric) color space discretization into color names belonging to a categorical variable, known as color vocabulary. Within the domain of cognitive science (see Figure 5), color naming was deeply investigated by linguistics. Central to this consideration is Berlin and Kay's landmark study of a "universal" inventory of eleven basic color (BC) words in twenty human languages: black, white, gray, red, orange, yellow, green, blue, purple, pink and brown [148]. Suitable for color naming of MS imagery, MS reflectance (hyper)space (hyper)polyhedralization is difficult to think of and impossible to visualize when the MS data space dimensionality is superior to three, see Figure 18. This is not the case of basic color (BC) names adopted in human languages [148], whose mutually exclusive and totally exhaustive perceptual polyhedra, neither necessarily convex nor connected, are intuitive to think of and easy to visualize in a 3D monitor-typical red-green-blue (RGB) data cube [149], see Figure 19.

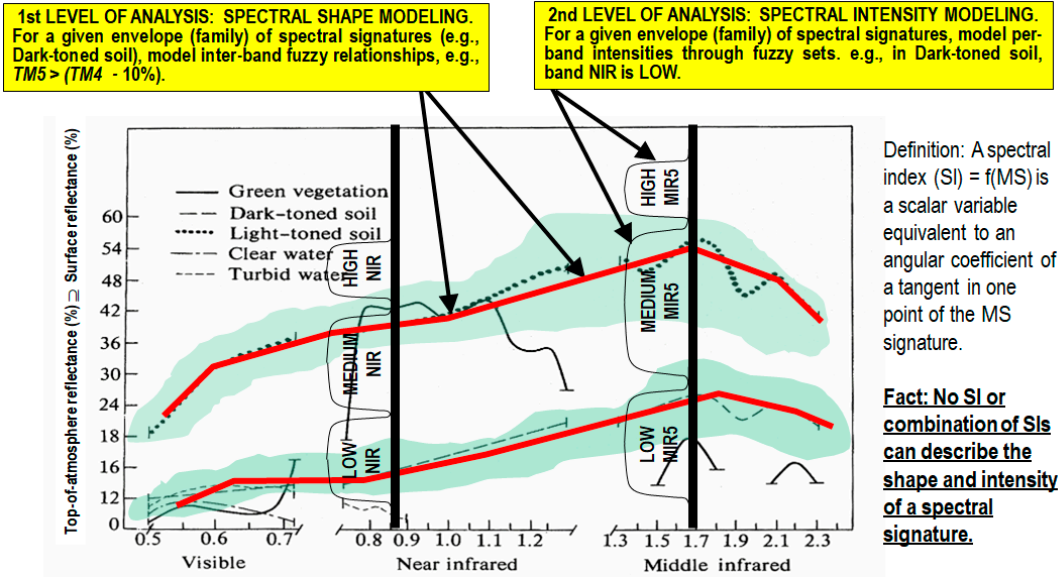
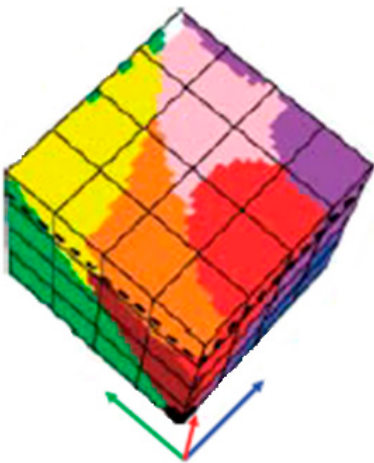



Figure 18. Examples of land cover (LC) class-specific families of spectral signatures [14] in top-of-atmosphere reflectance (TOARF) values, which include surface reflectance (SURF) values as a special case in clear sky and flat terrain conditions [150], i.e., in general, TOARF ⊇ SURF, where TOARF ≈ SURF (depicted as an ideal "noiseless" spectral signature in red) + atmospheric and topographic noise. A within-class family of spectral signatures (e.g., dark-toned soil) in TOARF or SURF values forms a buffer zone (hyperpolyhedron, envelope, manifold, joint distribution), depicted in light green. Like a vector quantity has two characteristics, a magnitude and a direction, any LC class-specific MS manifold is characterized by a multivariate shape and a multivariate intensity information component. In the RS literature, typical prior knowedge-based spectral decision trees for MS reflectance space hyperpolyhedralization into a finite and discrete vocabulary of MS color names, such as Sen2Cor's [9], MAJA's [39, 41], ATCOR's [63, 64], LEDAPS' [120, 121, 124) and LaSRC's [120, 121, 137], typically adopt either a multivariate analysis of spectral indexes or a logical (AND, OR) combination of univariate variables, such as scalar spectral indexes or spectral channels, considered mutually independent. A typical spectral index is a scalar band ratio or band-pair difference equivalent to an angular coefficient of a tangent to the spectral signature in one point. It is well known that infinite functions can feature the same tangent value in one point. In practice, no spectral index or combination of spectral indexes can reconstruct the multivariate shape and multivariate intensity information components of a spectral signature. As a viable alternative to traditional static (non-adaptive to data) spectral rule-based decision trees found in the RS literature, the Satellite Image Automatic Mapper (SIAM)'s prior knowledge-based spectral decision tree [10, 11, 12, 76, 78, 79, 91, 113, 114, 138, 139, 151] adopts a convergence-of-evidence approach to model any target family (ensemble) of spectral signatures, forming a hypervolume of interest in the MS reflectance hyperspace, as a combination of multivariate shape information with multivariate intensity information components. For example, as shown above, typical spectral signatures of dark-toned soils and typical spectral signatures of light-toned soils form two MS envelopes in the MS reflectance hyperspace that approximately share the same multivariate shape information component, but whose pair of multivariate intensity information components does differ.

1208 **Figure 19.** Courtesy of [149]. Unlike a MS reflectance space hyperpolyhedralization difficult to
1209 think of and impossible to visualize when the number of channels is superior to three, an RGB
1210 data cube polyhedralization is intuitive to think of and straightforward to display. For example,
1211 based on psychophysical evidence, human basic color (BC) names can be mapped onto a
1212 monitor-typical RGB data cube. Central to this consideration is Berlin and Kay's landmark
1213 study of a "universal" inventory of eleven BC words in twenty human languages: black, white,
1214 gray, red, orange, yellow, green, blue, purple, pink and brown [148].

1215 In vision, where a (2D) image-domain and a 4D geospatial-temporal scene-domain co-
1216 exist, see Figure 6, a community-agreed discrete and finite vocabulary of MS color names,
1217 pertaining to a MS color space in the image-domain, such as the eleven BC names proposed by
1218 Berlin and Kay in the image-domain of visible bands RGB [148], see Figure 19, should never be
1219 confused with a vocabulary of classes of real-world objects in the scene-domain, such as a
1220 discrete and finite legend of LC classes. Color names provide a categorical (nominal)
1221 representation of the numeric photometric variable associated as attribute to any LC class
1222 belonging to a discrete and finite LC class legend. Encoded as a categorical variable, a color
1223 attribute of a (categorical) LC class should never be confused with the LC class it belongs to.
1224 On the hand, the same color name can be shared by several LC classes. On the other hand, a
1225 single LC class can feature several color names as photometric attribute [11, 12]. For example,
1226 Adams et al. correctly observed that discrete spectral endmembers typically adopted in hyper-
1227 spectral image interpretation "cannot always be inverted to unique LC class names" [152]. This
1228 commonsense knowledge is sketched in Table 6, where set A = DictionaryOfColorNames with
1229 cardinality |A| = a = ColorVocabularyCardinality = 11 and set B = LegendOfObjectClassNames
1230 with cardinality |B| = b = ObjectClassLegendCardinality = 3. Between the vocabulary A of
1231 color names in the image-domain and the vocabulary B of LC classes in the scene-domain there
1232 is a binary relationship, R: A ⇒ B, subset of the 2-fold Cartesian product A × B, i.e., R: A ⇒ B
1233 ⊆ A × B, where A ≠ B in general. The Cartesian product A × B is a set whose elements are
1234 ordered pairs of each instance of set A combined with each instance of set B. Hence, the size of
1235 the Cartesian product A × B is rows × columns = a × b, where a ≠ b in general. Noteworthy,
1236 Cartesian product A × B ≠ FrequencyCount(A × B), where FrequencyCount(A × B) is known as
1237 two-way contingency table, association matrix, cross tabulation, bivariate table, bivariate
1238 frequency table (BIVRFTAB) or confusion matrix [10, 22, 153-157].

1239
1240 **Table 6.** . Example of a binary relationship R: A ⇒ B ⊆ A × B from set A =
1241 DictionaryOfColorNames, with cardinality |A| = a = ColorVocabularyCardinality = 11, and
1242 the set B = LegendOfObjectClassNames, with cardinality |B| = b =
1243 ObjectClassLegendCardinality = 3, where A × B is the 2-fold Cartesian product between sets A
1244 and B. The Cartesian product of two sets A × B is a set whose elements are ordered pairs. The
1245 size of A × B is rows × columns = a × b. The dictionary LegendOfObjectClassNames is a

1246 superset of the typical taxonomy of land cover (LC) classes adopted by the RS community.
1247 "Correct" entry-pairs (marked with √) must be: (i) selected by domain experts based on a *hybrid*
1248 combination of deductive prior beliefs with inductive evidence from data and (ii) community-
1249 agreed upon, to be used by members of the community [11, 12].

			Target classes of individuals (entities in a conceptual model for knowledge representation built upon an ontology language)		
			Class 1, Water body	Class 2, Tulip flower	Class 3, Italian tile roof
Basic color (BC) names	black	■		√	
	blue	■	√	√	
	brown	■	√	√	√
	grey	■			
	green	■	√	√	
	orange	■		√	
	pink	■		√	
	purple	■		√	
	red	■		√	√
	white			√	
	yellow	■		√	

1250
1251   In spite of this commonsense knowledge, see Table 6, and in contrast with the
1252 unquestionable true-fact that, in vision, primary spatial information dominates secondary
1253 color information in both the image-domain and the scene-domain [10-12, 20], see Section 2,
1254 prior knowledge-based spectral decision trees for MS color space hyperpolyhedralization into
1255 color names implemented by the RS community typically assume that the vocabulary A of
1256 color names in the image-domain and the vocabulary B of LC classes in the scene-domain do
1257 coincide in cardinality, semantics and order of presentation. If hypothesis A = B holds true,
1258 then the binary relationship R: A $\Rightarrow$ B $\subseteq$ A $\times$ B $\neq$ FrequencyCount(A $\times$ B) becomes a bijective
1259 (injective and surjective) function, while the two-way confusion matrix FrequencyCount(A $\times$
1260 B) becomes square, where the main diagonal guides the interpretation process [11, 12]. This is
1261 the case of the LC map legend of an SCM product generated as output by a static decision tree
1262 of spectral rules implemented by the ATCOR software toolbox, see Table 3 and Table 4, and
1263 by the Sen2Cor software toolbox, see Table 1, where there is one spectral rule (or OR-
1264 combination of spectral rules) per LC class and vice versa. Since the unrealistic assumption
1265 that categorical variable A = DictionaryOfColorNames, with cardinality |A| = a =
1266 ColorVocabularyCardinality, coincides with categorical variable B =
1267 LegendOfObjectClassNames, with cardinality |B| = b = ObjectClassLegendCardinality, is
1268 undertaken at the abstract level of system understanding of information/knowledge
1269 representation, considered the linchpin of success of any information processing system [10,
1270 16, 69], it heavily affects all subsequent levels of abstraction, specifically, system design,
1271 algorithm and implementation, in agreement with well-known information principle known
1272 as garbage in, garbage out (GIGO), synonym of error propagation through an information
1273 processing chain [11, 12]. This error propagation effect affecting CV $\supset$ EO-IU applications
1274 increases as the imaging sensor's spatial resolution becomes finer, such as in EO high spatial
1275 resolution (HR, in range [1 m, 30 m]) and very high spatial resolution (VHR, < 1 m) image
1276 understanding tasks [14].
1277   In addition to the unrealistic assumption that vocabularies of color names and of LC class
1278 names do coincide at the level of system understanding of information/knowledge
1279 representation, the implementation of the Sen2Cor's spectral rules for MS reflectance space
1280 hyperpolyhedralization into MS color names appears inadequate to define mutually exclusive
1281 and totally exhaustive hyperpolyhedra, where each hyperpolyhedron is a multivariate data
1282 distribution (joint probability).
1283   First, Sen2Cor models each joint probability by a single spectral rule as an and/or-
1284 combination of univariate distributions. It assumes a joint distribution is equivalent to a
1285 product of univariate distributions, which holds true if and only if univariate variables are

mutually independent. This assumption is completely unrealistic when modeling manifolds of LC class-specific spectral signatures sampled at a discrete and finite set of spectral bands. These spectral band values are not statistically independent because they are LC class-dependent. In general, multivariate data statistics are more informative than a combination of univariate data statistics. For example, maximum likelihood data classification, accounting for multivariate data correlation and variance (covariance), is typically more accurate than parallelepiped data classification whose rectangular decision regions, equivalent to a concatenation of univariate data constraints, poorly fit multivariate data in the presence of bivariate cross-correlation [24]. As a typical example of this critical point, see Table 4, showing the SPECL decision tree instantiation proposed by the ATCOR software toolbox, where each spectral rule is implemented as a sequence of univariate statistics.

Second, in the Sen2Cor's spectral rule-based decision tree a spectral rule implementation typically adopts many spectral indexes in the definition of each hyperpolyhedron equivalent to a MS color name. Like a vector quantity has two characteristics, a magnitude and a direction, any LC class-specific MS manifold is characterized by a multivariate shape and a multivariate intensity information component, see Figure 18. A spectral index, implemented as either a band difference or a band ratio, is conceptually equivalent to the angular coefficient of a tangent to the spectral signature in one point. It is well known that infinite functions can pass through the same point with the same angular coefficient. It means that, although appealing due to its conceptual and numerical simplicity [2], any scalar spectral index, equivalent to a spectral slope, is a MS shape descriptor independent of the MS intensity. It is unable *per se* to represent either the multivariate shape information component or the multivariate intensity information component of a MS signature [10]. In other words, no spectral index or combination of spectral indexes provides a lossless reconstruction of the multivariate shape and multivariate intensity information components of a spectral signature [10-12]. Because of this spectral information loss, the number of scalar spectral indexes proposed in the RS literature is ever increasing [2], in the unrealistic attempt to extract a more informative scalar variable from a multivariate spectral signature. In some published works, the misuse of spectral indexes reaches its peak when, to simplify a multivariate spectral signature both conceptually and numerically, the number of univariate spectral indexes extracted from the spectral signature is superior to the number of spectral bands; in this no-win situation, no MS data compression is accomplished while a loss in spectral information is guaranteed. As a typical example of this second critical point, see Table 4, showing the SPECL decision tree instantiation proposed by the ATCOR software toolbox where several band ratios are employed for modeling a MS hyperpolyhedron of interest in the MS data hyperspace.

4. Methods

To contribute toward the RTD of the first (never accomplished to date) CV system in operating mode capable of systematic ESA EO Level 2 product generation from multi-sensor, multi-temporal and multi-angular EO *big data* cubes (see Section 1), where conceptual drawbacks and failure modes of available CV subsystems for cloud and cloud-shadow detection in EO optical imagery, such as Sen2Cor and MAJA, must be overcome (see Section 3), an inherently ill-posed AutoCloud+ CV subsystem development [10, 36, 67] was constrained in accordance to the following system requirements specification, at the levels of system understanding of information/knowledge representation, system architecture, algorithm and implementation in the multidisciplinary domain of cognitive science (see Figure 5).

- Combined deductive and inductive (hybrid) inference, where deductive and inductive inference alternate at increasing levels of specialization, see Figure 12, to take advantage of each and overcome their shortcomings [10, 76, 78, 79].

- 2D (retinotopic, spatial topology-preserving) image analysis, see Figure 16, where primary spatial information and secondary color information in vision are combined according

1337 to a convergence-of-evidence approach, consistent with human symbolic reasoning, where
1338 eventually weak, but independent sources of evidence can be combined to infer strong
1339 conjectures, and where it is acknowledged that "vision goes symbolic almost immediately"
1340 [10, 69].

1341 • Provided with feedback loops (feedback system) [89, 97-102, 181]. A hybrid feedback CV
1342 system is alternative to inductive feedforward CV systems, such as popular DCNNs [31],
1343 deductive feedforward CV systems, such as Sen2Cor, see Figure 11 and Table 5, and hybrid
1344 feedforward CV systems, such as the "augmented" ATCOR software toolbox, see Figure 12.

1345 • Suitable for cloud and cloud-shadow quality layer detection in MS imagery under
1346 operational constraints of "universality". To be considered "universal" in operating mode
1347 (refer to Section 2), AutoCloud+ was required to be: (i) "fully automated", i.e., it requires no
1348 human-machine interaction to run and no labeled data set for supervised inductive learning-
1349 from-data, (ii) near real-time, e.g., its computational complexity increases linearly with image
1350 size, (iii) robust to changes in MS imaging sensor's spatial and spectral resolution
1351 specifications, and (iv) robust to changes in radiometric *Cal* metadata parameters, i.e.,
1352 AutoCloud+ can be input with multi-source MS imagery whether or not radiometrically *Cal*
1353 into TOARF, SURF or surface albedo values, such as EO image acquired by multiple platforms,
1354 either spaceborne or airborne, including unmanned aerial vehicles (UAVs), whose lightweight
1355 imaging sensors typically do not provide any radiometric *Cal* metadata file.

1356 At the level of system understanding of information/knowledge representation, the
1357 AutoCloud+ CV subsystem adopts a convergence-of-evidence approach equivalent to a
1358 Bayesian naïve classifier. According to the Bayesian law, a naive Bayes classifier assumes the
1359 "naive" conditional independence of input features F_i, $i = 1, \ldots, I$, hence,

1360 $$p(c|F_i, \ldots, F_I) = p(c) \prod_{i=1}^{I} p(F_i|c).\tag{1}$$

1361 Deeply investigated by the CV community [16, 69, 88] and by those portions of the RS
1362 community involved with traditional content-based image retrieval (CBIR) [87, 158] and with
1363 EO-IU applications for Earth surface variables estimation [10-12, 20, 78, 79, 159-161], well-
1364 known visual features are: (i) color values, typically discretized by humans into a finite and
1365 discrete vocabulary of BC names [148, 149]; (ii) planar shape [10, 160-163]; (iii) texture, defined
1366 as the perceptual spatial grouping of texture elements known as texels [10, 16, 164-166] or
1367 tokens [69]; (iv) inter-object spatial topological relationships, e.g., adjacency, inclusion, etc., and
1368 (v) inter-object spatial non-topological relationships, e.g., spatial distance, angle measure, etc.
1369 [16-21]. In vision, color is the sole visual property available at the imaging sensor's spatial
1370 resolution, i.e., at pixel level. In other words, pixel-based information is spatial context-
1371 independent, i.e., per-pixel information is exclusively related to color properties. Among the
1372 aforementioned visual variables, per-pixel color values are the sole non-spatial (spatial context-
1373 insensitive) numeric variable. It is easy to prove that, irrespective of their Pearson inter-feature
1374 cross-correlation, if any, individual sources of visual evidence, specifically, color, local shape,
1375 texture and inter-object spatial relationships, are statistically independent because, in general,
1376 Pearson's linear cross-correlation does not imply causation [10, 16, 27, 30, 57, 58]. In the CV
1377 domain, if priors are ignored because considered equiprobable in a maximum class-conditional
1378 likelihood inference approach, alternative to a maximum *a posteriori* optimization criterion, and
1379 if a canonical interpretation based on frequentist statistics can be relaxed by fuzzy logic [167]
1380 into a membership function m(·) ∈ [0, 1], then the naive Bayes classifier, see Equation (1),
1381 becomes [10-12]:

1382 p(c| ColorValue(x), ShapeValue(x), TextureValue(x), SpatialRelationships(x,
1383 Neigh(x))) =
1384 $p(c|F_i, \ldots, F_{I=4}) = p(c) \prod_{i=1}^{I=4} p(F_i|c) \propto$
1385 m(c| ColorValue(x), ShapeValue(x), TextureValue(x), SpatialRelationships(x,
1386 Neigh(x))) ∝

1387 $\quad$ $\min\{\sum_{ColorName=1}^{ColorVocabularyCardinality} m_1(ColorValue(x)|ColorName) \cdot m_2(ColorName|c),$

1388 $\quad$ $m_3(ShapeValue(x)|c), m_4(TextureValue(x)|c), m_5(SpatialRelationships(x,$

1389 $\quad$ $Neigh(x))|c)\} =$

1390 $\quad$ $\min\{m_2(ColorName^*|c), m_3(ShapeValue(x)|c), m_4(TextureValue(x)|c),$

1391 $\quad$ $m_5(SpatialRelationships(x, Neigh(x))|c)\}, c = 1, …, ObjectClassLegendCardinality,$

1392 $\quad$ where ColorName* ∈ {1, ColorVocabularyCardinality}, such that $m_1(ColorValue(x)|$

1393 $\quad$ ColorName*) = 1 and $m_2(ColorName^*|c) \in \{0, 1\}$, $\hfill$ (2)

1394 where x is a spatial unit in the (2D) image-domain, either 0D point, 1D line or 2D polygon [168],
1395 Neigh(x) is a generic 2D spatial neighborhood of spatial unit x, colorValue(x) belongs to a MS
1396 measurement space $\Re^{MS}$, i.e., $ColorValue(x) \in \Re^{MS}$, and color space $\Re^{MS}$ is partitioned into a
1397 set of mutually exclusive and totally exhaustive hyperpolyhedra, equivalent to a discrete and
1398 finite vocabulary (categorical variable) of static color names, with ColorName = 1, ..,
1399 ColorVocabularityCardinality. About Equation (2), the following considerations hold.

1400 $\quad$ ➢ $\quad$ Each numeric ColorValue(x) is one point in a MS color space $\Re^{MS}$. In the MS color space,
1401 $\quad\quad$ each point belongs to a single color name (hyperpolyhedron), identified as ColorName* in
1402 $\quad\quad$ the static color vocabulary, i.e., $\forall$ $ColorValue(x) \in \Re^{MS}$, then
1403 $\quad\quad$ equality $\sum_{ColorName=1}^{ColorVocabularyCardinality} m_1(ColorValue(x)|ColorName) =$
1404 $\quad\quad$ $m_1(ColorValue(x)|ColorName^*) = 1$ holds, where $m_1(ColorValue(x)|ColorName) \in \{0, 1\}$
1405 $\quad\quad$ is a binary (crisp) membership function, with ColorName = 1, …,
1406 $\quad\quad$ ColorVocabularyCardinality.

1407 $\quad$ ➢ $\quad$ Equation (2) shows that any convergence-of-evidence approach is more selective than each
1408 $\quad\quad$ individual source of evidence, in line with a focus-of-visual attention mechanism [88].

1409 $\quad\quad$ ➢ $\quad$ In Equation (2), a vocabulary of color names, physically equivalent to a partition of
1410 $\quad\quad$ a numeric color hyperspace $\Re^{MS}$ into color name-specific hyperpolyhedra, is conceptually
1411 $\quad\quad$ equivalent to a latent/ hidden/ hypothetical categorical variable, see Figure 20. In statistics,
1412 $\quad\quad$ the popular concept of latent/hidden categorical variable was introduced to fill the
1413 $\quad\quad$ information gap from input numeric observables, e.g., sub-symbolic sensory data such as
1414 $\quad\quad$ color values in a color space, to an output categorical variable, such as a discrete and finite
1415 $\quad\quad$ dictionary of LC classes in the mental model of the physical world (world ontology, world
1416 $\quad\quad$ model). Latent/hidden variables are categorical variables not directly measured, but
1417 $\quad\quad$ inferred from lower level variables. "The terms hypothetical variable or hypothetical
1418 $\quad\quad$ construct may be used when latent variables correspond to abstract concepts, like
1419 $\quad\quad$ perceptual categories or discrete mental states" [10, 169, 170]. Hence, to fill the semantic
1420 $\quad\quad$ gap from input sub-symbolic sensory data to an output categorical variable of symbolic
1421 $\quad\quad$ quality, an hypothetical categorical variable, such as a discrete and finite dictionary of BC
1422 $\quad\quad$ names [148, 149], is considered to be of "semi-symbolic" quality, i.e., superior to zero
1423 $\quad\quad$ semantics, but inferior to the symbolic quality of the output categorical variable.

1424

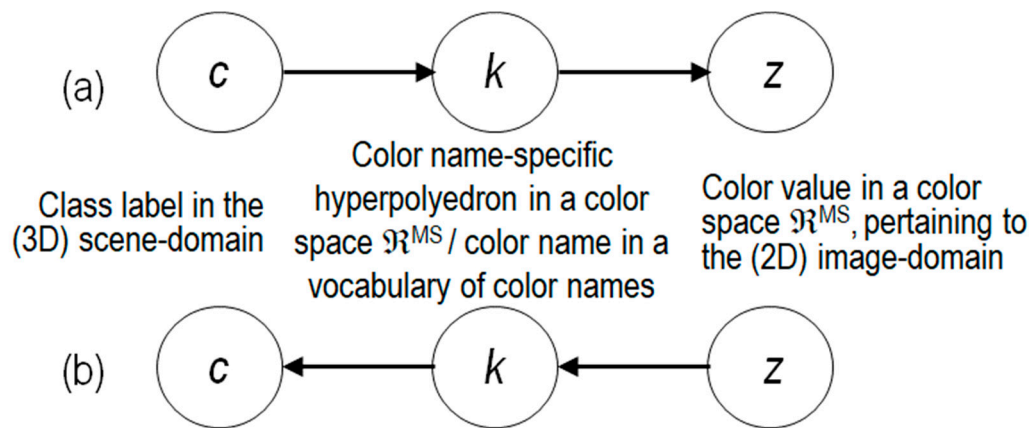
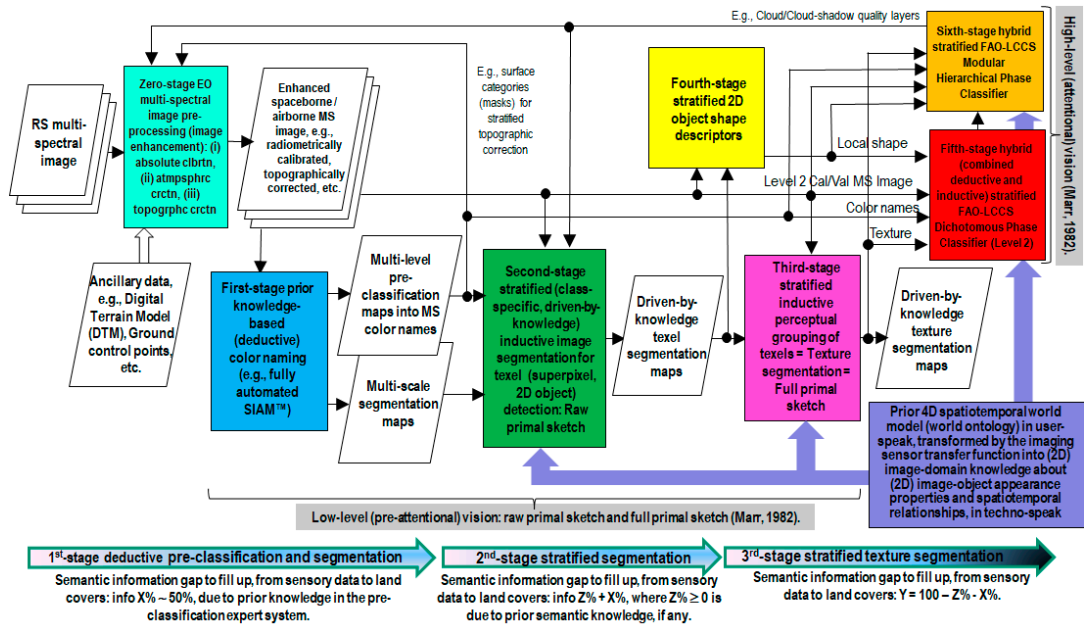



Figure 20. Graphical model of color naming, adapted from [169]. Let us consider z as a (sub-symbolic) numeric variable, such as MS color values of a population of spatial units, with vector data $z \in \mathfrak{R}^{MS}$, where $\mathfrak{R}^{MS}$ represents the MS data space, while c represents a categorical variable of symbolic classes in the physical world, with c = 1, …, ObjectClassLegendCardinality. (a) According to Bayesian theory, posterior probability p(c|z) $\propto$ p(z|c)p(c) = p(c) $\sum_{ColorName=k=1}^{ColorVocabularyCardinality}$ p(z|k)p(k|z), where color names, equivalent to color hyperpolyhedra in a numeric color space $\mathfrak{R}^{MS}$, provide a partition of the domain of change, $\mathfrak{R}^{MS}$, of numeric variable z. (b) For discriminative inference, the arrows in the graphical model are reversed using Bayes rule. Hence, a vocabulary of color names, physically equivalent to a partition of a numeric color space $\mathfrak{R}^{MS}$ into color name-specific hyperpolyhedra, is conceptually equivalent to a latent/ hidden/ hypothetical variable linking observables (sub-symbolic sensory data) in the real world, specifically, color values, to a categorical variable of semantic (symbolic) quality in the mental model of the physical world (world ontology, world model).

➢ Equation (2) shows that for any spatial unit x in the image-domain, when a hierarchical CV classification approach estimates posterior m(c| ColorValue(x), ShapeValue(x), TextureValue(x), SpatialRelationships(x, Neigh(x))) starting from an *a priori* knowledge-based near real-time color naming first stage, where condition m_1(ColorValue(x)| ColorName*) = 1 holds, if condition m_2(ColorName*| c) = 0 is true according to a static community-agreed binary relationship R: VocabularyOfColorNames $\Rightarrow$ LegendOfObjectClassNames (and vice versa) known *a priori* at first stage, see Table 6, then the output membership function m(c| ColorValue(x), ShapeValue(x), TextureValue(x), SpatialRelationships(x, Neigh(x))) = 0 irrespective of any second-stage assessment of spatial terms ShapeValue(x), TextureValue(x) and SpatialRelationships(x, Neigh(x)), whose computational model is typically difficult to find and computationally expensive. Intuitively, Equation (2) shows that first-stage static color naming of any spatial unit x in the image-domain, either (0D) pixel, (1D) line or (2D) polygon, allows the color-based stratification of unconditional multivariate spatial variables into color class-conditional multivariate data distributions, in agreement with the statistic stratification principle [130] and the divide-and-conquer (*dividi-et-impera*) problem solving approach [29, 30, 32]. Well known in statistics, the principle of statistic stratification guarantees that "stratification will always achieve greater precision provided that the strata have been chosen so that members of the same stratum are as similar as possible in respect of the characteristic of interest" [130].

At the levels of understanding of system design, algorithm and implementation, the AutoCloud+ CV subsystem requirements specification listed above is ambitious, but realistic, because AutoCloud+ can rely upon several CV ⊃ EO-IU software units (modules) already implemented, tested (by their authors) and validated (by independent third-parties) in recent years [10-12]. These existing CV ⊃ EO-IU software units can be combined according to an

1464   original six-stage hybrid feedback EO-IU system architecture, hereafter identified as
1465   QuickMap™ technology, see Figure 21, in compliance with the engineering principles of
1466   modularity, hierarchy and regularity considered mandatory by structured system design to
1467   guarantee scalability [32]. The six-stage hybrid feedback CV ⊃ EO-IU system shown in Figure
1468   21 constitutes the core of the inference engine required by the closed-loop EO-IU4SQ system
1469   to systematically provide sub-symbolic EO *big data* cubes with meanings (semantics,
1470   intelligence), to transform a traditional DIAS into an innovative AI4DIAS framework, see
1471   Figure 4 [10, 82-86].
1472

1473   **Figure 21.** Six-stage hybrid (combined deductive and inductive) feedback EO image
1474   understanding (EO-IU) system design, identified as QuickMap™ technology. It is based on a
1475   convergence-of-evidence approach to vision, consistent with Bayesian naïve classification
1476   ([10]). Alternative to inductive feedforward EO-IU system architectures adopted by the RS
1477   mainstream, such as Deep Convolutional Neural Networks (DCNNs) [31], it supports a
1478   hierarchical approach to a first-stage general-purpose, sensor-, application- and user-
1479   independent EO image understanding (classification) subsystem, followed by a second-stage
1480   sensor-, application- and user-specific EO image understanding subsystem, consistent with
1481   the standard two-stage fully-nested Land Cover Classification System (LCCS) taxonomy
1482   promoted by the Food and Agriculture Organization (FAO) of the United Nations, where a
1483   first-stage 3-level 8-class Dichotomous Phase (DP) is preliminary to a second-stage Modular
1484   Hierarchical Phase (MHP) [13]. For the sake of visualization, each of the six data processing
1485   stages plus stage-zero for EO data pre-processing (enhancement) is depicted as a rectangle
1486   with a different color fill. Visual evidence stems from multiple information sources,
1487   specifically, numeric color values quantized into categorical color names, local shape, texture
1488   and inter-object spatial relationships, either topological or non-topological. An example of
1489   first-stage general-purpose, user- and application-independent EO image classification
1490   taxonomy required by an ESA EO Level 2 Scene Classification Map (SCM) product is the 3-
1491   level 8-class FAO LCCS-DP legend, in addition to quality layers cloud and cloud-shadow.
1492   Second-stage EO image classification is user- and application-specific, where an SCM product
1493   of Level 3 or superior is provided with a map legend consistent with the FAO LCCS-MHP
1494   taxonomy [13], see Figure 1. In this figure, acronym SIAM stays for Satellite Image Automatic
1495   Mapper (SIAM), a lightweight computer program for MS reflectance space
1496   hyperpolyhedralization into a static vocabulary of MS color names, superpixel detection and
1497   vector quantization (VQ) quality assessment [10, 11, 12, 76, 78, 79, 91, 113, 114, 138, 139, 151].

1498   At the levels of system understanding of algorithm and implementation, the six-stage
1499 hybrid feedback CV ⊃ EO-IU system architecture shown in Figure 21 employs the following
1500 ensemble of existing CV software functions.

1501 (i)   Zero-stage EO image pre-processing (enhancement), to guarantee multi-source multi-
1502   temporal and multi-angular EO image harmonization and interoperability [10].

1503   I.   Required input data set: MS image provided with a radiometric *Cal* metadata file,
1504   in agreement with the GEO-CEOS QA4EO *Cal* requirements [3].

1505   a. Absolute radiometric *Cal* of DNs into TOARF values, based on
1506   radiometric *Cal* gain and offset metadata parameters.

1507   b. Automated stratified atmospheric, stratified adjacency and stratified
1508   topographic correction (STRATCOR) of TOARF into SURF values, see
1509   Figure 22, e.g., refer to [10, 76].

1510

(a)   (b)

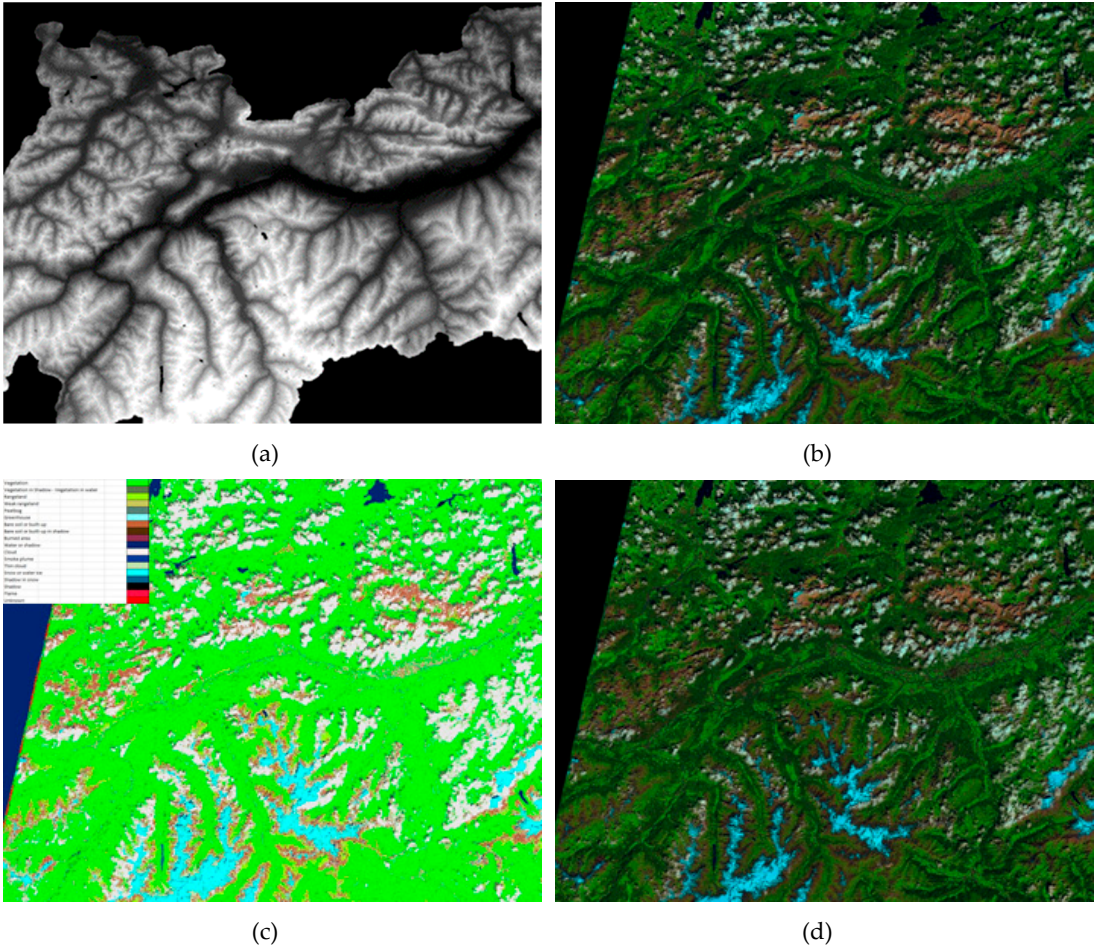
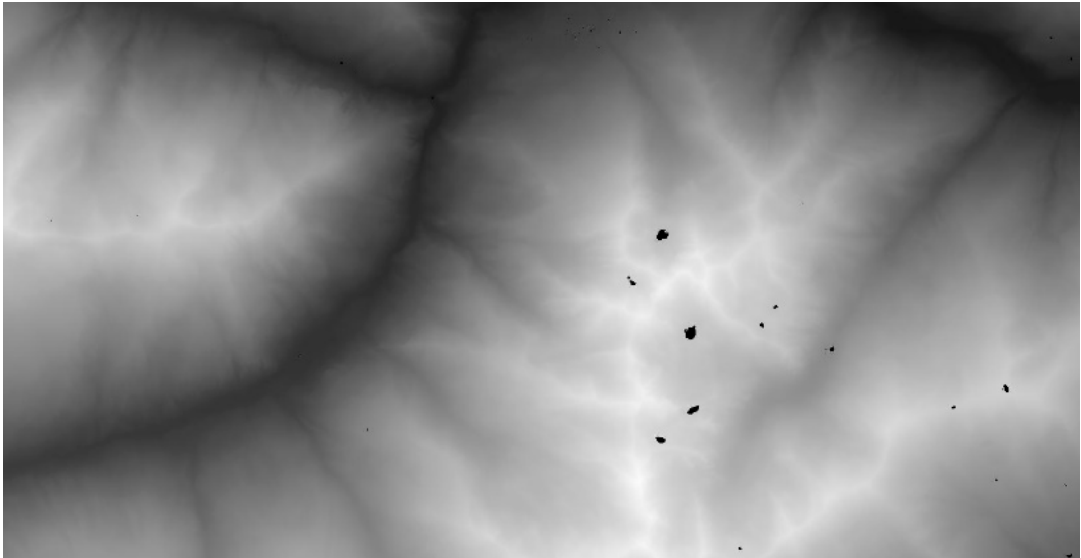
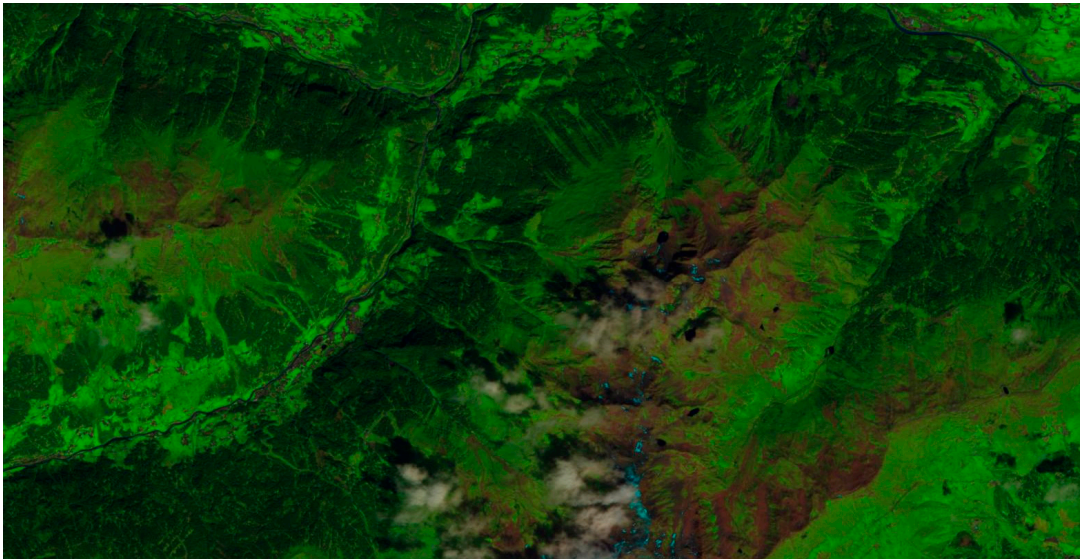
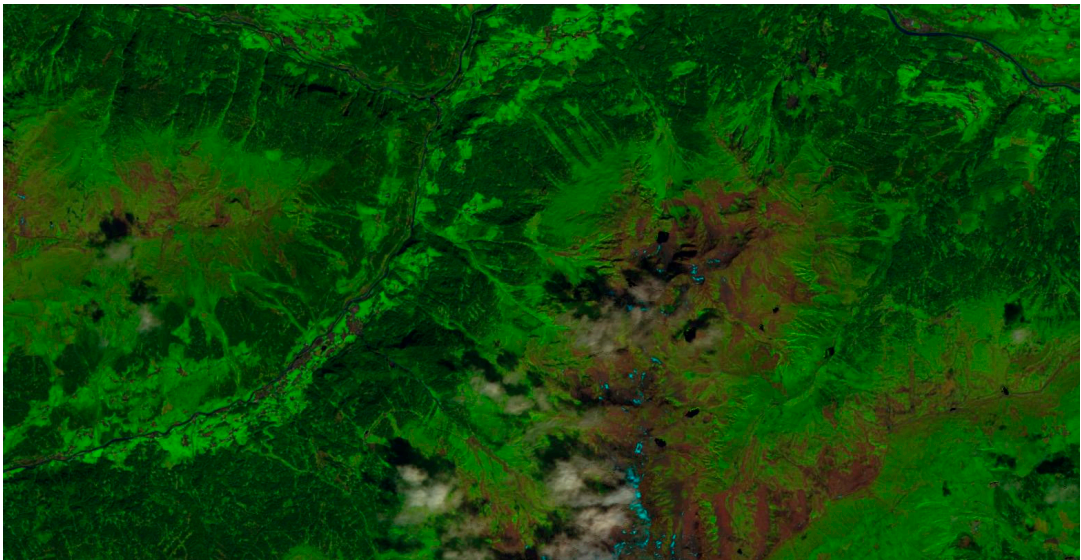
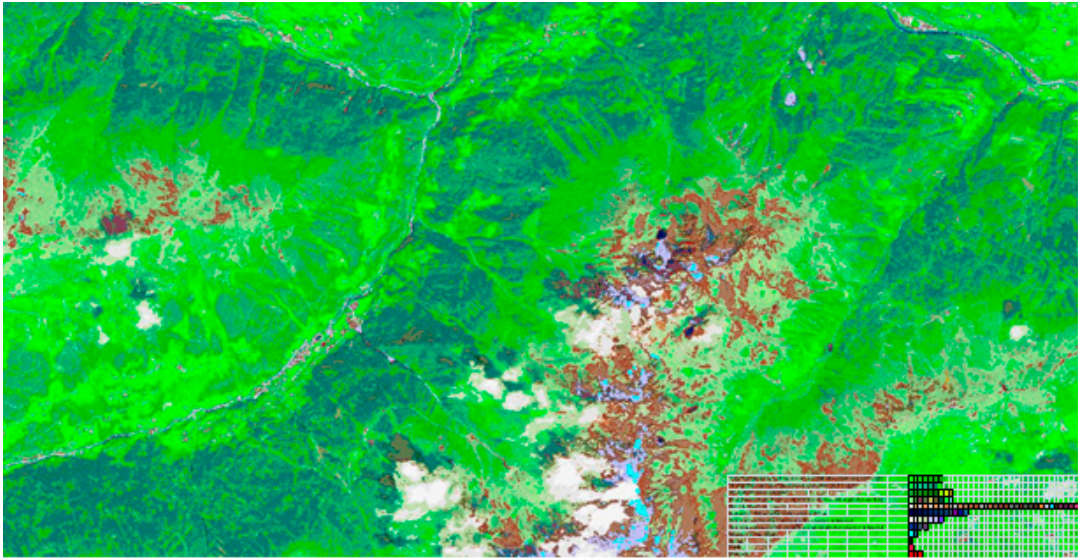
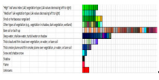
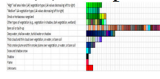
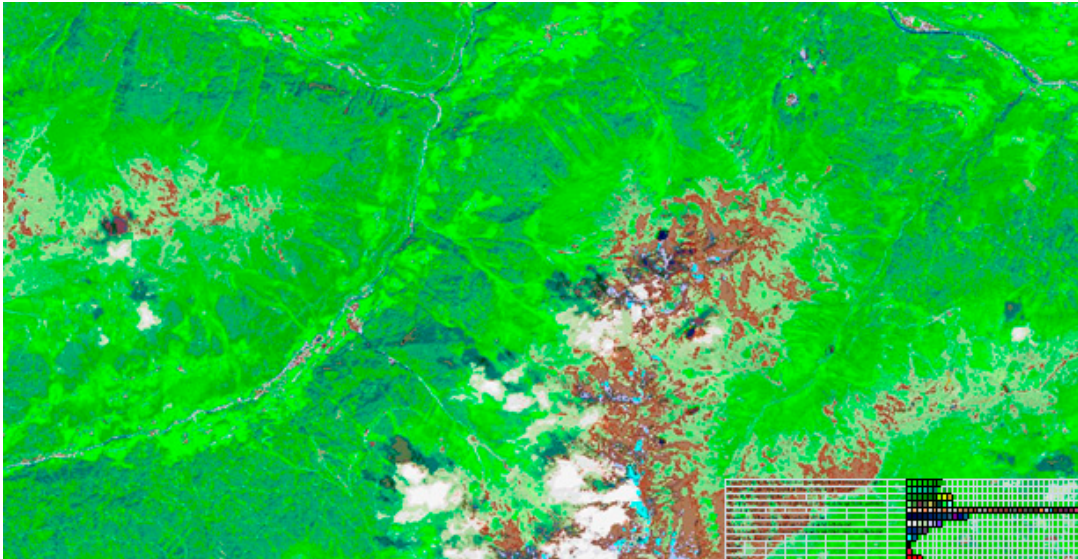
(c)   (d)

(e)

(f)

(g)

(h)

(i)

Figure 22. Automated stratified topographic correction (STRATCOR) of TOARF into SURF values, see Figure 23, e.g., refer to [10, 76]. (a) Input data set 1: Digital Terrain Model (DTM) of Tirol, Austria. 10 m resolution. (b) Input data set 2: Sentinel-2 Multi-Spectral Instrument (MSI) image of Austria (acquisition date: 2016-08-07), depicted in false colors (R: band MIR, G: band NIR, B: band Blue), 10 m resolution, calibrated into TOARF values. No histogram stretching is applied for visualization purposes. (c) Input data set 3: Satellite Image Automatic Mapper (SIAM) output map in color names [10, 11, 12, 76, 78, 79, 91, 113, 114, 138, 139, 151], automatically generated from the input Sentinel-2 image in TOARF values. Output map legend at coarse granularity, 18 spectral categories, depicted in false colors: (d) STRATCOR output product. Sentinel-2 MSI image of Austria, depicted in false colors (R: band MIR, G: band NIR, B: band Blue), 10m resolution, radiometrically calibrated into TOARF values and automatically corrected for topographic effects. No histogram stretching is applied for visualization purposes. (e) Zoom-in of the DTM shown in (a). (f) Zoom-in of the input Sentinel-2 image in TOARF values shown in (b). (g) Zoom-in of the STRATCOR output image shown in (d). To be compared with the zoomed-in input image shown in (f). (h) Zoom-in of the SIAM output map in color names, automatically generated from the input Sentinel-2 image in TOARF values shown in (b). Output map legend at coarse granularity, 96 spectral

1528 categories, depicted in false colors: . (i) Zoom-in of the SIAM output map in
1529 color names, automatically generated from the STRATCOR output image in TOARF values
1530 shown in (d). Output map legend at coarse granularity, 96 spectral categories, depicted in false

1531 colors: . To be compared with the zoomed-in SIAM map generated from the
1532 input image, shown in (h).This is an example of the radiometric correction approach sketched
1533 in Figure 12, where preliminary classification and correction stages alternate in a hierarchical
1534 sequence.

1535   II.   Required input data set: panchromatic image or MS image provided with no
1536   radiometric *Cal* metadata file. In this case, an inherently ill-posed color constancy
1537   algorithm must be run to guarantee input data interoperability through time,
1538   space and sensors, see Figure 23. In human vision, color constancy ensures that
1539   the perceived color of objects remains relatively constant under varying
1540   illumination conditions, so that they appear identical to a "canonical" (reference)
1541   image subject to a "canonical" (known) light source (of controlled quality), e.g.,
1542   under a white light source [171]. In short, solution of the color constancy problem
1543   is the recovery "of an illuminant-independent representation of the reflectance
1544   values in a scene" [172]. In practice, color constancy supports image
1545   harmonization and interoperability when a radiometric *Cal* metadata file is not
1546   available to transform dimensionless DNs into a physical unit of radiometric
1547   measure. Computational color constancy is an under-constrained problem in the
1548   Hadamard sense [171, 173], hence it is difficult to solve. Since it does not have a
1549   unique solution, color constancy requires *a priori* knowledge in addition to sensory
1550   data to become better conditioned for numerical solution [30]. Unfortunately,
1551   biophysical mechanisms of color constancy remain largely unknown to date.
1552   Hence, inherently ill-posed computational color constancy cannot be constrained
1553   by prior knowledge of biological color constancy mechanisms to become better
1554   posed for numerical solution. For these reasons, there has been a large number of
1555   alternative color constancy algorithms proposed in the CV literature in the last 30
1556   years [171, 173]. Inspired by human vision, a novel self-organizing statistical
1557   algorithm for multi-band image color constancy was recently implemented, as
1558   reported in [10, 174, 175].

1559

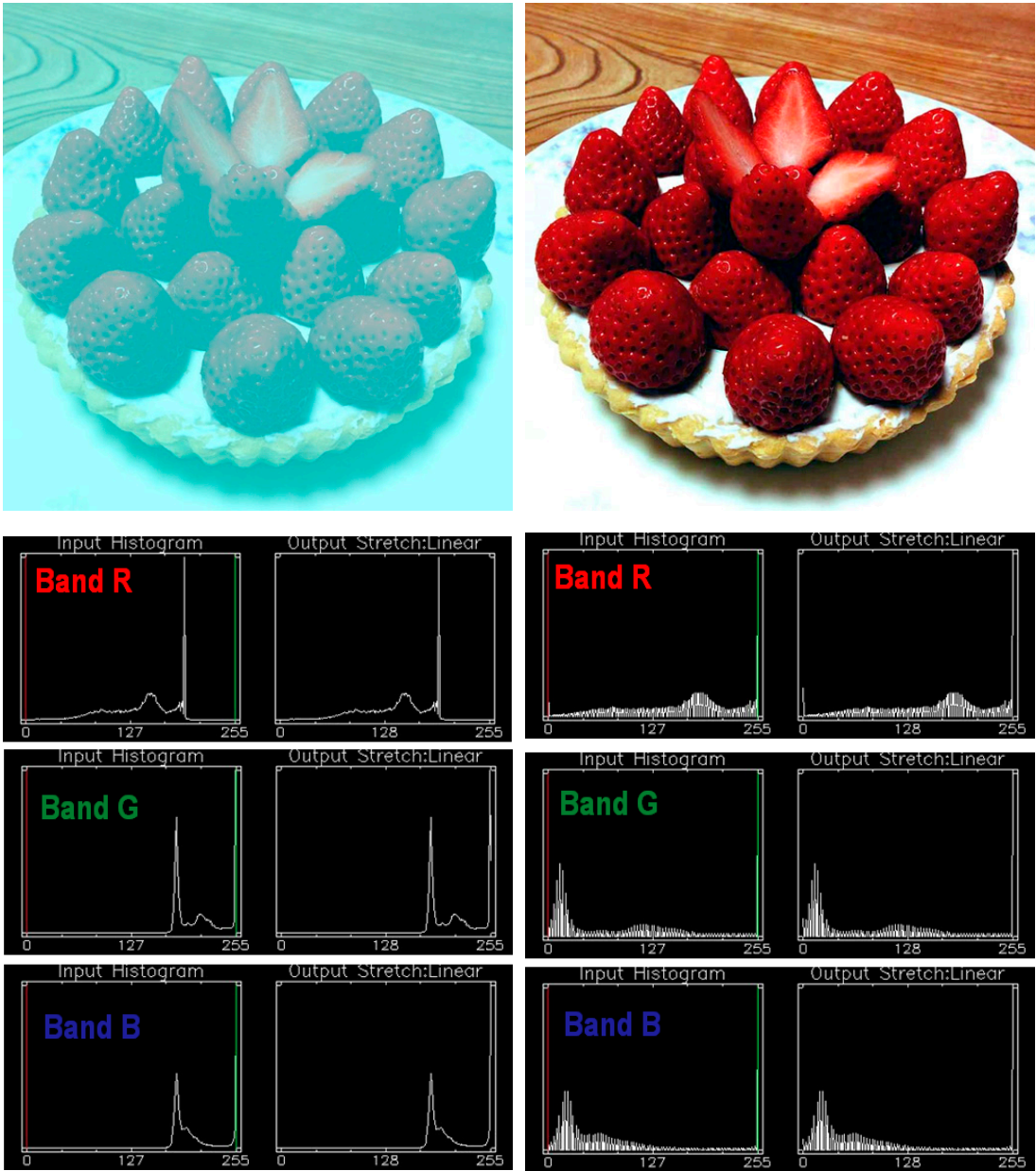



Figure 23. Top left: RGB image, source: Akiyoshi Kitaoka @AkiyoshiKitaoka, web page: http://nymag.com/selectall/2017/02/strawberries-look-red-without-red-pixels-color-constancy.html. Strawberries appear to be reddish, though the pixels are not. No histogram stretching is applied for visualization purposes, see the monitor-typical RGB input-output histograms shown at bottom left. Top right: Output of the self-organizing statistical model-based color constancy algorithm, as reported in [10, 174, 175], input with the image shown top left. No histogram stretching is applied for visualization purposes, see the monitor-typical RGB input-output histograms shown at bottom right.

(ii)　First-stage prior knowledge-based color naming.

　　　I.　Required input data set: MS image radiometrically *Cal* into TOARF, SURF or surface albedo values. Proposed to the RS community in recent years, the Satellite Image Automatic Mapper™ (SIAM™) is a lightweight computer program for automated near real-time MS reflectance space hyperpolyhedralization into MS color names, superpixel detection and vector quantization (VQ) quality assessment [10, 11, 12, 76, 78, 79, 91, 113, 114, 138, 139, 151], see Figure 24, Figure

1575   25 and Figure 26. SIAM claims its scalability to MS imaging sensors featuring
1576   different spectral resolution specifications, see Table 7 and Table 8. Moreover, the SIAM's
1577   spectral knowkedge-based decision tree for MS reflectance space hyperpolyhedralization
1578   outperforms its counterparts in terms of spectral quantization capability at different
1579   quantization levels, parameterized by the total number of detected color names at different
1580   levels of color gramularity, see Table 7 in comparison with Table 1 for Sen2Cor, Table 3 for
1581   ATCOR and Table 4 for ATCOR-SPECL. To accomplish a superior scalability to changes in
1582   sensor's specifications and a superior robustness to changes in input data, ranging from
1583   TOARF to SURF and surface albedo values, SIAM features a superior degree of redundancy
1584   of its multivariate spectral rules [11, 12]. For example, among the spectral knowledge-based
1585   decision trees under comparison, the SIAM decision tree is the sole to adopt two different
1586   sets of spectral rules to model the multivariate shape and the multivariate intensity
1587   components of a target manifold (hypervolume) of MS signatures, see Figure 18.

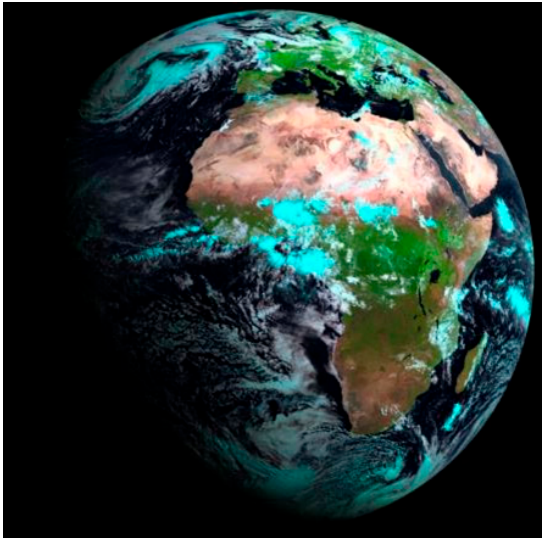
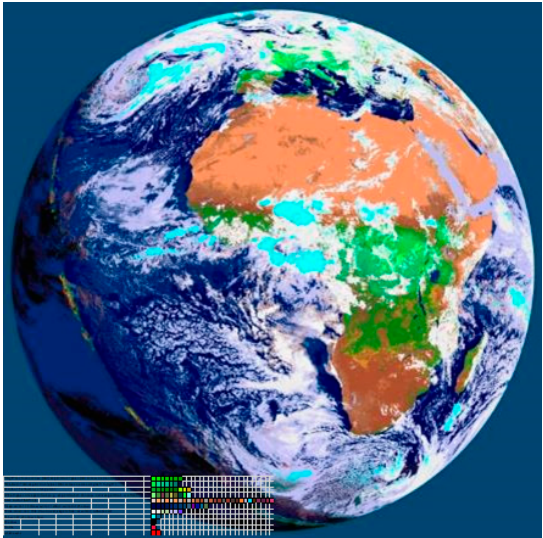
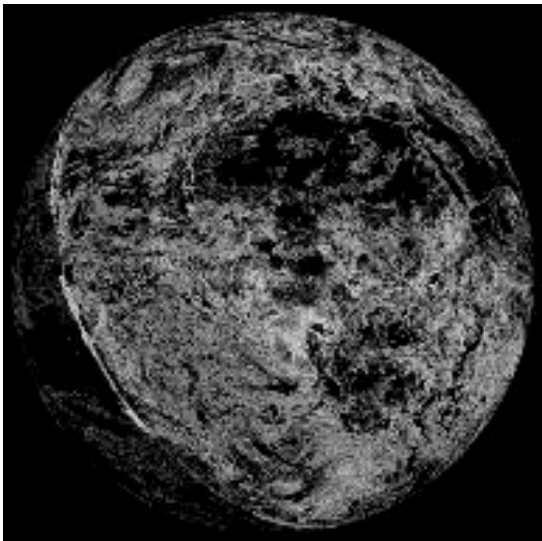
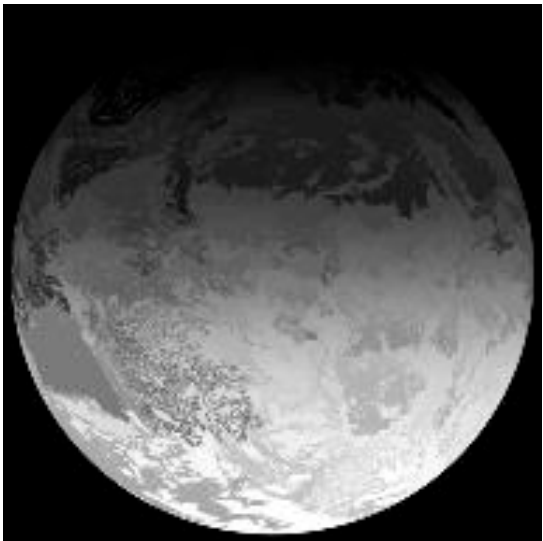
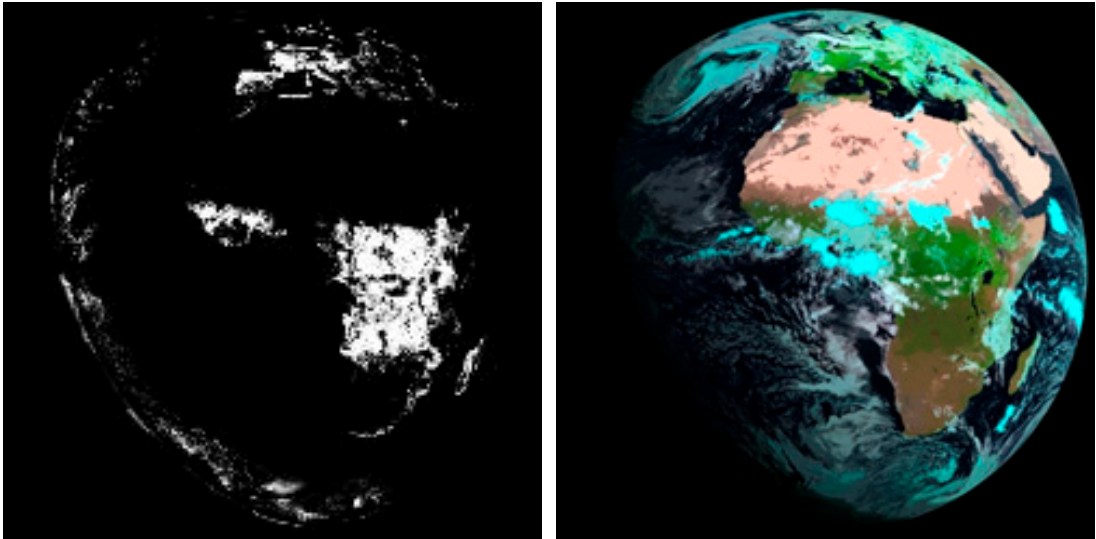
1588

(a)   (b)

(c)   (d)

(e)   (f)

Figure 24. (a) Meteosat Second Generation (MSG) SEVIRI image acquired on 2012-05-30, radiometrically calibrated into TOARF values and depicted in false colors (R: band MIR, G: band NIR, B: band Blue), spatial resolution: 3 km. No histogram stretching is applied for visualization purposes. (b). Advanced Very High Resolution Radiometer (AVHRR)-like SIAM (AV-SIAM) hyperpolyhedralization of the MS reflectance hyperspace and prior color map of the input MS image. The AV-SIAM map legend, consisting of 83 spectral categories, see Table 7, is depicted in pseudocolors. Map legend, similar to Table 8: (c) To visualize contours of image-segments automatically detected in the multi-level SIAM color map-domain by a deterministic two-pass connected-component multi-level image labeling algorithm [16, 176], see Figure 25, an automatic 4- or 8-adjacency cross-aura measure is estimated in linear time, see Figure 26. (d) Segmentation map deterministically generated from the SIAM multi-level output map shown in (b). Each segment is identified by a monotonically increasing (from top to bottom) integer number. (e). Vegetation binary mask, automatically generated from the SIAM multi-level output map shown in (b). On the left, pixel candidates for vegetation belong to spectral category: vegetation in shadow. (f) Piecewise-constant image reconstruction, where each pixel is replaced by the mean reflectance value of the segment that pixel belongs to ("object-mean view"). If the color quantization error (equal to the per-pixel absolute difference between the input image and the piecewise-constant image reconstruction) is "low", then the quality of the prior knowledge-based SIAM's color space partitioning is "high". No histogram stretching is applied for visualization purposes.

Multi-level image, e.g., 3-level thematic map with map legend {1. Water, Acronym = W, Pseudocolor = Blue; 2. Vegetation, Acronym = V, Pseudocolor = Green; 3. Bare soil, Acronym = BS, Pseudocolor = Brown}.

Segmentation map = {S1, ..., S9}. Acronym S means segment.

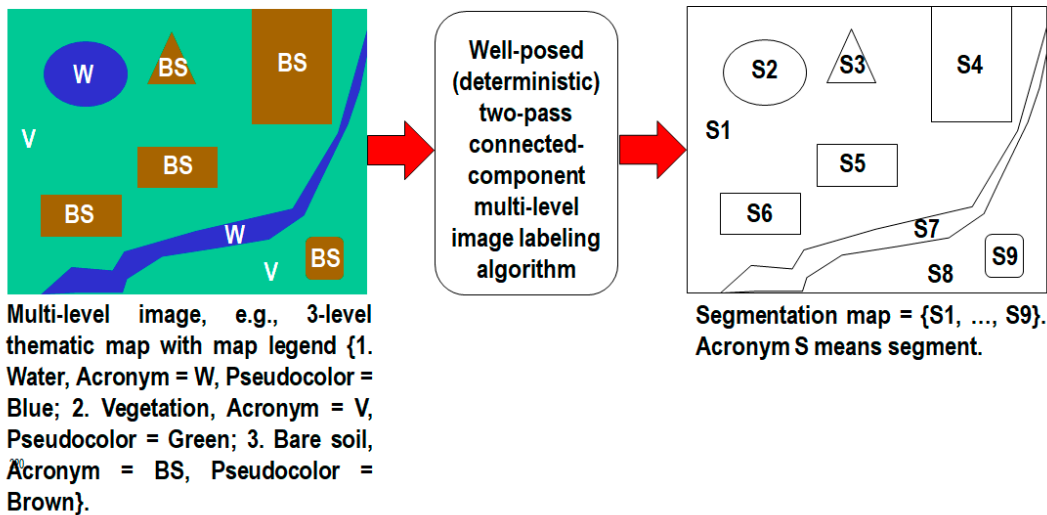



Figure 25. One segmentation map is deterministically generated from one multi-level (e.g., binary) image, such as a thematic map, but the vice versa does not hold, i.e., many multi-level images can generate the same segmentation map. To accomplish the determinist task of segmentation map generation from a multi-level image, the two-pass connected-component multi-level image labeling algorithm [16, 76] requires two raster scans of the input data set. In the figure above, as an example, nine image-objects/segments S1 to S9 can be detected in the 3-level thematic map shown at left. Each segment (image-object) consists of a connected set of pixels sharing the same multi-level map label. An image-object is either (0D) pixel, (1) line or (2D) polygon [168]. Each stratum/layer/level consists of one or more segments, e.g., stratum Vegetation (V) consists of two disjoint segments, S1 and S8. In general, a stratum is a multi-part polygon. Hence, in any multi-level (categorical, nominal, qualitative) image domain, three labeled spatial primitives (spatial units) co-exist and are provided with parent-child relationships: (i) pixel with a level-label and a pixel identifier (ID, e.g., the row-column coordinate pair), (ii) segment (either 0D, 1D, or 2D) with a level-specific label and a segment ID, and (iii) stratum (multi-part polygon) with a level-specific label, equivalent to a stratum ID. This overcomes the ill-fated dichotomy between traditional unlabeled sub-symbolic pixels versus labeled sub-symbolic segments in the numeric (quantitative) image domain traditionally coped with by the object-based image analysis (OBIA) paradigm [7].

1627

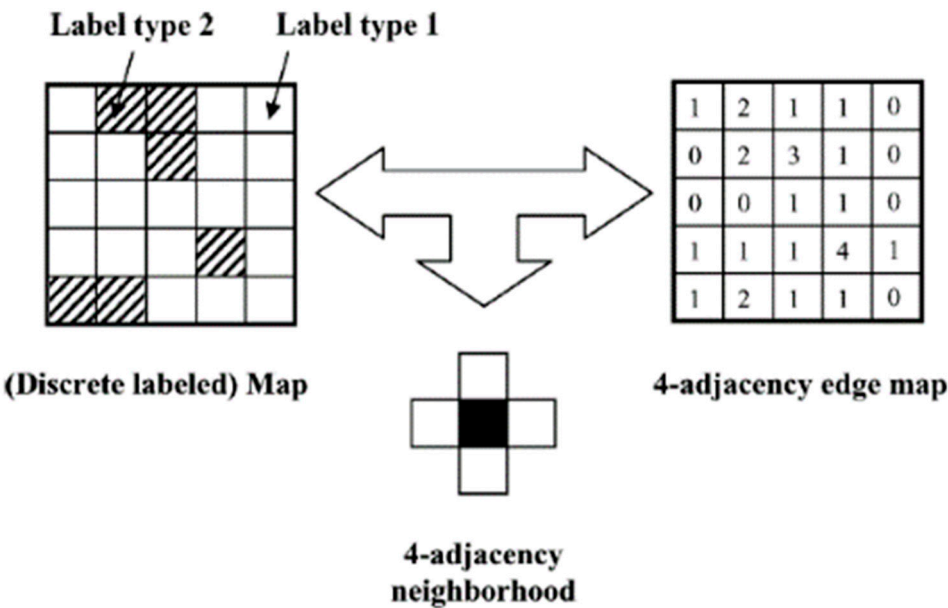



Figure 26. Example of a 4-adjacency cross-aura map, shown at right, generated in linear time from a two-level (binary) image shown at left.

Table 7. The SIAM computer program (release 88 version 7) is an EO system of systems scalable to any past, existing or future MS imaging sensor provided with radiometric calibration metadata parameters. It encompasses the following subsystems. (i) 7-band Landsat-like SIAM™ (L-SIAM™), with input channels Blue (B), Green (G), Red (R), Near Infra-Red (NIR), Medium IR1 (MIR1), Medium IR2 (MIR2), and Thermal IR (TIR). (ii) 4-band (channels G, R, NIR, MIR1) SPOT-like SIAM™ (S-SIAM™). (iii) 4-band (channels R, NIR, MIR1, and TIR) Advanced Very High Resolution Radiometer (AVHRR)-like SIAM™ (AV-SIAM™). (iv) 4-band (channels B, G, R, and NIR) QuickBird-like SIAM™ (Q-SIAM™) [10, 11, 12, 76, 78, 79, 91, 113, 114, 138, 139, 151].

SIAM, r88v7	Input bands	Prior knowledge-based color map legends: Number of output spectral categories = Vocabulary of multi-spectral (MS) color names			
		Fine discretization levels	Intermediate discretization levels	Coarse discretization levels	Inter-sensor discretization levels (*)
L-SIAM	7 – B, G, R, NIR, MIR1, MIR2, TIR	96	48	18	**33 (*):** employed for inter-sensor post-classification change/no-change detection
S-SIAM	4 – G, R, NIR, MIR1	68	40	15	
AV-SIAM	4 – R, NIR, MIR1, TIR	83	43	17	
Q-SIAM	4 – B, G, R, NIR	61	28	12	

Table 8. Legend (vocabulary) of the prior knowledge-based color map generated from a 7-band MS image (consisting of Landsat-like bands B, G, R, NIR, MIR1, MIR2 and TIR), radiometrically calibrated into TOARF, SURF or surface albedo values, by the Landsat-like SIAM (L-SIAM™, release 88 version 7) implementation, also refer to Table 7. For the sake of representation compactness, pseudo-colors of the 96 spectral categories are gathered along the same raw if they share the same parent spectral category in the decision tree, e.g., "strong" vegetation, equivalent to a spectral end-member. The pseudo-color of a spectral category (color name) is chosen as to mimic natural colors of pixels belonging to that spectral category.

II.   Required input data set: RGB image, either true- or false-color, not provided with radiometric calibration metadata parameters, but submitted to color constancy. The RGB Image Automatic Mapper (RGBIAM) is a lightweight computer program capable of color naming, superpixel detection and VQ quality assessment in non-calibrated RGB imagery, either true- or false-color, submitted to color constancy as a mandatory non-calibrated RGB image pre-processing stage for image harmonization through space, time and sensors. RGBIAM comprises a novel expert system capable of partitioning a three-band RGB data cube, either true- or false-color, into a pre-defined vocabulary of RGB color names at two different

1659   quantization levels of color granularity, specifically 50 and 11 color names, see
1660   Figure 27 and Table 9 [10, 174, 175].

1661

(a)

(b)

(c)

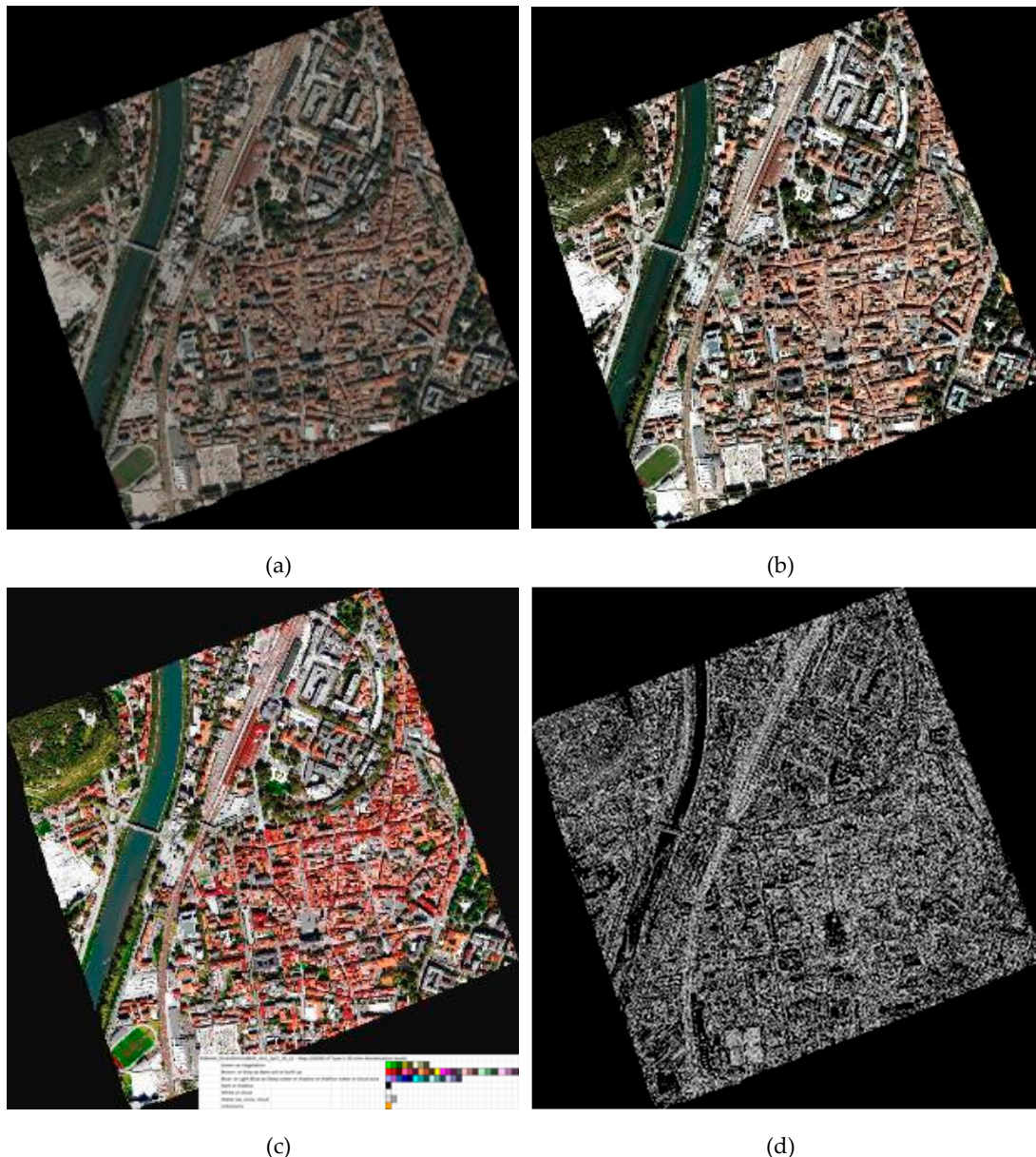
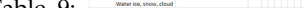
(d)

1662   **Figure 27.** (a) Airborne 10 cm resolution true-color RGB orthophoto of Trento, Italy, 4017 x
1663   4096 pixels in size x 3 bands, acquired in 2014 and provided with no radiometric calibration
1664   metadata file. No histogram stretching is applied for visualization purposes. (b) Same RGB
1665   orthophoto subject to self-organizing statistical color constancy. (c) RGBIAM
1666   polyhedralization of the RGB color space and prior color map of the RGB image subject to
1667   color constancy. The RGBIAM map legend, consisting of 50 spectral categories, is depicted in

1668   pseudocolors. Map legend, shown in Table 9: _________________. Input
1669   parameters: none. Processing time (one-pass, IDL implementation) = 2 min. (d) To visualize
1670   contours of image-segments automatically detected in the multi-level RGBIAM color map-
1671   domain by a deterministic two-pass connected-component multi-level image labeling
1672   algorithm [16, 176], see Figure 25, an automatic 4- or 8-adjacency cross-aura measure is
1673   estimated in linear time, see Figure 26.

Table 9. Legend (vocabulary) of the prior knowledge-based color map generated from a 3-band RGB image, in either true- or false-color and submitted to a color constancy pre-processing stage, by the RGBIAM (release 6 version 2) implementation at fine color quantization with 50 color names (versus 11 basic color names at coarse color space partitioning). For the sake of representation compactness, pseudo-colors of the 50 spectral categories are gathered along the same raw if they share the same parent spectral category in the decision tree. The pseudo-color of a spectral category (color name) is chosen as to mimic natural colors of pixels belonging to that spectral category.

(iii)   Second stage for image-contour detection and image-segmentation (raw primal sketch) [10, 69]. Based on an original physical model-based multi-scale multi-orientation 2D wavelet-based spatial filter bank [10], equivalent to a prior knowledge-based (deductive) CNN alternative to inductive learning-from-data DCNNs currently dominating the CV literature and whose architecture is based on heuristics, the proposed physical model-based CNN is consistent with constraint 'Human vision → CV ⊃ EO-IU', see Figure 5. In particular, a CV ⊃ EO-IU subsystem is tested on complex EO spaceborne/airborne images if and only if it performs in agreement with human visual perception on test cases at increasing levels of signal complexity, e.g., 1D synthetic signal, (2D) synthetic image, natural panchromatic imagery and natural color imagery whose "ground truth" is intuitive to understand, etc. Perceptual criteria to comply with are: (I) the Mach bands visual illusion [133], see Figure 9, and (II) the perceptual true-fact that human panchromatic and chromatic vision mechanisms are nearly as effective in scene-from-image reconstruction and understanding (refer to Section 2). The proposed physical model-based multi-scale multi-orientation CNN design and implementation satisfy these perceptual requirements. To reduce computation time, each 2D wavelet-based spatial filter, 2D_f($\cdot$), assumed to be W × W pixels in size, is implemented as a 2D separable filter, such that 2D_f(x,y) = 1D_g(x) × 1D_g(y), where the 1D wavelet-based spatial filter, 1D_g($\cdot$), is W pixels in size [140]. Hence, computational complexity of separable per-filter image convolution is 2W × image size in pixels, rather than W × W × image size in pixels. On theory, four filter orientations (0°, 45°, 90° and 135°) and four dyadic spatial scales, s = 0, ...3, are employed, where filter size $W_0 = 2^0 \times 3 = 3$ pixels, $W_1 = 2^1 \times 3 \approx 7$ pixels, $W_2 = 2^2 \times 3 \approx 13$ pixels and $W_3 = 2^3 \times 3 \approx 25$ pixels. In practice, the same W_0 filter is applied hierarchically upon the raw image at scale 0 and on three low-pass filtered images, downscaled by a factor of 2 at scales 1 to 3 [140]. The implemented CNN is capable of: (a) near-orthogonal image analysis/decomposition and lossless image synthesis/reconstruction, in analogy with the well-known multi-scale Gaussian and Laplacian pyramid proposed in [140], (b) automated (requiring no human-machine interaction) zero-crossing (ZX) image-contour detection ([10]), in line with the biologically plausible CV system proposed by Marr [69], and (c) automated ZX image-segment (blob, closed-contour) detection [10], never accomplished by Marr in his seminal work [69], see

1714   Figure 28. Such a deductive CNN implementation is alternative to inductive learning-
1715   from-data algorithms for image-contour detection and image segmentation, either 2D
1716   image analysis approaches, such as DCNNs [31], or 1D image analysis approaches, e.g.,
1717   [134-136], which are inherently semi-automatic and site-specific [2].

1718

(a)   (b)   (c)

(d)   (e)

(f)   (g)   (h)

(i)   (l)

(m)   (m)   (o)

(p)   (q)

(r)   (s)   (t)

(u)   (v)

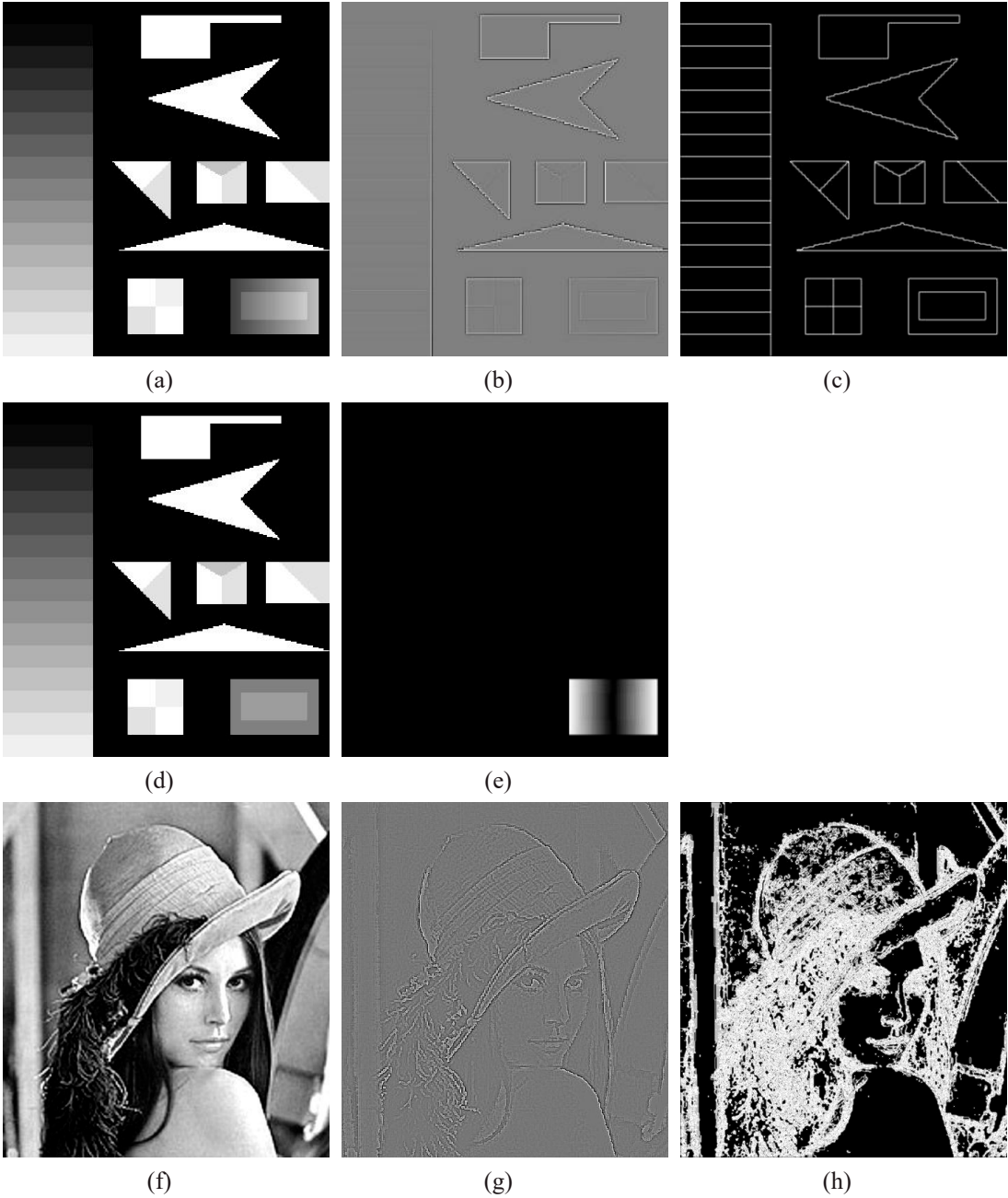
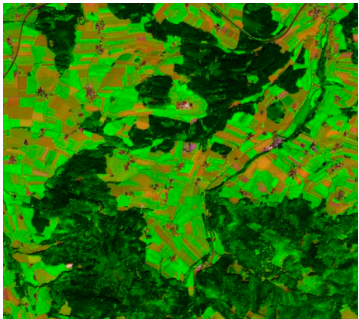
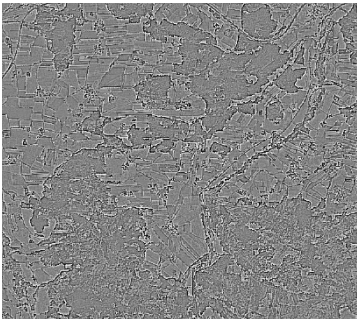
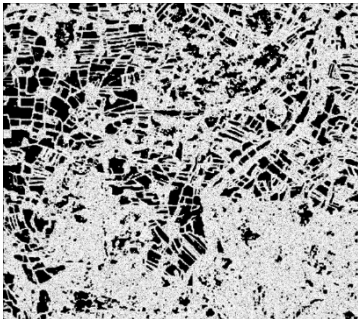
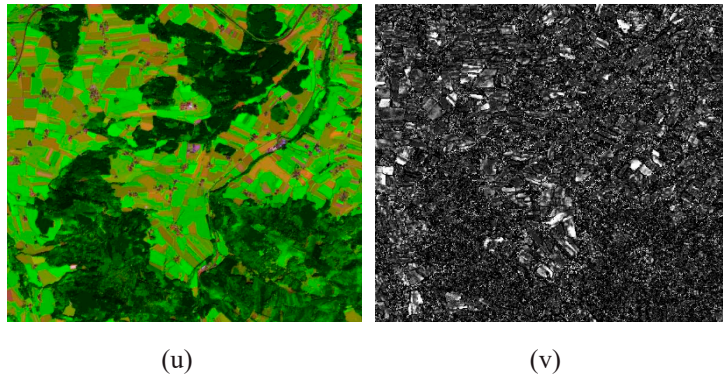



Figure 28. To comply with constraint 'Human vision → CV ⊃ EO-IU', see Figure 5, an EO-IU subsystem for image-contour detection and image segmentation is tested on complex EO spaceborne/airborne images if and only if it performs in agreement with human visual perception, such as: (1) the Mach bands visual illusion [133] and (2) the perceptual true-fact that human panchromatic and chromatic vision mechanisms are nearly as effective in scene-from-image reconstruction and understanding, starting from simpler test of increasing signal complexity, e.g., 1D synthetic signal, 2D synthetic image, natural panchromatic and natural chromatic images of intuitive "ground truth", etc. (a) SUSAN synthetic panchromatic image, byte coded in range {0, 255}. Step edges and ramp edges at known locations (the latter forming the two inner rectangles visible at the bottom right corner) form angles from acute to obtuse. According to human vision, 31 image-segments can be detected as reference "ground-truth". (b) Sum (synthesis) of the wavelet-based near-orthogonal multi-scale multi-orientation image decomposition. Filter value sum in range [-255, +255]. (c) Automated (requiring no human-machine interaction) image segmentation into zero-crossing (ZX) segments generated from ZX pixels detected by a multi-scale multi-orientation spatial filter bank, equivalent to a prior knowledge-based CNN, different from Marr's single-scale isotropic ZX pixel detection ([69]). Exactly 31 image-segments are detected with 100% contour accuracy. Segment contours depicted with 8-adjacency cross-aura values in range {0, 8}, see Figure 26. (d) Image-object mean view = object-wise constant input image reconstruction. (e) Object-wise constant input image reconstruction compared with the input image, per-pixel root mean square error (RMSE) in range [0, 255]. (f) Natural panchromatic image of Lenna. (g) Same as (b). (h) Same as (c), there is no CV system's free- parameter to be user-defined. (i) Same as (d). (l) Same as (e). (m) Natural RGB-color image of Lenna. (n) Same as (b). (o) Same as (c), there is no CV system's free- parameter to be user-defined. (p) Same as (d). (q) Same as (e). (r) Zoom-in of a Sentinel-2A MSI Level-1C image of the Earth surface south of the city of Salzburg, Austria. Acquired on 2015-09-11. Spatial resolution: 10 m. Radiometrically calibrated into top-of-atmosphere reflectance (TOARF) values in range {0, 255}, it is depicted as a false color RGB image, where: R = Medium InfraRed (MIR) = Band 11, G = Near IR (NIR) = Band 8, B = Blue = Band 2. Standard ENVI histogram stretching applied for visualization purposes. (s) Same as (b). (t) Same as (c), there is no CV system's free- parameter to be user-defined. (u) Same as (d). (v) Same as (e).

(iv)   Third stage for planar shape indexing. An original minimally dependent and maximally informative (mDMI) set of planar shape (geometric) indexes was conceived and implemented as described in [10, 163]. The proposed mDMI set of geometric functions comprises scale-invariant roundness (compactness and no holiness) ∈ [0, 1], elongatedness (and no holiness) ≥ 1, multi-scale straightness of boundaries ∈ [0, 1], simple connectivity (no holiness) ∈ [0, 1], rectangularity (and no holiness) ∈ [0, 1] and convexity (and no holiness) ∈ [0, 1], to be estimated per image-object in addition to size and orientation, see Figure 29.

1758

Segment Number	Chromatic	Panchromatic	Segment	Convexity and No Hole	Elongatedness	Polygon-Based Approximate Rectangularity	Roundness and No Hole	Simple-Connectivity	Straightness of Boundary	Angle of MER (in degrees)	Area (in pixels)	Average Contrast Along Boundary	Morphological multiscale characteristic	Mean Panchromatic intensity
1				0.96	1.10	1.00	0.90	1.00	0.68	90.00	81	30.53	5.59	94.95
2				0.85	2.92	0.95	0.66	1.00	0.63	75.07	666	1.91	15.14	57.62
3				0.86	4.68	1.00	0.62	1.00	0.77	-16.50	237	36.07	5.12	113.29
4				0.87	8.72	1.00	0.53	0.72	0.83	74.05	1406	27.24	7.27	89.84
5				0.78	4.86	1.00	0.58	0.89	0.89	73.30	1812	27.00	15.62	76.79
6				0.48	9.24	0.78	0.44	1.00	0.79	155.85	461	7.83	16.31	59.71
7				0.35	50.42	0.72	0.22	1.00	0.89	-105.95	727	17.72	9.64	54.21
8				0.67	22.35	0.05	0.33	1.00	0.85	-5.57	340	20.51	8.87	55.27
9				0.84	9.61	1.00	0.54	0.93	0.85	167.83	555	20.22	8.67	49.82

Figure 29. Screenshot of a graphic user interface (GUI) specifically developed to show a human expert values of the proposed minimally dependent and maximally informative (mDMI) set of geometric attributes. In this GUI, darker cells correspond to: (i) higher values of geometric attributes and (ii) lower values of photometric attributes, like the panchromatic mean intensity shown at the rightest column. In this figure, for reasons of readability only nine segments are shown simultaneously for comparison. Detected in the spaceborne VHR test image of an urban area, segments 1 through 6 correspond to buildings or parts of buildings while segments 7 through 9 belong to roads. These two families of segments appear easy to discriminate based on different combinations of ranges of change of their geometric attributes.

(v)   Fourth stage for texture segmentation (full primal sketch), synonym of perceptual spatial grouping of texture elements, texels [10, 16, 164-166] or tokens [69]. Based on a multi-scale texture binary profile accomplished in linear time complexity via multi-scale window-based analysis of the spatial distribution of binary image-contours [10]. To date, automated image-texture segmentation is an open problem [20, 68, 141, 177].

(vi)   Fifth stage for general-purpose, user- and application-independent ESA EO Level 2 product generation (refer to Section 1), based on a 2D convergence-of-visual-evidence approach, according to Equation (2) [10], where the proposed ESA EO Level 2 SCM product taxonomy consists of a standard 3-level 8-class FAO LCCS-DP legend augmented with quality layers cloud and cloud-shadow, see Table 2 and Figure 1.

(vii)   Sixth stage for user- and application-dependent FAO LCCS-MHP classification (see Figure 1), based on a 2D convergence-of-visual-evidence approach, according to Equation (2) [10].

At the levels of system understanding of information/knowledge representation, system architecture, algorithm and implementation, the AutoCloud+ subsystem requirements specification is considered a specialized version of the aforementioned six-stage hybrid feedback CV system baseline. In greater detail, the AutoCloud+ flow chart, algorithm and implementation are specialized as follows [10, 36, 67], see Figure 30.

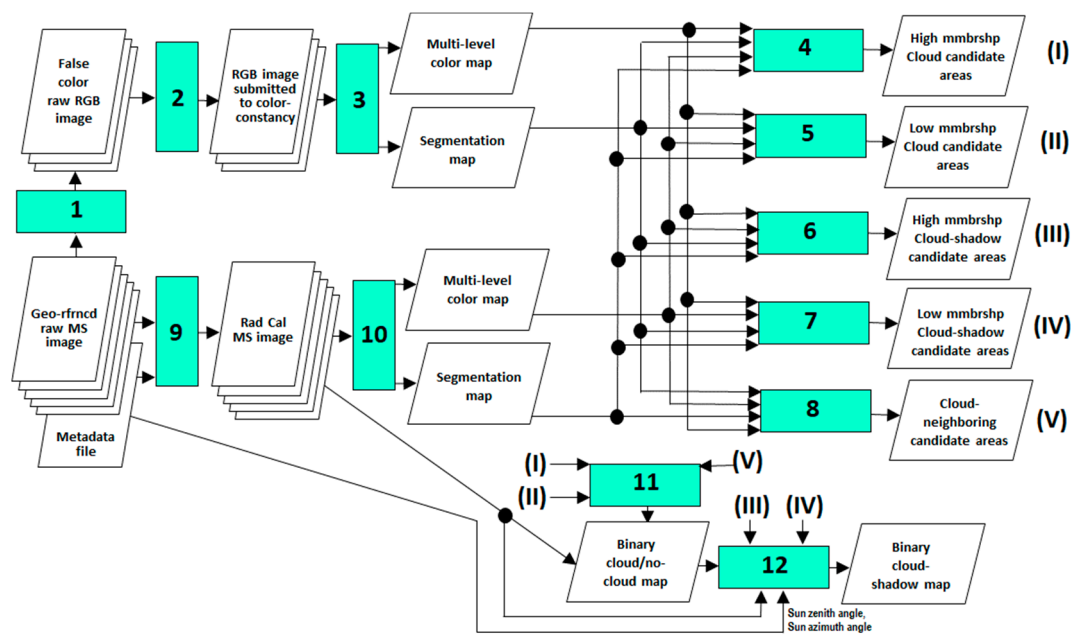



Figure 30. AutoCloud+ hybrid (combined physical model-based and statistical model-based) 2D image analysis system design (architecture) for spatial context-sensitive and spatial topology-preserving cloud/cloud-shadow detection. (1) True- or false-color RGB channel selection. (2) Statistical self-organizing color constancy algorithm. (3) RGBIAM lightweight computer program for RGB color space polyhedralization into color names, superpixel detection and vector quantization (VQ) quality assessment. (4) Candidate cloud areas, based on convergence-of evidence. (5) Candidate no-cloud areas, based on convergence-of evidence. (6) Candidate cloud-shadow areas, based on convergence-of evidence. (7) Candidate no-cloud-shadow areas, based on convergence-of evidence. (8) Candidate cloud neighboring areas, based on convergence-of evidence. (9) Radiometric calibration of digital numbers (DNs) into TOARF, SURF or surface albedo values. (10) SIAM lightweight computer program for MS reflectance space hyperpolyhedralization into color names, superpixel detection and vector quantization (VQ) quality assessment. (11) Spatial cloud modeler: Clouds detected from candidate cloud and cloud neighboring areas. (12) Spatial bidirectional cloud and cloud-shadow modeler: Physical model-based cloud shadow detection, moving from cloud to cloud-shadow candidates and vice versa, for mutual reinforcement learning.

First, the two SIAM and RGBIAM lightweight computer programs for color space discretization into color names are combined to pursue convergence of photometric evidence, whether or not the color space is radiometrically calibrated. For example, when the input MS image is radiometrically calibrated into TOARF, SURF or surface albedo values, then SIAM and RGBIAM are run in parallel upon, respectively, the original input MS image (calibrated) and an RGB false-color selection of the input (calibrated) image pre-processed by the self-organizing algorithm for color constancy (similar to image enhancement by histogram stretching). The two output multi-level maps in color names generated as output by SIAM and RGBIAM can be deterministically segmented in linear time complexity with image size into labeled image-objects (super-pixels, connected sets of pixels featuring the same color label) by a well-known two-pass connected-component multi-level image labeling algorithm [16, 176], see Figure 25. In the SIAM and RGBIAM output segmentation map, each connected-component is labeled with one segment identifier (ID) together with one semi-symbolic color name. For example, an image-object, either (0D) pixel, (1) line or (2D) polygon [168], is provided with a segment ID, a MS color name 'green-as-Vegetation' by SIAM and an RGB color name 'Green' by RGBIAM.

Second, candidate cloud-objects and candidate cloud-shadow-objects are detected in the image-domain based on a convergence of spatial shape and size properties with SIAM and

RGBIAM color names. Stratified (class-conditioned) by color names, a driven-by-knowledge shape-preserving dilation process is run to fill small spatial gaps within cloud/cloud-shadow candidate areas (e.g., due to self-occlusion and shadow-casting phenomena occurring in clouds, where solar luminance decreases) and along boundaries of candidate cloud areas.

Third, the cloud candidate image-objects can grow in the image-domain to include spatially adjacent image-objects whose color, shape and size properties are consistent with the hypothesis of belonging to class crown-of-cloud.

Finally, a bidirectional physical knowledge-based data modeler is run for spatial reasoning (see software blocks identified as 11 and 12 in Figure 30). Making use of OBIA concepts for spatial information modeling [178], it is capable of cloud/cloud-shadow image-object pair spatial matching in shape while accounting for the sun position, to reduce cloud and cloud-shadow false positives and false negatives. In the quest for a unidirectional cloud-to-cloud-shadow relationship, candidate cloud-shadows are searched for in the (2D) image-domain to be matched in shape by each individual candidate cloud-object adopted as starting position. This spatial search moves from the candidate cloud-object of interest, in the direction of the sun azimuth angle (known from the input EO image metadata file) with orientation away from the sun, for a spatial length estimated as a dependent variable of the cloud height, which is the sole unknown independent physical variable to cope with, estimated according to an *a priori* physical knowledge-based model of real-world cloud heights, in line with Sen2Cor [9]. The dual quest for object-pair matching in shape deals with the cloud-shadow-to-cloud unidirectional relationship, which is usually neglected in other cloud/cloud-shadow detectors, such as Sen2Cor [9] and MAJA ([39]). A simplified flow chart of the hybrid AutoCloud+ software toolbox for automated cloud/cloud-shadow detection is shown in Figure 31.

At the implementation level of system understanding, it is worth mentioning that candidate cloud/cloud-shadow image-object pairs are matched in shape according to the mDMI set of planar shape functions implemented in [10, 163]. The AutoCloud+ software blocks identified as data processing units 4 to 8, 11 and 12 in Figure 30 were implemented within the Trimble's eCognition Developer commercial software environment for fast prototyping OBIA solutions [179].

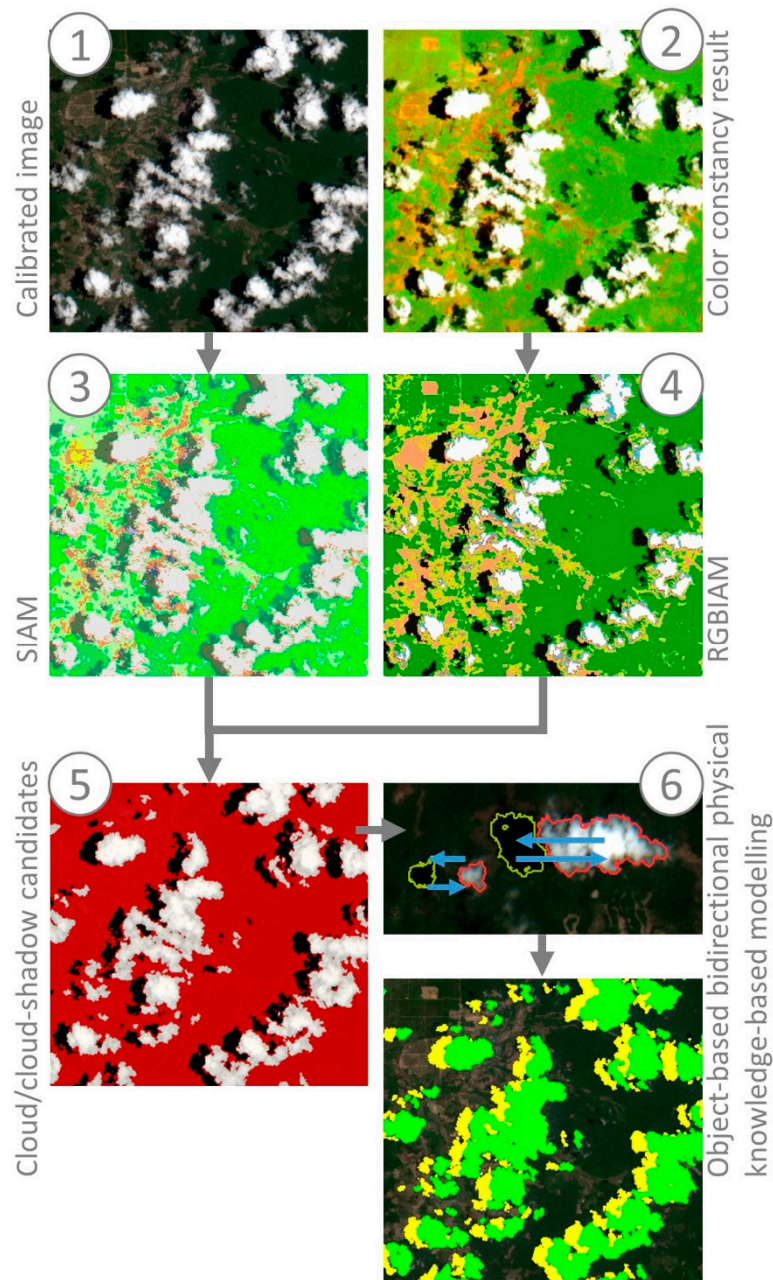
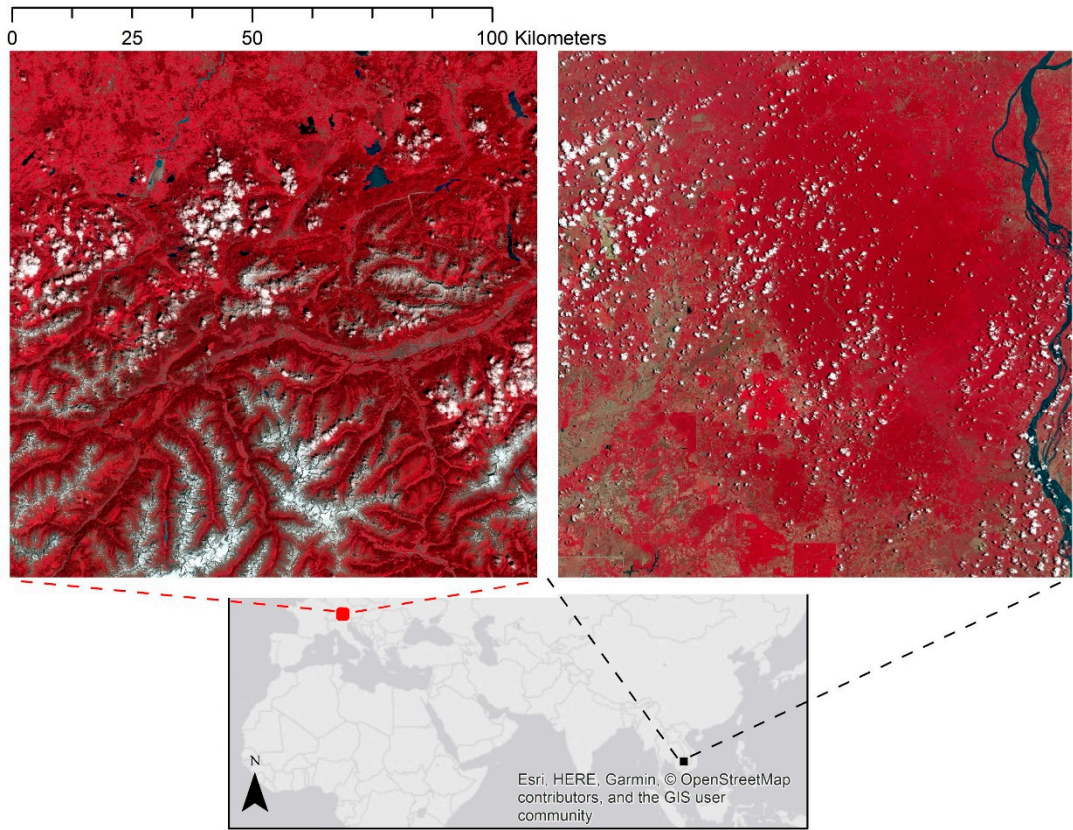



Figure 31. Simplified flow chart of the hybrid AutoCloud+ software toolbox for automated cloud/cloud-shadow detection, see Figure 30. This intuitive workflow is depicted as a sequence of two input data sets (1) and (2), generated from the same input MS image, one submitted to radiometric *Cal* (when radiometric *Cal* metadata parameters are available) and the other submitted to statistical color constancy, followed by four intermediate information products (3) to (6), and one final output product, specifically, a three-level output map with semantic layers cloud/cloud-shadow/others (rest of the world), shown at bottom right.

5. Materials

For testing purposes, the novel AutoCloud+ algorithm for cloud/cloud-shadow detection in single-date MS imagery was compared in quantitative terms of outcome and process quality, in compliance with the mDMI set of EO OP-Q²Is proposed in Section 2, with alternative standard CV software solutions available open source or free of cost, specifically, the single-date Sen2Cor and multi-temporal MAJA software toolboxes.

1863   As proof of concept, AutoCloud+, Sen2Cor and MAJA (refer to Table 5) were compared
1864 upon two test images, selected to be representative of the complexity of the cloud/cloud-
1865 shadow detection problem in real-world situations.
1866   The first test image was selected as a Sentinel-2 A (S2A) Multi-Spectral Instrument (MSI)
1867 Level 1C image radiometrically calibrated into TOARF values, depicting an Earth-surface area
1868 located   in   Cambodia   (Product   ID:
1869 S2A_MSIL1C_20170421T031541_N0204_R118_T48PWV_20170421T033212), see Figure 32, for
1870 which Sen2Cor and MAJA results were available for download from the Copernicus Open
1871 Access Hub (https://scihub.copernicus.eu/dhus/#/home) and from the CNES website
1872 (https://theia.cnes.fr/atdistrib/rocket/#/search?collection=SENTINEL2) respectively. This first
1873 test image shows a large variety of cloud patterns, varying in shape, size and cloud height,
1874 together with typical critical elements in cloud and cloud-shadow detection, such as cloud-
1875 shadows projected over water and vegetated surface types, in addition to occluded cloud and
1876 cloud-shadow phenomena.
1877   A second test image was identified in a Sentinel-2 B (S2B) Level 1C image radiometrically
1878 calibrated into TOARF values, depicting an Earth-surface area located in the Alpine area
1879 between   Austria   and   Germany   (Product   ID:
1880 S2B_MSIL1C_20180616T102019_N0206_R065_T32TPT_20180616T154713), see Figure 32. In
1881 this second test image, potential critical elements in cloud/cloud-shadow detection are the
1882 presence of snow/ice patterns in high-elevation areas, eligible for confusion with ice clouds, of
1883 clouds located next to visible snow/ice surface types and of Earth surfaces affected by shadows
1884 casted by mountains rather than clouds. For this second test image no MAJA mapping result
1885 was available for download from the aforementioned CNES website.
1886

1887   **Figure 32.** Overview of the two study areas adopted for testing (Left: Cambodia test site, Right:
1888   Alpine test site located between Austria and Germany). EO images of these two test areas must
1889   be selected to be representative of the complexity of the cloud/cloud-shadow detection
1890   problem in real-world situations. At left, second test image, identified as a Sentinel-2 B Multi-
1891   Spectral Instrument (MSI) Level 1C image, radiometrically calibrated into TOARF values,

1892 depicted in false-colors: monitor-typical RGB channels are selected as R = Near InfraRed (NIR)
1893 channel, G = Visible Red channel, B = Visible Green channel. Histogram stretching is applied
1894 for   visualization   purposes.   Product   ID:
1895 S2B_MSIL1C_20180616T102019_N0206_R065_T32TPT_20180616T154713. At right, first test
1896 image, identified as Sentinel-2 B MSI Level 1C image, radiometrically calibrated into TOARF
1897 values, and depicted in false-colors as the second test image shown at left. Product ID:
1898 S2A_MSIL1C_20170421T031541_N0204_R118_T48PWV_20170421T033212.

1899 ## 6. Results

1900 Alternative cloud/cloud-shadow output maps generated by the single-date AutoCloud+,
1901 single-date Sen2Cor and multi-temporal MAJA software toolboxes were compared in
1902 quantitative terms of outcome quality, as defined in the mDMI set of EO OP-Q²Is proposed in
1903 Section 2. In addition, the AutoCloud+ CV subsystem was assessed in quantitative terms of
1904 process quality, as defined in the mDMI set of OP-Q²Is proposed in Section 2. Noteworthy, this
1905 quality assurance strategy goes far beyond the traditional quantitative assessment policy of
1906 EO-IU systems presented in the RS literature, almost exclusively limited to mapping accuracy.
1907 With regard to process quality indicators defined in the mDMI set of OP-Q²Is proposed
1908 in Section 2, such as degree of automation and computation time, AutoCloud+ ran
1909 automatically (without human-machine interaction) upon the two test images, where Step 1 to
1910 Step 5 in Figure 31 were computed in near real-time, specifically, in linear time complexity
1911 with image size, equal to around 2 minutes per Sentinel-2 image in a standard laptop
1912 computer. The bidirectional object-based cloud/cloud-shadow spatial modeler, shown as Step
1913 6 in Figure 31 and prototyped in the Trimble eCognition software, depends on the number of
1914 cloud and cloud-shadow candidate image-objects detected in the image-domain at Step 5. For
1915 each of the two Sentinel-2 images selected for testing, the eCognition software prototype at
1916 Step 6 ran in a reasonable time frame, around 5 minutes per input image.
1917 With regard to outcome quality indicators defined in the mDMI set of OP-Q²Is proposed
1918 in Section 2, enhanced input images, either radiometrically calibrated into TOARF values or
1919 submitted to color constancy, and intermediate SIAM's and RGBIAM's output maps generated
1920 from the first Sentinel-2 test image are shown in Figure 33. Final cloud/cloud-shadow output
1921 maps generated by the single-date AutoCloud+, single-date Sen2Cor and multi-temporal
1922 MAJA computer programs are shown in Figure 34, zoomed-in in Figure 35 and extra zoomed-
1923 in in Figure 36. Table 10 reports on the total area of clouds and cloud-shadows detected by the
1924 three algorithms of interest in the first test image.
1925

(a)   (b)

(c)　　　　　　　　　　　　　　　　　　　(d)

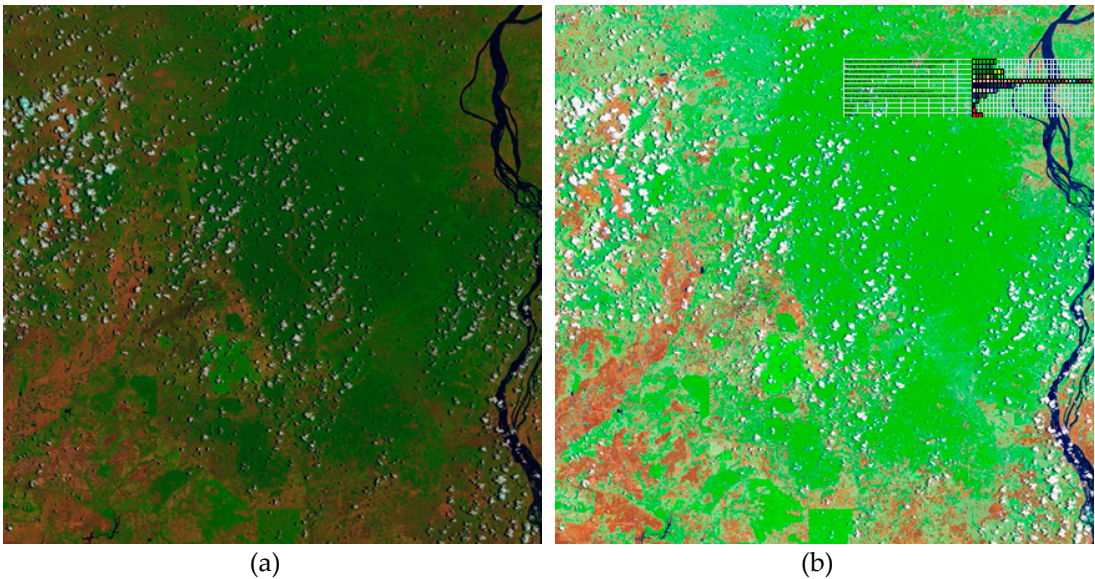
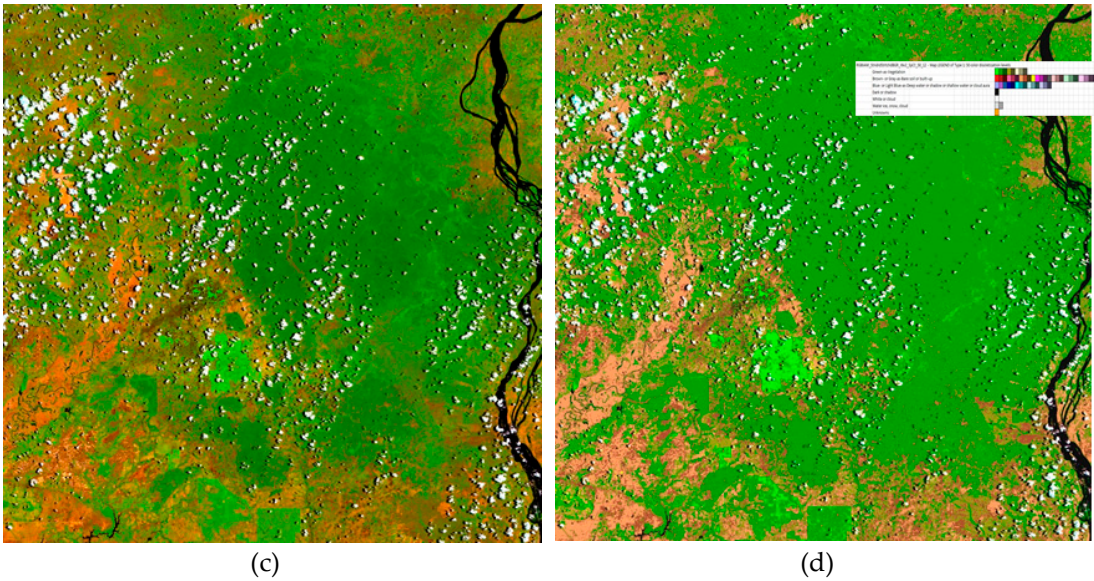
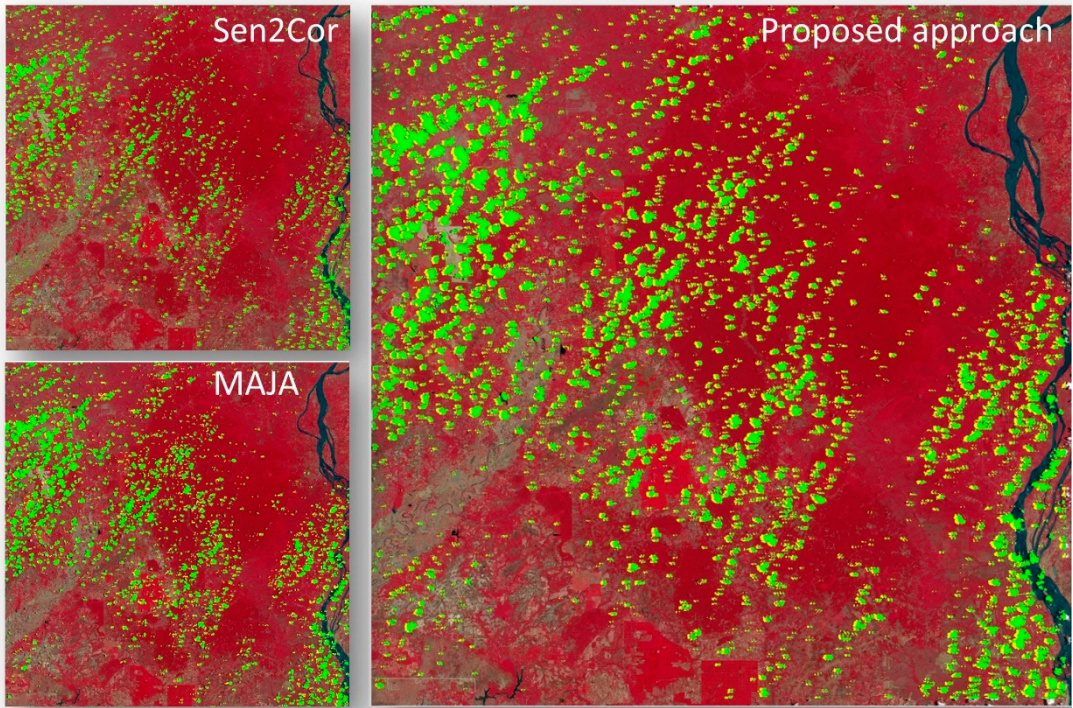
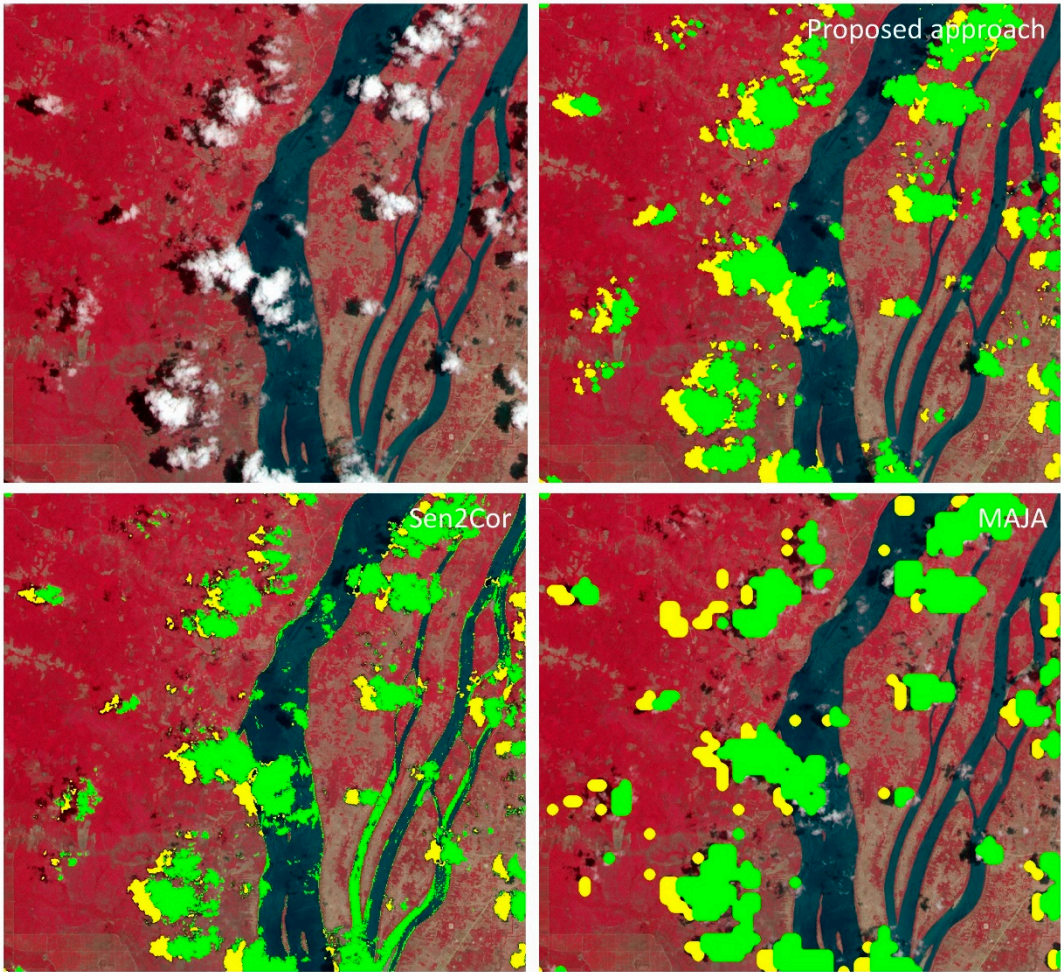



Figure 33. (a) First test image of a Cambodia site. Sentinel-2 A Multi-Spectral Instrument (MSI) Level 1C image of Cambodia, radiometrically calibrated into TOARF values. Acquisition date: 2018-06-16. Depicted in false-colors: monitor-typical RGB channels are selected as R = Medium InfraRed (MIR) channel, G = Near InfraRed (NIR) channel, B = Visible Blue channel. No histogram stretching is employed for visualization purposes. (b) SIAM map of the test S2 image shown in (a). SIAM map legend: 96 color names, depicted in pseudocolors as follows (see Table 8): (c) First test image, shown in (a), submitted to a self-organizing color constancy algorithm. Depicted in false-colors: monitor-typical RGB channels are selected as R = Medium InfraRed (MIR) channel, G = Near InfraRed (NIR) channel, B = Visible Blue channel. No histogram stretching is employed for visualization purposes. (d) RGBIAM map of the test S2 image shown in (c). RGBIAM map legend: 50 color names, depicted in pseudocolors as follows (see Table 9): .

1939   **Figure 34.** First test image of a Cambodia site. Final 3-level cloud/cloud-shadow/others maps
1940   generated by the three algorithms under comparison, specifically, single-date AutoCloud+,
1941   single-date Sen2Cor and multi-date MAJA, where class "others" is overlaid with the input
1942   Sentinel-2 A Multi-Spectral Instrument (MSI) Level 1C image, radiometrically calibrated into
1943   TOARF values and depicted in false-colors: monitor-typical RGB channels are selected as R =
1944   Near InfraRed (NIR) channel, G = Visible Red channel, B = Visible Green channel. Histogram
1945   stretching is applied for visualization purposes. Output class cloud is shown in a green
1946   pseudocolor, class cloud-shadow in a yellow pseudocolor.
1947

1948   **Figure 35.** First test image of a Cambodia site. Zoom-in of the final 3-level cloud/cloud-
1949   shadow/others maps generated by the three algorithms under comparison, where class
1950   "others" is overlaid with the input Sentinel-2 A Multi-Spectral Instrument (MSI) Level 1C
1951   image, radiometrically calibrated into TOARF values and depicted in false-colors: monitor-
1952   typical RGB channels are selected as R = Near InfraRed (NIR) channel, G = Visible Red channel,
1953   B = Visible Green channel. Histogram stretching is applied for visualization purposes. Output
1954   class cloud is shown in a green pseudocolor, class cloud-shadow in a yellow pseudocolor.
1955   Based on qualitative photointerpetation, Sen2Cor appears to underestimate cloud-shadows,
1956   although some water areas are misclassified as cloud-shadows. In addition, some river/river
1957   beds are misclassified as clouds. These two cases of cloud false positives and cloud-shadow
1958   false positives are highlighted in blue circles. MAJA overlooks some clouds small in size (in
1959   relative terms), as highlighted in red circles.

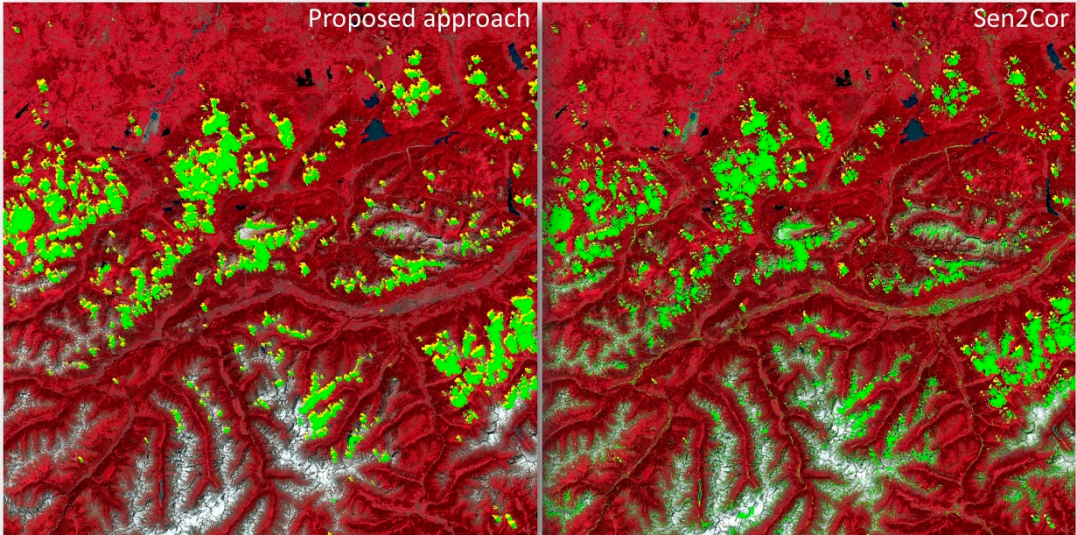
1960

1961   **Figure 36.** First test image of a Cambodia site. Extra zoom-in of the final 3-level cloud/cloud-
1962   shadow/others maps generated by the three algorithms under comparison, where class
1963   "others" is overlaid with the input Sentinel-2 A Multi-Spectral Instrument (MSI) Level 1C
1964   image, radiometrically calibrated into TOARF values and depicted in false-colors: monitor-
1965   typical RGB channels are selected as R = Near InfraRed (NIR) channel, G = Visible Red channel,
1966   B = Visible Green channel. Histogram stretching is applied for visualization purposes. Output
1967   class cloud is shown in a green pseudocolor, class cloud-shadow in a yellow pseudocolor.
1968   Based on qualitative photointerpetation, Sen2Cor appears to underestimate cloud-shadows,
1969   although some detected cloud-shadows are false positives because of misclassified water
1970   areas. To reduce false positives in cloud-shadow detection, MAJA adopts a multi-date
1971   approach. Nevertheless, MAJA misses some instances of cloud-over-water. Overall, MAJA
1972   cloud/cloud-shadow results look more "blocky" (affected by artifacts in localizing true
1973   boundaries of target image-objects).

1974   **Table 10.** First test image of a Cambodia site. Comparison of cloud/cloud-shadow total areas
1975   detected by the three tested algorithms.

	Cloud area (ha)	**Cloud-shadow (ha)**
AutoCloud+	103811	49368
Sen2Cor*	82589	20296
MAJA	109373	25896

*including all cloud probability classes

1976

1977   Final cloud/cloud-shadow output maps generated from the second Sentinel-2 test image
1978 by the single-date AutoCloud+ and single-date Sen2Cor computer programs are shown in
1979 Figure 37, zoomed-in in Figure 38. Table 11 reports on the total area of clouds and cloud-
1980 shadows detected by the two algorithms of interest in the second test image.
1981

1982 **Figure 37.** Second test image of an Alpine site. Final 3-level cloud/cloud-shadow/others maps
1983 generated by the two algorithms under comparison, specifically, single-date AutoCloud+ and
1984 single-date Sen2Cor, where class "others" is overlaid with the input Sentinel-2 B Multi-
1985 Spectral Instrument (MSI) Level 1C image, radiometrically calibrated into TOARF values and
1986 depicted in false-colors: monitor-typical RGB channels are selected as R = Near InfraRed (NIR)
1987 channel, G = Visible Red channel, B = Visible Green channel. Histogram stretching is applied
1988 for visualization purposes. Output class cloud is shown in a green pseudocolor, class cloud-
1989 shadow in a yellow pseudocolor.

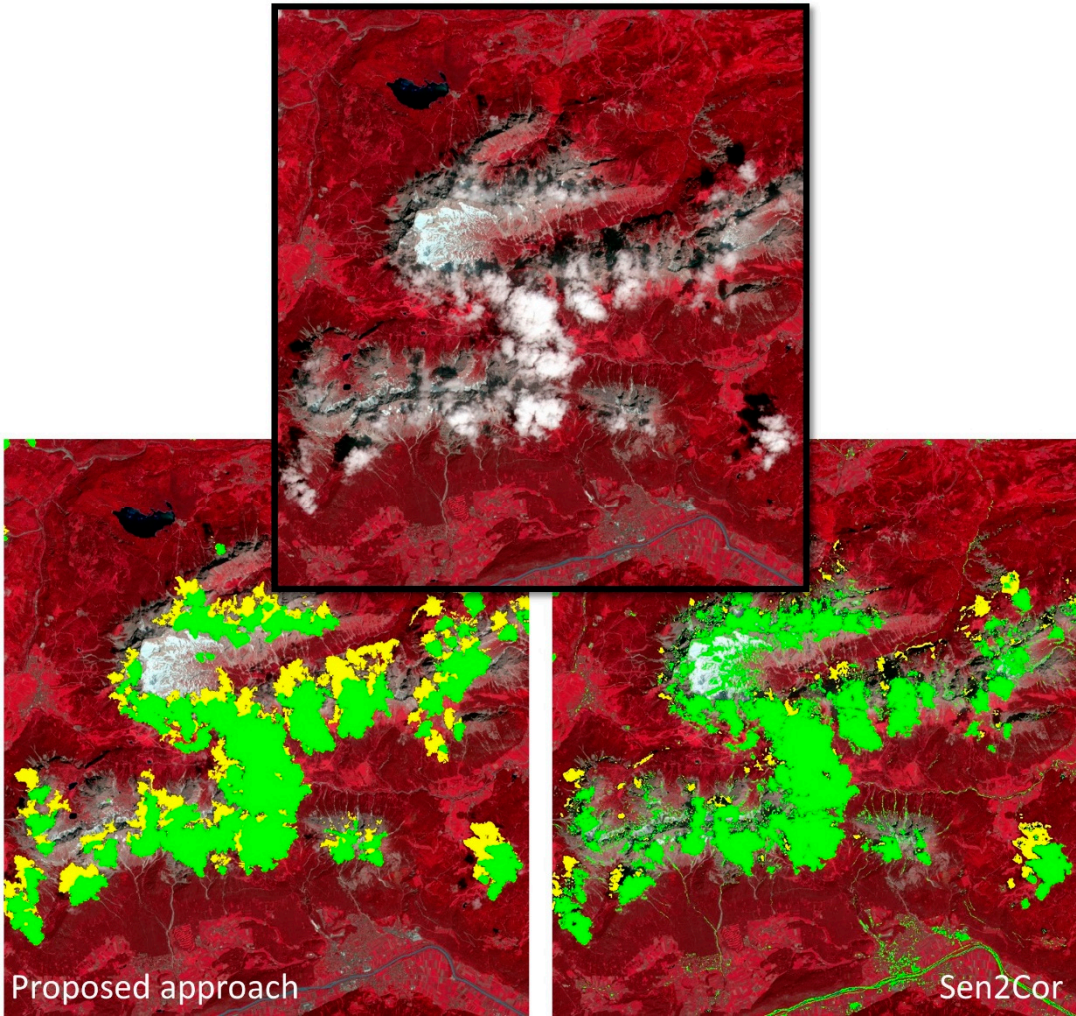



Figure 38. Second test image of an Alpine site. Zoom-in of the final 3-level cloud/cloud-shadow/others maps generated by the two algorithms under comparison, where class "others" is overlaid with the input Sentinel-2 A Multi-Spectral Instrument (MSI) Level 1C image, radiometrically calibrated into TOARF values and depicted in false-colors: monitor-typical RGB channels are selected as R = Near InfraRed (NIR) channel, G = Visible Red channel, B = Visible Green channel. Histogram stretching is applied for visualization purposes. Output class cloud is shown in a green pseudocolor, class cloud-shadow in a yellow pseudocolor. Based on qualitative photointerpetation, Sen2Cor appears to underestimate cloud-shadows, while some detected clouds are false positives because of misclassified water in river/river beds or misclassified snow/ice areas. AutoCloud+ avoids most of these cloud false positives and cloud-shadow false negatives by means of 2D spatial reasoning, specifically, by means of physical model-based modeling of cloud/cloud-shadow 2D shape properties and of bidirectional cloud/cloud-shadow 2D spatial relationships in the image-domain.

Table 11. Second test image of an Alpine site. Comparison of cloud/cloud-shadow total areas detected by the two tested algorithms.

	Cloud area (ha)	Cloud-shadow (ha)
AutoCloud+	106667	37905
Sen2Cor*	75180	14526

*including all cloud probability classes

7. Discussion

For the first Sentinel-2 test image of a Cambodia site, qualitative photointerpretation of the AutoCloud+ intermediate output products shown in Figure 33, in combination with quantitative vector quantization (VQ) error maps automatically generated as output by both the SIAM and RGBIAM software toolboxes (not shown), confirmed the validity of the SIAM and RGBIAM software solutions for color space quantization into categorical color names, in line with the existing literature [11, 12, 174], where SIAM was validated at continental scale on multi-year annual Landsat image mosaics. At first glance, qualitative photointerpretation of the output 3-level cloud/cloud-shadow/others thematic maps generated by the three AutoCloud+, Sen2Cor and MAJA algorithms under comparison, shown in Figure 34, look satisfactory. No cloud large in size (in relative terms) is missed by any algorithm. Actually, detected clouds appear somehow overestimated rather than underestimated in AutoCloud+ and MAJA. On a second glance, differences between the three approaches become notable. Table 10 reports on the total area of extracted clouds and cloud-shadows. AutoCloud+ and MAJA score similar values for the cloudy total area (~103 000 ha / 109 000 ha), but the former scores twice as much in the total amount of cloud-shadow areas (~50 000 ha). The Sen2Cor results reveal the lowest total area values (~ 82 500 ha / 20 000 ha) for both classes cloud and cloud-shadow respectively. These summary statistics are confirmed by visual analysis of zoomed-in subsets of Figure 34, as shown in Figure 35 and Figure 36. AutoCloud+ and MAJA appear to dilate (grow, on purpose) clouds at cloud boundaries to include an aura of thin cloud regions, if any, which are typically ignored by Sen2Cor. In this spatial neighboring analysis, MAJA results look more "blocky", i.e., affected by artifacts in localizing perceptually true boundaries of target image-objects, whereas the AutoCloud+ approach "naturally" follows the shape of clouds as they are perceived by a human photo interpreter. Sen2Cor largely underestimates cloud-shadows, although some water areas are misclassified as cloud-shadow, see Figure 35. In addition, some river/river beds are misclassified as clouds, see Figure 35 and Figure 36. To avoid such false positives in cloud and cloud-shadow detection affecting Sen2Cor, MAJA adopts a multi-temporal approach. Nevertheless, MAJA misses some instances of clouds small in size (in relative terms), which are detected quite well by the other two algorithms, see Figure 35. In addition, MAJA misses some instances of cloud-over-water, see Figure 36. Overall, on closer inspection of the first test case, single-date AutoCloud+ scores qualitatively "high" in cloud and cloud-shadow mapping accuracy, both affected by few false positives and few false negatives. Hence, single-date AutoCloud+ is considered superior to single-date Sen2Cor in mapping accuracy and superior to multi-date MAJA in mapping accuracy as well as ease-of-use and timeliness, since the former requires less input data to run.

For the second Sentinel-2 test image of an Alpine site, qualitative photointerpretation of the output 3-level cloud/cloud-shadow/others thematic maps generated by the two AutoCloud+ and Sen2Cor algorithms under comparison, shown in Figure 37 and zoomed-in Figure 38, confirm conclusions drawn from the first test case. Sen2Cor appears to underestimate cloud-shadows, while some detected clouds are false positives because of misclassified water in river/river beds or misclassified snow/ice areas. AutoCloud+ avoids most of these cloud false positives and cloud-shadow false negatives. Image-wide statistics collected in Table 11 confirm that AutoCloud+ tends to detect 30% to 40% more cloud-affected areas (in ha) and 50% more areas affected by cloud-shadow phenomena than Sen2Cor.

To be validated in future works at large spatial scale and multiple time samples according to the GEO-CEOS *Val* guidelines [180], these preliminary results confirm the well-groundedness of our working hypothesis (see Figure 3) in the multi-disciplinary domain of cognitive science (see Figure 5), starting from the AutoCloud+ project requirements specification summarized in Section 4 at the levels of system understanding of information/knowledge representation, system architecture, algorithm and implementation.

AutoCloud+ process quality indicators belonging to the mDMI set of EO OP-Q²Is proposed in Section 2, but not investigated in this experimental session are robustness to changes in input data, at large spatial scale and time span in agreement with the GEO-CEOS

2058   *Val* guidelines [180], and scalability to changes in imaging sensor's spatial and spectral
2059   resolutions. In agreement with theory (refer to Section 2 and Section 3), AutoCloud+ is
2060   expected to score "high" in all EO OP-Q²Is proposed in Section 2.

2061   **8. Conclusions**

2062   To contribute toward filling an analytic and pragmatic information gap from EO *big data,*
2063   characterized by the five Vs of volume, variety, veracity, velocity and value [4], to timely,
2064   comprehensive and operational EO data-derived value-adding products and services (VAPS),
2065   in compliance with the GEO-CEOS QA4EO Calibration/Validation (*Cal/Val*) requirements [3]
2066   and with the yet-unaccomplished visionary goal of a GEO's implementation plan for years
2067   2005-2015 of a Global Earth Observation System of Systems (GEOSS) [5], the present research
2068   and technological development (RTD) work provides several original contributions.

2069   Conceptual in nature and pertaining to the interdisciplinary domain of cognitive science,
2070   see Figure 5, our first original contribution coincides with our working hypothesis, see Figure
2071   3. In symbols of the standard Unified Modeling Language (UML) for graphical modeling of
2072   object-oriented software [81], our working hypothesis reads as:

2073   'Human vision → CV ⊃ EO-IU in operating mode ⊃ NASA EO Level 2 product →
2074   ESA EO Level 2 product → [EO-SCBIR + SEIKD = AI4DIAS] → GEOSS',

2075   where symbol '→' means *part-of* dependence, pointing from the supplier to the client, whereas
2076   symbol '⊃' means *subset-of* relationship (with inheritance), pointing from the superset to the
2077   subset. This working hypothesis postulates that necessary not sufficient pre-condition for
2078   multi-sensor EO *big data* cube analytics is systematic generation of ESA EO Level 2 product [8,
2079   9] (for definition, refer to Section 1), encompassing cloud and cloud-shadow quality layer
2080   detection in operating mode (for definition, refer to Section 2). In greater detail, systematic
2081   ESA EO Level 2 product generation was regarded as necessary not sufficient pre-condition for
2082   developing timely, comprehensive and operational EO data-derived VAPS, such as semantic
2083   content-based image retrieval (SCBIR) + semantics-enabled information/knowledge discovery
2084   (SEIKD) = artificial intelligence (AI) for Data and Information Access Services at the ground
2085   segment (AI4DIAS), in multi-sensor, multi-temporal and multi-angular EO *big data* cubes as
2086   *part-of* the GEO-CEOS visionary goal of a GEOSS, never accomplished to date by the RS
2087   community. This dependence relationship means that the GEOSS open problem, whose *part-*
2088   *of* are the still-unsolved (open) subproblems of SCBIR and SEIKD, cannot be accomplished
2089   until the necessary not sufficient pre-condition of computer vision (CV) ⊃ EO image
2090   understanding (EO-IU) in operating mode, specifically, systematic ESA EO Level 2 product
2091   generation, is accomplished in advance.

2092   Generation of an ESA EO Level 2 product from MS imagery is a CV ⊃ EO-IU task.
2093   Synonym of scene-from-image reconstruction and understanding [20], vision is a cognitive
2094   *(information-as-data-interpretation)* problem [15] very difficult to solve because (i) non-
2095   polynomial (NP)-hard in computational complexity [88, 89], (ii) and inherently ill-posed [20,
2096   68], because affected by: (I) a 4D-to-2D data dimensionality reduction, from the scene-domain
2097   to the image-domain, and (II) a semantic information gap, from the sub-symbolic
2098   (quantitative, unequivocal) image-domain to the symbolic (qualitative, uequivocal) scene-
2099   domain [15, 20]. Hence, vision ⊃ CV ⊃ ESA EO Level 2 product generation are inherently ill-
2100   posed problems very difficult to solve, that require *a priori* knowledge in addition to sensory
2101   data to become better conditioned for numerical solution [30]. If dependence 'ESA EO Level
2102   2 product → [EO-SCBIR + SEIKD = AI4DIAS] → GEOSS' holds, according to our working
2103   hypothesis, then the complexity of GEOSS ≥ complexity of [SCBIR + SEIKD] ≥ complexity of
2104   ESA EO Level 2 product, which is inherently ill-posed and NP-hard. This conclusion is a
2105   derived conjecture. It highlights, first, the potential relevance of ESA EO Level 2 product as
2106   information primitive for EO data-derived VAPS development in Space Economy 4.0 [59].
2107   Second, it justifies the present work, whose overarching goal is systematic ESA EO Level 2

2108 product generation, starting from cloud and cloud-shadow quality layer detection in
2109 operating mode.

2110 Our second original contribution is both conceptual and pragmatic in the codification of
2111 RS best practice, which is the focus of efforts made by, for example, intergovernmental
2112 organizations such as GEO and CEOS. ESA EO Level 2 product is regarded as baseline
2113 information primitive suitable for an "augmented" EO Analysis Ready Data (ARD) format
2114 specification, more restrictive (in terms of output product requirements specification) and
2115 more informative (in terms of physical and conceptual/semantic quality of numeric and
2116 categorical output products, respcetively), but more difficult to be inferred from EO sensory
2117 data than the U.S. Landsat ARD [120, 121] and CEOS ARD for Land (CARD4L) formats [119].

2118 To contribute toward filling an analytic and pragmatic information gap from EO *big data*
2119 to timely, comprehensive and operational EO data-derived VAPS, starting from the necessary
2120 not sufficient pre-condition of systematic ESA EO Level 2 product generation, our final original
2121 contribution deals with the RTD and the preliminary quality assessment of AutoCloud+, a
2122 "universal" software toolbox for cloud and cloud-shadow quality layer detection in multi-
2123 sensor multi-angular EO single-date MS imagery, either uncalibrated or radiometrically
2124 calibrated into TOARF, SURF or surface albedo values.

2125 Noteworthy, jointly detection of cloud and cloud-shadow is a typical example of physical
2126 model-based cause-effect relationship, expected to be very difficult to solve by inductive
2127 machine learning-from-data algorithms, such as increasingly popular DCNNs [31], suitable for
2128 learning complex correlations between input and output features, but capable of no inherent
2129 representation of causality ([57]; [27]), in agreement with the well-known dictum that
2130 correlation does not imply causation and vice versa ([10]; [30]; [57]; [27]; [58]; [16]).

2131 In comparison with standard cloud/cloud-shadow software toolboxes available either
2132 open source or free of cost, such as FMask, Sen2Cor and MAJA, AutoCloud+ features several
2133 degrees of novelty at the Marr's four levels of system understanding, specifically,
2134 information/knowledge representation, system architecture, algorithm and implementation.
2135 For example, at the levels of abstraction of information/knowledge representation and system
2136 architecture, typically considered the linchpin of success of an information processing system,
2137 AutoCloud+ is the sole hybrid (combined physical and statistical model-based) 2D image
2138 analysis (spatial context-sensitive and spatial topology-preserving) approach provided with
2139 feedback loops, in agreement with biological cognitive systems [71, 77, 89].

2140 In a proof-of-concept conducted by qualitative photointerpretation of two Sentinel-2 test
2141 images, AutoCloud+ outperformed the standard single-date Sentinel-2 sensor-specific
2142 Sen2Cor and the multi-date MAJA computer program in terms of cloud and cloud-shadow
2143 mapping quality indicators, such as true positive, false positive, true negative and false
2144 negative occurrences.

2145 With regard to an mDMI set of EO OP-Q^2Is proposed in Section 2, potential advantages
2146 of the proposed hybrid feedback 2D image analysis approach for cloud and cloud-shadow
2147 detection are summarized below.

2148 (1) The AutoCloud+ computer program is "fully automated".

2149   i.   It requires no system's free-parameter to be user-defined based on heuristics,
2150   unlike mainstream CV and EO-IU algorithms, such as inductive DCNNs, where *a*
2151   *priori* knowledge is encoded by design based on empirical criteria, a.k.a. trial-and
2152   error. Hence, its ease of use cannot be surpassed by alternative algorithms.

2153   ii.   It requires no supervised data set for inductive learning-from-data. Hence, its costs
2154   in manpower for collecting training data samples are zero and its timeliness from
2155   data collection to output product generation is reduced to computation time.

2156 (2) It employs a known hybrid feedback CV system design (architecture, see Figure 21) whose
2157   processing units have been to a large degree implemented in operating mode and
2158   tested/validated in previous works, e.g., refer to the SIAM and RGBIAM lightweight
2159   computer programs for color space partitioning into color names, superpixel detection and

2160   VQ quality assessment ([10]; [11], [12]; [174]). Hence, its RTD costs in manpower are greatly
2161   reduced.

2162 (3) In line with biological vision, the AutoCloud+ CV subsystem exploits dominant spatial
2163   topological and spatial non-topological information components in combination with
2164   secondary color information, discretized into color names, according to a convergence-of-
2165   evidence approach consistent with symbolic human reasoning, where potentially weak
2166   independent sources of evidence are typically combined to infer strong conjectures. For
2167   example, in two Sentinel-2 test images selected for proof-of-concept, the single-date
2168   AutoCloud+ computer program scored better than standard software toolboxes, such as
2169   the single-date Sen2Cor and the multi-date MAJA, in both cloud and cloud-shadow
2170   detection in terms of true positive, false positive, true negative and false negative
2171   estimates. In addition to showing a superior mapping accuracy, based on theory
2172   AutoCloud+ is expected to score also "high" in robustness to changes in input data and in
2173   scalability to changes in MS imaging sensor's specifications.

2174   To underpin the operational readiness of the AutoCloud+ software toolbox in support of
2175 systematic ESA EO Level 2 product generation, a complete AutoCloud+ software
2176 transcodification into the C programming language is planned for efficiency reasons, together
2177 with GEO-CEOS stage 3 and stage 4 *Val* campaigns to be conducted by independent means at
2178 large spatial extent and time span with multiple MS imaging sensors, in compliance with the
2179 GEO-CEOS *Val* guidelines [180] and the GEO-CEOS QA4EO *Cal/Val* requirements [3].

2180 **Author Contributions:** Conceptualization, A.B. and D.T.; Methodology, A.B. and D.T.; Software, A.B.
2181 and D.T.; Validation, D.T.; Formal Analysis, A.B. and D.T.; Investigation, A.B. and D.T.; Resources, A.B.
2182 and D.T.; Data Curation, D.T.; Writing-Original Draft Preparation, A.B.; Writing-Review & Editing, A.B.
2183 and D.T.; Visualization, A.B. and D.T.; Supervision, D.T.; Project Administration, D.T.; Funding
2184 Acquisition, D.T.

2185 **Funding:** This research was funded in part by the Austrian Science Fund (FWF) through the Doctoral
2186 College GIScience (DK W1237- N23) and by the Austrian Research Promotion Agency (FFG) with regard
2187 to project Sentinel-2 Semantic Data Cube (Sen2cube).

2188 **Acknowledgments:** Andrea Baraldi thanks Prof. Raphael Capurro for his hospitality, patience, politeness
2189 and open-mindedness. The authors wish to thank the Editor-in-Chief, Associate Editor and reviewers
2190 for their competence, patience and willingness to help.

2191 **Conflicts of Interest:** The authors declare no conflict of interest. The funders had no role in the design of
2192 the study; in the collection, analyses, or interpretation of data; in the writing of the manuscript, and in
2193 the decision to publish the results.

References

2195 1.   [1] G. Schaepman-Strub, M. E. Schaepman, T. H. Painter, S. Dangel, and J. V. Martonchik,
2196   "Reflectance quantities in optical remote sensing - Definitions and case studies," Remote Sens.
2197   Environ., vol. 103, pp. 27–42, 2006.

2198 2.   [2] Liang, S. 2004. Quantitative Remote Sensing of Land Surfaces. Hoboken, NJ, USA: John Wiley
2199   and Sons.

2200 3.   [3] Group on Earth Observation / Committee on Earth Observation Satellites (GEO-CEOS). (2010).
2201   A Quality Assurance Framework for Earth Observation, version 4.0.
2202   http://qa4eo.org/docs/QA4EO_Principles_v4.0.pdf

2203 4.   [4] Yang, C., Huang, Q., Li, Z., Liu K. & Hu, F., (2017). Big Data and cloud computing: innovation
2204   opportunities and challenges, International Journal of Digital Earth. 10:1, p.13-53, DOI:
2205   10.1080/17538947.2016.1239771

2206 5.   [5] Group on Earth Observation (GEO). (2005). The Global Earth Observation System of Systems
2207   (GEOSS) 10-Year Implementation Plan. Available online:
2208   http://www.earthobservations.org/docs/10-Year%20Implementation%20Plan.pdf (accessed on 19
2209   January 2012).

6.   [6] Blaschke, T. and Lang, S. (2006). Object based image analysis for automated information extraction-a synthesis. Measuring the Earth II ASPRS Fall Conference, 6-10.

7.   [7] Blaschke, T., G. J. Hay, M. Kelly, S. Lang, P. Hofmann, E. Addink, R. Queiroz Feitosa, F. van der Meer, H. van der Werff, F. van Coillie, and D. Tiede. 2014. "Geographic object-based image analysis - towards a new paradigm." ISPRS J. Photogram. Remote Sens. 87: 180–191.

8.   [8] European Space Agency (ESA). 2015. Sentinel-2 User Handbook, , Standard Document, Issue 1 Rev 2.

9.   [9] Deutsches Zentrum für Luft- und Raumfahrt e.V. (DLR) and VEGA Technologies. 2011. Sentinel-2 MSI – Level 2A Products Algorithm Theoretical Basis Document. Document S2PAD-ATBD-0001, European Space Agency.

10.   [10] Baraldi, A. 2017. "Pre-processing, classification and semantic querying of large-scale Earth observation spaceborne/airborne/terrestrial image databases: Process and product innovations". Ph.D. dissertation in Agricultural and Food Sciences, University of Naples "Federico II", Department of Agricultural Sciences, Italy. Ph.D. defense: 16 May 2017. DOI: 10.13140/RG.2.2.25510.52808. Accessed 30 Jan. 2018. https://www.researchgate.net/publication/317333100_Pre-processing_classification_and_semantic_querying_of_large-scale_Earth_observation_spaceborneairborneterrestrial_image_databases_Process_and_product_innovations

11.   [11] Baraldi, A., Humber, M. L., Tiede, D., & Lang, S. (2018a). GEO-CEOS stage 4 validation of the Satellite Image Automatic Mapper lightweight computer program for ESA Earth observation Level 2 product generation – Part 1: Theory. Cogent Geoscience, Accepted manuscript, 1467357. https://doi.org/10.1080/23312041.2018.1467357

12.   [12] Baraldi, A., Humber, M. L., Tiede, D., & Lang, S. (2018b). GEO-CEOS stage 4 validation of the Satellite Image Automatic Mapper lightweight computer program for ESA Earth observation Level 2 product generation – Part 2: Validation. Cogent Geoscience, Accepted manuscript, 1467254. https://doi.org/10.1080/23312041.2018.1467254

13.   [13] Di Gregorio, A., and L. Jansen. 2000. Land Cover Classification System (LCCS): Classification Concepts and User Manual. FAO: Rome, Italy, FAO Corporate Document Repository. Accessed 10 February 2015. http://www.fao.org/DOCREP/003/X0596E/X0596e00.htm

14.   [14] Swain, P. H., & Davis, S.M. (1978). Remote Sensing: The Quantitative Approach. New York: McGraw-Hill.

15.   15] Capurro, R. & Hjørland, B. (2003). The concept of information. Annual Review of Information Science and Technology, 37, 343-411.

16.   [16] Sonka, M., Hlavac, V., & Boyle, R. (1994). Image Processing, Analysis and Machine Vision. London, U.K.: Chapman & Hall.

17.   [17] Fonseca, F., Egenhofer, M., Agouris, P. & Camara, G. (2002). Using ontologies for integrated geographic information systems, *Trans. in GIS*, 6, 231–257.

18.   [18] Growe, S. (1999). Knowledge-based interpretation of multisensor and multitemporal remote sensing images. *Int. Archives of Photogrammetry and Remote Sensing*, 32, Part 7–4–3 W6, Valladolid, Spain, 3–4 June, 1999.

19.   [19] Laurini, R. & Thompson, D. (1992*). Fundamentals of Spatial Information Systems*. London, UK: Academic Press.

20.   [20] Matsuyama, T., and V. S. Hwang. (1990). SIGMA – A Knowledge-based Aerial Image Understanding System. New York, NY: Plenum Press.

21.   [21] Sowa, J. (2000). Knowledge representation: Logical, philosophical, and computational foundations, Brooks Cole Publishing Co., Pacific Grove (CA).

22.   [22] Ahlqvist, O. 2005. "Using uncertain conceptual spaces to translate between land cover categories." *Int. J. Geographic. Info. Science* 19: 831–857.

23.   [23] Bossard, M., J. Feranec and J. Otahel. 2000. "CORINE land cover technical guide – Addendum 2000, Technical report No 40." European Environment Agency.

24.   [24] Lillesand, T., & Kiefer, R. (1979). *Remote Sensing and Image Interpretation*, New York: John Wiley & Sons.

2263   25.   [25] Belward, A., Ed. 1996. The IGBP-DIS Global 1Km Land Cover Data Set "DISCover": Proposal
2264   and Implementation Plans. IGBP-DIS Working Paper 13. International Geosphere Biosphere
2265   Programme. European Commission Joint Research Center, Ispra, Varese, Italy.
2266   26.   [26] Dumitru C. O., S. Cui, G. Schwarz, and M. Datcu. 2015. "Information content of very-high-
2267   resolution SAR images: Semantics, geospatial context, and ontologies." IEEE J. Selected Topics
2268   Applied Earth Obs. Remote Sens. 8(4): 1635 – 1650.
2269   27.   [27] Marcus, G. 2018. "Deep Learning: A Critical Appraisal." arXiv: 1801.00631. Accessed: 16 Jan.
2270   2018. https://arxiv.org/ftp/arxiv/papers/1801/1801.00631.pdf
2271   28.   [28] Gutman, G., A. C. Janetos, C. O. Justice, E. F. Moran, J. F. Mustard, R. R. Rindfuss, D. Skole, B.
2272   L. Turner, M. A. Cochrane, Eds. 2004. *Land Change Science*. Dordrecht, The Netherlands: Kluwer.
2273   29.   [29] Bishop, C. M. 1995. Neural Networks for Pattern Recognition. Oxford, U.K.: Clarendon.
2274   30.   [30] Cherkassky, V., and F. Mulier. 1998. Learning from Data: Concepts, Theory, and Methods. New
2275   York, NY: Wiley.
2276   31.   [31] Cimpoi M., S. Maji, I. Kokkinos I., and A. Vedaldi. 2014. "Deep filter banks for texture
2277   recognition, description, and segmentation," CoRR, abs/1411.6836.
2278   32.   [32] Lipson, H. 2007. "Principles of modularity, regularity, and hierarchy for scalable systems,"
2279   Journal of Biological Physics and Chemistry, 7, 125–128.
2280   33.   [33] Wolpert, D. H. (1996). The lack of a priori distinctions between learning algorithms. Neural
2281   Computation, 8(7), 1341–1390.
2282   34.   [34] Wolpert, D. H., & Macready, W. G. (1997). No free lunch theorems for optimization. IEEE Trans
2283   on Evolutionary Computation, 1(1), 67–82.
2284   35.   [35] Zhaoxiang, Z.; Iwasaki, A.; Guodong, X.; Jianing, S. Small Satellite Cloud Detection Based On
2285   Deep Learning and Image Compression. Preprints 2018, 2018020103 (doi:
2286   10.20944/preprints201802.0103.v1.   Available   online:
2287   https://www.preprints.org/manuscript/201802.0103/v1 (accessed on 7 Aug. 2018).
2288   36.   [36] Baraldi, A. (2015). "Automatic Spatial Context-Sensitive Cloud/Cloud-Shadow Detection in
2289   Multi-Source Multi-Spectral Earth Observation Images – AutoCloud+," Invitation to tender
2290   ESA/AO/1-8373/15/I-NB – "VAE: Next Generation EObased Information Services", DOI:
2291   10.13140/RG.2.2.34162.71363.   arXiv: 1701.04256.   Date: 8 Jan. 2017.   Available online:
2292   https://arxiv.org/ftp/arxiv/papers/1701/1701.04256.pdf
2293   37.   [37] Gascon, F., Bouzinac, C., Thépaut, O., Jung, M., Francesconi, B., Louis, J., Lonjou, V., Lafrance,
2294   B., Massera, S., Gaudel-Vacaresse, A., Languille, F., Alhammoud, B., Viallefont, F., Pflug, B.,
2295   Bieniarz, J., Clerc, S., Pessiot, L., Trémas, T., Cadau, E., De Bonis, R., Isola, C., Martimort, P.,
2296   Fernandez, V., (2017). Copernicus Sentinel-2A Calibration and Products Validation Status. Remote
2297   Sens. Vol. 9, no. 6, 584. doi:10.3390/rs9060584
2298   38.   [38] Goodwin, N.R., Collett, L.J., Denham, R.J., Flood, N., Tindall, D., (2013). Cloud and cloud-
2299   shadow screening across Queensland, Australia: An automated method for Landsat TM/ETM+ time
2300   series. Remote Sensing of Environment, 134, p.50–65. http://dx.doi.org/10.1016/j.rse.2013.02.019.
2301   39.   [39] Hagolle, O., Huc, M., Desjardins, C., Auer, S., Richter, R., (2017a). *MAJA Algorithm Theoretical
2302   Basis Document*, Available at: https://zenodo.org/record/1209633#.W2ffFNIzZaQ.
2303   40.   [40] Olivier Hagolle, Bastien Rouquié, Camille Desjardins, Aliaksei Makarau, Magdalena Main-
2304   knorn, Guillaume Rochais, Bringfried Pug, 2017b. Recent advances in cloud detection and
2305   atmospheric correction applied to time series of high resolution images, RAQRS, 19 Sept 2017.
2306   Available   online:
2307   https://www.researchgate.net/profile/Olivier_Hagolle2/publication/320402521_Recent_advances_i
2308   n_cloud_detection_and_atmospheric_correction_applied_to_time_series_of_high_resolution_imag
2309   es/links/59eeee074585154350e83669/Recent-advances-in-cloud-detection-and-atmospheric-
2310   correction-applied-to-time-series-of-high-resolution-images.pdf (accessed on 8 Aug. 2018)
2311   41.   [41] Hagolle, O., Huc, M., Pascual, D.V., Dedieu, G., (2010). A multi-temporal method for cloud
2312   detection, applied to FORMOSAT-2, VENμS, LANDSAT and SENTINEL-2 images. Remote Sensing
2313   of Environment, 114(8), p.1747–1755. doi:10.1016/j.rse.2010.03.002
2314   42.   [42] Hollstein, A., Segl, K., Guanter, L., Brell, M., Enesco, M., (2016). Ready-to-use methods for the
2315   detection of clouds, cirrus, snow, shadow, water and clear sky pixels in Sentinel-2 MSI images.
2316   Remote Sensing. 8, p. 1–18. doi:10.3390/rs8080666

2317   43.   [43] C. Huang, N. Thomas, S. N. Goward, J. G. Masek, Z. Zhu, J. R. G. Townshend, and J. E.
2318   Vogelmann, "Automated masking of cloud and cloud shadow for forest change analysis using
2319   Landsat images," International Journal of Remote Sensing, vol. 31, no. 20, pp. 5449-5464, 2010.
2320   44.   [44] Hughes M.J. & Hayes, D.J. (2014). Automated detection of cloud and cloud shadow in single-
2321   date Landsat imagery using neural networks and spatial post-processing. Remote Sensing, 6(6),
2322   4907–4926. doi:10.3390/rs6064907
2323   45.   [45] R. R. Irish, J. L. Barker, S. N. Goward, and T. Arvidson, "Characterization of the Landsat-7 ETM+
2324   Automated Cloud-Cover Assessment (ACCA) Algorithm," Photogrammetric Engineering &
2325   Remote Sensing, vol. 72, no. 10, pp. 1179–1188, October 2006.
2326   46.   [46] Ju, J.; Roy, D.P. The availability of cloud-free Landsat ETM+ data over the conterminous United
2327   States and globally. Remote Sens. Environ. 2008, 112, 1196–1211.
2328   47.   [47] K. V. Khlopenkov and A.P. Trishchenko, "SPARC: New cloud, snow, and cloud shadow
2329   detection scheme for historical 1-km AVHRR data over Canada," Journal of Atmospheric and
2330   Oceanic Technology, vol. 24, pp. 322–343, 2007.
2331   48.   [48] S. Le Hégarat-Mascle and C. André, "Reduced false alarm automatic detection of clouds and
2332   shadows on SPOT images using simultaneous estimation," Proc. SPIE, 6748-46, vol. 1, p. 1-12, 2010.
2333   49.   [49] Lück W., and A. van Niekerk. 2016. "Evaluation of a rule-based compositing technique for
2334   Landsat-5 TM and Landsat-7 ETM+ images," Int. J. of Applied Earth Observation and
2335   Geoinformation 47: 1–14.
2336   50.   [50] Luo, Y., Trishchenko, A.P. & Khlopenkov, K. V., (2008). Developing clear-sky, cloud and cloud-
2337   shadow mask for producing clear-sky composites at 250-meter spatial resolution for the seven
2338   MODIS land bands over Canada and North America. Remote Sensing of Environment, 112(12),
2339   p.4167–4185. http://dx.doi.org/10.1016/j.rse.2008.06.010.
2340   51.   [51] Zhu, Z. & Woodcock, C.E., (2012). Object-based cloud and cloud-shadow detection in Landsat
2341   imagery. Remote Sensing of Environment, 118, p.83–94.
2342   52.   [52] Zhu, Z., Wang, S. & Woodcock, C.E., (2015). Improvement and expansion of the Fmask
2343   algorithm : cloud , cloud-shadow , and snow detection for Landsats 4 – 7 , 8 , and Sentinel 2 images.
2344   Remote Sensing of Environment, 159, p.269–277. http://dx.doi.org/10.1016/j.rse.2014.12.014
2345   53.   [53] Zhe Zhu and Woodcock, C. E. 2014. Automated cloud, cloud shadow, and snow detection in
2346   multitemporal Landsat data: An algorithm designed specifically for monitoring land cover change.
2347   Remote Sens. Environ., vol. 152, pp. 217–234, 2014.
2348   54.   [54] Main-Knorn, M., Louis, J., Hagolle, O., Müller-Wilm, U., and Alonso, K., The Sen2Cor and
2349   MAJA cloud masks and classification products, 2nd Sentinel-2 Validation Team Meeting, 29-31 Jan.
2350   2018, ESA-ESRIN, Frascati (Rome), Italy.
2351   55.   [55] Foga, S., P. Scaramuzza, S. Guo, Z. Zhu, R.Dilley Jr, T. Beckmann, G. Schmidt, J. Dwyer, M.
2352   Hughes, and B. Laue. 2017. "Cloud detection algorithm comparison and validation for operational
2353   Landsat data products." Remote Sens. Environ. 194: 379–390.
2354   56.   [56] U.S. Geological Survey (USGS), 2018d. U.S. Landsat Analysis Ready Data (ARD) Artifacts.
2355   Available online: https://landsat.usgs.gov/us-landsat-ard-artifacts (accessed on 15 July 2018).
2356   57.   [57] Chomba Bupe, Is Deep Learning fundamentally flawed and hitting a wall? Was Gary Marcus
2357   correct in pointing out Deep Learning's flaws?, Quora, 2018. Available online:
2358   https://www.quora.com/Is-Deep-Learning-fundamentally-flawed-and-hitting-a-wall-Was-Gary-
2359   Marcus-correct-in-pointing-out-Deep-Learnings-flaws. Accessed 7 Aug. 2018.
2360   58.   [58] Pearl, J. 2009. Causality: Models, Reasoning and Inference. New York (NY): Cambridge
2361   University Press.
2362   59.   [59] Mazzuccato M. and D. Robinson. (2017). Market Creation and the European Space Agency.
2363   European Space Agency (ESA) Report. Available at: https://marianamazzucato.com/wp-
2364   content/uploads/2016/11/Mazzucato_Robinson_Market_creation_and_ESA.pdf
2365   60.   [60] A. Baraldi, "Impact of radiometric calibration and specifications of spaceborne optical imaging
2366   sensors on the development of operational automatic remote sensing image understanding
2367   systems," IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens., vol. 2, no. 2, pp. 104-134, June 2009.
2368   61.   [61] Pacifici, F. Atmospheric Compensation in Satellite Imagery. US Patent 9396528B2, 19 July 2016.

2369 62. [62] F. Pacifici, N. Longbotham, and W. J. Emery, "The Importance of Physical Quantities for the
2370   Analysis of Multitemporal and Multiangular Optical Very High Spatial Resolution Images," IEEE
2371   Trans. Geosci. Remote Sens., 2014.
2372 63. [63] Richter, R., & Schläpfer, D. (2012a). Atmospheric / Topographic Correction for Satellite Imagery
2373   –   ATCOR-2/3   User   Guide,   Version   8.2   BETA.   Available   online:
2374   http://www.dlr.de/eoc/Portaldata/60/Resources/dokumente/5_tech_mod/atcor3_manual_2012.pdf.
2375   Accessed 12 April 2013.
2376 64. [64] Richter, R., & Schläpfer, D. (2012b). Atmospheric / Topographic correction for airborne imagery
2377   –   ATCOR-4   User   Guide,   Version   6.2   BETA.   Available   online:
2378   http://www.dlr.de/eoc/Portaldata/60/Resources/dokumente/5_tech_mod/atcor4_manual_2012.pdf.
2379   Accessed 12 April 2013.
2380 65. [65] Dorigo, W., Richter, R., Baret, F., Bamler, R., & Wagner, W. (2009). Enhanced automated canopy
2381   characterization from hyperspectral data by a novel two step radiative transfer model inversion
2382   approach. Remote Sens., 1, 1139-1170.
2383 66. [66] Schläpfer, D., Richter, R., & Hueni, A. (2009). Recent developments in operational atmospheric
2384   and radiometric correction of hyperspectral imagery, in Proc. 6th EARSeL SIG IS Workshop, 16-19
2385   March 2009. Retrieved July, 14, 2012 from http://www.earsel6th.tau.ac.il/~earsel6/CD/PDF/earsel-
2386   PROCEEDINGS/3054%20Schl%20pfer.pdf
2387 67. [67] Andrea Baraldi and Dirk Tiede, "AutoCloud+, a "universal" single-date multi-sensor physical
2388   and statistical model-based spatial context-sensitive cloud/cloud-shadow detector in multi-spectral
2389   Earth observation imagery," GEOBIA 2018, Montpellier, France, June 18-22, 2018.
2390 68. [68] M. Bertero, T. Poggio, and V. Torre, "Ill-posed problems in early vision," Proc. of the IEEE, vol.
2391   76, no. 8, pp. 869–889, Aug. 1988.
2392 69. [69] Marr, D. (1982). Vision. New York, NY: Freeman and C.
2393 70. [70] Serra, R. and Zanarini, G. (1990). Complex Systems and Cognitive Processes, Berlin: Springer-
2394   Verlag.
2395 71. [71] D. Parisi, "La scienza cognitiva tra intelligenza artificiale e vita artificiale," in Neurosceinze e
2396   Scienze dell'Artificiale: Dal Neurone all'Intelligenza, Bologna, Italy: Patron Editore, 1991.
2397 72. [72] G. A. Miller, "The cognitive revolution: a historical perspective", in Trends in Cognitive
2398   Sciences, vol. 7, pp. 141-144, 2003.
2399 73. [73] F. J. Varela, E. Thompson, and E. Rosch, The Embodied Mind: Cognitive Science and Human
2400   Experience. Cambridge, MA: MIT Press, 1991.
2401 74. [74] F. Capra and P. L. Luisi, The Systems View of Life: A Unifying Vision. Cambridge, UK:
2402   Cambridge University Press, 2014.
2403 75. [75] Castelletti, D., Pasolli, L,. Bruzzone, L., Notarnicola, C., & Demir, B. (2016). A novel hybrid
2404   method for the correction of the theoretical model inversion in bio/geophysical parameter
2405   estimation. *IEEE Trans. Geosci. Remote Sens.*, XX(Y), 1–11.
2406 76. [76] Baraldi, A., Gironda, M., & Simonetti, D. (2010c). Operational two-stage stratified topographic
2407   correction of spaceborne multi-spectral imagery employing an automatic spectral rule-based
2408   decision-tree preliminary classifier. IEEE Trans. Geosci. Remote Sensing, 48(1), 112-146.
2409 77. [77] J. Piaget, Genetic Epistemology, New York: Columbia University Press, 1970.
2410 78. [78] A. Baraldi and L. Boschetti, "Operational automatic remote sensing image understanding
2411   systems: Beyond   Geographic   Object-Based   and   Object-Oriented   Image   Analysis
2412   (GEOBIA/GEOOIA) - Part 1: Introduction," Remote Sens., vol. 4, no. 9, pp. 2694-2735, 2012.
2413 79. [79] A. Baraldi and L. Boschetti, "Operational automatic remote sensing image understanding
2414   systems: Beyond   Geographic   Object-Based   and   Object-Oriented   Image   Analysis
2415   (GEOBIA/GEOOIA) - Part 2: Novel system architecture, information/knowledge representation,
2416   algorithm design and implementation," Remote Sens., vol. 4, no. 9, pp. 2768-2817, 2012.
2417 80. [80] National Aeronautics and Space Administration (NASA). (2016). Data Processing Levels.
2418   [Online].   Available: https://science.nasa.gov/earth-science/earth-science-data/data-processing-
2419   levels-for-eosdis-data-products.   Accessed on December 20, 2016.
2420 81. [81] Fowler, M. (2003). UML Distilled Third Edition. Boston, MS, USA: Addison-Wesley.
2421 82. [82] A. Baraldi, D. Tiede, M. Sudmanns, M. Belgiu, and S. Lang, "Automated near real-time Earth
2422   observation Level 2 product generation for semantic querying," GEOBIA 2016, 14-16 Sept. 2016,

University of Twente Faculty of Geo-Information and Earth Observation (ITC), Enschede, The Netherlands.

83. [83] A. Baraldi, D. Tiede, M. Sudmanns, and S. Lang, 2017a. "Systematic ESA EO Level 2 product generation as pre-condition to semantic content-based image retrieval and information/knowledge discovery in EO image databases," 2017 Conf. on Big Data From Space, BiDS'17, Toulouse, France, 28-30 March, 2017.

84. [84] D. Tiede, A. Baraldi, M Sudmanns, M. Belgiu and S. Lang, "Architecture and prototypical implementation of a semantic querying system for big earth observation image bases," European Journal of Remote Sensing, vol. 50, no. 1, pp. 452-463, Aug. 2017. DOI: 10.1080/22797254.2017.1357432

85. [85] Augustin, H., Sudmanns, M., Tiede, D., and Baraldi, A., 2018. A semantic Earth observation data cube for monitoring environmental changes during the Syrian conflict. AGIT 2018, Salzburg, Austria, July 3-6, 2018, GI_Forum Journal, vol. 1, pp. 214–227. doi:10.1553/giscience2018_01_s214

86. [86] M. Sudmanns, D. Tiede, S. Lang and A. Baraldi, "Semantic and syntactic interoperability in online processing of big Earth observation data", Int. J. of Digital Earth, vol. 11, no. 1, pp. 95-112, 2018. DOI: 10.1080/17538947.2017.1332112.

87. [87] Smeulders, A., Worring, M., Santini, S., Gupta, A., & Jain, R. (2000). Content-based image retrieval at the end of the early years. *IEEE Trans. Pattern Anal. Machine Intell.*, 22(12), 1349-1380.

88. [88] Frintrop, S. (2011). Computational visual attention. In *Computer Analysis of Human Behavior, Advances in Pattern Recognition*, A. A. Salah and T. Gevers, Eds,. Springer .

89. [89] Tsotsos, J. K. (1990). Analyzing vision at the complexity level. *Behavioral and Brain Sciences*, 13, 423-469.

90. [90] Hadamard, J. 1902. "Sur les problemes aux derivees partielles et leur signification physique," Princet. Univ. Bull. 13: 49–52.

91. [91] Baraldi, A., L. Boschetti, and M. Humber. 2014. "Probability sampling protocol for thematic and spatial quality assessments of classification maps generated from spaceborne/airborne very high resolution images." IEEE Trans. Geosci. Remote Sens. 52(1): 701-760.

92. [92] J. D. Shepherd and J. R. Dymond, "BRDF correction of vegetation in AVHRR imagery," Remote Sens. Environ., vol. 74, no. 3, pp. 397–408, Dec. 2000.

93. [93] T. Danaher, "An empirical BRDF correction for landsat TM and ETM+ imagery," in Proc. 11th Aust. Remote Sens. Photogramm. Conf., Brisbane, Australia, 2002, pp. 2654–2657.

94. [94] A. Wu, Z. Li, and J. Cihlar, "Effects of land cover type and greenness on advanced very high resolution radiometer bidirectional reflectances: Analysis and removal," J. Geophys. Res., vol. 100, no. D5, pp. 9179–9192, Jan 1995.

95. [95] Ghahramani, Z. (2011). Bayesian nonparametrics and the probabilistic approach to modelling. *Phil. Trans. R. Soc. A*, 1–27.

96. [96] Wikipedia (2017). Bayesian inference. [Online]. Available: https://en.wikipedia.org/wiki/Bayesian_inference. Accessed 14 March 2017.

97. [97] DiCarlo, J. (2017). The Science of Natural intelligence: Reverse engineering primate visual perception. Keynote, CVPR17 Conference. Date 5 Jan. 2018. Available online: https://www.youtube.com/watch?v=ilbbVkIhMgo

98. [98] du Buf ,H., & Rodrigues, J. (2007). Image morphology: from perception to rendering. In *IMAGE - Computational Visualistics and Picture Morphology*.

99. [99] Kosslyn, S. M. (1994). *Image and Brain*. Cambridge, MA: MIT Press.

100. [100] Serre, T., Wolf, L., Bileschi, S., Riesenhuber, M., & Poggio, T. (2007). Robust object recognition with cortex-like mechanisms, *IEEE Trans. Pattern Anal. Machine Intell.*, 29(3), 411–426.

101. [101] Sylvester, J., & Reggia, J. (2016). Engineering neural systems for high-level problem solving. Neural Networks, 79, 37–52.

102. [102] Vecera, S., & Farah, M. (1997). Is visual image segmentation a bottom-up or an interactive process?. *Percept. Psychophys.*, 59, 1280–1296.

103. [103] P. Quinlan, "Marr's Vision 30 years on: From a personal point of view", Perception, vol. 41, pp. 1009 – 1012, 2012.

2475   104. [104] Duke University, (2016). Measurement: Process and Outcome Indicators, Duke Center for
2476   Instructional   Technology.   [Online]   Available:
2477   http://patientsafetyed.duhs.duke.edu/module_a/measurement/measurement.html

2478   105. [105] Boschetti, L., S. P. Flasse, and P. A. Brivio. 2004. "Analysis of the conflict between omission
2479   and commission in low spatial resolution dichotomic thematic products: The Pareto boundary."
2480   Remote Sens. Environ. 91: 280–292.

2481   106. [106] Couclelis, H. (2012). What GIScience is NOT: Three theses. Invited speaker, GIScience '12 Int.
2482   Conf., Columbus, Ohio, Sept. 18-21, 2012.

2483   107. [107] Etzioni, O. (2017). What shortcomings do you see with deep learning?. Quora. Date: 8 Jan.
2484   2018. Available online: https://www.quora.com/What-shortcomings-do-you-see-with-deep-
2485   learning

2486   108. [108] C. Shannon, "A mathematical theory of communication," Bell System Technical Journal, vol.
2487   27, pp. 379–423 and 623–656, 1948.

2488   109. [109] Tobler, W.R., 1970, A computer movie simulating urban growth in the Detroit Region.
2489   Economic Geography, 46, pp. 234–240.

2490   110. [110] P. A. Longley, M. F. Goodchild, D. J. Maguire D. W. Rhind, Geographic Information Systems
2491   and Science, Second Edition. New York: Wiley, 2005.

2492   111. [111] Andrea Baraldi, Stefan Lang, Dirk Tiede, and Thomas Blaschke, 2018c. "Earth observation big
2493   data analytics in operating mode for GIScience applications – The (GE)OBIA acronym(s)
2494   reconsidered," GEOBIA 2018, Montpellier, France, June 18-22, 2018.

2495   112. [112] Stefan Lang, Andrea Baraldi, Dirk Tiede, Geoffrey Hay and Thomas Blaschke, "Towards a
2496   (GE)OBIA 2.0 manifesto – Achievements and open challenges in information & knowledge
2497   extraction from big Earth data," GEOBIA 2018, Montpellier, France, June 18-22, 2018.

2498   113. [113] Baraldi, A., Durieux, L., Simonetti, D., Conchedda, G., Holecz, F., & Blonda, P. (2010a).
2499   Automatic spectral rule-based preliminary classification of radiometrically calibrated SPOT-4/-
2500   5/IRS, AVHRR/MSG, AATSR, IKONOS/QuickBird/OrbView/GeoEye and DMC/SPOT-1/-2 imagery
2501   – Part I: System design and implementation," IEEE Trans. Geosci. Remote Sensing, 8(3), 1299 - 1325.

2502   114. [114] Baraldi, A., Durieux, L., Simonetti, D., Conchedda, G., Holecz, F., & Blonda, P. (2010b).
2503   "Automatic spectral rule-based preliminary classification of radiometrically calibrated SPOT-4/-
2504   5/IRS, AVHRR/MSG, AATSR, IKONOS/QuickBird/OrbView/GeoEye and DMC/SPOT-1/-2 imagery
2505   – Part II: Classification accuracy assessment," IEEE Trans. Geosci. Remote Sensing, vol. 48, no. 3, pp.
2506   1326 - 1354, March 2010.

2507   115. [115] Goodchild, M. F., Yuan M., & Cova, T. J. (2007). Towards a general theory of geographic
2508   representation in GIS, International Journal of Geographical Information Science, 21:3, 239-260.

2509   116. [116] Bernus P., and O. Noran. 2017. Data Rich – But Information Poor. In: Camarinha-Matos L.,
2510   Afsarmanesh H., Fornasiero R. (eds) *Collaboration in a Data-Rich World*: PRO-VE 2017. IFIP Advances
2511   in Information and Communication Technology, 506: 206-214.

2512   117. [117] European Union, Copernicus Observer - The upcoming Copernicus Data and Information
2513   Access Services (DIAS), 26/05/2017. Available online: http://copernicus.eu/news/upcoming-
2514   copernicus-data-and-information-access-services-dias (accessed on 15 July 2018).

2515   118. [118] European Union, The DIAS: User-friendly Access to Copernicus Data and Information, June
2516   2018.   Available   online:
2517   http://copernicus.eu/sites/default/files/Data_Access/Data_Access_PDF/Copernicus_DIAS_Factshee
2518   t_June2018.pd (accessed on 15 July 2018).

2519   119. [119] Committee on Earth Observation Satellites (CEOS), 2018. CEOS Analysis Ready Data – CEOS
2520   Analysis Ready Data for Land (CARD4L) products. Available online: http://www.ceos.org/ard/
2521   (accessed on 4 May 2018).

2522   120. [120] U.S. Geological Survey (USGS), 2018b. U.S. Landsat Analysis Ready Data (ARD). Available
2523   online: https://landsat.usgs.gov/ard (accessed on 15 July 2018).

2524   121. [121] U.S. Geological Survey (USGS), 2018a. U.S. LANDSAT ANALYSIS READY DATA (ARD)
2525   DATA FORMAT CONTROL BOOK (DFCB) Version 4.0, January 2018. Available online:
2526   https://landsat.usgs.gov/sites/default/files/documents/LSDS-1873_US_Landsat_ARD_DFCB.pdf
2527   (accessed on 15 July 2018).

2528   122. [122] Guoqing Zhou, Architecture of Future Intelligent Earth Observing Satellites (FIEOS) in 2010
2529   and Beyond, National Aeronautics and Space Administration Institute of Advanced Concepts
2530   (NASA-NIAC), Technical Report, 2001.
2531   123. [123] GISCafe News, Earth-i led consortium secures grant from UK Space Agency, July 19, 2018.
2532   Available   online:
2533   https://www10.giscafe.com/nbc/articles/view_article.php?section=CorpNews&articleid=1600936
2534   (accessed on 15 July 2018).
2535   124. [124] Vermote, E., and N. Saleous. 2007. "LEDAPS surface reflectance product description - Version
2536   2.0." University of Maryland at College Park /Dept Geography and NASA/GSFC Code 614.5
2537   125. [125] D. Riaño, E. Chuvieco, J. Salas, and I. Aguado, "Assessment of different topographic
2538   corrections in Landsat TM data for mapping vegetation types," IEEE Trans. Geosci. Remote Sens.,
2539   vol. 41, no. 5, pp. 1056–1061, May 2003.
2540   126. [126] C. Leprieur, J. M. Durand, and J. L. Peyron, "Influence of topography on forest reflectance
2541   using Landsat Thematic Mapper and digital terrain data," Photogramm. Eng. Remote Sens., vol. 54,
2542   pp. 491–496, 1988.
2543   127. [127] A. G. Thomson and C. Jones, "Effects of topography on radiance from upland vegetation in
2544   North Wales," Int. J. Remote Sens., vol. 11, pp. 829–840, 1990.
2545   128. [128] Bishop, M. P. & Colby, J. D. 2002. Anisotropic reflectance correction of SPOT-3 HRV imagery.
2546   Int. J. Remote Sens., 23(10), 2125-2131.
2547   129. [129] Bishop, M. P., Shroder, J. F., & Colby, J. D. 2003. Remote sensing and geomorphometry for
2548   studying relief production in high mountains," Geomorphology, 55(1–4), 345-361.
2549   130. [130] N. Hunt and S. Tyrrell, Stratified Sampling. Coventry University, 2012. [Online] Available:
2550   http://www.coventry.ac.uk/ec/~nhunt/meths/strati.html
2551   131. [131] T. Poggio, "The Levels of Understanding framework, revised," Computer Science and
2552   Artificial Intelligence Laboratory, Technical Report, MIT-CSAIL-TR-2012-014, CBCL-308, May 31,
2553   2012.
2554   132. [132] Q. Iqbal and J. K. Aggarwal, "Image retrieval via isotropic and anisotropic mappings," in Proc.
2555   IAPR Workshop Pattern Recognit. Inf. Syst., Setubal, Portugal, Jul. 2001, pp. 34–49.
2556   133. [133] Pessoa, L. (1996). Mach Bands: How Many Models are Possible? Recent Experimental Findings
2557   and Modeling Attempts. *Vision Res.*, 36(19), 3205–3227.
2558   134. [134] Baatz, M. & Schäpe, A. (2000). Multiresolution Segmentation: An Optimization Approach for
2559   High Quality Multi-Scale Image Segmentation. In Angewandte Geographische
2560   Informationsverarbeitung XII; Strobl, J., Ed.; Herbert Wichmann Verlag: Berlin, Germany, 58, 12–
2561   23.
2562   135. [135] Espindola, G. M., Camara, G., Reis, I. A., Bins, L. S., & Monteiro, A. M. (2006). Parameter
2563   selection for region-growing image segmentation algorithms using spatial autocorrelation, Int. J.
2564   Remote Sens., 27(14), 3035–3040.
2565   136. [136] Canny, J. (1986). A computational approach to edge detection. *IEEE Trans. Pattern Anal. Mach.*
2566   *Intell.*, 8(6), 679–698.
2567   137. [137] U.S. Geological Survey (USGS), 2018c. Landsat Surface Reflectance Code (LaSRC) v1.2.0.
2568   Available online: https://github.com/USGS-EROS/espa-surface-reflectance/tree/lasrc_v1.2.0/
2569   (accessed on 15 July 2018).
2570   138. [138] Baraldi, A., & Humber, M. (2015). Quality assessment of pre-classification maps generated
2571   from spaceborne/airborne multi-spectral images by the Satellite Image Automatic Mapper™ and
2572   Atmospheric/Topographic Correction™-Spectral Classification software products: Part 1 – Theory.
2573   *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, 8(3), 1307-1329.
2574   139. [139] Baraldi, A., Humber, M., & Boschetti, L. (2013). Quality assessment of pre-classification maps
2575   generated from spaceborne/airborne multi-spectral images by the Satellite Image Automatic
2576   Mapper™ and Atmospheric/Topographic Correction™-Spectral Classification software products:
2577   Part 2 – Experimental results. *Remote Sens.*, 5, 5209-5264.
2578   140. [140] Burt, P. and Adelson, E. (1983). The laplacian pyramid as a compact image code. IEEE Trans.
2579   Commun. 31(4): 532–540.
2580   141. [141] Jain, A. & Healey, G. (1998). A multiscale representation including opponent color features for
2581   texture recognition. IEEE Trans. Image Process., 7, 124–128.

142. [142] Axios (2017), Artificial intelligence pioneer, Geoffrey Hinton, says we need to start over, Sept. 15, 2017. Date: 8 Jan. 2018. Available online: https://www.axios.com/artificial-intelligence-pioneer-says-we-need-to-start-over-1513305524-f619efbd-9db0-4947-a9b2-7a4c310a28fe.html

143. [143] Meer, P. (2012). Are we making real progress in computer vision today?, Image and Vision Computing, 30, 472–473.

144. [144] Nguyen, A., Yosinski, J., & Clune, J. (2014). Deep Neural Networks are Easily Fooled: High Confidence Predictions for Unrecognizable Images. arXiv: 1412.1897. Date: 8 Jan. 2018. Available online: https://arxiv.org/pdf/1412.1897.pdf

145. [145] Szegedy, C., Zaremba, W., Sutskever, I., Bruna, J., Erhan, D., Goodfellow, I., Fergus, R. (2013). Intriguing properties of neural networks. arXiv: 1312.6199. Date: 8 Jan. 2018. Available online: https://arxiv.org/pdf/1312.6199.pdf

146. [146] Bartoš, M. 2017. Cloud and Shadow Detection in Satellite Imagery. Master Thesis in Computer Vision and Image Processing, Faculty of Electrical Engineering, Department of Cybernetics, Czech Technical University in Prague.

147. [147] Martinetz, T., Berkovich, G., & Schulten, K. (1994). Topology representing networks. Neural Networks, 7(3), 507–522.

148. 148] Berlin, B., and P. Kay. 1969. "Basic color terms: their universality and evolution." Berkeley: University of California.

149. [149] Griffin, L. D. 2006. "Optimality of the basic color categories for classification." *J. R. Soc. Interface* 3: 71–85.

150. [150] Chavez, P. (1988). An improved dark-object subtraction technique for atmospheric scattering correction of multispectral data. *Remote Sens. Environ*. 24:459–479.

151. [151] Baraldi, A., V. Puzzolo, P. Blonda, L. Bruzzone, and C. Tarantino. (2006). Automatic spectral rule-based preliminary mapping of calibrated Landsat TM and ETM+ images. IEEE Trans. Geosci. Remote Sens. 44, p. 2563-2586.

152. [152] Adams, J. B., Donald, E. S., Kapos, V., Almeida Filho, R., Roberts, D. A., Smith, M. O., & Gillespie, A. R. (1995). Classification of multispectral images based on fractions of endmembers: Application to land-cover change in the Brazilian Amazon. Remote Sens. Environ. 52, 137-154.

153. [153] Kuzera, K., and R. G. Pontius jr. 2008. "Importance of matrix construction for multiple-resolution categorical map comparison." *GIScience and Remote Sens*. 45: 249–274.

154. [154] Pontius Jr., R.G., and J. Connors. 2006. "Expanding the conceptual, mathematical and practical methods for map comparison," in Caetano, M. and Painho, M. (Eds). Proceedings of the 7th International Symposium on Spatial Accuracy Assessment in Natural Resources and Environmental Sciences, Lisbon. Instituto Geográfico Português, 5 – 7 July 2006, 64-79.

155. [155] Stehman, S. V., and R L. Czaplewski. 1998. "Design and analysis for thematic map accuracy assessment: Fundamental principles." Remote Sens. Environ. 64: 331-344.

156. [156] Congalton, R. G., and K. Green. 1999. *Assessing the Accuracy of Remotely Sensed Data*. Boca Raton, FL, USA: Lewis Publishers.

157. [157] Lunetta, R., and D. Elvidge. 1999. *Remote Sensing Change Detection: Environmental Monitoring Methods and Applications*. London, UK: Taylor & Francis.

158. [158] Tyagi, V. (2017). Content-Based Image Retrieval Ideas, Influences, and Current Trends, Singapore: Springer.

159. [159] Nagao, M., & Matsuyama, T. (1980). *A Structural Analysis of Complex Aerial Photographs*. Plenum Press, New York.

160. [160] Shackelford, K., and C. H. Davis. 2003a. "A combined fuzzy pixel-based and object-based approach for classification of high-resolution multispectral data over urban areas, " IEEE Trans. Geosci. Remote Sens. 41(10): 2354 – 2363.

161. [161] Shackelford, K., and C. H. Davis. 2003b. "A hierarchical fuzzy classification approach for high-resolution multispectral data over urban areas," IEEE Trans. Geosci. Remote Sens. 41(10): 1920–1932.

162. [162] Wenwen Li, M. F. Goodchild and R. L. Church. 2013. "An efficient measure of compactness for 2D shapes and its application in regionalization problems." Int. J. of Geographical Info. Science 1-24.

163. [163] Andrea Baraldi, João V. B. Soares, "Multi-Objective Software Suite of Two-Dimensional Shape Descriptors for Object-Based Image Analysis," Subjects: Computer Vision and Pattern Recognition

(cs.CV), arXiv:1701.01941. Date: 8 Jan. 2017. [Online] Available: https://arxiv.org/ftp/arxiv/papers/1701/1701.01941.pdf

164. [164] Julesz, B. 1986. "Texton gradients: The texton theory revisited," in *Biomedical and Life Sciences Collection*, Springer, Berlin / Heidelberg, 54(4-5).

165. [165] Julesz, B., E. N. Gilbert, L. A. Shepp, and H. L. Frisch, 1973. "Inability of humans to discriminate between visual textures that agree in second-order statistics – revisited," *Perception* 2: 391-405.

166. [166] Yellott, J. (1993). Implications of triple correlation uniqueness for texture statistics and the Julesz conjecture. *Optical Society of America*, 10(5), 777-793.

167. [167] Zadeh, L.A. 1965. "Fuzzy sets," *Inform. Control* 8: 338–353.

168. [168] Open Geospatial Consortium (OGC) Inc. 2015. *OpenGIS® Implementation Standard for Geographic information - Simple feature access - Part 1: Common architecture*. Accessed 8 March 2015. http://www.opengeospatial.org/standards/is

169. [169] Shotton, J., Winn, J., Rother, C. and Criminisi, A. 2009. "Texton-Boost for Image Understanding: Multi-Class Object Recognition and Segmentation by Jointly Modeling Texture, Layout, and Context," Int. J. Comp. Vision. 81(1): 2–23.

170. [170] Wikipedia, 2018. Latent variable. Accessed 30 Jan. 2018. Available: https://en.wikipedia.org/wiki/Latent_variable.

171. [171] Gijsenij, A., Gevers, T., & van de Weijer, J. (2010). Computational color constancy: Survey and experiments. *IEEE Trans. Image Proc.*, 20(9), 2475-2489.

172. [172] Finlayson, G. D., Hordley, S. D., & Hubel, P. M. (2001). Color by correlation: A simple, unifying framework for color constancy. *IEEE Trans. Pattern Anal. Machine Intell.*, 23(11), 1209-1221.

173. [173] Gevers, T., Gijsenij, A., van de Weijer, J., & Geusebroek, J-M. (2012). *Color in Computer Vision.* Hoboken, New Jersey, USA: Wiley.

174. [174] Vo, A.-V., Truong-Hong, L., Laefer, D. F., Tiede, D., d'Oleire-Oltmanns, S., Baraldi, A., & Shimoni, M. (2016). Processing of extremely high resolution LiDAR and optical data: Outcome of the 2015 IEEE GRSS Data Fusion Contest. Part–B: 3D contest. *IEEE J. Selected Topics Applied Earth Obs. Remote Sens*. 9(12): 5560-5575.

175. [175] Baraldi, A., Tiede, D., & Lang, S. (2017b). Automated Linear-Time Detection and Quality Assessment of Superpixels in Uncalibrated True- or False-Color RGB Images. arXiv:1701.01940. Date: 8 Jan. 2017. [Online] Available: https://arxiv.org/ftp/arxiv/papers/1701/1701.01940.pdf

176. [176] Dillencourt, M. B., H. Samet, and M. Tamminen. 1992. "A general approach to connected component labeling for arbitrary image representations," *J. Association for Computing Machinery* 39: 253-280.

177. [177] Purdue University, 2018. Perceptual Grouping,. Available online: http://cs.iupui.edu/~tuceryan/research/ComputerVision/perceptual-grouping.html (accessed on 10 August 2018).

178. [178] Tiede, D., (2014). A new geospatial overlay method for the analysis and visualization of spatial change patterns using object-oriented data modeling concepts. Cartography and Geographic Information Science, 41(3), p.227–234. http://dx.doi.org/10.1080/15230406.2014.901900

179. [179] Trimble (2015). eCognition® Developer 9.0 Reference Book.

180. [180] Group on Earth Observation / Committee on Earth Observation Satellites (GEO-CEOS) - Working Group on Calibration and Validation (WGCV). 2015. *Land Product Validation* (LPV). Accessed March 20, 2015. http://lpvs.gsfc.nasa.gov/

181. [181] Slotnick, S. D., Thompson, W. L., & Kosslyn, S. M. (2005). Visual mental imagery induces retinotopically organized activation of early visual areas, *Cerebral Cortex*, 15, 1570-1583.

182. [182] Planet Labs, PLANET SURFACE REFLECTANCE PRODUCT, Feb. 2018. Available online: https://assets.planet.com/marketing/PDF/Planet_Surface_Reflectance_Technical_White_Paper.pdf (accessed on 11 July 2018).