Peer-reviewed version available at *Remote Sens.* **2020**, *12*, 495; doi:10.3390/rs12030495

Article

Knowledge and Data Driven Mapping of Environmental Status Indicators from Remote Sensing and VGI

Alessia Goffi [1,2], Gloria Bordogna [1,*], Daniela Stroppiana [1], Mirco Boschetti [1], Pietro Alessandro Brivio [1]

[1] IREA CNR; surname.name_initial@irea.cnr.it
[2] Terraria s.r.l.; a.goffi@terraria.it
* Correspondence: bordogna.g@irea.cnr.it; Tel.: +39 02 23699 299

Abstract: The paper proposes a human explainable artificial intelligence approach for mapping the status of environmental phenomena from multisource geo data. It is both knowledge and data driven: it exploits remote sensing expert's knowledge to define the contributing factors from which partial evidence of the environmental status can be computed. Furthermore, it aggregates the partial evidences to compute a map of the environmental status by adapting to a region of interest through a learning mechanism exploiting Volunteered Geographic Information (VGI), both from in situ observations and photointerpretation. The approach is capable to capture the specificities of local context as well as to cope with the subjectivity and incompleteness of expert's knowledge. The proposal is exemplified to map the status of standing water areas (i.e. water bodies and river, human driven or natural hazard flooding) by considering satellite data and geotagged observations. Results of the validation experiments were performed in three areas of Northern Italy, characterized by distinct ecosystems. Results of the proposed methodological framework showed better performances than traditional approaches based on single spectral indexes thresholding. The use of expert's knowledge, possibly imprecise/uncertain and incomplete, the need of few ground truth data for learning, and finally the explainability of learned rules are the distinguishing characteristics of the proposal with respect to traditional machine learning methods.

Keywords: soft constraints; Ordered Weighted Averaging Operators; Volunteered Geographic Information; standing water area mapping; decision attitude modeling

1. Introduction

In the age of big geo data we are faced with the new challenge of exploiting multisource information for designing new strategies aimed at a more effective and timely monitoring of the status of environmental phenomena such as prevention, preparedness and mitigation of natural disaster events.

Multisource big geo data can be obtained from authoritative data bases, for example territorial risk maps, from the analyses of remote sensing images, from volunteers creating geotagged observations either in situ, by means of apps installed on their mobile devices, or by using geographic information systems, i.e., Volunteered Geographic Information (VGI), and finally from the crowd interacting within social networks and exchanging information on territory observations.

Nowadays, ICT technologies are mature to manage and share big geo data on the Web by coping with huge volumes, variable creation rates, great variety and complexity of both data structures, formats, and semantics. As far as the storage of huge volumes of big geo data, the scale out architecture has established over the scale up architecture for its easiest expansibility at lower costs; the distributed processing paradigm, based on distributed file system and map-reduce [1], allows to process huge volumes of data efficiently; NoSQL databases have demonstrated to effectively manage

Peer-reviewed version available at Remote Sens. 2020, 12, 495; doi:10.3390/rs12030495

geospatial data with different data structure, both in the form of semi-structured information, and grid and raster data; finally, Web geo services of the Open Geospatial Consortium (OGC) provide standards interoperable means for sharing on the Web distributed and heterogeneous big geo data.

Nevertheless, potential stakeholders of big geo data, namely territorial administrators, might encounter the so called "data overloading situation" when analyzing multi-source geo big data to capture the status of environmental phenomena. Redundancy of geo data might carry inconsistent information, which may cause doubts on both data reliability and suitability for taking decisions to benefit territory management.

What is needed to solve such impasse, that is driving decision makers to a deadlock especially with respect to the exploitation of the huge data flow of remote sensing derived information, are flexible and mainly explicable approaches for big geo data synthesis capable to generate environmental status maps in near-real time.

In this paper, we propose a knowledge and data-driven explainable synthesis of Environmental Status Indicator (ESI) maps, obtained from the aggregation of remote sensing data and geotagged observations. The approach exploits remote sensing expert's knowledge, possibly ill-defined to derive partial evidence maps of the environmental phenomenon of interest, thus keeping humans in-the-loop. Further, it learns an aggregation function from geotagged observations, available in a region of interest (ROI), to compute the ESI map synthesizing the phenomenon. Moreover, the approach can linguistically represent the semantics of the learned aggregation function, that describes a decision attitude, thus achieving explainability. Finally, it is scalable, and suited for a distributing processing implementation framework.

The proposal is exemplified to map the status of standing water (i.e. water bodies and river, human driven or natural hazard flooding) in three regions of interest in Northern Italy, characterized by distinct environmental conditions. In this respect, partial evidence of standing water is computed from contributing factors, specifically derived from multiple spectral indexes (SI) [2,3,4,5,6]. The most suitable indices are suggested by remote sensing experts based on the literature, as indicators enhancing standing water areas presence and computed from remote sensing optical images acquired by Sentinel-2 MSI (Multispectral Instrument). The aggregation of partial evidence is performed through an Ordered Weighted Averaging (OWA) operator [7] generated by applying a machine learning approach exploiting limited ground truth, obtained by either in situ observations or photointerpretation from volunteers, i.e., VGI.

The paper is structured as follows. The next section recalls the materials and methods used to formalize the proposed approach. Section 3 describes the results, i.e., the motivation of our proposal its main characteristics. Section 4 introduces the discussion by a study case and reports results of experiments to detect standing water areas by exploiting geotagged observations. Conclusions comments the experimental results and summarize the main achievements and foreseen ongoing work.

2. Materials and Methods: the fuzzy notion

In the following subsections we define the basic concepts of fuzzy set theory that are the basic materials and methods used to model the process of Environmental Status Indicator mapping.

2.1 Soft constraints

Fuzzy sets were introduced by Zadeh in 1965 [8] to represent concepts characterized by unsharp boundaries, i.e., where the transition between membership and non-membership is gradual rather than abrupt. A fuzzy set A on a universe D is characterized by a membership function $\mu_A: D \rightarrow [0,1]$, assigning a membership degree, $\mu_A(d) \in [0,1]$, to each element d of the domain D. $\mu_A(d)$ provides an estimation of the belonging of d to A.

An elastic or soft constraint C on a domain D of a *variable* v, i.e., a *contributing factor*, is defined by a membership function of a fuzzy subset C of D. When we apply the soft constraint to a value $d \in D$ of the variable v its membership degree $\mu_C(d)$ indicates the degree of satisfaction of C: $\mu_C(d)=1$ means

Peer-reviewed version available at Remote Sens. 2020, 12, 495; doi:10.3390/rs12030495

that d fully satisfies C; $\mu_C(d)=0$ means that d does not satisfy C at all; $0<\mu_C(d)<1$ means that d partially satisfies C.

In the case of incomplete and imprecise expert's knowledge on a phenomenon, a soft constraint can be defined by the domain expert to specify a *criterion* to compute a *partial evidence* of the phenomenon, given the information on the value of a variable v, selected as a *contributing factor* of the phenomenon. In this case, the expert cannot state precisely which subset of values of the domain D of the contributing factor v provides evidence of the phenomenon, but can state imprecise/fuzzy subsets of D.

A simple definition of a soft constraint μ_C can be specified with a trapezoidal shape by a tuple (a,b,c,d,e,f), with $a,b,c,d \in [0,1]$ and $e,f>0$ as follows:

$$\mu_C(x) = \begin{cases} 0 & x < a \quad x > d \\ ((x-a)/(b-a))^e & a \le x < b \\ 1 & b \le x \le c \\ ((d-x)/(d-c))^f & c < x \le d \end{cases} \quad (1)$$

By setting $a = b = -\infty$ or $c = d = +\infty$ we obtain the special cases of L-functions (not increasing) and R-Functions (not decreasing) as the one depicted in Figure 1.

Complex soft constraints can be defined by combining soft constraints either by conjunction ("C1 and C2" is defined by $\min(\mu_{C1}(x), \mu_{C2}(y))$ $\forall x \in X$ and $\forall y \in Y$), by disjunction ("C1 or C2" is defined by $\max(\mu_{C1}(x), \mu_{C2}(y))$, $\forall x \in X$ and $\forall y \in Y$) and by negation ("Not C" is defined by the complement $1-\mu_C(x)$).

Finally, when $\mu_{C1}(x) \subseteq \mu_{C2}(x)$, $\forall x \in X$, C1 is included in C2, i.e. C1 is stricter than C2. When defining a soft constraint to compute the partial evidence degree of a critical phenomenon in need of prompt identification, the stricter the soft constraint, the more we may miss the phenomenon, i.e., we have a risky attitude, and tolerate false negatives. Conversely, by defining a relaxed soft constraint we have a precautionary attitude, but may set false alarms of the phenomenon, i.e., we tolerate false positives.

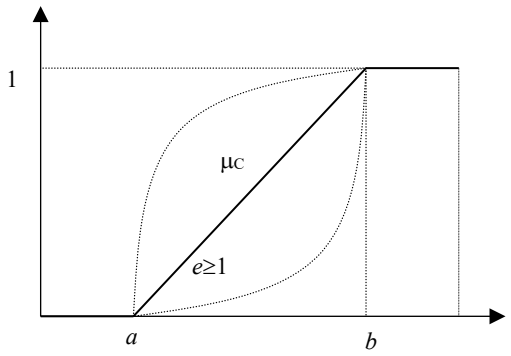


Figure 1: R-function defined by (a<b<1, c=d=+ ∞, e=1) representing the semantics of a soft constraint.

2.2 Ordered Weighted Averaging (OWA) operators

The seminal paper [9], stemming from the consideration that "*the efficient use of decision support systems (DSSs) is to assist and help humans arrive at a proper decision, but by no means, to replace humans*" proposes to introduce some synergy between the human and machine. To this end, the author defines the fuzzy logic-based calculi of linguistically quantified propositions as a viable means for expressing human interpretable decisions.

Linguistic quantifiers were first introduced in [10] as fuzzy subsets of the positive real numbers or of the unit interval [0,1] according to the fact that they express an absolute quantity, such as *many*, or a relative quantity, such as *most*.

In [11] the problem to define an *overall decision function* aggregating degrees of satisfaction of multiple criteria (in our context, partial evidence degrees computed by soft constraints defined on the domain of some variables), was proposed based on the Ordered Weighted Averaging (OWA)

Peer-reviewed version available at Remote Sens. 2020, 12, 495; doi:10.3390/rs12030495

operators. An OWA of dimension N and weighting vector W, with $\sum_{i=1,...N} w_i=1$, aggregates N values $[d_1,\ldots,d_N]$, and computes an aggregated value a in [0,1] as follows [11, 12]:

$$\text{OWA: } [0,1]^N \rightarrow [0,1]$$

such that

$$a = \text{OWA}([d_1, \ldots,d_N]) = \sum_{i=1,...N} w_i * g_i \quad (2)$$

in which g_i is the ith largest value of the d_1, d_N.

A fundamental aspect of the OWA is the reordering of its arguments so that the weight w_i is not associated with an argument d_i but rather with a particular rank of the arguments in decreasing order. A known property of the OWA operators is that they include the Max, Min and arithmetic mean operators for the appropriate selection of the weighting vector W:

For W=[1, 0, ...,0], OWA($[d_1, \ldots,d_N]$)= max($[d_1,\ldots, d_N]$)

For W=[0, ...,0,1], OWA($[d_1, \ldots,d_N]$)= min($[d_1,\ldots, d_N]$)

For W=[1/N,....,1/N], OWA($[d_1, \ldots,d_N]$)= $\frac{1}{N}\sum_{j=1}^{N} d_j$

It can be proved that OWA operators satisfy the commutativity, monotonicity and idempotency and are bounded by Min and Max operators [1]:

Min_($[d_1, \ldots,d_N]$) ≤ OWA($[d_1, \ldots,d_N]$) ≤ Max($[d_1, \ldots,d_N]$)

2.3 Characterizing the OWA semantics

To characterize the decision attitude modelled by an OWA operator with weighting vector W two measures have been introduced in [12]: orness and dispersion.

The *orness*(W)∈[0,1] measure is defined as follows:

$$orness(W) = \frac{1}{N-1}\left(\sum_{j=1}^{N}(N-j) * w_j\right) \quad (3)$$

This measure characterizes the degree to which the aggregation is like an *OR* (Max) operator. It can be shown that, when the argument values $d_1, \ldots,d_N$ are degrees of partial evidence of an anomaly of an environmental phenomenon from N distinct sources, i.e., the greater they are the more severe the anomaly, we have the following interpretations [13, 14]:

- *orness*[*1,…,0*] = 1 indicates a **pessimistic attitude adversing risks** (i.e., nothing is disregarded, any single source alone is trusted and taken into consideration to plan preparedness and mitigation interventions so as to minimize the occurrence of risky events);
- *orness*[*0,…,1*] = 0 indicates an **optimistic attitude towards tolerating risks** (i.e., prioritizing preparedness and mitigation interventions only to anomaly situations pointed out by all sources);
- *orness*[*1/N,…,1/N*] = 0.5 indicates a balanced and **neutral attitude** towards risk-prone and risk-adverse.

Another measure used to qualify the semantics of an OWA operator depending on the form of the weighting vector is the dispersion. This measure represents how much of the information in all the arguments is used by an OWA with weighting vector W. The idea behind its definition is that the greater the dispersion the more democratic is the aggregation of the correspondent OWA since it uses information from more sources. Several dispersion measures have been proposed, the first of which is based on the concept of entropy of W. We adopted the *dispersion(W)* ∈ [0,1] measure of an OWA operator as proposed in [15]:

$$dispersion(W) = 1 - Max_{i=1,...N} w_i \quad (4)$$

Peer-reviewed version available at Remote Sens. 2020, 12, 495; doi:10.3390/rs12030495

We see that *dispersion(W)* is clearly symmetric and when N is large it is defined in [0,1). When it is zero it means that only one source is considered, i.e., we have the less democratic or dictatorial aggregation. The larger its value, the more the result is determined by additional sources, and thus we have a more democratic, i.e. less dictatorial aggregation.

In order to explicit the semantics modeled by an OWA aggregation with weighting vector W one computes the degrees of *orness*(W) and *dispersion*(W) as defined in formulae (3) and (4) respectively, and then compares these values with the conditions in Table 1 to select the correspondent decision attitude. Notice that in Table 1, the value of an argument to aggregate is a degree of evidence of a critical phenomenon; then, a high value is considered a pessimistic evaluation of what is occurring, while a low value is an optimistic evaluation. The rationale is that since the values to aggregate are evidence degrees of some undesired phenomenon, such as flood occurrence, wild fire occurrence, etc., high values/low values have a negative/positive flavor. Thus, the interpretation of optimism and pessimism reported in Table 1 are complemented with respect to the context of multi-criteria decision making (pessimistic becomes optimistic, and viceversa).

Table 1: Decision attitude as a function of *orness* and *dispersion* in the case of aggregation of N=8 partial Evidence degrees of critical/anomalous event/phenomenon.

		Δ Dispersion (W)				
N = 8		0	> Δ >	0.44	> Φ >	**0.88**
Φ Orness (W)	0	Dictatorial & Optimistic				
	> Φ >	Dictatorial & Towards Optimistic	Semi Dictatorial & Towards Optimistic	Semi Dictatorial/ Democratic & Towards Optimistic	Semi Democratic & Towards Optimistic	Democratic & Towards Optimistic
	0.5	Dictatorial & Neutral	Semi Dictatorial & Neutral	Semi Dictatorial/D emocratic & Neutral	Semi Democratic & Neutral	Democratic & Neutral
	> Φ >	Dictatorial & Towards Pessimistic	Semi Dictatorial & Towards Pessimistic	Semi Dictatorial/ Democratic & Towards Pessimistic	Semi Democratic & Towards Pessimistic	Democratic & Towards Pessimistic
	1	Dictatorial & Pessimistic				

When *orness*(W) > 0.5 and *dispersion*(W) is close to 0, the decision is risk adverse since one mostly trusts the most pessimistic/towards pessimistic sources and almost disregards the optimistic ones. Nevertheless, in doing this, one may obtain many False Positives.

When *orness*(W) < 0.5 and *dispersion*(W) is close to 0, the decision attitude is risk prone since one mostly trusts the few sources that are optimistic. In this case one may miss potential alerting sources, and may thus generate many False Negatives.

A balanced decision attitude, characterized by *orness*(W)=0,5 and *dispersion*(W) = (N-1)/ N , takes into account equally all sources, then is both neutral and democratic. Intermediate values of *orness* and *dispersion* characterize different blends of both pessimism/optimism and democracy/dictatorship.

Peer-reviewed version available at Remote Sens. 2020, 12, 495; doi:10.3390/rs12030495

2.4 Learning OWA semantics from observations

One important issue in the domain of partial evidence aggregation is the determination of the OWA operator modelling the aggregation, since generally experts do not have clear idea of the decision attitude they should apply. If reference data are available, i.e. in a specific domain ground truth on occurrence of a phenomenon at certain locations of the region of interest (ROI), these can be used to learn the weighting vector of the OWA operator.

Such information can be available from in situ observations generated by the use of mobile applications and used by volunteers to send information on the status of a given phenomenon. An example is Space4Agri project (http://space4agri.irea.cnr.it/it/progetto/struttura/ambito-in-situ-1), in which agronomists were asked to tag crop fields with the observed crop and stage of growth and, eventually, they could indicate if rice paddies were inundated or not [16]. Another possible source of ground truth on the occurrence of critical situations can be extracted from social media posts by event detection techniques [17]. Finally, ground truth can be generated by volunteer experts by photointerpretation like in the Humanitarian OpenStreetMap project [18].

To this end we propose the application of a machine learning approach [19] exploiting VGI assumed as ground truth to learn the best OWA operator for a given ROI, by iteratively minimizing error between OWA results at cycle *t* with respect to the observations described by VGI.

Given K georeferenced observations a_1, … a_K, for example VGI elements, by knowing their geographic coordinates we can associate with each observation the partial evidence values [a_{i1}, …a_{iN}] having the same coordinates such that we obtain the following antecedent-consequent rules that must be satisfied:

$$\begin{aligned} &a_{11}, ..., a_{N1} \rightarrow a_1 \\ &... \\ &a_{1K}, ..., a_{NK} \rightarrow a_K \end{aligned} \tag{5}$$

In principle the observations a_1, … a_K, can be specified on a continuous scale [0,1], to quantify the extent of the phenomenon in the specific location; nevertheless, in practical situations a discrete scale such as {0, 0.5, 1}, or even a binary scale {0, 1} is used where 0 means absence of the phenomenon and 1 presence.

The learning mechanism starts at cycle L=0 by assuming as initial OWA operator the weighted average (balanced and neutral attitude), that is defined with weighting vector W0= [1/N, …1/N]. Then it iteratively tries to minimize a certain error until either convergence or a predefined maximum number of cycles is reached: at each iteration cycle L, it determines the weighting vector $W_L = [w_{1L}, …, w_{NL}]$ of OWA_L that minimizes the error existing between the results of its application to all the antecedents of the rules in (5). Specifically, for a number of cycles L=0, …, Lmax Minimize($OWA_L(a_{1K}, ..., a_{NK}) - a_K)^2/2$) ∀k=1,…K, which means repeating the minimization until L= Lmax or until the error do decrease significantly (by a small value ε) with respect to the preceding cycle.

Formally, this is equivalent to applying the following rule:

$$\text{select } W_L \text{ such that } |\Lambda_i(L) - \Lambda_i(L+1)| < \varepsilon \approx 0 \quad \text{or } L = L_{max} \tag{6}$$

where

$$\Lambda_i(L+1) = \Lambda_i(L) - \beta\, w_{iL}\,(argmax_i(a_{1k}, ..., a_{Nk}) - OWA_L(a_{1k}, ..., a_{Nk}))*(OWA_L(a_{1k}, ...,a_{Nk}) - a_k) \tag{7}$$

in which $\beta \in (0,1]$ is a learning rate parameter and the i*th* weighting vector element at cycle L is defined as follows:

$$w_{i\,L} = e^{\Lambda i(L)} / \sum_{j=1,...N} e^{\Lambda j(L)} \quad \forall i=1,...,N \tag{8}$$

Peer-reviewed version available at Remote Sens. 2020, 12, 495; doi:10.3390/rs12030495

3. : Results: motivations of the proposal and description of main phases and characteristics

3.1. Rationale for the Human Explainable approach

Environmental monitoring based on Earth Observation (EO) data consists in assessing the status of the environment at timestamps, and in comparing its changes in time by detecting possible anomalies in the dynamic evolution with respect to the normality. The objective is providing decision-makers with synthetic information, for example environmental status indicator (ESI) maps, to help them understanding ongoing critical conditions.

Current most up-to-date approaches for geo big data dimensionality reduction are data-driven: they apply machine learning techniques as black boxes, namely deep and convolutional neural networks [20] for scene classification purposes [21, 22, 23]. They require a preliminary training phase in which, given a set of ground truth observations, they learn the classifier which is subsequently applied on the entire ROI. Although these approaches demonstrated to be very successful in several contexts, they are opaque mechanisms, which do not explicit the classification rules. Moreover, in order to train properly the algorithms, training data sets must be large enough and representative in order to avoid overfitting [24]. Nevertheless, in many real cases of EO data applications over large areas, representative and spatially distributed date sets for training are not available. Finally, when changing the ROI one generally needs to repeat the training phase with new ground truth data, since transfer of a pre-trained network greatly depends on the choice of a proper network architecture for the target purpose [25]. In order to overcome limitations imposed by the training phase of the algorithm, expert's knowledge based approaches are widely used.

Knowledge-driven approaches for environmental status assessment rely on expert's knowledge, possibly ill-defined, about the physical interaction of the electromagnetic radiation remotely sensed with the characteristics of the analyzed target, that are hints of the undergoing environmental phenomenon of interest, to define mapping rules. Widely used approaches are based on Spectral Indexes (SI) which are defined on a real domain to integrate reflectance measurements at different wavelengths into a synthetic feature that can highlight some aspects of the environmental status: by applying a function combining the band signals, SI maps can be generated and then segmented to identify vegetation presence and vigor (biomass presence, Leaf Area Index, Chlorophyll content, etc), bare soil condition and soil properties composition, burned area or water presence, and so on [26,27,28,29,30]. Although this approach has the advantage of being human explainable, it is often ineffective to describe complex phenomena, such as delineation of flooded areas, for several reasons: many environmental phenomena have a different appearance when changing the geographic context and observation conditions (presence of clouds, shadows, specific land covers, density/fractional cover of the target, etc.); thus a single SI may be not sufficient to capture all aspects of a given phenomenon. For example, to identify water surfaces, composed by different types of standing water targets such as shallow water, deep water, wetlands, river and inland water bodies, rice flooded fields, many distinct SIs have been defined [2, 3, 4, 5, 6, 31]. Using distinct SIs to map a given phenomenon may result in redundant or conflicting maps. Furthermore, not all SIs are defined on the same domain, so one needs to normalize their values in the same domain to compare them. An accurate calibration phase is required for determining the proper threshold on the SI values that allows segmenting the phenomenon footprint, i.e., the spatial extent of the phenomenon with an acceptable accuracy in each ROI. This calibration is significantly dependent on local environmental conditions, and often one must engage with several trial and error phases to find the best threshold(s) that minimizes commission and omission errors. Besides the SIs, other contributing factors may constrain and influence the environmental phenomenon under study. For example, floods generally occur in mostly flat regions while cannot occur in areas with steep slope. Finally, knowledge-based approaches lack automatic adaptation mechanisms to ROIs exploiting machine learning and available observations.

In the following paragraph we describe our proposal for a human explainable approach to map ESIs from EO data which is both knowledge-driven, flexibly exploiting imprecise and uncertain domain knowledge to define approximate mapping rules, and data-driven from multi-sources,

Peer-reviewed version available at Remote Sens. 2020, 12, 495; doi:10.3390/rs12030495

namely volunteers, to adapt the expert's rules aggregation to a specific ROI. The overall objective is to overcome both the drawbacks of black-box approaches and the limitation of knowledge-driven approaches discussed above.

3.2. The knowledge & data driven adaptive approach

Expert's knowledge assessment of an environmental phenomenon is often performed evaluating multiple criteria which contribute to distinct extent to determine or influence its occurrence: these criteria are hereafter named contributing factors. To compute ESI maps describing the spatial evidence of a studied phenomenon, we proposed a fuzzy approach based on the fusion of multiple partial evidence derived by the analysis of several contributing factors, that can concur or complement one another [32]. This proposed approach stems from the way in which synthetic maps are created by means of a traditional Geographic Information System (GIS): generally, first several layers of thematic information are loaded into the GIS; then, from each layer, constraints are defined to perform selections of features, and, finally, all features are aggregated in a synthetic map by applying a Boolean operator, namely intersection or union. This approach suffers for the defects outlined in the previous subsection derived by the rigidness of both the constraints and the aggregation operators which are crisp and thus admit only Boolean satisfaction degrees.

To overcome such deficiencies, we generalized this traditional GIS operative modality by proposing an approach that allows the propagation of imprecision to the end of the process by allowing the specification of soft constraints, i.e., soft selection conditions admitting degrees of satisfaction, and fuzzy aggregation operators with behaviour that can be flexibly tuned in between that of the intersection and union. The original proposal [32] applied for many distinct purposes [33,34,35,36] heavily relies on the imprecise and incomplete expert's knowledge to identify the thematic maps, i.e., the contributing factors, and to define both the soft constraints and the aggregation operator to generate the ESI map.

A preliminary phase of the proposed approach is the selection of contributing factors that can influence the phenomenon: these contributing factors are physical variables whose values are computed in all spatial units of a ROI to create a thematic maps. They are identified by experts based on domain knowledge; a statistic analysis of the values of the contributing factors is performed on a classified data set in order to define soft constraints which better discriminate the class of interest from the others; soft constraints satisfaction degrees are interpreted as degrees of partial evidence of the phenomenon due to a specific contributing factor. In this phase, also an importance for each factor can be computed proportional to the degree of separability between the classes achieved by applying the soft constraints on the classified data set. Alternatively a degree of reliability or trust can be deemed for each factor depending on the knowledge of the phenomenon or reliability of the data source.

Notice that this preliminary phase does not need to be performed for each ROI, each time the algorithm is applied on new data. It can be done once and for all, while leaving to the automatic algorithm the task of adapting the mapping to the specific ROI by exploiting local ground truth.

The automatic algorithm depicted in Figure 2 is structured into two phases. In the first phase, automatic mapping starts by evaluating the input soft constraints for each contributing factor on the input map. This phase produces partial evidence (PE) maps in which each unit element, a pixel in the illustrated implementation, is associated with a degree in [0,1].

In the second phase, in order to generate the synthetic ESI map, PE maps are aggregated by applying a fuzzy approach, namely an OWA operator. OWA operators allow to define fusion strategies with distinct mean like semantics ranging from the minimum to the maximum of the values they aggregate. The mean like nature of fusion strategies has been outlined by many authors and is recognized as particularly useful in the context of spatial decision making [13, 37, 38, 39, 40, 41, 42].

A weakness of our original proposal [32] was its strict knowledge-driven characteristics requiring experts to define both the soft constraints for deriving the partial evidence maps, and the aggregation operator to fuse them. While for soft constraints definition they can rely on the thresholds on the distinct SIs defined by the literature, to find the best aggregation operator they generally apply

Peer-reviewed version available at Remote Sens. 2020, 12, 495; doi:10.3390/rs12030495

the most common fuzzy operators, that are the min, i.e., the AND, the max, i.e., the OR, and the average on a set of points for which ground truth is available. Then they select the aggregation operator that provides the best accuracy to map the whole ROI. Clearly, with a few trials one cannot be sure to have identified the best classifier.

In the present paper we propose the evolution of this method [32] as depicted in Figure 2, in which the first phase in mainly knowledge-driven, while the second phase incorporates a data-driven algorithm exploiting VGI to learn the best OWA operator to be used for the aggregation of the partial evidence maps.

The remarkable aspects of this data-driven phase is that the algorithm does not require a huge amount of VGI to converge, and the semantics of the learned aggregation operator is human explicable. Indeed, in many real cases some VGI is created in situ by the use of smart applications by volunteers to highlight critical situations, or in citizen science projects such as Humanitarian OSM [18] it is generated by photointerpretation.

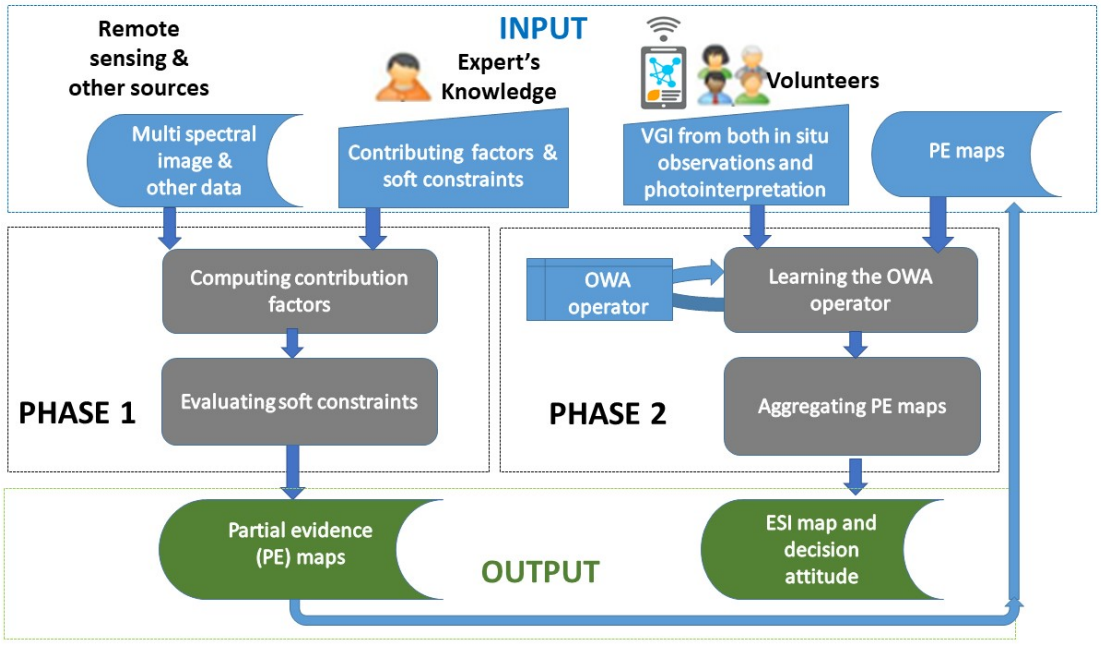


Figure 2: Workflow of the proposed fuzzy adaptive approach for computing Environmental Status Indicator (ESI) maps from remote sensing multispectral images, thematic information and VGI. While phase 1 exploits expert's knowledge, phase 2 is data driven exploiting VGI. The two phases are decoupled and communicate via input layer.

The ESI computed by the proposed approach has the following characteristics:

- the ESI map can be computed at spatial unit level, i.e., either a pixel or a larger spatial unit: this means that for each individual pixel (spatial unit) a single ESI value is computed;
- the ESI map can be tuned to a local context and observation conditions;
- ESI value is defined in [0,1];
- finally, the ESI computation is feasible in a distributed processing framework on big geo data in order to achieve scalability.

The algorithm computing the ESI can model distinct needs by inheriting the properties of OWA operators [13]: the distinct credit of the contributing factors and reliability of their sources; the possibility to take into account the local agreement of contributing factors to determine their influence on the result; the possibility to model distinct attitudes towards risks in between the two extreme optimistic and pessimistic; the possibility to explain the decision attitude represented by the learned aggregation operator.

Peer-reviewed version available at Remote Sens. 2020, 12, 495; doi:10.3390/rs12030495

Phase 2 of the algorithm explores the space of OWA operators in order to identify the best ESI map for each given ROI in accordance with the available VGI. Such a flexibility enables the possibility of adapting the learning to a ROI based on a given spatial stratification: for example, in homogeneous area with respect to given land covers, the OWA operator can be learned independently from other areas. The identified OWA operator can then be applied to spatial units in the stratified ROI.

Moreover, learning the aggregation operator can make results less sensible to slight different definitions of both the contributing factors and the soft constraints. In facts, when defining distinct contributing factors and soft constraints the first phase computes different partial evidence (PE) maps; thus the process learns the best OWA operator given the current evidence maps for each ROI in which VGI is available. This allows copying with distinct subjectivity of expert's knowledge.

By discussing a real application of standing water mapping we show that the proposed approach can adapt to distinct ill-defined knowledge of the experts and to distinct local contexts achieving good accuracy. Finally, we discuss how the knowledge acquired by the learning can be interpreted to refine the knowledge of the contributing factors interactions.

3.3. Scalability of the approach

The ESI computation described in the previous section can be implemented in a distributed processing framework represented by the schema depicted in Figure 3.

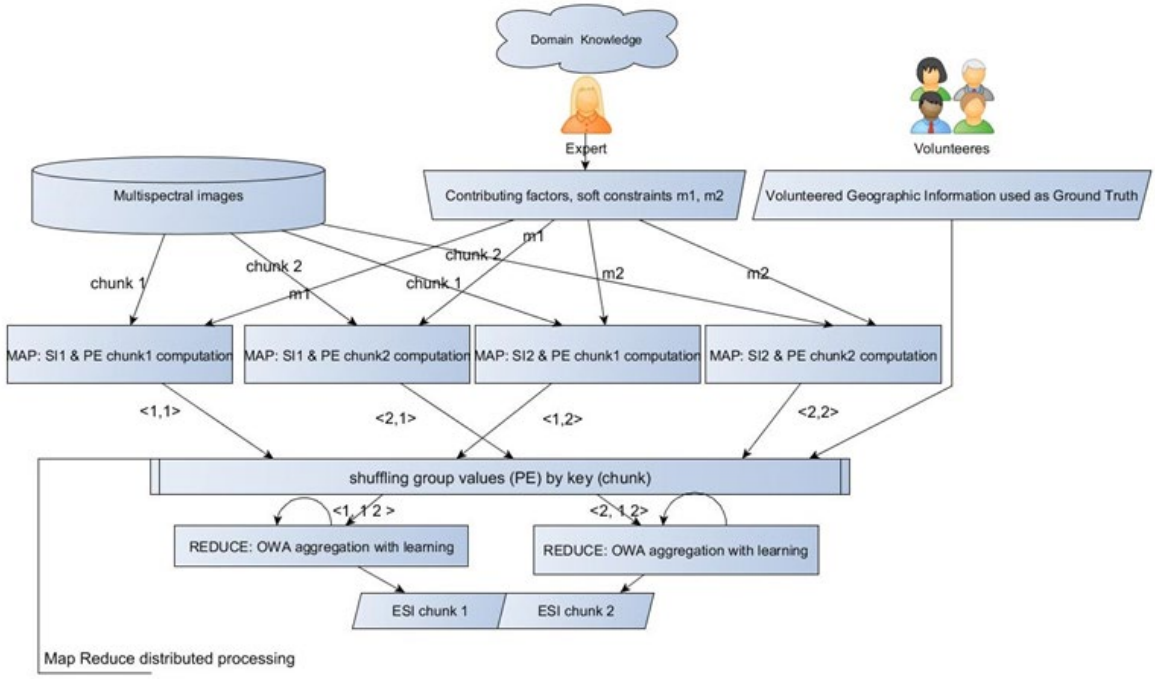


Figure 3: Map-reduce distributed process computation of the ESI map.

Since the ESI computation is performed independently for each spatial unit, and is organized in two subsequent phases, we can implement it in a single round of a Map-Reduce framework [1].

The Map-Reduce framework is inspired by the "Map" and "Reduce" functions used in functional programming. Computational processing occurs on data stored in a distributed file system or within a database, which takes a set of input key-values pairs and produces a set of output key-values pairs [43].

A *mapper* M is a Turing machine M (<k,v>) → (<k1′,v1′>, …, <ks′,vs′>) which accepts as input a single key-value pair <k,v> and produces a list of key-value pairs <k1′,v1′>, …, <ks′,vs′>.

A shuffle is performed on the outputs of the mappers so as to group the values with the same key: <k1′,v1′, …, vr1′ >,…., <kR′,v1′, …, vrR′ >.

Peer-reviewed version available at Remote Sens. 2020, 12, 495; doi:10.3390/rs12030495

A *reducer* R is a Turing machine R: <k′ , v1′, …, vr′ > → <k′,v″> which accepts as input a pair <k′ , v1′, …, vr′ > and produces as output the same key k′ and a new value v″.

It is well known that a bottleneck of Map-Reduce distributed processing is the number of rounds needed to implement an algorithm. Nevertheless, for implementing the ESI computation, we do not need multiple rounds. Simply distinct mappers can be instructed to execute the first phase for generating distinct partial evidence maps by processing in parallel the given image chunks based on expert's knowledge provided in input as both contributing factors and soft constraints definitions. A mapper can be instructed by its input parameters to compute more contributing factors and to evaluate more soft constraints on the same chunk: the input key k identifies either a single pixel or a spatial unit in a multispectral image chunk, while the associated value v is the information associated with the input chunk (e.g. the bands and theme values such as VGI), plus parameters (the contributing factors names ans definitions the mapper has to compute) and the tuples (a,b,c,d,e,f) defining the soft constraints membership functions according to definition (1).

A mapper can compute for each pixel in the input chunk the key-value pairs <k1′,v1′>, …, <ks′,vs′> where ki′ identifies the chunk and vi′ are the computed degrees of partial evidence of the SIs in the chunk.

Successively, the reducers execute the second phase by aggregating the partial evidence maps v1′, …, vr1′ of the same chunk ki′ in parallel so as to compute the ESI map v″ for the chunk.

Chunks are finally recombined by mosaic at the end of the process.

The values v″ is computed by applying in each pixel or spatial unit of the chunk the OWA operator learned by leveraging VGI in the ROI covered by the chunk. This way, each reducer can learn a distinct OWA operator, thus adapting the ESI computation to the local context and observation conditions. Notice that the learning process is performed within each reducer module which applies on its input chunk the OWA operator with the weighting vector learned at time cycle L based on the subset of VGI included in the input chunk. There is no need to upload the input at each cycle, since the evidence maps do not change from cycle to cycle; once the optimal OWA has been determined ESI map can be computed and stored on disk.

Map reduce was designed for processing massive data sets, so programs require that every reducer only has access to a portion of the input, and the strict modularization prohibits reducers from communicating within a round. These conditions are satisfied by our proposed algorithm which does not need any communication to occur among mappers and among reducers during a round.

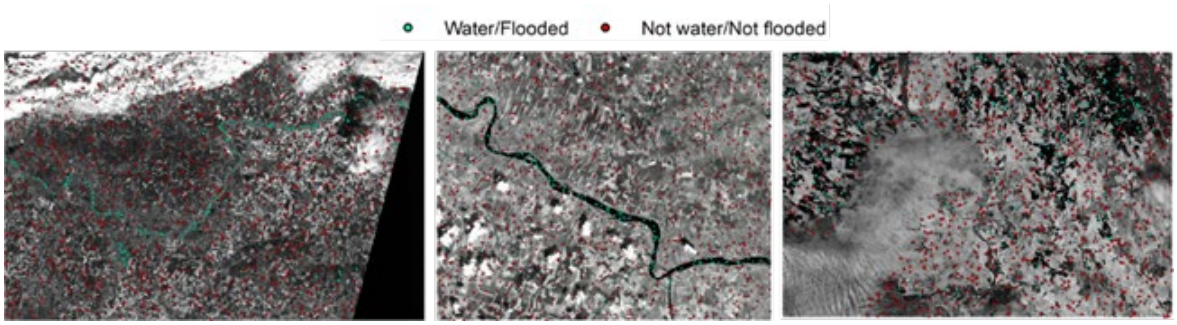
4. Discussion: Case Study

In this section a case study illustrating the application of the proposed approach for defining and mapping an ESI for standing water presence (e.g. inland water bodies, flooded area by human activities or natural hazard) is presented and described.

4.1. Study area, Data Sources, and Data

The case study is relative to the monitoring of a territory in Northern Italy with respect to the mapping of standing water, which can occur due to controlled inundations (irrigation), extreme event floods and natural water reservoirs. Specifically, the three sites in Figure 4 were selected as ROIs in Northern Italy where VGI obtained by both in situ observation and photointerpretation classifying water and no water were available. This VGI was assumed as ground truth for three distinct objectives: i) the definition of soft constraints, ii) learning the OWA aggregation and iii) validation of algorithm performance. Sites were selected to cover different conditions of standing water in order to capture variable spectral characteristics: flooded area due to extreme heavy rainfall (Cal/Val_1), river bed (Cal/Val_2) and flooded rice fields (Cal/Val_3) (Table 2). The latter site was selected, although flooding was not due to a natural event, to train and validate the algorithm over heterogeneous conditions of shallow water surface (< 50 cm) mixed with soil patches and vegetation (emerging rice plants); besides, for this site a large dataset of in situ observations was available from fields surveys.

Peer-reviewed version available at Remote Sens. 2020, 12, 495; doi:10.3390/rs12030495



Cal/Val_1 Cal/Val_2 Cal/Val_3

Figure 4: Study sites in Northern Italy with VGI assumed as ground truth points (blue water – red non water) used for ii) learning OWA aggregation and iii) validation of the algorithm.

Table 2: Location/extent of the study sites and characteristics/conditions of the surface water areas.

Site Name	Location (Italy)	Latitude (North)	Longitude (West)	Surface water conditions	Dimension (km^2)
Cal/Val_1	Emilia (IT)	44.968861	10.649674	Flooded areas due to severe rainfall events	2090
Cal/Val_2	Po Valley (IT)	44.992491	11.377019	River in standard conditions	546
Cal/Val_3	Rice paddies (IT)	45.278927	8.527552	Flooded rice fields	1937

Table 3 reports for each site EO satellite data and acquisition dates, the number of ground truth pixels used for soft constraints definition in the preliminary phase, for learning the OWA operator in phase two of the algorithm, and for validation of the computed ESI maps.

Table 3: Number of pixels (w/nw stand for water/not water) for Cal/Val sites used for soft constraints definition (S), OWA learning (L) and 10-fold cross typical/atypical validations (V). At each validation cycle, 10% (90%) of the ground truth pixels not used for (S) were randomly selected for (L) and the remaining 90% (10%) were used for (V) in the typical (atypical) validation settings, respectively.

Name	Area	Dates	# ground truth pixels for (S)	# ground truth pixels (w/nw) for (L)	# ground truth pixels (w/nw) for (V)
Cal/Val_1	Emilia (site 1)	S2A 13/12/2017	144689	(16/71) tot=87	(141/638) tot=779
Cal/Val_2	Po valley (site 2)	S2A 14/05/2017	51014	(19/94) tot=113	(173/845) tot=1018
Cal/Val_3	Rice paddies (site 3)	S2A 22/04/2016	42015	(17/85) tot=102	(153/768) tot=921

Peer-reviewed version available at Remote Sens. 2020, 12, 495; doi:10.3390/rs12030495

Specifically, the remote sensing data source used in all sites is Sentinel 2 (S2) of the Copernicus Earth Observation program led by the European Commission and operated by the European Space Agency (https://earth.esa.int/web/sentinel/home, accessed November 2019). S2 mission operates as part of a two-satellite system (A&B) providing high resolution multispectral optical imagery since June 2015 (A) and March 2017 (B). MSI measures the Earth's reflected radiance in 13 spectral bands from VIS/NIR to SWIR with a spatial resolution ranging from 10 m to 60 m. The study case was built on S2 data collected for post-event assessment (after flooding occurrence Cal/Val_1 and Cal/Val_2 and after the rice field survey for Cal/Val_3). Level-2A S2 images were downloaded and pre-processed with sen2r toolbox [40]. The details of the preprocessing operations are described in [36]. For Cal/Val_1, Cal/Val_2, Level-2A S2 imagery was downloaded as Bottom of Atmosphere (BOA) reflectance through the Copernicus Open Access Hub, and pre-processing consisted in clipping images to our area of interest and masking clouds using Scene Classification (SC) product: pixels classified as high and medium cloud probability were masked out, while pixels belonging to different classes were retained to avoid masking out water pixels. For Cal/Val_3, BOA image was not available at the desired dates of the field survey in the Copernicus archive, so it was necessary to download the Top of Atmosphere Level-1C products and apply atmospheric correction by using Sen2Cor algorithm of the sen2r toolbox library [44].

4.2.Contributing factors identification and soft constraints definition based on Expert's Knowledge

In order to exploit the huge literature based on single spectral index to map flooded areas, standing water, and vegetation cover from remote sensing, seven spectral indexes have been selected as contributing factors from which partial evidence of standing water can be computed (see Table 4). Besides spectral indexes, also Hue (H) and Value (V) dimensions of the HSV color space, derived by transforming the components SWIR2, NIR, RED, were selected to define the reduced space Hue-Value (HV) as further contributing factor; in this transformed space, standing water surfaces can be separated from land surfaces by means of empirical thresholds as defined in [6].

In the definitions of the formulae in Table 4, parameters are defined as follows: C1=4, C2 = 0.25, C3 = 2.75, D1=2.5, D2 = 1.5, D3=0.25, L=0.5 and S2 MSI bands are BLUE=band2 (490 nm), GREEN=band3 (560 nm), RED=band4 (665 nm), NIR=band8 (842 nm), SWIR1=band11 (1610 nm), SWIR2=band12 (2190 nm).

The transformation function f: SWIR2 × NIR × RED → H × V is a Standardized colorimetric transformation from RGB to HV components of the HSV color space, where SWIR2=R, NIR=G and RED=B respectively, defined as in [45]:

$$f(R,G,B)\begin{cases} V = \max(R,G,B) \\ H = \begin{cases} \left(60° * \frac{G-B}{V-\min(R,G,B)} + 360°\right) mod\ 360° & if\ V = R \\ \left(60° * \frac{B-R}{V-\min(R,G,B)} + 120°\right) & if\ V = G \\ \left(60° * \frac{R-G}{V-\min(R,G,B)} + 240°\right) & if\ V = B \end{cases} \end{cases} \quad (11)$$

For each contributing factor/spectral index, a soft constraint is defined on its domain by the expert by analysing statistical distribution of each SI value for the pixels corresponding to standing water with respect to the ones of non-water surfaces according to the ground truth data in the three Cal/Val sites. The soft constraints are defined with a trapezoidal shape, basically L and R functions, as defined in formula (1). In the case of the contributing factor HV a single bi-dimensional soft constraint on the domain H×V has been defined as a fuzzy relation combing by minimum the soft constraints on the two dimensions. The details of this activity preliminary to the execution of the algorithm phase 1 are reported in [36].

In order to set up a validation experiment aimed at testing the stability of the approach when changing expert's knowledge, we performed phase 1 twice, by exploiting two knowledge bases by

Peer-reviewed version available at Remote Sens. 2020, 12, 495; doi:10.3390/rs12030495

two experts (A and B), hereafter named KB_A and KB_B, respectively. They defined different soft constraints on the same set of contributing factors with distinct decision attitudes by interpreting available data as illustrated in Figure 5. It can be noticed that the soft constraints of expert A are generally stricter than those defined by expert B on the same SI, i.e., the membership functions defined by A are generally included in those of expert B. It follows that Expert A (Figure 5) has a more optimistic attitude towards mapping standing water areas (considered as an undesired phenomenon) ; he/she accepts the risk of generating omission errors by partially disregarding "shadows over water areas". Conversely, Expert B (Figure 5) takes a more pessimistic, in this context, precautionary, attitude by defining soft constraints so as not to miss "shadows over water areas" which belong to the support of the membership functions (i.e. have not null membership degree).

Table 4: Selected Contributing factors.

Contributing factors	Formula	Category	Reference
AWEI	C1 * (GREEN − SWIR1) − (C2 * NIR + C3 * SWIR2)	Water SI	[3]
AWEIsh	BLUE + D1 * GREEN – D2 * (NIR + SWIR1) – D3 * SWIR2	Water SI	[3]
mNDWI	(GREEN – SWIR1) / (GREEN + SWIR1)	Water SI	[31]
NDWI	(GREEN – NIR) / (GREEN + NIR)	Water SI	[5]
NDFI	(RED – SWIR2) / (RED + SWIR2)	Flooding SI	[2]
SAVI	(1 + L) * (NIR – RED) / (NIR + RED + L)	Vegetation SI	[4]
WRI	(GREEN + RED) / (NIR + SWIR2)	Water SI	[46]
HV	*f*(SWIR2, NIR, RED)	Water indicator	[6]

Peer-reviewed version available at Remote Sens. 2020, 12, 495; doi:10.3390/rs12030495

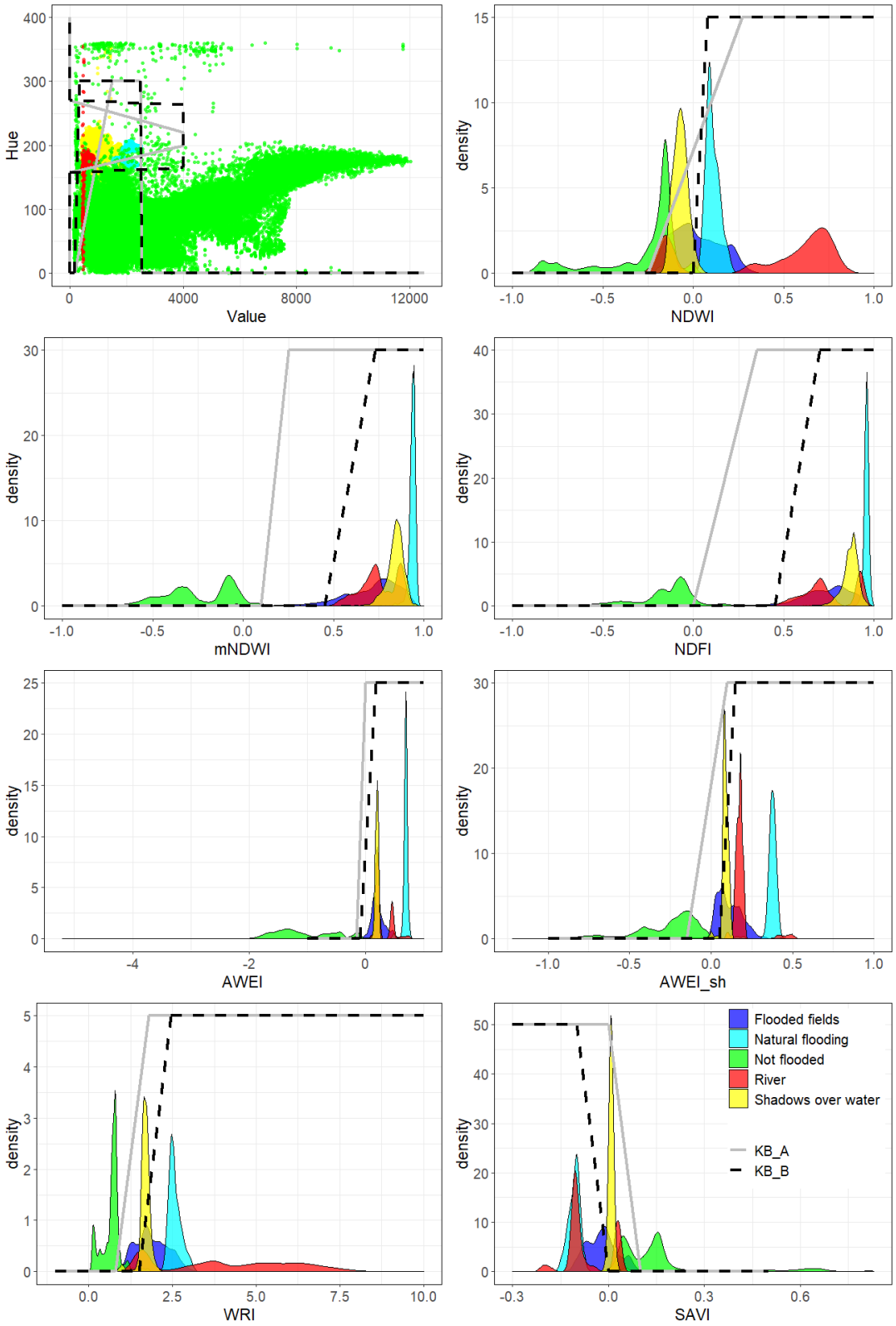


Figure 5: black dotted lines identify soft constraints defined by Expert A with *risky attitude*, i.e., optimistic w.r.t. mapping standing water regarded as a negative phenomenon, by taking into account the ability of soft constraints to separate the distributions of standing water (comprising the three classes "Natural flooding", "Flooded fields", "River") with respect to the "Not flooded" class. The **grey continuous lines** identify the soft constraints defined by Expert B with a *precautionary attitude*, i.e., pessimistic, by taking into account the ability of soft constraints to separate the distributions of standing water (comprising the four classes "Natural flooding", "Flooded fields", "River" and "Shadows over water") with respect to the "Not flooded" class.

Peer-reviewed version available at Remote Sens. 2020, 12, 495; doi:10.3390/rs12030495

4.3 Phase 1: Computation of partial evidence maps of environmental status

In phase 1 of the algorithm, by taking as input the pre-processed images, and the definitions of contributing factors and soft constraints in one of the knowledge base (either KB_A or KB_B), partial evidence maps are computed for each contributing factor. These are provided as output that is successively used by phase 2 to the aim of first learning the OWA operator and then computing the overall ESI map.

Phase 1 was executed twice: the first execution by using the soft constraints in KB_A, and the second by using KB_B, respectively. Thus, we obtained two distinct sets of partial evidence maps, indicated hereafter by PE_A and PE_B.

4.4 Phase 2: Learning the aggregation based on VGI and ESI computation

Phase 2 of the algorithm takes as input one set of partial evidence maps generated by a run of phase 1, either PE_A or PE_B, and a subset of VGI, and computes an ESI map. This consists in aggregating either PE_A or PE_B maps by applying the OWA operator that is learned by an iterative process using VGI assumed as ground truth. Output of this phase are: the ESI map, the weighting vector of the OWA operator, its *orness* and *dispersion* measures, and the correspondent label (as defined in Table 1) representing the decision attitude associated with the OWA.

By changing either the partial evidence maps in input or the VGI, different ESI maps can be computed for the same ROI. As it is described in the following subsection phase 2 was executed several times with distinct VGI subsets.

4.5 Validation experiments

The validation experiment was designed with the following objectives:

a) to compare the accuracy of the proposed approach with respect to traditional approaches based on single SI, on each single ROI;
b) to investigate the stability of results with respect to changing the ROI;
c) to investigate the stability of results with respect to changing expert knowledge (KB A and KB B);
d) to investigate the adaptability to local context by changing knowledge base in each ROI;
e) to investigate the performance achieved by the proposal by downscaling the dimension of the ground truth data set for learning the OWA aggregation.

The algorithm phase 1 was executed twice on the three ROIs to the aim of testing objectives c) and d), each one using the set of soft constraints in KB_A and KB_B, respectively. Accuracy of each single contributing factor in mapping standing water was evaluated by computing accuracy metrics True Positive (TP), True Negatives (TN), False Positives (FP) and False Negatives (FN) from the confusion matrix, commission (CE=FP/(FP+TP)) and omission (OE=FN/(FN+TP)) errors and F-score defined as follows:

$$F - score = \frac{2TP}{(2TP+FN+FP)} \quad (12)$$

Figure 6 reports the diagram of variation of the F-score measure in the three ROIs (the Cal/Val sites in Figure 4) obtained by using single contributing factors; reference set is composed of around 1000 VGI independent elements in each ROI as reported in third column of Table 3. Values of the F-score were computed by defining increasing thresholds on the SI domains with a 0.1 step and normalized in [0,1], and by assuming as "standing water" pixels with SI value exceeding the threshold. It can be noticed that F-score curves are not increasing with the threshold values. This is

Peer-reviewed version available at Remote Sens. 2020, 12, 495; doi:10.3390/rs12030495

because by increasing the threshold we are more strict on the selection of standing water pixels, thus we may increase omission errors by missing true standing water areas.

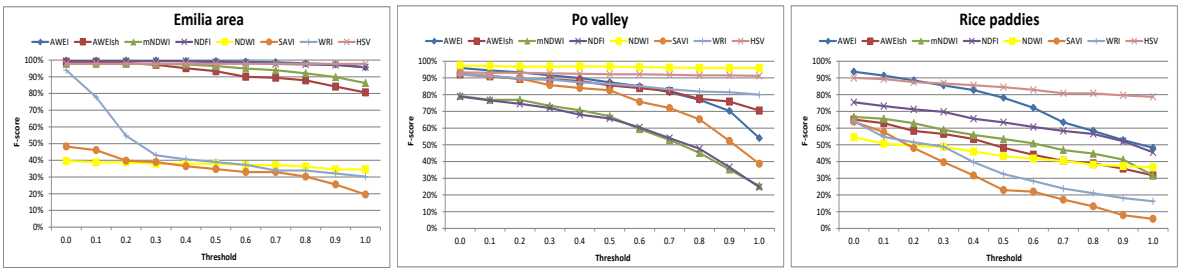


Figure 6: Diagrams show the variation of the F-score for eight distinct SIs defined in the literature for mapping standing water areas obtained by considering as watered those pixels whose SI value is above a threshold varying in [0,1].

It can be observed that in the three sites which are characterized by distinct land covers and water conditions (water depth, colour, fractional cover, plant/soil patches presence etc.) , a different SI presents the best performance (greatest F-score) for given values of the thresholds; this confirms our intuition that a single SI cannot capture all types of standing water conditions.

In Emilia area, AWEI, NDFI and HV have the best comparable performance for all thresholds; in the Po valley area the best index is NDWI followed by HV. Finally, in Rice paddies area, AWEI and HV are the best indices for threshold values below and above 0.3, respectively.

These results confirm the need of an aggregation phase capable to automatically select the best performing contributing factor for each pixel in each ROI in order to compute an ESI map.

This is achieved in phase 2 that applies an adaptation of the algorithm to a specific ROI depending on the used knowledge base, by exploiting available ground truth .

In order to achieve objectives a) – e) we designed two k-fold cross validation experiments, by using in input both KB_A and KB_B on each ROI.

We recall that a k-fold cross validation is a statistical method aimed at evaluating the performance of a learning algorithm by changing the training set; by doing so is it possible to compute both average performance metrics and standard deviation to assess its sensitivity.

In each experiment using either KB_A or KB_B, phase 2 was executed 10 times (k=10), thus generating ten weighting vectors of the OWA and consequently 10 distinct ESI maps for site: at each run a different subset of both ground truth data for learning and testing were selected by applying stratified random sampling. In these two experiments, we first used 90% of ground truth VGI elements for learning the OWA aggregation and 10% for testing as in the standard validation methods of machine learning. These experiments are named typical (T) k-fold cross validation

To achieve objective e) we performed other two experiments based on 10-fold cross validation, by using in input both KB_A and KB_B on each ROI but a different proportion of the learning and testing sets: differently than in the typical validation, this time we used a small subset of VGI elements for learning (only 10% of the available ground truth pixels) while we used the remaining 90% for testing. Stratified random sampling was applied to select the two subsets. This validation is called atypical (AT) and was aimed at investigating the stability of the results by decreasing drastically the learning set thus simulating a realistic situation with a small set of ground data.

Performance achieved on each ROI by the typical and atypical 10-fold cross validations is shown in Figure 7: the ten F-score diagrams in each area are relative to the ten ESI maps produced as a result of the ten executions of the algorithm phase 2 taking as input either KB_A or KB_B. Table 5 summaries average performance of the algorithm over all runs and all thresholds in both the typical and the atypical evaluation when using KB_A and KB_B, and when using the single best SI for each site. Table

Peer-reviewed version available at Remote Sens. 2020, 12, 495; doi:10.3390/rs12030495

6 reports the learned OWA operator averaged over the 10 runs when using both KB_A and KB_B in both the typical (T) and atypical (AT) validation settings.

Finally, Figure 8 illustrates for each Cal/Val site two maps highlighting in blue "standing water" areas identified by values of ESI > 0.5 computed based on either KB_A or KB_B.

5. Conclusions

The proposed approach for ESI map definition and computation applies artificial intelligence methodologies at distinct levels: it represents and manages the semantics of ill-defined expert's knowledge by means of soft constraints; aggregation strategies are applied whose semantics can be explained as modeling distinct decision attitudes; finally, OWA operators implementing the aggregation are defined by a machine learning mechanism to adapt the algorithm to the local context by copying with the subjectivity of expert's knowledge and exploiting a few VGI available. In the following we discuss the results achieved by the validation experiments carried out on the case study by targeting the single objectives of the validation experiments.

5.1. Objective a) comparison with traditional approaches based on single SIs

By looking at Table 5 we can observe that our proposal achieves results with performance equal or better than the results yielded by the best single SI in all the three ROIs. Besides this, an advantage of our proposal is that it can select automatically the best SIs for each single pixel in each site to determine the ESI value. This operation is not possible when using current approaches of multi criteria aggregation based on weighted average in which the weight is always associated with the same criterion for all pixels in a ROI. This means that our algorithm is adaptive to the local conditions and can recognize the distinct aspect of standing water.

Table 5: Average, standard deviation and average minimum F-score values over all 10 runs of the algorithm and all thresholds in the typical (T) and atypical (AT) validations using KB A and KB B and based on the best performing SI on the ROI. Best results are highlighted in bold.

Accuracy summary	10-fold cross	Average F-score (KB_A)	Average F-score (KB_B)	Std Dev (KB_A)	Std Dev (KB_B)	minimum F-score (KB_A)	minimum F-score (KB_B)
Rice Paddies	ESI (T)	0.896	**0.904**	0.043	0.024	0.823	0.865
	ESI (AT)	0.894	0.896	0.011	0.006	0.865	**0.886**
	HV	0.842		0.039		0.787	
Po valley	ESI (T)	**0.970**	0.960	0.027	0.012	0.920	0.933
	ESI Atypical	0.964	0.959	0.005	0.003	0.955	0.951
	NDWI	0.966		0.005		**0.959**	
Emilia area	ESI (T)	0.987	**0.992**	0.009	0.008	0.972	**0.978**
	ESI (AT)	0.949	0.961	0.023	0.018	0.930	0.947
	AWEI	0.988		0.013		0.957	

5.2. Objective b) investigating performance of the algorithm w.r.t changing the ROI

In Figure 7 we can observe that in all sites and in the typical validation setting, F-score diagrams are quite stable and maintain high values for all thresholds using both knowledge base KB_A and KB_B.

In the atypical setting the F-scores for some of the 10 runs decrease for high thresholds (above 0.9), especially in Emilia and Rice paddies areas. This may depend on the small dimension of the

Peer-reviewed version available at Remote Sens. 2020, 12, 495; doi:10.3390/rs12030495

learning set used that in some executions may not represent well all types of surface water. Nevertheless, as seen in Table 5, in all ROIs average F-scores are high in both the typical and atypical validations: in the Emilia and Po valley average F-scores are above 0.95 while in the Rice paddies site are above 0.90.

5.3. Objective c) investigating performance w.r.t changing knowledge base

From Table 5 we can notice that by using KB_B we get slightly better results than by using KB_A in the Emilia and Rice paddies sites, while KB_A performs better in the Po valley site. This is confirmed in both the typical and atypical validation settings. A reason could be that since in Emilia and Rice paddies sites flooded areas and inundated rice fields are present covering a mixture of situations (different water depth and patches of soil and vegetation), a precautionary attitude is more appropriate to map standing water while the risky attitude of expert A generates more False Negatives. In the in Po valley site, where we have a river water without influence of signal due to substrate or vegetation but only of water optical properties, a more optimistic, i.e., risky, attitude is the best choice since it does not generate so many False Positives. Furthermore, the stability of the results in terms of standard deviation are also better when using KB_B in both typical and atypical validations and in all the three sites. If we observe Figure 7 in the atypical validation a more consistent decreasing trend of some diagrams can be seen for KB_A with respect to KB_B. This again can be interpreted as due to the more optimistic attitude towards mapping an undesired phenomenon, i.e., risky attitude, of expert A: by considering as "standing water" only pixels with ESI > 0.9 more omission errors are produced when using KB_A with respect to KB_B which applies a more precautionary attitude.

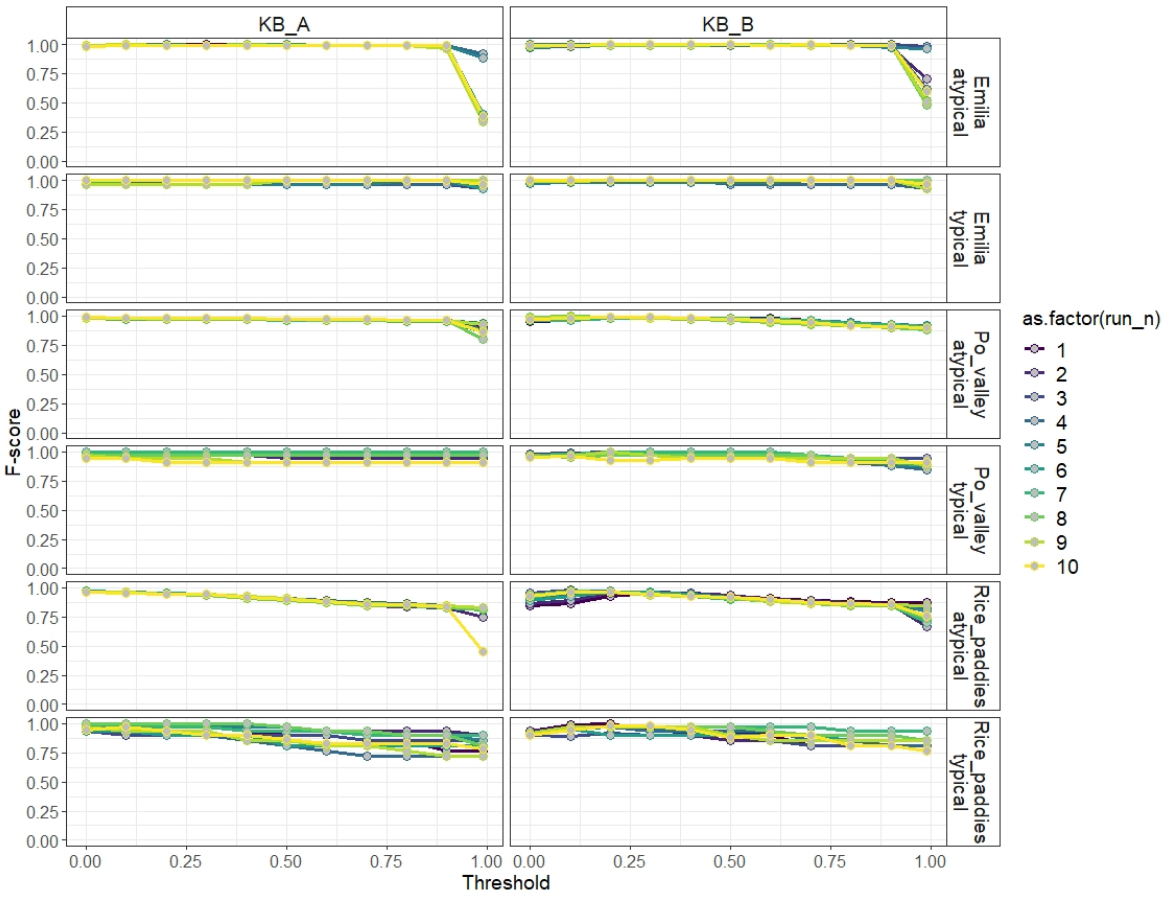


Figure 7: F-score diagrams on the three ROIs in the typical and atypical 10-fold cross validations using (a) KB_A and (b) KB_B. Parameters used for k-fold cross validations: k = 10, learning rate = 0.5, number of iterations = 500.

Peer-reviewed version available at Remote Sens. 2020, 12, 495; doi:10.3390/rs12030495

5.4. Objective d) investigating adaptability to local context and to the knowledge base

Table 6 reports for each site the average OWA operator learned in both the typical and atypical validations when considering both KB_A and KB_B. Specifically, it shows the weighting vector averaged over the runs of the 10-fold cross validation, the average *orness* of the learned OWAs, the correspondent standard deviation, the *dispersion* measure and the decision attitude (selected based on *orness* and *dispersion* values as listed in Table 1). It can be noticed that the algorithm can adapt the aggregation to site's characteristics by learning a distinct OWA weighting vector which is also stable (i.e. low standard deviation in each ROI). Finally, the decision attitude of the learned OWA operators does not change in the typical and atypical validation settings when using the same knowledge base.

Table 6: Learned weighting vectors of the OWA operator in each ROI averaged w.r.t. the 10 runs of both the Typical (T) and Atypical (AT) 10-fold cross validation when using KB_A and KB_B. The table also reports the values of the approximated weighting vector, the average *Orness* (Θ), the *Orness* Standard deviation (STD(Θ)), the *Dispersion* (Δ) and correspondent decision attitude.

10-fold cross with KB_A	**Learned OWA vector (Averaged over 10 runs on KB_A)**								Θ	STD(Θ)	Δ	**Decision attitude**
	w1	**w2**	**w3**	**w4**	**w5**	**w6**	**w7**	**w8**				
Emilia area (T)	0.25	0.43	0.3	0.015	0.005	0	0	0	0.8	0.030	0.6	*Semi Democratic & Towards Pessimism*
Emilia area (AT)	0.4	0.2	0.3	0.1	0	0	0	0	0.8	0.098	0.6	*Semi Democratic & Towards Pessimism*
Po valley (T)	1	0	0	0	0	0	0	0	1	0	0	*Dictatorial & Pessimistic*
Po valley (AT)	1	0	0	0	0	0	0	0	1	0	0	*Dictatorial & Pessimistic*
Rice Paddies (T)	1	0	0	0	0	0	0	0	1	0	0	*Dictatorial & Pessimistic*
Rice Paddies (AT)	1	0	0	0	0	0	0	0	1	0	0	*Dictatorial & Pessimistic*
10-fold cross with KB_B	**Learned OWA vector (Averaged over 10 runs on KB_B)**								Θ	STD(Θ)	Δ	**Decision attitude**
	w1	**w2**	**w3**	**w4**	**w5**	**w6**	**w7**	**w8**				
Emilia area (T)	0	0	0.7	0.3	0	0	0	0	0.7	0.000	0.3	*Semi Dictatorial & Towards Pessimism*
Emilia area (AT)	0	0.2	0.4	0.4	0	0	0	0	0.7	0.005	0.6	*Semi democratic & towards Pessimism*
Po valley (T)	0	0.8	0.2	0	0	0	0	0	0.8	0.000	0.2	*Semi Dictatorial & Towards Pessimism*
Po valley (AT)	0	0.7	0.3	0	0	0	0	0	0.8	0.002	0.3	*Semi Dictatorial & Towards Pessimism*
Rice Paddies (T)	0.1	0.3	0.6	0	0	0	0	0	0.8	0.000	0.4	*Semi Dictatorial & Towards Pessimism*
Rice Paddies (AT)	0.1	0.3	0.6	0	0	0	0	0	0.8	0.001	0.4	*Semi Dictatorial & Towards Pessimism*

When using KB_A in Emilia area, we obtained an OWA characterized by an average *orness* of 0.8 and a correspondent *dispersion* of 0.6 in both the typical and atypical validations, respectively, which corresponds to the attitude "*semi democratic and toward pessimism*". The weighting vector applies an aggregation mostly using the greatest five (four) arguments in the typical (atypical) settings, although with a different proportion. This means that there is not a unique partial evidence map that alone determines ESDI map. This ESI value can be explained by stating that it can increase depending on at most the greatest 5 (4) partial evidence degrees.

In the other two sites, the *orness* is 1 and the *dispersion* is 0, that corresponds to a "*dictatorial and pessimistic*" attitude in which the greatest partial evidence degree is the only determining ESI values.

Peer-reviewed version available at Remote Sens. 2020, 12, 495; doi:10.3390/rs12030495

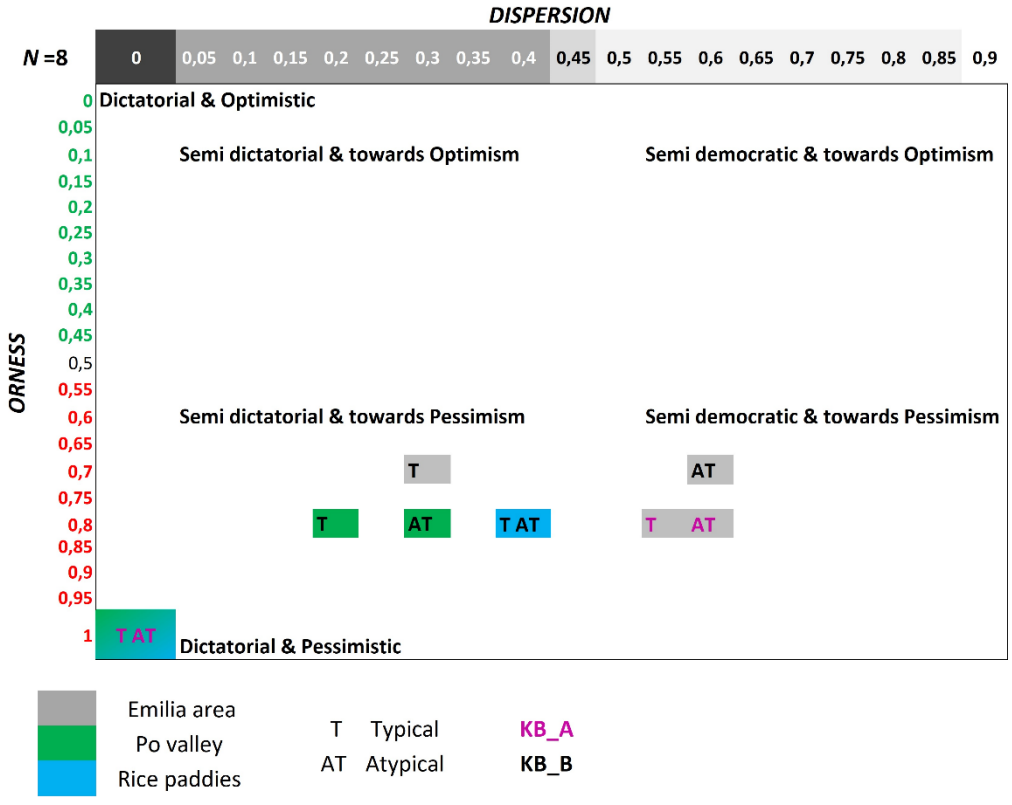
This means that in each pixel, ESI is determined by a single partial evidence; yet, due to the nature of the OWA, the partial evidence that contributes can derive from a different factor from pixel to pixel in the same site, thus allowing to capture the distinct aspects of standing water.

On the other side, when using KB_B, the algorithm adaptation mechanism can cope with the used knowledge by learning different OWA aggregations in each ROI.As shown in table 7, with KB_B the learned OWAs are stable in two out of the three ROIs. In the Po valley and Rice paddies sites, the generated OWA corresponds to "*semi dictatorial and toward pessimism*" decision attitude, meaning that based on the new knowledge base, we need a more synergic aggregation to optimize the ESI computation with respect to KB_A where in these sites the learned OWA was "*dictatorial & pessimistic*".

As far as Emilia site, we can observe that in the typical and atypical validations we obtain two OWA operators which differ just for the *dispersion* and not for the *orness*. Nevertheless, they are associated with two different decision attitudes although close one another: in the typical validation we obtain "*semi dictatorial and toward pessimism*" and in the atypical "*semi democratic & towards pessimism*". Both these aggregations determine a positive ESI value only if there are at least 2 (3) positive partial evidence degrees.

Looking at table 7, a general observation is that in using KB_B with respect to KB_A we generate OWA operators with smaller values of *orness* in all sites, i.e., OWA_B is more optimistic than OWA_A, and greater *dispersion*, i.e., OWA_B more democratic than OWA_A, in Po valley and rice paddies sites, except in the typical validation of Emilia area in which the dispersion decreases.

Table 7: *Orness-Dispersion* space in which the OWA operators learned in each ROI (identified by rectangles with distinct colors: grey Emilia area, green Po valley, light blue Rice paddies) in the Typical (T) and Atypical (AT) validations using KB_A (in violet) and KB_B (in black) are positioned according to their *Orness* and *Dispersion* measures.



Peer-reviewed version available at Remote Sens. 2020, 12, 495; doi:10.3390/rs12030495

These findings explicit the information on the roles that the partial evidence maps play depending on the knowledge base used. In the case of KB_A, in Po valley and Rice paddies sites, there is a full trust of each partial evidence map, so that the most pessimistic one determines the ESI. This allows to counterbalance the optimistic and risky attitude of expert A who defined strict soft constraints by disregarding "Shadows over water": too many omission errors could be generated by applying a synergic aggregation, which instead is appropriate in the Emilia area.

In the case of KB_B, more partial evidence maps are aggregated in a synergic way to compute ESI maps, so as not to obtain too many commission errors which may cause false alarms.

This reveals how the learning mechanism can cope with the subjectivity of the expert's knowledge. Figure 8 shows for each site two ESI maps generated by using KB_A and KB_B. It can be seen that the standing water areas are quite similar in each pair of maps of the same site.

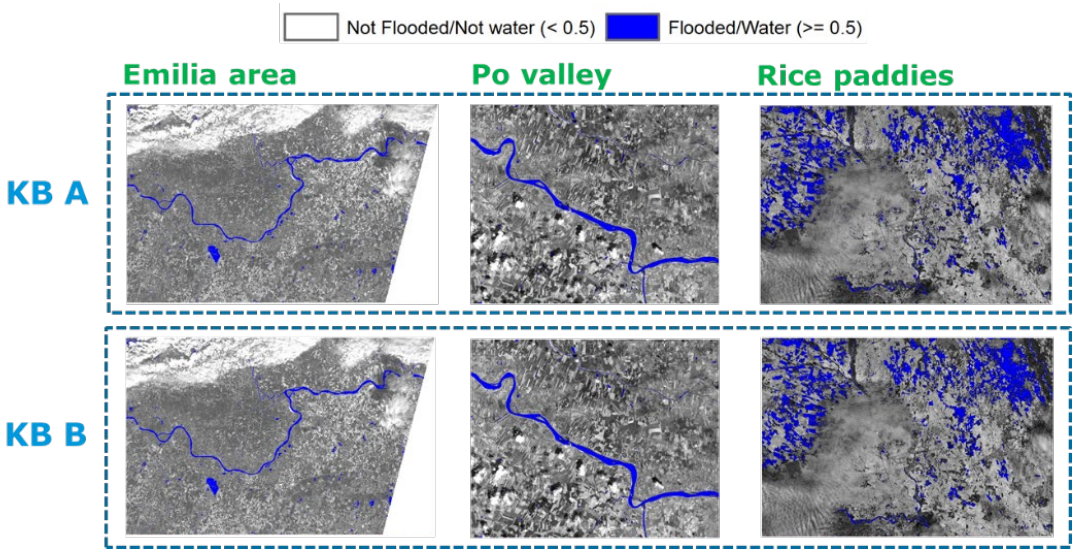


Figure 8: ESI maps in the three ROIs obtained by averaging the weighting vectors of the OWA operators learned on 10 run of algorithm in the atypical setting using the two knowledge bases A and B: values considered as standing water are in blue and are obtained by a threshold on ESI > 0.5.

5.5. Objective e) comparing the performance of the typical and atypical validations

This objective is aimed at assessing if, in a realistic situation in which the available ground truth is scarce, below 100 VGI elements, our automatic algorithm can still produce acceptable results. Of course, by having a larger training set, one gives the algorithm a better chance of understand underlying patterns, rather than just learning to identify specific examples from the training set. Nevertheless, learning from specific examples available in a local context can be effective if we assume that a local context is characterized by the presence of a specific kind of standing water, and if we use the result just for that area. If we compare the F-score diagrams in Figure 7 we can observe that independently of the used knowledge base and up to a threshold of 0.9, the F-score values are very high in both validation settings. For thresholds above 0.9, F-score for atypical validation has a greater decreasing rate compared to the typical values. This means that when we have a small ground truth set for the adaptation to the ROI, one must be careful when segmenting the resulting ESI map to identify the standing water areas. We have to choose a threshold which is below 0.9 to avoid generating too high omission error rates. On the other side, when considering a threshold on ESI > 0, we can be confident on the high accuracy of the ESI mapping obtained even when performing a learning based on a small ground truth set.

Future work needs to be done for confirming these findings. First of all, a more extensive validation on other sites is needed to evaluate the robustness of the approach depending on the ROI. Second, the application of the approach should be experienced by exploiting both different contributing factors and a reduced number of them. We intend exploring the use of all single spectral

Peer-reviewed version available at Remote Sens. 2020, 12, 495; doi:10.3390/rs12030495

bands as contributing factors and directly defining the soft constraints on their domains so as to obtain a fuzzy spectral signatures. Furthermore, the literature on the thresholds used to segment the SIs for identifying standing water, would be worth considering as an alternative knowledge base for defining the soft constraints.

Finally, the approach should be applied for different purposes, such as to detect burned areas, ice and snow area on glaciers, evidence of draughts sites, and crop damage/stress in agriculture.

Author Contributions: "conceptualization, G.B.; methodology, G.B., M.B., P.A.B., and D.S; software and validation, A.G.; data curation, A.G., D.S., and M.B; writing and editing, G.B.; review All; funding acquisition, G.B."

Funding: This work has been conducted in the frame of the Fondazione CARIPLO project "STRESS: Strategies, Tools and new data for REsilient Smart Societies" #2016-0766, "Bando Fondazione Rst - Ricerca dedicata al dissesto idrogeologico 2016" and SIMULATOR-ADS project, # 137287, co-funded by Regione Lombardia & FESR "Linea R&S per Aggregazioni".

Conflicts of Interest: "The authors declare no conflict of interest."

Peer-reviewed version available at Remote Sens. 2020, 12, 495; doi:10.3390/rs12030495

References

1. Dean, J.; Ghemawat, S. Mapreduce: simplified data processing on large clusters. *Commun. ACM* **2008**, 51(1), 107–113.
2. Boschetti, M.; Nutini, F.; Manfron, G.; Brivio, P.A.; Nelson, A. Comparative analysis of normalised difference spectral indices derived from MODIS for detecting surface water in flooded rice cropping systems, *PLoS One* **2014**, 9, 10.1371/journal.pone.0088741
3. Feyisa, G.L.; Meilby, H.; Fensholt, R.; Proud, S.R. Remote sensing of environment automated water extraction index: a new technique for surface water mapping using Landsat imagery, *Remote Sens. Environ.* **2014**, 140, 23-35, 10.1016/j.rse.2013.08.029
4. Huete, A.R. A soil-adjusted vegetation index (SAVI), *Remote Sens. Environ*. **1988**, 25, 295-309.
5. McFeeters, S.K. The use of the normalized difference water index (NDWI) in the delineation of open water features, *Int. J. Remote Sens*. **1996**, 17, pp. 1425-1432, doi: 10.1080/01431169608948714
6. Pekel, J.F.; Vancutsem, C.; Bastin, L.; Clerici, M.; Vanbogaert, E.; Bartholomé, E.; Defourny, P. A near real-time water surface detection method based on HSV transformation of MODIS multi-Spectral time series data, *Remote Sens. Environ*. **2014**, 140, 704-716.
7. Yager, R.R. On ordered weighted averaging aggregation operators in multi-criteria decision making. *IEEE Trans. on Systems, Man and Cybernetics* **1988**, 18, 183-190.
8. Zadeh, L.A. Fuzzy Sets, *Information and Control* **1965**, 8 (3), 338-353
9. Kacprzyk, J. Fuzzy Logic with Linguistic Quantifiers: A Tool for Better Modeling of Human Evidence Aggregation Processes?, *Advances in Psychology* **1988**
10. Zadeh, L.A. A computational approach to fuzzy quantifiers in natural languages, *Comps. & Maths. with Appls.* **1983**, 9, 149-184.
11. Yager, R.R., Quantifier guided aggregation using OWA operators. *International Journal of Intelligent Systems* **1996**, 11, 49–73.
12. Yager, R.R. On OWA operators in multi-criteria decision making, *IEEE Trans. Syst. Man, Cybern* **1988**.
13. Yager, R.R. New modes of OWA information fusion, *Int. J. Intell. Sys*. **1998**
14. Bordogna, G.; Pagani, M; Pasi, G. Imperfect Multisource Spatial data Fusion based on a Local Consensual Dynamics, in Uncertainty Approaches for Spatial Data Modeling and Processing, Kacprzyk, J., Petry F. E., Yazici, A. (eds), Springer Verlag, ISBN: 978-3-642-10662-0, 2010, pp 79-94.
15. Yager, R.R. On the dispersion measure of OWA operators, *Information Sciences* **2009**, 179, 3908–3919.
16. Bordogna, G.; Frigerio, L.; Kliment, T.; Brivio, P.A.; Hossard, L.; Manfron, G.; Sterlacchini, S. "Contextualized VGI" Creation and Management to Cope with Uncertainty and Imprecision. *ISPRS Int. J. Geo-Inf.* **2016**, 5, 234.
17. Arcaini, P.; Bordogna, G. Geotemporal Querying of Social Networks and Summarization, Encyclopedia of Social Network Analysis and Mining, **2018** Edition, Alhajj R. and Rokne J (eds), https://doi.org/10.1007/978-1-4939-7131-2_110156
18. Humanitarian Open Street Map https://www.hotosm.org/docs/
19. Filev, D.P.; Yager, R.R. On the issue of obtaining OWA operator weights, *Fuzzy Sets and Systems*, **1998**, 94, 157-169.
20. Lecun, Y.; Bengio, Y.; Hinton, G.E. Deep Learning. *Nature*, **2015**, 521, 436–444.
21. Reichstein, M.; Camps-Valls, G.; Stevens, B; *et al.* Deep learning and process understanding for data-driven Earth system science. *Nature* **2019**, 566**,** 195–204 doi:10.1038/s41586-019-0912-1
22. Zhang, L.; Zhang, L.; Du, B. Deep Learning for Remote Sensing Data: A Technical Tutorial on the State of the Art. *IEEE Geosci. Remote Sens. Mag*. **2016**, 4, 22–40.
23. Castelluccio, M.; Poggi, G.; Sansone, C.; Verdoliva, L. Land Use Classification in Remote Sensing Images by Convolutional Neural Networks. arXiv, arXiv:1508.00092, **2015**.
24. Larochelle, H.; Bengio, Y.; Louradour, J.; Lamblin, P. Exploring strategies for training deep neural networks. *J. Mach. Learn. Res*. **2009**, 10, 1–40.
25. Han, X.; Zhong, Y.; Cao, L.; Zhang, L. Pre-Trained AlexNet Architecture with Pyramid Pooling and Supervision for High Spatial Resolution Remote Sensing Image Scene Classification, *Remote Sens.* **2017**, 9(8), 848; https://doi.org/10.3390/rs9080848
26. Acharya, T.D.; Subedi, A.; Lee, D.H. Evaluation of water indices for surface water extraction in a Landsat 8 scene of Nepal, *Sensors* **2018**, 18, p. 2580

Peer-reviewed version available at Remote Sens. 2020, 12, 495; doi:10.3390/rs12030495

27. Ji, L.; Zhang L.; Wylie B.; Analysis of Dynamic Thresholds for the Normalized Difference Water Index, Photogrammetric Engineering & Remote Sensing, **2009,** 11, 1307-317, https://doi.org/10.14358/PERS.75.11.1307.
28. Estoque, R.C.; Murayama, Y.; Classification and change detection of built-up lands from Landsat-7 ETM+ and Landsat-8 OLI/TIRS imageries: A comparative assessment of various spectral indices, Ecological Indicators, **2015,** 56, 205-217, ISSN 1470-160X, https://doi.org/10.1016/j.ecolind.2015.03.037.
29. Dempewolf, J.; Trigg, S.; DeFries, R.S.; Eby, S., Burned-Area Mapping of the Serengeti–Mara Region Using MODIS Reflectance Data, IEEE Geoscience and Remote Sensing Letters, **2007**, 4(2), 312-316, doi: 10.1109/LGRS.2007.894140.
30. Qiu, B.; Zhang, K.; Tang, Z.; Chen, C.; Wang, Z; Developing soil indices based on brightness, darkness, and greenness to improve land surface mapping accuracy, GIScience & Remote Sensing, **2017**, 54(5), 759–777, https://doi.org/10.1080/15481603.2017.1328758.
31. Xu, H. Modification of normalised difference water index (NDWI) to enhance open water features in remotely sensed imagery, *Int. J. Remote Sens*. **2006**, 27, 3025-3033, doi: 10.1080/01431160600589179
32. Bordogna, G.; Pagani, M.; Pasi, G. A Flexible Decision support approach to model ill-defined knowledge in GIS. In *Geographic Uncertainty in Environmental Security*, A. Morris and S. Kokham (eds.), NATO Science for Peace and Security series (Springer) ISBN 978-1-4020-6437-1, **2006,** 133-152.
33. Brivio, P.A.; Boschetti, M.; Carrara, P.; Stroppiana, D.; Bordogna, G. Fuzzy integration of satellite data for detecting environmental anomalies across Africa. In *Advances in Remote Sensing and Geoinformation Processing for Land Degradation Assessment*, J. Hill and A. Roeder (eds), (London: Taylor & Francis) **2006**.
34. Carrara, P.; Bordogna, G.; Boschetti, M.; Brivio, P. A.; Nelson, A. D.; Stroppiana, D. A flexible multi-source spatial-data fusion system for environmental status assessment at continental scale. *Int. J. of Geographical Information Science* **2008**, 22(7), 781-799. https://doi.org/10.1080/13658810701703183.
35. Bordogna, G.; Boschetti, M.; Brivio, P.A.; Carrara, P.; Pagani, M.; Stroppiana, D. Fusion Strategies based on the OWA Operator in Environmental Applications, in *Recent Developments in the Ordered Weighted Averaging Operators: Theory and Practice*, Kacprzyk J., Yager R.R., Beliakov G., (eds), Sprinter Verlag Vol. 265, 1st Edition., ISBN: 978-3-642-17909-9 2011, **2011**, XIV, 298
36. Goffi, A.; Stroppiana, D.; Brivio, P.A.; Bordogna, G.; Boschetti, M. 2020, Towards an automated approach to map flooded areas from Sentinel-2 MSI data and soft integration of water spectral features, *Int. Journal of Applied Earth Obs. and Geoinfo.* **2020**, 84, 101951.
37. Bloch, I.; Maître, H. Information combination operators for data fusion: A comparative review with classification. *IEEE Transactions on Systems, Man, and Cybernetics, part A: systems and humans* **1996**, 26(1), 52-67.
38. Bone, C.; Dragicevic, S.; Roberts, A. Integrating high resolution remote sensing, GIS and fuzzy set theory for identifying susceptibility areas of forest insect infestations. *Int. Journal of Remote Sens*. **2005**, 26(21), 4809-4828.
39. Chanussot, J.; Mauris, G.; Lambert, P. Fuzzy fusion techniques for linear features detection in multitemporal SAR images. *IEEE Trans. on Geosci. and Remote Sens*. **1999**, 37(3), 1292-1305.
40. Jiang, H.; Eastman, J.R. Application of fuzzy measures in multi-criteria evaluation in GIS. *Int. Journal of Geographical Information Science* **2000**, 14(2), 173-184.
41. Malczewski, J. GIS-based multicriteria decision analysis: a survey of the literature. *Int. Journal of Geographical Information Science* **2006**, 20(7), 703-726.
42. Robinson, P.B. A perspective on the fundamentals of fuzzy sets and their use in Geographic Information Systems. *Transactions in GIS* **2003**, 7(1), 3-30.
43. Karloff, H.; Suri, S.; Vassilvitskii, S. A model of computation for MapReduce. In *Proc. of the twenty-first annual ACM-SIAM Symp. on Discrete algorithms* (SODA '10). Society for Industrial and Applied Mathematics, Philadelphia, PA, USA, **2010**, 938-948.
44. Ranghetti, L.; Busetto, L. Sen2r: An R Toolbox to Find, Download and Preprocess Sentinel-2 Data. R Package Version 1.0.0, **2019**, 10.5281/zenodo.1240384 URL: http://sen2r.ranghetti.info
45. Smith, A.R. Color gamut transform pairs, *Proc. of the 5th Annual Conf. Comput. Graph. Interact. Tech.*, **1978**, 23–25, 12–19.
46. Shen, L.; Li, C. Water body extraction from landsat ETM+ imagery using adaboost algorithm. *18th Int. Conf. Geoinformatics*, **2010**, 10.1109/GEOINFORMATICS.2010.5567762, Geoinformatics 2010 1–4