

Per un uso del suolo dinamico: classificazione di serie storiche di immagini Sentinel-2

Laura Annovazzi-Lodi¹[0000-0003-3939-9170], Vittorio Casella¹[0000-0003-2086-7931], Christian Massimiliano Baldin¹[0000-0002-8513-3472], Alice Bernini², Odunayo David Adeniyi², Michael Maerker²

¹ DICAR – Dip. Ingegneria Civile e Architettura - Università degli Studi di Pavia, (laura.annovazzilodi; vittorio.casella; christianmassimiliano.baldin)@unipv.it

² Dip. Scienze della Terra e dell’Ambiente – Università degli Studi di Pavia, michael.maerker@unipv.it

Abstract. Nell’ambito del progetto della Regione Lombardia denominato CE4WE, viene effettuata la classificazione di lunghe e frequenti serie storiche di immagini Sentinel-2, allo scopo di ottenere una mappa di uso del suolo dinamica. La classificazione è stata effettuata, per ora, nella sola modalità a immagine singola. Sono stati tuttavia considerati diversi scenari per l’istruzione del classificatore, che può essere a tempo singolo o a tempi multipli. Inoltre sono stati considerati tre scenari per le bande usate per la classificazione: uso delle sole bande; uso delle bande e degli indici; uso dei soli indici. Per quanto riguarda il classificatore adottato, in questa fase dello studio, è stato utilizzato l’algoritmo SVM, frequentemente impiegato in ambito Machine Learning. Successivamente verrà testato il Deep-Learning. Il presente articolo illustra alcuni primi risultati di tale attività.

Parole chiave: Classificazione, Land Use, SVM, serie storiche, Machine Learning

1 Introduzione

La Regione Lombardia ha attivato nell’anno 2019 un’iniziativa denominata “Hub Ricerca e Innovazione” volto a finanziare proposte e iniziative di ricerca e innovazione intraprese da gruppi composti da centri di ricerca e aziende.

Fra i progetti finanziati, vi è CE4WE, acronimo per *Circular Economy for Water and Energy* - Approvvigionamento energetico e gestione della risorsa idrica nell’ottica dell’Economia Circolare [1]. Esso ha come coordinatore l’Università di Pavia e come partner industriali due importanti realtà nel campo del Ciclo Idrico Integrato, CAP e A2A-ciclo idrico, un colosso nel campo dell’energia, ENI, e due PMI a forte vocazione innovativa, NeoruraleHub e MOGU. Il Progetto è iniziato il 1 febbraio 2020 e dovrebbe terminare a luglio 2022, anche se uno slittamento dovuto alla pandemia è ipotizzabile.

Il CE4WE mira a sviluppare soluzioni innovative su aspetti chiave della gestione del ciclo idrico e la valorizzazione ai fini energetici di materiali di scarto.

Per gestire al meglio le risorse coinvolte, esso è operativamente organizzato in 3 pilastri (P), linee di azione (WP) e azioni all’interno di ciascuna linea. Il pilastro P1 si

occupa del Ciclo dell'Acqua, il P2 riguarda la produzione di biocarburanti e materiali a valore aggiunto da prodotti di scarto, mentre il P3 concerne la sostenibilità integrata.

Mentre il P1 e il P2 sono di tipo tematico, le attività riguardanti il P3 sono di tipo trasversale rispetto ai primi due e produrranno una valutazione della sostenibilità e delle norme in vigore, promuovendo una forma evoluta di partenariato capace di esaltare le specificità di ciascun partner e favorire un approccio gestionale che ottimizzi le ricadute sul territorio delle attività di progetto.

Il Laboratorio di Geomatica dell'Università degli Studi di Pavia fa parte del pilastro P1 e ha come linea d'azione il WP 1.3. Essa intende definire buone pratiche per la costruzione di modelli per la parametrizzazione del primo sottosuolo e la simulazione numerica di flusso idrodinamico attraverso l'interfaccia suolo, anche con eventuale sviluppo di sensoristica dedicata. Le attività in corso comprendono l'utilizzazione di sistemi di machine learning per creare mappe di uso del suolo con input satellitari. Le mappe sono classificate in modalità tempo-dipendente per l'istruzione di algoritmi sulla variabilità temporale. Conseguentemente, si prevede la costruzione di mappe digitali delle caratteristiche del suolo nelle aree di studio in esame, attraverso i principali parametri chimico-fisici. È prevista anche attività sul campo per la misura e il campionamento attraverso stazioni di monitoraggio.

Tutte le informazioni vengono gestite per il progetto e per il pubblico con WebGIS e banche dati spaziali.

2 Scopo

Come linee d'azione, il Laboratorio di Geomatica include la rilevazione delle caratteristiche superficiali spettrali con tecniche di proximal e remote sensing e il post-processamento di dati ottenuti dalle analisi di laboratorio e di campagna, tramite una omogeneizzazione e georeferenziazione degli stessi.

Più in dettaglio, tra i vari obiettivi, il Progetto prevede per il WP 1.3 l'applicazione di un modello che simula la dinamica del bilancio idrico del suolo per studiare in particolare la diffusione di inquinanti nelle falde in varie zone della Regione. In questo quadro, verrà modellizzato in modo dettagliato il flusso delle acque superficiali mediante un modello che simula il bilancio idrico del suolo. Tale modello, il Soil Water Assessment Tool (SWAT), richiede specifici dati di input quali topografia, suolo, uso del suolo, dati climatici e gestione dei terreni [2]. Visto il dettaglio con cui tale modellazione verrà effettuata, per la caratterizzazione del terreno, verrà utilizzato un approccio di Digital Soil Mapping, al fine di ottenere una carta del suolo accurata.

Inoltre, il modello SWAT richiede mappe di uso del suolo minuziose e dinamiche, poiché lo stato di alcune parti del territorio varia anche più volte nello stesso anno, per esempio: loietto, terreno nudo, mais, terreno nudo o, ancora, terreno nudo, orzo, terreno nudo, mais.

Tali mappe verranno calcolate mediante l'applicazione di tecniche di classificazione supervisionate alle immagini acquisite da Sentinel-2, un sistema satellitare nato nell'ambito del programma europeo Copernicus [3].

3 Aree di studio

L'algoritmo di classificazione è stato istruito e collaudato su un'area test-site, individuata per un progetto precedente [4]. Essa si trova a circa 15 km a nord ovest della città di Pavia, prossima al paese di Bereguardo. Il sito considerato appartiene al territorio della Pianura Padana. L'area di studio contiene diverse coperture d'uso del suolo come terreni coltivati, boschi, aree industriali e urbane e un tratto del fiume Ticino, compresi i suoi meandri.

In seguito, i classificatori creati sono stati utilizzati per creare le mappe di uso del suolo dei siti identificati per il CE4WE. Essi si trovano a nord del test-site, tra il bacino di Scavizzolo e il bacino di Abbiategrasso. Queste aree sono state scelte per avere una caratterizzazione di un territorio molto particolare, con un sistema di irrigazione sofisticato, che si estende sui vari livelli dei terrazzi fluviali del fiume Ticino. La zona scelta copre sia la sponda destra che quella sinistra del Ticino, dove l'uso del terreno e, in particolare, il suolo stesso presentano delle caratteristiche differenti.

Contestualmente il Laboratorio di Geomatica ha sviluppato un WebGIS, attraverso la Web AppBuilder di ArcGIS Pro [5], per l'identificazione dell'area test e dei siti oggetti di studio, che nel tempo è diventato un grande contenitore di altri diversi dati prodotti dal progetto, con la possibilità di un costante ampliamento con ulteriori inserimenti.

L'immagine (Fig.1) illustra il WebGIS creato con le 3 zone utilizzate per il nostro WP. Il test-site, identificato con un rettangolo in nero, è denominato area_Bereguardo, mentre Bacino_Scavizzolo e Bacino_Abbiategrasso, rispettivamente in verde e in rosso, concernono le aree del Progetto.

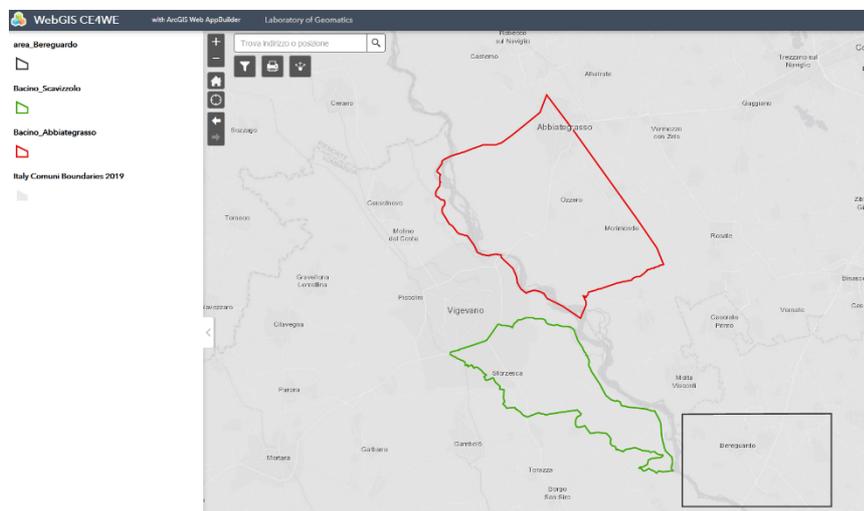


Fig. 1- WebGIS contenente il test site e le aree di studio del progetto CE4WE.

4 Metodologia

È stato deciso, al momento, di prendere in considerazione un intervallo temporale di un anno. È stata costruita una verità al suolo dinamica, cioè variabile con il tempo. Essa è frutto dell'evoluzione di uno studio precedentemente effettuato su singola immagine [4]. Lo strumento adottato è quello della mappa GIS multi temporale che assegna, a ognuno dei poligoni considerati, uno stato diverso a seconda della data considerata. La risoluzione temporale della mappa è pari a un giorno.

Intervistando i proprietari delle aziende agricole, sono state raccolte le informazioni di verità al suolo per ogni appezzamento agricolo dell'area. È stato necessario uno studio approfondito circa la fenologia delle colture agricole presenti, prendendo in considerazione la cronologia e la relazione tra i fattori climatici e gli eventi stagionali nel tempo. Per questo motivo è stato indispensabile creare un cronogramma che considerasse le date di semina, emergenza, sviluppo e senescenza.

I poligoni di verità al suolo così creati sono stati suddivisi e utilizzati in parte per la fase di istruzione dell'algoritmo e la classificazione (training sample) e in parte per la fase di validazione (test set).

Sono state scaricate e pre-elaborate tutte le immagini Sentinel-2 acquisite nel corso dell'anno 2017 e aventi buona qualità, senza copertura di nuvole. In totale, sono state selezionate 14 immagini. È da precisare che ogni tile di Sentinel-2 copre un'area di 100 km x 100 km. Con l'obiettivo di alleviare il carico di dati durante le fasi di elaborazione della classificazione, il tile è stata ritagliato in modo da circoscrivere solo l'area di interesse per il nostro studio. La correzione dell'atmosfera non è stata necessaria perché l'immagine era chiara all'interno del sito di studio. Grazie al terreno pianeggiante e la buona precisione di geolocalizzazione, non è stata nemmeno necessaria la pre-elaborazione geometrica.

Le quattro bande spettrali del visibile (B2, B3, B4) e NIR (B8) hanno una risoluzione spaziale di 10 m. Le restanti sei bande acquisite a 20 m, ovvero le tre di Red Edge (B5, B6 e B7), la banda Narrow NIR (B8A) e le due di SWIR (B11 e B12), e le tre bande atmosferiche (B1, B9 e B10), aventi risoluzione spaziale pari a 60 m, sono state ricampionate. Questo è stato fatto per ottenere un layer stack di 13 bande spettrali a 10 m.

Agli usuali livelli delle immagini Sentinel-2, sono stati aggiunti indici relativi alla vegetazione, alla minimizzazione dell'effetto del suolo nudo, alla distinzione tra acqua e urbano e alla discriminazione fra vegetazione e costruito; in particolare: NDVI, SAVI, VOG1, PRI, NDII, e NDBI.

L'indice NDVI (Normalized Difference Vegetation Index) è comunemente utilizzato per denotare la quantità e il vigore della vegetazione in un'immagine e per differenziare le aree vegetate e non vegetate. Esso è un indicatore numerico che utilizza le bande del visibile e del vicino infrarosso dello spettro elettromagnetico [6].

L'indice SAVI (Soil-Adjusted Vegetation Index) è molto simile a NDVI. I due indici spettrali sono calibrati in modo tale che le variazioni dei suoli siano normalizzate e non influenzino le misurazioni della chioma vegetale. SAVI è stato

sviluppato per ridurre al minimo le influenze della luminosità del suolo, incorporando un fattore di aggiustamento [7].

Gli indici VOG1 (Vogelmann Index) [8], PRI (Photochemical Reflectance Index) [9] e NDII (Normalized Difference Infrared Index) [10] sono utilizzati per applicazioni che includono studi sulla fenologia della vegetazione, produttività, monitoraggio della chioma e rilevamento dello stress. In particolare, NDII si è rivelato un proxy molto forte per la denotazione del deficit di accumulo e stoccaggio dell'umidità del suolo, che è una componente cruciale dei modelli idrologici. Esso sembra essere un indicatore affidabile della distribuzione temporale e spaziale delle condizioni di siccità.

Infine, l'indice NDBI (Normalized Difference Built-up Index) è impiegato per estrarre l'area edificata, sulla base del fatto che le aree urbane hanno una risposta di riflettanza nello SWIR superiore alle altre bande spettrali [11].

La classificazione supervisionata è stata effettuata, per ora, nella sola modalità a immagine singola sull'area test, di cui si conosce la verità a terra. Sono stati tuttavia considerati diversi scenari per l'istruzione del classificatore, che può essere a tempo singolo o a tempi multipli. Inoltre sono stati considerati tre scenari per le bande e i parametri usati per la classificazione: uso delle sole bande originarie, previo ricampionamento; uso dei soli indici; uso delle bande originarie, previo ricampionamento, e degli indici. Per questo step, abbiamo utilizzato il software ERDAS IMAGINE di Hexagon, creando dei modelli appositi attraverso lo Spatial Modeler [11].

Per quanto riguarda il classificatore adottato, in questa fase dello studio, è stato utilizzato l'algoritmo SVM (Support Vector Machine), frequentemente impiegato in ambito Machine Learning. Il vantaggio principale di questo algoritmo di apprendimento automatico è che può funzionare con successo con un numero esiguo di esempi [13], come nel nostro caso. Sviluppato da Vapnik e dai suoi collaboratori, invece di stimare le densità di probabilità delle classi, risolve direttamente il problema di interesse determinando i confini di classificazione tra le classi [14].

Senza la fase di validazione, la mappa classificata finale non può essere considerata affidabile e, pertanto, la sua applicabilità risulta limitata. L'accuratezza dell'immagine classificata è valutata confrontando la mappa ottenuta dal processo di classificazione con i dati di riferimento (test set). Questa fase fornisce informazioni sulla qualità del prodotto e, analizzando la matrice di confusione generata, individua probabili fonti di errori osservabili valutando alcune misure statistiche comuni, quali l'accuratezza complessiva, l'accuratezza del produttore, l'accuratezza dell'utente, gli errori di omissione e commissione e il coefficiente Kappa.

Considerate le 14 immagini telerilevate utilizzate per istruire l'algoritmo, sono stati così creati e validati 14 classificatori. Tenendo presente che i territori dell'area-test (Beregardo) e dei due siti di studio del progetto CE4WE (Scavizzolo e Abbiategrasso) sono prossimi geograficamente e simili per orografia, si è preceduto a classificare le immagini Sentinel-2 dell'anno 2020, inerenti le aree di Scavizzolo e Abbiategrasso, utilizzando i classificatori istruiti sull'area di calibrazione di Bereguardo. Il classificatore applicato a ciascuna nuova immagine è stato scelto con un criterio relativo alla data più prossima, pari a 20 giorni.

Infine, le immagini classificate sono state post-elaborate: le aree affette da rumore sono state migliorate. I pixel contigui appartenenti alla stessa classe sono stati identificati e raggruppati, e, attraverso un'operazione detta di majority filter, tutte le aree più piccole di un certo parametro di soglia sono state eliminate e lo spazio vuoto è stato assegnato alla classe predominante nei dintorni dell'area stessa. Le immagini così ottenute risultano ripulite in gran parte dell'effetto cosiddetto "sale e pepe".

5 Risultati

Si illustrano i risultati ottenuti nella classificazione delle immagini a tempo singolo per tutto il dataset relativo a un anno solare. A partire dalle singole matrici di confusione, si estraggono valori di sintesi in modo da quantificare la performance delle varie strategie menzionate sopra.

In prima istanza, si considerino i tre diversi scenari: uso delle sole bande originarie, previo ricampionamento; uso dei soli indici; uso delle bande e degli indici.

E' stato fatto un confronto per la data 17 maggio 2017. L'accuratezza totale (OA: Overall Accuracy) ottenuta utilizzando le sole bande spettrali è stata pari 0.8491. E' stato deciso di unire le classi di Urbano, Industriale e Asfalto in un'unica categoria denominata Artificiale, per una migliore discriminazione, in quanto l'algoritmo non era in grado di distinguere le classi di Asfalto e Industriale ed era presente confusione tra Terreno nudo e Urbano.

Esaminando il secondo scenario, ovvero la classificazione istruendo l'algoritmo con l'utilizzo dei soli sei indici, l'OA risulta pari a 0.7087. Il coefficiente Kappa è molto debole ed è pari a 0.462. Le classi di Orzo, Artificiale e Prato Permanente non vengono riconosciute.

Il terzo scenario, quello che considera l'istruzione dell'algoritmo usufruendo sia degli indici che delle bande spettrali, risulta di gran lunga il più favorevole e robusto.

Analizzando la matrice di confusione (vedi tab.1), si nota che la classe dell'Artificiale viene misuratamente confusa con quella del Terreno Nudo, mentre la classe dell'Orzo viene confusa con quella del Prato e Terreno Nudo portando a conseguenti valori discreti rispettivamente di UA (User's Accuracy) pari a 0.689 e PA (Producer's Accuracy) pari a 0.501. Tuttavia, i restanti parametri di qualità della classificazione risultano molto buoni e soddisfacenti. L'accuratezza totale che ne consegue è pari a 0.8942, con un coefficiente Kappa molto forte pari a 0.827.

E' stato valutato anche il solo impiego del solo indice NDVI affiancato all'utilizzo delle bande, che ha portato ad un'OA pari a 0.89. Si può concludere quindi che lo sfruttamento di tutti e sei gli indici favorisce una migliore qualità della classificazione.

I risultati dei tre scenari sono riassunti nella tab. 2.

Table 1. Matrice di confusione generata dalla validazione dello scenario che prevede l'istruzione del classificatore utilizzando sia gli indici che le bande spettrali S-2 per la data del 17 maggio 2017.

		Reference Dataset						Classified totals	UA	Commission Error
		Artificial	Bare soil	Barley	Grassland	Woodland	Water			
Classified Data	Artificial	800	265	2	1	2	0	1080	0.7407	0.2592593
	Bare soil	234	5404	19	57	11	0	5825	0.9277	0.0722747
	Barley	2	23	206	67	1	0	299	0.689	0.310368
	Grassland	0	75	62	964	9	0	1110	0.8685	0.13535
	Woodland	0	27	22	69	131	0	149	0.9238	0.076782
Water	0	0	0	0	0	137	137	1	0	
Reference Totals		1036	5794	411	168	1464	137	10000		
PA		0.7722	0.932689	0.5012	0.83247	0.977459	1	OA	0.8942	
Omission Error		0.2278	0.067311	0.4988	0.16753	0.022541	0	Kappa	0.827789	

Table 2. Riassunto dei valori di accuratezza e coefficiente Kappa per i vari scenari presi in esame per la data del 17 maggio 2017.

Scenario	Accuratezza Totale	Coeff. Kappa
Bande	0.8491	0.736
Indici	0.7087	0.462
Bande + Indici	0.8942	0.827
Bande + NDVI	0.8907	0.821

Di seguito vengono presentate le mappe di verità a terra e le conseguenti mappe classificate di copertura del suolo ottenute con il terzo scenario per due date del 2017 relative al test site, 17 maggio (fig.2) e 6 luglio (fig. 3).

Le verità al suolo qui rappresentate includono sia i poligoni di calibrazione dell'algoritmo che di validazione.

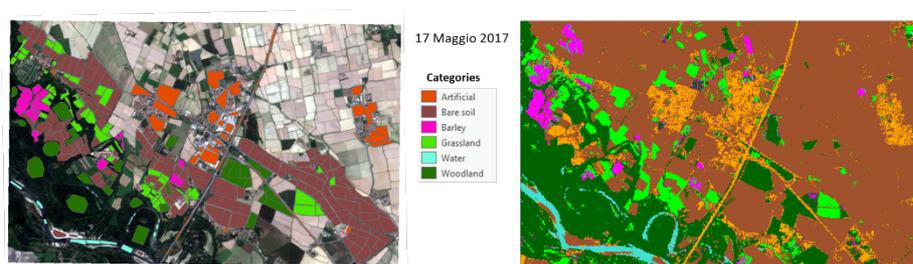


Fig. 2: A sinistra poligoni di verità al suolo per la data 17 maggio 2017. A destra la corrispondente mappa classificata.

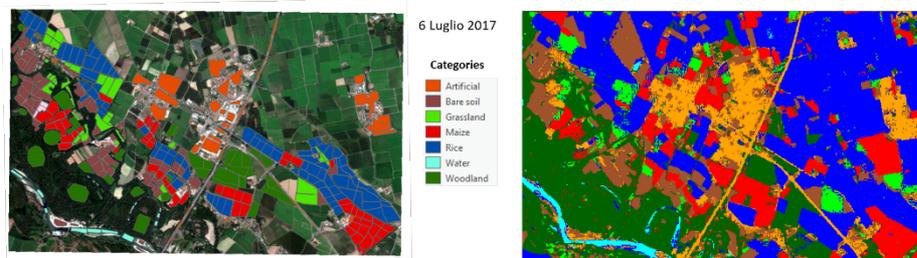


Fig. 3: A sinistra poligoni di verità al suolo per la data 6 luglio 2017. A destra la corrispondente mappa classificata.

Vengono inoltre riportate due mappe classificate che concernono i siti di studio del Progetto CE4WE relative all'anno 2020, alle date 21 maggio (fig.4) e 5 luglio (fig.5). Esse sono state classificate utilizzando i classificatori istruiti sul test site rispettivamente alle date 17 maggio e 6 luglio 2017.

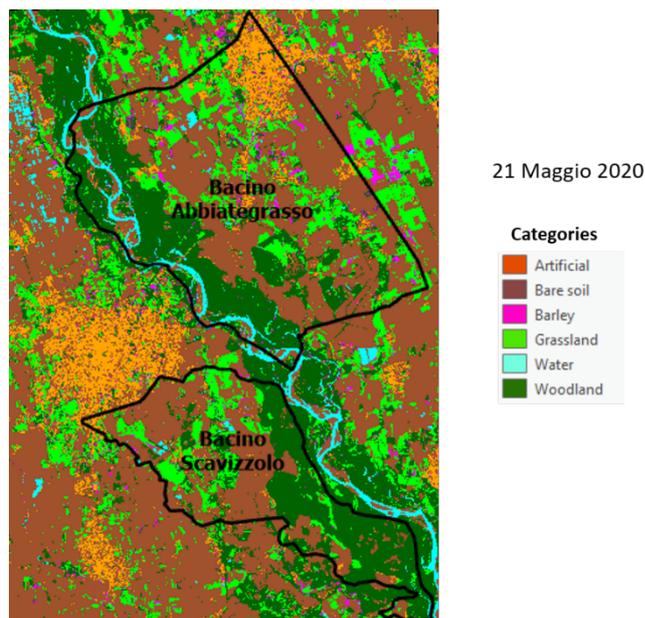


Fig. 4: Mappa classificata relativa al 21 maggio 2020, ottenuta utilizzando il classificatore istruito alla data 17 maggio 2017.

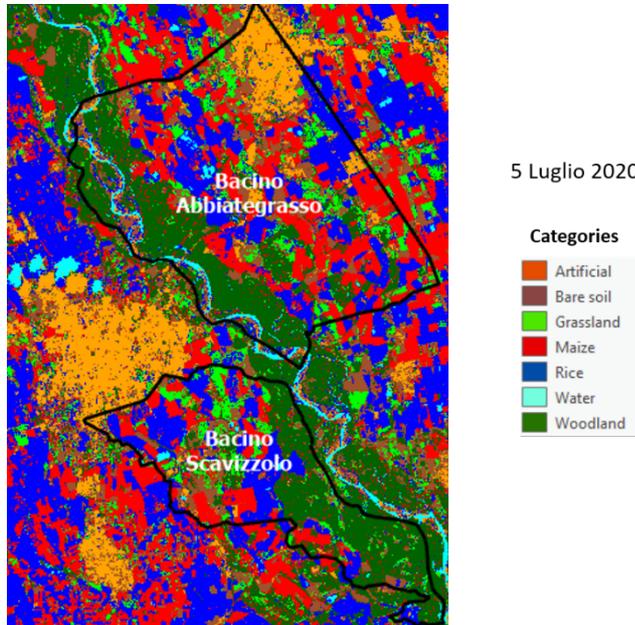


Fig. 5: Mappa classificata relativa al 5 luglio 2020, ottenuta utilizzando il classificatore istruito alla data 6 luglio 2017.

6 Conclusioni

Sono stati illustrati gli obiettivi del Laboratorio di Geomatica all'interno del Progetto CE4WE.

La classificazione supervisionata è stata effettuata, per ora, nella sola modalità a immagine singola sull'area test, di cui si conosce la verità a terra. Sono stati presentati i risultati per i tre scenari di istruzione considerati: uso delle sole bande originarie, previo ricampionamento; uso dei soli indici; uso delle bande, previo ricampionamento, e degli indici.

Esaminandoli, lo scenario che prevede una migliore qualità di classificazione risulta essere quello che prevede l'utilizzo delle bande congiuntamente agli indici.

I classificatori istruiti sull'area test sono stati impiegati per le due aree del progetto CE4WE.

7 Ulteriori sviluppi

Se sarà possibile stabilire buone relazioni con i gestori dei terreni delle aree del progetto CE4WE, verranno raccolte le informazioni relative alla verità a terra dei suddetti siti di studio e verrà creato un GIS multi-temporale analogo a quello costruito per il test-site.

Verrà certamente testato il Deep-Learning. La scelta delle bande da usare verrà affinata tenendo conto dei risultati contenuti nel presente articolo. Verrà infine adottato un metodo di classificazione multi-temporale: tutte le immagini del dataset saranno classificate in un unico processo.

Riferimenti bibliografici

1. URL: <http://osa.unipv.it/progetto-circular-economy-for-water-and-energy-ce4we>.
2. Douglas-Mankin, K.R., Srinivasan, R., Arnold, J.G. Soil and water assessment tool (SWAT) model: Current developments and applications. *Trans. ASABE* 53, 1423-1431 <https://doi.org/10.13031/2013.34915>, (2010)
3. URL ESA: <https://sentinel.esa.int/web/sentinel/missions/sentinel-2>
4. Annovazzi-Lodi, L., Franzini, M., Casella, V., Mapping Land Cover Types using Sentinel-2 Imagery: A Case Study. *Proceedings of GISTAM 2019*, ISBN: 978-989- 758-371-1. DOI:10.5220/0007738902420249, p. 242-249, (2019)
5. URL: <https://doc.arcgis.com/it/web-appbuilder/>
6. Rouse, J.W., Haas, R.H., Schell, J.A. and Deering, D.W., Monitoring vegetation systems in the Great Plains with ERTS. In 3rd ERTS Symposium, NASA SP-351 I, pp. 309–317, (1973).
7. Huete, A. R., A soil-adjusted vegetation index (SAVI), *Remote Sensing of Environment*, Volume 25, Issue 3, Pages 295-309, ISSN 0034-4257, [https://doi.org/10.1016/0034-4257\(88\)90106-X](https://doi.org/10.1016/0034-4257(88)90106-X), (1988)
8. Vogelmann, J. E., Rock, B. N., Moss, D. M., Red Edge spectral measurements from sugar maple leaves, *International Journal of Remote Sensing*, 14, 1563- 1575, <https://doi.org/10.1080/01431169308953986>, (1993)
9. Gamon, J., Serrano, L. and Surfus, J., The photochemical reflectance index: an optical indicator of photosynthetic radiation use efficiency across species, functional types, and nutrient levels, *Oecologia* 112, 492-501, (1997)
10. Sriwongsitanon, N. et al., The Normalized Difference Infrared Index (NDII) as a proxy for soil moisture storage in hydrological modelling, *Hydrology and Earth System Sciences Discussions* 12(8):8419-8457, (2015)
11. Zha, Y., Gao, J., and Ni, S., Use of normalized difference built-up index in automatically mapping urban areas from TM imagery, *Int. J. Remote Sens.* 24, 583–594, (2003)
12. URL: <https://www.hexagongeospatial.com/products/power-portfolio/erdas-imagine>
13. Taskin Kaya, G., Musaoglu, N., Ersoy, O. K., Damage assessment of 2010 Haiti earthquake with post-earthquake satellite image by support vector selection and adaptation. *Photogrammetric Engineering & Remote Sensing*, 77(10), 1025-1035 (2011).
14. Vapnik, V., Estimation of Dependences Based on Empirical Data. Nauka, Moscow, pp. 5165–5184, 27 (in Russian) (English translation: Springer Verlag, New York, 1982), (1982)