

Classificazione automatica del grado di mescolanza della copertura forestale con dati multispettrali Sentinel-2 e software open-source

Federica Murgia, Antonio Floris, Fabrizio Clementel, Sandro Zanotelli, Patrizia Gasparini
Consiglio per la Ricerca in agricoltura e l'Analisi dell'Economia agraria (CREA) – Centro di
ricerca Foreste e Legno, Trento (IT)
federica.murgia@crea.gov.it
antonio.floris@crea.gov.it
fabrizio.clementel@crea.gov.it
sandro.zanotelli@crea.gov.it
patrizia.gasparini@crea.gov.it

Riassunto

Abstract. Nei moderni inventari forestali, al fine di migliorare l'efficienza del campionamento, ottenere un monitoraggio più frequente del territorio e integrare le statistiche con prodotti di tipo cartografico, i rilievi al suolo vengono affiancati da analisi ed elaborazioni di dati telerilevati. Una informazione fondamentale legata alla copertura forestale riguarda il grado di mescolanza delle aree classificate come bosco, ossia la distinzione tra bosco puro di conifere, bosco puro di latifoglie, bosco misto di conifere e latifoglie. Le immagini satellitari sono considerate tra le fonti di dati più importanti per la mappatura dell'uso e copertura del suolo, grazie alla loro ampia copertura geografica ed a tempi di rivisita ridotti. Il presente contributo si propone l'obiettivo di presentare modalità ed esiti della classificazione del grado di mescolanza della copertura forestale mediante l'utilizzo di dati multispettrali Sentinel-2 nel territorio delle province di Trento e Bolzano (13605 km²); le immagini utilizzate sono riferite al 2018 e precedenti alla tempesta Vaia, che ha interessato l'area di studio nell'ottobre dello stesso anno. Come metodo di classificazione è stato usato Random Forest e per la feature selection è stato testato l'algoritmo Boruta che, in base alla letteratura reperita, non risulta essere stato mai applicato in ambito forestale. È stata operata anche una validazione esterna indipendente, su un territorio con caratteristiche forestali analoghe a quelle delle aree di addestramento rappresentato dalle province di Belluno e Vicenza. Lo studio è stato svolto utilizzando esclusivamente software Open Source (SNAP, QGIS, R). L'accuratezza complessiva valutata sul dataset di test è risultata pari a 82.7%, con un valore di kappa pari a 0.74 e un F1 medio pari a 0.80. La validazione esterna ha portato a un'accuratezza complessiva pari a 74.1%, con valori di kappa pari a 0.65 e un F1 medio pari a 0.72.

Parole chiave

Sentinel-2, inventari forestali, tipi forestali, Random Forest

1. Introduzione

Negli ultimi decenni si è assistito a un indubbio aumento della richiesta di mappe di uso e copertura del suolo così come di mappe dei tipi forestali. Un tipo forestale è una distinta associazione di specie arboree distribuite su un ampio range geografico, ed è generalmente associato con la specie arborea dominante caratteristica del tipo stesso, sebbene altre specie spesso si manifestino in misura minore. L'informazione sulla copertura forestale e sulle sue caratteristiche è essenziale per il monitoraggio delle risorse forestali. Una prima classificazione che viene riportata nelle attività di reporting nazionale e internazionale sulle foreste e in prodotti cartografici di carattere generale riguarda la ripartizione della copertura forestale in formazioni a prevalenza di conifere, di latifoglie o miste.

Al terzo livello della classificazione gerarchica della copertura del suolo, la più recente mappa Corine Land Cover [1] fornisce la distribuzione delle foreste europee ripartite in: a prevalenza di latifoglie, decidue o sempreverdi, a prevalenza di conifere e foreste miste (dove sia le conifere sia le latifoglie siano presenti in misura almeno pari al 25%, secondo la definizione FAO [2], a una risoluzione spaziale di 100 m e con un'unità minima di cartografabile pari a 25 ettari [3]. Questa mappa è caratterizzata da una risoluzione grossolana, se confrontata con le immagini Sentinel-2, e dalla presenza di cambiamenti di copertura del suolo mancanti rispetto alla mappa CLC del 2012 [4]. Inoltre, la definizione di foresta non è pienamente coerente con quella della FAO, poiché la copertura minima richiesta è del 30% (o 500 soggetti per ha) invece che del 10%.

Il servizio Copernicus High Resolution Layers (HRL,[5]), relativo alle caratteristiche della copertura del suolo ottenuto a partire dai dati Sentinel 2 per l'anno di riferimento 2018, include una mappa della copertura forestale per gruppo di specie prevalente (foresta di conifere e latifoglie) ad una risoluzione di 10 m coerente con la definizione di foresta della FAO. Il servizio fornisce inoltre una mappa del grado di mescolanza aggregato ad una risoluzione di 100 m (foresta di conifere, latifoglie e mista) che copre 39 Paesi europei.

Una mappa a scala europea, con risoluzione di 10 m, delle foreste di conifere e latifoglie per l'anno 2017 è stata recentemente derivata dai dati Sentinel-1 e confrontata con i dataset forestali di Copernicus HRLs e con i dati a scala nazionale (mappe forestali, dataset NFI e dati provenienti dalle università) per dodici Paesi [6]. Oltre alle mappe a scala europea, alcuni Paesi europei hanno prodotto mappe del grado di mescolanza di conifere e latifoglie per i rispettivi territori (esempi in [7], [8], [9]). La Carta Forestale d'Italia (CFI2020) in scala 1:10.000 recentemente diffusa in un formato non ancora definitivo (<https://sinfor.sian.it/#/carta>) prevede tre diversi strati informativi relativi alla composizione specifica (tipi forestali della classificazione europea, della classificazione dell'Inventario Forestale Nazionale INFC e della CFI2020 ottenuta per armonizzazione di tipologie regionali e locali) senza un livello di aggregazione superiore in tipi a prevalenza di conifere, di latifoglie o misti.

La classificazione della vegetazione è uno degli scopi principali dell'impiego dei dati telerilevati in ambito forestale [10], [11]. La classificazione della vegetazione da

immagini telerilevate si basa prevalentemente sulle caratteristiche spettrali del segnale osservato [11]. In questo ambito, la missione Sentinel-2 del programma Copernicus dell’Agenzia Spaziale Europea (ESA) offre ampie prospettive nello studio del territorio e della vegetazione. Il satellite Sentinel-2 (S2) trasporta un sensore multispettrale (MSI) a 13 bande operanti nelle regioni del visibile e dell’infrarosso. S2 acquisisce informazioni sulla superficie terrestre con ampia copertura (ampiezza di swath di 290 km) e un tempo di rivisita di cinque giorni, ottenuto combinando i passaggi di due satelliti gemelli S2-A e S2-B. In ambito internazionale, un ampio panorama di applicazioni forestali a scala locale è stato studiato e realizzato con i dati S2, quali: la quantificazione della biomassa epigea-AGB [12] e del Leaf Area Index- LAI su superfici di piccola estensione [13], la stima del volume legnoso [14], la classificazione del grado di mescolanza (boschi di conifere, boschi di latifoglie, boschi misti) [7], la classificazione della copertura del suolo [15] e, la classificazione della composizione in specie arboree [15], [16].

Molte delle applicazioni finalizzate alla classificazione della vegetazione forestale si basano sull’uso di Random Forest (RF) [17]; si tratta di un algoritmo di regressione e classificazione supervisionato e non parametrico ampiamente usato con dati da remote sensing, in particolare nella copertura del suolo e nella caratterizzazione delle foreste [7], [12], [16], [18], [19], [20], [21]. RF è spesso impiegato con dati iperspettrali perché ha mostrato di essere poco sensibile al cosiddetto fenomeno della “curse of dimensionality”, che si verifica quando il numero dei campioni (osservazioni) di addestramento per ciascuna classe è significativamente minore della dimensione del “feature space” (nello specifico, numero di bande spettrali) [17]. RF è stato usato con successo anche con dati LiDAR e multispettrali. L’uso di metriche per la feature selection prodotte dall’algoritmo (mean decrease Gini importance e la mean decrease accuracy), si è accresciuto negli ultimi anni. Ciò nonostante, scarso interesse è stato posto nella selezione delle metriche più appropriate da applicare ai dati telerilevati. Infatti, le misure di importanza delle variabili fornite da RF sono poco adatte a situazioni in cui i potenziali predittori adottano scale di misura o numero di categorie diverse fra loro [22].

Il presente lavoro illustra i risultati della classificazione automatica del grado di mescolanza di conifere e latifoglie ottenuta da dati Sentinel-2 per i territori delle province di Trento e Bolzano mediante l’utilizzo del classificatore RF e dell’algoritmo di feature selection Boruta [23], applicato al set di variabili predittive insieme a un’analisi di cross-correlazione. L’analisi e l’elaborazione dei dati sono state realizzate esclusivamente attraverso l’uso di software open-source (SNAP – [24], R – [25] e QGIS – [26]) ottenendo una mappa ad elevata risoluzione dei boschi di conifere, di latifoglie e misti per l’area di studio. La classificazione è stata validata in un’area territoriale esterna, avente caratteristiche forestali analoghe e superficie comparabile con quelle dell’area coinvolta nella messa a punto del modello.

2. Materiali e metodi

2.1 Area studio

Lo studio è stato condotto in Trentino-Alto Adige per quanto riguarda l'addestramento e il test del modello di classificazione, e in due province del Veneto (Belluno e Vicenza) per la sua validazione esterna (Figura 1). Si tratta di un territorio prevalentemente montuoso, situato nell'area alpina, caratterizzato da una vegetazione forestale in cui prevalgono l'abete rosso, il faggio, il larice e il pino silvestre. L'estensione complessiva dell'area studio è di 19900 km², con un indice di boscosità complessivo del 56.7% [27].

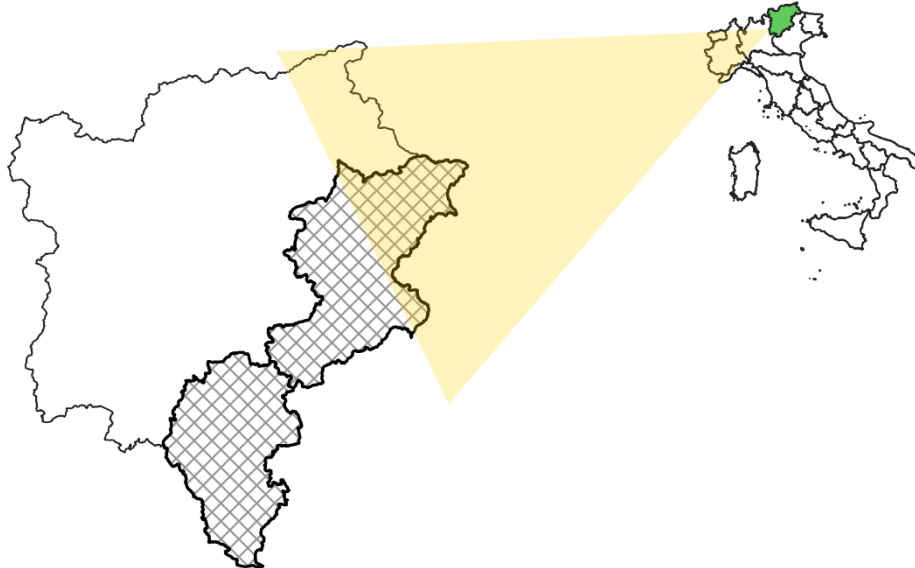


Fig. 1 – Area di studio per l'addestramento e test del modello di classificazione, corrispondente al territorio delle province di Trento e Bolzano: l'area retinata a destra, coincidente con le province di Belluno e Vicenza, è stata usata per la validazione esterna dell'algoritmo di classificazione.

2.2 Dati di verità a terra

I dati relativi alla classificazione della copertura forestale in a prevalenza di conifere, di latifoglie e mista di conifere e latifoglie secondo le definizioni FAO [2] provengono in parte dai rilievi al suolo di seconda fase realizzati per il secondo e il terzo inventario forestale nazionale italiano INFC (www.inventarioforestale.org) e in parte dalla

fotointerpretazione ex-novo di altri punti di verità, circa 1560, condotta su ortofoto da servizi di mappa (WMS) Google Satellite e Bing, per un totale di 8671 punti di addestramento e test del modello di classificazione e di 978 punti di validazione esterna. Il gap temporale intercorso tra il rilievo della verità a terra (rilievo al suolo o epoca di ripresa dell'ortofoto) e l'acquisizione del dato S2 è tale da non comportare significativi cambiamenti nella mescolanza, che non ha una dinamica temporale rapida. Nell'INFC l'area in cui valutare questa caratteristica, tramite rilievo al suolo, è individuata in un intorno circolare con raggio di 25 m denominato AdS25 centrato sul punto di campionamento [28]. La scelta di usare ortofoto Google satellite e Bing per la fotointerpretazione dei punti di verità aggiuntivi è motivata dalla disponibilità pubblica di questo tipo di immagini per tutti gli utenti, sebbene la loro qualità e risoluzione possa essere inferiore a quella di altre ortofoto. L'unità considerata nella fotointerpretazione o fotoplot è rappresentata, coerentemente con INFC, da un reticolo di celle quadrate di lato 50 m corrispondente a 5x5 pixel S2 e con essi spazialmente sincro (Figura 2). La fotointerpretazione è stata eseguita da quattro fotointerpreti esperti e preceduta da una prima fase di armonizzazione, con alcune classificazioni eseguite in panel e blind training su 100 punti (classificazione individuale sugli stessi punti senza conoscere le classificazioni altrui e successiva discussione delle discordanze). La tabella 1 riporta la consistenza totale del campione studiato, ripartita nelle classi osservate.

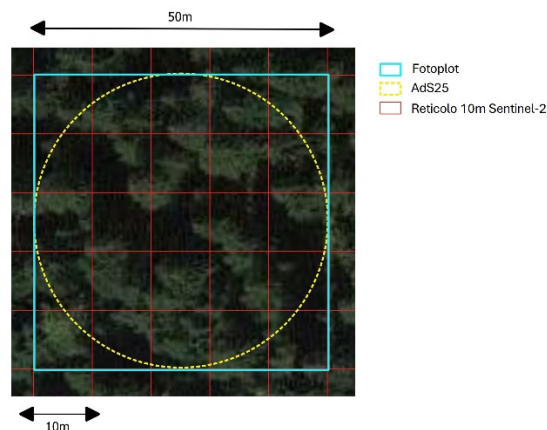


Fig. 2. Rappresentazione dell'unità di fotointerpretazione in relazione al reticolo di pixel Sentinel-2 e all'unità di rilievo al suolo AdS25.

Tabella 1. Consistenza del campione usato nello studio per lo sviluppo (training) e per la validazione interna (test) ed esterna (validation) del modello e sua ripartizione nelle classi considerate.

	Totale	Training set	Test set	Validation set
Boschi a prevalenza di conifere	4341	3084	1257	243

Boschi a prevalenza di latifoglie	1377	959	418	266
Boschi misti	1925	1328	597	198
Non bosco	1028	734	294	265

2.3 Dati Sentinel 2

Dal Open Access Hub del portale copernicus.eu [29] sono stati scaricati i frame (granules) ricoprenti l'area studio per un set multistagionale di epoche ritenute utili alla messa a punto degli algoritmi di classificazione. Nello specifico, sono state selezionate le immagini acquisite in epoche corrispondenti mediamente alle diverse fasi vegetative delle specie arboree che popolano l'area di studio (Tabella 2). I dati scaricati, caratterizzati da una copertura nuvolosa inferiore al 20% dell'intera immagine, sono al livello 2A (Bottom of Atmosphere – BOA) e relativi a 11 bande spettrali. Sono state infatti escluse dall'analisi due bande a 60 m di risoluzione, la banda 1 correlata all'aerosol costiero e la banda 9 relativa al vapore acqueo. Ogni frame ha dimensioni di 100 x 100 km e può essere ottenuto nel sistema di riferimento planimetrico WGS84-UTM, zona 32N nel nostro caso. Per mantenere la necessaria coerenza di classificazione con i dati INFC2015, sono state selezionate immagini relative all'anno 2018 e precedenti la tempesta Vaia che, come noto, ha prodotto ingentissimi danni alle foreste dell'area interessata [30]. Tutte le bande S2 sono state ricampionate a 10 m di risoluzione, mosaicate e ritagliate sulla regione di interesse (ROI) con i tool della Sentinel Application Platform (SNAP).

Tabella 2. Elenco dei granuli Sentinel-2 scaricati dal Open Access Hub del portale Copernicus. Tutti i file sono in formato .safe.

Inverno 2018	S2A_MSIL2A_20180119T101331_N0206_R022_T32TPR_20180119T135441
	S2A_MSIL2A_20180129T101251_N0206_R022_T32TPS_20180129T135502
	S2A_MSIL2A_20180129T101251_N0206_R022_T32TQS_20180129T135502
Primave ra 2018	S2A_MSIL2A_20180419T101031_N0207_R022_T32TPS_20180419T111252
	S2A_MSIL2A_20180419T101031_N0207_R022_T32TPR_20180419T111252
	S2B_MSIL2A_20180414T101019_N0207_R022_T32TQS_20180414T122253
Estate 2018	S2A_MSIL2A_20180718T101031_N0208_R022_T32TPR_20180718T180146
	S2B_MSIL2A_20180713T101029_N0208_R022_T32TPS_20180713T144147
	S2B_MSIL2A_20180723T101019_N0208_R022_T32TPR_20180723T161930
Autunno 2018	S2A_MSIL2A_20181019T102031_N0209_R065_T32TPR_20181019T132325
	S2A_MSIL2A_20181019T102031_N0209_R065_T32TPS_20181019T132325
	S2B_MSIL2A_20181021T101039_N0209_R022_T32TPS_20181021T151822

In corrispondenza della finestra di analisi (fotoplot) 50x50 m costruita attorno a ciascun punto del set di addestramento e validazione sono stati estratti i vettori dei

valori di riflettanza (media aritmetica dei 25 pixel S2 contenuti in ciascun fotoplot) per le 11 bande spettrali utilizzate, e la quota, nonché calcolati alcuni indici di vegetazione. Dei numerosi indici di vegetazione presenti in letteratura si è scelto di utilizzare un sottoinsieme rappresentativo e non ridondante in termini di bande spettrali coinvolte, di indici: Red-Edge Normalized Difference Vegetation index (RENDVI), Enhanced Vegetation Index (EVI) e Ratio Vegetation Index (RVI). Nello specifico RENDVI ed EVI contribuiscono ad indicare la presenza di vegetazione sana, senza problemi di saturazione legati a elevati valori di biomassa, mentre l'indice RVI mitiga gli effetti indesiderati legati all'orografia e all'atmosfera, specialmente in aree montuose.

L'analisi spaziale del dato è stata interamente svolta in QGIS (v. 3.28.3), mentre il calcolo degli indici è stato effettuato in R (v.4.4.0).

2.2.1 Algoritmi di classificazione e di estrazione delle variabili

I dati di verità a terra sono stati suddivisi in modo casuale in set di addestramento e validazione interna, utilizzando rispettivamente il 70% e il 30% dei dati totali disponibili. Quale algoritmo di classificazione è stato utilizzato RF, un algoritmo di classificazione supervisionato non parametrico costituito da un insieme di alberi decisionali, in grado di ridurre la varianza del classificatore e al contempo contenere il bias di quest'ultimo. La determinazione della divisione in ogni nodo si basa sul criterio di Gini. Per ciascun nodo viene selezionato casualmente un sottoinsieme di variabili e da questo sottoinsieme (il campione bootstrap) viene scelta la variabile che permette la migliore suddivisione. In questo modo, la correlazione tra gli alberi diminuisce e conseguentemente anche l'errore di generalizzazione. Infine, la classificazione dei nuovi dati si ottiene prendendo il voto di maggioranza tra i risultati della classificazione raccolti da tutti gli alberi decisionali. L'errore di classificazione viene stimato utilizzando, ad ogni iterazione bootstrap, tutti i campioni che non fanno parte del suddetto campione, ossia il dataset out-of-bag (OOB). Ogni albero decisionale viene utilizzato per classificare i campioni del corrispondente set di dati OOB. A questo punto, per ogni campione del set di dati originale, il voto di maggioranza tra i risultati della classificazione degli alberi decisionali coinvolti viene confrontato con l'etichetta della classe vera, fornendo una stima dell'errore di classificazione [17]. Considerato l'obiettivo di sviluppare un algoritmo di classificazione multi-temporale, sono state considerate tutte le bande spettrali di Sentinel-2 (11 bande spettrali per ogni stagione, quindi complessivamente 44 bande) e 3 indici di vegetazione per ogni fase vegetativa o stagione. Come variabile esplicativa aggiuntiva è stata inoltre considerata la quota di ciascun punto di verità a terra. In totale sono quindi state utilizzate 57 variabili predittive.

Per l'implementazione del modello è richiesta l'impostazione di due parametri, il numero di alberi da generare k ed il numero di variabili esplicative m da utilizzare ad ogni nodo. Generalmente, nel caso di utilizzo di RF per classificare, si imposta m pari a $\lceil \sqrt{p} \rceil$, essendo p il numero di variabili esplicative. Nel caso specifico è stato impostato k pari a 500 e m pari ad 8. Il modello fornisce due metriche per valutare l'importanza delle variabili esplicative, la diminuzione media dell'accuratezza (MDA) e la

diminuzione media di Gini (MDG). La MDA quantifica la perdita di accuratezza subita dal modello escludendo iterativamente ciascuna variabile. Più l'accuratezza ne risente, più la variabile è importante per una migliore classificazione. La MDG è una misura di come ogni variabile contribuisce all'omogeneità dei nodi di ciascun albero decisionale. Al fine di testare un algoritmo di feature selection esterno al modello, le variabili esplicative utilizzate sono state analizzate mediante l'algoritmo di selezione delle variabili denominato Boruta [23] che non risulta sia stato applicato prima nell'ambito della classificazione di coperture forestali. Boruta valuta l'importanza dell'insieme di variabili predittive confrontando ciascun Z-score ad esse associato (ossia il rapporto tra la perdita media e la sua deviazione standard), con il migliore di un insieme di "variabili-sonda" fittizie. Solo le variabili che superano il test possono essere considerate importanti. Il risultato del test ha considerato tutte le variabili come importanti ai fini della classificazione. Si è deciso quindi di fare una scelta conservativa e, pur essendo a conoscenza della correlazione tra alcune bande contigue, di non escludere alcuna variabile esplicativa.

3. Risultati

Come si può notare dalla matrice di confusione in Tabella 3, la maggior parte degli errori riguardano la classe dei boschi misti classificati come conifere e la classe dei boschi di latifoglie classificati come boschi misti. Alcune combinazioni di errata classificazione sono considerate più lievi rispetto ad altre, in particolare quelle che riguardano la confusione tra boschi misti e boschi puri di uno dei due gruppi di specie.

La performance dell'algoritmo di classificazione è stata valutata, sul test set, tramite matrice di confusione fra valori osservati e predetti e calcolo dei principali indici di accuratezza. La sensitivity risulta molto più elevata per le i boschi a prevalenza di conifere rispetto a quelli a prevalenza di latifoglie e ai boschi misti. Ciò potrebbe essere spiegato dalla tendenza delle conifere a formare, nell'area di studio, popolamenti puri con scarsa presenza di latifoglie, in particolare nel piano dominante (che è quello maggiormente intercettato dai dati spettrali).

Tabella 3. Risultati della validazione interna: matrice di confusione (in alto) e indici di accuratezza (in basso).

		Classificazione da dati di verità a terra			
		Conifere	Latifoglie	Misto	Non bosco
Classificazione da modello	Conifere	1161	11	103	7
	Latifoglie	6	289	93	12
	Misto	69	106	397	2
	Non bosco	21	12	4	273

	Conifere	Latifoglie	Misto	Non bosco
PA - Sensitivity	0,92	0,69	0,66	0,93
UA - Precision	0,91	0,72	0,69	0,88
F1-Score	0,91	0,71	0,68	0,90
OA - Overall Accuracy	0,83			

Tabella 4. Risultati validazione esterna: matrice di confusione (in alto) e indici di accuratezza (in basso).

Classificazione da dati di verità a terra					
	Conifere	Latifoglie	Misti	Non bosco	
Classificazione da modello	Conifere	229	11	79	24
	Latifoglie	1	180	10	28
	Misti	11	62	109	4
	Non bosco	2	13	0	209

	Conifere	Latifoglie	Misti	Non bosco
PA - Sensitivity	0,94	0,68	0,55	0,79
UA - Precision	0,67	0,82	0,59	0,93
F1-Score	0,78	0,74	0,57	0,85
OA - Overall Accuracy	0,75			

Nel set di validazione esterna sono da segnalare due tipi di errata classificazione di una certa rilevanza: boschi misti classificati come a prevalenza di conifere (79 casi) e boschi a prevalenza di latifoglie classificati come boschi misti (62 casi); entrambi sono da ascrivere al gruppo degli errori lievi, per i motivi già esposti.

Discussione e conclusioni

I dati satellitari multispettrali ad alta risoluzione stanno dimostrando di essere un insostituibile ausilio alla produzione di modelli di classificazione delle foreste e del loro stato di salute, oltre che di stima di alcune importanti variabili quantitative come il volume legnoso, la biomassa e il carbonio trattenuto. L'elevata risoluzione temporale garantita da alcune piattaforme, intesa come intervallo di rivisita ristretto (ravvicinato), è non solo importante per cogliere quei fenomeni caratterizzati da una rapida dinamica temporale (si pensi a determinati attacchi parassitari), ma permette di ridurre, attraverso processi di fusione di immagini acquisite in momenti diversi, gli effetti negativi della copertura nuvolosa e delle ombre, particolarmente rilevanti nelle aree settentrionali e montuose del nostro Paese.

I dati forniti dalla missione Sentinel-2 del programma ESA-Copernicus costituiscono un ottimo esempio al riguardo: vengono rilasciati pubblicamente e gratuitamente, possono essere pre-elaborati grazie a tool in continuo sviluppo all'interno del programma stesso e beneficiano della costante produzione di procedure, algoritmi e modelli di processing e analisi prodotti da una vasta comunità di utenti di alto livello appartenenti al settore R&S.

Questo lavoro ha presentato le modalità di produzione di mappe della composizione delle foreste a livello di gruppi specifici (conifere, latifoglie, boschi misti), in accordo con le definizioni inventariali adottate nel nostro e in altri Paesi, a partire da set multitemporali di immagini Sentinel-2 e da informazioni acquisite con rilievi al suolo e da fotointerpretazione, entrambe attività tipiche degli inventari forestali.

I principali aspetti caratterizzanti le procedure proposte sono stati: l'uso di dati telerilevati pubblici e di software Open Source; il ricorso a tecniche di feature selection in grado di mitigare gli effetti dell'autocorrelazione fra descrittori e lo sbilanciamento del peso con cui essi agiscono; la validazione esterna della mappa prodotta, con dati e in un territorio non coinvolto nell'addestramento degli algoritmi.

I risultati, in termini di accuratezza di classificazione, sono stati piuttosto buoni e indicano la possibilità che un prodotto di questo tipo, senza voler sostituire i rilievi per campionamento tipici delle tradizionali procedure inventariali in ambito forestale, possa essere molto utile per applicazioni di pre- e post- stratificazione e per estendere a un contesto continuo, accettandone i limiti, classificazioni ottenute da osservazioni puntuali. Anche la validazione esterna ha avuto esito complessivamente positivo, con risultati che non si discostano particolarmente da quelli della validazione interna.

La perdita di risoluzione spaziale dovuta a un'unità di classificazione di 50x50 m invece della più elevata risoluzione nativa S2 (10 m) è solo apparente, in quanto a) non avrebbe senso valutare la composizione in specie forestali in un intorno troppo ristretto in cui potrebbero ricadere anche solo 1-2 chiome arboree; b) in questo modo si compensa l'incertezza spaziale delle immagini telerilevate e la presenza di pixel non nettamente attribuibili ad una sola classe.

Ringraziamenti

Lo studio è stato realizzato nell'ambito della borsa di studio sul tema "Elaborazione di dati telerilevati per applicazioni nello studio di coperture forestali" finanziata dall'accordo di collaborazione IFNI3-RIL2&3 tra il Comando Unità Forestali, Ambientali e Agroalimentari Carabinieri e il Consiglio per la ricerca in agricoltura e l'analisi dell'economia agraria per attività di supporto alla progettazione, realizzazione e gestione del terzo Inventario Nazionale delle Foreste e dei serbatoi forestali di Carbonio (INFC2015).

Riferimenti bibliografici

1. Corine Land Cover (CLC, <https://land.copernicus.eu/pan-european/corine-land-cover>) ultimo accesso 2024/08/29.
2. FAO Land use/forest type classification: <https://www.fao.org/3/ae578e/AE578E03.htm>. Ultimo accesso 2024/09/25.
3. Kosztra, B., Büttner, G., Hazeu, G., & Arnold, S.: Updated CLC illustrated nomenclature guidelines. European Environment Agency: Wien, Austria, 1-124 (2017).
4. Paris, C., Gasparella, L., Bruzzone, L.: A scalable high-performance unsupervised system for producing large-scale hr land cover maps: The italian country case study. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 15, 9146-9159 (2022).
5. CLMS, <https://land.copernicus.eu/en> ultimo accesso 2024/08/29.
6. Dostálová, A., Lang, M., Ivanovs, J., Waser, L. T., Wagner, W.: European wide forest classification based on sentinel-1 data. *Remote Sensing*, 13(3), 337 (2021).
7. Waser, L., Rütetschi, M., Psomas, A., Small, D. and Rehush, N.: Mapping dominant leaf type based on combined Sentinel-1/-2 data - Challenges for mountainous countries. *ISPRS Journal of Photogrammetry and Remote Sensing*. 180. 209-226 (2021).
8. Waser L. T., Ginzler C., Rehush N.: Wall-to-Wall Tree Type Mapping from Countrywide Airborne Remote Sensing Surveys. *Remote Sens.*, 9, 766 (2017).
9. Lang M., Kaha M., Laarmann D., Sims A.: Construction of tree species composition map of Estonia using multispectral satellite images, soil map and a random forest algorithm. *Forestry Studies*, Vol. 68, Pages 5–24 (2018).
10. Korpela I., Individual tree measurements by means of digital aerial photogrammetry, *Silva Fennica Monographs*, vol. 32004 (2004).
11. Heikkinen, V., Tokola, T., Parkkinen, J., Korpela, I., Jämskäinen, T.: Simulated multispectral imagery for tree species classification using support vector machines., *IEEE Trans. Geosci. Remote Sens.* 48, 1355–1364. <https://doi.org/10.1109/TGRS.2009.2032239> (2010).
12. Luo, K., Wei, Y., Du, J., Liu, L., Luo, X., Shi, Y., Tang, X.: Machine learning-based estimates of aboveground biomass of subalpine forests using Landsat 8 OLI and Sentinel-2B images in the Jiuzhaigou National Nature Reserve, Eastern Tibet Plateau. *Journal of Forestry Research*, 1-12 (2022).
13. Clevers, J. G., Kooistra, L., Van den Brande, M. M.: Using Sentinel-2 data for retrieving LAI and leaf and canopy chlorophyll content of a potato crop. *Remote Sensing*, 9(5), 405 (2017).
14. Chrysafis, I., Mallinis, G., Tsakiri, M., & Patias, P.: Evaluation of single-date and multi-seasonal spatial and spectral information of Sentinel-2 imagery to assess growing stock volume of a Mediterranean forest. *International Journal of Applied Earth Observation and Geoinformation*, 77, 1-14 (2019).
15. Immitzer, M., Vuolo, F., Atzberger, C.: First Experience with Sentinel-2 Data for Crop and Tree Species Classifications in Central Europe, *Remote Sensing*, 8, (3), (2016).
16. Puletti, N., Chianucci, F., & Castaldi, C.: Use of Sentinel-2 for forest classification in Mediterranean environments. *Annals of Silvicultural Research*, 42, 32-38 (2018).
17. Breiman, L.: Random Forests. *Machine Learning*, 45, 5-35 (2001).
18. Immitzer, M., Atzberger, C., Koukal, T.: Tree species classification with random forest using very high spatial resolution 8-band WorldView-2 satellite data. *Remote sensing*, 4(9), 2661-2693, (2012).
19. Rodriguez-Galiano, V. F., Chica-Olmo, M., Abarca-Hernandez, F., Atkinson, P. M., & Jeganathan, C. J. R. S. E.: Random Forest classification of Mediterranean land cover using

- multi-seasonal imagery and multi-seasonal texture, *Remote Sensing of Environment*, 121, 93-107 (2012).
20. Fassnacht, F. E., Latifi, H., Stereńczak, K., Modzelewska, A., Lefsky, M., Waser, L. T., ... & Ghosh, A.: Review of studies on tree species classification from remotely sensed data. *Remote sensing of environment*, 186, 64-87 (2016).
 21. Thanh Noi, P., Kappas, M.: Comparison of random forest, k-nearest neighbor, and support vector machine classifiers for land cover classification using Sentinel-2 imagery. *Sensors*, 18(1), 18 (2017).
 22. Strobl, C., Boulesteix, AL., Zeileis, A. et al.: Bias in random forest variable importance measures: Illustrations, sources and a solution. *BMC Bioinformatics* 8, 25 <https://doi.org/10.1186/1471-2105-8-25> (2007).
 23. Kursa, M. B., Rudnicki, W. R.: Feature Selection with the Boruta Package. *Journal of Statistical Software*, 36(11), 1–13 (2010).
 24. SNAP - ESA Sentinel Application Platform v10.0.0
 25. R Core Team R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL <https://www.R-project.org/> (2021).
 26. QGIS.org, QGIS Geographic Information System. QGIS Association. <http://www.qgis.org>
 27. Gasparini, P., Di Cosmo, L., Floris, A.: Area and characteristics of Italian forests. In P. Gasparini, L. Di Cosmo, A. Floris, D. De Laurentis (Eds.), *Italian National Forest Inventory—Methods and results of the third survey*. Springer Tracts in Civil Engineering. Springer. https://doi.org/10.1007/978-3-030-98678-0_7 (2022).
 28. Floris A., Di Cosmo L., Rizzo M., Patrone A.: Field assessment – Survey Protocols and Data Collection. In: Gasparini, P., Di Cosmo, L., Floris, A., De Laurentis, D. (eds) *Italian National Forest Inventory—Methods and Results of the Third Survey*. Springer Tracts in Civil Engineering. Springer, Cham, pp. 67-101. https://doi.org/10.1007/978-3-030-98678-0_2 (2022).
 29. ESA Open Access Hub – Copernicus, <https://dataspace.copernicus.eu/> ultimo accesso il 2024/09/24.
 30. Giovannini, L., Davolio, S., Zaramella, M., Zardi, D., Borga, M.: Multi-model convection-resolving simulations of the October 2018 Vaia storm over Northeastern Italy. *Atmospheric Research*, 253, 105455 (2021).