

Classificazione delle coperture forestali con immagini satellitari multispettrali e dati *laser scanner* aereo

Francesca Bottalico, Davide Travaglini

GeoLAB – Laboratorio di Geomatica. Dipartimento di Gestione dei Sistemi Agrari, Alimentari e Forestali, Università degli Studi di Firenze. Via San Bonaventura, 13 - 50145 Firenze, Italia. Tel. 055 2755656, fax 055 319179, e-mail: davide.travaglini@unifi.it

Riassunto

In questo studio è stata valutata la potenzialità del telerilevamento satellitare multispettrale e *laser scanner* aereo per classificare l'uso del suolo, i tipi forestali e i tipi strutturali in un'area di studio dell'Italia centrale. La procedura applicata prevede l'utilizzo di tecniche di segmentazione delle immagini telerilevate e l'impiego del metodo di classificazione *Multinomial Logistic Regression*. L'accuratezza complessiva della classificazione, valutata su un campione di aree *test*, è risultata dell'81% per l'uso del suolo, del 57% per i tipi forestali e del 71% per i tipi strutturali.

Abstract

In this study multispectral satellite data and aerial laser scanning data have been tested to classify land use, forest types and forest structure in a study area in central Italy. The method is based on a object-oriented approach and the Multinomial Logistic Regression classifier. The overall accuracy of the classification computed on a sample of test site was 81% for land use, 57% for forest types and 71% for forest structure.

Introduzione

Il telerilevamento laser scanner aereo (*Airborne Laser Scanning* - ALS) con strumentazione LiDAR è in grado di fornire informazioni di dettaglio sulla struttura delle coperture forestali. Vari studi hanno evidenziato le potenzialità dei dati ALS per la stima di variabili quantitative di interesse inventariale e assestamentale.

In questo studio si presentano i primi risultati di una sperimentazione in corso presso il laboratorio di geomatica forestale dell'Università di Firenze. L'obiettivo generale è quello di testare le potenzialità del telerilevamento multispettrale e *laser scanner* per la caratterizzazione delle coperture forestali al fine di generare informazioni di supporto per la pianificazione di livello aziendale e comprensoriale. In particolare, l'obiettivo di questo lavoro è testare le potenzialità d'uso combinato dei dati satellitari multispettrali ad altissima risoluzione spaziale e dei dati *laser scanner* aereo a bassa densità di punti laser per classificare l'uso del suolo, i tipi forestali e i tipi strutturali in un'area di studio dell'Italia centrale.

Materiali e metodi

Area di studio

Lo studio è stato condotto in Provincia di Firenze nel complesso forestale di Monte Morello (43°50'5''-43°51'14'' N, 11°11'23''-11°15'20'' E). L'area esaminata copre una superficie di 842 ettari ed è situata nei pressi della città di Firenze e dei centri abitati di Sesto Fiorentino e Calenzano (Figura 1). Le quote oscillano tra 70 e 700 m s.l.m., con pendenze che raramente superano il 60%. L'esposizione sud è prevalente. Il substrato è di natura calcarea. Il clima è tipicamente mediterraneo; secondo i dati della stazione di Calenzano la precipitazione annua è di 890 mm e la

temperatura media annua è di 14,6 °C. L'uso del suolo è dominato da formazioni forestali a prevalenza di specie quercine e da rimboschimenti di conifere effettuati all'incirca nel periodo 1909-1984. Sono inoltre presenti aree a uso agricolo (oliveti, vigneti e seminativi) e alcune superfici urbanizzate.



Figura 1 – Localizzazione dell'area di studio visualizzata su una immagine a colori naturali ottenuta combinando le bande del blu, del verde e del rosso del satellite Ikonos, fuse con la banda pancromatica.

Dati

Il dato satellitare ad altissima risoluzione spaziale utilizzato in questo lavoro è una immagine Ikonos di archivio acquisita in data 7 luglio 2008. La copertura nuvolosa dichiarata dal fornitore è di circa il 3% ed è situata all'esterno dell'area selezionata per la sperimentazione. È una immagine Preciso® Italia prodotta e ortorettificata da Planetek Italia s.r.l. utilizzando il modello di ortorettifica *Rational Polynomial Coefficients* e il metodo di ricampionamento *Cubic convolution*. L'immagine consiste in una banda pancromatica ricampionata alla risoluzione spaziale di 1 m e in quattro bande multispettrali (blu, verde, rosso e infrarosso) ricampionate alla risoluzione di 4 m.

Il dato *laser scanner* aereo è stato acquisito nell'inverno del 2008 per indagini di carattere idrogeologico. Il volo è stato eseguito a una quota media di 3000 m s.l.m. utilizzando un sensore Altm (*Airborne Laser Terrain Mapper*) Gemini. Il sensore ha registrato due ritorni per ciascun impulso con una densità media di punti laser di 0,4 punti per m². I dati ALS sono stati filtrati e classificati per separare i punti *ground* dai punti *above ground*. A tal fine i punti caratterizzati da errori generati dal sensore sono stati precedentemente identificati e eliminati dal dataset ALS utilizzando il software TerraScan. In seguito, i punti *ground* sono stati separati dai punti *above ground* utilizzando l'algoritmo di Axelsson (2000) e il software TerraScan. I punti così ottenuti sono stati sottoposti ad una procedura di interpolazione spaziale e sono stati ricampionati in formato *raster* con una risoluzione spaziale di 4 m per produrre, rispettivamente, il Modello Digitale del Terreno (DTM) e il Modello Digitale della Superficie (DSM). Il Modello Digitale delle Chiome (CHM) è stato ottenuto per sottrazione algebrica del DTM dal DSM. A partire dal DTM sono state generate le immagini della pendenza (SLO) e della esposizione (ASP) utilizzando il software Idrisi Selva.

Le linee elettriche presenti nell'area di studio sono state mascherate sulle immagini telerilevate sulla base di una fotointerpretazione del CHM.

Nel periodo 2013-2014, nell'area di studio è stata condotta una campagna di rilevamento a terra. Complessivamente sono state realizzate 69 aree di saggio. Di queste, 41 aree sono state materializzate a terra con GPS Trimble Juno 3B Handheld con 2-5 m di accuratezza; in ciascuna area è stato eseguito il rilievo dei diametri a 1,3 m da terra con cavalletto dendrometrico e la misura delle altezze con ipsometro Vertex III di tutte le piante con diametro a petto d'uomo superiore a 2,5 cm su una superficie circolare di 531 m². I dati sono stati utilizzati per classificare l'uso del suolo, il tipo forestale e il tipo strutturale di ciascuna area di saggio secondo la legenda riportata in Tabella 1. I boschi puri sono stati distinti dai boschi misti in funzione dell'area basimetrica utilizzando una soglia del 75%. Le restanti 28 aree sono state materializzate con GPS e sono state classificate a vista facendo sempre riferimento ad una superficie circolare di 531 m².

Legenda	Aree di saggio		Training site		Test site	
	Num.		Num.	%	Num.	%
<i>Uso del solo</i>						
1 Conifere	20		14	70	6	30
2 Latifoglie	12		8	67	4	33
3 Misto conifere e latifoglie	27		19	70	8	30
4 Non bosco	10		7	70	3	30
Tot.	69		48	70	21	30
<i>Tipo forestale</i>						
1 Querceto di roverella	6		4	67	2	33
2 Rimboschimento di cipresso	11		8	73	3	27
3 Rimboschimento di pino nero	3		2	67	1	33
4 Querceto misto	6		4	67	2	33
5 Rimboschimento di conifere varie	6		4	67	2	33
6 Rimboschimento di cipresso misto a latifoglie	18		13	72	5	28
7 Rimboschimento di pino nero misto a latifoglie	2		1	50	1	50
8 Rimboschimento di conifere varie misto a latifoglie	7		5	71	2	29
9 Non bosco	10		7	70	3	30
Tot.	69		48	70	21	30
<i>Tipo strutturale</i>						
1 Monoplano	3		2	67	1	33
2 Biplano	38		27	71	11	29
3 Stratificato	18		12	67	6	33
4 Non bosco	10		7	70	3	30
Tot.	69		48	70	21	30

Tabella 1 – Legenda utilizzata per la classificazione e ripartizione delle aree di saggio in training site e test site.

Classificazione delle coperture forestali

La procedura di classificazione testata in questo studio prevede 4 fasi consecutive: a) segmentazione delle immagini; b) selezione dei *training site* e dei *test site*; c) estrazione delle metriche *ALS-based* e delle variabili multispettrali; d) classificazione e valutazione dell'accuratezza.

a) Segmentazione delle immagini

La finalità di questa fase è quella di delineare i principali usi del suolo presenti nell'area di studio e di separare le coperture forestali in oggetti (poligoni) con proprietà simili in relazione alle loro caratteristiche compositive e strutturali. A tal fine è stato applicato l'algoritmo di segmentazione del software Definiens (vers. 5). È stato generato un primo livello di segmentazione (Figura 2) utilizzando le bande multispettrali Ikonos per delineare gli usi del suolo secondo la legenda riportata in Tabella 1. In seguito è stato generato un secondo livello di segmentazione (Figura 2) utilizzando le bande multispettrali Ikonos, il DTM, il DSM e il CHM (Ke et al., 2010) per delineare soprassuoli con differenti caratteristiche compositive e strutturali (Tabella 1). I parametri utilizzati per la

segmentazione (Baatz, Schäpe, 2000), riportati in Tabella 2, sono stati definiti con metodo empirico sulla base di una analisi visiva.

Livello della segmentazione	Scale factor	Color	Shape	Smoothness	Compactness	Immagini	Peso
1	75	0,9	0,1	0,5	0,5	Blu	1
						Verde	1
						Rosso	1
						Infrarosso	1
2	50	0,9	0,1	0,5	0,5	Blu	1
						Verde	1
						Rosso	1
						Infrarosso	1
						DTM	1
						CHM	1

Tabella 2 – Parametri utilizzati per la segmentazione delle immagini.

b) Selezione dei *training site* e dei *test site*

Le aree di saggio a terra sono state ripartite in *training site* e *test site* (Figura 2). Il 70% delle aree sono state utilizzate come *training* e il restante 30% come *test* (Tabella 1).

c) Estrazione delle metriche *ALS-based* e delle variabili multispettrali

Per ciascun poligono generato con la segmentazione di primo e secondo livello è stato calcolato il valore medio e la deviazione standard dei pixel delle bande multispettrali Ikonos inclusi in ciascun poligono utilizzando il software Definiens. In modo analogo è stato calcolato per ciascun poligono il valore medio e la deviazione standard dei pixel delle immagini generate a partire dal dataset ALS (DTM, DSM, CHM, SLO, ASP). Non sono state considerate variabili tessiturali perché è stato osservato che comportano un allungamento dei tempi di elaborazione senza apportare un contributo significativo per la classificazione delle coperture forestali (Ke et al., 2010; Machala, Zejdová, 2014). Complessivamente sono state calcolate 18 variabili per i poligoni di primo livello e 18 per quelli di secondo livello. Le variabili così ottenute sono state estratte sui *training site* e sui *test site* utilizzando una procedura di intersezione spaziale tra i punti *training* e *test* e i poligoni di primo e secondo livello.

d) Classificazione e valutazione dell'accuratezza

La classificazione delle coperture forestali in classi di uso del suolo, tipi forestali e tipi strutturali è stata effettuata utilizzando il metodo *Multinomial Logistic Regression* (MLR) e il software Idrisi Selva. Il metodo MLR viene usato per predire le classi di una variabile categorica sulla base di variabili continue utilizzate come variabili indipendenti. A esempio, il metodo MLR è stato impiegato per esaminare la distribuzione delle classi di vegetazione (Levasseur et al., 2013) e per classificare l'uso del suolo (McRoberts, 2009) e i cambiamenti (Næsset et al., 2013) utilizzando dati inventariali e immagini telerilevate. Per maggiori dettagli sulla tecnica MLR si rimanda a Agresti (2007). Qui ci limitiamo a ricordare che: i) il metodo non prevede restrizioni circa il numero di variabili indipendenti ma è necessario che le variabili non siano correlate le une con le altre; ii) nel caso di una variabile categorica composta da M classi ($M > 2$), una delle classi, in genere la prima o l'ultima, è utilizzata come classe di riferimento e il metodo MLR produce $M-1$ equazioni, una per ciascuna classe, per descrivere la probabilità di appartenenza ad una classe in funzione delle variabili indipendenti; iii) la classificazione finale in genere viene effettuata assegnando agli elementi da classificare la classe con probabilità maggiore.

È stata eseguita una analisi di correlazione lineare tra le 18 variabili estratte sui *training site* al fine di escludere le variabili con coefficiente di correlazione superiore a 0,5. Le variabili indipendenti

così selezionate sono: valore medio della banda del rosso (Ik3m), valore medio della banda dell'infrarosso (Ik4m), valore medio del DSM (DMSm), deviazione standard del DSM (DSMsd).

I *training site* con i valori delle variabili indipendenti estratti dai poligoni generati con il primo livello di segmentazione sono stati utilizzati per classificare l'uso del suolo (Tabella 1) con il metodo MLR, mentre, i valori delle variabili indipendenti estratti dai poligoni di secondo livello sono stati utilizzati per classificare il tipo forestale e il tipo strutturale (Tabella 1). In tutti i casi la classe "Non bosco" è stata utilizzata come classe di riferimento.

La bontà del modello MLR è stata valutata sui *training site* sulla base delle seguenti statistiche: *McFadden pseudo R-square*, *Overall accuracy* (OA). Infine, l'accuratezza della classificazione è stata validata sui *test site* utilizzando le seguenti statistiche: *Overall accuracy*, *Producer's accuracy* (PA), *User's accuracy* (PA), *Kappa Index of Agreement* (KIA).

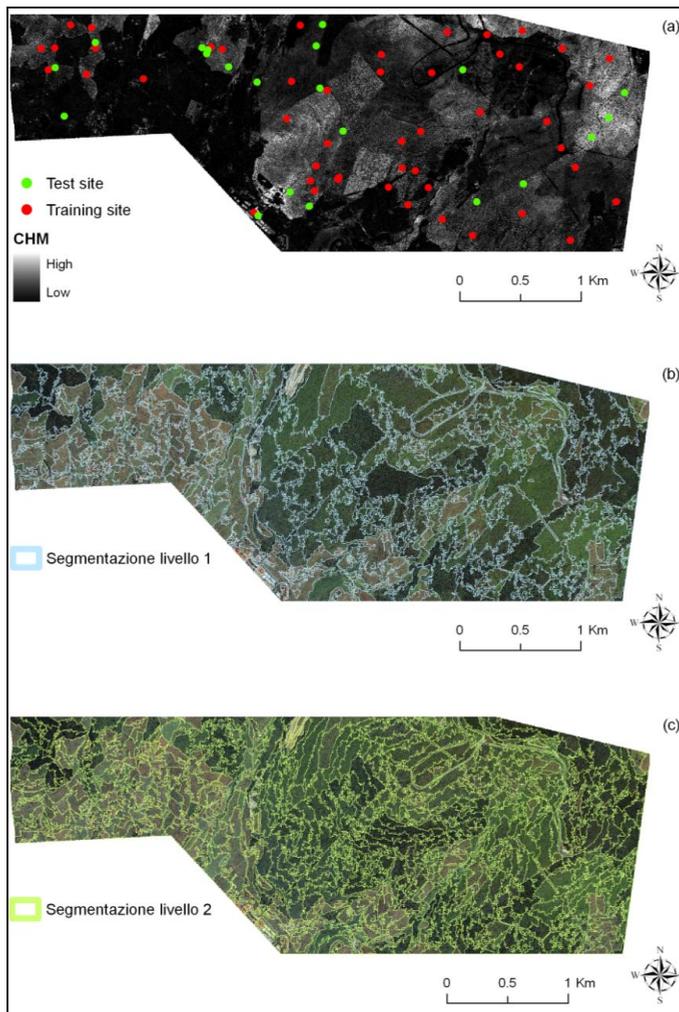


Figura 2 – a) Distribuzione nell'area di studio dei training site e test site visualizzati sul modello digitale delle chiome (CHM). b) Poligoni generati con la segmentazione di livello 1. c) Poligoni generati con la segmentazione di livello 2.

Risultati

I modelli MLR per la classificazione dell'uso del suolo, del tipo forestale e del tipo strutturale, elaborati sulla base di 48 aree di *training* e di 4 variabili indipendenti derivate dalle bande multispettrali del satellite Ikonos e dal dataset ALS, sono risultati statisticamente significativi ($p < 0,001$). I valori del *McFadden pseudo R-square* sono risultati in tutti i casi superiori a 0,2, soglia oltre la quale il modello può essere considerato relativamente buono (Clark, Hosking, 1986), e sono supportati dai valori di *Overall accuracy* (Tabella 3).

Modello MLR	Intercetta	Coefficienti delle variabili indipendenti			
		Ik3m	Ik4m	DSMm	DSMsd
<i>Usa del solo</i>					
1 Conifere	84,2768	-0,2957	-0,0282	0,0239	0,2233
2 Latifoglie	71,6366	-0,4639	0,0407	0,0279	0,1277
3 Misto conifere e latifoglie	90,5103	-0,3713	0,0004	0,0088	0,1270
N=48					
McFadden pseudo R-square=0,7780					
Overall accuracy=89,6					
<i>Tipo forestale</i>					
1 Querceto di roverella	580,9090	-9,2985	1,4535	0,4758	27,7301
2 Rimboschimento di cipresso	613,0361	-3,3001	-0,6952	0,4644	42,9732
3 Rimboschimento di pino nero	-112,0250	0,3474	-1,2475	0,8044	52,6497
4 Querceto misto	3336,4912	-20,1698	1,4521	0,0241	-18,5510
5 Rimboschimento di conifere varie	52,5158	-4,6799	0,2637	0,9054	50,9454
6 Rimboschimento di cipresso misto a latifoglie	1559,9663	-9,6660	1,1489	-1,5617	42,6476
7 Rimboschimento di pino nero misto a latifoglie	48,2645	-0,2600	-0,8228	0,7568	28,0254
8 Rimboschimento di conifere varie misto a latifoglie	1542,3768	-9,5907	1,1484	-1,5633	42,8288
N=48					
McFadden pseudo R-square=0,9072					
Overall accuracy=91,7					
<i>Tipo strutturale</i>					
1 Monoplano	74,0062	0,1139	-0,3232	0,0665	4,2520
2 Biplano	110,5983	-0,6276	0,0710	-0,0167	0,9960
3 Stratificato	111,3956	-0,6309	0,0721	-0,0177	0,8832
N=48					
McFadden pseudo R-square=0,5013					
Overall accuracy=72,9					

Tabella 3 – Modello Multinomial Logistic Regression (MLR) e coefficienti della regressione logistica
(Ik3m=media della banda del rosso; Ik4m=media della banda dell'infrarosso;
DSMm=media del DSM; DSMsd=deviazione standard del DSM).

La validazione effettuata sui *test site* denota tuttavia delle imprecisioni nella classificazione. Nel caso dell'uso del suolo, l'accuratezza complessiva della classificazione è risultata dell'81%, con un KIA pari a 0,729. La classificazione dei tipi forestali e dei tipi strutturali ha prodotto, rispettivamente, una accuratezza complessiva del 57,1% e del 71,4%, con un KIA rispettivamente pari a 0,478 e a 0,484. I valori di *User's accuracy* e di *Producer's accuracy* indicano, in sintesi, quanto segue: nel caso dell'uso del suolo (Tabella 4), si registra una concordanza relativamente buona tra la classificazione e le osservazioni a terra per le classi "Conifere" e "Misto conifere e latifoglie" e una concordanza moderata per la classe "Latifoglie"; nel caso dei tipi forestali (Tabella 5), si osserva una concordanza moderata tra la classificazione e la verità a terra per le classi "Rimboschimento di cipresso", "Rimboschimento di cipresso misto a latifoglie", "Rimboschimento di conifere varie misto a latifoglie" e una concordanza bassa per le altre classi; infine, nel caso dei tipi strutturali (Tabella 6), la concordanza tra la classificazione e il riscontro a terra è risultata buona per i soprassuoli con profilo monoplano, moderata per quelli con profilo biplano e bassa per le coperture con profilo stratificato.

Classificazione	Verità				UA (%)
	1	2	3	4	
1	5				100,0
2		2	1		66,7
3	1	2	7		70,0
4				3	100,0
PA (%)	83,3	50,0	87,5	100,0	OA (%)=81,0

Tabella 4 – Accuratezza della classificazione dell'uso del suolo valutata sui test site. La legenda dell'uso del suolo è riportata in Tabella 1 (OA=Overall accuracy, PA=Producer's accuracy, UA=User's accuracy).

Classificazione	Verità									UA (%)
	1	2	3	4	5	6	7	8	9	
1										0,0
2		2								100,0
3					1					0,0
4										0,0
5			1		1		1			33,3
6	2	1		2		5		1		45,5
7										0,0
8								1		100,0
9									3	100,0
PA (%)	0,0	66,7	0,0	0,0	50,0	100,0	0,0	50,0	100,0	OA (%)=57,1

Tabella 5 – Accuratezza della classificazione dei tipi forestali valutata sui test site. La legenda dei tipi è riportata in Tabella 1 (OA=Overall accuracy, PA=Producer's accuracy, UA=User's accuracy).

Classificazione	Verità				UA (%)
	1	2	3	4	
1	1				100,0
2		11		6	64,7
3					0,0
4				3	100,0
PA (%)	100,0	100,0	0,0	100,0	OA (%)=71,4

Tabella 6 – Accuratezza della classificazione dei tipi strutturali valutata sui test site. La legenda dei tipi è riportata in Tabella 1 (OA=Overall accuracy, PA=Producer's accuracy, UA=User's accuracy).

Discussione e conclusioni

In questo studio è stata valutata la potenzialità dei dati satellitari multispettrali e dei dati *laser scanner* aereo per classificare l'uso del suolo, i tipi forestali e i tipi strutturali in un'area di studio dell'Italia centrale caratterizzata dalla presenza di rimboschimenti di conifere e di boschi di specie quercine. La metodologia descritta prevede l'utilizzo delle bande multispettrali del satellite Ikonos, del DTM, DSM e CHM ottenuti da dati ALS a bassa densità di punti acquisiti nel periodo invernale e l'impiego di procedure di segmentazione delle immagini telerilevate. La classificazione, effettuata con metodo MLR, si basa sulle bande Ikonos del rosso e infrarosso e sulle metriche estratte dal DSM.

I risultati ottenuti, validati su un campione di aree *test*, evidenziano una buona concordanza tra la classificazione dell'uso del suolo e i dati a terra (accuratezza complessiva dell'81%), una concordanza moderata tra la classificazione dei tipi strutturali e la verità a terra (accuratezza complessiva del 71%) e una bassa concordanza tra la classificazione dei tipi forestali e le osservazioni a terra (accuratezza complessiva del 57%). McRoberts (2009), utilizzando dati multispettrali Landsat e il metodo MLR per classificare 4 classi di uso del suolo in Minnesota (USA), analoghe a quelle qui considerate, riporta una accuratezza complessiva della classificazione del 74%. Alberti et al. (2013) hanno testato in ambiente alpino le potenzialità dei dati LiDAR a bassa densità di punti per classificare i boschi di conifere, di latifoglie e misti e per identificare 6

diversi tipi strutturali, ottenendo, rispettivamente, una accuratezza complessiva della classificazione del 73% e del 68%.

Nelle condizioni esaminate, le difficoltà riscontrate nella classificazione dei tipi strutturali, e in particolare dei soprassuoli a profilo stratificato, sono probabilmente riconducibili alla bassa densità dei punti che caratterizzano il dato ALS utilizzato in questo studio, mentre, per quanto riguarda i tipi forestali, oltre alla densità dei punti e al periodo di acquisizione del dato ALS (invernale), la dimensione relativamente bassa del campione di aree a terra può avere influito sulle prestazioni del classificatore MLR (Levasseur et al., 2013). Ulteriori indagini sono perciò necessarie per valutare la capacità dei dati satellitari multispettrali e *laser scanner* da aereo di classificare le coperture forestali in relazione alle loro caratteristiche compositive e strutturali.

Ringraziamenti

I dati ALS sono stati gentilmente concessi dalla Regione Toscana. Il lavoro è stato svolto in parte con fondi di ricerca di Ateneo anno 2012 (responsabile scientifico: D. Travaglini) e in parte con fondi del progetto PRIN 2012 “NEUFOR: Modelli innovativi di analisi dei servizi ecosistemici nell'ambito di formazioni boschive urbane e periurbane” (coordinatore nazionale: G. Sanesi) finanziato dal Ministero dell'Istruzione, Università e Ricerca. Si ringraziano E. Altomonte, N. Camarretta, S. Fiorentini, A. Mariottini e F.P. Piemontese per avere contribuito nelle fasi di rilevamento a terra.

Bibliografia

- Agresti A. (2007), *An introduction to categorical data analysis*, Hoboken, NJ, Wiley-Interscience.
- Alberti G., Boscutti F., Pirotti F., Bertacco C., De Simon G., Sigura M., Cazorzi F., Bonfanti P. (2013), “A LiDAR-based approach for a multi-purpose characterization of Alpine forests: an Italian case study”, *iForest*, 6: 156-168.
- Axelsson P. (2000), “DEM generation from laser scanner data using adaptive TIN models”, *International Archives of Photogrammetry and Remote Sensing*, 33: 111-118 (B4/1; Part 4).
- Baatz M., Schäpe A. (2000), *Multiresolution segmentation - an optimization approach for high quality multi-scale image segmentation*, in: Strobl, J., Blaschke, T., Greisebener, G. (Eds.), *Angewandte Geographische Informationsverarbeitung XII*. Wichmann-Verlag, Heidelberg, Germany, 12-23.
- Clark W.A., Hosking P.L. (1986), *Statistical Methods for Geographers*, New York, John Wiley & Sons.
- Ke Y., Quackenbush L.J., Im J. (2010), “Synergistic use of QuickBird multispectral imagery and LIDAR data for object-based forest species classification”, *Remote Sensing of Environment*, 114: 1141-1154.
- Levasseur G., Vrac M., Roche D.M., Paillard D., Guiot J. (2013), “An objective methodology for potential vegetation reconstruction constrained by climate”, *Global and Planetary Change*, 104: 7-22.
- Machala M., Zejdová L. (2014), “Forest Mapping Through Object-based Image Analysis of Multispectral and LiDAR Aerial Data”, *European Journal of Remote Sensing*, 47: 117-131.
- McRoberts R.E. (2009), “A two-step nearest neighbors algorithm using satellite imagery for predicting forest structure within species composition classes”, *Remote Sensing of Environment*, 113: 532-545.
- Næsset E., Bollandsås O.M., Gobakken T., Gregoire T.M., Ståhl G. (2013), “Model-assisted estimation of change in forest biomass over an 11 year period in a sample survey supported by airborne LiDAR: A case study with post-stratification to provide “activity data””, *Remote Sensing of Environment*, 128: 299-314.