

## Stime non parametriche di attributi forestali con dati inventariali e immagini telerilevate

Davide Travaglini (\*), Giuseppe De Simon (\*\*), Nicola Puletti (\*), Giorgio Alberti (\*\*),  
Alessandro Peressotti (\*\*), Gherardo Chirici (\*\*\*), Piermaria Corona (\*\*\*\*)

(\*) Dipartimento di Economia, Ingegneria, Scienze e Tecnologie Agrarie e Forestali, Università di Firenze  
Via San Bonaventura, 13 – 50145 Firenze. Tel.: +39 055 3288618, Fax: +39 055 319179,  
e-mail: davide.travaglini@unifi.it

(\*\*) Dipartimento di Scienze Agrarie e Ambientali, Università di Udine. Via delle Scienze, 208 – 33100 Udine  
Tel.: +39 0432 558664, Fax: +39 0432 558603, e-mail: giuseppe.desimon@uniud.it

(\*\*\*) Dipartimento di Scienze e Tecnologie per l'Ambiente e il Territorio, Università del Molise  
C.da Fonte Lappone – 86090 Pesche (IS), Italy. Tel.: +39 0874 404138, Fax: +39 0874 404123,  
e-mail: piermaria.corona@unitus.it

(\*\*\*\*) Dipartimento di Scienze dell'Ambiente Forestale e delle sue Risorse, Università della Tuscia  
Via S. Camillo de Lellis – 01100 Viterbo. Tel.: +39 0761 357425, Fax: +39 0761 357389,  
e-mail: piermaria.corona@unitus.it

### Riassunto

In questo lavoro è stata sperimentata l'applicazione del metodo non parametrico *k-Nearest Neighbors* (*k*-NN) con l'obiettivo di produrre stime su vasta scala di attributi biofisici di interesse forestale. Sono stati utilizzati dati inventariali rilevati a terra e immagini satellitari multispettrali. Le prestazioni di diverse configurazioni dell'algoritmo *k*-NN sono state valutate per confronto con procedura *leave-one-out*. La configurazione ottimale è stata utilizzata per produrre una stima spazializzata della provvigione legnosa unitaria. La validazione delle stime è stata effettuata utilizzando dati estratti dai piani di assestamento forestale. La sperimentazione è stata condotta nella regione Friuli Venezia Giulia. Nelle condizioni esaminate l'errore quadratico medio si riduce significativamente applicando filtri di tipo passa-basso alle immagini satellitari. I risultati ottenuti evidenziano una correlazione significativa tra i valori di provvigione predetti con *k*-NN e i valori di provvigione riportati nei piani di assestamento.

### Abstract

*In this paper the k-Nearest Neighbors (k-NN) method was applied in Friuli Venezia Giulia region (Italy) to produce estimations of biophysical forest attributes on large territories. Forest inventory data and multispectral satellite images were used. Several k-NN configurations were assessed by leave-one-out techniques. The best k-NN configuration was applied to produce a forest growing stock volume map. k-NN estimates were validated by data extracted from forest management plans. Under the examined conditions the RMSE was significantly reduced by applying mean (low-pass) filter to the satellite images. Achieved results show a significant association between growing stock volume predicted by k-NN and data reported by forest management plans.*

### Introduzione

La pianificazione forestale richiede la conoscenza di informazioni accurate sullo stato e la consistenza delle risorse forestali di un determinato territorio. Tradizionalmente tali informazioni sono acquisite per mezzo di inventari forestali condotti a scala nazionale, regionale o comprensoriale; a scala aziendale le caratteristiche delle formazioni forestali sono invece, in genere, acquisite in fase di redazione dei piani di assestamento e di gestione forestale.

Negli ultimi anni si registra una crescente richiesta di dati cartografici rappresentativi della variabilità spaziale di attributi biofisici dei soprassuoli forestali (a es., provvigione legnosa per unità di superficie). I dati telerivati sono stati utilizzati per estendere a interi territori i valori di variabili biofisiche misurati in corrispondenza delle unità inventariali di campionamento a terra. Questa operazione, che in gergo tecnico è chiamata *spazializzazione*, viene in genere realizzata mediante procedure che prevedono l'impiego di metodi di stima di tipo parametrico o non parametrico (Corona, 2010). Tra i metodi parametrici la regressione lineare multipla è quello più comunemente utilizzato, mentre tra i metodi non parametrici quello più utilizzato in applicazioni di carattere operativo è il *k-Nearest Neighbors* (a es: Tomppo et al., 2008).

Obiettivo del presente lavoro è la sperimentazione del metodo *k-Nearest Neighbors* in regione Friuli Venezia Giulia al fine di spazializzare la provvigione legnosa unitaria ( $\text{m}^3\text{ha}^{-1}$ ) attraverso la combinazione di dati inventariali e immagini telerilevate con sensori multispettrali.

### **k-Nearest Neighbors**

Il metodo *k-Nearest Neighbors* (*k*-NN) permette di stimare, in modo non parametrico, il valore di una variabile  $Y$  per gli  $N$  elementi di una popolazione per i quali sia noto il valore vero di variabili ausiliarie (ancillari) correlate con  $Y$ , posto che per un campione di  $n$  elementi sia noto anche il valore vero di  $Y$ .

In genere la popolazione è costituita dai pixel di un'immagine telerilevata multispettrale, la variabile  $Y$  è misurata a terra in corrispondenza degli  $n$  pixel del campione (detto *reference set*) e per tutti gli  $N$  pixel sono noti i valori di variabili ancillari rappresentate dai *digital number* (DN) delle singole bande spettrali, da indici ottenuti dalla combinazione di queste ultime e da altre eventuali informazioni correlate con i valori di  $Y$  (a es., quota, esposizione ecc.). Il valore incognito  $\tilde{y}_j$  della variabile  $Y$  per ciascun  $j$ -esimo pixel dell'insieme  $N$ - $n$  (detto *target set*) può essere stimato come media pesata dei valori di  $Y$  misurati in corrispondenza dei  $k$  pixels del *reference set* più vicini al  $j$ -esimo pixel nello spazio multidimensionale definito dalle variabili ancillari:

$$\tilde{y}_j = \frac{\sum_{i=1}^k w_{ij} y_i}{\sum_{i=1}^k w_{ij}}$$

dove il peso  $w$  può essere posto pari a  $1/k$  (in questo caso il valore  $\tilde{y}_j$  è pari alla media aritmetica dei valori di  $Y$  misurati nei  $k$  pixel del *reference set* più vicini al  $j$ -esimo pixel) o, come avviene più frequentemente, può essere calcolato in modo inversamente proporzionale alla distanza multidimensionale tra il  $j$ -esimo pixel e ciascuno dei  $k$  pixel del *reference set* a esso più vicini.

La distanza multidimensionale è misurata nello spazio definito dalle variabili ancillari considerate e può essere calcolata attraverso diversi tipi di misure, la più semplice delle quali è la distanza euclidea (De Maesschalck et al., 2000).

Attraverso una stratificazione per mezzo di informazioni ancillari è possibile limitare la ricerca dei  $k$  pixel più vicini ai quei pixel del *reference set* che presentano determinati requisiti (a es., differenza di quota rispetto al  $j$ -esimo pixel non superiore a una determinata soglia).

La scelta delle variabili ancillari, del tipo di distanza multidimensionale e di  $k$  è in genere condotta empiricamente attraverso una procedura *leave-one-out* (LOO) di valutazione dell'accuratezza delle stime prodotte. Questo tipo di procedura prevede la stima mediante *k*-NN del valore della variabile  $Y$  per ciascun  $i$ -esimo pixel del *reference set*, con l'accortezza di escludere, ai fini della stima stessa, il valore vero  $y_i$  corrispondente a quel pixel: si ottengono così  $n$  valori stimati  $\tilde{y}_i$  che confrontati con i corrispondenti valori veri  $y_i$  permettono di valutare l'accuratezza delle stime prodotte. Sulla base dei risultati della procedura LOO viene definita la configurazione dell'algoritmo *k*-NN (in termini di variabili ancillari considerate, tipo di distanza multidimensionale, valore di  $k$ ) che può fornire, nel caso indagato, le stime più accurate e che quindi viene applicata per la stima di  $\tilde{y}_j$  sui pixel del *target set*.

## Materiali e metodi

### Area di studio

La sperimentazione è stata condotta sull'intera regione Friuli Venezia Giulia (FVG). Il patrimonio forestale del FVG copre una superficie di 323.832 ha, pari al 41% del territorio regionale (INFC, 2007) e si dispone quasi interamente nella fascia alpina e prealpina oltre che sul Carso. Nella pianura, un tempo interamente boscata, è rimasto solo qualche relitto di foresta planiziale, mentre in collina e nella fascia prealpina buona parte del territorio è ricoperta da boschi di neof ormazione derivanti dall'abbandono dell'attività agricola. Circa 120.000 ha di superficie forestale ricadono all'interno di 105 piani di assestamento, perlopiù in aree di proprietà pubblica. Le maggior parte dei boschi della regione sono governati a fustaia. Le principali categorie forestali sono le faggete (68.000 ha) e le peccete che, assieme ai piceo-faggeti e piceo-abieteti, coprono una superficie di 63.000 ha. Molto estese sono anche le pinete di pino nero e pino silvestre (36.000 ha) e gli orno-ostrieti e ostrio-querceti (32.000 ha).

### Dati

Nella presente sperimentazione sono stati combinati dati di provvigione legnosa unitaria ( $m^3ha^{-1}$ ) misurati nelle unità inventariali di campionamento a terra con dati satellitari acquisiti da sensori multispettrali ad alta risoluzione geometrica.

I rilievi a terra sono stati eseguiti nel biennio 2005-2006 in 378 unità di campionamento inventariale distribuite sul territorio regionale con metodo sistematico non allineato secondo una maglia di 1 km di lato; i dati dendrometrici sono stati forniti dalla Direzione Centrale Risorse Agricole, Naturali e Forestali della Regione Autonoma FVG e sono stati rilevati su aree di saggio di  $530 m^2$  di superficie con soglia di cavallettamento di 4,5 cm. La biomassa e la provvigione legnosa sono quindi stati calcolati utilizzando la metodologia proposta da Anfodillo et al. (2006); la provvigione oscilla tra 2 e  $793 m^3ha^{-1}$  con un valore medio di  $189 m^3ha^{-1}$ .

I dati telerilevati corrispondono a una scena satellitare Landsat 5 TM del 10/09/2006, con risoluzione spaziale pari a 30 m, e a una scena satellitare IRS-P6, acquisita dal sensore LISS-III il 12/06/2006, con risoluzione spaziale pari a 20 m. La scena Landsat è libera da nuvole e copre quasi interamente il territorio regionale, ad eccezione dell'estremità occidentale corrispondente al comune di Erto e Casso. La ripresa IRS, invece, presenta una parziale copertura nuvolosa e non copre la zona della Venezia Giulia, parte delle Alpi orientali e buona parte della pianura. Sono state utilizzate come variabili ancillari tutte le bande multispettrali a eccezione della banda 6 del satellite Landsat; inoltre, per valutare le prestazioni di stima dovute all'impiego di differenti *set* di variabili ancillari sono state eseguite le seguenti elaborazioni dei dati telerilevati: le bande satellitari sono state filtrate con filtri passa-basso e a ciascun pixel dell'immagine è stato attribuito un valore di DN pari al valore medio dei DN dei pixels inclusi all'interno di finestre mobili di dimensione  $3 \times 3$  e  $5 \times 5$ ; le bande multispettrali Landsat sono state sottoposte ad un'analisi delle componenti principali (PCA) per estrarre la prima componente (PCA1); le bande Landsat del rosso (TM3) e dell'infrarosso vicino (TM4, TM5) sono state combinate per calcolare l'indice NVDI (*Normalized Difference Vegetation Index*) e l'indice IRI (*Infrared index*):

$$NDVI = \frac{(TM\ 4 - TM\ 3)}{(TM\ 4 + TM\ 3)} \qquad IRI = \frac{(TM\ 4 - TM\ 5)}{(TM\ 4 + TM\ 5)}$$

Dal Catalogo dei Dati Ambientali e Territoriali della Regione Autonoma FVG è stata acquisita la carta vettoriale dei tipi forestali regionali in scala 1:5.000 aggiornata all'anno 1998; la carta è stata utilizzata per individuare le aree bosco all'interno dell'area di studio. Inoltre, dal medesimo Catalogo sono stati acquisiti i dati particellari dei piani di assestamento forestale aggiornati all'anno 2007; i dati particellari sono stati utilizzati per effettuare la validazione delle stime prodotte con metodo *k*-NN.

**Configurazioni k-NN**

In Tabella 1 sono indicate le configurazioni dell'algoritmo di stima *k*-NN messe a confronto in funzione delle variabili ancillari utilizzate, del metodo impiegato per il calcolo della distanza multidimensionale e di *k*.

In particolare sono state effettuate prove di stima utilizzando per confronto le seguenti variabili ancillari: 6 bande Landsat (1); 6 bande Landsat combinate con gli indici NDVI e IRI (2); PCA1 estratta dalle 6 bande Landsat (3); 6 bande Landsat filtrate con filtri passa-basso 3x3 (4) e 5x5 (5). Inoltre sono state effettuate prove utilizzando per confronto le bande satellitari Landsat e IRS filtrate (7-8-10-11) e non (6-9). In questo caso, dal *reference set* originale costituito da 378 unità inventariali sono state scartate le unità di campionamento che non risultavano coperte dalla scena IRS e le unità che sulle immagini IRS ricadevano in aree coperte da nuvole e da loro ombre; complessivamente il numero di unità inventariali scartate è risultato pari a 163 unità.

Tutte le prove sono state effettuate testando valori di *k* compresi tra 1-20 e utilizzando per confronto i seguenti metodi di calcolo della distanza multidimensionale: ED – distanza euclidea (De Maesschalck et al., 2000), MD – distanza di Mahalanobis (Holmstrom et al., 2001), FD – distanza di Mahalanobis modificata con pesi *fuzzy* (Chirici et al., 2008).

Numero di unità inventariali	Variabili ancillari	Distanza multidimensionale	<i>k</i>
378	1) 6 bande Landsat 5 TM 2) 6 bande Landsat 5 TM + NDVI+IRI 3) PCA1 4) 6 bande Landsat 5 TM filtrate con filtro 3x3 5) 6 bande Landsat 5 TM filtrate con filtro 5x5	ED, MD, FD	1-20
215	6) 6 bande Landsat 5 TM 7) 6 bande Landsat 5 TM filtrate con filtro 3x3 8) 6 bande Landsat 5 TM filtrate con filtro 5x5 9) 4 bande IRS 10) 4 bande IRS filtrate con filtro 3x3 11) 4 bande IRS filtrate con filtro 5x5	ED, MD, FD	1-20

Tabella 1 – Elenco delle configurazioni dell'algoritmo *k*-NN sperimentate.

**Valutazione della affidabilità delle stime**

La affidabilità delle stime prodotte dalle singole configurazioni dell'algoritmo *k*-NN è stata valutata sul *reference set* calcolando l'errore quadratico medio (RMSE) con procedura LOO. La migliore configurazione è stata utilizzata per effettuare la spazializzazione della provvigione legnosa unitaria nei pixels bosco del *target set*.

Le stime per pixel sono state poi incrociate con i confini particellari estratti dai piani di assestamento forestale e per ciascuna particella è stata calcolata la provvigione legnosa unitaria ottenuta come media delle stime *k*-NN dei pixels inclusi in ciascuna particella. Le stime così prodotte sono state raggruppate in classi di ampiezza di 10 m<sup>3</sup>ha<sup>-1</sup> e sono state confrontate con i valori di provvigione riportati nei piani per le sole particelle sottoposte a cavallettamento totale. Prima di effettuare il confronto è stato però necessario armonizzare i dati assestamentali e inventariali. Infatti, considerato che i piani di assestamento utilizzano una soglia di cavallettamento di 17,5 cm rispetto alla soglia di 4,5 cm utilizzata dall'inventario, la provvigione stimata con *k*-NN è stata ridotta utilizzando la seguente relazione empirica individuata sulla base dei dati inventariali:  $v=70,64x^{-1,17}$  ( $R^2=0,94$ ), dove *v* corrisponde alla proporzione di volume delle piante di diametro < 17,5 cm e *x* è il volume delle piante di diametro > 4,5 cm. Inoltre è stato necessario aggiornare i dati di provvigione dei piani di assestamento all'anno di acquisizione delle immagini satellitari. Tale aggiornamento è stato effettuato utilizzando i dati di incremento percentuale indicati nei piani; nel

caso di particelle forestali sottoposte a interventi di taglio è stato considerato l'anno d'intervento e la massa asportata ad ettaro. Sulla base dei dati disponibili nei piani è stato effettuato l'aggiornamento della provvigione legnosa unitaria di 216 particelle forestali rilevate con cavallettamento totale.

### Risultati

In Tabella 2 sono riportate le migliori prestazioni di stima delle diverse configurazioni dell'algoritmo  $k$ -NN valutate con procedura LOO. Dai risultati ottenuti si osserva che l'applicazione di filtri passa-basso alle immagini satellitari determina una riduzione significativa dell'errore quadratico medio, pari all'8% nel caso di immagini Landsat: come già evidenziato da altri Autori, l'impiego di immagini filtrate permette di ridurre possibili errori di coregistrazione tra le immagini satellitari e le unità di campionamento a terra (Mäkelä e Pekkarinen, 2001). L'utilizzo degli indici NDVI e IRI non ha comportato un sostanziale miglioramento delle stime e l'impiego della PCA1 non ha prodotto risultati positivi. Nelle condizioni esaminate le immagini satellitari IRS hanno prodotto risultati simili a quelli ottenuti con le immagini Landsat.

<i>Variabili ancillari</i>	<i>Distanza multidimensionale</i>	<i>k</i>	<i>R<sup>2</sup></i>	<i>RMSE m<sup>3</sup>ha<sup>-1</sup></i>
1) 6 bande Landsat 5 TM	ED	8	0,44	140,9
2) 6 bande Landsat 5 TM + NDVI+IRI	ED	10	0,46	139,2
3) PCA1	MD	6	0,28	155,9
4) 6 bande Landsat 5 TM filtrate con filtro 3x3	FD	8	0,50	135,9
5) 6 bande Landsat 5 TM filtrate con filtro 5x5	FD	10	0,56	130,2
6) 6 bande Landsat 5 TM	ED	12	0,45	135,1
7) 6 bande Landsat 5 TM filtrate con filtro 3x3	FD	10	0,49	131,6
8) 6 bande Landsat 5 TM filtrate con filtro 5x5	ED	11	0,54	127,1
9) 4 bande IRS	FD	8	0,54	127,1
10) 4 bande IRS filtrate con filtro 3x3	MD	9	0,56	124,7
11) 4 bande IRS filtrate con filtro 5x5	FD	9	0,57	124,0

Tabella 2 – Valutazione delle stime  $k$ -NN con procedura LOO.

La stima spazializzata della provvigione legnosa unitaria sui pixel del *target set* è stata effettuata utilizzando come variabili ancillari le bande multispettrali Landsat filtrate con filtro 5x5 e applicando la distanza multispettrale di Mahalanobis modificata con pesi *fuzzy* e un valore di  $k$  pari a 10 (Figura 1a). Con questo tipo di configurazione è stata ottenuta una correlazione significativa ( $R^2=0,82$ ;  $P<0,001$ ) tra i valori simulati e quelli misurati sulle unità inventariali del *target set*, anche se la pendenza della retta di regressione (0,67) denota una sottostima del dato reale per valori di provvigione unitaria superiore a 400 m<sup>3</sup>ha<sup>-1</sup>, probabilmente dovuta alla saturazione dei valori di riflettanza delle immagini satellitari per le classi di provvigione maggiori.

I risultati della validazione effettuata con i dati dei piani di assestamento forestale sono riportati in Figura 1b. La correlazione tra i valori simulati di provvigione legnosa unitaria e i valori di provvigione riportati nei piani di assestamento è significativa ( $R^2=0,73$ ;  $P<0,001$ ). Anche in questo caso i valori di provvigione superiori a 350-400 m<sup>3</sup>ha<sup>-1</sup> risultano sottostimati a causa della saturazione dei valori di riflettanza delle immagini satellitari. Infatti, se si considerano solo i valori di provvigione compresi tra 100 e 350 m<sup>3</sup>ha<sup>-1</sup> si ottiene una retta di regressione con una pendenza prossima a 1, una costante prossima 0 e  $R^2=0,80$ .

### Conclusioni

Informazioni aggiornate sullo stato e la consistenza delle risorse boschive sono necessarie per la pianificazione e il monitoraggio forestale e per le attività di *reporting* internazionale. Dati misurati a terra negli inventari forestali possono essere proficuamente utilizzati insieme a dati telerilevati per

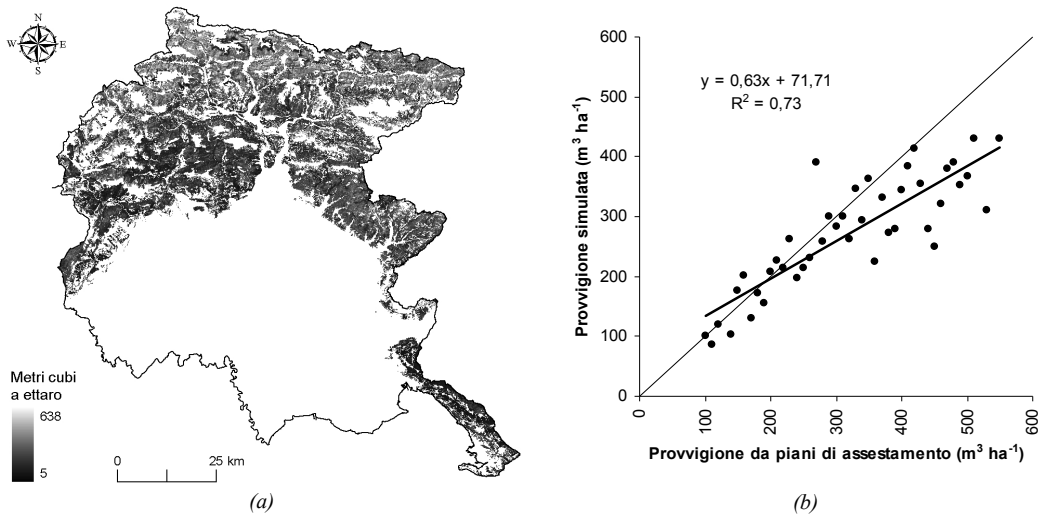


Figura 1 – a) Stima spazializzata della provvigione legnosa unitaria. b) Confronto tra i valori di provvigione simulati e i valori riportati nei piani di assestamento forestale.

realizzare stime spazializzate di attributi forestali su vaste superfici (Corona, 2010), altrimenti difficilmente ottenibili.

I risultati conseguiti dalla sperimentazione condotta in Friuli Venezia Giulia sono incoraggianti ed evidenziano la potenziale applicabilità del metodo non parametrico *k-Nearest Neighbors* finalizzato alla produzione di informazioni di supporto alla pianificazione forestale regionale.

## Bibliografia

- Anfodillo T., Pilli R., Salvatori I. (2006), Indagine preliminare sullo stock di carbonio nelle foreste del Veneto. Regione Veneto: 125 pp.
- Chirici G., Barbati A., Corona P., Marchetti M., Travaglini D., Maselli F., Bertini R. (2008), Non-parametric and parametric methods using satellite images for estimating growing stock volume in alpine and Mediterranean forest ecosystems. *Remote Sensing of Environment*, 112: 2686–2700.
- Corona P. (2010), Integration of forest mapping and inventory to support forest management. *iForest*, 3: 59-64 [online: 2010-05-17] URL: <http://www.sisef.it/iforest/show.php?id=531>.
- De Maesschalck R., Jouan-Rimbaud D., Massart D.L. (2000), The Mahalanobis distance. *Chemometrics and Intelligent Laboratory Systems*, 50: 1–18.
- Holmstrom H., Nilsson M., Ståhl, G. (2001), Simultaneous estimations of forest parameters using aerial photograph-interpreted data and the k nearest neighbor method. *Scandinavian Journal of Forest Research*, 16: 67–78.
- INFC (2007), Le stime di superficie 2005 – Prima parte. Autori G. Tabacchi, F. De Natale, L. Di Cosmo, A. Floris, C. Gagliano, P. Gasparini, L. Genchi, G. Scrinzi, V. Tosi. *Inventario Nazionale delle Foreste e dei Serbatoi Forestali di Carbonio*. MiPAF – Corpo Forestale dello Stato – Ispettorato Generale, CRA – ISAF, Trento. [on line] URL: <http://www.infc.it>.
- Katila M., Tomppo E. (2001), Selecting estimation parameters for the Finnish multisource National Forest Inventory. *Remote Sensing of Environment*, 76: 16–32.
- Mäkelä H., Pekkarinen A. (2001), Estimation of timber volume at the sample plot level by means of image segmentation and Landsat TM imagery. *Remote Sensing of Environment*, 77: 66–75.
- Tomppo E., Olsson H., Ståhl G., Nilsson M., Hagner O., Katila M. (2008), Combining national forest inventory field plots and remote sensing data for forest databases. *Remote Sensing of Environment*, 112: 1982–1999.