

STIMA CON METODO NON PARAMETRICO DI ATTRIBUTI FORESTALI A FINI ASSESTAMENTALI

Orazio CIANCIO (*), Davide TRAVAGLINI (*), Roberta BERTINI (*), Gherardo CHIRICI (**),
Nicola PULETTI (*)

(*) geoLAB – Laboratorio di Geomatica, Dipartimento di Scienze e Tecnologie Ambientali Forestali, Università degli Studi di Firenze, Via S. Bonaventura 13, I-50145 Firenze. Tel.: +39 055 3288621, Fax: +39 055 319179, e-mail: davide.travaglini@unifi.it

(**) EcoGeoFor - Laboratorio di Ecologia e Geomatica Forestale, Dipartimento di Scienze e Tecnologie per l'Ambiente e il Territorio, Università del Molise, Contrada Fonte Lappone snc, I-86090 Pesche (Isernia). Tel.: +39 0874 404113, Fax: +39 0874 404123, e-mail: gherardo.chirici@unimol.it

Riassunto

Il presente contributo illustra i primi risultati ottenuti dall'applicazione del metodo non parametrico *k-Nearest Neighbors* (*k*-NN) per la stima della provvigione legnosa nelle Riserve Naturali Statali Biogenetiche di Vallombrosa e dei Tomboli di Cecina. Il metodo *k*-NN è stato applicato per la stima della massa legnosa di popolamenti di abete bianco e di pino domestico attraverso l'integrazione di misure a terra, dati telerilevati dal satellite Landsat 5 TM e altre variabili ancillari. La configurazione locale ottimale dell'algoritmo *k*-NN nei due casi di studio è stata definita sulla base dell'accuratezza delle stime prodotte con procedura di validazione *leave-one-out*. La configurazione ottimale è stata quindi applicata per stimare la massa legnosa e i risultati sono stati validati attraverso il confronto con i valori riportati nei piani di gestione delle due Riserve. I primi risultati ottenuti in termini di accuratezza delle stime nelle due aree (errore quadratico medio percentuale delle stime pari al 21% e al 30% per Cecina e Vallombrosa rispettivamente) permettono di ipotizzare un uso operativo del metodo in assestamento forestale.

Abstract

This contribution describes the application of a non-parametric method for the estimation of forest growing stock within two National Nature Reserves in Tuscany (Italy): Vallombrosa (Florence) and Tomboli di Cecina (Livorno). The *k*-Nearest Neighbors (*k*-NN) method was tested for the estimation of silver fir and mediterranean pine stand volumes integrating field data, Landsat 5 TM imagery and auxiliary data. The local optimal *k*-NN configuration was defined by leave-one-out cross validation technique, comparing spectral distances of increasing complexity. The locally optimal *k*-NN configuration was used for estimations and the results were finally compared with data provided by forest management plans. Achieved results in the two study areas (average percent root mean square error of estimations of 21% and 30% for Cecina and Vallombrosa respectively) led to the possible operative use of the method in forest management.

Introduzione

Per la redazione di piani di assestamento e di gestione forestale di proprietà pubbliche o private è richiesta l'acquisizione di informazioni descrittive della condizione dei boschi attraverso indagini qualitative e quantitative. Tali informazioni forniscono al tecnico un quadro conoscitivo aggiornato dello stato dei soprassuoli forestali e supportano la scelta delle più idonee linee d'intervento.

L'indagine qualitativa consiste nella descrizione della stazione, delle caratteristiche compositive e strutturali dei popolamenti e delle loro tendenze evolutive. Essa fornisce tra l'altro le conoscenze di base per impostare la compartimentazione della superficie da assestare. Per compartimentazione

s'intende la suddivisione del bosco in aree omogenee per tipologia colturale, definite *unità colturali*, e l'aggregazione di più unità colturali in *comparti colturali*, unità di gestione caratterizzate da confini fisiografici.

L'indagine quantitativa consiste nel rilievo dendrometrico di una serie di attributi quali il numero, il diametro, l'altezza e la massa legnosa delle piante che compongono il soprassuolo. In genere i dati dendrometrici vengono esaminati a livello di unità colturale, di comparto e di intera foresta, ma il costo dei rilievi può determinare importanti limitazioni. Per questo motivo, negli ultimi anni, almeno in Italia, i rilievi per cavallettamento totale (misura dei diametri di tutti gli alberi) di intere unità vengono eseguiti solo in limitati casi e sono sostituiti, spesso, con rilievi su aree di saggio di dimensione ridotta (Corona, 2007).

In questo studio il metodo non parametrico *k-Nearest Neighbors* (*k-NN*) è stato applicato per la stima della massa legnosa (provvigione) di soprassuoli forestali su scala aziendale, integrando misure a terra rilevate su aree di saggio con dati telerilevati da satellite e altre variabili ancillari. L'algoritmo, previa ottimizzazione, è stato applicato per la stima della provvigione delle singole unità colturali presenti nelle due aree di studio; i dati stimati sono stati poi confrontati con i valori contenuti nei piani di gestione forestale ottenendo quindi una valutazione dell'accuratezza delle stime.

Materiali e Metodi

Metodo *k-Nearest Neighbors*

Il metodo *k-NN* permette di stimare in modo non parametrico il valore di una variabile Y per gli N elementi di una popolazione per la quale sia noto il valore vero di variabili ausiliarie correlate con Y , posto che per un campione di n elementi della popolazione N sia noto anche il valore vero di Y .

Nella sperimentazione la popolazione N è costituita dalle unità colturali dei piani di gestione, la variabile Y è la provvigione ad ettaro misurata a terra in corrispondenza di n unità del campione (detto *reference set*) e per tutte le N unità sono noti i valori di variabili ausiliarie rappresentate dai valori medi dei *digital number* (DN) delle singole bande spettrali del satellite Landsat 5 TM, da indici ottenuti dalla combinazione di queste ultime e da altre eventuali informazioni ancillari correlate con i valori di Y (a esempio, quota, esposizione, tipo di suolo, ecc).

Il valore incognito della variabile Y per ciascuna j -esima unità dell'insieme $N-n$ (detto *target set*) può essere stimato come media pesata dei valori di Y misurati in corrispondenza delle k unità del *reference set* più vicine alla j -esima unità nello spazio multidimensionale definito dalle variabili ausiliarie:

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

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 nelle k unità del *reference set* più vicini alla j -esima unità) o, come avviene più frequentemente, può essere calcolato in modo inversamente proporzionale alla distanza multidimensionale tra la j -esima unità e ciascuna delle k unità del *reference set* a essa più vicine.

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

All'aumentare di k tende in genere ad aumentare l'accuratezza della stima \tilde{y}_j . Con valori alti di k la variabilità nei valori stimati per le singole unità tende però a essere minore di quella reale a causa dell'effetto di livellamento prodotto dalla ponderazione. In genere, per la maggior parte di applicazioni di questa procedura su immagini telerilevate ad alta risoluzione possono essere orientativamente consigliati valori di k compresi tra 1 e 15.

La scelta delle variabili ausiliarie, del tipo di distanza multidimensionale e di k è spesso 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 ciascuna j -esima unità del *reference set* con l'accortezza di escludere, ai fini della stima stessa, il valore vero y_j corrispondente a quell'unità: si ottengono così n valori stimati \hat{y}_j che confrontati con i corrispondenti valori veri y_j permettono di valutare l'accuratezza delle stime prodotte.

Sulla base dei risultati della procedura LOO viene definita la configurazione localmente ottimale dell'algoritmo k -NN (in termini di variabili ausiliarie considerate, tipo di distanza multidimensionale, valore di k) capace di fornire le stime più accurate sulle unità del *target set*.

Aree di studio

La sperimentazione è stata condotta in Toscana, su due aree di studio rappresentative di ambienti forestali diversi, quello appenninico e quello litoraneo. In particolare, il metodo k -NN è stato applicato nelle Riserve Naturali Statali Biogenetiche di Vallombrosa (Firenze) e dei Tomboli di Cecina (Livorno).

La foresta di Vallombrosa copre una superficie di 1273 ettari a quote comprese tra 470 e 1440 metri s.l.m. È caratterizzata da boschi di abete bianco (669 ettari), faggio (197 ettari), pinete di pino nero e laricio (153 ettari), e boschi misti di latifoglie (120 ettari). Inoltre, sono presenti soprassuoli di douglasia (30 ettari), abete rosso (5 ettari) e castagno (38 ettari). Secondo la classificazione di Pavari la foresta rientra nella zona fitoclimatica del *Castanetum* per tutta l'area al di sotto dei 1000 metri e nella zona del *Fagetum* al di sopra di tale altitudine.

La foresta dei Tomboli di Cecina occupa una superficie di 406 ettari a nord e a sud della foce del fiume Cecina, distribuita su una fascia lunga circa 15 chilometri, di larghezza variabile tra 100 e 600 metri. La foresta è composta prevalentemente da pinete di pino domestico con presenza di leccio e macchia (288 ettari); la fascia di protezione a macchia e pino marittimo si è ridotta nel corso degli anni a causa dall'erosione costiera e oggi si estende su una superficie di 82 ettari. Secondo la classificazione fitoclimatica di Pavari la foresta ricade nella zona del *Lauretum*, sottozona media.

Dati a terra, popolazione sperimentale e dimensionamento del campione

I dati a terra provengono dalle campagne di rilievo condotte nel periodo 2004-2006 nell'ambito della realizzazione dei piani di gestione delle Riserve Naturali Statali Biogenetiche di Vallombrosa e dei Tomboli di Cecina (Ciancio, 2007a; Ciancio, 2007b). Complessivamente a Vallombrosa sono state rilevate 85 aree di saggio circolari di raggio 20 metri, distribuite tra i popolamenti forestali rappresentativi delle tipologie colturali presenti in foresta. A Cecina, il rilievo quantitativo ha interessato solo i popolamenti di pino domestico: complessivamente sono state realizzate 43 aree di saggio circolari di raggio 20 metri. In entrambi i casi il centro delle aree è stato georeferenziato con GPS con precisione sub metrica. Nei due piani di gestione la provvigione delle diverse tipologie forestali è stata determinata attraverso i dati misurati nelle unità di rilevamento a terra. Per le unità colturali aventi un'area di saggio al loro interno, la provvigione è stata posta uguale a quella misurata nelle aree di saggio, nelle altre unità colturali la provvigione è stata determinata a partire dai dati misurati nelle aree di saggio, tenendo conto dei valori di incremento percentuale indicati nelle tavole alsometriche locali. Tali dati sono stati poi confrontati, a parità di età, densità e fertilità, con la massa riportata nelle tavole alsometriche, riscontrando una buona corrispondenza.

Ai fini di questo studio a Vallombrosa la popolazione sperimentale N è costituita da 285 unità colturali, per complessivi 535,84 ettari, tutte a dominanza di abete bianco: 38 unità (113,94 ettari) costituiscono il *reference set* e 247 unità (421,90 ettari) costituiscono il *target set*.

A Cecina l'indagine si è focalizzata sulle pinete di pino domestico di età uguale o superiore a 70 anni. La popolazione sperimentale N è così costituita da 49 unità colturali (218,67 ettari): di queste, 25 (112,08 ettari) costituiscono il *reference set* e 24 (106,59 ettari) il *target set*.

In entrambi le aree di studio il *reference set* è costituito esclusivamente da unità colturali aventi al loro interno un'area di saggio. Le stime prodotte nei *target set* con il metodo *k*-NN sono confrontate con i valori riportati nei piani di gestione.

Dati satellitari e variabili ancillari

Le caratteristiche spettrali del territorio investigato sono state derivate da una scena Landsat 5 TM acquisita nell'estate del 2006. Dalla intera scena satellitare sono state tagliate due finestre in corrispondenza delle aree di studio che successivamente sono state ortorettificate utilizzando un modello digitale del terreno con passo di 30 metri e oltre 40 *Ground Control Points* individuati per confronto su ortofoto digitali a un metro di risoluzione geometrica. L'errore commesso in fase di georeferenziazione è risultato inferiore alla dimensione del pixel Landsat (30 metri). In entrambi i casi la copertura satellitare sulle aree di indagine era libera da nuvole e da relative ombre.

Le variabili ancillari utilizzate ai fini della stima con metodo *k*-NN derivano da cartografie tematiche realizzate in ambiente GIS sulla base dei dati di età, fertilità, quota, pendenza e esposizione contenuti nei piani di gestione delle due Riserve (Tabella 1). Altre variabili ausiliarie sono state prodotte per combinazione delle bande satellitari, come nel caso degli indici IRI (*InfraRed Index*) calcolato come rapporto $(Tm4-Tm5)/(Tm4+Tm5)$ e NDVI (*Normalized Difference Vegetation Index*) calcolato come rapporto $(Tm4-Tm3)/(Tm4+Tm3)$.

<i>Dati satellitari e variabili ancillari</i>	<i>Cecina</i>	<i>Vallombrosa</i>
Landsat 5 Tm1	x	x
Landsat 5 Tm2	x	x
Landsat 5 Tm3	x	x
Landsat 5 Tm4	x	x
Landsat 5 Tm5	x	x
Landsat 5 Tm7	x	x
Iri	x	x
Ndvi	x	x
Età	x	x
Distanza dalla costa	x	
Quota		x
Fertilità		x
Pendenza		x
Esposizione		x

Tabella 1 – Elenco dei dati satellitari e delle variabili ancillari utilizzati come variabili ausiliari dell'algoritmo di stima k-NN

Configurazione dell'algoritmo di stima e valutazione dell'accuratezza

Nella presente sperimentazione le variabili acquisite da telerilevamento sono state ottenute a livello di unità colturale come media dei valori di *digital number* misurati per ciascuna banda considerata nei pixels inclusi in ciascuna unità colturale.

In fase di ottimizzazione con LOO sono stati testati valori di *k* compresi tra 1 e 15 e cinque diversi metodi di calcolo della distanza multidimensionale: euclidea (De Maesschalck et al., 2000), di Mahalanobis (Holmstrom et al., 2001), modificata con pesi fuzzy (Maselli, 2001), modificata con metodo regressivo multivariato e con pesi non parametrici (Maselli et al., 2005).

Sulla base dei risultati ottenuti dal LOO sono state individuate le configurazioni localmente ottimali dell'algoritmo *k*-NN per la stima della massa legnosa delle unità colturali del *reference set*. L'algoritmo *k*-NN in tal modo ottimizzato è stato quindi applicato per stimare il valore di massa legnosa delle unità colturali appartenenti al *target set*. Successivamente i valori stimati sono stati aggregati a livello dei comparti colturali e di intera foresta. L'accuratezza delle stime prodotte sul *target set* è stata valutata attraverso il calcolo dello scarto quadratico medio e dello scarto medio,

ottenuti confrontando i valori di massa totale (m^3) riportati nei piani di gestione delle Riserve di Vallombrosa e Cecina con quelli derivati con metodo k -NN per ogni unità colturale; i valori dello scarto quadratico medio relativo e dello scarto medio relativo sono stati calcolati come proporzione del valore medio delle stime (Mäkelä, Pekkarinen, 2004).

Lo stesso confronto è stato ripetuto anche a livello di comparti colturali e di intera foresta.

Risultati

I risultati della fase di ottimizzazione con LOO dell'algoritmo k -NN sul *reference set* sono riportati in Figura 1. A Cecina, la configurazione migliore è stata ottenuta con la distanza modificata con metodo regressivo multivariato e con un valore di k uguale a 6; a Vallombrosa il migliore risultato è stato ottenuto con la distanza modificata con pesi fuzzy e con un valore di k uguale a 2.

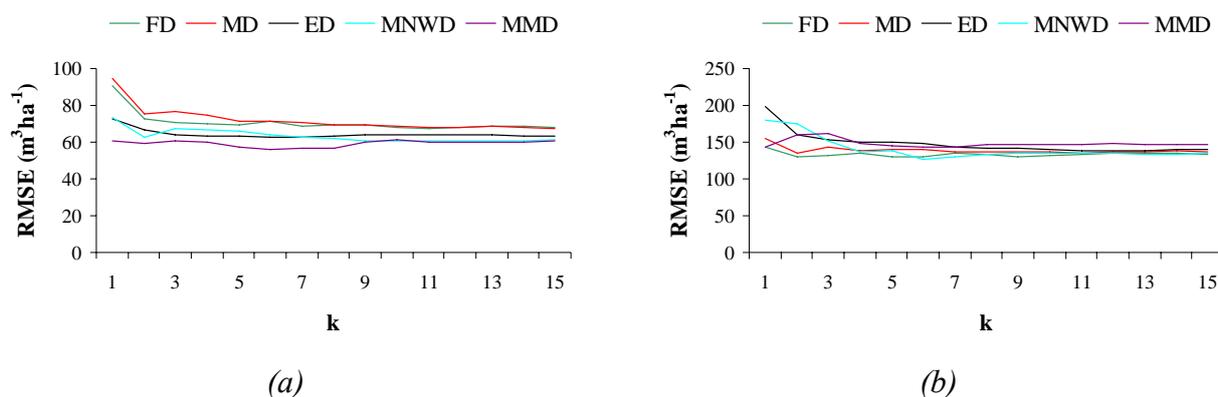


Figura 1 – Risultato del leave-one-out (LOO): a. Cecina; b. Vallombrosa (ED: distanza euclidea; FD: distanza modificata con pesi fuzzy; MD: distanza di Mahalanobis; MMD: distanza modificata con metodo regressivo multivariato; MNWD: distanza modificata con pesi non parametrici)

I risultati dell'applicazione dell'algoritmo k -NN in tal modo ottimizzato per la stima della provvigione legnosa sulle unità colturali dei *target set* delle due aree di studio sono riportati in Tabella 2.

Nelle due aree di studio lo scarto quadratico medio relativo e lo scarto medio relativo calcolati a livello di unità colturale sono risultati rispettivamente del 20-30% e del 10%.

I valori di scarto quadratico medio relativo si riducono del 19% a Cecina e del 47% a Vallombrosa quando i valori di massa totale delle unità colturali sono aggregati a livello dei comparti colturali.

Quando i valori di massa totale delle unità colturali di pino domestico e di abete bianco sono stati aggregati a livello di intera foresta, sia a Cecina sia a Vallombrosa si è registrata una sovrastima del 12% rispetto valore calcolato con i dati riportati nei rispettivi piani di gestione.

	CECINA				VALLOMBROSA			
	RMSE m^3	RMSE _r %	bias m^3	bias _r %	RMSE m^3	RMSE _r %	bias m^3	bias _r %
Unità colturali	285	21	148	11	362	30	124	10
Comparti colturali	526	17	336	11	1577	16	955	10

Tabella 2 - Accuratezza delle stime di massa totale a livello delle unità colturali e dei comparti colturali (RMSE: scarto quadratico medio; RMSE_r: scarto quadratico medio relativo; bias: scarto medio; bias_r: scarto medio relativo)

Conclusioni

Nei casi di studio esaminati le accuratezze di stima con metodo k -NN della massa legnosa di popolamenti di pino domestico e abete bianco presentano valori simili a quelli ottenuti in esperienze analoghe condotte nel nostro Paese (Bertini et al., 2007). Ulteriori miglioramenti sono attesi, almeno a Vallombrosa, aumentando la dimensione del campione a terra (*reference set*).

In termini assoluti i valori di scarto quadratico medio ottenuti a Cecina sono inferiori a quelli registrati a Vallombrosa, ma questo in parte è dovuto al fatto che le pinete litoranee di pino domestico sono caratterizzate, in genere, da provvigioni inferiori rispetto a quelle delle abetine appenniniche. Infatti, in termini relativi gli errori riscontrati nelle due aree di studio sono simili.

Sebbene si ritenga necessario approfondire la sperimentazione sulla potenzialità del metodo k -NN per applicazioni a fini assestamentali, i risultati di questo lavoro consentono di esprimere alcune prime conclusioni. In tutti quei casi in cui la derivazione di parametri biofisici quantitativi dei popolamenti forestali non possa essere basata sul cavallettamento totale per l'eccessivo costo di questo rilievo, il metodo k -NN appare essere una valida alternativa ai metodi di stima tradizionali. L'algoritmo k -NN, ottimizzato localmente secondo la metodologia descritta, anche basato su immagini telerilevate di non altissima risoluzione (come quelle Landsat), si è infatti dimostrato capace di produrre stime comparabili con quelle operativamente utilizzate nei piani di gestione delle due Riserve indagate. Pertanto, visti i risultati conseguiti, si ritiene che in futuro il metodo k -NN potrà essere applicato a Vallombrosa e a Cecina in fase di revisione e aggiornamento dei rispettivi piani di gestione per la stima della provvigione legnosa delle abetine e delle pinete di pino domestico.

Ringraziamenti

Gli Autori ringraziano il gruppo di lavoro che ha eseguito i rilievi a terra nell'ambito dei piani di gestione di Vallombrosa e dei Tomboli di Cecina: Stefano Bologna, Giuseppe Bonanno, Francesca Bottalico, Paola Brundu, Alessio Burini, Valentina Cappelli, Carmela Madeo, Federico Maetzke, Davide Melini, Ilaria Napoli, Franco Piemontese.

Si ringrazia inoltre il Prof. Piermaria Corona dell'Università della Tuscia per la revisione del testo inerente la descrizione del metodo k -NN.

Bibliografia

Bertini R., Chirici G., Corona P., Travaglini D. (2007), *Confronto di metodi parametrici e non-parametrici per la spazializzazione della provvigione legnosa tramite integrazione di misure a terra, dati telerilevati e informazioni ancillari*. Forest@ 4 (1): 110-117.

Ciancio O. (2007a), *La Riserva Naturale Statale Biogenetica di Vallombrosa. Piano di Gestione e Silvomuseo: 2006-2025*. In corso di pubblicazione.

Ciancio O. (2007b), *Piano di Gestione della Riserva Naturale Statale Biogenetica dei Tomboli di Cecina 2007-2021*.

Corona P. (2007), *Metodi di inventariazione delle masse e degli incrementi legnosi in assestamento forestale*. Aracne Editrice, Roma.

De Maesschalck R., Jouan-Rimbaud D., Massart D.L. (2000), *The Mahalanobis distance*. Chemometrics and Intelligent Laboratori System, 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*. Scand. J. For. Res., 16: 67-78.

Mäkelä H., Pekkarinen A. (2004), *Estimation of forest stand volume by Landsat TM imagery and stand-level field-inventory data*. Forest Ecology and Management, 196: 245-255.

Maselli F. (2001), *Extension of environmental parameters over the land surface by improved fuzzy classification of remotely sensed data*. International Journal of Remote Sensing, 17: 3597-3610.

Maselli F., Chirici G., Bottai L., Corona P., Marchetti M. (2005), *Estimation of Mediterranean forest attributes by the application of k -NN procedures to multitemporal Landsat ETM+ images*. International Journal of Remote Sensing, 17: 3781-3796.