

Stima del *Leaf Area Index* da satellite con modelli OLS (*Ordinary Least Square*) ed ANN (*Artificial Neural Network*)

Flavio BORFECCHIA (*), Ludovica GIORDANO (**), Luigi DE CECCO (*),
Claudia TROTTA (**), Gabriele SCHINO (**)

(*) Dipartimento Ambiente, Cambiamenti Globali e Sviluppo Sostenibile - ACS, Unità Osservazioni Aerospaziali della Terra e Sistemi Informativi Territoriali, ENEA CR-Casaccia, Via Anguillarese 301, 00123 S. Maria di Galeria (RM) 0630486042/4931; borfecchia@casaccia.enea.it

(**) Dipartimento Biotecnologie, Agroindustria e protezione della Salute - BAS, Gruppo Lotta alla Desertificazione, ENEA CR-Casaccia, Via Anguillarese 301, 00123 S. Maria di Galeria (RM), 0630483316

Riassunto

In questo lavoro sono presentati alcuni risultati dell'attività di ricerca per la stima estensiva di parametri biofisici della vegetazione tramite tecniche di telerilevamento satellitare, in aree del sud Italia a rischio di desertificazione. In particolare, si descrive la messa a punto di modelli per la stima del *Leaf Area Index* (LAI), su aree di vegetazione naturale e semi-naturale in zone caratterizzate da forte impatto dell'agricoltura intensiva e con possibilità di stress idrico dovuto alle diminuite precipitazioni. Per tener conto dell'estrema eterogeneità degli ecosistemi studiati si è fatto uso dei dati multispettrali ad alta risoluzione acquisiti tramite il sensore satellitare IKONOS, con 4 m di risoluzione geometrica. Le campagne di rilievo per le misure a terra sono state effettuate nel luglio 2004 e nel maggio 2005. Sono stati inizialmente testati modelli *Ordinary Least Squares* (OLS) utilizzando sia vari indici di vegetazione, sia regressioni multiple con i dati di riflettanza delle 4 bande di acquisizione IKONOS, dopo correzione degli effetti atmosferici. In generale i modelli con le regressioni multiple hanno dato buoni risultati ($r^2_{\text{corretto}}=0.78$, $N=55$, $P<0.00001$) per il 2004, mentre per il 2005 la capacità predittiva è risultata inferiore ($r^2_{\text{corretto}}=0.30$, $N=56$, $P=0.0002$). Avendo quindi ipotizzato, per il 2005, un aumento delle componenti non lineari nelle relazioni tra riflettanze e LAI, sono stati testati modelli alternativi tramite *Artificial Neural Network* (ANN). Con questi modelli sono state ottenute correlazioni elevate in entrambe gli anni (2004: $r^2_{\text{corretto}}=0.88$, $N=55$, $P<0.00001$; 2005: $r^2_{\text{corretto}}=0.84$, $N=56$, $P<0.00001$). Successivamente i modelli ANN sono stati ulteriormente migliorati introducendo anche variabili relative alla *tessitura*. In questo contesto, l'approccio con tecniche ANN ha mostrato, complessivamente, una efficacia superiore rispetto ai più tradizionali modelli regressivi.

Abstract

This paper reports the results of research activities aimed at estimating the spatial distribution of vegetation biophysical parameters by means of satellite remote sensing techniques, in areas affected by desertification risk. This study focused on the development and validation of semi-empirical models for retrieving *Leaf Area Index* (LAI). The study areas, characterised by semi-natural vegetation, were located in Southern Italy where frequent water stress is due to the changing precipitation regime and increased water demand by intensive agricultural activities. Two study areas, both in the Basilicata region, were selected: the first, located along the Ionian coast, is mainly characterised by an hygrophilous forest, whereas the second is situated in an inner zone with a mean altitude of 500-600 m a.s.l. and is covered by typical Mediterranean vegetation, such as maquis, mixed broadleaf forest and live oak forest. In order to account for the extremely heterogeneous ecosystems, high resolution IKONOS data were used. In 2004 and 2005 field surveys were carried

out in order to provide ground truth LAI data. In both years, the acquisition of the corresponding IKONOS images was originally planned to be nearly synchronous with ground sampling. However, cloud cover and satellite temporal resolution constraints resulted in satellite images being acquired about one month after ground sampling.

First, OLS (*Ordinary Least Squares*) models using atmospherically corrected IKONOS reflectance data were applied. In particular, simple regressions using LAI as the dependent variable and several spectral Vegetation Indices (VIs) as the independent variable, as well as multiple regressions using multiple spectral bands, were performed. The latter provided rather good results ($r^2=0.78$, $N=55$, $P<0.00001$) for the 2004 data set including both study areas, whereas for 2005 data a considerable decrease ($r^2=0.30$, $N=56$, $P=0.0002$) of the models' predictive capacity was observed. Then, hypothesizing an increase of non-linear components in the LAI-reflectances relationships in 2005, ANN (*Artificial Neural Network*) models were tested. In particular, a MLP (*MultiLayer Perceptron*) ANN version, using a single "hidden" layer and a back-propagation learning algorithm was used. This simple configuration allowed to significantly improve results ($r^2=0.88$, $N=55$, $P<0.00001$ for 2004 and $r^2=0.84$, $N=56$, $P<0.00001$ for 2005). Subsequently, ANN models were further improved by including some additionally input variables. In this context, the ANN-based approach showed to be more effective than the much more widely used regression models.

Introduzione

Le misure *in situ* delle grandezze biofisiche della vegetazione costituiscono la base per lo sviluppo e la calibrazione di modelli da utilizzare per la stima estensiva e sinottica delle distribuzioni spaziali e temporali delle grandezze stesse a partire da dati satellitari. Fra le variabili di maggiore interesse vi è il *Leaf Area Index* o LAI, il quale gioca un ruolo preminente nella caratterizzazione degli ecosistemi vegetali ed è un importante parametro di input di molti modelli climatologici (Sellers et al. 1997), idrologici (D'Urso et al., 1999) ed ecologici (Chen et al, 2000).

L'utilizzo di modelli semiempirici basati su regressione Ordinary Least Square (OLS) per la calibrazione dei dati telerilevati tramite misure a terra, risulta notevolmente diffuso ed adeguato, in una molteplicità di situazioni, a fornire stime affidabili delle distribuzioni spaziali delle variabili biofisiche d'interesse, partendo solitamente da dati di riflettanza utilizzati in forma di singole bande o di loro combinazioni in indici spettrali di vegetazione (Lawrence & Ripple, 1998, Colombo et al., 2003, Soudani et al., 2006). Tuttavia permangono l'onere e le ovvie difficoltà ad ottenere dati a terra che garantiscano un'efficace calibrazione dei modelli semiempirici da utilizzare per la spazializzazione delle variabili di interesse. Un ulteriore limite dell'approccio OLS sta nella scarsa stabilità spazio-temporale delle relazioni osservate che ne limita la loro estrapolazione. Spesso l'impiego dei dati di riflettanza sotto forma di indici spettrali di vegetazione, basati prevalentemente sulle riflettanze nel rosso (Red) e vicino infrarosso (NIR) e formulati per minimizzare gli effetti di rumore dovuti all'atmosfera ed alla topografia, limita ulteriormente la capacità predittiva a causa della scarsa dimensionalità dello spazio delle variabili indipendenti (riflettanze) usate.

Fatto salvo il vantaggio dell'impiego del telerilevamento, in relazione alla possibilità di ottenere mappe estensive con un dispendio di risorse sostenibile, l'incidenza di questi aspetti negativi si accentua laddove la vegetazione è caratterizzata da notevole eterogeneità e frammentazione, situazione frequente in ambito mediterraneo ed in particolare nelle aree sottoposte a disturbo (pascolo, incendi, attività agricole, urbanizzazione ecc.). Per ovviare a queste problematiche, sono stati sviluppati nel tempo diversi approcci, compresa una modellizzazione fisica del processo di trasferimento radiativo della *canopy* vegetale (Jacquemoud et al., 1995; Qin, Liang, 2000; Fang et al., 2003). In particolare, tra i diversi metodi testati, l'utilizzo di reti neurali artificiali (ANN) ha trovato, soprattutto negli anni recenti, sempre maggiore spazio grazie alla loro capacità di descrivere dipendenze complesse tra i dati senza la necessità di dover fare assunzioni a priori sui dati o sul tipo di relazioni (Foody & Boyd, 2002; Jensen, 2000, Walthall et al., 2004; Ingram et al., 2005).

Nelle sezioni successive sono presentati i risultati di un'attività di ricerca che ha riguardato la messa a punto di modelli OLS e ANN con i quali stimare distribuzioni spaziali del LAI a partire da dati telerilevati da satellite e rilievi *in situ*, effettuati in due anni consecutivi (2004 e 2005) su aree di vegetazione naturale e semi-naturale in zone del sud Italia.

Aree di studio ed acquisizione dei dati

Le due aree oggetto di questo studio ricadono nel settore ionico della Basilicata, nella Provincia di Matera. La prima, denominata bosco Pantano, è un'area costiera, pressoché pianeggiante, di circa 500 ha, caratterizzata dalla presenza prevalente di bosco igrofilo e macchia mediterranea. L'altra (Monte Coppolo), di circa 800 ha, è posta sui rilievi collinari interni tra i 500 e gli 800 m s.l.m. e si differenzia per la presenza di macchia mediterranea, bosco caducifoglio e lecceta. Per tener conto dell'estrema eterogeneità degli ecosistemi investigati si è fatto uso dei dati telerilevati ad alta risoluzione spaziale (1 m pancromatico, 4 m multispettrale) acquisiti tramite il sensore satellitare IKONOS. A causa di vincoli legati alle condizioni di copertura nuvolosa e di programmazione dell'acquisizione i dati satellitari sono stati ripresi, in entrambi gli anni, dopo circa un mese rispetto all'esecuzione dei rilievi a terra. In particolare, le campagne di misura del LAI sono state effettuate nel luglio 2004 e nel maggio 2005 mentre le immagini IKONOS sono state acquisite rispettivamente il 2 agosto 2004 e il 20 giugno 2005. Il LAI è stato misurato mediante LAI-2000 *PCA-Plant Canopy Analyzer* (Li-COR, 1992) in 55 plot nel 2004 e 56 nel 2005, indipendenti, della dimensione di 12x12 m confrontabile con la risoluzione spaziale del dato multispettrale IKONOS. I plot sono stati localizzati tramite GPS Trimble *ProXRS* con una precisione, a seguito di correzione differenziale, di circa 1,5-2 m, e caratterizzati dal valore medio di una serie di letture ($n=4\div 6$) di LAI eseguite al loro interno. Una caratterizzazione approfondita delle aree di studio e del protocollo per le misure a terra è possibile trovarla in Trotta et al., 2005.

Elaborazione dei dati e risultati

I dati satellitari sono stati opportunamente pre-processati, ovvero corretti geometricamente, radiometricamente ed atmosfericamente, in modo da poter ricavare i valori di riflettanza in corrispondenza dei siti di misura del LAI. Per ogni banda IKONOS, i valori della riflettanza sono stati estratti su finestre di 3x3 pixel, corrispondenti ai 12x12 m² dei plot di misura, e ne è stato calcolato il valor medio.

Il set di dati analizzato si compone oltre che dei valori di riflettanza in ogni banda multispettrale, di alcuni parametri di *tessitura* ricavati dall'elaborazione dell'immagine pancromatica (*range, media, varianza, entropia, skewness*) nonché del valore dell'angolo di incidenza, $\cos(i)$, tra illuminazione solare e direzione di osservazione del sensore (derivato da DEM ad adeguata risoluzione per tener conto della topografia locale) ed infine dei valori di LAI misurati a terra.

Prima di analizzare le relazioni tra dati satellitari e dati misurati a terra, è stata eseguita l'analisi della covarianza (ANCOVA) sul LAI, utilizzando come variabile indipendente la variabile radiometrica (diverse variabili sono state testate a questo scopo, tra cui le singole bande e l'NDVI ($(nIR-red)/(nIR+red)$) per valutare eventuali interazioni tra anno e LAI e tra area di studio e LAI allo scopo di individuare il *dataset* da utilizzare nelle successive investigazioni. L'ANCOVA ha mostrato l'esistenza di una interazione significativa fra variabile radiometrica ed anno di osservazione, per esempio nel caso del NDVI: $F=41.71$, $df=1/105$, $P<0.0001$. Questo risultato ha messo in evidenza che la relazione fra i dati di riflettanza e di LAI non è la stessa nei due anni di osservazione, per cui è stato necessario sviluppare modelli separati per i due anni.

Un altro aspetto preso in considerazione è il cosiddetto effetto di saturazione (Broge et al., 2000) per cui ad alti valori di LAI la relazione tra LAI e NDVI (e/o riflettanza) tende ad essere logaritmica piuttosto che lineare. Di fatto, nelle nostre aree, avendo valori di copertura vegetale e quindi di LAI anche elevati (caratteristici di zone boschive e macchia mediterranea), utilizzando $\ln(LAI)$ anziché la variabile LAI nelle regressioni si sono ottenuti valori di correlazione più alti.

Gli approcci con OLS hanno riguardato sia modelli univariati, lineari e non, che multivariati.

Nel primo caso si sono cercate relazioni tra i valori della variabile biofisica misurata e quelli di diverse variabili radiometriche, tra cui le singole bande, varie loro combinazioni sotto forma di indici spettrali di vegetazione, quali l'NDVI, il SR (*Simple Ratio*: nIR/red), SAVI (*Soil Adjusted Vegetation Index*: $nIR-red)/(1+L)/(nIR*red*L$, $L=0.5$), il SARVI 2 (*Soil and Atmospherically Resistant Vegetation Index*: $2.5*(nIR-red)/(1+nIR+6*red-7.5/blue)$), nonché i parametri di tessitura sopra citati.

Nel caso dei modelli OLS multivariati sono state provate diverse combinazioni delle variabili a disposizione, includendo anche il dato relativo alle condizioni di illuminazione in funzione della topografia locale ($\cos(i)$). I risultati migliori sono stati ottenuti con le regressioni multivariate utilizzando tutte le bande ed il $\cos(i)$. E' da notare che l'introduzione dell'informazione tessiturale non ha apportato alcun miglioramento nei risultati. Inoltre, mentre nel 2004 i modelli ottenuti hanno mostrato una buona capacità predittiva ($r^2_{\text{corretto}}=0.78$, $N=55$, $P<0.00001$), nel 2005 si è avuto un calo notevole della percentuale di varianza spiegata ($r^2_{\text{corretto}}=0.30$, $N=56$, $P=0.0002$).

Ipotizzando quindi un aumento delle componenti non lineari dovute ai differenti periodi dei rilievi, e con l'obiettivo di migliorare la stima, in particolare per i dati acquisiti nel 2005, sono stati elaborati modelli alternativi tramite ANN. In analogia con lo schema delle connessioni cerebrali degli organismi biologici, il modello ANN è basato su strutture di elementi interconnessi detti *neuroni*, la cui funzione di risposta è assimilabile ad un'espressione regressiva multivariata. L'introduzione di strati di neuroni di tipo "*hidden*", a funzione di attivazione tipicamente sigmoidale, mette in grado il sistema di simulare adeguatamente risposte anche non lineari, dopo opportuno training per il calcolo dei vari fattori di peso corrispondenti alle connessioni. E questo senza alcuna ipotesi sulla forma della funzione approssimata tramite risposte impulsive dei singoli neuroni.

Nel nostro caso è stata utilizzata una versione di ANN di tipo *MultiLayer Perceptron* (MLP), comunemente utilizzata nell'ambito applicativo del telerilevamento satellitare (Carvalho, 2001) e risultata particolarmente efficace in studi analoghi (Jensen, Binford, 2004). Il meccanismo di apprendimento utilizzato è stato quello di tipo *back-propagation*, che altri studi hanno dimostrato essere preferibile quando occorre modellare relazioni non lineari (Jensen, 2000). Al fine di sviluppare un modello analitico non eccessivamente complicato ed oneroso da implementare per la spazializzazione dei dati, si è scelto di utilizzare un numero di neuroni non troppo elevato. Il numero di neuroni *hidden* è stato quindi inizialmente posto eguale a quello delle variabili in ingresso, salvo diminuirlo o aumentarlo a seguito di una prima fase di addestramento, cancellando i neuroni che risultavano non attivati od introducendone di nuovi se necessario. Dopo la variazione del numero di neuroni veniva ripetuto il training sino alla stabilizzazione dell'errore quadratico medio tra i dati LAI stimati e misurati (differenza tra RMSE di cicli successivi $<$ di 10^{-4}). In particolare sono state testate due reti MLP. La prima consta di un layer di neuroni in input costituito da 5 neuroni, corrispondenti alle 4 bande derivate da IKONOS più il $\cos(i)$ (ovvero alle stesse variabili utilizzate nel caso dei risultati migliori ottenuti tramite OLS), mentre nella seconda sono stati aggiunti altri 5 neuroni, corrispondenti ai 5 parametri di tessitura sopra citati. Utilizzando 5 neuroni in input è stato ottenuto un primo miglioramento nella correlazione ($r^2_{\text{corretto}}=0.88$, $N=55$, $P<0.00001$, anno 2004; $r^2_{\text{corretto}}=0.84$, $N=56$, $P<0.00001$, anno 2005) rispetto a quella ottenuta tramite modello OLS. Successivamente il modello ANN è stato ulteriormente migliorato introducendo le variabili tessiturali ($r^2_{\text{corretto}}=0.99$, $N=55$, $P<0.00001$, 2004; $r^2_{\text{corretto}}=0.95$, $N=56$, $P<0.00001$, 2005) risultate invece scarsamente efficaci nei modelli OLS. In fig. 1 si riporta la mappa di LAI ottenuta dal modello analitico fornito dall'applicazione della rete neuronale a 10 neuroni.

Dalle analisi è inoltre risultato che l'importanza relativa delle singole variabili indipendenti nel modello multivariato OLS risulta differente da quella nel modello ANN in considerazione della capacità di quest'ultimo di tener conto di dipendenze non lineari.

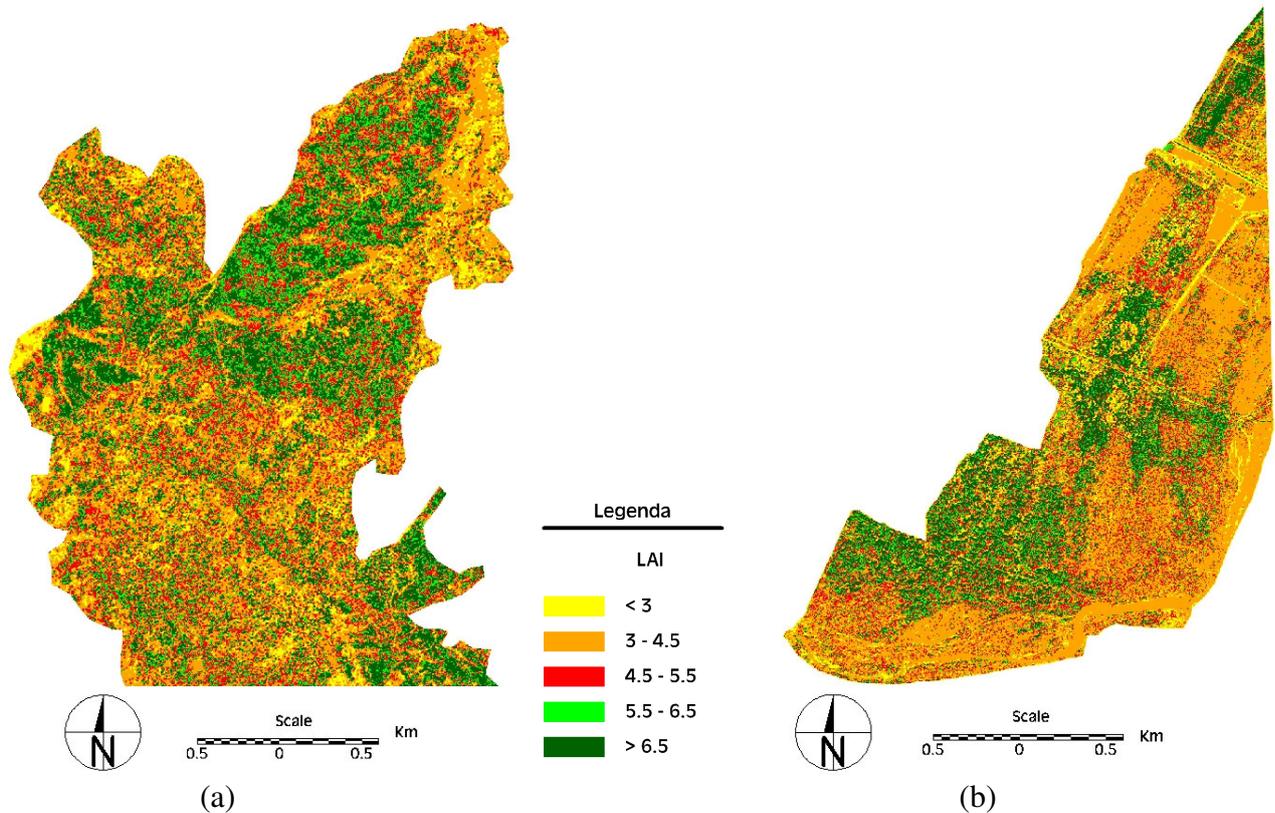


Figura 1- Distribuzione del LAI nelle due aree di studio Monte Coppolo (a) e Pantano (b) ottenute tramite il modello ANN.

Conclusioni

L'utilizzo di modelli OLS multivariati per la stima del LAI di vegetazione seminaturale ha dato buoni risultati per i dati relativi al 2004, rilevati durante la campagna a terra effettuata nel mese di luglio. Invece per i rilievi del 2005, acquisiti durante il mese di maggio, si è avuto un calo notevole della loro capacità predittiva.

L'utilizzo degli stessi dati in modelli di tipo ANN ha permesso di ottenere un aumento significativo del livello di correlazione per entrambe le campagne. Già con la configurazione di base delle variabili in input (4 bande + $\cos(i)$) sono stati ottenuti dei miglioramenti nelle correlazioni. Inoltre, i modelli ANN sono stati ulteriormente migliorati utilizzando le variabili ausiliarie tessiturali che erano risultate scarsamente efficaci nei modelli OLS.

In questo contesto complessivamente l'approccio con tecniche ANN ha dimostrato un'efficacia superiore rispetto ai più tradizionali modelli regressivi OLS, in particolare nel 2005, quando a causa di una discrepanza temporale significativa risulta debole l'ipotesi di dipendenza lineare tra riflettanze e LAI. Tale condizione potrebbe derivare dal periodo di campionamento a terra nel maggio 2005 coincidente con una fase di sviluppo vegetativo ancora significativo nella maggior parte delle associazioni vegetali presenti nelle aree di test, che ha comportato una effettiva discordanza tra quanto misurato in situ e quanto osservato da satellite un mese più tardi. Viceversa è ipotizzabile che questo sviluppo fosse già ultimato nel 2004 essendo i rilievi in quell'anno relativi al mese di luglio (e quindi a fronte di uno sfasamento temporale simile, di circa una mese, si è avuta comunque maggior consistenza tra misure a terra e dato telerilevato).

Poiché lo sfasamento tra i dati osservati e misurati a terra è una condizione piuttosto frequente, va posta una particolare attenzione nella scelta del periodo di studio, privilegiando quando possibile una fase avanzata dello sviluppo vegetativo per la maggior parte delle specie presenti, in modo da

far sì che ciò che viene misurato in campo non sia, in termini di biomassa, troppo differente da ciò che viene osservato da satellite in un momento leggermente diverso.

Bibliografia

- Broge N.H., E. Leblanc, (2000), “Comparing prediction power and stability of broadband and hyperspectral vegetation indices for estimation of green leaf area index and canopy chlorophyll density”, *Remote Sensing of Environment*, 76, 156-172.
- Carvalho, L. M. T. (2001), *Mapping and monitoring forest remnants: A multiscale analysis of spatio-temporal data*. Doctoral thesis. Wageningen University, Wageningen, The Netherlands.
- Chen. W. Chen J.M. and Chilar. J., (2000), “Integrated terrestrial ecosystem carbon-budget model based on changes in disturbance, climate, and atmospheric chemistry”. *Ecol. Model.* 135:55-79.
- Colombo R., Bellingeri D., Fasolini D., Marino C.M. (2003), “Retrieval of leaf area index in different vegetation types using high resolution satellite data”. *Remote Sensing of Environment*, 86: 120–131.
- D'Urso G., Menenti M., Santini A.,(1999), “Regional application of one-dimensional water flow models for irrigation management”. *Agricultural Water Management*, 40: 291±302.
- Fang HL, Liang SL, Kuusk A. (2003), “Retrieving leaf area index using a genetic algorithm with a canopy radiative transfer model”, *Remote Sensing of Environment*, 85 (3): 257-270.
- Foody G. M., D. S. Boyd (2002), “Sharpened Mapping of Tropical Forest Biophysical Properties from Coarse Spatial Resolution Satellite Sensor Data”, *Neural Computing & Applications*, 11 (1): 62-70.
- Jacquemoud, S., Baret, F., Andrieu, B., Danson, F. M., & Jaggard, K. (1995), “Extraction of vegetation biophysical parameters by inversion of the PROSPECT + SAIL models on sugar beet canopy reflectance data. Application to TM and AVIRIS sensors”, *Remote Sensing of Environment*, 52: 163–172.
- Jensen, J. R. (2000), *Remote Sensing of the Environment: An Earth resources perspective*, Upper Saddle River, NJ: Prentice-Hall.
- Jensen R. R., Binford M.W. (2004), “Measurement and comparison of Leaf Area Index estimators derived from satellite remote sensing techniques”, *Int. J. Remote Sensing*, vol. 25, NO. 20: 4251–4265
- Lawrence RL, Ripple WJ (1998), “Comparisons among vegetation indices and bandwise regression in a highly disturbed heterogeneous landscape: Mount St Helens, Washington”, *Remote Sensing of Environment*, 64: 91–102
- LI-COR (1992), *LAI-2000 Plant Canopy Analyser: Operating Manual*. Lincoln. NE. USA.
- Qin W. and Liang S. (2000), “Plane-parallel canopy radiation transfer modelling: Recent advances and future directions”, *Remote Sensing of Environment*, 18: 281-306.
- Sellers P. J., Dickinson R. E., Randall D. A., Betts A. K., Hall F. G., Berry J. A., Collatz G. J., Denning A. S., Mooney H. A., Nobre C. A., Sato N., Field C. B., Henderson-Sellers A. (1997). “Modeling the Exchanges of Energy, Water and Carbon Between Continents and the Atmosphere”, *Science*, 275. 502–509.
- Soudani K., Christophe F., Le Maire G., Le Dantec V., Dufrên E. (2006), “Comparative analysis of IKONOS, SPOT, and ETM+ data for leaf area index estimation in temperate coniferous and deciduous forest stands”, *Remote Sensing of Environment*, 102: 161–175.
- Trotta C., Giordano L., Menegoni P., Borfecchia F., De Cecco L. (2005), “Leaf area index retrieval using vegetation indices derived from very high resolution remotely sensed data”, *Colloques Phytosociologiques*, XXIX (in stampa).
- Walthall C., W. Dulaney, M. Anderson, J. Norman, H. Fang, S. Liang, (2004), “A comparison of empirical and neural network approaches for estimating corn and soybean leaf area index from Landsat ETM+ imagery”, *Remote Sensing of Environment*, 92:465–474.