La propagazione dell'errore negli indici di vegetazione derivati da immagini multispettrali: NDVI e PVI a confronto

Enrico Borgogno Mondino (*)

(**) Dipartimento DEIAFA, Università di Torino, Via L. da Vinci 44, 10095 Grugliasco (TO), tel: 011 6705523, fax: 011 6705516, enrico.borgogno@unito.it

Riassunto

Nell'ambito del telerilevamento satellitare (e non), finalizzato al monitoraggio della componente vegetazionale, grande successo riscuotono gli indici di vegetazione. Al di là della filosofia che ciascuno di essi sottende il problema che sollevano, spesso trascurato, è di come essi propaghino l'inevitabile incertezza di misura delle riflettanze da cui dipendono e di come poi questa si propaghi ulteriormente lungo le differenze di indici successivi lungo serie temporali. La soluzione del primo quesito consente di definire la sensibilità di apprezzamento dell'indice, cioè di stabilire quali differenze siano significative tra pixel appartenenti alla stessa scena. La risposta al secondo quesito consente invece di definire quali pixel siano davvero mutati nel tempo, cioè quale differenza di indice, calcolata sullo stesso pixel di scene diverse, sia da ritenersi significativa. Un indice può essere considerato una grandezza indiretta derivabile da misure dirette: le radianze al sensore, l'irradianza solare e gli eventuali fattori correttivi topografici e d'atmosfera. Il rigore scientifico imporrebbe di investigare dettagliatamente queste voci, aggregandole nelle forme previste per il calcolo delle riflettanze ed investigando come le loro varianze (assunte come incertezza/sensibilità di misura) si propaghino all'interno delle formulazioni adottate. Non è al momento scopo di questo lavoro spingersi a questo livello di dettaglio. Le esperienze condotte dimostrano come interessanti considerazioni possono già essere fatte affrontando il problema a partire dalle riflettanze ed ipotizzando, per le stesse, una ragionevole varianza che le caratterizzi. In questo lavoro due indici, l'NDVI (*Normalized Difference Vegetation Index*) e il PVI (*Perpendicular Vegetation Index*) rappresentativi rispettivamente delle due famiglie *slope-* e *distance-based* vengono considerati evidenziandone: a) i comportamenti rispetto alla propagazione della varianza delle riflettanze lungo la loro formulazione; b) i comportamenti lungo la loro differenza nel tempo. L'indagine è stata condotta mediante applicazione della legge di Propagazione della Varianze per variabili indipendenti supposte decorrelate.

Abstract

Many remote sensing applications make a wide use of vegetation indices and a huge number of users is very familiar with them. Nevertheless important issues, concerning their correct use, are often neglected: which is their precision? Is it possible to estimate it? How does the uncertainty related to the reflectances defining them act inside the index formulas? From a practical point of view the answer to these questions permits to define which index differences can be considered meaningful and which not over the same scene or along time series. The proposed approach is based on the statistical concept of the Variance Propagation Law describing how the error of direct measurements (reflectances) conditions the precision of those indirect measurements (indices) that depend on them. Rigorously speaking all of the factors conditioning the reflectance value recovering (sensor features, atmospheric model, topography, etc.) should be taken into consideration and all of the error components should be followed along the adopted formulas. At the moment this level of detail is not reached, but the test demonstrate that interesting conclusions

can be given even basing the reasoning on an hypothetical error value affecting the reflectances. In this work two vegetation indices, NDVI (Normalized Difference Vegetation Index) and PVI (Perpendicular Vegetation Index), respectively representing the slope- and distance-based families of indices are considered. According to their definition error propagation is evaluated both along the index definition and along the differences between indices of different times.

Gli indici di vegetazione

Gli indici di vegetazione costituiscono bande sintetiche ottenute per combinazione algebrica, più o meno complessa, di bande caratteristiche dello spettro elettromagnetico in grado di aggregare le informazioni relative all'attività ed alle condizioni della vegetazione. La maggior parte degli indici sfrutta la differenza di riflettanza che le piante, fotosinteticamente attive, mostrano fra visibile e vicino infrarosso, bande che risultano fortemente correlate alla quantità di biomassa vegetale presente. Il vantaggio di usare gli indici piuttosto che le semplici bande spettrali risiede nel fatto che essi minimizzano gli effetti del suolo e del *background* in genere, forniscono un certo grado di standardizzazione per il confronto ed enfatizzano il segnale legato alla vegetazione (Curran, 1981; Goward, 1989; Malingreau, 1989). Le formulazioni che li definiscono sono numerose e varrebbe forse la pena considerarle tutte. Tuttavia, a titolo esemplificativo ed in forza del loro successo operativo, in questo lavoro ne vengono presi in considerazione solamente due: l'NDVI (*Normalized Difference Vegetation Index*) e il PVI (*Perpendicular Vegetation Index*), come rappresentanti rispettivamente delle famiglie di indici *slope-based* e *distance-based* (Jackson e Huete, 1991) alle quali appartengono. Si ricorda che i primi sono basati sulla pendenza della relazione, supposta lineare, che lega le bande del rosso (RED) e dell'infra-rosso vicino (NIR) nell'ipotesi che tale parametro sia rappresentativo della forza vegetativa delle superfici osservate. I secondi invece considerano la distanza euclidea che separa la posizione dell'oggetto (*pixel*) osservato dalla retta che, nello spazio delle bande RED e NIR, rappresenta la risposta del suolo nudo. Più un *pixel* è lontano dalla linea dei suoli tanto più alta è la probabilità che esso sia attivo dal punto di vista vegetativo. Purtroppo, poiché non esistono parametri universalmente validi per definire la retta dei suoli, è necessario, per ogni scena, calibrare una regressione opportuna che tenga in conto le mutate condizioni di lavoro (ad es. tipo di sensore, condizioni di illuminazione, tipologia di territorio).
Le formulazioni a cui si fa riferimento in questo lavoro sono quella di *Rouse et al*.(1974) per l'NDVI [1] e quella di *Richardson e Wiegand* (1977) per il PVI [2].

$$NDVI = \frac{NIR - RED}{NIR + RED} \qquad [1]$$

$$PVI = [\sin(\alpha) \cdot NIR] - [\cos(\alpha) \cdot RED] \qquad [2]$$

dove, NIR = riflettanza nella banda del vicino infrarosso e RED = riflettanza nella banda del rosso; α è l'angolo formato dalla retta del suolo con l'asse del RED, assunto come noto.
Entrambe le formulazioni evidenziano come gli indici dipendano strettamente dalle riflettanze, le quali, come noto, soffrono di un'incertezza di misura a cui contribuiscono le operazioni che caratterizzano le fasi del processo che ne permettono la determinazione. La riflettanza, di per sé, non è una grandezza direttamente misurabile, ma è il risultato del calcolo del rapporto tra il valore di radianza al sensore e quello di irradianza solare al suolo. La radianza al sensore è nota a meno delle incertezze di calibrazione connesse: a) al grado di dettaglio con il quale sono definite le curve di calibrazione del sensore; b) alla sensibilità radiometrica del sensore stesso; c) al modello radiativo scelto in fase di pre-processamento dei dati (spesso fortemente approssimato). L'irradianza solare è, invece, nota al limitare esterno dell'atmosfera, ma, una volta portata al suolo, soffre anch'essa di tutte le approssimazioni introdotte dal modello di trasferimento radiativo adottato. Risulta pertanto ragionevole ammettere che i valori di riflettanza utilizzabili in fase di

calcolo degli indici siano affette da un'incertezza non trascurabile, la quale è destinata a propagarsi al loro interno. Il rigore scientifico imporrebbe di investigare dettagliatamente, lungo il processo di calcolo, gli effetti generati dagli errori di misura di tutte le componenti in gioco direttamente misurabili. Si ritiene che un tale approccio, per quanto scientificamente elegante, continuerebbe a soffrire di irrinunciabili approssimazioni e semplificazioni con il solo effetto di scaricare alcune ipotesi ad un livello precedente del processo. La scelta fatta in questa ricerca è stata, dunque, quella di fissare una precisione ipotetica per le riflettanze (ipotizzate già calcolate), e procedendo da quel punto in avanti a verificarne gli effetti nel calcolo degli indici.
Dal punto di vista operativo le domande alle quali questo contributo intende rispondere sono le seguenti: a) qual è una ragionevole stima della sensibilità di apprezzamento dell'indice di vegetazione, cioè quali sue differenze sono significative se considerate tra pixel appartenenti alla stessa scena? b) qual è una ragionevole stima della sensibilità con la quale si possono apprezzare differenze nel tempo degli indici, cioè quali *pixel* sono davvero mutati tra due periodi di osservazione?

Metodologia
Uno strumento statistico che può rispondere alle esigenze appena esposte è la legge di Propagazione della Varianza (LPV, Bevington et al.,2002), la cui adozione è auspicabile laddove si conducano misure indirette, cioè determinabili solo attraverso relazioni funzionali tra misure dirette.
Nel caso specifico essa verrà applicata alle formulazioni dei due indici di interesse, NDVI e PVI, dando per risolto, come già detto, il problema della determinazione della precisione delle riflettanze che concorrono a definirli. Questa viene assunta come nota e considerata uguale per tutte le bande ($\sigma_\rho = \sigma_{RED} = \sigma_{NIR}$).
L'applicazione della LPV, la cui formulazione è data in [3], è stata condotta ipotizzando che le due grandezze (RED e NIR) che partecipano alla definizione degli indici siano variabili statistiche indipendenti decorrelate.

$$\sigma_Y{}^2 = \left(\frac{\partial y}{\partial x_1}\right)^2_{\overline{x_1}} \cdot \sigma_{x_1}{}^2 + \left(\frac{\partial y}{\partial x_2}\right)^2_{\overline{x_2}} \cdot \sigma_{x_2}{}^2 + ... + \left(\frac{\partial y}{\partial x_n}\right)^2_{\overline{x_n}} \cdot \sigma_{x_n}{}^2 \qquad [3]$$

dove $Y = f(x_1, x_2, ..., x_n)$ è la misura indiretta che dipende dalle misure dirette x_i e dal loro scarto quadratico medio σ_i assunto come misura di precisione. Il valore delle derivate parziali deve essere calcolato in corrispondenza del valore medio (del valore) assunto dalle variabili indipendenti in quel punto. La prima conseguenze diretta di tale formulazione è che, in via generale, valori diversi della stessa grandezza sono caratterizzati da precisione (incertezza) diversa.
L'applicazione della LPV, come evidente da [3], impone il calcolo delle derivate parziali degli indici rispetto alle riflettanze che li definiscono.
Nel caso dell'NDVI le derivate parziali risultano:

$$\frac{\partial NDVI}{\partial \rho_{NIR}} = \frac{1 - NDVI}{(\rho_{NIR} + \rho_{RED})} \quad ; \quad \frac{\partial NDVI}{\partial \rho_{RED}} = \frac{1 + NDVI}{(\rho_{NIR} + \rho_{RED})} \qquad [4]$$

La formulazione completa della LPV per il calcolo dell'NDVI è esplicitata in [5].

$$\sigma^2{}_{NDVI} = \left[\frac{1 - NDVI}{(\rho_{NIR} + \rho_{RED})}\right]^2 \cdot \sigma_\rho{}^2 + \left[-\frac{1 + NDVI}{(\rho_{NIR} + \rho_{RED})}\right]^2 \cdot \sigma_\rho{}^2 = \frac{2\sigma_\rho{}^2}{(\rho_{NIR} + \rho_{RED})^2}\left(1 + NDVI^2\right) = f(X,Y) \quad [5]$$

La [5] mostra come la varianza dell'NDVI vari in accordo ai suoi valori locali (quindi pixel per pixel) e a quelli di riflettanza delle due bande RED e NIR. Questo significa che ogni pixel della scena sarà soggetto ad un diverso valore di $\sigma^2{}_{NDVI}$ da calcolare localmente sulla scena processata.

La stessa procedura può essere applicata alla formulazione del PVI. In tal caso le derivate parziali risultano essere le seguenti:

$$\frac{+PVI}{+.\ NIR}) \sin , ! \quad ; \quad \frac{+PVI}{+.\ RED}) \%\cos , ! \qquad [6]$$

La formulazione completa della LPV è esplicitata in [7] nell'ipotesi semplificativa che il valore di , non porti contributo d'errore.

$$/^{2}_{PVI}) \sin^{2} , !2/^{2}_{.} \# \cos^{2} , !2/^{2}_{.}) /^{2}_{.}) \text{ costante} \qquad [7]$$

Sorprendentemente, a differenza di quanto avvenuto per la $/^{2}_{NDVI}$, si osserva che $/^{2}_{PVI}$ è indipendente dal valore dell'indice e da quello delle bande che partecipano alla sua definizione. Questo significa che è costante sull'intera scena, cioè per qualunque *pixel*. Una seconda osservazione riguarda l'entità di $/^{2}_{PVI}$ che si mantiene uguale a quella delle bande da cui deriva: non si assiste ad alcuna propagazione dell'errore.
Qualora si proceda al confronto multitemporale di indici successivi è lecito domandarsi, in riferimento alle precisioni che caratterizzano gli indici nei due periodi, quali differenze siano ascrivibili a reali variazioni e quali invece rientrino nelle fluttuazioni proprie dell'incertezza della differenza. Ancora una volta, in modo molto semplice, può essere applicata la LPV alla differenza d'indice. Nel caso dell'NDVI essa risulta:

$$/^{2}_{0NDVI}) /^{2}_{NDVI(t)} \# /^{2}_{NDVI(t\#0t)}) f(X,Y) \qquad [8]$$

Nuovamente, vista la stretta dipendenza della [8] dal valore di $/^{2}_{NDVI}$, è possibile osservare che ogni *pixel* della coppia di scene multitemporali in gioco, manifesta un diverso grado di apprezzabilità della differenza d'indice che lo caratterizza.
A tal fine è stato messo a punto uno strumento software in grado di generare una immagine di $/_{0NDVI}(X,Y)$, avente le stesse dimensioni e risoluzione geometrica delle scene considerate, e di confrontarla con l'immagine delle differenze

$$0_{NDVI}(X,Y)) \ NDVI(X,Y,t\#0t) \%NDVI(X,Y,t\#0t) \qquad [9]$$

per filtrare i pixel della scena, evidenziando solo quelli per cui la differenza registrata nel periodo considerato è significativa. La procedura, una volta calcolata l'immagine $/_{0NDVI}(X,Y)$, verifica, *pixel* per *pixel*, il soddisfacimento o meno della seguente condizione:

$$0_{NDVI}(X_i,Y_i) 1\ 2/_{0NDVI}(X_i,Y_i) \qquad [10]$$

Il fattore 2 che moltiplica $/_{0NDVI}$ ha l'effetto di ampliare l'intervallo di confidenza (quindi di aumentare il grado di certezza che la differenza calcolata sia veramente significativa) portandolo, nell'ipotesi di normalità della distribuzione delle differenze da cui il valore registrato si suppone estratto, prossimo al 95%.
Nel caso del PVI, essendo la $/^{2}_{PVI}$ costante per l'intera scena, la LPV applicata alla differenza d'indice risulta:

$$/^{2}_{0PVI}) \ 2/^{2}_{.}) \text{ costante} \qquad [11]$$

Pertanto, in questo caso, non si assiste ad alcuna variabilità spaziale dell'incertezza. Non occorre alcuna specifica *routine* per trasferire il concetto in ambito applicativo.

E' sufficiente ricorrere agli operatori matriciali disponibili all'interno dei software commerciali imponendo la verifica della [12].

$$0_{PVI}(X_i, Y_i) \; 1 \; 2/_{0PVI}) \; 2\sqrt{2}/. \qquad [12]$$

Risultati

Per investigare come la precisione di misura dell'NDVI risponda alle diverse condizioni che lo determinano, cioè come sia condizionato dai valori originari di $._{NIR}$ e di $._{RED}$, uno specifico simulatore software è stato sviluppato in linguaggio IDL (*Interactive Data Language*). Questo ha consentito di far variare, con passo pari a 0.01, le riflettanze $._{NIR}$ e $._{RED}$, e di calcolare, secondo la [5], i corrispondenti valori di NDVI e della relativa precisione. Il simulatore ammette la possibilità di controllare il valore di $/_{.}$, che, per la simulazione qui presentata è stata fissata a 0.05 (5 %).

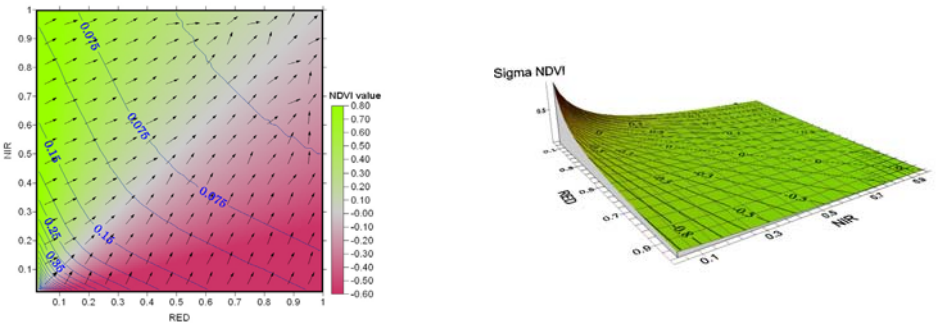

Figura 1 – (sinistra): comportamento di $/_{NDVI}$ (isolinee e campo vettoriale) in funzione del valore di NDVI (scala di colore) e dei valori di riflettanza RED e NIR che concorrono a determinarlo (assi X e Y). (destra): rappresentazione alternativa dell'andamento di $/_{NDVI}$ (mesh) in funzione dei valori di riflettanza di RED e NIR (assi) e del valore di NDVI (isolinee).

Una prima immediata considerazione che la simulazione condotta consente di fare è che $/_{NDVI}$ tende a diminuire all'aumentare delle riflettanze $._{NIR}$ e $._{RED}$. Questo sembra suggerire che, laddove il segnale registrato dal sensore risulta più elevato, la misura di NDVI appare più precisa, probabilmente in forza del miglior rapporto segnale/rumore.

La seconda considerazione possibile è più banale, ma non meno importante a livello operativo. Una contenuta incertezza nella determinazione delle riflettanze determina, soprattutto in carenza di segnale, un non trascurabile decadimento della precisione dell'NDVI. La corrispondente $/_{NDVI}$ può addirittura aumentare di un ordine di grandezza arrivando, già per riflettanze di 0.2, ad essere 10 volte il valore della $/_{.}$ da cui dipende.

La terza evidente considerazione è che lo stesso valore di NDVI, ottenuto da differenti valori di RED e NIR, è caratterizzato da precisioni diverse. Quindi, qualora si voglia qualificare un'immagine NDVI definendone appunto la precisione, dovrà essere predisposta una procedura che agisca, pixel per pixel, applicando la [5] a partire dalle riflettanze di NIR e RED.

Nessuna simulazione è invece richiesta per comprendere le dinamiche di propagazione dell'errore all'interno della formulazione dell'indice PVI, in quanto l'indipendenza dalle riflettanze RED e NIR ne assicura la costanza.

Conclusioni

In questo lavoro si è proceduto ad indagare, da un punto di vista strettamente teorico, gli effetti della propagazione degli errori relativi all'incertezza propria del processo di calcolo della riflettanza lungo le equazioni che definiscono due indici di vegetazione assunti come rappresentativi delle famiglie a cui appartengono: l'NDVI per la famiglia *slope-based* e il PVI per la famiglia *distance-based*. L'approccio adottato è quello suggerito dalla statistica per la stima della precisione di misura di grandezze indirette a partire dalla precisione nota di misure dirette: la Propagazione della Varianza. Dal punto di vista pratico questo significa poter disporre di uno strumento in grado di definire una stima ragionevole del grado di sensibilità con la quale è possibile apprezzare il valore degli indici presi in considerazione sulla singola scena o confrontati nel tempo, mediante differenza, lungo serie temporali. Lo sviluppo delle equazioni richieste per la soluzione del problema ha evidenziato, in primo luogo, che la varianza relativa all'NDVI è dipendente dal valore dell'NDVI stesso e dai valori della riflettanze delle bande che concorrono a definirlo (quindi variabile sulla scena), mentre la varianza relativa al PVI risulta indipendente dal suo valore, costante per l'intera scena e soprattutto coincidente con quella delle riflettanze che lo definiscono (non vi è propagazione).

La simulazione condotta sull'NDVI ha inoltre evidenziato: a) come vi sia una forte accelerazione del processo di degrado della sua precisione al diminuire del segnale riflesso; b) come lo stesso valore di NDVI possa essere caratterizzato da precisioni diverse a seconda dei valori assunti dalle riflettanze delle bande utilizzate.

Allo stato attuale, procedure software apposite sono state sviluppate in linguaggio IDL 6.4.1 per poter applicare quanto esplorato a livello teorico, sulle immagini. Alcuni test sono già stati condotti su immagini SPOT 2 nell'ambito della *change detection*. Risultati di tale esperienza costituiranno oggetto di futuri contributi.

Bibliografia

Bevington and Robinson (2002), "Data Reduction and Error Analysis for the Physical Sciences", 3rd Ed., McGraw-Hill.

R.D. Jackson and A.R. Huete (1991), "Interpreting vegetation indices", Prev. Vet. Med. **11** (1991).

Rouse, J.W., Haas, R.W., Schell, J.A., Deering, D.W., Harlan, J.C., 1974. "Monitoring the vernal advancement and retrogradation (greenwave effect) of natural vegetation", NASA/GSFCT Type III Final Report, Greenbelt, MD, USA.