Realizzazione ed analisi di un database di firme spettrali basato sull'utilizzo di classi in base alla legenda Corine Land Cover 2000.

Manuela D'AMEN (*), Nico BONORA (*), Valter SAMBUCINI (*), Marco MARCHETTI (**), Michele MUNAFO' (*)

(*) SINANet, Servizio Informativo Nazionale Ambientale (APAT) – Via Vitaliano Brancati – 00144, Roma
(**)Laboratorio di Ecologia e Geomatica Forestale, Dipartimento di Scienze e Tecnologie per l'Ambiente e il Territorio - STAT, Università degli Studi del Molise, Isernia.

Introduzione

Il progetto *Image & Corine Land Cover 2000* (I&CLC2000) si propose di aggiornare, a dieci anni di distanza, la base dati relativa alla copertura/uso del suolo relativa al 1990 e di mappare i cambiamenti territoriali sulla base di livelli gerarchici con risoluzione a 5 ha. Per lo sviluppo del progetto sono state utilizzate immagini Landsat 7 ETM+ acquisite per lo più tra Maggio e Settembre degli anni 1999-2001. In ambito nazionale il progetto ha subito un ulteriore approfondimento tematico mediante la creazione di un IV livello gerarchico, che interessa le *superfici agricole utilizzate* ed i *territori boscati ed ambienti semi-nauirali* (classi 2 e 3 della legenda CLC2000). La classificazione *Corine* si basa su analisi fotointerpretative e di campo e classifica il territorio in base ad una presenza minima dell'80% del tema considerato. Per ovviare a questa indeterminazione ed aumentare il dettaglio tematico, si è cercato di avvalersi dell'utilizzo delle firme spettrali di alcune classi di copertura del suolo. Per la strutturazione iniziale dello studio si è scelto di indagare la classe 3.1.1.5 del IV livello della legenda *Corine*, che identifica una copertura territoriale a faggete. Questa classe presenta caratteristiche uniformi per continuità di copertura, distribuzione sul territorio nazionale e range altimetrico di presenza della specie, condizioni che hanno permesso di procedere con un'analisi piuttosto semplificata. Considerata la scala di lavoro, è stata tenuta in considerazione l'influenza dei fattori naturali (stagionalità, condizioni climatiche e posizione geografica).

Abstract

Aims of the *Image & Corine Land Cover 2000* (I&CLC2000) project were the CLC90 database updating and the land-use changes verifying after ten years, by using Landsat 7 ETM+ images acquired during the summer of 1999-2001. An Italian peculiarity of I&CLC2000 project is the 4th level nomenclature, concerning "*Agricultural areas*" and "*Forests and semi-natural areas*" (classes 2 and 3 of CLC2000). This classification is founded on computer-aided photointerpretation and field validation; these areas are classified on the basis of the presence of characteristic land cover structures and overall reliability must be 80 % or more. In order to obviate this lack of preciseness and to improve the thematic accurateness, a feasibility study was undertaken to evaluate the potential employment of the spectral signatures to the land cover identifying, with the final intent of a forest formations spectral signatures database creating about the 3.1.1.5 class (beechwood dominance), taking in account also the natural factors influence (seasonal variation, climatic conditions, geographic location) on spectral variability.

Scopo dello studio

Il presente lavoro mira alla costruzione di un database di firme spettrali relativo a zone boscate sulla base del IV livello gerarchico della copertura CLC2000, al fine di migliorare, nei limiti della

precisione che le scene Landsat ETM+ permettono, l'identificazione del tematismo considerato ovviando all'indeterminazione della classificazione legata alla prevalenza della specie che, nel peggiore dei casi, presenta un errore del 20%. Poiché la risposta spettrale delle specie vegetazionali dipende dai cambiamenti fenologici che avvengono con la stagionalità, il progetto ha anche lo scopo di verificare la variabilità spettrale che caratterizza il territorio italiano in stagioni ed in anni differenti. Lo studio cerca inoltre di fornire uno strumento capace di apportare un contributo nell'identificazione di quelle aree che, durante specifici periodi dell'anno, sono soggette a rischi naturali od antropici, quali incendi, processi di desertificazione e degrado della componente vegetazionale.

Metodologia

La classe di copertura del suolo identificata per la strutturazione iniziale del database è il tematismo 3.1.1.5 (*boschi a prevalenza di faggi*) della legenda CLC2000. Sono state analizzate, tramite l'utilizzo del software Envi 4.2 (RSinc.), 21 scene Landsat ETM+ che includono tale tematismo: queste immagini hanno una data di acquisizione compresa tra i mesi di Maggio e Settembre tra gli anni 1999-2002. Di ogni immagine sono state analizzate le prime 7 bande. Per facilitare le operazioni di calcolo, per ogni immagine è stato creato un *resize* tramite opportuno mascheramento. La maschera, di tipo vettoriale, è stata ricavata dalla copertura CLC 2000. Tramite l'ausilio di precedenti controlli puntuali sul campo, usati per la validazione dell'affidabilità della copertura CLC2000, è stato selezionato ed analizzato un insieme di pixel della copertura a faggete identificante una terna di valori RGB. Questo ha permesso di poter espandere la selezione ad altri insiemi di pixel presentanti valori analoghi. La selezione dei pixel è avvenuta tramite l'utilizzo di *regioni di interesse* (*ROI*), identificanti porzioni di immagine sulla base di un poligono tracciato manualmente, che permettono di procedere ad un'analisi delle sole aree selezionate escludendo la restante parte dell'immagine. E' stato quindi possibile definire, a scala nazionale, la variabilità spettrale del tema considerato. Questo ha permesso di poter procedere ad un'analisi dei valori minimi, medi, massimi e di deviazione standard nella distribuzione e del valore del DN. Come secondo obiettivo, ci si è proposti di analizzare quelle porzioni di territorio coperte da più scene, al fine di poter condurre un'analisi multitemporale. Per poter ovviare ad eventuali errori legati ad una nuova selezione di pixel, si è proceduto tramite il riutilizzo di quelle aree precedentemente selezionate. Questa fase ha comportato diversi passaggi per il raggiungimento dello scopo preposto, in quanto le *ROI* sono associate ad una determinata scena e non sono esportabili da un immagine all'altra. Per poter riutilizzare esattamente le stesse aree selezionate si è agito tramite procedure di vettorializzazione che hanno permesso l'esporazione della selezione. Raggiunto l'obiettivo di poter analizzare aree comuni a più scene, si sono potute confrontare le distribuzioni dei valori di *DN* da un punto di vista statistico, tenendo in considerazione sia l'anno ed il mese di acquisizione e sia la condizione climatica al momento dell'acquisizione.

Analisi statistica

Sono state analizzate 21 scene, per un totale di 29166 *pixel* (tra 1069 e1845 *pixel* per ogni singola scena). Per ogni scena singola e per ogni singola banda la distribuzione dei valori di DN è stata analizzata dal punto di vista statistico. Per ciascuna banda di ogni scena sono stati determinati i valori minimo (*Min scena*) e massimo (*Max scena*) di DN ottenuti, l'ampiezza dell'intervallo (*Range scena*), il valore medio (*Mean*), la deviazione standard (*St.Dev.*) ed il valore di DN corrispondente al picco di frequenza dei *pixel* (*Picco scena*). Prendendo in considerazione complessivamente tutte le scene, per ogni singola banda, per ciascuno dei suddetti parametri sono stati calcolati il minimo, il massimo e la mediana, al fine di determinare i caratteristici intervalli di DN e dei picchi.

Sono stati effettuati due tipi di confronto tra le bande. In primo luogo è stata comparata l'ampiezza degli intervalli di DN ottenuti nelle diverse bande. In seguito dopo aver raggruppato le scene sia in base alla loro distribuzione geografica (Nord, Centro e Sud), sia in base al mese di acquisizione

(fine maggio-giugno; luglio; agosto e settembre) sono stati confrontati tra i gruppi, per ogni singola banda, i valori minimo e massimo di DN, il Range ed il Picco.
Poiché in entrambi i casi, i valori non seguono una distribuzione normale, i confronti sono stati effettuati attraverso test non parametrici, in particolare utilizzando il Kruskal-Wallis ANOVA. Il test di Kruskal-Wallis analizza le differenze tra più di 2 campioni e rappresenta l'alternativa non parametrica del test F per l'analisi della varianza ad un criterio di classificazione.
La metodologia del test di Kruskal-Wallis può essere sintetizzata nei seguenti passaggi: a) si assegna un rango a tutte le osservazioni dei gruppi di dati, considerate come una serie unica e, se sono presenti misure uguali, si assegna un rango medio; b) si calcolano le somme dei ranghi di ogni gruppo R_i e quella totale; c) si calcolano le medie dei ranghi di ogni gruppo; d) si calcola il valore di un indice $\mathbf{H}_{[1]}$

$$H = A \, \Sigma_i \, [\, R_i^2 / n_i \,] - 3(N+1) \quad [1]$$

dove la sommatoria è estesa a tutti i k gruppi a confronto e n_i è il numero di dati del gruppo i-esimo, N è il numero totale di osservazioni, $A = N(N+1)/12$.
Il parametro H si distribuisce approssimativamente come la distribuzione χ^2 con (k-1) gradi di libertà. Quando il valore di H calcolato è superiore a quello riportato nella corrispondente tabella, alla probabilità α prefissata, si rifiuta l'ipotesi nulla e quindi esiste almeno una differenza significativa tra le medie dei ranghi .
Anche nella statistica non parametrica è possibile ricorrere a confronti multipli per individuare quali sono i gruppi sono significativamente diversi tra loro. In questo caso è stato utilizzato il test di Tukey applicato ai ranghi (Soliani, 2005).
In tutte le analisi statistiche effettuate, è stato considerato un limite di confidenza di 0,95.

Descrizione dei risultati
Nella maggior parte delle scene studiate, i valori di DN estratti dalle bande1, 2, 3, 6 e 7 seguono una distribuzione gaussiana, in cui è chiaramente distinguibile un valore di picco di frequenza dei DN. In particolare lo spettro ottenuto dalla banda 3 risulta avere un andamento tipo gaussiano nel 92% delle scene analizzate, mentre nella banda 6 i valori di DN estratti dalle scene seguono tale andamento solo nel 64% dei casi. Al contrario, analizzando lo spettro ottenuto dalle bande 4 e 5 di ogni e per tutte le scene, si è osservato un ampio scattering dei valori dei DN, che non permetteva né di individuare qualche particolare andamento, né tanto meno di identificare chiaramente alcun picco di frequenza. Al fine di verificare la bontà delle distribuzioni dei DN ottenuti, la procedura di selezione delle aree, tramite *ROI*, è stata ripetuta per alcune scene scelte a caso. Il confronto dei valori dei DN estratti dalle nuove aree selezionate con quelli ottenuti durante la prima elaborazione non hanno mai dato uno scarto maggiore di 5, confermando in tal modo la rappresentatività delle aree scelte. Solo in una scena (l'unica acquisita nel 2002) la distribuzione dei DN si scosta notevolmente dalla media di tutte le altre scene, e si è deciso quindi di escluderla dalle analisi statistiche. Si può ipotizzare che la diversa risposta possa essere imputabile alla correzione atmosferica, o da peculiari condizioni meteoclimatiche al momento dell'acquisizione. In tabella 1 sono riportati i risultati ottenuti per le singole bande: da essi emerge un intervallo medio caratteristico dei valori di DN per ciascuna banda. Tuttavia, per nessuna banda è stato possibile individuare un intervallo minimo di valori di DN comune a tutte le scene. Esiste una differenza significativa tra i range dei valori di DN osservati nelle diverse bande ($H_{(6,\ 140)}$=97,97, p=0,0000. Kruskal-Wallis ANOVA test). Il *post-hoc* test, ovviamente, ha evidenziato che differenze significative sono presenti nel confronto tra le bande 4 e 5 con tutte le altre.
Infatti l'intervallo medio dei valori di DN è piuttosto ristretto nelle bande 1, 2, 3, 6 e 7 (da un minimo di 12 nella banda 6 ad un massimo di 20 nella banda 7), mentre nelle bande 4 e 5, si è osservato, in ogni caso, un ampio scattering della distribuzione dei valori del DN, con un range medio intorno a 40. Pertanto, esclusivamente per le bande 1, 2, 3, 6 e 7 è possibile identificare una

caratteristica firma spettrale. Nella tabella 1 ed in figura 1 è riportato anche l'intervallo di valori corrispondenti al picco di frequenza dei DN per tutte le bande, tranne che per la banda 4 e la banda 5, per le quali, come già evidenziato, non è stato possibile individuare alcun picco di frequenza.

		minimo	massimo	mediana	dev.st.
banda 1	*Min*	**48**	66	55	4,5499
	Max	58	**86**	73	7,3505
	range	9	31	14	6,5221
	picco	53	71	62,5	5,3797
banda 2	*Min*	**37**	52	42,5	4,3723
	Max	49	**66**	60,5	4,6935
	range	11	26	16	4,1948
	picco	41	58	50	4,1577
banda 3	*Min*	**26**	41	29,5	3,9203
	Max	39	**57**	45	5,7334
	range	10	27	14,5	4,5736
	picco	30	47	35,5	4,5409
banda 4	*Min*	**92**	122	104	6,6697
	Max	122	**190**	154,5	16,8302
	range	25	72	48,5	12,2224
	picco				
banda 5	*Min*	**56**	83	75,5	5,7831
	Max	104	**129**	115	5,9427
	range	29	56	39	6,7482
	picco				
banda 6	*Min*	**110**	132	123	6,1867
	Max	123	**148**	134	7,4130
	range	6	18	13	2,7657
	picco	118	138	128	5,9456
banda 7	*Min*	**26**	37	31,5	6,1900
	Max	45	**60**	52,5	7,0931
	range	15	29	20	4,1627
	picco	38	47	41	5,7822

Tabella 1 - Intervallo dei valori corrispondenti al picco di frequenza dei DN per tutte le bande.

Nelle bande 1, 2, 3 e 6 il picco di frequenza osservato ricade in un intervallo di ampiezza 14-20 valori di DN. Nel caso della banda 7, invece, il range osservato è molto più ristretto pari ad un intervallo di soli 9 valori di DN. Al fine di evidenziare le eventuali variazioni della firma spettrale in funzione delle condizioni atmosferiche sono state valutate anche alcune variabili meteoclimatiche che potrebbero influenzare le caratteristiche spettrali delle scene acquisite. Gli indicatori climatologici considerati sono: a) la bagnatura fogliare media, b) il bilancio idrico giornaliero medio e c) l'umidità relativa media. I valori mensili negli anni 1999-2002 di tali grandezze sono resi disponibili dal Sistema nazionale per la raccolta, l'elaborazione e la diffusione di dati Climatologici di Interesse Ambientale (SCIA). Tenendo conto di tali grandezze, sono state confrontate scene acquisite nello stesso mese di anni diversi ma con medesima localizzazione geografica (considerando come suddivisione Nord, Centro e Sud Italia): dal confronto dei parametri misurati dei DN non emergono differenze apprezzabili in nessun caso, né tanto meno si osservano differenze degne di nota tra i valori degli indicatori climatologici.

Analisi delle componenti climatologico-geografiche

Il picco di frequenza dei DN tende ad assumere valori più alti nelle scene che rappresentano le regioni dell'Italia meridionale in tutte le bande considerate (1, 2, 3 e 6) tranne che nella banda 7. Tale differenza però non risulta statisticamente significativa in alcun caso. Un simile andamento è riscontrabile considerando il valore Minimo dei DN di tutte le scene: essi assumono valori maggiori dei DN nelle scene dell'Italia meridionale in tutte le bande considerate (1, 2, 3 e 6), tranne che nella banda 7. In questo caso la differenza risulta statisticamente significativa nelle bande 1 ($H_{(2,20)}$ =7,1645, p = 0,0278. Kruskal-Wallis ANOVA test), 2 ($H_{(2,20)}$ = 9,4541, p = 0,0089. Kruskal-Wallis ANOVA test), e 3 ($H_{(2,20)}$ = 8,6716, p = 0,0131. Kruskal-Wallis ANOVA test). Il confronto statistico dei parametri sopra riportati per tutte le altre bande (Min, Max, Range e Picco) non ha

evidenziato differenze significative tra le scene distribuite al Nord, Centro e Sud (i valori del test Kruskal-Wallis ANOVA, sono riportati in tabella 2). Per quanto riguarda lo studio dell'influenza della stagionalità sulla firma spettrale è stato effettuato un analogo confronto tra le scene raggruppate in base al periodo di acquisizione. L'analisi statistica non ha messo in evidenza differenze significative per i parametri considerati (i valori del test Kruskal-Wallis ANOVA, sono riportati in tabella 4).

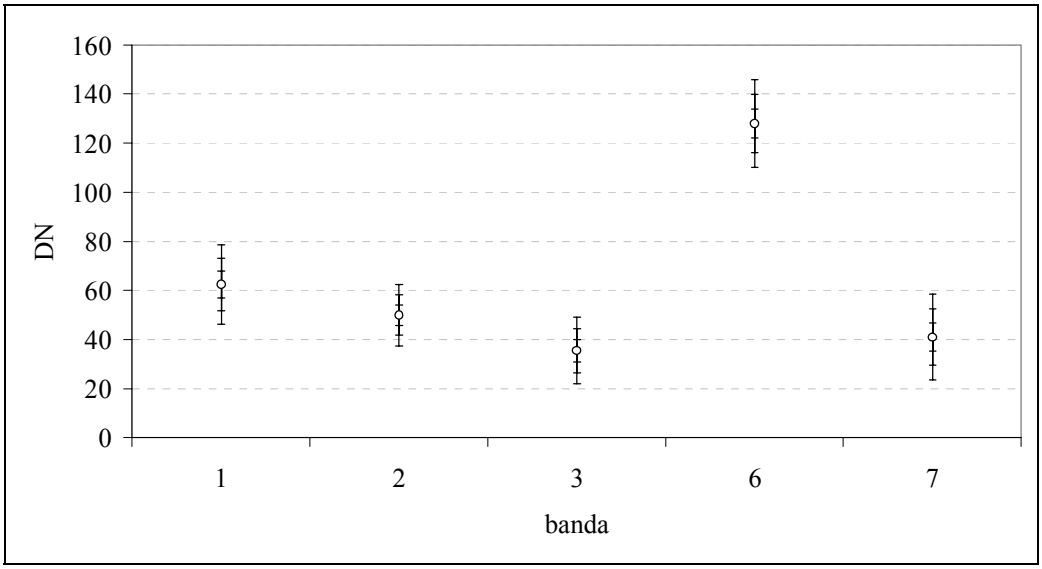

*Figura 1 - Picco di frequenza dei DN (mediana) per le bande in cui esso è riconoscibile. Le barre di errore indicano la deviazione standard*1,*2 e *3.*

	DNmax		DNmin		DNmax-Dnmin		picco	
	$H_{(2,20)}$	**p**	**$H_{(2,20)}$**	**p**	**$H_{(2,20)}$**	**p**	**$H_{(2,20)}$**	**p**
banda 1	3,3124	0,1909	7,1645	0,0278	3,9247	0,1405	2,9487	0,2289
banda 2	3,8454	0,1462	9,4541	0,0089	4,1019	0,1286	5,8627	0,0533
banda 3	4,9621	0,0837	8,6716	0,0131	4,7789	0,0917	3,7398	0,1541
banda 6	5,3771	0,068	3,1085	0,2113	0,5743	0,7504	5,6444	0,0595
banda 7	0,6432	0,725	2,3648	0,4162	2,4288	0,2969	1,3287	0,5146

Tabella 2 - Valori del test di Kruskal-Wallis ANOVA e livello di significatività ottenuti per il confronto tra le scene raggruppate in base alla distribuzione geografica.

Per quanto riguarda lo studio dell'influenza della stagione sulla firma spettrale è stato effettuato un analogo confronto tra le scene raggruppate in base al periodo di acquisizione. L'analisi statistica ha messo in evidenza differenze significative per quanto riguarda il minimo nelle bande 2 ($H_{(2,20)}$ = 10,8287; p = 0,0127. Kruskal-Wallis ANOVA test), 6 ($H_{(2,20)}$ = 9,8682; p = 0,0197. Kruskal-Wallis ANOVA test) e 7 ($H_{(2,20)}$ = 8,8008; p = 0,0321. Kruskal-Wallis ANOVA test). In tutti i casi dal test *post-hoc* è emerso che solo i valori ottenuti a settembre sono significativamente differenti da quelli ottenuti negli altri mesi,: che potrebbe essere imputabile alle differenti condizioni climatologiche di fine estate. Gli altri valori ottenuti dal test Kruskal-Wallis ANOVA, per quanto riguarda le differenze stagionali, sono riportati in tabella 3.[M1]. Per un'analisi più accurata delle firme spettrali, anche per quelle porzioni di territorio coperte da più scene sono stati tenuti in considerazione i valori degli indicatori climatologici ed i periodi di acquisizione. In particolare per tutte le coppie di scene sovrapposte sono stati confrontati i seguenti parametri dei valori di DN ottenuti per le bande 1, 2, 3, 6 e 7: *Min scena*; *Max scena*; *Range scena* e *Picco scena*. Come atteso, confrontando scene sovrapposte ottenute nello stesso mese dello stesso anno si osserva uno scarto minimo (DN compreso tra 0 e 5). Risultati simili si ottengono dal confronto delle scene acquisite nei mesi di luglio e agosto dello stesso anno. Comparando invece scene acquisite in primavera con quelle acquisite nell'estate dello stesso anno (fine di maggio con l'inizio di agosto e con fine luglio) si osserva uno scarto più ampio, anche se comunque abbastanza limitato (compreso tra 1 e 23 per le bande1, 2, 3, 6 e 7).

	DNmax		DNmin		DNmax-Dnmin		picco	
	$H_{(2,20)}$	p	$H_{(2,20)}$	p	$H_{(2,20)}$	p	$H_{(2,20)}$	p
banda 1	2,3319	0,5064	6,6086	0,0855	0,4604	0,9275	4,8577	0,1825
banda 2	3,6412	0,3029	10,8287	0,0127	0,6527	0,8843	10,1738	0,0171
banda 3	2,4713	0,4805	6,0504	0,1092	2,1853	0,5348	5,1625	0,1603
banda 6	4,7466	0,1913	9,8682	0,0197	2,3886	0,4958	5,5906	0,1333
banda 7	6,1194	0,1059	8,8008	0,0321	5,523	0,1373	10,2073	0,0169

Tabella 3 - Valori del test di Kruskal-Wallis ANOVA e livello di significatività ottenuti per il confronto tra le scene raggruppate in base al periodo di acquisizione.

Differenze si riscontrano anche tra scene acquisite a metà e fine estate: si osservano picchi ed intervalli con valore più basso a settembre rispetto alle stesse aree acquisite in agosto dello stesso anno. Infine considerando scene acquisite nello stesso mese, ma in anni differenti, lo scarto ottenuto è compreso tra 0 e 7, risultato molto simile a quello ottenuto per le scene dello stesso anno. In tutti i casi si non ci sono differenze apprezzabili tra i valori degli indicatori climatologici considerati nei periodi ed aree confrontate.

Discussione

La continuazione di questo studio non si propone di estendere l'analisi a tutte le specie di carattere arboreo a scala nazionale, in quanto per alcune di esse, presentando diradamenti, i valori di radianza risulterebbero inficiati dalla presenza del suolo nudo o manti erbosi (si pensi alla distribuzione degli ulivi o l'organizzazione dei vigneti), che altererebbero le caratteristiche spettrali della copertura considerata. Questi fattori non permetterebbero un'analisi analoga, in quanto, anche se si riuscisse a ripulire il dal disturbo del suolo nudo, bisognerebbe considerare la variabilità dello stesso su un territorio che presenta differenti tipologie pedologiche. Il proseguimento di questo lavoro verterà quindi all'applicazione della metodologia utilizzata in aree boscate che si prestano ad uno studio del *canopy,* cioè della parte apicale presentante una densità tale da escludere la componente del sottobosco o suolo nudo. Inoltre, le immagini pervenuteci sono state trattate per uno studio mirato all'identificazione fotointerpretativa di classi del suolo con unità minima cartografabile di 5ha (CLC2000), che ha contemplato, oltre ad una correzione atmosferica, georeferenziazione e ortorettifica, il ricampionamento del pixel a 25 m tramite convoluzione cubica che, di fatto, ha alterato ed appiattito i valori di radianza. Considerando però che le aree campione sono ubicate in aree ad elevata omogeneità, si assume che il ricampinamento dell'imagine abbia inciso in maniera lieve sul reale valore reale di radianza del gruppo di pixel selezionati. Un fattore non trascurabile che potrebbe compromettere studi futuri relativi all'identificazione dei cambiamenti delle classi considerate è l'inoperatività del sensore utilizzato nel presente studio, che utilizza 6 bande con intervallo 0.45-2.35µm più un infrarosso 10.4-12.5µm, una copertura per scena di 185 x 185 km e costi accessibili. L'alternativa potrebbe trovare riscontri nell'utilizzo del precedente Landsat 5 TM, da poco nuovamente operativo, coprente la stessa area ma presentante una minore risoluzione geometrica nel termico, oppure tramite l'impiego del sensore SPOT, che presenta una minore estensione di copertura, un numero di bande limitato a 4 con intervallo più ristretto (0.50-1.75µm), un angolo di acquisizione maggiore che porta ad una maggiore distorsione ai lati della scena ma una migliore risoluzione geometrica (20 m).

Bibliografia

Soliani, L. (2005), "Manuale Di Statistica Per La Ricerca E La Professione. Statistica Univariata E Bivariata Parametrica E Non-Parametrica Per Le Discipline Ambientali E Biologiche".