

USO DI DATI ANCILLARI A LIVELLO REGIONALE PER LA CLASSIFICAZIONE DI IMMAGINI SATELLITARI DI ARCHIVIO

Andrea GALLI (*), Ernesto MARCHEGGIANI (*),
Giovanni CIABOCCO (*), Mauro TIBERI (**)

(*) Dipartimento SAIFET – Università Politecnica delle Marche, 60131, Ancona, Tel +39 071 2204965,
a.galli@univpm.it, e.marcheggiani@univpm.it , g.ciabocco@univpm.it

(**) Centro Operativo Servizio Suoli – ASSAM (Agenzia per i Servizi Settore Agroalimentare delle Marche),
Via Alpi 21, Ancona, Tel +39 071 8081, info-suoli@regione.marche.it

Abstract

In the lack of a specific set of ancillary data dedicated to the supervised classification of a multi-spectral high-resolution IKONOS imagery, aiming to the realization of a land cover map, at the 1:10'000 scale, authors has tested to retrieve necessary information through the reuse of several prexistant information from existing regional information systems and monitoring networks.

Abstract

In assenza di un opportuno set di dati ancillari per la classificazione supervisionata di immagini multispettrali ad alta risoluzione IKONOS, al fine della realizzazione di una carta di uso/copertura delle terre, alla scala 1:10'000, si è sperimentato il recupero delle necessarie informazioni mediante il riuso di una serie di dati e informazioni di archivio derivanti dalle banche dati dei preesistenti sistemi informativi e dalle reti di monitoraggio regionali.

1. Introduzione e obiettivi

I servizi tecnici delle pubbliche amministrazioni nell'espletamento delle ordinarie funzioni a cui sono deputati specificamente, o per rispondere alle nuove istanze che sempre più frequentemente vengono poste loro in merito alla razionalizzazione delle risorse, hanno crescenti esigenze di operare valutazioni e analisi a scala territoriale, integrando il costante flusso informativo derivante da verifiche, misure o monitoraggi operate in sito o mediante dati di archivio. Questa evenienza può nascere da una reale esigenza di operare su dati pregressi per analisi ex-post, o dalla volontà di recuperare e valorizzare dati acquisiti per diversi scopi al fine di produrre nuova informazione. In tale contesto, il tema della conoscenza aggiornata dell'uso/copertura del suolo, rappresenta un banco di prova adatto alla sperimentazione di tecniche per utilizzare e valorizzare dati di archivio oramai riccamente presenti nelle pubbliche amministrazioni italiane.

L'aggiornamento e la corretta gestione delle informazioni alla scala territoriale rappresentano esigenze particolarmente importanti in primo per i settori della pubblica amministrazione dedicati all'agricoltura e ai servizi ambientali in genere. Per essi la disponibilità di informazioni aggiornate e affidabili è alla base della programmazione delle loro attività. Le più recenti politiche europee in materia ambiente e salute dei cittadini, espresse dalle direttive UE, e in particolare, le più recenti normative in materia di Politica Agricola Comunitaria (PAC), sono rapportabili alla definizione di procedure atte a valutare l'impatto ambientale e le ricadute socio-economiche derivate dalle misure del Piano di Sviluppo Rurale (PSR). Il rispetto di tale volontà politica, espressa a livello comunitario, crea l'esigenza di una rete di che metta a disposizione tutte le informazioni disponibili al fine di produrre le necessarie valutazioni e analisi.

In tale contesto, bisogna in primo luogo rilevare come la situazione italiana, in generale, e quella della regione Marche, a cui gli autori si riferiscono, sia contraddistinta da un profondo ritardo sia tecnologico, sia concettuale. Uno dei nodi particolarmente critici è rappresentato proprio dall'assenza di efficaci reti di monitoraggio interoperanti. In particolare, la programmazione agroforestale regionale Marchigiana, in ambito PAC e in merito al PSR, evidenzia una sempre crescente necessità di *territorializzazione* degli interventi. La disponibilità di informazioni sullo stato delle coperture diviene, in tale senso, necessaria per una corretta valutazione ex-ante degli interventi.

	Uso Agricolo	% su SAU totale	% su Area totale	ha SAU
1	Cereali (Frumento)	39,79	28,87	4785
2	Barbabetola	6,25	4,53	752
3	Erba Medica	5,69	4,13	685
4	Vite	3,24	2,35	390
5	Orzo	3,15	2,28	379
6	Olivo	1,28	0,93	155
7	Arboree (Frutteti)	0,35	0,25	43
8	Pascolo Polifita	0,30	0,22	37
9	Prato Pascolo Permanente	0,27	0,19	33
Num.	Uso Forestale			
1	Rimboschimenti Di Conifere			
2	Querceti Di Roverella			
3	Orno-Ostrieti			
4	Boschi Di Latifoglie Miste			
5	Leccete			
6	Formazioni Riparie			
7	Cerrete			
8	Arbusteti (Gariga Arborata)			

Tabella 1 - Categorie di uso agricolo e forestale presenti nell'area presa in esame, in particolare, gli usi sono correlati dalla loro ripartizione percentuale in funzione, rispettivamente, della Superficie Utile e Totale.

2. Materiali e metodi

2.1. Definizione del set di verità a terra

L'immagine utilizzata si riferisce all'area agricola prossima al Monte Conero ed al centro abitato della città di Ancona, questa è stata acquisita dal sensore IKONOS-2 in data 07 maggio 2006, con i seguenti parametri di ripresa: Sun Angle Azimuth 153.7476°, Sun Angle Elevation 61.39686°, Scan Azimuth 271.41° e Cross & Along Scan 0.82 m.

La prima fase del lavoro ha portato all'ortorettificazione dell'immagine, in accordo con il sistema di proiezione UTM 33 Nord, e Datum WGS84. La procedura effettuata mediante *PCI – Orthoengine®* utilizzando quale set di controllo a terra (GCPs), ha permesso di definire 15 punti selezionati riferendosi alla Carta Tecnica Regionale (CTR) serie 1:10.000, relativa all'anno 2000, correlandoli ai punti individuati nell'immagine pancromatica. Il risultato ottenuto presenta un errore RMS (*Root Mean Square*) inferiore al pixel, quindi al metro a terra. Tale risultato è basato sulla costruzione di un dettagliato set di verità a terra GTs (Ground Truths) riferita allo stato delle coperture del suolo per relative all'anno 2008; due anni dalla presa dell'immagine. Con particolare attenzione alle aree agricole e forestali. Mentre per la realizzazione dello stato di copertura delle aree urbane si è proceduto per classica fotointerpretazione, sfruttandone la relativa stabilità nel tempo. Le informazioni inerenti aree agricole e forestali sono state derivate dalla integrazione delle informazioni derivate da differenti sistemi informativi presenti all'interno dei servizi tecnici regionali. In particolare, sono state utilizzate: (1) il servizio catastale di visualizzazione on-line (Fogli e Parcelle), (2) gli estratti delle dichiarazioni, effettuate annualmente dagli imprenditori agricoli, inerenti le colture praticate, in regime di pagamento del premio PAC, relative alla campagna agricola 2006, (3) il *Sistema Informativo Agricolo Nazionale*, SIAN, in continuo

aggiornamento, e infine (4) il *Sistema Informativo Forestale Regionale*, SIFR Marche relativo all'anno 2000.

Per la classificazione delle coperture si è preso a riferimento legenda Corine Land Cover - European Environmental Agency (EEA) adattandola al caso di studio. L'approccio diacronico rappresenta un elemento essenziale per caratterizzare la variabilità nel tempo delle coperture all'interno dell'area in esame. In merito all'uso agricolo del territorio, l'area in esame mostra uno stato delle Superfici Agricole Utilizzate, SAU di 12.025 ettari, contraddistinte da circa 75 colture differenti. Le colture sono state raggruppate in 16 classi potenziali, in base all'andamento fenologico, di cui solo 9 classi: Frumento, Orzo, Barbabietola, Erba medica, Vite, Olivo, Frutteti, Prato polifita, Pascolo sono risultate essere presenti all'interno dell'area oggetto di studio, con un'estensione pari a circa il 60% della SAU totale.

Mentre, l'analisi delle coperture forestali effettuata utilizzando i dati della Carta dei Tipi Forestali della Regione Marche alla scala 1:25.000, realizzata nell'ambito del progetto del SIFR Marche, ha portato alla definizione di 23 tipi forestali, raggruppati in 8 categorie forestali, riportate in tabella 1.

Per la realizzazione del set di GTs è stata seguita una procedura basata su interrogazioni (overlay querying) a carico degli strati informatizzati inerenti le coperture catastali registrati (georiferiti) all'immagine IKONOS. Lo strato digitale georeferenziato dei fogli catastali è stato realizzato dal servizio Suoli dell'ASSAM (Agenzia per i Servizi Agroalimentari delle Marche), attraverso la digitalizzazione delle cartografie comunali e successiva georeferenziazione e trasformazione. Successivamente, il record relativo a ciascuna particella catastale è arricchita attribuendole la cultura praticata sfruttando il servizio WebGIS del SIAN.

Per le aree forestali si è proceduto invece all'overlay tra carta dei Tipi forestali e immagine IKONOS, definendo, per ogni categoria forestale, una serie di poligoni suddivisi in aree test (*trainig-site*) e aree di controllo. Come precedentemente descritto, le classi di uso/copertura del suolo afferenti alle classi 1 e 5 del I° livello gerarchico della legenda Corine Land Cover, sono state classificate direttamente sull'immagine IKONOS per fotointerpretazione. Di seguito (Tab. 3) si riporta la legenda del set di verità a terra collezionato per la classificazione dell'immagine in oggetto.



Figura 1 - Confronto tra l'immagine IKONOS 2006 con lo strato dei fogli catastali (destra) e l'ortofotocarta con i particellari nel SIAN (sinistra).

2.2. Analisi di separabilità spettrale tra il set di verità a terra

Al fine della determinazione della migliore combinazione tra bande spettrali e del set ottimale di GTs da utilizzare nelle successive fasi di classificazione supervised, è stata eseguita l'analisi della separabilità mediante algoritmo Jeffries-Matusita (JM), implementato sul software ENVI®. Tale metodologia permette di valutare il grado di separabilità per coppie di classi spettrali basandosi sulla distanza di Bhattacharyya. In particolare l'algoritmo JM fornisce valori oscillanti tra 0 e 2 in cui l'estremo superiore, pari a 2, rappresenta la massima separabilità (Richards J.A., 1999).

L'analisi della separabilità è stata condotta sulla base dell'immagine IKONOS a 11bit utilizzando dapprima il set completo di ROI a disposizione (24 livelli), e ripetuto per tutte le possibili combinazioni tra bande spettrali. Ciò ha permesso di individuare quale migliore combinazione quella in cui tutte 4 le bande spettrali a disposizione sono utilizzate. In figura 2 sono illustrati i risultati della matrice least to most che rappresenta in maniera sinottica tutti i rapporti di separabilità tra classi. Si noti come l'uso di 24 livelli risulti essere ridondante in considerazione del fatto che molte combinazioni mostrano un grado di separabilità estremamente scarso.

Scegliendo quale soglia accettabile un valore di JM pari a 1.8 sono state eliminate le classi con valori inferiori a tale limite (1201, 1202, 1203, 1204, 1205, 1208, 1210, 1301, 1302, 1306, 1307, 1402), eliminando così le firme spettrali che generano confusione all'interno del set di verità dato. La procedura di valutazione della separabilità tra classi spettrali è stata ripetuta utilizzando le restanti 12 classi. In questo caso i risultati hanno fornito valori di separabilità accettabili ($JM > 1.8$) per tutte le classi. Il set di ROI selezionato mediante esclusione delle classi che non hanno raggiunto il limite di scoring di separabilità è stato utilizzato per le successive fasi di classificazione supervised.

Num.	Codice	Classe	Codice Clc	Classe Clc Corrispondente
1	1101	edifici con tetti in coppi, tegole	1.1	zone urbanizzate di tipo residenziale
2	1102	superfici di materiali riflettenti (cemento...)	1.2.1	aree industriali, commerciali e dei servizi pubblici e privati
3	1103	superfici asfaltate, edifici con coperture bituminose e ferrovie	1.2.2	reti stradali, ferroviarie e infrastrutture tecniche
4	1201	Cereali	2.1.1	seminativi in aree non irrigue
5	1202	Vite	2.2.1	vigneti
6	1203	Olivo	2.2.3	oliveti
7	1204	Arboree	2.2.2	frutteti e frutti minori
8	1205	erba medica	2.1.1	seminativi in aree non irrigue
9	1207	Orzo	2.1.1	seminativi in aree non irrigue
10	1208	Foraggi	2.1.1	seminativi in aree non irrigue
11	1209	Barbabetola	2.1.1	seminativi in aree non irrigue
12	1210	prato/pascolo di polifite	2.3.1	prati stabili (foraggiere permanenti)
13	1211	suolo nudo umido	2.1	seminativi
14	1212	suolo nudo secco	2.1	seminativi
15	1301	rimboschimento di conifere	3.1.2	boschi di conifere
16	1302	lecceta xerofila	3.1.1	boschi di latifoglie
17	1303	gariga arborata	3.2.4	aree a vegetazione boschiva ed arbustiva in evoluzione
18	1304	Cerrete	3.1.1	boschi di latifoglie
19	1305	formazioni riparie	3.1.1	boschi di latifoglie
20	1306	orno-ostrieti	3.1.1	boschi di latifoglie
21	1307	querceti di roverella	3.1.1	boschi di latifoglie
22	1401	corsi d'acqua	5.1.1	corsi d'acqua, canali e idrovie
23	1402	Laghi	5.1.2	bacini d'acqua
24	1403	Mari	5.2.3	mari e oceani

Tabella 3 – Legenda del set di verità a terra utilizzato per la classificazione dell'immagine e corrispondenza di questa con la legenda CLC.

2.3. Classificazione

La procedura di classificazione *supervised* (Curran, P. J. and Williamson, H. D., 1985), eseguita mediante classificatori classici: Maximun Likelihood, ML e Spectral Angle Mapper SAM, (Lillesand, T. M. and Kiefer, R. W., 1994, Kruse F.A *et Al.* 1993), è stata ripetuta utilizzando il medesimo set di verità a terra, sia per il set a 12 ottenuta mediante analisi JM, sia quello originale a 24 classi. In entrambi i casi si sono mantenuti invariati i parametri di input di default dall'algoritmo classificatore (ENVI®, Research Systems, Inc.). La classificazione ML utilizzando l'intero set di 24 classi, ha prodotto una accuratezza (*Overall Accuracy*) pari a 86.08% (*Kappa Coefficient* uguale a 0.7247) e nessun pixel non classificato, mentre il classificatore SAM ha prodotto una accuratezza pari a 81.94% (*Kappa Coefficient* uguale a 0.6419) con il 25,85% di pixel non classificati.

Nella classificazione mediante ML, le uniche classi che hanno bassi errori sia del primo sia del secondo ordine (prossimo a 0) sono la 1402 e la 1403 (rispettivamente Water Bodies and Sea), le quali hanno di conseguenza elevata Producer User accuracies (prossimi al 100%). All'opposto altre classi (ad esempio 1203, 1204, 1302 e 1210) presentano valori di ommissione e commissione molto

alti (prossimi al 100%) con Producer and User accuracies molto basse (inferiori al 15%). Altre classi appaiono relativamente vicine al primo gruppo (1102, 1103, 1211, 1212) presentando valori di omissione e commissione relativamente bassi (inferiori al 30%) e Producer and User accuracies relativamente alte (intorno all'80%). Le rimanenti classi presentano valori intermedi e accuratezza inferiori o di poco superiori al 70%. Complessivamente, come era nelle attese tale classificazione non può essere considerata soddisfacente. Similmente, per il classificatore SAM si ritrovano i due gruppi estremi. Da una parte troviamo le classi con bassi errori ed elevate accuratezze (1402 e 1403), dall'altra quelle con elevati errori e basse accuratezze (tra cui 1203,1204, 1302 e 1210), e progressivamente le situazioni intermedie a questi due estremi. Si deve infine notare come il non classificato mostri una numerosità di pixel pari al 22.84%, in cui la classe 1102 mostra una numerosità decisamente più attendibile (1.46%). Alcune considerazioni devono essere fatte in merito agli errori di prima specie (Omission) e di seconda specie (Commission) e alle accuratezze (Producer Accuracy and User Accuracy) relative alle singole classi (Congalton, R.G., 1991, Wickham, J.D., et Al., 2004, Stehman, S.V., et Al., 2003). E' noto che gli errori di omissione (*Omission Errors*) sono dovuti al rifiuto del classificatore dell'ipotesi assunta come vera di appartenenza di un certo pixel a una certa classe (ipotesi definita dal set di GTs) calcolati a partire dai valori sulle colonne e fuori diagonale della matrice di confusione (*Confusion Matrix*). Mentre gli errori di commissione (*Commission Errors*) sono dovuti all'accettazione da parte del classificatore di false ipotesi di appartenenza di un pixel a una certa classe, calcolati a partire dai valori sulle righe e fuori diagonale della matrice di confusione. Di conseguenza, le omissioni comportano una sottostima delle classi d'informazione mentre le commissioni ne comportano una sovrastima (Swain, P. H., 1978).

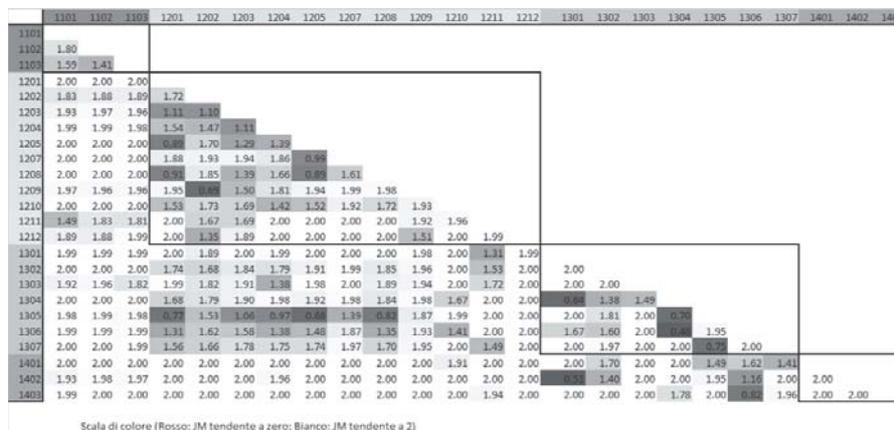


Figura 2 – Analisi di separabilità mediante algoritmo Jeffries-Matusita, classe per classe sul set a 24 classi. La gradazione di colore identifica il progressivo grado di allontanamento da valori accettabili di separabilità.

La classificazione del set a 12 classi mediante ML ha prodotto una Overall Accuracy pari a 94.13% (Kappa Coefficient uguale a 0.8550) e nessun pixel non classificato, mentre mediante SAM si è ottenuta una Overall Accuracy pari a 89.60% (Kappa Coefficient uguale a 0.7422) ed il 25,43% di pixel non classificati. Più nello specifico, nella classificazione ML si osserva un primo gruppo di classi (1304, 1403) con bassi errori di commissione e omissione (prossimi a 0) ed elevate Producer and User accuracies (superiori al 90%) e un secondo gruppo (1103, 1207, 1209, 1303, 1305) che presenta Producer and User accuracies abbastanza alte (intorno a 80%) ed errori di prima e di seconda specie relativamente contenuti, inferiori al 35%. Inoltre, le classi 1211 e 1212, pur avendo errori relativamente contenuti (inferiori al 30%) manifestano Producer and User accuracies più basse dei due gruppi precedenti. Infine la classe 1401 ha un comportamento particolare che la differenzia dalle altre, soprattutto per la bassa Producer Accuracy. Nella classificazione SAM si osservano alcune differenze rispetto ai gruppi precedenti. Un primo gruppo di classi (1403) con bassi valori commissione e omissione (prossimi a 0) ed elevate Producer and User accuracies

(prossime al 100%) e un secondo gruppo (1209, 1303, 1304) che presenta Producer and User accuracies relativamente alte (intorno a 70%) ed errori di prima e di seconda specie relativamente contenuti (inferiori al 35%). Il terzo gruppo di classi (1101, 1103, 1211, 1212, 1401) presenta sia errori sia Producer and User accuracies non omogenei (Commission alto e Omission basso oppure viceversa) con conseguenti accuratezze variabili. Da questi è facile osservare la correlazione inversa che sussiste tra Omission Error e Producer Accuracy da una parte, e tra Commission Error e User Accuracy dall'altra. Infine la classe 1305 ha entrambi i tipi di errore abbastanza elevati e mostra le peggiori accuratezze.

3. Conclusioni

Il presente lavoro dimostra come in assenza di specifici investimenti per l'implementazione di uno specifico servizio (regionale) dedicato al monitoraggio permanente dello stato di utilizzazione del territorio e delle sue dinamiche di trasformazione nel tempo, gli enti in servizio alla pubblica amministrazione possono trarre vantaggio dal riutilizzo delle informazioni già presenti nelle loro banche dati. Attraverso l'integrazione delle informazioni preesistenti e attraverso le classiche tecniche di Telerilevamento, utilizzando anche immagini satellitari a basso costo, è stato possibile produrre utili strati informativi di supporto alle classiche fasi di classificazione supervisionata di un set IKONOS multispettrale ad alta risoluzione. La piena attuazione di un tale sistema, a una adeguata scala di rappresentazione (ad esempio 1:10'000), non è di facile implementazione nella Regione Marche. L'inadeguatezza strutturale e la scarsa attitudine all'innovazione tecnologica dei dirigenti regionali, impedisce a molti servizi di pubblica utilità di sfruttare le potenzialità offerte oggi dalle moderne tecnologie, nel nostro caso, di gestione dell'informazione geografica (GIS) e di trattamento dei dati telerilevati. Tale situazione pregiudica gravemente la maggior parte dei servizi del territorio regionale che gli enti pubblici devono fornire obbligatoriamente sia ai cittadini sia alle altre istituzioni.

Bibliografia

- Congalton, R.G. (1991), "A review of assessing the accuracy of classifications of remotely sensed data", *Remote Sensing of Environment*, 37, 35-46
- Curran, P. J. and Williamson, H. D. (1985), "The accuracy of ground data used in remote-sensing investigations", *International Journal of Remote Sensing*, 6 (10), 1637-1651.
- Kruse F.A., Lefkoff A.B., Boardman J.W., Heiderbrecht K.B., Shapiro A. T., Barloon P.J., et. al. (1993), "The spectral image processing system (SIPS), Interactive visualization and analysis of imaging data", *Remote Sensing of the Environment*, 44: 145-163
- Kruse F.A., Lefkoff A.B., Dietz J.B. (1993), "Expert System-based mineral mapping in northern Death Valley, California/Nevada using the Airborne Visible/Infrared Imaging Spectrometer", *Remote Sensing of the Environment*, 44:309-336
- Lillesand, T. M. and Kiefer, R. W. (1994), *Remote Sensing and Image Interpretation*, John Wiley and Sons, Inc., p. 468
- Richards, J. A. (1995), *Remote Sensing Digital Image Analysis. An Introduction*, 2nd ed., Springer-Verlag.
- SIFR, (2000), "Sistema Informativo Forestale Regionale", IPLA Torino, Regione Marche.
- Stehman, S.V., J.D. Wickham, J.H. Smith and L. Yang (2003), "Thematic Accuracy of the 1992 National Land-Cover Data for the Eastern United States: Statistical Methodology and Regional Results", *Remote Sensing of Environment*, 86: 500-516.
- Swain, P. H. (1978), *Fundamentals of pattern recognition in remote sensing. In Remote Sensing: The Quantitative Approach*, McGraw-Hill, pp. 136-187.
- Wickham, J.D., S.V. Stehman, J.H. Smith and L. Yang (2004), "Thematic Accuracy of the 1992 National Land-Cover Data for the Western United States", *Remote Sensing of Environment*, 91: 452-468.