

Accuratezza nell'interpolazione dei DTM: analisi dei fattori determinati.

Gabriele Garnero, Danilo Godone

Università di Torino, DEIAFA Sez. Topografia Via L. Da Vinci 44, 10095 Grugliasco (TO)
gabriele.garnero@unito.it, danilo.godone@unito.it

Riassunto

I modelli digitali del terreno sono strumenti primari di analisi e gestione del territorio per il fatto di poter essere direttamente impiegabili nei sistemi GIS e in varie applicazioni specialistiche, quali le modellizzazioni idrauliche, le analisi geotecniche, le progettazioni stradali, lo studio delle localizzazioni per gli impianti di telecomunicazione, ecc..

Sono disponibili specifiche tecniche ormai consolidate finalizzate all'acquisizione dati per la generazione di TIN a partire dalle varie tecniche di misura, mentre non sono ancora definite le più idonee modalità di interpolazione per passare dal modello a facce triangolari al grigliato regolare (DTM). In altri termini, non è disponibile una metodologia condivisa, che possa essere suggerita e applicata all'interno delle produzioni cartografiche per la generazione dei modelli numerici.

Nel presente lavoro verranno presi in considerazione vari algoritmi di calcolo, analizzando differenti tecniche di interpolazione (deterministiche e geostatistiche) per l'individuazione di una metodologia ottimale di stima, le interpolazioni verranno applicate ad una nuvola di punti ottenuta da Laser Scanner elitrasportato.

Nel caso specifico si procederà anche all'analisi, tramite tecniche statistiche di classificazione, all'individuazione delle criticità dei dati di input che maggiormente influiscono, nella fase di interpolazione, alla riduzione della qualità del risultato finale.

Abstract

Digital terrain models are key tools in land analysis and management as they are directly employable in GIS systems and other specific applications like hydraulic modelling, geotechnical analyses, road planning, telecommunication...

TIN generation, from different kind of measurement techniques data, is already ruled by specific regulations while interpolation techniques, to compute a regular grid from a TIN, are still lacking. A shared methodology has not already been developed in order to be used in cartographic production while generating digital models.

In the present study different algorithms (stochastic and geostatistic) will be analysed in order to spot an optimal interpolation methodology. Computations will be carried out on a point cloud obtained by helicopter-borne Laser Scanner.

Moreover, a statistic analysis will be carried out, by classification techniques, with the purpose of highlighting critical input data that influences, in the interpolation phase, the quality reduction of the final output.

Introduzione

La produzione di un modello digitale, sia derivato da una ripresa LiDAR sia ottenuto attraverso una produzione basata su collimazioni fotogrammetriche, oggi in massima parte ottenute con tecniche di autocollimazione, ha come fase critica la produzione del grigliato attraverso tecniche di interpolazione che, com'è esperienza di tutti gli operatori, producono risultati anche notevolmente

differenti e quindi, in diversa misura, vanno a compromettere, per l'utente finale, la qualità dell'elaborato dal punto di vista metrico.

Vari documenti sono oggi disponibili, per la comunità scientifica e per il mondo degli operatori, per definire le modalità per la produzione dei dati per i modelli digitali, a partire dalle caratteristiche delle riprese aeree alle quote di volo ed alle precisioni richieste per i sensori: negli ultimi anni, vengono adottate in modo relativamente generalizzato le specifiche approvate in ambito CNIPA e finalizzate alla produzione dei DTM e delle ortoimmagini.

Né queste né altre specifiche prevedono invece indicazioni per quanto concerne la metodologia ottimale da utilizzarsi per il passaggio dai dati di misura ai dati della griglia, definendo le procedure ed i parametri per l'interpolazione mediante i quali produrre il modello digitale, fissando o per lo meno consigliando fasi operative che sono ancora lasciate alla totale autonomia dei singoli operatori.

In sostanza si tratta di rispondere ai seguenti quesiti:

- qual è la metodologia di interpolazione ottimale, al fine di produrre un grid che rispetti al meglio i dati disponibili, e possa quindi essere utilizzato in sostituzione dei dati rilevati garantendo l'utente che il prodotto rappresenta al meglio la qualità che è stata raggiunta con i sofisticati mezzi di rilievo oggi disponibili? Ovvero, in altri termini, qual è il metodo di interpolazione che consente di conservare l'elevata qualità delle informazioni acquisite in fase di misura?
- quali sono i parametri fisici e geometrici che possono influire sulla qualità degli elaborati finali?

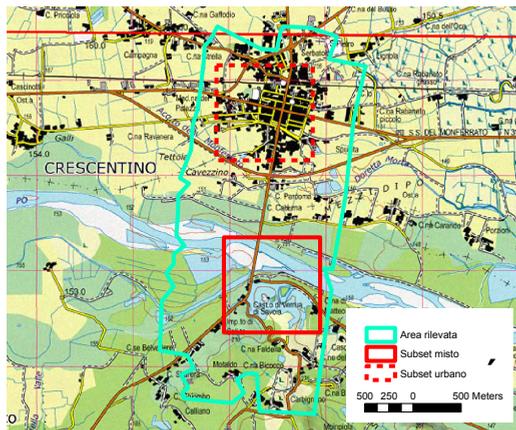


Figura1 – Zona di studio e subsets.

Materiali e metodi

Per la realizzazione dello studio in oggetto è stata utilizzata una ripresa LiDAR nelle disponibilità nel nostro Dipartimento, realizzato all'interno di uno studio sperimentale realizzato qualche anno addietro: la zona di studio è collocata sul tratto di confluenza tra il Po e la Dora Riparia, in prossimità dell'abitato di Crescentino (TO). Il volo, con decollo dall'aeroporto di Casale Monferrato, è stato eseguito in data 21/2/2001 (Figura 1), impiegando uno strumento Saab TopEye. L'orientamento ed il posizionamento del velivolo sono stati determinati tramite l'impiego di una piattaforma inerziale Honeywell H764 a 50 Hz e di antenne GPS montate a bordo, con l'uso di un ricevitore GPS a doppia frequenza - Trimble 4000 Ssi – come stazione di riferimento a terra, posizionato su vertici di coordinate note.

Il sensore è costruito dalla British Aerospace Australia: modello Eagle 2 con una lunghezza d'onda $\lambda = 1063,6 - 1064,5 \text{ nm}$, il laser emesso ha una frequenza di 7 KHz operante in una Classe 3B secondo gli standard Europei EN 60 825.

Gli algoritmi di interpolazione impiegati per la valutazione dell'accuratezza di stima sono i seguenti, suddivisi nei due principali gruppi di classificazione:

1. Algoritmi di interpolazione con metodi stocastici:
 - a. IDW (Tomeczak, 2003);
 - b. Spline (Johnston et al, 2001),
 - c. Natural Neighbors (Watson, 1999);
2. Algoritmi geostatistici:
 - a. Kriging (Matheron, 1970).

Elaborazioni

Le operazioni di interpolazione e di validazione sono state effettuate in ambiente ArcGis, con l'ausilio dell'extensions "Geostatistical Analyst" e "3D Analyst"; le successive fasi di calcolo sono state svolte in ambiente Excel.

Data la mole di dati a disposizione, si è proceduto ad individuare due zone test di superficie pari a 1 Km² dove condurre le sperimentazioni. Una zona, denominata "subset urbano" (1015170 punti), è stata collocata sull'abitato di Crescentino per valutare l'interpolazione di un modello digitale in un territorio altamente antropizzato caratterizzato da una morfologia semplice, ma con la presenza di edifici ed altri ostacoli.

La seconda area ("subset misto"- 1233304 punti) è invece stata posizionata in prossimità della collina di Verrua Savoia e dell'alveo del fiume Po, con l'obiettivo di studiare gli effetti di una morfologia più complessa e varia sulla procedura di generazione del modello.

Il dataset iniziale è stato suddiviso in due insiemi indipendenti con lo scopo di effettuare sia l'interpolazione che la validazione del risultato. Il tool "Create subsets..." ha infatti permesso di creare un set di training (97% dei dati) da impiegare come input nelle fasi di interpolazione e un test set indipendente (3% dei dati) per la validazione. Il dataset originale è stato così suddiviso per non impoverire l'insieme di dati destinato all'interpolazione e non compromettere l'integrità del dato originale (Bater Coops, 2009).

Le singole interpolazioni sono state effettuate impiegando, quando richiesti, i parametri calcolati ai minimi quadrati dal sistema, in modo da non introdurre variazioni dovute all'intervento dell'operatore. I raster sono stati elaborati con una risoluzione pari a 1 metro.

Al termine della procedura di interpolazione, il software ha estratto nella posizione dei punti impiegati come set di validazione il valore interpolato confrontandolo con quello associato al punto, compilando una tabella di scarti. Tale dato, importato su un foglio di lavoro, ha permesso di calcolare le statistiche descrittive dei valori di scarto e dei moduli.

Al fine di determinare l'influenza di parametri morfologici e caratteristici della misura laser sull'errore di interpolazione è stata condotta un'ulteriore serie di analisi: la procedura impiegata è il CART (Classification Analysis and Regression Tree) e consiste nell'analizzare l'andamento di una variabile dipendente in funzione di altre variabili indipendenti (Breiman et al, 1984).

Il CART è un metodo robusto che non richiede assunzioni particolari quali la normalità della distribuzione dei dati o altro; è inoltre in grado di processare variabili categoriche e continue anche in un'unica analisi.

L'output della procedura è un diagramma ad albero che suddivide il dataset iterativamente in subsets di crescente omogeneità (Urban, 2002): l'algoritmo individua nelle singole variabili un valore, detto "di *split*", che permette la suddivisione in due sottogruppi adeguatamente significativi; viene così a crearsi il diagramma precedentemente illustrato, dove ogni ramificazione è associata ad una variabile indipendente e ad un determinato valore soglia.

Al fine di non suddividere eccessivamente il diagramma in ramificazioni di scarsa significatività, è possibile definire una soglia (*tree pruning*) al di sotto della quale l'algoritmo non suddivide ulteriormente una ramificazione (Kittler and Devijver, 1982).

Nel presente studio gli scarti ottenuti dal dataset di validazione, che ha fornito il risultato migliore, sono stati impiegati come variabile dipendente e testati con le seguenti variabili indipendenti (la denominazione delle variabili tra parentesi quadre è quella utilizzata nel modello):

1. Intensità dell'impulso laser di ritorno [INTENSITY];
2. Densità di impulsi per metro quadrato [DENSITY];
3. Pendenza [SLOPE];
4. Curvatura [CURV];
5. Discontinuità morfologica (espressa come deviazione standard della quota calcolata con una finestra mobile 3x3) [ST_DEV].

Il risultato atteso è la suddivisione del dataset in sottogruppi omogenei di crescente qualità nell'accuratezza di stima e la conseguente individuazione dei valori soglia delle variabili indipendenti che determinano tali classificazioni.

Le variabili indipendenti, ad eccezione dell'intensità (già associata ad ogni punto), sono state elaborate dal dataset iniziale generando dei raster dai quali è stato estratto il valore da associare ai singoli punti del subset di validazione. Le elaborazioni sono state compiute in ambiente ArcGis con l'ausilio delle extensions "Spatial Analyst" e "3D Analyst". Le operazioni di analisi delle variabili sono state effettuate su SPSS.

Misto

	IDW	K	SPLINE	NN
MEDIA	0.041	0.011	0.011	0.017
RMS	2.360	2.625	2.625	2.611
MIN	-64.772	-64.771	-64.771	-23.030
MAX	21.931	22.003	21.988	64.763

	IDW	K	SPLINE	NN
MEDIA	1.003	1.024	1.024	1.005
RMS	2.137	2.417	2.417	2.410
MIN	-	-	-	-
MAX	64.772	64.771	64.771	64.763

Urbano

	IDW	K	SPLINE	NN
MEDIA	0.017	-0.015	0.004	0.027
RMS	1.669	1.852	1.701	2.002
MIN	-20.604	-19.990	-23.861	-24.442
MAX	37.480	21.170	35.899	24.818

	IDW	K	SPLINE	NN
MEDIA	0.754	1.004	0.619	0.753
RMS	1.489	1.556	1.585	1.854
MIN	-	-	-	-
MAX	37.480	21.170	35.899	24.818

Tabella 1 – Statistiche descrittive degli scarti ottenuti dai dataset di validazione e dei relativi moduli.

Risultati e discussione

L'analisi degli scarti sulle due aree sperimentali ha consentito di calcolare le statistiche descrittive dei valori ottenuti e dei moduli. tutti gli algoritmi hanno fornito risultati interessanti, ma il migliore è quello ottenuto, in entrambi le aree, dall'impiego dell'IDW (Tabella 1).

Data l'alta densità dei punti, l'elevata qualità nei risultati dei differenti algoritmi era comunque attesa: i valori riscontrati sono molto simili e risulta difficile stilare una classificazione gerarchica degli stimatori in seguito all'individuazione del risultato ottimale. La differenza riscontrata tra i valori di RMS dei due subset è invece imputabile alla complessità e varietà morfologica del subset "misto".

L'analisi con il metodo CART è stata effettuata su entrambi i subsets impiegando come variabile dipendente gli scarti dell'IDW, in entrambi i casi è stato prodotto un diagramma ad albero che raggruppa i valori degli scarti in gruppi, in funzione delle variabili illustrate in precedenza (Figura 2).

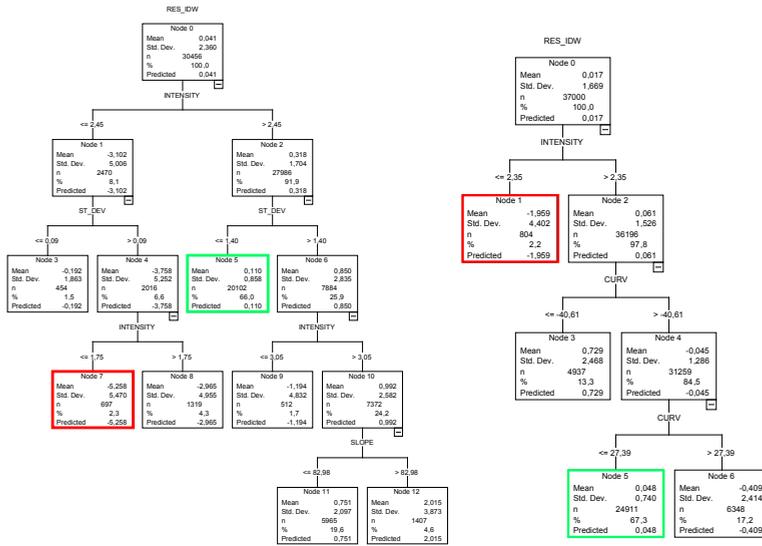


Figura 2 – Diagrammi CART dell’analisi degli scarti ottenuti dall’interpolatore IDW su “Subset Misto” (sinistra) e “Subset Urbano” (destra).

In entrambi i diagrammi è possibile notare le variabili e relative soglie che suddividono il dataset: dopo ogni suddivisione è presente un box che riassume le caratteristiche del sottogruppo e nel caso specifico identifica un incremento o decremento dell’accuratezza di stima. I valori di media e deviazione standard dei sottogruppi consentono di stabilire le porzioni di dataset affetto da errori di stima molto elevati o molto ridotti e l’analisi delle ramificazioni del diagramma ad albero consente di individuare quelle variabili che maggiormente hanno influito, e con quale valore limite, alla determinazione di quel particolare sottogruppo.

Nella Figura 3 è illustrata l’importanza delle variabili indipendenti derivata dal diagramma del CART: in entrambi i casi si nota una marcata influenza dell’intensità dell’impulso; nel subset urbano è inoltre altamente influente la variabile rappresentata dalla curvatura. Le altre variabili sono invece caratterizzate da un ridotto contributo alla determinazione del diagramma.

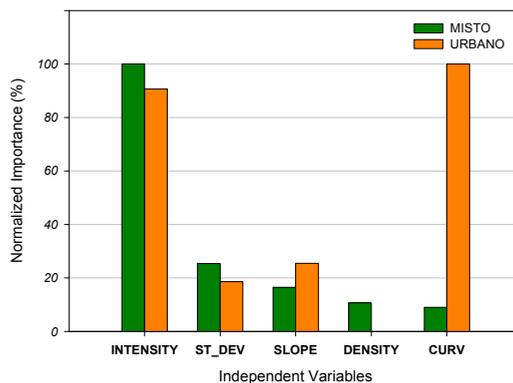


Figura 3 – Importanza delle variabili indipendenti nella classificazione.

Il diagramma consente inoltre di estrarre da ogni gruppo la sintassi SQL per effettuare queries in ambiente GIS e determinare l'estensione e collocazione delle differenti zone: di seguito è riportata a titolo di esempio la sintassi del nodo n° 5 del diagramma del subset "misto".

```
(( ( NOT (INTENSITY = -9999)) AND (INTENSITY > 2,45)) OR ((INTENSITY IS NULL) OR (INTENSITY = -9999)) ) AND ((( NOT (ST_DEV = -9999)) AND (ST_DEV <= 1,40324122)) OR ((ST_DEV IS NULL) OR (ST_DEV = -9999)))
```

Dall'analisi dei due diagrammi si possono quindi evidenziare i valori di soglia che hanno portato alla definizione dei gruppi caratterizzata da scarti più ridotti ed elevati (Tabella 2).

In entrambi i subset gli errori di stima più bassi si sono verificati dove l'intensità (INTENSITY) dell'impulso laser era più elevata e dove la complessità morfologica era più ridotta. Nel subset misto tale condizione è rappresentata da ridotti valori del parametro ST_DEV che significano discontinuità morfologiche di minore entità; nel subset urbano invece si osserva una maggiore importanza nel parametro CURV che rappresenta, tramite la derivata seconda dei valori di quota, l'andamento concavo, convesso o piano del terreno (Zeverbergen et al, 1987; Moore et al, 1991); nel caso specifico le condizioni favorevoli si verificano nelle porzioni caratterizzate da andamento pianeggiante. Al contrario nei sottogruppi caratterizzati da valori di scarto più elevati si riscontra un ridotto valore di intensità (INTENSITY) e, nel caso del subset misto, di una maggiore presenza di discontinuità morfologiche (parametro ST_DEV), come nell'area collinare o nelle sponde del fiume Po.

Misto		Urbano	
Ridotto errore di stima INTENSITY > 2.45 ST_DEV ≤ 1.40	Elevato errore di stima INTENSITY ≤ 2.45 ST_DEV > 0.09 INTENSITY ≤ 1.75	Ridotto errore di stima INTENSITY > 2.35 CURV > -40.61 CURV ≤ 27.39	Elevato errore di stima INTENSITY ≤ 2.35

Tabella 2 – Valori di soglia desunti dai differenti nodi dei diagrammi CART.

L'analisi CART ha evidenziato una serie di parametri e relative soglie significative nel determinare la qualità dell'interpolazione e per fornire elementi utili alla progettazione di nuove campagne. Il valore di curvatura (CURV) ha evidenziato che l'interpolatore ha minori difficoltà di calcolo nel ricostruire forme pianeggianti; risultato confermato dall'indice di complessità morfologica (ST_DEV) che associa ridotto un errore di stima a porzioni territoriali piane o comunque non caratterizzate da morfologie complesse. Parametri quali l'intensità del segnale di ritorno (INTENSITY) invece, possono essere impiegati nella pianificazione di successive campagne di rilevamento; in particolare, nella scelta del sensore da impiegare: al variare dello strumento, e conseguentemente della lunghezza d'onda dell'impulso, varia l'intensità del segnale di ritorno. Obiettivo futuro del presente lavoro è la ripetizione dell'analisi CART su raster a differenti risoluzioni, interpolati dal medesimo dataset, per verificare, ulteriormente, l'attendibilità dei risultati ottenuti.

Conclusioni

L'impiego del CART consente quindi di effettuare un'analisi approfondita sui risultati di una procedura di interpolazione: il metodo va oltre la semplice valutazione degli scarti ottenuti da un dataset di validazione e permette di individuare, attraverso lo studio di variabili coinvolte direttamente o indirettamente nel processo di elaborazione, fattori predisponenti un degrado di qualità dell'elaborato finale.

Bibliografia

Bater C. W., Coops N. C., 2009. Evaluating error associated with Lidar derived DEM interpolation. *Computer & Geosciences* 35, pp. 289 – 300.

- Tomeczak M., 2003. Spatial Interpolation and its Uncertainty using Automated Anisotropic Inverse Distance Weighing (IDW) – Cross-validation/Jackknife Approach in *Mapping radioactivity in environment, Spatial Interpolation Comparison* eds. G. Dubois, J. Malczewski, M. de Court, European Commission – Joint Research Centre, pp. 51-62-
- Urban D. L., 2002. Classification and Regression Trees in McCune B., Grace J. B., *Analysis of Ecological Communities*, MjM Software Design Gleneden Beach, Oregon USA, ISBN 0-9721290-0-6, pp. 222 – 232.
- Johnston K., Ver Hoef J., Krivoruchko K., Neil L., 2001. *Using ArcGIS™ Geostatistical Analyst*, ESRI™, USA
- Watson D., 1999. The natural neighbor series manuals and source codes, *Computers & Geosciences*, Elsevier, pp. 463-466.
- Moore, I.D., R. B. Grayson, and A. R. Landson., 1991. Digital Terrain Modelling: A Review of Hydrological, Geomorphological, and Biological Applications. *Hydrological Processes* 5, pp. 3–30.
- Zeverbergen, L. W., and C. R. Thorne., 1987. Quantitative Analysis of Land Surface Topography. *Earth Surface Processes and Landforms* 12, pp. 47 - 56.
- Breiman, L., J. H. Friedman, R. A. Olshen, and C. J. Stone., 1984. *Classification and Regression Trees*. New York: Chapman & Hall/CRC.
- J. Kittler and P.A. Devijver, 1982. Statistical Properties of Error Estimators in Performance Assessment of Recognition Systems. *IEEE Trans. Pattern Analysis and Machine Intelligence*, 4(2), pp. 215 - 220.
- Matheron, G., 1970. The Theory of Regionalized Variables and its Applications, *Les Cahiers du Centre de Morphologie mathématique*, Fascicule V, Ecole des Mines de Paris:211