

Il controllo di qualità nella produzione di cartografia collaborativa: la nuova carta del consumo di suolo della Regione Toscana

Cinzia Licciardello^(a), Antonio Di Marco^(b), Diego Palazzuoli^(c), Khalil Tayeh^(d)

^(a) Agenzia Regionale per la Protezione Ambientale della Toscana (ARPAT), 055.320.6413, c.licciardello@arpat.toscana.it

^(b) ARPAT, 055.320.6416, a.dimarco@arpat.toscana.it

^(c) ARPAT, 055.320.6380, d.palazzuoli@arpat.toscana.it

^(d) ARPAT, 055.320.6376, k.tayeh@arpat.toscana.it

Riassunto

Nell'ambito dei lavori propedeutici alla pubblicazione del rapporto sul consumo di suolo in Italia relativo all'anno 2016 a cura dell'Istituto Superiore per la Protezione e la Ricerca Ambientale (ISPRA) alcune agenzie regionali, tra le quali l'Agenzia Regionale per la Protezione Ambientale della Toscana (ARPAT), hanno provveduto all'aggiornamento della carta 2015 sulla base dell'indicatore NDVI prodotto da ISPRA su scala nazionale a partire dall'elaborazione di foto satellitari Sentinel-2. A seguito della decisione di ripartire l'attività di fotointerpretazione su più operatori, il territorio regionale è stato suddiviso in 59 quadranti 1:25.000 assegnati in modo casuale tra gli operatori coinvolti. A supporto del processo di controllo di qualità del *dataset*, finalizzato alla valutazione di eventuali difformità dovute alla quota di soggettività nella fotointerpretazione, sono stati definiti alcuni indicatori basati su indagini statistiche e verifiche indipendenti a campione condotte dagli operatori.

Abstract

The 2016 Italian national report on soil consumption has been published by the Italian National Institute for Environmental Protection and Research (ISPRA), starting from preliminary work by some regional environmental agencies on 2015 soil consumption update basing on the national NDVI index derived from Sentinel-2 images. At the Regional Agency for Environmental Protection of Tuscany (ARPAT), in order to detect and correct the product uncertainty due to the subjectivity of the photointerpretation process (59 quads at 1:25.000 scale randomly assigned to four experts in manual digitalization) some indicators have been built basing on statistical analysis and independent verification on random samples by the operators: such indicators have provide valuable aid for quality assessment of the whole soil consumption dataset.

La nuova carta del consumo di suolo

L'Istituto Superiore per la Protezione e la Ricerca Ambientale (ISPRA) gestisce dal 2014 la pubblicazione del rapporto annuale sul consumo di suolo a livello nazionale, avvalendosi della collaborazione delle agenzie regionali per la protezione ambientale nella fase di test della cartografia derivata da

fotointerpretazione manuale. Per l'edizione 2017 è stata messa a punto una metodologia di rilevamento automatico delle zone a maggiore probabilità di presenza di suolo consumato, basata sull'utilizzo delle immagini multispettrali Sentinel-2 e sul calcolo dell'indice di vegetazione NDVI su più immagini. Il Settore Sistema Informativo regionale Ambientale (SIRA) dell'Agenzia Regionale per la Protezione Ambientale della Toscana (ARPAT) ha provveduto alla fotointerpretazione delle variazioni di suolo consumato intercorse tra gli anni 2016 e 2017, distribuendo l'attività in parti uguali tra quattro tecnici esperti: il territorio regionale è stato suddiviso in 59 quadranti di lato pari a 25km (Fig. 1), distribuiti casualmente tra i membri del gruppo in al fine di ripartire nel modo più omogeneo possibile le inevitabili variabilità nei risultati.

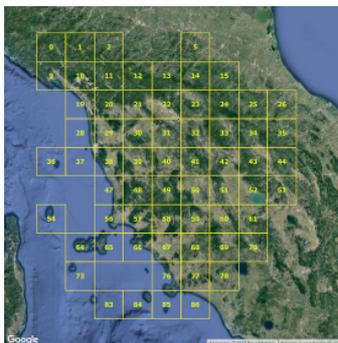


Figura 1 – Ripartizione del territorio regionale



Figura 2 – Dettaglio immagine Sentinel-2

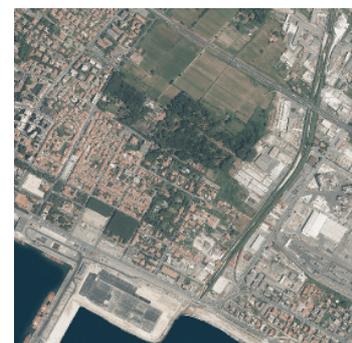


Figura 3 – Dettaglio ortofoto 1:10.000

La complessità del territorio e la bassa risoluzione spettrale e spaziale delle immagini Sentinel-2 (Fig. 2) ai fini di una precisa rilevazione del suolo consumato e gli inevitabili elementi di soggettività nell'adesione alle regole definite capitolato tecnico predisposto dal livello nazionale della rete SINANET (Rete del Sistema Informativo Nazionale Ambientale) hanno fatto emergere la necessità di una verifica a posteriori sulla qualità del dato finale, con il duplice obiettivo della rimozione di eventuali errori macroscopici (falsi positivi e/o negativi nella rilevazione del suolo consumato) e dell'identificazione di *pattern* comportamentali ricorrenti tra i tecnici responsabili del processo di digitalizzazione.

Criticità nella classificazione

Dal punto di vista dell'omogeneità della classificazione tra i tecnici coinvolti, le principali criticità hanno riguardato soprattutto le zone ad edificato diffuso con presenza di ampliamenti recenti, elementi infrastrutturali o urbanistici particolari quali le nuove strade bianche in zone rurali e le serre nelle zone vivaistiche, ed aree ad elevato tasso di variabilità quali aree di cava, discariche e cantieri. Le dimensioni minime delle zone impermeabilizzate, l'inclusione o meno di strade a servizio di zone urbanizzate o zone agricole, l'individuazione della permanenza e della stagionalità delle serre sono stati i principali elementi che hanno contribuito ad aumentare il tasso di variabilità dei risultati della classificazione da parte dei quattro tecnici, sia in termini di falsi positivi e/o negativi (appartenenza o meno alla classe del suolo consumato) che in termini quantitativi (entità della superficie classificata come

suolo consumato). Nelle figure che seguono sono riportati i casi più significativi riscontrati nel corso del controllo di qualità al quale è stato sottoposto il *dataset*, che illustrano alcuni contesti nei quali appare maggiore la probabilità di sovrastimare o sottostimare le variazioni di suolo consumato.



Figura 4 – Casi significativi (area di cava). Da sinistra a destra: superfici digitalizzate dai tecnici (3 su 4, un falso negativo), ortofoto in scala 1:10.000 e ortofoto con scala compresa tra 1:5.000 e 1:7.000.

Si noti in figura 4 il diverso aspetto della zona a scale diverse: in questo caso il falso negativo riscontrato e le disomogeneità della digitalizzazione sono dovuti alla scala utilizzata (1:10.000), che ha comportato la mancata riconoscibilità dell'area estrattiva interna a una zona a suolo nudo. Anche nei casi riportati di seguito si nota l'influenza della scala alla quale avvengono riconoscimento e digitalizzazione delle geometrie: nella figura 5 il fondo del piazzale adiacente all'abitazione è difficilmente distinguibile a piccola scala da quello delle strade adiacenti, e può portare a sovrastimare la porzione di suolo consumato. In figura 6 la disponibilità di foto ad alta risoluzione appare invece indispensabile per una corretta distinzione del suolo impermeabilizzato nel caso di campi sportivi (fondo sintetico o manto erboso) e serre (serre stabili o temporanee).



Figura 5 – Abitazione rurale. Da sinistra a destra: foto e superfici digitalizzate dai tecnici. Si noti il diverso comportamento nell'inclusione del piazzale non asfaltato adiacente all'abitazione.



Figura 6 – (a) Campo sportivo e (b) area agricola con serre. Si notino: (a) la non immediata distinguibilità delle superfici naturali da quelle sintetiche (b) i diversi criteri adottati nell'aggregazione delle serre alle aree consumate.

Nei paragrafi che seguono vengono riportate la metodologia utilizzata nel controllo di qualità del *dataset*, basata su un'analisi statistica in ambiente R e su un collaudo a campione in ambiente GIS, e i risultati dell'indagine.

I dati sono stati analizzati a due distinti livelli di aggregazione:

- a livello di quadrante, con l'obiettivo di ottenere degli indicatori su scala regionale della distribuzione delle variazioni di suolo consumato in relazione alle classi *Corine Land Cover* (CLC)
- a livello di singola geometria, con l'obiettivo di valutare in dettaglio le tipologie di anomalie eventualmente riscontrate. In quest'ultimo caso è stato condotto un test a campione su 100 geometrie.

Indicatori su scala regionale

Come variabili di indagine del processo di digitalizzazione sono state utilizzate la superficie del quadrante, il tecnico fotointerprete, il numero e la superficie delle geometrie che rappresentano il nuovo suolo consumato, superficie e superficie percentuale rispetto al totale di tre macroclassi *Corine Land Cover* costruite a partire dalla classificazione di livello 1 (1: superfici artificiali, 2: aree agricole, 0: aree incolte, zone umide e corpi idrici). Da una prima analisi di base si nota come oltre il 98% delle variazioni si concentra su metà dei quadri del campione: dalla verifica del contributo dei singoli quadri per ciascun fotointerprete (conteggio delle geometrie digitalizzate e della superficie globale classificata come consumata) si nota come la maggior parte della variazione sia concentrata su due/tre quadri al massimo, ad eccezione del tecnico C.

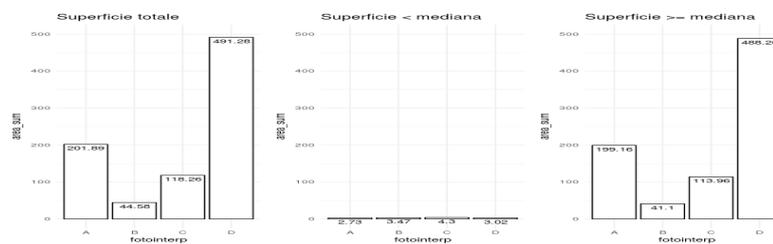


Figura 7 – superfici (ha): totale, inferiore e superiore al valore mediano del campione

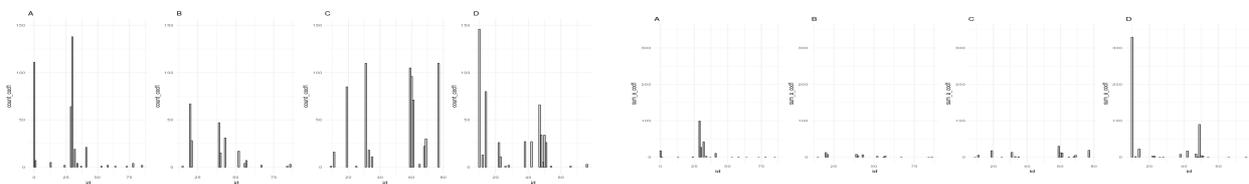


Figura 8 – conteggio geometrie e superfici digitalizzate (tecnici A, B, C, D) sui quadri assegnati

L'analisi dei dati sulle altre variabili evidenzia, oltre all'attesa correlazione della nuova superficie consumata con la superficie dei quadranti fotointerpretati, una correlazione con la maggiore o minore presenza di elementi attribuibili alle macroclassi 1-2 del *dataset Corine Land Cover* (CLC). In particolare si evidenzia come la maggior parte di suolo consumato si trovi in quadri a bassa urbanizzazione (<10% di copertura macroclasse 1) e con copertura delle macroclassi 1-2 tra il 40 e l'80%, ad eccezione di tre quadri (11,0,14) nei quali si riscontrano variazioni rilevanti (superfici >5ha). Tale comportamento è riscontrabile per tutti e 4 i tecnici coinvolti.



Figura 10 – superficie variata con riferimento alle macroclassi CLC 1 e 1+2

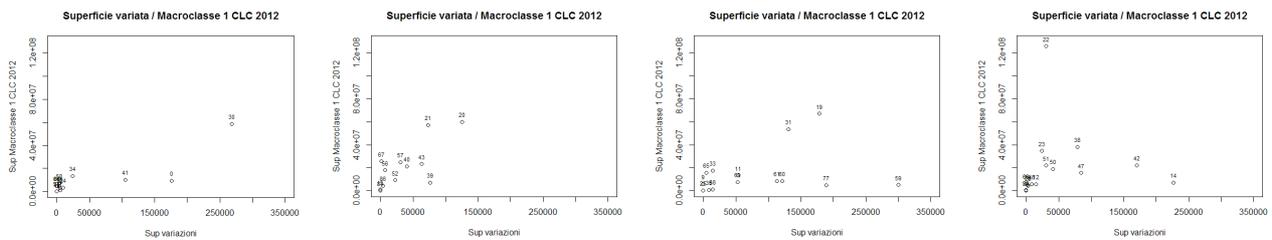


Figura 11 – superficie variata con riferimento alla macroclasse CLC 1 (tecnici A, B, C, D)

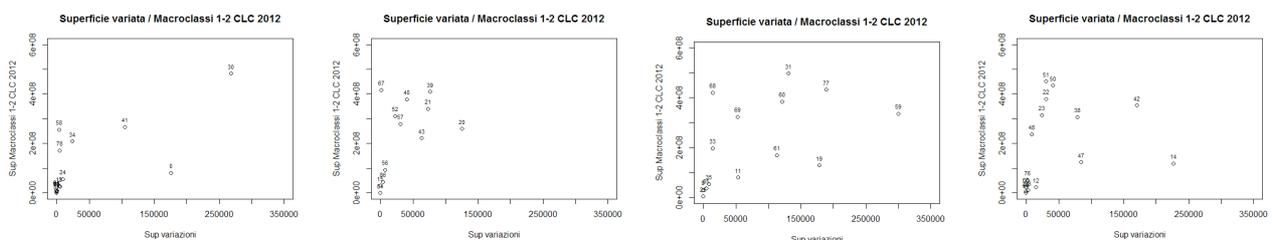


Figura 12 – superficie variata con riferimento alle macroclassi CLC 1-2 (tecnici A, B, C, D)

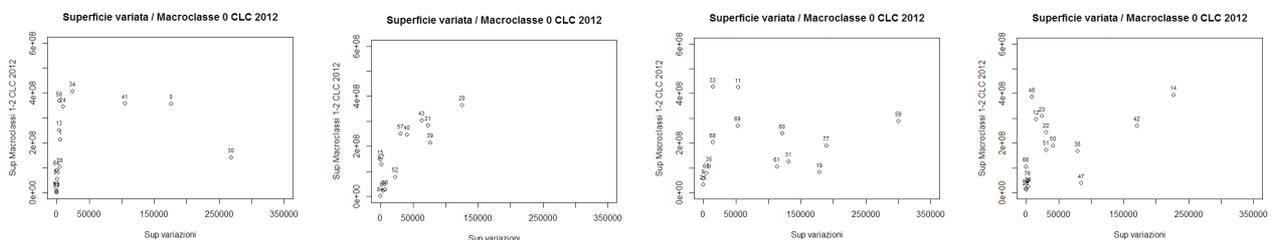


Figura 13 – superficie variata con riferimento alla macroclasse CLC 0 (tecnici A, B, C, D)

Test a campione

Il test è stato condotto sul circa il 6% del totale (100 su 1724 geometrie digitalizzate dai quattro tecnici coinvolti, con superficie di 200 ettari su un totale di 980). Il campione è stato digitalizzato *ex novo* da tutti i quattro tecnici: in tabella si riportano il numero di geometrie classificate come corrette da ciascun tecnico, oltre alla superficie complessiva e percentuale rispetto al totale del campione suddivisa in due classi (superficie totale <2 ha).

| Tecnico | Geometrie confermate | | % geometrie confermate | | Superficie (ettari) | | % superficie originale | | Superficie (media) | | Superficie (varianza) | |
|---------|----------------------|-----------|------------------------|------------|---------------------|--------------|------------------------|-------------|--------------------|--------------|-----------------------|---------------|
| | tot | <2ha | tot | <2ha | tot | <2ha | tot | <2ha | tot | <2ha | tot | <2ha |
| A | 67 | 64 | 67% | 64% | 25.13 | 8.87 | 8.2 | 4.4 | 0.3751 | 0.1386 | 1.2989 | 0.1982 |
| B | 97 | 86 | 97% | 86% | 214.3 | 31.19 | 107.3 | 15.6 | 2.2102 | 0.367 | 120.02 | 0.3144 |
| C | 58 | 57 | 58% | 57% | 16.39 | 14.46 | 8.2 | 6.1 | 0.2536 | 0.2824 | 0.3505 | 0.2750 |
| D | 70 | 63 | 70% | 63% | 87.91 | 12.99 | 44 | 7.2 | 1.2484 | 0.2062 | 4,2491 | 0.2807 |

Dall'analisi della dimensione delle superfici per le singole geometrie emerge come il contributo prevalente nel conteggio del suolo consumato sia imputabile alle aree a superficie superiore ai due ettari. Al netto dei contributi superiori a 2 ettari numero e superficie delle geometrie confermate presentano un andamento assai più omogeneo; tale andamento risulta confermato dal calcolo delle medie e delle varianze delle superfici digitalizzate da ciascun tecnico. Nel caso del tecnico B, il test ha evidenziato una anomalia nella percentuale di variazione della superficie complessiva, che una successiva verifica ha attribuito alla sovrastima di superfici in area di cava. Nella tabella che segue sono riportate le superfici digitalizzate dai tecnici per le tre geometrie digitalizzate dal tecnico B a superficie maggiore: al netto della superficie dei tre campioni, classificati come falsi positivi dai tecnici A e C, le discrepanze tra l'entità della superficie censita si riducono considerevolmente.

| ID Campione | Tecnico | Superficie (ettari) | % superficie (tecnico) | % superficie originale | Tecnico | Superficie campioni | Superficie residua |
|-------------|---------|---------------------|------------------------|------------------------|---------|---------------------|--------------------|
| 1278 | A | - | - | - | A | - | 25.13 |
| | B | 116.4 | 54.3% | 58% | B | 147.4 | 66.9 |
| | C (nc) | 24 | 40.8% | 12% | C | - | 16.39 |
| | D | 30.4 | 34.6% | 15.2% | D | 34.01 | 53.9 |
| 1296 | A | - | - | - | | | |
| | B | 11 | 18.5% | 5.5% | | | |
| | C (nc) | 10 | 18.8% | 5.2% | | | |
| | D | 6.8 | 11.6% | 3.4% | | | |
| 1299 | A | - | - | - | | | |
| | B | 20 | 34% | 10% | | | |
| | C (nc) | 0.4 | 0.6% | 0.2% | | | |
| | D | 16.7 | 28.3% | 8.4% | | | |

Delle 100 geometrie testate solamente 30 sono state confermate da tutti i quattro tecnici; anche in questo caso le differenze evidenziate sull'attività del tecnico B sono riferibili ai falsi positivi da attribuire alla classificazione precedente e alla sovrastima delle superfici impermeabilizzate nelle aree di cava. Nelle figure che seguono sono state confrontate le superfici delle geometrie ottenute dai tecnici, con riferimento al *subset* del campione con superficie inferiore ai 2 ettari e ai campioni con superficie rispettivamente inferiore a 5.000 e 1.500 mq.

L'inclinazione della retta di regressione consente di identificare eventuali trend sulle superfici digitalizzate con riferimento all'attività di ciascun tecnico: più la retta di regressione si discosta dalla bisettrice dei quadranti, più è evidenziabile un trend di sovrastima/sottostima attribuibile all'uno o all'altro tecnico. La minore o maggiore dispersione dei campioni rispetto alla bisettrice dei quadranti evidenzia invece il tasso di variabilità nella delimitazione delle superfici: maggiore è la dispersione attorno alla retta, maggiore la discrepanza nelle modalità di discriminazione delle superfici da perimetrare.

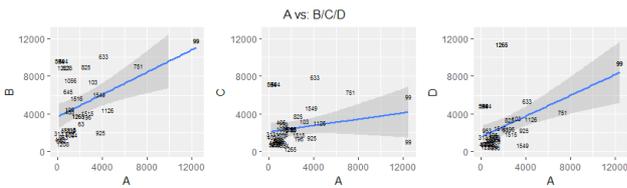


Figura 5 – Tecnico A a raffronto con tecnici B, C e D

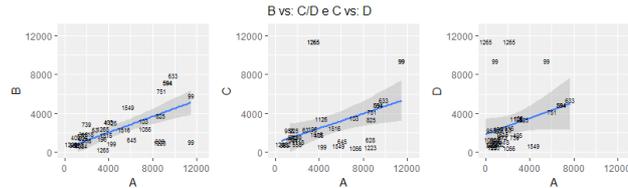


Figura 6 – Tecnico B a raffronto con tecnici C, D e tecnico C a raffronto con D

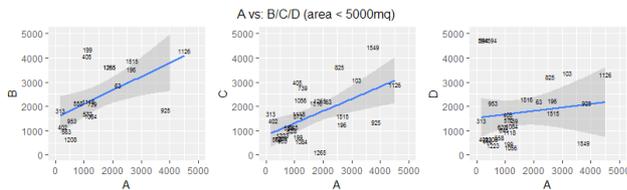


Figura 7 – Tecnico A a raffronto con tecnici B, C e D (< 5.000mq)

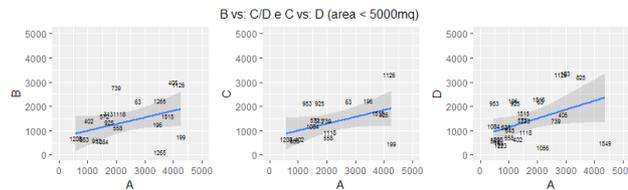


Figura 8 – Tecnico B a raffronto con tecnici C, D e tecnico C a raffronto con D (< 5.000mq)

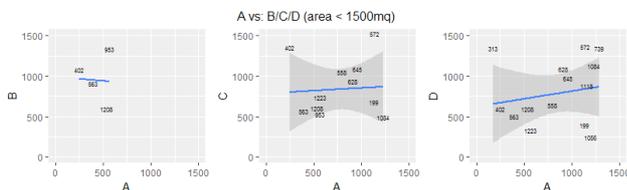


Figura 9 – Tecnico A a raffronto con tecnici B, C e D (< 1.500mq)

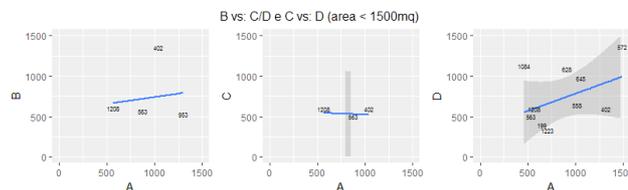


Figura 10 – Tecnico B a raffronto con tecnici C, D e tecnico C a raffronto con D (< 1.500mq)

Conclusioni

Le tecniche di indagine statistica costituiscono, in abbinamento alle metodologie tradizionali basate sull'utilizzo di strumenti GIS, un valido ausilio nel controllo di qualità dei *dataset* spaziali laddove si possano identificare più variabili ausiliarie in grado di evidenziare eventuali comportamenti anomali, e appare sicuramente da raccomandare in tutti i contesti nei quali è necessario un controllo rigoroso delle incertezze. Nel caso della banca dati delle variazioni del consumo di suolo, l'elevata variabilità associata alla digitalizzazione di grandi porzioni di territorio da parte di ciascun tecnico costituisce un elemento di incertezza da gestire con attenzione, dal momento che gli elementi legati alla soggettività del tecnico vengono amplificati a dismisura rispetto alle abituali variazioni di suolo consumato, inferiori ai due ettari (parcheggi, centri commerciali, abitazioni private...). Per la quantificazione di zone ad elevato tasso di variazione nel consumo di suolo è pertanto raccomandabile ricorrere all'integrazione di *dataset* certificati di terze parti, quali ad esempio i censimenti ufficiali delle aree di cava e dei cantieri delle grandi opere, piuttosto che ricorrere a una digitalizzazione *ex novo*, soprattutto laddove la superficie venga censita per la prima volta.

Bibliografia

F. Tomaello, M. Favaretto, A. Fiduccia, "La valutazione degli indici di qualità delle banche dati degli strumenti urbanistici locali nella Regione del Veneto", in *Atti 17a Conferenza Nazionale ASITA (Riva del Garda 5-7 novembre 2013)*, ASITA, Milano, 1231-1238

V. Liguori, C. Tafi, "Metodologia per la valutazione qualitativa del database topografico in base alla conformità con le caratteristiche reali del territorio", in *Atti 18a Conferenza Nazionale ASITA (Firenze 14-16 ottobre 2014)*, ASITA, Milano, 759-766

Sentinel User Handbook and Exploitation Tools (SUHET) (2015), "Sentinel-2 user handbook", *ESA Standard Document*, European Space Agency (ESA)

A. Gatti, A. Bertolini (2015), "Sentinel-2 Products Specification Document", *ESA Standard Document*, Advanced Computer System (ACS)

S. J Baillarin, A. Meygret, C. Dechoz, B. Petrucci, S. Lacherade, T. Tremas, C. Isola, P. Martimort, F. Spoto (2012), "Sentinel-2 level 1 products and image processing performances", in *International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences, Volume XXXIX-B1, XXII ISPRS Congress, 25 August – 01 September 2012*, Melbourne, 197-202

Istituto Superiore per la Protezione e la Ricerca Ambientale (ISPRA) (2017), "Consumo di suolo, dinamiche territoriali e servizi ecosistemici", *ISPRA, Rapporti 266/2017*, Roma, ISBN: 978-88-448-0831-0

Berti C., Di Marco A., Munafò M. (2013), "Tecniche per lo studio del consumo di suolo in Toscana. Metodologie a confronto", in *Atti 17a Conferenza Nazionale ASITA (Riva del Garda 5-7 novembre 2013)*, ASITA, Milano, 175-184

Munafò M. (2013), "Il monitoraggio del consumo di suolo in Italia", *Ideambiente*, 62:20-31.

Malagesi S., Baiocco F., Marinosci I., Munafò M. (2012), "Un'applicazione per il monitoraggio del consumo di suolo", in *Atti 16a Conferenza Nazionale ASITA (Vicenza 6-9 novembre 2012)*, ASITA, Milano, 913-918.

Munafò M., Ferrara A. (2012), Consumo di suolo: proposte di tassonomia e misura, in *XXXIII Conferenza italiana di scienze regionali. Atti* (Roma, 13-15 settembre 2012).

Munafò M., Lupia F., Marinosci I. (2012), "Valutazioni sul consumo di suolo mediante dati di copertura e telerilevati", *Geomedia*, 6:38-40.

R Core Team (2013). "R: A language and environment for statistical computing", *R Foundation for Statistical Computing*, Vienna, Austria.

H. Wickham (2009), "ggplot2: Elegant Graphics for Data Analysis", Springer-Verlag, New York, ISBN: 978-0-387-98140-6

Baptiste Auguie (2015). gridExtra: Miscellaneous Functions for "Grid" Graphics. R package version 2.0.0.