

## Metodi quantitativi per la redazione di flore (parte I): la gestione e la cartografia dei dati floristici

M. D'Antraccoli, G. Bedini, L. Peruzzi

Come ampiamente assodato negli studi floristici, il primo passo verso la redazione di una flora è un'accurata ricognizione e successiva aggregazione di tutte le segnalazioni ricadenti nel territorio oggetto di indagine. Oltre ai dati reperibili in letteratura e alla consultazione di erbari, negli ultimi anni si è aggiunta una sempre maggior mole di dati archiviati in geodatabase online, come ad esempio il progetto Wikiplantbase (Peruzzi, Bedini 2015). Qualunque sia la sorgente del dato, ogni segnalazione individua la presenza di un taxon in una determinata posizione geografica e in un determinato periodo (Soberón, Peterson 2004). Ne consegue che il dato floristico porta in modo intrinseco due tipi di informazione: quella relativa all'incertezza spaziale e quella temporale. In termini pratici, si pensi ad esempio ad una segnalazione riferita genericamente a un toponimo esteso diversi chilometri quadrati rispetto ad una georeferenziata con un moderno dispositivo GPS, oppure a come un dato antico debba essere trattato con maggiore cautela rispetto a uno recente. In seguito a una mirata indagine, non ci risultano nella letteratura floristica approcci cartografici che tengano in considerazione queste fonti di incertezza dei dati, attualmente gestiti e cartografati in modo acritico. Scopo del presente contributo è quindi esplorare un approccio metodologico da introdurre nella ricerca floristica che permetta di gestire in modo oggettivo le incertezze spazio-temporali dei dati, al fine di: (i) elaborare strumenti cartografici capaci di modellizzare le effettive conoscenze floristiche di un'area e (ii) fornire stime probabilistiche circa la presenza dei taxa nell'area d'indagine. Il rilevante valore teorico e pratico dello strumento cartografico nella ricerca floristica è già stato evidenziato da Scoppola, Blasi (2005). Tuttavia, riteniamo necessario lo sviluppo di un nuovo approccio che renda espliciti i gradi di incertezza dei propri dati e risultati, a maggior ragione considerando che, per dirla con Bert Friesen (Allard et al. 2013): "Una volta che viene predisposta una mappa le persone tendono ad accettarla come realtà di fatto". Per ottenere i risultati sopra menzionati si è quindi elaborato un algoritmo che necessita di due elementi: un poligono con la delimitazione dell'area di studio e una tabella delle segnalazioni, completa di accuratezza geografica e anno di ciascun dato. Per la restituzione cartografica, le coordinate delle segnalazioni vengono trasformate dall'algoritmo da punti a buffer circolari con raggio equivalente al valore dell'accuratezza geografica. Contemporaneamente, viene assegnata a queste nuove aree circolari un'importanza decrescente all'incremento dell'area del cerchio. Al valore della segnalazione viene poi applicata una funzione di decadimento della sua importanza in base all'età del dato. A fine processo, ogni buffer avrà un valore finale, espressione del proprio 'peso' cartografico, derivante dalla combinazione dell'incertezza spaziale e di quella temporale. I due casi estremi dello spettro delle possibilità sono: segnalazioni recenti con alta precisione spaziale, che avranno nella mappa un peso rilevante (valore elevato) ma localizzato, e segnalazioni datate con alta incertezza spaziale, che ricopriranno nella mappa aree maggiori ma avranno un minor 'peso' cartografico (valore basso). La mappa finale deriva infine dalla sommatoria aritmetica di tutti i buffer che insistono in un determinato punto della mappa: nella scala relativa, i valori alti saranno quelli con la maggior conoscenza floristica, mentre le aree con valori bassi saranno quelle con maggiori lacune di conoscenze, che evidenziano quindi maggiore ignoranza floristica. La mappa in Fig. 1 è stata ottenuta per il Parco Regionale di Migliarino-San Rossore-Massaciuccoli a partire dalle 165.911 segnalazioni archiviate al 4 Settembre 2018 nella piattaforma Wikiplantbase #Toscana (Peruzzi, Bedini 2015). Contrariamente ai classici metodi cartografici, su essa influiscono anche segnalazioni le cui coordinate ricadono al di fuori dell'area d'indagine ( $n = 9.238$ ), ma il cui buffer d'incertezza ha una porzione più o meno estesa intersecante l'area studio. La seconda parte del metodo misura il grado di sovrapposizione del buffer della segnalazione con l'area studio e, incrociando l'informazione con l'anno del dato, calcola la probabilità che quella segnalazione sia effettivamente presente nell'area studio. Ad esempio: una segnalazione attuale con buffer completamente incluso nell'area studio restituirà una probabilità di presenza del 100% (che diminuirà progressivamente all'aumentare dell'antichità del dato), mentre un buffer che non ha nessuna relazione spaziale con l'area studio verrà escluso (probabilità nulla). Nel caso che più segnalazioni dello stesso taxon, ciascuna con una propria probabilità di presenza, ricadano nell'area studio, l'algoritmo applica il principio di inclusione-esclusione, uno dei principi base del calcolo combinatorio (Sane 2013). Questo permette di calcolare la probabilità complessiva che l'evento 'presenza del taxon nell'area studio' si realizzi a partire dalle singole segnalazioni, che rappresentano eventi multipli non mutualmente esclusivi. L'output finale dell'approccio è quindi un elenco floristico dell'area studio dove ad ogni taxon è associato il grado di confidenza percentuale che esso

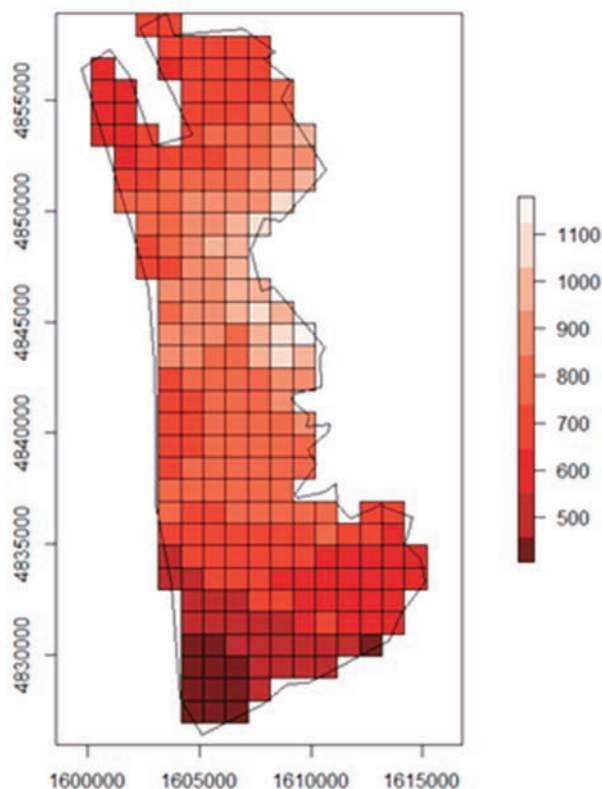


Fig. 1

Mappa delle conoscenze floristiche del Parco Regionale di Migliarino-San Rossore-Massaciuccoli (celle 1×1 Km).

sia effettivamente presente nell'area studio. Oltre a un maggior rigore metodologico e alla possibilità di ridurre la soggettività nella restituzione cartografica delle conoscenze floristiche, il nostro metodo può rappresentare un importante strumento per l'esplorazione floristica del territorio, mettendo a disposizione dello studioso un elenco dei taxa attesi da confrontare in itinere con i propri dati di campo, nonché una mappa per permettere di concentrare gli sforzi di campionamento nelle aree sulle quali insiste una maggiore ignoranza floristica (D'Antraccoli et al. 2018).

#### Letteratura citata

- Allard D, Chilès JP, Delfiner P (2013) *Geostatistics: Modeling Spatial Uncertainty*. John Wiley & Sons, Inc.: 147.
- D'Antraccoli M, Bacaro G, Tordoni E, Bedini G, Peruzzi L (2018) Metodi quantitativi per la redazione di flore (parte II): strategie di indagine floristica basate su approcci probabilistici. *Notiziario della Società Botanica Italiana* 2(1): 9.
- Peruzzi L, Bedini G (eds) (2015) *Wikiplantbase #Toscana v2.1* <http://bot.biologia.unipi.it/wpb/toscana/index.html>
- Sane SS (2013) The inclusion-exclusion principle. In: *Combinatorial Techniques. Texts and Readings in Mathematics*, vol 65. Hindustan Book Agency, Gurgaon.
- Scoppola A, Blasi C (eds) (2005) *Stato delle Conoscenze sulla Flora Vascolare d'Italia*. Palombi editori, Roma. 256 pp.
- Soberón JM, Peterson AT (2004) Biodiversity informatics: managing and applying primary biodiversity data. *Philosophical transactions of the Royal Society of London. Series B, Biological sciences* Royal Society 359: 689-98.

#### AUTORI

Marco D'Antraccoli ([marco.dantraccoli@biologia.unipi.it](mailto:marco.dantraccoli@biologia.unipi.it)), Gianni Bedini ([gianni.bedini@unipi.it](mailto:gianni.bedini@unipi.it)), Lorenzo Peruzzi ([lorenzo.peruzzi@unipi.it](mailto:lorenzo.peruzzi@unipi.it)), Dipartimento di Biologia, Università di Pisa, Via Derna 1, 56126 Pisa

Autore di riferimento: Marco D'Antraccoli