

Fabio Squarcini Berti*, Lorena Bravi**

* Dipartimento di Civiltà e Forme del Sapere, Università di Pisa.

** Dipartimento di Scienze dell'Antichità, Sapienza Università di Roma.

DAI *LEGACY DATA* ALLA DIGITALIZZAZIONE E ANALISI. UN DATABASE PER LO STUDIO DELLA DISTRIBUZIONE DEI RINVENIMENTI DI EPOCA ETRUSCA NELL'AREA DEL MEDIO VALDARNO

Abstract: This contribution aims to analyse the use of legacy data, extracted from bibliographic sources to study and represent, through graphic-spatial tools, the distribution of archaeological finds in the Middle Valdarno. Through exploratory spatial analyses conducted with QGIS and R, and organised by chronological phases, it has been possible to identify a strongly clustered settlement pattern, spanning from the Iron Age to the late Hellenistic period. In particular, the analyses conducted using Kernel

Density Estimation (KDE), Point Pattern Analysis (PPA), and Nearest Neighbour Index (NNI) have highlighted significant variations in the areas of settlement concentration: in the earlier periods, finds are distributed in clusters in areas of higher altitude, while in later periods, a shift towards lower elevations is observed, accompanied by increasing spatial dispersion.

Keywords: Archaeology, Etruscology, Spatial Analysis, QGIS, R.

1. Introduzione

Il presente contributo si propone di offrire una rappresentazione grafico-spaziale dei *legacy data* raccolti durante un lavoro di tesi (Squarcini Berti, 2022), in cui sono stati documentati tutti i ritrovamenti archeologici nell'area del medio Valdarno. Per rappresentare in modo dettagliato le zone di rinvenimento dei reperti, è stato necessario digitalizzare e strutturare i dati ottenuti tramite ricerca bibliografica, esaminando anche la distribuzione dei ritrovamenti. Questo approccio ha richiesto tecniche avanzate di sistematizzazione dei dati, consentendo un'analisi accurata della loro distribuzione con strumenti di analisi spaziale esplorativa, al fine di individuare le aree di maggiore concentrazione per le diverse fasi cronologiche.

L'analisi dei *legacy data* presenta sfide metodologiche, in particolare per la loro interpretazione e affidabilità. Come sottolineato da Huggett (2022), tali dati non

sono oggettivi e immutabili, essendo soggetti a interpretazioni che ne influenzano la natura. Inoltre, la digitalizzazione e le infrastrutture tecnologiche possono alterarne la forma originaria, con interpretazioni nuove e differenziate. Huggett (2022) evidenzia anche l'importanza di un approccio critico nell'impiego di questi dati, bilanciando le opportunità offerte dalle tecnologie digitali con la consapevolezza delle loro limitazioni.

L'integrazione dei *legacy data* tramite tecniche GIS consente di visualizzare l'evoluzione insediativa nel paesaggio antico, ma deve essere fatta con un atteggiamento critico, come sottolineano Casarotto et al. (2019). Per questo motivo è stato fondamentale identificare metodologie efficaci per l'estrazione, la gestione e la valorizzazione dei dati, riducendo al minimo l'errore interpretativo, andando a sviluppare tecniche di digitalizzazione che si adattassero alla variabilità e incertezza di queste informazioni, estraendone il massimo valore possibile.

Inoltre, la localizzazione delle aree di interesse presenta problematiche poiché i dati storici non sono dotati di coordinate precise. È stato quindi necessario sviluppare una metodologia per migliorare l'accuratezza spaziale e garantire una localizzazione più precisa delle aree, permettendo analisi che identificassero zone di maggiore concentrazione. Ciò ha comportato l'adozione di tecniche avanzate di georeferenziazione, integrando informazioni spaziali con quelle storiche e archeologiche. L'obiettivo finale è stato ottenere una rappresentazione accurata della distribuzione dei reperti, e fornire un supporto per future analisi spaziali e interpretazioni sulle dinamiche insediative del medio Valdarno, tenendo conto delle variabili cronologiche.

2. Materiali e Metodi

I dati bibliografici sono stati organizzati in un censimento numerato, strutturato in base alla localizzazione dei ritrovamenti e suddiviso per macroaree geografiche, che corrispondono alle diverse valli della Valdera, del Valdarno, della Val d'Egola e della Val d'Elsa. L'area del medio Valdarno (fig. 1), situata nella Toscana nord-occidentale, si estende tra Pontedera e la confluenza del fiume Era nell'Arno a ovest, Fucecchio a nord, San Miniato e la Val d'Egola a est, e Casciana Terme a sud.

Questa regione, denominata la 'Terra dei Quattro Fiumi', è un'area geograficamente complessa, racchiusa da tre corsi d'acqua principali. Il fiume Arno, che scorre al centro della valle, è circondato da una serie di rilievi: a nord dalle Cerbaie e a sud da una dorsale collinare. Il paesaggio è solcato dai fiumi Egola, Chiecina ed Era, che definiscono la struttura idrografica prima che l'Arno si apra verso la pianura di Pisa.

A nord, ai piedi delle Cerbaie, scorre l'Usciana, un fiume incanalato in rettilinei creati dalle bonifiche granducali del Cinquecento, che confluisce nell'Arno poco a monte di Pontedera. Questo fiume conserva ancora le caratteristiche di un antico corso d'acqua, il cui letto è mutevole e meandriforme. Intorno all'anno Mille, il fiume Usciana era noto con il nome etrusco di *Arme*, che indicava il corso d'acqua che drenava le acque della Nievole e delle due Pescie (di Pescia e di Collodi).

A ovest scorre il fiume Era, che, insieme ai suoi affluenti Roglio e Cascina, si dirige verso l'acropoli di Volterra e verso la Toscana centro-meridionale (Ciampoltrini & Grassini, 2011, pp. 15-16). Infine, a est, l'area oggetto di indagine è delimitata dal

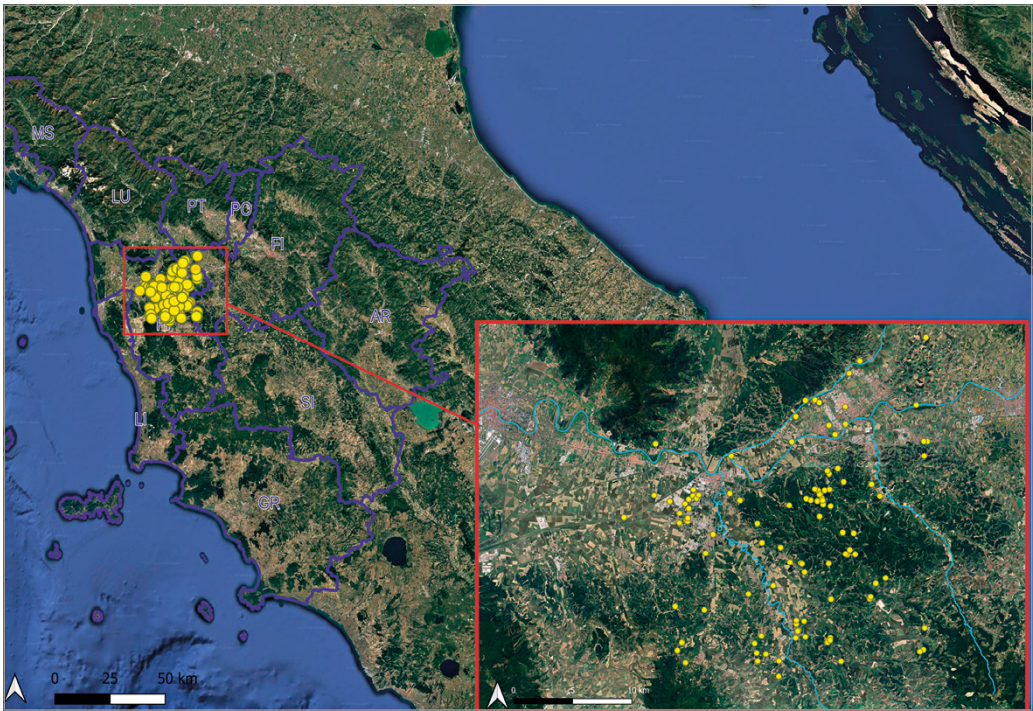


fig. 1. Mappa geografica dell'area del medio Valdarno.

fiume Egola, che confluisce anch'esso nell'Arno, contribuendo ulteriormente alla definizione dei confini geografici della regione.

Le aree di interesse sono state catalogate seguendo una successione geografica da ovest verso est. Ogni sito archeologico è contrassegnato dalla sigla 'S.' seguita da un numero progressivo arabo, e accompagnato da una scheda dettagliata. Questa scheda include diverse sezioni, ognuna delle quali fornisce informazioni approfondite sull'area archeologica in esame: comune, tipologia e anno di indagine, ubicazione e descrizione del sito, materiali, periodo/fase, interpretazione e bibliografia (Squarcini Berti, 2022).

Il progetto nasce dalla consapevolezza della necessità di migliorare la chiarezza e la completezza delle informazioni raccolte, che richiedono un lavoro di organizzazione più approfondito per essere adeguatamente utilizzate ed analizzate. Per rispondere a questa esigenza è sembrato opportuno digitalizzare i dati e organizzarli in un DB, riconosciuto come lo strumento più idoneo per la sua capacità di garantire flessibilità e integrazione di informazioni eterogenee, adattandosi così alle diverse esigenze analitiche del progetto. Questa scelta non solo consente di rendere i dati più coerenti, ma favorisce anche una consultazione rapida e un'analisi approfondita, facilitando la lettura critica e l'individuazione di correlazioni. Permette, inoltre, di apportare aggiornamenti futuri e di migliorarne l'interoperabilità con altri strumenti digitali.

In questo contesto, sono stati esplorati metodi di strutturazione e gestione dei dati che hanno permesso di preservare l'integrità di ciascun dato, integrando anche strumenti GIS per geolocalizzare i ritrovamenti, creare visualizzazioni ed analisi che consentano di valorizzare la loro rilevanza storica e archeologica. L'adozione di

linguaggi di programmazione, come *R*, ha consentito, tramite l'attuazione sempre di analisi spaziali, di trasformare questi dati in risorse attive, capaci di fornire nuovi spunti interpretativi. Questo approccio, basato su dati concreti e oggettivi, senza sovra interpretazioni, consente di costruire conoscenze più approfondite e sfumate, con particolare riferimento, nel caso in oggetto, al contesto dell'Etruria nord-occidentale.

2.1 Dal Catalogo al *Database*

Per raggiungere l'obiettivo preposto, è stato necessario trasferire il catalogo, contenente 94 'Siti', attraverso un processo di digitalizzazione manuale in un database più efficiente, in grado di strutturare le informazioni per renderle rapidamente e intuitivamente comprensibili, consultabili, gestibili, ricercabili, analizzabili e condivisibili.

Questo processo è stato svolto mediante l'utilizzo del *software* Microsoft Excel¹, che ha facilitato una strutturazione dell'informazione che fosse sia coerente con gli obiettivi sia rivolta a un pubblico generico che ne potesse trarre una facile lettura e utilizzo. Il database creato si compone di diciotto campi, che sono stati successivamente collegati a uno SHP poligonale generato con il *software open source* QGIS, utilizzando una chiave primaria (*QGIS overview · QGIS Web Site*, 2025). La scelta di utilizzare uno SHP poligonale è stata dettata dalla necessità di rappresentare in modo più preciso possibile le aree di ritrovamento, soprattutto nei casi in cui la posizione esatta dei reperti non fosse nota. In particolare, per i siti con localizzazione incerta, è stato deciso di delimitare un'area di possibile interesse piuttosto che attribuire un punto arbitrario, evitando così un'interpretazione eccessiva dei dati. Questa scelta metodologica, adottata come prima fase del progetto, ha permesso di ridurre al minimo il rischio di sovra-interpretazione e di mantenere la coerenza con l'incertezza delle informazioni disponibili.

I campi del database sono i seguenti:

- ID: campo numerico che definisce una chiave primaria univoca assegnata a ciascun ritrovamento, che ne consente l'identificazione individuale.
- LOCALITÀ: campo testuale che indica la località del ritrovamento.
- COMUNE: campo testuale che indica il comune del ritrovamento.
- PROVINCIA: campo testuale che indica la provincia del ritrovamento.
- ANNO_INDAGINE: campo numerico che riporta l'anno o gli anni in cui si è svolta l'indagine.
- CRONO_INIZIALE: campo numerico che indica l'anno d'inizio della datazione dei reperti.
- CRONO_FINALE: campo numerico che indica l'anno di fine della datazione dei reperti.
- PERIODO: campo testuale che indica il *range* cronologico in secoli a cui si data il ritrovamento.
- RITROVAMENTI: campo testuale che descrive i reperti trovati nel sito di interesse.
- SITUAZIONE_RITROVAMENTO: campo testuale che indica il contesto in cui sono stati rinvenuti i reperti.

¹ <https://www.microsoft.com/it-it/microsoft-365/excel?market=it>

- **AFFIDABILITÀ:** campo numerico che esprime l'indice di affidabilità dei dati.
- **DESCRIZIONE:** campo testuale che riporta la descrizione del ritrovamento e la sua ubicazione.
- **INTERPRETAZIONE:** campo testuale che fornisce un'interpretazione della tipologia dei ritrovamenti.
- **FASE CRONOLOGICA:** campo testuale che definisce il periodo cronologico di riferimento per i reperti.
- **BIBLIOGRAFIA:** campo testuale che elenca i testi utilizzati per la compilazione della scheda.
- **AUTORE:** campo testuale che contiene il nome e il cognome dell'autore del progetto.
- **COMPILATORE:** campo testuale che contiene il nome e il cognome del compilatore del database.
- **ID_MAGOH:** campo numerico che riporta il numero d'intervento ripreso dal portale MAGOH per ogni ritrovamento.

Inoltre, per rendere i dati uniformi e consistenti, è stato scelto di utilizzare vocabolari standardizzati provenienti dal portale MAGOH², per il campo 'Interpretazione' (II livello) e il vocabolario OGT dal portale GNA³ (Geoportale Nazionale Archeologia) per il campo 'Situazione di Ritrovamento'. Questo ha consentito di rendere i dati uniformi, consistenti, interoperabili, efficienti e facilmente ricercabili, rendendo il database uno strumento più potente e affidabile.

È stato inoltre ritenuto opportuno aggiungere un campo denominato "Affidabilità" per indicare il livello di attendibilità associato al riferimento di ciascun ritrovamento. Come riferimento, sono stati consultati i gradi di affidabilità definiti dal sistema Mappa GIS, basati, principalmente su tre livelli, divisi in sei categorie. Dato che la documentazione proviene esclusivamente da fonti bibliografiche, è stato scelto di utilizzare come riferimento la categoria 'Relazione' secondo Fabiani & Gattiglia (2012, pp. 49-50). Partendo da questa base, sono stati sviluppati tre criteri per definire il grado di affidabilità dei rinvenimenti nell'area oggetto di analisi:

- **Grado 1:** conoscenza limitata del ritrovamento, ottenuta tramite una ricerca circoscritta al singolo sito. La descrizione del luogo di rinvenimento è sufficiente, ma manca materiale archeologico; quindi, l'interpretazione si concentra soprattutto sul contesto generale.
- **Grado 2:** conoscenza moderata del ritrovamento, acquisita attraverso una ricerca su una piccola porzione di territorio. La descrizione del materiale rinvenuto è buona, permettendo un'interpretazione che si basa principalmente sul materiale stesso.
- **Grado 3:** conoscenza approfondita del ritrovamento, ottenuta mediante una ricerca su un ampio territorio. La descrizione del materiale rinvenuto è dettagliata, consentendo un'interpretazione completa basata sia sul contesto che sui reperti archeologici.

² <https://magoh.cfs.unipi.it/en/web/guest>

³ https://gna.cultura.gov.it/wiki/index.php/Pagina_principale

Durante la digitalizzazione del catalogo, è emersa una problematica: molti ‘Siti’ contenevano al loro interno più aree di ritrovamento, molto vicine tra loro, ma identificate con localizzazioni differenti dentro la voce ‘Ubicazione e Descrizione del sito’. Per risolvere l’incongruenza è stato necessario suddividere tali ‘Siti’ in più vettorializzazioni poligonali, ciascuna corrispondente a una specifica area di ritrovamento. Questa scelta ha permesso di condurre analisi più dettagliate e precise. Sebbene le aree di rinvenimento condividano gli stessi attributi, si differenziano per la loro specifica localizzazione spaziale⁴. Al termine del processo di digitalizzazione e strutturazione, il DB è stato salvato in formato CSV, facilitando così il suo inserimento come file di testo delimitato all’interno di QGIS.

2.2 Dal *Database* a QGIS

Una volta completato il DB, è iniziato il processo di visualizzazione grafico-spaziale dei dati. È stata condotta una ricerca topografica e geografica per identificare il corretto posizionamento dei siti nel catalogo. Utilizzando il portale MAGOH (Anichini et al., 2023), sono stati scaricati in formato JSON i siti già vettorializzati. Per quelli non identificabili, invece, sono stati utilizzati portali come Retore⁵ per l’analisi toponomastica e Castore⁶ per verificare la presenza di siti nel Catasto Leopoldino del 1765. Inoltre, sono stati impiegati il DTM della provincia di Pisa e la carta idrica (SHP lineare) della Toscana, scaricati da Open Toscana⁷.

Il processo ha seguito un *workflow* metodologico che ha integrato strumenti GIS e analisi spaziali avanzate sviluppate in R. Un nuovo progetto è stato creato in QGIS 3.28.14, con il sistema di riferimento WGS84 (EPSG:4326). I dati da MAGOH sono stati importati in formato JSON, e le aree non identificate sono state vettorializzate manualmente tramite ricerca toponomastica, creando uno SHP poligonale per delimitare l’area di interesse. Durante la vettorializzazione è stato generato un campo attributi numerico chiamato ‘ID’ per semplificare l’associazione con il database. I file di MAGOH, ciascuno dotato di un ID univoco associato al relativo ritrovamento, sono stati importati singolarmente nel *software* e successivamente uniti tramite un’operazione di *merge* complessivo, includendo anche quelli vettorializzati manualmente nel *layer* denominato “poligonale_ID”. Infine, è stato importato il DB come file di testo delimitato. Grazie alla chiave primaria, è stato eseguito un processo di *join*, che ha consentito di combinare i due formati. Il risultato è stato esportato in un terzo file poligonale, denominato “poligonale_ritrov_finale”, corredato di tutte le informazioni presenti nel DB.

Questo processo ha permesso di visualizzare i siti presenti nel catalogo sia in forma grafica che descrittiva, migliorando notevolmente la comprensione e la fruibilità delle informazioni.

⁴ Questa operazione è stata applicata agli ID: 7, 10, 13, 18, 23, 33, 36, 38, 40, 44, 48, 55, 62, 64, 74.

⁵ https://www502.regione.toscana.it/searcherlite/retore_start.jsp

⁶ <https://www502.regione.toscana.it/castoreapp/>

⁷ <https://dati.toscana.it/dataset/>

2.3 Analisi spaziali in QGIS

Per condurre le analisi spaziali esplorative tramite *Point Pattern Analysis* (PPA) (Orton, 2004) in QGIS è stato necessario convertire i poligoni in centroidi, poiché gli algoritmi di *Kernel Density Estimation* (KDE) (Hu, 2025; Nakoinz & Knitter, 2018, pp. 77-81; Ripley, 2005, p. 152; Illian et al., 2008) richiedono di utilizzare solo elementi puntuali per calcolare distribuzioni e concentrazioni. La conversione è stata effettuata tramite il *plugin* ‘Centroidi’, che ha permesso di assegnare un elemento puntuale a ciascun ritrovamento analizzato.

2.3.1 Nearest Neighbor Index

Le analisi di PPA consistono in analisi statistiche volte a comprendere ed evidenziare eventuali tendenze spaziali in una distribuzione di punti, verificando, attraverso coefficienti, se questi siano raggruppati, dispersi in modo regolare o distribuiti casualmente (Basile & Campus, 2023; Gatrell et al., 1996; Kempf & Günther, 2023). Questa analisi è stata condotta utilizzando l'algoritmo *Nearest Neighbor Index* (NNI), reperibile nella Toolbox di Processing di QGIS, applicato al *layer* puntuale filtrato per ciascuna fase cronologica. Dopo aver configurato i parametri richiesti, sono stati calcolati i principali indicatori per definire la tipologia di distribuzione dei punti: la distanza media osservata, la distanza media attesa in condizioni di CSR (*complete spatial randomness*) e il corrispondente NNI, integrato dal test Z-Score (Moore et al., 2009).

2.3.2 Kernel Density Estimation

Dopo aver applicato i filtri per fase cronologica, è stata condotta un'analisi di concentrazione spaziale mediante l'algoritmo di KDE all'interno di QGIS. Per garantire un'elevata precisione, è stata adottata una funzione di kernel quartica, con una risoluzione dei pixel elevata e un raggio di influenza di 5500 m⁸. Quest'ultimo è stato determinato empiricamente dopo diverse prove, in quanto si è rivelato il più adatto a rappresentare la distribuzione dei ritrovamenti nell'area di studio. L'obiettivo principale dell'analisi era l'individuazione di *cluster* locali, motivo per cui si è preferito un valore di sigma relativamente basso. La scelta di questo parametro è stata effettuata in modo empirico, valutando il miglior compromesso tra il livello di dettaglio e la generalizzazione della distribuzione spaziale. L'analisi è stata anche parametrizzata utilizzando il campo ‘Affidabilità’, al fine di considerare il grado di certezza delle evidenze archeologiche. Tali analisi hanno consentito di mettere in luce le aree di maggiore concentrazione dei siti archeologici nelle diverse epoche cronologiche (villanoviana, orientalizzante, arcaica, classica ed ellenistica), tenendo conto dell'affidabilità dei rinvenimenti.

⁸ Inizialmente le analisi sono state effettuate in EPSG:4326 (coordinate geografiche in gradi), ma poiché gli strumenti di analisi spaziale in *spatstat* (R) (§2.4.3) richiedono un sistema metrico, i dati sono stati riproiettati in EPSG:3003. Per garantire la confrontabilità dei risultati, anche in QGIS le analisi KDE sono state eseguite con la stessa proiezione.

2.4 Analisi statistiche e spaziali in R

Per confermare o confutare le analisi condotte in QGIS e garantire un'interpretazione oggettiva, le stesse analisi sono state eseguite anche nel software R-4.4.2⁹. L'elaborazione ha coinvolto sempre l'uso di SHP puntuali, su cui sono state eseguite varie operazioni di analisi spaziale, tra cui:

- *Kernel Density Estimation*: per identificare aree di maggiore concentrazione dei ritrovamenti, pesando i dati in base all'affidabilità;
- *Nearest Neighbor Index*: per valutare la tendenza dei ritrovamenti a disporsi in maniera casuale, aggregata o regolare;
- Funzione K di Ripley: per valutare la densità cumulativa dei punti e per analizzare la presenza di aggregazione o regolarità dei ritrovamenti rispetto a un modello di CSR;
- Funzione G (*nearest neighbor distance distribution*): per misurare la distribuzione delle distanze minime tra ciascun ritrovamento e il suo vicino più prossimo;
- Funzione F (*empty space function*): per analizzare la distribuzione delle distanze del ritrovamento più vicino a un punto scelto casualmente all'interno del campo;
- Funzione J: derivata dalle funzioni F e G, per aiutare a interpretare se i *pattern* qualora identificati siano aggregati, casuali o regolari.

2.4.1 Pre-elaborazione dei dati spaziali

I dati, vettorializzati in QGIS, sono stati importati in R con il pacchetto *sf*, verificando la presenza di attributi chiave (coordinate, altimetria, affidabilità). In una fase preliminare, i dati originariamente in EPSG:4326 sono stati convertiti nel sistema metrico EPSG:3003 (Monte Mario/Italy zone 1), necessario per garantire la corretta interpretazione delle distanze spaziali nelle analisi condotte con il pacchetto *spatstat*. Successivamente, per integrare il dato altimetrico, è stato utilizzato in QGIS l'algoritmo "Contour", applicato al DTM del comune di Pisa, associando a ciascun punto il valore della curva di livello più vicina.

Una volta strutturati i dati, i punti sono stati trasformati in oggetti di tipo PPP, formato richiesto per analisi spaziali quali NNI, KDE e le funzioni K, G, F e J. È stata quindi condotta una fase di normalizzazione e pulizia, con gestione dei valori mancanti e attribuzione di un peso ai ritrovamenti sulla base del loro grado di affidabilità. Infine, è stata definita una finestra spaziale unica per l'intero *dataset*, calcolando il *bounding box* globale e ampliandolo di 3000 m, al fine di ridurre gli effetti di bordo nelle analisi.

2.4.2 Calcolo dell'NNI

Per ogni periodo considerato, i dati sono stati letti in R tramite le librerie *sf*, *spatstat* e *spatstat.geom*. I singoli SHP sono stati importati e convertiti in formato *sf*, dal quale sono state estratte le coordinate (x, y) necessarie per le analisi spaziali. Successivamente, per ciascun insieme di punti è stata definita la 'finestra di osservazione' (*window*) corrispondente al *bounding box* dei dati, e attraverso la funzione PPP è stato creato l'oggetto PPP.

⁹ <https://cran.r-project.org/bin/windows/base/>

Una volta definito il *pattern* di punti, sono state calcolate le distanze dal NNI per ciascun punto (funzione *nn-dist()*), ottenendo la distanza media osservata (*obs_mean*).

Per stimare la densità puntuale, è stata calcolata l'area del *bounding box*, che ha fornito la stima di λ . La distanza media attesa in condizioni di completa casualità spaziale è stata determinata come: $\frac{1}{2\sqrt{\rho}}$ (Clark & Evans, 1954, p. 451).

Infine, il NNI è stato derivato come rapporto tra la distanza media osservata e quella attesa. I valori di NNI inferiori a 1 indicano una tendenza all'aggregazione (*clustering*), i valori intorno a 1 suggeriscono una distribuzione casuale, mentre i valori superiori a 1 denotano una disposizione dei punti più uniforme. Per ciascun periodo, i risultati (numero di punti, area, distanze osservate e attese e relativo NNI) sono stati salvati e stampati, consentendo così un confronto diretto tra i diversi periodi.

2.4.3 Calcolo della *Kernel Density Estimation*

Nelle analisi di KDE (Venables & Ripley, 2002), i dati sono stati ponderati in base all'affidabilità dei ritrovamenti. Inoltre, quest'ultima è stata normalizzata su un intervallo [1, peso], assegnando i valori minimi e massimi rispettivamente ai ritrovamenti meno e più affidabili, evitando così l'assegnazione di valori negativi. La scelta della banda di *smoothing* (σ) è stata ottimizzata testando valori compresi tra 2000 e 5000 m, con incrementi di 500 m, per bilanciare dettaglio locale e generalizzazione spaziale, al fine di ottenere una rappresentazione della densità spaziale il più accurata possibile. In aggiunta, l'affidabilità è stata normalizzata su una scala applicando un peso¹⁰ maggiore ai punti più affidabili. Questo ha migliorato la precisione della KDE, riducendo l'impatto dei dati meno affidabili e ha consentito di focalizzarsi sulle aree "più sicure", migliorando ulteriormente l'accuratezza della densità spaziale stimata. Le mappe KDE sono state generate per ogni periodo, valore di σ e peso di affidabilità, e visualizzate utilizzando il pacchetto *ggplot2* (Wickham, 2006), con una scala cromatica progressiva. I risultati sono stati esportati in formato PNG per consentire un confronto visivo tra le diverse combinazioni di parametri.

2.4.4 Analisi integrata delle funzioni spaziali ponderate (K di Ripley, G, F, J) per l'affidabilità dei ritrovamenti

Per approfondire la comprensione delle caratteristiche della distribuzione spaziale dei ritrovamenti archeologici, i dati originali in formato *shapefile* sono stati importati e gestiti con la libreria *sf*. Successivamente, le coordinate spaziali (x, y) sono state estratte per creare oggetti di classe PPP, integrando finestra spaziale predeterminata definita per assicurare uniformità e comparabilità tra i vari periodi storici. Poiché ciascun punto corrisponde a un ritrovamento archeologico caratterizzato da un livello variabile di affidabilità, è stato indispensabile, anche in questo caso, attribuire a ogni singolo punto un peso specifico. Pertanto, è stato applicato un processo di normalizzazione dei valori di affidabilità, scalati nell'intervallo [0,1]. Successivamente, tramite un parametro definito "peso_values"¹¹, è stato possibile modificare la scala di questi

¹⁰ `c(0, 0.1, 0.2, 0.3, 0.4, 0.5, 0.6, 0.7, 0.8, 0.9, 1, 2, 3, 4, 5, 6, 7, 8, 9)`.

¹¹ `Peso_values <- c(0, 0.1, 0.2, 0.3, 0.4, 0.5, 0.6, 0.7, 0.8, 0.9, 1, 2, 3, 4, 5, 6, 7, 8, 9)`.

valori normalizzati, producendo pesi ponderali (*marks*) capaci di enfatizzare in modo differenziale i ritrovamenti più affidabili.

L'analisi è stata svolta implementando congiuntamente quattro differenti funzioni spaziali fondamentali: G, F, J e K di Ripley (Diggle, 2013; Baddeley et al., 2016) ciascuna con specifiche configurazioni di calcolo. In dettaglio, per ogni periodo e valore di peso considerato, la funzione G(r) è stata stimata mediante la funzione *Gest()* del pacchetto *spatstat*, applicando una correzione per i bordi (*border*) per ridurre distorsioni dovute agli effetti di margine. Analogamente, la funzione F(r) è stata calcolata con la funzione *Fest()*, mentre la funzione J(r), derivata dalla relazione tra le funzioni precedenti, è stata stimata utilizzando la funzione *Jest()*. Infine, la funzione K di Ripley (K(r)) è stata determinata tramite la funzione *Kest()* con correzione di tipo Ripley. Le funzioni K e G sono state calcolate su intervalli di distanza compresi tra 0 e 5000 m, con incrementi regolari di 50 m, mentre le funzioni F e J sono state calcolate sempre con gli stessi intervalli di distanza ma con incrementi regolari di 20 m per garantire una risoluzione ottimale delle informazioni spaziali e una comparabilità accurata tra i diversi periodi storici e valori di peso considerati.

Per verificare la robustezza statistica dei risultati ottenuti e valutare in modo accurato eventuali deviazioni dalla casualità spaziale, sono stati generati *envelope* di confidenza mediante simulazioni Monte Carlo con un numero variabile di iterazioni (99, 1000 e 5000 simulazioni). Tale strategia metodologica consente, infatti, di ottenere intervalli di confidenza progressivamente più stabili e precisi all'aumentare del numero di simulazioni, riducendo l'incertezza legata al campionamento casuale del processo di riferimento (CSR). Ad esempio, con sole 99 simulazioni si ottengono bande di confidenza ampie, il che rende meno affidabile l'individuazione di *pattern* significativi. Aumentando il numero di iterazioni a 1000 o 5000, tali bande si restringono sensibilmente, consentendo una maggiore precisione nell'identificazione di fenomeni spaziali non casuali. In aggiunta, la scelta di un passo di discretizzazione delle distanze rappresenta un compromesso ottimale tra risoluzione spaziale e computazionale. Le funzioni K e G si basano su calcoli cumulativi o integrali, che generalmente risultano meno sensibili alla discretizzazione della variabile di distanza r. In altre parole, incrementi di 50 m sono sufficienti per ottenere stime stabili e accurate. Al contrario, le funzioni F e J si basano sulla distribuzione delle distanze minime (o sulla loro derivata) e sono molto più sensibili alla risoluzione del vettore r. Se il passo è troppo ampio (come 50 m), la griglia non è abbastanza fine per catturare tutte le variazioni locali, portando a stime imprecise o addirittura alla restituzione di valori NA. Per questi metodi è necessario utilizzare incrementi molto più piccoli (ad esempio, intorno a 20-25 m) per garantire una discretizzazione sufficientemente fine e ottenere una stima accurata.

Il confronto diretto tra le curve empiriche ottenute dai dati ponderati e gli *envelope* generati dalle simulazioni ha permesso una precisa identificazione dei fenomeni di aggregazione o dispersione spaziale, chiarendo le dinamiche spaziali intrinseche dei ritrovamenti archeologici a diverse scale di analisi. Infine, per facilitare una lettura efficace e immediata dei risultati, sono stati predisposti report interattivi in formato HTML, generati mediante le librerie *knitr* e *rmarkdown*. Questi documenti mostrano chiaramente le curve empiriche, le linee teoriche di riferimento e gli *envelope* di confidenza per ogni periodo e per ciascun parametro di peso analizzato, supportando così un'interpretazione chiara e una valutazione critica e approfondita dei *pattern* spaziali emersi dall'analisi.

3. Risultati

3.1 KDE in QGIS e R

Le analisi di KDE condotte in QGIS (**fig. 2**) rivelano come, nella fase villanoviana, i ritrovamenti archeologici si concentrino principalmente nelle aree collinari di Peccioli (Podere Ortaglia, ID 38) e Palaia (Montacchita, ID 44), con altitudini comprese tra 150 e 190 m s.l.m. In contrasto, le aree di bassa quota come Pontedera (Giuncaiola, ID 21) a 20 m s.l.m., mostrano una distribuzione più rarefatta. Durante il periodo orientalizzante, Palaia e Ponsacco (Podere S. Eugenio, ID 23) mantengono una buona densità di concentrazione, alle quali si aggiunge l'area di Terricciola dove si registra una crescita significativa e Montopoli in Val d'Arno, con particolare attenzione alla zona di Marti (160 m s.l.m.) che inizia ad emergere con una densità moderata. Nel periodo arcaico, Peccioli, Palaia e Terricciola continuano a dominare la distribuzione collinare mentre l'area di Pontedera, quella di bassa quota. Inoltre, emergono nuove aree di frequentazione nella zona di Montopoli in Val d'Arno e più a nord nella zona di Castelfranco di Sotto. Nel periodo classico si confermano le aree di Peccioli, Palaia, Terricciola e Pontedera che mantengono un'elevata densità di ritrovamenti, mentre Montopoli in Val d'Arno registra una presenza più moderata ma costante. Infine, nel periodo ellenistico, pur continuando a essere predominanti le aree collinari, si osserva un cambiamento nella distribuzione con una maggiore frammentazione e un'espansione verso le pianure, soprattutto nell'area geografica compresa tra Montopoli in Val d'Arno e Terricciola.

L'analisi spaziale realizzata in R, mediante la KDE (**fig. 3**), adottando un valore di sigma ottimizzato attraverso prove iterative e validato dal calcolo della funzione G a 4400 m, ha evidenziato una trasformazione progressiva e significativa nei modelli di distribuzione dei rinvenimenti.

Nel periodo villanoviano, i ritrovamenti sono maggiormente concentrati nelle aree collinari intermedie (150-170 m s.l.m.), con località come Ghizzano e Montacchita che rappresentano i poli principali. Tuttavia, c'è una presenza significativa di reperti anche nelle aree pianeggianti di Pontedera e Ponsacco (20 m s.l.m.) che mantengono in tutti i periodi una forte densità. Le altitudini più elevate, come a Parlascio (290 m s.l.m.), registrano invece una scarsa densità di ritrovamenti, indicando una minore preferenza per le aree di quota molto elevata. Durante il periodo orientalizzante, la concentrazione si mantiene stabile, con un'elevata densità nei bassopiani, ma si assiste a una crescita più marcata delle zone collinari, in particolare a Marti e Terricciola (160-170 m s.l.m.). Nel periodo arcaico, la distribuzione dei ritrovamenti si stabilizza nelle aree pianeggianti, ma si nota un rafforzamento della presenza nelle colline tra Marti, Peccioli e Terricciola (140-190 m s.l.m.). Durante il periodo classico, si osserva un ulteriore consolidamento nelle colline, con l'area tra Terricciola e Ceppato (140-170 m s.l.m.) che registra una delle più alte densità, mentre le altitudini oltre i 250 m s.l.m. rimangono scarsamente occupate. Nel periodo ellenistico, la tendenza alla crescita della densità nelle colline continua, con una riduzione della presenza nei bassopiani. Le aree principali di concentrazione si trovano a Pontedera e in tutta l'area tra Marti e Peccioli.

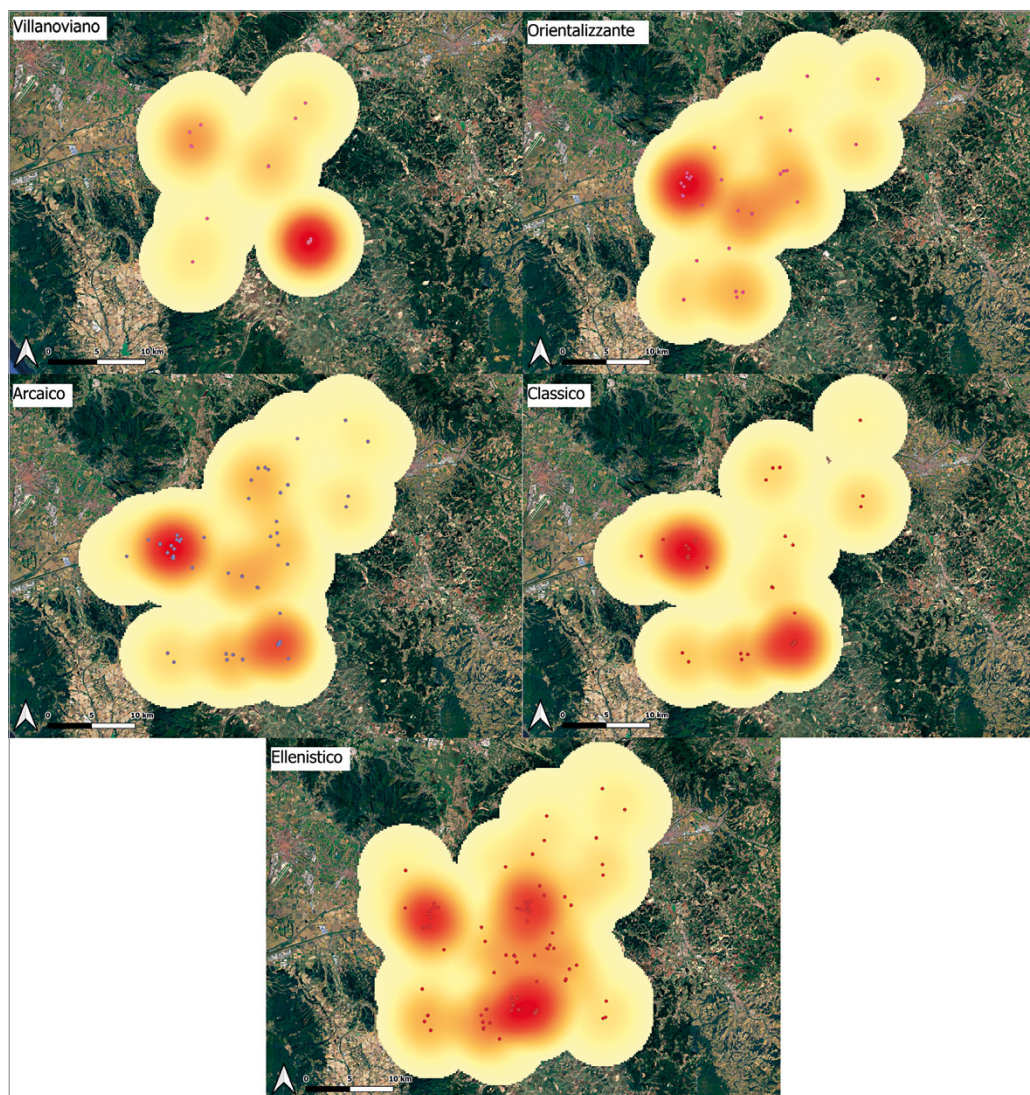


fig. 2. Analisi di *Kernel Density Estimation* (KDE) realizzata tramite QGIS, mostrando la distribuzione spaziale dei ritrovamenti archeologici suddivisi per periodi cronologici (villanoviano, orientalizzante, arcaico, classico, ellenistico).

3.2 Analisi del NNI in ambiente *R* e GIS

L'analisi spaziale dei ritrovamenti archeologici è stata approfondita mediante il calcolo del NNI, sia in ambiente GIS che tramite linguaggio *R*. I risultati delle analisi condotte in ambiente GIS sono sintetizzati nella Tabella 1 (tab. 1).

Parallelamente, l'analisi effettuata in ambiente *R* ha confermato pienamente i risultati ottenuti con QGIS (tab. 2), supportando ulteriormente la tendenza alla formazione di *cluster*.

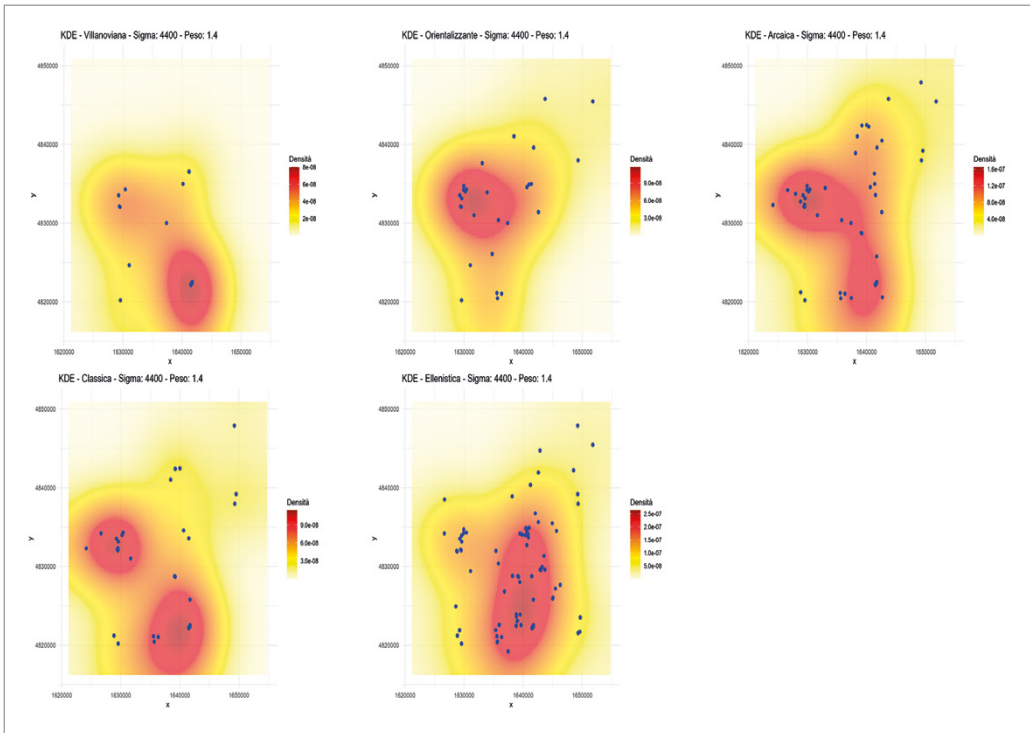


fig. 3. Analisi di *Kernel Density Estimation* (KDE) realizzata tramite R, mostrando la distribuzione spaziale dei ritrovamenti archeologici suddivisi per periodi cronologici (villanoviano, orientalizzante, arcaico, classico, ellenistico).

La forte concordanza dei risultati ottenuti dai due strumenti metodologici ne rafforza ulteriormente la solidità. In particolare, il periodo villanoviano evidenzia un *clustering* moderato e statisticamente significativo, mentre il periodo orientalizzante presenta una distribuzione quasi casuale, con un livello di aggregazione inferiore. Nei periodi arcaico e classico, invece, si osserva un'intensificazione marcata del *clustering*, mentre il periodo ellenistico mostra una graduale attenuazione dell'aggregazione.

3.3 Analisi funzioni spaziali K Ripley, G, F, J

L'analisi spaziale dei ritrovamenti archeologici è stata approfondita adottando un approccio integrato, basato sulle funzioni spaziali K di Ripley, G, F e J implementate in R. L'obiettivo principale era esplorare la distribuzione spaziale dei siti archeologici, tenendo conto del grado di affidabilità delle informazioni. Durante la fase preliminare, in particolare nell'applicazione della funzione G corredata di *envelope* ottenuti da 5000 simulazioni Monte Carlo, è emersa una soglia critica di circa 4400 m. Oltre tale distanza, la funzione G iniziava infatti a restituire valori NA, segnalando che le stime non erano più affidabili. Questo fenomeno, lungi dall'essere un mero "errore tecnico", fornisce un'informazione metodologica preziosa: i 4400 m rappresentano presumibilmente il limite ottimale per analizzare la distribuzione dei ritrovamenti in

Periodo	Numero di punti	Distanza media osservata (m)	Distanza media attesa CSR (m)	NNI	Z-Score
Villanoviano	16	1037.426	1783.6867	0.5816	-3.2015
Orientalizzante	30	1985.373	2191.3849	0.9059	-0.9850
Arcaico	56	929.279	1847.9182	0.5028	-7.1168
Classico	35	997.654	2240.0758	0.4453	-6.2772
Ellenistico	85	919.185	1455.848	0.6313	-6.5016

tab. 1. Risultati dettagliati dell’analisi *Nearest Neighbor Index* (NNI) condotta in QGIS, suddivisi per periodo.

Periodo	Numero di punti	Distanza media osservata (m)	Distanza media attesa CSR (m)	NNI
Villanoviano	16	1037.24	1783.68	0.582
Orientalizzante	30	1985.05	2191.38	0.906
Arcaico	56	929.13	1847.92	0.503
Classico	35	997.50	2240.08	0.445
Ellenistico	85	919.04	1455.85	0.631

tab. 2. Risultati dettagliati dell’analisi *Nearest Neighbor Index* (NNI) condotta in R, suddivisi per periodo storico, con comparazione dei valori di NNI per ciascun periodo.

modo statisticamente robusto. Un ulteriore e rilevante aspetto riguarda l’incremento utilizzato nel vettore di *r* per le funzioni come *Fest()*, che richiedono una discretizzazione sufficientemente fine. È stato constatato come uno *step* di 50 m risulti troppo ampio, causando una griglia eccessivamente grossolana e la comparsa di valori NA. Riducendo l’incremento a 20 m, è stato invece possibile catturare con maggiore accuratezza le variazioni locali, specialmente nelle funzioni *F* e *J*. Per quanto riguarda l’assegnazione dei pesi nel campo “Affidabilità”, i valori compresi tra 0,9 e 2 hanno mostrato il miglior equilibrio tra la capacità di distinguere correttamente i livelli di affidabilità e l’evitare distorsioni. Pesi troppo bassi avrebbero di fatto annullato la capacità di differenziare i singoli dati, mentre pesi elevati avrebbero enfatizzato in modo eccessivo un numero ristretto di punti, compromettendo la robustezza delle analisi.

Per definire un unico valore da impiegare in maniera uniforme, è stata quindi calcolata una media dei pesi più rilevanti, adottata come peso unico per l’intero *set* di funzioni. Questa scelta garantisce una coerenza metodologica nei confronti di tutte le analisi successive, evitando di sovra- o sotto- rappresentare specifici ritrovamenti e consentendo una lettura più equilibrata dei dati. Inoltre, l’impiego di 5000 simulazioni Monte Carlo per il calcolo degli *envelope* ha permesso di ottenere stime di confidenza molto più stabili rispetto alle 99 o 1000 simulazioni, consentendo di individuare variazioni anche minime con una maggiore sicurezza statistica.

I risultati delle analisi spaziali evidenziano, in tutti i periodi considerati (villanoviano, orientalizzante, arcaico, classico, ellenistico), un marcato *pattern* di *clustering*. In particolare, la funzione *K* (fig. 4) resta al di sopra della curva CSR su tutte le distanze

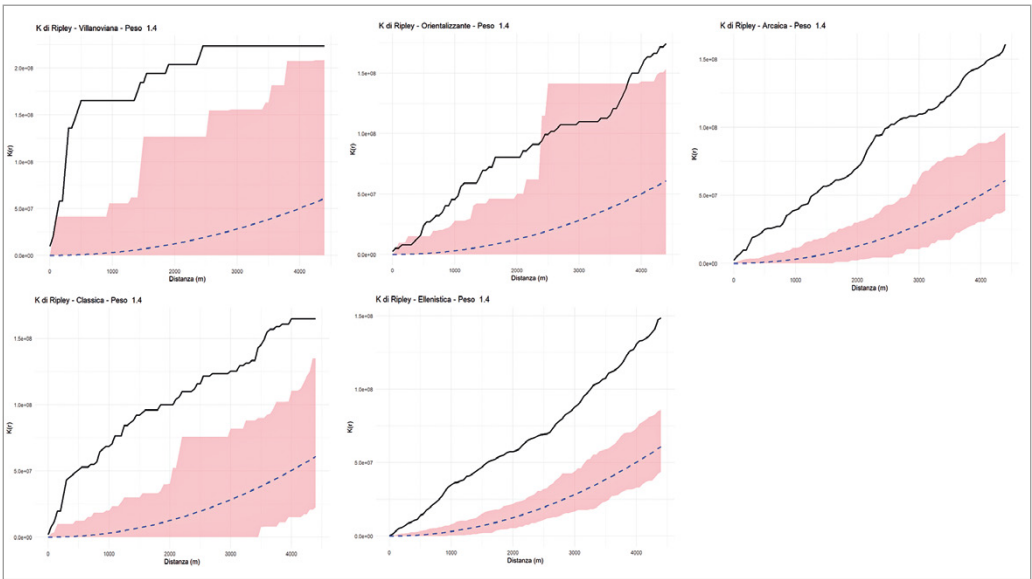


fig. 4. Visualizzazione della funzione K di Ripley per i periodi villanoviano, orientalizzante, arcaico, classico ed ellenistico, ottenuta tramite analisi in R, con assegnazione di un peso pari a 1,4 e 5000 simulazioni Monte Carlo.

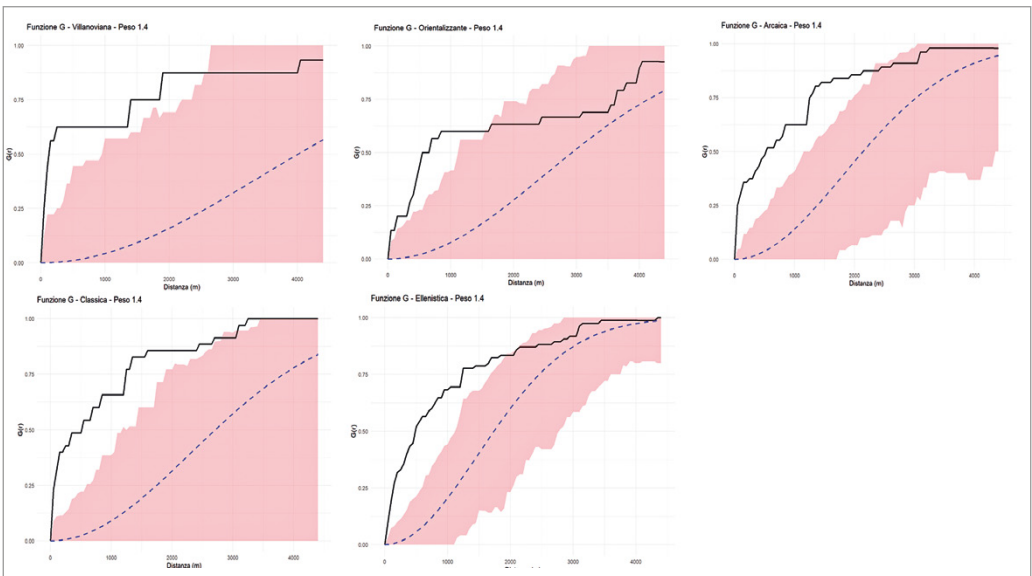


fig. 5. Visualizzazione della funzione G per i periodi villanoviano, orientalizzante, arcaico, classico ed ellenistico, ottenuta tramite analisi in R, con assegnazione di un peso pari a 1,4 e 5000 simulazioni Monte Carlo.

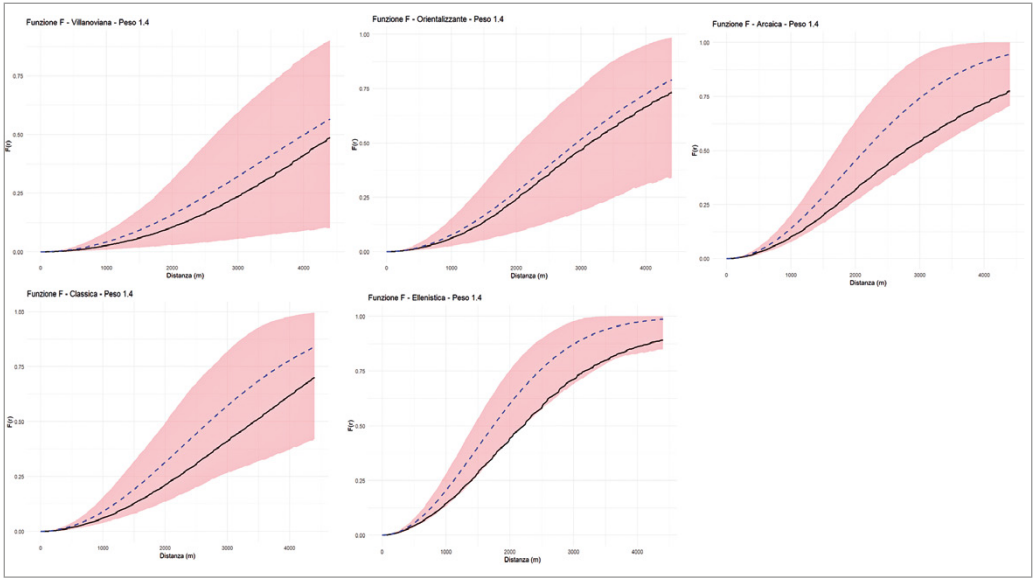


fig. 6. Visualizzazione della funzione F per i periodi villanoviano, orientalizzante, arcaico, classico ed ellenistico, ottenuta tramite analisi in R, con assegnazione di un peso pari a 1,4 e 5000 simulazioni Monte Carlo.

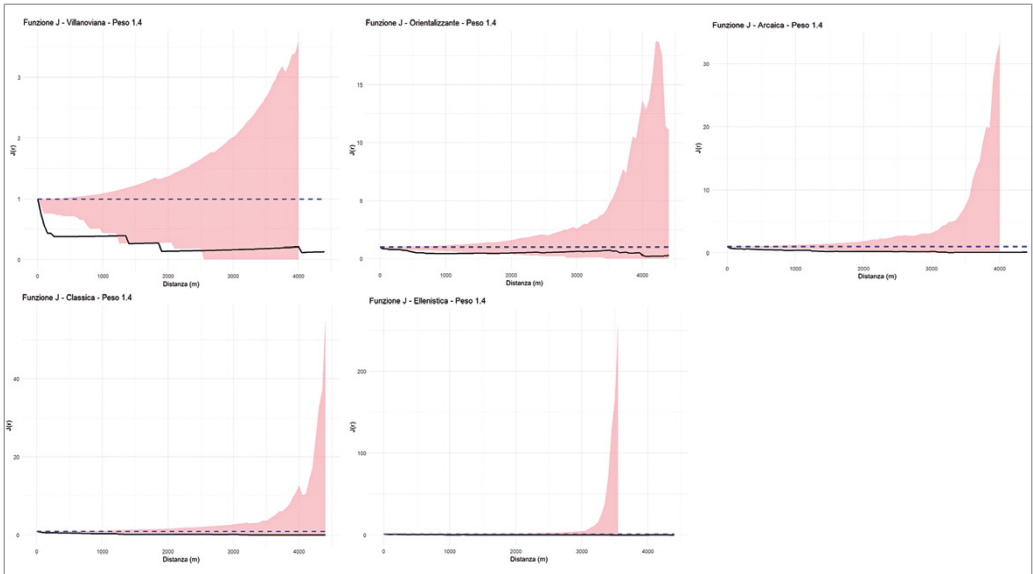


fig. 7. Visualizzazione della funzione J per i periodi villanoviano, orientalizzante, arcaico, classico ed ellenistico, ottenuta tramite analisi in R, con assegnazione di un peso pari a 1,4 e 5000 simulazioni Monte Carlo.

analizzate, confermando un'elevata densità di siti e un addensamento su scala locale e media. La funzione G (fig. 5) mostra un rapido aumento della curva, superando la curva CSR a distanze relativamente brevi e indicando che la maggior parte dei siti ha vicini molto più prossimi di quanto previsto da una distribuzione casuale, un segno di *clustering* significativo, soprattutto su scala piccola e media (fino a circa 2000 m). La funzione F (fig. 6), mantenendosi sotto la curva CSR, suggerisce che, nonostante il *clustering*, esistano ampie aree vuote lontane dai nuclei di ritrovamenti, indicando che questi tendono a concentrarsi in alcune zone specifiche. La funzione J (fig. 7), infine, rimane sotto il valore 1 per quasi tutto l'intervallo di distanze, confermando il pattern aggregato e l'assenza di una distribuzione regolare dei siti.

4. Discussione dei risultati

La presente ricerca mette in luce l'importanza cruciale della gestione e valorizzazione dei *legacy data* per una solida analisi spaziale, nonostante le sfide poste dalla natura eterogenea e spesso incompleta dei dati archeologici. La digitalizzazione e sistematizzazione dei dati provenienti da fonti bibliografiche, inizialmente privi di coordinate precise, ha richiesto un approccio metodologico rigoroso basato su tecniche avanzate di vettorializzazione e georeferenziazione. L'integrazione di strumenti GIS e analisi spaziali esplorative come la KDE, il NNI e le funzioni spaziali K di Ripley, F, G e J hanno consentito di identificare modelli di distribuzione e tendenze insediative significative nel medio Valdarno tra il periodo villanoviano e quello ellenistico.

Uno degli obiettivi principali del progetto è stato quello di trasformare un catalogo bibliografico in un DB strutturato e georeferenziato, in grado di supportare analisi spaziali dettagliate. La scelta di digitalizzare i dati mediante Microsoft Excel e successivamente integrarli in QGIS, garantendo l'interoperabilità attraverso l'utilizzo di vocabolari standardizzati (come quelli provenienti dai portali MAGOH e GNA), ha reso i dati uniformi e facilmente integrabili con ulteriori risorse digitali, facilitando aggiornamenti futuri e confronti interdisciplinari. Questa strategia ha consentito di preservare l'integrità dei dati, nonostante la presenza di problematiche inerenti alla natura dei *legacy data*, soggetti a interpretazioni e possibili alterazioni nel processo di digitalizzazione (Huggett, 2022).

Le analisi spaziali, condotte con strumenti complementari come QGIS e R, hanno fornito un quadro dettagliato della distribuzione dei ritrovamenti archeologici attraverso le diverse fasi cronologiche, suggerendo possibili dinamiche di insediamento nel medio Valdarno. L'analisi di KDE ha evidenziato una notevole variabilità nei periodi storici: nel periodo villanoviano, i ritrovamenti si concentrano principalmente in aree collinari come Peccioli (Podere Ortaglia, 150-190 m s.l.m.) e Palaia (Montacchita, 160-190 m s.l.m.), il che potrebbe indicare un insediamento strategico, probabilmente per esigenze difensive e di controllo territoriale. Nel periodo orientalizzante, si osserva un'espansione dei ritrovamenti verso nuove aree come Terricciola e Montopoli in Val d'Arno, suggerendo una crescita della densità insediativa e lo sviluppo di strutture più permanenti, che potrebbe riflettere una maggiore stabilità e intensificazione delle attività economiche e di traffici (Alberti et al., 2004, pp. 7-8). Nel periodo arcaico,

emergono nuovi siti di interesse, con attività agricole e commerciali più intense, in aree come Pineta del Maltufo, Montopoli in Val d'Arno e Castelfranco di Sotto. Tuttavia, alcuni siti sembrano essere stati abbandonati, forse a causa di fattori ecologici o altri eventi che ne hanno influito sull'abitabilità (Ciampoltrini, 2010, pp. 135-144). Nel periodo classico, la distribuzione dei ritrovamenti diventa più equilibrata tra alture e pianure fertili, con un incremento nelle attività agricole e commerciali e una maggiore attenzione al controllo delle vie fluviali, come dimostrano i ritrovamenti di anfore greco-italiche. Infine, nel periodo ellenistico, si registra un'espansione significativa dei ritrovamenti verso le aree di bassa quota, come Pontedera e Ponsacco, suggerendo un cambiamento nelle dinamiche insediative, con un'accentuata attenzione alle risorse fluviali e alle pianure fertili (Bruni, 2008, pp. 567-569; Ciampoltrini, 2014, pp. 37-40).

Il caso specifico dell'area di Pontedera rappresenta una problematica particolare, caratterizzata da un'elevata concentrazione di reperti anche nei periodi più antichi, nonostante la bassa altitudine. Ciò suggerisce l'influenza di fattori locali specifici, forse legati al controllo delle rotte fluviali o alla disponibilità di risorse idriche.

Anche l'applicazione dell'algoritmo NNI ha evidenziato una tendenza generale al *clustering* nei periodi analizzati, con valori inferiori a 1. Questo comportamento, confermato sia in QGIS che in *R*, indica che i ritrovamenti non sono distribuiti casualmente, ma si concentrano in aree specifiche, suggerendo insediamenti strategicamente organizzati, in particolare nei periodi arcaico e classico.

Dinamica confermata dall'analisi integrata delle funzioni K, G, F, e J dei vari periodi storici che ha rivelato un pattern di *clustering* evidente, variabile in base alle distanze e al periodo considerato. Nel periodo villanoviano, il *clustering* è particolarmente marcato a distanze inferiori ai 2000 m, mentre durante l'orientalizzante, i siti tendono a raggrupparsi con distanze tra i vicini spesso inferiori ai 1000 m, pur mostrando un eccesso di densità fino a 3000-4000 m. Nel periodo arcaico, il *pattern* di *clustering* persiste, con una concentrazione maggiore tra i 2000 e i 2500 m. Nel periodo classico, la distribuzione continua a mostrare un raggruppamento compatto, con distanze medie contenute tra i siti. Infine, nel periodo ellenistico, il fenomeno si mantiene forte su scala locale (fino a circa 2000 m) e si estende anche a distanze maggiori, con una densità superiore rispetto a un modello casuale.

In sintesi, le quattro funzioni analizzate confermano un *pattern* di *clustering* costante in tutti i periodi. Le funzioni G e K risultano superiori alla curva CSR, evidenziando distanze più brevi tra i siti e una maggiore densità locale rispetto a una distribuzione casuale. La funzione J, con valori inferiori a 1, suggerisce una distribuzione aggregata piuttosto che regolare, mentre la funzione F mostra valori sotto la curva CSR, indicando la presenza di ampie aree prive di insediamenti e confermando la concentrazione dei siti in nuclei distinti. L'analisi delle distanze mostra che già entro i primi 1000-2000 m i siti sono significativamente più vicini rispetto a una distribuzione casuale, con il *clustering* che si estende fino a circa 3000-4000 m, pur con differenze tra i vari periodi.

Nonostante i risultati ottenuti offrano una visione dettagliata della distribuzione spaziale nel medio Valdarno, emergono alcune criticità. La mancanza di coordinate storiche precise ha richiesto metodologie di georeferenziazione che, sebbene efficaci, introducono margini di errore interpretativo. Ulteriori sviluppi potrebbero perfezionare tali tecniche, integrando dati provenienti da altre fonti storiche e cartografiche.

Inoltre, l'aggiornamento continuo del DB con nuove scoperte, insieme all'utilizzo di tecnologie digitali avanzate, risulta essenziale per migliorare ulteriormente la veridicità delle analisi spaziali. L'adozione di modelli statistici più avanzati, inclusi approcci di *machine learning* per l'identificazione di *pattern*, potrebbe fornire nuove chiavi di lettura delle dinamiche insediative del medio Valdarno.

5. Conclusioni

Questo studio ha permesso di fornire una rappresentazione grafica e spaziale dei dati raccolti, restituendo un possibile quadro dell'evoluzione della distribuzione dei ritrovamenti archeologici nell'area del medio Valdarno, attraverso un approccio metodologico integrato che unisce tecniche GIS con analisi statistico-spaziali avanzate. L'applicazione delle metodologie di KDE, NNI e le funzioni spaziali (K di Ripley, G, F, J) ha confermato un marcato *pattern* di *clustering* dei ritrovamenti, evidenziando variazioni significative nelle distanze tra i siti e nelle aree di maggiore concentrazione nei diversi periodi storici.

I risultati hanno rivelato un progressivo consolidamento e spostamento dei ritrovamenti: inizialmente concentrati in aree collinari durante il periodo villanoviano, si espandono gradualmente verso le pianure nei periodi successivi, con una densità crescente in epoca ellenistica. L'analisi delle distanze ha inoltre suggerito che, in tutti i periodi, i ritrovamenti risultano essere più ravvicinati rispetto a una distribuzione casuale, indicando una possibile organizzazione territoriale strategica influenzata da fattori geografici e ambientali. Un dato di particolare interesse emerge dai siti di Pontedera e Ponsacco, che mostrano una concentrazione elevata di ritrovamenti a bassa altitudine anche nei periodi più antichi. Questo fenomeno potrebbe suggerire dinamiche specifiche, probabilmente di natura economica, sociale o ambientale, che meritano ulteriori approfondimenti.

In conclusione, sebbene l'uso dei *legacy data* presenti alcune limitazioni, come incertezze derivanti dalla mancanza di coordinate precise e potenziali errori interpretativi (§1), i processi di digitalizzazione, georeferenziazione e analisi hanno consentito di valorizzare efficacemente le informazioni disponibili. Questo ha permesso di ottenere risultati significativi nello studio della distribuzione dei ritrovamenti archeologici nel medio Valdarno.

Per il futuro, sarà fondamentale continuare ad aggiornare il DB con nuove scoperte e migliorare le metodologie analitiche. Inoltre, l'integrazione del DB in un sistema web-GIS interattivo permetterebbe una più ampia diffusione dei dati, facilitando l'accesso e la consultazione da parte di studiosi e ricercatori, e favorendo nuove collaborazioni per approfondire ulteriormente il tema del popolamento dell'area.

Bibliografia

- Alberti, A., Baldassari, M., Bisio, E., & Ciampoltrini, G. (2004). *Dai villaggi etruschi ai castelli medievali: Palaia e il suo territorio nei secoli*. Bandecchi & Vivaldi.
- Anichini, F., Gattiglia, G., & Saponara, A. R. (2023). MAGOH: Un nuovo strumento per la gestione e la consultazione dei dati archeologici del Nord della Toscana. *Archeologia e Calcolatori*, 34(2), 277-296.

- Baddeley, A., Rubak, E., & Turner, R. (2016). *Spatial point patterns: Methodology and applications with R* (Vol. 1). CRC press.
- Basile, S., & Campus, A. (2023). Integrating Point Pattern Analysis and Logistic Regression approaches for exploring the settlement pattern of the Versilia and Garfagnana mountains in Roman times. *Archeologia e Calcolatori*, 34(2), 87-104.
- Bruni, S. (2008). Località Le Serre, Podere Ortaglia: Scavi nel santuario etrusco. *Notiziario della Soprintendenza per i Beni Archeologici della Toscana*, 3(2007), 496-711.
- Casarotto, A., Pelgrom, J., & Stek, T. D. (2019). A systematic GIS-based analysis of settlement developments in the landscape of Venusia in the Hellenistic-Roman period. *Archaeological and Anthropological Sciences*, 11, 735-753.
- Ciampoltrini, G. (2010). Edilizia rurale tra Valdarno e Valle del Serchio: la colonizzazione etrusca tra VI e V secolo a.C. e le deduzioni coloniali d'età tardorepubblicana. In M. Bentz & Ch. Reusser (eds.), *Etruskisch-italische und römisch-republikanische Häuser* (pp. 135-144). Reichert Verlag.
- Ciampoltrini, G. (2014). *Gli etruschi della bassa Valdera tra Pisa e Volterra. Prolegomeni all'edizione dello scavo della Giuncaiola di Pontedera* (2011-2012). Tipografia La Grafica Pisana.
- Ciampoltrini, G., & Grassini, F. (eds.). (2011). *Gli etruschi e le vie d'acqua. L'insediamento della Granchiaia fra Montopoli in Val d'Arno e Marti*. Edizioni ETS.
- Clark, P. J., & Evans, F. C. (1954). Distance to Nearest Neighbor as a Measure of Spatial Relationships in Populations. *Ecology*, 35(4), 445-453. <https://doi.org/10.2307/1931034>
- Diggle, P. J. (2013). *Statistical analysis of spatial and spatio-temporal point patterns*. CRC press.
- Fabiani, F., & Gattiglia, G. (2012). La struttura informatica di archiviazione. In Anichini, F. Fabiani, G. Gattiglia, & M. L. Gualandi (eds.). *MAPPA. Metodologie applicate alla predittività del potenziale archeologico. Volume I* (pp. 41-71). Nuova Cultura.
- Gatrell, A. C., Bailey, T. C., Diggle, P. J., & Rowlingson, B. S. (1996). Spatial point pattern analysis and its application in geographical epidemiology. *Transactions of the Institute of British Geographers*, 21(4), 256-274.
- Hu, H. (2025). Spatial synergy between cultural heritage and metro networks: A case study of distribution patterns and value assessment in Beijing. *Sustainability*, 17(4), 1666.
- Huggett, J. (2022). Data legacies, epistemic anxieties, and digital imaginaries in archaeology. *Digital*, 2(2), 267-295.
- Illian, J., Penttinen, A., Stoyan, H., & Stoyan, D. (2008). *Statistical analysis and modelling of spatial point patterns*. John Wiley & Sons.
- Kempf, M., & Günther, G. (2023). Point pattern and spatial analyses using archaeological and environmental data: A case study from the Neolithic Carpathian Basin. *Journal of Archaeological Science: Reports*, 47, 103747.
- Moore, D.S., McCabe, G.P., & Craig, B.A. (2009). *Introduction to the practice of statistics* (Vol. 4). W. H. Freeman.
- Nakoinz O., & Knitter D. (2018). *Modelling human behaviour in landscapes: Basic concepts and modelling elements*. Springer.
- Orton, C. (2004). Point pattern analysis revisited. *Archeologia e Calcolatori*, 15, 299-315.
- Ripley, B. D. (2005). *Spatial statistics*. John Wiley & Sons.
- Squarcini Berti, F. (2022). *Il popolamento del medio Valdarno tra l'Età del Ferro e l'età ellenistica*. [Tesi di laurea non pubblicata]. Università di Pisa.
- Venables, W. N., & Ripley, B. D. (2002). *Modern applied statistics with S* (4th ed.). Springer.
- Wickham, H. (2006). An introduction to ggplot: An implementation of the grammar of graphics in R. *Statistics*, 1, 1-8.